SocBiz IITR Analytics

December 2023

Credit Card Default Prediction

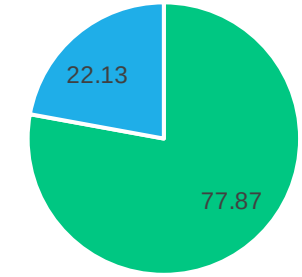
Dhruv Agrawal

Executive Summary

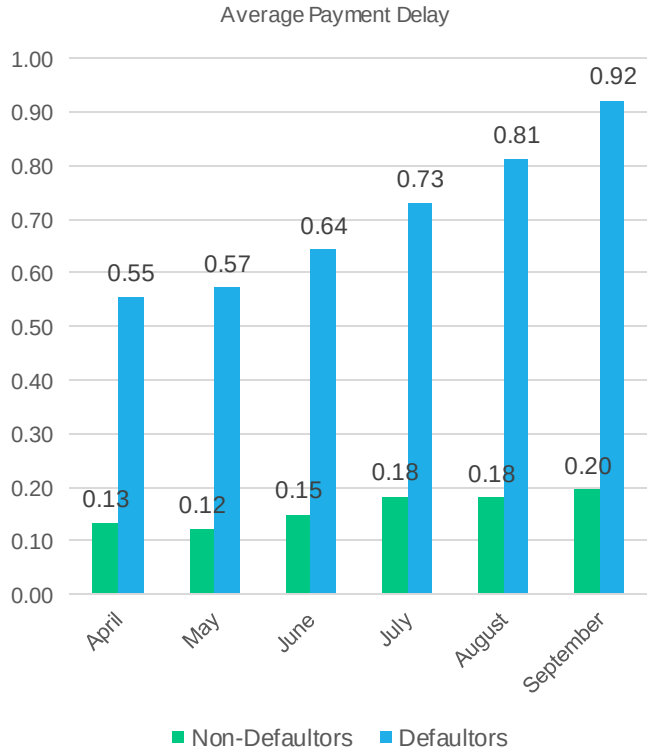
- The dataset portrays diverse financial behaviours and a significant proportion of defaulters, indicating potential challenges due to class imbalance and varying payment delay patterns influencing credit defaults.
- The dataset underscores the impact of payment behavior on credit defaults, highlighting higher bill amounts among males, although females, despite being more numerous, exhibit lower default rates and a consistent increase in billing from April to September for both genders
- Default rates exhibit an escalating trend with increasing age, while married individuals consistently display higher bill amounts, potentially indicating greater household financial responsibilities contributing to elevated billing over consecutive months.
- The dataset demonstrates a considerable imbalance in class distribution, influencing the preference for F1 score over accuracy in model evaluation, where Decision Tree with Cross Validation achieves the best performance.

The dataset portrays diverse financial behaviours and a significant proportion of defaulters, indicating potential challenges due to class imbalance and varying payment delay patterns influencing credit defaults.

	Mean	Standard Deviation	Description
Limit Balance	167k	78% of Mean	Significant dispersion of credit amounts around the mean.
Pay Status(Sept)	0.35	0.76	Significant average delay in payments among individuals, alongside a wide dispersion of payment delay durations.
Amount(Sept)	51k	143% of Mean	Denotes considerable dispersion in bill amounts among individuals.
Default Rate	0.22	0.41	Significant proportion of credit card users are at risk of defaulting on payments.

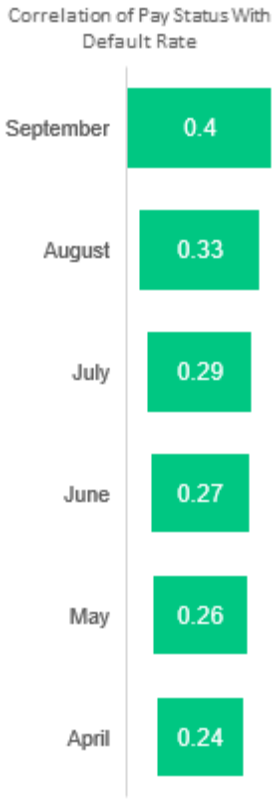


- The dataset exhibits a **substantial class imbalance**.
- This class imbalance may impact the model's predictive performance, potentially leading to biased learning patterns.

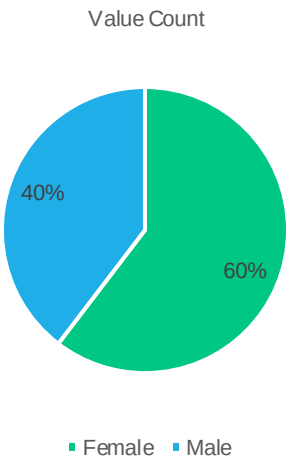


- **Defaulters** exhibit a notably **higher average delay in payments** compared to **non-defaulters**.
- This difference in payment delay durations might signify a critical behavioral pattern, suggesting that prolonged payment delays could serve as a significant indicator to the likelihood of defaulting on credit card payments

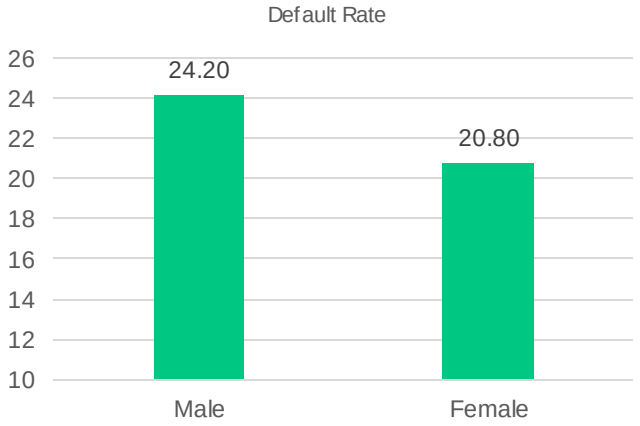
The dataset underscores the impact of payment behavior on credit defaults, highlighting higher bill amounts among males, although females, despite being more numerous, exhibit lower default rates and a consistent increase in billing from April to September for both genders



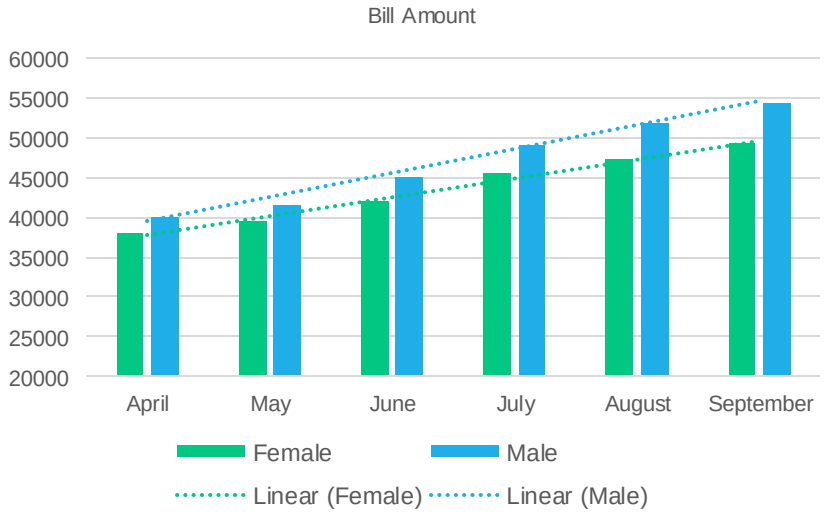
As the payment delays increase over time, there is a corresponding escalation in the correlation with default rates, emphasizing the crucial influence of payment behaviour on credit defaults.



In the dataset, **Females were more in number than males** suggesting that more females tend to use credit cards than males.

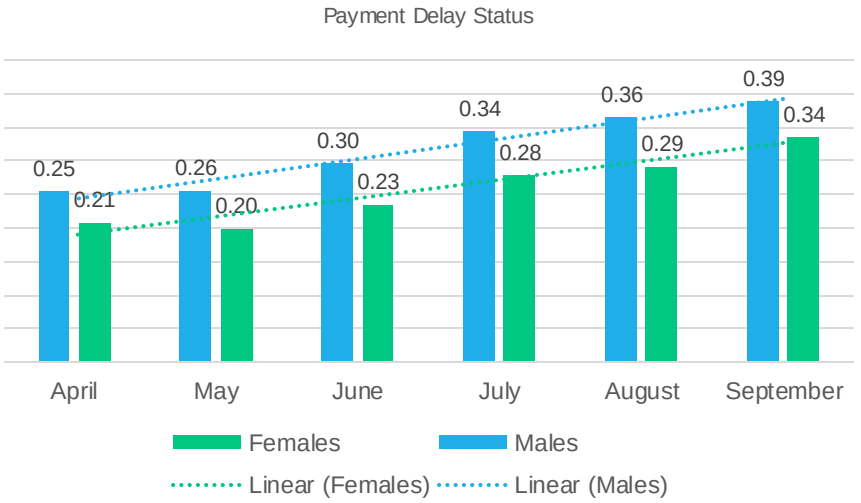


Despite females being more numerous in number, their **default rate is less than males**.

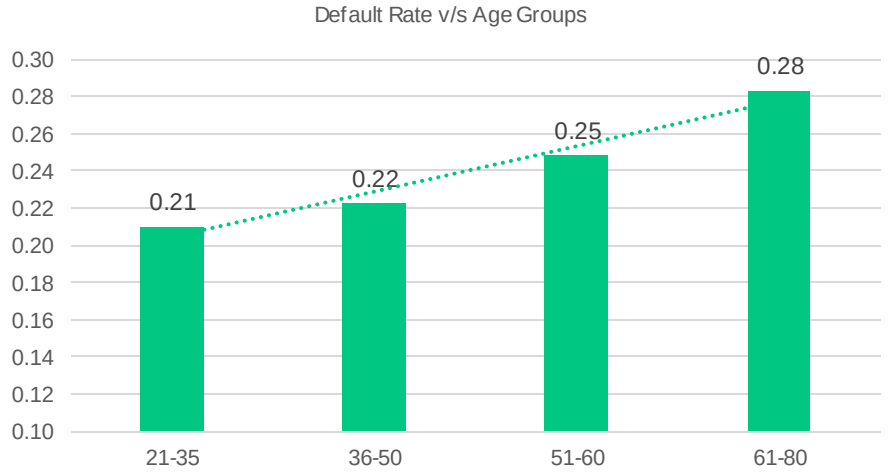


- In this dataset, **males tend to have higher bill amounts** than females during specific months.
- Additionally, both **genders show an increasing trend in bill amounts** from April to September, indicating a consistent rise in billing over successive months.

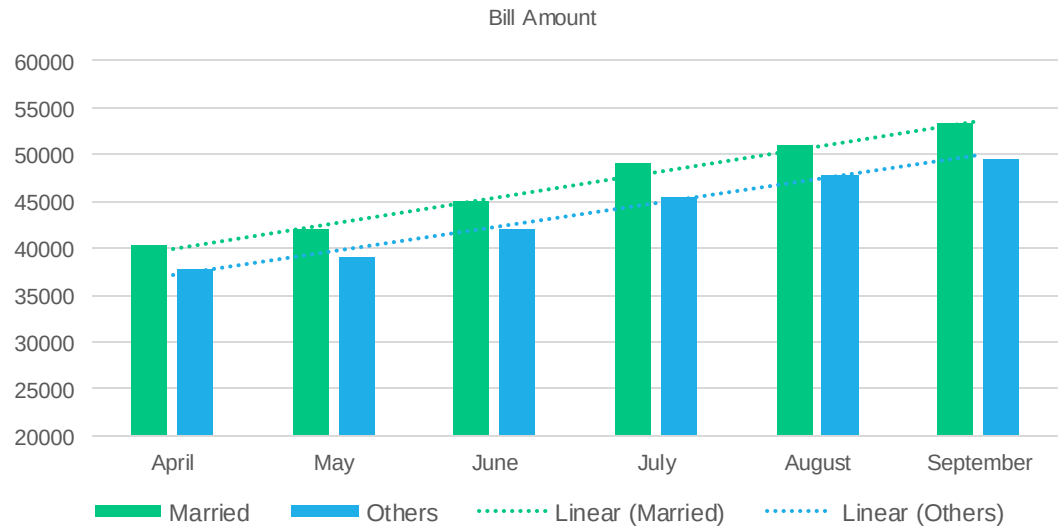
Default rates exhibit an escalating trend with increasing age, while married individuals consistently display higher bill amounts, potentially indicating greater household financial responsibilities contributing to elevated billing over consecutive months.



The dataset reveals that **males tend to have higher payment delay** statuses compared to females, showing a consistent increase for both genders from April to September.

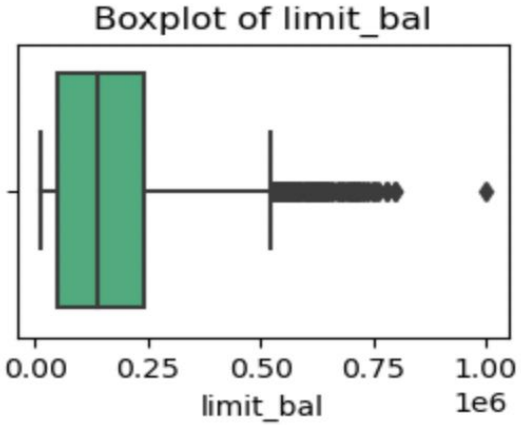


The analysis highlights a trend where **default rates** demonstrate an **increase with advancing age groups**.



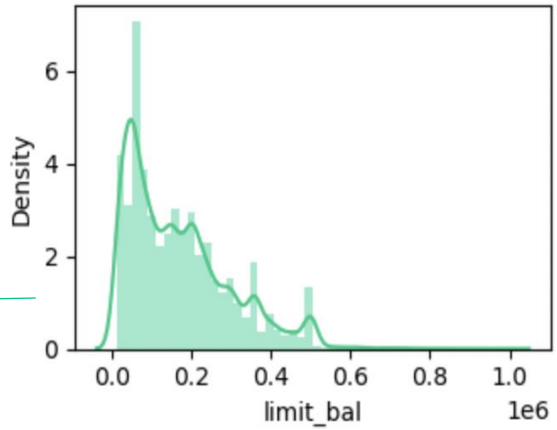
- The analysis reveals that **married individuals consistently exhibit higher bill amounts** than others, showing a continual increase in billing from April to September.
- This trend might be attributed to shared financial responsibilities or larger household expenditures among married individuals, leading to higher bill amounts over time.

The dataset demonstrates a considerable imbalance in class distribution, influencing the preference for F1 score over accuracy in model evaluation, where Decision Tree with Cross Validation achieves the best performance.



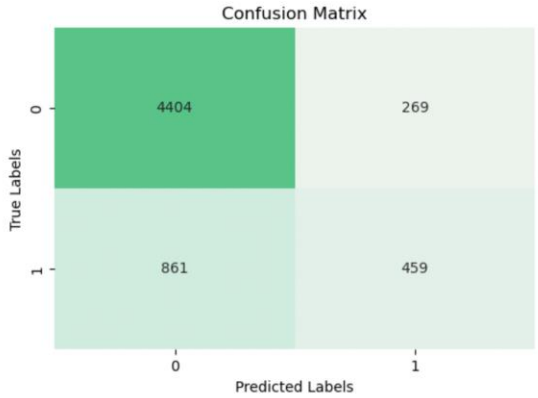
There are a **large number of outliers** in the dataset.

Due to the **skewed data distribution** of most features, **IQR technique** is used to remove the outliers.



Model	Accuracy	F1 Score
kNN	0.752545	0.719248
kNN with Cross Validation	0.769106	0.715607
Logistic Regression	0.752545	0.715607
Logistic Regression with Cross Validation	0.641749	0.667501
Decision Tree	0.720174	0.722907
Decision Tree with Cross Validation	0.811447	0.789811
SVM	0.779743	0.683244
Naive Bayes	0.370265	0.359950

- The dataset has high class imbalance, **F1 score provides more reliable evaluation** of the model's performance compared to accuracy.
- **Decision Tree with Cross Validation** gives the highest accuracy and f1 score among all models.



Accuracy: 81.1%

F1 Score: 78.9%

Recall: 34.7%

Precision: 63%

Thank You