

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Informatica e dell'Automazione



TESI DI LAUREA

**Progettazione e implementazione di una campagna di data science
per l'analisi del mercato degli smartphone**

**Design and implementation of a data science campaign for the
analysis of smartphone market**

Relatore

Prof. Domenico Ursino

Correlatore

Dott. Michele Marchetti

Candidato

Luca Busilacchi

ANNO ACCADEMICO 2025-2026

"Siamo annegati nelle informazioni, ma affamati di conoscenza."

John Naisbitt

Sommario

La presente tesi affronta l'analisi dei dati relativa all'andamento del mercato degli smartphone con l'obiettivo di individuare pattern significativi mediante dashboard descrittive, per l'analisi del mercato attuale, e dashboard predittive, per la previsione delle tendenze future. La tesi si focalizza sull'utilizzo di Power BI per l'elaborazione e la visualizzazione di dataset pubblici eterogenei, mediante lo sviluppo di dashboard interattive orientate all'analisi descrittiva e predittiva. È stata realizzata una prima fase di ETL (Extract, Transform and Load) con l'obiettivo di trattare i dati grezzi tramite la pulizia, la trasformazione e l'integrazione dei dati provenienti da fonti eterogenee. Particolare attenzione è stata rivolta, successivamente, alla progettazione di visualizzazioni efficaci, alla scrittura di formule DAX per l'analisi avanzata e all'impiego di tecniche di previsione (forecasting) per l'identificazione di trend significativi. Il lavoro svolto dimostra come l'integrazione di processi ETL, dashboard interattive e tecniche di analisi predittiva costituisca un efficace supporto ai processi decisionali, consentendo di trasformare dati eterogenei in informazioni utili per l'analisi strategica del mercato degli smartphone.

Keyword: Data Analytics, Business Intelligence, Power BI, ETL, Data Visualization, Predictive Analytics, Smartphone Market, Dashboard.

Introduzione	1
1 Introduzione alla Data Analytics	3
1.1 Introduzione al concetto di Data Science	3
1.1.1 Breve cenni storici	4
1.2 Big Data	5
1.3 Le 5 "V" dei Big Data	7
1.3.1 Volume	7
1.3.2 Velocità	7
1.3.3 Varietà	7
1.3.4 Veridicità	7
1.3.5 Valore	8
1.4 La varietà dei dati coinvolti nei Big Data	9
1.4.1 Dati strutturati	10
1.4.2 Dati non strutturati	11
1.4.3 Dati semi-strutturati	11
1.4.4 Metadati	12
1.5 Data Analytics	12
1.6 Tipologie della Data Analytics	14
1.6.1 Analisi descrittiva	14
1.6.2 Analisi diagnostica	14
1.6.3 Analisi predittiva	14
1.6.4 Analisi prescrittiva	15
1.7 Differenza tra Data Analytics e Data Analysis	15
1.8 Ciclo di vita della Big Data Analytics	16
1.8.1 Business Case Evaluation	17
1.8.2 Data Identification	17
1.8.3 Data Acquisition and Filtering	18
1.8.4 Data Extraction	19
1.8.5 Data Validation and Cleansing	19
1.8.6 Data Aggregation and Representation	20
1.8.7 Data Analysis	21
1.8.8 Data Visualization	22
1.8.9 Utilization of Analysis Results	22
1.9 Piramide DIKW	23

2	Power BI	28
2.1	Introduzione a Power BI	28
2.2	Infrastruttura di Power BI	29
2.2.1	Architettura	30
2.2.2	Flusso di lavoro	32
2.3	Visualizzazioni di Power BI	37
2.4	Tecnologie di programmazione integrate	41
3	Descrizione dei dataset di partenza	42
3.1	Dataset utilizzati	42
3.1.1	Dataset 1: Smartprix Smartphones Dataset	42
3.1.2	Dataset 2: MySmartPrice Mobile Dataset	43
3.2	Processo di ETL	44
3.2.1	Extract	46
3.2.2	Transform	46
3.2.3	Load	47
3.3	Differenza tra ETL e ELT	48
3.4	Strumenti ETL	49
3.5	Processo ETL applicato ai dataset	50
3.5.1	Extract	50
3.5.2	Transform	51
3.5.3	Load	56
4	Progettazione e realizzazione delle dashboard di base	59
4.1	Data Visualization	59
4.2	Data Visualization in Power BI	61
4.2.1	Grafici disponibili in Power BI	61
4.2.2	Misure DAX in Power BI	63
4.3	Misure DAX implementate	63
4.4	Descrizione e analisi delle dashboard di base realizzate	71
4.4.1	Dashboard 1 - Panoramica generale del mercato degli smartphone	72
4.4.2	Dashboard 2 - Analisi delle caratteristiche tecniche degli smartphone	74
4.4.3	Dashboard 3 - Analisi delle caratteristiche tecniche per fascia di prezzo	76
4.4.4	Dashboard 4 - Analisi dei sistemi operativi negli smartphone	77
4.4.5	Dashboard 5 - Analisi della soddisfazione degli utenti e del rapporto qualità/prezzo	79
5	Progettazione e realizzazione delle dashboard avanzate	86
5.1	Introduzione all'analisi predittiva	86
5.2	Analisi predittiva in Power BI	87
5.3	Introduzione all'analisi predittiva nel contesto del mercato degli smartphone	92
5.3.1	Dashboard 6 - Analisi predittiva dell'andamento dei prezzi e del numero di modelli di smartphone	93
5.3.2	Dashboard 7 - Analisi predittiva dell'evoluzione hardware degli smartphone	100
6	Discussione	103
6.1	Punti di Forza	103
6.1.1	Approccio completo di Data Analytics	103
6.1.2	Processo di ETL	103
6.1.3	Utilizzo di Power BI	104

6.1.4	Analisi di dashboard descrittive	104
6.1.5	Analisi di dashboard predittive	104
6.2	Limiti e vulnerabilità	104
6.2.1	Qualità dei dati	104
6.2.2	Integrazione complessa	105
6.2.3	Difficoltà di confronto tra sistemi operativi	105
6.2.4	Aggiornamento nel tempo	105
6.2.5	Limiti dell'analisi predittiva	105
Conclusioni		106
Bibliografia		108
Sitografia		109
Ringraziamenti		111

Elenco delle figure

1.1	Crescita esponenziale dei dati nel tempo	6
1.2	Caratteristiche dei Big Data	8
1.3	Grafico Valore-Veridicità e Valore-Tempo	9
1.4	Dati generati dall'uomo	9
1.5	Dati generati dalle macchine	10
1.6	Classificazione dei dati	12
1.7	Tipologie della Data Analytics	14
1.8	Ciclo di vita della Big Data Analytics	17
1.9	Esempio validazione e pulizia di dati	19
1.10	Esempio di aggregazione di dati	20
1.11	Stesso dato memorizzato in due formati diversi	21
1.12	Piramide DIKW	23
1.13	Processo aziendale	26
2.1	Infrastruttura di Power BI	32
2.2	Sorgenti di dati gestibili da Power BI	33
2.3	Trasformazione dei dati	34
2.4	Sezioni relative alla trasformazione dei dati	34
2.5	Oggetti visivi disponibili in Power BI Desktop	35
2.6	Esempio di report	36
2.7	Differenza tra dashboard e report	37
2.8	Lato sinistro di Power BI Desktop per le visualizzazioni	38
2.9	Visualizzazione dei report	38
2.10	Sezioni laterali relative alla visualizzazione dei report	39
2.11	Visualizzazione Dati in Power BI	40
2.12	Visualizzazione delle Relazioni	41
3.1	Visualizzazione iniziale della pagina di report vuota	50
3.2	Importazione dei dati in Power BI	51
3.3	Visualizzazione dell'anteprima per opzioni di caricamento o trasformazione dei dati	51
3.4	Scelta dell'opzione "Unisci query come nuova" per la creazione di una terza tabella creata tramite operazione di merge	57
3.5	Processo di scelta dei criteri per l'operazione di Merge	58

3.6	Scelta delle colonne del secondo dataset da mantenere nel nuovo dataset ottenuto tramite operazione di Merge	58
4.1	Dashboard 1 - Panoramica generale del mercato degli smartphone	73
4.2	Dashboard 2 - Analisi delle caratteristiche tecniche degli smartphone	75
4.3	Dashboard 3 - Analisi delle caratteristiche tecniche per fascia di prezzo	76
4.4	Dashboard 4 - Analisi dei sistemi operativi negli smartphone	78
4.5	Dashboard 5 - Analisi della soddisfazione degli utenti e del rapporto qualità/prezzo	80
5.1	Menu della sezione Analisi nelle Visualizzazioni	89
5.2	Selezione della funzionalità Previsione nel pannello Analisi	89
5.3	Opzioni configurabili nella sezione Previsione	90
5.4	R e Python visuals	90
5.5	Processo di installazione della libreria <i>scikit-learn</i> utilizzando il terminale Windows	92
5.6	Dashboard 6 - Analisi predittiva dell'andamento dei prezzi e del numero di modelli di smartphone	93
5.7	Dashboard 7 - Analisi predittiva dell'evoluzione hardware degli smartphone	101

	Elenco delle tabelle
--	----------------------

4.1	Principali funzioni DAX in Power BI	63
-----	---	----

3.1	Modifica del tipo di colonne nel Dataset 1	52
3.2	Creazione della colonna Price_Euro	52
3.3	Creazione della colonna Price_Category	52
3.4	Creazione della colonna Price_USD	53
3.5	Ridenominazione delle colonne 5G, IR e NFC	53
3.6	Split della colonna dedicata all'informazione sul sistema operativo	53
3.7	Ridenominazione delle colonne relative al sistema operativo e al prezzo iniziale	53
3.8	Aggiunta della colonna Screen_size_pollici	53
3.9	Ridenominazione della camera	54
3.10	Normalizzazione della camera	54
3.11	Creazione della colonna ModelKey	54
3.12	Colonna personalizzata brand_name del dataset 2	55
3.13	Colonna personalizzata model del dataset 2	55
3.14	Modifica del tipo per la colonna release_date	56
3.15	Colonna personalizzata ModelKey del dataset 2	56
3.16	Modifica del tipo delle colonne del dataset 2	56
4.1	Misura DAX per Average Smartphone Price (INR)	64
4.2	Misura DAX per Average Smartphone Price (EUR)	64
4.3	Misura DAX per Average Smartphone Price (USD)	64
4.4	Misura DAX per Cheapest Brand	64
4.5	Misura DAX per Cheapest Model	64
4.6	Misura DAX per Cheapest Price (EUR)	65
4.7	Misura DAX per Cheapest Price (INR)	65
4.8	Misura DAX per Cheapest Price (USD)	65
4.9	Misura DAX per Most Expensive Brand	65
4.10	Misura DAX per Most Expensive Model	65
4.11	Misura DAX per Most Expensive Price (EUR)	65
4.12	Misura DAX per Most Expensive Price (INR)	66
4.13	Misura DAX per Most Expensive Price (USD)	66
4.14	Misura DAX per Average Battery	66
4.15	Misura DAX per Average RAM	66
4.16	Misura DAX per Average Storage	66
4.17	Misura DAX per Average Rear Camera	66
4.18	Misura DAX per Average Front Camera	66
4.19	Misura DAX per Average Screen Size	66

4.20	Misura DAX per Android models	67
4.21	Misura DAX per iOS models	67
4.22	Misura DAX per Norm RAM	67
4.23	Misura DAX per Norm Storage	67
4.24	Misura DAX per Norm Battery	67
4.25	Misura DAX per Norm Camera	67
4.26	Tabella DAX Metrics	68
4.27	Misura DAX Selected Metric Value	68
4.28	Misura DAX per Selected Metric Android	68
4.29	Misura DAX per Selected Metric iOS	68
4.30	Misura DAX per Average Rating	69
4.31	Misura DAX per Rated Models	69
4.32	Misura DAX per High Rated Percentage	69
4.33	Colonna DAX Release Year	69
4.34	Misura DAX per Total Models 2024	70
4.35	Misura DAX per Predicted Models 2027	70
4.36	Misura DAX per Average Price 2024	70
4.37	Misura DAX per Predicted Average Price 2027	71
4.38	Misura DAX per CAGR Models	71
4.39	Misura DAX per CAGR Prices	71
4.40	Tabella DAX Smartphone_Value_Unique	80
4.41	Misura DAX Rating Price Score	82
4.42	Tabella DAX Feature Impact	83
4.43	Misura DAX Feature Impact Value	83
5.1	Script Python per la realizzazione del grafico <i>Avarage Launch Price Forecast by Brand until 2027</i>	95
5.2	Script Python per la realizzazione del grafico <i>Predicted Smartphone Releases by Brand until 2027</i>	96

La crescente digitalizzazione della società ha trasformato profondamente le modalità con cui le persone interagiscono con il mondo che le circonda. In tale contesto, lo smartphone ha assunto un ruolo centrale, diventando uno strumento indispensabile per la comunicazione, il lavoro, l'intrattenimento e l'accesso a un numero sempre crescente di servizi digitali. In questo contesto, un numero sempre maggiore di servizi risulta accessibile attraverso la rete e le aziende produttrici investono continuamente nello sviluppo di nuove tecnologie, con l'obiettivo di soddisfare le esigenze di una società sempre più digitalizzata. Negli ultimi anni, ciò ha influenzato enormemente il mercato degli smartphone, che ha conosciuto un'evoluzione rapida, caratterizzata da un continuo progresso tecnologico, dall'introduzione di dispositivi sempre più performanti e da una crescente competitività tra i principali produttori.

L'elevato numero di modelli commercializzati, l'eterogeneità delle caratteristiche hardware e software e la disponibilità di grandi quantità di dati rendono sempre più complessa l'analisi delle dinamiche che caratterizzano questo settore. In tale contesto, la capacità di raccogliere, integrare e interpretare dati provenienti da fonti differenti rappresenta un elemento fondamentale per comprendere l'evoluzione del mercato, per individuare trend significativi e per supportare processi decisionali basati sui dati.

È in questa prospettiva che si colloca il presente lavoro di tesi, il cui obiettivo consiste nella progettazione e realizzazione di una campagna di Data Analytics finalizzata all'analisi del mercato degli smartphone mediante l'utilizzo di Power BI. L'obiettivo del lavoro consiste nel trasformare grandi quantità di dati grezzi in informazioni utilizzabili attraverso tecniche di analisi descrittiva e predittiva. Inoltre l'utilizzo di tecniche di Business Intelligence permette di sintetizzare queste informazioni complesse mediante dashboard interattive e visualizzazioni efficaci, favorendo una comprensione immediata dei fenomeni analizzati e l'individuazione di pattern difficilmente osservabili attraverso un'analisi tradizionale.

La scelta di Power BI è motivata dal fatto che si tratta di una piattaforma sviluppata da Microsoft per supportare le attività di Business Intelligence attraverso la realizzazione di visualizzazioni interattive. Uno dei punti di forza di Power BI è la capacità di integrare dati provenienti da software, servizi cloud e applicazioni differenti, consentendo di raccogliere, trasformare, analizzare e visualizzare informazioni provenienti da sorgenti eterogenee in maniera semplice e intuitiva. Inoltre, Power BI mette a disposizione un'ampia gamma di visualizzazioni, oltre a funzionalità integrate per la realizzazione di previsioni, come la possibilità di includere script Python e R per realizzare modelli predittivi basati sui dati storici, nonché funzionalità di forecasting direttamente integrate nella piattaforma.

Lo sviluppo del lavoro è stato organizzato secondo un approccio strutturato, articolato

in più fasi. In un primo momento, sono stati selezionati due dataset pubblici disponibili sulla piattaforma Kaggle, che mette a disposizione numerosi dataset provenienti da differenti contesti applicativi. I dataset selezionati contengono informazioni relative alle caratteristiche tecniche degli smartphone presenti sul mercato. Successivamente tali dati sono stati sottoposti a un processo di ETL (Extract, Transform and Load) finalizzato alla pulizia, trasformazione e integrazione dei dati provenienti da fonti eterogenee, prestando particolare attenzione alla qualità e alla coerenza delle informazioni. In seguito, dopo la preparazione dei dati, il lavoro si è concentrato sulla fase di analisi durante la quale abbiamo effettuato analisi di tipo descrittivo e predittivo. L'analisi descrittiva ha portato alla realizzazione di dashboard dedicate all'esplorazione del mercato, con l'obiettivo di evidenziarne le principali caratteristiche e le relazioni tra le variabili considerate. Successivamente sono state sviluppate dashboard di analisi predittiva, basate su tecniche di forecasting e sull'integrazione di script Python, con l'obiettivo di individuare possibili evoluzioni future del mercato degli smartphone.

Nell'ambito dell'attività di analisi, le dashboard sono state suddivise in due categorie: dashboard di base, dedicate all'analisi descrittiva, e dashboard avanzate, dedicate all'analisi predittiva. Ogni dashboard è stata progettata per evidenziare il mercato degli smartphone da una prospettiva differente.

Il contributo della tesi consiste nel dimostrare come un approccio strutturato di Data Analytics, supportato da strumenti di Business Intelligence, consenta di trasformare dati eterogenei in informazioni utili per analizzare nel dettaglio il mercato degli smartphone, individuare pattern significativi e supportare le strategie aziendali.

Questa tesi è strutturata come di seguito specificato:

- Il Capitolo 1 introduce i concetti di Data Science e Business Intelligence, presenta il paradigma dei Big Data e ne descrive le principali caratteristiche, per poi approfondire la Data Analytics, le sue principali tipologie e il relativo ciclo di vita, concludendo con la presentazione della piramide DIKW, quale modello di trasformazione dei dati in informazione, conoscenza e saggezza.
- Il Capitolo 2 presenta Power BI, la piattaforma di Business Intelligence utilizzata per lo sviluppo del presente lavoro, descrivendone l'infrastruttura, il flusso di lavoro, le principali visualizzazioni e le tecnologie di programmazione integrate.
- Il Capitolo 3 presenta i dataset utilizzati per l'analisi. Esso introduce il processo di ETL (Extract, Transform and Load) distinguendolo dal processo di ELT, e mettendone in evidenza le principali differenze. Infine, descrive come il processo ETL sia stato applicato ai dataset selezionati.
- Il Capitolo 4 introduce i principi della Data Visualization e presenta le cinque dashboard sviluppate per l'analisi descrittiva del mercato degli smartphone, evidenziandone le principali caratteristiche e le relazioni tra le variabili considerate.
- Il Capitolo 5 introduce le principali tecniche di analisi predittiva e descrive le due dashboard sviluppate mediante funzionalità di forecasting e integrazione di script Python, finalizzate a prevedere l'evoluzione del mercato degli smartphone negli anni successivi.
- Infine, il Capitolo 6 espone una riflessione critica sul lavoro svolto, valutandone diversi punti di forza ma anche evidenziandone limitazioni e vulnerabilità.

Introduzione alla Data Analytics

In questo capitolo verranno introdotti i concetti fondamentali della Data Science, evidenziandone le caratteristiche principali e le differenze rispetto ad altri ambiti correlati, come il Machine Learning, l'Intelligenza Artificiale e la Business Intelligence. Successivamente, verrà analizzato l'oggetto principale della Data Science, ovvero i Big Data, descrivendo il significato del termine e le principali caratteristiche che li contraddistinguono. A partire da questo contesto, sarà poi approfondito il concetto di Data Analytics, mettendone in evidenza il ruolo strategico nell'analisi dei dati e nella trasformazione di grandi quantità di informazioni grezze in conoscenza utile a supportare i processi decisionali e l'innovazione in ambito aziendale e sociale. Nel corso del capitolo verranno, inoltre, presentate le quattro principali tipologie di Data Analytics, evidenziando le differenze rispetto alla Data Analysis tradizionale. Infine, saranno descritti il ciclo di vita della Big Data Analytics e il modello della piramide DIKW, utilizzato per rappresentare il processo di trasformazione dei dati in informazione, conoscenza e saggezza direttamente utilizzabile dalle aziende.

1.1 Introduzione al concetto di Data Science

La Data Science è una disciplina che combina principi metodologici basati sul metodo scientifico e tecniche multidisciplinari. È volta a interpretare ed estrarre conoscenza dai dati attraverso una fase di analisi svolta da un esperto, chiamato Data Scientist.

I metodi della Data Science si basano su tecniche provenienti da varie discipline, principalmente dalla matematica, dalla statistica, dalla scienza dell'informazione, dall'informatica e dalle scienze sociali, includendo sottodomini come basi di dati, Business Intelligence, Intelligenza Artificiale e Machine Learning.

Tuttavia, occorre precisare che Data Science, AI e Machine Learning vengono spesso usati come sinonimi. In realtà non lo sono, pur essendo tre concetti strettamente correlati.

- L'“AI – Intelligenza Artificiale”, è quel campo dell'informatica che ha come obiettivo costruire “macchine” che imitino il comportamento umano comprendendo sia la capacità di “pensare” sia quella di “agire”.
- Il “Machine Learning” è l'apprendimento (automatico) della “macchina”; il sistema informatico impara a riconoscere schemi ricorrenti, detti pattern, attraverso algoritmi artificiali.
- La “Data Science” è la disciplina scientifica di costruzione di modelli statistici e di tecniche di analisi che servono a trarre valore dai dati. È meno tecnica e più ampia

del Machine Learning, occupandosi anche del “cosa” e non solo del “come”. Viene definita un sottoinsieme dell’Intelligenza Artificiale perché i risultati della Data Science concorrono a costruire agenti intelligenti.

Tuttavia, la Data Science potrebbe essere facilmente confusa con il termine “*Business Intelligence*” (BI). Entrambi si riferiscono ai dati di un’organizzazione e all’analisi di tali dati, ma differiscono nell’obiettivo.

Business Intelligence (BI) è un termine generico che indica l’insieme delle tecnologie che consentono la preparazione, il data mining, la gestione e la visualizzazione dei dati. Gli strumenti e i processi di Business Intelligence consentono agli utenti finali di identificare le informazioni utili dai dati non elaborati, facilitando il processo decisionale basato sui dati all’interno delle organizzazioni di vari settori. Sebbene gli strumenti di Data Science si sovrappongano in parte a queste funzioni, la Business Intelligence si concentra maggiormente sui dati del passato, e gli insight degli strumenti di BI sono di natura più descrittiva. La BI è orientata verso dati statici (immutabili) che solitamente sono strutturati.

Sebbene la Data Science utilizzi dati descrittivi, in genere lo fa per individuare variabili predittive, che vengono, quindi, utilizzate per classificare i dati o per fare previsioni.

1.1.1 Breve cenni storici

La storia della Data Science può essere descritta attraverso i seguenti momenti fondamentali:

- Nel 1962, John Tukey descrisse un campo che chiamò “data analysis”, che richiama l’attuale Data Science.
- Il termine “Data Science” è stato introdotto per la prima volta nel 1974 dall’informatico danese Peter Naur nel suo libro “Concise Survey of Computer Methods” come evoluzione del concetto di datalogy, utilizzato dallo stesso Naur qualche anno prima, per contrapporlo al concetto, ritenuto più limitativo, di informatica. In questa prima definizione, Naur intende la data science come una disciplina relativa alla gestione e alla manipolazione dei dati così come si presentano, attribuendo poca importanza alla possibilità di estrarre, dagli stessi dati, informazioni di valore.
- La Data Science è stata riconosciuta come disciplina a sé stante (quindi non più una branca dell’informatica e della statistica) soltanto nel 2001, quando William Cleveland ne delineò i campi di competenza, individuando sei aree principali: la ricerca multidisciplinare, i modelli, l’elaborazione dei dati, la pedagogia, la valutazione degli strumenti e la teoria.
- Con l’avvento dei Big Data e dell’idea di “valore del dato”, propria di questo paradigma, si è evoluto il concetto stesso di Data Science, che diviene, così, una scienza olistica, il cui principio fondante non è la mera gestione del dato, ma una più ampia valorizzazione della grande mole di dati eterogenei provenienti da diverse fonti (Data warehouse, sensori, web). La Data Science, al giorno d’oggi, va, quindi, intesa come una disciplina trasversale che comprende informatica, statistica e matematica.
- Nel 2018, il “Big Data Analytics & Business Intelligence Observatory” del Politecnico di Milano ha svolto un’analisi delle offerte di lavoro individuando le figure professionali più richieste dalle aziende nel settore delle Data Science. Da quest’analisi è stato possibile individuare almeno tre tipi di profili professionali distinti: il data scientist, il data engineer e il data analyst.

La data science, quindi, è una parte importante della cosiddetta "Industria 4.0", che descrive il rapido cambiamento della tecnologia, dei modelli e dei processi sociali nel 21° secolo, a causa della crescente interconnettività e dell'automazione intelligente.

1.2 Big Data

La comunità delle basi di dati è sempre stata consapevole dell'esigenza di gestire grandi quantità di dati (il congresso Very Large Data Bases ha questo nome dal 1975), ma solo negli ultimi anni il termine "Big Data" ha assunto un grandissimo rilievo, entrando nell'uso comune.

Ciò è dovuto al fatto che, da un lato, la quantità di informazioni generate ha raggiunto livelli straordinari grazie alle più svariate fonti; dall'altro lato, strumenti informatici sempre più evoluti sono in grado di memorizzare, interrogare e analizzare tali informazioni.

L'enfasi iniziale sui Big Data può essere attribuita a Jim Gray, che, all'inizio di questo secolo, ha collaborato con gli astronomi contribuendo alla costruzione di una gigantesca mappa stellare. Nel libro dedicato alla sua memoria (intitolato "Il quarto paradigma"), vengono delineate le principali problematiche del settore; in particolare vengono descritti quattro paradigmi scientifici successivi, tra cui il quarto evidenzia un concetto molto importante, quello di *data-driven*, che caratterizza l'attuale evoluzione tecnologica e scientifica.

Da un punto di vista filosofico, l'approccio data-driven alla soluzione dei problemi si contrappone all'approccio model-driven. Assumendo che i dati siano sufficientemente numerosi e che la loro significatività statistica garantisca la validità dei risultati sperimentali, non è più necessario disporre di un modello scientifico e la causalità non risulta più strettamente necessaria. La sola osservazione dei dati può infatti fornire risposte sufficienti.

L'evoluzione dell'approccio data-driven e la continua crescita della quantità di dati disponibili hanno portato alla diffusione del concetto di "Big Data", oggi centrale nell'ambito della Data Science e della moderna analisi dei dati.

Infatti, sebbene i Big Data possano sembrare una disciplina recente, in realtà si basano su problematiche studiate da molti anni. Già in passato, infatti, la gestione e l'analisi di grandi dataset rappresentavano una sfida, ad esempio nei censimenti. Il concetto di Big Data si è evoluto proprio a partire da queste radici, sviluppando nuove tecniche e strumenti.

Accanto ai metodi statistici tradizionali, i Big Data introducono nuove tecniche basate sull'uso di elevate risorse computazionali e algoritmi avanzati. Questo cambiamento è dovuto al fatto che i dati sono diventati:

- sempre più grandi (volume);
- più eterogenei (varietà);
- più complessi e spesso generati in tempo reale (streaming).

In passato, l'analisi si basava su campionamenti di dati. Oggi, grazie ai progressi della tecnologia, è possibile elaborare interi dataset, rendendo il campionamento spesso non necessario.

I confini dei Big Data sono in continua evoluzione, influenzati dallo sviluppo delle tecnologie hardware e software. Ciò significa che ciò che oggi non è considerato Big Data potrebbe diventarlo in futuro (e viceversa).

Le applicazioni dei Big Data sono innumerevoli; tra queste troviamo:

- *Profilazione degli utenti.* Attraverso l'analisi dei dati provenienti da motori di ricerca, social network e dispositivi telefonici, le aziende possono ottenere informazioni

dettagliate sui comportamenti degli utenti. Questo consente di prevedere le abitudini dei consumatori e di proporre pubblicità mirate.

- *Ottimizzazione dei processi produttivi.* Nell'ambito industriale, l'analisi dei dati permette di migliorare l'efficienza dei processi produttivi, riducendo costi e tempi. I modelli predittivi consentono, inoltre, di ottimizzare la gestione della logistica, dei magazzini e delle catene di distribuzione.
- *Miglioramento della salute.* La raccolta e l'analisi di grandi quantità di dati sanitari consentono di correlare l'insorgenza delle malattie con fattori ambientali e comportamentali, migliorando la qualità delle cure e permettendo lo sviluppo di terapie personalizzate basate su modelli statistici e predittivi.
- *Aumento della sicurezza.* Incrociando informazioni provenienti da diverse fonti, è possibile individuare attività criminali, rilevare frodi e prevenire fenomeni di corruzione o terrorismo attraverso tecniche di analisi avanzata dei dati.
- *Miglioramento dell'organizzazione delle città.* I Big Data permettono di sviluppare servizi più efficienti nella smart city, migliorando sistemi amministrativi, trasporti e gestione urbana. Inoltre, l'analisi dei dati consente di prevedere eventi critici, come fenomeni meteorologici, e ridurre gli effetti.

I Big Data, oggi, sono un argomento interessante per molte aziende, le quali negli ultimi anni hanno investito in questa tecnologia più di 15 miliardi di dollari, finanziando lo sviluppo di software per la gestione e l'analisi dei dati. Ciò è avvenuto perché le economie più forti sono fortemente interessate all'analisi di enormi quantità di dati: basti pensare che ci sono oltre 4,6 miliardi di smartphone attivi e circa 2 miliardi di persone hanno accesso a Internet.

Dal 1986 ad oggi il volume dei dati in circolazione è aumentato in maniera esponenziale:

- nel 1986 i dati erano 281 Petabyte;
- nel 1993 i dati erano 471 Petabyte;
- nel 2000 i dati erano 2,2 Exabyte;
- nel 2007 i dati erano 65 Exabyte;
- nel 2014 i dati sono arrivati oltre i 650 Exabyte.

La Figura 1.1 mostra la crescita esponenziale dei dati nel corso degli anni.

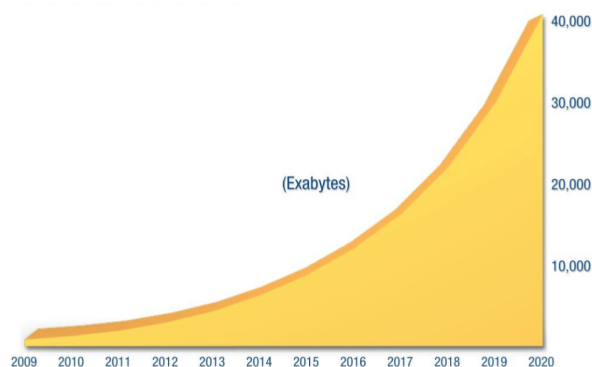


Figura 1.1: Crescita esponenziale dei dati nel tempo

1.3 Le 5 "V" dei Big Data

I Big Data vengono comunemente descritti attraverso quattro proprietà caratterizzate dalla stessa iniziale, denominate comunemente le quattro "V": Volume, Velocità, Varietà, Veridicità, alle quali si aggiunge spesso una quinta proprietà: il Valore.

1.3.1 Volume

È la principale caratteristica dei Big Data. Si riferisce alla grande quantità di dati generati. I Big Data si contraddistinguono per le loro dimensioni, che possono essere molto maggiori rispetto a quelle dei dati tradizionali. La gestione e l'analisi di grandi volumi di dati richiedono infrastrutture e strumenti specifici.

Ad esempio, secondo alcune stime, nel 2022, in un solo minuto venivano prodotte 5,9 milioni di ricerche su Google, 1,7 milioni di contenuti su Facebook, 500 ore di video su YouTube e 231 milioni di messaggi.

Per la gestione del volume, bisogna investire in infrastrutture adatte. Occorre assicurarsi di avere la capacità di archiviare e elaborare grandi volumi di dati, e ciò potrebbe richiedere l'adozione di tecnologie come il cloud computing o l'implementazione di soluzioni di storage scalabili.

Inoltre, è opportuno utilizzare tecniche di compressione dei dati per ridurre la dimensione senza perdere informazioni significative.

1.3.2 Velocità

Indica la rapidità con cui avviene la generazione e la raccolta dei dati. Nel contesto dei Big Data, i dati possono essere generati in real-time o in alta frequenza. La capacità di elaborare e analizzare i dati in tempo reale diventa, quindi, cruciale. Ciò può includere l'utilizzo di strumenti di streaming analytics o di piattaforme di elaborazione dei dati in real-time. Lo scopo è di ottenere informazioni utili e maggiore rapidità nei processi decisionali aziendali.

1.3.3 Varietà

Si riferisce ai diversi tipi di dati attualmente utilizzati. Mentre in passato l'attenzione principale era rivolta ai dati strutturati, generalmente memorizzati in tabelle, oggi la maggior parte dei dati sono non strutturati (testi, immagini, audio, video).

Infatti, i dati possono provenire da diverse fonti ed essere strutturati, non strutturati o semi-strutturati. La gestione e l'analisi di questa varietà di dati richiedono soluzioni flessibili e adattabili. Per gestire tale varietà, si consiglia, quindi, di utilizzare strumenti di *data integration*.

1.3.4 Veridicità

Si riferisce alla possibilità di estrarre informazioni affidabili dai dati, nonostante possano contenere errori, imprecisioni e incompletezze. Nel contesto dei Big Data, è importante garantire che i dati raccolti siano accurati, completi e privi di errori.

Per garantire la veridicità dei dati, si consiglia di implementare processi di controllo della qualità. Ciò può includere l'utilizzo di tecniche di *data cleaning* o di algoritmi di rilevamento degli errori. Inoltre, per mantenere la veridicità dei dati, è importante implementare anche misure di sicurezza adeguate per proteggere i dati stessi da accessi non autorizzati o modifiche indebite.

La Figura 1.2 mostra le caratteristiche dei Big Data.



Figura 1.2: Caratteristiche dei Big Data

1.3.5 Valore

Rappresenta il potenziale beneficio che può essere ottenuto dai dati. I Big Data offrono l'opportunità di analizzare e sfruttare i dati per ottenere informazioni preziose, identificare tendenze, modelli e correlazioni, migliorare i processi decisionali, individuare nuove opportunità di business e offrire un'esperienza personalizzata ai clienti.

Per estrarre valore dai dati, si rende necessario utilizzare strumenti di analisi avanzati come il Machine Learning, l'Intelligenza Artificiale o la visualizzazione dei dati.

I risultati dell'analisi dei dati devono essere integrati nelle decisioni aziendali per ottenere un impatto concreto. Ciò richiede una collaborazione tra i team di analisi dei dati e le figure responsabili dei processi decisionali.

È importante osservare che la caratteristica del valore è strettamente legata a quella della veridicità.

Più i dati sono accurati, affidabili e privi di errori, maggiore sarà il valore che possono offrire all'azienda. Dati di scarsa qualità, invece, portano a decisioni sbagliate, e quindi a una perdita di valore.

Tuttavia, il valore non dipende solo dalla qualità dei dati, ma anche dal tempo.

Infatti, i risultati ottenuti dall'analisi dei dati hanno spesso una sorta di "data di scadenza". Un'informazione può essere molto utile se arriva subito, ma può diventare inutile se arriva in ritardo. Per questo motivo si può affermare che valore e tempo sono inversamente correlati: più tempo si impiega per trasformare i dati in informazioni, minore è il valore per l'azienda.

In conclusione, i Big Data rappresentano una grande opportunità, ma anche una sfida. Da un lato permettono analisi più profonde e decisioni più informate; dall'altro richiedono tecnologie, competenze e strategie adeguate per essere gestiti efficacemente. Capire e saper utilizzare i dati rappresenta oggi un vantaggio competitivo fondamentale per il futuro.

Nella Figura 1.3 possiamo osservare i legami Valore-Veridicità e Valore-Tempo.

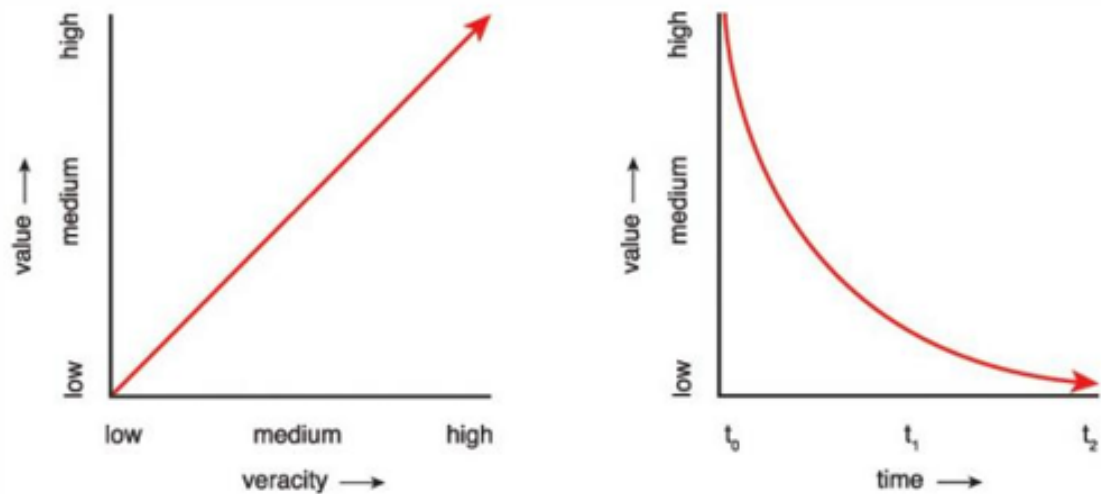


Figura 1.3: Grafico Valore-Veridicità e Valore-Tempo

1.4 La varietà dei dati coinvolti nei Big Data

Come già spiegato precedentemente, i dati coinvolti nei Big Data provengono da fonti estremamente differenti.

I dati processati da soluzioni basate su Big Data possono essere generati dagli uomini o dalle macchine, sebbene la responsabilità ultima nella generazione dei risultati analitici sia delle macchine.

Possiamo, quindi, analizzare separatamente i dati generati dall'uomo, da quelli generati dalle macchine.

Dati generati dall'uomo

I dati generati dall'uomo sono il risultato dell'interazione umana con i sistemi, ad esempio nei servizi online e nei dispositivi digitali.

La Figura 1.4 mostra esempi di dati generati dall'uomo (social media, post nei blog, e-mail, condivisione di foto e messaggi).

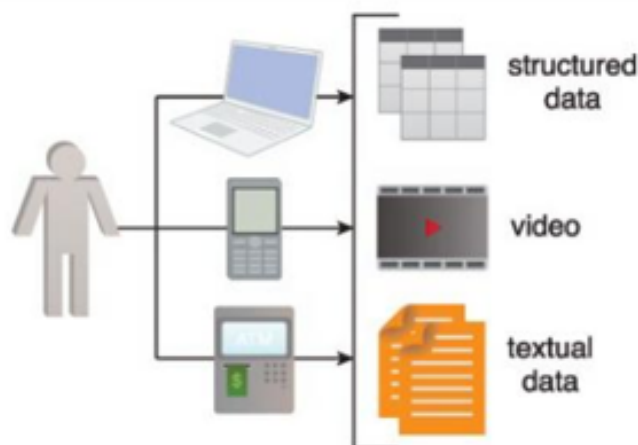


Figura 1.4: Dati generati dall'uomo

Dati generati dalla macchina

I dati generati dalla macchina sono prodotti da programmi software e dispositivi hardware in risposta a eventi del mondo reale.

Per esempio, un file di log registra una decisione di autorizzazione fatta da un servizio di sicurezza, un sistema di un punto vendita genera una transazione sull’inventario per riflettere i prodotti acquistati da un cliente, e così via.

Da una prospettiva hardware, un esempio di dati generati dalle macchine è rappresentato dalle informazioni restituite dai numerosi sensori nelle celle telefoniche; queste possono riportare informazioni come la posizione e l’intensità del segnale.

La Figura 1.5 fornisce una rappresentazione visiva di diverse tipologie di dati generati dalle macchine (web log, dati relativi ai sensori, dati di telemetria, dati di misurazione intelligenti e dati sugli utilizzi degli apparecchi).

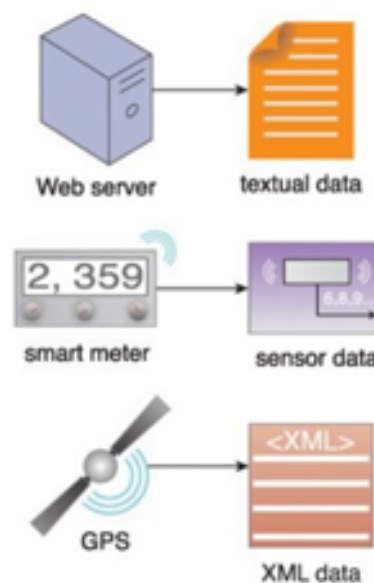


Figura 1.5: Dati generati dalle macchine

Classificazione dei dati

I dati generati dall’uomo e quelli generati dalla macchina provengono da una varietà di sorgenti e possono essere rappresentati in vari formati o tipi.

In base alla loro organizzazione interna, i dati vengono generalmente classificati in tre principali categorie: dati strutturati, dati non strutturati e dati semi-strutturati. A queste si aggiungono i metadati, fondamentali per descrivere le caratteristiche e l’origine dei dati stessi.

1.4.1 Dati strutturati

I dati strutturati sono quei dati che seguono un modello ben definito, cioè uno schema preciso. Questo significa che sono organizzati in modo ordinato, spesso sotto forma di tabelle, con righe e colonne. Grazie a questa struttura, è molto facile gestirli, analizzarli e, soprattutto, mettere in relazione diverse entità. Proprio per questo motivo, i dati strutturati vengono generalmente memorizzati all’interno di database relazionali. Questi dati sono molto diffusi

nel contesto aziendale; infatti vengono generati quotidianamente da applicazioni e sistemi informativi, come gli ERP e i CRM.

Un altro vantaggio importante è che, essendo così diffusi, esistono numerosi strumenti e database progettati appositamente per gestire dati strutturati. Di conseguenza, questi dati non richiedono particolari complessità né nella memorizzazione né nell'elaborazione.

1.4.2 Dati non strutturati

I dati non strutturati sono tutti quei dati che non seguono uno schema o un modello predefinito. A differenza dei dati strutturati, non sono organizzati in tabelle e non possono essere facilmente inseriti in database relazionali.

Si stima che questi dati rappresentino fino all'80% dei dati presenti all'interno di un'azienda, e la loro crescita è ancora più rapida rispetto a quella dei dati strutturati.

Tra gli esempi più comuni troviamo:

- file di testo, come e-mail, tweet o post sui blog;
- audio;
- video.

Questi dati sono generalmente salvati in file che sono autocontenuti e non relazionali, cioè non organizzati secondo relazioni precise tra elementi. Anche se tecnicamente essi hanno una struttura definita dal loro formato, ciò che li rende "non strutturati" è il fatto che il contenuto non segue uno schema facilmente interrogabile.

Per questo motivo, i dati non strutturati richiedono una logica di elaborazione più complessa e dipendente dal contesto. Ad esempio, per interpretare un file video è necessario utilizzare il codec corretto. Un'altra caratteristica importante è che questi dati non possono essere interrogati direttamente tramite SQL, come avviene per i dati strutturati; per tale motivo essi vengono spesso gestiti tramite database NoSQL.

1.4.3 Dati semi-strutturati

I dati semi-strutturati rappresentano una via di mezzo tra le due tipologie precedenti. Infatti, possiedono un certo livello di struttura e organizzazione, ma non seguono il rigido schema dei database relazionali. Questi dati sono generalmente non relazionali e organizzati in forma gerarchica o a grafo. Questo significa che le informazioni sono strutturate, ma in modo più flessibile rispetto alle tabelle tradizionali.

Spesso i dati semi-strutturati sono memorizzati in file di testo, e tra i formati più comuni troviamo XML e JSON.

Questi formati permettono di rappresentare i dati in modo leggibile e organizzato, mantenendo, però, una certa flessibilità. Grazie a questa struttura parziale, i dati semi-strutturati risultano più facili da elaborare rispetto ai dati non strutturati, anche se richiedono comunque strumenti e tecniche specifiche.

Esempi di fonti comuni includono:

- file EDI (Electronic Data Interchange);
- spreadsheet;
- RSS feed;
- dati provenienti da sensori.

Un aspetto importante è che questi dati spesso necessitano di operazioni di *pre-processing*, cioè attività preliminari prima dell'analisi.

Ad esempio, un file XML può essere validato per verificare che rispetti il proprio schema.

1.4.4 Metadati

I metadati sono, in parole semplici, dati sui dati. Forniscono informazioni aggiuntive riguardo alle caratteristiche, alla struttura e al contenuto di un dataset. Queste informazioni non rappresentano il dato principale, ma aiutano a comprenderlo meglio. Ad esempio, possono indicare chi ha creato un documento, quando è stato creato, oppure le caratteristiche tecniche di un file.

I metadati sono spesso generati automaticamente dalle macchine e vengono associati direttamente ai dati a cui si riferiscono.

Un aspetto molto importante è il tracking dei metadati, cioè la capacità di tracciarne l'evoluzione. Questo è fondamentale nei Big Data perché permette di conoscere il cosiddetto "pedigree" dei dati, cioè la loro origine e il percorso che hanno seguito durante le varie fasi di elaborazione.

Questo aiuta a garantire qualità, affidabilità e trasparenza nell'analisi dei dati.

Esempi concreti di metadati includono:

- i tag XML, che indicano autore e data di creazione di un documento;
- gli attributi di un file, come dimensione o risoluzione di un'immagine.

Infine, le soluzioni di Big Data si basano fortemente sui metadati, soprattutto quando si lavora con dati semi-strutturati e non strutturati, dove la struttura non è immediatamente evidente.

Nella Figura 1.6 è mostrato un semplice schema che illustra la distribuzione delle varie tipologie di dati.

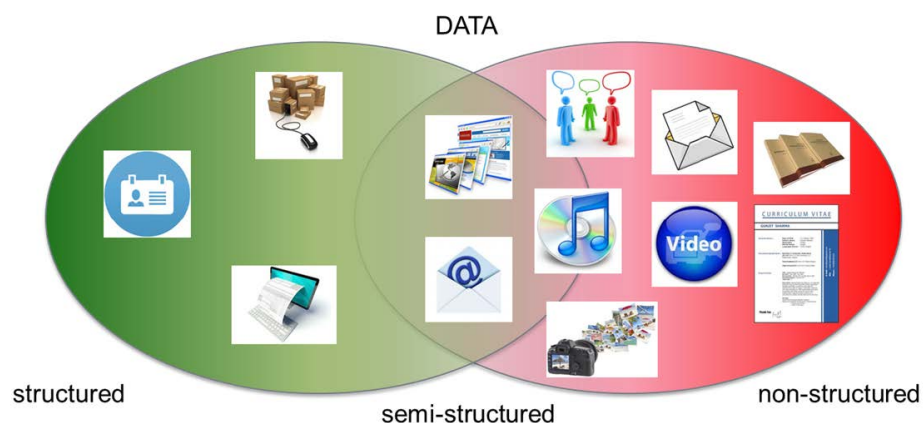


Figura 1.6: Classificazione dei dati

1.5 Data Analytics

La Data Analytics è il processo di analisi, interpretazione ed elaborazione grafica di insiemi di dati complessi e voluminosi, allo scopo di individuare pattern significativi e supportare decisioni strategiche. Si articola in una varietà di tecniche che vanno dall'analisi

statistica al machine learning, finalizzate all'analisi dei dati passati e attuali per prevedere le tendenze future.

La Data Analytics è una disciplina importante perché permette di estrarre valore dai dati, oggi considerati una vera e propria materia prima.

Per le organizzazioni che cercano di conquistare e mantenere un vantaggio competitivo, la Data Analytics è fondamentale per il mantenimento del successo.

Gli strumenti e i team specializzati in analisi dei dati svolgono un ruolo essenziale in questa sequenza, raccogliendo, elaborando e analizzando minuziosamente i dati.

Questa disciplina trova applicazione in un'ampia gamma di settori come:

- finanza;
- sanità;
- settore manifatturiero;
- commercio;
- marketing.

Le tecniche di Data Analytics possono essere suddivise in quattro principali categorie:

- analisi descrittiva;
- analisi diagnostica;
- analisi predittiva;
- analisi prescrittiva.

Come già detto, all'inizio degli anni 2000, i progressi nelle funzionalità software e hardware hanno reso possibile per le organizzazioni la raccolta e la gestione di grandi quantità di dati non strutturati.

Grazie a questa esplosione di dati utili, le comunità open source hanno sviluppato framework per i Big Data per memorizzare ed elaborare tali dati. Questi framework vengono utilizzati per lo storage distribuito e l'elaborazione di set di dati di grandi dimensioni in una rete di computer.

Insieme ad altri strumenti e librerie, i framework di Big Data possono essere utilizzati per:

- *modellazione predittiva* tramite l'integrazione di Intelligenza Artificiale (AI) e algoritmi statistici;
- *analisi statistica* per l'esplorazione approfondita dei dati;
- *analisi "what-if"* per simulare diversi scenari ed esplorare i potenziali risultati;
- *elaborazione di diversi set di dati*, tra cui dati strutturati, semi-strutturati e non strutturati provenienti da varie fonti.

In questo contesto nasce un'estensione della Data Analytics, ovvero la Big Data Analytics.

La Big Data Analytics non è altro che l'elaborazione e l'analisi sistematiche di grandi quantità di dati e set di dati complessi, noti come Big Data, allo scopo di scoprire tendenze, modelli e correlazioni tra i dati per aiutare gli analisti a prendere decisioni informate e a individuare insight preziosi.

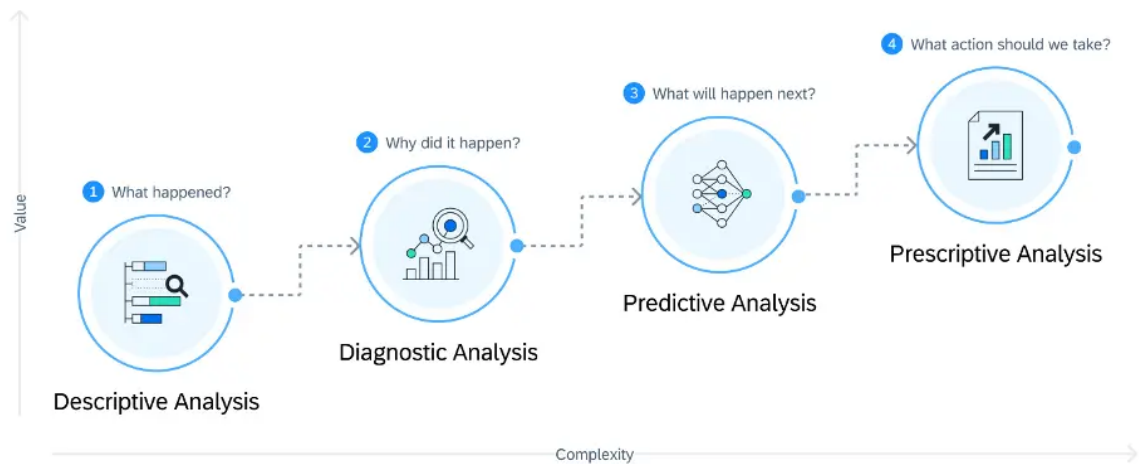


Figura 1.7: Tipologie della Data Analytics

1.6 Tipologie della Data Analytics

La Figura 1.7 illustra le principali tipologie della Data Analytics.

1.6.1 Analisi descrittiva

L'analisi descrittiva costituisce lo stadio preliminare della Data Analysis. Risponde al quesito: "che cosa è successo?". L'analisi descrittiva precede l'analisi diagnostica, a sua volta seguita dall'analisi predittiva e dall'analisi prescrittiva.

Ha lo scopo di descrivere ciò che è già accaduto. Questo tipo di analisi si concentra sull'organizzazione e sulla sintesi dei dati storici, per fornire una visione chiara della situazione. Viene, spesso, utilizzata tramite report e dashboard.

Per quanto riguarda il valore aggiunto, l'analisi descrittiva fornisce i risultati di minor valore. I report hanno generalmente una natura statica e mostrano dati storici presentati sotto forma di griglie o grafici.

1.6.2 Analisi diagnostica

L'analisi diagnostica ha lo scopo di determinare la causa di un fenomeno accaduto nel passato utilizzando domande che si focalizzano sulla ragione che stava dietro al fenomeno.

Affronta il quesito "Perché è successo?" utilizzando tecniche come il drill-down, l'esplorazione dei dati, il *data mining* e le correlazioni.

L'analisi diagnostica fornisce più valore rispetto all'analisi descrittiva ma richiede un insieme di skill più avanzate.

I risultati dell'analisi diagnostica vengono visualizzati tramite tool di visualizzazione interattiva che consentono agli utenti di identificare trend e pattern.

1.6.3 Analisi predittiva

L'analisi predittiva si avvale di modelli statistici e del Machine Learning per prevedere i risultati futuri sulla base dei dati storici. Risponde alla domanda "Cosa potrebbe succedere?" individuando le tendenze e prevedendo gli eventi futuri.

Con l'analisi predittiva, le informazioni vengono arricchite di significato al fine di generare conoscenza e individuare correlazioni tra i dati.

L'analisi predittiva mira a prevedere eventi futuri; le previsioni vengono effettuate sulla base di pattern, trend ed eccezioni individuati nei dati storici e correnti.

Questo tipo di analisi coinvolge l'uso di grandi dataset che comprendono dati interni ed esterni nonché varie tecniche di analisi dei dati.

1.6.4 Analisi prescrittiva

L'analisi prescrittiva suggerisce azioni in base all'analisi dei dati. Affronta il quesito "Cosa dovremmo fare?" affidandosi a tecniche quali gli algoritmi di ottimizzazione e simulazione.

Il focus è non solo su qual è l'azione migliore da prescrivere, ma anche sul perché; in altre parole, l'analisi prescrittiva fornisce risultati su cui si possono effettuare ragionamenti con lo scopo di comprendere lo scenario sottostante.

Pertanto, questo tipo di analisi può essere utilizzata per ottenere un vantaggio o per mitigare un rischio.

L'analisi prescrittiva fornisce più valore rispetto a qualunque altra tipologia di analisi e richiede le skill più avanzate, come pure software e tool specializzati

1.7 Differenza tra Data Analytics e Data Analysis

Quando in generale si parla di Data Analytics e Data Analysis comunemente i due termini vengono utilizzati come sinonimi; tuttavia, a livello tecnico, pensare che siano la stessa cosa è un errore. La distinzione generalmente accettata è la seguente:

- Data Analytics è l'ampio campo che si occupa dell'utilizzo di dati e strumenti per prendere decisioni aziendali.
- Data Analysis è un sottoinsieme della data analytics e si occupa della estrazione di informazione dai dati.

La Data Analysis rappresenta una sottoattività della Data Analytics e si focalizza sull'analisi di uno o più dataset strutturati, mediante operazioni di pulizia, trasformazione, interrogazione e modellazione dei dati, con l'obiettivo di estrarre informazioni significative e utili all'interpretazione dei fenomeni analizzati. Si può quindi esprimere che l'analisi dei dati è il processo che consiste nell'esaminare i dati per individuare fatti, relazioni, pattern e trend significativi con l'obiettivo principale di supportare il processo decisionale (decision making) in modo efficace.

D'altra parte la Data Analytics è un termine più vasto che comprende la Data Analysis; è una disciplina che include la gestione del ciclo di vita completo dei dati; quest'ultimo consiste nelle seguenti fasi:

- raccolta;
- pulizia;
- organizzazione;
- memorizzazione;
- analisi;
- gestione e governance dei dati.

Oltre agli aspetti operativi e gestionali, la Data Analytics comprende anche differenti tipologie di analisi, utilizzate per estrarre valore dai dati e supportare decisioni strategiche precedentemente illustrate, come l'analisi descrittiva, l'analisi diagnostica, l'analisi predittiva e l'analisi prescrittiva.

Pertanto, mentre la Data Analysis si concentra principalmente sull'esame diretto dei dati e sull'estrazione di informazioni significative, la Data Analytics adotta un approccio più ampio e strategico, integrando metodologie, tecnologie e diversi modelli analitici per generare valore lungo l'intero ciclo di vita del dato.

1.8 Ciclo di vita della Big Data Analytics

L'analisi dei Big Data si differenzia dall'analisi tradizionale dei dati principalmente per il volume, la velocità e la varietà dei dati elaborati. Queste caratteristiche rendono il processo analitico molto più complesso e richiedono strumenti, tecnologie e approcci differenti rispetto a quelli tradizionali.

Per gestire efficacemente grandi quantità di dati, è necessario adottare una metodologia step-by-step, in grado di organizzare tutte le attività e i task coinvolti nel processo analitico. Tale metodologia interessa tutte le principali fasi del trattamento dei dati:

- acquisizione dei dati;
- elaborazione;
- analisi;
- riutilizzo dei dati.

Proprio per questo motivo viene introdotto il concetto di Data Analytics Lifecycle, ovvero un ciclo di vita della Data Analytics, che consente di organizzare e gestire in modo strutturato tutte le attività legate all'analisi dei Big Data.

Il ciclo di vita della Big Data Analytics può essere suddiviso nelle seguenti nove fasi:

- Business Case Evaluation;
- Data Identification;
- Data Acquisition & Filtering;
- Data Extraction;
- Data Validation & Cleansing;
- Data Aggregation & Representation;
- Data Analysis;
- Data Visualization;
- Utilization of Analysis Results.

Nella Figura 1.8 viene mostrato il ciclo di vita della Big Data Analytics. Nelle prossime sottosezioni esamineremo in dettaglio ciascuna di queste fasi:

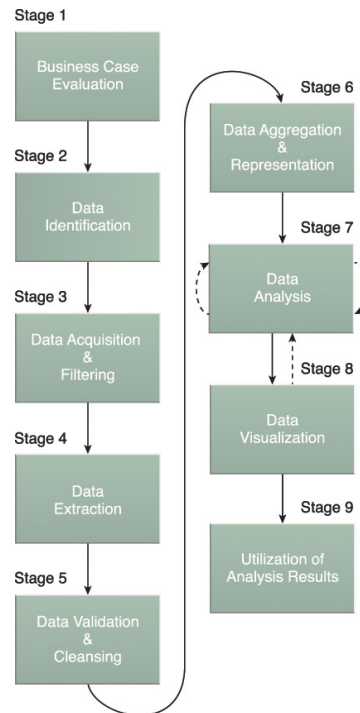


Figura 1.8: Ciclo di vita della Big Data Analytics

1.8.1 Business Case Evaluation

La Business Case Evaluation rappresenta la prima fase del ciclo di vita della Big Data Analytics e consiste nella valutazione preliminare degli obiettivi, della fattibilità e del valore di un progetto di analisi dei dati. In questa fase vengono definiti il problema aziendale, le risorse necessarie e i risultati attesi.

Tra le attività principali rientrano:

- definizione del problema e degli obiettivi;
- identificazione degli stakeholder coinvolti;
- valutazione della disponibilità dei dati;
- individuazione dei KPI per la valutazione dei risultati.

Gli obiettivi devono essere formulati secondo il modello SMART, quindi devono essere specifici, misurabili, raggiungibili, rilevanti e tempestivi.

In questa fase viene, inoltre, verificato se il problema possa essere considerato un problema di Big Data, ossia se risulti legato ad almeno una delle caratteristiche principali: volume, velocità o varietà.

Infine, viene definito il budget necessario considerando costi relativi a strumenti, hardware e formazione del personale.

1.8.2 Data Identification

Lo stadio di Data Identification ha l'obiettivo di identificare i dataset necessari per il progetto di analisi e le relative sorgenti dei dati.

L'identificazione di una maggiore varietà di fonti di dati può aumentare la probabilità di trovare modelli e correlazioni nascosti. Ad esempio, per ottenere informazioni più

approfondite, può essere utile identificare il maggior numero possibile di tipologie di fonti correlate, soprattutto quando non è chiaro esattamente cosa cercare.

A seconda della portata del progetto di analisi e della natura dei problemi aziendali da affrontare, i dataset necessari e le relative fonti possono essere interni e/o esterni all'azienda.

Durante la scelta dei dataset è importante valutare con attenzione la qualità dei dati, verificare la pertinenza rispetto agli obiettivi del progetto e controllare la disponibilità e l'accessibilità delle fonti.

1.8.3 Data Acquisition and Filtering

Durante lo stadio di Data Acquisition and Filtering, come suggerisce il nome della fase, il processo può essere suddiviso in due attività principali:

- acquisizione dei dati;
- filtraggio dei dati.

In questa fase i dati vengono recuperati dalle sorgenti identificate nello stadio precedente. Una volta individuate le fonti sulle quali si intende lavorare, i dati vengono estratti e, nella maggior parte dei casi, combinati in un unico dataset.

Successivamente, i dati acquisiti vengono sottoposti a un processo di filtraggio automatico finalizzato alla rimozione di dati corrotti o non rilevanti per l'analisi. In particolare, i dati possono essere valutati, puliti e trasformati per migliorarne la qualità.

I dati considerati corrotti possono includere:

- valori mancanti;
- dati inconsistenti o privi di significato;
- tipi di dati non validi.

Durante questa fase possono, inoltre, essere aggiunti automaticamente dei metadati sia ai dati provenienti da sorgenti interne sia a quelli provenienti da sorgenti esterne. Tali metadati possono includere:

- dimensione e struttura del dataset;
- informazioni sulla sorgente;
- data e ora di creazione;
- informazioni specifiche del linguaggio.

È fondamentale che i metadati siano interpretabili dalla macchina e utilizzabili nelle successive fasi di analisi.

L'aggiunta dei metadati consente, inoltre, di mantenere la tracciabilità dei dati lungo tutto il ciclo di vita della Big Data Analytics, contribuendo a preservarne accuratezza e qualità.

1.8.4 Data Extraction

Durante l'analisi dei Big Data può accadere che alcuni dati identificati come input siano disponibili in formati incompatibili con la soluzione adottata.

Questo problema si verifica soprattutto quando i dati provengono da sorgenti esterne, poiché spesso risultano eterogenei per struttura e formato.

Per questo motivo, lo stadio di Data Extraction ha l'obiettivo di estrarre dati provenienti da diverse sorgenti e trasformarli in un formato compatibile con il sistema di Big Data utilizzato.

In questa fase vengono, quindi, eseguite operazioni di conversione, trasformazione e normalizzazione dei dati, così da renderli utilizzabili nelle successive attività di analisi.

1.8.5 Data Validation and Cleansing

Nell'analisi dei Big Data, i dati non validi possono inquinare e falsificare i risultati, rendendo l'analisi poco affidabile.

A differenza dei dati aziendali tradizionali, in cui la struttura dei dati è predefinita e i dati sono prevalidati, i dati inseriti nelle analisi dei Big Data possono essere non strutturati e privi di qualsiasi indicazione di validità. La loro complessità può, inoltre, rendere difficile definire un insieme di vincoli di validazione adeguati.

Lo stadio di Data Validation and Cleansing ha, quindi, il compito di:

- stabilire regole di validazione, spesso complesse;
- rimuovere o correggere dati non validi.

Inoltre, le soluzioni Big Data spesso ricevono dati ridondanti provenienti da diversi dataset. Questa ridondanza può essere sfruttata per esplorare dataset interconnessi al fine di definire parametri di validazione e colmare le lacune relative ai dati validi mancanti.

Un esempio di validazione e pulizia dei dati viene illustrato nella Figura 1.9. In esso:

1. il primo valore nel Dataset B viene confrontato con il suo valore corrispondente nel Dataset A;
2. il secondo valore nel Dataset B non corrisponde al valore presente nel Dataset A;
3. se un valore è mancante, viene ricavato dal Dataset A.

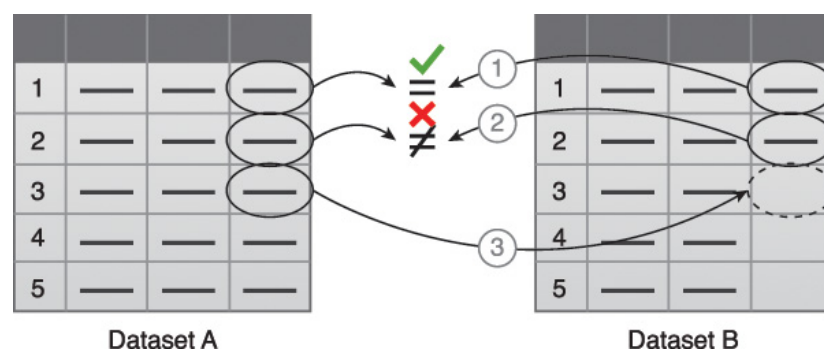


Figura 1.9: Esempio validazione e pulizia di dati

1.8.6 Data Aggregation and Representation

Durante l'analisi dei Big Data, i dati possono essere distribuiti su più dataset, per cui è necessario collegarli attraverso campi comuni, come ad esempio una data o un identificatore. In altri casi, gli stessi dati possono comparire in dataset diversi, creando duplicazioni o incongruenze. Per questo motivo è necessario:

- utilizzare un metodo di riconciliazione dei dati, oppure
- determinare quale dataset rappresenta il valore corretto.

Lo stadio di Data Aggregation and Representation è dedicato a integrare più dataset per arrivare a una visione unificata. L'esecuzione di questa fase può diventare complessa a causa delle differenze presenti tra i dataset, in particolare per quanto riguarda:

- *la struttura dei dati*: sebbene il formato dei dati possa essere lo stesso, il modello dei dati può risultare diverso;
- *la semantica*: valori etichettati in modo diverso in dataset differenti possono avere lo stesso significato, come nel caso di "surname" e "last name".

Gli elevati volumi di dati elaborati dalle soluzioni di Big Data possono rendere l'aggregazione dei dati un'operazione particolarmente onerosa in termini di tempo e risorse. Per questo motivo, in questa fase, è necessario considerare anche i futuri requisiti di analisi, così da favorire il riutilizzo dei dati nelle successive attività analitiche.

Sia nel caso in cui l'aggregazione dei dati sia necessaria sia nel caso in cui non lo sia, è importante comprendere che gli stessi dati possono essere memorizzati in molte forme differenti. La modalità di rappresentazione dei dati influisce, infatti, sulle successive attività di analisi. Alcune rappresentazioni risultano più adatte di altre a seconda del tipo di analisi da effettuare; in particolare:

- dati in formato BLOB sono difficili da interrogare;
- dati organizzati in colonne sono facilmente interrogabili campo per campo.

La Figura 1.10 mostra un semplice esempio di aggregazione dei dati in cui due dataset vengono integrati utilizzando il campo Id.

La Figura 1.11 mostra lo stesso dato memorizzato in due formati diversi. Il Dataset A contiene il dato desiderato, ma fa parte di un BLOB non facilmente interrogabile tramite query. Il Dataset B contiene lo stesso dato organizzato in una struttura a colonne, che consente di interrogare ciascun campo singolarmente.



Figura 1.10: Esempio di aggregazione di dati

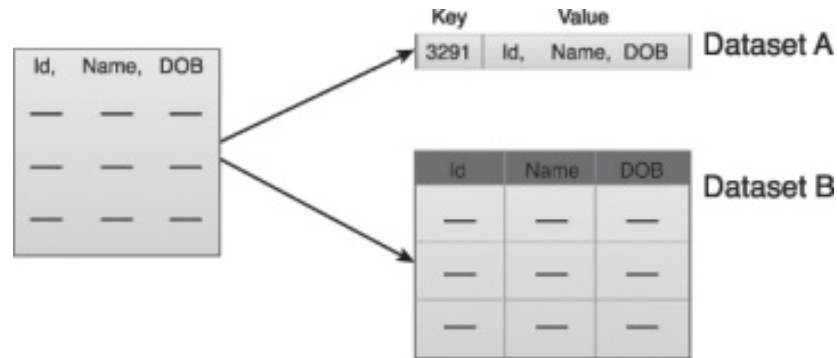


Figura 1.11: Stesso dato memorizzato in due formati diversi

1.8.7 Data Analysis

La fase di Data Analysis è dedicata all'esecuzione dell'analisi vera e propria, che, in genere, prevede uno o più tipi di analisi.

L'analisi dei dati è un processo iterativo che prevede l'estrazione di informazioni significative da dataset ampi e complessi. Lo studio viene ripetuto fino a quando non vengono individuati pattern o associazioni significative.

A seconda del tipo di risultato richiesto, questo stadio può assumere diversi livelli di complessità:

- *semplice*: ad esempio interrogare un dataset ed eseguire aggregazioni per confrontare i dati;
- *complessa*: combinare tecniche di *data mining* e statistica avanzata per:
 - trovare pattern nascosti;
 - individuare anomalie;
 - costruire modelli matematici o statistici.

L'analisi dei dati può essere classificata in analisi confermativa o analisi esplorativa, quest'ultima legata al data mining. In particolare:

- *L'analisi confermativa dei dati*: rappresenta un approccio deduttivo in cui la causa del fenomeno analizzato viene ipotizzata a priori. L'ipotesi formulata viene successivamente verificata attraverso l'analisi dei dati, con l'obiettivo di confermarla o smentirla e fornire risposte specifiche a determinate domande. In questo approccio vengono spesso utilizzate tecniche di campionamento dei dati, mentre eventuali anomalie o risultati inattesi tendono a essere ignorati poiché l'analisi segue una causa già prestabilita.
- *L'analisi esplorativa dei dati*: rappresenta invece un approccio induttivo, strettamente collegato al *data mining*. In questo caso non vengono formulate ipotesi iniziali, ma i dati vengono esplorati direttamente per comprendere le possibili cause dei fenomeni osservati. Sebbene non fornisca sempre risposte definitive, questo metodo permette di individuare pattern, correlazioni e anomalie utili a orientare le successive attività di analisi.

1.8.8 Data Visualization

La capacità di analizzare enormi quantità di dati e ricavarne informazioni utili ha poco valore se gli unici in grado di interpretarne i risultati sono gli analisti.

Lo stadio di Data Visualization è dedicato ad usare tecniche e tool di visualizzazione dei dati per comunicare graficamente i risultati dell'analisi consentendo, così, un'interpretazione efficace da parte degli utenti aziendali.

Gli utenti aziendali devono essere in grado di comprendere i risultati per poter trarre valore dall'analisi e, di conseguenza, avere la possibilità di fornire feedback.

I risultati ottenuti completando la fase di visualizzazione dei dati offrono agli utenti la possibilità di eseguire analisi visive, consentendo di scoprire risposte a domande che non hanno ancora nemmeno formulato.

Gli stessi risultati possono essere presentati in diversi modi, il che può influenzarne l'interpretazione.

Per questo motivo è fondamentale:

- scegliere la tecnica di visualizzazione più adatta;
- tenere conto del contesto di business.

Un ulteriore aspetto importante riguarda il *drill-down*, ossia la possibilità di partire da dati aggregati per poi scendere progressivamente nel dettaglio. Questo consente agli utenti di:

- comprendere come sono stati generati i risultati;
- confrontare statistiche semplici;
- approfondire l'analisi dei dati.

1.8.9 Utilization of Analysis Results

Una volta che i risultati dell'analisi sono stati resi disponibili agli utenti aziendali per supportare il processo decisionale, ad esempio tramite dashboard, possono emergere ulteriori opportunità di utilizzo dei risultati ottenuti.

A seconda della natura dei problemi affrontati, i risultati dell'analisi possono generare modelli che racchiudono nuove conoscenze sui pattern e sulle relazioni presenti nei dati analizzati.

Un modello può assumere la forma di un'equazione matematica oppure di un insieme di regole e può essere utilizzato per migliorare la logica dei processi aziendali e dei sistemi applicativi.

Le principali aree applicative esplorate in questa fase includono:

- *Input per i sistemi aziendali*: i risultati dell'analisi possono essere inseriti automaticamente o manualmente nei sistemi aziendali per migliorarne il comportamento e le prestazioni.
- *Ottimizzazione dei processi aziendali*: pattern, correlazioni e anomalie individuate, durante l'analisi possono essere utilizzati per migliorare e ottimizzare i processi aziendali.
- *Avvisi*: i risultati dell'analisi possono essere utilizzati per generare notifiche e avvisi automatici, ad esempio tramite e-mail o SMS, al fine di segnalare eventi che richiedono azioni correttive.

1.9 Piramide DIKW

La piramide DIKW, o piramide della conoscenza, è legata al concetto sempre più diffuso di *data-driven*. Oggi si parla spesso di aziende (ma anche di società) “guidate dai dati”, mettendo, così, in luce il valore dei dati nel processo che porta alla consapevolezza per poi prendere decisioni. Si tratta di un processo che parte dai dati grezzi e che arriva alla saggezza.

Questo processo è ben rappresentato dalla piramide:

- Data;
- Information;
- Knowledge;
- Wisdom.

Si tratta di un modello gerarchico che mostra i vari modi per estrarre intuizioni e valore.

La piramide della conoscenza è uno schema teorico volto a spiegare il processo conoscitivo e di apprendimento, che mette in relazione gerarchica in maniera lineare quattro elementi: dati, informazioni, conoscenza, saggezza.

Questo modello raffigura il processo della conoscenza come una piramide costituita da una base molto ampia di dati grezzi i quali, andando verso la cima della piramide, vengono sottoposti a un processo di aggregazione-contestualizzazione (informazione) e applicazione-sperimentazione (conoscenza). Infine, sulla cima della piramide si colloca la saggezza che presuppone un livello conoscitivo che esula dall’ambito di una specifica applicazione. Tali stati conoscitivi sono quindi stati connessi in modo gerarchico immaginando che tra essi ci possa essere una transizione ordinata dal basso verso l’alto.

Il passaggio da un livello all’altro rappresenta un processo in cui vengono aggiunti elementi di valore rispetto al livello precedente. Data, Information, Knowledge, and Wisdom è un modello per comprendere il libero flusso di dati all’interno di un’organizzazione, che lo rende utile nella gestione di progetti data-driven.

Nella Figura 1.12 vengono mostrati la piramide DIKW e il passaggio da un livello ad un altro.

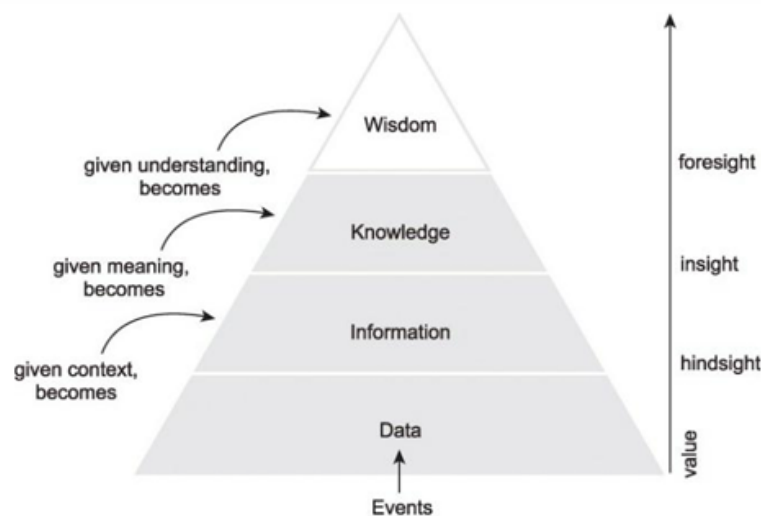


Figura 1.12: Piramide DIKW

Alla base della piramide troviamo i dati. Questi ultimi sono elementi grezzi, non ancora interpretati: numeri, valori ed eventi. Da soli, però, non hanno un significato completo.

Possiamo affermare che i dati grezzi siano contemporaneamente fondamentali e inutili:

- Fondamentali, perché rappresentano il punto di partenza del percorso che conduce alle informazioni, alla conoscenza e, infine, alla capacità di prendere decisioni corrette in ambito aziendale. Per questo motivo è essenziale che un'organizzazione sia in grado di raccogliere e tracciare i dati.
- Inutili, perché considerati singolarmente non possiedono un significato concreto e non generano alcun miglioramento. Si presentano semplicemente come numeri, valori o stringhe memorizzate all'interno di un database.

Questo apparente paradosso non dipende dai dati stessi, ma dall'utilizzo che viene fatto di essi. Finché i dati vengono soltanto raccolti senza strumenti adeguati per analizzarli, non è possibile estrarre informazioni utili né avanzare nei livelli superiori della piramide.

Diventa, quindi, necessario disporre di strumenti in grado di collegare, organizzare e analizzare i dati.

È proprio in questo contesto che entra in gioco la Business Intelligence, la quale permette di trasformare i dati grezzi in informazioni utili e di iniziare a generare valore dai dati raccolti.

- *Information (dai dati grezzi all'informazione)*

Un dato, di per sé, non ha un valore intrinseco fino a quando non viene contestualizzato; solo nel momento in cui viene messo a confronto con altri dati è possibile trarne delle conclusioni e quindi ottenere informazioni.

A questo livello di elaborazione si può trovare risposta a domande specifiche e puntuali come "cosa?", "dove?", "quando?", dando un primo significato ai dati raccolti.

Il lavoro del consulente di Business Intelligence consiste proprio nell'aggregare e contestualizzare dati per trasformarli in informazioni.

- *Knowledge (dalle informazioni alla conoscenza)*

Per la prima volta l'attenzione si sposta da qualcosa di tecnico e informatico (dati e sistemi di *data warehousing*) all'intervento di analisi umana.

L'output del lavoro della Business Intelligence è rappresentato da dashboard e report realizzati dai consulenti di BI mediante l'aggregazione di enormi quantità di dati. I risultati vengono presentati sotto forma di tabelle che collegano informazioni diverse oppure tramite grafici e dashboard sintetici.

Quando alle informazioni viene attribuito un significato più profondo e vengono collegate tra loro, si ottiene conoscenza.

Questo permette di comprendere situazioni e relazioni, non solo di descriverle.

- *Wisdom (Dai dati alla saggezza)*

La saggezza rappresenta la capacità di utilizzare la conoscenza per prendere decisioni corrette e consapevoli, basate su esperienza e giudizio.

La piramide mostra, quindi, un percorso evolutivo che conduce:

- dai dati grezzi;
- alle informazioni;

- alla conoscenza;
- fino alla saggezza.

Man mano che si sale nella piramide, aumenta il valore dei dati e cresce la capacità di supportare decisioni consapevoli ed efficaci.

Tuttavia, è fondamentale fare una distinzione importante: i sistemi informatici possono raccogliere ed elaborare dati, trasformarli in informazioni e, in parte, supportare la costruzione della conoscenza.

Oltre questo livello, però, diventa necessario l'intervento umano, poiché solamente le persone, grazie all'esperienza, all'intuizione e alle competenze maturate nel tempo, sono in grado di interpretare correttamente i risultati e prendere decisioni consapevoli.

Un aspetto particolarmente rilevante riguarda quindi il ruolo dei sistemi ICT all'interno della piramide DIKW:

- i sistemi informatici risultano molto efficaci nella gestione dei dati, delle informazioni e, parzialmente, della conoscenza;
- la saggezza, invece, non viene generata dalla tecnologia, ma deriva dall'esperienza e dalla capacità decisionale delle persone.

Come mostrato anche nella rappresentazione della piramide, il livello della saggezza viene spesso illustrato solamente tramite un contorno, proprio per evidenziare che essa esiste ma non può essere prodotta direttamente dai sistemi ICT.

I cosiddetti *knowledge worker*, infatti, forniscono intuizione, esperienza e capacità interpretativa, elementi indispensabili per contestualizzare la conoscenza disponibile e trasformarla in saggezza.

Per questo motivo, all'interno delle organizzazioni:

- la tecnologia supporta la raccolta, la gestione e l'analisi dei dati fino alla generazione della conoscenza;
- il personale umano rimane responsabile delle decisioni finali, facendo leva su competenze, esperienza e capacità di giudizio.

In conclusione, la piramide DIKW evidenzia come i dati, da soli, non siano sufficienti a generare valore: è necessario un processo progressivo di trasformazione che conduca dai dati grezzi fino alla conoscenza e, infine, alla saggezza, necessaria per supportare decisioni efficaci e strategiche.

Dal punto di vista aziendale, la piramide DIKW può essere utilizzata per descrivere il modo in cui un'organizzazione trasforma i dati raccolti durante le attività operative in conoscenza utile per supportare le decisioni strategiche. In questo contesto, l'azienda può essere vista come un sistema organizzato su tre livelli principali: operativo, manageriale e strategico.

Nella Figura 1.13 viene mostrato come i tre livelli di un'azienda comunicano seguendo un processo continuo.

Il livello operativo rappresenta lo strato più basso ed è quello in cui vengono eseguiti i processi aziendali quotidiani. In questa fase vengono generati grandi volumi di dati e metriche che descrivono ciò che sta accadendo all'interno dell'organizzazione. I dati raccolti vengono contestualizzati attraverso i processi di business e trasformati in informazioni.

Dati → Informazione

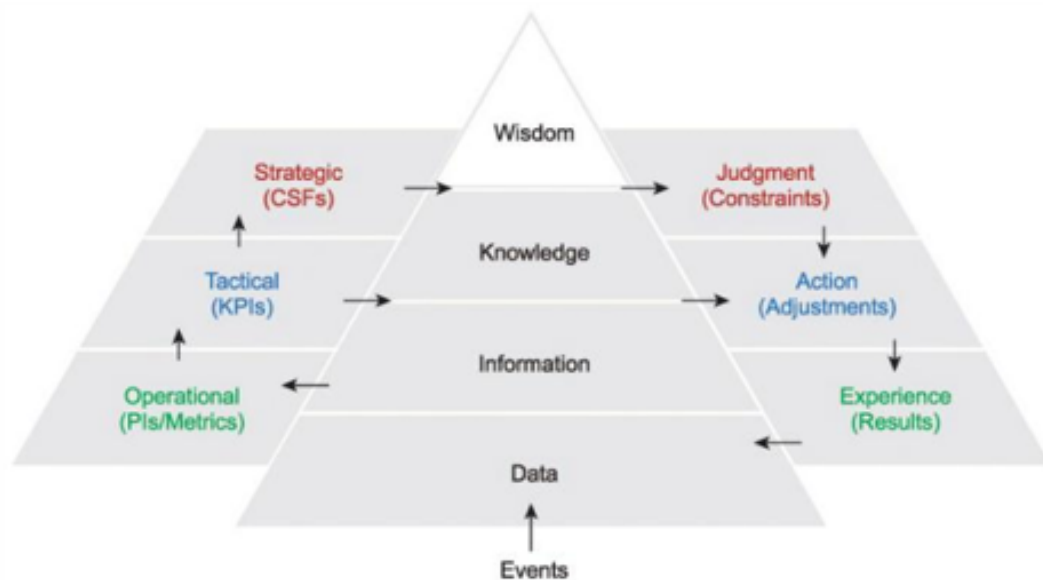


Figura 1.13: Processo aziendale

Dal basso verso l'alto, le informazioni prodotte vengono raccolte sotto forma di PI (Performance Indicator), che rappresentano indicatori relativi alle attività operative dell'azienda.

Il livello manageriale o tattico utilizza queste informazioni per monitorare l'andamento del business, analizzare le performance e comprendere le cause dei risultati ottenuti. In questa fase le informazioni vengono interpretate e acquistano significato, trasformandosi in conoscenza.

Informazione → Conoscenza

I PI vengono quindi aggregati in KPI (Key Performance Indicator), utilizzati dal management per valutare l'efficacia delle attività aziendali e verificare l'allineamento con gli obiettivi definiti dal livello strategico.

Il livello strategico rappresenta, invece, lo strato superiore dell'organizzazione ed è occupato dai dirigenti e dai gruppi decisionali. In questa fase la conoscenza viene integrata con esperienza, intuizione e capacità di giudizio, così da supportare le decisioni aziendali e la definizione delle strategie future. È in questo livello che avviene il passaggio dalla conoscenza alla saggezza.

Conoscenza → Saggezza

I KPI vengono confrontati con i CSF (Critical Success Factor) per comprendere se l'azienda stia raggiungendo i propri obiettivi strategici e per individuare eventuali modifiche da apportare alla strategia aziendale.

I tre livelli organizzativi non operano alla stessa velocità. Il livello operativo è il più rapido, poiché gestisce attività quotidiane e processi continui; il livello manageriale presenta una velocità intermedia; mentre il livello strategico è il più lento, ma fornisce stabilità, direzione e coordinamento all'intera organizzazione.

Inoltre, ogni livello utilizza differenti strumenti di monitoraggio e visualizzazione:

- il livello strategico usa le *balanced scorecard*;

- il livello manageriale utilizza le dashboard e l'analisi dei KPI;
- il livello operativo usa il monitoraggio dei processi.

Il funzionamento complessivo dell'organizzazione segue, quindi, un processo continuo. Le decisioni strategiche vengono comunicate al livello manageriale che le traduce in azioni operative da applicare ai processi aziendali. Le attività operative producono nuovi dati e metriche che vengono nuovamente raccolti, analizzati e trasformati in informazioni e conoscenza, generando, così, un ciclo continuo di miglioramento e feedback all'interno dell'azienda.

In questo capitolo verrà presentato Power BI, uno degli strumenti maggiormente utilizzati nell'ambito della Business Intelligence. Dopo una panoramica generale delle sue caratteristiche principali, saranno descritti i componenti che costituiscono l'infrastruttura della piattaforma e il relativo flusso operativo per la creazione, la pubblicazione e la condivisione dei report. Infine, verranno analizzate le principali visualizzazioni disponibili e le tecnologie di programmazione integrate nello strumento.

2.1 Introduzione a Power BI

Power BI è un'applicazione di Business Intelligence (BI) sviluppata da Microsoft. Il suo obiettivo è quello di supportare la Business Intelligence attraverso visualizzazioni interattive con un'interfaccia semplice, in modo da dare all'utente la possibilità di creare report e dashboard.

Power BI è composto da un insieme di software, servizi cloud e applicazioni che consentono di raccogliere, trasformare, analizzare e visualizzare i dati provenienti da diverse sorgenti.

Grazie alla connessione a un'ampia varietà di sorgenti, tra cui database, file Excel e servizi cloud, Power BI consente di aggregare e trasformare i dati grezzi in informazioni utilizzabili. Le sue funzionalità avanzate, come la modellazione dei dati e l'analisi in tempo reale, realizzano visualizzazioni interattive e report personalizzati.

Power BI favorisce la collaborazione tra gruppi di lavoro perché rende accessibili le dashboard e i report su diversi dispositivi. Esso rappresenta uno strumento particolarmente diffuso nelle organizzazioni che adottano un approccio data-driven, favorendo una maggiore comprensione dei dati e il monitoraggio delle attività aziendali.

A differenza di altri strumenti di analisi, Power BI è accessibile sia ad utenti tecnici sia a manager non esperti, grazie a un'interfaccia intuitiva e a funzioni di drag-and-drop.

Infatti, sia che si tratti di un utente aziendale, di un autore di report o di uno sviluppatore, Power BI offre strumenti e servizi integrati per connettersi, visualizzare e condividere i dati nell'organizzazione.

Power BI offre numerosi vantaggi che ne hanno favorito l'ampia diffusione nell'ambito della Business Intelligence. Tra i principali si possono evidenziare i seguenti:

- *Diffusione e affidabilità della piattaforma:* essendo sviluppato da Microsoft, Power BI rappresenta una delle soluzioni di Business Intelligence più utilizzate a livello mondiale, sia in ambito aziendale sia accademico.
- *Facilità d'uso:* l'interfaccia intuitiva e le funzionalità di tipo drag-and-drop consentono anche agli utenti con limitate competenze tecniche di realizzare report e dashboard interattive.
- *Ampia disponibilità di risorse e supporto:* la presenza di una vasta comunità di utenti, di documentazione ufficiale e di contenuti formativi facilita l'apprendimento e la risoluzione delle problematiche più comuni.
- *Qualità delle visualizzazioni:* Power BI mette a disposizione numerosi strumenti grafici che consentono di rappresentare i dati in modo chiaro, efficace e interattivo, favorendo l'analisi e l'interpretazione delle informazioni.

Nonostante i numerosi punti di forza, Power BI presenta anche alcune limitazioni:

- *Costi delle funzionalità avanzate:* sebbene sia disponibile una versione gratuita, alcune funzionalità avanzate richiedono l'acquisto di licenze specifiche, con conseguente incremento dei costi per le organizzazioni.
- *Complessità nella gestione di progetti estesi:* in contesti caratterizzati da numerosi report e dataset, la gestione delle risorse può diventare complessa e richiedere adeguate strategie di governance.
- *Dipendenza dall'ecosistema Microsoft:* l'integrazione ottimale con i servizi Microsoft rappresenta un vantaggio, ma può limitare la flessibilità in ambienti basati su piattaforme differenti.
- *Limitazioni prestazionali con dataset molto complessi:* all'aumentare della quantità e della complessità dei dati, possono emergere problematiche legate ai tempi di elaborazione e alle prestazioni delle analisi.

2.2 Infrastruttura di Power BI

Come già detto, Power BI rappresenta una soluzione avanzata di Business Intelligence costituita da un insieme di strumenti e servizi interconnessi, progettati per offrire un'esperienza di analisi dei dati integrata, flessibile e fruibile su molteplici dispositivi.

Tra i principali servizi offerti da Power BI troviamo:

- Power BI Desktop;
- Power BI Service;
- Power BI Gateway;
- Power BI Mobile Apps;
- Power BI Visuals Marketplace.

2.2.1 Architettura

Power BI Desktop

Power BI Desktop è l'applicazione gratuita di Microsoft dedicata all'analisi dei dati e alla creazione di report interattivi. L'applicazione può essere utilizzata per:

- connettersi a diverse fonti di dati;
- modellare e trasformare i dati tramite Power Query, al fine di pulire i dati grezzi, rimuovere errori, formattare le tabelle e integrare le informazioni in un unico modello coerente;
- creare misure DAX (Data Analysis Expressions) per ottenere indicatori e informazioni personalizzate;
- progettare report interattivi, dashboard e visualizzazioni.

Power BI Service

Power BI Service è la componente cloud di Power BI accessibile tramite Browser.

Serve a pubblicare, organizzare e condividere in sicurezza i report interattivi creati, permettendo al team di visualizzare dashboard aggiornate in tempo reale da qualsiasi browser o dispositivo mobile.

In particolare, Power BI Service consente di:

- pubblicare e condividere report creati con Power BI Desktop;
- creare dashboard personalizzate in real-time;
- collaborare con altri utenti.

Power BI Gateway

Power BI Gateway, sviluppato da Microsoft, è un'importante soluzione che facilita il collegamento tra le sorgenti di dati locali e Power BI Service.

Questo strumento, disponibile gratuitamente, riveste un ruolo fondamentale nel consentire agli utenti di importare e condividere agevolmente le informazioni archiviate localmente attraverso i report di Power BI, garantendo, al contempo, elevati standard di sicurezza.

Il gateway funge da ponte tra i servizi cloud di Microsoft e l'infrastruttura on-premise di un'azienda, che comprende software e risorse fisiche.

Esso offre la possibilità di configurare le modalità di trasferimento dei dati tra le sorgenti locali e i servizi cloud. A tale scopo, Power BI mette a disposizione due differenti versioni del gateway:

- *Versione Personale di Power BI Gateway*; progettata per i singoli utenti, consente di mantenere aggiornati i dati importati in Power BI attraverso query eseguite direttamente tra la piattaforma cloud e il dispositivo personale dell'utente.
- *Versione Standard di Power BI Gateway*; pensata per ambienti aziendali e multiutente, permette il collegamento a diverse sorgenti di dati *on-premise*. Inoltre, può essere installata su più macchine per creare un cluster locale, garantendo ridondanza e maggiore affidabilità del servizio. Questa soluzione è particolarmente adatta a scenari complessi in cui disponibilità e continuità operativa rivestono un ruolo fondamentale.

Power BI Gateway presenta diversi vantaggi e alcune limitazioni che devono essere considerati durante il suo utilizzo.

I principali vantaggi di Power BI Gateway sono:

- *Programmazione dell'aggiornamento dei dati*: consente di pianificare aggiornamenti su base oraria, giornaliera o settimanale e supporta aggiornamenti in tempo reale tramite la modalità *DirectQuery*, garantendo la disponibilità di dati sempre aggiornati per le analisi.
- *Controllo centralizzato delle connessioni*: permette di gestire tutte le connessioni ai dati aziendali da un unico punto, migliorando il controllo, la sicurezza e il monitoraggio degli accessi alle sorgenti di dati.

Tra le principali limitazioni si possono evidenziare:

- *Necessità di una connessione Internet*: il trasferimento dei dati dalle sorgenti locali al servizio Power BI richiede una connessione Internet attiva.
- *Numero limitato di refresh*: la frequenza di aggiornamenti dipende dalla licenza Power BI utilizzata e dalla tipologia di sorgente dati connessa, rappresentando un possibile vincolo in scenari con elevati volumi di dati o necessità di aggiornamenti frequenti.

Power BI Mobile Apps

Power BI Mobile Apps comprende le applicazioni disponibili per smartphone e tablet (iOS, Android, Windows).

Queste App consentono agli utenti di accedere ai contenuti di Power BI direttamente nei loro browser o su dispositivi mobili tramite un semplice link, eliminando la necessità di navigare nella struttura Workspace all'interno dell'interfaccia Microsoft Fabric.

Il vantaggio dell'utilizzo delle Power BI App risiede nella loro capacità di presentare agli utenti finali una vista personalizzata che evidenzia solo i contenuti finalizzati e verificati. Inoltre, le app sono impostate in modalità di sola lettura, rendendole perfette per gli utenti che non necessitano di modificare il contenuto.

Quindi, una Power BI App funziona come un'interfaccia consolidata per accedere ai report all'interno di uno spazio di lavoro, consentendo agli utenti di passare agevolmente tra i report senza la necessità di aprire più finestre del browser.

Power BI Visuals Marketplace

Power BI Visuals Marketplace è uno store integrato che offre una vasta gamma di grafici e strumenti di visualizzazione aggiuntivi rispetto a quelli predefiniti in Power BI. È un portale online dove è possibile scaricare e integrare visualizzazioni personalizzate create dalla community o da partner Microsoft.

L'utilizzo di visualizzazioni personalizzate consente di ampliare le possibilità di rappresentazione dei dati, adattando i report a specifiche esigenze analitiche e di comunicazione. Le visualizzazioni disponibili possono essere installate direttamente all'interno dell'ambiente Power BI e successivamente utilizzate nella progettazione di report e dashboard interattive.

Nella Figura 2.1 viene mostrata l'infrastruttura di Power BI.

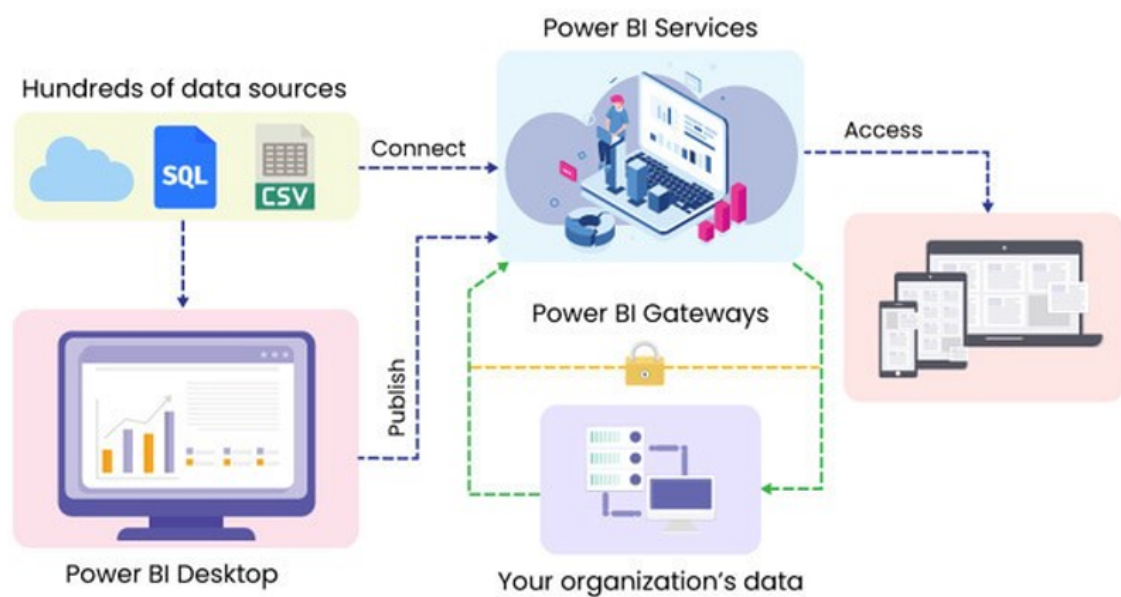


Figura 2.1: Infrastruttura di Power BI

2.2.2 Flusso di lavoro

Il flusso di lavoro tipico di Power BI comprende le seguenti attività:

- connettersi ai dati;
- trasformare e pulire i dati per creare un modello di dati (Data Cleaning);
- creare oggetti visivi che forniscano rappresentazioni dei dati;
- creare report che raccolgano gli oggetti visivi in una o più pagine di report;
- condividere i report con altri utenti tramite il servizio Power BI.

Ciascuna delle attività elencate svolge un ruolo specifico nel processo di trasformazione e analisi dei dati. Di seguito viene fornita una descrizione delle singole fasi.

Connettersi ai dati

Con Power BI Desktop è possibile connettersi a un'ampia varietà di sorgenti dei dati. La piattaforma supporta file locali, come Excel, CSV, XML e JSON, database relazionali, quali SQL Server, MySQL, Oracle e PostgreSQL, servizi cloud, come SharePoint, OneDrive, Azure, Google Analytics e Salesforce, oltre ad API e sorgenti dati personalizzate.

La possibilità di integrare dati provenienti da fonti eterogenee rappresenta uno dei principali punti di forza di Power BI, poiché consente di centralizzare informazioni distribuite in sistemi differenti all'interno di un unico ambiente di analisi. Una volta stabilita la connessione, i dati possono essere importati, trasformati e integrati in un modello di dati coerente, pronto per le successive attività di elaborazione e visualizzazione.

La Figura 2.2 mostra un esempio delle sorgenti di dati che possono essere gestiti in Power BI Desktop attraverso la funzione *Recupera dati*.

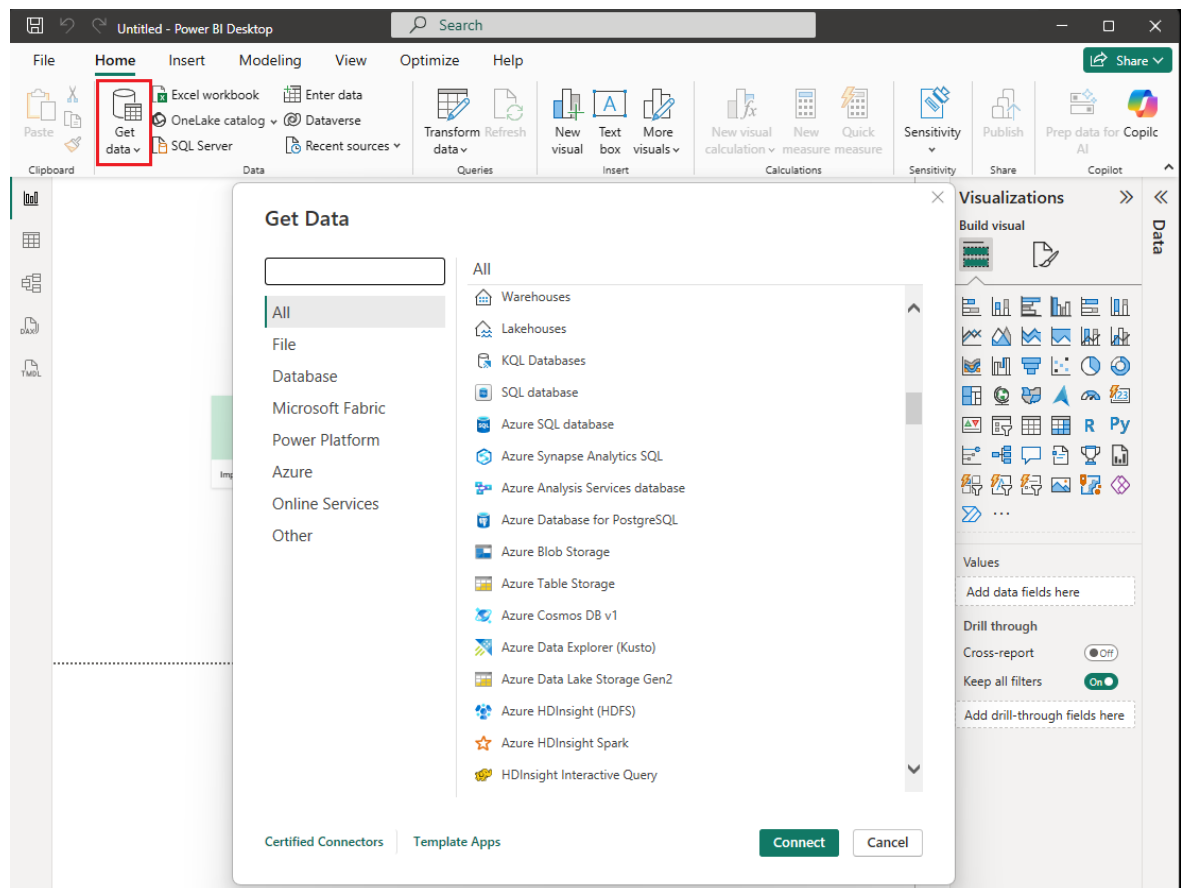


Figura 2.2: Sorgenti di dati gestibili da Power BI

Data Cleaning

In Power BI è possibile realizzare operazioni di ETL sui dati attraverso l'Editor di query, che permette di estrarre, trasformare e pulire i dati per poterli analizzare correttamente.

Le principali operazioni di trasformazione dei dati permettono di:

- rinominare le tabelle;
- unire le tabelle;
- creare colonne calcolate o condizionali;
- eliminare colonne non necessarie;
- filtrare righe e colonne;
- modificare il tipo di dato;
- correggere errori e rimuovere record duplicati;
- sostituire valori specifici;
- gestire eventuali valori nulli sui dati.

Power BI Desktop mette a disposizione l'editor di Power Query. Con questo editor è possibile connettersi a una o più sorgenti di dati, eseguire il data shaping e trasformare i dati in base alle esigenze, per poi caricare il modello in Power BI Desktop.

Per accedere all'editor di Power Query, è necessario selezionare il pulsante *Trasforma dati* presente nella scheda Home di Power BI Desktop, come mostrato nella Figura 2.3.

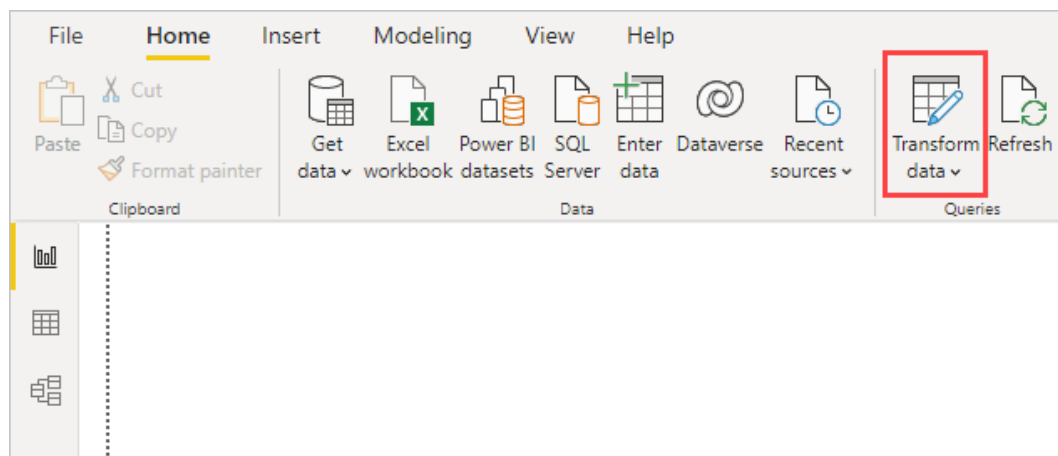


Figura 2.3: Trasformazione dei dati

Dopo aver stabilito una connessione ai dati, l'editor di Power Query presenta quattro principali componenti dell'interfaccia:

1. *Barra multifunzione*: contiene strumenti e comandi che consentono di interagire con i dati e applicare operazioni di trasformazione alle query.
2. *Riquadro dei dataset*: situato nella parte sinistra dell'interfaccia, mostra l'elenco dei dataset disponibili, che possono essere selezionati, visualizzati e modificati.
3. *Area centrale*: visualizza i dati relativi al dataset selezionato e rappresenta l'area principale in cui vengono effettuate le operazioni di modellazione e trasformazione dei dati.
4. *Riquadro Impostazioni query*: riporta le proprietà del dataset e l'elenco dei passaggi applicati, consentendo di monitorare e modificare le trasformazioni eseguite.

Nella Figura 2.4 sono mostrate le sezioni sopra elencate.

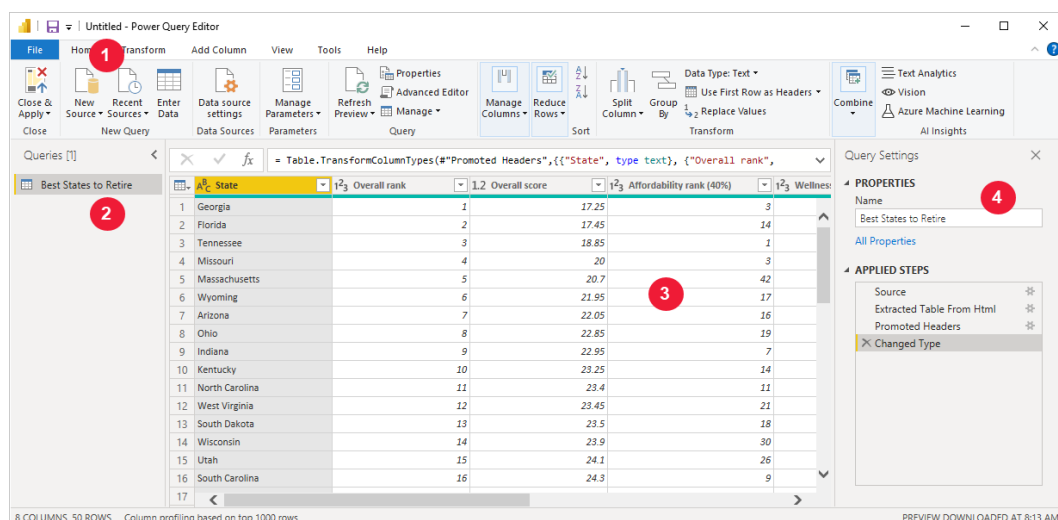


Figura 2.4: Sezioni relative alla trasformazione dei dati

In Power BI è presente anche un editor avanzato che consente di visualizzare il codice generato automaticamente da Power Query durante le operazioni di trasformazione dei dati. Ogni passaggio applicato ai dati viene, infatti, tradotto in istruzioni scritte nel linguaggio M, permettendo di comprendere nel dettaglio le trasformazioni eseguite.

L'editor avanzato offre, inoltre, la possibilità di modificare direttamente il codice generato o di scrivere istruzioni personalizzate nel linguaggio M, consentendo di realizzare trasformazioni più complesse rispetto a quelle disponibili tramite l'interfaccia grafica. Questa funzionalità risulta particolarmente utile quando è necessario automatizzare operazioni specifiche o implementare logiche di trasformazione avanzate.

L'utilizzo del linguaggio M consente, inoltre, di rendere il processo di preparazione dei dati più trasparente e facilmente replicabile. Le trasformazioni applicate vengono, infatti, salvate come una sequenza di istruzioni che può essere modificata, riutilizzata e documentata nel tempo.

Creare oggetti visivi

Dopo avere creato un modello di dati, i campi possono essere trascinati nell'area di disegno del report per realizzare oggetti visivi. Nel pannello Visualizzazioni sono disponibili numerosi oggetti visivi utilizzabili per la progettazione di report e dashboard, tra cui:

- grafici di diverse tipologie (a barre, a linee e a torta);
- tabelle, matrici e mappe geografiche;
- indicatori di KPI, slicer e timeline;
- visualizzazioni personalizzate disponibili nel marketplace di Power BI.

La Figura 2.5 mostra gli oggetti visivi disponibili in Power BI Desktop.

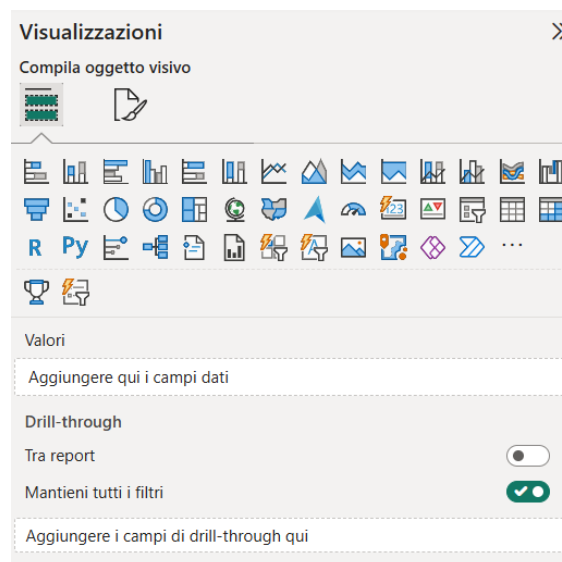


Figura 2.5: Oggetti visivi disponibili in Power BI Desktop

Creare report

La fase successiva consiste nella creazione di una raccolta di oggetti visivi che illustrano i diversi aspetti dei dati utilizzati per la costruzione del modello in Power BI. Una raccolta di oggetti visivi prende il nome di *report*.

Quindi, è possibile utilizzare diversi elementi grafici per realizzare un report interattivo attraverso il quale l'utente può esplorare e analizzare le informazioni disponibili. I report possono essere organizzati in una o più pagine e consentono un'esplorazione dettagliata dei dati mediante filtri, slicer e altre funzionalità di analisi.

Diversamente, una *dashboard* è costituita da un'unica pagina che raccoglie gli elementi più significativi. Il suo scopo principale è fornire una panoramica sintetica e immediata delle informazioni più rilevanti, risultando particolarmente utile per il monitoraggio delle prestazioni e il supporto ai processi decisionali.

Una dashboard di Power BI, spesso chiamata *canvas*, consente di presentare in modo sintetico le informazioni principali attraverso rappresentazioni grafiche dei dati. Una dashboard ben progettata evidenzia gli aspetti più importanti dell'analisi, mentre i dettagli possono essere approfonditi consultando i report correlati.

Grazie alla loro struttura sintetica, le dashboard permettono di individuare rapidamente indicatori chiave e informazioni rilevanti senza dover analizzare grandi quantità di dati. Per questo motivo vengono frequentemente utilizzate per il monitoraggio continuo delle prestazioni aziendali. Le visualizzazioni presenti all'interno di una dashboard sono chiamate riquadri. Questi ultimi derivano dai report e sono basati sui relativi set di dati. Una dashboard può, pertanto, essere considerata una via d'accesso ai report e ai modelli dei dati sottostanti.

Un esempio di report è mostrato nella Figura 2.6, mentre la Figura 2.7 riporta una sintesi delle principali differenze tra dashboard e report.

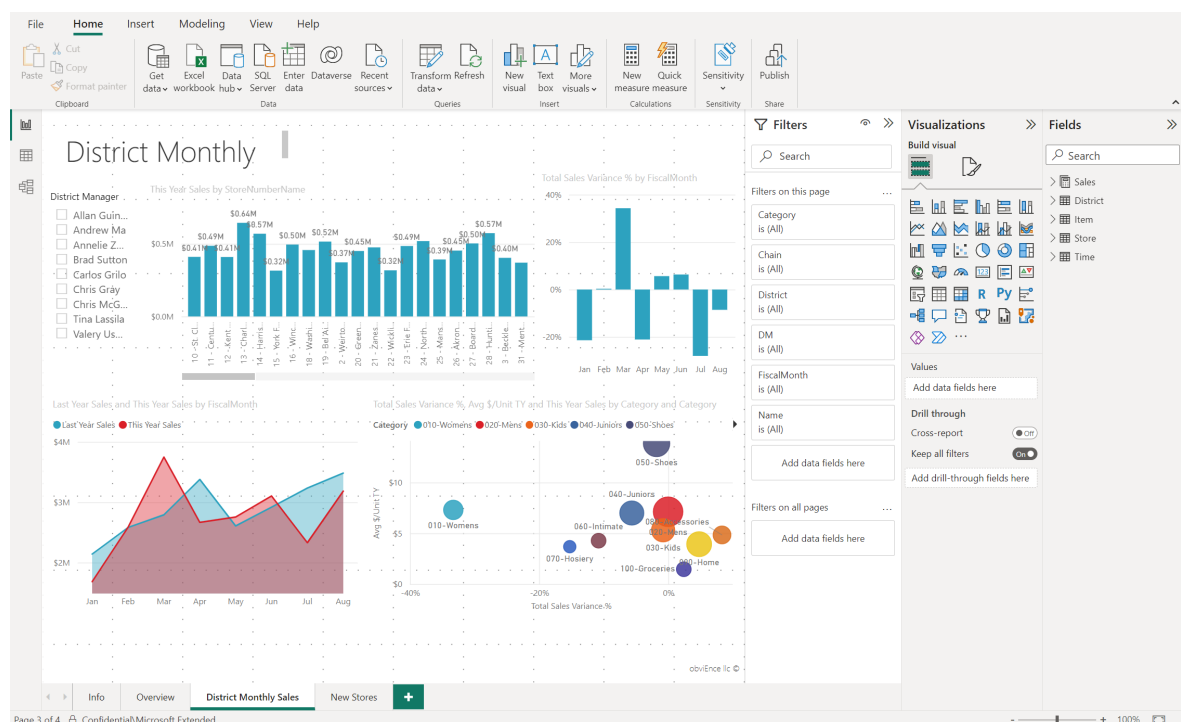


Figura 2.6: Esempio di report

Condividere report

Una volta che un report è pronto per essere condiviso con altri utenti, è possibile usare il comando **Pubblica** per pubblicare il report nel Power BI Service.

Il flusso di lavoro di Power BI consente, quindi, di trasformare dati provenienti da sorgenti eterogenee in report e dashboard interattive, facilitando il processo decisionale basato sui dati.

	Dashboard	Report
Pagina	Una pagina	Una o più pagine
Origini dati	Uno o più report e uno o più set di dati per dashboard	Un singolo set di dati per report
Disponibile in Power BI Desktop	No	Sì
Applicazione di filtri	No. Non è possibile filtrare o sezionare	Sì. Molti modi diversi di filtrare ed evidenziare
Impostazione di avvisi	Sì. Disponibile per i riquadri del dashboard in determinate circostanze	No
Permette di visualizzare i campi e le tabelle del set di dati sottostante	No	Sì

Figura 2.7: Differenza tra dashboard e report

2.3 Visualizzazioni di Power BI

Power BI Desktop mette a disposizione tre diverse visualizzazioni, ciascuna dedicata a una specifica fase del processo di analisi dei dati:

- *Visualizzazione Report*: consente di creare e organizzare visualizzazioni interattive a partire dai dati elaborati, strutturandole in una o più pagine di report che possono essere successivamente condivise con altri utenti.
- *Visualizzazione Dati*: permette di esplorare i dati sotto forma di modello tabellare, offrendo la possibilità di creare misure, aggiungere nuove colonne e gestire le relazioni tra le diverse tabelle.
- *Visualizzazione Relazioni*: fornisce una rappresentazione grafica del modello dei dati e delle relazioni esistenti tra le tabelle, consentendo di visualizzarle, modificarle e gestirle in modo intuitivo.

Per accedere a queste visualizzazioni è necessario selezionare una delle tre icone presenti nella barra laterale sinistra di Power BI Desktop. Ciascuna icona consente di passare rapidamente da una modalità di lavoro all'altra, permettendo di visualizzare i dati da prospettive differenti e di svolgere specifiche attività di analisi e modellazione.

Nella Figura 2.8 viene mostrata la barra laterale sinistra di Power BI Desktop, nella quale è selezionata la visualizzazione Report, identificata dalla barra evidenziata accanto alla relativa icona.

Visualizzazione report

Power BI Desktop include una visualizzazione dei report che consente di creare pagine interattive contenenti differenti visualizzazioni dei dati e di progettare dashboard personalizzate. Grazie a queste visualizzazioni, gli utenti possono esplorare, analizzare e sintetizzare i dati in modo efficace, individuando pattern, relazioni e informazioni significative che difficilmente riuscirebbero a individuare dall'analisi dei dati grezzi.

La Figura 2.9 mostra la sezione dedicata alla visualizzazione dei report.

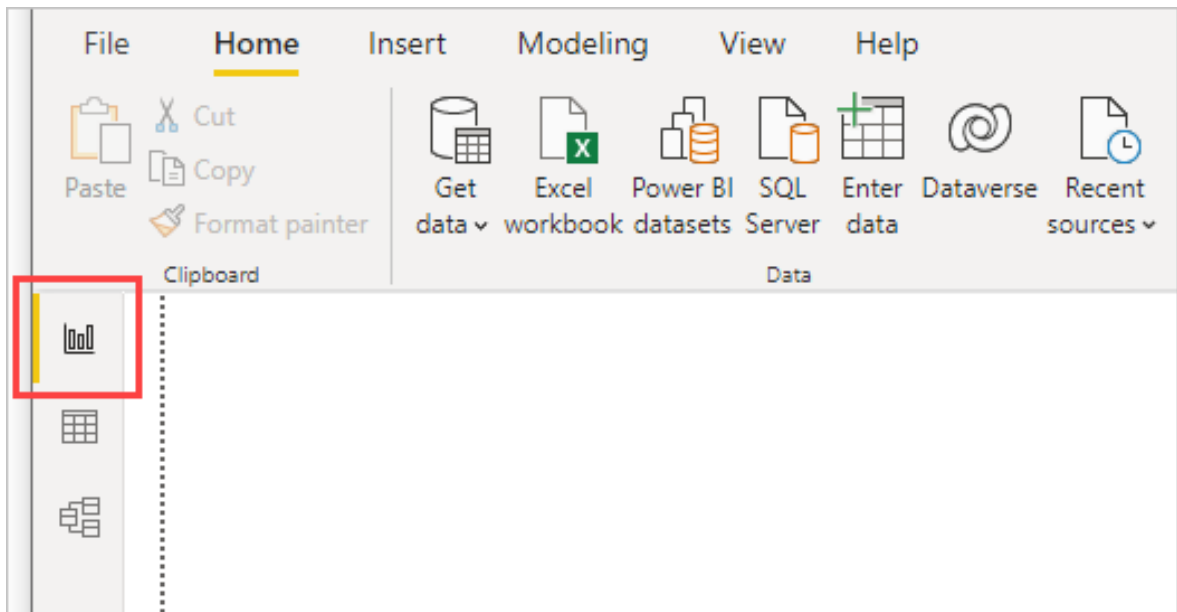


Figura 2.8: Lato sinistro di Power BI Desktop per le visualizzazioni

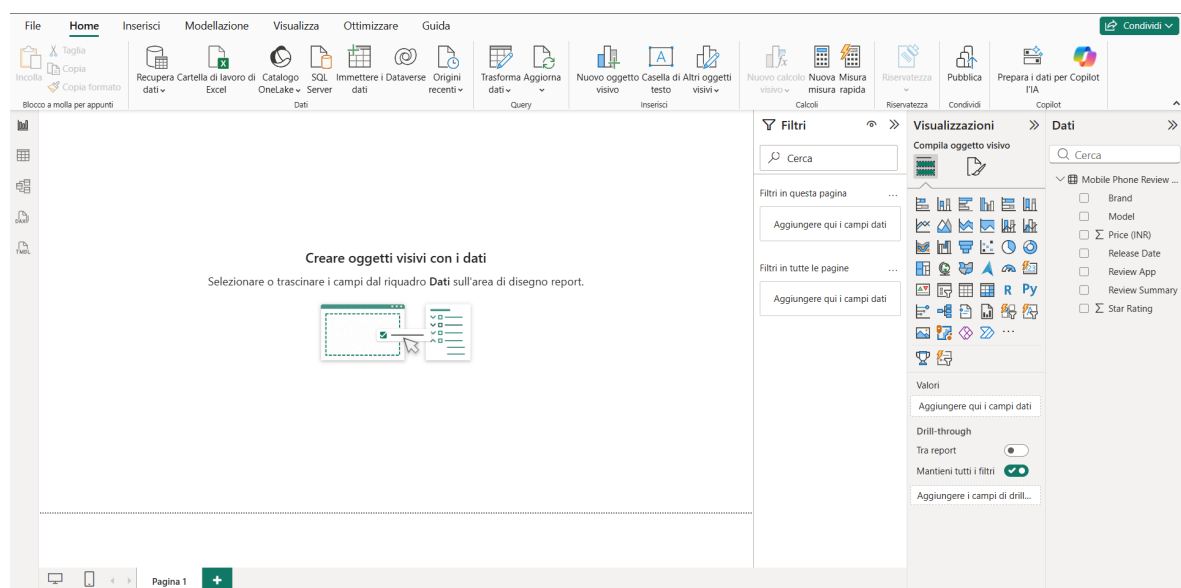


Figura 2.9: Visualizzazione dei report

La pagina di visualizzazione dei report è organizzata in tre sezioni laterali che aiutano la generalizzazione dei report visivi e la costruzione di oggetti visivi:

- *Sezione Filtri:* consente di applicare specifici filtri a livello di report, pagina o singoli oggetti visivi, limitando la visualizzazione dei dati alle condizioni definite dall'utente.
- *Sezione Visualizzazioni:* contiene l'elenco degli oggetti visivi disponibili per la realizzazione dei report. Questa sezione è, a sua volta, suddivisa nelle seguenti sottosezioni:
 - *Campi;* è l'area nella quale vengono trascinate le colonne e le misure da visualizzare nel grafico, distinguendo tra valori, categorie e legende.

- *Formato*; consente di personalizzare l'aspetto dell'oggetto visivo selezionato, modificando attributi quali colori, stile, titoli, etichette, dimensioni del testo, bordi e sfondo.
- *Analisi*; permette di inserire elementi di analisi automatica, come linee di media, linee di tendenza, intervalli di riferimento o costanti, utili per evidenziare pattern e comportamenti nei dati.
- *Sezione Dati*: mostra i modelli dei dati disponibili e le relative colonne utilizzabili per la realizzazione dei componenti visivi nei report. In questa sezione vengono, inoltre, visualizzate le misure create tramite DAX e le colonne calcolate, rendendo possibile la gestione completa degli elementi che compongono il modello dei dati.

La Figura 2.10 mostra le tre sezioni laterali della visualizzazione Report.

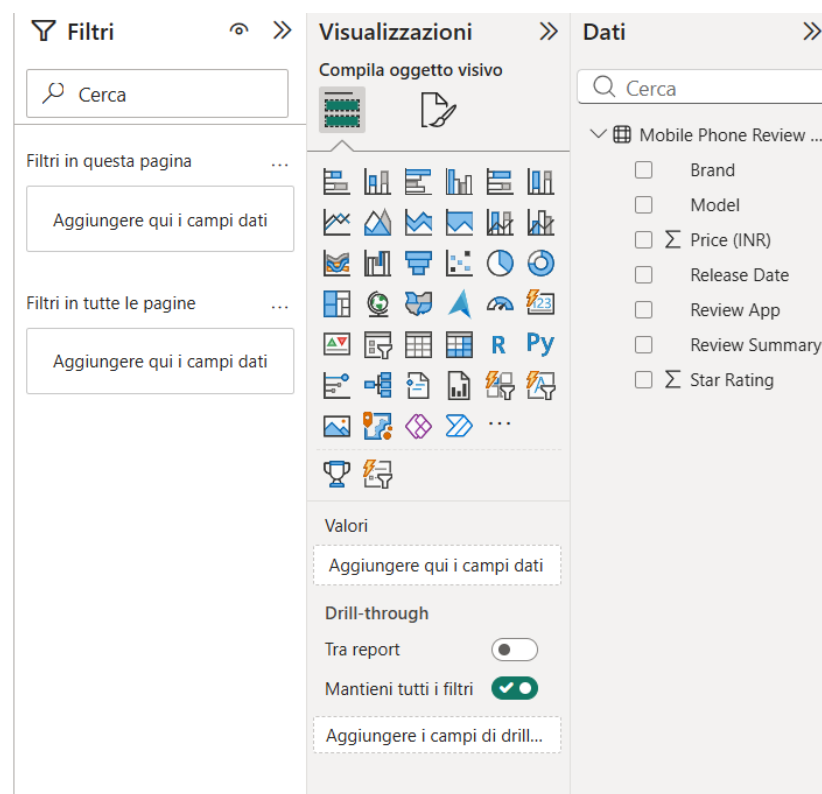


Figura 2.10: Sezioni laterali relative alla visualizzazione dei report

Visualizzazione Dati

La Visualizzazione Dati consente di esaminare, esplorare e comprendere i dati presenti nel modello di Power BI Desktop dopo il loro caricamento. Essa si differenzia dall'editor di Power Query, che viene utilizzato per eseguire operazioni di estrazione, trasformazione e pulizia dei dati prima del loro inserimento nel modello.

Attraverso questa visualizzazione è possibile consultare direttamente le informazioni presenti nel modello dei dati e verificarne il contenuto senza dover creare preventivamente un oggetto visivo nell'area di disegno del report.

Durante la fase di modellazione può risultare utile analizzare i dati al termine del processo di ETL, verificando il risultato delle trasformazioni applicate e la corretta presenza delle informazioni all'interno del modello. Questa funzionalità risulta particolarmente utile

durante la creazione di misure e colonne calcolate, nonché per l'identificazione di eventuali anomalie nei dataset.

La Visualizzazione Dati è particolarmente utile quando si desidera:

- controllare il risultato delle trasformazioni applicate tramite Power Query;
- verificare i dati calcolati o derivati da colonne personalizzate;
- individuare eventuali valori nulli o anomali nei dataset.

La Figura 2.11 mostra la sezione dedicata alla Visualizzazione Dati.

Brand	Model	Review Summary	Star Rating	Release Date	Price (INR)	Review App
Samsung	Galaxy S22	Excellent camera and performance	49	2/25/2022	66337	GSMARena
Apple	iPhone 14	Great build, but not a huge upgrade	48	9/16/2022	74617	TechRadar
Asus	ROG Phone 6	Best for gaming, bulky design	47	7/5/2022	62917	PhoneArena
OnePlus	Nord 2T	Fast charging and clean UI	46	5/19/2022	31457	PhoneArena
Xiaomi	Redmi Note 11	Value for money, solid battery life	45	1/26/2022	16517	Android Authority
Realme	GT Neo 3	Excellent gaming performance	44	3/22/2022	38927	GSMARena
OpPO	Reno8	Good all-rounder, competitive pricing	43	7/18/2022	35607	NDTV Gadgets
Vivo	V25 Pro	Stylish design, decent camera	42	8/17/2022	33117	91Mobiles
Motorola	Edge 30	Lightweight and smooth display	41	5/12/2022	28967	Android Central
Lenovo	Legion Y90	Extreme performance, heavy build	4	3/1/2022	53867	Android Central
Sony	Xperia 10 IV	Great battery, outdated design	39	6/1/2022	35607	GSMARena
Infinix	Zero Ultra	Great charging speed, average camera	38	10/5/2022	24837	TechRadar
Tecno	Camon 19 Pro	Solid camera at low price	37	6/15/2022	18177	Android Authority
Lava	Blaze 5G	Affordable 5G option	36	11/3/2022	14857	91Mobiles
Nokia	X30	Eco-friendly build, stock Android	35	9/1/2022	28137	GSMARena
Google	Pixel 7	Excellent software and camera	47	10/13/2022	45999	GSMARena
Samsung	Galaxy A53	Good mid-range with AMOLED display	43	3/17/2022	27999	TechRadar
Apple	iPhone SE 2022	Compact, powerful for price	44	3/18/2022	42999	PhoneArena
OnePlus	10T	Flagship performance, fast charging	45	8/3/2022	42999	NDTV Gadgets
Xiaomi	POCO X4 Pro	Strong specs for budget	42	1/10/2022	17999	Android Authority
Realme	10 Pro+	Good camera and design	41	9/14/2022	24999	91Mobiles
Vivo	Y75 5G	Affordable 5G with good battery	39	5/10/2022	14999	GSMARena
Motorola	G Stylus 5G (2022)	Stylus support, budget-friendly	38	4/28/2022	17999	TechRadar
Nokia	G11 Plus	Long battery life, clean software	37	7/7/2022	9999	Android Authority
Infinix	Note 12 Pro	Good display and fast charging	4	8/25/2022	15999	PhoneArena
Tecno	Spark 9T	Fast charging, basic camera	36	3/15/2022	8999	91Mobiles
Lava	Agni 2	Entry-level, decent for calls	35	1/5/2022	6999	GSMARena

Figura 2.11: Visualizzazione Dati in Power BI

Visualizzazione Relazioni

La Visualizzazione Relazioni consente di rappresentare graficamente la struttura del modello dei dati e di definire i collegamenti tra le diverse tabelle. Attraverso questa modalità è possibile comprendere come le informazioni provenienti dai vari dataset siano connesse tra loro e come vengano utilizzate nelle analisi e nei report.

Questa vista risulta particolarmente utile nei modelli caratterizzati da numerose sorgenti di dati, poiché permette di avere una visione d'insieme dell'organizzazione del modello e dei collegamenti esistenti tra i diversi elementi.

Ogni tabella viene rappresentata mediante un riquadro contenente le relative colonne, mentre i collegamenti tra le tabelle sono visualizzati tramite linee che uniscono i campi chiave coinvolti nelle relazioni.

La Visualizzazione Relazioni consente, inoltre, di creare e modificare le relazioni esistenti, definendone le principali proprietà, come la cardinalità (uno-a-molti o molti-a-uno) e la direzione del filtro incrociato (*cross filter direction*).

La Figura 2.12 mostra un esempio della Visualizzazione Relazioni, evidenziando i collegamenti definiti tra le diverse tabelle del modello dei dati.

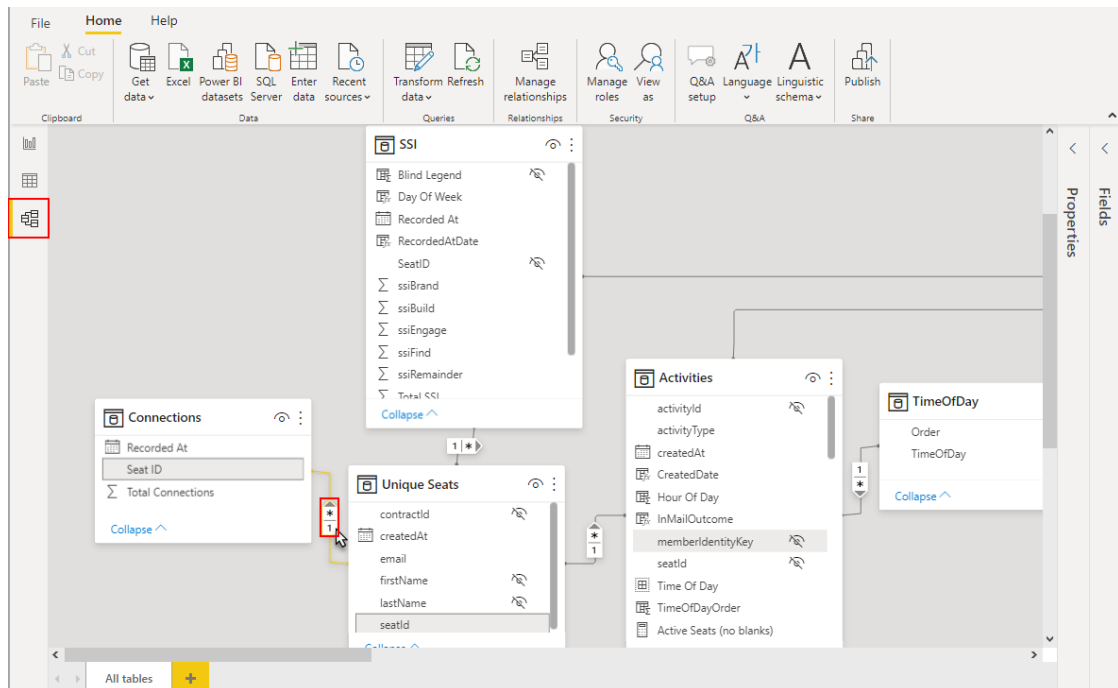


Figura 2.12: Visualizzazione delle Relazioni

2.4 Tecnologie di programmazione integrate

Power BI integra diversi linguaggi di programmazione che consentono di estendere le funzionalità della piattaforma nelle attività di trasformazione, modellazione e analisi dei dati. Tra i principali linguaggi supportati vi sono DAX (Data Analysis Expressions), M (Power Query Formula Language), Python e R.

Ciascuno di essi svolge un ruolo specifico all'interno dell'ambiente Power BI, contribuendo, rispettivamente, alla realizzazione di calcoli avanzati, alle operazioni di estrazione e trasformazione dei dati e all'integrazione di funzionalità statistiche e di analisi avanzata. Più specificatamente:

- *DAX (Data Analysis Expressions)* è un linguaggio costituito da funzioni, operatori e costanti utilizzati per creare formule ed espressioni in grado di calcolare e restituire uno o più valori. Esso consente di generare nuove informazioni a partire dai dati già presenti nel modello attraverso calcoli, aggregazioni e filtri avanzati.
- *M (Power Query Formula Language)* è il linguaggio utilizzato da Power Query per le operazioni di estrazione, trasformazione e caricamento dei dati (ETL). Consente di filtrare, modificare, pulire e combinare dati provenienti da sorgenti differenti. Ogni trasformazione eseguita tramite l'interfaccia grafica viene automaticamente tradotta in codice M, che può essere visualizzato e modificato attraverso l'editor avanzato.
- *R e Python*: si tratta di linguaggi ampiamente utilizzati nell'ambito dell'analisi dei dati e pienamente integrati in Power BI. Essi consentono di sviluppare analisi statistiche avanzate, modelli predittivi e soluzioni di machine learning, oltre alla realizzazione di visualizzazioni personalizzate mediante script dedicati.

Descrizione dei dataset di partenza

In questo capitolo verrà presentato il processo di ETL descrivendo nel dettaglio le sue fasi di estrazione, trasformazione e caricamento. Successivamente, verranno illustrate le principali differenze del processo ELT rispetto al classico processo ETL. Infine, verrà mostrato come il processo ETL sia stato applicato ai dataset scelti dalla piattaforma Kaggle allo scopo di creare un dataset affidabile e coerente per una successiva analisi con l'obiettivo di creare dashboard descrittive e predittive.

3.1 Dataset utilizzati

Per l'analisi condotta tramite Power BI sono stati utilizzati due dataset pubblici reperibili sulla piattaforma Kaggle, contenenti informazioni relative al mercato degli smartphone. Entrambi i dataset sono costituiti da dati strutturati, organizzati in formato CSV, e comprendono specifiche tecniche, caratteristiche hardware e informazioni commerciali relative a diversi modelli di dispositivi mobili.

3.1.1 Dataset 1: Smartprix Smartphones Dataset

Il primo dataset utilizzato, denominato *Latest Smartphones 2026: Price, Specs & VFM Score*, è basato su dati relativi a circa 1000 smartphone, con informazioni riguardanti il prezzo, le specifiche tecniche e un indicatore personalizzato di rapporto qualità-prezzo, denominato *Value-for-Money Score*. Il dataset è stato costruito a partire da dati provenienti dal sito Smartprix, una piattaforma online specializzata nella comparazione di smartphone e dispositivi elettronici. Su Kaggle il dataset viene presentato come una raccolta di dati relativa agli smartphone più recenti del 2026, con prezzi, specifiche e punteggi VFM.

Il dataset iniziale è composto da 997 record e 26 colonne. Le colonne presenti prima dell'applicazione del processo ETL sono le seguenti:

1. `brand_name`: nome del brand dello smartphone;
2. `model`: modello del dispositivo;
3. `price_category`: categoria di prezzo a cui appartiene lo smartphone;
4. `price`: prezzo del dispositivo;
5. `spec_score`: punteggio complessivo relativo alle specifiche tecniche;

6. `vfm_score`: punteggio del rapporto qualità-prezzo;
7. `vfm_label`: etichetta descrittiva del valore del dispositivo;
8. `has_5G`: presenza o assenza del supporto alla rete 5G;
9. `has_NFC`: presenza o assenza della tecnologia NFC;
10. `has_IR`: presenza o assenza del sensore a infrarossi;
11. `processor_brand`: marca del processore;
12. `processor_name`: nome del processore;
13. `num_core`: numero di core del processore;
14. `processor_speed`: velocità del processore;
15. `ram`: quantità di memoria RAM;
16. `memory`: capacità di memoria interna;
17. `battery_capacity (mAh)`: capacità della batteria espressa in mAh;
18. `fast_charging (W)`: potenza della ricarica rapida espressa in Watt;
19. `charging_ratio`: rapporto relativo alla velocità di ricarica;
20. `charging_speed_type`: tipologia di velocità di ricarica;
21. `screen_size`: dimensione dello schermo;
22. `refresh_rate`: frequenza di aggiornamento del display;
23. `rear_camera`: risoluzione della fotocamera posteriore;
24. `front_camera`: risoluzione della fotocamera anteriore;
25. `rear_camera_count`: numero di fotocamere posteriori;
26. `os`: sistema operativo del dispositivo.

3.1.2 Dataset 2: MySmartPrice Mobile Dataset

Il secondo dataset utilizzato, denominato *Mobile price and specs [2025]*, è anch'esso disponibile sulla piattaforma Kaggle. Esso contiene informazioni relative a smartphone e dispositivi mobili, ottenute mediante attività di web scraping del sito MySmartPrice.com, una piattaforma online specializzata nella pubblicazione di specifiche tecniche, prezzi, recensioni e confronti tra dispositivi mobili. Secondo la descrizione fornita su Kaggle, la raccolta comprende un ampio insieme di dati relativi ai telefoni cellulari provenienti da tale piattaforma.

Il dataset iniziale è composto da 5263 record e 15 colonne. Le colonne presenti prima dell'applicazione del processo ETL sono le seguenti:

1. `Unnamed: 0`: indice numerico della riga;
2. `mobile_name`: nome dello smartphone;
3. `release_date`: data di rilascio del dispositivo;

4. `price`: prezzo dello smartphone;
5. `avg_rating`: valutazione media del dispositivo;
6. `total_ratings`: numero totale di valutazioni ricevute;
7. `cpu`: informazioni relative al processore;
8. `rear_camera`: caratteristiche della fotocamera posteriore;
9. `front_camera`: caratteristiche della fotocamera anteriore;
10. `display`: informazioni relative al display;
11. `ram_and_storage`: informazioni su RAM e memoria interna;
12. `battery_and_charging_speed`: capacità della batteria e velocità di ricarica;
13. `operating_system`: sistema operativo del dispositivo;
14. `5G|NFC|Fingerprint`: presenza del supporto alle tecnologie 5G e NFC e del sensore di impronte digitali;
15. `expert_view`: valutazione o descrizione sintetica fornita dagli esperti.

L'utilizzo congiunto dei due dataset consente di disporre di una base dati ampia ed eterogenea, utile per analizzare il mercato degli smartphone sia dal punto di vista tecnico sia da quello commerciale. Tuttavia, poiché i dati provengono da fonti differenti e presentano strutture non perfettamente omogenee, è stato necessario applicare un processo di ETL finalizzato alla pulizia, alla trasformazione e all'integrazione delle informazioni.

3.2 Processo di ETL

Il termine ETL, ovvero Extract, Transform and Load (estrazione, trasformazione e caricamento), è il processo di raccolta dei dati da un gran numero di sorgenti eterogenee e della loro successiva organizzazione e centralizzazione in un unico repository. Tale processo consiste nell'utilizzo di una serie di regole aziendali per pulire e organizzare i dati grezzi provenienti da più fonti differenti e prepararli per l'archiviazione, l'analisi e il Machine Learning. L'analisi dei dati consente di rispondere a esigenze specifiche di Business Intelligence, come la previsione dell'esito delle decisioni aziendali o la generazione di report e dashboard.

Le organizzazioni moderne dispongono di grandi quantità di dati strutturati e non strutturati provenienti da diverse fonti, tra cui:

- dati dei clienti raccolti tramite sistemi di pagamento online e sistemi di *Customer Relationship Management* (CRM);
- dati relativi all'inventario e alle operazioni provenienti dai sistemi dei fornitori;
- dati generati dai sensori installati su dispositivi dell'Internet of Things (IoT);
- dati di marketing ottenuti dai social media e dai feedback dei clienti;
- dati dei dipendenti provenienti dai sistemi interni di gestione delle risorse umane.

Secondo uno studio, due terzi delle aziende sono in grado di ottenere benefici limitati dall'utilizzo dei dati, se non addirittura nessun beneficio. I dati tendono a rimanere bloccati; il processo ETL rende tali dati disponibili estraendoli per finalità di business.

Attraverso il processo di ETL è possibile estrarre i dati dalle diverse sorgenti, trasformarli in un formato coerente e adeguato all'analisi e, infine, caricarli in un sistema centralizzato. Questo processo consente di preparare i singoli dataset grezzi in una struttura più organizzata e fruibile, facilitando l'analisi dei dati e permettendo di ottenere informazioni e approfondimenti più significativi a supporto dei processi decisionali aziendali.

Vantaggi dell'ETL per la Business Intelligence

La tecnologia di *Extract, Transform and Load* (ETL) migliora i processi di Business Intelligence e di analisi dei dati, rendendo la gestione delle informazioni più affidabile, accurata ed efficiente. Tra i principali vantaggi e ambiti di applicazione dell'ETL si possono individuare i seguenti:

- *Contesto storico*: l'ETL consente di integrare dati storici con informazioni provenienti da sistemi e applicazioni più recenti, offrendo una visione completa dell'evoluzione aziendale nel tempo. Questa integrazione permette di confrontare dati legacy e dati correnti, supportando analisi longitudinali e processi decisionali più consapevoli.
- *Visione integrata dei dati*: attraverso l'unificazione di dati provenienti da fonti eterogenee, l'ETL fornisce una vista consolidata delle informazioni aziendali. Ciò riduce la complessità nella gestione di molteplici dataset, migliora la qualità dei dati e facilita attività di analisi, reporting e visualizzazione.
- *Accuratezza delle analisi*: i processi di trasformazione, pulizia e validazione dei dati contribuiscono a garantire l'affidabilità delle informazioni. L'integrazione con strumenti di controllo della qualità consente, inoltre, di soddisfare requisiti normativi e standard di conformità.
- *Automazione delle operazioni*: l'ETL automatizza le attività ripetitive di raccolta, trasformazione e caricamento dei dati, riducendo l'intervento manuale e aumentando l'efficienza operativa. Gli aggiornamenti possono essere eseguiti periodicamente o in tempo reale, consentendo una gestione più dinamica delle informazioni.
- *Migrazione dei dati*: l'ETL facilita il trasferimento di dati tra applicazioni e piattaforme differenti, garantendo la continuità operativa e la corretta conservazione delle informazioni.
- *Replica e backup*: i processi di ETL supportano la replica dei dati per finalità di backup e ridondanza, contribuendo alla sicurezza e alla disponibilità delle informazioni aziendali.
- *Integrazione dei sistemi operativi*: l'ETL consente il trasferimento e l'arricchimento dei dati tra sistemi aziendali, come CRM e Operational Data Store (ODS), migliorando la qualità e la coerenza delle informazioni.
- *Supporto alla Business Intelligence*: mediante il caricamento dei dati all'interno di un Data Warehouse, l'ETL favorisce l'organizzazione e la trasformazione dei dati in informazioni strategiche a supporto delle attività decisionali.
- *Migrazione verso il Cloud*: l'ETL supporta il trasferimento di dati e applicazioni da infrastrutture locali verso ambienti cloud, ibridi o multi-cloud, facilitando i processi di innovazione tecnologica.

- *Sincronizzazione dei sistemi*: l'ETL garantisce l'allineamento continuo dei dati tra sistemi aziendali differenti, assicurando coerenza, aggiornamento e disponibilità delle informazioni.

Il processo di ETL è generalmente suddiviso in tre fasi:

- *Estrazione (extract)*;
- *Trasformazione (transform)*;
- *Caricamento (load)*.

Esse saranno esaminate in dettaglio nelle prossime sottosezioni.

3.2.1 Extract

Durante la fase di estrazione i dati grezzi vengono prelevati da una vasta gamma di sorgenti (database relazionali, file CSV, API o sistemi cloud) e successivamente archiviati in un'area di staging. Quest'ultima rappresenta un'area di archiviazione intermedia utilizzata per conservare temporaneamente i dati estratti. Le aree di staging sono spesso transitorie, poiché i loro contenuti vengono eliminati una volta completata l'estrazione. Tuttavia, possono anche mantenere una copia dei dati a fini di monitoraggio e risoluzione dei problemi.

Le principali sorgenti dalle quali possono essere acquisiti i dati sono le seguenti:

- *Database relazionali e NoSQL*: contengono dati strutturati e semi-strutturati relativi alle applicazioni e ai servizi aziendali, come informazioni sugli utenti, transazioni e configurazioni di sistema.
- *Log di sistema*: registrano il comportamento dei sistemi operativi e delle infrastrutture IT, includendo eventi di avvio, arresto, errori e messaggi diagnostici.
- *Registri di attività (activity logs)*: tracciano le operazioni eseguite dagli utenti o dalle applicazioni, ad esempio accessi, modifiche a risorse e operazioni amministrative.
- *Report di anomalie e prestazioni*: raccolgono metriche e segnalazioni riguardanti il funzionamento delle applicazioni, come tempi di risposta elevati, utilizzo anomalo delle risorse o malfunzionamenti.
- *Eventi di sicurezza*: comprendono informazioni generate da firewall, software antivirus e altri strumenti di cybersecurity, utili per identificare minacce e comportamenti sospetti.

3.2.2 Transform

La fase di trasformazione del processo di ETL, considerata la più critica, consiste nell'applicazione di regole e operazioni ai dati estratti al fine di migliorarne la qualità, standardizzarne il formato, integrarne il contenuto e prepararli per l'analisi. In questa fase, i dati presenti nell'area di staging vengono trasformati e consolidati per essere successivamente trasferiti nel data warehouse di destinazione.

Il processo di trasformazione dei dati può essere suddiviso in due categorie principali, in funzione delle operazioni applicate ai dati stessi:

- trasformazioni dei dati di base;
- trasformazioni dei dati avanzata.

Trasformazioni dei dati di base

Le trasformazioni di base migliorano la qualità dei dati attraverso la rimozione degli errori, la correzione di dati incompleti o inconsistenti e la semplificazione delle informazioni. Di seguito sono riportati alcuni esempi di tali trasformazioni.

- *Pulizia dei dati*: consiste nella rimozione degli errori e nell'adattamento dei dati al formato di destinazione. Ad esempio, i valori mancanti possono essere sostituiti con il valore 0 o con altri valori predefiniti coerenti con il contesto applicativo.
- *Deduplicazione dei dati*: consiste nell'individuazione e nella rimozione dei record duplicati.
- *Revisione del formato dei dati*: consiste nella conversione dei dati, quali insiemi di caratteri, unità di misura e valori temporali, in un formato coerente e uniforme. Ad esempio, un'azienda alimentare potrebbe disporre di database diversi contenenti ricette con ingredienti espressi sia in chilogrammi sia in libbre. Durante la fase di trasformazione, è possibile utilizzare query scritte nel linguaggio M di Power Query per convertire automaticamente tutti i valori in una medesima unità di misura, ad esempio in chilogrammi.

Trasformazioni dei dati avanzata

Le trasformazioni avanzate sfruttano regole aziendali per ottimizzare i dati allo scopo di facilitarne l'analisi. Di seguito sono riportati alcuni esempi di tali trasformazioni.

- *Derivazione*: consiste nell'applicazione di regole aziendali ai dati per calcolare nuovi valori a partire da quelli esistenti.
- *Unione*: consiste nel collegamento di dati corrispondenti provenienti da sorgenti diverse.
- *Divisione*: consiste nella suddivisione di una colonna di dati in più colonne nel sistema di destinazione.
- *Sintesi*: consiste nella riduzione di un elevato numero di valori in un insieme di dati più compatto, al fine di facilitarne l'analisi. Ad esempio, i valori delle fatture degli ordini dei clienti possono presentare numerosi importi differenti. In tal caso, è possibile sintetizzare i dati aggregandoli nell'arco di un determinato periodo per creare un indicatore di Customer Lifetime Value (CLV).
- *Crittografia*: consiste nella protezione dei dati sensibili mediante tecniche di cifratura, al fine di rispettare la normativa vigente in materia di protezione dei dati e tutela della privacy prima del trasferimento verso il sistema di destinazione.

3.2.3 Load

La fase di caricamento consiste nel trasferimento dei dati trasformati dall'area di staging al data warehouse di destinazione, dove risultano pronti per le successive attività di analisi, visualizzazione e creazione di dashboard e report.

Per la maggior parte delle organizzazioni che vi ricorrono, si tratta di un processo automatizzato, ben definito, continuo e basato su batch. Di seguito sono descritti due metodi per il caricamento dei dati.

Caricamento completo

Nel caso del caricamento completo, l'intero contenuto del sistema sorgente viene trasformato e trasferito nel data warehouse di destinazione. Questa modalità viene generalmente utilizzata durante il primo popolamento del data warehouse o quando è necessario effettuare una sincronizzazione completa dei dati. Tuttavia, richiede tempi di esecuzione più lunghi e un maggiore impiego di risorse rispetto ad altre modalità di caricamento.

Caricamento incrementale

Nel caso del caricamento incrementale, lo strumento di ETL trasferisce a intervalli regolari esclusivamente le modifiche intervenute rispetto all'ultima esecuzione del processo. A tal fine, viene memorizzata la data dell'ultimo caricamento, così da individuare e trasferire soltanto i record aggiunti o modificati. Questa modalità risulta più efficiente rispetto al caricamento completo ed è utilizzata nella maggior parte degli scenari operativi, in particolare nei processi eseguiti periodicamente, nei quali le variazioni dei dati tra un'esecuzione e la successiva sono limitate. Esistono due principali modalità di implementazione del caricamento incrementale:

- *Caricamento incrementale in streaming*: il caricamento incrementale in streaming viene adottato quando i volumi di dati sono contenuti e le modifiche devono essere trasferite continuamente verso il data warehouse di destinazione. Quando la velocità di generazione dei dati raggiunge milioni di eventi al secondo, è possibile ricorrere all'elaborazione dei flussi di eventi (event stream processing) per monitorare ed elaborare i dati in tempo reale, consentendo decisioni più tempestive.
- *Caricamento incrementale in batch*: il caricamento incrementale in batch viene utilizzato in presenza di grandi volumi di dati. Le modifiche vengono raccolte e trasferite periodicamente secondo intervalli temporali prestabiliti. Durante ciascuna finestra di caricamento, i sistemi di origine e di destinazione rimangono sincronizzati mediante l'esecuzione delle operazioni di aggiornamento pianificate.

Durante il processo di caricamento dei dati, una delle attività fondamentali è rappresentata dal monitoraggio e dalla gestione degli errori. La presenza di dati di scarsa qualità in uno dei sistemi può compromettere l'affidabilità delle informazioni provenienti da altre sorgenti. Inoltre, le operazioni di estrazione e caricamento possono non andare a buon fine qualora uno o più sistemi risultino temporaneamente indisponibili.

Per tale motivo, risulta fondamentale implementare meccanismi di logging e auditing, strategie di rollback per il recupero da eventuali guasti e notifiche automatiche per segnalare agli amministratori eventuali anomalie durante l'esecuzione del processo.

In genere, il processo di caricamento viene eseguito durante le ore di minore attività dei sistemi, quando il traffico sui sistemi di origine e sul data warehouse è ridotto al minimo.

3.3 Differenza tra ETL e ELT

ETL ed ELT sono due metodi di integrazione dei dati; essi sono i metodi più comuni anche se, esistono altri approcci utilizzati per facilitare i workflow di integrazione dei dati.

Il processo di estrazione, caricamento e trasformazione (ELT) è un'estensione di quello di estrazione, trasformazione e caricamento (ETL) con un'inversione dell'ordine delle operazioni. La differenza più evidente è che il processo ELT esegue la fase di caricamento prima di quella di trasformazione, invertendo la seconda e la terza fase del processo ETL. Nel processo

ETL, i dati vengono prima trasformati, adattandoli alle esigenze dell'analisi, per poi essere caricati nel sistema di destinazione. Nel processo ELT, invece, i dati grezzi vengono caricati integralmente nel sistema di destinazione e trasformati solo successivamente.

Tuttavia, l'ordine dei passaggi non è l'unica differenza. Nell'ELT il sistema di archiviazione dei dati di destinazione può essere un data warehouse, ma più spesso si tratta di un data lake. Quest'ultimo è un grande archivio centrale progettato per contenere dati strutturati e non strutturati su larga scala. I data lake vengono gestiti tramite una piattaforma Big Data (come Apache Hadoop) o un sistema di gestione di dati NoSQL distribuito. Sebbene possano supportare attività di Business Intelligence, essi vengono impiegati principalmente per applicazioni di Intelligenza Artificiale, Machine Learning, analisi predittiva ed elaborazione di dati in tempo reale.

Oltre all'ordine delle operazioni, esistono ulteriori differenze tra ETL ed ELT. Poiché nell'ETL i dati vengono trasformati prima del trasferimento nel repository centrale, questo approccio può facilitare il rispetto delle normative in materia di protezione dei dati e tutela della privacy rispetto all'ELT. Ad esempio, se gli analisti non trasformano i dati sensibili prima del loro utilizzo, questi potrebbero rimanere inosservati nel data lake. Tuttavia, i data scientist potrebbero preferire l'ELT, poiché consente loro di effettuare sperimentazioni in una *sandbox* contenente dati non elaborati e di applicare trasformazioni personalizzate in funzione delle specifiche esigenze analitiche. In conclusione, la scelta tra ETL ed ELT dipende dalle esigenze operative dell'organizzazione, dalle risorse disponibili e dalle caratteristiche dei dati da elaborare.

3.4 Strumenti ETL

In passato, le organizzazioni scrivevano il proprio codice ETL. Oggi sono disponibili molti strumenti ETL open source e commerciali, così come servizi basati sul cloud, tra cui scegliere. Gli strumenti ETL sono software progettati per estrarre, trasformare e caricare dati da una o più sorgenti verso un sistema o database di destinazione. Sono progettati per automatizzare e semplificare il processo di estrazione dei dati da diverse fonti, trasformarli in un formato coerente e pulito e caricarli nel sistema di destinazione in modo tempestivo ed efficiente.

Le principali caratteristiche da ricercare in uno strumento ETL sono le seguenti:

- *Open-source*: le soluzioni open-source offrono generalmente una maggiore flessibilità e permettono di evitare il fenomeno del *vendor lock-in*, ossia la dipendenza da un singolo fornitore.
- *Portabilità*: con la crescente diffusione degli ambienti cloud ibridi, è fondamentale poter sviluppare le integrazioni una sola volta ed eseguirle successivamente su infrastrutture differenti in base alle esigenze operative.
- *Semplicità d'uso*: gli strumenti ETL dovrebbero essere intuitivi, facili da apprendere e dotati di interfacce grafiche che garantiscono la creazione e la gestione visuale delle pipeline di integrazione dei dati, spesso tramite drag-and-drop.
- *Funzionalità avanzate*: è importante che la piattaforma supporti funzionalità quali la parallelizzazione dei processi, la gestione degli errori e dei meccanismi di *failover*, il logging dettagliato e il monitoraggio in tempo reale delle attività.
- *Compatibilità con il cloud*: lo strumento dovrebbe operare in modo nativo in ambienti *single-cloud*, *multi-cloud* e ibridi, garantendo elevata flessibilità e interoperabilità.

3.5 Processo ETL applicato ai dataset

Nel contesto di questo lavoro sono stati analizzati due dataset distinti: *Latest Smartphones 2026: Price, Specs & VFM Score* e *Mobile price and specs [2025]*, descritti nella Sezione 3.1. Su entrambi i dataset è stato applicato un processo di ETL finalizzato alla preparazione, all'integrazione e alla successiva analisi dei dati.

3.5.1 Extract

La prima fase del processo ETL è quella di *Extract*, ovvero l'estrazione e l'acquisizione dei dati provenienti dalle fonti selezionate.

In questa fase sono stati importati in Power BI i due dataset, entrambi in formato CSV e reperibili sulla piattaforma Kaggle.

La fase di estrazione si è articolata nei seguenti passaggi:

1. *Step 1*: I dataset sono stati scaricati dalla piattaforma Kaggle in formato CSV.
2. *Step 2*: Entrambi i dataset sono stati poi importati in Power BI per consentire le operazioni di pulizia, trasformazione, integrazione e visualizzazione dei dati.

Per l'importazione dei dataset in Power BI è stata seguita la procedura descritta di seguito:

- all'avvio di Power BI viene visualizzato un report vuoto, come mostrato nella Figura 3.1;

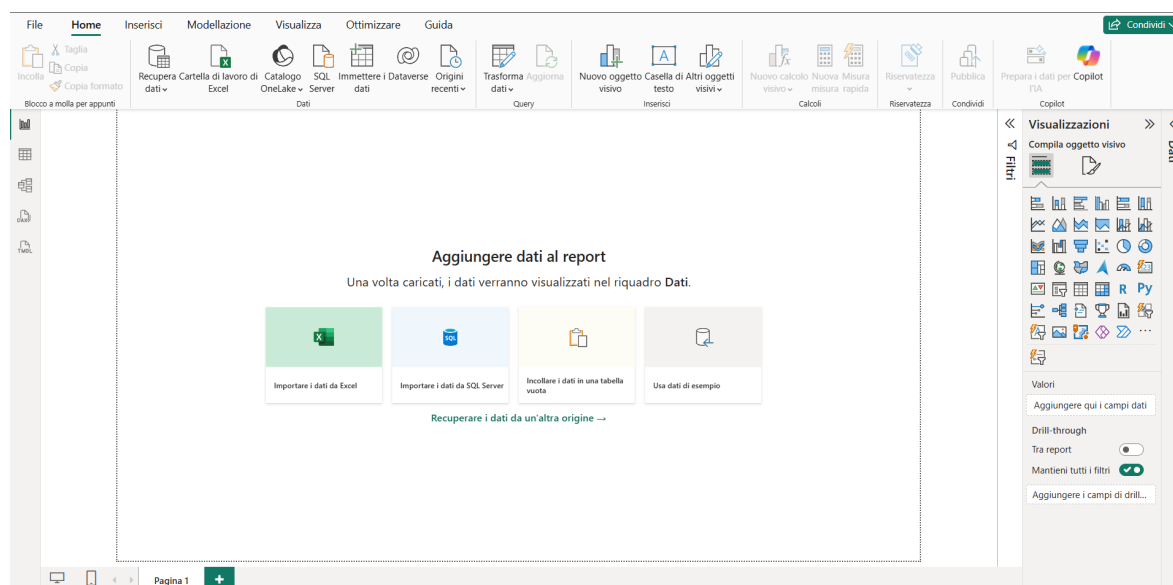


Figura 3.1: Visualizzazione iniziale della pagina di report vuota

- dal menu *Home* viene selezionata la voce *Recupera dati* e, successivamente, l'opzione *Testo/CSV*, come mostrato nella Figura 3.2;
- viene, quindi, richiesto di selezionare il file CSV presente sul dispositivo locale;
- dopo la selezione del file, Power BI visualizza un'anteprima del relativo contenuto;
- infine, è possibile scegliere tra due modalità operative:
 - *Carica*, che consente di importare direttamente i dati nel modello;

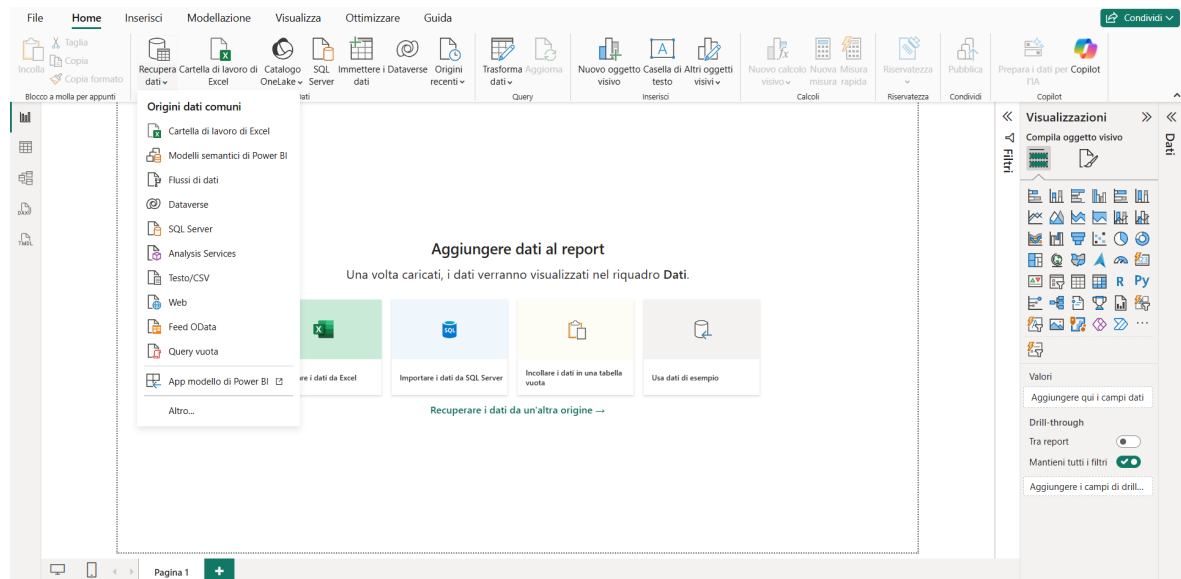


Figura 3.2: Importazione dei dati in Power BI

- *Trasforma dati*, che permette di accedere al *Power Query Editor* per eseguire operazioni di pulizia e trasformazione prima del caricamento dei dati.

L'anteprima viene mostrata nella Figura 3.3.

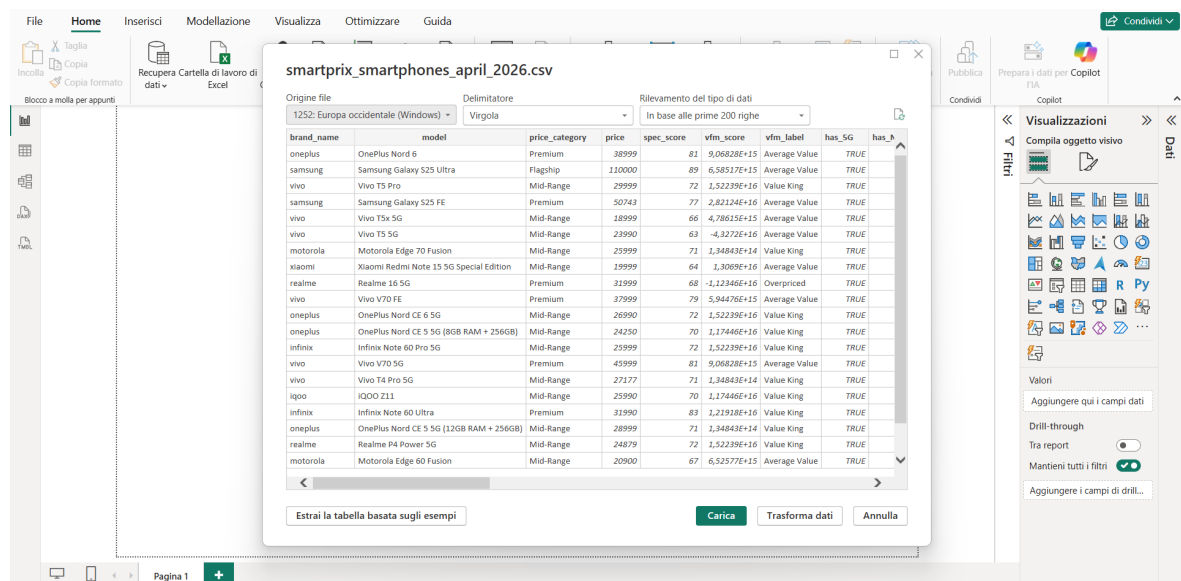


Figura 3.3: Visualizzazione dell'anteprima per opzioni di caricamento o trasformazione dei dati

3.5.2 Transform

In Power BI, la fase di trasformazione dei dati avviene attraverso lo strumento *Power Query Editor*, accessibile immediatamente dopo l'importazione dei file, selezionando l'opzione *Trasforma dati* anziché *Carica*. Questo ambiente consente di effettuare una serie di operazioni sui dati al fine di renderli coerenti, completi e pronti per l'analisi.

Dataset 1: Smartprix Smartphones Dataset

Durante la fase di trasformazione dei dati in Power Query sono stati applicati diversi passaggi di pulizia e standardizzazione, con l'obiettivo di rendere il dataset coerente e utilizzabile nelle successive analisi.

Come prima operazione di trasformazione, sono stati corretti i tipi di dato delle principali colonne numeriche e testuali, così da garantirne il corretto utilizzo nelle successive elaborazioni, analisi e visualizzazioni realizzate in Power BI.

Il codice utilizzato per effettuare tale operazione è riportato nel Listato 3.1.

```

1   = Table.TransformColumnTypes("#Intestazioni alzate di livello",
2   [{"brand_name", type text},
3   {"model", type text},
4   {"price_category", type text},
5   {"price", Int64.Type},
6   {"spec_score", Int64.Type},
7   {"vfm_score", Int64.Type},
8   {"vfm_label", type text},
9   {"has_5G", type logical},
10  {"has_NFC", type logical},
11  {"has_IR", type logical},
12  {"processor_brand", type text},
13  {"processor_name", type text},
14  {"num_core", Int64.Type},
15  {"processor_speed", Int64.Type},
16  {"ram", Int64.Type},
17  {"memory", Int64.Type},
18  {"battery_capacity(mAh)", Int64.Type},
19  {"fast_charging(W)", Int64.Type},
20  {"charging_ratio", Int64.Type},
21  {"charging_speed_type", type text},
22  {"screen_size", Int64.Type},
23  {"refresh_rate", Int64.Type},
24  {"rear_camera", Int64.Type},
25  {"front_camera", Int64.Type},
26  {"rear_camera_count", Int64.Type},
27  {"os", type text}])

```

Listato 3.1: Modifica del tipo di colonne nel Dataset 1

Successivamente sono state aggiunte tre nuove colonne:

- Price_EUR, ottenuta a partire dalla colonna iniziale del dataset in cui era indicato il prezzo degli smartphone in rupie indiane, come indicato nel Listato 3.2;
- Price_Category, ottenuta definendo dei criteri di selezione per identificare la categoria di prezzo dello smartphone, come indicato nel Listato 3.3;
- Price_USD, ottenuta a partire dalla colonna iniziale del dataset in cui era indicato il prezzo degli smartphone in rupie indiane, come riportato nel Listato 3.4.

```

1   = Table.AddColumn("#Modificato tipo",
2   "Price_EUR",
3   each [price] * 0.011)

```

Listato 3.2: Creazione della colonna Price_Euro

```

1   = Table.AddColumn("#Modificato tipo colonna Price_EUR",
2   "Price_Category_EUR",
3   each if [Price_EUR] < 300 then "Low Budget"

```

```

4     else if [Price_EUR] <= 700 then "Mid Budget"
5     else "Premium Budget")

```

Listato 3.3: Creazione della colonna Price_Category

```

1     = Table.AddColumn("#Rimosse colonne5", "Price_USD", each [price_INR] / 83)

```

Listato 3.4: Creazione della colonna Price_USD

Alcune colonne iniziali sono state rinominate per rendere il dataset più chiaro e immediatamente utilizzabile. In particolare, sono state modificate le colonne relative alle funzionalità 5G, IR e NFC, come descritto nel Listato 3.5. Inoltre, poiché è stata prevista la creazione di colonne dedicate al prezzo in euro e in dollari, la colonna originaria contenente il prezzo espresso in rupie indiane è stata rinominata per distinguerla dalle altre variabili relative al costo, come riportato nel Listato 3.7.

La colonna relativa al sistema operativo è stata, inoltre, suddivisa in due colonne separate: una contenente il nome del sistema operativo e l'altra contenente la relativa versione. Questa operazione consente di effettuare analisi più dettagliate sulle diverse versioni dei sistemi operativi presenti nel dataset. Le operazioni di suddivisione e rinominazione delle colonne sono riportate nei Listati 3.6 e 3.7.

```

1     = Table.RenameColumns("#Aggiunta colonna Price_Category_EUR",
2     {"has_5G", "5G"},
3     {"has_NFC", "NFC"},
4     {"has_IR", "IR"})

```

Listato 3.5: Ridenominazione delle colonne 5G, IR e NFC

```

1     = Table.SplitColumn("#Modificato tipo 5G,NFC,IR", "os",
2     Splitter.SplitTextByDelimiter(" ", QuoteStyle.Csv), {"os.1", "os.2"})

```

Listato 3.6: Split della colonna dedicata all'informazione sul sistema operativo

```

1     = Table.RenameColumns("#Modificato tipo os",
2     {"os.1", "os.Name"},
3     {"os.2", "os.Version"},
4     {"price_category", "price_category_INR"},
5     {"price", "price_INR"})

```

Listato 3.7: Ridenominazione delle colonne relative al sistema operativo e al prezzo iniziale

Sono state, inoltre, effettuate operazioni di normalizzazione sulle colonne relative alla RAM, alla memoria e alle altre specifiche hardware, convertendo i valori nelle corrette unità di misura e rendendoli omogenei per le successive analisi. Attraverso tali trasformazioni è stato possibile, quindi, ottenere valori realistici e uniformi, nella unità di misura desiderata, rendendo i dati correttamente interpretabili nelle successive analisi statistiche e visualizzazioni.

È stata anche aggiunta la colonna Screen_Size_pollici, effettuando, anche in questo caso, operazioni di normalizzazione sui valori iniziali, che presentavano differenti unità di misura e formati non uniformi. Tale operazione è mostrata nel Listato 3.8.

```

1     = Table.AddColumn("#Rimosse colonne2", "Screen_size_pollici",
2     each if [screen_size] > 100
3     then [screen_size] / 100
4     else [screen_size] / 10)

```

Listato 3.8: Aggiunta della colonna Screen_size_pollici

Successivamente, sono state effettuate operazioni di sostituzione dei valori mancanti o non coerenti.

Un'operazione analoga di normalizzazione è stata effettuata anche sulle colonne `rear_camera` e `front_camera`, che sono state inizialmente rinominate per rendere esplicita l'unità di misura utilizzata e successivamente normalizzata, come mostrato nei Listati 3.9 e 3.10.

```
1 = Table.RenameColumns("#Rimosse colonne3",
2   {"rear_camera", "rear_camera_MP"},
3   {"front_camera", "front_camera_MP"})
```

Listato 3.9: Ridenominazione della camera

```
1 = Table.TransformColumns("#Rinominate colonne camera",
2   {"rear_camera_MP", each _ / 10, type number})
3
4 = Table.TransformColumns("#Divisa colonna rear_camera",
5   {"front_camera_MP", each _ / 10, type number})
```

Listato 3.10: Normalizzazione della camera

Infine, per alcune colonne testuali, come `os.Name`, `os.Version` e `processor_brand`, i valori nulli o mancanti sono stati sostituiti con la dicitura *Unknown*. In questo modo, è stato possibile evitare la presenza di valori mancanti nelle visualizzazioni e nei filtri.

Per le colonne numeriche relative alle fotocamere, come `rear_camera_MP` e `front_camera_MP`, i valori mancanti sono stati sostituiti con valori di riferimento, così da mantenere complete le informazioni tecniche necessarie per le analisi comparative e predittive.

Sono state eseguite operazioni di normalizzazione anche sui nomi dei brand e dei modelli. Alcuni valori scritti in modo non uniforme, come, ad esempio, *samaung* o *redmagicmagic*, sono stati corretti, rispettivamente, in *samsung* e *redmagic*, per evitare duplicazioni logiche e garantire una corretta aggregazione dei dati per brand.

Infine, la colonna relativa alla tecnologia 5G è stata trasformata sostituendo i valori numerici con etichette testuali più interpretabili. In particolare, il valore *1* è stato convertito in *Yes*, mentre il valore *0* è stato convertito in *No*. Tale trasformazione ha reso il campo più leggibile nelle dashboard e nei filtri interattivi. La stessa trasformazione è stata applicata anche alle colonne `IR` e `NFC`.

Come ultima operazione nel dataset 1, è stata aggiunta la colonna `ModelKey`, utile per l'integrazione del Dataset 1 con il Dataset 2. Tale operazione viene mostrata nel Listato 3.11.

```
1 = Table.AddColumn("#Modificato tipo3", "ModelKey", each
2   let
3     nome = Text.Lower([model]),
4     senzaParentesi = try Text.BeforeDelimiter(nome, "(") otherwise nome,
5     pulito1 = Text.Replace(senzaParentesi, " 5g", ""),
6     pulito2 = Text.Replace(pulito1, " 128gb", ""),
7     pulito3 = Text.Replace(pulito2, " 256gb", ""),
8     pulito4 = Text.Replace(pulito3, " 512gb", ""),
9     pulito5 = Text.Replace(pulito4, " 1tb", ""),
10    pulito6 = Text.Replace(pulito5, " 8gb", ""),
11    pulito7 = Text.Replace(pulito6, " 12gb", "")
12  in
13    Text.Trim(pulito7)
14  )
```

Listato 3.11: Creazione della colonna `ModelKey`

Dataset 2: MySmartPrice Mobile Dataset

Rispetto al primo dataset, il secondo conteneva due attributi aggiuntivi:

- `release_date`, ovvero la data di rilascio dello smartphone;
- `avg_rating`, ovvero la valutazione media espressa dagli utenti relativamente ad uno specifico modello.

L'integrazione di tali informazioni nel primo dataset consente di effettuare analisi temporali, predittive e sulla soddisfazione degli utenti.

Tuttavia, il secondo dataset non presentava due colonne distinte per il brand e per il modello dello smartphone, come invece accadeva nel primo dataset, bensì una singola colonna denominata `mobile_name` contenente entrambe le informazioni.

Per tale motivo, si è deciso di trasformare il secondo dataset al fine di renderlo strutturalmente coerente con il primo dataset, creando separatamente le colonne:

- `brand_name`;
- `model`.

Le due colonne ottenute sono state utilizzate per creare una chiave composta (`ModelKey`), formata dalla concatenazione degli attributi `brand_name` e `model`, utilizzata nelle fasi successive dell'analisi per effettuare il merge tra i due dataset.

Per il secondo dataset, si è deciso di concentrare l'analisi sui brand maggiormente diffusi e rappresentativi del mercato degli smartphone. In particolare, sono stati selezionati i seguenti marchi:

- Samsung;
- Apple;
- Google;
- Realme;
- Vivo;
- Xiaomi.

In una prima fase è stata creata la colonna `brand_name`, come riportato nel Listato 3.12.

```
1 = Table.AddColumn("#Filtrate righe", "brand_name", each
2     if Text.StartsWith(Text.Lower([mobile_name]), "samsung") then "Samsung"
3     else if Text.StartsWith(Text.Lower([mobile_name]), "apple") then "Apple"
4     else if Text.StartsWith(Text.Lower([mobile_name]), "google") then "Google"
5     else if Text.StartsWith(Text.Lower([mobile_name]), "realme") then "Realme"
6     else if Text.StartsWith(Text.Lower([mobile_name]), "xiaomi") then "Xiaomi"
7     else if Text.StartsWith(Text.Lower([mobile_name]), "vivo") then "Vivo"
8     else "Other")
```

Listato 3.12: Colonna personalizzata `brand_name` del dataset 2

A partire dalla colonna `brand_name` è stata successivamente generata la colonna `model`, ottenuta rimuovendo il nome del brand dalla colonna `mobile_name`, come indicato nel Listato 3.13.

```
1 = Text.Trim(
2     Text.Replace([mobile_name], [brand_name], "")
3 )
```

Listato 3.13: Colonna personalizzata `model` del dataset 2

Dopo la definizione delle informazioni identificative del dispositivo, è stata effettuata la pulizia della colonna `release_date`.

La colonna presentava inizialmente valori nel formato:

Release: 22 Jan, 2025

Per rendere la colonna utilizzabile nelle successive analisi temporali e predittive, è stata effettuata un'operazione di sostituzione della stringa *Release:* con una stringa vuota.

Successivamente, dopo la rimozione degli eventuali spazi superflui, la colonna è stata convertita nel formato *Date* illustrato nel Listato 3.14.

```
1 = Table.TransformColumnTypes("#Sostituito Release con 'vuoto'",
2 {"release_date", type date})
```

Listato 3.14: Modifica del tipo per la colonna `release_date`

Poiché del secondo dataset risultavano rilevanti ai fini dell'analisi soltanto le colonne:

- `release_date`;
- `avg_rating`;

tutti gli altri attributi già presenti nel primo dataset sono stati eliminati, con l'obiettivo di semplificare il processo di integrazione e ridurre la ridondanza informativa.

A questo punto è stata costruita la colonna `ModelKey`, necessaria per il merge tra i due dataset. Tale operazione viene mostrata nel Listato 3.15.

```
1 Table.AddColumn("#Rimosse colonne", "ModelKey", each
2     let
3         nome = Text.Lower([mobile_name]),
4         senzaParentesi = try Text.BeforeDelimiter(nome, "(") otherwise nome,
5         pulito1 = Text.Replace(senzaParentesi, " 5g", ""),
6         pulito2 = Text.Replace(pulito1, " 128gb", ""),
7         pulito3 = Text.Replace(pulito2, " 256gb", ""),
8         pulito4 = Text.Replace(pulito3, " 512gb", ""),
9         pulito5 = Text.Replace(pulito4, " 1tb", ""),
10        pulito6 = Text.Replace(pulito5, " 8gb", ""),
11        pulito7 = Text.Replace(pulito6, " 12gb", "")
12    in
13        Text.Trim(pulito7)
14 )
```

Listato 3.15: Colonna personalizzata `ModelKey` del dataset 2

Infine, sono stati corretti i tipi di dato delle principali colonne utilizzate nelle successive analisi, come `ModelKey`, `brand_name` e `model`, come mostrato nel Listato 3.16.

```
1 = Table.TransformColumnTypes("#Riordinate colonne1", {
2 {"ModelKey", type text},
3 {"brand_name", type text},
4 {"model", type text}
5 })
```

Listato 3.16: Modifica del tipo delle colonne del dataset 2

3.5.3 Load

Una volta completata la fase di trasformazione tramite *Power Query Editor*, il caricamento dei dati è stato effettuato selezionando l'opzione *Chiudi e carica*. In questo modo, i dati vengono importati nel modello di Power BI, dove è possibile utilizzarli per la costruzione delle

dashboard, la definizione di misure personalizzate tramite linguaggio DAX e la realizzazione di report interattivi.

Nel contesto di questo progetto, i due dataset relativi agli smartphone sono stati caricati nel modello di Power BI dopo aver applicato le opportune trasformazioni e operazioni di pulizia.

In particolare, il primo dataset rappresenta la base principale dell'analisi, in quanto contiene le caratteristiche tecniche dei dispositivi, mentre il secondo dataset è stato utilizzato per arricchire le informazioni disponibili nel primo dataset attraverso l'aggiunta delle informazioni riguardanti la data di rilascio e la valutazione media degli utenti.

Per realizzare l'integrazione dei due dataset è stata eseguita un'operazione di merge; l'integrazione è stata realizzata utilizzando la colonna `ModelKey`, precedentemente creata in entrambe le tabelle tramite la normalizzazione del nome del modello. È stato adottato un *Left Outer Join*, mantenendo tutti i record del primo dataset e aggiungendo, quando disponibili, le informazioni provenienti dal secondo dataset.

Come abbiamo già detto, dal secondo dataset sono stati conservati esclusivamente gli attributi `release_date` e `avg_rating`, ritenuti utili per le analisi successive. Il risultato dell'operazione è stato la creazione di una nuova tabella, denominata `Smartphone_dataset_Merged`, che integra le caratteristiche tecniche degli smartphone con le informazioni relative alla data di rilascio e alla valutazione media espressa dagli utenti.

Descrizione del procedimento per effettuare il merge tra i due dataset

Dopo la creazione della colonna `ModelKey` in entrambi i dataset, è stata selezionata l'opzione *Unisci query come nuova* dal menu *Home* di *Power Query Editor*, al fine di preservare le tabelle originali e generare una nuova tabella derivata, come mostrato nella Figura 3.4.

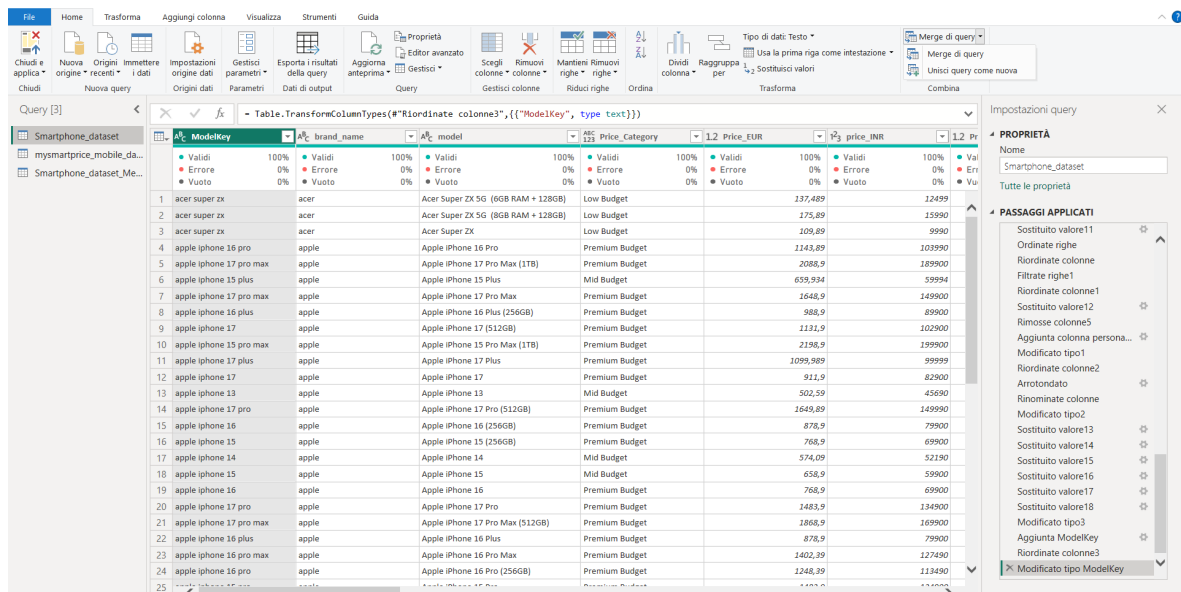


Figura 3.4: Scelta dell'opzione "Unisci query come nuova" per la creazione di una terza tabella creata tramite operazione di merge

Nella finestra di configurazione del merge sono stati selezionati i due dataset e la colonna `ModelKey` come chiave di corrispondenza. È stato, quindi, applicato un *Left Outer Join*.

La configurazione dell'operazione di merge è illustrata nella Figura 3.5.

Nel corso dell'operazione di merge, Power Query ha individuato circa 250 corrispondenze tra le 766 righe presenti nella prima tabella. Questo risultato evidenzia che, a causa delle

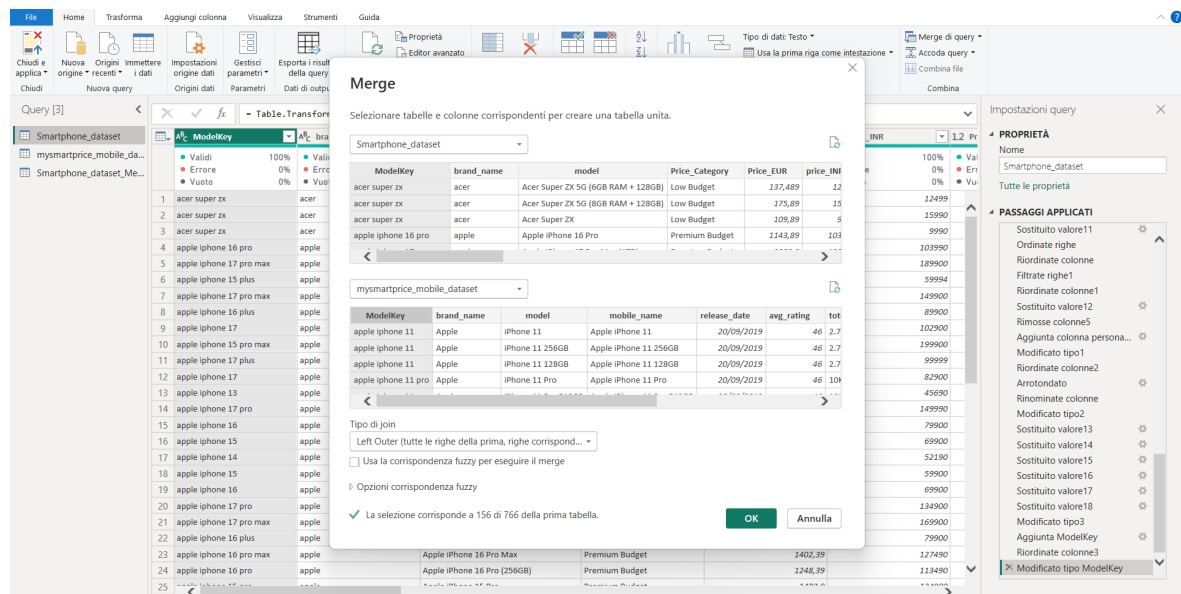


Figura 3.5: Processo di scelta dei criteri per l'operazione di Merge

diverse fonti di provenienza dei dataset, non tutti gli smartphone del primo dataset presentano una corrispondenza nel secondo.

Successivamente sono state espanse solamente le colonne `release_date` e `avg_rating` del secondo dataset, ottenendo il dataset finale utilizzato per la realizzazione delle dashboard descrittive e predittive.

La Figura 3.6 mostra le colonne del secondo dataset selezionate per l'integrazione.

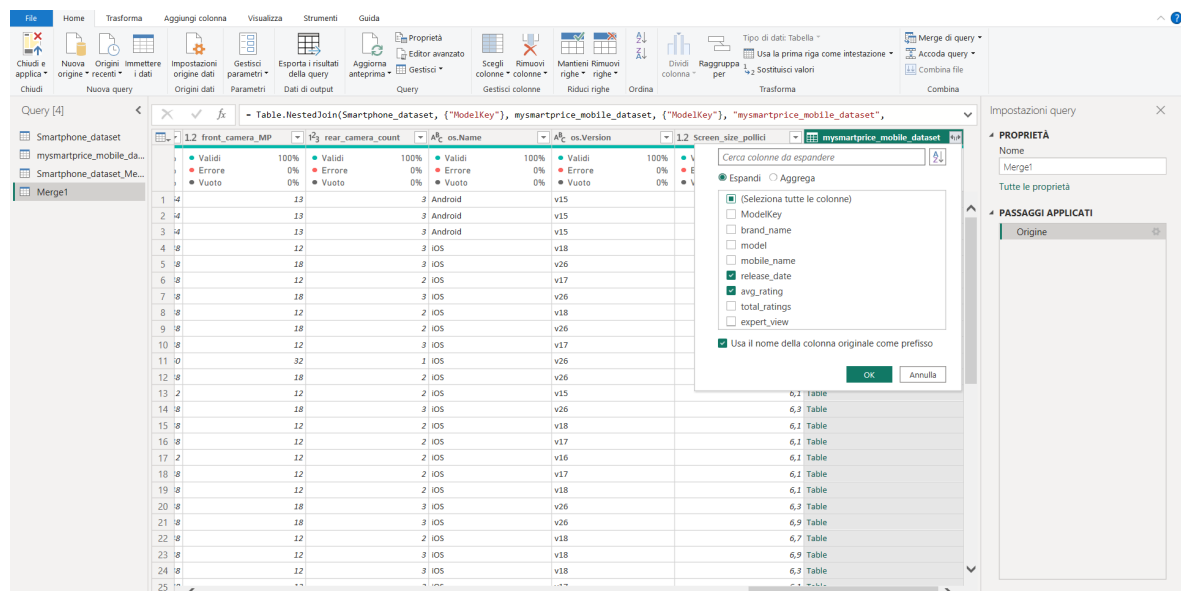


Figura 3.6: Scelta delle colonne del secondo dataset da mantenere nel nuovo dataset ottenuto tramite operazione di Merge

Progettazione e realizzazione delle dashboard di base

In questo capitolo verranno presentati il concetto di Data Visualization e il suo ruolo fondamentale nella realizzazione delle dashboard descrittive. Successivamente, saranno illustrati i principali grafici disponibili in Power BI e le potenzialità offerte dall'utilizzo delle misure DAX, che consentono di arricchire il dataset con informazioni personalizzate in grado di apportare ulteriore valore all'analisi. Infine, verranno presentate le dashboard descrittive realizzate.

4.1 Data Visualization

La Data Visualization consiste nella rappresentazione grafica dei dati mediante elementi visivi, quali grafici, mappe, tabelle e dashboard, con l'obiettivo di facilitarne l'esplorazione, la comprensione e la comunicazione. Attraverso una rappresentazione visiva efficace è possibile individuare più facilmente trend, anomalie, relazioni e pattern che risulterebbero difficili da cogliere osservando i dati in forma tabellare.

L'efficacia di una visualizzazione non dipende esclusivamente dalla qualità dei dati rappresentati, ma anche dal modo in cui gli individui percepiscono ed elaborano gli stimoli visivi. Per questo motivo risulta utile richiamare alcuni concetti della psicologia della percezione.

Cenni alla psicologia della percezione

La psicologia della percezione è la branca della psicologia che studia i processi attraverso i quali gli individui elaborano e interpretano le informazioni provenienti dagli stimoli sensoriali, in particolare quelli visivi e uditivi. Un aspetto cruciale di questa disciplina riguarda la diversa memorizzazione e la comprensione delle informazioni presentate visivamente rispetto a quelle presentate attraverso il linguaggio scritto o parlato.

Uno dei principali fattori che contribuiscono alla facilità con cui le immagini vengono memorizzate e comprese è il modo in cui il cervello elabora le informazioni visive. La teoria della codifica duale, sviluppata da Allan Paivio negli anni '70, suggerisce che il cervello elabora le informazioni in due modalità distinte: una modalità verbale, che si occupa delle parole e del linguaggio, e una modalità visiva, che si occupa delle immagini e delle rappresentazioni spaziali.

Secondo questa teoria, quando le persone vedono un'immagine, il cervello elabora l'informazione sia nella modalità visiva sia nella modalità verbale, creando una doppia

rappresentazione della stessa informazione. Questo processo di codifica duale aumenta la probabilità che le informazioni vengano memorizzate a lungo termine e facilmente richiamate, poiché le informazioni possono essere recuperate sia dalla memoria visiva sia da quella verbale.

Oltre alla teoria della codifica duale, un altro concetto rilevante è il cosiddetto effetto di superiorità delle immagini, fenomeno ampiamente documentato in psicologia, che dimostra che le persone tendono a ricordare meglio le immagini rispetto alle parole. Numerosi studi hanno mostrato che le immagini sono più facilmente riconoscibili e richiamabili rispetto alle parole, specialmente quando sono colorate, dettagliate e con un alto grado di concretezza.

Tra i principali contributi della psicologia della percezione assumono particolare rilevanza le Leggi della Gestalt, che descrivono i principi attraverso cui il cervello organizza automaticamente gli stimoli visivi e attribuisce loro significato.

Collegamento tra la psicologia della percezione e la Data Visualization

Chi si occupa di Data Visualization sa che il proprio lavoro si compone di diverse fasi che vanno dalla raccolta dei dati fino alla rappresentazione dei risultati. Grazie alla sua competenza è in grado di estrapolare trend, raggruppamenti e strutture all'interno di grandi quantità di dati; paradossalmente, però, è il passaggio successivo quello più delicato, ovvero riuscire a rendere questi stessi risultati accessibili anche ai non specialisti, presentandoli nel modo migliore possibile.

La psicologia della percezione ha elaborato una serie di principi che possono essere applicati alla progettazione di visualizzazioni efficaci. Tali principi sono comunemente noti come Leggi della Gestalt e costituiscono uno dei principali riferimenti teorici nella progettazione delle visualizzazioni informative.

Leggi della Gestalt

Per comprendere l'origine delle Leggi della Gestalt è necessario fare riferimento agli studi condotti nei primi decenni del Novecento da Max Wertheimer, Wolfgang Köhler e Kurt Koffka. Nel 1912 Wertheimer diede origine alla corrente psicologica della Gestalt (termine tedesco traducibile come "forma"), focalizzata sullo studio della percezione e dell'esperienza. Secondo l'approccio gestaltico, la comprensione del comportamento umano passa attraverso lo studio dei processi percettivi e dell'esperienza. Wertheimer sviluppò l'approccio gestaltico negli anni successivi insieme a Köhler e Koffka. Del loro lavoro, in questa sede ricordiamo in particolare le Leggi della Gestalt, enunciate nel 1923.

Le Leggi indicano i principi percettivi innati che la nostra mente utilizza per elaborare gli stimoli ricevuti dando loro ordine e forma. Conoscerli, quindi, ci aiuta a comprendere il modo in cui il nostro cervello "legge" ciò che lo circonda (nel nostro caso, le immagini).

Tra i principi maggiormente rilevanti per la progettazione delle visualizzazioni informative si possono individuare sei leggi percettive fondamentali:

- vicinanza;
- uguaglianza;
- connessione;
- continuità;
- simmetria;
- chiusura.

I principi della Gestalt trovano applicazione concreta negli strumenti moderni di Business Intelligence, tra cui Power BI. La disposizione degli elementi all'interno di una dashboard, la scelta dei colori, l'uso di grafici appropriati e l'organizzazione delle informazioni seguono principi percettivi che influenzano direttamente la capacità degli utenti di interpretare correttamente i dati.

L'efficacia della Data Visualization dipende dalla capacità di rendere comprensibili informazioni complesse anche a utenti privi di competenze specialistiche. Un insieme disordinato di dati grezzi rende particolarmente difficile l'identificazione di informazioni significative e di relazioni rilevanti. Per questo l'obiettivo della Data Visualization è quello di rendere intuibili (o meglio ancora evidenti) tutti quei pattern, quei raggruppamenti o quelle strutture che permettono di interpretare un fenomeno nel modo corretto e, di conseguenza, di prendere decisioni migliori.

Le rappresentazioni visive favoriscono una comprensione più rapida delle informazioni rispetto alla semplice presentazione testuale dei dati. Di conseguenza, la progettazione delle visualizzazioni riveste un'importanza centrale nel processo di comunicazione dei risultati dell'analisi.

La Data Visualization non consiste esclusivamente nella rappresentazione grafica dei dati, ma richiede un'attenta progettazione degli elementi visivi utilizzati. La scelta dei colori, della disposizione degli oggetti, delle dimensioni e delle forme influenza, infatti, il modo in cui gli utenti percepiscono e interpretano le informazioni rappresentate.

4.2 Data Visualization in Power BI

Come illustrato nel Capitolo 2, Microsoft Power BI è un software di Business Intelligence che consente agli utenti di creare report interattivi in grado di semplificare la lettura e la condivisione dei dati. Esso consente di trasformare grandi volumi di dati, provenienti da fonti differenti, in report e dashboard dinamiche, offrendo diversi metodi e strumenti per la rappresentazione visiva delle informazioni raccolte.

Power BI offre una vasta gamma di grafici e visualizzazioni che consentono la realizzazione di dashboard interattive. Attraverso menù dinamici, gli utenti possono effettuare selezioni specifiche e filtrare i dati in base a determinati criteri, come il brand o la fascia di prezzo degli smartphone. Le visualizzazioni vengono aggiornate automaticamente in funzione delle selezioni effettuate, favorendo un'esplorazione più approfondita dei dati. Grazie alle funzionalità di cross-filtering e drill-down, Power BI rappresenta uno strumento particolarmente efficace per le analisi descrittive e diagnostiche.

Inoltre, Power BI consente di definire misure personalizzate attraverso il linguaggio DAX (Data Analysis Expressions), sfruttando i dati presenti nei dataset a disposizione. Tali misure permettono di creare calcoli e metriche personalizzate utili ad arricchire le visualizzazioni con indicatori chiave di performance (KPI). Gli utenti possono, quindi, progettare dashboard capaci di fornire informazioni sintetiche e facilmente interpretabili, consentendo una rappresentazione chiara sia degli andamenti storici sia di eventuali previsioni future.

4.2.1 Grafici disponibili in Power BI

Power BI mette a disposizione numerose visualizzazioni per la realizzazione di dashboard. Di seguito vengono descritte le principali tipologie disponibili.

- *Grafico a colonne (Column Chart)*: utilizza barre verticali per confrontare valori tra diverse categorie. Sono disponibili varianti raggruppate e impilate.

- *Grafico a barre (Bar Chart)*: rappresenta i dati tramite barre orizzontali; è particolarmente utile quando le etichette delle categorie sono numerose o di lunghezza elevata.
- *Grafico a linee (Line Chart)*: consente di visualizzare l'andamento di uno o più indicatori nel tempo, evidenziando trend e variazioni.
- *Grafico ad area (Area Chart)*: simile al grafico a linee, ma con l'area sottostante colorata per enfatizzare la grandezza delle variazioni nel tempo.
- *Grafico combinato (Combo Chart)*: combina grafici a colonne e grafici a linee, permettendo il confronto simultaneo di misure differenti.
- *Grafico a nastri (Ribbon Chart)*: mostra l'evoluzione del ranking delle categorie nel tempo, evidenziando i cambiamenti di posizione.
- *Grafico a cascata (Waterfall Chart)*: rappresenta l'effetto cumulativo di incrementi e decrementi che conducono da un valore iniziale a un valore finale.
- *Grafico a torta (Pie Chart)*: visualizza la distribuzione percentuale delle categorie rispetto a un totale.
- *Grafico ad anello (Donut Chart)*: variante del grafico a torta caratterizzata da un foro centrale che consente di visualizzare informazioni aggiuntive all'interno del grafico.
- *Mappa ad albero (Treemap)*: rappresenta dati gerarchici mediante rettangoli di dimensioni proporzionali ai valori associati.
- *Grafico ad imbuto (Funnel Chart)*: rappresenta processi articolati in fasi sequenziali, come pipeline di vendita o conversione.
- *Grafico a dispersione (Scatter Chart)*: evidenzia la relazione tra due o più variabili numeriche e permette di individuare correlazioni e anomalie.
- *Grafico a bolle (Bubble Chart)*: estensione del grafico a dispersione che utilizza la dimensione delle bolle per rappresentare una terza variabile.
- *Tabella (Table)*: consente di visualizzare dati dettagliati organizzati in righe e colonne.
- *Matrice (Matrix)*: visualizzazione simile a una tabella pivot che supporta gerarchie e funzionalità di drill-down per analisi multidimensionali.
- *Mappe geografiche (Map Visualizations)*: permettono la rappresentazione spaziale dei dati mediante mappe standard, mappe di forma e integrazione con ArcGIS.
- *Card e Multi-row Card*: consentono di evidenziare uno o più indicatori sintetici all'interno di una dashboard.
- *Indicatore KPI*: visualizza le prestazioni di una metrica rispetto a un obiettivo prestabilito.
- *Grafico a misuratore (Gauge Chart)*: rappresenta il livello di raggiungimento di un obiettivo attraverso un indicatore radiale.
- *Visualizzazione Obiettivi (Goals)*: consente di monitorare metriche aziendali e il progresso verso target definiti.
- *Albero di scomposizione (Decomposition Tree)*: visualizzazione basata sull'Intelligenza Artificiale che permette di analizzare un indicatore attraverso differenti dimensioni.

- *Slicer*: elemento interattivo che consente di filtrare dinamicamente le visualizzazioni presenti nel report.
- *Visualizzazioni personalizzate (Custom Visuals)*: grafici aggiuntivi scaricabili dal marketplace di Power BI, tra cui Radar Chart, Box Plot, Heatmap e numerose altre visualizzazioni specialistiche.

4.2.2 Misure DAX in Power BI

In Power BI è possibile creare misure personalizzate con il linguaggio delle formule DAX (Data Analysis Expressions). La misura DAX è un calcolo creato utilizzando una formula che sfrutta le colonne del modello di dati a disposizione. Una volta create, le misure possono essere utilizzate nei report di Power BI come qualsiasi altra colonna del modello di dati, ad esempio per creare visualizzazioni o per filtrare i dati.

Le principali funzioni DAX che utilizzeremo successivamente per la creazione delle DAX personalizzate sono mostrate nella Tabella 4.1.

Funzione DAX	Significato
AVERAGE	Calcola la media aritmetica dei valori di una colonna.
MIN	Restituisce il valore minimo presente in una colonna o espressione.
MAX	Restituisce il valore massimo presente in una colonna o espressione.
CALCULATE	Modifica il contesto di filtro ed esegue un'espressione DAX nel nuovo contesto.
FIRSTNONBLANK	Restituisce il primo valore non vuoto di una colonna.
COUNT	Conta il numero di valori non vuoti in una colonna.
DIVIDE	Esegue una divisione gestendo automaticamente eventuali divisioni per zero.
DATATABLE	Crea una tabella definita manualmente direttamente in DAX.
SWITCH	Valuta più condizioni e restituisce il risultato corrispondente.
SELECTEDVALUE	Restituisce il valore selezionato in un contesto di filtro, se unico.
COUNTROWS	Conta il numero di righe di una tabella.
FILTER	Restituisce una tabella filtrata secondo una condizione specifica.
ISBLANK	Verifica se un valore è vuoto (blank).
NOT	Inverte il valore logico di un'espressione.
YEAR	Estrae l'anno da una data.
DISTINCTCOUNT	Conta il numero di valori distinti in una colonna.
SUMMARIZE	Crea una tabella riepilogativa raggruppando i dati.
ALLSELECTED	Rimuove alcuni filtri mantenendo il contesto delle selezioni effettuate.
AVERAGEX	Calcola la media di un'espressione valutata riga per riga.
SUMX	Calcola la somma di un'espressione valutata riga per riga.
POWER	Eleva un numero a una potenza specificata.
ROUND	Arrotonda un numero a un determinato numero di cifre.
VAR	Definisce una variabile temporanea all'interno di una misura DAX.
RETURN	Restituisce il risultato finale di una misura o espressione DAX.

Tabella 4.1: Principali funzioni DAX in Power BI

4.3 Misure DAX implementate

Nel corso della presente analisi sono state sviluppate e implementate diverse misure DAX (Data Analysis Expressions), con l'obiettivo di esaminare e rappresentare in modo approfondito i dati relativi al mercato degli smartphone.

Tali misure, definite come formule personalizzate, permettono di effettuare calcoli dinamici e di realizzare aggregazioni complesse direttamente all'interno dell'ambiente Power BI.

Le espressioni DAX sono state progettate per supportare la costruzione di dashboard interattive, ciascuna focalizzata su specifici aspetti del mercato.

In particolare, l'analisi ha riguardato elementi quali l'andamento dei prezzi, il confronto tra i diversi brand, la distribuzione dei modelli e le principali caratteristiche dei dispositivi disponibili sul mercato.

Nelle prossime sottosezioni vengono descritte le principali misure sviluppate per ciascun dataset, evidenziandone le finalità analitiche e il contributo informativo nell'ambito della visualizzazione e interpretazione dei dati.

Misure DAX del dataset **Smartprix Smartphones Dataset**

Le misure DAX elencate successivamente sono state fondamentali per la realizzazione delle prime quattro dashboard.

- **Average Smartphone Price (INR):** calcola la media del prezzo degli smartphone sul mercato in Rupie indiane

```
1 Average Smartphone Price (INR) =
2 AVERAGE('Smartphone_dataset'[price_INR])
```

Listato 4.1: Misura DAX per Average Smartphone Price (INR)

- **Average Smartphone Price (EUR):** calcola la media del prezzo degli smartphone sul mercato in Euro

```
1 Average Smartphone Price (EUR) =
2 AVERAGE('Smartphone_dataset'[Price_EUR])
```

Listato 4.2: Misura DAX per Average Smartphone Price (EUR)

- **Average Smartphone Price (USD):** calcola la media del prezzo degli smartphone sul mercato in Dollari

```
1 Average Smartphone Price (USD) =
2 AVERAGE('Smartphone_dataset'[Price_USD])
```

Listato 4.3: Misura DAX per Average Smartphone Price (USD)

- **Cheapest Brand:** individua il brand associato allo smartphone con il prezzo più basso presente nel dataset

```
1 Cheapest Brand =
2 VAR MinPrice =
3     MIN('smartphone_dataset'[Price_EUR])
4 RETURN
5     CALCULATE(
6         FIRSTNONBLANK('smartphone_dataset'[brand_name], 1),
7         'smartphone_dataset'[Price_EUR] = MinPrice
8     )
```

Listato 4.4: Misura DAX per Cheapest Brand

- **Cheapest Model:** individua il modello di smartphone con il prezzo più basso presente nel dataset

```
1 Cheapest Model =
2 VAR MinPrice =
3     MIN('smartphone_dataset'[Price_EUR])
4 RETURN
5     CALCULATE(
6         FIRSTNONBLANK('smartphone_dataset'[model], 1),
7         'smartphone_dataset'[Price_EUR] = MinPrice
```

```
8      )
```

Listato 4.5: Misura DAX per Cheapest Model

- Cheapest Price (EUR): calcola il prezzo minimo degli smartphone sul mercato in Euro

```
1 Cheapest Price (EUR) =
2 MIN('smartphone_dataset'[Price_EUR])
```

Listato 4.6: Misura DAX per Cheapest Price (EUR)

- Cheapest Price (INR): calcola il prezzo minimo degli smartphone sul mercato in Rupie indiane

```
1 Cheapest Price (INR) =
2 MIN('smartphone_dataset'[price_INR])
```

Listato 4.7: Misura DAX per Cheapest Price (INR)

- Cheapest Price (USD): calcola il prezzo minimo degli smartphone sul mercato in Dollari

```
1 Cheapest Price (USD) =
2 MIN('smartphone_dataset'[Price_USD])
```

Listato 4.8: Misura DAX per Cheapest Price (USD)

- Most Expensive Brand: individua il brand associato allo smartphone con il prezzo più elevato presente nel dataset

```
1 Most Expensive Brand =
2 VAR MaxPrice =
3     MAX('smartphone_dataset'[Price_EUR])
4 RETURN
5     CALCULATE(
6         FIRSTNONBLANK('smartphone_dataset'[brand_name], 1),
7         'smartphone_dataset'[Price_EUR] = MaxPrice
8     )
```

Listato 4.9: Misura DAX per Most Expensive Brand

- Most Expensive Model: individua il modello di smartphone con il prezzo più elevato presente nel dataset

```
1 Most Expensive Model =
2 VAR MaxPrice =
3     MAX('smartphone_dataset'[Price_EUR])
4 RETURN
5     CALCULATE(
6         FIRSTNONBLANK('smartphone_dataset'[model], 1),
7         'smartphone_dataset'[Price_EUR] = MaxPrice
8     )
```

Listato 4.10: Misura DAX per Most Expensive Model

- Most Expensive Price (EUR): calcola il prezzo massimo degli smartphone sul mercato in Euro

```
1 Most Expensive Price (EUR) =
2 MAX('smartphone_dataset'[Price_EUR])
```

Listato 4.11: Misura DAX per Most Expensive Price (EUR)

- **Most Expensive Price (INR):** calcola il prezzo massimo degli smartphone sul mercato in Rupie indiane

```
1 Most Expensive Price (INR) =  
2 MAX('smartphone_dataset'[price_INR])
```

Listato 4.12: Misura DAX per Most Expensive Price (INR)

- **Most Expensive Price (USD):** calcola il prezzo massimo degli smartphone sul mercato in Dollari

```
1 Most Expensive Price (USD) =  
2 MAX('smartphone_dataset'[Price_USD])
```

Listato 4.13: Misura DAX per Most Expensive Price (USD)

- **Average Battery:** calcola la capacità media della batteria degli smartphone in mAh

```
1 Average Battery =  
2 AVERAGE('Smartphone_dataset'[battery_capacity_mAh])
```

Listato 4.14: Misura DAX per Average Battery

- **Average RAM:** calcola la memoria RAM media degli smartphone in GB

```
1 Average RAM =  
2 AVERAGE('Smartphone_dataset'[ram_GB])
```

Listato 4.15: Misura DAX per Average RAM

- **Average Storage:** calcola la capacità media di archiviazione degli smartphone in GB

```
1 Average Storage =  
2 AVERAGE('Smartphone_dataset'[memory_GB])
```

Listato 4.16: Misura DAX per Average Storage

- **Average Rear Camera:** calcola la risoluzione media della fotocamera posteriore degli smartphone in megapixel

```
1 Average Rear Camera =  
2 AVERAGE('Smartphone_dataset'[rear_camera_MP])
```

Listato 4.17: Misura DAX per Average Rear Camera

- **Average Front Camera:** calcola la risoluzione media della fotocamera frontale degli smartphone in megapixel

```
1 Average Front Camera =  
2 AVERAGE('Smartphone_dataset'[front_camera_MP])
```

Listato 4.18: Misura DAX per Average Front Camera

- **Average Screen Size:** calcola la dimensione media dello schermo degli smartphone in pollici

```
1 Average Screen Size =  
2 AVERAGE('Smartphone_dataset'[Screen_size_pollici])
```

Listato 4.19: Misura DAX per Average Screen Size

- **Android Models:** calcola il numero di modelli di smartphone che utilizzano come sistema operativo Android

```

1 Android Models =
2     CALCULATE (
3         COUNT (' Smartphone_dataset' [model]),
4         ' Smartphone_dataset' [os.Name] = "Android"
5     )

```

Listato 4.20: Misura DAX per Android models

- iOS Models: calcola il numero di modelli di smartphone che utilizzano come sistema operativo iOS

```

1 iOS Models =
2     CALCULATE (
3         COUNT (' Smartphone_dataset' [model]),
4         ' Smartphone_dataset' [os.Name] = "iOS"
5     )

```

Listato 4.21: Misura DAX per iOS models

- Norm RAM: calcola il valore medio della RAM normalizzato rispetto al valore massimo presente nel dataset.

```

1 Norm RAM =
2     DIVIDE (
3         [Average RAM],
4         MAX (' Smartphone_dataset' [ram_GB])
5     )

```

Listato 4.22: Misura DAX per Norm RAM

- Norm Storage: calcola il valore medio della capacità di archiviazione normalizzato rispetto al valore massimo presente nel dataset.

```

1 Norm Storage =
2     DIVIDE (
3         [Average Storage],
4         MAX (' Smartphone_dataset' [memory_GB])
5     )

```

Listato 4.23: Misura DAX per Norm Storage

- Norm Battery: calcola la capacità media della batteria normalizzata rispetto al valore massimo presente nel dataset.

```

1 Norm Battery =
2     DIVIDE (
3         [Average Battery],
4         MAX (' Smartphone_dataset' [battery_capacity_mAh])
5     )

```

Listato 4.24: Misura DAX per Norm Battery

- Norm Camera: calcola la risoluzione media della fotocamera posteriore normalizzata rispetto al valore massimo presente nel dataset.

```

1 Norm Camera =
2     DIVIDE (
3         [Average Rear Camera],
4         MAX (' Smartphone_dataset' [rear_camera_MP])
5     )

```

Listato 4.25: Misura DAX per Norm Camera

Nel processo di costruzione del grafico radar, è stata creata una tabella di supporto denominata `Metrics`, contenente l'elenco delle principali caratteristiche hardware considerate nell'analisi (RAM, Storage, Batteria e Fotocamera).

Questa tabella ha lo scopo di definire le categorie dell'asse del grafico radar, consentendo di rappresentare in modo strutturato e confrontabile le diverse metriche.

La tabella `Metrics` è stata implementata tramite la seguente espressione DAX:

```
1 Metrics =
2 DATATABLE(
3     "Metric", STRING,
4     {
5         {"RAM"},
6         {"Storage"},
7         {"Battery"},
8         {"Camera"}
9     }
10 )
```

Listato 4.26: Tabella DAX `Metrics`

- `Selected Metric Value`: misura dinamica che seleziona il valore corretto da visualizzare nel grafico radar in base alla metrica corrente (RAM, Storage, Batteria o Fotocamera).

```
1 Selected Metric Value =
2 SWITCH(
3     SELECTEDVALUE(Metrics[Metric]),
4     "RAM", [Norm RAM],
5     "Storage", [Norm Storage],
6     "Battery", [Norm Battery],
7     "Camera", [Norm Camera]
8 )
```

Listato 4.27: Misura DAX `Selected Metric Value`

- `Selected Metric Android`: calcola il valore normalizzato della metrica selezionata considerando solo gli smartphone con sistema operativo Android.

```
1 Selected Metric Android =
2 CALCULATE(
3     [Selected Metric Value],
4     'Smartphone_dataset'[os.Name] = "Android"
5 )
```

Listato 4.28: Misura DAX per `Selected Metric Android`

- `Selected Metric iOS`: calcola il valore normalizzato della metrica selezionata considerando solo gli smartphone con sistema operativo iOS.

```
1 Selected Metric iOS =
2 CALCULATE(
3     [Selected Metric Value],
4     'Smartphone_dataset'[os.Name] = "iOS"
5 )
```

Listato 4.29: Misura DAX per `Selected Metric iOS`

Misure DAX del dataset **Smartphone dataset Merged**

Le seguenti misure DAX sono state implementate per la realizzazione della Dashboard 5, dedicata all'analisi della valutazione media degli utenti e del rapporto tra soddisfazione, prezzo e caratteristiche tecniche degli smartphone.

Esse si basano sul dataset ottenuto dal merge tra il primo e il secondo dataset, secondo il procedimento descritto nel Capitolo 3.

- **Average Rating:** calcola la valutazione media degli smartphone considerando i modelli per cui è disponibile il valore di rating.

```
1 Average Rating =
2     AVERAGE(Smartphone_dataset_Merged[avg_rating])
```

Listato 4.30: Misura DAX per Average Rating

- **Rated Models:** calcola il numero di modelli di smartphone per cui è disponibile una valutazione media degli utenti.

```
1 Rated Models =
2     COUNTROWS (
3         FILTER (
4             Smartphone_dataset_Merged,
5             NOT (ISBLANK(Smartphone_dataset_Merged[avg_rating]))
6         )
7     )
```

Listato 4.31: Misura DAX per Rated Models

- **High Rated Percentage:** calcola la percentuale di smartphone con valutazione media pari o superiore a 4 stelle rispetto al totale dei modelli valutati.

```
1 High Rated Percentage =
2     DIVIDE (
3         COUNTROWS (
4             FILTER (
5                 Smartphone_dataset_Merged,
6                 Smartphone_dataset_Merged[avg_rating] >= 4
7             )
8         ),
9         COUNTROWS (
10            FILTER (
11                Smartphone_dataset_Merged,
12                NOT (ISBLANK(Smartphone_dataset_Merged[avg_rating]))
13            )
14        )
15    )
```

Listato 4.32: Misura DAX per High Rated Percentage

- **Release Year:** crea una colonna calcolata che estrae l'anno dalla data di rilascio dello smartphone. Questa colonna viene utilizzata nelle successive misure predittive per analizzare l'andamento temporale dei modelli e dei prezzi.

```
1 Release Year =
2     YEAR(Smartphone_dataset_Merged[release_date])
```

Listato 4.33: Colonna DAX Release Year

- **Total Models 2024:** calcola il numero totale di modelli distinti rilasciati nel 2024, considerando la colonna ModelKey per evitare duplicazioni dovute a varianti dello stesso modello.

```

1      Total Models 2024 =
2          CALCULATE (
3              DISTINCTCOUNT (Smartphone_dataset_Merged[ModelKey]),
4              Smartphone_dataset_Merged[Release Year] = 2024
5          )

```

Listato 4.34: Misura DAX per Total Models 2024

- **Predicted Models 2027:** stima il numero di modelli previsti nel 2027 tramite una regressione lineare implementata in DAX. La misura utilizza l'anno di rilascio come variabile indipendente e il numero di modelli distinti come variabile dipendente.

```

1      Predicted Models 2027 =
2          VAR TabellaRegressione =
3              SUMMARIZE (
4                  FILTER (
5                      ALLSELECTED (Smartphone_dataset_Merged),
6                      NOT (ISBLANK (Smartphone_dataset_Merged[Release Year]))
7                  ),
8                  Smartphone_dataset_Merged[Release Year],
9                  "Y", DISTINCTCOUNT (Smartphone_dataset_Merged[ModelKey])
10             )
11          VAR AvgX =
12              AVERAGEX (TabellaRegressione, Smartphone_dataset_Merged[Release Year])
13          VAR AvgY =
14              AVERAGEX (TabellaRegressione, [Y])
15          VAR Slope =
16              DIVIDE (
17                  SUMX (
18                      TabellaRegressione,
19                      (Smartphone_dataset_Merged[Release Year] - AvgX) * ([Y] - AvgY)
20                  ),
21                  SUMX (
22                      TabellaRegressione,
23                      POWER (Smartphone_dataset_Merged[Release Year] - AvgX, 2)
24                  )
25              )
26          VAR Intercept =
27              AvgY - Slope * AvgX
28          RETURN
29              ROUND (Intercept + Slope * 2027, 0)

```

Listato 4.35: Misura DAX per Predicted Models 2027

- **Average Price 2024:** calcola il prezzo medio degli smartphone rilasciati nel 2024, considerando il prezzo espresso in Euro.

```

1      Average Price 2024 =
2          CALCULATE (
3              AVERAGE (Smartphone_dataset_Merged[Price_EUR]),
4              Smartphone_dataset_Merged[Release Year] = 2024
5          )

```

Listato 4.36: Misura DAX per Average Price 2024

- **Predicted Average Price 2027:** stima il prezzo medio previsto degli smartphone nel 2027 tramite una regressione lineare implementata in DAX. La

misura utilizza l'anno di rilascio come variabile indipendente e il prezzo medio annuale come variabile dipendente.

```

1 Predicted Average Price 2027 =
2 VAR TabellaRegressione =
3     SUMMARIZE (
4         FILTER (
5             ALLSELECTED (Smartphone_dataset_Merged),
6             NOT (ISBLANK (Smartphone_dataset_Merged[Release Year])) &&
7             NOT (ISBLANK (Smartphone_dataset_Merged[Price_EUR]))
8         ),
9         Smartphone_dataset_Merged[Release Year],
10        "Y", AVERAGE (Smartphone_dataset_Merged[Price_EUR])
11    )
12 VAR AvgX =
13     AVERAGEX (TabellaRegressione, Smartphone_dataset_Merged[Release Year])
14 VAR AvgY =
15     AVERAGEX (TabellaRegressione, [Y])
16 VAR Slope =
17     DIVIDE (
18         SUMX (
19             TabellaRegressione,
20             (Smartphone_dataset_Merged[Release Year] - AvgX) * ([Y] - AvgY)
21         ),
22         SUMX (
23             TabellaRegressione,
24             POWER (Smartphone_dataset_Merged[Release Year] - AvgX, 2)
25         )
26     )
27 VAR Intercept =
28     AvgY - Slope * AvgX
29 RETURN
30     Intercept + Slope * 2027

```

Listato 4.37: Misura DAX per Predicted Average Price 2027

- CAGR Models: calcola il tasso annuo composto di crescita del numero di modelli tra il 2024 e il valore previsto per il 2027.

```

1 CAGR Models =
2     POWER (
3         DIVIDE ([Predicted Models 2027], [Total Models 2024]),
4         1 / 3
5     ) - 1

```

Listato 4.38: Misura DAX per CAGR Models

- CAGR Prices: calcola il tasso annuo composto di crescita del prezzo medio tra il 2024 e il prezzo medio previsto per il 2027.

```

1 CAGR Prices =
2     POWER (
3         DIVIDE ([Predicted Average Price 2027], [Average Price 2024]),
4         1 / 3
5     ) - 1

```

Listato 4.39: Misura DAX per CAGR Prices

4.4 Descrizione e analisi delle dashboard di base realizzate

In questa sezione vengono presentate le dashboard realizzate a partire dai dataset analizzati e dalle misure DAX implementate:

- *Dashboard 1*: fornisce una visione d'insieme del mercato degli smartphone, sintetizzando i principali indicatori legati ai prezzi e alla distribuzione dei modelli. In particolare, consente di analizzare i prezzi medi in diverse valute, la distribuzione dei dispositivi per brand e fascia di prezzo e di individuare gli estremi dell'offerta, rappresentati dai modelli più economici e più costosi. L'obiettivo è offrire una lettura immediata delle dinamiche generali del mercato.
- *Dashboard 2*: è dedicata all'analisi delle principali caratteristiche tecniche degli smartphone, con l'obiettivo di evidenziare la distribuzione delle specifiche hardware e le relazioni tra esse. In particolare, vengono esaminate variabili quali configurazione delle fotocamere, memoria, dimensione dello schermo, numero di core e diffusione di alcune tecnologie. La dashboard consente di individuare pattern e tendenze nelle configurazioni dei dispositivi presenti nel dataset.
- *Dashboard 3*: analizza le caratteristiche tecniche degli smartphone in relazione alla fascia di prezzo di appartenenza. L'obiettivo è mettere in evidenza come le principali specifiche hardware variano tra dispositivi economici, di fascia media e premium, consentendo un confronto tra le diverse categorie di mercato. La dashboard, permette quindi, di comprendere l'impatto del prezzo sulle configurazioni dei dispositivi.
- *Dashboard 4*: è focalizzata sull'analisi dei sistemi operativi negli smartphone, con l'obiettivo di esaminare la distribuzione dei dispositivi tra Android e iOS e le principali differenze tra le due piattaforme. In particolare, consente di analizzare la diffusione dei sistemi operativi, la loro distribuzione nelle diverse fasce di prezzo e il confronto delle caratteristiche tecniche medie dei dispositivi.
- *Dashboard 5*: è dedicata all'analisi della relazione tra prezzo, valutazione degli utenti e caratteristiche tecniche degli smartphone, con l'obiettivo di comprendere quali fattori influenzano maggiormente la customer satisfaction. In particolare, consente di analizzare la correlazione tra prezzo e valutazione media, di individuare i dispositivi con il rapporto qualità/prezzo migliore e peggiore, e di valutare l'impatto delle principali caratteristiche hardware sul rating degli utenti. L'analisi permette, quindi, di evidenziare i modelli che offrono il maggiore valore percepito rispetto al prezzo di vendita.

4.4.1 Dashboard 1 - Panoramica generale del mercato degli smartphone

Questa dashboard ha l'obiettivo di fornire una visione complessiva e sintetica delle principali caratteristiche del mercato degli smartphone. Lo scopo è permettere un'analisi immediata degli indicatori chiave, come i prezzi medi espressi in diverse valute (USD, EUR e INR), la distribuzione dei modelli per brand e fascia di prezzo, nonché l'individuazione dei dispositivi più economici e più costosi.

La dashboard consente, inoltre, di esplorare i dati in modo dinamico attraverso filtri dedicati ai diversi produttori, facilitando il confronto tra le strategie di posizionamento dei brand e la varietà dell'offerta disponibile sul mercato.

Nella Figura 4.1 è mostrata la Dashboard 1 – Panoramica generale del mercato degli smartphone.

Key Performance Indicator (KPI) utilizzati

Considerando l'intero dataset, senza applicare filtri relativi a specifici brand, si ottengono i seguenti valori:

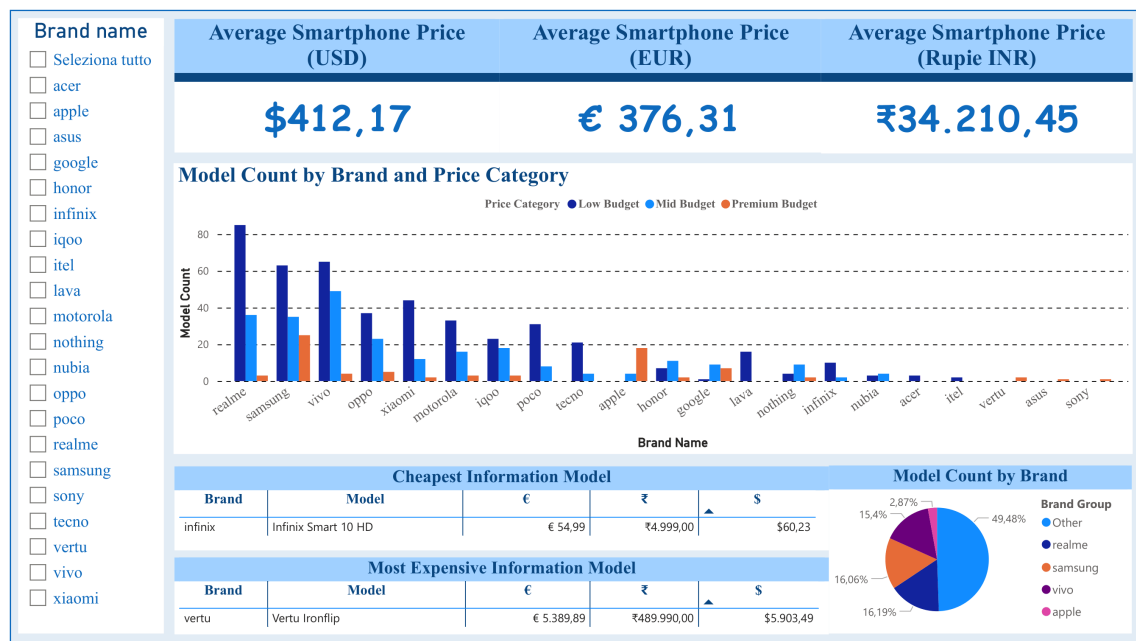


Figura 4.1: Dashboard 1 - Panoramica generale del mercato degli smartphone

- *Average Smartphone Price (USD):* \$412,17;
- *Average Smartphone Price (EUR):* €376,31;
- *Average Smartphone Price (INR):* ₹34.210,45.

Grafici utilizzati

- *Grafico a colonne raggruppate:* rappresenta il numero di modelli per ciascun brand suddivisi per categoria di prezzo, evidenziando come i diversi produttori si distribuiscono tra fasce economiche, medie e premium.
- *Tabelle riepilogative:* mostrano, rispettivamente, il modello più economico e quello più costoso presente nel dataset, permettendo di individuare rapidamente gli estremi dell'offerta di mercato.
- *Grafico a torta:* illustra la distribuzione dei modelli per brand, evidenziando la percentuale relativa di ciascun produttore all'interno del mercato analizzato.

Nota: per migliorare la leggibilità del grafico sono stati visualizzati esclusivamente i brand maggiormente rappresentati nel dataset (Samsung, Apple, Realme, Vivo), aggregando tutti gli altri produttori nella categoria *Other*.

Conclusioni sull'analisi della Dashboard 1

L'analisi della Dashboard 1 ha permesso di ottenere una visione d'insieme delle principali caratteristiche del mercato degli smartphone. I KPI relativi ai prezzi medi consentono di osservare il valore medio degli smartphone espresso nelle diverse valute considerate (USD, EUR e INR), facilitando la lettura dei dati in contesti geografici differenti.

L'analisi per brand mostra come alcuni produttori concentrino la propria offerta in specifiche fasce di prezzo, mentre altri adottano strategie più diversificate, coprendo contemporaneamente segmenti economici, medi e premium.

Questo comportamento evidenzia differenti approcci di posizionamento competitivo all'interno del mercato.

Inoltre, l'individuazione dei modelli più economici e più costosi consente di comprendere meglio l'ampiezza dell'offerta disponibile, mettendo in luce la presenza di dispositivi con caratteristiche e prezzi estremamente variabili.

Nel complesso, la dashboard si configura come uno strumento efficace per l'analisi descrittiva del mercato, permettendo di individuare rapidamente pattern rilevanti e supportando una comprensione immediata delle dinamiche competitive tra i diversi brand.

4.4.2 Dashboard 2 - Analisi delle caratteristiche tecniche degli smartphone

La Dashboard 2 consente di analizzare la distribuzione delle principali caratteristiche tecniche degli smartphone e le relazioni esistenti tra esse, permettendo di individuare configurazioni ricorrenti e pattern tecnologici nel mercato.

In particolare, essa consente di esaminare:

- La distribuzione delle principali tecnologie di connettività tra cui:
 - 5G (Fifth Generation): è la quinta generazione delle reti mobili, che consente velocità di connessione molto elevate, bassa latenza e una maggiore capacità di gestione dei dispositivi connessi.
 - NFC (Near Field Communication): è una tecnologia di comunicazione a corto raggio che permette lo scambio di dati tra dispositivi vicini, utilizzata principalmente per pagamenti contactless e trasferimento rapido di informazioni.
 - IR (Infrared): è una tecnologia basata sulla trasmissione di segnali tramite raggi infrarossi, utilizzata negli smartphone per controllare dispositivi elettronici come televisori e condizionatori.
- La configurazione delle fotocamere (rear camera e front camera).
- La dimensione degli schermi in pollici.
- La combinazione tra memoria RAM e capacità di archiviazione.
- Il numero di core dei processori degli smartphone presenti sul mercato.

Attraverso l'utilizzo di grafici diversificati, è possibile ottenere una visione dettagliata delle caratteristiche più diffuse sul mercato, individuando le configurazioni maggiormente adottate dai produttori e comprendendo meglio le tendenze tecnologiche attuali. L'analisi congiunta delle diverse visualizzazioni consente, inoltre, di evidenziare relazioni significative tra le principali componenti hardware, mettendo in luce le configurazioni che caratterizzano i dispositivi più diffusi e supportando una comprensione più approfondita dell'evoluzione tecnologica del settore degli smartphone.

Nella Figura 4.2 è mostrata la Dashboard 2 – Analisi delle caratteristiche tecniche degli smartphone.

Grafici utilizzati

- *Grafici a torta*: rappresentano la distribuzione delle principali tecnologie di connettività (5G, NFC e IR), evidenziando la percentuale di dispositivi che supportano ciascuna funzionalità.

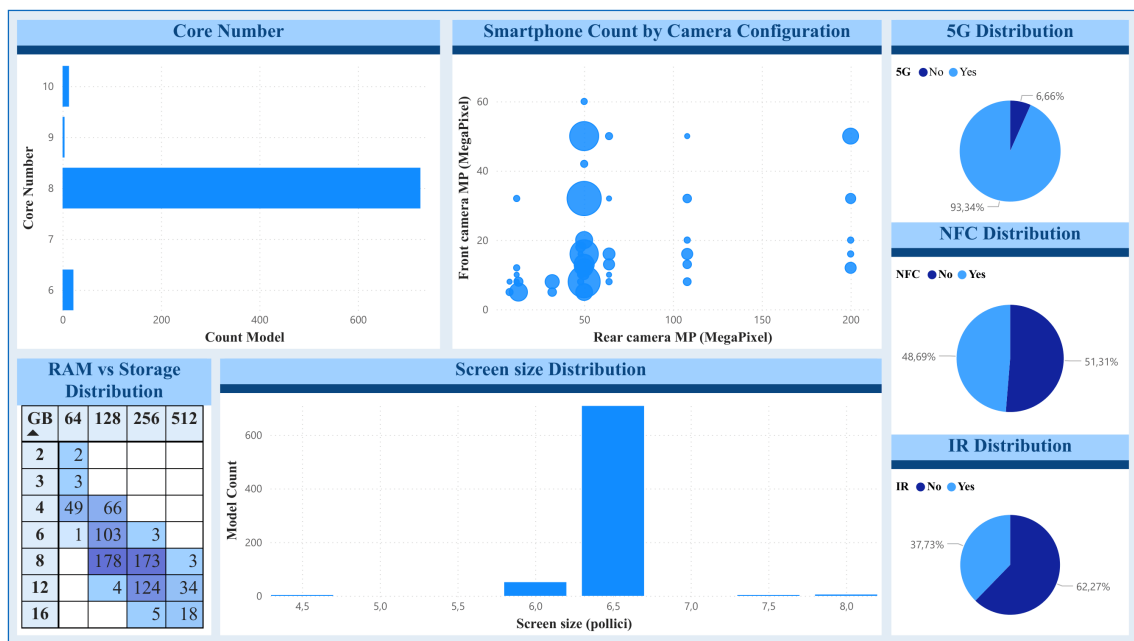


Figura 4.2: Dashboard 2 – Analisi delle caratteristiche tecniche degli smartphone

- *Grafico a dispersione (Scatter Plot):* mostra la relazione tra la risoluzione della fotocamera posteriore e quella della fotocamera frontale, permettendo di individuare eventuali correlazioni tra le due caratteristiche.
- *Heatmap (matrice):* evidenzia le combinazioni più frequenti tra memoria RAM e capacità di archiviazione (Storage), consentendo di individuare le configurazioni più diffuse sul mercato.
- *Istogramma:* rappresenta la distribuzione della dimensione degli schermi (in pollici), mostrando la concentrazione dei dispositivi nelle diverse fasce dimensionali.
- *Grafico a barre:* illustra la distribuzione del numero di core dei processori, fornendo una panoramica sulle configurazioni hardware più comuni.

Conclusioni sull'analisi della Dashboard 2

L'analisi della Dashboard 2 evidenzia come alcune caratteristiche tecniche siano ormai standard nel mercato degli smartphone. In particolare, la tecnologia 5G risulta ampiamente diffusa, mentre funzionalità come NFC e IR presentano una distribuzione più eterogenea, indicando una differenziazione nell'adozione delle tecnologie tra i vari dispositivi.

La relazione tra le fotocamere mostra una certa variabilità, suggerendo che una maggiore risoluzione della fotocamera posteriore non implica necessariamente prestazioni analoghe della fotocamera frontale. Questo evidenzia come le scelte progettuali dei produttori possano privilegiare specifici aspetti hardware rispetto ad altri.

La heatmap relativa a RAM e memoria evidenzia una forte concentrazione nelle configurazioni intermedie, in particolare nelle combinazioni da 8 GB di RAM con 128 GB o 256 GB di archiviazione, indicando una preferenza del mercato per dispositivi bilanciati tra prestazioni e costo.

Per quanto riguarda la dimensione degli schermi, la maggior parte degli smartphone si concentra nella fascia compresa tra 6 e 6.5 pollici, confermando una tendenza verso dispositivi con display ampi e uniformità nelle dimensioni.

Infine, l'analisi del numero di core mostra una predominanza di processori a 8 core, evidenziando uno standard consolidato nelle configurazioni hardware dei dispositivi moderni.

Nel complesso, la dashboard consente di comprendere in modo chiaro le principali tendenze tecnologiche del mercato, evidenziando le configurazioni più diffuse e supportando un'analisi comparativa delle caratteristiche tecniche degli smartphone.

4.4.3 Dashboard 3 - Analisi delle caratteristiche tecniche per fascia di prezzo

Questa dashboard ha l'obiettivo di analizzare le principali caratteristiche tecniche degli smartphone in relazione alla fascia di prezzo di appartenenza; quest'ultima può essere:

- Low Budget,
- Mid Budget,
- Premium Budget.

Lo scopo è comprendere come le specifiche hardware e le prestazioni dei dispositivi variano al variare del prezzo, evidenziando eventuali differenze tra le diverse categorie di mercato. La dashboard consente, inoltre, di esplorare i dati in modo dinamico attraverso un filtro dedicato alla *Price Category*, permettendo di focalizzare l'analisi su specifiche fasce di prezzo e facilitando il confronto tra dispositivi economici, di fascia media e premium.

Nella Figura 4.3 è mostrata la Dashboard 3 – Analisi delle caratteristiche tecniche per fascia di prezzo.

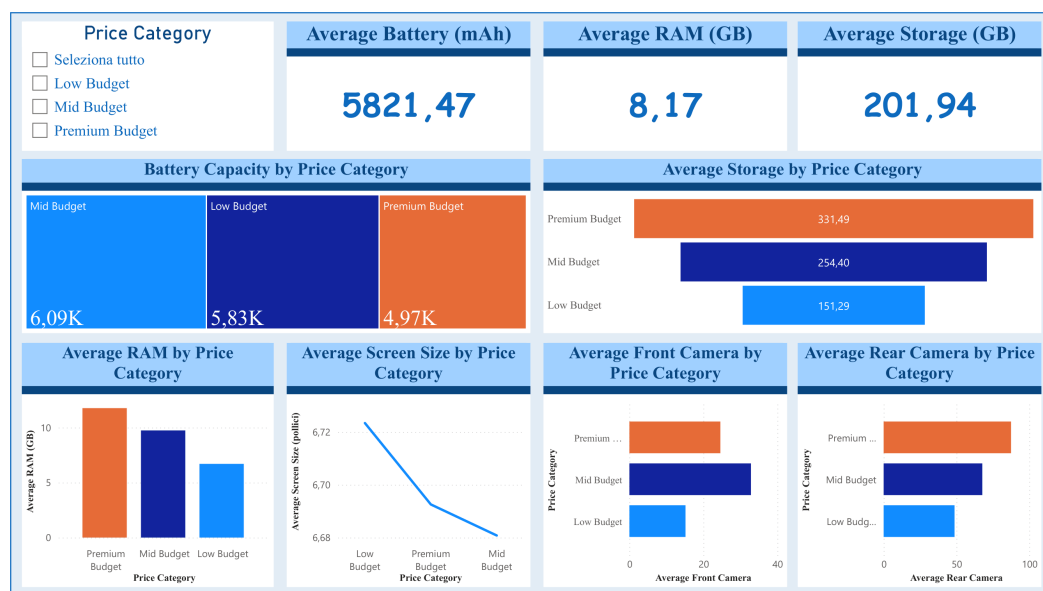


Figura 4.3: Dashboard 3 - Analisi delle caratteristiche tecniche per fascia di prezzo

Key Performance Indicator (KPI) utilizzati

Considerando l'intero dataset senza applicare filtri specifici, i principali indicatori risultano:

- *Average Battery Capacity (mAh)*: 5.821,47 mAh;
- *Average RAM (GB)*: 8,17 GB;
- *Average Storage (GB)*: 201,94 GB.

Grafici utilizzati

- *Treemap*: rappresenta la capacità media della batteria per ciascuna fascia di prezzo.
- *Grafico a barre orizzontali*: mostra la capacità media di archiviazione (storage) per fascia di prezzo.
- *Grafico a colonne*: illustra la quantità media di RAM per ciascuna categoria di prezzo.
- *Grafico a linee*: rappresenta la dimensione media dello schermo (in pollici) per fascia di prezzo.
- *Grafici a barre orizzontali*: mostrano la qualità media delle fotocamere (anteriore e posteriore) per fascia di prezzo.

Conclusioni sull'analisi della Dashboard 3

L'analisi della Dashboard 3 evidenzia differenze significative tra le diverse fasce di prezzo in termini di caratteristiche tecniche.

In particolare, emerge come i dispositivi di fascia premium offrano generalmente prestazioni superiori in termini, ad esempio, di RAM e capacità di archiviazione, risultando orientati a garantire elevate performance complessive.

Al contrario, i dispositivi di fascia media presentano una capacità della batteria mediamente più elevata rispetto ai modelli premium, suggerendo una diversa strategia di progettazione orientata all'autonomia piuttosto che alle prestazioni.

Questo fenomeno è riconducibile al posizionamento di mercato dei dispositivi premium, progettati per offrire elevate prestazioni. In tali dispositivi, l'attenzione si concentra maggiormente su aspetti come velocità e capacità di multitasking, mentre la capacità della batteria, pur rimanendo un elemento fondamentale, può risultare meno prioritaria rispetto ad altri componenti hardware.

Per quanto riguarda la dimensione dello schermo, non si osservano variazioni particolarmente marcate tra le diverse categorie, indicando come questa caratteristica sia relativamente uniforme tra i dispositivi analizzati.

Nel complesso, la dashboard consente di comprendere come il prezzo influenzi direttamente le specifiche tecniche degli smartphone, evidenziando un chiaro trade-off tra prestazioni, autonomia e caratteristiche hardware nelle diverse fasce di mercato.

4.4.4 Dashboard 4 - Analisi dei sistemi operativi negli smartphone

La Dashboard 4 è dedicata all'analisi dei sistemi operativi utilizzati negli smartphone presenti nel dataset, con l'obiettivo di comprendere la distribuzione dei dispositivi tra Android e iOS e le principali differenze in termini di caratteristiche tecniche e posizionamento nel mercato.

In particolare, la dashboard consente di esaminare:

- la distribuzione complessiva dei modelli per sistema operativo;
- la diffusione delle diverse versioni di Android e iOS;

- la distribuzione dei sistemi operativi nelle diverse fasce di prezzo;
- il confronto tra le caratteristiche hardware medie dei dispositivi Android e iOS.

La dashboard permette, quindi, di analizzare non solo la predominanza di un sistema operativo rispetto all'altro, ma anche le differenze strutturali tra le due piattaforme in termini di specifiche tecniche e segmentazione di mercato.

Nella Figura 4.4 è mostrata la Dashboard 4 – Analisi dei sistemi operativi negli smartphone.

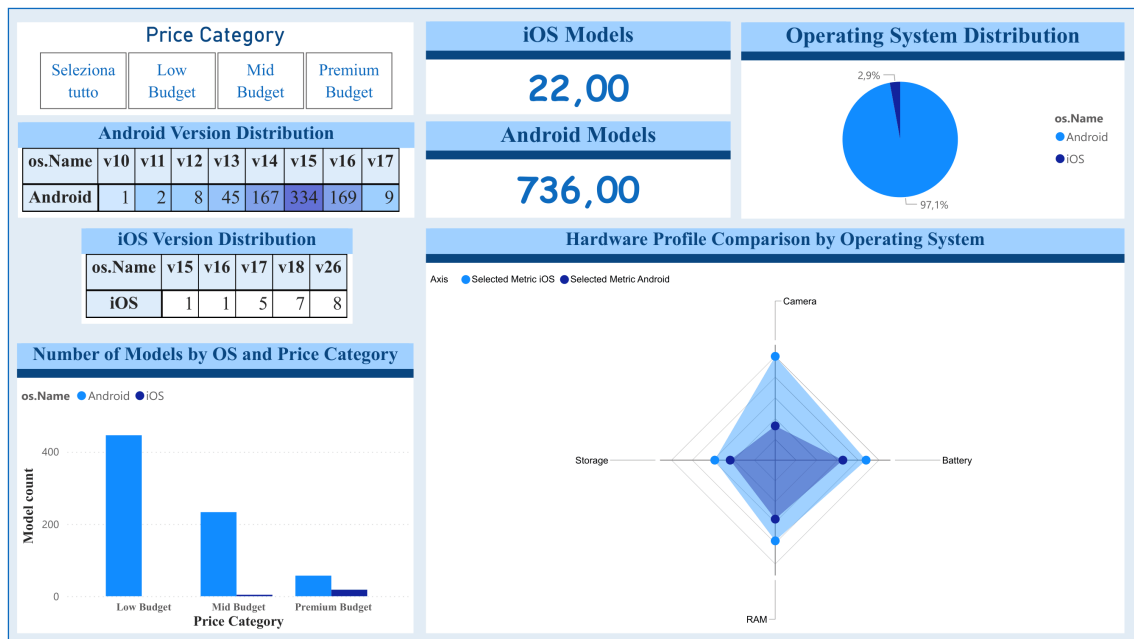


Figura 4.4: Dashboard 4 - Analisi dei sistemi operativi negli smartphone

Key Performance Indicator (KPI) utilizzati

Considerando l'intero dataset, e in assenza di filtri applicati, si ottengono i seguenti valori:

- *Android Models*: 736;
- *iOS Models*: 22.

Grafici utilizzati

- *Grafico a torta*: rappresenta la distribuzione complessiva dei modelli per sistema operativo, evidenziando la netta predominanza dei dispositivi Android rispetto a quelli iOS.
- *Grafico a colonne*: illustra il numero di modelli per sistema operativo e fascia di prezzo, permettendo di osservare la maggiore diffusione dei dispositivi Android in tutte le categorie.
- *Heatmap (matrice)*: mostra la diffusione delle diverse versioni dei sistemi operativi Android e iOS, evidenziando una maggiore frammentazione nel caso di Android rispetto a iOS.

- *Grafico radar (o a ragnatela)*: consente di confrontare il profilo medio delle principali caratteristiche hardware (RAM, capacità di archiviazione, batteria e fotocamera posteriore) tra dispositivi Android e iOS, utilizzando valori normalizzati per garantire la comparabilità tra grandezze eterogenee.

Conclusioni sull'analisi della Dashboard 4

L'analisi della Dashboard 4 evidenzia, innanzitutto, una forte predominanza dei dispositivi Android all'interno del dataset, che rappresentano la quasi totalità dei modelli analizzati, mentre i dispositivi iOS risultano presenti in una quota marginale.

Tale risultato è riconducibile al fatto che la maggior parte delle case produttrici utilizza il sistema operativo Android per i propri dispositivi, mentre iOS è adottato esclusivamente da Apple.

Per quanto riguarda la distribuzione per fascia di prezzo, emerge come i dispositivi iOS siano concentrati principalmente nel segmento premium, mentre Android copre in modo più uniforme tutte le categorie di mercato, dai dispositivi economici a quelli di fascia alta.

L'analisi delle versioni dei sistemi operativi evidenzia, inoltre, una maggiore frammentazione nel caso di Android, caratterizzato dalla presenza di numerose versioni differenti, mentre iOS mostra una distribuzione più concentrata e omogenea.

Il grafico radar consente di confrontare il profilo medio delle principali caratteristiche hardware tra dispositivi Android e iOS, evidenziando valori medi più elevati per i dispositivi iOS. Tuttavia, tali risultati devono essere interpretati con cautela. Il dataset analizzato risulta, infatti, sbilanciato: Android comprende dispositivi appartenenti a tutte le fasce di prezzo, mentre iOS è rappresentato esclusivamente da dispositivi di fascia alta. Questa differenza nella composizione comporta un innalzamento delle medie per iOS e una riduzione di quelle per Android. Pertanto, il confronto tra i due sistemi operativi risulta influenzato dalla distribuzione dei dispositivi nel dataset, piuttosto che da una reale superiorità hardware di una piattaforma rispetto all'altra.

Nel complesso, la Dashboard 4 consente di evidenziare come le differenze tra Android e iOS non riguardino esclusivamente la diffusione sul mercato, ma coinvolgono anche il posizionamento strategico dei produttori e le caratteristiche tecniche dei dispositivi associati alle due piattaforme.

4.4.5 Dashboard 5 - Analisi della soddisfazione degli utenti e del rapporto qualità/prezzo

La Dashboard 5 è dedicata all'analisi della soddisfazione degli utenti relativamente ai principali modelli di smartphone presenti nel mercato. L'obiettivo della dashboard è quello di comprendere:

- la relazione tra prezzo e valutazione media degli utenti;
- il comportamento dei principali brand di smartphone;
- quali caratteristiche hardware risultano maggiormente associate alla soddisfazione degli utenti;
- quali smartphone presentano il miglior e il peggior rapporto qualità/prezzo.

Nella Figura 4.5 è mostrata la Dashboard 5 – Analisi della soddisfazione degli utenti e del rapporto qualità/prezzo.

Per la realizzazione della dashboard è stato utilizzato il dataset ottenuto tramite il merge tra:

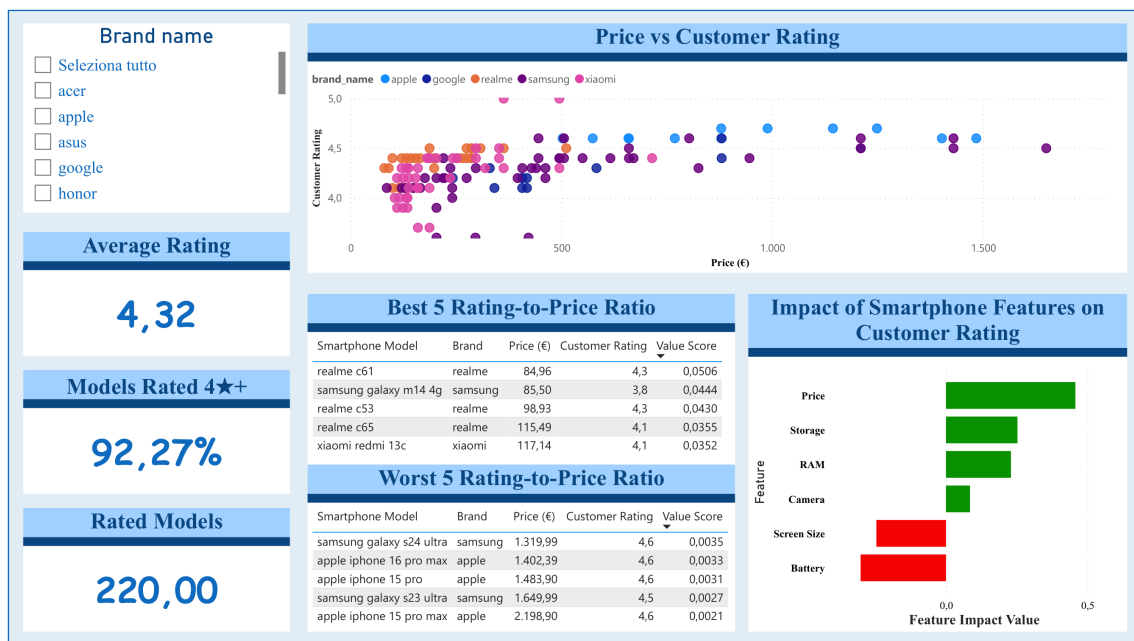


Figura 4.5: Dashboard 5 - Analisi della soddisfazione degli utenti e del rapporto qualità/prezzo

- il dataset principale contenente le specifiche hardware degli smartphone;
- il dataset contenente valutazioni degli utenti, recensioni e data di rilascio dei dispositivi.

Per evitare la presenza di modelli duplicati dovuti alle differenti configurazioni hardware, è stata successivamente creata una tabella aggregata contenente una sola riga per ciascun modello di smartphone.

Creazione della tabella aggregata

La tabella `Smartphone_Value_Unique` è stata implementata con l'obiettivo di:

- eliminare i duplicati presenti nel dataset merged;
- ottenere una singola riga per ogni smartphone;
- calcolare valori medi relativi a prezzo e valutazione.

```

1 Smartphone_Value_Unique =
2 SUMMARIZE (
3     Smartphone_dataset_Merged,
4     Smartphone_dataset_Merged[ModelKey],
5     Smartphone_dataset_Merged[brand_name],
6     "avg_rating", AVERAGE (Smartphone_dataset_Merged[avg_rating]),
7     "Price_EUR", AVERAGE (Smartphone_dataset_Merged[Price_EUR]),
8     "Rating Price Score",
9     DIVIDE (
10         AVERAGE (Smartphone_dataset_Merged[avg_rating]),
11         AVERAGE (Smartphone_dataset_Merged[Price_EUR])
12     )
13 )

```

Listato 4.40: Tabella DAX `Smartphone_Value_Unique`

Successivamente è stata creata una relazione tra:

- Smartphone_dataset_Merged;
- Smartphone_Value_Unique;

tramite la colonna:

ModelKey

in modo da rendere interattivi slicer e filtri presenti nella dashboard.

Key Performance Indicator (KPI) utilizzati

Per sintetizzare rapidamente le principali informazioni della dashboard sono stati implementati diversi KPI tramite misure DAX *Average Rating*, *Rated Models* e *High Rated Percentage* descritte in precedenza nella sezione dedicata alle misure DAX. Tali misure sono:

- *Average Rating*: rappresenta la valutazione media degli utenti relativamente ai modelli presenti nel dataset.
- *Rated Models*: rappresenta il numero totale di smartphone per cui è disponibile una valutazione media.
- *Models Rated 4+*: rappresenta la percentuale di smartphone caratterizzati da una valutazione pari o superiore a 4 stelle.

Grafici utilizzati

- *Scatter Plot*: rappresenta la relazione tra il prezzo degli smartphone e la valutazione media espressa dagli utenti. L'asse X rappresenta il prezzo dello smartphone espresso in Euro, mentre l'asse Y rappresenta la valutazione media degli utenti. I punti sono, inoltre, differenziati tramite colore in base al brand dello smartphone, permettendo di confrontare il comportamento dei principali produttori presenti sul mercato. Questa visualizzazione consente di osservare il comportamento dei diversi brand e di individuare eventuali relazioni tra fascia di prezzo e soddisfazione degli utenti.
- *Tabelle comparative*: rappresentano gli smartphone caratterizzati dal miglior e peggior rapporto qualità/prezzo. Le tabelle mostrano:
 - il modello dello smartphone;
 - il brand;
 - il prezzo medio;
 - la valutazione media;
 - l'indicatore del rapporto qualità/prezzo.

Queste visualizzazioni permettono di individuare sia gli smartphone più convenienti rispetto alla qualità percepita dagli utenti, sia i dispositivi caratterizzati da un prezzo elevato rispetto al livello di soddisfazione ottenuto.

- *Grafico a barre orizzontali*: rappresenta il coefficiente di correlazione tra le principali caratteristiche hardware degli smartphone e la valutazione media espressa dagli utenti. Il grafico permette di osservare quali caratteristiche risultano maggiormente associate alla soddisfazione degli utenti.

Le barre verdi rappresentano correlazioni positive, mentre le barre rosse rappresentano correlazioni negative. Tra le caratteristiche considerate vi sono:

- la RAM;
- lo storage;
- la batteria;
- la fotocamera;
- la dimensione dello schermo;
- il prezzo.

Analisi del rapporto qualità/prezzo

Per identificare gli smartphone con il miglior e peggior rapporto qualità/prezzo è stata implementata la misura `DAX Rating Price Score`.

Questa misura è risultata fondamentale poiché ha permesso di costruire un indicatore sintetico in grado di combinare due aspetti molto importanti nella valutazione di uno smartphone:

- la valutazione media espressa dagli utenti;
- il prezzo medio del dispositivo.

Il rapporto qualità/prezzo rappresenta, infatti, uno dei criteri più rilevanti nel processo di acquisto di uno smartphone, poiché consente di valutare se il livello di soddisfazione degli utenti sia proporzionato al costo del dispositivo.

Per questo motivo, sulla base della misura `Rating Price Score`, sono state realizzate due tabelle comparative:

- *Top 5 Best Rating-to-Price Ratio*;
- *Worst 5 Rating-to-Price Ratio*.

La prima tabella consente di individuare gli smartphone che offrono il miglior rapporto tra valutazione media e prezzo, mentre la seconda permette di evidenziare i dispositivi che presentano un rapporto meno vantaggioso tra costo e soddisfazione degli utenti.

La misura `DAX` implementata è la seguente:

```
1 Rating Price Score =
2 DIVIDE(
3     AVERAGE(Smartphone_dataset_Merged[avg_rating]),
4     AVERAGE(Smartphone_dataset_Merged[Price_EUR])
5 )
```

Listato 4.41: Misura `DAX Rating Price Score`

La misura calcola il rapporto tra la valutazione media dello smartphone e il prezzo medio del dispositivo. In particolare, al numeratore viene considerata la media della valutazione degli utenti, mentre al denominatore viene considerato il prezzo medio espresso in Euro.

Valori più elevati della misura indicano smartphone caratterizzati da una valutazione media alta rispetto al prezzo sostenuto, e quindi da un miglior rapporto qualità/prezzo. Al contrario, valori più bassi indicano dispositivi per i quali il prezzo risulta meno vantaggioso rispetto alla soddisfazione media espressa dagli utenti.

È importante sottolineare che la misura `Rating Price Score` non ha l'obiettivo di identificare semplicemente lo smartphone "migliore" in termini assoluti, ma piuttosto, di valutare quale dispositivo sia in grado di offrire il livello più elevato di soddisfazione degli utenti rispetto al prezzo sostenuto.

L'analisi evidenzia come molti smartphone appartenenti alle fasce economiche e di fascia media presentino un rapporto qualità/prezzo particolarmente vantaggioso, ottenendo valutazioni elevate da parte degli utenti nonostante un costo relativamente contenuto.

Al contrario, gli smartphone flagship e premium, pur caratterizzati da valutazioni mediamente superiori, presentano un valore relativo inferiore a causa del prezzo significativamente più elevato.

Questo fenomeno può essere interpretato come un esempio di *diminishing returns*: all'aumentare del prezzo del dispositivo, il miglioramento percepito dagli utenti non cresce in maniera proporzionale rispetto al costo sostenuto.

La dashboard realizzata evidenzia, quindi, come il valore percepito dagli utenti non dipenda esclusivamente dalla qualità assoluta del dispositivo, ma anche dal rapporto tra prestazioni offerte e prezzo di acquisto.

Impact of Smartphone Features on Customer Rating

L'ultima visualizzazione implementata consiste in un grafico a barre orizzontali utilizzato per analizzare la relazione tra le principali caratteristiche hardware degli smartphone e la valutazione media espressa dagli utenti.

L'obiettivo di questa analisi è comprendere quanto una determinata caratteristica tecnica risulti associata al livello di soddisfazione degli utenti.

Per realizzare questa visualizzazione è stata inizialmente creata una tabella di supporto contenente le caratteristiche hardware considerate nell'analisi:

```
1 Feature Impact =
2 DATATABLE(
3     "Feature", STRING,
4     {
5         {"Camera"},
6         {"Battery"},
7         {"RAM"},
8         {"Storage"},
9         {"Screen Size"},
10        {"Price"}
11    }
12 )
```

Listato 4.42: Tabella DAX Feature Impact

Successivamente è stata implementata la misura DAX Feature Impact Value, utilizzata per calcolare il coefficiente di correlazione tra la valutazione media degli utenti e ciascuna caratteristica hardware selezionata.

La misura DAX implementata è la seguente:

```
1 Feature Impact Value =
2 VAR SelectedFeature =
3     SELECTEDVALUE('Feature Impact'[Feature])
4
5 VAR FeatureTable =
6     FILTER(
7         ADDCOLUMNS(
8             Smartphone_dataset_Merged,
9             "X",
10            SWITCH(
11                SelectedFeature,
12                "Camera", Smartphone_dataset_Merged[rear_camera_MP],
13                "Battery", Smartphone_dataset_Merged[battery_capacity_mAh],
14                "RAM", Smartphone_dataset_Merged[ram_GB],
15                "Storage", Smartphone_dataset_Merged[memory_GB],
```

```

16         "Screen Size", Smartphone_dataset_Merged[Screen_size_pollici],
17         "Price", Smartphone_dataset_Merged[Price_EUR]
18     ),
19     "Y", Smartphone_dataset_Merged[avg_rating]
20 ),
21     NOT (ISBLANK([X])) &&
22     NOT (ISBLANK([Y]))
23 )
24
25 VAR AvgX =
26     AVERAGEX (FeatureTable, [X])
27
28 VAR AvgY =
29     AVERAGEX (FeatureTable, [Y])
30
31 VAR Numerator =
32     SUMX (
33         FeatureTable,
34         ([X] - AvgX) * ([Y] - AvgY)
35     )
36
37 VAR Denominator =
38     SQRT (
39         SUMX (FeatureTable, POWER([X] - AvgX, 2)) *
40         SUMX (FeatureTable, POWER([Y] - AvgY, 2))
41     )
42
43 RETURN
44     DIVIDE (Numerator, Denominator)

```

Listato 4.43: Misura DAX Feature Impact Value

La misura restituisce un coefficiente di correlazione compreso tra -1 e 1. Quest'ultimo rappresenta il grado di relazione esistente tra due variabili. In questo caso viene misurata la relazione tra una specifica caratteristica hardware dello smartphone e la valutazione media espressa dagli utenti.

In particolare:

- valori positivi indicano una correlazione diretta;
- valori negativi indicano una correlazione inversa;
- valori prossimi allo zero indicano una correlazione debole o poco significativa.

Una *correlazione positiva* indica che all'aumentare della caratteristica hardware considerata tende ad aumentare anche la valutazione media degli utenti.

Ad esempio, una correlazione positiva tra RAM e avg_rating suggerisce che smartphone dotati di una maggiore quantità di memoria RAM tendono ad ottenere valutazioni mediamente più elevate da parte degli utenti.

Al contrario, una *correlazione negativa* indica che all'aumentare della caratteristica considerata tende a diminuire la valutazione media degli utenti.

Ad esempio, una correlazione negativa tra dimensione dello schermo e valutazione media suggerisce che nel dataset analizzato gli smartphone con display più grandi tendono ad ottenere valutazioni mediamente inferiori.

È importante sottolineare che la correlazione non implica necessariamente un rapporto di causa-effetto, ma rappresenta esclusivamente una misura statistica della relazione tra due variabili.

Nel grafico finale:

- le barre verdi rappresentano correlazioni positive;
- le barre rosse rappresentano correlazioni negative.

Questa visualizzazione consente, quindi, di identificare rapidamente quali caratteristiche hardware risultano maggiormente associate alla soddisfazione degli utenti nel dataset analizzato.

Conclusioni sull'analisi della Dashboard 5

La dashboard realizzata ha permesso di ottenere una panoramica dettagliata del livello di soddisfazione degli utenti relativamente ai principali modelli di smartphone presenti sul mercato.

L'integrazione tra caratteristiche hardware, prezzo dei dispositivi e valutazioni espresse dagli utenti, ha consentito di analizzare il mercato degli smartphone non soltanto dal punto di vista tecnico, ma anche dal punto di vista della percezione e dell'esperienza degli utenti finali.

Dall'analisi effettuata emerge una tendenza secondo cui gli smartphone appartenenti alle fasce premium tendono ad ottenere valutazioni medie più elevate, suggerendo una possibile relazione tra qualità percepita e prezzo del dispositivo. Tuttavia, la dashboard ha anche permesso di individuare diversi modelli appartenenti a fasce di prezzo più contenute, che presentano un ottimo rapporto qualità/prezzo, risultando particolarmente competitivi rispetto alla soddisfazione espressa dagli utenti.

L'analisi delle correlazioni ha, inoltre, evidenziato che alcune caratteristiche hardware, come RAM, storage e prezzo, risultano maggiormente associate a valutazioni positive da parte degli utenti, mentre altre caratteristiche mostrano relazioni meno significative, o addirittura inverse, nel dataset analizzato.

Le tabelle comparative implementate hanno permesso di identificare rapidamente:

- gli smartphone caratterizzati dal miglior rapporto tra valutazione media e prezzo;
- i dispositivi con rapporto qualità/prezzo meno vantaggioso;
- eventuali differenze di comportamento tra i vari brand presenti nel mercato.

Dal punto di vista della Business Intelligence, la dashboard realizzata rappresenta uno strumento utile per:

- confrontare il posizionamento dei differenti brand;
- comprendere le preferenze degli utenti;
- identificare le caratteristiche maggiormente apprezzate nel mercato degli smartphone;
- supportare eventuali decisioni strategiche o analisi comparative.

L'utilizzo combinato di KPI, scatter plot, tabelle comparative e grafici di correlazione ha permesso di realizzare una dashboard interattiva, dinamica e facilmente interpretabile, in grado di fornire una visione completa del mercato degli smartphone dal punto di vista della customer satisfaction e del rapporto qualità/prezzo.

Progettazione e realizzazione delle dashboard avanzate

In questo capitolo verrà illustrato il concetto di analisi predittiva e si esaminerà il suo utilizzo in Power BI attraverso l'analisi delle dashboard predittive realizzate.

5.1 Introduzione all'analisi predittiva

L'analisi predittiva rappresenta una branca dell'analitica avanzata e costituisce uno degli ambiti più rilevanti della moderna Business Intelligence. Essa consiste nell'utilizzo di dati storici, algoritmi statistici e tecniche di Machine Learning per stimare la probabilità del verificarsi di eventi futuri, supportando, così, i processi decisionali aziendali. Le aziende, attraverso tecniche di forecasting e modelli statistici, utilizzano l'analisi predittiva per costruire modelli predittivi con lo scopo di identificare rischi e opportunità, individuare trend, prevedere comportamenti futuri e supportare decisioni strategiche basate sui dati.

L'analisi predittiva può essere implementata in vari settori per diversi problemi aziendali. Tra i principali ambiti applicativi si possono individuare i seguenti casi d'uso per settore, che illustrano come l'analisi predittiva possa informare il processo decisionale in situazioni reali:

- **Settore bancario:** i servizi finanziari utilizzano l'apprendimento automatico e strumenti quantitativi per fare previsioni sui clienti potenziali e quelli attuali. L'impiego di tali modelli consente agli istituti finanziari di stimare il rischio di insolvenza, classificare i clienti in base al profilo di rischio, individuare i clienti più redditizi e rilevare potenziali operazioni fraudolente.
- **Sanità:** nel settore sanitario l'analisi predittiva viene utilizzata per rilevare e gestire la cura dei pazienti affetti da malattie croniche, oltre che per monitorare infezioni specifiche, favorendo interventi preventivi e una migliore pianificazione delle risorse sanitarie.
- **Risorse umane (HR):** i team HR utilizzano l'analisi predittiva e le metriche dei sondaggi dei dipendenti per trovare potenziali candidati, ridurre il turnover e aumentare il coinvolgimento dei dipendenti. Questa combinazione di dati quantitativi e qualitativi consente alle aziende di ridurre i costi di assunzione e migliorare il livello di soddisfazione dei dipendenti, il che è particolarmente utile nei momenti di volatilità del mercato del lavoro.

- *Marketing e vendite*: i team di marketing utilizzano l'analisi predittiva per analizzare l'andamento storico delle vendite. Inoltre, l'analisi predittiva consente alle aziende di personalizzare le strategie di interazione con i clienti lungo l'intero ciclo di vita.
- *Supply chain*: le aziende utilizzano comunemente l'analisi predittiva per gestire l'inventario dei prodotti e impostare strategie di determinazione dei prezzi. L'analisi predittiva aiuta le aziende a soddisfare la domanda dei clienti senza sovraccaricare i magazzini. Essa consente, inoltre, alle aziende di valutare il costo e il rendimento dei loro prodotti nel tempo.

Un'organizzazione che sa cosa aspettarsi in base ai modelli passati ha un vantaggio aziendale nella gestione delle scorte, della forza lavoro, delle campagne di marketing e della maggior parte degli altri aspetti operativi. Oltre alle applicazioni nei diversi settori produttivi, l'analisi predittiva viene impiegata per supportare specifiche attività aziendali, tra cui:

- *Il rilevamento delle frodi*: la combinazione di più metodi di analisi può migliorare l'individuazione di schemi ("pattern") e prevenire i comportamenti criminali. Con il crescente numero di minacce informatiche, l'analisi comportamentale ad alte prestazioni permette di esaminare in tempo reale tutte le azioni avvenute su una rete per individuare anomalie che potrebbero indicare frodi, falle di sicurezza e minacce persistenti avanzate.
- *L'ottimizzazione delle campagne di marketing*: l'analisi predittiva viene utilizzata per determinare le risposte o gli acquisti dei clienti, nonché per promuovere opportunità di cross-selling. I modelli predittivi aiutano le aziende ad attirare, fidelizzare e incrementare il valore dei clienti più redditizi.
- *Il miglioramento delle operazioni*: molte aziende utilizzano i modelli predittivi per prevedere le rimanenze di magazzino e gestire le risorse. Le compagnie aeree utilizzano l'analisi predittiva per determinare i prezzi dei biglietti. Gli hotel cercano di prevedere il numero di ospiti per una determinata notte in modo da massimizzare la percentuale di occupazione e aumentare i ricavi. Questi esempi evidenziano come l'analisi predittiva consenta di ottimizzare l'impiego delle risorse e di migliorare l'efficienza operativa in differenti contesti applicativi.
- *Riduzione dei rischi*: i punteggi di affidabilità creditizia che vengono utilizzati per valutare la probabilità di inadempienza di un acquirente sono un ben noto esempio di analisi predittiva. Il "credit score" è, in sostanza, un numero generato da un modello predittivo che incorpora tutti i dati relativi all'affidabilità creditizia di una persona. Altri usi connessi ai rischi includono indennizzi assicurativi e riscossioni.

5.2 Analisi predittiva in Power BI

Power BI non è limitato all'analisi dei dati correnti, ma può essere utilizzato anche nella previsione delle tendenze future. Esistono, principalmente, due strumenti con cui realizzare analisi predittive in Power BI:

- *Funzionalità di forecasting integrate*: Power BI mette a disposizione dell'utente un insieme di strumenti per l'analisi temporale al fine di esplorare trend e andamenti storici e fare previsioni basate su modelli predittivi. Le funzionalità di forecasting sono accessibili e intuitive per gli utenti, poiché permettono di modificare i parametri di previsione, come lunghezza, stagionalità e intervallo di confidenza, direttamente dal menù grafico nel pannello delle visualizzazioni.

- *Estensioni integrative per analisi previsionali:* Power BI è in grado di aggiungere e utilizzare al suo interno strumenti di analisi come R e Python, consentendo agli utenti di sfruttare le librerie di tali linguaggi per creare modelli personalizzati di apprendimento automatico e tecniche statistiche per generare previsioni.

Forecasting integrato in Power BI

Power BI offre molteplici funzionalità di analisi avanzata dei dati. Tra le più interessanti evidenziamo il built-in forecasting. Questo strumento permette di effettuare un'analisi predittiva in modo semplice e rapido, mantenendo la stessa granularità dei dati.

Dopo aver importato il dataset, l'analisi temporale tramite la funzionalità integrata comporta l'utilizzo di un grafico a linee (line chart); questo perché è indispensabile avere un riferimento temporale continuo nell'asse delle ascisse. Nell'asse delle ordinate viene inserita la misura aggregata, rispetto a livello di granularità del punto dati per la quale vogliamo calcolare la previsione. Questa funzionalità permette di generare automaticamente previsioni future a partire da serie temporali di dati storici, senza la necessità di implementare manualmente modelli statistici complessi.

Le previsioni in Power BI si basano su un algoritmo di forecast molto complesso denominato Exponential Smoothing. Per essere precisi, esistono due modelli di Exponential Smoothing: Seasonal Data (ETS AAA), che tiene conto di trend e stagionalità additive, e Non-Seasonal Data (ETS AAN), modello più semplice che considera solo i trend. Il modello ETS AAA è anche conosciuto come algoritmo di Holt-Winters. È il modello più complesso dei due perché ricerca i parametri ottimali per minimizzare lo scarto quadratico medio (MSE) nelle previsioni del modello di addestramento, prendendo in considerazione non solo le previsioni più "vicine" ma anche quelle con un orizzonte temporale più lungo, al fine di ridurre la variazione e preservarne la tendenza. Essenzialmente, il modello applica una media ponderata delle osservazioni storiche, dando più peso alle osservazioni recenti rispetto a quelle più remote. In entrambi i casi, Power BI applica automaticamente l'algoritmo migliore a seconda dei dati a disposizione.

Per utilizzare la funzionalità di forecasting integrata in Power BI è necessario configurare opportunamente il grafico contenente la serie temporale di interesse. Al termine della procedura, il software genera una proiezione della serie storica verso periodi futuri, accompagnata da una banda di confidenza che rappresenta l'incertezza associata alle previsioni. La procedura da seguire è la seguente:

1. creare un grafico a linee;
2. inserire una variabile temporale nell'asse X;
3. inserire la misura da analizzare nell'asse Y;
4. accedere al pannello *Analisi*, identificabile tramite l'icona della lente di ingrandimento nella sezione *Visualizzazioni*, come mostrato nella Figura 5.1;
5. selezionare la funzionalità *Forecast* presente nel pannello *Analisi*, come evidenziato nella Figura 5.2;
6. configurare i parametri della previsione, come mostrato nella Figura 5.3; tra questi parametri abbiamo:
 - orizzonte temporale della previsione;
 - stagionalità;
 - intervallo di confidenza.

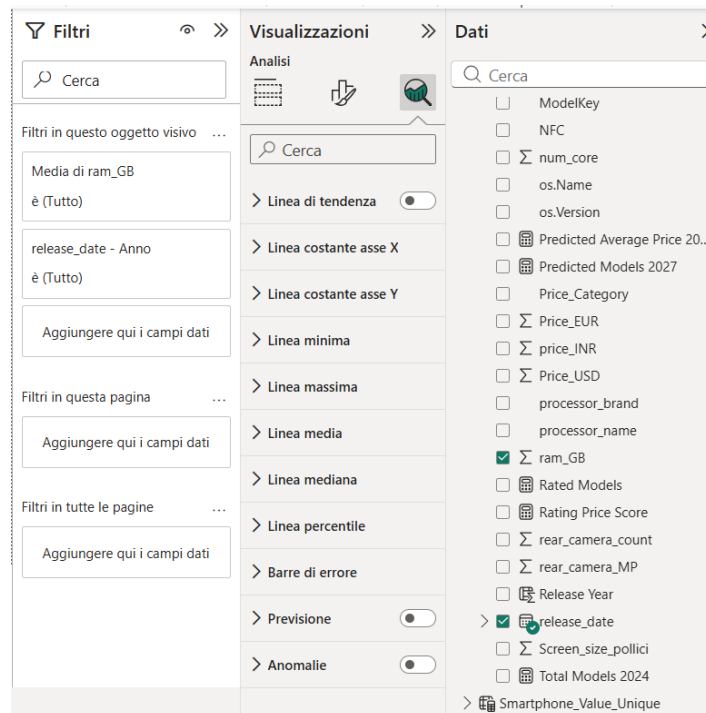


Figura 5.1: Menu della sezione Analisi nelle Visualizzazioni



Figura 5.2: Selezione della funzionalità Previsione nel pannello Analisi

Questa modalità risulta particolarmente utile per realizzare rapidamente analisi esplorative e dashboard interattive direttamente all'interno di Power BI.

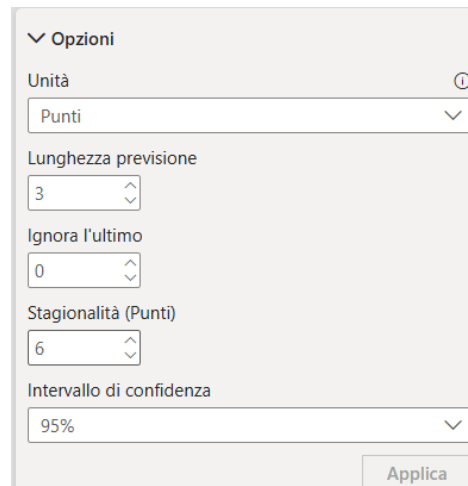


Figura 5.3: Opzioni configurabili nella sezione Previsione

Analisi predittiva avanzata con Python e R

Power BI offre la possibilità di utilizzare librerie e strumenti di R e Python come estensioni aggiuntive agli strumenti già presenti. Per entrambi i linguaggi di programmazione sono disponibili specifici visual che abilitano l’inserimento diretto del codice per creare grafici personalizzati, che possono, poi, essere facilmente incorporati nelle dashboard di Power BI, come mostrato nella Figura 5.4.

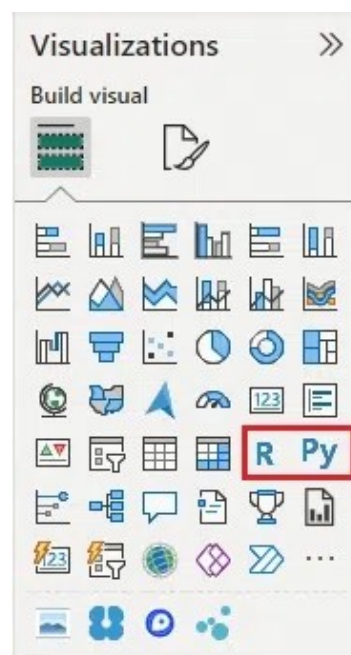


Figura 5.4: R e Python visuals

Quando si seleziona l'icona di uno di questi oggetti visivi nel pannello delle visualizzazioni di Power BI, si genera un nuovo elemento nell'area di lavoro e si apre l'editor dedicato per inserire il codice necessario allo sviluppo del modello previsionale. Per integrare i dati nel proprio script R, è sufficiente trascinare i campi di interesse nella sezione "Valori" del pannello di configurazione della visualizzazione, seguendo lo stesso

procedimento applicabile a qualsiasi altro elemento visivo. È importante notare che soltanto i campi inseriti in questa area saranno accessibili e utilizzabili nello script R.

L'integrazione con Python e R consente, ad esempio, di:

- implementare modelli statistici personalizzati;
- eseguire operazioni avanzate di data preprocessing;
- utilizzare algoritmi di machine learning;
- realizzare visualizzazioni personalizzate;
- sviluppare analisi previsionali più sofisticate.

Tuttavia, per utilizzare R e/o Python all'interno di Power BI, sono necessari alcuni prerequisiti; infatti, bisogna accertarsi di aver correttamente installato R e/o Python nel proprio computer locale, oltre alle librerie necessarie all'analisi. Inoltre, andrà abilitato lo scripting R e/o Python in Power BI Desktop in *Opzioni File > Opzioni e impostazioni > Opzioni > Globale* selezionando Script R o Script Python a seconda di quello che si vuole abilitare.

In particolare, per utilizzare script Python o R in Power BI è generalmente necessario:

1. installare il linguaggio e le librerie necessarie;
2. configurare Python o R nelle impostazioni di Power BI;
3. inserire una visualizzazione Python o R all'interno del report;
4. scrivere ed eseguire lo script desiderato.

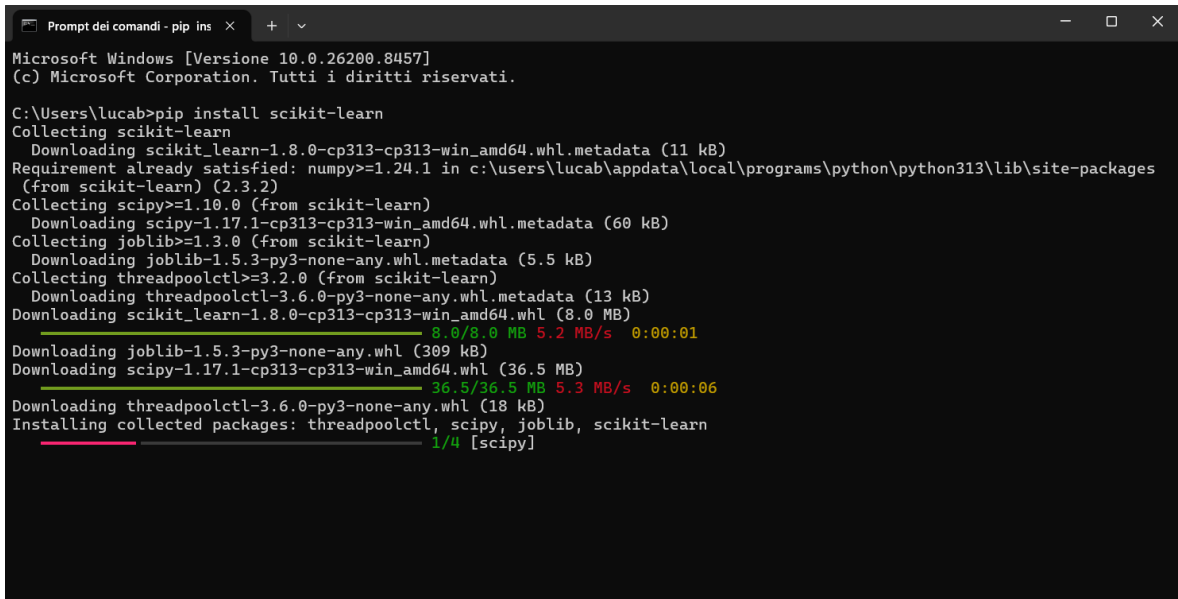
Nel caso di Python, alcune delle librerie maggiormente utilizzate per analisi predittive includono:

- *scikit-learn*;
- *statsmodels*;
- *Prophet*;
- *pandas*;
- *matplotlib*.

Ciascuna libreria risponde a esigenze differenti, dalla preparazione dei dati fino alla costruzione di modelli predittivi e alla visualizzazione dei risultati. In particolare, per la realizzazione di alcuni grafici successivi utilizzati nella Dashboard 6, è stato utilizzato uno script Python installando la libreria *scikit-learn* dal terminale Windows, tramite il comando *pip install scikit-learn*. Il processo viene mostrato nella Figura 5.5, che illustra il terminale Windows con il processo di installazione della libreria *pip install scikit-learn*.

L'utilizzo di Python e R offre una maggiore libertà nella costruzione dei modelli previsionali, permettendo di controllare direttamente:

- la logica del modello;
- le variabili utilizzate;
- gli algoritmi di previsione;
- la modalità di visualizzazione dei risultati.



```
Microsoft Windows [Versione 10.0.26200.8457]
(c) Microsoft Corporation. Tutti i diritti riservati.

C:\Users\lucab>pip install scikit-learn
Collecting scikit-learn
  Downloading scikit_learn-1.8.0-cp313-cp313-win_amd64.whl.metadata (11 kB)
Requirement already satisfied: numpy>=1.24.1 in c:\users\lucab\appdata\local\programs\python\python313\lib\site-packages
(from scikit-learn) (2.3.2)
Collecting scipy>=1.10.0 (from scikit-learn)
  Downloading scipy-1.17.1-cp313-cp313-win_amd64.whl.metadata (60 kB)
Collecting joblib>=1.3.0 (from scikit-learn)
  Downloading joblib-1.5.3-py3-none-any.whl.metadata (5.5 kB)
Collecting threadpoolctl>=3.2.0 (from scikit-learn)
  Downloading threadpoolctl-3.6.0-py3-none-any.whl.metadata (13 kB)
Downloading scikit_learn-1.8.0-cp313-cp313-win_amd64.whl (8.0 MB)
 8.0/8.0 MB 5.2 MB/s 0:00:01
Downloading joblib-1.5.3-py3-none-any.whl (309 kB)
Downloading scipy-1.17.1-cp313-cp313-win_amd64.whl (36.5 MB)
 36.5/36.5 MB 5.3 MB/s 0:00:06
Downloading threadpoolctl-3.6.0-py3-none-any.whl (18 kB)
Installing collected packages: threadpoolctl, scipy, joblib, scikit-learn
 1/4 [scipy]
```

Figura 5.5: Processo di installazione della libreria *scikit-learn* utilizzando il terminale Windows

Tuttavia, le visualizzazioni generate tramite script Python o R presentano, generalmente, un livello di interattività inferiore rispetto alle visualizzazioni native di Power BI, soprattutto per quanto riguarda:

- tooltip dinamici;
- drill-through;
- interazioni avanzate tra grafici.

Nonostante queste limitazioni, l'integrazione con linguaggi esterni rappresenta una soluzione particolarmente flessibile per realizzare dashboard predittive avanzate e analisi personalizzate.

5.3 Introduzione all'analisi predittiva nel contesto del mercato degli smartphone

Per l'analisi predittiva del mercato degli smartphone sono state realizzate le seguenti dashboard:

- *Dashboard 6:* è focalizzata sull'analisi predittiva dell'andamento dei prezzi e del numero di modelli rilasciati dai principali brand di smartphone. L'obiettivo è stimare l'evoluzione futura del mercato fino al 2027 attraverso tecniche di forecasting e regressione lineare implementate in Power BI e Python. In particolare, consente di analizzare la previsione del prezzo medio degli smartphone, la stima del numero di modelli futuri e gli indicatori CAGR relativi a prezzi e quantità di dispositivi rilasciati. Nel complesso, la dashboard consente, quindi, di comprendere i possibili trend evolutivi del mercato degli smartphone e le strategie future dei principali produttori.
- *Dashboard 7:* è dedicata all'analisi predittiva dell'evoluzione delle principali caratteristiche hardware degli smartphone. L'obiettivo è stimare come specifiche tecniche, quali RAM, capacità di archiviazione, batteria, fotocamera posteriore e

dimensione dello schermo, potrebbero evolvere nei prossimi anni. In particolare, consente di analizzare i trend previsionali delle componenti hardware fino al 2027, anche in relazione alle differenti fasce di prezzo dei dispositivi. Nel complesso, la dashboard consente di comprendere l'evoluzione tecnologica prevista del mercato degli smartphone e l'evoluzione delle configurazioni hardware.

5.3.1 Dashboard 6 - Analisi predittiva dell'andamento dei prezzi e del numero di modelli di smartphone

La Dashboard 6 ha come obiettivo principale l'analisi predittiva dell'evoluzione del mercato degli smartphone, con particolare attenzione:

- all'andamento futuro dei prezzi medi di lancio;
- all'evoluzione del numero di modelli rilasciati dai principali brand;
- al tasso di crescita previsto del mercato;
- al confronto tra le strategie produttive delle principali aziende produttrici di smartphone.

La dashboard consente, quindi, non solo di analizzare i dati storici presenti nel dataset, ma anche di stimare l'andamento futuro del mercato mediante tecniche di forecasting implementate attraverso script Python integrati in Power BI. Nella dashboard sono stati considerati i principali brand presenti nel dataset, quali Apple, Samsung, Google, Realme e Xiaomi.

L'intera dashboard è interattiva grazie alla presenza di uno slicer dedicato al brand, che permette di filtrare dinamicamente tutte le visualizzazioni presenti nella pagina.

La Dashboard 6 - Analisi predittiva dell'andamento dei prezzi e del numero di modelli di smartphone è riportata nella Figura 5.6.

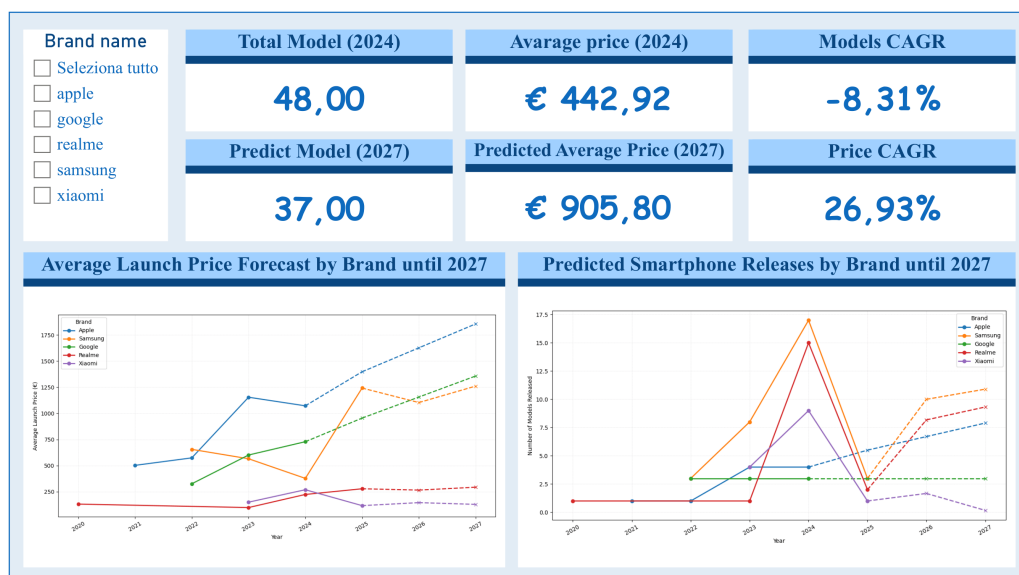


Figura 5.6: Dashboard 6 - Analisi predittiva dell'andamento dei prezzi e del numero di modelli di smartphone

Key Performance Indicator (KPI)

Per sintetizzare rapidamente le principali informazioni della dashboard sono stati implementati sei KPI principali.

- *Total Models (2024)*: rappresenta il numero totale di modelli smartphone distinti rilasciati nel 2024.
- *Predicted Models (2027)*: rappresenta il numero di modelli previsto per il 2027 sulla base del trend storico osservato nel dataset.
- *Average Price (2024)*: rappresenta il prezzo medio degli smartphone presenti nel dataset nel 2024.
- *Predicted Average Price (2027)*: rappresenta il prezzo medio degli smartphone previsto nel 2027.
- *Models CAGR*: rappresenta il tasso annuo composto di crescita del numero di modelli tra il 2024 e il 2027.
- *Price CAGR*: rappresenta il tasso annuo composto di crescita del prezzo medio tra il 2024 e il 2027.

Per la realizzazione dei KPI previsionali sono state implementate due misure DAX utilizzate per stimare il numero di modelli smartphone previsto nel 2027 e il prezzo medio previsto degli smartphone nel 2027.

Le misure DAX utilizzate sono quelle illustrate nel dettaglio nei capitoli precedenti, ovvero:

- *Predicted Models 2027*;
- *Predicted Average Price 2027*.

Grafici utilizzati - Analisi predittiva tramite regressione lineare

I grafici utilizzati per l'analisi predittiva sono i seguenti:

- *Grafico lineare predittivo del prezzo medio*: rappresenta l'andamento storico e la previsione futura del prezzo medio di lancio degli smartphone per ciascun brand considerato. Il grafico consente di confrontare le differenti strategie di posizionamento dei produttori e di osservare l'evoluzione prevista dei prezzi nel mercato degli smartphone. Le linee continue rappresentano i dati storici, mentre le linee tratteggiate rappresentano le previsioni fino al 2027.
- *Grafico lineare predittivo del numero di modelli*: rappresenta il numero storico di modelli di smartphone rilasciati annualmente dai principali brand e la relativa previsione fino al 2027. Questa visualizzazione consente di osservare le differenti strategie produttive dei brand, evidenziando eventuali aumenti o riduzioni nel numero di dispositivi rilasciati nel tempo.

Per realizzare le previsioni relative al prezzo medio degli smartphone e al numero di modelli rilasciati, sono stati utilizzati script Python integrati direttamente all'interno di Power BI. Le previsioni sono state ottenute mediante un modello di regressione lineare implementato tramite la libreria *scikit-learn*. Il modello utilizza:

- l'anno di rilascio come variabile indipendente;
- il prezzo medio (nel grafico dedicato alla previsione del prezzo medio) o il numero di modelli (nel grafico dedicato alla previsione del numero di modelli) come variabile dipendente.

Di seguito è possibile visualizzare i due script Python implementati per la realizzazione dei due grafici predittivi.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from sklearn.linear_model import LinearRegression
5
6 df = dataset.copy()
7
8 df = df.dropna(subset=['brand_name', 'Anno', 'Price_EUR'])
9 df['Year'] = df['Anno']
10 df['brand_name'] = df['brand_name'].str.lower()
11
12 brands = ['apple', 'samsung', 'google', 'realme', 'xiaomi']
13 df = df[df['brand_name'].isin(brands)]
14
15 yearly = (
16     df.groupby(['brand_name', 'Year'])['Price_EUR']
17     .mean()
18     .reset_index()
19 )
20
21 plt.figure(figsize=(12, 6))
22
23 for brand in brands:
24     data = yearly[yearly['brand_name'] == brand].sort_values('Year')
25
26     if len(data) < 2:
27         continue
28
29     X = data[['Year']]
30     y = data['Price_EUR']
31
32     model = LinearRegression()
33     model.fit(X, y)
34
35     last_year = int(data['Year'].max())
36     last_price = float(y.iloc[-1])
37
38     future_years = np.arange(last_year + 1, 2028).reshape(-1, 1)
39     future_preds = model.predict(future_years)
40
41     line, = plt.plot(
42         data['Year'],
43         y,
44         linewidth=2,
45         marker='o',
46         label=brand.capitalize()
47     )
48
49     color = line.get_color()
50
51     plt.plot(
52         future_years.flatten(),
53         future_preds,
```

```

54         linewidth=2,
55         linestyle='--',
56         marker='x',
57         color=color
58     )
59
60     plt.plot(
61         [last_year, future_years[0][0]],
62         [last_price, future_preds[0]],
63         linestyle='--',
64         linewidth=2,
65         color=color
66     )
67
68     plt.xlabel('Year')
69     plt.ylabel('Average Launch Price')
70
71     plt.xticks(np.arange(yearly['Year'].min(), 2028, 1), rotation=30)
72     plt.grid(True, linestyle='--', linewidth=0.5, alpha=0.3)
73
74     plt.legend(title='Brand')
75
76     plt.tight_layout()
77     plt.show()

```

Listato 5.1: Script Python per la realizzazione del grafico *Avarage Launch Price Forecast by Brand until 2027*

```

1  import pandas as pd
2  import numpy as np
3  import matplotlib.pyplot as plt
4  from sklearn.linear_model import LinearRegression
5
6  df = dataset.copy()
7
8  df = df.dropna(subset=['brand_name', 'Anno', 'ModelKey'])
9  df['Year'] = df['Anno']
10 df['brand_name'] = df['brand_name'].str.lower()
11
12 brands = ['apple', 'samsung', 'google', 'realme', 'xiaomi']
13 df = df[df['brand_name'].isin(brands)]
14
15 yearly = (
16     df.groupby(['brand_name', 'Year'])['ModelKey']
17     .nunique()
18     .reset_index(name='Number_of_Models')
19 )
20
21 plt.figure(figsize=(12, 6))
22
23 for brand in brands:
24     data = yearly[yearly['brand_name'] == brand].sort_values('Year')
25
26     if len(data) < 2:
27         continue
28
29     X = data[['Year']]
30     y = data['Number_of_Models']
31
32     model = LinearRegression()
33     model.fit(X, y)
34

```

```

35 last_year = int(data['Year'].max())
36 last_value = float(y.iloc[-1])
37
38 future_years = np.arange(last_year + 1, 2028).reshape(-1, 1)
39 future_preds = model.predict(future_years)
40 future_preds = np.maximum(future_preds, 0)
41
42 line, = plt.plot(
43     data['Year'],
44     y,
45     linewidth=2,
46     marker='o',
47     label=brand.capitalize()
48 )
49
50 color = line.get_color()
51
52 plt.plot(
53     future_years.flatten(),
54     future_preds,
55     linewidth=2,
56     linestyle='--',
57     marker='x',
58     color=color
59 )
60
61 plt.plot(
62     [last_year, future_years[0][0]],
63     [last_value, future_preds[0]],
64     linestyle='--',
65     linewidth=2,
66     color=color
67 )
68 plt.xlabel('Year')
69 plt.ylabel('Number of Models Released')
70
71 plt.xticks(np.arange(yearly['Year'].min(), 2028, 1), rotation=30)
72 plt.grid(True, linestyle='--', linewidth=0.5, alpha=0.3)
73
74 plt.legend(title='Brand')
75
76
77 plt.tight_layout()
78 plt.show()

```

Listato 5.2: Script Python per la realizzazione del grafico *Predicted Smartphone Releases by Brand until 2027*

L'obiettivo della regressione lineare è individuare il trend medio storico dei dati e proiettarlo negli anni successivi fino al 2027.

In particolare:

- le linee continue rappresentano i dati storici presenti nel dataset;
- le linee tratteggiate rappresentano, invece, i valori previsionali stimati dal modello.

La previsione non rappresenta una certezza assoluta sull'andamento futuro del mercato, ma una stima statistica basata sui dati storici disponibili nel dataset.

Regressione lineare e CAGR

La regressione lineare è stata utilizzata sia nelle misure DAX sia negli script Python con finalità differenti. Dal punto di vista matematico, sia le misure DAX sia gli script Python utilizzano una regressione lineare semplice basata sul metodo dei minimi quadrati ordinari (Ordinary Least Squares, OLS).

Nelle misure DAX, la regressione è stata impiegata per ottenere un valore puntuale da visualizzare nei KPI della dashboard. In particolare, le misure *Predicted Models 2027* e *Predicted Average Price 2027* stimano, rispettivamente, il numero di modelli e il prezzo medio previsti per il 2027.

Negli script Python, invece, la regressione lineare è stata applicata separatamente per ciascun brand considerato. In questo caso, il modello non produce soltanto una singola previsione, ma genera una serie di valori previsionali per gli anni successivi all'ultima osservazione disponibile, fino al 2027. Tali valori sono stati utilizzati per costruire i grafici predittivi, nei quali le linee continue rappresentano i dati storici e le linee tratteggiate rappresentano le stime future.

Pertanto, la differenza principale non riguarda il modello matematico utilizzato, ma la finalità dell'applicazione: le misure DAX producono valori sintetici per i KPI, mentre gli script Python consentono di visualizzare graficamente l'evoluzione prevista nel tempo e il confronto tra i diversi brand.

L'equazione utilizzata per la realizzazione della regressione lineare è la seguente:

$$\hat{y} = \beta_0 + \beta_1 x$$

dove:

- $\hat{y}$ rappresenta il valore stimato dal modello;
- x rappresenta l'anno di rilascio;
- β_0 rappresenta l'intercetta della retta;
- β_1 rappresenta il coefficiente angolare della retta.

I coefficienti della retta vengono determinati mediante il metodo dei minimi quadrati ordinari (*Ordinary Least Squares*), che minimizza la somma dei quadrati degli errori tra i valori osservati e quelli stimati.

Il coefficiente angolare è calcolato come:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

dove:

- x_i rappresenta il valore della variabile indipendente per l'osservazione i ;
- y_i rappresenta il valore osservato della variabile dipendente;
- $\bar{x}$ rappresenta la media dei valori della variabile indipendente;
- $\bar{y}$ rappresenta la media dei valori della variabile dipendente;
- n rappresenta il numero totale di osservazioni considerate.

L'intercetta è, invece, determinata dalla seguente relazione:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Una volta calcolati i parametri della retta, il valore previsto per il 2027 viene ottenuto sostituendo $x = 2027$ nell'equazione di regressione:

$$\hat{y}_{2027} = \beta_0 + \beta_1 \cdot 2027$$

Per analizzare, invece, il tasso di crescita previsto del mercato sono state implementate le misure:

- *Models CAGR*;
- *Price CAGR*.

Il CAGR (*Compound Annual Growth Rate*) rappresenta il tasso annuo composto di crescita e consente di misurare la crescita media annuale di un determinato valore in un intervallo temporale.

La formula matematica del CAGR è la seguente:

$$CAGR = \left(\frac{\text{Valore Finale}}{\text{Valore Iniziale}} \right)^{\frac{1}{n}} - 1$$

dove:

- il valore finale rappresenta il valore previsto nel 2027;
- il valore iniziale rappresenta il valore osservato nel 2024;
- n rappresenta il numero di anni considerati nell'analisi.

Il CAGR consente, quindi, di sintetizzare in un unico indicatore la crescita media annuale prevista per il numero di modelli di smartphone e per il prezzo medio dei dispositivi.

Conclusioni sull'analisi della Dashboard 6

Dall'analisi effettuata emerge come il mercato degli smartphone sia caratterizzato dalle seguenti tendenze:

- una crescita significativa del prezzo medio previsto;
- una possibile riduzione del numero totale di modelli rilasciati;
- una progressiva concentrazione verso dispositivi di fascia premium.

L'aumento previsto del prezzo medio risulta coerente con le dinamiche osservate nel mercato degli smartphone negli ultimi anni, caratterizzato da una tendenza crescente verso i dispositivi premium e da un incremento dei costi medi dei flagship.

Tale crescita dei prezzi può essere, inoltre, associata alla continua evoluzione delle caratteristiche hardware degli smartphone. Negli ultimi anni i dispositivi hanno, infatti, registrato un aumento di:

- quantità di memoria RAM;
- capacità di storage;
- qualità dei comparti fotografici;

- prestazioni dei processori;
- tecnologie legate a display e batterie.

Tale fenomeno può essere ricondotto anche alla crescente integrazione di funzionalità basate sull'Intelligenza Artificiale, oggi sempre più presenti nei principali smartphone di fascia alta.

I sistemi basati sull'Intelligenza Artificiale vengono impiegati principalmente per:

- un'elaborazione fotografica avanzata;
- l'utilizzo di assistenti intelligenti;
- la generazione automatica di contenuti;
- le traduzioni in tempo reale;
- l'ottimizzazione delle performance e dei consumi.

Di conseguenza, tale evoluzione sta contribuendo ad aumentare sia la complessità tecnologica dei dispositivi sia il loro posizionamento verso fasce di prezzo più elevate.

La riduzione del numero di modelli previsti non deve essere interpretata come una certezza assoluta sull'evoluzione futura del mercato, ma come una proiezione statistica derivante dai dati storici presenti nel dataset.

Dal punto di vista della Business Intelligence, la dashboard realizzata rappresenta uno strumento utile per:

- analizzare tendenze future del mercato degli smartphone;
- confrontare le strategie dei differenti brand;
- comprendere l'evoluzione prevista dei prezzi;
- supportare analisi strategiche e previsionali.

L'utilizzo combinato di KPI previsionali, regressione lineare, grafici Python integrati e indicatori CAGR ha consentito la realizzazione di una dashboard predittiva interattiva e facilmente interpretabile, in grado di fornire una visione evolutiva del mercato degli smartphone fino al 2027.

5.3.2 Dashboard 7 - Analisi predittiva dell'evoluzione hardware degli smartphone

La Dashboard 7 è dedicata all'analisi predittiva dell'evoluzione delle principali caratteristiche hardware degli smartphone, con l'obiettivo di comprendere come le specifiche tecniche dei dispositivi possano evolvere nei prossimi anni.

In particolare, la dashboard consente di analizzare:

- l'evoluzione della memoria RAM media;
- la crescita della capacità media di archiviazione;
- l'andamento della capacità media delle batterie;
- l'evoluzione della risoluzione media delle fotocamere posteriori;
- la variazione della dimensione media dei display.

La dashboard permette, quindi, di osservare come il mercato degli smartphone sia orientato verso dispositivi caratterizzati da specifiche hardware sempre più avanzate, evidenziando i principali trend tecnologici previsti fino al 2027.

Nella Figura 5.7 è mostrata la Dashboard 7 - Analisi predittiva dell'evoluzione hardware degli smartphone.

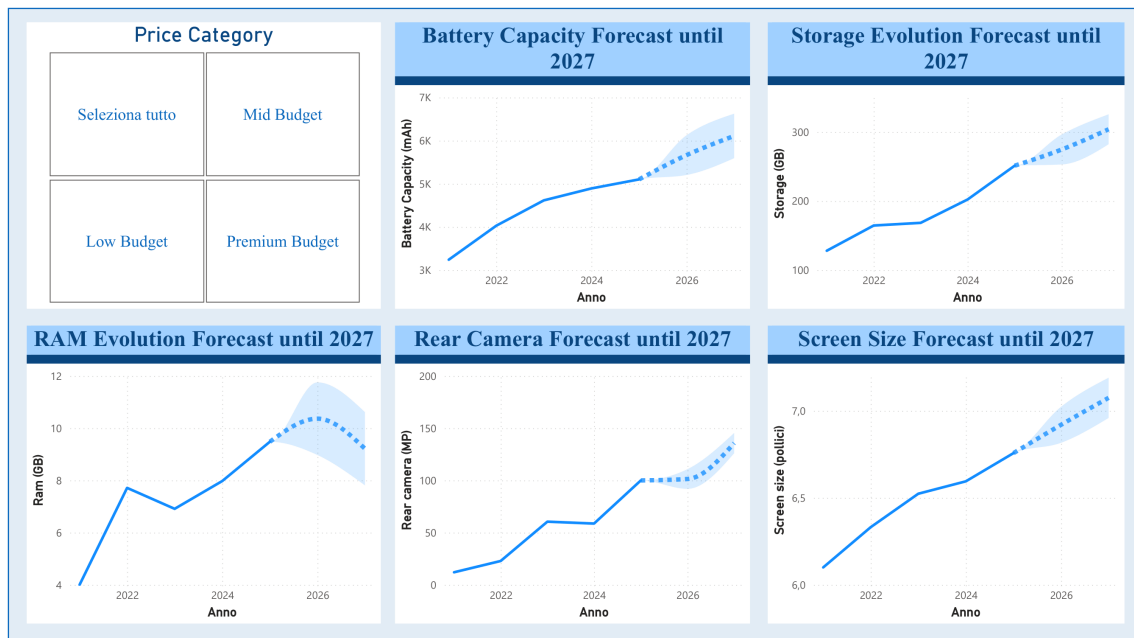


Figura 5.7: Dashboard 7 - Analisi predittiva dell'evoluzione hardware degli smartphone

Grafici utilizzati

Per la realizzazione della dashboard sono stati utilizzati grafici a linee con funzionalità di *forecasting* integrata in Power BI. In particolare:

- i grafici mostrano i dati storici relativi all'evoluzione delle differenti componenti hardware;
- la parte tratteggiata rappresenta la previsione stimata fino al 2027;
- l'area ombreggiata rappresenta l'intervallo di confidenza generato automaticamente dal modello previsionale di Power BI.

La dashboard include, inoltre, un filtro relativo alla categoria di prezzo (*Price Category*), che consente di confrontare l'evoluzione tecnologica tra smartphone appartenenti a differenti fasce di mercato, ovvero:

- *Low Budget*;
- *Mid Budget*;
- *Premium Budget*.

Conclusioni sull'analisi della Dashboard 7

L'analisi della dashboard evidenzia un trend generale di crescita delle principali caratteristiche hardware degli smartphone. In particolare:

- la memoria RAM media mostra un incremento progressivo nel tempo;
- la capacità di storage evidenzia una crescita significativa, coerente con la crescente richiesta di spazio per applicazioni, contenuti multimediali e funzionalità AI;
- la capacità delle batterie tende ad aumentare gradualmente per supportare dispositivi sempre più performanti;
- le fotocamere posteriori mostrano un incremento della risoluzione media, confermando l'importanza crescente del comparto fotografico nel mercato degli smartphone;
- la dimensione media dei display evidenzia una tendenza verso schermi sempre più ampi.

I risultati sono coerenti con l'evoluzione osservata nel mercato degli smartphone negli ultimi anni, caratterizzato da una continua crescita delle prestazioni hardware e da una progressiva introduzione di funzionalità avanzate basate sull'Intelligenza Artificiale.

Nel complesso, la dashboard consente di ottenere una visione evolutiva delle principali tecnologie hardware nel settore degli smartphone, supportando analisi strategiche e attività di pianificazione basate su scenari previsionali.

Il presente capitolo è dedicato a una valutazione critica del percorso metodologico e analitico sviluppato nel corso della tesi, con l'obiettivo di evidenziarne i principali punti di forza e le criticità riscontrate. Dopo aver affrontato le diverse fasi del processo di Data Analytics, comprendenti l'estrazione, la trasformazione e il caricamento dei dati (ETL), la realizzazione di dashboard interattive tramite Power BI e lo svolgimento di analisi descrittive e predittive, risulta opportuno dedicare questo capitolo per esaminare il contributo di tali attività alla comprensione del fenomeno analizzato, ma anche i limiti e le difficoltà che hanno condizionato l'efficacia e la robustezza dei risultati.

6.1 Punti di Forza

6.1.1 Approccio completo di Data Analytics

Uno dei principali punti di forza del progetto consiste nell'aver sviluppato un processo di Data Analytics completo, affrontando in maniera strutturata tutte le fasi del ciclo relativo all'analisi dei dati. In particolare, il lavoro ha previsto l'acquisizione dei dati, la loro preparazione mediante il processo di ETL, l'integrazione dei dataset, la modellazione dei dati e la realizzazione di dashboard dedicate alle analisi descrittive e predittive. L'adozione di questo approccio ha consentito di garantire continuità e coerenza tra le diverse fasi del progetto, riducendo il rischio di propagare errori nelle fasi successive e migliorando l'affidabilità dei risultati ottenuti. Di conseguenza, il lavoro non si è limitato alla sola visualizzazione dei dati, ma ha coperto l'intero processo di trasformazione del dato grezzo in informazioni utili a supportare l'analisi del mercato degli smartphone.

6.1.2 Processo di ETL

Un altro punto di forza, strettamente collegato al precedente, è aver dedicato particolare attenzione alla preparazione dei dati applicando un processo di *extract, transform e load*, generando, a partire dai dati disponibili nella fase di estrazione, un dataset pulito, coerente e affidabile, riducendo le incongruenze nei formati e nelle unità di misura, pronti per la realizzazione di dashboard basate su dati affidabili. Il processo di ETL ha permesso, quindi, di migliorare la qualità del dataset attraverso la normalizzazione, la conversione delle valute, la standardizzazione dei nomi, la gestione delle colonne e la creazione di nuove variabili, aumentando la coerenza delle analisi.

6.1.3 Utilizzo di Power BI

Power BI ha rappresentato uno strumento fondamentale per lo sviluppo dell'intero progetto, offrendo un ambiente integrato per la modellazione dei dati, la costruzione delle dashboard e lo sviluppo delle analisi descrittive e predittive. Uno degli strumenti messi a disposizione da Power BI è rappresentato dalle misure DAX, che consentono di ottenere indicatori sintetici e confrontabili. Power BI garantisce l'integrazione con gli ambienti R e Python e la funzionalità di forecasting integrata per la realizzazione di analisi predittive. La possibilità di utilizzare grafici combinati di diverse tipologie per la realizzazione di dashboard interattive ha consentito la realizzazione di dashboard che forniscono una comunicazione chiara ed efficace dei risultati, anche a un pubblico non specializzato. Power BI offre, inoltre, strumenti di esplorazione, quali filtri, slicer e funzionalità di drill-down, migliorando la leggibilità delle informazioni rappresentate e individuando rapidamente pattern e anomalie significative all'interno dei dataset.

6.1.4 Analisi di dashboard descrittive

La realizzazione delle dashboard descrittive ha rappresentato uno dei principali punti di forza del progetto, poiché ha consentito di trasformare il dataset elaborato in uno strumento di analisi capace di fornire una visione completa e immediatamente interpretabile del mercato degli smartphone. Attraverso la realizzazione delle prime dashboard, è stato possibile analizzare il mercato attuale da prospettive differenti, evidenziando sia aspetti legati all'offerta disponibile sia approfondimenti relativi alle caratteristiche tecniche dei dispositivi, alla loro distribuzione nelle diverse fasce di prezzo, alla diffusione dei sistemi operativi e al rapporto tra prezzo e soddisfazione degli utenti. Tali analisi hanno permesso di confrontare produttori, smartphone di diverse categorie di prezzo e diversi aspetti delle specifiche hardware, facilitando l'individuazione di pattern, correlazioni e differenze che sarebbero risultate meno evidenti attraverso la semplice consultazione dei dati grezzi.

6.1.5 Analisi di dashboard predittive

L'integrazione di dashboard dedicate all'analisi predittiva costituisce un ulteriore punto di forza del progetto, in quanto, ha consentito di estendere l'analisi oltre la semplice descrizione dei dati storici, introducendo un'analisi con l'obiettivo di evidenziare l'evoluzione futura del mercato. Questo aspetto è stato possibile grazie alla capacità di Power BI di integrare librerie R e Python per sviluppare modelli previsionali, e alla funzione di forecasting integrata in Power BI stesso, che consente di stimare l'andamento dei prezzi di lancio, del numero di modelli prodotti, e l'evoluzione delle principali caratteristiche hardware, come memoria RAM, capacità di archiviazione, batteria, fotocamera e dimensione dello schermo.

6.2 Limiti e vulnerabilità

6.2.1 Qualità dei dati

Una delle prime difficoltà affrontate ha riguardato la qualità e la completezza dei dataset utilizzati. Nonostante le sorgenti di dati fossero pubbliche e affidabili, durante il processo di ETL sono emerse numerose incongruenze che hanno richiesto un'approfondita fase di trasformazione e pulizia dei dati prima di poter procedere con le analisi. In particolare, durante la fase di trasformazione del processo di ETL, è stato necessario intervenire sulla standardizzazione dei formati, sulla conversione dei prezzi nelle diverse valute, sulla normalizzazione delle unità di misura, sulla gestione dei valori mancanti e sulla rimozione di

attributi ridondanti o non significativi. Tali operazioni hanno richiesto un notevole impegno, ma si sono rivelate indispensabili per ottenere un dataset omogeneo, coerente e adatto alla realizzazione delle dashboard. Nonostante il processo di ETL abbia consentito di correggere gran parte delle anomalie riscontrate, eventuali errori residui, informazioni mancanti o imprecisioni presenti nei dataset di partenza potrebbero aver influenzato, almeno in parte, l'affidabilità dei risultati ottenuti.

6.2.2 Integrazione complessa

Sulla base dei due dataset scelti per effettuare l'analisi del mercato degli smartphone, per avere a disposizione maggiori informazioni da utilizzare per confrontare variabili e realizzare analisi più approfondite, è stata effettuata un'operazione di merge tra i due dataset, in particolare un'operazione di *Left Outer Join*, basata sulla chiave identificativa dei modelli. Tuttavia, nel corso dell'operazione di merge, sono state individuate soltanto circa 250 corrispondenze tra le 766 righe presenti nella prima tabella. Questo risultato ha evidenziato che, a causa delle diverse fonti di provenienza dei dataset, non tutti gli smartphone del primo dataset presentano una corrispondenza nel secondo, limitando il numero di analisi effettuabili sulle variabili presenti in entrambe le sorgenti.

6.2.3 Difficoltà di confronto tra sistemi operativi

Nel mercato attuale degli smartphone, la quasi totalità dei produttori presenti nel dataset utilizza Android, a differenza di Apple che utilizza il proprio sistema operativo. Un esempio concreto è nel grafico radar, rappresentato nella Dashboard 4, che consente di confrontare il profilo medio delle principali caratteristiche hardware tra dispositivi Android e iOS. Tale grafico tende a mostrare valori medi più elevati per dispositivi iOS. Tali risultati dipendono dal fatto che, come già evidenziato, i dataset risultano sbilanciati; infatti Android comprende dispositivi appartenenti a tutte le fasce di prezzo, mentre iOS è rappresentato esclusivamente da dispositivi di fascia alta. Questa differenza nella composizione comporta un innalzamento delle medie per iOS e una riduzione di quelle per Android. Pertanto, il confronto diretto tra i due sistemi operativi risulta influenzato dalla distribuzione dei dispositivi nel dataset, piuttosto che da una reale superiorità hardware di una piattaforma rispetto all'altra.

6.2.4 Aggiornamento nel tempo

Un ulteriore limite, specifico del mercato degli smartphone, è legato al fatto che, nell'attuale contesto tecnologico, il mercato degli smartphone evolve molto rapidamente, con il conseguente rischio di ridurre nel tempo l'attualità dei risultati ottenuti, se i dataset non venissero aggiornati periodicamente e integrati con dati più recenti.

6.2.5 Limiti dell'analisi predittiva

L'analisi predittiva basata su modelli di regressione lineare, realizzata mediante script Python integrati in Power BI, o mediante la funzionalità di forecasting integrata, potrebbe rappresentare un limite in riferimento al mercato degli smartphone. Entrambi gli approcci sono basati esclusivamente sui dati storici dei dataset e, pertanto, potrebbero non garantire previsioni pienamente accurate poiché non considerano eventi imprevedibili, cambiamenti tecnologici repentini e non sono in grado di prevedere nuove strategie dei produttori.

La presente tesi ha avuto come obiettivo la progettazione e l'implementazione di una campagna di Data Science finalizzata all'analisi del mercato degli smartphone, integrando le attività di acquisizione, trasformazione, modellazione e visualizzazione dei dati attraverso la piattaforma Microsoft Power BI.

Partendo da un'analisi teorica e approfondita dei principali concetti relativi alla Data Science, ai Big Data e alla Business Intelligence, sono stati selezionati due dataset relativi alle caratteristiche tecniche degli smartphone presenti sul mercato, con lo scopo di estrarre informazioni utili a supporto delle decisioni aziendali e analizzare le preferenze dei consumatori.

È stato applicato ai due dataset un processo di ETL con l'obiettivo di effettuare operazioni di pulizia dei dati al fine di ottenere informazioni immediatamente utilizzabili. Successivamente, mediante un'operazione di merge, i due dataset sono stati integrati in un unico modello di dati con l'obiettivo di disporre di maggiori informazioni ed eseguire analisi più approfondite per individuare nuovi pattern e relazioni tra le variabili.

L'utilizzo delle funzionalità di Power BI, delle misure DAX e di strumenti di analisi predittiva ha consentito la realizzazione di dashboard interattive capaci di rappresentare sia lo stato attuale del mercato sia i possibili scenari futuri.

Le dashboard sviluppate hanno permesso di evidenziare le principali caratteristiche del mercato degli smartphone, analizzando aspetti quali la distribuzione dei prezzi, le specifiche hardware, la diffusione dei sistemi operativi, il rapporto qualità-prezzo e il livello di soddisfazione degli utenti.

L'esperienza maturata nel corso dello sviluppo del progetto ha permesso di comprendere come la qualità dei dati sia fondamentale per ottenere analisi affidabili e come una rappresentazione chiara e corretta delle dashboard sia indispensabile per consentire una visualizzazione efficace dei dati, favorendo un'immediata comprensione delle analisi svolte e dei risultati ottenuti.

Questa esperienza ha permesso di acquisire competenze nell'utilizzo di Power BI per la gestione e l'analisi di dataset complessi attraverso dashboard e visualizzazioni interattive. Power BI si è dimostrato uno strumento completo e versatile, capace di integrare le attività di preparazione dei dati, modellazione, visualizzazione e analisi avanzata all'interno di un unico ambiente di lavoro.

Dal punto di vista professionale, il lavoro svolto ha consentito di consolidare e approfondire le competenze acquisite nell'ambito della Data Analytics, dei Big Data, della Business Intelligence, dei processi ETL, della modellazione dei dati e dello sviluppo di dashboard interattive.

Il lavoro svolto ha evidenziato come, in un mondo che richiede velocità e immediatezza, ma, allo stesso tempo, affidabilità e accuratezza, l'utilizzo di dashboard capaci di fornire una sintesi visiva efficace rappresenti un importante vantaggio competitivo in ambito professionale, poiché consente di cogliere rapidamente i principali pattern e gli aspetti fondamentali del mercato analizzato.

Per quanto riguarda i possibili sviluppi futuri, l'analisi del mercato degli smartphone, alla luce delle vulnerabilità discusse nel Capitolo 6, potrebbe essere migliorata mediante l'integrazione di dataset costantemente aggiornati, così da consentire analisi più aderenti all'evoluzione del mercato. Potrebbero essere introdotti modelli di Machine Learning più avanzati per migliorare l'accuratezza delle previsioni, affiancando o sostituendo le tecniche di forecasting utilizzate. Inoltre, l'analisi potrebbe essere estesa mediante l'integrazione di un numero maggiore di dataset, al fine di aumentare il numero di corrispondenze tra i dati durante il processo di integrazione ed analizzare, quindi, non solo un piccolo gruppo di brand per le analisi predittive, ma tutto il mercato generale. In conclusione, ulteriori sviluppi potrebbero riguardare l'integrazione di attributi aggiuntivi provenienti da dataset differenti, quali dati di vendita, recensioni degli utenti, dati provenienti dai social media o indicatori economici, allo scopo di ottenere una visione ancora più completa delle dinamiche del mercato e supportare processi decisionali sempre più efficaci.

- ATZENI, P., CERI, S., FRATERNALI, P., PARABOSCHI, S. e TORLONE, R. (2009), *Basi di dati*, McGraw-Hill, 4 ed.
- FERRARI, A. e RUSSO, M. (2016), *Introducing Microsoft Power BI*, Microsoft Press.
- FEW, S. (2013), *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring*, Analytics Press, 2 ed.
- FRICKÉ, M. (2009), «The Knowledge Pyramid: A Critique of the DIKW Hierarchy», *Journal of Information Science*, vol. 35 (2), p. 131–142.
- IANDIORIO, E. (2019), *Big data. Cosa sono, come analizzarli e utilizzarli per fare marketing*, Dario Flaccovio Editore.
- KIMBALL, R. e CASERTA, J. (2004), *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, Wiley.
- MAURO, A. D. (2019), *Big Data Analytics: Analizzare e interpretare dati con il machine learning*, Apogeo Editore.
- PROVOST, F. e FAWCETT, T. (2013), *Data Science for Business*, O'Reilly Media.
- RUNKLER, T. A. (2025), *Data Analytics: Models and Algorithms for Intelligent Data Analysis*, Springer, 4 ed.
- SALVAGGIO, A. (2023), *Business Intelligence con Microsoft Power BI. Guida completa per l'analisi e la visualizzazione dei dati*, Edizioni LSWR.
- SHARDA, R., DELEN, D. e TURBAN, E. (2024), *Business Intelligence, Analytics, Data Science, and AI: A Managerial Perspective*, Pearson, 5 ed.
- SIMITSIS, A., VASSILIADIS, P. e SELLIS, T. (2005), «Optimizing ETL Processes in Data Warehouses», in «Proceedings of the 21st International Conference on Data Engineering Workshops (ICDEW)», p. 564–575, IEEE.
- SIMITSIS, A., VASSILIADIS, P. e SKIADOPOULOS, S. (2002), «Conceptual Modeling for ETL Processes», in «Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP», p. 14–21.
- WEBB, C. (2014), *Power Query for Power BI and Excel*, Apress.

- **Amazon** – <https://aws.amazon.com/it/what-is/etl/>
- **Data Storytelling** – <https://data-storytelling.it/approfondimenti/com-e-ottenere-una-migliore-data-visualization-grazie-alle-leggi-della-gestalt/>
- **Data Storytelling** – <https://data-storytelling.it/approfondimenti/percezione-visiva-e-memoria-perche-le-immagini-sono-piu-facili-da-ricordare-delle-parole/>
- **Dev4Side** – <https://www.dev4side.com/blog/power-bi-app/>
- **Dev4Side** – <https://www.dev4side.com/blog/power-bi-gateway/>
- **IBM** – <https://www.ibm.com/it-it/think/topics/big-data-analytics>
- **IBM** – <https://www.ibm.com/it-it/think/topics/data-science>
- **IBM** – <https://www.ibm.com/it-it/think/topics/elt>
- **IBM** – <https://www.ibm.com/it-it/think/topics/predictive-analytics>
- **InformIT** – <https://www.informit.com/articles/article.aspx?p=2473128&seqNum=11>
- **InGage** – <https://ingage.it/power-bi-cose-come-funziona-analisi-aziendale/>
- **Kaggle** – <https://www.kaggle.com/>
- **Kaggle - Dataset MySmartPrice** – <https://www.kaggle.com/datasets/vashisthrahul13/mobile-price-and-specs-2025>
- **Kaggle - Dataset Smartprix** – <https://www.kaggle.com/datasets/showmik121/smartphones-dataset-2026-1000-devices>
- **Microsoft** – <https://learn.microsoft.com/it-it/power-bi/>
- **Microsoft** – <https://learn.microsoft.com/it-it/power-bi/transform-model/desktop-query-overview>

- Progesa – <https://www.progesa.com/it-it/microsoft-power-bi-vantaggi-limiti-e-confronto-con-qlik-sense.aspx>
- SAP – <https://www.sap.com/italy/resources/what-is-data-analytics>
- Sas – https://www.sas.com/it_it/insights/analytics/predictive-analytics.html
- Talend – <https://www.talend.com/it/resources/what-is-etl/>
- UniverseIT – <https://universeit.blog/piramide-dikw/>
- Visualitics – <https://visualitics.it/analisi-predittiva-power-bi/>
- Wikipedia – https://it.wikipedia.org/wiki/Big_data
- Wikipedia – https://it.wikipedia.org/wiki/Scienza_dei_dati
- ZeroUno – <https://www.zerounoweb.it/big-data/5-v-dei-big-data-cosa-sono-quale-ruolo-rivestono/>
- ZeroUno – <https://www.zerounoweb.it/big-data/analytics-big-data-cosa-sono-e-come-sono-usati/>
- ZeroUno – <https://www.zerounoweb.it/big-data/data-science/>
- ZeroUno – <https://www.zerounoweb.it/big-data/la-piramide-dikw-cosa-sono-e-cosa-rappresenta/>

Ringraziamenti

Ci tengo a ringraziare il Prof. Domenico Ursino, che con la sua disponibilità e la sua passione mi ha accompagnato nello sviluppo di questo lavoro, offrendomi un supporto prezioso e costante durante tutto il percorso di tesi. Un sentito ringraziamento va anche al Dott. Michele Marchetti, il cui aiuto nella fase realizzativa è stato fondamentale per la buona riuscita del progetto.

Un ringraziamento speciale va a Caterina, per essermi stata accanto con pazienza e supporto nei momenti più difficili di questo percorso.

Ringrazio i miei genitori, per aver sempre creduto in me, sostenendo ogni mia scelta e dandomi gli strumenti per raggiungere questo traguardo.

Ringrazio mio fratello, per essermi sempre stato accanto con la sua presenza, il suo incoraggiamento e il suo costante sostegno.

Infine, rivolgo un pensiero speciale a mia nonna, per la sua vicinanza, la sua premura e il sostegno che non mi ha mai fatto mancare.