



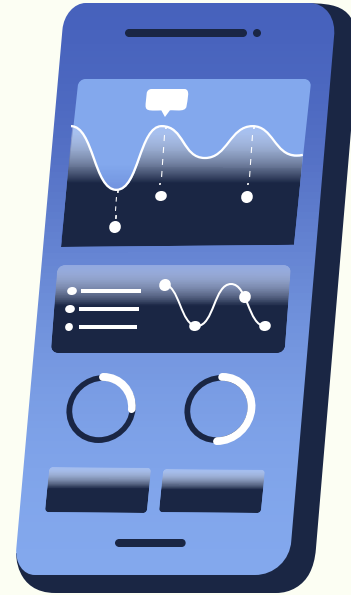
Credit Approval Project

By: Isabel Ernzen, Claire Linn, and Cameron Porath



Introduction and Problem

- 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

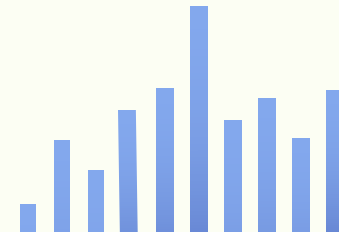
Financial

- 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

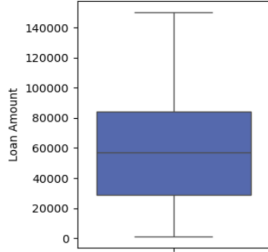


Analysis and Modeling

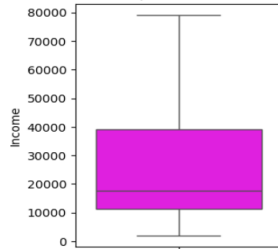
Boxplot

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

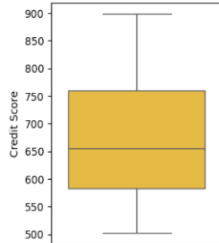
Boxplot of Loan Amount



Boxplot of Income



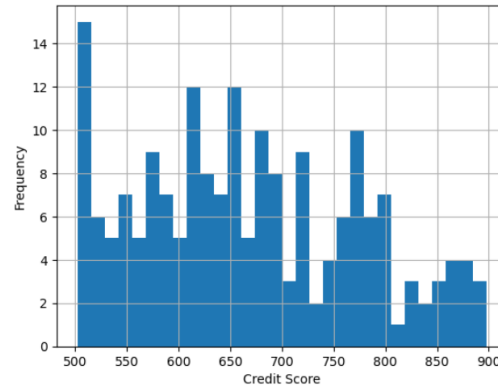
Boxplot of Credit Score



Histogram

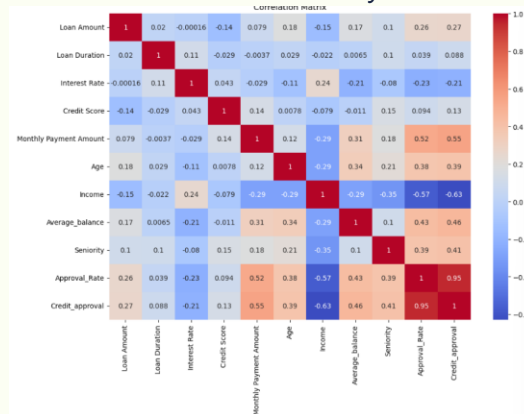
- Shows how credit scores are distributed among all of the applicant's
- X-axis = credit score range
- Y-axis – how many people fall in each range
- Most have scores between 500 and 800

Distribution of Credit Score



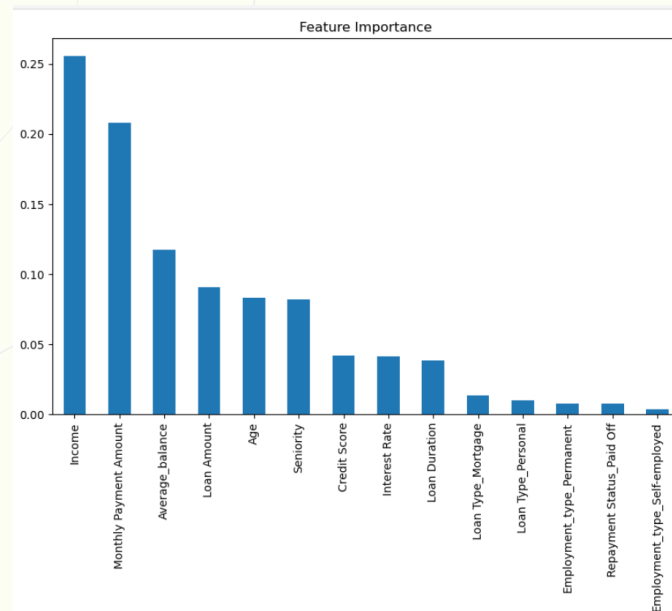
Correlation Heatmap

- 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
 - o Income
 - o 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
 - o Marital Status
 - o Education Level
 - o 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

