LENDING CLUB CASE STUDY

SUBMISSION

Group Names: Shyam Basa & Ritesh Rai

Project Introduction

- **Project brief:**

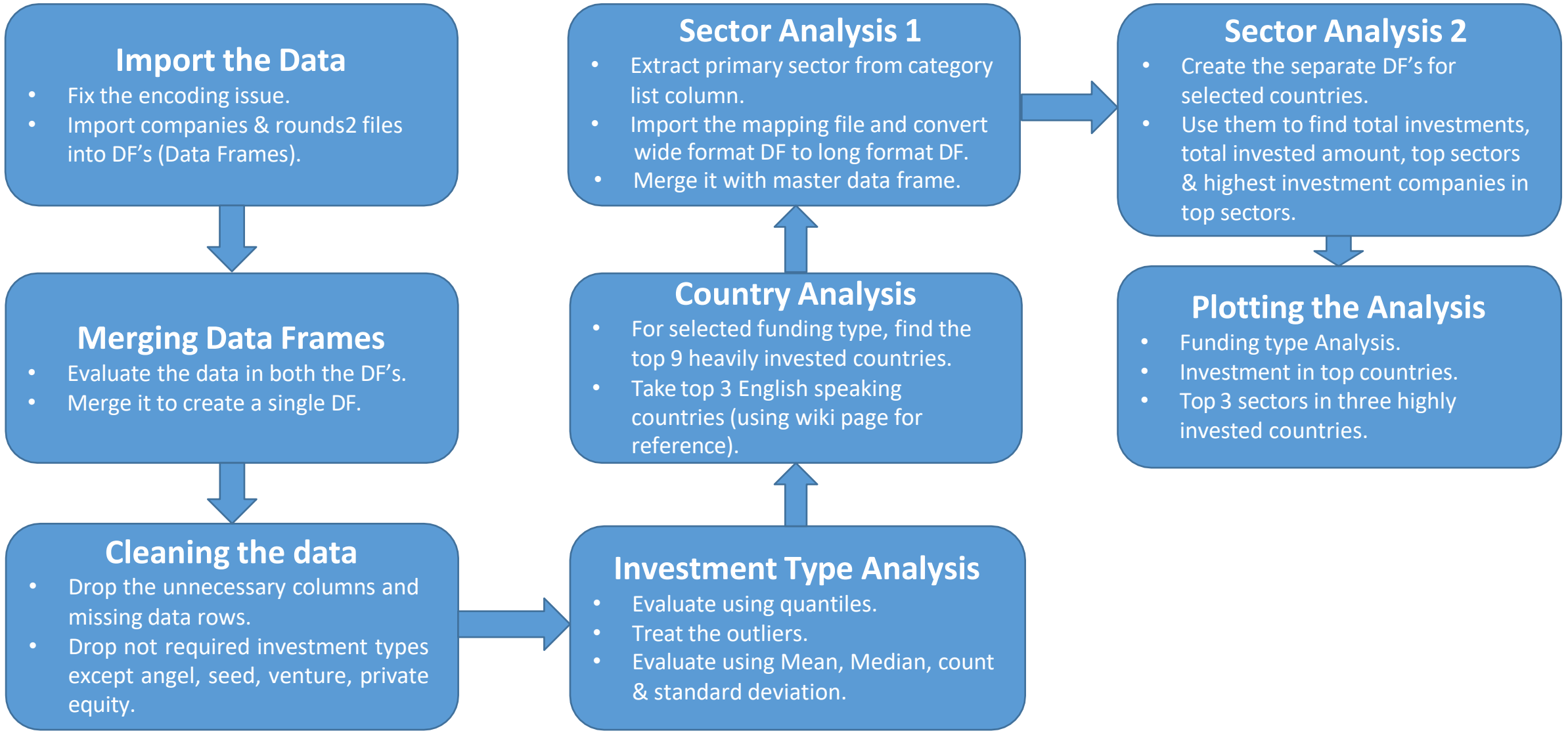
Lending club is an consumer finance company which specialises in lending various types of loans to urban customers.

  - 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 LendingClub 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.
  - LendingClub can reduce its default rates and increase profitability.

Flowchart of the Analysis Process



Data Understanding

- Dataset having some unwanted columns and records which needs to removed from dataset for the accurate analysis.
  - Null value columns
  - column who's values are less than 25% of data
  - Unwanted columns
- 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

- 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.

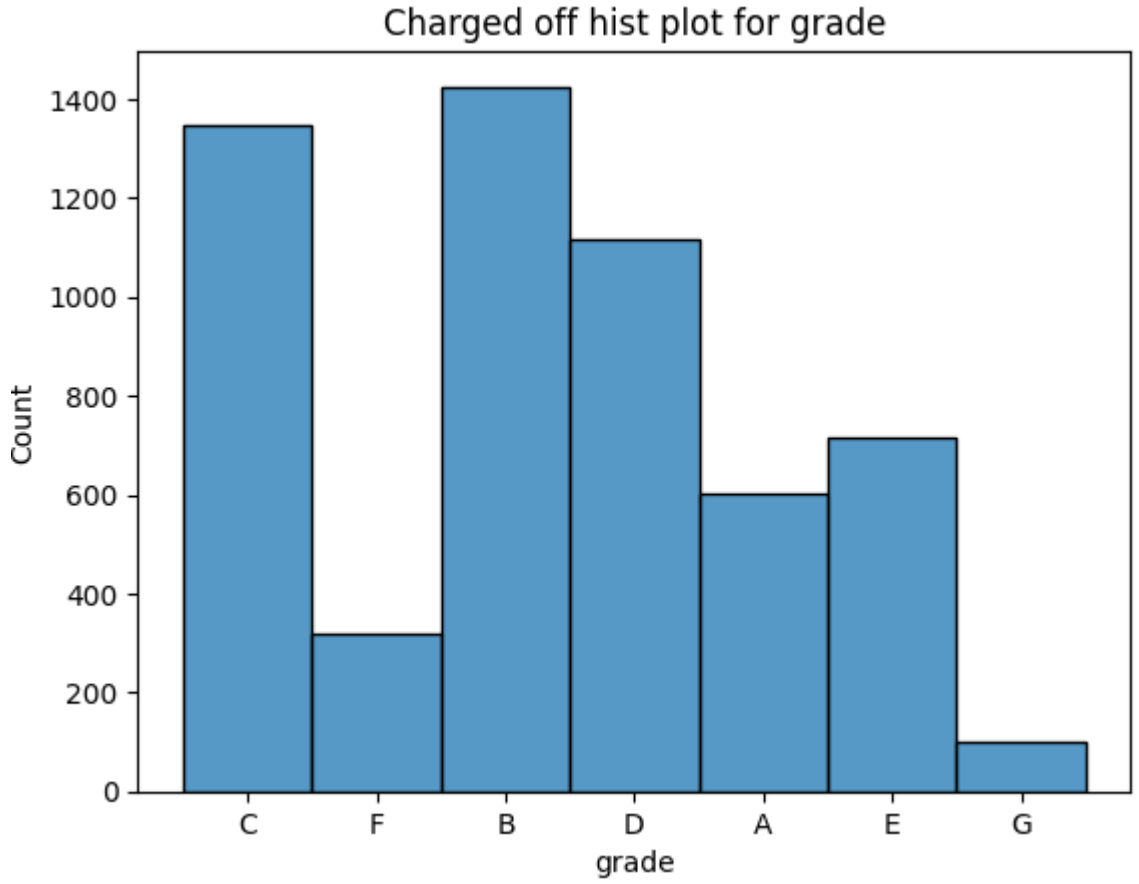
Loan Status analysis

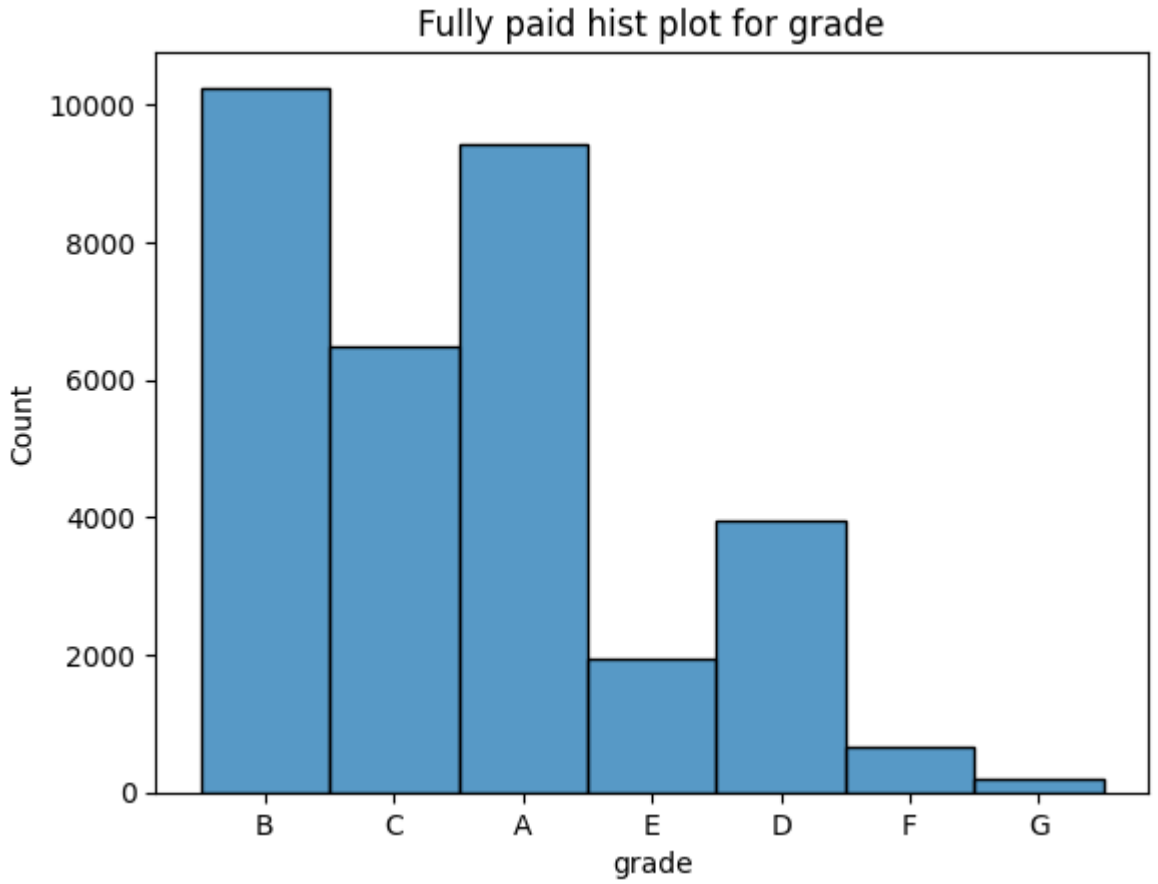
- LC wants to analyze the based on the loan status default customers, based on this we have filtered the data accordingly from dataset
- This data will be useful plotting insights.

Loan Status	Total Invested Amount
Fully Paid	32937
Charged Off	5626

Data Analysis

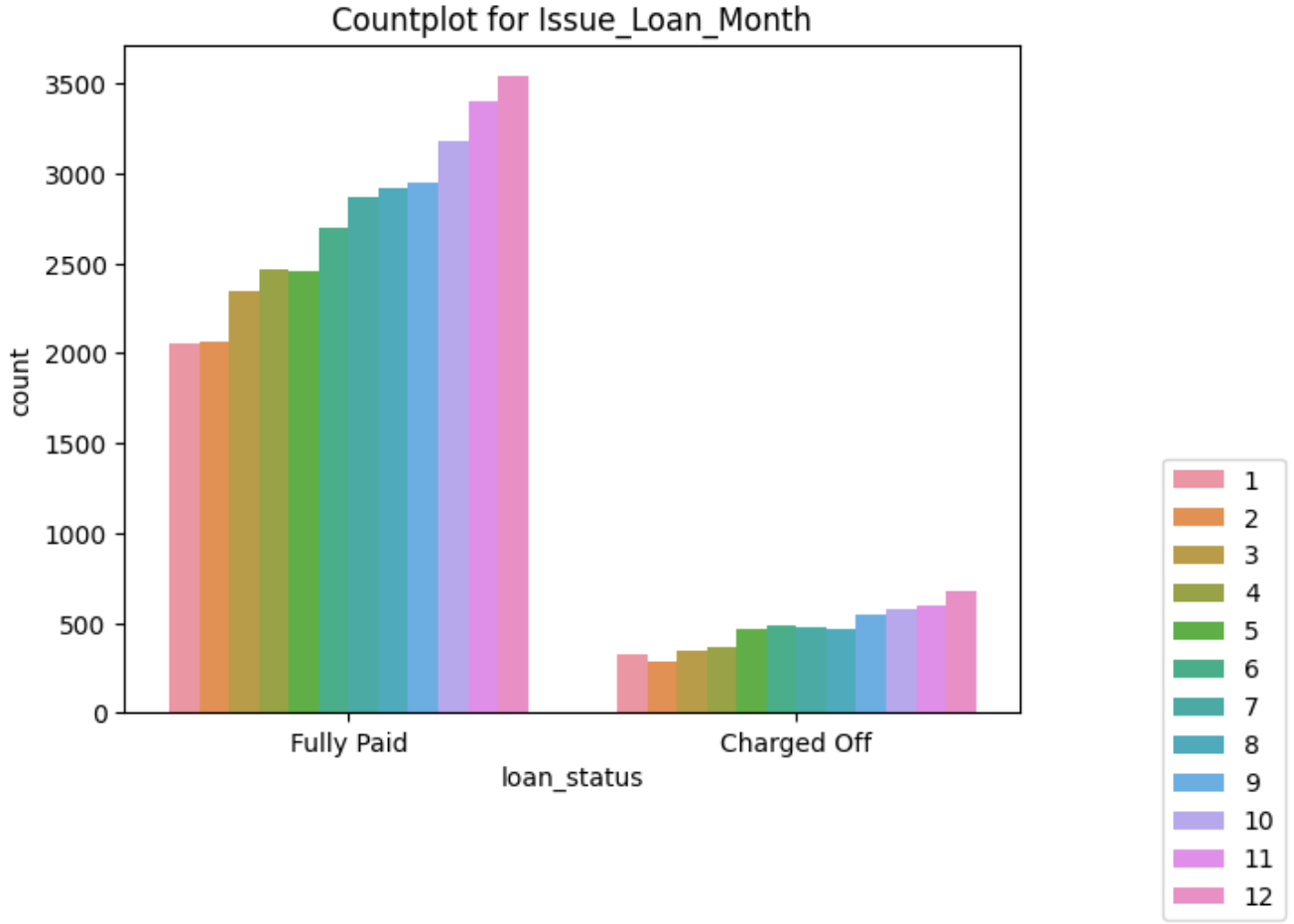
- Histogram plotted for all the 'loan_amnt','annual_inc','grade','dti' against same to check the loan status catogory and found some below statistics.





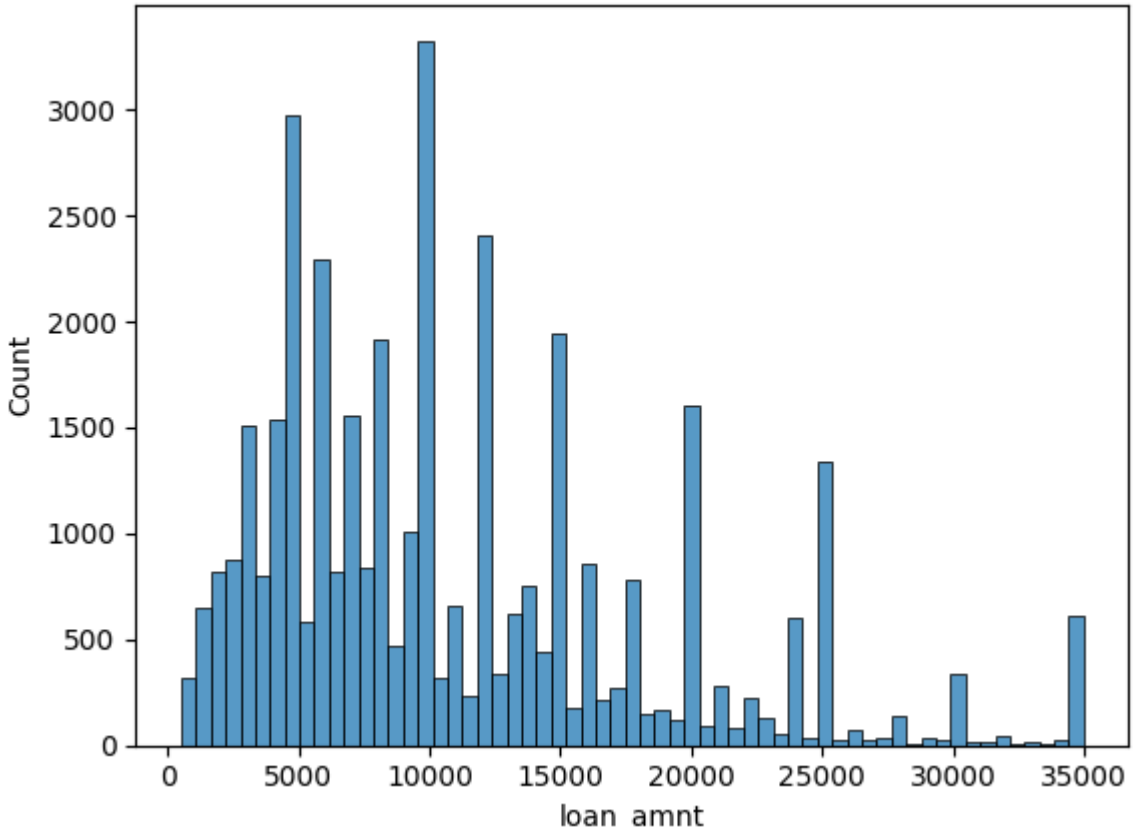
Data Analysis

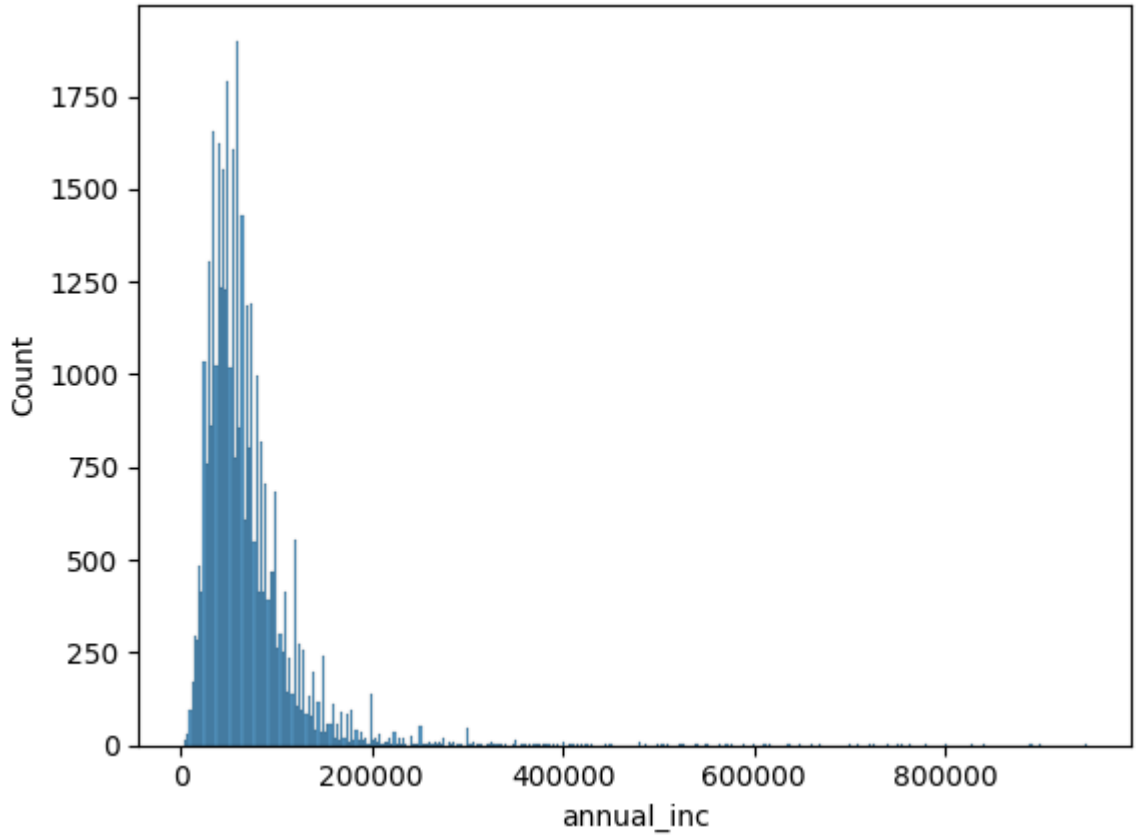
- You can look at the trend for customers who are defaulters for issue loan month from JAN, NOV and DEC



Data Analysis

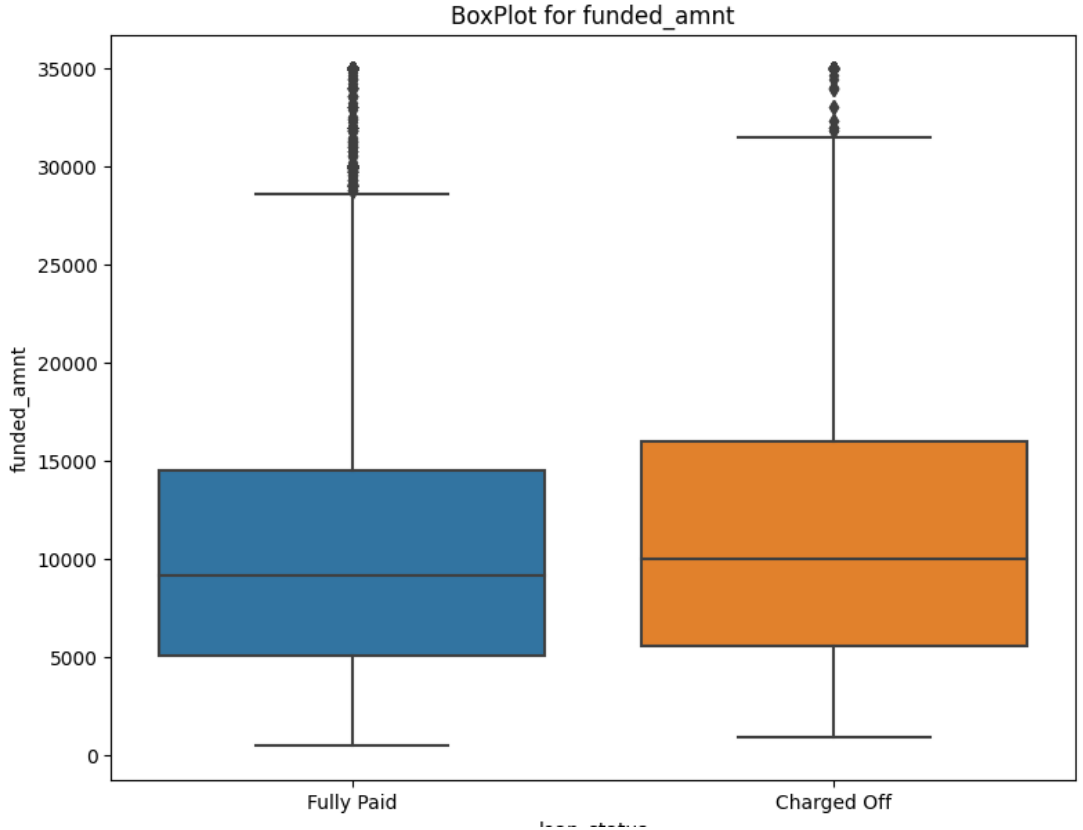
- Loan avg loan amount is disbursing is from 5000 to 15000 and can view avg annual income.

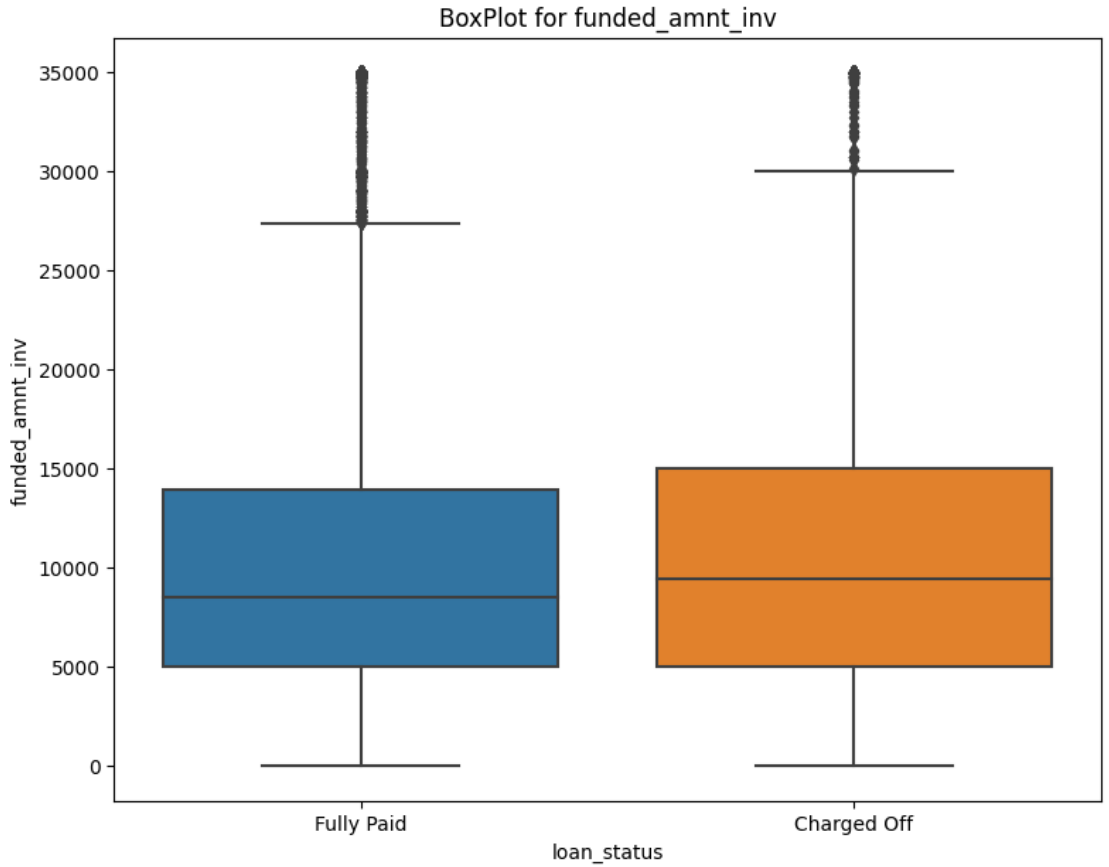




Data Analysis

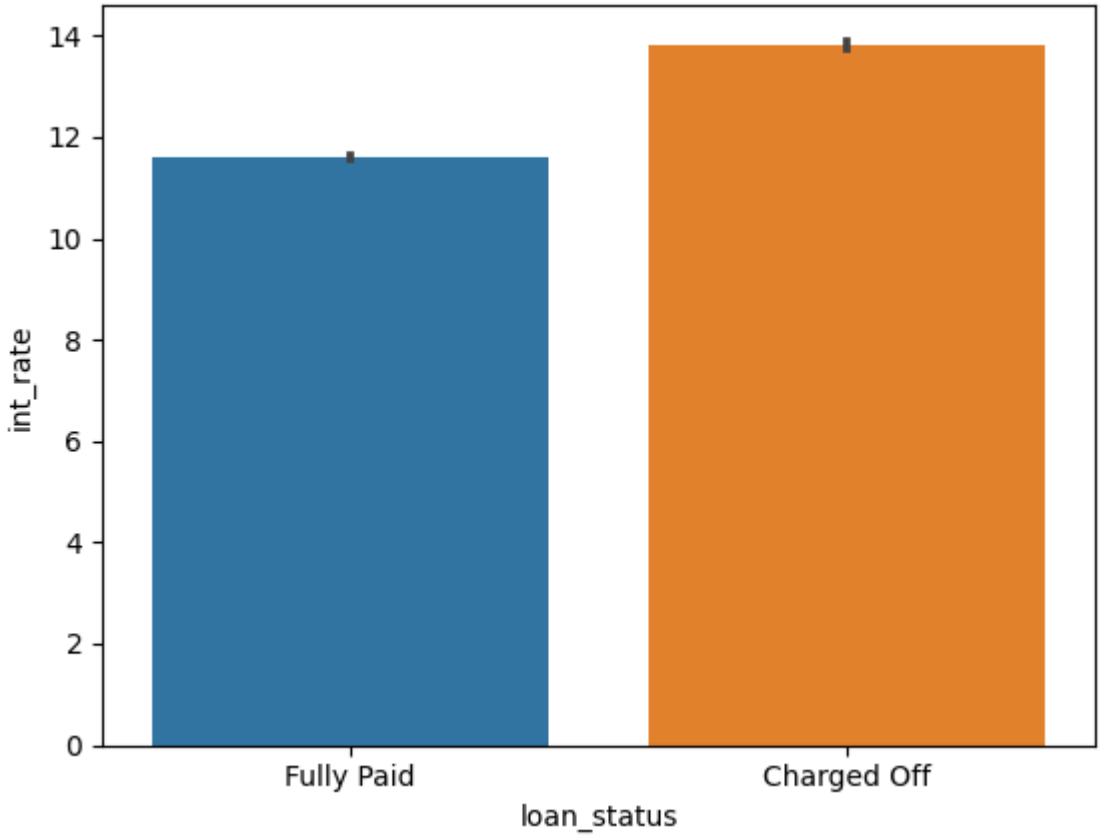
- It has been observed that whose funded amount is Q3 are more defaulters the amount approved by the companies those who are defaulters.

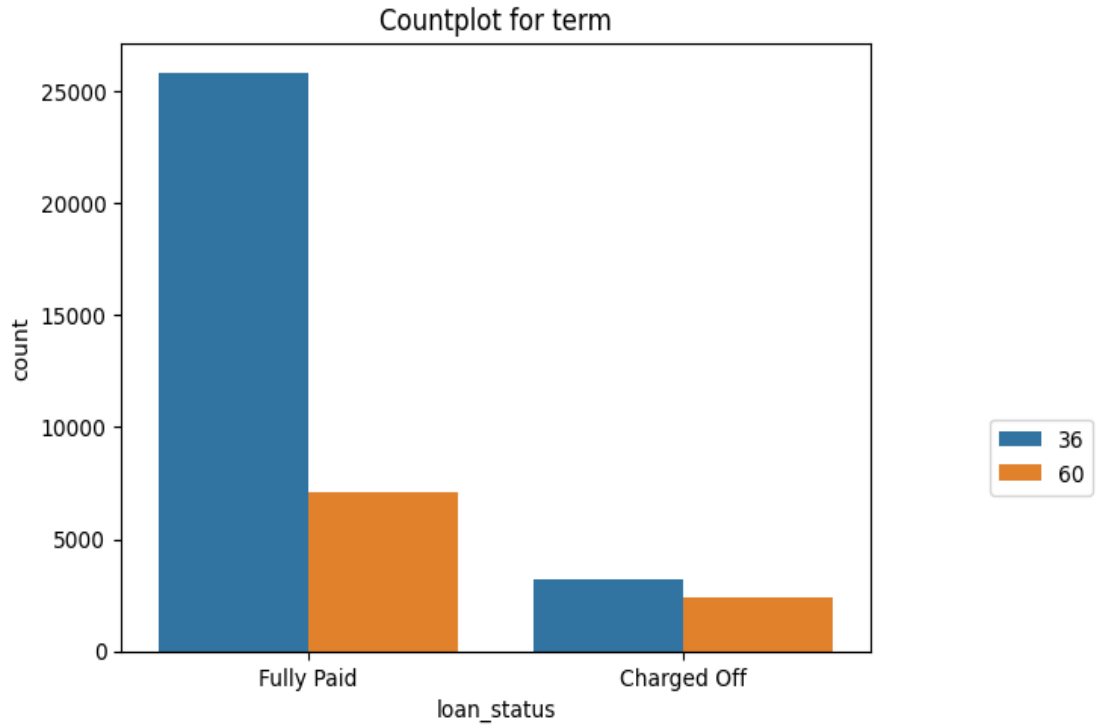




Data Analysis

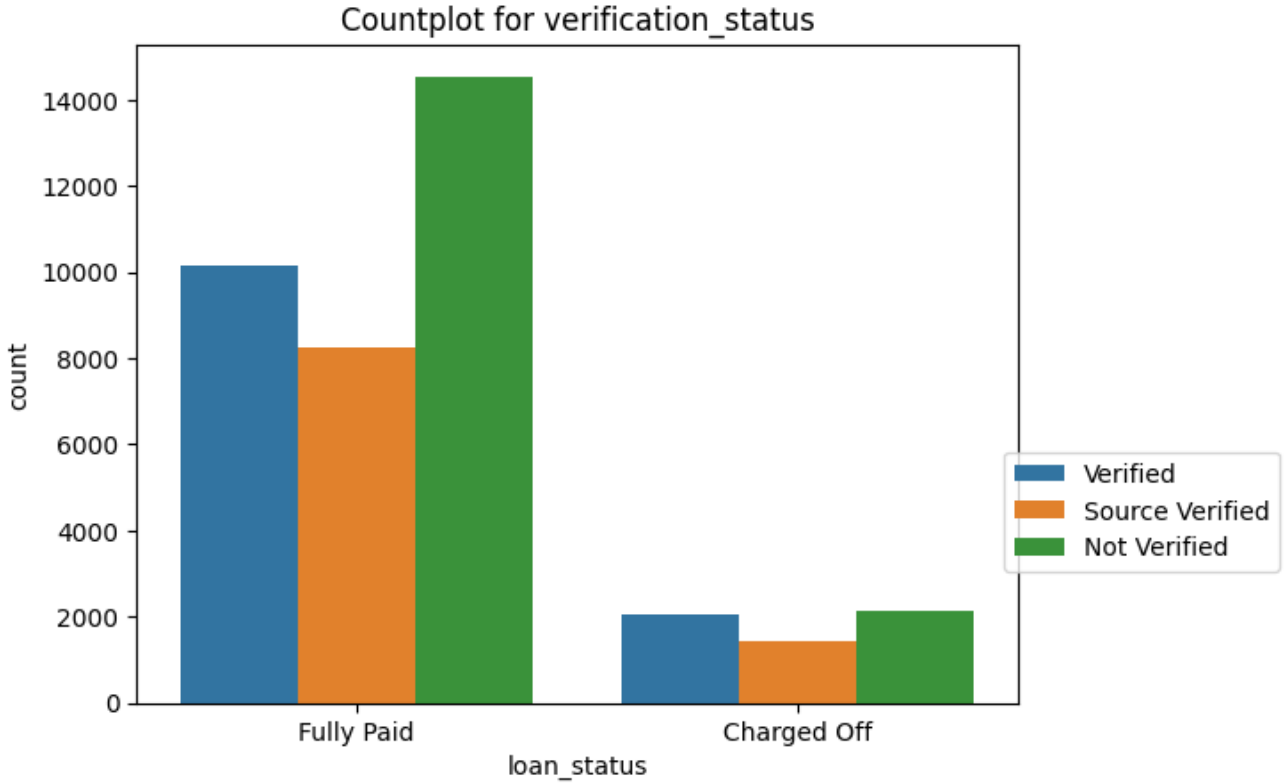
- Highest interest rate customers are charged off more and we 36 months term loans are get paid off more in number.

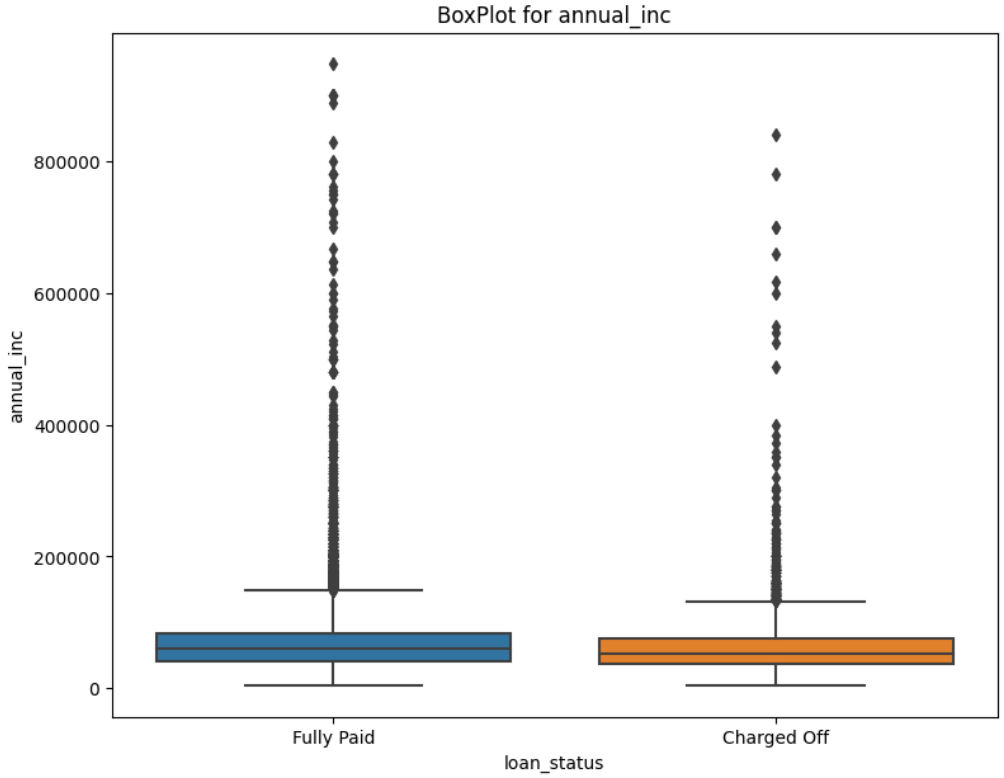




Data Analysis

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





Conclusions/Recommendations

- We can provide the loans to customers who are fell 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 are typically paid loans.
- Loan avg loan amount is disbursing is from 5000 to 15000 is significantly high.