Lending Club Case Study

Shubham Chauhan & Bhumika Jain

Problem Statement

Lending Club, a consumer finance marketplace specializing in offering a variety of loans to urban customers. When the 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:

1.If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

2.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

When a person applies for a loan, there are two types of decisions that could be taken by the company:

1.Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
2. **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
3. **Charged-off:** Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

2.Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)
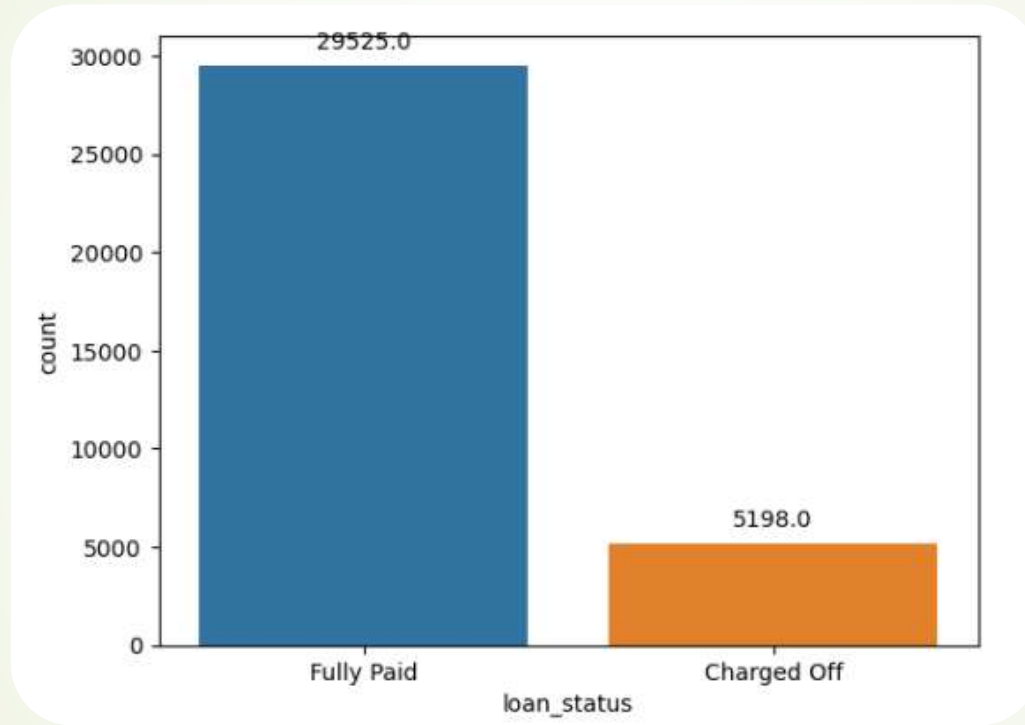
Business Objectives

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default, therefore objective is to point out applicants who have risk of defaulting on loans, enabling them to reduce business loss.

Data Cleaning & Pre-processing

Data Import to dataset

Drop null value columns

Dropping uniform value features

Removing irrelevant columns

Dropping rows where loan status is ongoing

Correcting data types and deriving new columns

Filter Data for requirement.
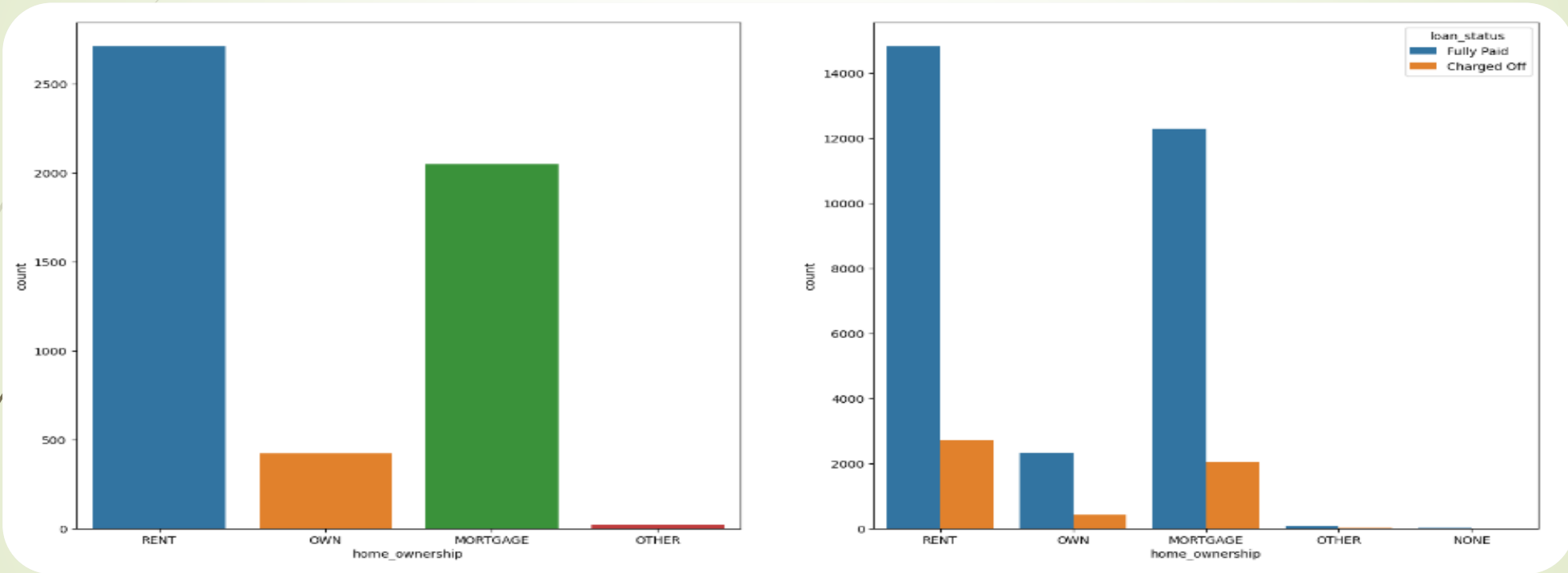
Removing outliers

Loan Status and Amount

- **Loan Status:** Defaulted loans are very low compared to Fully Paid.
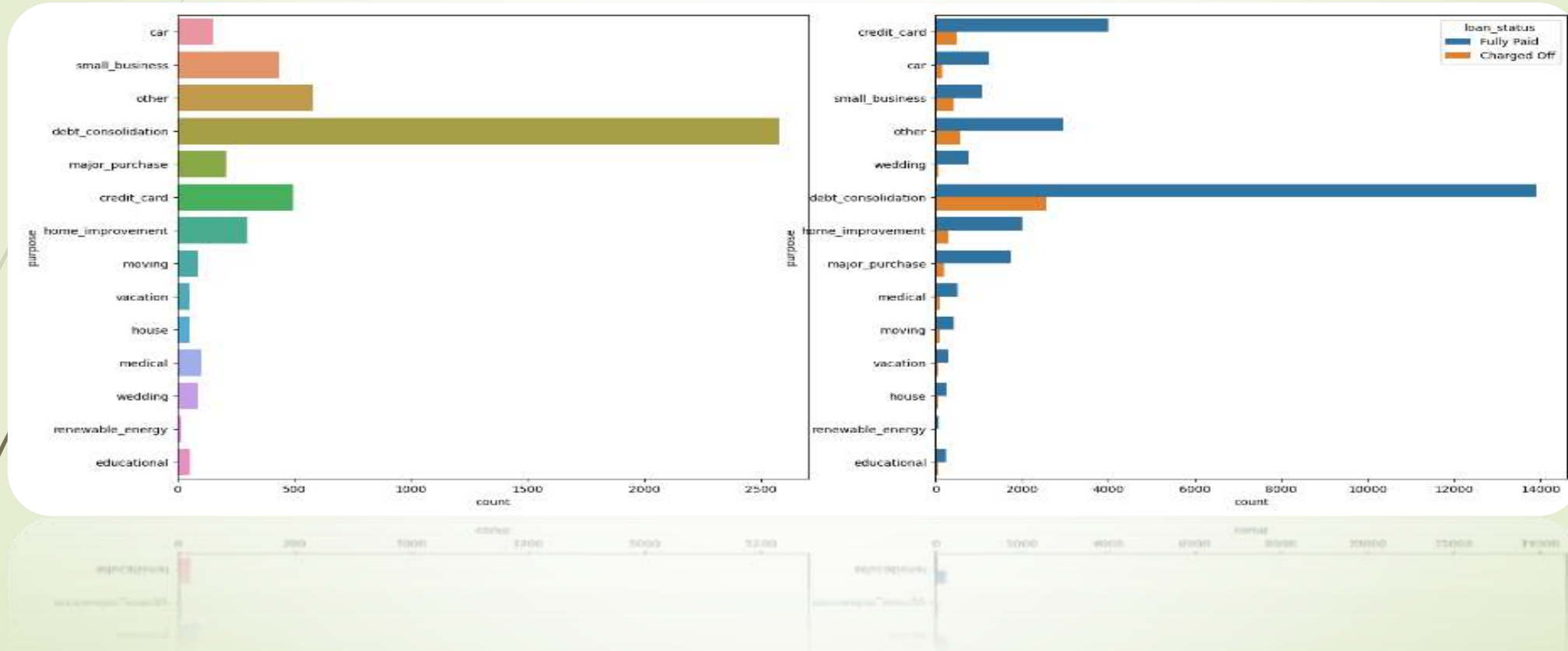
36 vs 60 Month comparison
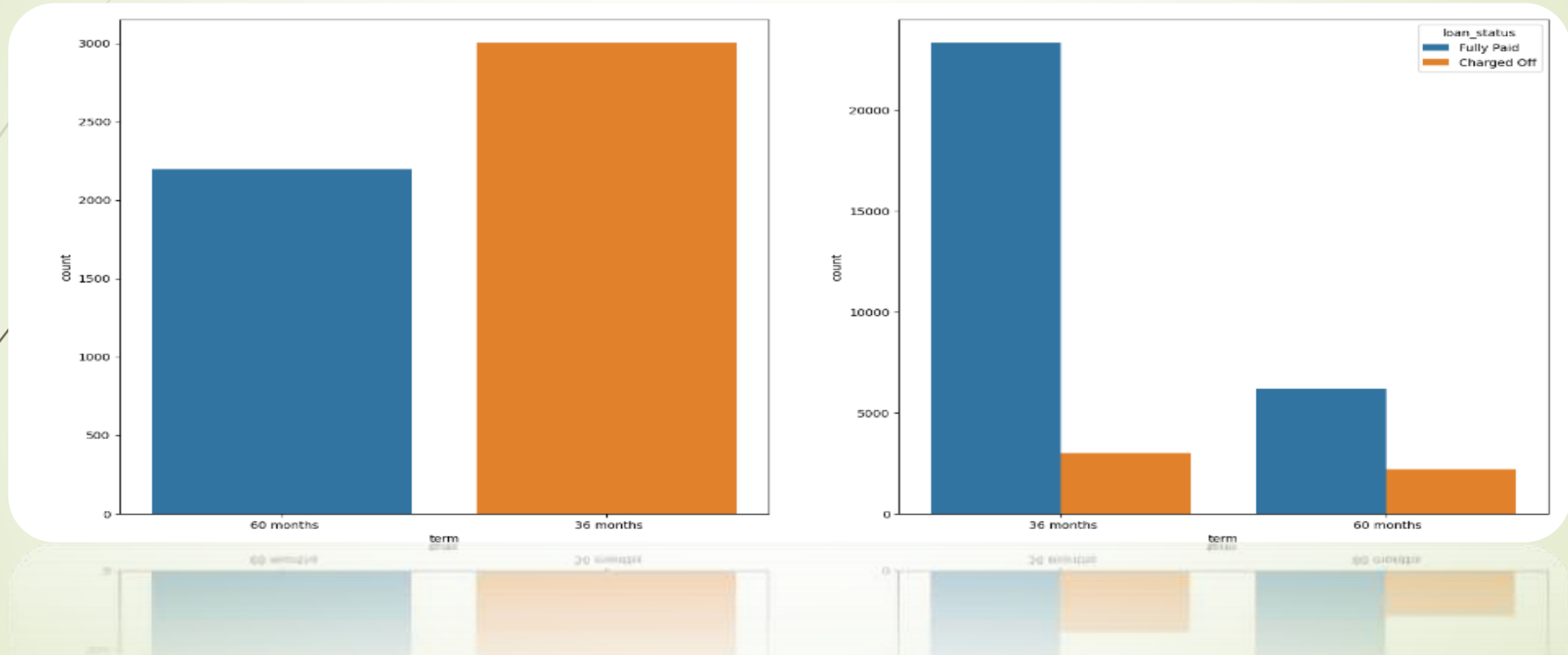
- **Observation :** short term loans (36 months) are likely to get defaulted more compared to 60 months
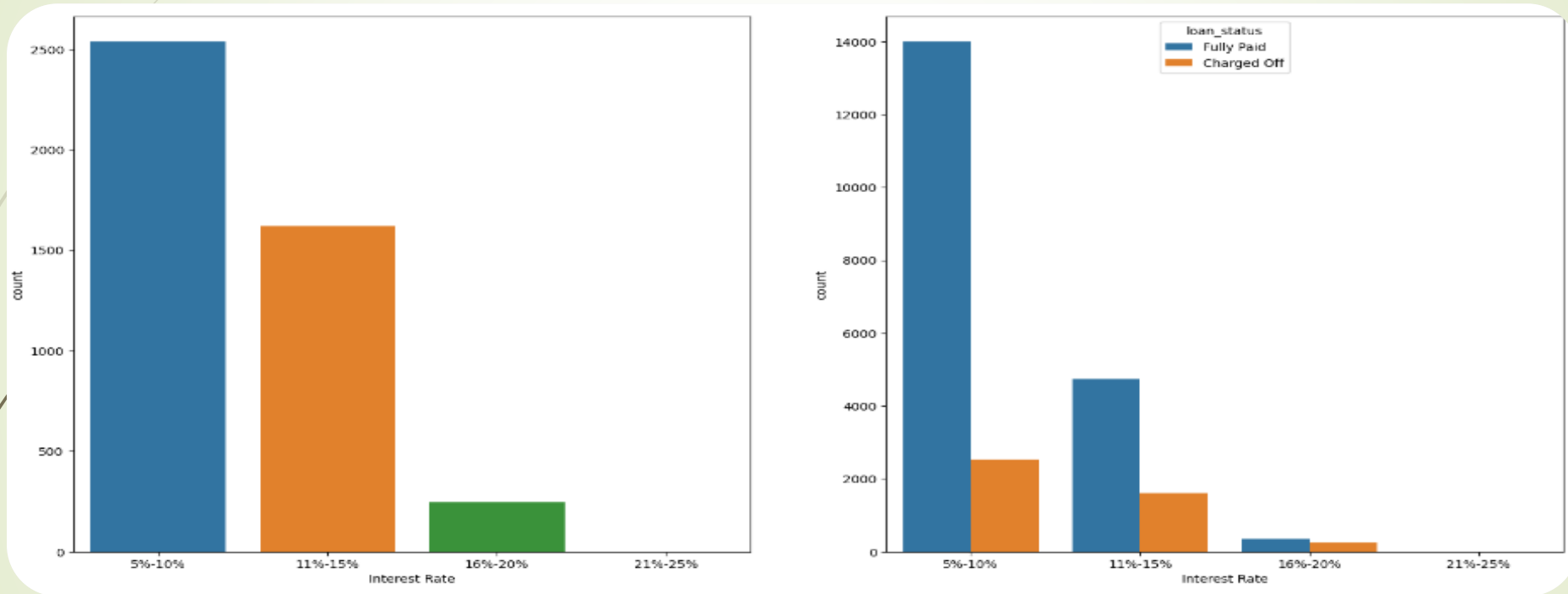
Impact of Ownership

- **Observation :** Rented candidates are more likely to default the loan compared to owned and others
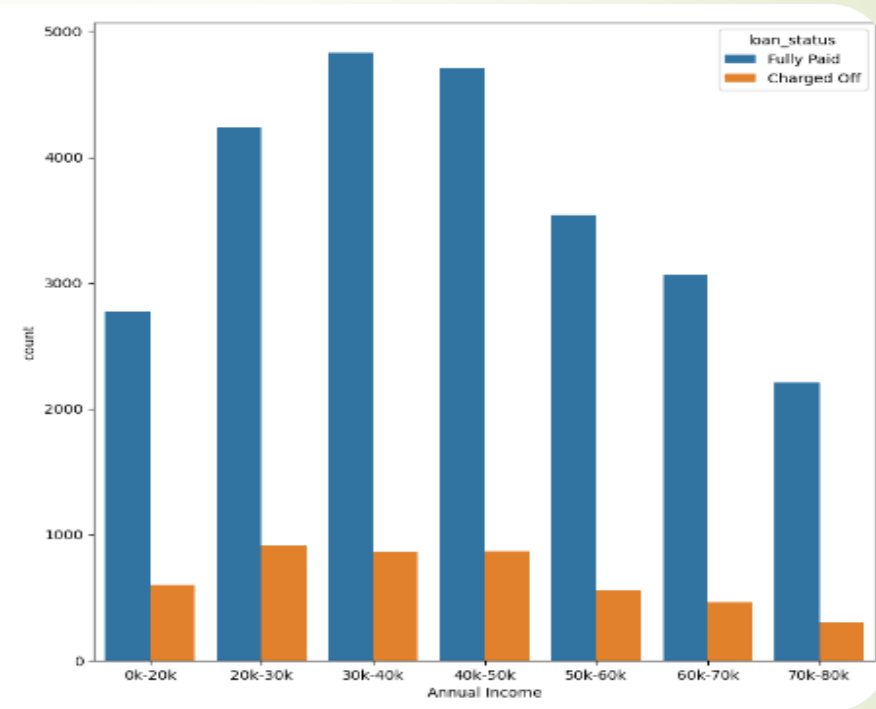
Impact of loan purpose
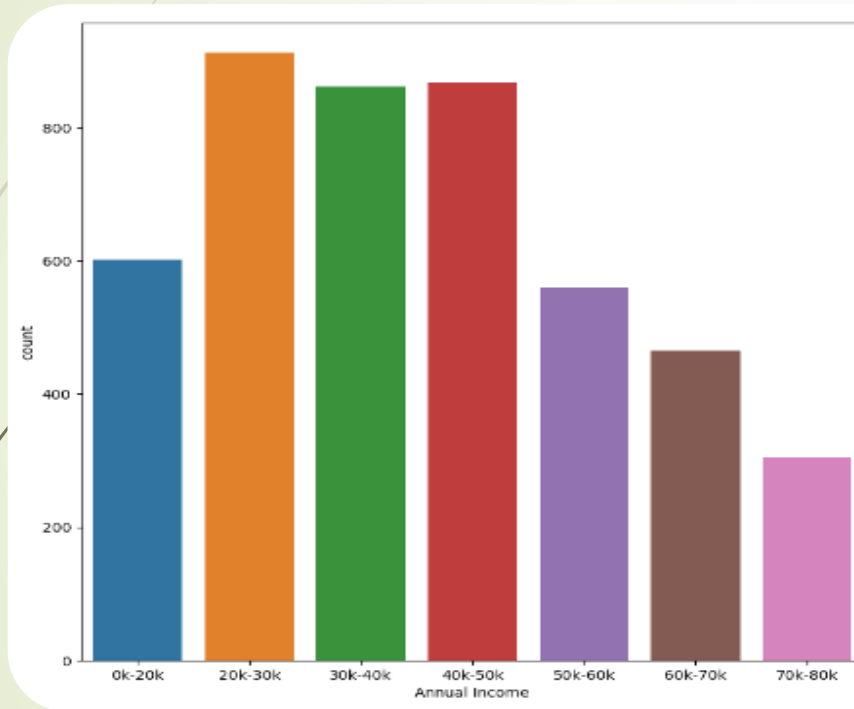

▪**Observation :** A large number of loans are taken for clearing other debts.

Impact of Interest rate

▪**Observation : fully paid** to **defaulters** candidate ratio in right figure shows that the higher the interest rate higher the chance of candidate getting defaulted

Impact of income

▪**Note:** Candidates with less 50000 annual income are more likely to default their loan than higher annual income.

DTI category analysis

Impact of dti on defaulted loans
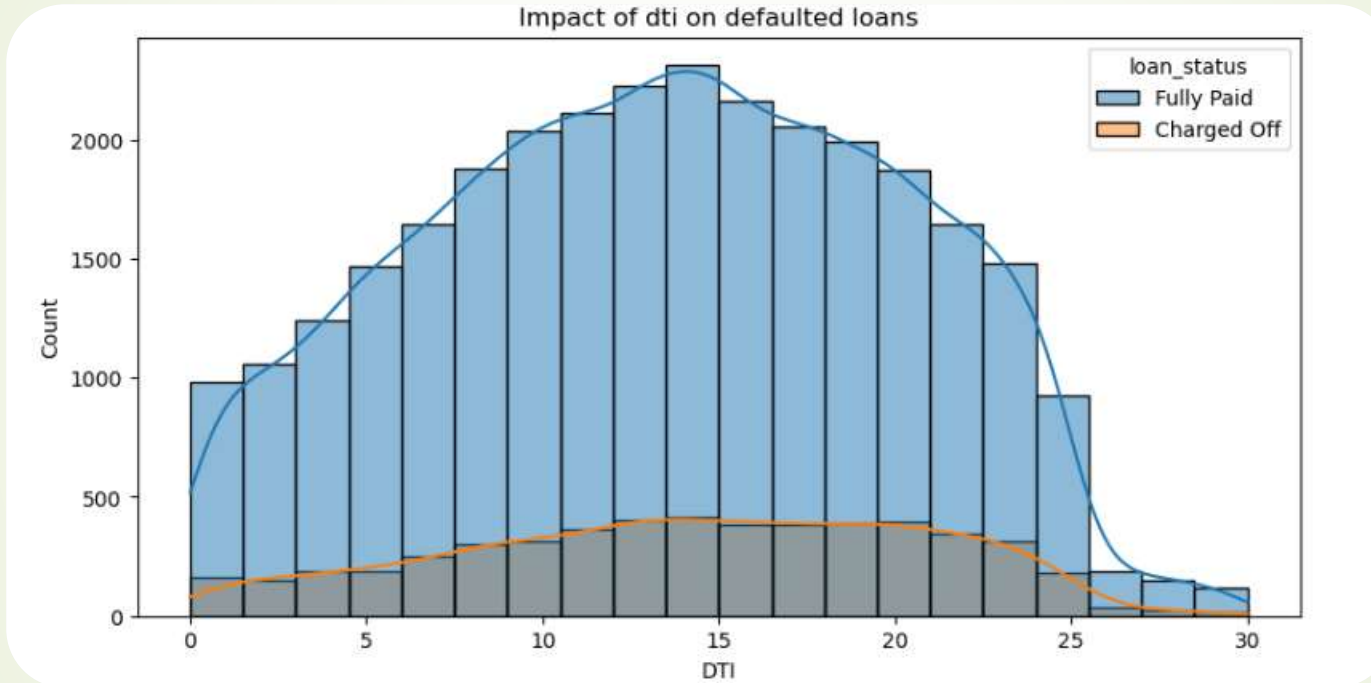
Note : slightly high in the mid (10-15 dti), doesn't give concrete observation that candidate will default if he/she has dti in between 10-15, however we could consider it as side pointer to narrow down our analysis.
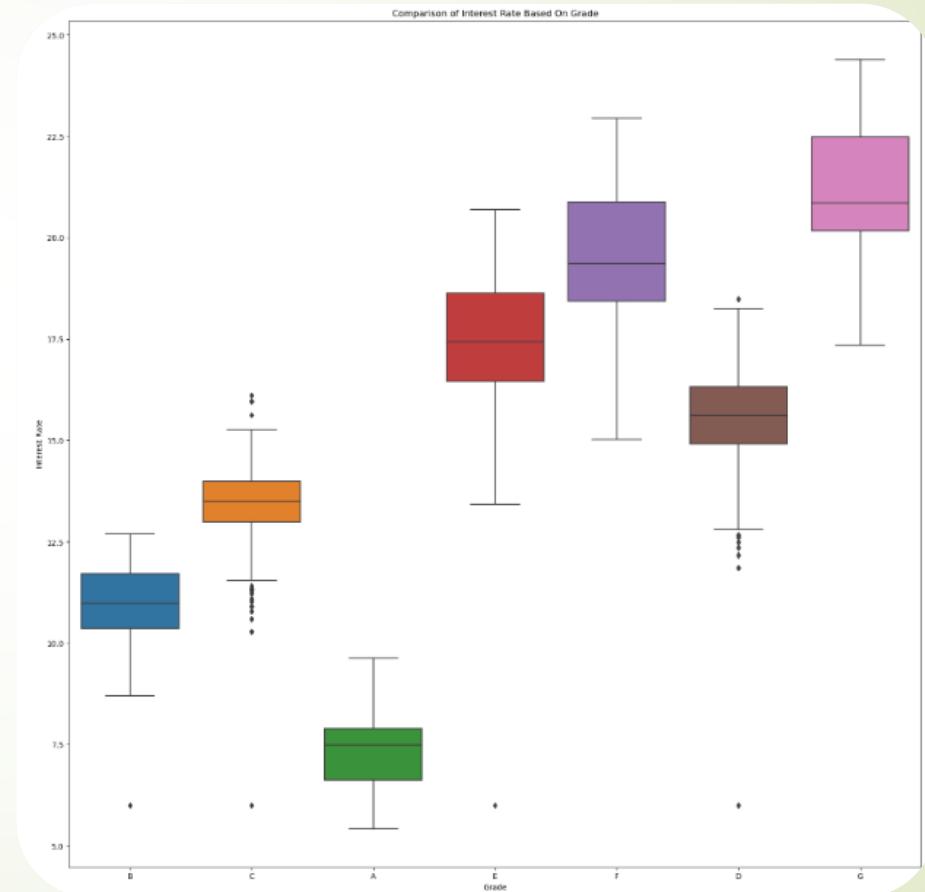
Employment Length

Observation : For the defaulted candidate, it is found that most of the candidates are of 10+ employment length.
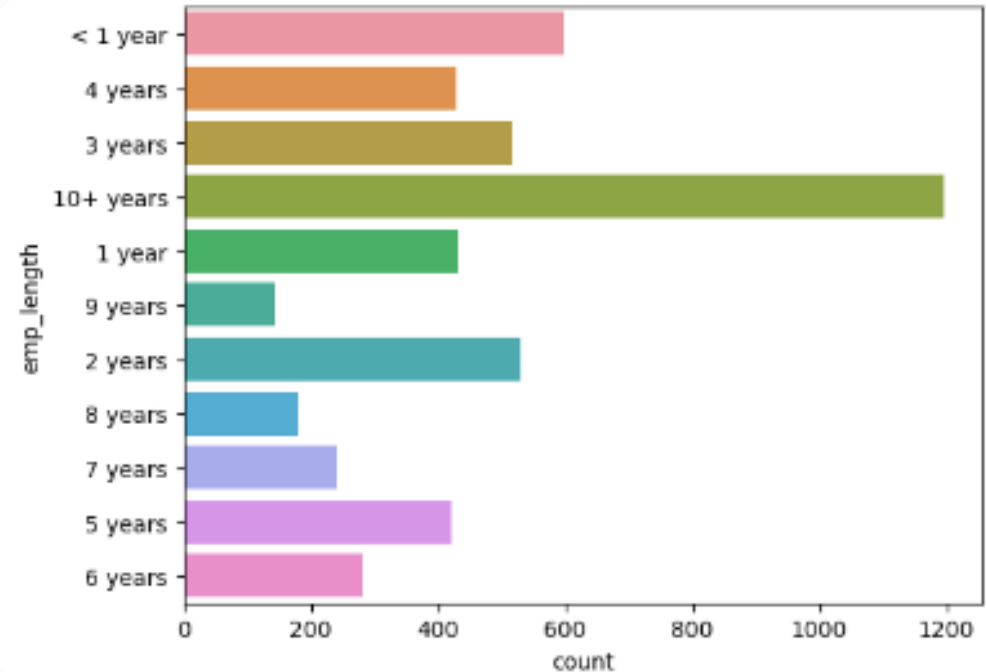
Interest Rate vs Grade Plot

- **Observation :** with increasing order of grade, we can observe that, lower the grade (considering G to be the lowest and A to be the highest grade) higher the risk of loan getting defaulted(as interest rates are getting higher and earlier univariate analysis already showed higher the interest rate, higher the chance of loan default).

Summary

Observations Comparison (36 Months - 60 Months - Fully Paid - Charged Off)
•short term loans (36 months) are likely to get defaulted more compared to 60 months.
Impact of Ownership
•Rented candidates are more likely to default the loan compared to owned and others.
Impact of loan purpose
•A large number of loans are taken for clearing other debts.
Impact of Interest rate
•"fully paid to defaulters" candidate ratio in above figure shows that the higher the interest rate higher the chance of candidate getting defaulted.
Employment Length
•For the defaulted candidate, it is found that most of the candidates are of 10+ employment length.
Interest Rate vs Grade Plot :
•with increasing order of grade, we can observe that, lower the grade (considering G to be the lowest and A to be the highest grade) higher the risk of loan getting defaulted.

Pointers
Impact of income
•Candidates with less 50000 annual income are more likely to default their loan than higher annual income.
DTI category analysis
•slightly high in the mid (10-15 dti), doesn't give concrete observation that candidate will default if he/she has dti in between 10-15, however we could consider it as side pointer to narrow down our analysis.