

Lending Club Case Study

GROUP MEMBERS:

BIBHUTI RANJAN SINHA
HEENA JAIN

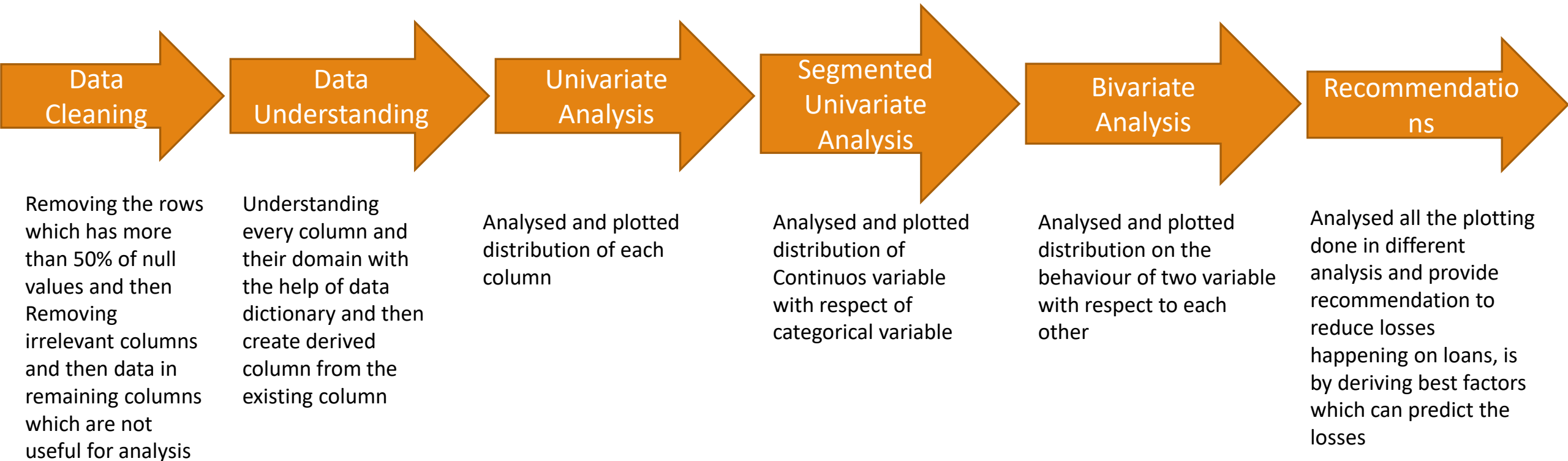
Business Understanding

- A consumer finance company which specialize 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.

Business Objectives and Goal

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- If one is able 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.
- The goal 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.

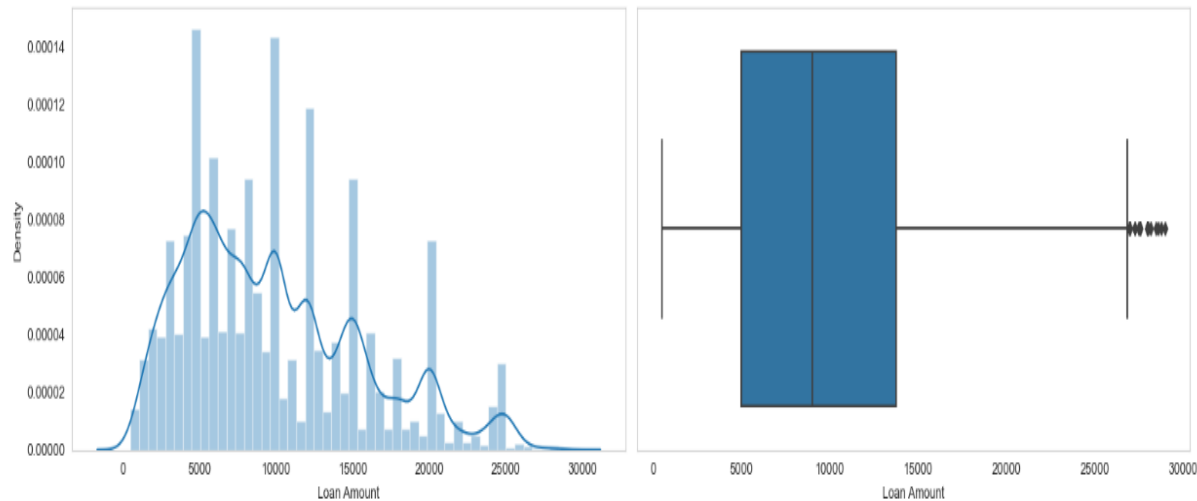
Problem Solving Methodology



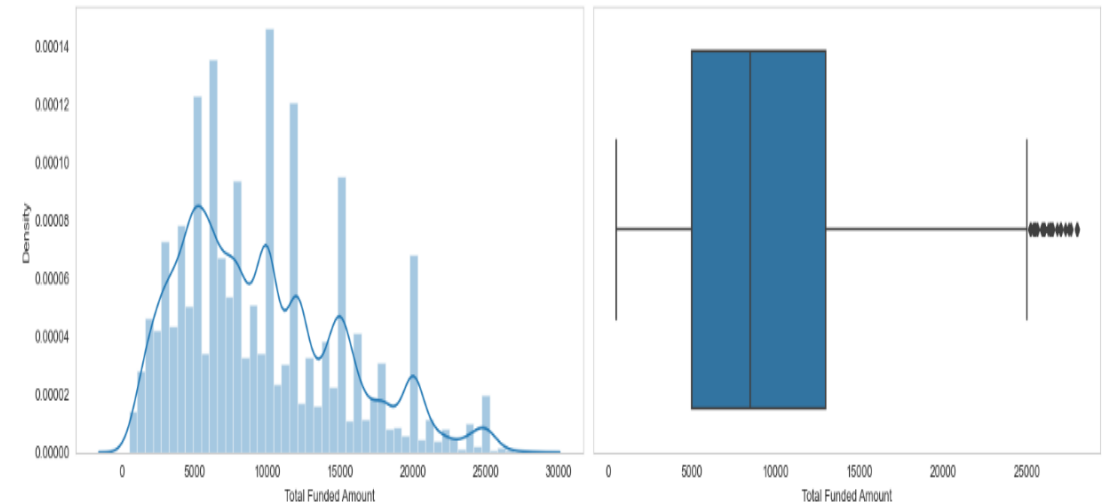
Univariate Analysis

Quantitative Variable Analysis

Loan Amount



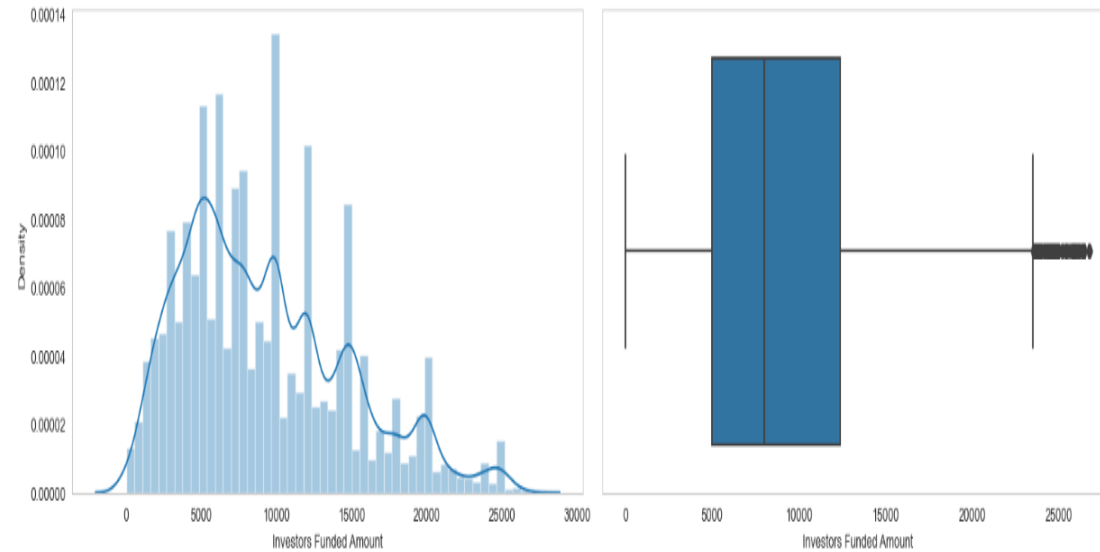
Total Funded Amount



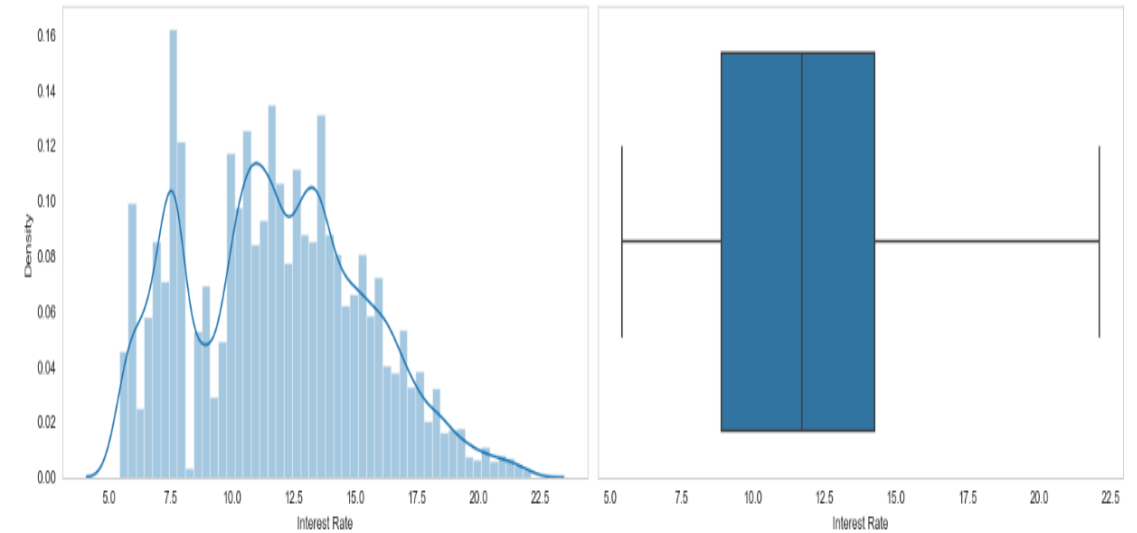
Inference:

- Maximum frequency of loan amount exists between 5k to 14k
- Maximum frequency of Funded amount exists between 5k to 13k

Investors Funded Amount



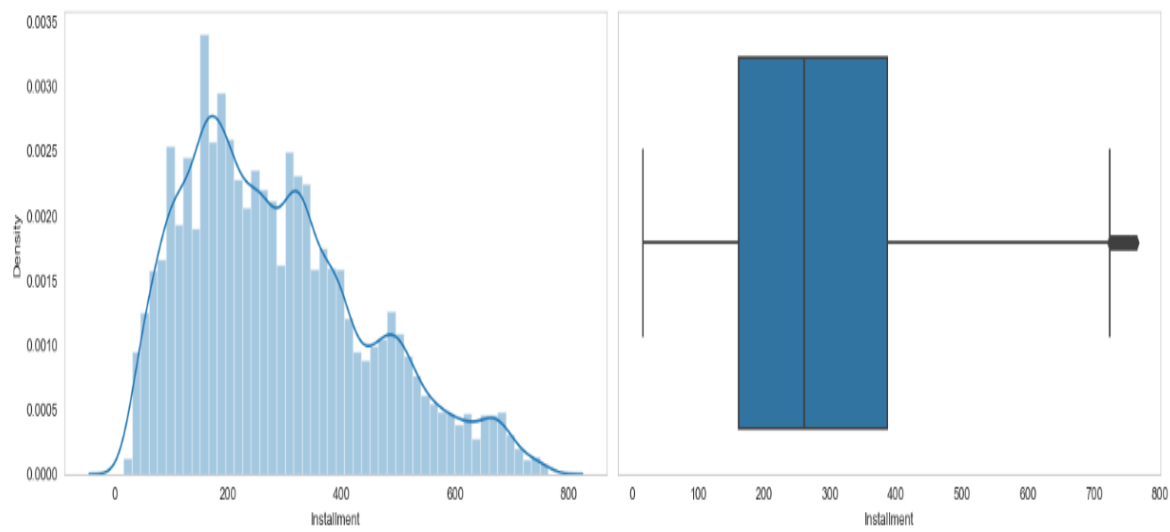
Interest Rate



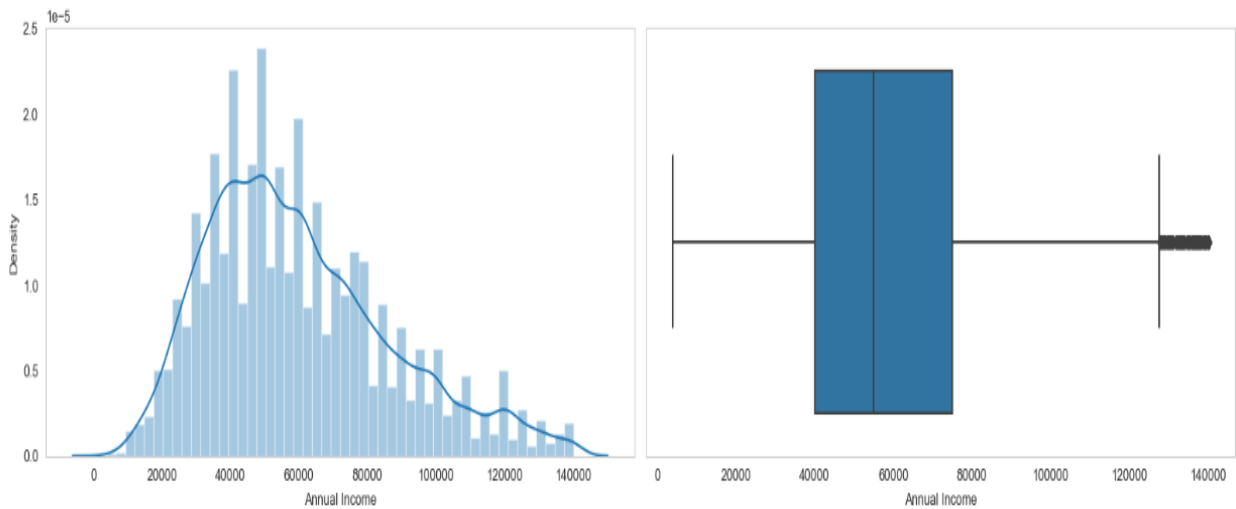
Inference:

- Maximum frequency of Investors Funded Amount exists between 5k to 12k
- Data for the interest rate is in the range of 5% to 16% going at the max to 22%

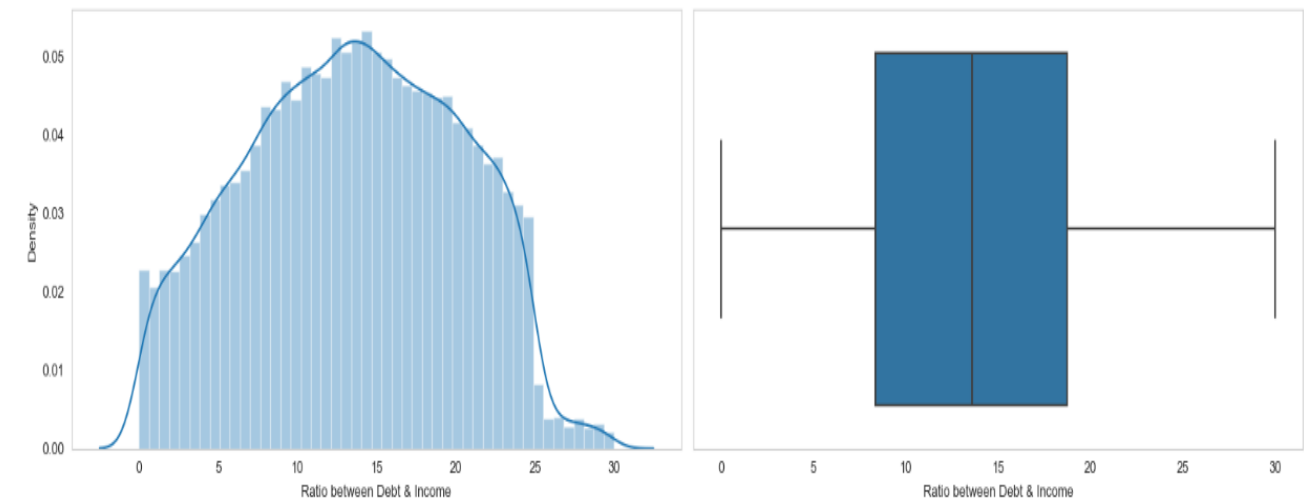
Installment



Annual Income



Ratio between Debt & Income

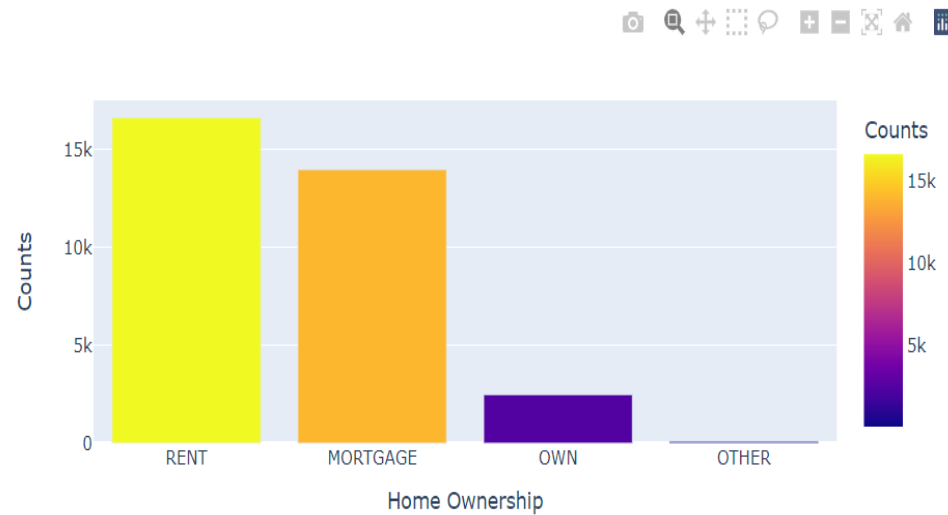


Inference:

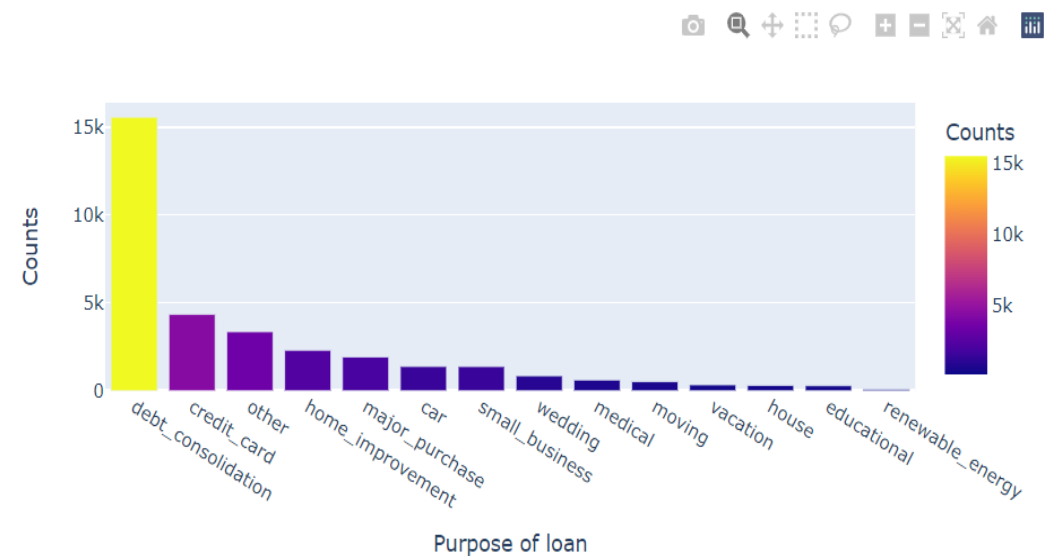
- Data for the installment is in the range of 20 to 400 going at the max to 700
- Majority of the annual income is in the range of 4k to 40k going at the max to 120k. This column required major outlier treatment
- Majority of the debt to income is in the range of 0 to 20 going at the max to 30

Unordered Categorical Variable Analysis

Home Ownership



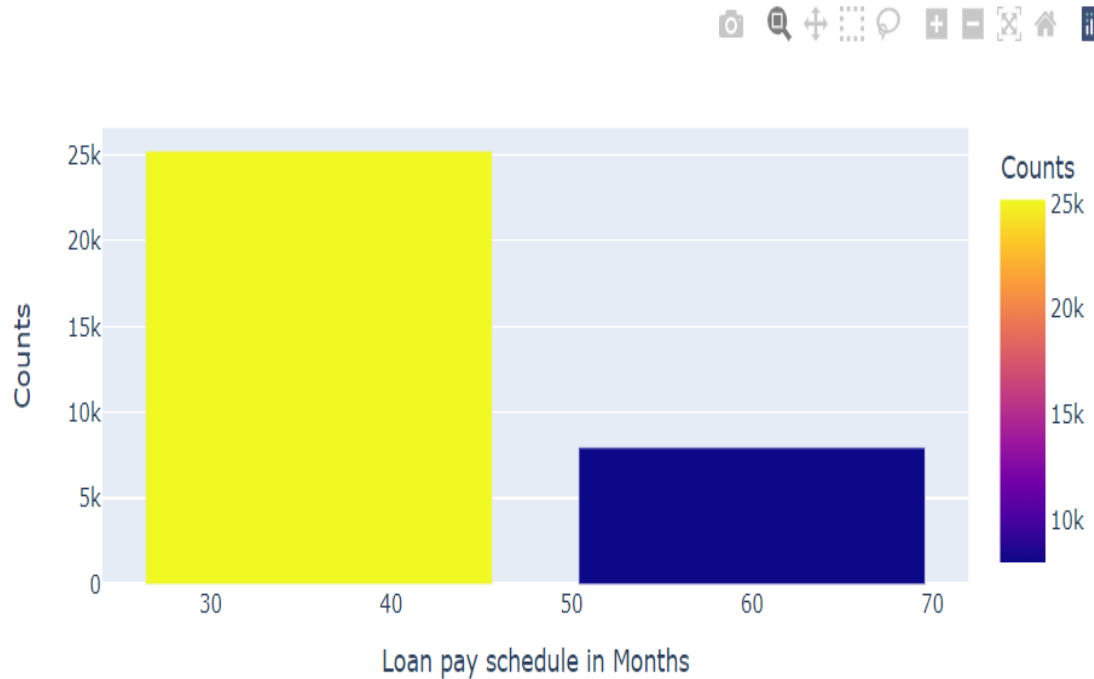
Purpose of loan



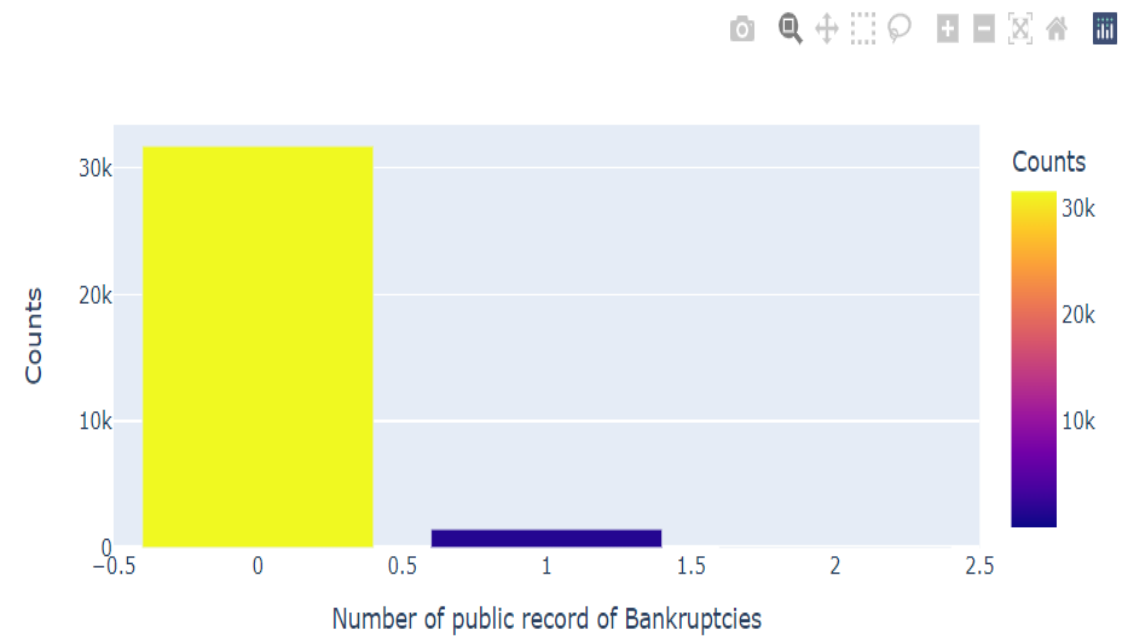
Inference:

- People who are living on rent or having mortgage having more loan
- Most of the loan has been taken for dept consolidation

Loan pay schedule in Months



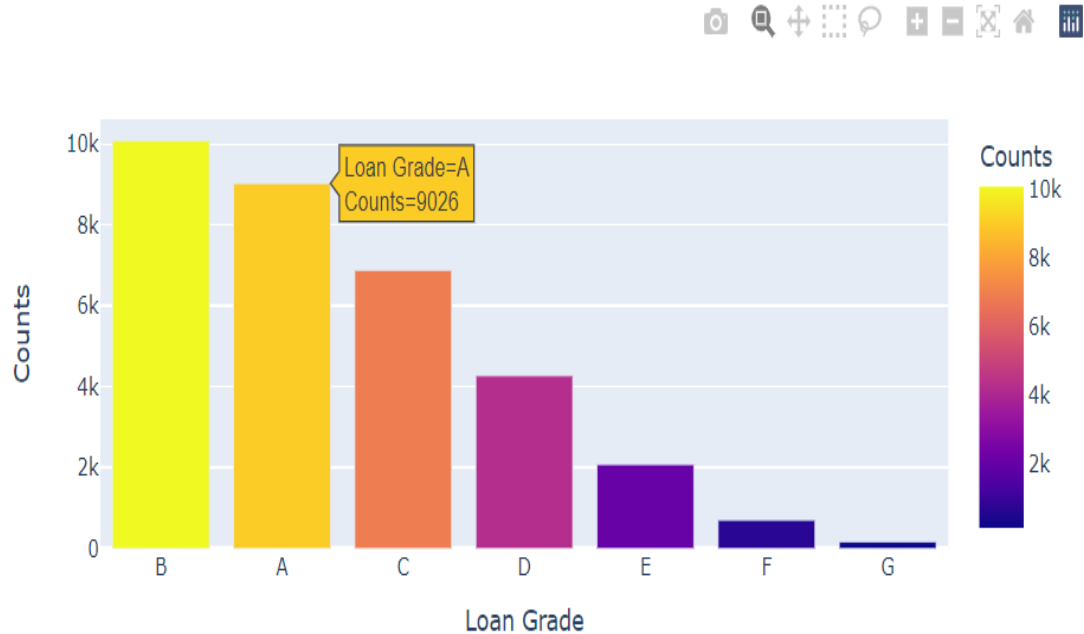
Number of public record of Bankruptcies



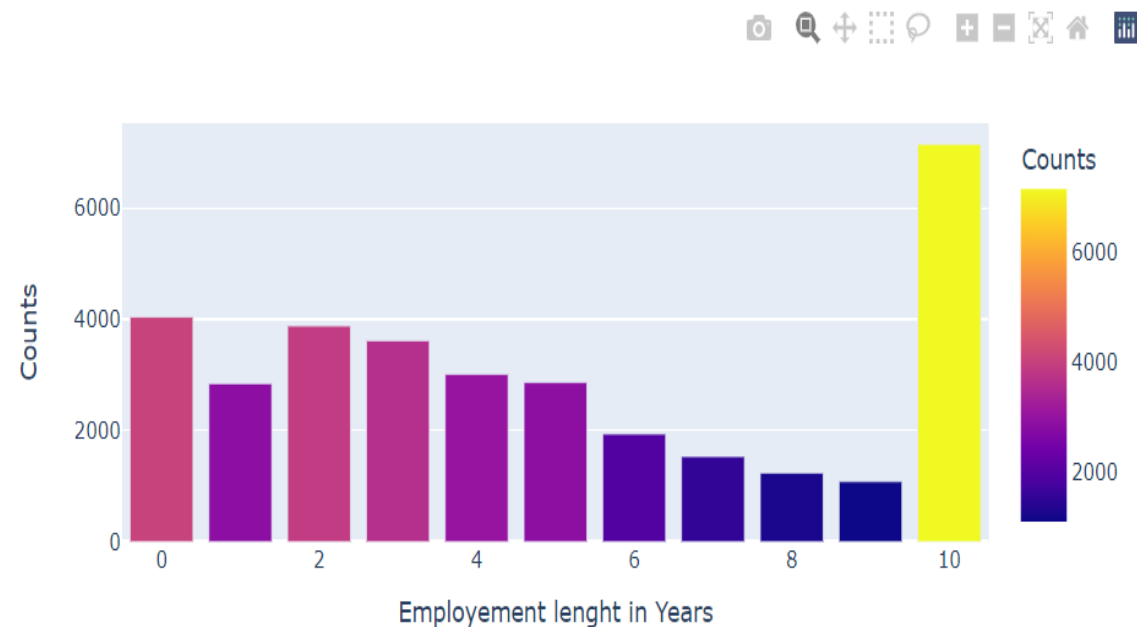
Inference:

- Most of the loan has been taken for 36 months duration
- Majority of the loan applicants are in the category of not having an public record of bankruptcies

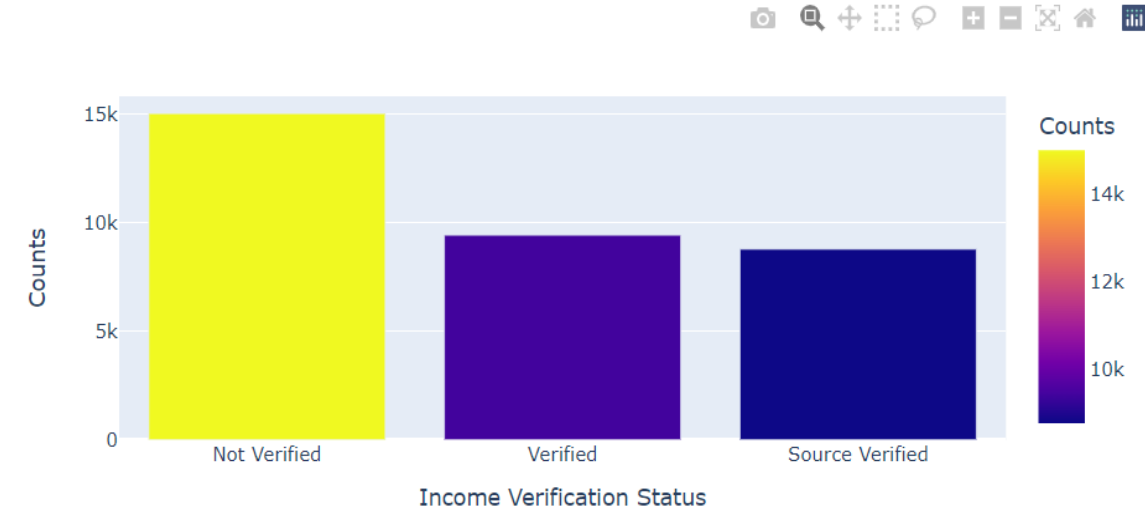
Loan Grade



Employment lenght in Years



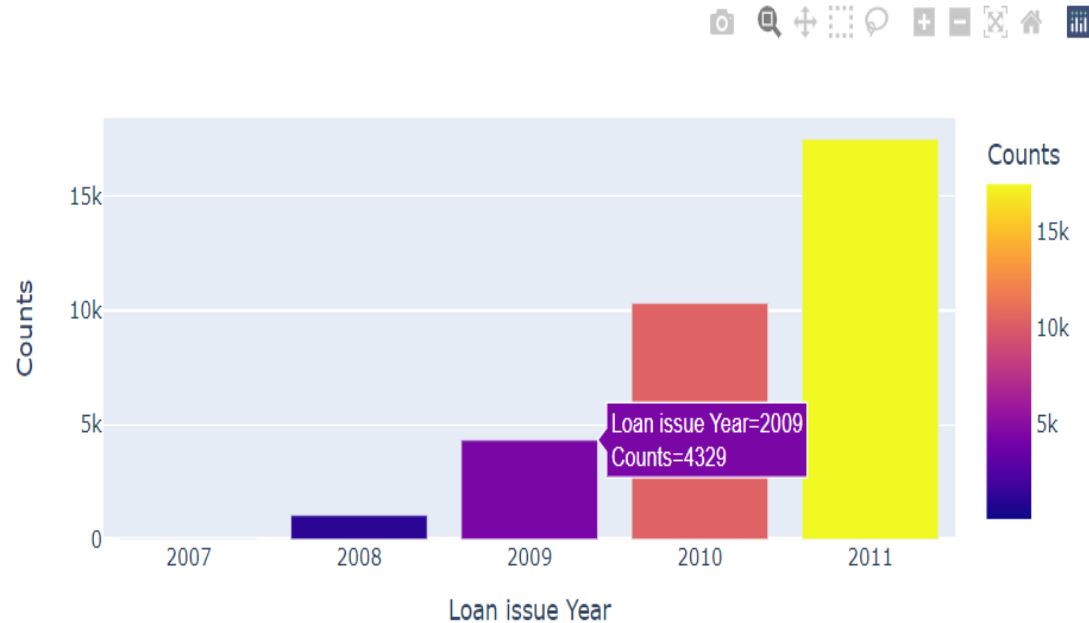
Income Verification Status



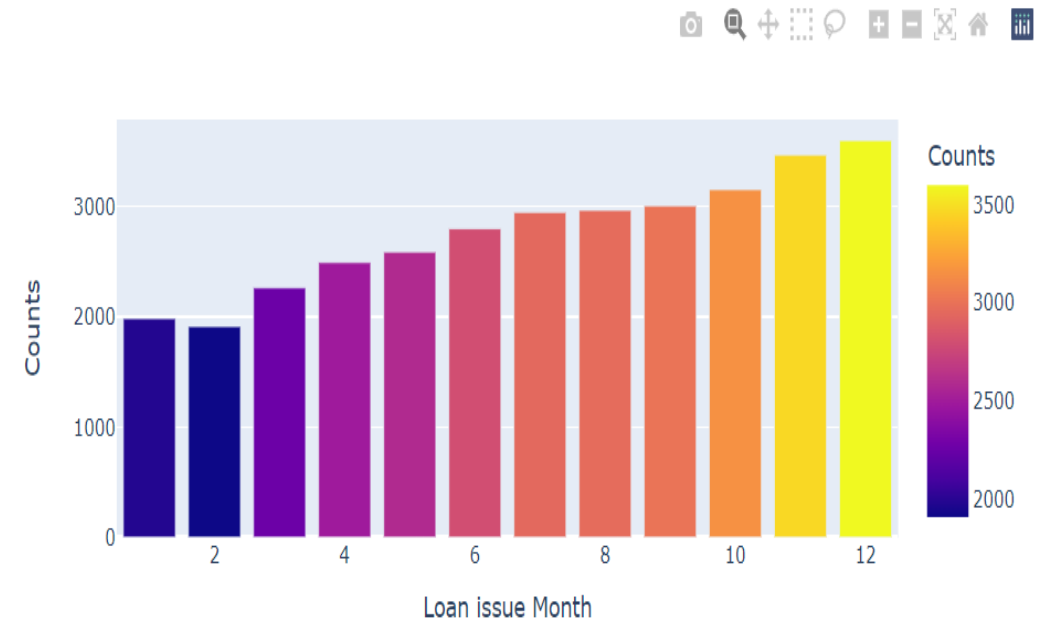
- Inference:
- Most of the loans falls under Grade B
 - Employees with 10 years of working experience having more number of loans
 - Data shows for most of the loan applications, Income Source was not verified. This could result the bad loan hence financial loss

Derived variables analysis

Loan issue Year



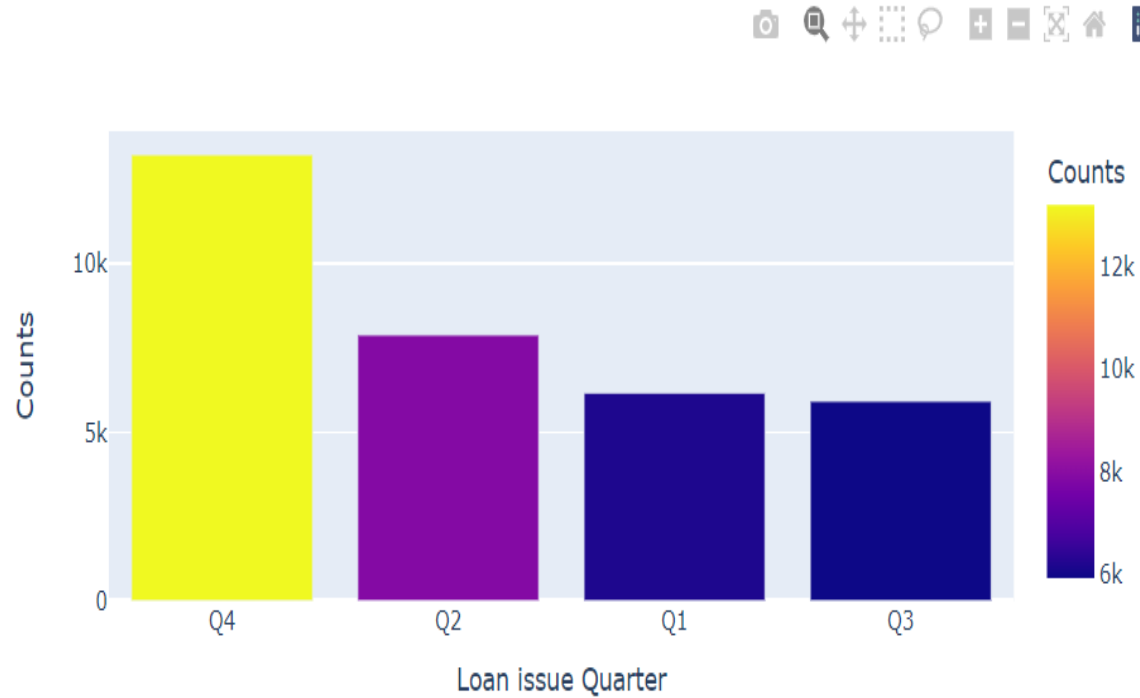
Loan issue Month



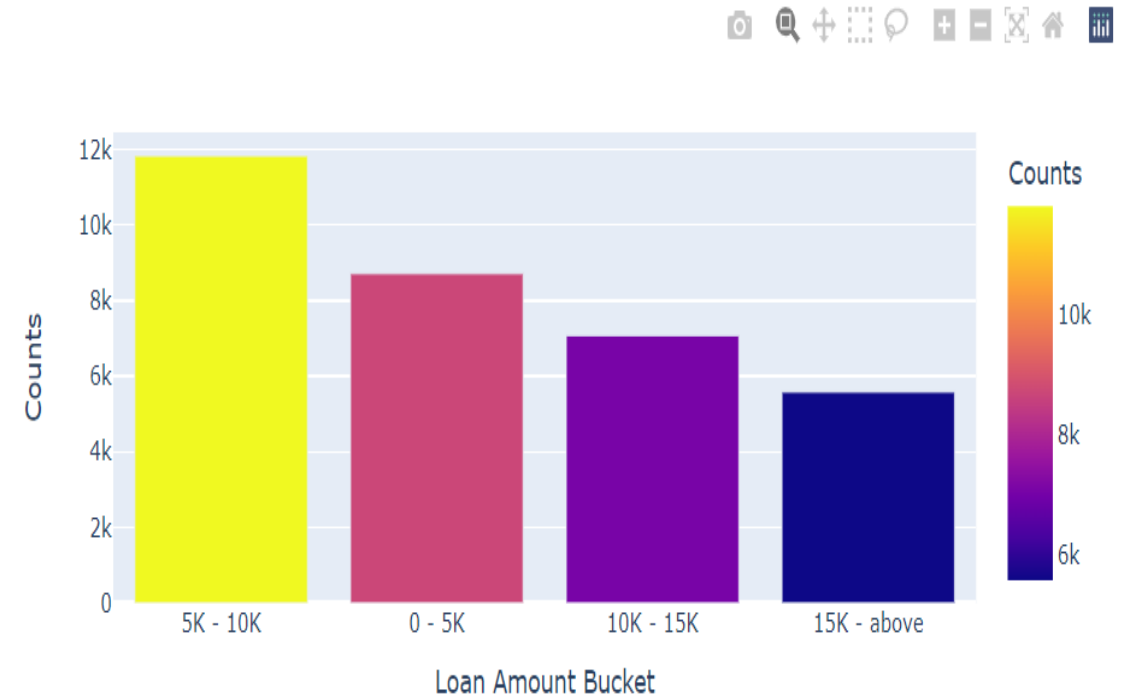
Inference:

- Number of loan issues in 2011 is very high. It's also indicate that number of loan applications is increasing yearly
- Chart shows that people takes more loans in month of Dec. Number of loans issued in end of year is more

Loan issue Quarter



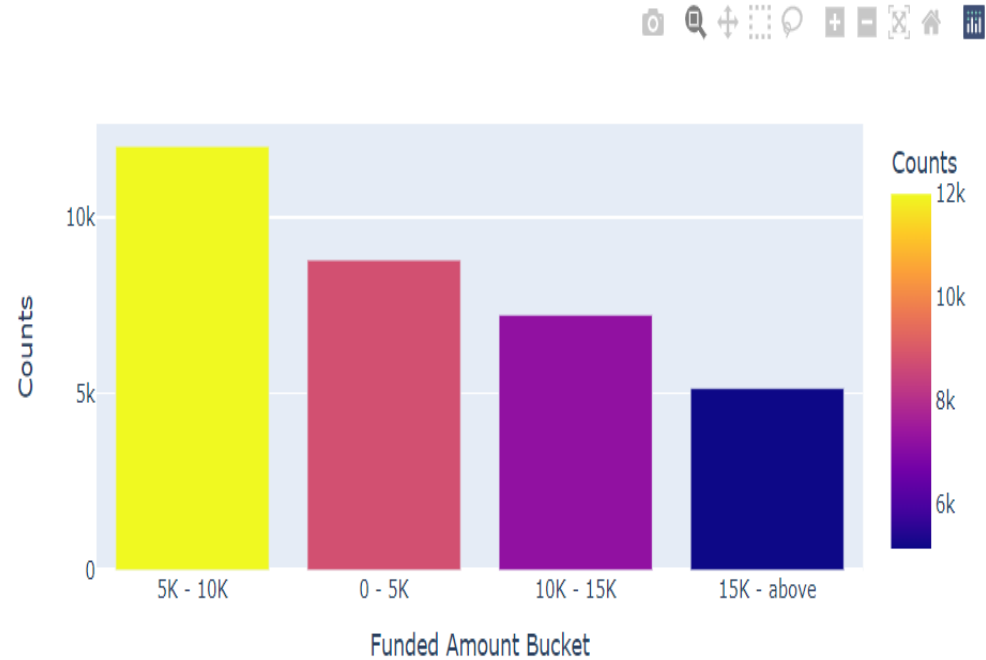
Loan Amount Bucket



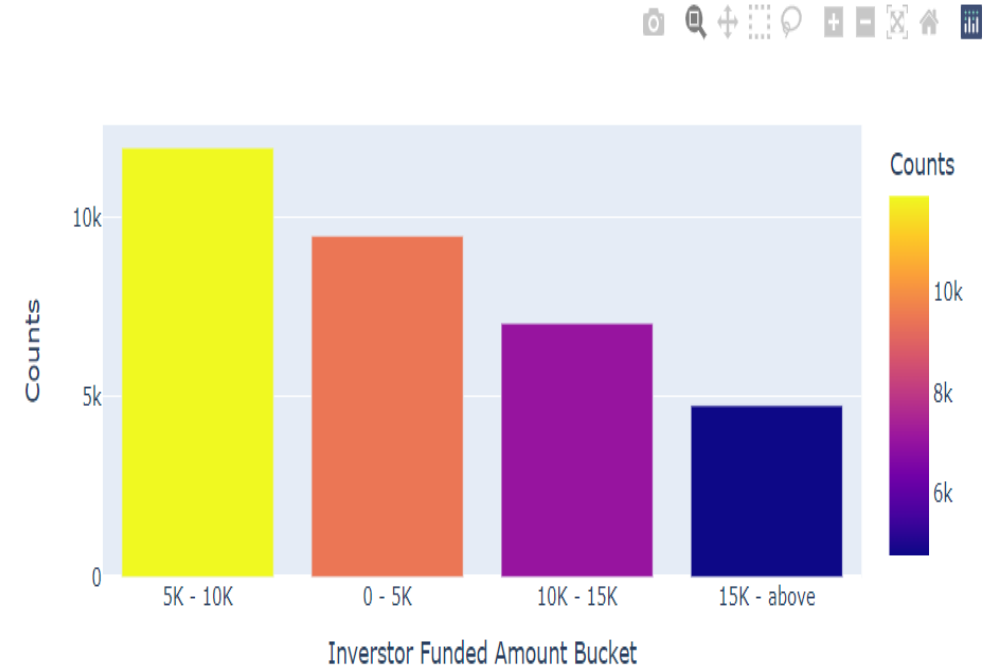
Inference:

- As we have seen in previous chart, this chart also represent number of issued loan is more in Q4 or end of the year
- Maximum number of loan amount lies between 5k to 10k where 15k and above loan amount are very less

Funded Amount Bucket



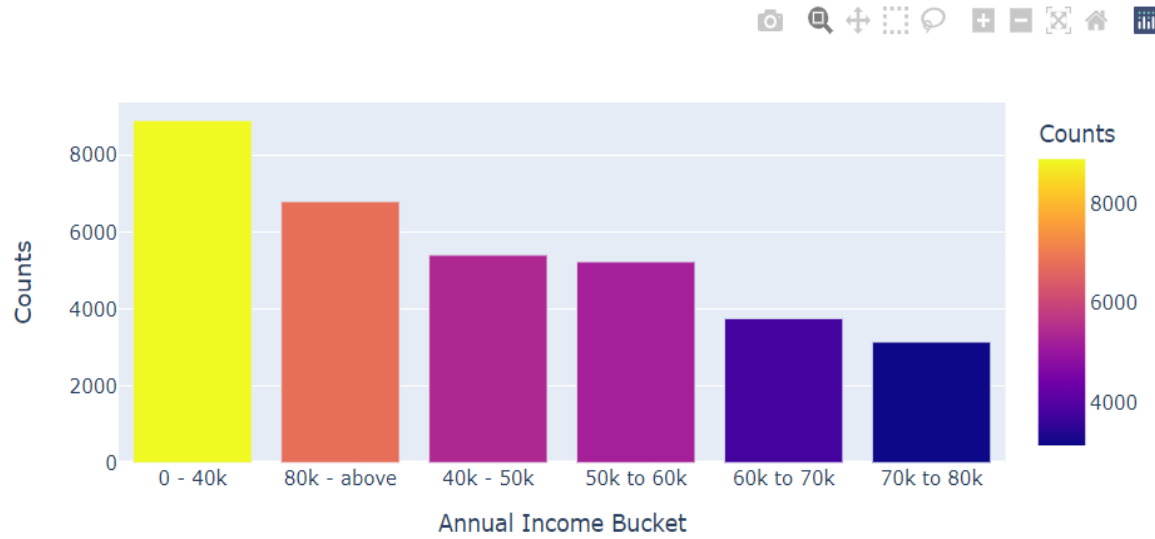
Investor Funded Amount Bucket



Inference:

- Maximum number of funded amount lies between 5k to 10k where 15k and above funded amount are very less
- Maximum number of Investor funded amount lies between 5k to 10k where 15k and above Investor funded amount are very less

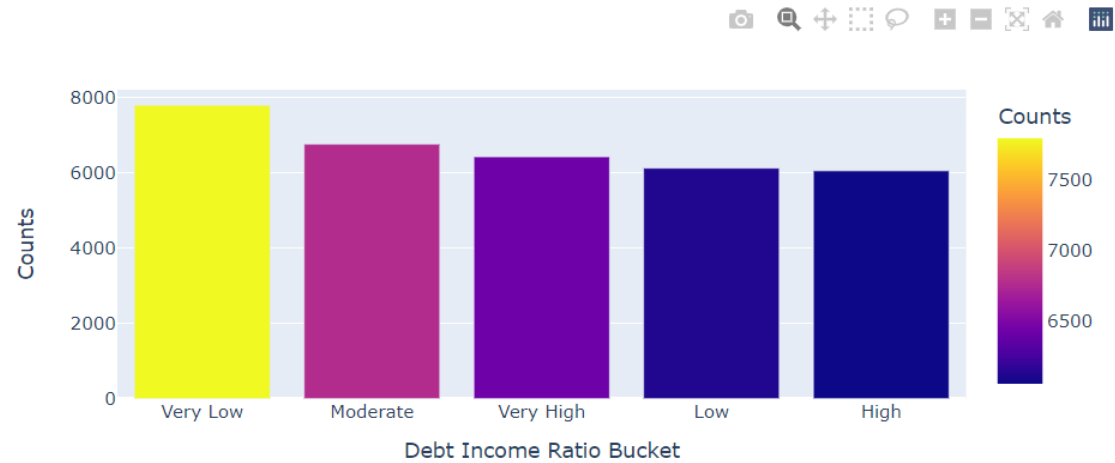
Annual Income Bucket



Interest Rate Bucket



Debt Income Ratio Bucket

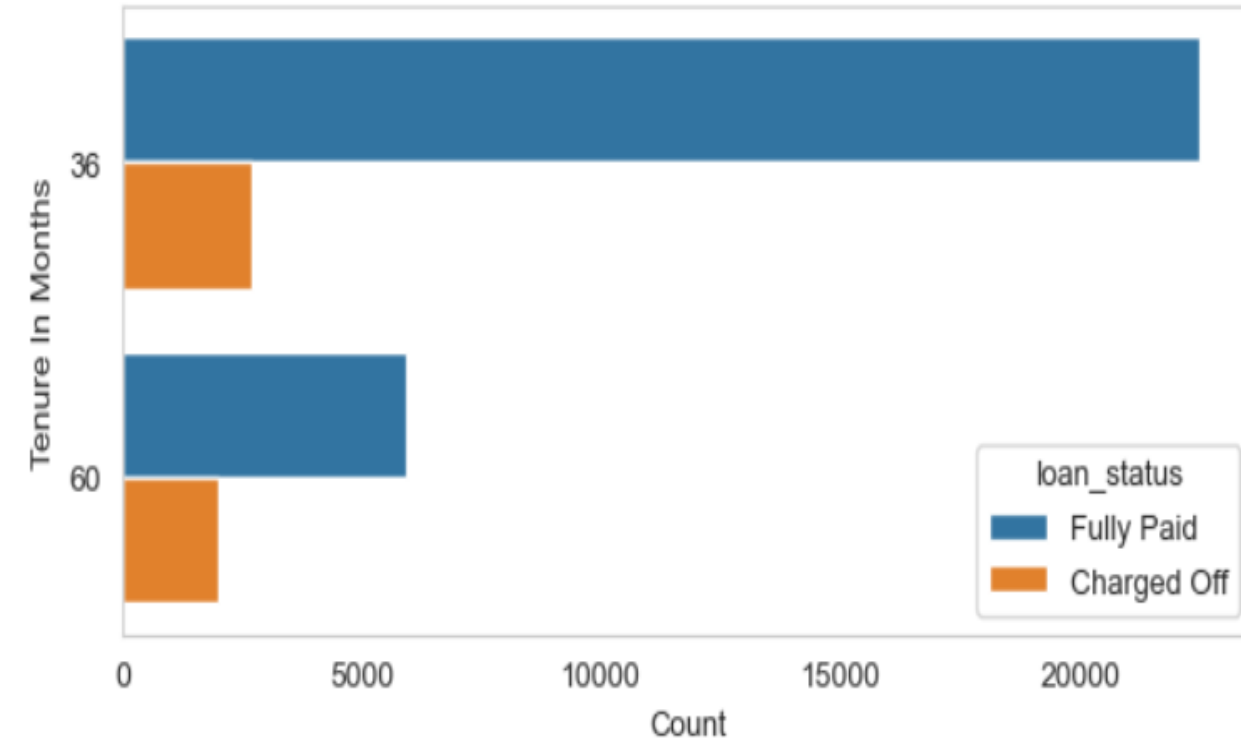


Inference:

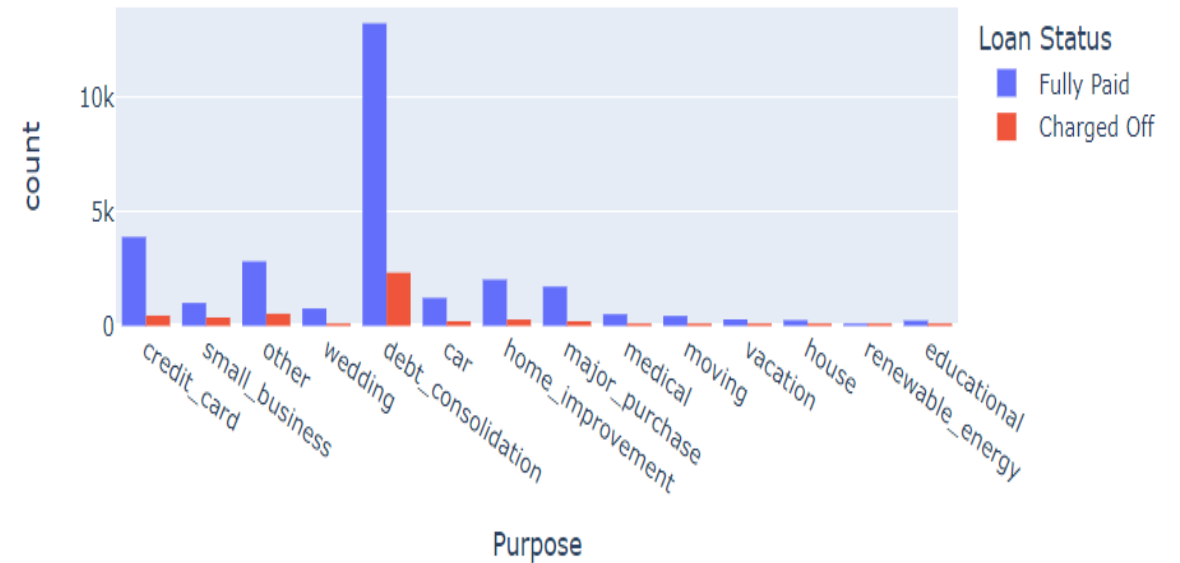
- This charts clearly shows the Annual Income and Number of loan has negative correlation
- Maximum number of loans are on low interest rate but Majority of loans are on Moderate and Very High Interest rate
- Majority of the loan applications are in Moderate debt to income ratio

Segmented Univariate Analysis

Loan status based on loan tenure



Loan Taken For Different Purposes



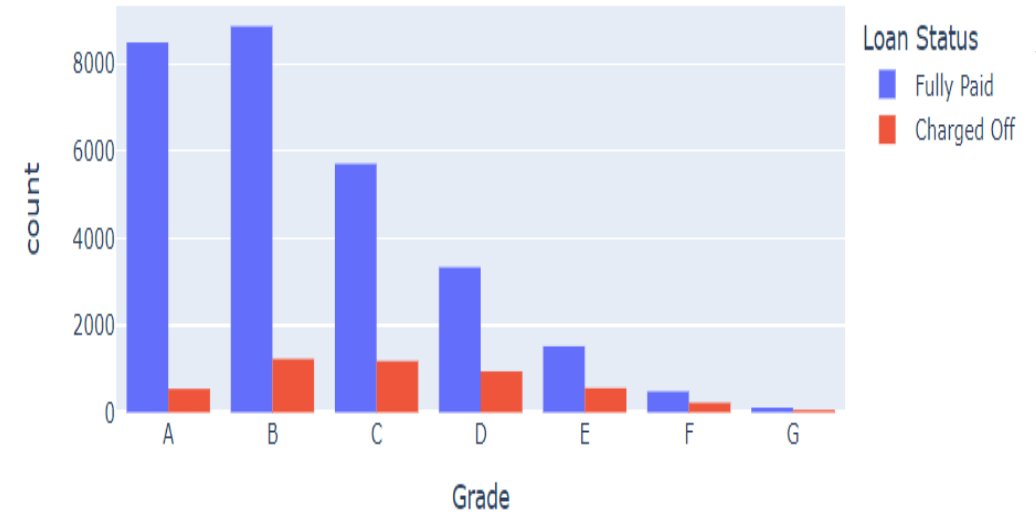
Inference:

- The longer tenure has highest chance to defaulting loan. In the above chart 60 month term has higher rate of defaulting than 36 month term whereas the 36 month term has higher rate of fully paid loan.
- Debt Consolidation is the most popular loan purpose and has highest number of fully paid loan and defaulted loan.

Distribution of Loan Amount For Loan Status



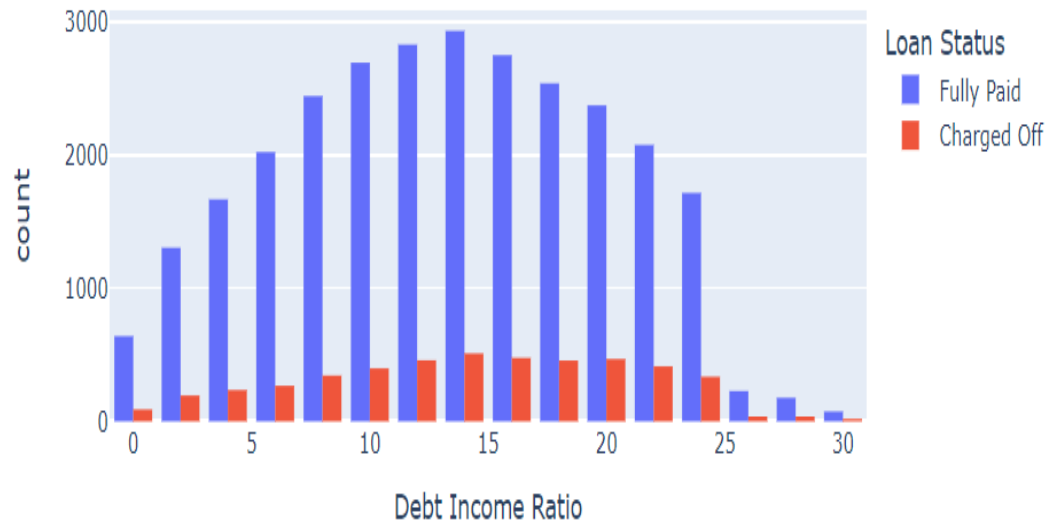
Loan status based on loan assigned grade



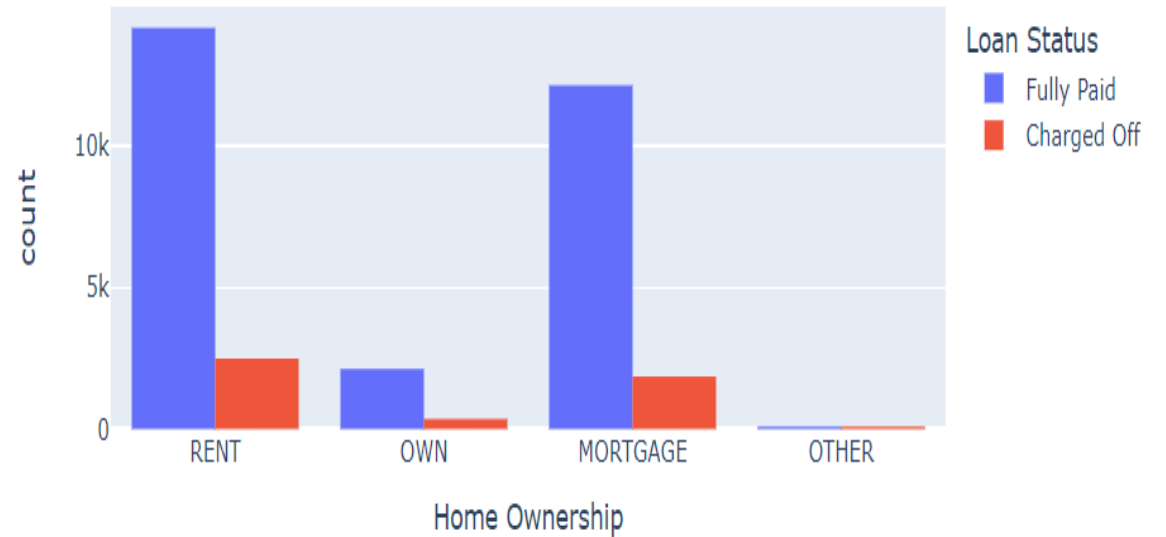
Inference :

- The mean and 25 percentile are almost same for both type of loan status but if we look for 75 percentile, it indicating that large amount of loan has higher chance of defaulting.
- The above chart indicating that Loan Grade B have Higher number of Fully Paid loans that Charged Off where E,D,F G, have comparatively higher chances of defaulted loans

Loan Status based on Debt Income Ratio



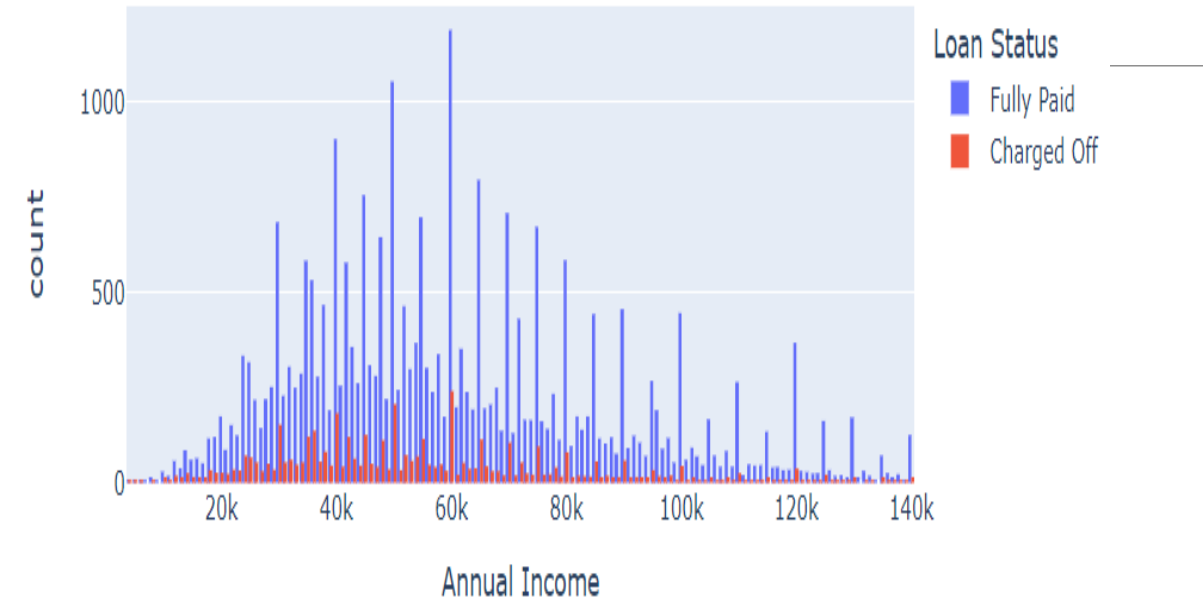
Loan Status based on Home Ownership



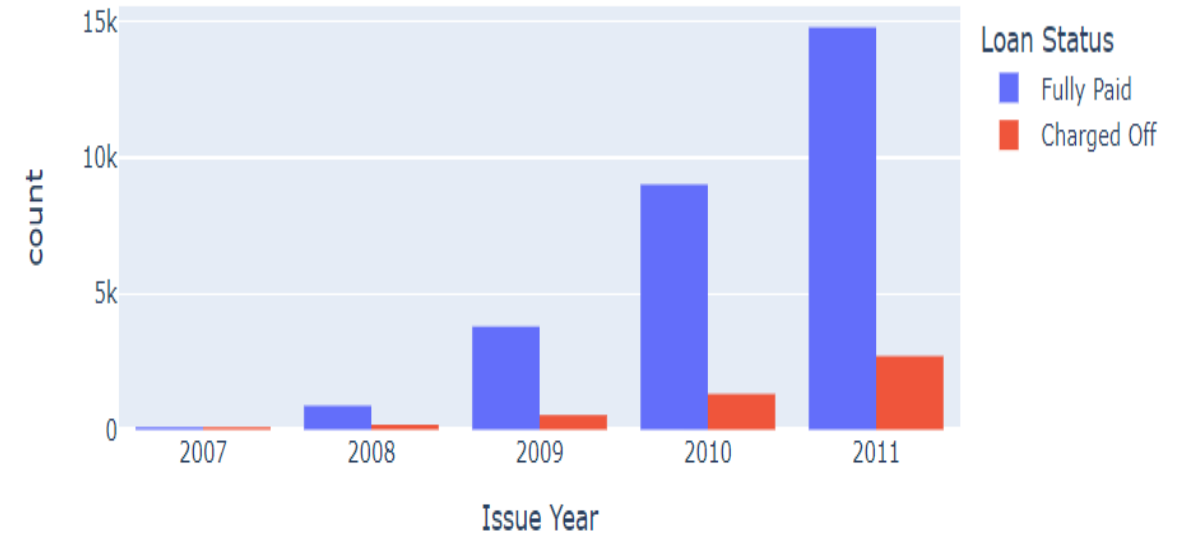
Inference:

- 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.
- This chart indicating that Customer on Rented or Customer with mortgage has high chance of Loan Defaulting

Loan Status based on Annual Income



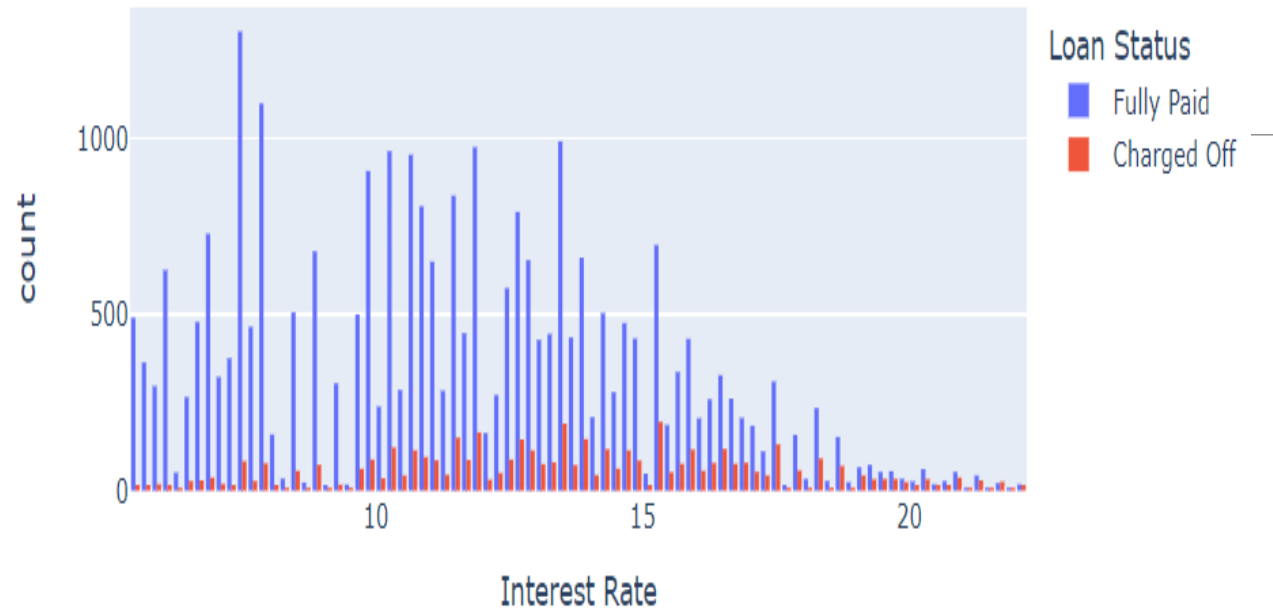
Loan Status based on Issue Year



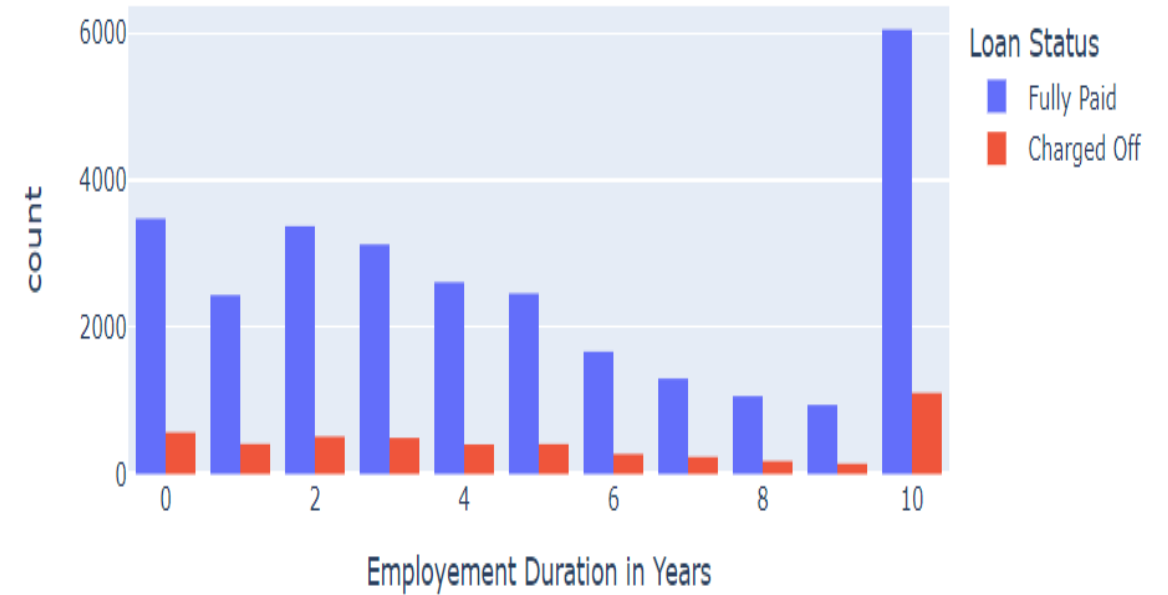
Inference:

- Applicants whose income is 70k and below is highly chance to do loan default
- The Fully paid loan are increasing exponentially with the time compared to defaulted loan

Loan Status based on Interest rate



Loan Status based on Employment Duration

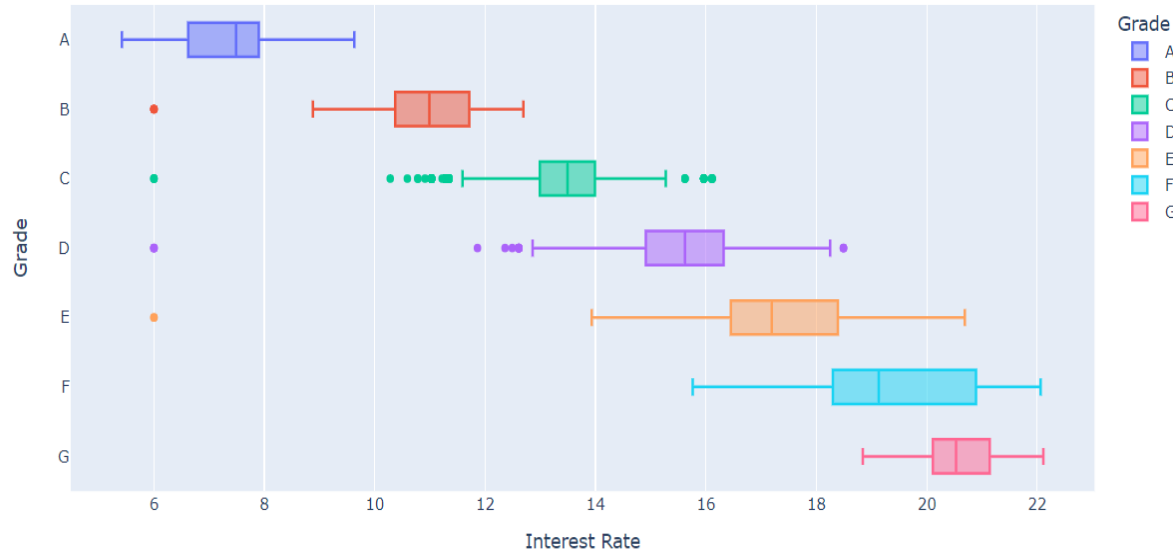


Inference:

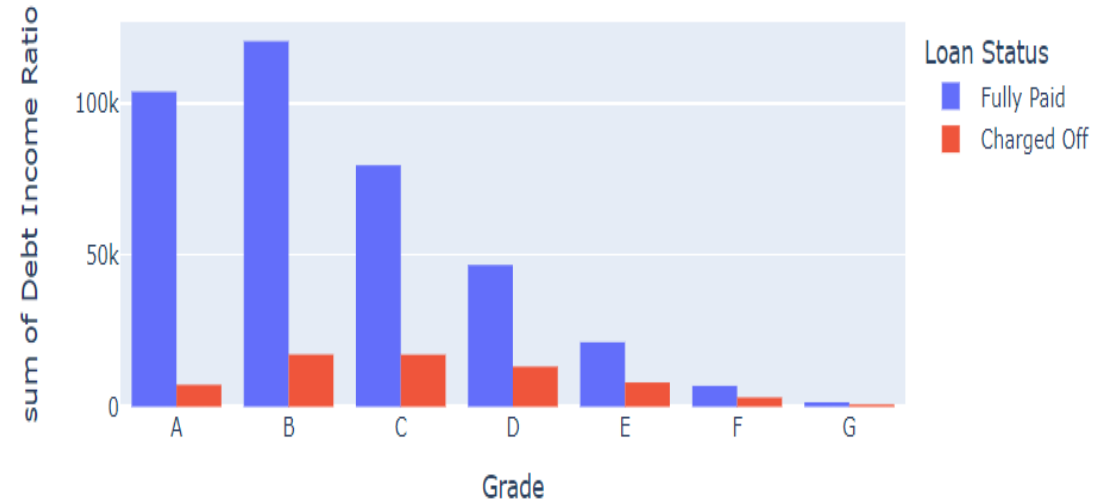
- Higher Interest rate has higher chance to loan default
- Employees with 10 and more years of employments are less likely to default the loan.

Bivariate Analysis

Comparison of Interest Rate Based On Grade

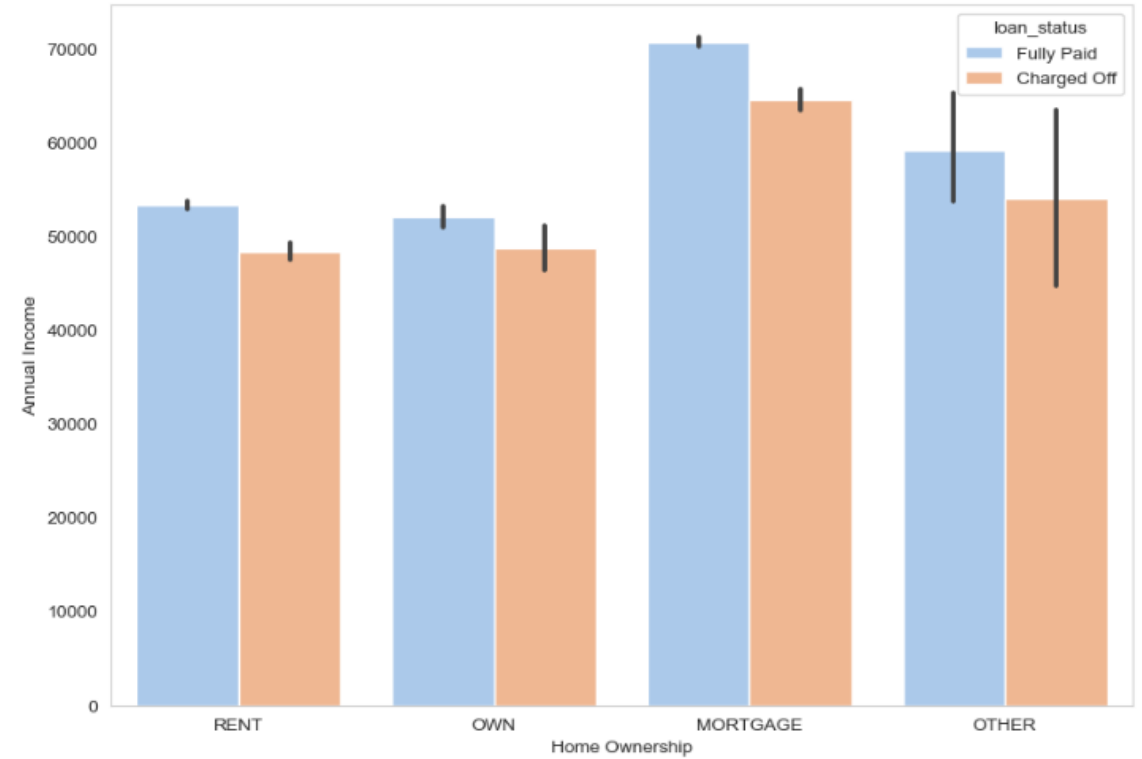
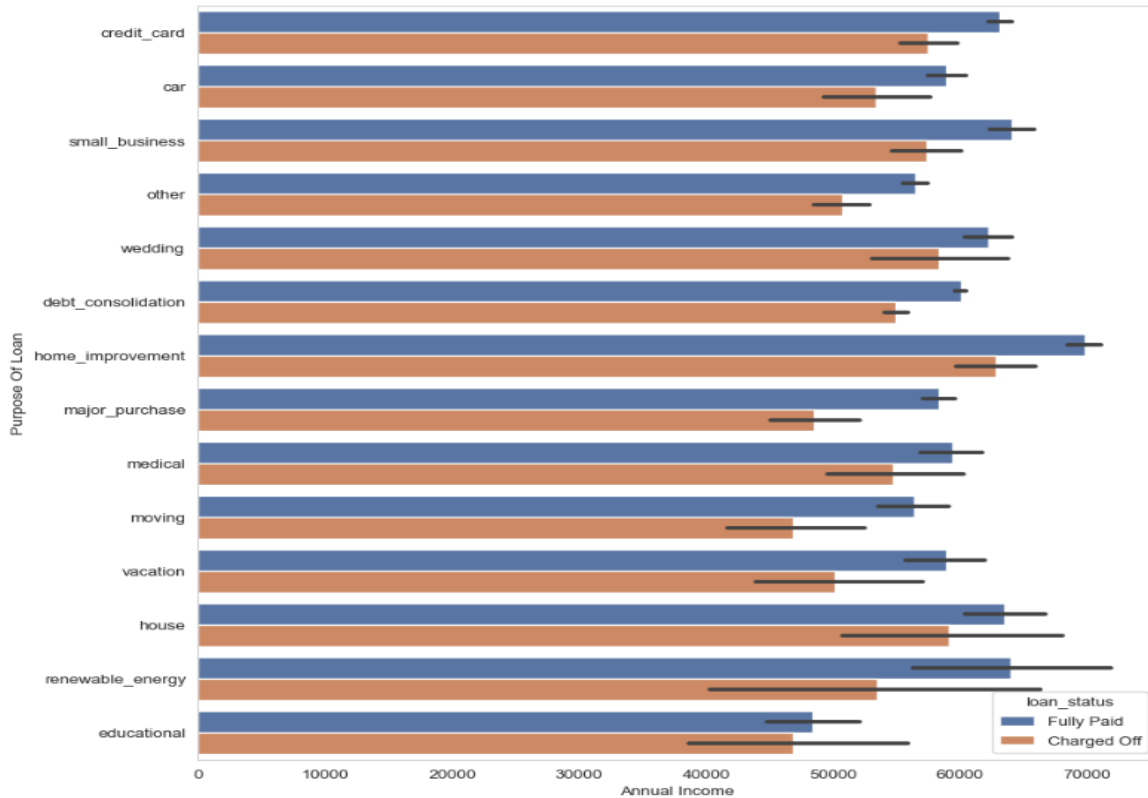


DTI over grade for loan status



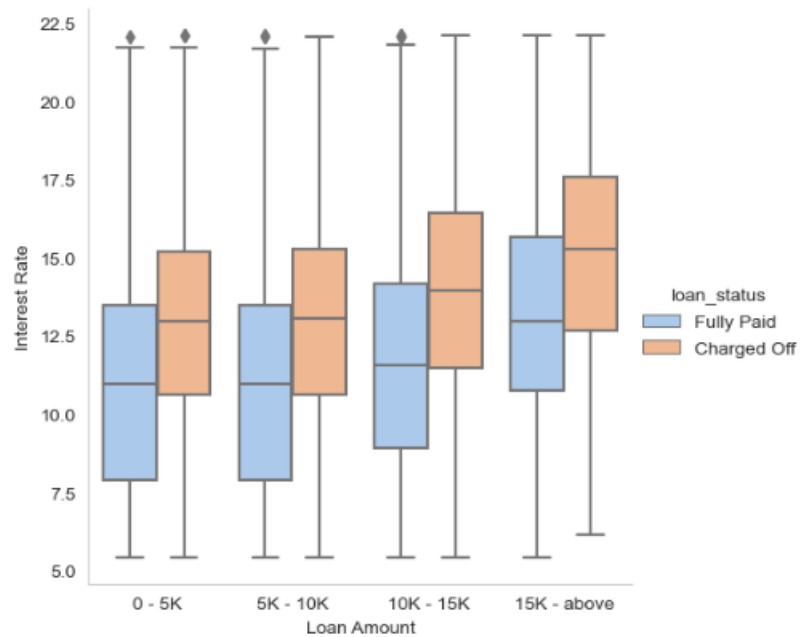
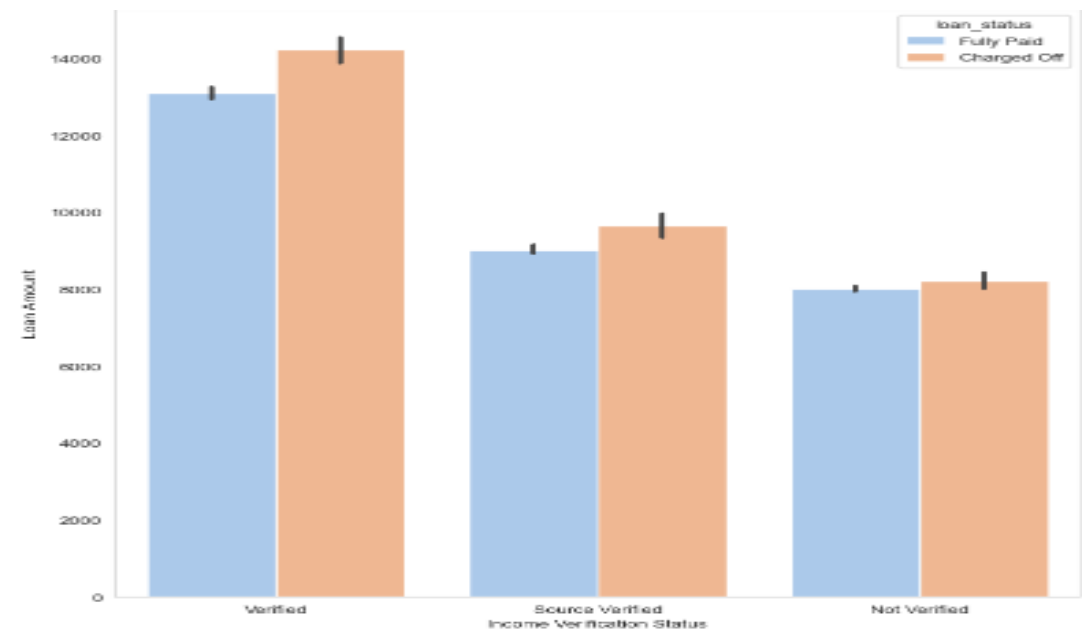
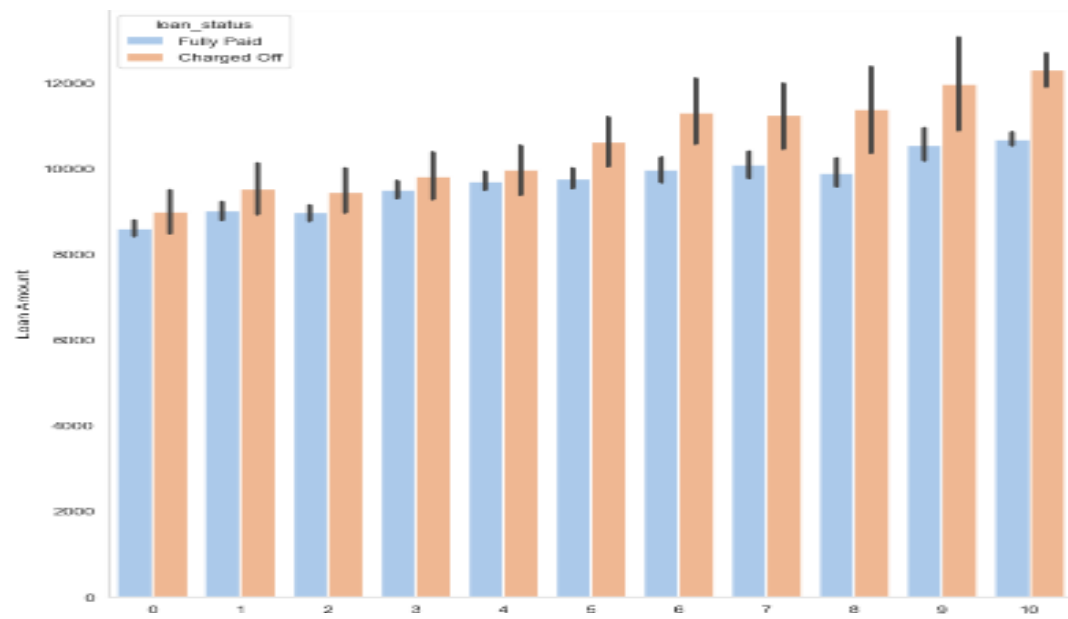
Inference:

- The Above chart show Grade as Risk Factor, As the Grade Decrease, Interest Rate Increase
- Grade A has Lower risk factor as well as lower rate of defaulted Loan



Inference:

- Though the number of loans applied and defaulted are the highest in number for "debt_consolation", the annual income of those who applied isn't the highest. Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses“.
- Loan Applicant with high income have high chance of loan fully paid.



Inference:

- Employees with longer working history got the loan approved for a higher amount.
- Looking at the verification status data, verified loan applications tend to have higher loan amount. Which might indicate that the firms are first verifying the loans with higher values.
- The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan_amount groups.
- This can be a pretty strong driving factor for loan defaulting.

Recommendations on the basis of analysis

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

- ☐ Annual income
- ☐ DTI
- ☐ Grades
- ☐ Pub_rec_bankruptcies
- ☐ Verification Status

Other considerations for 'defaults' :

- ☐ Applicants with working experience 10+ years.
- ☐ Applicants having annual income in the range 50k-100k.
- ☐ Applicants having Public Recorded Bankruptcy.
- ☐ Applicants with least grades like E,F,G which indicates high risk.
- ☐ Applicants with very high Debt to Income value.