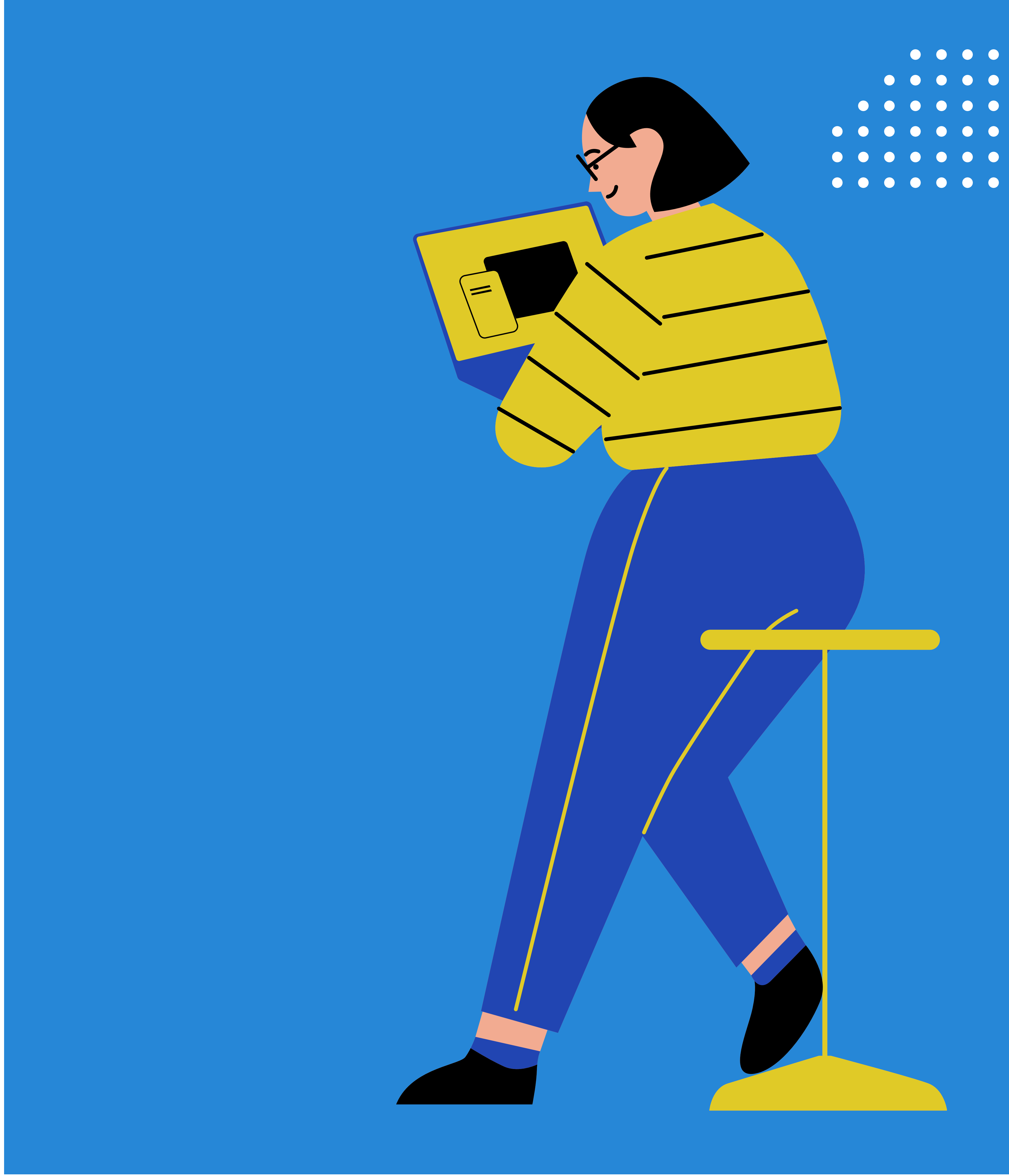
LENDING CLUB CASE STUDY

Exploratory Data Analysis

Team
- Kushagra Mittal
- Aviraj Khare

Data Analysis



01 - Problem Statement

This company 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.

In this case study, develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Data

Analysis

02 - Objective & Approach

Identify the "risky loan applicants," which can help minimize loans and, in turn, reduce credit loss. This case study aims to pinpoint such applicants using Exploratory Data Analysis (EDA).

To achieve the objective of this case study, we will utilize data visualization tools such as histograms, scatter plots, heatmaps, and clustering techniques as part of our Exploratory Data Analysis approach.

Data

Analysis

03 - Data Cleaning

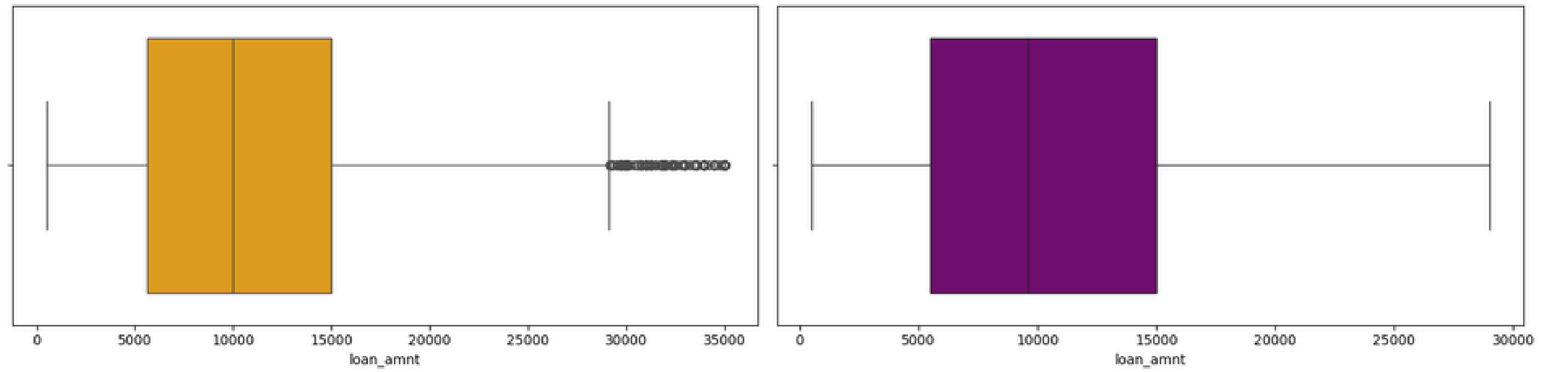
- ## Missing Value Treatment
  - Removing Columns which have more than 64% missing values.
  - Columns which have very less missing values those rows were drop.
  - Columns like “pub_rec_bankruptcies” having very less missing values are imputated with median value.

- ## Outliers Treatment
  - Columns like ‘loan_amnt’ and ‘int_rate’ outliers are removed using IQR (P25,P75) Approach.

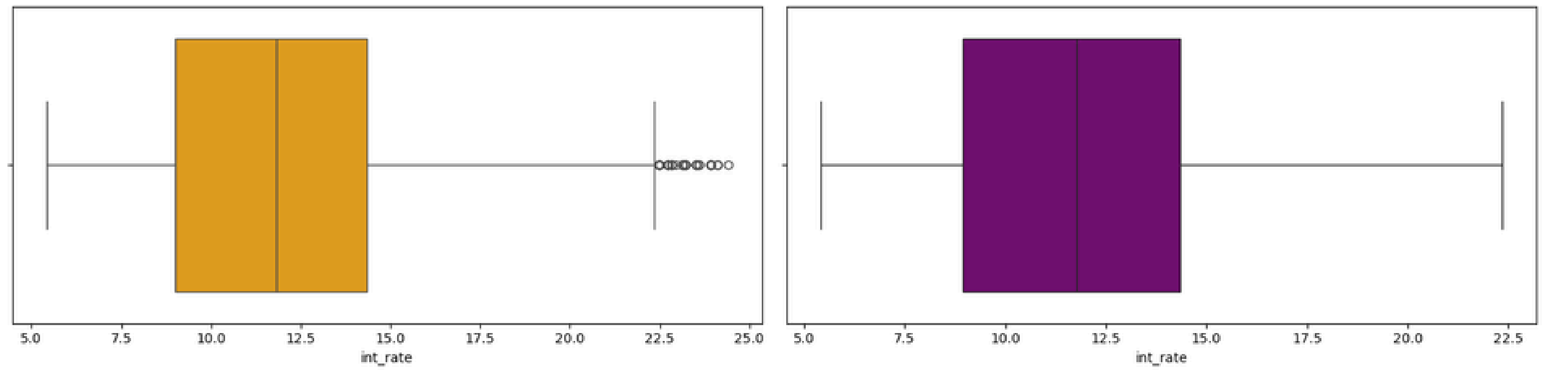
- ## Standardizing Values
  - Removing % from float values.
  - Converting values into DateTime

Before & After Removing Outliers

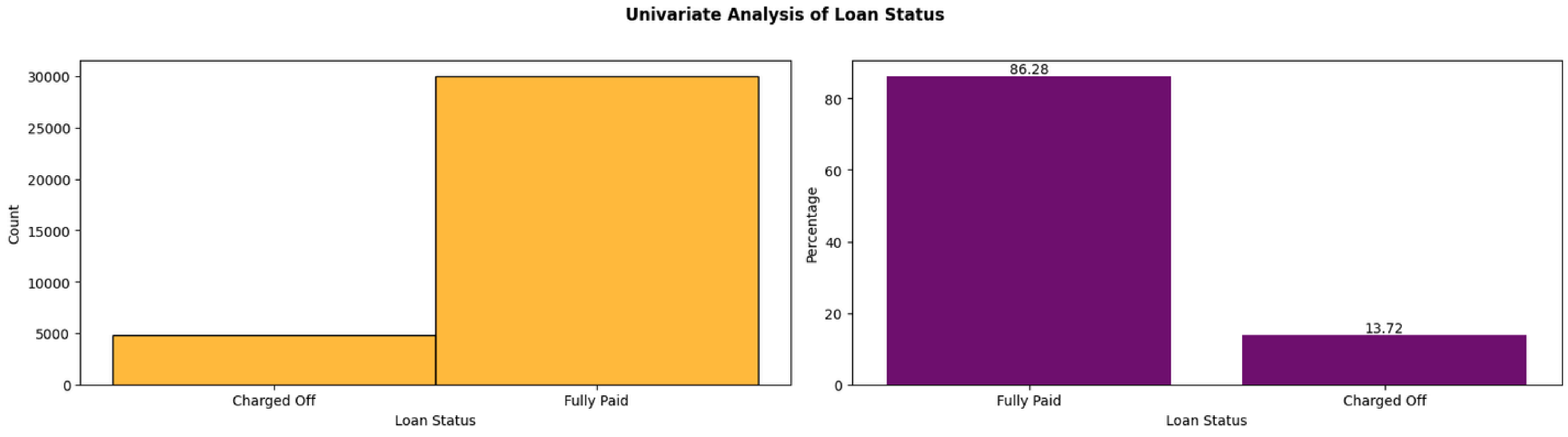
Loan Amount



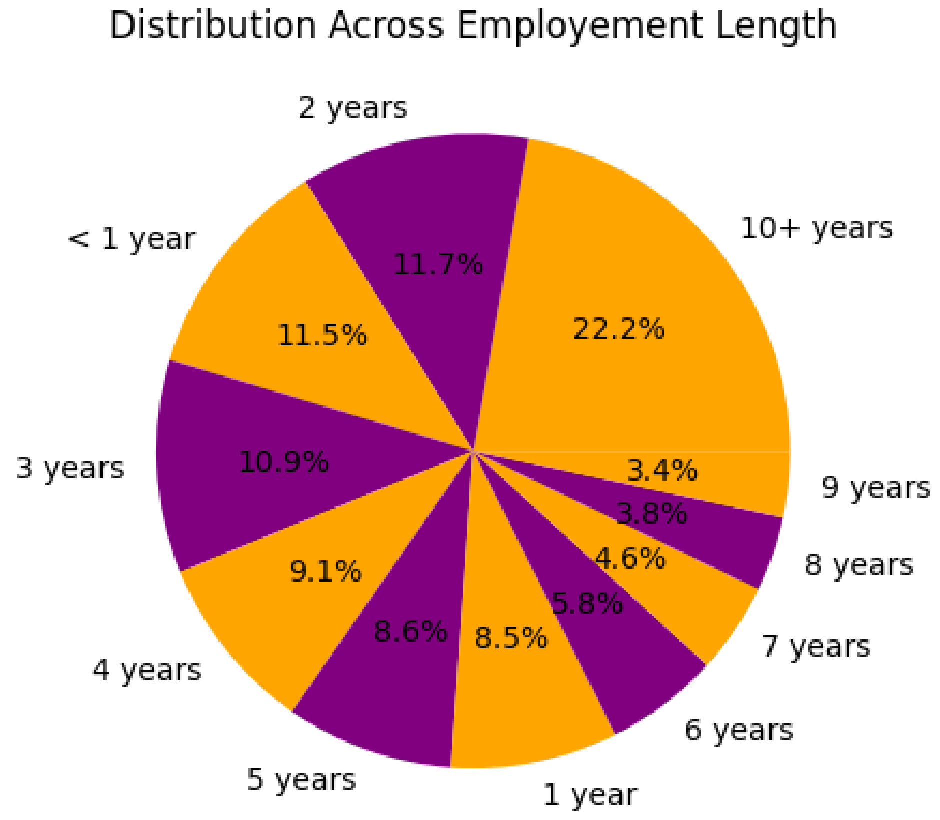
Interest Rate



04 - Univariate Categorical Analysis

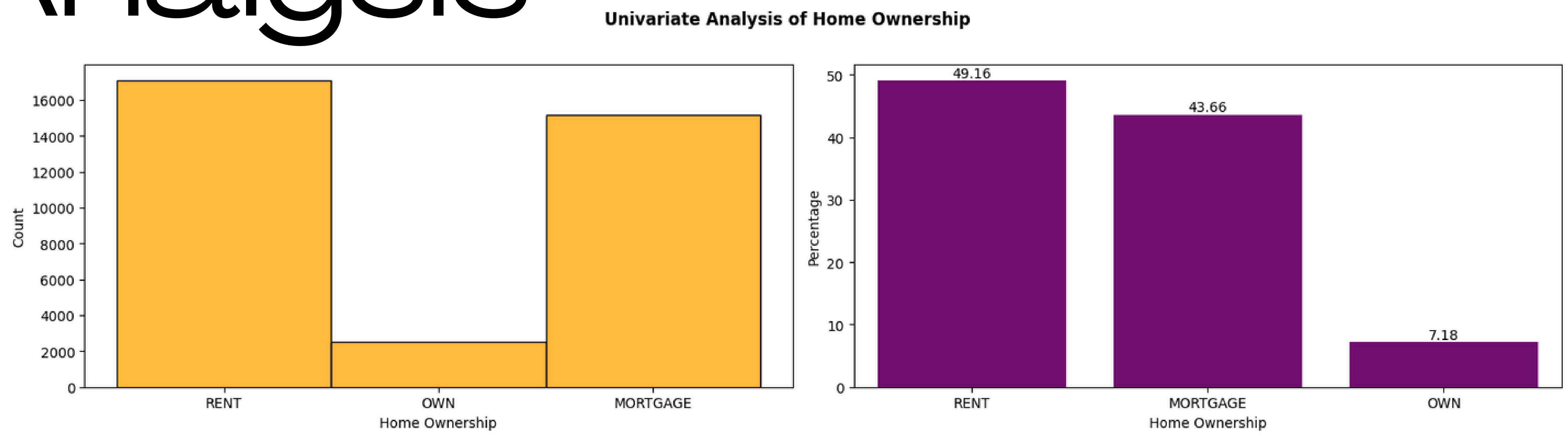


- 85% of the Applicants have paid the loan
- 15% of the Applicants have defaulted the loan



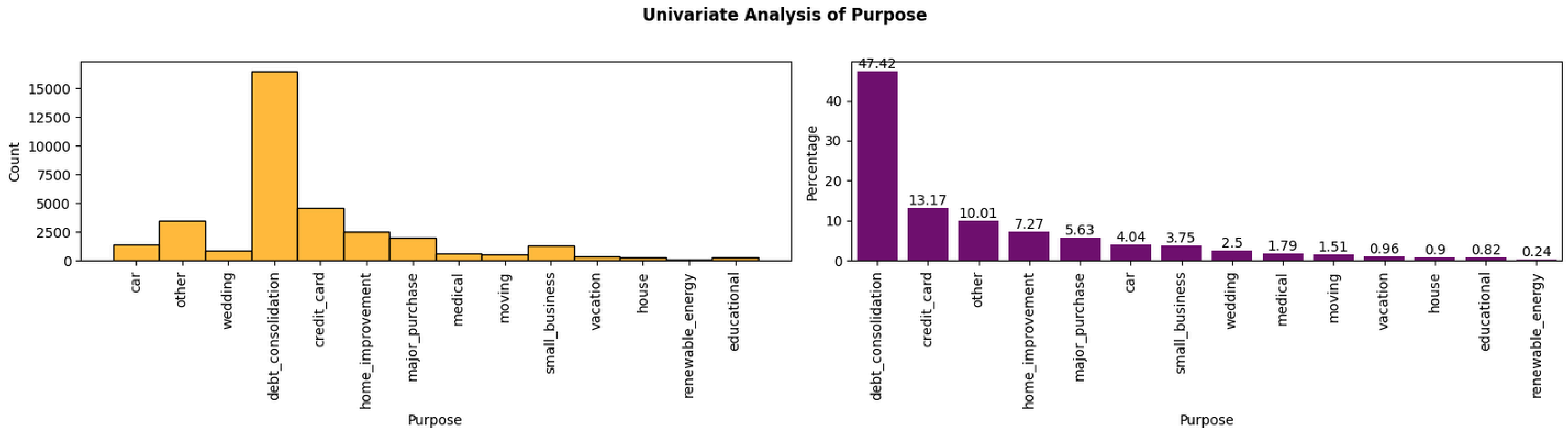
- 23% of the Applicants have Employment length of 10+ years.
- 11% of the Applicants have Employment length of <1 year.
- Only 3.3% of the Applicants have Employment length of 9 years.

04 - Univariate Categorical Analysis

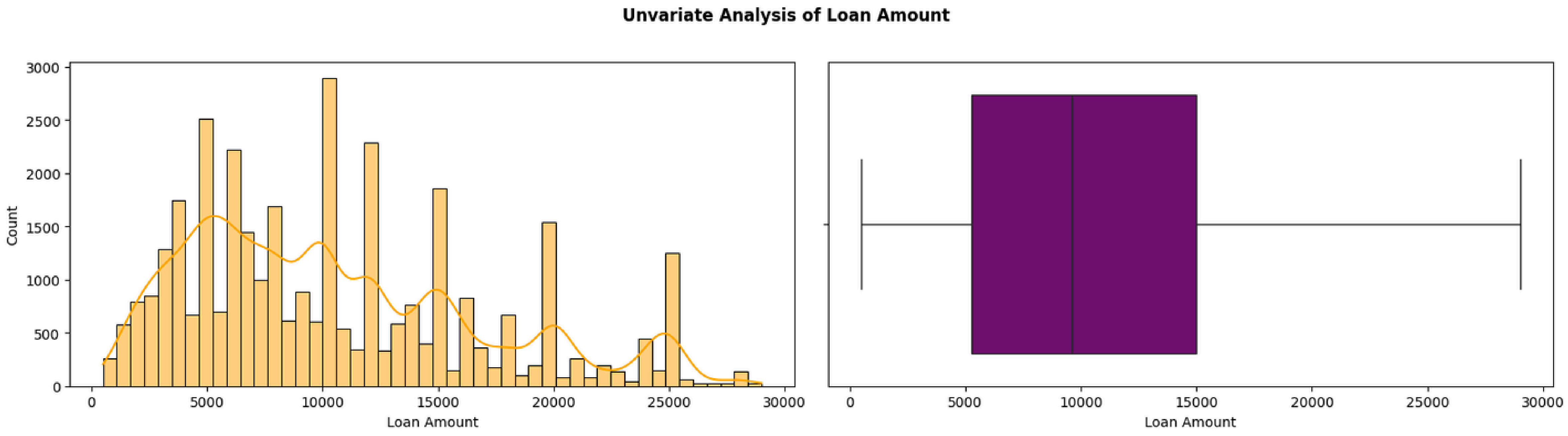


- 49% of the Applicants have Rented Home.
- 42% of the Applicants have Mortgage Home.
- 7.12% of the Applicants have Owned Home.

- 47.42% of the Applicants had applied for loan for Debt Consolidation.
- 13.17% of the Applicants had applied for loan for Credit Card.
- Renewable Energy has the least number of Applicants.

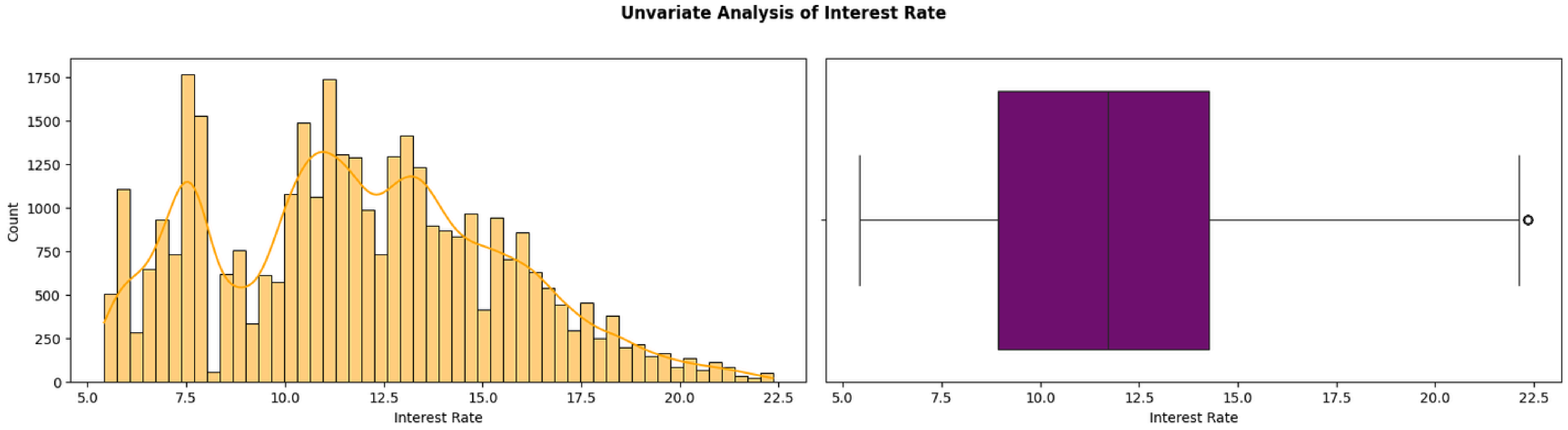


04 - Univariate Continuous Analysis

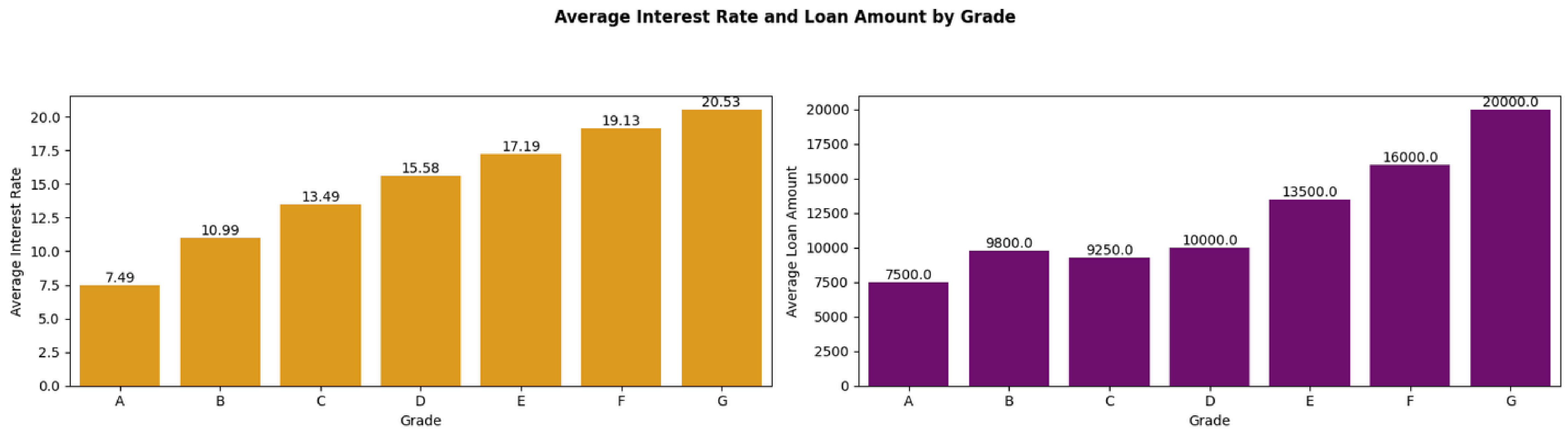


- Highest number of Applicants had applied for loan amount between 5000 to 10000 .
- Least number of Applicants had applied for loan amount greater than 25000.
- Average Loan Amount is 9600.

- Highest number of Applicants had applied for loan which had interest rate between 10.45 to 11.03.
- Least number of Applicants had applied for loan which had interest rate greater than 20.



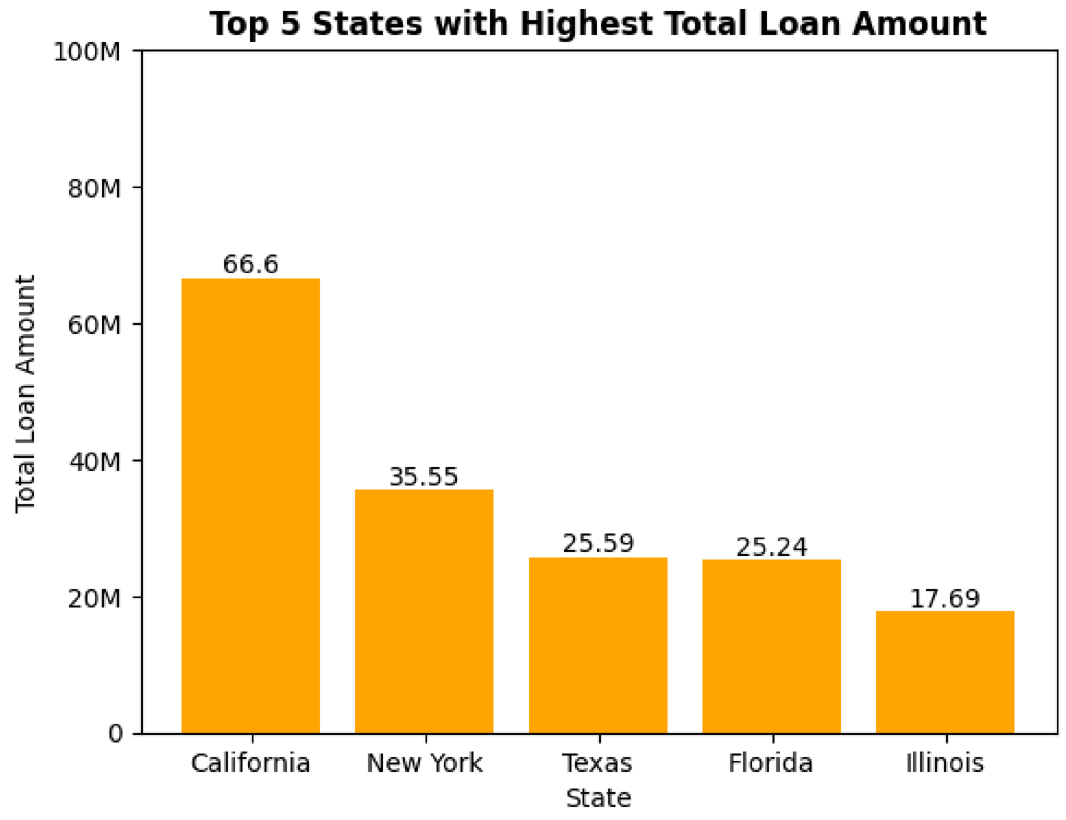
04 - Segmented Analysis



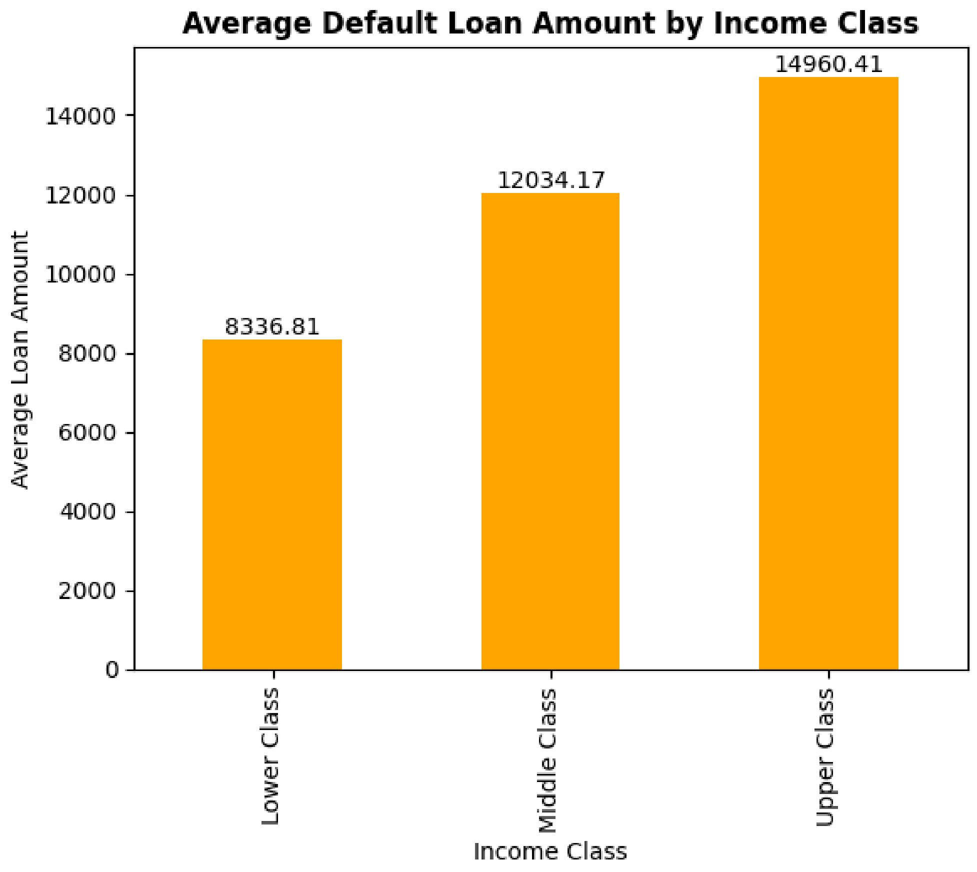
- Loan's which are graded `F` have High Interest Rates.
- Loan's which are graded `A` have Low Interest Rates.
- `Interest Rates` are gradually increasing from `Grade A` to `Grade G`.

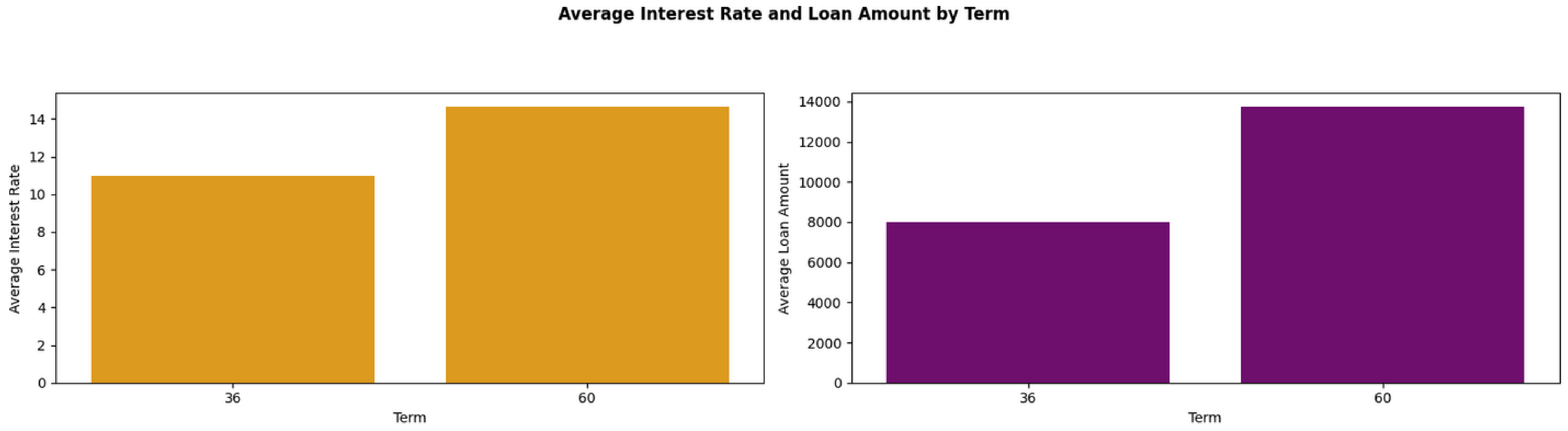
- `Grade F` have highest `Average Loan Amount`.
- `Grade A` have lowest `Average Loan Amount`.
- `Grade B,C,D ` have almost same `Average Loan Amount`.

- `CA` has highest Total Loan Amount.
- `NY` has second highest Total Loan Amount.
- `TX` has third highest Total Loan Amount.
- `FL` has fourth highest Total Loan Amount.
- `NJ` has fifth highest Total Loan Amount.



04 - Segmented Analysis





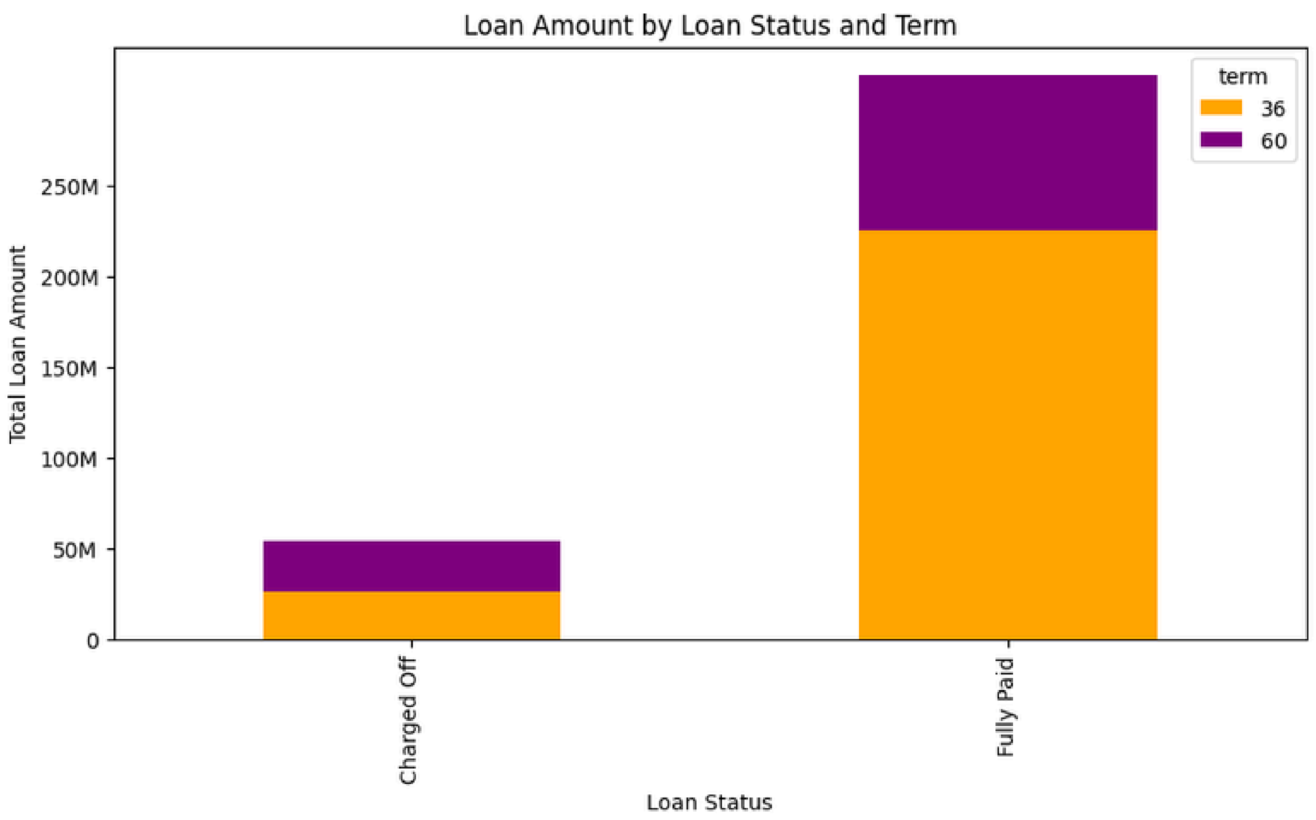
- Upper Class have highest Average Default Loan Amounts.
- Lower Class have lowest Average Default Loan Amounts.

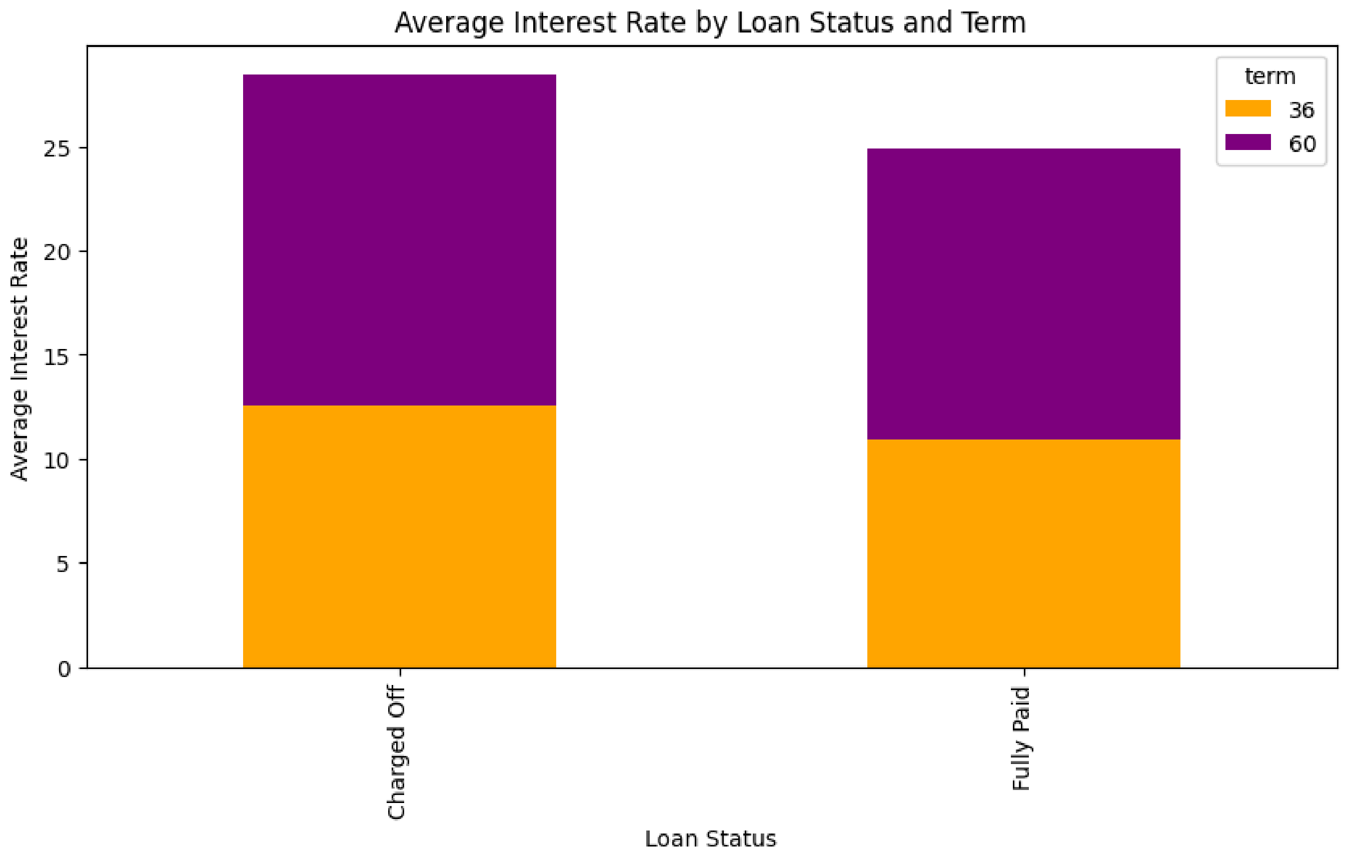
- Long Term Loans (60 months) have high `Interest Rates` due to high risk involved so lenders expect high returns.
- Short Term Loans (36 months) have low `Interest Rates` due to low risk involved so lenders expect low returns.

- Total Loan Amount is high for `36 months` term loans which shows Lenders are more interested in giving loans for 36 months term where risk is low and interest rates are low.

04 - Bivariate Categorical Analysis

- Borrowers with "Fully Paid" loans dominate in both 36-month and 60-month terms.
- Borrowers with "Charged Off" loans are riskier, as the amounts are smaller and skewed towards the longer (60-month) term.

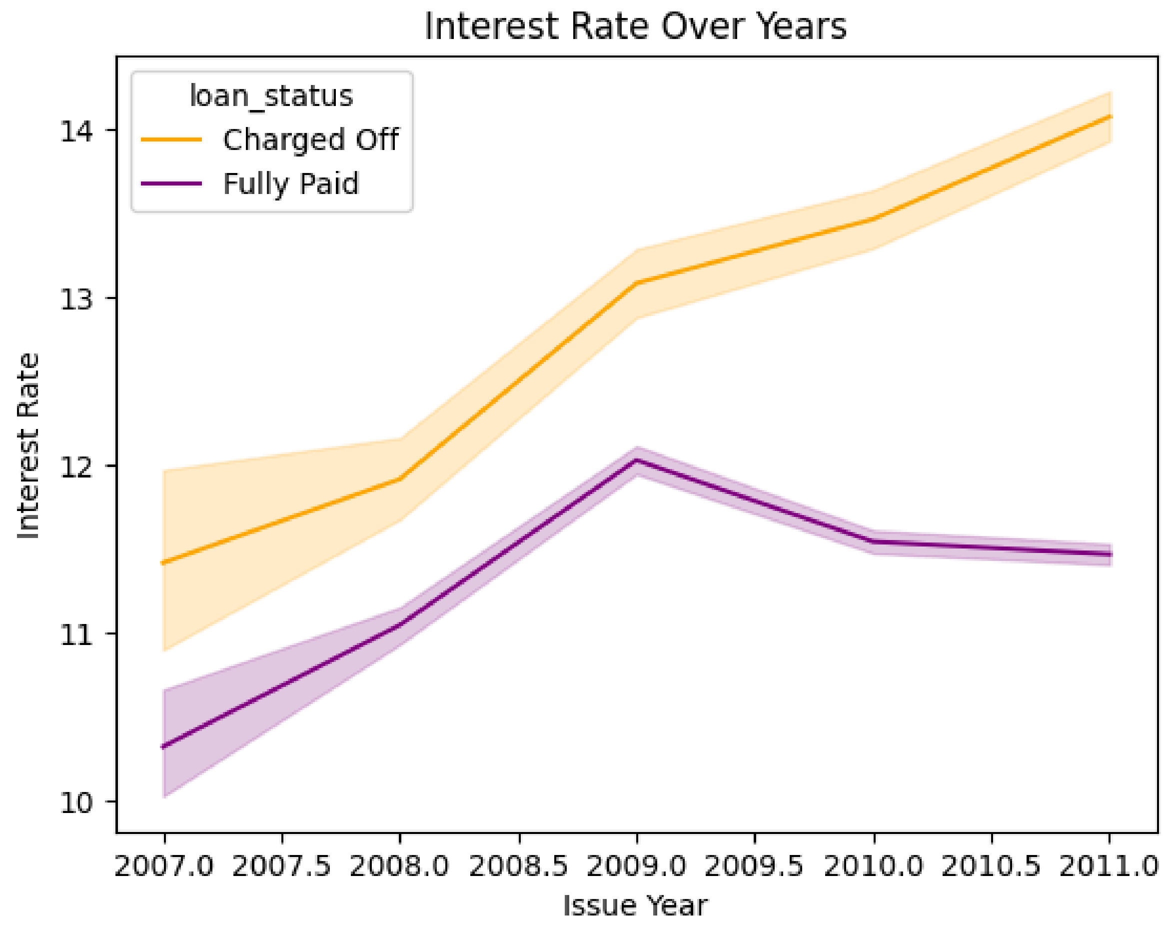


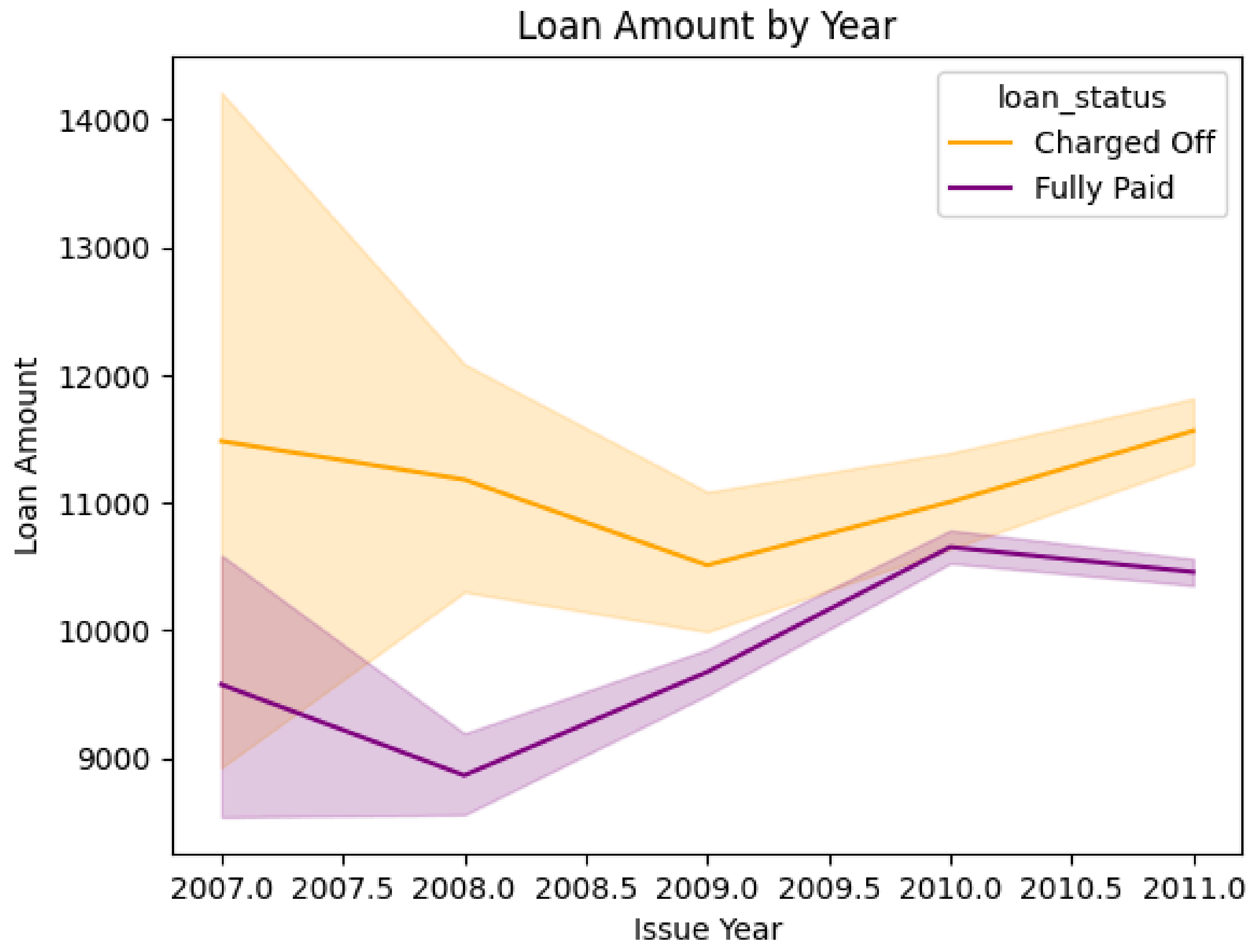


- 60 Months Loan Term have high default rate where Lenders have high risk and high return.
- 36 Months Loan Term have low default rate where Lenders have low risk and low return.

04 - Bivariate Continuous Analysis

- Charged Off :
  - Interest rates have been increasing over the years.
  - The interest rates for charged off loans are higher than fully paid loans.
- Fully Paid :
  - Interest rates increases till 2009 and then decreases.
  - The interest rates for fully paid loans are lower than charged off loans.

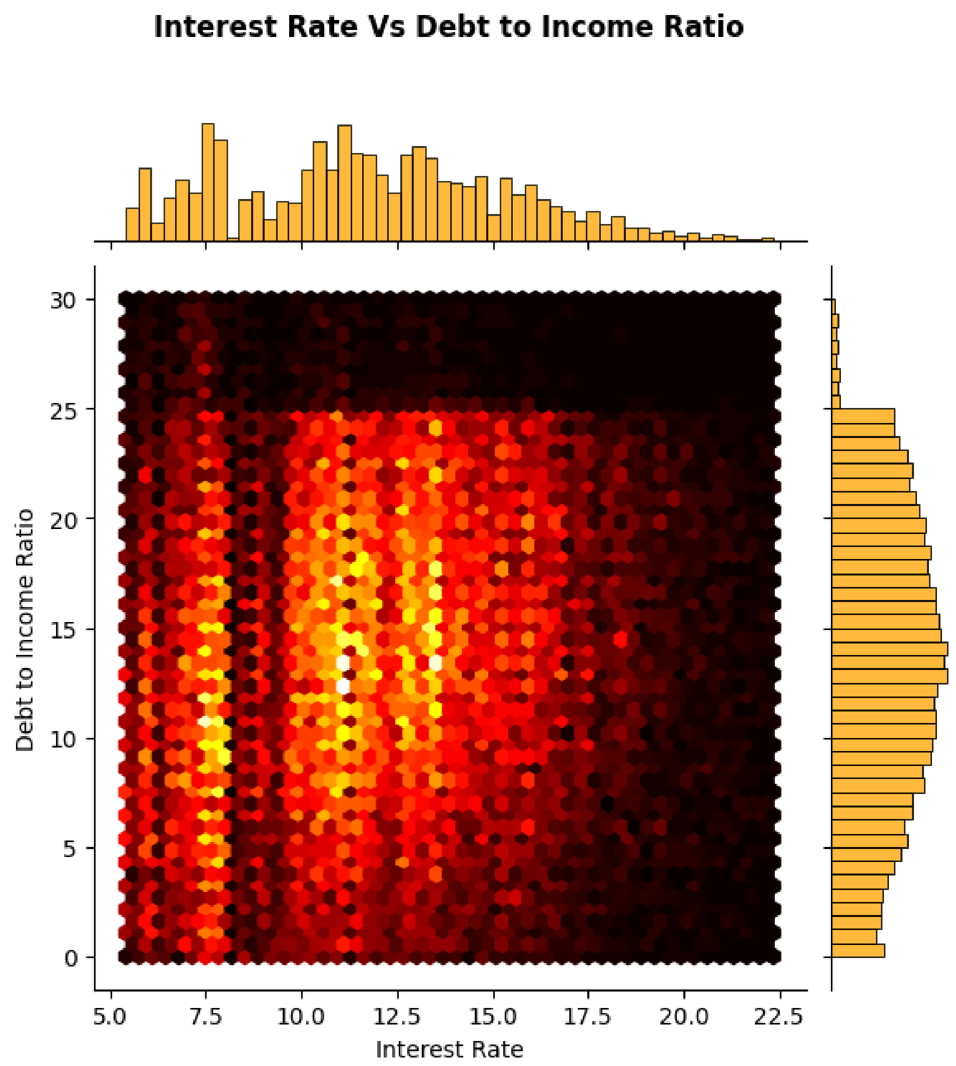


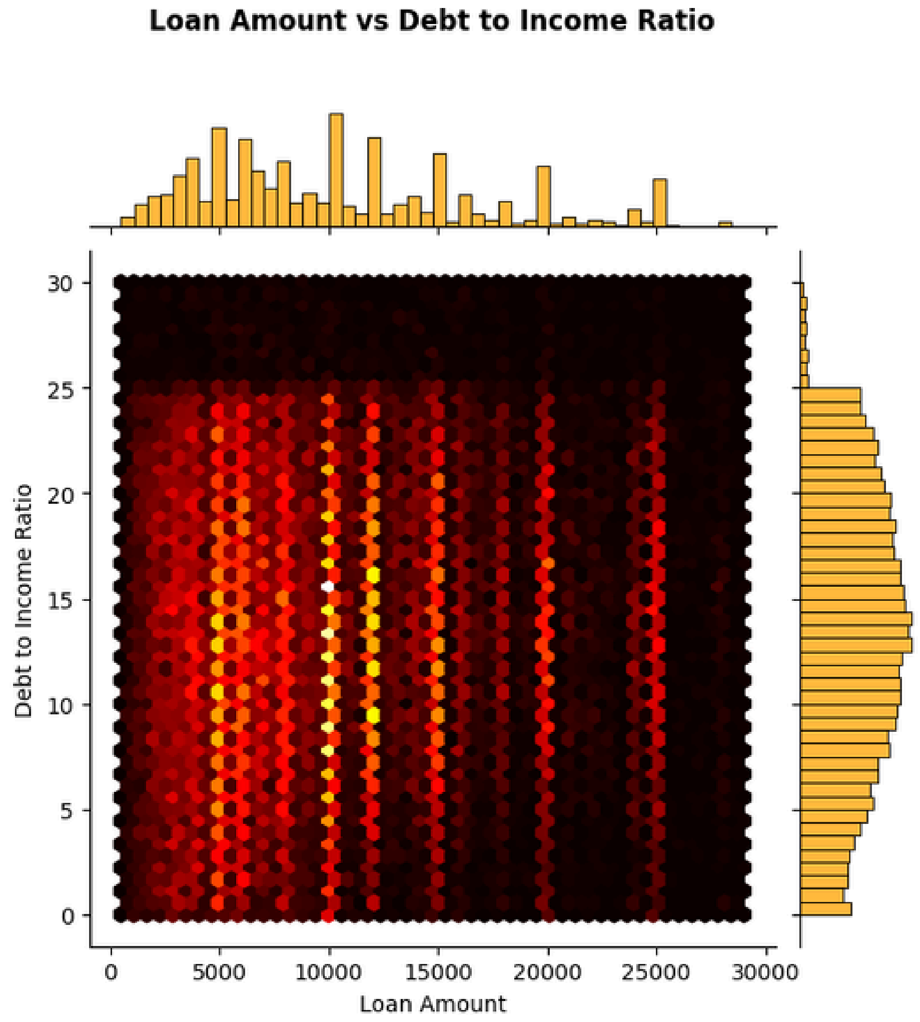


- The line for "Charged Off" loans starts at a higher level and gradually decreases. This suggests that initially, larger loans were more likely to be charged off.
- The line for "Fully Paid" loans starts at a lower level and gradually increases. This suggests that initially, smaller loans were more likely to be fully paid.

04 - Bivariate Continuous Analysis

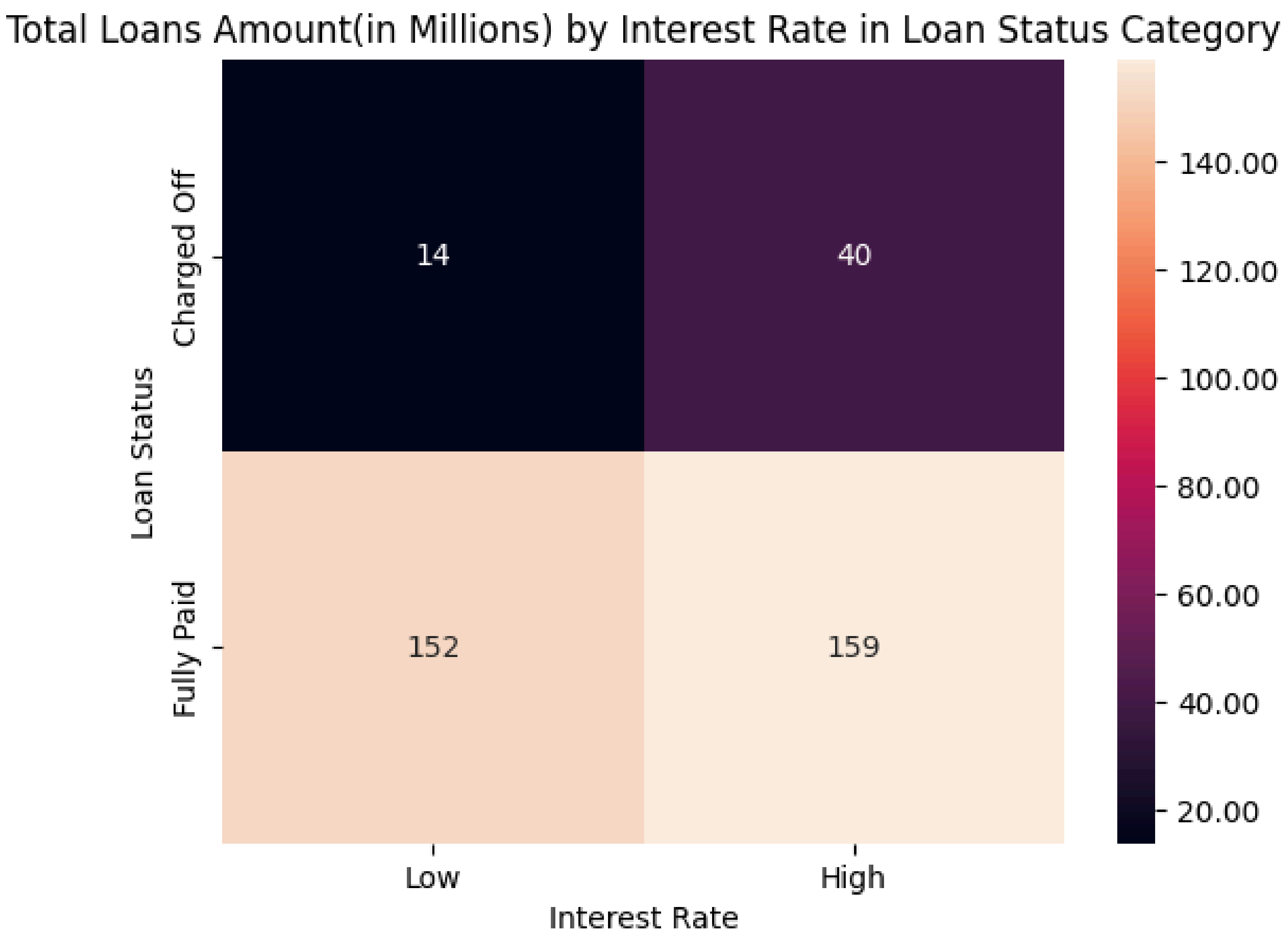
- There is a weak positive correlation between interest rate and debt-to-income ratio. This suggests that borrowers with higher debt-to-income ratios tend to have higher interest rates. Lenders might perceive borrowers with higher debt-to-income ratios as having a higher risk of default.
- Concentration of data where moderate int_rate values (10 to 13) correspond to moderate dti values (10 to 15). This could indicate a relationship or a common range for both variables



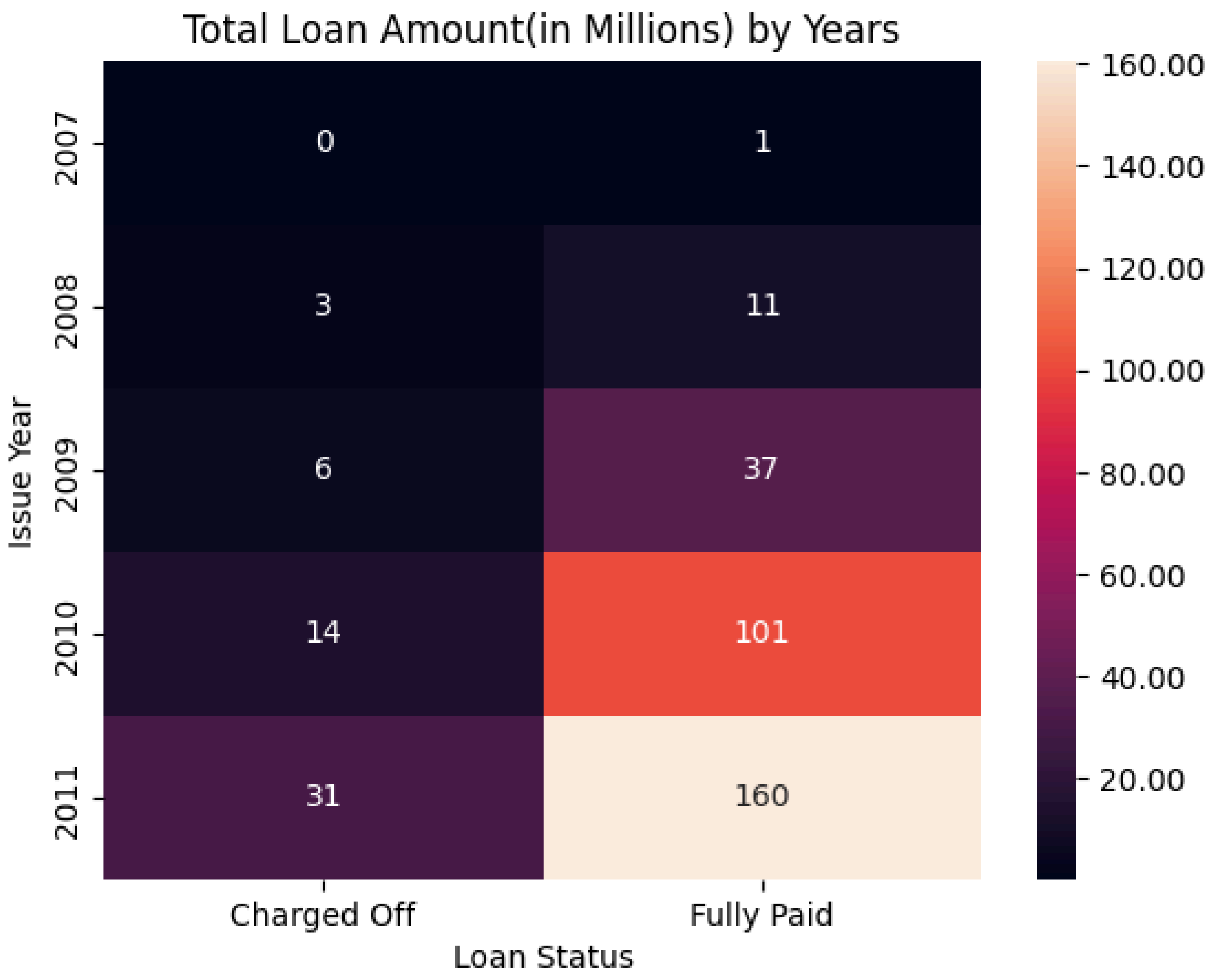


- The vertical segmentation in the hexbin plot and the spikes in the loan_amnt histogram indicate that specific loan sizes are more commonly issued (e.g., 5,000, 10,000, 15,000).
- Most borrowers have a debt-to-income ratio between 10 and 15, and loans are frequently issued in the 10,000–15,000 range.

04 - Multivariate Analysis

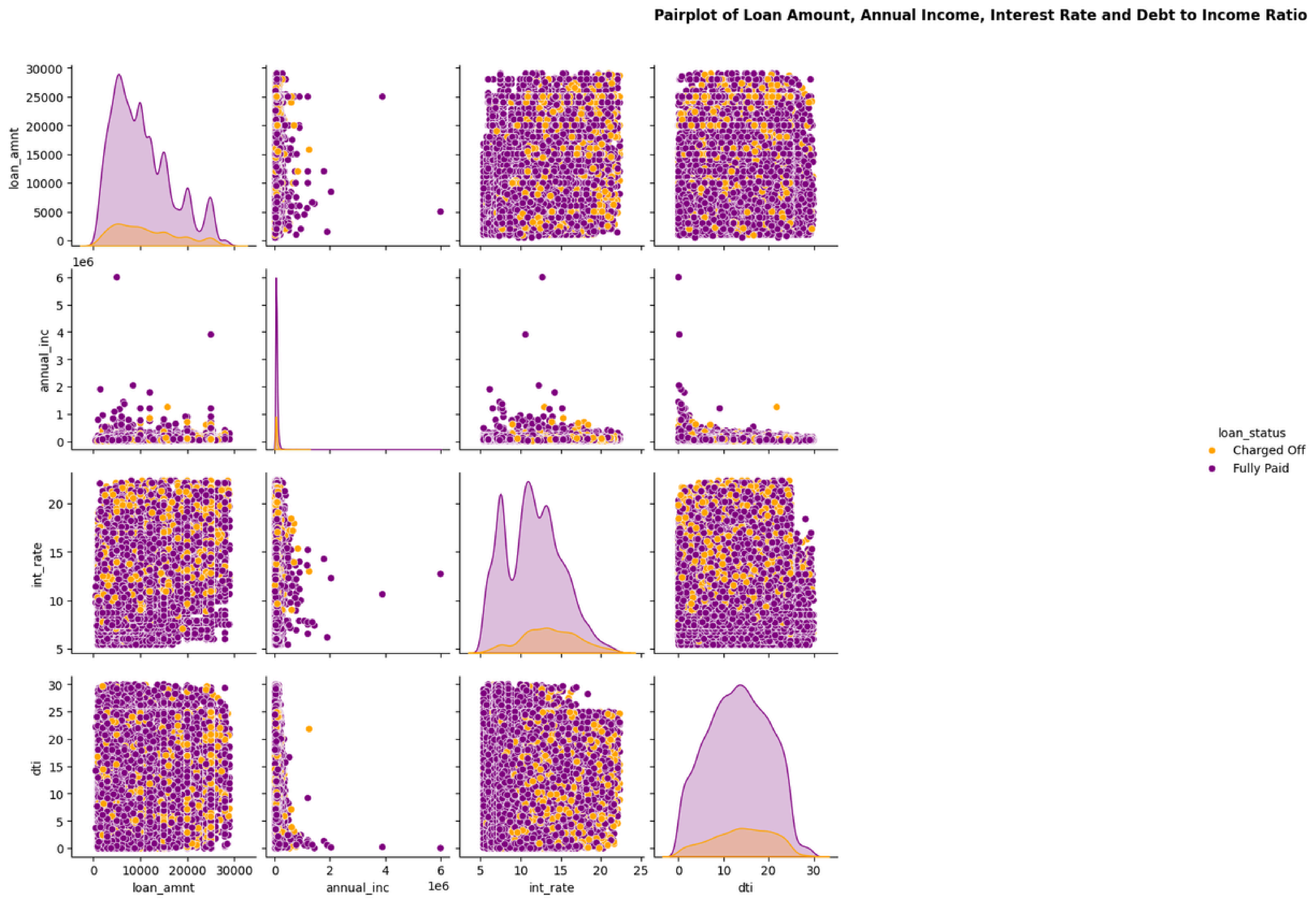


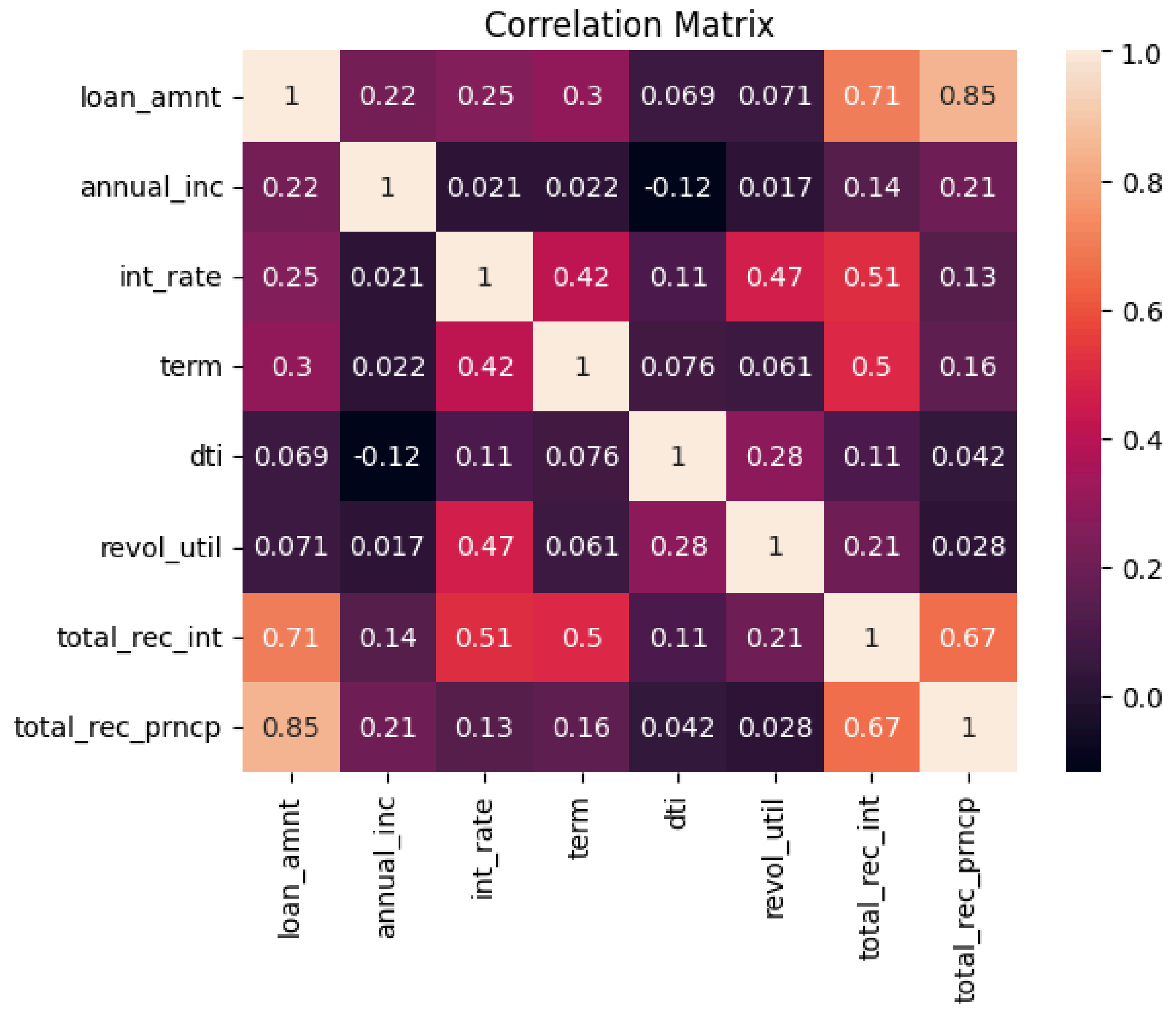
- Defaulters have high `Loan Amount` and high `Interest Rate`.
- Non-Defaulters have low `Loan Amount` and low `Interest Rate`.



- `Fully Paid` loans amount increases with increase in `issue_d` year.
- `Charged Off` loans amount increases with increase in `issue_d` year.

04 - Multivariate Analysis





Both graphs giving relationship between Continuous Variables in Dataset.

05 - Conclusions

1. **Loan Status and Interest Rates**:
  - Defaulters tend to have higher interest rates compared to non-defaulters.
  - High interest rates are more likely to result in loan defaults.

2. **Loan Amount and Loan Status:**
  - Higher loan amounts are associated with higher risk of default.
  - Applicants with higher loan amounts should be scrutinized more closely.

3. **Employment Length and Loan Status:**
  - Applicants with unstable employment lengths and low loan amounts are riskier.
  - Despite stable employment lengths, some applicants with high loan amounts still default.

4. **Home Ownership and Loan Status:**
  - Applicants with rented homes have a higher risk of default compared to those with mortgage or owned homes.

5. **Purpose of Loan and Loan Status:**
  - Certain loan purposes, such as debt consolidation, have higher total loan amounts and higher default rates.
  - Applicants applying for loans for high-risk purposes should be evaluated carefully.

6. **Income Class and Loan Status:**
  - Upper-class applicants have the highest average default loan amounts.
  - Lower-class applicants have the lowest average default loan amounts.

7. **Geographical Analysis:**
  - States like California (CA) and New York (NY) have the highest total loan amounts and highest default amounts.
  - Applicants from these states should be assessed with additional caution.

8. **Grade and Loan Status:**
  - Loans graded `F` have the highest interest rates and are more likely to default.
  - Loans graded `A` have the lowest interest rates and are less likely to default.
  - Interest rates gradually increase from `Grade A` to `Grade G`.

9. **Loan Term and Loan Status:**
  - Long-term loans (60 months) have higher default rates due to higher risk.
  - Short-term loans (36 months) have lower default rates due to lower risk.

Thanks