Lending Club Risk Analysis

Name: Snehal Gunde and Vikas Rane

Overview

Lending is a Consumer Finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company must decide loan approval based on the applicant's profile.

Objective

- To analyze what are the parameters to be considered in the Loan Approval / Rejection decisions.

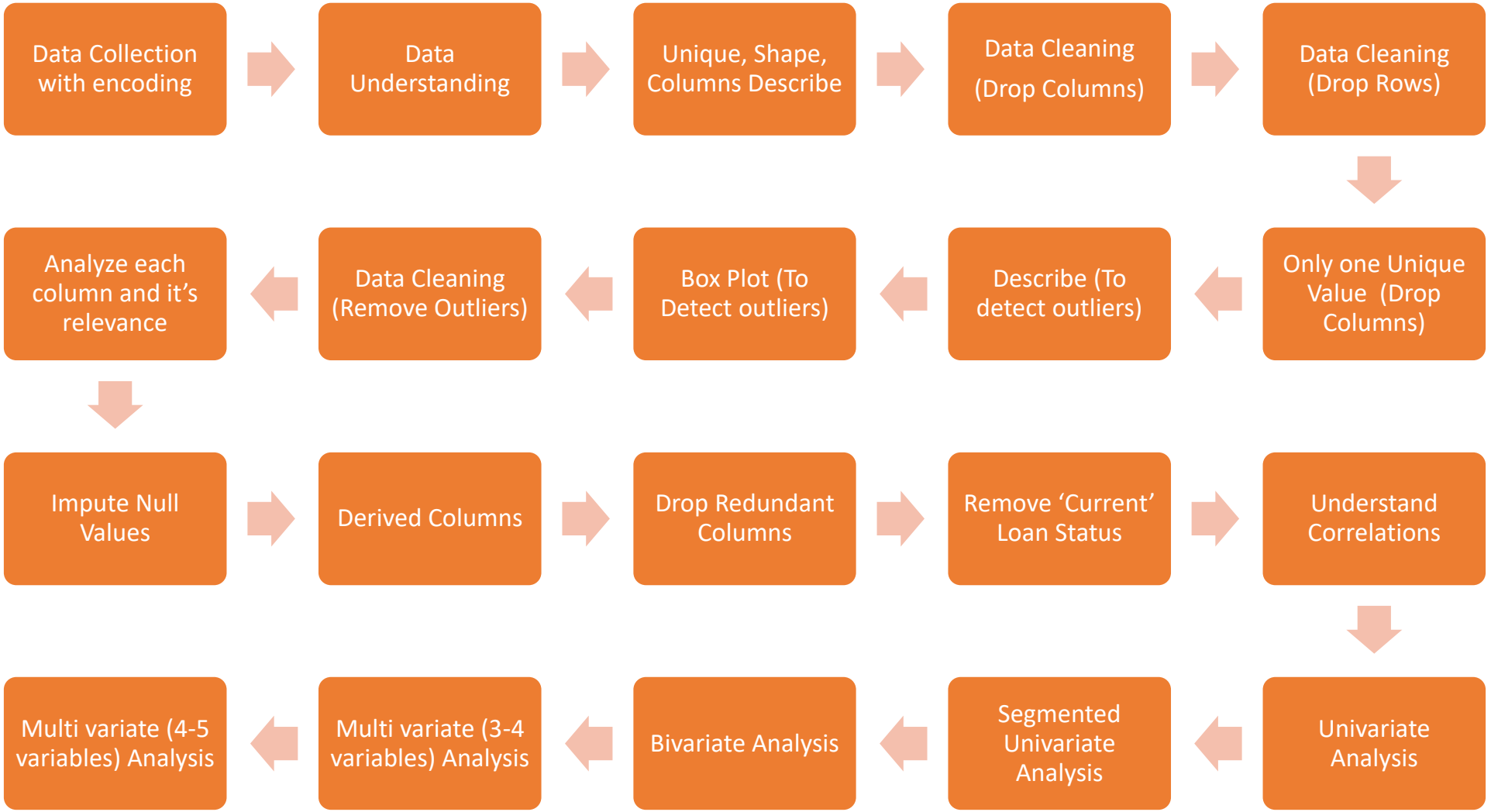
Strategy

- To take precautionary measures while processing loan applications to avoid Charge offs.

Risks involved

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Methodology



Analysis : Data Collection & Understanding

- Loan.csv file is with ascii encoding
- For reading use encoding standard as "ISO-8859-1"
- Loan_df provides 39717 observations and 111 Attributes
- Use of the Following functions provides better understanding of the Datasets
  - Info
  - Shape
  - Describe
  - Nunique
  - Sample / Head

- Here are the observations
  - Many Attributes contains 100% Null Values
  - Some of the Attributes contains only one Unique Value
  - It's safe to drop the above-mentioned Columns
- Loan_status is identified as a Target variable against which the entire analysis would be performed

Analysis : Data cleaning

- After dropping the Redundant columns, below is the output.
  - ❑ 39717 Observations
  - ❑ 45 Attributes
- There are still Null values in 8 Columns and outliers
- Since most of the Null Values are in Categorial Variables, instead of dropping them we may impute with meaningful Representative values e.g. 'Not Disclosed'
- There are columns with special characters e.g., % such special characters were removed for further analysis.

- Rather than addressing Outliers in one shot we addressed one at a time on need basis.
- One of the Outliers E.g., If Annual Income is more than $million then some one spending time on loan process for a $10k to $25k is very rare and such outliers and hence can be easily removed.
- Converted Attributes to standardized format Converted dates from Object format to Date Time Format.

Analysis : Univariate

- Analyzed one Attribute at a time
- For Further Processing analyzed variables into Following Categories
  - ❑ Ordered Categorical
  - ❑ Unordered categorical variables
  - ❑ Quantitative / numeric variables
- Factors need further Analysis
  - ❑ Customer Behavior
  - ❑ Loan Characteristics
  - ❑ Applicant Demographic

- Customer Behavior
  - ❑ Non-Disclosure of Emp Length and Title,
  - ❑ No Income Verification,
  - ❑ Public Bankruptcy,
  - ❑ High Credit revolving Balance
- Applicant Demographic
  - ❑ Low Income, delinq_2yrs,
  - ❑ Nevada State, Loan Purpose
  - ❑ Non-Rental Non-Own Home Ownership,
- Loan Characteristics
  - ❑ Interest Rate, Term,
  - ❑ Installments,
  - ❑ High DTI,
  - ❑ Age of Credit history,
  - ❑ Funded Amount < Loan Amount

Analysis : Segmented Univariate

- loan_status was used for segmentation and Analysis was performed against ‘charged-off’ and ‘fully paid’. Few examples are listed below.
- Loan_status vs loan_amount : Loan amount from 30-35 K have a greater number of charged-off. As the amount of loan increasing the percent of Charged Off is increasing.
- Addr_state v/s loan_Status Percentage of Charge offs are higher in NE, NIU, AE
- Revolving Balance v/s loan_Status .With increase in revolving balance utilization charge off rate is increasing

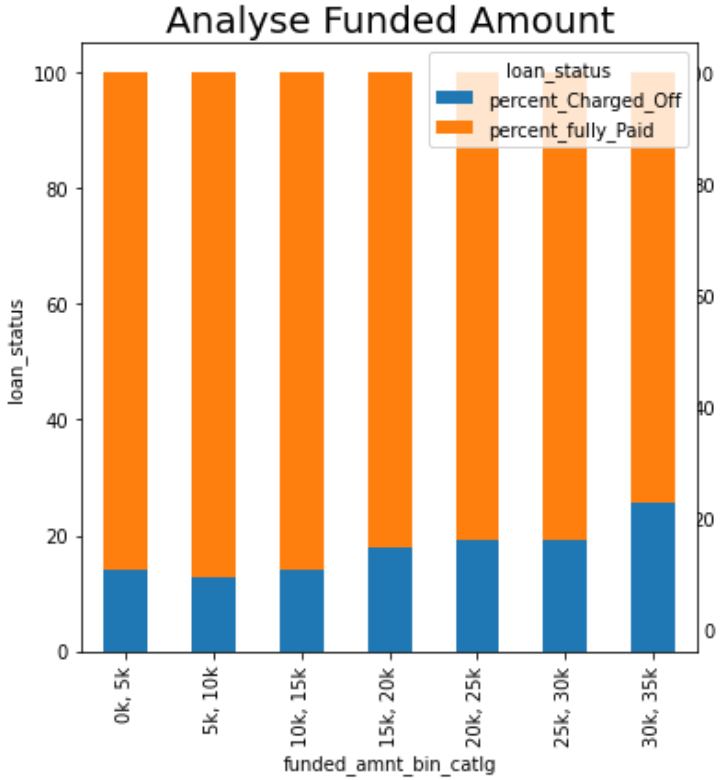
- Tenure of the loan (36 or 60 Months) vs loan_status: Higher Tenure results into higher percentage of Charge Off.
- Home_ownership vs loan_status: The home ownership with 'other' are charging off the loans at higher rate.
- Purpose vs loan_status : Small Business and Educational loans are highly charged off
- Funded Amount v/s loan_Status Higher the Funded Amount of the Loan higher the percentage of Charged off.

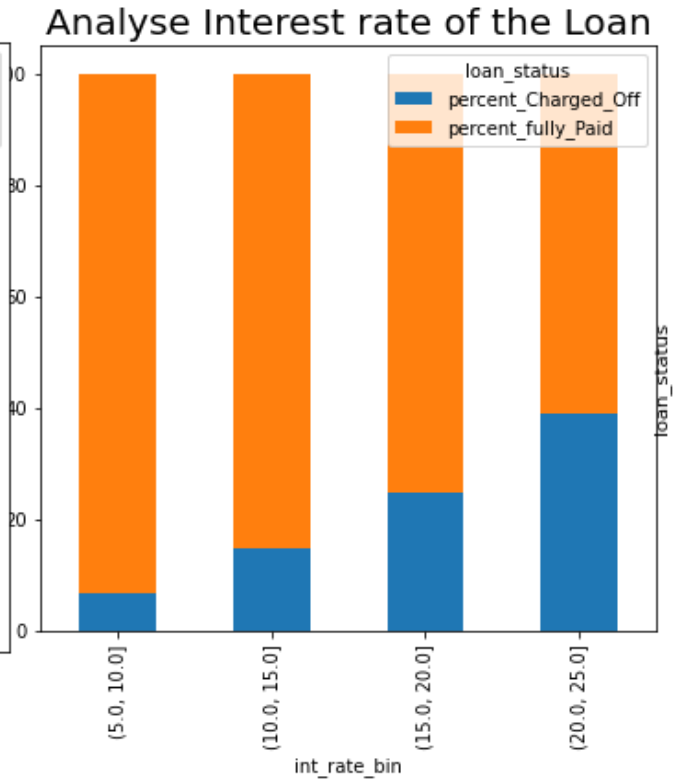
Analysis : Bivariate

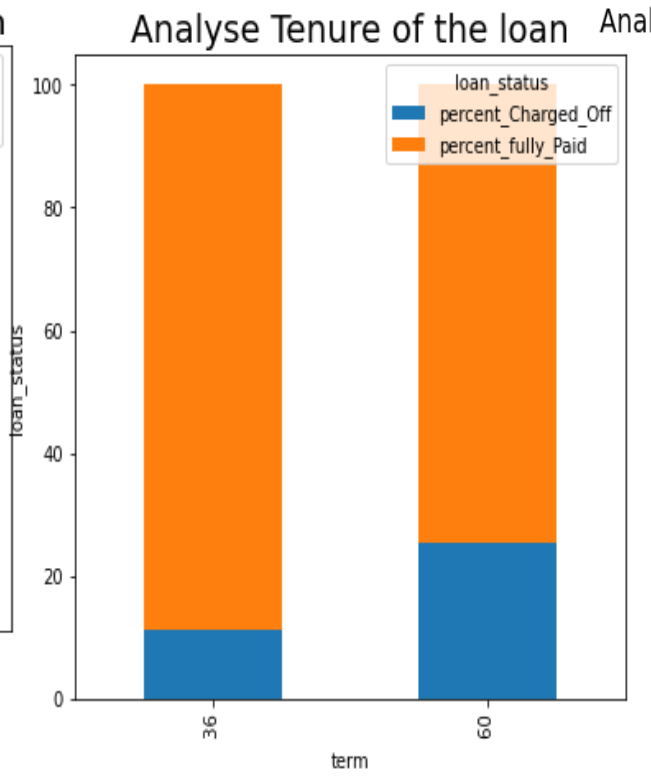
- In Bivariate / Multivariate Analysis, we have considered the impact of two or more variables e.g., Bivariate (Interest rate and Annual income against charge off %) and Multivariate (Term, Purpose, Home ownership and Income Verification against Charge off %).
- To achieve this Impact Analysis, we have formed Pivot tables for 2 data sets (1. Charge off 2. Total) and on the Variables on which Bi / Multi Variate Analysis to be performed.
- we have taken percentage-based approach so that it's easier for comparison

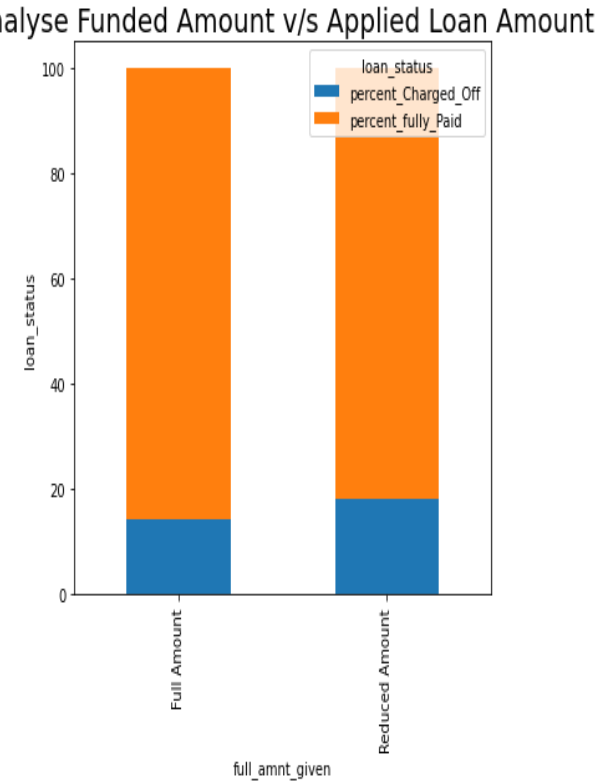
- Now the Graph indicates,
  - ❑ If interest rate is same then, charge off % decreases with increase in annual income
  - ❑ If income is same then, charge off % increases with increase in interest rate
- Similar Analysis performed on other variables.
- Annual income and DTI vs Loan Status
- Term, Purpose, Home ownership, income verification against Charge off %.

Univariate Analysis :









Higher the Funded Amount of the Loan, higher the percentage of Charged off.

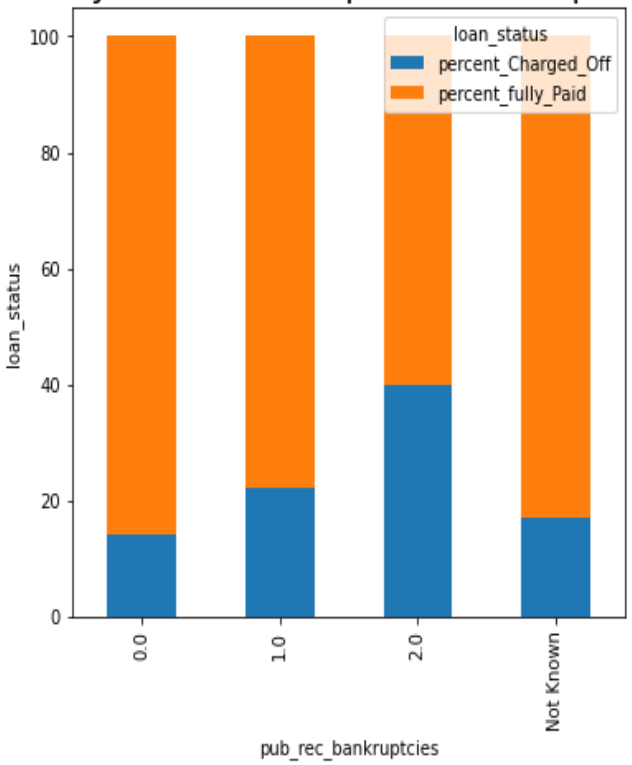
With increse in interest rate is, the percentage of Charged off increases.

Higher the Tenure results into higher percentage of Charge Off.

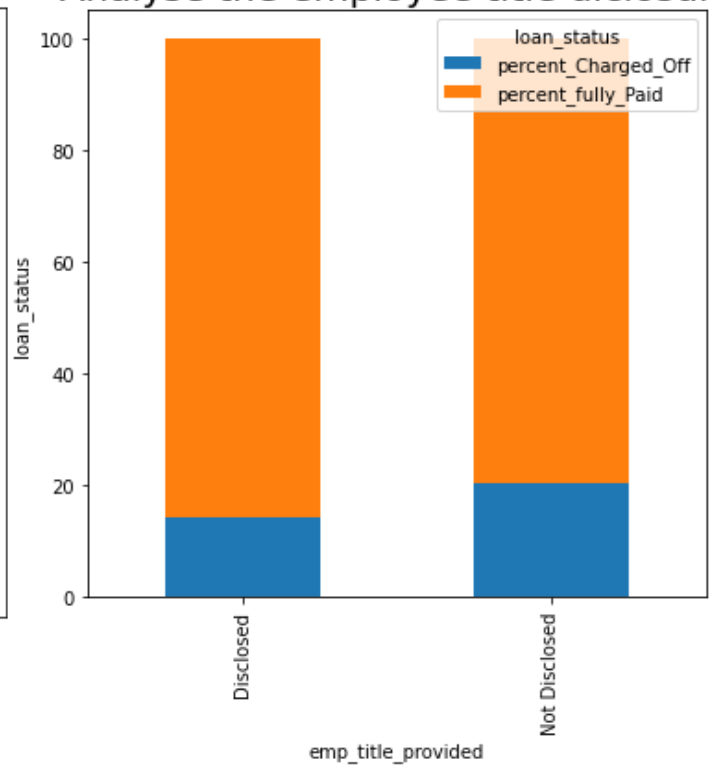
Higher chances of charge offs when the reduced amount is funded.

Univariate Analysis

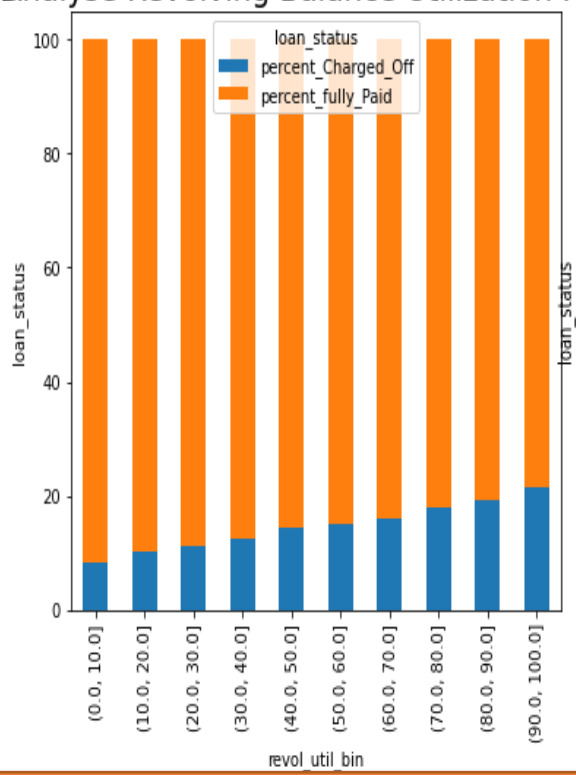
Analyse Number of public bankruptcies



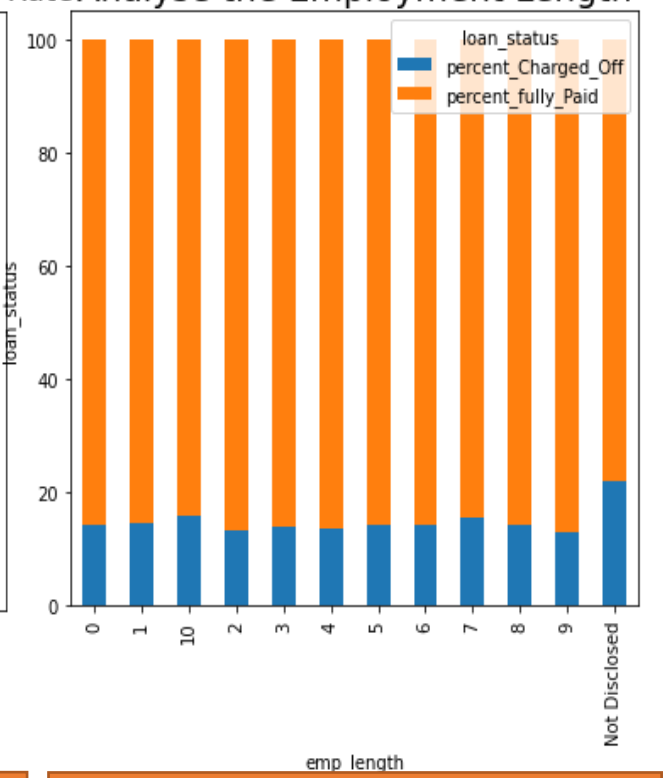
Analyse the employee title diclosure



Analyse Revolving Balance Utilization Rate



Analyse the Employment Length



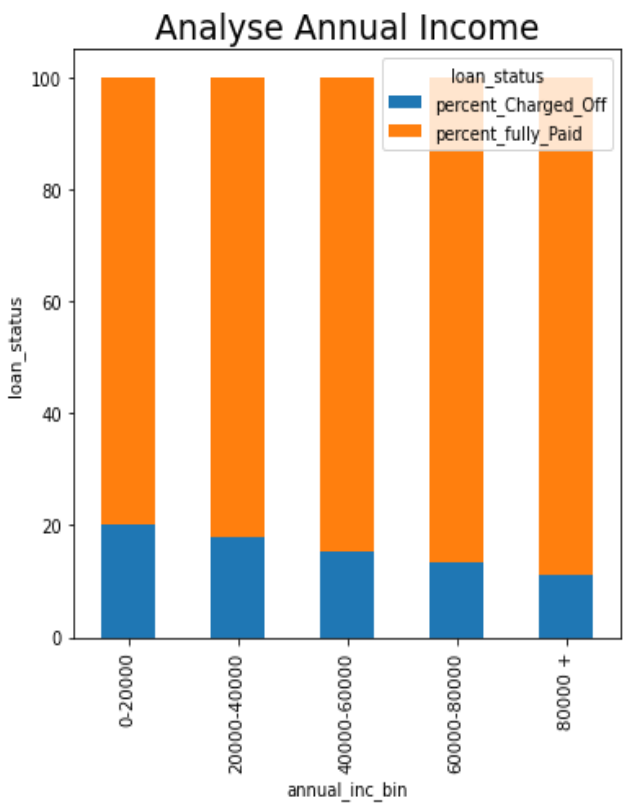
The charge off rate is increasing with increase in Bankruptices

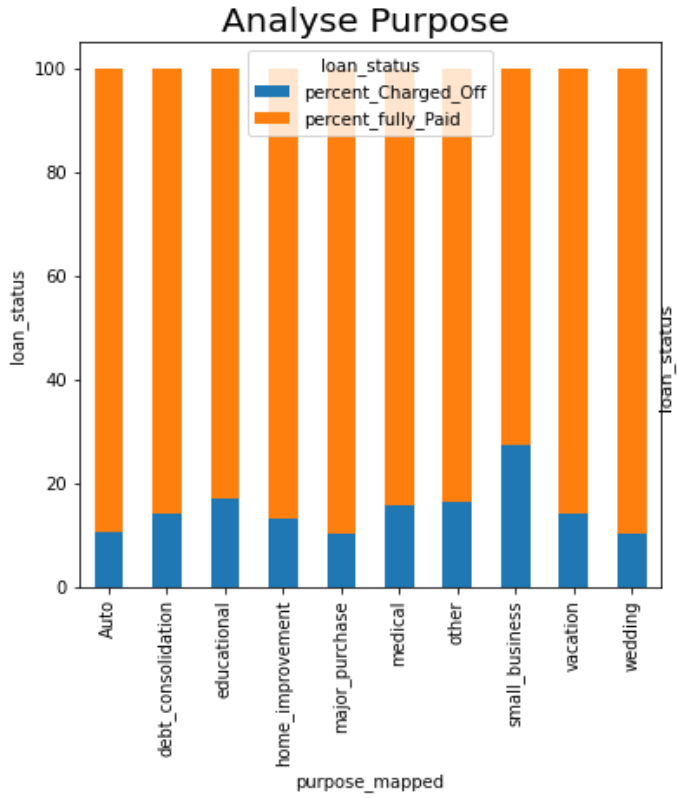
If the Employee Title is not disclosed, then charged off percentage is higher.

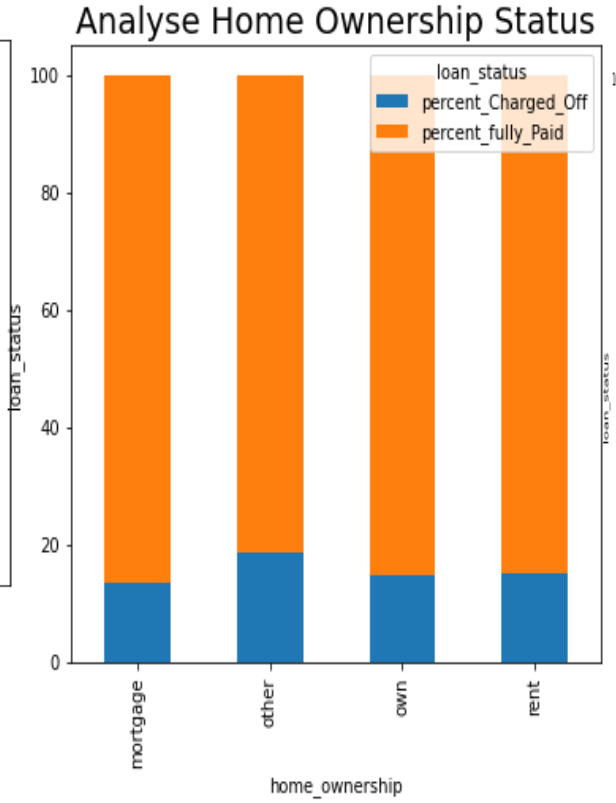
With increase in revolving balance utilization, charge off percentage is increasing

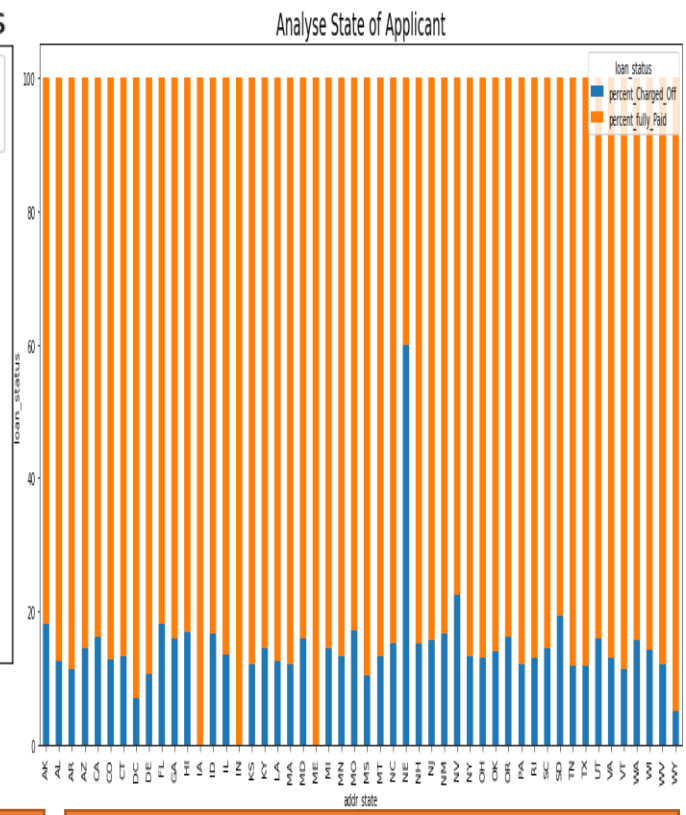
If employment length is not disclosed, then Charge Off rate is higher

Univariate Analysis :









Charge offs decreases with increase in Annual Income.

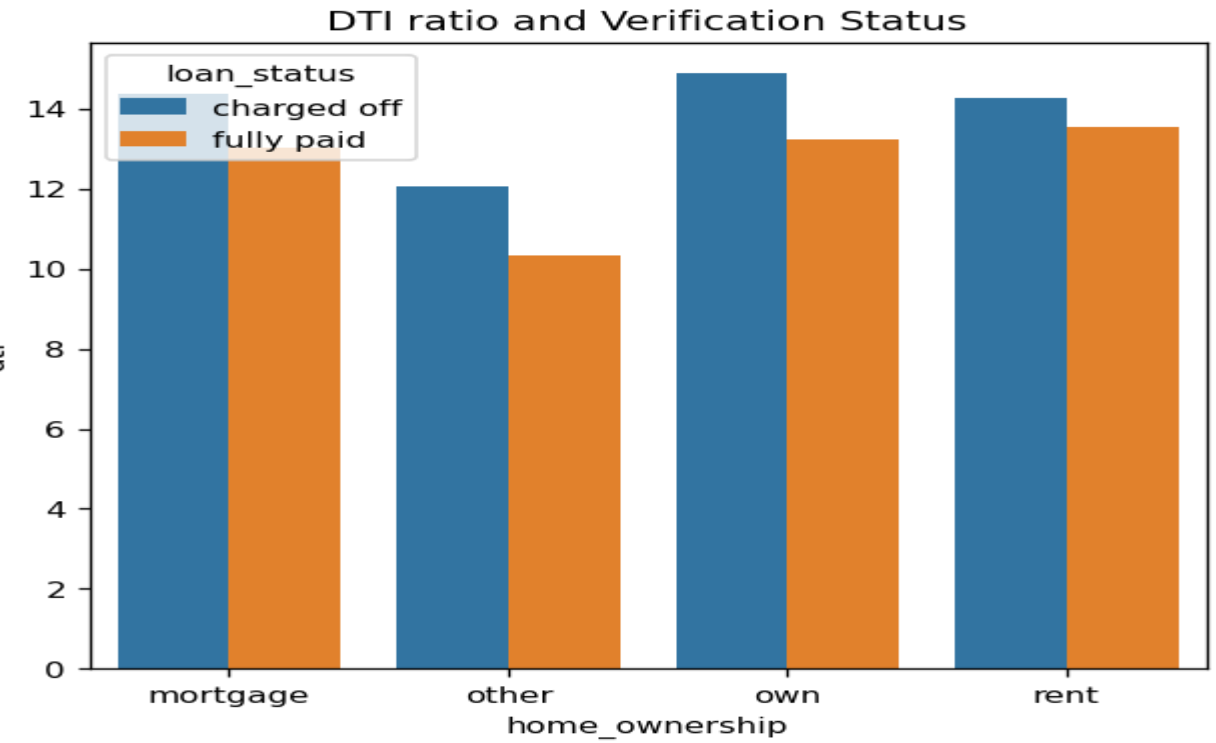
Small Business and Educational loans are highly charged off.

"Other" Home ownership has higher percentage of Charged off.

Percentage of Charge offs are higher in NE.

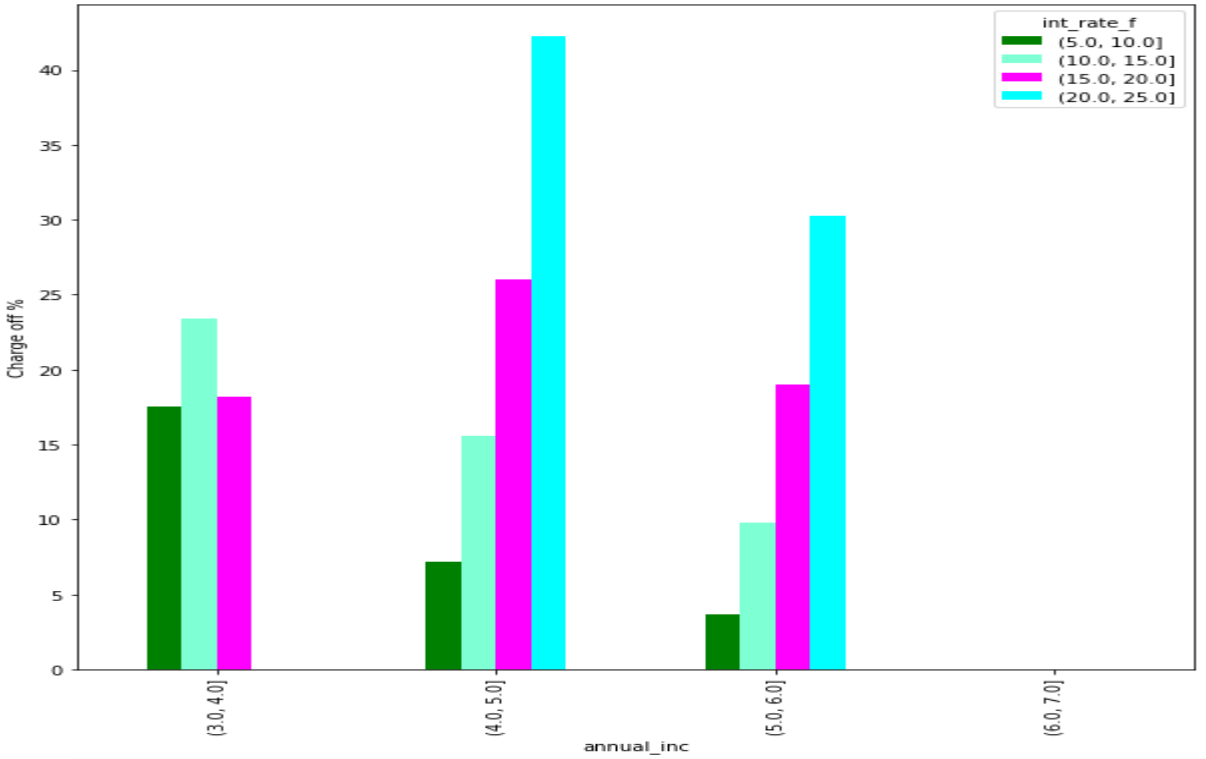
Bivariate Analysis v/s Loan Status / Charge off %

Home ownership and Dti



Observation : Percentage of defaulting loan is higher in the below mentioned scenario. If the DTI is higher and The home ownership is 'own'

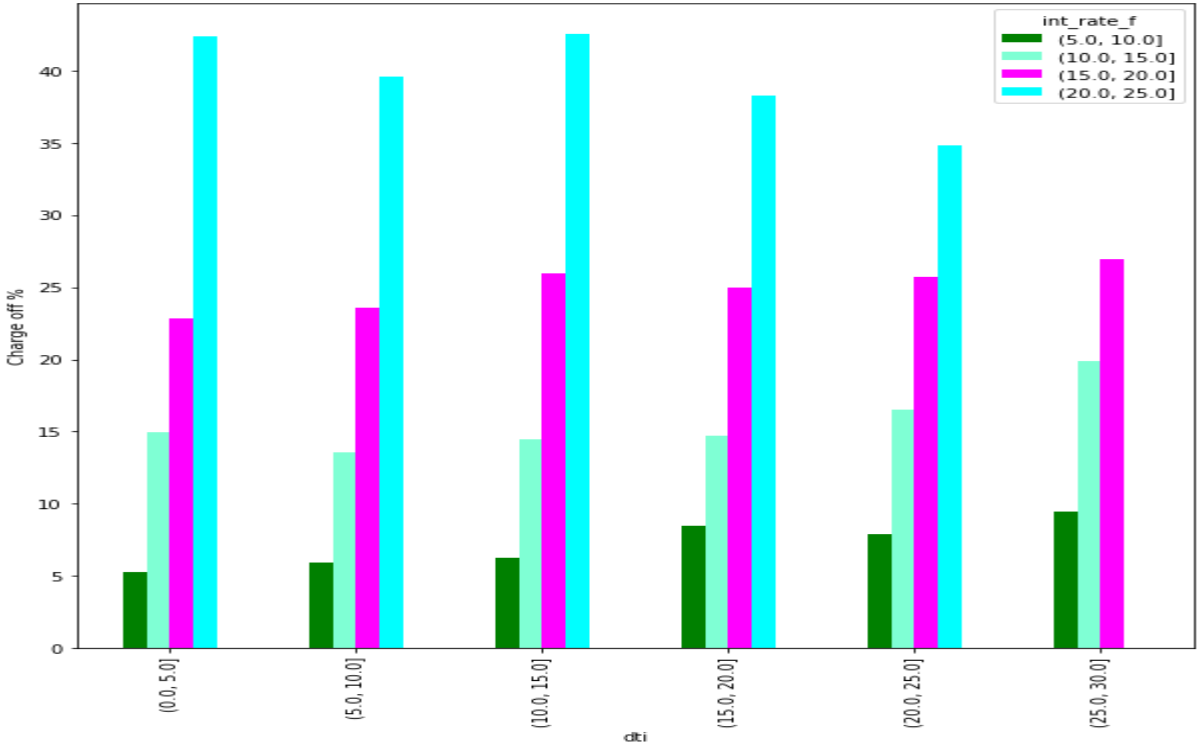
Annual income and Interest rate



Observation : If interest rate is same then, charge off rate decreases with increase in annual income.
If income is same then, charge off increases with increase in interest rate

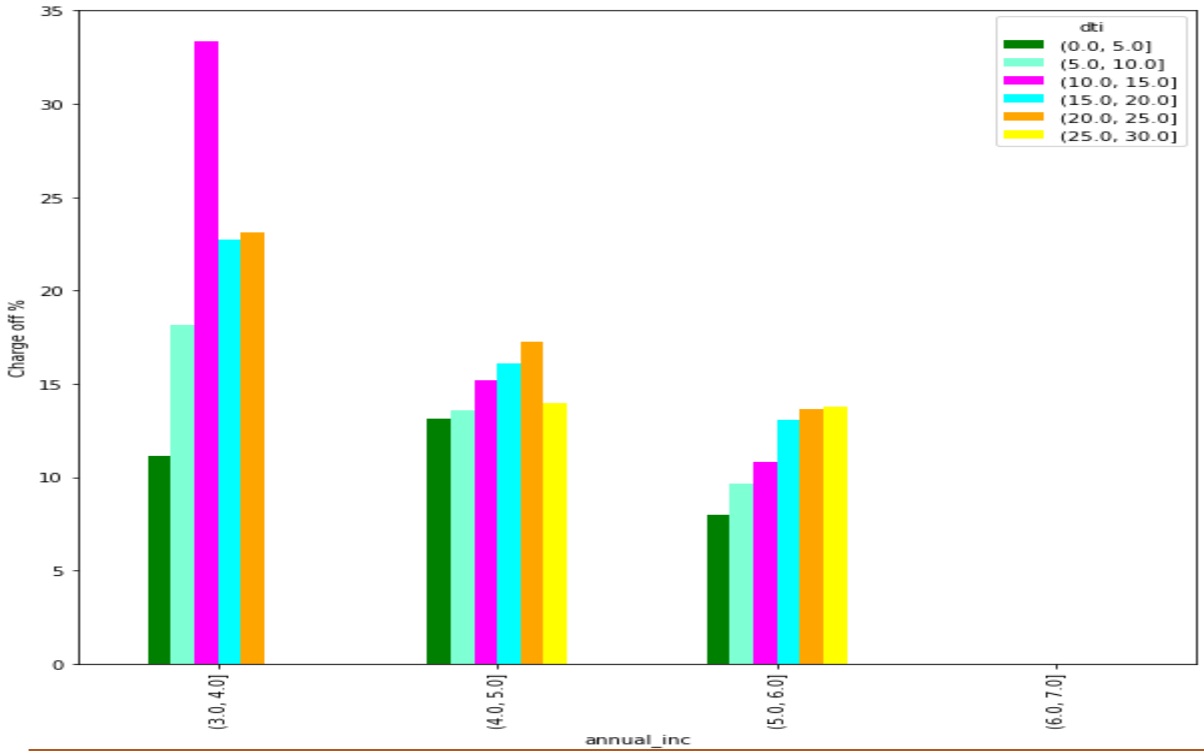
Bivariate Analysis against Charge Off %

DTI and Interest Rates



Observation:
If dti is same then, higher interest rates result in more charge-offs Higher interest rates have significant impact on charge offs

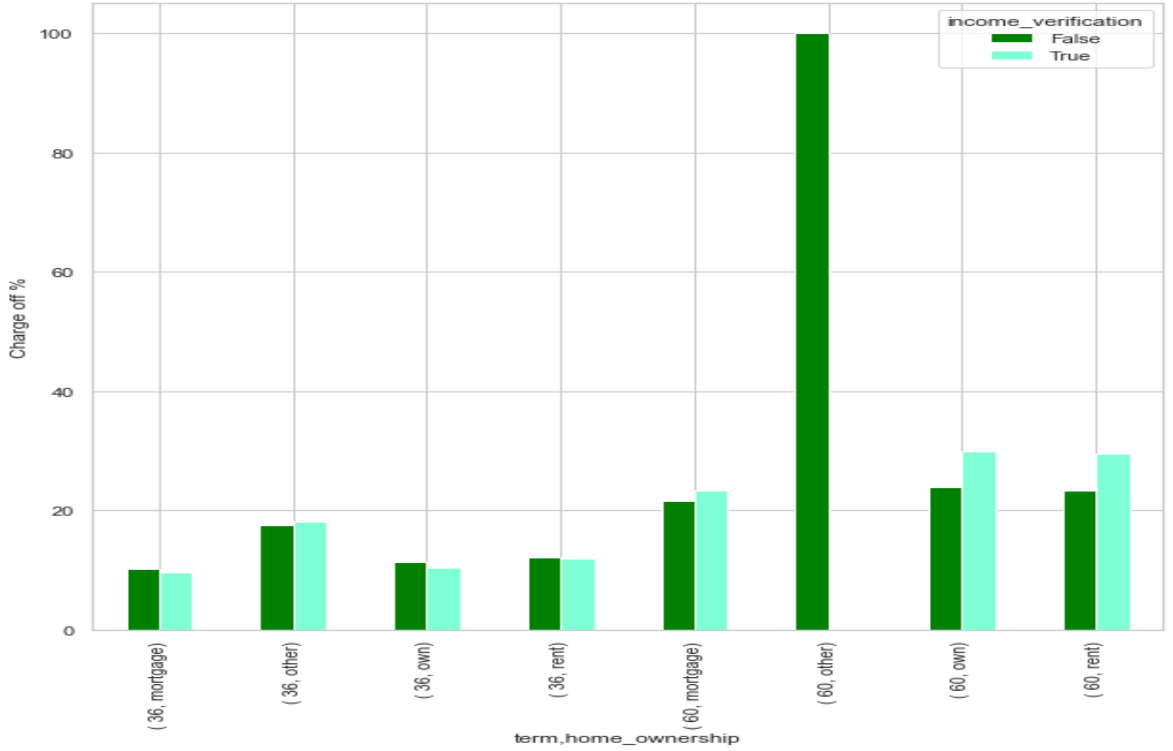
Annual income and DTI



Observation :
If income is same then, charge off increases with increase in DTI
If income is same then, charge off decreses with increase in salary

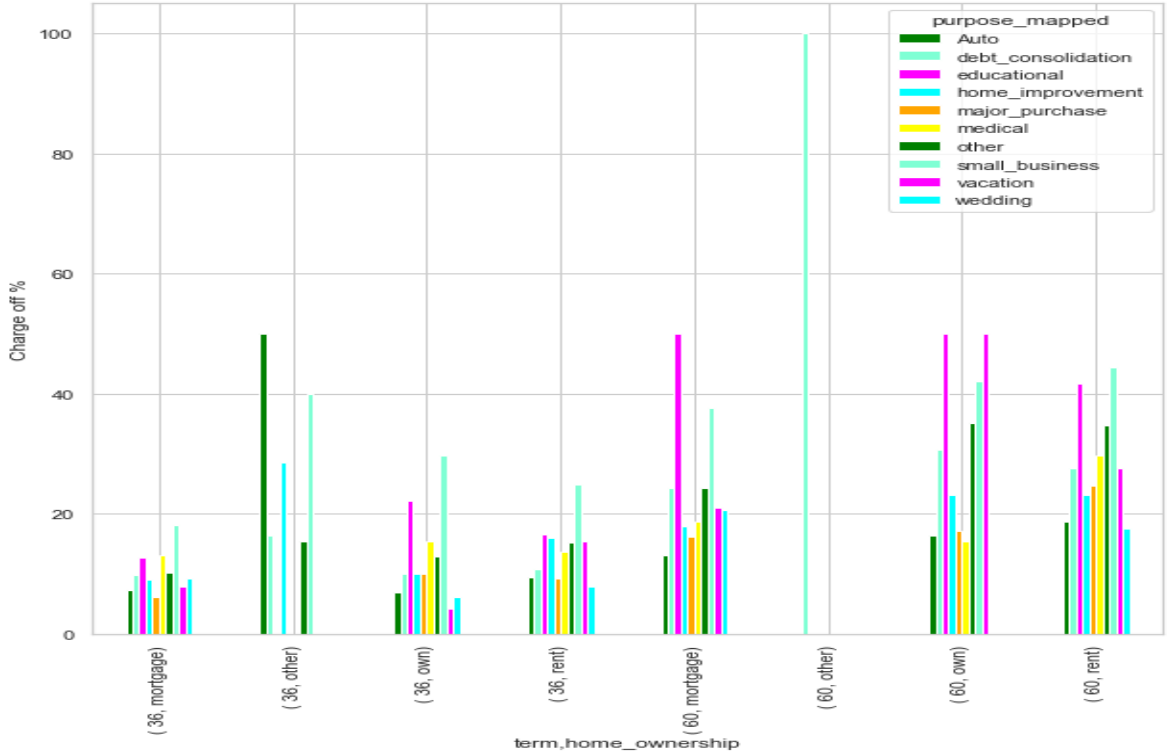
Multivariate Analysis

Term, Home Ownership and Income Verification vs loan status



Observation: Highest Charge off % in the following scenario When income is not verified, and Home ownership is 'Other' and Term is 60 Months

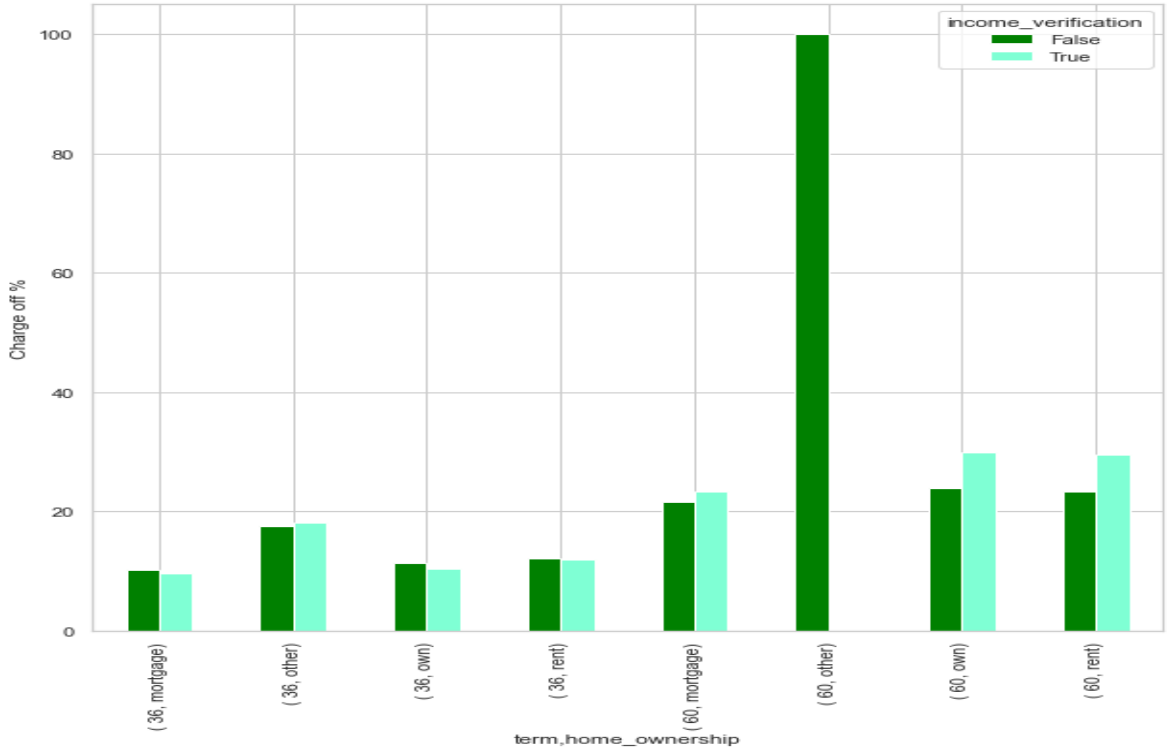
Term, purpose and home_ownership vs loan status



Observation: Highest Charge Offs % in the following scenario Purpose is Debt Consolidation and Term is 60 Months and Home ownership is 'Other'

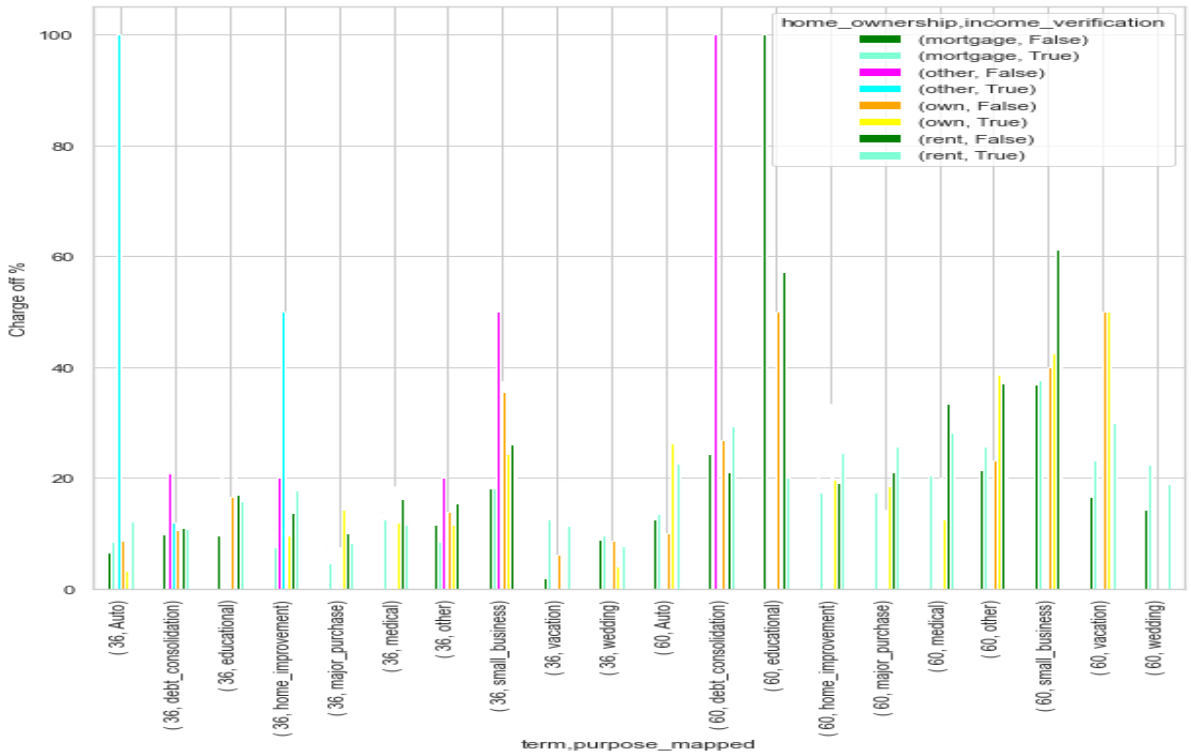
Multivariate Analysis

Term, Home ownership, Income verification v/s Loan Status



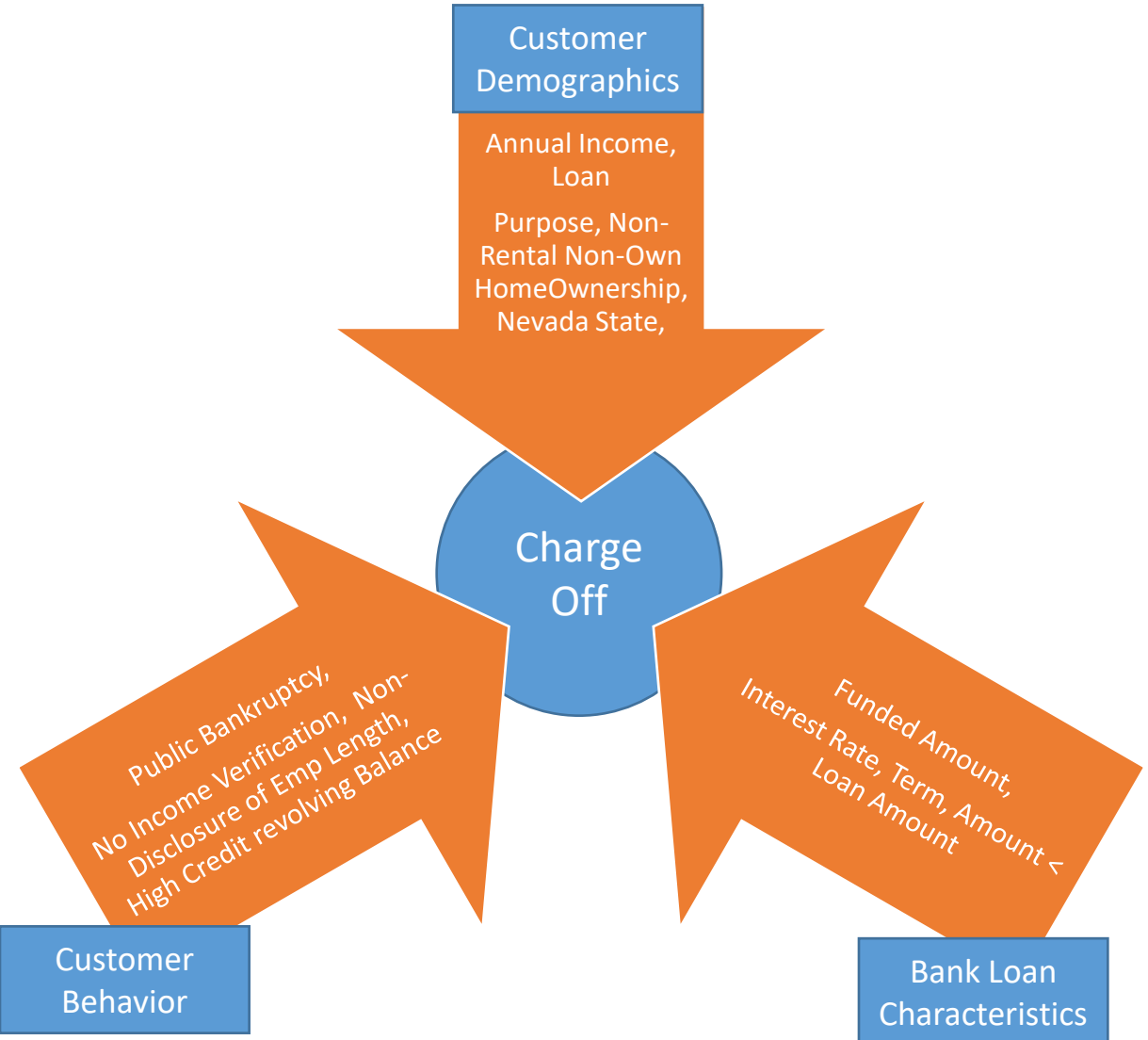
Charge offs % is highest for the below mentioned scenario.
Term is 60 and Home ownership is other and income verification

Term, purpose, Home ownership, Income verification v/s Loan Status



Charge off % is highest in the below mentioned scenario.
When Term is 60 (Purpose is Debt Con, Education) and (Home Ownership is 'other' Mortgage) and (income not Verified)

Conclusion : Critical Observations and Recommendations



Critical Observations : **Charge offs % is higher for the below mentioned scenarios.**

1. When income is not verified, and Purpose is educational, and Term is 60 Months

2. Term is 60 Months and Purpose is Debt Consolidation and Home ownership is Other

3. Term is 60 and Home ownership is other and income verification

4. Term is 60, Purpose is Debt Consolidation and Home Ownership is other and income Not verified,

5.Term is 60, purpose is educational, and Home Ownership is mortgage and income verification is False

Recommendations: Before taking Go–No-Go lending decisions, Club should analyze various such Combinations

Thank You