Lending Club Case Study Submission

MOHIT PORWAL

Problem statement

To analyse the dataset containing details about past loan applicants using EDA to investigate how customer attributes and loan attributes can affect the loan default.

Focus of this case study:

- Provides an idea about how exploratory data analysis can be used in solving real life business problems.
- It also develops a basic understanding of risk analytics in banking and financial services.
- How the data can be used to minimize the loss of money while lending it to customers.
- It improves our understanding of the data visualization techniques.
- Get an understanding for which variables are important, view summary statistics, and visualize the data

Business Understanding

You work for the `LendingClub` company which specialises in lending various types of loans to urban customers. 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

The data given contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

> When a person applies for a loan, there are two types of decisions that could be taken by the company:

> 1. `Loan accepted`: If the company approves the loan, there are 3 possible scenarios described below:

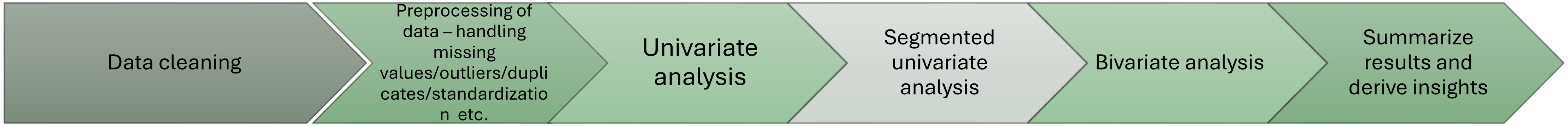
- `Fully paid`: Applicant has fully paid the loan (the principal and the interest rate)

- `Current`: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

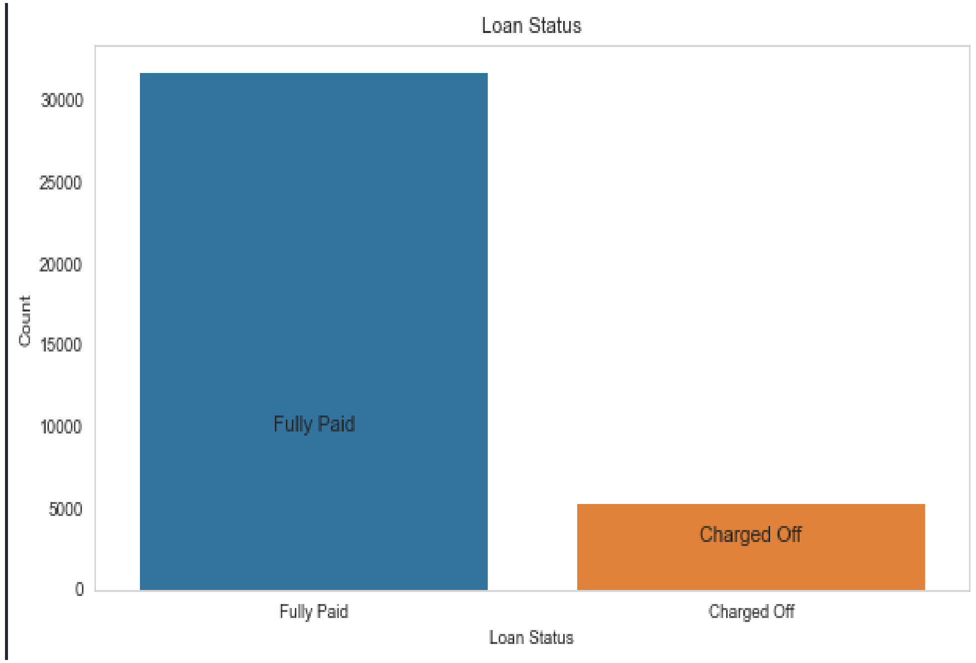
- `Charged-off`: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

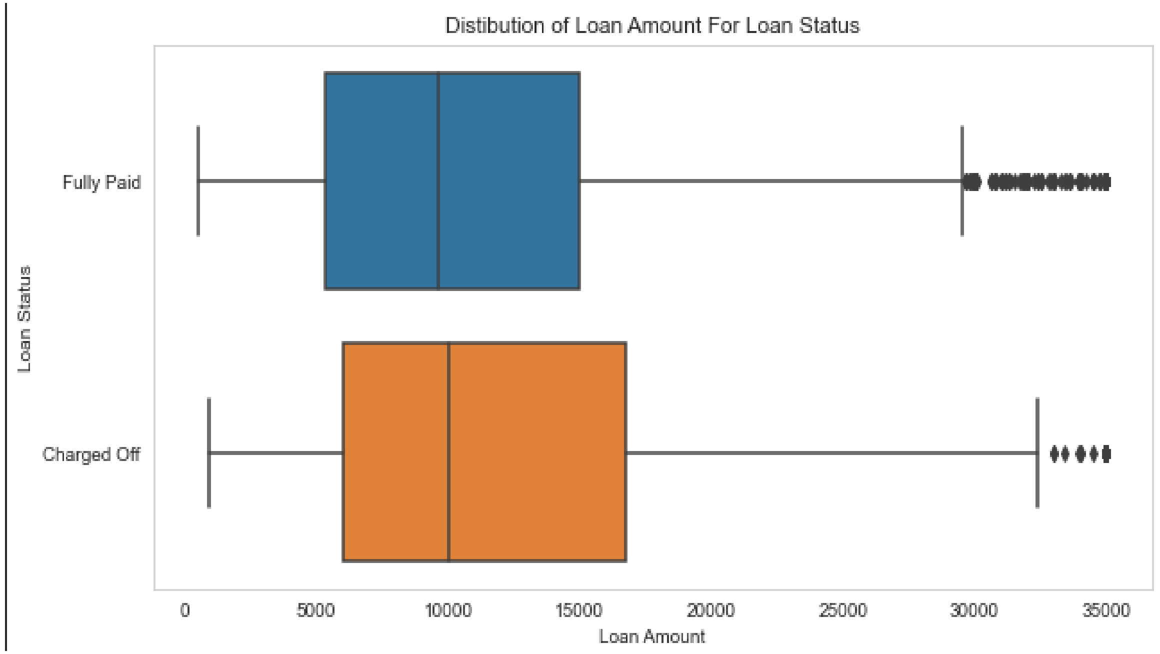
> 2. `Loan rejected`: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Analysis flow



Analysis – Overall Loan Status

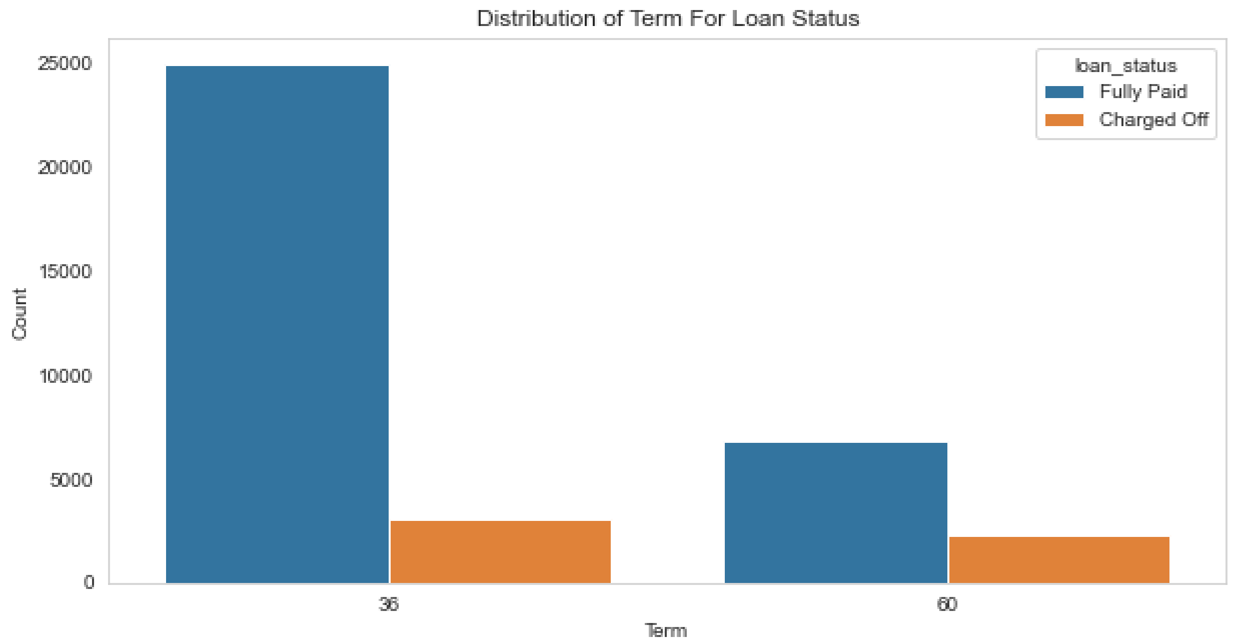


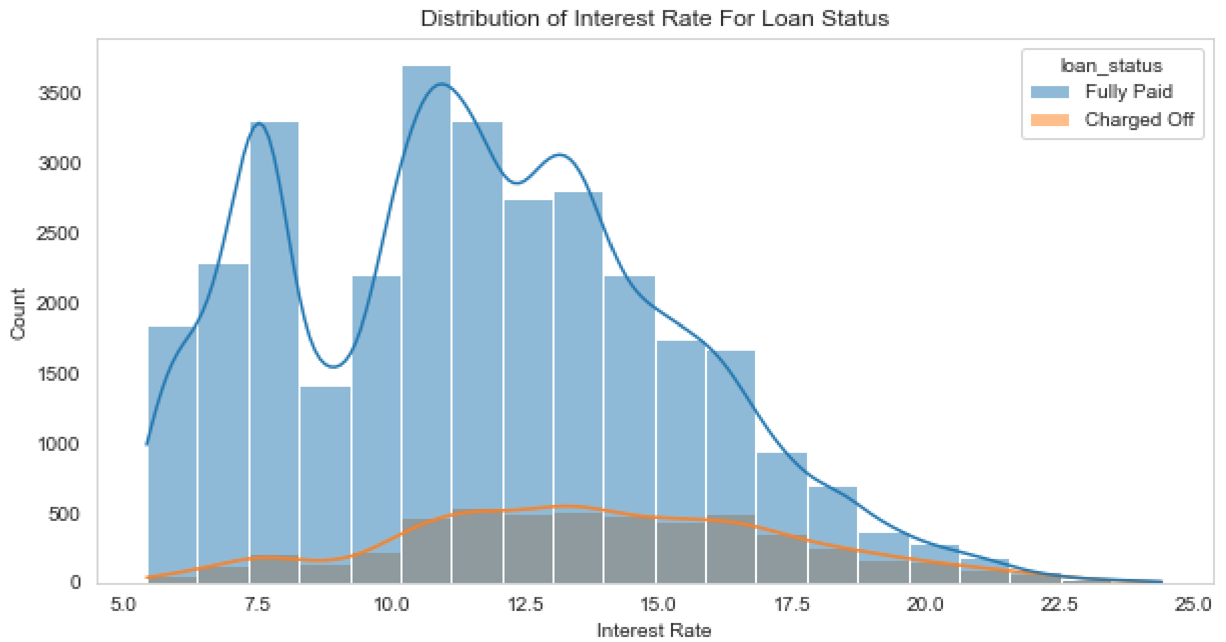


- **Loan Status: Approximately 14% of loans are defaulted.**

- **Loan Amount:** It ranges from 500 to 35000 with a median of 10000. Very few clients have taken large loans and larger it goes we have higher chance of defaulting.

Analysis – Default by term and interest rates

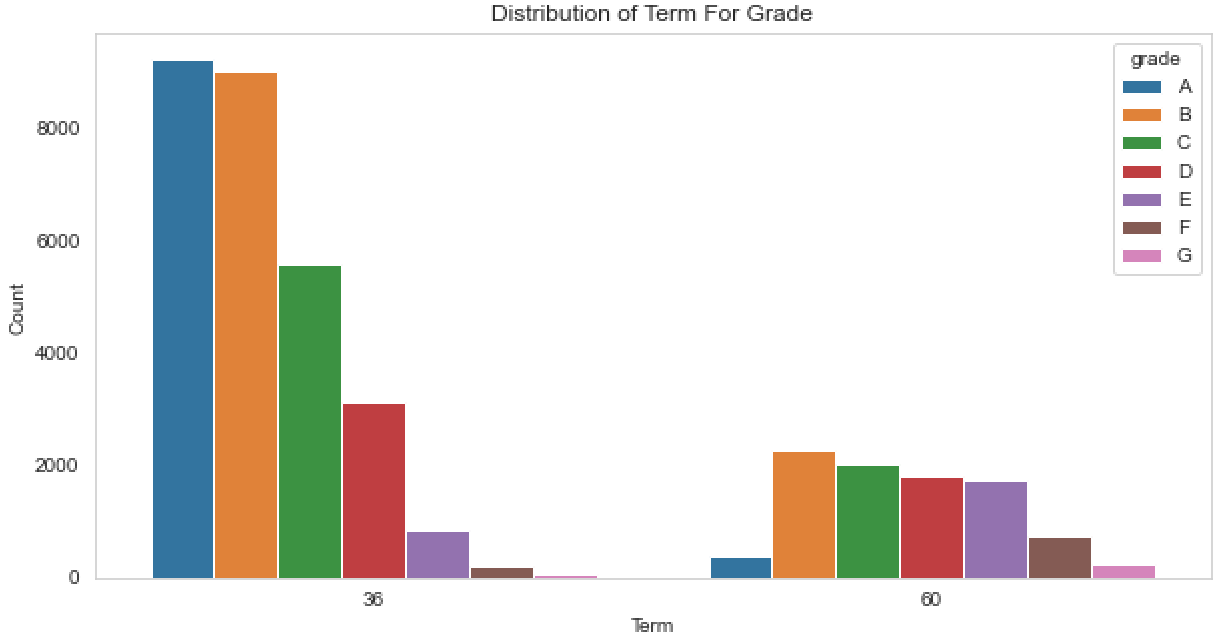


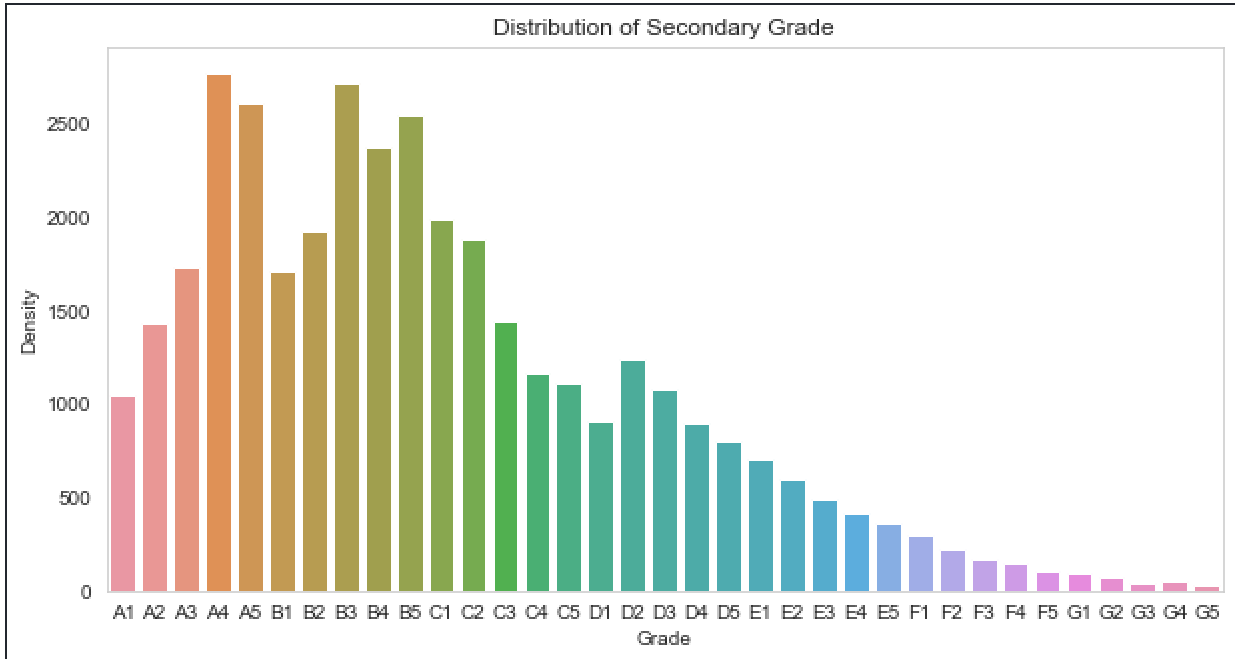


- **Loan Term:** The loan term of 36 months are much more than 60 months and have lower chance of defaulting.

- **Interest Rate:** The count of loan taken varies with interest rate showing peak around in 5-15 bracket and decreasing slowly where as the chance of defaulting increases with interest rate.

Grade and Sub-Grade

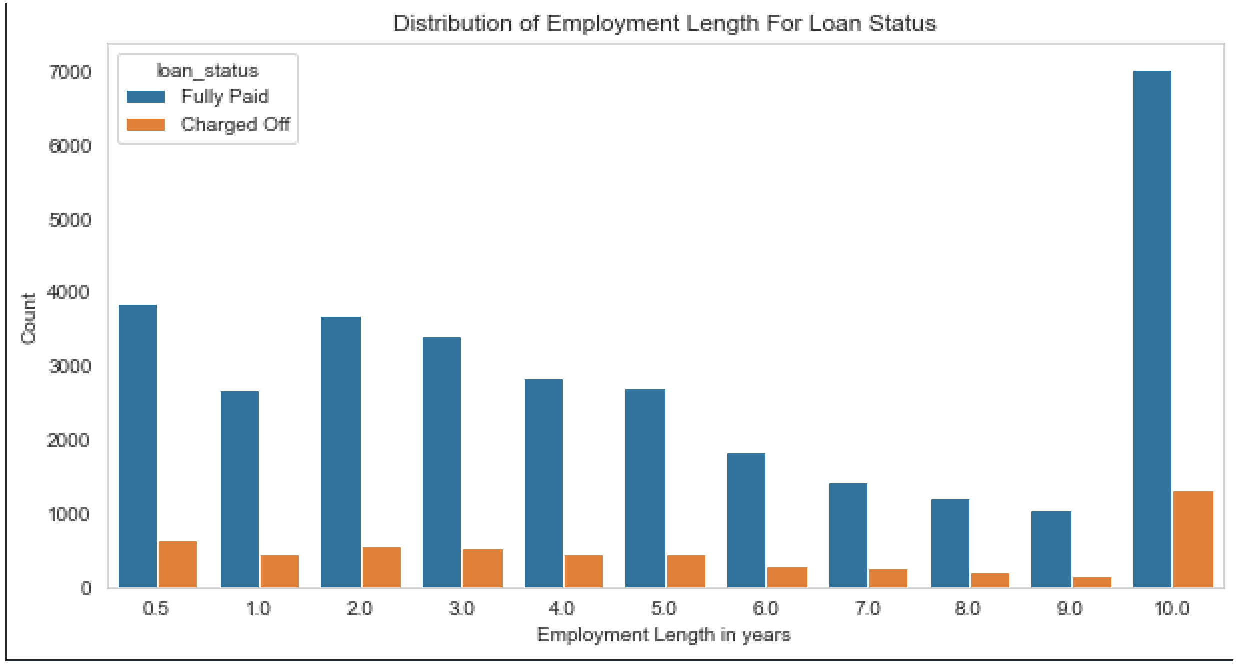


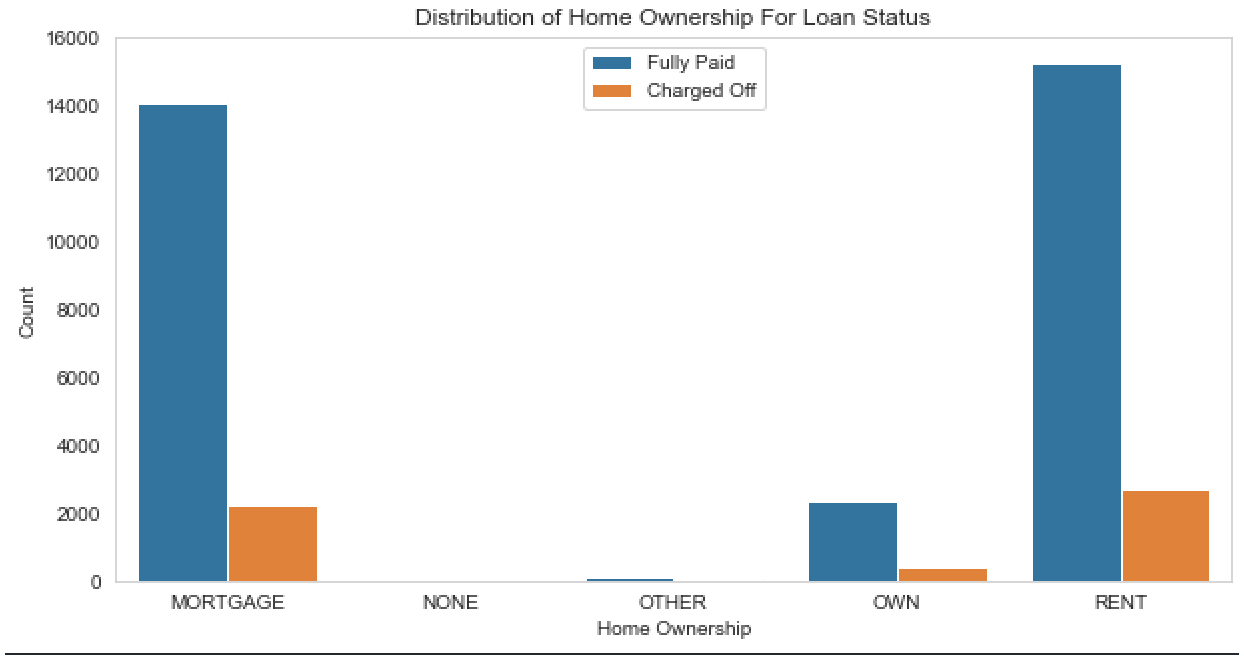


- **Grade:** Majority of the loans approved are of higher-grade quality. 60-month term loans have larger number of lower grade loans with high risk of defaulting.

- **Sub Grade:** This provides more insight that the loans within grade are more skewed towards lowered sub grades.

Employment Length & Homeownership

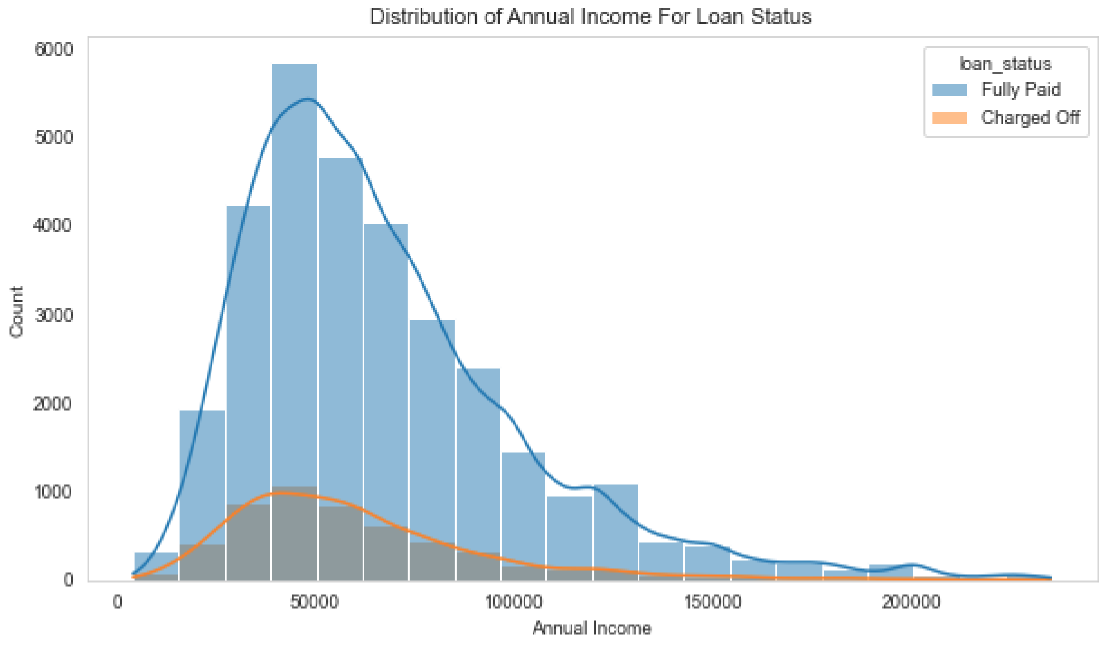


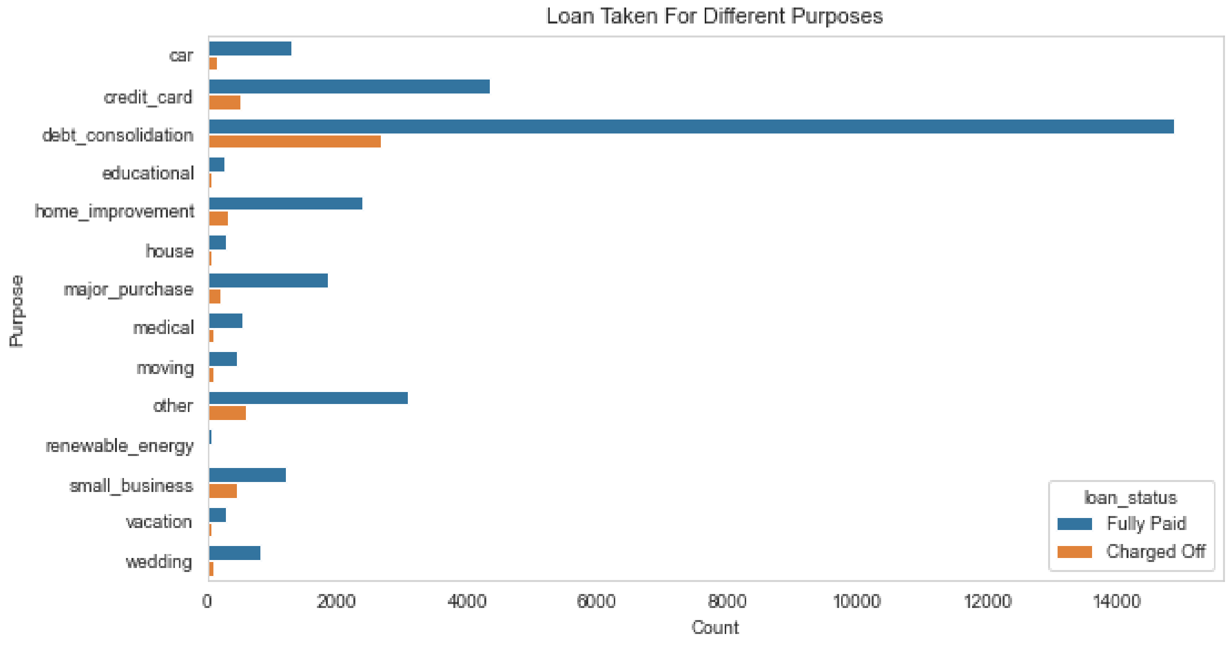


- **Employment Length:** Majority of clients have 10+ years of experience and has highest number of defaulted loan.

- **Home Ownership:** Majority of clients are lacking ownership of any property and are on rent or mortgage and have a higher chance of defaulting.

Annual Income & Purpose

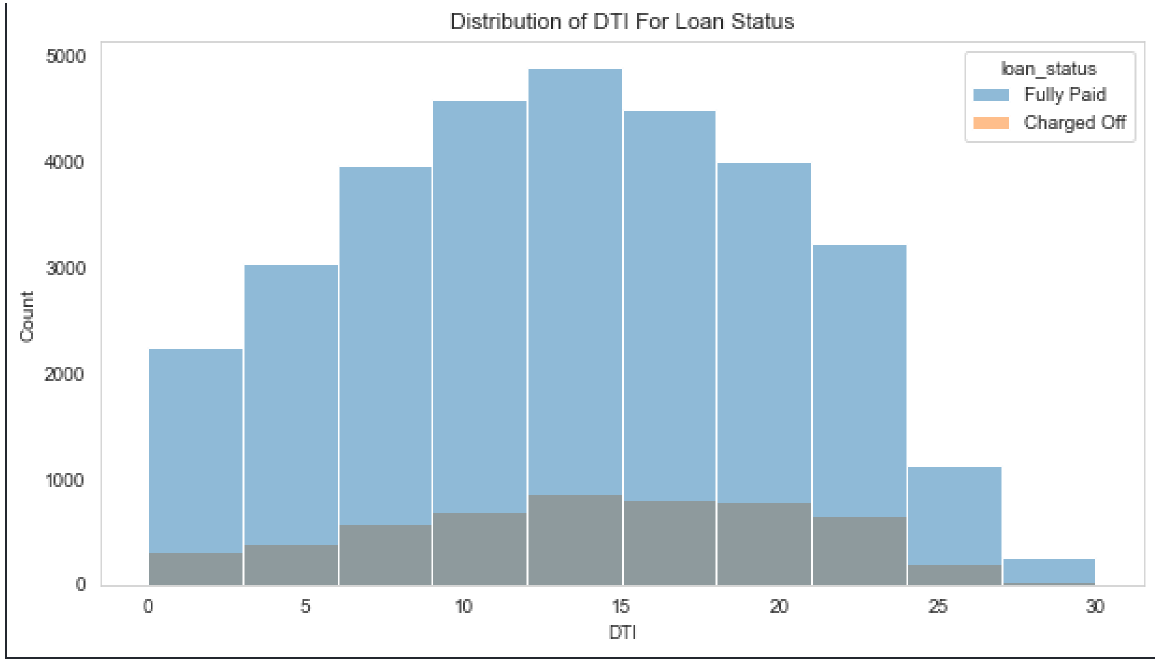


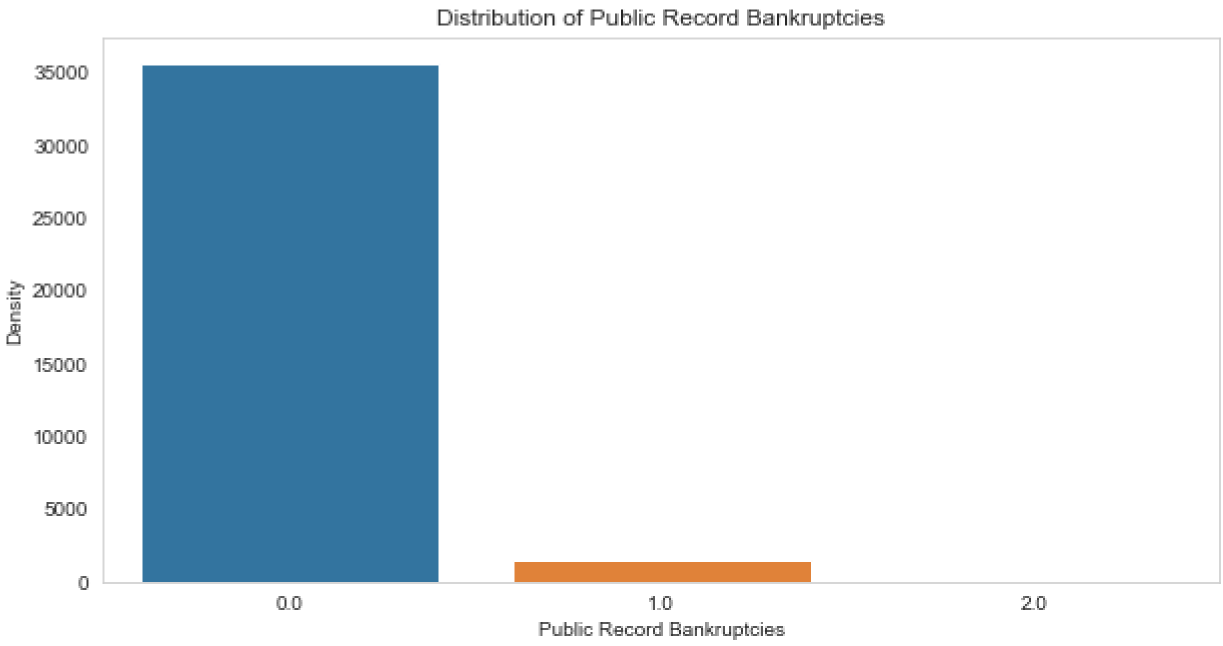


- **Annual Income :** The Majority of clients have low annual income compared to rest and income lower than 50k has higher chance of defaulting.

- **Purpose:** Maximum no. of loans are for debt consolidation, followed by credit card payment. Whereas the debt consolidation has highest fully paid loan but also has highest defaulted loans as well.

DTI ratio & Bankruptcy

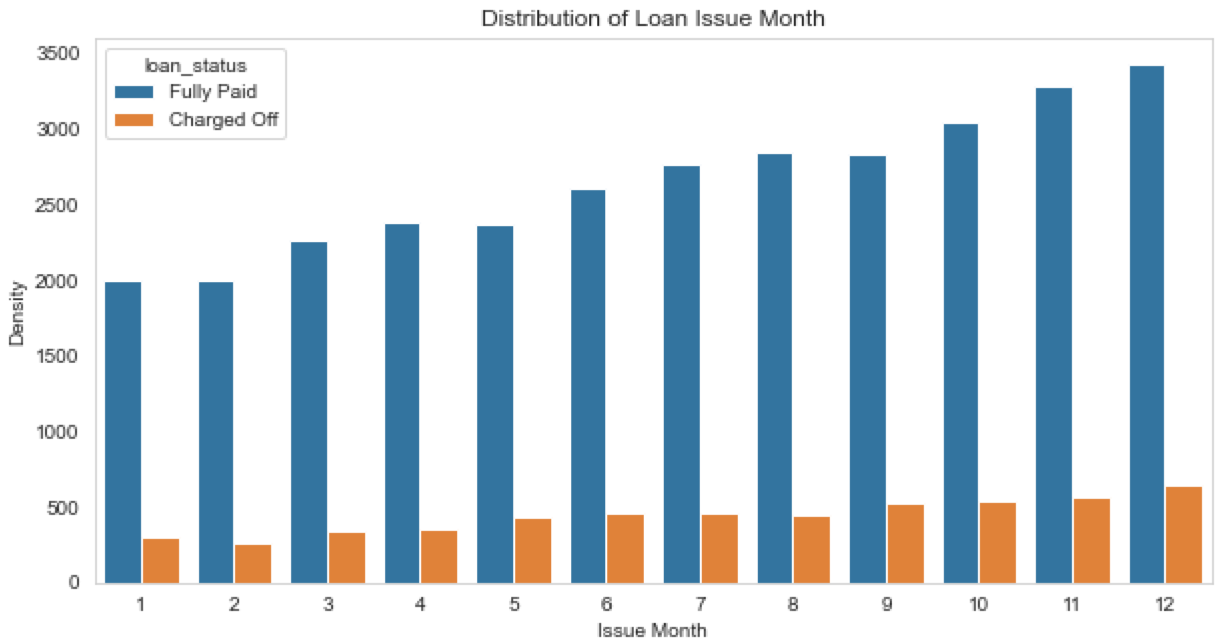




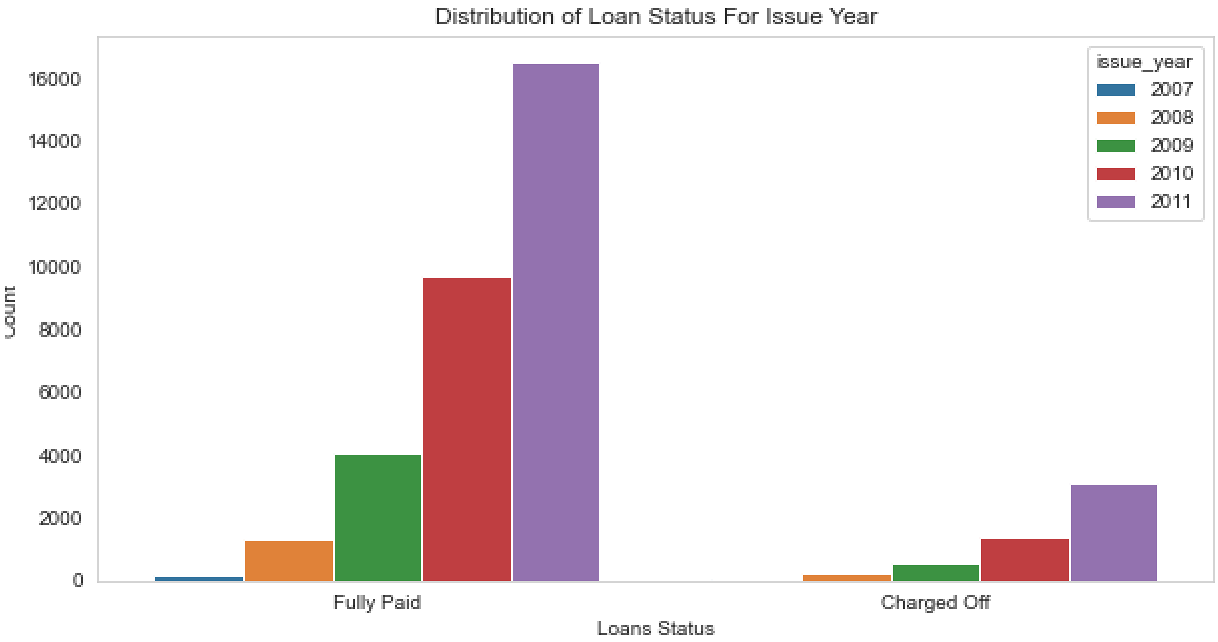
- **DTI:** The large percentage of Clients have a large Debt to Income ratio which shows that lending to such clients can be very risky.

- **Public Recorded Bankruptcy:** Majority of clients have no record of declaring bankruptcy.

Loan Trend over years

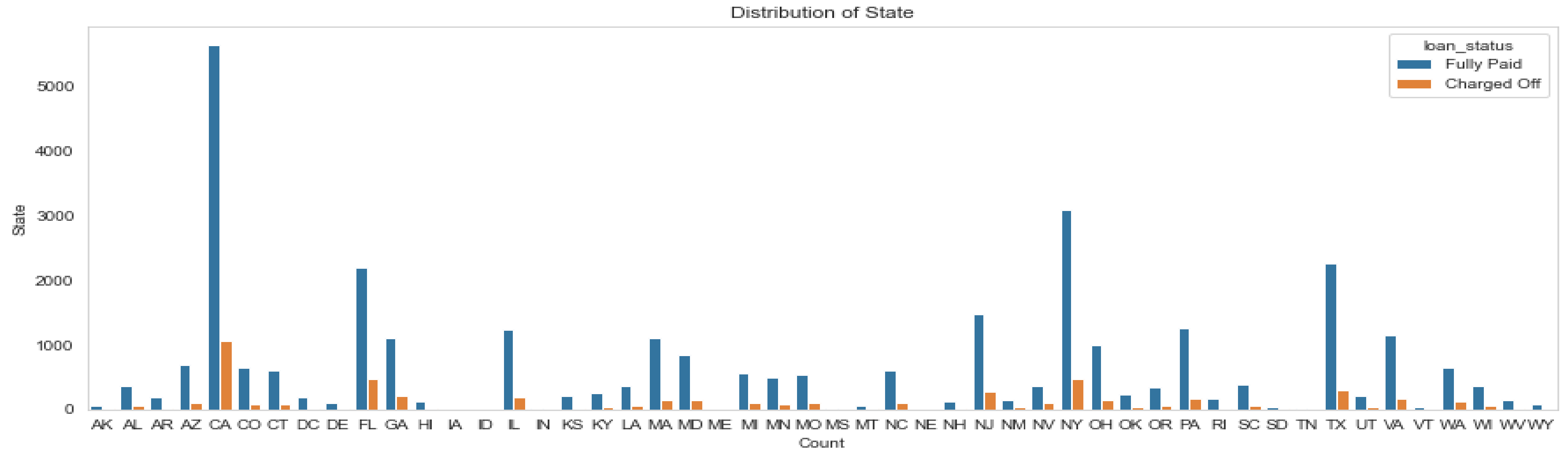


We see a gradual increase in loan taken through the year, with lesser defaulting rate in April ,August, December quarter wise and better more late in year.



With each passing year loan taken are increasing exponentially which indicate we are seeing large increase in DTI ratio and decrease in defaulting rate.

Location Based



For large metropolitan cities we see large number of loans, with higher number of defaulted loans like California, New York, Texas, Florida but have a lower chance of defaulting.

Summary of the findings

Findings:
Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

1. DTI
2. Grades
3. Verification Status
4. Annual income
5. Pub_rec_bankruptcies

Other considerations for 'defaults' :

1. Borrowers not from large urban cities like California, new york, texas, florida etc.
2. Borrowers having annual income in the range 50000-100000.
3. Borrowers having Public Recorded Bankruptcy.
4. Borrowers with least grades like E,F,G which indicates high risk.
5. Borrowers with very high Debt to Income value.
6. Borrowers with working experience 10+ years.