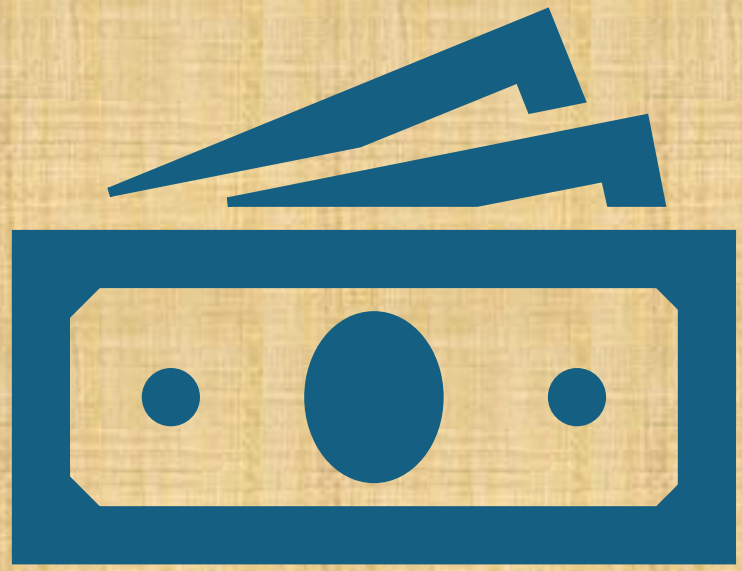
Lending Club Case Study

Analysis of Risky Loan Applicants



- Ajit Mishra
- Pravin Shivaji Ahire

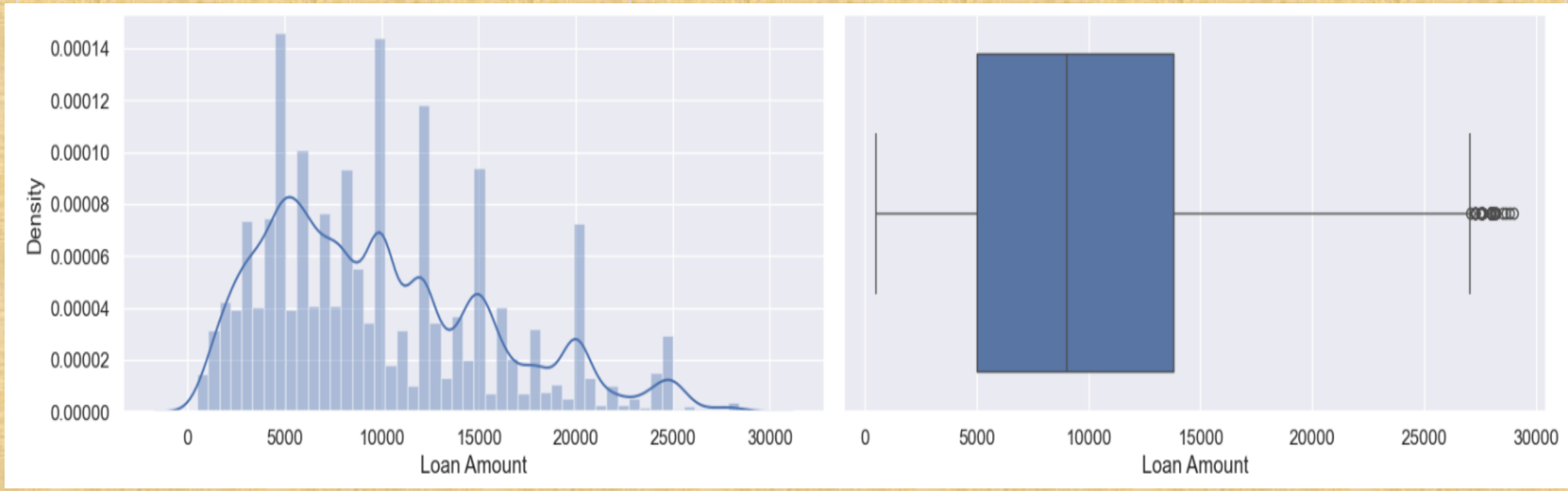
Objective

- This analysis aims to identify patterns and factors that indicate if a person is likely to default on a loan, using data from past loan applicants. The goal is to help the company make informed decisions to reduce credit loss.

Loan_status

loan_status	
Fully Paid	32950
Charged Off	5627

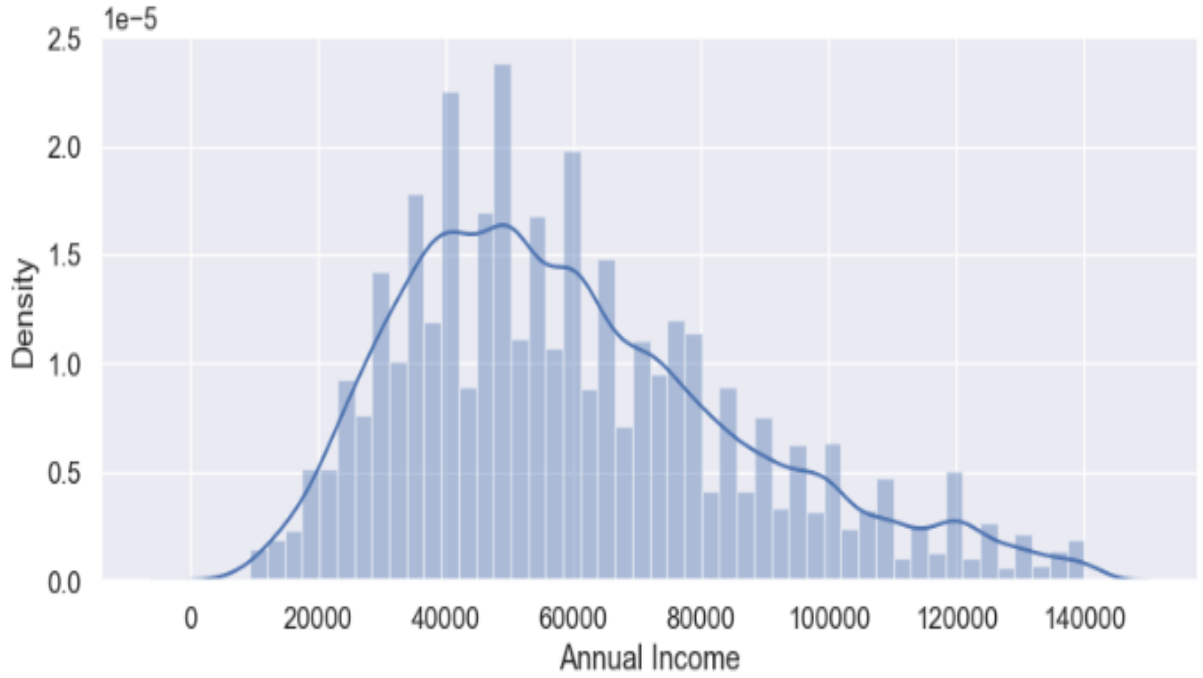
Loan Amount:

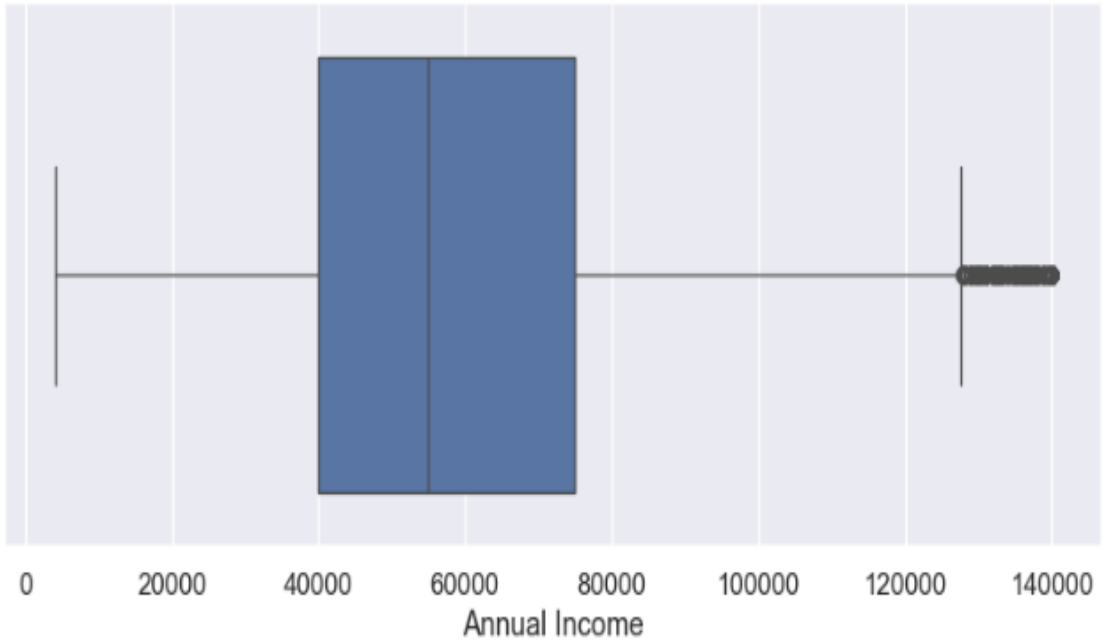


Graph Observations:

- The distribution of loan amounts is right-skewed, with a majority of loans falling between $5,000 and 15,000.
- The mean loan amount is around $11,000.
- There are a few outliers on the higher end, indicating some loans significantly larger than the average.

Annual Incomes

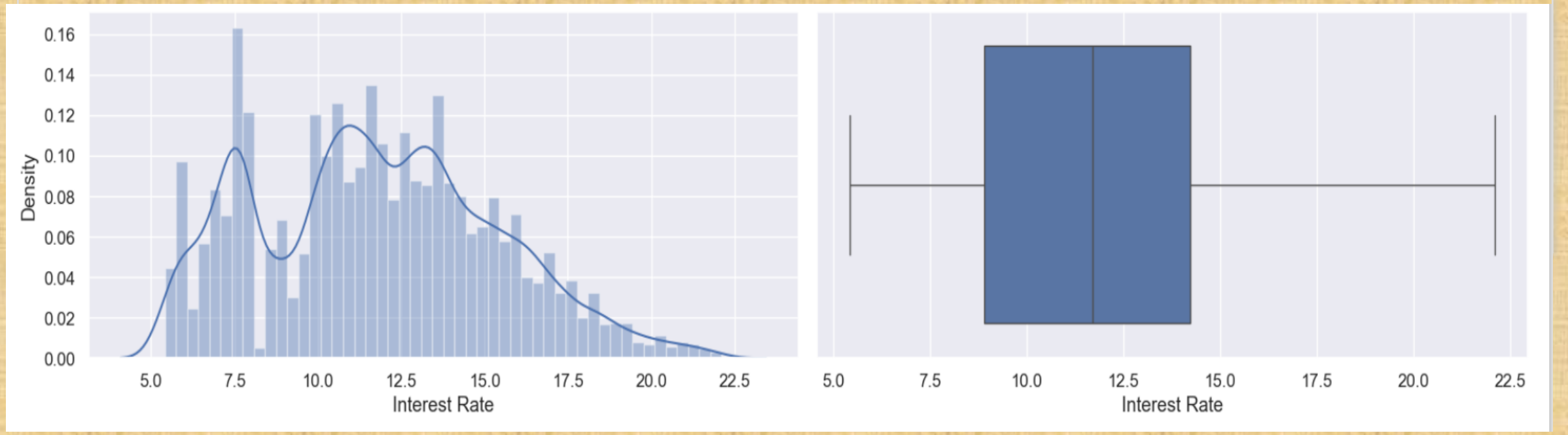




Graph Observations:

- The distribution of annual incomes is heavily right-skewed, indicating a concentration of borrowers with lower incomes and a long tail of high earners.
- The mean annual income is around $65,000, but the median is likely lower due to the skew.
- There are potential outliers on the higher end, suggesting a small number of borrowers with significantly higher incomes compared to the majority.

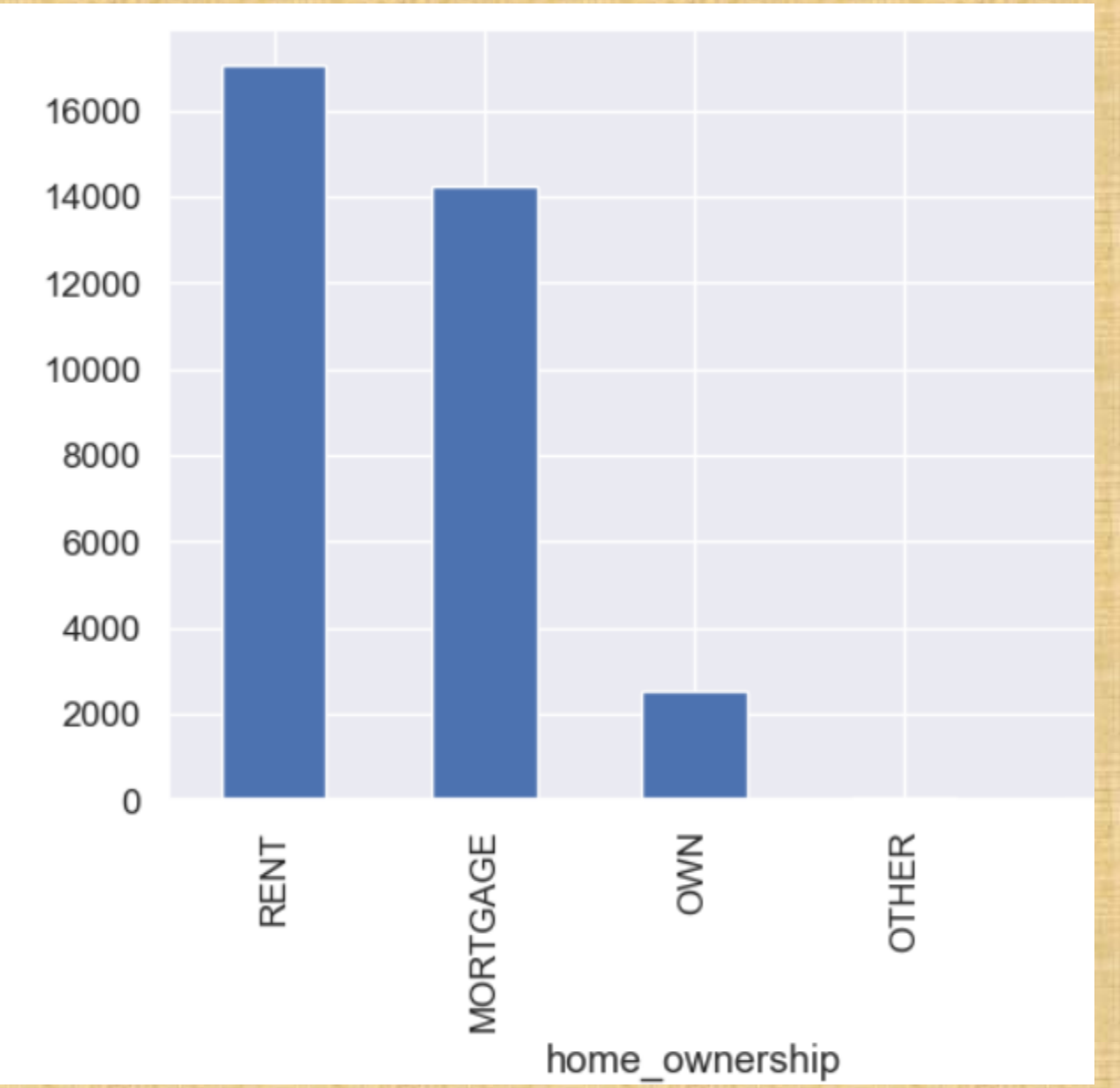
Interest Rates



Graph Observations:

- The distribution of interest rates appears to be bimodal, with peaks around 10% and 13%.
- The mean interest rate is approximately 12%.
- There are no significant outliers in the interest rate distribution.
- The interest rate for most of the loans lies between 9%-14%.

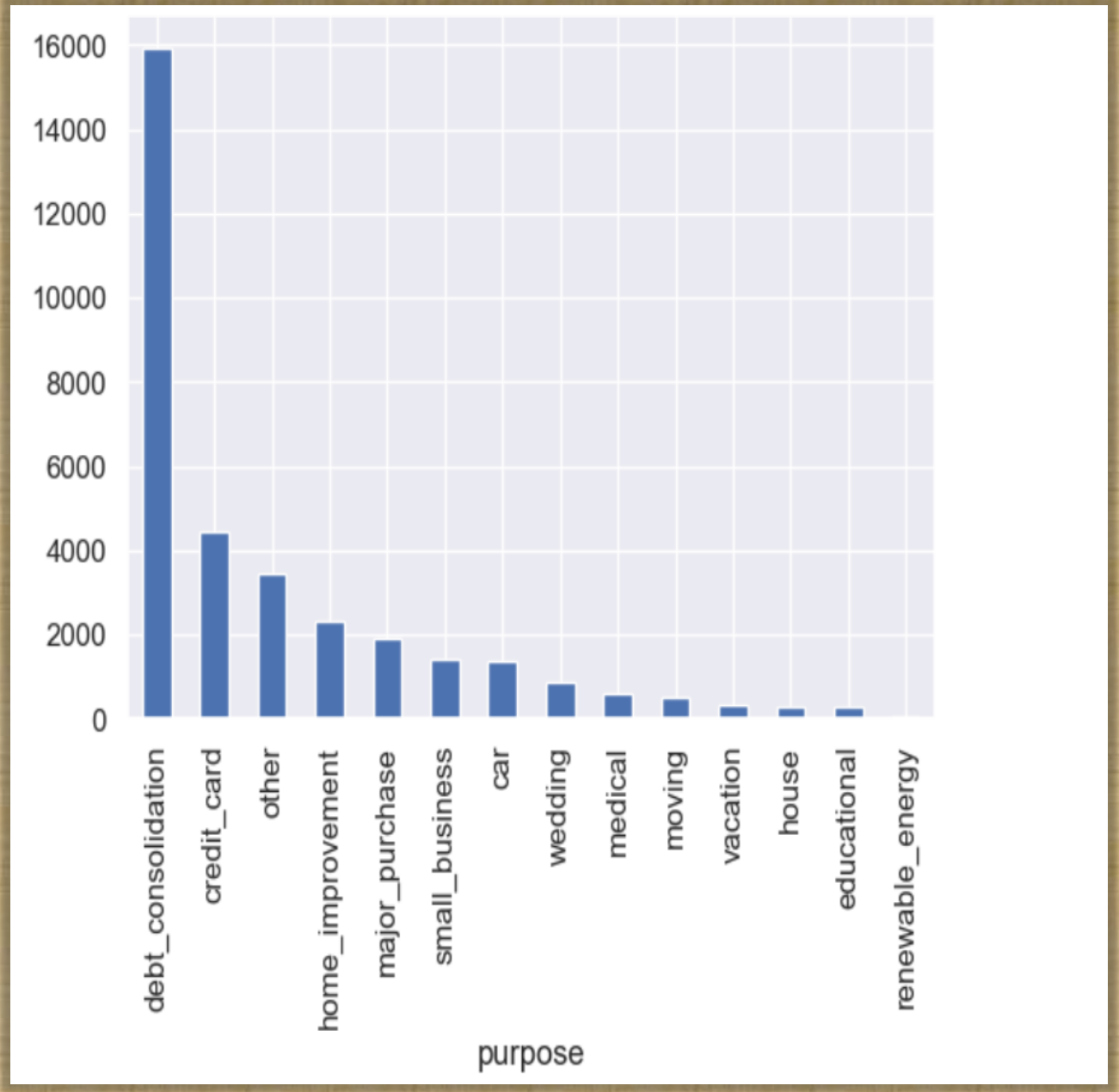
Home Ownership



Graph Observation:

Most of the loan applicants either living in rented house or mortgaged their house.

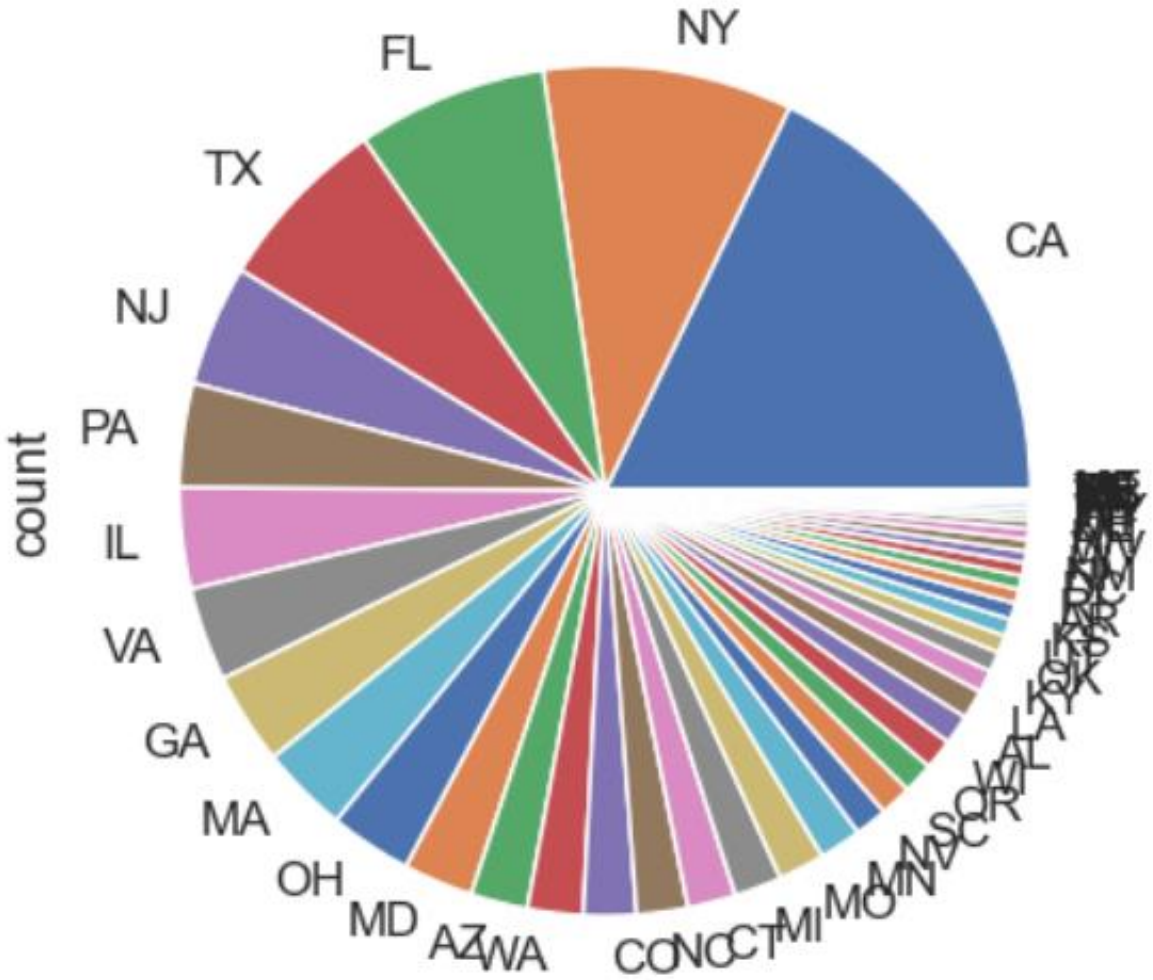
Purpose



Graph Observation:

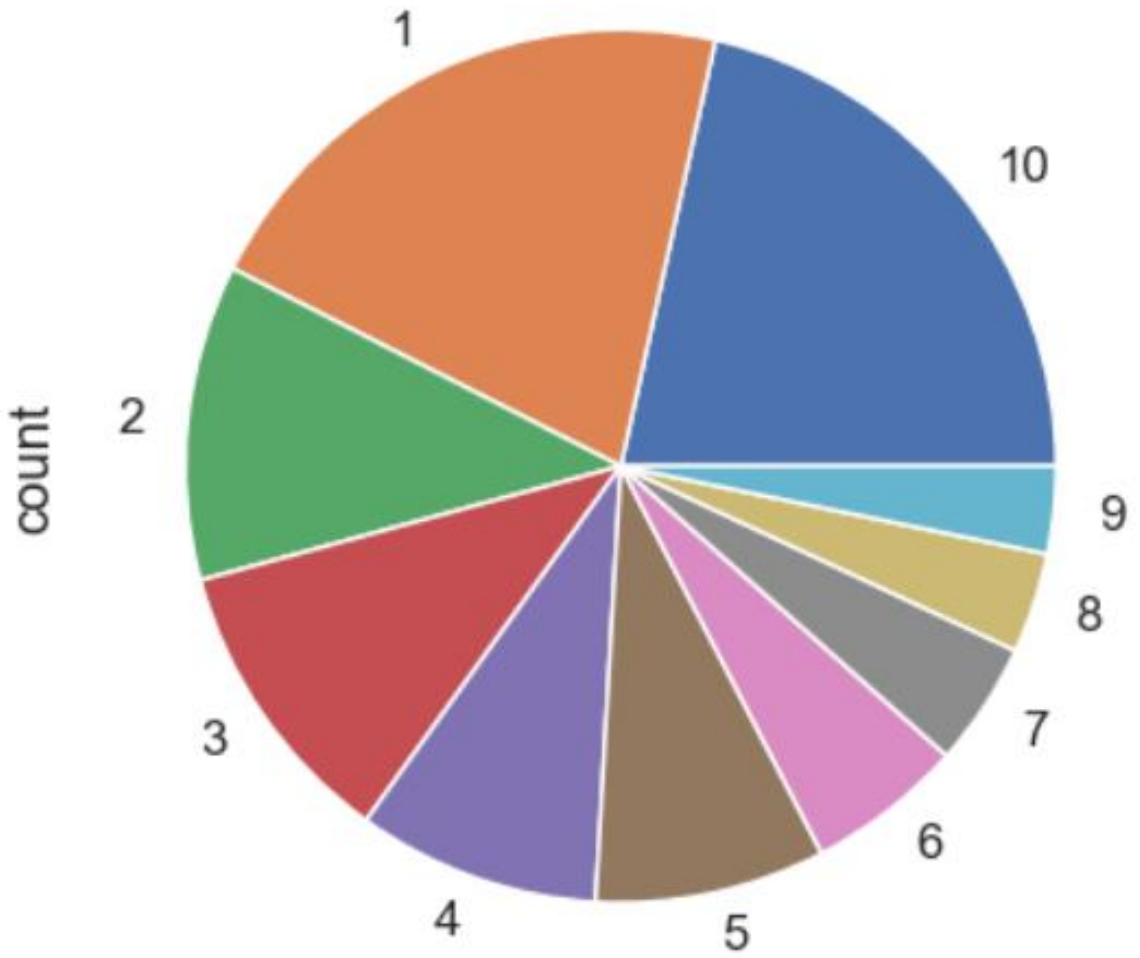
Most of the loan applicants are for debt_consolidations.

Purpose



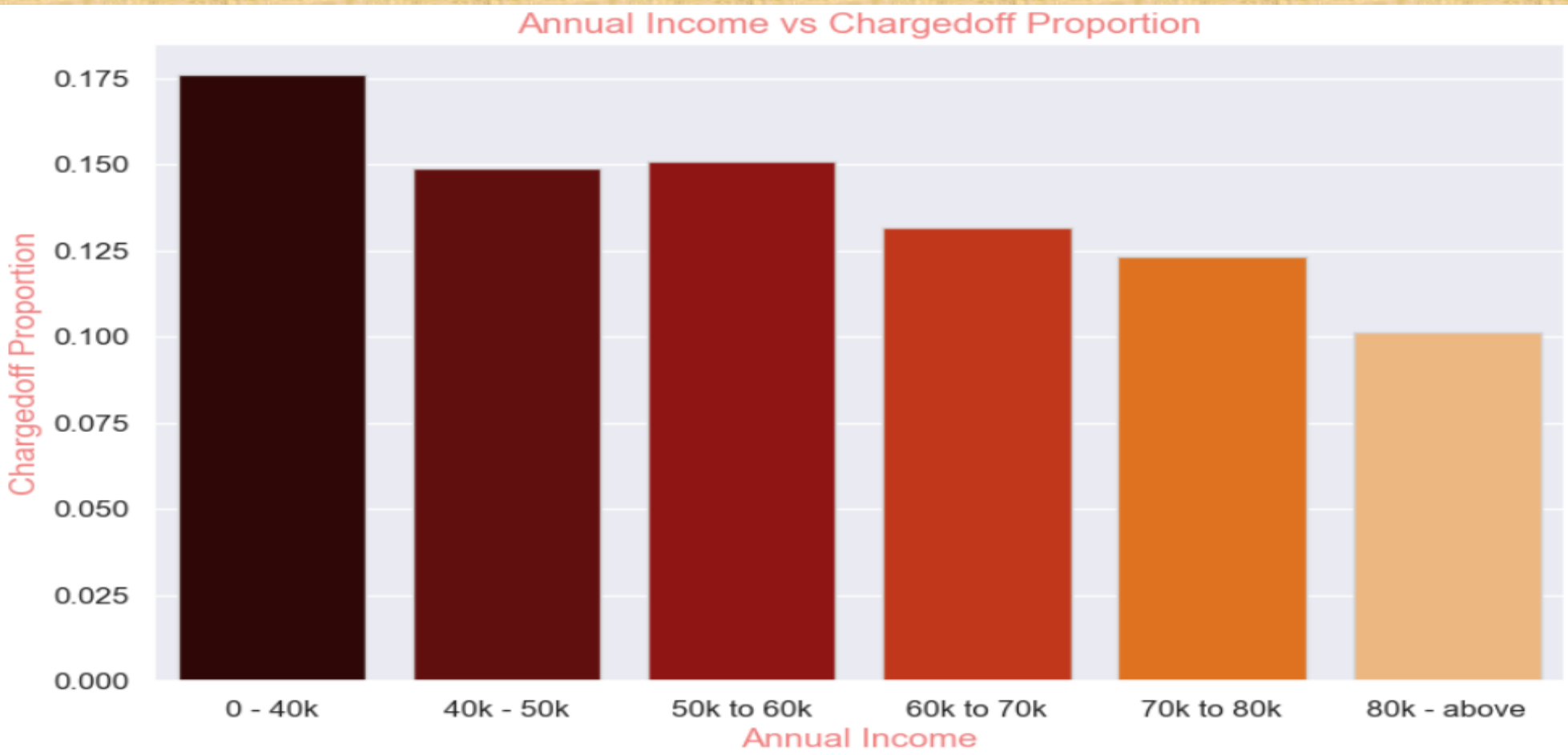
Graph Observation: Most of the loan applicants are from CA state.

Employee Length



Graph Observation: Most of the applications are having 10+ yrs of Exp.

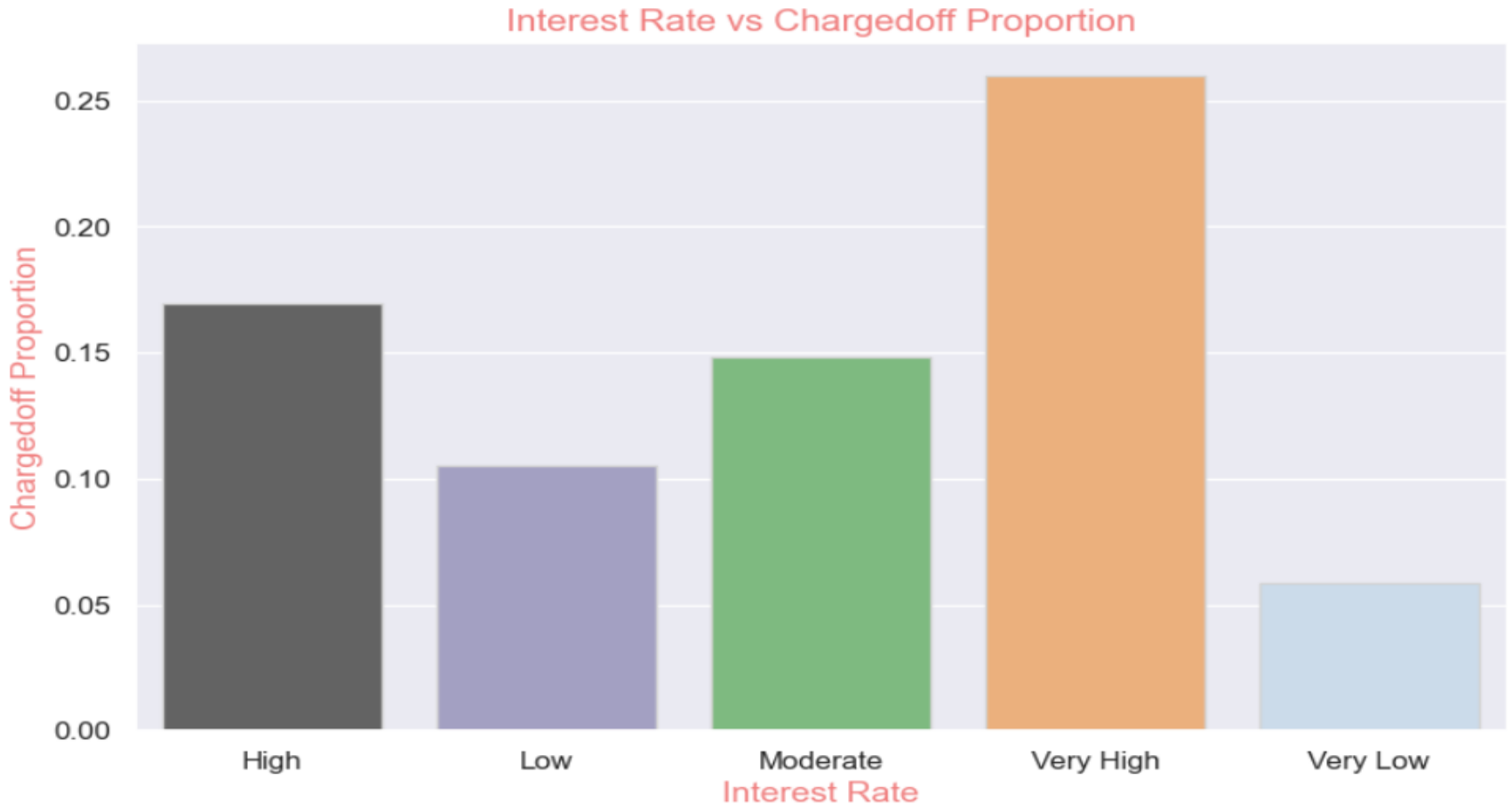
Bivariate Analysis



Graph Observation:

- Applicants with lower annual income are more likely to default on loans compared to those with higher annual incomes.
- This suggests that income level is a significant factor in assessing credit risk.
- Income range 80000+ has less chances of charged off.
- Income range 0-40000 has high chances of charged off.
- Increase in annual income charged off proportion got decreased.

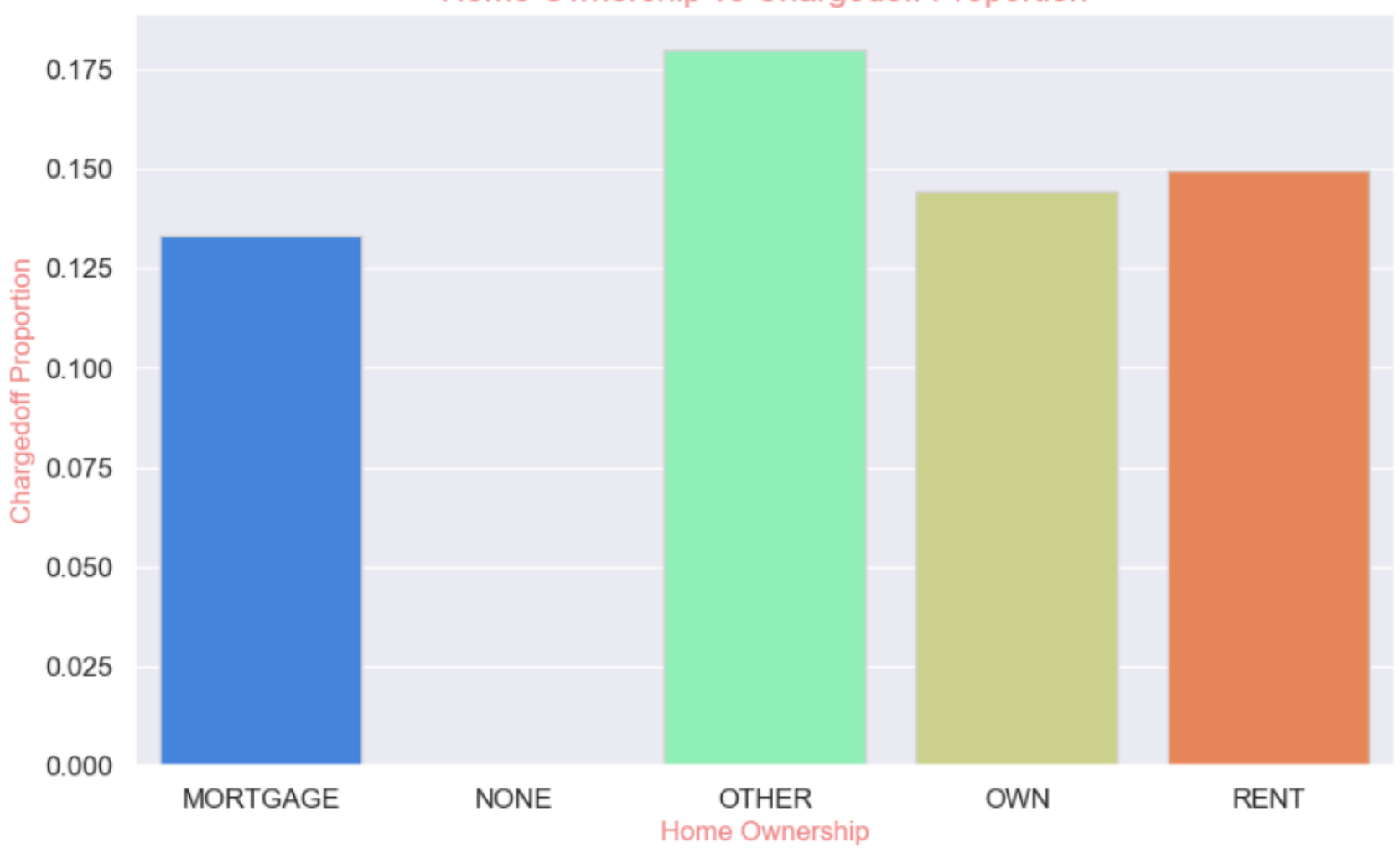
Interst Rate Vs Chargedoff Proportion



Graph Observation:

- Loans with higher interest rates have a significantly higher proportion of charge-offs.
- This indicates that interest rate is a strong predictor of default risk, with higher rates associated with increased likelihood of loan default.
- interest rate less than 10% or very low has very less chances of charged off. Intrest rates are starting from minimum 5 %.
- interest rate more than 16% or very high has good chnaces of charged off as compared to other category intrest rates.
- Charged off proportion is increasing with higher intrest rates.

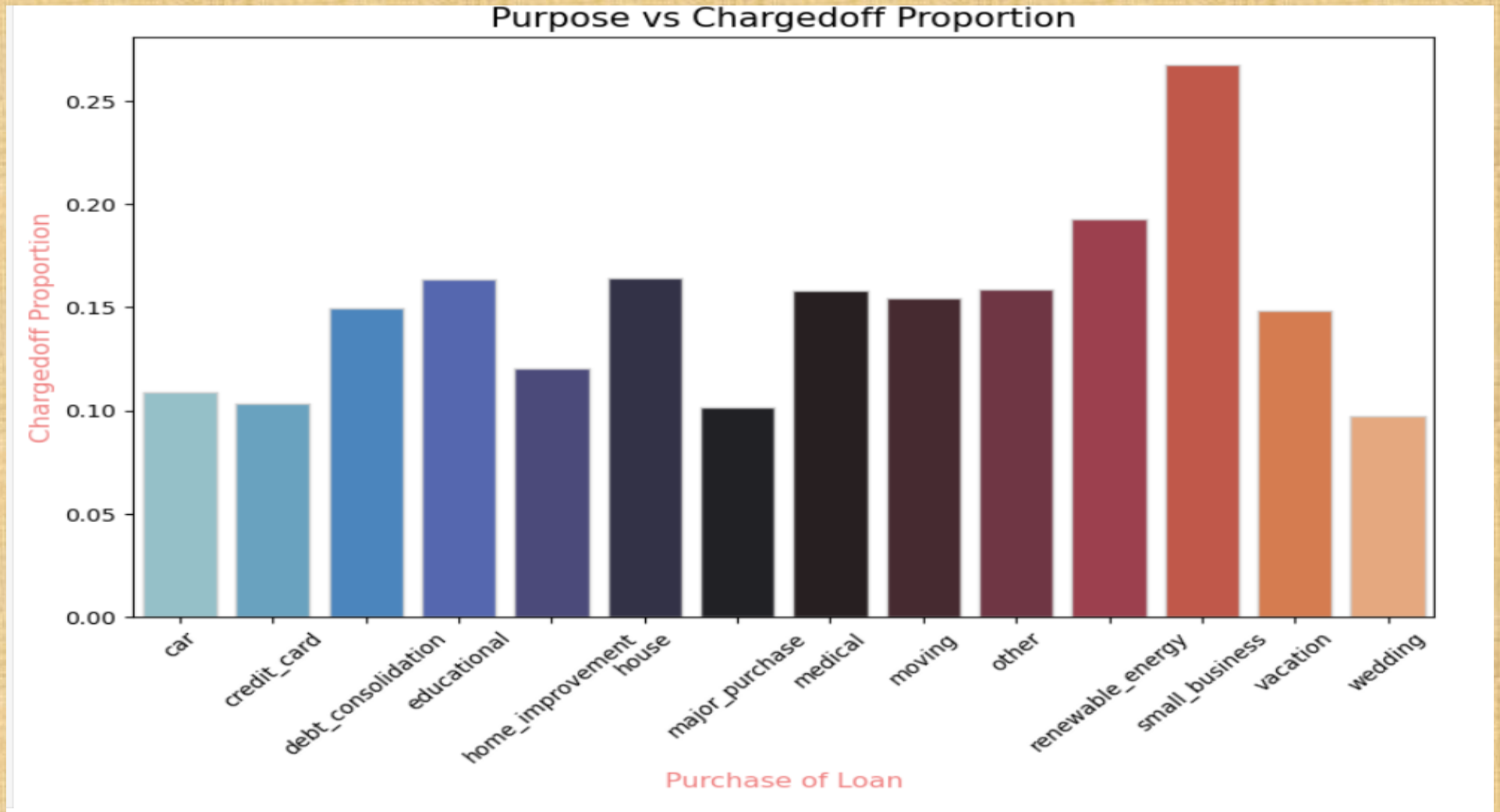
Ownership Vs Chargedoff Proportion



Graph Observation:

- Those who are not owning the home is having high chances of loan defaults.

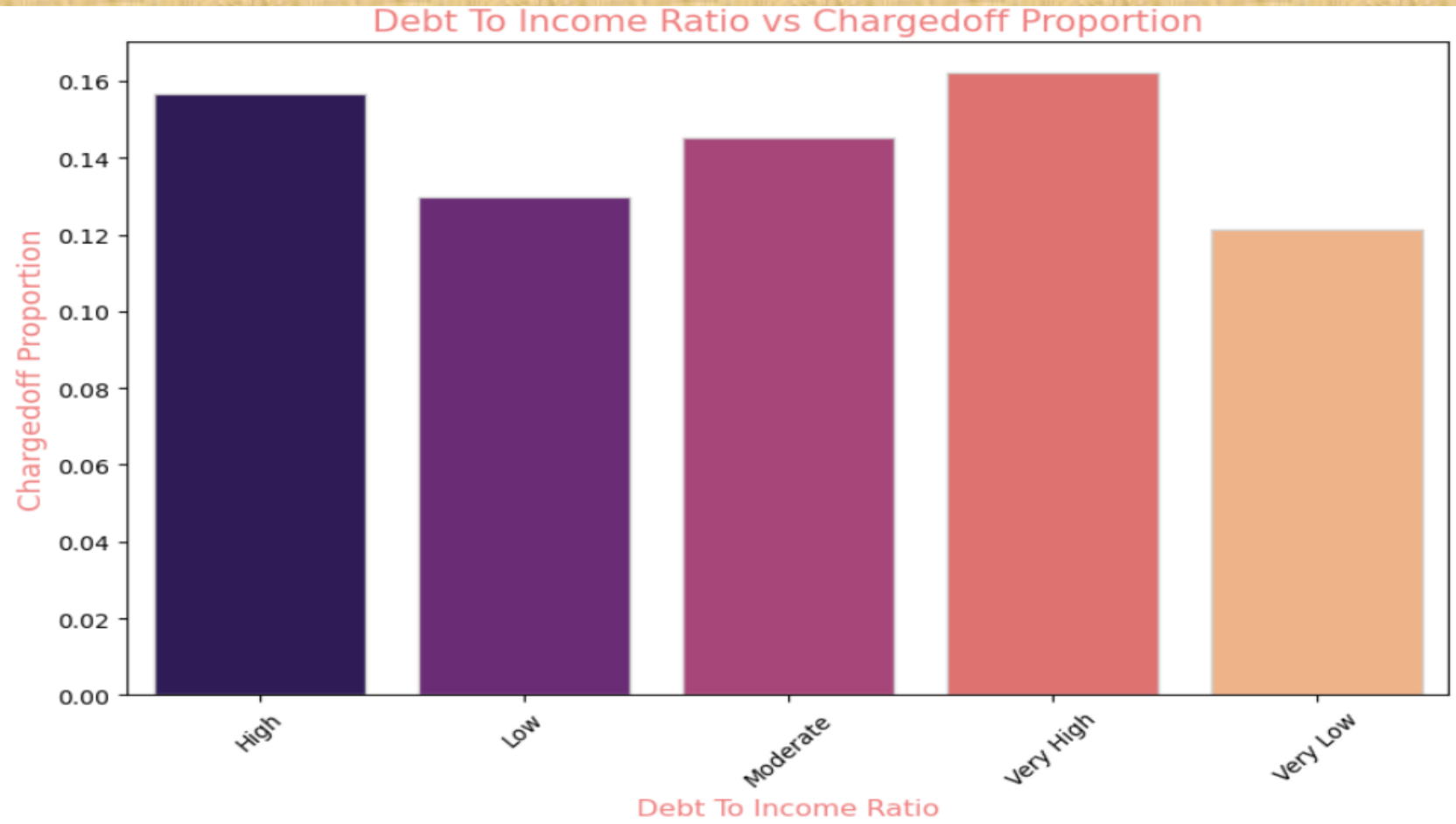
Purpose Vs Chargedoff Proportion



Graph Observation:

- Those applicants who is having home loan is having low chances of loan defaults.
- Those applicants having loan for small bussiness is having high chances for loan defaults.

DTI Vs Chargedoff Proportion

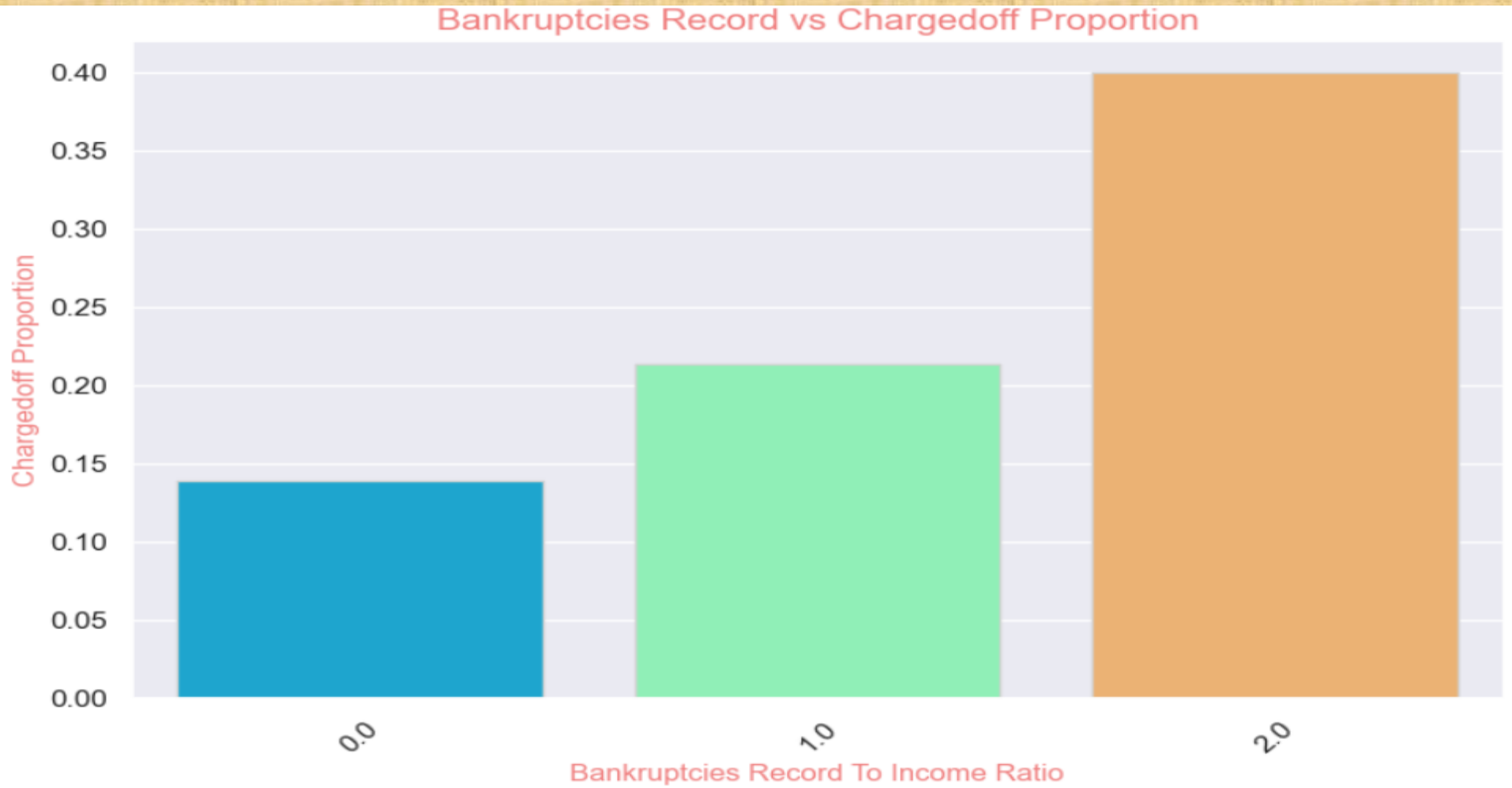


Graph Observation:

- High DTI value having high risk of defaults.
- Lower the DTO having low chances loan defaults

This suggests that borrowers with higher DTI ratios are more likely to default, indicating a higher risk associated with lending to individuals with a larger debt burden relative to their income.

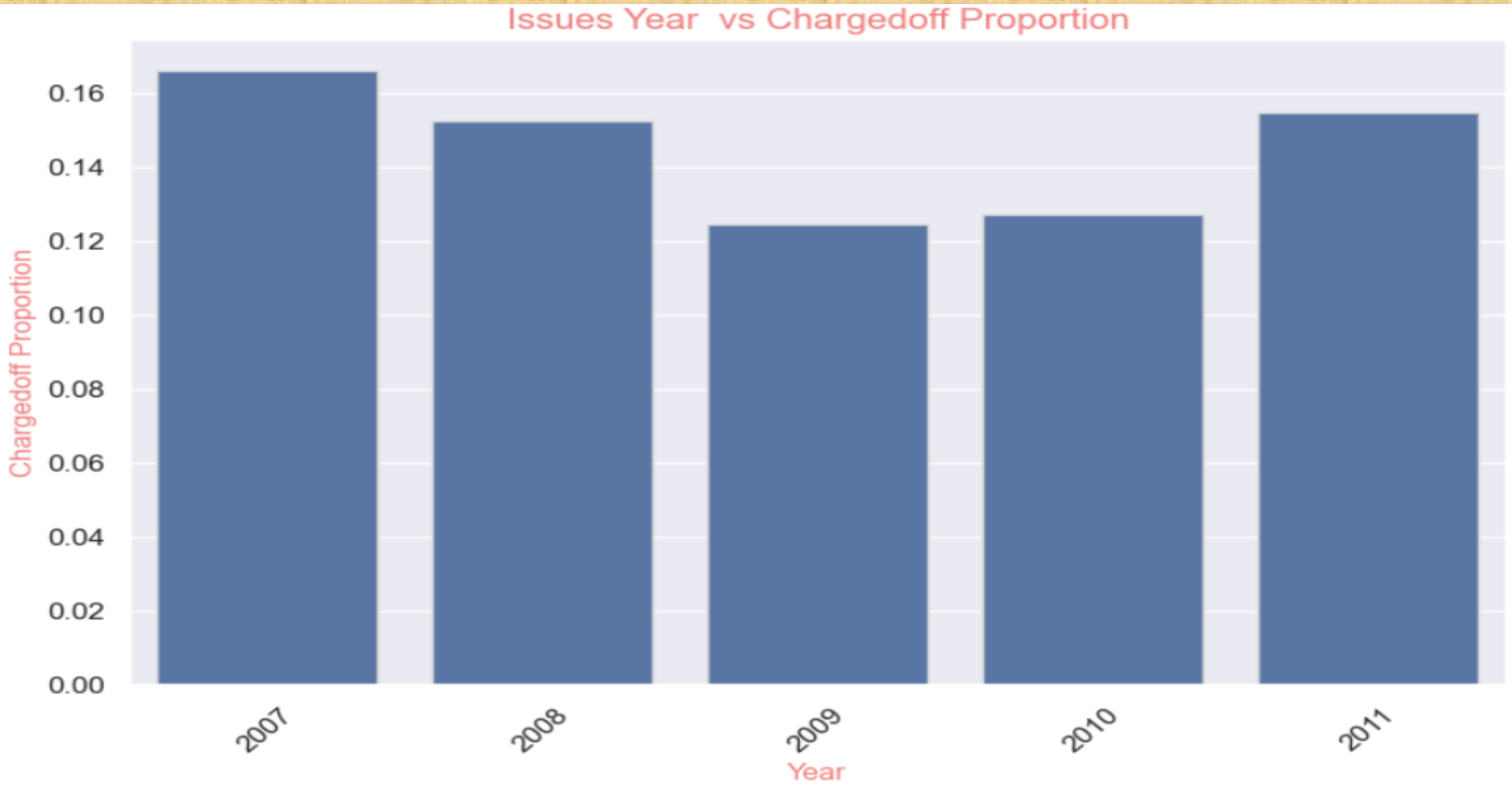
Bankruptcies Vs Chargedoff Proportion



Graph Observation:

- Bankruptcies Record with 2 is having high impact on loan defaults.
- Bankruptcies Record with 0 is low impact on loan defaults.
- Lower the Bankruptcies lower the risk.

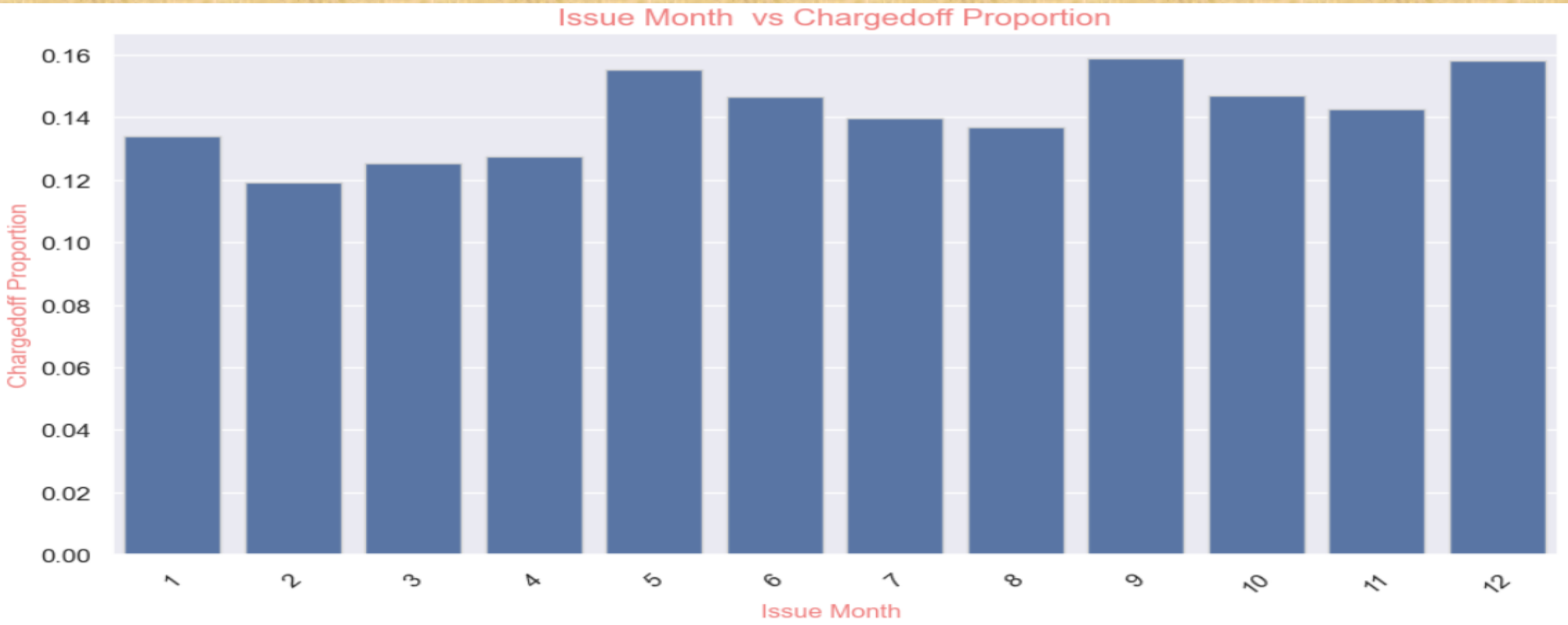
Issues Year Vs Chargedoff Proportion



Graph Observation:

- Loans issued in earlier years (2007 and 2008) show a higher proportion of charge-offs compared to those issued in later years.
- Year 2009 is having lowest loan defaults.
- This could be attributed to various factors, including the economic conditions prevalent during those years or changes in lending practices over time.

Issues Month Vs Chargedoff Proportion

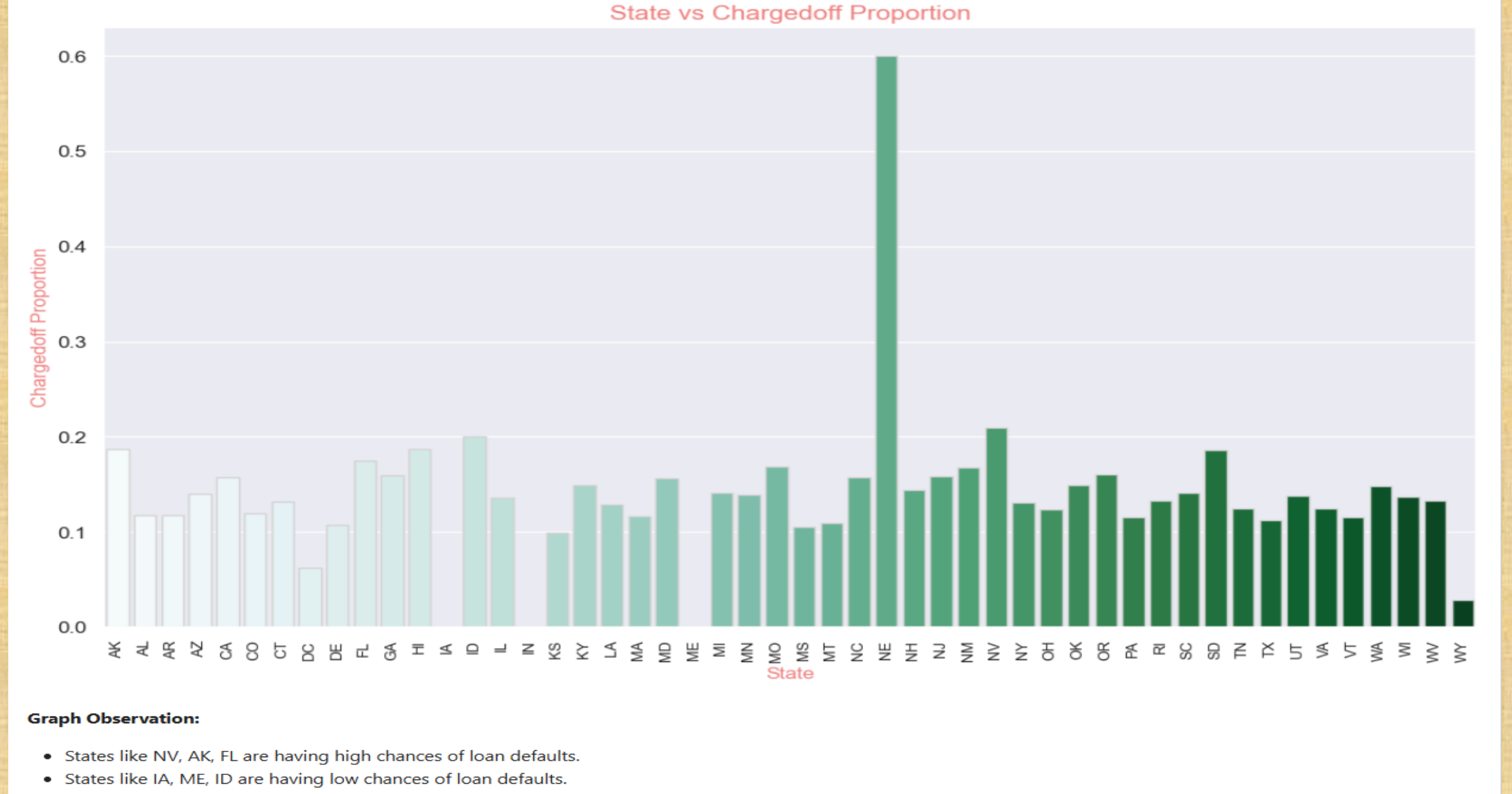


Graph Observation:

- Those loan has been issued in May, September and december is having high number of loan defaults.
- Those loan has been issued in month of February is having low number of loan defaults.
- Majority of loan defaults comig from applicants whose loan has been approved from September-to Deceber.

The month in which a loan is issued might not be a strong predictor of default risk.

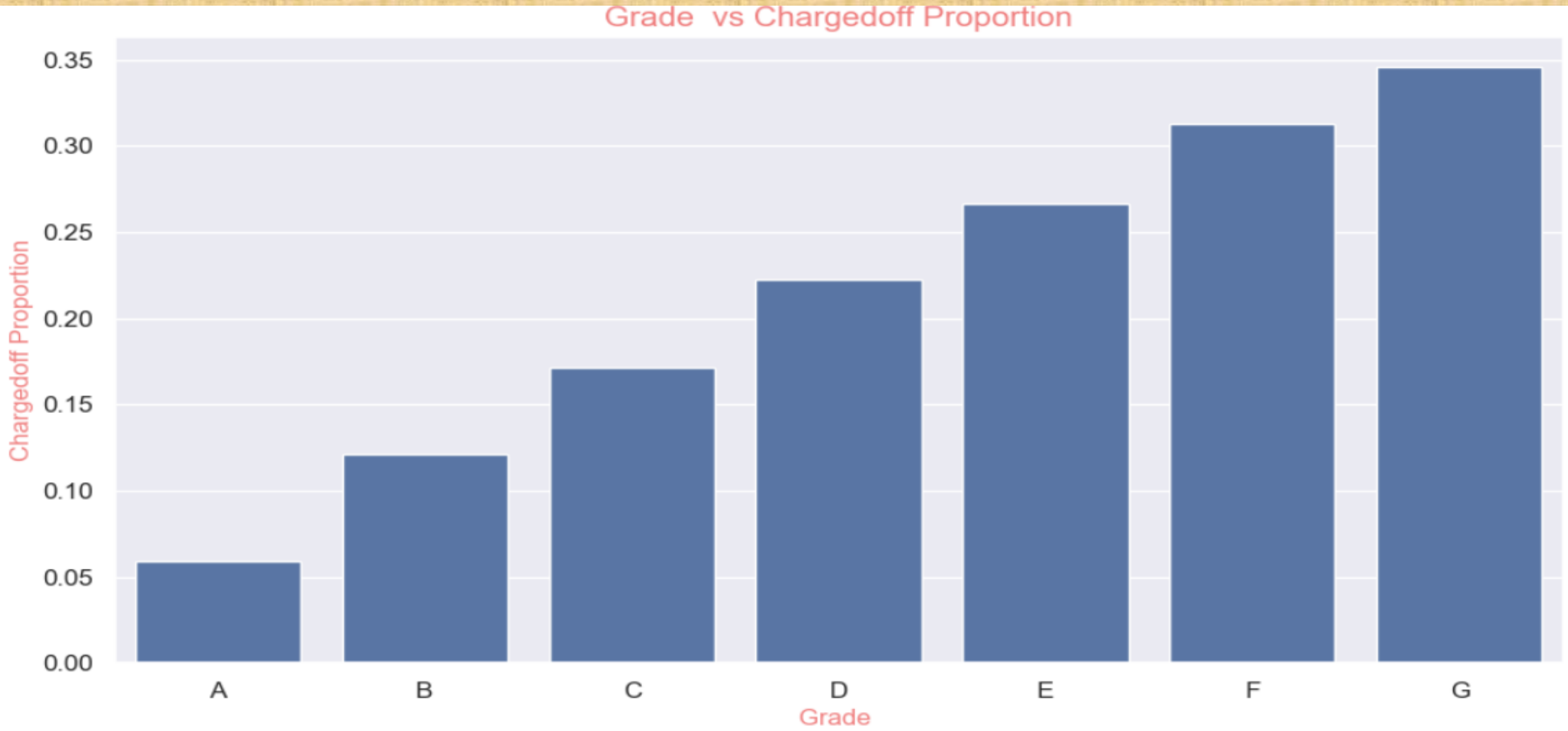
State Vs Chargedoff Proportion



Graph Observation:

- States like NV, AK, FL are having high chances of loan defaults.
- States like IA, ME, ID are having low chances of loan defaults.

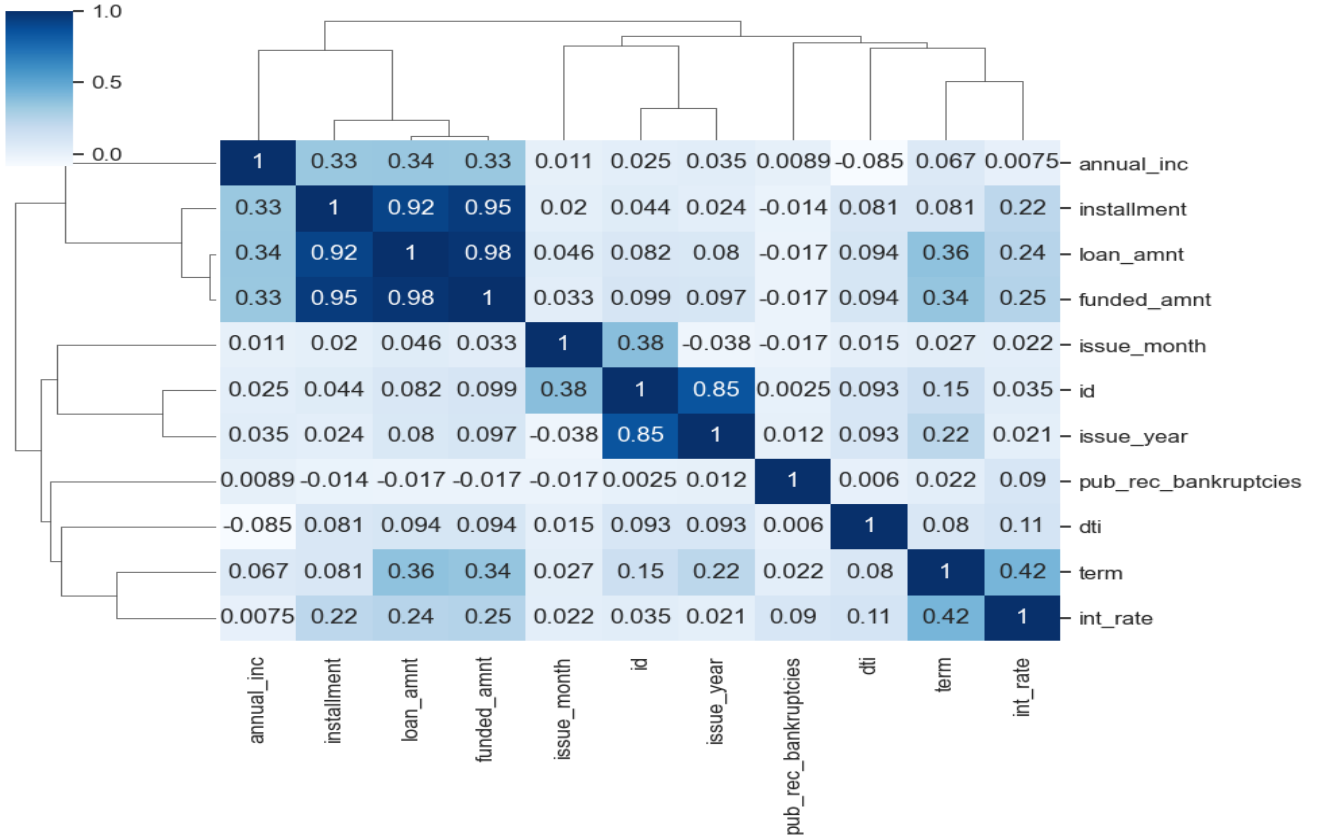
Grade Vs Chargedoff Proportion



Graph Observation:

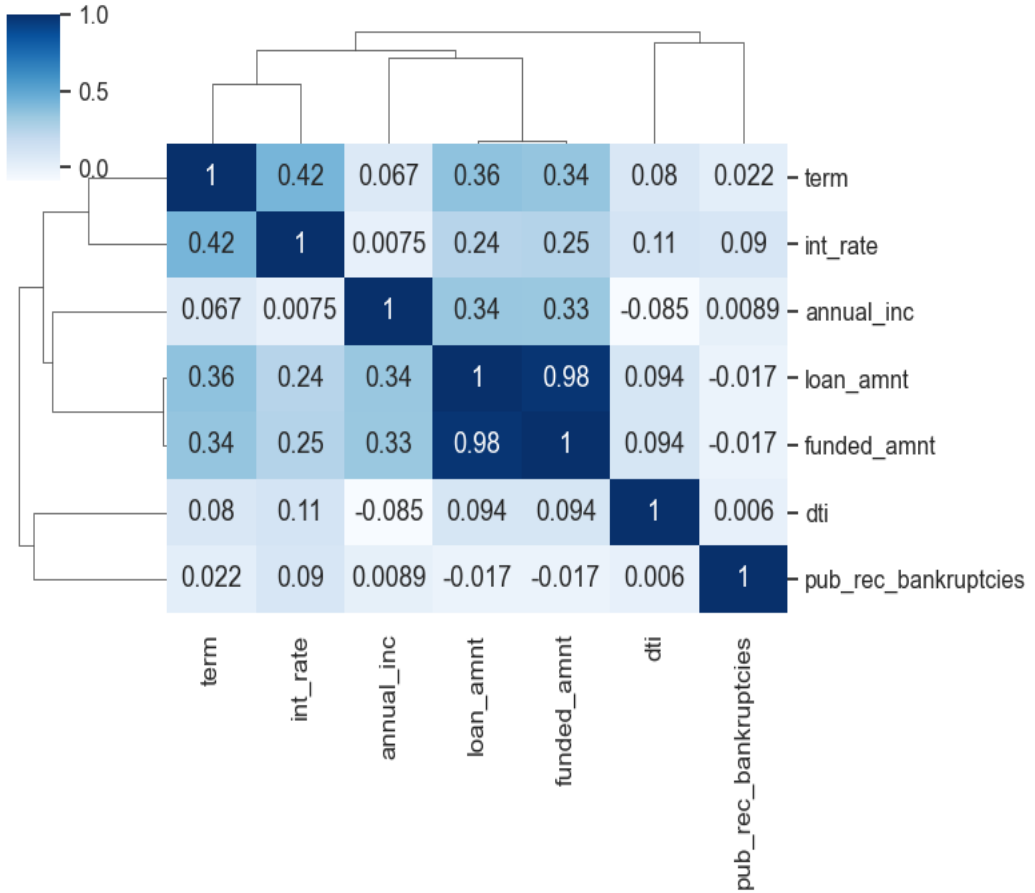
- The Loan applicants with loan Grade G is having higest Loan Defaults and loan applicants with loan A is having lowest Loan Defaults.

Correlation heatmap



Observations on correlation graph

1. Positive Correlations:
   - Loan amount, funded amount and investor-funded amount are strongly positively correlated, indicating that larger loan requests generally receive higher funding.
   - Interest rate shows a moderate positive correlation with loan amount, suggesting that larger loans tend to have higher interest rates.
   - Installment amount is strongly correlated with loan amount and funded amount, reflecting the direct relationship between loan size and monthly payments.
2. Negative Correlations:
   - Public record bankruptcies exhibit a weak negative correlation with annual income, implying that individuals with higher incomes are less likely to have a history of bankruptcies.
3. Overall:
   - The heatmap reveals the interdependencies among various numerical variables in the loan dataset.
   - These correlations can be leveraged to gain insights into factors influencing loan approval, interest rates, and the likelihood of default.



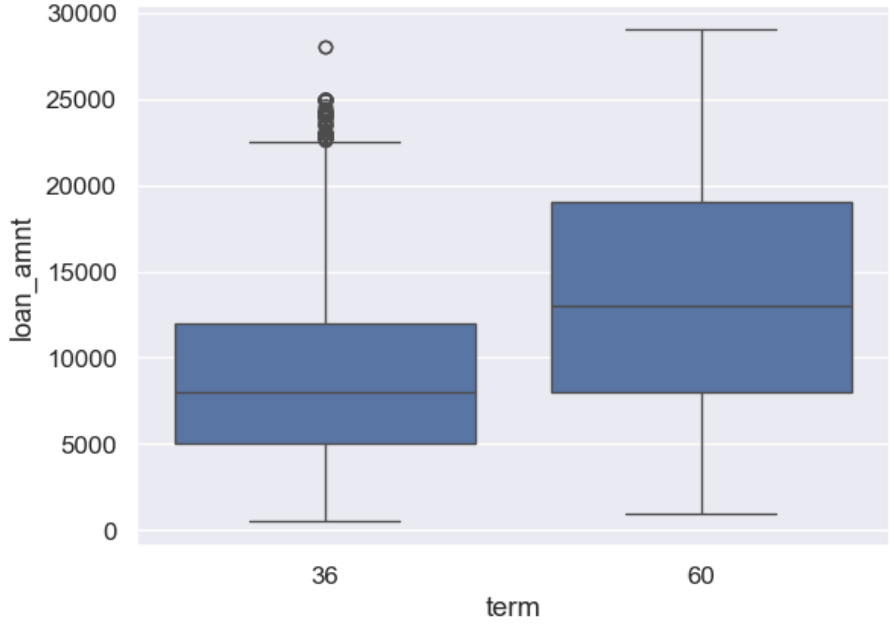
- Negative Correlation:

```
1.loan_amnt has negative correlation with pub_rec_bankrupticies
2.annual income has a negative correlation with dti
```

- Strong Correlation:

```
1.term has a strong correlation with loan amount
2.term has a strong correlation with interest rate
3.annual income has a strong correlation with loan_amount
```

Term Vs Loan Amount



Observation on graph

This graph indicates a bivariate analysis of the relationship between loan term ('term') and loan amount ('loan_amnt'). Specifically, it's a box plot visualization that allows us to compare the distribution of loan amounts for different loan terms (e.g., 36 months vs. 60 months).

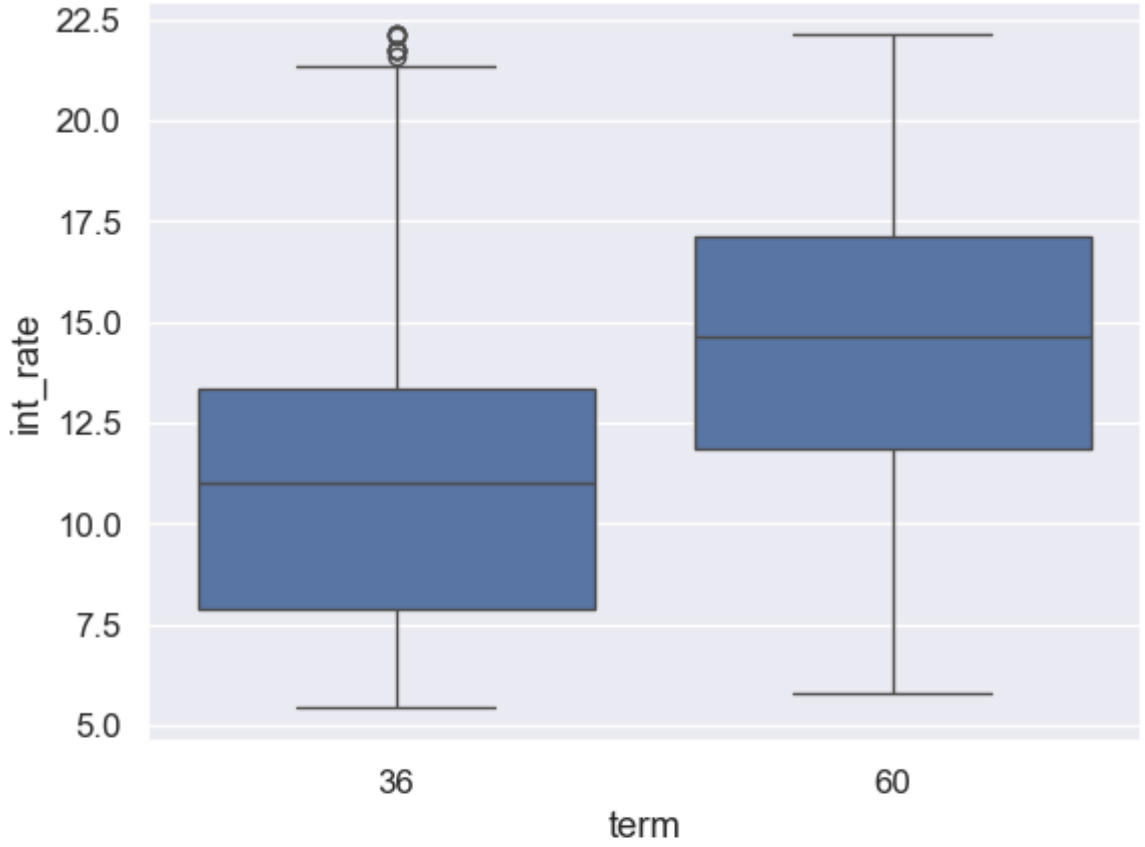
Key insights that can be derived from this box plot include:

- **Median Loan Amount**: The horizontal line within each box represents the median loan amount for that term. We can observe whether loans with longer terms tend to have higher or lower median amounts.
- **Interquartile Range (IQR)**: The box itself spans the interquartile range (IQR), which contains the middle 50% of the data. This gives us an idea of the spread or variability of loan amounts within each term category.
- **Outliers**: The dots or circles beyond the "whiskers" of the box plot represent outliers, which are loan amounts that are significantly higher or lower than the majority of loans for that term.
- **Distribution Shape**: The shape of the box and whiskers can provide clues about the distribution of loan amounts. For instance, a longer upper whisker might suggest a right-skewed distribution, indicating a greater number of larger loan amounts.

By analyzing this box plot, we can gain a deeper understanding of how loan term and loan amount are related and whether there are any notable differences in loan amount distributions for different terms.

From the above graph we can see that if term is increasing then loan amount is also increasing.

Term Vs Interest Rate



36 – Short Term
60 – Long Term

Observation:

- The last two boxplots indicate a comparison of loan amount and interest rate across different loan terms (likely short-term and long-term).
- This analysis helps understand how the loan amount and interest rate vary based on the chosen loan term.

Summary

- Loan applicants with lower income and higher interest rates are more likely to default.
- Borrowers who don't own homes and have a history of bankruptcies also pose higher risk.
- Loan purpose influences default rates, with small business loans being riskier.
- Loans issued in earlier years (2007-2008) show higher charge-off rates.
- States like NV, AK, FL exhibit higher default rates compared to IA, ME, ID.

Thank you..... !!!