E-Commerce and Retail B2B Case Study

Shraddha

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

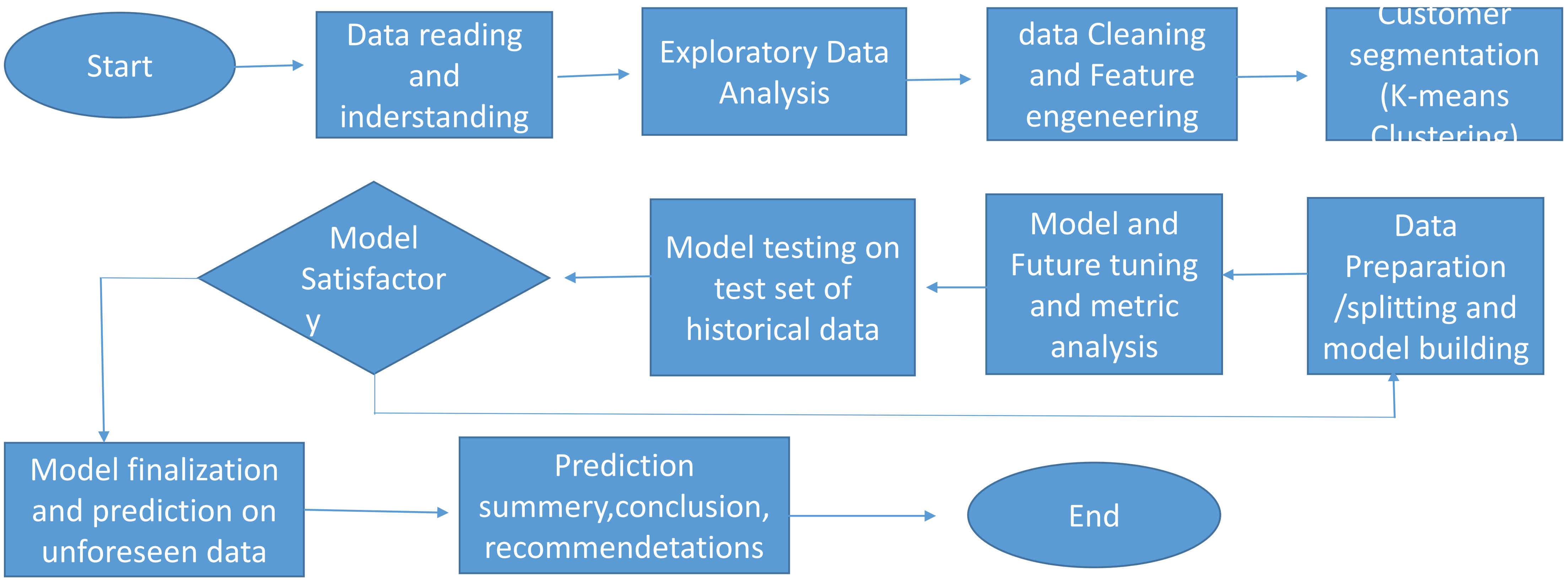
:-**Business Objectives**

• 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

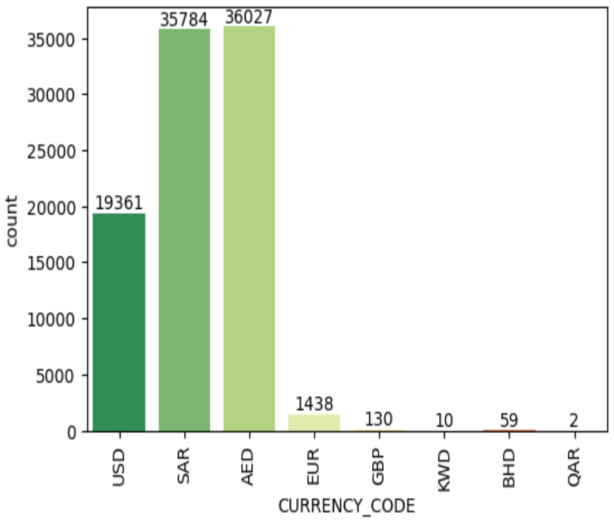
• The company requires the prediction for better resource delegation, quicker credit recovery and reduction of low value-adding activities.

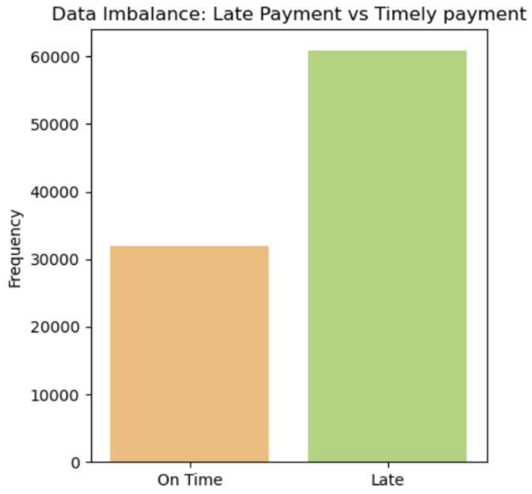
Approach Strategy to the Problem

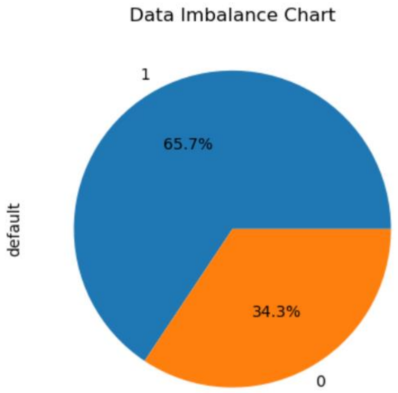


Class imbalance and transaction insights

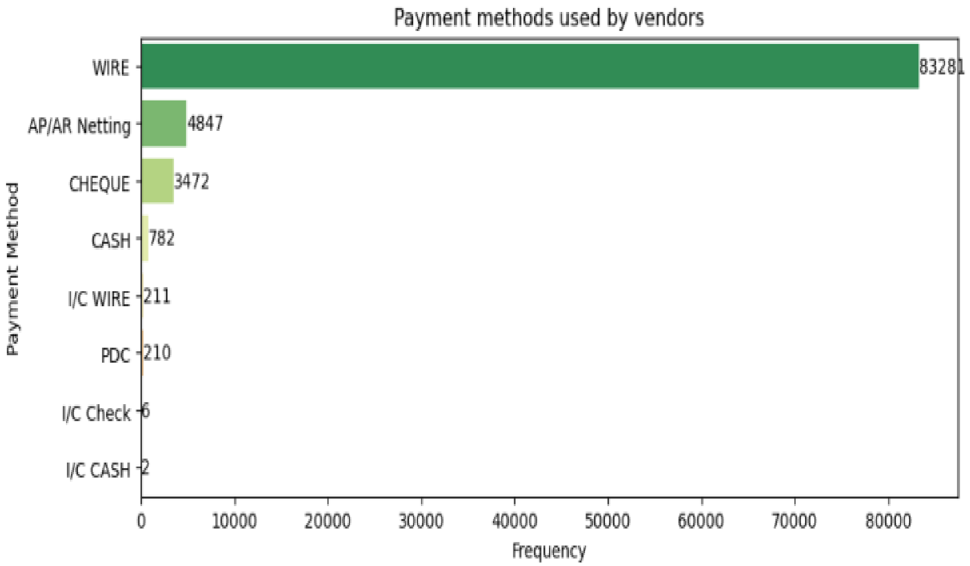
- The class imbalance is 65.7% towards payment delayers which is an acceptable imbalance and does not need imbalance treatment
- The top three currencies in which the company deals are AED, SAR and USD with AED as the most dealt currency suggesting greater transactions with the middle-east







Class imbalance and transaction insights



From Fig. 1, we observe, • Wire payment method is the most common payment method received by the company, followed by netting , cheque and cash

From Fig. 2 and 3: • Goods type invoices comprise of the major share of invoices generated • The major invoice class is 'Invoice' with the rest having very low percentages of the shar

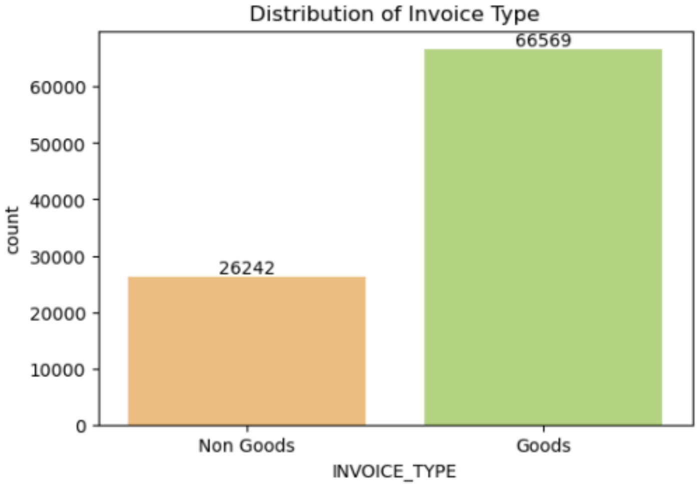


Fig. 2

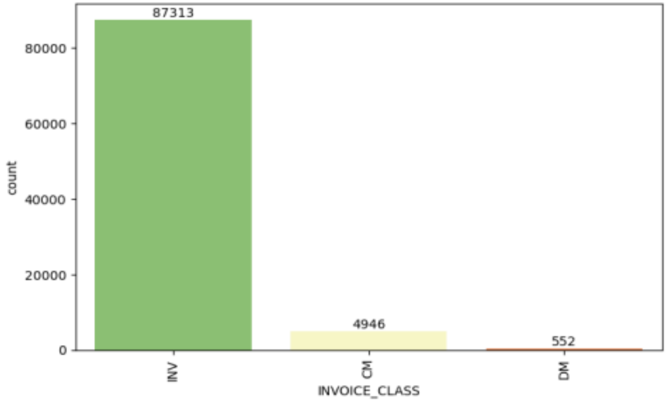


Fig. 3

Identifying characteristics of defaulter payment types

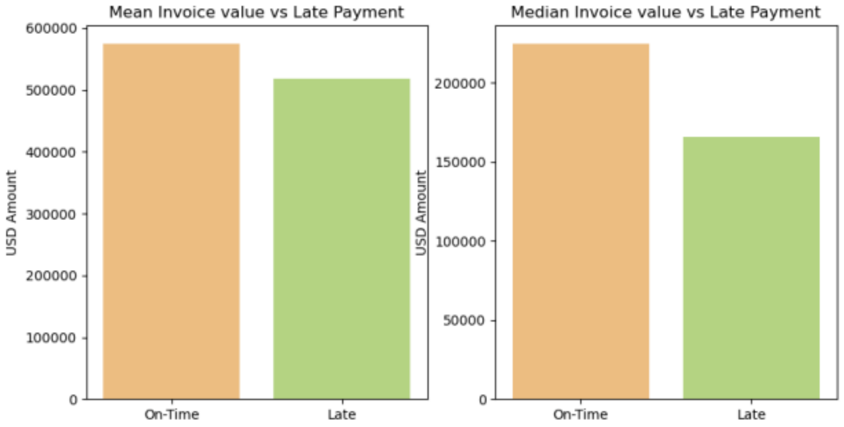


Fig. 1

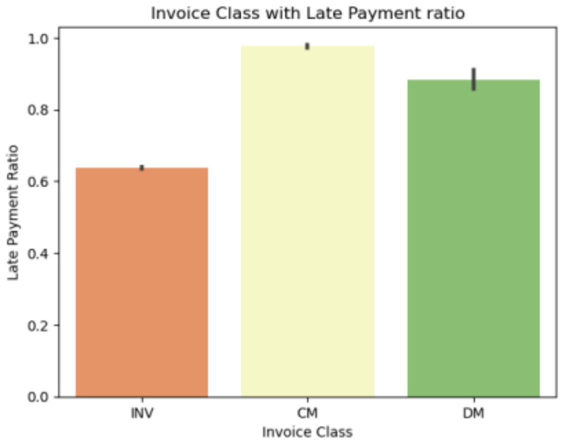


Fig. 2

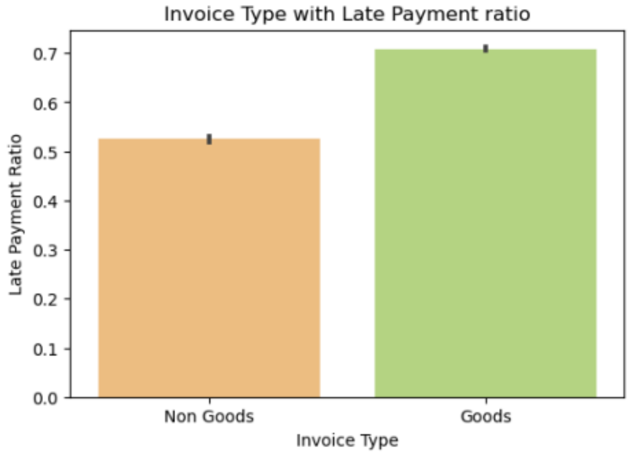


Fig. 3

From fig. 1, the mean and median of the payment amount is higher for payers who pay on time than late, suggesting that higher value transactions show lesser delay risk than lower value transactions

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

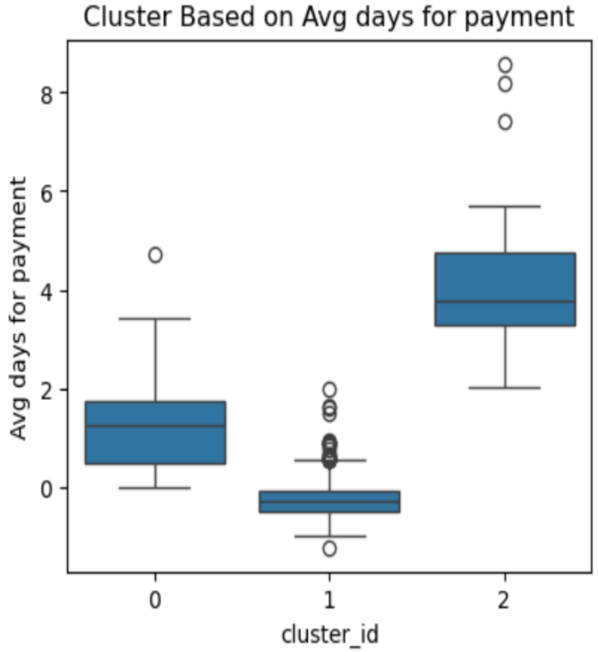
From fig. 3, Goods type invoices show greater late payment ratio than non-goods hence showing increased chances of payment delay

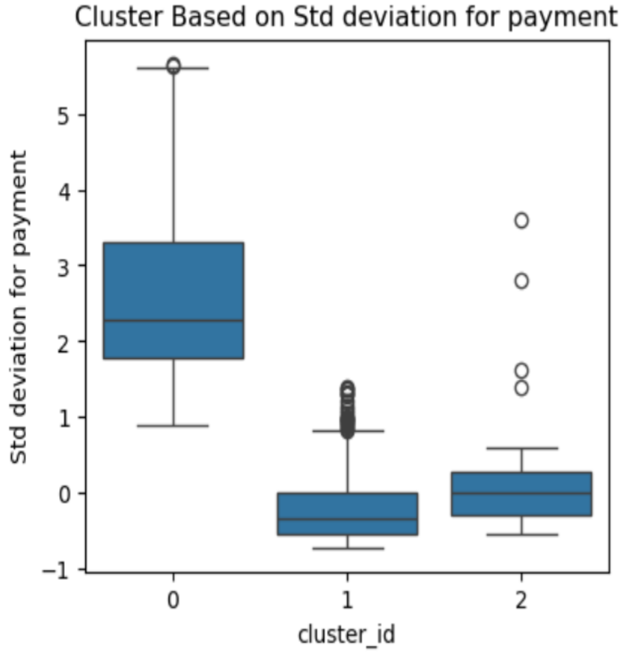
Customer segmentatio

The number of clusters were decided to be 3 since with increase in clusters post 3, there was a significant decrease in silhouette score

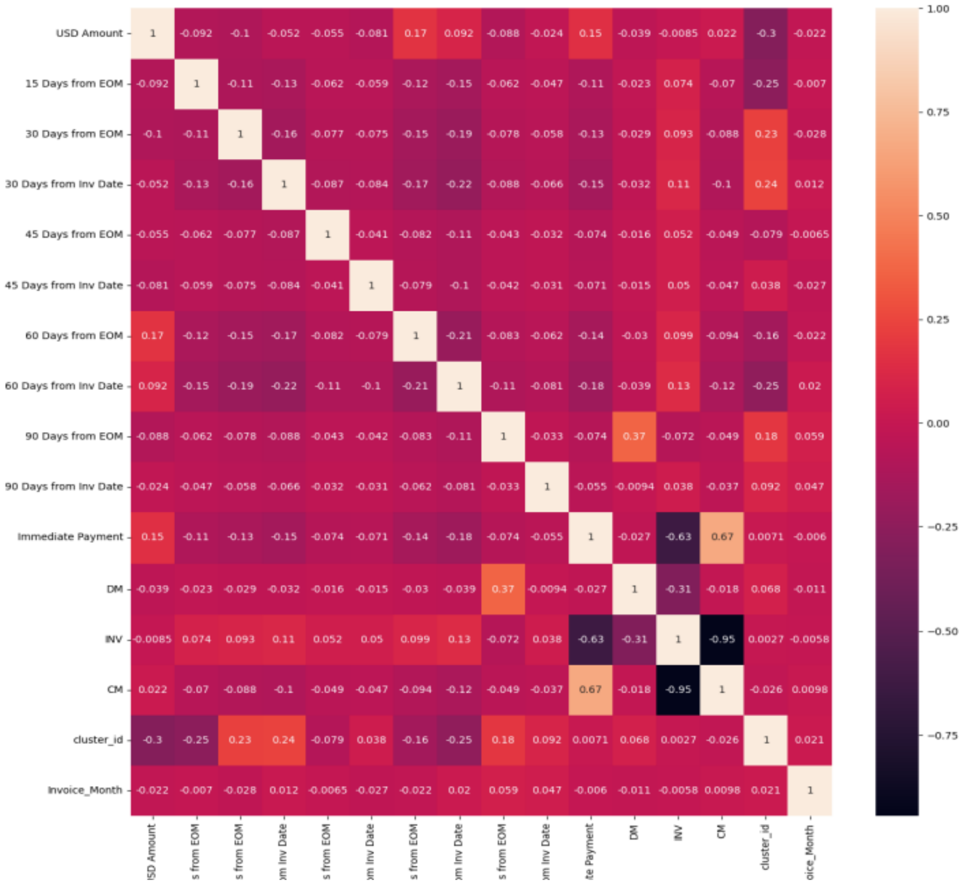
The number of clusters were decided to be 3 since with increase in clusters post 3, there was a significant decrease in silhouette score

```
For n_clusters=2, the silhouette score is 0.7557759850933141
For n_clusters=3, the silhouette score is 0.7491797445652462
For n_clusters=4, the silhouette score is 0.6097388985555463
For n_clusters=5, the silhouette score is 0.6173540681032771
For n_clusters=6, the silhouette score is 0.3980238443004184
For n_clusters=7, the silhouette score is 0.4012628375918799
For n_clusters=8, the silhouette score is 0.41457849738976615
```





Model Building



- CM & INV, INV & Immediate Payment, DM & 90 days from EOM has high multicollinearity, hence dropping these columns to prevent multicollinearity effect

Comparison between two models, logistic regression and random forests

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

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

Incomplete

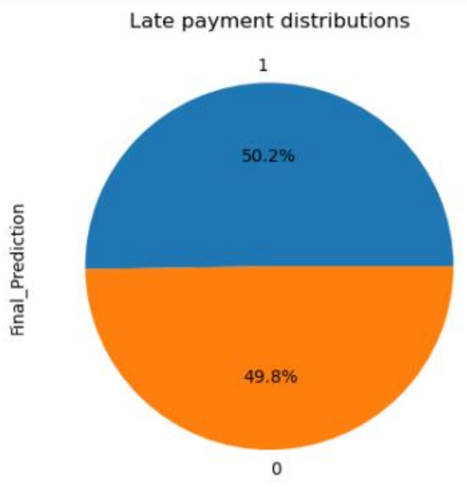
Random Forest Feature Ratings

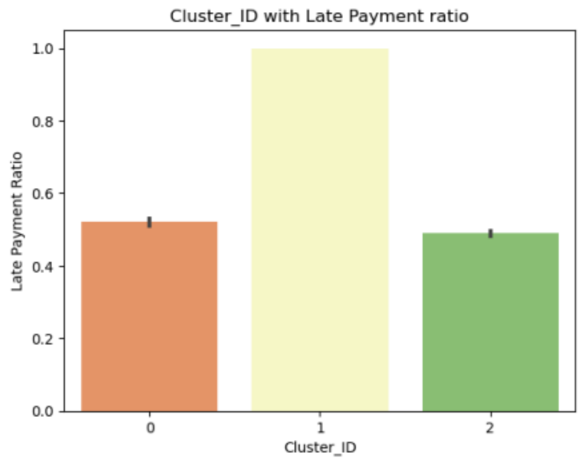
- The random forest was then used to find out the feature rankings which shows that the top 5 features to predict delay which included • USD Amount • Invoice Month • 60 Days from EOM (Payment Term variable) • 30 Days from EOM (Payment Term variable) • Cluster-ID (which in turn is dependent on average and standard deviation of days required to make payment)

The customers segmented with cluster ID was then applied to the open-invoice data as per the customer name and predictions were made

50% payments predicted to be delayed as per Open invoice data, prolonged payment 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
- 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 out comesto business operations.





Customers with the highest delay probabilities

- 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.

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