

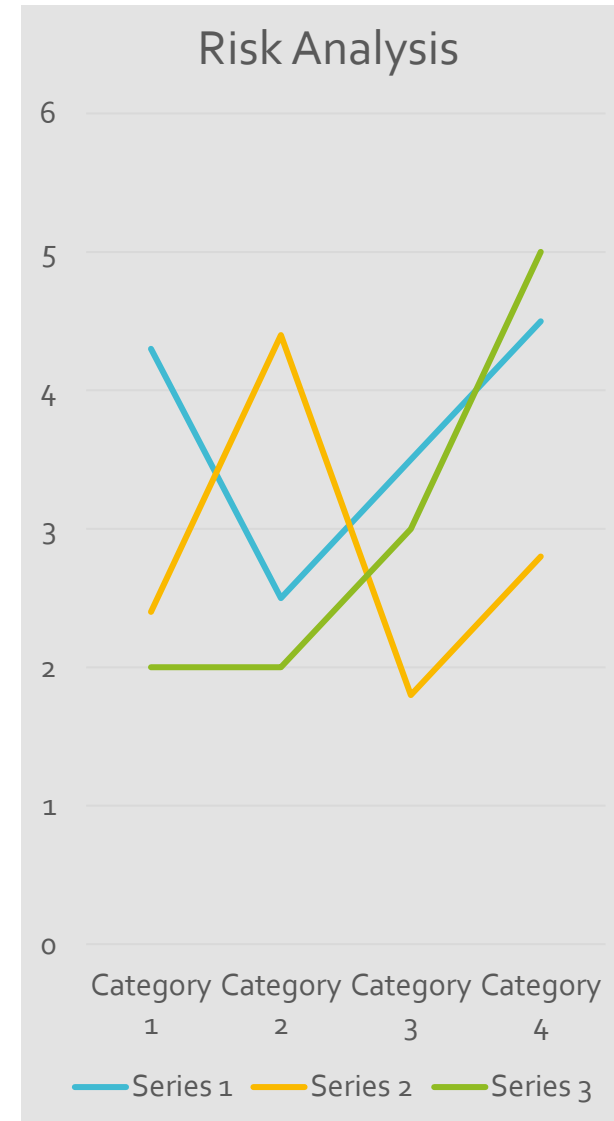
Lending Club Case Study

(Risk Factors Analysis for lending Loans)

Group Members :-

Aniket Singla

Smita Behera



Abstract

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures .
- Borrowers can easily access lower interest rate loans through a fast online interface
- The objective of analysis is to use the information about past loan applicants and find whether they 'defaulted' or not.

Problem Solving Methodology

➤ Data Comprehension

Working with the Data Dictionary and getting an in-depth understanding of all the columns and their domain specific uses. The data dictionary along with a glimpse into data, indicates that some of the numerical variables like interest rate, loan amount, loan status, verification status, employment length are useful.

➤ Data Cleaning

Removing the null valued columns, columns with 85% missing data, data which has a unique value and unnecessary variables which do not aid risk analysis.

➤ Univariate Analysis

Analyzing each column, plotting the distributions of each columns, checking **skewness** of data. It was observed that the default history compromises of only 14% of data. To counter the imbalance, percentile-based metrics were derived to better compare the variables. Another method possible was to **resample**, but percentage method worked well for EDA.

Problem Solving Methodology

➤ Segmented Univariate Analysis

Analyzing the continuous data columns with respect to the categorical column. Numerical continuous variables were divided into categories and provided great in-sights into the data in hand.

➤ Bivariate Analysis

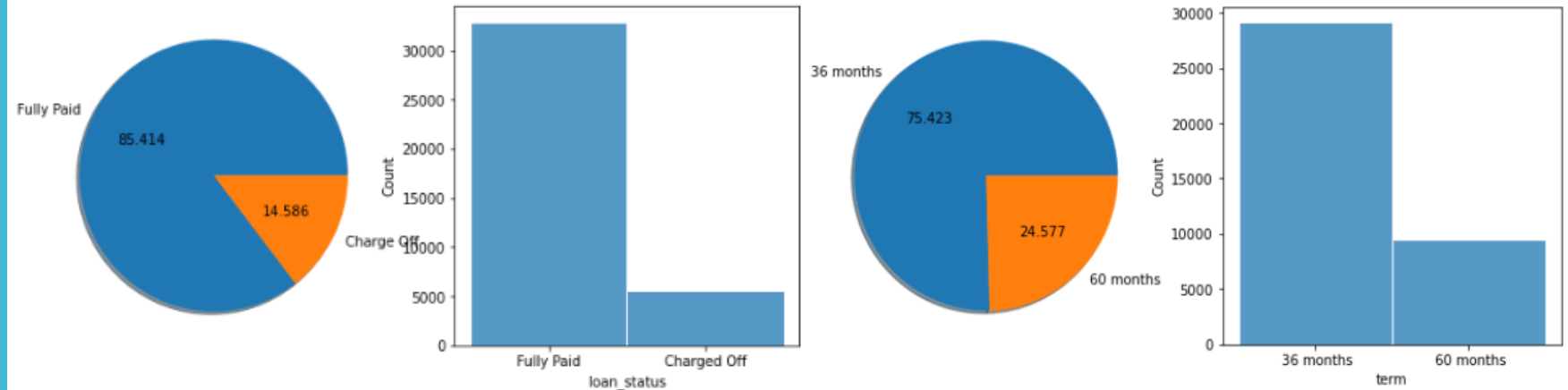
Based on insights from the above univariate and segmented analysis, the major drivers were apparent, and these could further be analyzed by grouping them together to understand the combined impact on data.

➤ Data Visualization

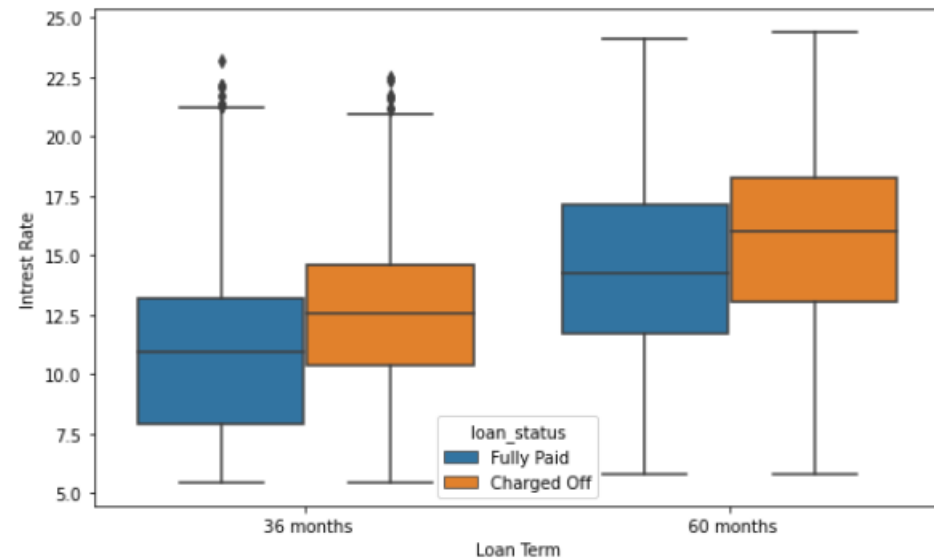
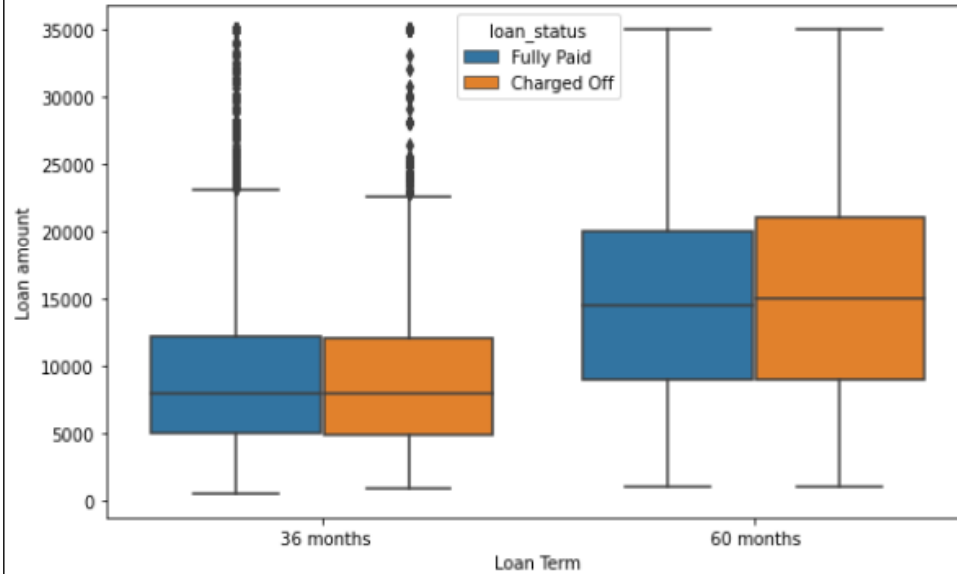
Various kinds of plot give visibility into data and heatmaps were used extensively to understand the impact of various variables.

➤ Recommendations : Analyzing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.

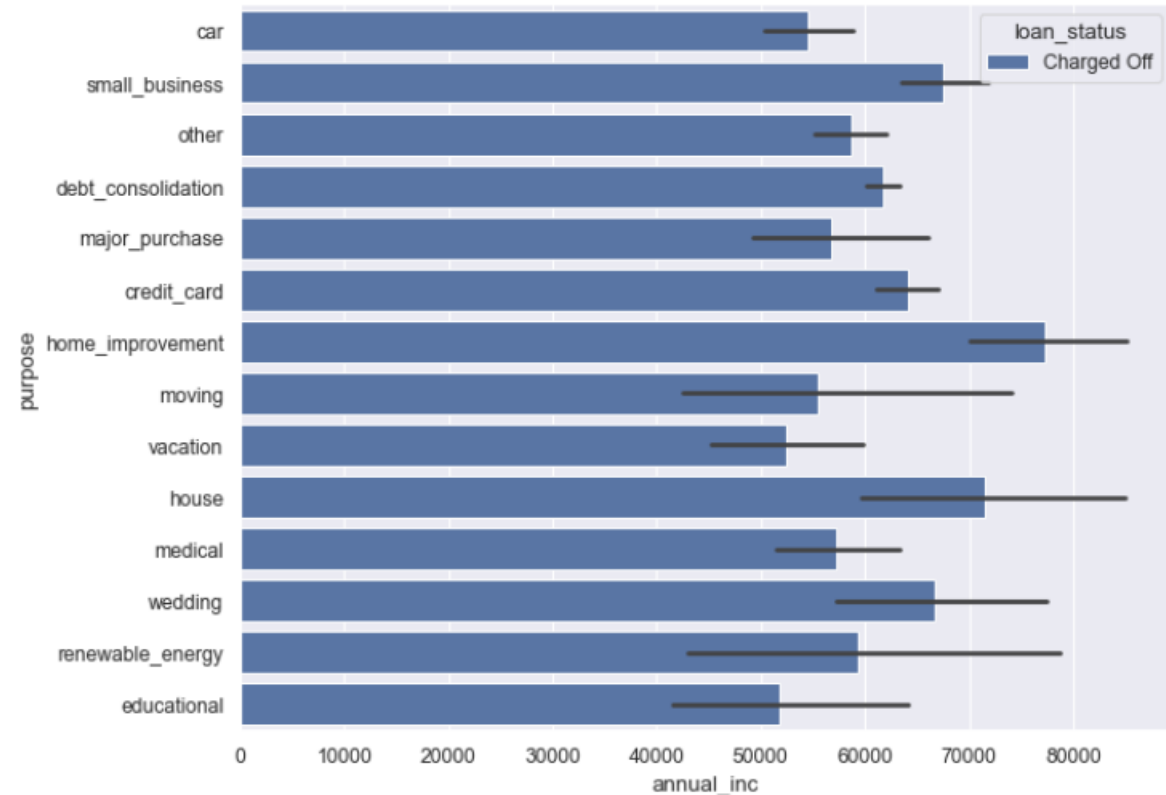
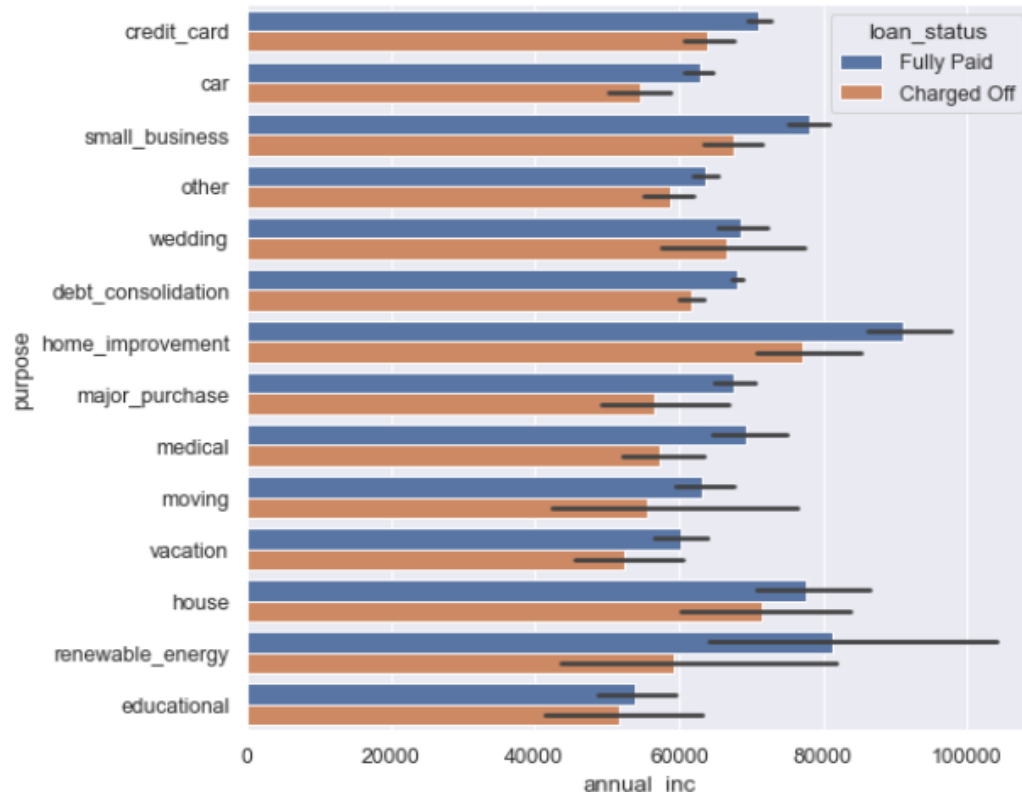
Exploratory Data Analysis



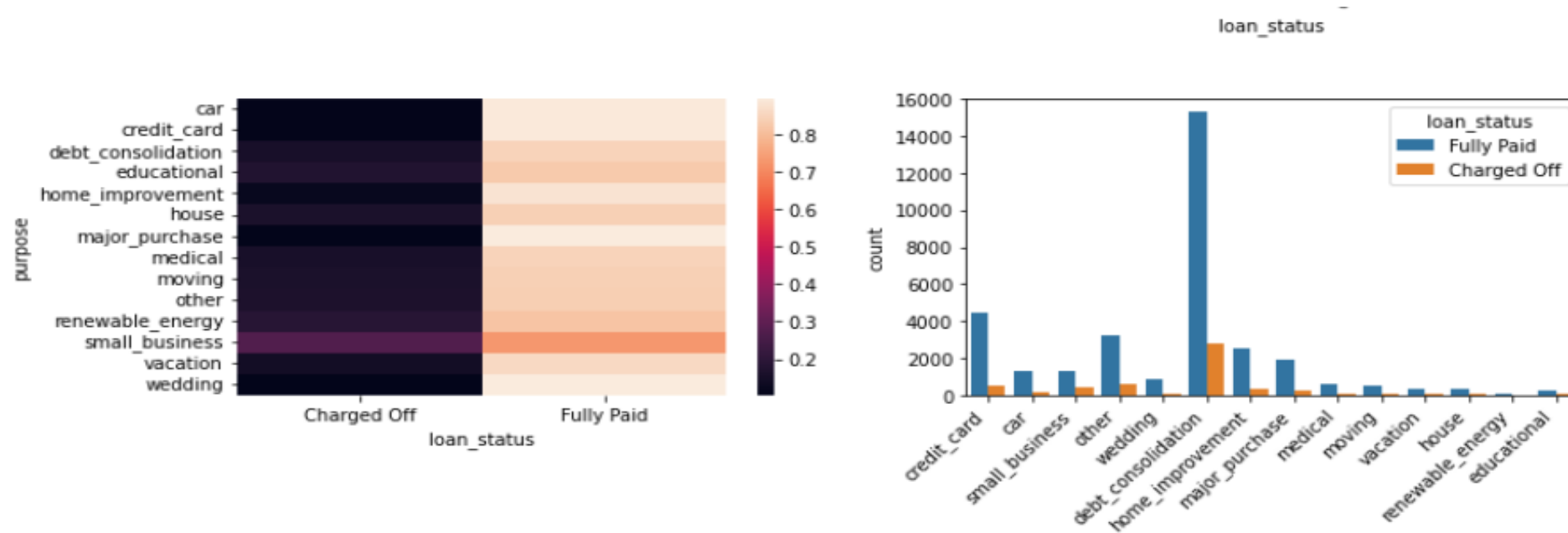
- There are two loan terms 36 and 60 months and approx.75% borrowers took loan for 36 months term .
- 85% of the loan had been paid off fully while 15% loans were charged off i.e defaulters



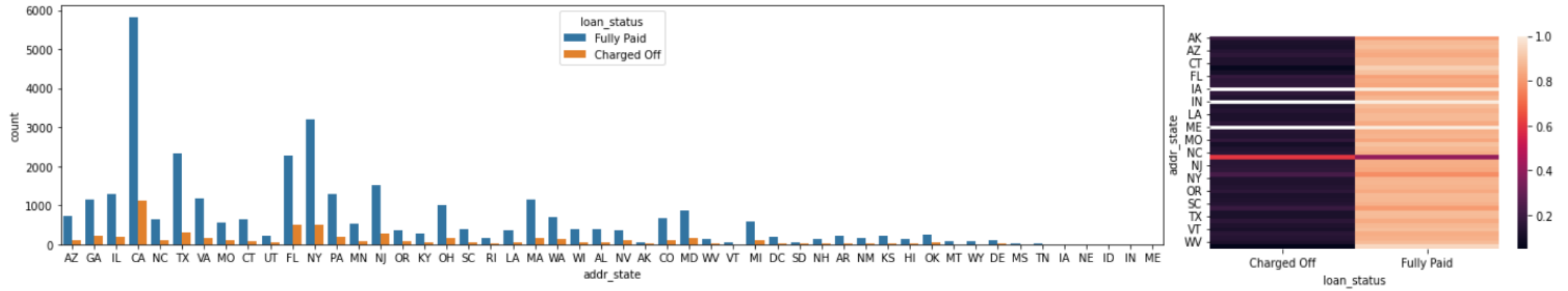
- Most of high amount loans for 60 months term leads to default .
- The default rate is high in 60 months tenure because most people took high loan amount with high interest rate and they faced difficulties , turning to defaulters
- When Loan Interest rate is high, borrowers are more likely to default



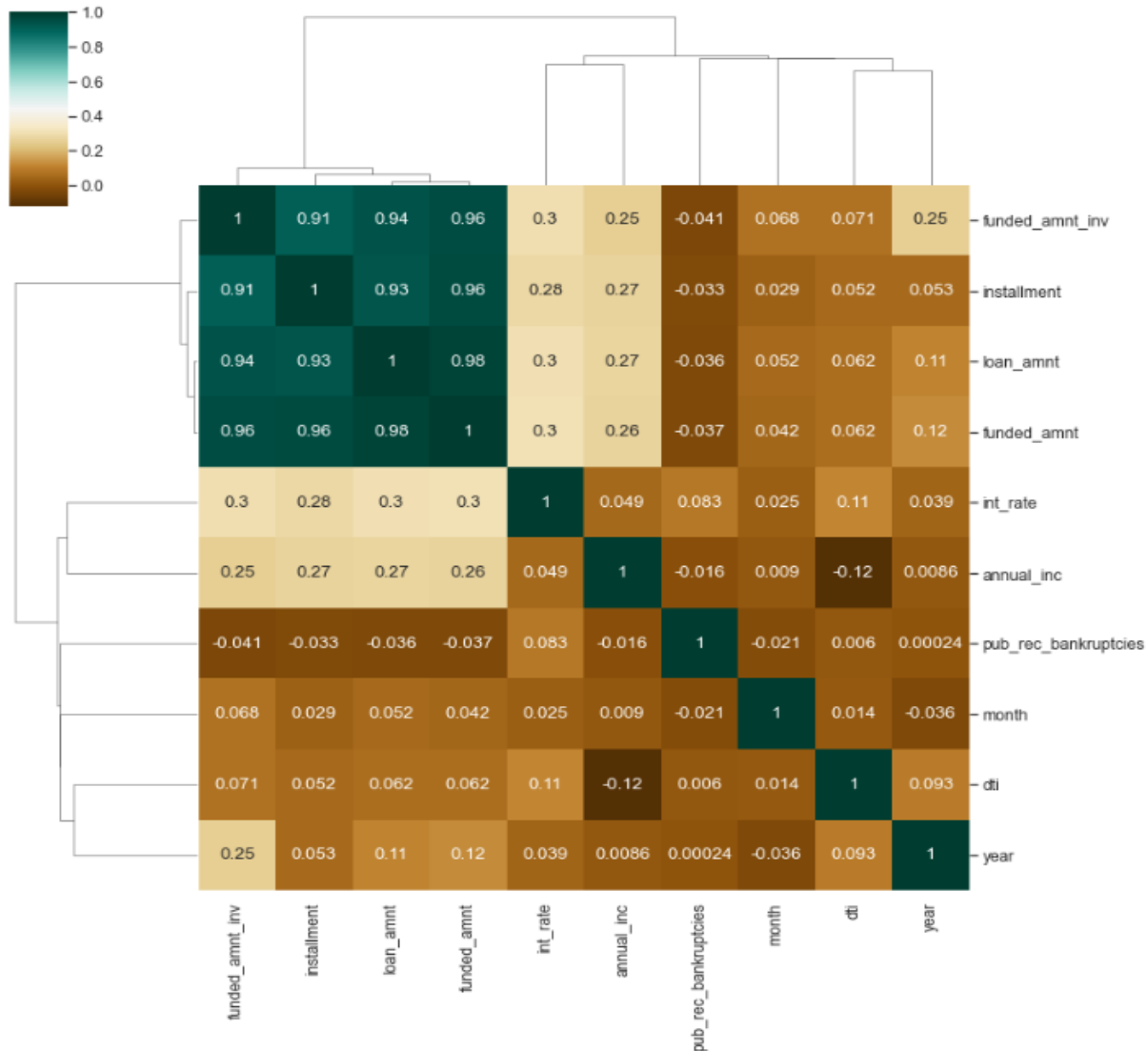
- Borrowers who took loan for the purpose of home improvements have most number of defaulters , followed by buying house and small business purposes ,
- Small business owners has the highest percentage of defaulters



- When purpose of loan is "small_business", the percentage of defaulters to fully paid loans is highest, followed by renewable_energy , educational and others.

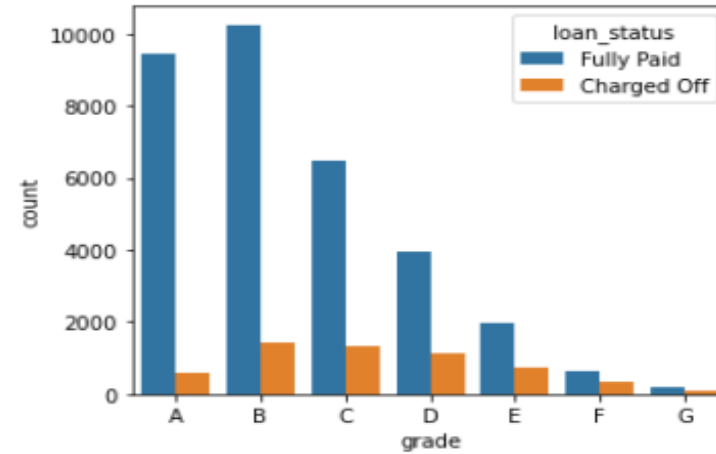
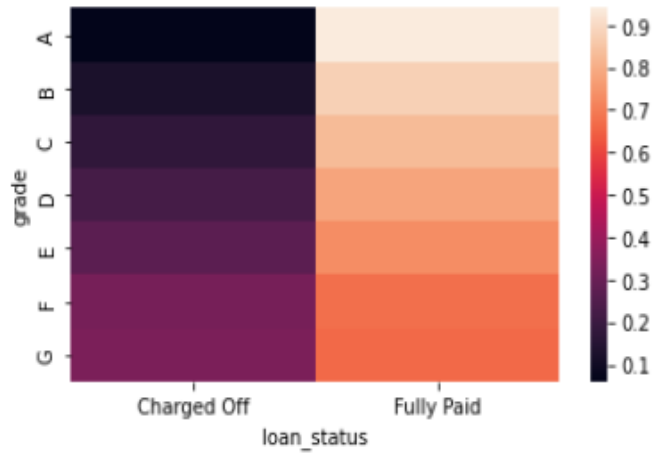


- The highest number of loans were given in CA, FL and NY states with highest number of defaulters as well .
- Highest percentage of charge offs are for WV state

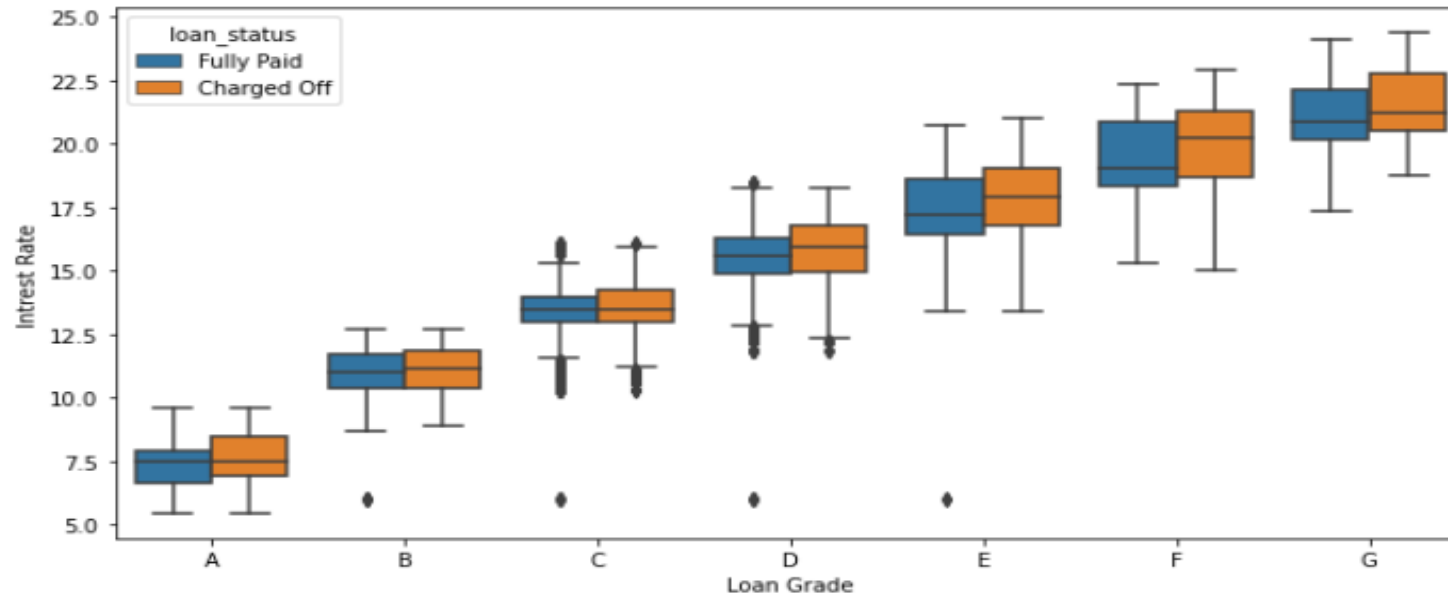


Correlation Matrix

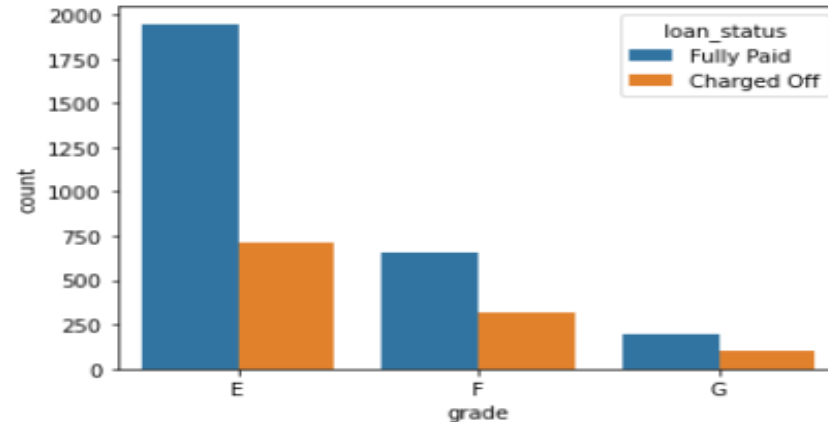
- Loan Installments , Amount, Investor amount, Funding amount are strongly correlated
- DTI (Debt-to-income ratio) and Annual income are negatively correlated, with low annual income having high DTI
- Annual income and employment years show positive correlation
- Loan Interest rate and Annual income show positive correlation



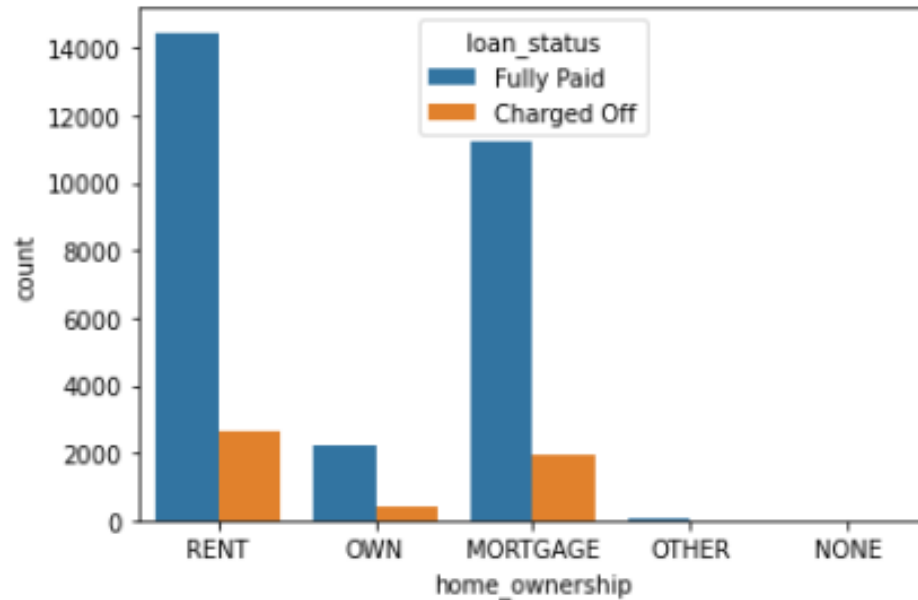
Heatmap for % of "Charge off" and "Fully Paid"



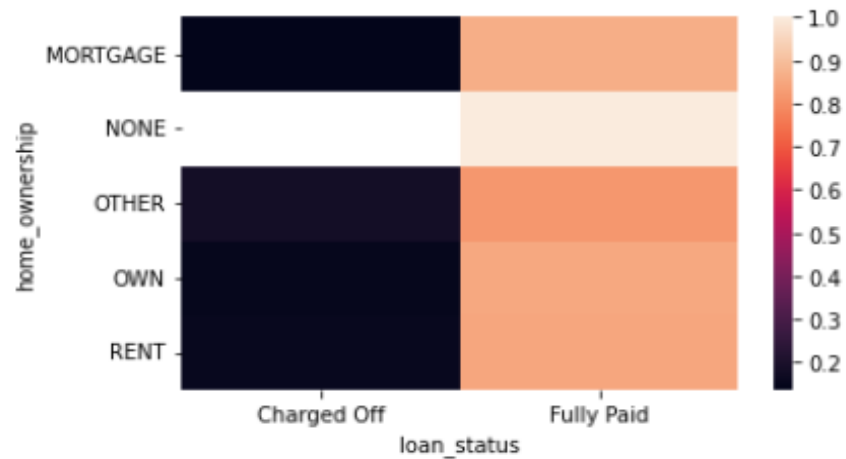
Boxplot of "Charge off" and "Fully Paid" with Interest rate and Loan Grade

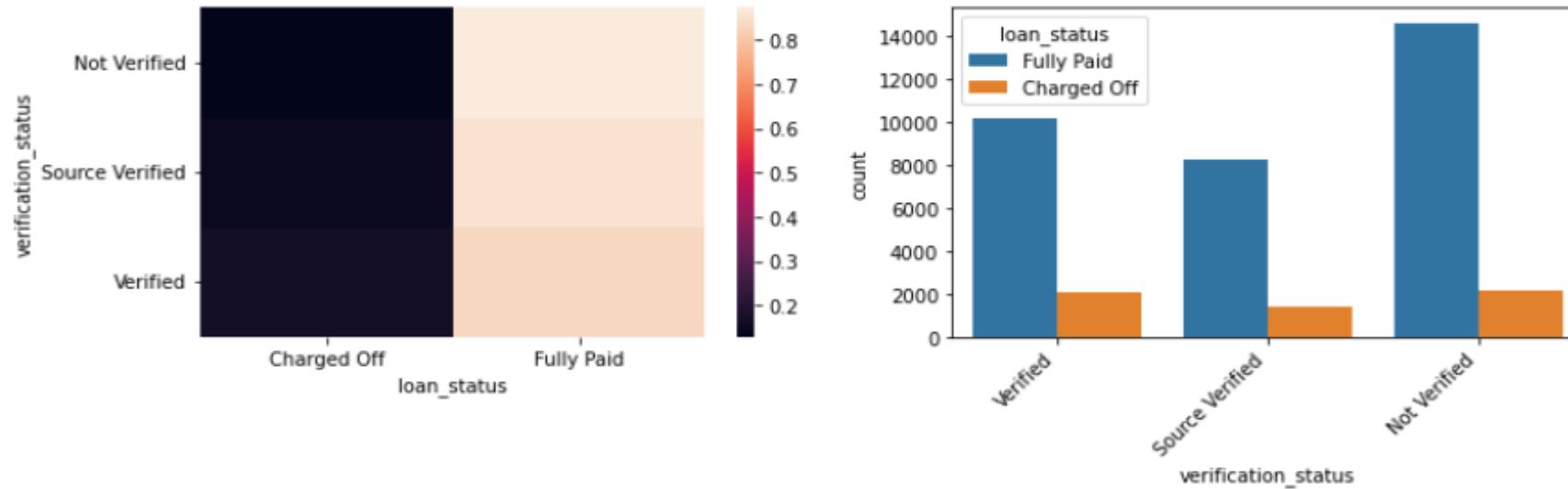


- Grades are very good category to tell the borrower probability of defaulting the loan.
- The Lower grades(E,F,G) have higher chances of defaulting the loan than Higher ones(A,B)
- Also the Lower grades are getting loans for higher interest rates which might be the cause for loan default.
- For different grades, we can draw a positive conclusion that defaulters are lowest in grade A inspite of a good number of loans. Charge offs are highest in B, C and D, but looking at percentage, we can conclude that likelihood of Charge offs is highest in grade G and F though the numbers of loans in this category are less.

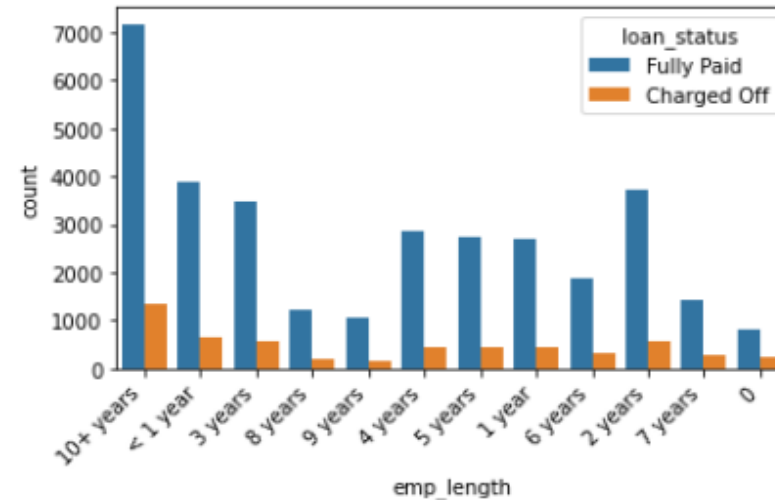
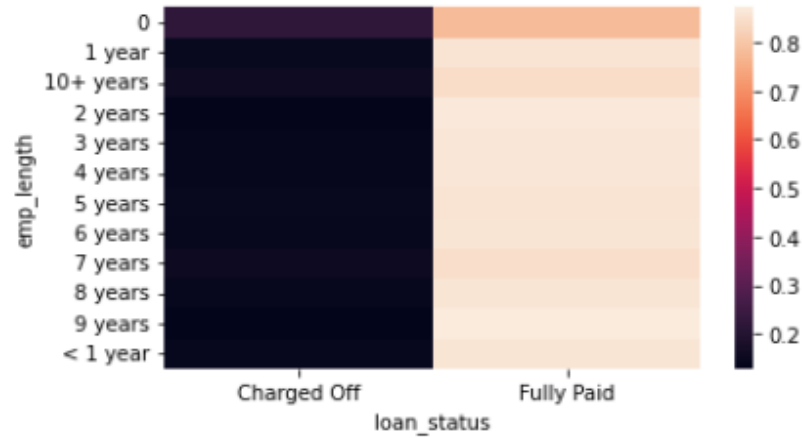


- Most of the borrowers live on rent or had a mortgage
- Home ownership shows highest charge offs for Mortgage and rent but percentage of chargeoff for each category is more or less same with category "OTHER" having a slightly more risky group.

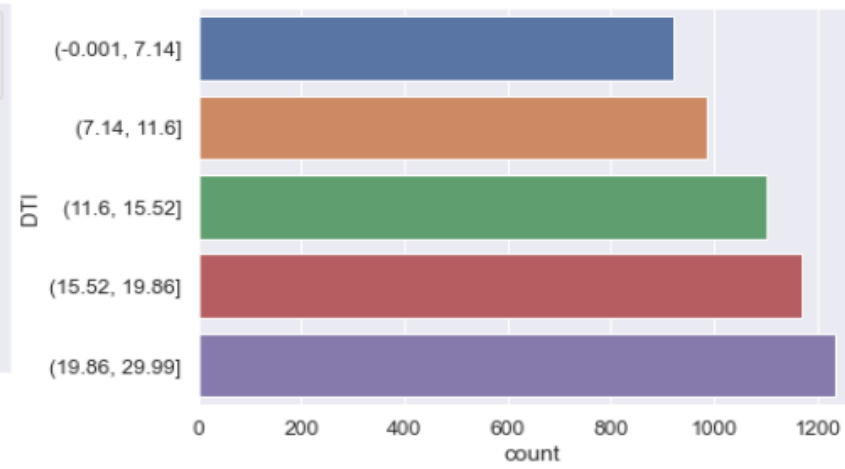
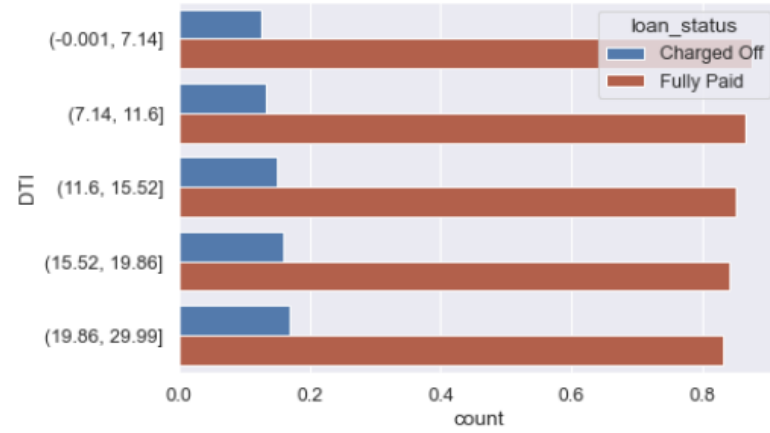




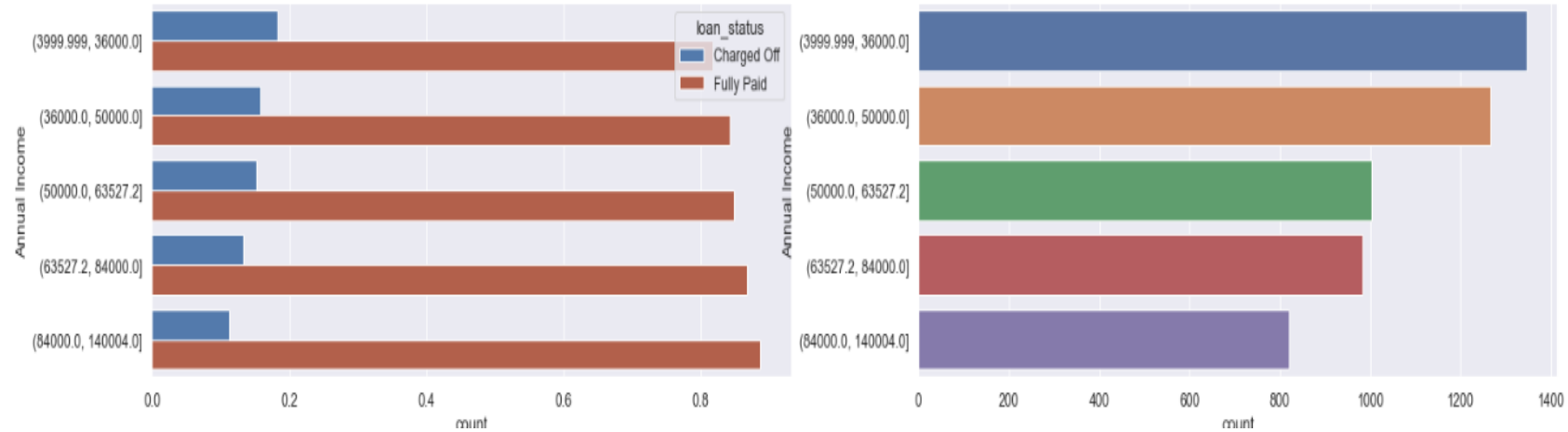
- Highest number of chargeoff are for "Not-verified" but the percentage is highest for "verified" indicating a major flaw in verification process.



- Though charge offs are highest for 10+ years but percentages indicate not very high impact of emp_length on charge offs though 10+ years are still on high number of defaulters.



- Charged off increase gradually with DTI (debt-to-income ratio)



- Charge offs is less for higher income groups and increase for lower income groups

Recommendations

➤ Key Recommendations

- The key factor to identifying defaulters is grade and sub-grade, with customer with grade A and corresponding sub-grade [A1 through A7] being the most reliable.
- Higher interest rate associated with high loan amount will lead to defaulting.
- Surprising number of charged offs belonged to category “Verified” for “verification_status” and the huge number of “Not_verified” status indicates a major need to revamp the verification process being followed.
- Public bankruptcy was a strong indicator of default
- High DTI should be a key deciding factor for lending.

➤ Optional Recommendations

- Purpose of loan like “education”, “small business” are likely to default
- Giving loans to lower grades at high interest rate can with highest percentage of defaulters are getting loans for higher interest rates which might be the cause for loan default
- Continue encouraging 36 months tenure as 60 months tenure had a higher default rate.