

Lending Club Data Analytics

Yiwen Xue
Aug 2018

Outline



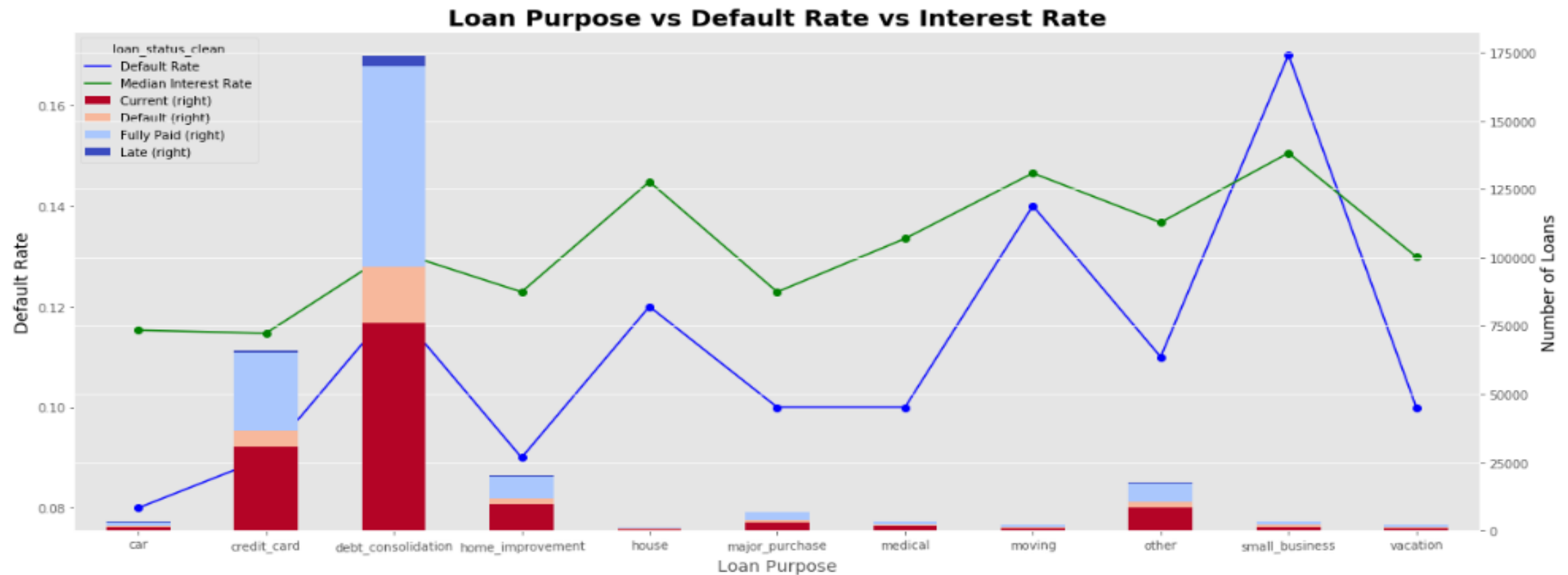
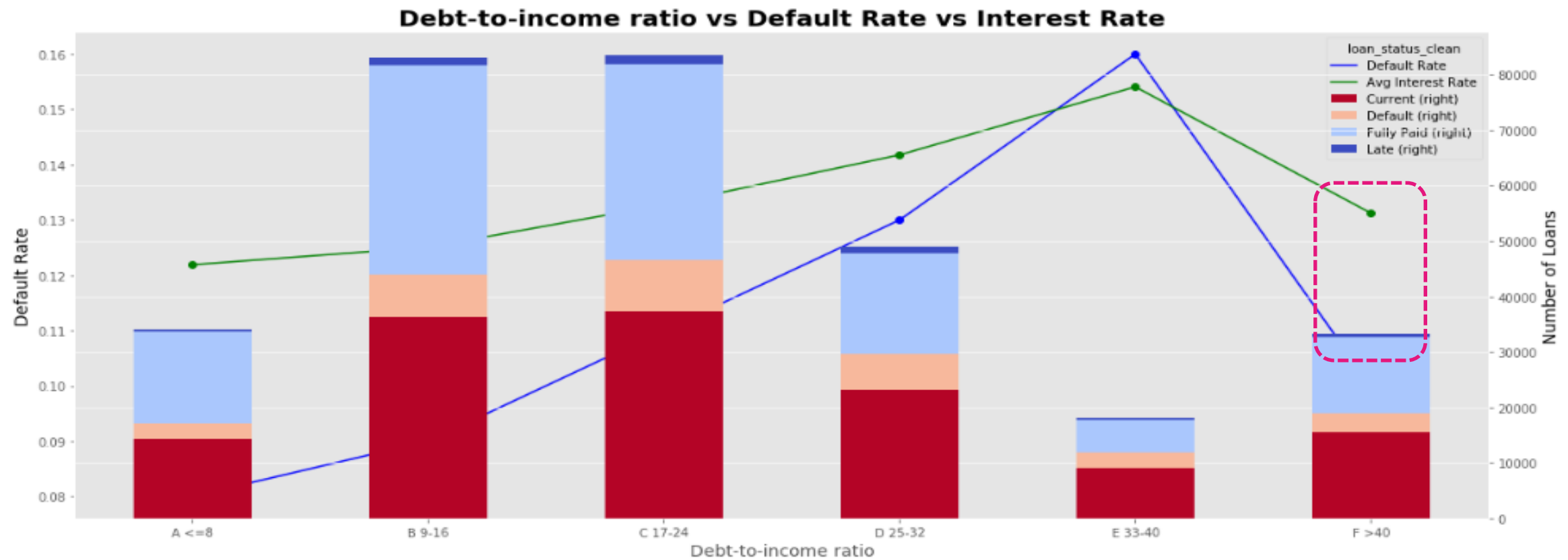
Exploratory Data Analysis & Findings

Predictive Model on Loan Default

Model Predicted Feature Importance

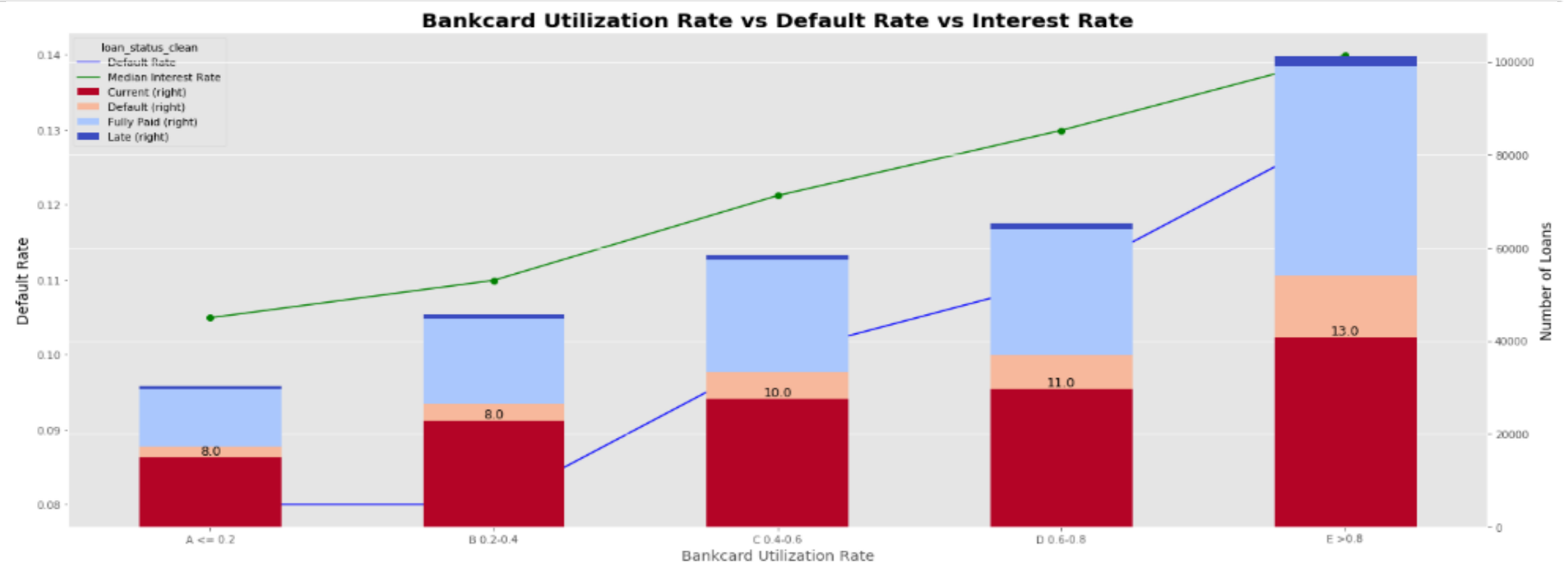
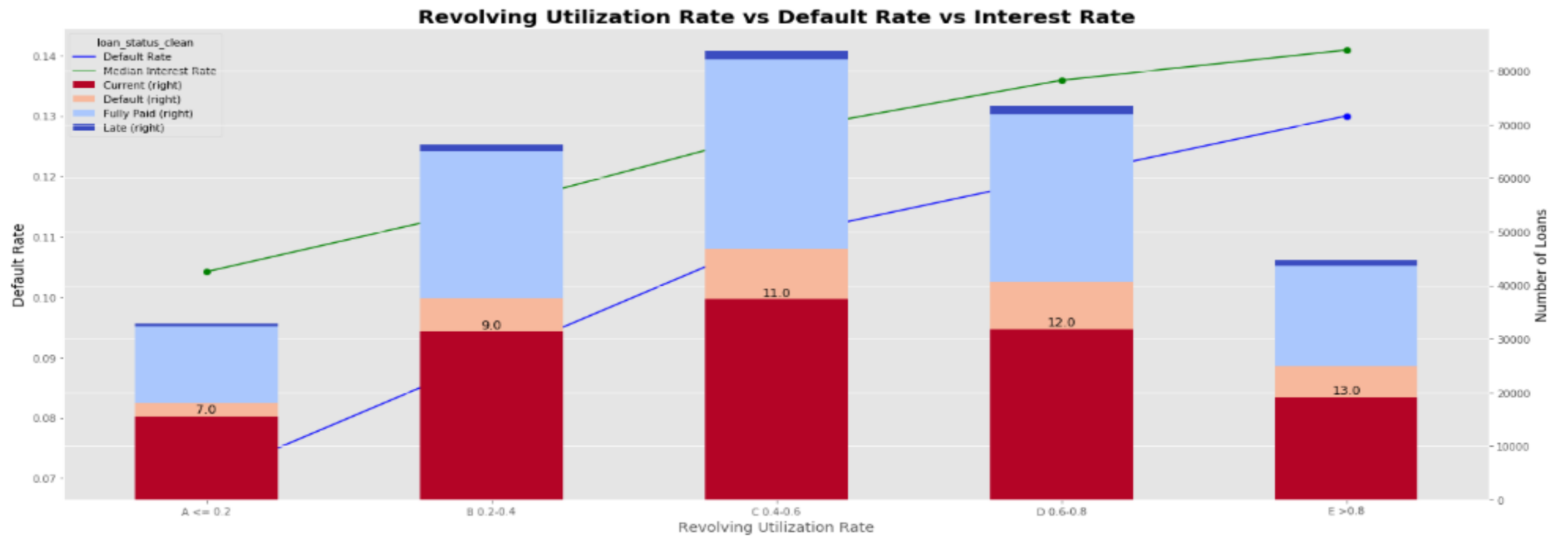
Exploratory Data Analysis & Findings

Debt-to-income ratio ≥ 40 has a lower default rate and interest rate than 33-40 bucket



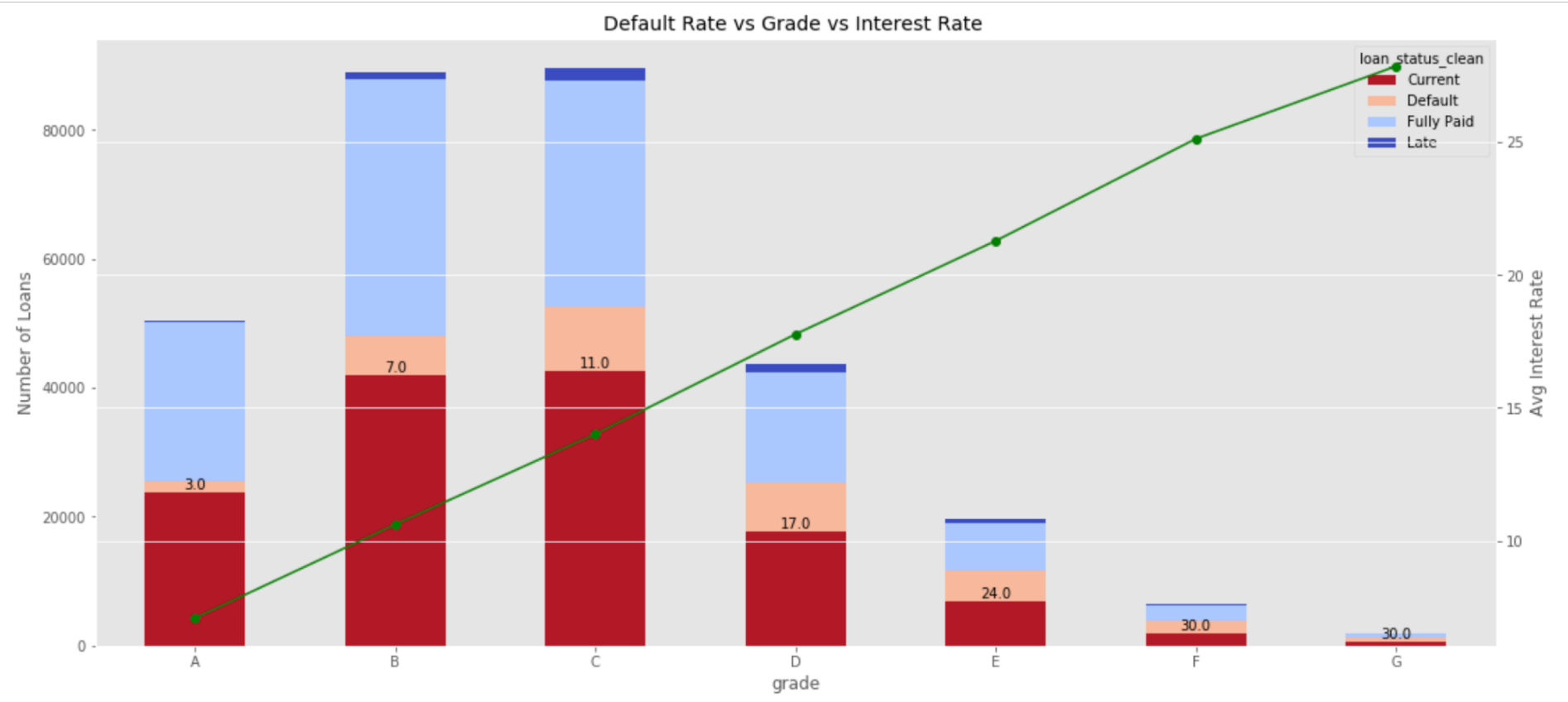
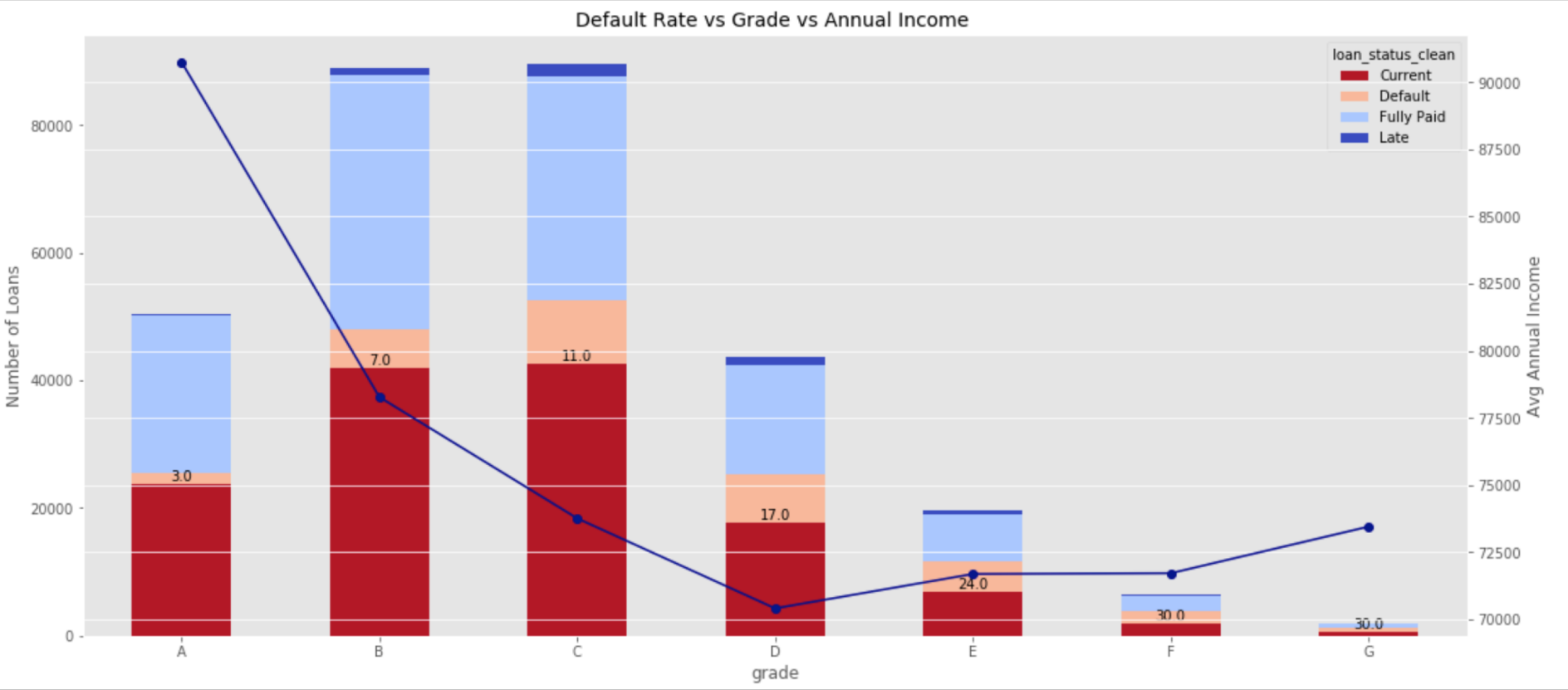
Exploratory Data Analysis & Findings

The higher the bankcard and revolving line utilization, the higher the default risk and interest rate



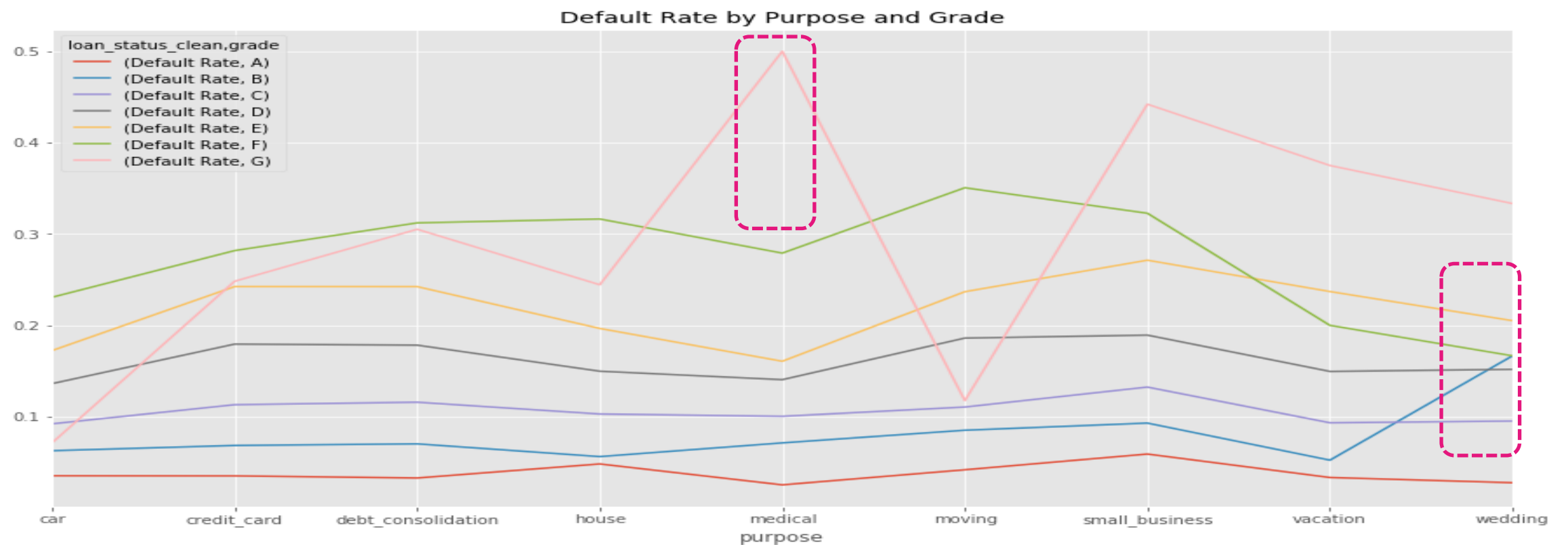
Exploratory Data Analysis & Findings

The lower the loan grade, the higher the default and interest rate. Average annual income is higher in grade G is higher than D-F



Exploratory Data Analysis & Findings

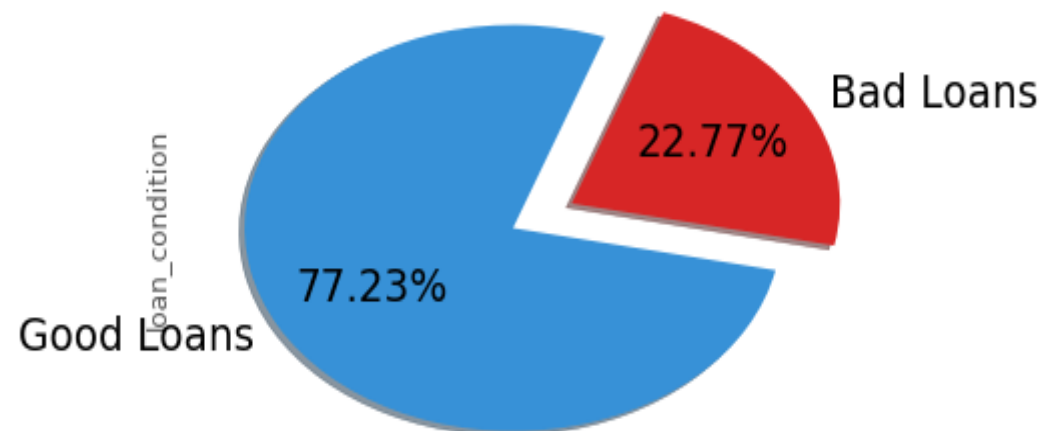
Grade G shows a very interesting default rate trend among different loan purposes – extremely high for medical purpose, lowest for moving. Grade B borrowers have a surprisingly high default rate in wedding



Predictive Model on Loan Default

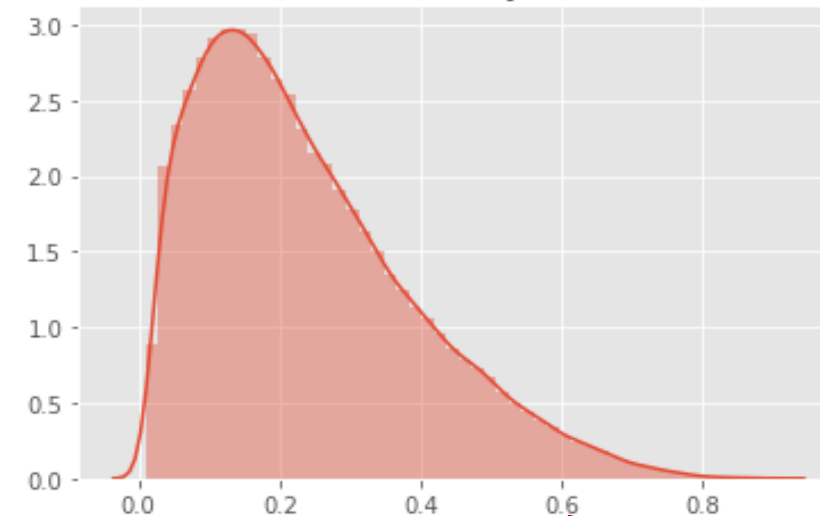
Random Forest and Gradient Boosting Tree models show similar model performance – 0.72 AUC (73% prediction accuracy)

Distribution of Loan Conditions



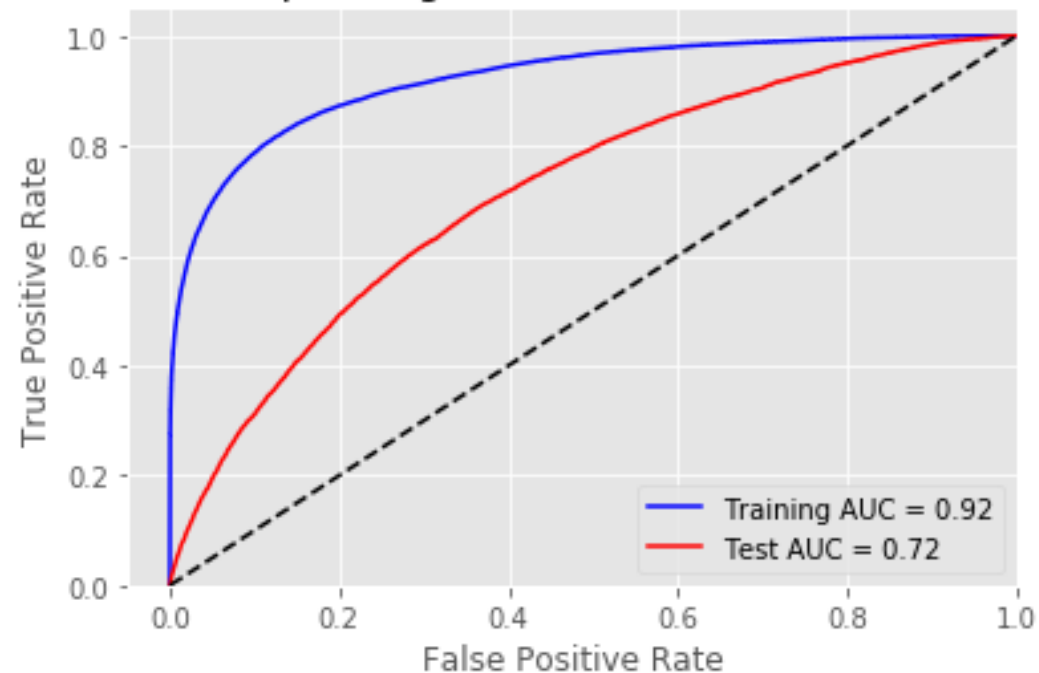
Distribution of loan conditions: good loans include only fully paid loans, bad loans include default, charged-off, late and loans in grade period. Current ongoing loans are not part of the pie chart.

Distribution of Predicted Probability of Default on current loans

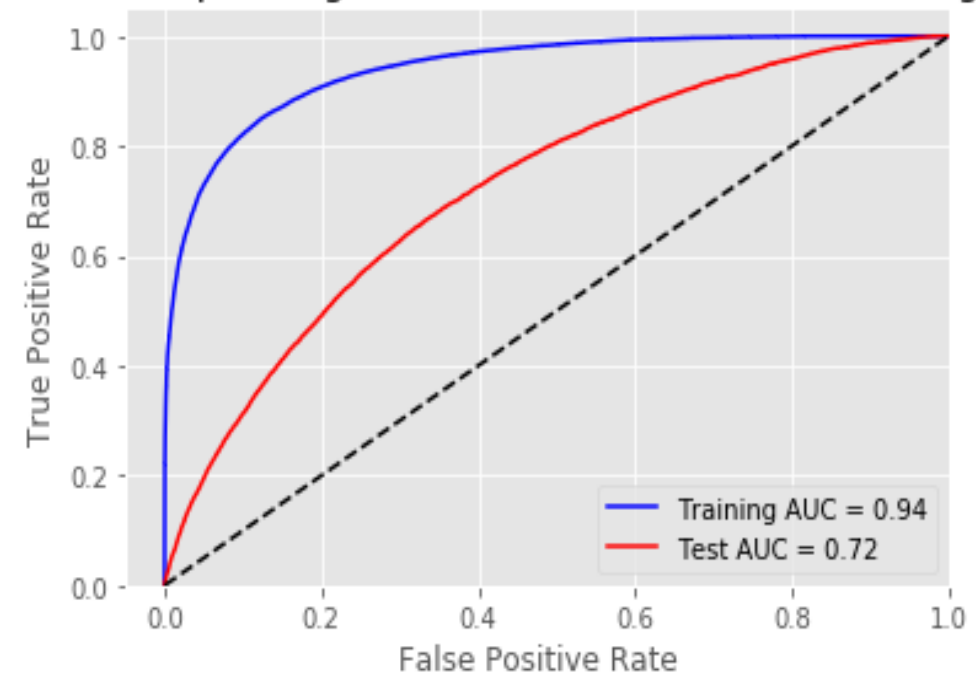


* Based on the distribution of predicted probability of default, it's suggested to choose cut-off probability to be in the range 0.3-0.35, which approx. cut the borrower population into 77/23. loans with default probability above the threshold classified as good loans, otherwise bad loans.

Receiver Operating Characteristic of Random Forest



Receiver Operating Characteristic of Gradient Boosting Tree



Model Predicted Feature Importance

Interest rate, dti, annual income, fico score and borrowers' historical credit record such as revolving balance and bankcard utilization tell a lot about their credibility to pay back the loan

