E-COMMERCE & RETAIL B2B CASE STUDY

Submitted By:-
- Shreevatsa Hegde
- Darshana M
- Chandrashekar Savadatti

E-COMMERCE & RETAIL B2B CASE STUDY

Problem Statement

Schuster is a multinational retail company dealing in sports goods and accessories. Schuster conducts significant business with hundreds of its vendors, with whom it has credit arrangements. Unfortunately, not all vendors respect credit terms and some of them tend to make payments late. Schuster levies heavy late payment fees, although this procedure is not beneficial to either party in a long-term business relationship. The company has some employees who keep chasing vendors to get the payment on time; this procedure nevertheless also 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.

To understand how to approach this problem using data science, let's first understand the payment process at Schuster now. Every time a transaction of goods takes place with a vendor, the accounting team raises an invoice and shares it with the vendor. This invoice contains the details of the goods, the invoice value, the creation date and the payment due date based on the credit terms as per the contract. Business with these vendors occurs quite frequently. Hence, there are always multiple invoices associated with each vendor at any given time.

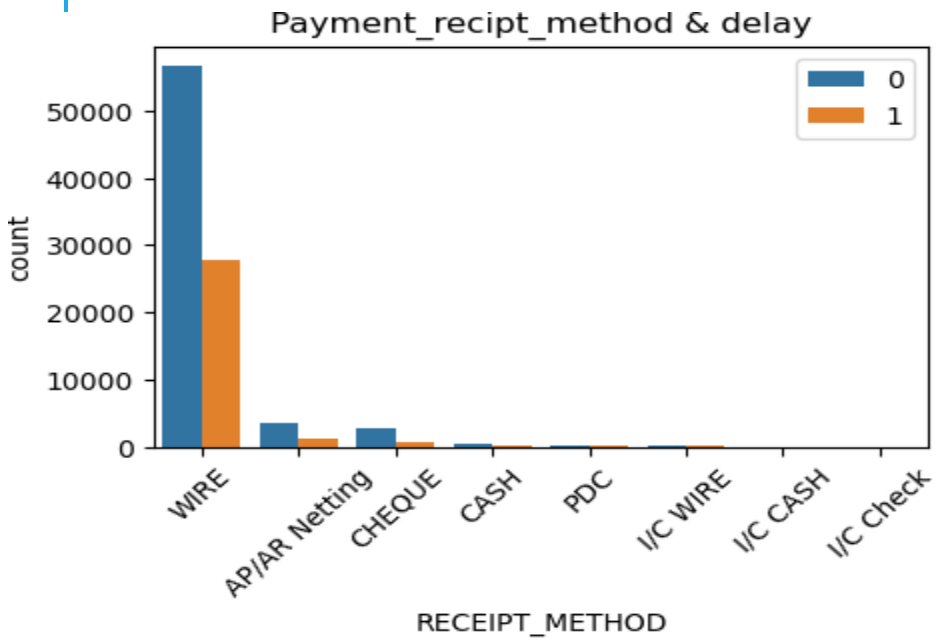
Business Goal

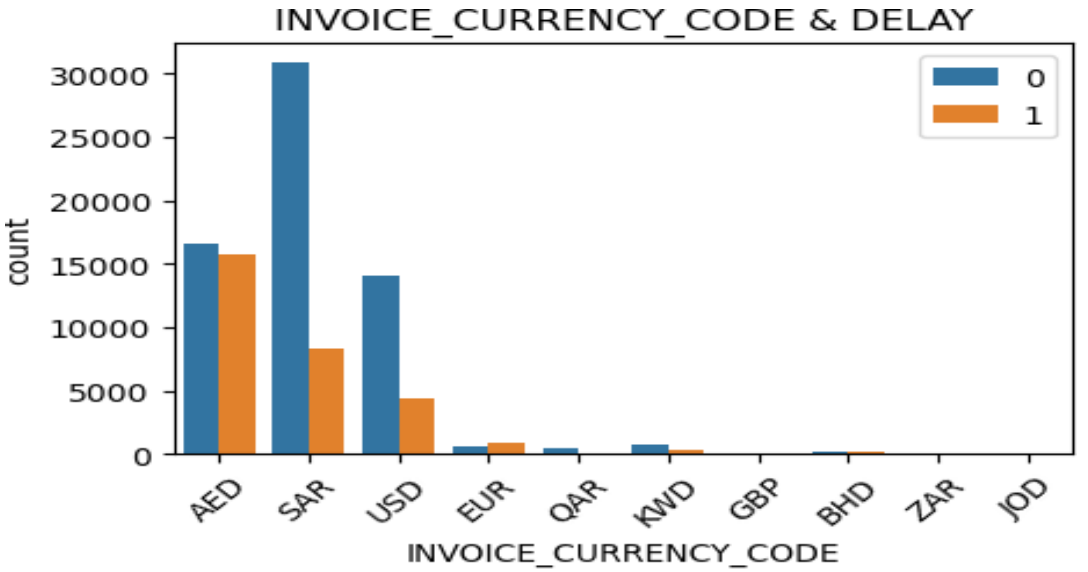
Schuster would like to better understand the customers' payment behavior based on their past payment patterns (customer segmentation).
Using historical information, it wants to be able to predict the likelihood of delayed payment against open invoices from its customers.
It wants to use this information so that collectors can priorities their work in following up with customers beforehand to get the payments on time.
To summarize, as a business analyst, you want to find the answer to these questions:
How can we analyze the customer transactions data to find different payment behaviors?
In which way can you segregate the customers based on their previous payment patterns/behaviours?
Based on the historical data, can you predict the likelihood of delayed payment against open invoices from the customers?
Can you draw any business insights based on your developed model?

STRATEGY

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Customer Segmentation
- Splitting the data into Test and Train dataset.
- Feature Scaling
- Building a logistic Regression model and calculate Delay probability. Evaluating the model by using different metrics - Precision and Recall.
- Applying the best model in Test data.
- Data cleaning of Open Invoice csv
- Feature Engineering & Data Preparation
- Final Predication on Open_Invoice_CSV

EXPLORATORY DATA ANALYSIS

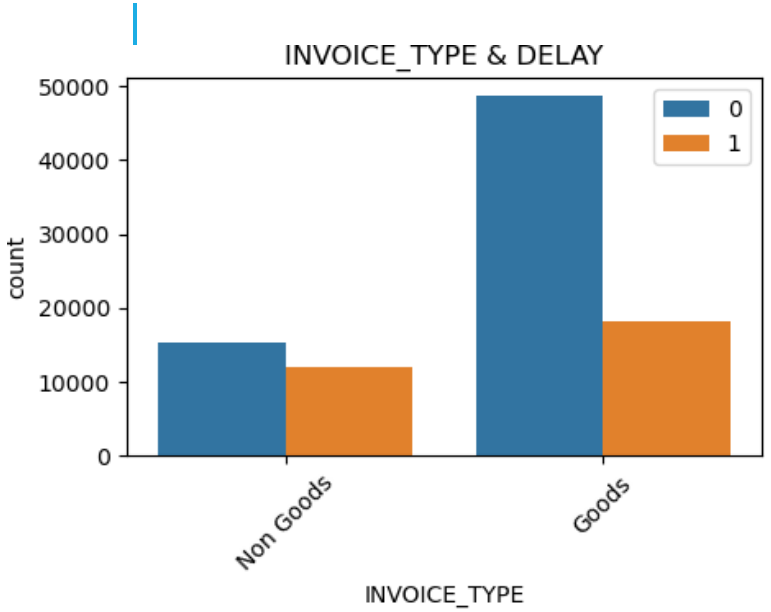


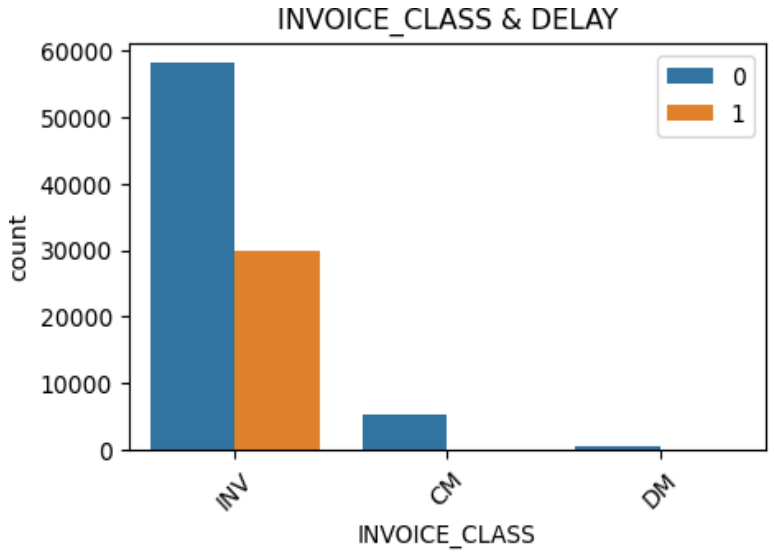


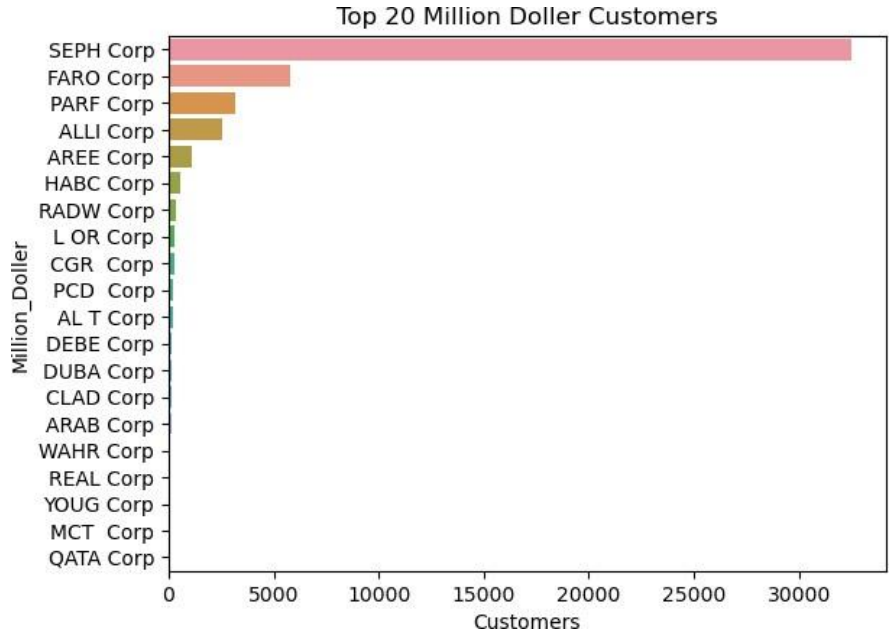
Inferences:-

- We can clearly see that majority of payments are being received via WIRE, AP/AR Netting or Cheques. Also we don't get enough information with respect to our target variable.
- We can clearly see that Payment Currency SAR have higher delays than other payment receipt currency.
- We can have some additional countermeasures for customers who pays in SAR, AED & USD.

EXPLORATORY DATA ANALYSIS



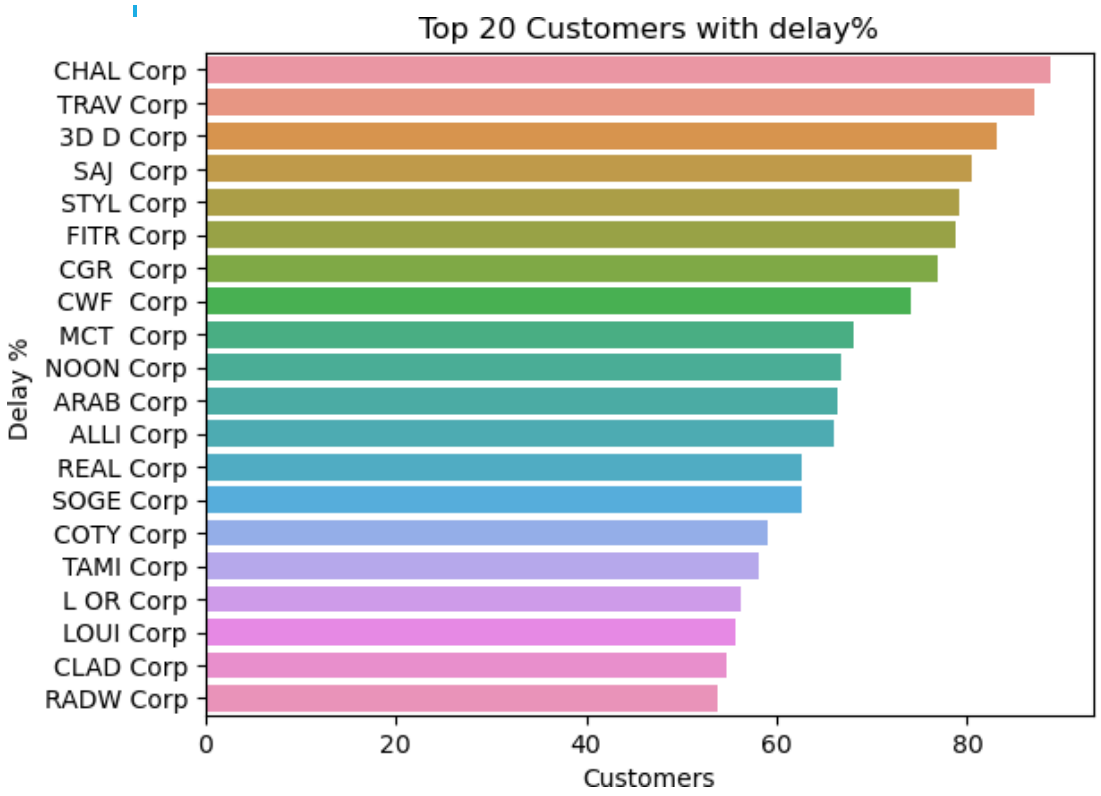


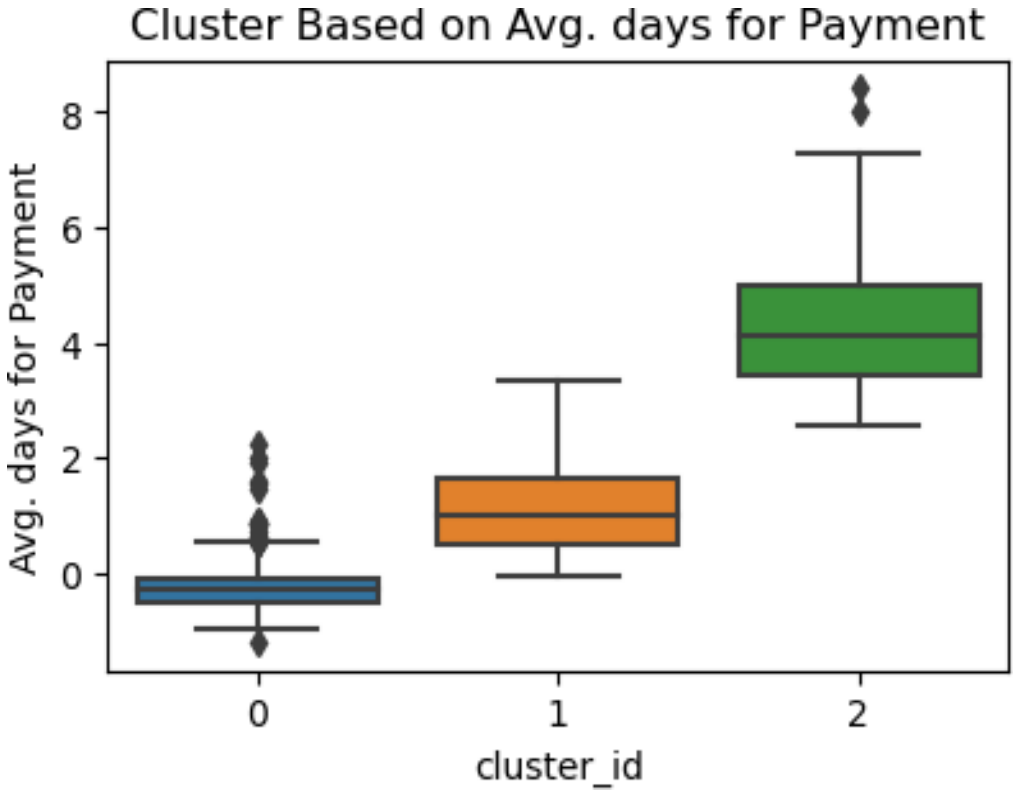


Inferences:-

- We can say that Goods category is having higher delays ration in comparison to Non Goods category.
- By looking it is very obvious that INV class invoices have only Delay.
- We can see SEPH Corp,FARO Corp,PARF corp,ALLI Corp and AREE Corp are the top customers.

EXPLORATORY DATA ANALYSIS





Inferences:-

- We can see top 5 companies which are having higher delay percentage are as below.
  - CHAL Corp
  - TRAV Corp
  - 3D D Corp
  - SAJ Corp
  - STYL Corp

- Based on Avg.days for Payment clusters ,we can see three clusters as below-

-'1' Cluster -- Fast Invoice Payment
-'2' Cluster -- Delay/Late Invoice Payment
-'0' Cluster -- Medium Invoice Payment

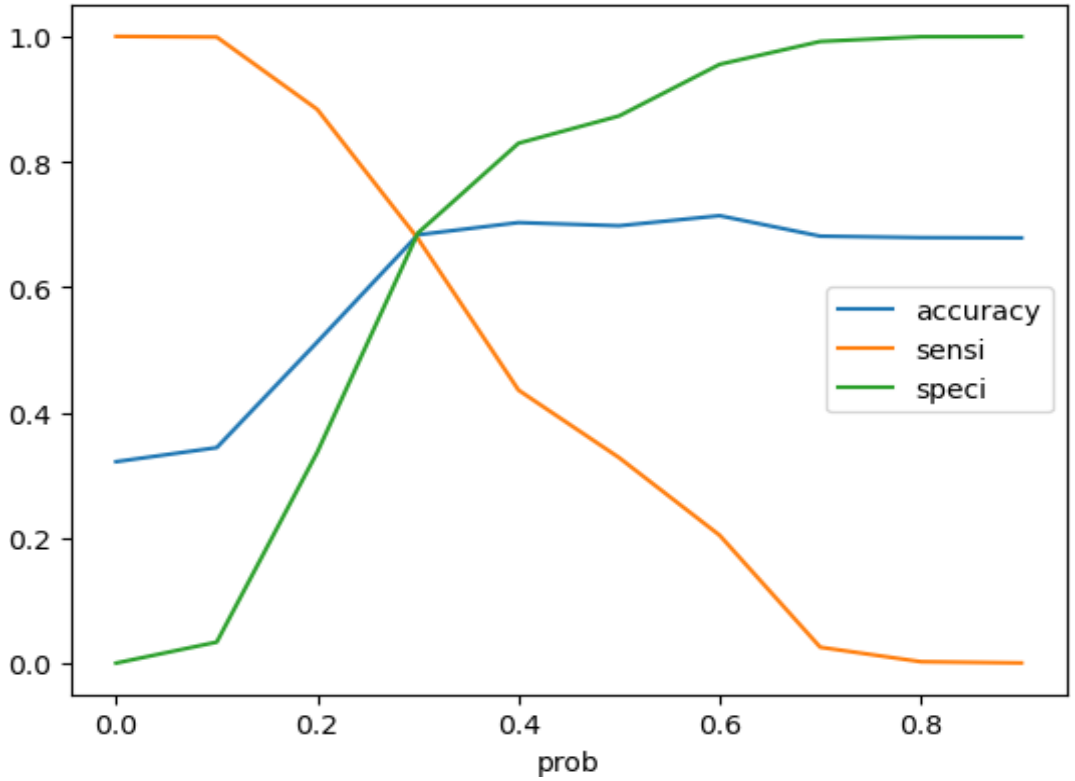
VARIABLES IMPACTING THE DELAY

Top 10 parameters mostly to be considered are as below(In decreasing order of importance)

Payment_Term_Days →	7.488
Usd Amount →	1.3225
Invoice_Currency_Code_Eur →	0.4954
Cluster_Id →	0.2907
Invoice_Currency_Code_Kwd →	-0.7207
Invoice_Currency_Code_Usd →	-0.9909
Invoice_Currency_Code_Sar →	-1.0766
Invoice_Currency_Code_Qar →	-1.2977
Invoice_Currency_Code_Gbp →	-1.3108
Invoice_Class_Dm →	-1.8664

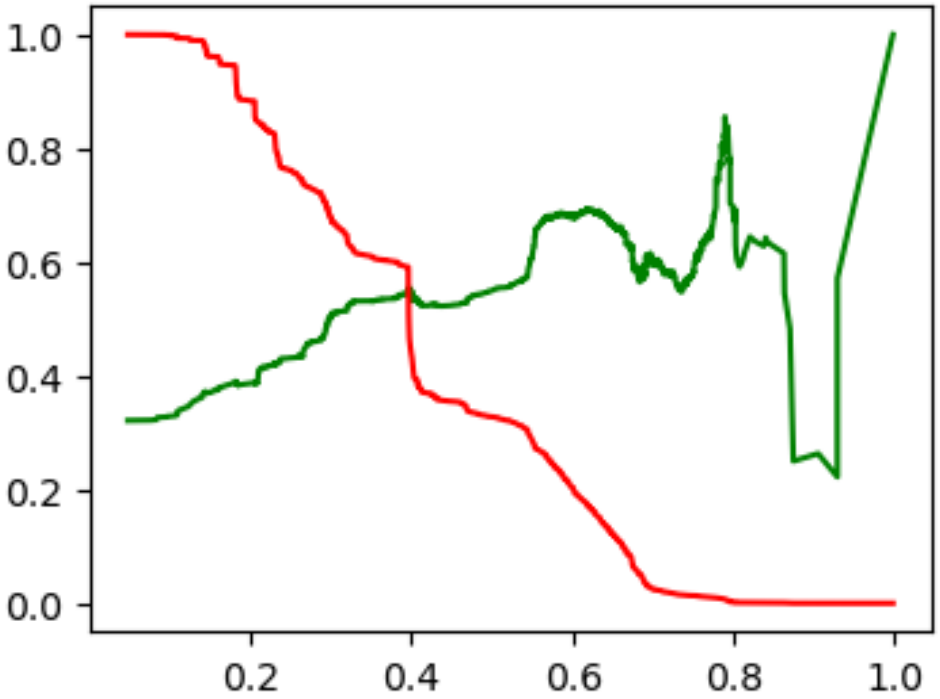
MODEL EVALUATION - ON TRAIN DATA & TEST DATA SET

Sensitivity and Specificity on Train Data Set



The graph depicts an optimal cut off of 0.3 based on Accuracy, Sensitivity and Specificity

Precision and Recall on Train Dataset



The graph depicts an optimal cut off of 0.4 based on Precision and Recall

CONCLUSION

From our Analysis we can make the following inference:-

- Customers in cluster 2 has the highest average payment days, so that might be the ones that make a late payment
- Customers in cluster 0 has the second highest average payment days followed by cluster .
- Customers in cluster 1 have the lowest average payment days and are more likely to make payments on time. We intend to use the above inference along with the predictions from the logistic regression to make decisions on which customers to focus.
- Payment Currency SAR have higher delays than other payment receipt currency.
- Company can have some additional countermeasures for customers who pays in SAR, AED & USD.
- Top 5 companies which are having higher delay percentage are as below on past data. Company should focus on these companies specially.
  - CHAL Corp
  - TRAV Corp
  - 3D D Corp
  - SAJ Corp
  - STYL Corp

- From Open_Invoice file we have predicted whether customer will delay(0) or pay in time(1) & provide on Customer Level Delay% expected.
- Company should invest time in higher delay% customers to get the payment on time.