Lending Club Case Study

Problem Statement

A consumer finance company specializing in lending various loans to urban customers. When the company receives a loan application, the company has to decide on 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 for the company
- If the applicant is not likely to repay the loan, i.e. he/she is expected to default, then approving the loan may lead to a financial loss to the company

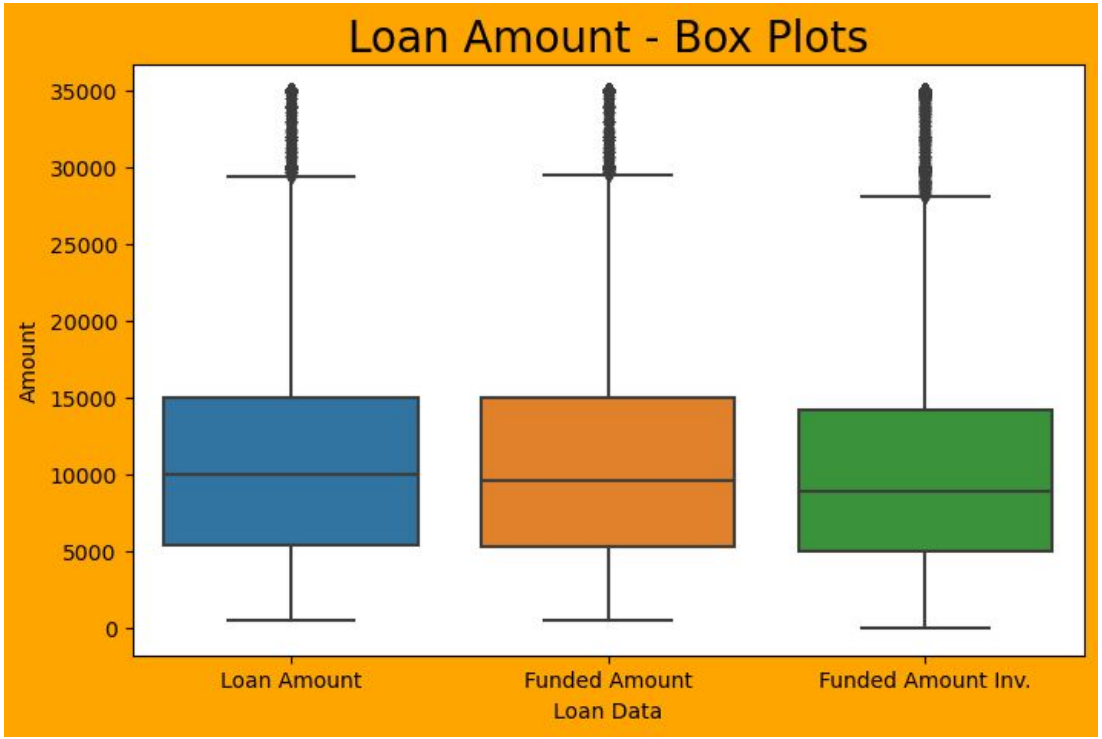
Appproach

- **Univariate Analysis**
- **Segmented Univariate Analysis**
- **Bivariate Analysis**

Univariate Analysis

Univariate analysis - Loan amount, Funded amount and Funded amound inv

Loan Amount - Box Plot



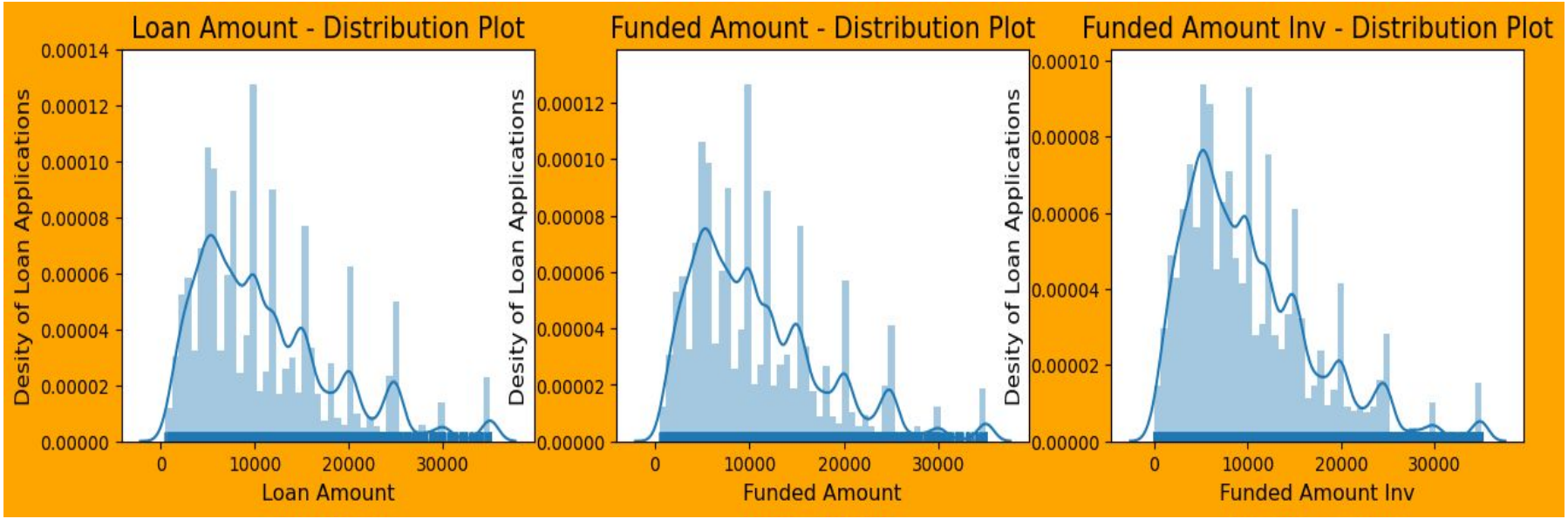
Univariate analysis - Loan amount, Funded amount and Funded amound inv

Observations

- Loan amount, funded amount and funded amount inv. shows almost the same pattern of data spread in box plots.
- Data near whisker line are also closely placed to the line.
- We can conclude that there are no outliers here

Univariate analysis - Loan amount, Funded amount and Funded amound inv

Distribution Plot

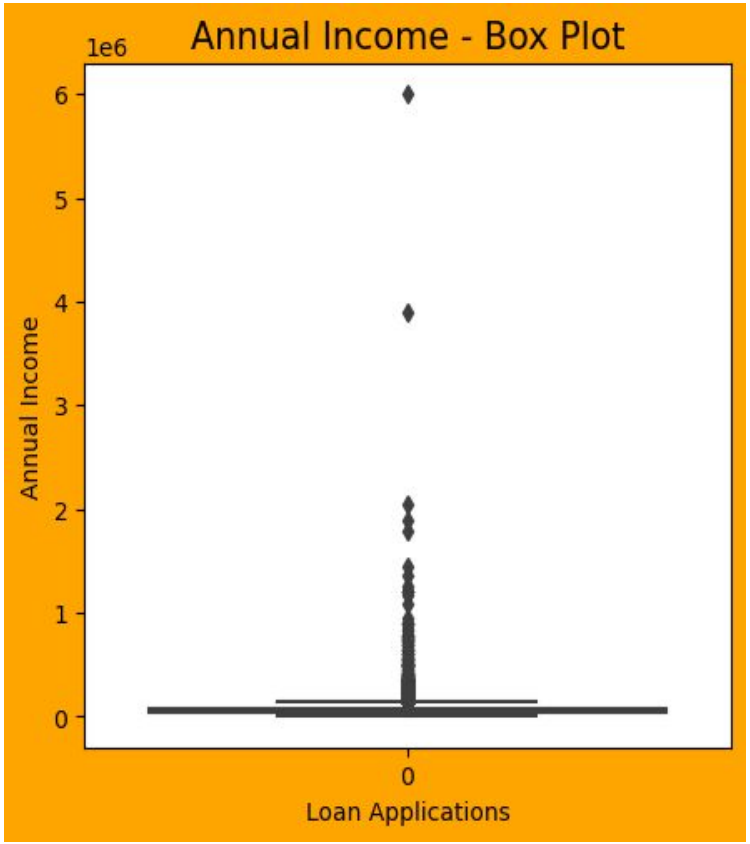


Univariate analysis - Loan amount, Funded amount and Funded amound inv

Observations

- There is almost similar pattern of data spread for all the three parameters. Funded Amount Inv shows a little higher curve near 10000 but it can be ignored.
- We can consider Loan amount field for further analysis against loan amount.

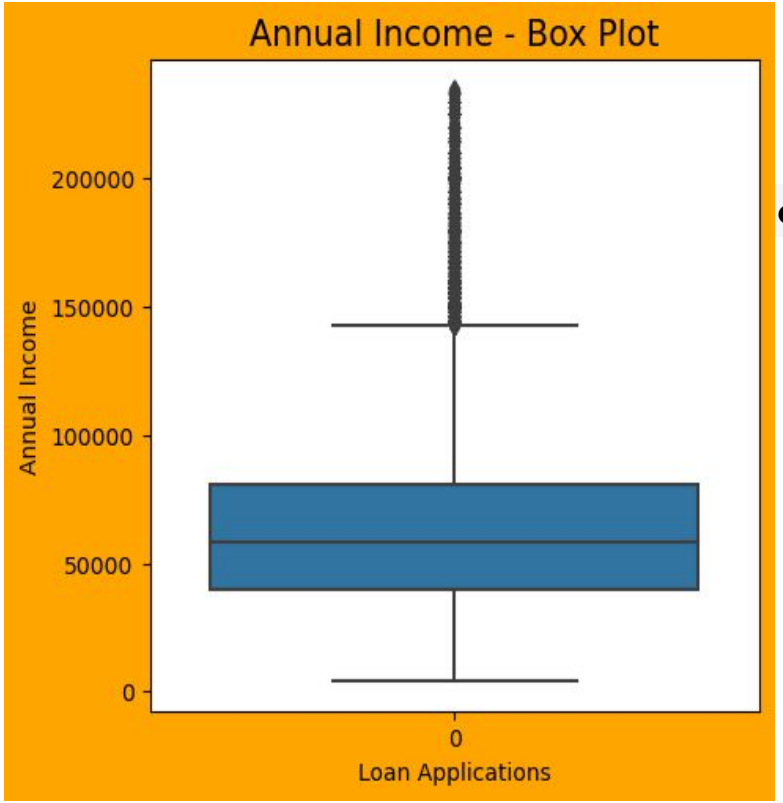
Univariate analysis - Annual income - Box Plot



Observations

- From the above box plot we can say that there are outliers in annual income. Some applicant's annual income is shown unexpectedly high.
- This can impact the analysis adversely. So we need to remove outliers.

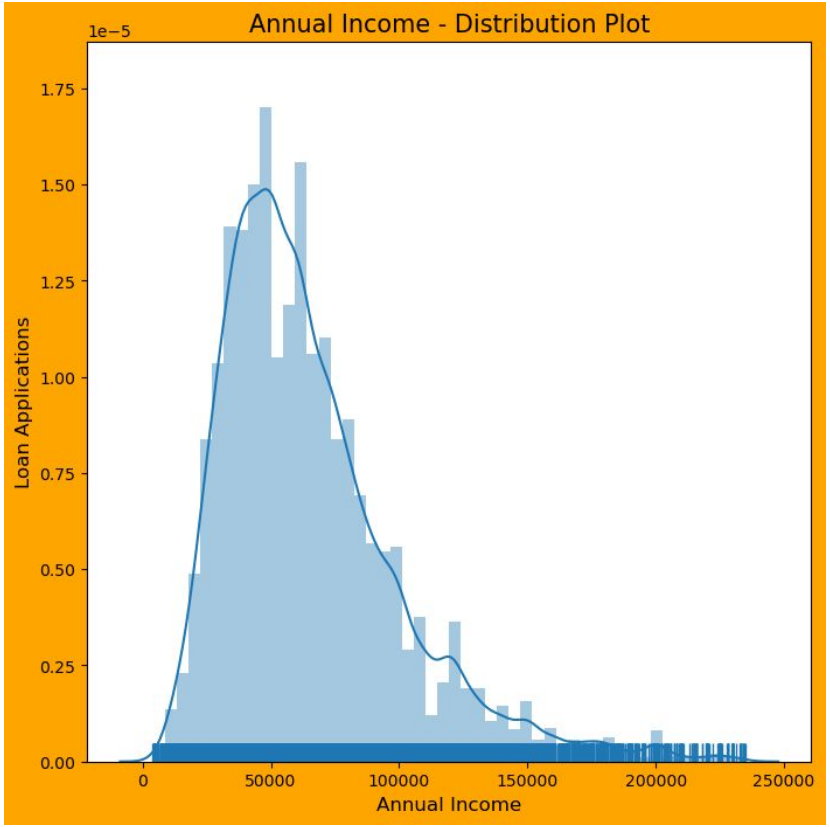
Annual Income - Box Plot (outliers dropped)



Observations

- After removing outliers we can see that most of the loan applicant's annual income is between 5000 - 10000

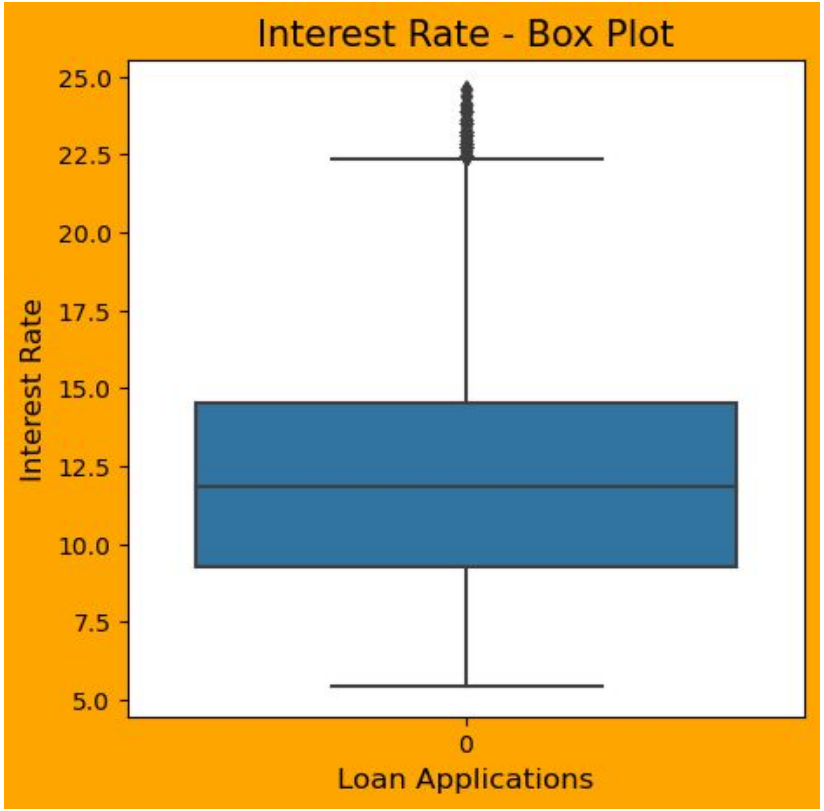
Annual Income - Box Plot (outliers dropped) - Dist. Plot

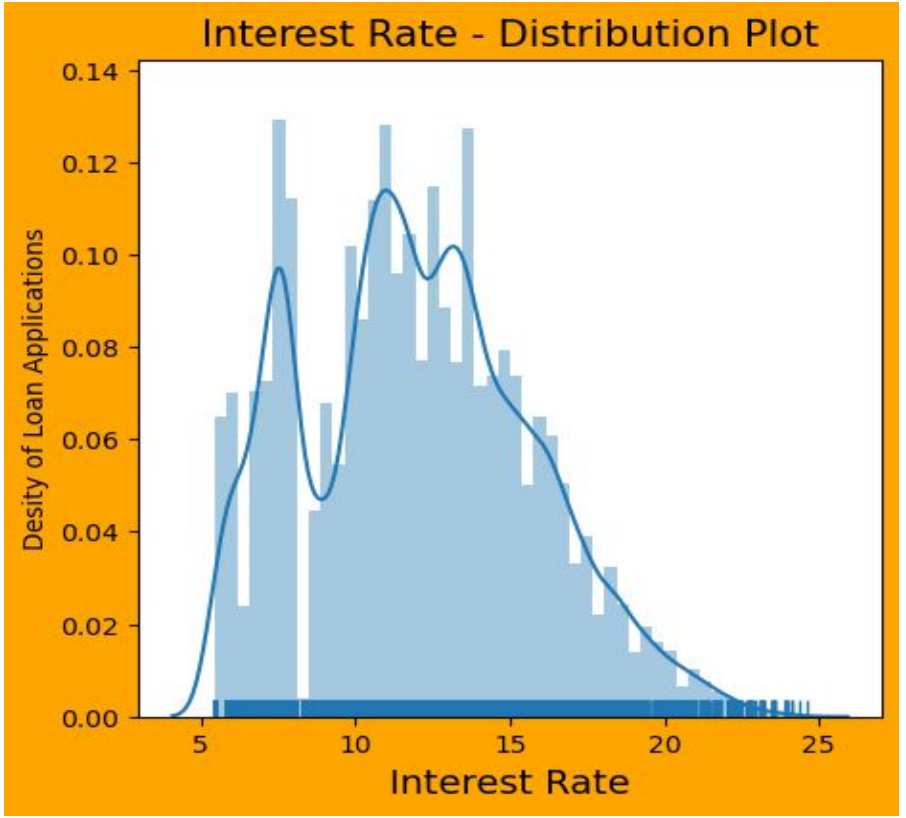


Observations

- As we observed in box plot, most of the loan applicant's annual income is between 5000 - 10000.
- As the annual income is increasing, it shows the drop in loan applications or say requirements of the loan.

Interest Rate - Box Plot and Distribution Plot

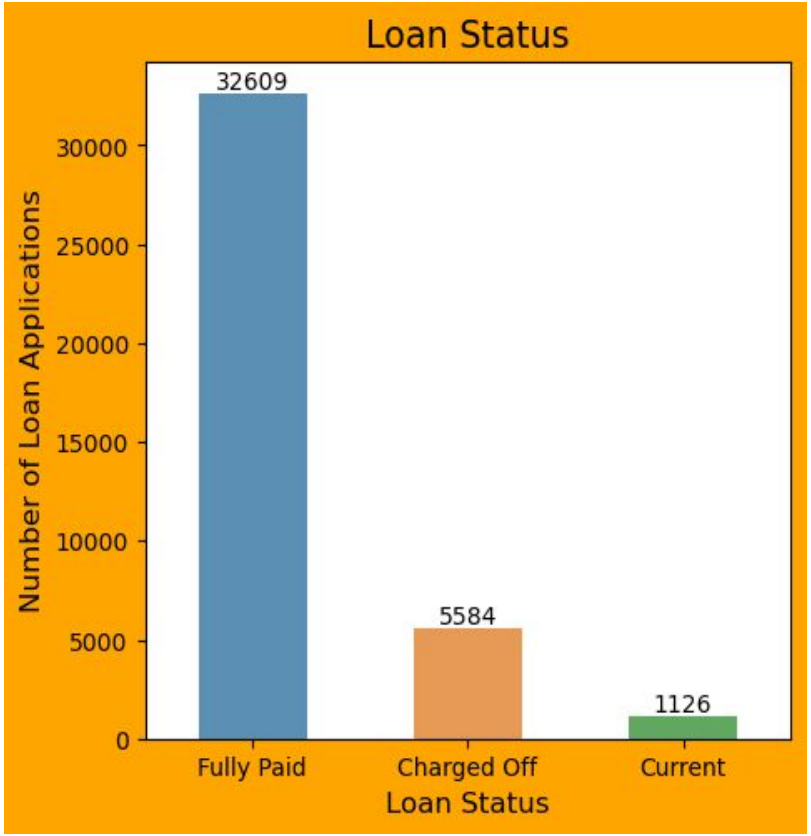




Interest Rate - Observations

- From box plot we can see there are some loans at higher interest rate between 22% - 25%. But we will not consider them outliers. It depends on the applicant’s profile.
- Form box plot and distribution plot, we can see most of the loans interest rate lies between 9% - 15%.

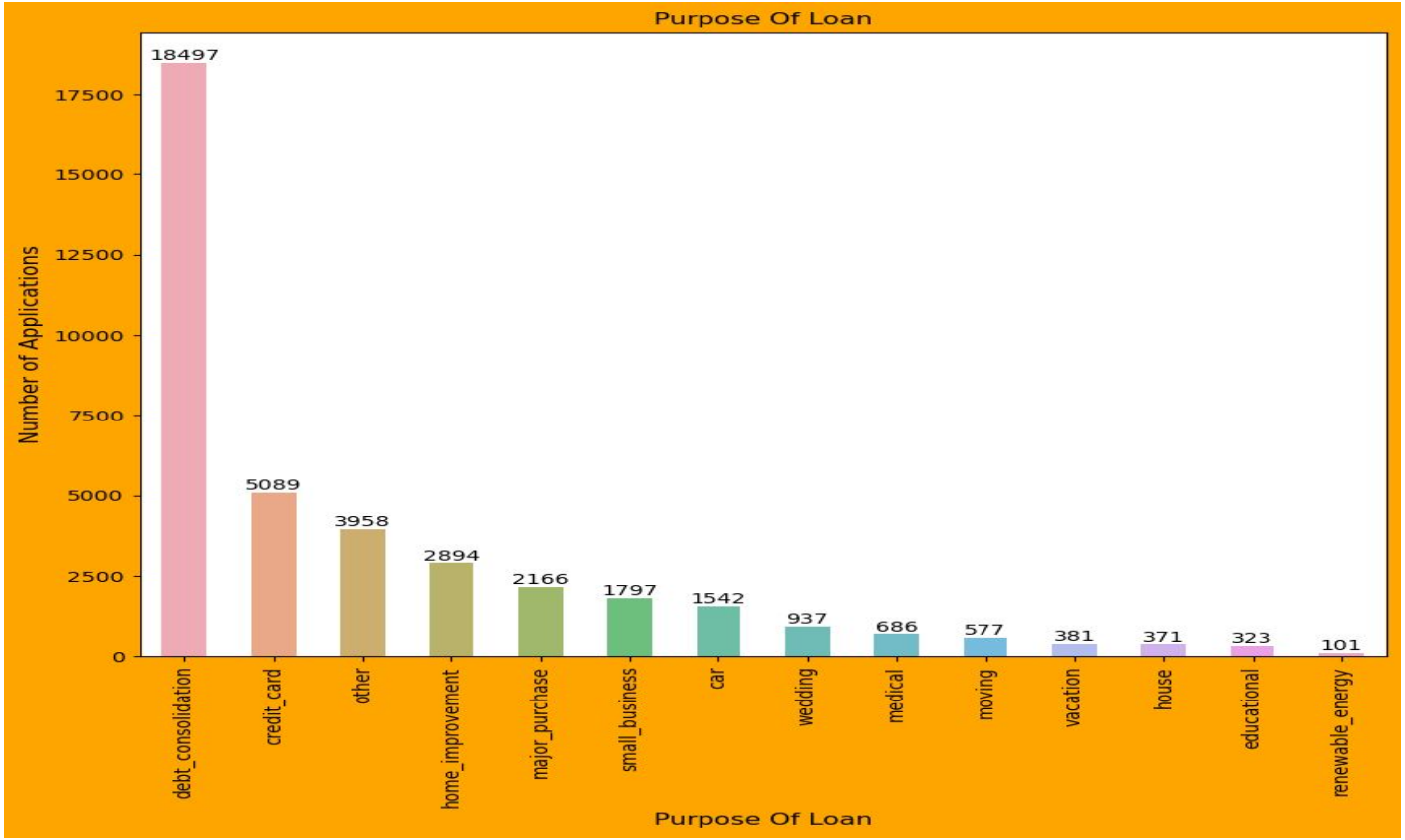
Loan Status - Box Plot



Observations

- There are three types of loan statuses can be seen throughout the data.
- Maximum loans are fully paid, but there is a significant number of charged off loans also.
- Our analysis will be based around the behaviour of charged off loans vs different parameters. It will help in identifying the risky applicants who are more likely to default.

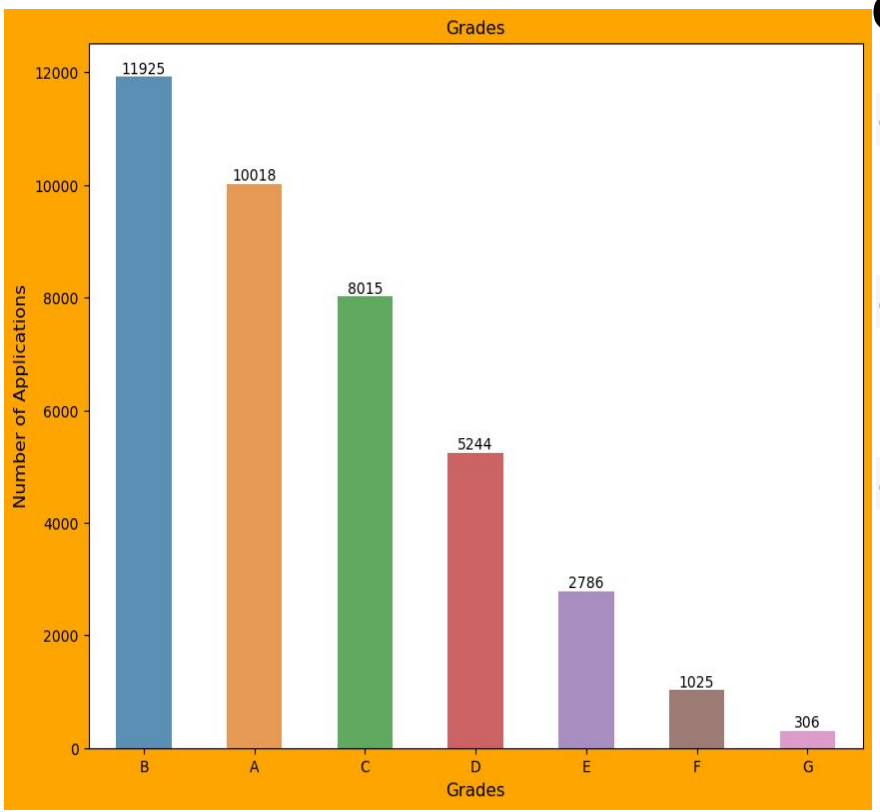
Purpose of the Loan - Bar Chart



Purpose of the Loan - Observations

- Maximum loans are taken for debt consolidation and credit cards bills repayment.

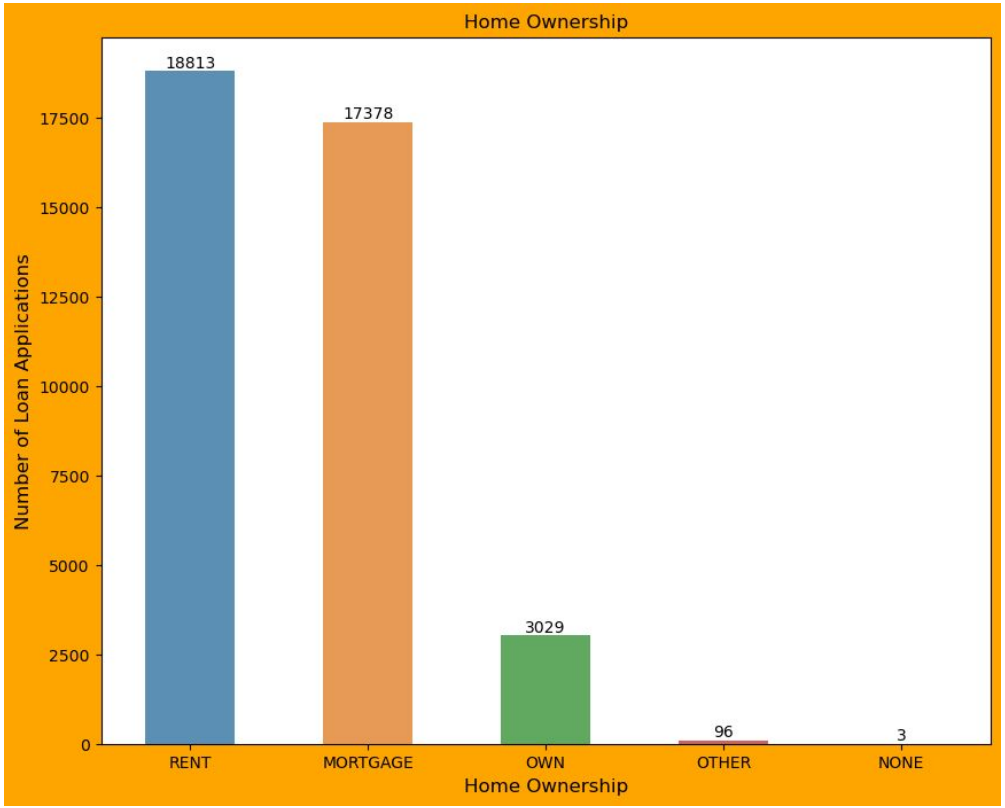
Grade - Bar Chart



Observations

- Grades are directly related to the interest rate. As you move from A to G, interest rate also increases.
- It is exactly aligned with the distribution chart of interest rate, lwhere loans are less for less than than 9% rates and greater than 15% rates.
- So we can say Grade A applies to lower interest rates, B is for 9% -15% interest rates and so on.

Home Ownership - Bar Chart

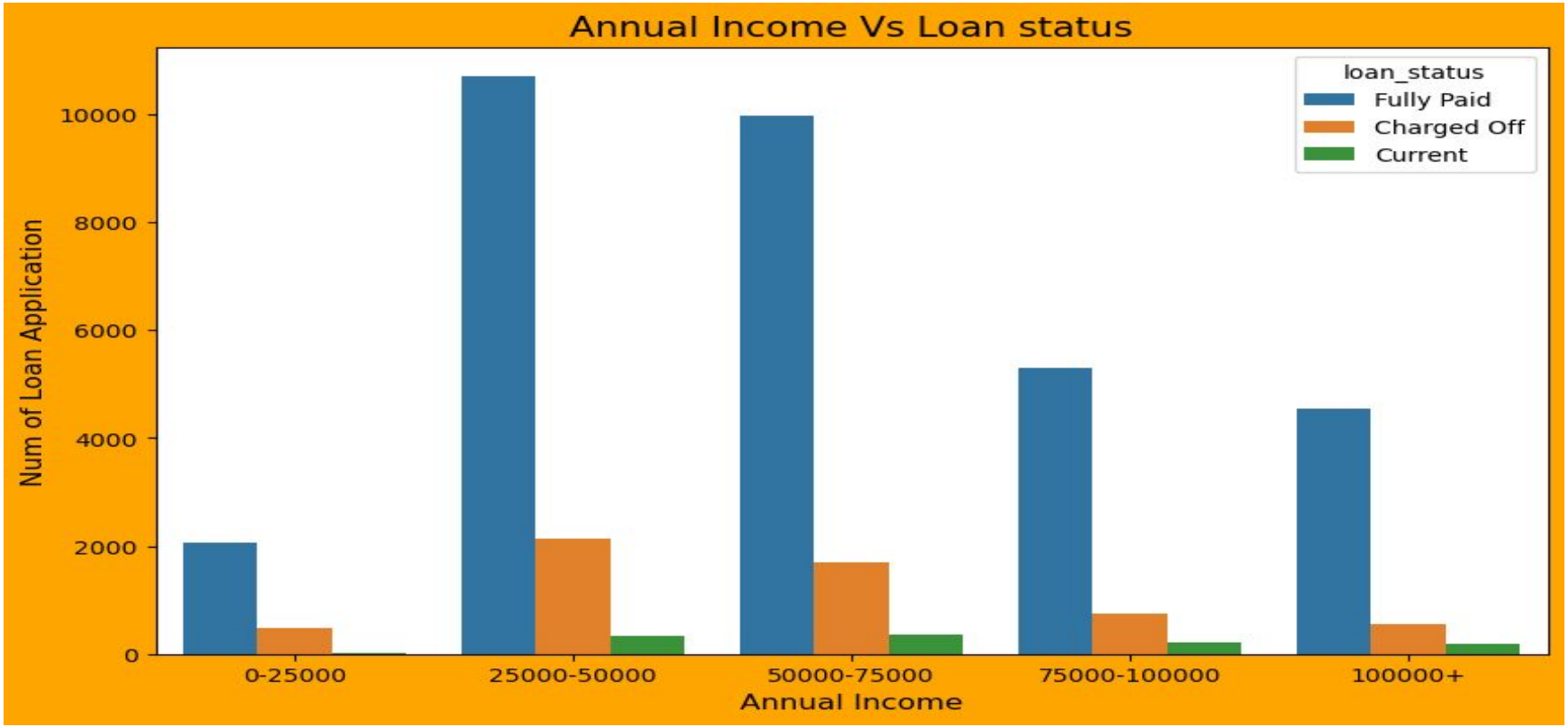


Observations

- Maximum applicants are staying in rented or mortgaged houses.

Segmented Univariate Analysis

Annual Income vs. Loan Status

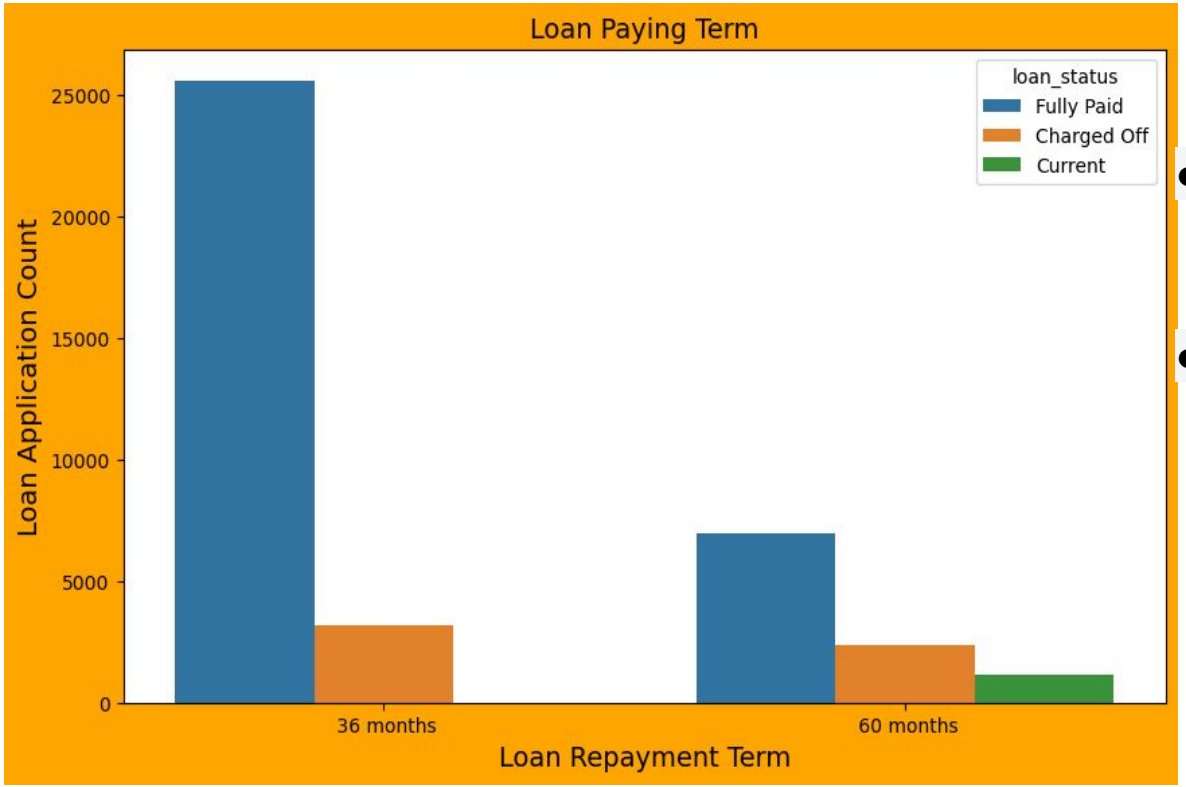


Annual Income vs. Loan Status

Observations

- Compared to distribution chart, it is more clear here that very few loan are approved for lower income applicants.
- Maximum loans are approved for the income range 25000-50000 and 50000-75000.
- As the income increases, it can be assumed that those applicants are less.

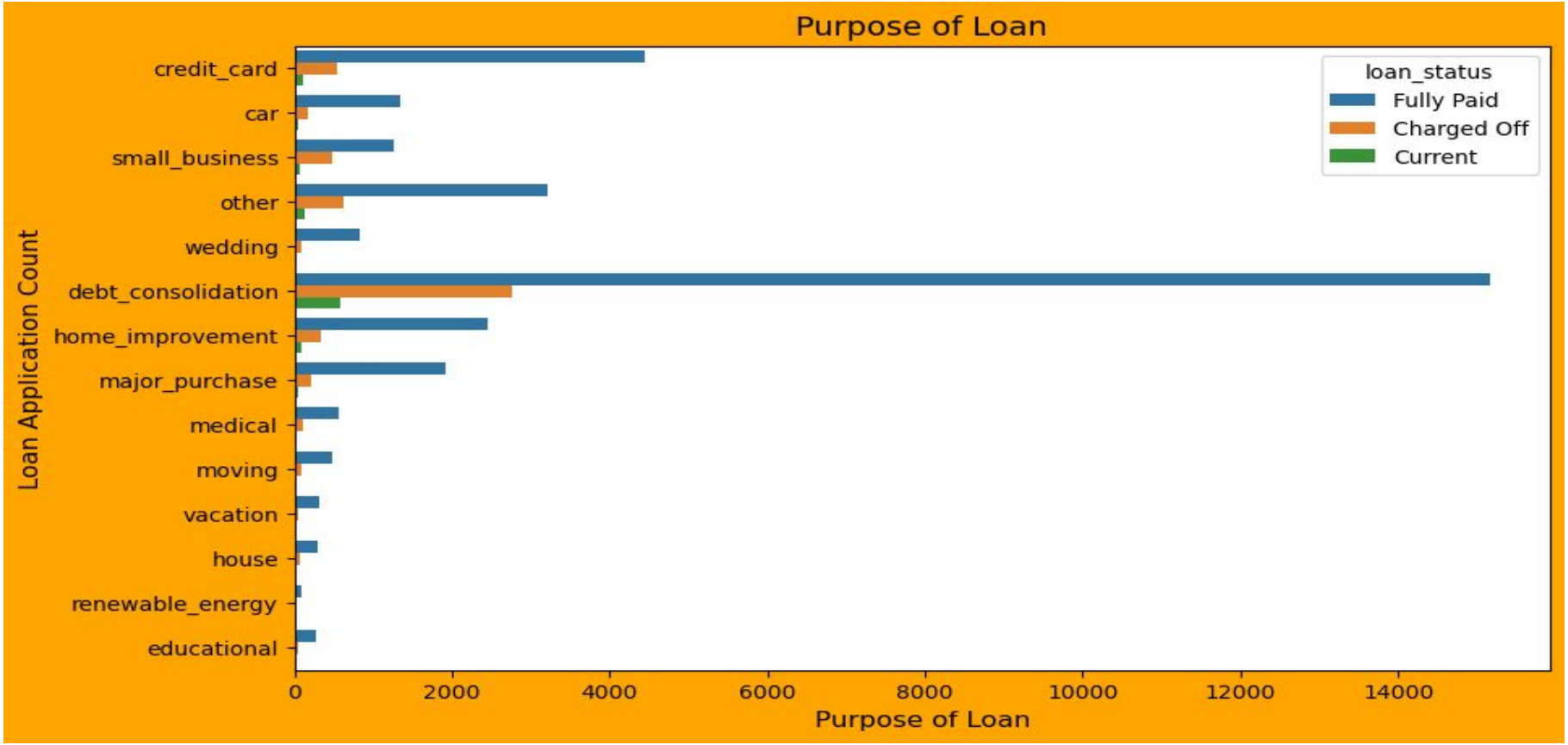
Term vs. Loan Status



Observations

- It shows longer the term for repayment, more are the chances to default the loan.
- Here with 60 months of repayment duration has more Chareged Off laons compared to the total loans in that segment.

Purpose vs. Loan Status



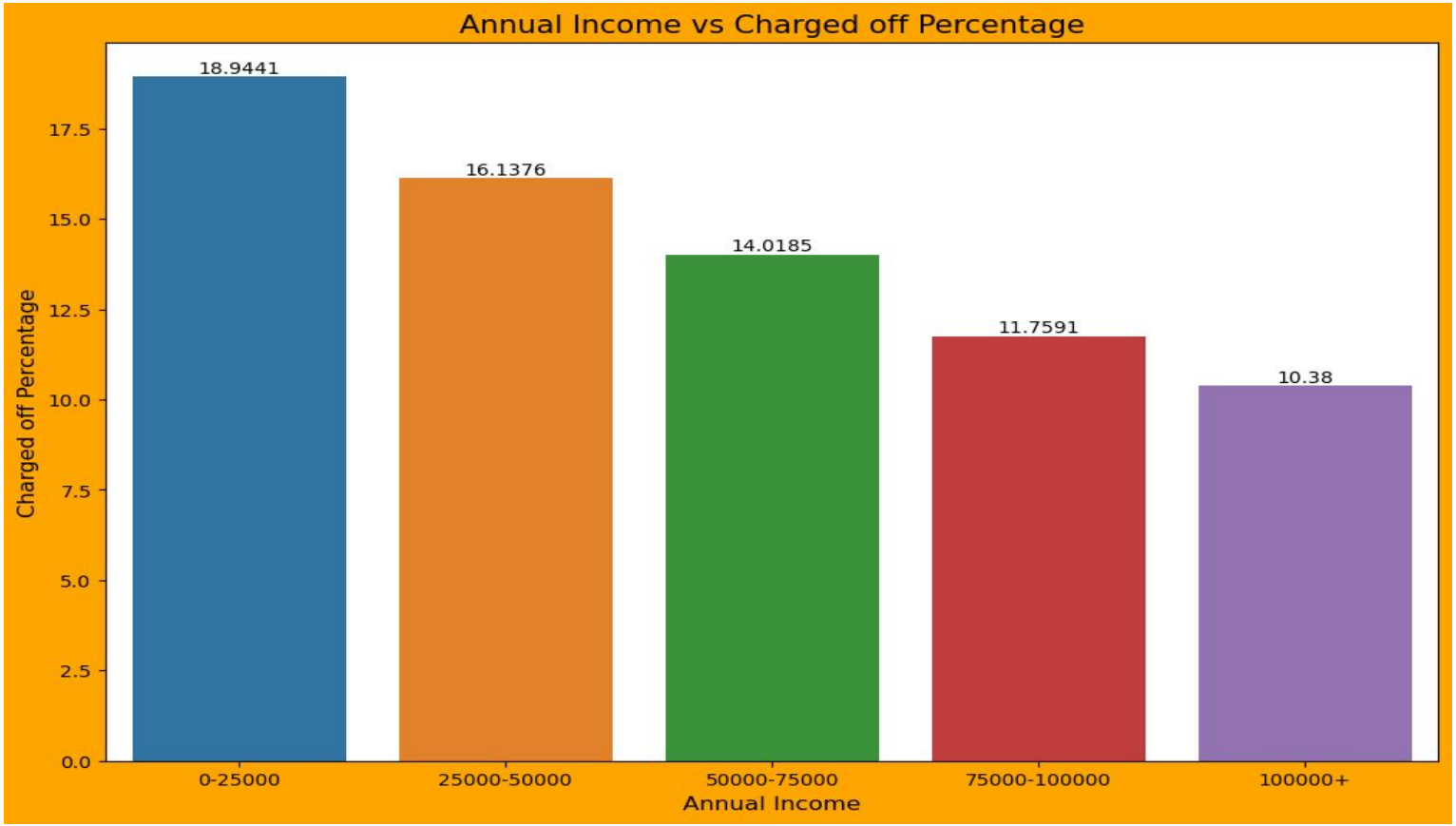
Purpose vs. Loan Status

Observations

- Maximum loans are taken for debt consolidation and creadit card bills payment.
- So as the numbe of charged-off loans are also high for both.
- If loan is taken for settling other dues or debts, there are high chances to be charged-off.

Bivariate Analysis

Annual Income vs. Charged-off Percentage

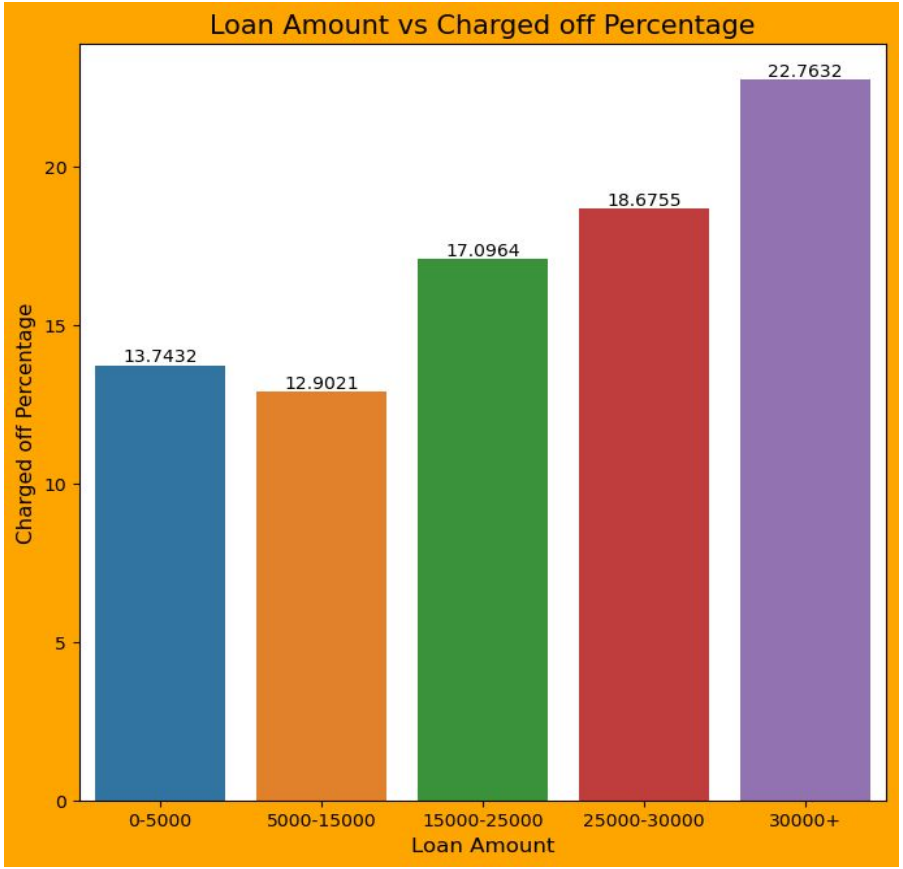


Annual Income vs. Charged-off Percentage

Observations

- It is clearly visible that there are more applicants in the income range of 0-25000 (there is no zero income for any applicant) who default to repay the loan compared to higher income ranges.
- As the annual income increases the probability of defaulting the loan is decreasing.
- It is safe to lend money to applicants with higher annual income

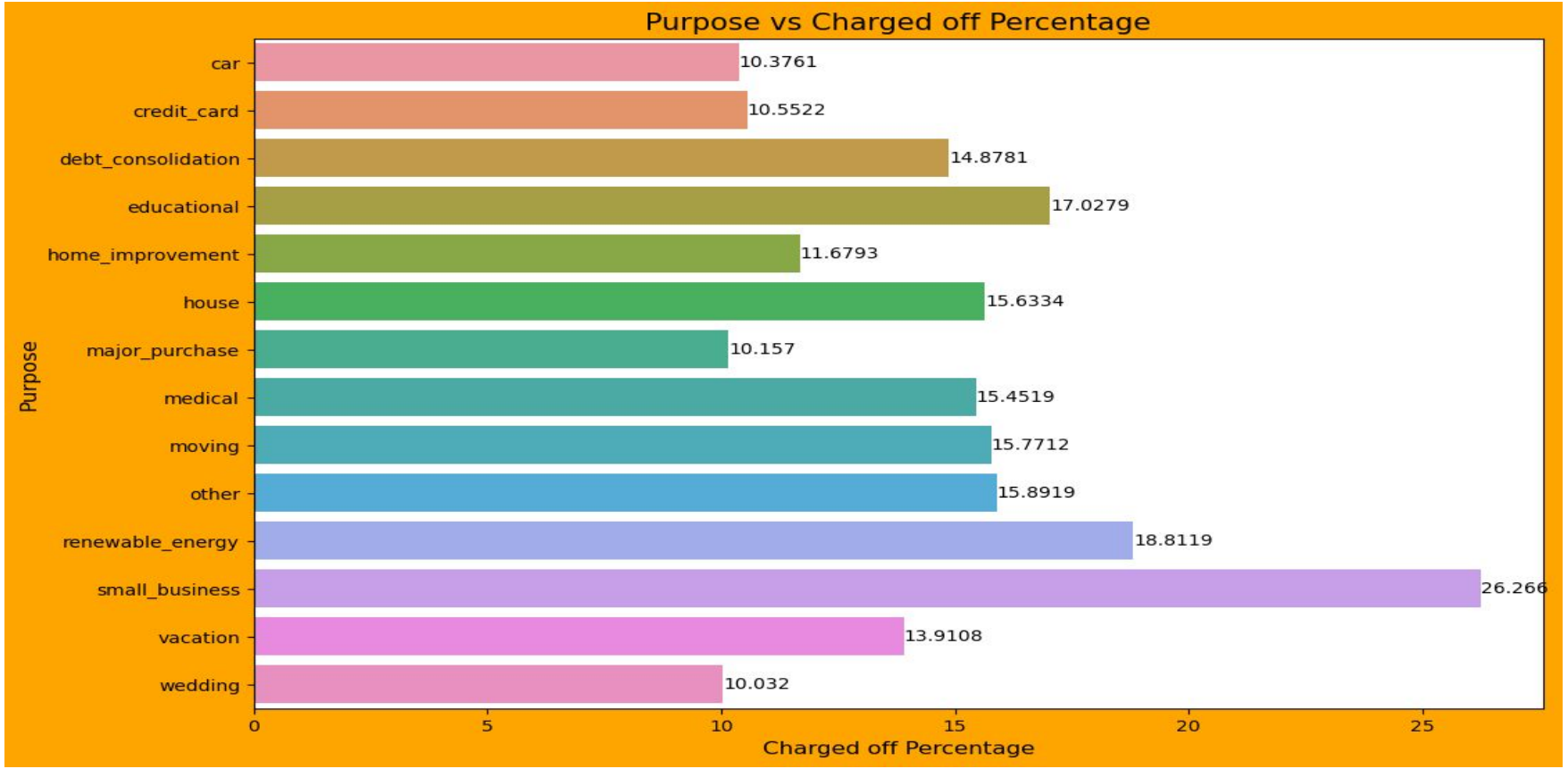
Loan Amount vs. Charged-off Percentage



Observations

- As the loan amount is increasing the charged-off rate is also increasing. For now data is very less on higher loan amount like 30000+ loan amount to conclude.
- But if we compare '5000-15000' and '15000-25000' above observation stands true.
- 0-5000 segment is showing more charged off percentage than 5000-15000, which is significant. These loan would be given to the applicants with low annual income and it is charged-off for low loan amount also.

Purpose vs. Charged-off Percentage

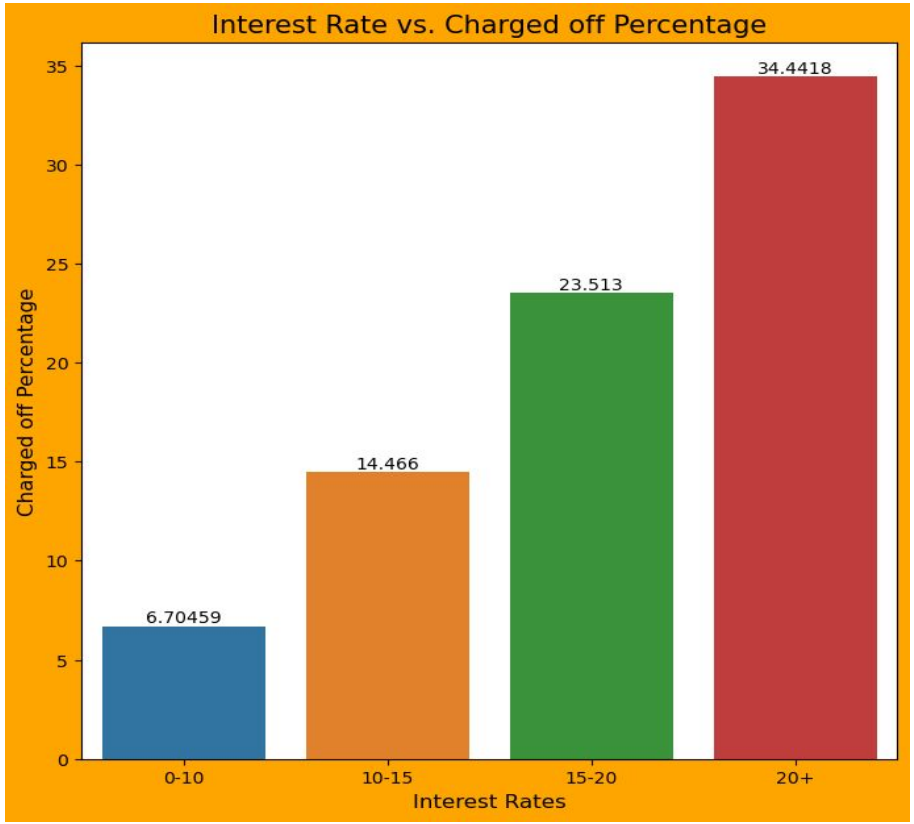


Purpose vs. Charged-off Percentage

Observations

- Small business loans are charged off more. That means small businesses taking loans are at higher risk to be charged-off due to not being able to make immediate returns on their investments or other reasons. But business loan data is too small to conclude.
- Compared to other reasons, "debt_consolidation", "Other" are more applicants and their percentage of charged off are close to 15%, which shows more risk than other purposes. Also the purpose is to pay off the existing debts which is shows the concern about their repayment capability.
- Car, credit card, major purchases and wedding looks more genuine reasons as their percentage of getting charged off is less.

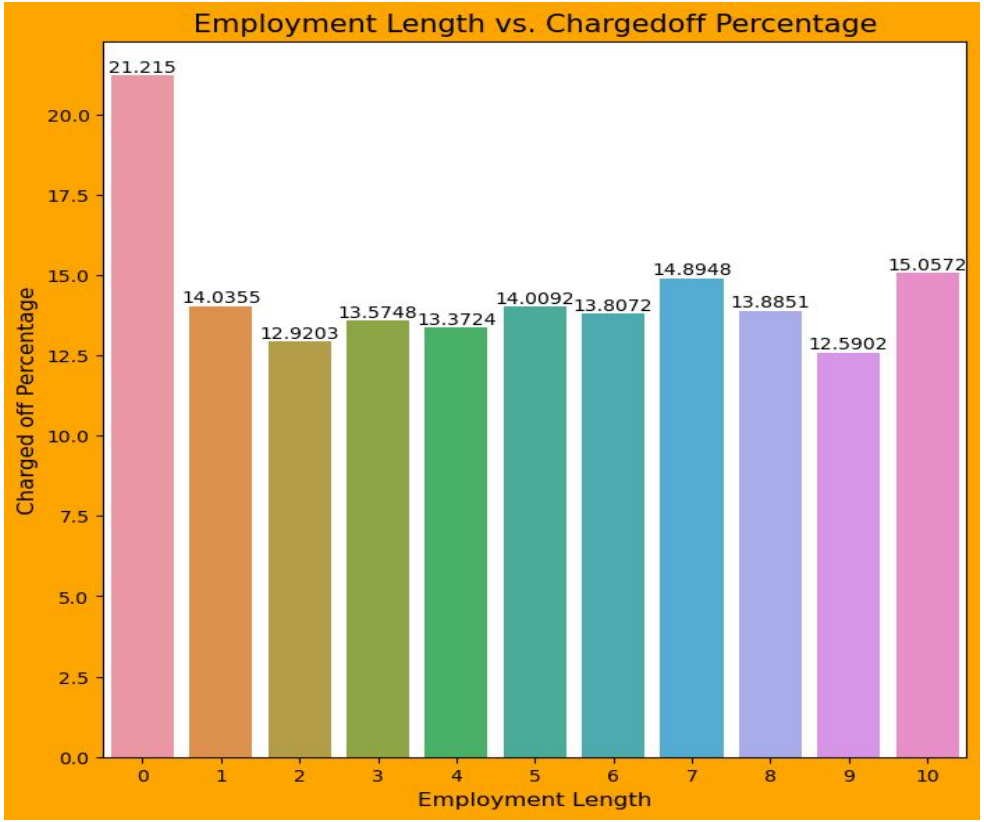
Interest Rate vs. Charged-off Percentage



Observations

- It clearly shows that with the increase in interest rate, the risk of charge-off is also increasing.
- Above 15% interest rate, charged off loans rate is drastically increasing.

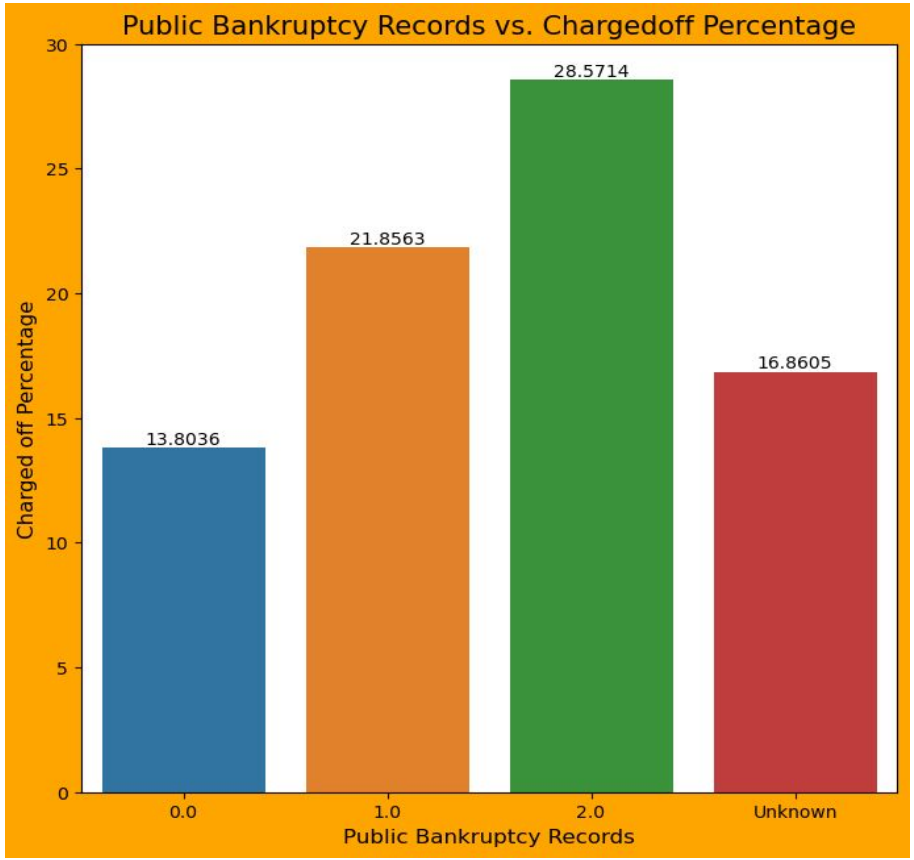
Employment Length vs. Charged-off Percentage



Observations

- If applicant's employment time zero or less than one month, he/she would not be earning and so the chances of charged-off the loan are more.
- Here it is clear that charged off rate is high for those applicants with less than one month of employment.
- For others it is more or less equal chances of charged-off.

Employment Length vs. Charged-off Percentage



Observations

- People who has 0 bankruptcy also defaults the loan. Number of loans are more with 0.
- People with 1 and 2 are having higher percentage of default the laon but numbers are low to conclude.
- But there are high chances of default for those who has record of bankruptcy.

Results From Observations

- Lower the annual income, higher the chances of charged-off .
- Higher the loan amount, higher the chances of charged-off.
- Higher the interest rate, higher the chances of charged-off.
- Higher the repayment term, higher the chances of charged-off.
- When purpose is to settle the other debts then chances are high that the loan will be charged-off.
- For new businesses also, possibilities are high to default as ROI is not guaranteed on new businesses.
- Applicants with no employment history may have more chances of default the loan.
- Applicants having bankruptcy records are having more chances to default on the loan.

Glossary

Data Source Details

Source of Data: loan.csv

Format: csv

Number of Rows: 39717

Each row is: giving the details of the existing loan applican’t characteristics.

Sampling Method: All the loan applications between Apr-2008 and Sep-2011

Technologies

Programming Language: Python, version: 3.10.9

Supporting Libraries: Numpy, verson: 1.23.5

Pandas, version: 1.5.3

Charting Tools: Matplotlib, version: 3.7.0

Seaborn: 0.12.2

Thank You