Lending Club Case Study

Presented by :

Ms. Nazneen Ansari

Mr. Chetan Desai

Batch : **ML C36**

Date : **9th February, 2022**

Problem Statement:

The lending company facilitates personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

However, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

Thus, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment before approving any loan application.

Methodology:

- In this case study, EDA will be used to understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- Identifying the two **types of risks** associated with the loan applications of the Bank
  - If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  - If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company
- Analysis has to be done on the given data which contains the information about past loan applicants and whether they 'defaulted' or not.

Types of Variables

- Quantitative variables :

  Rate of Interest, Annual Income, DTI, Funded loan amount, Open to Total account ratio

- Categorical variables :
  - Unordered :

    Loan Status, Purpose of Loan, Home Ownership, Verification status
  - Ordered :

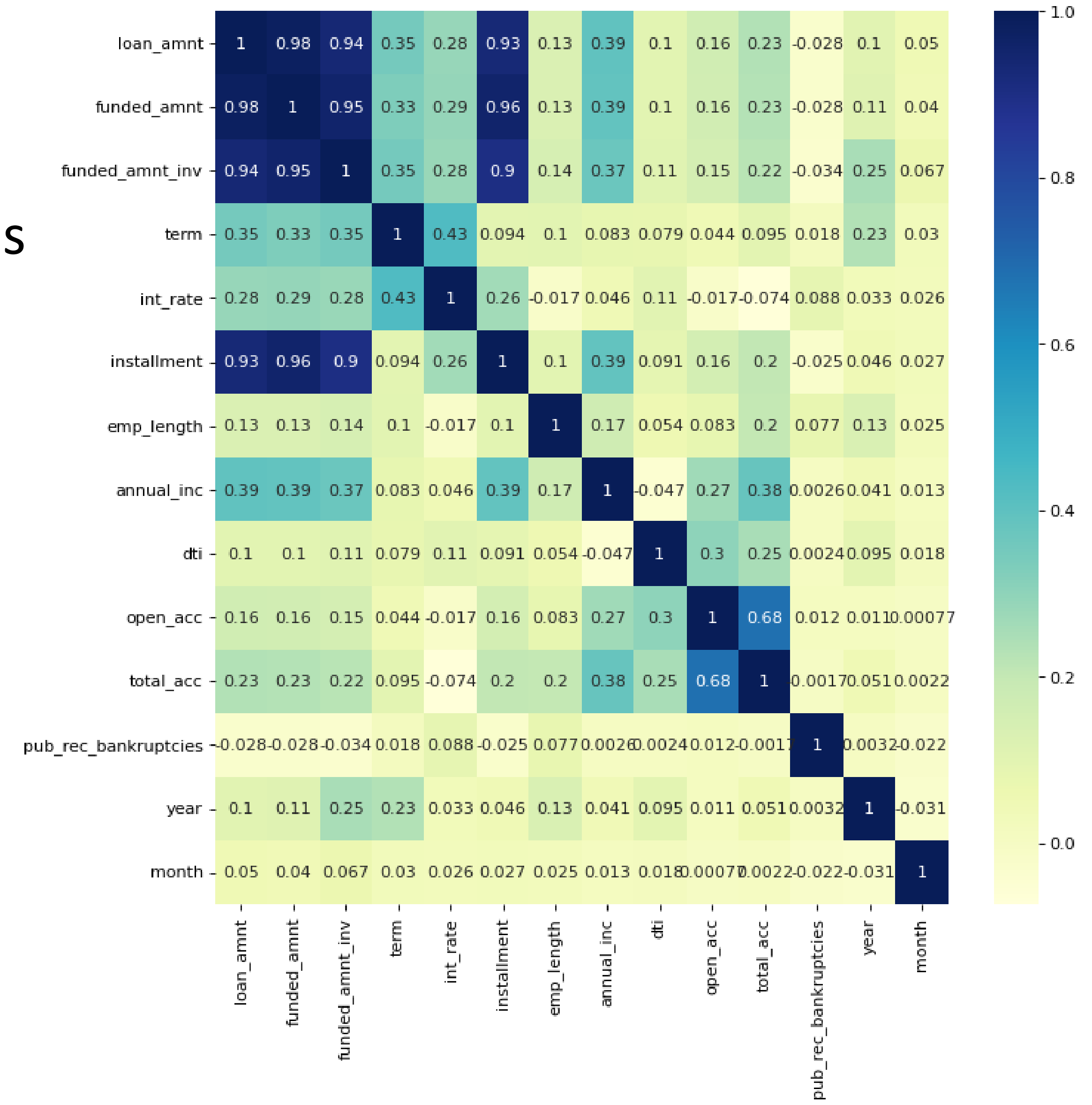
    Year/Month, Term

Data Cleaning & Manipulation

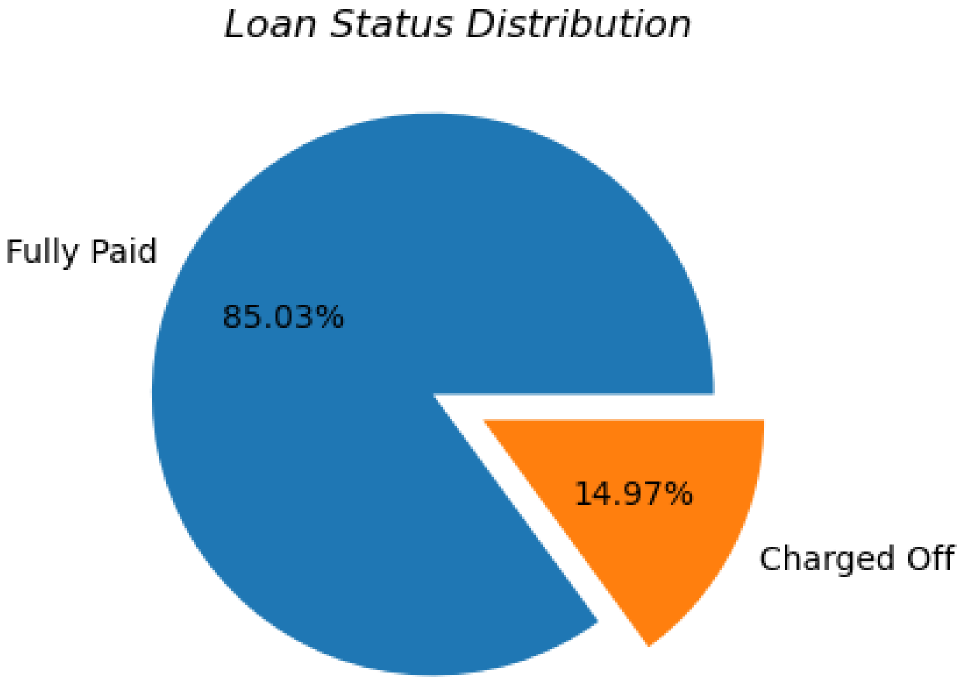
- Columns eliminated :
  - All the columns with values NA/null
  - All the customer behavioural variables
  - All the columns with least distinctive values
- Number of columns before data cleaning : 111
- Number of columns after data cleaning : 20
- Outliers have been removed for Annual income
- In Employment length column, the values which are less than 1 year are taken as 1 year and 10+ are taken as 10 years

Correlation

Correlation between attributes

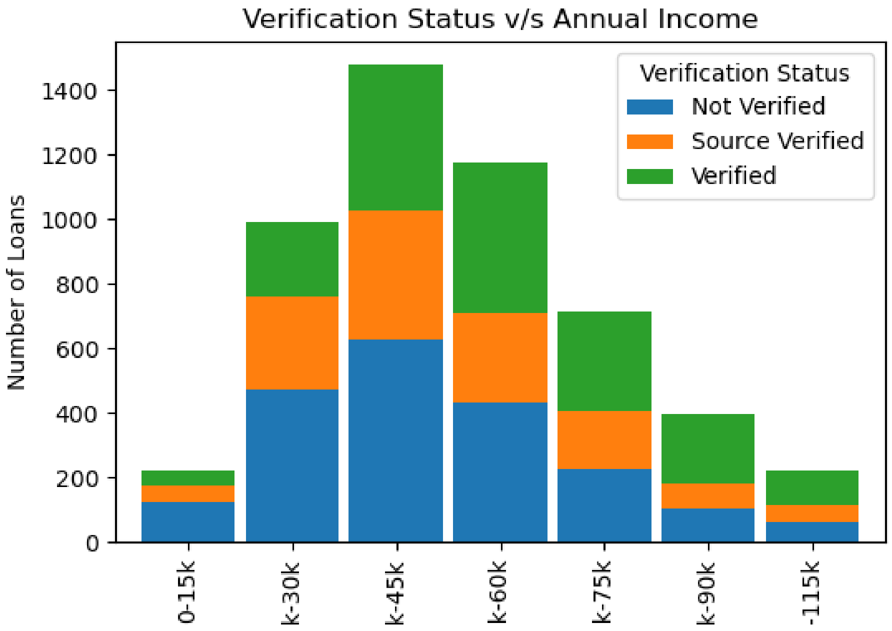


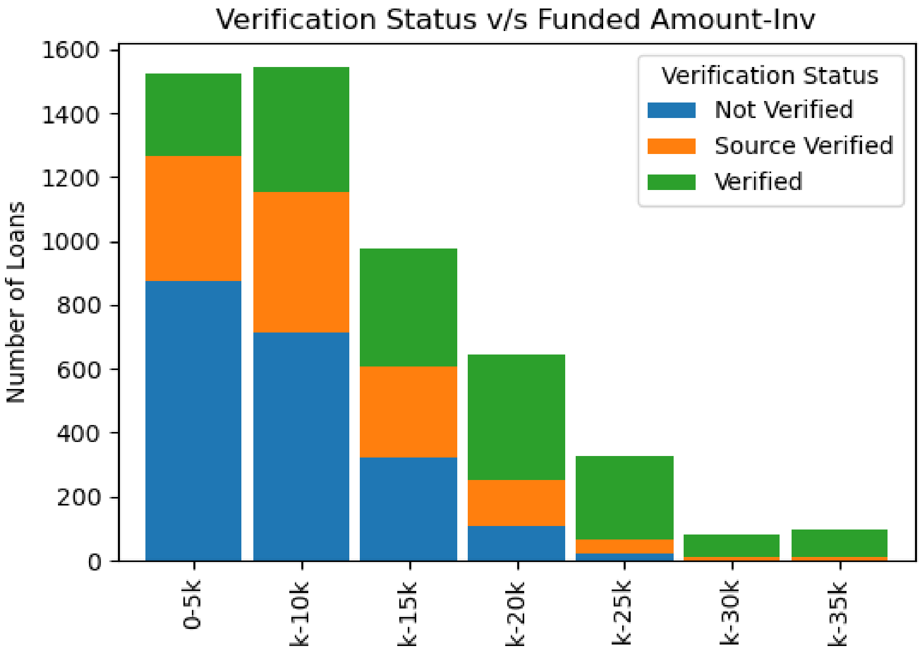
Loan Status Distribution



14.97% loans out of total loans are Charged Off

Variable : Verification status

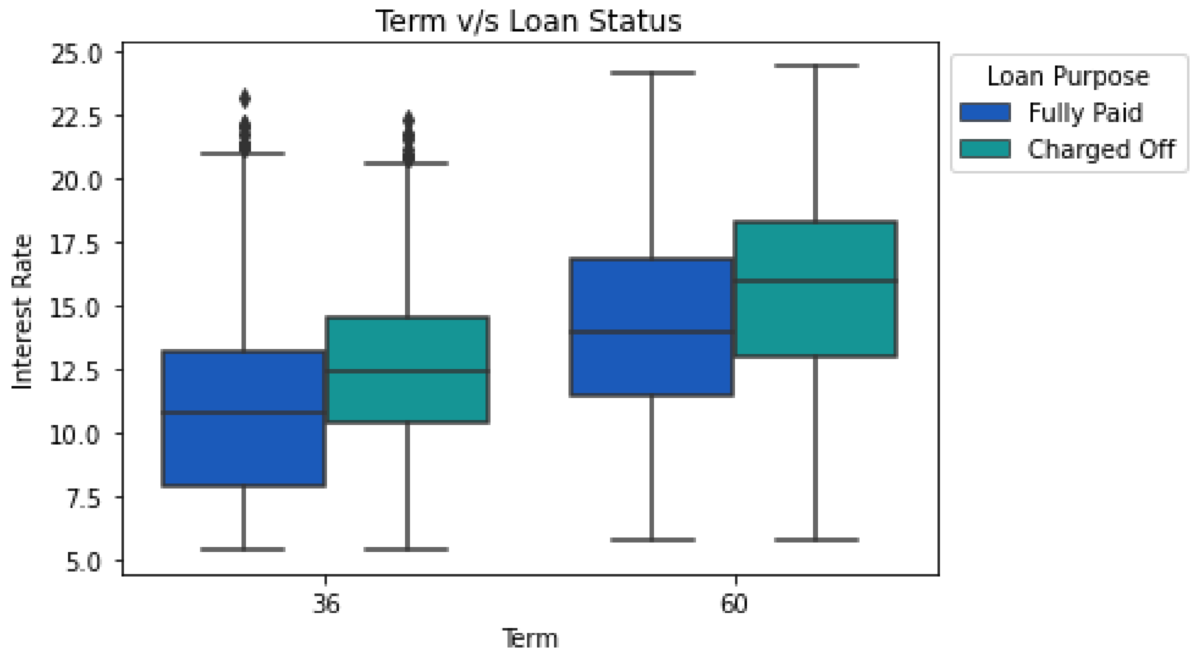


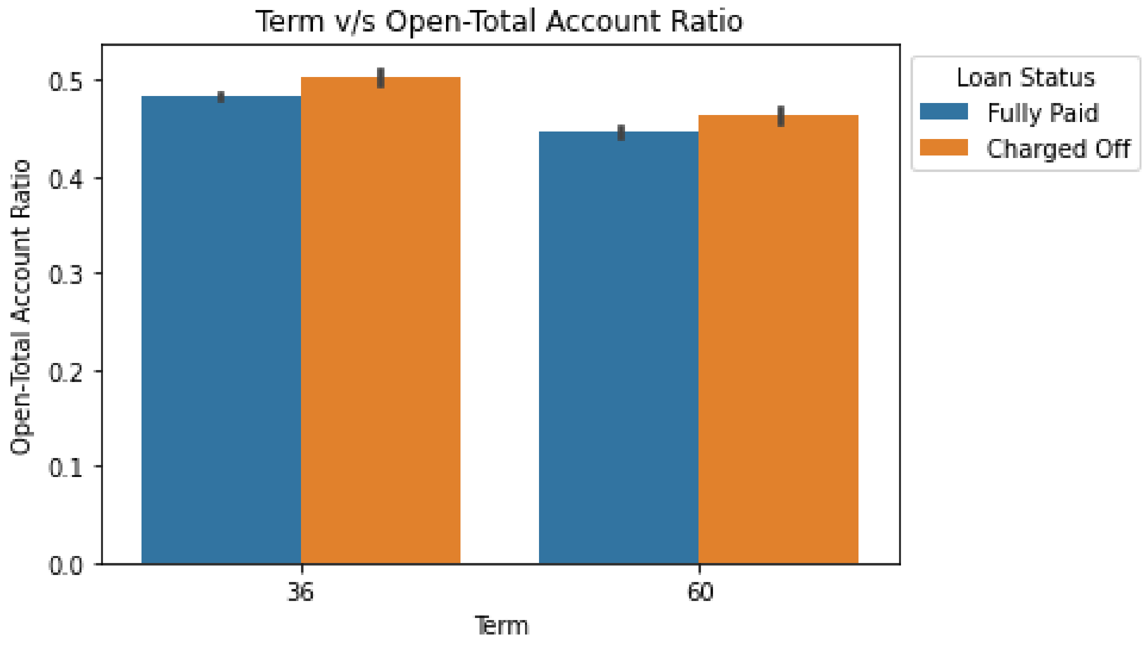


Account holders with annual income between 30k-45k with verification status as Not Verified tend to default

Account holders who got amount funded by investors between 0-5k and verification status as Not Verified tend to default

Variable : Term

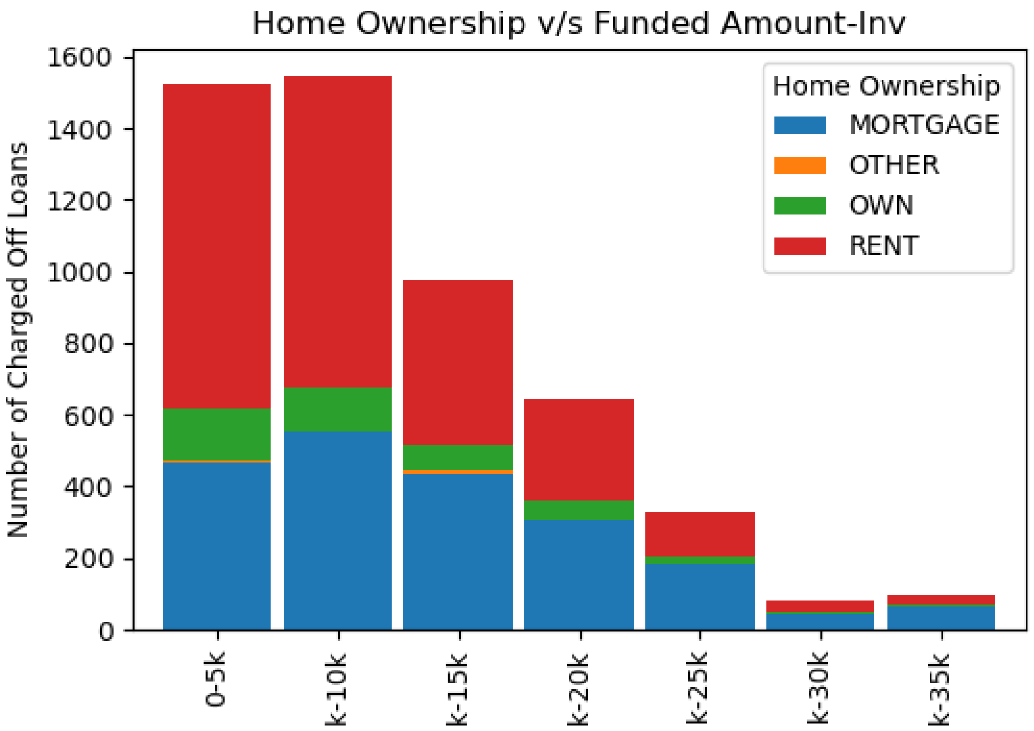




The ratio of Open-to-Total Account is higher for loans with loan term of 36 months. The above graph represent that borrowers with more open accounts, taking loan for 36 months tend to default.

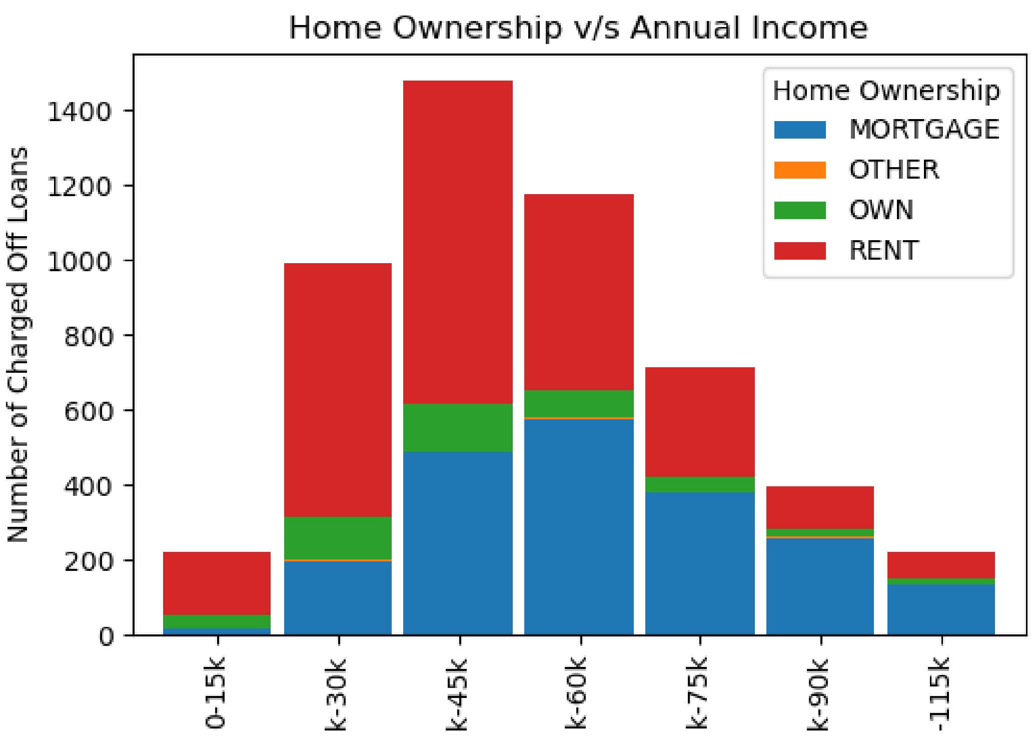
From the above two plots, it is evident that loans with lower term (36 months) and lower interest rate (~ 12.5%) tend to default comparatively.

Variable :
Home Ownership Vs Funded Amount



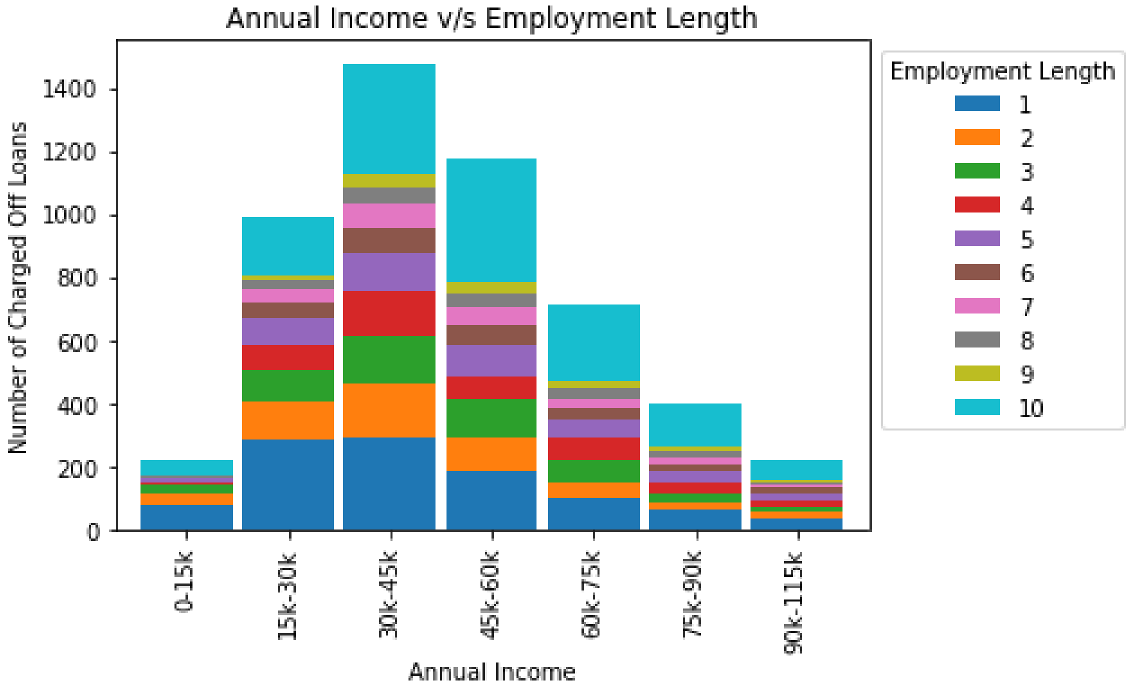
People with Mortgage home ownership tend to default with loan amount between 5k-10k, whereas people with Rent home ownership default with loan amount between 0-5k.

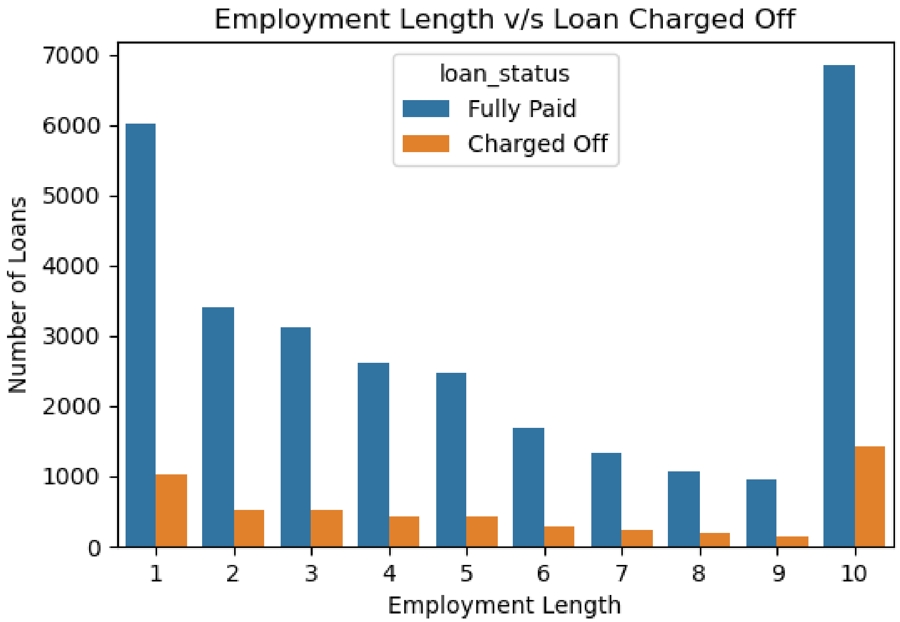
Variable :
Home Ownership Vs Annual Income



People with Annual Income between 30k-45k staying in Rented Accommodation , followed by people with Annual Income between 45k-60k having Mortgage home ownership.

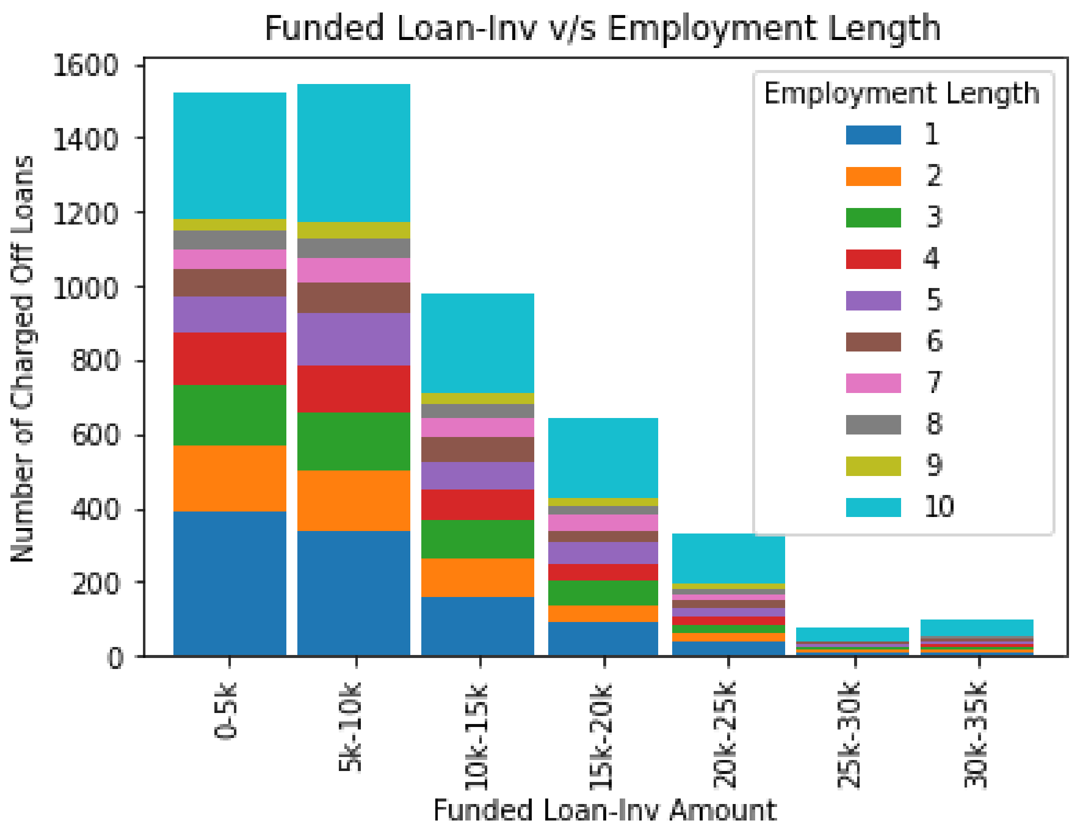
Variable : Employment length Vs Annual Income





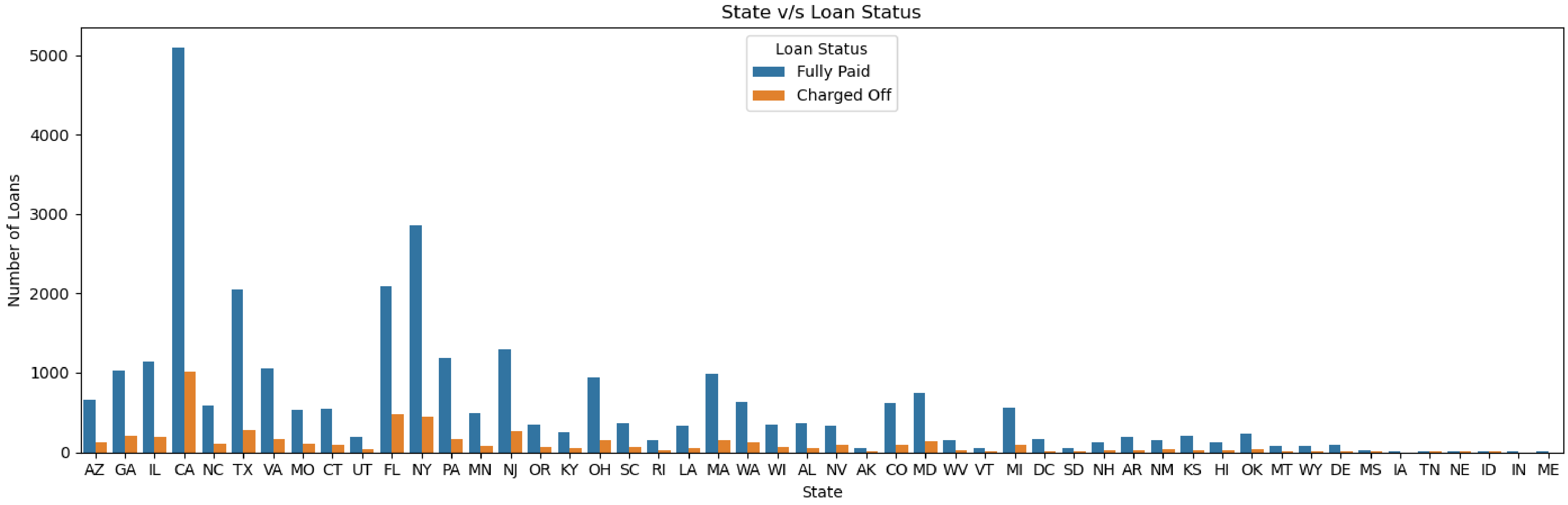
Borrowers with 10+ years of work experience and annual income in the range of 45k-60k tend to default. However, it is also visible that there are considerable number of people with 1 or less year of work experience and annual income in the range of 15k-45k who contribute to defaulters list

Variable : Employment length Vs Funded Loan Amount



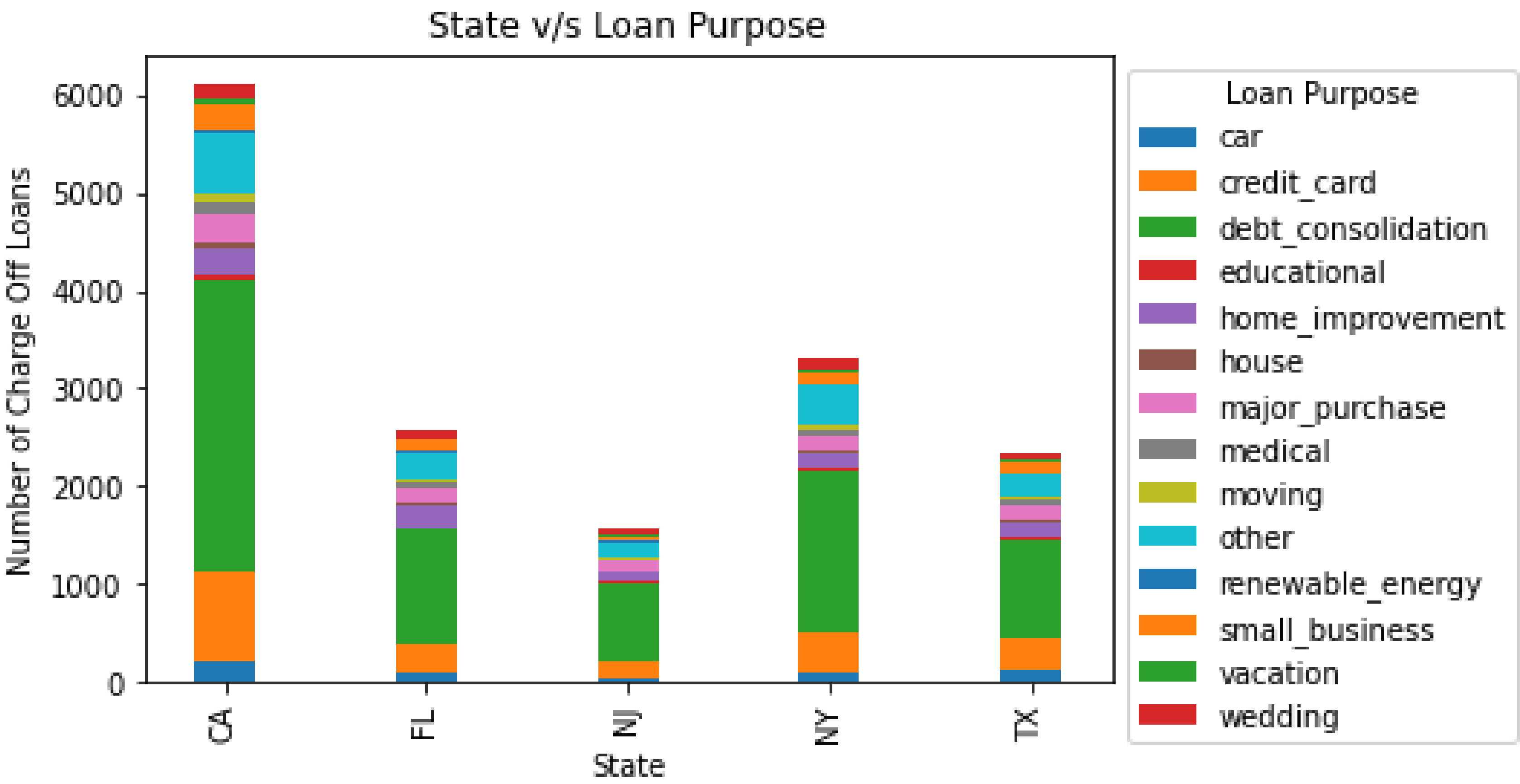
Borrowers with 10+ years of work experience and funded loan amount in the range of 5k-10k tend to default. Also, people with 1 or less year of work experience and funded loan amount in the range of 0k-5k contribute to defaulters list.

Variable : States



Maximum loan is borrowed by the people living in California(CA), followed by New York(NY), Florida(FL), Texas(TX) and New Jersey(NJ). The charged off proportion is also in the same order of the loan borrowed for these states.

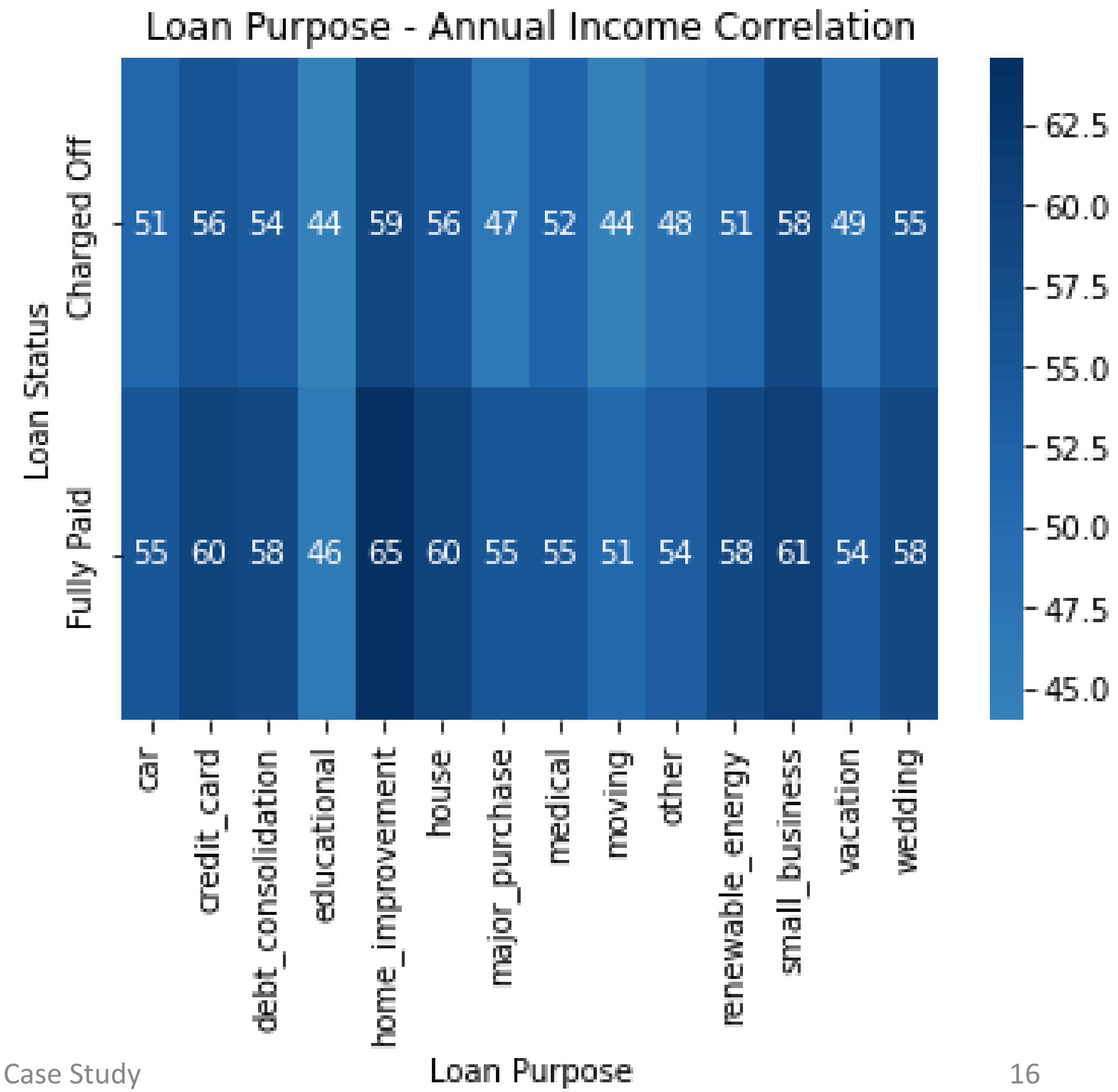
Variable : States



People living in the top 5 states of the U.S., which have the maximum number of Charged Off loan, borrowed loan for debt consolidation.

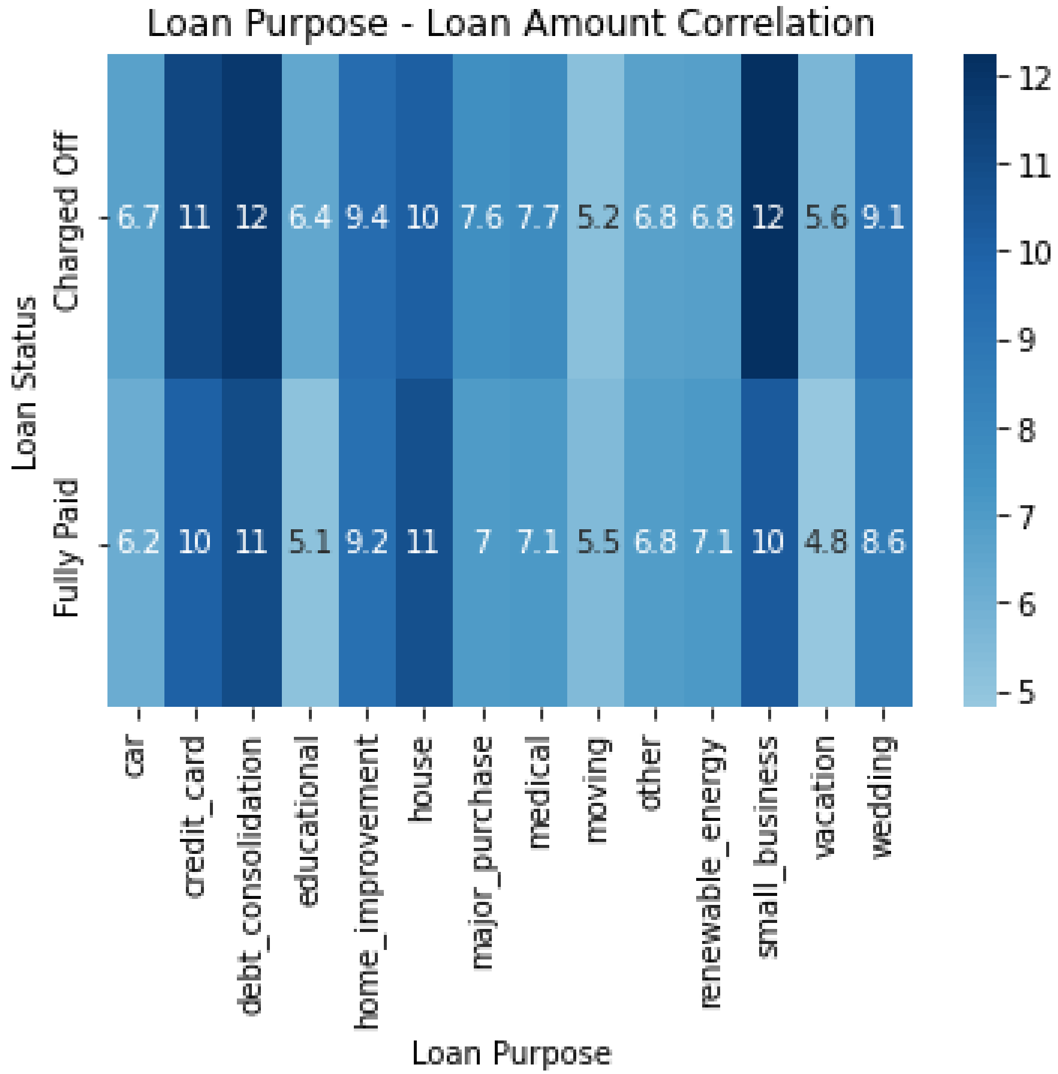
Correlation –
Loan Purpose Vs Annual Income

The number of loans borrowed for home improvement & small business with the annual income around 59k & 58k respectively contribute to defaulters.

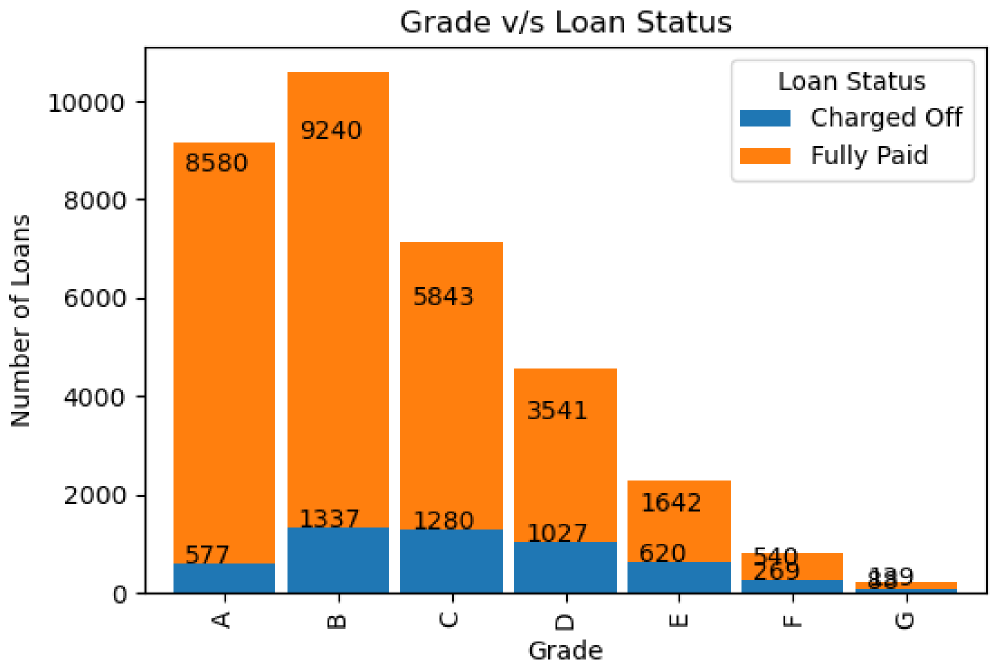


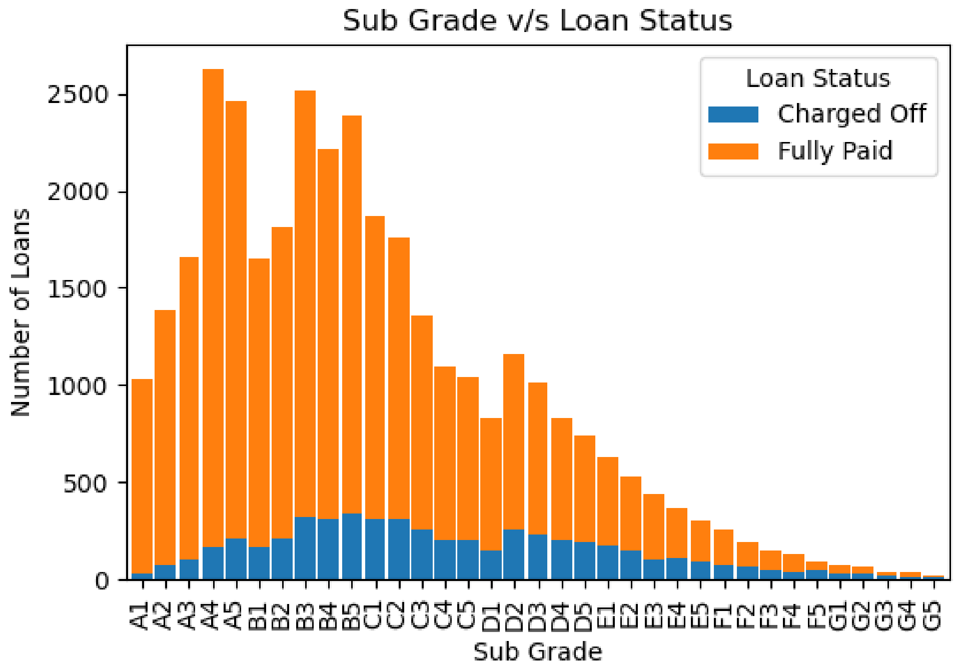
Correlation –
Loan Purpose Vs Funded Loan Amount

The number of loans taken for debt consolidation & small business with the funded loan amount around 12k tend to Charged Off.



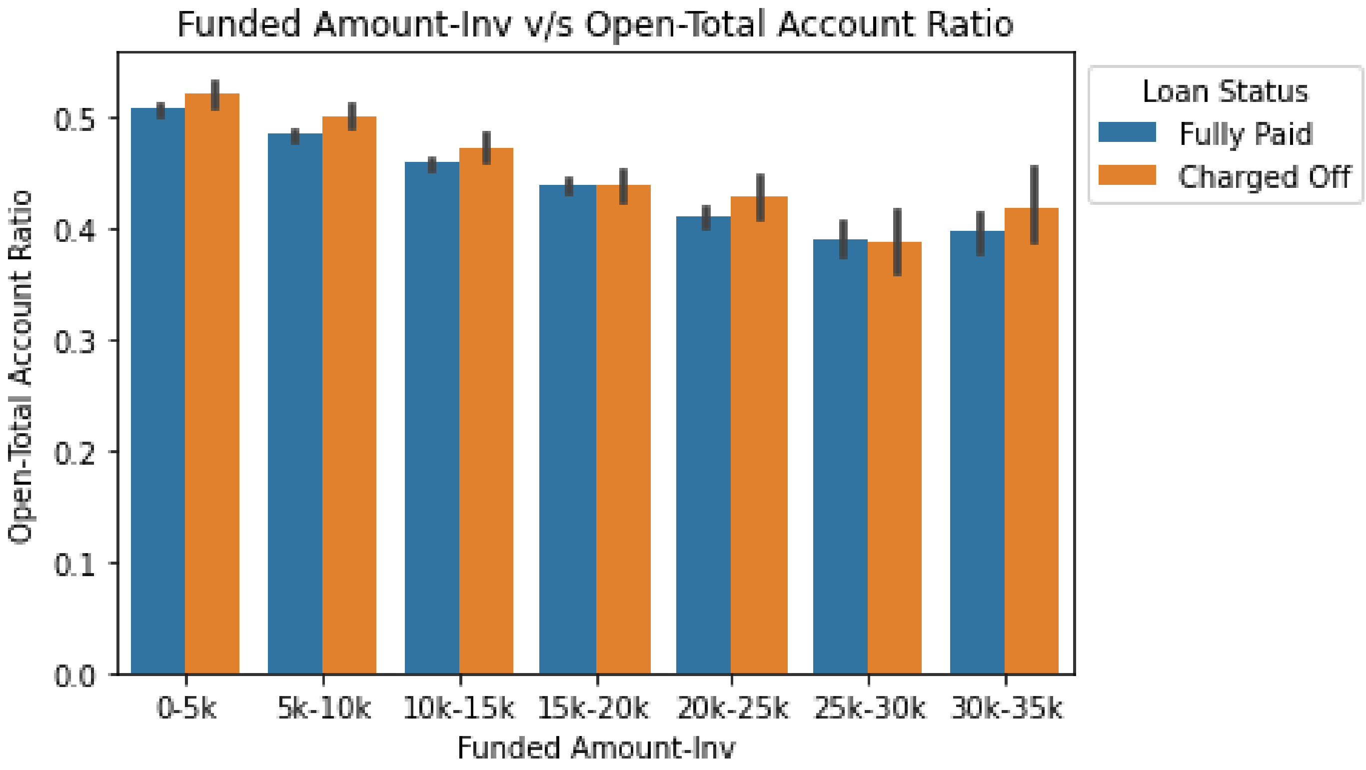
Variable : Grades and Subgrades





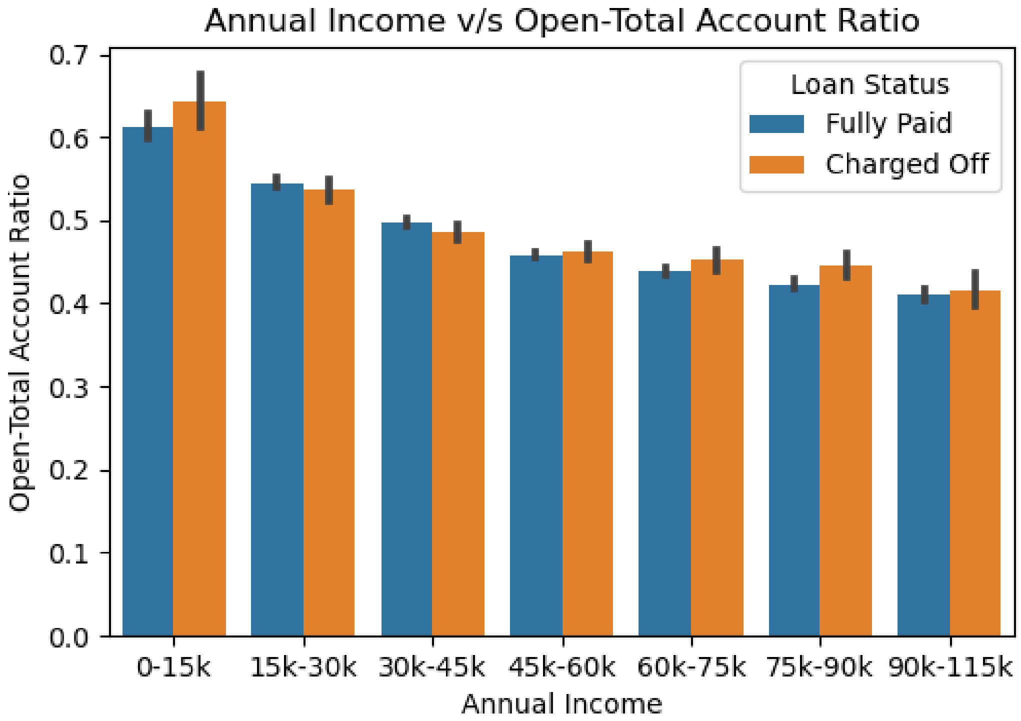
From the above two plots, it can be deduced that Loan with grade B and sub-grade B5 comparatively Charged Off more.

Variable : Funded Amount Vs Open to Total account



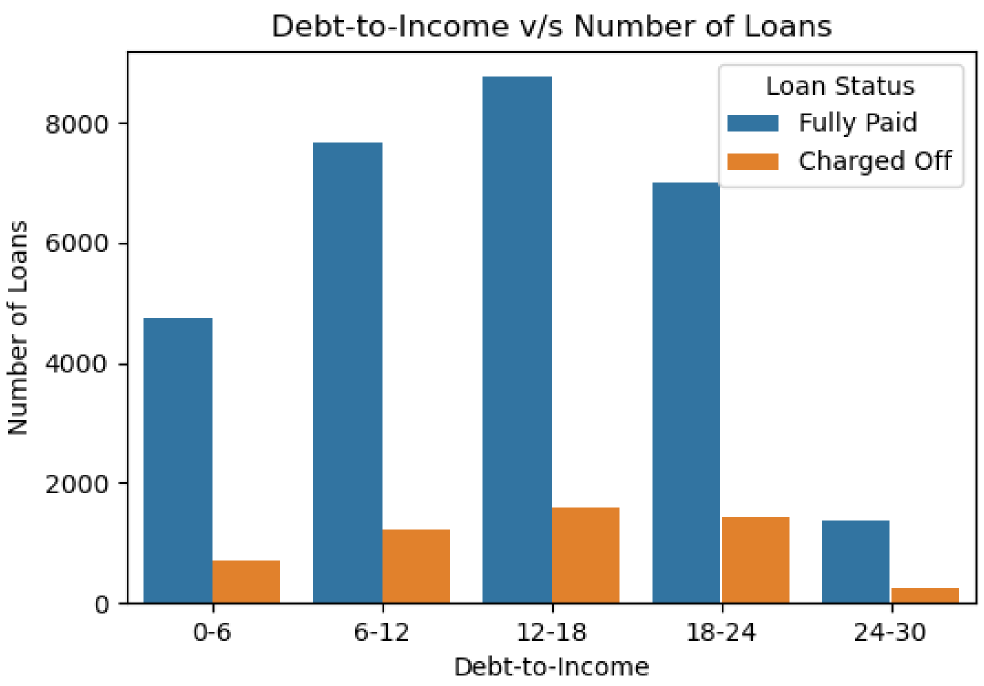
Charged Off loans are comparatively higher for customers having Open-to-Total Account Ratio more than 0.5, which means a customer having more Open accounts with funded loan amount in the range of 0-5k tend to default.

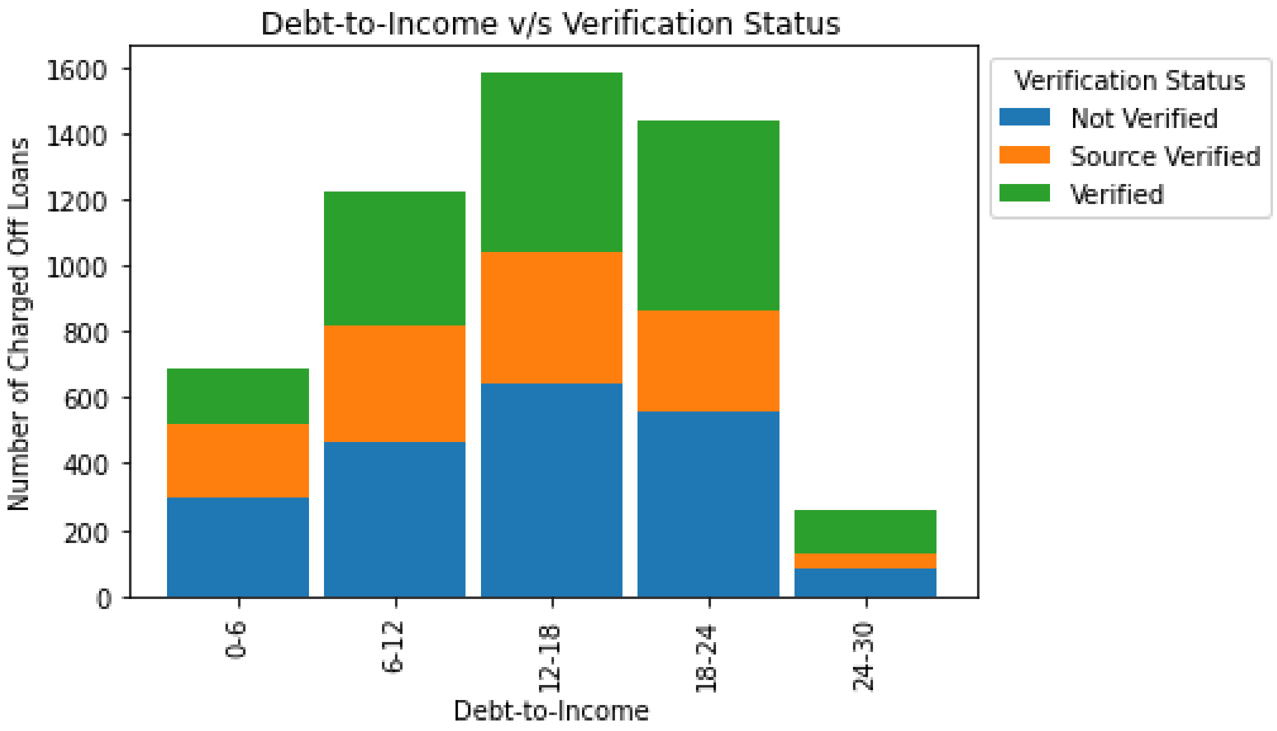
Variable :
Annual Income Vs Open to Total account



Charged Off loans are comparatively higher for customers having Open-to-Total Account Ratio more than 0.6, which means a customer having more Open accounts with annual income in the range of 0-15k tend to default.

Variable : Debt to Income





The maximum number of Charged Off loans are in the DTI (debt-to-income) range of 12-18. Loans which are Not Verified in the DTI range of 12-18 tend to Charged Off.

Observations & Recommendations

- Loans with funded amount between **0-5k** & loans with borrowers annual income between **30k-45k** needs to be **verified.**
- Borrowers with high open to total account ratio and interest rate **(~ 12.5%)** for the loan term of **36 months**, tend to default, hence need to be monitored
- Borrowers with **10+ years** and **less than 1 year** of work experience with annual income in the range of **45k-60k** and **15k-45k** respectively, and funded amount **5k-10k** and **0-5k** respectively, tend to default, hence to be asked to submit some security against the loan
- Citizens from **California** who borrowed loan for **debt consolidation** need to be watched

Observations & Recommendations

- The borrowers with **grade B** and **sub-grade B5** are tend to be Charged Off more.
- People having **rented home** tend to charged off, followed by ones having home ownership as **Mortgage**.
- Account holders with high **open to total account ratio** with funded loan amount of **0-5k** and an annual income of **0-15k** are more prone to miss their repayment on time
- The maximum number of Charged Off loans are in the **DTI (debt-to-income)** range of **12-18**. Loans which are **Not Verified** in the DTI range of **12-18** tend to Charged Off.

Thank you...!