Lending Club Case Study

Analysis By: Sarika Sakore

Submission Date: 8th May, 2022

Case Summary:

- You work for a **consumer finance company** which specializes in lending various types 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:
- 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
- When a person applies for a loan, there are **two types of decisions** that could be taken by the company:
- **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
  - **Fully paid**
  - **Current**
  - **Charged-off**:
- **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)

Problem Statement:

- Need to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study. In other words, 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 utilize this knowledge for its portfolio and risk assessment.

- **Steps:**

1. Data Understanding
2. Data Cleaning
3. Data Manipulation
4. Data Analysis
   - Univarient Analysis
   - Bivarient Analysis
   - Multivarient Analysis
5. Recommendations

Data Understanding:

- There are many features in provided loan data, some important features are
  - Annual Income: Based on annual income of borrower, it can be identified whether loan can be provided to borrower or not
  - Loan Amount: The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
  - Loan Status: Current status of the loan. There are 3 types of loan status
    - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
    - **Current**: Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
    - **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
  - Interest Rate: Interest rate applied on the loan
  - Employee Length: Employee experience in years
  - Purpose: Purpose for taking loan
  - Term: Period in months within loan need to repay
  - Pub_Rec_Bankruptcies: Number of public records where borrower is failed to repay the loan
  - DTI(Debt to Income ratio): Ratio of how much borrower owe to how much borrower earn
  - Home Ownership: Borrower's home ownership like rent, own, mortgage etc.

Data Cleaning:

- There are multiple features with NA data in whole column, removed that columns
- Some features exists with unique value, it wouldn't help in analysis removed that columns
- Some features will have user specific info like id, member_id, url, desc etc, which are not adding any valuable insight in our analysis, removed the same
- Identifying the columns which are useful in analysis, check the missing value percentage in that columns and either identify the correct value to replace missing values or remove missing value rows.

  e.g. emp_length missing value percentage id ~2.6% , replacing it with '0' years

  Removing missing value rows from last_pymnt_d, revol_util, last_credit_pull_d, as % of missing value is less . It will not impact our analysis

Data Manipulation:

- There are 3 status of loans Current, Fully_paid and Charged_Off, so Current status is neither fully_paid nor defaulted, removing the rows with Current loan status
- Removing % character from int_rate and revol_util columnf and converted to numeric data
- Removing years from emp_length and converted to numberic data
- Find out the outlier from various features like annual_inc, loan _amnt and remove the data where maximum variation is observed

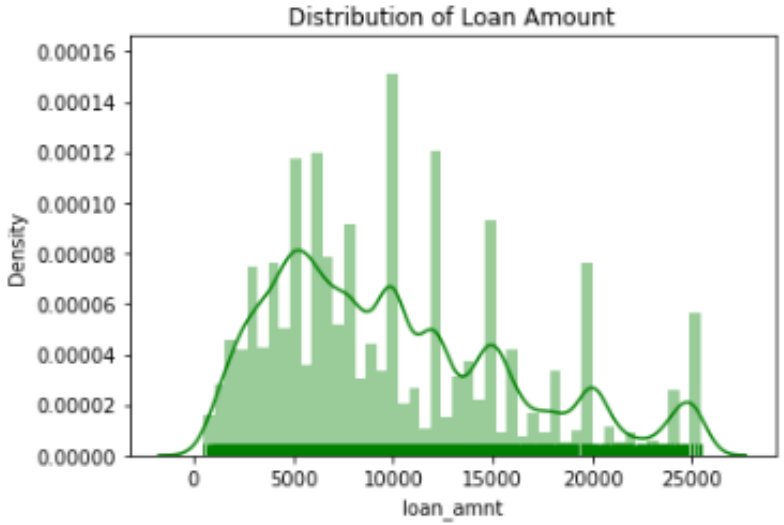
Data Analysis:

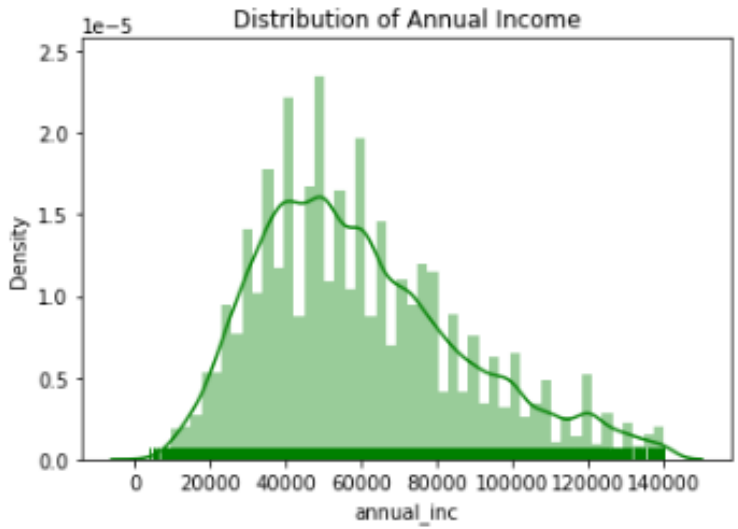
- **Univariate Analysis:**
- **Distribution of Loan Amount:**

  Observed spike in between 5000 and 15000, majority of borrowers applying in between these amount

- **Distribution of Annual Income:**

  Maximum borrowers who are taking loans having annual income of 30000 to 600000



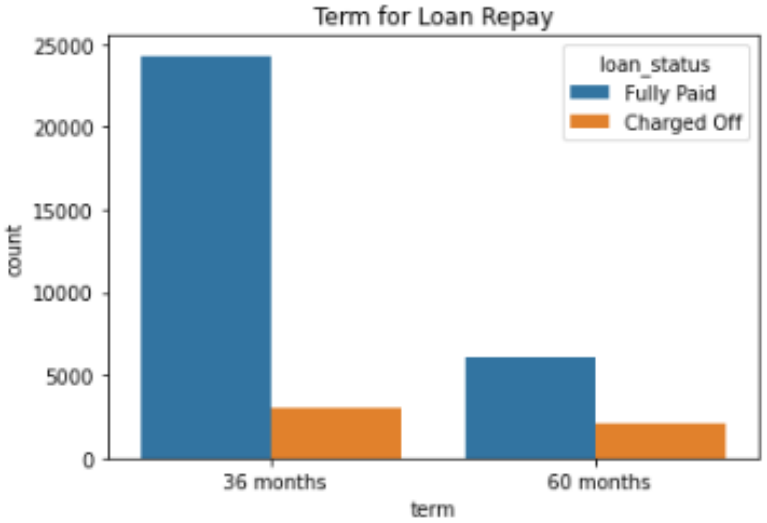


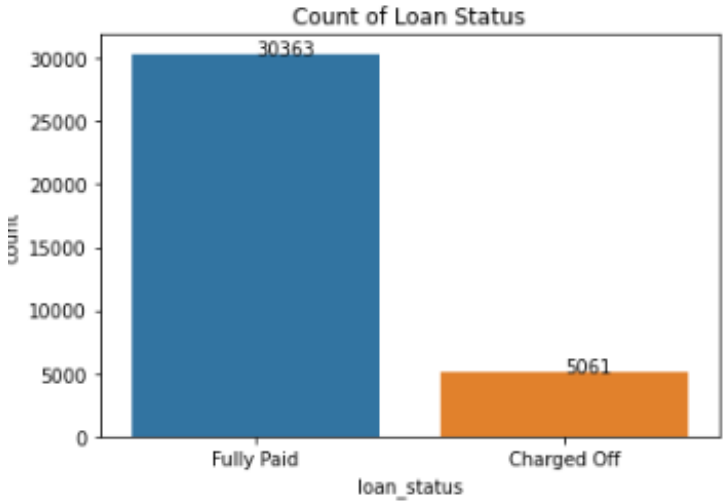
Univariate Analysis:

- **Distribution based on Loan Repay Term:**

  Borrower who had taken the loan for maximum time period, there are high chances that borowwer will not pay the installments on time

- **Distribution of Loan Status:**
  - Based on loan status, observed that 14% borrowers didn't pay the loan installments within due time for long period of time.



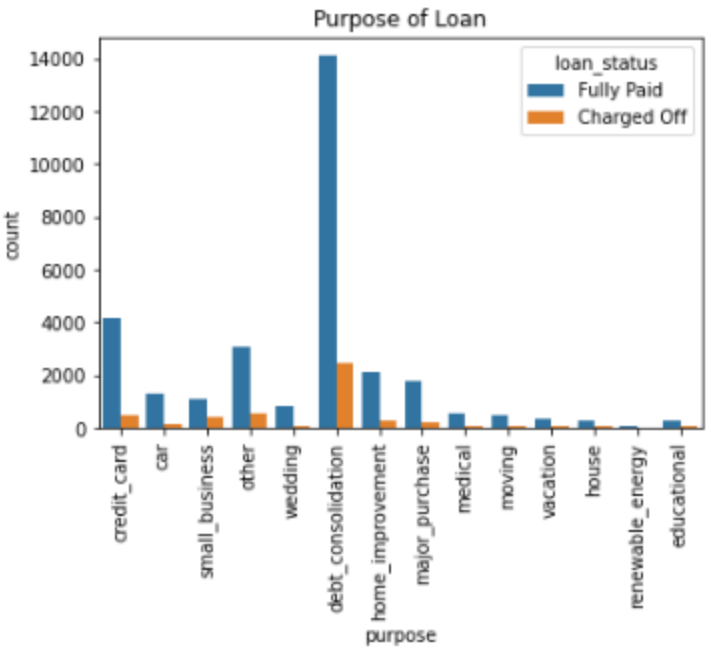


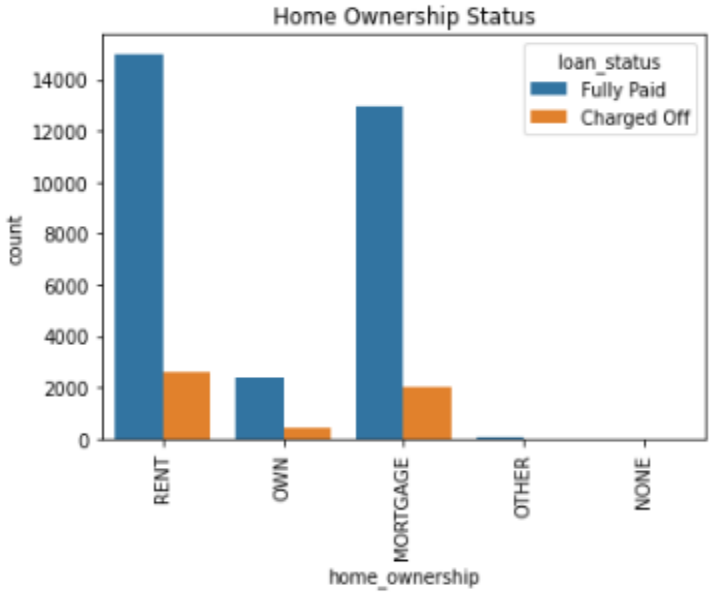
- **Univariate Analysis:**
- **Distribution based on Loan Purpose:**

  Majorly borrowers are taking loan for debt consolidation and paying credit card bills and high chances of charged-off is for these categories only

- **Distribution of Loan based Home Ownership:**

  Observed that majority of borrowers who are taking loan are living in rented house or mortgage their house and even high chances of charged-off for these categories, as compared to others.





Data Analysis:

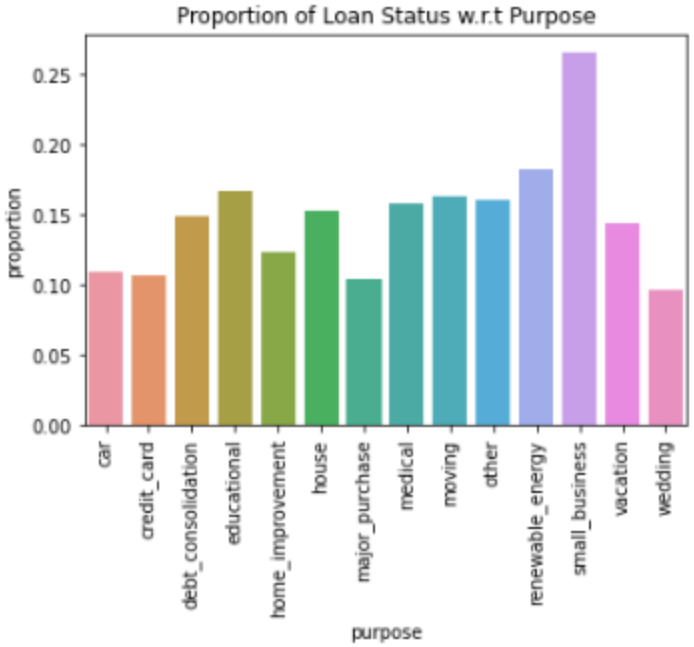
- **Bivariate Analysis:**
- **Proportion of Loan Status w.r.t Purpose:**

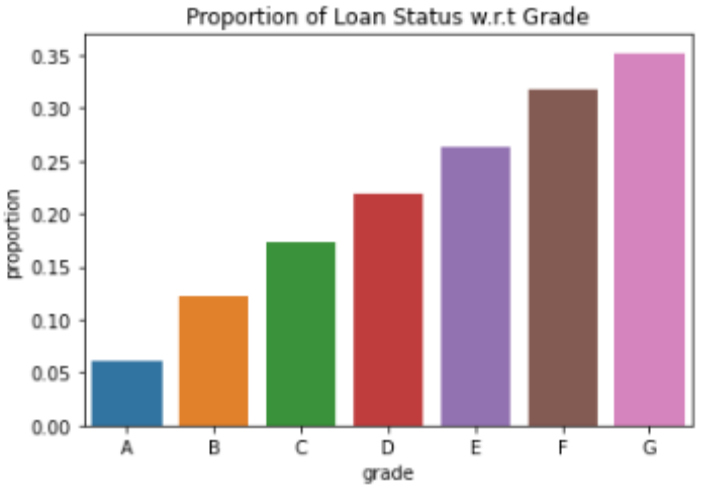
  Based on proportion of loan status , it can be identified that borrowers who are taking loan for small business are having maximum chances of charged-off.

- **Distribution of Annual Income:**

  Based on proportion of loan status , chances of charged-off is increasing with grade moving form A to G.

  Borrowers with grade A having less chances of charged-off while borrowers with grade G are having maximum chances of charged-off





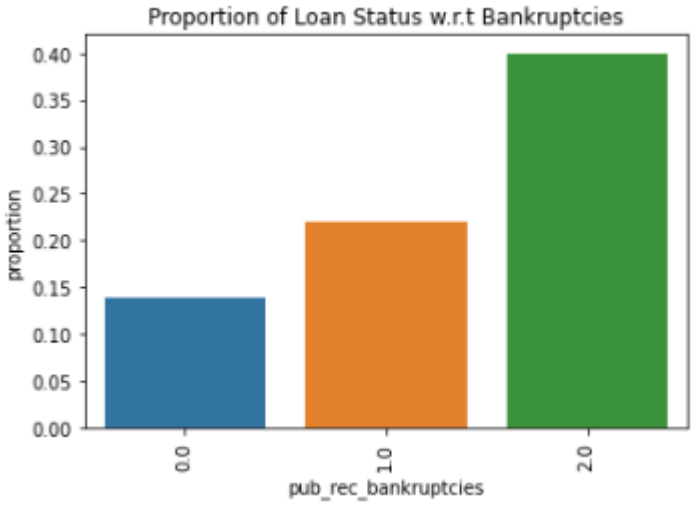
- **Bivariate Analysis:**

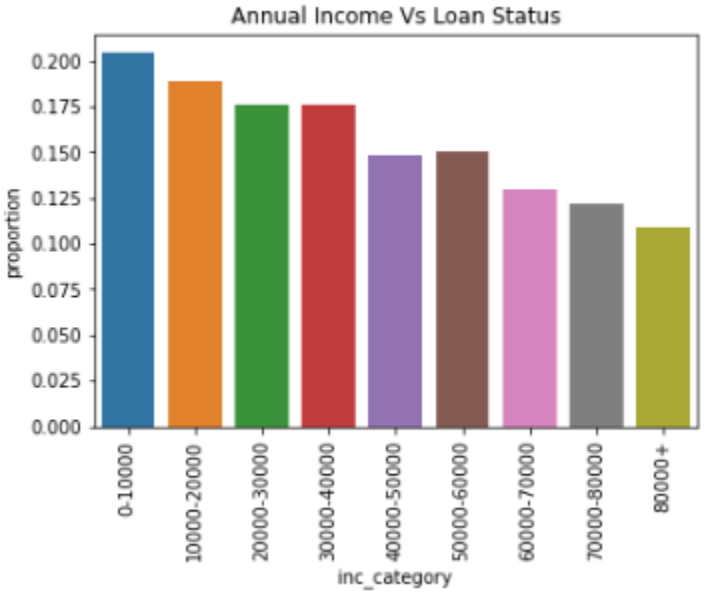
❖ **Proportion of Loan Status w.r.t Bankruptcies:**

From below plot observed spike at 2, it means the borrowers who had charged-off earlier there maximum chances of charged-off again

❖ **Annual Income vs. Loan Status :**

Observed from below plot, chances of charged-off is decreasing as annual income is going to be increased. Borrowers who are having less income high chances of charged-off.





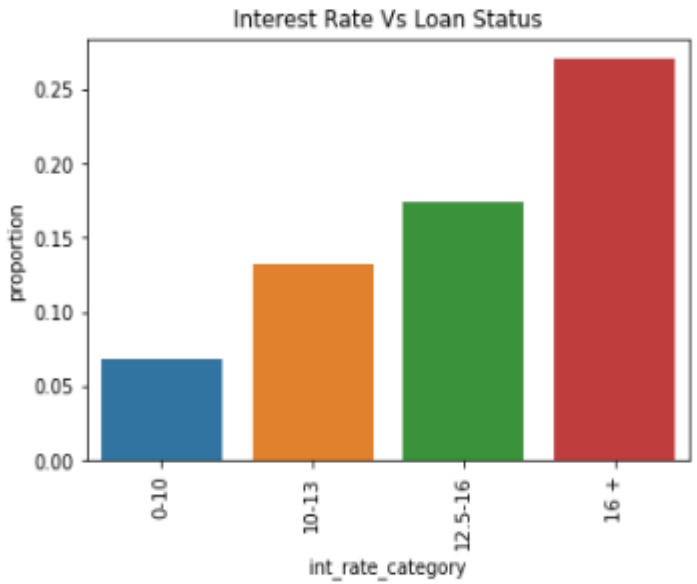
- **Bivariate Analysis:**

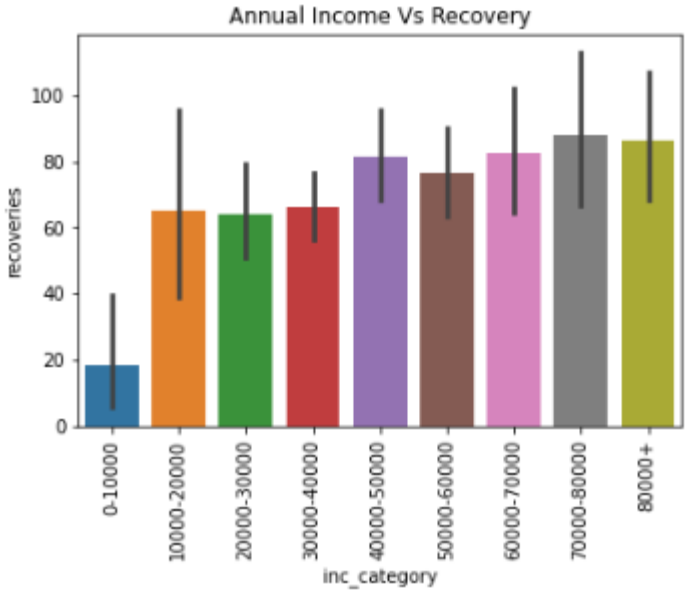
❖ **Interest Rate vs. Loan Status:**

From below plot observed that increased in interest rate leads to increased in charged-off proportion

❖ **Annual Income vs. Recovery:**

Observed that loan amount is recovered when annual income is high.





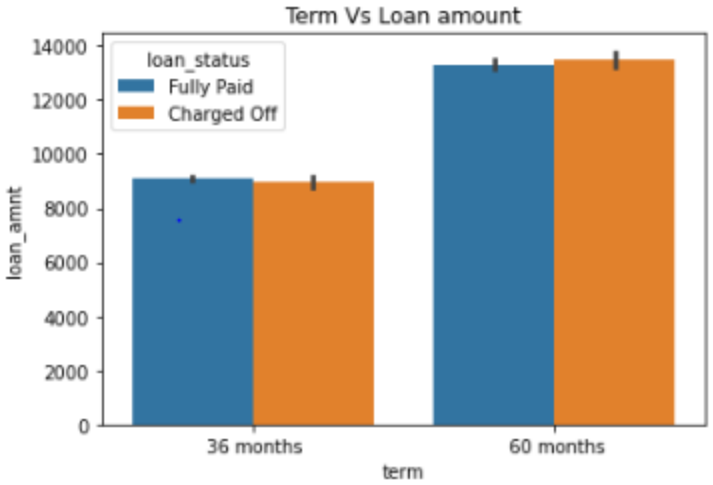
- **Bivariate Analysis:**

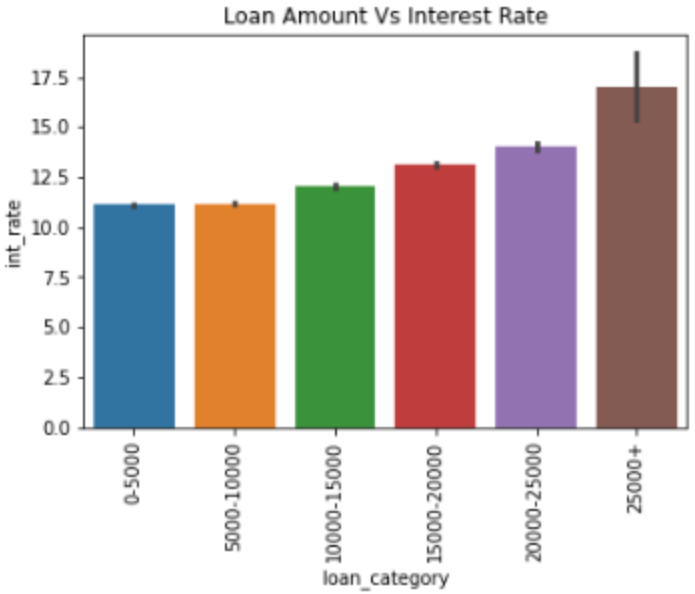
❖ **Term vs. Loan Status:**

Loan amount is directly proportion to term, means if loan amount is high then term is also high

❖ **Loan Amount vs. Interest Rate:**

Loan amount is directly proportion to interest rate, means if loan amount is increased then interest rate is also high





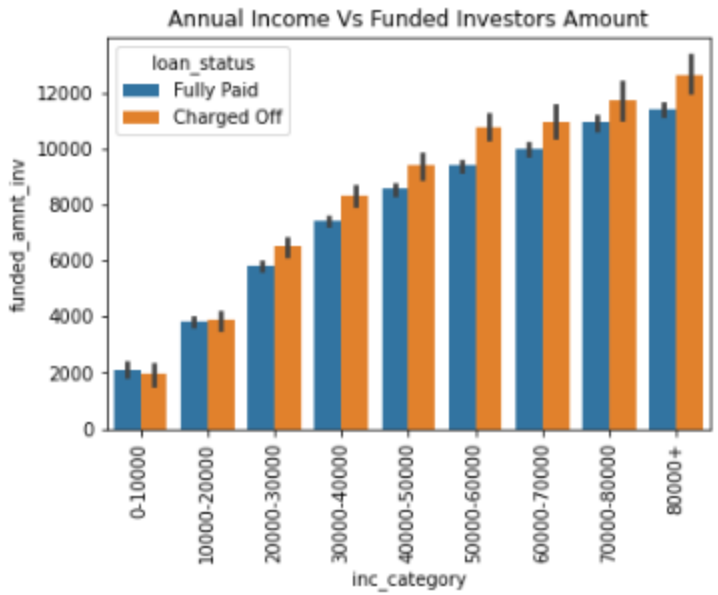
- **Bivariate Analysis:**

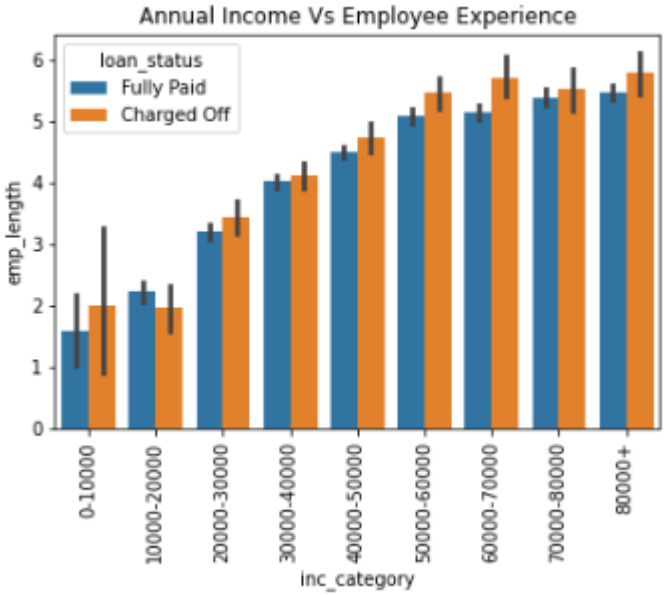
❖ **Annual Income vs. Funded Investors Amount:**

If annual income is high then loan amount approved by investors is also high and there is high chances of chances of charged-off for highly funded amount.

❖ **Annual Income vs. Employee Experience:**

If annual income is low and employee experience is less then high chances of charged-off

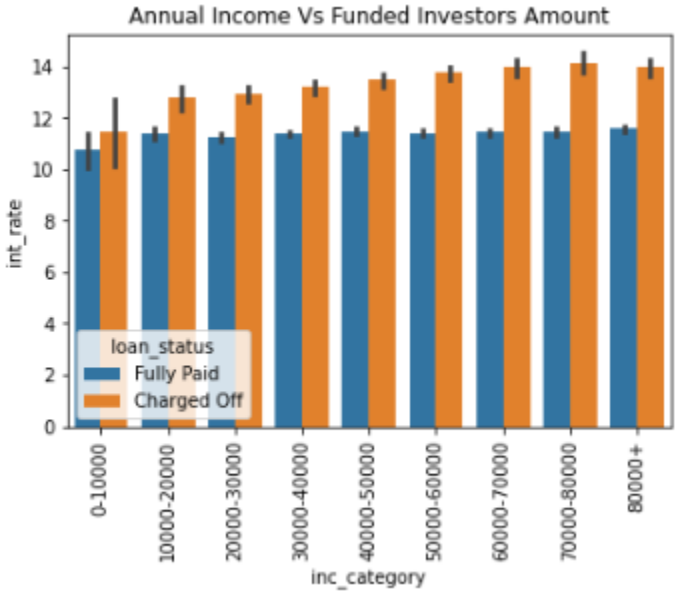




- **Bivariate Analysis:**

❖ **Annual Income vs. Interest Rate:**

If annual income is above 60000 and interest rate is above 12% then maximum chances of charged-off.



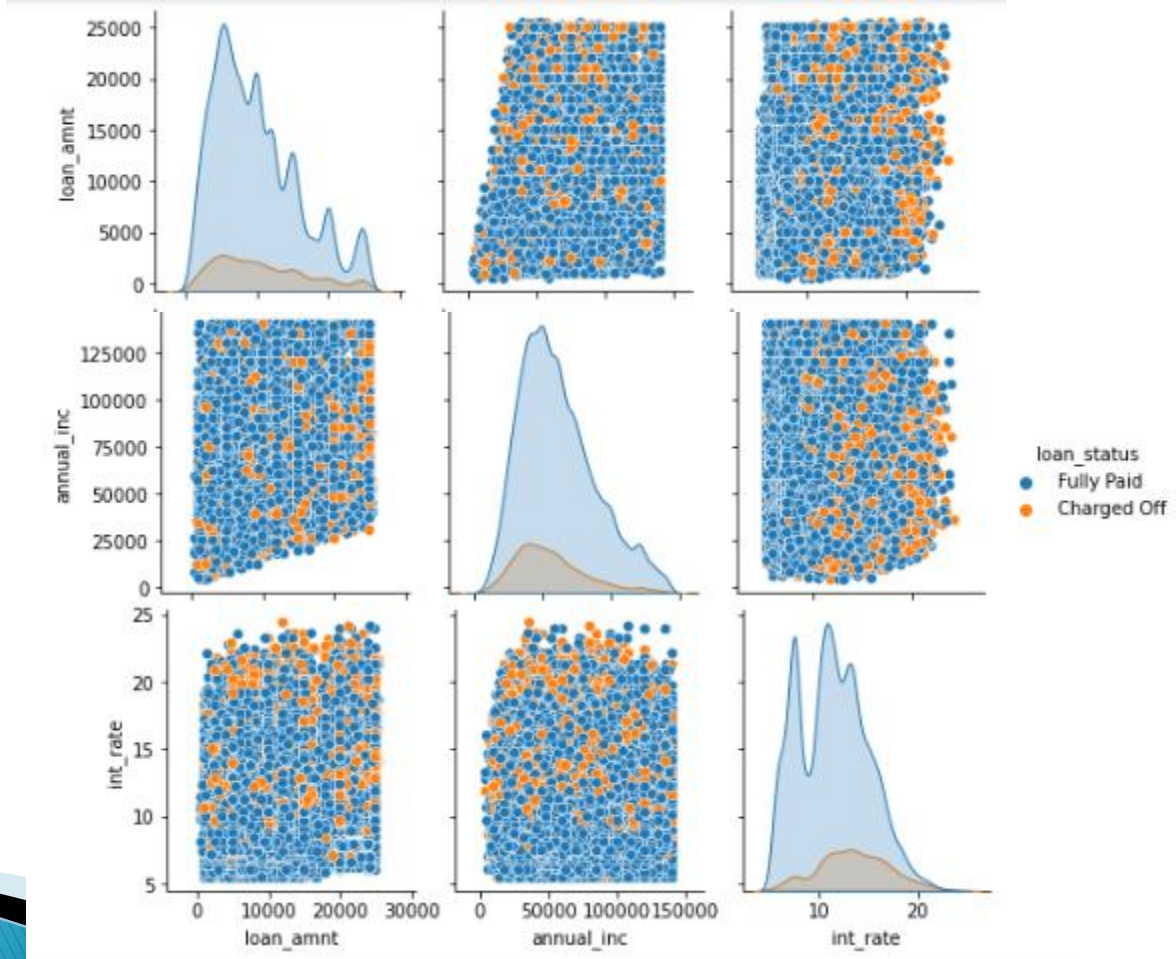
Data Analysis:

- **Multiivariate Analysis:**

❖ **Distribution of Loan Amount:**

1. Maximum chances of charged off if loan amount is high because for high loan amount interest rate is also high

2. Higher annual income is getting high loan amount



Recommendations:

- **Please find below recommendations, based on analysis done to avoid applicants who may be become default**
  1. If applicant is already default and receives loan application again then it should be rejected, as high chances of charged-off again
  2. If applicant is having annual income below 60K and loan amount with interest rate 12% and above
  3. If interest rate is above 15% then high chances of charge-off
  4. If applicant is taking loan of 5K and above with interest rate 15% and above
  5. If applicants home ownership is either Rented or Mortgage
  6. Applicants who are taking loan for small business having higher chances of charged-off
  7. Applicant who are taking loan for large time period maximum chances of charged-off
  8. Applicants whose grade is G
  9. Applicants whose work experience is less having maximum chances of charged-off