

E-Commerce and Retail B2B Case Study

Ramapriya Sasikumar
Alisha Mishra

Addressing the issue and defining objectives

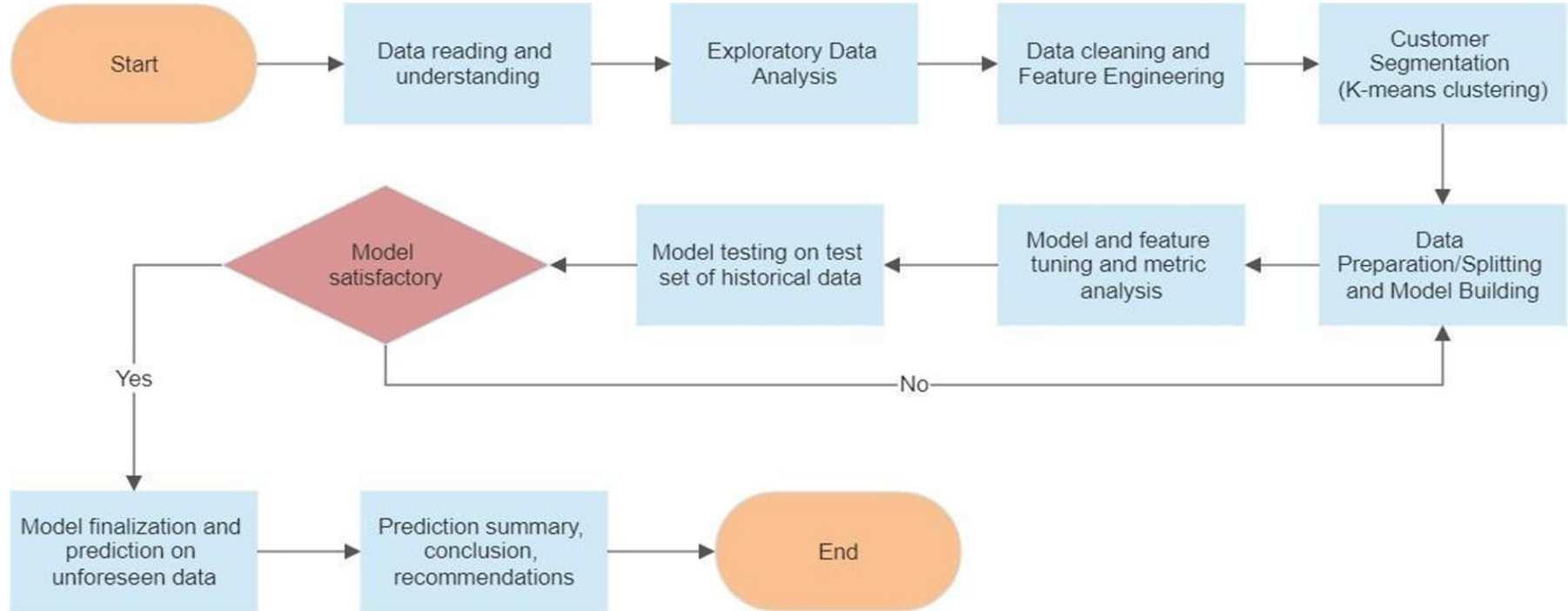
Problem identification

- A sports retail company Schuster dealing in B2B transactions often deals with vendors on a credit basis, who might or might not respect the stipulated deadline for payment
- Vendors delaying their payments result in financial lag and loss which becomes detrimental to smooth business operations
- Additionally, company employees are set up chasing around for collecting payments for a long period of time resulting in no value-added activities and wasteful resource expenditure

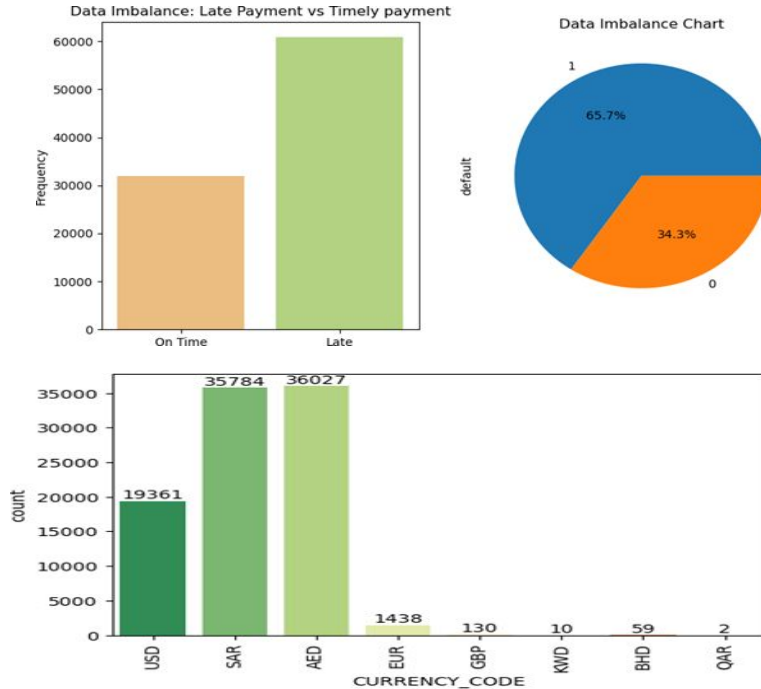
Business Objectives

- Customer segmentation to understand the customer's payment behavior
- Using historical information, the company requires prediction of delayed payment against an unforeseen dataset of transactions with due date yet to be crossed
- The company requires the prediction for better resource delegation, quicker credit recovery and reduction of low value-adding activities

Approach Strategy to the Problem

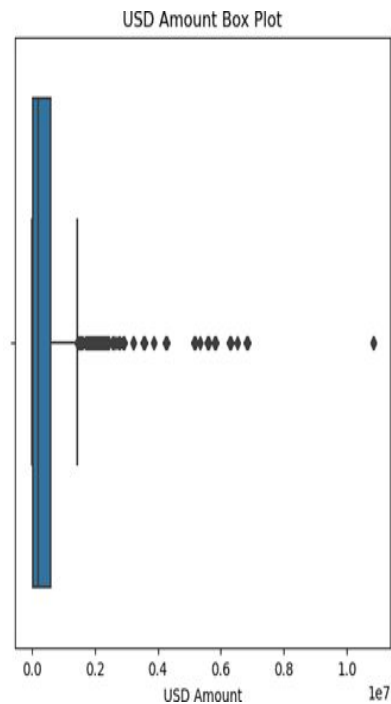
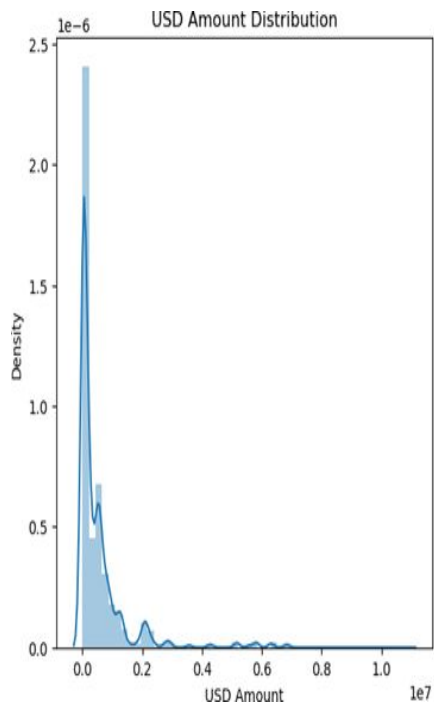


Class imbalance and transaction insights (univariate)



From Fig. 1 and 2:

- The class imbalance is 65.7% towards payment delayers which is an acceptable imbalance and does not need imbalance treatment
- The top three currencies in which the company deals are AED, SAR and USD with AED as the most dealt currency suggesting greater transactions with the middle-east

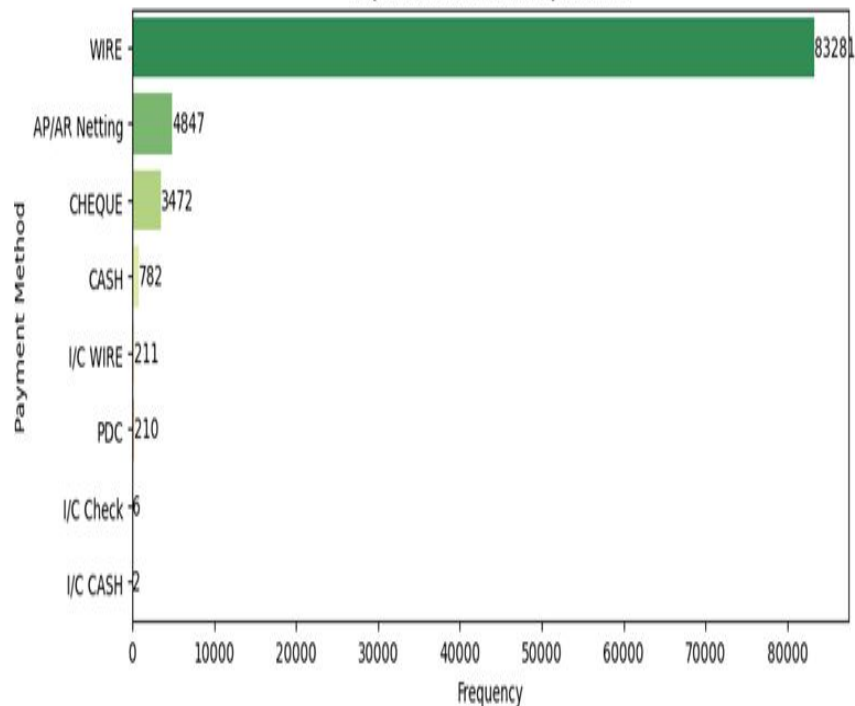


From fig 3.

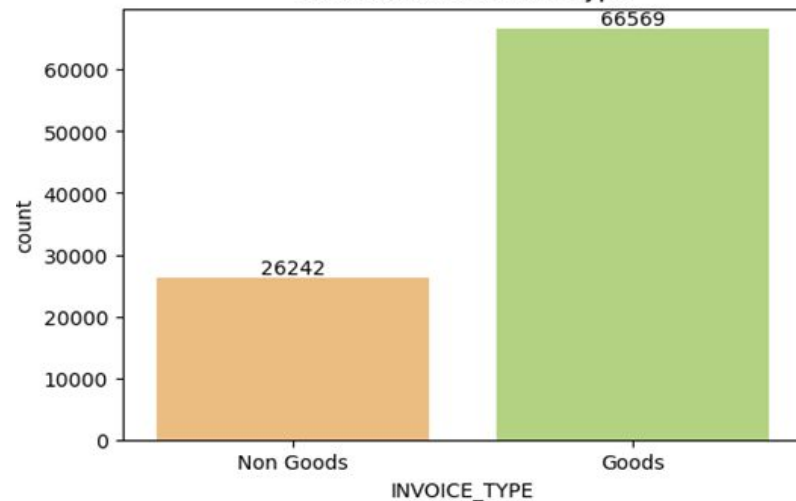
- The transaction values seem to lie between a range of \$1m to \$3m.
- The transaction values are most frequent below \$1.75m

Class imbalance and transaction insights

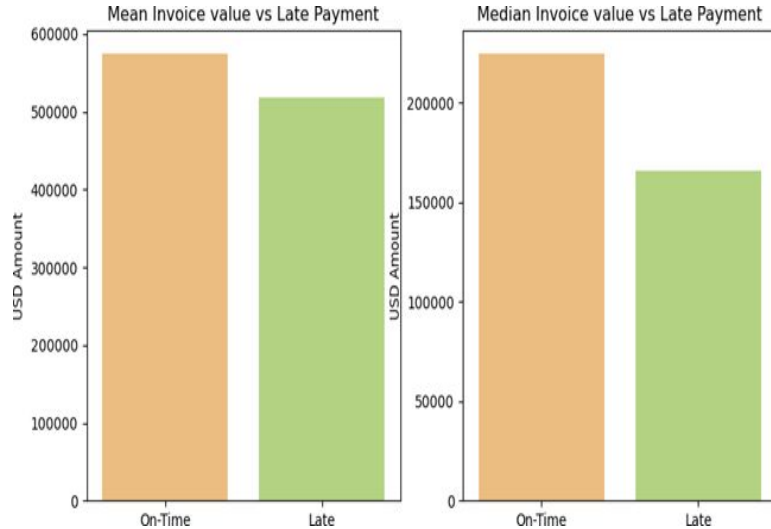
Payment methods used by vendors



Distribution of Invoice Type



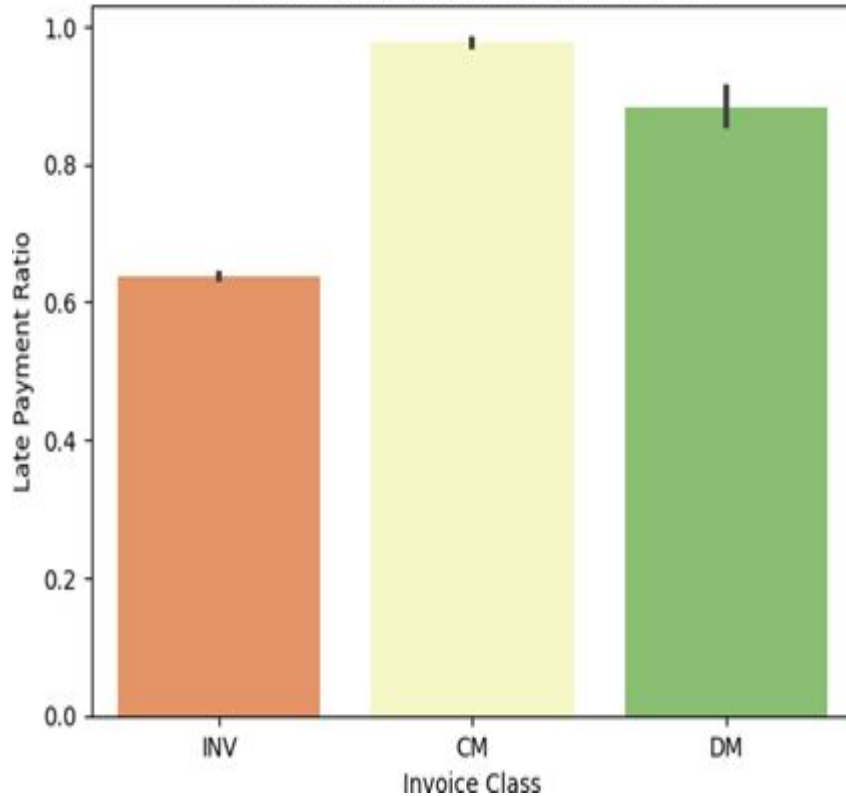
Identifying characteristics of defaulter payment types (Bivariate)



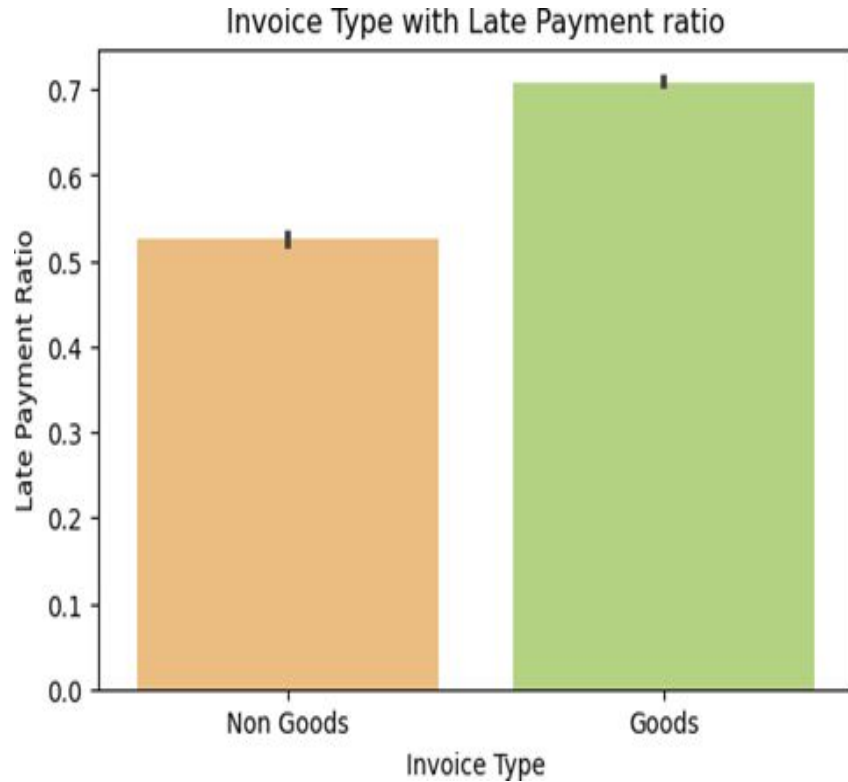
From the given figure,

- The mean and median of the payment amount is higher for payers who pay on time than late, suggesting higher value transaction show lesser delay risk than lower value transactions

Invoice Class with Late Payment ratio

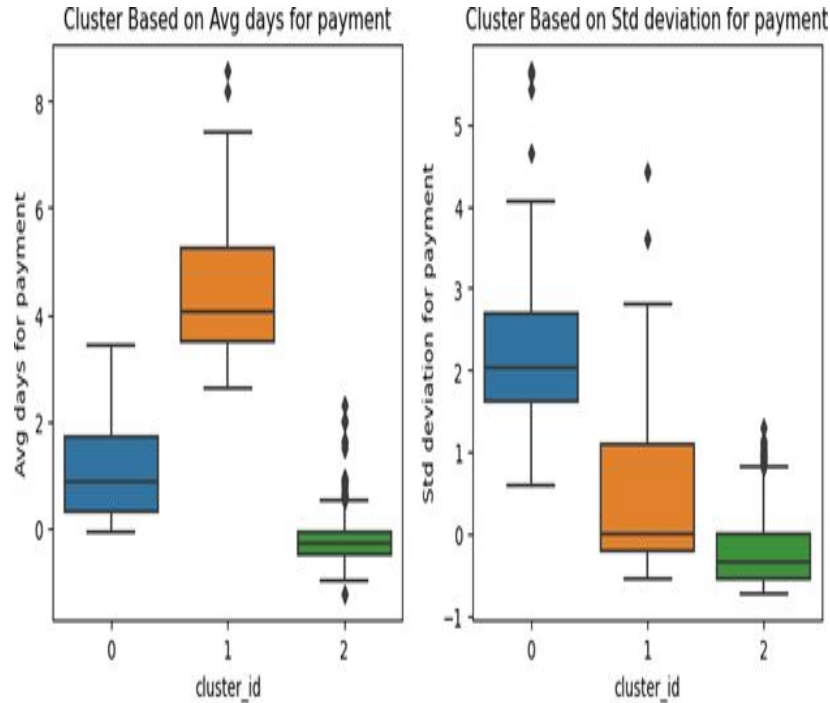


From fig. 2, late payment ratio for Credit Note transaction types are maximum, followed by Debit Note and Invoice suggesting higher delay risk in Credit and Debit note invoice classes



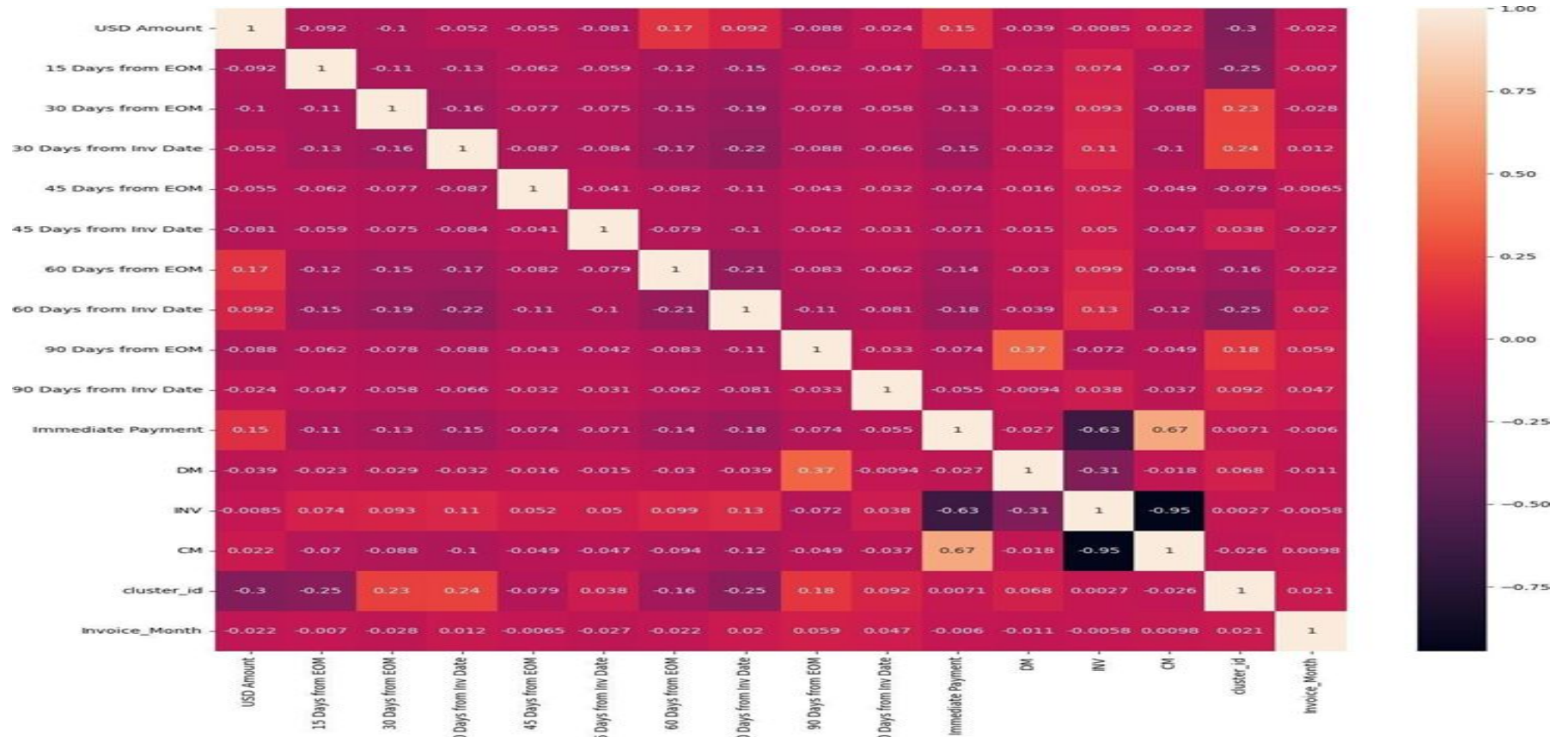
Goods type invoices show greater late payment ratio than non goods hence showing increased chances of payment delay.

Customer segmentation using K- Means Clustering

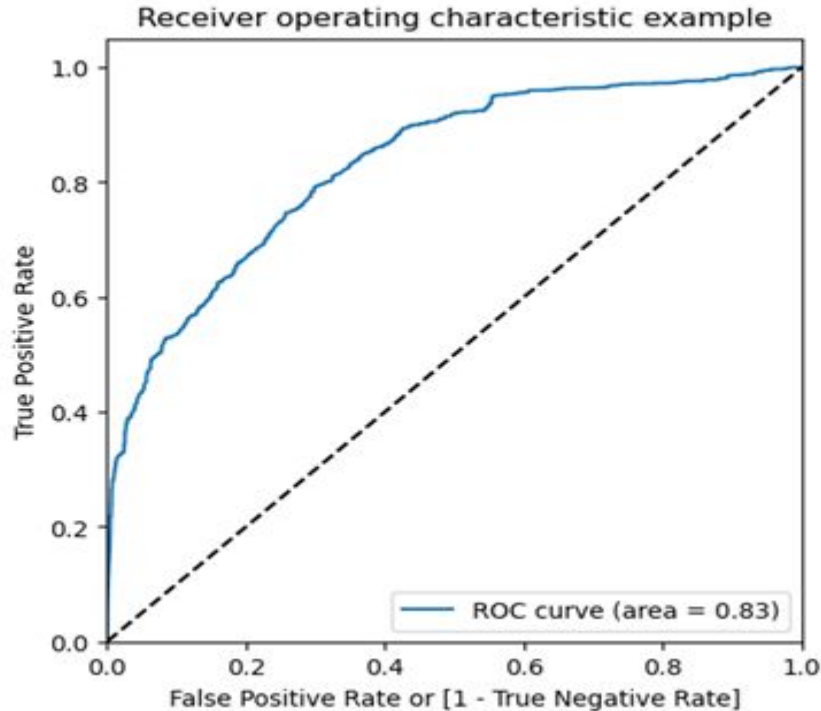


- The category 2 were early payers with least number of average days taken to pay and category 1 were prolonged payers with greatest number of average days taken to pay.
- The number of clusters were decided to be 3 since with increase in clusters post 3, there was a significant decrease in silhouette score.
- It was also observed that prolonged players historically have significantly greater rates of delay in payment than early or medium payment durations.

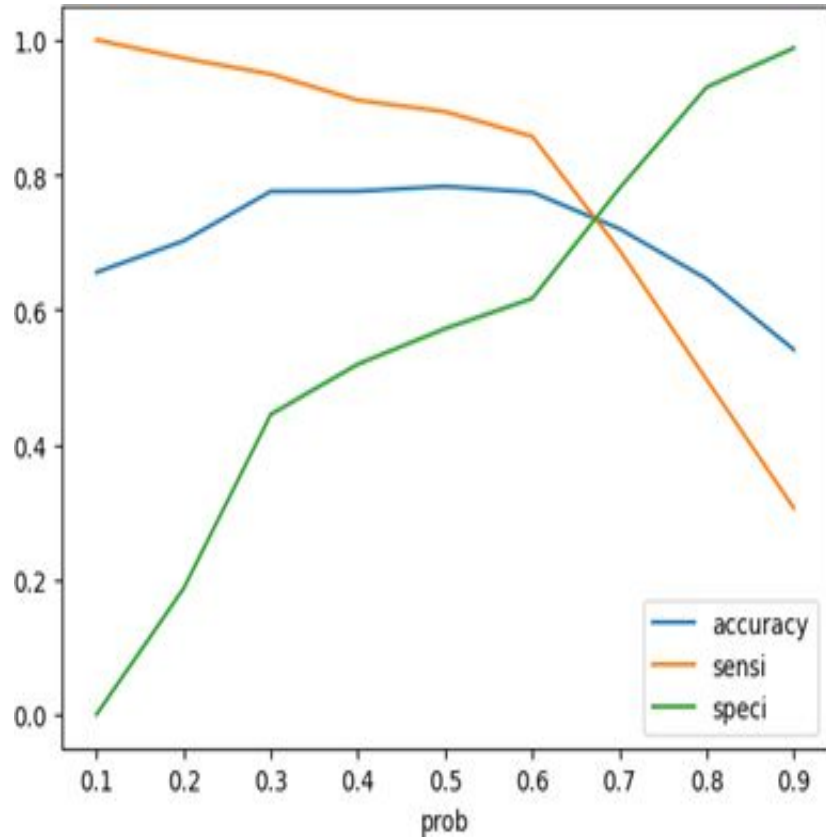
Model Evaluation



Comparison between two models, Logistic Regression and Random forest



- Logistic Regression model formed after dropping multicollinearity and unnecessary variables resulted in remaining variables with acceptable p-value and VIF figures, hence retained the remaining features with no further feature elimination and a good ROC curve area of 0.83.



- The trade off plot between accuracy, sensitivity and specificity revealed an optimum probability cutoff of ~ 0.6 , which was used further to predict which transactions would result in delayed payments in the received payments dataset.

Comparison between two models, logistic regression and random forest

- A random forest model was built using the same parameters as the logistic regression with hyper parameter tuning which resulted in following parameters,

```
Best hyperparameters: {'max_depth': 30, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 150}
```

```
Best f1 score: 0.9394084954678357
```

Using the above parameters, a random forest was built, whose metrics were compared to the logistic regression model and the final model was finalized before.

Random forest found better than Logistic Regression

```
# Let's check the overall accuracy.  
accuracy_score(y_pred_final.default, y_pred_final.final_predicted)  
  
0.7754632955035196
```

```
#precision score  
precision_score(y_pred_final.default, y_pred_final.final_predicted)  
  
0.8115658179569116
```

```
# Recall Score  
recall_score(y_pred.default, y_pred.final_predicted)  
  
0.8569416073818412
```

	precision	recall	f1-score	support
0	0.92	0.85	0.88	9502
1	0.93	0.96	0.94	18342
accuracy			0.92	27844
macro avg	0.92	0.91	0.91	27844
weighted avg	0.92	0.92	0.92	27844

It can be observed that the overall precision and recall scores of the Random forest model far exceeded the logistic regression model. Also, recall scores were most important in this case since it was important to increase the percentage prediction of late players to be targeted. Since the data is heavy on categorical variables, random forest is better suited to the job than logistic regression.

Random Forest Feature Ratings

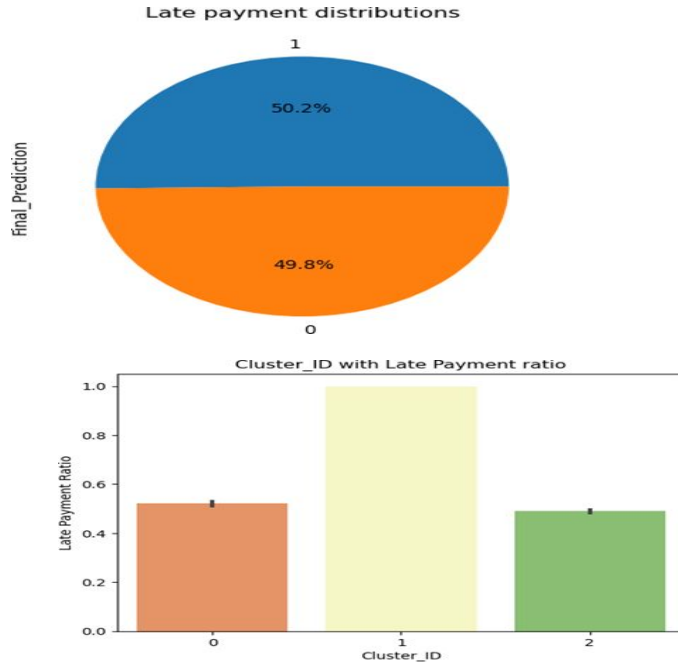
Feature ranking:

1. USD Amount (0.465)
2. Invoice_Month (0.130)
3. 60 Days from EOM (0.113)
4. 30 Days from EOM (0.105)
5. cluster_id (0.053)
6. Immediate Payment (0.042)
7. 15 Days from EOM (0.027)
8. 30 Days from Inv Date (0.015)
9. 60 Days from Inv Date (0.013)
10. 90 Days from Inv Date (0.008)
11. INV (0.007)
12. 90 Days from EOM (0.006)

The random forest was then used to find out the feature rankings which shows that top 5 features to predict delay which included,

- USD Amount
- Invoice Month
- 60 days from EOM (Payment Term Variable)
- 30 days from EOM (Payment Term Variable)
- Cluster ID

50% payments predicted to be delayed as per Open Invoice data, prolonged payments days to observe alarmingly high delay rates



- Predictions made by the final model suggests that there is a probable 50.2% transactions where payment delay can be expected, which can cause a shocking lag to business operations
- Customer segment with historically prolonged payment days are anticipated to have the most delay rate (~100%) than historically early or medium days payment transactions, this is similar to the result found based on historical outcomes

Customers with the highest delay probabilities

Customer_Name	Delayed_Payment	Total_Payments	Delay%
ALSU Corp	7	7	100.0
LVMH Corp	4	4	100.0
MILK Corp	3	3	100.0
MUOS Corp	3	3	100.0
MAYC Corp	3	3	100.0
ROVE Corp	3	3	100.0
AMAT Corp	3	3	100.0
TRAF Corp	3	3	100.0
CITY Corp	3	3	100.0
DAEM Corp	3	3	100.0

- Predictions suggest that the companies presented in the table to the left has the maximum probability of default with maximum number of delayed and total payments

Recommendations

Customer_Name	Delayed_Payment	Total_Payments	Delay%
ALSU Corp	7	7	100.0
LVMH Corp	4	4	100.0
MILK Corp	3	3	100.0
MUOS Corp	3	3	100.0
MAYC Corp	3	3	100.0
ROVE Corp	3	3	100.0
AMAT Corp	3	3	100.0
TRAF Corp	3	3	100.0
CITY Corp	3	3	100.0
DAEM Corp	3	3	100.0

From our clustering analysis we can make the following inference

- Credit Note Payments observe the greatest delay rate, compared to Debit Note or Invoice type invoice classes, hence company policies on payment collection could be made stricter around such invoice classes
- Goods type invoices had significantly greater payment delay rates than non-goods types and hence can be subjected to stricter payment policies
- Since lower value payments comprise of the majority of the transactions, also late payments are seen more on lower value payments, it is recommended to focus more on those. The company can apply penalties depending on billing amount, the lesser the bill, the greater the percentage of penalty on late payments. Of course this has to be last resort
- Credit Note Payments observe the greatest delay rate, compared to Debit Note or Invoice type invoice classes, hence company policies on payment collection could be made stricter around such invoice classes
- Goods type invoices had significantly greater payment delay rates than non-goods types and hence can be subjected to stricter payment policies
- Since lower value payments comprise of the majority of the transactions, also late payments are seen more on lower value payments, it is recommended to focus more on those. The company can apply penalties depending on billing amount, the lesser the bill, the greater the percentage of penalty on late payments. Of course this has to be last resort