

Università Politecnica delle Marche

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Informatica e dell'Automazione



Analisi con Tableau sulla vendita di telefoni in Indonesia

DOCENTI

Prof. Ursino Domenico

Prof. Luca Virgili

STUDENTI

Mori Nicola

Sospetti Mattia

Zitoli Francesca

Anno accademico 2021-2022

Indice

1	Introduzione	2
2	Prodotti	2
3	Luogo	4
4	Analisi temporale	6
5	Previsioni	8

1 Introduzione

Per quanto riguarda l'analisi del dataset in Tableau, abbiamo deciso di ricalcare la suddivisione in viste già affrontata nel Capitolo relativo a Qlik, pertanto le viste analizzeranno i prodotti, i luoghi ed una analisi temporale.

A quanto già fatto, si aggiunge un'ulteriore visualizzazione, che sfrutta la caratteristica chiave del tool Tableau, vale a dire Previsioni, che, per quanto il nome sia già di per sé esplicativo, fa una analisi delle previsioni offerte da Tableau in termini di prodotti più venduti e vendite giornaliere.

Le seguenti sottosezioni si occuperanno di presentare e descrivere approfonditamente le viste singolarmente.

2 Prodotti

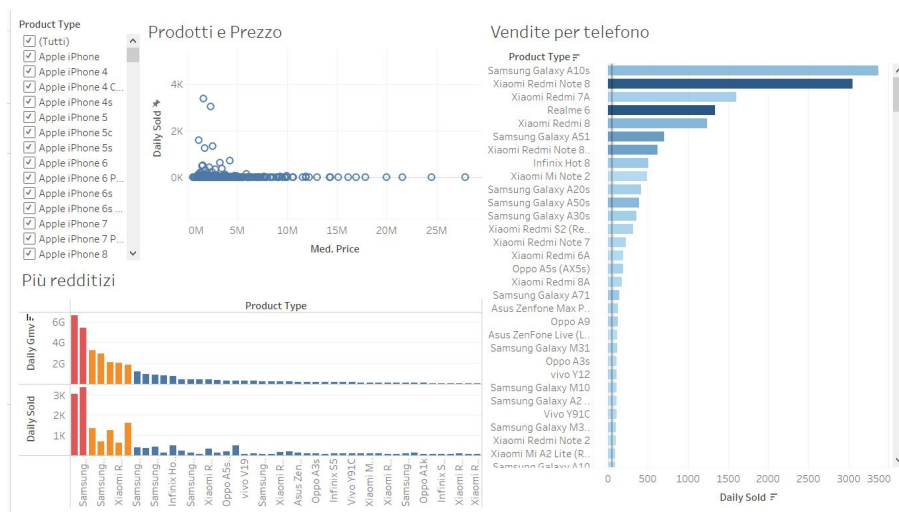


Figura 1: Dashboard sulle vendite dei prodotti

In questa dashboard (Figura 1) si può notare come si è voluta porre l'attenzione sul prodotto venduto, sulla sua varietà e alcune curiosità che porta con sé.

La dashboard è corredata di un filtro, mostrato nella colonna di sinistra, che permette di selezionare solo una sottosezione di telefoni venduti, mostrando così i risultati che ognuno di noi vuole vedere, ignorando eventuali informazioni poco rilevanti in termini di analisi.

Scorrendo dal filtro verso destra incontriamo lo scatterplot “Prodotti e Prezzo”, che, come suggerisce il nome, mostra in che modo sono relazionati il prezzo di vendita e le quantità vendute.

I risultati mostrati non ci stupiscono, sono in linea con quanto ci potevamo aspettare, e non sono presenti stravolgimenti: del resto notiamo come ci siano numerosi prodotti che si accumulano nella zona in basso a sinistra, a mostrare come in realtà ci sia una grande varietà di telefoni, tra cui però ne emergono solo pochi, che sono quelli invece in alto, che si distinguono dagli altri per un valore maggiore del “Daily Sold”. I telefoni presenti in quella zona sono quelli che troviamo anche nell’istogramma alla sua destra, con il titolo “Vendite per telefono”, che riporta in valore assoluto il numero di unità vendute per telefono (Figura 2).

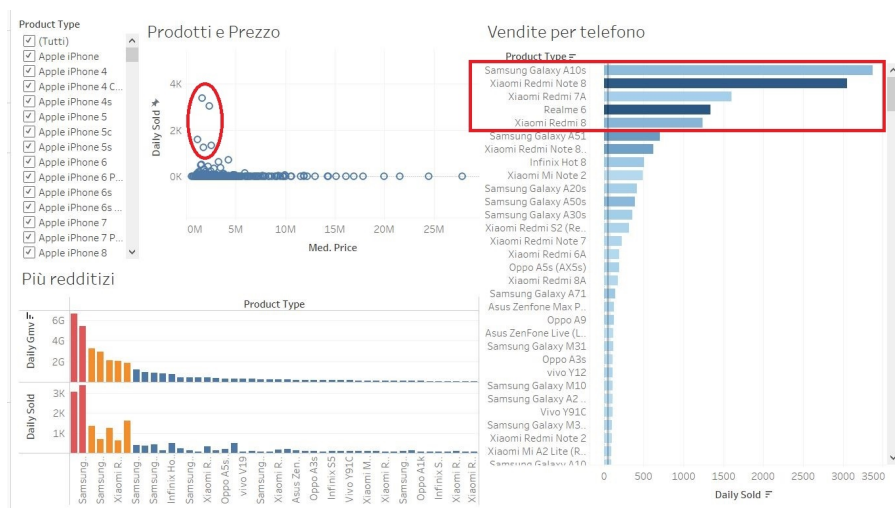


Figura 2: Focus sui prodotti più venduti

Qui evidenziamo quindi i top 5 telefoni venduti, che corrispondono alle prime 5 righe dell’istogramma, questo a dimostrazione della correttezza dello scatterplot.

Continuando a focalizzarci sullo scatterplot, possiamo avanzare anche delle ipotesi, o per lo meno delle considerazioni, dal punto di vista del prezzo del singolo prodotto: infatti possiamo vedere come i prodotti più venduti abbiano un prezzo relativamente basso (sono nell’estremità sinistra del grafico, quindi con un valore di “Med. Price” basso), mentre invece i prodotti nella zona a destra del grafico, con un prezzo sensibilmente maggiore, abbiano venduto davvero poco.

Ma possiamo dire che è un risultato perfettamente in linea con le abitudini economiche di un paese come quello dell’Indonesia, che presenta un PIL pro

capite di 3.869,59 USD, un valore che di certo rende i telefono più costosi, come iPhone o Samsung Galaxy Fold, assolutamente fuori portata a meno di rare eccezioni.

Infine il grafico in basso mostra i prodotti più redditizi, ovvero i prodotti che, al netto delle vendite, hanno fatto intascare più denaro nelle casse dei commercianti. Abbiamo deciso di effettuare anche un clustering qui, per mostrare 3 nette categorie di prodotti, colorati in modo diverso (rosso, arancio, blu) nel grafico in base alla redditività di ciascuno di essi.

Chiaramente il primo cluster è composto dai 2 telefoni più venduti per distacco, ovvero il Samsung Galaxy A10 ed il Xiaomi Redmi Note 8.

Il secondo cluster presenta quei prodotti con un tasso di redditività medio, al cui interno non sono presenti stravolgimenti se confrontiamo i risultati con i grafici posti sopra.

Infatti contiene al suo interno i 5 telefoni che occupano dalla terza alla settima posizione in termini di unità vendute, ma con delle piccole discrepanze a livello di ordinamento: infatti notiamo come ad esempio il Samsung Galaxy A51 sia il sesto dispositivo più venduto, ma guadagna ben 2 posizioni in termini di redditività, e questo guadagno può essere facilmente spiegato da un prezzo leggermente maggiore rispetto ai prodotti che lo seguono, pertanto per un dispositivo del genere basta una quantità minore di unità vendute per generare lo stesso guadagno ai commercianti.

Il terzo ed ultimo cluster infine contiene i prodotti meno redditizi, e quindi la loro vendita è tendenzialmente sconsigliata ai commercianti, preferendo i prodotti appartenenti alle prime due categorie sopra esposte.

3 Luogo

La seconda dashboard (Figura 3) si focalizza invece sull'aspetto geografico delle vendite dei prodotti.

Lo scatterplot in alto, che riporta il titolo "Vendite città prezzo medio" mostra la relazione che intercorre tra il prezzo medio delle unità vendute nelle città e le quantità vendute.

Anche qui abbiamo deciso di effettuare una attività di clustering, pensando fosse interessante categorizzare le città in 3 macrocategorie, ovvero non proficue, proficue, fuori target.

La prima categoria, rappresentata dal colore blu nel grafico, presenta quelle città che purtroppo presentano un valore di Daily Sold abbastanza basso (meno di 2k unità vendute) a fronte di un valore di "Med. Price" che invece dovrebbe essere abbordabile per i cittadini locali.

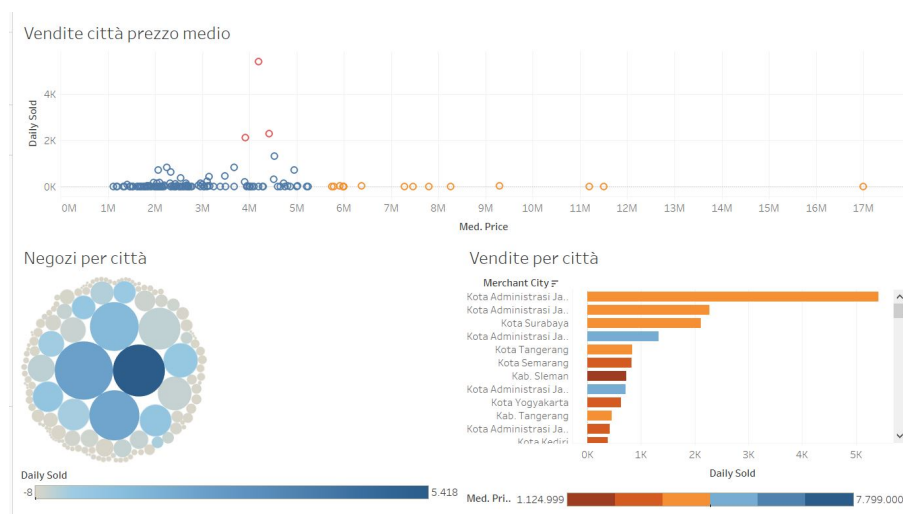


Figura 3: Dashboard sulla localizzazione

Un risultato di questo tipo può essere spiegato con diverse ipotesi plausibili: possiamo pensare ad una densità bassa di abitanti in quelle città, che quindi spiegherebbe le poche vendite, oppure una posizione geografica poco convincente, che quindi spingerebbe le persone a spostarsi verso la capitale (Jakarta) per fare acquisti.

Infatti il cluster “proficue” (in colore rosso) presenta solo 3 località, di cui 2 sono quartieri della capitale, ovvero Kota Administrasi Jakarta Pusat e Barat, che occupano le prime due posizioni dell’istogramma a destra dello scatterplot. Ciò ci spinge a pensare che le persone siano tendenzialmente concentrate nella capitale e nelle zone limitrofe, schiacciando verso il basso il numero di vendite nelle altre città del paese.

Il terzo ed ultimo cluster, “fuori target” (in colore arancio nello scatterplot), invece mostra tutte le città con poche unità vendute ed un prezzo medio degli smartphone venduti alto, quindi dei prodotti assolutamente fuori dalla portata dei cittadini indonesiana, e questo a giustificazione del nome che abbiamo deciso di dare al cluster.

Nell’istogramma a destra si ponga l’accento su un outlier, Kota Administrasi Jakarta Utari, che, essendo di colore blu, ha un valore di prezzo medio di vendita maggiore di quelli che gli stanno immediatamente sopra e sotto. Dunque vende tanto anche se i prezzi sono più alti di altre città, e questo potrebbe essere spiegato nel caso in cui ipotizzassimo che il quartiere di Utari sia il quartiere più ricco di Jakarta, con quindi abitanti disposti a spendere di più per questi beni.

Concludiamo l’analisi di questa dashboard con il grafico in basso a si-

nistra, “Negozi per città”, che quindi mette in chiaro la concentrazione di negozi di elettronica per ogni città. Abbiamo deciso di rendere più informativo quel grafico, dotando ogni bolla di un colore in base alla quantità di unità vendute. Dal grafico dunque si evince come l’andamento segua una power law, ovvero tanti negozi che vendono poco e molti negozi che vendono molto.

4 Analisi temporale

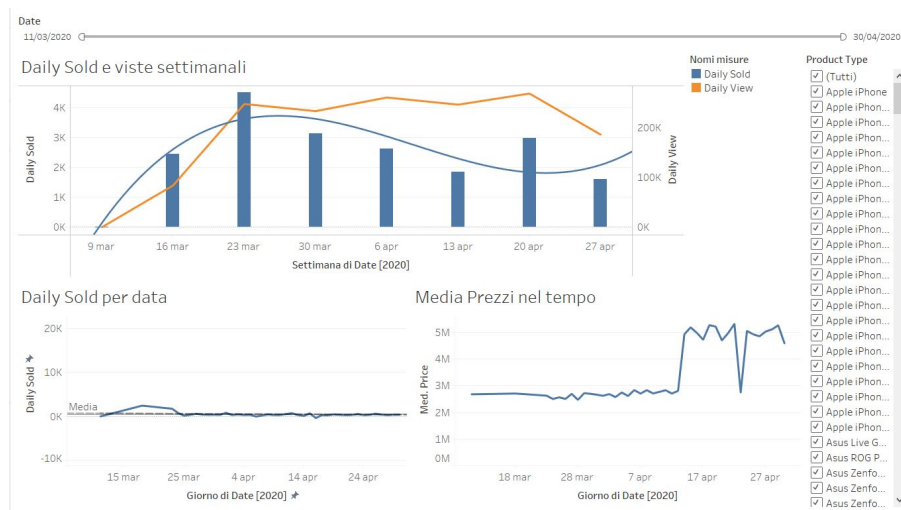


Figura 4: Dashboard sull’analisi temporale

La terza dashboard (Figura 4) mette in chiaro le informazioni temporali contenute nel nostro dataset.

Dunque il primo grafico posto in alto mostra l’andamento delle vendite settimanali in termini di unità vendute, corredato di anche informazioni circa le unità totali visualizzate nella singola settimana.

Abbiamo aggiunto anche una linea che corrisponde alla tendenza, che segue bene gli andamenti delle vendite una volta scelto l’andamento polinomiale di terzo grado, con un p-value di 0.1, superiore alla soglia minima (0.05) ma accettabile a cuor leggero nel momento in cui si nota che i dati di cui si ha a disposizione sono pochi (si parla di un mese e mezzo di dati di vendite) per permettere al software di tracciare una linea di tendenza che rispetti bene l’andamento dei dati.

Il grafico a sinistra invece illustra l'andamento delle vendite giorno per giorno: ad un boom iniziale (che corrisponde alla barra più alta dell'istogramma) segue invece un andamento relativamente stabile, che si assesta intorno alla media (linea tratteggiata nel grafico). Notiamo invece dei dati che fanno storcere il naso: in data 16/4 c'è un valore puntuale del Daily Sold di -376 unità.

A seguito di numerose ricerche, abbiamo trovato alcune informazioni circa un terremoto di portata molto importante occorso in data 10/4. Abbiamo avanzato l'ipotesi che quel valore negativo sia la somma delle perdite dei telefoni caduti in ogni negozio durante la scossa, ma rimaniamo aperti a discussioni future per approfondire i risultati.

Un altro grafico che presenta dei dati molto interessanti è sicuramente "Media prezzi nel tempo", che appunto mette in luce l'andamento giorno per giorno della media dei prezzi dei prodotti venduti nel paese.

Notiamo come segua un andamento relativamente stabile fino alla data 14/4, data a partire dalla quale il valore schizza verso l'alto, raddoppiando il prezzo medio (da 2.807.851 rupie del giorno precedente a 4.928.866 rupie).



Figura 5: Andamento temporale della valuta indonesiana (rupie)

A seguito di una ricerca, abbiamo visto come l'andamento della rupia indonesiana nel tempo (Figura 5) abbia subito una forte depressione a partire proprio dal 10/4 (data della scossa), concludendo come questa sia la motivazione alla base dell'aumento dei prezzi medi di vendita dei telefoni nel paese.

5 Previsioni

Nell'ultima dashboard (Figura 6) presentiamo la sezione relativa alle Previsioni, realizzate grazie al tool di Tableau.

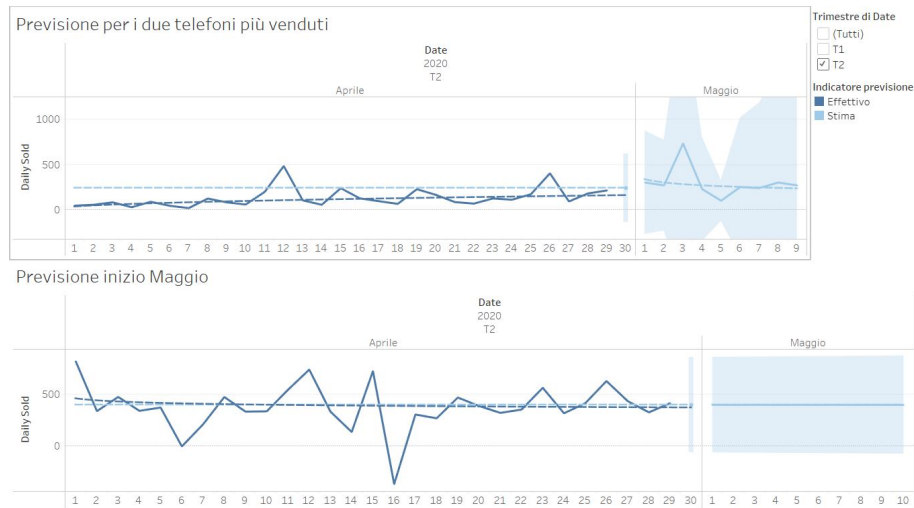


Figura 6: Previsioni delle vendite nel tempo

Il primo grafico, come suggerisce il titolo, mostra le previsioni per i due telefoni più venduti.

Abbiamo ottenuto un risultato di questo tipo andando innanzitutto a filtrare tutto il dataset concentrandoci solamente sui top 2 telefoni venduti, ovvero il Samsung Galaxy A10s e Xiaomi Redmi Note 8.

Come si può notare, per inizio Maggio è stato possibile estrarre una previsione dell'andamento delle vendite nel tempo, considerando solamente da Aprile in poi (infatti a destra si può notare come il filtro abbia permesso un focus particolare solamente sul secondo semestre T2).

Un risultato di questo tipo è stato possibile sicuramente grazie al fatto che è possibile identificare una certa regolarità dei dati, il che ha permesso a Tableau di azzardare delle previsioni (seppur prendendosi degli intervalli di confidenza importanti).

Per quanto riguarda invece l'andamento delle vendite per tutti i telefoni, purtroppo i risultati sono molto diversi: in questo caso l'andamento è decisamente più irregolare, il che, sommato ad una quantità limitata di dati ed un intervallo temporale di solo un mese e mezzo, non dà la facoltà a Tableau di fare una previsione, e pertanto si limita a stabilire la media in futuro, pari a 398 unità giornaliere vendute.

Come detto, le ragioni dietro un risultato di questo tipo sono molteplici, e pensiamo che siano principalmente le seguenti:

- forte irregolarità dei dati;
- quantità limitata di dati a nostra disposizione;
- intervallo temporale insufficiente per fare previsioni.

Chiaramente siamo aperti a discussioni, ma confidiamo nell'idea che se il dataset fosse più ricco di informazioni sarebbe sicuramente possibile estrarre delle previsioni soddisfacenti.

In ogni caso, oltre alle previsioni, abbiamo inserito anche delle informazioni riguardanti i trend. Focalizzandoci sui primi due telefoni, il p-value ottenuto con una trend line che segue una power law è di 0.002, rendendo il risultato significativo dal punto di vista statistico, ma poco accurato perché il coefficiente di determinazione si assesta a solo 0.29, un valore purtroppo basso, causato probabilmente dalle stesse motivazioni di cui sopra.

Purtroppo i risultati invece allargando l'orizzonte su tutti i telefoni non sono tra i più rosei: anche qui abbiamo un basso coefficiente di determinazione (0.02), accompagnato da un p-value ben oltre la soglia (0.48). Ciò ci permette di concludere che le analisi dal punto di vista temporale sono da escludersi per questo dataset, in base a come ci è stato reso disponibile dalla fonte, ma può essere sicuramente perfezionato nel caso in cui i dati fossero in futuro arricchiti.