E-Commerce & Retail B2B Case Study

- Harshal Naik
- Mehul Mehtani
- Kamesvarer S

Problem Statement

Analyse the customer transactions data to find different payment behaviours.

Segregate the customers based on their previous payment patterns

Based on the historical data, predict the likelihood of delayed payment against open invoices from the customers.
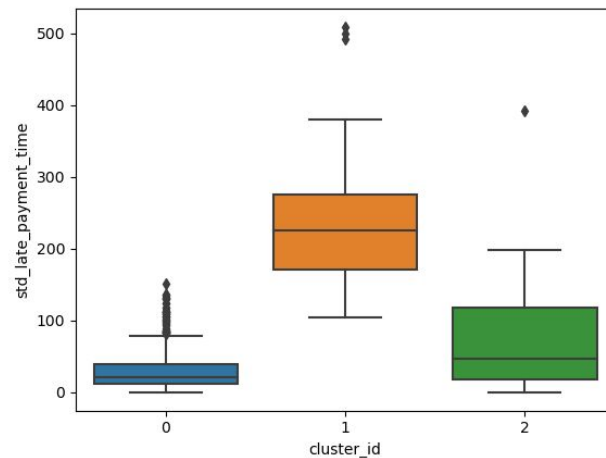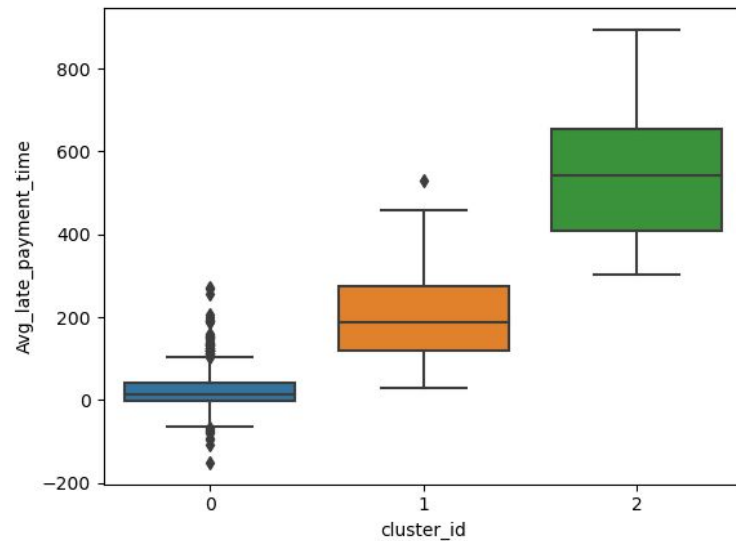
Draw Business insights .

Available
Datasets

- Received_Payments_Data.csv
  - Past transactions data

- Data+Dictionary.xlsx
  - Description of Dataset Columns

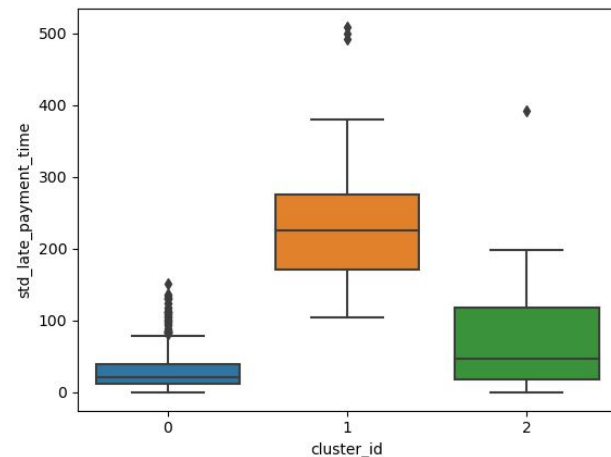- Open_Invoice_data.csv
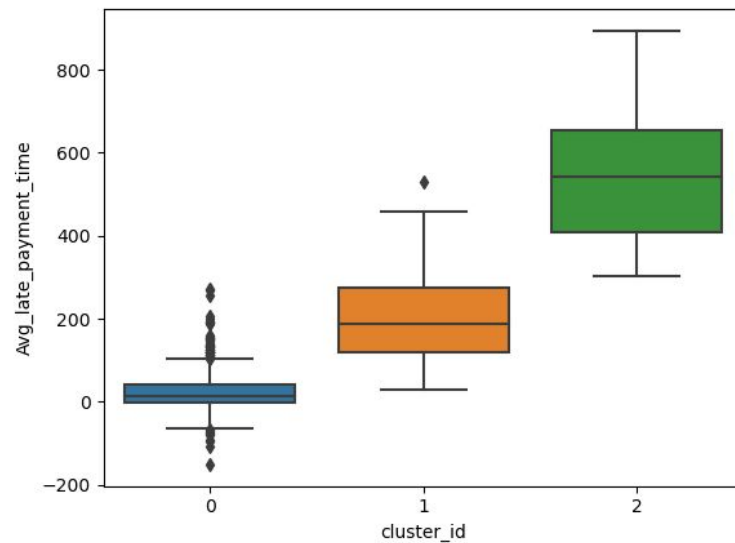 - Present/future Transactions data

- Cluster 0 :
  - regularly paying on time

- Cluster 1 :
  - Intermittently paying late

- Cluster 2 :
  - Regularly paying late

Clustering

- Cluster 0 :
  - Prime Customers

- Cluster 1 :
  - General Customers

- Cluster 2 :
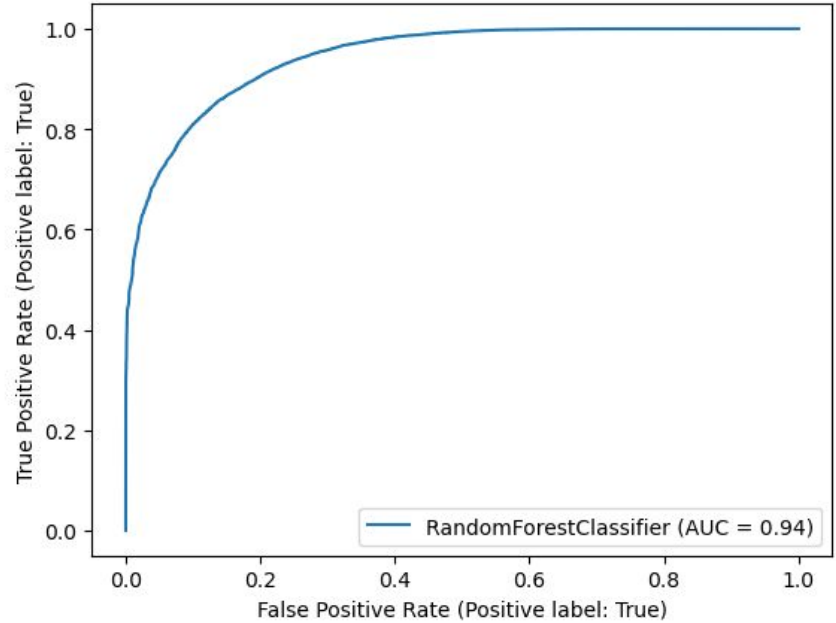 - Problematic Customers

Clustering

The Random Forest model was built using following parameters
- Payment_term_days
- USD Amount

cross-validated accuracy : 82 %

Accuracy of model = ~ 84%
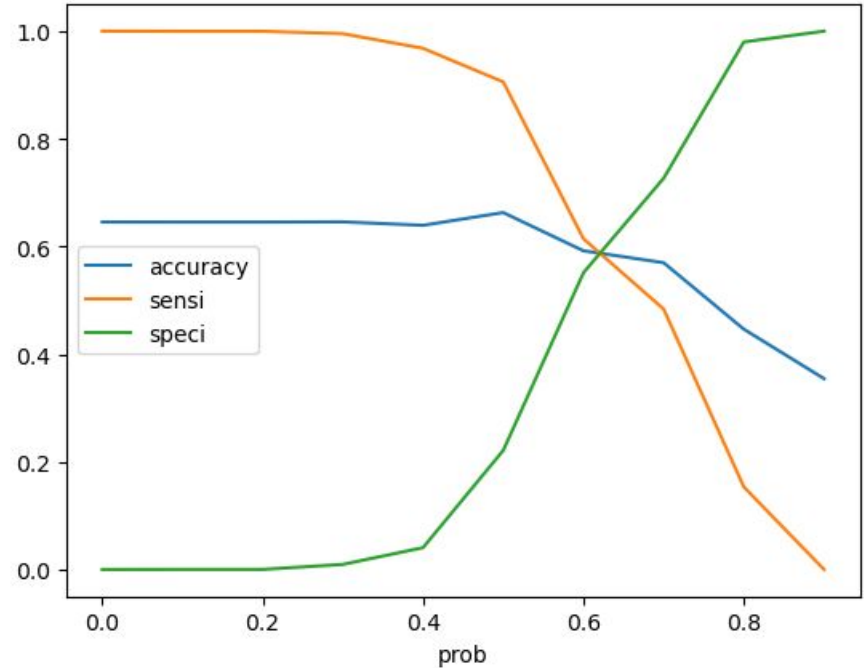
Random forest model

The Logistic regression model was built using following parameters
- Payment_term_days
- USD Amount

Training Accuracy : 66 %

Test Accuracy = ~ 65%
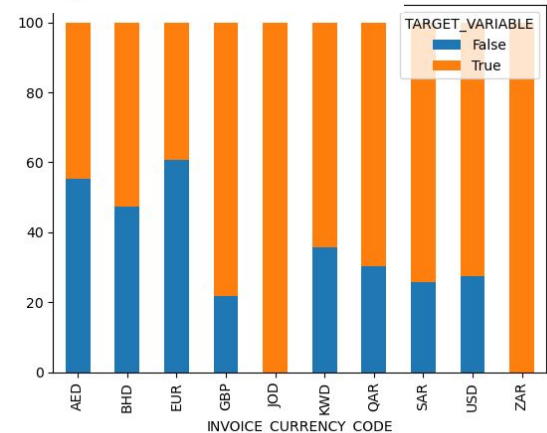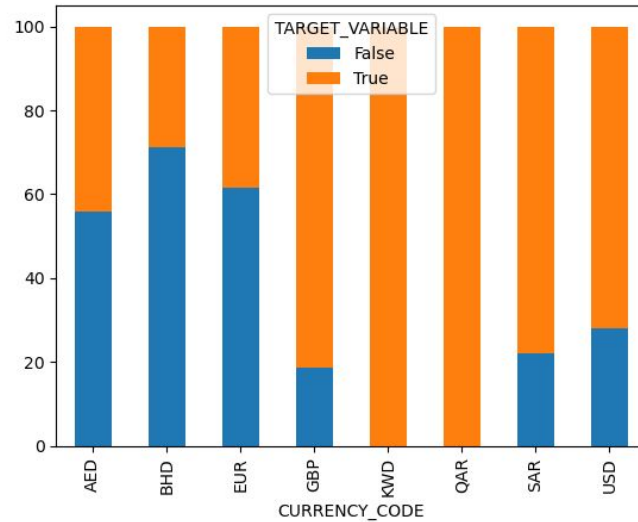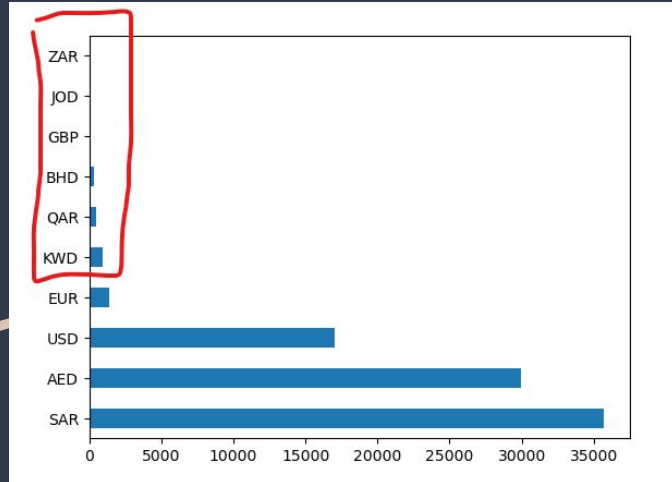
Logistic Regression Model

CHOSEN CUTOFF : 0.5

Insights

Less common currencies are associated with higher late payment rates.

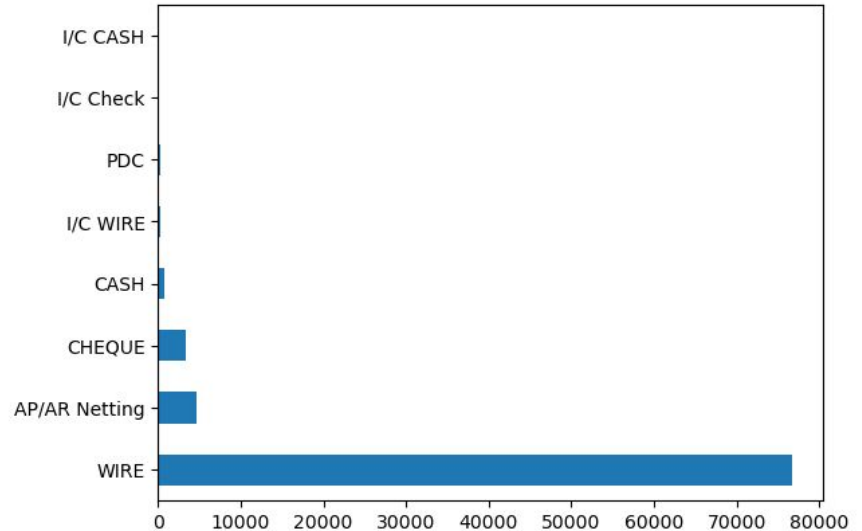Action : investigate whether there is issue related to currency in transaction process

True : Late Payment

Insights

Most of the customer chooses Wire transfer as mode of payment .

Action : make wire transfer mode more feasible/faster , if possible.

Insights

In 2761 out of 85915 cases, the payment was done before the invoice was generated.

Action : Investigate the process of Invoice generation.

#Insights

Using Random Forest Model that we've built in jupyter notebook , We will predict the customers who are more likely to make late payment.

We will ensure timely payment follow-up for these customers to facilitate the recovery of cash flow.

Thank You.!