## I. Problem Statement:

- Lending Club is an online loan marketplace, facilitating various types of online loans.
- Given a historical loan dataset, the loan lending company wants to understand the possible variables which can cause loan defaulting using EDA.
- The company wants to utilize this knowledge / information, for its portfolio and risk assessment.

## II. Assumptions:

- Columns where more than 30% of the fields are NaN / Empty, as these fields might not indicate right data. So, all such columns are removed from analysis.
- Columns, which are associated with post approval, as they can't form the basis for initial loan approval process. So, all such columns are removed from the list [Details captured in Jupyter Notebook].
- ***loan_status*** column is identified as target column, as loan status parameter(Charged Off) indicates the defaulter.
- Loans with status as "Current" are not considered for analysis, as the process of rejecting the loan is only for new applicants, and "Current" applicants are not impacted.

## III. Analysis Approach:

- In this case study, as an initial approach the data dictionary is checked to understand the data thoroughly.
- Once the relevant fields are understood, data cleaning was done following below criteria:
    - Columns where more than 30% of the fields are NaN / Empty
    - Columns which do not impact the final result, such as zip_code etc.
    - Columns, which are associated with post approval
- Detecting Outliers:
    - Analysed the relevant fields to detect outliers, if any, and corrected some of the fields accordingly.

- EDA:
  - ***loan_status*** column is identified as target column, as loan status parameter indicates whether borrower can be defaulter (status "Charged Off") or not.
  - In current analysis, loans with status as "Current" are not considered for analysis, as the process of rejecting the loan is only for new applicants, and "Current" applicants are not impacted.
  - Univariate Analysis:
    - Analysis done on the impact of some of the parameters with loan_status column to see, how various factors impacting defaulters.
  - Bivariate Analysis:
    - Analysis is done, on the impact of loan status with combination of variables e.g.: how grade, loan_amnt impacts the loan_status.

## IV. Results of Univariate Analysis:

- Majority of the loans are taken under the category of "debt_consolidation", and the percentage off charged off users is more under this category.
- For loan amount in excess of $10K has higher chances of defaulters.
- For loans taken with higher interest rates (>14%) has higher chances of defaulters.
- Loans with DTI of 12-20% has higher chances of defaulting.

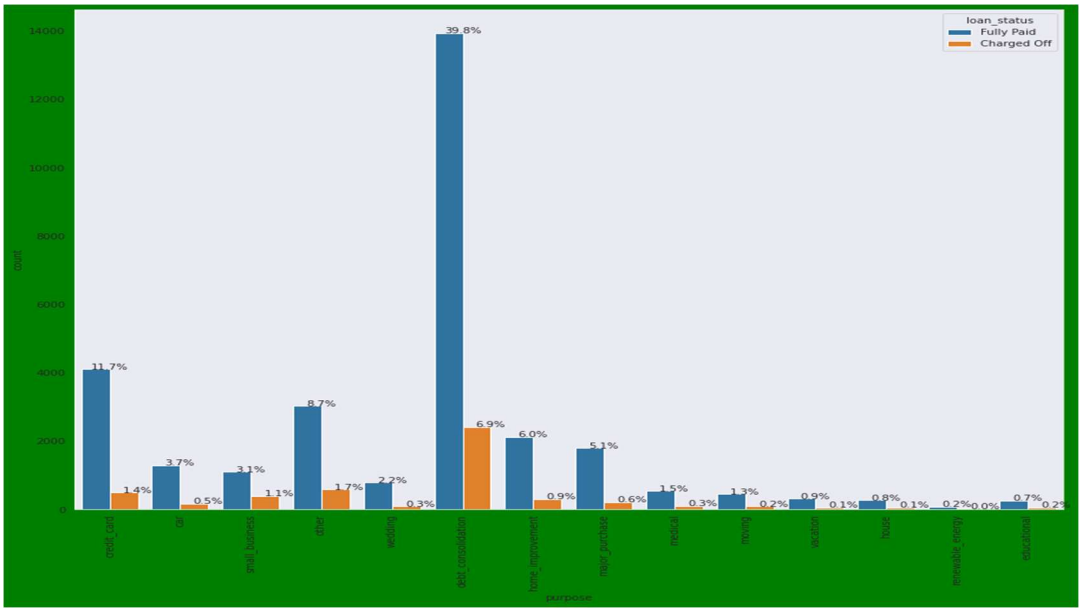
## V. Results of Bivariate Analysis:

- Loans with Grade B has higher chances of Defaulting. Chances of defaulting is more with lower subgrades (4 or 5).
- Loans with smaller amount (<8K), with loan purpose of moving (or) renewable energy has lesser chances of defaulting.
- Loans taken up to 10K, for purposes of debt_consolidation, small business has more chances of defaulting.
- Borrowers with home_ownership as "MORTGAGE" with higher annual_inc has higher chances of defaulting.
- loan_amnt of >15K, and int_rate > 14% has higher chances of defaulting.

- Loans taken for longer term (60 months) with higher loan amount (>10K) has higher chances of defaulting.
- Borrowers, whose employment is > 6years, and loan amount > 10K has higher chances of defaulting.

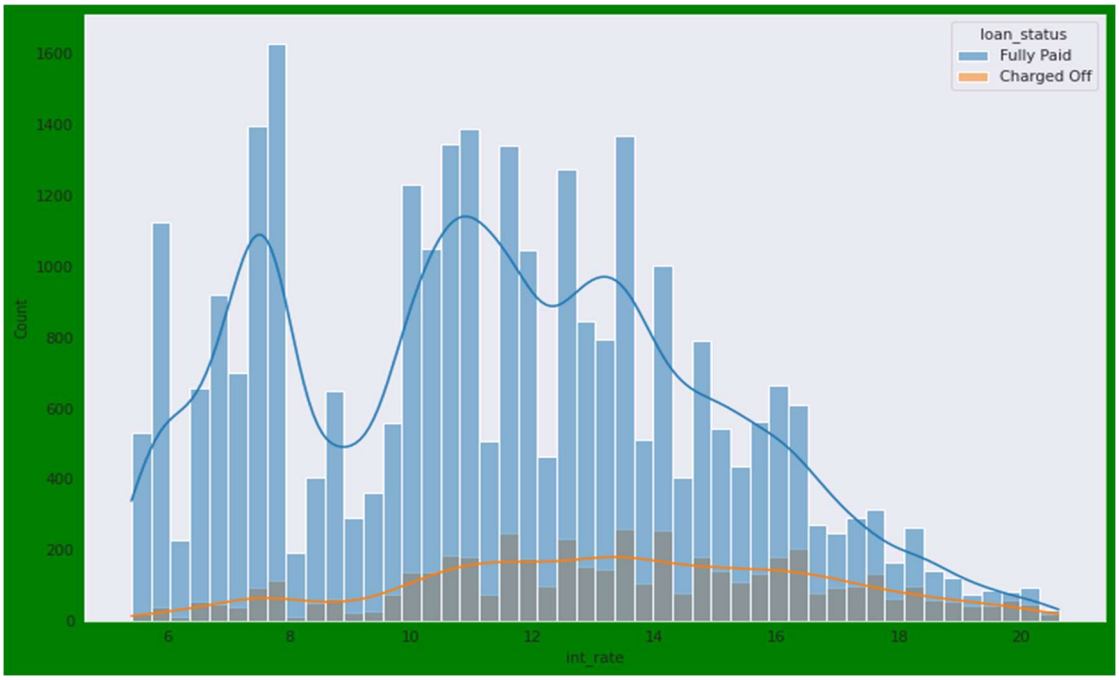
## VI. Visualizations & Summary:

- Captured some of the visualizations of univariate, Bivariate analysis.
- **Univariate Analysis:**
  - Loan purpose vs loan_status



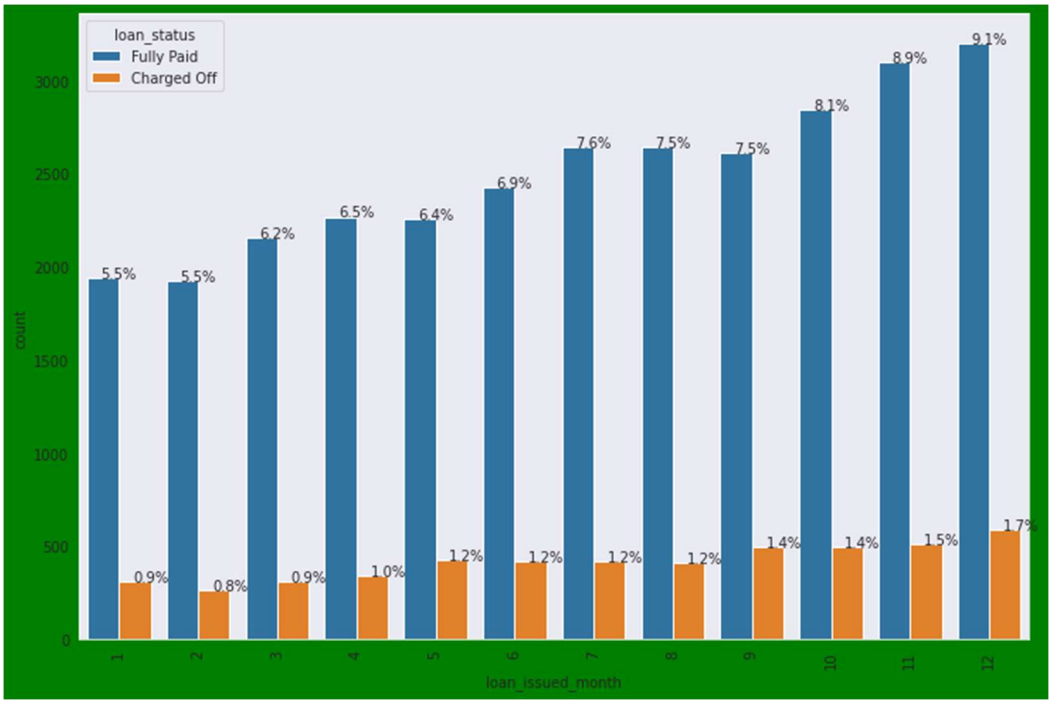

    - **Observations:** Majority of the loans are taken under debt_consolidation, followed by other category and there are more defaulters (charged off) under that category.

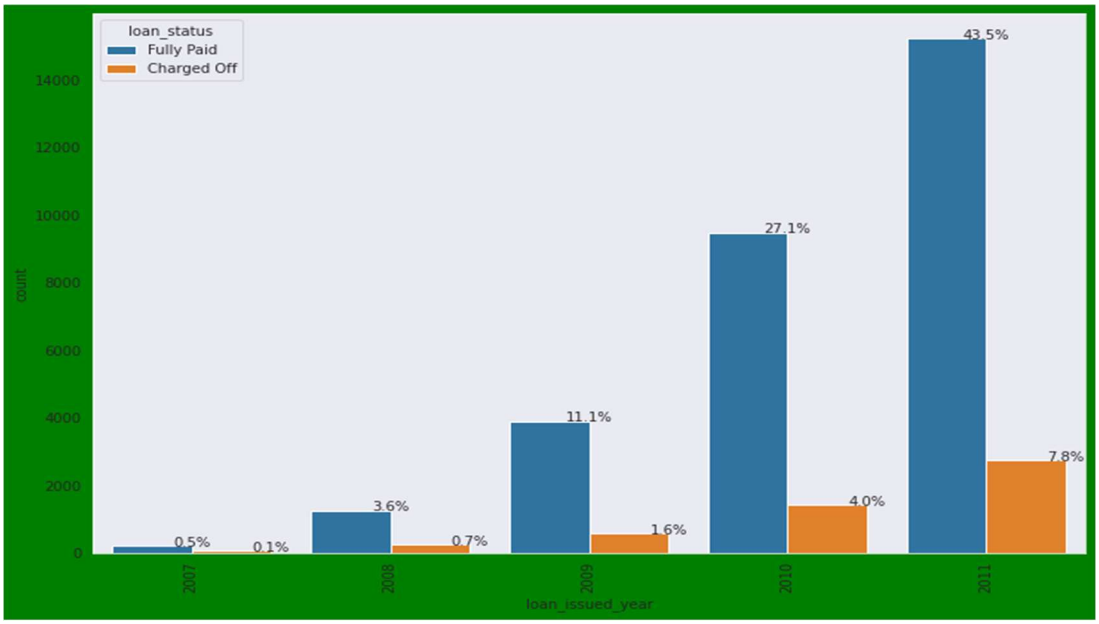
  - Int_rate Vs loan_status




- Observations: For interest rates > 14% there are more chances of defaulters.

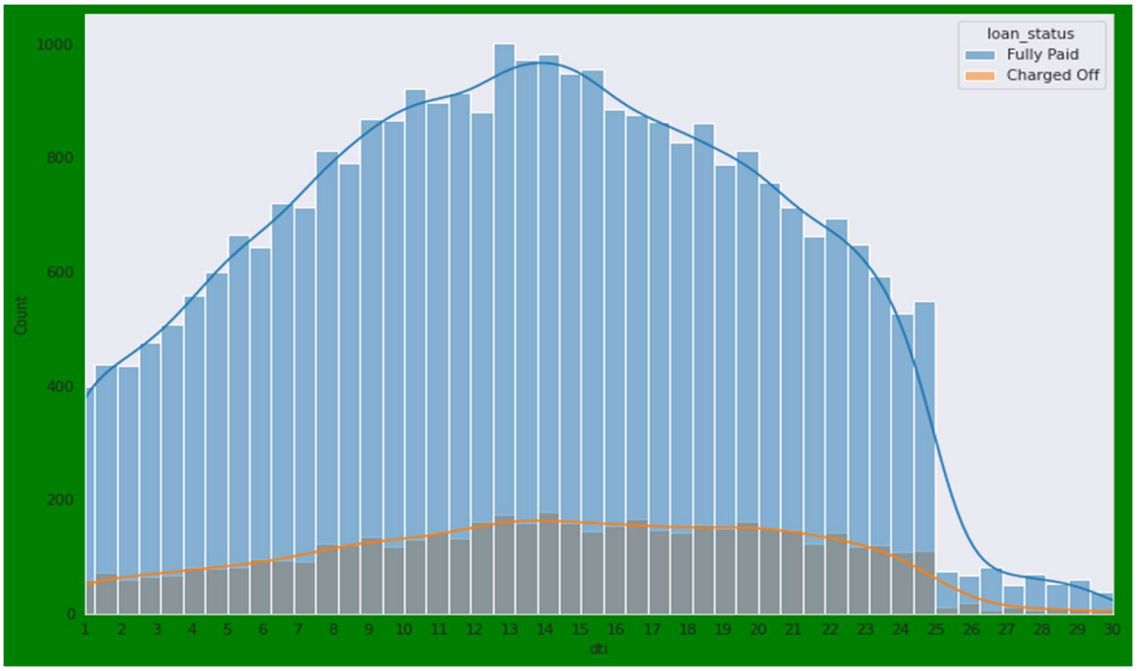
- loan_issued_month, loan_issued_year vs loan_status





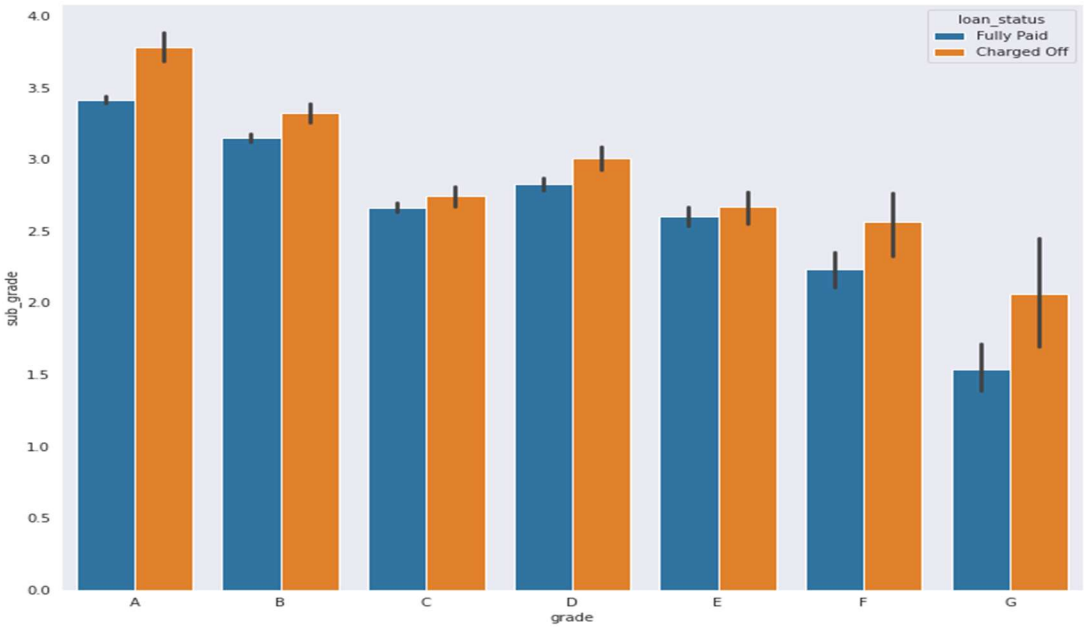

- Observations: There are more defaulters for the loans issued in the months of December, November as compared to other months. There are more defaulters for loans issued in the year of 2011, followed by 2010, as compared to other years.

- dti vs loan_status

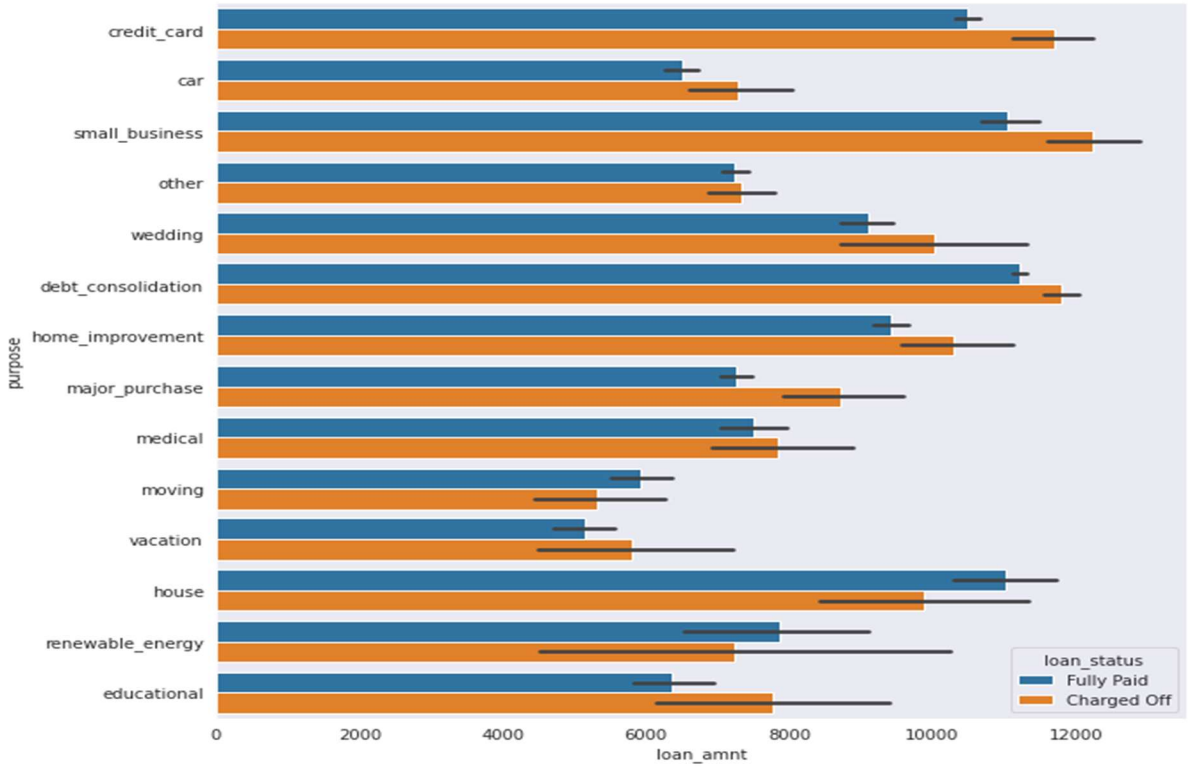



- Observations: For DTI (Debit to Income Ratio) of 12-20% has high chances of charging of as compared to others. For DTI > 25% there are very less chances of charging off.

- Bivariate Analysis:
  - grade, subgrade vs loan_status

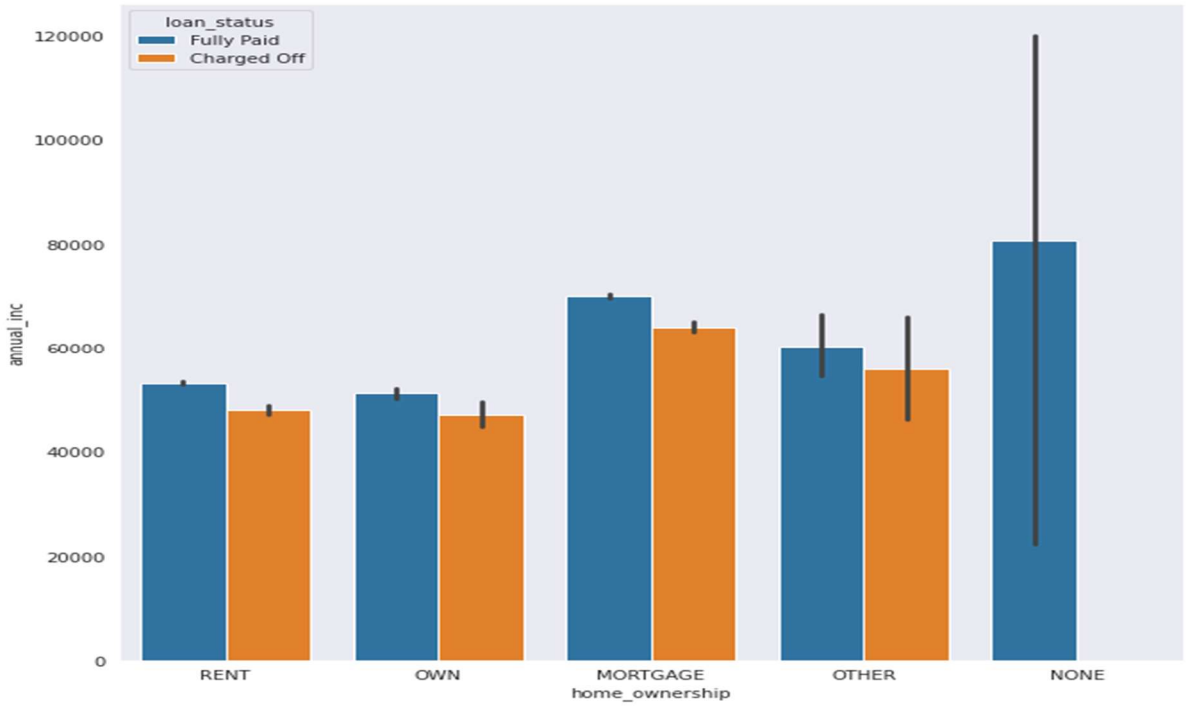



    - Observations: The loans of Grade B have higher chances of defaulting. There are higher chances of defaulting in case of loans with sub grade 4 or 5 in majority of the grades. Moving from grade A to grade, G the chances of defaulting increases even for lower grades (2 or 3) grades as well.
  - loan_amnt, purpose vs loan_status:



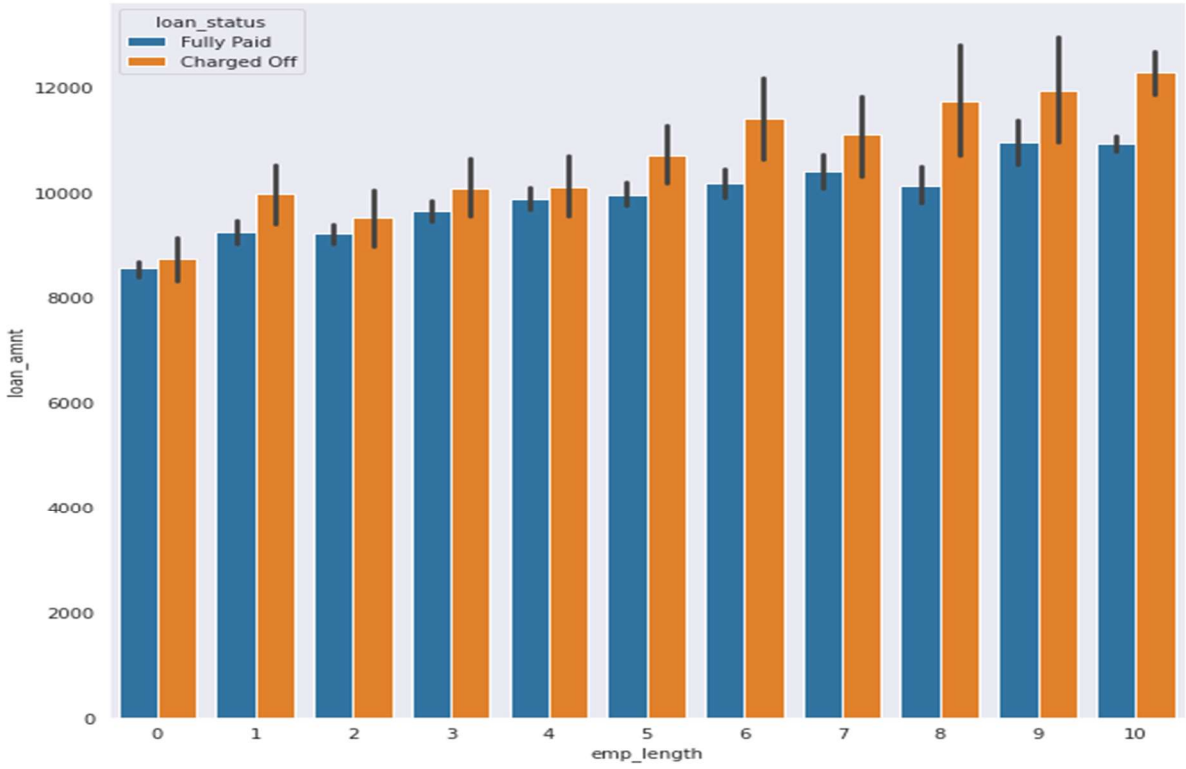

- Observations:
  - Loans taken for the below category are smaller and the charged off possibility is lesser (<8K)
    - moving, renewable energy
  - Loans taken for moderate amount for below category has more chances of charged off (up to 10K+)
    - debt consolidation, small business, Credit card, medical, home improvement
  - Loans taken for the purpose of "house" has generally been with more loan amount, and the possibility of charged off less.

- annual_inc, home_ownership Vs loan_status




- Observations: Borrowers with higher annual income the home ownership of "MORTGAGE" or "NONE". The possibility of charged off is more for home ownership category "MORTGAGE" or "OTHER".

- emp_length, loan_amnt impact Vs loan_status



- Observations; Employees with more experience have requested for longer loan_amnt, and in this case the defaulters are more as the employee experience and loan amount increases. For loan amount > 10K, and experience, greater than 6 years the defaulter percentage is relatively higher.

## Insights & Recommendations

- From the analysis so far, there are higher chances of **defaulting** in below conditions:
  a. Borrower, annual income is 60000 or below and take loans for the purposes of debt_consolidation.
  b. Borrower loan interest rate is 14% and loan amount > 15K or above
  c. Borrower loan term is longer (60 months)
  d. Borrower employee experience is>6 years and loan amount > 10K
  e. Borrower has house on MORTGAGE or OTHER home ownership category has taken loan amounts $10K+
  f. Loan issued in the months of Dec, Nov, and loans issued in the year of 2011
  g. Loan issued for the purposes of Debt Consolidation, Other
  h. When the Borrower loan grade is B, F, G or any loans given with subgrade of 4 or below.
  i. Borrowers whose DTI ranges from 12-20%
  j. When installment variable is ranging from 160-350
- Recommendations: The below recommendations are based on above data analysis. In following cases there are higher chances for borrower to pay full loan amount.
  a. Borrowers with home_ownership category of "OWN"
  b. Borrowers with years of experience < 6 years and loan amount requested is <10K
  c. Borrowers with requested loan amount <10K, and are falling under <=14%b loan interest rate.

d. Borrowers with grade A

e. Borrowers whose loan amount is < 10K and requesting for the purposes of Moving, renewable energy.

f. Borrowers whose annual income is > 60K

g. Borrowers with DTI >= 25%