

Introduction

This document is intended for the presentation of a short case study. The candidate is required to share their work through slides in a presentation lasting 15 to 20 minutes. The estimated time for completing the case study is approximately 3 to 4 hours. The dataset, in CSV format, will be sent via email.

2. Data

The dataset in question represents data related to quotations requested by Esprinet customers. A quotation is a request for an offer from the customer. The customer adds desired products to their cart and can request a quotation, which is essentially a discount for each product in the cart.

Here's an explanation of the variables and their meanings:

- **DEALKEY**: Quotation header identifier (string)
- **DEALDETKEY**: Quotation line identifier. A quotation can have multiple lines, each referring to a product. (string)
- **CUSTOMERID**: Customer ID (string)
- **CUSTOMER CREATE DATE**: Date of the customer's first registration (date)
- **CHANNELID**: Sales channel ID (string)
- **CUSTOMER DISCOUNT PERS**: Customized customer discount (float)
- **HAS REAL ORDER (target variable)**: 1/0, transformed/untransformed quotation into an order (boolean)
- **DEALVALUE**: Quotation line value, ORDER PRICE multiplied by DEALQTY (float)
- **ORIGINAL PRICE**: Article price (float)
- **ORIGINAL PRICE NO DISCOUNT PERS**: Article price without personalized discount (float)
- **ORDER PRICE**: Article price after quotation and personalized discount (DISCOUNT PCT, CUSTOMER DISCOUNT PERS) (float)
- **DEALQTY**: Number of pieces (integer)
- **QUOTATION DATE**: Quotation date (date)
- **DISCOUNT PCT**: Quotation discount (float)
- **ARTID**: Article identifier (string)
- **ARTID DATE CREATE**: Article's first registration date (date)
- **ARTDSC**: Article description (string)
- **BRANDID**: Article brand (string)
- **CATEGORYID**: Merchandise category (string)
- **CATEGORYDSC**: Merchandise category description (string)
- **BRANDDSC**: Brand description (string)

CUSTOMER DISCOUNT PERS: Some Esprinet customers may have a dedicated discount on certain brands.

3. Questions

The objective of the case study is to analyze the dataset in order to extract meaningful insights that can be communicated and shared. Additionally, attempt to develop a model for predicting the acceptance or rejection of a quotation (HAS REAL ORDER).

You are free to make assumptions about the data and use any techniques for analysis. The case study should be structured following typical data science processes, including data acquisition, data understanding, data processing, modeling, and model evaluation.

Python should be used for completing the case study. You will be requested to share the code used for analyzing the data.

In case of any issues with data import, the following command should yield a positive outcome and load the data into memory:

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