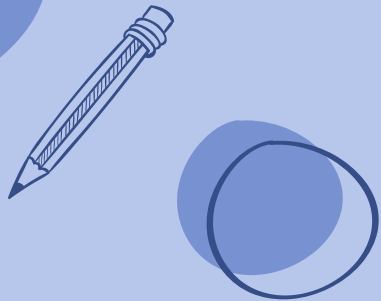


Predicting Loan Defaults

By Stephanie Duncan
11th March 2021





CONTENTS

- Introduction to Data & Client Business Question
- Data Cleaning & Challenges Encountered
- Data Analysis
- Predictive Models
- Summary of Findings
- Questions



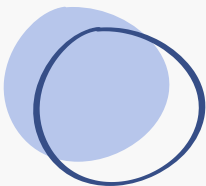
Introduction to Data

- Provided by an American online loan provider
- Over 5 year period: 2007 - 2011
- Client is concerned about default rate on their loans
 - Trend by Employment Length
 - Home Ownership
 - Fico Score
- Who should be lend to in the future?



Data Cleaning

- Project was carried out in Python
- Bias - zip code was removed to avoid bias by location
- Missing Values
 - Many rows/columns fully missing - removed
 - Employment length
- Over 100 variables - removed those with single values

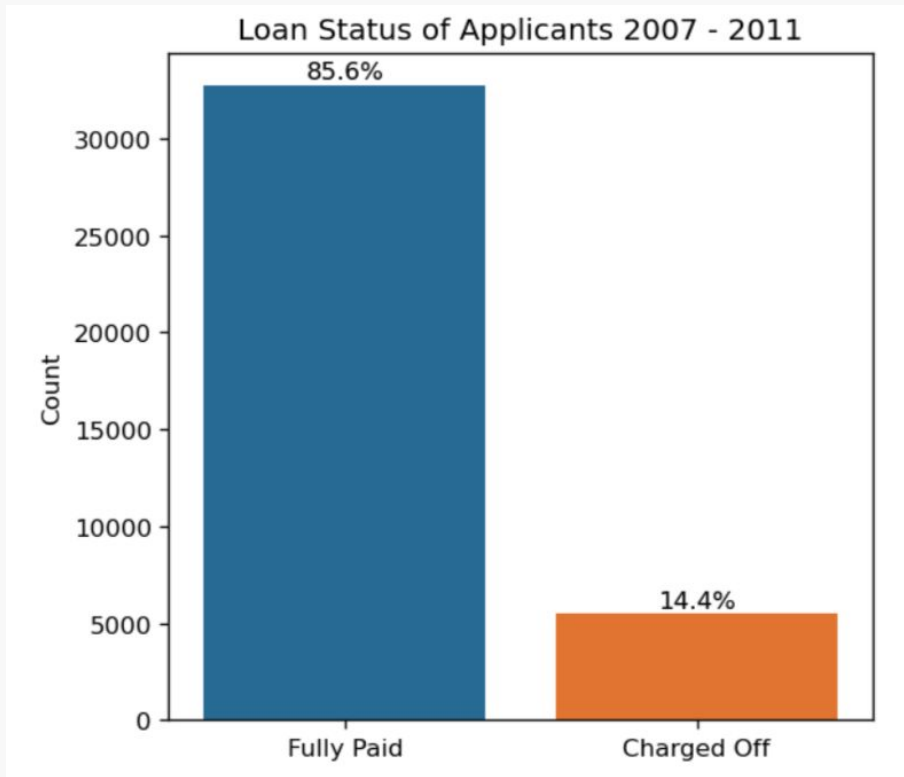




DATA ANALYSIS

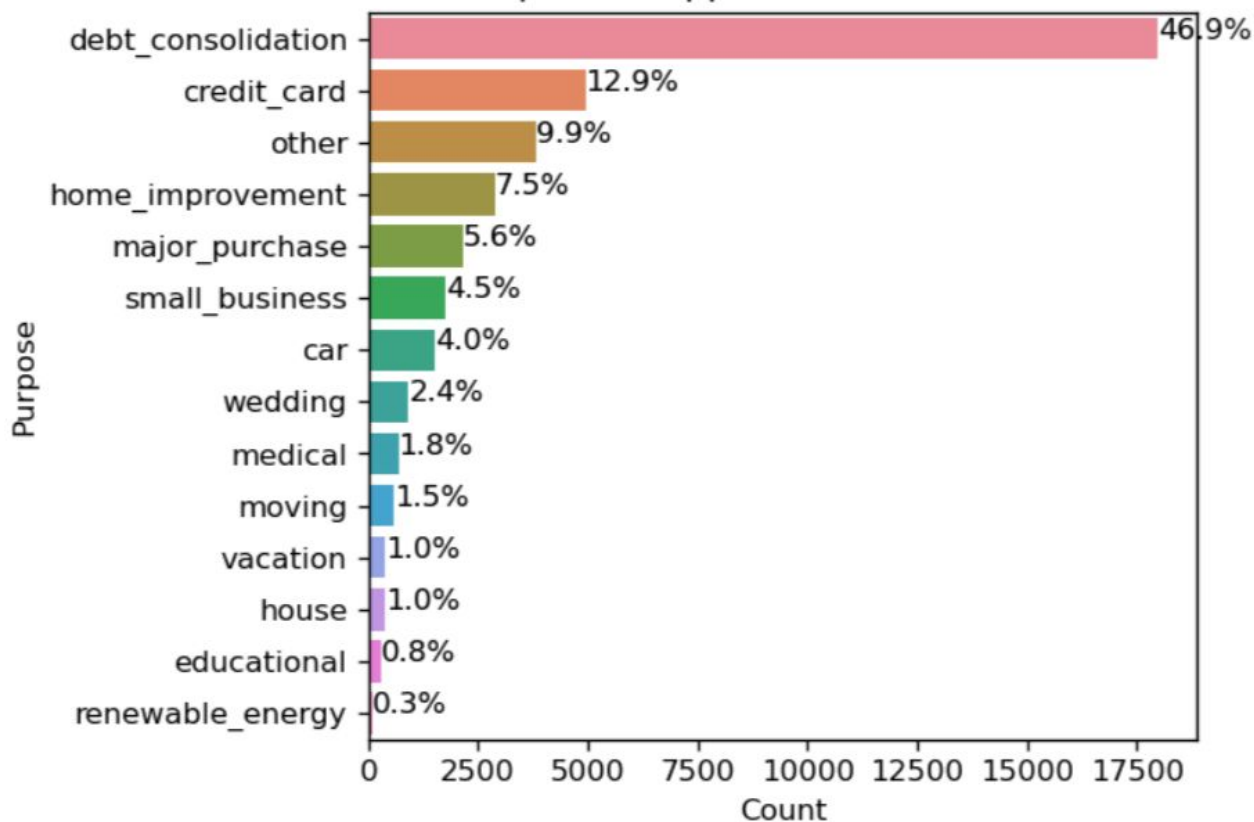


Loan Status

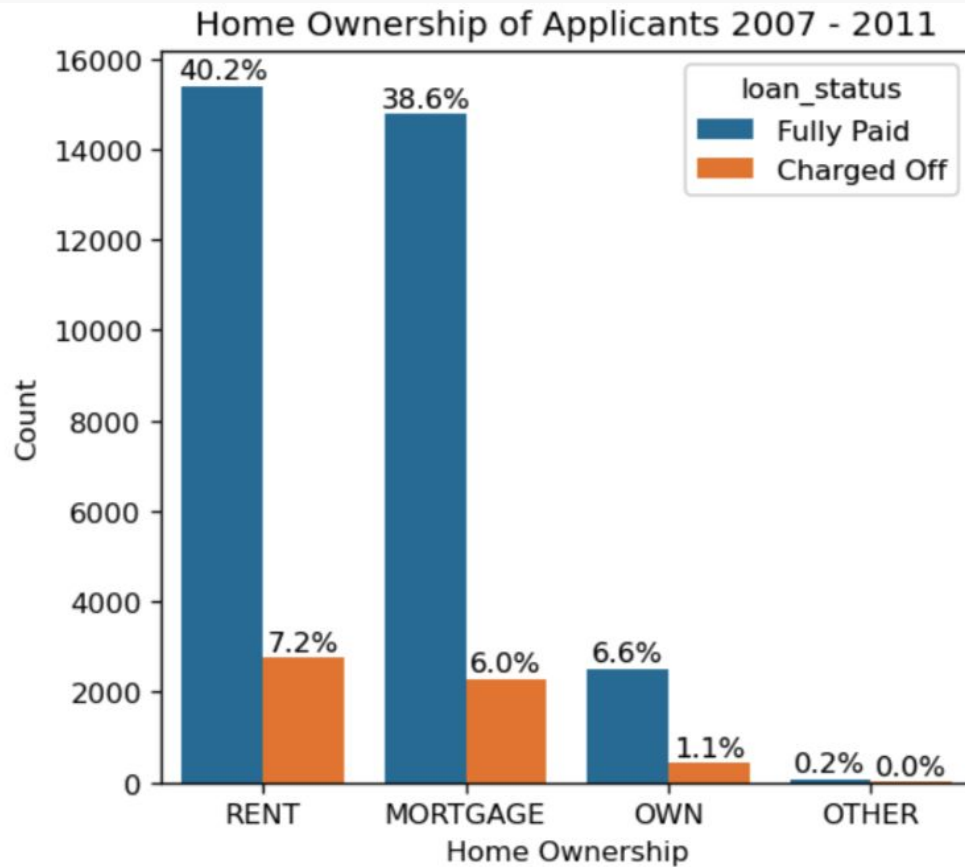


Purpose of Application

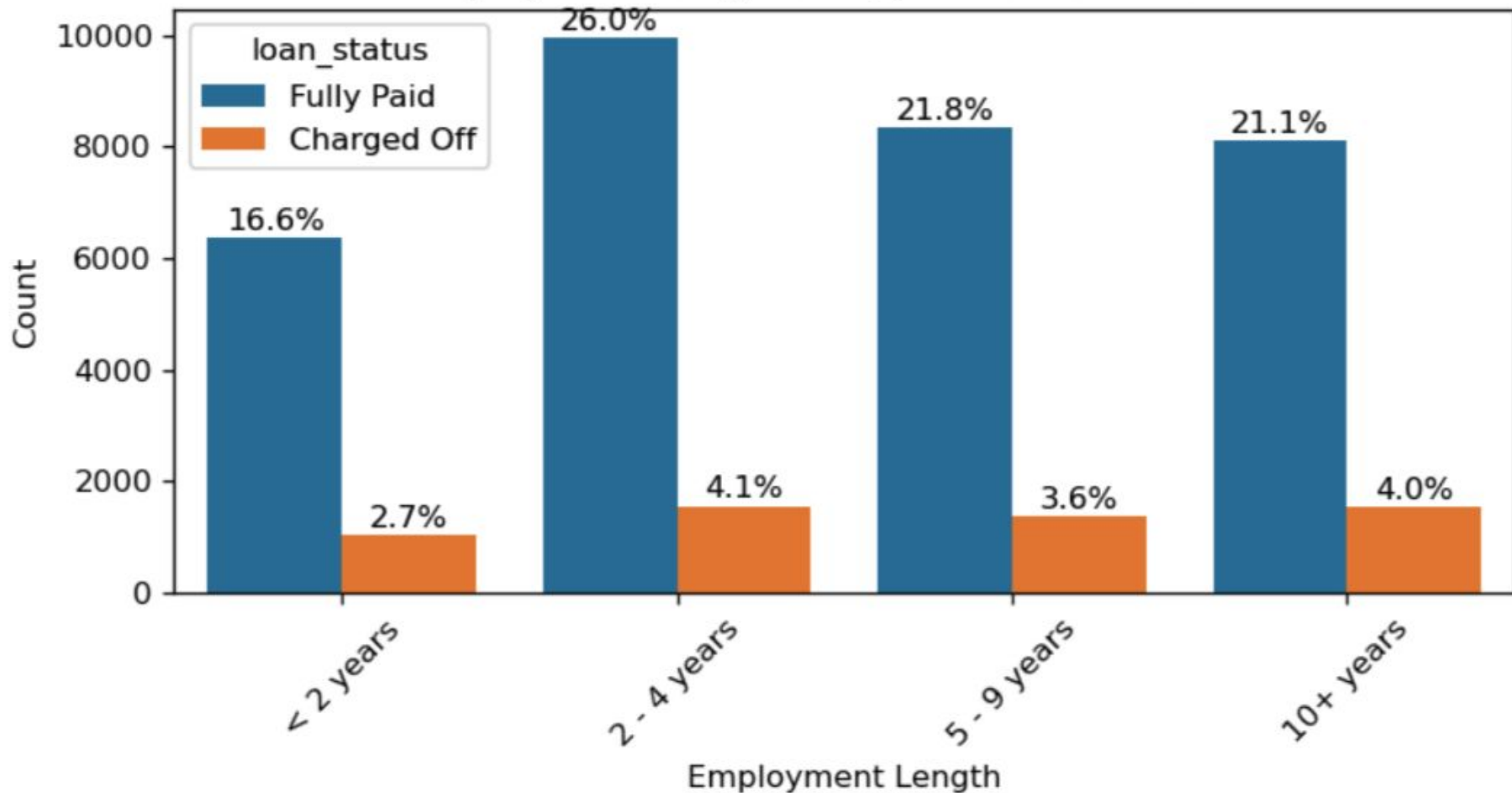
Purpose of Applications 2007 - 2011



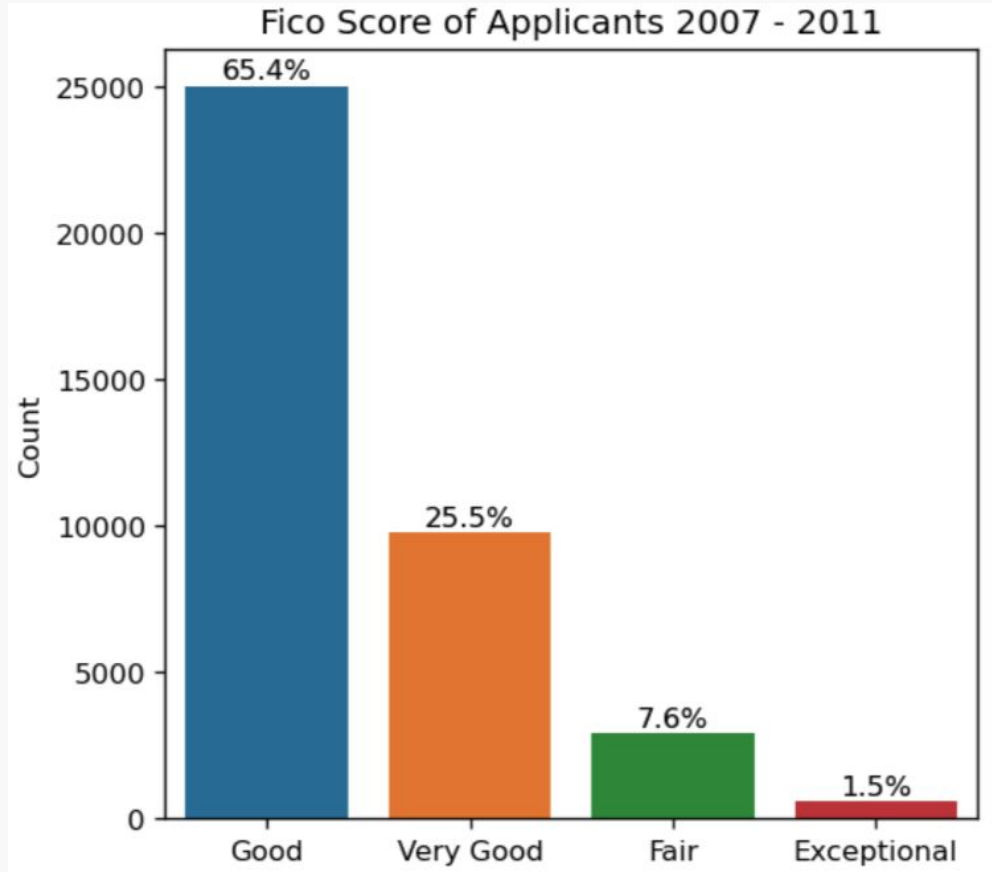
Home Ownership



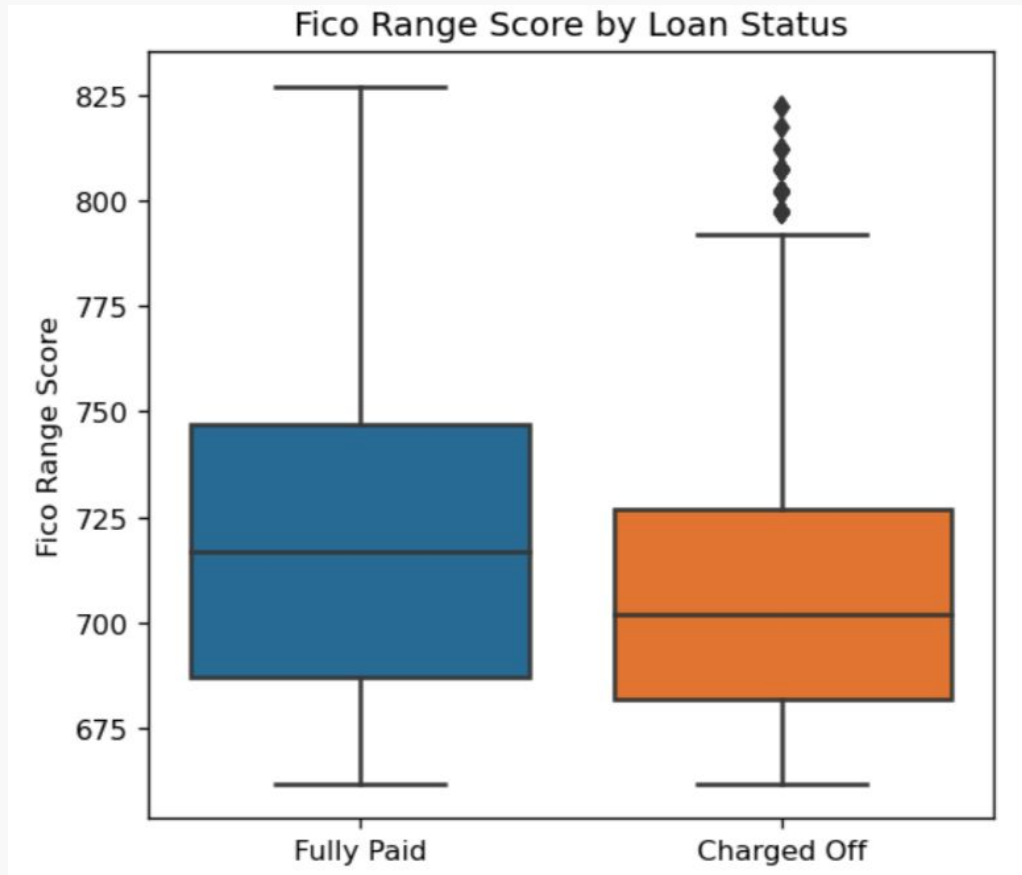
Employment Length of Applicants 2007 - 2011



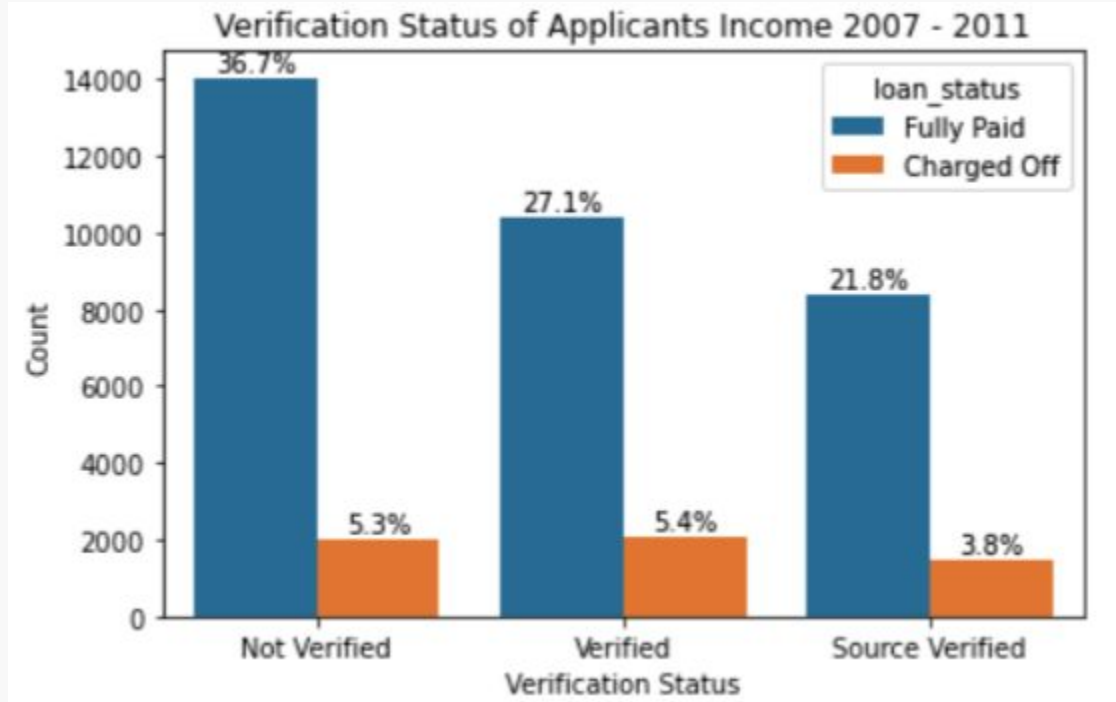
Fico Score



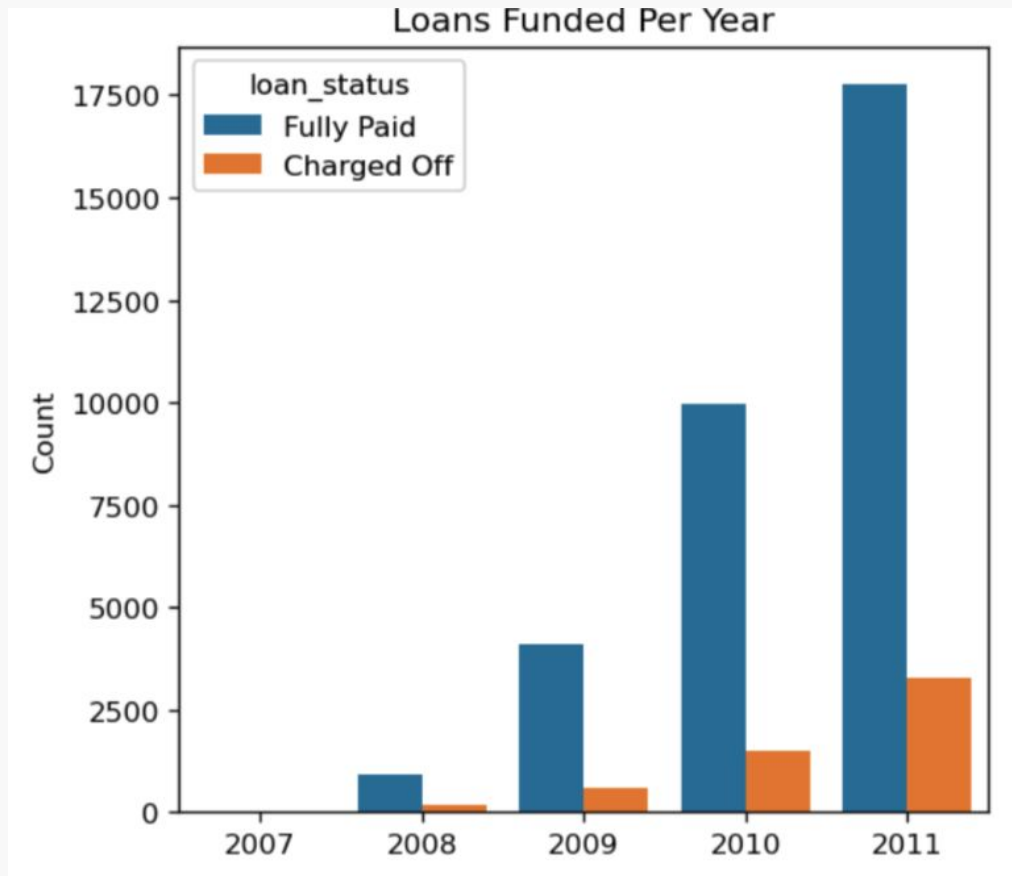
Fico Score



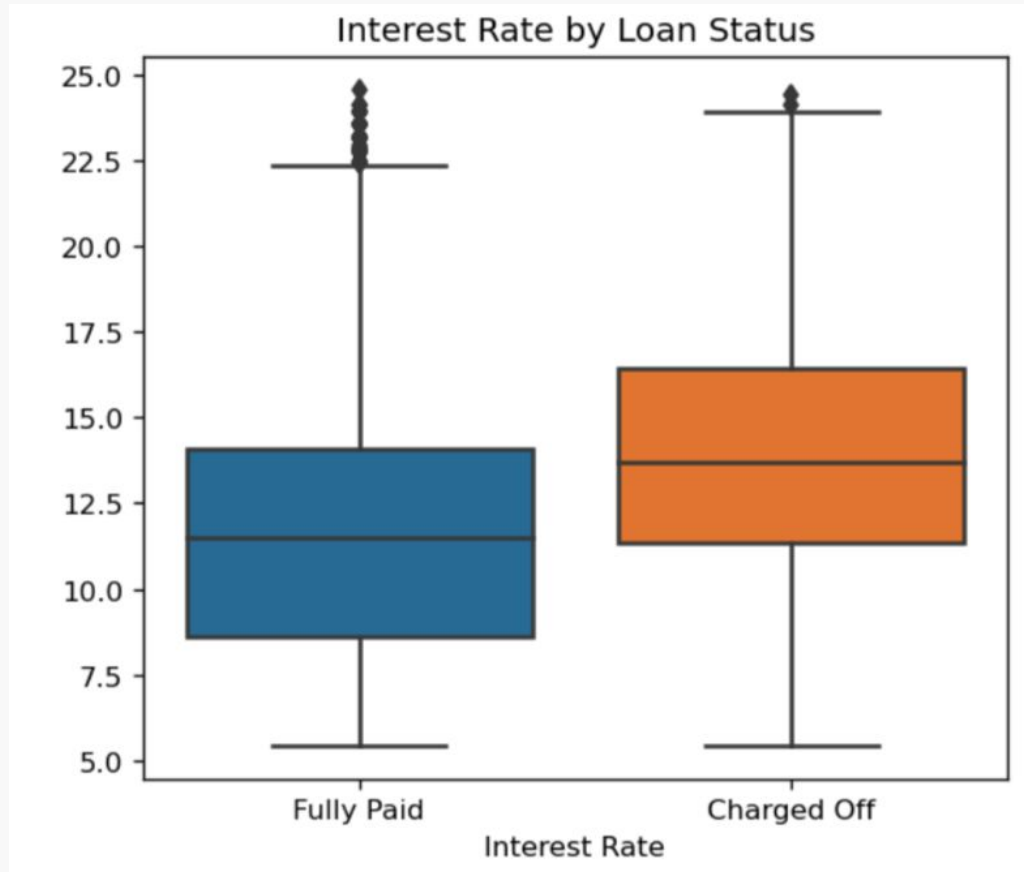
Verification Status



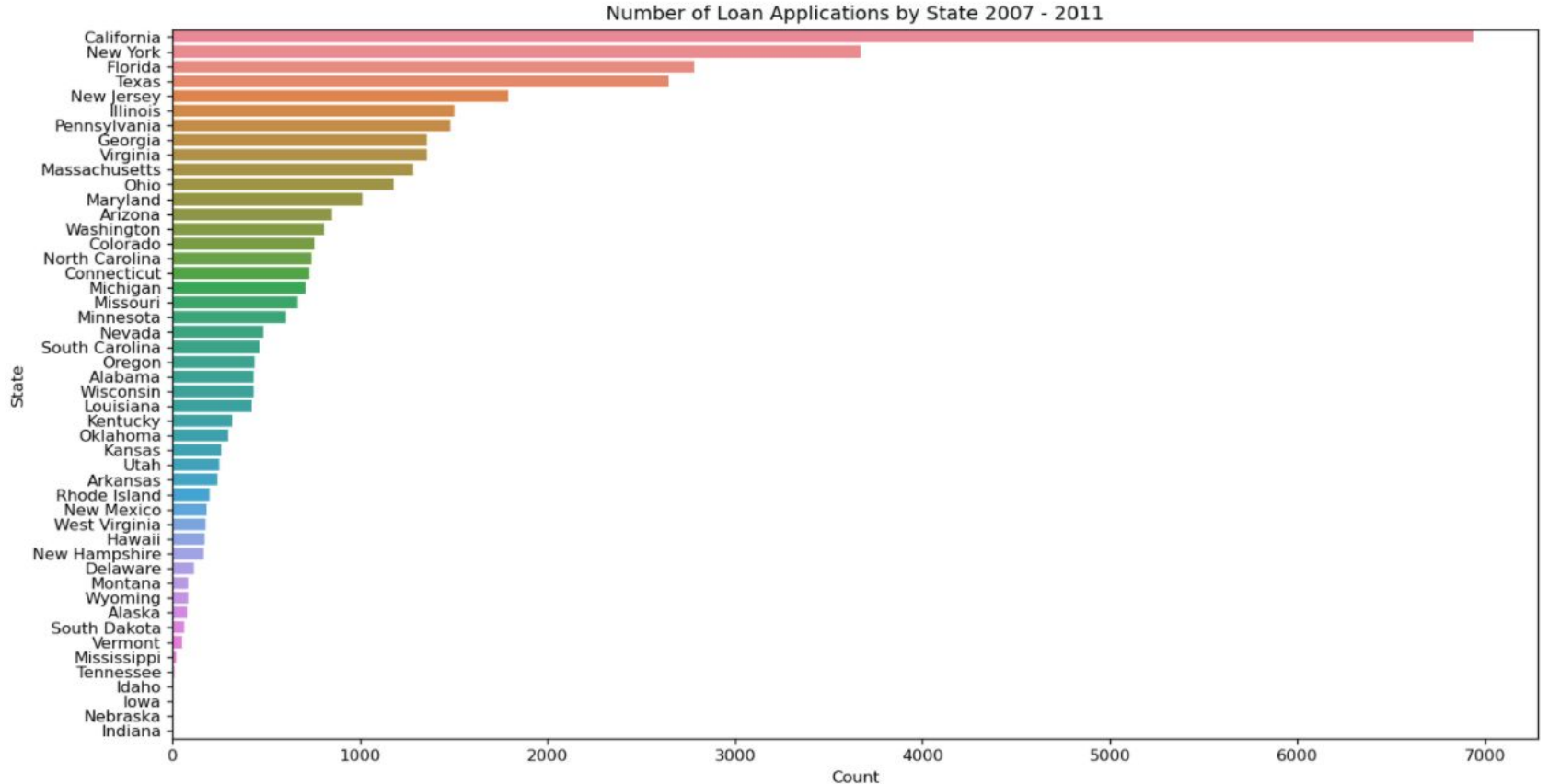
Applications over Time



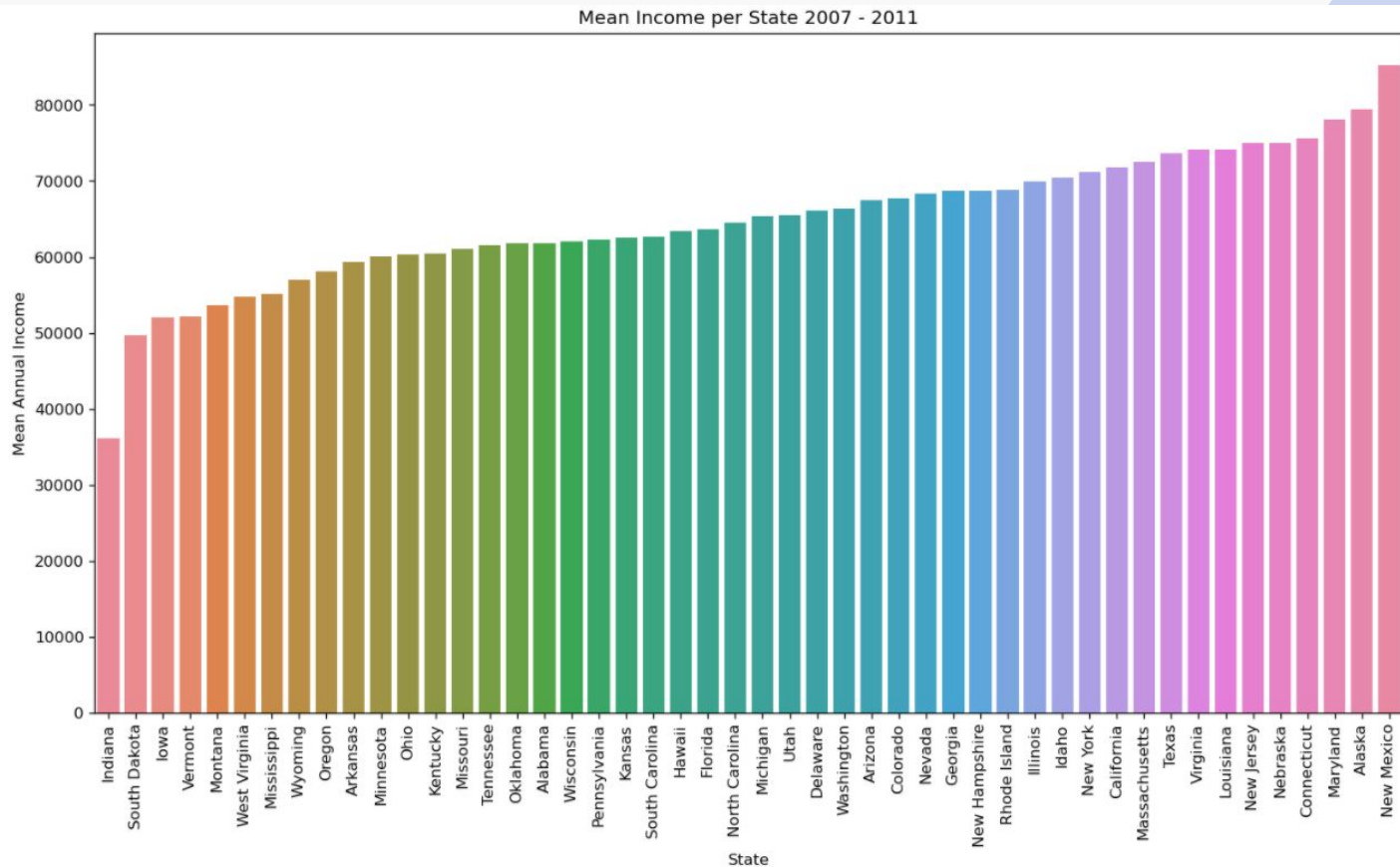
Interest Rate



Location



Mean Income by Location





PREDICTIVE MODELS **With Machine Learning**

Target Variable

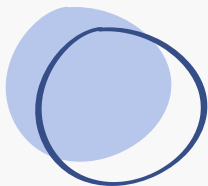
- Loan Status - Fully Paid or Charged Off where encoded as Fully Paid == 1 and Charged Off == 0 for the purpose of modelling.

Predictors

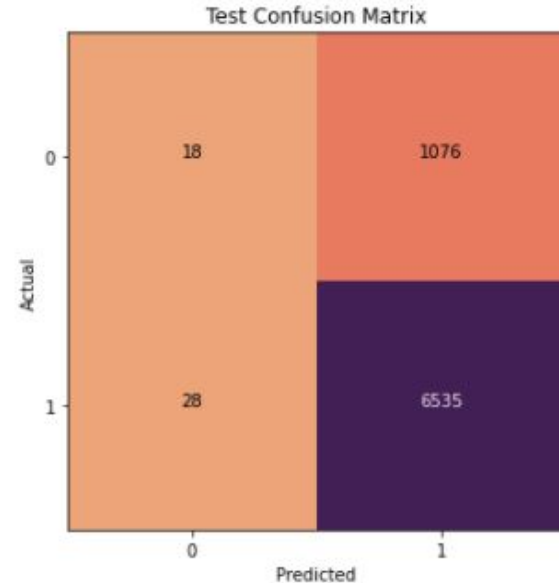
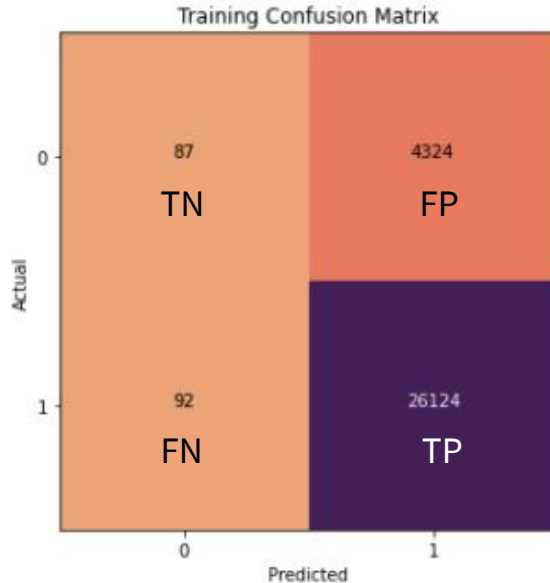
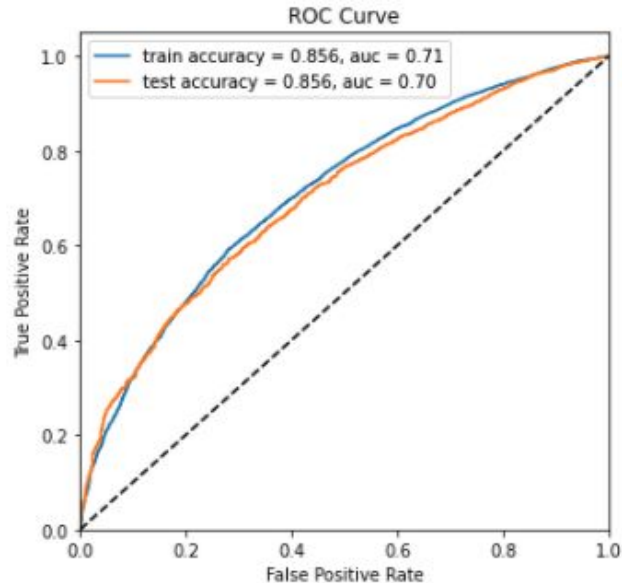
- Categorical variables were converted to dummy variables.
- Panda's Profiling showed the annual income variable is highly skewed so the log-transform of the values was taken.

Train/Test Split

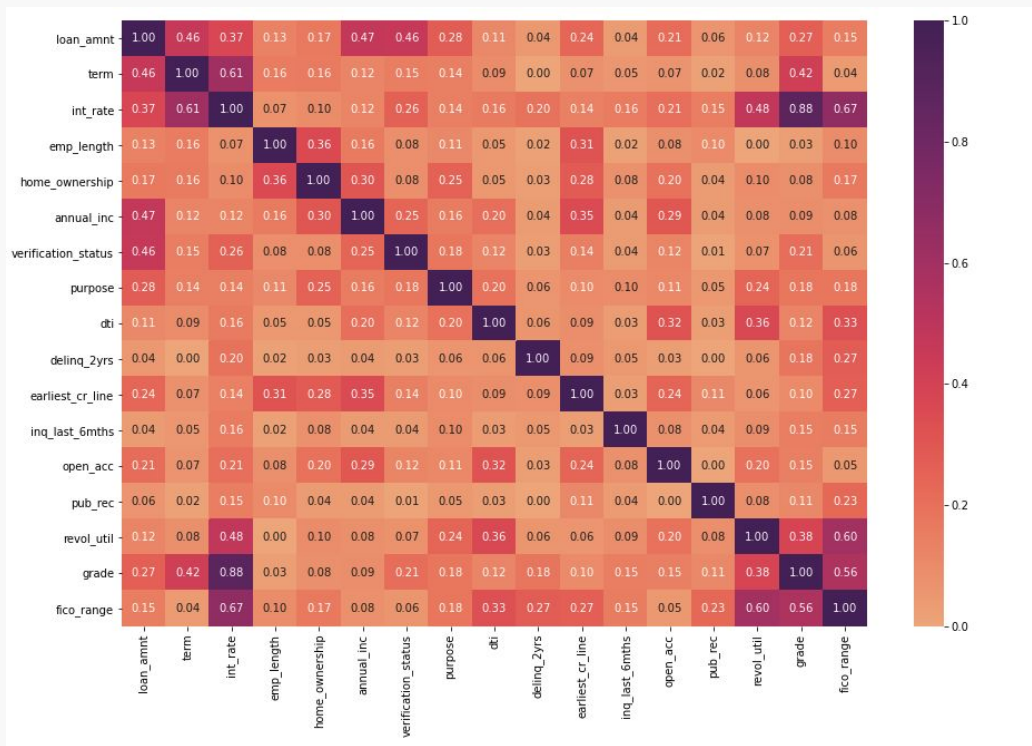
- 80/20 conventional split



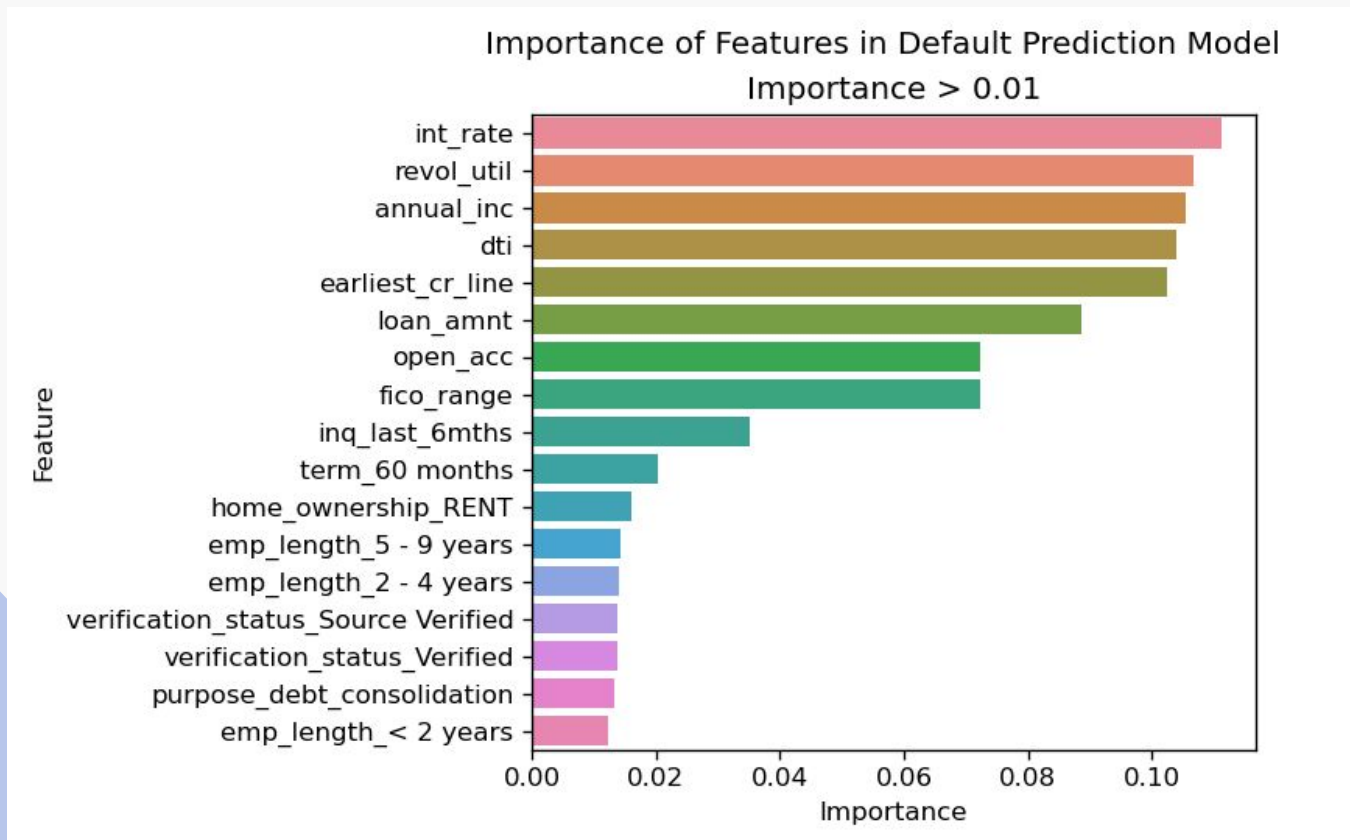
Logistic Regression Model Results



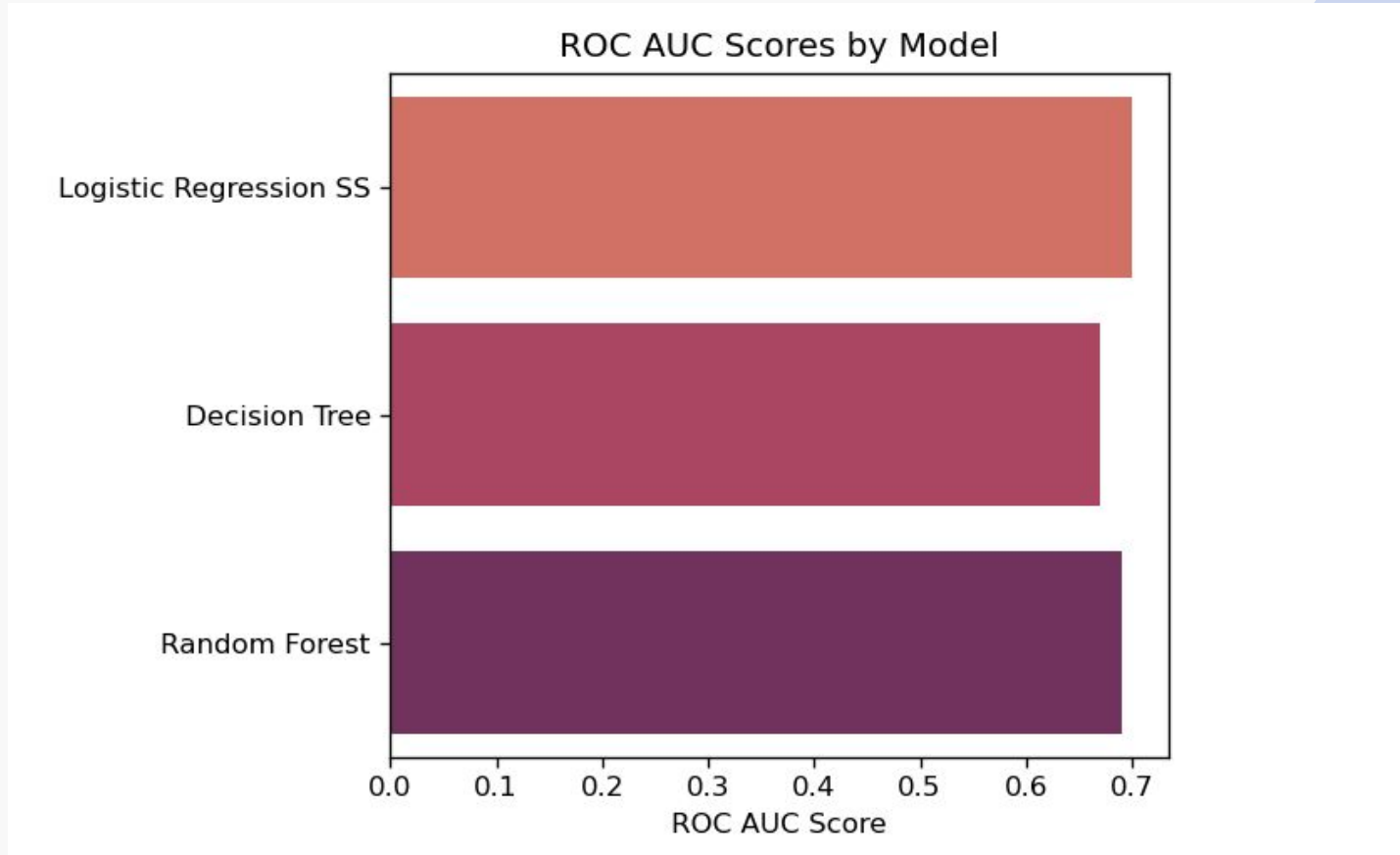
Correlations between Predictors



Feature Importance

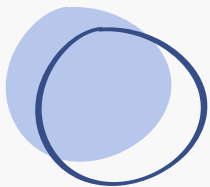


Model Comparison



Summary

- Interest Rate - loans with higher rates more likely to be charged off
- Other features of high importance in default prediction:
 - Revolution Credit Utilization Rate
 - Annual Income
 - Debt to Income Ratio





Thank you for Listening Question Time