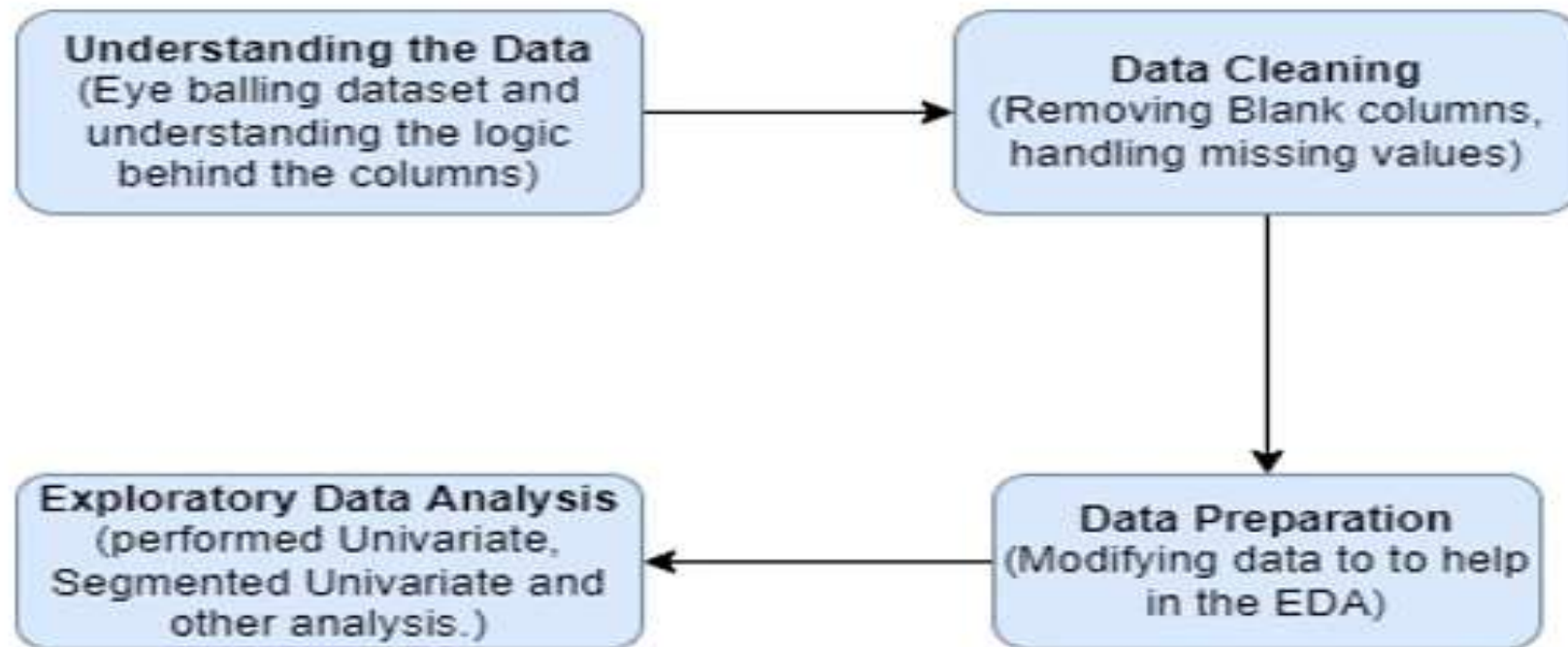# Case Study:

LendingClub

Group Members:

1.Rajat Raina

Objectives

- To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

- Organization can utilise this knowledge for its portfolio and risk assessment.

EDA Process

Data Cleaning

Below steps were performed to clean the data

- Removing columns with all Null Values.
- Finding Missing values and handling them depending on the type of the columns if it is Categorical or Continuous Numerical
- Removing columns that doesn't carry out enough information for the analysis.
- Removing duplicate rows if any.

Analysis

To perform EDA below analysis was done on the dataset after cleaning and processing it.

- **Univariate Analysis:** Individual columns were targeted to get some insights using histogram, countplots.
- **Segmented Univariate:** Single column was segmented into buckets then using second column insights were derived using barplots.
- **Bivariate Analysis:** Two columns were considered to see the impact of one column over the other using boxplots, line graphs.
- **Multivariate Analysis:** Multiple columns were used to derive insights likw which columns are highly correlated using Heatmaps.
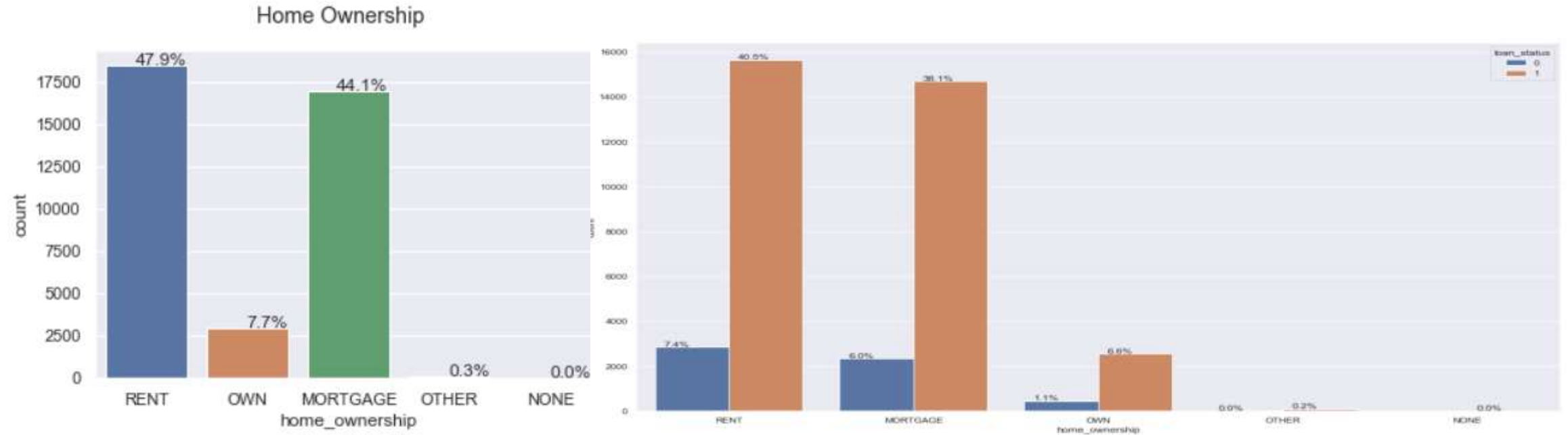
Employee Service Length Analysis

Employment Length
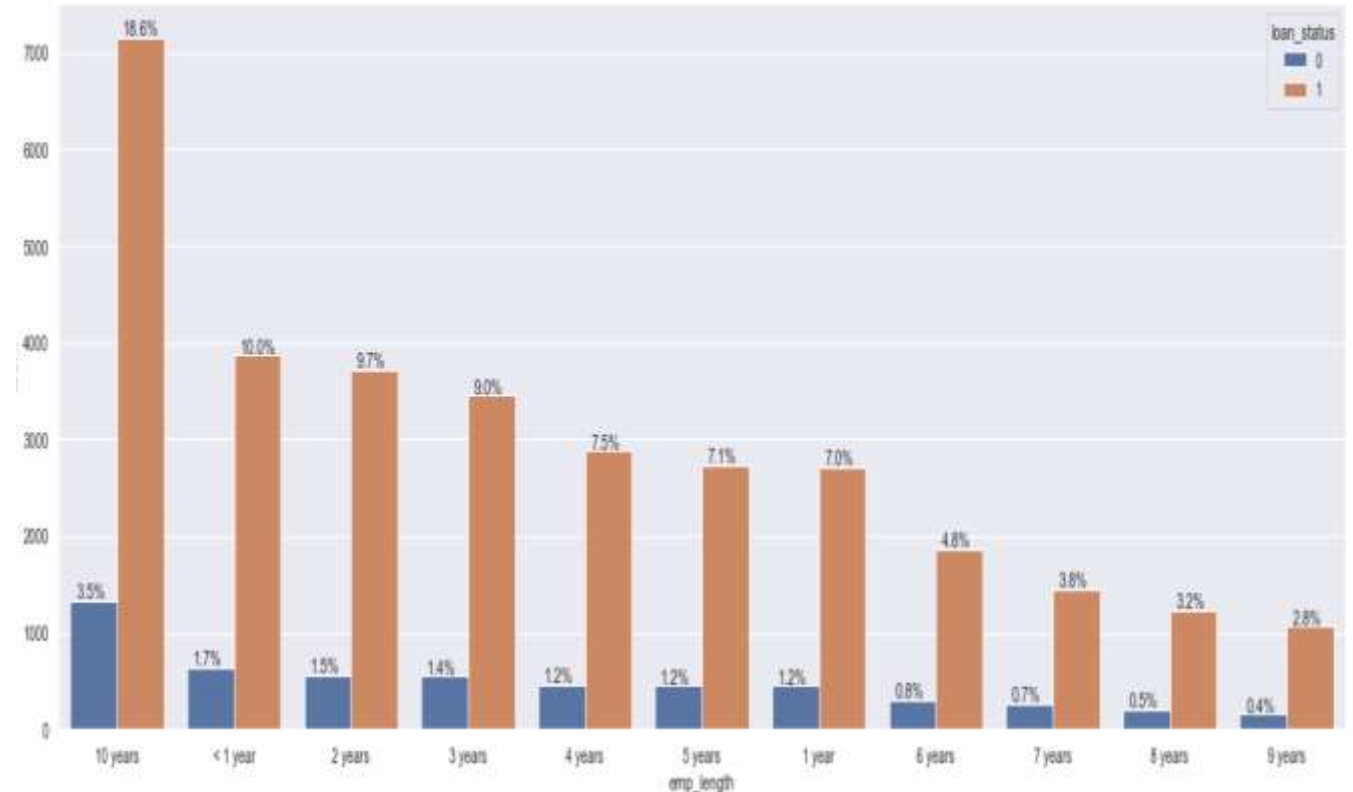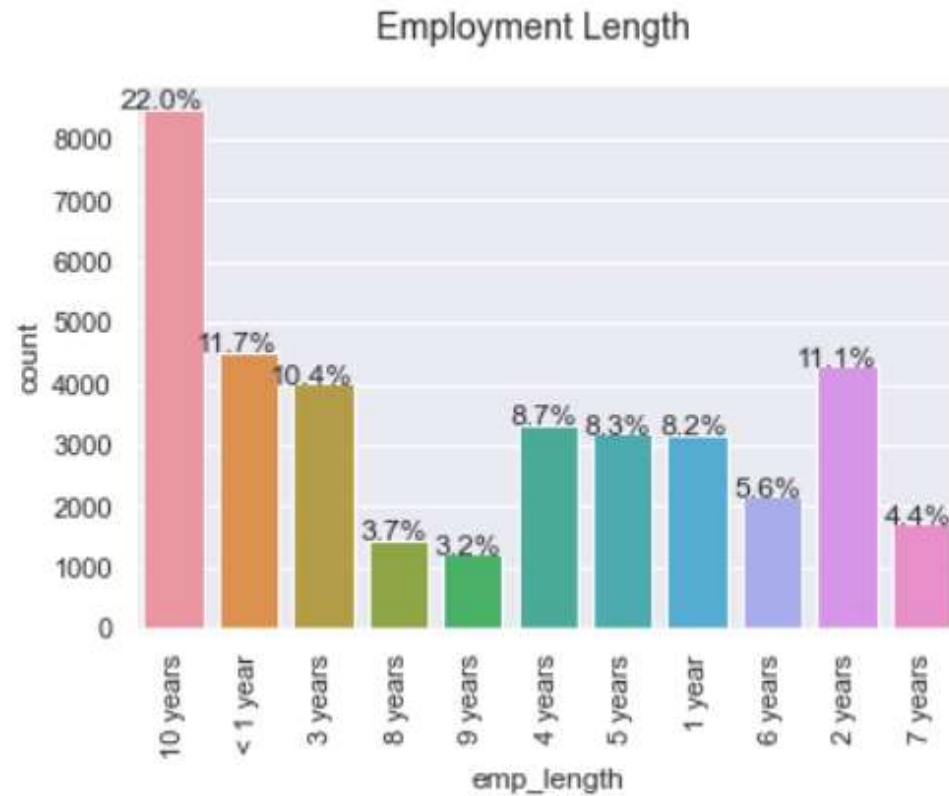
Insights:

1. Most of the loans have been taken by people having experience of 10 years.
2. Also , it's worth noting that most of the loans which are Fully paid belong to people having 10 years of experience.

Home Ownership Analysis

Insights:

1. People who are staying on Rent and having any Mortgage have taken majority of the loans.
2. Also, most of the Fully Paid loan has been paid by people living on Rent and having any Mortgage, The reason we can think of this behaviour is that people who own a house usually have better financial condition than others.
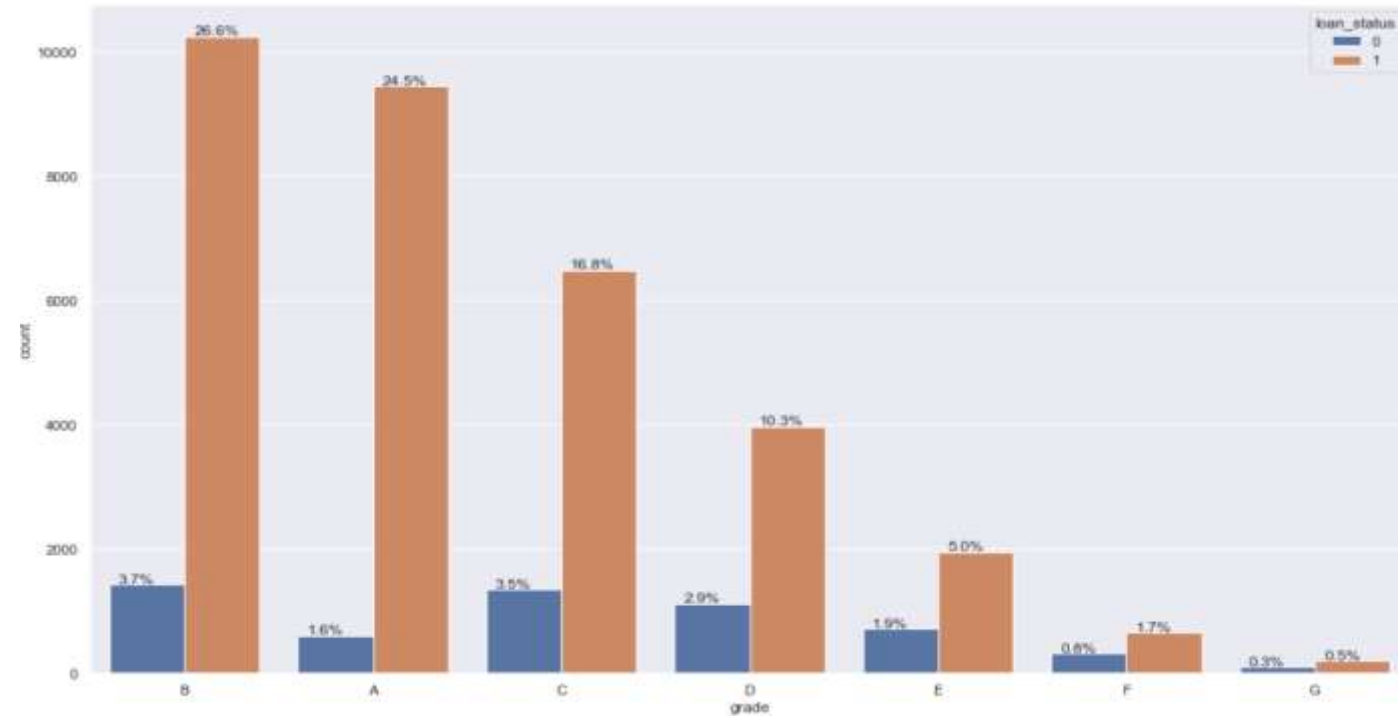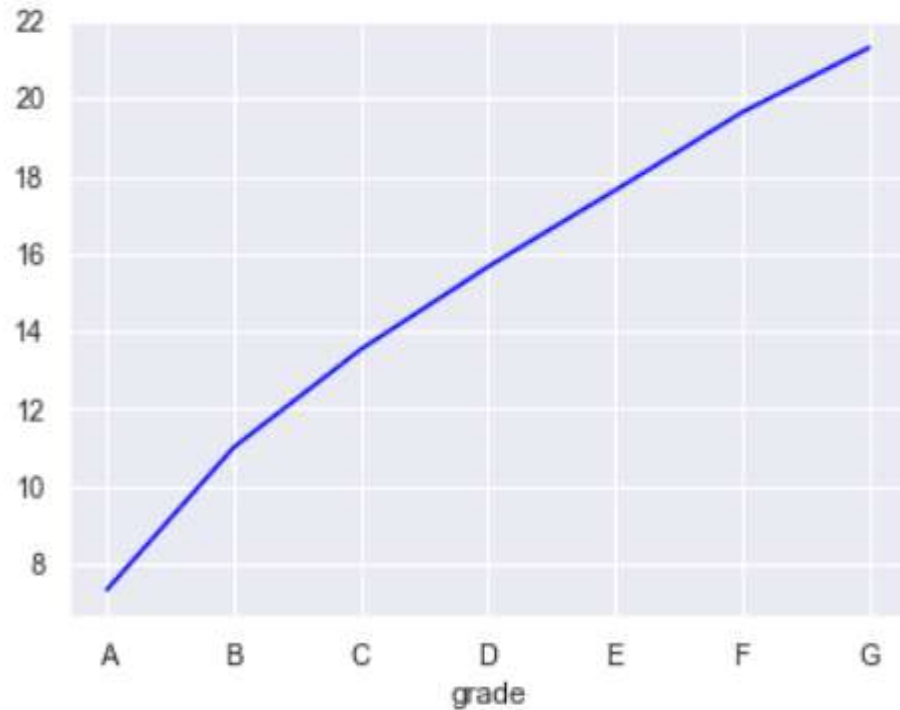
Verification Status Analysis
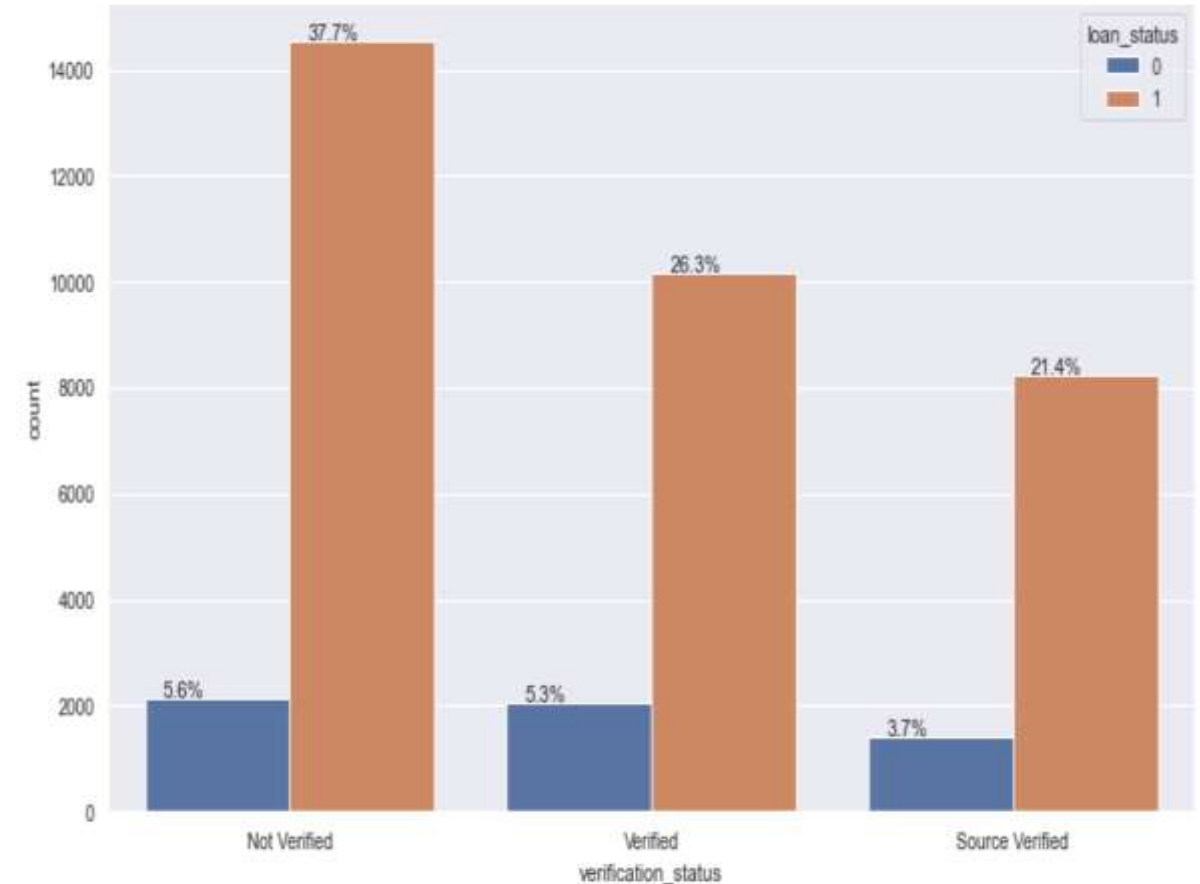
Insights:

1. Customers who are not Verified have taken most of the Loans.
2. Its slightly difficult to digest but the fact is that majority of the Loans which are Fully paid belong to people who are not verified, so company need to check this behaviour.
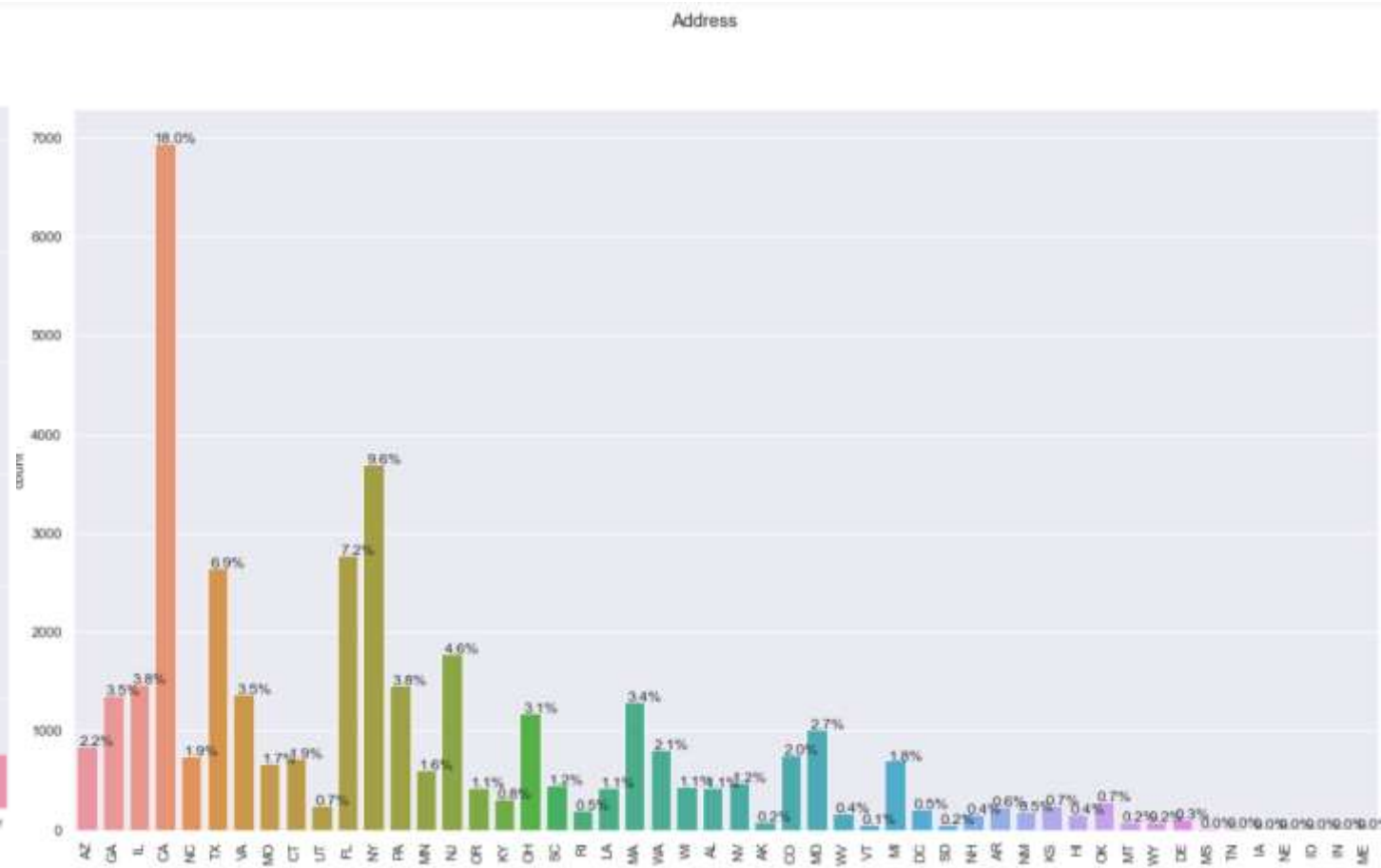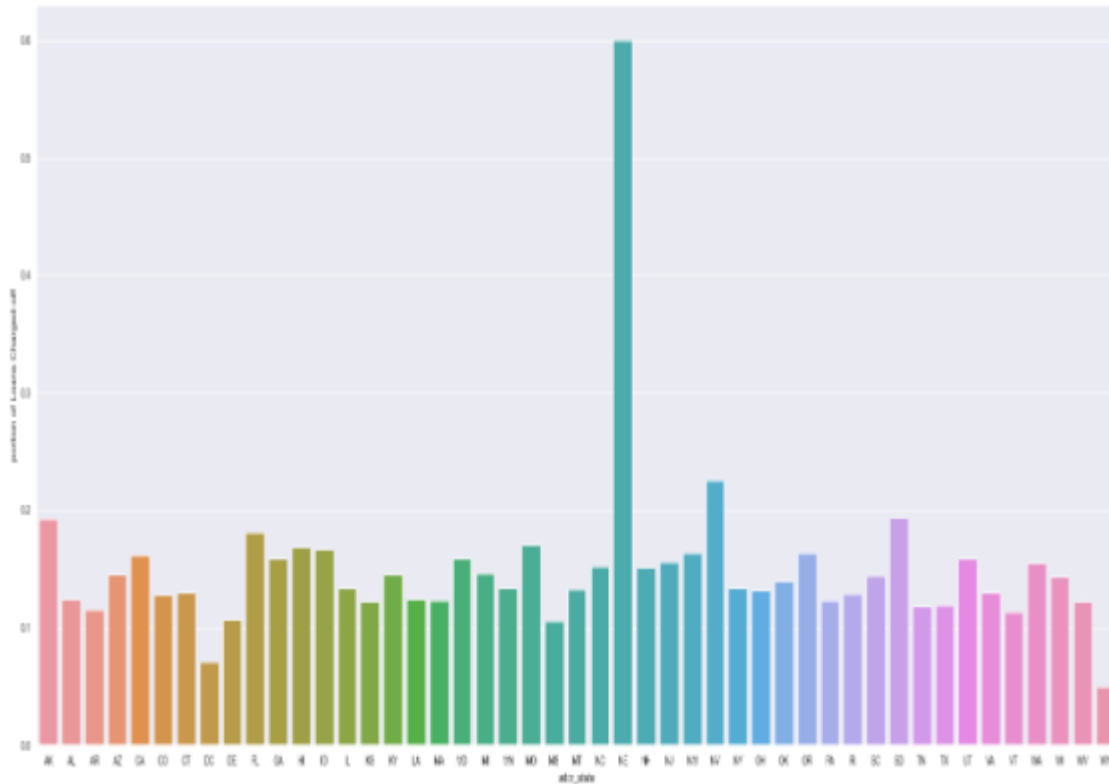
Grade-Interest Rate Analysis

Insights:

1. As the Grade increases from A-G the interest rate also increases.
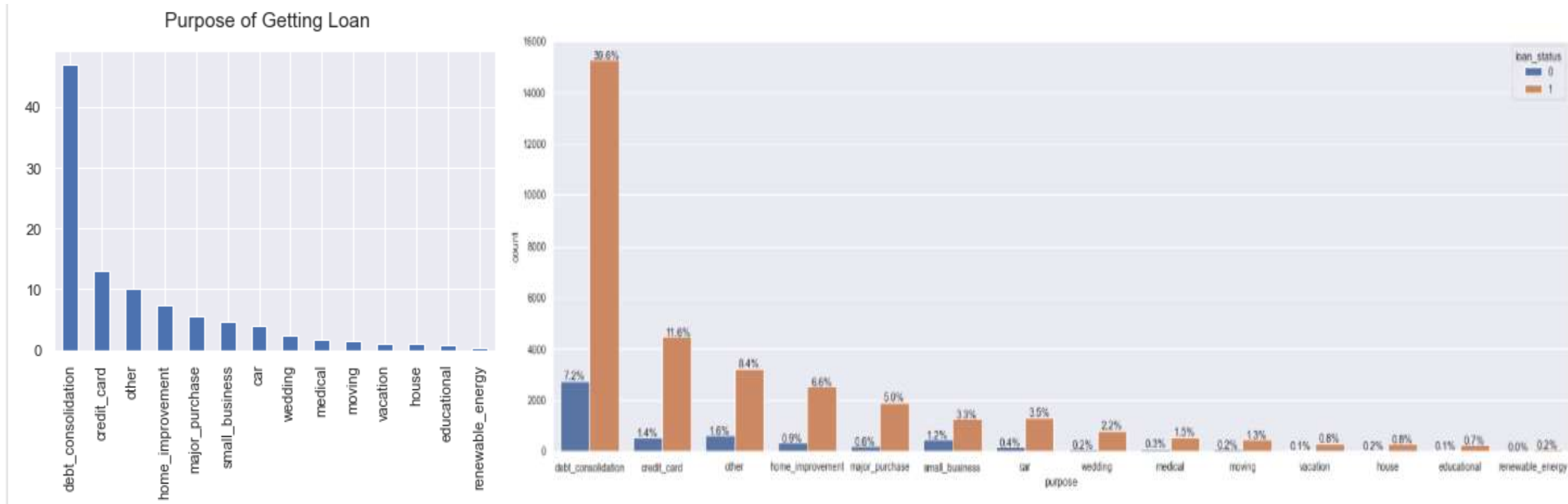2. Also, as the grade increases the default rate also increases.

Customer Address Analysis

Insights:

1. Most of the loan has been taken by people from CA.
2. Also interesting fact is that most of the loans which are Charged Off belong to people from NE.
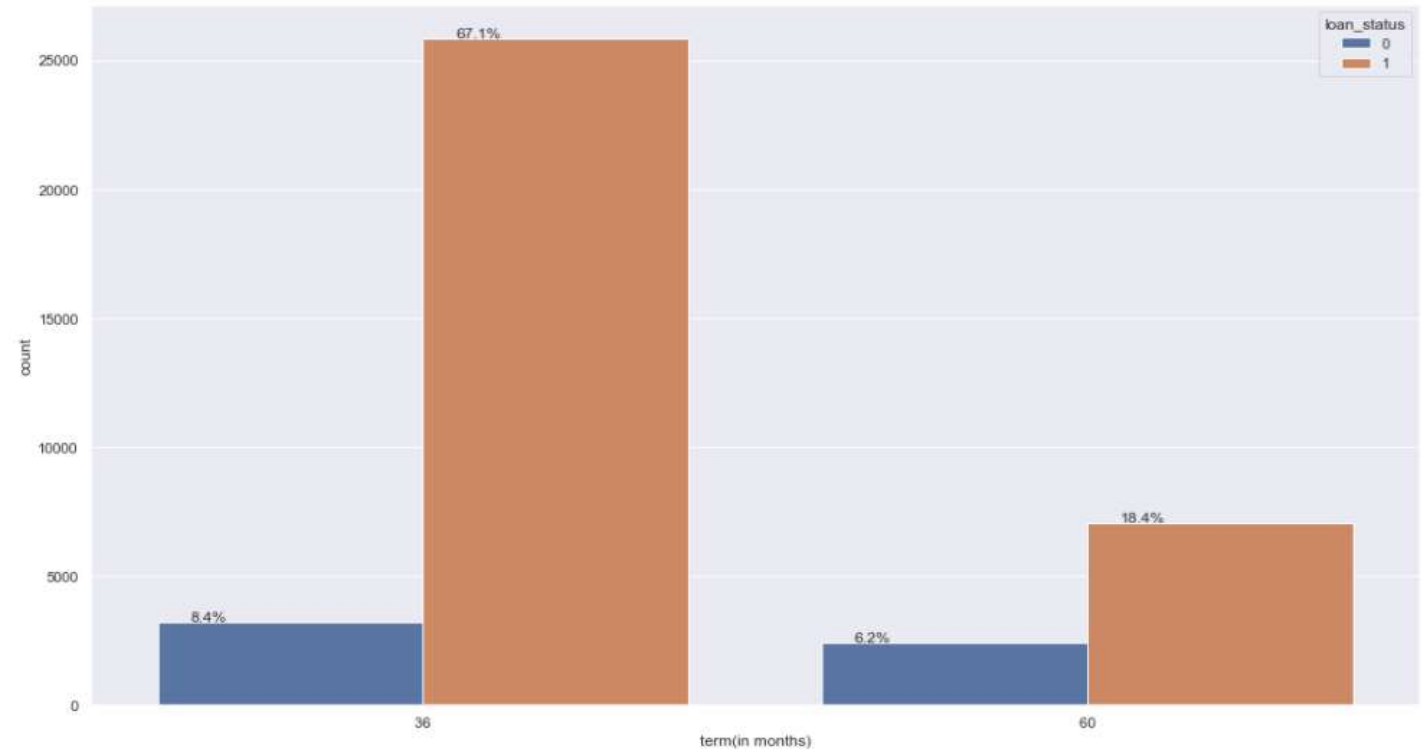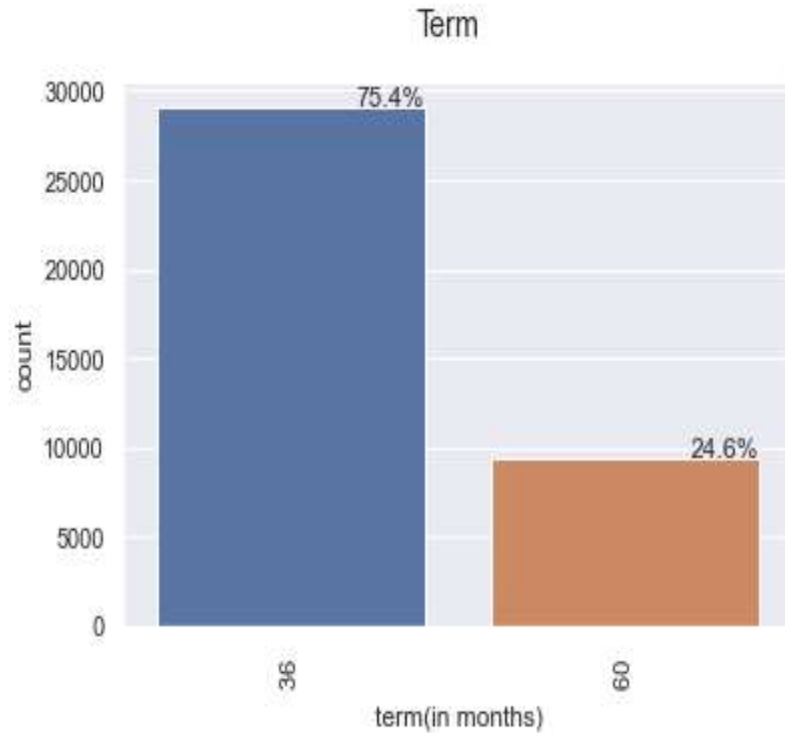
Loan Purpose Analysis

Purpose of Getting Loan

Insights:

1. People prefer to take loan for the purpose of debt_consolidation.
2. Majority of the loan that was fully paid is coming from people who took loan for debt_consolidation but the dafualts are also majorly from debt_consolidation.
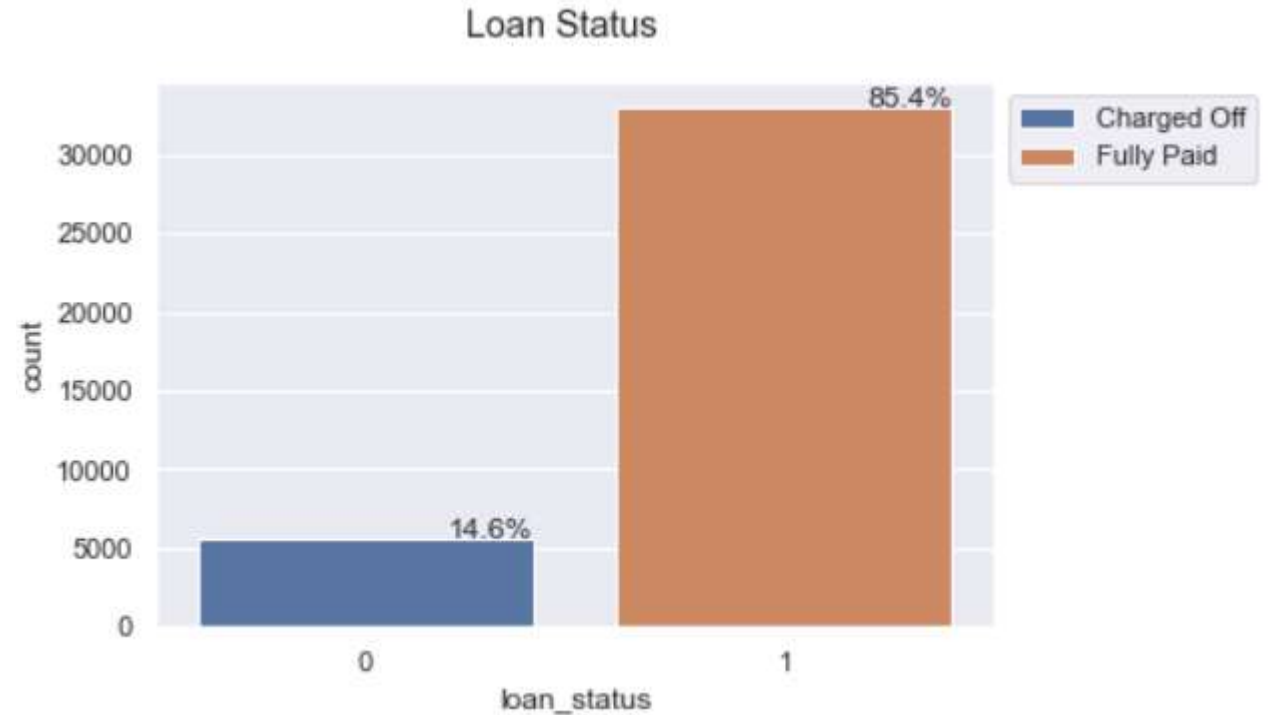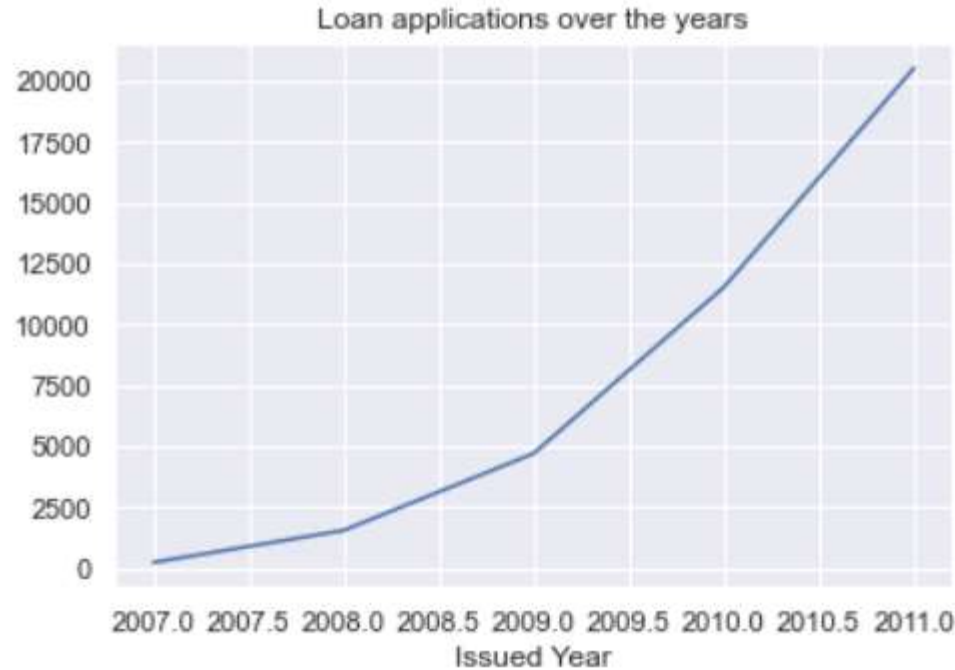
Loan Tenure Analysis

Insights:

1. Customers are taking most of the loans for a tenure of 36 months.
2. While most of the loans taken for a tenure of 36 months are fully paid there is also a slightly high tendancy to default.

Loan Applications Analysis

Insights:

1. Year by year the number of loan applications where loan was granted also increased.
2. While the number of the loans increased , majority of the them are fully paid while 14% of the loans are charged off.

Conclusion

Even though we did fair amount of analysis on the given dataset but we conclude our analysis with below pointers.

1.While giving loans to people at high interest rate the organization should do a thorough check on the background of the customer.

2. If organizations are giving loan to people from different states then customers annual salary and the type of city should be considered(rural or urban).

3. As the number of loan applications are increasing year by year, if the loan purpose is debt consolidation, applicant should be checked thoroughly as it's has the highest default.