E- Commerce and Retail B2B Case study

Ms. Malini S
Mr. Manish Kumar
Ms. Meghana

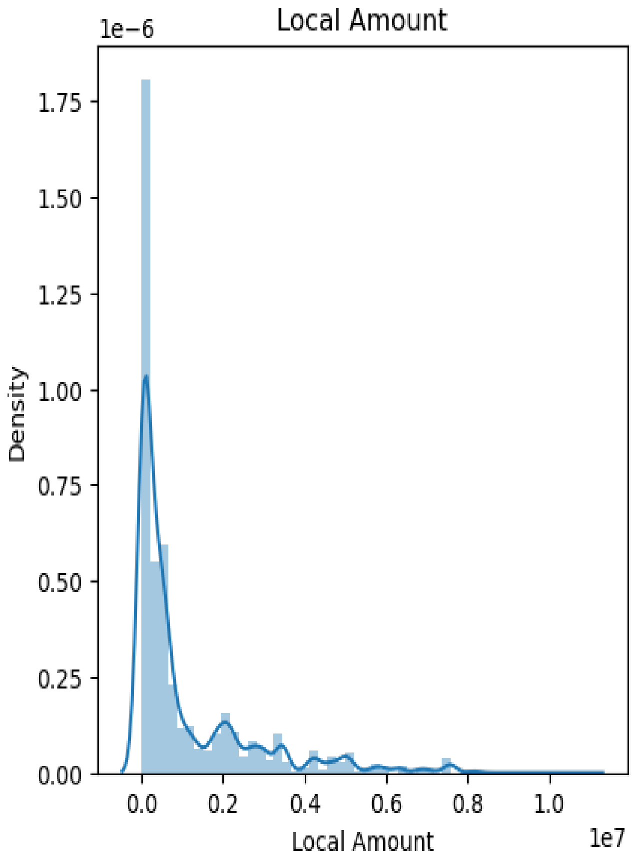
Goal:

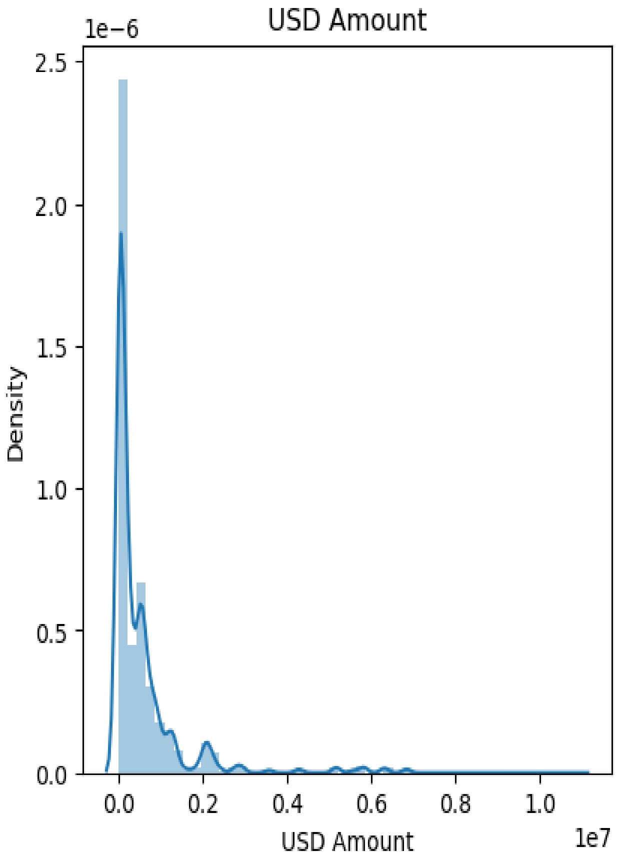
- Schuster would like to better understand the customers' payment behaviour based on their past payment patterns (customer segmentation).
- Using historical information, it wants to be able to predict the likelihood of delayed payment against open invoices from its customers.
- It wants to use this information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

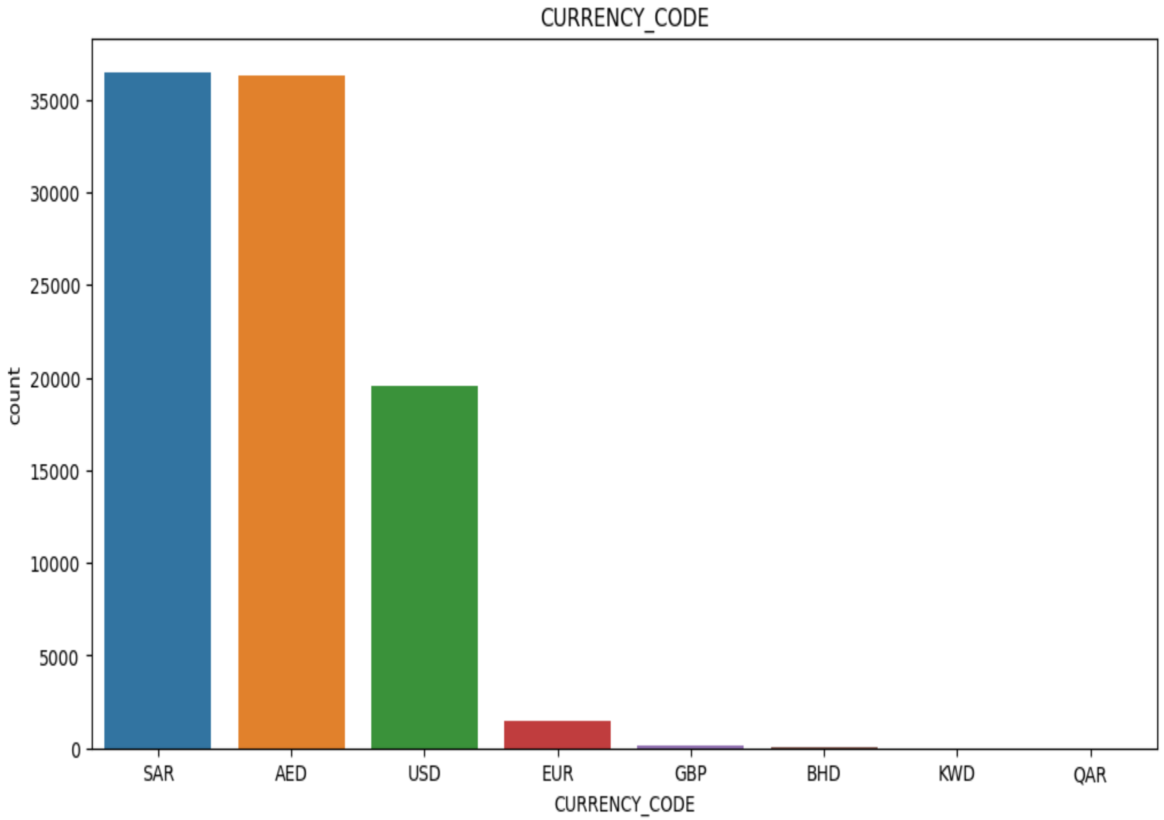
Algorithm

1. Reading the dataset of Received_Payments_Data
2. Data Cleaning
3. Exploratory Data Analysis
4. Feature Engineering
5. Reading, understanding and cleaning of open voice data
6. Segmentation on customers and clustering
7. Data preparation for Modeling
8. Model Building - Logistic Regression
9. Model Building - RandomForest
10. Adding Open_Invoice dataset for prediction
11. Results and Suggestions

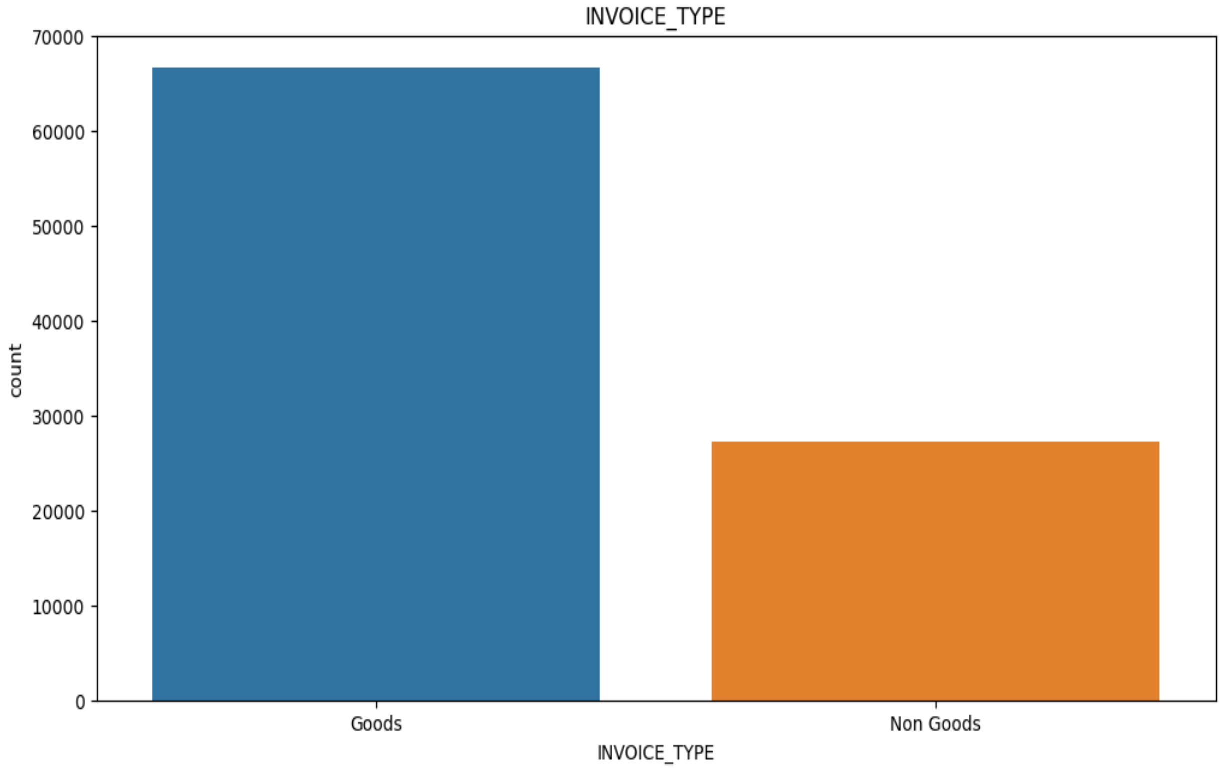
Uni variate Analysis for Local Amount, USD Amount and CURRENCY_CODE

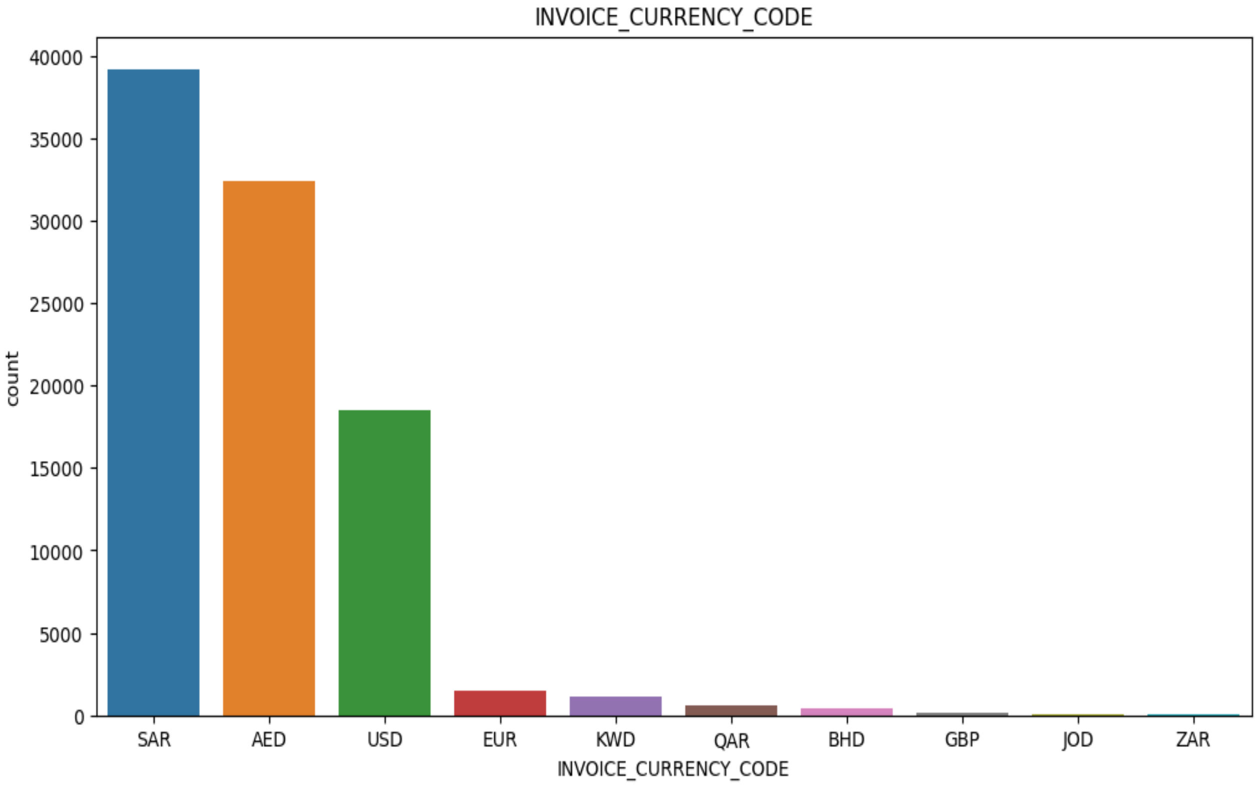






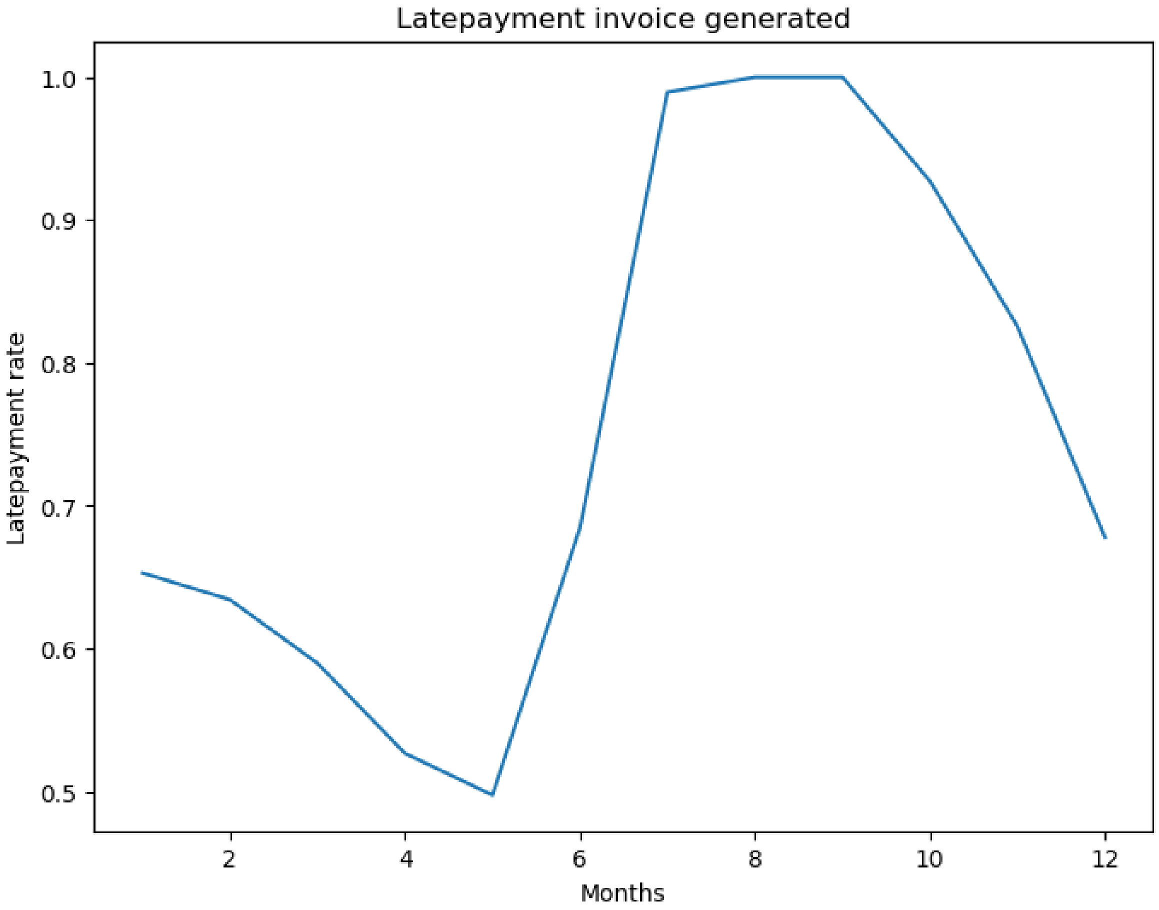
Univariate Analysis for INVOICE_TYPE and INVOICE_CURRENCY_CODE



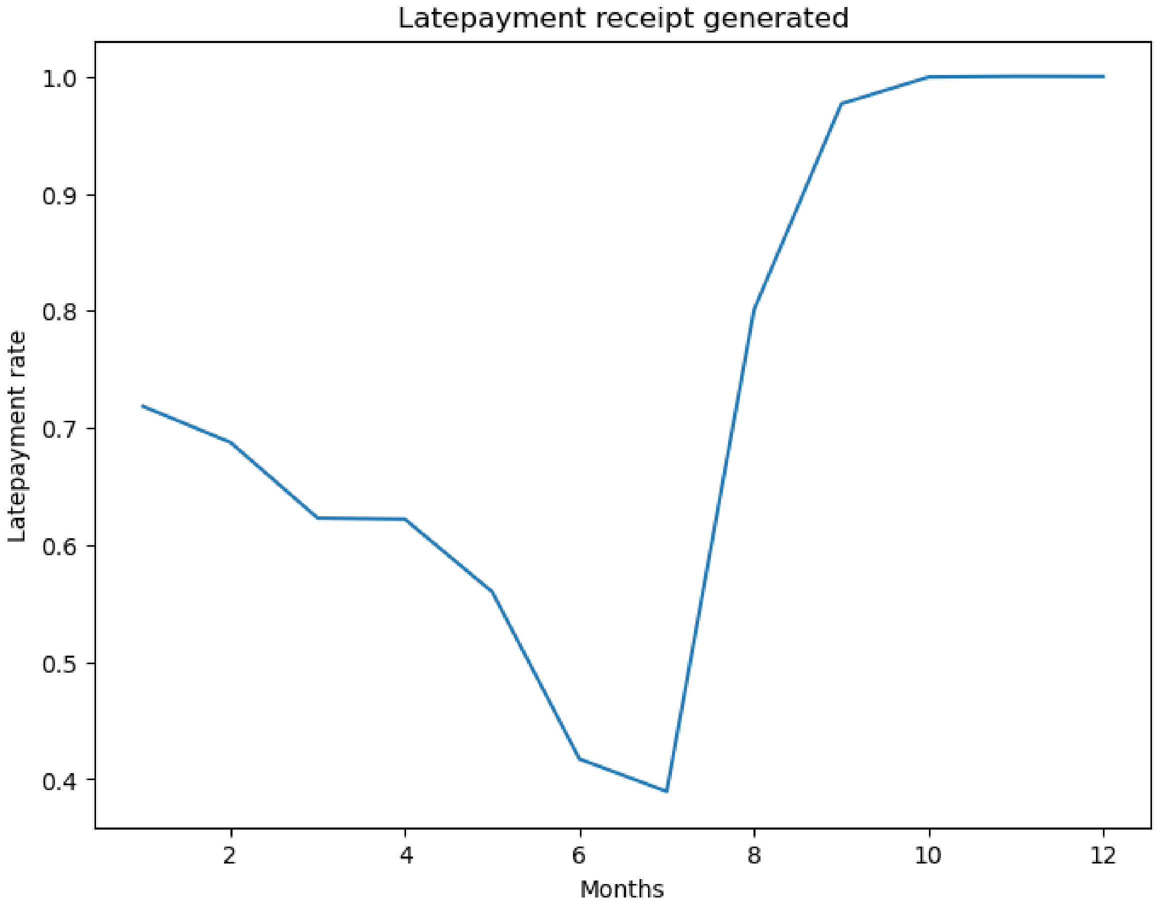


Bi-variate Analysis

Latepayment customers Vs INVOICE_CREATION_DATE

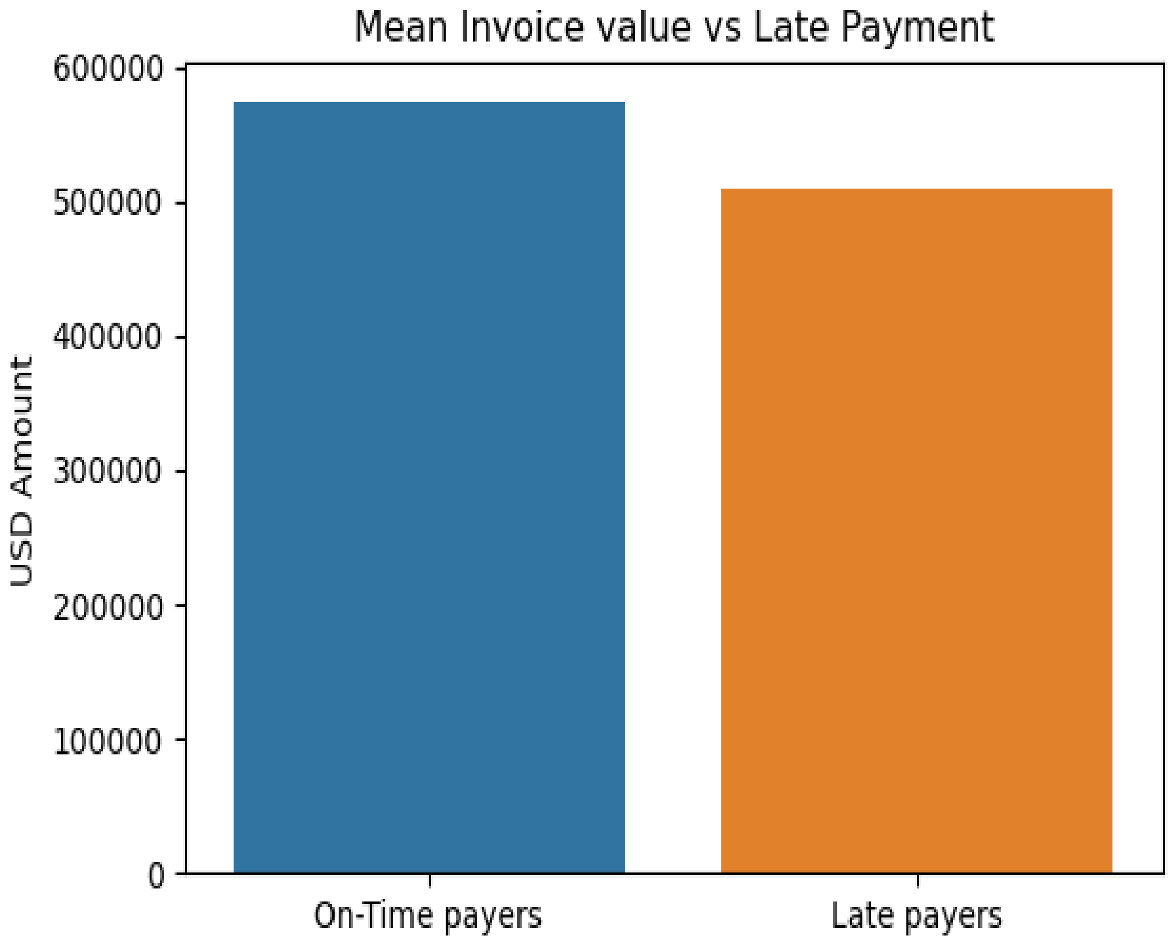


Latepayment customers Vs RECEIPT_DATE

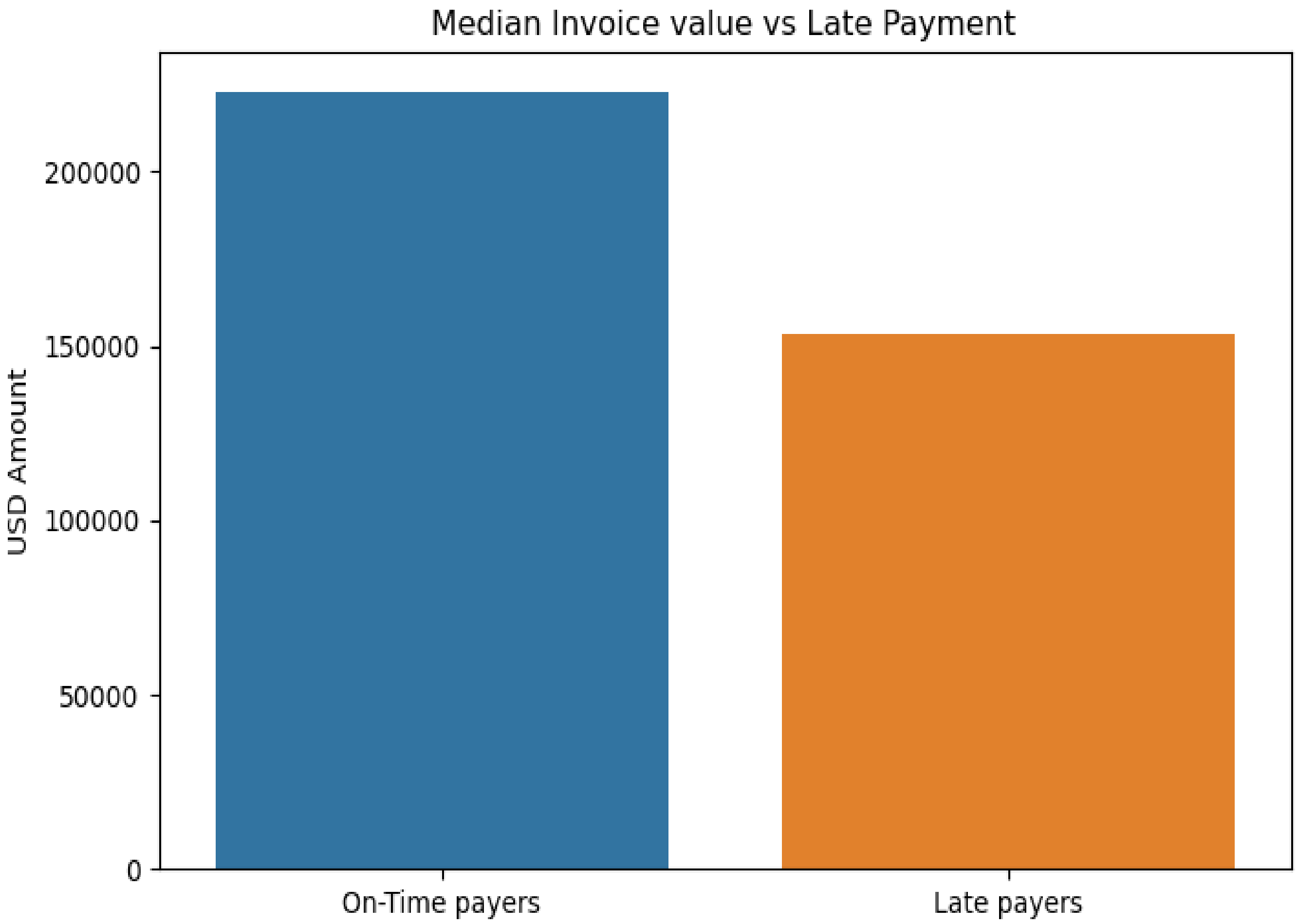


Bi-variate Analysis

Mean Invoice value Vs Late payment



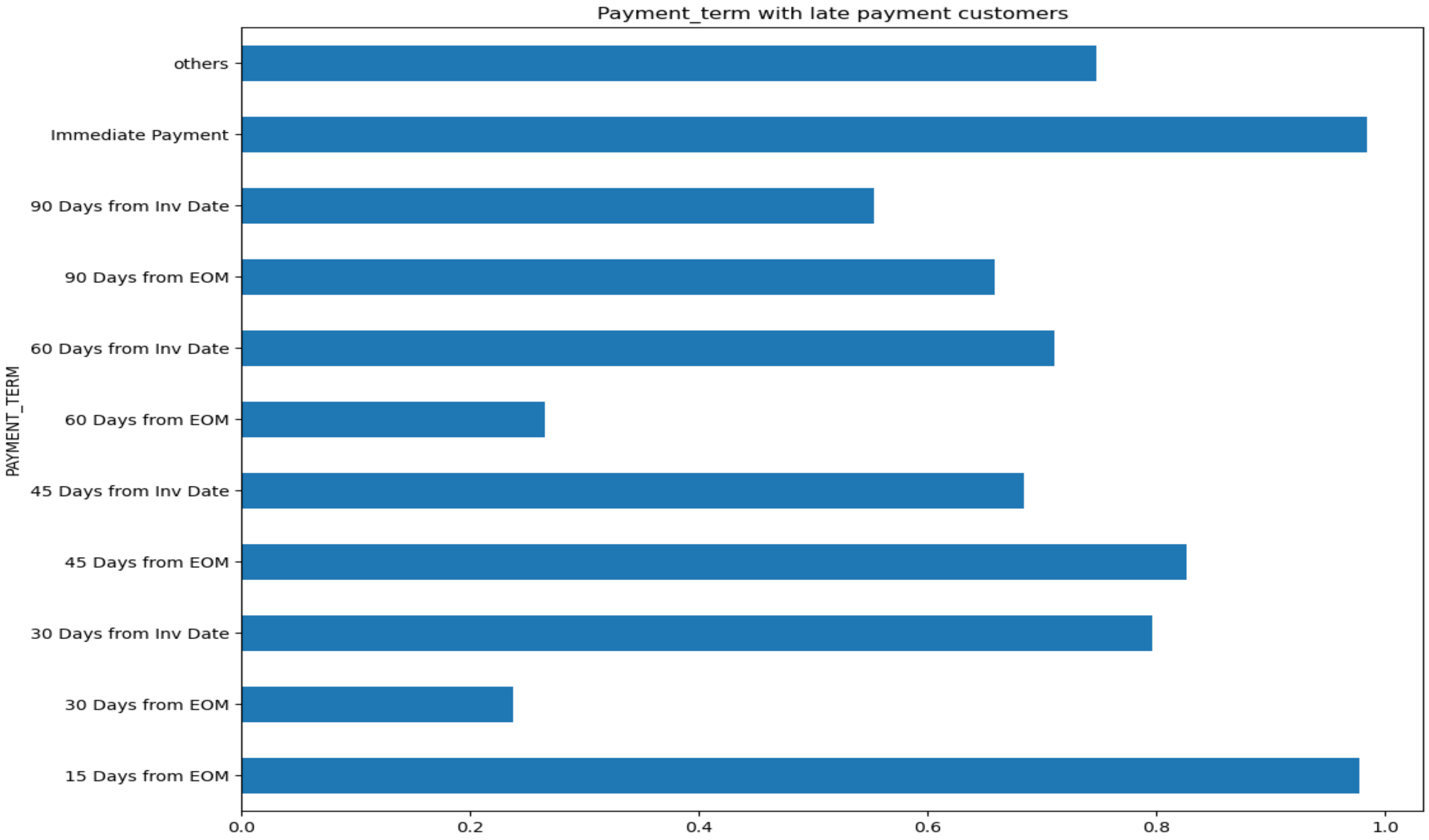
Median Invoice value Vs Late payment



Inferences from Uni-variate and Bi-variate analysis

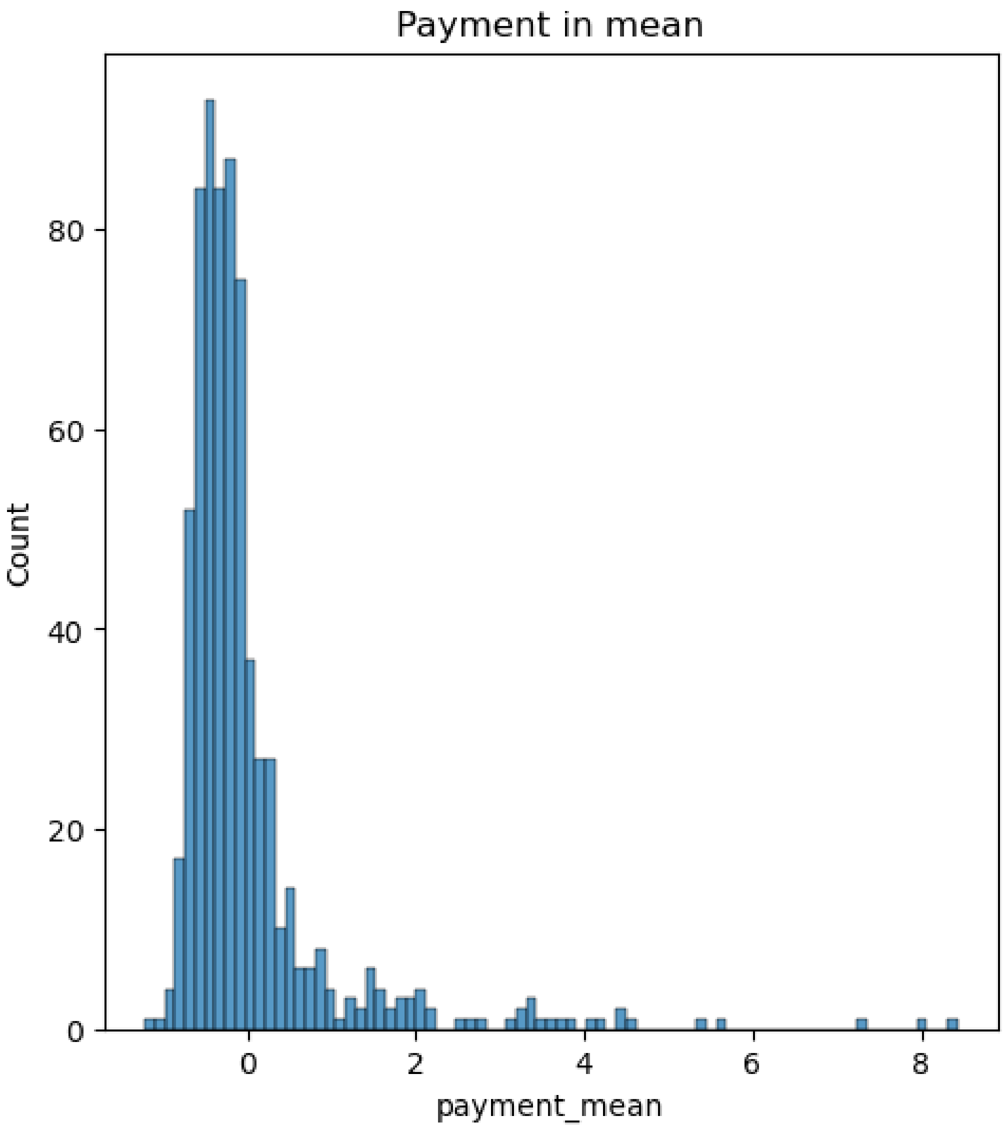
- 90% of vendors use "WIRE" as payment method
- The top customer/vendor is SEPH Corp
- CLASS has only one value "PMT" and this feature could be ignored while modeling
- Most of the vendors uses "SAR" as currency code
- In the month 7, the late payment rate is low and it get increased gradually till month 12
- In month 5, the Late payment invoice generated is less and it is peak in the month 8
- In the month 7, the receipt generated is less and gradually increases from month 7 to month 12
- The mean and median values of On-time payers are higher than the late payers
- The late payment rate is high through CM (credit memo) class
- Late payment rate is higher in "Goods"

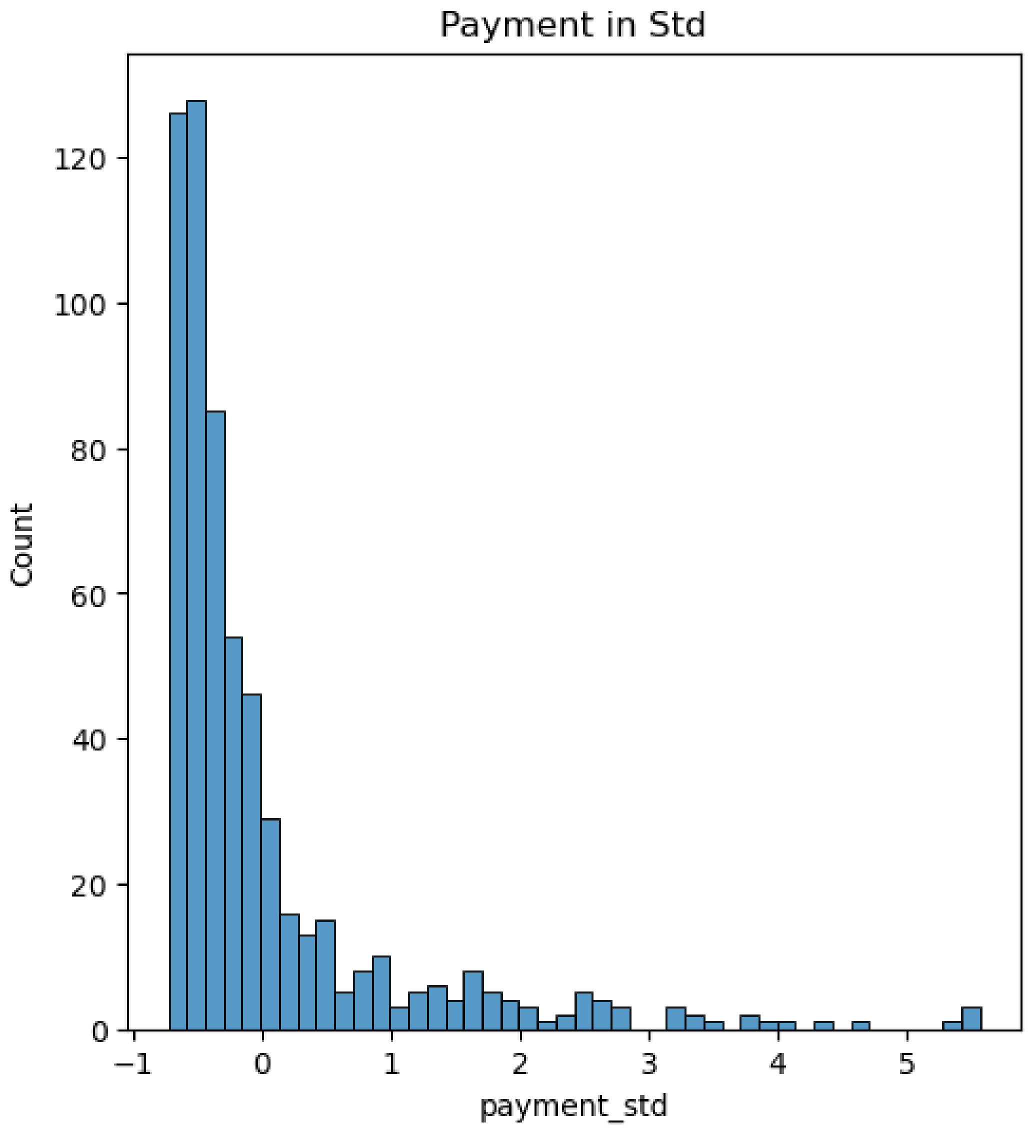
Feature Engineering on PAYMENT_TERM



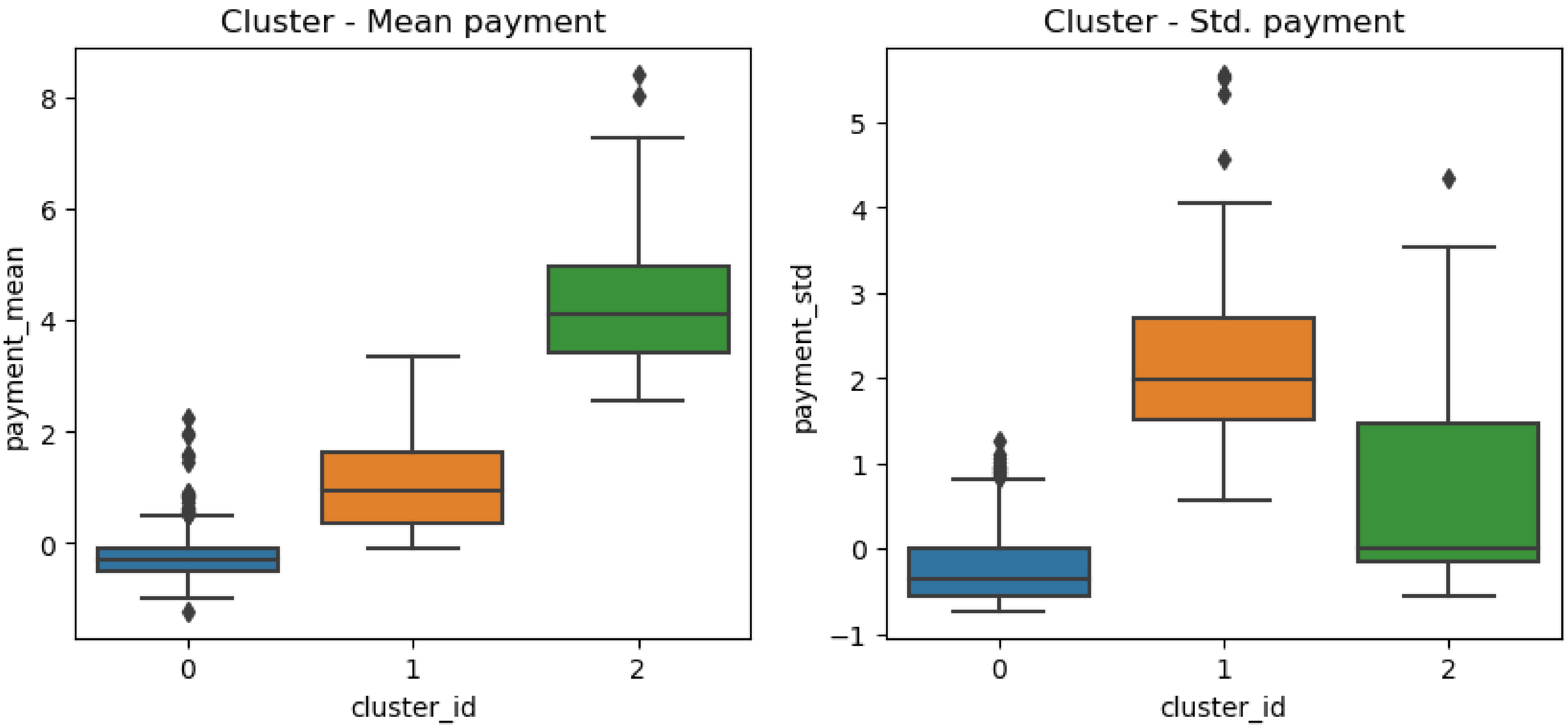
- **60 Days from Inv Date**
- **30 Days from Inv Date**
- **60 Days from EOM**
- **30 Days from EOM**
- **Immediate Payment**
- **15 Days from EOM**
- **90 Days from EOM**
- **45 Days from EOM**
- **others**
- **45 Days from Inv Date**
- **90 Days from Inv Date**

Segmentation on Customers

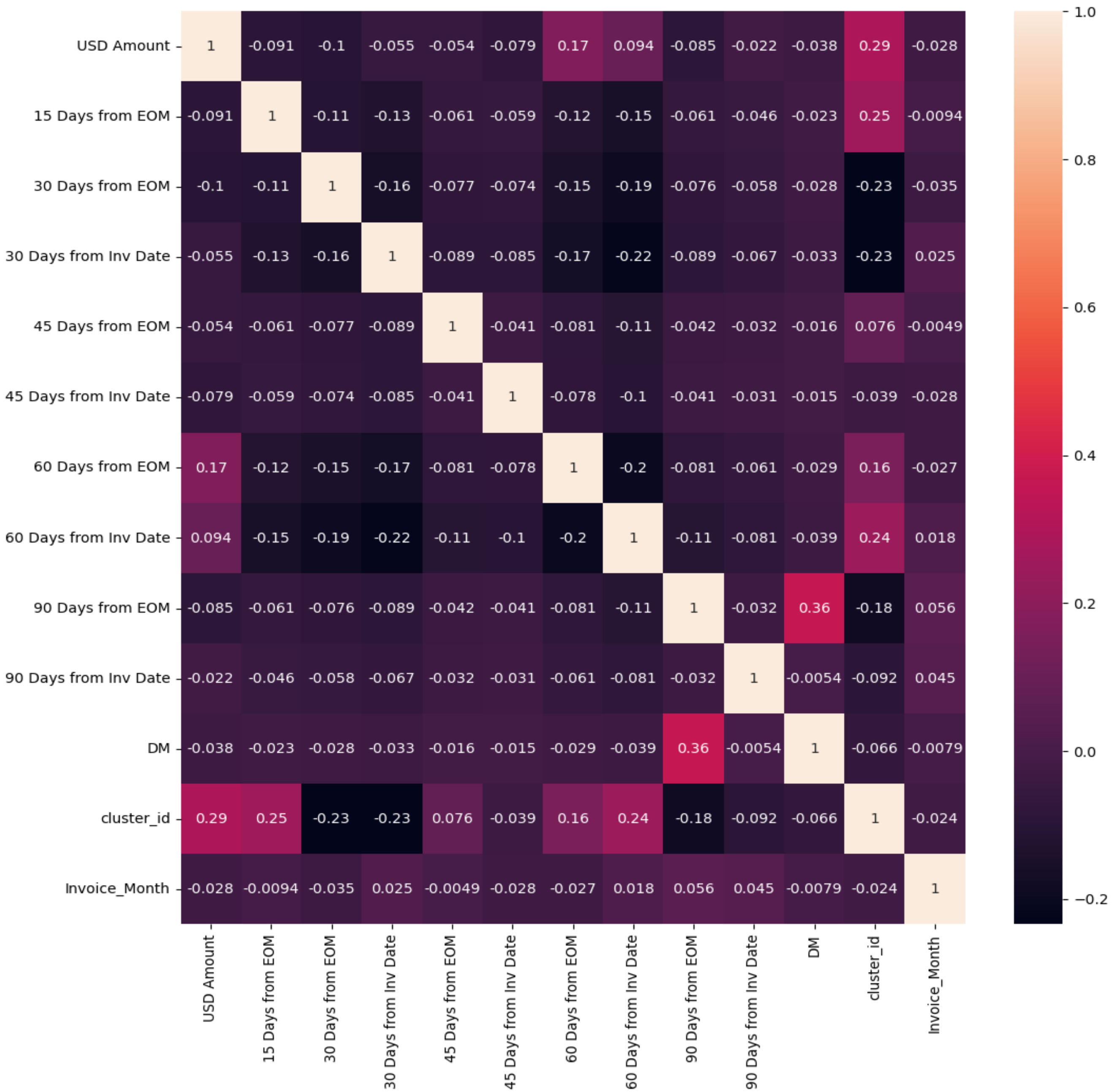




Clustering

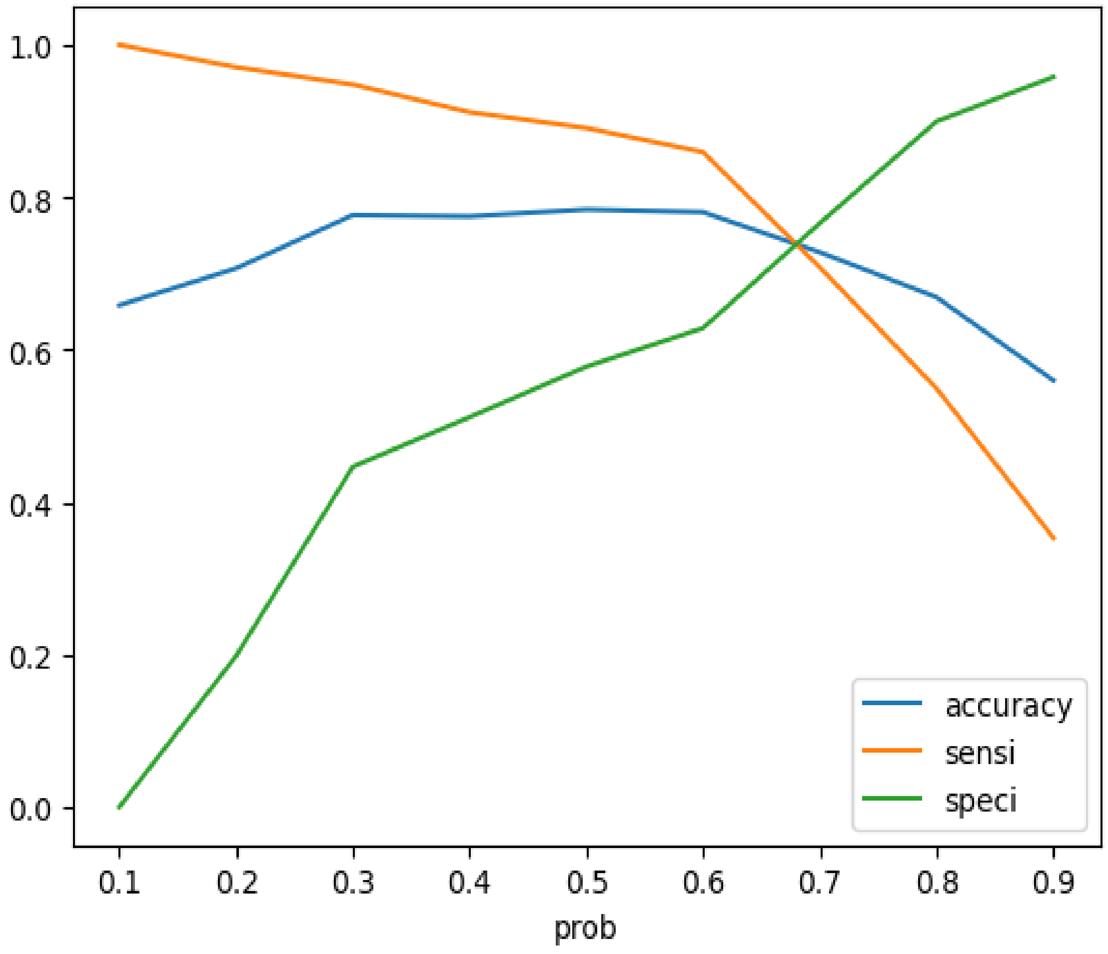

Cluster - Mean payment
payment_mean
8
6
4
2
0
0
1
2
cluster_id
Cluster - Std. payment
payment_std
5
4
3
2
1
0
−1
0
1
2
cluster_id

Heat map for Train dataset

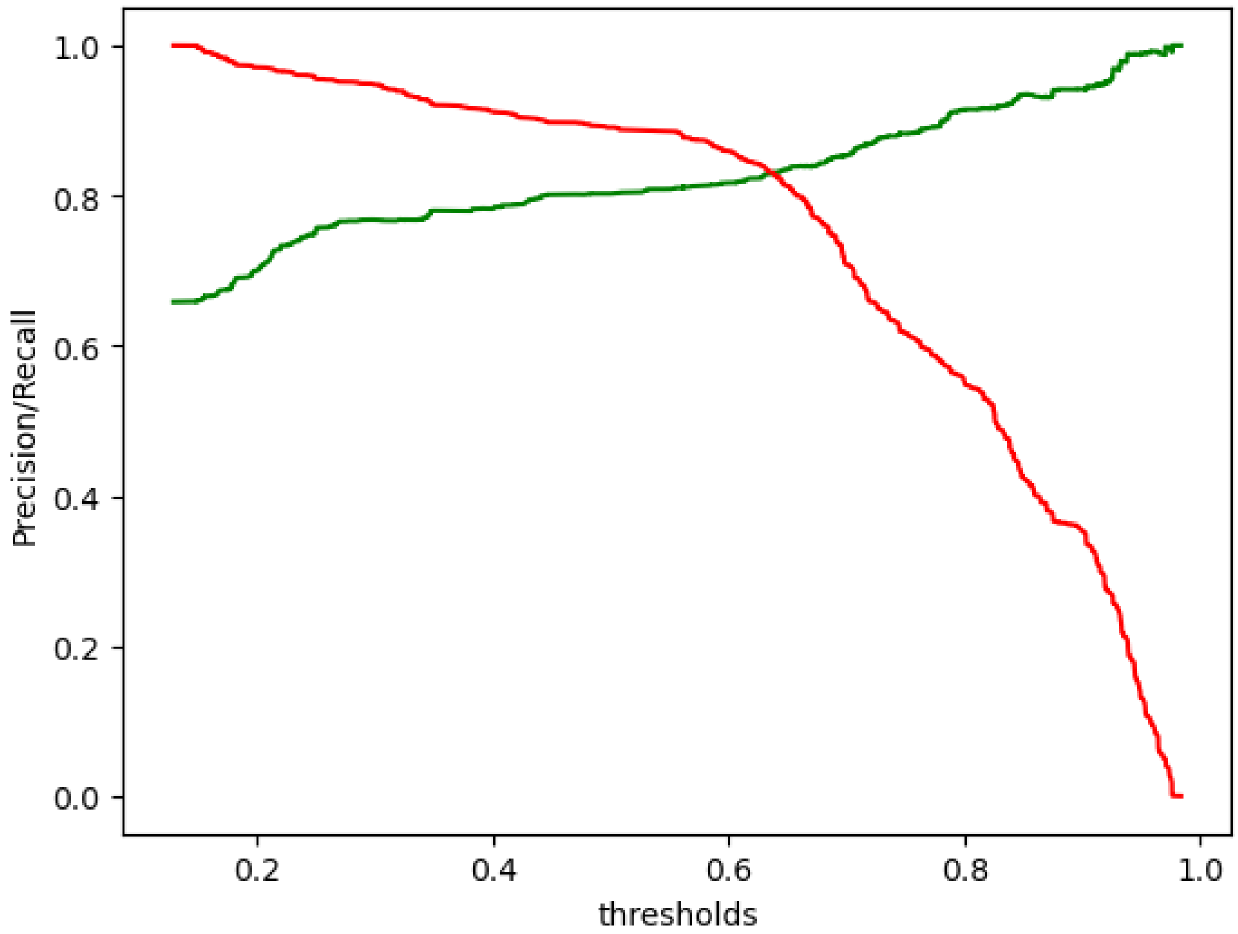


Model Building - Logistic Regression

Plotting accuracy, sensitivity and specificity

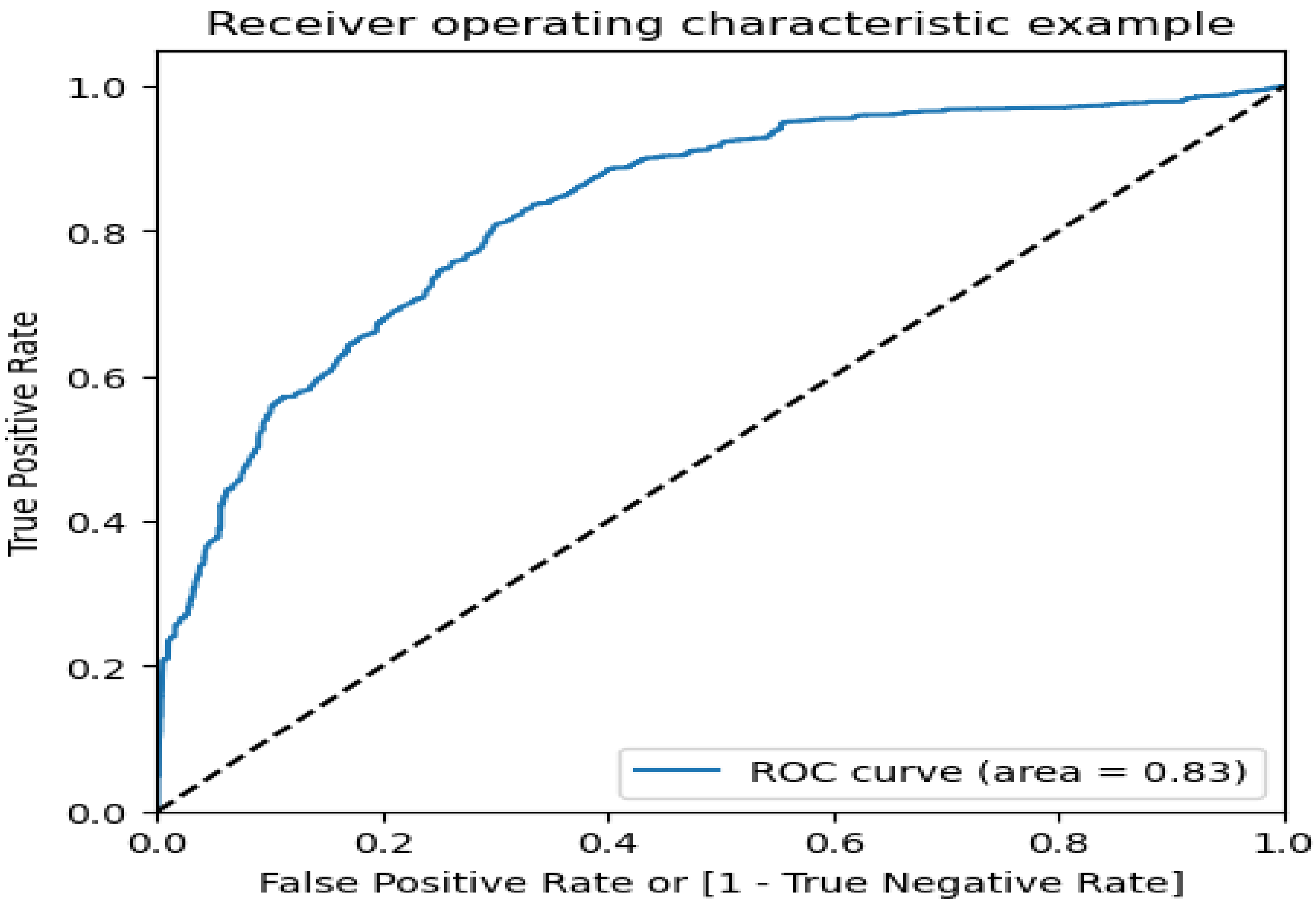


Precision – Recall Trade off curve



Model Building - Logistic Regression

Plotting ROC curve



Results

For Train dataset:

- **Accuracy : 74.76**
- **Sensitivity: 84.37**
- **Specificity: 75.69**

For test dataset:

- **Accuracy: 77.33**
- **Sensitivity: 83.05**
- **Specificity: 82.75**

Model Building – RandomForest classifier

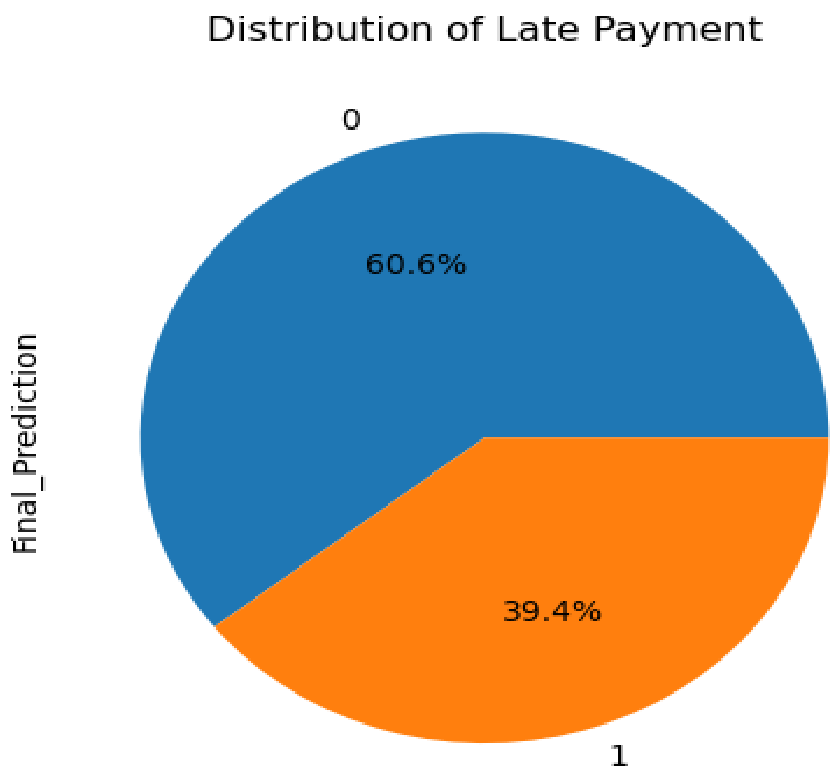
Evaluation metrics for Randomforest Classifier

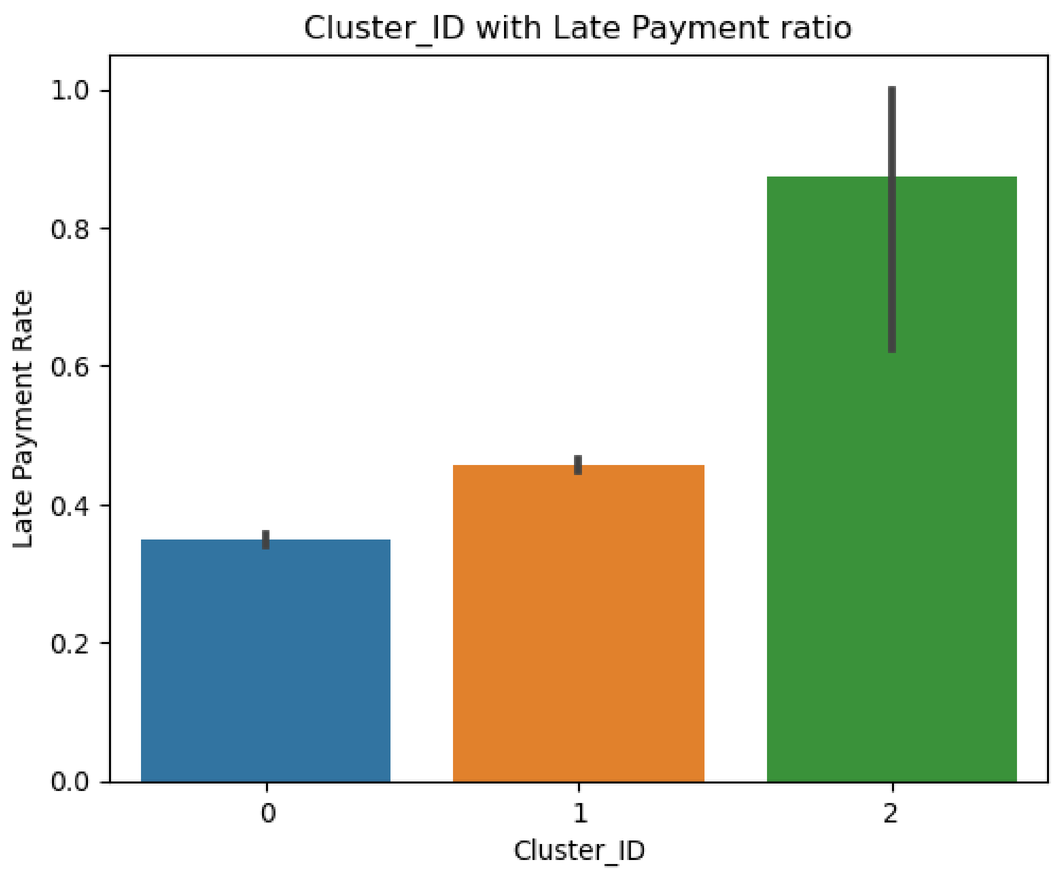
Classification Report	precision	recall	f1-score	support
0	0.90	0.81	0.85	9496
1	0.91	0.95	0.93	18686
accuracy			0.91	28182
macro avg	0.90	0.88	0.89	28182
weighted avg	0.91	0.91	0.90	28182

Feature Ranking

1. USD Amount (0.357)
2. 60 Days from EOM (0.174)
3. 30 Days from EOM (0.155)
4. Invoice_Month (0.138)
5. cluster_id (0.067)
6. Immediate Payment (0.035)
7. 15 Days from EOM (0.024)
8. 60 Days from Inv Date (0.011)
9. 90 Days from Inv Date (0.010)
10. 30 Days from Inv Date (0.009)
11. 90 Days from EOM (0.008)
12. 45 Days from EOM (0.006)
13. 45 Days from Inv Date (0.004)
14. INV (0.001)
15. DM (0.001)
16. CM (0.000)

Open_Invoice dataset for prediction





- 60.6% payments of negative aged values in the open invoice data indicates that due date does not crossed.
- Cluster id 2 has higher late payment rate

Cluster 0 - early invoice payment, Cluster 1 - medium invoice payment, Cluster 2 - prolonged invoice payment

Results

Metrics	Logistic Regression		Random Forest
	Train dataset	Test dataset	
Accuracy	74.91	77	92.41
Precision	84.28	82.67	90.0
Recall	76.06	82.63	81.0

1. The metrics of RandomForest are higher than the LogisticRegression.
2. Cluster id 2 has higher late payment rate

   Where

   Cluster 0 - early invoice payment

   Cluster 1 - medium invoice payment

   Cluster 2 - prolonged invoice payment

Results

3. Features that corresponds for delayed payments

1.USD Amount (0.357)
2. 60 Days from EOM (0.174)
3. 30 Days from EOM (0.155)
4. Invoice_Month (0.138)
5. cluster_id (0.067)
6. Immediate Payment (0.035)
7. 15 Days from EOM (0.024)
8. 60 Days from Inv Date (0.011)
9. 90 Days from Inv Date (0.010)
10. 30 Days from Inv Date (0.009)

4. Top 15 customers who makes payment with more delay

Customer_Name	Delayed_Payment	Total_Payments	Delay%
ALSU Corp	7	7	100.0
FEME Corp	5	5	100.0
SUND Corp	4	4	100.0
LVMH Corp	4	4	100.0
MAYC Corp	3	3	100.0
MUOS Corp	3	3	100.0
ZAIN Corp	3	3	100.0
VENI Corp	3	3	100.0
MILK Corp	3	3	100.0
ROVE Corp	3	3	100.0
AL Y Corp	2	2	100.0
MAWA Corp	2	2	100.0
X TR Corp	2	2	100.0
DAMA Corp	2	2	100.0
CTC Corp	2	2	100.0

Recommendations

- Create a prioritized list of customers based on the model's predictions and allocate more resources to follow up with these high-risk accounts.The customers like ALSU Corp, FEME Corp, SUND Corp, LVMH Corp, MAYC Corp, MUOS Corp, ZAIN Corp, VENI Corp, MILK Corp, ROVE Corp, AL Y Corp, MAWA Corp, X TR Corp, DAMA Corp, CTC Corp should be prioritized because these customer have higher delay payment

- Proactive Communication: Send reminder emails or make phone calls a few days before the due date. Highlight the payment terms and any late fees that might apply.

- Develop customized payment plans or installment options for customers who might struggle with large one-time payments. This can help in securing at least partial payments on time.

- The late payment rate is higher in CM (credit memo) class. So, the company should check the vendors who pay through credit. If delay is still high, they could introduce penalty.

- The company should check the customers whose invoice_type is physical goods since these customers have higher late payment rate

- Customers were segregated based on their past payment patterns (customer segmentation) and we can see that customers in Cluster 2 need to be focussed on as their probability of defaulting on making timely payments is higher.

- The payment features which are responsible for delayed payments are 60 Days from EOM, 30 Days from EOM, 15 Days from EOM, 60 Days from Inv Date, 90 Days from Inv Date, 30 Days from Inv Date. So, the company should check and may give warnings from these payment terms.

- Incentivize Timely Payments-Offer loyalty discounts for customers who pay their invoices before the due date to motivate timely payments.

- Late payments are more likely on SAR, AED currencies followed by USD. This group can incentivized to pay in a timely manner.

Thank you!