**REPORT sito e-commerce “*Cioccorane*”**

**Gruppo Horror**

# INTRODUZIONE

Il report ha come obiettivo quello di analizzare i dataset messi a disposizione dell’azienda e-commerce Cioccorane, distributrice di articoli di elettronica, abbigliamento e libri.

Le analisi sono state condotte su cinque dataset, comprensivi dei dati sulle vendite dell’anno 2022 (“*transazioni*”), sul catalogo dei prodotti disponibili (database “*prodotti*”), sul database storico dei clienti (“*clienti*”), sulle recensioni di tutti i prodotti commercializzati (database “*ratings*”) e infine sulle spedizioni relative alle vendite dell’anno 2022 (dataset “*spedizioni*”).

I risultati finali delle analisi sono stati suddivisi in quattro sezioni:

1. prodotti;
2. clienti;
3. spedizioni;
4. vendite e transazioni.

Per una comprensione migliore del report, occorre chiarire alcune scelte riguardo i dati considerati nella analisi.

In tutte le analisi è stato utilizzato come valore di riferimento il **ricavo**, ottenuto moltiplicando la quantità dei prodotti acquistata per il prezzo unitario. Questa scelta è stata dettata dalla necessità di focalizzare le analisi sui ricavi effettivi dell’azienda.

La colonna dell’importo della transazione è stata considerata come **costo** delle **spese di spedizione**, in quanto si sarebbe venuta a creare una discrepanza con i dati; infatti, in alcuni casi il costo della transazione risulta addirittura inferiore al prezzo unitario di alcuni prodotti.

Le analisi hanno riguardato solo **l’anno 2022**, rappresentativo di tutti i valori inseriti come data di transazione (*transazioni*). Siccome *Cioccorane* è un’azienda e-commerce, solitamente in questa tipologia di esercizi commerciali, la ricevuta di acquisto viene emessa in automatico al momento della transazione online e, di conseguenza, viene dichiarato nel fatturato di quell’anno.

Infine, tutti i valori relativi ad alcuni grafici, specialmente quelli che illustrano dati sulle vendite, sono espressi in migliaia.

Fatte le dovute precisazioni, si passa ora ad illustrare le analisi effettuate.

# ANALISI PRODOTTI

**1.1 Analisi Categoria Prodotto**

Questa analisi è stata affrontata con due diverse modalità: tramite Excel e MySQL Workbench.

In **Excel** sono stati importati i dataset su *transazioni* e *prodotti*. I dati sono stati inseriti in una tabella pivot, al fine di visualizzare le vendite per categoria di prodotto ed è stato poi creato un grafico a barre per meglio visualizzare i risultati.

Lo scopo dell’analisi era quello di trovare la **categoria** più venduta e quella con un ricavo maggiore. La scelta di considerare la categoria, e non i singoli prodotti, è dipesa dalla natura del dataset; infatti, i nomi dei prodotti non sono risultati informativi riguardo la natura degli stessi (es. nel dataset i prodotti sono chiamati “Prodotto508”, non Jeans Bianco Modello WideLeg).

Dalle analisi è emerso che i **libri** rappresentano la categoria più venduta e quella con un ricavo maggiore, seguita dall’abbigliamento, ultima l’elettronica.

In **SQL** è stato interrogato il dataset dei prodotti e delle transazioni (vedi query5, file Zip).

I risultati sono in linea con le analisi effettuate su Excel.

Nello specifico sono state vendute:

1. 865 unità della categoria **libri**, con un ricavo di **€ 239˙790;**
2. 834 prodotti della categoria **abbigliamento**, con un ricavo di **€ 222˙102;**
3. 768 prodotti della categoria **elettronica**, con un ricavo di **€ 189˙085**.

Le analisi hanno permesso di individuare in maniera immediata la categoria dei prodotti più richiesta dai clienti e, di conseguenza, quella che ha garantito all’azienda il ricavo maggiore.

Una volta individuata la categoria più richiesta, si è voluto individuare anche i prodotti più richiesti.

Per questa analisi più mirata, è stata eseguita una Query in SQL (vedi query2, file Zip).

Dalla Join tra i dataset *prodotti* e *transazioni*, emerge che i 3 prodotti più acquistati sono stati:

1. ***Prodotto 3477***, categoria Libri, 17 unità;
2. ***Prodotto 756***, categoria Libri, 16 unità;
3. ***Prodotto 3923***, categoria Elettronica, 13 unità.
   1. **Analisi Recensioni dei prodotti**

Per analizzare la qualità dei prodotti dell’e-commerce, sono state eseguite delle analisi sulle recensioni dei clienti, utilizzando sempre due diversi strumenti.

Iniziando da **Excel**, è stato importato il dataset *ratings*, i dati di interesse sono stati inseriti in una tabella pivot e poi rappresentati mediante un grafico a barre raggruppate.

Come prima operazione, è stata aggiunta la categoria “**Fuori Listino**” in quanto molti dei prodotti recensiti non ricadevano in nessuna delle categorie del dataset *“prodotti*” analizzato.

Si è ipotizzato che questi fossero prodotti fuori produzione, appartenenti ad un vecchio catalogo.

I punteggi delle recensioni sono su scala di punteggio da 1-5.

Dal grafico è visibile come la categoria **Fuori Listino** abbia in media un punteggio di recensione di poco superiore, si noti infatti come la scala sia da 2 a 2,7.

Visto che la differenza nei punteggi è risultata minima, è stato deciso di inserire delle statistiche descrittive dei punteggi, divisi per categoria

In generale, le recensioni medie sono molto basse (**Fuori Listino**, 2,61; **Abbigliamento** e **Libri**, 2,49; **Elettronica** 2,47). I prodotti “Fuori Listino” hanno, in media, recensioni più alte rispetto alle altre categorie, seppur con un numero totali di recensioni minore (conteggio: 184), soprattutto se messe a confronto con il numero delle recensioni dei libri (conteggio: 3163).

|  |  |
| --- | --- |
| *Rating Elettronica* | |
|  |  |
| Media | 2,47 |
| Mediana | 2 |
| Moda | 0 |
| Minimo | 0 |
| Massimo | 5 |
| Conteggio | 3270 |

|  |  |
| --- | --- |
| *Rating Abbigliamento* | |
|  |  |
| Media | 2,49 |
| Mediana | 3 |
| Moda | 3 |
| Minimo | 0 |
| Massimo | 5 |
| Conteggio | 3383 |

|  |  |
| --- | --- |
| *Rating Libri* | |
|  |  |
| Media | 2,49 |
| Mediana | 2 |
| Moda | 0 |
| Minimo | 0 |
| Massimo | 5 |
| Conteggio | 3163 |

|  |  |
| --- | --- |
| *Rating "Fuori Listino"* | |
|  |  |
| Media | 2,61 |
| Mediana | 3 |
| Moda | 3 |
| Minimo | 0 |
| Massimo | 5 |
| Conteggio | 184 |

Il punteggio più frequente per la categoria **libri** ed elettronica è di **0**, quindi molti prodotti hanno una recensione del tutto negativa. Invece, per **abbigliamento** e **fuori listino**, la moda è **3**.

Si è tentato individuare il prodotto con la recensione media più alta (query 8), ma a causa della natura del dataset, questa analisi risulta essere poco informativa dato che sono presenti molti prodotti con media di punteggio 5.

* 1. **Analisi Magazzino Prodotti**

Per condurre un’indagine esaustiva sui prodotti, è stato analizzato il **magazzino** e in particolare sono state analizzate le **quantità disponibili** di prodotti, la **quantità media** per ogni categoria e il **prodotto più raro**.

Tramite SQL è stato interrogato il dataset *prodotti*.

Il totale dei prodotti disponibili in magazzino è di **253˙065** (vedi Query16).

In generale, la produzione risulta nettamente sovrastimata e questo ha portato ad un accumularsi di merce invenduta: per cui a fronte di 2467 prodotti venduti, il magazzino consta di 253˙065 prodotti. Insomma, è stato venduto solo lo **0,97%** della produzione.

Il valore della merce in magazzino ammonta a € 64˙402˙499 a fronte di un ricavo annuo di soli

€ 650˙976, ovvero il **venduto rappresenta in proporzione l’1% del valore del magazzino**.

Per conoscere, nello specifico, la disponibilità media dei prodotti per ogni categoria, è stata eseguita una query (vedi query10). Il risultato mostra che, in media, sono disponibili circa **50 prodotti** per ogni categoria. Come era da attendersi, la categoria con maggior disponibilità è quella Elettronica (media 50,82, disponibilità totale 20˙821), seguita da Libri (media 50,68, disponibilità totale 20˙152) e infine Abbigliamento (media 50,36, tot. 22˙373), seppur la differenza di disponibilità tra le tre categorie è davvero minima.

I prodotti più rari, ottenuti considerando solo quelli la cui quantità è inferiore alla media, sono i prodotti (vedi Query13):

* **Prodotto 1358**, categoria Abbigliamento;
* **Prodotto 1414**, categoria Libri;
* **Prodotto 144**, categoria Elettronica;
* **Prodotto 1482**, categoria Elettronica;
* **Prodotto 1495**, categoria Abbigliamento.

Tali risultati sono stati ottenuti mettendo in ordine alfabetico il nome dei prodotti, dato quelli con quantità disponibile uguale a 1 erano molti.

Concludendo, si rileva una relazione direttamente proporzionale alle quantità vendute.

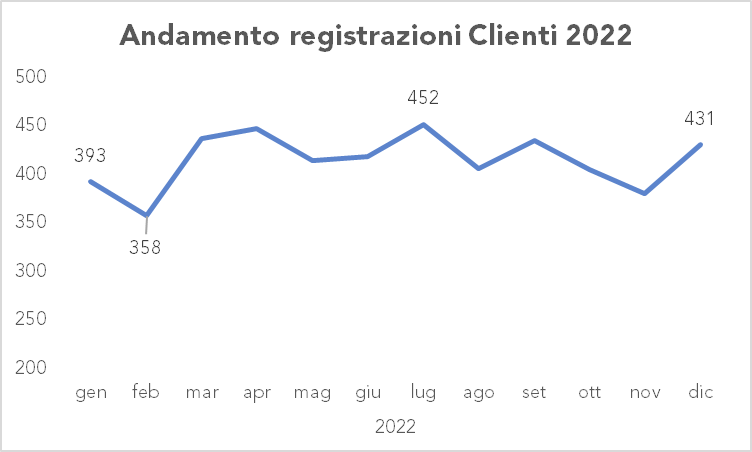
# ANALISI CLIENTI

L’analisi del comportamento dei clienti è stata condotta allo scopo di identificarne il comportamento d’acquisto. Inoltre, sono stati identificati i clienti più fedeli e quelli che, ad oggi, non risultano aver ancora effettuato nessuna transazione. Tutte le analisi sono state condotte in SQL.

Innanzitutto, sul sito si sono registrati 5˙000 clienti, con una percentuale di registrazione di circa il 9% sul totale dei clienti al mese.

Il mese che ha registrato più iscrizioni è quello di **luglio 2022**, con **452** clienti registrati; il mese con meno registrazioni risulta quello di **febbraio** del 2022 (358). (query12).

Il mese di luglio incide per lo 0,9% sul totale annuo delle registrazioni del 2022.



Dei 5˙000 clienti registrati, quelli che ancora non hanno effettuato **nessuna transazione** sono **4˙525** (query17), numero piuttosto elevato.

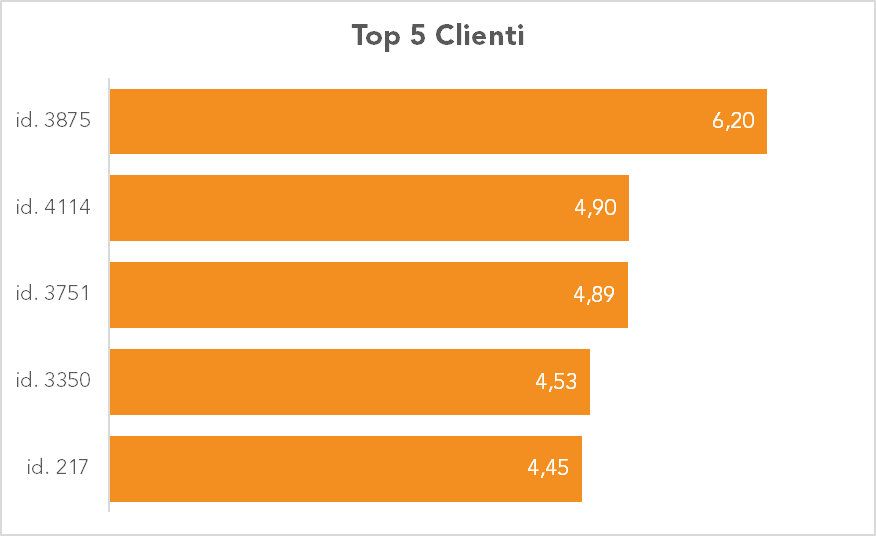
Per analizzare la frequenza di acquisto, sono state conteggiate le transazioni effettuate.

Da questa analisi è emerso che i clienti che hanno acquistato più volte sono stati i clienti con **ID 700** e **ID** **4063**, che hanno effettuato un totale di 3 transazioni (vedi query3).

Per approfondire ulteriormente quest’aspetto è stato identificato anche il cliente che ha speso di più e che quindi ha effettuato la transazione con un valore più alto, con l’intento di fare un confronto tra il numero di transazioni e la spesa totale.

Il cliente con la transazione di valore più alta è quello con **ID 3875**, con un valore totale di **€ 6˙199,83**. (vedi query6). Nello specifico il cliente ha acquistato due prodotti: il **prodotto 519** e il **51** (analisi 3 di Excel).

Infine, sono stati individuati i 5 clienti che hanno speso di più, così come è visibile nel grafico di seguito (analisi 5 Excel e query 14 SQL).



È stato creato anche un grafico in cui vengono riportati i nomi dei prodotti acquistati dai nostri top cinque clienti (Analisi 5 Excel).

Il grafico non è molto indicativo, in quanto il nome dei prodotti è troppo generico e non è subito visibile il cliente che ha effettuato la transazione con valore più alto.

# ANALISI SPEDIZIONI

Per valutare il servizio **logistica**, sono state condotte analisi che hanno permesso di confrontare i due metodi di spedizioni (**Posta Prioritaria** e **Corriere Express**) e l’incidenza delle spedizioni avvenute correttamente in confronto a quelle che risultano ancora in “Consegna”. Infine, è stato individuato il Metodo di Spedizione più utilizzato.

È stato possibile così constatare come la logistica presenti delle problematiche.

Come mostra il grafico di seguito, più della metà dei prodotti spediti non sono stati ancora consegnati. (analisi 2 Excel e query7 e 19 SQL). Delle spedizioni effettuate solo il **48,7%** sono state consegnate.

Analizzando i due metodi di spedizione, il servizio logistico con tasso di spedizione di riuscita maggiore è Posta Prioritaria (**51%**). La differenza con Corriere Express è comunque minima.

Successivamente, attraverso SQL, è stato individuato il metodo di spedizioni più utilizzato. Dall’analisi è emerso che il metodo di spedizioni più utilizzato è Posta Prioritaria. (vedi query11)

In conclusione, Posta Prioritaria sembra essere non solo il metodo più utilizzato ma anche quello più efficiente, con un tasso di successo maggiore rispetto al Corriere Express. Tuttavia, dall’analisi si evince un grosso problema legato alla logistica, infatti la media dei tempi di attesa affinché la merce venga consegnata è di 375 giorni (vedi query3 aggiuntiva). Inoltre, dal dataset emerge che in media i clienti attendono 377,58 giorni prima che la merce venga loro spedita, con una mediana pari a 372 giorni, con conseguente aumento di costi di magazzino.

# ANALISI VENDITE e TRANSAZIONI

Sono state eseguite una serie di analisi per monitorare i dati sulle transazioni, quali vendite di prodotti, importi di acquisto e variazioni mensili e trimestrali.

Innanzitutto, l’analisi è partita dalla ricerca del valore medio di ogni transazione, suddivisa per categoria. In SQL (vedi Query4), è emerso che la categoria libri, anche considerando il valore medio della transazione, si riconferma come quella più redditizia, con un importo medio di € 1˙444, seguita dalla categoria dell’abbigliamento (€ 1˙269) e dall’elettronica (€ 1˙189).

Dalla singola transazione, si è passato poi a fare un’analisi annuale delle vendite (query18).

In totale, nel **2022** c’è stato un ricavo di **€ 650˙976,55**.

È stato analizzato l’andamento delle vendite mensili in SQL (query1) e in Excel (analisi 4).

Dal grafico si può osservare come il mese con il ricavo maggiore è stato quello di **agosto**, con un ricavo di € 67˙857,31; seguono marzo e maggio con un ricavo piuttosto simile (rispettivamente,

€ 64˙096,96 e € 64˙021,33). Il mese con il **ricavo minore** è stato **luglio** (€ 28˙844,91).

Si osserva inoltre come l’andamento sia stato costante nei mesi da febbraio a maggio e da settembre a dicembre, picchi minimi si osservano invece nel mese di febbraio e luglio. In definitiva il mese di **agosto** si conferma essere il mese migliore per le vendite. (vedi query 15, il mese migliore per le vendite). Riveliamo inoltre un rapporto inversamente proporzionale con le registrazioni dei clienti, in cui si registra un picco nel mese di luglio. Occorrerebbe condurre un’ulteriore analisi per capire la natura di questa relazione.

Per un ulteriore approfondimento è stato creato un grafico aggiuntivo che mostra l’andamento delle vendite raggruppato per categoria (analisi 4 Excel).

È stata condotta poi un’analisi degli acquisti dei clienti nel tempo, focalizzandosi sui trimestri dell’anno 2022 (analisi 6 Excel). Dal grafico si può osservare come il **secondo trimestre** sia per l’azienda il più redditizio, con un ricavo di € 184˙004 mentre il trimestre meno redditizio è stato il **primo**, con un ricavo di €147˙950.

Infine, è stata calcolata la variazione mensile delle vendite, non tenendo in conto il mese di gennaio in quanto non disponiamo dei dati dell’anno 2021.

Dal grafico (query9) è possibile notare, come il terzo trimestre sia quello che presenta variazioni più significative, avendo sia il picco minimo a luglio che il massimo ad agosto.

# CONCLUSIONI

Dalle indagini sulle categorie, non è emersa nessuna categoria con un ricavo nettamente superiore rispetto alle altre. In ogni caso, visto che comunque la categoria libri rappresenta, anche se di poco, quella più venduta, si potrebbe provare ad investire su una pubblicità maggiore per le altre categorie, aumentandone la visibilità. Per *Cioccorane,* l’eccessiva diversificazione delle merci, invece che rappresentare un punto di forza potrebbe invece essere un punto di debolezza, in quanto i ricavi totali non compensano affatto i costi di magazzino.

Infatti, quest’ultimi risultano insostenibili se messi in relazione ai ricavi, con il venduto che rappresenta in proporzione l’1% del valore del magazzino. Per diminuire i costi del magazzino e per compensare i mesi in cui si sono registrate meno vendite, si può valutare di fare delle offerte, considerando anche che il prezzo unitario dei singoli prodotti è molto elevati.

Si suggerisce di fare una campagna di sconti mirata, una volta individuati i clienti che hanno fatto più acquisti e che hanno speso di più. Inoltre, per incentivare gli acquisti, dato la maggior parte dei clienti registrati non ha ancora effettuato alcuna transazione, si potrebbe pensare di aggiungere uno sconto ulteriore per chi effettua la prima transazione.

La valutazione dei prodotti risulta inoltre sotto la media per tutte le categorie.

Potrebbe essere interessante svolgere un approfondimento sulle recensioni, in modo da capire il motivo di punteggi così bassi, se dovuti alla spedizione o altro.

In effetti le spedizioni risultano un altro punto critico dell’azienda, sia per i tempi di attesa troppo lunghi, sia per il tasso di riuscita delle consegne.

In conclusione, l'analisi dei dataset ha permesso di ottenere una panoramica dettagliata sull'azienda, evidenziando punti di forza e criticità. Ha inoltre fornito informazioni importanti per comprendere il mercato e il comportamento dei clienti dell'azienda *Cioccorane*. Le conclusioni possono essere utilizzate per prendere decisioni strategiche e migliorare l'efficienza delle operazioni aziendali. Si ritiene, infatti, che effettuate ulteriori analisi e ricerche, le informazioni ottenute si riveleranno utili soprattutto per facilitare gli stakeholders nell’orientare le strategie future dell’azienda per migliorare la soddisfazione dei clienti e accrescere così il business.

**Federica Branca**

**Margot Stefanetti**

**Marco Cerri**

**Martina Albano**

**Sandy Camargo**