Lending Club Case Study – Exploratory Data Analysis (EDA) SUBMISSION

Group Members:

➢ Anurag Bombarde

➢ Ameet Topare

Business Understanding

A **consumer finance company** receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two **types of risks** are associated with the bank's decision:

- 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

Business Objectives

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default

Loan Rejected
(Not considered in dataset)
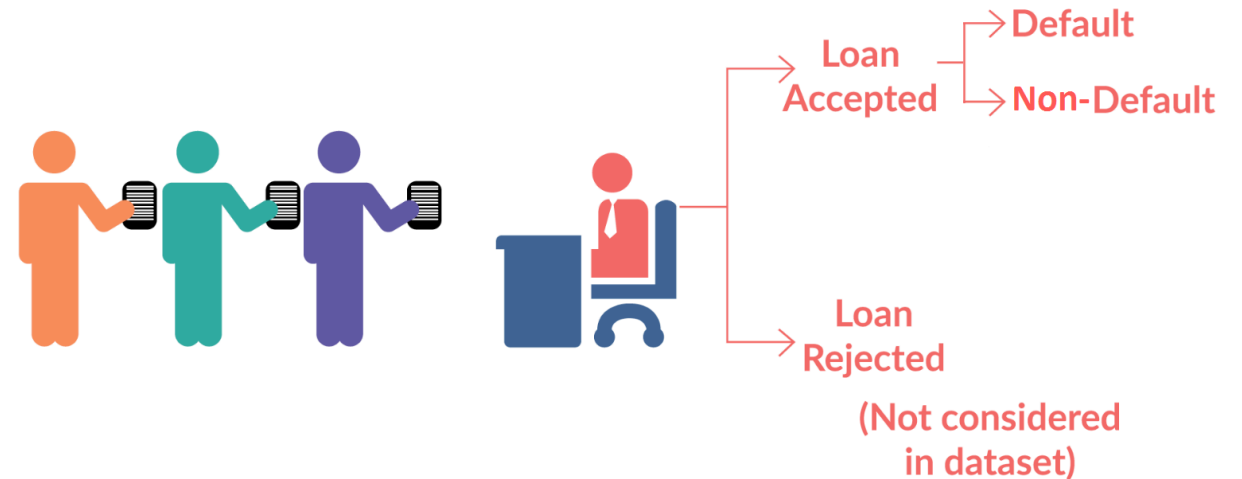
The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Analysis Methodology - EDA

4
Bivariate
Analysis

5
Conclusion

3
Univariate and
Segmented
Univariate
Analysis

2
Data
Cleaning

1
Data /
Business and
Domain
Understanding

Univariate Analysis
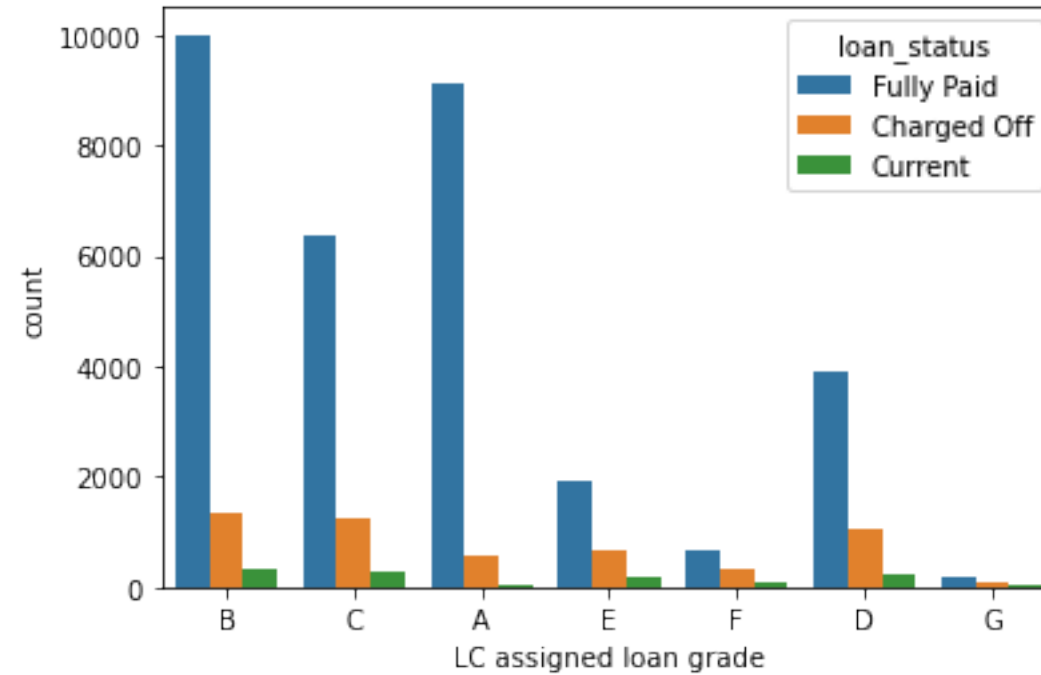
Univariate Analysis

Inference

- Less than 3 years experienced people and more than 10 Years are going for loan most frequently.

- More than 10 years exp folks are less likely to become defaulter
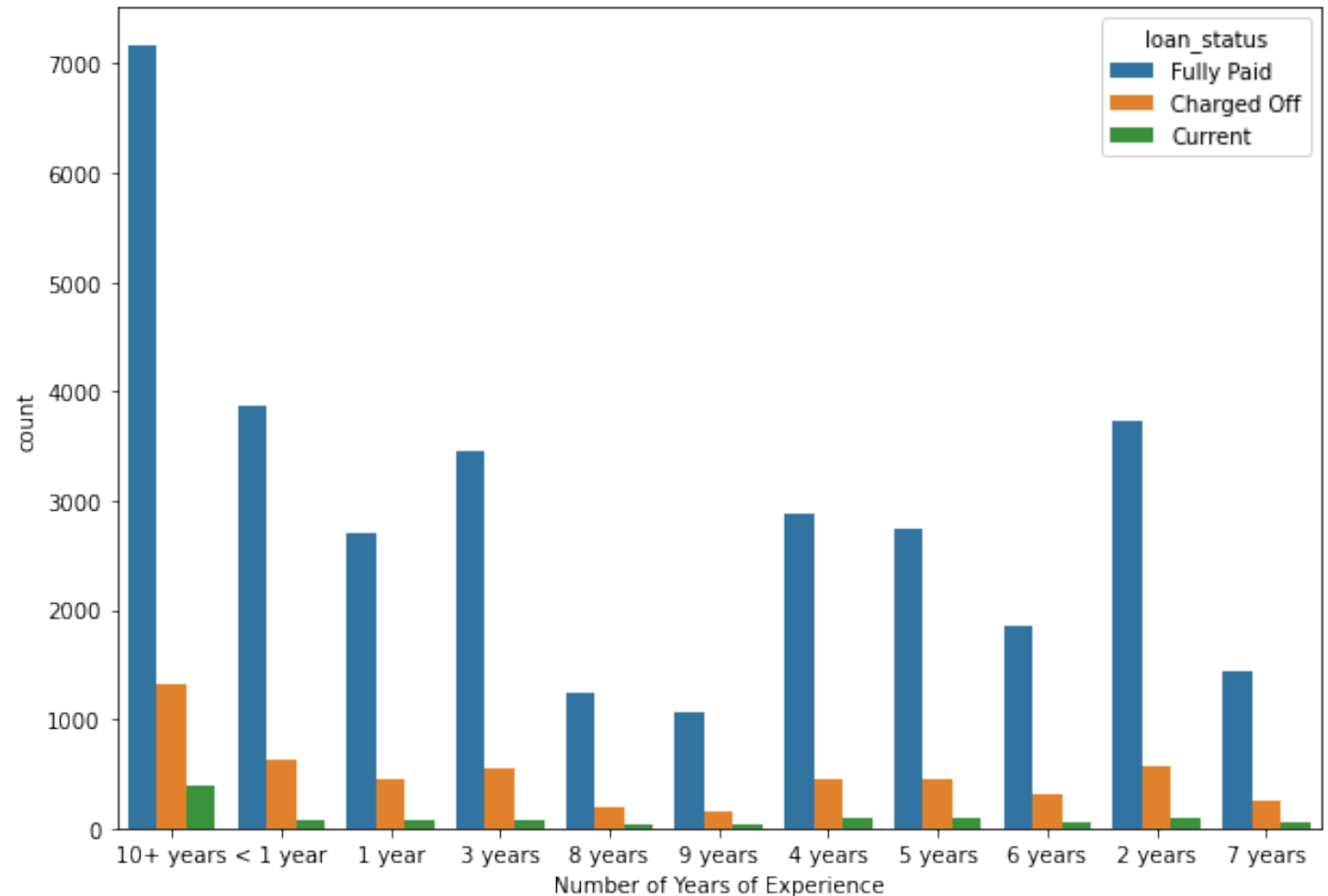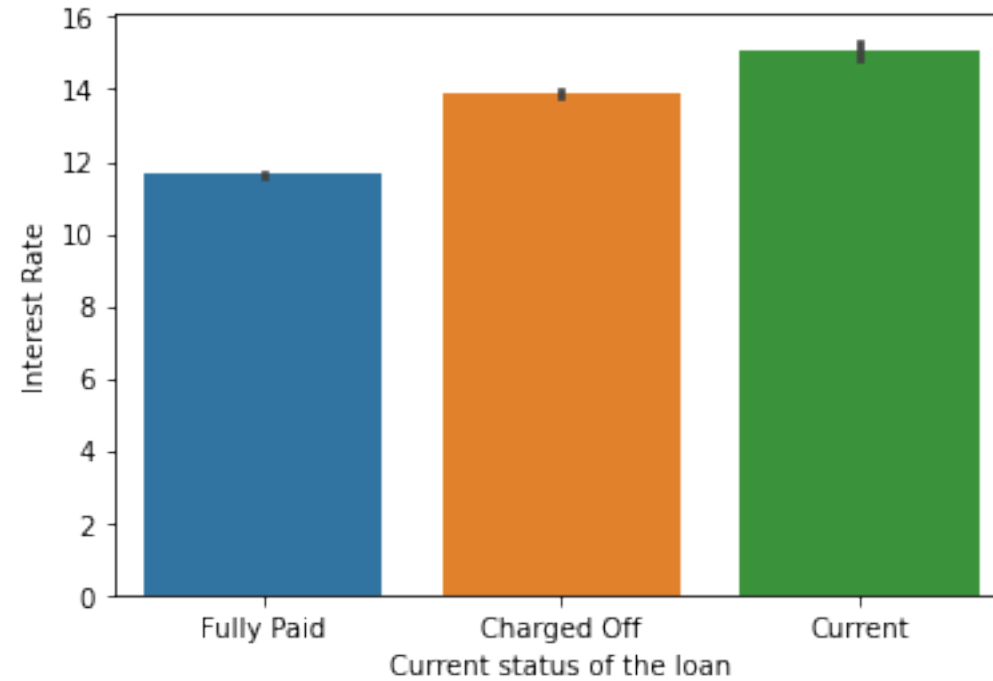
Univariate Analysis

Inference

Looking at the plot it can be concluded that LC assigned loan Grades B,C,D are having more numbers of defaulters
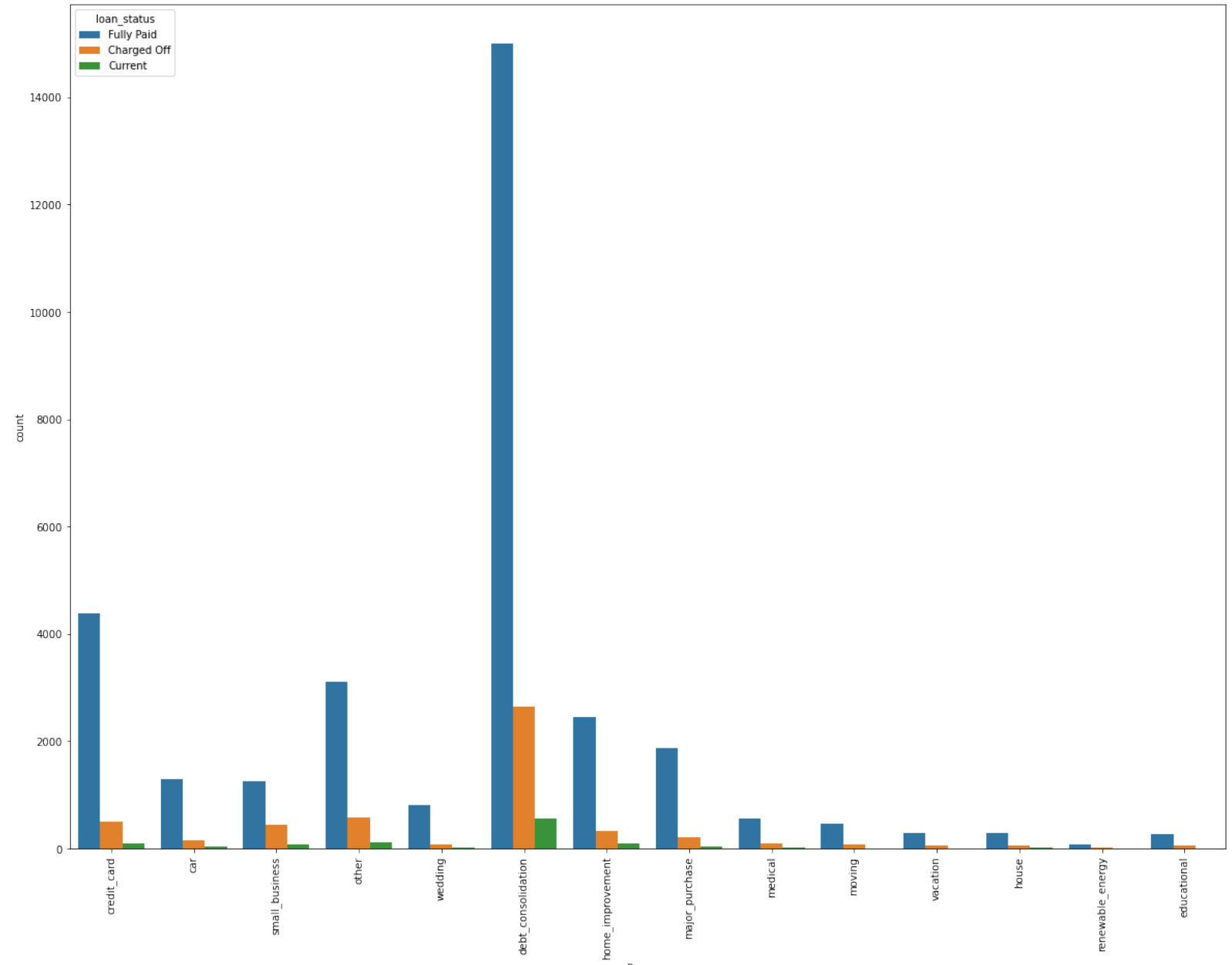
Univariate Analysis

Inference

This shows if interest rate is more than 12% then its most likely that number of defaulters will increase

Univariate Analysis

Inference-

Loan for "Debt consolidation" purpose is given frequently and fully paid as well
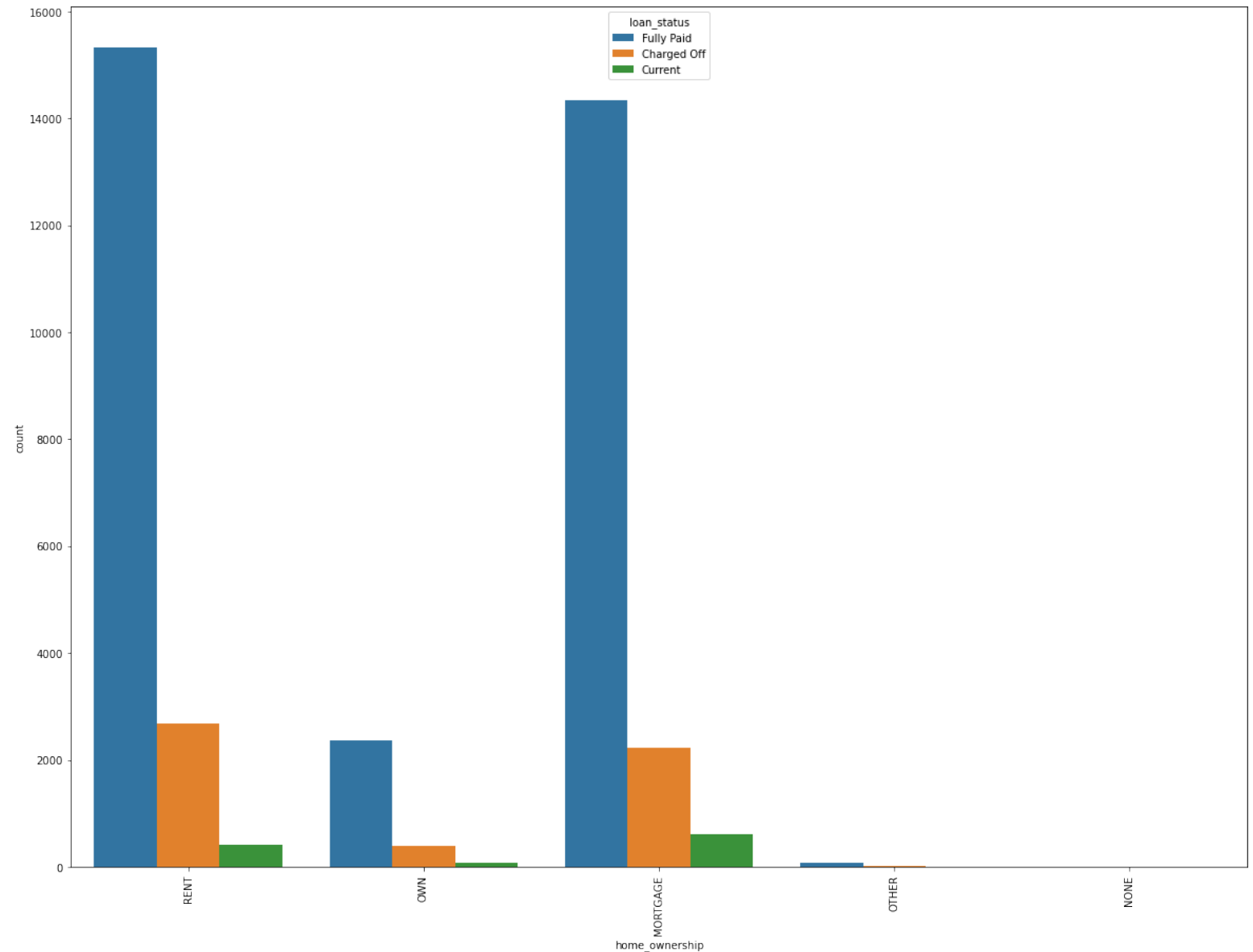
Segmented Univariate Analysis

Segmented Univariate Analysis

Inference

This shows that home ownership types RENT and MORTGAGE have high frequency of fully paid loan
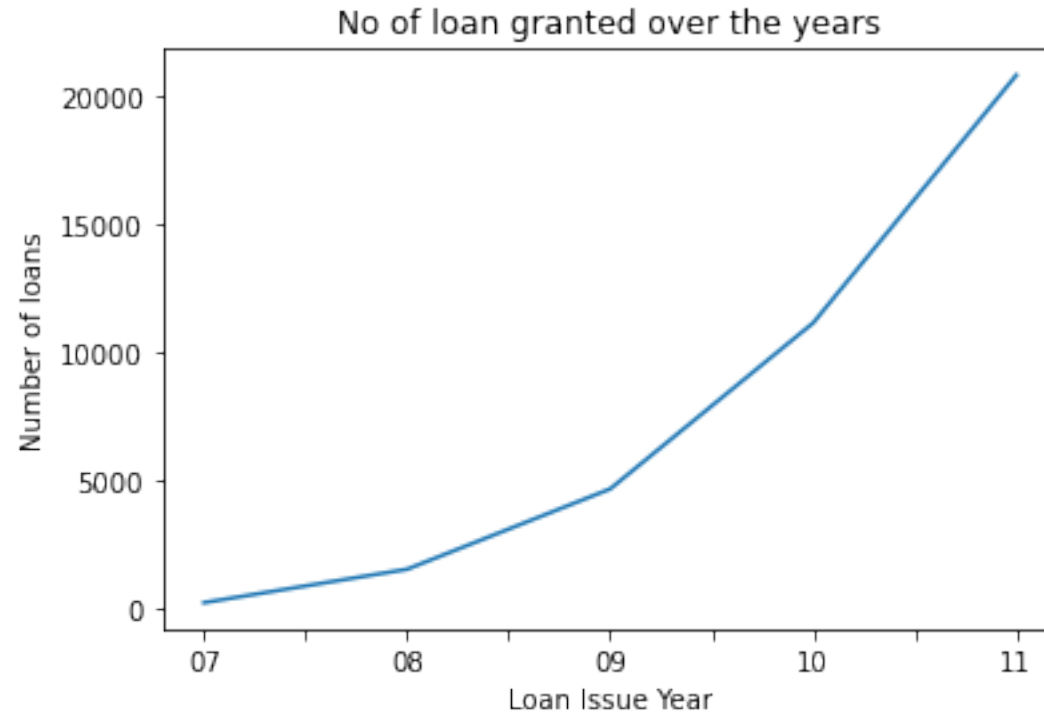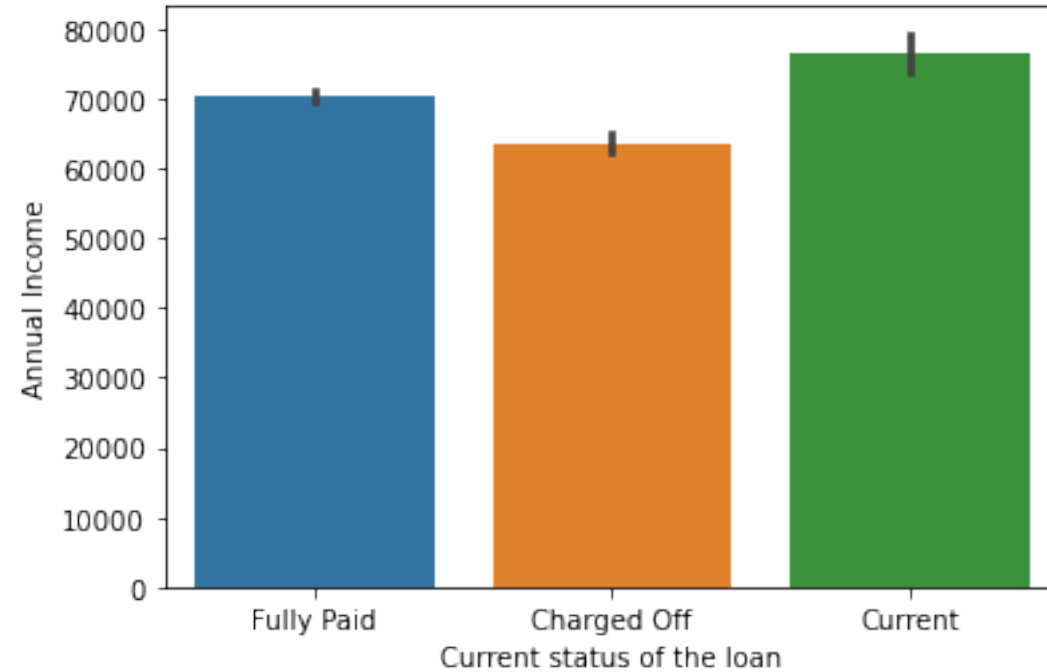
Bivariate Analysis

Bivariate Analysis

Inference-

Within period of 2007 to 2011 loan granted count is also increasing

No of loan granted over the years

Bivariate Analysis

Inference

This shows bivariate relation that, with increasing annual income fully paid frequency is increasing and defaulter rate is decreasing

Conclusion

5 important driver variables

- **Employment Length** - More than 10 years experienced people are less likely to default

- **Grade** - It can be concluded that employees with LC assigned loan Grade B,C,D are having more numbers of defaulters

- **Interest rate** - If interest rate is more than 12% then its most likely that number of defaulters will increase

- **Purpose** - Debt consolidation purpose loan is given frequently and fully paid as well

- **Home Ownership type** - Home ownership types RENT and MORTGAGE have high frequency of fully paid loan

Bivariate -
➢ From year 2007 to 2011 loan granted count is also increasing
➢ With increasing annual income fully paid frequency is increasing and defaulter rate is decreasing