

Prima esplorazione

1. Introduzione

Questo documento spiega la prima esplorazione che si è voluta fare nei dati ormai puliti.

L'idea è che si possano acquisire insight e grafici interessanti anche senza una modellazione particolare. Sono stati creati delle nuove colonne unendo informazioni di dati diversi, quindi applicando quello che viene definito “Ingegnerizzazione delle caratteristiche”.

Nel mio caso specifico ho voluto creare tre grafici che ho ritenuto essere potenzialmente interessanti, due di questi parametrici attraverso delle semplici funzioni.

I grafici sono i seguenti:

- Numero di vendite di un'oggetto specifico in un mese specifico: `QtaObjMonth(df, obj, month);`
- I top 10 prodotti acquistati da un utente nello storico totale: `QtaObjMonth(df, user);`
- La provenienza dei clienti, in questo specifico caso, si è scelto di mostrare i primi 5 per frequenza in quanto più significativi. Molti stati avevano da 1 a massimo 5 clienti mentre paesi come UK valori poco sotto i 4000 clienti.

2. Numero vendite in uno specifico mese

La funzione non ritorna nulla, crea semplicemente un grafico. Sull'asse delle x troviamo la data scelta attraverso la variabile **month** della funzione, mentre sull'asse delle y la quantità degli oggetti (**obj**) venduti lungo il periodo.

Ho voluto riportare due esempi del grafico creato dalla funzione per mostrare come cambia il grafico con diversi oggetti.

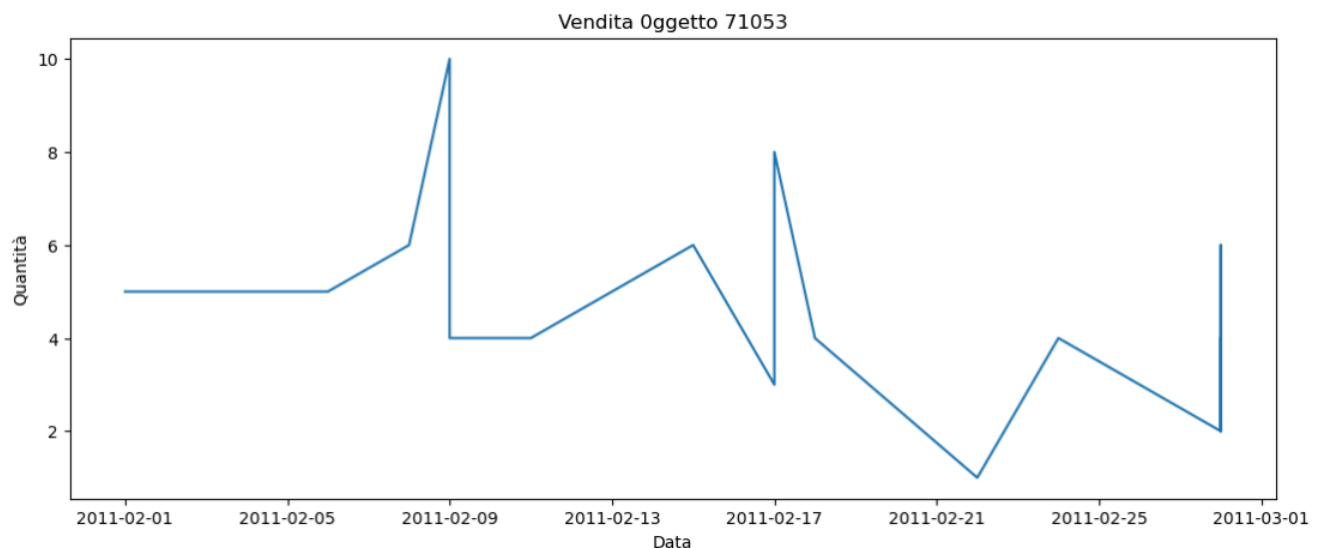


Figure 1: Vendite oggetto 71053 a febbraio.

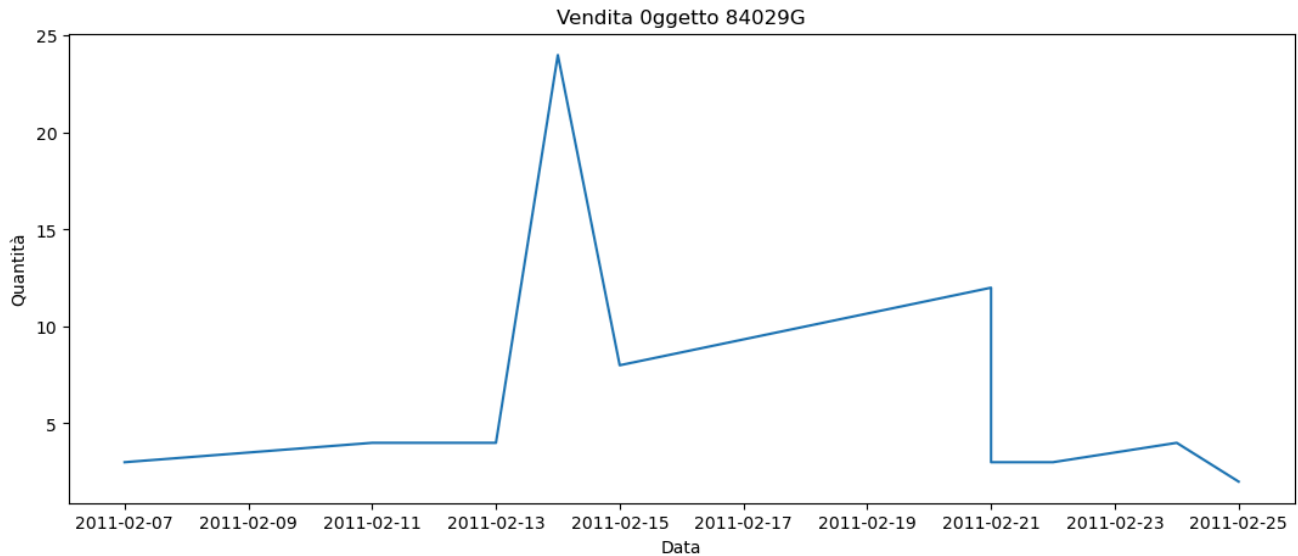


Figure 2: Vendite oggetto 84029G a febbraio.

3. Top 10 prodotti per cliente

Anche in questo caso la funzione è void, ovvero non ritorna nulla. Sull'asse delle x troviamo il prodotto la quantità invece si trova sull'asse delle y venduti dall'utente scelto (**user**) durante tutto lo storico del dataset.

In questo modo è facile capire l'interesse di un utente per diverse categorie di prodotto oppure i suoi interessi.

Anche in questo caso ho riportato due grafici di utenti diversi per mostrare come variano tra di loro.

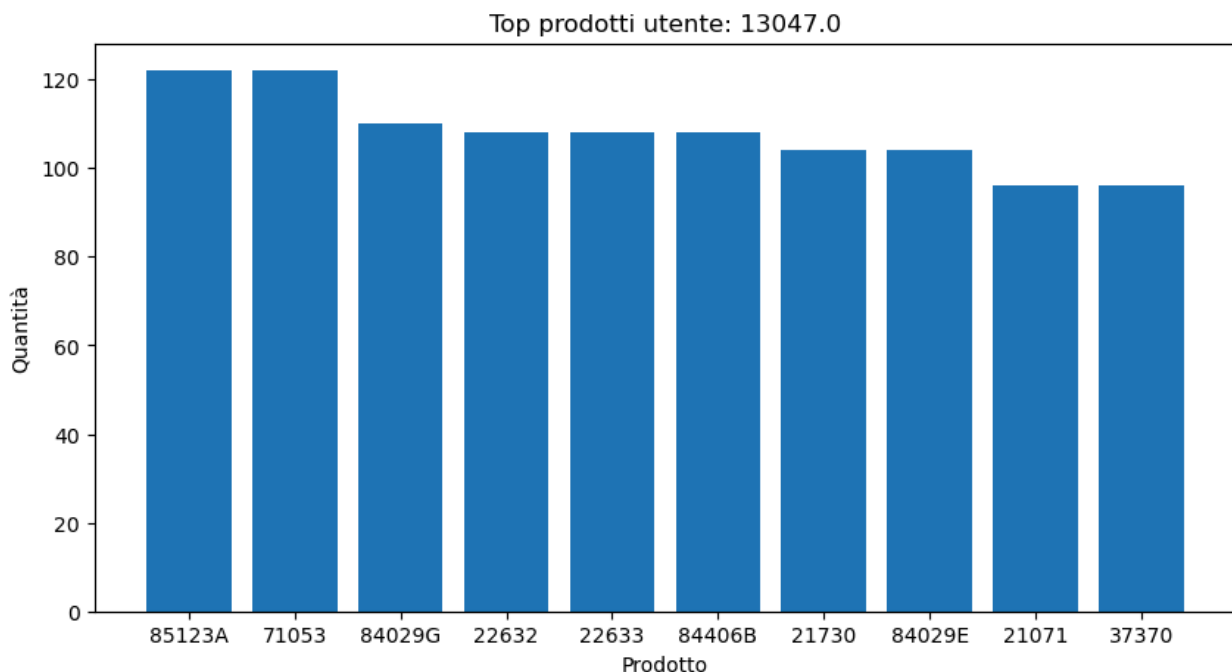


Figure 3: Top prodotti utente 13047.

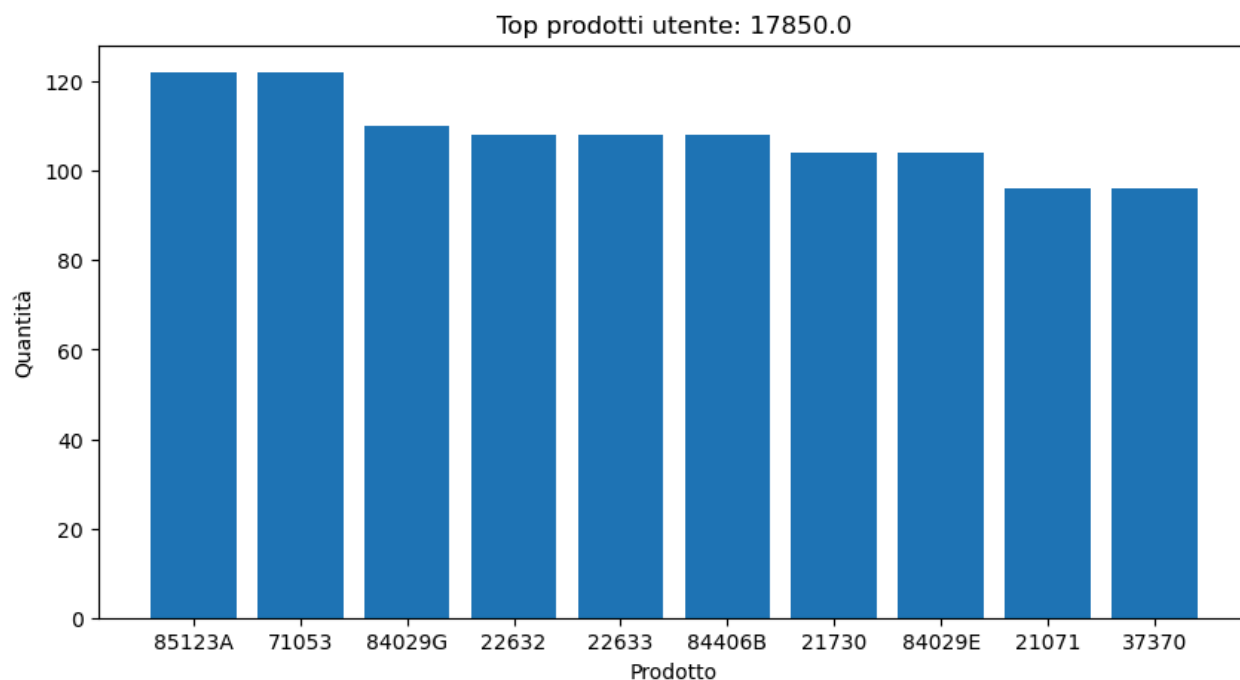


Figure 4: Top prodotti utente 17850.

4. Provenienza dei clienti

Questo grafico non è una funzione bensì un plot a se stante che ho voluto fare perchè ritengo possa essere utile come analisi di mercato.

In questo caso il grafico risulta essere poco significativo per due distinti motivi:

- Il grafico è stato ricavato da un dataset ottenuto online che rappresenta un business di nostro interesse, ma più una prova di fattibilità e di “template”;
- La maggioranza dei clienti sono residenti in UK, il che rende molto piccoli le altre quattro nazioni con maggior numero di acquirenti, rendendole quasi non significative sul totale.

Sottostante ho riportato il grafico della clientela per paese.

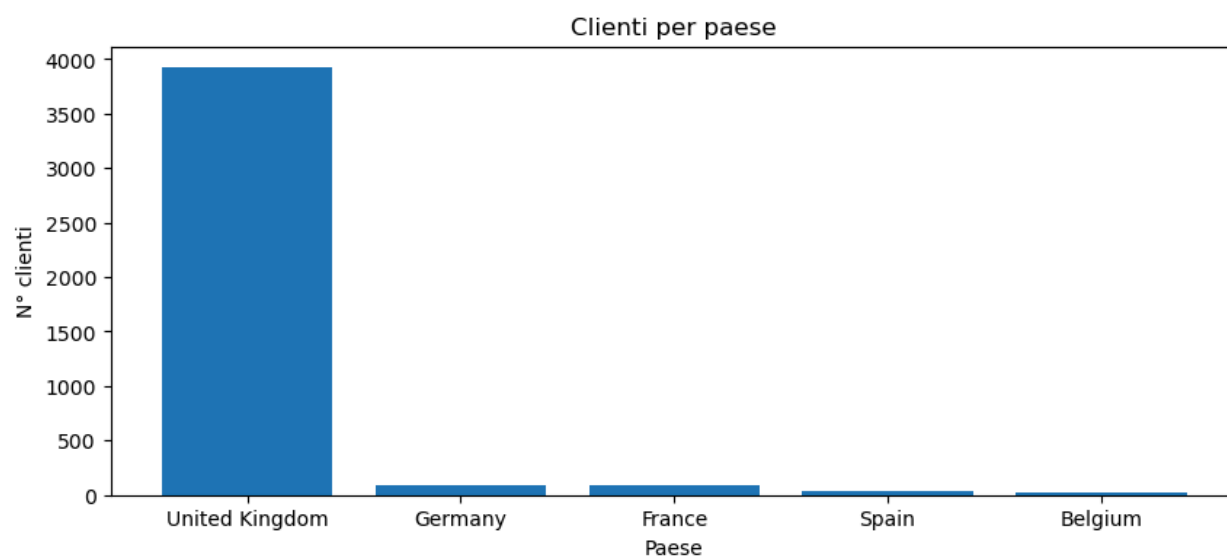


Figure 5: Cliente per paese.