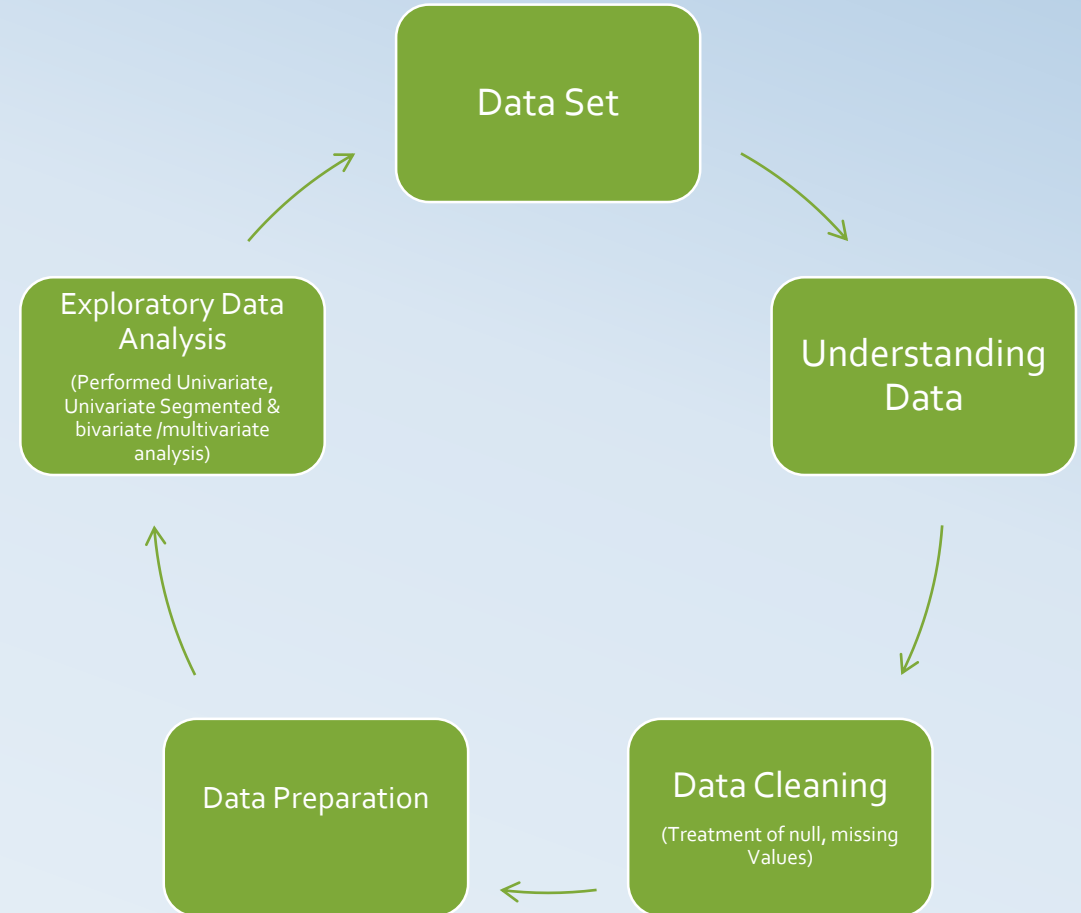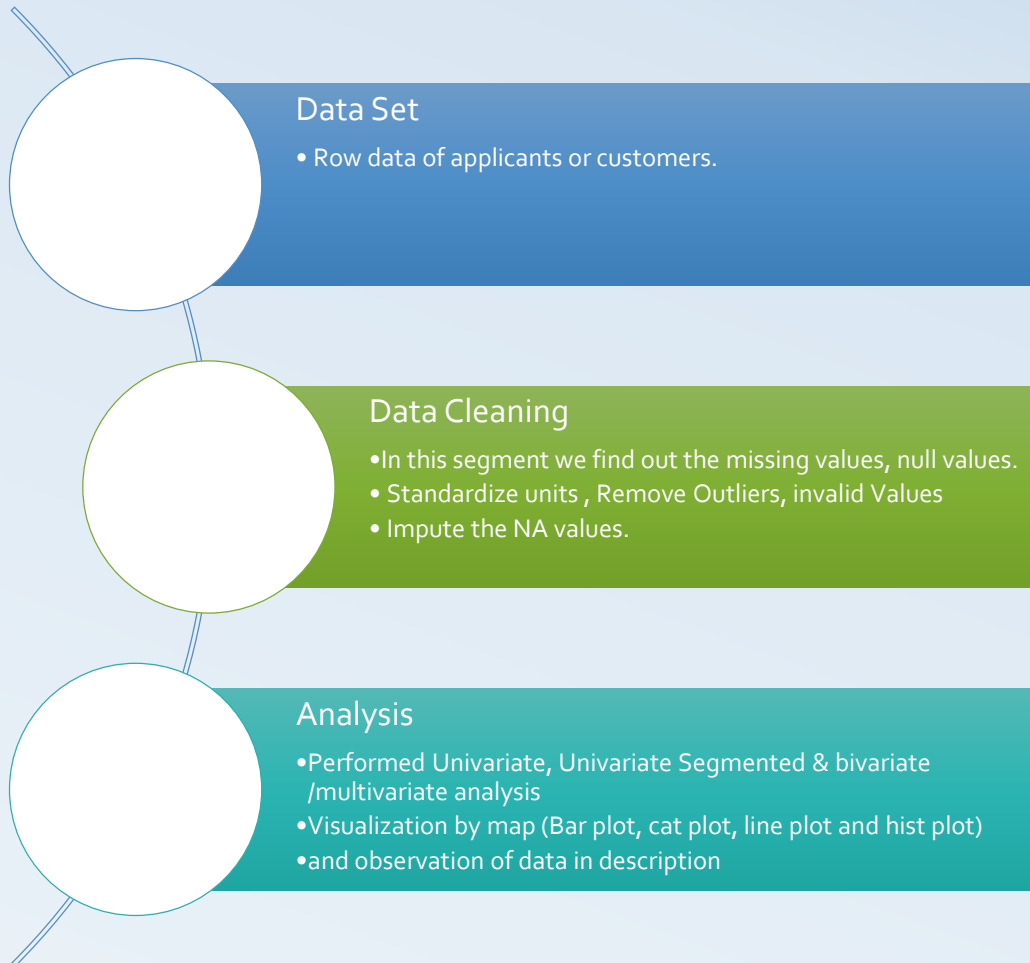# Lending Club Case Study

Group Members:

1. Dharmendra Kumar
2. Anuja Gurav

Case Study Objectives

- In this case, We will understand how quality of data is used to minimize the risk of loosing money while sanctioning loan to customers

- Sanctioning loans to applicants who is not able to pay the loan amount is the largest source of financial loss or NPA

- Identification of such applicants using EDA is the aim of this case study.

- Perform an analysis to understand the driving factors (or driver variables) behind loan default

- The main objective is to be able to identify these bad loan applicants, By avoiding this type of bad loans will reduced the amount of loss in business.

Problem Solving Methodology

Data Set
- Row data of applicants or customers.

Data Cleaning
- In this segment we find out the missing values, null values.
- Standardize units , Remove Outliers, invalid Values
- Impute the NA values.

Analysis
- Performed Univariate, Univariate Segmented & bivariate /multivariate analysis
- Visualization by map (Bar plot, cat plot, line plot and hist plot)
- and observation of data in description

Data Set

Understanding Data

Data Cleaning
(Treatment of null, missing Values)

Data Preparation

Exploratory Data Analysis
(Performed Univariate, Univariate Segmented & bivariate /multivariate analysis)

Data Cleaning Steps

Step 1: Data Cleaning
      a. Fixing rows and column
      b. Null Values/Missing Values

Step 2: Standardizing values
      a. Standardize units
      b. Remove Outliers
      c. Invalid Values
      d. Filtering data

Step 3: Univariate Analysis

Step 4: Segmented Univariate Analysis

Step 5: Bivariate/Multivariate Analysis

Step 6: Visualizing

Step 7: Results/observation

Analysis

● The main goal of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

● We performed data cleaning and preparation on the Loan dataset:

      a. Imputed the NA values for all the variables

      b. Created Five new columns:

            ■ installment_group

            ■ loan_amt_group

            ■ funded_amnt_inv_group

            ■ dti_groups

            ■ int_rate_groups

● During Univariate analysis we have created:

      a. Count Plots charts to check out the no of defaulters.

      b. Box plots to detect the Outliers

      c. Hist plots for last six month enquiry.

● During Bivariate/Multivariate analysis we have created:

      a. Line Plots charts to check loan distribution.

      b. Count plots to check loan distribution over several attributes.

      c. Box plots to check interest rate between 'charged off' and 'fully paid'.
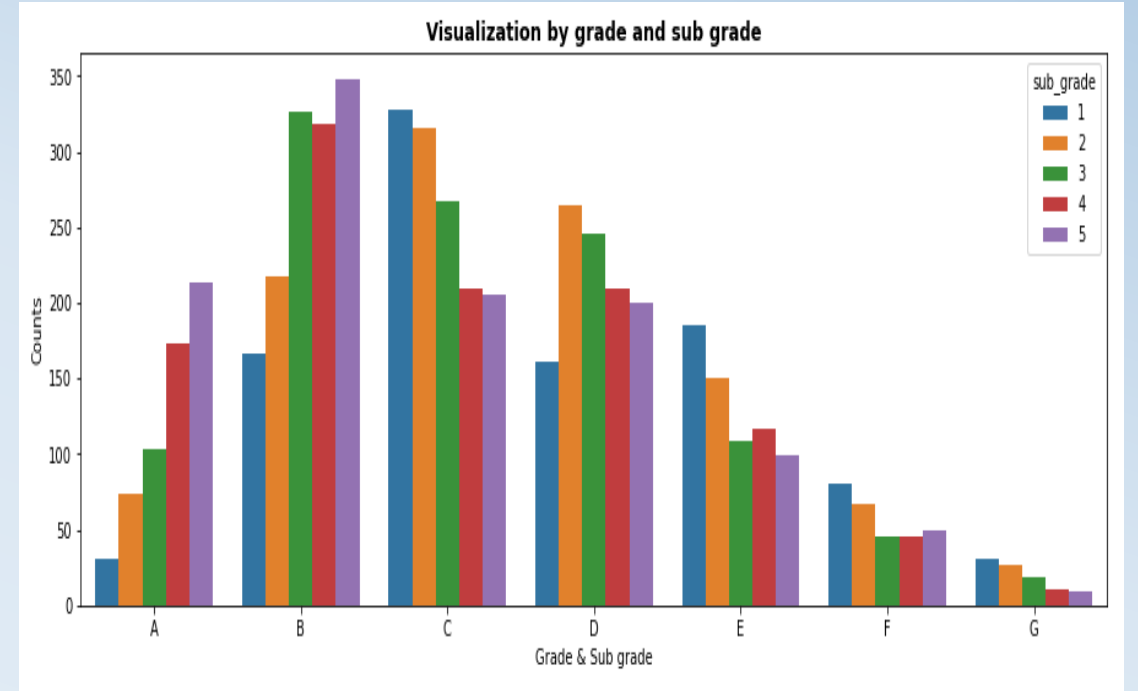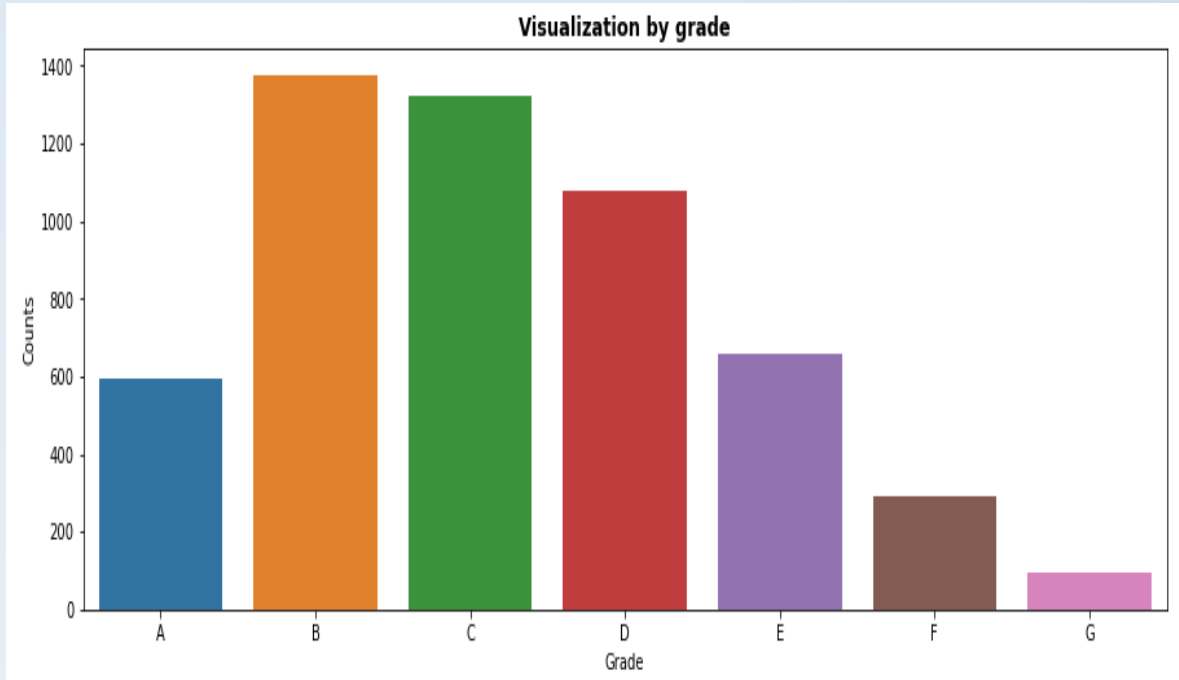
Loan Status & Loan Amount Analysis
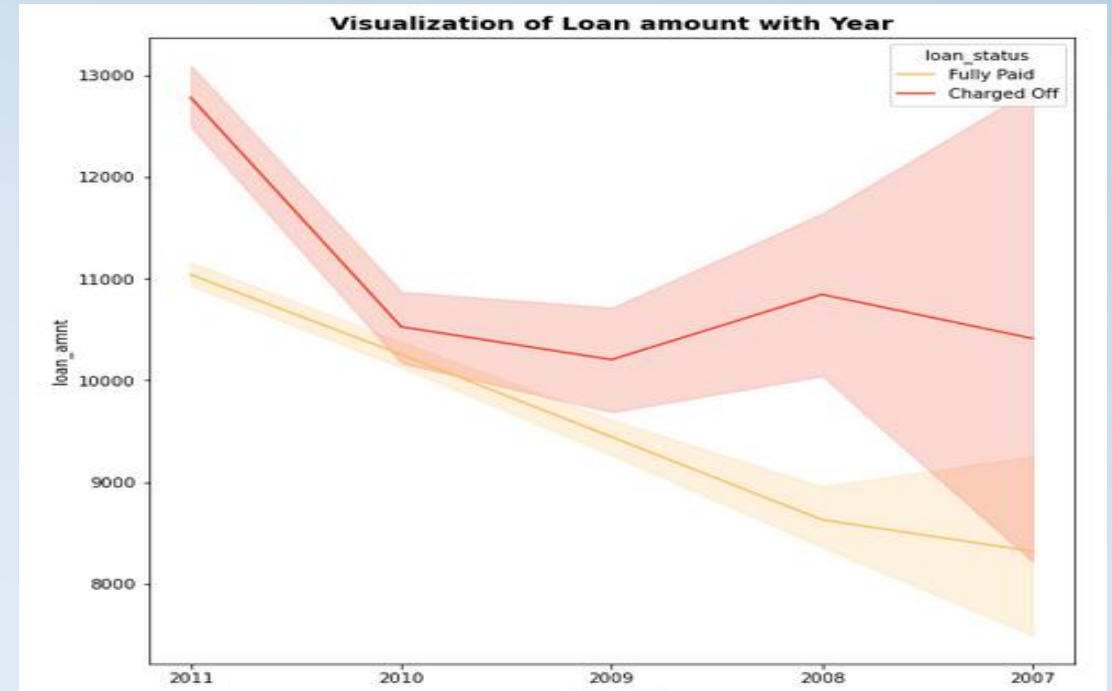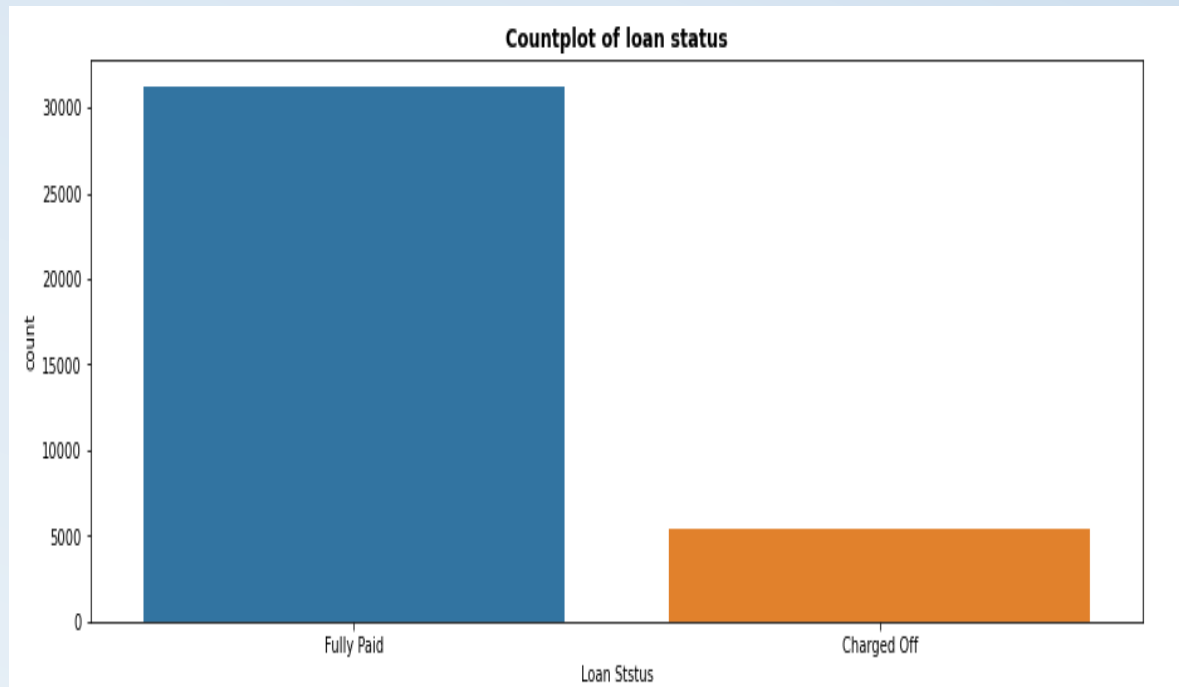
Observations:

1. Most of the loans are Fully Paid.
2. About 15% of loan are having status as defaulters.
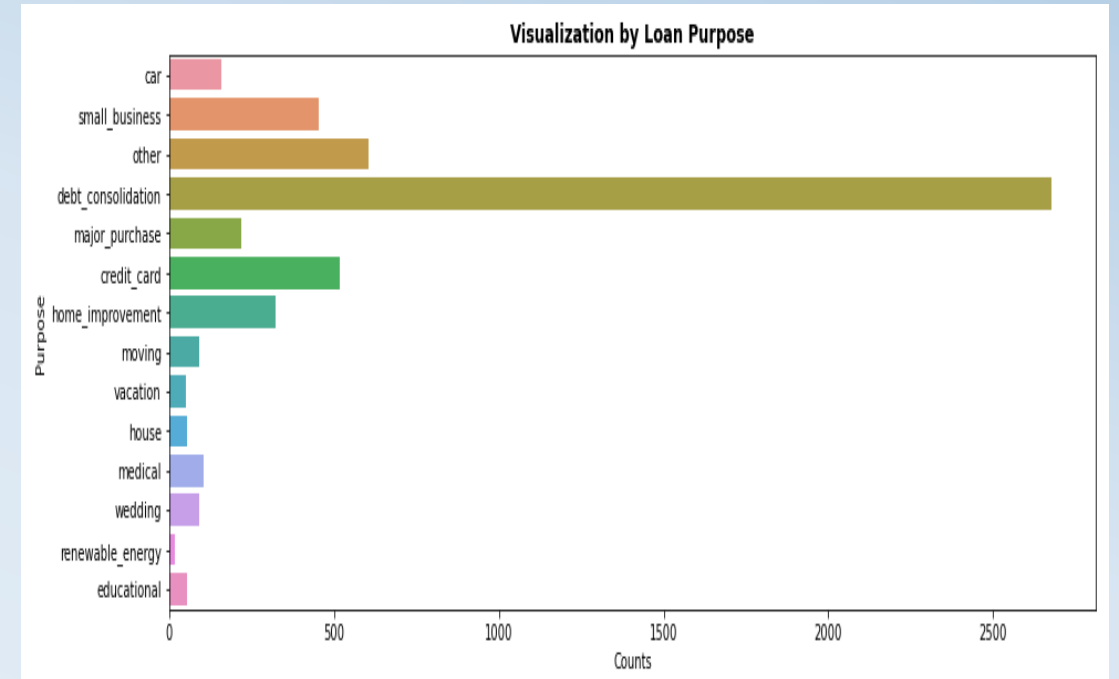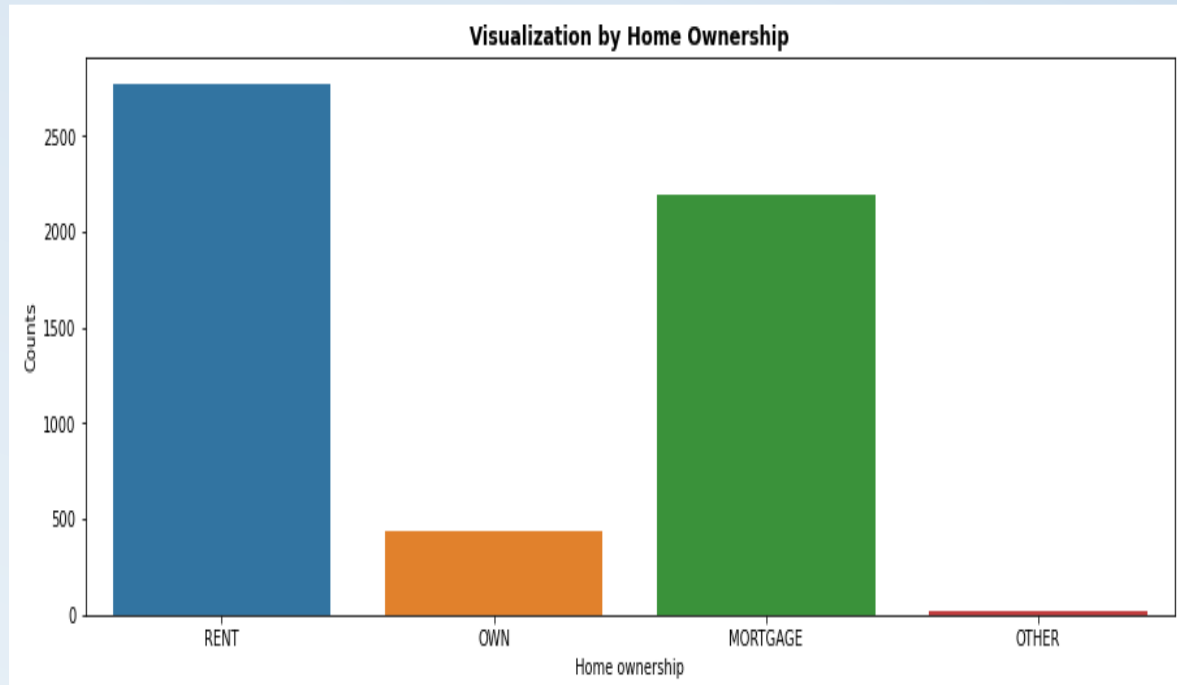3. The loan has been increasing exponentially over the years.

Grade & Sub Grade Analysis

Observations:

1. There is a 'More' probability of defaulting the loan if applicant is taken loan for debt consolidation purpose.
2. There is a "More" probability of defaulting the loan if applicant is belong to 'Grade B' with 'Sub Grade 5 '.

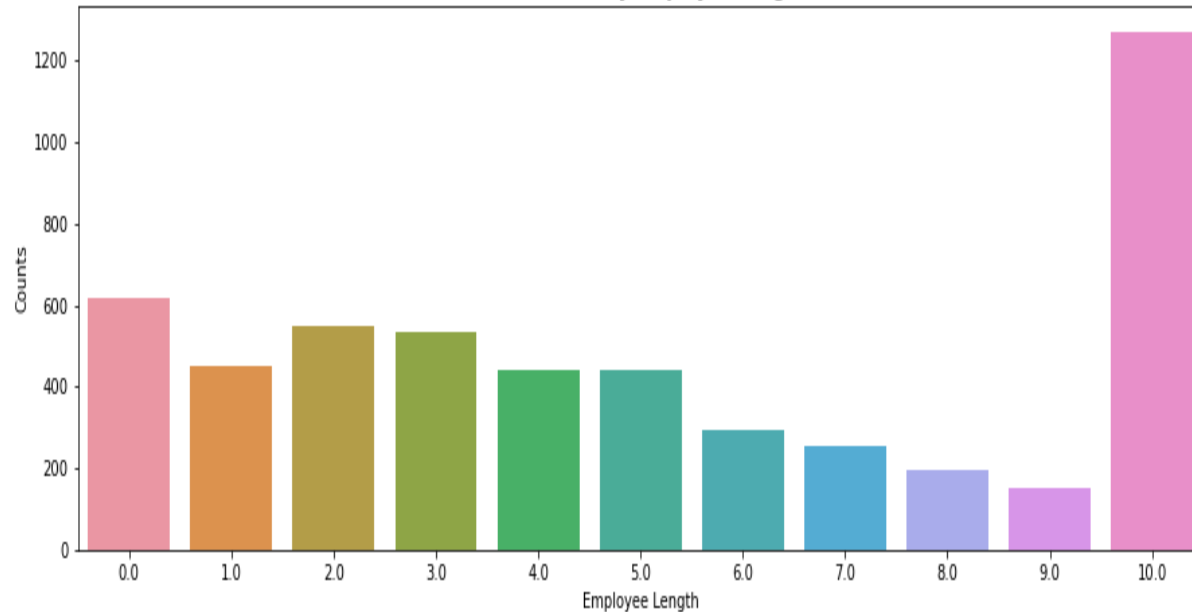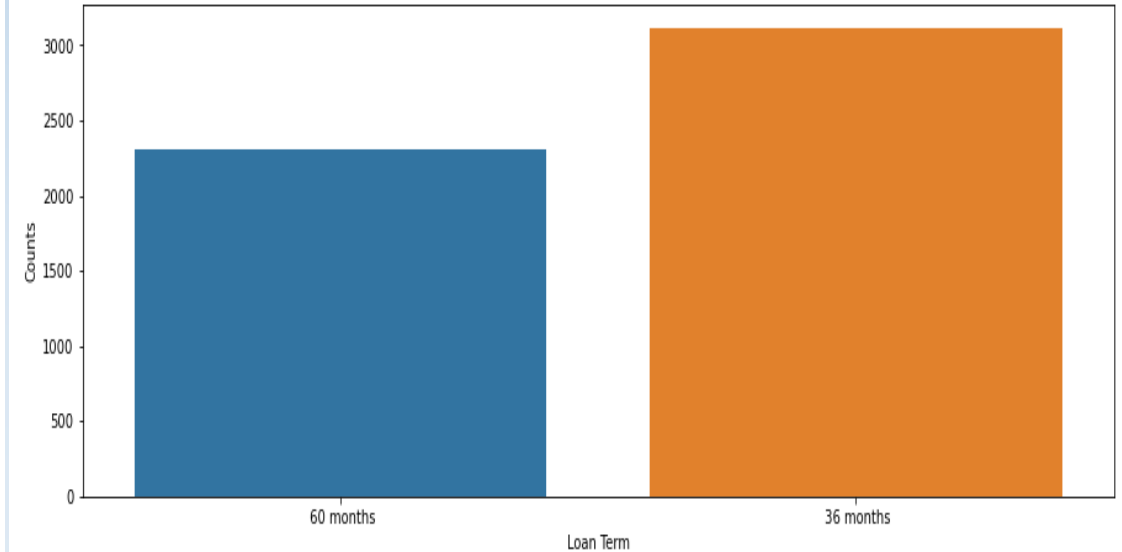# Home Ownership & Loan Purpose Analysis

Observations:

1. Applicant having home ownership is 'Rent' There is a 'More' probability of defaulting loan.
2. There is a 'More' probability of defaulting the loan if applicant is taken loan for debt consolidation purpose.

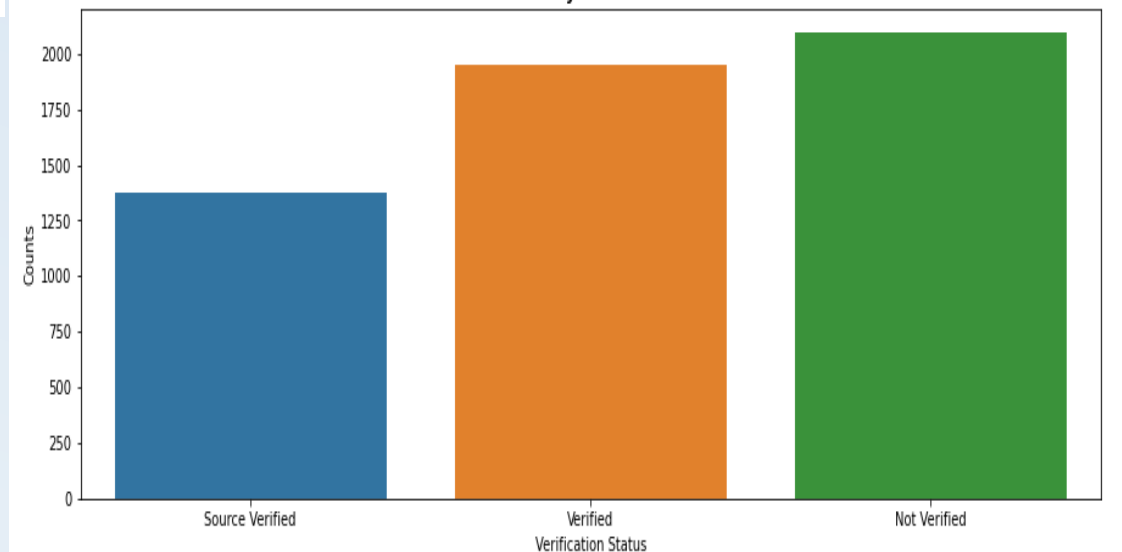# Employee Service length, Loan Term & Verification Status Analysis
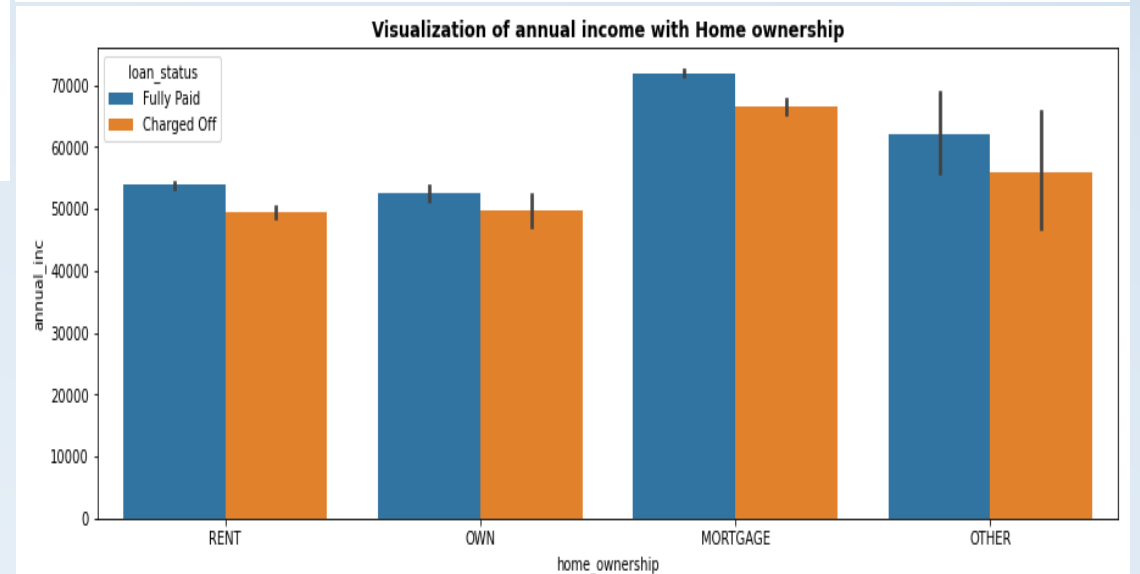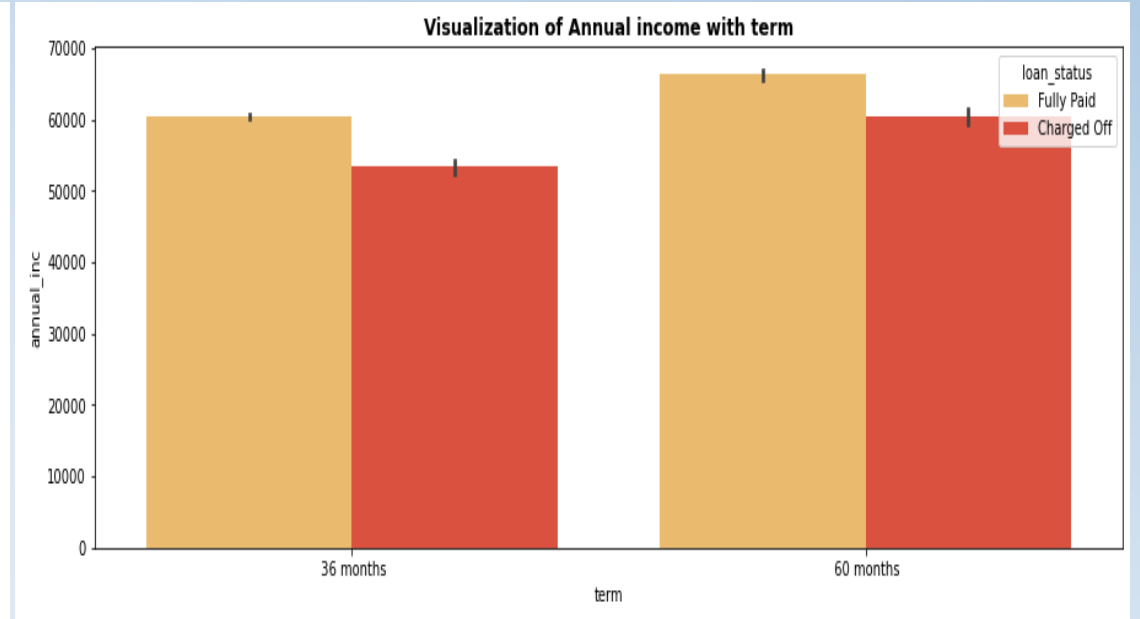
Observations:

1. Applicants with employment length of 10 is more probability of defaulting.
2. Applicants with Loan term of 36 months is more probability of defaulting.
3. There is a 'More' probability of defaulting When the loan status is Not verified.

# Annual income with Loan Purpose, Term & Home ownership

Visualization of Annual income with loan purpose

Visualization of Annual income with term

Visualization of annual income with Home ownership
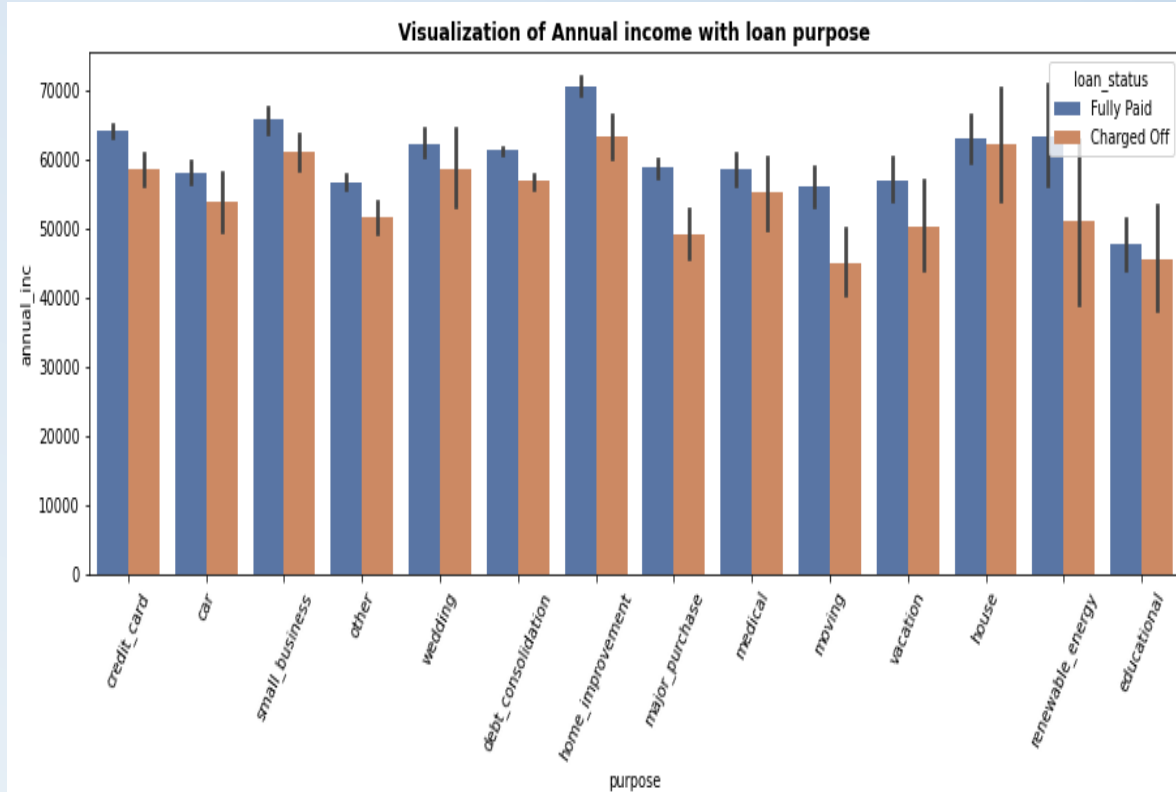
Observations:

1. Applicants more than 60k annual income applied for 'Home improvement' & 'House' loan.

2. The ratio of loan defaulter is highest with annual income>50K and who applied for 'Home improvement', 'House', 'credit card', 'small business', 'wedding' and 'debt consolidation'.

3. Applicants with high annual income taken loan for maximum no of month(Long term).

4. . Ratio of defaulting the loan is higher in long term loan as compared to short term loan and who's annual income is high also.

Annual income with Loan Amount & interest rate
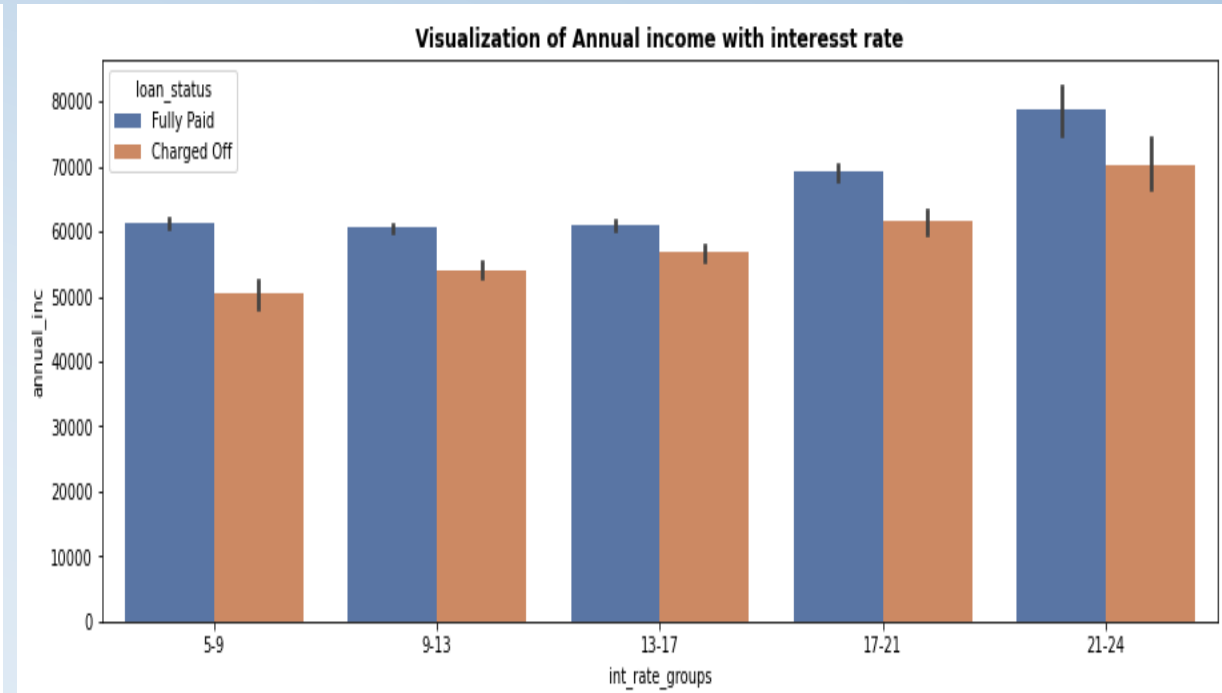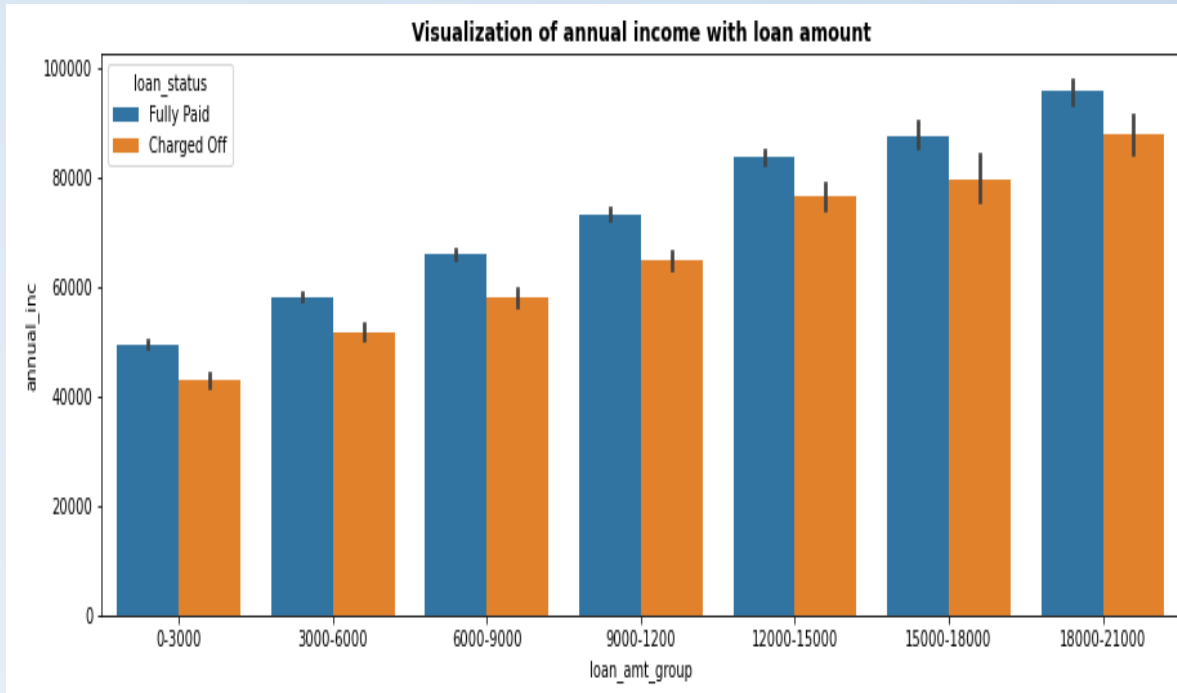
Observations:

1. We can clearly see higher the annual income higher the loan is sanctioned.

2. Maximum loan amount with higher annual income, the defaulting ratio is higher.

3. Applicants who have higher income paying high interest rate.

4. The ratio of defaulting the loan is higher for applicants who have high income with higher interest rate.

Loan Amount with Month and Year
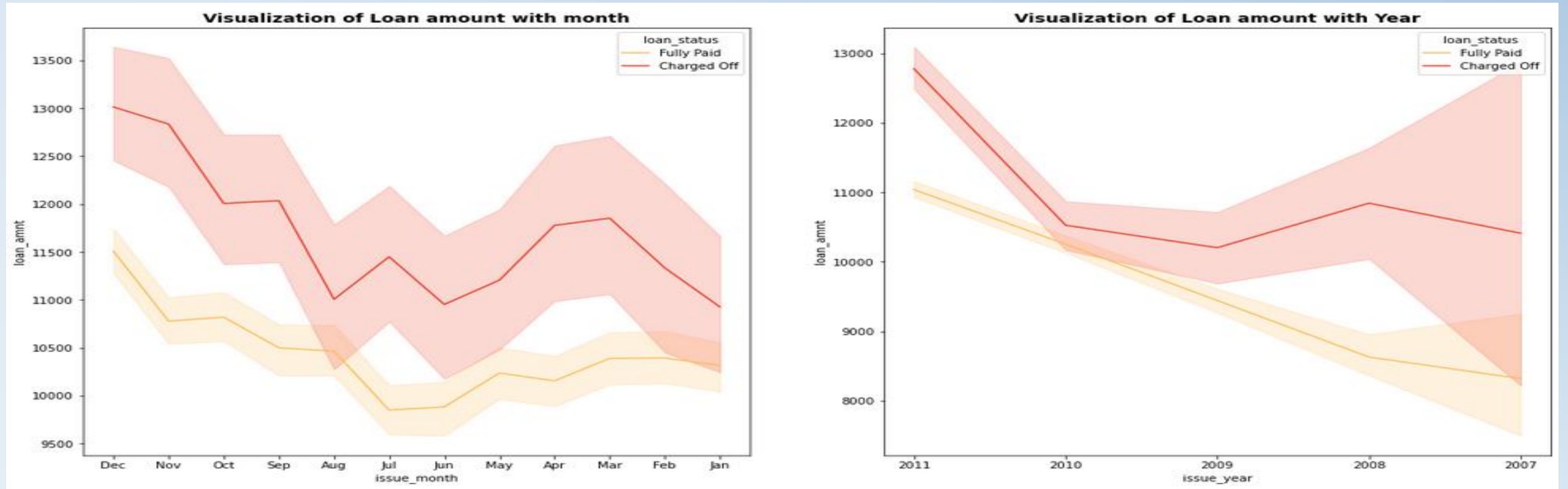
Observations:

1. Maximum number "charged off" loan was sanctioned/issued in December as compared to other month.

2. Loan issued is maximum in the year 2011 as compared to other years.

3. "Charged off amount is also maximum in the year 2011 as compared to other years.

4. we can see December is the month which has the highest number of loan applications per year and also have the highest default ratio. May be people take loan for Christmas party   and then they are not able to pay the amount.

5. We can clearly see maximum no of defaulter in year '2011', this may be due to 'RECESSION' and job loss.

Loan Amount with Purpose & Home Ownership
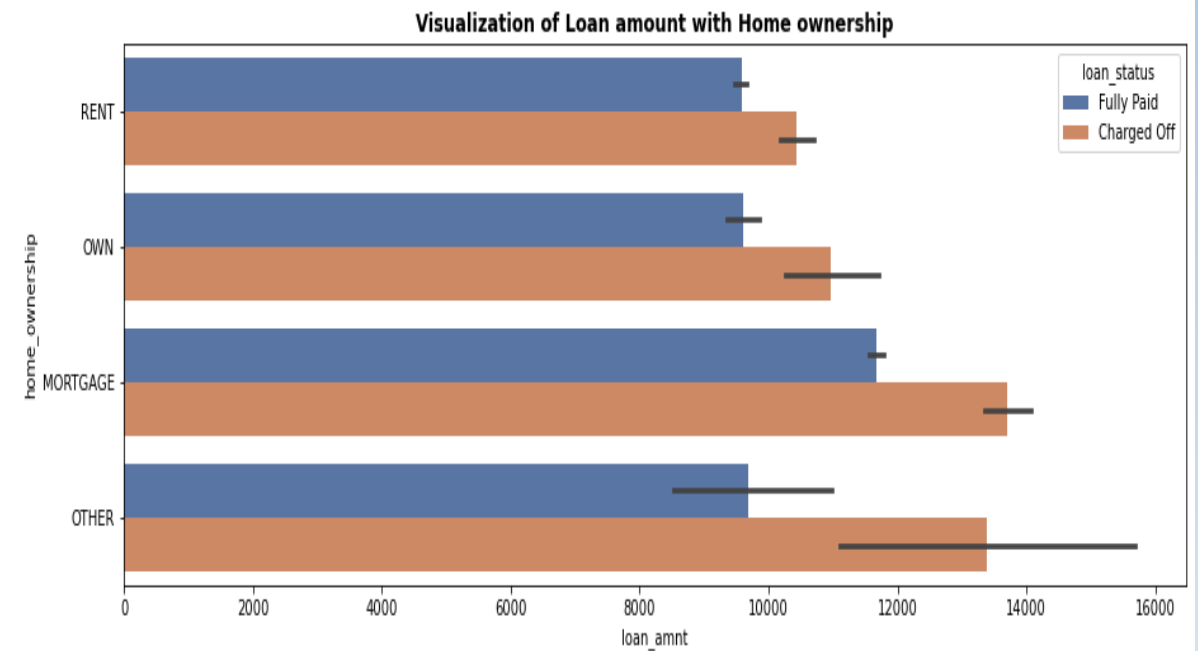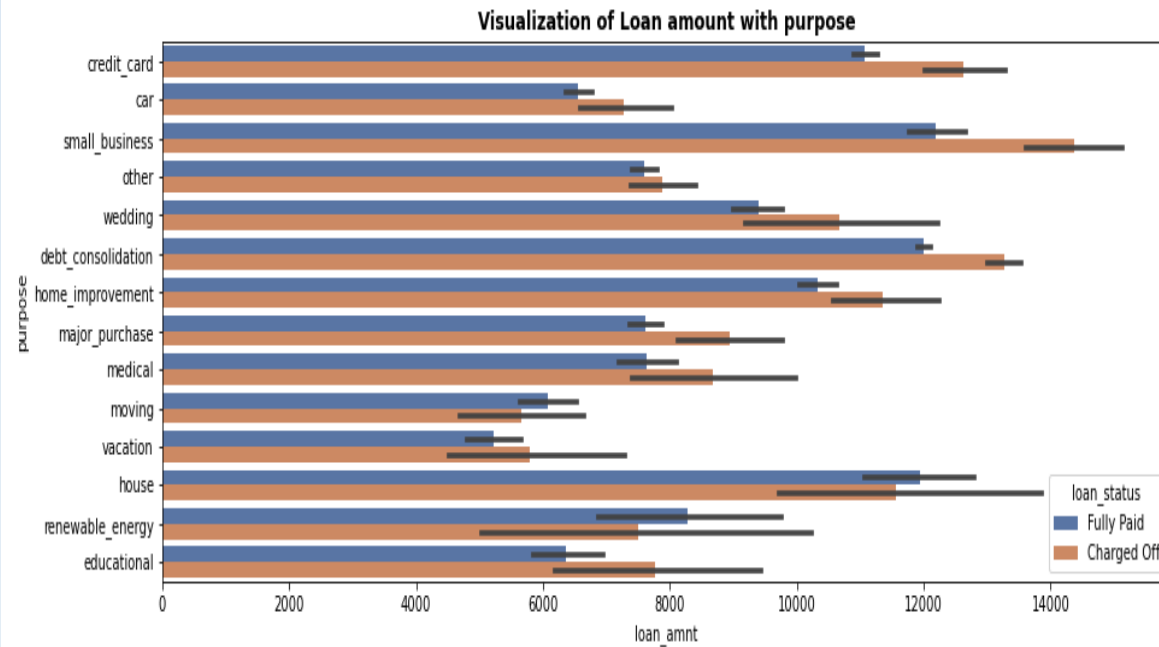
Observations:

1. We can see Maximum loan is taken for 'Small Business' purpose.

2. There are more probability of defaulting whose taken loan for 'small business' purpose.

3. defaulting ratio is highest of 'small business' this may be due to 'shutdown' of business.

4. Applicants whose home ownership is 'MORTGAGE' have taken maximum amount as compared to other.

5. Applicants whose home ownership is 'MORTGAGE' have highest defaulting ratio.

6. Applicants whose home ownership is 'MORTGAGE' have paid maximum loan amounts as compared to other.

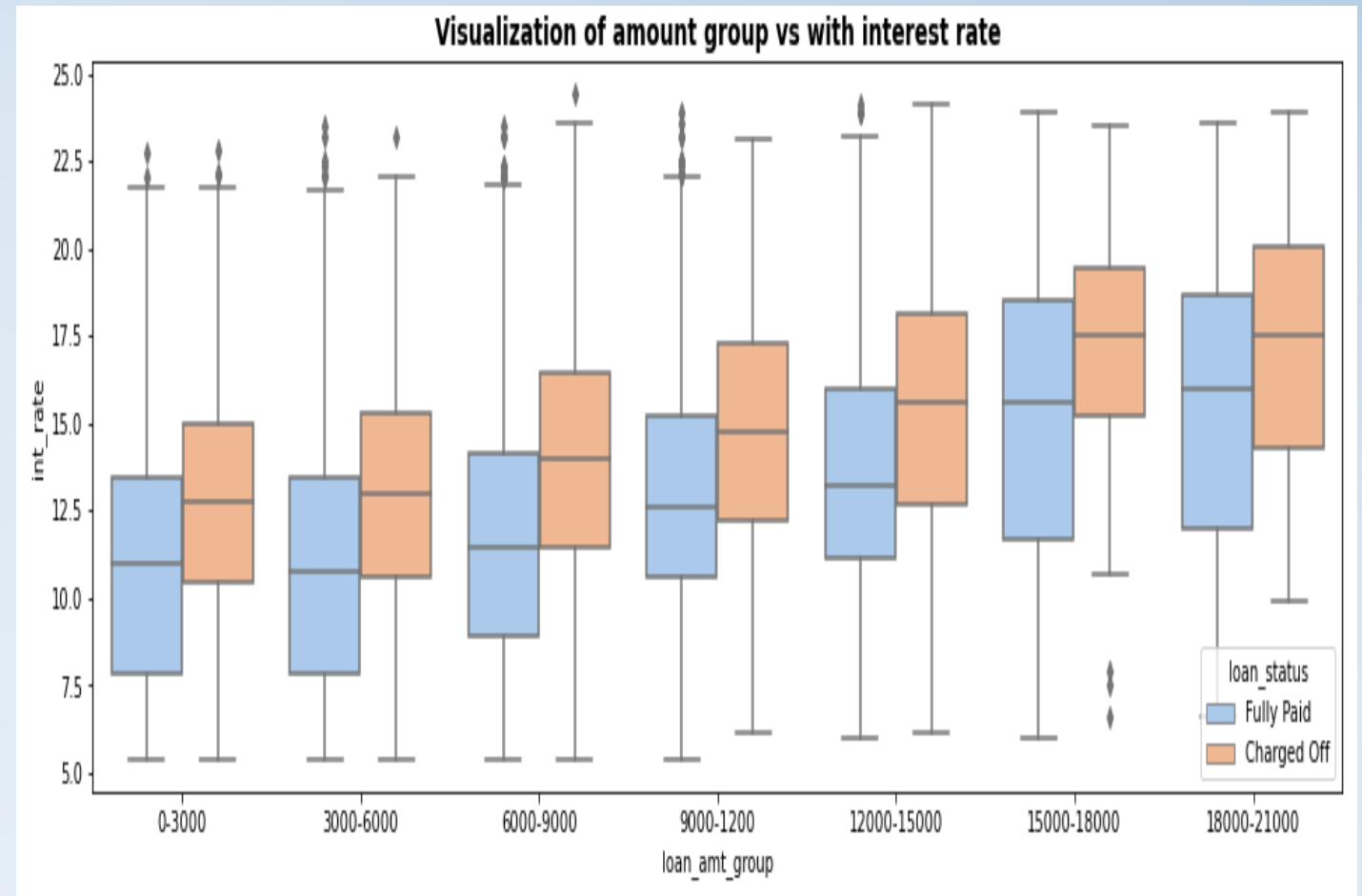# Loan Amount with Grade & Emp service length
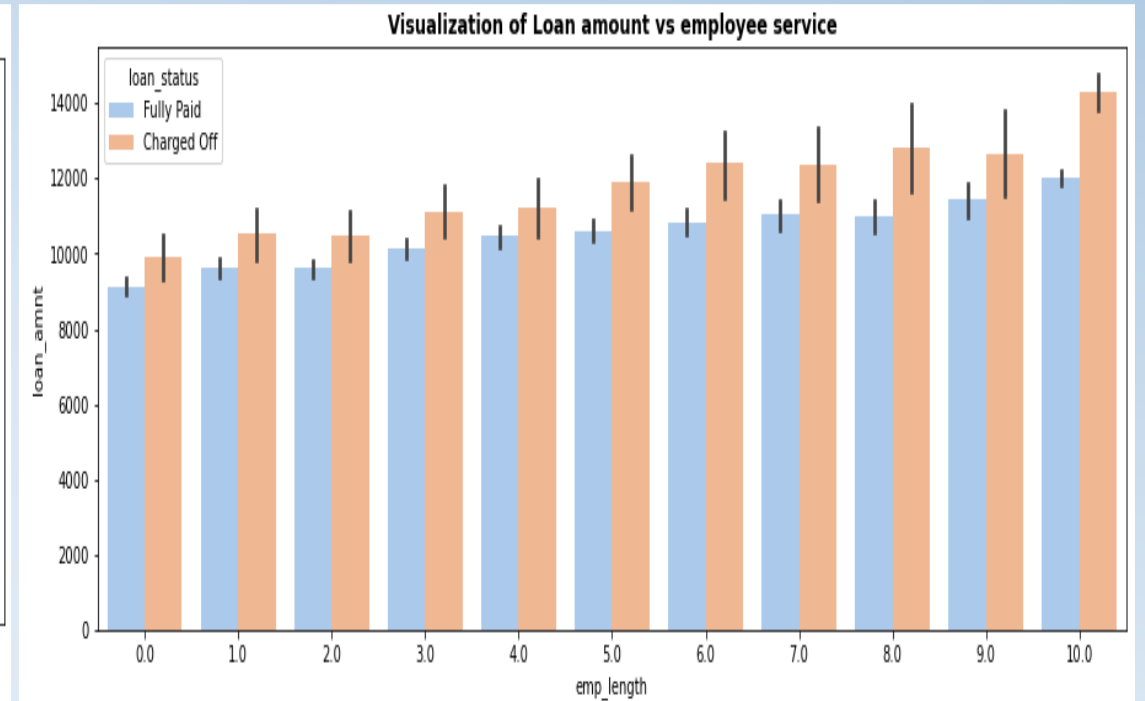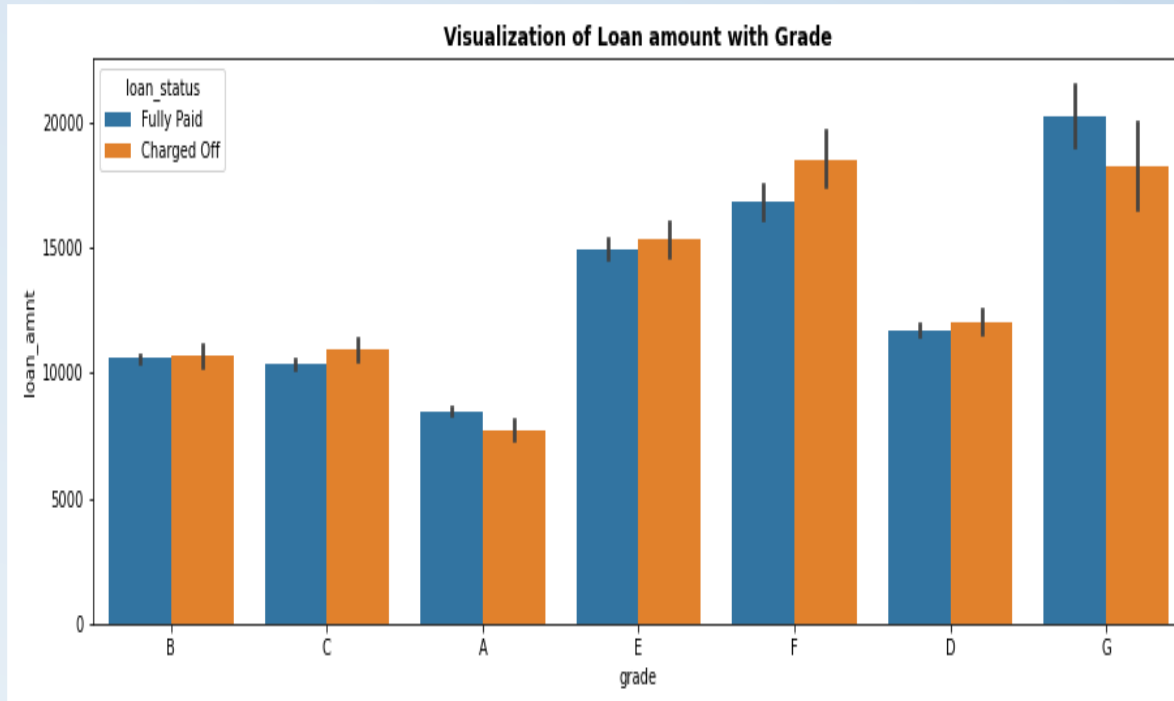
Observations:

1. Applicants with Grade 'G' have take maximum loan amount.
2. Applicants with Grade 'G' have 'fully paid' loan amount is more than 'charged off'.
3. Applicants with grade 'F' have more 'Charged off' loan compared to grade 'G'.
4. Applicants with longer employment length have applied for higher loan amount .
5. Applicants of all employee length group have more probability to get defaulter as compared to 'fully paid' applicants.
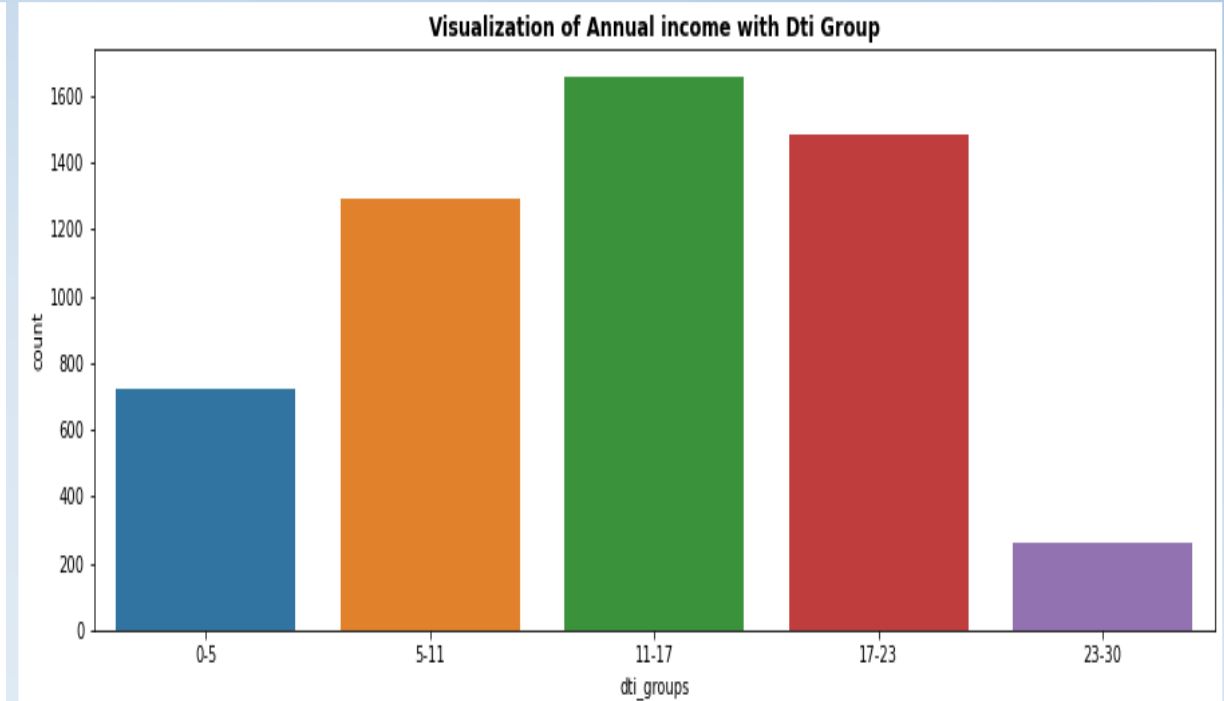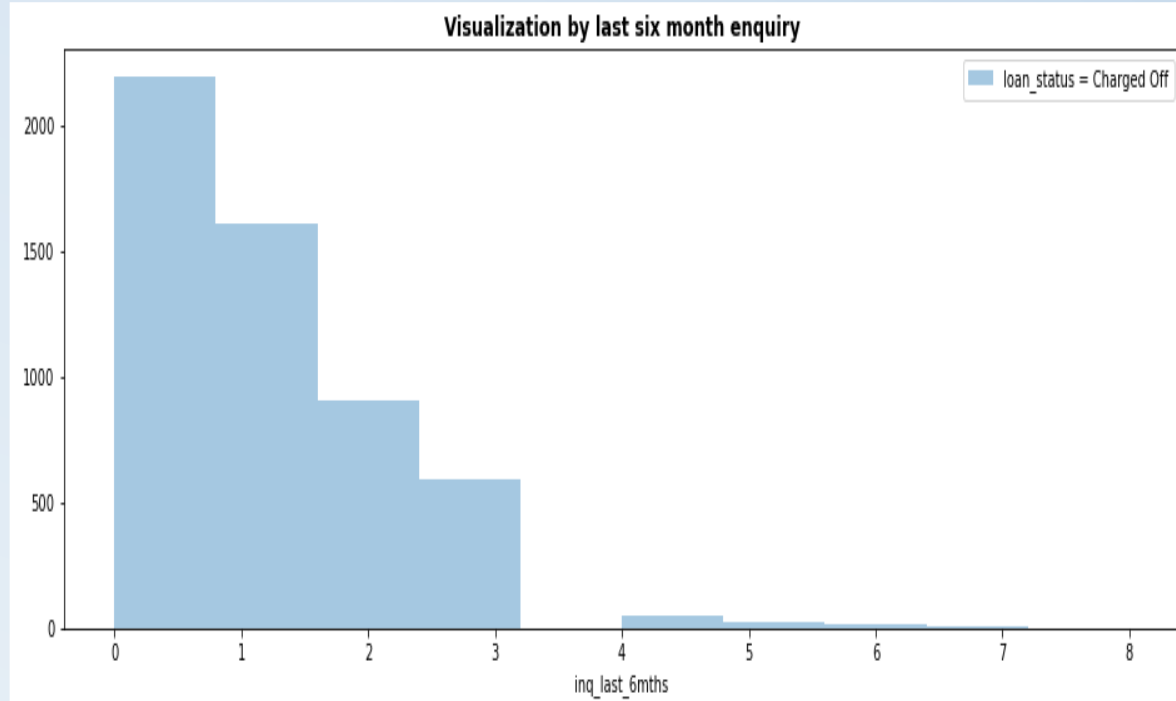
Loan Amount Group with Interest Rate

Observation

1 .The interest rate for charged off loans is high than that of fully paid loans in all the loan_amount groups.

2. This can be a strong factor for loan defaulting.

Visualization of amount group vs with interest rate

Loan Amount with Grade & Emp service length

Observations:

1. There is a 'More' probability of defaulting When the no of enquiries in last 6 months is 0.
2. Maximum number of defaults occurred When Dti is between 11-17.

Conclusion/Results

The above analysis with respect to the Charged off/Bad Loans. So there are more probability of defaulting when :

• Maximum number "charged off" loan was sanctioned/issued in December as compared to other month.

• Loan issued is maximum in the year 2011 as compared to other years.

• "Charged off amount is also maximum in the year 2011 as compared to other years.

• we can see December is the month which has the highest number of loan applications per year and also have the highest default ratio. May be people take loan for Christmas party and then they are not able to pay the amount.

• We can clearly see maximum no of defaulter in year '2011', this may be due to 'RECESSION' and job loss .

• Applicants whose home ownership is 'MORTGAGE' have highest defaulting ratio.

• Applicants of all employee length group have more probability to get defaulter as compared to 'fully paid' applicants.

• The interest rate for charged off loans is high than that of fully paid loans in all the loan amount groups. This can be a strong factor for loan defaulting.

• Ratio of defaulting the loan is higher in long term loan as compared to short term loan and who's annual income is high also.

• The ratio of loan defaulter is highest with annual income>50K and who applied for 'Home improvement', 'House', 'credit card', 'small business', 'wedding' and 'debt consolidation'.

• Maximum number of defaults occurred When monthly installments are between 145-274.