

LENDING CLUB CASE STUDY

SUBMISSION

Group Members: Pavindan Natrajan & Raghavendra Derangula

Project Introduction

➤ Project brief:

- Lending club is a consumer finance company which specializes in lending various types of loans to urban customers.
- Based on applicant's profile, the company has to make decision for loan approval.
- Company has to be very careful and risk involved in getting the loan approved from bank.

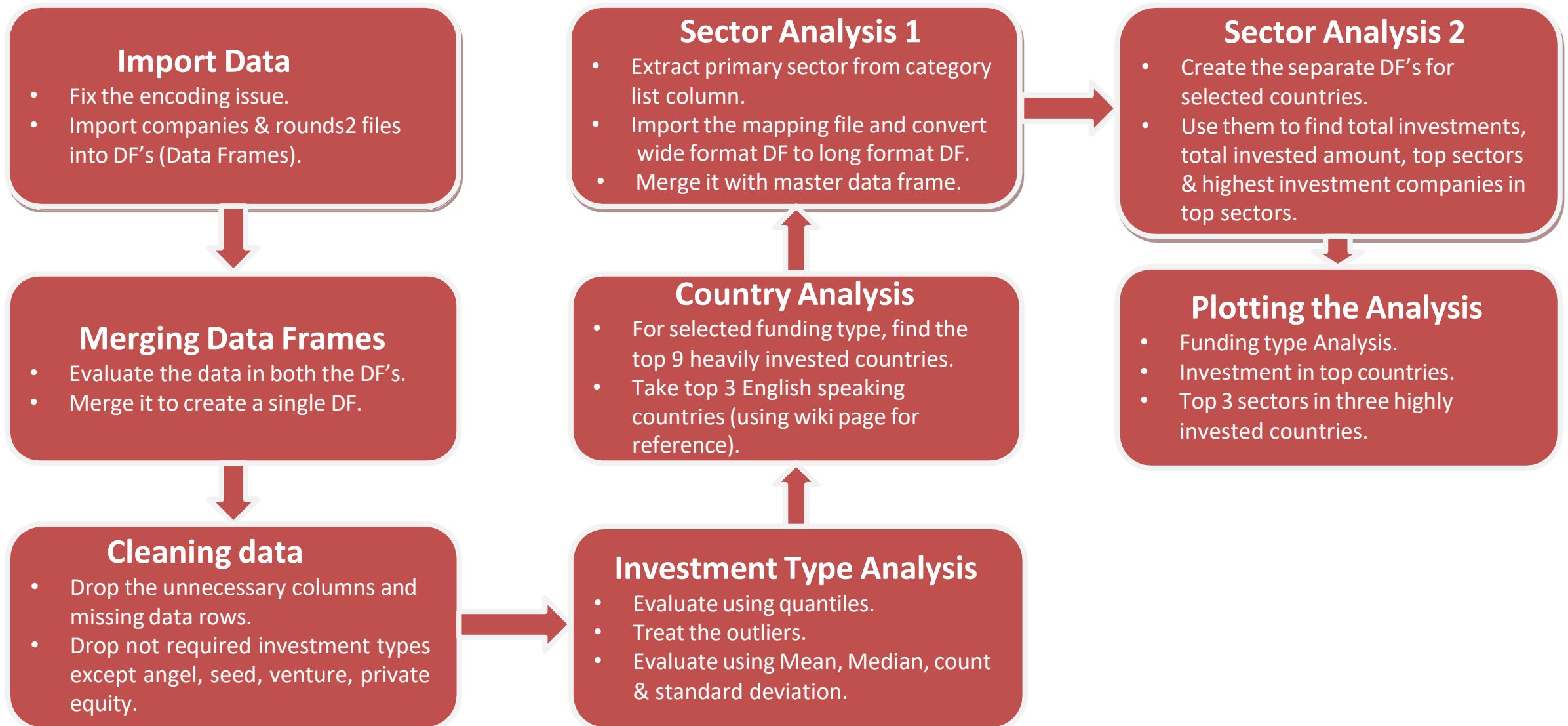
➤ Business Objectives:

- Identify the key factors that contribute to default risk, allowing Lending Club to improve its loan underwriting process and reduce the risk of default.
- Identify the largest amount of loss to the lenders are default borrowers.

➤ Goals of data analysis:

- By building a predictive model that accurately identifies the default risk for loan applications.
- Lending Club can reduce its default rates and increase profitability.

Overall Analysis Process



Data Understanding

- Dataset having some unwanted columns and records which needs to remove from dataset for the accurate analysis.
 - Null value columns
 - Unwanted columns
 - column whose values are less than 25% of data
- Data has some with the values.
- Dataset has some outliers.
- After analyzing all the statistical measures, the data seems to be positively skewed. But as there are no more outliers, considering 'Mean' as the representative value is right.

Data Cleaning and Manipulation

- Dropped some columns having null values. With the help of below code and removed unwanted columns as well.

```
for i in empty_cols:  
    print(df[i].isnull().sum())
```

- Excluded the data for Current customers.
- Annual income has some outliers, eliminated outlier data and added Issue_Loan_Month variable from issue_d variable.
- Extracted only numeric value from int_rate column and updated same column.
- Assumed the 'term' values will be numeric to plot the graph in appropriate manner.
- Created to different dataset one having "Fully Paid" and other "Charged Off" to differentiate the data.

Loan Status analysis

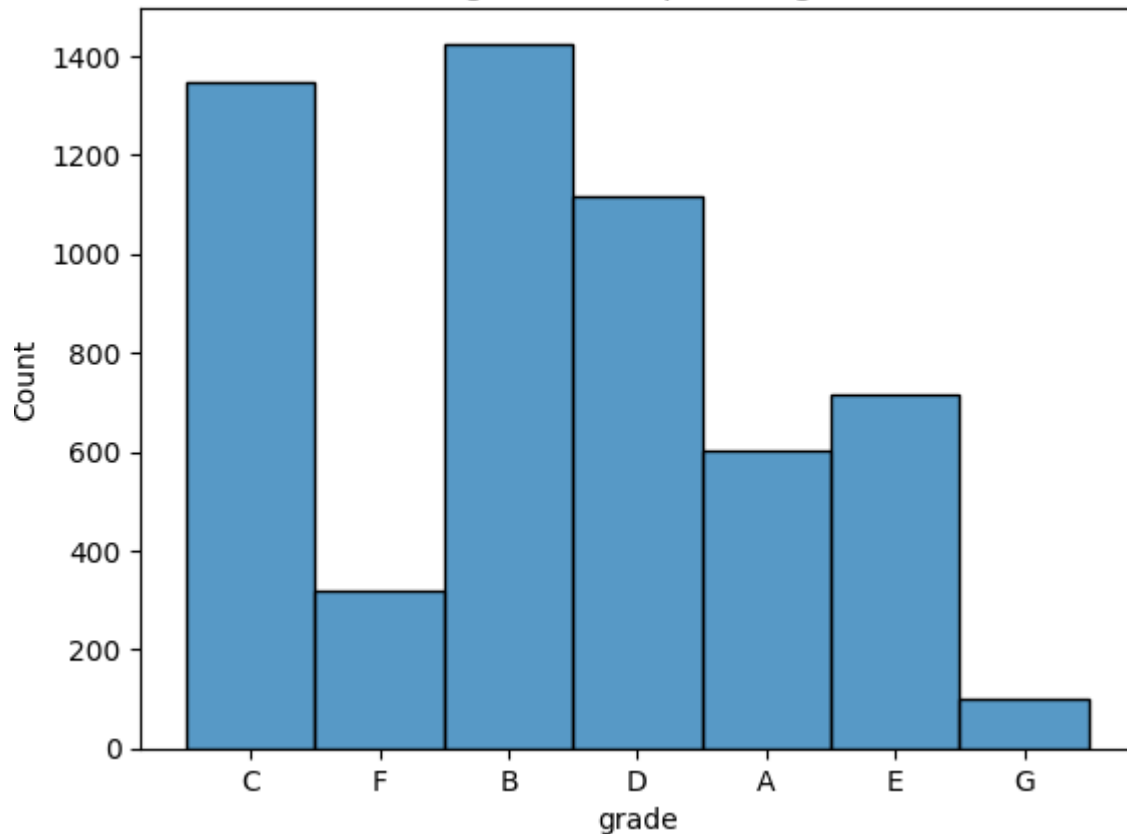
- LC wants to analyze the based on the loan status default customers, based on this we have filtered the data accordingly from dataset
- This data will be useful for plotting insights.

Loan Status	Total Invested Amount
Fully Paid	32937
Charged Off	5626

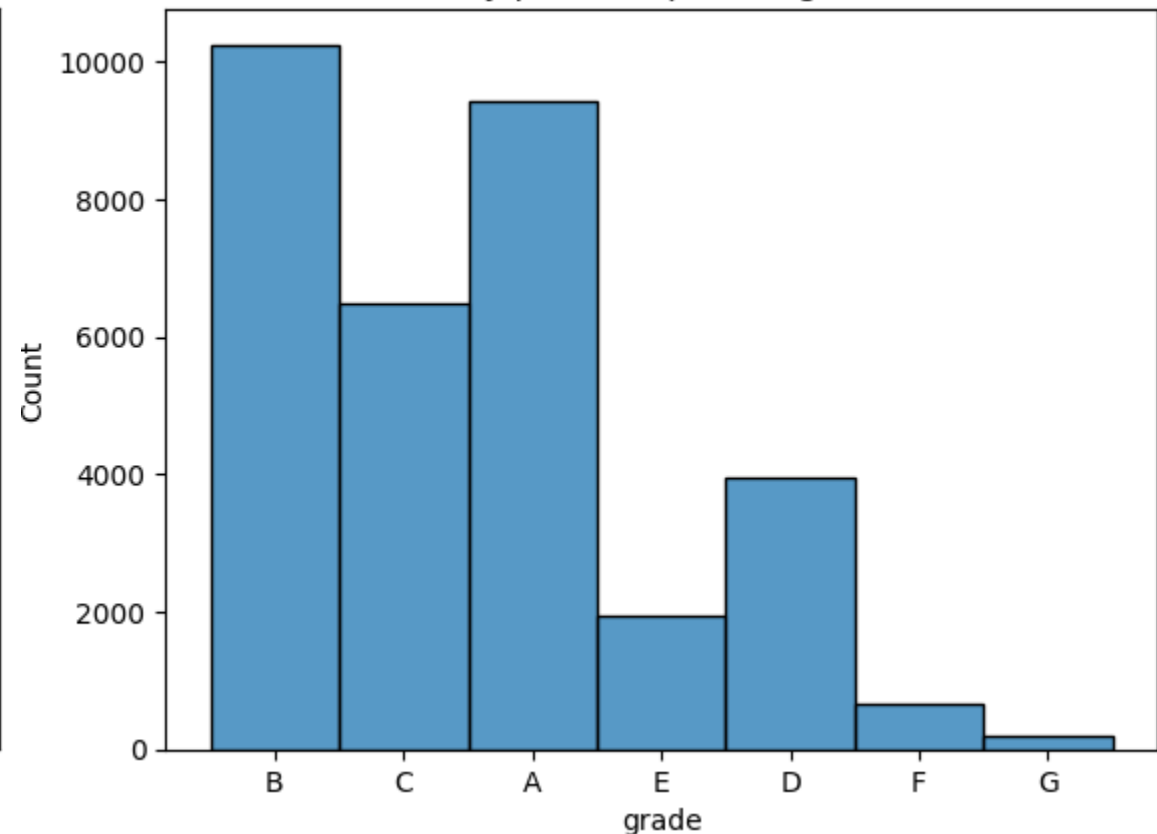
Data Analysis

- Histogram plotted for all the 'loan_amnt', 'annual_inc', 'grade', 'dti' against same to check the loan status category and found some below statistics.

Charged off hist plot for grade

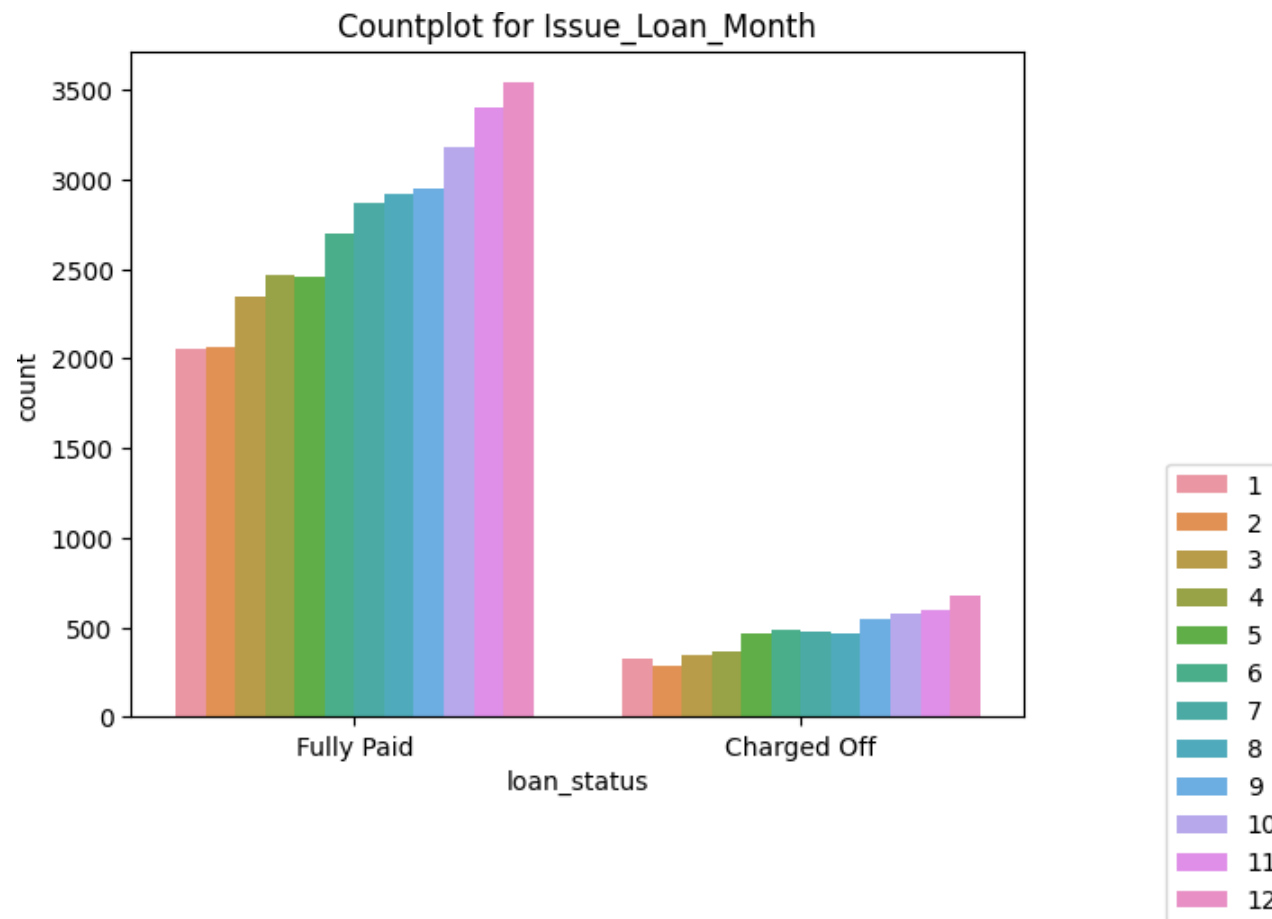


Fully paid hist plot for grade



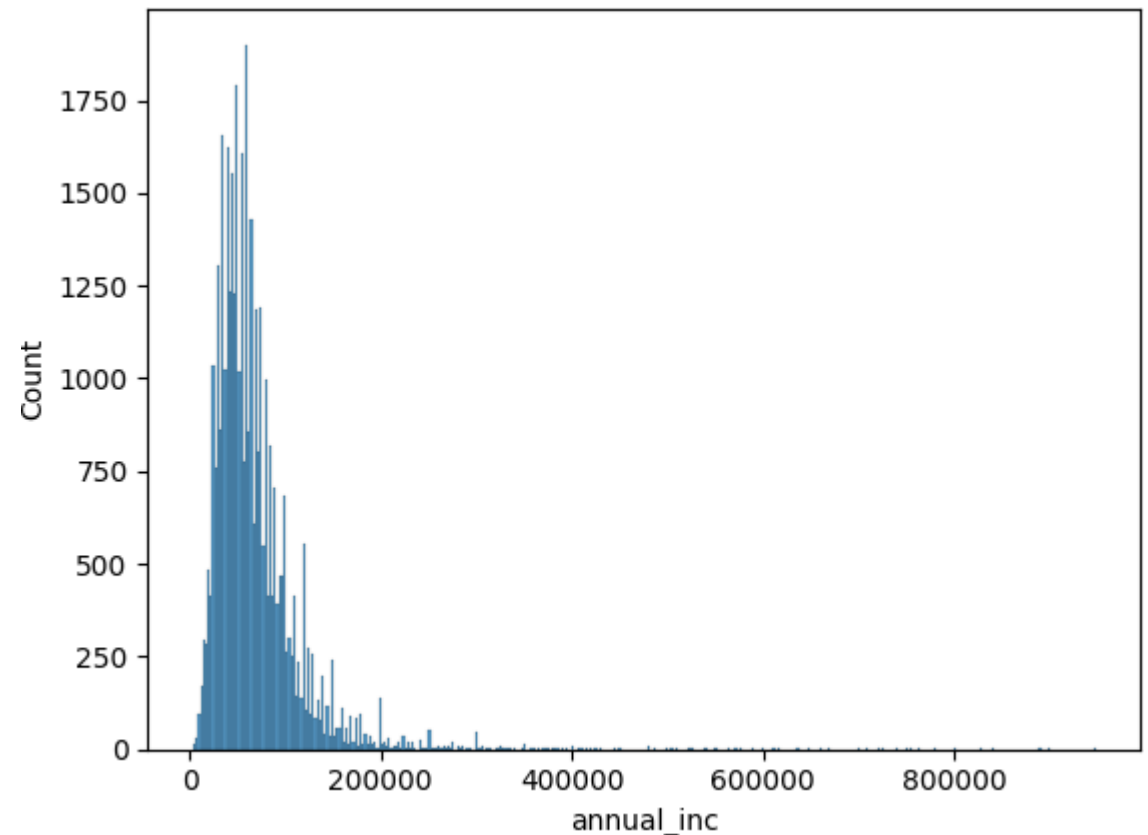
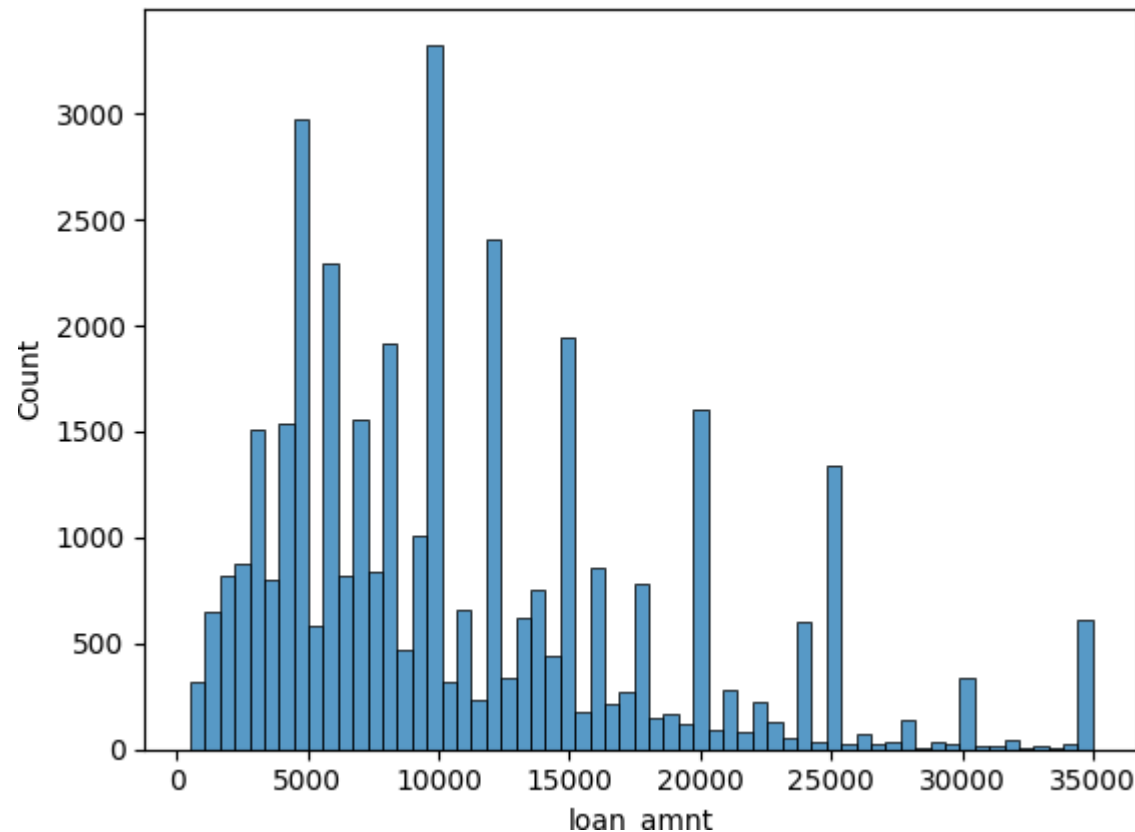
Data Analysis

- You can look at the trend for customers who are defaulters for issue loan month from JAN, NOV and DEC



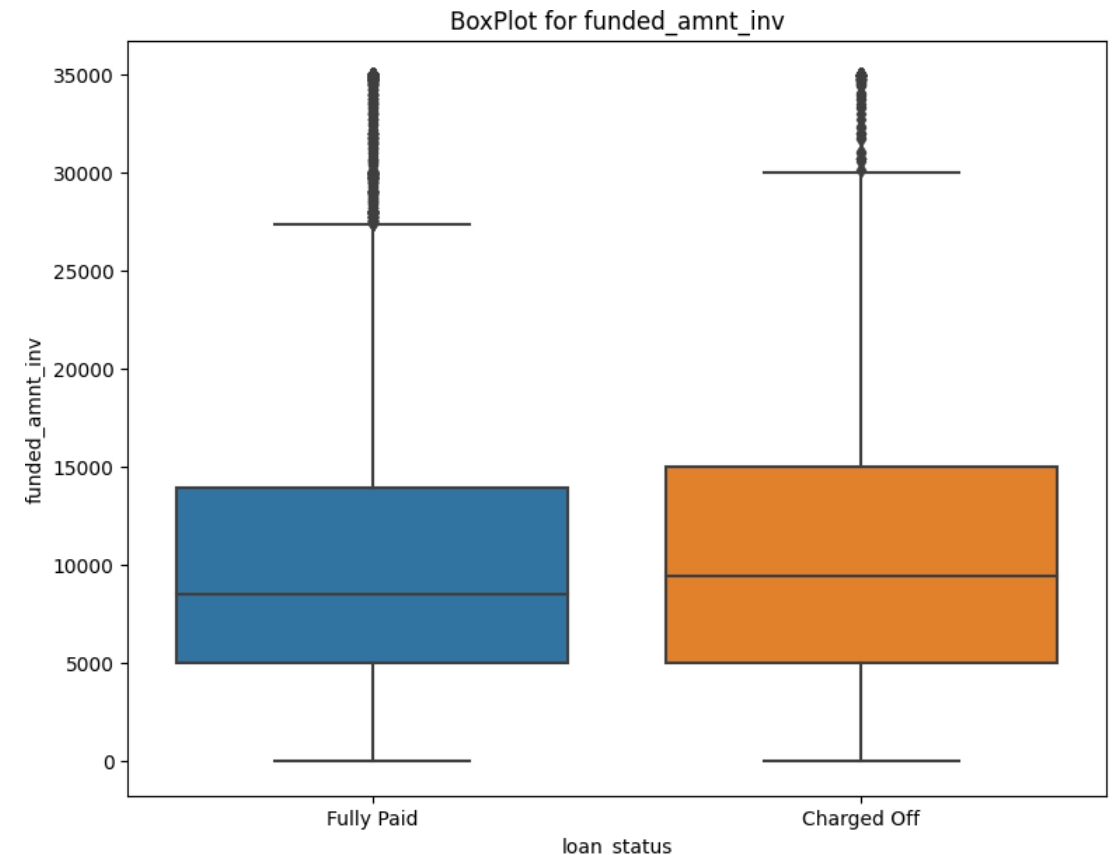
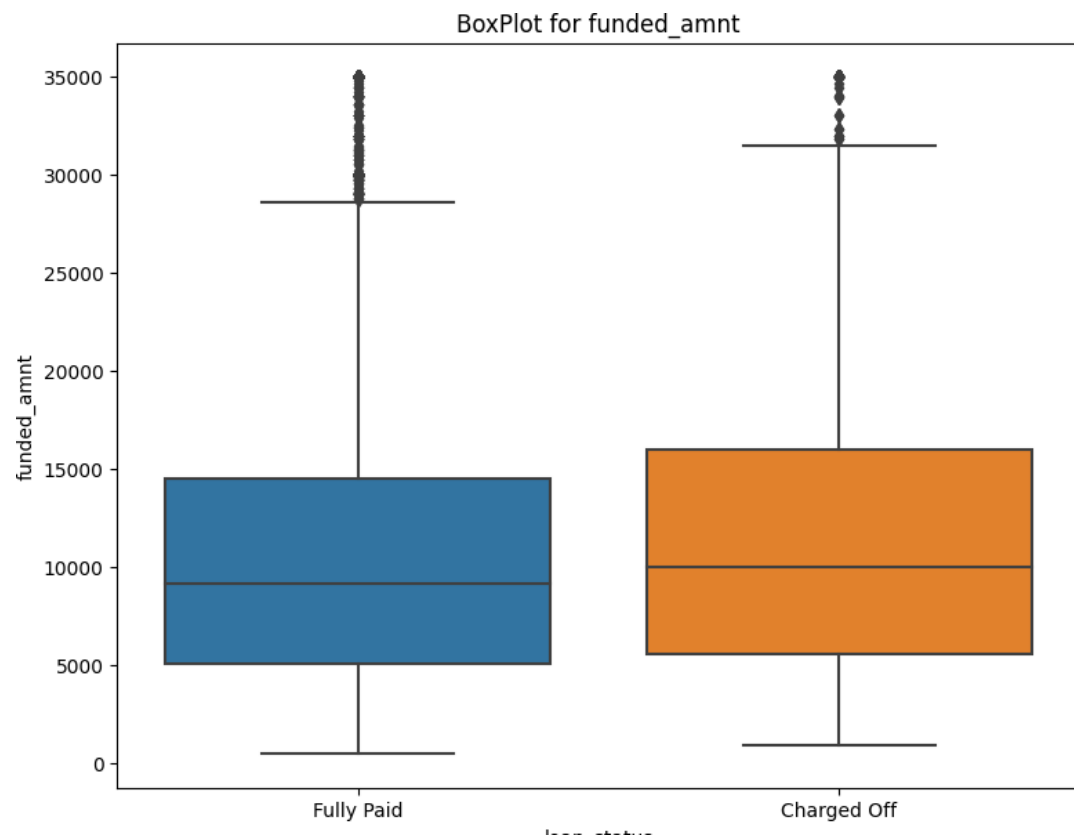
Data Analysis

- Loan avg loan amount is disbursing is from 5000 to 15000 and can view avg annual income.



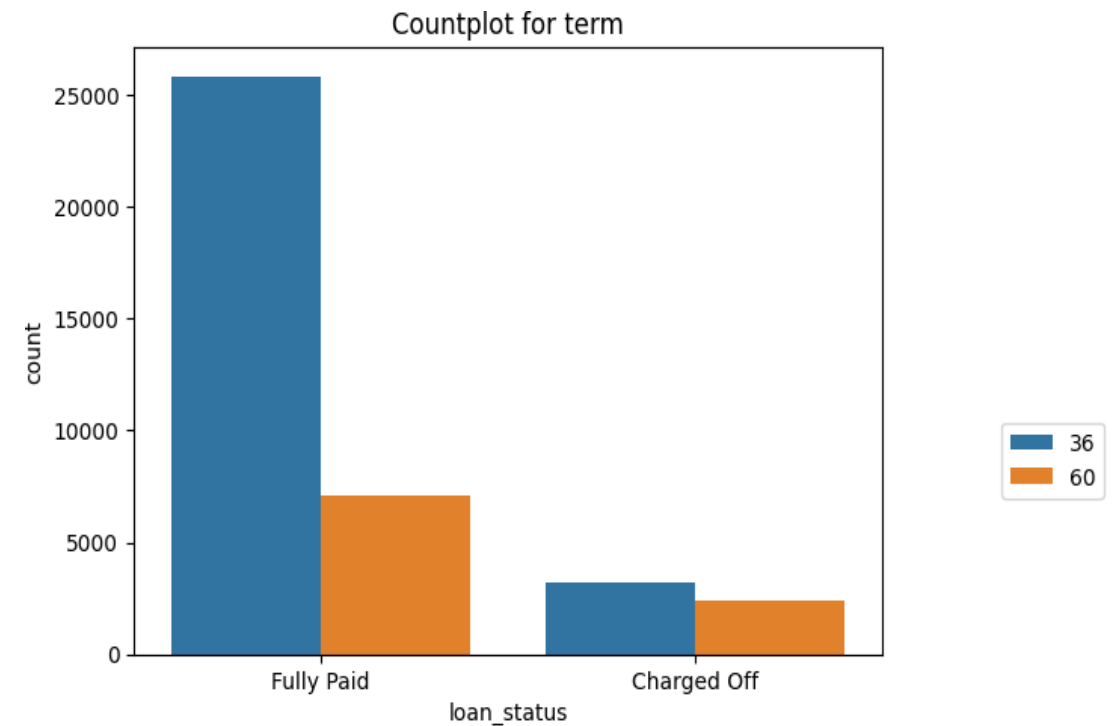
Data Analysis

- It has been observed that individuals who receive funding amounts in the third quartile (Q3) are more likely to default compared to the amounts approved by the companies for those who do default.



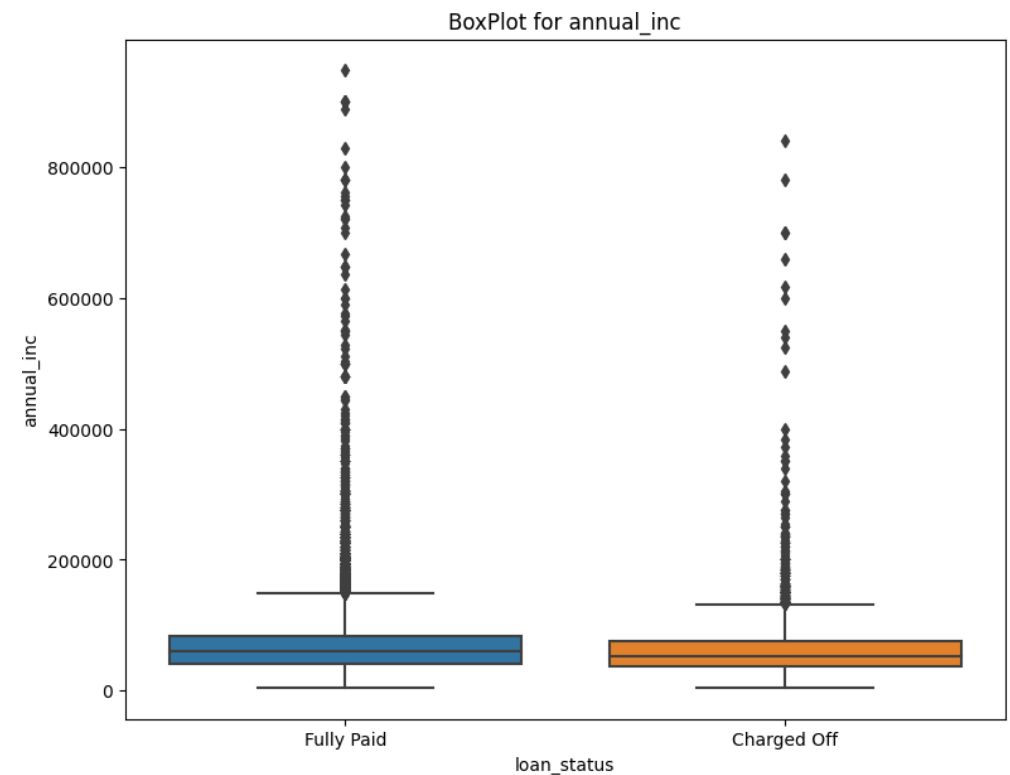
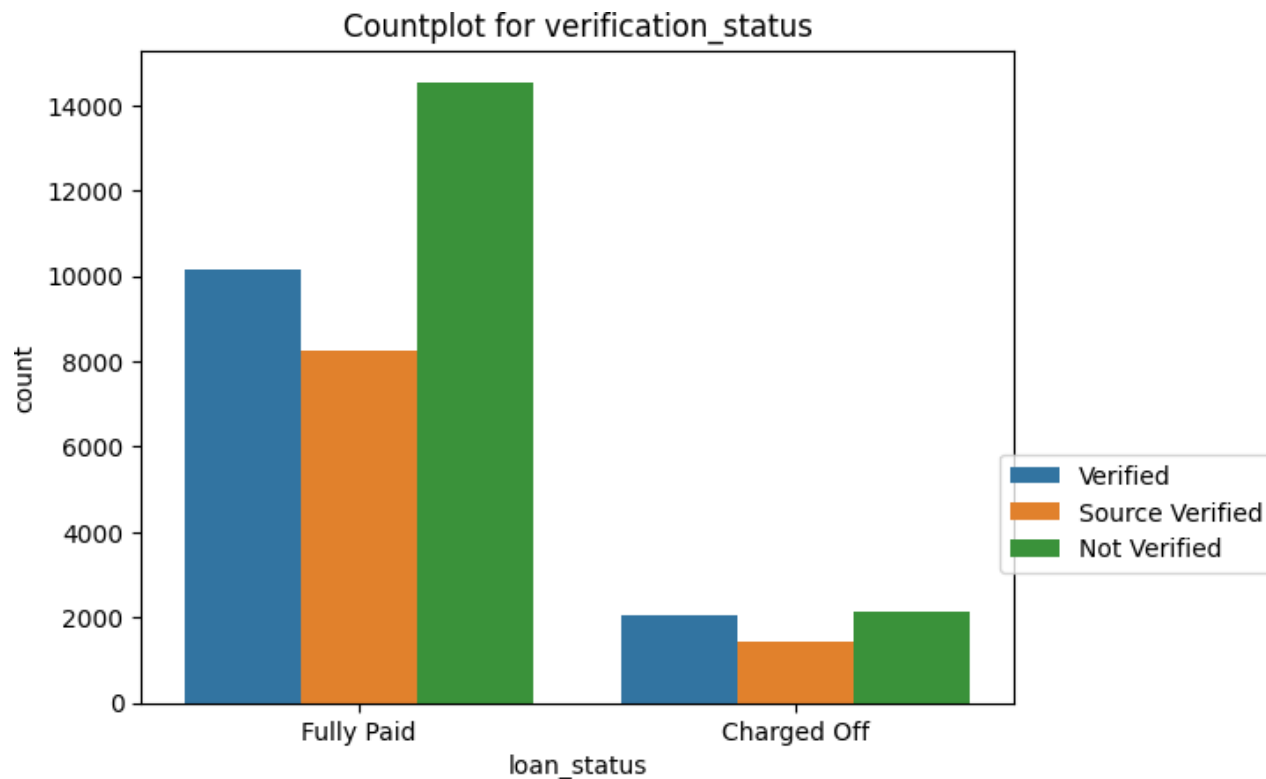
Data Analysis

- Customers with the highest interest rates are charged off more frequently, while loans with a 36-month term are paid off more often



Data Analysis

- In here we can see the source verified customers are paying the amount higher than other two categories and those who are annual income less the defaulters are more in number.



Recommendations

- Loans can be provided to customers who fall in grade A and ignore of Grade D customers.
- Source verified customers are less defaulters.
- The loan disbursed in Jan, Nov and month are significant defaulters.
- It is observed that the loan term of 36 months is typically paid loans.
- Loan avg loan amount is disbursing is from 5000 to 15000 is significantly high.