Lending Club Case Study - EDA

Submitted By:

- ❖ Prasad P R
- ❖ Mayuri Ladi

Introduction & Objective

* 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.

* Borrowers who **default or fail to repay the money back** cause the largest amount of loss to the lending club marketplace.

* The company wants to understand the **driving factors (or driver variables)** behind loan default.

* The main objective of this case study is to apply EDA techniques and knowledge to analyze lending club loan data, understand the driving factors which is causing more loan defaults, observe the data patterns and provide risk assessment.

Process for Exploratory Data Analysis

| Identify Business Objective | Data Sourcing | Data Cleaning | Data Understanding | EDA Analysis (Univariate, BiVariate, Multivariate) |

Observations

There are few key observations from analysis done based on columns:

- Most of loan was taken for **36 months** i.e. for short term.
- Large number of **Unverified loan data** exists which can be one of the reason for increasing number of defaulters.
- Most of the loans were issued in **December** .
- There is continuous increase in loan disbursement every year.
- Highest number of **B grade** type loans were given, followed by **A grade** .
- Highest number of loans were given to the **10+ years** employed people.
- **85.97%** of loans were fully paid whereas only **14.03%** loans were defaulted.
- **California** state has the highest number of people who have taken loan i.e. approximately about **28%**, followed by **New York** which is approx **15%**.
- **49%** loans were taken for **debt consolidation** which is highest among all other purposes, followed by **credit card** with **13%** loans.
- About **50.36%** of loans were taken by people living in **Rented** house whereas **42.11%** loans were taken by people with **mortgage** home.

Observations

- **19.45%** of loans were taken by employees working for **US Army**, followed by employees of **Bank Of America** with **17.21%**.
- Majority of the **loan amount** was in the range of **5000 to 15000**
- Majority of the **annual income** ranging in between **40k to 80K**
- **Interest rate** was ranging between **8.5 - 15%**
- **Installment** ranges between **200 to 400**
- Median loan amount provided to the defaulters is 9800, while median recovered is about 162.79, making **the percentage of unrecovered amount as 98.34%.**
- Highest recovery (median, in terms of money recovered) = **renewable energy**. Highest recovery(number of recoveries) = **debt consolidation**

Analysis
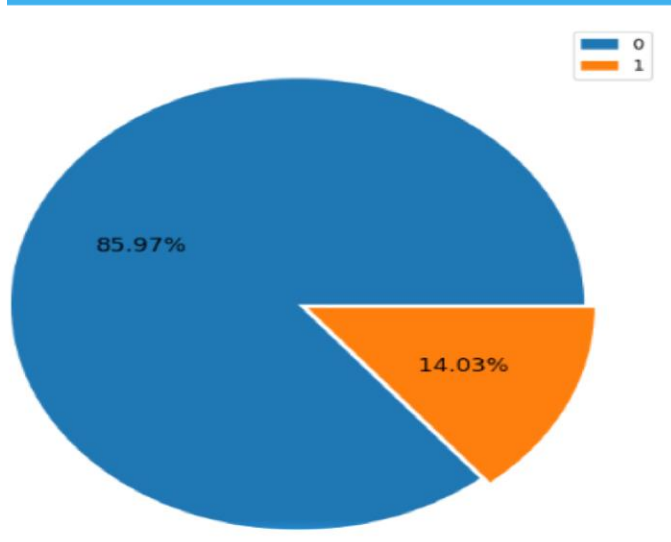
- **85.97%** of loans were fully paid whereas only **14.03%** loans were defaulted.
- **California** state has the highest number of people who have taken loan i.e. approximately about **28%**, followed by **New York** which is approx **15%**.
- **49%** loans were taken for **debt consolidation** which is highest among all other purposes, followed by **credit card** with **13%** loans.
- About **50.36%** of loans were taken by people living in **Rented** house whereas **42.11%** loans were taken by people with **mortgage** home.
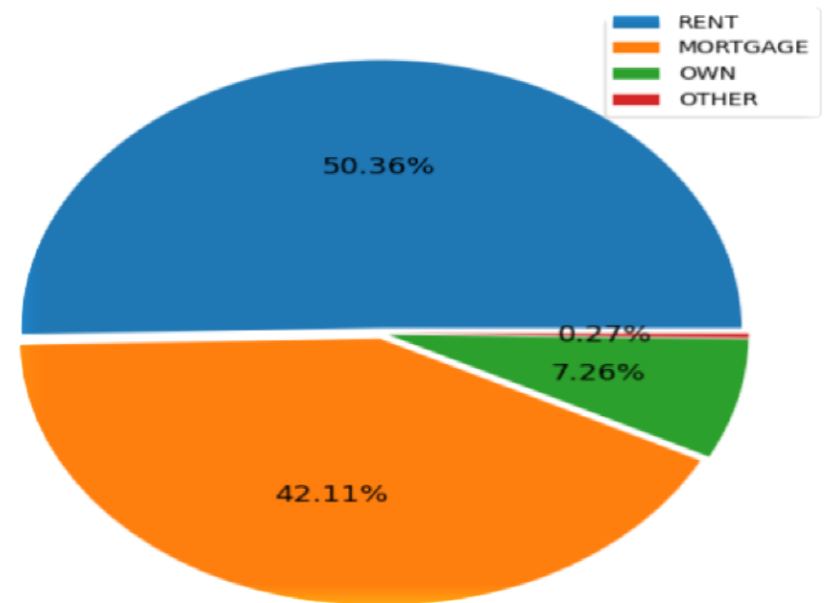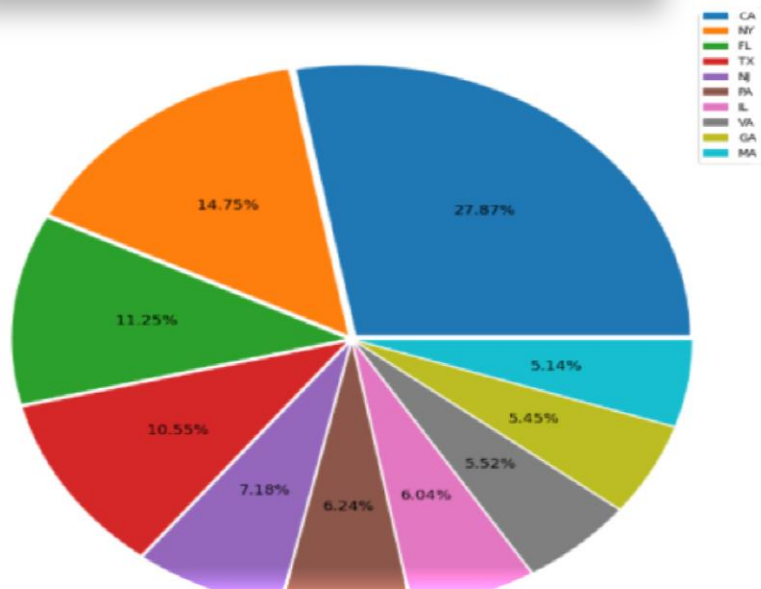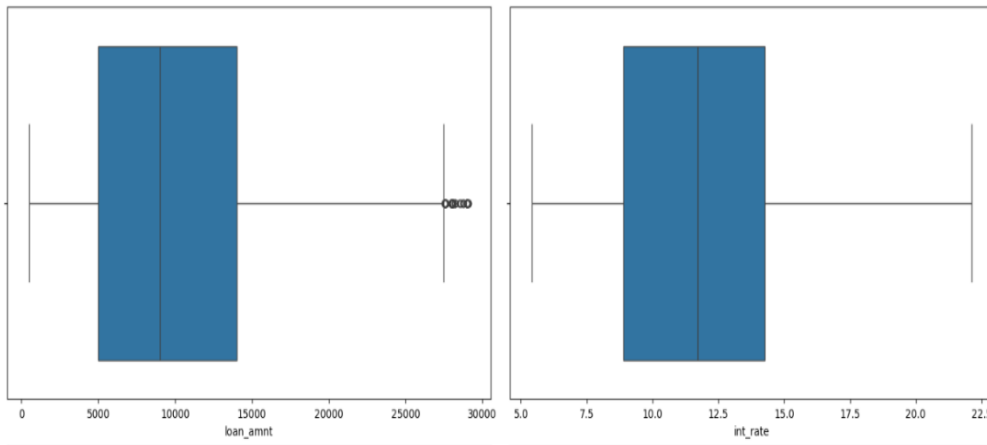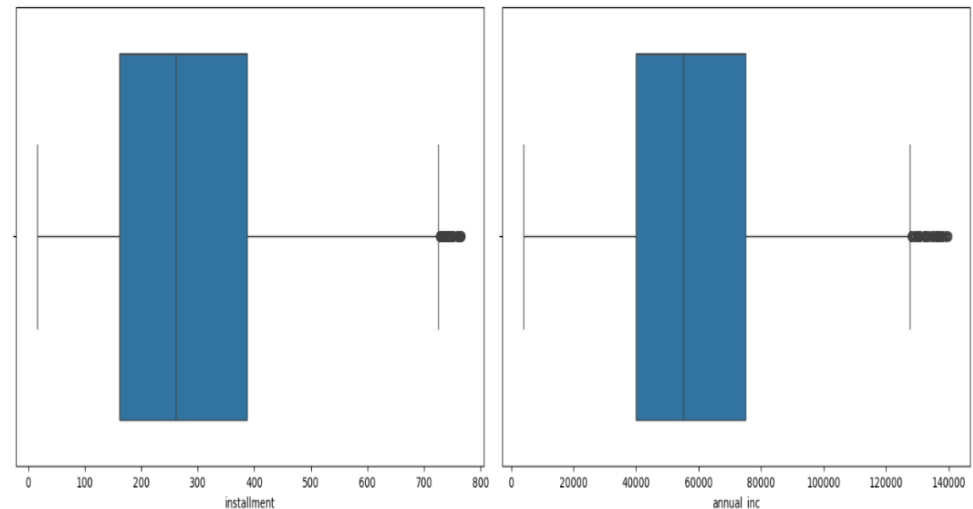
Analysis
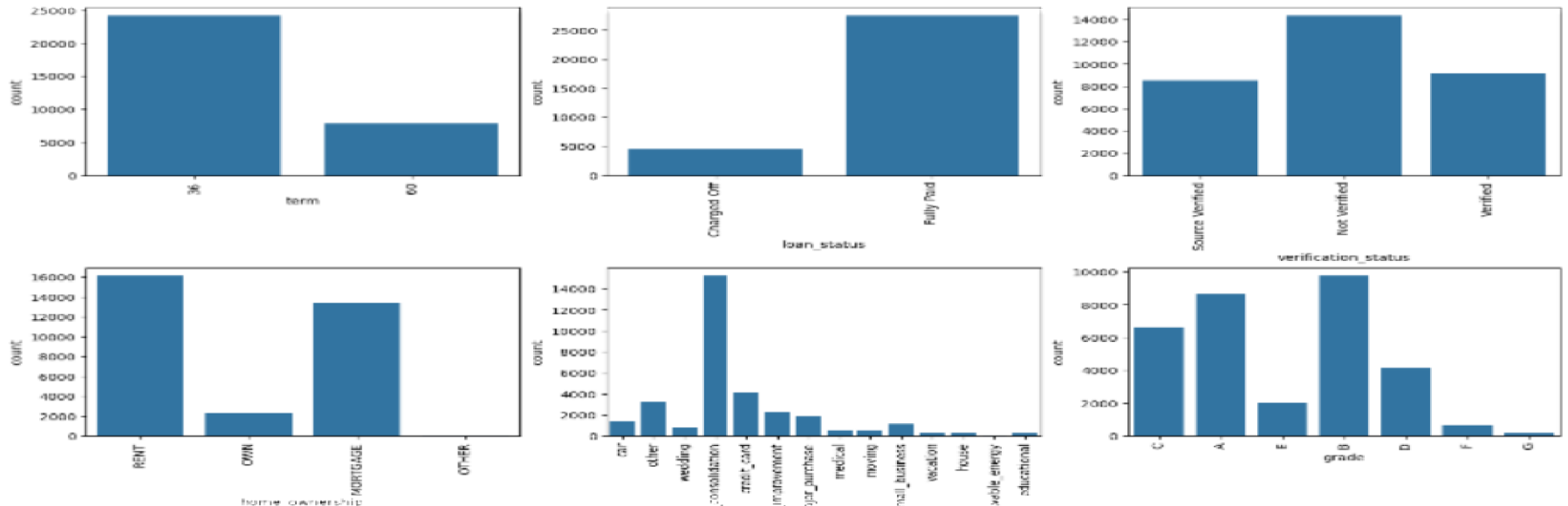
- Most of loan was taken for **36 months** i.e. for short term.
- Large number of **Unverified loan data** exists which can be one of the reason for increasing number of defaulters.
- Most of the loans were issued in **December** .
- Highest number of **B grade** type loans were given, followed by **A grade** .
- Highest number of loans were given to the **10+ years** employed people.
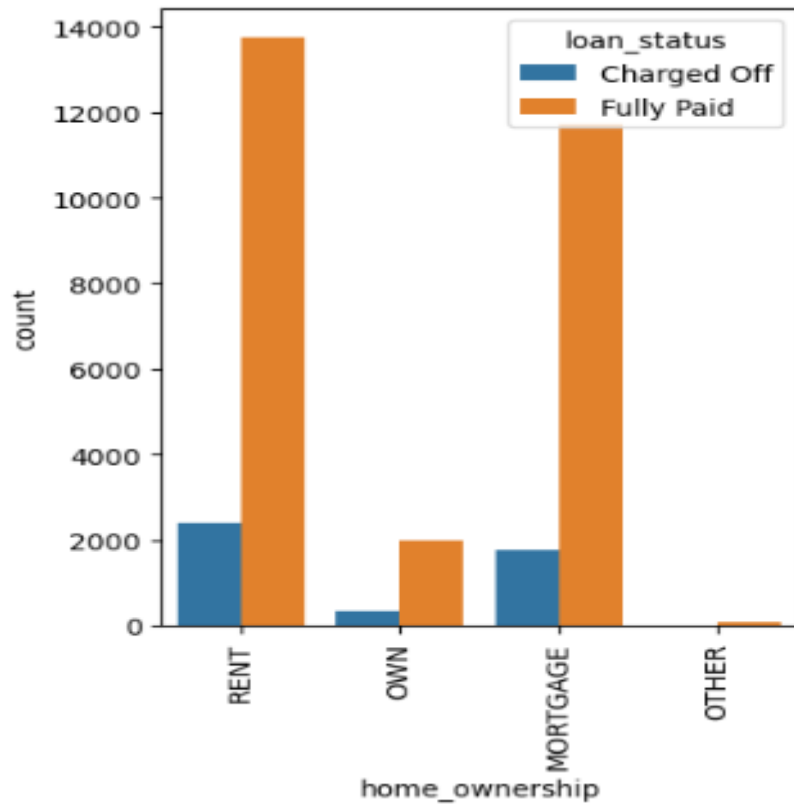
Analysis

- Majority of the **loan amount** was in the range of **5000 to 15000**
- Majority of the **annual income** ranging in between **40k to 80K**

- **Interest rate** was ranging between **8.5 - 15%**
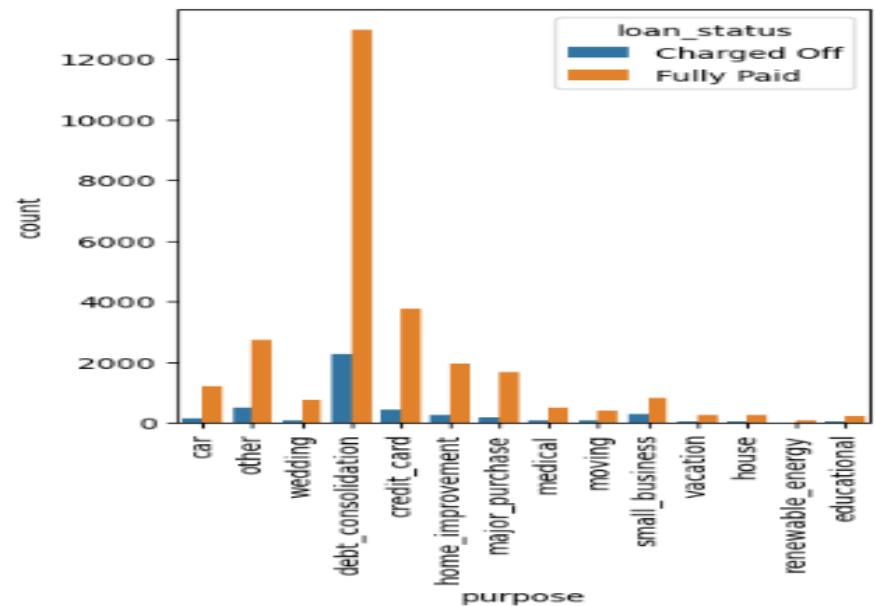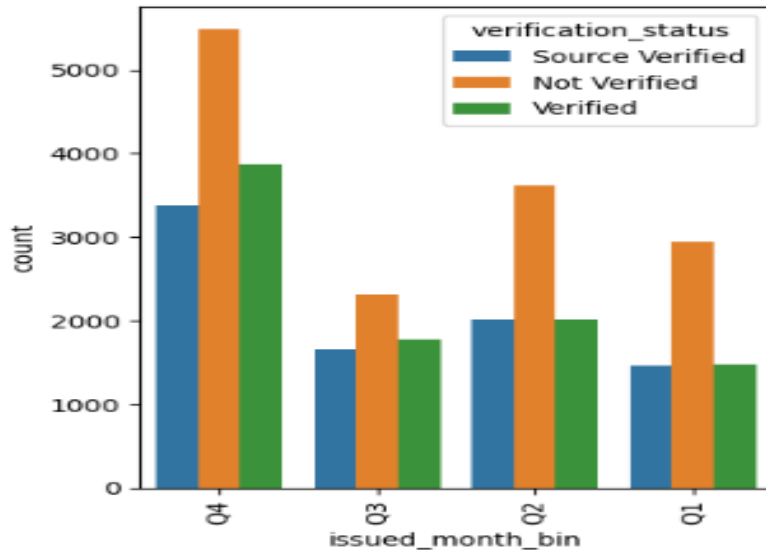- **Installment** ranges between **200 to 400**

Analysis

- People living on rent or whose house is on mortgage are likely to charge off loan in comparison to the one who owns the house.

- Loan taken for debt consolidation purpose have higher chances for loan getting charged off but highest number of loans are also fully paid for same purpose.
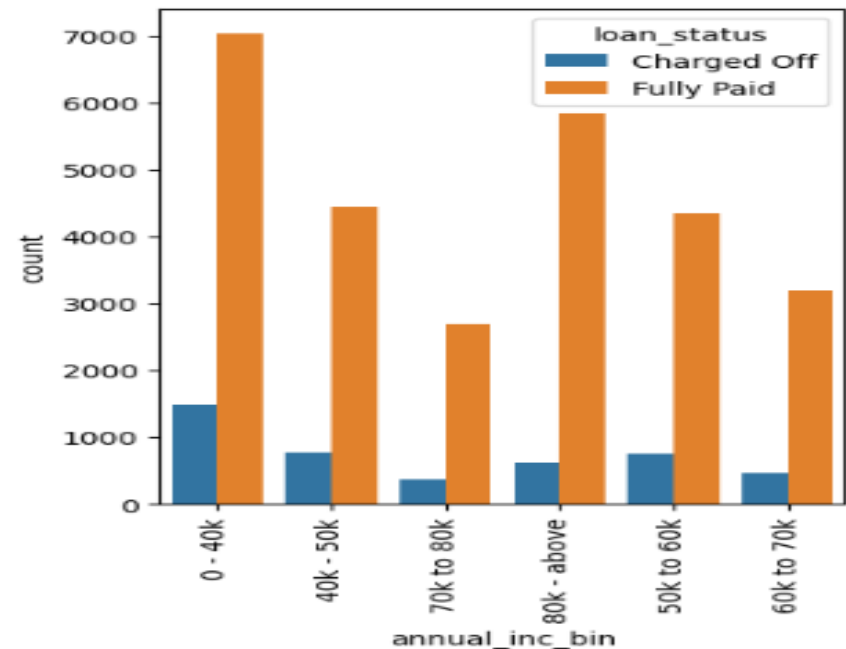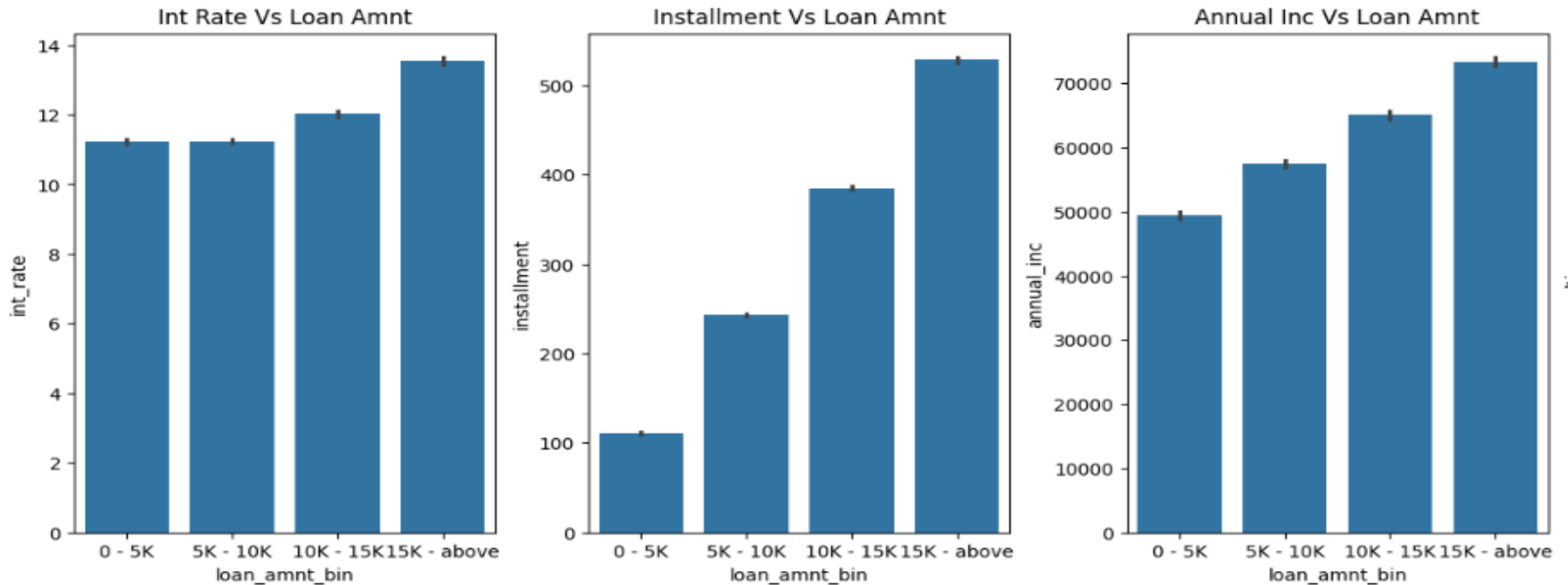
Analysis

- In Q4, loan was given without verifying the details of person. It could be done to meet the year end quota. This might be one of the factor for more defaulters and making loss.

- There are more chances of people with annual income less than 50k becoming defaulters.
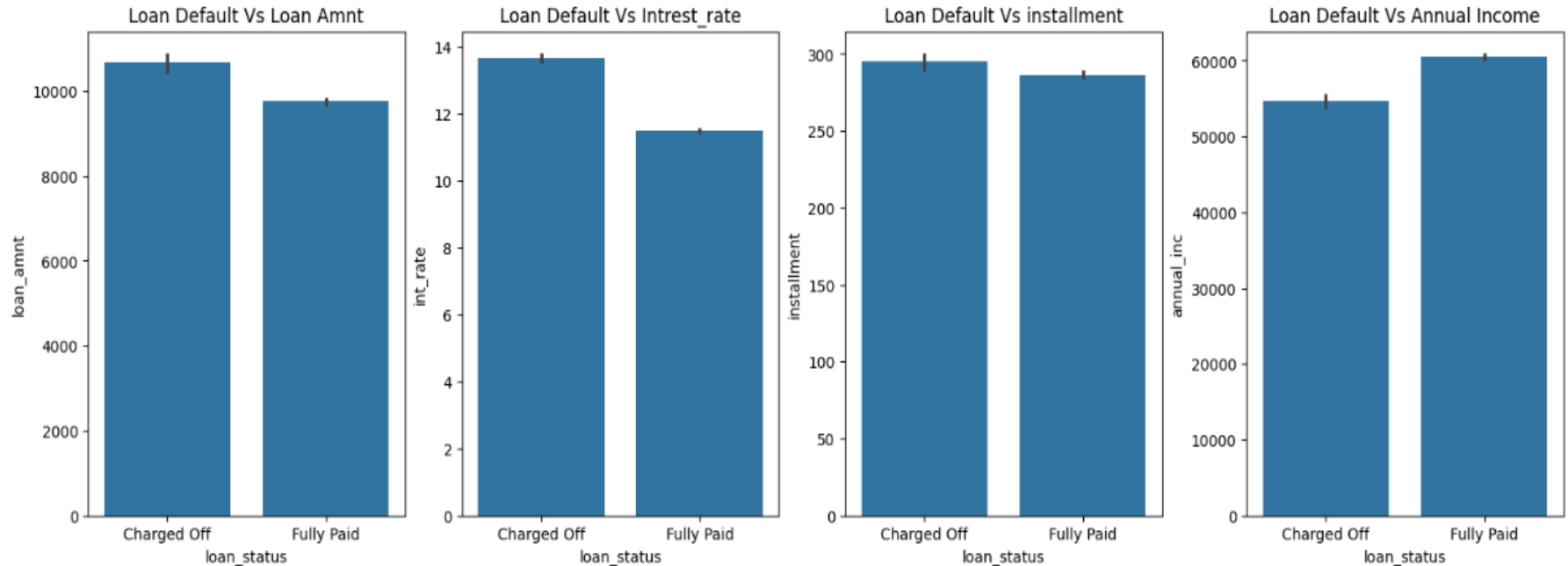
Analysis

- Interest Rate is higher for higher loan amount
- Installments or emi are higher for higher loan amount
- People with higher annual income gets higher loan amount.

Bivariate Analysis on Numerical columns

- There is higher chances of loan getting defaulted if loan amount, interest rate or installments are higher and annual income is less.

Recommendations

- Instead of trying to meet targets, perform background checks on the loans being provided and then provide loan appropriately.
- Try to get more of upper grade loans and shorter term loans instead of lower grade loans and larger term loans which has more probability of defaulting.
- People living on mortgage or rental properties have a tendency to default. need more background verification on providing loans to them.
- Unverified loans provided tend to default most, followed by verified loans. Best to get the information of the source of income of the party taking the loans.
- People who are employed for more than 10 years and having less annual income(less than 50k) are more likely be defaulters.
- When loans are provided for debt consolidation, need to take care of the collaterals provided for the loans, so as to improve the chances of recovering the same upon default.
- Attract more customers for other purpose loans as well, by providing some offers, lower interest rates, etc.
- Higher the loan amount, followed by higher interest rates result in a higher probability of defaulting of loans. There is higher chances of loan getting defaulted if loan amount, interest rate or installments are higher and annual income is less.
- Increase the number of recoveries.

THANK YOU..!!