

Lending Club Case Study

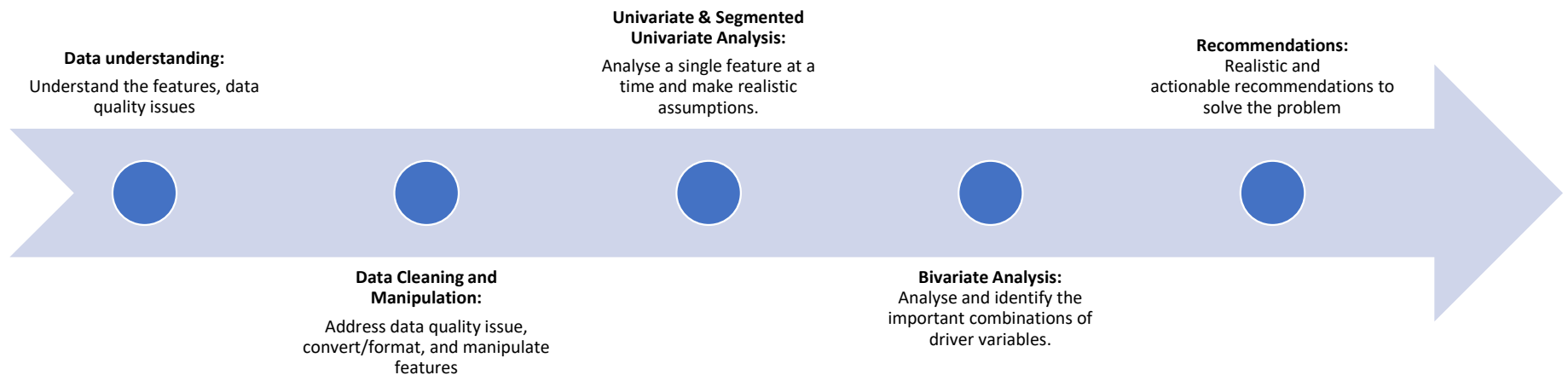
Biju Puthan Veetil

Executive PG Programme in Machine Learning & AI - March 2022

Problem statement

- The Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Borrowers can easily access lower interest rate loans through a fast online interface.
- lending loans to 'risky' applicants is the largest source of financial loss (called credit loss)
- The objective is understand the **driving factors (or driver variables)** behind loan default. The Lending Club can utilise this knowledge for its portfolio and risk assessment.

Analysis Approach



Data Understanding

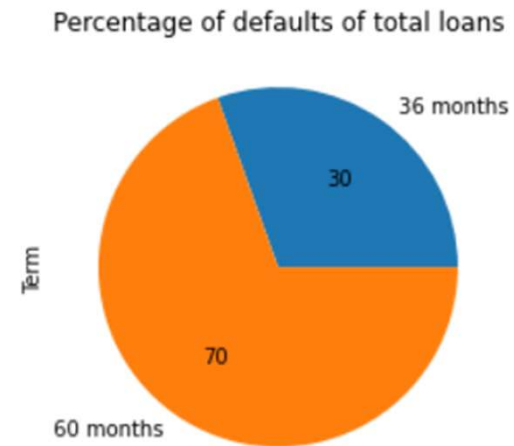
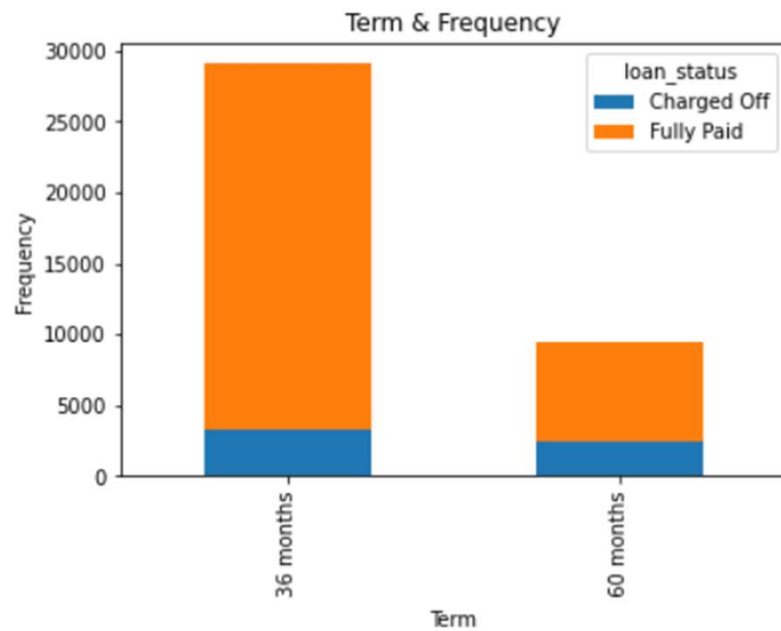
- Dataset: The complete loan data for all loans issued through the time period 2007 to 2011.
 - 111 features
 - 39719 records
- Data quality
 - 54 features with all record value NaN
 - 9 features with single unique value
 - 2 features with more than 25% NaN values
- Data Source:
 - loan.csv,
 - Data_Dictionary.csv

Data Cleaning and Manipulation

- Data Cleaning
 - Removed features with all records NaN
 - Removed feature with single unique values
 - Removed features with more than 25% NaN values
 - Removed behavior features not available at the time of application
 - Removed ID's and features containing only ID information as this does not contribute to analysis
- Feature transformation
 - Int_rate to float64
 - New feature int_rate_bins as bin of int_rate
 - New feature dti_q4 as bin of dti
 - New feature annual_inc_q4 as bin of annual_inc
 - New feature prncp_rem as remaining principle
 - Data frame df_bad contains only bad loans
 - Data frame df_loan contains only bad and good loans

Univariate & Segmented Univariate Analysis

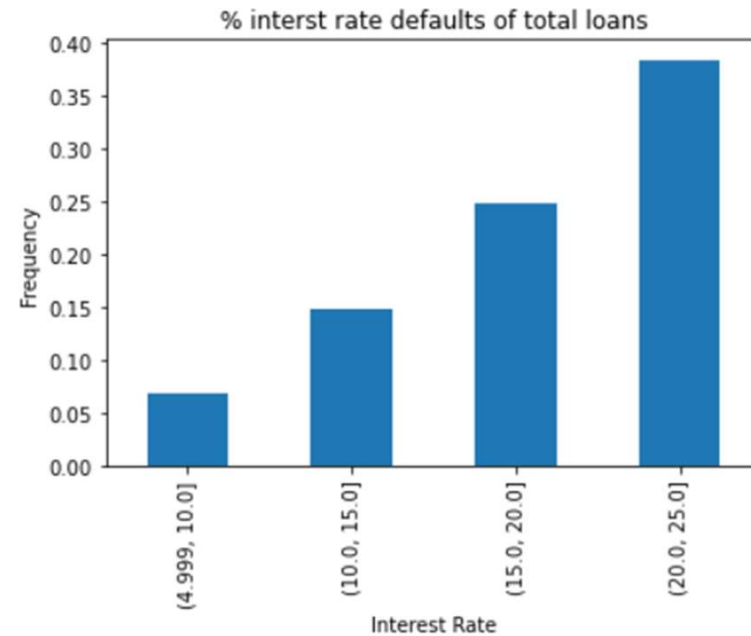
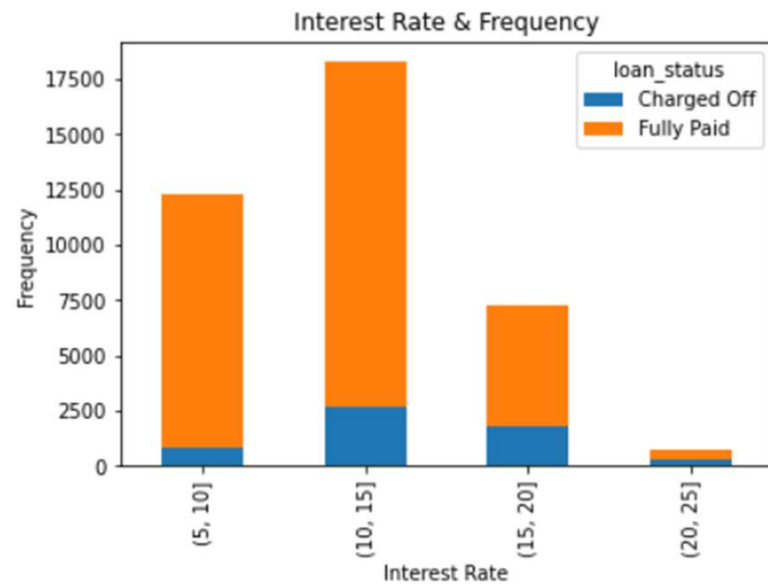
Term



Observation:

- Majority of the loans are issued with 36 months term.
- Significantly high percent of the loan defaults from total loans are with 60 months term.

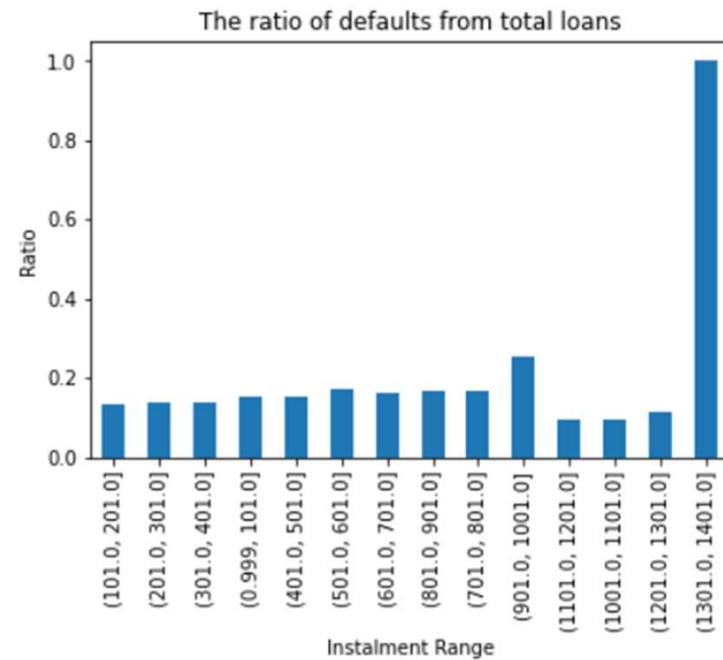
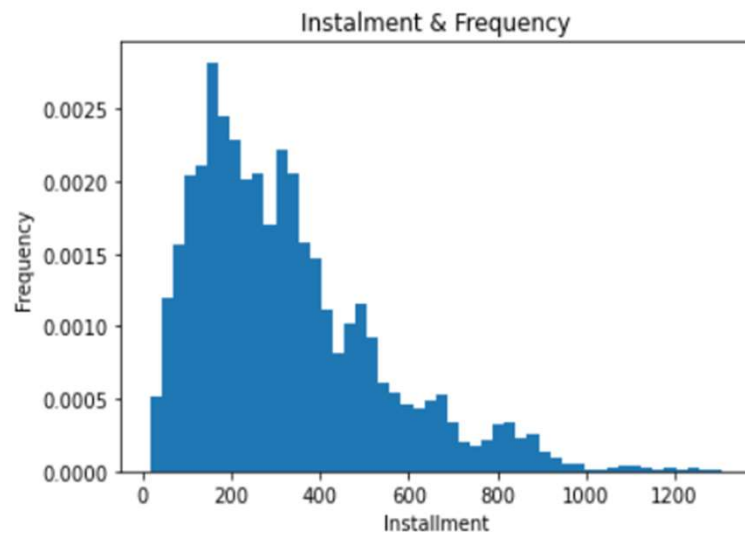
Interest rate



Observation:

- Higher number of loans are issued between interest rate 5% to 15%
- There is gradual increase in defaults when the interest rate increases. Higher the interest rate, higher the percent of defaults.

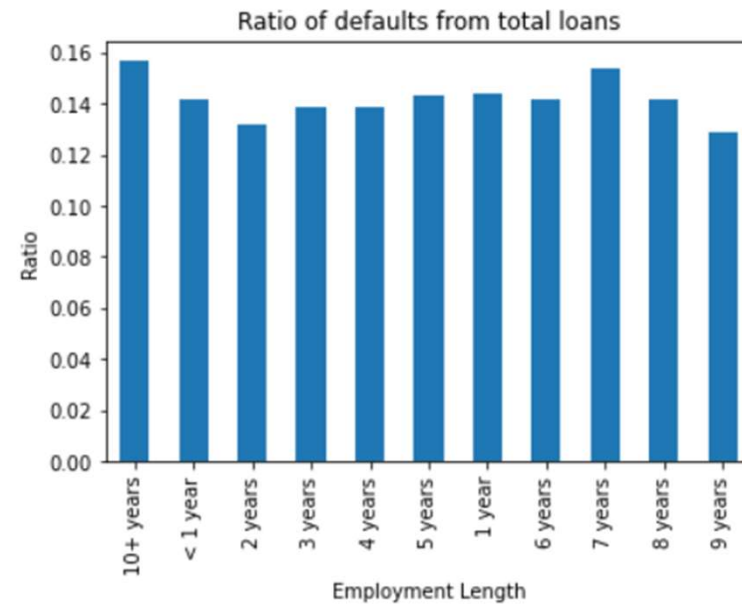
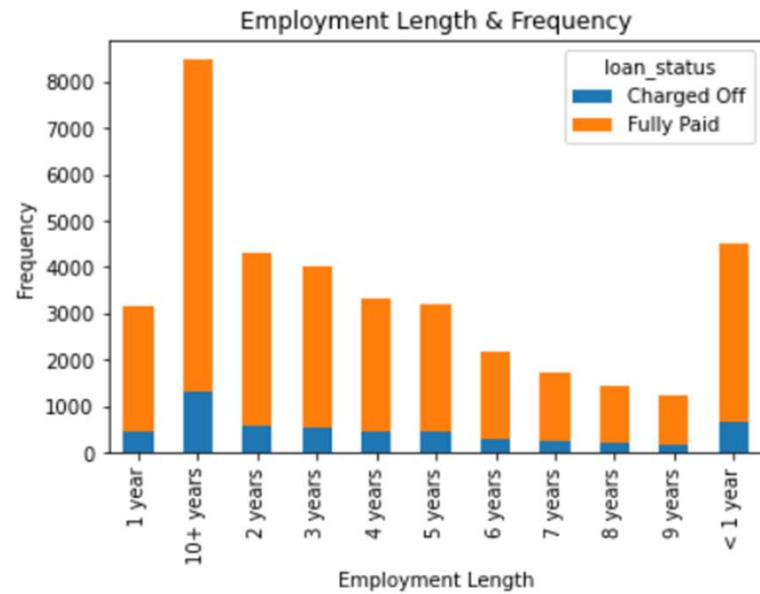
Monthly Instalment



Observation:

- Majority of the instalments are between \$150 and \$350
- Instalment range \$1300 and \$1400 has 100% defaults(total 2 loan)

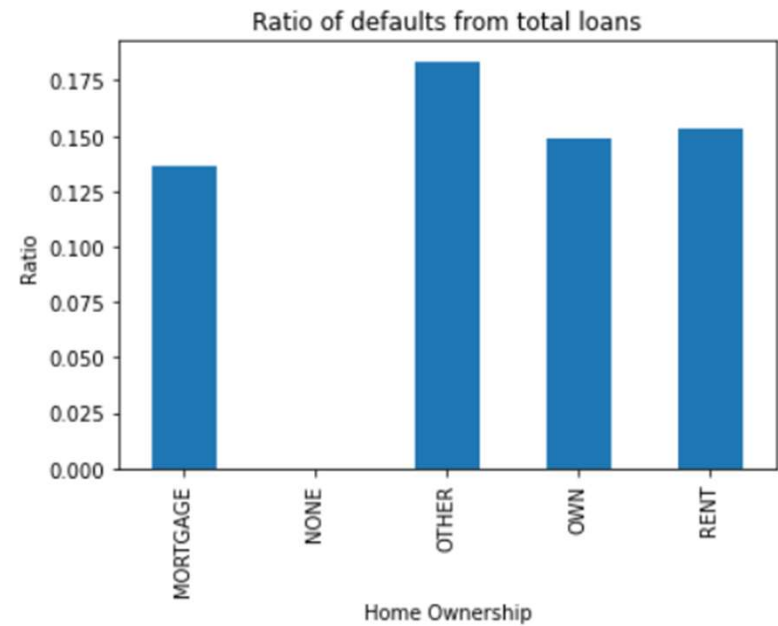
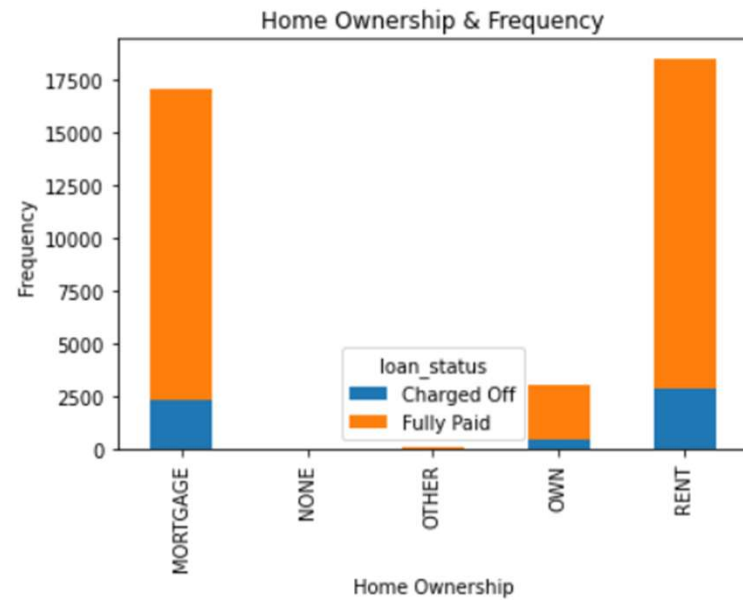
Employment Length



Observation:

- People with experience less 4 years and over 10 years are tend to take loans
- There is no significant difference in the ratio of defaults across difference employment groups

Home Ownership



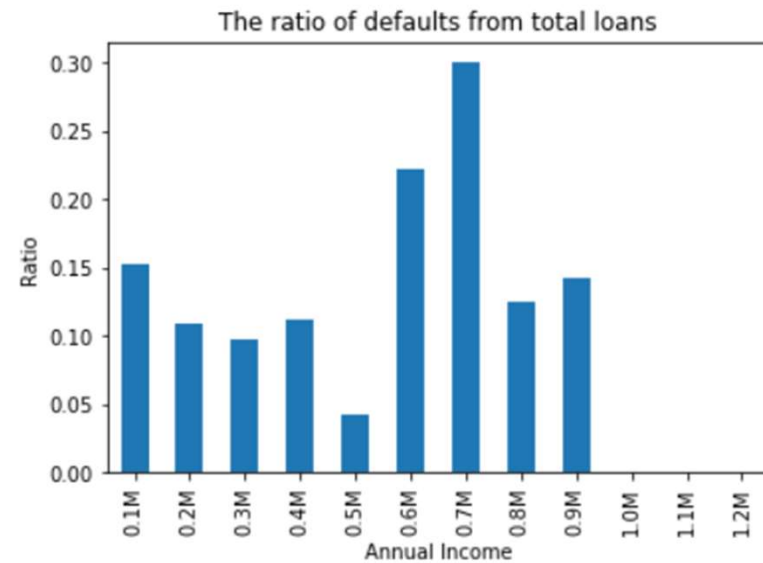
Observation:

- Higher number of loans are from people staying on RENT and MORTGATE
- There is no significant difference in ratio of default from this category

Annual Income

Overall Max Annual income: 6.0M,
Defaulter Max Annual Income: 1.25M

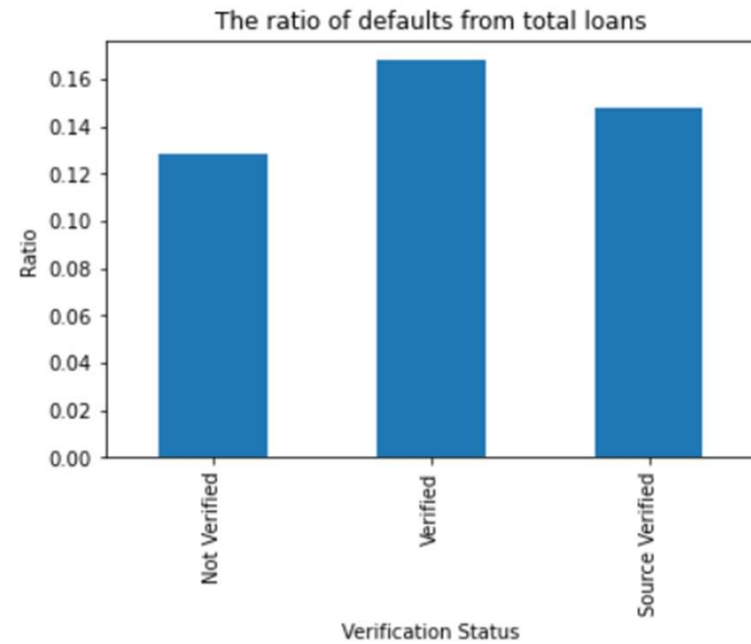
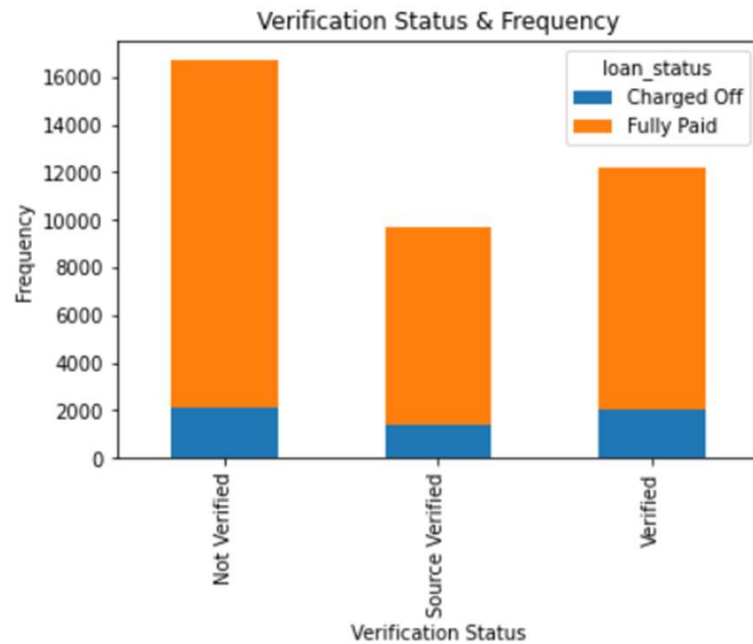
```
count    3.857700e+04  
mean     6.877797e+04  
std      6.421868e+04  
min      4.000000e+03  
25%      4.000000e+04  
50%      5.886800e+04  
75%      8.200000e+04  
80%      9.000000e+04  
max      6.000000e+06  
Name: annual_inc, dtype: float64
```



Observation:

- Most of the loan defaults are with annual income below \$90K. 80% of total loans are from this group.
- Between annual income 1.25M to 6M all loans are fully paid.
- Loan with annual income of 50K to 70K are among the highest defaulters.

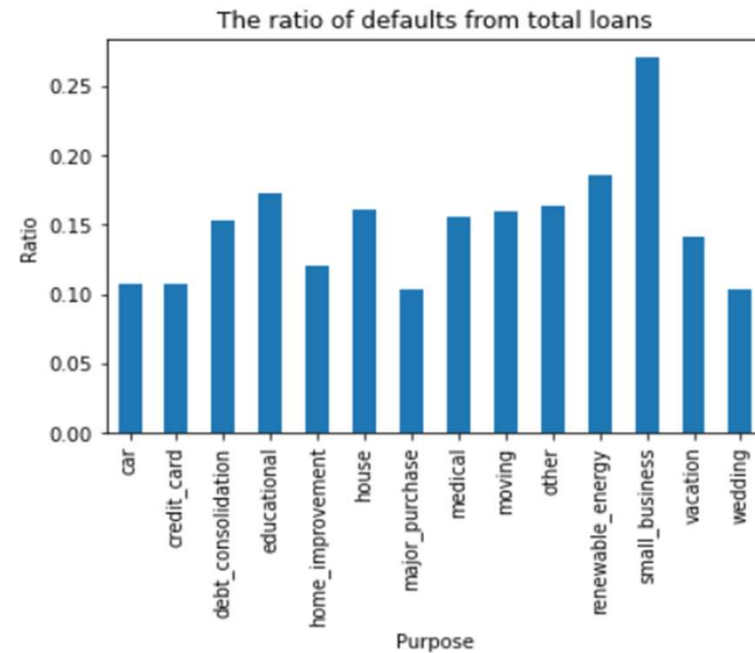
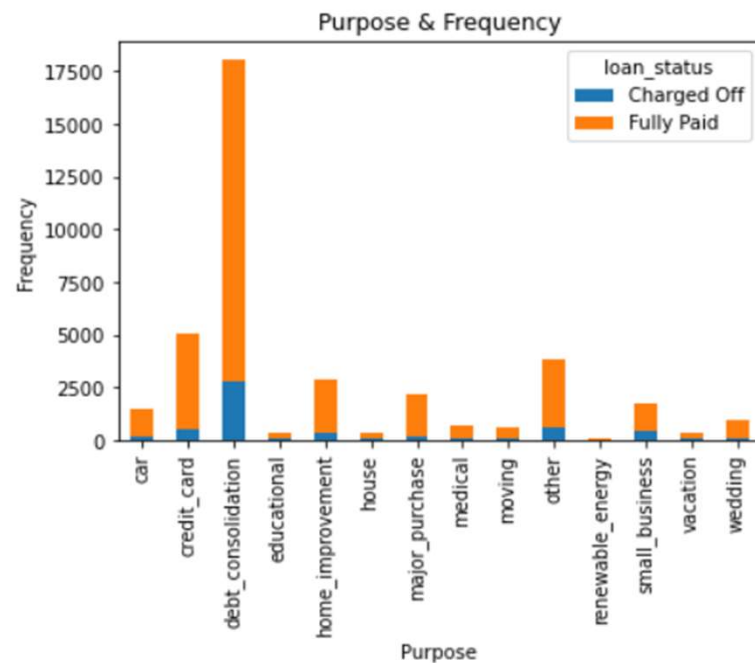
Income Verification



Observation:

- Most of the loan issues are not income verified.
- Verified and Source Verified has slightly higher fraction of defaults.

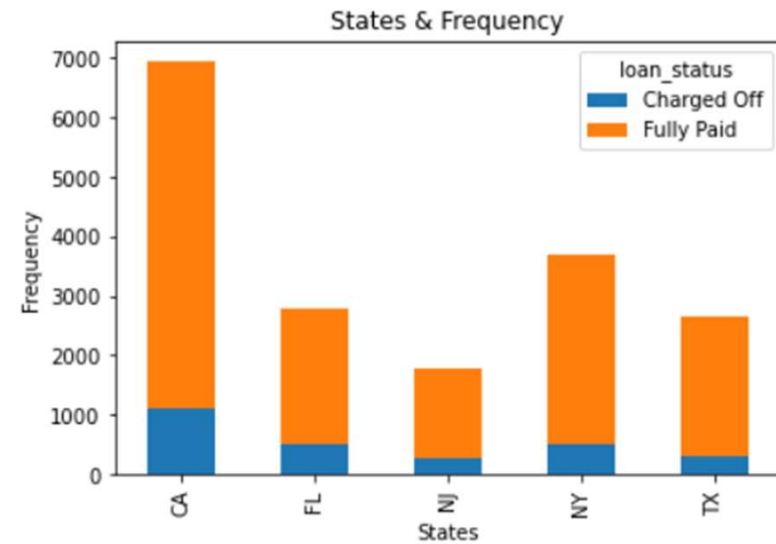
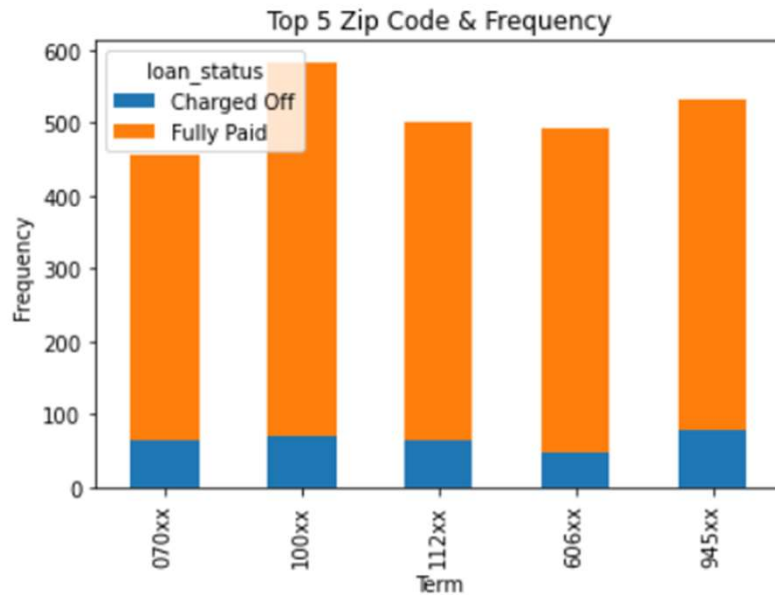
Purpose Of Loan



Observation:

- Nearly half of the entire loan issued are for debt consolidation.
- Debt consolidation is the highest contributor of loan defaults.
- Small Business has highest ratio of defaults

Zip Code and State

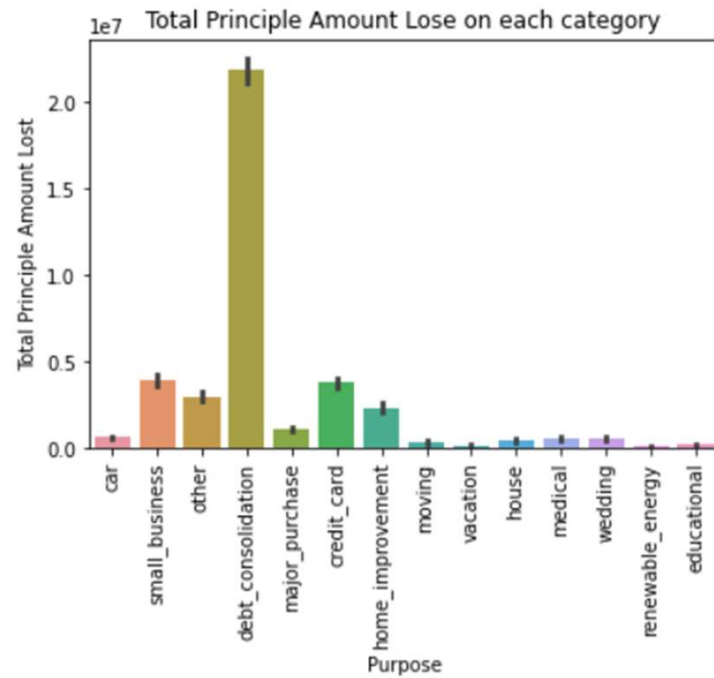


Observation:

- Zip code 100xx is where the maximum loan issued
- Zip code 945xx is where maximum defaults happened among the max loan issued
- Majority of the loans are issued on state CA.
- FL and CA are the top two defaulters from the among all states.

Bivariate Analysis

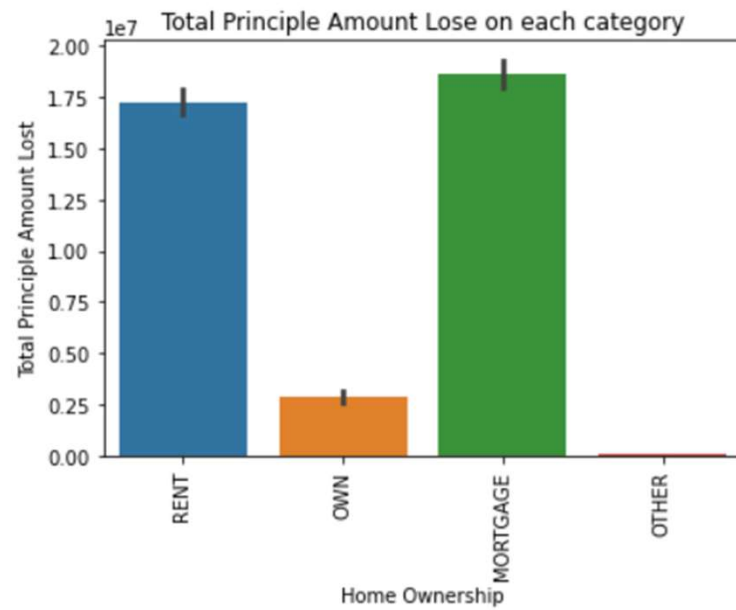
Principle Lost vs Loan Purpose



Observation:

- Loan for Debt Consolidation is the highest lose making business

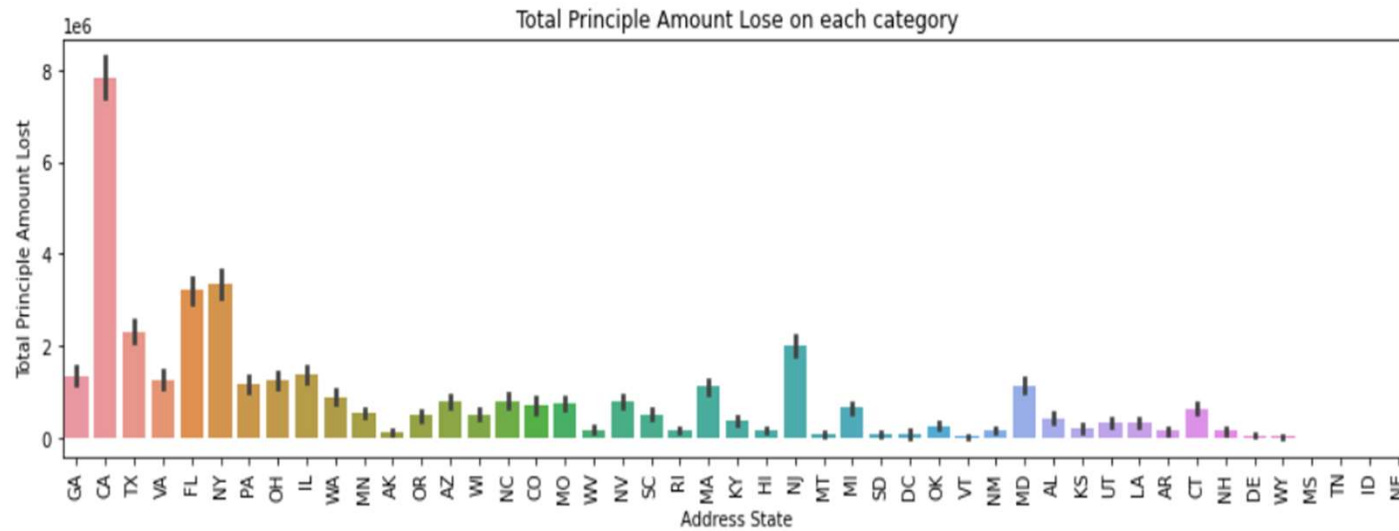
Principle Lost vs Home Ownership



Observation:

- It is safe to give loan with people staying their own homes

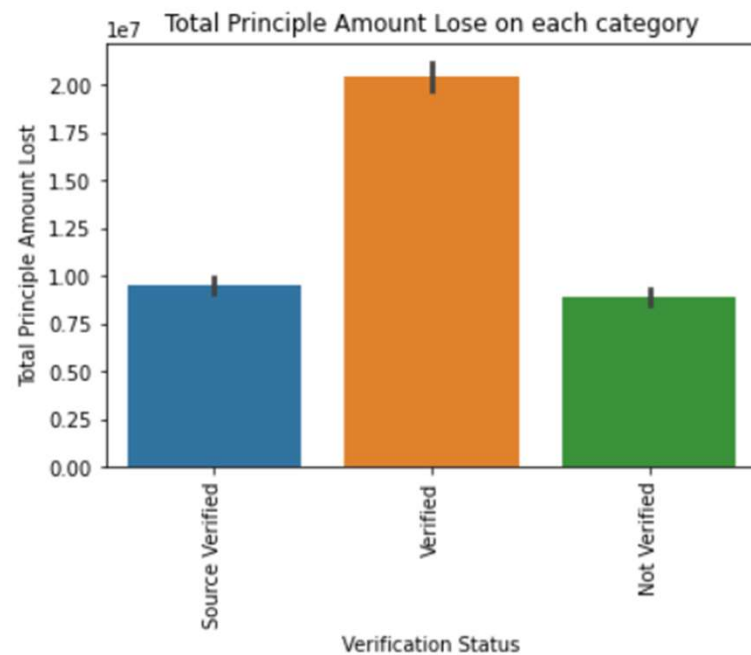
Principle Lost vs Address State



Observation:

- Majority of the money lost from state CA, NY and FL

Principle Lost vs Address State



Observation:

- Income Verification seems not thorough

Recommendations

- Loan issued for debt consolidation purposes seems highly risky as more than 50% of the investor's money was lost from this category.
- Verification Status as verified contributed 50% of the money lost hence needs a review of the verification process.
- There is gradual increase in defaults when the interest rate increases. Higher the interest rate, higher the percent of defaults.
- All loans from people from annual income 1.25M to 6M are fully paid hence it is safe to issue loans to this group of people.
- It is safe to give loans to people staying in their own homes.
- Majority of the money lost from states CA, NY, and FL.
- Most of the loan issues are not income verified.
- Zip code 100xx is where the maximum loan issued.
- Zip code 945xx is where maximum defaults happened among the max loan issued.
- Most of the loan defaults are with annual income below \$90K. 80% of total loans are from this group.
- Loan with annual income of 50K to 70K are among the highest defaulters.
- Significantly high percent of the loan defaults from total loans are with 60 months term.