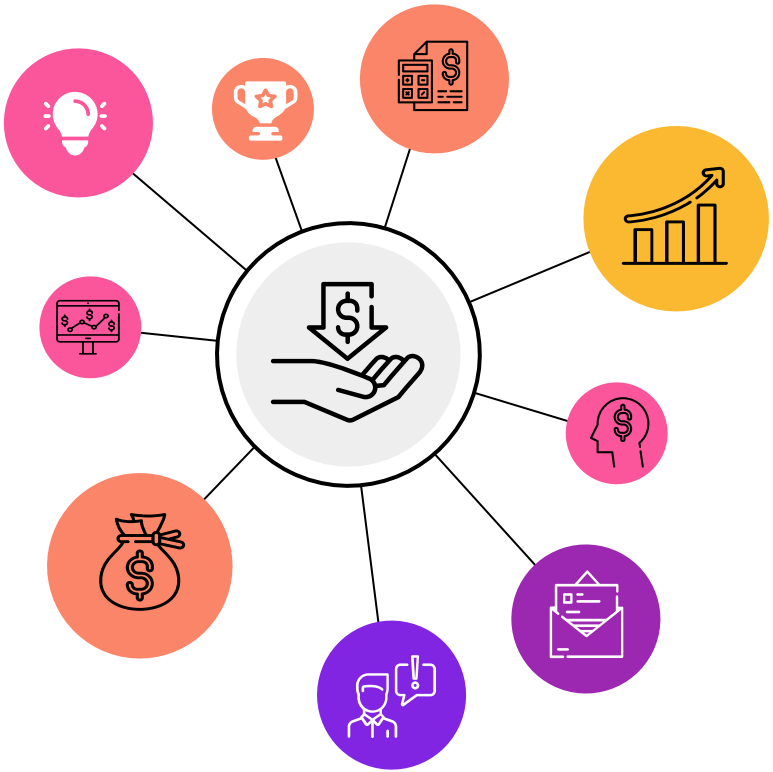
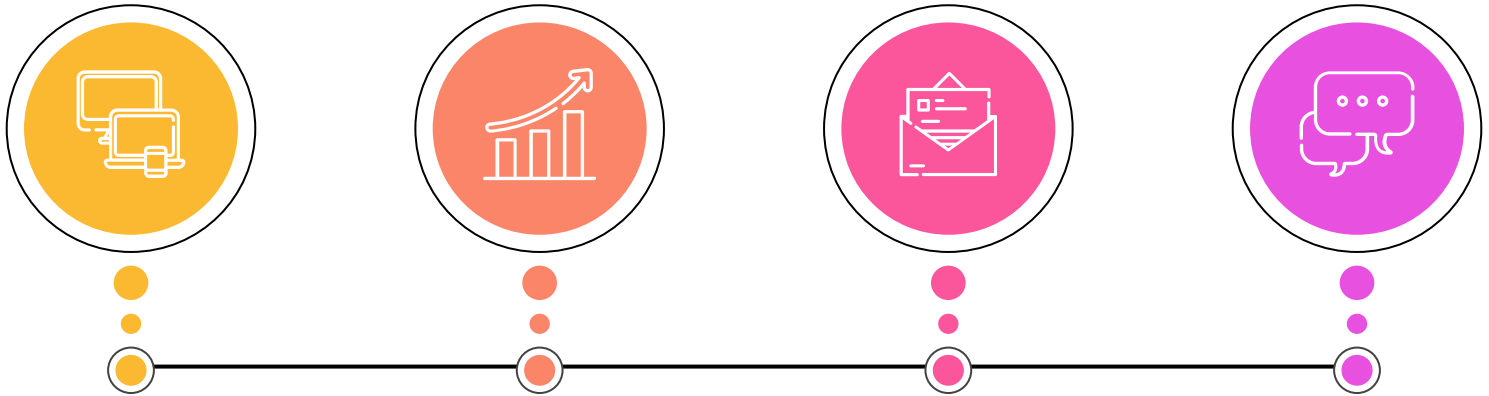


Lending Club Case Study

- *Lena Arora*

Table of Contents



Problem Statement & Approach

This describes the problem statement and approach used during the analysis.

Data Cleaning & Sanity Checks

This part involves cleaning data and doing sanity checks to make the analysis more accurate.

Univariate & Bivariate Analysis

Certain charts are used to describe the results of analysis which has been detailed in this part.

Observations & Conclusion

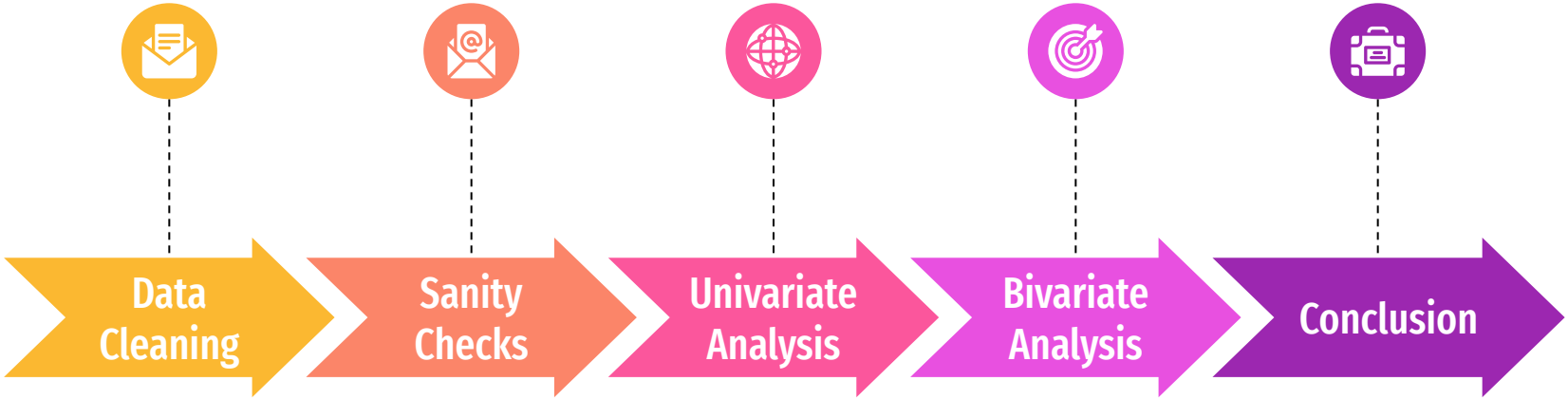
Observations and conclusions have been given so as to determine what company should do to lower the rate of Defaulters.

Problem Statement

A consumer finance company aims to identify patterns in loan applicant data to predict loan default. By analyzing historical data, the company seeks to understand the factors that influence default risk.

This knowledge will enable them to make informed decisions regarding loan approvals, interest rates, and risk mitigation strategies. The goal is to minimize financial losses by reducing the number of loans granted to high-risk applicants.

Approach



1. Null Data Treatment
2. Outliner Treatment

1. Sanity Checks

1. Box Plot
2. Distribution Plots

1. Bar Chart
2. Box Plot

1. Data Inference
2. Recommend ations

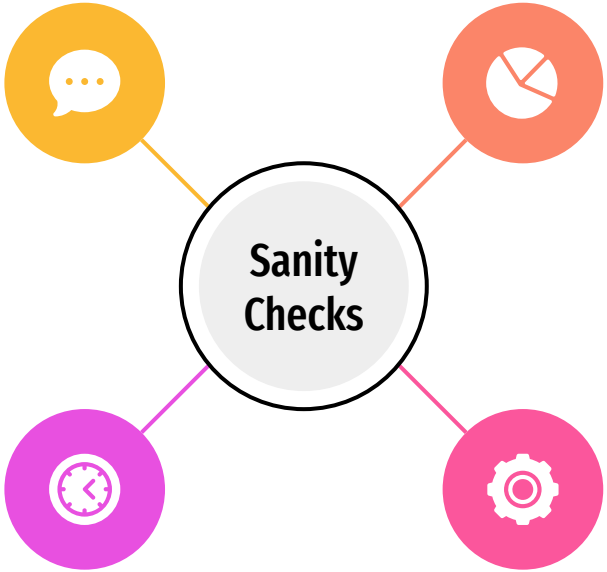
Data Cleaning

1.NaN Columns

Columns with null values has been removed to keep the data clean.

2. Dropping unnecessary Columns

Unnecessary columns which cannot be made factors in loan decision has been removed.



3.Missing Value Treatment

Missing values have either been removed or either been filled up with mode.

4. Checking

Info() and describe() attributes have been taken to check and draw summaries from the data.

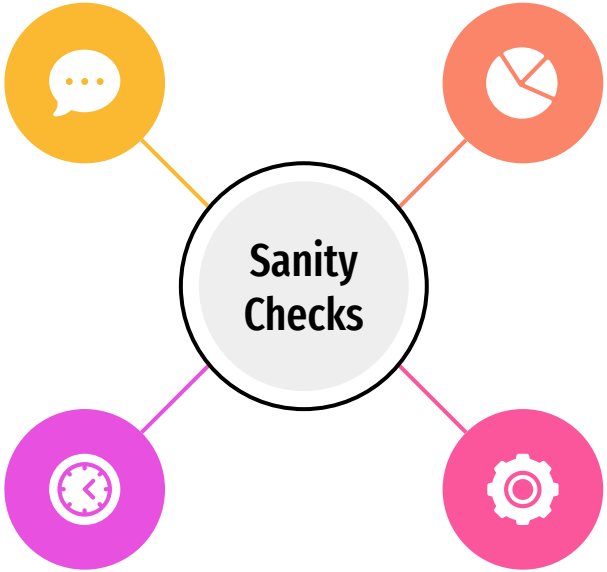
Sanity Checks

1.Loan Status

Sanity check has been performed in loan_status to check the unique target values, whereas continuous variables has been removed.

2. Cleaning revol_util

Percentage sign has been removed from this column to analyse the correct trend.



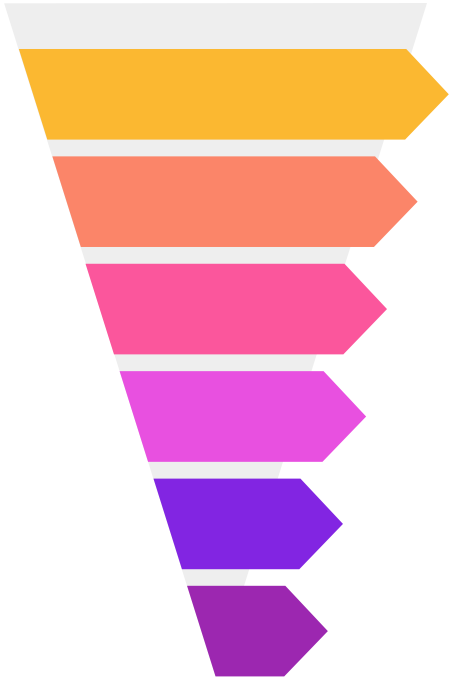
3. Interest Rate

Similar percentage sign has been removed from Interest Rate to analyse better.

4. Home Ownership

The columns with NaN values has been applied to Others column.

Univariate Analysis & Bivariate Analysis

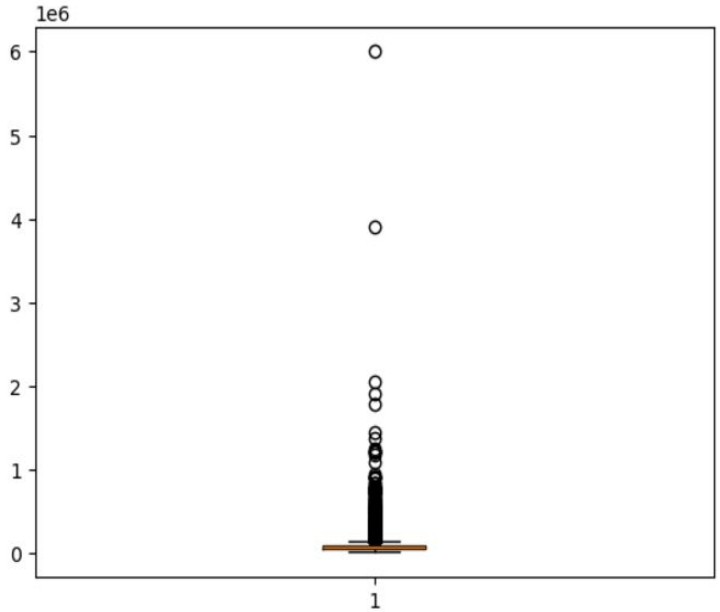


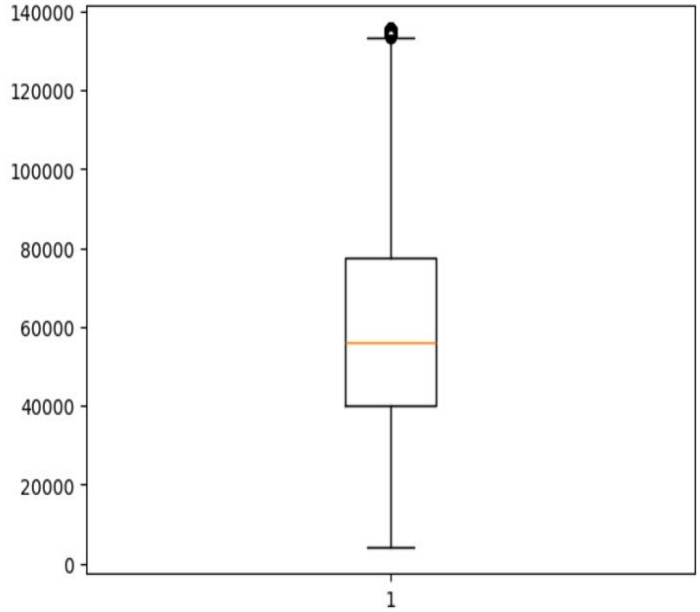
Outliner Treatment	A simple plan for organizing code and logic flow.
Box Plot	Graph showing distribution, median, and outliers of data set.
Bar Chart	Chart displaying data with rectangular bars representing different values.
Distplot	Graph showing distribution of data with histogram and kde.
Count Plot	Graph showing counts of categorical data using bars.
Scatter Plot	Graph showing counts of categorical data using bars.
Correlation Matrix	Table showing relationships between multiple variables with values.

Outliners

Annual Income

- Outliner Treatment has been performed.
- All outliners have been removed.
- Annual Income is taken till 1,35,000 p.a





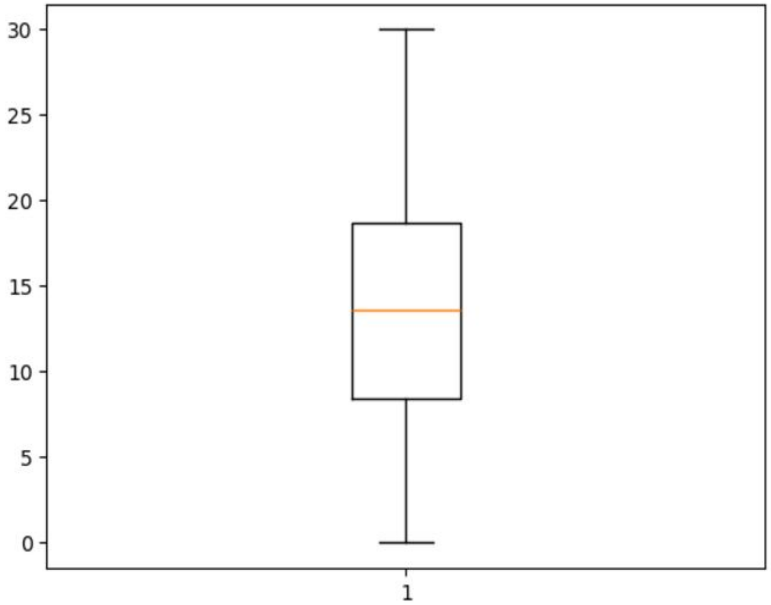
Outliners

DTI

- A ratio calculated using the borrower's total monthly debt payments on total debt obligations, excluding mortgages and the requested LC Loan, divided by borrower's self reporting income.
- The DTI ratio helps lenders assess the risk involved in lending money to a borrower. A lower DTI indicates that a borrower has more income available to take on additional debt.

Observation:¶

Here it is observed that dti is largely above 1, so an impression can be drawn that the borrower is over-leveraged or relying too heavily on credit to meet their financial needs. This could indicate underlying financial issues that may need to be addressed.



Analysing loan Status

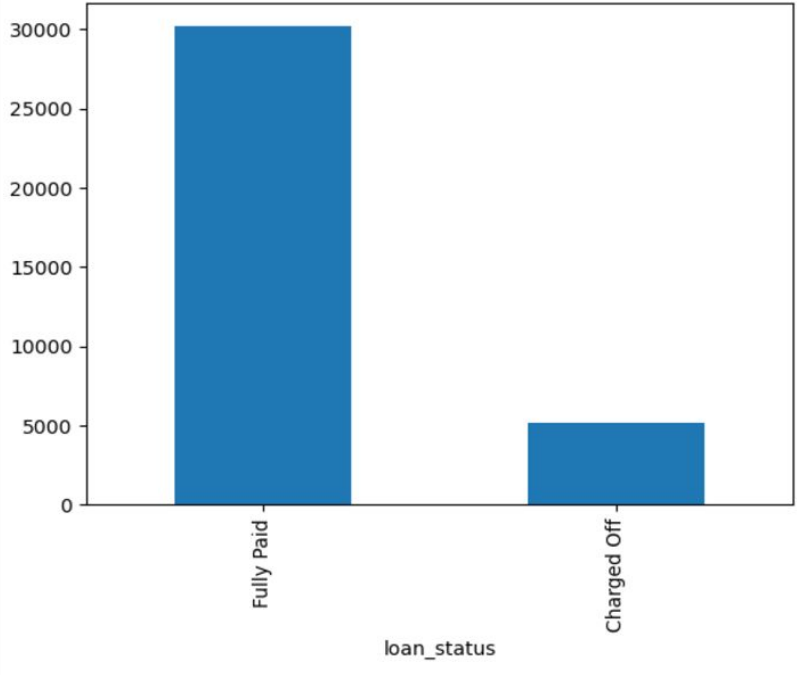
Loan Status

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

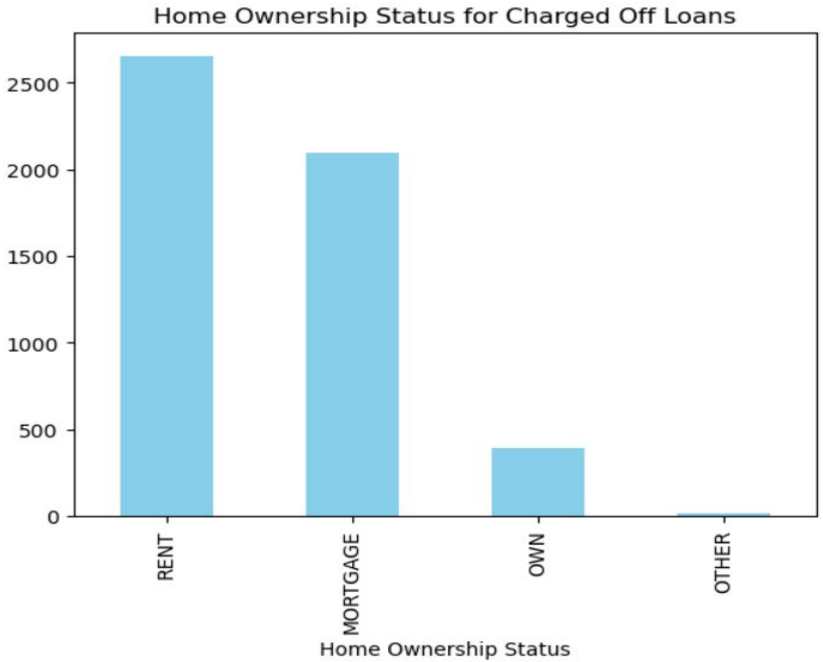
- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

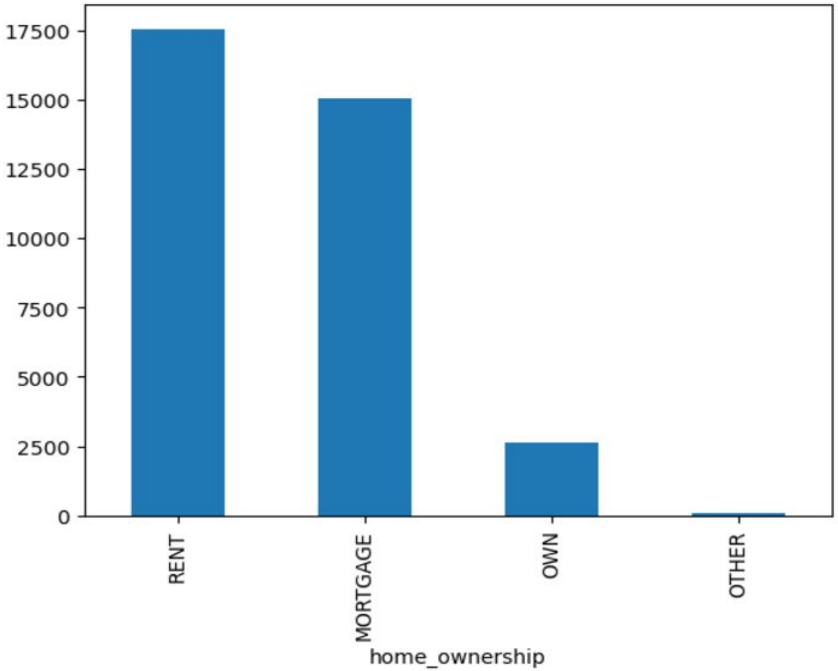
Observation:

Charged off loans are nearly 4500 - 5000 over 30000 loan takers, which means that nearly 14-16% are defaulters.



Analysing Home Ownership

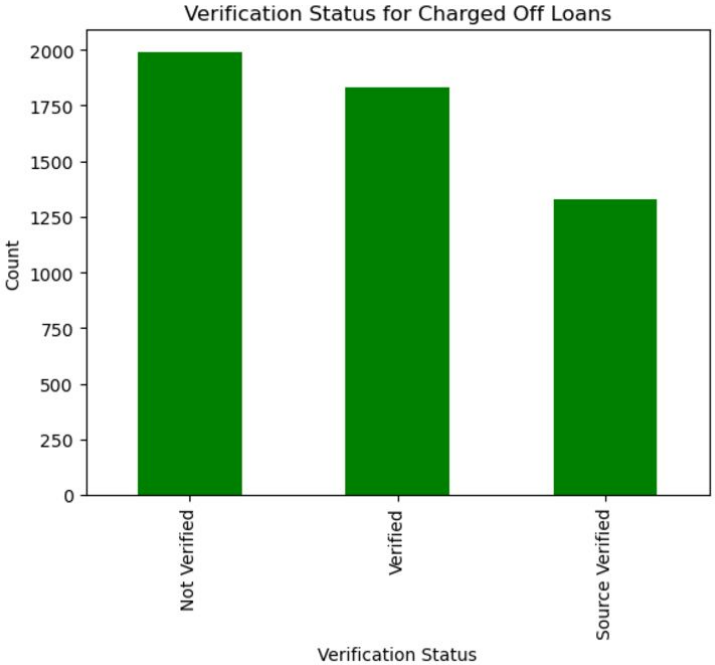




Observation:

It has been observed that people who live in rent, tend to charge off loans. Second category belongs to people who have mortgaged their loans

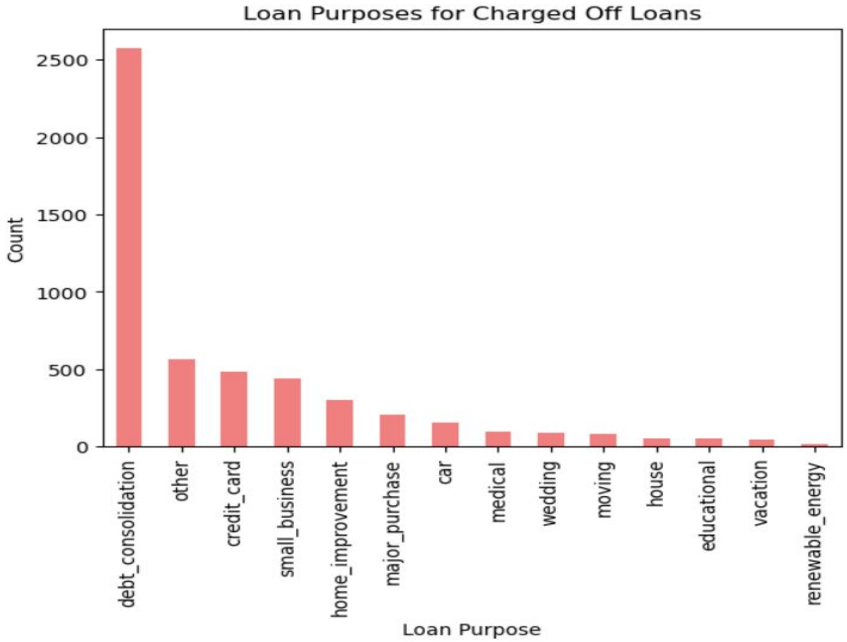
Verification Status



Observation:

People who are not verified by the team tend to charge off their loans.

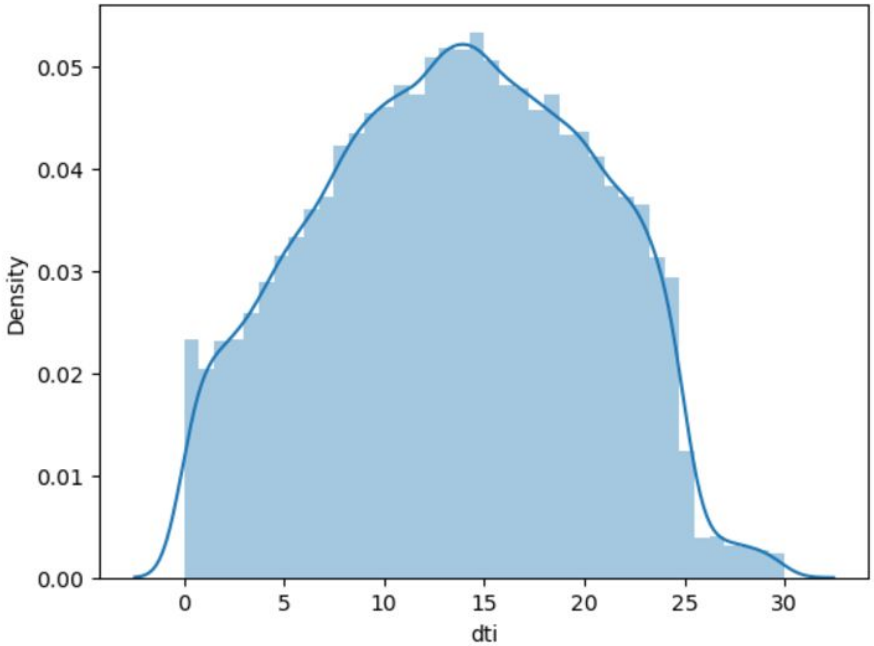
Loan Purpose

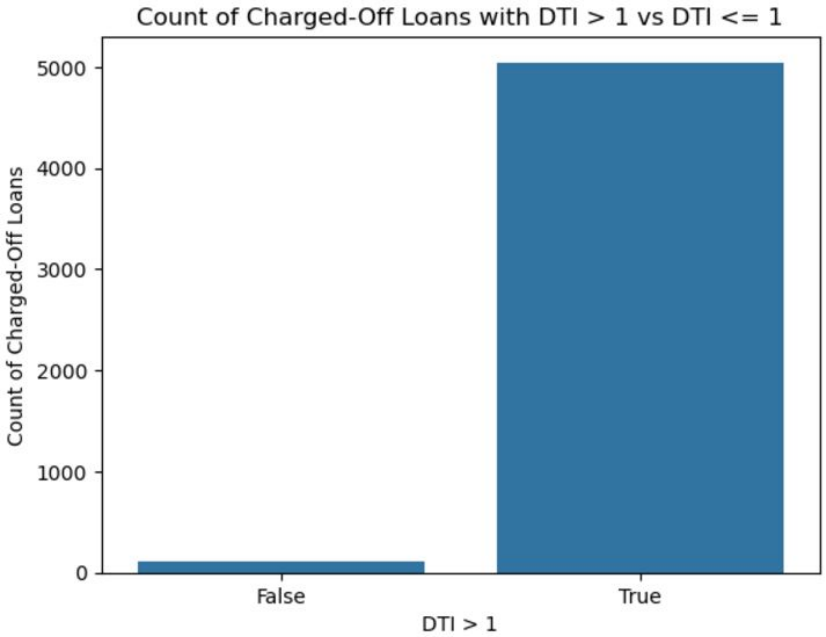


Observation:

People who have taken loan for debt consolidation tend to default more than others.

DTI v/s Charged Off





DTI v/s Charged Off

Observation:

True (DTI > 1): The majority of charged-off loans fall into this category, suggesting that most of the loans that were charged off have a DTI greater than 1.

False (DTI ≤ 1): Fewer charged-off loans have a DTI of 1 or less, which indicates that lower DTI values are less common among the loans that were charged off.

Since most of the charged-off loans have a DTI greater than 1, it could imply that borrowers with a higher DTI are more likely to default or have their loans charged off. This makes sense because a higher DTI suggests that the borrower has more debt relative to their income, which could indicate financial strain and difficulty in repaying the loan.

Loans with a DTI greater than 2 show a substantial increase in charge-offs, while loans with DTI below 1 are less likely to be charged off.

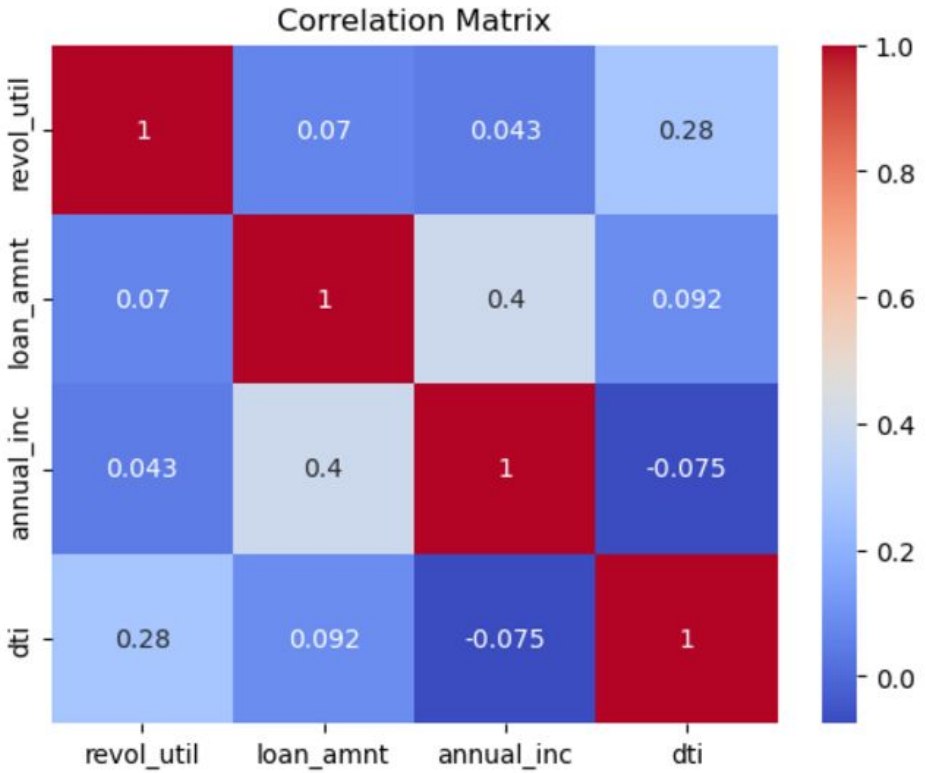
Relationship between Annual Income & Int Rate by Home status

Observation:

1. An observation can be drawn from here that people with an annual income of between 25000 - 75000 are likely to default more.

2. Secondly, we can even see that people default more when the rate of interest is between 10% - 17.5%.

Correlation Matrix



Observation:¶

The negative correlation of -0.075 between annual income and revolving utilization rate suggests that higher income slightly reduces the need to use credit, but the relationship is weak.

Other factors, such as debt-to-income ratio (DTI) and loan amount, show stronger positive correlations with utilization, indicating that individuals with higher debt burdens or larger loans are more likely to use more of their available credit.

Although annual income plays a role in credit utilization, it is less significant compared to DTI and other financial factors. Therefore, charge-off risk is more closely related to these other variables than to income alone.

Results Provided

- All codes is mentioned in one well-commented Python file; briefly mentioned the insights and observations from the analysis
- Presented the overall approach of the analysis in a presentation:
  - Mentioned the problem statement and the analysis approach briefly
  - Explained the results of univariate, bivariate analysis etc. in business terms
  - Included visualisations and summarised the most important results in the presentation