

# Università Politecnica delle Marche

Facoltà di Ingegneria

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica e dell'Automazione

---



**Tesi di Laurea**

**Valutazione degli effetti del terremoto del Centro Italia nel  
consumo dei farmaci a partire dai dati di una farmacia di  
Amatrice**

**Evaluation of the effects of the earthquake in Central Italy  
on drug consumption starting from data from a pharmacy in  
Amatrice**

Relatore

Prof. Domenico Ursino

Correlatore

Luca Virgili

Candidato

Fabio Russo

---

**Anno Accademico 2020-2021**



---

# Indice

<b>Introduzione</b> .....	11
<b>1 Il terremoto del Centro Italia del 2016 e del 2017</b> .....	13
1.1 La sequenza sismica .....	13
1.2 Danni materiali, sociali ed economici .....	15
<b>2 Strumenti software utilizzati</b> .....	17
2.1 Software di Business Intelligence .....	17
2.1.1 Qlik Sense .....	18
2.1.2 Tableau .....	19
2.1.3 Power BI .....	20
2.2 Statsmodels .....	21
<b>3 Struttura dei dati di partenza</b> .....	23
3.1 Dataset delle vendite di una farmacia di Amatrice .....	23
3.2 Descrizione dei campi .....	24
<b>4 Analisi descrittiva e diagnostica</b> .....	29
4.1 Andamento delle vendite della farmacia .....	29
4.2 Effetti del terremoto del Centro Italia sulle vendite dei prodotti da farmacia .....	34
<b>5 Analisi predittiva</b> .....	39
5.1 Previsione del ricavato nel biennio 2021-2022 .....	39
5.2 Previsione delle vendite di specifiche categorie di prodotto .....	41
<b>6 Analisi prescrittiva e Machine Learning</b> .....	45
6.1 Premessa .....	45
6.2 Correlazione .....	45
6.2.1 Correlazione tra le categorie di prodotto nel periodo pre-terremoto .....	46
6.2.2 Correlazione tra le categorie di prodotto nel periodo post-terremoto .....	46
6.3 Classificazione .....	48

4	Indice	
	6.3.1	Classificazione dei farmaci in base al prezzo di vendita . . . . . 48
	6.3.2	Classificazione dei farmaci in base alla quantità venduta . . . . . 49
6.4	Clustering . . . . .	52
	6.4.1	Segmentazione dei prodotti e analisi degli outlier . . . . . 52
<b>7</b>	<b>Analisi delle serie temporali . . . . .</b>	<b>55</b>
	7.1	Premessa . . . . . 55
	7.2	Analisi dei componenti della serie temporale . . . . . 56
	7.3	Costruzione di un modello per la previsione della serie temporale . . . 58
<b>8</b>	<b>Conclusioni . . . . .</b>	<b>63</b>
	<b>Riferimenti bibliografici . . . . .</b>	<b>65</b>
	<b>Ringraziamenti . . . . .</b>	<b>67</b>

---

## Elenco delle figure

1.1	Mappa epicentrale della sequenza sismica in Italia centrale aggiornata al 6 aprile 2018.....	14
1.2	Il centro di Amatrice dopo il sisma del 24 agosto 2016.....	15
2.1	Esempio di dashboard prodotta da Qlik Sense.....	18
2.2	Esempio di dashboard prodotta da Tableau.....	19
2.3	Esempio di dashboard prodotta da Power BI.....	20
2.4	Logo ufficiale di Statsmodels.....	22
4.1	Key Performance Indicator relativi alle entrate effettive <b>(a)</b> e stimate <b>(b)</b> nel periodo 2015-2020 (Qlik Sense).....	30
4.2	Diagramma a torta delle percentuali di ricavo per categoria di prodotto (Power BI).....	30
4.3	Grafico a barre relativo agli incassi annui dal 2015 al 2020 (Qlik Sense)	31
4.4	Grafico lineare dell'andamento mensile degli incassi per ogni anno (Qlik Sense).....	31
4.5	Diagramma ad anello relativo alle percentuali di quantità venduta per categoria di prodotto (Power BI).....	32
4.6	Mappa ad albero relativa al numero di confezioni vendute per categoria di prodotto ed anno (Power BI).....	33
4.7	Grafico a dispersione delle vendite nella fascia oraria 9-13:30 (Power BI).....	33
4.8	Grafico a dispersione delle vendite nella fascia oraria 16-19:30 (Power BI).....	34
4.9	Grafico a barre dell'andamento delle vendite di farmaci ansiolitici (Tableau).....	35
4.10	Grafico a barre dell'andamento delle vendite di farmaci antidepressivi (Tableau).....	35
4.11	Grafico a barre multiplo dell'andamento delle vendite di farmaci e parafarmaci veterinari (Tableau).....	36
4.12	Grafico a barre dell'andamento delle vendite dei prodotti cosmetici (Tableau).....	37

4.13	Grafico a barre multiplo dell'andamento delle vendite dei prodotti per bambini (Tableau) . . . . .	37
4.14	Grafico a barre dell'andamento delle vendite degli integratori (Tableau) . . . . .	38
5.1	Previsione dell'andamento mensile del ricavato nel biennio 2021-2022 (Tableau) . . . . .	40
5.2	Previsione del ricavato per categoria di prodotto per il biennio 2021-2022 (Tableau) . . . . .	40
5.3	Previsione dell'andamento delle vendite degli ansiolitici per il biennio 2021-2022 (Power BI) . . . . .	41
5.4	Previsione dell'andamento delle vendite degli antidepressivi per il biennio 2021-2022 (Power BI) . . . . .	42
5.5	Previsione dell'andamento delle vendite dei farmaci veterinari per il biennio 2021-2022 (Power BI) . . . . .	42
5.6	Previsione dell'andamento delle vendite dei prodotti per bambini per il biennio 2021-2022 (Power BI) . . . . .	43
6.1	Mappa di calore delle correlazioni tra le categorie di prodotto nel periodo pre-terremoto (Tableau) . . . . .	47
6.2	Mappa di calore delle correlazioni tra le categorie di prodotto nel periodo post-terremoto (Tableau) . . . . .	47
6.3	Tabella di classificazione delle categorie di prodotto in base al ricavato totale (Tableau) . . . . .	50
6.4	Tabella di classificazione delle categorie di prodotto in base alla quantità venduta (Tableau) . . . . .	51
6.5	Segmentazione dei prodotti in base al ricavato e alla quantità venduta (Power BI) . . . . .	52
7.1	Grafico della serie temporale delle vendite della farmacia di Amatrice	57
7.2	Grafici dei componenti della serie temporale ottenuti tramite decomposizione moltiplicativa ed additiva . . . . .	58
7.3	Equazioni del metodo di Holt . . . . .	59
7.4	Metodo additivo di Holt-Winters . . . . .	59
7.5	Metodo moltiplicativo di Holt-Winters . . . . .	60
7.6	Applicazione del modello predittivo della serie temporale . . . . .	61

---

## Elenco delle tabelle

3.1	Tabella dei campi del dataset <code>DatiVenditeFarmacia.csv</code> .....	24
3.2	Tabella dei valori del campo <i>Suddivisione Merceologica</i> .....	27





---

## Elenco dei listati

6.1	Definizione delle misure <b>Quantile 0.33 Prezzo</b> e <b>Quantile 0.66 Prezzo</b> .....	49
6.2	Definizione della dimensione <b>Classe Di Prezzo</b> .....	49
6.3	Definizione delle misure <b>Quantile 0.33 Quantità</b> e <b>Quantile 0.66 Quantità</b> .....	50
6.4	Definizione della dimensione <b>Classe Di Vendita</b> .....	50
7.1	Generazione del grafico della serie temporale .....	56
7.2	Decomposizione moltiplicativa e additiva della serie temporale .....	57
7.3	Definizione del modello predittivo della serie temporale.....	60



---

## Introduzione

Il 24 agosto 2016, alle ore 3:36, un terremoto di magnitudo 6.0, con epicentro ad Accumoli (RI), ha dato inizio ad una sequenza sismica che, nel giro di appena quattro mesi, ha portato alla distruzione di molti comuni di Lazio, Marche, Umbria e Abruzzo, situati nell'Appennino Centrale.

In particolare, Amatrice è risultato il paese maggiormente colpito da questa immane catastrofe naturale, con un numero di vittime pari a 238 e con il 95% del patrimonio edilizio danneggiato. Di conseguenza, l'assenza di abitazioni e di strutture che potessero ospitare i turisti ha comportato un ingente spopolamento negli anni successivi, che ha causato un notevole decremento dei profitti delle numerose attività commerciali ed artigianali del posto.

Tale cataclisma ha prodotto non solo danni a livello materiale ed economico, ma anche a livello psicologico e sociale. Il disturbo da stress post-traumatico ha, infatti, inciso in maniera preponderante sui comportamenti e sulle abitudini della popolazione, cambiamenti che si sono riflessi principalmente sui consumi.

Per tutti questi motivi, si è deciso di effettuare uno studio dei dati delle vendite di una farmacia di Amatrice, con l'obiettivo di valutare in che modo il sisma abbia inciso sul consumo di determinati farmaci e prodotti. L'analisi dei dati permette, infatti, di ottenere importanti risultati di business, come:

- comprendere la reazione del mercato a seguito di eventi dal forte impatto;
- descrivere situazione economica attuale e passata di un'impresa;
- identificare lo status emotivo e psicologico dei clienti;
- individuare delle correlazioni tra le vendite di specifici prodotti;
- prevedere l'andamento futuro delle vendite.

Il dataset degli scontrini emessi nel periodo 2015-2020, che tale farmacia di Amatrice ha messo a disposizione per questo progetto di tesi, è stato ritenuto utile al fine di evidenziare il cambiamento dello status economico dell'attività commerciale dovuto al sisma, e di riflettere i problemi che quest'ultimo ha indotto nel territorio.

In questo modo, i risultati ottenuti dall'analisi del suddetto dataset potrebbero essere utilizzati per definire strategie di marketing con l'obiettivo di incrementare le rendite della farmacia e di individuare delle soluzioni che possano favorire la ripresa del paese sotto diversi contesti.

Per questo lavoro di tesi sono stati impiegati tre software di *Business Intelligence*, ovvero Qlik Sense, Tableau e Power BI, per generare dei grafici utili a descrivere l'evoluzione della situazione economica dell'attività commerciale nel periodo considerato e a comprendere le cause alla base delle variazioni delle vendite di particolari prodotti.

Si è rilevato che il sisma del Centro Italia ha provocato un decremento di oltre un terzo del fatturato della farmacia, dovuto, principalmente, ad un abbandono del paese da parte sia dei residenti che dei turisti.

Inoltre, lo stress conseguente ad un evento catastrofico di tale portata e le difficili condizioni di vita hanno, purtroppo, prodotto un incremento del consumo di psicofarmaci da parte della popolazione.

Un altro problema emerso dallo studio condotto riguarda le difficoltà riscontrate dalle numerose aziende agricole, le quali costituiscono un settore molto importante dell'economia del territorio, nel mantenimento dei capi di bestiame, che si riflette in una diminuzione del numero di farmaci veterinari venduti.

Il passo successivo è stato quello di fornire delle previsioni riguardanti l'andamento delle vendite nel biennio 2021-2022, e, alla luce dei risultati ottenuti, si può presumere che sia possibile un moderato aumento del fatturato, probabilmente dovuto al completamento dei primi progetti di ricostruzione e ristrutturazione delle abitazioni.

Nell'ultima fase dell'analisi, partendo dalle informazioni ottenute in precedenza, si è cercato, ricorrendo a delle tecniche di *Machine Learning*, di individuare le categorie di prodotti sulle quali la farmacia dovrebbe fondare le sue strategie di marketing.

Infine, per studiare i trend e i pattern stagionali delle vendite, è stata utilizzata la libreria di Python `Statsmodels`, la quale include diverse funzioni per l'analisi delle serie temporali.

La presente tesi è strutturata come di seguito specificato:

- Nel Capitolo 1 si parlerà della sequenza sismica del 2016-2017 e dei principali danni ad essa conseguenti.
- Nel Capitolo 2 verranno introdotti gli strumenti software utilizzati per svolgere l'analisi.
- Nel Capitolo 3 si descriverà il dataset delle vendite della farmacia oggetto dell'analisi.
- Nel Capitolo 4 verranno illustrati i risultati ottenuti nelle fasi descrittiva e diagnostica.
- Nel Capitolo 5 saranno fornite delle previsioni sulle vendite del biennio 2021-2022.
- Nel Capitolo 6 sarà trattata la fase di analisi prescrittiva, con una particolare attenzione ai metodi di Machine Learning adottati.
- Nel Capitolo 7 verranno discussi i risultati ottenuti dall'analisi delle serie temporali delle vendite della farmacia.
- Nel Capitolo 8 saranno tratte le conclusioni riguardanti il lavoro svolto.

## Il terremoto del Centro Italia del 2016 e del 2017

*In questo primo capitolo verrà descritta la sequenza sismica del terremoto del Centro Italia del 2016 e del 2017, illustrando le caratteristiche delle principali scosse. Si passerà, poi, ad un'analisi più dettagliata delle conseguenze del sisma, riportando i danni più gravi a livello materiale, sociale ed economico.*

### 1.1 La sequenza sismica

Gli eventi sismici del Centro Italia del 2016 e 2017, definiti dall'*Istituto Nazionale di Geofisica e Vulcanologia (INGV)* sequenza sismica Amatrice-Norcia-Visso, hanno avuto inizio ad agosto 2016 con epicentri situati tra l'alta valle del Tronto, i Monti Sibillini, i Monti della Laga e i Monti dell'Alto Aterno.

La prima forte scossa si è avuta il 24 agosto 2016, alle ore 3:36 e ha avuto una magnitudo di 6.0, con epicentro nel comune di Accumoli e ipocentro alla profondità di 8 km. La durata è stata di 15-20 secondi. Per quanto concerne l'accelerazione del suolo, nota come *peak ground acceleration*, si sono registrati valori di 0,45 g, con un picco massimo di 0,86 g registrato nella sola Amatrice, il che ha amplificato la violenza del sisma e aggravato i danni al patrimonio edilizio. Durante la notte sono state registrate numerose scosse nella zona norcina e in quella reatina, tra queste, varie superiori ai 4 gradi. Alle ore 4:33 una scossa di 5.3 gradi è stata registrata a Norcia, in provincia di Perugia.

La zona dell'evento sismico si trova in un'area sismologica molto attiva dell'Italia che comprende anche L'Aquila, dove il terremoto del 6 aprile 2009 (Mw 6.3) provocò oltre 300 morti e circa 65000 sfollati, oltre alle Marche e all'Umbria stessa, che subirono il terremoto del 26 settembre 1997 (Mw 6.0).

Due potenti repliche sono avvenute il 26 ottobre 2016 con epicentri al confine umbro-marchigiano, tra i comuni della provincia di Macerata di Visso, Ussita e Castelsantangelo sul Nera, la prima alle 19:11 con magnitudo 5.4, e la seconda alle 21:18 con magnitudo 5.9. Sono seguite una serie di scosse, le più forti di magnitudo compresa tra 3.0 e 4.5.

Il 30 ottobre alle 07:40 si è verificata una scossa di magnitudo momento 6.5, percepita in gran parte della penisola italiana, in Austria e sulle coste balcaniche, risultando la più forte scossa di terremoto in Italia dal sisma dell'Irpinia del 1980. Il

sisma si è originato a 9,4 km di profondità, con epicentro in provincia di Perugia tra i paesi di Norcia, Preci e Castelsantangelo sul Nera. In questo caso, i valori relativi all'accelerazione del suolo sono stati di 0,48 g, con picchi massimi di 0,76 g registrati ad Arquata del Tronto. In seguito si sono verificate repliche di magnitudo compresa tra 3.5 e 4.8.

Il 18 gennaio del 2017 si sono verificate quattro scosse molto intense in un arco di tempo poco superiore alle quattro ore: la prima alle 10:25 di magnitudo 5.1 con epicentro a Montereale; la seconda di magnitudo 5.5 alle 11:14 con epicentro a Capitignano; la terza alle 11:25 con magnitudo 5.4 con epicentro a Pizzoli; la quarta di magnitudo 5.0 alle 14:33 con epicentro a Cagnano Amiterno. Queste nuove scosse hanno coinvolto tutti i paesi già colpiti dai precedenti eventi sismici dell'Abruzzo, delle Marche, del Lazio e, in parte, dell'Umbria. Non si può escludere una correlazione, anche se non vi sono prove certe, con una valanga che si è prodotta nelle ore seguenti il sisma e che, alle 17:40 circa, si è abbattuta sull'Hotel Rigopiano a Farindola, in provincia di Pescara, distruggendolo e intrappolando 40 persone al suo interno.

Questo insieme di eventi provocò in tutto circa 41000 sfollati, 388 feriti e 303 morti, dei quali 3 morirono per via indiretta.

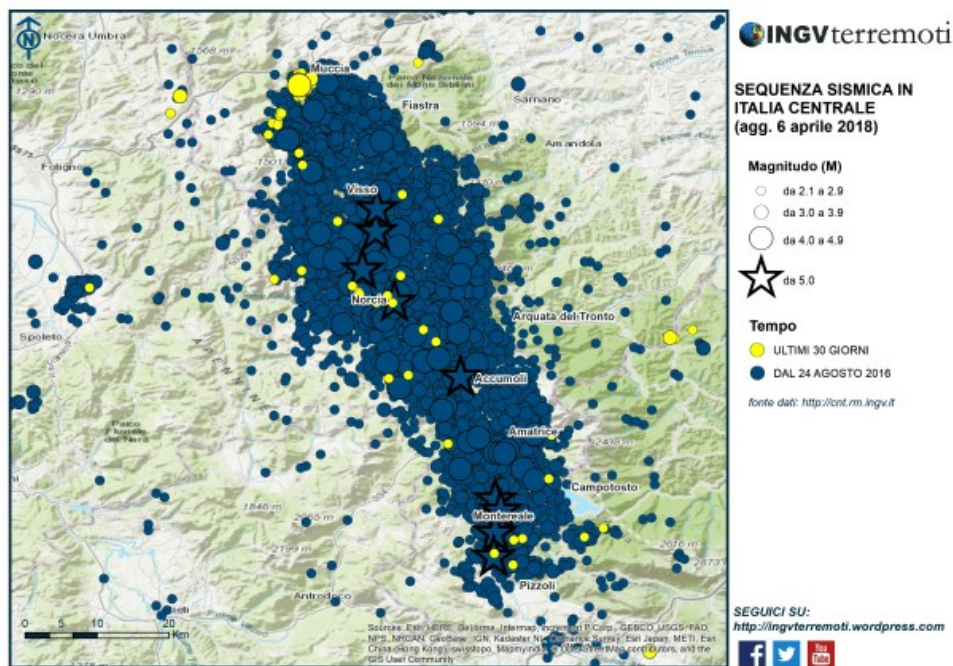


Figura 1.1: Mappa epicentrale della sequenza sismica in Italia centrale aggiornata al 6 aprile 2018

## 1.2 Danni materiali, sociali ed economici

Le aree più colpite dagli eventi sismici sono l'alta valle del Tronto, sede del sisma di magnitudo 6.0 avvenuto il 24 agosto, e la zona dei monti Sibillini, tra Umbria e Marche, dove sono avvenute le scosse del 26 e del 30 ottobre. Nei comuni di Amatrice e Arquata del Tronto si sono raggiunti danni pari all'XI grado della Scala Macrosismica Europea (*EMS*).

La Protezione Civile riporta che le vittime sono state 299, mentre sono state estratte vive dalle macerie 238 persone, alcune delle quali sono decedute in seguito; i feriti portati in ospedale sono invece 388. Il numero delle vittime è stato particolarmente alto perché i territori colpiti, che d'inverno sono abitati da un modesto numero di residenti, erano in quel momento nel pieno della stagione turistica e ospitavano un numero molto maggiore di persone, specialmente ad Amatrice. Tra le vittime, infatti, ci sono stati numerosi turisti ed ex residenti in villeggiatura nelle seconde case, provenienti soprattutto da Roma.



Figura 1.2: Il centro di Amatrice dopo il sisma del 24 agosto 2016

La scossa del 24 agosto, di magnitudo 6.0, distrusse i comuni di Amatrice (RI), Accumoli (RI) e Arquata del Tronto (AP), con le relative frazioni, in particolare Pescara del Tronto, frazione di Arquata, che fu completamente rasa al suolo, anche a causa degli eventi dei mesi successivi, che causarono una grossa frana. Diversi furono i crolli e le lesioni anche in molti comuni circostanti, compresi sempre tra le province di Rieti e Ascoli Piceno. È stato gravemente danneggiato, e conseguentemente evacuato, il principale ospedale dell'area laziale-marchigiana, il "Francesco Grifoni" di Amatrice.

Tra le vie di comunicazione hanno subito danni: la strada statale 4 Via Salaria, dove alcuni viadotti subirono spostamenti e in vari tratti la carreggiata fu ristretta

dalle frane; la strada regionale 260 Picente, con il traffico tra Amatrice e L'Aquila deviato sulla regionale 577 del Lago di Campotosto; la strada statale 685 delle Tre Valli Umbre, della quale furono chiusi ampi tratti.

Il terremoto ha causato anche vasti danneggiamenti al patrimonio culturale della zona. Ad Amatrice la via principale del corso umbertino è diventata il simbolo della devastazione della scossa del 24 agosto: soltanto la torre civica rimase in piedi, assieme alla chiesa di Sant'Agostino, alle porte del centro, benché con la facciata parzialmente crollata. Ad Accumoli ci furono vari crolli nel centro storico, tra i quali quello più gravoso del campanile della chiesa parrocchiale. Anche il borgo di Arquata del Tronto risultò gravemente danneggiato, con il crollo della chiesa parrocchiale e delle case del centro storico. Soltanto la Rocca di Arquata fu interessata in maniera meno grave, con la caduta delle merlature delle torri.

Le due maggiori scosse di 5.4 e 5.9 hanno causato gravi danni ai comuni di Castelsantangelo sul Nera, Visso ed Ussita, nonché il crollo totale della chiesa di Santa Maria delle Grazie e della chiesa abbaziale di San Salvatore in frazione Campi di Norcia. A Camerino crollò il campanile del santuario di Santa Maria in Via. A Tolentino (MC) un anziano morì colpito da infarto.

La scossa del 30 ottobre ha creato danni gravissimi al patrimonio artistico e infrastrutturale di Norcia, come l'esonazione del fiume Nera sulla strada statale della Valnerina. Il simbolo dei danni subiti dalla città di Norcia è la rovina quasi totale della Basilica di San Benedetto, di cui restò in piedi solo la facciata, così come le mura medievali, la Cattedrale, la semidistruzione della frazione di Castelluccio di Norcia e della Chiesa di Sant'Andrea di Campi.

A causa delle frequenti scosse di terremoto, ad Amatrice vennero prodotti nuovi danni, come il collasso del palazzo municipale, rimasto in piedi, e della sommità della torre civica al corso. Anche un lato della chiesa di Sant'Agostino crollò, mentre una frana inghiottì una porzione del borgo di Arquata del Tronto, già pesantemente danneggiato dal sisma del 26 ottobre. Con la scossa del 30 ottobre, inoltre, venne ulteriormente danneggiato l'Ospedale "Vittorio Emanuele II" di Amandola, il quale fu dichiarato inagibile con la conseguente evacuazione e chiusura. Il sisma ha provocato uno spostamento di due lobi: uno verso est di 40 cm nell'area di Montegallo e l'altro verso ovest di 30 cm nella zona di Norcia. Sul Monte Vettore si aprì un'enorme spaccatura causando una frana.

Il sisma del 18 gennaio 2017 ha creato danni contenuti nei comuni di Montereale, Campotosto, Capitignano, Cortino e Valle Castellana. Campotosto è risultato il comune più colpito, con il crollo della torretta del palazzo municipale degli anni '50 e danni alla facciata della chiesa parrocchiale, già lesionata dal terremoto del 2009 e mai restaurata. Alcuni capannoni e rimesse di campagna sono crollate a Castel Castagna, dove si è registrata l'unica vittima del sisma. Ad Amatrice crollò definitivamente il campanile della chiesa di Sant'Agostino.

Nel complesso per l'intera sequenza sismica i danni ammontano a 23 miliardi e 530 milioni di euro, di cui 12,9 miliardi si riferiscono ai danni relativi agli edifici privati e 1,1 miliardi di euro agli edifici pubblici. La stima comprende i danni diretti, sia pubblici sia privati, ossia quelli che hanno provocato la distruzione di edifici, di infrastrutture, di raccolti e anche quelli che hanno colpito industrie e imprese, il patrimonio culturale, le reti di distribuzione dell'energia, del gas, dell'acqua e i costi eleggibili, sostenuti dallo Stato per far fronte all'emergenza.



## Strumenti software utilizzati

*Nel capitolo corrente verranno presentati i software di Business Intelligence utilizzati per questo progetto di tesi, ovvero Qlik Sense, Tableau e Power BI, descrivendone le principali funzionalità. Successivamente, verrà presentata Statsmodel, la libreria Python tramite la quale è stata effettuata l'analisi delle serie temporali.*

### 2.1 Software di Business Intelligence

Negli ultimi anni l'uso di strumenti di *Business Intelligence (BI)* è aumentato notevolmente, soprattutto grazie alla crescita dei volumi di dati a disposizione delle aziende, i cosiddetti *Big Data*, e dello sviluppo di nuove tecnologie nell'ambito della *Business Analysis*, anche su Cloud.

Per Business Intelligence si intendono, quindi, tutti i processi e gli strumenti attraverso i quali un'azienda riesce a raccogliere dati di diversa natura per analizzarli e trarne decisioni strategiche. Ogni giorno, infatti, le aziende e le organizzazioni producono una quantità davvero incredibile di dati, che spesso, però, sono inutili per la società, se non per lo scopo originario per cui sono stati creati. Tuttavia, con l'avanzamento di strumenti di BI come la Business Analytics, le aziende riescono ad utilizzare i dati raccolti per prendere decisioni strategiche in maniera sempre più veloce ed efficace.

Uno dei vantaggi più interessanti delle piattaforme di Business Intelligence è che vengono progettate appositamente per essere utilizzate anche da persone senza conoscenze tecniche. Ciò avviene grazie all'utilizzo di dashboard aziendali e indicatori di performance di facile comprensione, con diagrammi e grafici, per permettere anche a chi non ha alcuna conoscenza di analisi dei dati di avere un'immagine istantanea di ogni tipo di attività in pochi secondi. La capacità di reporting sotto forma di dashboard visive è una funzionalità fondamentale di molte piattaforme di BI, ed è stata sviluppata appositamente per essere compresa dai soggetti decisori.

Nella restante parte della sezione verranno descritti i 3 tool di Business Intelligence utilizzati nella parte di *Data Analytics*, ovvero Qlik Sense, Tableau e Power BI.

### 2.1.1 Qlik Sense

*Qlik Sense* è una piattaforma di *Business Analytics* che fornisce funzioni di Business Intelligence self-service per sviluppare velocemente dashboard completamente personalizzabili, in grado di fornire rapidamente informazioni utili sui dati a disposizione e di supportare executive, business leader e chiunque altro debba rimanere sempre aggiornato.

Tale software semplice, ma allo stesso tempo potente, consente di combinare, caricare, visualizzare ed esplorare facilmente i dati di interesse, indipendentemente dalle loro dimensioni. Tramite Qlik Sense è possibile porre domande, trovare risposte e seguire percorsi che permettono di ottenere informazioni strategiche per supportare quotidianamente il processo decisionale della propria organizzazione, creando un'azienda a tutti gli effetti orientata ai dati.

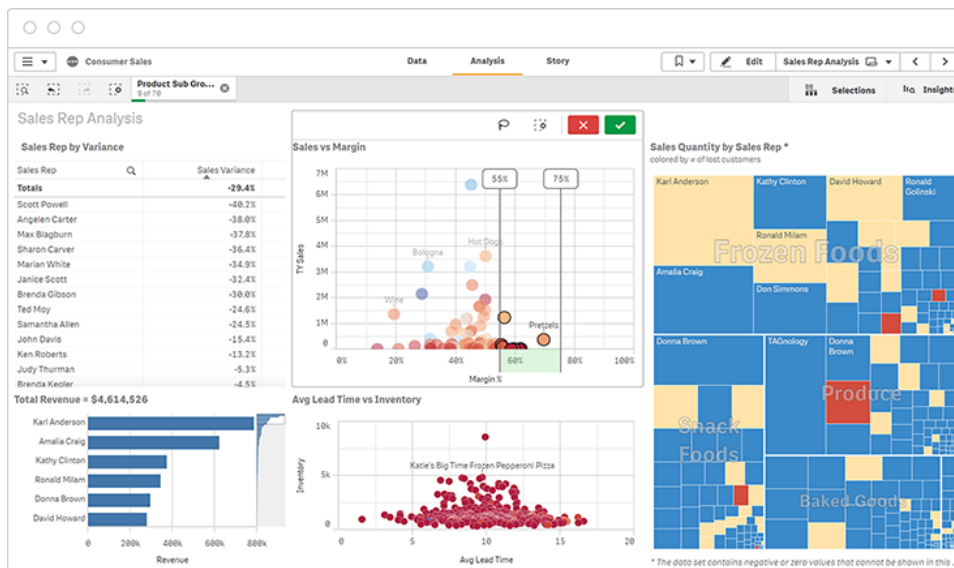


Figura 2.1: Esempio di dashboard prodotta da Qlik Sense

Il motore associativo di Qlik Sense permette agli utenti di esplorare liberamente in qualsiasi direzione, ricalcolando le analytics ed evidenziando le relazioni tra i dati dopo ogni click. Tutti i grafici, le tabelle e gli oggetti creati con Qlik Sense sono interattivi e si aggiornano istantaneamente in base al contesto attuale e a seguito di ogni azione.

Qlik Sense consente, inoltre, di collaborare con i colleghi, indipendentemente dal momento e dal luogo in cui si trovano. Tutte le funzionalità di Qlik Sense, compresa l'esperienza associativa e di collaborazione, sono disponibili sui dispositivi mobili. Con Qlik Sense è possibile porre domande, e quindi reperire risposte, anche per le domande di follow-up, con i colleghi, indipendentemente dall'ubicazione.

### 2.1.2 Tableau

*Tableau* è una piattaforma di analisi visiva che permette agli utenti di esplorare e gestire facilmente i dati, nonché di individuare e condividere velocemente informazioni di business funzionali ad uno sviluppo strutturato dell'organizzazione. Tale software fornisce funzionalità per l'esecuzione di query su database relazionali, cubi di elaborazione analitica online, database in Cloud e fogli di calcolo per generare visualizzazioni di dati di tipo grafico interattive ed altamente personalizzabili.

Fondata nel 2003, Tableau è una società nata, in realtà, come progetto universitario di Computer Science all'interno di *Stanford* che aveva l'obiettivo, fin dall'inizio, di migliorare il flusso di analisi e rendere i dati più accessibili per le persone attraverso le visualizzazioni e le rappresentazioni grafiche dei dati. I cofondatori Chris Stolte, Pat Hanrahan e Christian Chabot hanno sviluppato e brevettato la tecnologia fondamentale di Tableau, *VizQL*, un sistema che "esprime" i dati visivamente, traducendo le azioni drag-and-drop in query di dati attraverso un'interfaccia intuitiva e facile da utilizzare.

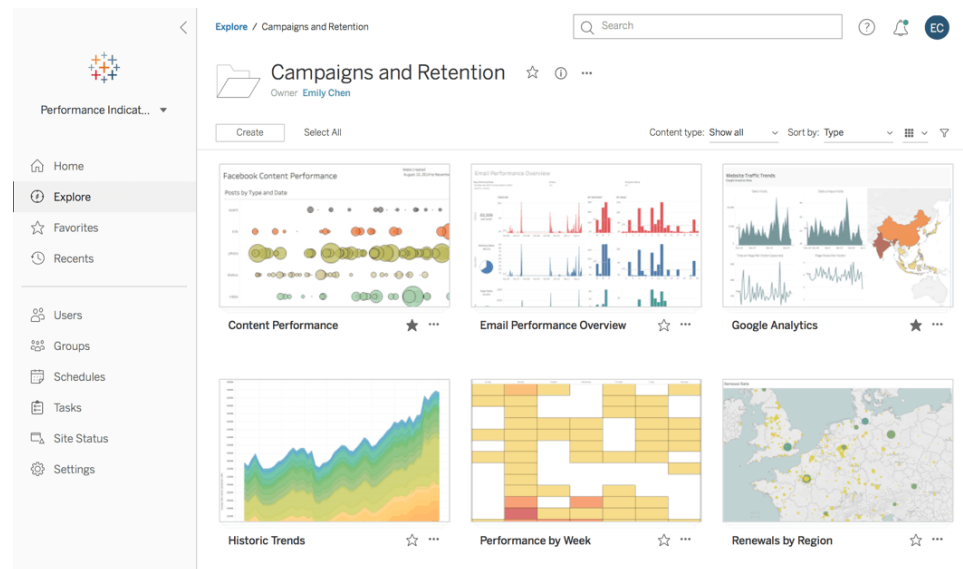


Figura 2.2: Esempio di dashboard prodotta da Tableau

La piattaforma consente di fare un'analisi end-to-end di tutti i dati aziendali; con funzionalità molto semplici di drag-and-drop, le persone possono raccogliere i dati da interrogare per condurre le analisi di cui necessitano da qualsiasi fonte, indipendentemente dal formato originario del dato. La piattaforma integra nativamente strumenti avanzati che consentono di normalizzare i dati, ossia di ripulirli ed integrarli in modo che possano sempre essere sfruttati da chi ne ha bisogno.

La parte di *Data Visualization* è ciò che più supporta le persone nei processi di Business Intelligence e analisi dei dati; la piattaforma ha potenti capacità di elaborazione dei dati che restituisce in differenti forme grafiche, come dashboard

interattive, grafici e cruscotti informativi di semplice interpretazione e condivisione, in grado di rispondere in modo intuitivo alle interrogazioni degli utenti, facendo, dunque, emergere le informazioni nascoste nei dati ed utili ai business manager.

### 2.1.3 Power BI

*Power BI* è una famiglia di strumenti per l'analisi aziendale prodotti da Microsoft, che forniscono visualizzazioni interattive e funzionalità di Business Intelligence con un'interfaccia semplice ed intuitiva, per consentire agli utenti di creare, a partire da dati aziendali, e condividere report ed informazioni a supporto delle decisioni. Tale software offre, inoltre, servizi di *Data Warehouse* tra cui preparazione dati, rilevazione dei dati e dashboard interattive.

L'applicazione è stata ideata da Thierry D'Hers e Amir Netz del team *SQL Server Reporting Services* di Microsoft. È stato progettato da Ron George nell'estate del 2010 e chiamato Project Crescent. Project Crescent si poteva scaricare già dall'11 luglio 2011, in bundle con SQL Server Codename Denali. Successivamente ribattezzato Power BI, è stato annunciato da Microsoft a settembre 2013 come Power BI per Office 365.

La prima versione di Power BI si basava su componenti aggiuntivi che utilizzavano Microsoft Excel: Power Query, Power Pivot e Power View. Dopo diverso tempo, Microsoft ha aggiunto molte funzionalità come FAQ, connettività dei dati a livello aziendale e opzioni di sicurezza tramite Power BI Gateway. Power BI è stato rilasciato per la prima volta il 24 luglio 2015.

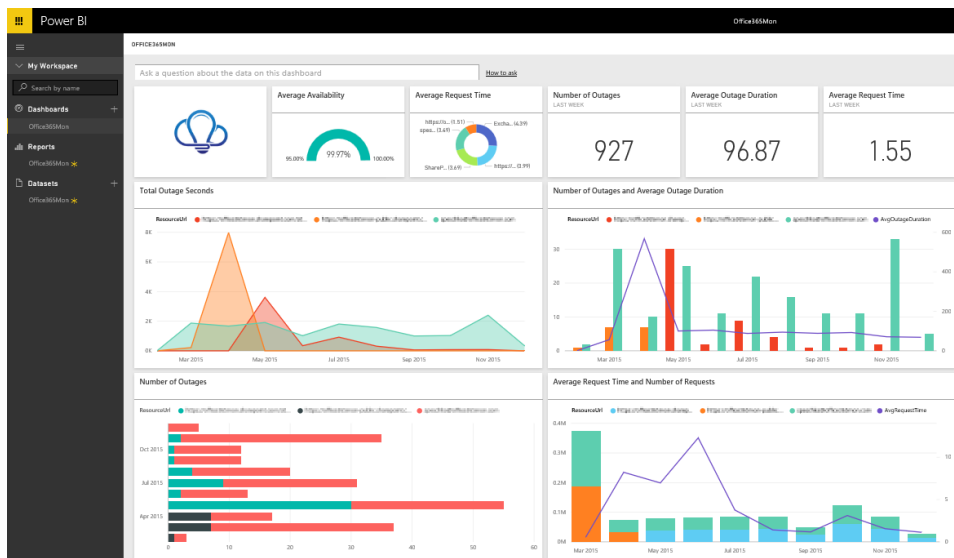


Figura 2.3: Esempio di dashboard prodotta da Power BI

Si definisce famiglia di strumenti perché gli elementi che compongono Power BI sono molteplici. I componenti chiave dell'ecosistema Power BI comprendono:

- *Power BI Desktop*: un'applicazione desktop scaricabile gratuitamente su PC per l'elaborazione dei dati e la costruzione dei report,
- *Power BI Service*: l'ambiente cloud dove vengono pubblicati, analizzati e condivisi i report creati con Power BI Desktop, accessibile attraverso un account Microsoft,
- *App per dispositivi mobili Power BI*: le app Power BI Mobile per dispositivi Android e iOS, cioè per telefoni e tablet Windows,
- *Power BI Gateway*: i gateway utilizzati per sincronizzare i dati esterni in entrata e in uscita da Power BI, necessari per gli aggiornamenti automatici,
- *Power BI integrato*: l'API REST di Power BI, utilizzabile per creare dashboard e report nelle applicazioni personalizzate che servono i diversi utenti,
- *Server di report Power BI*: una soluzione di reporting Power BI locale per le aziende che non archiviano o non possono archiviare dati nel servizio Power BI basato su Cloud,
- *Mercato di Visual BI Power*: un mercato di elementi visivi personalizzati e di elementi visivi basati su R.

## 2.2 Statsmodels

*Statsmodels* è un modulo *Python* che fornisce classi e funzioni per la stima di molti modelli statistici diversi, così come per il calcolo di test statistici e l'esplorazione di dati. Un ampio elenco di statistiche dei risultati è disponibile per ogni estimatore.

*Statsmodels* è costruito sopra *NumPy*, *SciPy* e *Matplotlib*, ma contiene funzioni più avanzate per i test statistici e una modellazione che non è presente nelle librerie numeriche *NumPy* e *SciPy*.

Di seguito sono elencate le principali feature contenute in *Statsmodels*:

- *Linear Regression*;
- *Generalized Linear Models*;
- *Generalized Estimating Equations*;
- *Robust Linear Models*;
- *Linear Mixed Effects Models*;
- *Regression with Discrete Dependent Variables*;
- *ANOVA*;
- ***Time Series Analysis***;
- *Models for Survival and Duration Analysis*;
- *Statistics*;
- *Nonparametric Methods*;
- *Generalized Method of Moments*;
- *Empirical Likelihood*;
- *Plotting Functions*;
- *Datasets Package*;

In particolare, per questo progetto di tesi, è stata utilizzata la feature `statsmodels.tsa`, la quale contiene classi di modelli e funzioni utili per l'analisi delle serie temporali. I modelli di base includono *modelli autoregressivi univariati (AR)*, *modelli au-*



Figura 2.4: Logo ufficiale di Statsmodels

*toregressivi vettoriali (VAR) e modelli di media mobile autoregressiva univariata (ARMA).*

I modelli non lineari includono la *Markov Switching Dynamic Regression* e l'autoregressione. Comprendono anche statistiche descrittive per serie temporali, ad esempio, l'autocorrelazione e la funzione di autocorrelazione parziale, nonché le corrispondenti proprietà teoriche di ARMA.

La struttura del modulo all'interno di `statsmodels.tsa` è:

- `stattools`: proprietà empiriche e test, acf, pacf, adf unit root test, kpss test, bds test, ljung-box test e altri;
- `armodel`: processo autoregressivo univariato;
- `arimamodel`: modello ARIMA univariato;
- `armaprocess`: proprietà dei processi ARMA con determinati parametri; essi includono strumenti per convertire tra ARMA, MA e rappresentazione AR, così come acf, pacf, ecc.;
- `tsatools`: funzioni helper aggiuntive, per creare array di variabili in ritardo, costruire regressori per trend, detrend e simili.

## Struttura dei dati di partenza

*In questo capitolo si parlerà del dataset di partenza, riguardante le vendite di una farmacia sita nel comune di Amatrice, sul quale è stata incentrata l'analisi relativa alla presente tesi. Successivamente si passerà ad una descrizione dettagliata dei campi di tale dataset.*

### 3.1 Dataset delle vendite di una farmacia di Amatrice

Il dataset utilizzato per condurre l'analisi degli effetti economici causati dal sisma del 24 agosto 2016 è nominato `DatiVenditeFarmacia.csv` e contiene i dati degli scontrini emessi dalla farmacia in esame dal 1 gennaio 2015 al 31 dicembre 2020. Tale periodo di tempo è stato scelto appositamente per confrontare le vendite dei vari prodotti pre e post terremoto del Centro Italia del 2016 e del 2017.

Sfortunatamente, a causa del crollo dell'edificio in cui era localizzata la suddetta farmacia, avvenuto il 24 agosto 2016, nel quale sono andati distrutti tutti i computer dell'esercizio commerciale, non è stato possibile effettuare il backup di fine mese e, perciò, risultano mancanti i dati dei prodotti venduti dal 27 luglio al 23 agosto 2016.

La farmacia è stata riaperta in un container provvisorio il 24 settembre 2016, data a partire dalla quale è stato possibile ottenere nuovamente dati utili ai fini dell'analisi. Alla fine del mese di settembre 2017 la farmacia in questione è stata trasferita presso un centro commerciale provvisorio, costruito dalla *Regione Lazio* nello stesso anno per permettere ad alcune attività commerciali del paese di ripartire.

Tale dataset è costituito da 231560 record articolati nei campi che verranno descritti nella prossima sezione.

## 3.2 Descrizione dei campi

Tutti i campi nei quali è strutturato il dataset `DatiVenditeFarmacia.csv` sono elencati e descritti nella Tabella 3.1, osservabile di seguito.

NOME ATTRIBUTO	DESCRIZIONE	FORMATO
Codice Prodotto	ID del prodotto	Numero intero
Descrizione	Nome del prodotto	Stringa
Suddivisione Merceologica	Categoria del prodotto	Stringa
Data	Data di emissione dello scontrino	Data
Ora	Ora di emissione dello scontrino	Ora
Totale Prezzo Pagato Dal Cliente	Prezzo in euro pagato per il prodotto nello scontrino	Numero decimale
Quantità Venduta	Numero di confezioni del prodotto nello scontrino	Numero intero

Tabella 3.1: Tabella dei campi del dataset `DatiVenditeFarmacia.csv`

Nella Tabella 3.2 sono, invece, descritti nel dettaglio i valori del campo *Suddivisione Merceologica*, il quale è risultato l'attributo più significativo nell'analisi delle vendite della farmacia.



SUDDIVISIONE MERCEOLOGICA	DESCRIZIONE
ALIMENTO FINI MEDICI SPECIALI	Prodotto alimentare per la gestione dietetica di pazienti con specifiche esigenze nutrizionali.
BIOCIDA	Sostanza, chimica o biologica, utilizzata come disinfettante per l'igiene umana, animale, alimentare e ambientale.
DISPOSITIVO MEDICO	Strumento utilizzato in medicina per finalità diagnostiche e/o terapeutiche.
FARMACO DA BANCO	Farmaco da automedicazione vendibile senza obbligo di prescrizione medica.
FARMACO ETICO	Farmaco vendibile esclusivamente sotto prescrizione medica.
FARMACO GENERICO	Farmaco bioequivalente rispetto a un altro medicinale.
FARMACO OSPEDALIERO ESITABILE	Farmaco prescrivibile solamente da centri ospedalieri e/o specialisti.
FARMACO VETERINARIO	Sostanza avente proprietà curative e profilattiche delle malattie animali.

MATERIA PRIMA	Materie prime per la creazione di farmaci.
MEDICINALE VETERINARIO PREFABBRICATO	Medicinale veterinario preparato in anticipo e che non corrisponde alla definizione delle specialità medicinali.
OMEOPATICO USO UMANO	Medicinale ottenuto a partire da sostanze denominate materiali di partenza per preparazioni omeopatiche.
PARAFARMACO ERBORISTICO	Prodotto di erboristeria, connesso con la salute, venduto prevalentemente nelle farmacie.
PARAFARMACO SANITARIO	Articolo sanitario venduto prevalentemente nelle farmacie.
PARAFARMACO USO UMANO	Prodotto comunemente venduto in farmacia, con funzione coadiuvante rispetto ai farmaci veri e propri.
PARAFARMACO USO VETERINARIO	Prodotto ad uso animale comunemente venduto in farmacia, con funzione coadiuvante rispetto ai farmaci veterinari veri e propri.

PREMISCELA MEDICATA VETERINARIA	Medicinale veterinario preparato in anticipo per la successiva fabbricazione di alimenti medicamentosi.
PREPARAZIONE MAGISTRALE	Medicinale preparato in farmacia in base ad una prescrizione medica.
PRESIDIO MEDICO CHIRURGICO	Prodotto o dispositivo contenente una o più sostanze disinfettanti, germicide, battericide, fungicide, insetticide, topicide da usare contro i corrispondenti organismi nocivi.

---

Tabella 3.2: Tabella dei valori del campo *Suddivisione Merceologica*

Dal momento che la farmacia oggetto dell'analisi è un'attività commerciale di piccole dimensioni, la gestione dei dati delle vendite non comporta eccessive difficoltà e, di conseguenza, il dataset descritto in questo capitolo risulta pulito.

Infine, per quanto riguarda la fase di *Extract, Transform, Load (ETL)*, l'unica operazione effettuata è stata la rimozione, tramite una funzione di *Microsoft Excel*, della colonna relativa al campo *Codice Prodotto*, in quanto non contenente informazioni di interesse ai fini dell'analisi.



## Analisi descrittiva e diagnostica

*In questo capitolo verranno dapprima mostrati e discussi dei grafici che descrivono l'andamento delle vendite della farmacia oggetto dell'analisi nel periodo 2015-2020, con particolare attenzione alle suddivisioni merceologiche. Nella seconda parte verranno illustrati ulteriori grafici, nei quali ci si è concentrati sulle tipologie di farmaci le cui vendite sono state maggiormente condizionate dal terremoto del Centro Italia.*

### 4.1 Andamento delle vendite della farmacia

Nella *Big Data Analytics* la prima tipologia di analisi è quella descrittiva, nella quale vengono sintetizzati e descritti, lungo tutte le loro dimensioni, i dati relativi agli eventi passati, creando report facilmente interpretabili e di particolare interesse per comprendere la situazione attuale dell'organizzazione. L'analisi descrittiva risulta la più semplice e diffusa al giorno d'oggi, e la sua utilità è dovuta al fatto che essa consente di apprendere dai comportamenti del passato e di capire in che modo questi ultimi potrebbero influenzare i risultati futuri.

La seconda tipologia di analisi dei *Big Data* è quella diagnostica, la quale si pone l'obiettivo di esaminare più approfonditamente i dati per tentare di comprendere le cause di eventi e comportamenti passati. Essa risulta utile per determinare, ad esempio, i fattori che hanno contribuito ad un dato risultato, sia esso negativo o positivo.

In questa sezione verranno approfonditi dei grafici che rappresentano l'andamento delle vendite della farmacia in questione nel periodo 2015-2020; essi permetteranno di effettuare un confronto tra la situazione economica pre e post terremoto del Centro Italia.

Innanzitutto sono state calcolate le entrate effettive nel periodo 2015-2020. La somma di denaro incassata nei giorni compresi tra il 27 luglio e il 23 agosto 2016 è stata stimata prendendo come riferimento i dati delle vendite del mese di agosto dell'anno precedente.

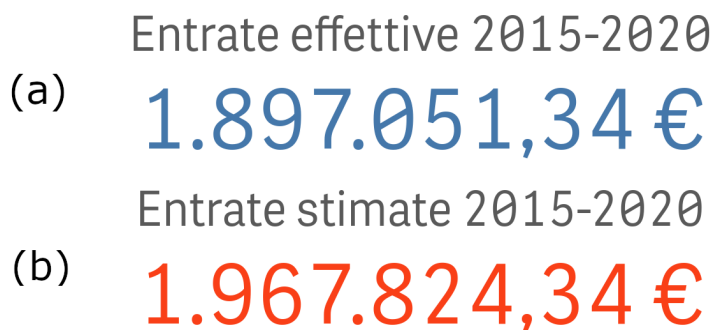


Figura 4.1: Key Performance Indicator relativi alle entrate effettive (a) e stimate (b) nel periodo 2015-2020 (Qlik Sense)

Nella Figura 4.2 è possibile osservare le percentuali di ricavo di ciascuna categoria di prodotto e gli equivalenti importi in euro. È da notare che oltre il 50% degli incassi proviene dalle vendite dei farmaci etici e dei parafarmaci ad uso umano, e il 20% dai dispositivi medici.

Percentuali di Ricavo per Categoria di Prodotto

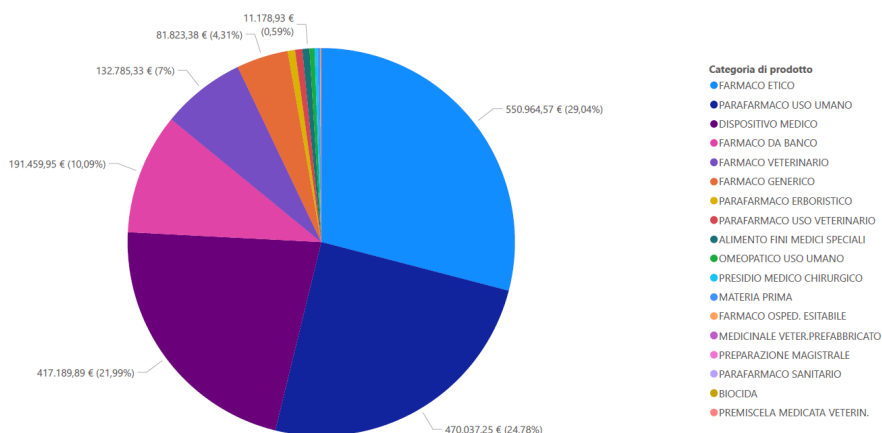


Figura 4.2: Diagramma a torta delle percentuali di ricavo per categoria di prodotto (Power BI)

La Figura 4.3 mostra l'andamento degli incassi nei 6 anni considerati in questa analisi. Si può notare come, a seguito del sisma del 24 agosto 2016, ci sia stata una considerevole diminuzione delle entrate, che hanno raggiunto il valore minimo nell'anno 2017, quando la maggior parte della popolazione del paese era ospitata in altre città.

Con il rientro della popolazione nelle SAE (*Soluzioni Abitative d'Emergenza*), avvenuto tra la fine del 2017 e il 2018, si è verificata una crescita degli incassi, che

si è poi stabilizzata nel biennio successivo. Confrontando il ricavato del 2015 con quello del 2020, è possibile osservare una perdita di circa il 37%.

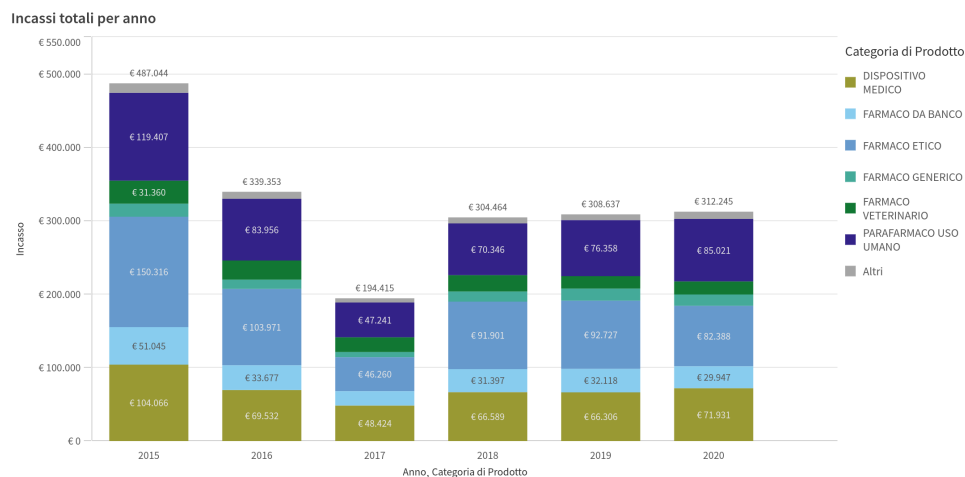


Figura 4.3: Grafico a barre relativo agli incassi annui dal 2015 al 2020 (Qlik Sense)

In Figura 4.4 è rappresentato, invece, l'andamento mensile degli incassi. Dal grafico si deduce che, ad eccezione dell'anno 2017, il picco massimo di ricavi si ha nel mese di agosto, poiché il paese è meta di turisti e di proprietari di seconde case. Sempre per lo stesso motivo, si hanno dei picchi, sebbene più bassi, in corrispondenza delle festività natalizie e pasquali. I valori minimi si registrano tra novembre e febbraio.

Il grafico evidenzia come le entrate nel mese di settembre 2016 siano molto vicine allo 0, a causa della chiusura dell'attività per il sisma.

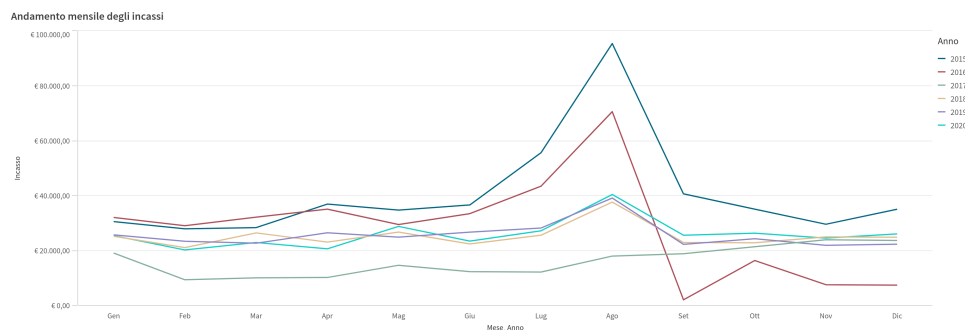


Figura 4.4: Grafico lineare dell'andamento mensile degli incassi per ogni anno (Qlik Sense)

Da questo punto in poi l'analisi si focalizzerà sulla quantità dei prodotti venduti.

Nella Figura 4.5 si possono osservare la percentuale di quantità venduta per ciascuna categoria di prodotto ed il relativo numero di confezioni. Anche in questo caso si nota che oltre il 60% dei prodotti venduti rientrano nelle categorie farmaco etico e parafarmaco ad uso umano.

Percentuali di Quantità Venduta per Categoria di Prodotto

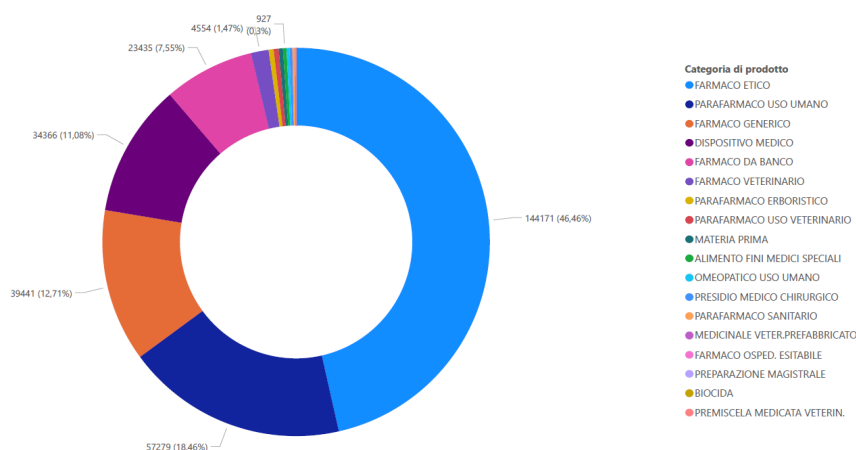


Figura 4.5: Diagramma ad anello relativo alle percentuali di quantità venduta per categoria di prodotto (Power BI)

Confrontando i diagrammi nelle Figure 4.2 e 4.5, si può osservare che i farmaci etici, nonostante costituiscano circa il 50% dei prodotti venduti, contribuiscono soltanto a meno del 30% del ricavato, in quanto sono medicinali vendibili con obbligo di prescrizione medica e rimborsati dal *SSN (Sistema Sanitario Nazionale)*, quindi non totalmente a carico del cliente.

Analogamente, i farmaci generici, anch'essi per la maggior parte rimborsati dal SSN, avendo un costo inferiore rispetto agli originali, contribuiscono a meno del 5% degli incassi, pur rappresentando circa il 13% dei prodotti venduti. Al contrario, i dispositivi medici e i parafarmaci ad uso umano, che costituiscono, rispettivamente, l'11% e il 18% dei prodotti venduti, contribuiscono al 22% e al 25% del ricavato, poiché sono caratterizzati da un prezzo di vendita abbastanza elevato.

La Figura 4.6 contiene una mappa ad albero che mostra più dettagliatamente il numero di confezioni vendute per le 5 principali categorie di prodotto nei diversi anni.

Confrontando i dati relativi agli anni 2015 e 2017, si evince chiaramente come il sisma del Centro Italia abbia causato un ingente calo delle vendite per tutte le categorie di prodotto prese in considerazione.

L'ultimo aspetto considerato nella fase di analisi descrittiva riguarda la distribuzione delle vendite all'interno della giornata lavorativa nel periodo preso in esame.



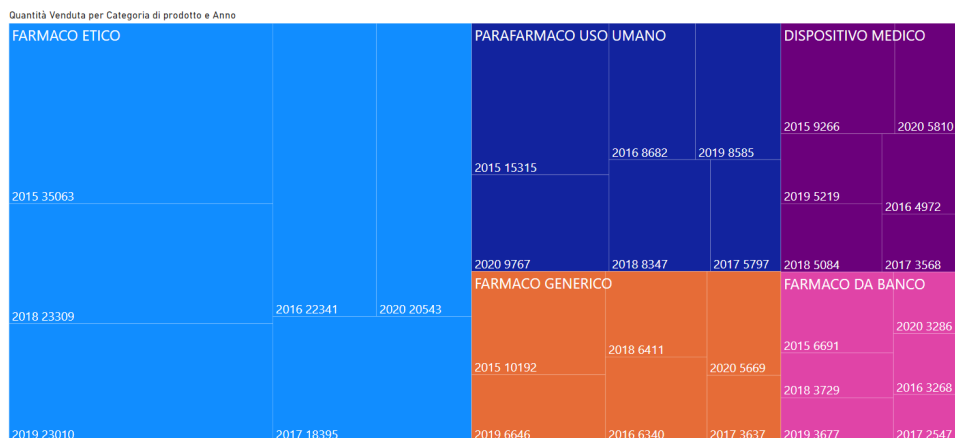


Figura 4.6: Mappa ad albero relativa al numero di confezioni vendute per categoria di prodotto ed anno (Power BI)

I grafici a dispersione nelle Figure 4.7 e 4.8 rappresentano, rispettivamente, le vendite nella fascia oraria 9-13:30 e 16-19:30. L'asse delle ascisse corrisponde alla quantità venduta, mentre quello delle ordinate corrisponde ai ricavi.

La dimensione di ogni cerchio è proporzionale al numero di scontrini in cui compaiono prodotti della categoria corrispondente.

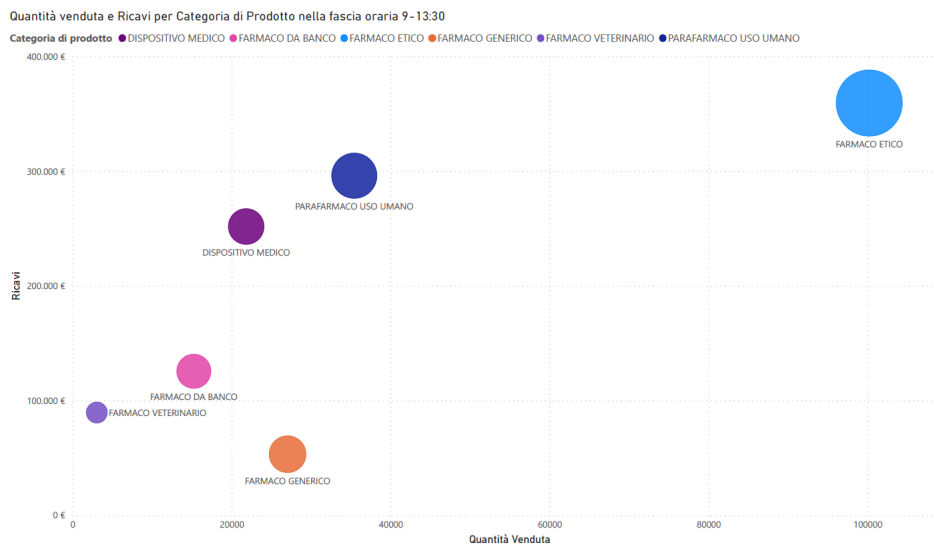


Figura 4.7: Grafico a dispersione delle vendite nella fascia oraria 9-13:30 (Power BI)

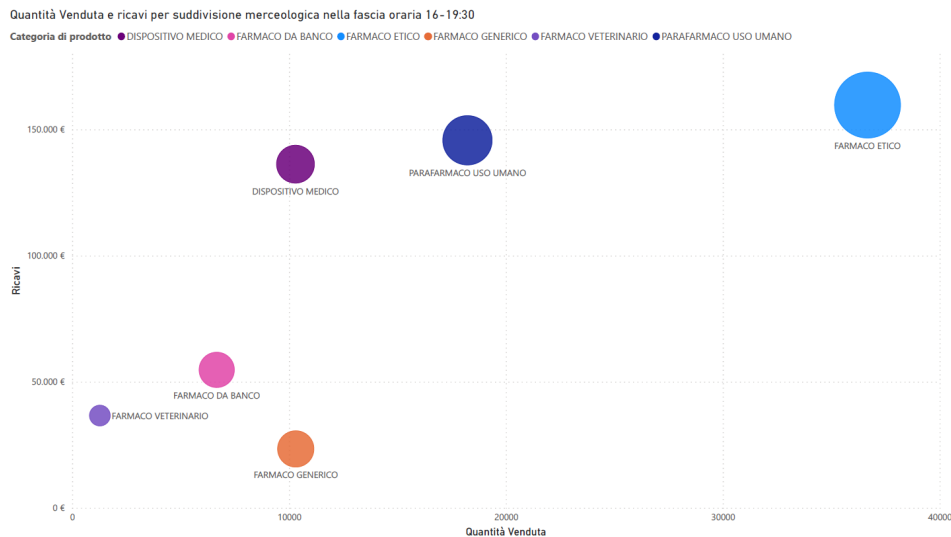


Figura 4.8: Grafico a dispersione delle vendite nella fascia oraria 16-19:30 (Power BI)

Osservando le due figure si può evincere che la maggior parte delle vendite si riscontra nella fascia oraria 9-13:30, poiché in essa sono fruibili tutte le attività commerciali e di servizio da parte degli abitanti del paese, che risiedono prevalentemente in frazioni distanti diversi chilometri dal centro.

## 4.2 Effetti del terremoto del Centro Italia sulle vendite dei prodotti da farmacia

In questa sezione verranno discussi alcuni grafici che mettono in evidenza come il sisma del Centro Italia abbia causato notevoli variazioni delle vendite di determinati prodotti.

Innanzitutto, nelle Figure 4.9 e 4.10 sono riportati, rispettivamente, gli andamenti delle vendite dei farmaci ansiolitici ed antidepressivi, focalizzando l’attenzione su quelli più utilizzati, quali *Lorazepam* e *Alprazolam* per i primi e *Sertralina* per i secondi.

I due grafici a barre mettono in evidenza che, nonostante il ricavato totale della farmacia sia diminuito di oltre un terzo dal 2015 al 2020, le vendite di farmaci ansiolitici non abbiano proporzionalmente subito lo stesso decremento. Al contrario, si rileva addirittura un aumento in termini assoluti dei ricavi derivanti dalle vendite dei farmaci antidepressivi.

Si ipotizza che questo risultato sia dovuto all’aumento di patologie ansioso-depressive post traumatiche causate dal terremoto del 24 agosto 2016 ed ulteriormente acute dalla pandemia di *COVID-19*.

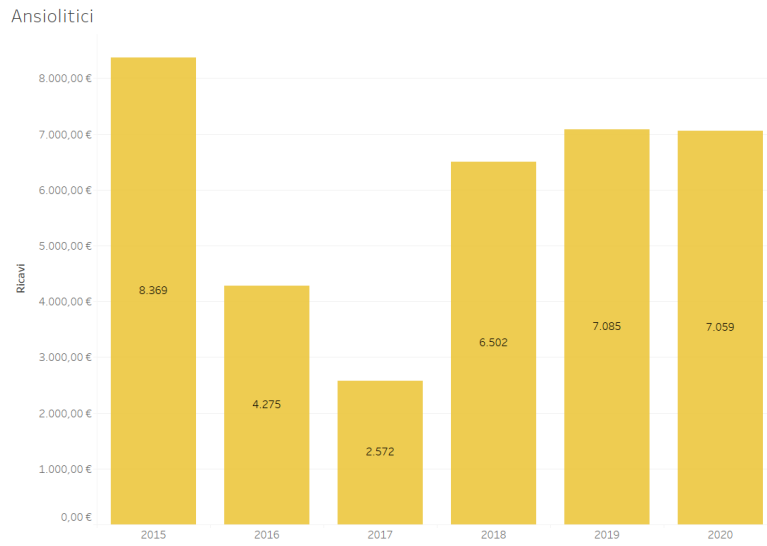


Figura 4.9: Grafico a barre dell'andamento delle vendite di farmaci ansiolitici (Tableau)

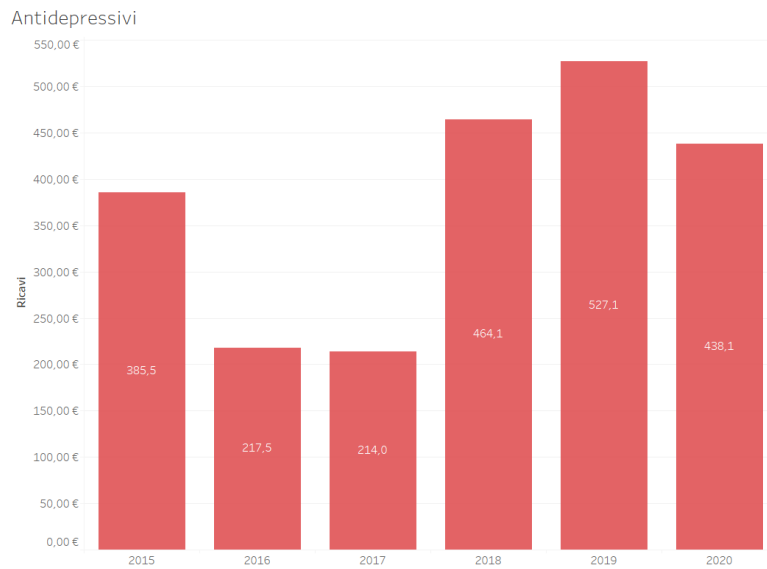


Figura 4.10: Grafico a barre dell'andamento delle vendite di farmaci antidepressivi (Tableau)

La Figura 4.11 mostra gli effetti negativi sulla vendita di farmaci e parafarmaci ad uso veterinario dovuti al sisma del Centro Italia, che ha causato gravi danni alle aziende agricole, con conseguenti difficoltà di mantenimento dei capi di bestiame, che, nei casi più estremi, hanno portato alla chiusura di tali attività.



Figura 4.11: Grafico a barre multiplo dell'andamento delle vendite di farmaci e parafarmaci veterinari (Tableau)

Un ulteriore effetto dovuto al sisma riguarda l'andamento delle vendite dei prodotti cosmetici, osservabile in Figura 4.12, nella quale si può notare un importante decremento negli anni 2016 e 2017.

Al contrario, a partire dal 2018, subito dopo la delocalizzazione della farmacia in un centro commerciale, è stato registrato un aumento progressivo delle vendite, che hanno raggiunto il picco nel 2020, grazie anche al lockdown imposto dal Governo Italiano per contrastare la diffusione del *COVID-19*. Infatti, durante tale periodo, le attività commerciali non essenziali, tra cui le profumerie, erano chiuse, e quindi la clientela doveva rivolgersi alle farmacie per effettuare determinati acquisti.

La conseguenza più grave causata dal sisma è, senza dubbio, lo spopolamento, rispecchiato dal grafico in Figura 4.13, rappresentante l'andamento delle vendite di alcuni tra i principali prodotti per bambini, ovvero giocattoli, omogeneizzati e pannolini.

Osservando il grafico si può evincere che molte famiglie con bambini, temendo che le difficili condizioni in cui versa il paese dopo il terremoto potessero influire negativamente sul proprio benessere economico e psicologico, hanno deciso di trasferirsi altrove in seguito a tale tragico evento.

Si nota, infatti, come, a partire dal 2017, le vendite dei prodotti per bambini abbiano subito un calo drastico. Tuttavia, nel 2020 si è verificato un leggero rialzo degli incassi derivanti dai pannolini, il quale potrebbe essere dovuto, oltre che alla nascita dei figli delle giovani coppie che hanno deciso di rimanere ad Amatrice, alla difficoltà di reperimento di tali articoli al di fuori delle farmacie, al contrario degli omogeneizzati e dei giocattoli.



Figura 4.12: Grafico a barre dell'andamento delle vendite dei prodotti cosmetici (Tableau)

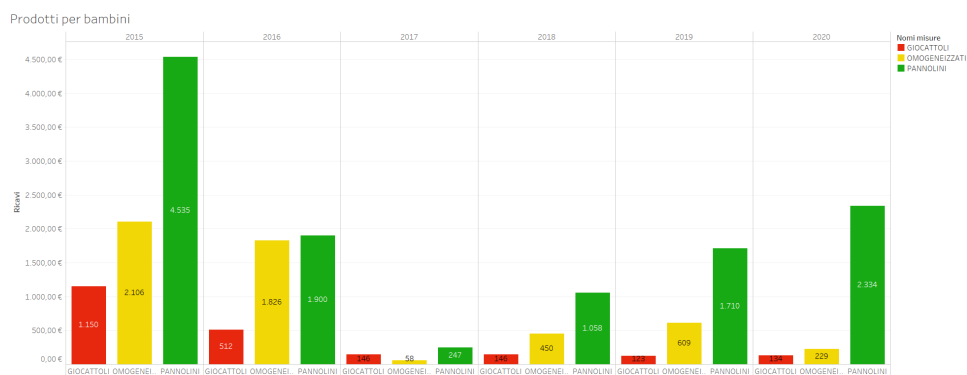


Figura 4.13: Grafico a barre multiplo dell'andamento delle vendite dei prodotti per bambini (Tableau)

Infine, sempre per fornire un'idea su quanto sia stato grande lo spopolamento provocato dal sisma, in Figura 4.14 si osserva l'andamento delle vendite degli integratori, caratterizzato da un brusco calo in corrispondenza degli anni 2016 e 2017.

Il rialzo delle vendite registrato nel 2018, con una progressiva diminuzione nei due anni successivi, può essere spiegato dalla necessità di molti clienti di acquisire integratori per contrastare lo stress causato proprio dal terremoto.

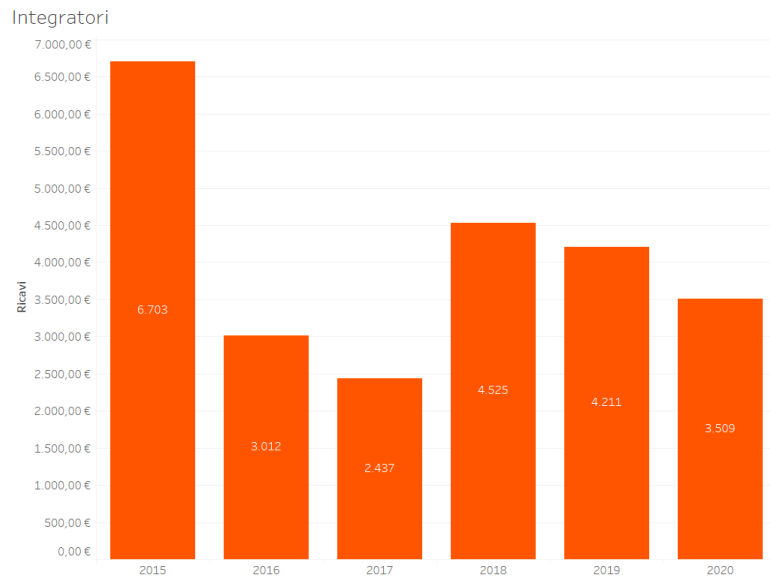


Figura 4.14: Grafico a barre dell'andamento delle vendite degli integratori (Tableau)

## Analisi predittiva

*Dopo aver illustrato gli effetti del sisma del Centro Italia sul consumo di determinati farmaci, nel capitolo corrente verranno discussi dei grafici riguardanti le previsioni dell'andamento delle entrate della farmacia oggetto dell'analisi nel biennio 2021-2022. In seguito, l'attenzione verrà focalizzata sulla stima delle vendite dei prodotti di maggior rilevanza nel medesimo periodo.*

### 5.1 Previsione del ricavato nel biennio 2021-2022

L'analisi predittiva è un metodo di analisi che esamina i dati attuali e storici allo scopo di prevedere gli eventi futuri. Si avvale di tecniche quali *Machine Learning*, modellazione statistica e *Data Mining*, per consentire alle aziende di identificare tendenze, comportamenti, risultati futuri e opportunità di business.

L'analisi predittiva offre lo strumento ideale per interpretare e utilizzare i *Big Data* allo scopo di ottenere informazioni aziendali approfondite e supportare il processo decisionale. I *Big Data* costituiscono l'elemento chiave dell'analisi predittiva poiché, per identificare pattern o tendenze e prendere decisioni veramente informate, servono grandi quantità di dati.

Basandosi su un set di variabili di input, l'analisi predittiva addestra un modello allo scopo di prevedere i valori futuri dei nuovi dati. Il modello identifica, quindi, le relazioni e i pattern che legano le variabili, assegnando anche un punteggio in base agli elementi che è stato addestrato a cercare. Tale punteggio può essere utilizzato per valutare il rischio o i potenziali vantaggi di una serie di condizioni e consente di determinare la probabilità che un determinato evento si verifichi.

In questo capitolo verranno discussi i risultati dell'analisi predittiva effettuata sul dataset delle vendite della farmacia illustrato precedentemente. Di seguito sono riportati i migliori modelli predittivi riguardanti le entrate dell'esercizio commerciale, ottenuti entrambi tramite *Tableau* con un intervallo di previsione del 99%.

In Figura 5.1 viene mostrata la previsione dell'andamento mensile degli incassi per il biennio 2021-2022, formulata basandosi sui dati relativi al triennio 2018-2020. Dall'analisi di tale grafico si può osservare che, come riferito nel capitolo precedente, il ricavato della farmacia si è stabilizzato a seguito del rientro di molti abitanti del paese nelle *SAE*.

Previsione del ricavato mensile nel biennio 2021-2022

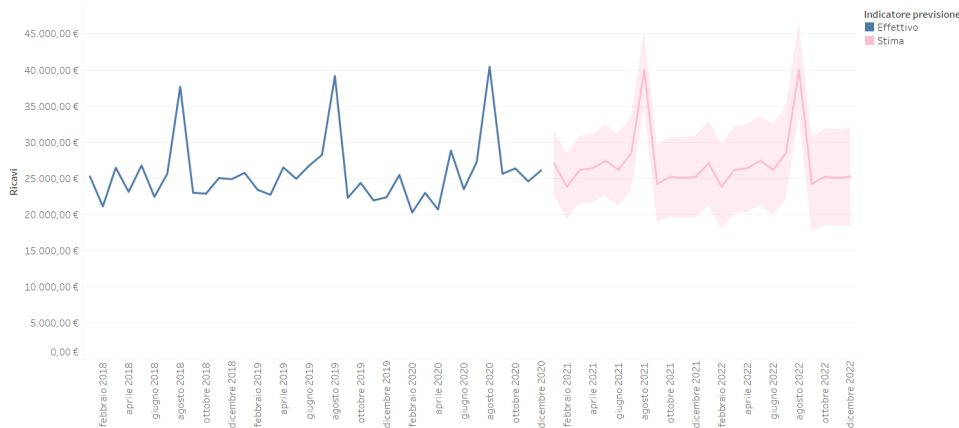


Figura 5.1: Previsione dell’andamento mensile del ricavato nel biennio 2021-2022 (Tableau)

Come si può notare dal grafico soprastante, si prevede che anche nei prossimi due anni, essendo state ultimate da poco le pratiche burocratiche per avviare i primi progetti di ricostruzione, l’afflusso di clienti presso l’attività commerciale in esame resterà pressoché invariato, con picco massimo registrato nel mese di agosto e valore minimo nel periodo invernale.

Il grafico a barre in Figura 5.2 rappresenta, invece, la previsione del ricavato per le 5 principali categorie di prodotto nel biennio 2021-2022, formulata sempre a partire dai dati relativi al triennio 2018-2020. Si prevede, osservando tale grafico, un leggero aumento degli incassi per ciascuna delle categorie considerate rispetto al 2020, specialmente per i parafarmaci ad uso umano.

Previsione del ricavato per categoria di prodotto nel biennio 2021-2022

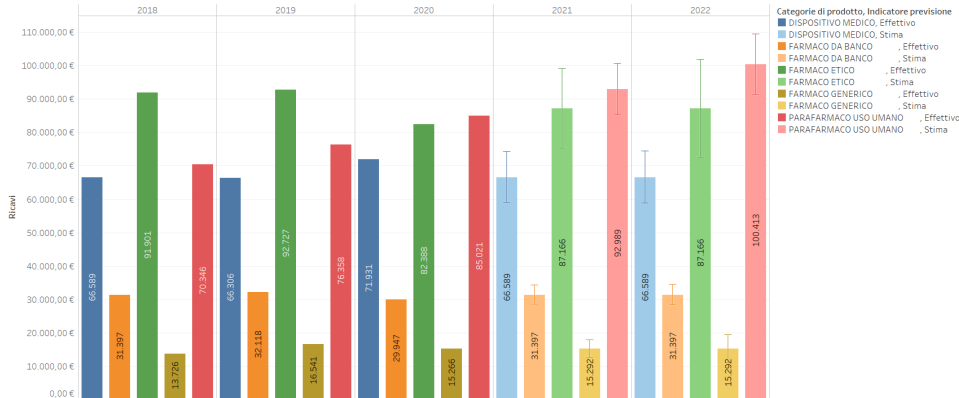


Figura 5.2: Previsione del ricavato per categoria di prodotto per il biennio 2021-2022 (Tableau)



Una possibile motivazione di tale risultato è la maggiore libertà di spostamento che si avrà quando una buona percentuale della popolazione italiana avrà ricevuto il vaccino anti *COVID-19*, e che porterà ad un piccolo incremento dell'afflusso turistico, essendo le attuali strutture alberghiere del paese poco capienti.

## 5.2 Previsione delle vendite di specifiche categorie di prodotto

Nella sezione corrente verranno illustrati ed analizzati dei grafici riguardanti le previsioni dell'andamento delle vendite delle categorie di prodotto maggiormente influenzate dal sisma del Centro Italia. Tali modelli predittivi sono stati creati con *Power BI*, impostando l'intervallo di previsione al 99%, dal momento che la serie temporale compresa tra il 2018 e il 2020 è risultata troppo breve per effettuare previsioni in *Tableau*.

Nelle Figure 5.3 e 5.4 sono riportate, rispettivamente, la previsione dell'andamento delle entrate provenienti dagli ansiolitici e dagli antidepressivi citati nel capitolo precedente. Si prevede, per il biennio 2021-2022, una crescita delle vendite notevole per i primi e leggera per i secondi, in quanto, essendo stati avviati da poco tempo i primi cantieri ad Amatrice, molti abitanti del paese dovranno aspettare ancora diversi anni per rientrare nelle proprie abitazioni, ed è intuibile che tale situazione di precarietà, nel lungo periodo, possa acuire i disturbi mentali di coloro che hanno subito traumi di questa tipologia.

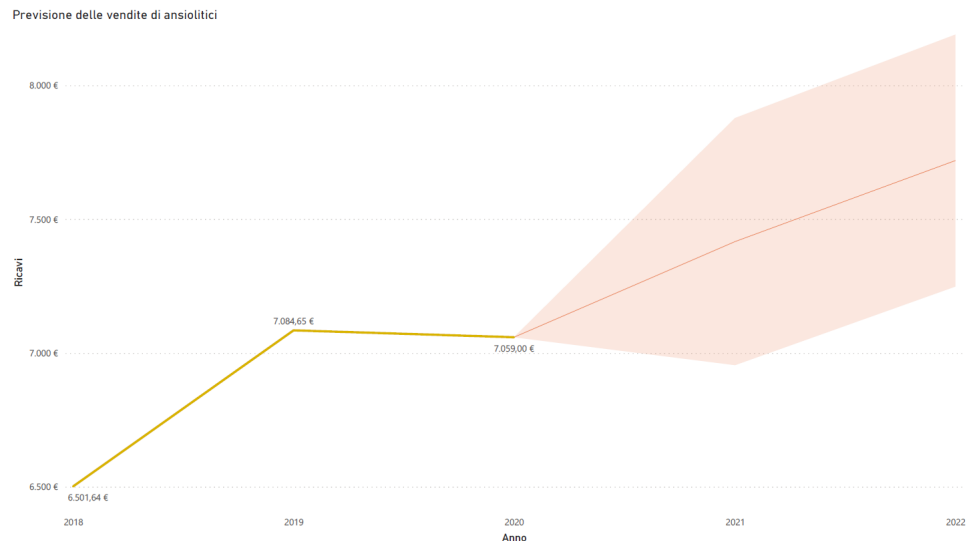


Figura 5.3: Previsione dell'andamento delle vendite degli ansiolitici per il biennio 2021-2022 (Power BI)

Previsione delle vendite di antidepressivi

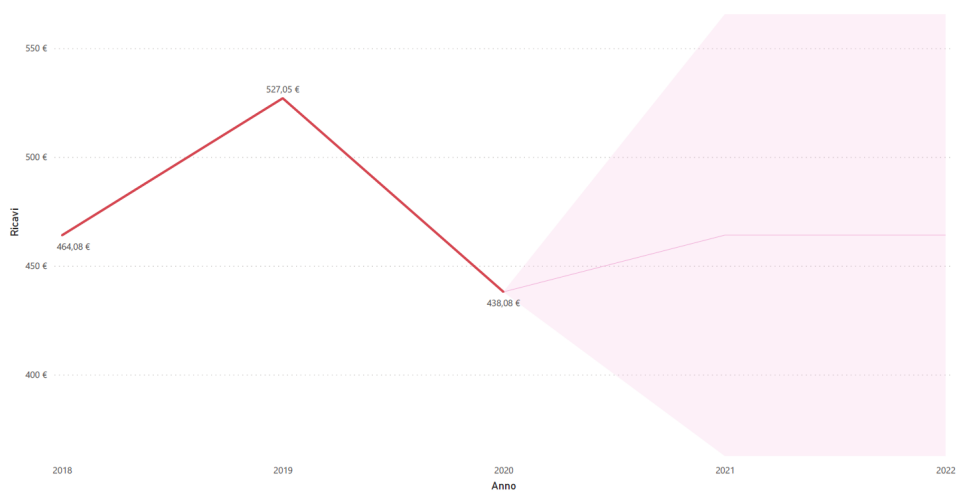


Figura 5.4: Previsione dell'andamento delle vendite degli antidepressivi per il biennio 2021-2022 (Power BI)

In Figura 5.5 è possibile osservare, invece, la previsione dell'andamento delle vendite dei farmaci veterinari nel biennio considerato.

Previsione delle vendite dei farmaci veterinari

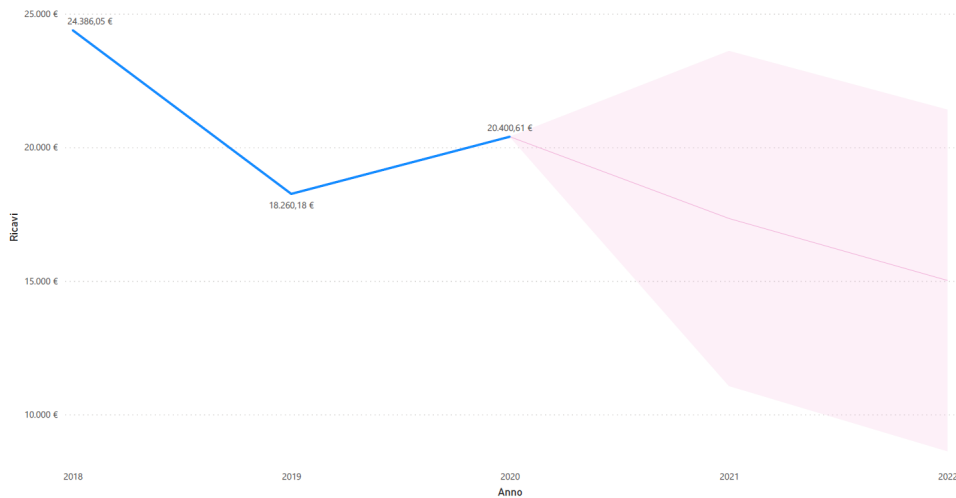


Figura 5.5: Previsione dell'andamento delle vendite dei farmaci veterinari per il biennio 2021-2022 (Power BI)

È previsto un drastico calo degli incassi derivanti da tale categoria di prodotto, il quale può essere spiegato dalle ingenti difficoltà nel ricostruire, entro il 2022, una buona porzione delle strutture per il ricovero di animali, di proprietà delle varie aziende agricole presenti sul territorio amatriciano.

In tal modo gli allevatori potrebbero vedersi costretti a vendere parte dei propri capi di bestiame, per evitare loro una condizione di disagio che possa influire negativamente sulla produzione di alimenti di origine animale, o a cessare le proprie attività, per impossibilità di sostenere le spese ad esse legate.

Per concludere, la Figura 5.6 rappresenta la previsione dell'andamento delle vendite dei prodotti per bambini.

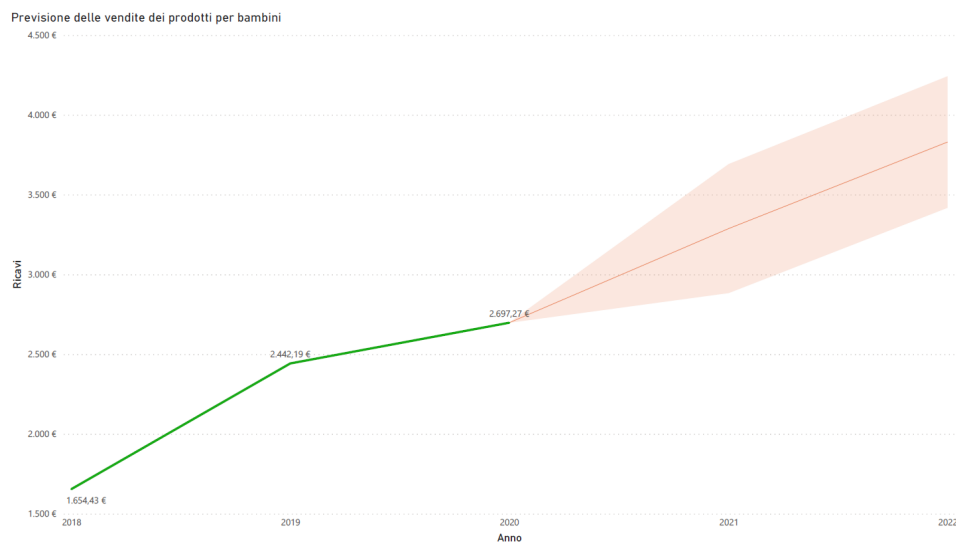


Figura 5.6: Previsione dell'andamento delle vendite dei prodotti per bambini per il biennio 2021-2022 (Power BI)

Tale aumento del ricavato, pronosticato per il biennio 2021-2022, è probabilmente dovuto all'aumento del tasso di natalità che si sta verificando negli ultimi anni nel comune di Amatrice. Si deduce, infatti, che, come accaduto a L'Aquila, l'attaccamento alla vita e l'istinto di sopravvivenza, che spesso seguono eventi di portata catastrofica, quale il sisma del Centro Italia, comportino la volontà, da parte delle giovani coppie, di concepire figli per alleviare il dolore derivante dalle ingenti perdite di vite umane.



---

## Analisi prescrittiva e Machine Learning

*In questo capitolo verranno discussi i risultati delle analisi più avanzate condotte sul dataset delle vendite della farmacia in esame, al fine di prescrivere delle azioni che potrebbero comportare un miglioramento dello status economico dell'esercizio commerciale. Dapprima si cercherà di individuare delle correlazioni tra specifiche tipologie di prodotti, per poi procedere ad una classificazione di questi ultimi in base al prezzo e alla quantità venduta. Infine verranno utilizzati degli algoritmi di Clustering per segmentare i prodotti e analizzare le caratteristiche dei vari gruppi.*

### 6.1 Premessa

L'analisi prescrittiva costituisce l'ultima fase della *Big Data Analytics*, e può essere definita come una combinazione di analisi matematica e statistica, esperimenti, ipotesi, modellazioni e simulazioni in grado di migliorare e guidare il processo decisionale e l'efficacia delle decisioni prese dagli esseri umani.

L'analisi prescrittiva, di fatto, si spinge ben oltre la previsione di eventuali scenari o risultati futuri, e fornisce ai decision maker, anche in modo automatizzato, raccomandazioni utili sulle decisioni da prendere e le azioni da compiere per raggiungere determinati risultati.

Al giorno d'oggi, per questa tipologia di analisi, si fa sempre più uso di sistemi di *Machine Learning* e tecniche di *Intelligenza Artificiale*, come nel caso dell'analisi predittiva, i quali, però, vengono addestrati ed utilizzati in modo differente, al fine di renderli in grado di spiegare le cause di determinati eventi. Infatti, mentre nell'analisi predittiva si deduce cosa è probabile che avvenga in futuro, nell'analisi prescrittiva si comprendono anche i motivi per i quali vengono forniti i suggerimenti di azione e le raccomandazioni utili per prendere decisioni efficaci e ottenere i risultati previsti.

### 6.2 Correlazione

La *correlazione* è una tecnica di analisi che ha lo scopo di determinare se esiste una relazione lineare tra due variabili. Tale relazione non dipende necessariamente da

un rapporto di causa-effetto quanto dalla tendenza di una variabile a cambiare in funzione di un'altra.

L'utilizzo della correlazione può servire a sviluppare una comprensione di un dataset e a trovare relazioni che possono assistere nella spiegazione di un fenomeno. La correlazione viene, pertanto, usata comunemente per il *Data Mining*, dove l'identificazione delle relazioni tra le variabili in un dataset porta alla scoperta di pattern e anomalie.

In questa prima sezione si cercherà di individuare delle correlazioni all'interno del dataset degli scontrini emessi dalla farmacia oggetto dell'analisi, al fine di comprendere come potrebbero evolvere in futuro le vendite di determinati prodotti in funzione di altri e di fornire dei suggerimenti sulle azioni da compiere al fine di aumentare i ricavi dell'attività commerciale.

### 6.2.1 Correlazione tra le categorie di prodotto nel periodo pre-terremoto

In questa sezione si analizzerà una mappa di calore, osservabile in Figura 6.1, rappresentante le correlazioni tra le diverse categorie di prodotto nel periodo gennaio 2015-luglio 2016. La misura utilizzata per la creazione di questa matrice è l'*indice di correlazione di Pearson*, un indice molto utilizzato in statistica che esprime un'eventuale relazione di linearità tra due variabili statiche. Esso ha un valore compreso tra +1 e -1, dove +1 corrisponde alla perfetta correlazione lineare positiva, 0 corrisponde ad un'assenza di correlazione lineare e -1 corrisponde alla perfetta correlazione lineare negativa.

Date due variabili statistiche X e Y, l'indice di correlazione di Pearson è definito come la loro covarianza divisa per il prodotto delle deviazioni standard delle due variabili:  $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ , dove  $\sigma_{XY}$  è la covarianza tra X e Y e  $\sigma_X$ ,  $\sigma_Y$  sono le due deviazioni standard. Nel caso in esame, ogni variabile statistica corrisponde al numero di confezioni vendute di una delle due categorie di prodotto coinvolte nella correlazione.

Osservando la Figura 6.1, si notano valori molto elevati dell'indice di Pearson per le seguenti correlazioni:

- Farmaco etico e farmaco generico: valore *0.832*;
- Farmaco ospedaliero esitabile e parafarmaco ad uso umano: valore *1*;
- Farmaco veterinario e presidio medico chirurgico: valore *1*.

Tuttavia, per trarre delle conclusioni da questi risultati, è necessario definire una matrice di correlazione analoga, ma relativa al periodo post-terremoto, e verificare se anche in essa siano presenti valori simili per le medesime correlazioni.

### 6.2.2 Correlazione tra le categorie di prodotto nel periodo post-terremoto

In Figura 6.2 è illustrata la mappa di calore delle correlazioni tra le diverse categorie di prodotto nel periodo settembre 2016-dicembre 2020.

Anche in questo caso si osservano valori molto alti per diverse correlazioni, quali:

Matrice di correlazione tra categorie di prodotto pre-terremoto

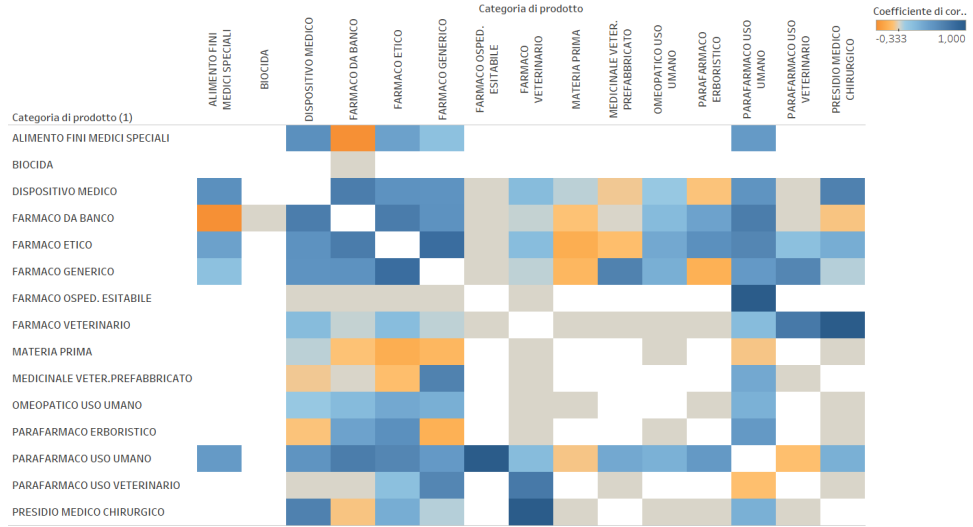


Figura 6.1: Mappa di calore delle correlazioni tra le categorie di prodotto nel periodo pre-terremoto (Tableau)

Matrice di correlazione tra categorie di prodotto post-terremoto

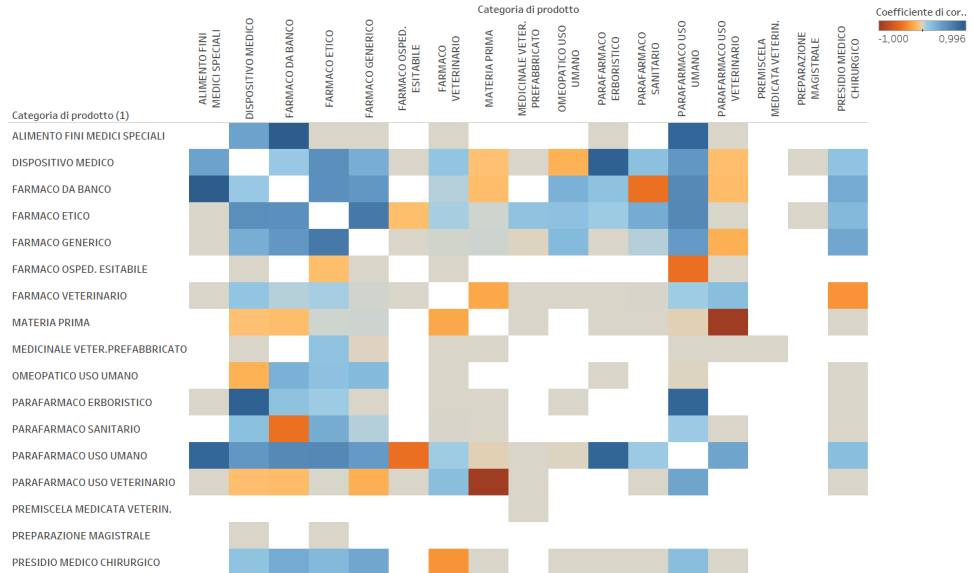


Figura 6.2: Mappa di calore delle correlazioni tra le categorie di prodotto nel periodo post-terremoto (Tableau)

- Farmaco etico e farmaco generico: valore  $0.747$ ;
- Alimento per fini medici speciali e parafarmaco ad uso umano: valore  $0.884$ ;
- Parafarmaco ad uso umano e parafarmaco erboristico: valore  $0.895$ ;
- Dispositivo medico e parafarmaco erboristico: valore  $0.933$ ;
- Alimento per fini medici speciali e farmaco da banco: valore  $0.996$ .

Confrontando i valori dell'indice di correlazione di Pearson presenti nelle due mappe di calore si può concludere che l'unica correlazione non dovuta alla casualità, e quindi significativa, sia quella tra farmaco etico e farmaco generico. In effetti, è plausibile pensare che molti clienti, che si rechino in farmacia per acquistare un farmaco mediante ricetta medica, preferiscano acquistare il generico piuttosto che l'originale, in quanto quest'ultimo è caratterizzato da un costo superiore. Il decremento di oltre il 25% del valore dell'indice di Pearson è, probabilmente, dovuto alla notevole diminuzione della clientela che si è verificata in seguito al sisma del Centro Italia.

## 6.3 Classificazione

La *classificazione* è una tecnica di *apprendimento supervisionato* tramite la quale i dati vengono suddivisi in categorie rilevanti, già conosciute a priori. Essa può essere effettuata per due o più categorie e consiste di due passi:

- Fornire al sistema dei dati di training già categorizzati o etichettati, in modo tale che esso possa sviluppare una comprensione delle diverse categorie.
- Fornire al sistema dati sconosciuti, ma simili, cosicché esso, basandosi sulla comprensione che ha precedentemente sviluppato a partire dai dati di training, possa classificare i nuovi dati.

La classificazione consente di estrarre modelli in grado di descrivere importanti classi di dati e predire futuri andamenti degli stessi. I modelli così ottenuti vengono rappresentati sotto forma di regole di classificazione, alberi di decisione, formule matematiche, etc.

In questa sezione verrà effettuata una classificazione delle categorie di prodotto in base al prezzo di vendita e alla quantità venduta, al fine di identificare quelle più richieste, e quindi più redditizie, e di comprendere, conseguentemente, quali siano i prodotti su cui puntare per aumentare il ricavato della farmacia.

### 6.3.1 Classificazione dei farmaci in base al prezzo di vendita

In questa sezione ci si pone l'obiettivo di assegnare, ad ogni categoria di prodotto, una delle seguenti tre classi, in base al ricavato totale generato da essa:

- 1: altamente redditizia;
- 2: mediamente redditizia;
- 3: scarsamente redditizia.



Per definire queste classi è stata utilizzata la funzione `MODEL_QUANTILE` di Tableau, la quale permette di calcolare il quantile predittivo posteriore, ovvero il valore previsto in un quantile specificato. Essa prevede tre parametri:

- *Quantile*: un numero compreso tra 0 e 1, che indica quale quantile deve essere previsto.
- *Espressione di destinazione*: la misura da prevedere o “destinazione”.
- *Espressione predittore*: il predittore utilizzato per eseguire la stima. I predittori possono essere dimensioni, misure, o entrambi.

Nel caso in questione, sono state definite, tramite la funzione `MODEL_QUANTILE`, le due misure mostrate nel Listato 6.1.

```
1 %Quantile 0.33 Prezzo
2 MODEL_QUANTILE(0.33, SUM([Totale Prezzo Pagato Dal Cliente]))
3
4 %Quantile 0.66 Prezzo
5 MODEL_QUANTILE(0.66, SUM([Totale Prezzo Pagato Dal Cliente]))
```

Listato 6.1: Definizione delle misure **Quantile 0.33 Prezzo** e **Quantile 0.66 Prezzo**

Successivamente i quantili così calcolati sono stati utilizzati per creare la dimensione **Classe Di Prezzo**, la quale permette proprio di assegnare una classe ad una categoria di prodotto in base al ricavato totale da essa generato. Tutto questo viene mostrato nel Listato 6.2.

```
1 %Classe Di Prezzo
2 IF SUM([Totale Prezzo Pagato Dal Cliente]) < [Quantile 0.33 prezzo] THEN '3'
3 ELSEIF SUM([Totale Prezzo Pagato Dal Cliente]) >= [Quantile 0.33 prezzo] AND
4 SUM([Totale Prezzo Pagato Dal Cliente]) <= [Quantile 0.66 prezzo] THEN '2'
5 ELSEIF SUM([Totale Prezzo Pagato Dal Cliente]) > [Quantile 0.66 prezzo] THEN '1' END
```

Listato 6.2: Definizione della dimensione **Classe Di Prezzo**

I risultati di questa prima classificazione sono mostrati in Figura 6.3 e confermano quanto affermato nella Sezione 4.1 riguardo la redditività delle diverse categorie di prodotto. Tuttavia, per poter fornire dei consigli in merito a quali prodotti dovrebbe puntare l'attività commerciale per aumentare le entrate economiche, è necessario effettuare una classificazione anche in base alla quantità venduta. Questa sarà l'argomento principale della prossima sezione.

### 6.3.2 Classificazione dei farmaci in base alla quantità venduta

Per quanto riguarda la quantità venduta, sono state identificate, allo stesso modo, le seguenti tre classi:

- *1*: altamente richiesta;
- *2*: mediamente richiesta;
- *3*: scarsamente richiesta.

Anche in questo caso è stata utilizzata la funzione `MODEL_QUANTILE` di Tableau per definire due quantili secondo quanto mostrato nel Listato 6.3.

## Classificazione delle categorie di prodotto in base al ricavato totale

Categoria di prodotto	Classe di prezzo	Ricavato totale
ALIMENTO FINI MEDICI SPECIALI	3	11.179
BIOCIDA	3	20
DISPOSITIVO MEDICO	1	417.190
FARMACO DA BANCO	1	191.460
FARMACO ETICO	1	550.965
FARMACO GENERICO	2	81.823
FARMACO OSPED. ESITABILE	3	1.039
FARMACO VETERINARIO	2	132.785
MATERIA PRIMA	3	4.062
MEDICINALE VETER.PREFABBRICATO	3	485
OMEOPATICO USO UMANO	3	7.997
PARAFARMACO ERBORISTICO	3	11.824
PARAFARMACO SANITARIO	3	65
PARAFARMACO USO UMANO	1	470.037
PARAFARMACO USO VETERINARIO	3	11.328
PREMISCELA MEDICATA VETERIN.	3	7
PREPARAZIONE MAGISTRALE	3	169
PRESIDIO MEDICO CHIRURGICO	3	4.616

Figura 6.3: Tabella di classificazione delle categorie di prodotto in base al ricavato totale (Tableau)

```

1 %Quantile 0.33 Quantità
2 MODEL_QUANTILE(0.33, SUM([Quantità Venduta]))
3
4 %Quantile 0.66 Quantità
5 MODEL_QUANTILE(0.66, SUM([Quantità Venduta]))

```

Listato 6.3: Definizione delle misure Quantile 0.33 Quantità e Quantile 0.66 Quantità

Dopodiché è stata creata la dimensione **Classe Di Vendita** seguendo lo stesso procedimento adottato in precedenza per la dimensione **Classe Di Prezzo**.

```

1 %Classe Di Vendita
2 IF SUM([Quantità Venduta]) < [Quantile 0.33 quantità] THEN '3'
3 ELSEIF SUM([Quantità Venduta]) >= [Quantile 0.33 quantità] AND
4 SUM([Quantità Venduta]) <= [Quantile 0.66 quantità] THEN '2'
5 ELSEIF SUM([Quantità Venduta]) > [Quantile 0.66 quantità] THEN '1' END

```

Listato 6.4: Definizione della dimensione Classe Di Vendita

Nella tabella in Figura 6.4 vengono elencati i risultati di questa seconda classificazione.

## Classificazione delle categorie di prodotto in base alla quantità venduta

Categoria di prodotto	Classe di vendita	Quantità Venduta
ALIMENTO FINI MEDICI SPECIALI	2	927
BIOCIDA	3	2
DISPOSITIVO MEDICO	1	34.366
FARMACO DA BANCO	2	23.435
FARMACO ETICO	1	144.171
FARMACO GENERICO	1	39.441
FARMACO OSPED. ESITABILE	3	76
FARMACO VETERINARIO	2	4.554
MATERIA PRIMA	2	1.079
MEDICINALE VETER.PREFABBRICATO	3	197
OMEOPATICO USO UMANO	2	758
PARAFARMACO ERBORISTICO	2	1.303
PARAFARMACO SANITARIO	3	668
PARAFARMACO USO UMANO	1	57.279
PARAFARMACO USO VETERINARIO	2	1.291
PREMISCELA MEDICATA VETERIN.	3	2
PREPARAZIONE MAGISTRALE	3	10
PRESIDIO MEDICO CHIRURGICO	2	739

Figura 6.4: Tabella di classificazione delle categorie di prodotto in base alla quantità venduta (Tableau)

Osservando le Figure 6.3 e 6.4, si può giungere alla conclusione che, dal momento che i dispositivi medici e i parafarmaci ad uso umano risultano due categorie di prodotto altamente redditizie e molto richieste, per poterne incrementare le vendite, e conseguentemente il ricavato, i titolari della farmacia potrebbero effettuare delle campagne promozionali, che prevedano degli sconti sui suddetti prodotti, o anche avvalersi di promoter, che incentivino i clienti all'acquisto.

Inoltre, per quanto riguarda i farmaci da banco, si potrebbe dedicare un angolo di esposizione ad essi, indicandone il prezzo praticato, lasciando al cliente la libera scelta dell'acquisto.

## 6.4 Clustering

Il *Clustering* è il processo di raggruppamento di oggetti di cui non si conosce l'etichetta di classe. Un cluster è un insieme di oggetti che sono simili l'un l'altro e sono dissimili dagli oggetti di altri cluster.

La classificazione richiede una costruzione e un'etichettatura del training set, che risultano spesso costose in termini operazionali, e, quindi, spesso può essere desiderabile procedere in senso inverso. Il clustering è adattabile ai cambiamenti e consente di scegliere quali sono le caratteristiche di interesse per distinguere i vari gruppi.

Il clustering è un esempio di *apprendimento non supervisionato*; esso rappresenta una disciplina giovane dove convergono aree di ricerca quali il *Data Mining*, la statistica, il *Machine Learning*, la tecnologia dei database spaziali, la biologia e il marketing.

In questa sezione si cercherà di suddividere in cluster l'insieme dei prodotti venduti dalla farmacia in esame, al fine di comprendere quali fra essi rappresentano degli *outlier*, ovvero degli elementi con valori anomali rispetto a quelli degli altri raggruppamenti, e analizzarne le caratteristiche.

### 6.4.1 Segmentazione dei prodotti e analisi degli outlier

In Figura 6.5 è rappresentato il risultato dell'applicazione dell'algoritmo di Clustering di Power BI sull'insieme dei prodotti che compaiono negli scontrini emessi dall'esercizio commerciale, utilizzando come feature il ricavato e il numero di confezioni vendute; il numero di cluster è stato impostato automaticamente dal software.

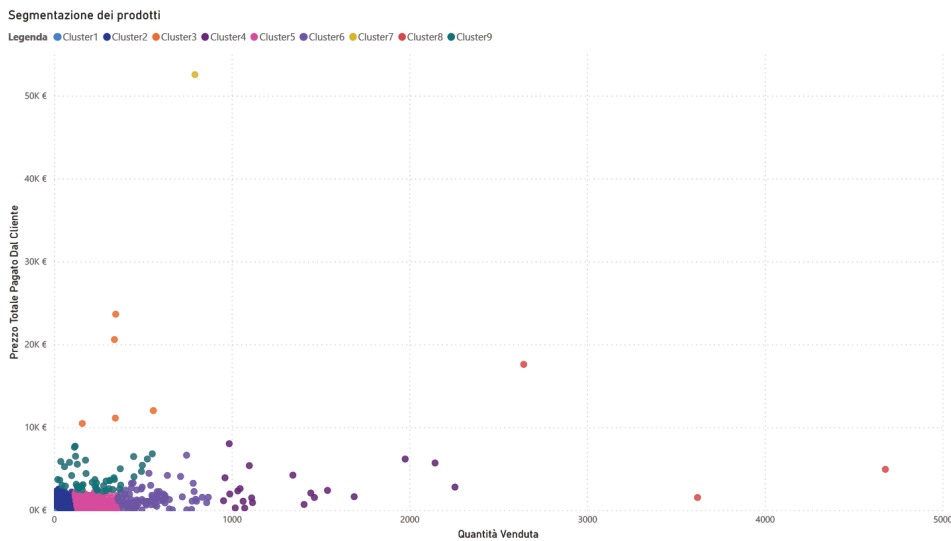


Figura 6.5: Segmentazione dei prodotti in base al ricavato e alla quantità venduta (Power BI)

Dal grafico si notano immediatamente due cluster molto interessanti, quali il 3 e l'8; il primo contiene prodotti quali Glucocard G Sensor, Contour Next, Novofine e Glucomen, che sono dispositivi medici per diabetici, e il Repen 250 ml, che è un antibiotico ad uso veterinario, mentre il secondo include la Tachipirina 1000 mg in compresse, il Lasix in compresse e la Cardioaspirina in compresse.

È evidente che i prodotti del cluster 3 producono un notevole incasso, sebbene la loro quantità venduta non sia numericamente elevata; quindi, si può definire tale raggruppamento come quello delle merci ad alto costo.

D'altro canto, si può affermare che il cluster 8 rappresenta i prodotti ad alta rotazione, ovvero caratterizzati da un'alta frequenza di vendita, ma da un basso costo; questo spiega la loro posizione marginale all'interno del grafico.

Di particolare rilievo è il cluster 7, il quale è composto da un unico prodotto, l'Accu-Chek Aviva, un ulteriore dispositivo medico per diabetici, che però non è stato inserito nel cluster 3 dal software Power BI, molto probabilmente per il prezzo che lo caratterizza, il quale risulta il più elevato in assoluto all'interno del dataset. Proprio per questo motivo si può considerare tale prodotto un outlier.

Per concludere, viste le caratteristiche di tali cluster, si consiglia ai titolari della farmacia di indirizzare gran parte delle vendite dei dispositivi medici per diabetici su un'unica referenza, in modo tale da ottenere una maggiore scontistica al momento dell'acquisto, e, di conseguenza, un guadagno più elevato.



## Analisi delle serie temporali

*Nel capitolo corrente verrà analizzata la serie temporale delle vendite della farmacia oggetto di studio, focalizzando l'attenzione sui trend e sulle stagionalità. Successivamente verrà costruito e valutato un modello per la previsione dell'andamento futuro della serie temporale.*

### 7.1 Premessa

L'*analisi delle serie temporali* raggruppa una serie di metodi statistici atti a indagare una serie temporale, determinare il processo alla base della stessa e trarre previsioni. Secondo l'approccio tradizionale, si assume che il processo abbia una parte deterministica, che consente di scomporlo in componenti tendenziali, cicliche e/o stagionali, e che la differenza tra i dati teorici del modello deterministico ed i dati osservati sia attribuibile ad una componente casuale residuale. Secondo l'approccio moderno, invece, si assume che il processo descritto sia stato generato da un processo stocastico descrivibile mediante un modello probabilistico di tipo parametrico.

L'analisi di una serie temporale può avere diversi obiettivi:

- Descrivere sinteticamente l'andamento nel tempo di un fenomeno; il grafico di una serie, in particolare, mette facilmente in evidenza sia eventuali regolarità, sia valori anomali.
- Spiegare il fenomeno, individuando il suo meccanismo generatore ed eventuali relazioni con altri fenomeni.
- Filtrare la serie; con ciò si intende la scomposizione della serie stessa nelle sue componenti non osservabili.
- Prevedere l'andamento futuro del fenomeno.

Nelle sezioni successive verrà utilizzata `Statsmodels`, la libreria di Python introdotta nel Capitolo 2, per analizzare la serie temporale delle vendite della farmacia di Amatrice.

## 7.2 Analisi dei componenti della serie temporale

Qualsiasi serie temporale può essere suddivisa nei seguenti componenti:

- *Livello di base*: i valori medi nella serie temporale;
- *Trend*: una pendenza crescente o decrescente nella serie temporale;
- *Stagionalità*: un pattern distinto ripetuto tra intervalli regolari, dovuto a fattori stagionali;
- *Errore*: le variazioni casuali nella serie temporale.

Non è obbligatorio che tutte le serie temporali debbano avere un trend e/o una stagionalità.

Potrebbero, dunque, esistere serie temporali, ovvero serie:

- con trend e stagionalità;
- con trend ma senza stagionalità;
- senza trend ma con stagionalità;
- senza trend e senza stagionalità.

A seconda della natura del trend e della stagionalità, una serie temporale può essere modellata come additiva oppure moltiplicativa. In una serie additiva ciascuna osservazione della serie può essere espressa come una somma dei componenti:

$$\text{valore} = \text{livello di base} + \text{trend} + \text{stagionalità} + \text{errore.}$$

In una serie moltiplicativa ciascuna osservazione della serie può essere espressa come un prodotto dei componenti:

$$\text{valore} = \text{livello di base} * \text{trend} * \text{stagionalità} * \text{errore.}$$

È possibile effettuare una decomposizione di una serie temporale nei suoi quattro componenti supponendo che la serie sia additiva oppure moltiplicativa.

Per analizzare le componenti della serie temporale delle vendite della farmacia in esame è stato creato un file `.csv`, chiamato `VenditeFarmaciaAmatrice.csv`, il quale risulta composto da due campi:

- **Data**, ovvero la coppia Mese/Anno;
- **Ricavato**, ovvero le entrate in euro ottenute in quello stesso mese.

Dapprima è stato generato il grafico della serie temporale, osservabile in Figura 7.1, tramite il codice mostrato nel Listato 7.1.

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import statsmodels.api as sm
4 from statsmodels.tsa.seasonal import seasonal_decompose
5 from dateutil.parser import parse
6
7 # Caricamento delle vendite della farmacia di Amatrice dal 2015 al 2020
8 df = pd.read_csv('gdrive/My Drive/data/VenditeFarmaciaAmatrice.csv', delimiter=';', parse_dates=["Data"],
9                 index_col="Data")
10
11 #Plotting del grafico della serie temporale
12 ax = df['Ricavato'].plot(figsize = (18,10))
13 ax.set(ylabel='Vendite', xlabel="Data", title='Vendite di una farmacia di Amatrice')
```

Listato 7.1: Generazione del grafico della serie temporale



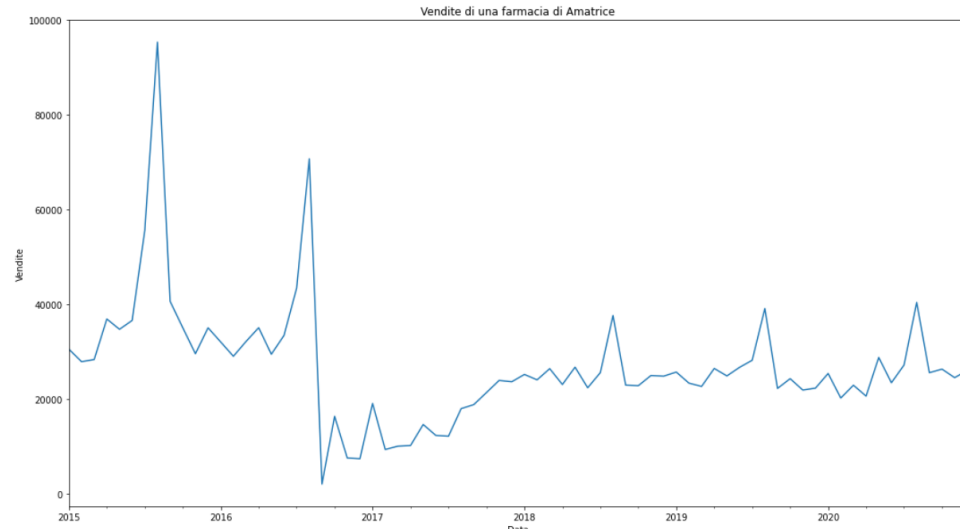


Figura 7.1: Grafico della serie temporale delle vendite della farmacia di Amatrice

Successivamente è stata applicata su tale serie temporale sia la decomposizione moltiplicativa che quella additiva; il codice relativo a queste due operazioni è illustrato nel Listato 7.2.

```

1 # Decomposizione Moltiplicativa
2 result_mul = seasonal_decompose(df['Ricavato'], model='multiplicative', extrapolate_trend='freq')
3
4 # Decomposizione Additiva
5 result_add = seasonal_decompose(df['Ricavato'], model='additive', extrapolate_trend='freq')
6
7 # Generazione dei grafici delle componenti della serie temporale
8 def plotseasonal(res, axes, title):
9     res.observed.plot(ax=axes[0], legend=False, title=title)
10    axes[0].set_ylabel('Osservata')
11    res.trend.plot(ax=axes[1], legend=False)
12    axes[1].set_ylabel('Trend')
13    res.seasonal.plot(ax=axes[2], legend=False)
14    axes[2].set_ylabel('Stagionalità')
15    res.resid.plot(ax=axes[3], legend=False)
16    axes[3].set_ylabel('Residuo')
17
18    fig, axes = plt.subplots(ncols=2, nrows=4, sharex=True, figsize=(18,10))
19
20    plotseasonal(result_mul, axes[:,0], 'Decomposizione Moltiplicativa')
21    plotseasonal(result_add, axes[:,1], 'Decomposizione Additiva')
22
23    plt.tight_layout()
24    plt.show()

```

Listato 7.2: Decomposizione moltiplicativa e additiva della serie temporale

In Figura 7.2 è possibile osservare i risultati sia della decomposizione moltiplicativa che di quella additiva.

Si nota subito che le due decomposizioni hanno restituito grafici con andamenti identici per il trend e molto simili per la stagionalità. Partendo dal trend, secondo quanto affermato nel Capitolo 4, si può pensare che la pendenza decrescente che si registra tra il primo trimestre del 2015 e quello del 2017 sia dovuta alla temporanea chiusura della farmacia e all'ingente spopolamento che ha colpito Amatrice a seguito del sisma del Centro Italia.

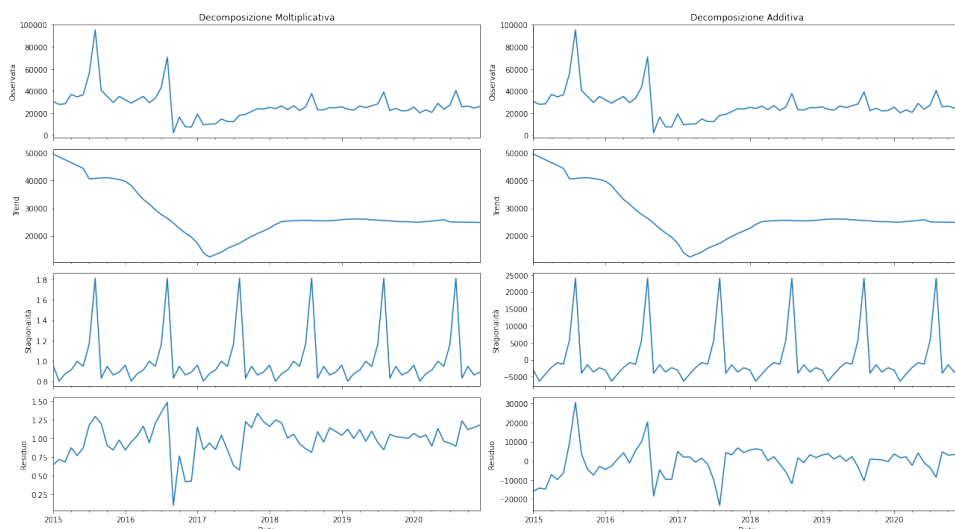


Figura 7.2: Grafici dei componenti della serie temporale ottenuti tramite decomposizione moltiplicativa ed additiva

D'altro canto, la pendenza positiva che si osserva tra l'inizio del 2017 e l'inizio del 2018 può essere spiegata dalla parziale ripresa dell'attività in seguito al trasferimento presso un centro commerciale provvisorio e alla costruzione delle SAE, che hanno permesso ad un'importante parte degli sfollati di tornare a vivere nel territorio amatriciano. Dal secondo trimestre del 2018 in poi il trend presenta un andamento quasi costante, in quanto, essendo i primi progetti di ricostruzione partiti soltanto nel 2021, non si sono registrate variazioni importanti della popolazione, e, conseguentemente, delle vendite.

Per quanto riguarda la stagionalità, entrambi i grafici confermano il pattern delle vendite già discusso nel Capitolo 4, il quale presenta il picco maggiore nel mese di agosto, mese in cui nel paese si rileva la massima affluenza di turisti e proprietari di seconde case in ferie, e picchi minori in corrispondenza delle festività natalizie e pasquali.

Infine, per quanto riguarda i residui, i quali consistono nelle differenze tra valori teorici e osservati, si può intravedere la presenza di pattern nella decomposizione additiva, mentre, in quella moltiplicativa, l'andamento è assolutamente casuale. Quindi, per tale serie temporale, si dovrebbe preferire la decomposizione moltiplicativa, in quanto, nell'approccio tradizionale, si assume che i residui siano dovuti al caso e, pertanto, assimilabili ad errori accidentali.

### 7.3 Costruzione di un modello per la previsione della serie temporale

Esistono diversi metodi per la previsione delle serie temporali, i quali si focalizzano principalmente sulle relazioni lineari. Dal momento che la serie temporale oggetto

dell'analisi risulta univariata con trend e stagionalità, è stato selezionato il metodo *Holt Winter's Exponential Smoothing (HWES)* per costruire un modello predittivo.

Il metodo *Double Exponential Smoothing*, altrimenti detto *metodo di Holt*, prevede un'equazione di previsione e due equazioni di *smoothing*, una per il livello di base e una per il trend. L'equazione di previsione contiene l'equazione di livello e l'equazione di trend: la prima è funzione di  $\alpha$ , ovvero il parametro di smoothing per il livello, e dei valori precedenti di livello e di trend, mentre la seconda è funzione di  $\beta^*$ , ovvero il parametro di smoothing per il trend, e dei valori precedenti di livello e di trend.

Poiché la stagionalità non è ancora considerata in questo metodo, il modello finale sarà una semplice linea retta inclinata che si estende dai punti di dati più recenti.

$$\begin{array}{ll} \text{Forecast equation} & \hat{y}_{t+h|t} = \ell_t + hb_t \\ \text{Level equation} & \ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \end{array}$$

Figura 7.3: Equazioni del metodo di Holt

Il metodo *Triple Exponential Smoothing*, altrimenti detto *metodo di Holt-Winters*, comprende un'equazione di previsione e tre equazioni di smoothing, una per il livello di base, una per il trend e una per la stagionalità, alle quali corrispondono, rispettivamente, i parametri di smoothing  $\alpha$ ,  $\beta^*$  e  $\gamma$ . La  $m$  indica la frequenza di stagionalità, ovvero il numero di stagioni in un anno. Ad esempio, per i dati trimestrali  $m = 4$ , mentre per i dati mensili  $m = 12$ .

Esistono due varianti di questo metodo che differiscono per la natura della componente stagionale, che può essere moltiplicativa o additiva. Di seguito sono illustrate le equazioni sia per il metodo additivo che per quello moltiplicativo. Il parametro  $k$  che compare nelle equazioni di previsione è la parte intera della frazione  $(h - 1)/m$ , la quale garantisce che le stime degli indici stagionali utilizzati per le previsioni provengano dall'ultimo anno del campione.

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\ \ell_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}, \end{aligned}$$

Figura 7.4: Metodo additivo di Holt-Winters

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$$

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m}$$

Figura 7.5: Metodo moltiplicativo di Holt-Winters

Da notare che il livello e il trend sono simili per i due metodi, mentre il calcolo della stagionalità è differente. Infatti, il metodo additivo mostra una relazione lineare tra il valore stimato  $y$  e il fattore di stagionalità  $s$ , mentre il metodo moltiplicativo mostra una relazione moltiplicativa tra i due.

In `Statsmodels` il metodo di Holt-Winters è implementato tramite la funzione `ExponentialSmoothing()`. L'utente può applicare sia il *Double Exponential Smoothing* che il *Triple Exponential Smoothing*, specificando, rispettivamente, i parametri “trend” e “seasonal” all'interno di tale funzione.

Nel Listato 7.3 è illustrato il codice tramite il quale è stato definito il modello predittivo per la serie temporale delle vendite della farmacia di Amatrice.

```

1 from statsmodels.tsa.holtwinters import ExponentialSmoothing
2
3 #Definizione del modello predittivo per la serie temporale tramite i metodi DoubleExponentialSmoothing e
  TripleExponentialSmoothing
4 df.index.freq = 'MS'
5 df['DES_12_add'] = ExponentialSmoothing(df['Ricavato'], trend='add').fit().fittedvalues.shift(-1)
6 df['DES_12_mol'] = ExponentialSmoothing(df['Ricavato'], trend='mul').fit().fittedvalues.shift(-1)
7 df['TES_12_add'] = ExponentialSmoothing(df['Ricavato'], trend='add', seasonal='add', seasonal_periods=12).fit().
  fittedvalues
8 df['TES_12_mol'] = ExponentialSmoothing(df['Ricavato'], trend='mul', seasonal='mul', seasonal_periods=12).fit().
  fittedvalues
9 df[['Ricavato', 'DES_12_add', 'DES_12_mol', 'TES_12_mol', 'TES_12_add']].iloc[-128:].plot(figsize=(12,6)).autoscale(
  axis='x', tight=True)

```

Listato 7.3: Definizione del modello predittivo della serie temporale

In Figura 7.6 è possibile osservare l'applicazione dei quattro metodi che costituiscono il modello predittivo della serie temporale in esame.

Da tale grafico si evince i metodi *TES* (*Triple Exponential Smoothing*) sono in grado di descrivere i dati della serie temporale più efficacemente dei metodi *DES* (*Double Exponential Smoothing*). Ciò era altamente prevedibile, dal momento che ci si può rendere conto, anche visivamente, della stagionalità che caratterizza il dataset a disposizione.

In particolare, il metodo `TES_12_mol` risulta avere le prestazioni migliori, mentre il metodo `TES_12_add` non sembra molto adatto per la previsione delle serie tem-

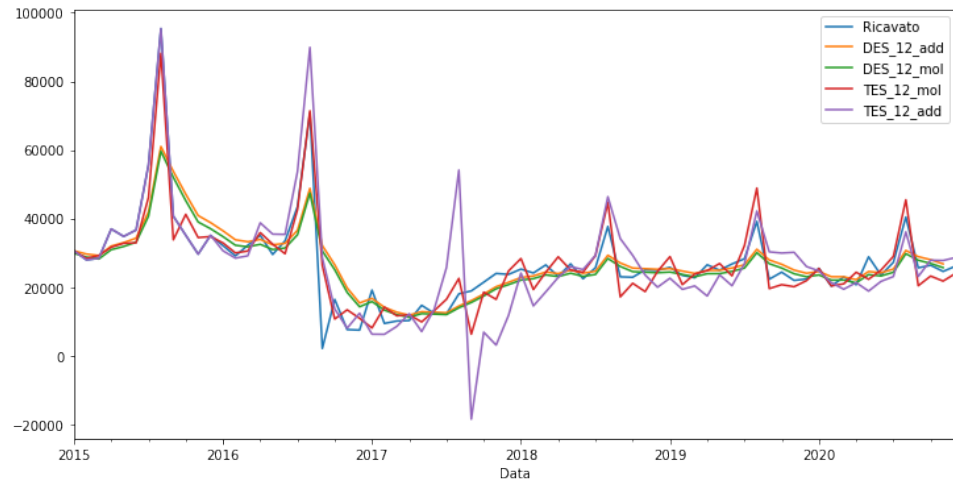


Figura 7.6: Applicazione del modello predittivo della serie temporale

porali che mostrano dei disturbi significativi della stagionalità a causa di eventi inaspettati e dal forte impatto, come proprio il sisma del Centro Italia.

D'altro canto, si può notare che i metodi *DES* forniscono delle prestazioni accettabili per la porzione di serie temporale relativa al 2017, nella quale la componente stagionale è venuta meno poiché gran parte della popolazione era alloggiata fuori Amatrice per la mancanza di abitazioni sul territorio, e, conseguentemente, l'afflusso di clienti presso la farmacia oggetto di studio era molto basso.



## Conclusioni

*In questo ultimo capitolo verranno esposte le conclusioni riguardanti l'analisi condotta.*

Nell'elaborato corrente è stata condotta un'analisi approfondita del dataset delle vendite di una farmacia di Amatrice tra il 2015 e il 2020, al fine di evidenziare i principali effetti causati dal terremoto del Centro Italia sul consumo dei farmaci.

Nel primo capitolo si è cercato di inquadrare il contesto dell'analisi, ripercorrendo l'intera sequenza sismica che danneggiò fortemente molti paesi del Centro Italia nel 2016 e nel 2017.

Successivamente, dopo aver introdotto i software di Business Intelligence utilizzati, e dopo aver descritto il dataset oggetto di studio, è stata illustrata la prima fase di analisi dei dati, che aveva l'obiettivo di descrivere lo status economico dell'attività commerciale nel corso del periodo considerato e di indagare le principali cause delle variazioni delle vendite, focalizzando l'attenzione su particolari famiglie di prodotti.

Come ci si poteva aspettare, il terremoto del Centro Italia, oltre a causare un'ingente diminuzione delle entrate, ha condizionato fortemente le vendite di alcuni articoli, quali, ad esempio, gli ansiolitici, gli antidepressivi, i farmaci veterinari e i prodotti per bambini.

Nel quinto capitolo sono state fornite, alla luce dei risultati ottenuti nel precedente, delle previsioni sull'andamento delle vendite della farmacia nel biennio 2021-2022.

L'ultima parte dell'elaborato è stata dedicata all'applicazione di metodi di Machine Learning, quali correlazione, classificazione e clustering, per prescrivere all'esercizio commerciale in questione delle azioni da intraprendere per migliorare il proprio status economico, e all'utilizzo della libreria di Python `Statsmodels` per analizzare la serie temporale delle vendite e costruire un modello predittivo della stessa.

Come possibile sviluppo futuro del progetto, si potrebbe ridefinire il dataset delle vendite, aggiungendo due campi, quali ID dello scontrino parlante ed età dell'acquirente, in modo tale da poter effettuare una profilazione dei clienti che frequentano la farmacia e, conseguentemente, pianificare campagne di marketing personalizzate.





---

## Riferimenti bibliografici

1. Che cos'è l'analisi predittiva. <https://www.redhat.com/it/topics/automation/how-predictive-analytics-improve-it-performance>.
2. Power BI. <https://powerbi.microsoft.com/en-us/>.
3. Qlik Sense. <https://www.qlik.com/it-it/products/qlik-sense>.
4. Statsmodels. <https://www.statsmodels.org/stable/index.html>.
5. Tableau. <https://www.tableau.com/it-it>.
6. Che cosa è la Business Intelligence? <https://www.extrasys.it/it/redblog/che-cosa-%C3%A8-business-intelligence>, 2016.
7. CNR : Appennino spezzato dopo il terremoto! Norcia si sposta di 30 cm a ovest, Montegallo di 40 verso est. <https://www.inmeteo.net/2016/11/03/cnr-appennino-spezzato-terremoto-norcia-si-sposta-30-ovest-montegallo-40-verso-est/>, 2016.
8. Dozens Dead in Italy After 6.2 Magnitude Earthquake Levels Entire Towns. <https://weather.com/safety/earthquake/news/central-italy-earthquake>, 2016.
9. La leggenda del sisma durato più di due minuti. <https://www.lastampa.it/cronaca/2016/09/01/news/la-leggenda-del-sisma-durato-piu-di-due-minuti-1.34797585>, 2016.
10. La piu' forte scossa di terremoto degli ultimi 35 anni in Italia: Magnitudo 6.5. <https://www.meteogiuliacci.it/meteo/articoli/cronaca/la-piu-forte-scossa-di-terremoto-degli-ultimi-35-anni-italia-magnitudo-65>, 2016.
11. Sequenza sismica tra le province di Rieti, Ascoli P., Perugia, Teramo e L'Aquila. <https://web.archive.org/web/20160827124928/http://terremoti.ingv.it/it/ultimi-eventi/1001-evento-sismico-tra-le-province-di-rieti-e-ascoli-p-m-6-0-24-agosto.html>, 2016.
12. Terremoto: 22mila le persone assistite. Renzi: 'Ce la faremo'. [https://www.ansa.it/sito/notizie/cronaca/2016/10/30/terremoto-oggi-nuova-scossa-di-magnitudo-6.1-epicentro-tra-norcia-e-precipi\\_ba5f3d9b-0a35-46c3-b4db-9a63e584cadf.html](https://www.ansa.it/sito/notizie/cronaca/2016/10/30/terremoto-oggi-nuova-scossa-di-magnitudo-6.1-epicentro-tra-norcia-e-precipi_ba5f3d9b-0a35-46c3-b4db-9a63e584cadf.html), 2016.
13. Terremoto, 281 morti e 388 i feriti Amatrice decimata: 221 le sue vittime 2.500 le persone senza una casa. [https://www.corriere.it/cronache/16\\_agosto\\_26/terremoto-terra-trema-ancora-nuova-bilancio-vittime-6d999598-6b4a-11e6-8bdd-2a860cc068c8.shtml](https://www.corriere.it/cronache/16_agosto_26/terremoto-terra-trema-ancora-nuova-bilancio-vittime-6d999598-6b4a-11e6-8bdd-2a860cc068c8.shtml), 2016.
14. Terremoto Centro Italia, 293 chiese, monumenti e palazzi storici danneggiati. Sopraluogo anche al Colosseo: "Tutto ok". <https://www.ilfattoquotidiano.it/2016/08/24/terremoto-centro-italia-chiese-e-monumenti-storici-danneggiati-sopraluogo-anche-al-colosseo-tutto-ok/2993874/>, 2016.

15. Terremoto, danni alle strade: sulla Salaria crolli e ponti inagibili. <https://www.romatoday.it/cronaca/terremoto-oggi-24-agosto-2016-salaria-situazione-strade.html>, 2016.
16. Terremoto di 6,5 tra Norcia e Preci. È il più forte in Italia dal 1980. "Si temono 100mila sfollati". [https://www.repubblica.it/cronaca/2016/10/30/news/terremoto\\_rieti-150898762/](https://www.repubblica.it/cronaca/2016/10/30/news/terremoto_rieti-150898762/), 2016.
17. Terremoto, forti scosse in Centro Italia. La prima di magnitudo 5,4, la seconda 5,9, terza 4,6. Sindaco di Ussita: "Paese finito". Tremano Roma e L'Aquila. [https://www.repubblica.it/cronaca/2016/10/26/news/terremoto\\_forte\\_scossa\\_sentita\\_a\\_roma\\_l\\_aquila\\_e\\_perugia-150644641/](https://www.repubblica.it/cronaca/2016/10/26/news/terremoto_forte_scossa_sentita_a_roma_l_aquila_e_perugia-150644641/), 2016.
18. Terremoto, la strage dei romani in ferie: settanta i morti. [https://roma.corriere.it/notizie/cronaca/16\\_agosto\\_27/terremoto-strage-romani-ferie-settanta-morti-273ec242-6bd3-11e6-8bdd-2a860cc068c8.shtml](https://roma.corriere.it/notizie/cronaca/16_agosto_27/terremoto-strage-romani-ferie-settanta-morti-273ec242-6bd3-11e6-8bdd-2a860cc068c8.shtml), 2016.
19. Terremoto: situazione drammatica ad Amatrice: i morti sono oltre 120 aumenta il numero dei dispersi. [https://www.ilmessaggero.it/rieti/rieti\\_terremoto\\_amatrice-1926688.html](https://www.ilmessaggero.it/rieti/rieti_terremoto_amatrice-1926688.html), 2016.
20. Hotel Rigopiano, estratti altri corpi. La speranza sopravvive dietro un ultimo muro. <https://www.rainews.it/dl/rainews/articoli/Hotel-Rigopiano-estratti-altri-corpi-da-albergo-sotto-valanga-soccorsi-vigili-del-fuoco-sopravvissuti-muro-da-abbattere-886b663b-d0be-4a7c-a4c9-ee9e21bbc137.html>, 2017.
21. I rilievi macrosismici dei terremoti del 2016-2017 in Italia centrale. <https://web.archive.org/web/20170923003232/https://ingvterremoti.wordpress.com/tag/terremoto-amatrice/>, 2017.
22. Terremoto, Centro Italia in ginocchio. Un morto nel teramano. Slavina su hotel, tre dispersi. [https://www.repubblica.it/cronaca/2017/01/18/news/scossa\\_di\\_terremoto\\_avvertita\\_anche\\_a\\_roma-156285725/](https://www.repubblica.it/cronaca/2017/01/18/news/scossa_di_terremoto_avvertita_anche_a_roma-156285725/), 2017.
23. Terremoto in Centro Italia, rapporto alla Ue: danni per oltre 23 miliardi. [https://www.repubblica.it/cronaca/2017/02/15/news/sisma\\_danni\\_per\\_oltre\\_23\\_miliardi-158355389/](https://www.repubblica.it/cronaca/2017/02/15/news/sisma_danni_per_oltre_23_miliardi-158355389/), 2017.
24. 24 agosto 2016, ore 3:36: due anni fa, il terremoto scuoteva il centro Italia. 730 giorni dopo, la ricostruzione è ferma al palo. <https://news-town.it/cronaca/21616-24-agosto-2016,-ore-3-36-due-anni-fa,-il-terremoto-scuoteva-il-centro-italia-730-giorni-dopo,-la-ricostruzione-%C3%A8-ferma-al-palo.html>, 2018.
25. Analisi descrittiva, predittiva e prescrittiva: le differenze e quali adottare per i Big Data Analytics. <https://www.bnova.it/blog/bnext/analisi-descrittiva-predittiva-prescrittiva/>, 2020.
26. Vedere, analizzare e comprendere i dati con Tableau. <https://www.bnova.it/blog/bnext/data-visualization-tableau/>, 2020.
27. Estela Bee Dagum. *Analisi delle serie storiche - modellistica, previsione e scomposizione*. Springer Verlag, Milano, 2002.
28. INGV. *Relazione di dettaglio: Perugia Mw 6.5 del 2016-10-30 06:40:17 UTC; versione del 2016-10-30 ore 11:27:08 UTC*. Roma, 30 ottobre, 2016.
29. INGV. *Relazione di dettaglio: Rieti Mw 6.0 del 2016-08-24 01:36:32 UTC; versione del 2016-08-24 ore 04:26:02 UTC*. Roma, 24 agosto, 2016.
30. Sheldon M. Ross. *Introduzione alla statistica, 2<sup>a</sup> ed.* Maggioli Editore, 2014.

---

## Ringraziamenti

*Il primo ringraziamento va alla mia famiglia, che mi ha permesso di portare a termine questo percorso universitario, mostrando una pazienza fuori dal comune e sostenendomi nei diversi periodi difficili che lo hanno caratterizzato.*

*Vorrei ringraziare il Prof. Ursino per avermi dato la possibilità di incentrare la mia tesi di laurea su di un argomento che ho molto a cuore e per la straordinaria disponibilità dimostrata nel corso della sua stesura.*

*Un ringraziamento speciale va al mio correlatore Luca Virgili, i cui consigli si sono rivelati fondamentali per ottenere dei buoni risultati a livello di analisi.*

*Ringrazio con affetto i miei amici di Amatrice per tutti i bei momenti passati insieme, che spero continueremo a vivere in futuro, e per esserci supportati a vicenda nei momenti tristi della nostra vita.*

*Ringrazio di cuore i miei compagni di corso, nonché amici, che, con tanti piccoli gesti e parole opportune, mi hanno aiutato ad affrontare e a risolvere i problemi che mi hanno afflitto in questi ultimi anni. È anche merito vostro se sono riuscito a raggiungere questo importante traguardo.*

*Ringrazio anche tutti gli amici che ho conosciuto in facoltà in questi anni, con i quali ho vissuto molti momenti divertenti e stimolanti che hanno reso la mia esperienza universitaria estremamente piacevole.*