Lending Club Case Study

Presentation Overview:

- Exploratory Data Analysis (EDA) techniques used to get an understanding of risk Associated with losing money while lending to the customers.

Metadata Description:

- Dataset has been shared by the team to perform the Exploratory Data Analysis.
- Metadata Description has been shared and taken reference from Data Dictionary.

Analysis Method:

- Risk analysis conducted through Exploratory Data Analysis (EDA) using Google Colab (Python).

Task Owners:
Ramandeep Mehra
Pullam Chetti Divakar

Objective

- Objective of this case study is implement Exploratory Data Analysis (EDA) techniques on loan data from a consumer finance company to get an understanding of risk analytics in banking and financial services and also to understand how data is used to minimize the risk of losing money while lending to the customers.

- To identify variables which are strong indicators of default and potentially use the insights in approval / rejection decision making.

- Business objective of this case study is to help company understand the driving factors behind loan default. The company wants to utilize this knowledge for its portfolio and risk assessment.

- Identification of risky loan applicants using EDA is the aim of this case study.

Business Understanding

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:

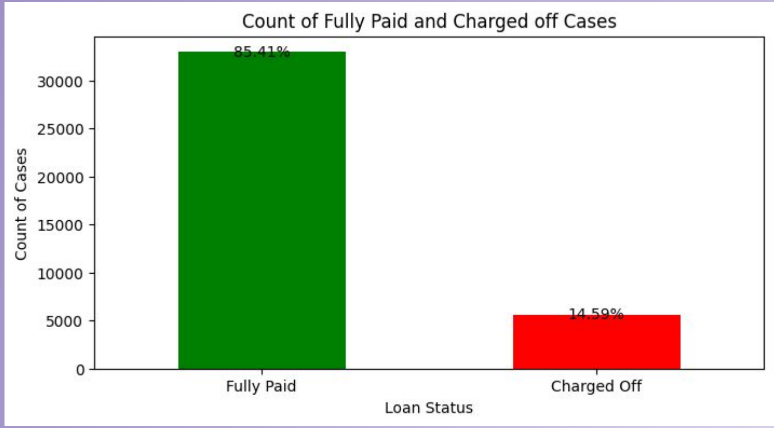
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

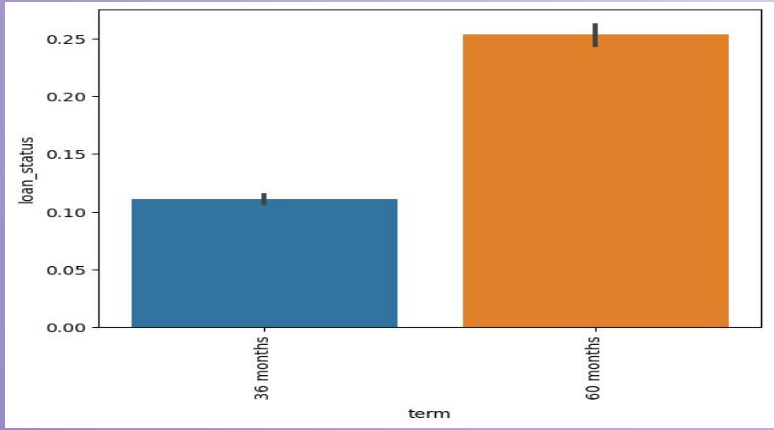
In this case study, we use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Data Cleanup and Preparation Steps/Process

- Importing relevant python libraries
- Importing the dataset
- Checking information of the dataset
- Removing columns with null values
- Removing columns with more than 50% values as unique
- Identifying & Removing irrelevant columns
- Fixing missing values and correcting data types
- Generating new columns from existing columns
- Checking and removing Outliers
- Performing EDA on prepared dataset

Loan Status & Term of Loan

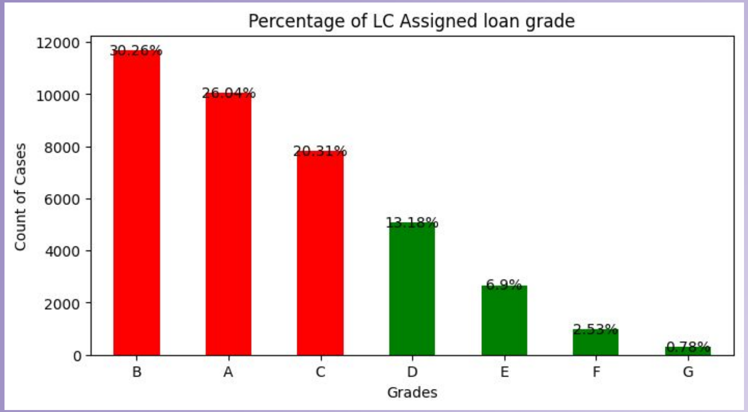


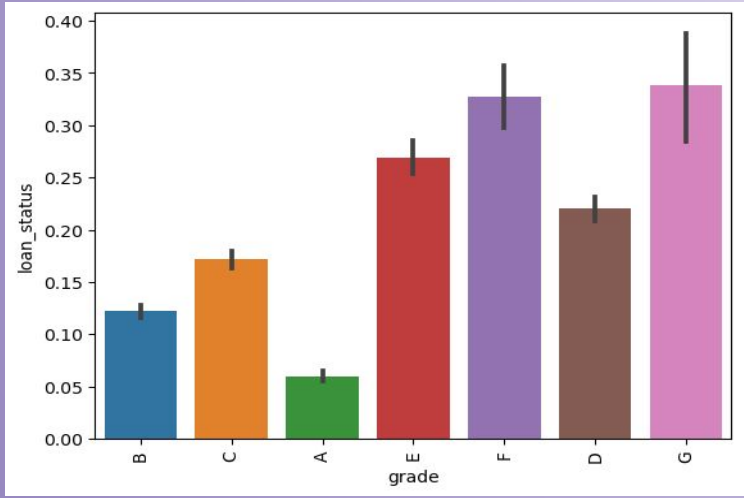


Observation:

- Defaulted Loans for Charged Off (14.6%) is much small as compared to the Fully Paid (85.4%) loans.
- Majority of the loans are approved for 36 months period.
- Default rate is high for 60 months of term.
- 2.15% Charged Off variance has been observed in 36 months and 60 months loans approval cycle. Though tenurity is also impacting the charged off as higher the term, higher the chances for default.

Loan Distribution against grades





Observation:

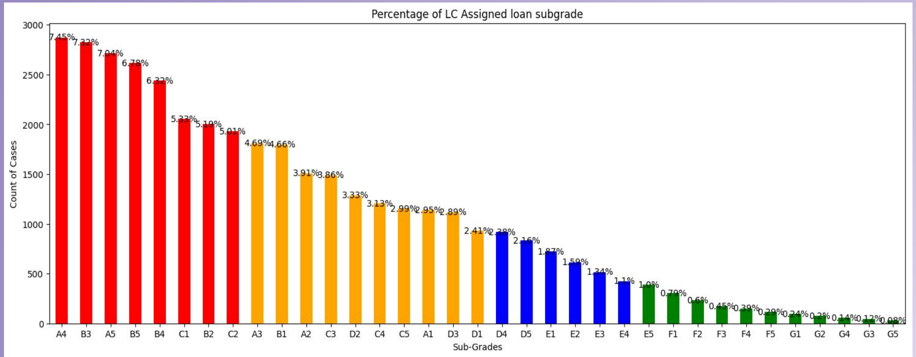
Grades:

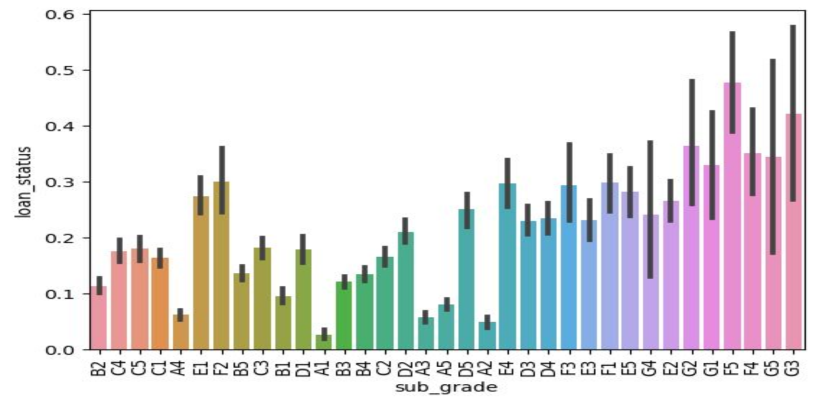
- A large amount of loans are with category **B, A and C** grade has contribution percentage of **76.61%** as compared to other grades, concluding that most loans are approved for high grade loans and contributing in overall loan.

Default Rate Observation:

- Higher default rate has been observed in category of grades **G, F, E and D.**

Loan Distribution against sub-grades





Observation:

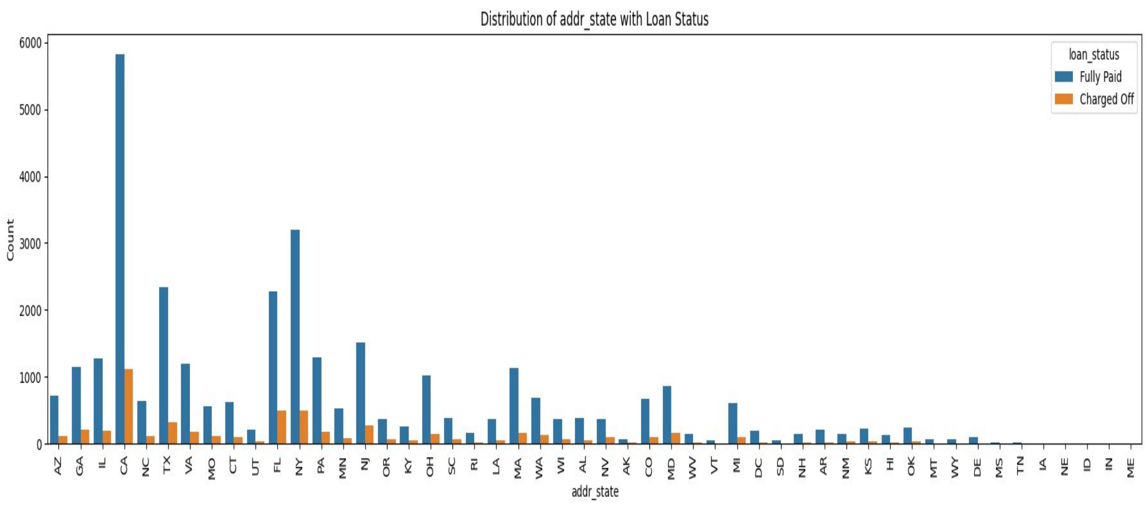
Grades:

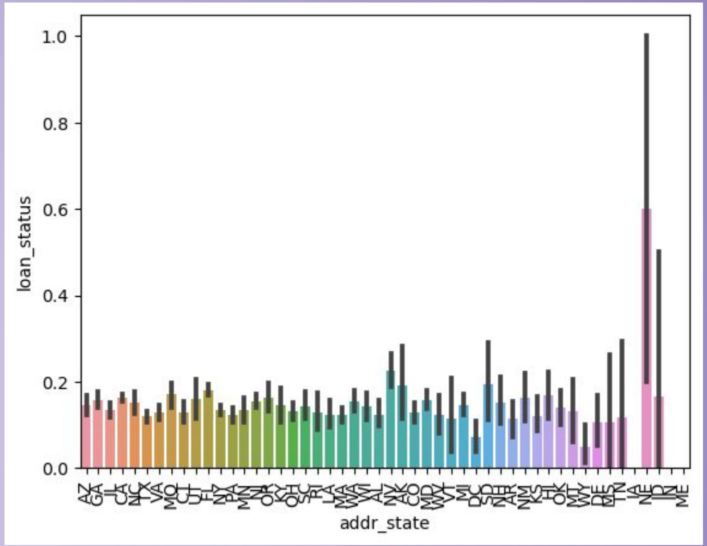
- A large amount of loans are with category **B, A and C sub - grade** has contribution percentage of **76.61%** as compared to other grades, concluding that most loans are high grade loans and contributing in overall loan.

Sub-Grades default rate:

- High default rate sub_grades has been observed in the chart are **F, G and E**.

State wise Loan Distribution

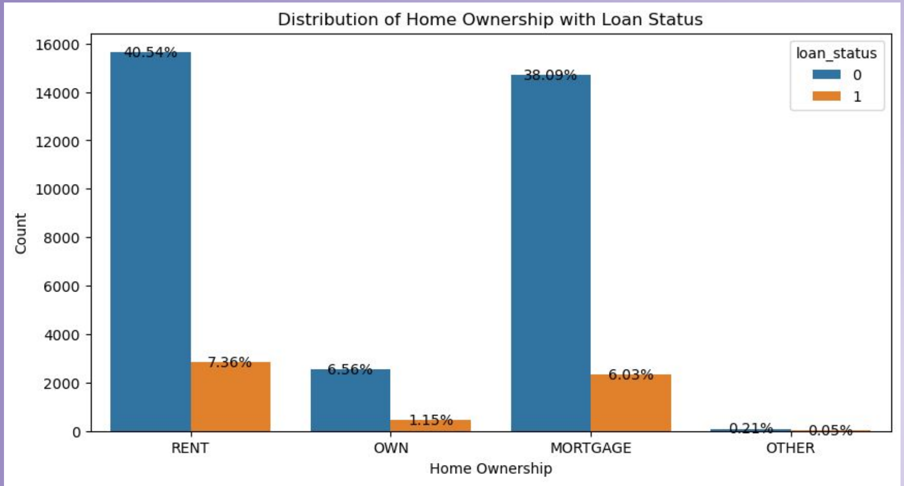


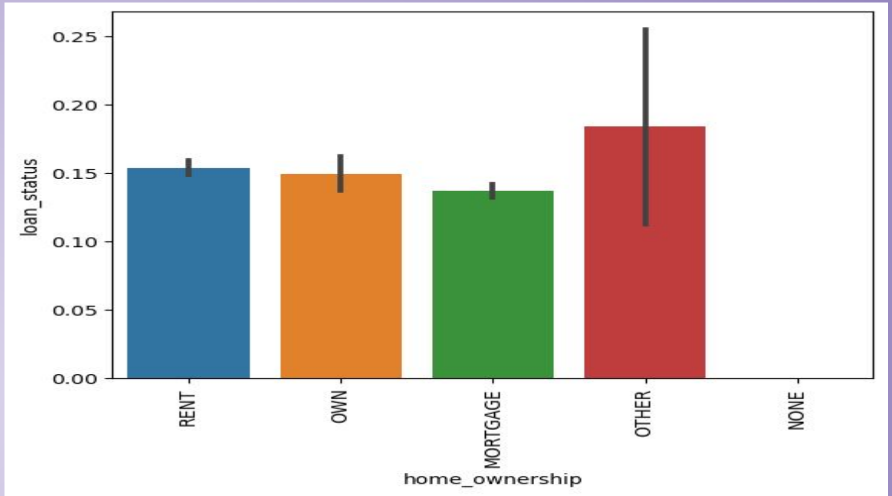


Observation:

- Looking at the distribution of loans along with states, we can conclude that majority of the borrowers are from big urban states like **California, New York, Florida, Texas.**
- There are majority of the borrowers are for charged Off category are also from big urban states like **California, New York, Florida, Texas**.
- There are high default rate observed in NE, NV, SD and ID state.

Home Ownership vs Loan Status





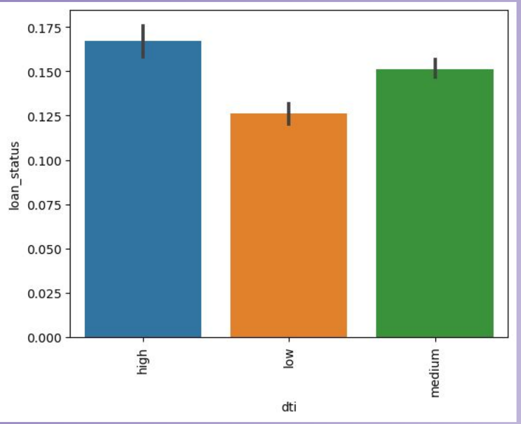
Observation:

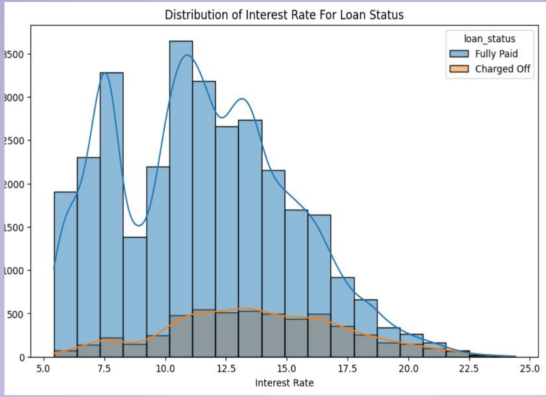
- **92%** borrowers does not possess any property and are on either "**RENT"** or **"MORTGAGE"**.
- **13.39%** Borrowers are in Charged Off category are also from **“RENT”** and **“MORTGAGE”**.
- The % of loan defaulters are more if the borrower is already having an ongoing **"MORTGAGE".**

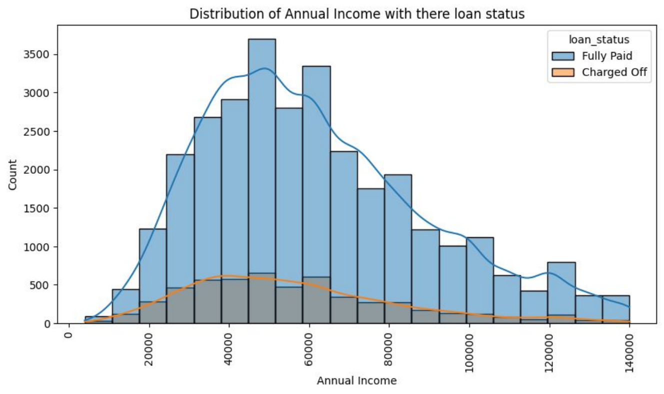
High Default Rate Observation:

- In **Other** Category the default rate is high and followed by **RENT** and **OWN** category.

dti ratio/Interest Rate and Annual Income vs Loan Status



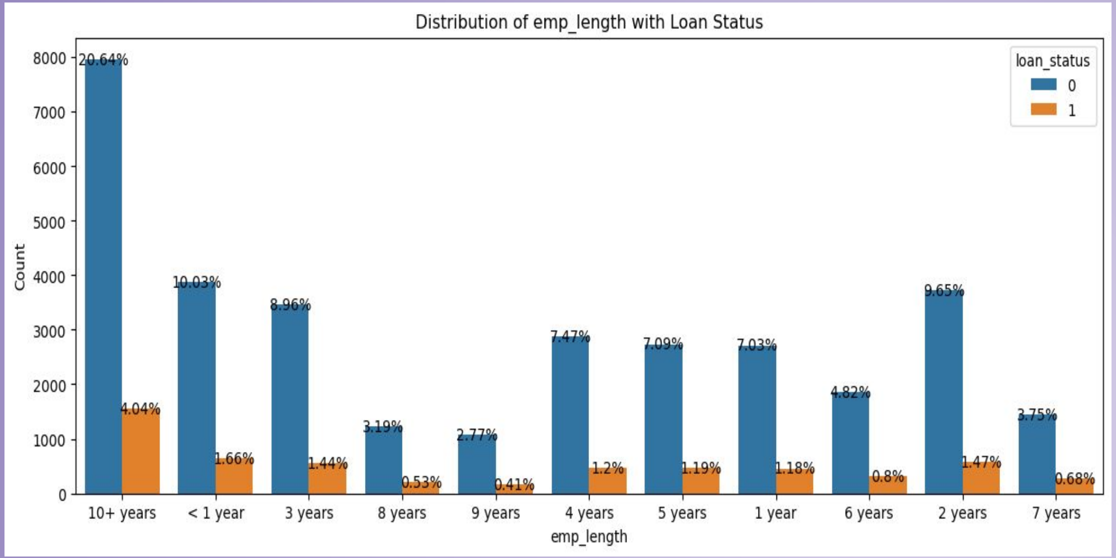


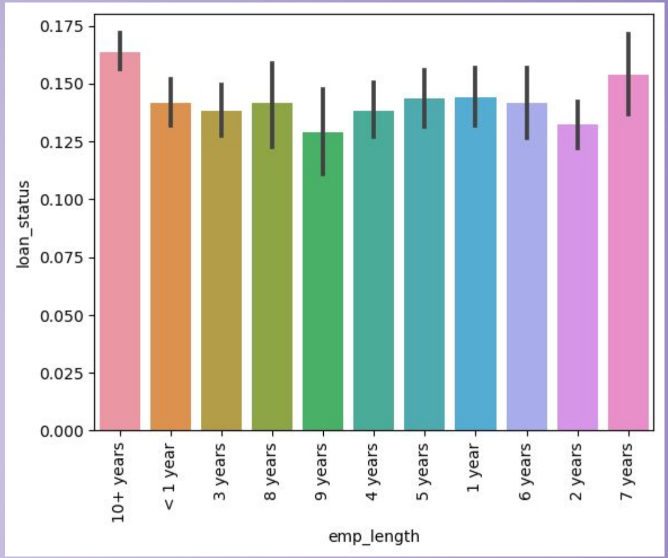


Observation:

- Looking at the distribution we can conclude that for majority of the borrowers, the debt to income ratio is very high.
- The distribution for Annual Income looks like a normal distribution but it is left skewed. So we can say that most of the borrowers have low annual income.
- Looking at the right hand chart: We can conclude that the borrowers with less than 80000 as the annual income are most likely to default the loan.
- **10, 12.5, 15, 17.5 and 20** interest rate are on higher side for charged Off category, which is stating that higher the interest rate chances of default are high.

Applicants Employment Length and Loan Status

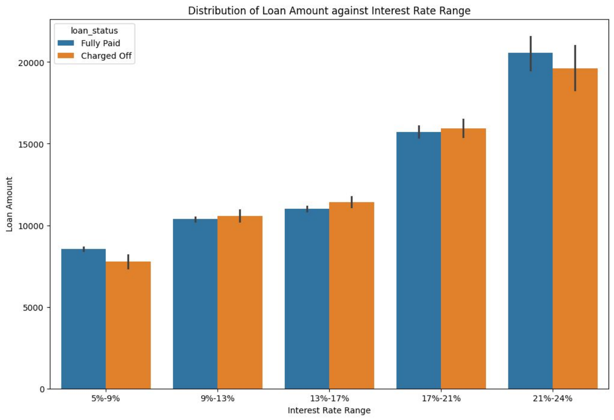


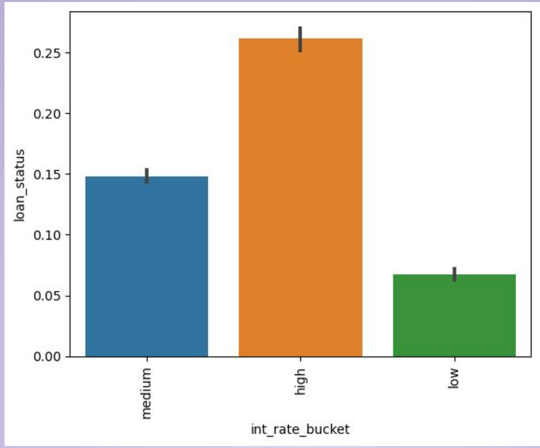


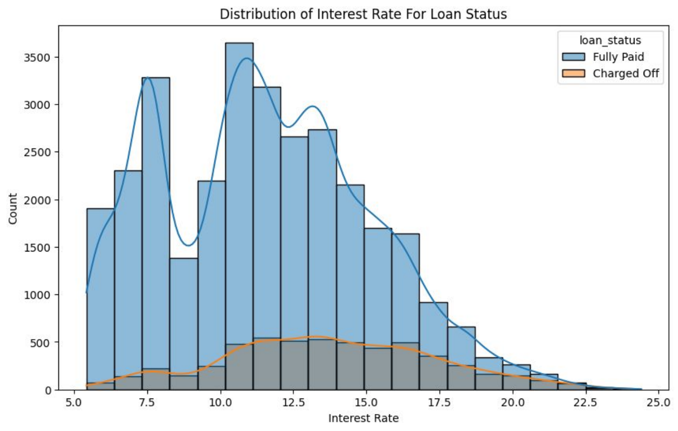
Observation:

- 24.68% data contribution os for 10+ years, and higher number of % for default rate is also higher for 10+ years, 7 year with second highest default rate and 1 Years.
- Most of the borrowers are having experience of 10+ years
- For Applicants with employment length of 10+ years, chances of defaulting at the loan payment is more when loan amount is more than 12k
- Most of the defaulters are having experience of 10+ years and as the emp length is greater than 5 the default rate is high.

Interest Rate Range and Loan Amount Distribution



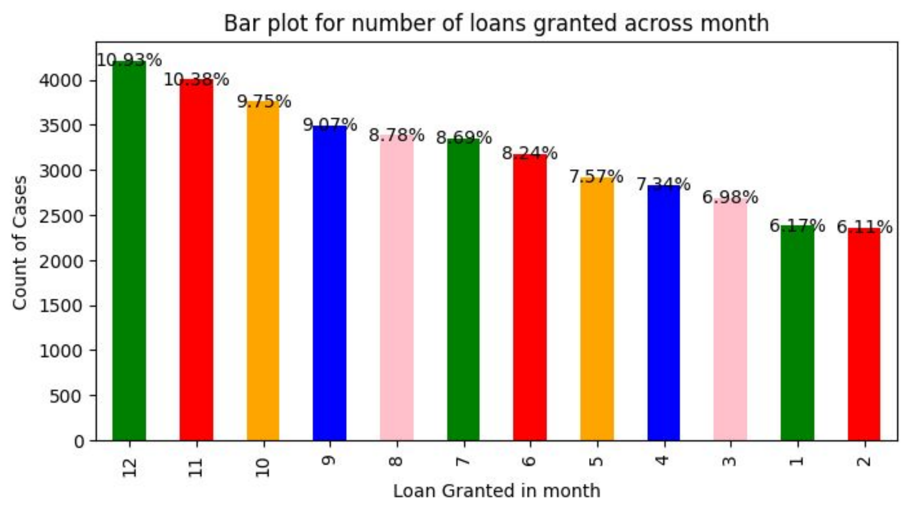


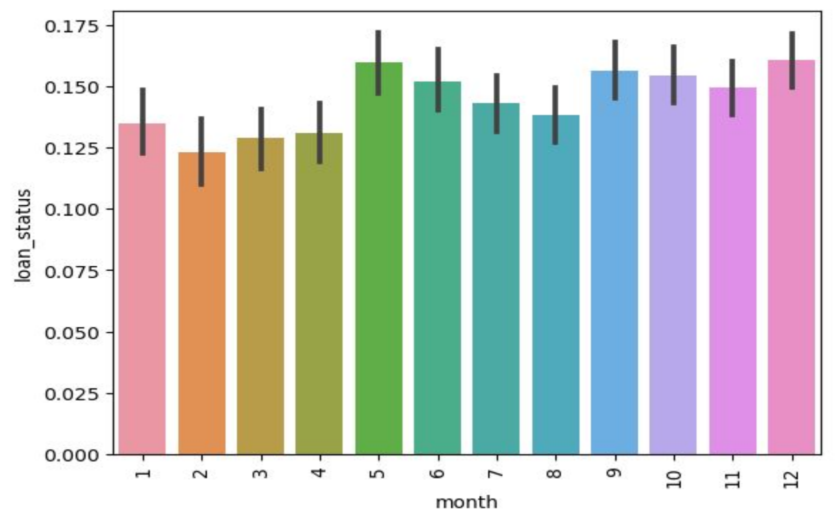


Observation:

- For most of the loans approved, the interest rate is between 9%-13% followed by 13%-17% and 5%-9%.
- Interest Rate increases as the value of loan amount increases.
- For the borrowers where the interest rate is high the defaulters % is more as compared to borrowers with other interest rate percentage
- As we can see from the graph that higher the interest rate higher the chances of default.
- If the interest rate is low, the default rate is also low.

Loan Trends – Month of the year



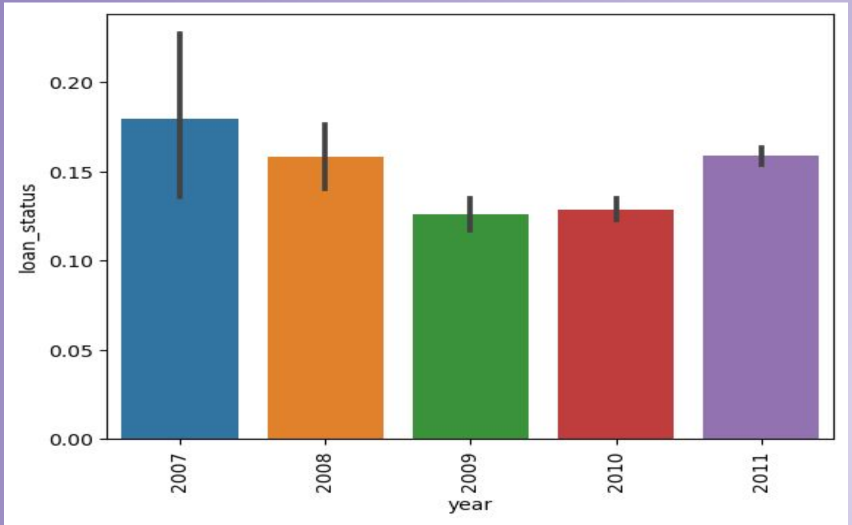


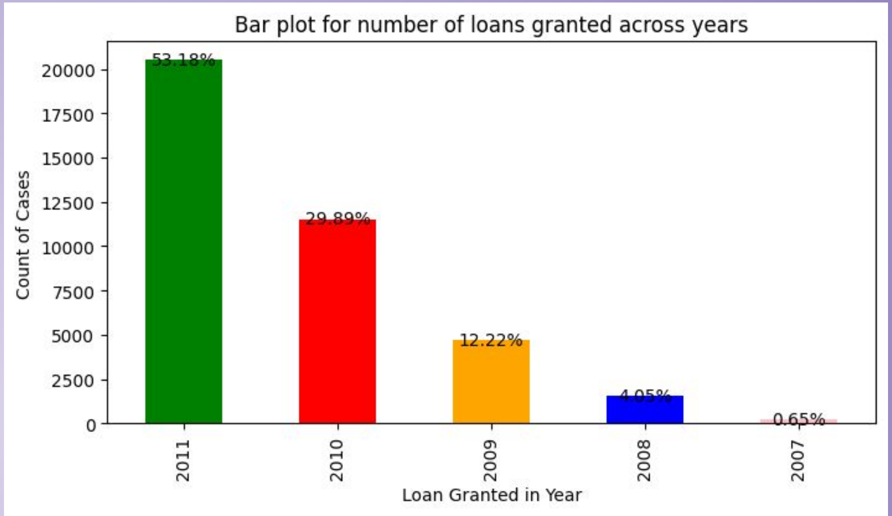
Observation:

- Looking at the distribution we can conclude that majority of the loans are processed in the last quarter of the year
- As the loans processed in last quarter of the year are more, the defaulters are also more from the last quarter of the year.
- The default rate is high for the below mentioned month

1. Dec month.
2. May month.
3. September month.
4. October month.
5. June month.

Loan Trends – Yearly





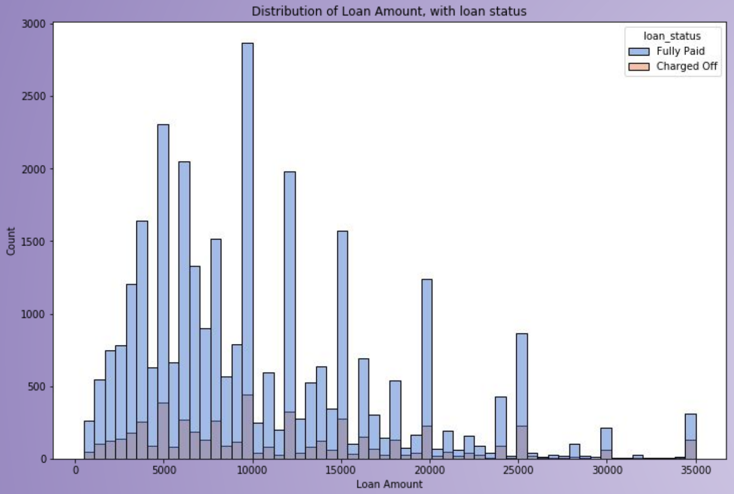
Observation:

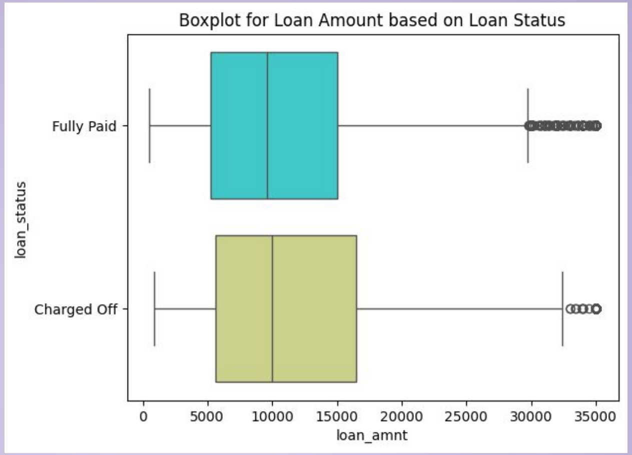
- Looking at the distribution we can conclude that the loan approval numbers increases positively year on year and the increase with time is exponential
- With the increase in number of loans approvals, the loan defaulters have also increased positively year on year
- 78% increment in applicants in 2011 compared to 2010.
- 145% increment in applicants in 2010 compared to 2009.

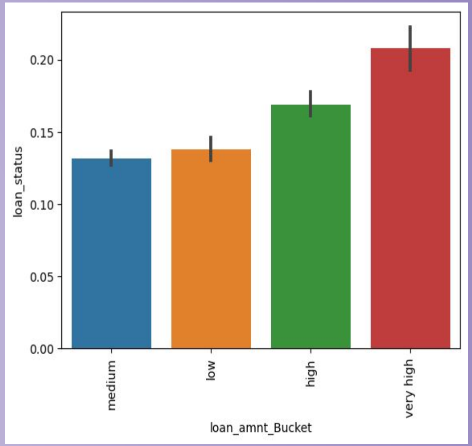
Default Rate:

- Default rate is high for the year of 2007, 2008.
- Default rate decreased in 2009 and 2010, later in the year of 2011 the default rate has increased.

Loan Amount against Loan Status







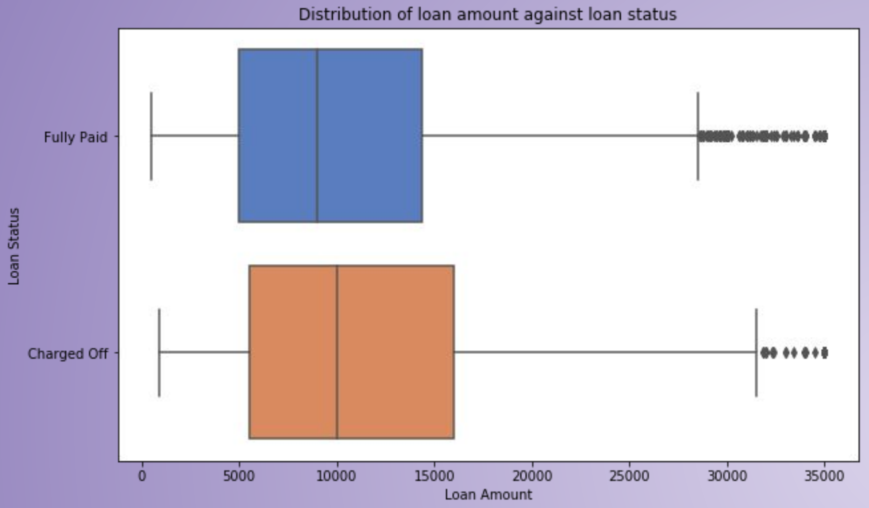
Observation:

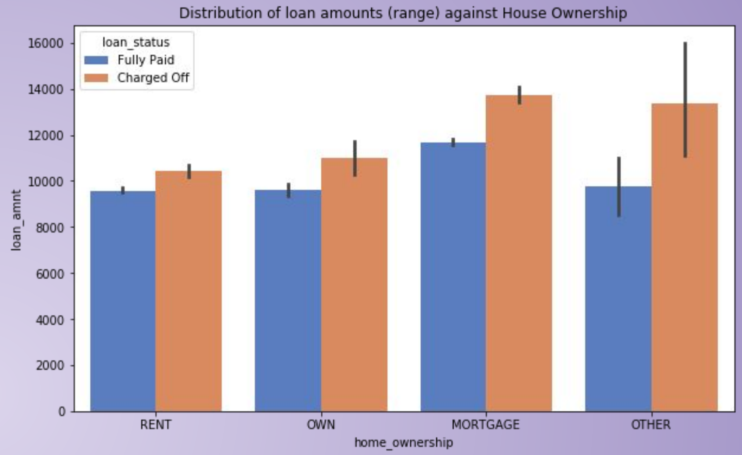
- As we can see in the box plot, higher the loan amount for charged off; higher the chances of default.
- Defaulter percentage is more when the loan amount is more than 10k.

Default Rate observation:

- Default rate is high for the higher loan amount and if the loan amount is medium than the default rate is on lower side.

Loan Amount against Loan Status & Home_ownership





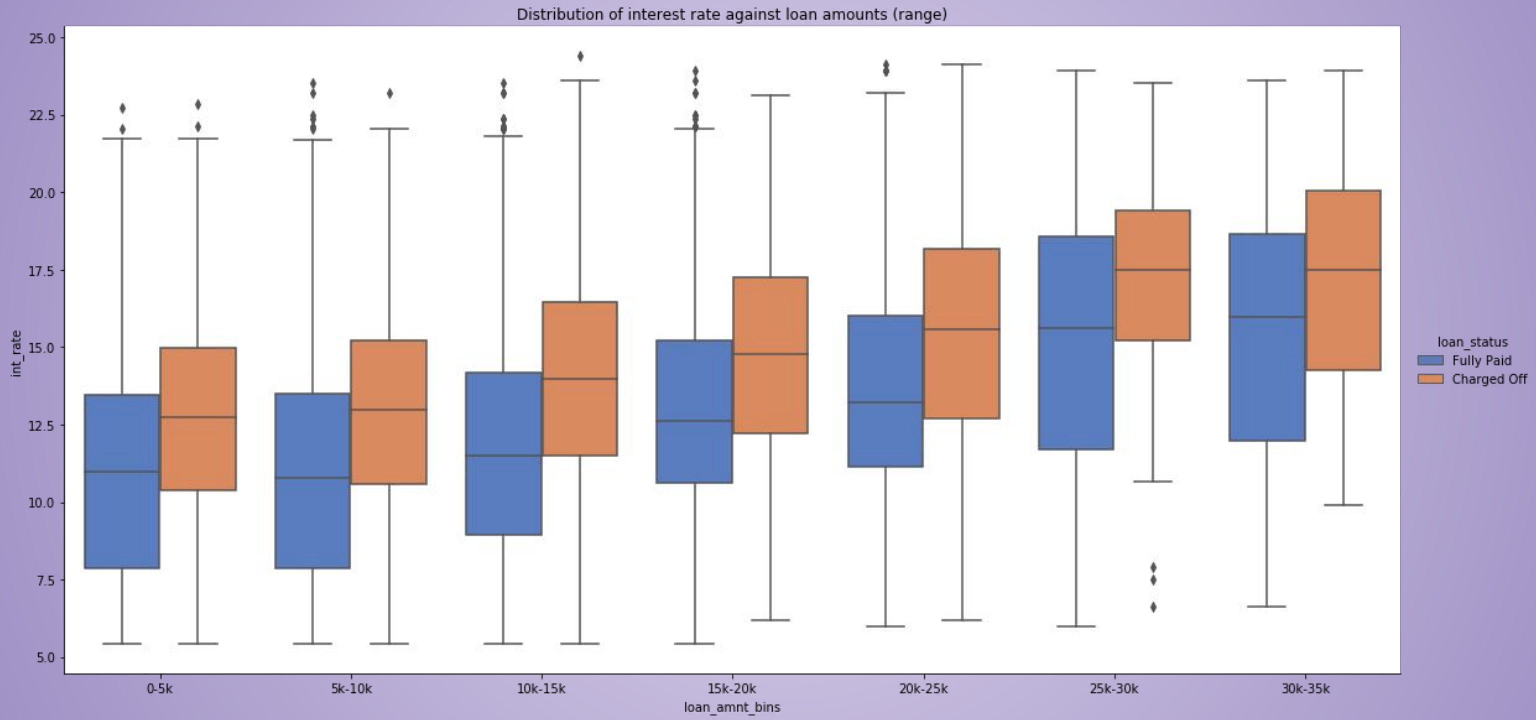
Observation:

- The median and Quartile deviation is on higher side for higher loan amount for Charged Off category.
- We see a larger 75% quantile in "Charged Off" loans - indicating large amount of loans have high chance of defaulting.

Home Ownership:

- Applicants with an existing OTHER have higher rate for defaulting at the loan payments.

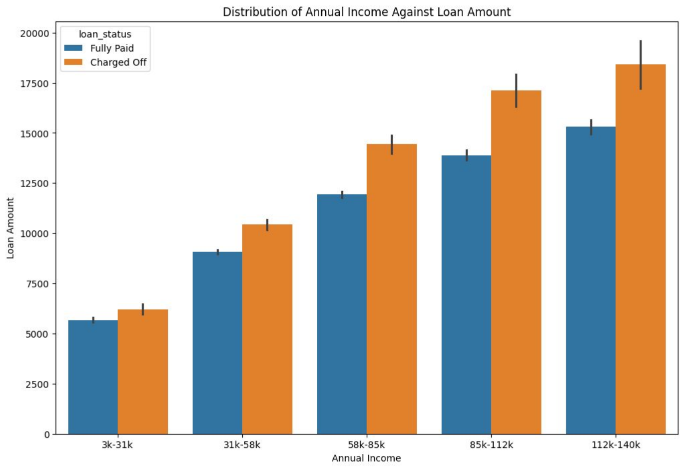
Interest Rate Distribution for loans

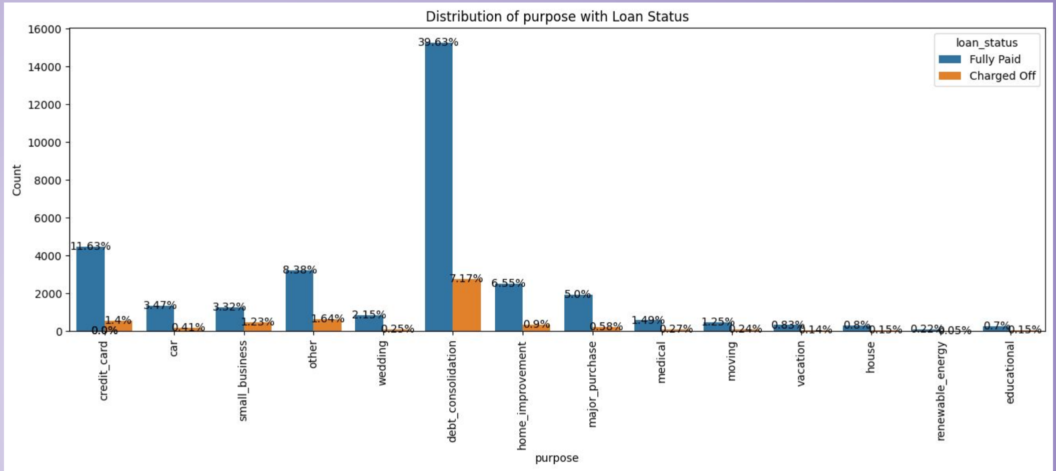


Observation:

- Interest rate for charged off loans is pretty high as compared to fully paid loans
- This can be a very strong factor for loan defaulting

Annual Income against Loan Purpose



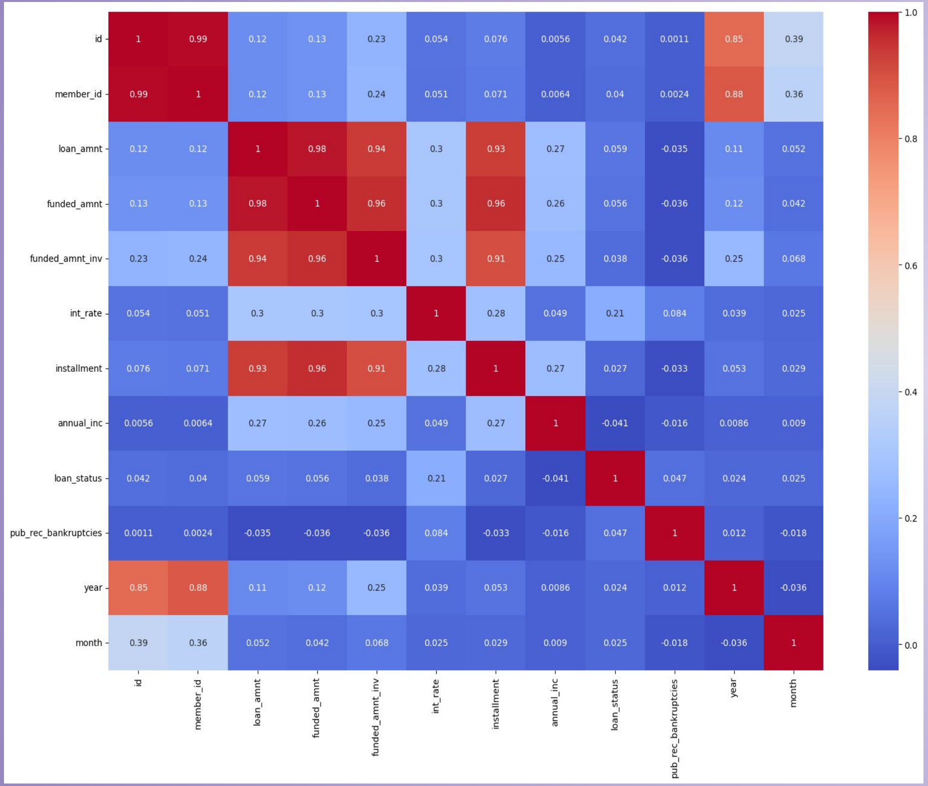


Observation:

- Higher the annual income higher the loan amount.
- The individuals with higher salary mostly applied loans for home improvement, small business and against credit cards
- We also see the highest number of loan defaulters against the debt consolidation, small business, credit card and home improvement category

1. Debt consolidation: 46.8% and high default rate 7.17%.
2. Credit Card: 13.03% and second highest default rate 1.4%.
3. Small Business: 13.03% and second highest default rate 1.4%.

Annual Income against Loan Purpose



Observation:

1. In correlation metrics, there is slight positive correlation in Interest rate and Loan Status, which is indicating that the higher the interest rate higher the chances of default.

2. Interest Rate is slightly positively correlated with Loan Amount, Funded Amount, Funded amount INV and installments. Which showing the relationship is the number of installments are high than the interest rate will be high.

3. There is high positive correlation of Loan Amount with Installments, Funded Amount and Funded amount INV

Impact Analysis

Following are few observations with respect to the Charged Off loans.

There is more chances of defaulting when:

- Applicants who use the loan to clear other debts, start a small business, clear off credit card bills or home improvement.
- Applicants who receive interest at the rate of 13-17%.
- Applicants where the employee length is 10+ years.
- Applicants with income less than 50000.
- Applicants for whom the dti is high.
- The Applicant is already having an ongoing "MORTGAGE"
- The loan amount is applied for a larger value (>10000)

Few more observations:

- The chances of defaulting increases for applicants:
  - Whose home ownership is "Mortgage" and loan value is in range of 14k - 16k
  - Employment length is 10+ years and loan amount is in range of 12k - 14k
  - When the interest is changed in the range of 15-17% and loan is in the range of 30k - 35k
- Interest rate for charged off loans is pretty high as compared to fully paid loans. This can be a strong factor for loan defaulting.

Recommendations

1. Verification and Default Rates:
   - 43.27% of loans were not verified, with 5.55% resulting in defaults.
   - Suggests a need for more thorough verification processes, especially for 10+ years term loans.
2. Impact of default Year 2007 - 2011:
   - The introduction of credit card loans in 2011 led to a higher percentage of loans granted.
   - Default rates increased, emphasizing the importance of caution when granting loans, especially for 36 months.
3. Home Ownership and Default Rates:
   - Majority of loans approved for RENT and MORTGAGE home ownership.
   - High default rates in these categories highlight in OTHER and RENT need for careful consideration when granting loans to these homeownership types.
4. Interest Rates and Default Risk:
   - Higher interest rates are associated with an increased risk of default.
   - Recommends thorough checks and verification before granting loans with higher interest rates.
5. Loan Amount and Default Rates:
   - Higher loan amounts correlate with higher default rates.
   - Observations show a higher Charged Off rate for larger loans, indicating a potential risk factor.

Post analysis suggests the importance of careful scrutiny and verification processes before approving loans, especially for longer terms, higher interest rates, and larger loan amounts. It also emphasizes the need for caution when dealing with certain categories of home ownership due to higher default rates.