EDA CASE STUDY

By

Priyal and Shubham

Problem Statement:

For given dataset for people taking loans from banks and need to find which person is likey to default.

The aim is to 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.

Approach:

- First see the data and filter it by removing unwanted columns and fill the missing rows.

- Try to use various analysis techniques to find the insights,patterns in the data.

- Plot the graph and various visualization to get better understanding of data.

- Make decision based on result got from analysis and visualization.

Univarient Analysis:

- In the data provided we have column which is loan_status it gives various insights like how many people are fully paid,charged off and current.

- By doing the count of them we can get insights like mostly people paid their loan on time and very few of them got defaulted.

- Below slide contains the graph form univarient analysis.

Histogram of Status Categories
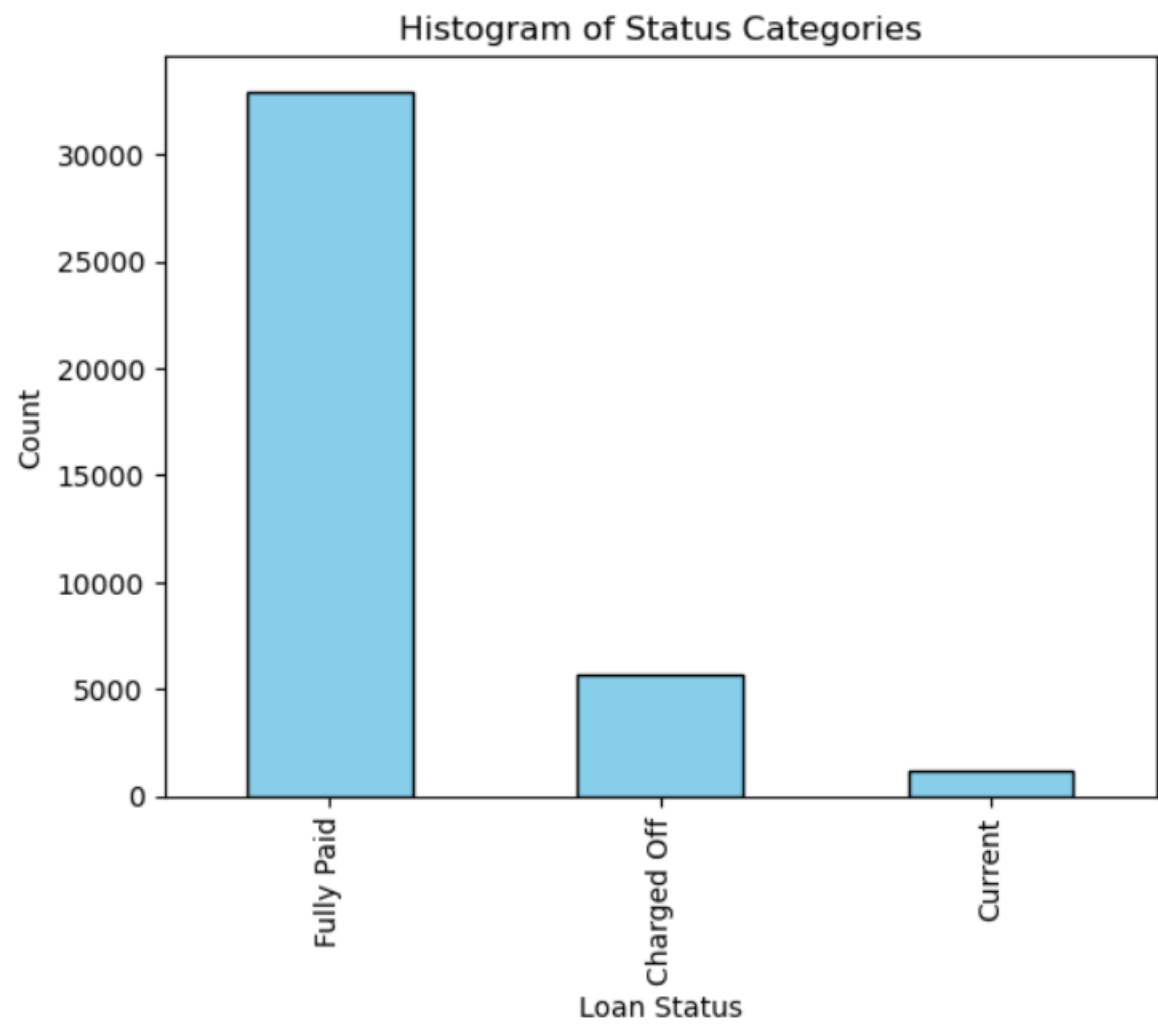
Bivarient Analysis:

- To get more insights from data taking two columns to see which person is more likely to be default.

- Taking loan_status and and home_ownership column which tells a lot about the people.
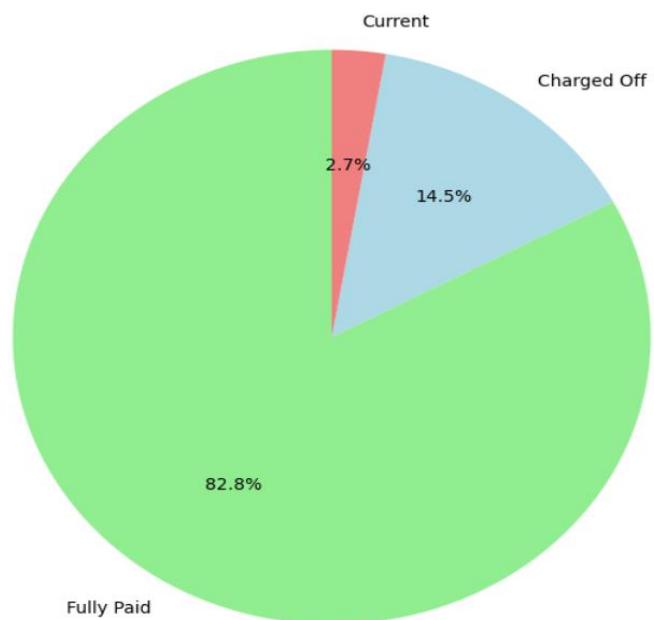
  Like-
    - Percentage wise in 15% people are charged off who are renting the house or living on mortgage.

  Conclusion :People living on rent and mortgage are having more chances to get default.
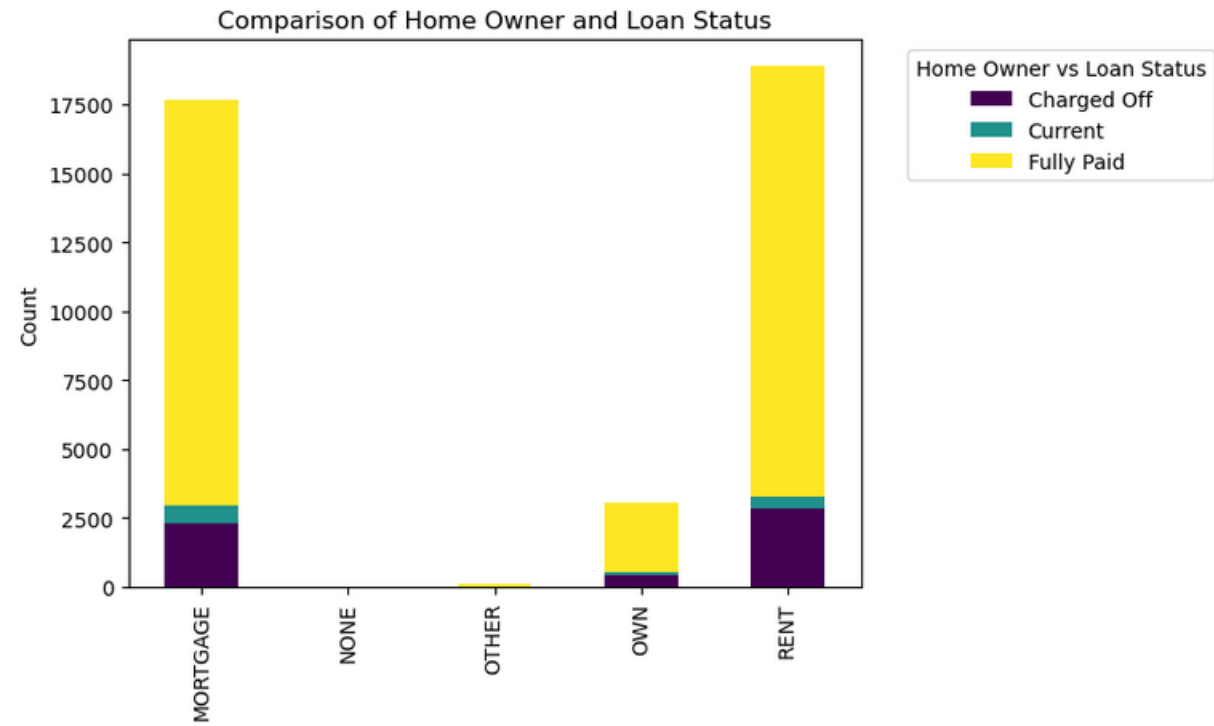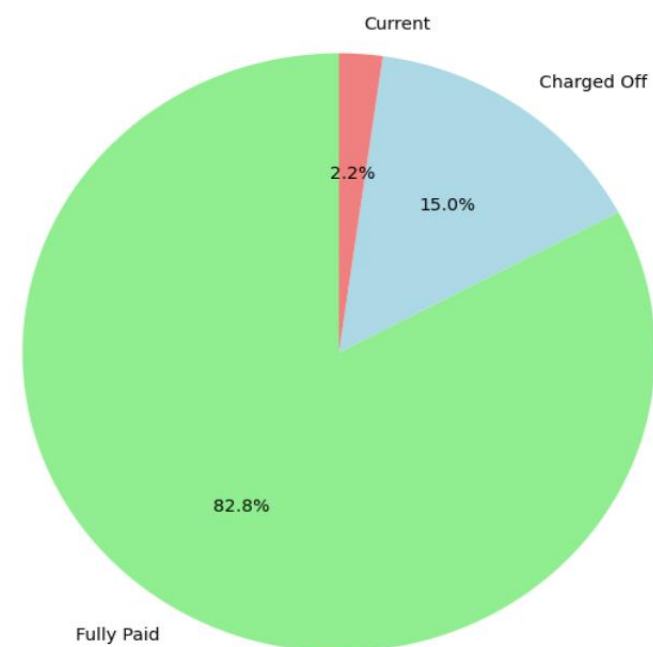
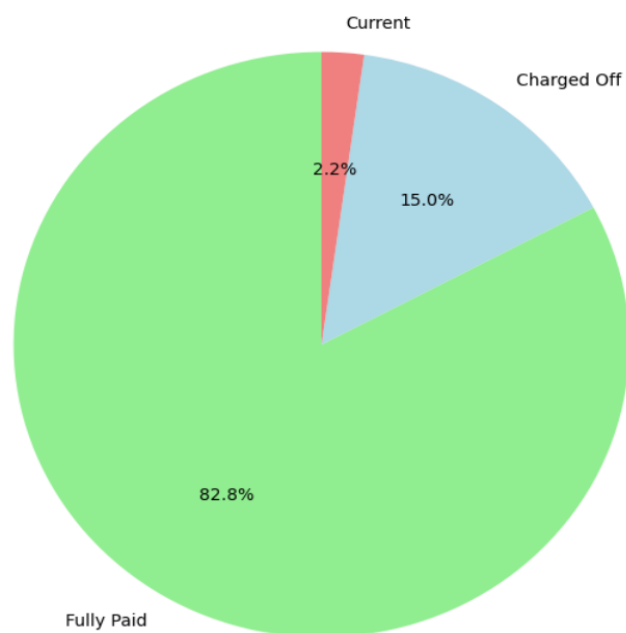This graph shows the comparision of home owner and their loan status

Distribution of Fully Paid Loans by Home Owner

Current — 2.7%
Charged Off — 14.5%
Fully Paid — 82.8%

Distribution of Fully Paid Loans by Mortgage

Current — 2.2%
Charged Off — 15.0%
Fully Paid — 82.8%

Distribution of Fully Paid Loans by Rent Category

Current — 2.2%
Charged Off — 15.0%
Fully Paid — 82.8%

Bivarient Analysis:
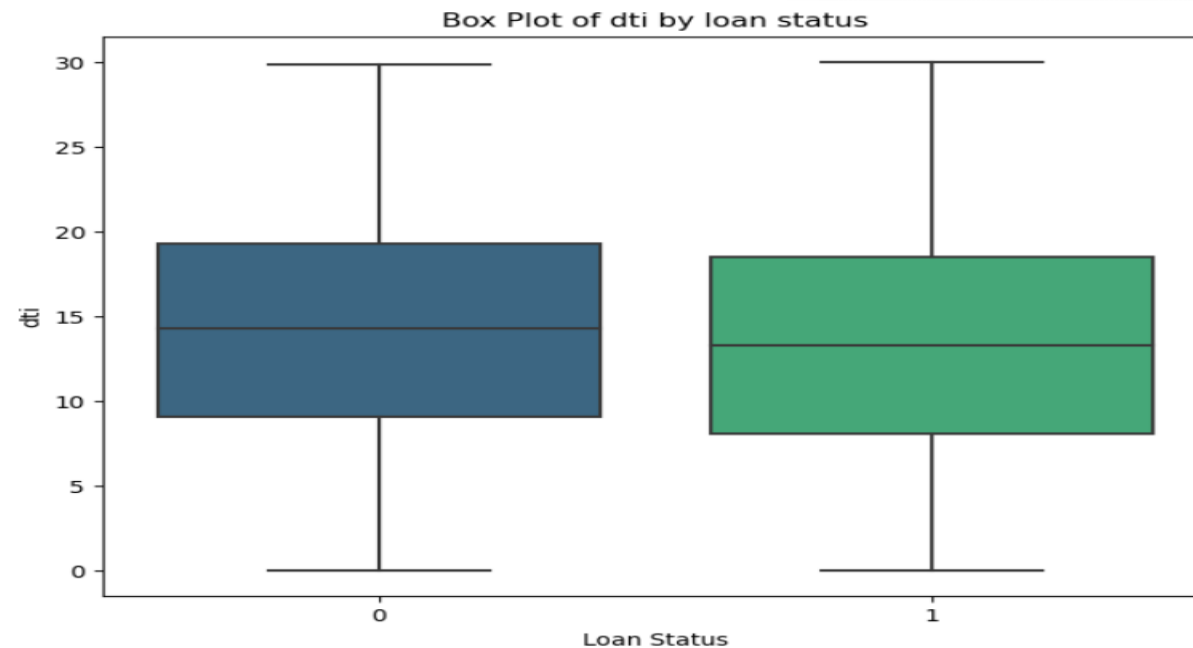
- Analysing verification_status with loan_status

  Conclusion:People are sourced verified are having higher changes of becoming defaulter than not verified and verified.
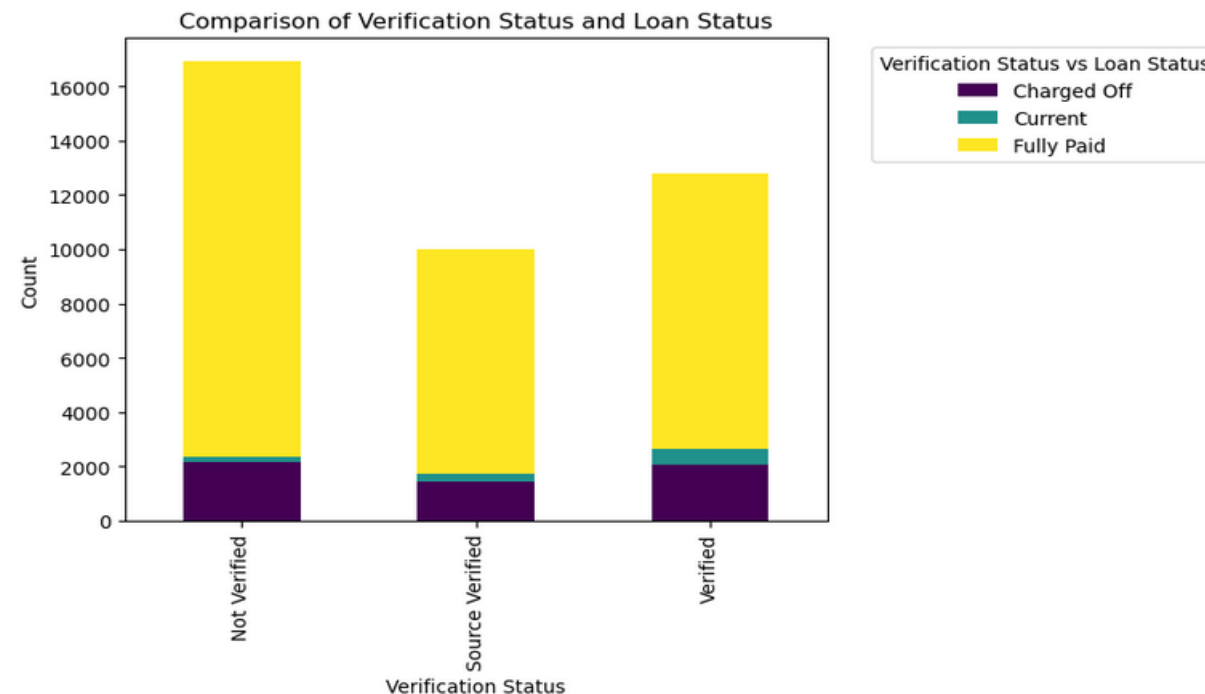
Bivarient Analysis:
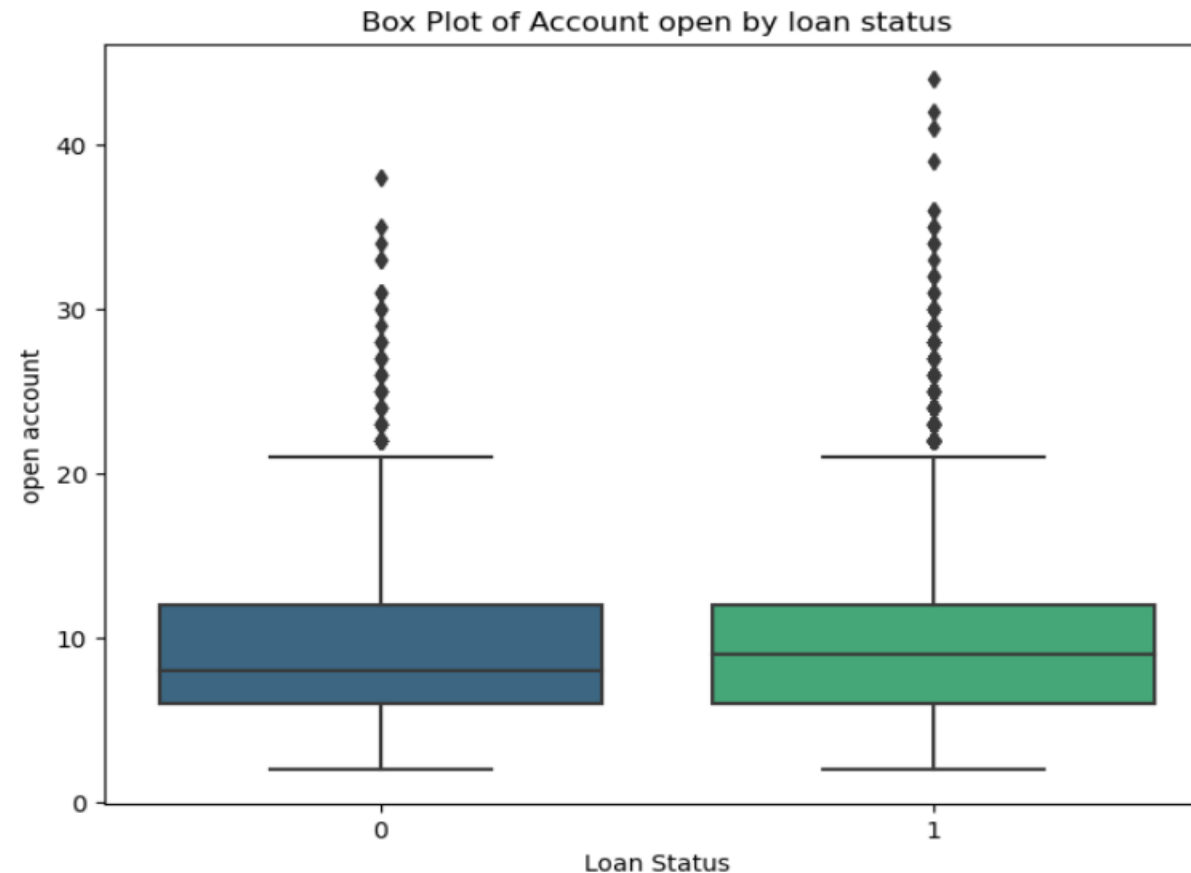
- Analyzing data between dti and loan status

  Conclusion: People who are having higher dti are more likely to get
  default.

Bivarient Analysis:

- Analysis with account open and loan status

  Conclusion:

CONCLUSION:

- People living on rent and mortgage are having more chances to get default.
- People are sourced verified are having higher changes of becoming defaulter than not verified and verified.
- People who are having higher dti are more likely to get default.