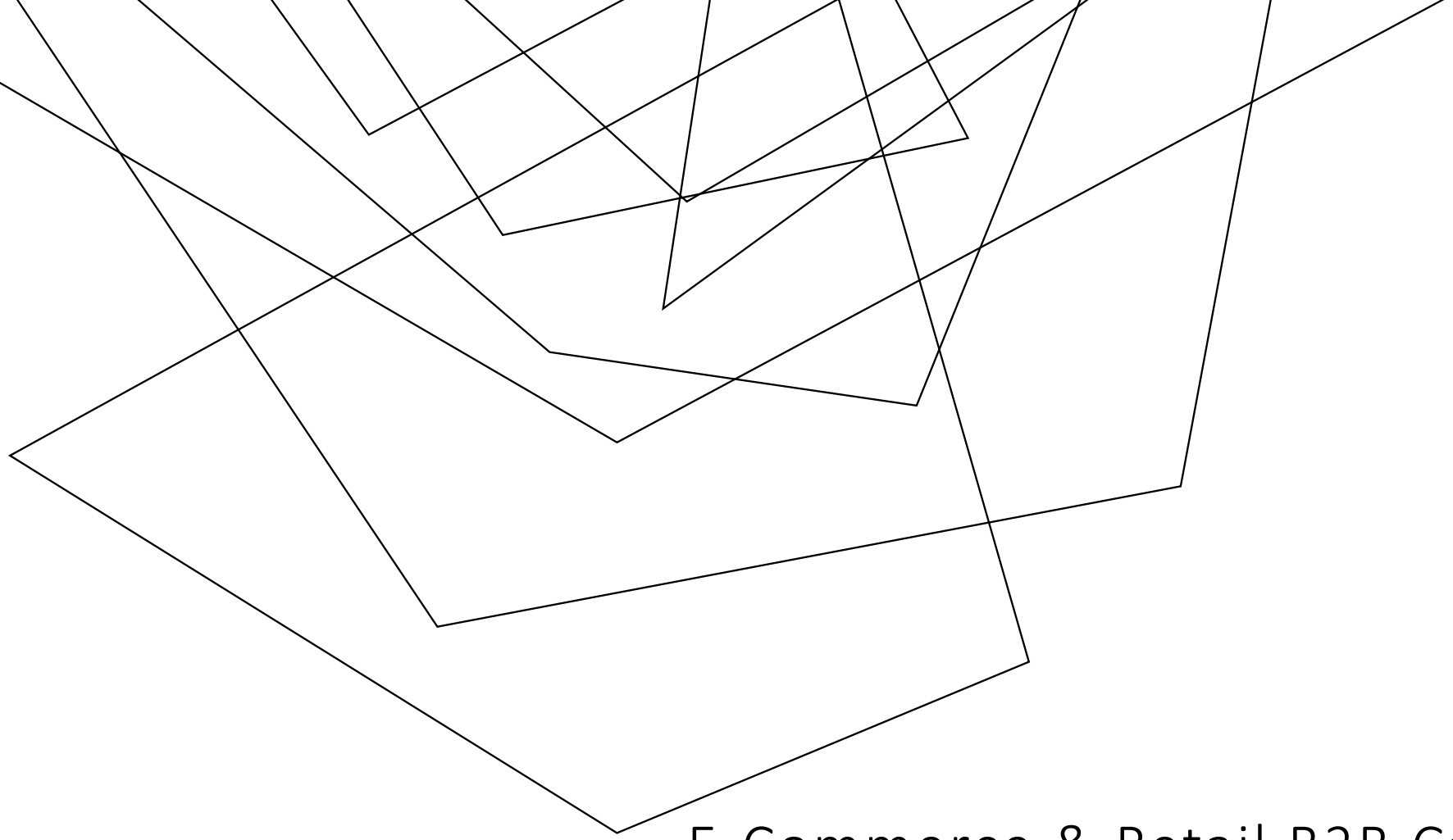


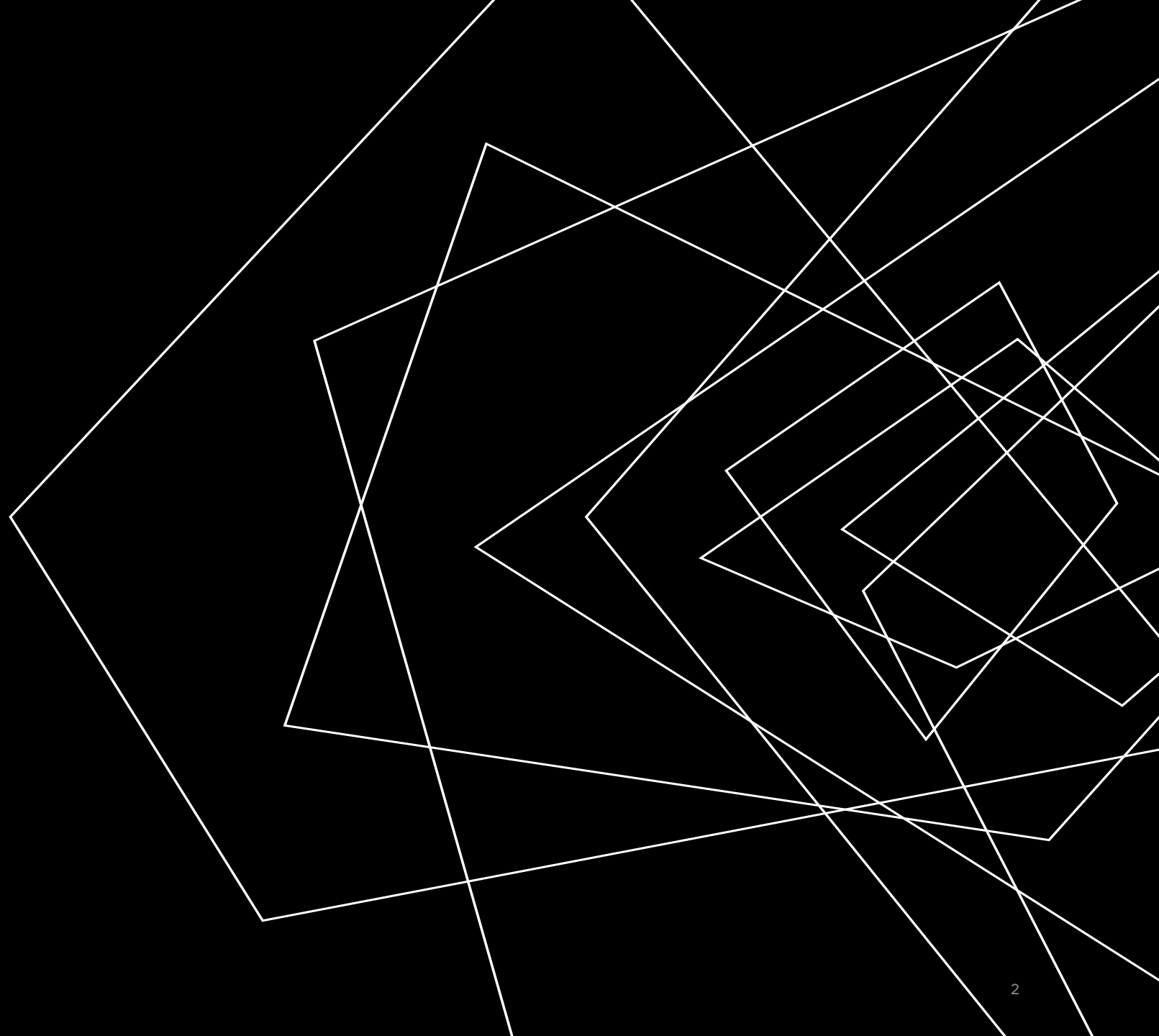
E-Commerce & Retail B2B Case Study

September 10, 2024

Vijay Viswanathan | Varsha K S

AGENDA

- Background and objectives
- Analysis Approach
- Executive Summary and Recommendations
- Technical results



Our understanding of the case study

Background

- Schuster is a multinational retail company dealing in sports goods and accessories.
- Significant business with hundreds of its vendors, with whom it has credit arrangements.
- They are facing issues with vendors who tend to make payments late. Chasing vendors to get the payment on time results in non-value-added activities, loss of time and financial impact.
- Schuster would thus try to understand its customers' payment behavior and predict the likelihood of late payments against open invoices so that they can take proactive action

Objectives

- Schuster would like to better understand the customers' payment behavior based on their past payment patterns
- It also wants to be able to predict the likelihood of delayed payment against open invoices from its customers so appropriate follow ups can be done to get the payments on time

Analysis approach

Data understanding

- Import Necessary Libraries
- Load the data and Data dictionary

Data Preprocessing and Preparation

- Missing value treatment
- Outlier treatment
- Dummy variable creation for categorical variables

Exploratory data Analysis

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Correlation analysis

Model building and Evaluation

- Test-train split of the data
- Feature Scaling
- Feature Selection
- Assessing the model

Results

- Final model that will help identify the customers likely to delay payments on open invoices

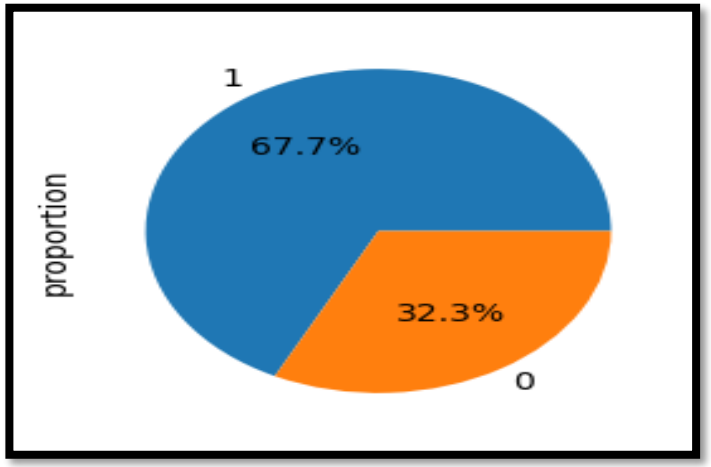
Executive Summary

- The model accuracy is ~88% i.e., the model can classify the open invoices which are likely to be delayed
- Customers can be mainly categorized into 3 groups
  - Category 0: Customers paying in time or having invoices with mean delay of 66 days
  - Category 1: Customers with invoices having a mean delay of 254 days
  - Category 2: Customers with invoices having mean delay of 757 days
- Majority of the customer invoices are of the first category
- Overall, 67% of the invoices are delayed with a mean invoice value of USD 514,966. The invoices paid on time have a mean invoice value of USD 585,115.
- Invoice types ‘goods’ have a higher percentage of delayed payment than ‘non-goods’.
- While invoice class INV has highest number of delayed payment, it has the lowest percentage of delayed payments.
- The invoice class CM (Credit memo) has the highest ratio of delayed payments
- Customers with the highest percentage of invoices likely to be delayed have been provided so that Schuster can take proactive action

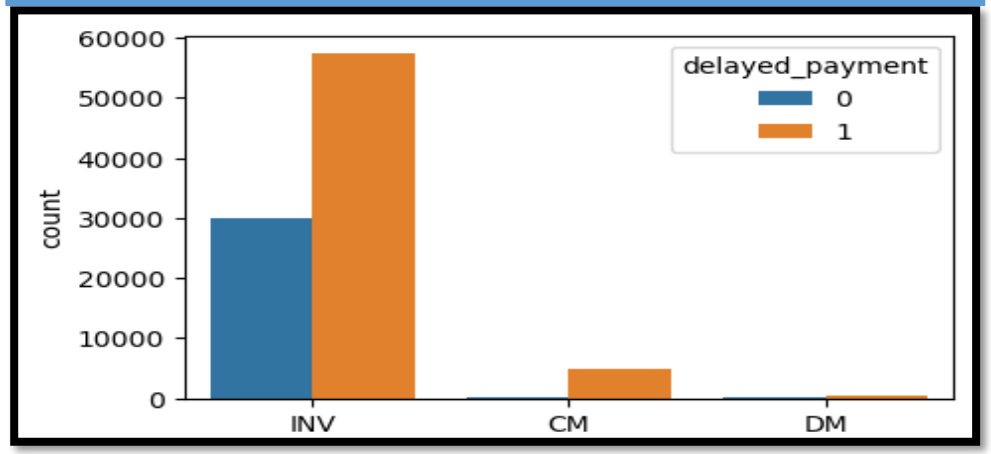
TECHNICAL RESULTS

INSIGHTS BASED EXPLORATORY DATA ANALYSIS OF THE RECEIVED PAYMENT DATA

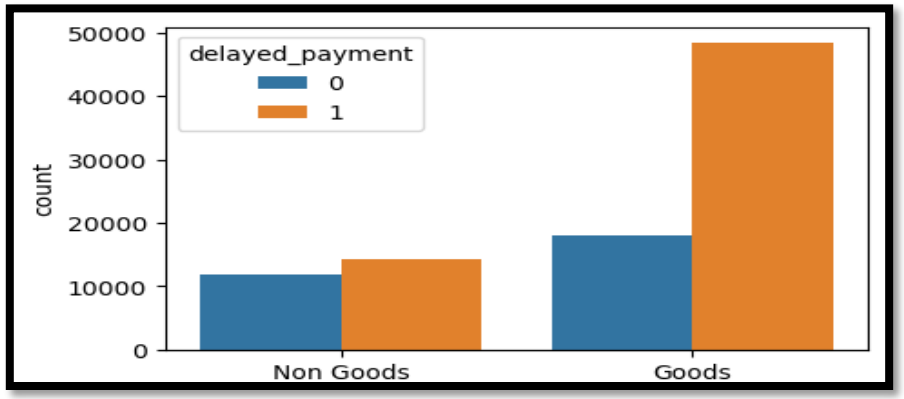
67% of all invoices are delayed



Invoice class 'INV' has the highest number of delayed payments



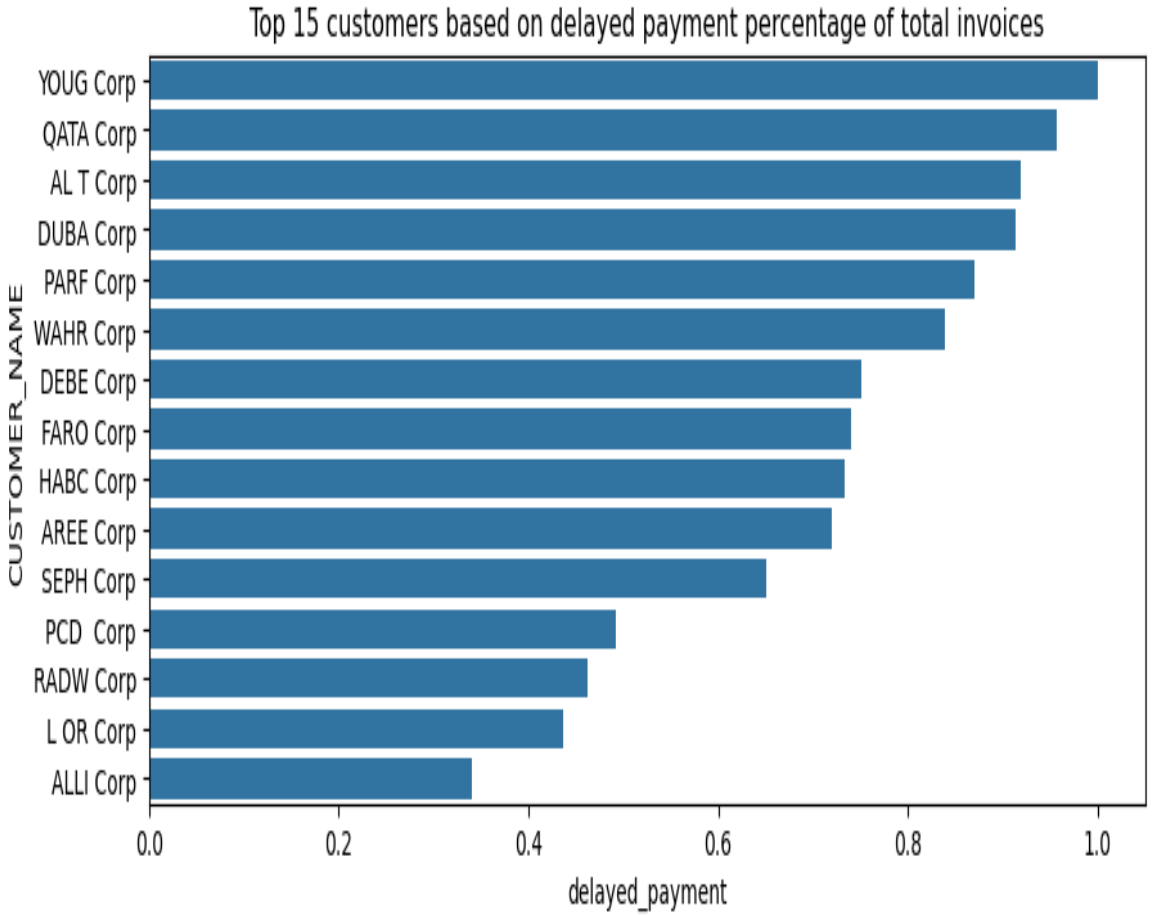
Customer purchasing 'Goods' tend to delay payments vs. Non-Goods'

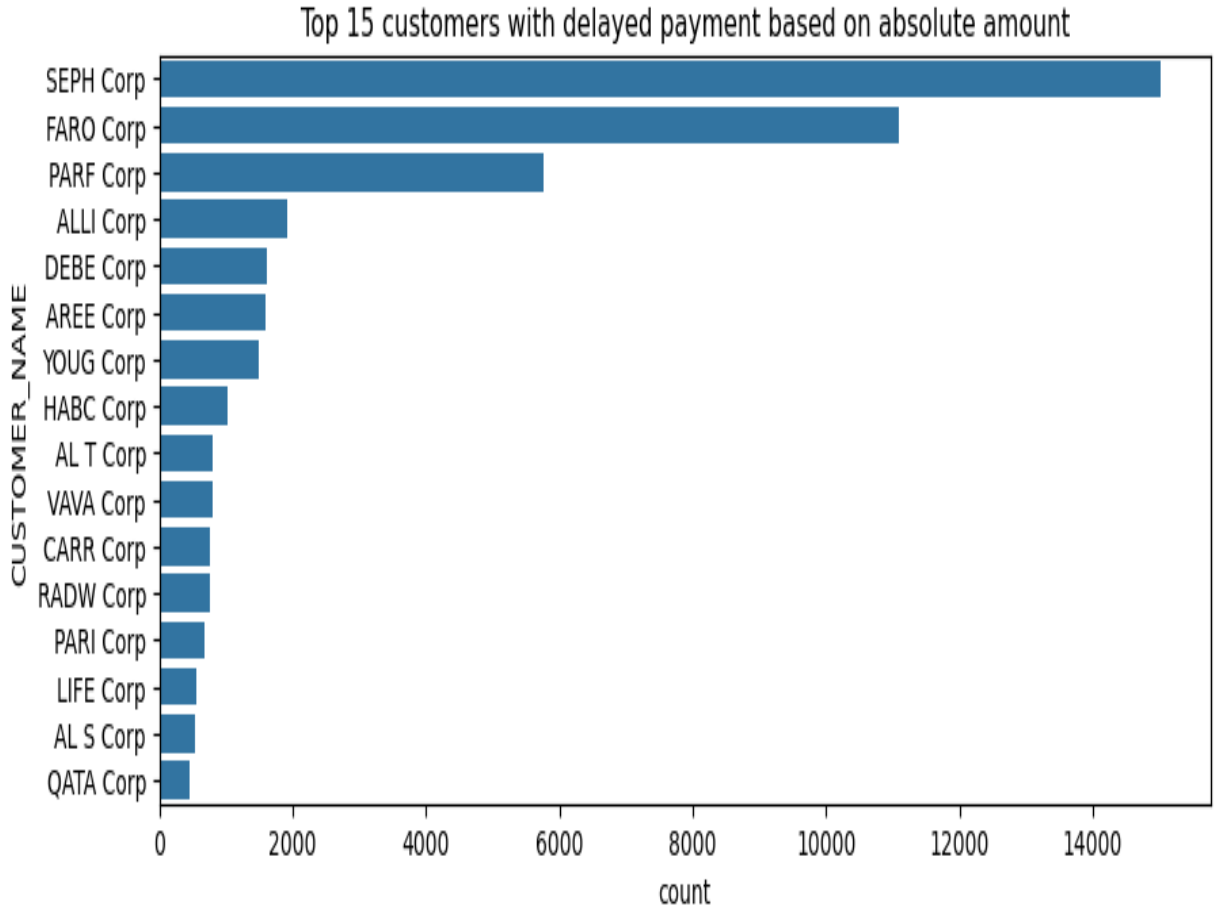


Payments for lower invoice values tend to more delayed

delayed_payment		mean	median	min	max
		USD Amount	USD Amount	USD Amount	USD Amount
0	0	585114.557398	234367.68	5.489449	6836735.61
1	1	514966.049743	164229.02	1.287743	10849828.80

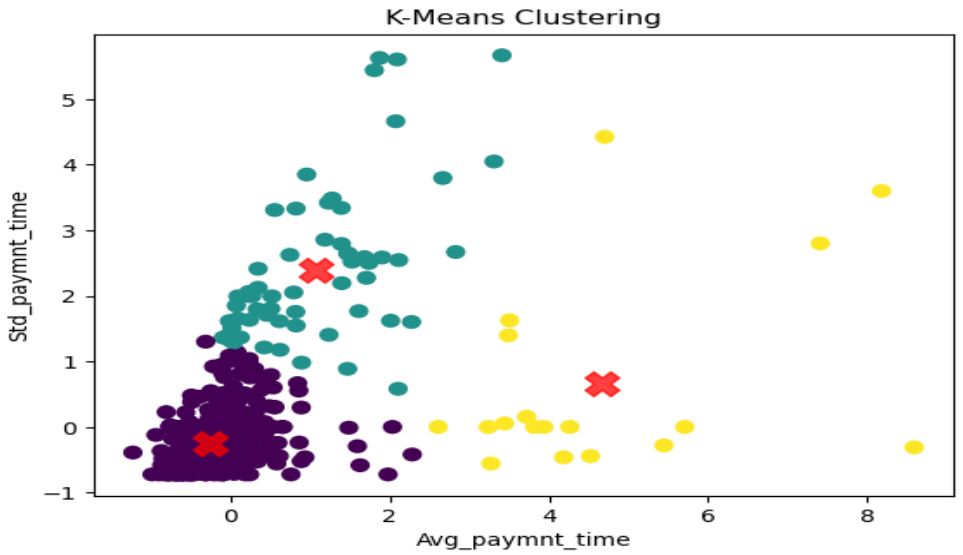
TOP CUSTOMERS WITH DELAYED PAYMENTS



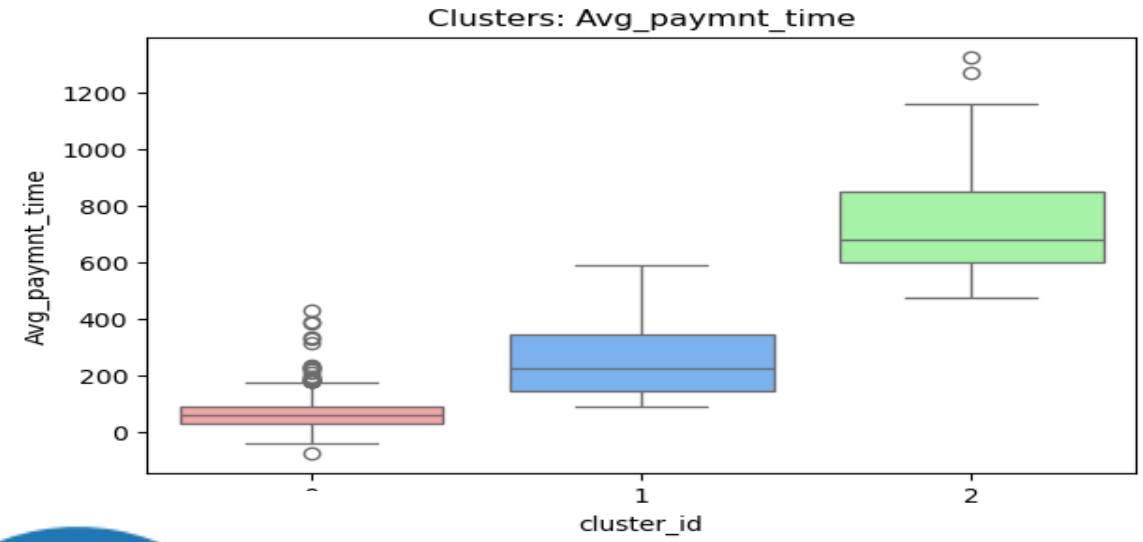


BASED ON AVERAGE PAYMENT DELAY CUSTOMER CAN BE CATEGORIZED INTO 3 GROUPS

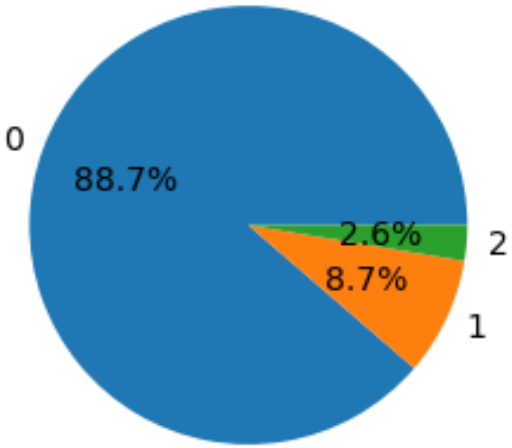
Customer are categorized into 3 different groups



A clear difference in the average payment delay is seen among the three groups



Majority of the customer invoices are of the first category

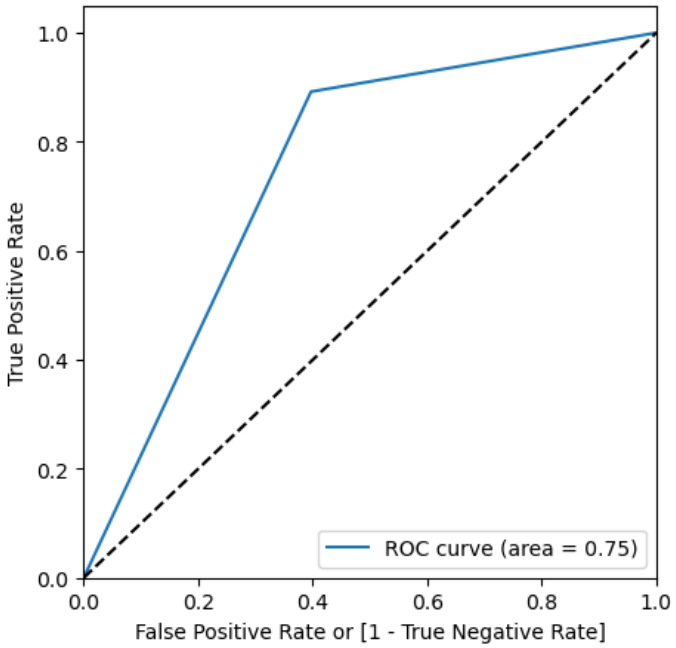


0: Customers paying in time or having invoices with mean delay of 66 days
1: Customers with invoices having a mean delay of 254 days
2: Customers with invoices having mean delay of 757 days

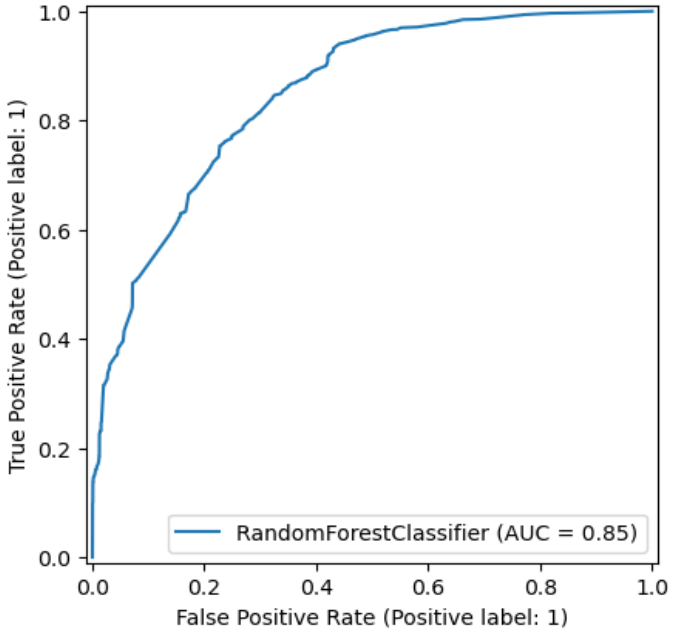
MODEL BUILDING AND EVALUATION

- Two classification models – Logistic regression and Random Forest were evaluated
- Random Forest had higher AUC score, and accuracy and hence was used for prediction

Logistic Regression



Random Forest



The following features appeared significant in predicting whether an invoice will be delayed on not

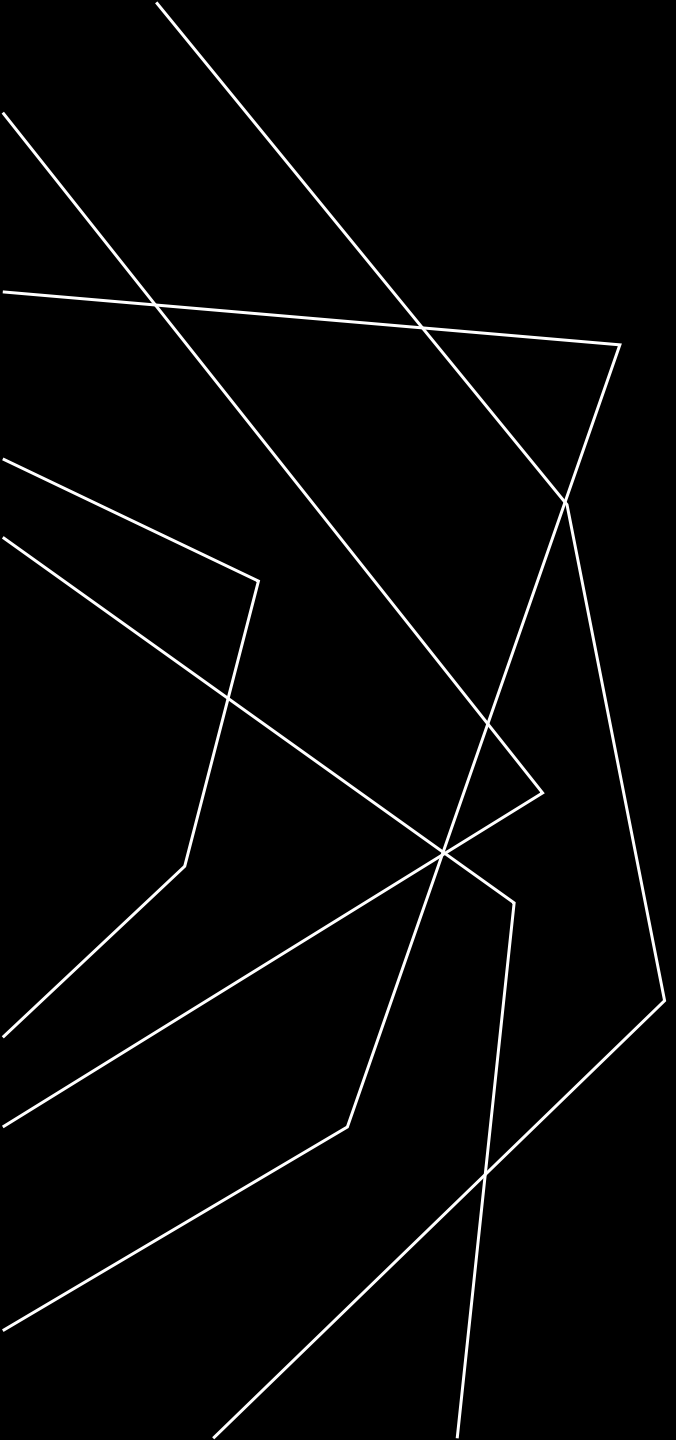
Importance rank	Feature name
1	Payment_Term_30 Days from EOM
2	Payment_Term_60 Days from EOM
3	Time_to_payment
4	USD Amount
5	Payment_Term_Immediate
6	cluster_id
7	Payment_Term_30 Days from Inv Date
8	Payment_Term_60 Days from Inv Date
9	SAR
10	Other_Curr
11	Payment_Term_Others
12	USD
13	Payment_Term_45 Days from EOM
14	Payment_Term_90 Days from EOM
15	INVOICE_CLASS_INV
16	Payment_Term_45 Days from Inv Date

Model Evaluation scores

Data	Overall Accuracy	Precision	Recall
Train data	89.64%	89.10%	96.48%
Test data	88.00%	88.01%	95.28%

Results for Schuster action

- The model predicts the open invoices that are likely to be delayed
- These have been aggregated at the customer level as a customer may have multiple invoices
- Customers with the highest percentage of open invoices that are likely to be delayed have been provided for Schuster to take appropriate action



THANK YOU