

# Lending Club Case Study

Nipun Vyas (facilitator & contributor)

Prashanth KS (Group member & contributor)

# Problem statement

- Develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.
- 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.
- Use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- Identify the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default

# Our Approach using EDA (Data cleansing and sanitization)

- The data set provided had a large number of columns with empty data. There were around 54 columns in which all entries were NaN/null. These were removed from the data set.
- Many columns such as url, address, desc etc were removed as part of data cleansing as they did not have much relevance to problem.
- Columns (customer behavior variables) which are related to current (loan status) were removed
- The rows where value of column[loan\_status] == "Current" have been removed. This data is of no significance for our analysis
- Certain columns (int\_rate, emp\_length, loan\_month) had % sign, + sign, < sign with numeric values these were sanitized by removing unwanted signs and preserving the float values or interpreting the values in certain way
  1. Int\_rate column had % sign and values were like 20%, 6.6% etc and these were sanitized as 20, 6.6
  2. Emp\_length column had values like <1 year, 10+ years, 6years etc. These were sanitized by removing year/years and <1 was treated as 0 and 10+ was treated as 10. Any other value between 1-10 was treated as normal float value
  3. loan\_month was treated as date

# Univariate Analysis

- Continuous variables for which univariate analysis was done are as follows
  1. loan\_amnt
  2. annual\_inc
  3. int\_rate
  4. dti
  5. installment
  6. funded\_amnt

# Univariate Analysis (Continuous variables)

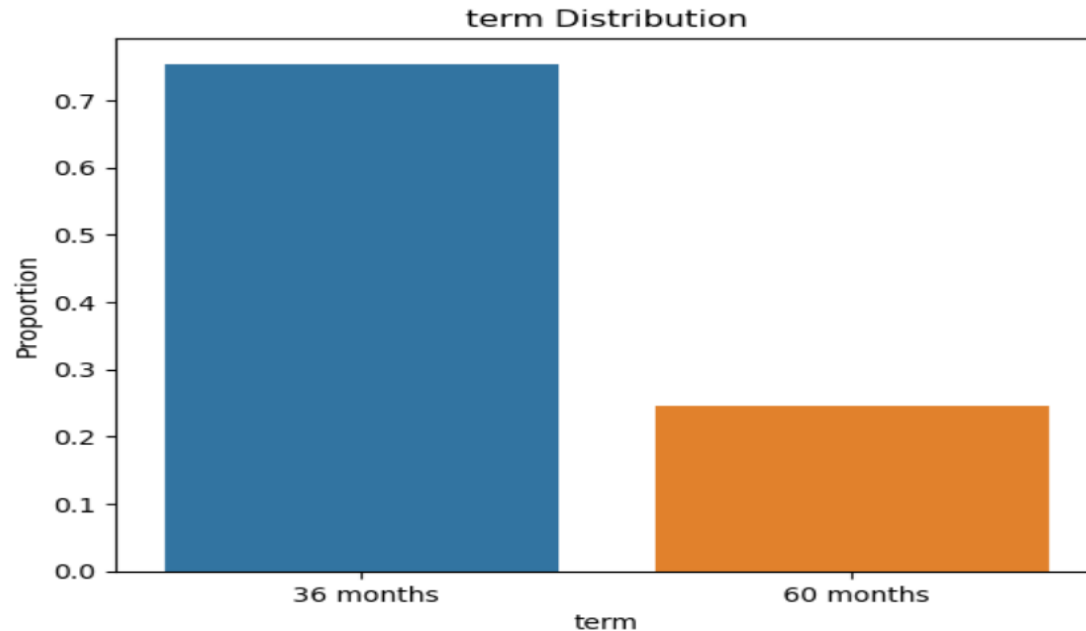
- **Observations on continuous variables :**

1. There were outliers in 95th to 99th quantile where loan amount was between 25000-35000, they were removed
2. 25th to 75th quantile for loan\_amnt exist between 5000-15000. Large number of loans of amount 10000 exist.
3. 99% people have annual income less than 234144.
4. Mostly int\_rate lies between 9 to 14%
5. Some loans have been taken at very high interest rates 22.5% to 25%
6. 99% people have annual income less than 234144.
7. Installment mostly lies between 150 to 400 with max being around 715

# Univariate Analysis (Categorical variables)

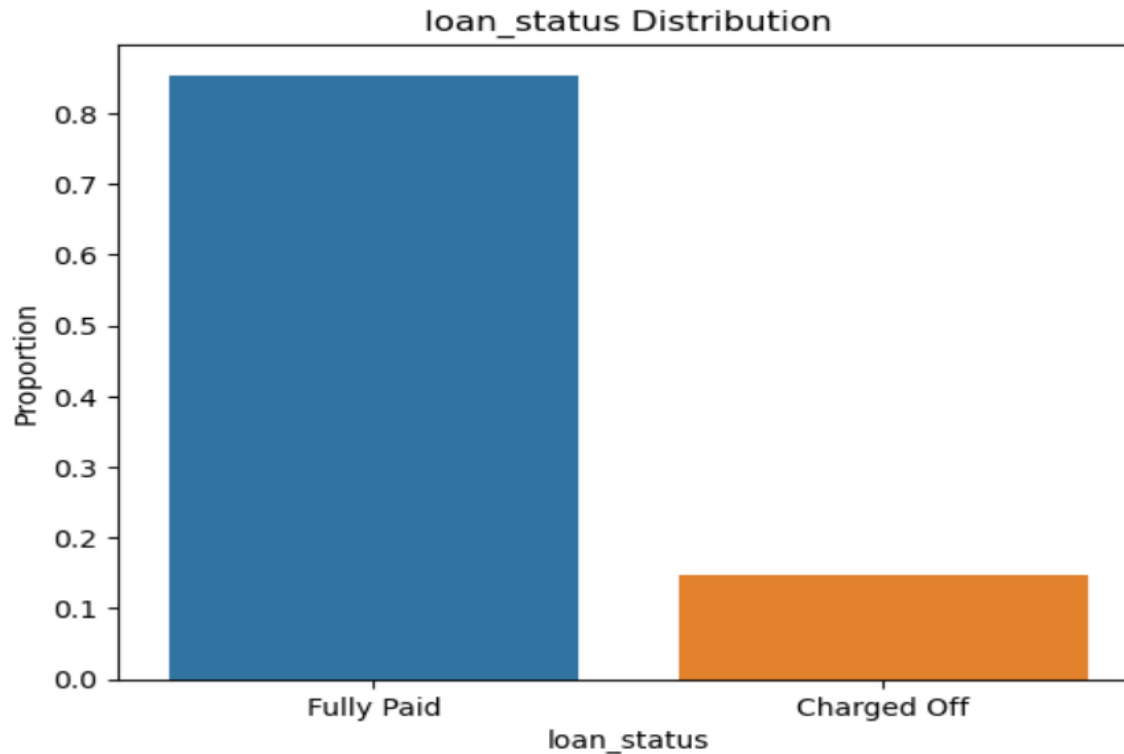
- **Observations on categorical variables :**

1. 36months term loans are almost 80% of total loans



# Univariate Analysis (Categorical variables)

- **Observations on categorical variables :**
  1. 85% of total loans are non-default



# Univariate Analysis (Categorical variables)

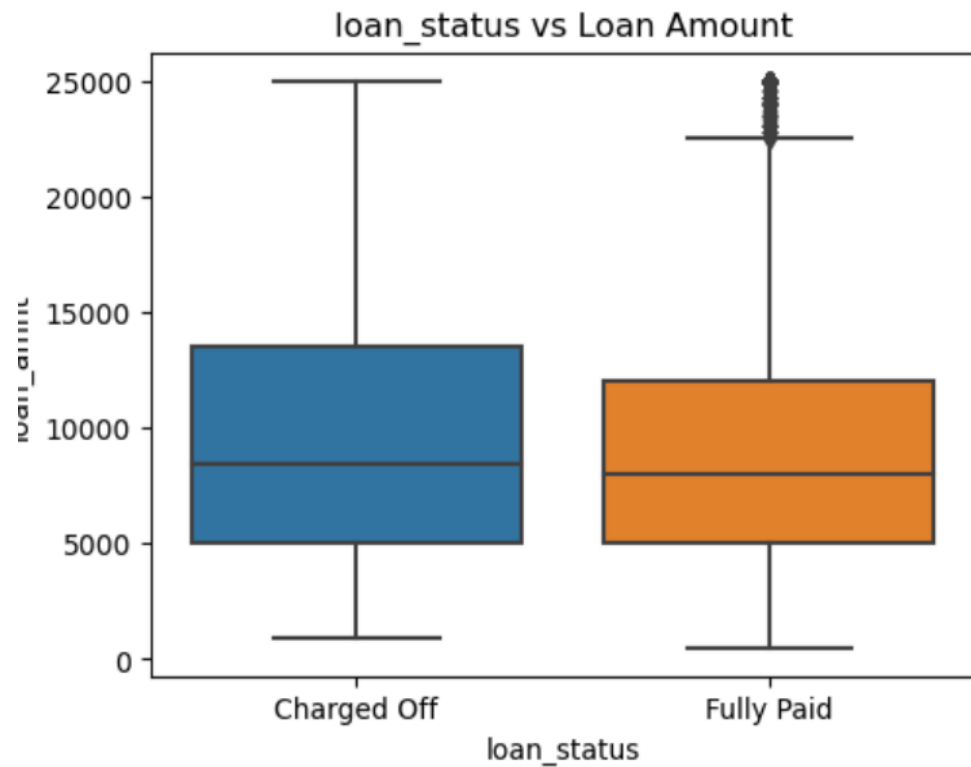
- **Observations on categorical variables :**

1. For approx. 50% loans the income of borrower was not verified by LC.
2. For approx. 25% loans the source of income was verified by LC.
3. For approx. 25% loans the income of borrower was verified by LC.
4. Very few people actually own a house
5. Most of the people are living in rented or mortgaged house
6. Highest percentage of employees who took loan have employee service of 10 or more years
7. Second highest percentage of employees who took loan have service tenure of less than a year
8. More than 50% borrowers fall under loan grade A,B
9. Around 50% of purpose given by borrowers is debt consolidation



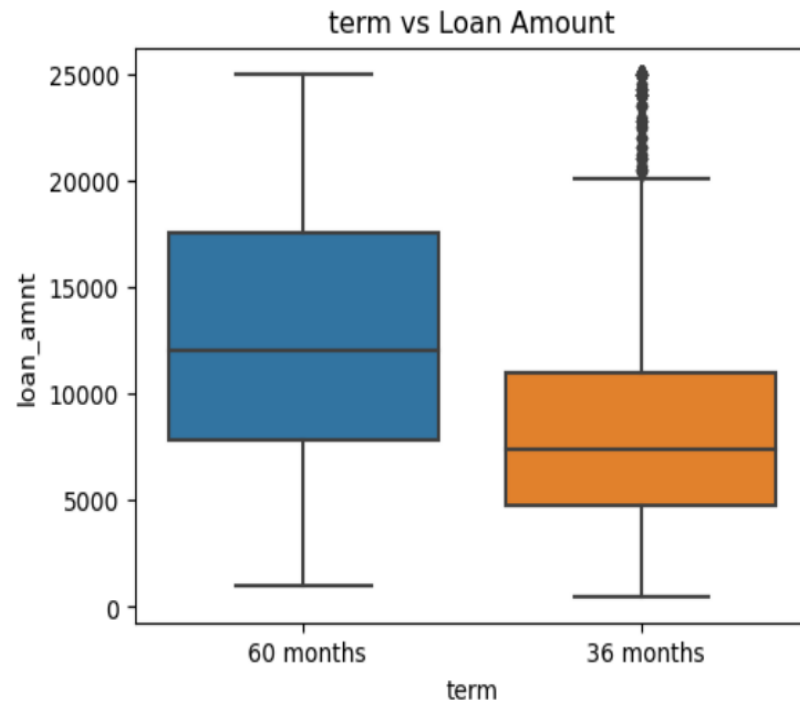
# Segmented Univariate Analysis

- **Observations on categorical variables :**
  - Default loans are of higher loan amounts



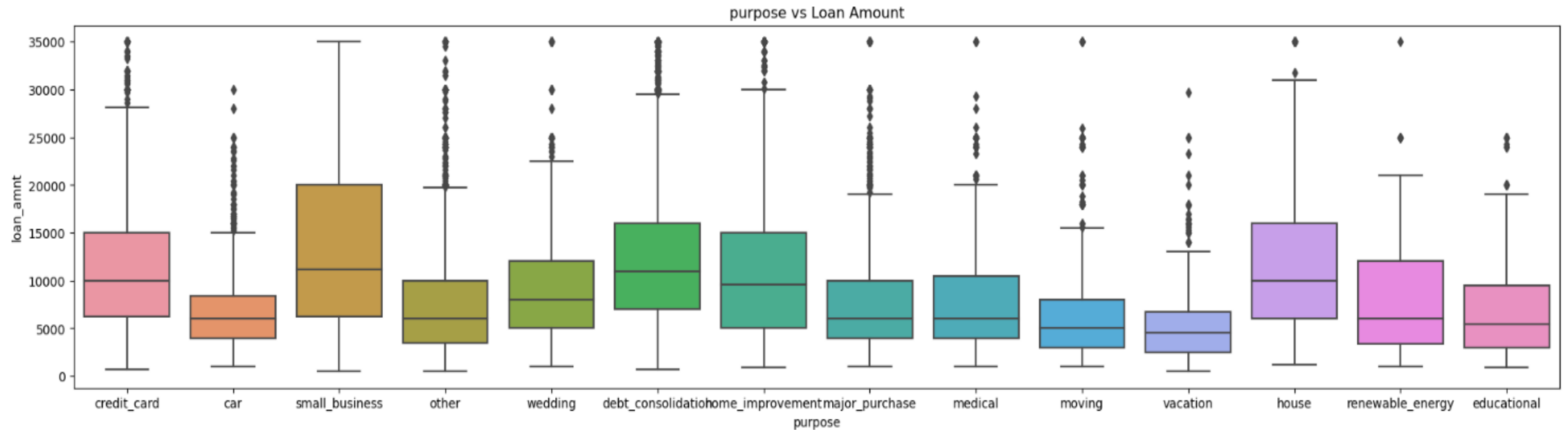
# Segmented Univariate Analysis (Contd)

- **Observations on categorical variables :**
  - Loan\_amnt are higher for 60months term



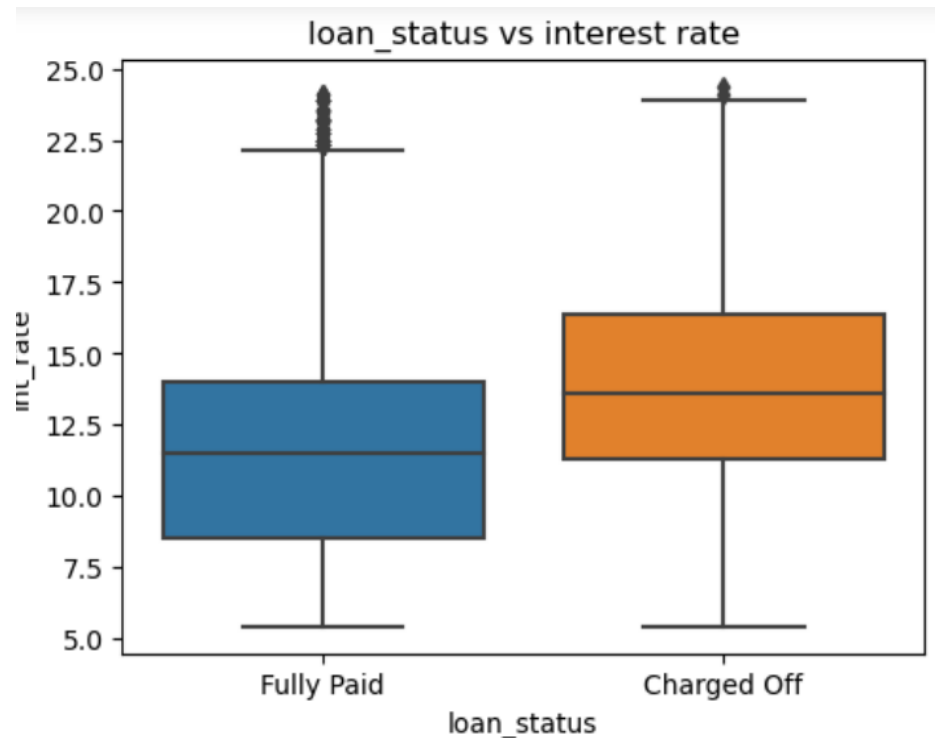
# Segmented Univariate Analysis (Contd)

- **Observations on categorical variables :**
  - Loan\_amnt is higher for small\_business purpose compared to other purposes



# Segmented Univariate Analysis (Contd)

- **Observations on categorical variables :**
  - Higher int\_rate results in more defaults



# Segmented Univariate Analysis (Contd)

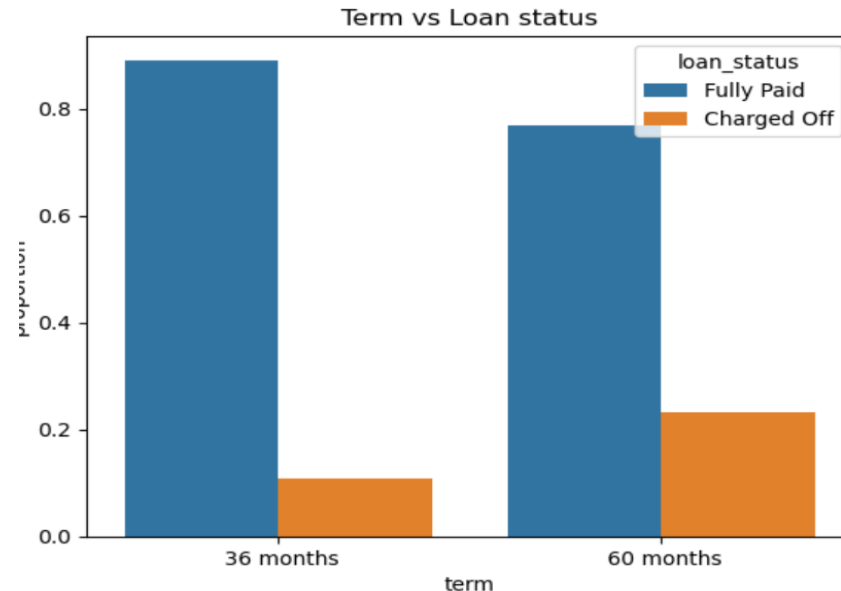
- **Observations on categorical variables :**

1. int\_rate for borrowers having own house or mortgage are less. not\_verified borrowers are getting less int\_rate on the loan
2. Small business and house loans are given at higher int\_rate
3. Annual income for fully paid and charged off loans is very slightly different. For charged off loans it is slightly less Annual income gradually increases with increase in emp\_length
4. Borrowers with higher income are taking loans which have lower grades F and G (A (high) -> G (low))
5. Borrowers with verified income have higher annual\_inc. Mortgage home\_ownership has higher annual income
6. Borrowers with high dti has bit more probability to default No particular observation on emp\_length

# Bivariate Analysis

- **Observations:**

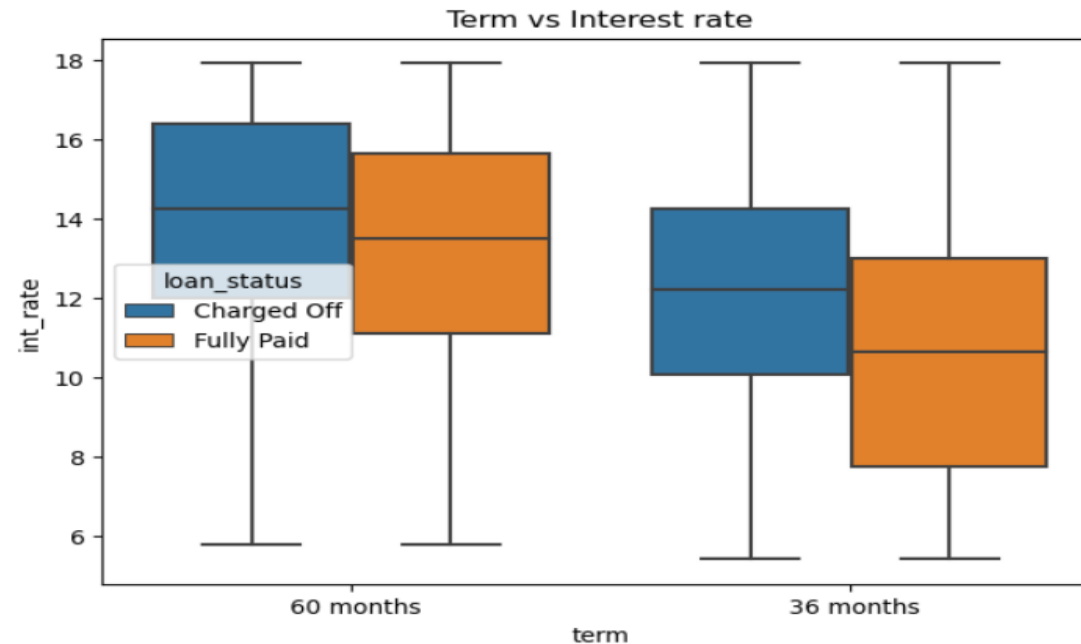
- 1 . 60months tenure loans are defaulted twice as much as 36 months
- 2 . 36months tenure loans are proportionately more in terms of Fully paid.  
Hence 36months loans are less likely to be defaulted



# Bivariate Analysis

- **Observations:**

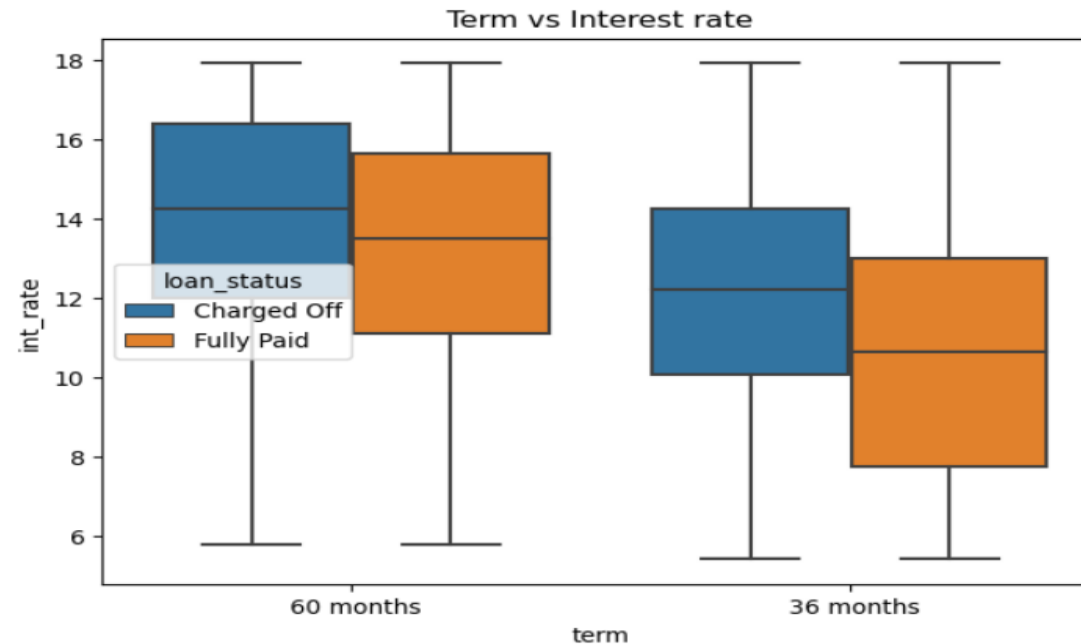
- With higher interest rate default loans are higher



# Bivariate Analysis

- **Observations:**

- With higher interest rate default loans are higher

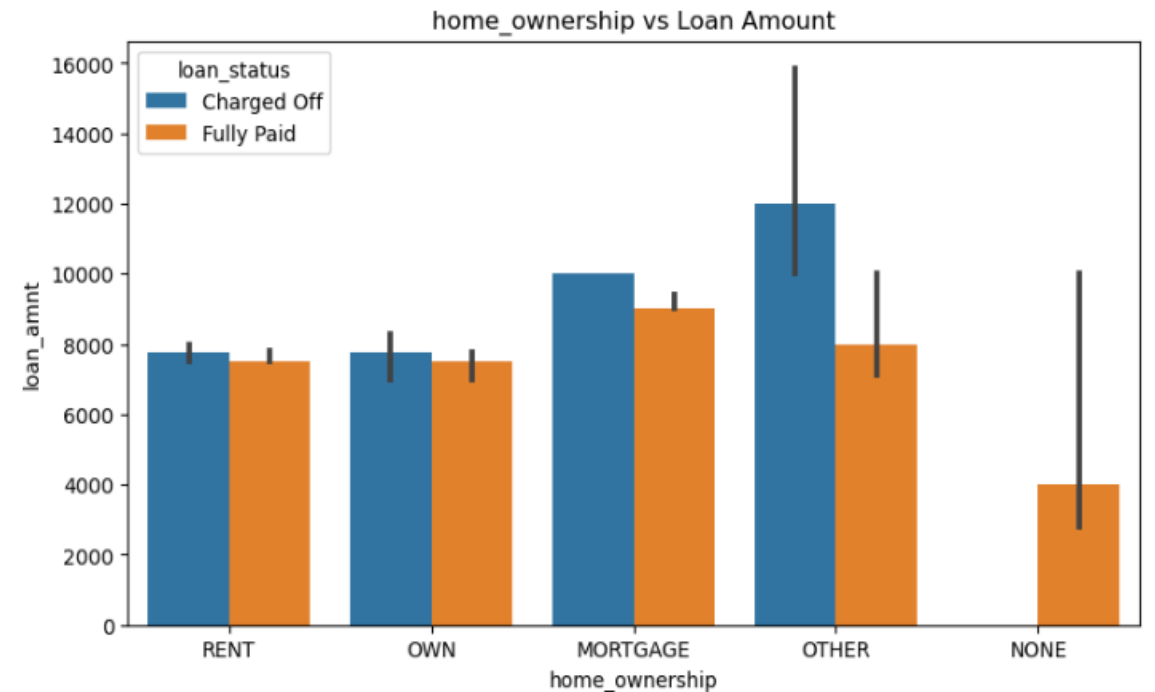
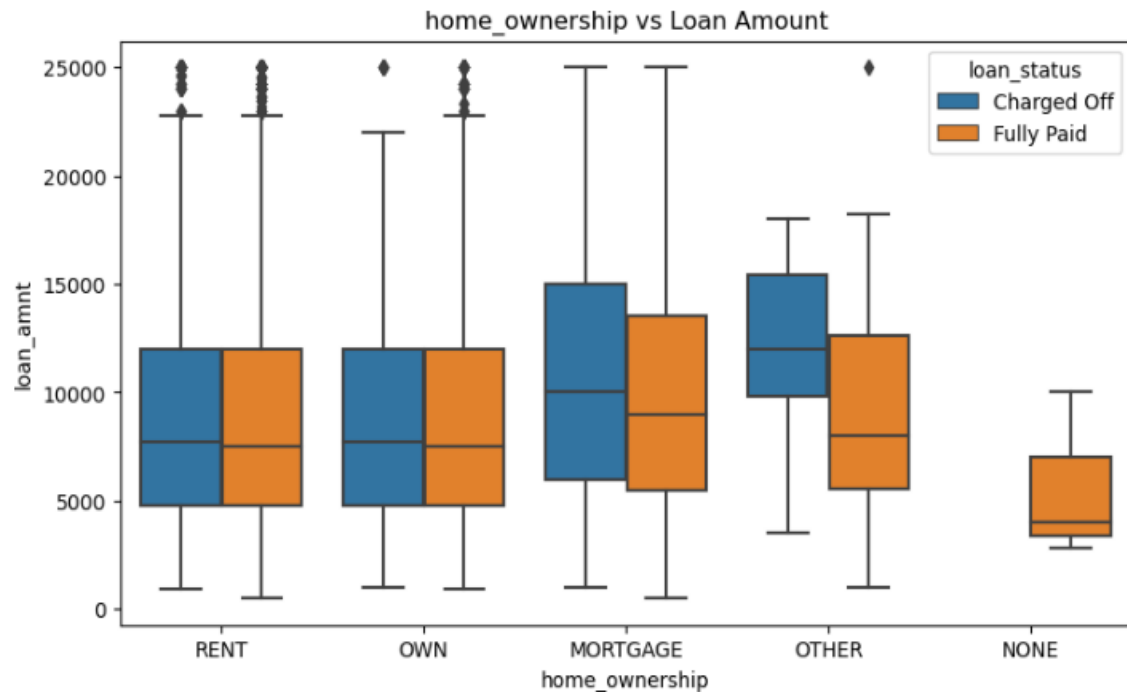




# Bivariate Analysis

- **Observations:**

- With increase in loan amount charge off is more especially if house ownership is mortgage or other



# Bivariate Analysis (Contd)

- **Observations :**

1. At higher values of dti charged off is more especially if home ownership is other
2. In income verified when loan amount is more there is more default  
In case of income verified more loans are given and with higher amount and still defaults are high
3. Higher the dti more is charged off rate
4. The default loans increases gradually as the grade goes from A->G
5. House loans are most defaulted when int\_rate is high
6. Credit\_card and debt\_consolidation purpose default loans have high dtis
7. Small business, credit\_card and debt\_consolidation purpose has more defaults when loan amount is high.
8. Higher the dti more is charged off rate

# Recommendations

- When providing loans dti (debt to income ratio) is towards higher side than such loans should be rejected or be executed with caution
- High amount loan should not be give for following purpose “**Small business, credit\_card and debt\_consolidation**”. They generally are defaulted
- House loans with high interest rates must be watched out for default
- Borrowers with high dti should be avoided
- Verification of income should be conducted using better processes and should be more strict

Thank you