Lending Club Case Study

Irfan Syed & Jereme Wilsander

Business Understanding

- 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.

- The data given in dataset loan contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Business Objective

- To understand the **driving factors (or driver variables) behind loan default**, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Analysis Approach

- Data Understanding
  - Check the dataset for the nature of Data
- Data Cleaning
  - Remove unwanted data that might influence our analysis
- Derived Variables
  - Derive useful information from existing data
- Univariate Analysis
  - Analyze Individual variables to get insights
- Bivariate Analysis
  - Analyze the relationship between 2 variables

Data Understanding

- Load the given 'loan.csv' to python data frame and check the data
  - It has 39717 rows and 111 columns
- Observations
  - We have quite a few columns that has ONLY NULL values
  - We have some columns that has only unique values
  - We have some columns calculated on customer behaviour and won't be useful for our analysis
  - We have some columns whose data are got after the loan approval process
  - We have some columns that has free text or links that doesn't help in our analysis
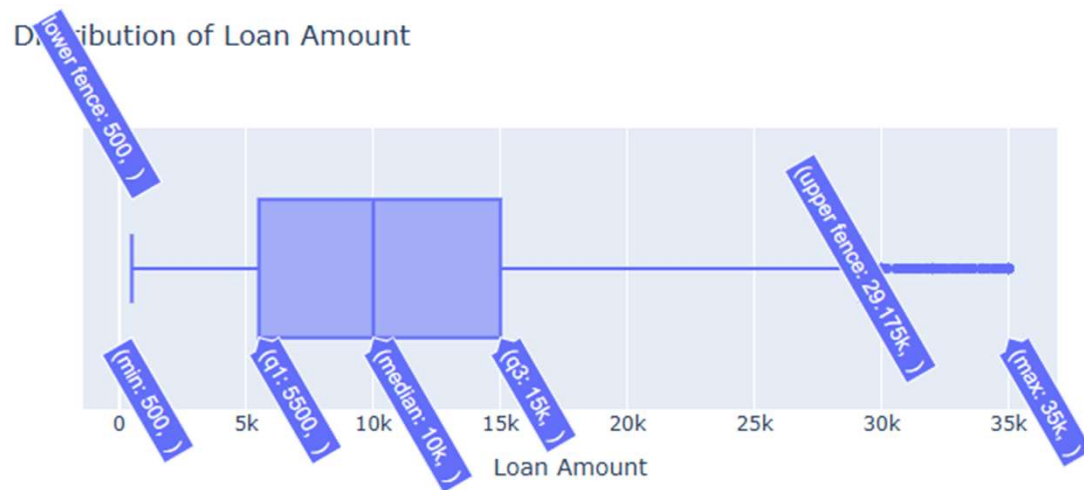
Data Cleaning

- Following steps are performed as part of data cleaning
  - Drop invalid/unnecessary columns
  - Fix/Filter rows
  - Missing Values
  - Standardize Numbers
  - Standardize Text
  - Fix Invalid Values
  - Filter Data

Derived Variables

- Derive month and year from issue_d column

Distribution of Loan Amount

Most of the customers got loan amount sanctioned between 5k-15k

Distribution of Interest Rate

Most of the loans have interest rate between 9%-15%

Trend of Annual Income

Trend of Annual Income

Approximately 80% of customers who availed loan had income less than 100k

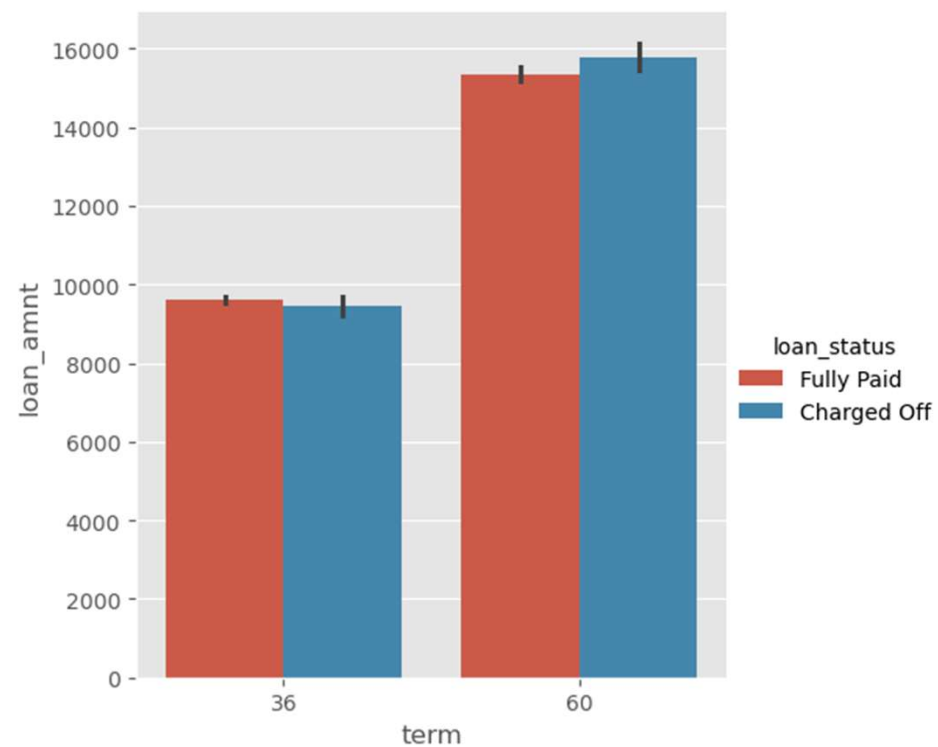# Distribution of Debt To Income Ratio

Most of the customers has debt to income ratio between 8 - 20
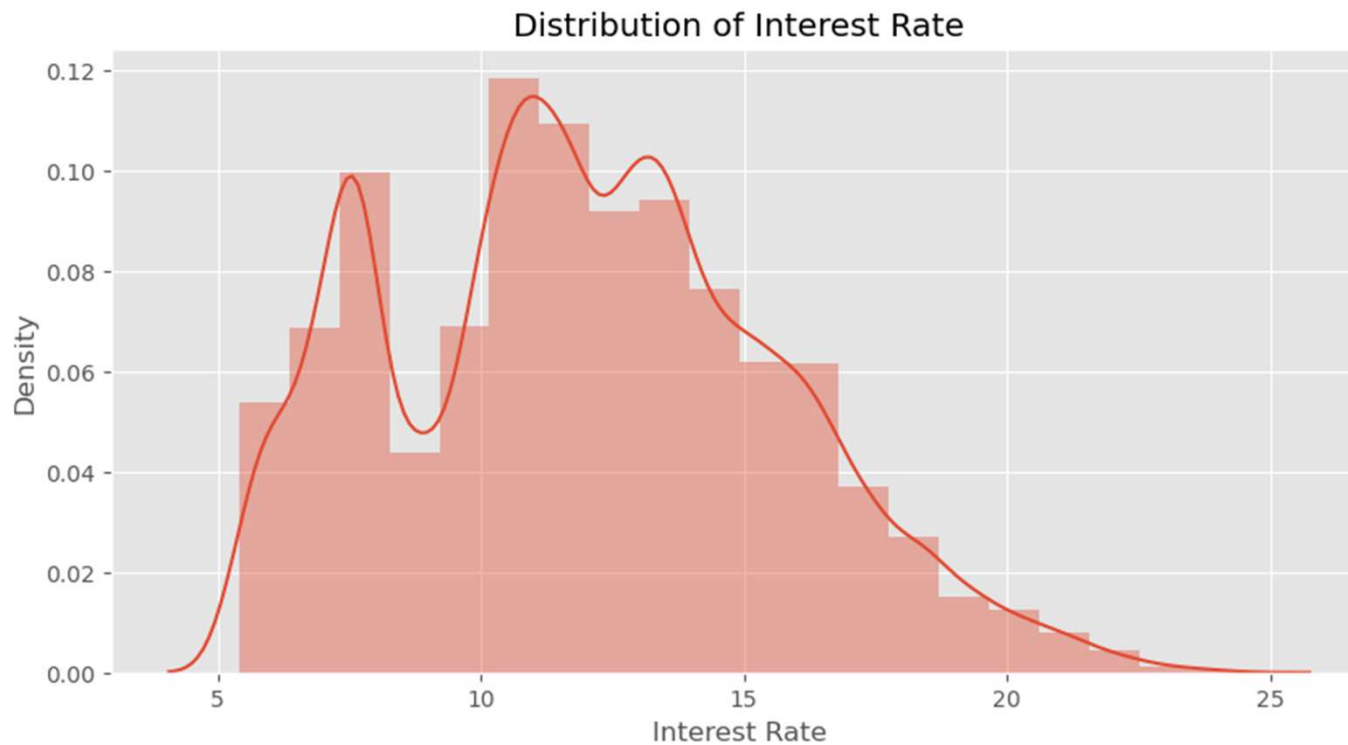
Loan status
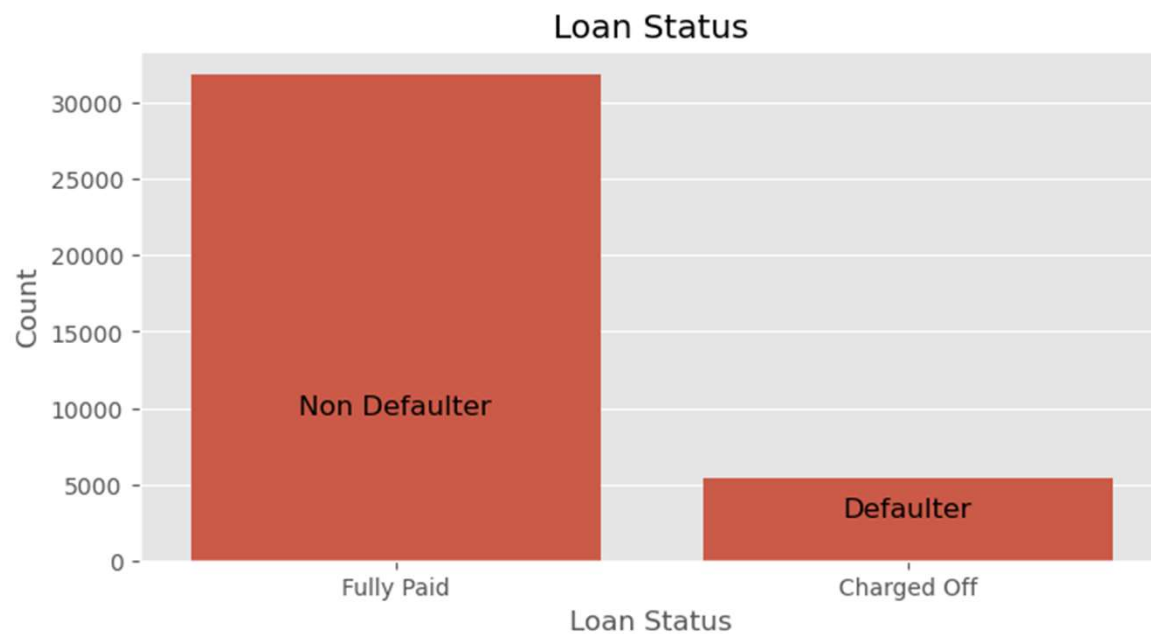
Approximately 13% of total customers defaulted their loans
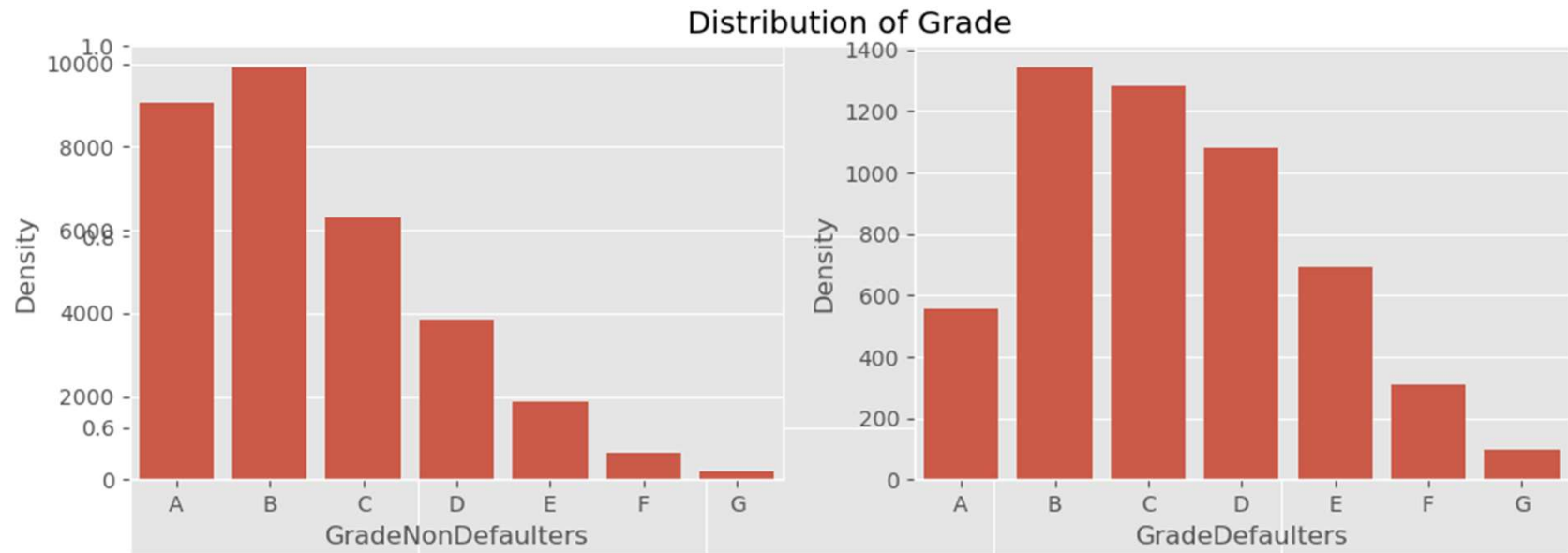
Term of loan

People who took loan for higher term tend to default more
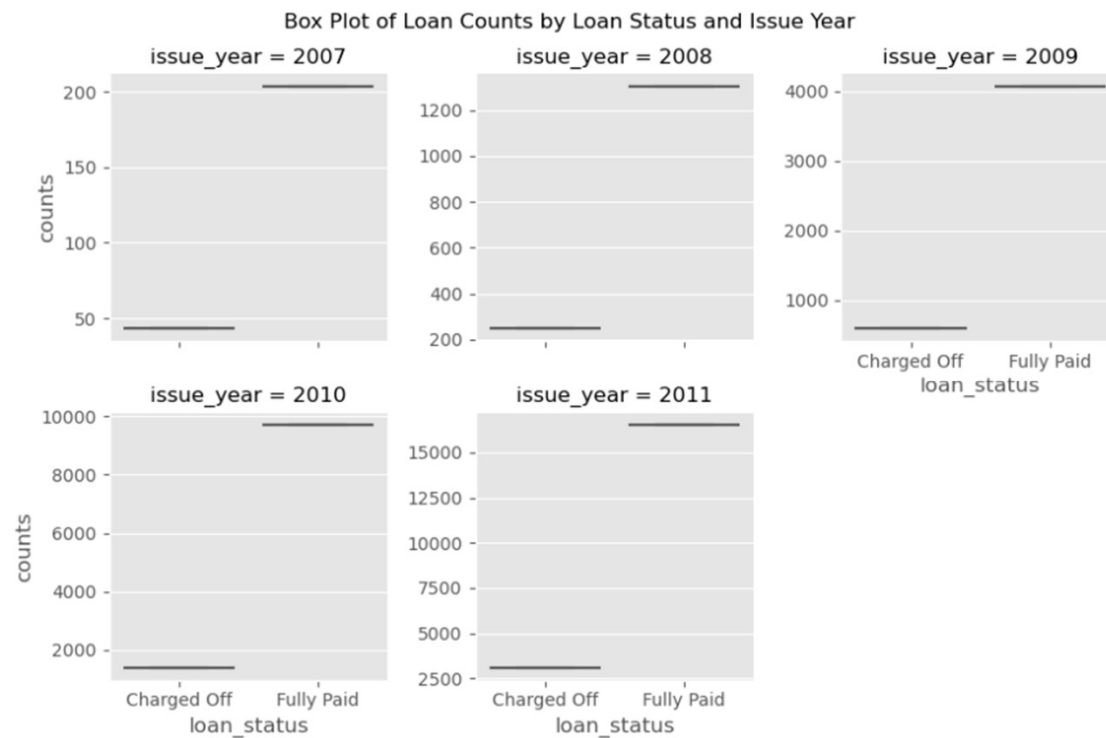
Distribution of interest rate

Most of the customers got interest rate between 10-14% and a
considerable amount of customers also got around 8%

Distribution of Grade

People in Grade 'A' default very less compared to other grades

Distribution of Loan Status For Issue Year

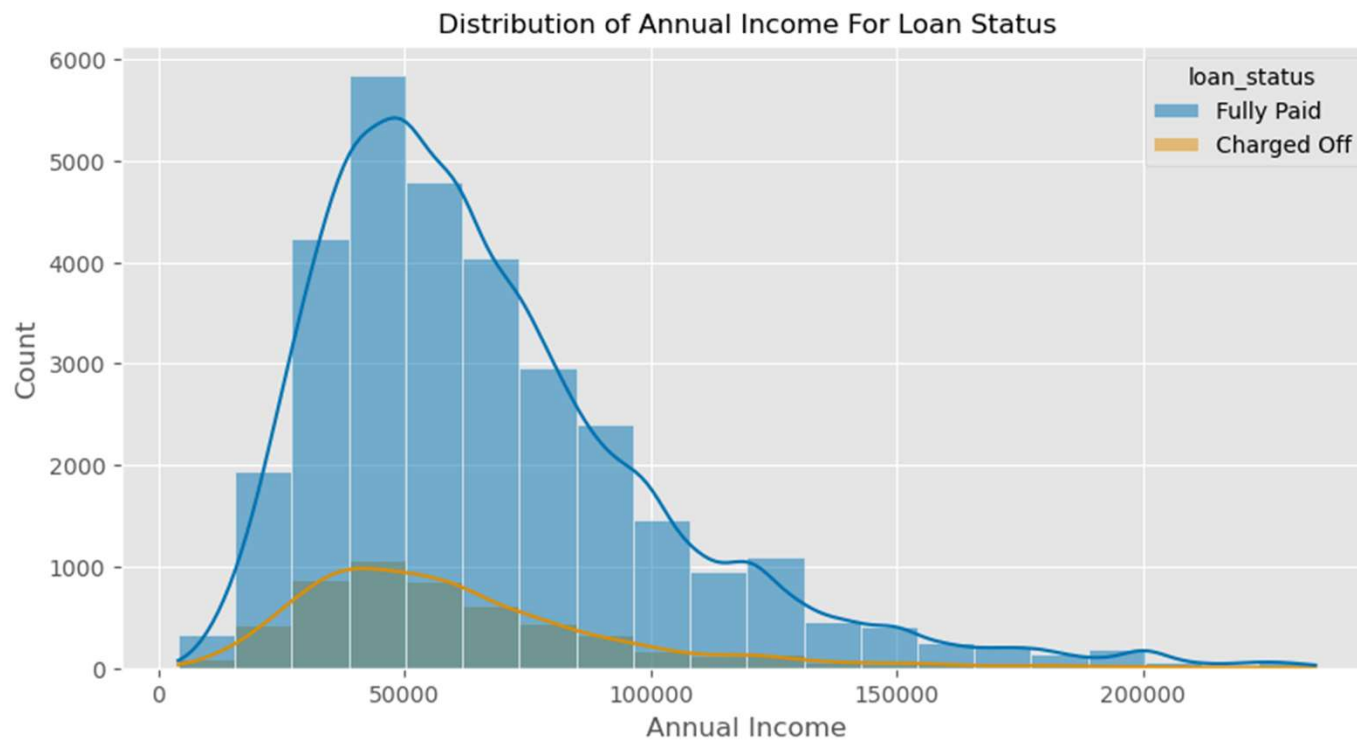Box Plot of Loan Counts by Loan Status and Issue Year

Volume of loans has vastly increased with each year, also indicates approvals have increased

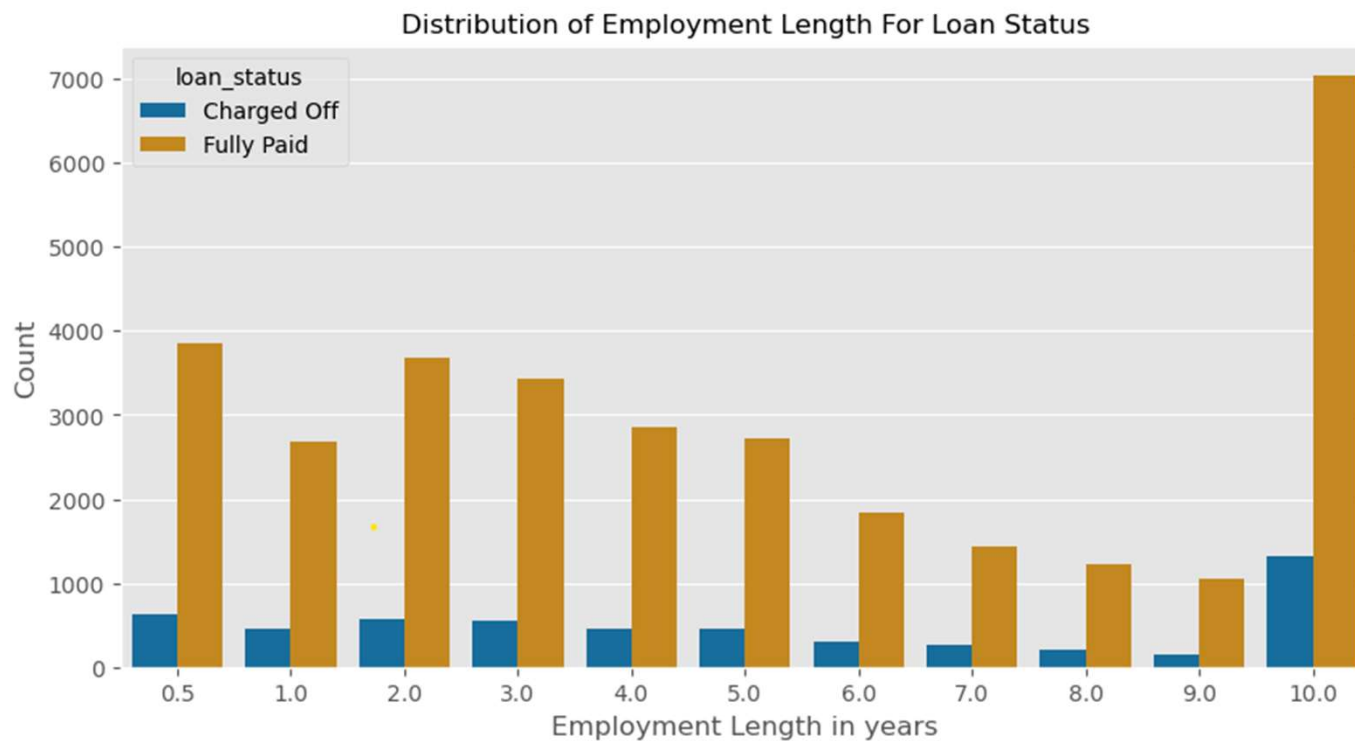# Distribution of defaulted loans across year of lending
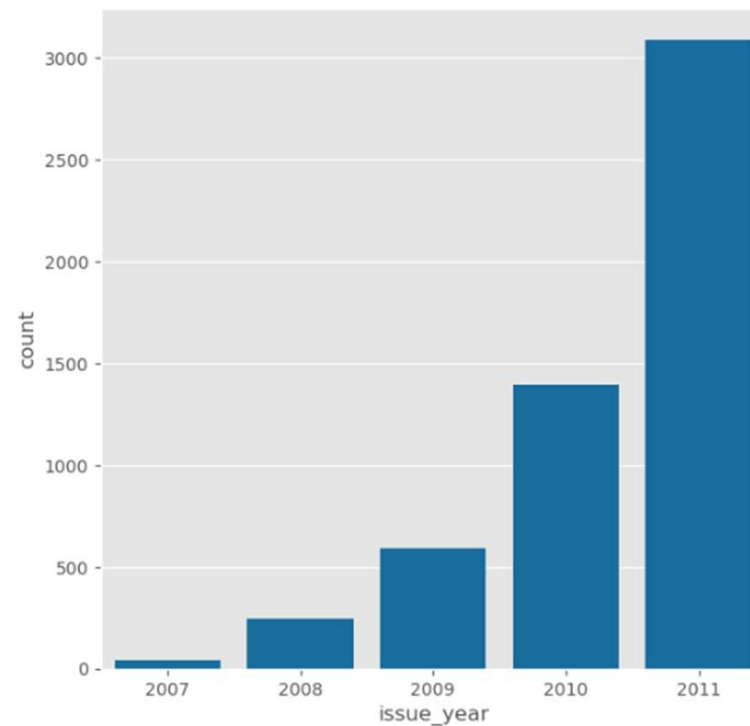
Year 2011 had a huge surge of loans resulting in huge amount of defaults

Distribution of Annual Income For Loan Status
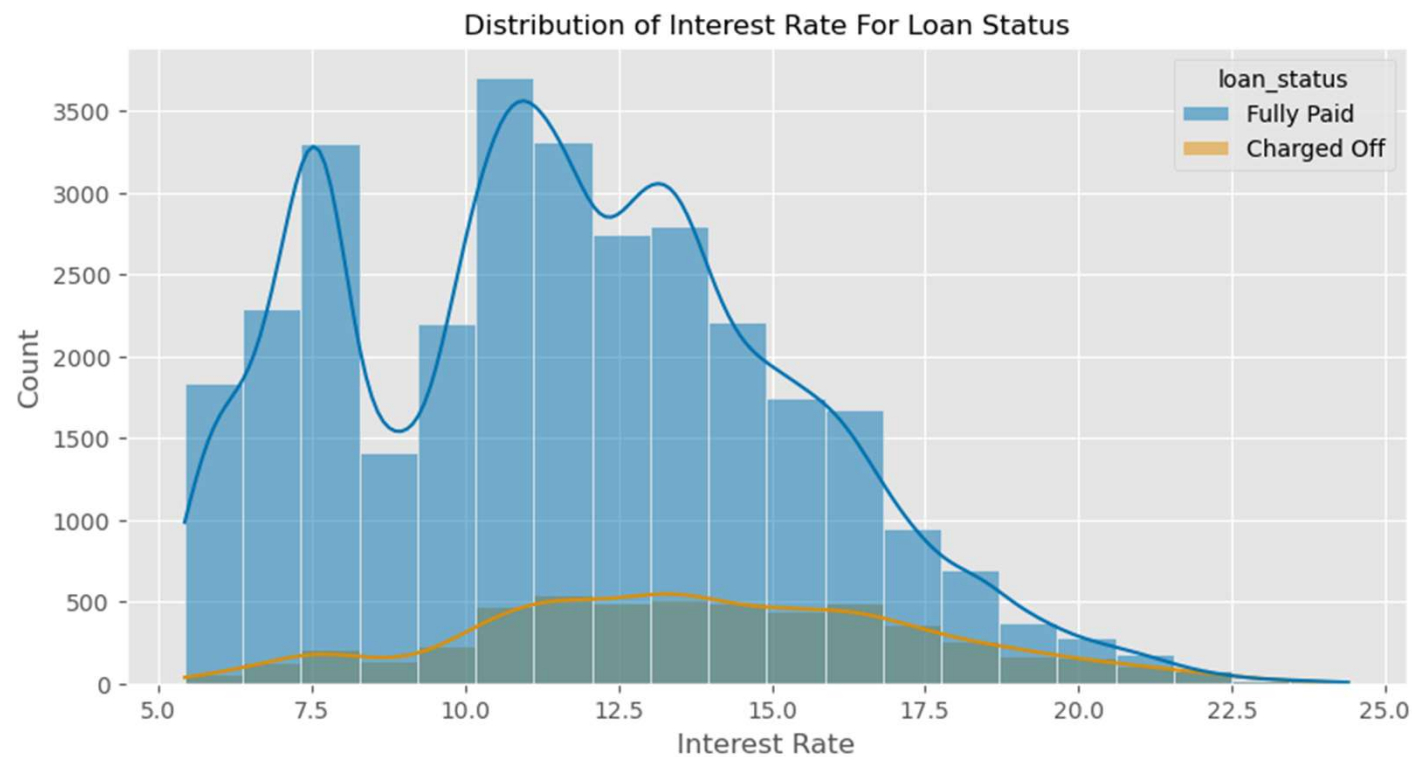

Borrowers below 50k have high almost half of the defaulters, this can be a risk factor

Distribution of Employment Length For Loan Status

Distribution of Employment Length For Loan Status

Borrowers with 10+ years have taken more loans and high number of fully paid, defaulters are also comparatively more

Distribution of Interest Rate For Loan Status
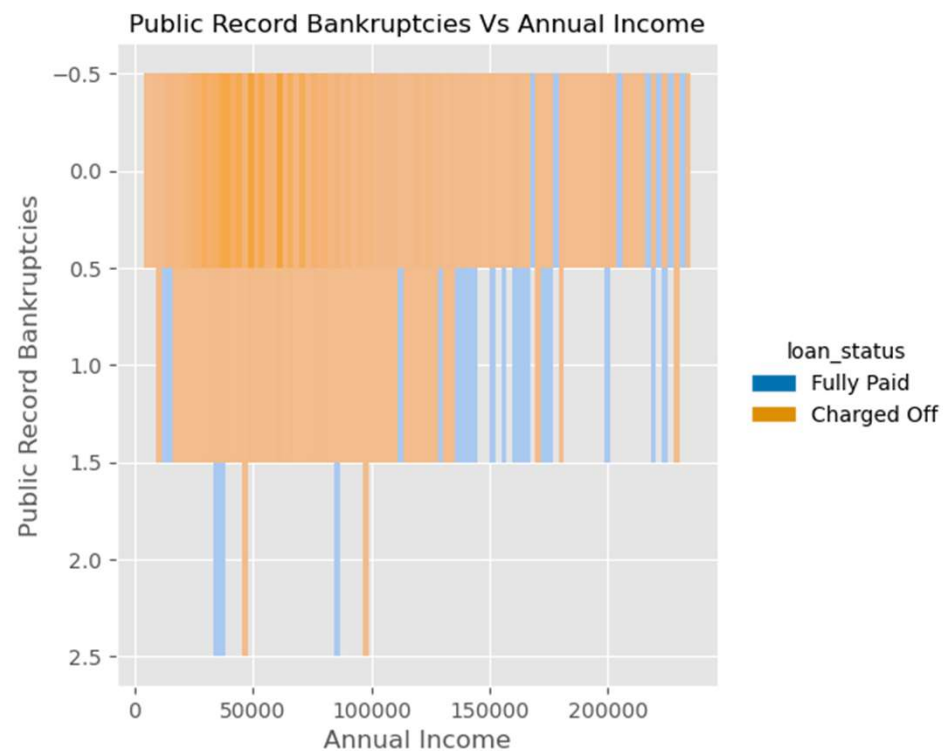

The interest rate has increased till 16% and then dropped from 17% for defaulters. For Fully paid , there is drop near 10%

Comparison of Interest Rate Based On Grade
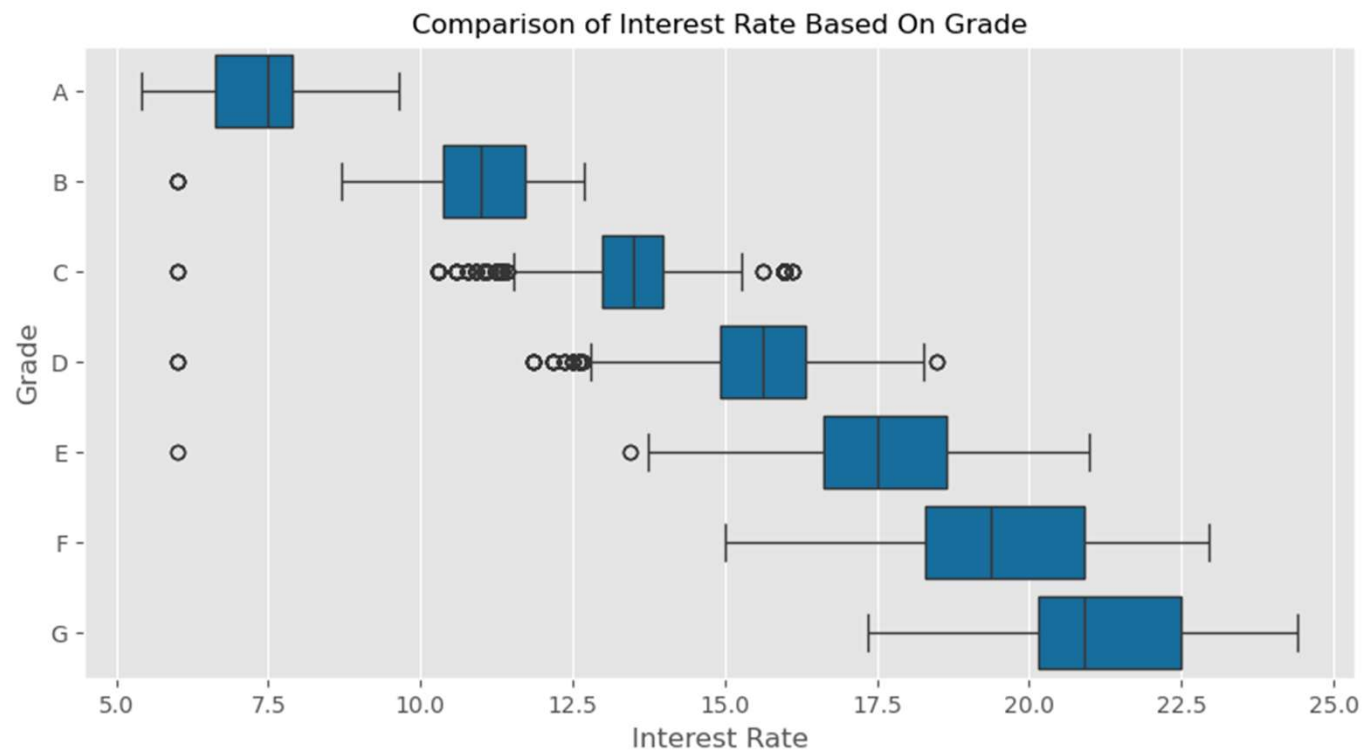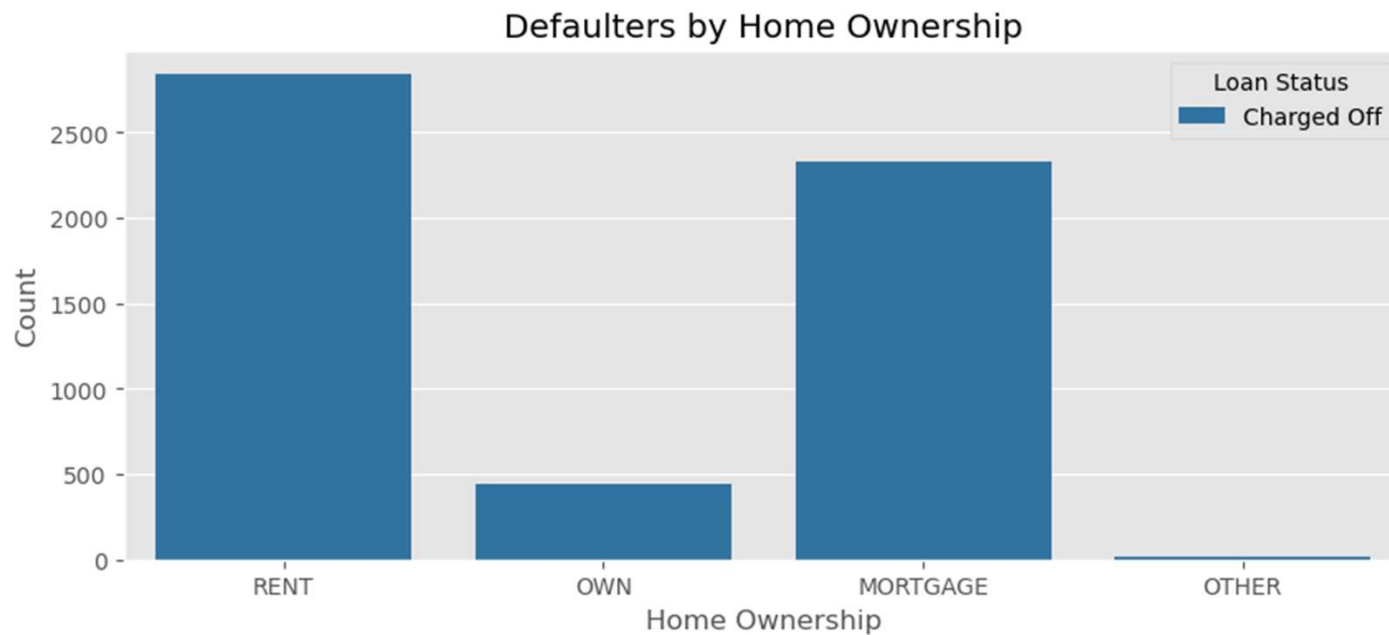

Higher the grade lower the rate of interest. People with good grades get lower interest rates.

Public Record Bankruptcies Vs Annual Income

Customers who have public record of Bankruptcies tend to default more

Defaulters by Home Ownership

Defaulters by Home Ownership

Defaulters with RENT and MORTGAGE Home Owners are most likely to default than with OWN homes

Distribution of DTI For Loan Status
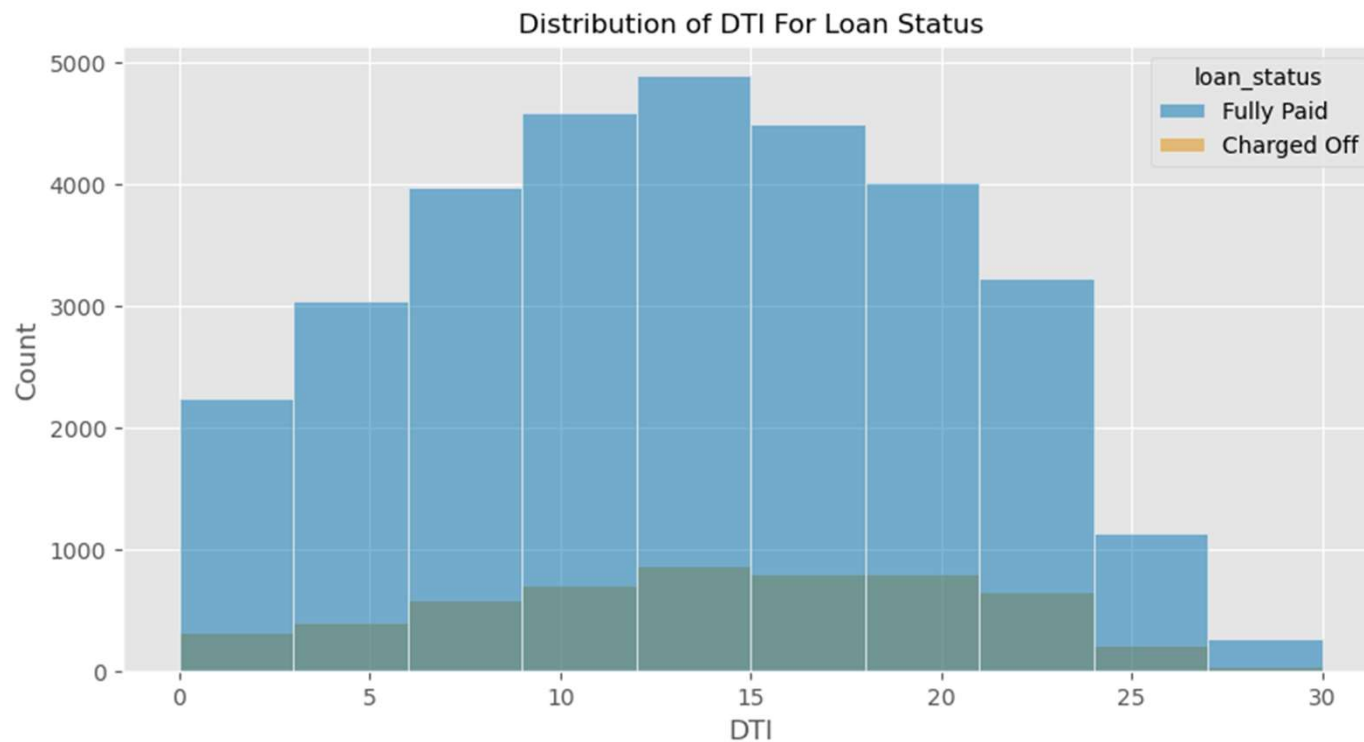
The Loan Status varies with DTI ratio, we can see that the loans in DTI ratio 10-15 have higher number of defaulted loan but higher dti has higher chance of defaulting

Distribution of Term For Grade
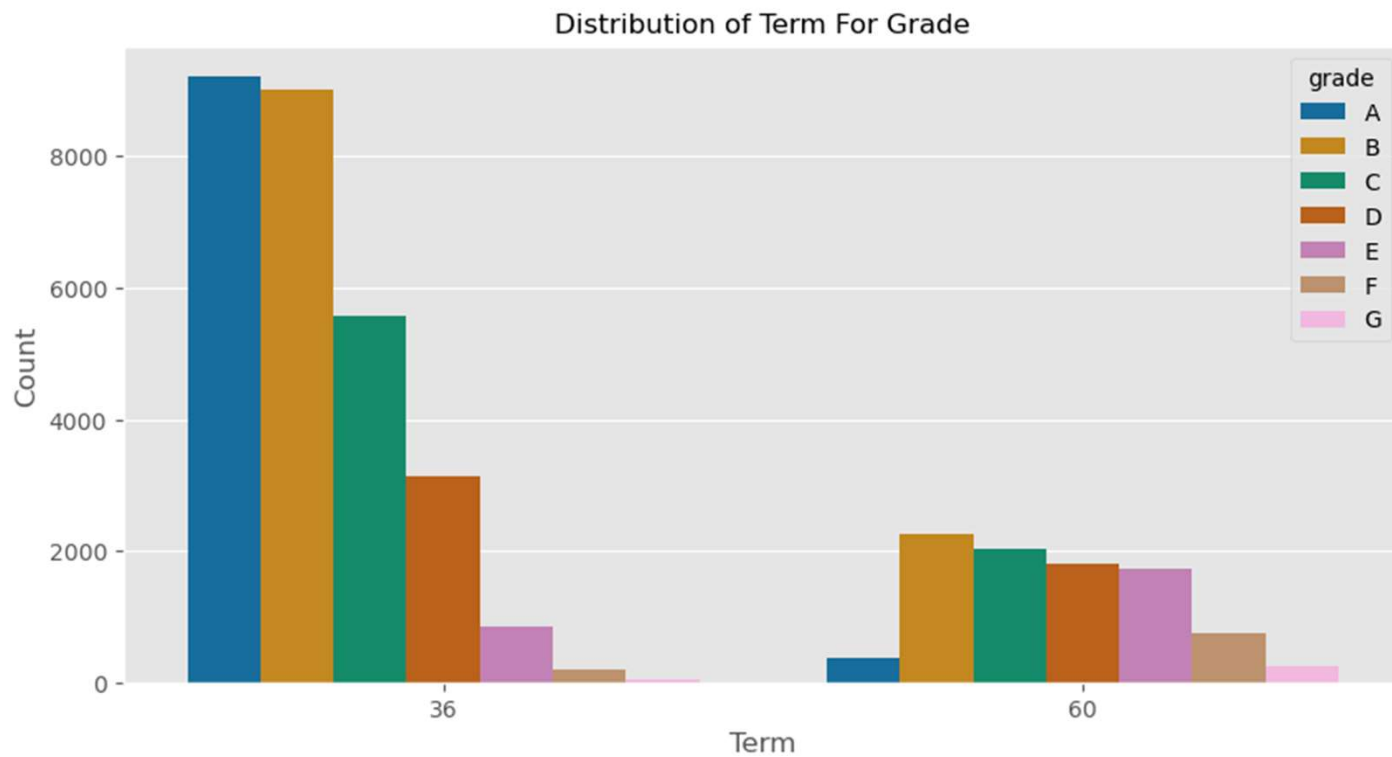
People with grade 'A' took loan for short term

Correlation Heatmap Including Categorical Variables

Correlation Heatmap Including Categorical Variables

Observations

- **Univariate Analysis**
  - Defaulter's loans are lower in numbers compared to Non-Defaulters (Fully Paid)
  - Loan amount has a range of loans from 500 to 35k with a mean of 9.8k
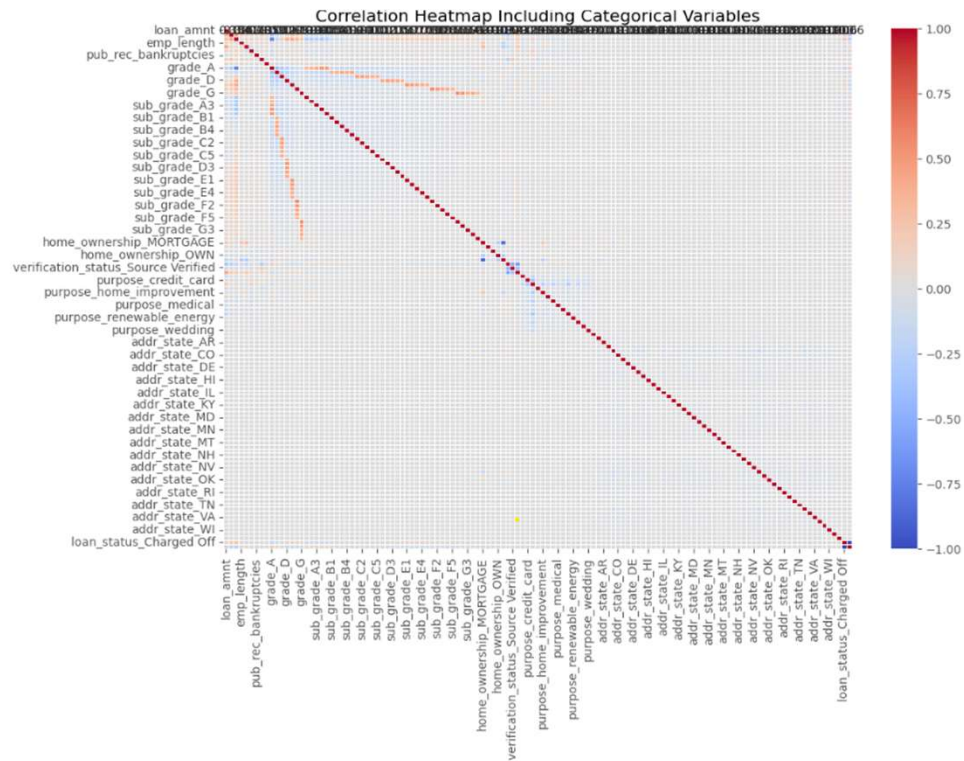 - The loans in 60 months term mostly consist of grade B, C and D loans whereas the loans in 36 months term majorly consist of grade A and B loans
  - More than half of the loan taken has term of 36 months compared to 60 months
  - Interest rate is more crowded between 5-10 and 10-15 along with a drop near 10%
  - Lot of loans are in Grade A and Grade B, meaning loans are approved more for higher grades

Observations

- **Segmented Univariate Analysis**
  - High number of loans are issued in 12th months and year 2011
  - Borrowers below 50k have high almost half of the defaulters, this can be a risk factor
  - Volume of loans has vastly increased with each year, also indicates approvals have increased
  - Borrowers with 10+ years have taken more loans and high number of fully paid, defaulters are also comparatively more
  - The interest rate has increased till 16% and then dropped from 17% for defaulters. For Fully paid, there is drop near 10%
  - Defaulters with RENT and MORTGAGE Home Owners are most likely to default than wit OWN homes
  - The Loan Status varies with DTI ratio; we can see that the loans in DTI of 10-15% have higher number of defaulted loan, higher the DTI more chance of defaulting.
  - A high percentage of loans are taken for the purpose of debt consolidation followed by credit card
  - Compared to Sub grades of A, more numbers of defaulters are in Sub Grade of B & C, this us a risk factor

Observations

- **Bivariate Analysis**
  - The rate of interest is inversely proportional to Grades meaning higher the grade, lower is the interest rate. Grade is a risk factor.
  - Most of the borrowers don't have bankruptcy records and are safe choice for loan approvals
  - The Grade A which is lowest risk also has lowest DTI ratio which we can say that higher grade has lower rate of default
  - The borrowers are mostly having no record of Public Recorded Bankruptcy and are safe choice for loan issue.

Factors influencing loan defaulters

- Loans against Debt Consolidation and Credit Card
- Debt to Income factor
- Grades
- Annual income
- Public Record Bankruptcies
- Home Ownership
- Higher Interest rates
- Length of the Employment

Rationale

- Borrowers having loans against debt consolidation and Credit Cards are more likely to get into debt trap with high interest rates complicating loan repayment and are likely to default
- Borrowers with very high Debt to Income ratio are likely to default, additional due diligence needs to be carried out
- Borrowers with highest grade A are safer compared to lower grades pose high risk
- Borrowers having annual income less than ~50000 are likely to default more posing risk of repayment
- Borrowers with Public Recorded Bankruptcies pose higher risk for loan repayment
- Borrowers with Rented and Mortgaged Home Ownership carry more disk to default loans
- Borrowers with high interest rates pose high risk to default
- Even though borrowers with working experience of 10+ years are likely to fully pay, they also pose risk to default loans given the fact borrowers with 10+ years experience are high consumers of loans