Lending Club Case Study

Presented By – Agniva Chakraborty

Objective

- The company wants to identify risky loan applicants and reduce the amount of credit loss.

- Identification of such applicant's using EDA is the aim of this case study.

- To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Data Cleaning

In this section the data was cleaned to remove any columns/rows which do not help with our analysis

- Columns having all Nan's were identified and deleted

- Columns having all zeroes were identified and deleted

- Columns with <3 unique values were deleted as they had very less information for our analysis.

- Keeping the 'ID' column rest all columns having count same as ID was deleted as 'ID' gives us the distinct count

- Other columns were analysed and deleted as they didn't help with the analysis on loan defaulters.

- Rows where Loan Status = 'Current' were removed as the applicants had already started paying their loans and this did not help with our analysis.

Data Standardization

In this section the data was standardized to get rid of inconsistencies and create categories for proper analysis
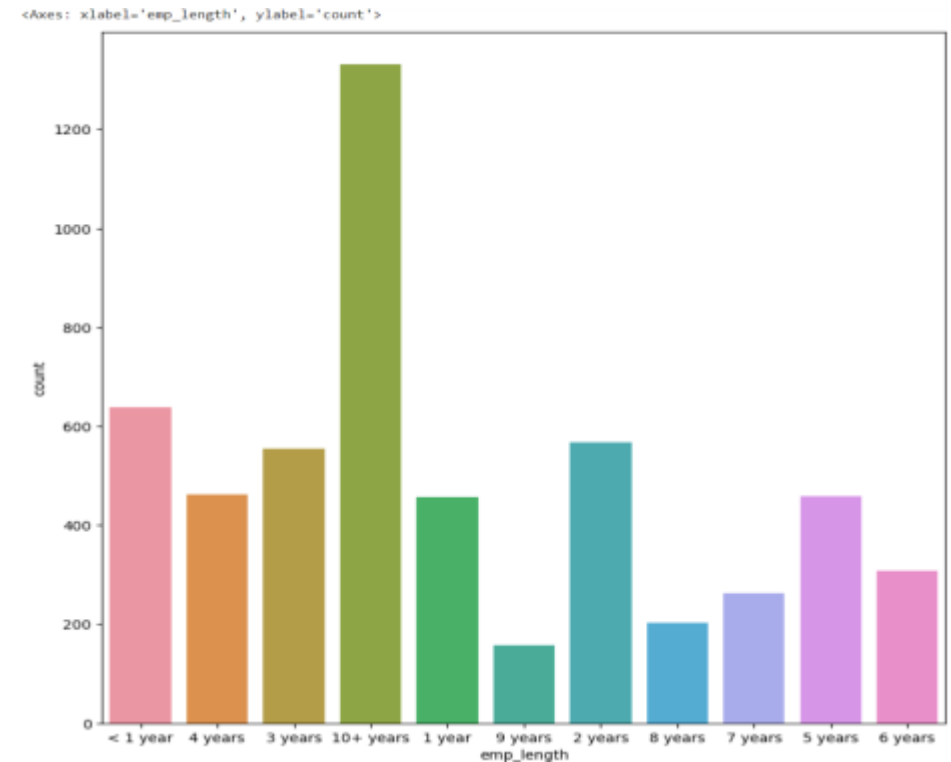
- Removing % from 'int_rate' column

- 'Verified' & 'Source Verified' meant the same in verification_status column so 'Source Verified' was replaced with 'Verified' .

- 4 new Columns were added to create buckets for 'int_rate', 'dti' , 'annual_inc' & 'loan_amnt'.
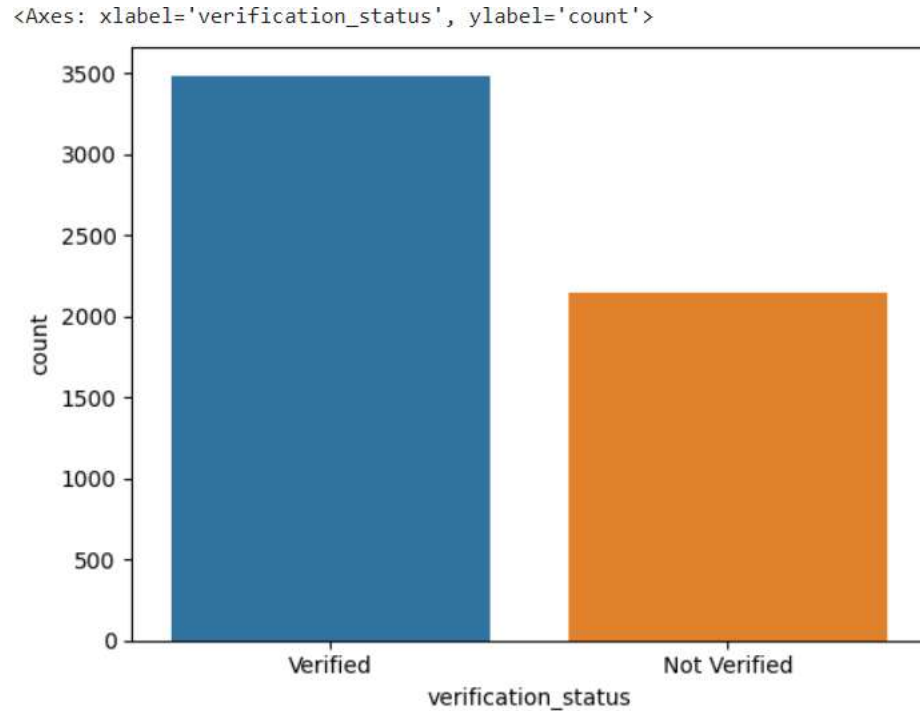
Home Ownership
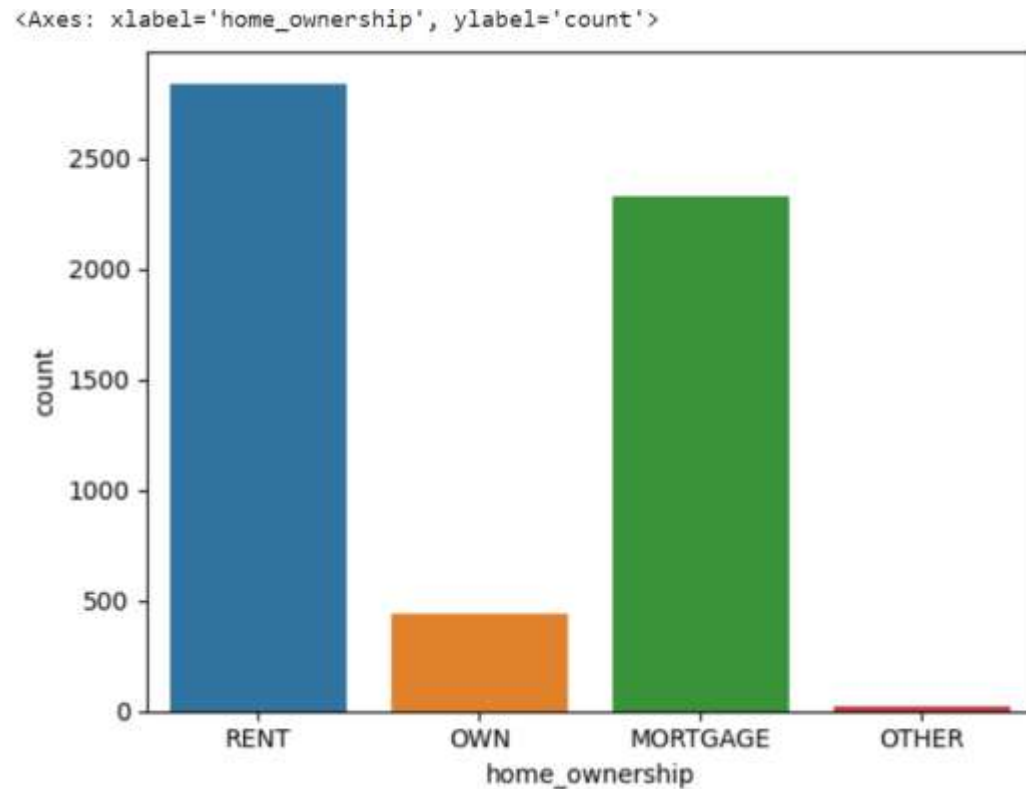
Home Ownership: Most defaulters are having 'RENT' followed by 'MORTGAGE' as home ownership
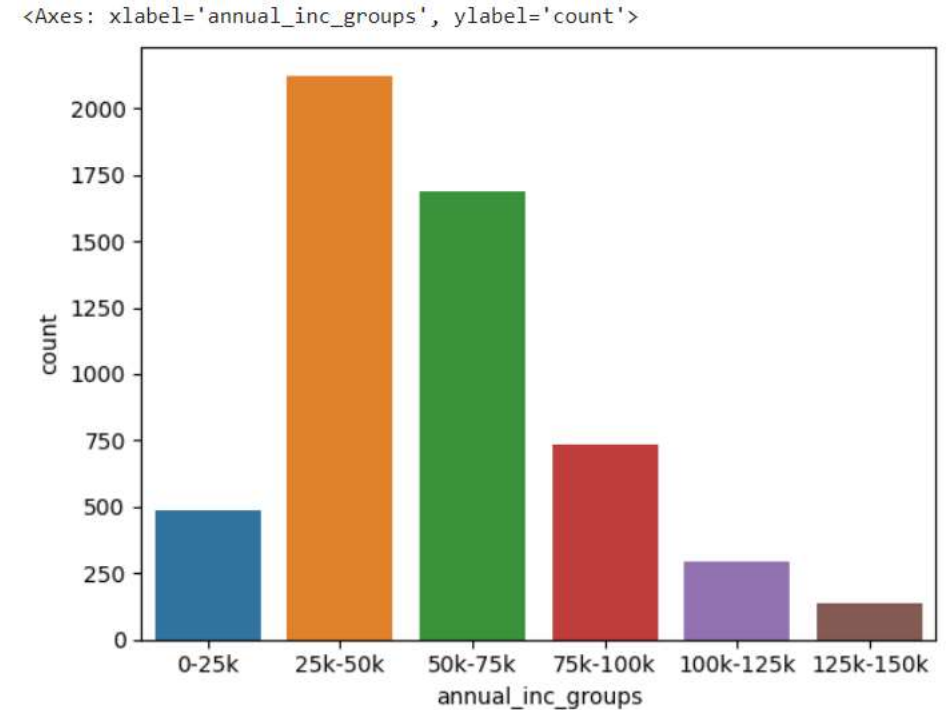
Employment Length

Employment Length : 10+ Years of experience have the highest defaulters.
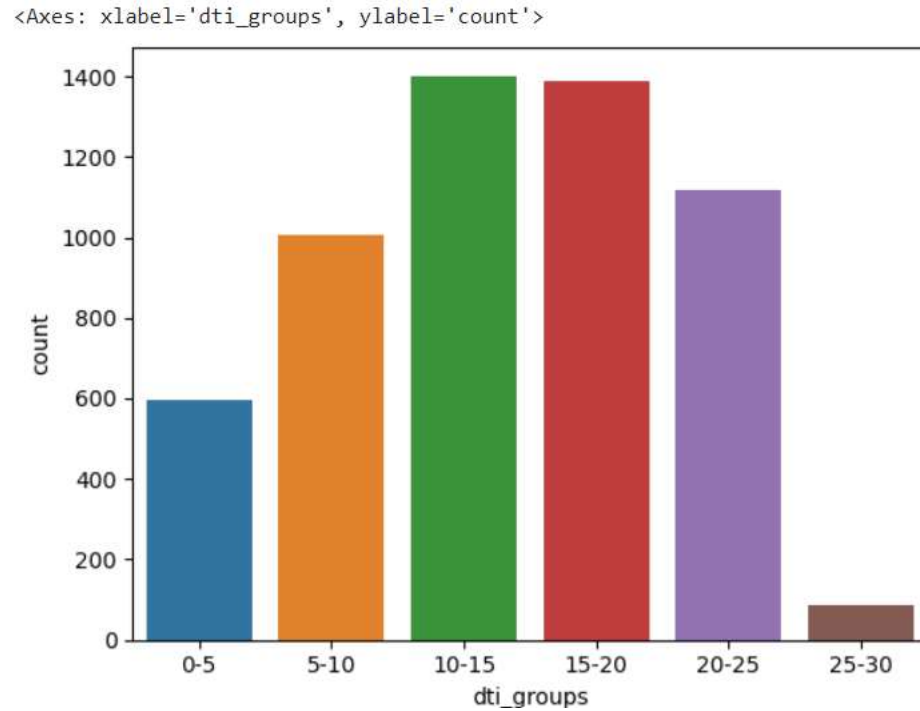
Verification Status

Annual Income

<Axes: xlabel='verification_status', ylabel='count'>

<Axes: xlabel='annual_inc_groups', ylabel='count'>

Verification Status: 'Verified' applicants are the highest defaulters
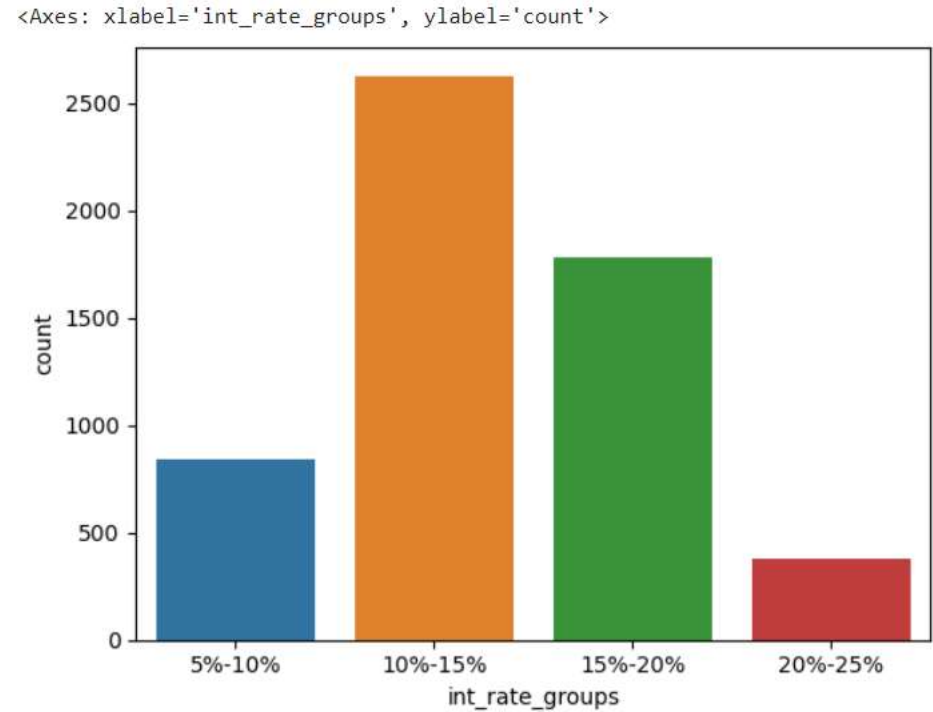
Annual Income: Annual Income range between 25-50k have the highest defaulters.
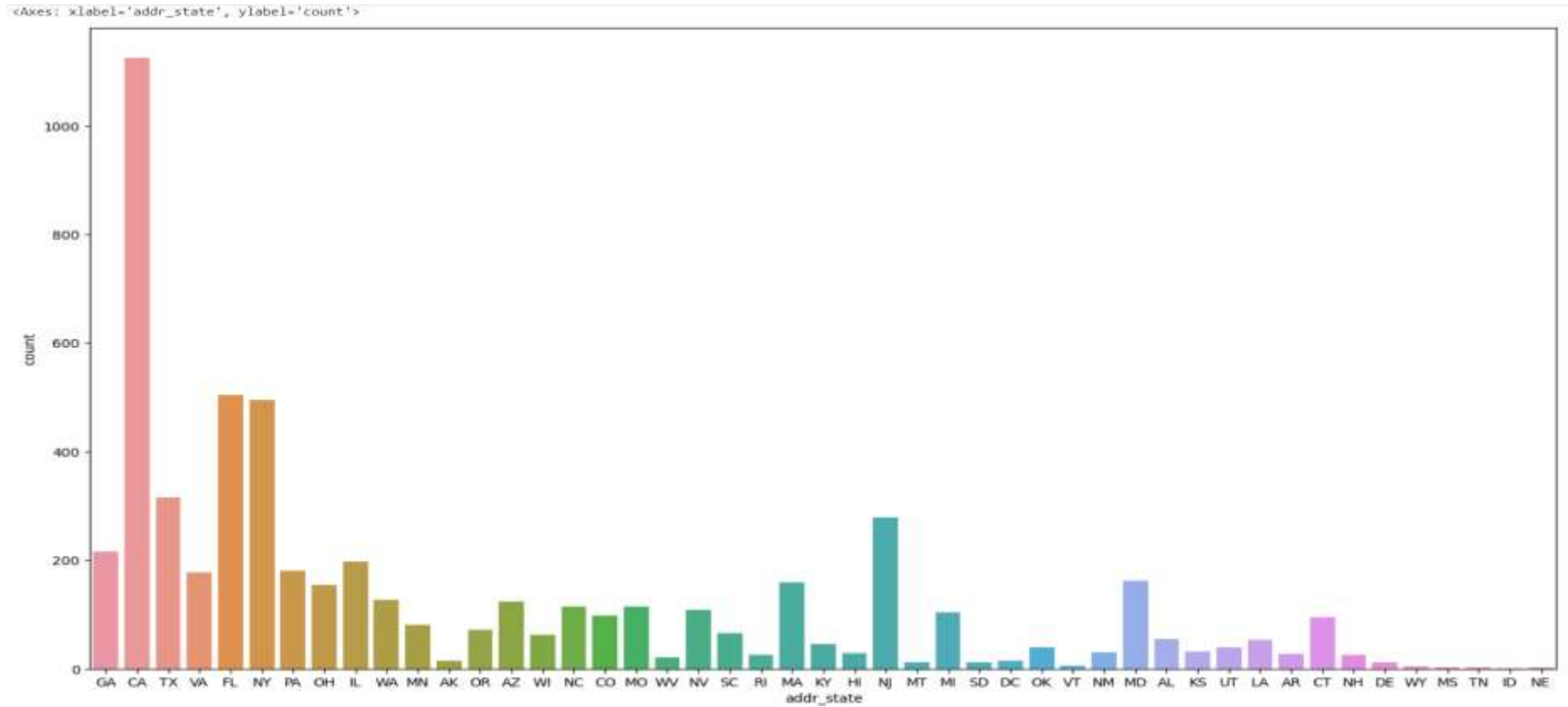
DTI

DTI: DTI range between 10-20 have the highest number of defaulters.

Interest Rate

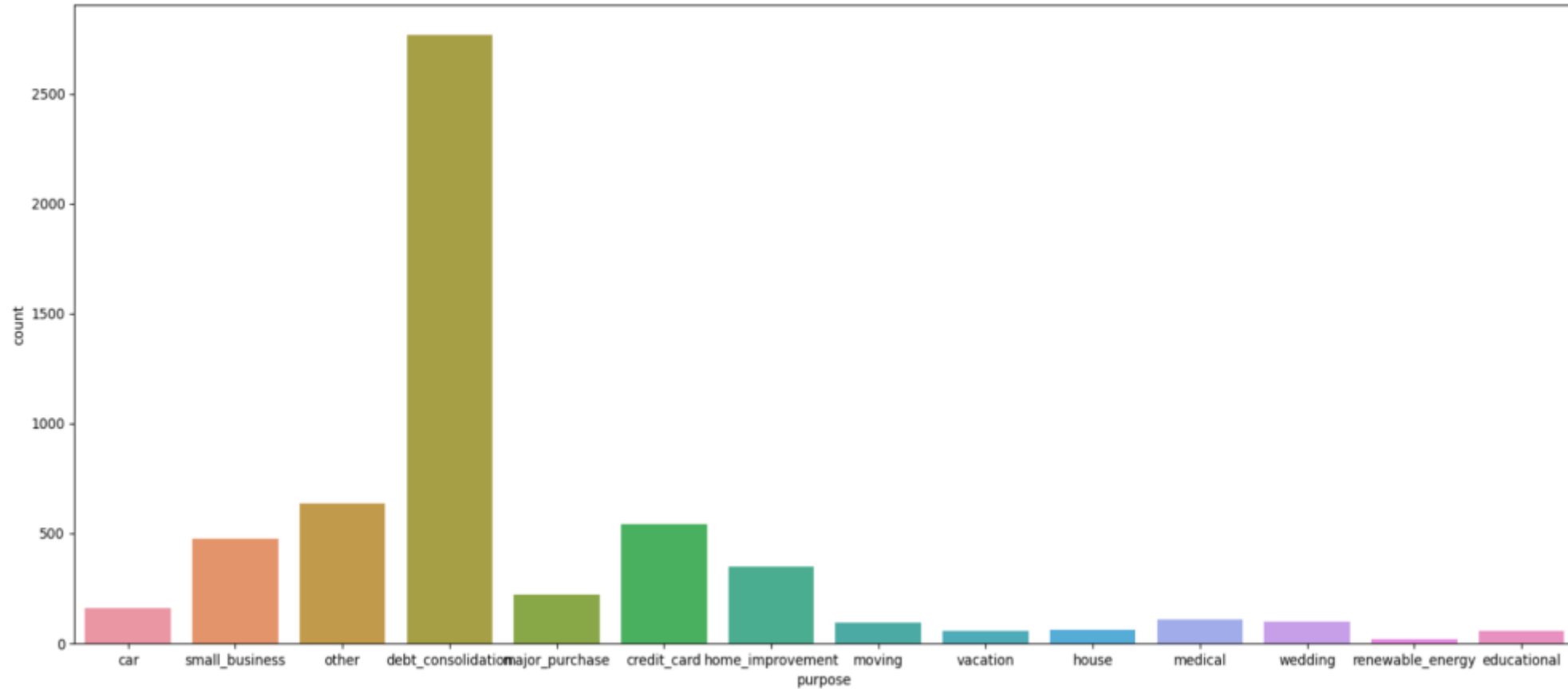**Interest Rate:** Interest Rate range between 10-15% have the highest defaulters.

Address State

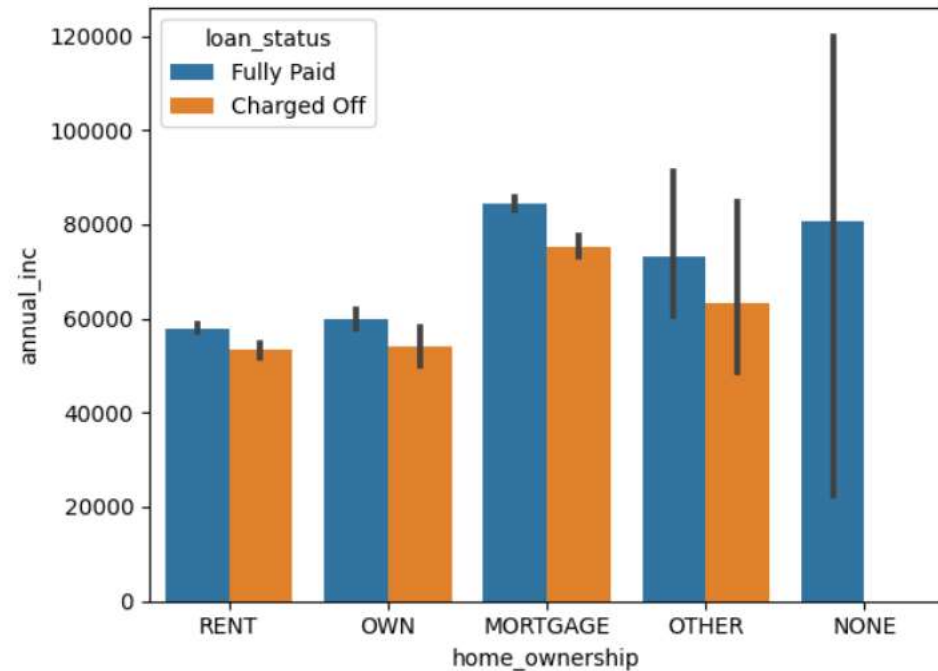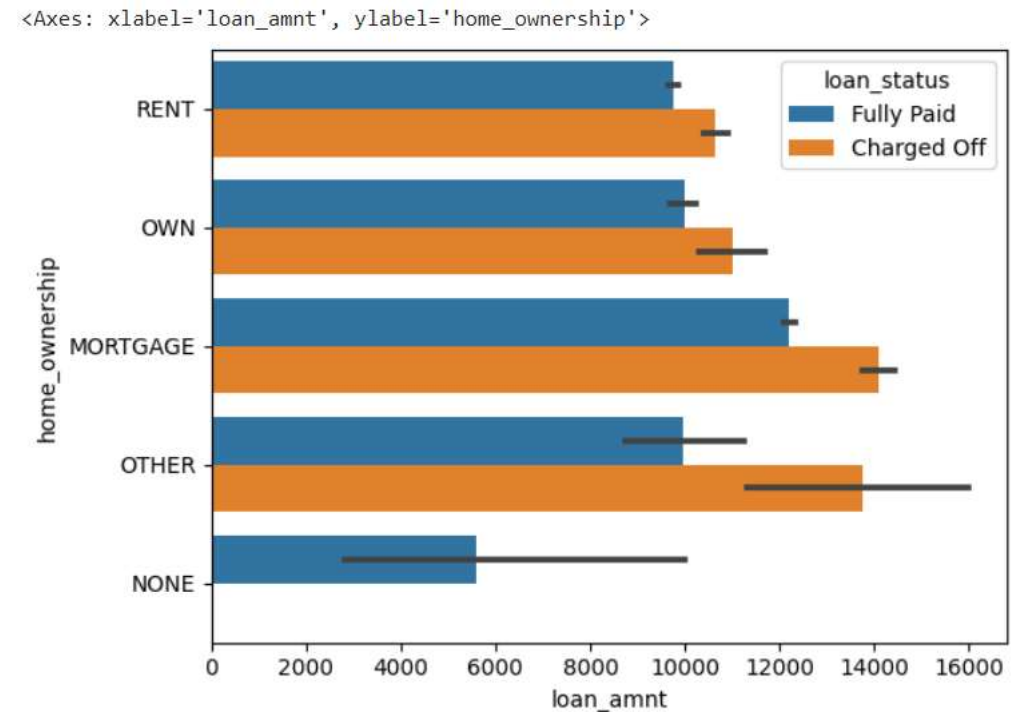**Address State :** CA/California have the highest number of defaulters.

Purpose

`<Axes: xlabel='purpose', ylabel='count'>`

Purpose : Loan taken with the purpose of 'Debt consolidation' have the highest number of defaulters.
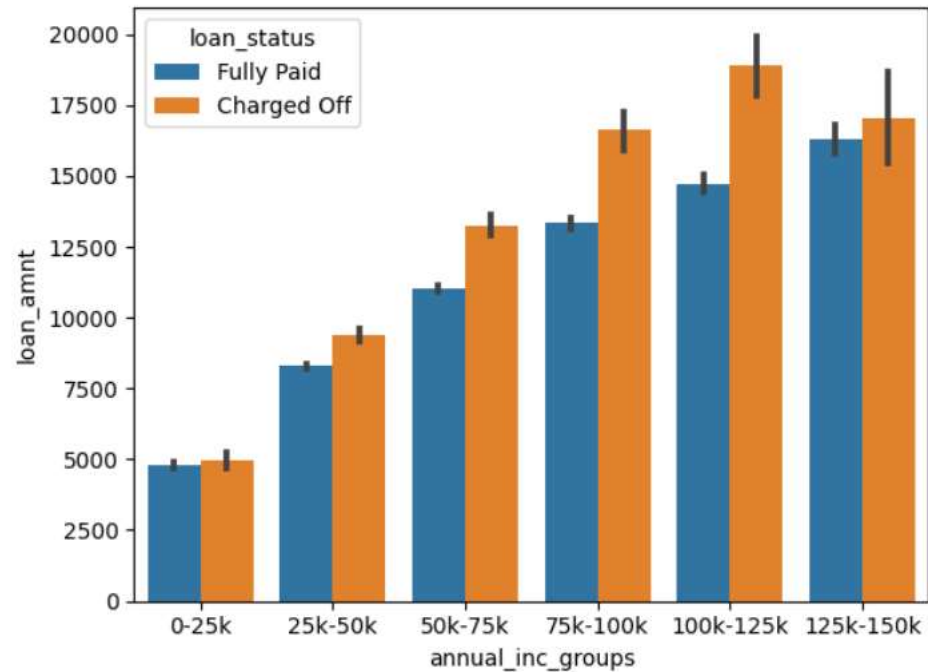
Home Ownership Vs Annual Income

Home Ownership Vs Loan Amount

Home Ownership is 'Mortgage' & Annual Income between 60-70k and Loan Amount between 12k-14k is showing the highest number of defaulters.
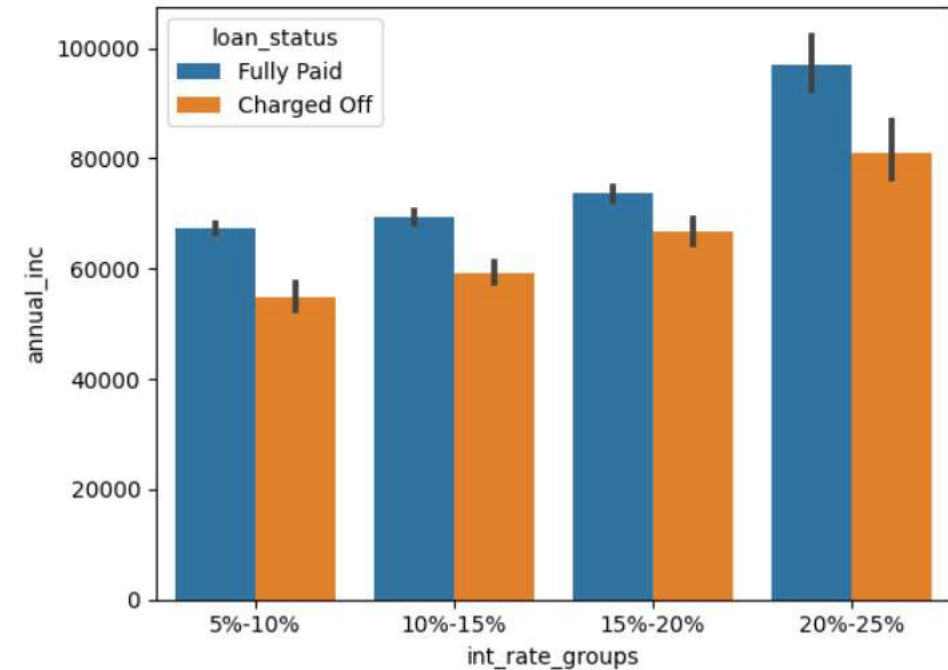
Loan Amount Vs Annual Income
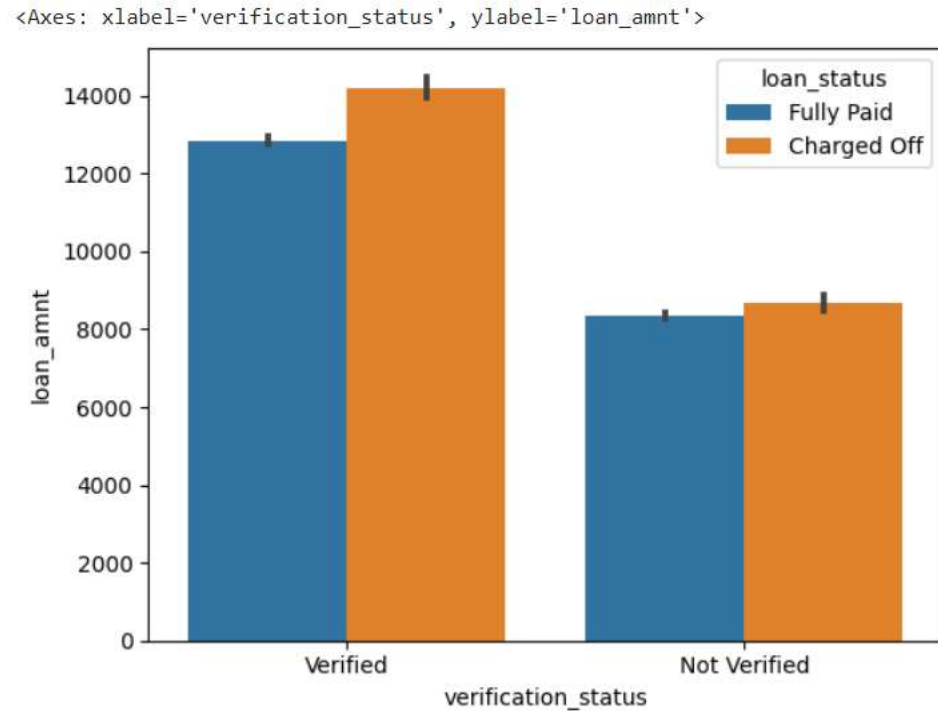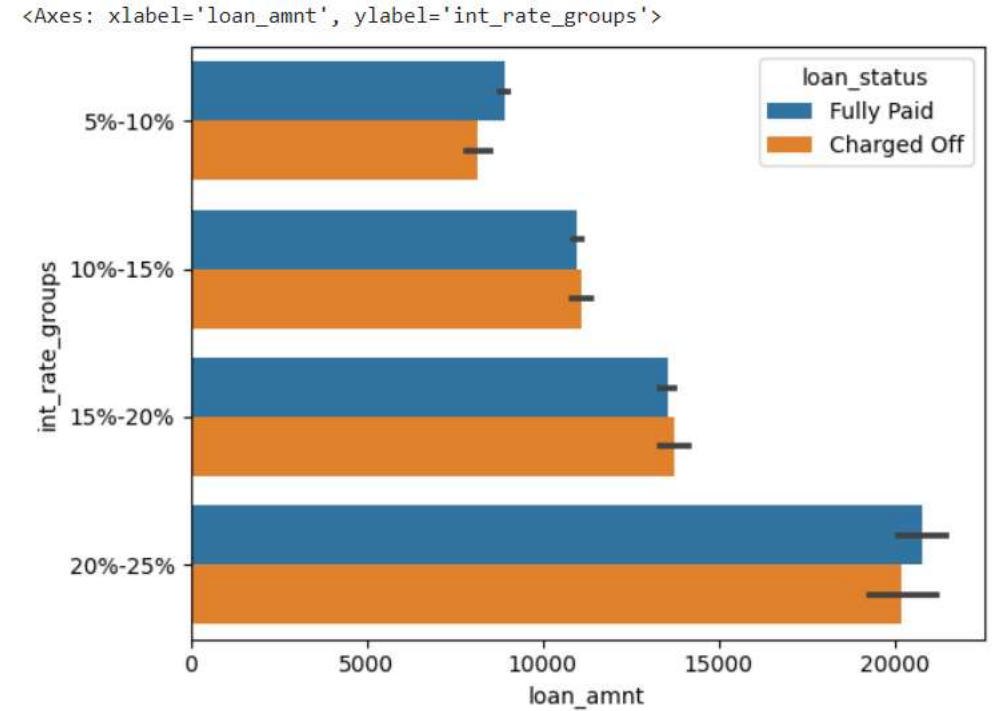
Interest Rate Vs Annual Income

Annual income between 100k-125k having **Loan Amount** >150k has higher defaulters.
Annual income between 60k-80k having **Interest Rate** between 20-25% has higher defaulters.

Loan Amount Vs Verification Status
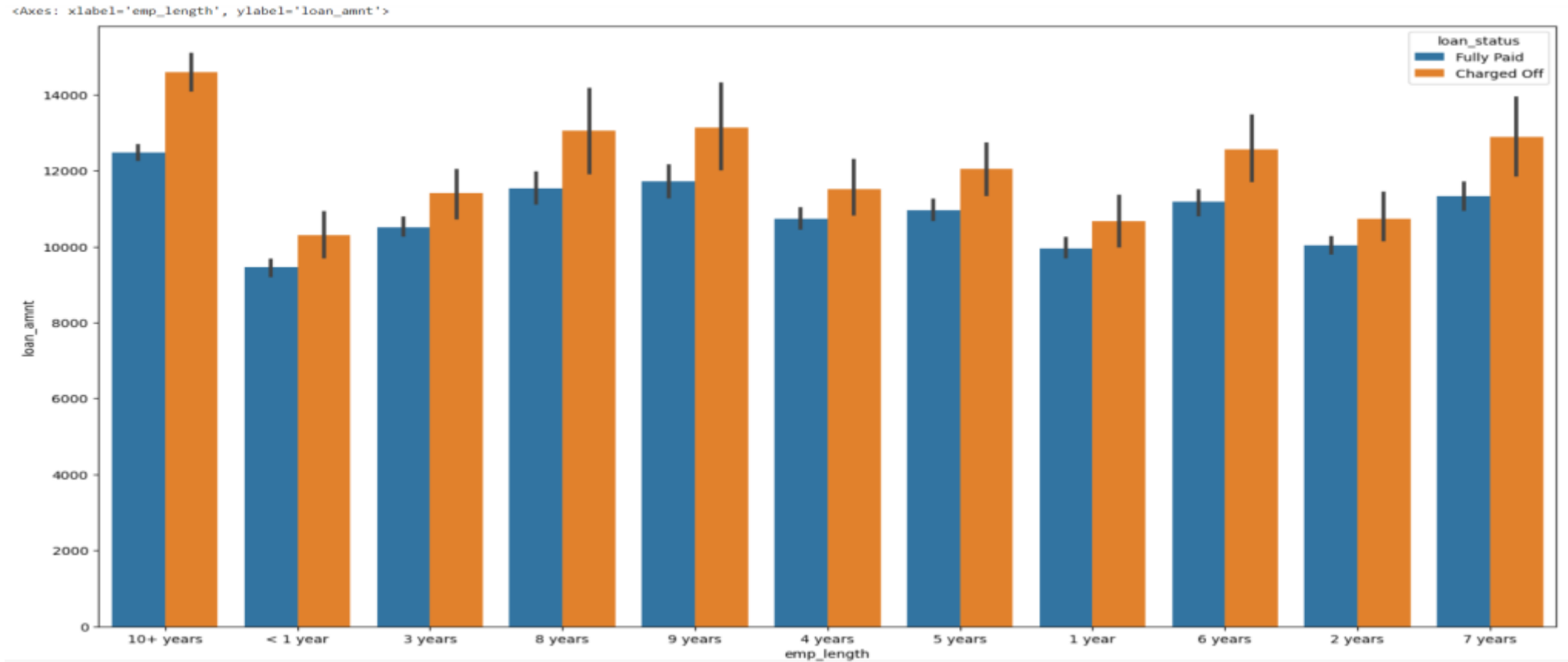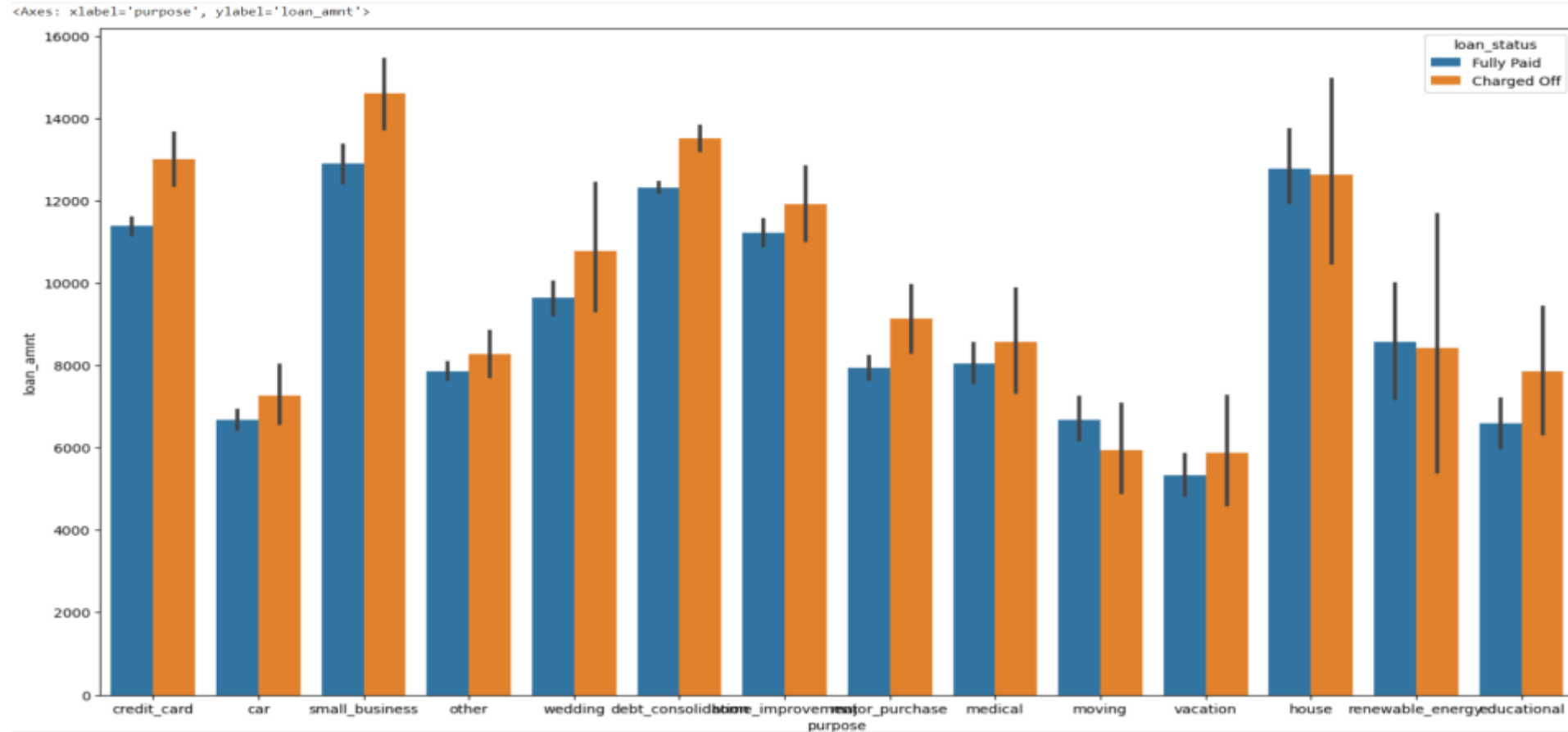
Interest Rate Vs Loan Amount

Loan Amount between 12k-14k having Verification Status as 'Verified' has higher defaulters.

Loan Amount Vs Employment Length

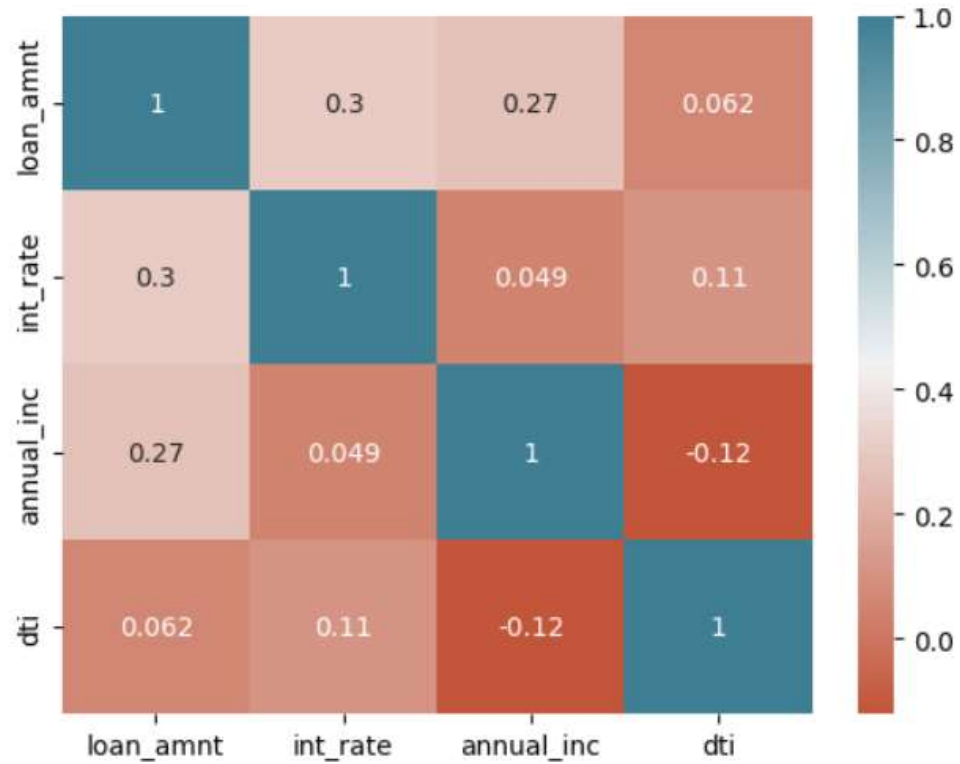Loan Amount between 120k-150k having Employment Length 10+ years has higher defaulters.

Loan Amount Vs Purpose

Loan Amount between 13k-15k taken with the **Purpose** of 'Small Business' has higher defaulters.

Correlation

Loan Amount, Interest Rate, Annual Income & DTI are showing a strong correlation implying that if all the components are increased it will lead to more 'Charged Off' scenarios.

Observations

- Most defaulters are having 'RENT' followed by 'MORTGAGE' as **Home Ownership**
- **Employment Lenth** 10+ Years have the most no. of defaulters
- Most no. of defaulters have **Verification Status as** 'Verified'
- Most no. of defaulters have **Interest Rate** between 10-20%
- Most no. of defaulters have **Annual Income** range between than 25k-50k
- Most no. of defaulters have **Address State** as California
- Loan taken with the **Purpose** of 'debt consolidation' have the most defaulters.
- Most no. of defaulters have **DTI** between 10-20.

Other Possibilities of Default

- **Home Ownership** = 'Mortgage' & **Annual Income** between 60-70k and **Loan Amount** between 12k-14k has highest defaulters.
- **Annual income** between 100k-125k having **Loan Amount** >150k has higher defaulters.
- **Annual income** between 60k-80k having **Interest Rate** between 20-25% has higher defaulters.
- **Loan Amount** between 12k-14k having **Employment Length** 10+ years has higher defaulters.
- **Loan Amount** between 13k-15k taken with the **Purpose** of 'Small Business' has higher defaulters.
- **Loan Amount** between 12k-14k having **Verification Status** as 'Verified' has higher defaulters.