

Case study - Esprinet S.p.A

Fumagalli Marco

Agosto 2023

1 Introduzione

Il presente documento è volto alla presentazione di un breve case study. Il candidato dovrà condividere il proprio lavoro attraverso slides in una presentazione della durata di 15/20 minuti. Il tempo stimato per il completamento del case study è di circa 3/4 ore.

Il dataset, in formato csv, verrà inviato tramite email.

2 Dati

Il dataset in oggetto rappresenta dati relative alle quotazioni richieste da clienti esprinet. La quotazione è una richiesta di offerta da parte del cliente.

Il cliente inserisce i prodotti desiderati all'interno del carrello e può richiedere ad Esprinet una quotazione, ovvero uno sconto per ciascun prodotto presente nel carrello.

Di seguito una spiegazione delle variabili e del loro significato:

- DEALKEY: identificativo quotazione testata (string)
- DEALDETKY: identificativo riga quotazione. Una quotazione può avere più righe, ciascuna riferita ad un prodotto. (string)
- CUSTOMERID: id cliente (string)
- CUSTOMER CREATE DATE: data di prima registrazione cliente (date)
- CHANNELID: id canale vendita (string)
- CUSTOMER_DISCOUNT_PERS: sconto personalizzato cliente (float)
- HAS_REAL_ORDER (target variable): 1/0, quotazione trasformata/non trasformata in ordine (bool)
- DEALVALUE: valore riga quotazione, ORDER_PRICE moltiplicato per DEALQTY (float)
- ORIGINAL_PRICE: prezzo articolo (float)

- ORIGINAL_PRICE_NO_DISCOUNT_PERS: prezzo articolo senza sconto personalizzato (float)
- ORDER_PRICE: prezzo articolo scontato da quotazione e da sconto personalizzato (DISCOUNT_PCT, CUSTOMER_DISCOUNT_PERS) (float)
- DEALQTY: numero pezzi. (int)
- QUOTATION_DATE: data quotazione (date)
- DISCOUNT_PCT: sconto quotazione (float)
- ARTID: identificativo articolo (string)
- ARTID_DATE_CREATE: data prima registrazione articolo (date)
- ARTDSC: descrizione articolo (string)
- BRANDID: brand articolo (string)
- CATEGORYID: categoria merceologica. (string)
- CATEGORYDSC: descrizione categ merceologica (string)
- BRANDDSC: descrizione brand (string)

CUSTOMER_DISCOUNT_PERS: Alcuni clienti in Esprinet possono avere uno sconto dedicato su determinati brand.

3 Domande

L'obiettivo del case study è quello di analizzare il dataset in oggetto, in modo da estrarre degli insight significativi che possono esser comunicati e condivisi. Prova inoltre ad esplicitare un modello per la previsione dell'accettazione o rifiuto della quotazione (HAS_REAL_ORDER).

Sei libero di fare delle assunzione sui dati in oggetto e di utilizzare qualsiasi tecnica per l'analisi.

Il case study deve essere strutturato seguendo le fasi tipiche di un processo di data science, come l'acquisizione dei dati, la comprensione dei dati, l'elaborazione dei dati, la modellizzazione e la valutazione dei modelli.

Si richiede di utilizzare python per il completamento del case study.

Verrà successivamente richiesto di condividere il codice utilizzato per analizzare i dati in oggetto.

Nel caso di problemi nell'importazione dei dati il comando

```
import pandas as pd
pd.read_csv('_case_study_anonymized.csv', sep='|', encoding="latin")
```

dovrebbe dare esito positivo e caricamento in memoria dei dati.