Schuster Retail B2B

Case Study

By

Ankit Garg

Ankur Kushwaha

Understanding the Problem Statement

- Schuster, a multinational retail corporation specializing in sports goods and accessories, maintains numerous credit arrangements with its vendors. However, certain vendors consistently delay payments, leading to significant financial losses for both parties. Despite imposing heavy late payment fees, this practice undermines long-term business relationships.

- To mitigate these challenges, Schuster seeks to analyze vendor payment patterns and forecast the probability of late payments for outstanding invoices. This proactive approach aims to reduce non-value-added activities, save time, and minimize financial repercussions.

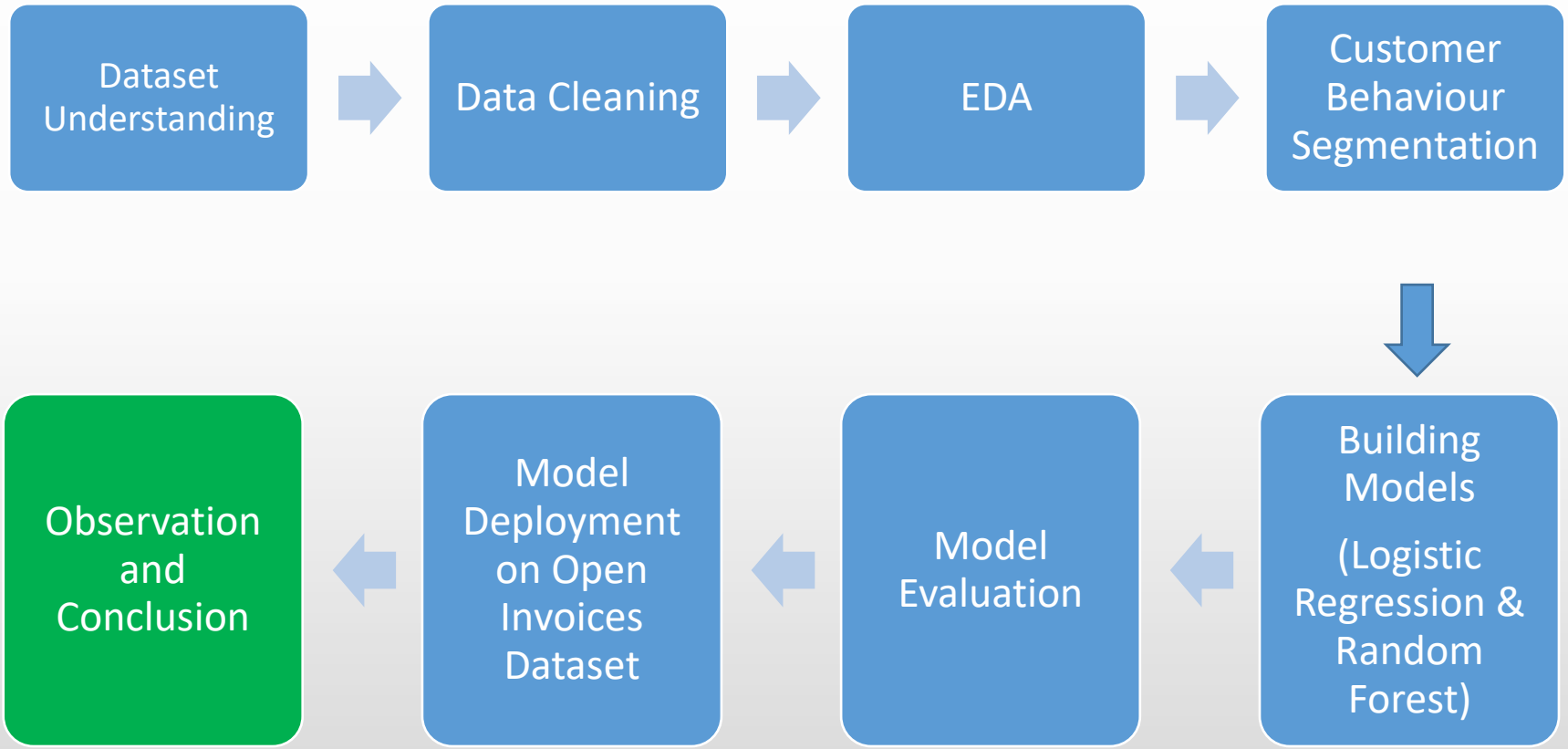
Business Objective

Schuster's business objective is to analyze past payment patterns to understand customer behavior and predict the likelihood of delayed payments for open invoices. This predictive capability allows collectors to prioritize follow-up efforts and ensure timely payments.

Business Goals

- Schuster aims to understand customer payment behavior through analysis of past payment patterns.
- The company seeks to segment customers based on their historical payment data.
- Using historical information, Schuster intends to predict the likelihood of delayed payments for open invoices from its customers.
- This predictive capability enables collectors to prioritize their efforts in following up with customers to ensure timely payments.

Approach Adopted



Data Understanding

We are given 2 datasets :

1. 'Received_Payment_Data.csv'
    - ❖ Contains the information of all the transactions that have been performed with various vendors in the past.
    - ❖ It has 93937 rows and 16 columns

2. ‘Open_Invoice_Data.csv’
    - ❖ This data contains the information of all the invoices that are open, i.e. that haven't been paid yet.
    - ❖ It has 88204 rows and 14 columns

Data Cleaning

Missing Data Identification:

- It is learnt that in the Received_Payment dataset, 54% records were duplicate.
- RECEIPT_DOC_NO contained 29 null values.

Action Taken:

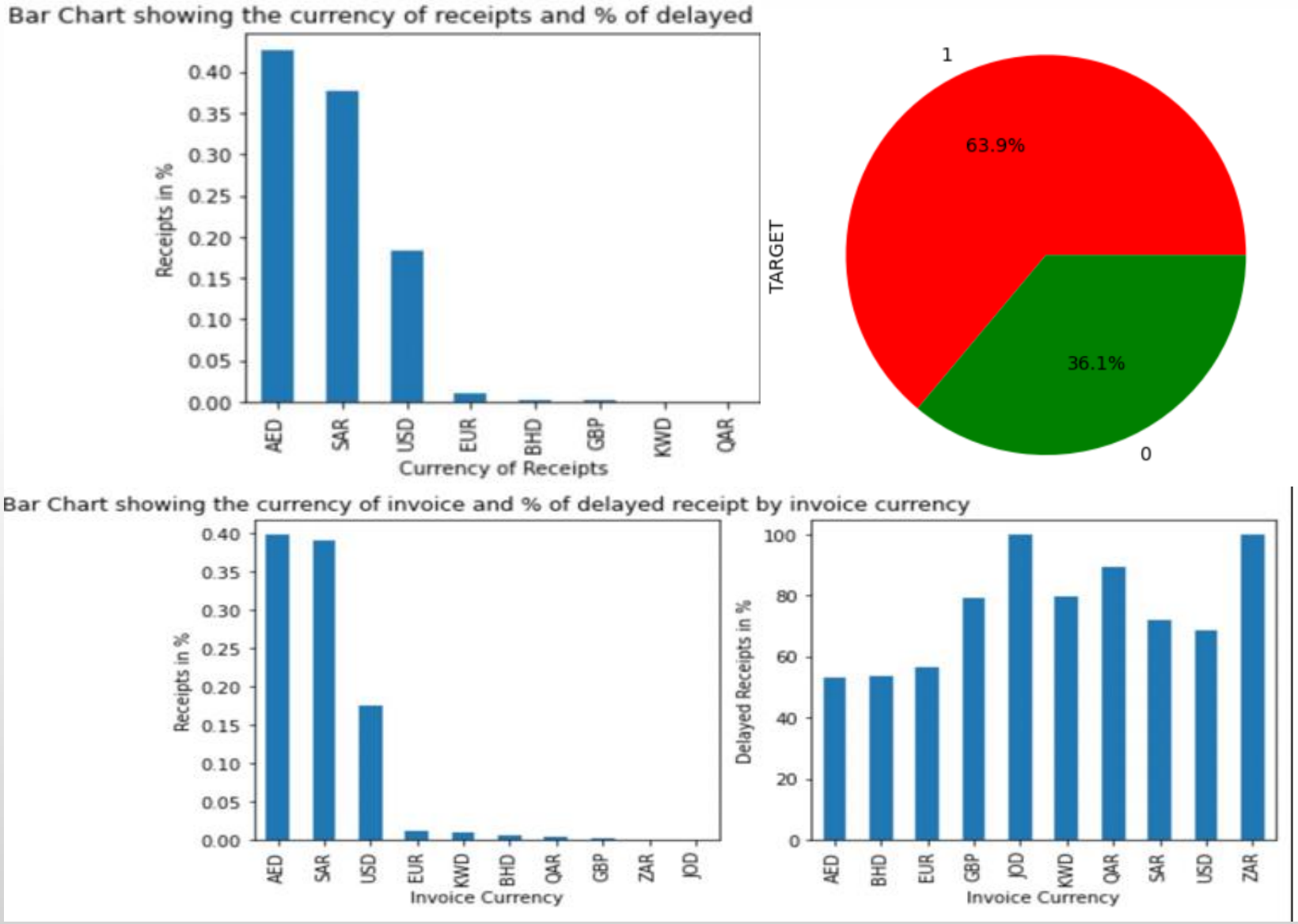
- Duplicate records were removed to avoid data ambiguity.
- Though the percentage of Null values in RECEIPT_DOC_NO column constitute 0.04% of the total records, still we can delete this column as it is not required for building the model.

Outliers

- As the invoices with 0 value are not viable for analysis, hence records were purged and only those invoices with Local Amount > 0 were kept for further analysis.
- Under USD Amount column, there are 4 txs above 8 million by SEPH Corp., however, as they are one of the Top Clients, we are retaining outlier values.

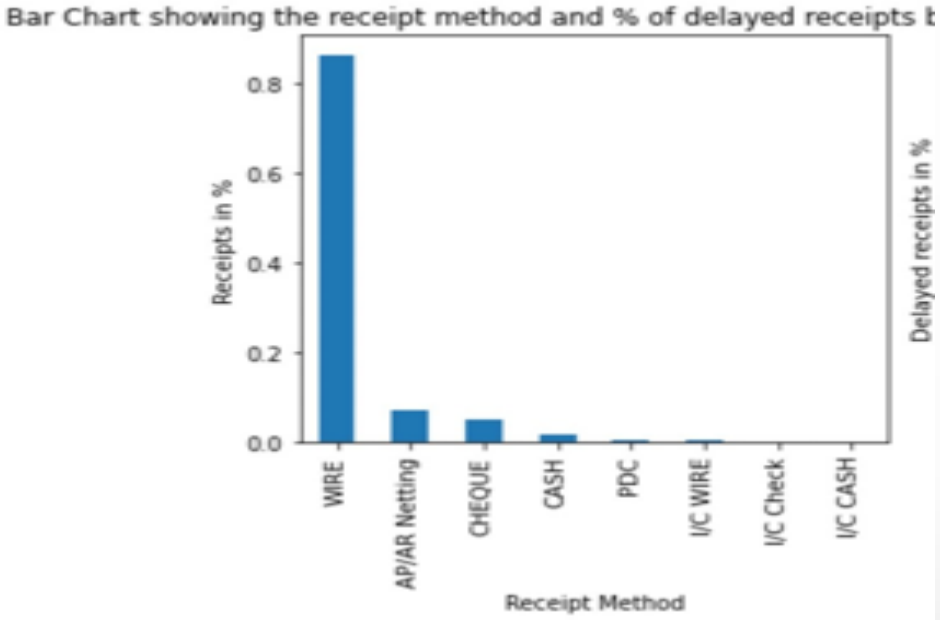
EDA Insights

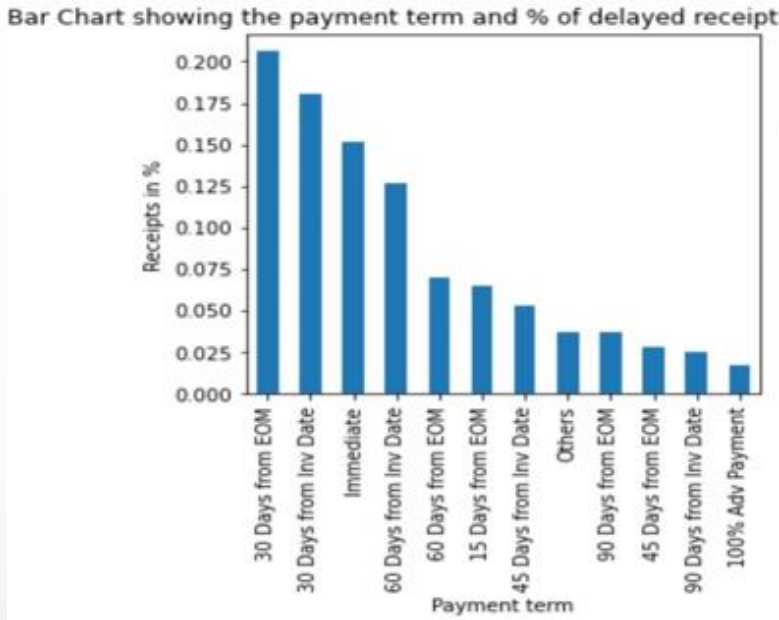
- Around 63.9 % of the times payment (corresponding to TARGET Variable 1) was delayed.
- Delays are high where payments are received in KWD followed by GBP
- No major data imbalance in the target variable.
- Invoices are mainly raised in AED, SAR and USD.
- Almost 100% delay where invoicing currency is JOD or ZAR

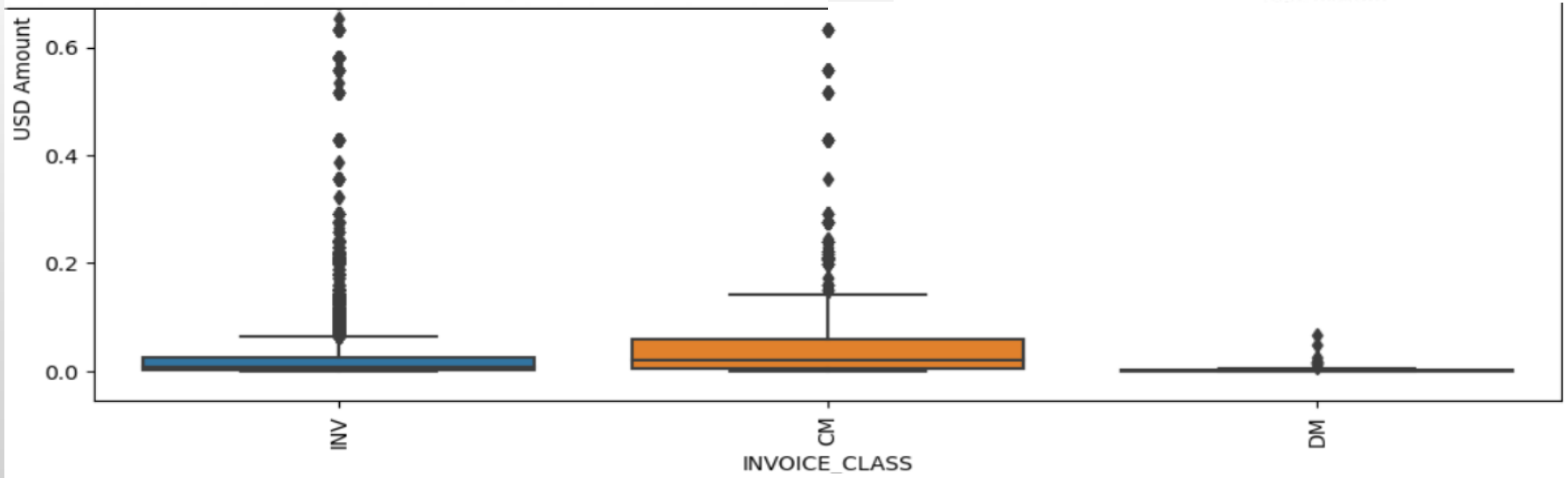


EDA Insights

- WIRE is popular method of payment, but delays are high in case of L/C Check
- Payment delays are high in case payment terms are 15 days from EOM or Immediate
- The median value of time taken for
- payment is high in case of 15 days from EOM

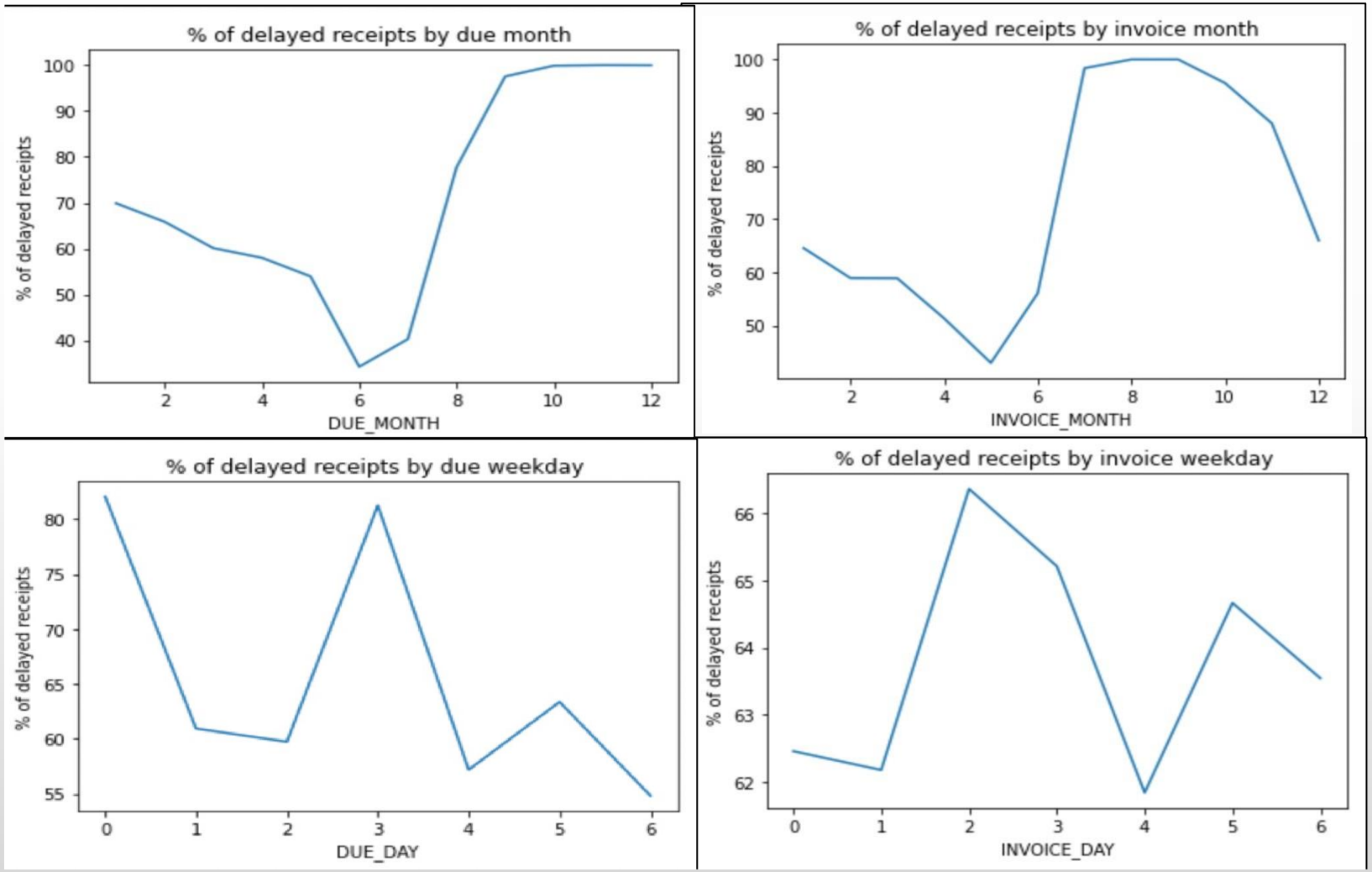






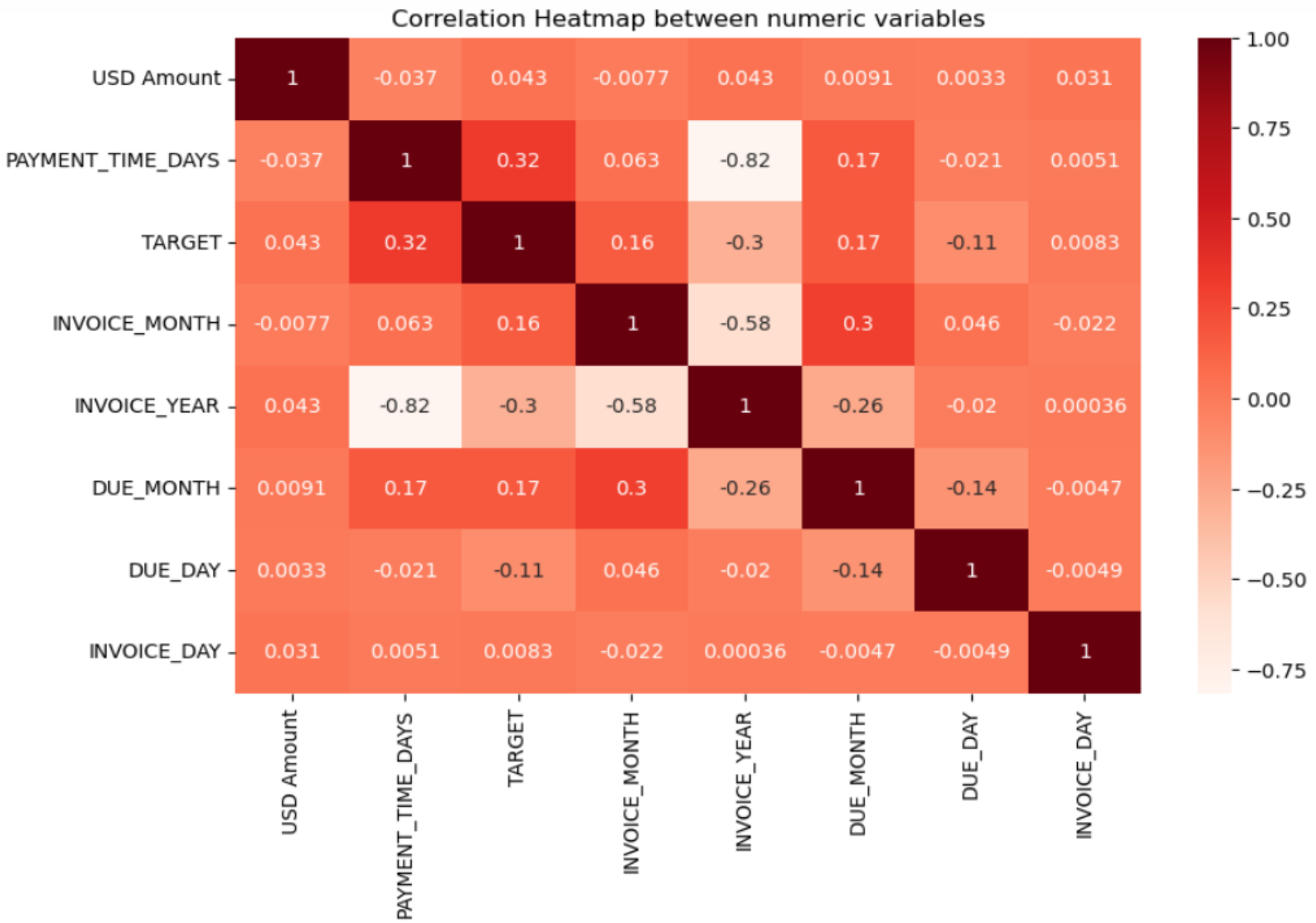
EDA Insights

- Most of Schuster's Billings are for value less than 1Mn
- Schuster has few high value customers
- No strong correlation found among the numeric variables



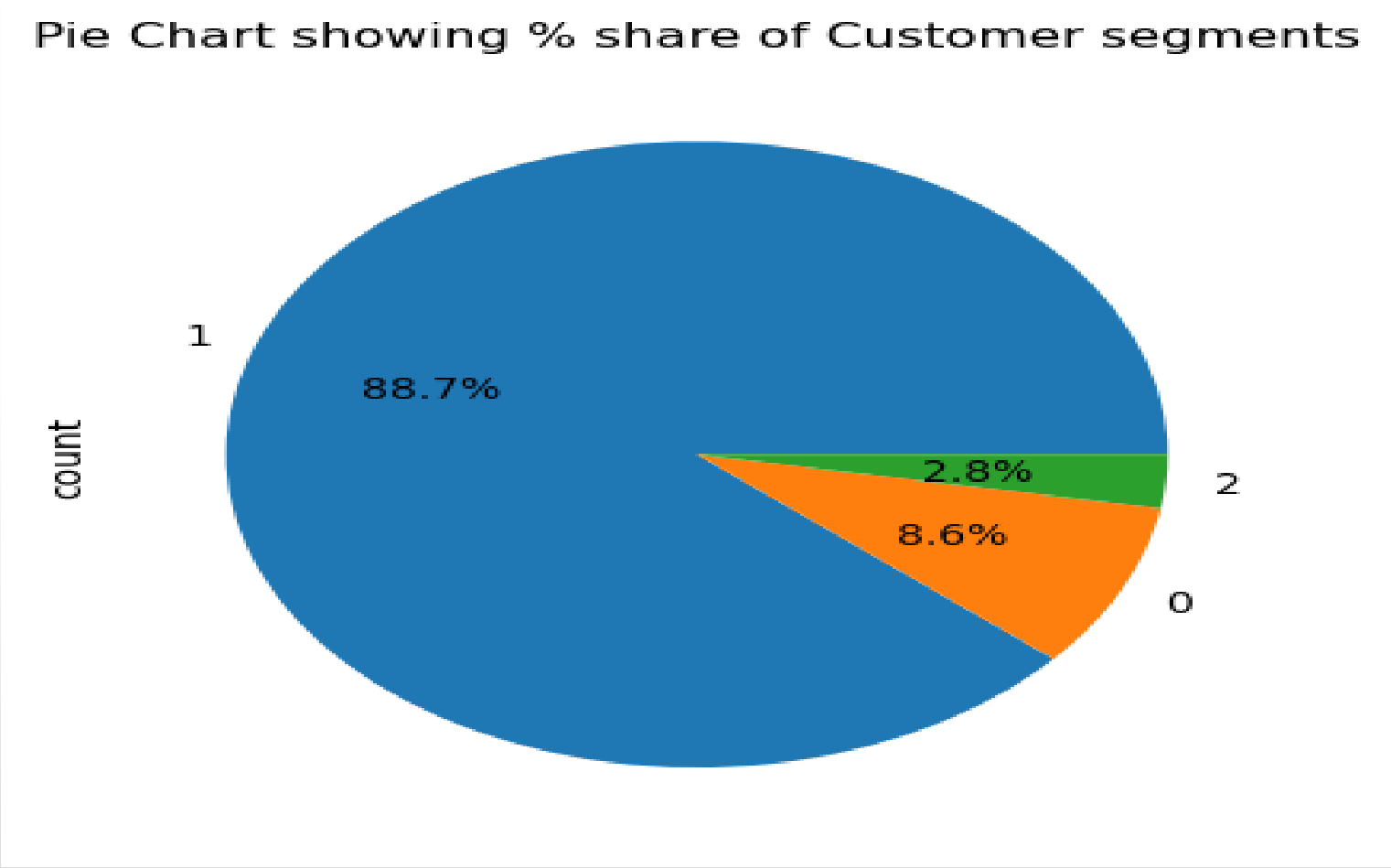
EDA Insights

- Significant delays in receipts where invoices pertain to Jul-Sep or invoices due in the period Oct-Dec
- Invoices due on Monday and Thursday of the week have been delayed by the customer
- % of delayed receipts is high where invoices are raised on Wednesday



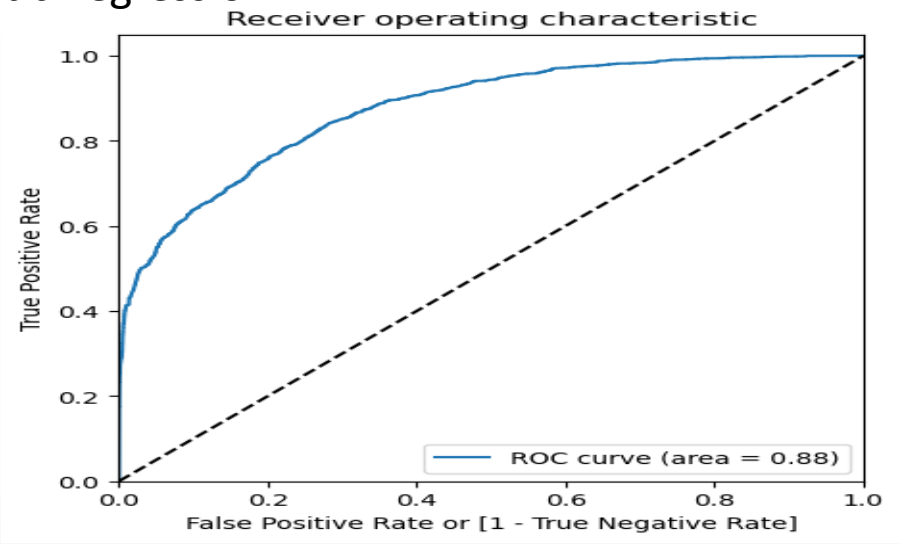
Customer Segmentation

- Majority of the customers are in Cluster 1, focus should be on Cluster 0 and 2.
- Cluster 2: Prolonged delays with moderate stand deviation
- Cluster 1: Early Payments with low std deviation
- Cluster 0: Moderate Delays with high standard deviation

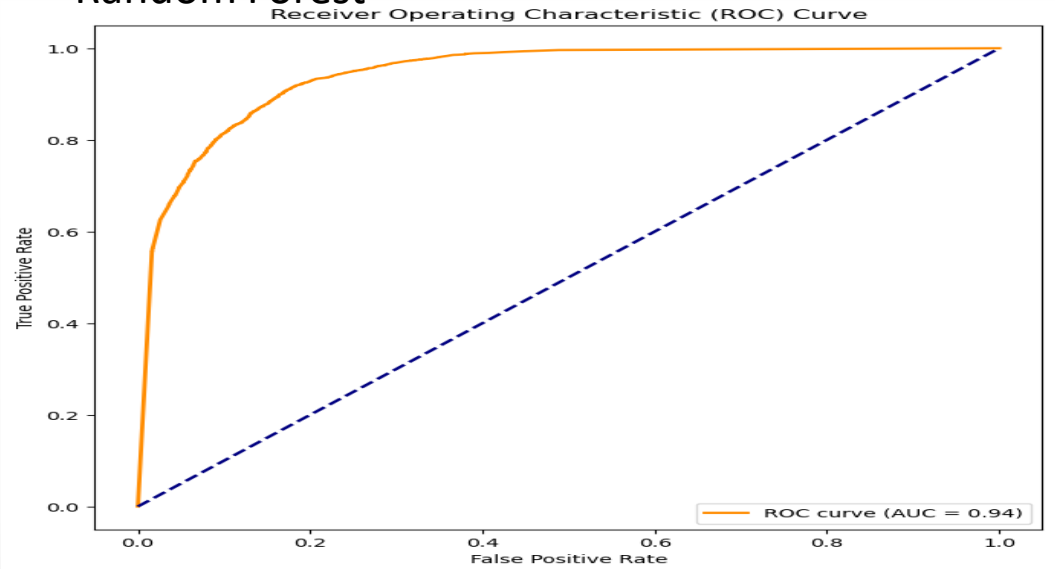


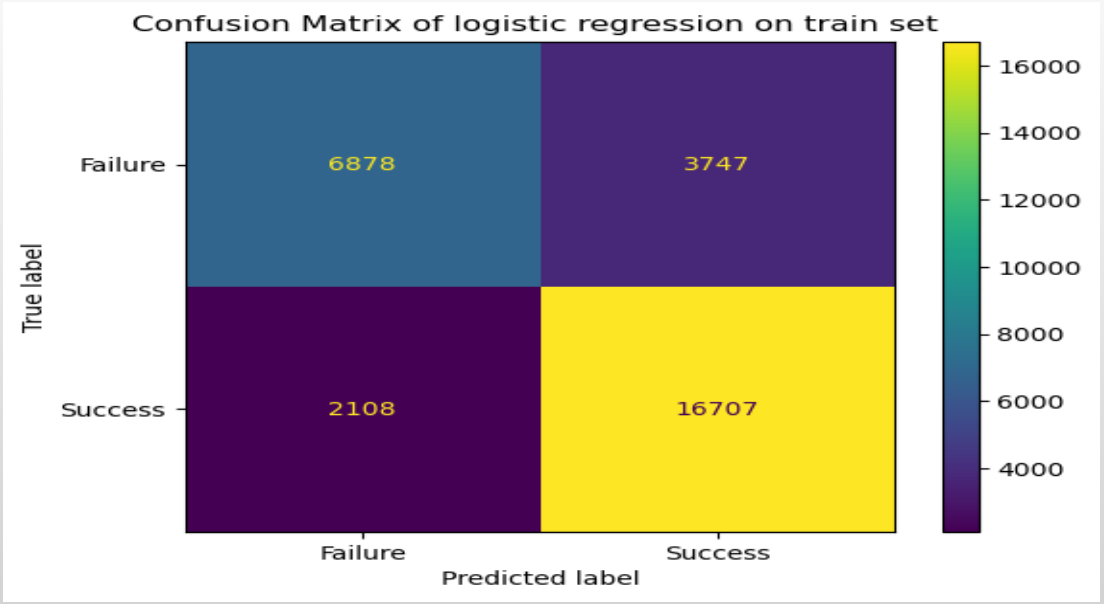
Random Forest Vs. Logistic Regression

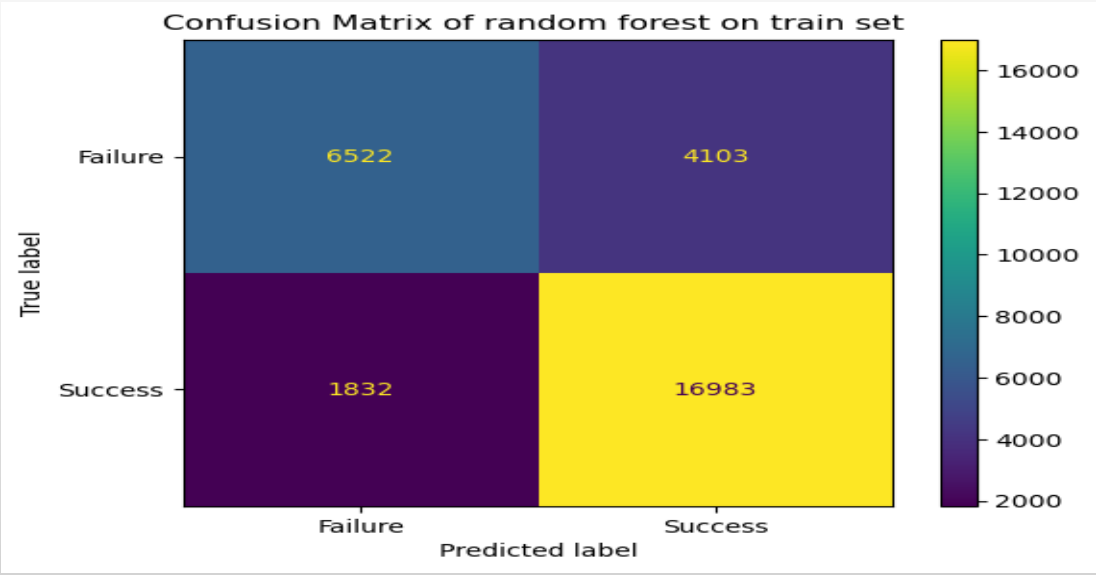
Logistic Regression



Random Forest



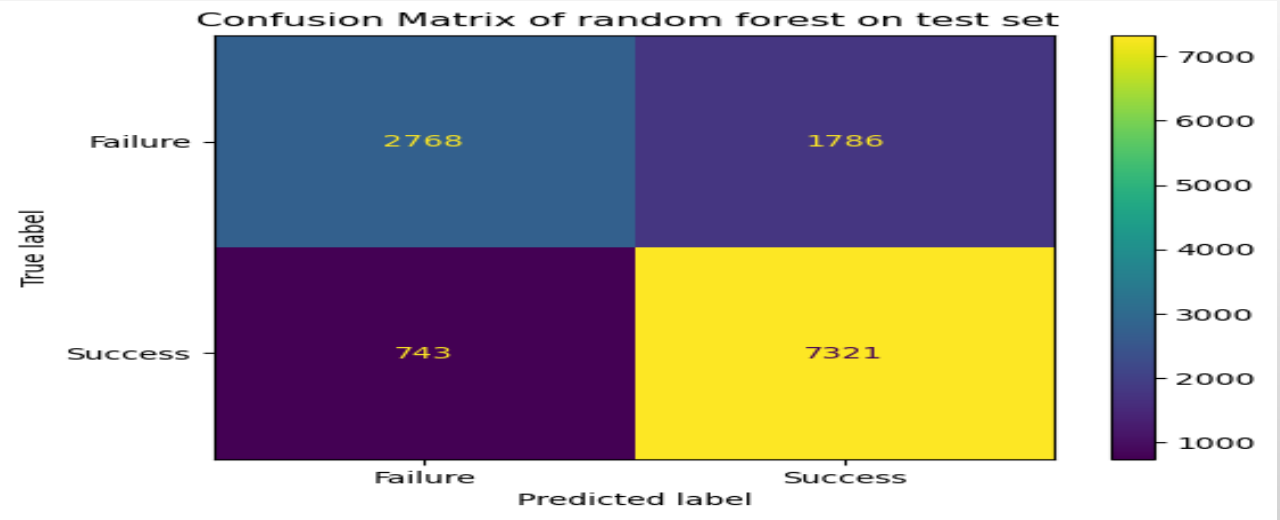




Random Forest vs Logistic Regression

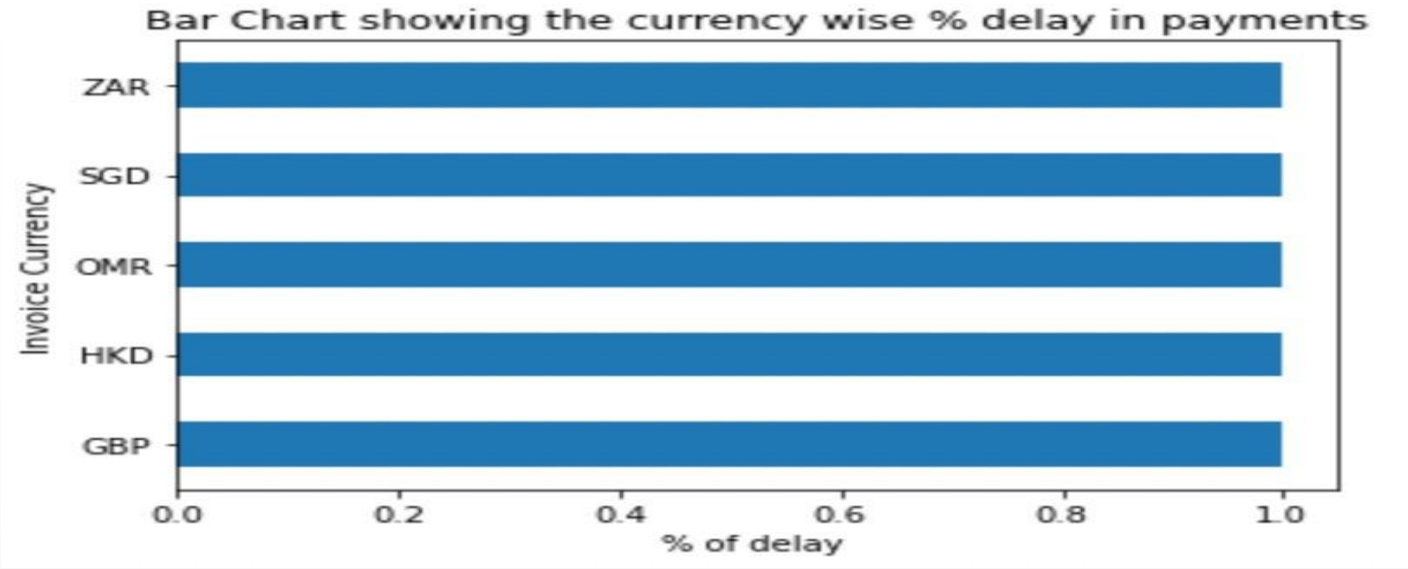
Random forest is the preferred model for deployment for this Case Study as:

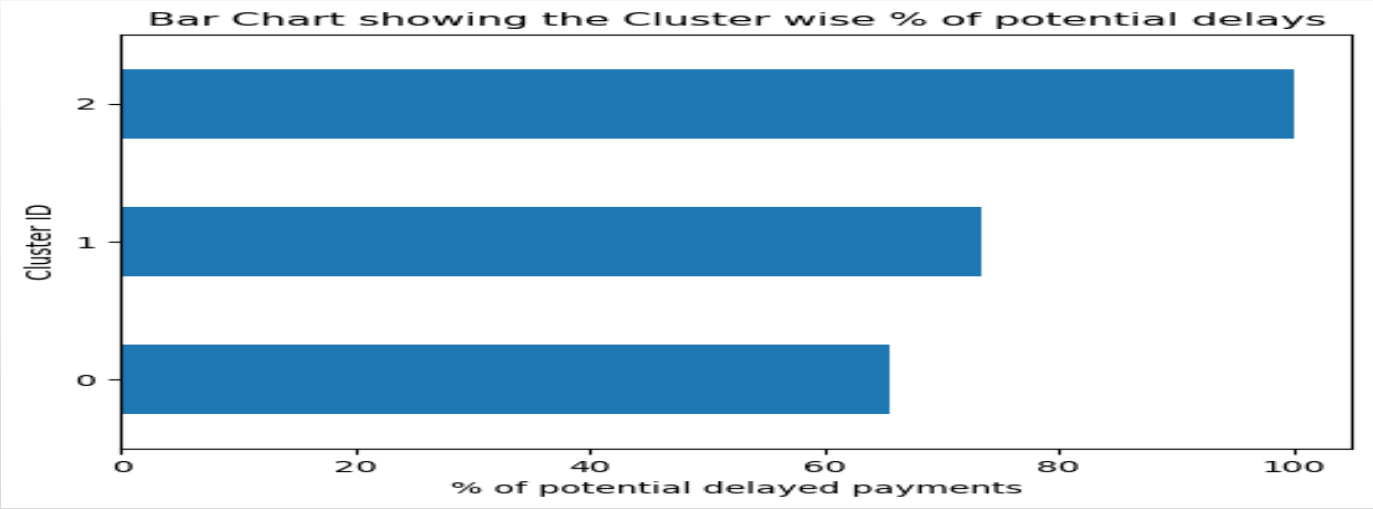
- Majority AUC of random forest 0.94 is better than logistic regression(0.88)
- Sensitivity(Recall) of random forest is 0.90 on both train and test sets
- Sensitivity(Recall) of logistic regression is 0.89 on the train and 0.88 on the test set.
- Precision 0.8 on train and test sets of Logistic Regression and Random forest
- Accuracy of random forest is also better than logistic regression.
- High sensitivity will ensure that the model will identify the cases of delayed payments much better thereby
- giving heads-up for early follow-up.



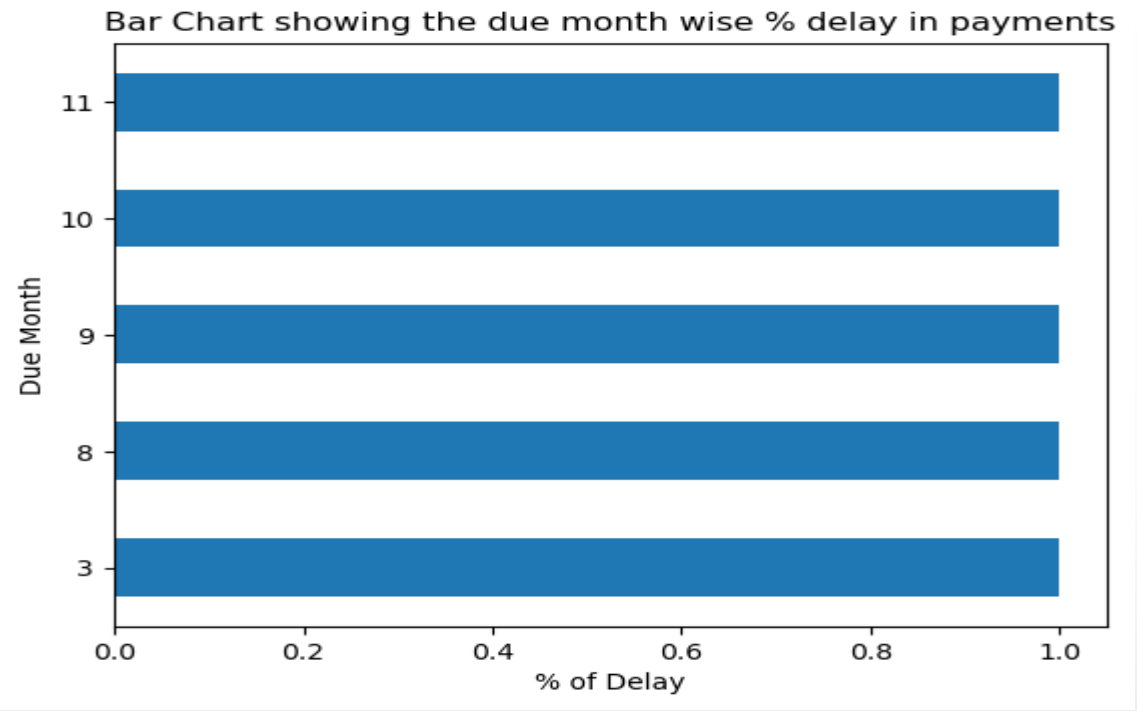
Predictions on Deployment

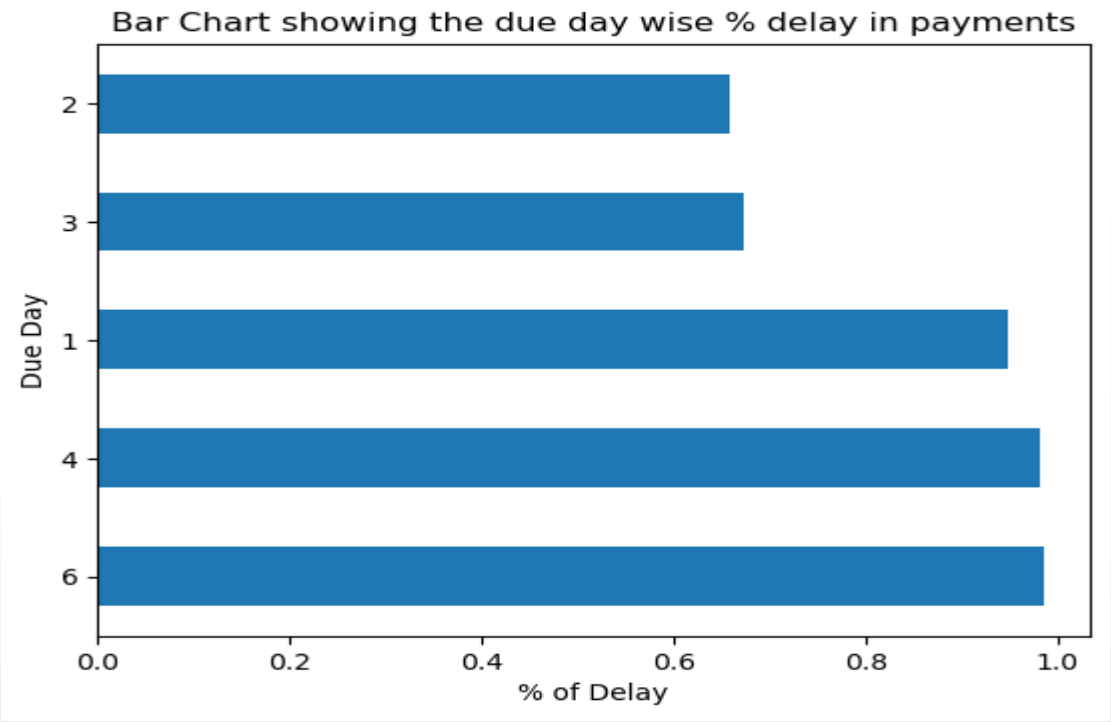
- On deployment, there's delay expected in 70.21% of the open invoices.
- 100% delays can be expected if:
  - Invoice currency is GBP,HKD,OMR,SGD and ZAR
  - Invoice due on Friday or Sunday
  - Customers belonging to Cluster ID 0





Predictions on Deployment





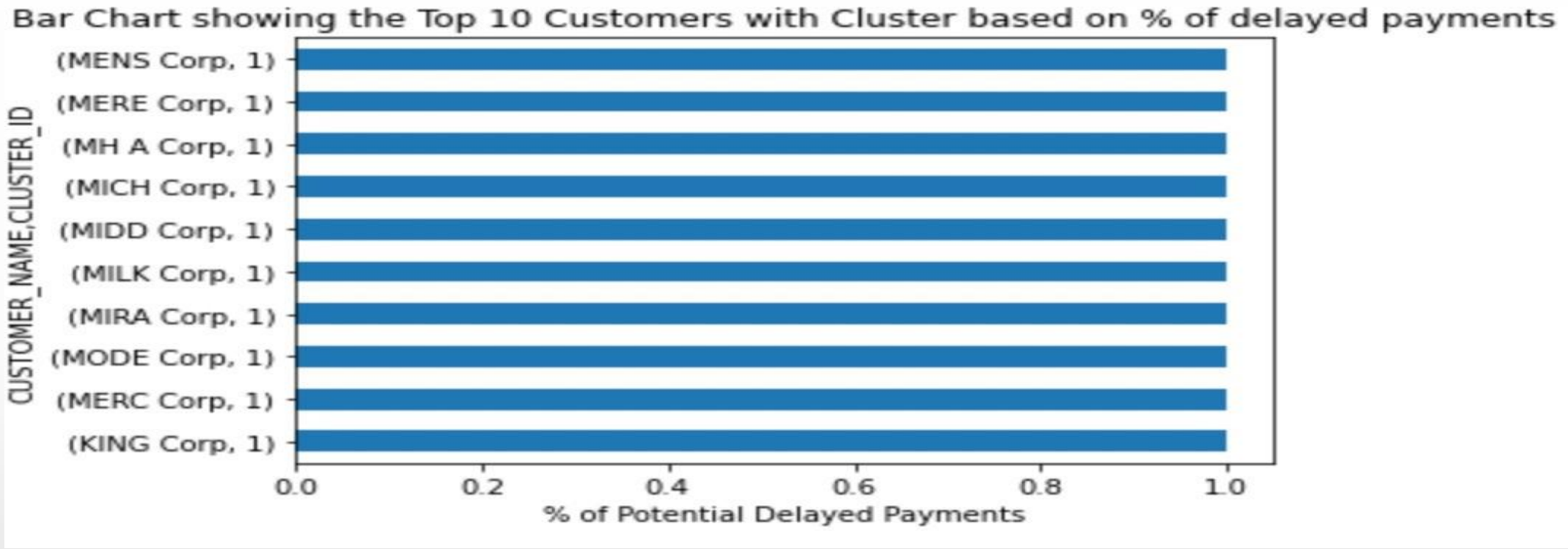
- Invoices due in March, Jul-Oct period will need advance follow-up
- Delays can be expected on Invoices due on Sunday and Friday, proactive follow-up is required.

Potential Customer

CUSTOMER_NAME	Number of Open Invoices	% of Predicted Delayed Invoices	Delayed Invoices Count
SEPH Corp	8260	0.648305	5355.0
FITR Corp	3454	0.866532	2993.0
PARF Corp	1717	0.845079	1451.0
AREE Corp	1117	0.482543	539.0
ALLI Corp	1042	0.435701	454.0
HABC Corp	517	0.696325	360.0
AL T Corp	584	0.578767	338.0
DEBE Corp	654	0.481651	315.0
RADW Corp	490	0.589796	289.0
CARR Corp	363	0.796143	289.0

- Customers with large volume of invoices will need better tracking
- Follow up with Customers having high value billing will help reduce the receivables

Customer with Delayed Payements



Customers with Potential Delayed Payments

Recommendation & Conclusion

- Data Cluster 0 as these are cases of prolonged delays, so focus on all those cases.
- Invoices with huge invoice values or huge volume of transactions cases, should be followed up more.
- Assign a dedicated resource at SEPH to manage follow-ups, considering the substantial business value and the volume of invoices involved..
- Customers who have a high probability of default or delay % is 100%, should be followed up more.
- During follow-ups, pay special attention to invoices due in March, July to October, and those falling on Fridays and Sundays..
- Take a proactive approach, particularly in cases where the invoicing currency is ZAR, HKD, SGD, OMR, and GBP.

Thank you!