

Lending Club Case Study

Exploratory Data Analysis

Sathish Kumar Ramasamy
Sudheer Balabadrani

Table of Contents

- Problem Statement
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Conclusion

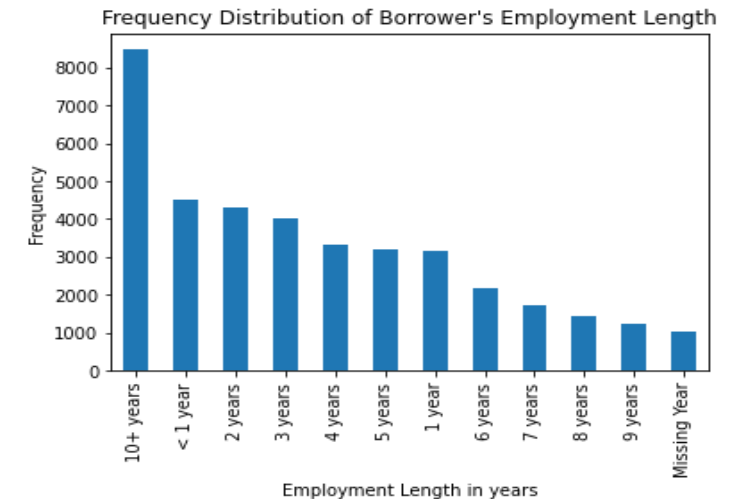
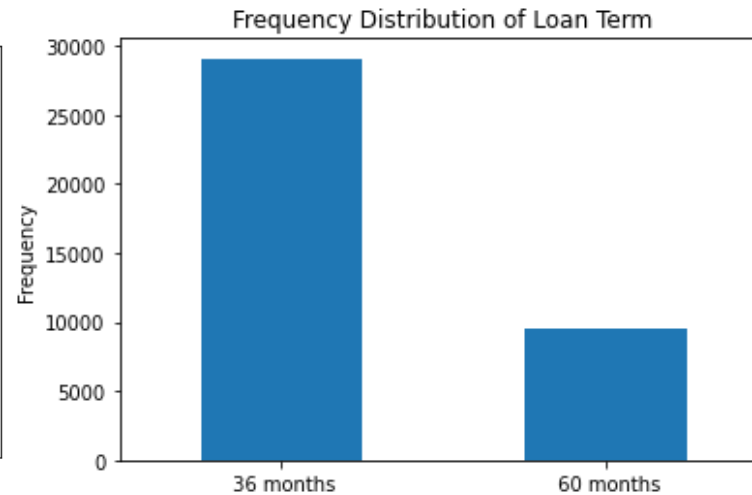
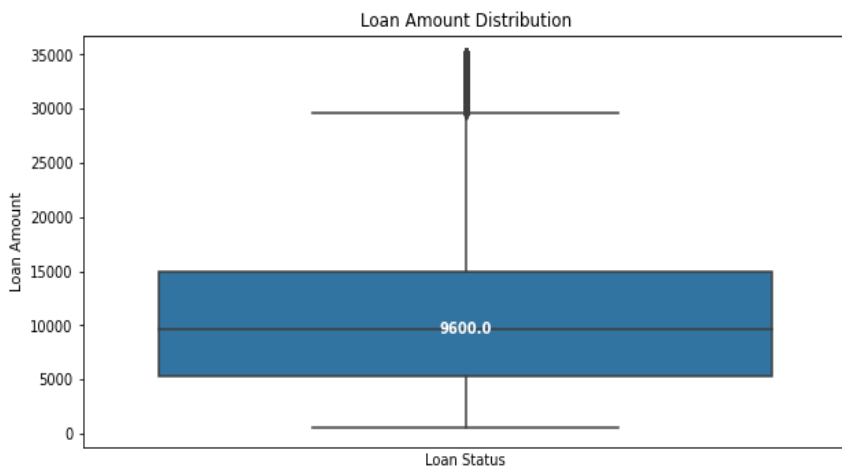
Problem Statement

- The company is the largest online loan marketplace lending borrowers loans for personal, business and medical needs.
- Lending loans to risky applicants is what contributes to the largest financial loss(credit loss) for the company
- The company wants to understand the driving factors(variables) behind loan default so that credit loss can be reduced

Univariate Analysis – Loan Amount, Term, Employment Length

Observations

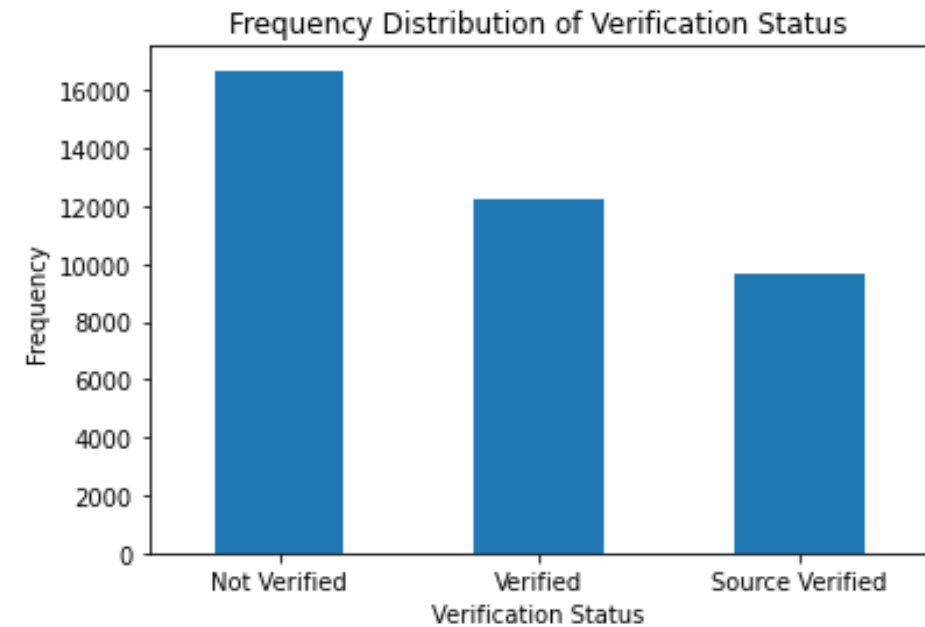
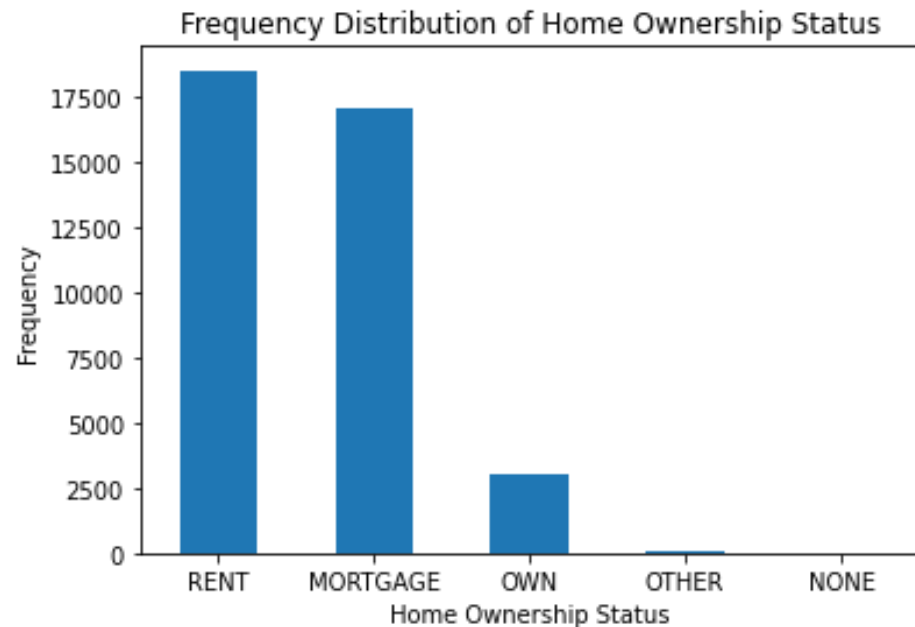
- 75% of the loan amount requested is within 15000
- About 75% of the loan is requested/granted for the term of 36 months
- Majority of the loan applicants have employment length of 10+ years



Univariate Analysis – Home Ownership, Verification Status, Purpose

Observations

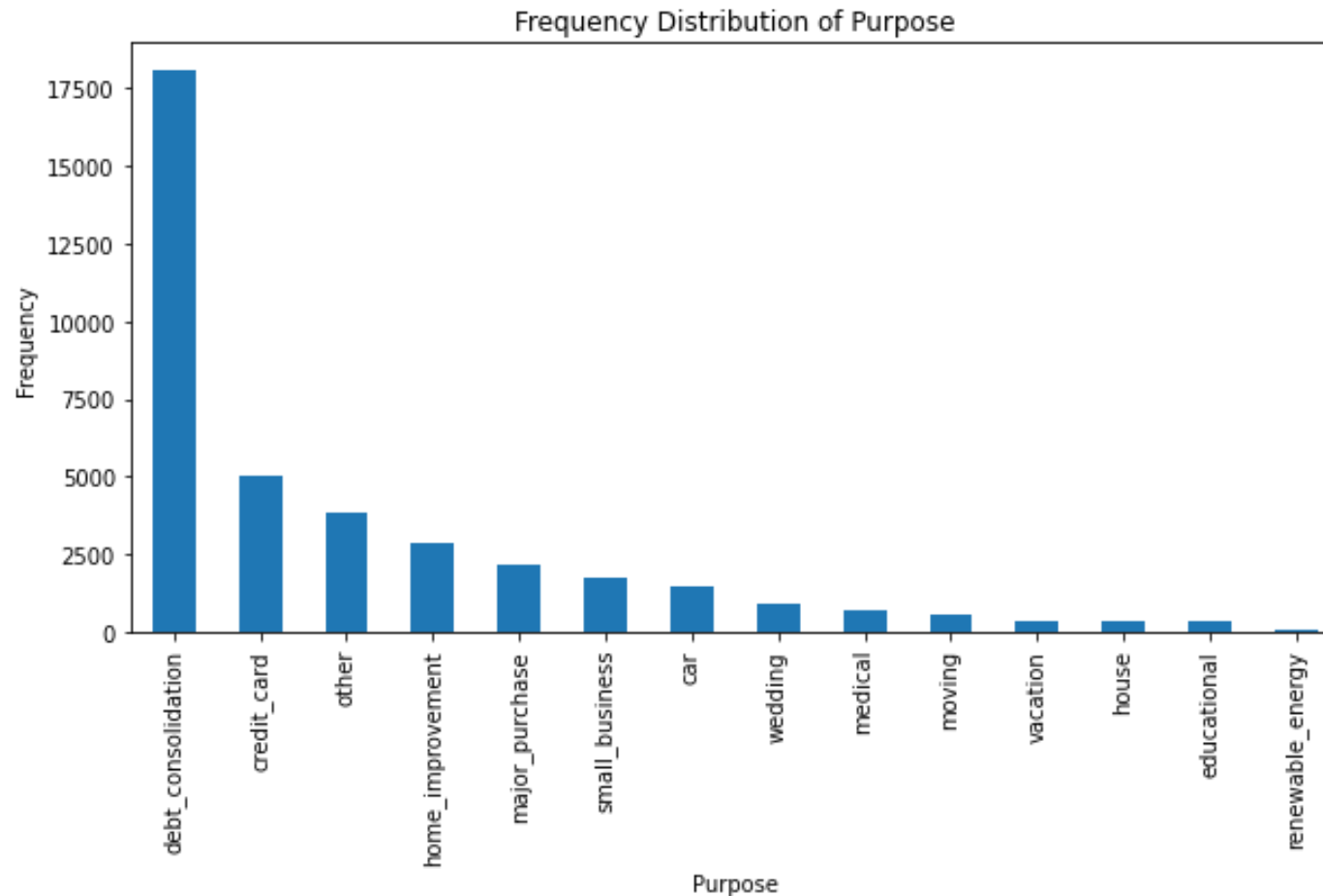
- About 48% and 44% of the loan applicants have respectively RENT and MORTGAGE as their home ownership
- About 43% of the loan applicants income were not verified



Univariate Analysis – Purpose

Observations

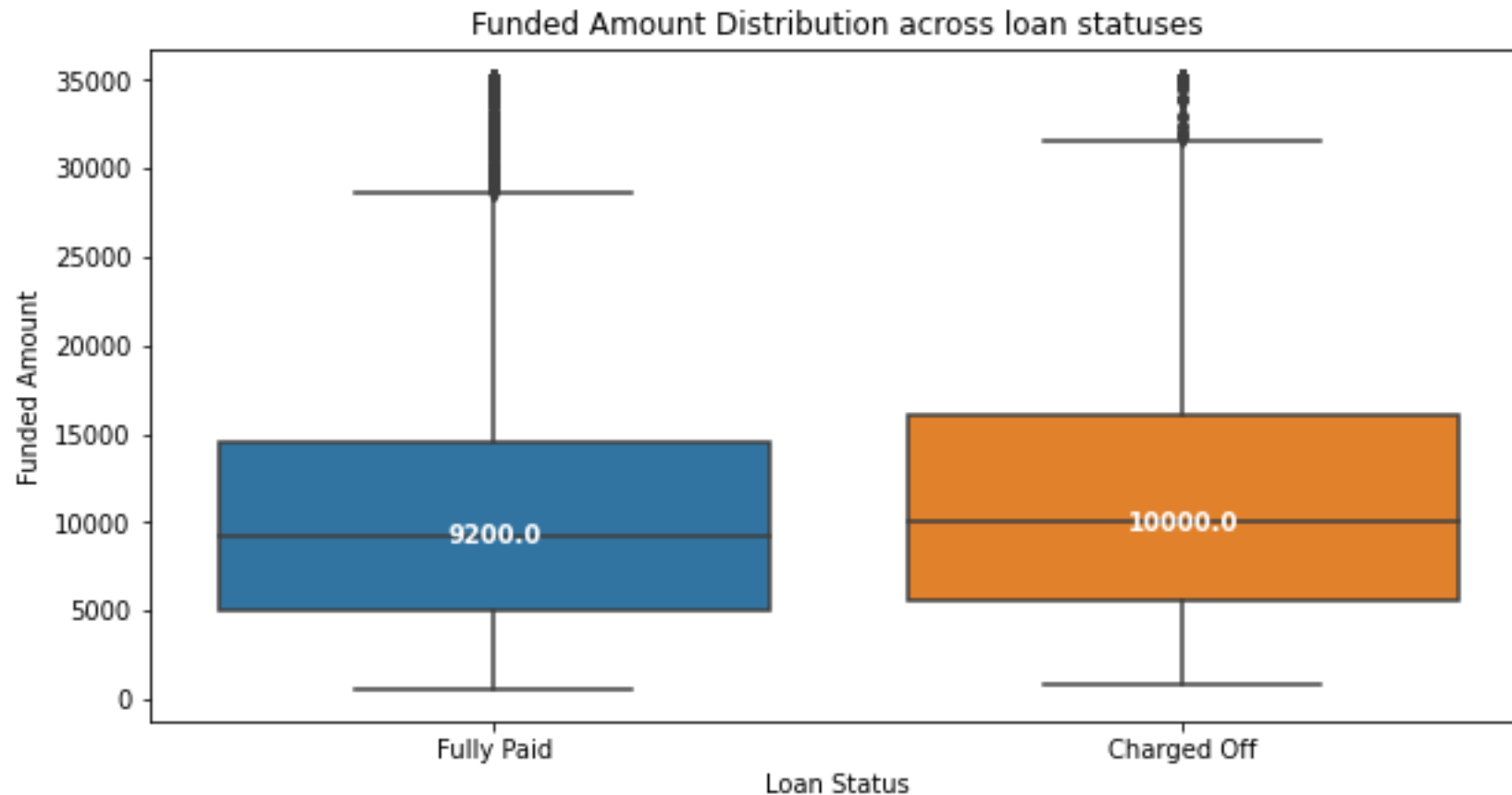
- About 46% of the loan applicants have specified purpose of loan as debt consolidation



Segmented Univariate Analysis – Funded Amount

Observations

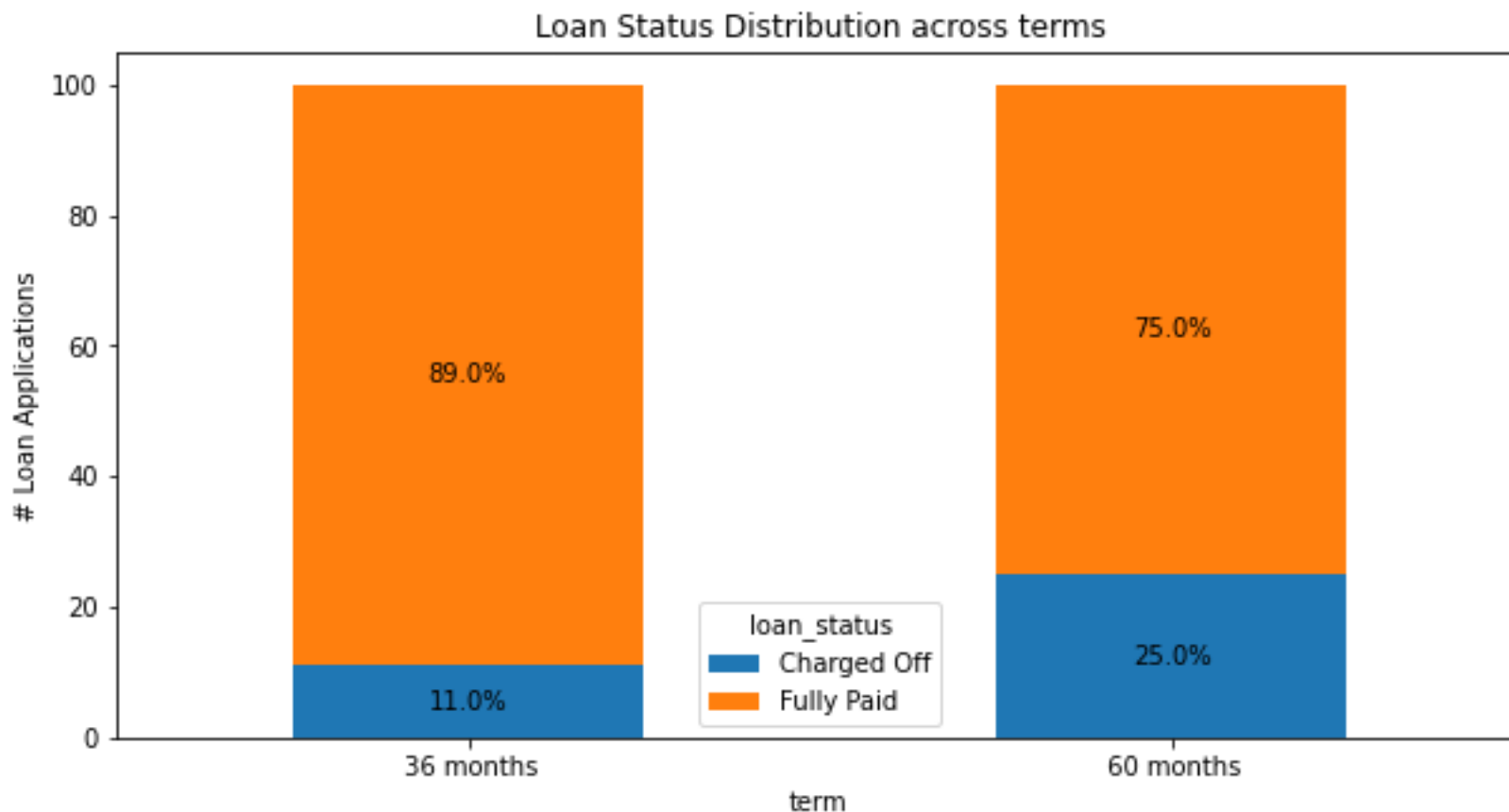
- Applicants granted with higher funded amount tend to default more than those who got lesser funded amount



Segmented Univariate Analysis – Loan Term

Observations

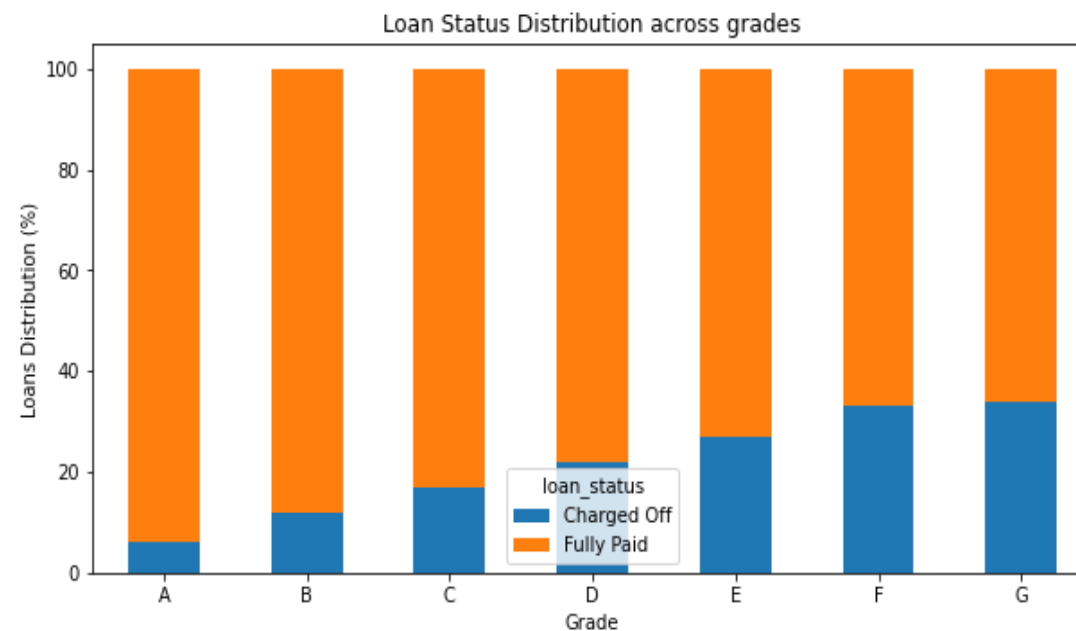
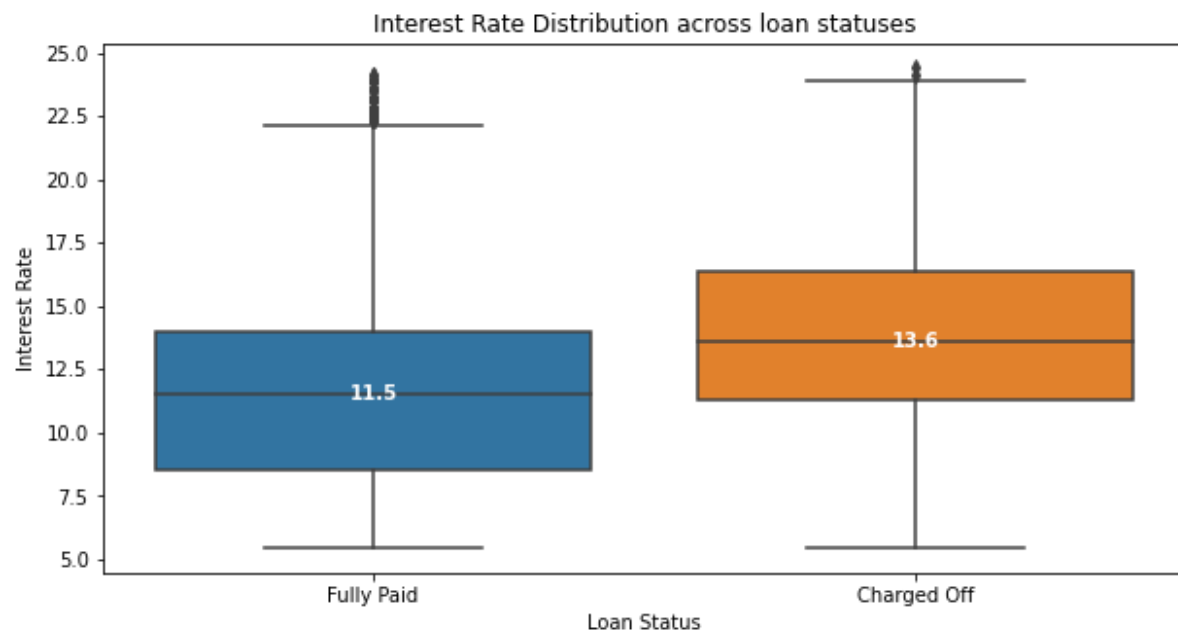
- Applicants who opted for higher term (60 months) tend to default more than those who opted for 36 months term. This is substantial considering the univariate analysis of term where 75% of the loan was requested/granted for 36 months.



Segmented Univariate Analysis – Interest Rate

Observations

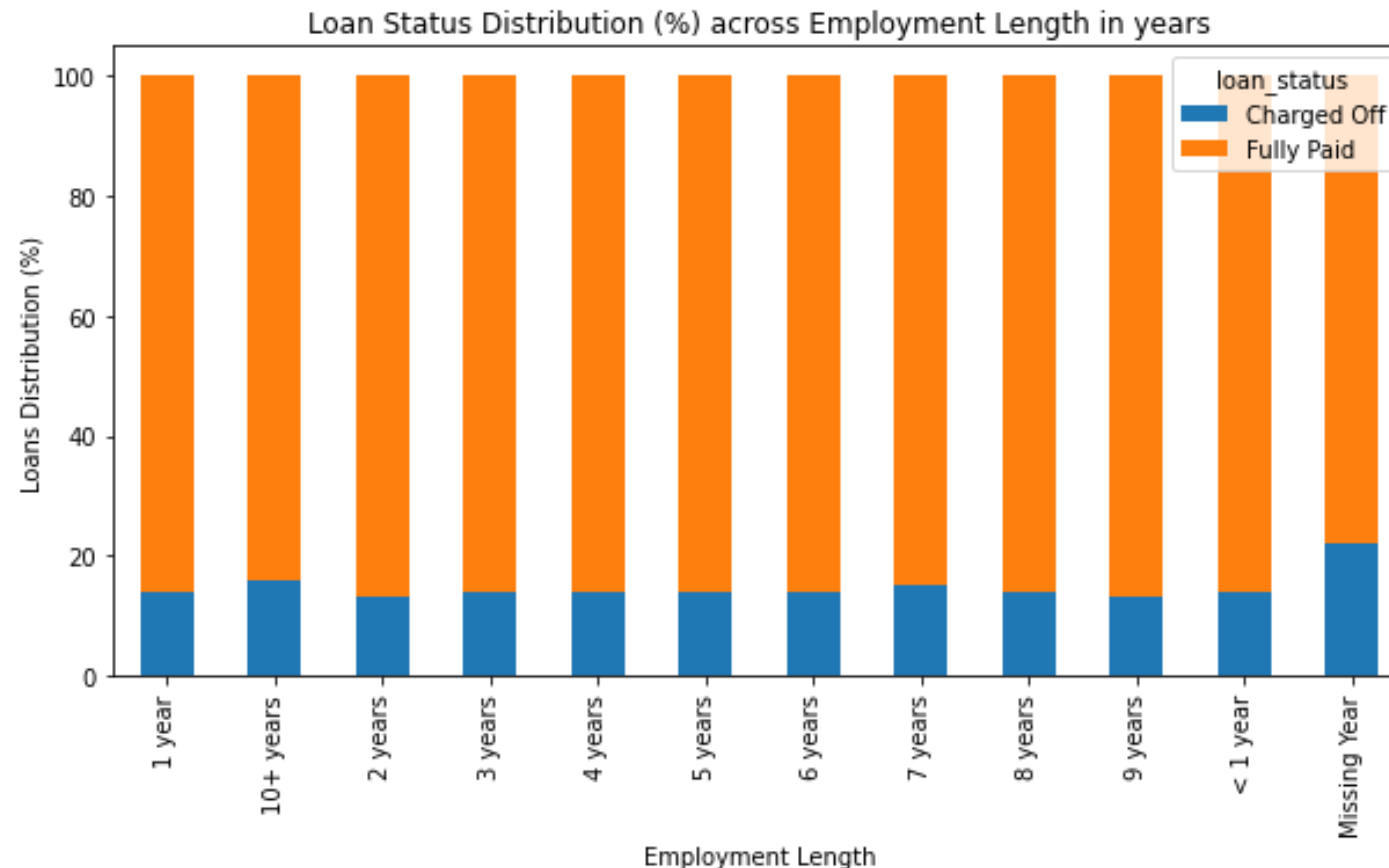
- Applicants who got loan for higher interest rate tend to default more than those who got it for lower interest rates
- Same behavior is observed for grades and sub grades where alphanumerically higher order grades and subgrades has more default rate. Business wises, grades and sub grades are correlated to interest rate. Hence, this observation is justified



Segmented Univariate Analysis – Employment Length

Observations

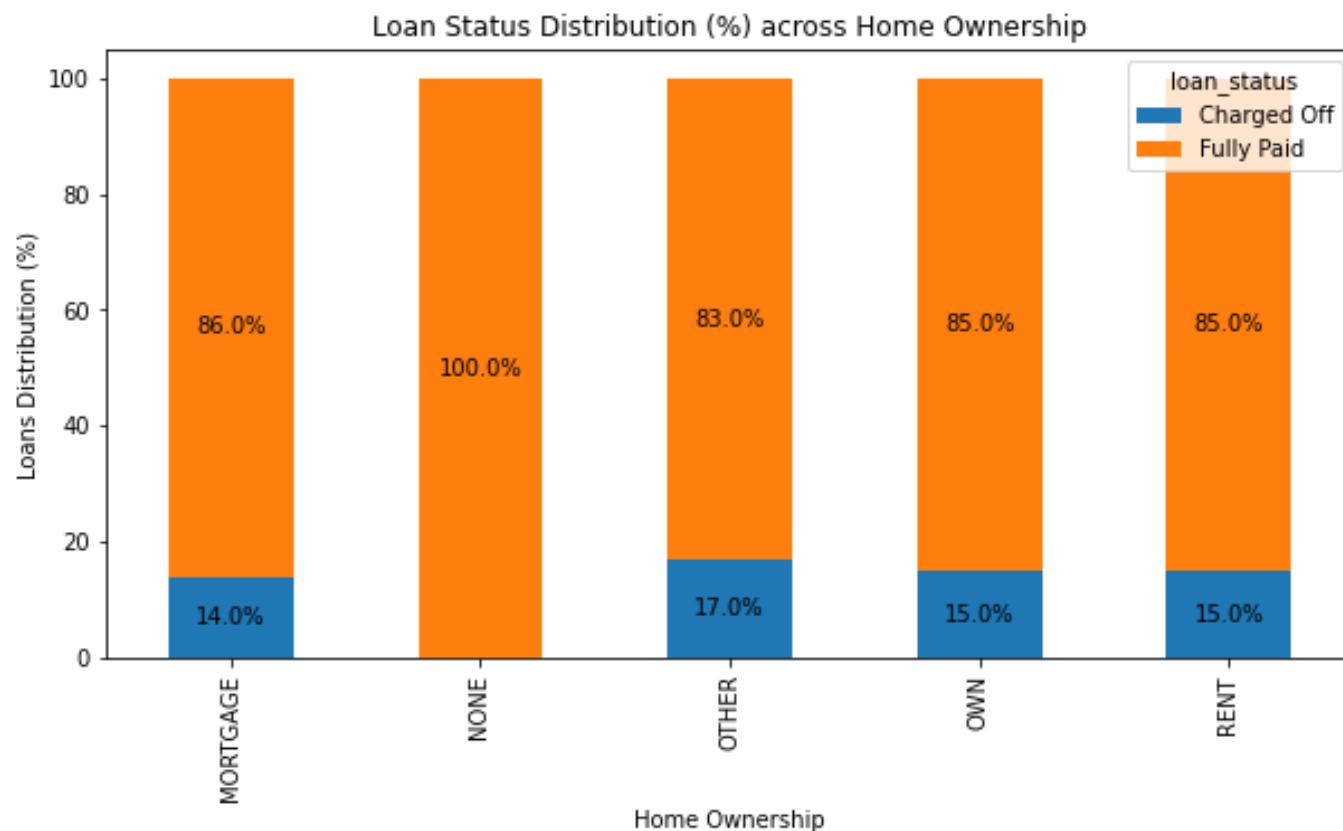
- Applicants with Employment Length of 10+ years tend to default more than other durations
- Applicants whose Employment Length is missing has higher factor of default rate(22%) than all others. This is substantial considering the univariate analysis of employment length where 10+years dominate the distribution



Segmented Univariate Analysis – Home Ownership

Observations

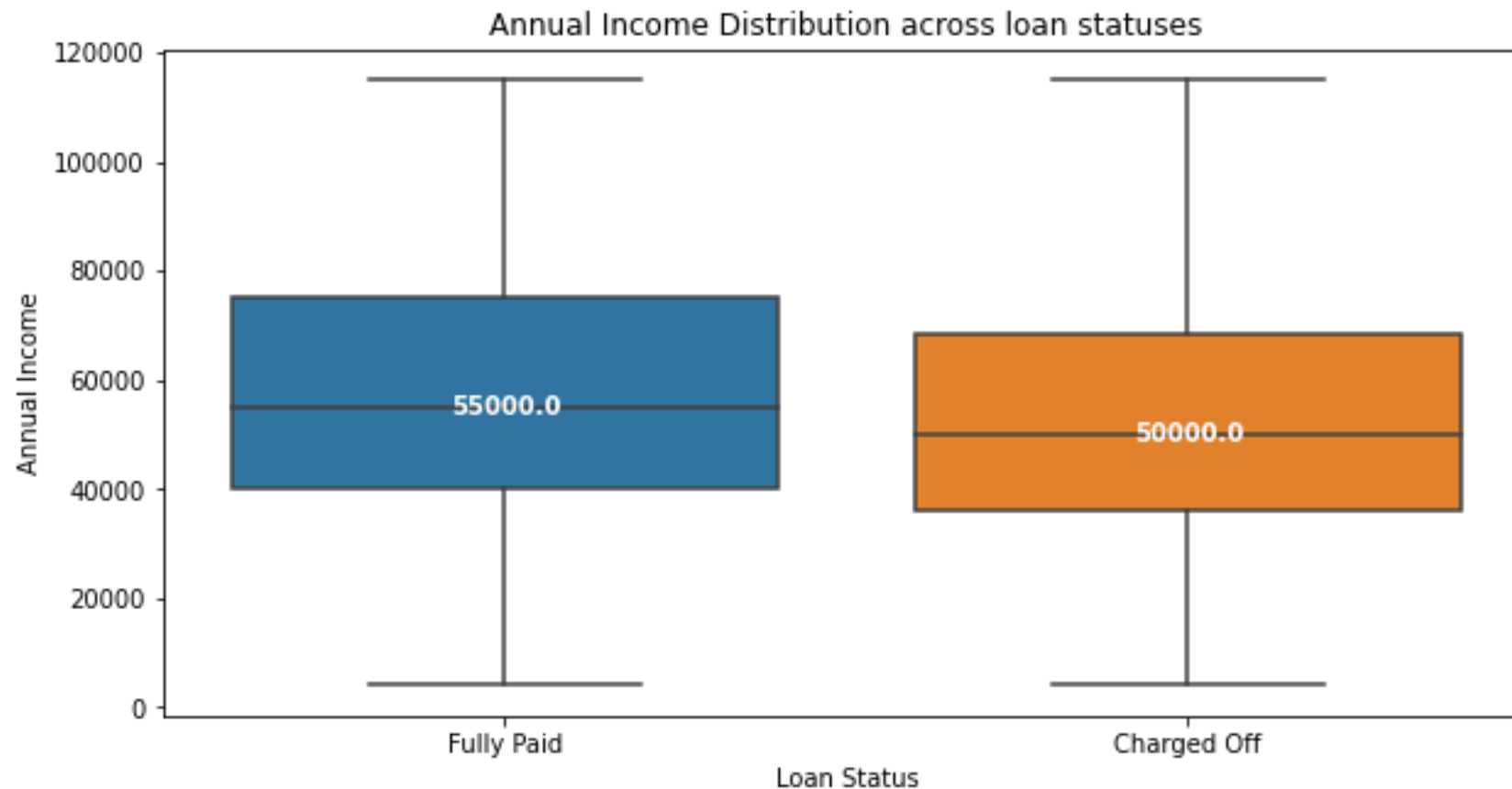
- Applicants with home ownership as “OTHER” tend to default more likely than rest of the categories. This is despite the fact that 92% of the applicants has either RENT or MORTGAGE as the home ownership.
- Interesting to note that there are no defaulters in category “NONE”



Segmented Univariate Analysis – Annual Income

Observations

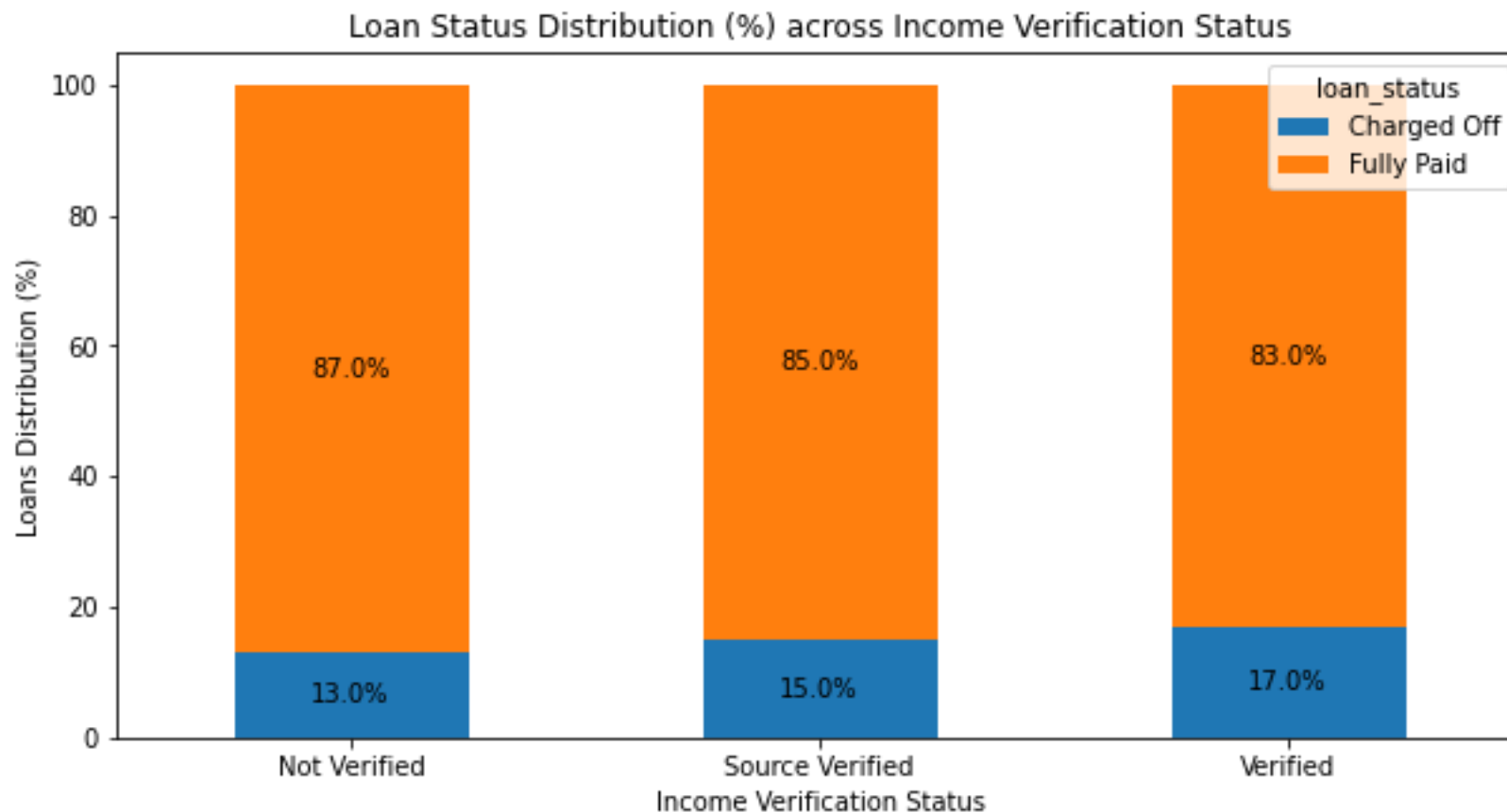
- Applicants with lesser Annual income tend to default more than those with higher levels of annual income



Segmented Univariate Analysis – Income Verification Status

Observations

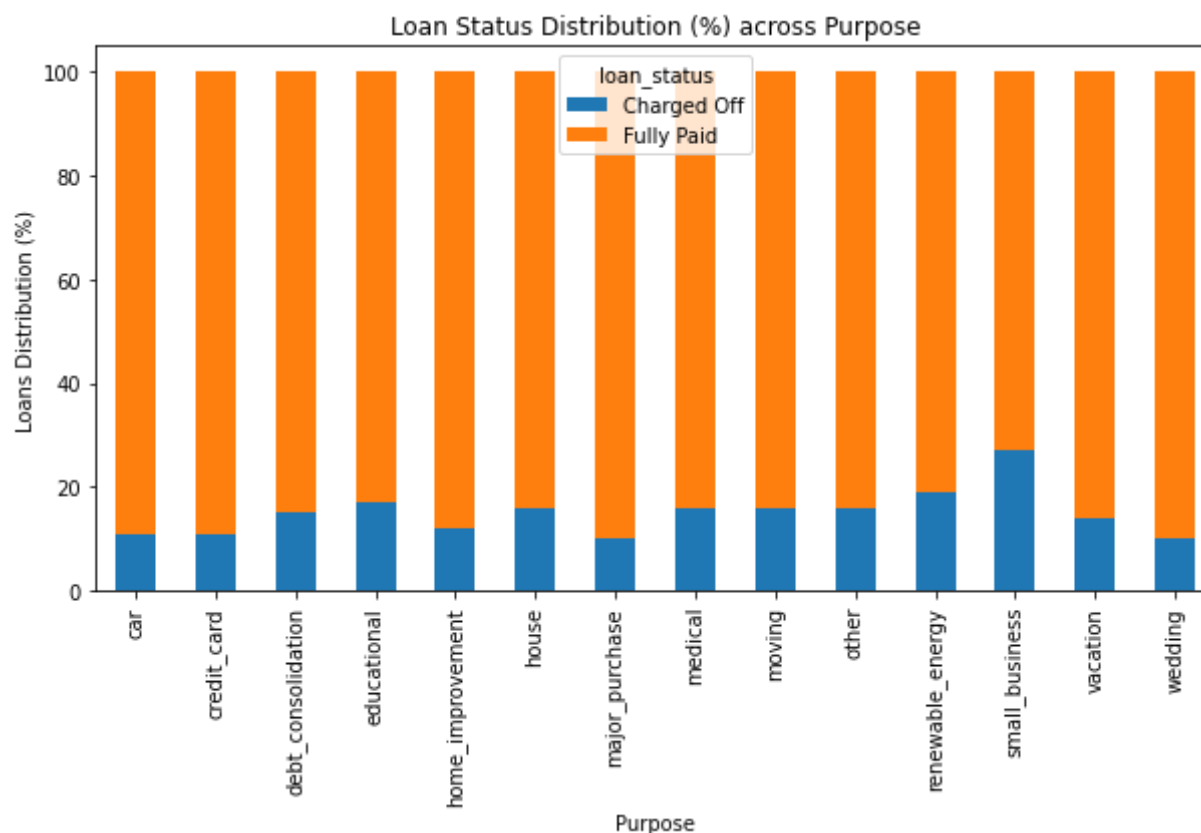
- Surprisingly, Loan Applications whose Income is verified defaults more than others by at least 2%. May be verification process needs to be revisited.
- This is substantial considering the fact that 43% of the applicants income were not verified.



Segmented Univariate Analysis – Purpose

Observations

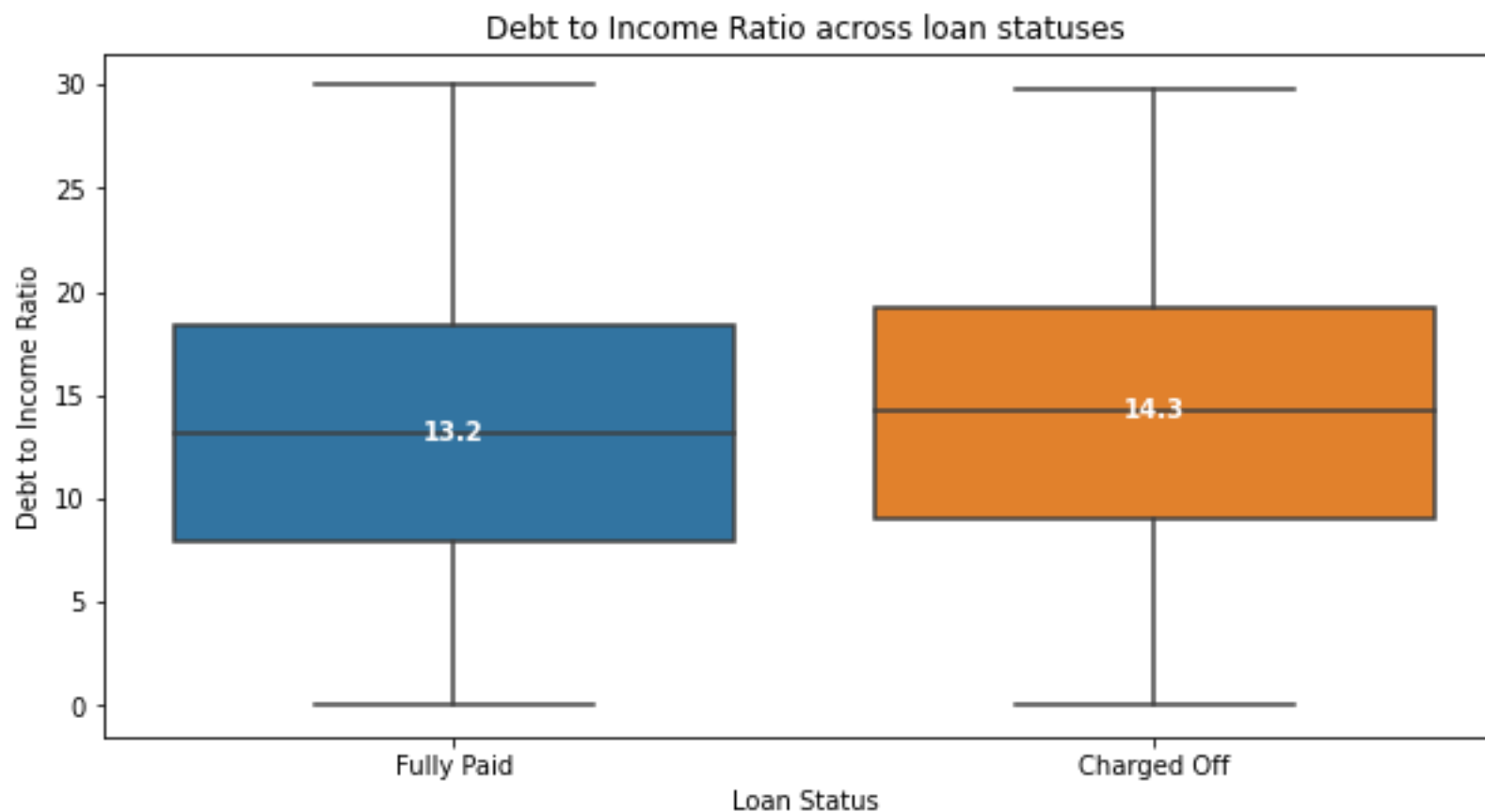
- Applicants whose purpose of loan is “educational” or “small business” tend to default more than other categories by huge difference between 8 % - 17 %.
- This is substantial considering the fact that purpose “debt consolidation” dominates the dataset (46% applicants)



Segmented Univariate Analysis – Debt to Income Ratio

Observations

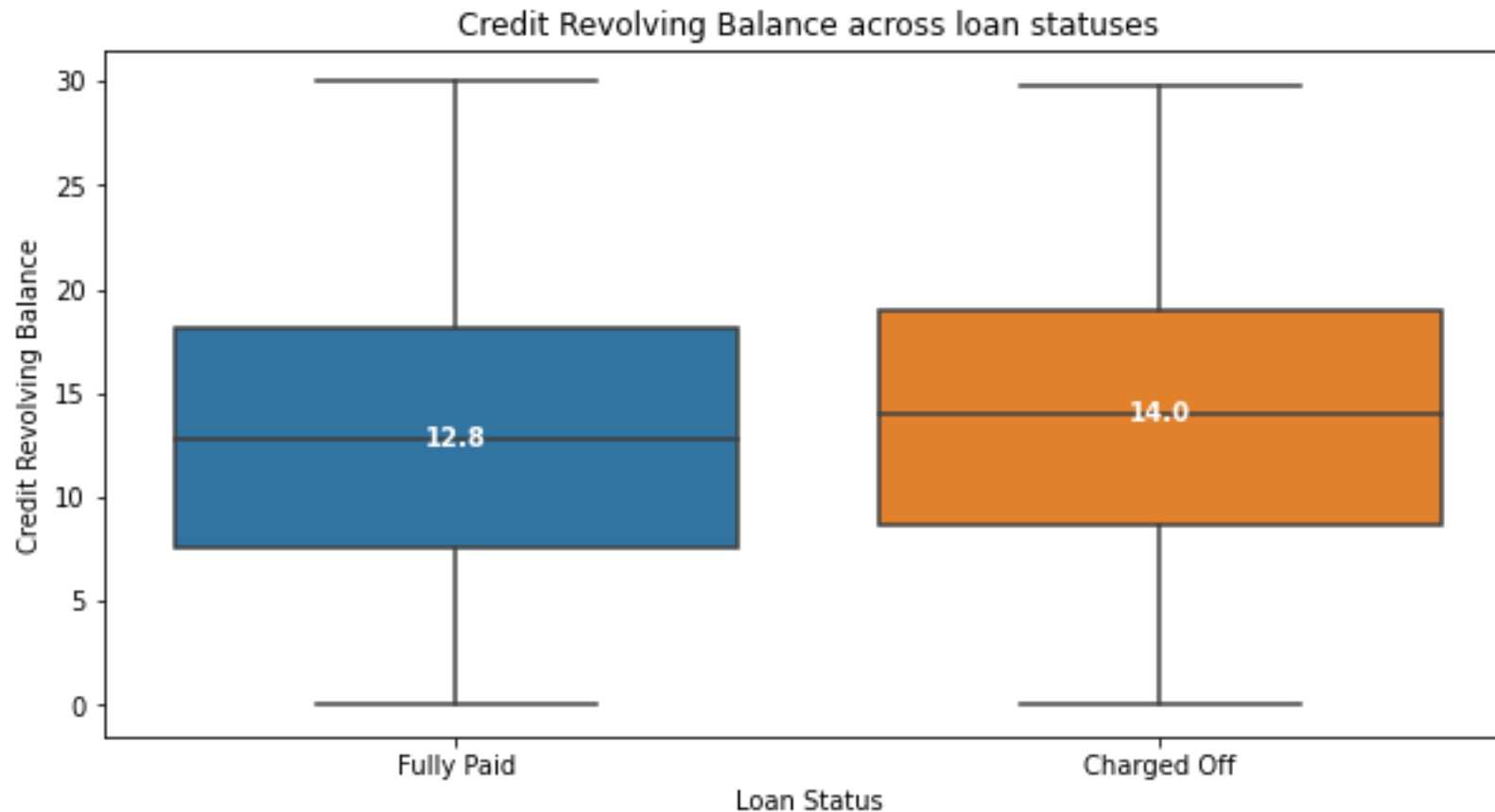
- Applicants with higher debt to income ratio tends to default more than those with lesser debt to income ratio. Although this is not a strong indicator as the difference in mean between them is merely 1



Segmented Univariate Analysis – Revolving Balance

Observations

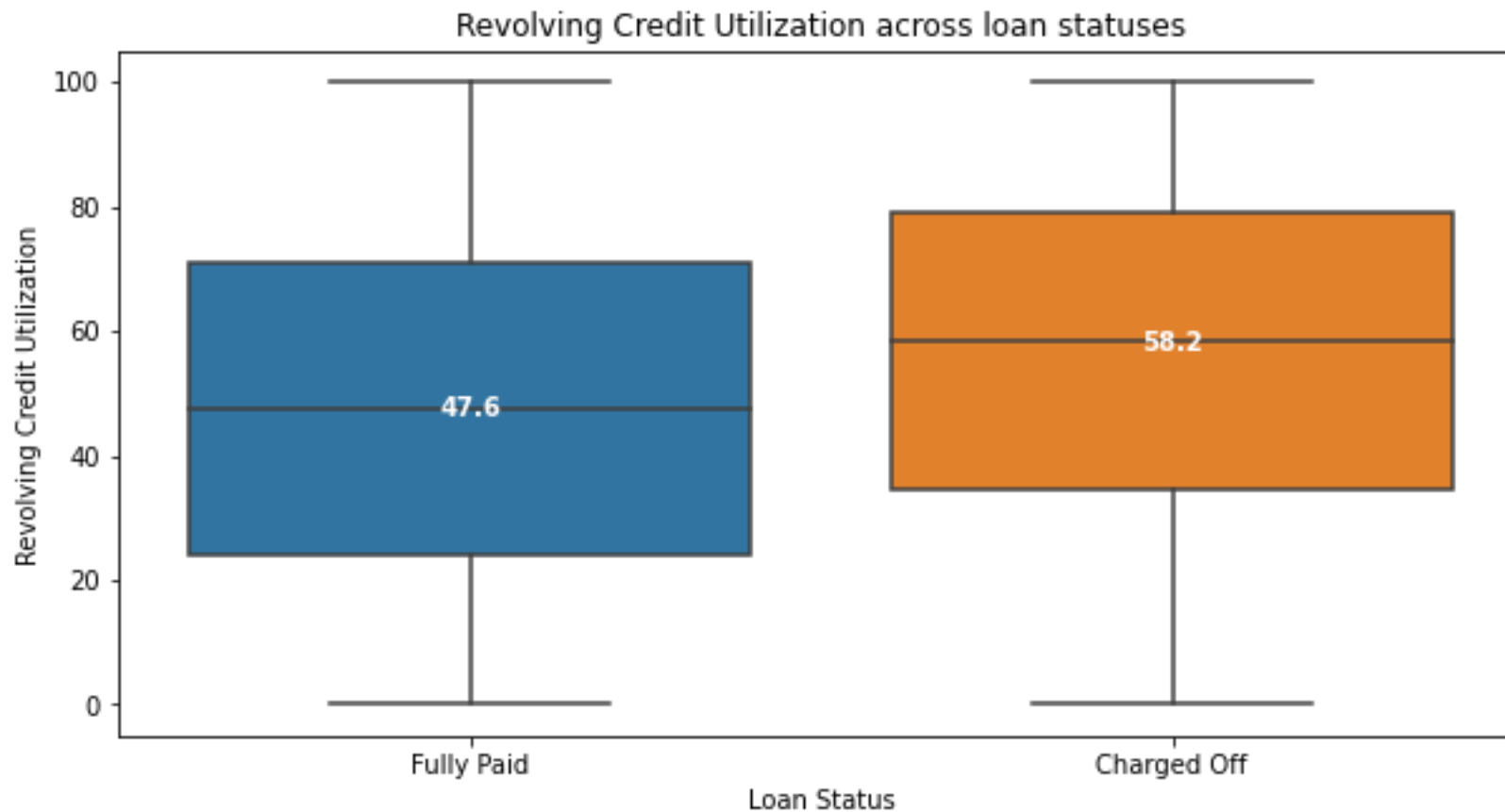
- Applicants with higher credit revolving balance tend to default more than those with lesser credit revolving balance



Segmented Univariate Analysis – Revolving Utilization

Observations

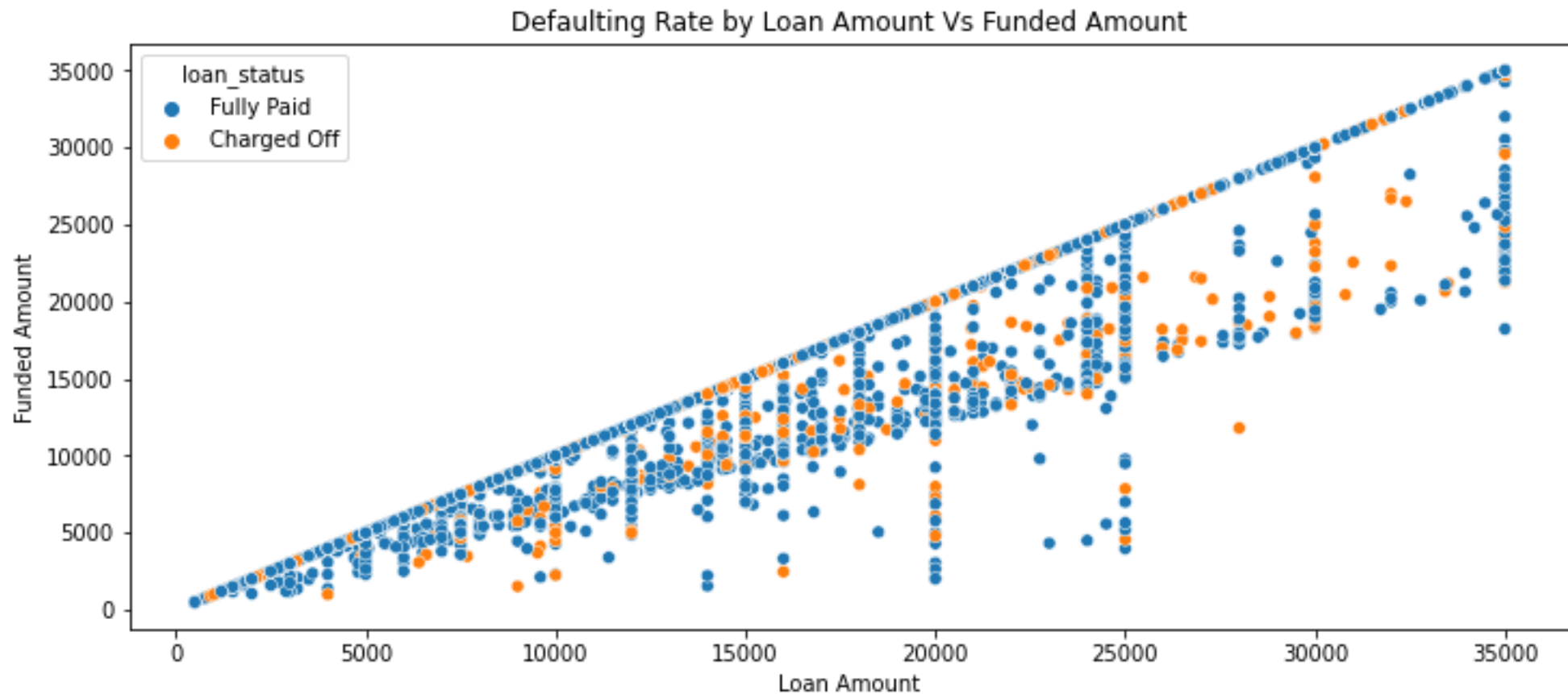
- Applicants with higher revolving credit utilization tend to default more than those with lesser credit utilization



Bivariate Analysis – Loan Amount vs Funded Amount

Observations

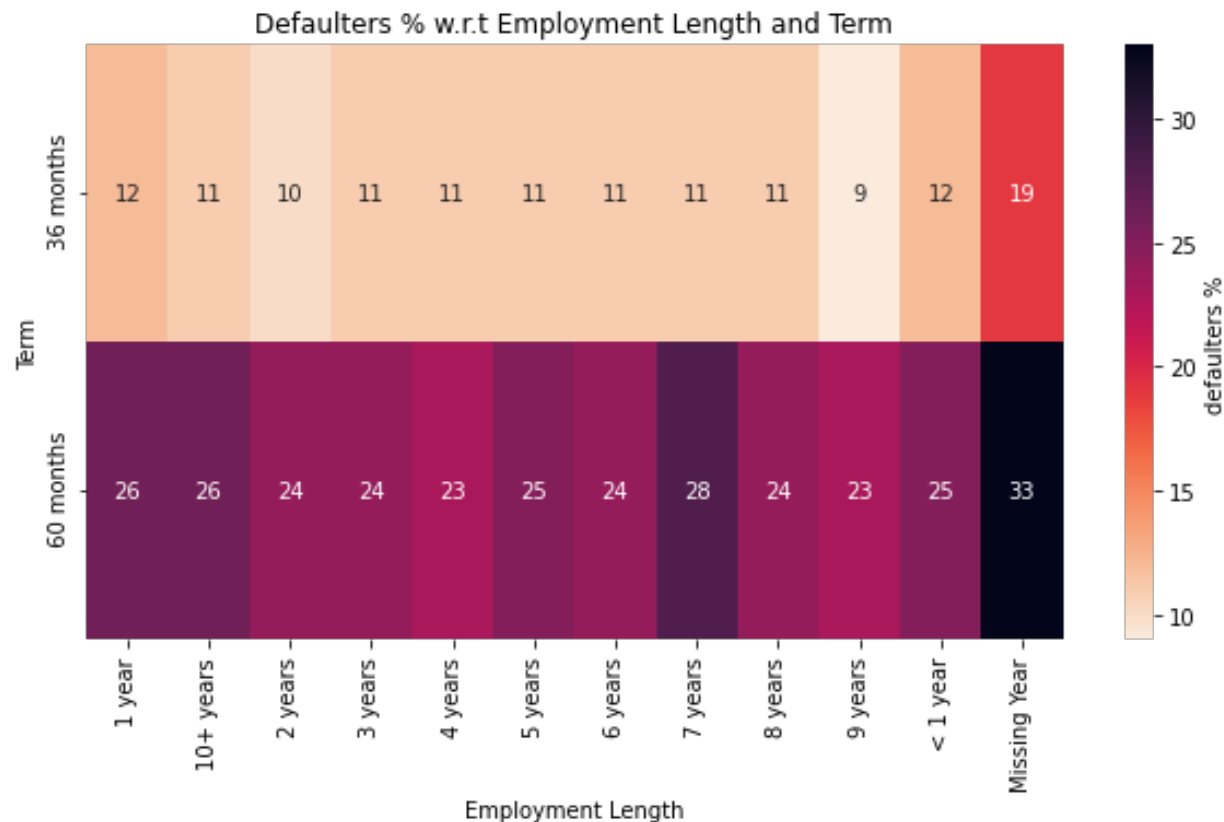
- There are very less defaulters when funded amount is nearly same as that of requested loan amount (linear relation)



Bivariate Analysis – Term vs Employment Length

Observations

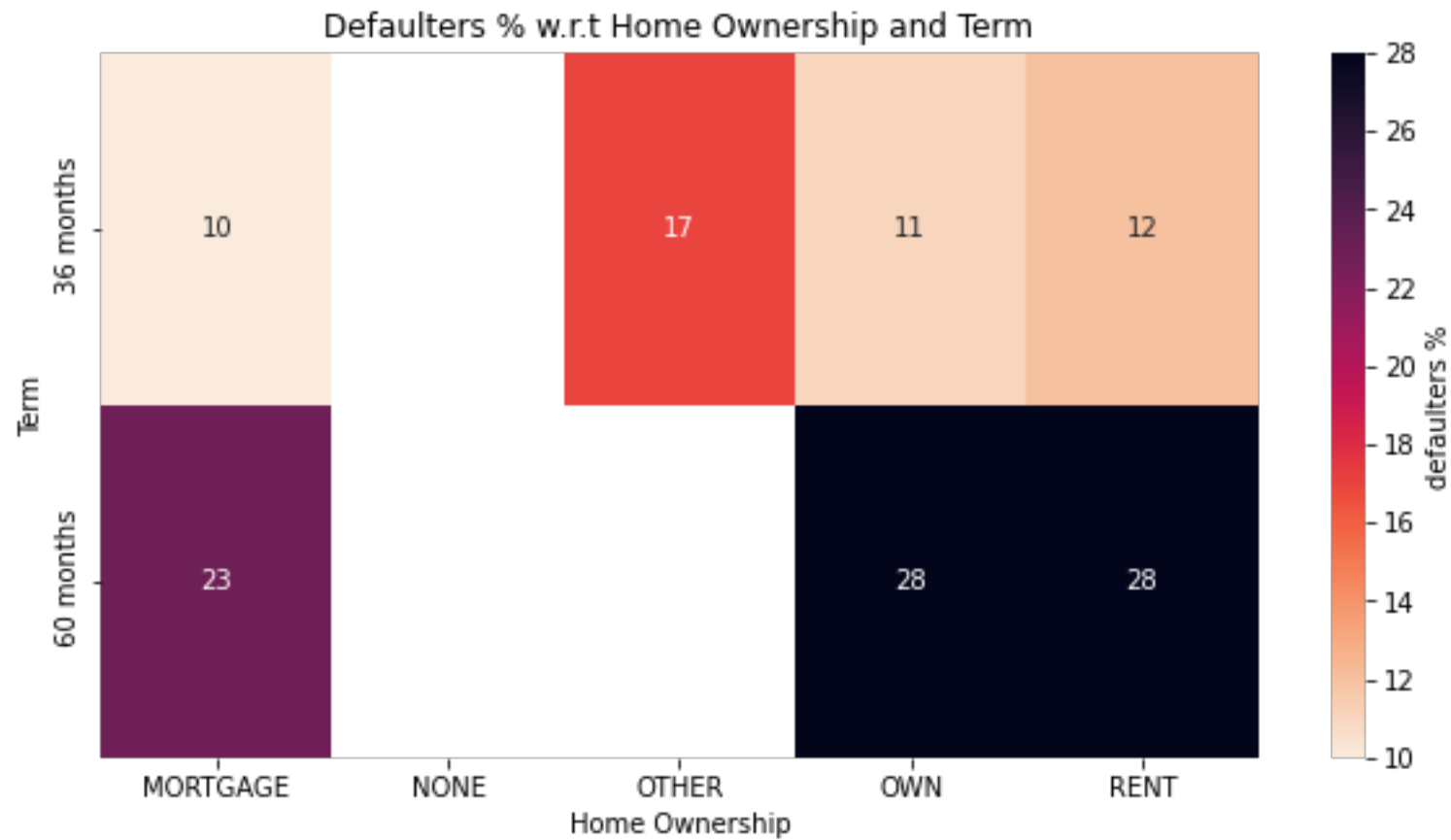
- Applicants with employment length of 7 years tend to default more than others when the term is 60 months
- Interesting to note that applicants whose employment length is missing also shows similar pattern for 60 months



Bivariate Analysis – Term vs Home Ownership

Observations

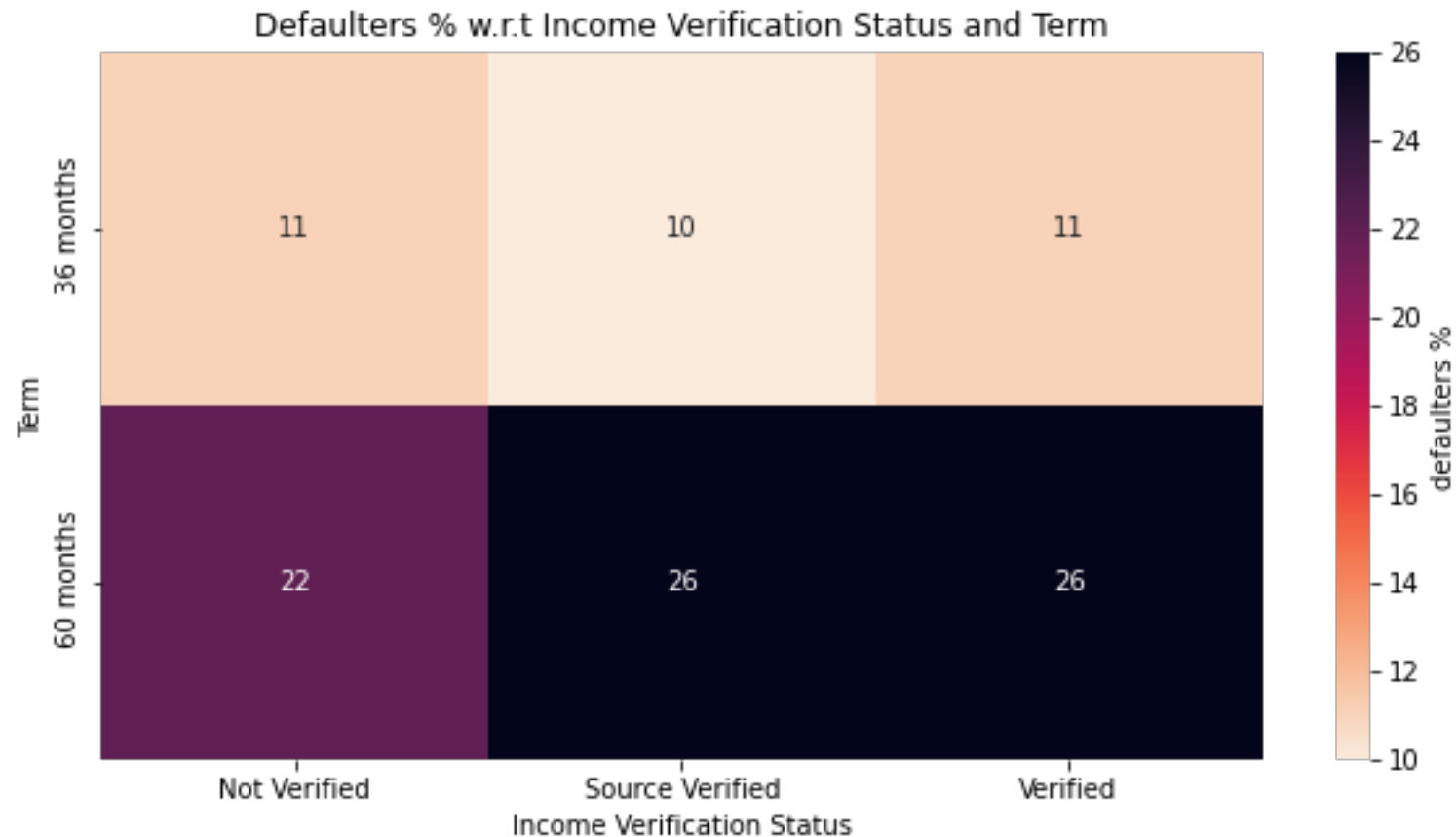
- Applicants with 60 months as term and Home ownership as OWN and RENT tend to default more likely than other combinations



Bivariate Analysis – Term vs Verification Status

Observations

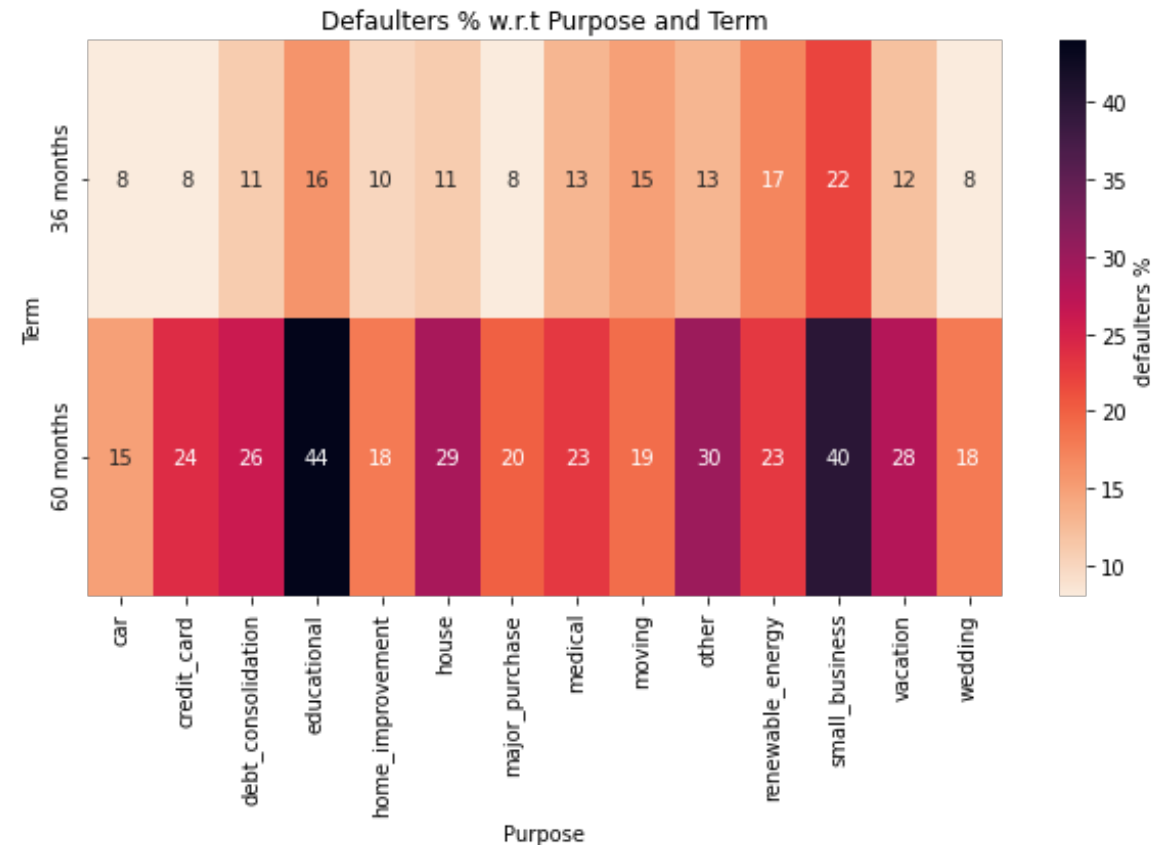
- Applicants with 60 months term and verification status as “Source Verified” and “Verified” tend to default more than other combinations



Bivariate Analysis – Term vs Purpose

Observations

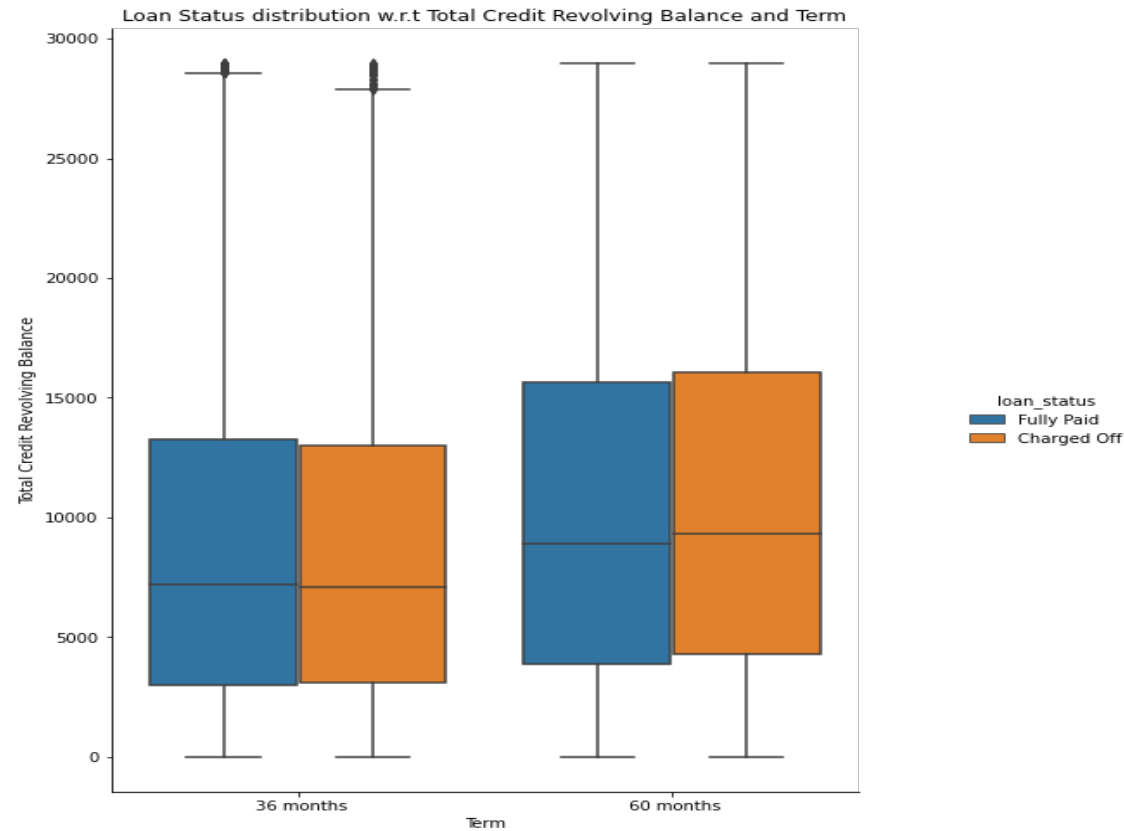
- Applicants with 60 months term and purpose as “educational” and “small business” tend to default more than other combinations



Bivariate Analysis – Term vs Revolving Balance

Observations

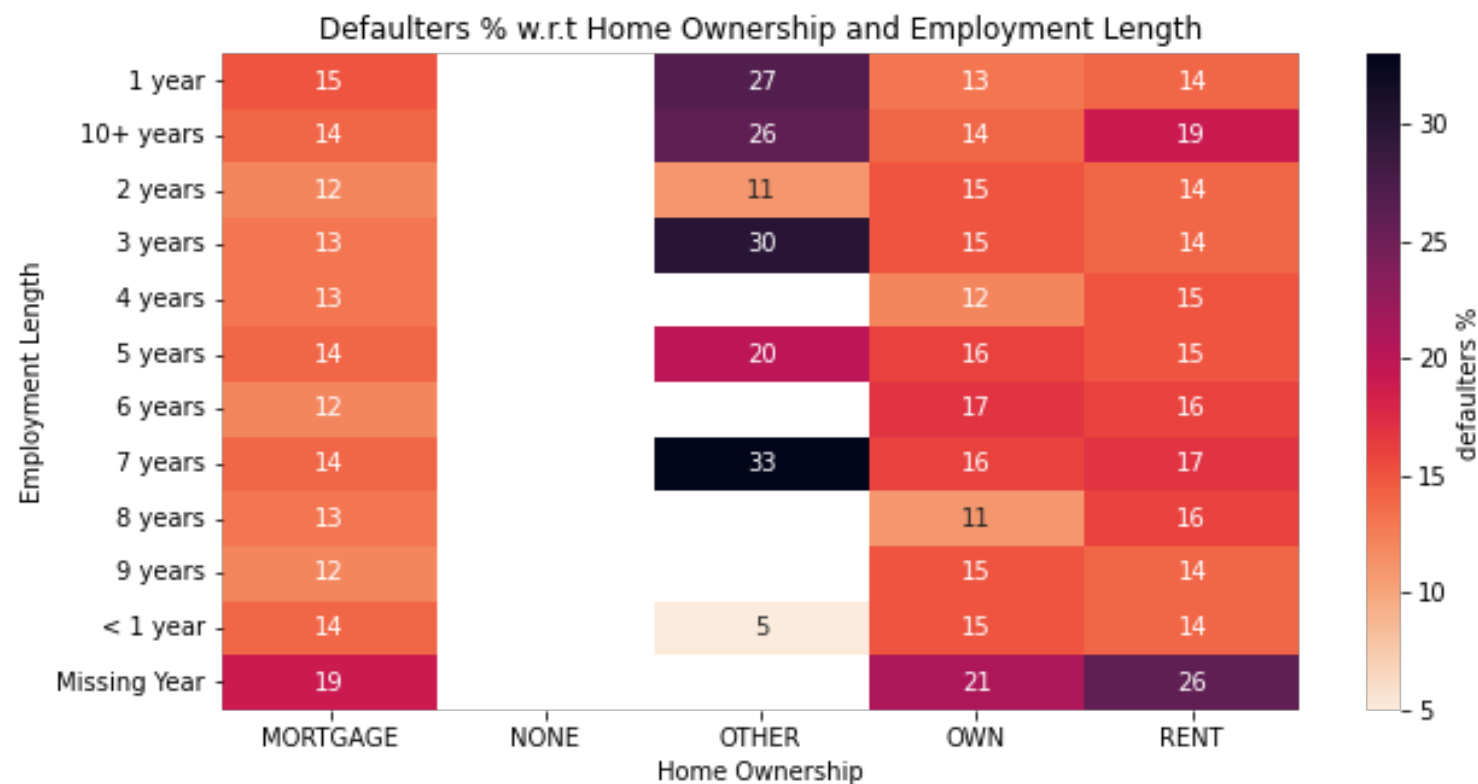
- Applicants with higher credit revolving balance tend to default likely only for 60 months term



Bivariate Analysis – Employment Length vs Home Ownership

Observations

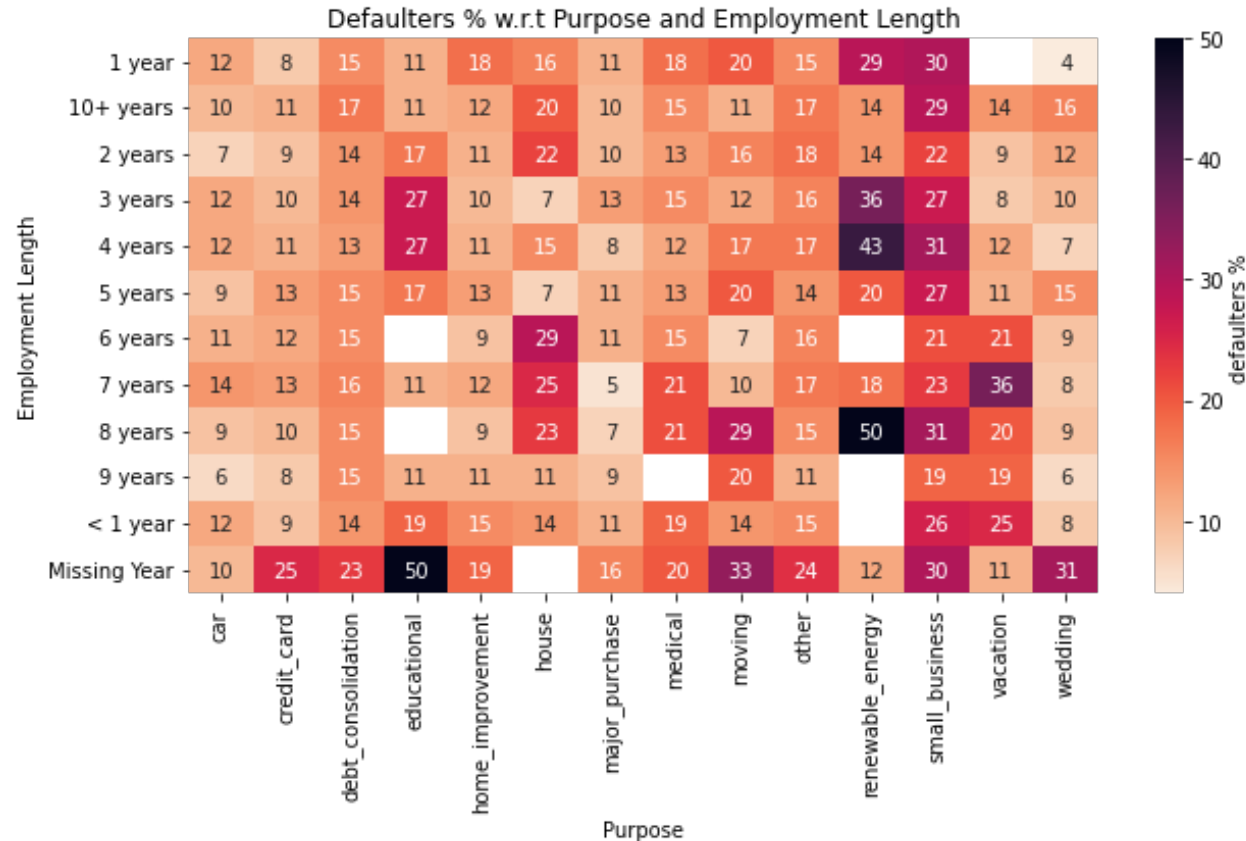
- Applicants with home ownership as “OTHER” and employment length as 7 years or 3 years tend to default more than other combinations



Bivariate Analysis – Employment Length vs Purpose

Observations

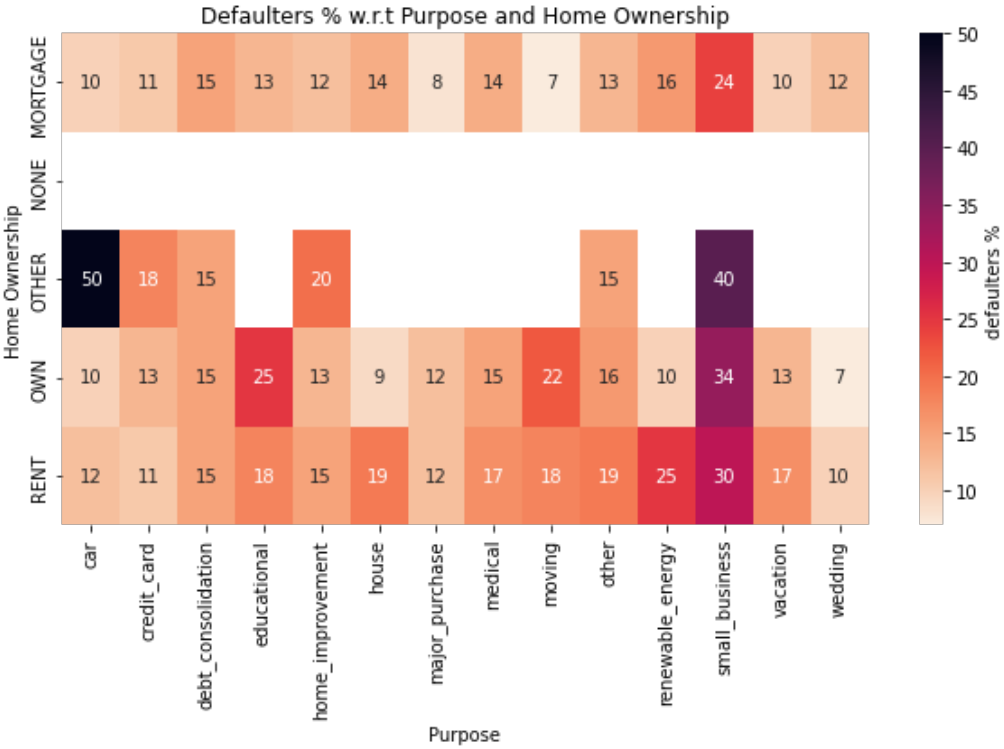
- Loan Applicants for purpose as renewable energy tend to likely default when their employment length is 4 years or 8 years.
- Interestingly, Loan Applicants with no employment length specified tend to default more for educational purpose



Bivariate Analysis – Home Ownership vs Purpose

Observations

- Applicants with the purpose of small business tend to default more likely when home ownership is “RENT”, “OWN” and “OTHER”
- Applicants with the purpose of car tend to default with home ownership as “OTHER”



Conclusion

- Applicants who opted for higher term (60 months) tend to default more than others.
- Applicants who got loan for higher interest rate tend to default more than others.
- Applicants with home ownership as “OTHER” tend to default more likely than rest of the categories. Also, there are no defaulters in category “NONE”
- Surprisingly, Loan Applications whose Income is verified defaults more than others by at least 2%. May be verification process needs to be revisited.
- Applicants whose purpose of loan is “educational” or “small business” tend to default more than other categories by huge difference between 8 % - 17 %.
- Applicants whose employment length is missing also shows similar pattern for 60 months
- Applicants with 60 months term and verification status as “Source Verified” and “Verified” tend to default more than other combinations
- Applicants with no employment length specified tend to default more for educational purpose