# SCHUSTER
# E – Commerce & RETAIL B2B CASE STUDY

Submitted By:

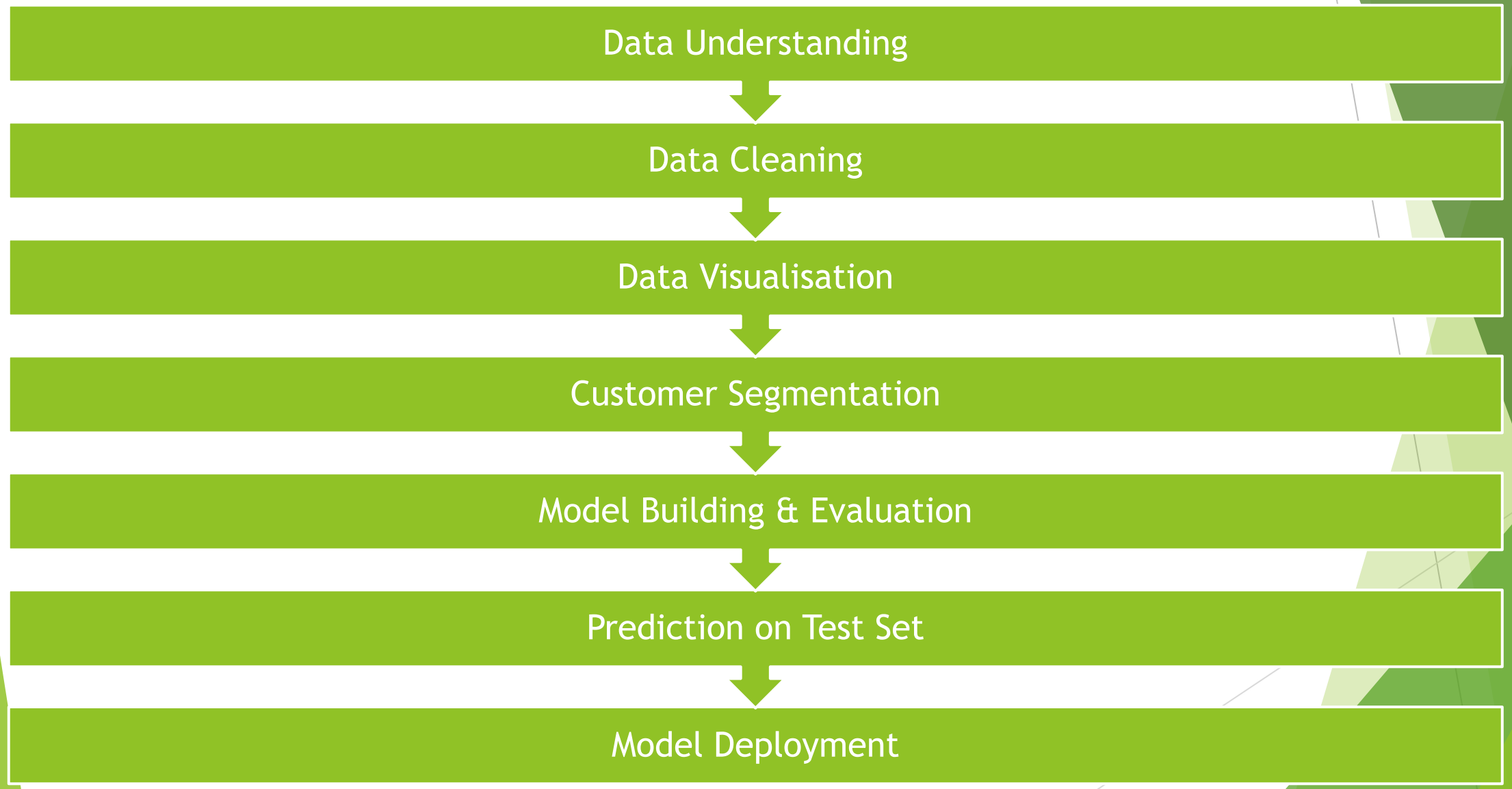
Amlesh Dwivedi

# 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.

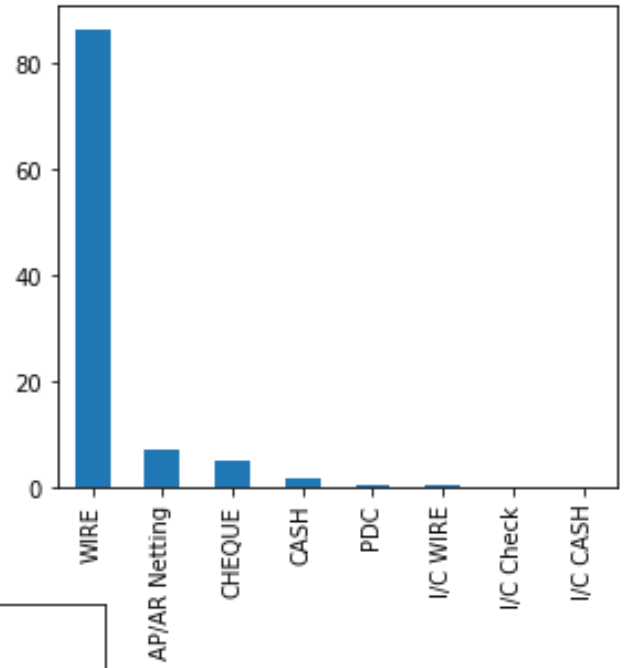
- Schuster would like to better understand the customers' payment behaviour based on their past payment patterns (customer segmentation).
- Using historical information, it wants to be able to predict the likelihood of delayed payment against open invoices from its customers.
- It wants to use this information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

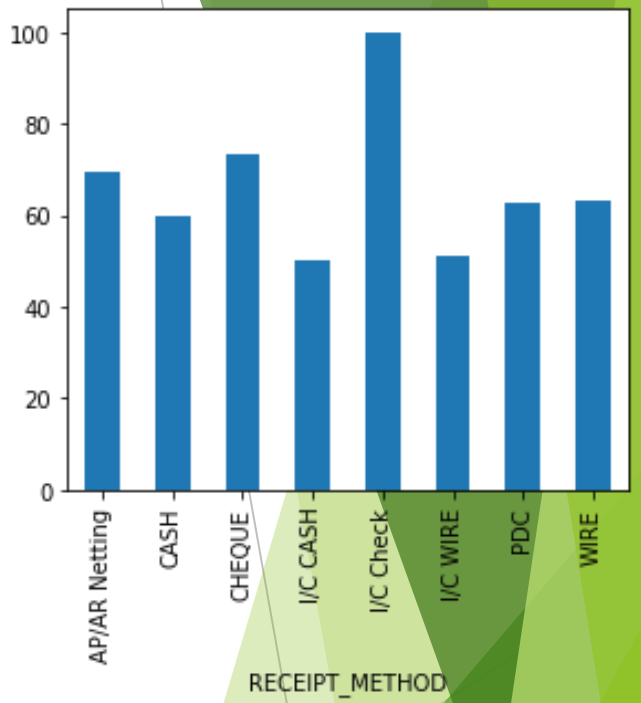
# Steps Involved

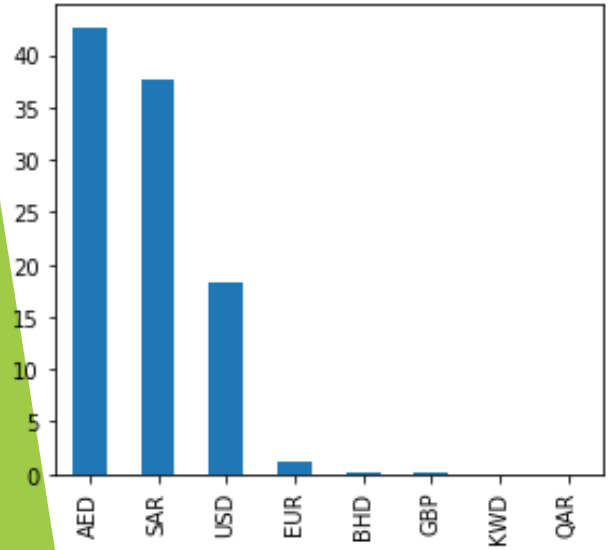


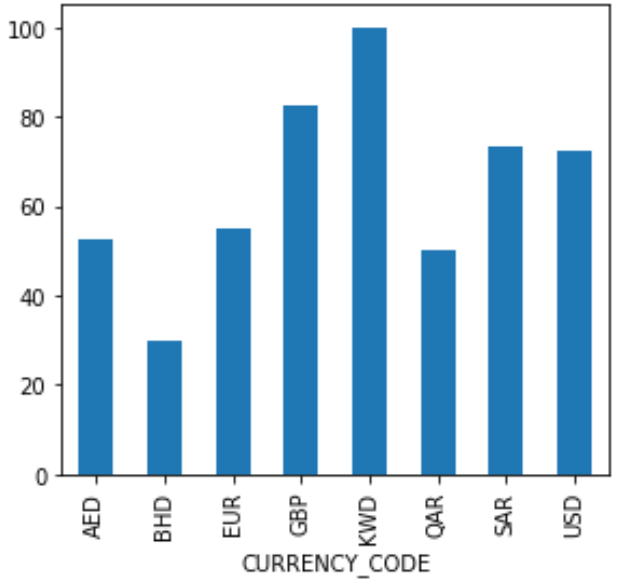
# EDA Insights

- Delays are high for I/C Check method
-Wire is the popular method of transfer



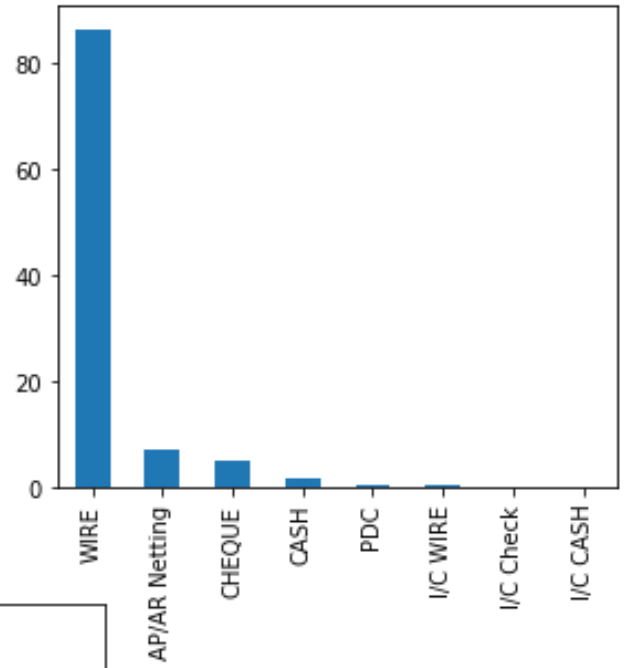



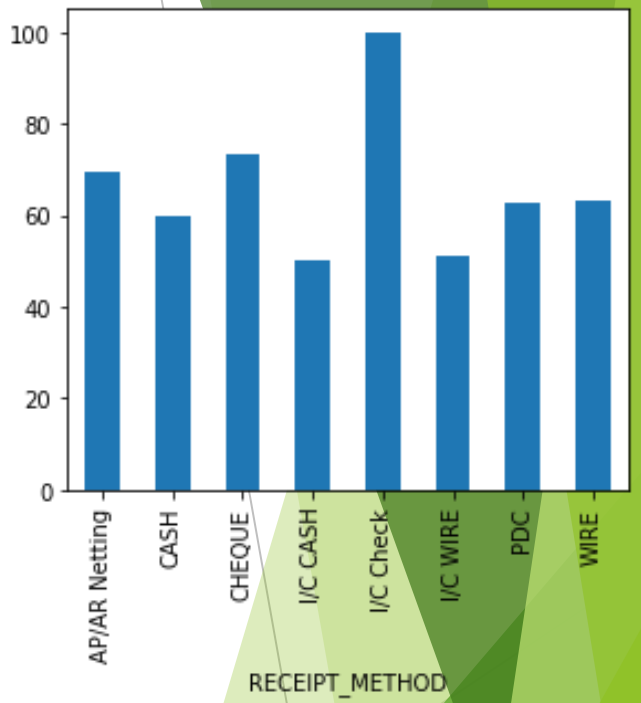



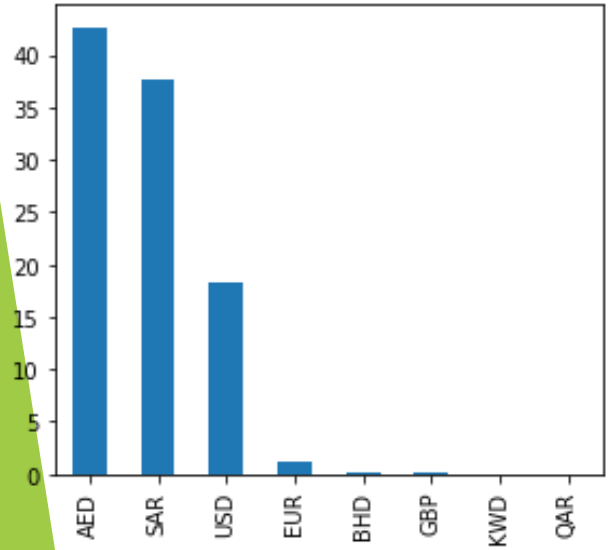

-Invoice are majorly generated in AED, SAR and USD
- Delay is max for KWD

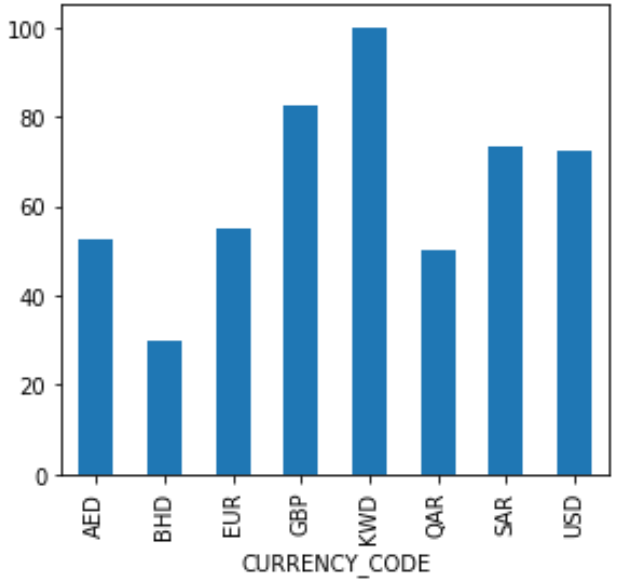
# EDA Insights

- Delays are high for I/C Check method
-Wire is the popular method of transfer



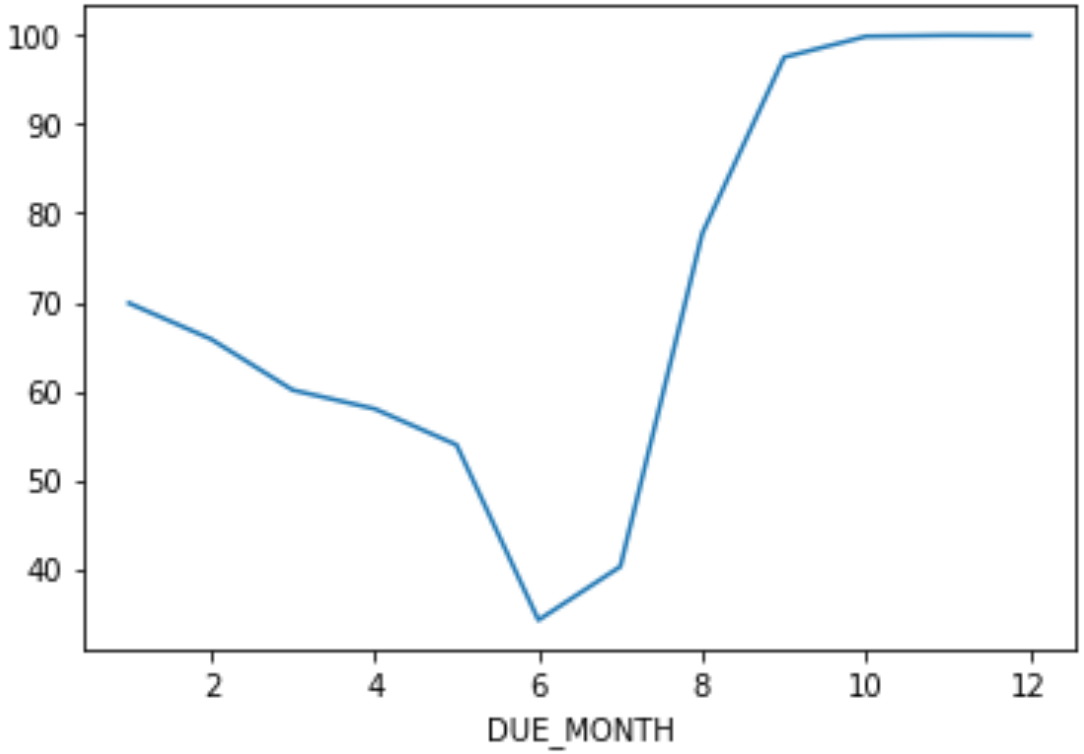



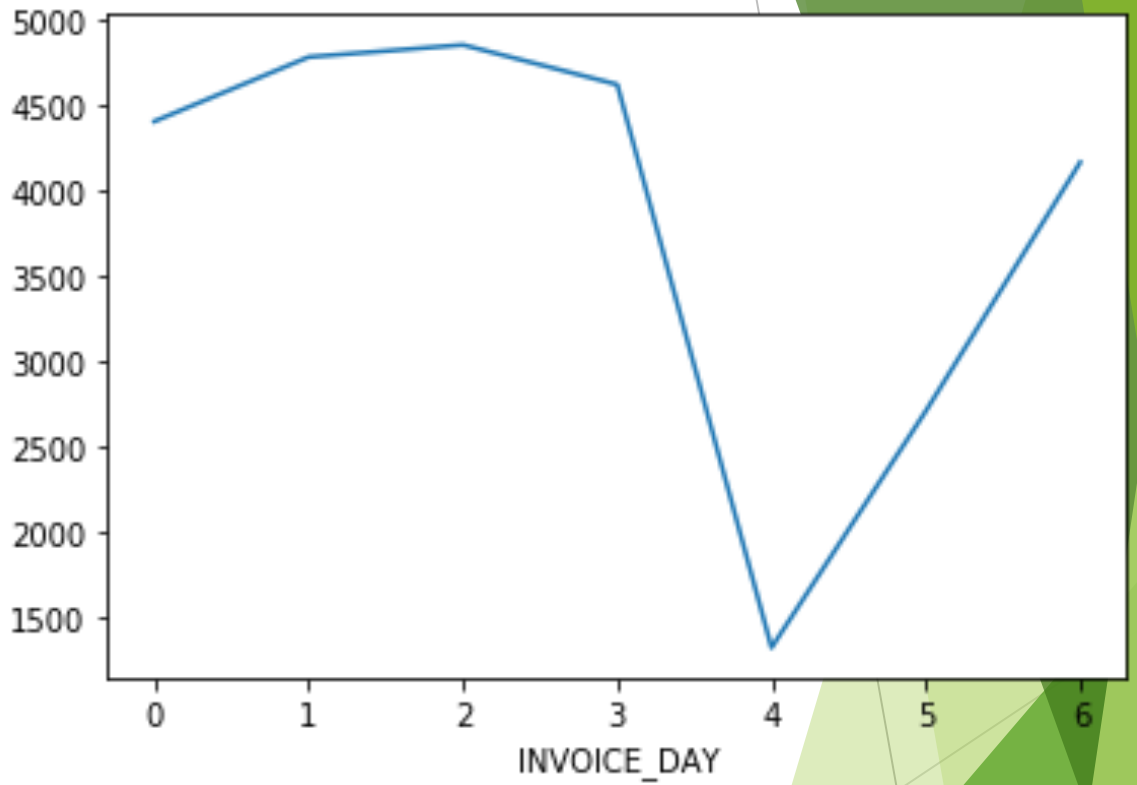




-Invoice are majorly generated in AED, SAR and USD
- Delay is max for KWD

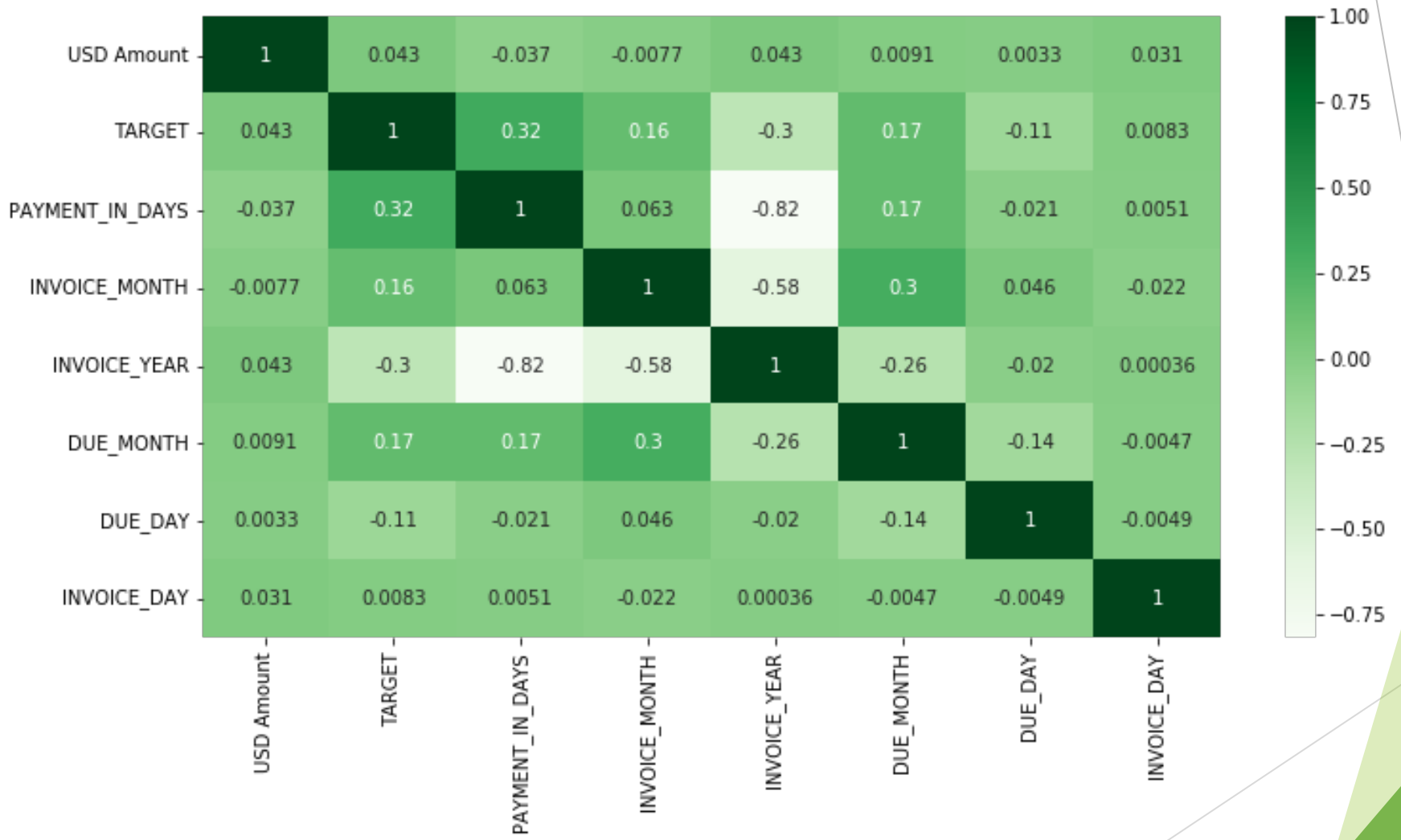
# EDA Insights



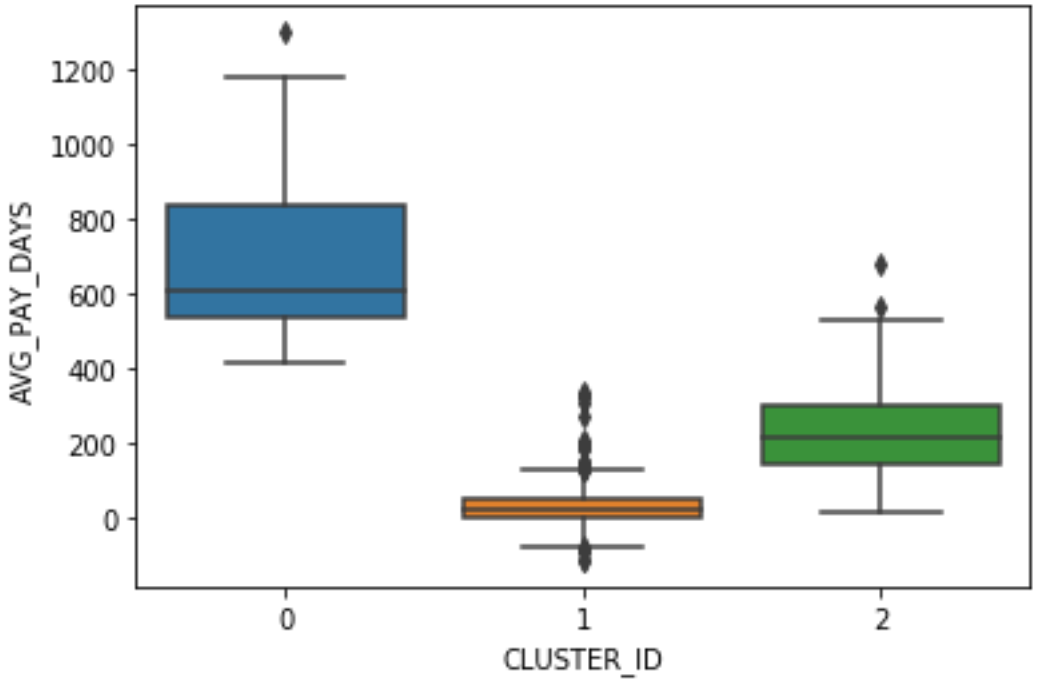

- Invoice for July Sep and Oct- Nov have significant delay
- Invoice for Friday are less

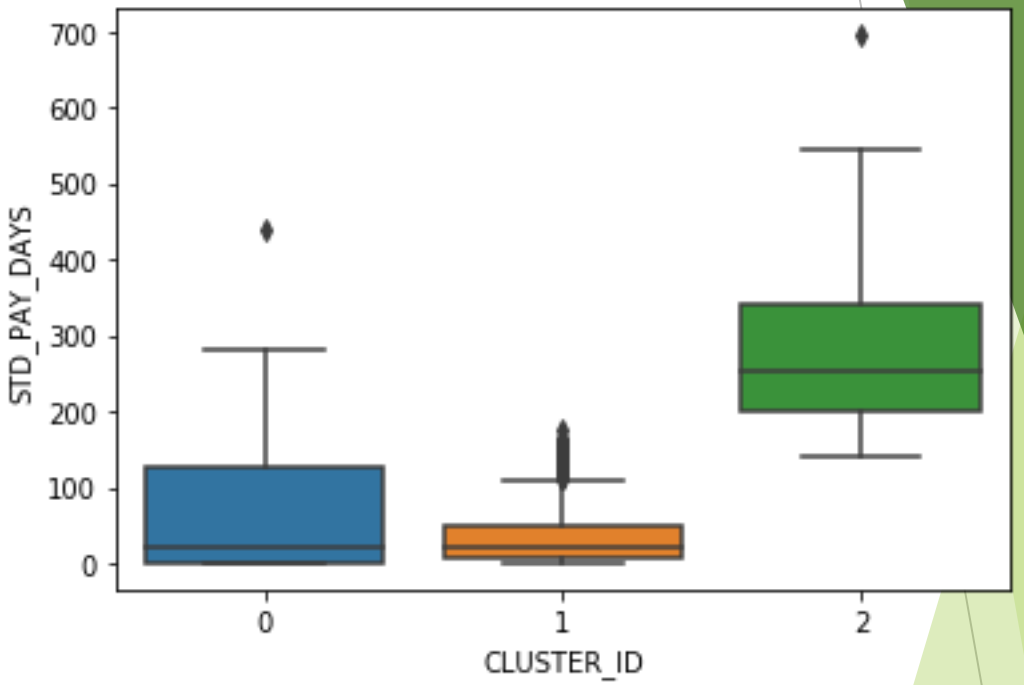
# EDA Insights




No strong correlation between the numerical variables

# Customer Segmentation




Cluster 0 has major delay with moderate SD Cluster 1 has early payments with small SD Cluster 2 has small delays but a bit high SD

# Logistic Regression Model

| | Features | VIF |
|---|---|---|
| 0 | const | 2.787946 |
| 1 | PAYMENT_TERM_15 Days from EOM | 1.154744 |
| 2 | PAYMENT_TERM_30 Days from EOM | 1.153192 |
| 3 | PAYMENT_TERM_60 Days from EOM | 1.073112 |
| 4 | PAYMENT_TERM_Immediate | 1.121179 |
| 5 | DUE_MONTH_6 | 1.024676 |
| 6 | DUE_MONTH_7 | 1.010495 |
| 7 | DUE_MONTH_9 | 1.011407 |
| 8 | DUE_MONTH_10 | 1.011082 |
| 9 | CLUSTER_ID_2 | 1.154136 |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| **const** | 0.8321 | 0.022 | 37.401 | 0.000 | 0.788 | 0.876 |
| **PAYMENT_TERM_15 Days from EOM** | 2.1928 | 0.126 | 17.371 | 0.000 | 1.945 | 2.440 |
| **PAYMENT_TERM_30 Days from EOM** | -2.2189 | 0.038 | -57.902 | 0.000 | -2.294 | -2.144 |
| **PAYMENT_TERM_60 Days from EOM** | -1.3068 | 0.051 | -25.675 | 0.000 | -1.407 | -1.207 |
| **PAYMENT_TERM_Immediate** | 2.9537 | 0.112 | 26.344 | 0.000 | 2.734 | 3.173 |
| **DUE_MONTH_6** | -1.3653 | 0.048 | -28.488 | 0.000 | -1.459 | -1.271 |
| **DUE_MONTH_7** | -1.7054 | 0.097 | -17.627 | 0.000 | -1.895 | -1.516 |
| **DUE_MONTH_9** | 2.8743 | 0.348 | 8.261 | 0.000 | 2.192 | 3.556 |
| **DUE_MONTH_10** | 5.5862 | 1.004 | 5.563 | 0.000 | 3.618 | 7.554 |
| **CLUSTER_ID_2** | 0.6292 | 0.034 | 18.645 | 0.000 | 0.563 | 0.695 |

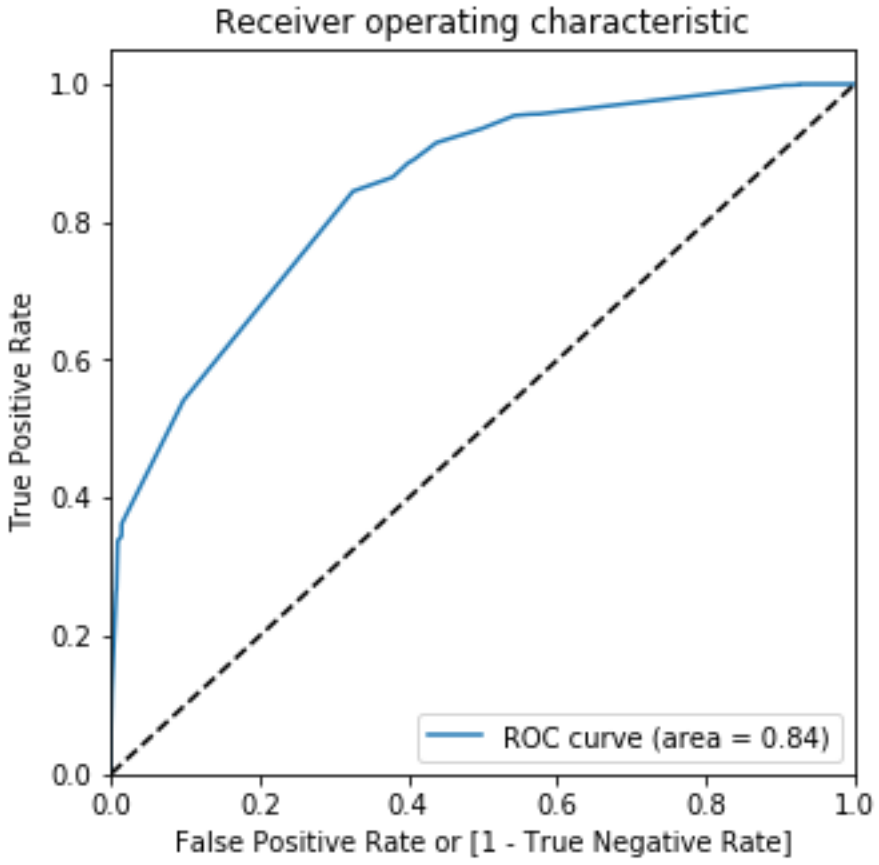
# Random Forest Model

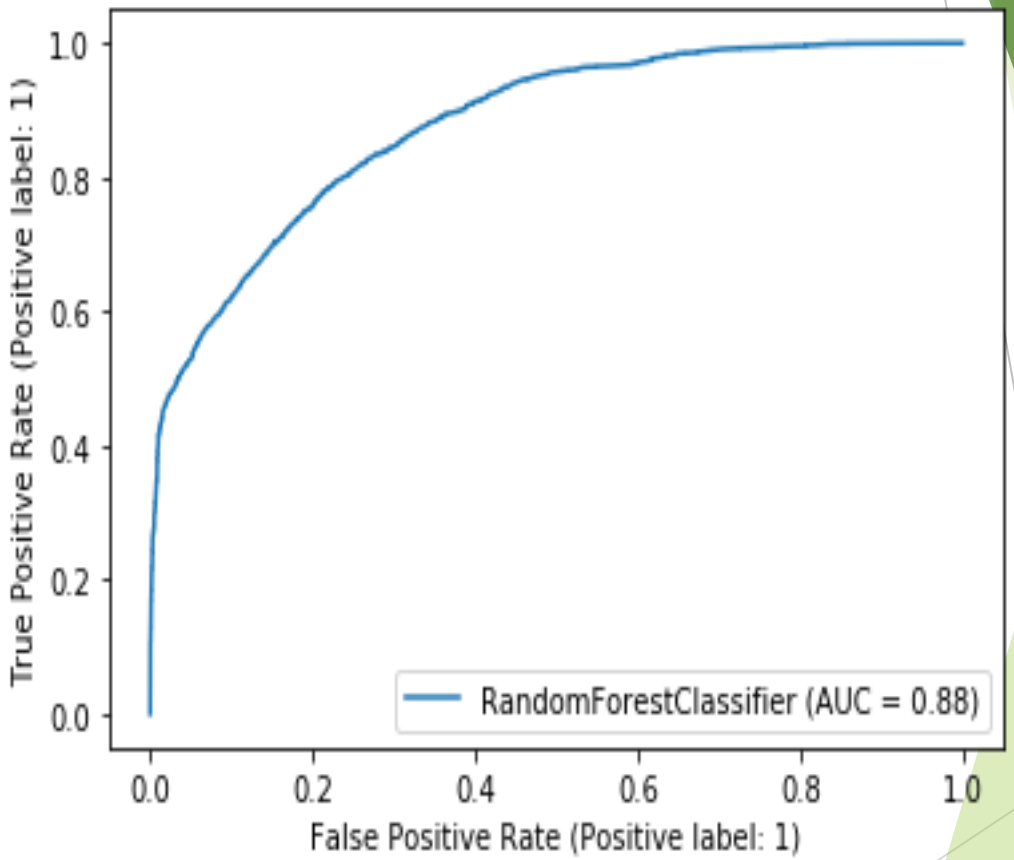
| | VarName | Imp |
|---|---|---|
| 2 | PAYMENT_TERM_30 Days from EOM | 0.412979 |
| 10 | PAYMENT_TERM_Immediate | 0.152829 |
| 33 | DUE_MONTH_12 | 0.068208 |
| 27 | DUE_MONTH_6 | 0.060389 |
| 6 | PAYMENT_TERM_60 Days from EOM | 0.048591 |
| 40 | CLUSTER_ID_1 | 0.045829 |
| 1 | PAYMENT_TERM_15 Days from EOM | 0.032649 |
| 41 | CLUSTER_ID_2 | 0.031238 |
| 0 | USD Amount | 0.028567 |
| 3 | PAYMENT_TERM_30 Days from Inv Date | 0.019036 |
| 13 | INVOICE_CLASS_INV | 0.017282 |
| 32 | DUE_MONTH_11 | 0.012081 |
| 20 | INVOICE_CURRENCY_CODE_SAR | 0.011598 |
| 35 | DUE_DAY_2 | 0.010823 |
| 28 | DUE_MONTH_7 | 0.007255 |

```
Fitting 5 folds for each of 256 candidates, totalling 1280 fits

GridSearchCV(cv=5,
             estimator=RandomForestClassifier(n_jobs=-1, oob_score=True,
                                              random_state=42),
             n_jobs=-1,
             param_grid={'max_depth': [5, 10, 15, 20],
                         'max_features': [5, 10, 15, 20],
                         'min_samples_leaf': [200, 500, 1000, 1500],
                         'n_estimators': [25, 50, 80, 100]},
             verbose=1)
```

# LR V/S Random Forest

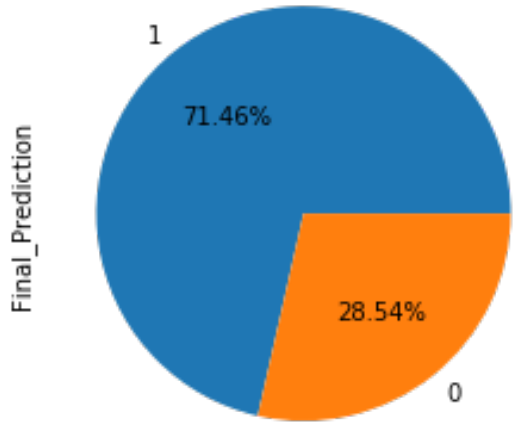


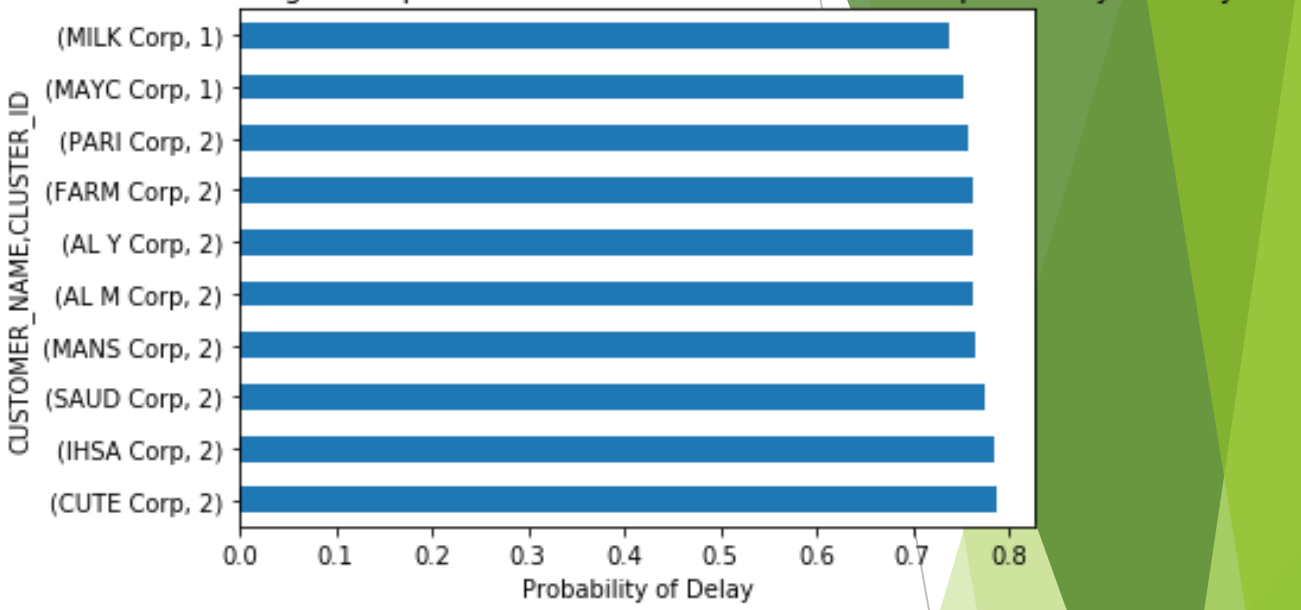

AUC is better for the Random Forest (0.88) than for LR (0.84)

# Deployment

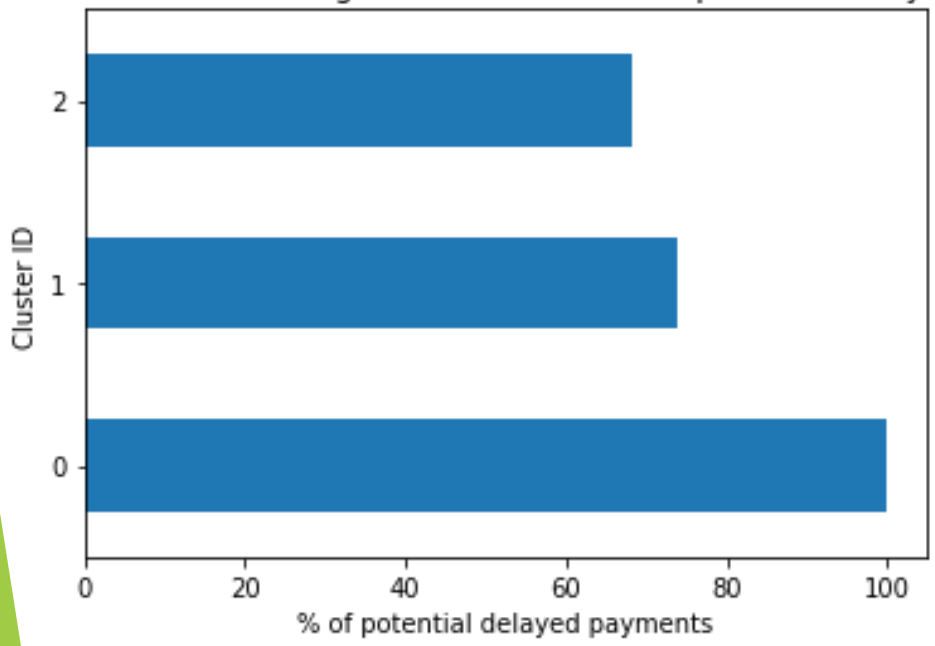
Pie Chart showing the % share of default, 1-Default 0-No Default



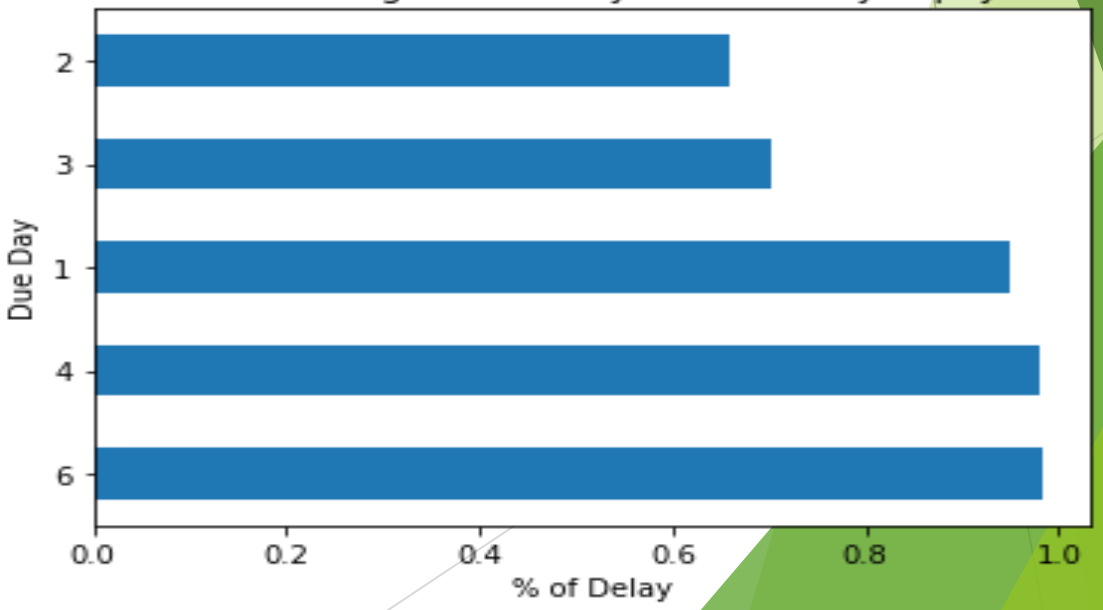

Bar Chart showing the Top 10 Customers with Cluster based on probability of delay



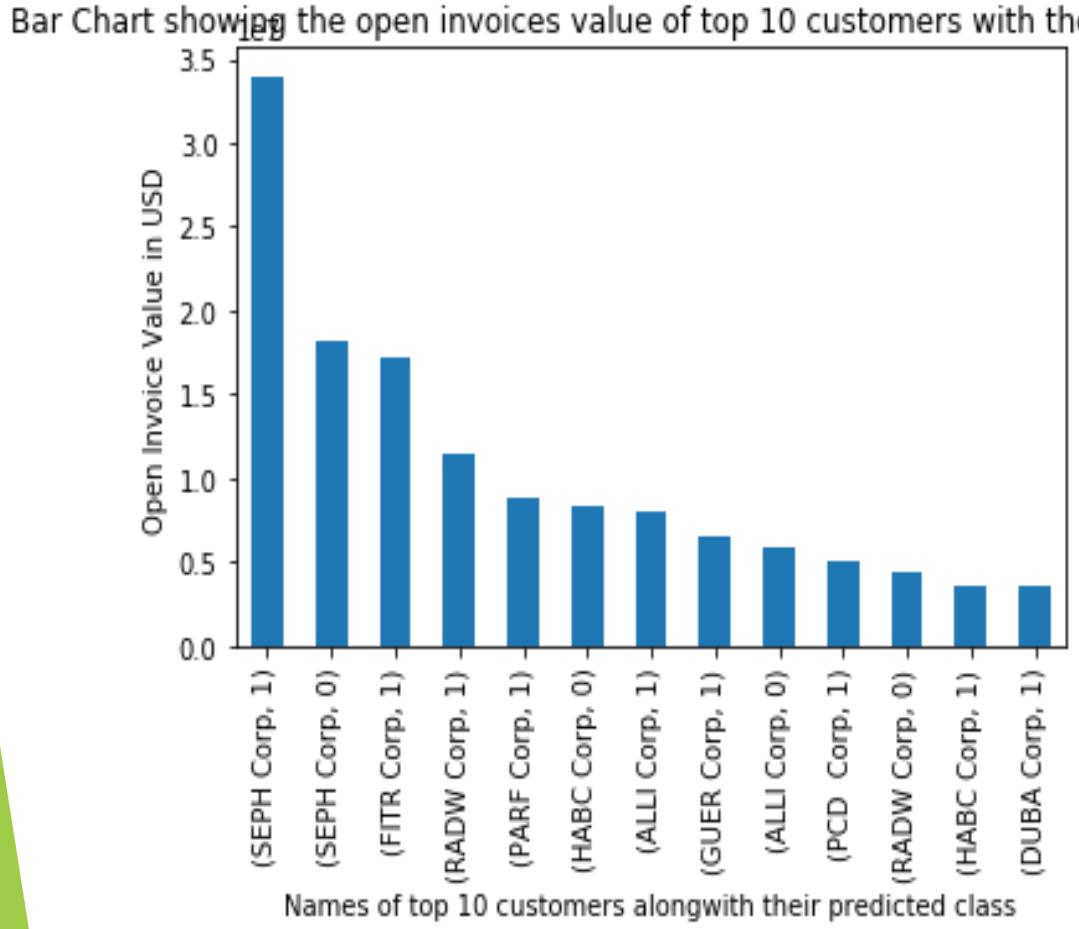

Bar Chart showing the Cluster wise % of potential delays




Bar Chart showing the due day wise % delay in payments



# Customer to Focus




| CUSTOMER_NAME | Number of Open Invoices | % of Predicted Delayed Invoices | Delayed Invoices Count |
|---|---|---|---|
| SEPH Corp | 8260 | 0.660048 | 5452.0 |
| FITR Corp | 3454 | 0.866532 | 2993.0 |
| PARF Corp | 1717 | 0.845079 | 1451.0 |
| AREE Corp | 1117 | 0.482543 | 539.0 |
| ALLI Corp | 1042 | 0.435701 | 454.0 |
| ... | ... | ... | ... |
| ST R Corp | 3 | 0.000000 | 0.0 |
| JIHA Corp | 2 | 0.000000 | 0.0 |
| J A Corp | 1 | 0.000000 | 0.0 |
| SONO Corp | 1 | 0.000000 | 0.0 |

# Recommendations

- Focus on Cluster 0 it has prolonged delays
- Dedicte people for follow up as volume of invoices is huge
- Invoices due in Mar , Jul-Oct and due on Friday and Sunday have higher chances of default
- Invoicing currency is ZAR,HKD,SGD,OMR and GBP should also be kept and eye on
- SEPH, FITR, PARF should be checked as their Delay invoice count is high