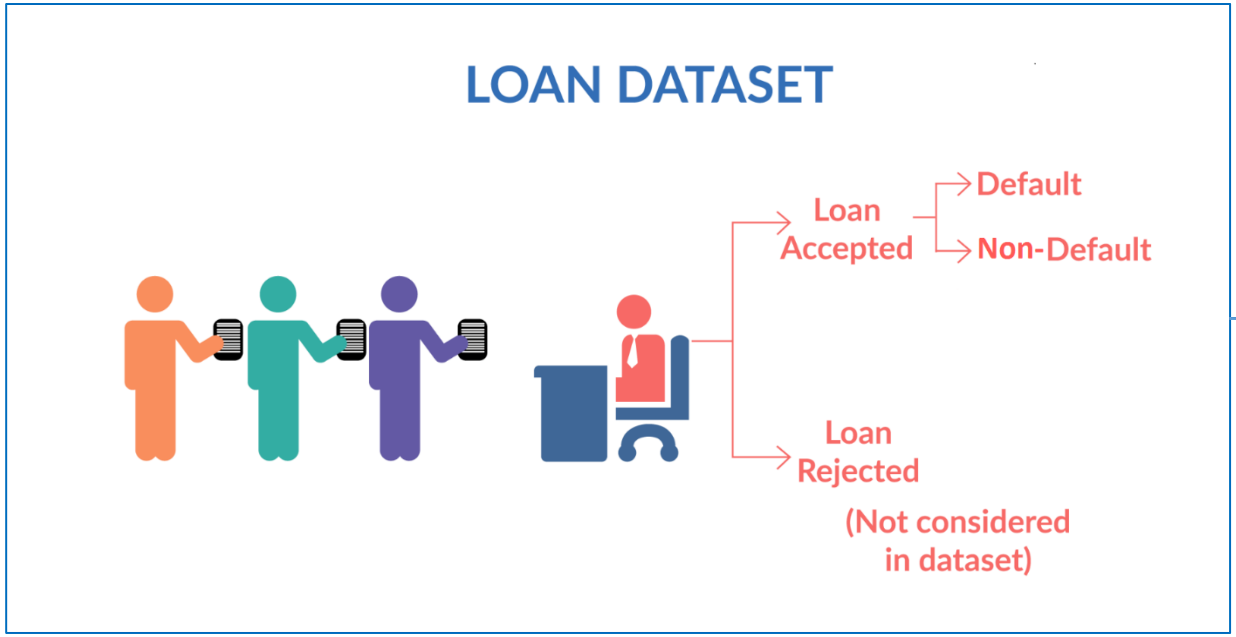
LendingClub

Case Study

Objective:



Perform EDA to identify the consumer attributes and loan attributes influence the tendency of default

Identify patterns which indicate if a person is likely to default

Help the organization to identify risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss

Process:

Data understanding

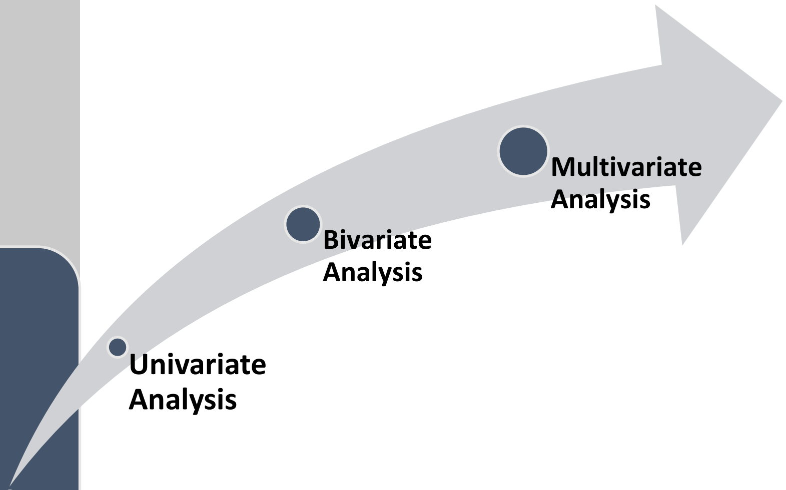
- Understanding the meaning of the variables

Data Cleaning and Manipulation

- Missing value imputation, outlier treatment

Exploratory Data analysis

- Missing value imputation, outlier treatment



Recommendations

- Observations and inferences from the analysis

Approach:

- **Understanding of Loan dataset :**
  - The given dataset comprises of 111 columns and 39717 rows.
  - Every detail of the person's loan history, break down of the balance amount and remaining amount, credit line details, delinquencies, public records, personal information etc. are covered.
  - Using this dataset we will be judging if a person is eligible for an extra loan or not.

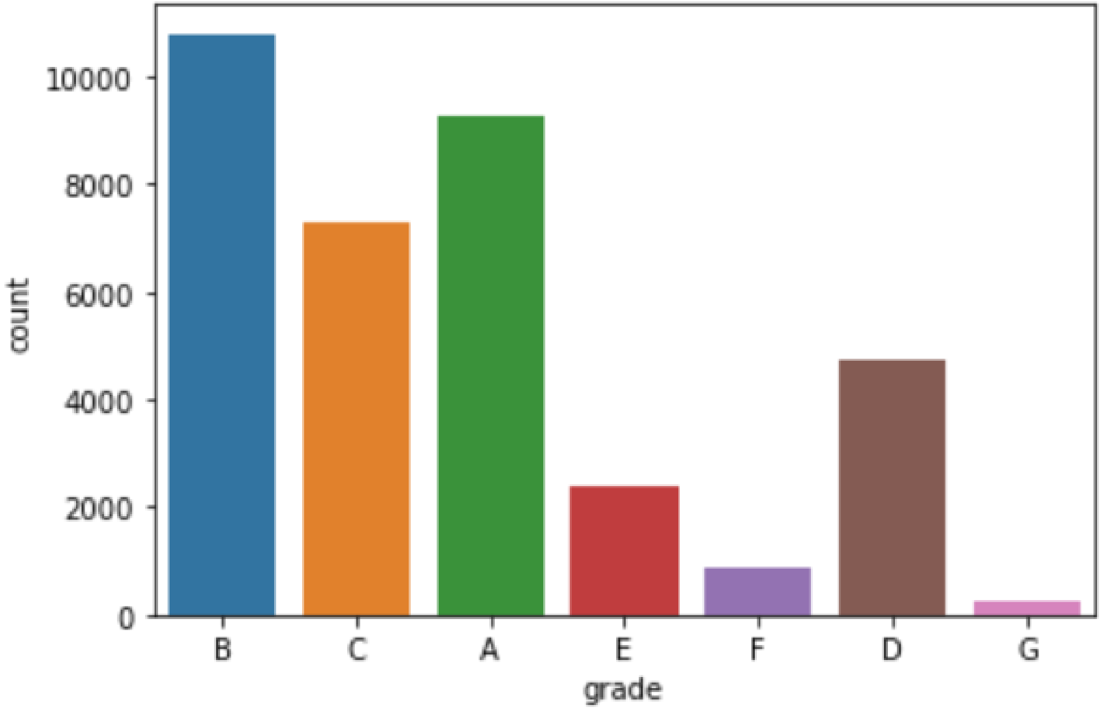
- **Data Cleaning and Manipulation :**
  - Removed the NA values for all the variables and derived multiple columns from the given data
  - Segregated date column into day, month and year
  - Created total amount column from term and installment
  - Created income range, interest range, dti range , total amount range , etc

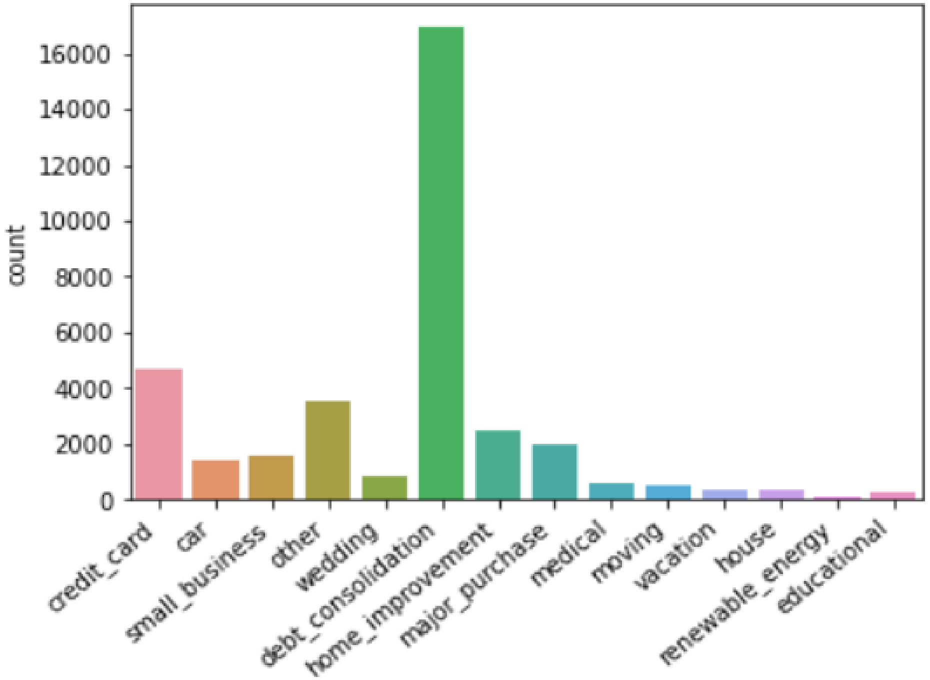
- **Univariate Analysis:**
  - Generated histograms, frequency plot to get information about the data.
  - Box plot for outliers detection

- **Bivariate Analysis:**
  - Generated bar plot to get behavior of a variable with charging off.

Univariate Analysis

Analysis on Grade and Purpose

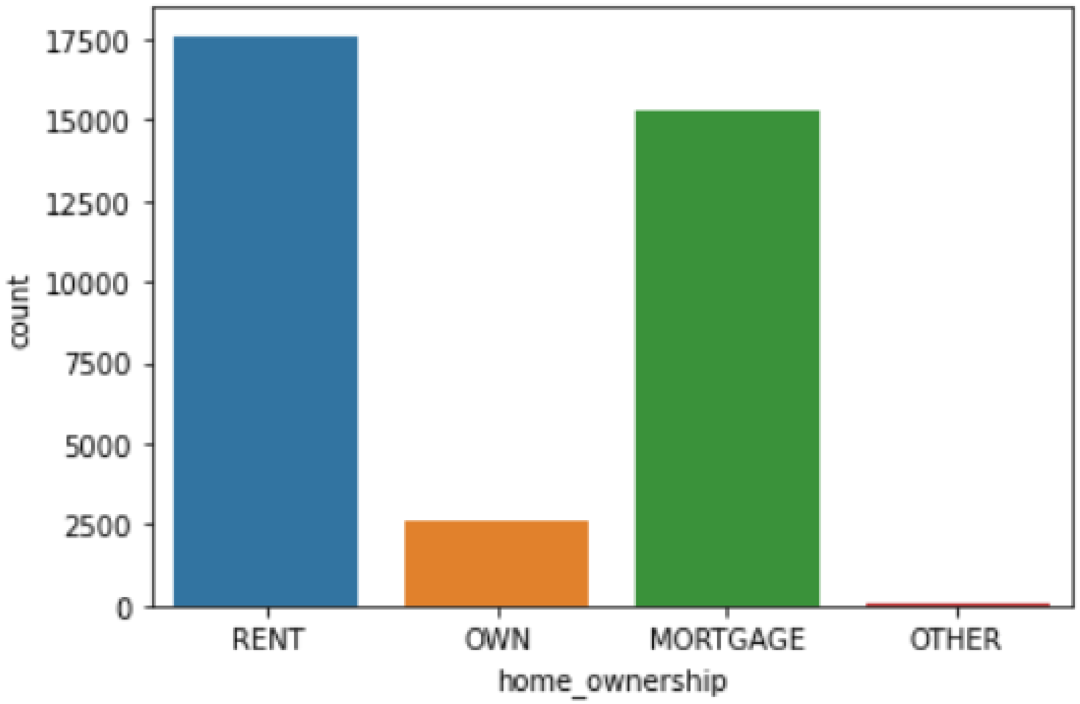


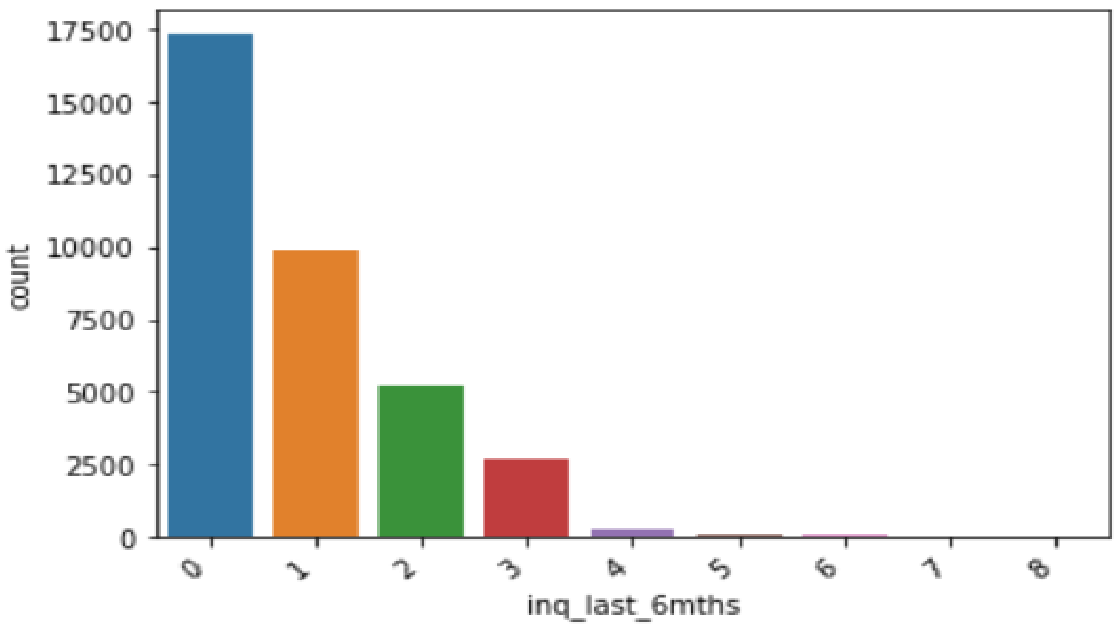


Observation :

- There are many customers with loan grade as B and A
- High number of loan is applied for debt consolidation

Charged off Loans Vs Interest Rates/Loan Grade



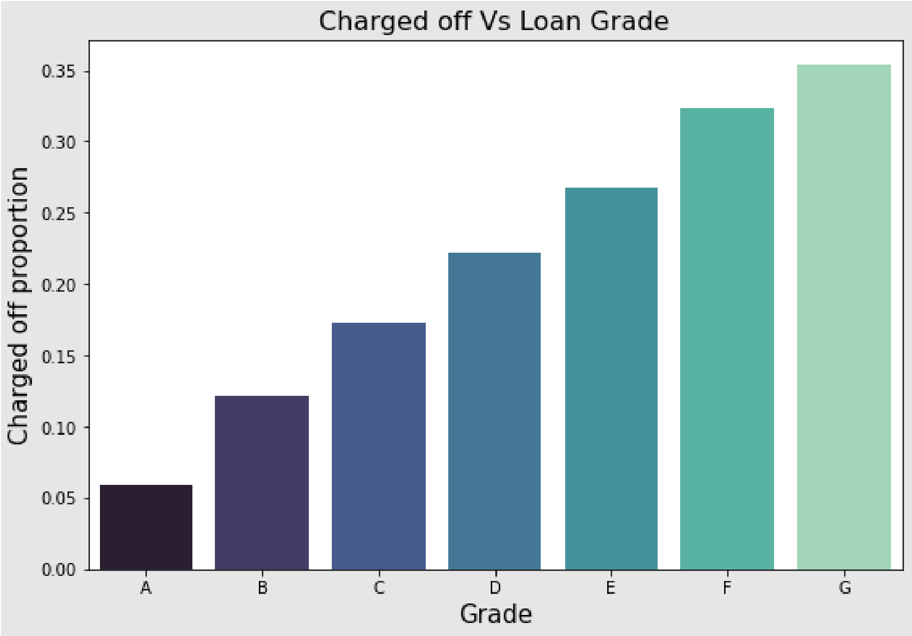


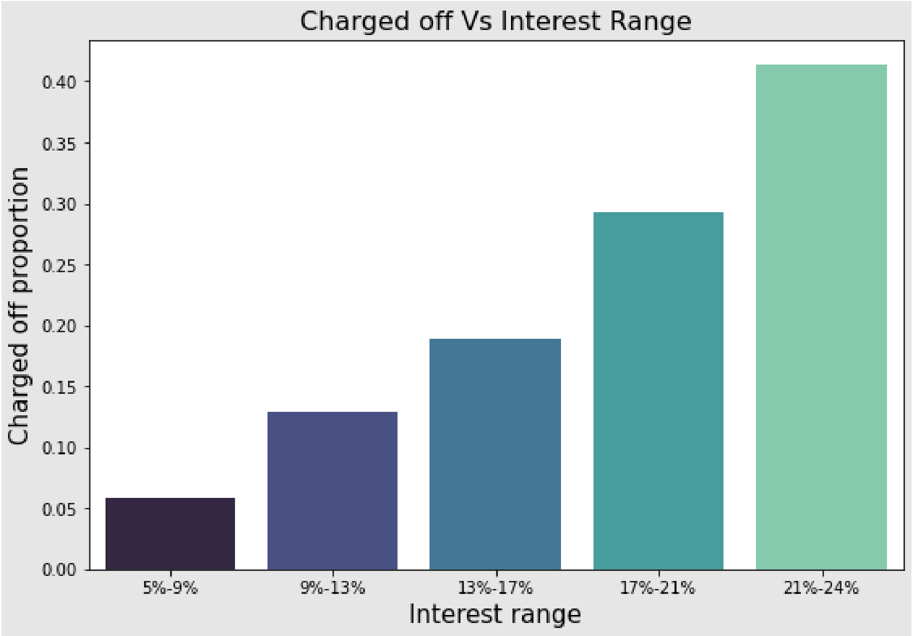
Observation :

- Many people with rented home and mortgage home applied for the loan
- There are many number of people with zero inquiry.

Bivariate Analysis

Charged off Loans Vs Interest Rates/Loan Grade

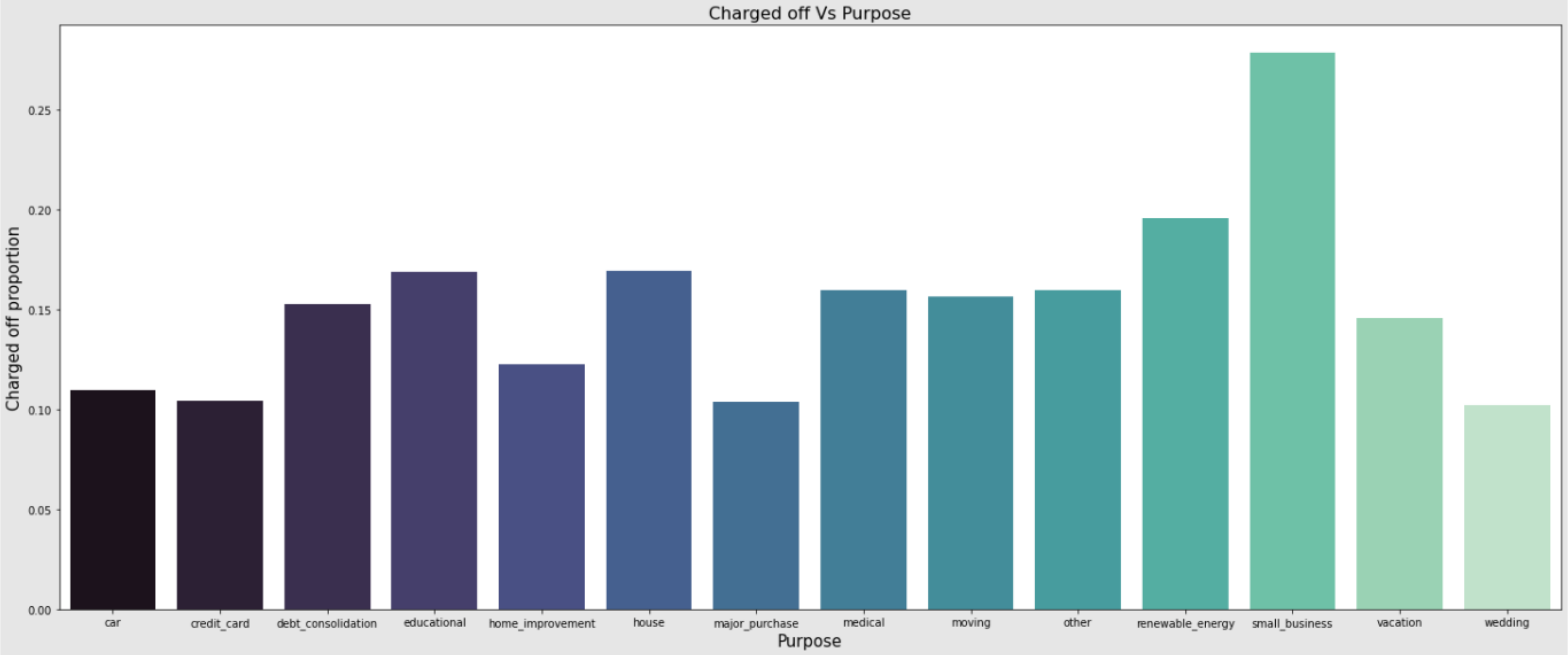




Observation :

- With higher interest rate increases the risk of charge off
- Charge off proportion increases with along the grades from A to G

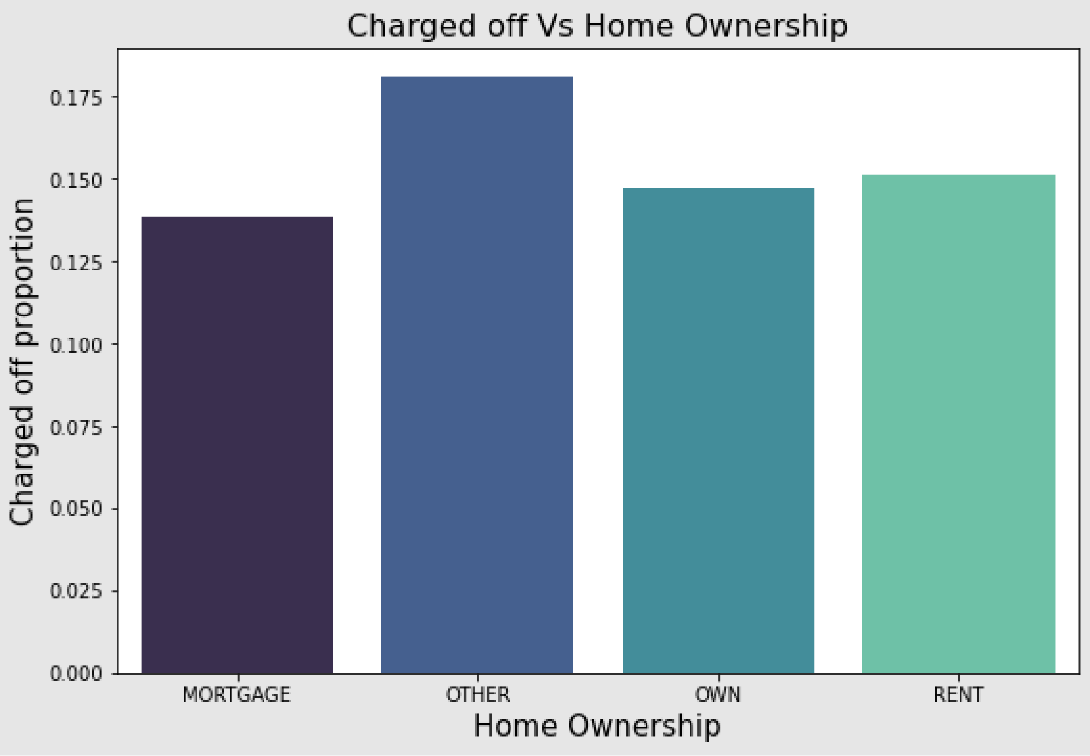
Charged off Loans Vs Purpose

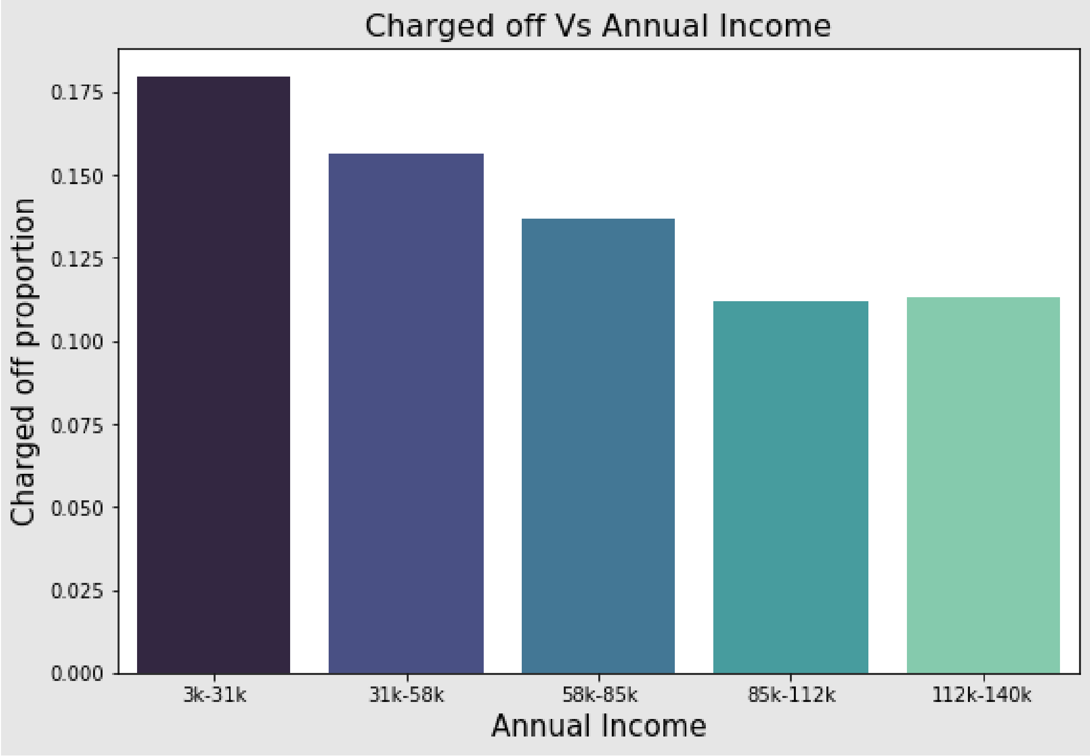


Observation :

- Charge off proportion is very high for small business loans followed by renewable energy purposes

Charged off Loans Vs Home Ownership/Annual Income

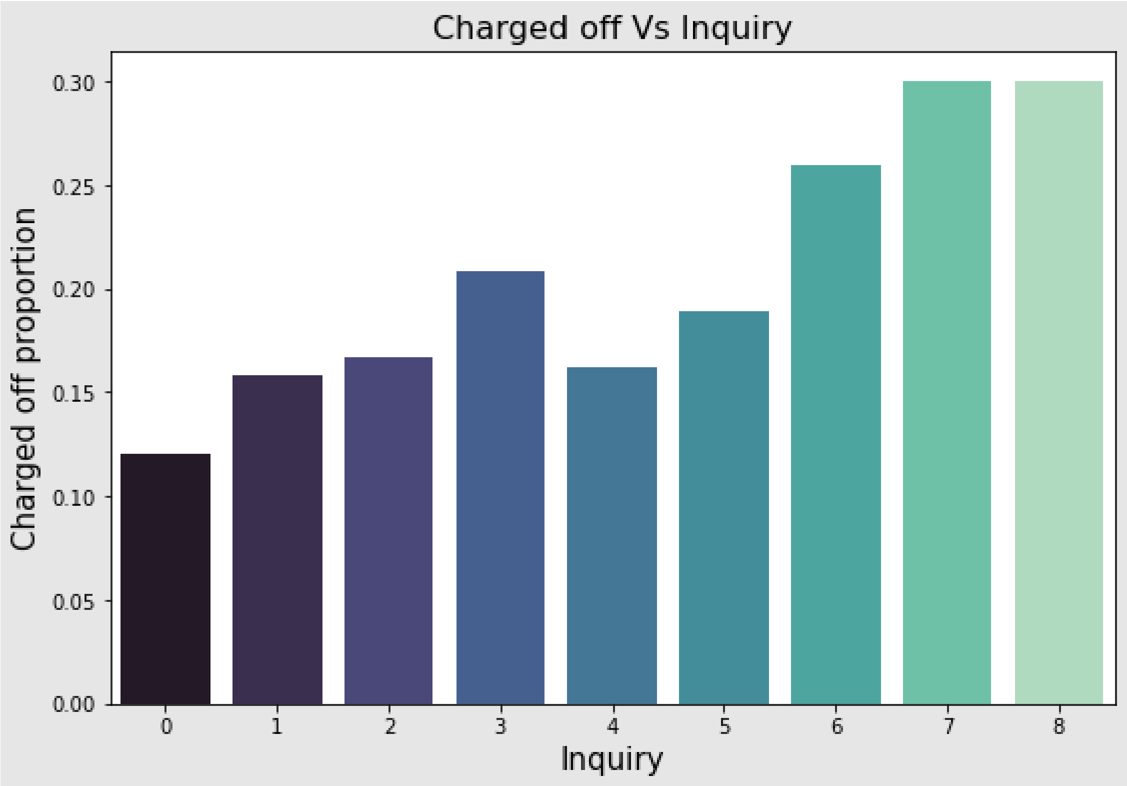


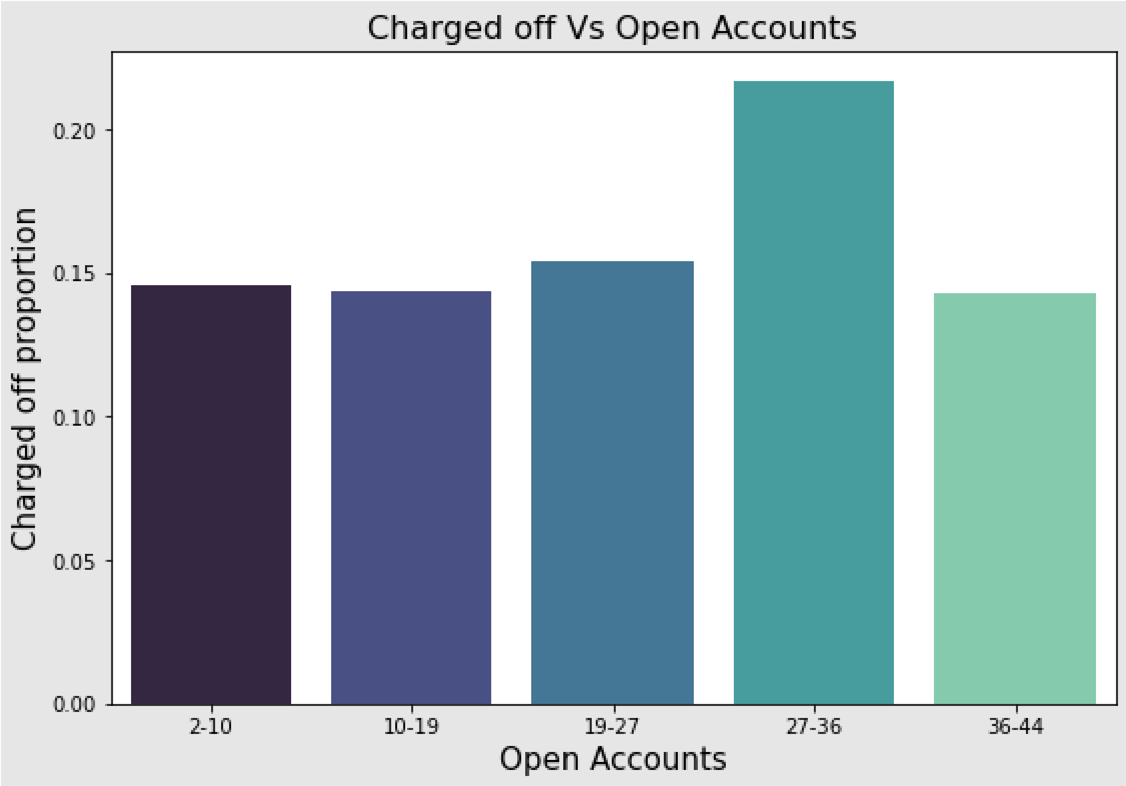


Observation :

- Charge off proportion is high when the home ownership is others
- Loans given to low income group have higher chances of defaulting.

Charged off Loans Vs Inquiry/Open Accounts





Observation :

- Charge off proportion is high when the number of inquiry is high.
- Charge off proportion is high when the customer has open accounts in the range of 27-36.

Recommendations

Insights :

- Loans with **higher interest** rate are having more defaulters. Background verification has to be thoroughly as a deciding factor before lending the loan.
- **High Debt** to Income ratio leads to Defaulting, this ratio must be verified for high risky applicants.
- **Lower graded** loans has more defaulters which shows the grading system is one of the important deciding factor for lending loans and lower graded loans should be handled with more verification.
- **Small businesses** loans has more defaulters, business loans should be validated through with the business plan and profitability
- **Low income** group loans have higher tendency of defaulting. The income sources needs to be verified thoroughly before processing such loans and it is recommended to keep the loan amount lower for such customers.