Lending Club Case Study

Raghavendran Ramakrishnan

Problem Statement

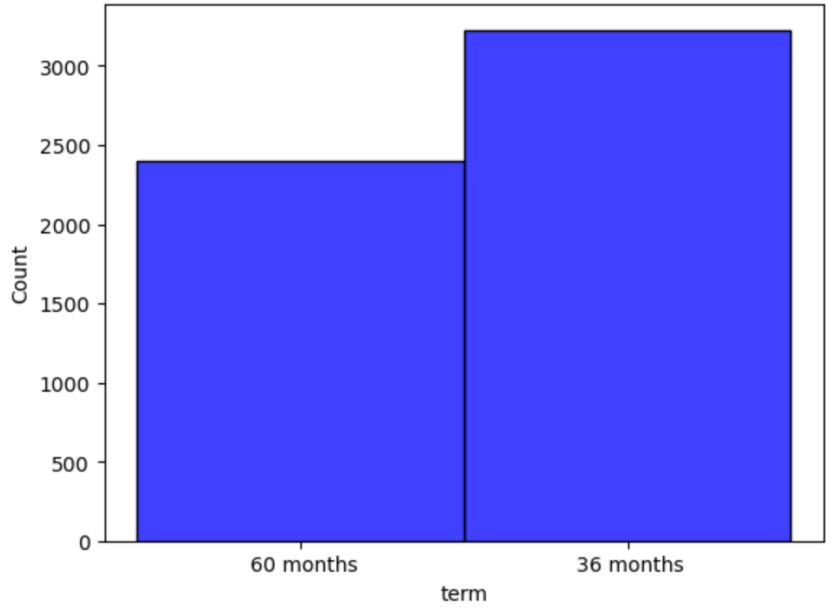
- Use EDA to analyze the given loan data set to determine the driving factors behind loan default and find out the variables that are strong indicators of default.
- Understand the given loan data set and find out the variables that are strong indicators of default using univariate and bivariate analysis.
- Validate the identified loan default driving factor variables with the help of graphs such as distribution, histogram plots and scatter plots.
- Provide Business reasons on how these variables are the driving factors behind loan default which would help the company to prevent giving loans to customers who could tend to default and giving loans to customers who would tend to repay the loan.

Analysis Approach

- Load the loan dataset into a data frame using Python.
- Check on columns that have null / NA values and remove these columns from the data set.
- Remove columns that have NA values and would not be possible indicators for loan default.
- Remove columns that have NA values and have the same values as these columns would not be possible indicators.
- Select columns that could be possible indicators for loan default for univariate analysis such as interest rate, loan term, annual income, home ownership for univariate analysis.
- Filter the dataset for defaulted loans (Loan status is Charged Off / Defaulted status) and perform univariate analysis on selected columns / variables.
- Peform data cleaning on the selected columns (conversion to integer / float)
- Perform Bivariate analysis on columns / variables that could possible driving indicators for loan default
- Document the analysis and perform business reasoning into how these variables can tend a loan to default.

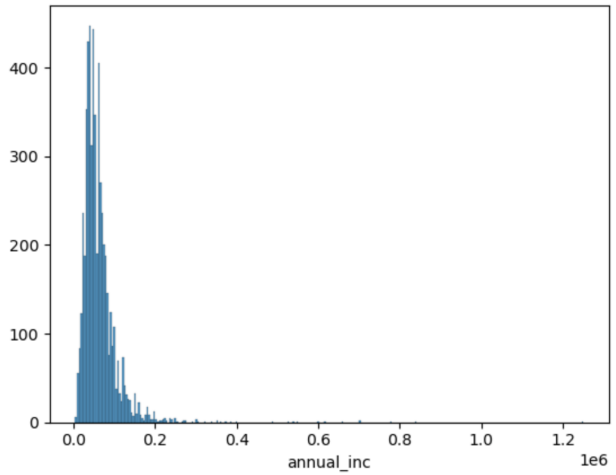
Univariate Analysis - Term

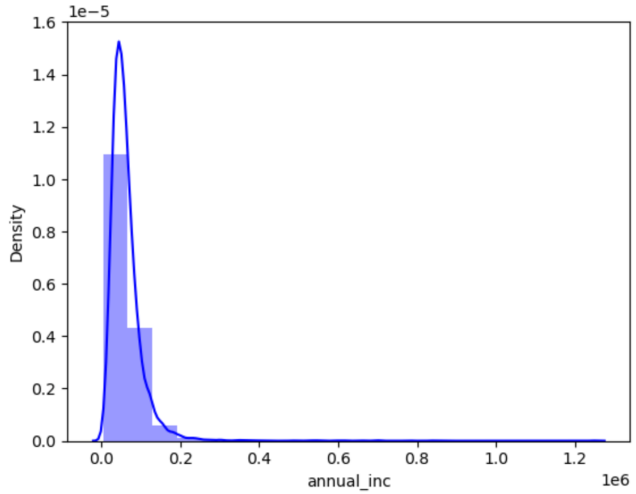
- Loan Term is the length of the loan and can be 36 months and 60 months.
- Loan Term at 36 months had more defaulters than 60 months.

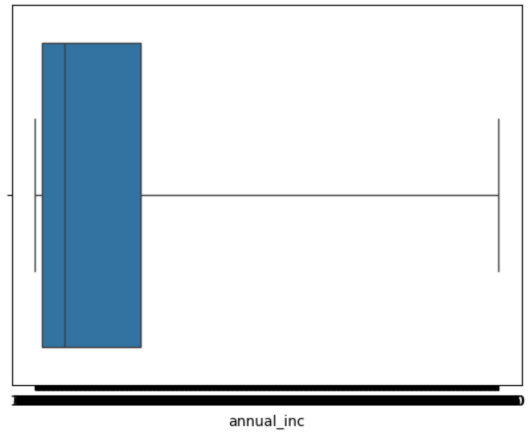


Univariate Analysis – Annual Income

- Annual income of the customer who has taken the loan and defaulted.
- Customer who had annual income less than 20,000 tended to default more.
- Box plot indicates a right skew.

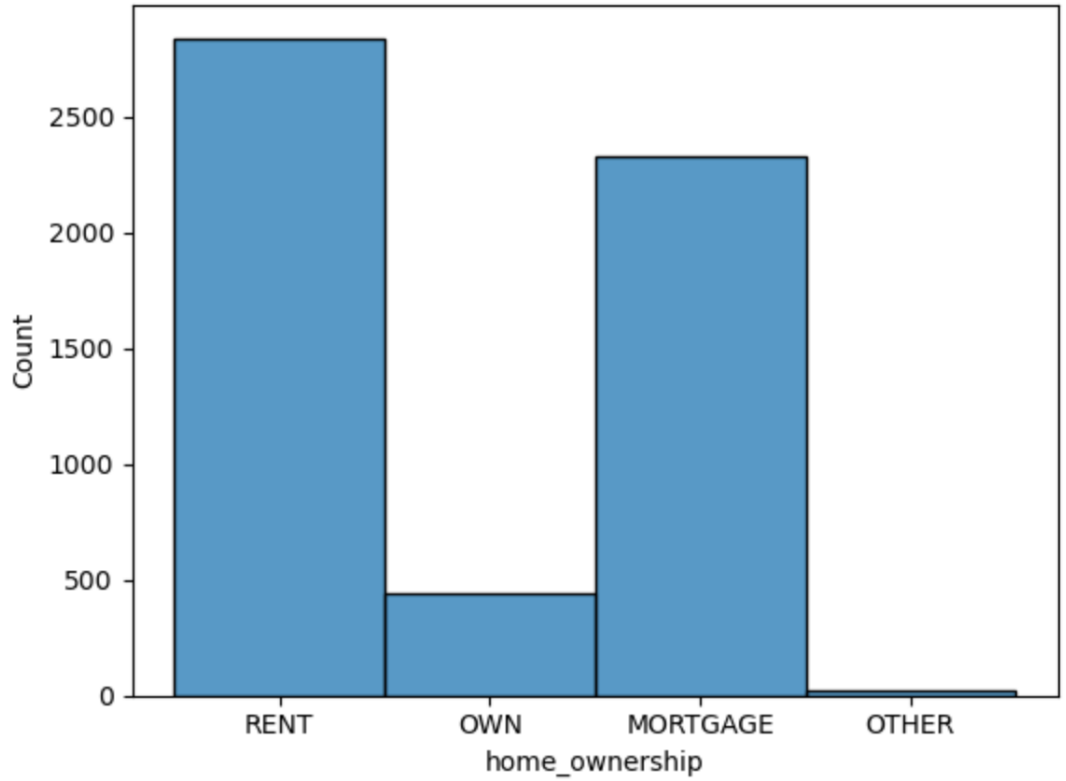






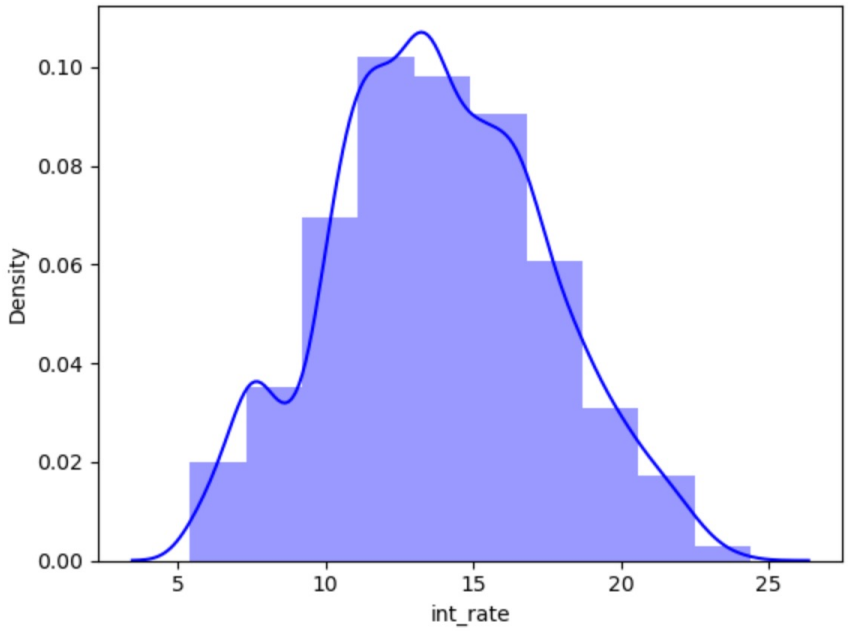
Univariate Analysis – Home Ownership

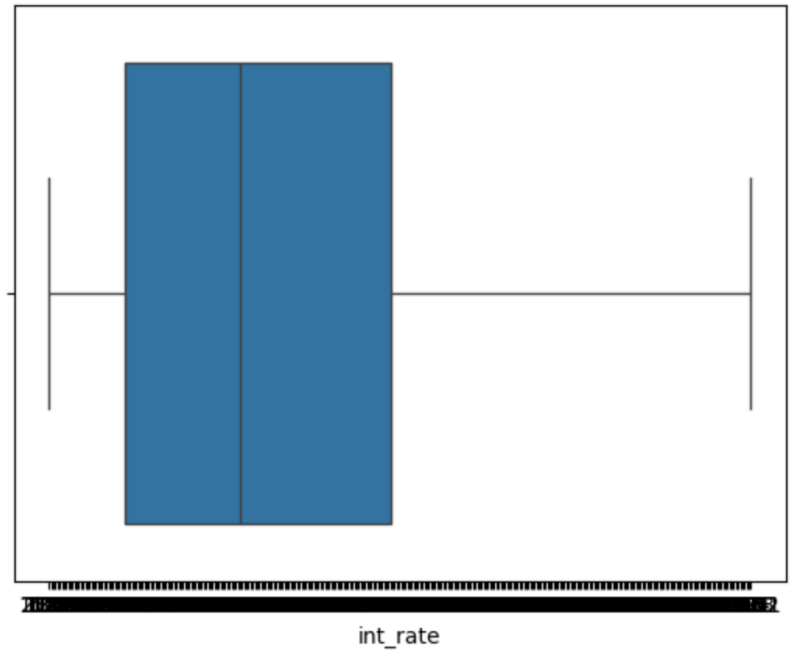
- Home ownership indicates the home ownership type of the customer such as Rent, Own, Mortgage and others.
- Customer who rented or had a Home Mortgage tended to default more on the loan than customers who owned their own home.



Univariate Analysis – Interest Rate

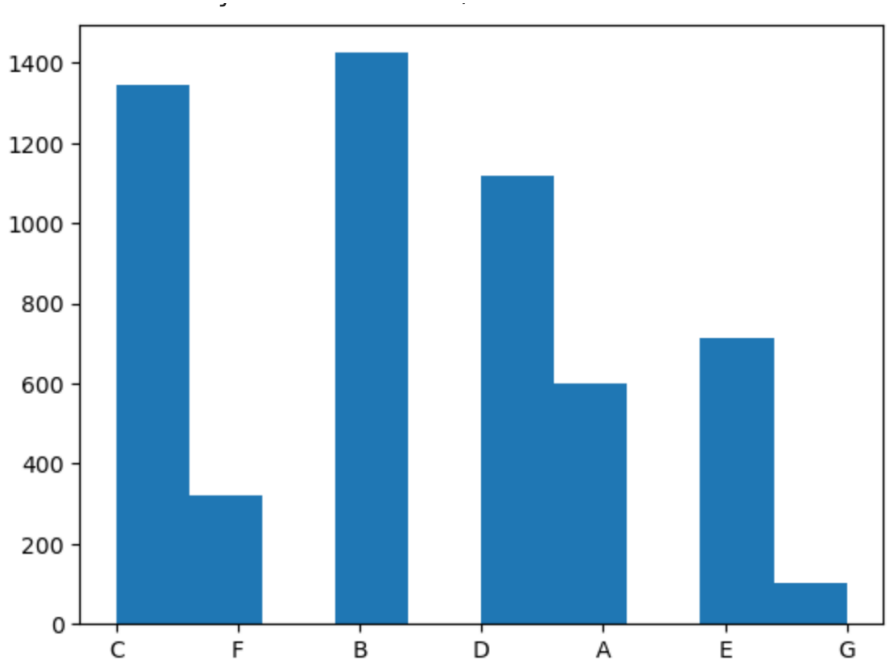
- Interest Rate is the rate of the interest of the for the loan that the customer had defaulted.
- Customers who had taken a loan with higher interest rate (12-17%) tend to default more.
- Box plot indicates a skew to the right.
- Mean = 13.82 / Median = 13.61





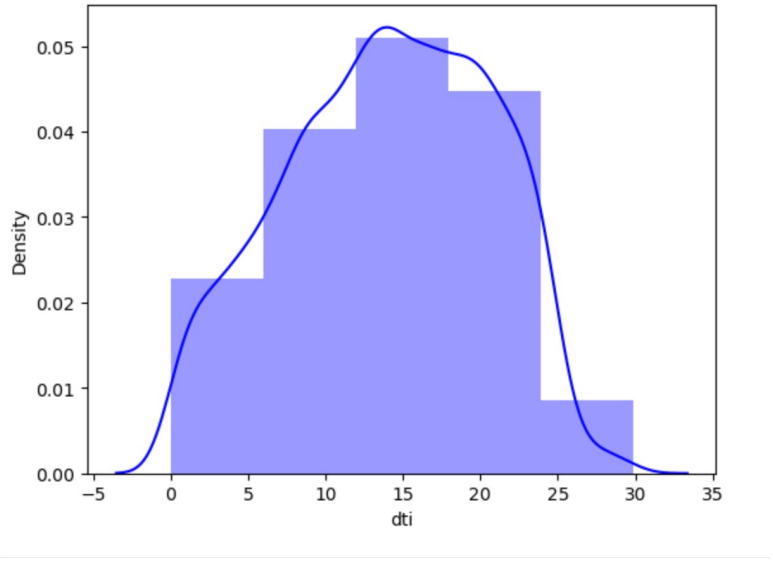
Univariate Analysis – Loan Grade

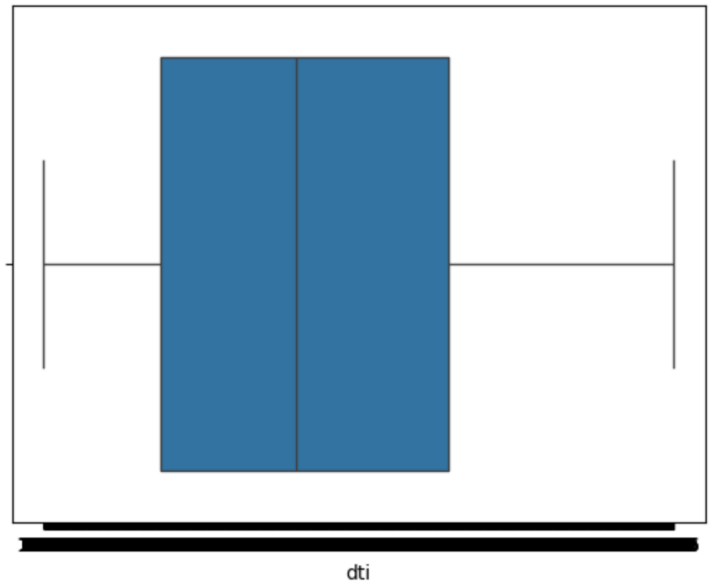
- Loan Grade is the LC assigned loan grade that the customer had taken and defaulted.
- Customers who were assigned loan grade B,C,D and E tended to default more than customers how were assigned loan grade G,F and A.



Univariate Analysis – DTI

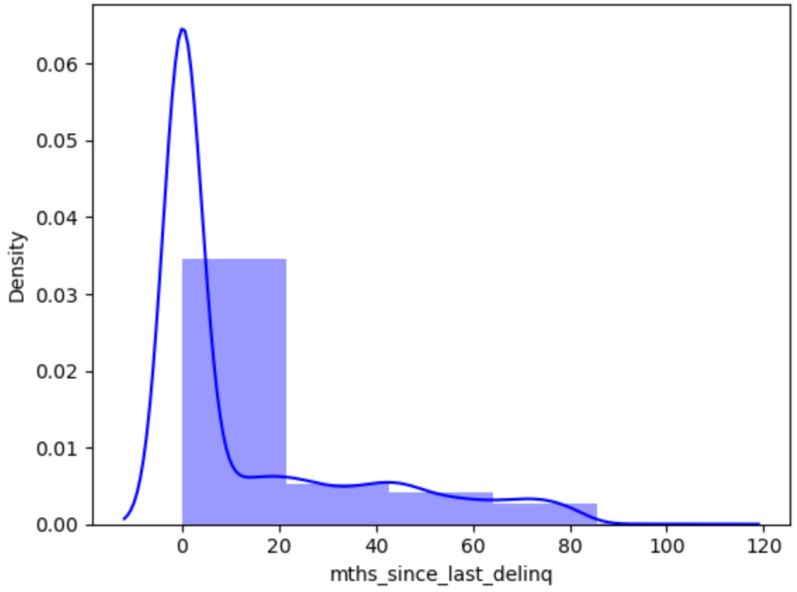
- DTI - A ratio calculated using the borrower's total monthly debt payments on the total debt obligations.
- Cutomers with DTI (10-23) tended to default more than customers who had DTI more than 25.
- DTI Mean = 14, DTI Median = 14.29

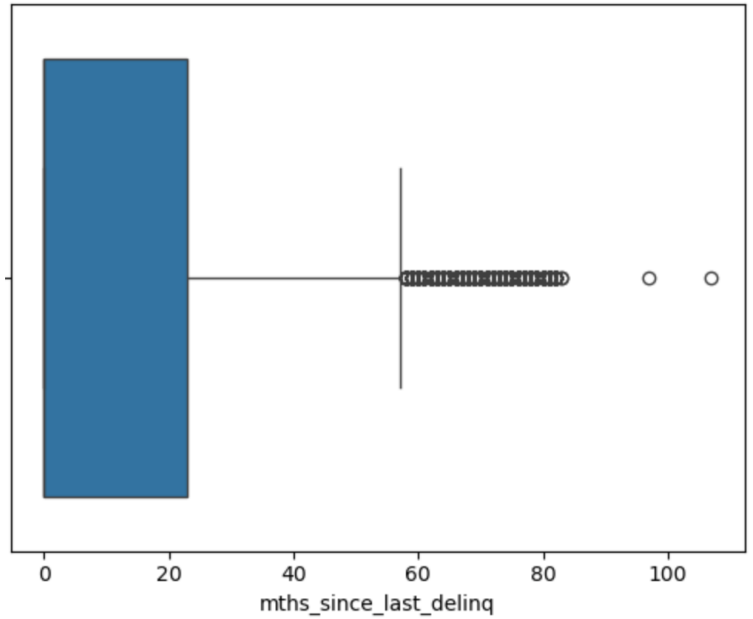




Univariate Analysis –mths_since_last_delinq

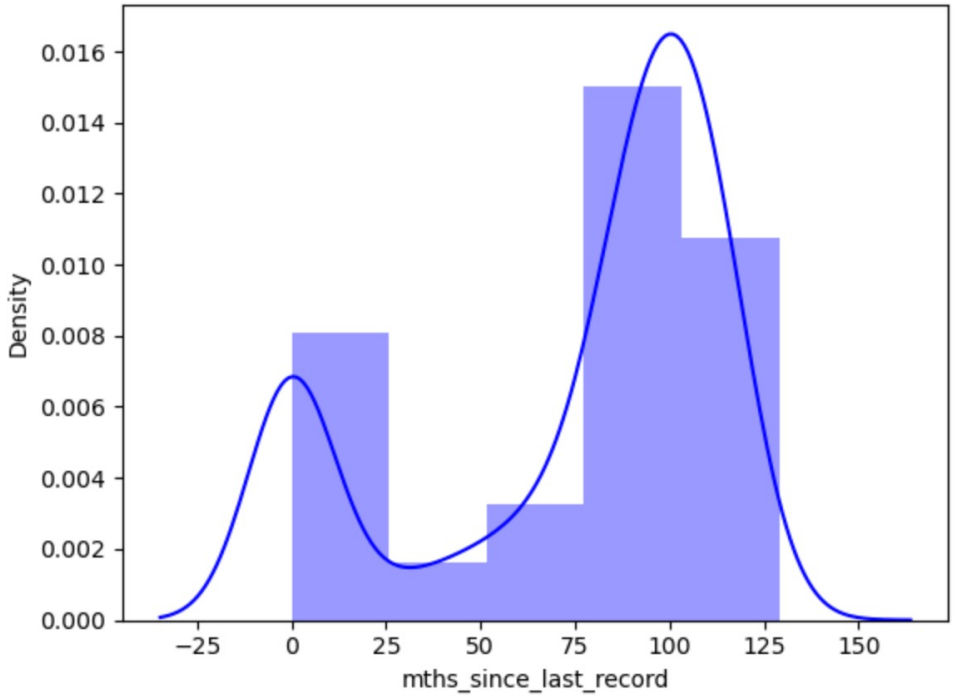
- mths_since_last_delinq - The number of months since the borrower's last delinquency.
- Customers who had below 20 months before the last delinquency tended to default more.





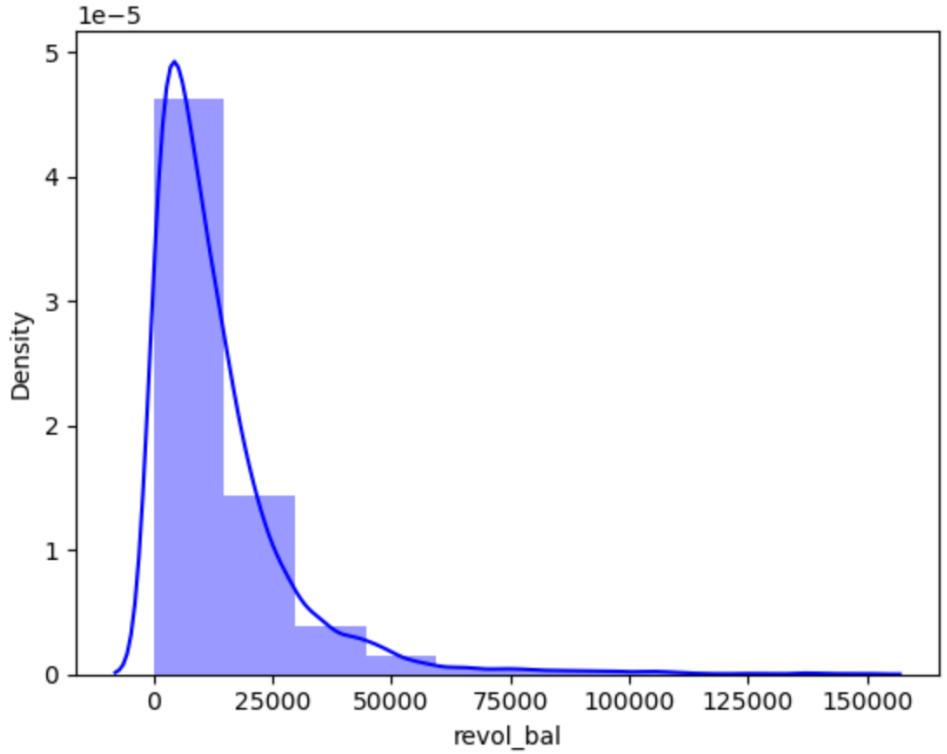
Univariate Analysis –mths_since_last_record

- mths_since_last_record - The number of months since the last public record
- Customers who had between 75-125 months since the last public record had tendency to default on the loan.



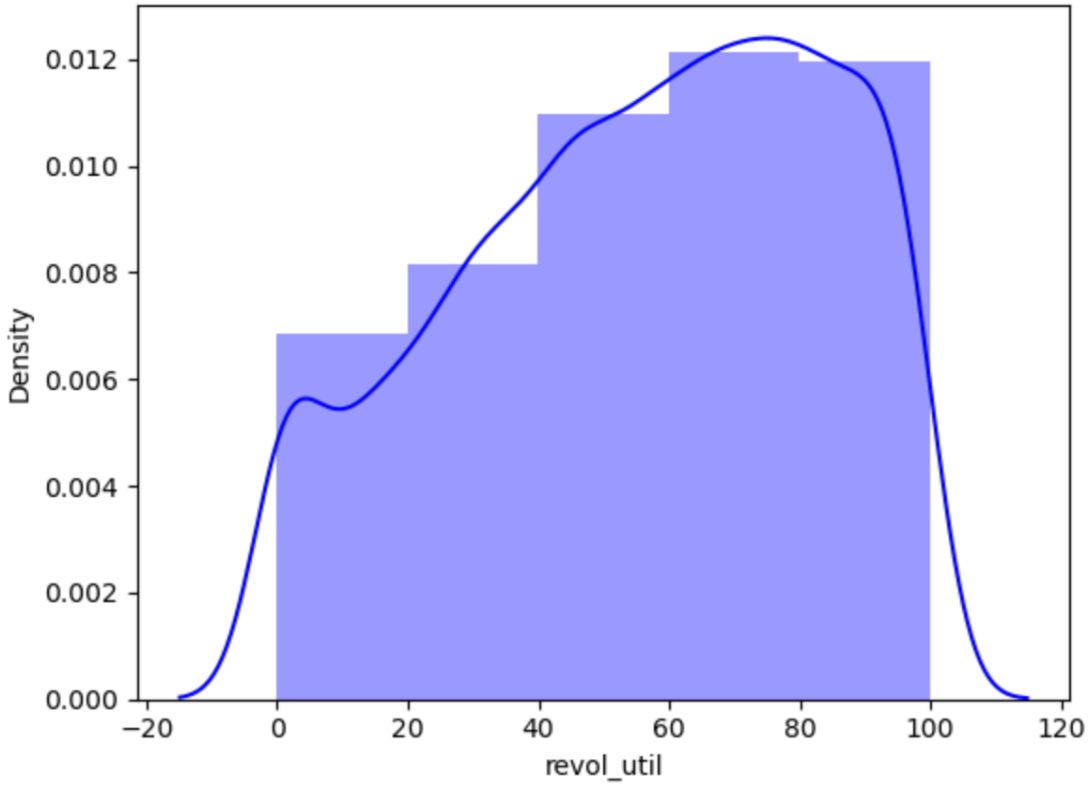
Univariate Analysis –revolving balance

- Revolving balance – Total credit revolving balance
- Customers who had total credit revolving lesser than 25000 tended to default more than those that had credit revolving balance greater than 50,000.



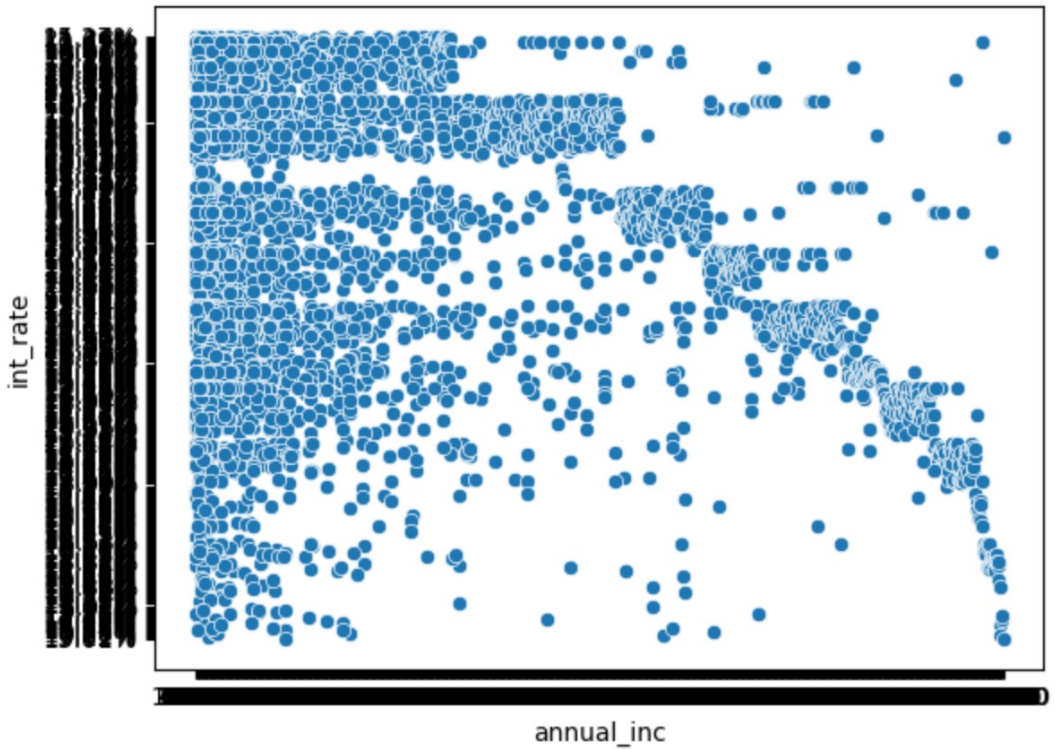
Univariate Analysis –revolving util

- Revolving Utilization - Revolving line utilization rate, or the amount of credit the borrower is using relative to all revolving available credit.
- Customers how have rev_util between 60 to 80% tended to default more.



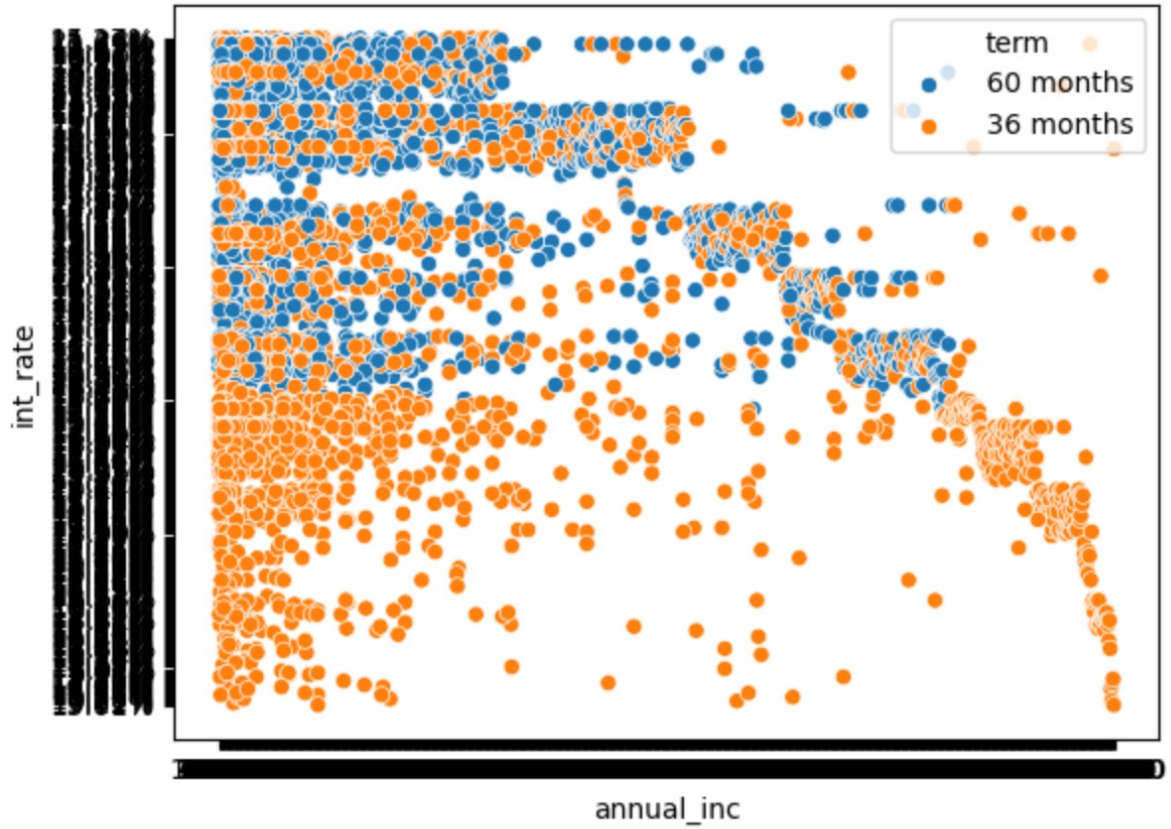
Bivariate Analysis – (Annual Income vs Interest Rate)

- Customers how had lesser annual income and loans with higher interest rate defaulted more.
- Customers who had more annual income have lesser tendency to default.



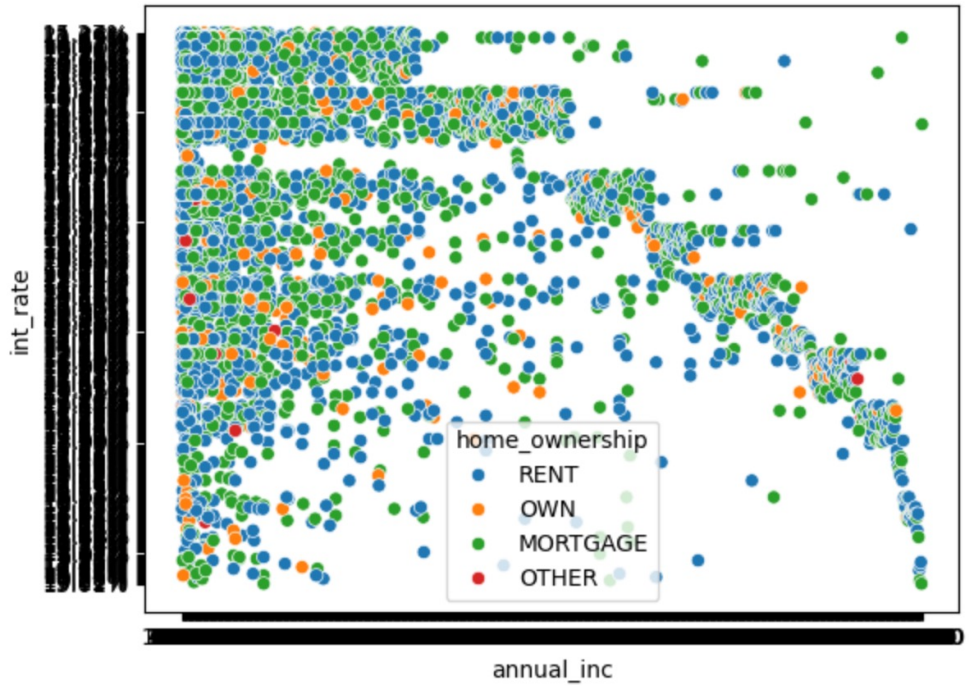
Bivariate Analysis – (Annual Income vs Interest Rate vs Loan Term)

- Customers had lesser annual income and loans with lower terms tended to default more.



Bivariate Analysis – (Annual Income vs Interest Rate vs Home Ownership)

- Customers had lesser annual income and rented / mortgaged their home and had taken loans with higher interest rate defaulted more.




```
loan_status_default_df.isnull().sum()
```

Business Analysis For Providing Loans

Loans should be provided to customers who meet the following criteria:

- Provide Loans to customers how have higher income (Higher interest rate is acceptable)
- Provide Loans with lower interest rate and longer terms (60 months) to customers with lower income
- Provide Loans to customers who own their own home (have already paid off their mortgage)
- Provide Loans to customers who were assigned loan grades G, F and A.
- Provide Loans to customers who have DTI more than 25.
- For customers who have delinquency in the past, consider those customers who had the delinquency more than 20 months in the past.
- Provide Loans to customer who have a credit revolving balance greater than 50,000.
- Provie Loans to customer with revolving util less than 40%.

Thank You