

# LENDING CLUB CASE STUDY

## EXPLORATORY DATAANALYSIS

Group Members

V. Santhosh Kumar

D.Kavitha

# Problem Statement

## Company Profile:

The Lending Club is one of the largest online marketplace for the loans. Consumer can easily apply for the loans in various categories. The loans are funded by Investors based on the Credit ratings of the Customers.

## Problem:

Lot of loans are Defaulted and Charged off. The company wants to understand the driving factors behind loan default. The company can utilise this knowledge for its portfolio and risk assessment to make better decisions.

## Objective:

As a Data Analyst, perform EDA to understand how customer attributes and loan attributes influence the tendency of loans default.

# Analysis Approach

- Data sourcing
- Data cleaning
- Outliers treatment
- Univariate analysis
- Bivariate analysis
- Conclusion

# Data Sourcing

- The available data contains the information about past loan applicants and whether they 'defaulted' or not. The dataset does not contain any information for the loans which were 'NOT Approved'.
- The provided dataset can be analysed using Exploratory Data Analysis (EDA)

## Datasets provided:

### loan.csv

The historical data for the approved loans which have either been fully-paid or Charged Off. It is in comma separated value (csv) file format

### Data\_Dictionary.xlsx

Contains information about the various columns. It is required for Data Understanding. It is in excel format.

# Data Cleaning

Data cleaning involves understanding and resolving quality issues with the data. This is the most time consuming tasks of data analysis.

- Drop columns with all null values/ single categorical value / random values
- Impute null values for important columns
- Convert values into proper data.
- Check distribution and frequencies for numerical and categorical data
- Analyse the variable against segments of other variables
- Create derived Matrices for year and month for date.
- Convert Quantitative values into categories.
- Calculate ratios and percentage and other derived metrics.
- Analyse how 2 variables influence each other.
- Check for the correlations
- Publish insights and observations from the analysis.

# Outlier Treatment

Finding the outliers in numeric columns like

➤ annual\_inc

➤ dti

➤ loan\_amnt

➤ funded\_amnt\_inv

In annual\_inc column observed the presence of outliers. so we can remove those outliers using IQR remove Outliers function returns a lower\_limit and upper limit. The values which are outside of those limits will be treated as outliers

# Univariate Analysis

## Visualizing and analysing the categorical Data

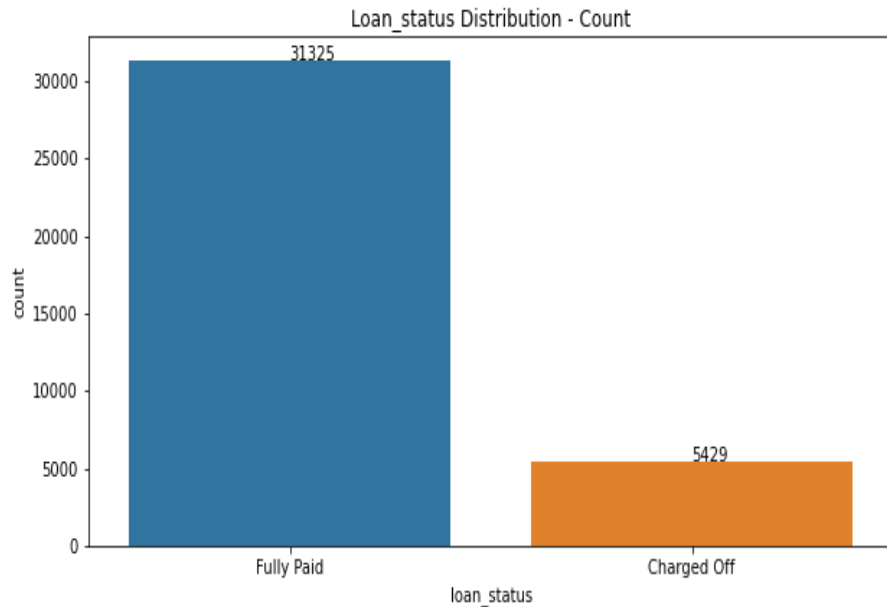
### Observations:

From analysis high chance of defaulting when the following apply loan

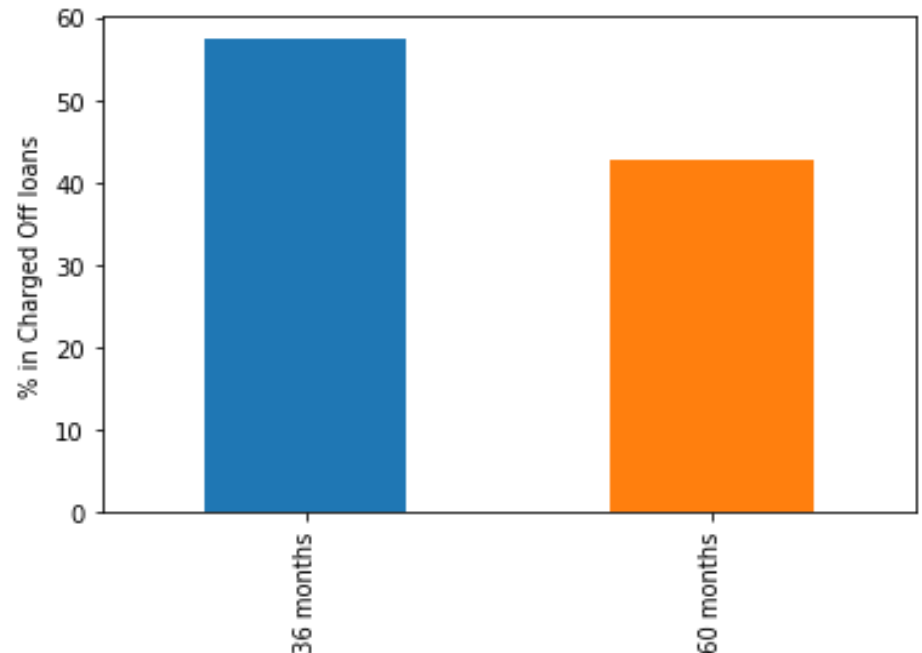
- People with Grade 'B' apply.
- Applications with tenure of '36 months'.
- People who took loan for 'debt\_consolidation'.
- People having house\_ownership as 'RENT'
- Applications with varification\_status as 'Not verified'
- Applications submitted in the month of December and November.
- Applications submitted in the year of '2011'
- Employees with '<1 year experience'( in graph employees with 10+year having higher count. This is because 10+ years will cover lot of applications people having experience in the range of 10-50 will also fall under this category. So we can assume <1 year experience having higher risk of defaulting)

# Sample Results

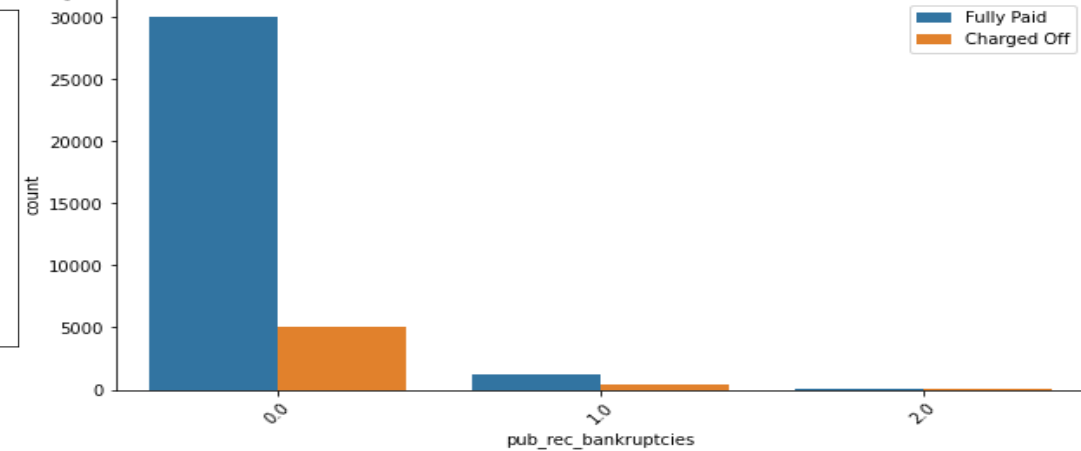
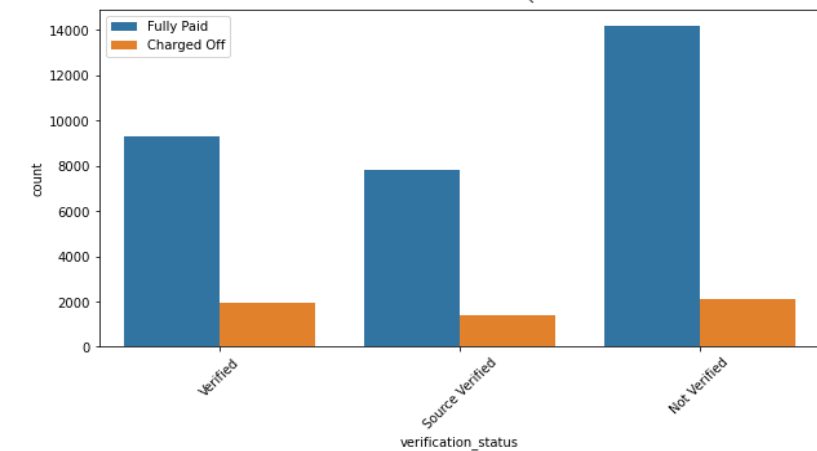
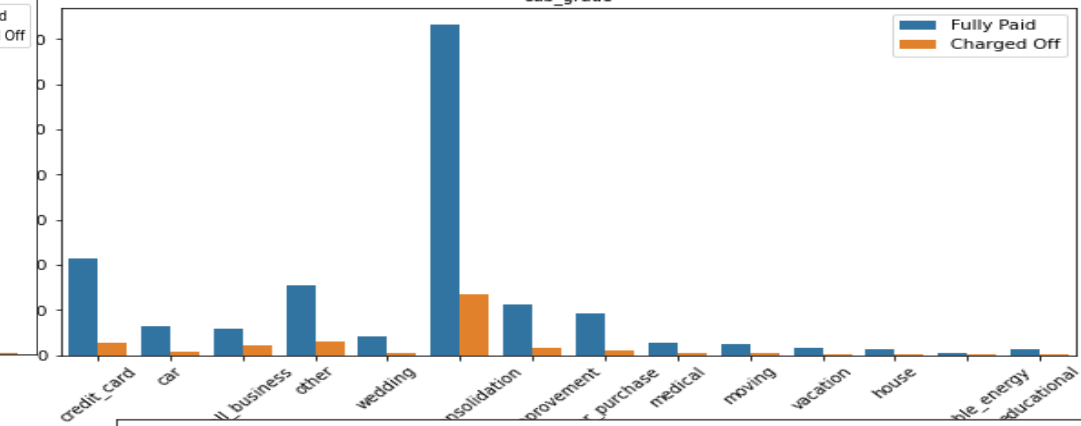
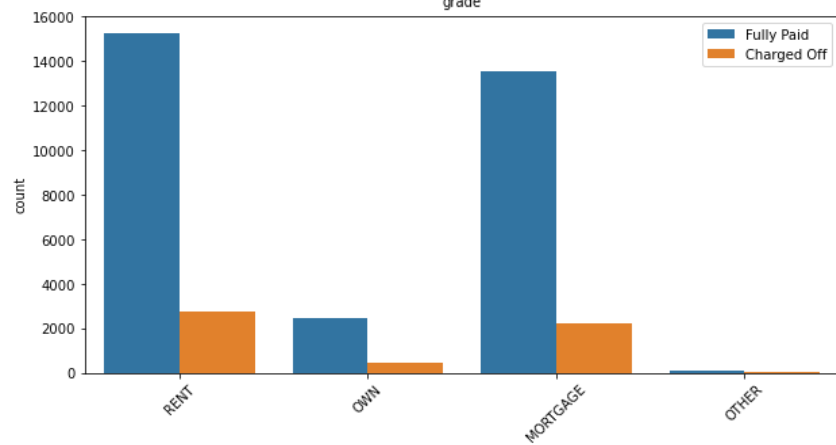
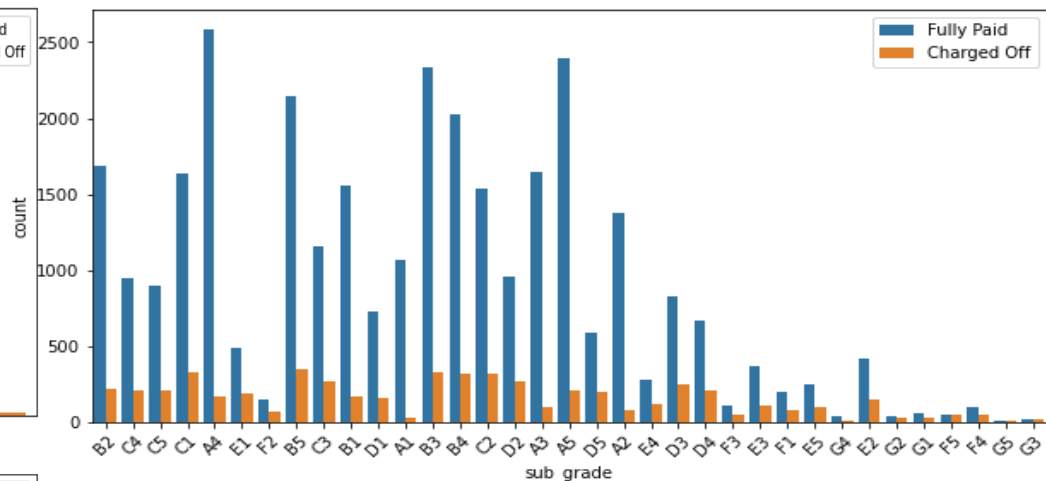
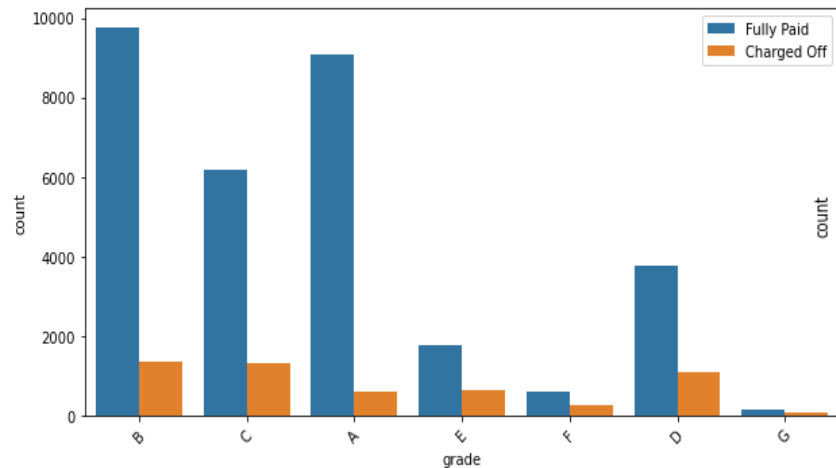
- ❖ In the given data 14.77 % of Loans( 5429 loans) are Defaulted

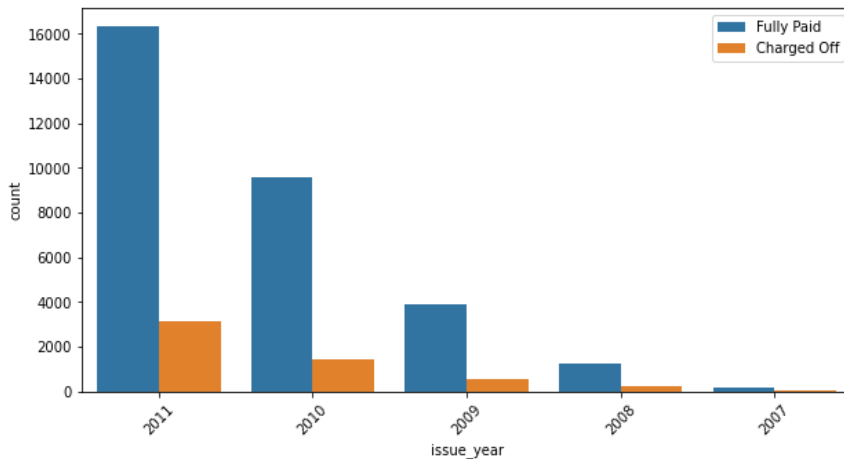
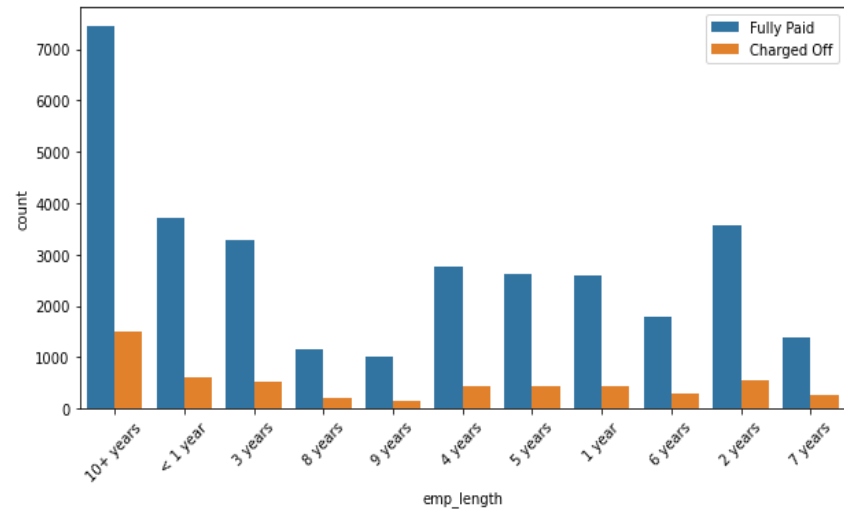
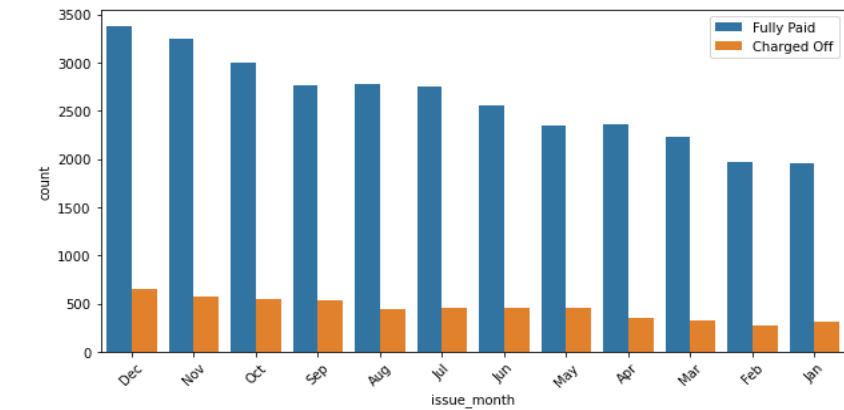


- ❖ Around 75% of the total loans are given for duration of 3 years. while just 25% of the loans are those given for 5 years.
- ❖ Among Charged Off loans, percentage of term 60 months rises to 45%. The higher term loans have a higher chance of default





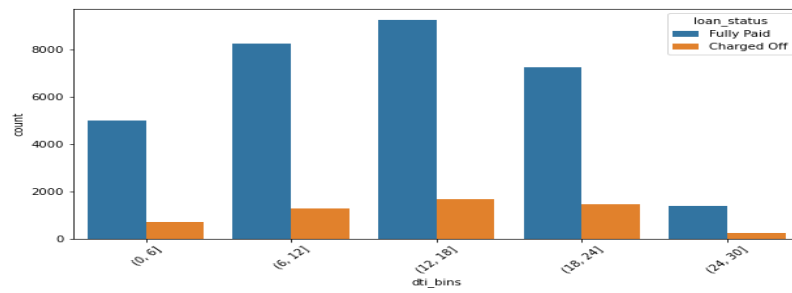
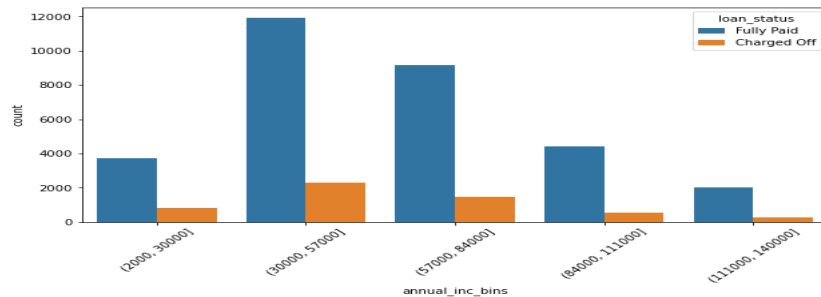
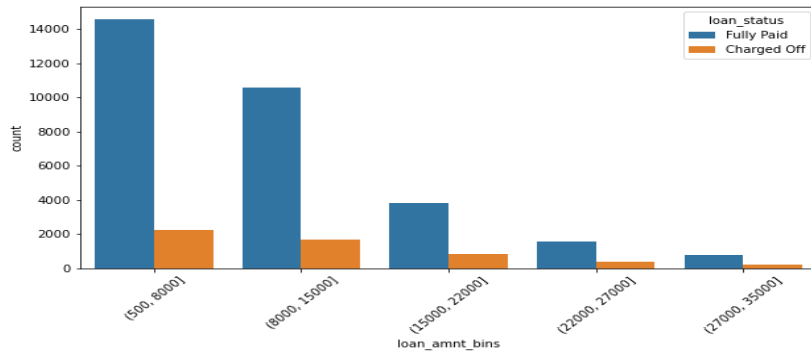
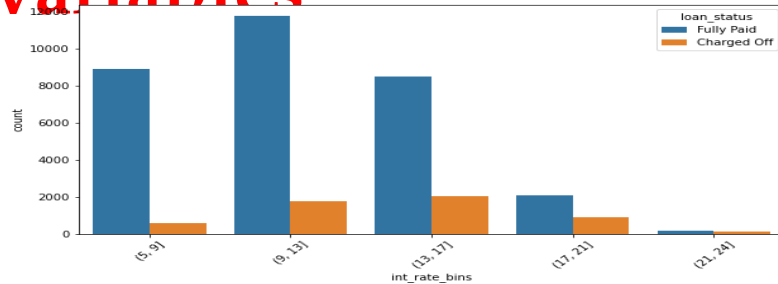




## Low chance of defaulting when :

- ✓ People with Grade 'G' apply.
- ✓ Applications with tenure of '60 months'.
- ✓ People who took loan for 'renewable\_energy','vacation','educational','house','wedding','moving'.
- ✓ People having house\_ownership as 'OWN'.
- ✓ Applications with varification\_status as 'Source verified'
- ✓ Applications submitted in the month of 'February 'and 'January'
- ✓ Applications submitted in the year of '2007'
- ✓ Applications of Employees with '8-9 years experience'

# Visualizing and analysing Continuous/Numeric variables



## Observations:

High chance of defaulting in case of

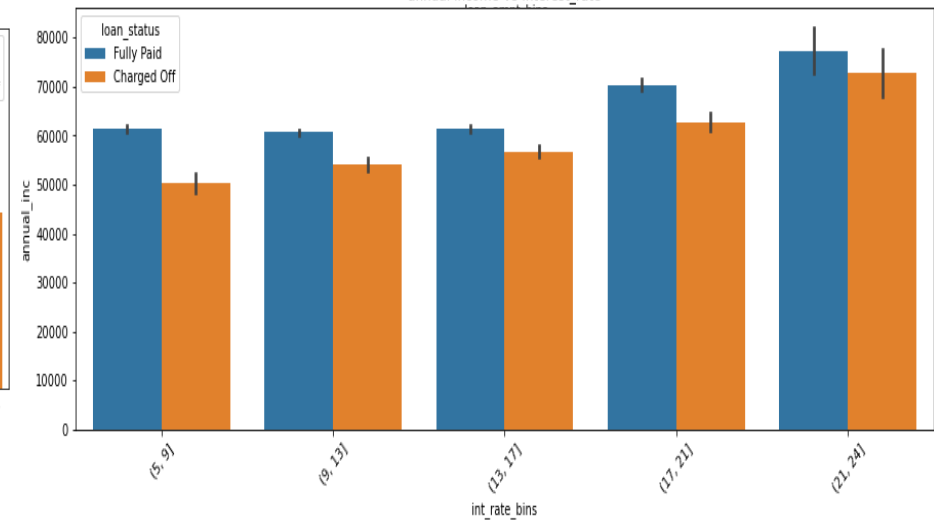
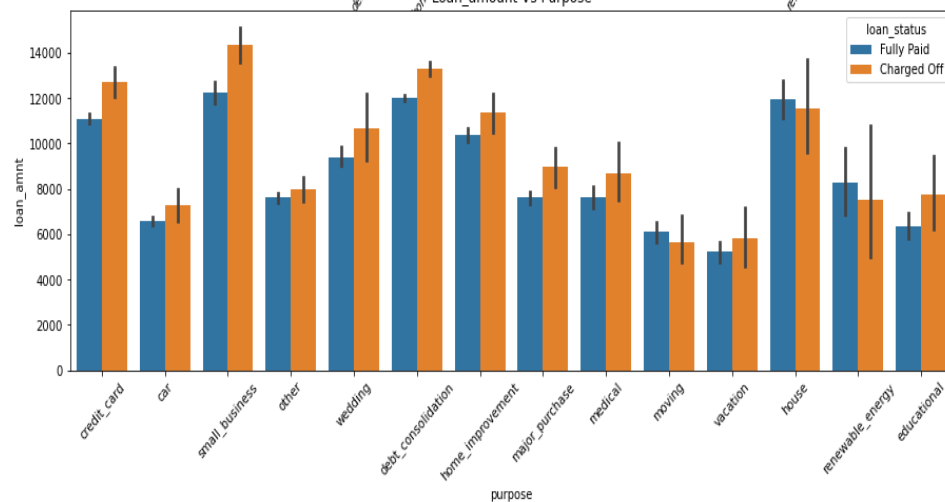
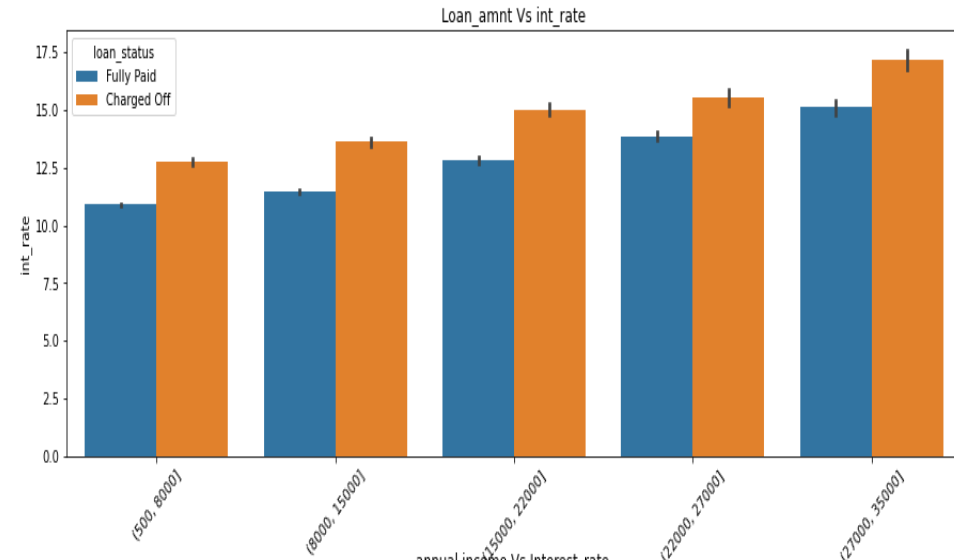
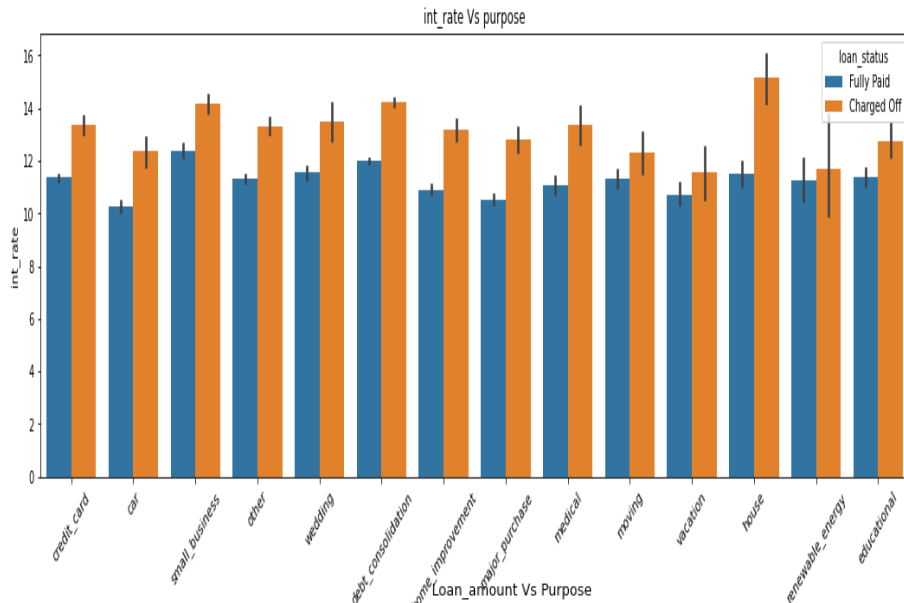
- ✓ Loans with interest rate of 13-17%
- ✓ Applicants who have an income of range 31000- 58000
- ✓ Loan amount in the range of 500 – 14000
- ✓ Dti is between 12-18

Low chance of defaulting when :

- ✓ Applicants who have an income of range 112000- 140000
- ✓ Loan amount in the range of 28000 – 35000
- ✓ Dti is between 24-30

# Bivariate Analysis

## Sample results



## Observations:

From analysis There is a high chance of defaulting when :

- ✓ People who took loan for house and having interest\_rate between 14-16%
- ✓ People who took loan for debt\_consolidation,small\_business and having interest\_rate between 12-14%
- ✓ Loan amount in the range 28k - 35k and interest rate at 15-17.5 %
- ✓ Loan for small business and the loan amount in between 14-16k
- ✓ Loan for debt\_consolidation and the loan amount in between 12-14k
- ✓ Having annual income between 70-80k and interest rate at 21-24%
- ✓ Loan status as verified and loan\_amount between 14k - 16k plus
- ✓ Home ownership is 'MORTGAGE' and loan amount between 12-14k

# Conclusions

- The % of charged off loans increases substantially as we go up the loan amount buckets. Most loans are below 20000 amount. The higher loans, though lesser in number, carry a substantially higher risk of default.
- Around 75% of the total loans are given for duration of 3 years. while just 25% of the loans are those given for 5 years. Among Charged Off loans, percentage of term 60 months rises to 45%. The higher term loans have a higher chance of default
- 26% of loans for small business are Charged Off. Making them the most risky purpose. Approximately ~49% of the loans are issued for the purpose of dept consolidation. 17% of the loans for renewable\_enrgy are charged Off, but the number is too less to be of significance.
- The percentage of Charged Off loans is more when the borrower has a prior record of bankruptcy.
- Overall, the intrest rate varies from 5.42% to 24.4% with average intrest rate of 11.8%. The intrest rate for Charged Off loans appear to be higher than for Fully paid. This is naturally expected. As, the risk increases the rate of interest imposed on the loan also increases
- As the intrest rate increases there is a higher chances of defaulting the loan
- the percentage of loans issued in the category of E,F,G are very less but percentage charged

- There is no impact of employment length to chance of loan default.
- There is no impact of home\_ownership to default on loan.
- Verified loans show more charged Off percentage as compared to Not verified loans
- When the dti is higher than 20, higher percentage of loans are Charged Off Higher the dti higher the chances of loan being Charged Off
- The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years values ranges between 0 and 11. There are higher chances of default if this variable is increasing from 1
- Higher loan amounts are Verified more often. We already know that larger loans are less in number, but see a higher charge off rate. This, combined with previous observation, explains why verified loans see a higher rate of default. It's not the verified status per se, it's the fact that higher loan amounts are riskier and are also verified more often by Lending Club.
- Higher loan amount are associated with lower grade for longer terms.
- revol\_util and grade(and therefore int\_rate) are correlated in some way. The revol\_util is positively correlated to the grade. As the grade goes from A to E the revol\_util also increases. This may be because higher loan amounts are associated with higher grades.
- loan charged off has more correlation with interest rate and then with revol\_util means higher the interest rate more the chances for charged off

*Thank You*