

E-Commerce and Retail B2B Case Study

Submitted by



Atul Mani

Harish Prabhakar

Nguyen Thi Thu Huong



Problem Statement

- Schuster is a multinational retail company dealing in sports goods and accessories.
- Schuster conducts significant business with hundreds of its vendors
- It has credit arrangements.
- Unfortunately, not all vendors respect credit terms and some of them tend to make payments late.
- Schuster would thus try to understand its customers' payment behaviour and predict the likelihood of late payments against open invoices.



Goal

- Schuster would like to better understand the customers' payment behaviour based on their past payments.
- Using historical information, it wants to be able to predict the likelihood of delayed payment against open invoices.
- It wants to use this information so that collectors can prioritize their work in following up with customers beforehand to get the payments on time.

A blue pen with a silver tip is resting on a document that features a bar chart with blue bars of varying heights. The pen is positioned diagonally across the frame, pointing towards the bottom right.

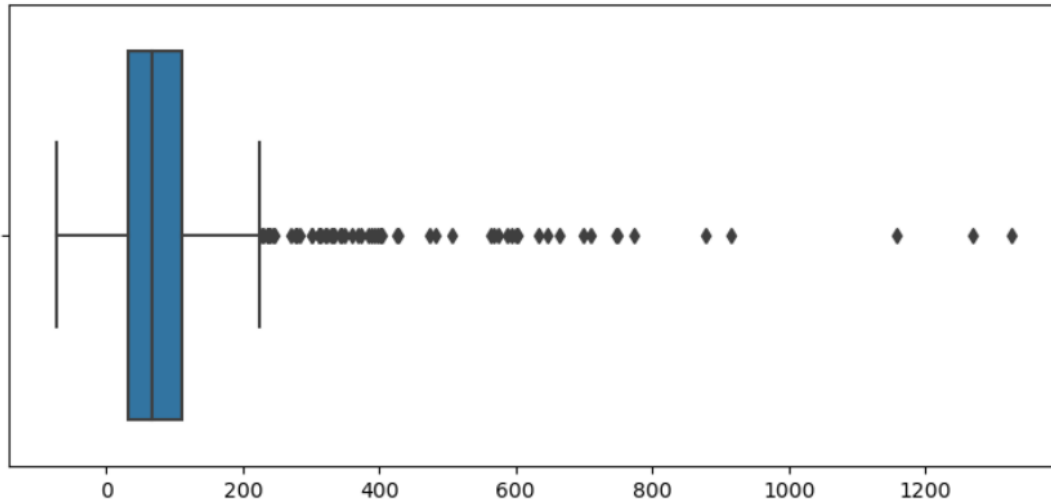
Solution Approach

- For this case study we're going to use logistic regression model to predict whether the customer is going to delay the payment or not.
- The steps involved for this case study are mentioned below:
 - Data Loading Data Exploration and pre-preprocessing which will include following steps:
 - a. Missing value handling
 - b. Outlier handling
 - c. Feature engineering
 - Model Building using statsmodel and VIF
 - Model Performance Evaluation using Probability Calibration, ROC Curve, Precision-Recall Curve
 - Final recommendation.

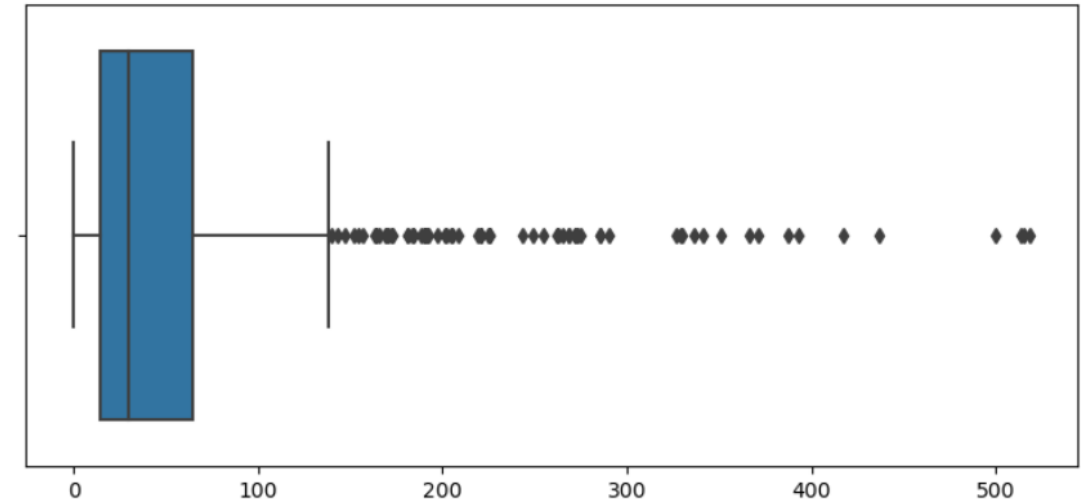
Customer Segmentation

- Based on the payment patters the customers have been segmented into 3 categories.
- The metric to segment is the mean time in payment and the standard deviation in payment time.

PAYMENT_TIME_MEAN

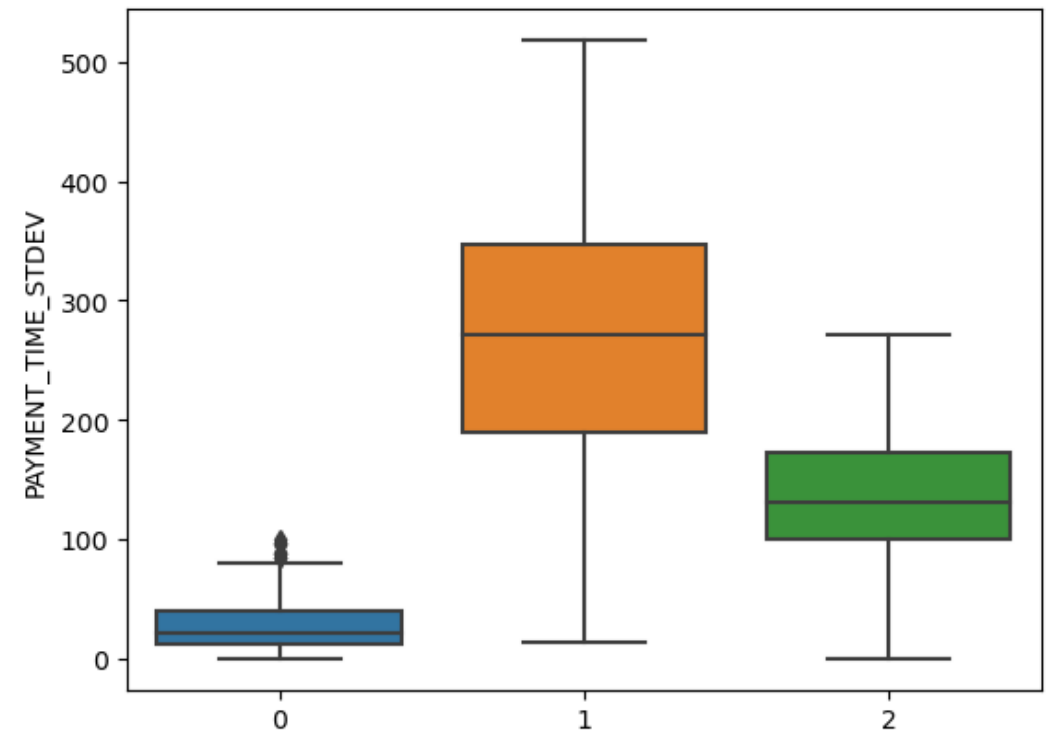
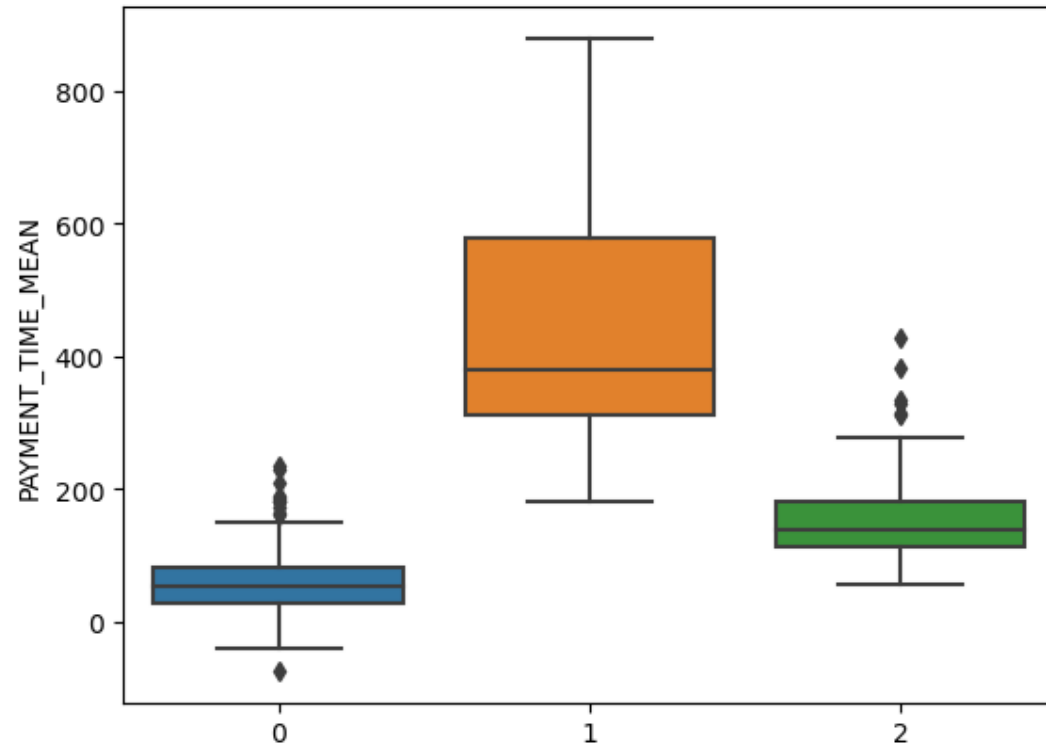


PAYMENT_TIME_STDEV



Customer Segmentation

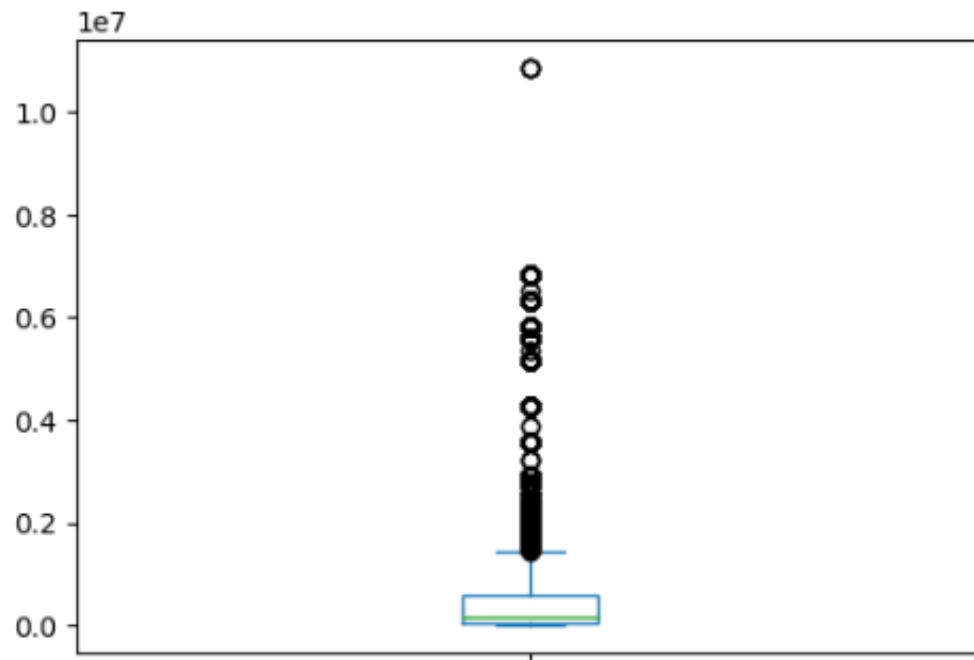
- Based on the payment patters the customers have been segmented into 3 categories.
- The metric to segment is the mean time in payment and the standard deviation in payment time.



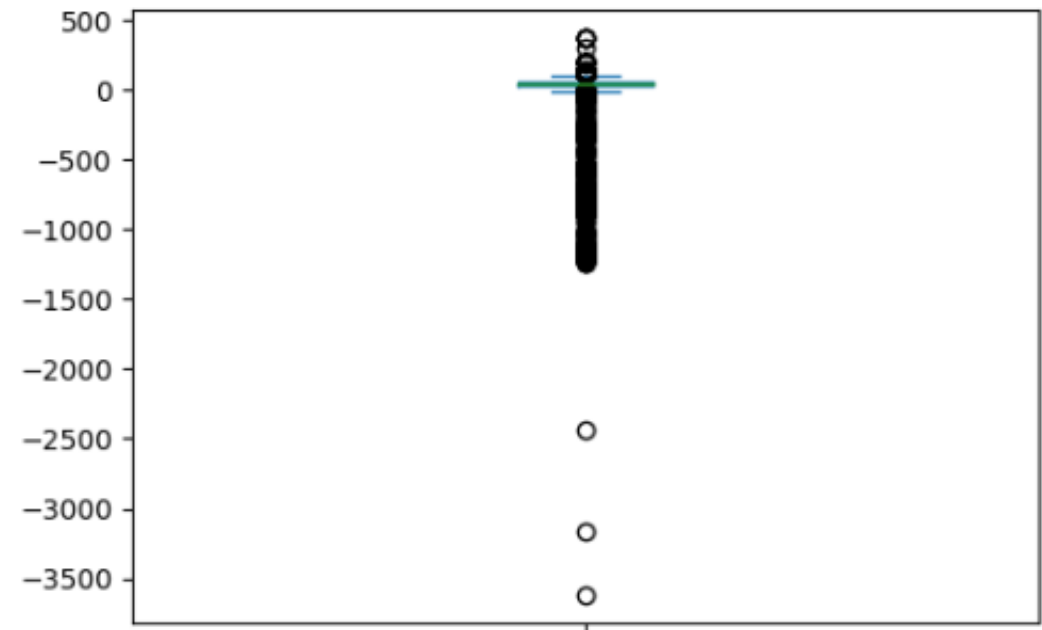
The Numeric Data

- The numeric data used in analysis are:
 - USD Amount
 - Payment_Term

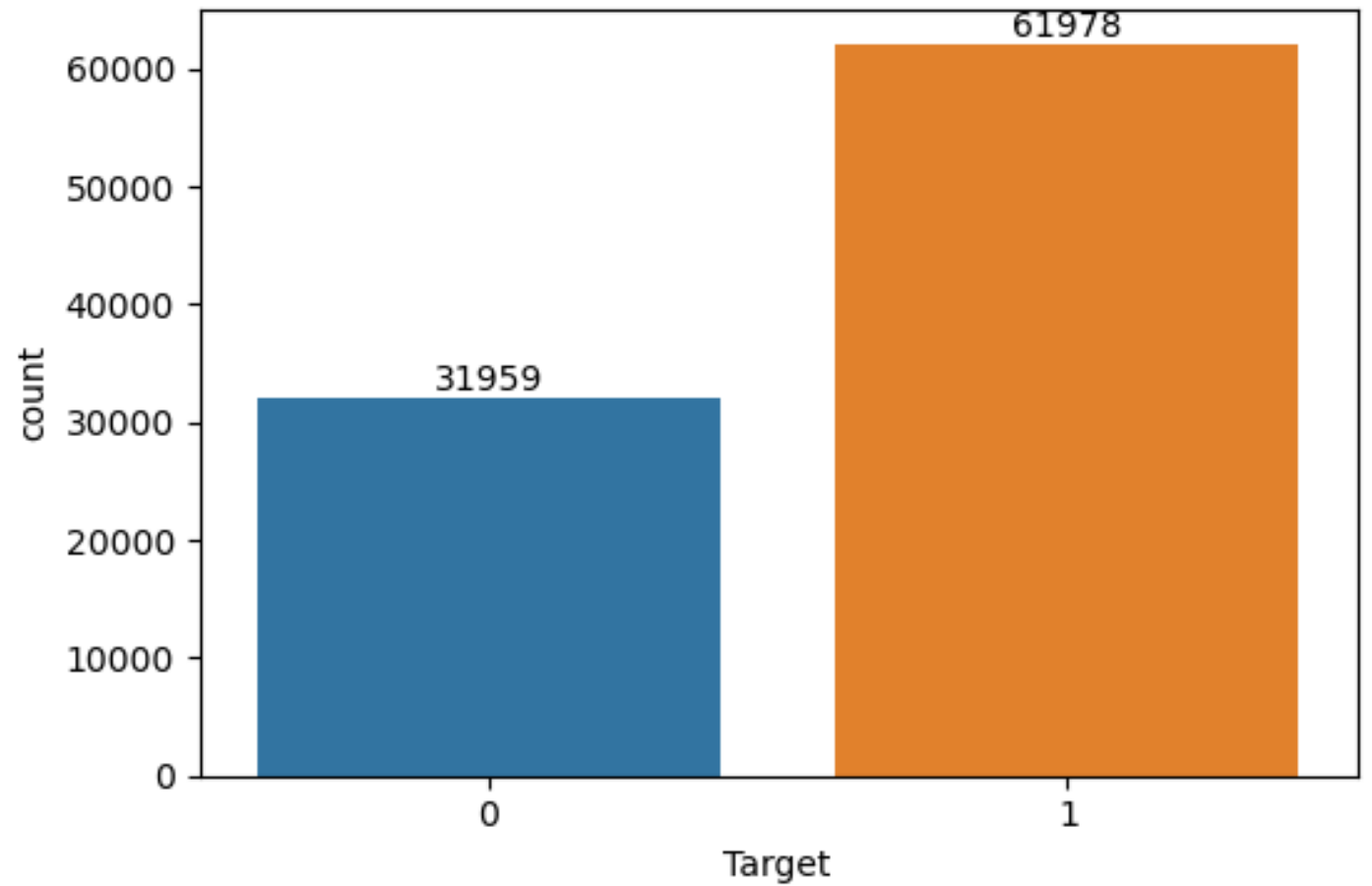
USD Amount



PAYMENT_TERM

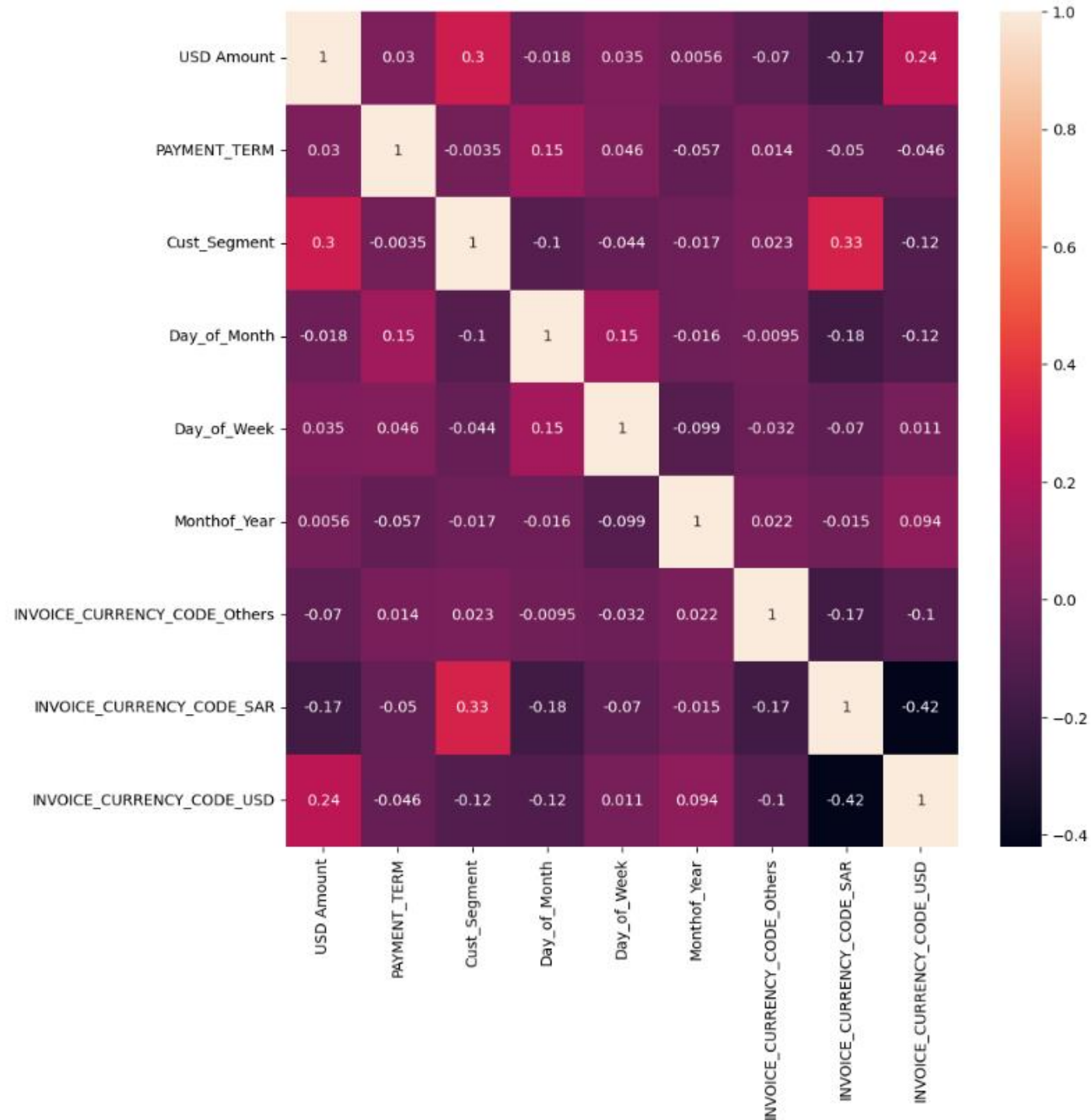


Data
Imbalance
for Target
Column

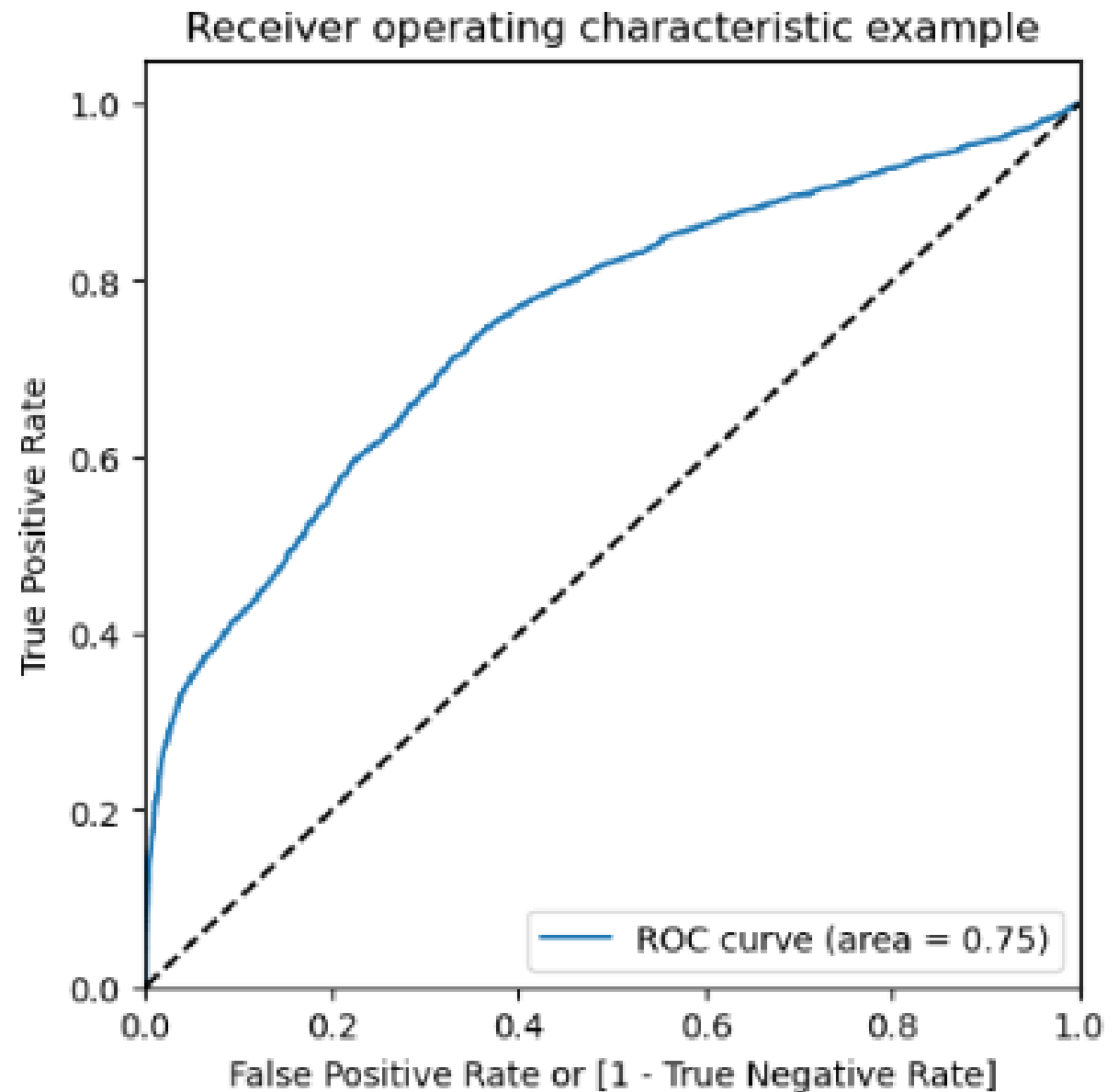


Correlation Matrix

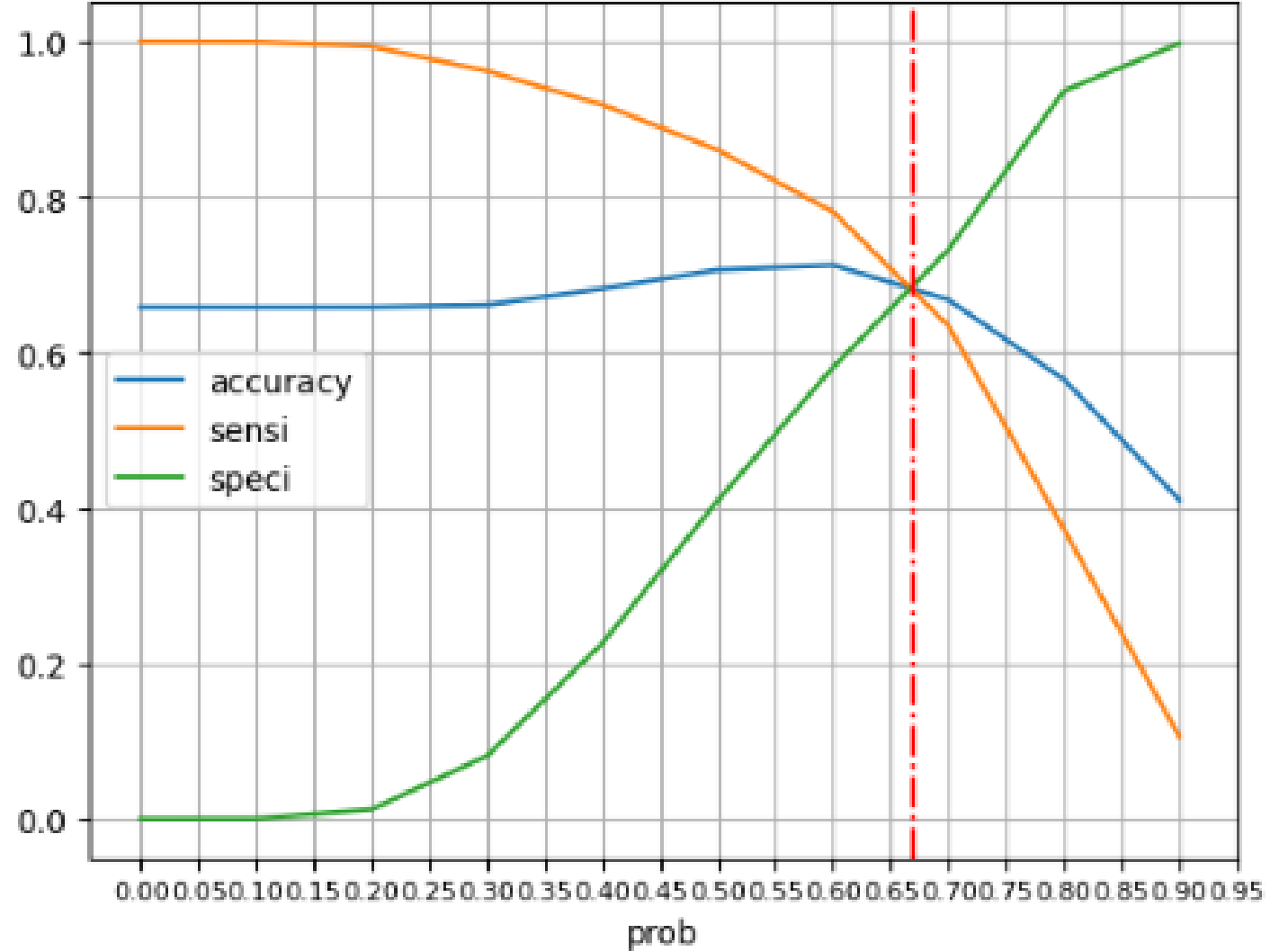
There are not much Correlation in variables selected



The ROC Curve



Finding Optimal Cut-off (as 0.67)



Recommendation



There were 88204 Open invoices.



Out of 88204 for 28080 invoices the payment is due in future.



Out of 28080 open invoices, 15440 are likely to delay their payment.



The Schuster should chase these invoice's customers before the due dates.

Thank you

Atul Mani
Harish Prabhakar
Nguyen Thi Thu Huong