Credit Approval Project

By: Isabel Ernzen, Claire Linn, and Cameron Porath

Introduction and Problem

- Credit approval: asking to borrow money from someone (bank) and the bank decides if they trust you to pay it back
- Purpose: gain a better understanding of the factors that influence credit approval
- Identify the trends that explain why some applications are accepted and some are denied
- Increase the consistency and equity of credit approval

Data Description

Columns

Financial
- Loan Amount
- Loan Duration
- Interest Rate
- Credit Score
- Monthly Payment Amount

Personal
- Age
- Income
- Employment Type (contract/ permanent/ self-employed)
- Seniority

- Repayment Status (ongoing/ paid off)
- Average Balance
- Credit Approval
- Approval Rate
- Loan Type

- Drop 'Credit approval' because it's the target we're trying to predict, and drop 'Approval Rate' since it's strongly correlated with the target and adds redundancy

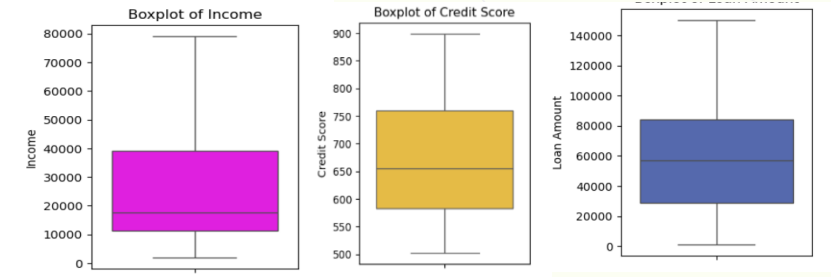
- We converted the 'Loan Type', 'Repayment Status', and 'Employment type' since they are categorical to binary so each sting input was its own column (Loan type: auto, personal, mortgage)
- The correlation to credit approval was not strong enough so we dropped them

Analysis and Modeling

Boxplot

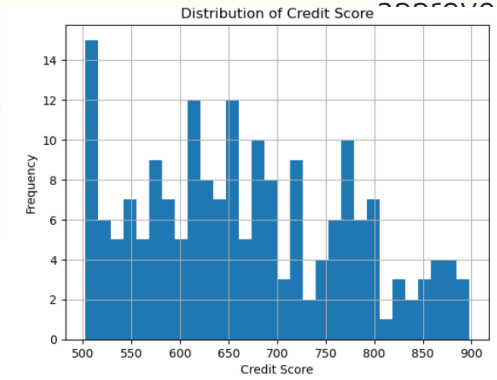
- Analyse Loan Amount, Income, and Credit Score
- Loan Amount and Income are widespread indicating significant variability
- Credit Score is more tightly grouped, less variation
- Helps to find unusual values(outliers)

Overall, Loan Amount and Income are more variable and contain more outliers, while Credit Score is more consistent across applicants. This tells us where more complex decision-making may be needed and where patterns might be easier to detect



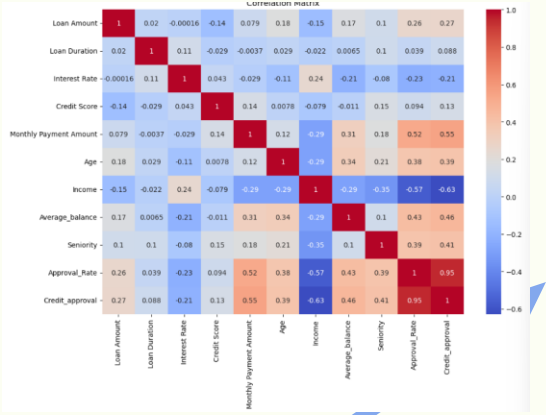
Histogram

- Following the credit score boxplot
- Shows how credit scores are distributed among all of the applicant's
- X-axis = credit score range
- Y-axis – how many people fall in each range
- Most have scores between 500 and 800



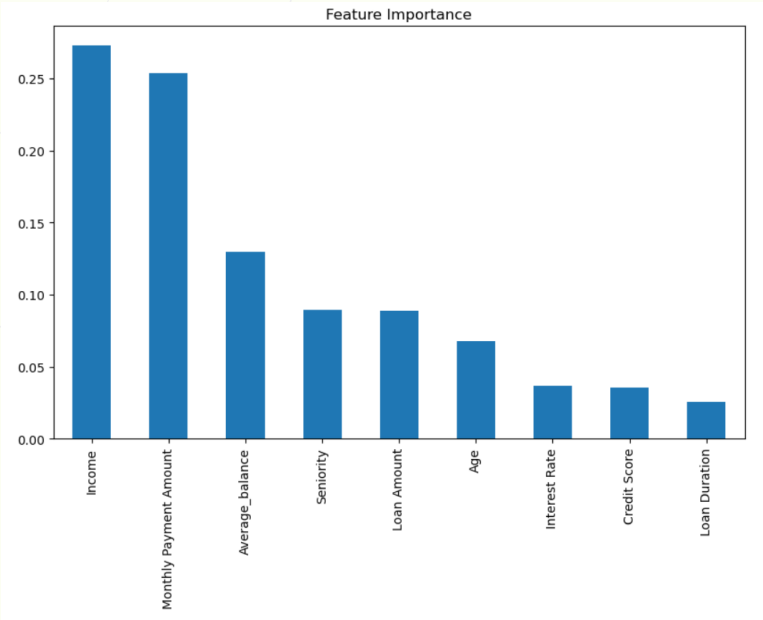
Correlation Heatmap

- Shows how different number-based columns in the dataset are related
- Red = values increase together
- Blue = one goes up while the other goes down
- White = not much connection
- We removed approval rate from data after looking at the heat map because they were highly correlated and it made it more complex
- Income has a negative relationship which means people with lower income were more likely to be approved



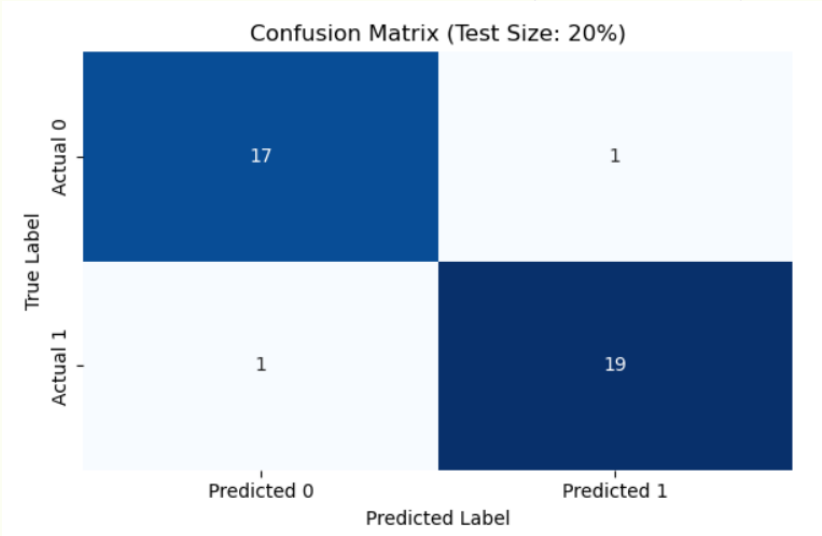
Results and Evaluation

- It showed strong performance in identifying patterns that influence credit approval decisions
- The feature importance graph shows the top features influencing the model's decisions
    - Income
    - Monthly Payment Amount
- This information can guide policy or criteria refinement based on data driven insights



Conclusion and Insights

- The confusion matrix analysis helped in identifying the trade-offs between false approvals and false rejections
- The data quality at the beginning wasn't good so we made sure to clean it up and get rid of duplicates
- The accuracy for the confusion matrix is based on 20% test size
- Exploring other classifiers could be interesting to see what other facts go into credit approval
    - Marital Status
    - Education Level
    - Security of Loan




```
Top Feature Model Evaluation:
Confusion Matrix:
 [[17  1]
 [ 1 19]]

Accuracy: 0.95

Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.94      0.94        18
           1       0.95      0.95      0.95        20

    accuracy                           0.95        38
   macro avg       0.95      0.95      0.95        38
weighted avg       0.95      0.95      0.95        38
```