

# Lending Club Case Study

Swaraj Kumar  
Gagan Banga

# Problem Statement

## Summary

- Solve a business problem using EDA for the consumer finance company which specializes in lending various types of loans to urban customers.

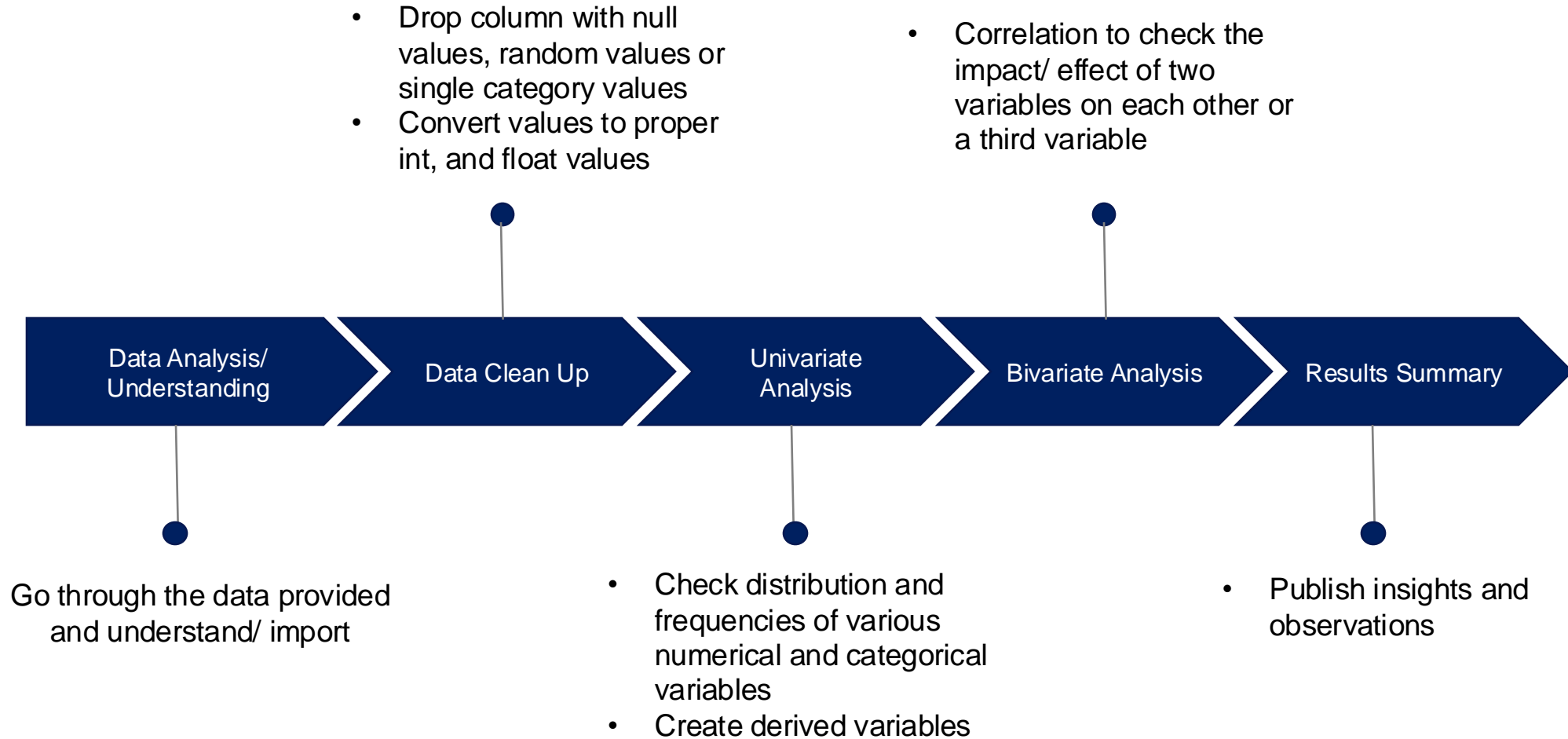
## Summary

- To develop a model that assists the company to make informed decision based on the past patterns of the applicants

## Outcome/ Approach

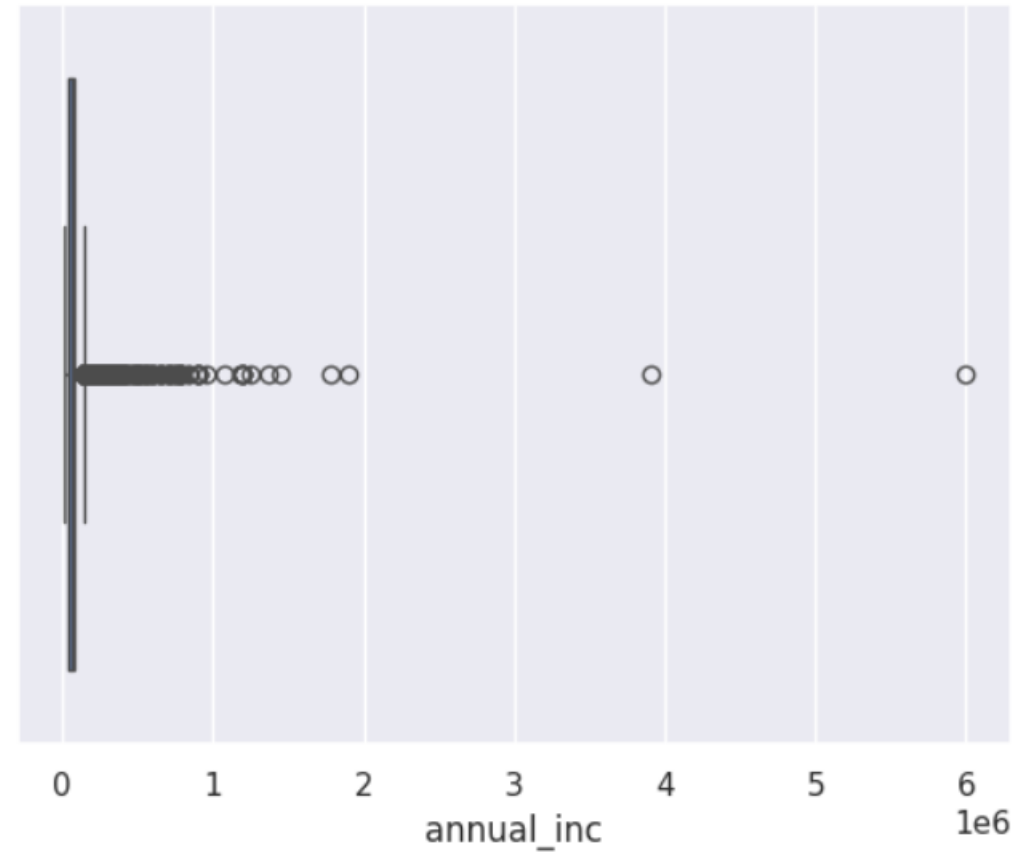
- Data Understanding/ Cleaning
- Data Analysis
- Recommendations

# Analysis Approach

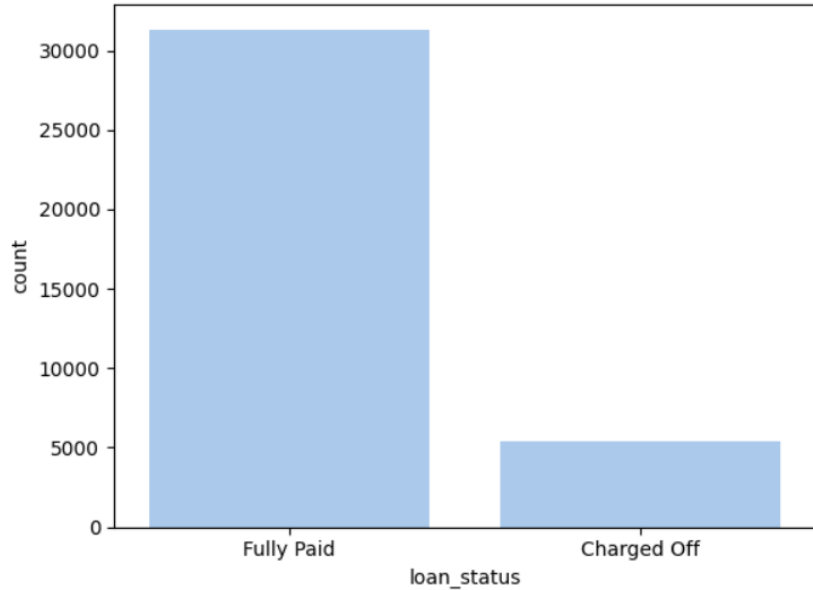


# Range of outliers

- Data distribution indicates presents of outliers, explore the range of outliers using box plot
- Outliers are above the 97th percentile, so we removed them



# Univariate Analysis

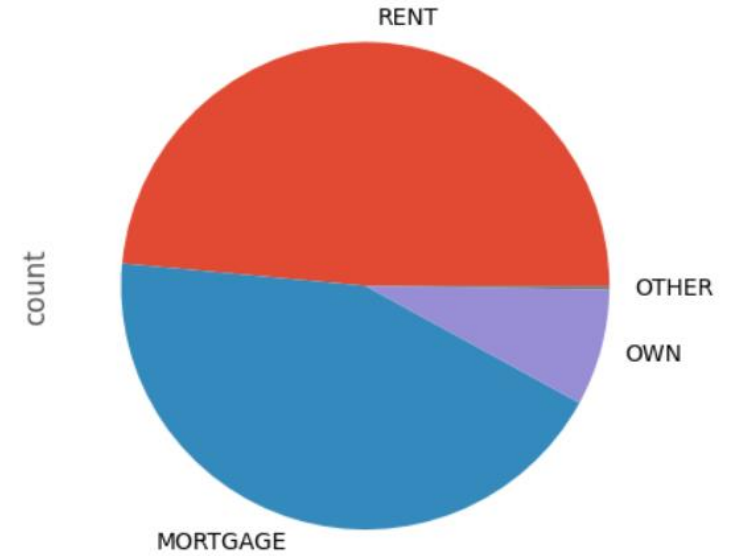
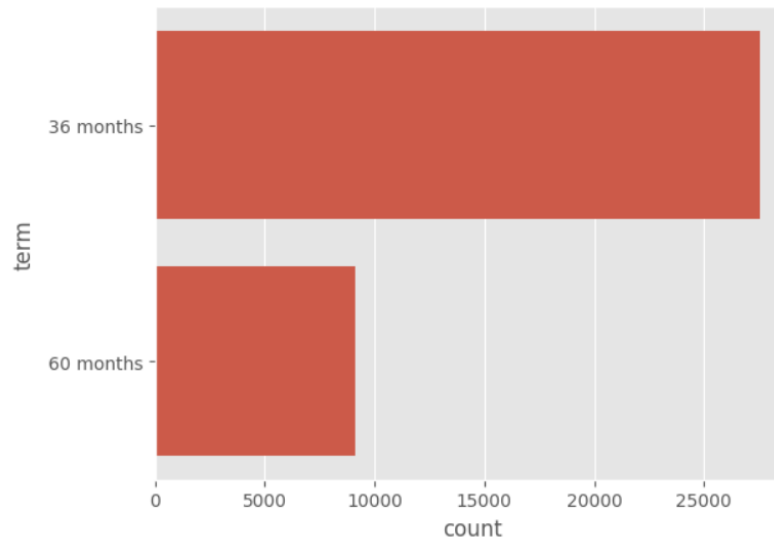


↑

This count plot visualizes the distribution of loan statuses in the dataset. It helps identify the proportion of loans that are fully paid, charged off, or in default.

↓

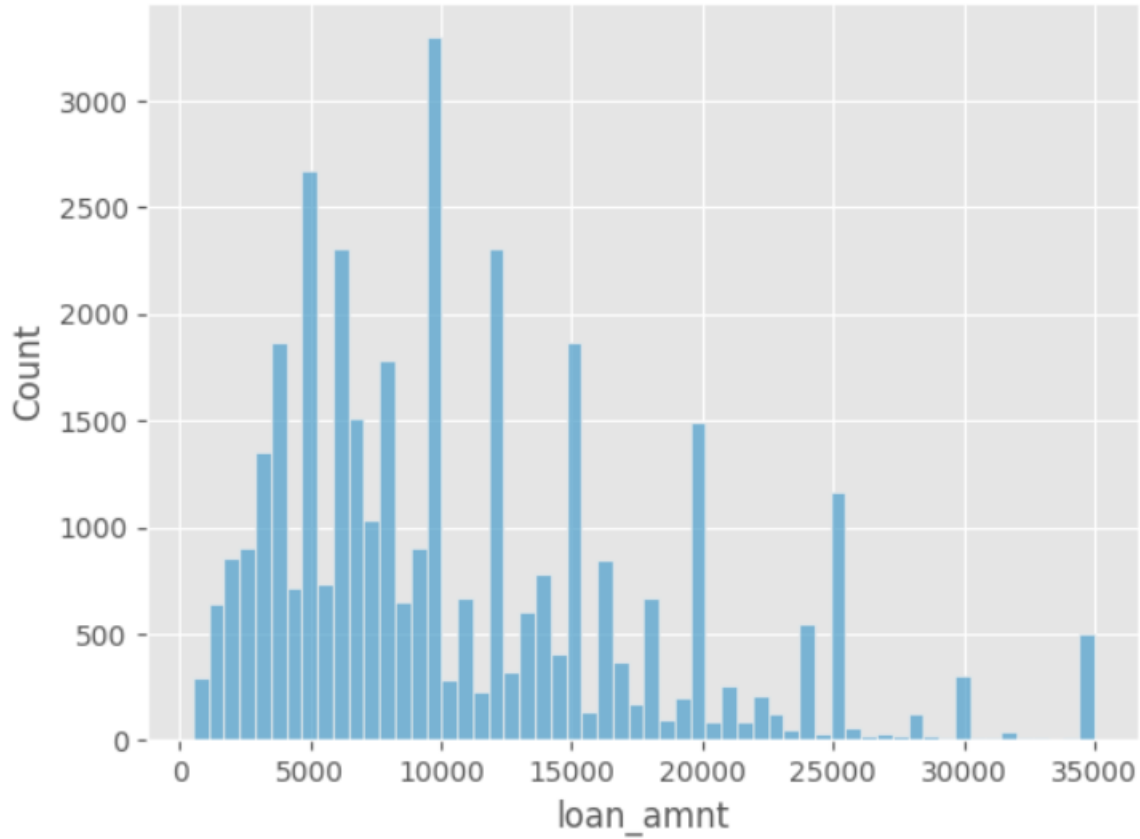
This count plot displays the distribution of loan terms. It helps understand the prevalence of different loan durations



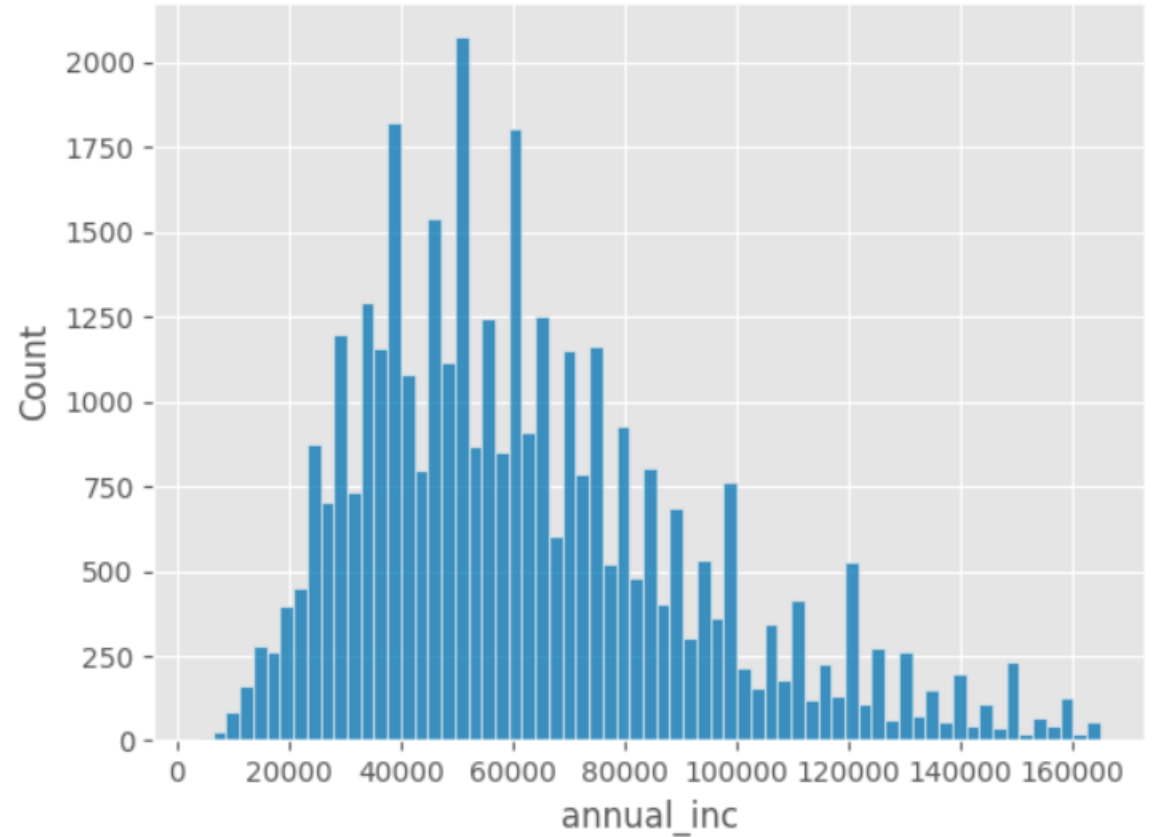
↑

This count plot visualizes the distribution of loan statuses in the dataset. The pie chart shows the distribution of home ownership statuses among borrowers. It provides insights into the most common types of home ownership in the dataset. identify the proportion of loans that are fully paid, charged off, or in default

# Univariate Analysis

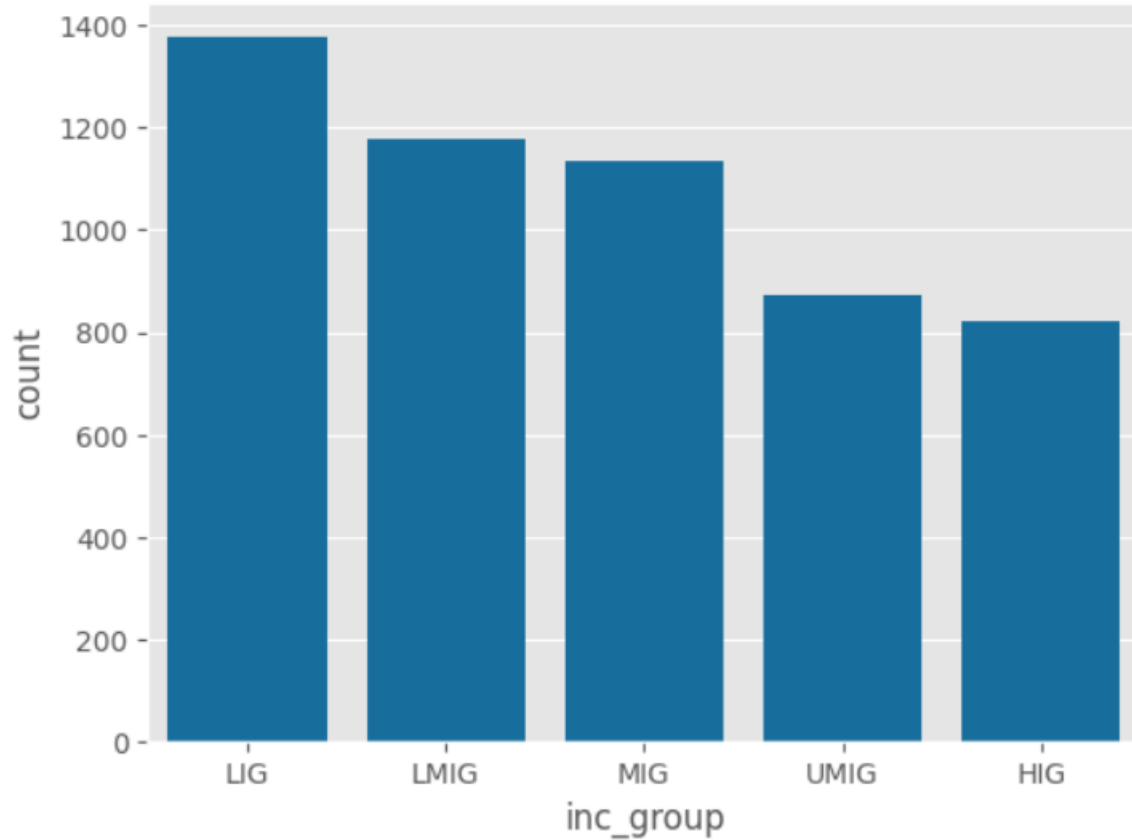


The histogram shows the distribution of loan amounts. It highlights the range and frequency of loan amounts requested by borrowers.

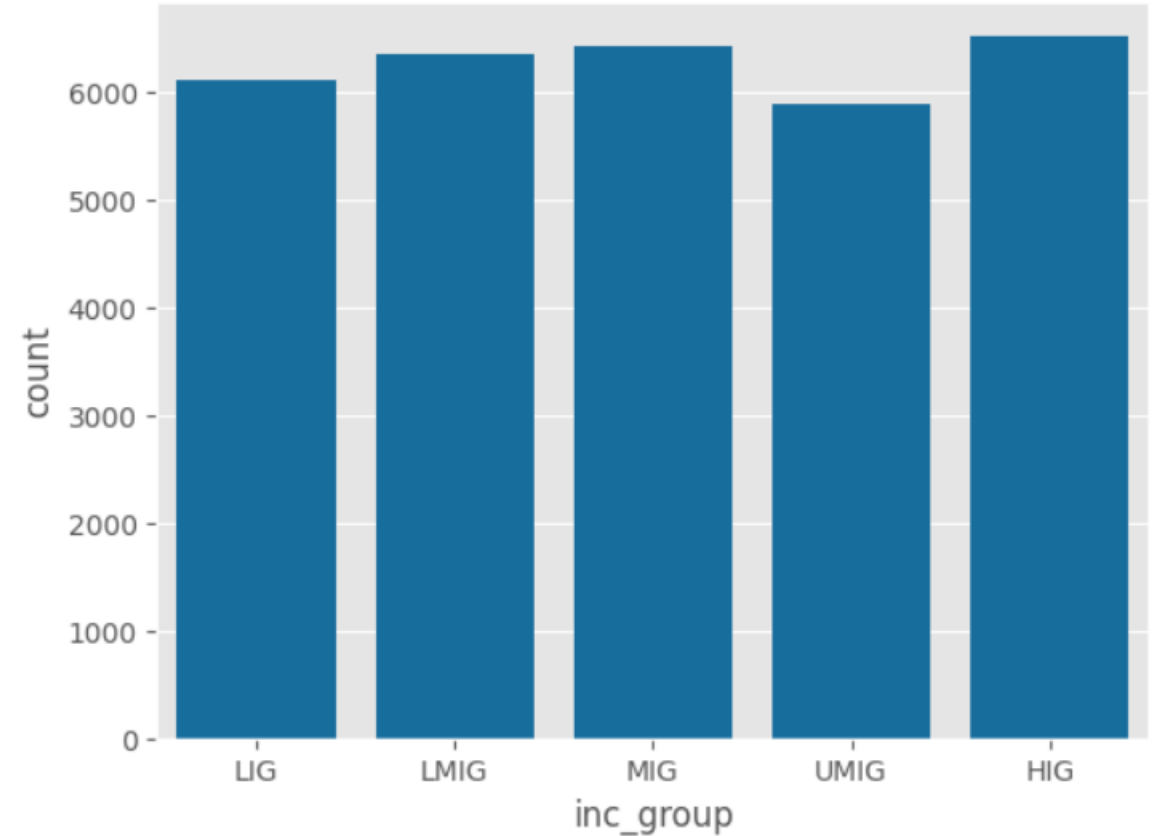


This histogram visualizes the distribution of annual incomes of borrowers. It helps identify the income levels of most borrowers.

# Datasets based on Loan Status for comparison



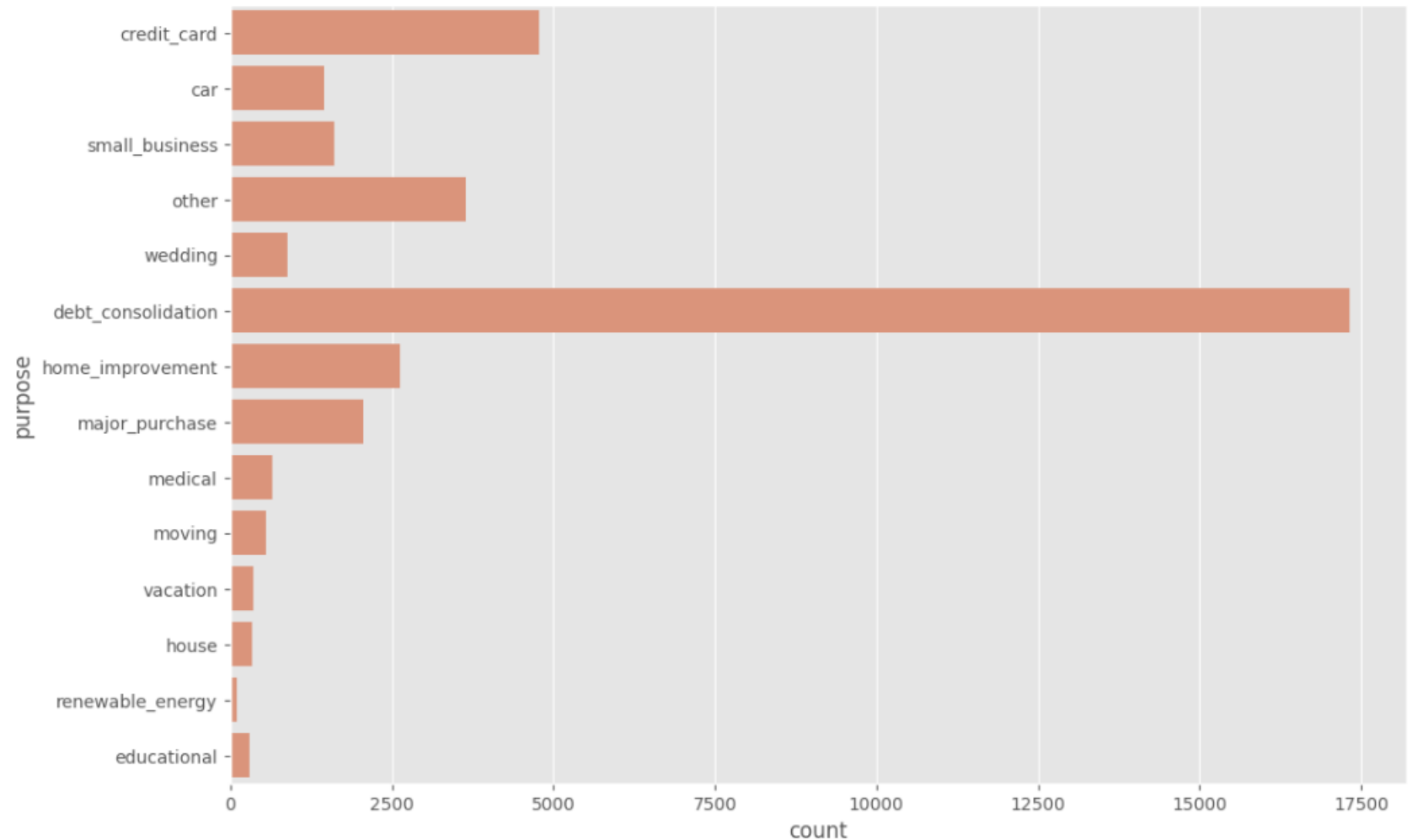
The count plot shows the distribution of income groups among defaulted loans. It provides insights into which income groups are more likely to default.



This count plot displays the distribution of income groups among paid-up loans. It helps understand which income groups are more likely to repay their loans.

# Univariate Analysis

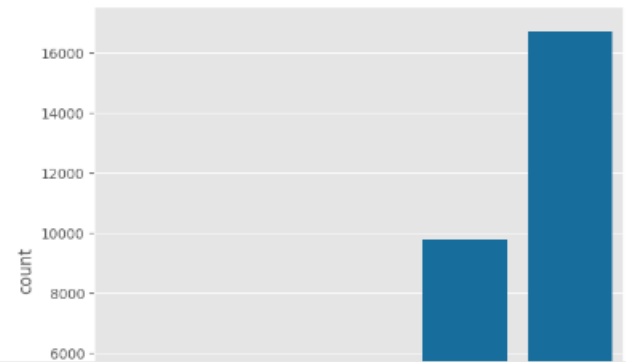
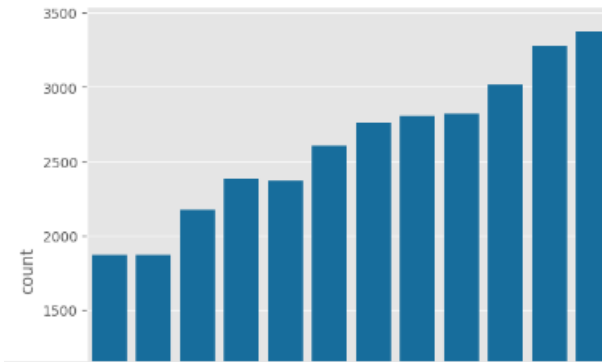
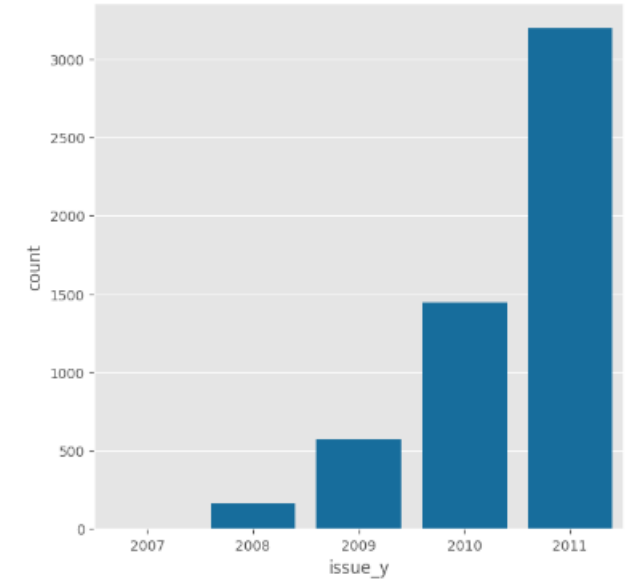
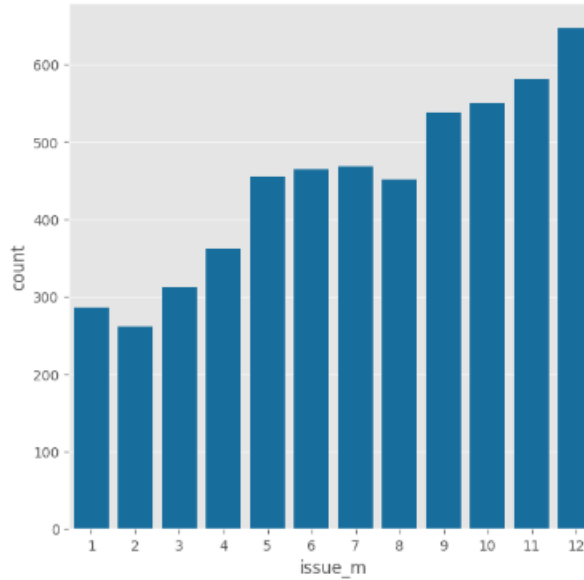
The count plot visualizes the distribution of loan purposes. It highlights the most common reasons borrowers take out loans.



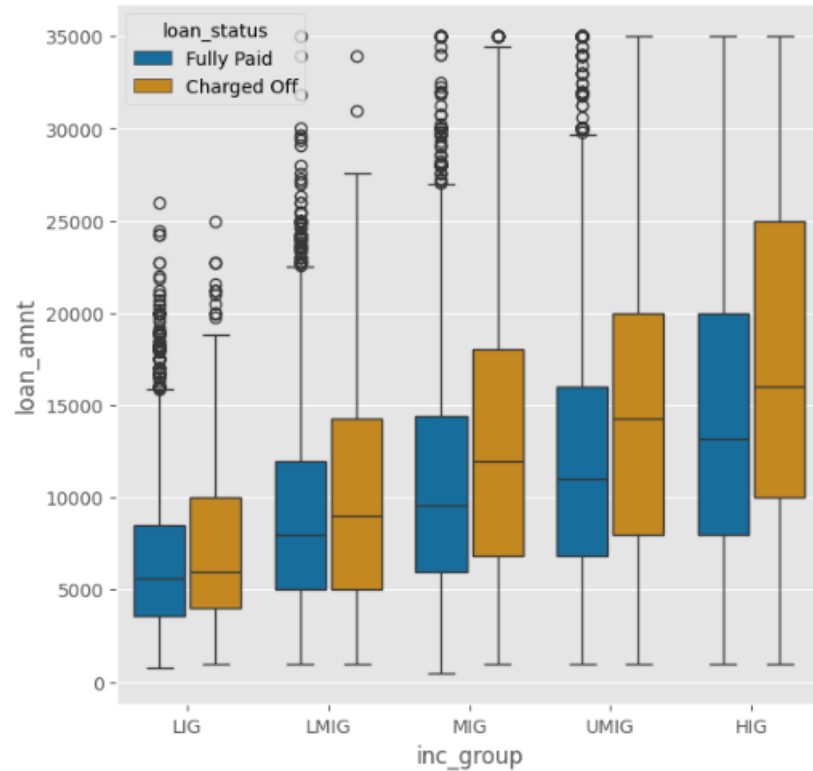


# Univariate Analysis

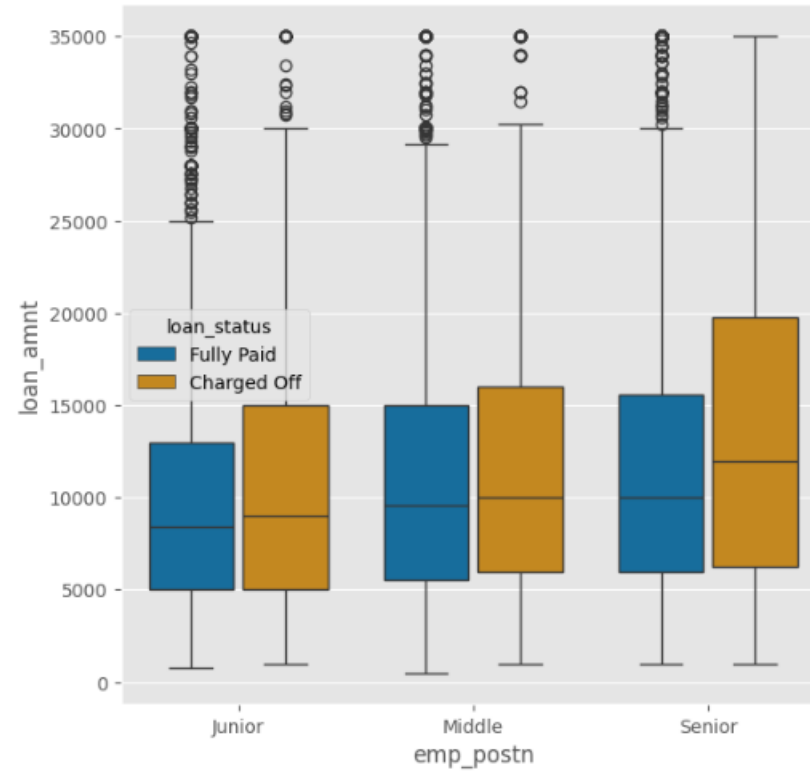
These subplots show the distribution of issue months and years for both defaulted and paid-up loans. They help identify any seasonal or yearly trends in loan issuance and repayment behavior.



# Bivariate Analysis



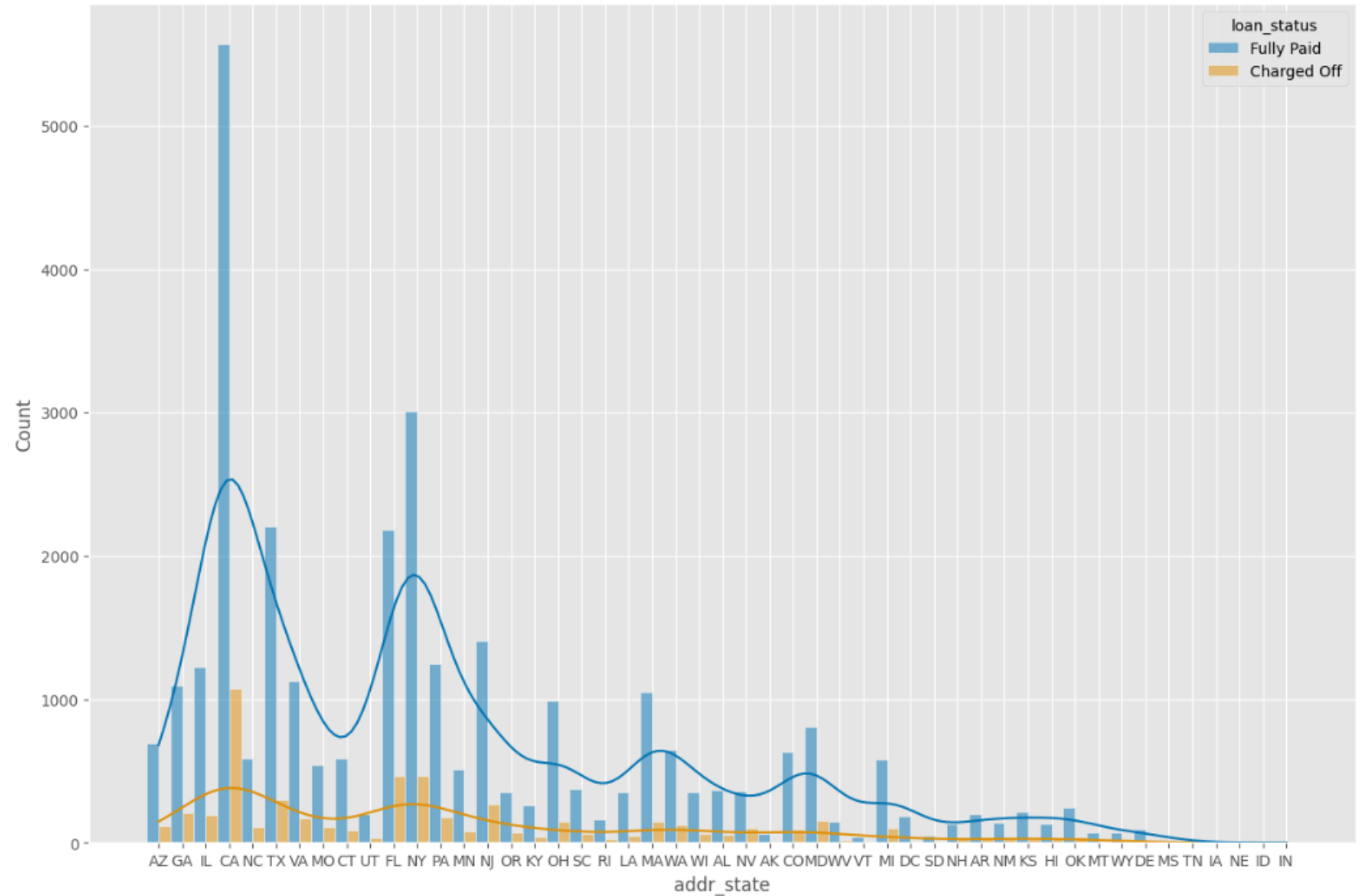
This plot shows the distribution of loan amounts across different income groups, segmented by loan status.



This plot illustrates the variation in loan amounts based on different employment positions.

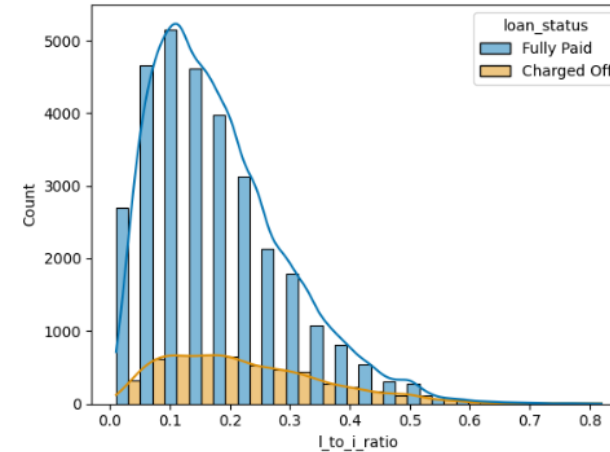
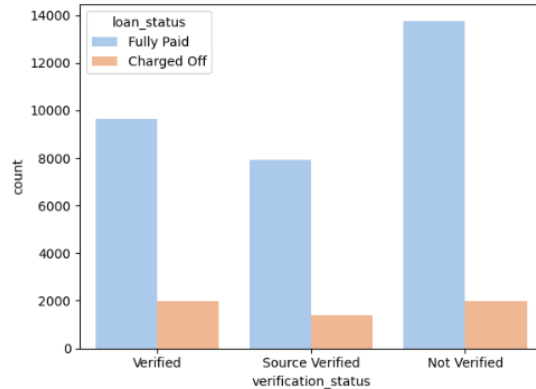
# Bivariate Analysis

This histogram displays the distribution of loans across different states, with a distinction between loan statuses



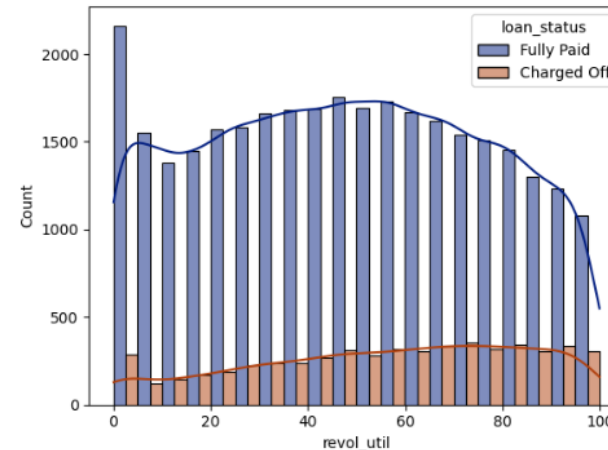
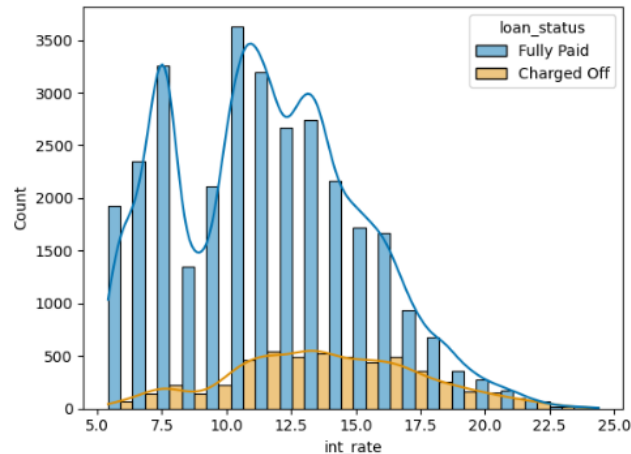
# Bivariate Analysis

This count plot shows the count of loans based on verification status, segmented by loan status



This histogram represents the distribution of the loan-to income ratio, categorized by loan status. It helps in understanding the relationship between borrowers' income and the loan amounts they receive

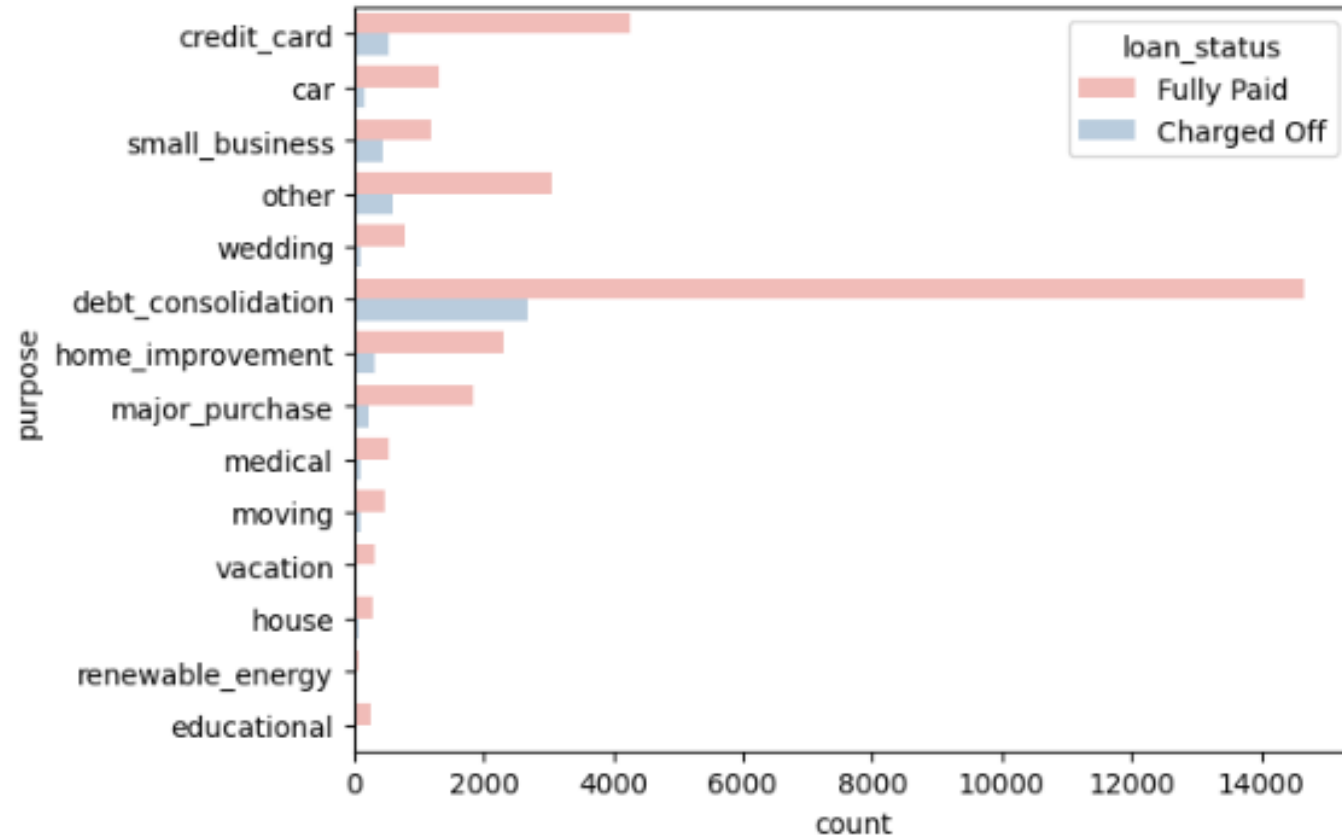
This plot shows the distribution of interest rates for loans, segmented by loan status



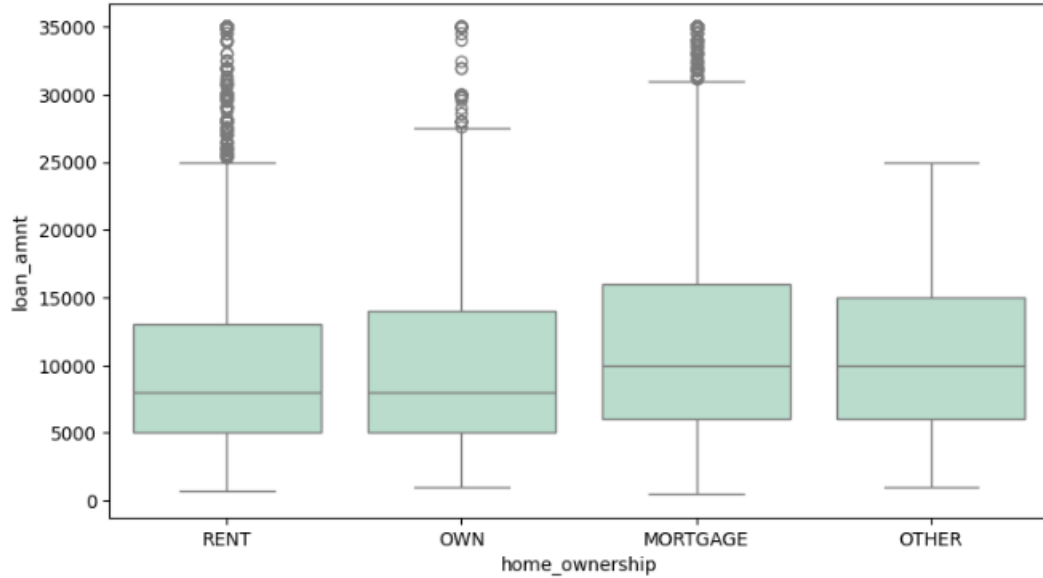
This histogram displays the distribution of revolving utilization rates, categorized by loan status

# Bivariate Analysis

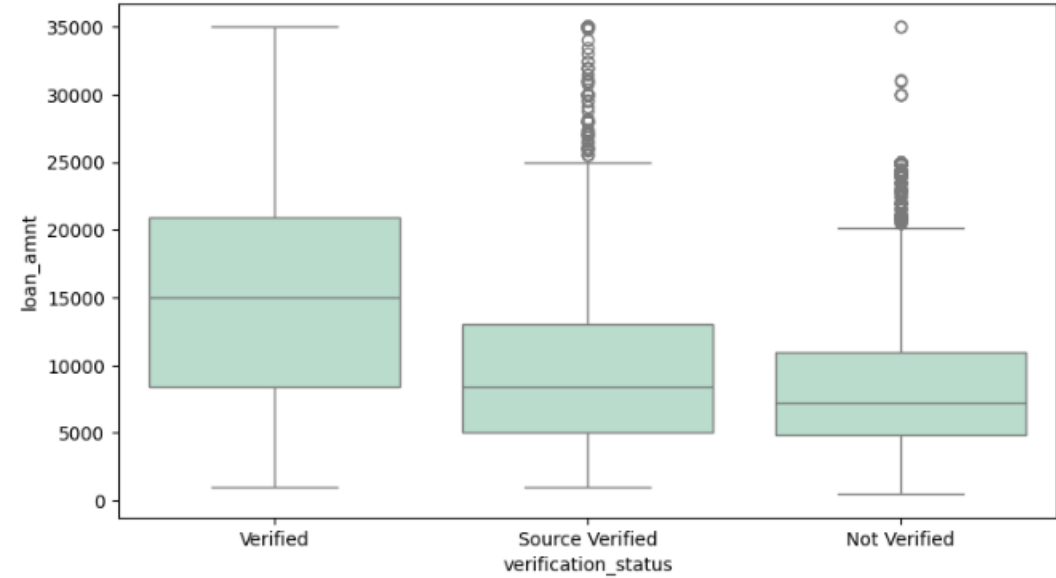
This count plot shows the count of loans for different purposes, segmented by loan status. It helps in identifying the most common reasons for taking loans



# Bivariate Analysis



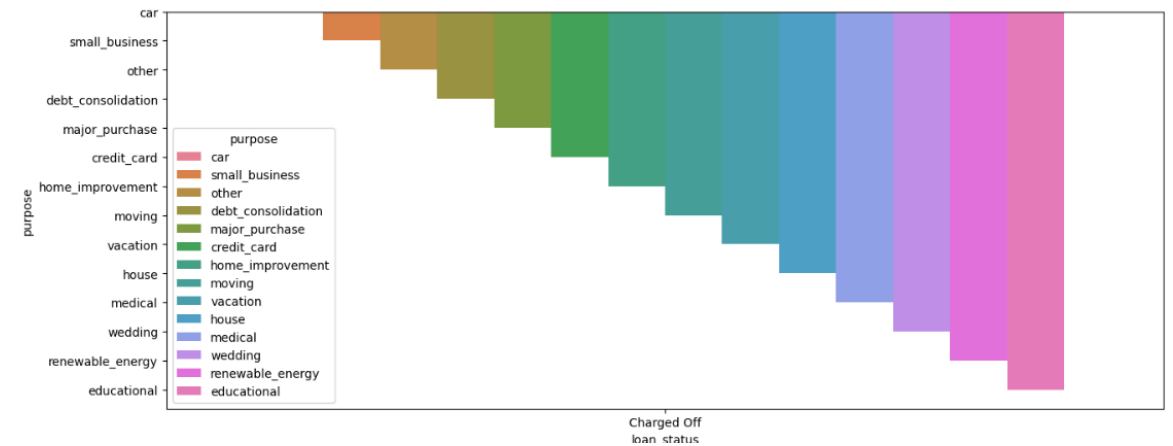
This plot illustrates the variation in loan amounts based on home ownership status. It provides insights into how owning a home affects the loan amounts borrowers receive



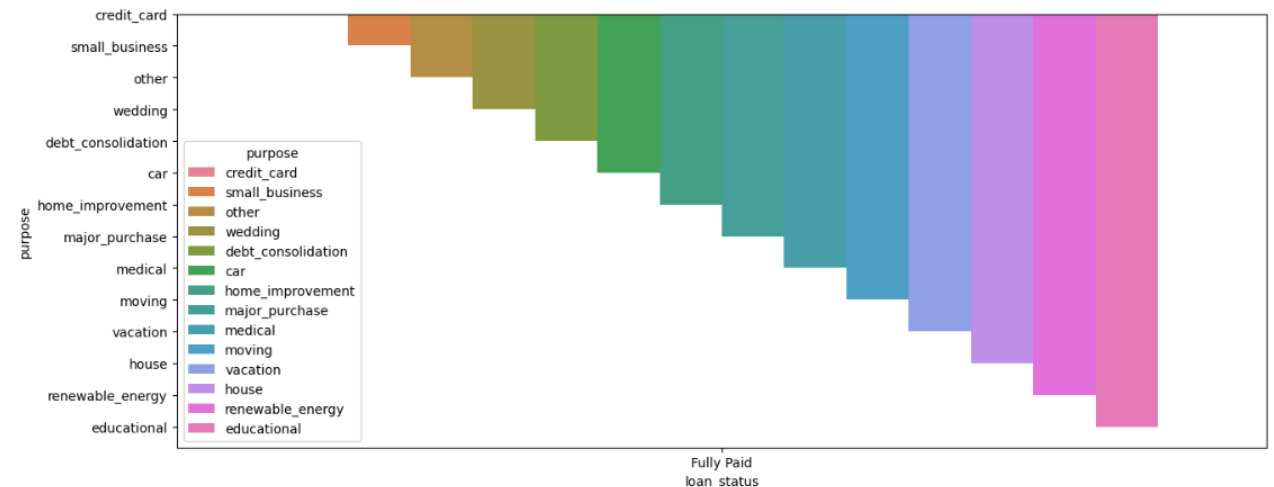
This plot shows the distribution of loan amounts based on verification status. It helps in understanding the impact of verification on the loan amounts.

# Bivariate Analysis

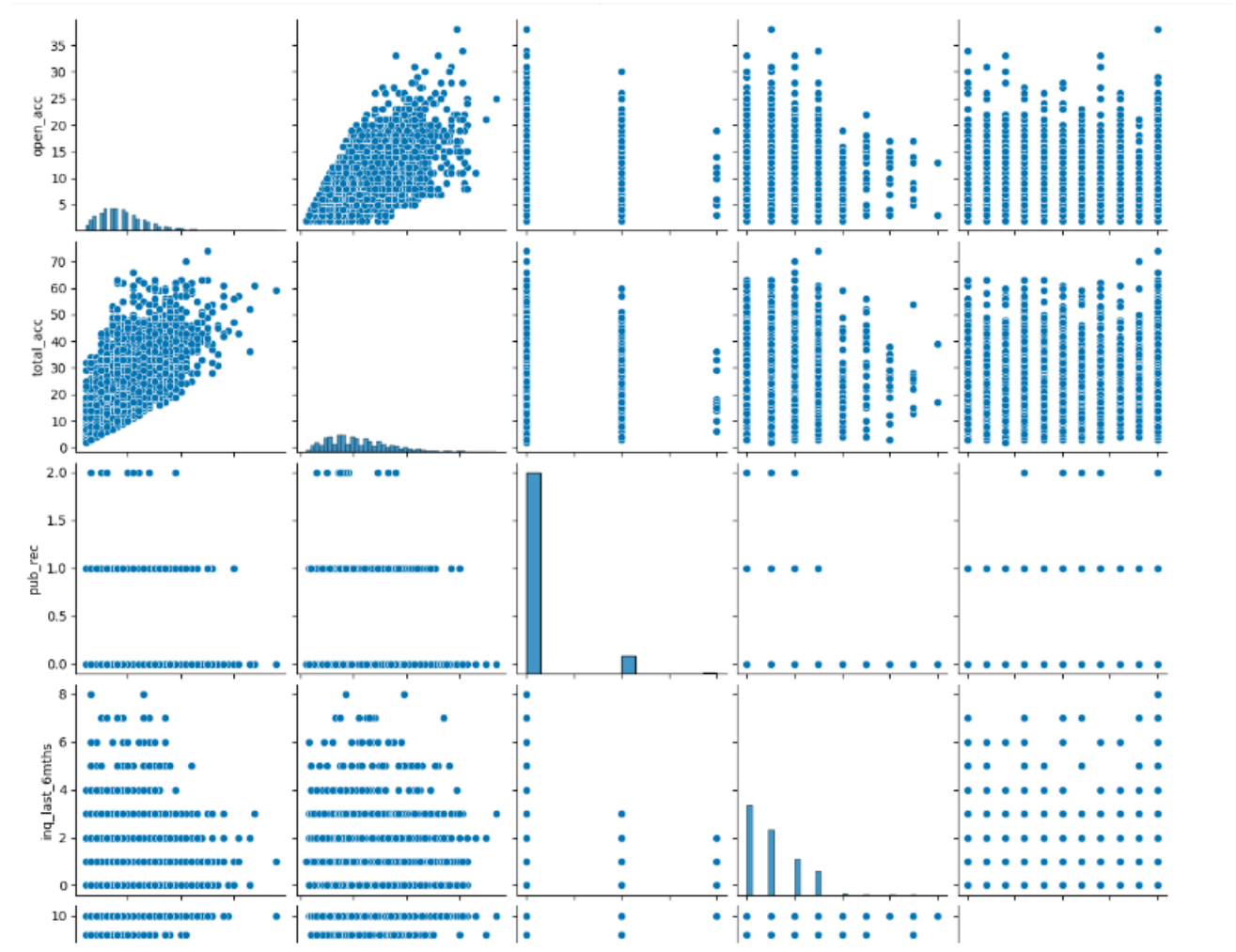
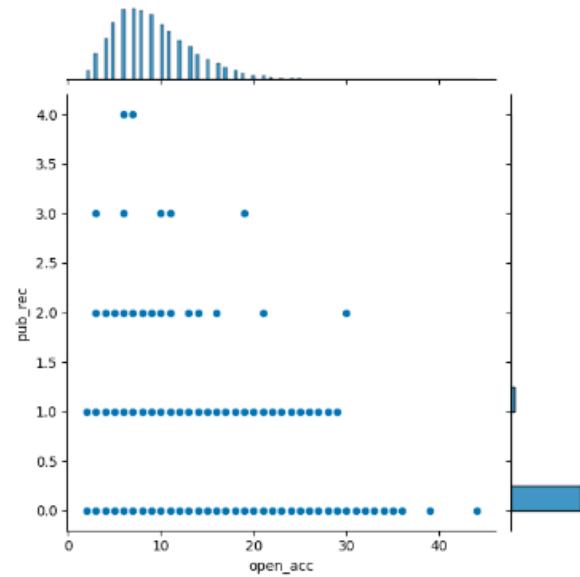
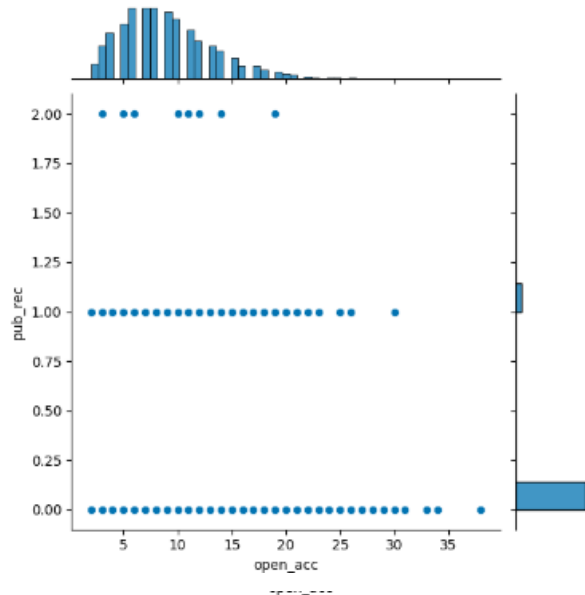
This bar plot displays the distribution of defaulted loans based on their purpose. It helps in identifying which loan purposes have higher default rates



This bar plot shows the distribution of paid-up loans based on their purpose. It provides insights into which loan purposes are more likely to be fully paid



# Bivariate Analysis





# Conclusion

- Based on the Exploratory Analysis of given data, we could infer that
- Borrowers who have a loan to income ratio of more than 30% are very likely to default
- Borrowers having high rate of revolving utilization are likely to default
- Loans taken for purposes Debt Consolidation, Credit Card, Small Business and Other are very likely to default
- Borrowers with employment tenure over have a risk of default
- Borrowers with over 100K annual income are likely to default compared to lower income group
- There is no relationship between the non demographic data and the risk of default

