

Loan Data Case Study

Group Members:

Pankit Bhat

Achyuthramaiah K

The problem

Company

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.

Context

Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

Problem statement

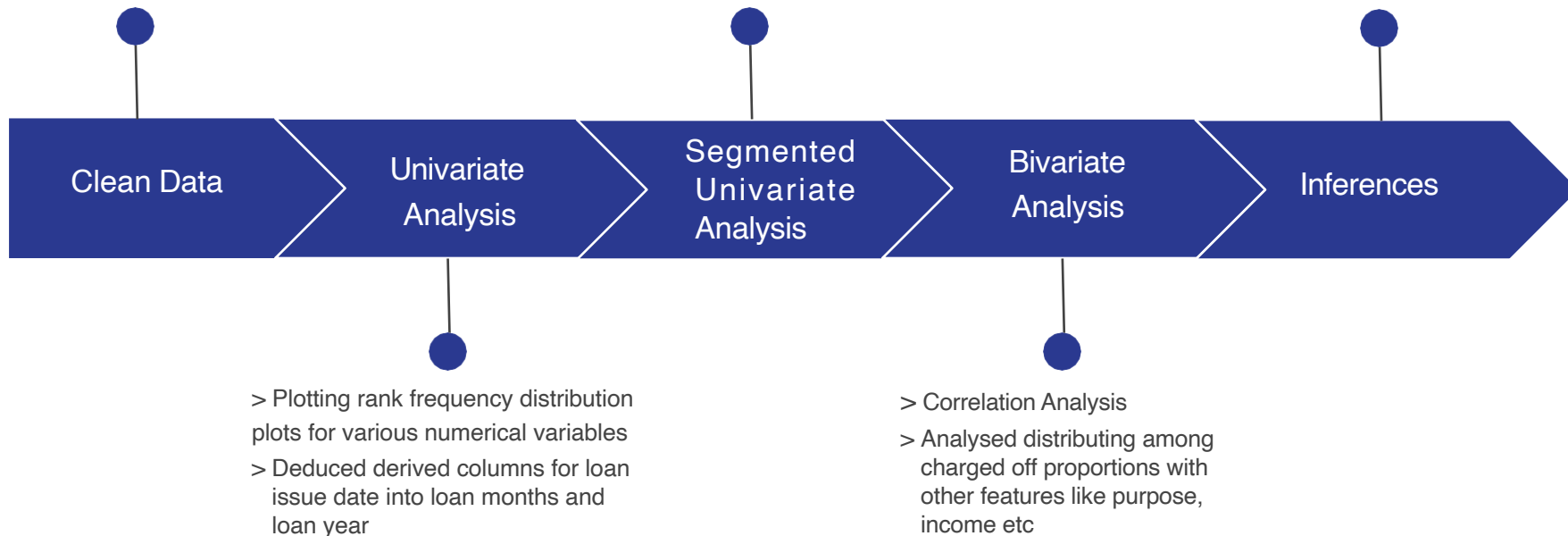
As a data scientist working for Lending Club analyse the dataset containing information about past loan applicants using EDA to understand how consumer attributes and loan attributes influence the tendency of default

Analysis Approach

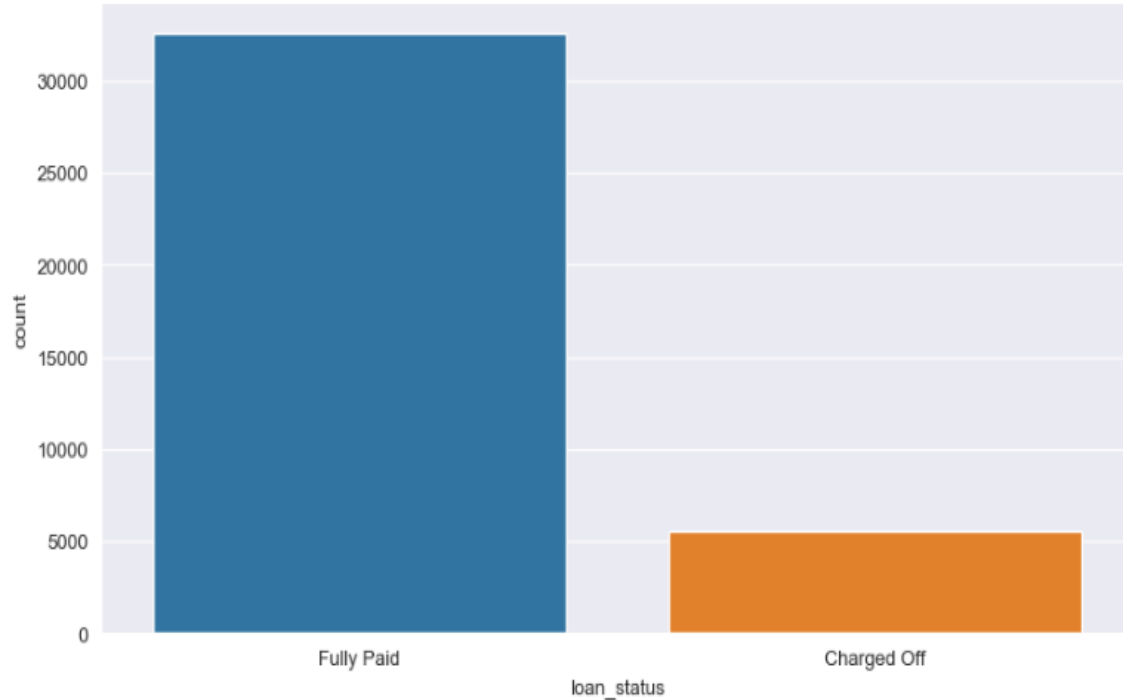
- > Drop Columns with Single valued entries , null values , irrelevant values, high cardinal values
- > Drop columns with 50% missing values, fixing the data type and removing outliers

- > Analysing purpose of loan w.r.t loan status etc
- > Plotting segmented categorical plots

Observation from the whole analysis



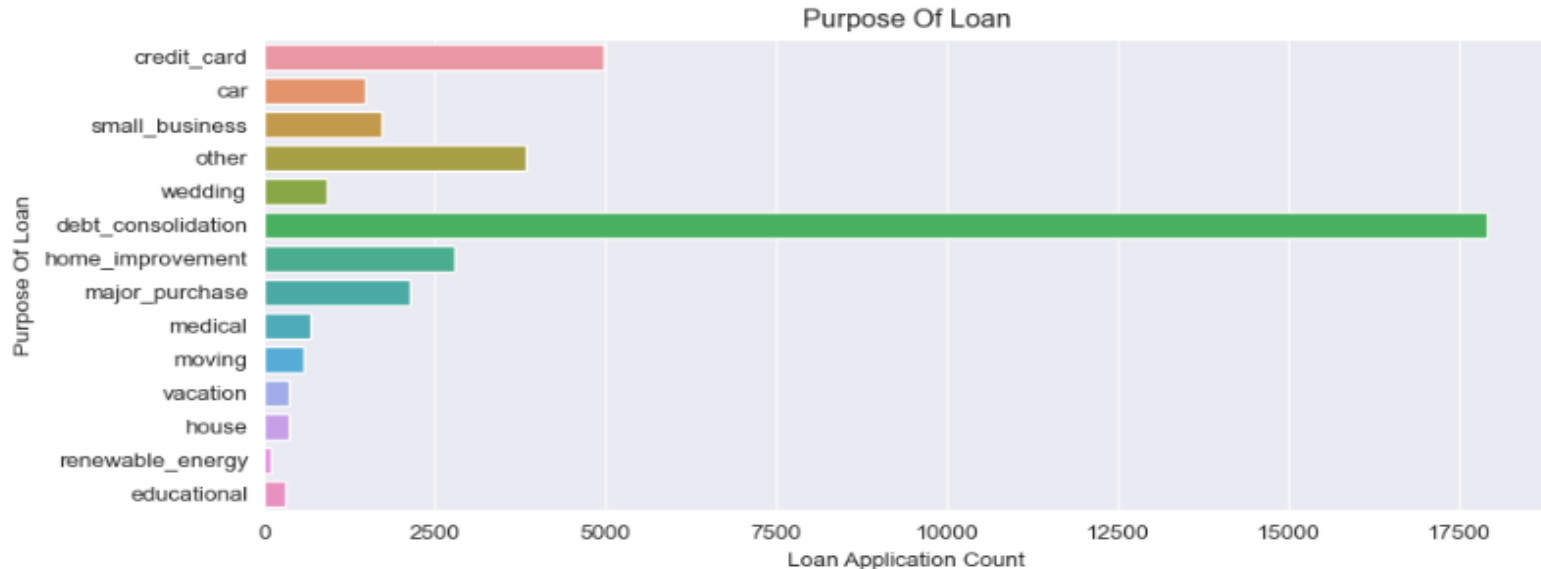
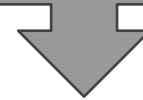
Analysis - Overall Loan Status



Charged Off loans are low in number as compared to Fully Paid Loans

Analysis - Understanding Loans

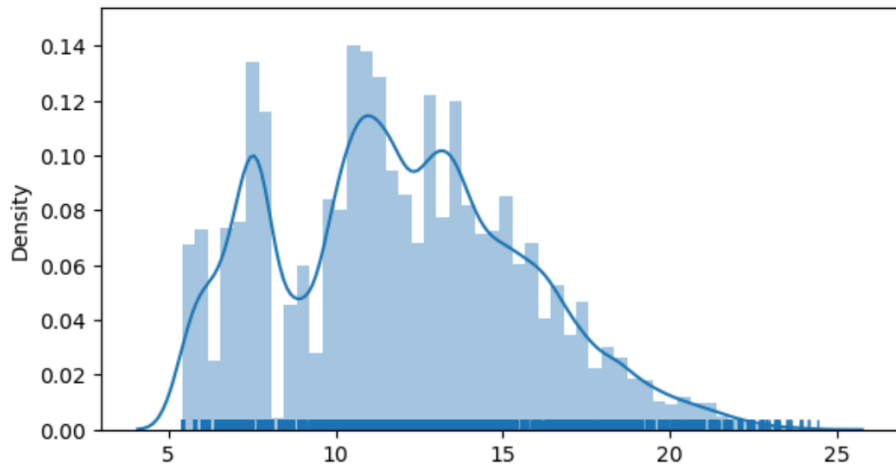
most of the loans were taken for the purpose of debt_consolidation and credit_card payment. The no. of charged off loans is also high for these loans



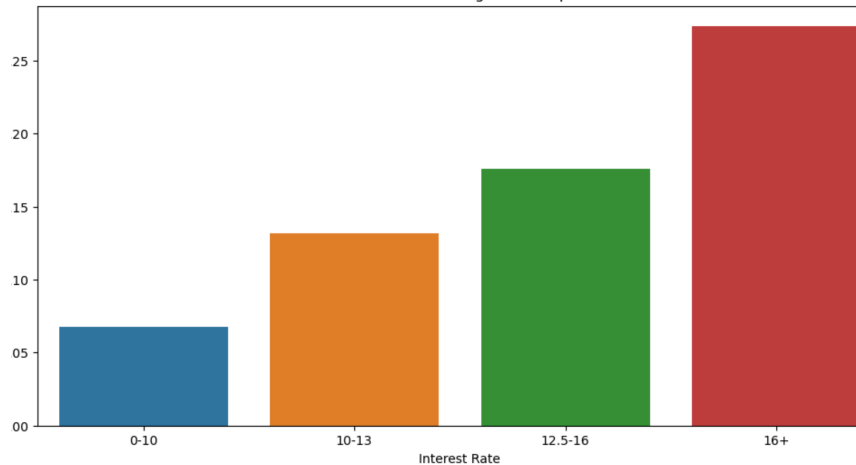
Analysis - Defaults by Interest Rate

Interest Rates which are less than 10% have very less chances of charged off. Charged Off Proportion increases with higher interest rates

Interest Rate-Distribution Plot

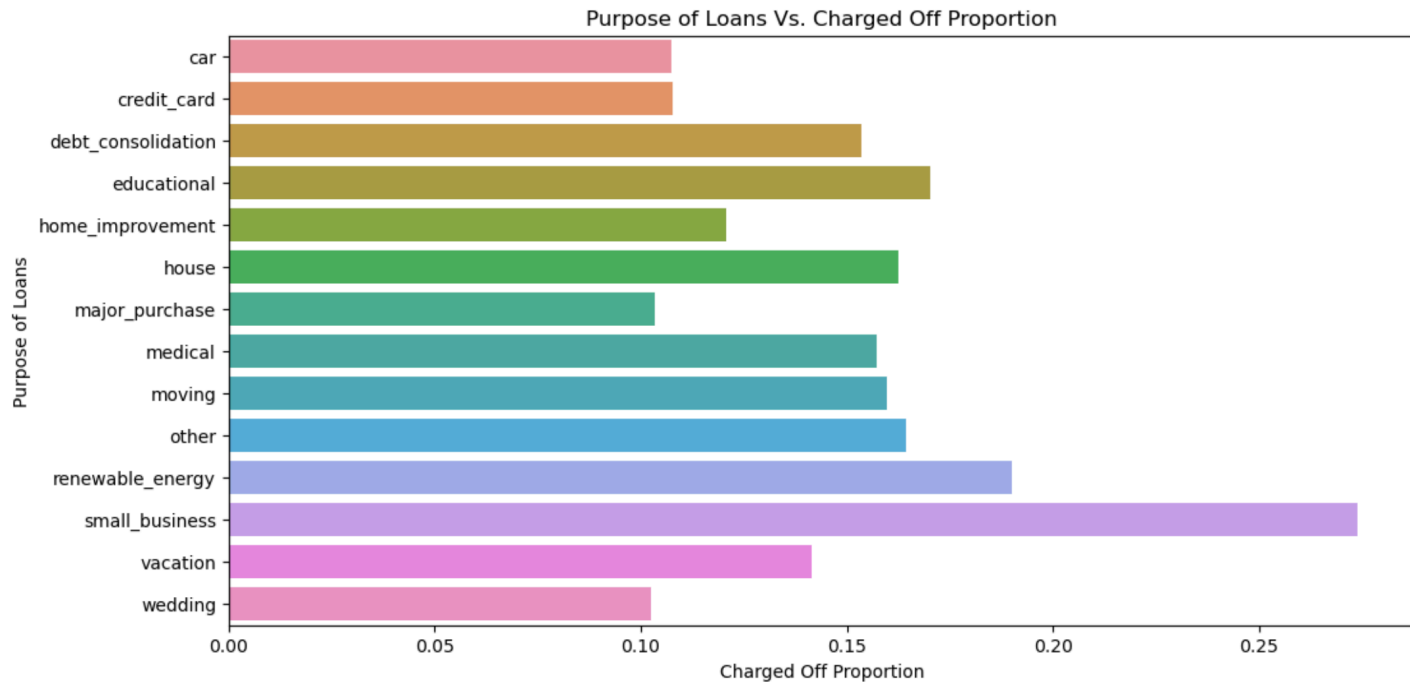


Interest Rates vs Charged Off Proportion



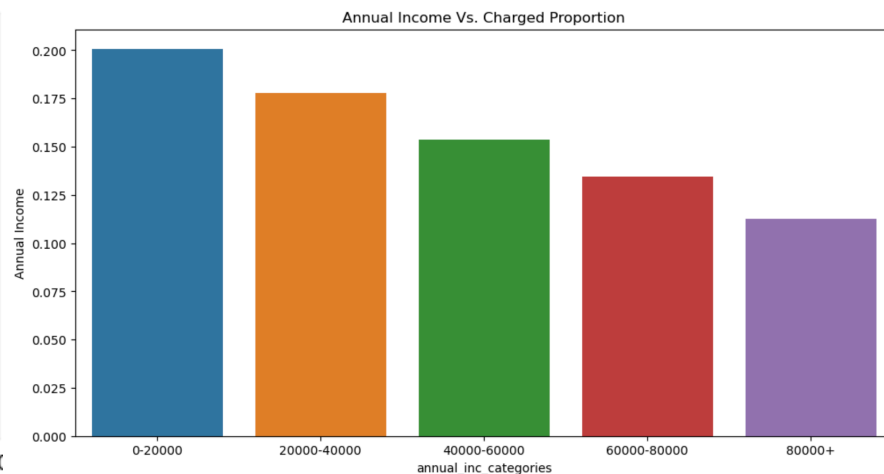
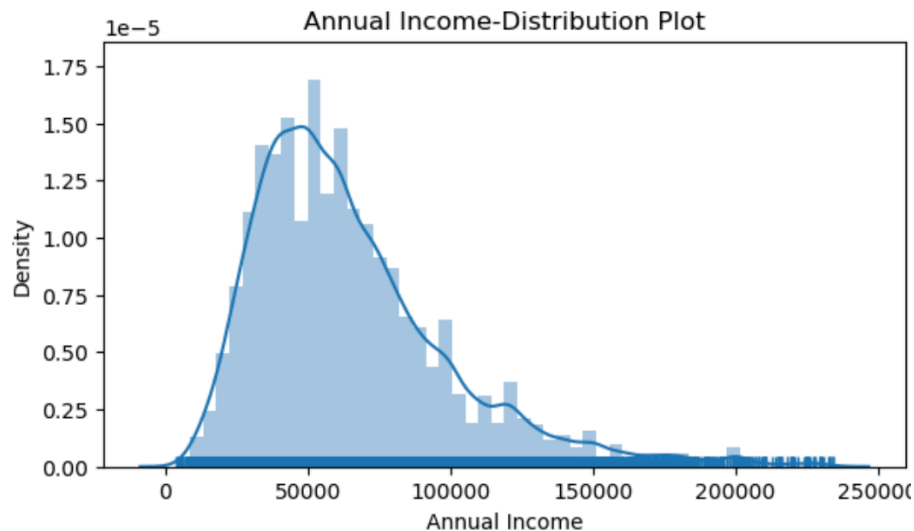
Analysis - Defaults by Loan Purpose

small business applicants have high chance of getting charged off. Also, renewable energy is better than other categories .

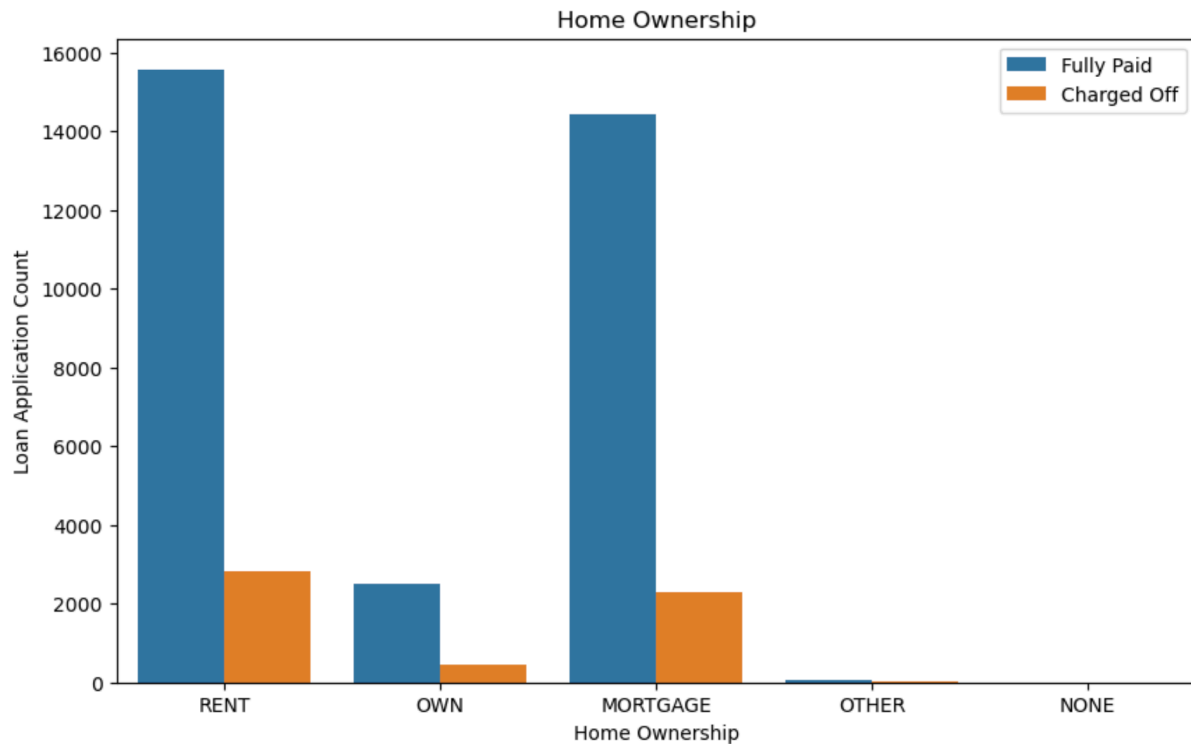


Analysis - Defaults by Borrower's Income

Annual Income range 0-20k has high chances of charged off. Annual Income range 80k+ have less chances of charged off. Annual income is inversely proportional to Charged off .



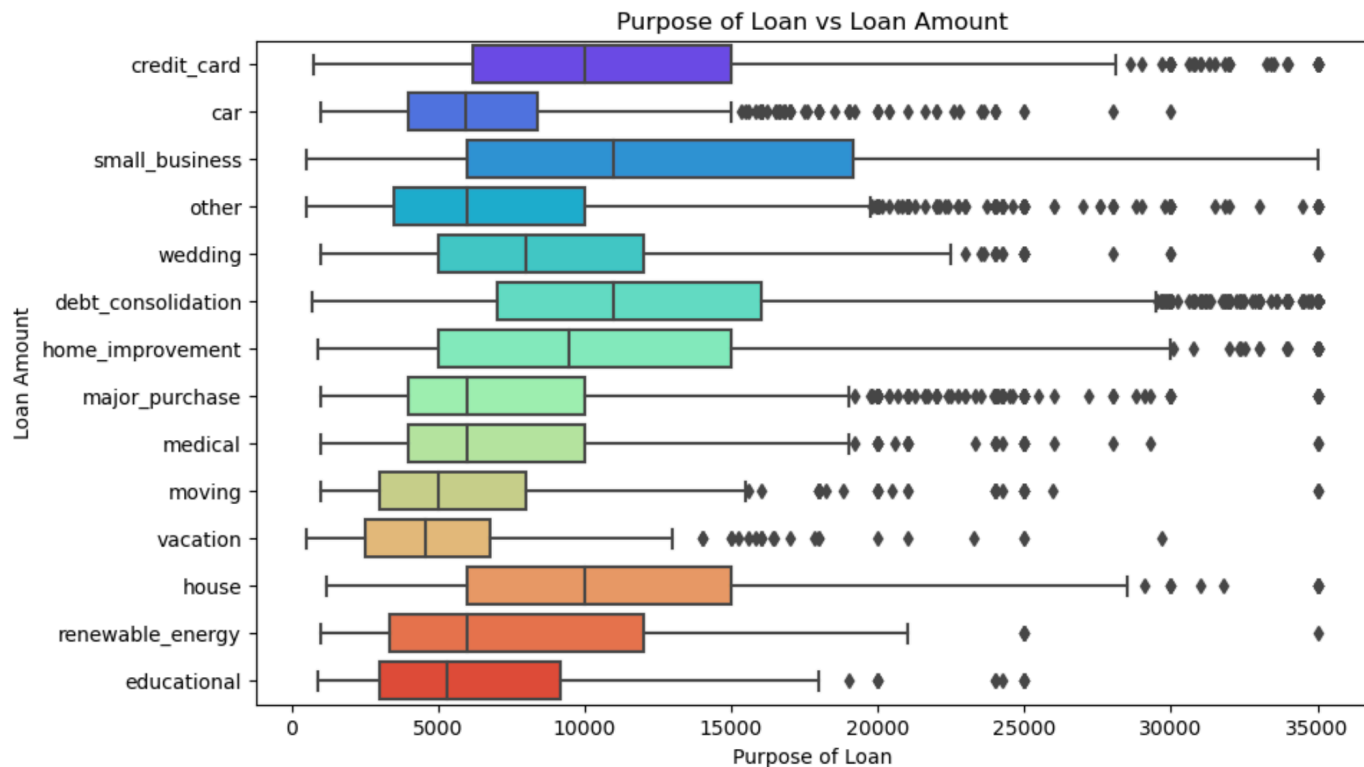
Analysis - Defaults by home ownership



most of them live in rented
home
or have mortgaged their home

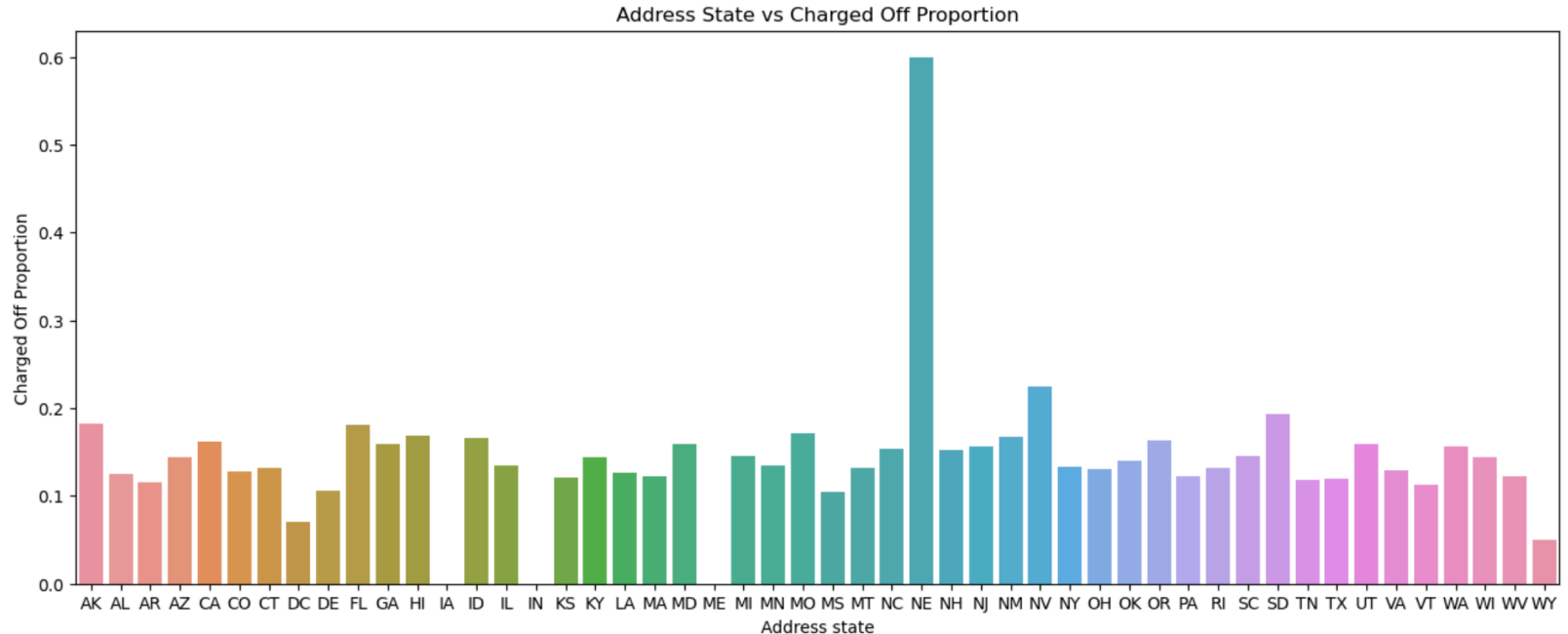
Analysis - Defaults by Purpose vs Loan Amount

Median, 95th percentile, 75th percentile of loan amount is highest for loan taken for small business purpose among all purposes. Debt consolidation is second and Credit card comes 3rd.



Analysis - Defaults by Address State

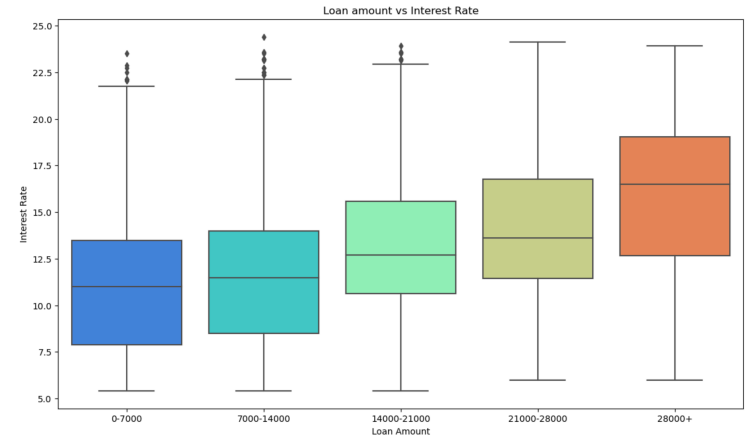
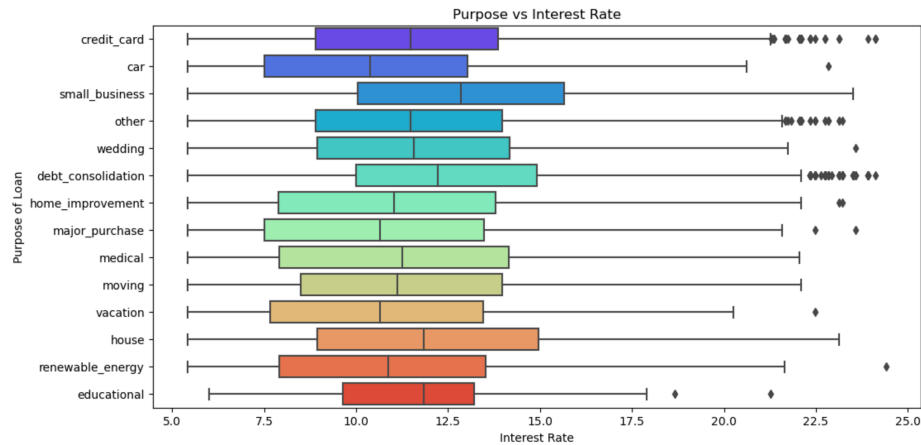
States NE has very high chances of charged off but number of applications are too low to make any decisions. States NV, CA and FL states shows good number of charged offs in good number of applications.



Analysis - Defaults by Purpose vs Interest Rate vs Loan Amount

The average interest rate is higher for small businesses. Debt consolidation is second .

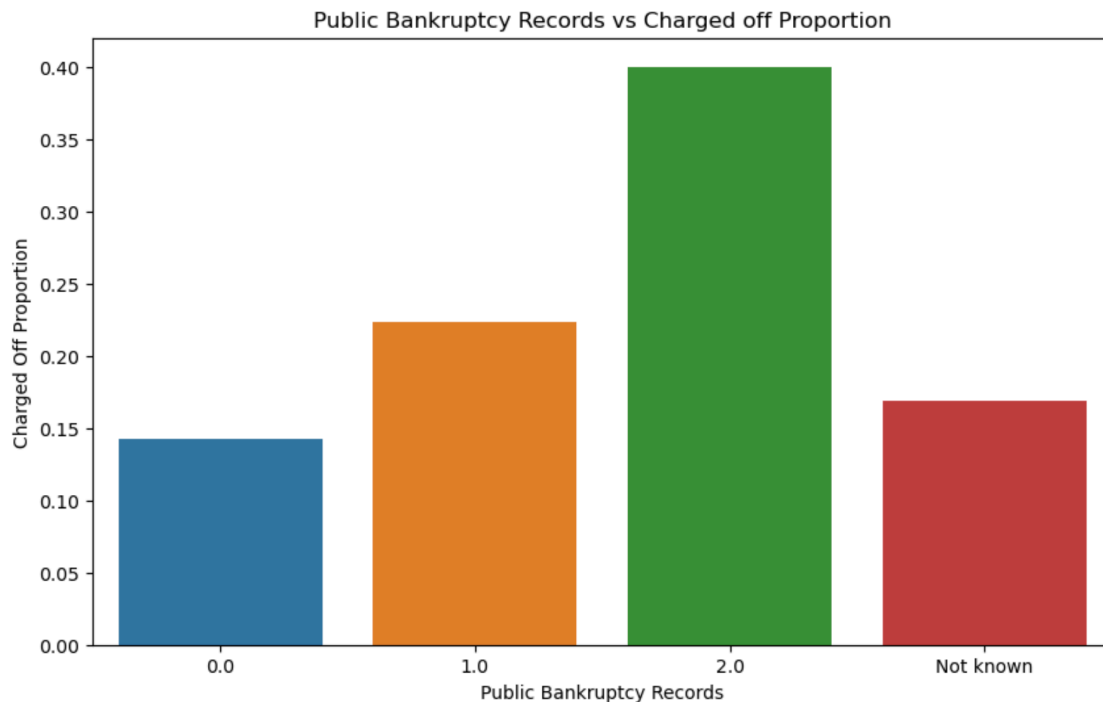
Loans taken for small businesses had to repay the loan with more interest rate as compared to others. Interest Rate is increasing with loan Amount. When the loan amount is more, it's probable that its taken for a longer term and that's why higher interest rate.



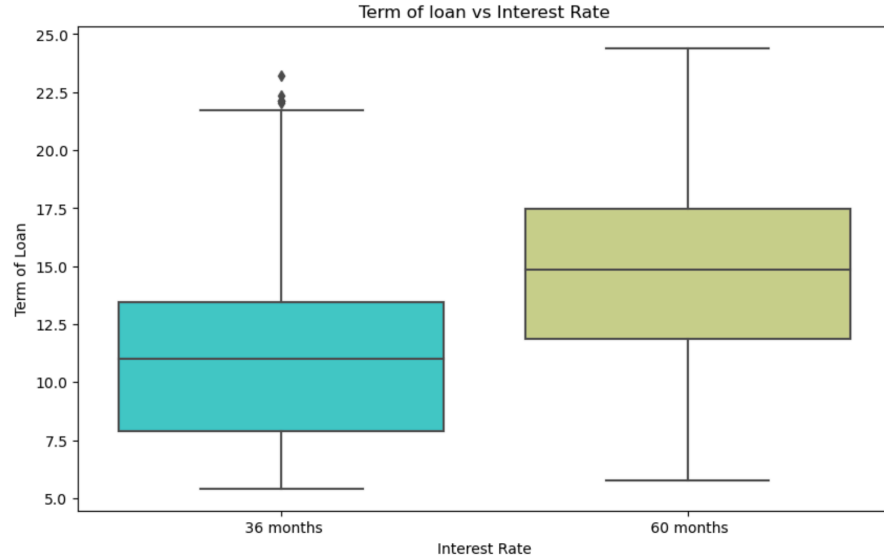
Analysis - Defaults by Bankruptcy Records

Those who already have `pub_rec_bankruptcies` value 1, have charged off proportion higher than who have no `pub_rec_bankruptcies`.
`pub_rec_bankruptcies` count 2 has even higher charged off proportion.

Those who have defaulted before have more chances of defaulting in future as well.

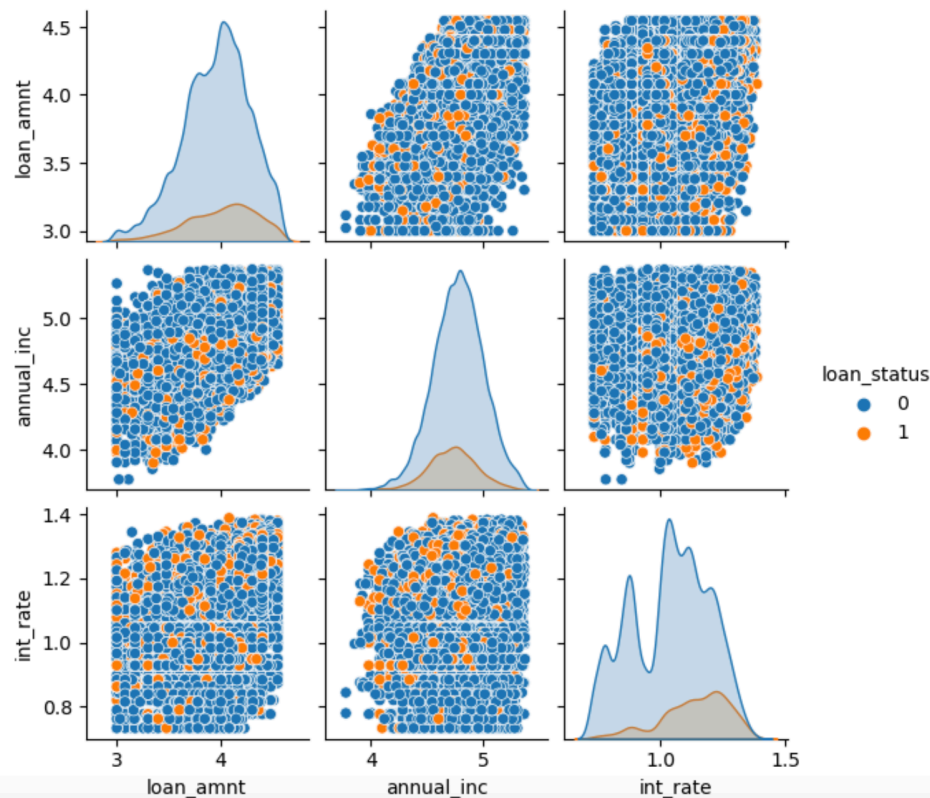


Analysis by Term of loan vs Interest Rate



The average interest rate is higher for 60 months loan term and most of the loans issued for longer term had higher interest rates of repayment

Analysis – Defaults by Interest Rate, Annual Inc, Loan Amount



Higher is the charged off ratio for higher interest rate. Also higher the annual income , higher the loan amount. Interest rate is increasing with loan amount and hence an increase in charged off

Inferences

There is a more probability of a loan app defaulting when :

1. Applicants whose home ownership is mortgage and have rented their home
2. Applications where the purpose of loan is debt consolidation & credit card payments
3. Applicants from small business , education and renewable energy
4. Applicants who have debt to income ratio as 10-15% with an annual income of 25k-35k
5. Applicants who have defaulted before have high chances of defaulting again
6. Applicants who have an annual income range of 0-20k
7. Applicants who have high loan amount thus resulting in higher interest rate and higher increase in charged off as well
8. Applicants from NV, CA and FL states have good number of charged off applications
9. Applicants repaying loan in 60 months term



Thank You