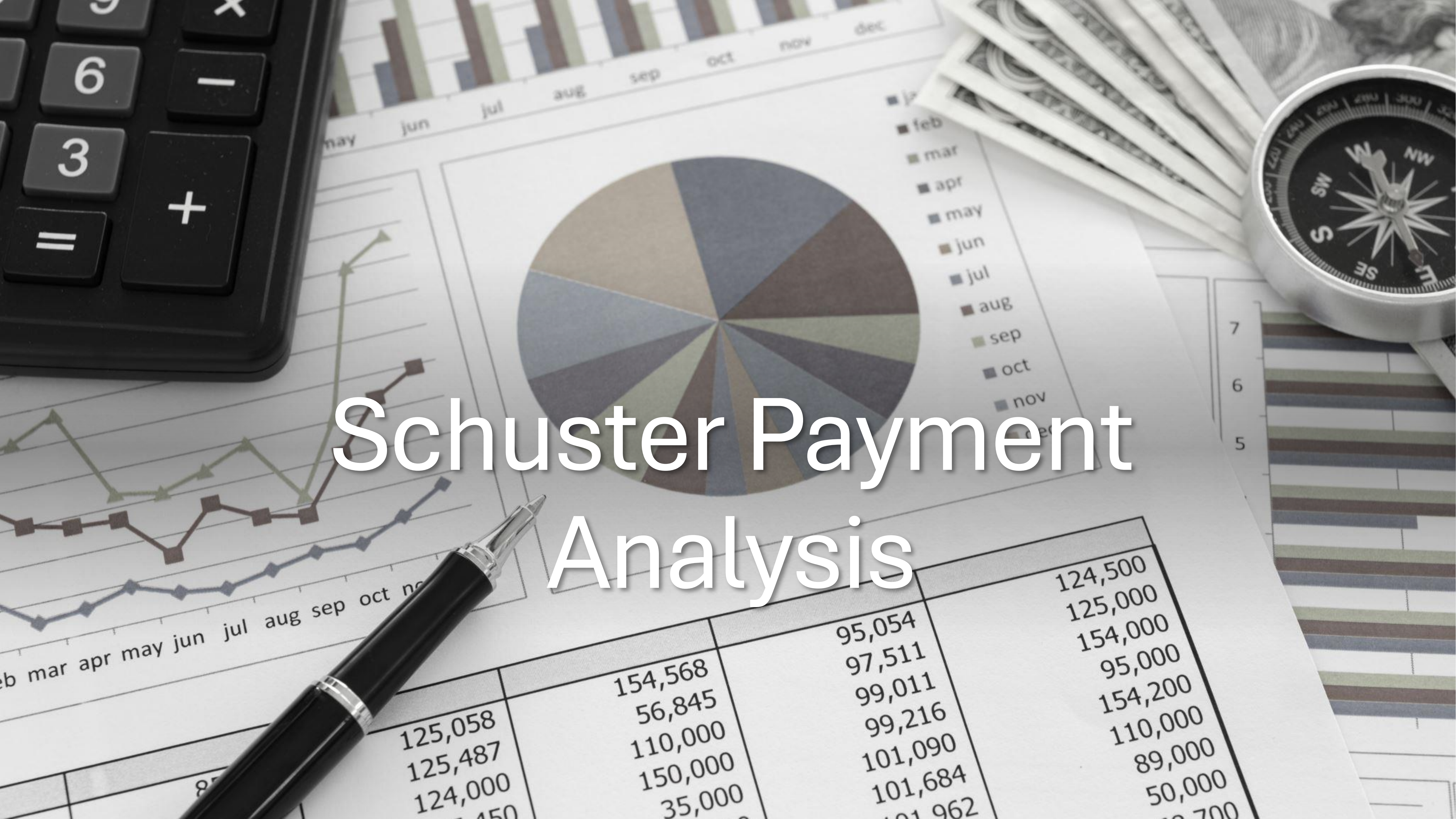
Schuster Payment Analysis



Content

Section1

Data Handling

Problem

- **Background**

Schuster is a multinational retail company specializing in sports goods and accessories. It regularly deals with numerous vendors on credit terms. However, inconsistencies in payment compliance by these vendors lead to inefficiencies.

- **Objective:**

To develop a predictive model that estimates the likelihood of late payments for open invoices. This will aid in prioritising collector activities, enhancing payment compliance, and sustaining healthy business relationships.

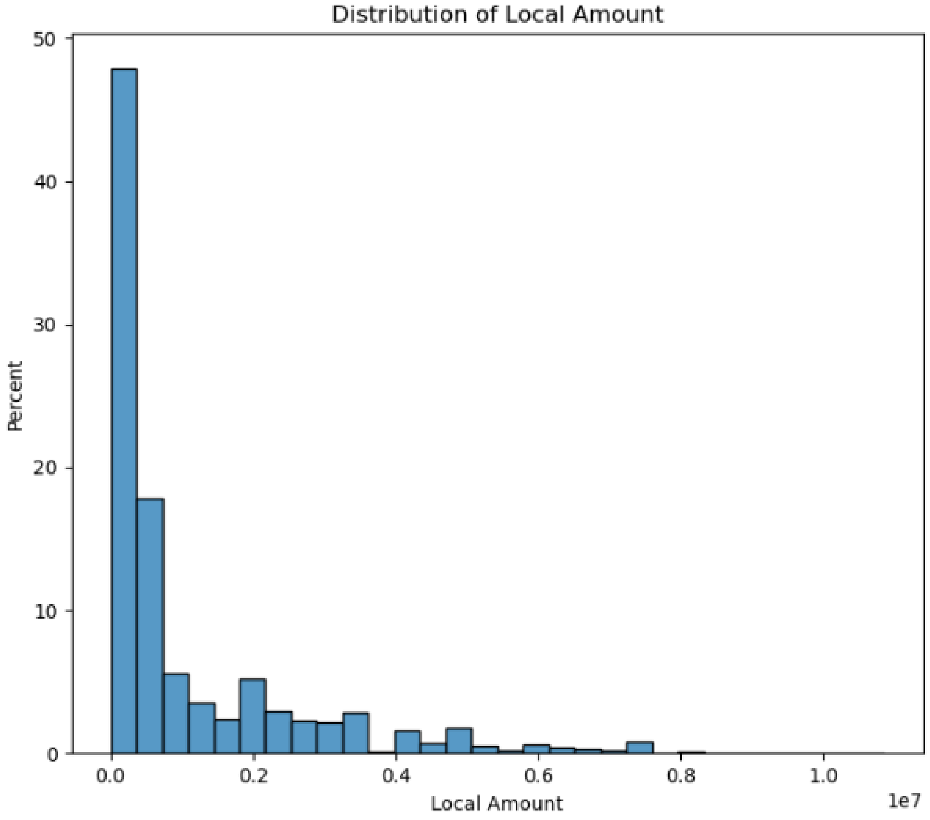
Received Payment Data			
RECEIPT_METHOD	In which method payments have been made	USD AMOUNT	Invoice Value converted to USD
CUSTOMER_NAME	Name of the customer/vendor	INVOICE_ALLOCATED	Invoice number that has been allocated to a particular vendor
CUSTOMER_NUMBER	Customer's unique identity number	INVOICE_CREATION_DATE	The date on which the invoice was created
RECEIPT_DOC_NO	Reference number of the payment receipt	DUE_DATE	The date by which the payment was to be made
RECEIPT_DATE	The date in which the payment has been made	PAYMENT_TERM	Days given to the vendor/customer for making the payments
CLASS	As the payment against these invoices have already been received so Transaction Class as PMT (short for Payment) assigned	INVOICE_CLASS	Three types of Invoice classes - Credit Memo or Credit Note (CM), Debit Memo or Debit Note (DM) or Invoice (INV)
CURRENCY_CODE	Currency used for the payment	INVOICE_CURRENCY_CODE	Currency code as per the invoice generated
Local Amount	Invoice value in local currency	INVOICE_TYPE	Invoice created for physical goods or services (non-goods)

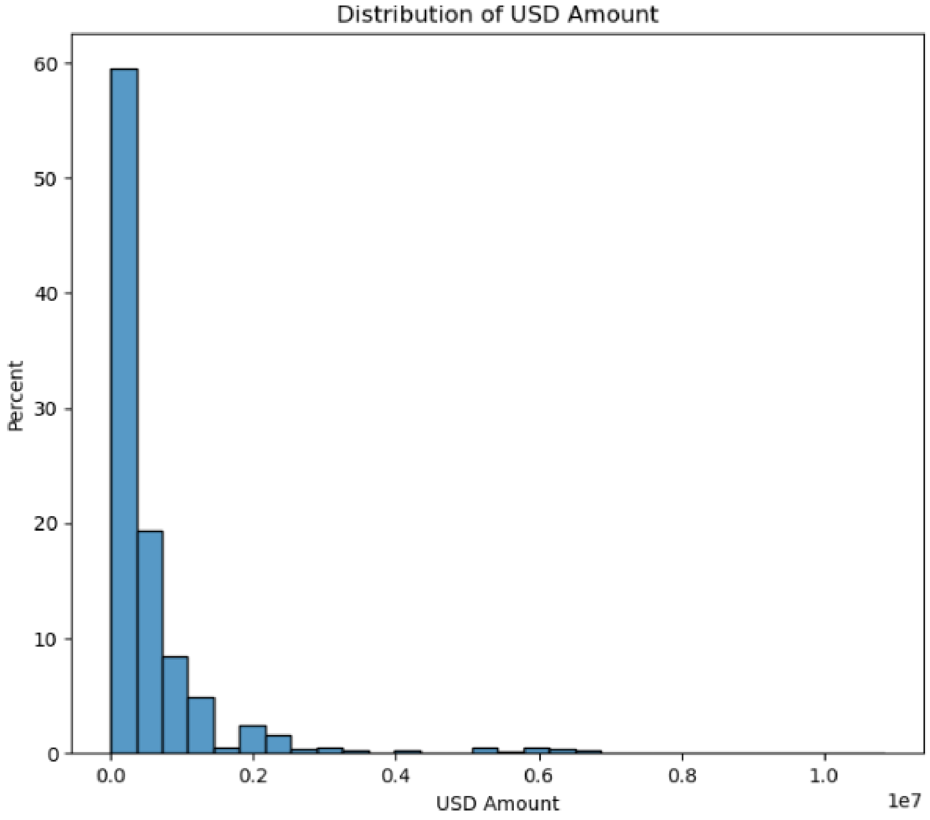
Data Description

- **Received Payments Data:** Contains historical transactions with details including payment terms, dates, amounts, and customer details. (93937 entries, total 16 columns)

Exploratory Data Analysis (EDA)

- The distribution of both local and USD amounts is skewed, with a concentration of values at the lower end and a long tail towards higher values (Received_payments_Data)



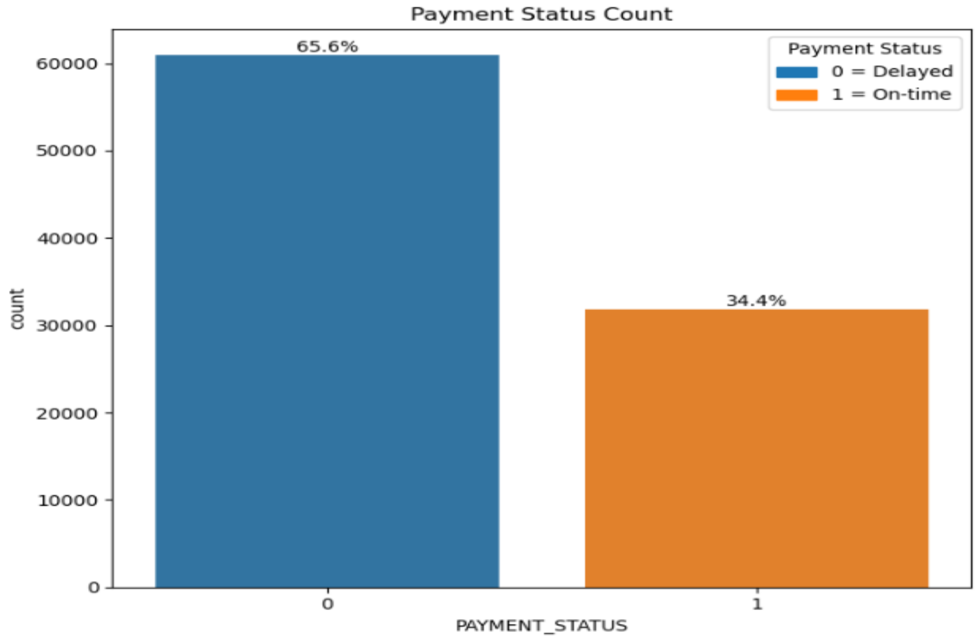


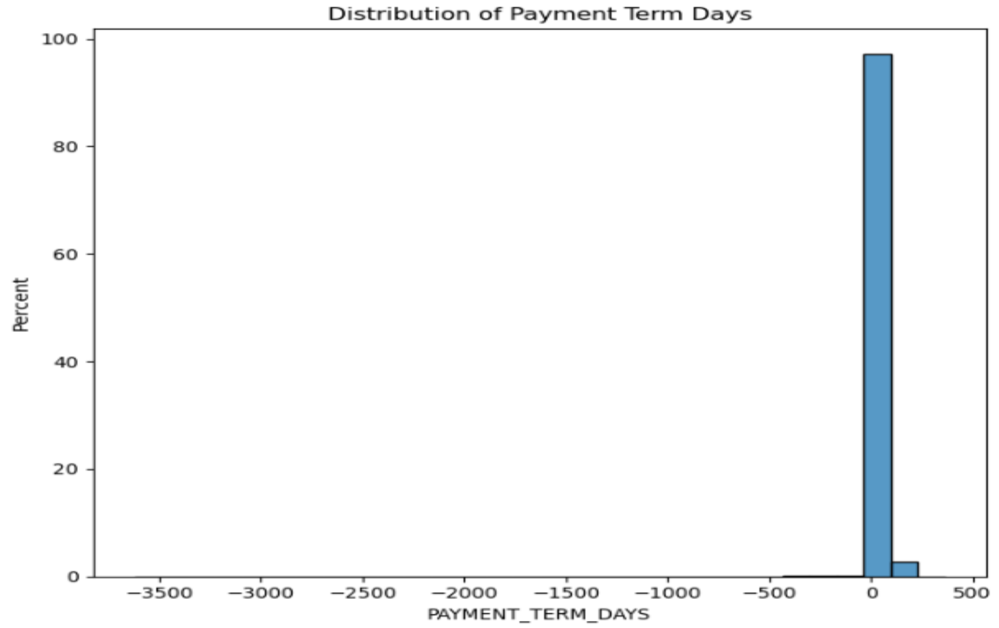
Exploratory Data Analysis (EDA)

- The average payment term is around 42 days, with a standard deviation of 61 days, indicating a wide range of payment terms.

- The distribution shows that most terms are concentrated around 30 to 60 days. Negative values might represent data entry errors or special cases where due dates are set before invoice creation dates.

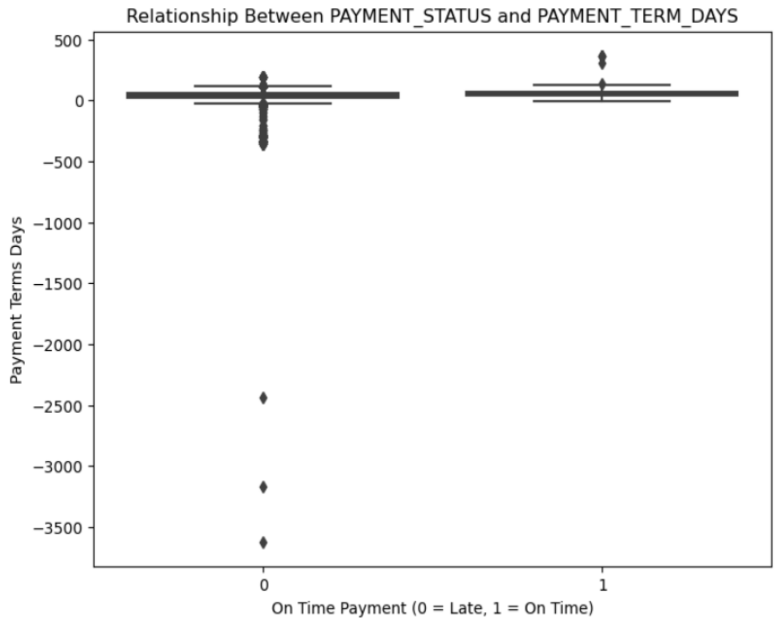
- Approximately 66% of the payments are late (represented by 1). Around 34% of the payments are on time (represented by 0).

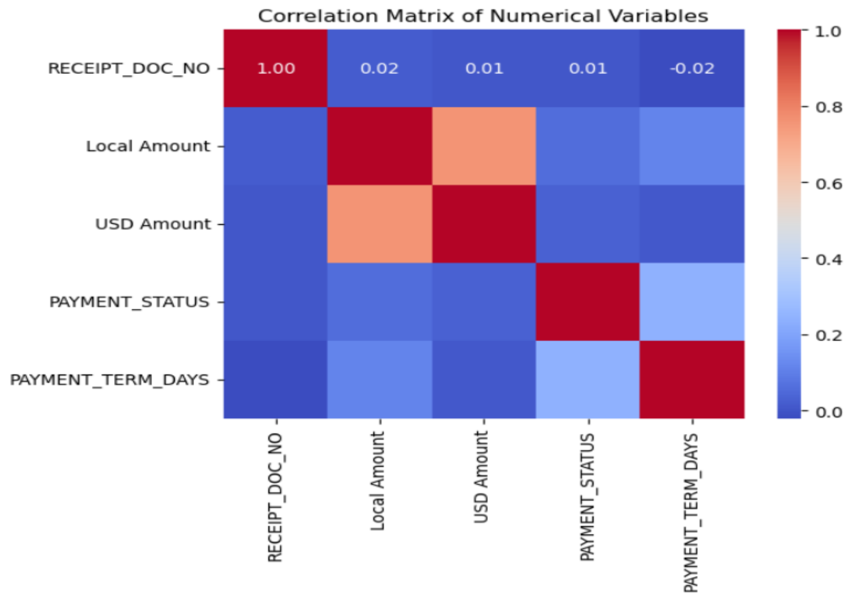




Exploratory Data Analysis (EDA)

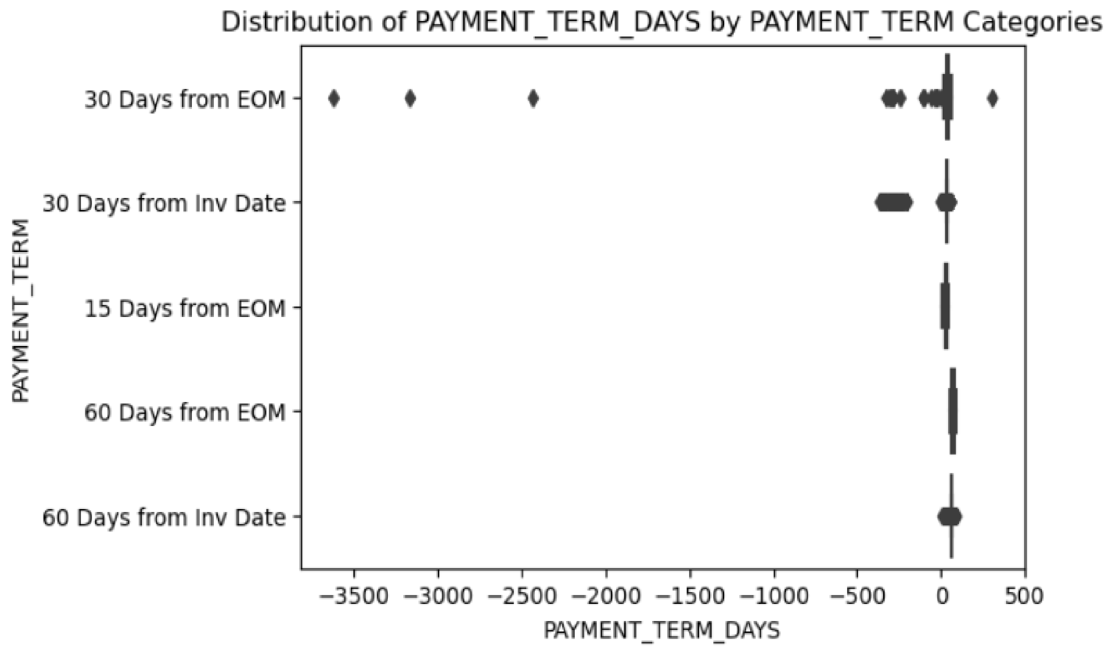
- The boxplot reveals that the distribution of payment terms days is quite similar for both on-time and late payments. However, it looks like on-time payments have a slightly higher median and a wider range in payment terms compared to late payments, suggesting that longer payment terms might be slightly more manageable for on-time payments.
- The heatmap of the correlation matrix provides insights into the relationships between numerical features: PAYMENT_STATUS and PAYMENT_TERM_DAYS show some degree of correlation, suggesting that as the number of days from invoice creation to due date increases, the likelihood of a payment being late also increases, though the correlation is not very strong.
- Local Amount and USD Amount are highly correlated, as expected since these are just currency conversions of the same underlying values. The other variables show less significant correlations with PAYMENT_STATUS, indicating that while useful, they are not as directly predictive of late payments as PAYMENT_TERM_DAYS.





Exploratory Data Analysis (EDA)

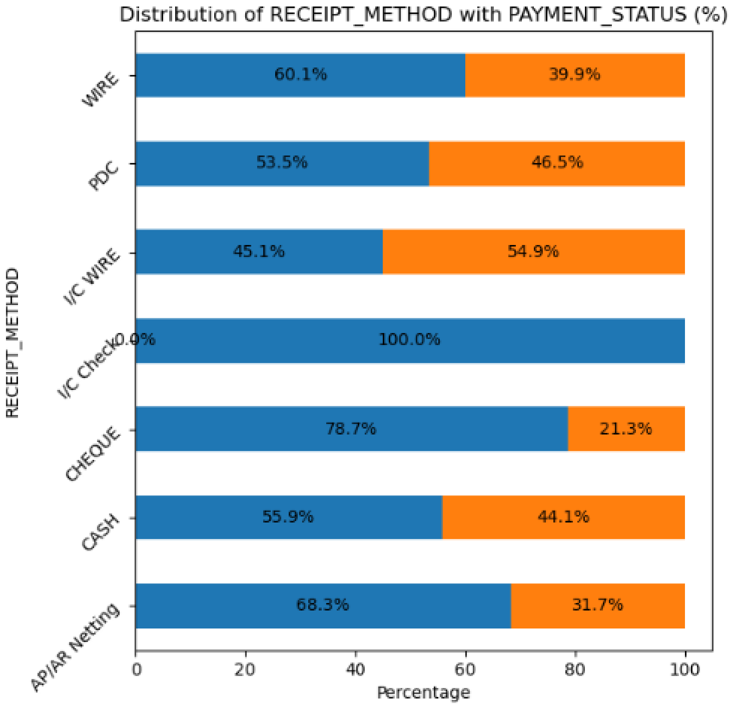
- Count: There are 3,258 entries with negative payment term days. RECEIPT_METHOD: Most payments were made via wire transfer. CUSTOMER_NAME: A diverse set of customers is involved, with 'YOUG Corp' being the most frequent. PAYMENT_TERM_DAYS: Values range from -1 to -3,622 days, with a mean of about -19.43 days. This suggests that invoices are often issued after the due date, or due dates are set retroactively. IS_LATE: Approximately 98.34% of these payments are classified as late, which is quite high and may indicate systemic issues with how due dates are set or invoices are processed.
- Distribution of Payment Terms: 'Immediate Payment' is the most common payment term associated with these entries, which might be misaligned with actual payment processes if the terms are not adhered to. Currencies: SAR (Saudi Riyal) appears most frequently, suggesting a regional pattern or specific business operations impacting these transactions.
- A significant number of invoices with negative payment term days and a high rate of late payments among them suggests either a procedural issue in how payment terms are set or in how transactions are recorded. This could be a critical area for improving business processes or data management practices to enhance payment efficiencies.

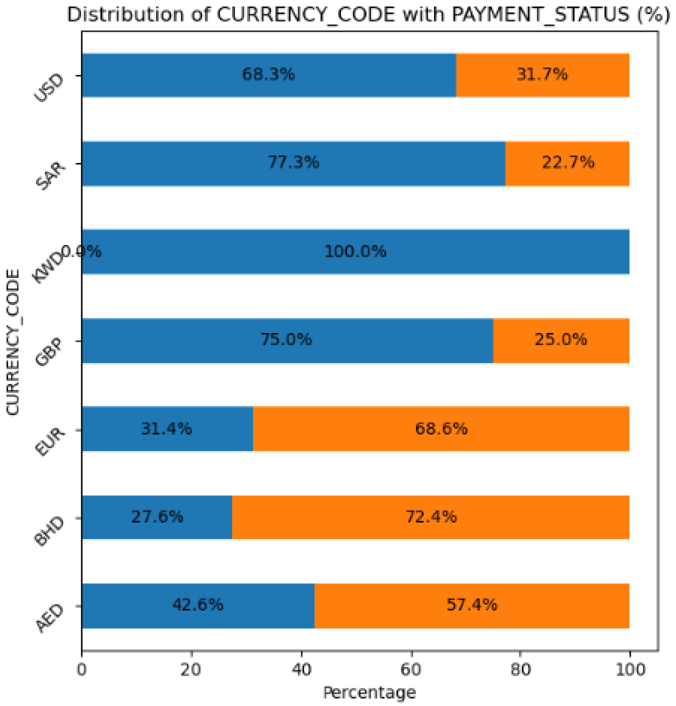


Exploratory Data Analysis (EDA)

- **RECEIPT_METHOD:**
The method of receipt seems to show some variation in payment timeliness. Certain methods might be more conducive to on-time payments, potentially due to the efficiency or ease of processing.

- **CURRENCY_CODE:**
Payment timeliness also appears to vary with the currency, which could reflect differences in banking processes or economic factors specific to regions using different currencies.

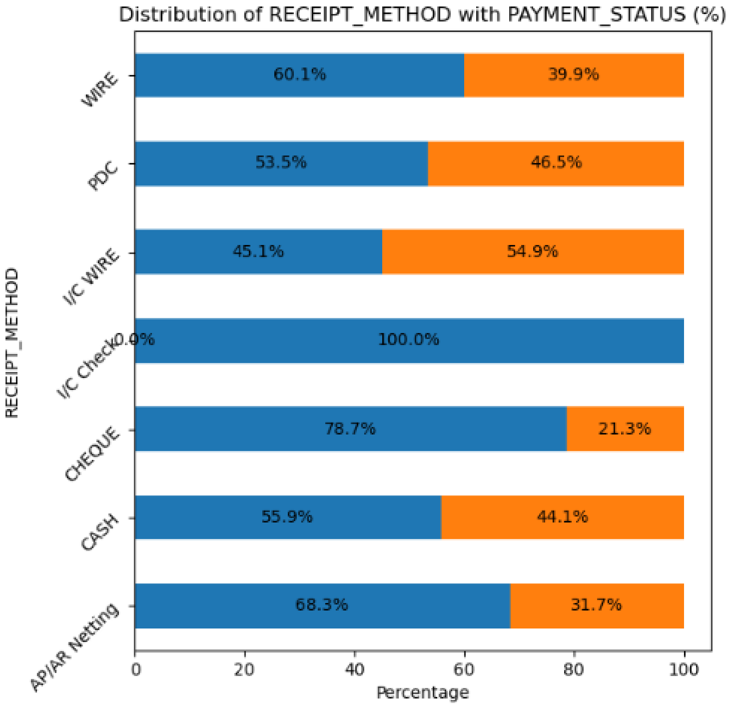


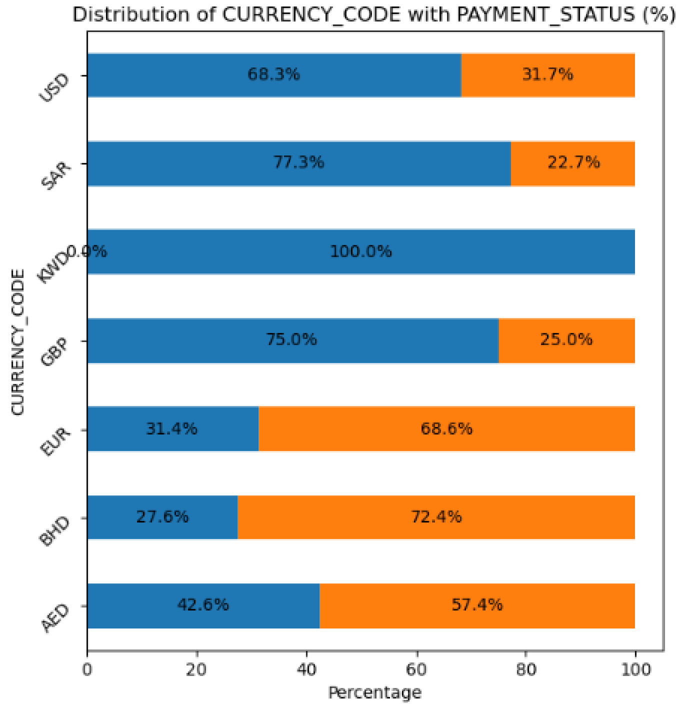


Exploratory Data Analysis (EDA)

- **INVOICE_CLASS:**
- Different classes of invoices show distinct patterns in their payment statuses, suggesting that the nature of the invoice (like goods, services, etc.) impacts payment behaviour.

- **INVOICE_TYPE:**
- Similarly, the type of invoice influences whether payments are made on time or are delayed, highlighting the importance of the transaction context.

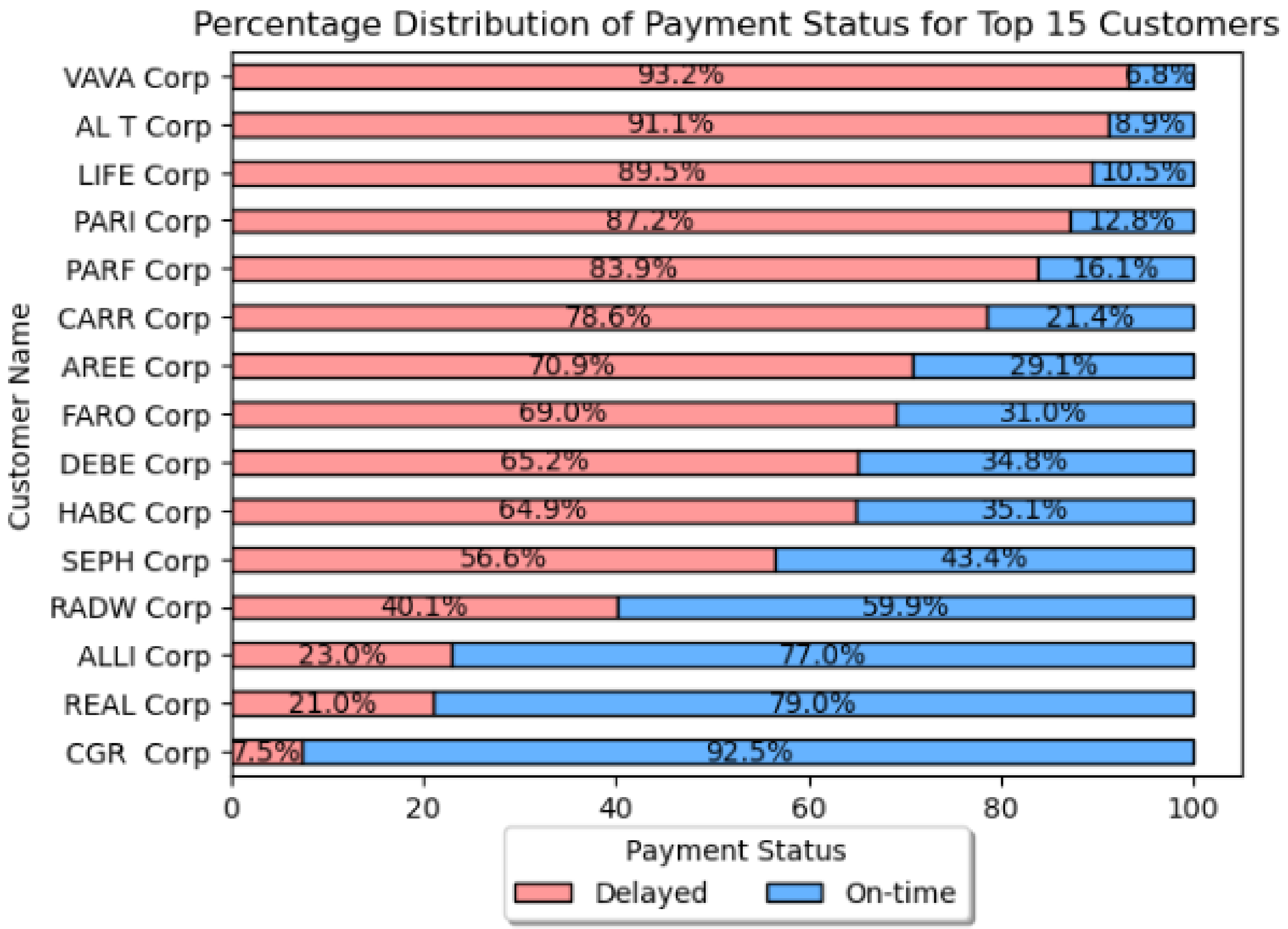




Exploratory Data Analysis (EDA)

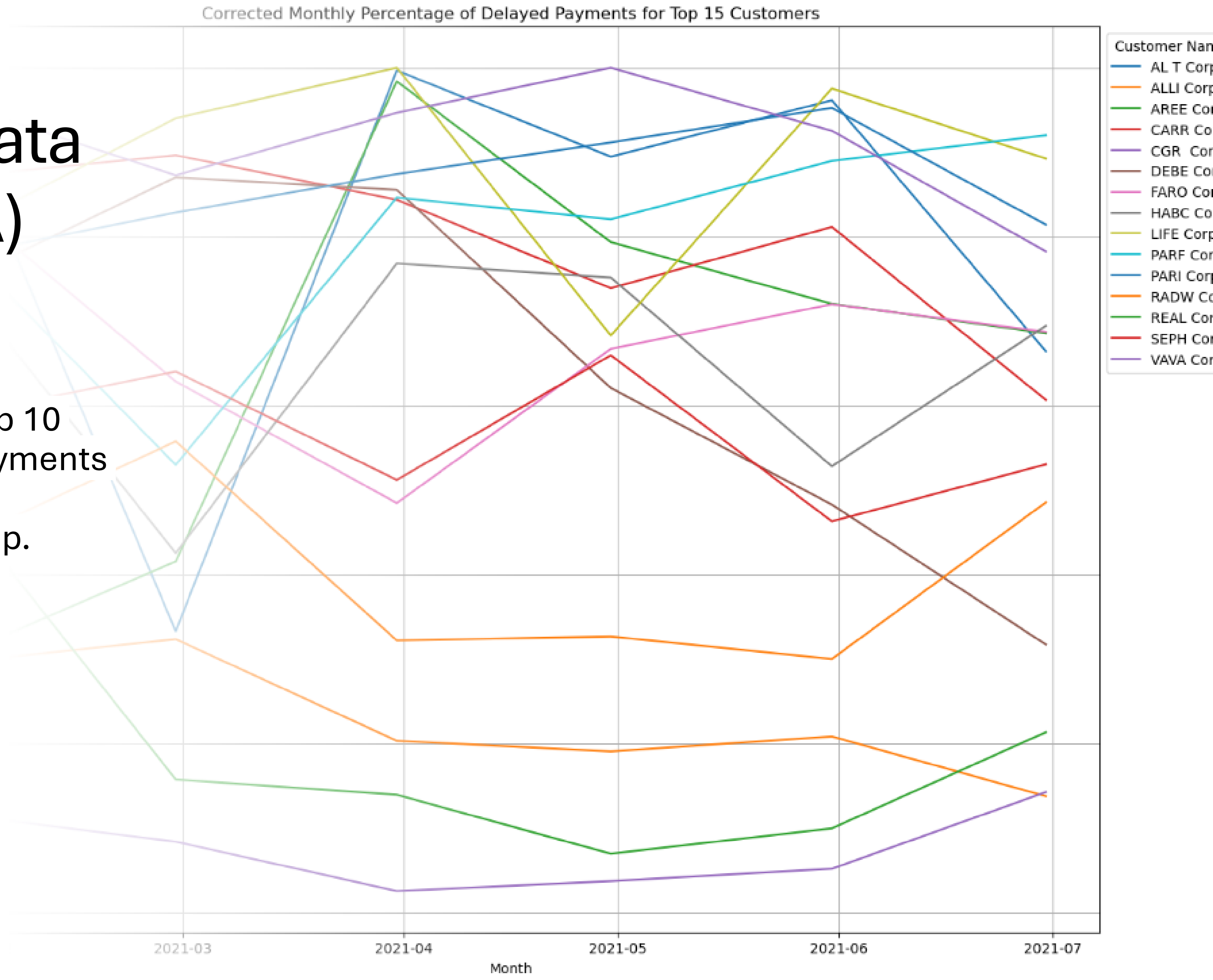
- IVAVA Corp has the highest % of Delayed payments among the top 15 customers with dealyed payments.

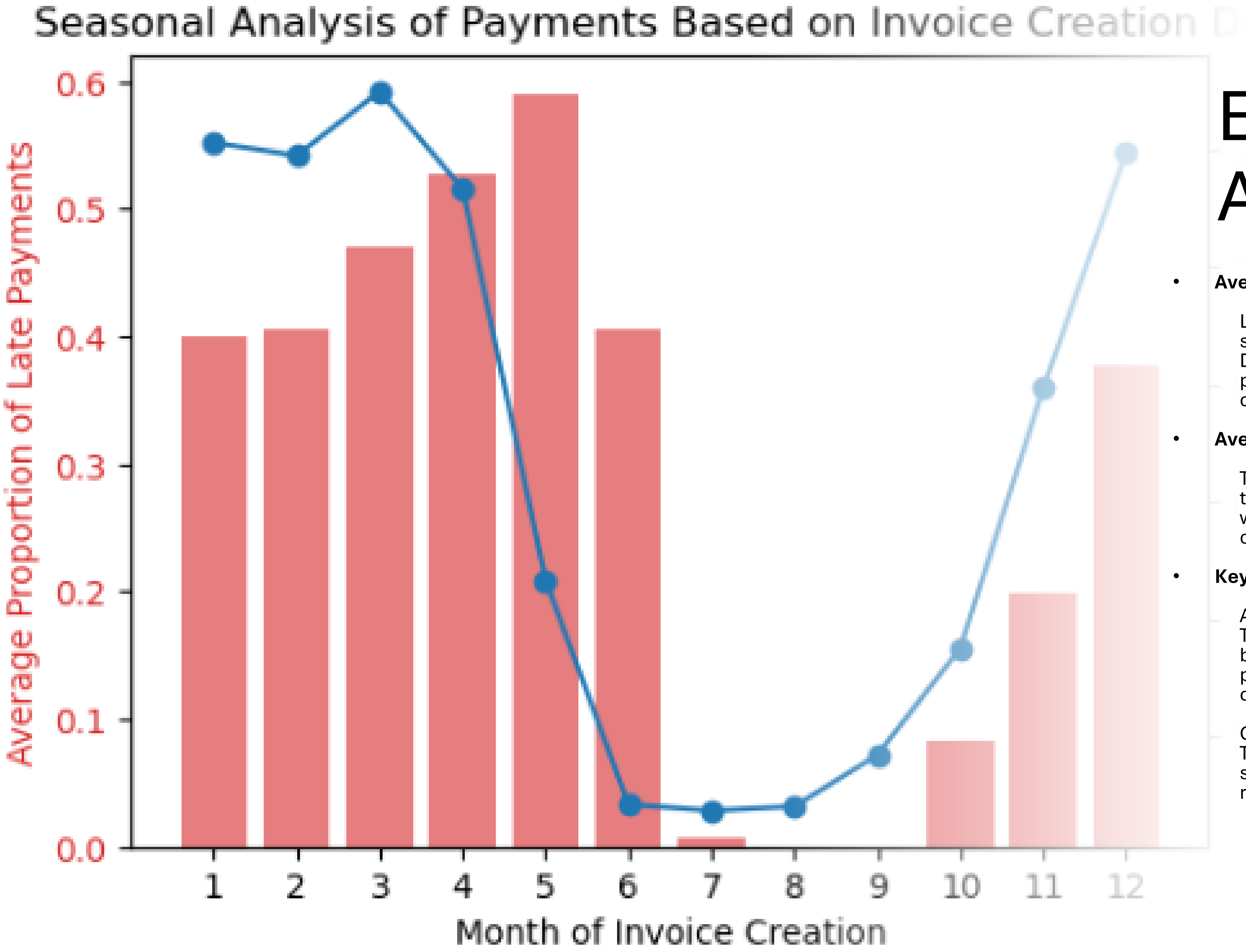
- CGR Group is the lowest % among the top 15 customers with dealyed payments.



Exploratory Data Analysis (EDA)

- It is observed for the Top 10 customers with late payments – there is no consistent pattern of delay as group.





Exploratory Data Analysis (EDA)

- **Average Proportion of Late Payments:**

  Late Payments (Red Bars): There is noticeable seasonality in late payments based on Invoice Creation Date. Certain months show higher averages of late payments, potentially indicating when billing practices or financial pressures affect payment behaviours.

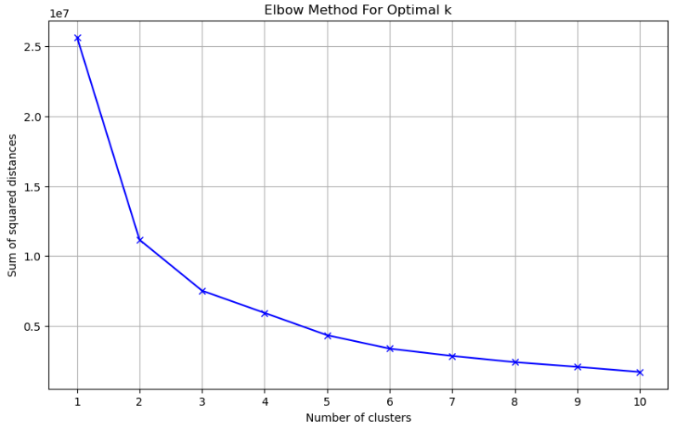
- **Average Transaction Volume:**

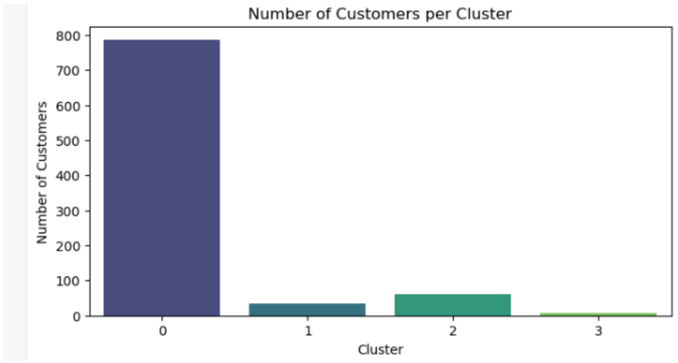
  Transaction Count (Blue Line): The pattern of transaction volumes also shows seasonal fluctuations, which might relate to business cycles, sales initiatives, or financial reporting periods.

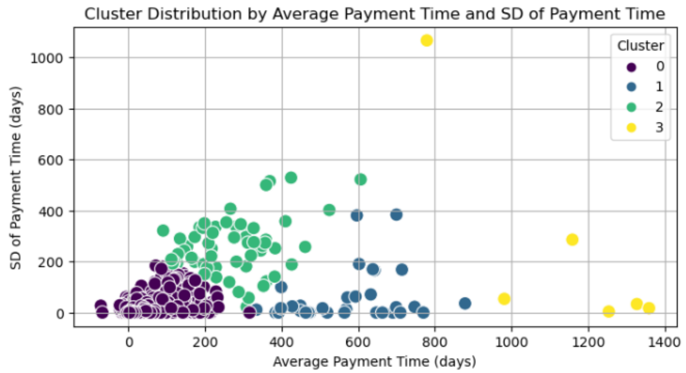
- **Key Observations:**

  Alignment with Business Cycles:
  The peaks in late payments might align with typical business cycles, such as end-of-quarter financial periods or seasonal sales, impacting the invoice creation practices.

  Cyclical Transaction Volumes:
  Transaction volumes vary across the year, suggesting strategic business activities like promotions or contract renewals often occur in specific months.







Exploratory Data Analysis (EDA)

- Customers are categorized into 4 Clusters as per their Average Payment Time and SD of Payment Time

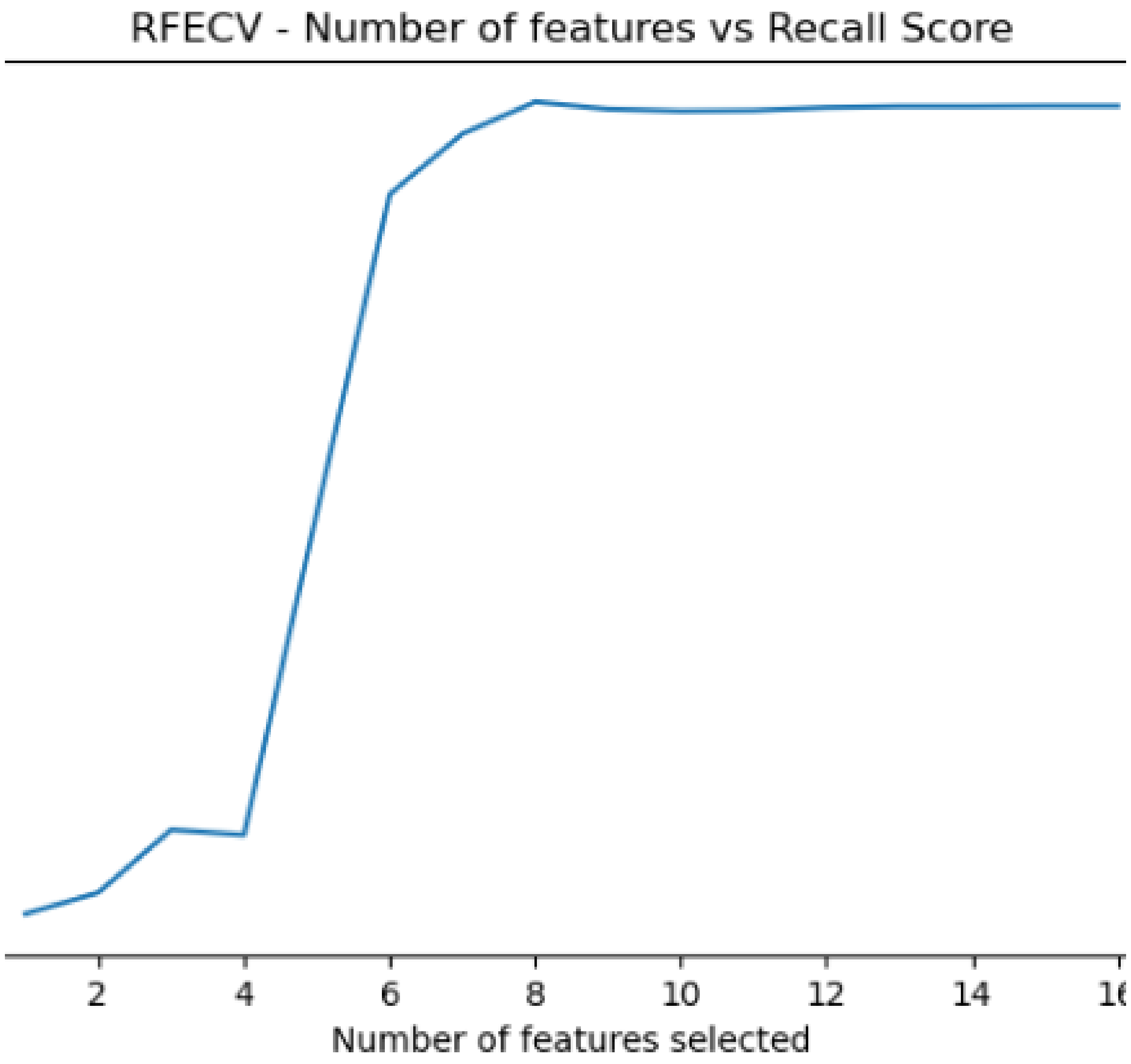
- Majority at Cluster -0 can be observed are making payments within 200 days

Section2

Model Building

Model – Logistic Regression

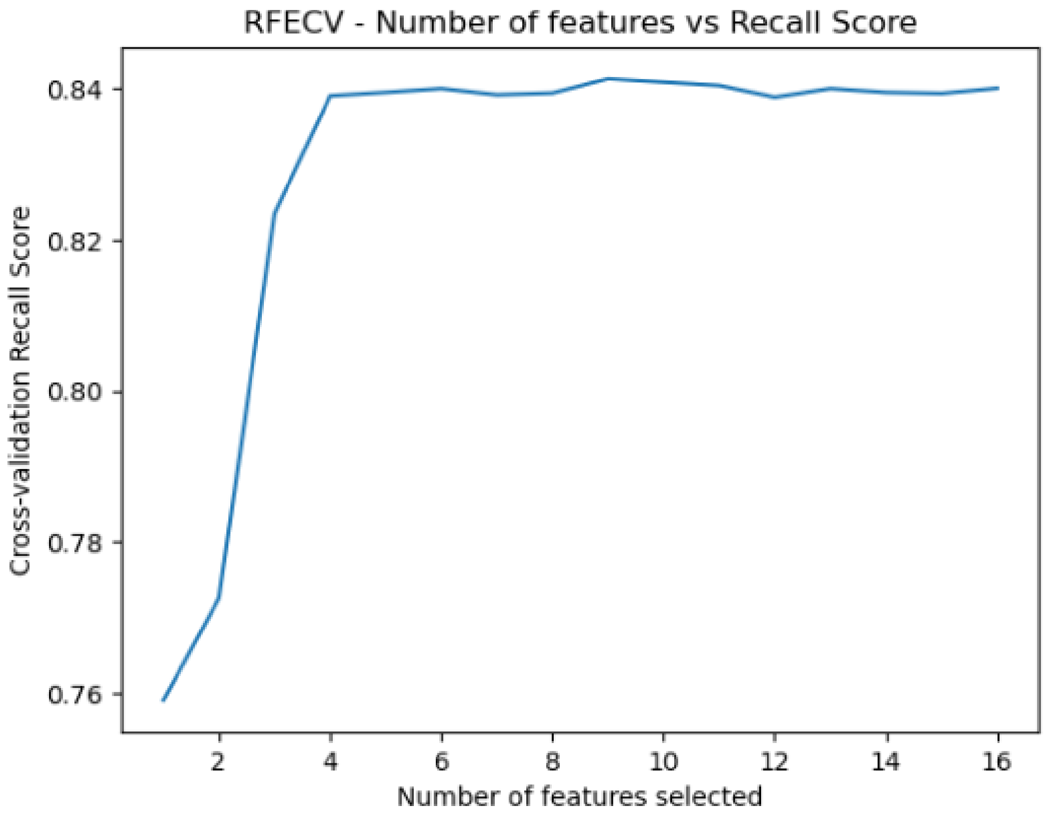
- Considering the intent to understand the pattern of late payments, recall is the metric for Model Evaluation.
- Model – 1: Logistic Regression Model is created for analysis.
    - Average Recall Score: 0.4503747444923915
    - Feature Selection conducted through RFECV
- Not satisfactory



Model – Random Forest

- Model – 2: Random Forest Model is created for analysis.
    - Average Recall Score (CV): 0.8411083352259823
    - Feature Selection conducted through RFECV – 10 features selected.
    - Hyperparameter tuning conducted - Best parameters found:

    {'bootstrap': False, 'max_depth': None, 'max_features': 6, 'min_samples_leaf': 3, 'min_samples_split': 2, 'n_estimators': 326}

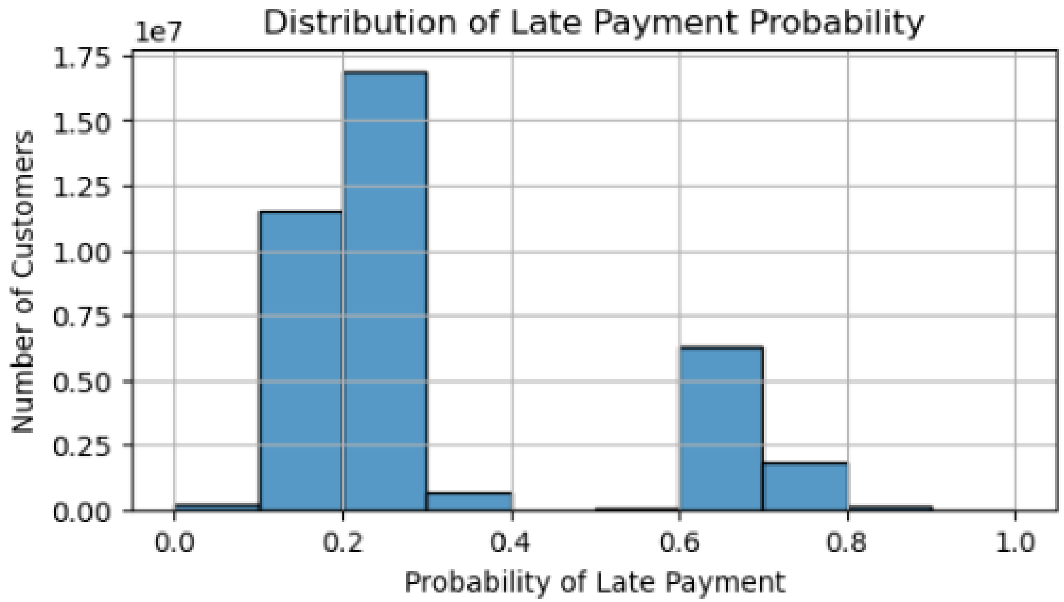


Model Evaluation

Classification Report on Training Set				
	precision	recall	f1-score	Support
0	0.93	0.97	0.95	34366
1	0.96	0.89	0.92	22015
Recall Score on Training Set: 0.8927549398137633				

Classification Report on Test Set				
	precision	recall	f1-score	Support
0	0.9	0.94	0.92	14930
1	0.9	0.84	0.87	9234
Recall Score on Test Set: 0.8391812865497076				

Inference from Training Dataset



Probability Threshold may be decided (e.g. 50%) - Focus can be provided to these customers on Late Payemnts.

Section3

Recommendations

Open Invoice Data

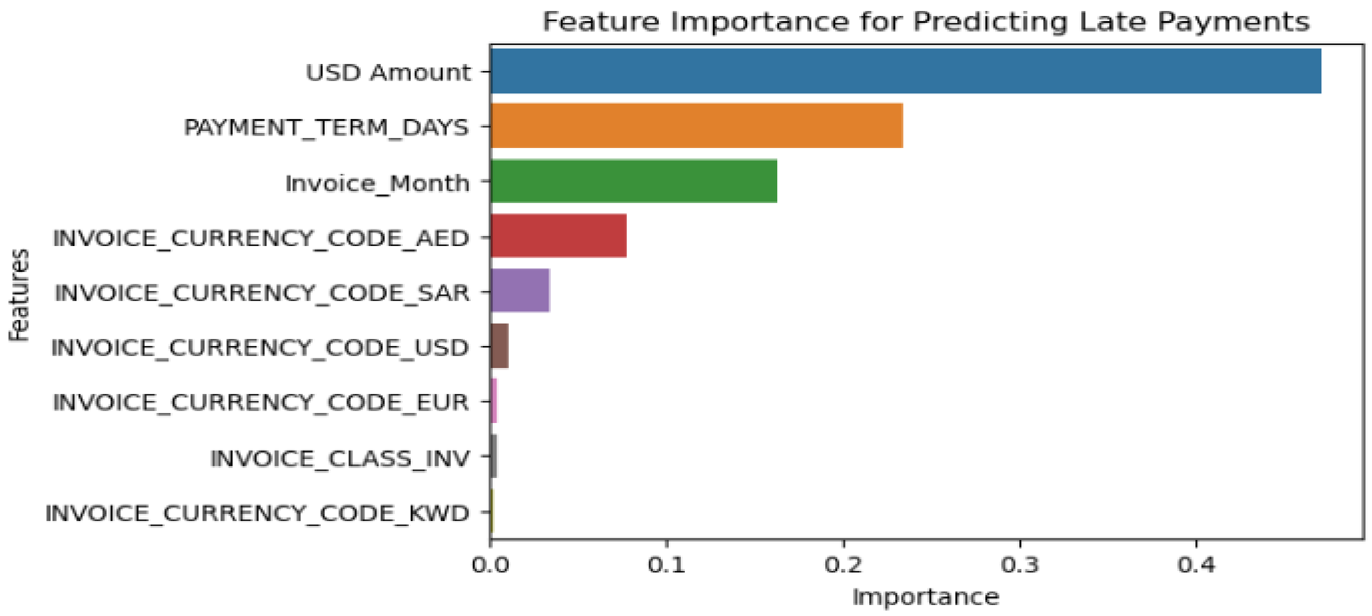
AS_OF_DATE	As of the current date that is 4th April, 2022 (when the report was extracted by Schuster)	DUE DATE	Date when the actual payment is due from the vendor
CUSTOMER TYPE	Whether the customer is a third party or holding a direct relationship with Schuster	TRANSACTION CURRENCY	In which currency the goods or services were sold
CUSTOMER_NAME	Name of the customer/vendor	LOCAL AMOUNT	Invoice value in local currency
CUSTOMER ACCOUNT NO	Customer's unique identity number	TRANSACTION CLASS	It is the invoice class - CM, DM, INV or PMT
TRANSACTION NUMBER	Transaction number against each invoice	AGE	It is the age of the open invoice. This is the difference between the Invoice Due date and the as-of-date
TRANSACTION DATE	This represents the date when the goods or services were sold by Schuster	USD AMOUNT	Invoice value in USD
PAYMENT TERM	Days given to the vendor/customer for making the payments	INV_CREATION_DATE	Date when the invoice was created

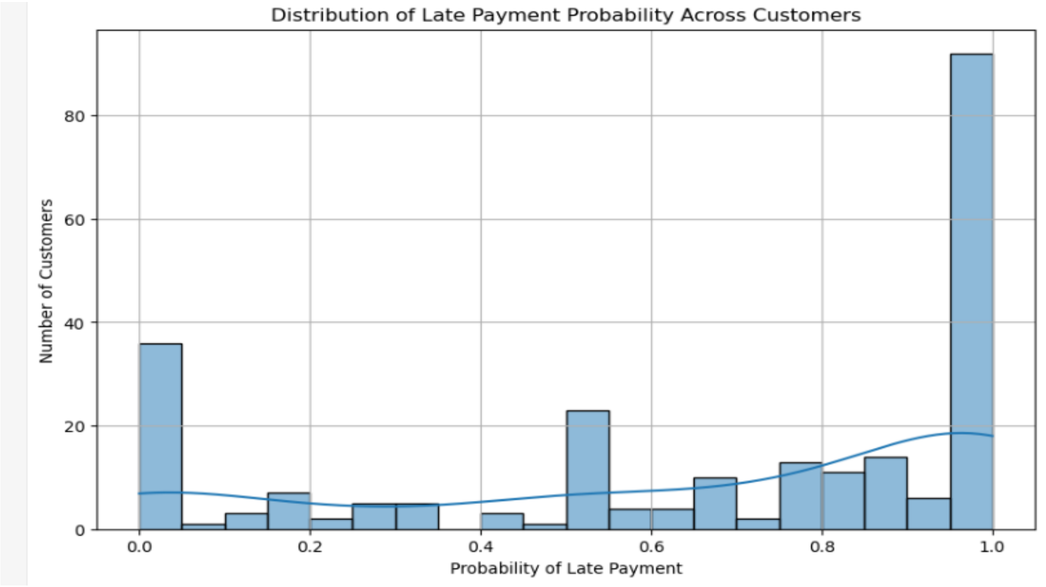
Data Description

- **Open Invoices Data:**

This contains data related to Open Invoices – to be used for testing the model built (88,204 entries and 14 columns)

Recommendations from Test Data





- It can be observed that Invoice Value (USD Amount) plays a major role in late payments.
- Payment term and Invoice Month are the next two most critical parameters to be reviewed by Schuster while identifying defaulters.

Defaulters risk profile:

Risk of Default and Range	No of Customers
Very Low (0-0.1)	37
Low (0.1-0.2)	10
Moderately Low (0.2-0.3)	6
Moderate (0.3-0.4)	6
Moderately High (0.4-0.5)	4
High (0.5-0.6)	25
Very High (0.6-0.7)	14

Thank You