E-Commerce and Retail B2B Case Study

ROHITH VADDETI

Addressing the issue and defining objectives

Problem identification •

- A sports retail company Schuster dealing in B2B transactions often deals with vendors on a credit basis, who might or might not respect the stipulated deadline for payment
- Vendors delaying their payments result in financial lag and loss which becomes detrimental to smooth business operations
- Additionally, company employees are set up chasing around for collecting payments for a long period of time resulting in no value-added activities and wasteful resource expenditure

Business Objectives

- Customer segmentation to understand the customer’s payment behavior
- Using historical information, the company requires prediction of delayed payment against an unforeseen dataset of transactions with due date yet to be crossed
- The company requires the prediction for better resource delegation, quicker credit recovery and reduction of low value-adding activities

Class imbalance and transaction insights (univariate)

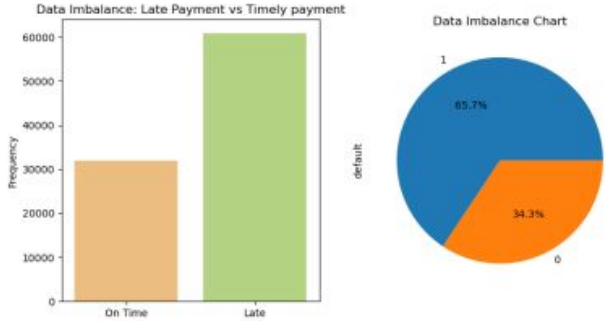


Fig. 1

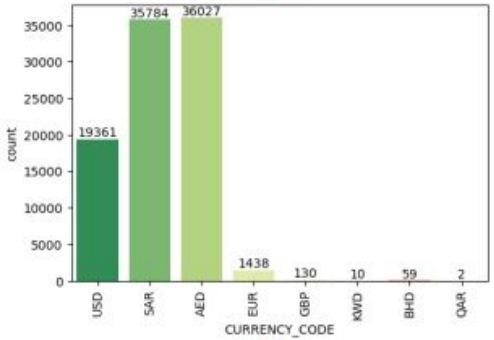
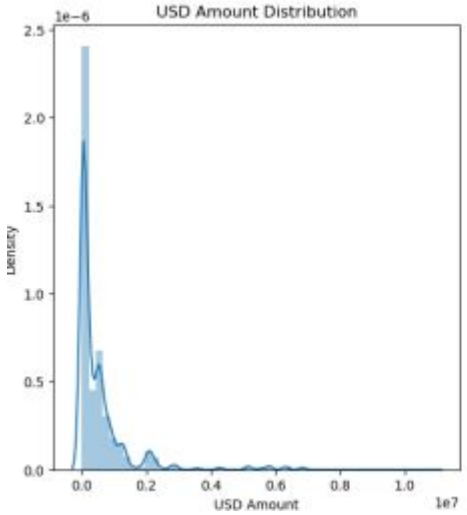


Fig. 2



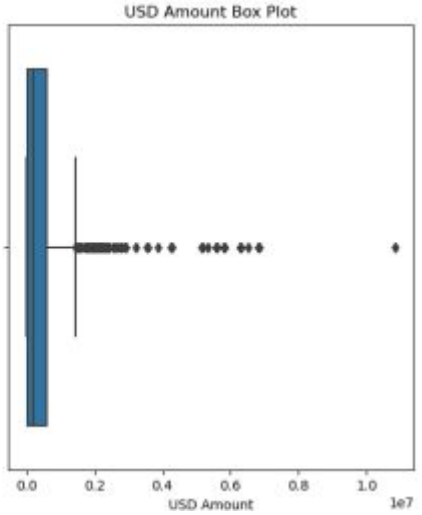


Fig. 3

From Fig. 1 and 2:

- The class imbalance is 65.7% towards payment delayers which is an acceptable imbalance and does not need imbalance treatment
- The top three currencies in which the company deals are AED, SAR and USD with AED as the most dealt currency suggesting greater transactions with the middle-east

From Fig. 3: we observe,

- The transaction values seem to lie between a range of $1 and $3m
- The transaction values are most frequent below ~$1.75m

Class imbalance and transaction insights (univariate)

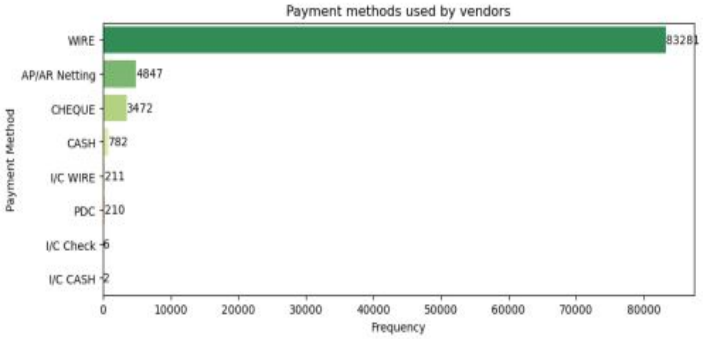


Fig. 1

From Fig. 1, we observe, Wire payment method is the most common payment method received by the company, followed by netting , cheque and cash

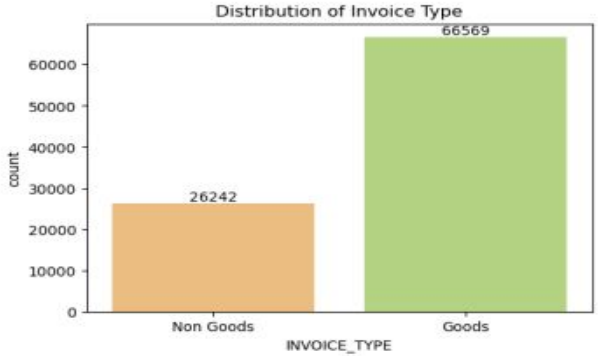


Fig. 2

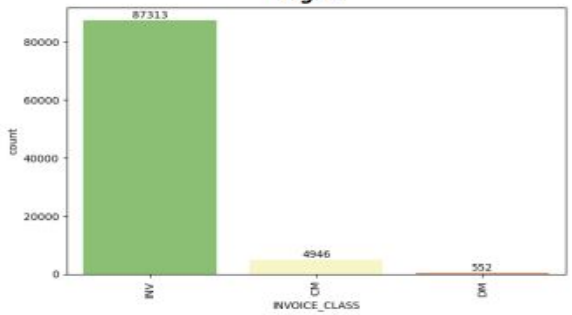


Fig. 3

From Fig. 2 and 3:

- Goods type invoices comprise of the major share of invoices generated
- The major invoice class is ‘Invoice’ with the rest having very low percentages of the share

Identifying characteristics of defaulter payment types (Bivariate)

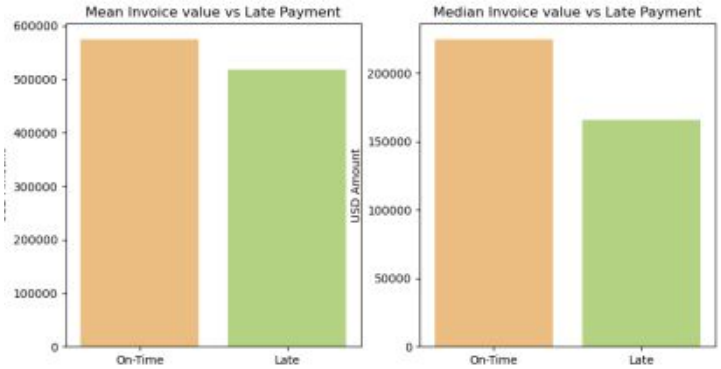


Fig. 1

The mean and median of the payment amount is higher for payers who pay on time than late, suggesting that higher value transactions show lesser delay risk than lower value transactions

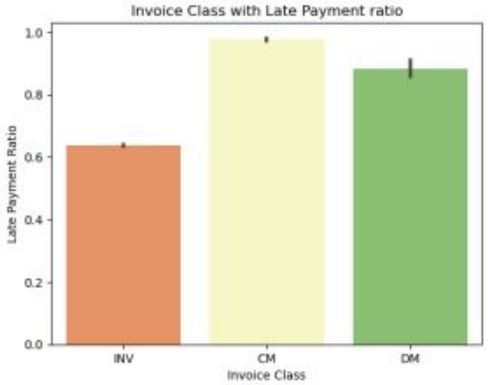


Fig. 2

late payment ratio for Credit Note transaction types are maximum, followed by Debit Note and Invoice suggesting higher delay risk in Credit and Debit note invoice classes

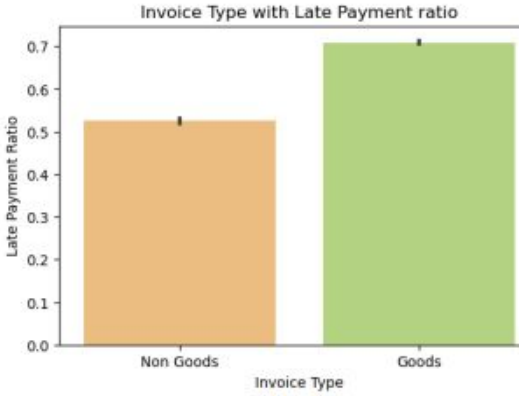
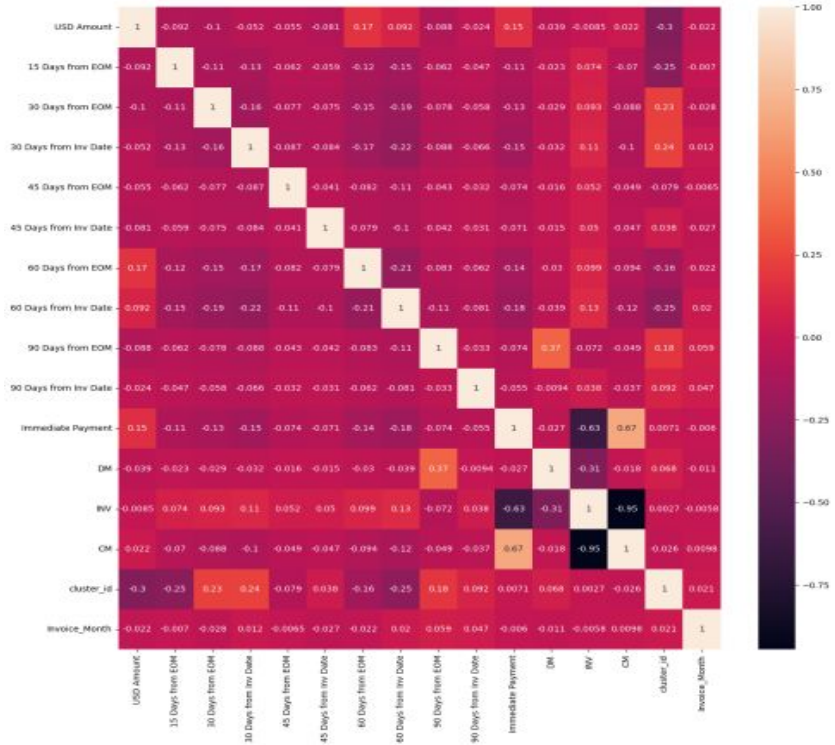


Fig. 3

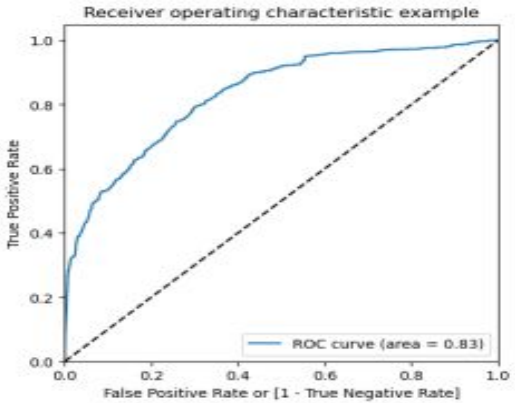
Goods type invoices show greater late payment ratio than non goods hence showing increased chances of payment delay

Model Building

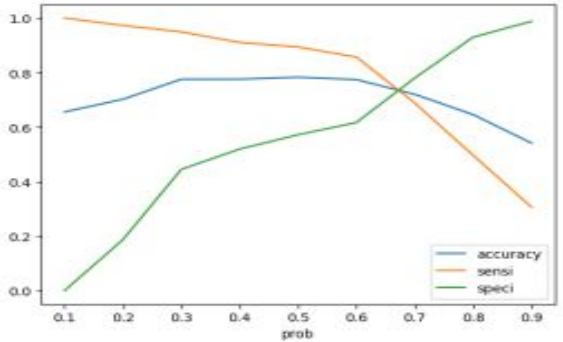


CM & INV, INV & Immediate Payment, DM & 90 days from EOM has high multicollinearity, hence dropping these columns to prevent multicollinearity effect

Comparison between two models, logistic regression and random forests

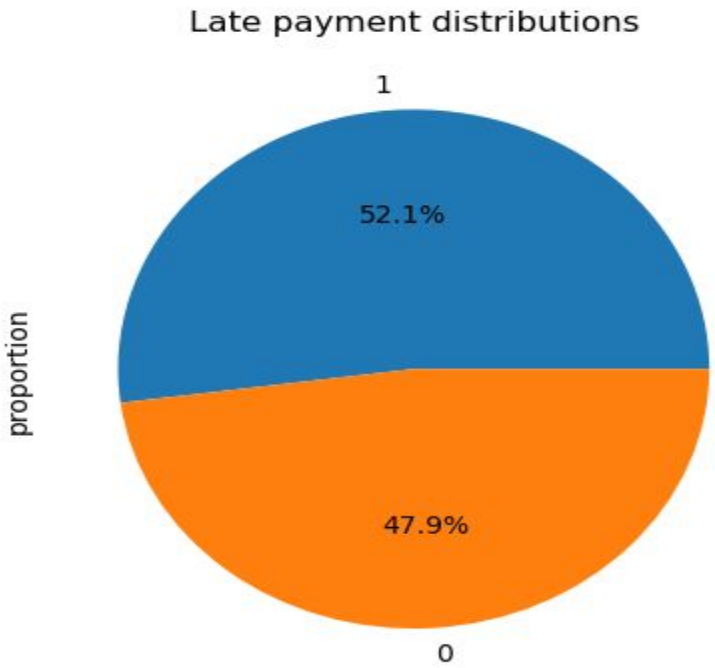


- Logistic regression model formed after dropping multicollinearity and unnecessary variables resulted in remaining variables with acceptable p-value and VIF figures, hence retained the remaining features with no further feature elimination and a good ROC curve area of 0.83

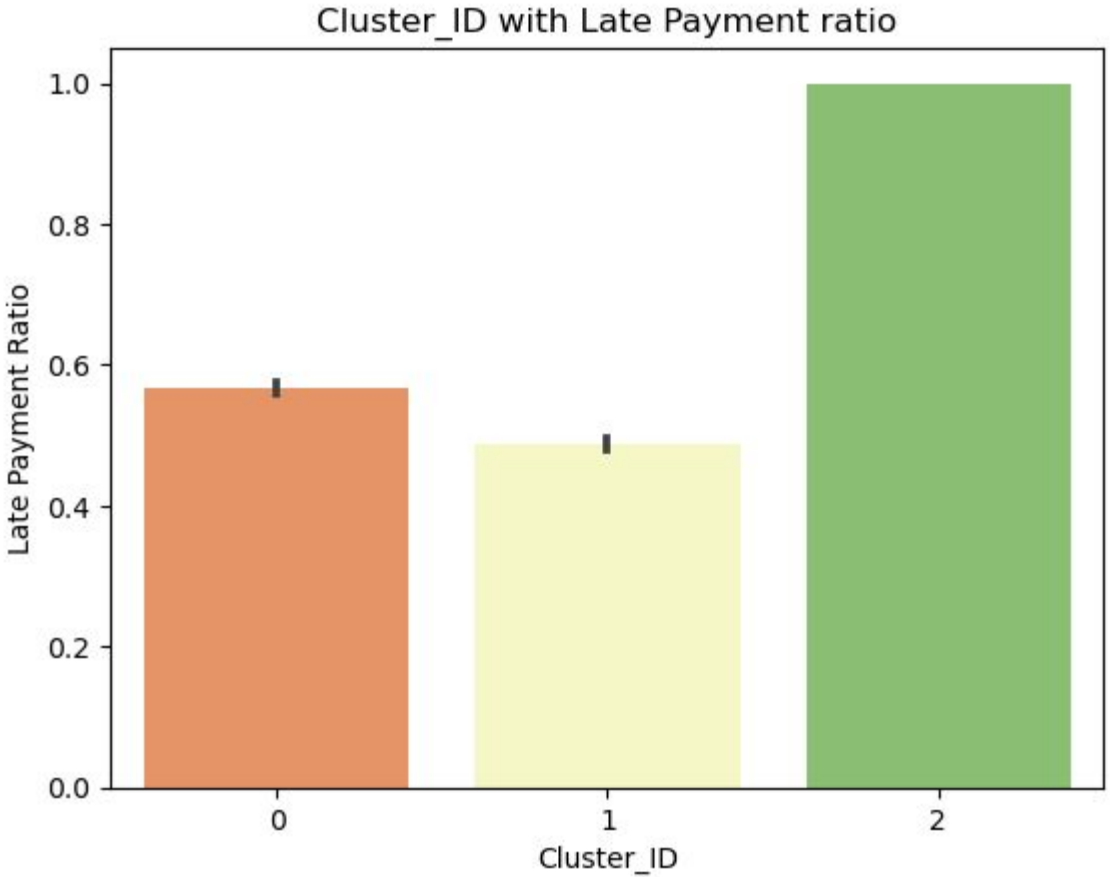


- The trade-off plot between accuracy, sensitivity and specificity revealed an optimum probability cutoff of ~0.6, which was used to further predict which transactions would result in delayed payments in the received payments dataset

52% payments predicted to be delayed as per Openinvoice data, prolonged payment days to observe alarmingly high delay rates



Predictions made by the final model suggests that there is a probable 52.1% transactions where payment delay can be expected, which can cause a shocking lag to business operations



Customer segment with historically prolonged payment days are anticipated to have the most delay rate (~100%) than historically early or medium days payment transactions, this is similar to the result found based on historical outcomes