

Lending Case Study

- 
- Submitted by
 - Mohita Mundra
 - ML-62 Batch / February 2024

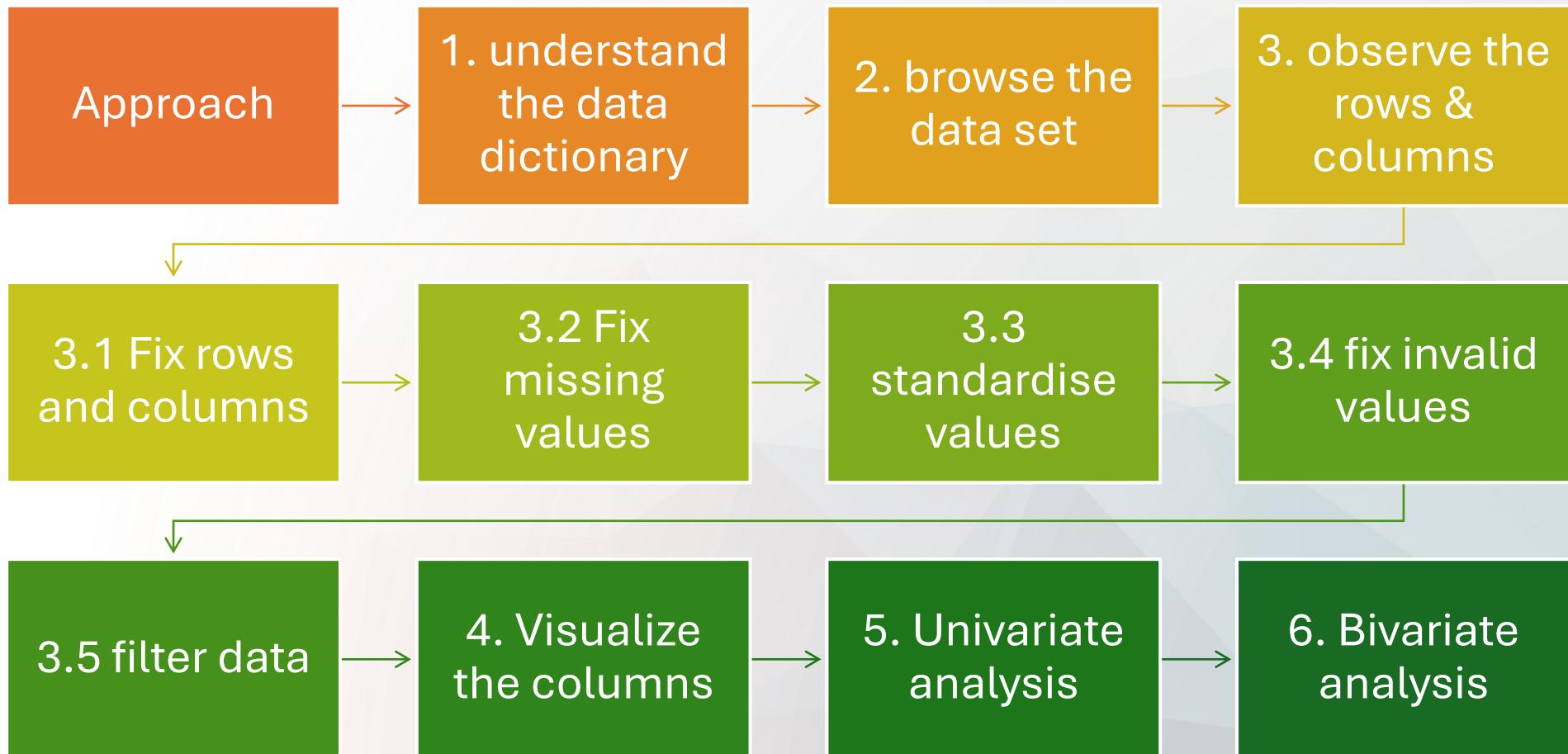
Business understanding

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

Problem statement

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.



Libraries used :

pandas

numpy

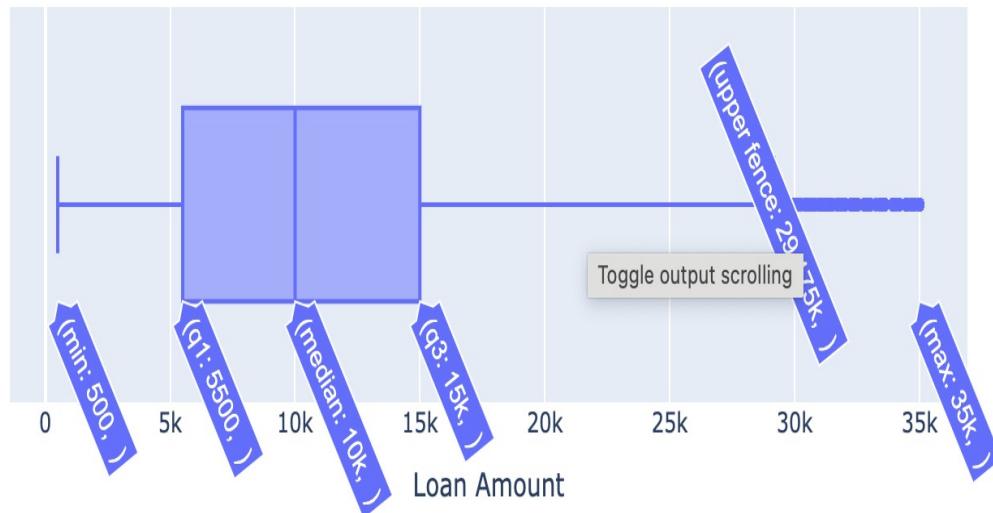
matplotlib

seaborn

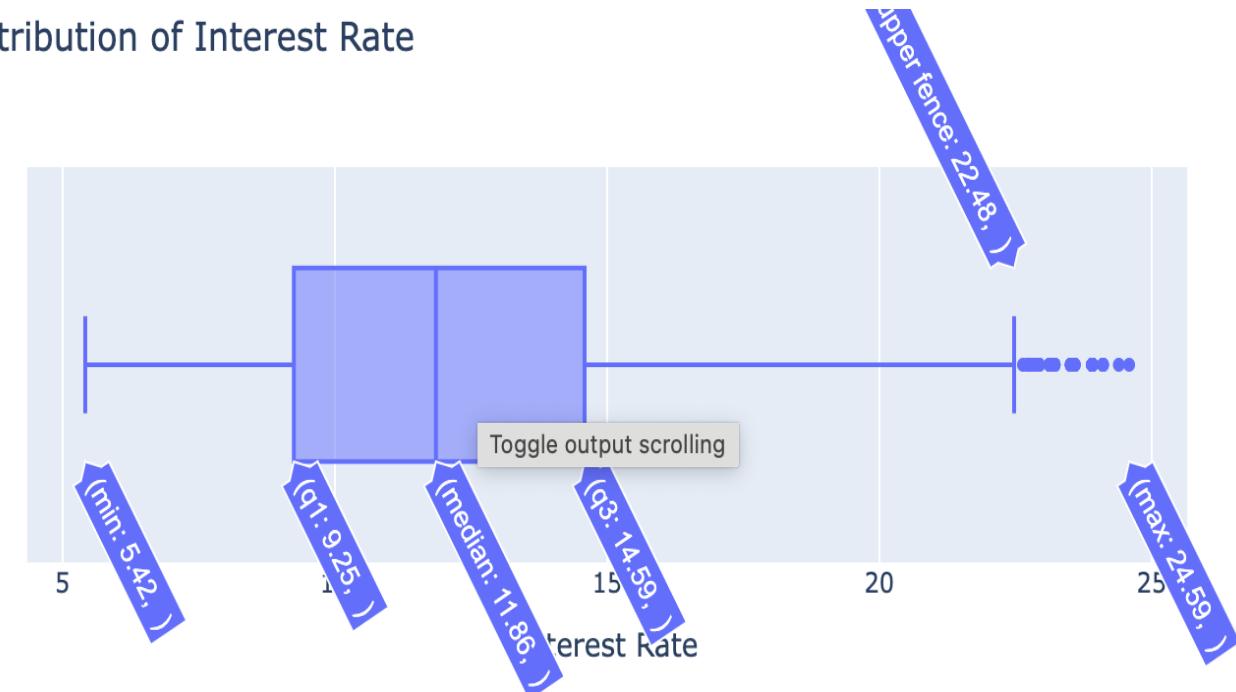
plotly

Identification process to remove outliers: Using Box Plot

Distribution of Loan Amount



Distribution of Interest Rate



Observation:

1. Upper fence turns out to be $29.175k = 29175$ whereas max is $35k = 35000$ which is not much more than upper fence thus will not have much impact on the analysis.

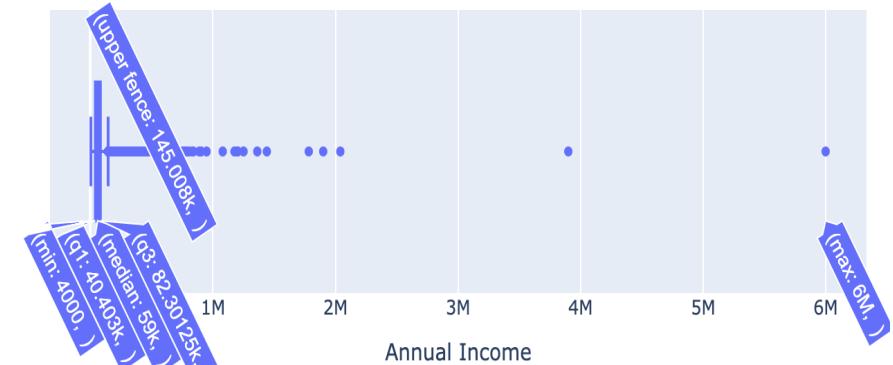
Observation:

1. Upper fence turns out to be 22.64 whereas max is 24.4 which is not much more than upper fence thus will not have much impact on the analysis.

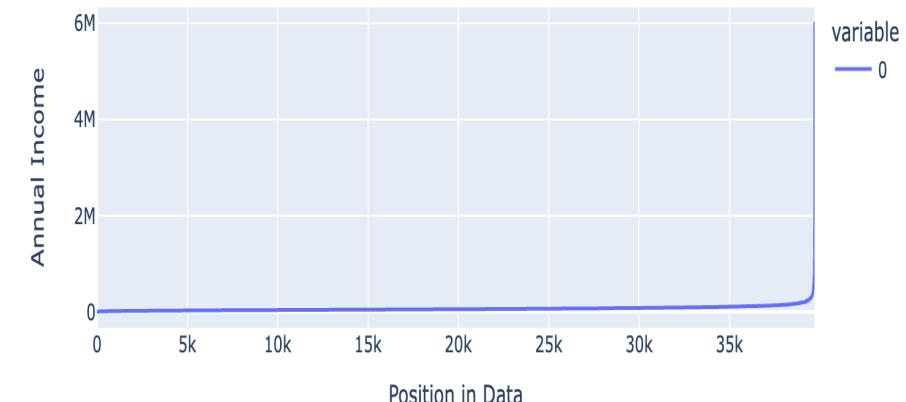
Identification process to remove outliers: Using Box Plot

- Observations:
 - Upper fence turns out to be 146k whereas max is 6000k which is much from upper fence thus we will remove the outliers in column annual_inc.
 - As it can be observed from the line chart, the annual_inc is increasing in exponential format around 99th percentile. Thus, we can remove values greater than 99th percentile.

Distribution of Annual Income of the Burrower

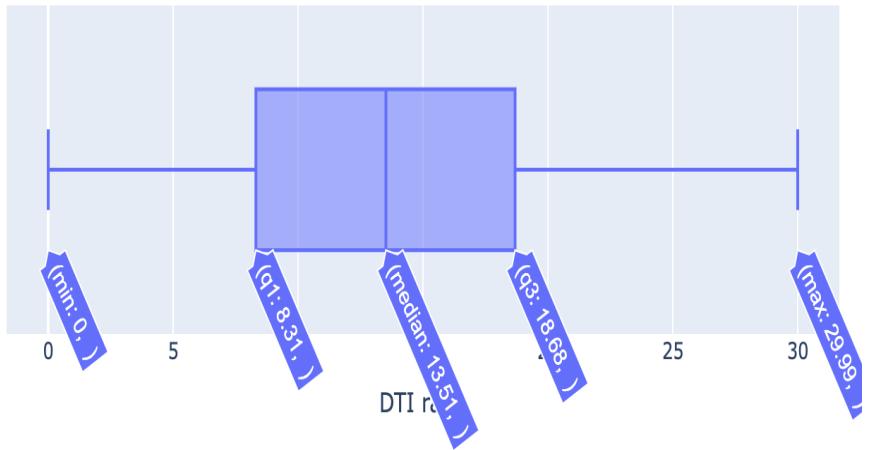


Trend of Annual Income

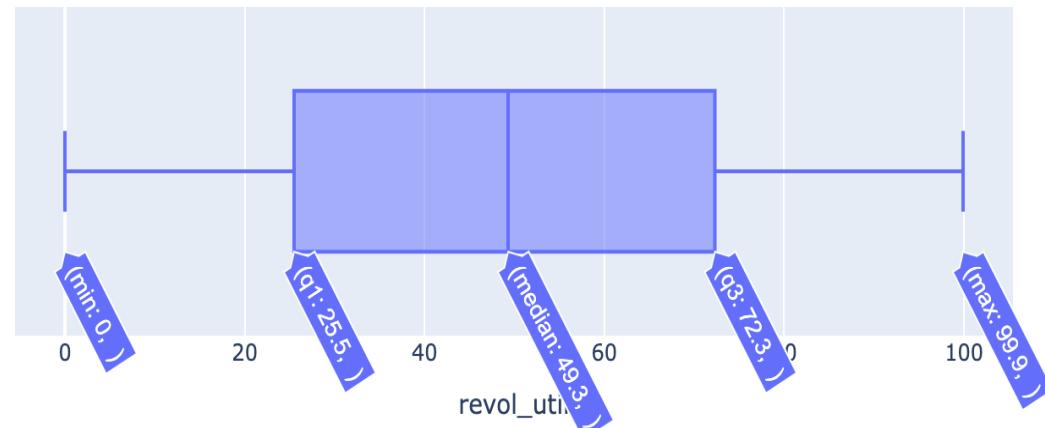


Identification process to remove outliers: Using Box Plot

Distribution of Debt To Income Ratio



the amount of credit the borrower is using relative to all available revolving credit



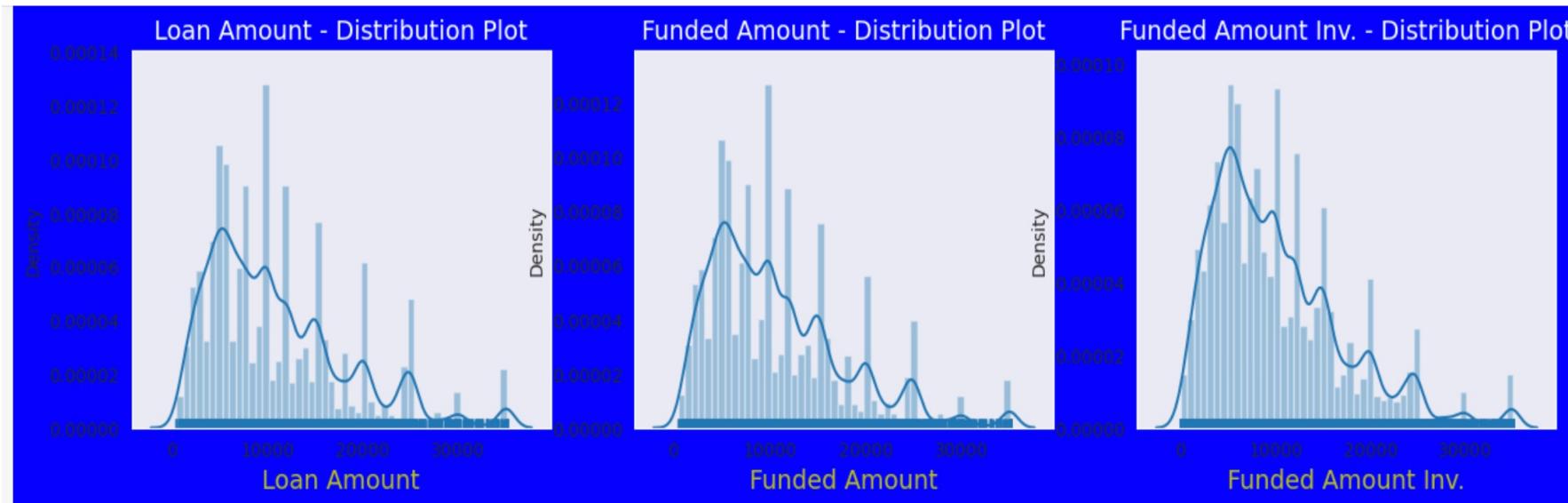
Observation:

There are no outliers in dti hence we can move ahead with analysis.

Observation:

There are no outliers in Revol_util hence we can move ahead with analysis. borrowers whose utilization risk is higher is more likely to default

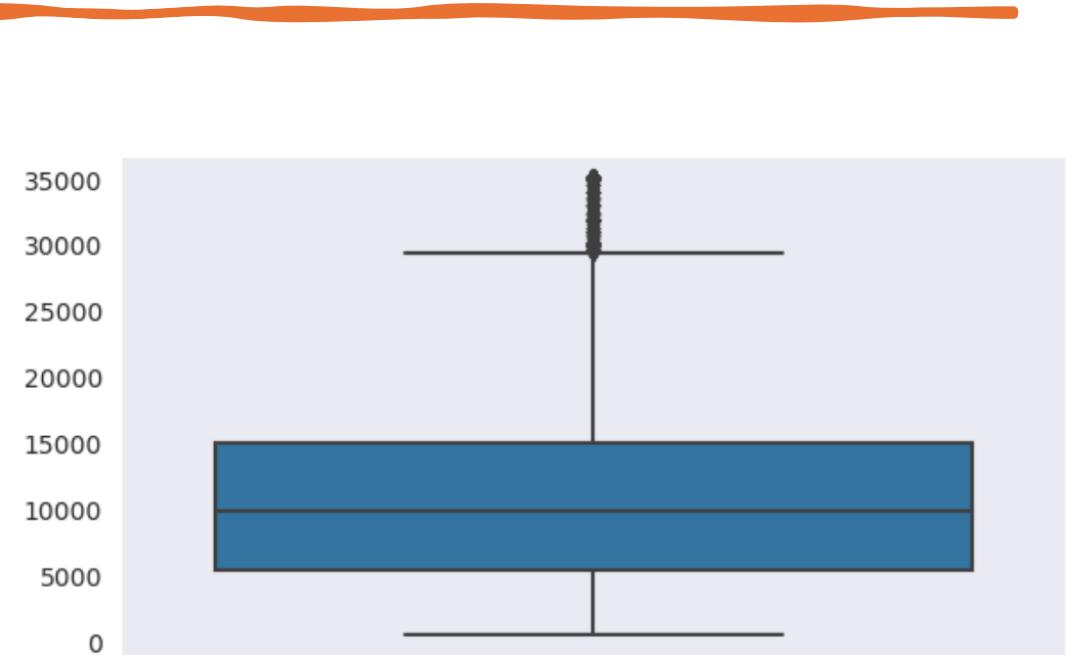
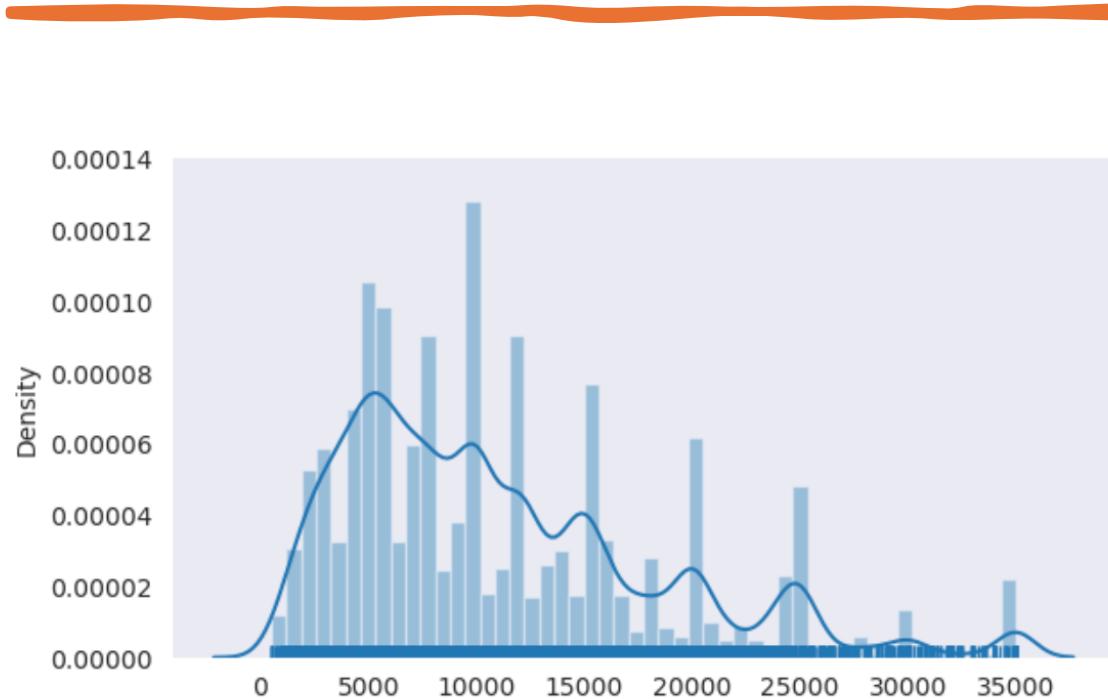
Univariate Analysis: Loan Amount



Observation:

1. Distribution of amounts for all three looks very much similar. We will work with only loan amount column for rest of our analysis

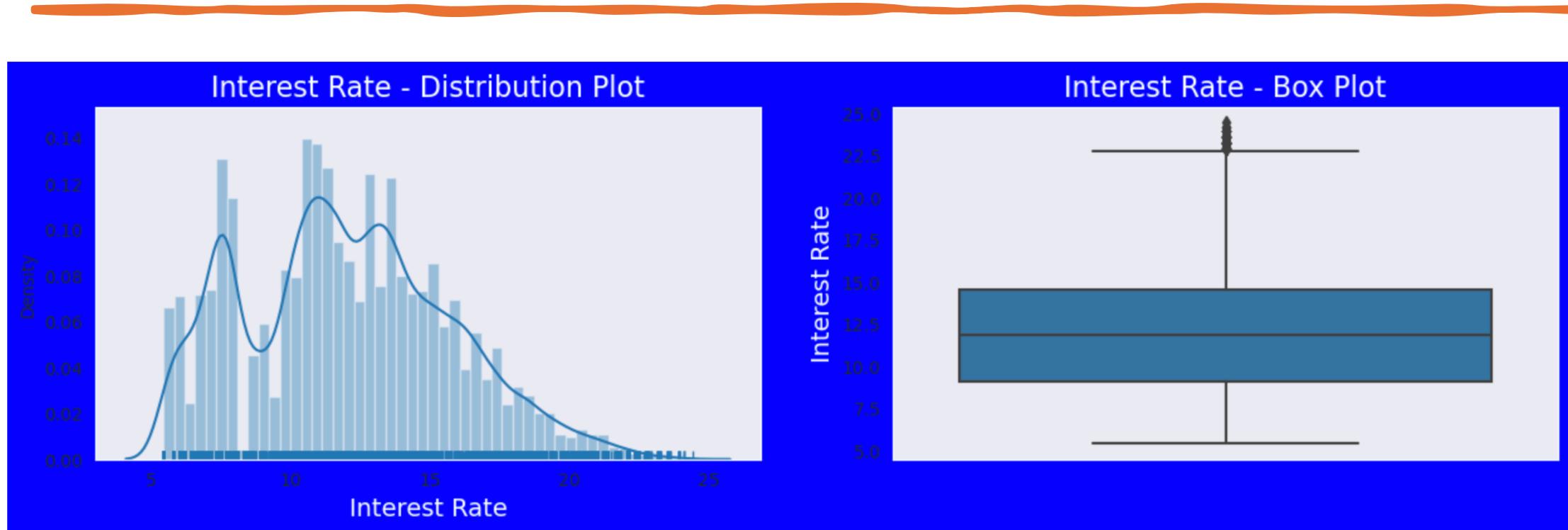
Univariate Analysis: Loan Amount



Observation:

1. plots show that most of the Loan amounts are in range of 5000 - 15000

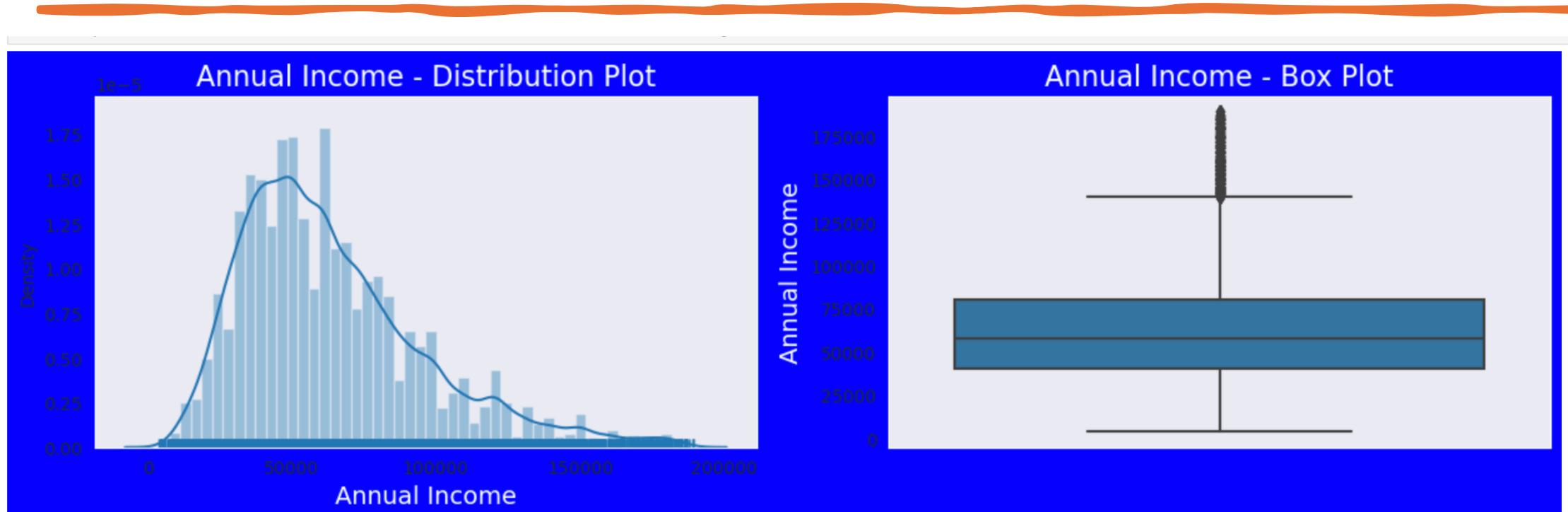
Univariate Analysis: Interest Rate



Observation:

1. plots show that most of the Interest Rates on loans are in range of 10% - 15%

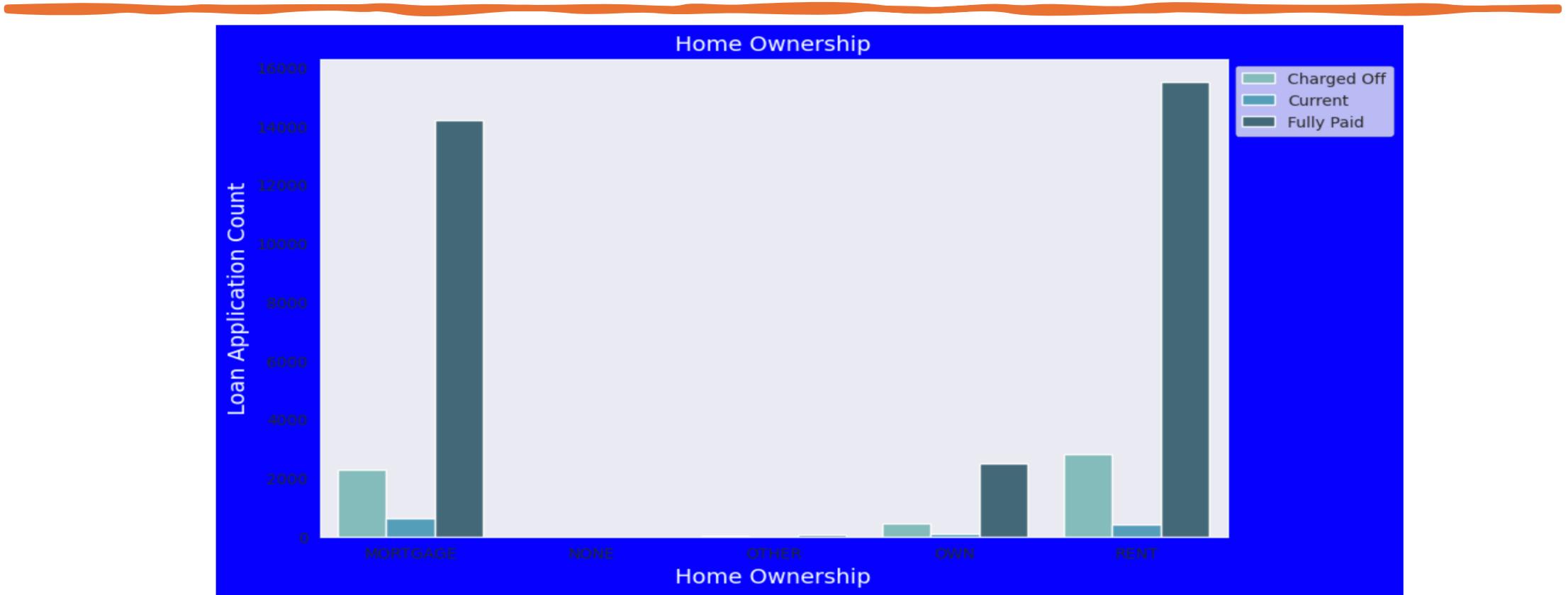
Univariate Analysis: Annual Income



Observation:

1. plots show that most the borrower's Annual incomes are in range of 40000- 80000

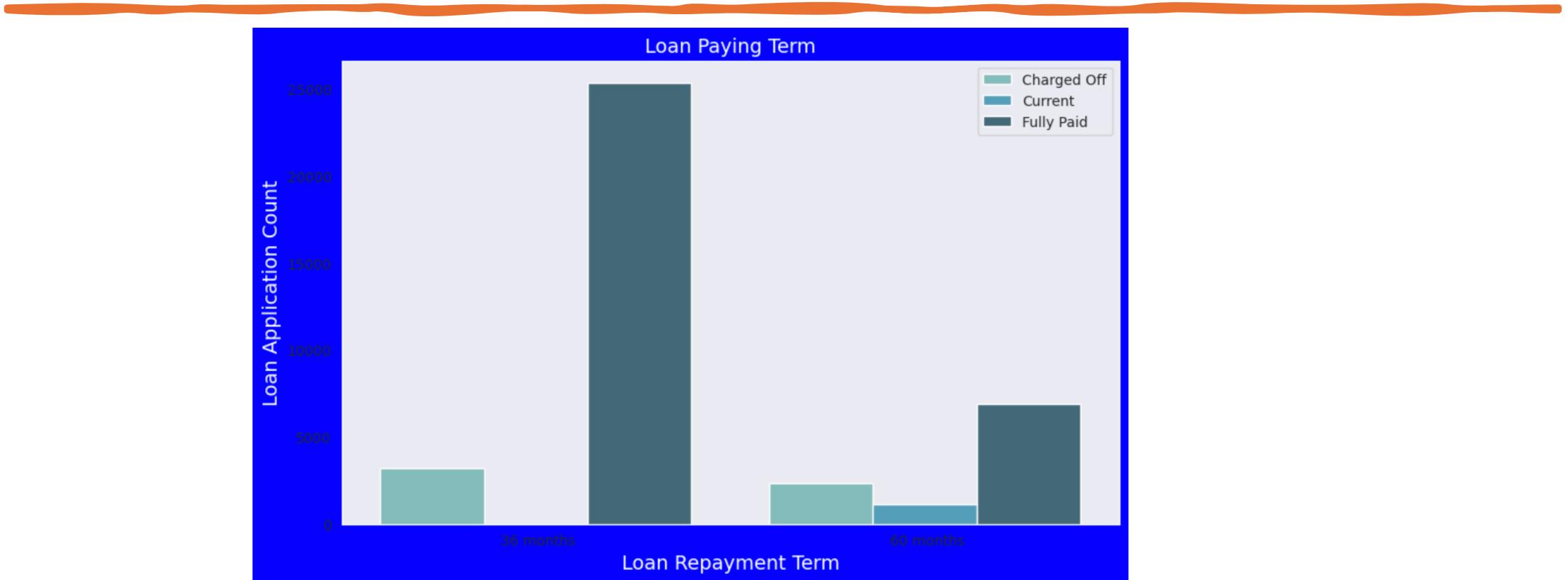
Univariate Analysis: unordered categorical Variable - Home Ownership



Observation:

1. most of them living in rented home or mortgaged their home have fully paid the loan
2. there is a significant number of applicant whose loan has been charged off with Mortgage and rent

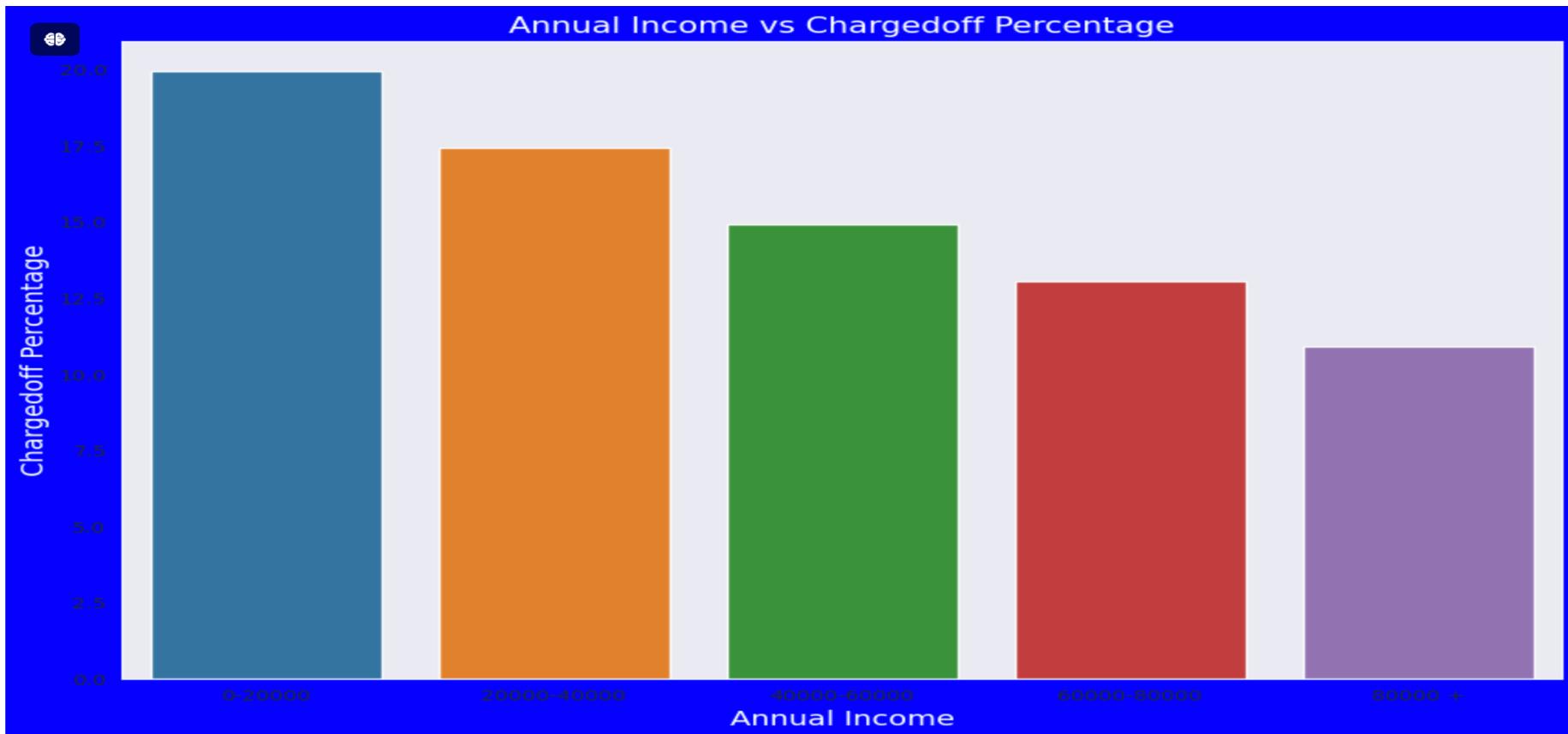
Univariate Analysis: unordered categorical Variable – Loan Paying term



Observation:

1. Those who had taken loan to repay in 36 months had more number of applicants getting charged off as compared to applicants who had taken loan for 60 months.

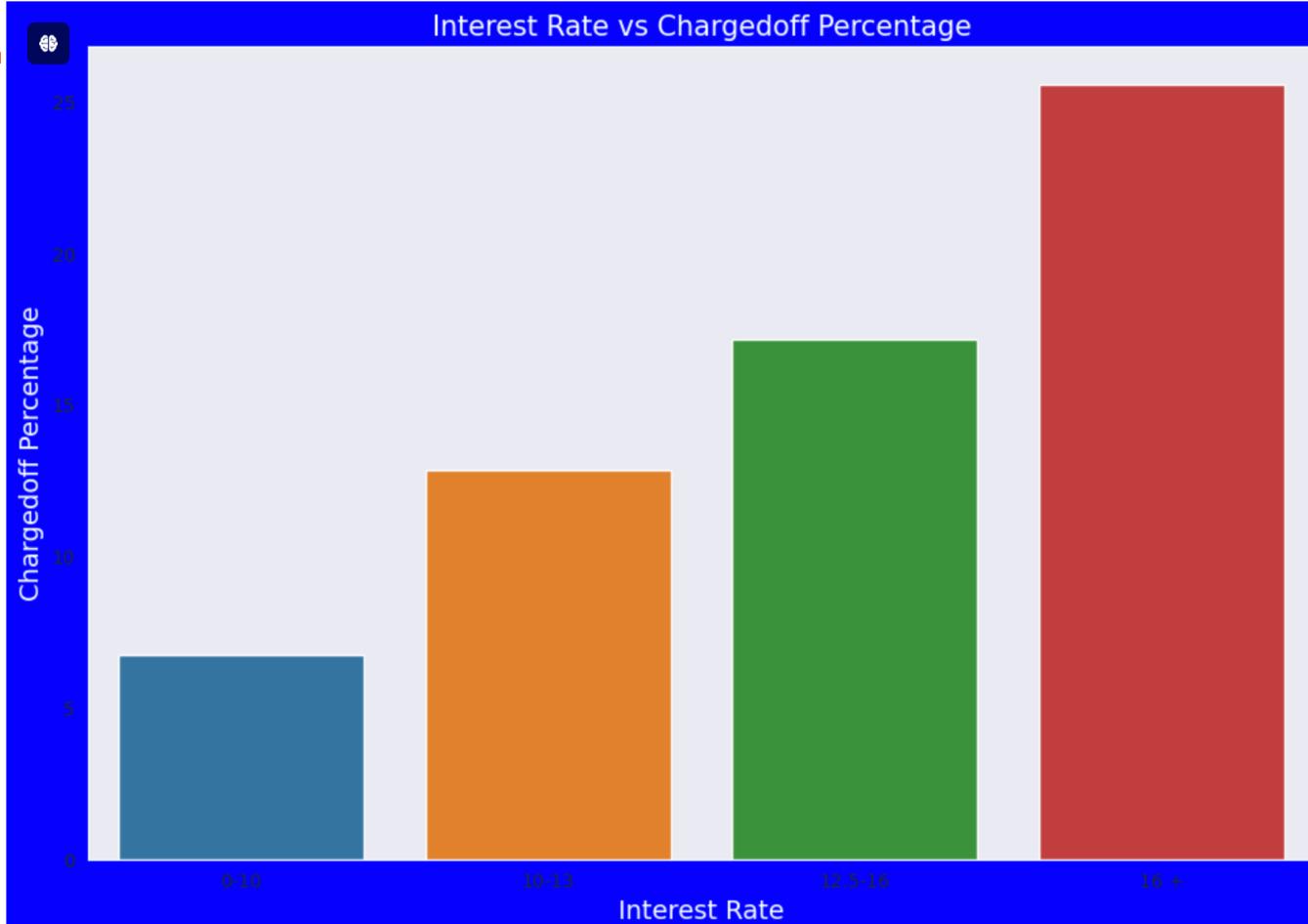
Bivariate Analysis: Annual Income v/s charged off



Observation:

1. Income range 0-20000 has high chances of charged off. People are likely to default paying loans
2. Income range 80000+ has less chances of charged off

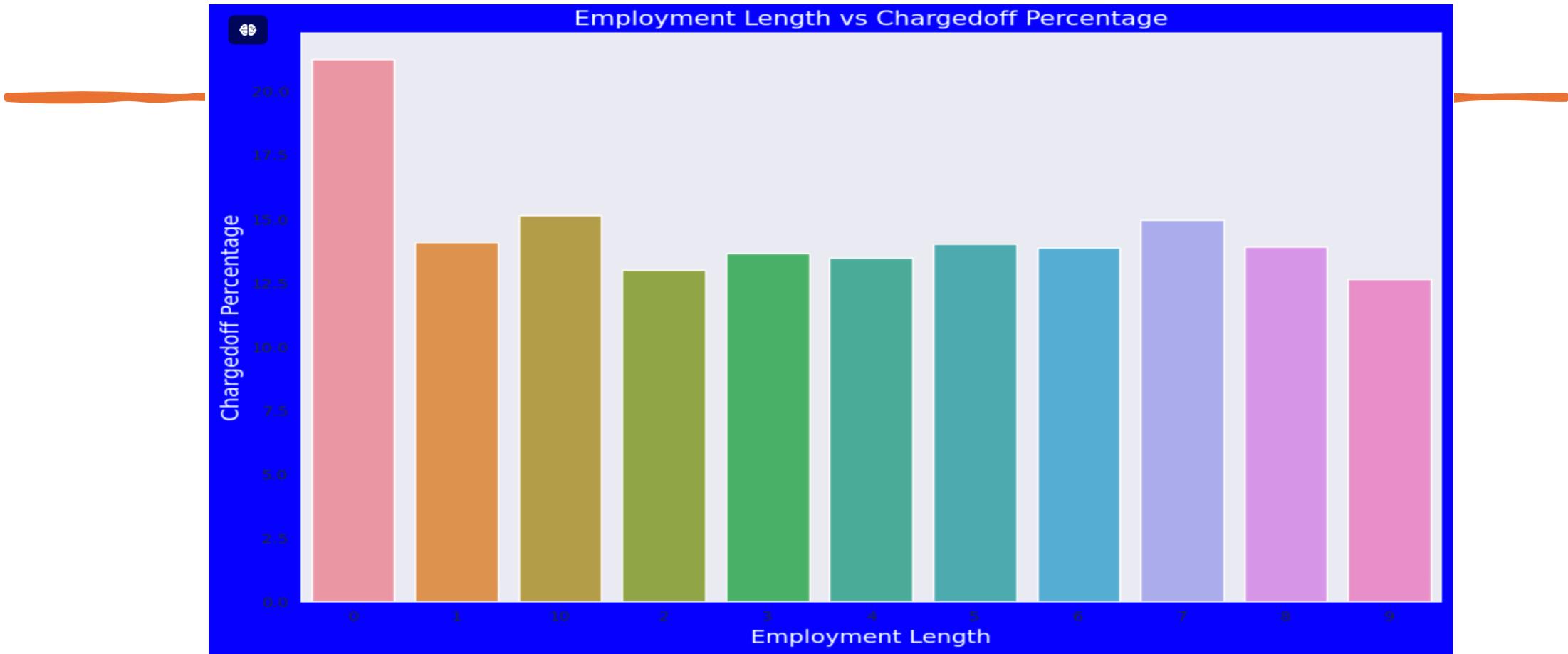
Bivariate Analysis: Interest Rate v/s charged off



Observation:

1. interest rate more than 16% has good chances of charged off as compared to other category interest rates.

Bivariate Analysis: Employment length



Observation:

1. Those who are not working or have less than 1 year of work experience have high chances of getting charged off.

Conclusion:

1. People with living in rented home or mortgaged their home have fully paid the loan and they are less likely to default
2. People with shorter interest replay term are more likely to get charged off
3. People whose annual income is not verified are more likely to charged off
4. Income range 0-20000 has high chances of charged off. The chances reduces for the people with higher income bracket