FRESCO RETAIL PROJECT

PROBLEM STATEMENT, VARIABLE DESCRIPTION & DELIVERABLES

Fresco Retail Project – Objective & Deliverables

Problem description: The data analyses Fresco Retail's customers' transaction data to predict return decision using various data points like ***customer background, payment modes, store types, product nature***, etc. The objective is to study the influence of different transaction data points and their impact on a customer's final decision to return the product(s) bought in a transaction.

Recommended Project Steps & Guidelines:

1. **Understand the data variables properly**. Refer to the page 2 of this document for the variable description.
2. **Clean the data:** Clean the data, that is, fill the missing values (if any), treat the outliers (or odd values), etc. Ensure each variable's data is as per the nature of the variable (eg – Date field should contain only date values, numeric column should be formatted as numeric, etc.).
3. **Conduct EDA (Exploratory Data Analysis) on the cleaned Data:** Summarize, explore the data and then decide your strategy. Make note of any important assumptions that you make.
4. **Uni-variate and Bi-variate Analysis:** Check the distribution of independent variables and also compare them with the dependent variable. **(Bonus Points)**
5. **Feature Engineering:** Create new meaningful features based on the existing features by applying some aggregation functions on them. **(Bonus Points)**
6. **Identify the most important variables (or data parameters) that affect the final decision:** Identify the impact of each variable on the final result graphically (correlation / scatter plots, regression plots, etc.). Keep those variables that affect the final outcome. **(Bonus Points)**
7. **Develop and Validate Samples:** Divide samples into 2 parts: Development Sample (70%) & Validation Sample (30%). Build your analysis model using the Development Sample, and validate it on the validation sample and then predict on test sample.
8. **Model Building:** Analyze the dependent variable and decide which technique out of regression or classification to use and hence build the model.
9. **Improving model accuracy:** Perform various iterations by eliminating or adding the variables to see if the model accuracy is improving or not. Also, you can apply various transformation like log transformation on dependent variable or independent variables or both to improve accuracy. **(Bonus Points)**
10. **Model Comparison:** Comparing the chosen model with other similar models that could have been used in this project **(Bonus Points)**
11. **Efficient coding:** Use parameterized functions to do certain tasks. **(Bonus Points)**

Project Deliverables:

1. **Explain your approach and steps with comments** – In form of comments in your Python notebook, document your approach, steps, model outputs, and findings. It should at least include the Data Cleaning steps & why, Missing Data Treatment, Statistical Model used & why (added bonus for comparing the chosen model with other similar models that could have been used in this project), Model Tests performed / Results, Model Validation Results, Statistical Conclusion and Business Recommendations (if any).
2. **Submit**:- Your Python notebook with comments and outputs in Excel (if any).

Variable Description:

transaction_id: Unique transaction number for each sale

Month_code: The month of transaction

prod_cat_code: Product category code. A product may have multiple sub-categories

prod_cat: Product category name mapped to prod_cat_code.

prod_subcat_code: Product sub-category code

prod_subcat: Product sub-category name mapped to product_subcat_code

Qty: Quantity of products bought / returned

Rate: Price per unit of a product in local currency

Amount: Total amount (Qty * Rate) without the taxes

Tax: Tax amount by local government

Delivery_chrgs: Delivery charges to deliver the products ordered

Payment_mode: Mode of payment used by the customer to pay for a transaction

Store_type: Type of store where the transaction took place. eShop means online commerce. MBR means Multi branded retail store. Flagship store means standalone fully owned shop by the retailer. Teleshop means orders received on retailer phone numbers.

Reviews: Review left by the customer on the product or transaction

Cust_id: Customer identification number

DOB: Customer's date of birth

Gender: Customer's gender

Education_Level_Code: Code assigned to different education levels of a customer

Level_Education: Education level mapped to Education_Level_Code

Profession_Code: Customer's profession code

Profession Type: Type of profession (Salaried, Self Employed, Others)

City_code: City identification code assigned to each city

RETURN: This is our **target variable** that stores the response of whether the transaction was finally returned by the customer.