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 analytics relativa ai decessi per overdose in uno stato degli Stati Uniti

Design and implementation of a data analytics campaign related to overdose death in a USA state

Relatore

Prof. Domenico Ursino

Correlatore

Dott. Michele Marchetti

Candidata

Gloria Santilli

ANNO ACCADEMICO 2025-2026

L'importante è non smettere mai di farsi domande. La curiosità ha la sua ragione di esistere.

Albert Einstein

Sommario

La crisi degli oppioidi rappresenta una delle emergenze sanitarie più gravi affrontate dagli Stati Uniti negli ultimi tre decenni, con oltre 500.000 vite perdute dal 1999 a oggi. Comprendere le dinamiche di questa epidemia richiede un approccio analitico rigoroso e multidimensionale, capace di trasformare dati complessi in informazioni chiare e utilizzabili per le politiche di intervento. Questa tesi propone un'analisi approfondita della crisi attraverso la Business Intelligence, utilizzando due dataset complementari: uno a livello federale (US Opioid Overdose Deaths, 1999-2014) e uno granulare nel Connecticut (Accidental Drug Related Deaths, 2012-2021). Il lavoro comprende una fase di Data Cleaning e ETL, seguita dallo sviluppo di sette dashboard interattive in Microsoft Power BI, supportate da misure DAX e colonne calcolate. L'analisi esplora quattro dimensioni chiave: la correlazione storica tra prescrizioni mediche e mortalità, il profilo sociodemografico delle vittime, l'architettura tossicologica del policonsumo e i pattern territoriali e stagionali dell'epidemia. Il presente lavoro dimostra come la visualizzazione interattiva dei dati possa trasformare informazioni complesse in narrazioni visive efficaci, supportando decisioni consapevoli nel campo della sanità pubblica, della prevenzione e della pianificazione degli interventi di emergenza.

Keyword: Data Analytics, Business Intelligence, Oppioidi, Crisi Sanitaria, Power BI, DAX, Fentanyl, Dashboard, Policonsumo

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

Negli ultimi tre decenni, gli Stati Uniti hanno affrontato una delle più gravi crisi sanitarie della loro storia moderna: l'epidemia degli oppioidi. Una catastrofe che, partendo dalla diffusione sistematica di analgesici farmaceutici prescritti con leggerezza negli anni Novanta, si è progressivamente trasformata in un'emergenza senza precedenti, alimentata dall'ingresso di sostanze sintetiche sempre più letali, primo fra tutte il Fentanyl. I numeri raccontano una storia di distruzione; infatti, oltre 500.000 vite perdute dal 1999 a oggi, un tasso di mortalità che continua a crescere, famiglie disintegrate, comunità intere devastate. Eppure, dietro questa tragedia numerica si celano milioni di storie individuali: persone che hanno iniziato con una prescrizione legittima, altre che hanno cercato sollievo in strade buie, adolescenti uccisi da pillole contraffatte che non sapevano contenessero veleno.

Ogni numero, in realtà, rappresenta una persona. E ogni analisi dei dati, se condotta con rigore e sensibilità, deve avere lo scopo di trasformare quei numeri in comprensione e, infine, in azione.

Il paradosso contemporaneo è che non mancano i dati. Le autorità sanitarie, le agenzie federali, i laboratori tossicologici raccolgono informazioni dettagliate su ogni aspetto della crisi, ad esempio il numero dei decessi per stato, quali sostanze sono coinvolte, quale età hanno le vittime, dove muoiono. Eppure, per lungo tempo, questi dati sono rimasti confinati in database pubblici, difficili da esplorare e ancora più difficili da comunicare in modo che possano effettivamente guidare le decisioni di policy e i programmi di prevenzione.

Come affermava il matematico francese Henri Poincaré, "la scienza è fatta di dati come una casa è fatta di pietre. Ma un ammasso di dati non è scienza più di quanto un mucchio di pietre sia una vera casa". I dati grezzi, senza trasformazione, rimangono muti.

Proprio da questa convinzione nasce il presente lavoro. L'obiettivo è doppio; infatti, da un lato serve per fornire una comprensione profonda multidimensionale della crisi degli oppioidi attraverso l'analisi rigorosa di due dataset complementari (uno a livello federale, uno granulare nel Connecticut); dall'altro dimostra come la Business Intelligence, con l'aiuto di tool quali Microsoft Power BI, possa trasformare masse di informazioni eterogenee in visualizzazioni interattive, chiare e immediatamente utilizzabili per il supporto decisionale.

Il Connecticut rappresenta un caso di studio particolarmente significativo. Pur essendo uno stato di medie dimensioni, ha registrato un'incidenza della crisi degli oppioidi tra le più alte della nazione, con una crescita esponenziale dei decessi nel decennio 2012-2021. Questa combinazione di gravità del fenomeno e disponibilità di dati dettagliati a livello individuale lo rende ideale per un'analisi che miri a illuminare sia le dinamiche nazionali che le specificità locali.

Nel corso di questa tesi, seguiremo un percorso analitico progressivo, strutturato secondo una logica Top-Down che parte dalla visione macroscopica nazionale per discendere fino alla realtà microscopica dei singoli decessi. Questo viaggio sarà guidato da strumenti che trasformano i dati in narrazione visiva, consentendo al lettore di comprendere non solo "cosa è successo", ma anche "perché è accaduto" e "dove le vulnerabilità sono più acute".

La presente tesi è composta da sette capitoli strutturati come di seguito specificato:

- Nel Capitolo 1 introdurremo il concetto di dato nell'era digitale, ripercorrendone l'evoluzione dal XIX secolo fino all'emergere dei Big Data. Definiremo il ruolo della Data Analytics come strumento di trasformazione dell'informazione in conoscenza strategica, illustrando il ciclo di vita della Big Data Analytics e l'evoluzione verso la Data Science.

- Nel Capitolo 2 presenteremo Microsoft Power BI, analizzando il suo posizionamento globale nel mercato della Business Intelligence, la sua architettura modulare, il flusso di lavoro tipico (ETL, modellazione, visualizzazione), e i linguaggi di programmazione integrati (DAX, M, Python, R) che ne garantiscono la flessibilità.

- Nel Capitolo 3 descriveremo i due dataset di partenza, introdurremo il contesto biologico e storico degli oppioidi, mapperemo la qualità dei dati grezzi e documenteremo tutte le operazioni di pulizia e trasformazione (Data Cleaning ed ETL) condotte in Power Query.

- Nel Capitolo 4 illustreremo le misure DAX e le colonne calcolate sviluppate per supportare le visualizzazioni, quindi presenteremo le prime tre dashboard, pensate come strumenti di esplorazione descrittiva; esse riguarderanno molteplici aspetti, dalla panoramica federale sulla correlazione tra prescrizioni e mortalità, al focus territoriale sul Connecticut, fino al profilo sociodemografico delle vittime.

- Nel Capitolo 5 descriveremo le quattro dashboard avanzate, caratterizzate da oggetti visivi più complessi. Tali dashboard approfondiscono la mobilità epidemiologica, l'architettura del policonsumo, la stagionalità dei decessi e l'impatto specifico del Fentanyl.

- Nel Capitolo 6 analizzeremo criticamente i risultati principali emersi dall'analisi lungo quattro dimensioni complementari: storica, demografica, tossicologica, territoriale e temporale. Il capitolo ripercorre l'architettura del progetto analitico e propone un'interpretazione critica di ciò che i dati rivelano riguardo le cause strutturali della crisi e le implicazioni per le politiche di intervento.

- Nel Capitolo 7, infine, esporremo le conclusioni, sintetizzando gli insegnamenti principali e discutendone le implicazioni per le politiche sanitarie e cliniche. Il capitolo affronta, inoltre, i limiti dello studio e le prospettive future di ricerca, collocando i risultati nel più ampio contesto scientifico, epidemiologico e di policy della crisi americana degli oppioidi.

In ultima analisi, questo lavoro vuole essere uno strumento. Uno strumento per chi lavora nella sanità pubblica, per chi gestisce i soccorsi, per chi disegna le politiche di prevenzione. Perché i dati, quando trasformati in conoscenza, possono salvare vite.

CAPITOLO 1

Introduzione alla Data Analytics

In questo capitolo viene introdotto il concetto di dato nell'era digitale, ripercorrendone l'evoluzione storica e le principali tipologie, fino al modello della piramide DIKW, che descrive il percorso dal dato grezzo alla conoscenza strategica. Si affronta, poi, il fenomeno dei Big Data, illustrando i limiti delle architetture tradizionali e il paradigma delle 5V — Volume, Velocità, Varietà, Veracità e Valore — come chiave per comprendere le sfide e le opportunità nella gestione di grandi volumi informativi. Successivamente, si approfondisce la Data Analytics come strumento essenziale per trasformare i dati in decisioni concrete, presentando i quattro livelli di analisi: descrittiva, diagnostica, predittiva e prescrittiva. Il capitolo si conclude con il ciclo di vita della Big Data Analytics e con una panoramica sull'evoluzione verso la Data Science, con l'integrazione di Machine Learning e algoritmi avanzati a supporto della strategia aziendale moderna.

1.1 Nascita e classificazione del dato nell'era digitale

Il concetto di "dato" ha assunto oggi una centralità tale da essere considerato il vero motore dell'economia e della società moderna. Il dato non è, però, una risorsa che si trova già pronta in natura; esso è, infatti, il risultato di un processo di osservazione, registrazione e astrazione della realtà. Nell'era digitale, la capacità di catturare ogni minimo evento e di trasformarlo in un valore binario ha cambiato radicalmente il modo in cui le persone interagiscono e le aziende operano.

1.1.1 Evoluzione storica del dato

L'esigenza di raccogliere e organizzare i dati non è un fenomeno nato con l'informatica, ma una necessità intrinseca alla civiltà umana per finalità amministrative, commerciali e di controllo sociale. Tuttavia, è con l'avvento delle tecnologie digitali che la natura del dato ha subìto una trasformazione radicale. Il primo vero salto tecnologico verso l'era moderna avviene alla fine del XIX secolo. Come evidenziato nella letteratura storica di settore, il punto di svolta è rappresentato dal censimento statunitense del 1890. In quell'occasione, Herman Hollerith introdusse l'uso delle macchine tabulatrici a schede perforate. Per la prima volta, i dati (come l'età o il sesso di un cittadino) venivano codificati in fori su cartoncino, permettendo a una macchina elettromeccanica di "leggere" e conteggiare le informazioni. Questa invenzione non solo velocizzò i tempi di calcolo, ma segnò la nascita dell'industria del trattamento automatico dei dati, portando alla fondazione della Tabulating Machine Company, oggi nota come IBM.

Con la nascita dei primi computer commerciali, la gestione dei dati divenne più complessa. Inizialmente, i dati venivano memorizzati su nastri magnetici, rendendo l'accesso sequenziale e inefficiente. La vera rivoluzione avvenne nel 1970 con Edgar F. Codd, che propose il modello relazionale. Questo modello permise di separare la rappresentazione logica dei dati (tabelle e relazioni) dalla loro implementazione fisica sull'hardware, garantendo l'indipendenza dei dati e facilitando la gestione di grandi volumi di informazioni in modo strutturato e coerente.

1.1.2 La digitalizzazione della realtà

Oggi la realtà viene costantemente "mappata" e trasformata in dati: ogni nostra azione, che sia un acquisto o un post sui social o uno spostamento tracciato dal GPS, lascia una traccia digitale. Come sottolinea Serena Sensini, stiamo vivendo un'esplosione di informazioni mai vista prima: se alla fine degli anni '80 i dati scambiati nel mondo si misuravano in Petabyte, oggi siamo passati agli Zettabyte. In pratica, il dato è diventato lo specchio digitale di tutto ciò che facciamo. Per capire come questa mole di dati venga gestita, dobbiamo immaginare l'azienda come un insieme di sottosistemi, dove ogni livello supporta quello superiore (Figura 1.1).

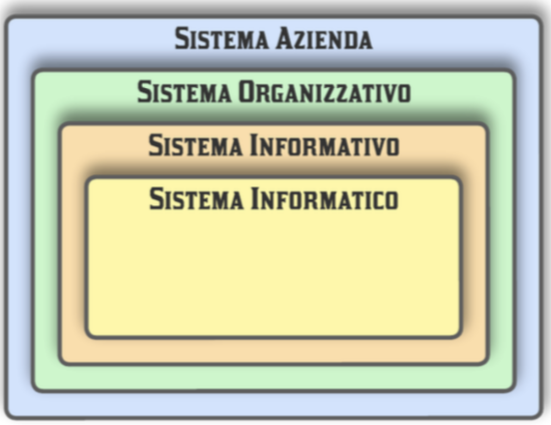


Figura 1.1: Rappresentazione dell'azienda come un sistema con sottosistemi

Questa enorme raccolta di dati trova la sua applicazione pratica all'interno delle organizzazioni, che possono essere viste come un insieme di sistemi annidati l'uno dentro l'altro. Al livello più esterno troviamo il Sistema Azienda, il cui scopo è generare valore trasformando risorse. Per gestire questa complessità, l'azienda si avvale di un Sistema Organizzativo (suo sottosistema), che coordina le persone e pianifica le attività.

A sua volta, all'interno del sistema organizzativo opera il Sistema Informativo; questo rappresenta il cuore logico dell'organizzazione e comprende tutte le risorse e le procedure necessarie per gestire le informazioni. Quest'ultimo non è da confondere con il Sistema Informatico, che ne è solo un ulteriore sottosistema. Mentre il Sistema Informativo esiste indipendentemente dalla tecnologia, il Sistema Informatico ne rappresenta la componente automatizzata, ovvero l'insieme di hardware e software che permette di elaborare i dati in modo digitale.

Con l'era digitale, il Sistema Informatico ha finito per assorbire quasi interamente il Sistema Informativo. Questo significa che l'operatività aziendale, come gli ordini o la gestione del magazzino, viene tradotta in flussi informativi, convertendo eventi fisici in dati strutturati all'interno dei sistemi gestionali. Questa evoluzione ha cambiato profondamente il modo di lavorare: oggi le aziende sono diventate entità data-driven, cioè guidate dai dati. In breve, il dato è la risorsa strategica che permette a tutti questi sottosistemi di comunicare; se i dati

sono ben organizzati nei DBMS (i sistemi di gestione dei database), l'azienda riesce a trovare la strada giusta in mezzo alla confusione delle informazioni moderne.

1.1.3 Tipologie di dato

I dati che circolano nei moderni sistemi informatici non sono tutti uguali; essi variano per complessità e, soprattutto, per il livello di struttura che possiedono. Una prima distinzione fondamentale riguarda la natura del dato, che può essere semplice o complesso. I dati semplici, spesso definiti "atomici", rappresentano le unità minime di informazione che non possono essere ulteriormente suddivise senza perdere il loro significato; ne sono un esempio i numeri interi, i singoli caratteri o i valori logici. Quando più dati semplici vengono aggregati tra loro, si ottengono i dati complessi: una stringa di testo, ad esempio, non è altro che una sequenza organizzata di caratteri, così come una data o un record anagrafico nascono dall'unione di più elementi elementari. Possiamo, quindi, classificare i dati complessi in tre categorie principali, a seconda del loro grado di strutturazione:

- *Dati strutturati*: sono le informazioni organizzate secondo schemi rigidi e predefiniti, tipicamente all'interno di tabelle composte da righe e colonne. Questo è il formato classico dei database relazionali, dove ogni campo ha un tipo di dato stabilito (ad esempio, una colonna dedicata solo a date o solo a valute). Questa struttura è ciò che permette ai software di eseguire ricerche e calcoli in modo estremamente veloce e preciso.
- *Dati non strutturati*: rappresentano la forma più comune di dato generata oggi, ma anche la più difficile da gestire. Si tratta di informazioni che non hanno un formato predefinito, come il contenuto di una e-mail, un post sui social media, una registrazione audio o un video. Sebbene questi dati siano ricchi di valore, richiedono tecnologie avanzate per essere interpretati, poiché non possono essere inseriti semplicemente nelle tabelle di un database tradizionale.
- *Dati semi-strutturati*: rappresentano una via di mezzo. Pur non essendo rigidi come i dati strutturati, contengono dei marcatori o delle etichette (noti come tag) che ne facilitano l'organizzazione. Formati come XML o JSON, appartengono a questa categoria: non prevedono una tabella fissa, ma usano le etichette per spiegare cosa rappresenta ogni singolo valore.

Saper distinguere tra queste tipologie è il primo passo per una corretta gestione dell'informazione aziendale. Infatti, a seconda del tipo di dato, cambierà la scelta degli strumenti tecnologici da utilizzare; infatti, si passa dai classici database relazionali per i dati strutturati, fino alle moderne architetture NoSQL e Big Data necessarie per gestire tutto ciò che è privo di una forma prestabilita.

1.1.4 Dal dato all'informazione: la piramide DIKW

Per comprendere il valore reale del dato nell'era digitale non basta classificarlo tecnicamente; occorre analizzare il percorso attraverso il quale esso viene raffinato per produrre valore. Questo processo è descritto dal modello della piramide DIKW (Data, Information, Knowledge, Wisdom), uno schema gerarchico che illustra come la conoscenza si costruisca partendo da elementi grezzi per arrivare a decisioni strategiche. Come si può osservare nella Figura 1.2, la piramide poggia su una base molto larga costituita dai Dati grezzi. In questa fase, i dati non sono stati ancora interpretati; sono frutto di una raccolta massiccia. Si viene a creare un paradosso interessante: i dati grezzi sono contemporaneamente fondamentali e

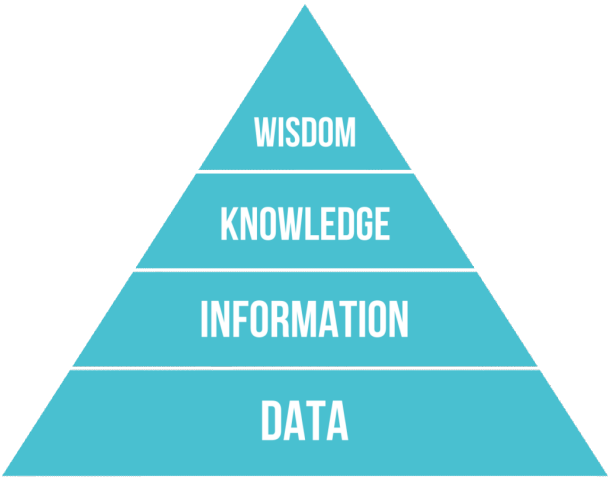


Figura 1.2: Piramide DIKW

inutili. Sono fondamentali perché rappresentano il punto di partenza imprescindibile per ogni analisi; sono inutili perché, se lasciati isolati in un database senza strumenti per leggerli, non hanno alcun significato autonomo e non portano ad alcun miglioramento aziendale.

Il primo salto di livello avviene con il passaggio all'Informazione. Come indicato nello schema, questo avanzamento si ottiene attraverso un processo di aggregazione e contestualizzazione. È qui che entra in gioco la Business Intelligence (BI) che, tramite processi di elaborazione, permette di mettere i dati a confronto tra loro. Solo quando il dato viene contestualizzato inizia ad avere un valore intrinseco.

Salendo ancora nella gerarchia si raggiunge la Conoscenza. Se l'informazione ci dice cosa è successo, la conoscenza, frutto di un processo di applicazione e sperimentazione, ci spiega il "come". In questa fase l'attenzione si sposta dai sistemi informatici all'intervento umano: i report e le dashboard prodotti dalla BI devono essere interpretati dai decision maker. Raggiungere questo gradino significa integrare i dati nella quotidianità aziendale, trasformandoli in modelli ricorrenti che guidano la comprensione dei fenomeni.

Al vertice della piramide si colloca, infine, la Saggezza. Mentre i livelli precedenti si basano spesso sull'analisi di dati passati, la saggezza è orientata al futuro e riguarda il "perché". Essa rappresenta la capacità di applicare la conoscenza acquisita per progettare azioni strategiche e sviluppare analisi predittive. Un'azienda che raggiunge questo stadio diventa una vera realtà data-driven, capace di usare i dati non solo come memoria storica, ma come framework per prevedere i trend e agire nel modo migliore possibile. La saggezza, dunque, è il punto in cui il dato smette di essere un semplice numero e diventa una guida consapevole per la strategia d'impresa.

1.2 La rivoluzione dei Big Data

Negli ultimi anni, il modo in cui le organizzazioni guardano ai dati è cambiato radicalmente. Non si parla più solo di "gestire informazioni", ma di affrontare una vera e propria ondata tecnologica chiamata "Big Data". Questa rivoluzione non riguarda solo la quantità di dati, ma rappresenta una sfida nel modo in cui questi vengono catturati, conservati e analizzati per generare valore.

1.2.1 I limiti delle architetture tradizionali

Per decenni, il mondo dell'informatica si è basato sui database relazionali. Questi sistemi sono stati perfetti per gestire dati precisi e ordinati, come i conti bancari o le anagrafiche dei dipendenti. Tuttavia, con l'esplosione del web e dei social media, le architetture tradizionali hanno iniziato a mostrare i primi limiti invalicabili. Il primo grande ostacolo è la rigidità dello schema. Nei sistemi tradizionali, prima di inserire un dato, bisogna definire esattamente il suo schema. Se oggi un'azienda riceve dati con una struttura diversa, deve modificare l'intero database, un'operazione lenta e costosa. In un mondo dove i dati cambiano ogni secondo, questa rigidità è diventata un freno. Il secondo limite riguarda la scalabilità. I database classici sono progettati per "crescere in verticale": se i dati aumentano, serve un computer più potente e costoso. Ma oltre un certo limite, non esistono computer abbastanza grandi per gestire i dati globali. Le architetture di Big Data, invece, sono pensate per "crescere in orizzontale", distribuendo il lavoro su centinaia di piccoli computer che operano insieme. Quindi, il passaggio ai Big Data nasce proprio dalla necessità di gestire flussi di informazioni che i sistemi di un tempo non riuscivano nemmeno a leggere.

1.2.2 Cosa sono i Big Data: il paradigma delle 5V

Spesso si commette l'errore di pensare che i "Big Data" siano semplicemente "tanti dati". In realtà, si tratta di un concetto molto più profondo che la letteratura tecnica riassume nel paradigma delle 5V. È interessante notare come questo modello si sia evoluto nel tempo: inizialmente, infatti, la definizione si poggiava solo su tre pilastri fondamentali (Volume, Velocità e Varietà), focalizzati principalmente sulle sfide tecnologiche dell'archiviazione e del calcolo. Solo in un secondo momento sono state aggiunte altre due dimensioni cruciali, la Veracità e il Valore, che hanno spostato l'attenzione sulla qualità del dato e sulla sua utilità strategica.

Queste cinque caratteristiche (Figura 1.3) spiegano perché gestire i Big Data richieda strumenti completamente nuovi rispetto al passato. Esse sono:

- *Volume*: è la caratteristica più banale. Parliamo di quantità di dati talmente grandi (Zettabyte) che non possono essere contenuti in un solo supporto fisico. La sfida qui è trovare spazio e modi efficienti per l'archiviazione.
- *Velocità*: i dati oggi vengono generati e devono essere analizzati quasi in tempo reale. Pensiamo ai sensori di un'auto a guida autonoma o ai tweet durante un evento mondiale: il dato è utile solo se viene processato nell'istante in cui arriva.
- *Varietà*: i Big Data sono un mix disordinato di foto, video, messaggi vocali, file di log e dati GPS. Questa varietà di tipologie di dati rende impossibile l'uso delle sole tabelle tradizionali.
- *Veracità*: con un numero così grande di dati aumenta il rischio di imprecisioni. Questa caratteristica riguarda proprio la qualità del dato, ovvero quanto un dato è affidabile
- *Valore*: è la caratteristica più importante per un'azienda. Avere miliardi di dati non serve a nulla se non si riesce a estrarre da essi informazioni per avere un vantaggio. Il valore è il fine ultimo che consiste nel trasformare quella massa di bit in conoscenza utile per produrre valore

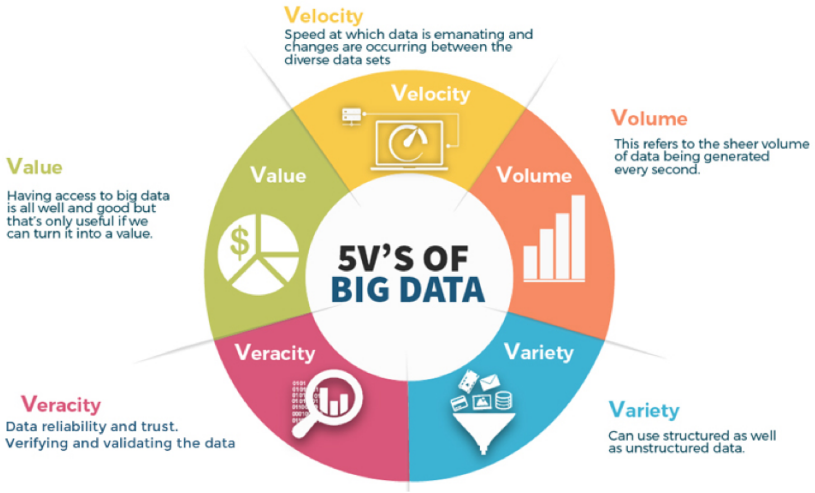


Figura 1.3: Paradigma delle 5V

1.3 La Data Analytics

Raccogliere un'enorme mole di dati, come abbiamo visto parlando di Big Data, non serve a molto se l'azienda non ha gli strumenti per leggerli e interpretarli. La Data Analytics è proprio l'insieme delle tecniche che permette di prendere decisioni, dopo aver effettuato l'analisi dei dati. L'obiettivo finale non è solo avere dei grafici colorati, ma creare valore per il business: si passa dal decidere basandosi solo sulle sensazioni o sull'esperienza passata, al decidere basandosi su prove oggettive estratte dai dati. L'analisi non è un processo unico, ma si divide in quattro livelli che diventano via via più complessi e utili, come si può vedere nella Figura 1.4.

Figura 1.4: Livelli di analisi

1.3.1 Analisi descrittiva

L'analisi descrittiva è il punto di partenza fondamentale per ogni azienda. Senza di essa, ci si muoverebbe al buio. Questa metodologia si occupa di rispondere alla domanda: "Che cosa è successo?".

In sostanza, l'analisi descrittiva riceve milioni di righe di dati e li trasforma in grafici. In azienda, questo si traduce solitamente in due strumenti, ovvero:

- *I report statici*: sono i classici documenti (come PDF o fogli Excel) che arrivano periodicamente via e-mail. Sono molto amati perché sono "comodi": i dati arrivano già pronti. In essi si possono trovare i KPI (Key Performance Indicator), ovvero quegli indicatori chiave che permettono di monitorare a colpo d'occhio l'andamento del business. Tuttavia, il limite principale del report è la sua staticità: l'utente può vedere solo ciò che è stato prestabilito in fase di creazione, senza la possibilità di "navigare" tra le informazioni o applicare filtri per approfondire dati correlati o scoprire nuove prospettive.
- *Le dashboard interattive*: sono strumenti più evoluti e dinamici (come quelli di Power BI), dove l'utente può interagire con i dati, passando da uno spettatore passivo a un esploratore attivo. Attraverso pagine digitali dotate di filtri, pulsanti e selettori, l'utente può interagire con i dati in tempo reale. Questo permette di personalizzare l'indagine in base alle proprie necessità, isolando specifici set di dati o approfondendo i dettagli che più incuriosiscono. In questo caso, non si riceve solo una fotografia fissa del passato, ma si ha a disposizione un ambiente di analisi autonomo in cui cercare attivamente le risposte che si cercano.

1.3.2 Analisi diagnostica

Una volta capito cosa è successo, la domanda successiva è naturale: "Perché è successo?". Qui entra in gioco l'analisi diagnostica. Il suo obiettivo è "scavare" sotto la superficie per trovare le relazioni tra i diversi dati.

Per fare questo, si utilizzano tecniche come l'analisi di correlazione o i test statistici. Un esempio molto importante è la rilevazione delle anomalie. Immaginiamo una linea di produzione industriale piena di sensori: l'analisi diagnostica è in grado di ispezionare continuamente i dati per trovare fenomeni strani o malfunzionamenti. Se un macchinario inizia a scaldarsi più del solito, il sistema invia un segnale d'allarme, permettendo ai tecnici di intervenire prima che il guasto blocchi tutto. In sintesi, l'analisi diagnostica serve a dare un senso ai trend che abbiamo osservato nella fase descrittiva.

1.3.3 Analisi predittiva

L'analisi predittiva rappresenta un salto di qualità perché sposta l'attenzione dal passato al futuro. L'obiettivo è rispondere alla domanda: "Che cosa succederà?". In questa fase non si usano solo semplici medie, ma algoritmi di Intelligenza Artificiale che cercano dei modelli (pattern) nei dati storici per calcolare delle probabilità su eventi futuri.

Grazie a questa analisi, le aziende possono fare:

- *Forecast* (Previsioni): un esempio potrebbe essere stimare quante vendite ci saranno il prossimo mese, quale sarà il prezzo di una materia prima o il livello di rischio di un investimento.
- *Propensity model* (Modelli di propensione): un esempio potrebbe essere prevedere quanto è probabile che un cliente accetti un'offerta o, al contrario, quanto è probabile che ci abbandoni per un concorrente.
- *Segmentazione*: un esempio potrebbe essere raggruppare i clienti in base a caratteristiche simili per offrire loro un servizio personalizzato.

Anticipare il futuro significa smettere di rincorrere gli eventi e iniziare a prepararsi in anticipo, ottenendo un enorme vantaggio competitivo.

1.3.4 Analisi prescrittiva

L'ultimo gradino, il più avanzato, è l'analisi prescrittiva; questa risponde alla domanda decisiva: "Che cosa dobbiamo fare?". Essa non si limita a prevedere un futuro possibile, ma suggerisce il miglior modo per raggiungere un obiettivo.

Esempi di analisi prescrittiva sono ovunque nella nostra vita digitale. Tra essi citiamo i seguenti:

- *Sistemi di raccomandazione*: quando Amazon o Netflix ci suggeriscono cosa comprare o guardare, un algoritmo ha già analizzato migliaia di opzioni e ha "scremato" per noi solo quelle più adatte ai nostri gusti.
- *Agenti autonomi*: nei casi più complessi, come il trading finanziario automatizzato o la pubblicità online, gli algoritmi non solo suggeriscono l'azione, ma la eseguono in tempo reale senza aspettare l'approvazione umana. Questi sistemi imparano continuamente dai propri successi e fallimenti, correggendo la strategia per ottenere il miglior risultato possibile.

1.4 Il ciclo di vita della Big Data Analytics

Il ciclo di vita della Big Data Analytics rappresenta l'intero processo che permette di estrarre informazioni e conoscenza di alto valore da contenitori di dati enormi e complessi. Questo percorso comprende ogni singola fase della gestione del dato, dalla sua raccolta iniziale fino all'utilizzo finale per prendere decisioni strategiche in azienda.

Come illustrato nella Figura 1.5, questo ciclo si articola in nove fasi ben distinte, che esamineremo nel dettaglio nei prossimi paragrafi.

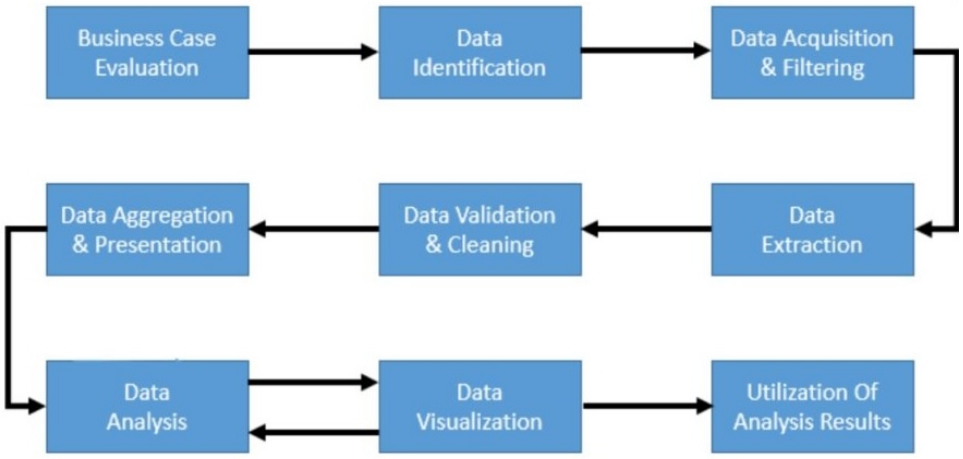


Figura 1.5: Ciclo di vita della Big Data Analytics

1.4.1 Business case evaluation

Questa è la fase iniziale del progetto. Prima ancora di mettere mano sui dati, o investire risorse economiche e tecnologiche, è fondamentale valutare la fattibilità e il potenziale del progetto. Le attività principali sono le seguenti:

- *Definizione del problema*: consiste nel descrivere chiaramente la sfida aziendale o l'opportunità che si intende cogliere. In questa fase è fondamentale tradurre un'esigenza di

business generica (spesso espressa in modo ambiguo dai manager) in una domanda analitica precisa.

- *Fissazione degli obiettivi*: consiste nello stabilire dei traguardi che si vogliono raggiungere (tramite KPI) per poter valutare, una volta terminato il progetto, il suo successo.
- *Coinvolgimento degli stakeholder*: consiste nell'identificare tutte le figure chiave, dai manager aziendali ai data scientist, fino agli utenti finali che useranno i risultati.
- *Valutazione preliminare*: consiste nel fare una primissima stima sulla reale disponibilità e qualità dei dati necessari per affrontare la sfida.

1.4.2 Data Identification

Una volta definito il "perché", bisogna capire "cosa" analizzare. L'identificazione dei dati serve a "mappare" tutte le informazioni a disposizione dell'azienda. I passaggi chiave di questa fase prevedono:

- *La mappatura delle fonti*: consiste nell'individuare dove risiedono i dati. Questi ultimi possono essere database interni, documenti di testo oppure fonti esterne, come API, social media e sensori IoT.
- *Il controllo di qualità e pertinenza*: consiste nel verificare che i dati individuati non solo siano di buona qualità, ma contengano effettivamente gli attributi utili a risolvere il problema iniziale.
- *La verifica di accessibilità*: consiste nell'accertarsi che i dati siano legalmente e tecnicamente utilizzabili, tenendo conto di eventuali vincoli contrattuali o di privacy.
- *Il data profiling*: consiste nell'eseguire una prima scansione delle caratteristiche dei dati per preparare il terreno alle elaborazioni successive.

1.4.3 Data Acquisition and Filtering

In questa fase si passa all'azione: i dati vengono materialmente raccolti e subiscono una prima scrematura. Il processo si divide in due macro-attività:

- *L'acquisizione*: i dati vengono estratti dalle fonti e vengono combinati in un unico ambiente. È in questo momento che si iniziano ad aggiungere i "metadati" (le etichette descrittive del dato), sia per le fonti interne che esterne.
- *Il filtraggio*: si esamina la coerenza delle informazioni appena acquisite. I dati subiscono una pulizia iniziale (rimozione di duplicati o valori palesemente anomali) e, se necessario, il volume del set di dati viene ridotto o trasformato in un formato più idoneo per alleggerire il lavoro dei server.

1.4.4 Data Extraction

I dati raccolti, specialmente nel mondo dei Big Data, sono spesso eterogenei e incompatibili tra loro. L'estrazione serve proprio a "tradurre" queste informazioni nel formato corretto per il sistema di analisi scelto. Per farlo, si definisce l'ambito di interesse e si scelgono le tecniche migliori: per i database classici si usano linguaggi come SQL, mentre per testi complessi o documenti si ricorre al Natural Language Processing (NLP) o al data parsing.

1.4.5 Data Validation and Cleansing

Questa è una delle fasi più delicate, poiché l'affidabilità delle decisioni future dipende dall'accuratezza dei dati di partenza. La validazione e la pulizia vanno molto più a fondo rispetto al primo filtraggio. Durante questo passaggio, i dati vengono pre-elaborati, integrati e confrontati. Un metodo classico è la validazione incrociata: se si uniscono due dataset (A e B), i valori del secondo vengono controllati rispetto al primo. Se ci sono delle mancanze o delle incongruenze, il sistema cerca di colmarle usando la fonte più affidabile. Solo i dati che superano questi rigidi controlli di qualità verranno passati alle fasi di sviluppo e analisi dei modelli. La fase di Data Validation and Cleansing rappresenta, quindi, un passaggio obbligato per garantire l'accuratezza, la coerenza e l'affidabilità delle informazioni. Come illustrato nella Figura 1.6, il Data Validation and Cleansing Cycle può essere rappresentato come un processo circolare e iterativo, suddiviso in diverse fasi fondamentali che garantiscono la qualità del dato.

Figura 1.6: Data Validation and Cleansing Cycle

Nello specifico, il ciclo si articola nei seguenti step:

- *Merge Data* (Integrazione dei dati): questa fase prevede l'unione e il consolidamento di informazioni provenienti da fonti o dataset differenti. L'obiettivo è creare una base di dati unica e coerente su cui operare.
- *Rebuild Missing* (Ricostruzione dei dati mancanti): uno dei problemi più comuni è la presenza di valori nulli o mancanti. In questo step si interviene per colmare queste lacune, eliminando le righe incomplete o, più frequentemente, utilizzando tecniche di imputazione (ad esempio sostituendo i valori mancanti con la media, la mediana o i modelli predittivi) per stimare i valori mancanti senza alterare la distribuzione statistica generale.
- *Verify Data and Validate* (Verifica e Validazione dei dati): i dati vengono sottoposti a rigidi controlli logici e strutturali per assicurarne l'accuratezza e la coerenza. Si verifica che le informazioni rispettino specifiche regole di dominio, identificando anomalie, errori di inserimento o valori fuori scala.
- *Remove Duplicate Data* (Rimozione dei duplicati): consiste nell'individuazione e rimozione di record ripetuti o ridondanti. La presenza di duplicati può falsare le analisi statistiche e sbilanciare l'addestramento dei futuri modelli di Machine Learning, rendendo questo passaggio cruciale.

- *Standardise and Normalise Data*(Standardizzazione e Normalizzazione): questa fase mira a uniformare il dataset. La standardizzazione assicura che i dati seguano formati comuni (ad esempio, formattazione univoca per le date o per le stringhe di testo). La normalizzazione, invece, interviene sui dati numerici, riportando variabili con unità di misura o scale molto diverse all'interno di range comparabili.

- *Import and Export Data* (Importazione ed Esportazione): rappresenta il flusso operativo del dato; i dati "grezzi" vengono importati nel sistema per essere elaborati. Una volta terminato il ciclo di pulizia, il dataset raffinato viene esportato.

1.4.6 Data Aggregation and Representation

A questo punto, i dati puliti devono essere uniti per creare una "vista unificata" del problema. Aggregare enormi volumi di dati presenta, però, due sfide principali:

- *Struttura*: due dataset potrebbero avere lo stesso formato (ad esempio un foglio di calcolo) ma modelli logici di archiviazione completamente diversi.

- *Semantica*: sistemi diversi potrebbero usare nomi differenti per indicare la stessa cosa (ad esempio, un database potrebbe usare l'etichetta "surname" e un altro "last name").

  Superare questi ostacoli richiede tecniche di modellazione e standardizzazione che permettano alle macchine di leggere i dati in modo univoco, basandosi spesso su campi chiave (come l'ID di un cliente) per unire le informazioni. Un esempio di questo passaggio è rappresentato nella Figura 1.7, che mostra come due dataset distinti possano essere combinati per creare una struttura dati standardizzata.

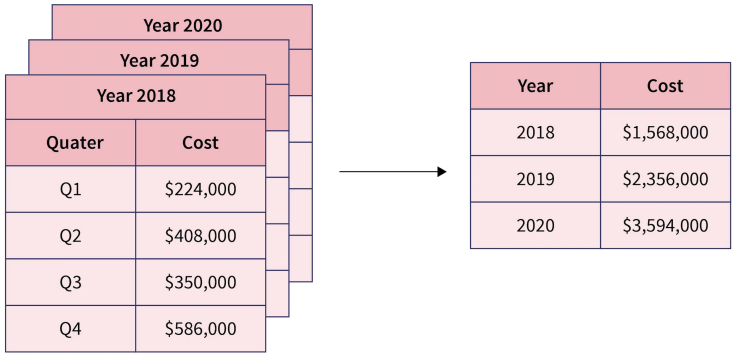


Figura 1.7: Aggregazione di dataset

1.4.7 Data Analysis

È il cuore pulsante dell'intero ciclo; si tratta di un processo tipicamente iterativo (che si ripete più volte) in cui gli analisti cercano di estrarre significato dalla massa di dati. Questo passaggio include:

- *L'esplorazione*: consiste in un'analisi preliminare per individuare valori anomali e capire la distribuzione delle variabili tramite statistiche descrittive.

- *Il Feature Engineering*: consiste nella creazione di nuove variabili o nella trasformazione matematica di quelle esistenti (come la conversione di categorie testuali in numeri) per favorire gli algoritmi.

- *L'Integrazione e il Monitoraggio*: una volta che i modelli matematici trovano dei pattern o generano insight utili, questi vengono implementati nei sistemi aziendali. Poiché il mondo cambia, le prestazioni di questi modelli dovranno essere monitorate costantemente nel tempo.

1.4.8 Data Visualization

Anche la migliore analisi del mondo risulta inutile se non viene compresa dai manager aziendali. La visualizzazione dei dati serve proprio a tradurre i modelli matematici (come grafici di regressione, alberi decisionali o matrici di confusione) in formati visivi chiari e immediati. Attraverso diagrammi a dispersione, mappe di calore o dashboard interattive, i Data Scientist raccontano una storia che trasmette le scoperte in modo chiaro. In molti casi, si creano ambienti self-service che permettono anche a chi non ha competenze tecniche di esplorare i dati in autonomia, monitorare allerte in tempo reale e prendere decisioni tempestive. Un esempio concreto di questa capacità comunicativa è rappresentato dalla dashboard nella Figura 1.8.

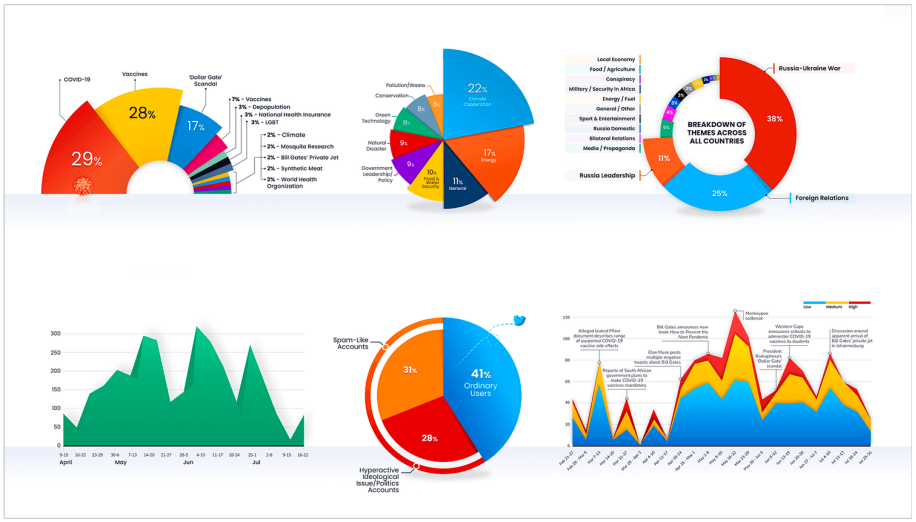


Figura 1.8: Esempio di Data Visualization

1.4.9 Utilization of Analysis Results

L'ultima fase chiude idealmente il ciclo e dà senso a tutto il lavoro svolto. Gli insight estratti non rimangono pura teoria, ma diventano il motore per le azioni aziendali. In base ai risultati ottenuti, i decision maker possono ottimizzare un processo, lanciare un nuovo prodotto o correggere una strategia. Subito dopo, si misurano nuovamente i KPI per valutare se la decisione presa ha avuto l'efficacia sperata rispetto ai risultati attesi. In un'ottica di miglioramento continuo, le lezioni apprese in questa fase diventano le nuove fondamenta per i successivi cicli di analisi.

1.5 Evoluzione verso la Data Science

Se la Big Data Analytics rappresenta il motore metodologico per gestire grandi volumi di informazioni, la Data Science ne costituisce l'evoluzione scientifica e multidisciplinare. Definibile come la "scienza che estrae valore dai dati", essa non si limita a osservare il passato, ma utilizza metodi statistici avanzati e tecniche di analisi per trasformare il dato grezzo in

informazioni strategiche. Questa disciplina ha radici profonde; il termine fu coniato nel 1974 da Peter Naur, ma è nel 2001 che William Cleveland ne ha delineato lo statuto scientifico, individuando sei aree di competenza: i fondamenti teorici, le indagini multidisciplinari, la costruzione di modelli statistici, il calcolo e l'elaborazione dei dati, la valutazione degli strumenti, la formazione.

1.5.1 Dal monitoraggio storico ai modelli dinamici

Il passaggio dalla Business Intelligence tradizionale alla Data Science segna il superamento del semplice monitoraggio storico. Mentre l'analisi tradizionale risponde alla domanda "Cosa è successo?", i modelli dinamici della Data Science si concentrano sul futuro e sull'ottimizzazione in tempo reale. L'importanza strategica dei dati trova conferma nei dati raccolti dall'Osservatorio Big Data and Business Analytics del Politecnico di Milano (Figura 1.9), che fotografano un'Italia in forte transizione digitale, nonostante le sfide poste dal periodo pandemico. Nelle grandi realtà industriali e dei servizi, la valorizzazione del dato è ormai diventata una priorità assoluta: la quasi totalità delle grandi aziende analizzate (il 96%) ha, infatti, avviato progetti strutturati per estrarre valore dal proprio patrimonio informativo. Tuttavia, il vero salto di qualità si osserva in quel 42% di imprese definite "mature". Queste realtà non si limitano più a guardare il passato, ma adottano le cosiddette Advanced Analytics, cioè metodologie che permettono di:

- *Prevedere scenari futuri*: non più solo report statici, ma simulazioni dinamiche che anticipano le tendenze di mercato.
- *Effettuare Manutenzione Predittiva*: monitorare impianti e macchinari in tempo reale per intervenire prima che si verifichi un guasto, riducendo i costi di riparazione.
- *Ottimizzare gli asset*: gestire le risorse aziendali eliminando sprechi e sovraccarichi grazie a flussi di dati costanti.

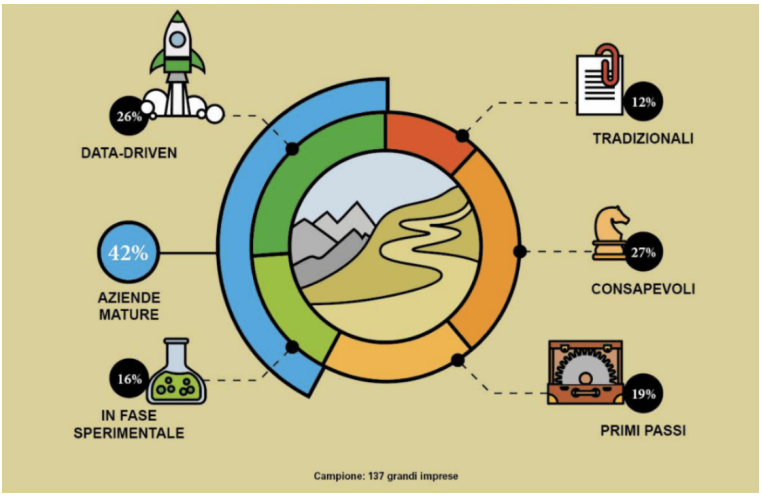


Figura 1.9: Maturità delle grandi imprese italiane nel mercato Analytics

Anche il settore delle Piccole e Medie Imprese, spina dorsale dell'economia italiana, sta mostrando segnali di forte dinamismo. Sebbene con risorse diverse rispetto ai grandi gruppi, nel 2020 una PMI su due ha deciso di investire concretamente in soluzioni di Analytics. Questo interesse si traduce in un impegno costante: oltre il 62% di queste imprese ha infatti attività di analisi dati correntemente in atto, dimostrando che la consapevolezza del dato

come "nuovo petrolio" ha ormai superato i confini delle multinazionali per diventare un asset fondamentale per la sopravvivenza e la crescita di ogni impresa nel territorio nazionale.

1.5.2 Integrazione di algoritmi e Machine Learning

Il cuore pulsante di questa evoluzione risiede nell'integrazione sinergica tra Intelligenza Artificiale, Machine Learning e Data Science. Sebbene nel linguaggio comune questi termini vengano spesso confusi, è fondamentale tracciarne una distinzione netta per comprenderne il legame gerarchico:

- *L'Intelligenza Artificiale* rappresenta il dominio più vasto, orientato allo sviluppo di sistemi capaci di simulare le facoltà cognitive umane sia nel pensiero che nell'azione.
- *La Data Science*, al suo interno, si configura come la disciplina metodologica che estrae valore dai dati attraverso modelli statistici e matematici complessi, occupandosi non solo dell'aspetto tecnico ma anche del contesto strategico.
- *Il Machine Learning* funge da braccio operativo, fornendo l'insieme di algoritmi che permette ai sistemi di apprendere autonomamente dai dati, migliorando le proprie prestazioni nel tempo senza necessità di una programmazione rigida per ogni singola operazione.

L'adozione di queste tecnologie segna il passaggio definitivo da sistemi informatici statici a modelli adattivi basati su diverse metodologie di apprendimento. In particolare, nell'apprendimento supervisionato, o supervised learning, il sistema viene addestrato su dataset etichettati per imparare a riconoscere pattern e prevedere eventi futuri. Al contrario, l'apprendimento non supervisionato interviene su dati privi di etichette per scoprire strutture nascoste; ne sono un esempio le tecniche di clustering, che raggruppano i dati per somiglianza, e il Data Mining, che ne estrae legami logici precedentemente ignoti.

Questa architettura tecnologica sta trasformando radicalmente i settori chiave dell'economia. In definitiva, la Data Science non rappresenta più un semplice supporto tecnico, ma si è consolidata come il pilastro centrale della strategia aziendale moderna, capace di orientare il management verso una gestione consapevole e predittiva del business.

CAPITOLO 2

Introduzione a Power BI

In questo capitolo viene analizzato il ruolo di Microsoft Power BI come piattaforma leader nel panorama della Business Intelligence. Si ripercorre dapprima l'evoluzione interna di Microsoft, partendo da una gestione frammentata dei propri dati, fino ad arrivare alla nascita di Power BI. Il posizionamento strategico della piattaforma viene certificato dall'analisi del Gartner Magic Quadrant 2025, che conferma Microsoft come Leader assoluto, grazie a tre pilastri fondamentali: l'integrazione in Microsoft Fabric, l'introduzione dell'Intelligenza Artificiale con Copilot e l'unificazione dei flussi operativi e analitici. Successivamente, si descrive l'architettura modulare della piattaforma, illustrando le componenti principali — Desktop, Service, Mobile, Gateway e Report Server — e il modo in cui esse cooperano per coprire l'intero ciclo di vita del dato. Si approfondisce, quindi, il flusso di lavoro tipico, dal processo ETL gestito da Power Query, alla modellazione semantica tramite relazioni e misure DAX, fino allo sviluppo di report e dashboard e alla loro condivisione nel cloud. Il capitolo si conclude con una panoramica sui linguaggi di programmazione integrati — DAX, M, Python e R — e sui principali vantaggi strategici offerti dalla piattaforma, tra cui la sicurezza granulare degli accessi, la versatilità nell'integrazione delle fonti e le funzionalità di Intelligenza Artificiale.

2.1 Il posizionamento nel mercato: Leader della Business Intelligence

Oggi la Business Intelligence non è più un semplice aiuto tecnico, ma è diventata il cuore pulsante delle decisioni aziendali. Essa, infatti, permette di capire subito cosa sta succedendo e come muoversi nel mercato. In questo settore, Microsoft è il punto di riferimento mondiale. Il successo di Power BI non è dovuto solo alla tecnologia, ma al fatto che è riuscito a "democratizzare" i dati, rendendoli accessibili a tutti e non solo agli esperti informatici. Come vedremo, Power BI nasce proprio dai problemi reali che Microsoft aveva al proprio interno; risolvendo i propri difetti, l'azienda ha creato uno standard che oggi è considerato il migliore al mondo per l'analisi dei dati.

2.1.1 L'evoluzione delle soluzioni Microsoft

L'evoluzione di Power BI rappresenta la risposta di Microsoft a una crisi profonda nella gestione dei propri dati interni. Prima di diventare il fornitore leader nel mercato degli Analytics, Microsoft ha vissuto un lungo percorso di trasformazione, affrontando sfide comuni a molte grandi organizzazioni, come definizioni di dati incoerenti, gerarchie contrastanti e una forte resistenza culturale verso la standardizzazione dei processi. Fino a pochi anni fa,

l'ecosistema informativo di Microsoft era caratterizzato da una proprietà frammentata del dato. Questo scenario, spesso definito "Shadow BI", portava gli analisti aziendali a investire oltre il 75% del proprio tempo nella raccolta e compilazione dei dati, piuttosto che nell'analisi. La mancanza di coerenza era tale che ogni area geografica o dipartimento utilizzava KPI personalizzati, generando confusione e portando alla creazione di circa 350 sistemi finanziari centralizzati e una spesa annuale di 30 milioni di dollari per applicazioni non coordinate. La svolta è avvenuta con l'adozione di un nuovo modello operativo basato su due pilastri fondamentali che sono, la "disciplina al centro" e la "flessibilità alla periferia". La disciplina al centro è garantita dall'IT, che cura un'unica fonte di dati principale, stabilendo tassonomie rigorose e permessi centralizzati. La flessibilità alla periferia, invece, permette ai team di Sales, Marketing e Finance di operare in modalità Self-Service BI (SSBI), analizzando i dati velocemente senza dipendere costantemente dal supporto tecnico. Questo percorso ha portato alla creazione di piattaforme interne come Starlight e il KPI Lake, un modello semantico tabulare che aggrega oltre 100 fonti di dati diverse. Proprio dopo queste esperienze di utilizzo interno dei propri prodotti prima del rilascio sul mercato è nato quello che oggi conosciamo come Power BI. Inizialmente, nel 2013, Power BI era "nascosto" all'interno di Excel come componenti aggiuntive (Power Query, Power Pivot e Power View); successivamente il software ha subìto una maturazione rapidissima, fino ad arrivare al rilascio ufficiale della versione il 24 luglio 2015.

2.1.2 Analisi del Gartner Magic Quadrant

Nel panorama competitivo delle piattaforme di Analytics e Business Intelligence, il "Magic Quadrant" elaborato annualmente dalla società di consulenza Gartner rappresenta lo standard globale per valutare il posizionamento strategico e tecnologico dei principali attori del mercato. Come evidenziato nell'edizione pubblicata a giugno 2025 (Figura 2.1), Microsoft si è confermata nel quadrante dei "Leader" per il diciottesimo anno consecutivo. Un traguardo ulteriormente rafforzato da un record specifico, ossia che per il settimo anno di fila, l'azienda si è posizionata al vertice assoluto di tutto il mercato sia per la "Completezza di Visione" sia per la "Capacità di Esecuzione". Questo solido primato, supportato oggi da una comunità di oltre 30 milioni di utenti attivi su base mensile, è il risultato di una profonda evoluzione tecnologica che ha trasformato Power BI da un semplice strumento di reportistica a una piattaforma integrata avanzata. Analizzando il report e le recenti innovazioni, emergono tre pilastri fondamentali che giustificano questa leadership, che sono:

- *L'integrazione ecosistemica in Microsoft Fabric*: nel suo decimo anno di vita come soluzione stand-alone, Power BI è diventato il cuore pulsante di Microsoft Fabric, una piattaforma dati SaaS (Software-as-a-Service) unificata. Questa convergenza permette agli utenti di passare dalla preparazione dei dati, all'analisi su larga scala (Big Data) all'interno di un unico ambiente, abbattendo i silos informativi tradizionali e sfruttando funzionalità avanzate come i notebook Python o i sistemi di allerta in tempo reale.

- *L'integrazione dell'Intelligenza Artificiale*: con l'introduzione di Copilot in Power BI la barriera tecnica per l'esplorazione dei dati è stata drasticamente ridotta. Funzionalità come il "Chat with your data" (dialoga con i tuoi dati) permettono agli utenti, anche quelli privi di competenze tecniche, di interrogare report e modelli semantici utilizzando il linguaggio naturale, delegando all'AI la stesura di complesse formule matematiche (DAX) e la creazione automatica di visualizzazioni.

- *L'unificazione dei flussi operativi e analitici*: tradizionalmente, le aziende utilizzano sistemi separati per le operazioni quotidiane e per l'analisi dei dati, generando latenza e

rallentando i processi decisionali. Oggi gli utenti possono intraprendere azioni operative direttamente dalle dashboard di Power BI.

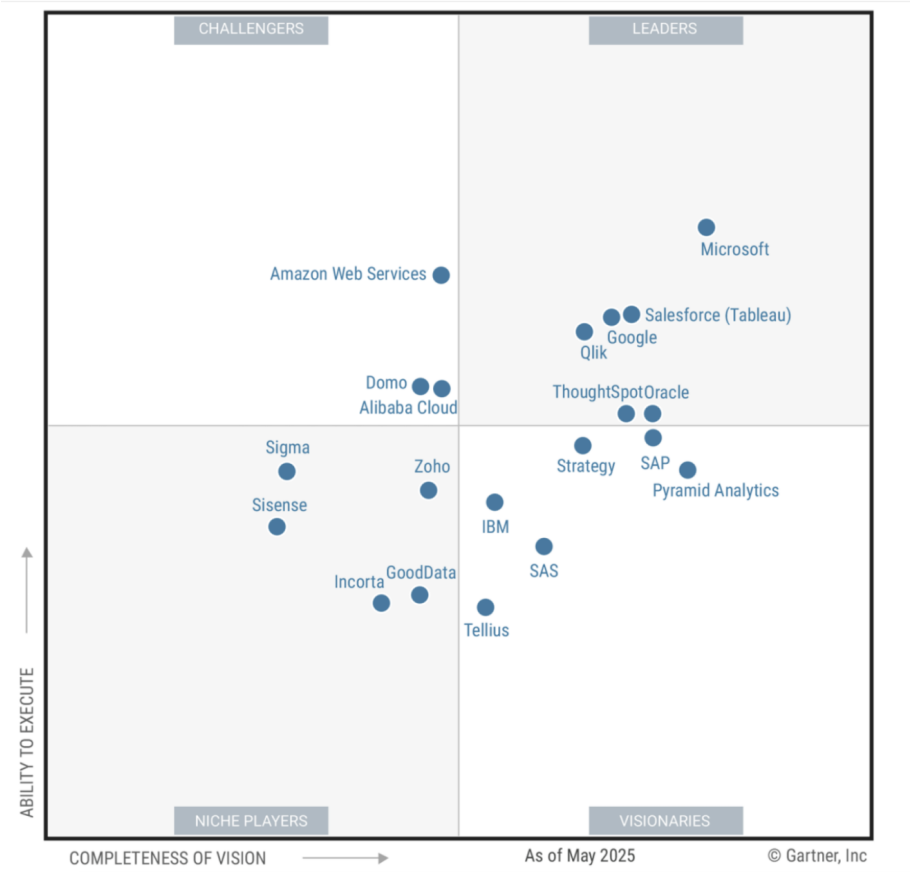


Figura 2.1: Gartner Magic Quadrant per Analytics e BI Platforms (2025)

L'impatto di questa visione "Leader" si riflette in vantaggi operativi concreti per le imprese. Realtà aziendali come Lumen Technologies hanno sfruttato questa architettura unificata per eliminare migliaia di ore di elaborazione manuale, ottenendo metriche aggiornate in tempo reale a costo zero di latenza. In ambito no-profit, la Make-A-Wish Foundation ha trasformato le proprie operazioni creando cruscotti direzionali in tempo reale per le sedi locali, ottimizzando la pianificazione strategica e aumentando l'impatto delle proprie iniziative sul territorio. In sintesi, l'analisi del Magic Quadrant 2025 certifica che il vantaggio competitivo di Microsoft risiede nella capacità di fondere l'accessibilità per l'utente finale (tramite l'AI) con la potenza architetturale (tramite Fabric), offrendo una soluzione capace di scalare dalle piccole necessità aziendali fino ai Big Data globali.

2.2 Componenti di Power BI

Il successo di Power BI come piattaforma di Business Intelligence risiede nella sua architettura modulare, composta da diversi strumenti che lavorano in modo sinergico per coprire l'intero ciclo di vita del dato, dalla raccolta alla distribuzione. Questa struttura permette di separare le fasi di sviluppo tecnico dall'utilizzo delle informazioni, garantendo flessibilità e sicurezza a ogni livello dell'organizzazione.

2.2.1 Power BI Desktop

Power BI Desktop rappresenta l'applicazione principale dedicata allo sviluppo e alla modellazione dei dati. Si tratta di un'interfaccia client gratuita, installabile localmente su sistemi Windows. Attraverso questo strumento è possibile connettersi a una vasta gamma di sorgenti di dati, combinandole tra loro per creare modelli complessi. L'applicazione si divide in sezioni distinte che seguono i passaggi logici di un progetto di analisi, ovvero l'importazione e la trasformazione dei dati, la creazione di relazioni tra le tabelle e, infine, la realizzazione di oggetti visivi interattivi. La caratteristica peculiare di questa componente è la sua capacità di trasformare dati grezzi e complessi in report grafici pronti per essere condivisi, costituendo, di fatto, il punto di partenza della maggior parte dei flussi di lavoro di Business Intelligence.

2.2.2 Power BI Service

Una volta completata la fase di creazione sul Desktop, i report vengono pubblicati nel Power BI Service, noto anche come Power BI Online. Questa è la componente SaaS (Software as a Service) basata sul cloud, che costituisce il cuore collaborativo della piattaforma. Mentre il Desktop è orientato alla creazione, il Servizio è progettato per la condivisione. All'interno di questo ambiente, gli utenti possono realizzare dashboard interattive che riuniscono in un'unica posizione i dati più rilevanti, permettendo un monitoraggio costante delle performance aziendali. Il Servizio offre, inoltre, funzionalità avanzate; infatti, è possibile creare aree di lavoro per collaborare con i colleghi, distribuire contenuti sotto forma di app aziendali e gestire i permessi di accesso per garantire che ogni membro del team visualizzi solo ciò che è di propria competenza. Questa natura cloud-based facilita l'aggiornamento automatico dei dati e la collaborazione in tempo reale.

2.2.3 Power BI Mobile

Per rispondere alle esigenze di una gestione aziendale dinamica, Microsoft offre Power BI Mobile. Si tratta di un' applicazione nativa per dispositivi iOS e Android che permette di accedere ai propri dati in versione mobile. Non è un semplice strumento di visualizzazione, ma uno strumento ottimizzato che permette di interagire con i report e ricevere avvisi tempestivi sulle performance aziendali anche quando si è fuori ufficio. Questa componente garantisce che il flusso informativo non si interrompa, favorendo un processo decisionale continuo e basato su insight sempre aggiornati.

2.2.4 Power BI Gateway

Sebbene non sia sempre visibile all'utente finale, il Power BI Gateway è un componente infrastrutturale critico. Esso funge da "ponte" sicuro tra i dati residenti on-premises (ovvero all'interno dei server fisici aziendali) e i servizi cloud di Power BI. Il Gateway permette di mantenere i dati dove risiedono originariamente, garantendo, al contempo, che i report pubblicati sul Web siano costantemente aggiornati tramite connessioni protette e criptate. Esso è lo strumento che collega in sicurezza i server dell'azienda con il cloud, permettendo di usare i dati locali senza esporli a rischi.

2.2.5 Power BI Report Server

Per le organizzazioni che hanno particolari esigenze di conformità normativa o che preferiscono non caricare i propri dati sul cloud pubblico, la soluzione ideale è Power BI Report Server. Si tratta di un'opzione che permette di ospitare e distribuire i report interattivi

interamente all'interno del perimetro di sicurezza dell'organizzazione. Questo componente offre un controllo completo sull'infrastruttura e sui dati, garantendo un'esperienza utente molto simile a quella del servizio cloud, e rappresenta il punto di incontro tra le esigenze di sicurezza tradizionale e le moderne capacità di visualizzazione analitica.

2.3 Flusso di lavoro in Power BI

2.3.1 Il processo ETL: estrazione, trasformazione e caricamento dei dati

Il cuore pulsante della preparazione dei dati in Power BI è rappresentato da Power Query, un motore che gestisce l'intero ciclo ETL (Extract, Transform, Load). A differenza dei sistemi tradizionali, Power Query unifica in un unico ambiente l'estrazione dai dati grezzi, la loro pulizia e il caricamento finale, permettendo di trasformare dati in informazioni pronte per essere analizzate.

Estrazione

Il primo e fondamentale passo di qualsiasi progetto in Power BI consiste nel collegare lo strumento alle fonti informative. La potenza di Power BI risiede proprio nella sua estrema versatilità; infatti, la piattaforma è progettata per "dialogare" con un'ampia varietà di sorgenti. Il punto di partenza di questo processo è il comando "Recupera dati", situato nella barra multifunzione principale di Power BI Desktop. Facendo click su questa opzione, si apre una finestra che offre l'accesso a centinaia di collegamenti predefiniti alle diverse sorgenti, organizzati per categorie. Come mostrato nella Figura 2.2, l'utente ha la possibilità di accedere a dati provenienti da sorgenti molto diverse tra loro; queste possono essere:

- *File locali,* come fogli di calcolo Excel, file CSV, XML o semplici documenti di testo.
- *Database,* come connessioni dirette a sistemi strutturati quali SQL Server, Oracle, MySQL o IBM DB2.
- *Piattaforme Cloud,* come l'integrazione nativa con l'ecosistema Azure, ma anche con servizi esterni quali Google Analytics, Salesforce e SharePoint.
- *Altre sorgenti;* è possibile estrarre dati direttamente da pagine web, feed OData o script di programmazione, come R e Python.

Power BI permette di scegliere tra diverse modalità di importazione: è possibile caricare fisicamente i dati nel modello (per massimizzare la velocità di analisi) oppure stabilire una connessione in tempo reale, dove lo strumento interroga la sorgente originale solo quando necessario. In questo modo, l'azienda può garantire che i report siano sempre allineati allo stato più recente delle proprie basi dati, eliminando i passaggi manuali e riducendo drasticamente il rischio di errore umano.

Trasformazione

Una volta stabilita la connessione con le sorgenti informative, i dati si presentano raramente in un formato pronto per l'analisi; essi, infatti, spesso contengono errori, duplicati o una struttura poco adatta alla creazione di report. In questa fase entra in gioco Power Query, lo strumento integrato in Power BI che consente di preparare e trasformare i dati prima dell'analisi. La sua operatività si sviluppa attraverso l'Editor di Power Query, un'interfaccia grafica che permette di modellare i dati in modo semplice e visivo. Il principale vantaggio di

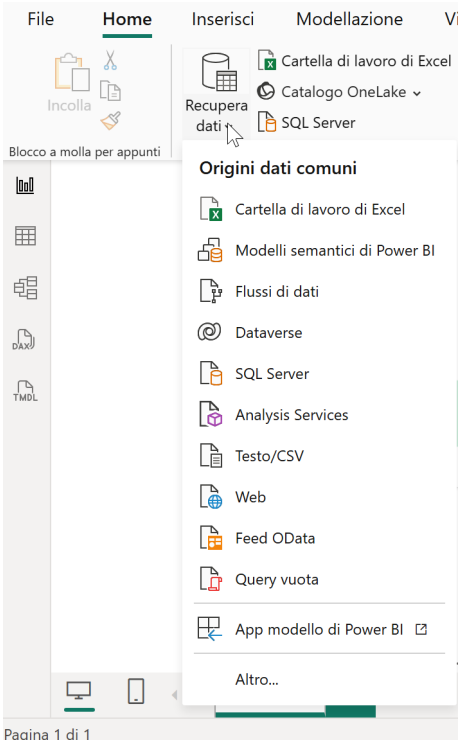


Figura 2.2: Interfaccia per la selezione delle sorgenti dei dati tramite il comando "Recupera dati"

questo strumento è che ogni operazione eseguita dall'utente non modifica il dato originale alla sorgente, ma viene registrata come un "passaggio applicato", riprodotto automaticamente a ogni aggiornamento del report. L'interfaccia, come mostrato nella Figura 2.3, è organizzata in una barra multifunzione che guida l'analista attraverso le diverse fasi di pulizia e raffinamento dei dati. Power Query mette a disposizione strumenti avanzati per dare una forma precisa allo schema dei dati; tali strumenti sono:

- *Pulizia e normalizzazione:* è possibile promuovere la prima riga a intestazione di tabella, gestire i valori mancanti, filtrare le informazioni inutili o eliminare i record duplicati per garantire l'integrità del database.
- *Ristrutturazione delle tabelle:* lo strumento permette di raggruppare i dati per categorie specifiche o di trasformare la struttura stessa della tabella tramite le funzioni di Pivot e Unpivot. In particolare, l'Unpivot consente di trasformare colonne che rappresentano valori della stessa variabile in righe, rendendo la tabella più ordinata e adatta all'analisi. Ad esempio, se i mesi dell'anno sono disposti in colonne separate, questa operazione li converte in un'unica colonna contenente il mese e in un'altra contenente il valore corrispondente, così da ottenere un formato molto più leggibile, flessibile e utilizzabile nei report e nei modelli analitici.
- *Arricchimento del dato:* oltre alla trasformazione di ciò che già esiste, Power Query consente di aggiungere nuove colonne basate su logiche personalizzate, fornendo, al contempo, strumenti di diagnostica per individuare eventuali colli di bottiglia o errori nel caricamento.

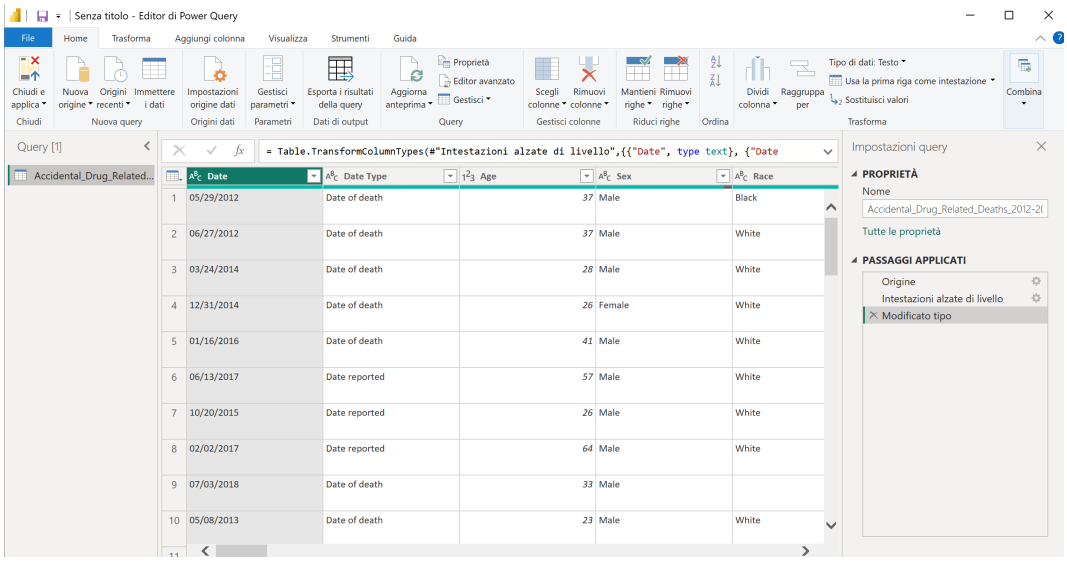


Figura 2.3: L'interfaccia dell'Editor di Power Query

Caricamento

Completata la fase di trasformazione, il passaggio finale consiste nel caricare i dati nel modello di Power BI. È sufficiente selezionare il comando "Chiudi e applica" nell'Editor di Power Query affinchè tutte le trasformazioni definite vengano applicate in sequenza e i dati risultanti siano resi disponibili per la modellazione e la creazione di report. Da questo momento, a ogni aggiornamento del file, Power BI rieseguirà automaticamente l'intero processo, dalla connessione alla sorgente fino al caricamento, garantendo che i dati siano sempre aggiornati.

2.3.2 Modellazione dei dati

Una volta caricati i dati nel modello, si entra nella fase di modellazione, che costituisce il nucleo logico dell'intero progetto svolto in Power BI. L'obiettivo è strutturare e mettere in relazione le tabelle importate in modo da costruire un modello semantico coerente, su cui i report possano basarsi per restituire risultati accurati e performanti. Il primo passaggio consiste nella definizione delle relazioni tra tabelle, un elemento fondamentale per garantire che i calcoli e i report restituiscano risultati corretti. Power BI permette di collegare tabelle diverse tramite campi comuni, rispecchiando la logica con cui i dati sono organizzati alla sorgente.

Un aspetto particolarmente utile è il rilevamento automatico delle relazioni. Infatti, quando vengono caricate più tabelle contemporaneamente, Power BI Desktop analizza i nomi delle colonne e, se individua corrispondenze con un sufficiente grado di attendibilità, crea le relazioni in modo autonomo. Questo meccanismo riduce significativamente il lavoro manuale nella maggior parte dei casi. Tuttavia, non sempre il rilevamento automatico è sufficiente; infatti, quando le corrispondenze tra colonne non sono univoche o il modello presenta strutture più complesse, è necessario intervenire manualmente.

Al centro della modellazione si trovano le misure DAX (Data Analysis Expressions) che nascono dall'esigenza di rispondere a domande analitiche che i dati grezzi, da soli, non sono in grado di soddisfare. Infatti, attraverso formule personalizzate è possibile definire calcoli come somme, medie, conteggi, percentuali o confronti temporali, i cui risultati si aggiornano dinamicamente in risposta ai filtri e alle interazioni dell'utente nel report. Inoltre, per chi si avvicina per la prima volta al linguaggio, Power BI mette a disposizione le misure rapide, cioè

calcoli preconfigurati selezionabili tramite un'interfaccia guidata, utili sia per velocizzare il lavoro sia per apprendere la sintassi DAX analizzando le formule generate automaticamente.

2.3.3 Sviluppo di report e dashboard

Per creare i report a partire dai dati importati con Power Query, è possibile sfruttare l'ampia gamma di visualizzazioni grafiche di Power BI. La piattaforma mette a disposizione strumenti potenti e aggiornati per mettere in luce i trend aziendali e ottenere, a colpo d'occhio, le informazioni chiave. Nel pannello "Visualizzazioni" della schermata Report si trovano numerosi tipi di grafico. Alcuni dei più utilizzati, sono i seguenti:

- *Grafico a barre*, per visualizzare un valore attraverso diverse categorie.
- *Scheda*, per mostrare uno o più valori singoli in evidenza.
- *Grafico a linee*, per mostrare l'andamento nel tempo di un certo fenomeno.
- *Filtro*, per selezionare l'anno o il segmento.
- *Grafico a Torta o Ciambella*, per mostrare il peso di determinati valori sul totale.
- *Mappa*, per visualizzare la distribuzione dei dati per aree geografiche.

Il pannello "Visualizzazioni" di Power BI offre un'ampia scelta di oggetti visivi, come mostrato nella Figura 2.4.

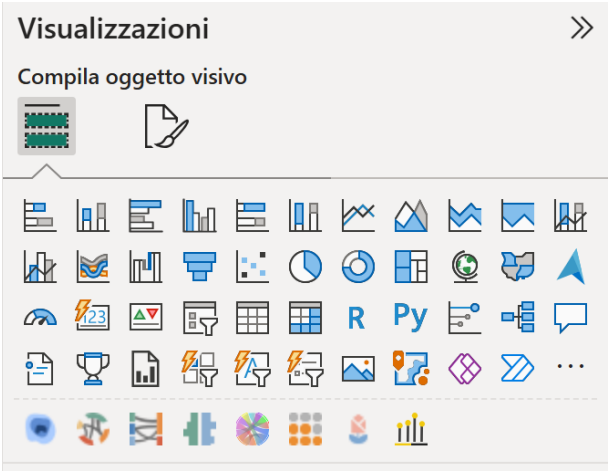


Figura 2.4: Oggetti visivi offerti da Power BI

Oltre alle visualizzazioni predefinite, Power BI mette a disposizione un AppSource marketplace, raggiungibile direttamente dall'interfaccia tramite il pulsante con i tre puntini in fondo al pannello Visualizzazioni, dove è possibile scaricare oggetti visivi aggiuntivi creati da sviluppatori terzi. L'offerta è in continua crescita e include grafici avanzati, mappe personalizzate, indicatori e molto altro, consentendo di estendere ulteriormente le capacità di analisi e rappresentazione dei dati.

Se nelle fasi di estrazione e trasformazione si svolge il lavoro "dietro le quinte", la dashboard rappresenta il culmine visivo e strategico di questo processo. Una dashboard non è un semplice raccoglitore di oggetti visivi, ma un'interfaccia avanzata che offre una panoramica sintetica e centralizzata delle performance aziendali. Essa si configura come il front-end dell'intero sistema BI; la sua essenza risiede nella capacità di tradurre serie numeriche complesse in una narrazione visiva coerente, rendendo l'analisi dei dati accessibile anche a utenti non

specialisti. Il suo scopo primario è democratizzare l'accesso alle informazioni, permettendo a tutti i livelli dell'organizzazione di allineare le operazioni quotidiane agli obiettivi strategici e di prendere decisioni data-driven, ovvero guidate dall'evidenza empirica.

La distinzione tra una dashboard moderna (come quelle realizzabili in Power BI) e la reportistica tradizionale rappresenta un radicale cambio di approccio nel modo di concepire l'analisi aziendale, come si può osservare nella Tabella 2.1.

Caratteristica	**Report Tradizionale**	**Dashboard di Business Intelligence**
Natura	Statico e retrospettivo.	Dinamico e interattivo.
Scopo	Risponde alla domanda: *"Cosa è successo?"*.	Risponde alle domande: *"Cosa sta succedendo ora?"* e *"Perché?"*.
Dati	Dati aggregati e storici, con un basso livello di granularità.	Dati in tempo reale o ad alta frequenza, consentendo il *drill-down* per analizzare i dettagli.
Fruizione	Lettura passiva.	Esplorazione attiva.

Tabella 2.1: Confronto tra Reportistica Tradizionale e Dashboard di Business Intelligence.

Per essere realmente efficace e portare risultati concreti, una dashboard deve fondarsi su alcuni pilastri fondamentali:

- *Interattività*: l'utente esplora i dati attraverso filtri dinamici e navigazione gerarchica, trasformando l'analisi in un processo di scoperta attiva.
- *Personalizzazione*: la vista può essere configurata in base alle specifiche responsabilità dell'utilizzatore, garantendo che ogni professionista riceva esattamente le informazioni necessarie per il proprio ruolo.
- *Visualizzazione intelligente*: l'impiego di principi di design cognitivo assicura che le informazioni vengano trasmesse con chiarezza, minimizzando il carico mentale e massimizzando la comprensione immediata.
- *Scalabilità*: la struttura deve garantire il mantenimento di prestazioni ottimali indipendentemente dal volume di dati processati o dalla crescita aziendale.

A seconda dello scopo analitico e degli utenti destinatari, le dashboard si differenziano in tre tipologie principali:

- *Dashboard Strategiche*: sono destinate al top management (CEO, CFO, direttori); offrono una visione di altissimo livello sugli Indicatori di Prestazione Chiave (KPI) che misurano la salute complessiva dell'azienda. L'orizzonte temporale è tipicamente mensile o annuale.
- *Dashboard Operative*: sono rivolte a figure che gestiscono le attività quotidiane dell'azienda, come i responsabili di magazzino, gli addetti all'assistenza clienti o i coordinatori della produzione, e monitorano in tempo quasi reale i processi operativi. Il loro fine è garantire il corretto svolgimento delle operazioni, ad esempio controllando il livello delle scorte o il numero di ticket di assistenza aperti, e segnalare tempestivamente eventuali anomalie che richiedono un intervento immediato.
- *Dashboard Analitiche*: sono utilizzate dai Data Analyst, offrono funzionalità avanzate di esplorazione per investigare le cause profonde di un andamento. Permettono di incrociare molteplici variabili per scoprire pattern nascosti e formulare ipotesi strategiche, come l'analisi dettagliata di una campagna di marketing.

2.3.4 Condivisione

Una volta completata la fase di creazione del report su Power BI Desktop, il passaggio successivo per rendere il lavoro fruibile ai decisori aziendali è la pubblicazione nel Power BI Service. Questa fase rappresenta il passaggio dal lavoro individuale alla collaborazione a livello organizzativo. Il flusso ha inizio con il comando "Pubblica", che trasferisce il file locale nel cloud di Microsoft, all'interno di un'area di lavoro (Workspace). Da qui, l'utente può gestire la distribuzione dei contenuti attraverso due modalità, ovvero:

- *La condivisione diretta:* è possibile condividere singoli report o intere dashboard selezionando l'opzione "Condividi". Il sistema permette di generare collegamenti accessibili a chiunque all'interno dell'organizzazione o, per una sicurezza maggiore, limitare l'accesso a utenti specifici tramite il loro indirizzo e-mail.
- *L'integrazione con l'ecosistema Microsoft:* poichè Power BI offre una profonda integrazione con gli strumenti di produttività, i report possono essere condivisi direttamente su Microsoft Teams, incorporati in presentazioni PowerPoint dinamiche o salvati su OneDrive e SharePoint per una gestione collaborativa dei file.

Inoltre, grazie alle app mobili, i report condivisi possono essere consultati e distribuiti anche da smartphone e tablet, garantendo la disponibilità del dato in ogni momento. Un aspetto importante della condivisione è la gestione della sicurezza e della governance del dato. Durante la generazione del collegamento, il proprietario del report può definire permessi granulari, decidendo, ad esempio, se consentire ai destinatari di ricondividere il contenuto con altri colleghi oppure se autorizzarli a utilizzare il set di dati originale per creare report personalizzati. In questo modo si favorisce una cultura del dato diffusa, ma al tempo stesso controllata e coerente con le politiche di accesso dell'organizzazione. Questa architettura assicura che le informazioni corrette raggiungano le persone giuste con il livello di autorizzazione appropriato, trasformando un semplice report statico in uno strumento di collaborazione dinamico e sicuro.

2.4 Linguaggi di Programmazione

L'estrema flessibilità di Power BI è garantita dall'integrazione di diversi linguaggi di programmazione, ognuno specializzato in una fase specifica del ciclo di vita del dato. Mentre alcuni si occupano della preparazione e della pulizia, altri sono progettati per l'analisi avanzata e la creazione di calcoli dinamici.

2.4.1 DAX (Data Analysis Expressions)

Il cuore pulsante della capacità analitica di Power BI è rappresentato dal linguaggio DAX (Data Analysis Expressions). Si tratta di un linguaggio di formule sviluppato da Microsoft, che offre una sintassi potente e intuitiva per eseguire trasformazioni e analisi complesse all'interno dei modelli di dati. Sebbene la sintassi possa apparire simile a quella delle formule di Excel, il DAX è progettato per operare su modelli di dati tabellari relazionali. Esso consente di definire e gestire i calcoli attraverso due strumenti fondamentali:

- *Misure*: sono calcoli valutati dinamicamente in base al contesto del report (filtri e selezioni dell'utente). Sono ideali per aggregazioni che devono adattarsi istantaneamente, come il calcolo del totale delle vendite mensili filtrato per una specifica area geografica.

- *Colonne Calcolate*: vengono calcolate riga per riga e memorizzate come parte del modello dei dati. Sono utilizzate principalmente per aggiungere nuove informazioni permanenti, come categorizzazioni o segmentazioni dei prodotti.

Le prestazioni eccezionali del DAX sono garantite dal motore VertiPaq, un sistema di compressione in-memory che permette di eseguire calcoli su milioni di righe in tempi rapidissimi, ottimizzando l'uso delle risorse computazionali. La Figura 2.5 mostra un esempio di una misura creata per il progetto relativo alla presente tesi, che calcola la percentuale di incidenza del Fentanyl sul totale dei decessi. La formula sfrutta la funzione CALCULATE per filtrare i decessi legati a una sostanza specifica, ALL per rimuovere i filtri attivi e ottenere il totale generale, e DIVIDE per restituire il rapporto in forma percentuale. Questo esempio illustra come DAX consenta di costruire calcoli articolati e contestuali in modo leggibile e strutturato.

```
1 % Incidenza Fentanyl =
2 VAR DecessiFentanyl = CALCULATE(DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
  'Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza] = "Fentanyl")
3 VAR TotaleDecessiGenerale = CALCULATE(DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]), ALL
  ('Accidental_Drug_Related_Deaths_2012-2021.'))
4 RETURN
5 DIVIDE(DecessiFentanyl, TotaleDecessiGenerale)
```

Figura 2.5: Esempio di misura DAX

2.4.2 M (Power Query)

Se DAX è il linguaggio dell'analisi, M è il linguaggio dell'acquisizione e della trasformazione. Conosciuto formalmente come Power Query Formula Language, "M" è il motore dietro l'interfaccia grafica di Power Query. La sua funzione principale è il cosiddetto Data Mashup, ovvero la capacità di filtrare, combinare e unire dati provenienti da una o più sorgenti in un unico flusso coerente. Ogni azione compiuta nell'interfaccia utente viene tradotta automaticamente in uno script M, rendendo il processo di preparazione dei dati completamente ripetibile e automatizzato su diverse piattaforme Microsoft, tra cui Excel e Dataverse.

2.4.3 Phyton e R

Per scenari di analisi ancora più avanzati, Power BI permette l'integrazione nativa dei due linguaggi leader nel mondo della Data Science, ovvero Python e R. Questa integrazione estende le capacità della piattaforma permettendo:

- *La manipolazione avanzata (ETL);* questa consente di pulire e modellare dati complessi che richiederebbero logiche troppo onerose per i soli linguaggi nativi.
- *L'analisi statistica e predittiva;* questa permette di utilizzare R per statistiche raffinate e Python per applicazioni di Machine Learning e Data Science generale.
- *Le visualizzazioni personalizzate;* queste consentono di creare grafici avanzati, offrendo una libertà espressiva totale nella rappresentazione degli insight.

2.5 Vantaggi di Power BI

L'adozione di Power BI all'interno di un'organizzazione non si limita alla semplice visualizzazione dei dati, ma porta con sé una serie di vantaggi strategici che trasformano il

modo in cui l'azienda gestisce e interpreta il proprio patrimonio informativo. Tra i principali punti di forza spiccano la sicurezza, la capacità di integrazione e le recenti innovazioni basate sull'Intelligenza Artificiale.

2.5.1 Robuste funzioni di sicurezza

La sicurezza e la privacy dei dati rappresentano una priorità assoluta per qualsiasi infrastruttura IT moderna. Power BI risponde a questa esigenza offrendo un ambiente altamente protetto e conforme ai più rigorosi standard internazionali, come certificato dal Microsoft Trust Center. La piattaforma permette agli amministratori di esercitare un controllo granulare sugli accessi.

Una delle caratteristiche più sicure dell'architettura di Power BI è la separazione tra la fruizione del report e l'accesso alla sorgente dati: quando un report o una dashboard viene condivisa, il destinatario ottiene l'autorizzazione a visualizzare le metriche elaborate, ma non necessita di credenziali dirette per accedere al database originale. Il sistema verifica l'identità dell'utente usando le sue credenziali aziendali, così da capire a quali dati può accedere. In genere, chi apre un report può vedere solo le informazioni per cui è stato autorizzato, senza entrare direttamente nella sorgente dati originale. Nelle connessioni in tempo reale, come quelle verso SQL Server Analysis Services (SSAS) tramite gateway locale, questo controllo diventa ancora più rigoroso: ogni richiesta viene verificata per il singolo utente e l'accesso è consentito solo se quest'ultimo dispone dei permessi necessari sul modello semantico.

2.5.2 Versatilità nell'integrazione delle fonti

Un ulteriore vantaggio competitivo di Power BI è la sua estrema flessibilità nell'acquisizione dei dati. La piattaforma è progettata per abbattere i cosiddetti "silos aziendali", permettendo di combinare informazioni eterogenee in un'unica visione d'insieme. L'integrazione nativa con l'intero ecosistema Microsoft (inclusi Excel, Azure e SharePoint) facilita l'importazione diretta di fogli di calcolo, la connessione a database cloud per analisi avanzate e la distribuzione dei risultati. Oltre all'ambiente Microsoft, l'ampia libreria di connettori supporta il collegamento a database relazionali di terze parti, servizi web e file locali. Questa versatilità garantisce che l'azienda non debba stravolgere la propria infrastruttura esistente, potendo, invece, collegare Power BI a qualsiasi sistema già in uso per estrarne valore immediato.

2.5.3 Funzionalità di AI integrata

Il panorama della Business Intelligence è stato recentemente rivoluzionato dall'introduzione dell'Intelligenza Artificiale generativa. In Power BI, questo si traduce nell'integrazione di Copilot in Microsoft Fabric, un assistente virtuale che migliora e accelera l'esperienza di analisi sia per gli sviluppatori che per gli utenti finali. A differenza di strumenti tradizionali, l'architettura di Copilot si adatta in modo dinamico all'esperienza richiesta dall'utente, seguendo un flusso di lavoro sofisticato descritto nel dettaglio nella Figura 2.6.

Come si può osservare dal diagramma, il sistema opera attraverso quattro fasi principali, che sono:

- *Input:* l'interazione può avvenire tramite prompt testuali (nel riquadro chat per interrogare i dati) o tramite azioni dirette (pulsanti per generare automaticamente descrizioni delle misure nel modello).
- *Pre-elaborazione e Grounding:* Copilot recupera i dati di base necessari. Se l'utente pone una domanda, il sistema di AI consulta lo schema del modello semantico; se quest'ul-

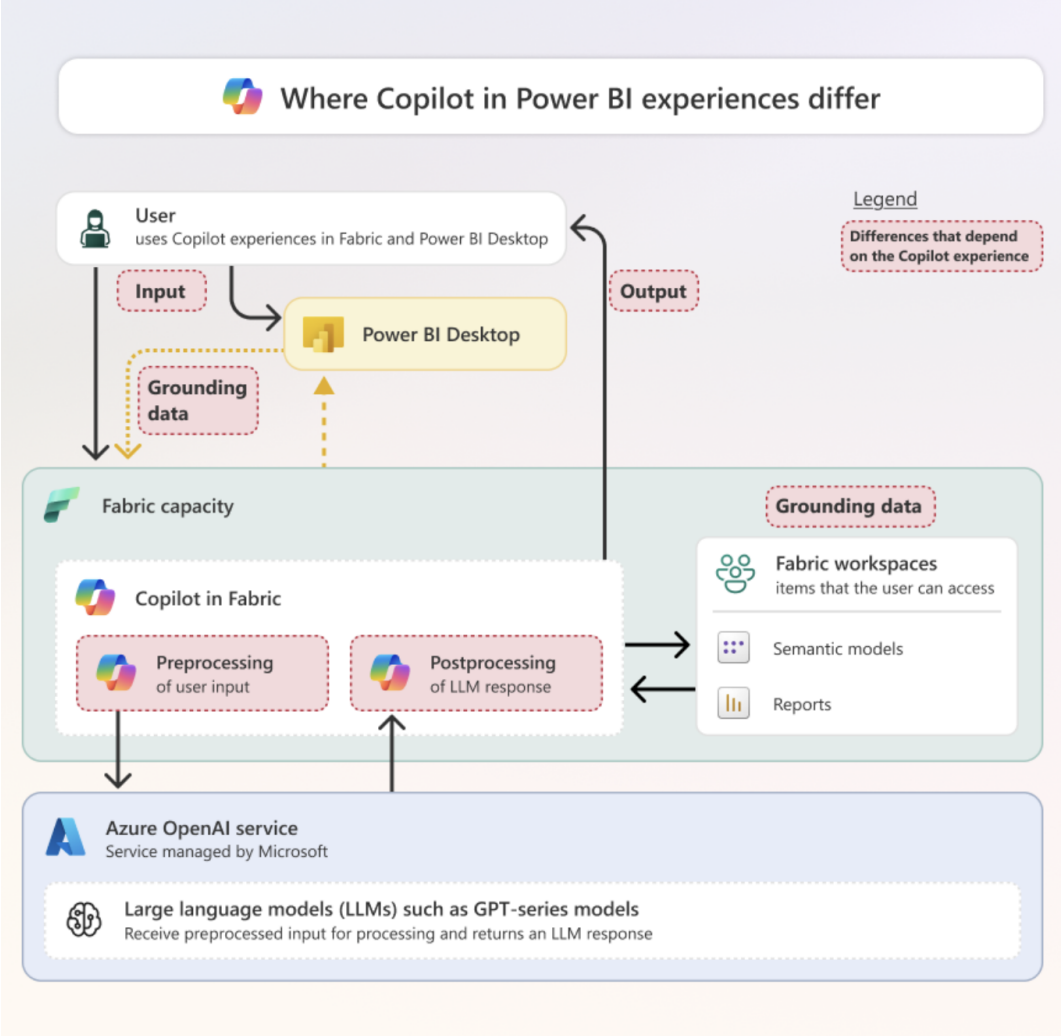


Figura 2.6: Architettura del flusso di lavoro di Copilot in Power BI

timo richiede un riassunto, Copilot analizza i metadati del report. Gli sviluppatori mantengono il controllo nascondendo i campi sensibili all'AI.

- *Post-elaborazione:* le risposte fornite dal Large Language Model (LLM) vengono validate. Ad esempio, una query DAX generata viene passata attraverso un parser per garantirne l'eseguibilità e correggere eventuali errori del modello.
- *Output:* il risultato finale si adatta alla richiesta, producendo spiegazioni in linguaggio naturale, codice DAX o intere nuove pagine di report coerenti con il tema aziendale.

L'introduzione di funzionalità di AI così avanzate richiede, tuttavia, una solida "cultura del dato". Per ottenere un reale vantaggio competitivo, l'uso di Copilot deve essere accompagnato da un'adeguata formazione e da una governance precisa, assicurando che l'automazione supporti efficacemente le decisioni umane.

CAPITOLO 3

Descrizione dei dataset di partenza

In questo capitolo vengono descritti e analizzati i dataset su cui si fonda l'intera ricerca sulla crisi degli oppioidi negli Stati Uniti. Si introduce dapprima il contesto necessario alla corretta interpretazione dei dati, cioè la natura farmacologica degli oppioidi, classificati per potenza e pericolosità, e le dinamiche storiche che hanno generato l'epidemia, dalla sovra-prescrizione degli anni Novanta fino all'esplosione degli oppiodi sintetici come il fentanyl. Vengono, poi, descritti i due dataset impiegati nell'analisi — il registro granulare dei decessi accidentali nello Stato del Connecticut (2012–2021) e il dataset aggregato a livello federale sulle overdose da oppioidi (1999–2014) — illustrandone la struttura, le variabili principali e la complementarità tra un'indagine microscopica sul singolo caso e una visione macroscopica del trend nazionale. Il capitolo affronta, quindi, l'analisi della qualità dei dati grezzi, mappando le criticità strutturali riscontrate, che sono la disomogeneità temporale e geografica tra le due fonti, la presenza diffusa di valori mancanti nei campi demografici e tossicologici, e le incoerenze nei formati di date, numeri e stringhe. Si documentano infine tutte le operazioni di pulizia e trasformazione condotte in Power Query, dalla gestione dei valori nulli e dall'operazione di unpivot delle colonne tossicologiche, alla normalizzazione dei formati, fino alla creazione di attributi derivati, come le coordinate geografiche e le macro-categorie dei luoghi di decesso, che hanno permesso di ottenere un modello relazionale coeso e pronto per la fase di visualizzazione in Power BI.

3.1 Introduzione ai dati

L'analisi proposta in questo elaborato si basa sulla convinzione che i dati, se correttamente interpretati, rappresentino uno strumento fondamentale per comprendere e contrastare fenomeni sociali complessi. Il tema centrale della ricerca è la crisi degli oppioidi, un'emergenza sanitaria che negli ultimi decenni ha trasformato radicalmente il panorama della salute pubblica negli Stati Uniti. Per affrontare questo studio attraverso le metodologie della Business Intelligence, la scelta è ricaduta su dataset di natura pubblica (Open Data) che consentono di osservare il problema da due angolazioni diverse e complementari, che sono:

- *Il contesto nazionale (USA)*: un'analisi macroscopica per identificare le correlazioni tra le politiche di prescrizione medica e l'aumento sistematico dei decessi a livello federale.
- *Il focus territoriale (Connecticut)*: un'indagine microscopica e granulare che, attraverso i dati sui singoli decessi accidentali, permette di mappare l'evoluzione delle sostanze coinvolte e il profilo demografico delle vittime.

Per interpretare correttamente i dati e le visualizzazioni prodotte nelle dashboard, è tuttavia necessario inquadrare la natura degli oppioidi e le dinamiche storiche che hanno generato l'epidemia.

3.1.1 Gli oppioidi

Gli oppioidi sono sostanze chimiche psicoattive ampiamente utilizzate nella medicina moderna come potenti analgesici per il trattamento del dolore acuto e cronico, in quanto capaci di produrre effetti farmacologici assimilabili a quelli della morfina. Queste molecole agiscono legandosi a specifici recettori nel sistema nervoso centrale, inibendo la percezione della sofferenza e, contestualmente, stimolando il sistema dopaminergico (complesso circuito neuronale che utilizza la dopamina come neurotrasmettitore per regolare funzioni chiave come motivazione, ricompensa, piacere, controllo motorio e funzioni cognitive). Questa stimolazione scatena intense sensazioni di euforia che portano il cervello a desiderare la sostanza, inducendo progressivamente l'organismo ad assuefarsi e spingendo l'assuntore a ricercare dosi sempre maggiori per ottenere lo stesso effetto. Da un punto di vista farmacologico e tossicologico, gli oppioidi si dividono in tre macro-categorie:

- *Oppioidi naturali* (oppiacei): derivano direttamente dal papavero da oppio e includono sostanze come la morfina, considerata il punto di riferimento clinico per misurare l'efficacia degli altri analgesici, e la codeina.
- *Oppioidi semisintetici*: sono composti ottenuti attraverso la modifica chimica degli oppiacei naturali al fine di potenziarne l'effetto. Rientrano in questa categoria l'ossicodone (ampiamente diffuso nella pratica clinica statunitense e circa una volta e mezza più potente della morfina) e l'eroina, divenuta una droga altamente pericolosa e diffusa nel mercato nero.
- *Oppioidi sintetici*: sono prodotti interamente in laboratorio e progettati per raggiungere potenze estreme. L'esponente più critico di questa categoria è il fentanyl, un farmaco da 50 a 100 volte più potente della morfina, oggi responsabile di gran parte dei decessi a causa del suo altissimo rischio di indurre una rapida e fatale depressione respiratoria.

Per comprendere appieno l'escalation di pericolosità delle diverse sostanze che verranno analizzate nei dataset, la Tabella 3.1 propone una classificazione dei principali oppioidi. La loro efficacia e tossicità sono espresse in relazione alla potenza della morfina, considerata il benchmark (parametro di riferimento oggettivo) clinico standard.

3.1.2 La genesi e lo sviluppo della crisi sanitaria negli Stati Uniti

La crisi sanitaria che attraversa gli Stati Uniti affonda le sue radici nella seconda metà degli anni '90, quando le grandi case farmaceutiche iniziarono a commercializzare in quantità considerevole potenti analgesici, come l'ossicodone, promuovendoli in modo ingannevole come soluzioni sicure e prive di rischi di dipendenza. Questa pratica generò una prima ondata di assuefazione su scala nazionale, intrappolando milioni di americani in una spirale di dipendenza nata da prescrizioni mediche del tutto legali. A partire dal 2013, il fenomeno ha subito un drammatico incremento con l'arrivo sul mercato clandestino degli oppioidi sintetici. Il fentanyl ha iniziato a essere miscelato a insaputa degli acquirenti con eroina o cocaina, innescando un'impennata di overdose. Secondo i dati dei Centers for Disease Control and Prevention (CDC), oggi il fentanyl è coinvolto in oltre il 70% dei decessi per overdose da oppioidi negli USA. Complessivamente, dagli anni '90 a oggi, la crisi ha stroncato più di 500.000 vite, raggiungendo la cifra record di oltre 81.000 vittime nel solo 2022.

Oppioide	Potenza vs Morfina	Note
1. Tramadolo	$0.1\times$	Oppioide debole, usato per dolori lievi
2. Codeina	$0.1\times$	Basso rischio di dipendenza, effetto blando
3. Tapentadolo	$0.4\times$	Via di mezzo, efficace per dolore moderato
4. Idrocodone	$1\times$	Equivalente alla morfina
5. Morfina	$1\times$	Riferimento standard per il confronto
6. Ossicodone	$1.5\times$	Più efficace della morfina, alto rischio di abuso
7. Eroina	$2-3\times$	Effetto intenso e rapido, legale solo in pochi Paesi
8. Metadone	$3\times$	Usato per dolore cronico e dipendenza
9. Idromorfone	$5\times$	Più efficace della morfina, usato in casi gravi
10. Buprenorfina	$25-100\times$	Molto potente ma agonista parziale, utile nella terapia sostitutiva
11. Fentanyl	$50-100\times$	Usato in anestesia e nel dolore grave da cancro, altissimo rischio se abusato

Tabella 3.1: Classificazione degli oppioidi per potenza relativa rispetto alla morfina.

3.2 Struttura dei dataset

Di seguito vengono descritti i due dataset impiegati nell'analisi. Il primo, relativo allo Stato del Connecticut, offre un quadro dettagliato di ogni singolo decesso accidentale legato all'abuso di sostanze. Il secondo, di portata nazionale, aggrega i dati per anno e per Stato, permettendo di osservare l'andamento generale della mortalità da oppioidi negli Stati Uniti nel corso di un quindicennio.

3.2.1 Accidental Drug Related Deaths (2012-2021)

Questo dataset raccoglie ogni decesso accidentale per overdose da sostanze stupefacenti registrato nel Connecticut tra il 2012 e il 2021, per un totale di 9.202 casi. I dati sono il risultato delle indagini condotte dall'Office of the Chief Medical Examiner del Connecticut, che integra il referto tossicologico, il certificato di morte e l'ispezione della scena del decesso. Ogni record rappresenta un singolo individuo deceduto, descritto attraverso 48 variabili suddivise in quattro aree principali:

- *Area anagrafica:* data del decesso, età, sesso e origine etnica della vittima.
- *Area geografica:* luogo di residenza, luogo dell'overdose e luogo del decesso, ciascuno declinato in città, contea e stato, con relative coordinate geografiche per il mapping territoriale.
- *Area circostanziale:* tipologia del luogo (ad esempio abitazione, strada, veicolo), causa ufficiale e modalità del decesso certificati dal medico legale, descrizione narrativa della dinamica ed eventuali patologie preesistenti.
- *Area tossicologica:* una serie di campi flag (valore "Y" o vuoto) indica le sostanze rilevate nel referto tossicologico, tra cui eroina, fentanyl, cocaina, ossicodone, benzodiazepine, metadone e molte altre. Il campo "Any Opioid" funge da indicatore aggregato nei casi in cui il medico legale non possa distinguere con certezza la sostanza oppioide di origine.

3.2.2 US Opioid Overdose Deaths (1999-2014)

Questo dataset, prodotto dal National Center for Health Statistics dei Centers for Disease Control and Prevention, offre una visione aggregata a livello nazionale del fenomeno. Il dataset copre il periodo 1999-2014, coincidendo con la prima e seconda ondata della crisi degli oppioidi. Tra il 2000 e il 2014, il tasso di overdose negli USA è cresciuto del 137%, con un incremento del 200% per i soli oppioidi. Ogni riga del dataset corrisponde a una specifica combinazione tra anno e Stato e riporta, per quella unità di osservazione, le principali informazioni relative alla mortalità da oppioidi e al volume delle prescrizioni. Le 9 variabili presenti possono essere descritte come segue:

- *Index*: identifica univocamente il record all'interno del dataset.
- *State*: indica lo Stato federale statunitense a cui il dato si riferisce e consente di confrontare l'andamento del fenomeno tra le diverse aree del Paese.
- *Year*: rappresenta l'anno di riferimento dell'osservazione e permette di seguire l'evoluzione temporale della crisi nel periodo analizzato.
- *Deaths*: indica il numero assoluto di decessi per overdose da oppioidi registrati in quello Stato e in quell'anno.
- *Population*: riporta la popolazione dello Stato nell'anno considerato, utile per contestualizzare il dato assoluto dei decessi.
- *Crude Rate, Crude Rate Lower, Crude Rate Upper*: esprimono, rispettivamente, il tasso grezzo di mortalità per 100.000 abitanti e i limiti inferiore e superiore del relativo intervallo di confidenza al 95%.
- *Prescriptions*: rappresenta il numero di prescrizioni di oppioidi dispensate dai retailer statunitensi nell'anno, espresso in milioni. Questa variabile è particolarmente rilevante perché mette in relazione il volume dei farmaci immessi nel mercato legale con l'andamento della mortalità, rendendo possibile un'analisi della prima fase della crisi, strettamente legata alla sovraprescrizione medica.

3.3 Analisi della qualità dei dati

La fase di estrazione dei dati grezzi, per quanto supportata da fonti autorevoli come Kaggle, rappresenta solo il punto di partenza di un progetto di Business Intelligence. I dataset del mondo reale sono intrinsecamente imperfetti; infatti, presentano incoerenze e lacune che, se non gestite in fase preventiva, rischiano di compromettere l'affidabilità delle dashboard finali e delle logiche decisionali che ne derivano. L'analisi della qualità dei dati è, quindi, un passaggio metodologico obbligato per mappare le criticità strutturali e pianificare le successive operazioni di pulizia e trasformazione. Nel contesto di questa indagine, l'esame incrociato del dataset nazionale sulle prescrizioni e di quello statale del Connecticut ha fatto emergere diverse categorie di anomalie, tipiche dei flussi informativi socio-sanitari.

3.3.1 Disomogeneità temporale e geografica

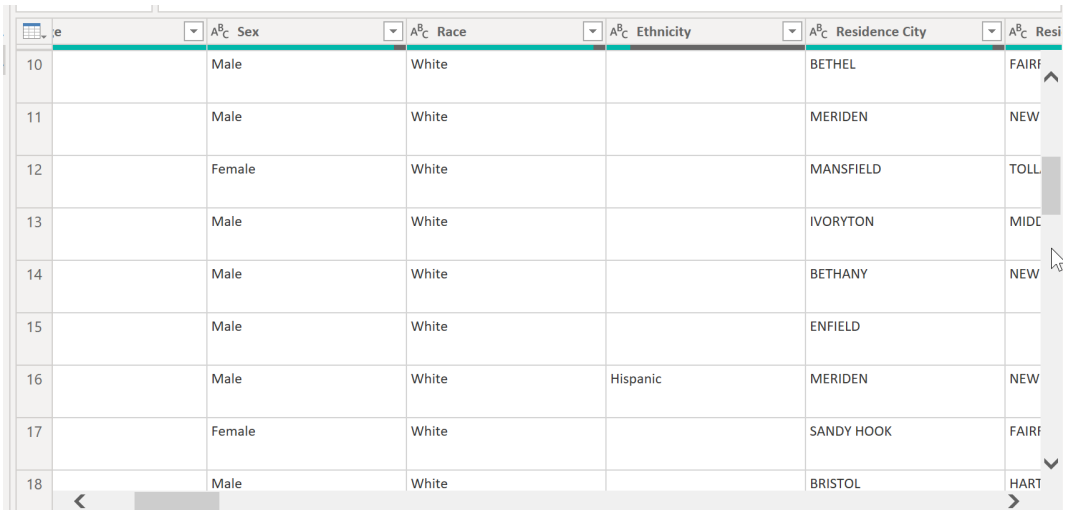
Un primo aspetto da considerare riguarda il diverso livello di dettaglio con cui i due dataset descrivono il fenomeno, sia dal punto di vista temporale che da quello geografico. Poiché si tratta di due fonti costruite con finalità differenti, è naturale che presentino strutture informative non perfettamente omogenee. Il dataset nazionale offre una rappresentazione

aggregata su base annuale, utile per individuare i macro-trend di lungo periodo nelle prescrizioni e nella mortalità. Il dataset locale del Connecticut, invece, registra i singoli eventi e riporta la data esatta del decesso, come, ad esempio "05/29/2012", consentendo un'analisi molto più granulare del fenomeno. Sul piano della disomogeneità geografica, l'analisi a livello locale nel dataset del Connecticut presenta sfide legate alla moltitudine di coordinate spaziali disponibili per un singolo evento. I record contengono colonne separate per il luogo di residenza, il luogo dell'incidente e il luogo del decesso. Queste informazioni spesso non coincidono, e la mole di valori vuoti sparsi tra queste colonne costringerà, in fase di modellazione, a definire una chiara regola logica per determinare quale coordinata geografica utilizzare per la costruzione delle mappe.

3.3.2 Analisi dei valori mancanti e incompleti

La presenza di valori nulli rappresenta un ostacolo primario, in modo particolare nel referto autoptico del Connecticut. Su oltre 10.000 record estratti, migliaia di righe presentano gravi lacune. Le assenze di dati si concentrano soprattutto in due ambiti:

- *Informazioni socio-demografiche e geografiche:* diverse colonne anagrafiche e territoriali risultano solo parzialmente compilate. Ad esempio, il campo Ethnicity (Figura 3.1) e alcune variabili geografiche, come Injury County o Residence State, presentano un numero elevato di valori mancanti.



	e	Sex	Race	Ethnicity	Residence City	Resi
10		Male	White		BETHEL	FAIRF
11		Male	White		MERIDEN	NEW
12		Female	White		MANSFIELD	TOLL
13		Male	White		IVORYTON	MIDD
14		Male	White		BETHANY	NEW
15		Male	White		ENFIELD	
16		Male	White	Hispanic	MERIDEN	NEW
17		Female	White		SANDY HOOK	FAIRF
18		Male	White		BRISTOL	HART

Figura 3.1: Valori mancanti nel campo "Ethnicity"

- *Informazioni tossicologiche:* le colonne binarie dedicate alle singole sostanze non seguono una codifica esplicita del tipo "Sì/No". Quando una sostanza viene rilevata compare la lettera "Y" (Figura 3.2), mentre nei casi in cui non viene rilevata la cella resta semplicemente vuota.

3.3.3 Criticità nei formati delle stringhe e delle date

Un ulteriore aspetto da considerare riguarda il formato con cui alcuni valori vengono importati all'interno del software di Business Intelligence. In diversi casi, infatti, il problema non risiede nel contenuto del dato ma nel modo in cui esso viene interpretato dalla piattaforma. Tra le situazioni più rilevanti si possono distinguere le seguenti:

- *Valori numerici letti come testo:* nel dataset nazionale, colonne come Deaths, Crude Rate e i relativi intervalli di confidenza vengono spesso importate come stringhe testuali

Heroin	Heroin death certificate (DC)	Cocaine
		Y
Y		
Y		
Y		
		Y
		Y
Y		
Y		

Figura 3.2: Flag tossicologici

anziché come valori numerici (Figura 3.3). Ciò dipende dalla presenza di diciture come "Suppressed" o "Unreliable", introdotte per ragioni di privacy statistica o di affidabilità del dato. Finché queste voci non vengono isolate, convertite o sostituite con valori nulli, non è possibile eseguire correttamente operazioni matematiche, come somme, medie o confronti quantitativi.

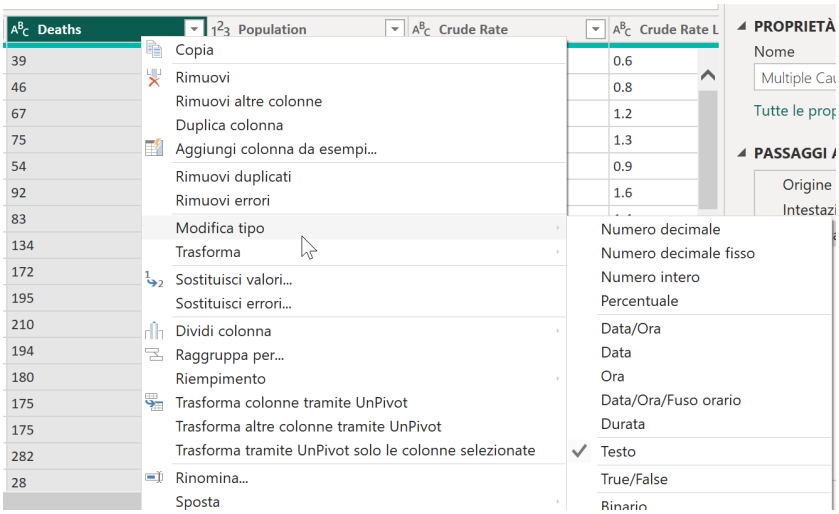


Figura 3.3: Valori numeri interpretati come "testo"

- *Date interpretate come testo:* nel dataset del Connecticut, la colonna "Date" è registrata nel formato americano MM/DD/YYYY, ma inizialmente viene letta come semplice testo anziché come data vera e propria (Figura 3.4). Senza una conversione esplicita nel formato corretto, la piattaforma non riesce a sfruttare le funzionalità di Time Intelligence, rendendo più difficile analizzare l'evoluzione del fenomeno per anni, trimestri e mesi.
- *Variabilità nelle descrizioni testuali:* nel dataset del Connecticut alcune colonne presentano valori inseriti manualmente, e quindi non del tutto uniformi. La colonna Fentanyl, ad esempio, non contiene soltanto il flag "Y", ma anche varianti come "Y POPS" o "Y

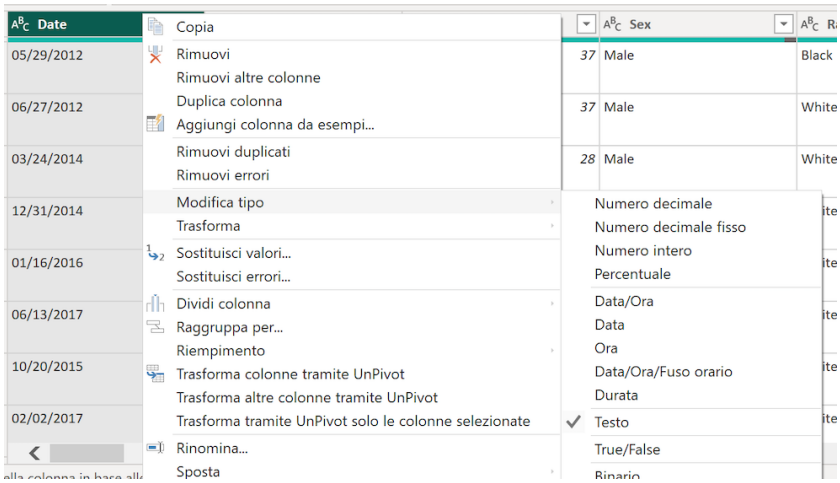


Figura 3.4: Colonna "Date" riconosciuta come testo

(PTCH)". Se queste diciture non vengono ricondotte a un unico valore standard, il software rischia di interpretarle come categorie differenti, compromettendo la correttezza delle analisi.

3.4 Pulizia dei dati

A seguito dell'individuazione delle criticità strutturali, è stata avviata la fase di pulizia e trasformazione dei dati (Data Cleaning ed ETL). Utilizzando l'ambiente di Power Query, i dataset grezzi sono stati sottoposti a una serie di passaggi sequenziali (Figura 3.5) volti a correggere le anomalie, standardizzare i valori e preparare le tabelle per la modellazione in Power BI. L'obiettivo primario di questa fase è stato quello di trasformare flussi informativi eterogenei in un modello relazionale coeso.

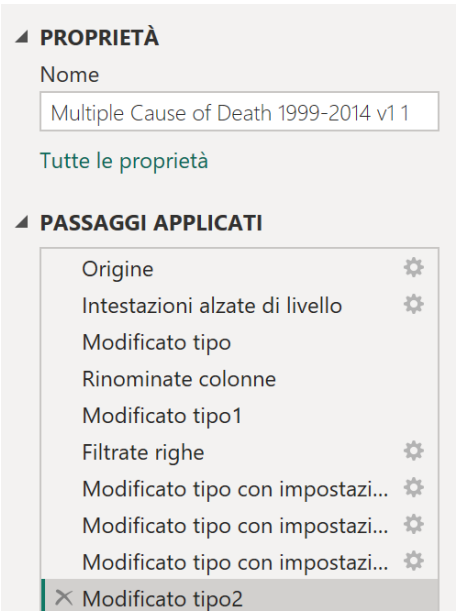


Figura 3.5: Esempio dei passaggi di trasformazione applicati in Power Query

3.4.1 Gestione dei valori nulli

La presenza di dati mancanti è un aspetto molto comune quando si lavora con dataset reali. Anche nel caso di questa analisi, è stato necessario prestare particolare attenzione ai valori assenti o incompleti, perché una loro gestione non corretta avrebbe potuto compromettere la lettura dei risultati e l'affidabilità delle dashboard finali. Per questo motivo, prima di procedere con la modellazione, è stato indispensabile individuare le principali lacune informative e definire le operazioni più adatte per trattarle in modo coerente. In particolare, il lavoro si è sviluppato nei seguenti passaggi:

- *Filtraggio dei dati invalidi:* nel dataset nazionale (US Opioid Overdose Deaths (1999-2014)) sono state escluse tutte le righe in cui il tasso di mortalità (Crude Rate) presentava le diciture "Suppressed" o "Unreliable", permettendo così di trattare la colonna come puramente matematica. Nel dataset locale (Accidental Drug Related Deaths (2012-2021)) sono state rimosse le righe anagrafiche prive di informazioni cardine, scartando i decessi privi della coordinata spaziale o segnati come "UNKNOWN".
- *Pulizia verticale:* sono state eliminate intere colonne ritenute ridondanti o caratterizzate da un'eccessiva percentuale di valori nulli (come Ethnicity, Fentanyl Analogue e Heroin death certificate).
- *Trasformazione Unpivot:* per rendere più leggibile e utilizzabile la parte tossicologica del dataset, la struttura originaria è stata trasformata da orizzontale (Figura 3.6) a verticale (Figura 3.7). Le diverse sostanze sono state raccolte in un'unica colonna denominata "Sostanza", mantenendo solo i casi con esito positivo ("Y"). In questo modo è stato possibile semplificare il dataset e superare il problema dei numerosi valori vuoti presenti nelle colonne tossicologiche.

Heroin	Cocaine	Fentanyl	Fentanyl Analogue	Oxycodone
	Y			
Y				
Y				
Y				
		Y		
	Y			
				Y

Figura 3.6: Struttura dei dati prima della trasformazione Unpivot

3.4.2 Normalizzazione dei formati data e testo

La seconda fase del lavoro ha riguardato la normalizzazione dei formati, la conversione dei tipi di dato e la traduzione di alcune informazioni, così da rendere il dataset più coerente e più semplice da analizzare in un contesto italiano. In questa fase, l'obiettivo non era soltanto correggere i formati, ma anche uniformare il linguaggio e preparare i dati alle successive visualizzazioni. Le principali operazioni svolte sono state le seguenti:

- *Allineamento temporale e numerico:* le colonne contenenti le date sono state convertite nel formato corretto utilizzando le impostazioni locali americane (Figura 3.8), necessarie

Sostanza	Presenza
Eroina	Y
Eroina	Y
Cocaina	Y
Ossicodone	Y

Figura 3.7: Struttura dei dati dopo la trasformazione Unpivot

per interpretare correttamente la struttura MM/DD/YYYY. Allo stesso tempo, i tassi di mortalità sono stati trasformati in valori numerici decimali, così da poter essere utilizzati senza errori nelle elaborazioni successive.

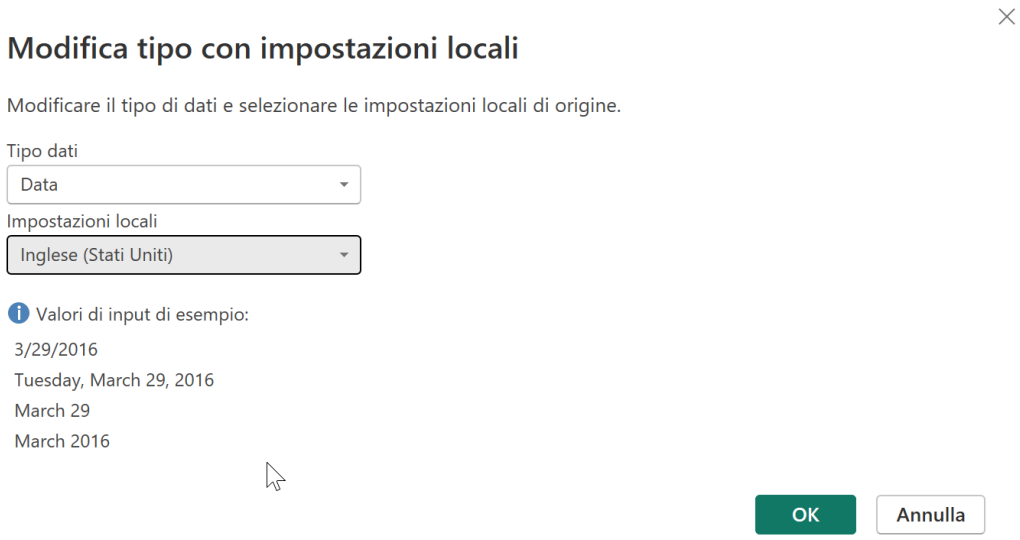


Figura 3.8: Modifica del tipo di formato con le impostazioni locali

- *Standardizzazione geografica:* le sigle statali e alcune denominazioni territoriali sono state uniformate per evitare incoerenze nella lettura geografica dei dati. Ad esempio, la sigla "CT" è stata convertita in "CONNECTICUT" e alcune differenze toponomastiche minori, come "EAST HARTFORD", sono state ricondotte a una forma più omogenea.
- *Traduzione e pulizia testuale:* è stata eseguita un'attività di sostituzione testuale per rendere più comprensibili le categorie socio-demografiche e i nomi delle sostanze. In questo modo, etichette come "Heroin", "Meth/Amphetamine" e "Any Opioid" sono state tradotte rispettivamente in "Eroina", "Meta/Anfetamina" e "Qualsiasi Oppioide".

3.4.3 Creazione di attributi derivati e aggregazioni

L'ultima fase del processo di ETL è stata dedicata alla creazione di nuovi attributi utili ad arricchire il modello e a supportare meglio l'analisi. A partire dai dati originali, sono quindi state ricavate alcune informazioni aggiuntive che hanno reso più efficace la costruzione delle

visualizzazioni e delle dimensioni di analisi. Le principali elaborazioni hanno riguardato i seguenti aspetti:

- *Estrazione della dimensione temporale:* dalla data completa dell'incidente è stata ricavata la colonna "Anno", necessaria per abilitare le analisi temporali e costruire grafici di andamento nel tempo.
- *Parsing delle coordinate spaziali:* a partire dalla colonna grezza DeathCityGeo, il testo contenuto tra parentesi è stato separato in due valori numerici distinti, corrispondenti a "LATITUDINE" e "LONGITUDINE". Questo passaggio è stato essenziale per poter utilizzare correttamente le visualizzazioni su mappa.
- *Aggregazione semantica dei luoghi:* la colonna relativa al luogo di ritrovamento presentava moltissime descrizioni eterogenee, come ad esempio "car", "Hilton" o "park". Attraverso una serie di script condizionali basati sulla ricerca di parole chiave in Figura 3.9, queste voci sono state ricondotte a sei categorie più leggibili e omogenee: "Casa", "Ospedale", "Hotel & Motel", "Veicolo", "Spazi Pubblici all'aperto" e "Aree Commerciali".

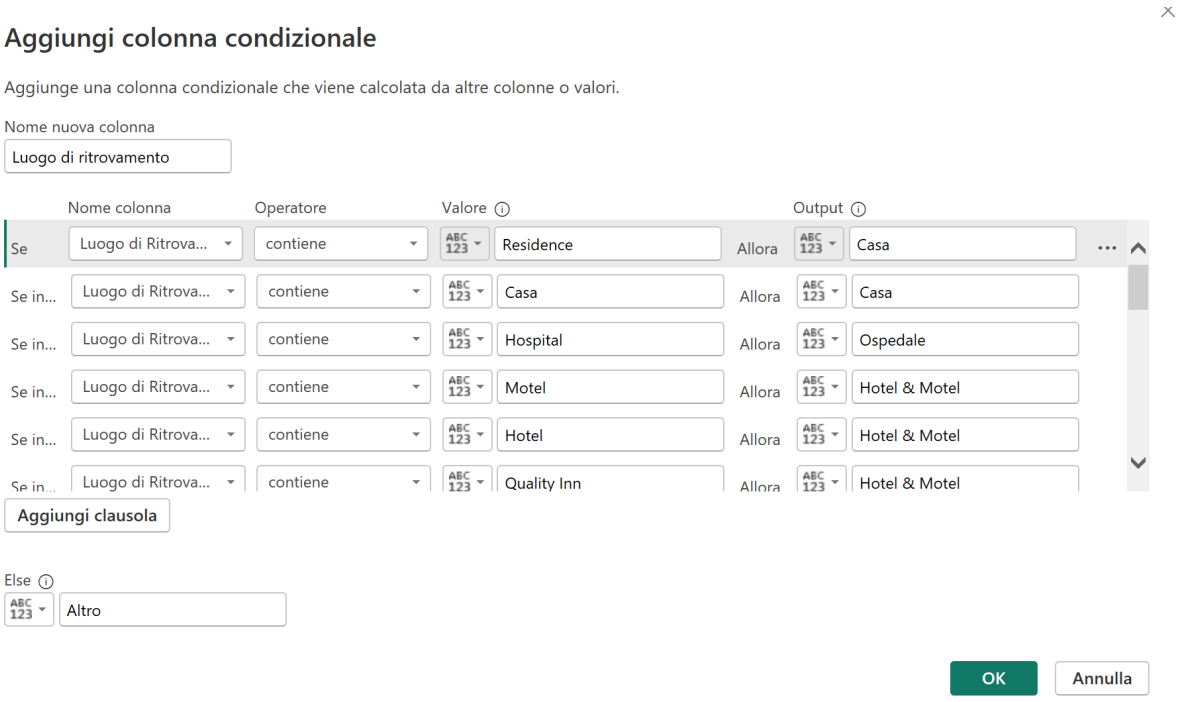


Figura 3.9: Creazione della colonna condizionale "Luogo di ritrovamento"

CAPITOLO 4

Progettazione e implementazione delle dashboard di base

In questo capitolo vengono descritte la progettazione e l'implementazione delle prime tre dashboard del progetto, dedicate all'esplorazione introduttiva della crisi degli oppioidi negli Stati Uniti. Si illustrano, dapprima, le misure DAX sviluppate per alimentare dinamicamente le visualizzazioni, dal conteggio dei decessi all'incidenza del Fentanyl, dalla variazione percentuale annua all'identificazione del genere dominante, nonchè e le colonne calcolate create per arricchire il dataset con attributi derivati quali la mobilità territoriale delle vittime e il numero di sostanze per caso. Si descrive, quindi, l'architettura narrativa delle dashboard, strutturata secondo una logica Top-Down che guida l'utente da una visione macroscopica nazionale fino all'analisi granulare del Connecticut. La prima dashboard - "Dalle prescrizioni all'emergenza: 15 anni di crisi sanitaria" - inquadra il fenomeno su scala federale, correlando il volume delle prescrizioni mediche con l'escalation della mortalità nel periodo 1999–2014. La seconda — "Focus territoriale: l'emergenza sanitaria sul Connecticut (1999-2014)" — isola le specificità dello Stato, confrontandone il tasso di mortalità con la media nazionale e analizzando la relazione diretta tra offerta farmaceutica e decessi. La terza — "Analisi dei decessi accidentali nel Connecticut (2012-2021)" — sposta il focus sul profilo sociodemografico delle vittime, esaminando la distribuzione per età, genere ed etnia nel decennio 2012–2021.

4.1 Misure DAX

Le implementazioni in linguaggio DAX (Data Analysis Expressions) realizzate all'interno del progetto sono state create principalmente per visualizzare in modo chiaro le informazioni all'interno delle dashboard. Queste possono essere suddivise in due macro-categorie essenziali: le misure, utilizzate per calcolare valori dinamici nei grafici e negli indicatori, e le colonne calcolate, impiegate per aggiungere nuove informazioni al dataset e facilitare la categorizzazione e il filtraggio dei dati.

4.1.1 Descrizione delle misure DAX implementate

Le misure DAX sono state utilizzate per calcolare e mostrare nelle dashboard valori aggiornati in modo dinamico. A differenza delle colonne, queste formule non occupano memoria fisica nel dataset, ma vengono calcolate "al momento" in base ai filtri selezionati dall'utente. Sono utili per eseguire aggregazioni, come somme, medie e percentuali, per effettuare confronti nel tempo e per aggiornare automaticamente i KPI quando si interagisce con i filtri della dashboard.

Nello specifico, per rispondere agli obiettivi di analisi del progetto, sono state sviluppate le seguenti misure:

- *Decessi*: rappresenta la misura cardine del progetto. Utilizza `DISTINCTCOUNT` sull'identificativo univoco del caso per evitare sovrapposizioni. L'operatore +0 serve a forzare la visualizzazione dello zero nei grafici anche in assenza di dati (Figura 4.1).

```
Decessi = DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice])+0
```

Figura 4.1: Decessi

- *Incidenza Fentanyl*: rappresenta l'impatto del Fentanyl sul totale dei decessi. La logica `ALL` applicata al denominatore permette di mantenere fisso il totale generale come termine di paragone (Figura 4.2).

```
% Incidenza Fentanyl =
VAR DecessiFentanyl = CALCULATE(DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.
'[Indice]), 'Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza] = "Fentanyl")
VAR TotaleDecessiGenerale = CALCULATE(DISTINCTCOUNT
('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]), ALL
('Accidental_Drug_Related_Deaths_2012-2021.'))
RETURN
DIVIDE(DecessiFentanyl, TotaleDecessiGenerale)
```

Figura 4.2: Incidenza Fentanyl

- *Incidenza Fentanyl per città*: questa misura calcola, a differenza della precedente, l'incidenza del Fentanyl in rapporto ai decessi totali della sola città visualizzata. Questo permette di capire quanto la sostanza sia prevalente in una specifica comunità, offrendo una visione contestualizzata e confrontabile tra diverse aree geografiche (Figura 4.3).

```
% Incidenza Fentanyl per città =
VAR DecessiFentanyl =
    CALCULATE(
        DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
        'Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza] = "Fentanyl"
    )

VAR TotaleDecessiDellaCitta =
    CALCULATE(
        DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
        ALL('Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza])
    )

RETURN
DIVIDE(DecessiFentanyl, TotaleDecessiDellaCitta, 0)
```

Figura 4.3: Incidenza Fentanyl per città

- *Età media:* rappresenta l'età media tra il totale dei morti, calcolata tramite `AVERAGEX`. Questa formula è fondamentale perché isola l'età per ogni singolo indice di decesso prima di farne la media, evitando errori dovuti alla presenza di più righe per lo stesso individuo (Figura 4.4).

```
Età Media =
AVERAGEX(
    DISTINCT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
    CALCULATE(MAX('Accidental_Drug_Related_Deaths_2012-2021.'[Età]))
)
```

Figura 4.4: Età media

- *Genere Dominante*: identifica statisticamente quale sesso è maggiormente colpito nel contesto filtrato. Utilizza `TOPN` per estrarre il valore con la frequenza più alta (Figura 4.5).

```
Genere Dominante =
VAR ConteggioPerSesso =
    TOPN(
        1,
        SUMMARIZE(
            'Accidental_Drug_Related_Deaths_2012-2021.',
            'Accidental_Drug_Related_Deaths_2012-2021.'[Sesso],
            "Conteggio", [Decessi]
        ),
        [Conteggio],
        DESC
    )
RETURN
MAXX(ConteggioPerSesso, [Sesso])
```

Figura 4.5: Genere dominante

- *Variazione % Anno:* confronta i decessi dell'anno selezionato con quelli dell'anno precedente (AnnoCorrente - 1), restituendo il trend di crescita o decrescita (Figura 4.6).

```
Variazione % Anno =
VAR AnnoCorrente = SELECTEDVALUE('Accidental_Drug_Related_Deaths_2012-2021.'[Anno])
VAR DecessiOggi = [Decessi]
VAR DecessiIeri =
    CALCULATE(
        [Decessi],
        ALL('Accidental_Drug_Related_Deaths_2012-2021.'),
        'Accidental_Drug_Related_Deaths_2012-2021.'[Anno] = AnnoCorrente - 1
    )
RETURN
IF(
    ISBLANK(DecessiIeri) || DecessiIeri = 0,
    BLANK(),
    DIVIDE(DecessiOggi - DecessiIeri, DecessiIeri)
)
```

Figura 4.6: Varazione percentuale annua

- *Verifica Mix (Dataset Sostanze_Mix)*: questa misura fa parte di una tabella che è stata creata come supporto nel modello per gestire le matrici di co-occorrenza. La sua fun-

zione è quella di permettere l'incrocio tra due diverse sostanze per contare quante volte compaiano insieme nel medesimo referto. Questa struttura è indispensabile per rompere i filtri di riga standard del dataset principale (Figura 4.7).

```
Verifica Mix =
VAR SostanzaSelezionataInColonna = SELECTEDVALUE('Sostanze_Mix'[Sostanza])
RETURN
CALCULATE(
    DISTINCTCOUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
    KEEPFILTERS(
        CALCULATETABLE(
            VALUES('Accidental_Drug_Related_Deaths_2012-2021.'[Indice]),
            'Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza] =
            SostanzaSelezionataInColonna,
            ALL('Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza])
        )
    )
)
```

Figura 4.7: Verifica Mix

- *Media Crude Rate USA:* questa misura è stata progettata per calcolare un valore di riferimento nazionale (benchmark) che rimanga costante indipendentemente dalle selezioni geografiche effettuate nel report. Attraverso l'impiego della funzione `CALCULATE`, combinata con la funzione `AVERAGE`, viene calcolato il tasso grezzo di mortalità medio. L'elemento distintivo è l'utilizzo di `ALLEXCEPT`, che istruisce il motore DAX a ignorare tutti i filtri attivi (come lo Stato o la causa del decesso) preservando esclusivamente il filtro sulla colonna "Anno". Ciò permette di confrontare dinamicamente il dato di una specifica area locale con la tendenza media nazionale per il medesimo periodo temporale (Figura 4.8).

```
Media Crude Rate USA =
CALCULATE(
    AVERAGE('Multiple Cause of Death 1999-2014 v1 1'[Crude Rate]),
    ALLEXCEPT('Multiple Cause of Death 1999-2014 v1 1', 'Multiple Cause of Death
    1999-2014 v1 1'[Anno])
)
```

Figura 4.8: Media Crude Rate USA

4.1.2 Descrizione delle Colonne Calcolate

Le colonne calcolate sono state inserite nel dataset del Connecticut per arricchire l'analisi demografica e geografica, trasformando i dati grezzi in dimensioni di analisi più significative. A differenza delle misure, queste colonne vengono elaborate durante il caricamento dei dati e diventano parte integrante della struttura del modello. Ciò ha permesso di segmentare la popolazione in classi specifiche e di standardizzare le informazioni geografiche, fornendo una base solida per le operazioni di filtraggio e categorizzazione all'interno dei report.

Nello specfico sono state realizzate le seguenti colonne DAX:

- *Località Pulita*: creata per garantire una corretta geolocalizzazione nelle mappe di Power BI (Figura 4.9). Questa concatena la città e lo stato con il suffisso "USA", standardizzando i dati per il motore di ricerca geografico.

```
1 LocalitaPulita = 'Accidental_Drug_Related_Deaths_2012-2021.'[Città di Malattia] & ",
  " & 'Accidental_Drug_Related_Deaths_2012-2021.'[Stato di Malattia] & ", USA"
```

Figura 4.9: Località pulita

- *Mese*: estrae il nome testuale del mese dalla data dell'evento. Essenziale per ordinare e visualizzare i trend mensili in modo leggibile (Figura 4.10).

```
1 Mese = FORMAT('Accidental_Drug_Related_Deaths_2012-2021.'[Data], "MMMM")
```

Figura 4.10: Mese

- *Mobilità:* confronta la città di residenza con quella in cui è avvenuta la malattia. Questo permette di distinguere tra casi "Locali" e casi "Fuori Comune", fornendo informazioni sugli spostamenti legati al consumo di sostanze (Figura 4.11).

```
1 Mobilità = IF('Accidental_Drug_Related_Deaths_2012-2021.'[Città di Residenza] =
  'Accidental_Drug_Related_Deaths_2012-2021.'[Città di Malattia], "Locale", "Fuori
  Comune")
```

Figura 4.11: Mobilità

- *Numero sostanze*: quantifica il grado di "complessità farmacologica" di ogni singolo caso; infatti, conta quante righe (sostanze) sono associate allo stesso Indice. È fondamentale per distinguere i decessi causati da una singola sostanza rispetto a quelli multi-sostanza (Figura 4.12).

```
1 Numero Sostanze =
2 CALCULATE(
3     COUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza]),
4     ALLEXCEPT('Accidental_Drug_Related_Deaths_2012-2021.',
      'Accidental_Drug_Related_Deaths_2012-2021.'[Indice])
5 )
```

Figura 4.12: Numero sostanze

- *Sostanze raggruppate*: categorizza dinamicamente i decessi in base alla complessità del referto tossicologico, raggruppandoli per numero di sostanze rilevate nel corpo della vittima. Attraverso l'utilizzo della funzione `SWITCH`, sono state definite delle classi di frequenza (ad esempio: "1", "2", fino alla classe aggregata "5-9"), che permettono di analizzare l'incidenza dei casi di poli-assunzione e semplificare la lettura dei dati all'interno delle dashboard (Figura 4.13).

```
Sostanze Raggruppate =
VAR Conteggio =
    CALCULATE(
        COUNT('Accidental_Drug_Related_Deaths_2012-2021.'[Sostanza]),
        ALLEXCEPT('Accidental_Drug_Related_Deaths_2012-2021.',
        'Accidental_Drug_Related_Deaths_2012-2021.'[Indice])
    )
RETURN
    SWITCH(
        TRUE(),
        Conteggio = 1, "1",
        Conteggio = 2, "2",
        Conteggio = 3, "3",
        Conteggio = 4, "4",
        Conteggio >= 5, "5-9",
        FORMAT(Conteggio, "0")
    )
```

Figura 4.13: Sostanze raggruppate

4.2 Sviluppo delle dashboard di base

Dopo la fase di modellazione e pulizia dei dati, il progetto è confluito nella creazione di un ecosistema di visualizzazione. L'obiettivo è trasformare dati tabellari complessi in informazioni chiare e immediatamente utilizzabili per il supporto decisionale.

Il progetto complessivo si compone di sette dashboard, suddivise in due gruppi in base alla loro complessità tecnica. In questa sezione analizzeremo le prime tre dashboard, pensate come strumenti di esplorazione generale. Seguendo un approccio "Top-Down", il percorso narrativo guida l'utente da una visione macroscopica nazionale fino alla realtà specifica del Connecticut:

- *Dashboard 1 – "Dalle prescrizioni all'emergenza: 15 anni di crisi sanitaria"*: focalizzata sull'inquadramento storico e sulla correlazione tra mercato farmaceutico e mortalità a livello federale (USA).
- *Dashboard 2 – "Focus territoriale: l'emergenza sanitaria sul Connecticut (1999-2014)"*: progettata come ponte analitico per confrontare i tassi di mortalità locali rispetto alla media nazionale statunitense.
- *Dashboard 3 – "Analisi dei decessi accidentali nel Connecticut (2012-2021)"*: un'esplorazione dettagliata delle dinamiche demografiche e tossicologiche interne allo Stato, basata su dati granulari e specifici.

Le restanti quattro dashboard, caratterizzate da logiche di calcolo più articolate e analisi avanzate, verranno trattate nel capitolo successivo. Questa struttura a tre livelli, nazionale, comparativa e locale, permette di costruire una narrativa progressiva, accompagnando l'utente verso un'analisi sempre più dettagliata e contestualizzata del fenomeno.

4.2.1 Dashboard 1: "Dalle prescrizioni all'emergenza: 15 anni di crisi sanitaria"

Questa dashboard è il punto di partenza dell'analisi e offre una panoramica del fenomeno a livello nazionale. Il titolo riassume il percorso narrativo della visualizzazione, cioè, mostrare

come la diffusione delle prescrizioni mediche di antidolorifici si sia trasformata in una delle più gravi crisi sanitarie degli Stati Uniti. L'arco temporale di circa quindici anni permette di osservare l'evoluzione storica dell'emergenza e di identificarne le tendenze principali (Figura 4.14).

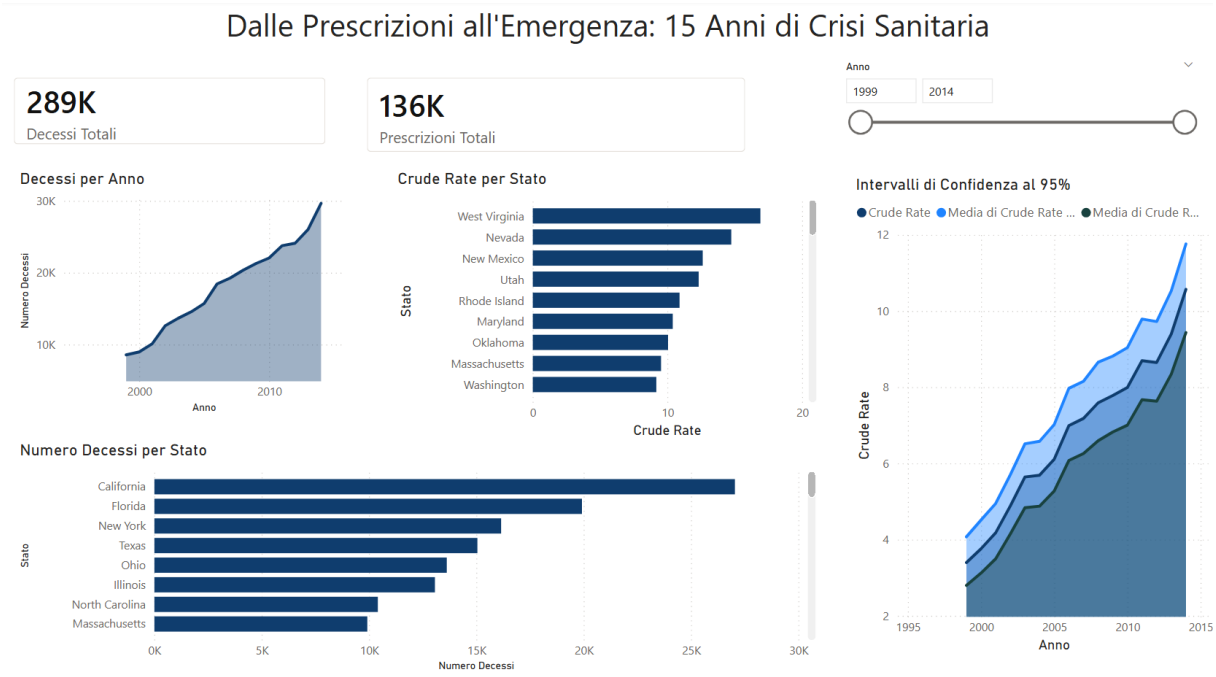


Figura 4.14: Dashboard 1 - Dalle prescrizioni all'emergenza: 15 anni di crisi sanitaria

Come mostrato nella Figura 4.14, la dashboard è organizzata in un insieme di componenti visive complementari, progettate per offrire una lettura immediata e progressiva del fenomeno. In particolare, gli elementi principali sono i seguenti:

- *KPI Card - Decessi Totali (289K):* questo è l'indicatore principale che riassume il numero complessivo di vittime tra il 1999 e il 2014. Questo dato offre un'immagine immediata della gravità dell'emergenza sanitaria, fornendo la base numerica necessaria per contestualizzare tutta l'analisi successiva

- *KPI Card - Prescrizioni Totali (136K):* mostra il numero complessivo di prescrizioni mediche di oppioidi erogate. È un dato fondamentale per contestualizzare la correlazione tra la distribuzione legale di farmaci e l'insorgere della crisi da abuso di sostanze.

- *Grafico ad area - Decessi per Anno:* questa visualizzazione temporale evidenzia il trend di crescita dei decessi su base annuale. Il passaggio da circa 10.000 a quasi 30.000 decessi annui mostra un'accelerazione costante, sottolineando come il fenomeno non abbia subito flessioni nel corso dei quindici anni considerati.

- *Grafico a barre - Numero Decessi per Stato:* questo istogramma classifica gli stati americani in base al numero assoluto di morti. Alcuni stati, come California, Florida e New York, appaiono in cima alla lista principalmente a causa della loro elevata densità demografica, identificando le aree geografiche con il maggior carico gestionale per il sistema sanitario.

- *Grafico a barre - Crude Rate per Stato:* a differenza del precedente, questo grafico normalizza il dato calcolando il tasso di mortalità ogni 100.000 abitanti. È l'indicatore più

fedele della gravità dell'epidemia; infatti, rivela che stati meno popolosi, come il West Virginia, presentano, in realtà, l'incidenza di rischio più alta.

- *Grafico a linee con area - Intervalli di Confidenza al 95%:* questo grafico rappresenta la solidità statistica dell'intera analisi. La linea centrale traccia la media del Crude Rate nazionale, mentre la fascia d'ombra che la circonda indica l'intervallo di confidenza (limiti superiore e inferiore). La ridotta ampiezza di questa fascia conferma un'elevata precisione dei dati; ciò dimostra che il trend di crescita osservato è statisticamente significativo e non imputabile a semplici oscillazioni casuali.
- *Slicer temporale (Anno):* questo strumento di filtro permette di isolare specifici intervalli temporali all'interno della dashboard. Consente un'analisi dinamica, permettendo al lettore di osservare come cambiano le classifiche degli stati e i tassi di mortalità muovendosi tra l'inizio (1999) e il picco (2014) della rilevazione.

4.2.2 Dashboard 2: "Focus territoriale: l'emergenza sanitaria sul Connecticut (1999-2014)"

Questa dashboard è stata progettata per offrire un focus sullo stato del Connecticut, permettendo di confrontare i dati locali con i trend nazionali e di indagare la correlazione diretta tra l'offerta di farmaci oppioidi e la mortalità registrata. L'obiettivo è isolare le specificità di un singolo stato per comprendere se le dinamiche della crisi sanitaria seguano modelli uniformi o presentino anomalie territoriali (Figura 4.15).

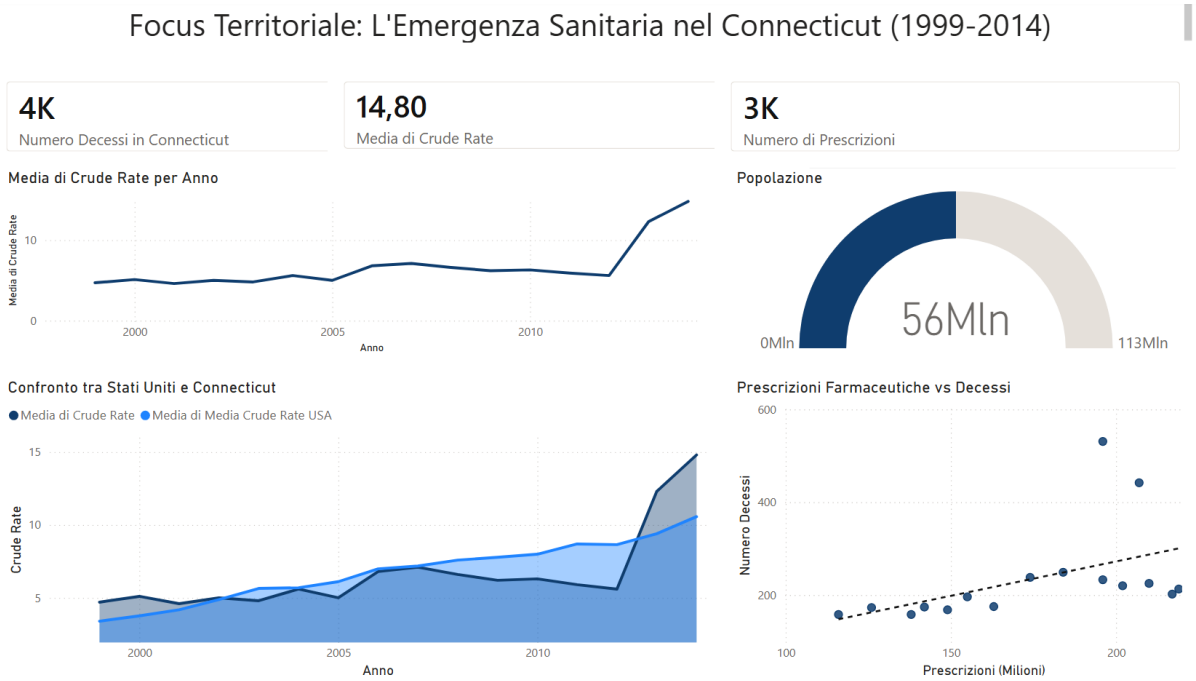


Figura 4.15: Dashboard 2: "Focus territoriale: l'emergenza sanitaria sul Connecticut (1999-2014)"

Di seguito vengono analizzati nel dettaglio gli oggetti visivi che compongono la dashboard:

- *KPI Card - Numero Decessi in Connecticut (4K):* espone il valore assoluto delle vittime nel periodo considerato all'interno dello stato.
- *KPI Card - Media di Crude Rate (14,80):* indica il tasso di mortalità medio per 100.000 abitanti, fornendo la misura della gravità dell'impatto sulla popolazione locale.

- *KPI Card - Numero di Prescrizioni (3K):* quantifica il volume di ricette per oppioidi emesse, dato essenziale per l'analisi della correlazione.
- *Grafico a linee - Media di Crude Rate per Anno*: visualizza l'evoluzione temporale della mortalità in Connecticut. È possibile osservare una fase di relativa stabilità fino al 2012, seguita da un'impennata drammatica negli ultimi due anni, segnale di un brusco cambiamento nelle dinamiche di abuso o nella potenza delle sostanze circolanti.
- *Tachimetro - Popolazione:* questo indicatore rappresenta la dimensione demografica del campione analizzato. Il valore di 56 milioni corrisponde alla somma dei residenti calcolata nell'arco temporale del dataset. La sua funzione è fondamentale per contestualizzare la portata del Crude Rate, ricordando che ogni tasso di mortalità viene calcolato in rapporto a questo volume di popolazione per garantirne la validità statistica.
- *Grafico ad area - Confronto tra Stati Uniti e Connecticut*: questo è uno degli oggetti più critici della dashboard. Mette a confronto il Crude Rate del Connecticut con la media degli interi Stati Uniti. L'area evidenzia come, dopo un lungo periodo in linea con il trend nazionale, il Connecticut abbia superato significativamente la media USA a partire dal 2013, diventando un caso di particolare urgenza sanitaria.
- *Grafico a dispersione (Scatter Plot) - Prescrizioni Farmaceutiche vs Decessi:* questo grafico analizza la relazione tra il numero di prescrizioni mediche (Asse X) e il numero di decessi (Asse Y). La linea di tendenza tratteggiata che punta verso l'alto dimostra visivamente una correlazione positiva: all'aumentare della disponibilità di prescrizioni legali è corrisposto, storicamente, un aumento dei decessi, supportando l'ipotesi della natura farmacologica iniziale della crisi.

4.2.3 Dashboard 3: "Analisi dei decessi accidentali nel Connecticut (2012-2021)"

Questa dashboard sposta il focus dall'analisi geografica a quella sociodemografica, tracciando il profilo delle persone colpite dalla crisi. Attraverso l'esame di variabili quali età, genere ed etnia, questa dashboard permette di identificare le fasce di popolazione più vulnerabili. L'obiettivo è fornire una base analitica per comprendere come i determinanti sociali influenzino l'esposizione al rischio di overdose accidentale nel contesto specifico del Connecticut (Figura 4.16).

Di seguito vengono analizzati nel dettaglio gli oggetti visivi presenti nella dashboard:

- *KPI Card - Decessi (6K):* rappresenta il volume totale dei casi analizzati tra il 2012 e il 2021.
- *KPI Card - Età Media (43):* identifica l'età media delle vittime, posizionando il fenomeno prevalentemente nella fascia adulta e produttiva della popolazione.
- *KPI Card - Genere Dominante (Maschio):* evidenzia immediatamente la forte disparità di genere nel fenomeno.
- *Grafico ad area - Decessi per Anno:* questa visualizzazione temporale evidenzia il trend di crescita dei decessi su base annuale, sottolineando come il fenomeno sia sempre in crescita.
- *Grafico a cascata - Variazione Annuale dei Decessi*: questo grafico è particolarmente efficace per mostrare l'incremento netto anno dopo anno. I blocchi verdi indicano l'aumento costante dei casi, permettendo di visualizzare come ogni anno si sia aggiunto un ulteriore carico di vittime rispetto al precedente, fino a raggiungere il totale cumulativo di 6.100 decessi.

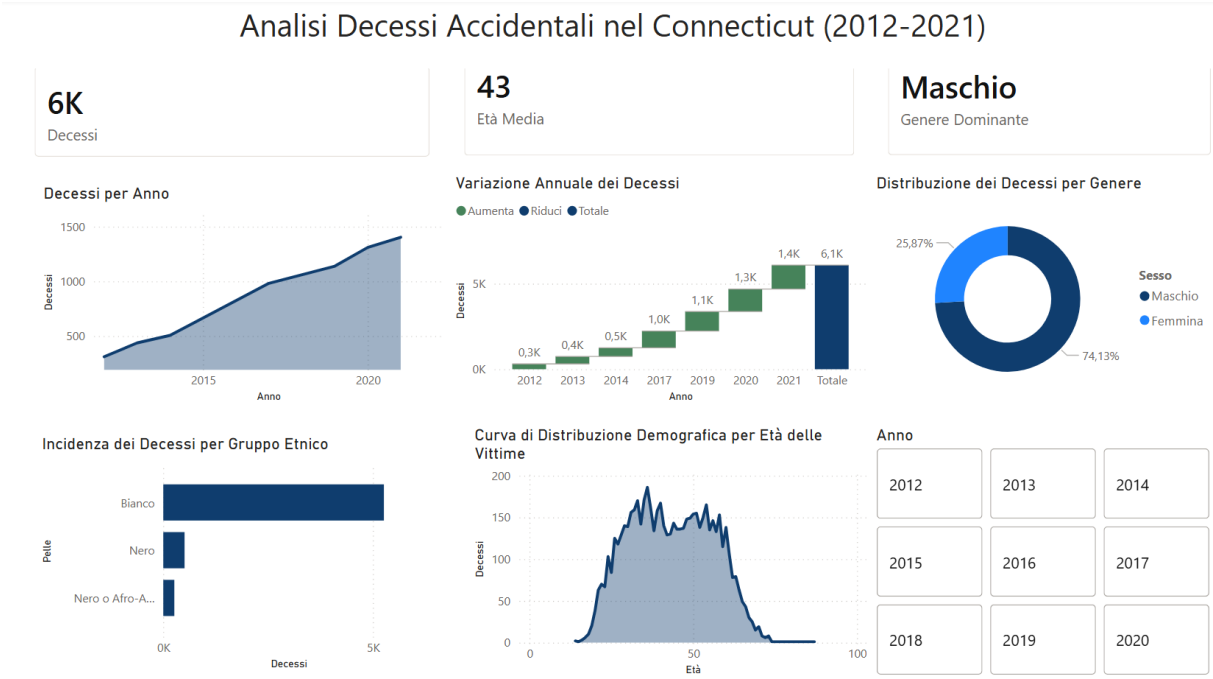


Figura 4.16: Dashboard 3: "Analisi dei decessi accidentali nel Connecticut (2012-2021)"

- *Grafico a ciambella - Distribuzione dei Decessi per Genere:* visualizza la ripartizione percentuale tra i sessi. Il dato mostra una netta prevalenza maschile (74,13% contro il 25,87% femminile), suggerendo la necessità di approcci di prevenzione differenziati per genere.
- *Grafico a barre - Incidenza dei Decessi per Gruppo Etnico*: questo istogramma analizza la distribuzione delle vittime in base all'appartenenza etnica. La netta prevalenza del gruppo "Bianco" rispetto agli altri gruppi riflette le dinamiche demografiche locali, ma permette anche di monitorare eventuali sproporzioni nel tasso di mortalità tra le diverse comunità.
- *Grafico ad area - Curva di Distribuzione Demografica per Età delle Vittime:* rappresenta la distribuzione di frequenza dei decessi rispetto all'età. La curva mostra un picco massimo intorno ai 35-55 anni, indicando che la crisi non colpisce solo i giovanissimi, ma si concentra pesantemente sugli adulti nella fase centrale della vita.
- *Slicer - Anno:* i pulsanti posizionati in basso a destra permettono di filtrare l'intera dashboard per ogni singolo anno dal 2012 al 2021. Questo strumento è essenziale per osservare se il profilo delle persone coinvolte sia cambiato nel tempo (ad esempio, se l'età media si sia abbassata con l'introduzione di sostanze più potenti come il Fentanyl).

CAPITOLO 5

Progettazione e implementazione delle dashboard avanzate

In questo capitolo vengono descritte la progettazione e l'implementazione delle quattro dashboard avanzate del progetto. Si introduce dapprima il cambio di prospettiva metodologica rispetto al capitolo precedente, illustrando le ragioni che hanno reso necessario il ricorso a oggetti visivi di maggiore complessità — il diagramma di Sankey, il chord diagram e la matrice di co-occorrenza — capaci di rappresentare flussi, interazioni multidimensionali e reti di relazioni tra variabili. Si descrivono, quindi, le quattro dashboard sviluppate, ciascuna dedicata a una dimensione chiave dell'emergenza sanitaria. La quarta dashboard — "Analisi Territoriale: Flussi e Luoghi del Decesso" — mappa la mobilità epidemiologica delle vittime tra comune di residenza e comune del decesso, identificando i principali hub urbani di attrazione e la prevalenza della mortalità in contesti domestici. La quinta — "Analisi Tossicologica: Interazioni tra Sostanze e Profili di Policonsumo" — indaga la rete del policonsumo attraverso la matrice di co-occorrenza e il chord diagram, certificando come la letalità della crisi sia quasi sempre il prodotto di interazioni sinergiche tra più sostanze. La sesta — "Analisi Temporale e Stagionalità" — scende a una granularità mensile per isolare pattern ciclici e picchi stagionali nel decennio 2012–2021. La settima e ultima — "Focus Monografico: L'impatto del Fentanyl" — traccia la crescita esponenziale dell'oppioide sintetico che è stato il motore assoluto della terza ondata epidemica.

5.1 Introduzione all'analisi approfondita

Se nel capitolo precedente l'attenzione si è concentrata sull'esplorazione descrittiva iniziale, tracciando la linea evolutiva della crisi su scala federale e il profilo sociodemografico generale delle vittime nel Connecticut, questo quinto capitolo segna un fondamentale cambio di passo metodologico. L'obiettivo della ricerca si sposta, infatti, da una prospettiva puramente descrittiva ("cosa è successo e chi è stato coinvolto?") a un approccio di tipo diagnostico e correlazionale ("come interagiscono i fattori e dove si concentrano le dinamiche?"), volto a far emergere i pattern e le relazioni complesse che caratterizzano il fenomeno dei decessi per overdose. A questo scopo sono state sviluppate le ultime quattro dashboard del progetto, caratterizzate da una maggiore complessità visiva e informativa rispetto alle precedenti. Queste dashboard sono state progettate per esplorare quattro dimensioni chiave dell'emergenza sanitaria, cioè i flussi di mobilità territoriale delle vittime, il fenomeno del policonsumo, l'andamento temporale e stagionale dei decessi e, infine, l'impatto specifico del Fentanyl come vero e proprio motore principale della crisi. Per riuscire a rappresentare visivamente queste dinamiche così articolate, si è reso necessario l'utilizzo di oggetti visivi

più complessi rispetto a quelli standard di Power BI, capaci di mostrare flussi e interazioni multidimensionali. In particolare, sono stati utilizzati:

- *Il diagramma di Sankey*: un grafico dedicato all'analisi dei flussi, in cui i punti di origine e di destinazione sono collegati da bande colorate. La larghezza di queste bande è proporzionale al volume dei dati, permettendo di capire a colpo d'occhio l'entità degli spostamenti.
- *Il diagramma a corda (chord diagram)*: una visualizzazione circolare in cui i diversi elementi sono disposti lungo la circonferenza e interconnessi da nastri. È lo strumento ideale per mostrare le relazioni reciproche e la loro intensità, evidenziando quali elementi tendono ad associarsi più frequentemente (come le sostanze assunte contemporaneamente).
- *La matrice di co-occorrenza*: una tabella a doppia entrata arricchita da una formattazione condizionale a sfumature di colore (heatmap). Consente di incrociare una lista di elementi con se stessa, mostrando, nel punto di intersezione, quante volte due variabili si presentano insieme nello stesso record.

5.2 Le dashboard di analisi e visualizzazione avanzata

Conclusa la panoramica introduttiva sui volumi complessivi e sulle caratteristiche demografiche del fenomeno, il focus della ricerca si sposta ora sulle dinamiche di dettaglio. In questa sezione verranno analizzate le ultime quattro dashboard, sviluppate e pensate come strumenti di analisi avanzata per l'esplorazione dei dati. Attraverso l'incrocio di variabili spaziali, temporali e chimiche, questi moduli visivi permettono di identificare pattern ricorrenti, correlazioni inaspettate e fattori di rischio specifici che un'analisi aggregata non potrebbe rivelare. Esplorando la geografia dei decessi, la struttura del policonsumo, la stagionalità degli eventi e il ruolo del Fentanyl, si cercherà di ricostruire l'architettura esatta di questa crisi. Il percorso analitico si strutturerà attraverso le seguenti dashboard:

- *Dashboard 4 — "Analisi Territoriale: Flussi e Luoghi del Decesso"*: mappa la mobilità epidemiologica delle vittime, tracciando i flussi reali tra il comune di residenza e il comune in cui è avvenuto il decesso, distinguendo la mortalità locale da quella fuori comune.
- *Dashboard 5 — "Analisi Tossicologica: Interazioni tra Sostanze e Profili di Policonsumo"*: indaga la complessa e letale rete del policonsumo attraverso la quantificazione delle co-occorrenze e delle relazioni tra i diversi principi attivi rinvenuti nei test tossicologici.
- *Dashboard 6 — "Analisi Temporale e Stagionalità"*: riesamina l'evoluzione storica dei decessi nel decennio 2012-2021 scendendo a un livello di granularità mensile, al fine di isolare eventuali anomalie cicliche, picchi stagionali o variazioni nei luoghi di ritrovamento in base al periodo dell'anno.
- *Dashboard 7 — "Focus Monografico: L'impatto del Fentanyl (2012-2021)"*: si configura come un approfondimento monografico dedicato interamente all'oppioide sintetico responsabile della più recente e devastante ondata della crisi, profilandone la crescita esponenziale, le specificità demografiche e i principali cluster urbani di incidenza.

5.2.1 Dashboard 4 — "Analisi Territoriale: Flussi e Luoghi del Decesso"

La quarta dashboard introduce la dimensione geografica e logistica della crisi degli oppioidi nel Connecticut, spostando l'oggetto di studio dall'analisi puramente quantitativa a quella

spazio-temporale. La dashboard è stata progettata per mappare non solo la distribuzione fisica delle overdose sul territorio, ma soprattutto per indagare il fenomeno della mobilità correlata al consumo. Attraverso l'incrocio tra la città di residenza ufficiale della vittima e il comune in cui è stato rinvenuto il corpo, l'analisi mira a verificare l'esistenza di "hub" urbani di attrazione per il consumo o lo spaccio, distinguendo tra dinamiche di mortalità locale e flussi di mobilità intercomunale. La configurazione grafica della dashboard viene presentata nella Figura 5.1.

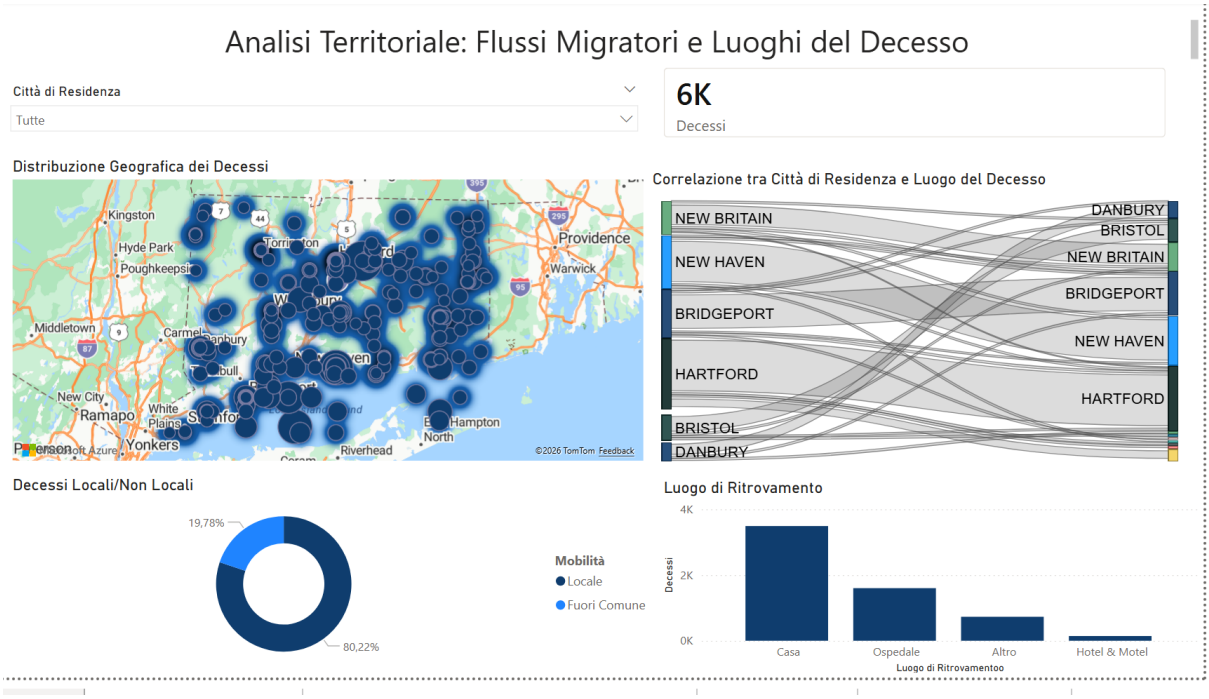


Figura 5.1: Dashboard 4 — "Analisi Territoriale: Flussi e Luoghi del Decesso"

Di seguito vengono analizzati nel dettaglio gli oggetti visivi presenti nella dashboard:

- *KPI Card - Decessi (6K)*: mostra il volume totale dei casi analizzati, fornendo la base statistica per le percentuali di mobilità e localizzazione.
- *Slicer - Città di Residenza*: menù a tendina posizionato in alto a sinistra che permette di filtrare l'intera pagina per singolo comune di origine; è essenziale per isolare i flussi migratori di specifiche comunità.
- *Mappa di Densità - Distribuzione Geografica dei Decessi*: mappa georeferenziata che utilizza la tecnica della heatmap (mappa di calore) per visualizzare i cluster di mortalità. I punti a maggiore densità cromatica (blu intenso) evidenziano la concentrazione delle overdose nei principali centri urbani, confermando che la crisi, pur essendo diffusa, risponde a precise geografie della vulnerabilità urbana.
- *Diagramma di Sankey - Correlazione tra Città di Residenza e Luogo del Decesso*: elemento ad alta complessità analitica. Questo grafico a flussi mette in relazione bidirezionale la colonna di sinistra (Città di Residenza) con quella di destra (Luogo del Decesso). La larghezza dei flussi grigi è proporzionale al numero di casi. Dal punto di vista epidemiologico, questo grafico permette di mappare la mobilità della crisi.
- *Grafico a Ciambella - Decessi Locali/Non Locali*: mostra che l'80,22% dei decessi avviene all'interno del proprio comune di residenza ("Locale"), mentre il 19,78%, quasi un

decesso su cinque, avviene "Fuori Comune". Questo circa 20% di mobilità intercomunale rappresenta un dato cruciale per le politiche sanitarie, poiché dimostra che il rischio di overdose non si esaurisce entro i confini amministrativi di residenza della vittima.

- *Grafico a Barre - Luogo di Ritrovamento*: classifica i contesti fisici del decesso. La larghissima prevalenza della categoria "Casa" (oltre 4.000 casi) rispetto a "Ospedale", "Hotel e Motel" e "Altro" evidenzia la natura isolata dell'overdose da oppioidi. Il fatto che la maggior parte delle morti avvenga in contesti domestici privati implica che il consumo avviene spesso in solitudine, riducendo drasticamente le probabilità di un intervento tempestivo di soccorso.

5.2.2 Dashboard 5 — "Analisi Tossicologica: Interazioni tra Sostanze e Profili di Policonsumo"

La quinta dashboard costituisce il nucleo strettamente clinico e tossicologico della ricerca, focalizzandosi in modo specifico sul fenomeno del policonsumo. Il presupposto scientifico che guida questa sezione è che l'elevata letalità dell'attuale crisi degli oppioidi non sia quasi mai riconducibile all'assunzione di una singola sostanza isolata. Al contrario, il tasso di mortalità risulta direttamente correlato alle interazioni biochimiche sinergiche che si innescano quando più molecole vengono assunte simultaneamente, spesso all'insaputa del consumatore stesso. Dal punto di vista fisiologico, infatti, combinare diverse molecole annulla le naturali capacità di difesa dell'organismo. Quando si assumono insieme più sostanze ad azione deprimente (come oppioidi uniti ad alcol o tranquillanti), i loro effetti non si limitano a sommarsi, ma si moltiplicano in modo esponenziale. Il sistema nervoso centrale subisce un sovraccarico inibitorio, cioè il cervello smette di inviare ai polmoni l'impulso vitale ed automatico del respiro. Il corpo scivola, così, in una rapida e silenziosa asfissia prima ancora di poter smaltire le tossine. Altrettanto insidiosa è l'associazione con sostanze stimolanti (come la cocaina); infatti, in questo caso, l'effetto eccitante "maschera" temporaneamente la sedazione, illudendo il soggetto di avere il controllo e spingendolo ad assumere dosi di oppioidi che si riveleranno fatali non appena la spinta dello stimolante si esaurisce. L'obiettivo di questa dashboard è esplorare i dati derivanti dai referti autoptici per mappare la co-occorrenza delle sostanze, operando una scomposizione incrociata per genere e fasce anagrafiche. Tale approccio permette di identificare gli abbinamenti letali più ricorrenti all'interno del mercato illecito, delineando l'architettura chimica dell'overdose (Figura 5.2).

Di seguito vengono analizzati nel dettaglio gli oggetti visivi presenti nella dashboard:

- *Grafico a Tornado - Distribuzione dei Decessi per Sostanza Principale e Genere*: un grafico a barre orizzontali contrapposte che scompone la presenza delle sostanze tra genere femminile (sinistra, viola) e maschile (destra, verde). Il Fentanyl e la categoria generale "Qualsiasi Oppioide" dominano nettamente le rilevazioni, seguiti da Cocaina ed Eroina. Visivamente, il grafico evidenzia come la sproporzione di genere (con i maschi che rappresentano circa il triplo dei casi in quasi tutte le sostanze) rimanga costante, tranne che per le Benzodiazepine, dove lo scarto tra i generi si riduce proporzionalmente, suggerendo pattern di consumo o prescrizione parzialmente differenti nelle donne.

- *Tabella a Matrice - Matrice di Co-occorrenza delle Sostanze Tossiche*: questa matrice incrocia le sostanze su entrambi gli assi per contare quante volte due molecole sono state trovate insieme nello stesso corpo. Leggendo le intersezioni, emergono combinazioni critiche; l'incrocio tra Fentanyl e Cocaina registra ben 1.570 casi, mentre quello tra Fentanyl ed Eroina ne segna 1.072. Questo dato certifica la diffusione dei mix di stimolanti e depressori, o la massiccia contaminazione del mercato della cocaina da parte del fentanyl sintetico.

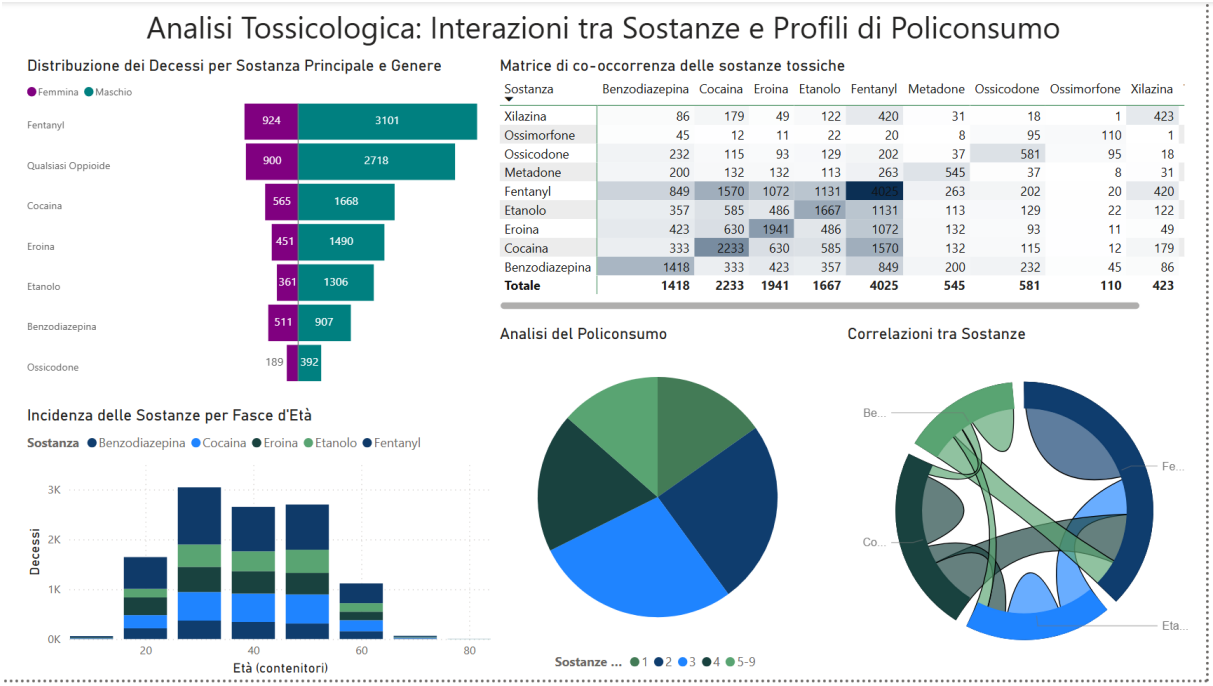


Figura 5.2: Dashboard 5 — "Analisi Tossicologica: Interazioni tra Sostanze e Profili di Policonsumo"

- *Grafico a Barre Raggruppate - Incidenza delle Sostanze per Fasce d'Età*: analizza come cambia la preferenza o l'esposizione alle sostanze al variare dell'età (20, 40, 60, 80 anni). Permette di notare come il Fentanyl (blu scuro) sia massicciamente presente nelle fasce giovanili e adulte (20-40), mentre in età più avanzate (60+) mantengano una quota rilevante sostanze più "tradizionali", o legate a prescrizioni croniche, come l'etanolo o determinati oppioidi farmaceutici.

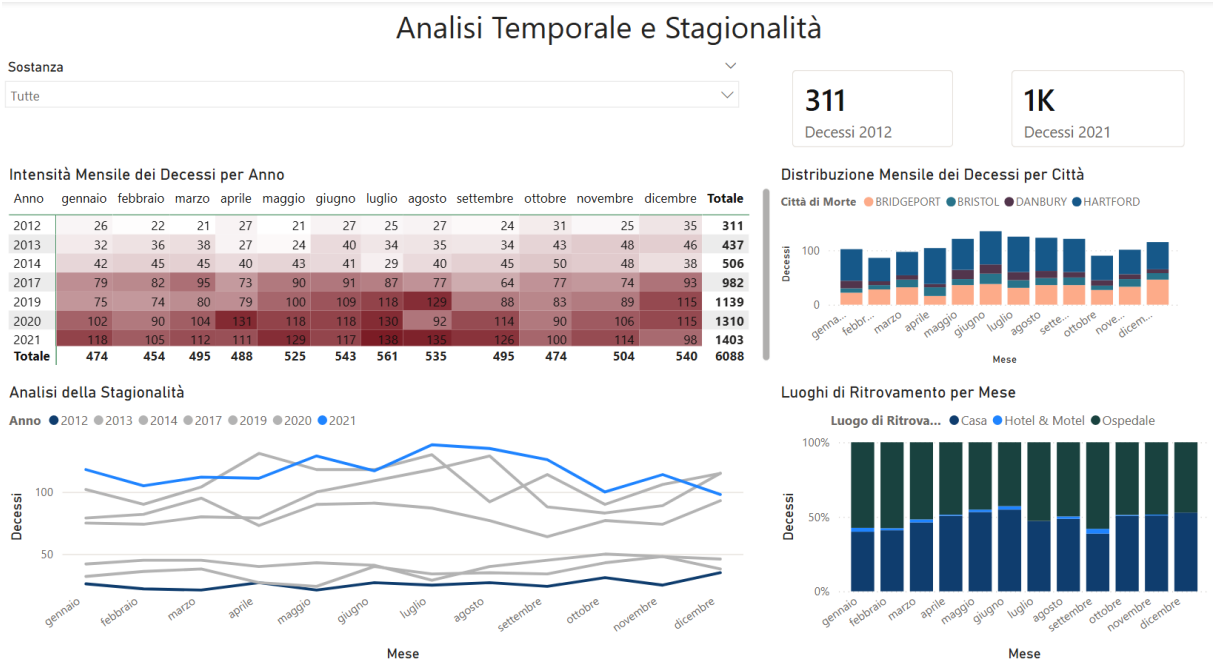
- *Grafico a Torta - Analisi del Policonsumo*: mostra la scomposizione dei decessi in base al numero esatto di sostanze rinvenute simultaneamente (da 1 a 9 sostanze). La quota rappresentata da una sola sostanza è minoritaria; la stragrande maggioranza dei decessi presenta una positività a 2, 3 o 4+ sostanze, confermando graficamente la tesi che l'overdose sia un fenomeno intrinsecamente multi-molecolare.

- *Chord Diagram - Correlazioni tra Sostanze* : questo diagramma circolare rappresenta visivamente la rete di interconnessioni tossicologiche. Ogni arco della circonferenza rappresenta una sostanza e lo spessore delle corde interne che collegano due archi è proporzionale alla frequenza del loro consumo congiunto. Visivamente, le corde più spesse collegano il Fentanyl alla Cocaina e all'Eroina. Questo grafico permette di cogliere istantaneamente la struttura "a rete" della dipendenza; il Fentanyl funge da vero e proprio nucleo centrale dell'epidemia, legandosi inestricabilmente a ogni altra sostanza presente sul mercato illecito.

5.2.3 Dashboard 6 — "Analisi Temporale e Stagionalità"

La sesta dashboard affronta la dimensione strettamente cronologica ed epidemiologica, investigando l'evoluzione dell'epidemia sia sul lungo periodo (trend pluriennale) sia sul breve periodo (variazioni stagionali e mensili). La scomposizione mensile del dato mira a superare la visione statica del fenomeno per esplorarne la micro-ciclicità. Isolare i pattern stagionali significa riconoscere che i picchi di overdose non si distribuiscono in modo casuale nel corso dell'anno, ma agiscono come sismografi di altre vulnerabilità. L'obiettivo è

rintracciare come le dinamiche climatiche (che esasperano l'isolamento fisico e i rischi per le popolazioni marginalizzate), i periodi di festività (spesso associati a un inasprimento del disagio psicologico) e le fluttuazioni logistiche nelle reti di approvvigionamento (che alterano improvvisamente il grado di purezza e tossicità delle sostanze in strada) impattino sul tasso di mortalità. La dashboard persegue, dunque, un obiettivo duplice e complementare: da un lato mappare la traiettoria macroscopica e storica del fenomeno; dall'altro fornire evidenze empiriche sui fattori di rischio temporali, offrendo uno strumento indispensabile per programmare campagne di prevenzione e interventi di salute pubblica nei periodi di massima criticità (Figura 5.3).



Anno	gennaio	febbraio	marzo	aprile	maggio	giugno	luglio	agosto	settembre	ottobre	novembre	dicembre	Totale
2012	26	22	21	27	21	27	25	27	24	31	25	35	311
2013	32	36	38	27	24	40	34	35	34	43	48	46	437
2014	42	45	45	40	43	41	29	40	45	50	48	38	506
2017	79	82	95	73	90	91	87	77	64	77	74	93	982
2019	75	74	80	79	100	109	118	129	88	83	89	115	1139
2020	102	90	104	131	118	118	130	92	114	90	106	115	1310
2021	118	105	112	111	129	117	138	135	126	100	114	98	1403
Totale	**474**	**454**	**495**	**488**	**525**	**543**	**561**	**535**	**495**	**474**	**504**	**540**	**6088**

Figura 5.3: Dashboard 6 — "Analisi Temporale e Stagionalità"

Di seguito vengono analizzati nel dettaglio gli oggetti visivi presenti nella dashboard:

- *KPI Cards - Decessi 2012 (311) vs Decessi 2021 (1K)*: questi due indicatori numerici affiancati offrono un impatto visivo immediato della traiettoria della crisi. Il passaggio da 311 decessi annui a 1K dimostra che in meno di un decennio la mortalità è quadruplicata.

- *Heatmap - Intensità Mensile dei Decessi per Anno*: elemento ad alta complessità analitica. Questa matrice incrocia gli anni sulle righe e i mesi sulle colonne, applicando una formattazione condizionale cromatica; infatti, più il colore rosso si fa scuro e intenso, più il numero di decessi è elevato. La lettura verticale mostra lo shock storico, cioè le celle passano dal rosa chiarissimo del 2012 (valori intorno ai 20-30 decessi al mese) al rosso profondo del 2021 (con picchi come i 138 decessi di luglio o i 135 di agosto). La lettura orizzontale permette invece l'analisi intra-annuale, rivelando come i mesi estivi (giugno, luglio, agosto) tendano a concentrare un maggior numero di eventi critici rispetto ai mesi invernali.

- *Grafico a Linee Sovrapposte - Analisi della Stagionalità*: riporta i mesi dell'anno in ascissa e il volume dei decessi in ordinata, dedicando una linea distinta a ogni annualità analizzata. Questa modalità di visualizzazione è cruciale poiché, isolando i trend stagionali dalle variazioni del volume assoluto, permette di confrontare direttamente la morfologia delle curve mensili. L'analisi evidenzia una chiara ciclicità: la maggior parte delle linee mostra

una flessione a inizio anno (febbraio-marzo), un'ascesa progressiva verso un picco estivo e una successiva recrudescenza nel mese di dicembre. Al contempo, il grafico restituisce l'impatto della crescita macroscopica del fenomeno nel corso del tempo, visibile nel progressivo slittamento verso l'alto delle traiettorie annuali, culminante nella curva del 2021 (evidenziata in blu) che si posiziona stabilmente al vertice della distribuzione.

- *Grafico a Barre in Pila - Distribuzione Mensile dei Decessi per Città*: scompone l'andamento mensile dei decessi in base alle principali città dell'evento (Hartford, New Haven, Waterbury, ecc.). Questo permette di verificare se la stagionalità colpisca in modo uniforme l'intero stato o se sia guidata da anomalie specifiche di singoli territori urbani.

- *Grafico a Barre - Luoghi di Ritrovamento per Mese*: questo istogramma a colonne standardizzate al 100% analizza la quota percentuale dei contesti fisici di decesso (Casa, Hotel, Ospedale) mese per mese. L'evidenza principale è l'estrema stabilità della struttura; la barra blu scuro relativa alla "Casa" mantiene una quota costante (circa il 50% o più) in ogni singolo mese dell'anno. Ciò dimostra che la tendenza a consumare e morire in isolamento domestico è una costante strutturale dell'epidemia, del tutto indipendente dalle condizioni climatiche o stagionali esterne.

5.2.4 Dashboard 7 — "Focus Monografico: L'impatto del Fentanyl (2012-2021)"

La settima e ultima dashboard costituisce il momento conclusivo dell'analisi, si tratta di un approfondimento monografico interamente dedicato all'impatto del Fentanyl. L'avvento di questo oppioide sintetico ad altissima potenza non ha rappresentato un semplice aggravamento del fenomeno, ma ha radicalmente ridefinito la natura biologica, sociale e criminale della crisi americana, segnando un vero e proprio spartiacque epidemiologico. La letteratura scientifica concorda nel suddividere l'emergenza in tre fasi storiche: se la prima ondata è stata innescata dalla sovraprescrizione di farmaci analgesici e la seconda dal conseguente passaggio all'eroina, questo modulo visivo mappa specificamente la cosiddetta 'terza ondata', caratterizzata dall'egemonia assoluta e letale dei composti sintetici. Isolando all'interno del dataset esclusivamente i referti tossicologici positivi a questa molecola, la dashboard si pone l'obiettivo di tracciarne il peculiare profilo demografico, misurarne la velocità esponenziale di penetrazione nel mercato illecito e mappare con precisione i cluster urbani a maggiore letalità (Figura 5.4).

Di seguito vengono analizzati nel dettaglio gli oggetti visivi presenti nella dashboard:

- *KPI Cards - Decessi (4K)*: indica che su circa 6.000 decessi totali registrati nel periodo, ben 4.000 coinvolgono direttamente il Fentanyl, quantificandone l'impatto epidemiologico devastante.

- *KPI Cards - Incidenza Fentanyl sul Totale (66%)*: rappresenta l'indicatore chiave di questa sezione. Certifica che il Fentanyl è presente in due decessi su tre all'interno dello Stato del Connecticut, sancendo il passaggio da un'epidemia diversificata a un regime di quasi-monopolio della letalità sintetica.

- *KPI Cards - Età della vittima più piccola (14)*: una statistica drammatica che evidenzia la penetrazione della sostanza nelle fasce d'età adolescenziali, legata alla diffusione di pillole contraffatte (fake pills) che imitano farmaci legittimi ma contengono dosi letali di Fentanyl.

- *Grafico a Linee - Analisi della Crescita del Fentanyl*: rappresenta l'andamento dei decessi da fentanyl dal 2012 al 2021. La traiettoria mostra una crescita quasi verticale a partire dal 2015, anno in cui il Fentanyl ha iniziato a sostituire l'eroina come principale agente

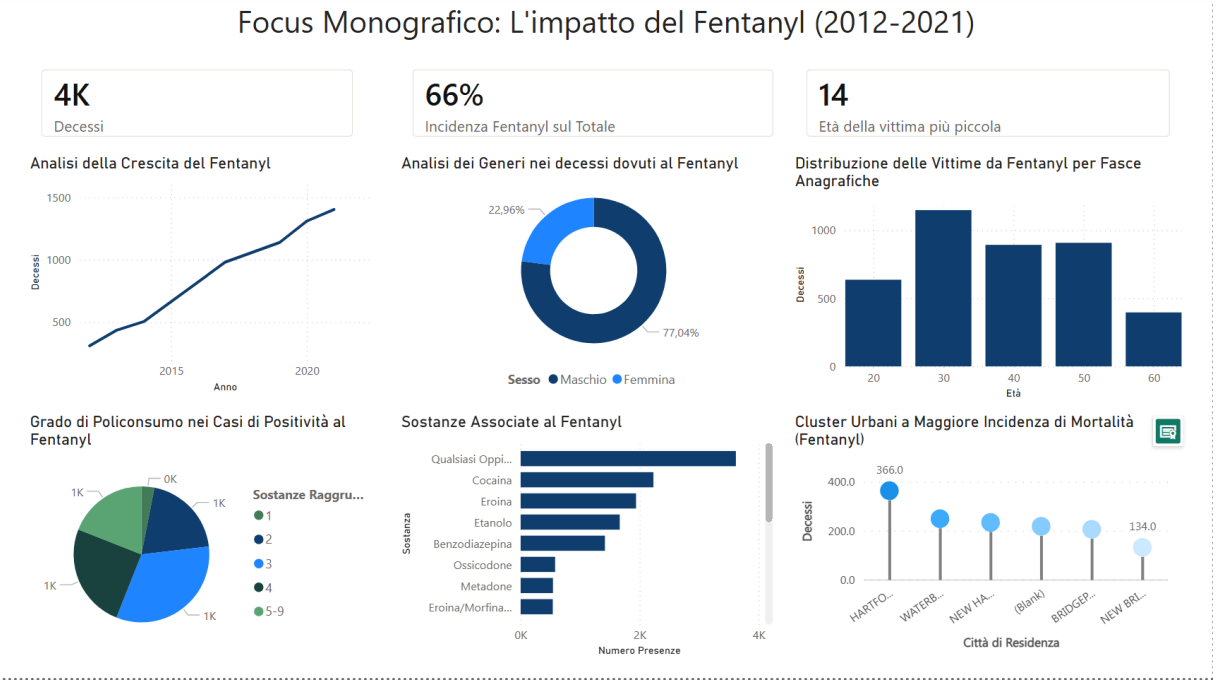


Figura 5.4: Dashboard 7 — "Focus Monografico: L'impatto del Fentanyl (2012-2021)"

di taglio o come sostanza primaria nel mercato di strada, determinando l'esplosione della mortalità generale vista nei capitoli precedenti.

- *Grafico a Ciambella - Analisi dei Generi nei decessi da Fentanyl*: mostra che il 77,04% delle vittime è di sesso maschile rispetto al 22,96% di sesso femminile. Questo scarto è leggermente più pronunciato rispetto alla media del dataset generale, confermando che la popolazione maschile presenta indici di vulnerabilità e stili di consumo legati agli oppioidi sintetici ancora più estremi.

- *Grafico a Barre - Distribuzione delle Vittime per Fasce Anagrafiche*: l'istogramma evidenzia come il picco assoluto dei decessi legati al Fentanyl si concentri nella fascia d'età dei 30-39 anni (i trentenni), seguita dalla fascia dei 40 e dei 50. Questo smentisce lo stereotipo mediatico che dipinge l'overdose come un problema confinato alle subculture giovanili o all'emarginazione cronica. Il Fentanyl colpisce prevalentemente la popolazione nel pieno dell'età lavorativa, riproduttiva e sociale.

- *Grafico a Torta - Grado di Policonsumo nei Casi di Positività al Fentanyl*: analizza quante altre sostanze si associano al Fentanyl quando questo è presente. La fetta relativa a "1 sostanza" (Fentanyl puro) è estremamente ridotta; la stragrande maggioranza dei grafici mostra combinazioni a 2, 3 o 4 sostanze raggruppate, indicando che il Fentanyl viene quasi sempre assunto in combinazione volontaria o involontaria con altre molecole.

- *Grafico a Barre Orizzontali - Sostanze Associate al Fentanyl*: elenca i co-attivatori chimici più frequenti nei decessi da Fentanyl. Le barre più lunghe mostrano la presenza concomitante di Cocaina (oltre 2.000 casi) ed Eroina. Questo incrocio rappresenta un punto di svolta analitico; infatti, dimostra che il Fentanyl si è infiltrato non solo nel mercato degli oppioidi, ma ha invaso massicciamente il mercato degli stimolanti. Ciò implica che una quota considerevole di vittime è deceduta per overdose da oppioidi senza avere una tolleranza biologica ad essi, essendo convinta di assumere esclusivamente uno stimolante, elemento che spiega l'esplosione improvvisa dei tassi di mortalità.

- *Lollipop Chart - Cluster Urbani a Maggiore Incidenza*: identifica i comuni più colpiti per la mortalità da Fentanyl. Hartford guida nettamente la classifica con 366 decessi, seguita da vicino da Waterbury, New Haven e Bridgeport. Questo grafico mappa con precisione chirurgica la geografia del rischio sintetico. Dimostra che la letalità del Fentanyl non si distribuisce in modo casuale sul territorio, ma è strettamente legata ai principali nodi urbani e ai mercati di strada storici del Connecticut.

CAPITOLO 6

Discussione

In questo capitolo vengono analizzati criticamente i risultati emersi dall'analisi quantitativa della crisi degli oppioidi nel Connecticut. Dopo una breve ricapitolazione dell'architettura del progetto, il capitolo organizza i risultati lungo quattro dimensioni complementari: storica, demografica, tossicologica, e territoriale-temporale. Per ciascuna dimensione, i dati grezzi sono posti in dialogo con il contesto scientifico ed epidemiologico più ampio, al fine di estrarre il significato interpretativo dei numeri e le implicazioni che essi comportano per la comprensione del fenomeno.

6.1 Architettura del percorso analitico

Il progetto si è articolato come un percorso di esplorazione progressiva, costruito attorno a una logica narrativa Top-Down, dal macroscopico al granulare, dal contesto federale alla singola vittima. La prima fase ha riguardato la raccolta, la pulizia e la modellazione dei dati. Sulla base di questo modello è stato costruito un ecosistema di sette dashboard interattive sviluppate in Power BI, arricchite da misure DAX create per rispondere agli specifici obiettivi del progetto. Le prime tre dashboard assumono una funzione descrittiva ed esplorativa, tracciando la linea evolutiva della crisi su scala federale e il profilo sociodemografico generale delle vittime nel Connecticut. Le successive quattro hanno, invece, operato a un livello diagnostico e correlazionale, indagando la mobilità territoriale delle vittime, l'architettura del policonsumo, la stagionalità dei decessi e l'impatto specifico del Fentanyl come principale motore della terza ondata epidemica.

6.2 Sintesi e interpretazione dei risultati principali

Per descrivere appieno la complessità multidimensionale della crisi degli oppioidi, la presentazione dei risultati è stata organizzata secondo quattro assi analitici complementari. Una dimensione storica illustra l'evoluzione temporale del fenomeno a livello nazionale e statale, tracciandone la progressione epidemiologica e il passaggio tra le tre ondate successive. Una dimensione demografica approfondisce i profili delle vittime (sesso, età, etnia) rivelando le vulnerabilità specifiche entro la popolazione. Una dimensione tossicologica disseziona l'architettura chimica dell'overdose contemporanea, mappando le sostanze, le loro combinazioni letali e il ruolo dominante del Fentanyl. Infine, una dimensione territoriale e temporale localizza geograficamente il fenomeno e ne descrive i ritmi stagionali e annuali.

Questa strutturazione consente di passare da una comprensione puramente numerica a un'interpretazione critica della realtà.

6.2.1 Dimensione storica

L'analisi del dataset nazionale ha documentato una crescita inarrestabile della mortalità da oppioidi nell'arco di un quindicennio. Tra il 1999 e il 2014, i decessi annui a livello federale sono passati da circa 10.000 a quasi 30.000, con un incremento costante e privo di flessioni apprezzabili. Il dato aggregato di 289.000 decessi totali nello stesso periodo restituisce la dimensione assoluta della catastrofe sanitaria. La ridotta ampiezza degli intervalli di confidenza al 95% esclude che la crescita osservata sia riconducibile a oscillazioni casuali, confermandone il carattere sistemico e strutturale.

A livello statale emerge una duplice lettura del fenomeno. In termini assoluti, stati demograficamente imponenti, come California, Florida e New York, registrano il maggior numero di decessi, identificando le aree con il carico gestionale più consistente. Tuttavia, normalizzando il Crude Rate per 100.000 abitanti, la graduatoria si inverte radicalmente. Stati meno popolosi, come il West Virginia, risultano tra i più colpiti in termini di rischio reale per la popolazione residente, dimostrando che la densità demografica non è il fattore determinante della vulnerabilità.

A livello locale, il Connecticut ha registrato una media di Crude Rate pari a 14,80 su una popolazione cumulata di 56 milioni di residenti nell'arco temporale considerato, con un volume complessivo di 4.000 decessi e 3.000 prescrizioni di oppioidi. L'andamento è stato caratterizzato da una fase di relativa stabilità fino al 2012, seguita da un'impennata drammatica negli anni successivi. Nel decennio successivo (2012-2021), la mortalità ha subito un'ulteriore accelerazione ancora più violenta. Si è passati da 311 decessi annui nel 2012 a oltre 1.000 nel 2021, con una quadruplicazione della mortalità in meno di dieci anni. Lo shock storico di questa escalation è reso particolarmente evidente dall'andamento mensile: dalle 20-30 vittime mensili del 2012 ai picchi di 138 decessi nel luglio 2021 e 135 ad agosto dello stesso anno.

6.2.2 Dimensione demografica

Il profilo sociodemografico delle vittime presenta caratteristiche marcate e coerenti. La vittima tipo è un uomo (74,13% dei casi) di etnia prevalentemente bianca, con un'età media di 43 anni. La distribuzione per fasce d'età mostra un picco di massima concentrazione nella fascia 35-55 anni, smentendo il cliché mediatico che associa l'overdose ai giovani. La crisi degli oppioidi colpisce soprattutto gli adulti in piena età lavorativa, con conseguenze devastanti sulle famiglie e sul tessuto economico delle comunità.

La disparità di genere è una costante trasversale. Gli uomini rappresentano circa il triplo dei casi in quasi tutte le categorie di oppioidi, cocaina ed eroina. L'unica eccezione parziale riguarda le benzodiazepine, dove lo scarto tra i generi si riduce, suggerendo pattern di consumo o di prescrizione differenti nelle donne, probabilmente legati a una maggiore frequenza di prescrizioni terapeutiche per ansia e disturbi del sonno. Nel profilo specifico dei decessi da Fentanyl, la sproporzione di genere si accentua ulteriormente: il 77,04% delle vittime è di sesso maschile. La distribuzione delle sostanze varia significativamente in funzione dell'età. Il Fentanyl domina in modo schiacciante nelle fasce giovanili e adulte (20-40 anni), mentre nelle età più avanzate (60 anni e oltre) mantengono una quota rilevante sostanze più tradizionali. Per il Fentanyl, nello specifico, il picco assoluto di mortalità si concentra nella fascia dei 30-39 anni. Il dato sulla vittima più giovane positiva al Fentanyl, 14 anni, segnala come la penetrazione della sostanza attraverso pillole contraffatte abbia

iniziato a raggiungere le fasce adolescenziali, aprendo scenari di rischio ancora difficili da quantificare pienamente.

6.2.3 Dimensione tossicologica

L'analisi tossicologica restituisce un quadro dettagliato dell'architettura chimica dell'overdose contemporanea. Il primo risultato strutturale riguarda la natura intrinsecamente multi-molecolare del fenomeno. La stragrande maggioranza dei casi presenta positività simultanea a 2, 3 o 4 sostanze o più, confermando che l'overdose letale è quasi sempre il prodotto di interazioni biochimiche sinergiche tra più principi attivi.

Il secondo risultato riguarda la posizione dominante e trasversale del Fentanyl nell'epidemia. Con una presenza certificata nel 66% di tutti i decessi registrati nel Connecticut tra il 2012 e il 2021, l'oppioide sintetico ha conquistato una posizione di quasi monopolio della letalità. La svolta drammatica avviene nel 2015, quando il Fentanyl irrompe sul mercato di strada sostituendo l'eroina, sia come sostanza primaria sia come agente di taglio. Da quel momento la crescita dei decessi diventa quasi verticale: su 6.000 morti totali registrate nel periodo, ben 4.000 sono legate direttamente a questa sostanza.

Il terzo risultato riguarda le combinazioni di sostanze più ricorrenti. L'incrocio tra Fentanyl e Cocaina registra 1.570 casi, mentre quello tra Fentanyl ed Eroina ne conta 1.072: questi due binomi rappresentano le combinazioni più letali. L'oppioide sintetico funge da nucleo centrale dell'intera rete tossicologica, legandosi a ogni altra sostanza presente sul mercato illecito.

6.2.4 Dimensione territoriale e temporale

L'analisi geografica ha rivelato come la crisi risponda a precise geografie della vulnerabilità urbana nel territorio del Connecticut. I cluster di mortalità si concentrano nei principali centri urbani dello stato, confermando che l'epidemia, pur essendo diffusa sul territorio, non si distribuisce in modo casuale ma segue le linee di forza dei mercati di approvvigionamento e delle reti di consumo.

Un dato significativo riguarda la mobilità epidemiologica: l'80,22% dei decessi avviene all'interno del comune di residenza della vittima, ma il restante 19,78%, quasi un decesso su cinque, si verifica fuori comune. Questo 20% di mobilità intercomunale rappresenta un dato di rilievo per le politiche sanitarie, dimostrando che il rischio di overdose non si esaurisce entro i confini amministrativi di residenza.

Il dato più significativo dell'intera analisi territoriale riguarda il contesto fisico del decesso. La categoria "Casa" registra oltre 4.000 casi su un totale di circa 6.000, con un distacco netto rispetto a tutte le altre categorie (Ospedale, Hotel e Motel, Altro). La prevalenza della morte domestica è una costante strutturale dell'epidemia che si mantiene attorno al 50% o più in ogni singolo mese dell'anno, indipendentemente dalle condizioni climatiche o dai periodi di festività. Questo segnala una dinamica di consumo prevalentemente solitaria e isolata, con conseguenze dirette sulle probabilità di un intervento di soccorso tempestivo.

I comuni a maggiore incidenza di mortalità sono Hartford (366 decessi), seguito da Waterbury, New Haven e Bridgeport. Questi quattro centri urbani coincidono con i principali nodi storici dei mercati di strada del Connecticut, confermando la stretta correlazione tra la presenza di infrastrutture di spaccio consolidate e la distribuzione geografica della mortalità.

Sul piano della stagionalità, emerge un ritmo annuale distinto. Gli eventi critici aumentano progressivamente da inizio anno, raggiungono il picco massimo in estate, tra giugno e agosto. I dati calano leggermente tra febbraio e marzo e tornano a salire a dicembre. Nel complesso, il fenomeno sta crescendo di anno in anno; infatti, i valori del 2021 si mantengono stabilmente al di sopra di tutti gli anni precedenti.

6.3 Interpretazione critica dei risultati

La correlazione tra volume di prescrizioni mediche e numero di decessi non rappresenta una semplice coincidenza statistica. Costituisce una conferma empirica della tesi, largamente condivisa nella letteratura scientifica, secondo cui la prima ondata della crisi è stata innescata da un sistema di prescrizione farmaceutica sistematicamente deregolamentato nel corso degli anni Novanta. Il mercato illegale degli oppioidi non è nato nel vuoto, ma ha prosperato su una dipendenza fisicamente prodotta da farmaci legali, per poi sviluppare una propria autonomia strutturale quando l'accesso legale è stato progressivamente ristretto.

Il passaggio dalle prescrizioni mediche all'eroina, e infine al Fentanyl sintetico, non è stato solo un aumento dei consumi, ma un cambiamento radicale della natura stessa del problema. Il Fentanyl ha riscritto le regole del rischio. È fino a cento volte più potente della morfina e difficilissimo da tracciare chimicamente. Inoltre, poiché viene mescolato con altre sostanze stimolanti, espone al rischio di overdose anche chi non ha mai usato oppioidi e non ha alcuna resistenza fisica. Questa contaminazione invisibile del mercato spiega, in gran parte, l'impennata improvvisa dei decessi registrata dal 2015 in poi.

I dati sulla prevalenza dei decessi in casa suggeriscono l'insufficienza delle strategie di intervento attuali. Se le persone muoiono soprattutto in contesti privati e isolati, i servizi concentrati nei luoghi pubblici o nei centri di aggregazione non bastano. La sfida è raggiungere le persone laddove si trovano, con strumenti preventivi e di riduzione del danno che operino anche a domicilio.

CAPITOLO 7

Conclusioni

In questa tesi abbiamo condotto un'analisi approfondita della crisi degli oppioidi nel Connecticut nel periodo 2012 - 2021, integrando dati storici nazionali dal 1999, con l'obiettivo di comprendere questo fenomeno nella sua multidimensionalità. L'analisi ha confermato che la crisi degli oppioidi non è riducibile a singole cause, bensì rappresenta un fenomeno strutturale e progressivamente mutevole. Le tre ondate epidemiche (prescrizioni mediche, eroina e Fentanyl), non costituiscono tre problemi distinti, ma tre capitoli di una medesima storia sanitaria, ciascuno generato dal precedente, con intensità crescente. Abbiamo osservato che il rischio non è uniformemente distribuito sulla popolazione, infatti, le geografie della mortalità rivelano concentrazioni precise attorno ai nodi storici dei mercati di approvvigionamento, mentre i profili demografici mostrano vulnerabilità specifiche legate al genere, all'età e, presumibilmente, alla classe socioeconomica. La complessità tossicologica contemporanea, dominata dalla diffusione del Fentanyl, richiede una ridefinizione radicale delle strategie di intervento, poiché il Fentanyl non è semplicemente una droga più potente ma un agente chimico che trasforma il significato stesso della contaminazione del mercato illecito. Infatti, una singola pillola contraffatta può risultare letale per chi non ha sviluppato tolleranza, alterando profondamente il calcolo del rischio individuale e collettivo. I dati relativi alle morti in contesto domestico spingono, inoltre, a riconsiderare le strategie tradizionali di riduzione del danno, in quanto le persone muoiono prevalentemente sole in casa, uno scenario che i servizi di riduzione del danno convenzionali non intercettano adeguatamente. Abbiamo, quindi, riconosciuto che la crisi riguarda la salute di tutta la popolazione e non solo i singoli pazienti affetti da dipendenza, pertanto, data l'enormità del fenomeno, sono necessari interventi strutturali che agiscano simultaneamente sul mercato illegale, sulla prescrizione medica, sulla disponibilità di abitazioni sicure e sul legame sociale tra le persone.

Le implicazioni per le politiche sanitarie che ne derivano indicano tre direzioni principali di intervento. In primo luogo, abbiamo evidenziato la necessità di bloccare le sostanze pericolose sul mercato attraverso una collaborazione internazionale per tracciare la provenienza delle sostanze chimiche illegali, affiancata da test rapidi e accessibili che consentano ai consumatori di scoprire se una droga sia stata contaminata. In secondo luogo, è indispensabile portare i servizi di aiuto direttamente sul territorio; poiché i dati si concentrano in zone specifiche, l'assistenza non può restare confinata in presidi ospedalieri lontani dalle comunità, ma deve radicarsi nei quartieri più a rischio attraverso operatori di strada, unità mobili per la distribuzione del naloxone e programmi di inserimento abitativo immediato. In terzo luogo, l'azione deve fondarsi su un'unione strategica tra cure mediche e dati reali,

prestando particolare attenzione alle differenze della popolazione; le donne, ad esempio, pur registrando meno decessi, presentano fragilità cliniche e sociali specifiche legate all'uso prolungato di farmaci prescritti, alla vergogna sociale e a una maggiore difficoltà nel chiedere aiuto.

Lo studio presentato contiene, tuttavia, alcuni limiti che è necessario esplicitare e che aprono prospettive future di ricerca. L'analisi utilizza dati di mortalità certificata, il che significa che esclude tutte le persone con dipendenza in vita: una comprensione più piena della crisi richiederebbe l'integrazione con dati su diagnosi di disturbo da uso di oppioidi, ricoveri e utilizzo di servizi. Il dataset, inoltre, non include variabili socioeconomiche dirette, come reddito, istruzione e situazione occupazionale, che, pur essendo in parte inferibili dalla geografia, meriterebbero un'analisi esplicita. Infine, questa ricerca si fonda su dati retrospettivi; infatti, per identificare i fattori di rischio con maggiore precisione e valutare l'efficacia reale degli interventi, sarebbe necessario uno studio prospettico che segua le persone nel tempo. Le prospettive future di ricerca includono, pertanto, l'estensione temporale dell'analisi fino al 2024, al fine di valutare l'impatto della pandemia da COVID-19 sui pattern di consumo e di mortalità, nonché un'analisi comparativa tra stati con diverse strategie di riduzione del danno, ad esempio il Connecticut e il Vermont, che potrebbe fornire insegnamenti sui modelli di intervento più efficaci.

Oltre ogni considerazione tecnica e epidemiologica, rimane un dovere morale: i 6000 decessi registrati nel Connecticut tra il 2012 e il 2021 non sono astrazioni statistiche, bensì vite interrotte, famiglie dilaniate, comunità ferite. La visualizzazione dei dati applicata alla sanità pubblica ha il compito di trasformare il dato in realtà percepibile e la realtà percepibile in pressione per l'azione. Questo lavoro intende contribuire a quella trasformazione. Le sette dashboard costruite nel corso del progetto, i risultati analitici qui discussi, le interpretazioni critiche proposte rimangono strumenti al servizio di un'intenzione semplice ma urgente: aiutare chi progetta e gestisce gli interventi sanitari a comprendere con precisione la geometria della crisi, per rispondere con strategie all'altezza della sua complessità. La sfida finale rimane quella di trasformare l'informazione in decisioni, e le decisioni in protezione effettiva delle vite vulnerabili.

Bibliografia

ATZENI, P., CERI, S., PARABOSCHI, S. e TORLONE, R. (2018), *Basi di dati*, McGraw-Hill Education, 5. ed. ed.

EMC EDUCATION SERVICES (2015), *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, John Wiley & Sons.

FERRARI, A. e RUSSO, M. (2016), *Introducing Microsoft Power BI*, Microsoft Press.

FEW, S. (2009), *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Analytics Press.

KNAFLIC, C. N. (2015), *Storytelling with Data: A Data Visualization Guide for Business Professionals*, John Wiley & Sons.

MACHIRAJU, S. e GAURAV, S. (2018), *Power BI Data Analysis and Visualization*, Apress.

MAURO, A. D. (2019), *Big Data Analytics. Analizzare e interpretare dati con il machine learning*, Apogeo.

REZZANI, A. (2017), *Big Data Analytics. Il manuale del data scientist*, Apogeo Education.

ROTHMAN, K. J. (2012), *Epidemiology: An Introduction*, Oxford University Press, 2. ed. ed.

SENSINI, S. (2021), *Basi di dati. Tecnologie, architetture e linguaggi per database*, Guida Completa, Apogeo.

TUFTE, E. R. (2001), *The Visual Display of Quantitative Information*, Graphics Press, 2. ed. ed.

Sitografia

- Analytics big data: cosa sono e come sono usati – https://www.zerounoweb.it/big-data/analytics-big-data-cosa-sono-e-come-sono-usati/m
- Big Data Analytics, tutte le prerogative di uno strumento di successo – https://www.zerounoweb.it/big-data/big-data-analytics-tutte-le-prerogative-di-uno-strumento-di-successo/g
- Che cos'è la big data analytics?– https://www.ibm.com/it-it/think/topics/big-data-analytics
- Che cos'è Microsoft Power BI–https://www.opta.it/digital-transformation/business-intelligence/microsoft-power-bi/
- Cosa sono i Big Data? – https://www.oracle.com/it/big-data/what-is-big-data/
- Data science: cos'è e in quali ambiti si utilizza – https://www.zerounoweb.it/big-data/data-science/
- Dataset 1: Accidental Deaths Related to Drugs(2012-2021) – https://www.kaggle.com/datasets/nvarisha/accidental-deaths-related-to-drugs
- Dataset 2: US Opioid Overdose Deaths – https://www.kaggle.com/datasets/thedevastator/us-opioid-overdose-deaths
- La doppia faccia degli oppioidi: tra medicina e dipendenza – https://www.marionegri.it/magazine/oppioidi$0
- Microsoft named a Leader in the 2025 Gartner® Magic Quadrant™ for Analytics and BI Platforms– https://learn.microsoft.com/it-it/power-bi/guidance/center-of-excellence-microsoft-business-intelligence-transformation
- Oppioidi: l'epidemia a stelle e strisce – https://www.saluteinternazionale.info/2024/11/oppioidi-lepidemia-a-stelle-e-strisce/
- Panoramica dell'integrazione di Copilot in Power BI – https://learn.microsoft.com/it-it/power-bi/create-reports/copilot-integration$0

- Power BI: Cos'è e Come Trasforma i Tuoi Dati in Azioni –https://finprimeinstitute.com/business-intelligence/power-bi-cose/#cos-e-power-bi
- Power Query: cos'è e perché è importante – https://visualitics.it/power-query/
- Sicurezza di Power BI – https://learn.microsoft.com/it-it/fabric/enterprise/powerbi/service-admin-power-bi-security$0
- Trasformazione BI di Microsoft– https://learn.microsoft.com/it-it/power-bi/guidance/center-of-excellence-microsoft-business-intelligence-transformation

Ringraziamenti

Un sincero ringraziamento va al Professor Domenico Ursino, la cui costante disponibilità e tempestività nelle risposte sono state una guida preziosa e rassicurante durante tutta la stesura di questo lavoro.

Desidero dedicare il più profondo pensiero alla mia famiglia, il mio porto sicuro e i miei fan numero uno. Mi avete permesso di intraprendere questo percorso e mi avete sostenuta in ogni momento di difficoltà, dandomi la forza e la motivazione per continuare a provarci sempre. Non immaginate quanto mi riempisse il cuore vedere la gioia nei vostri occhi, fieri e sorridenti, ogni volta che festeggiavamo un esame andato bene.

Un ringraziamento speciale è dedicato a mia nonna Iole: i tuoi pranzi domenicali non sono stati solo una tradizione, ma una vera e propria ricarica di energia, fondamentale per affrontare i giorni di studio. Tra un piatto e l'altro, mi hai insegnato i segreti della cucina e, soprattutto, mi hai trasmesso le lezioni di vita più importanti.

Questo viaggio, tuttavia, non sarebbe stato lo stesso senza gli amici straordinari che ho avuto la fortuna di avere accanto e che mi hanno fatto venire voglia di andare all'università ogni giorno (persino quando non c'erano lezioni). Grazie a Sara, Sere e Bea per aver colorato questi anni di risate inopportune, passeggiate strategiche tra i corridoi e pause caffè al bar, ormai diventate un rituale. E grazie per i pranzi di Sere, sempre molto discutibili ma memorabili. Un grazie speciale a te, Rania: abbiamo condiviso tutto fin dal primo giorno, le stesse materie e gli stessi esami. Non hai mai esitato un istante a sostenermi, nemmeno nei momenti di assoluta stanchezza, e te ne sono immensamente grata.

Altrettanto fondamentale è stato il sostegno di Cristian, una persona importantissima, che è sempre stata presente in questo percorso. Mi hai ascoltata ripetere per gli esami, vista piangere quando andavano male, e hai vissuto la mia agitazione prima e dopo. Nonostante tutto, hai sempre creduto in me, anche quando io non ci riuscivo. Grazie per quei messaggi la sera prima di ogni esame, e grazie soprattutto per aver gioito dei miei successi come se fossero i tuoi.

Infine, il ringraziamento più importante lo devo a me stessa. Se oggi posso guardare a questo traguardo con orgoglio e profonda soddisfazione, è solo grazie alla mia determinazione, alla mia costanza e alla mia capacità di non arrendermi mai di fronte agli ostacoli.