Lending Club Case Study Submission
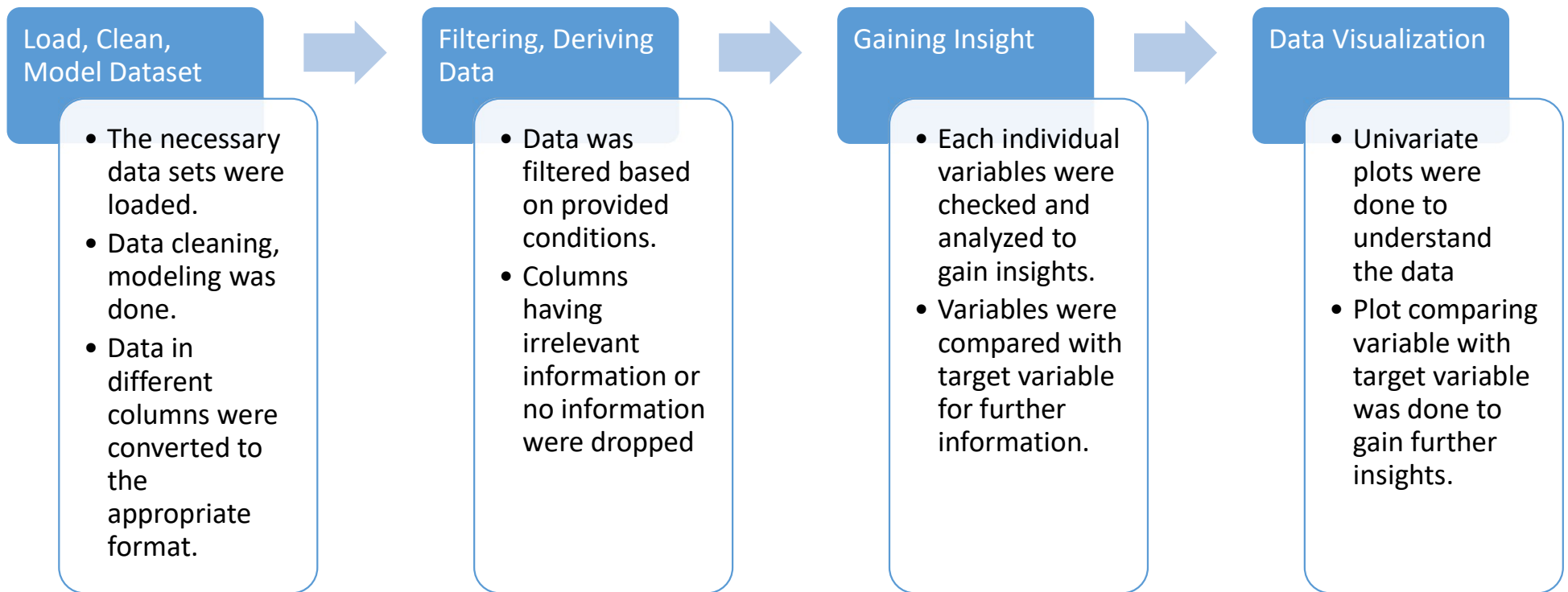
Name: DIPAYAN PRAMANIK AND RIDDHI TOKDAR
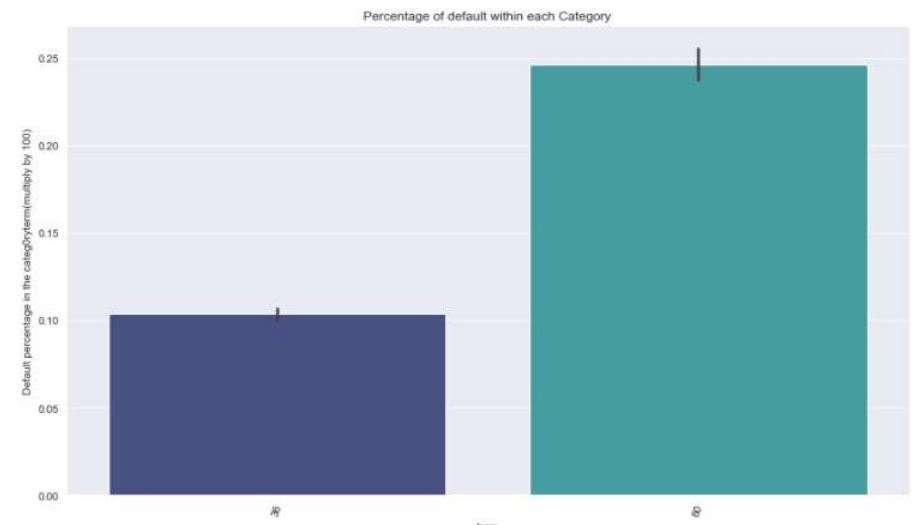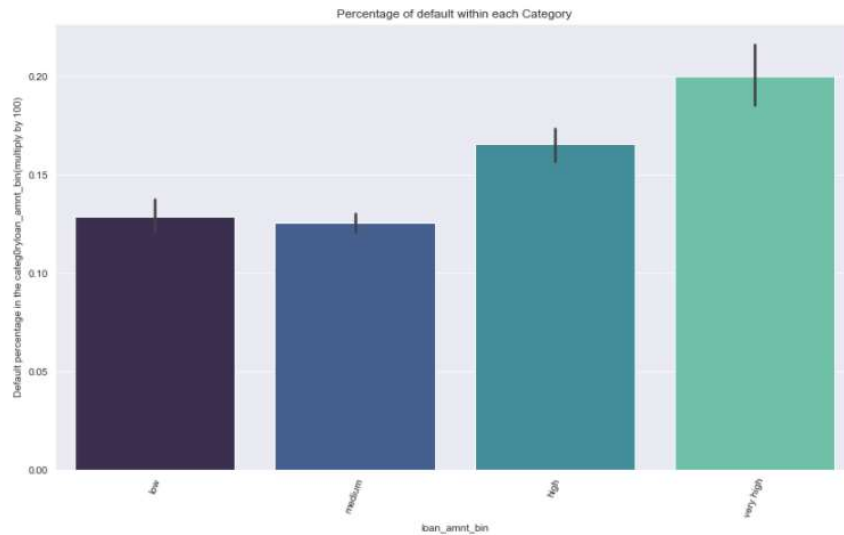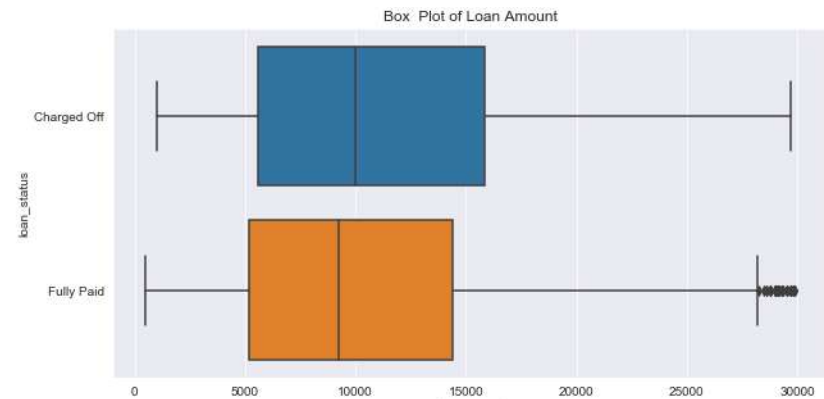
Abstract

The data contains the complete loan data for all loans issued through the time period 2007 to 2011. The goal is to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Problem solving methodology

Load, Clean, Model Dataset

- The necessary data sets were loaded.
- Data cleaning, modeling was done.
- Data in different columns were converted to the appropriate format.

Filtering, Deriving Data

- Data was filtered based on provided conditions.
- Columns having irrelevant information or no information were dropped

Gaining Insight

- Each individual variables were checked and analyzed to gain insights.
- Variables were compared with target variable for further information.

Data Visualization

- Univariate plots were done to understand the data
- Plot comparing variable with target variable was done to gain further insights.

Analysis

- Most of the loans are between 5000 to 15000. Higher the loan amount, higher the default rate.

- Loans having term of 60 months has higher default rate (24.6%%) compared to loans with 36 months tenure(10.3%)

Box Plot of Loan Amount

Percentage of default within each Category


Percentage of default within each Category

Analysis

- The loans that have defaulted has higher interest rate compared to those paid.
- Percentage of defaults increases with the increase of interest rates in loan.
- The default rate of loans increase with the increase of installment amount.
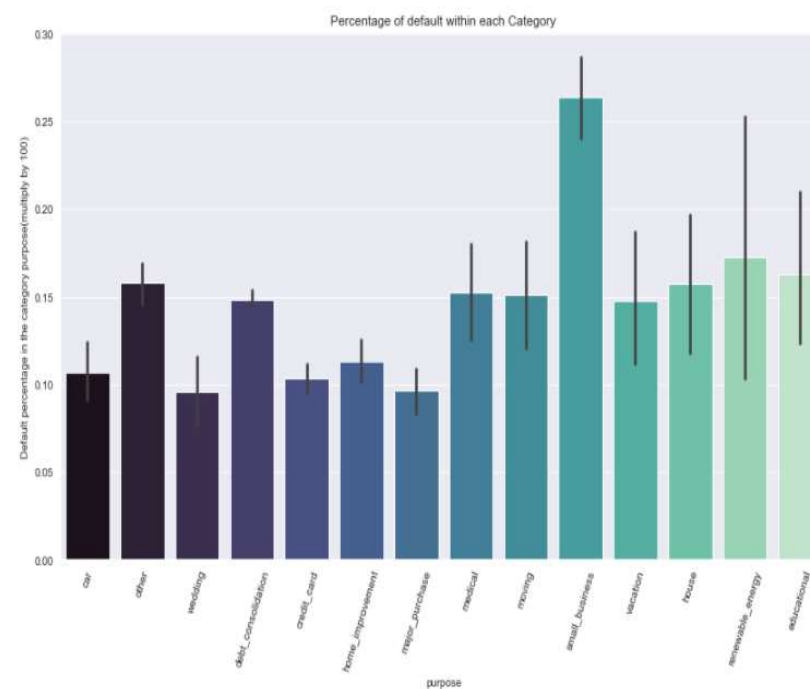
Analysis
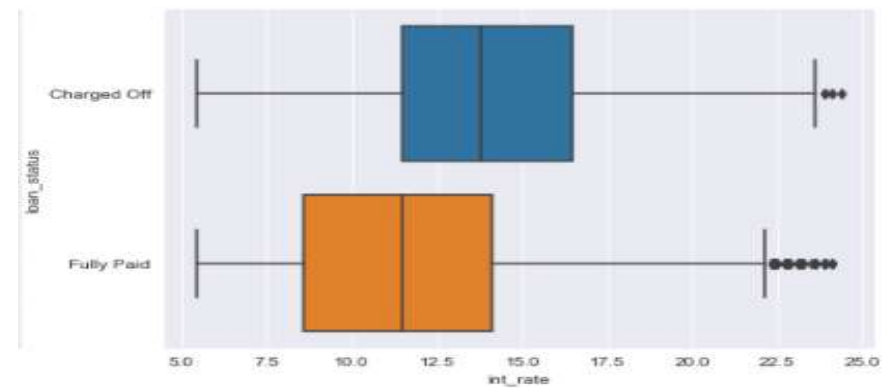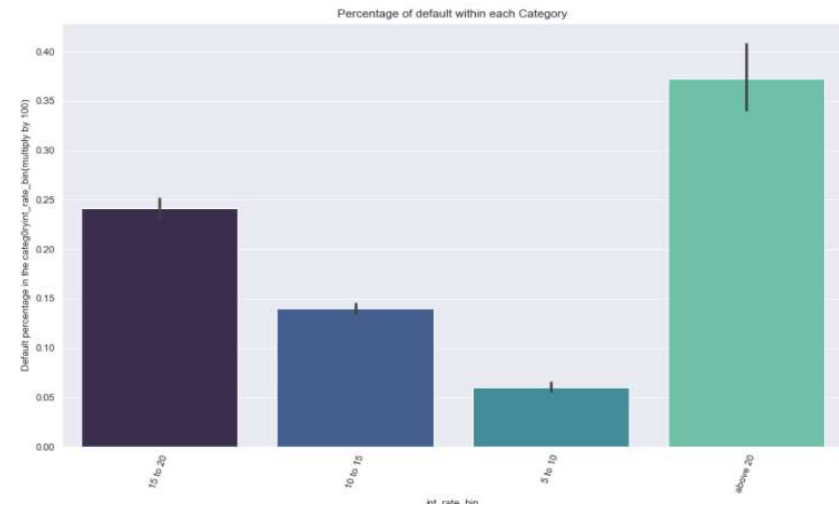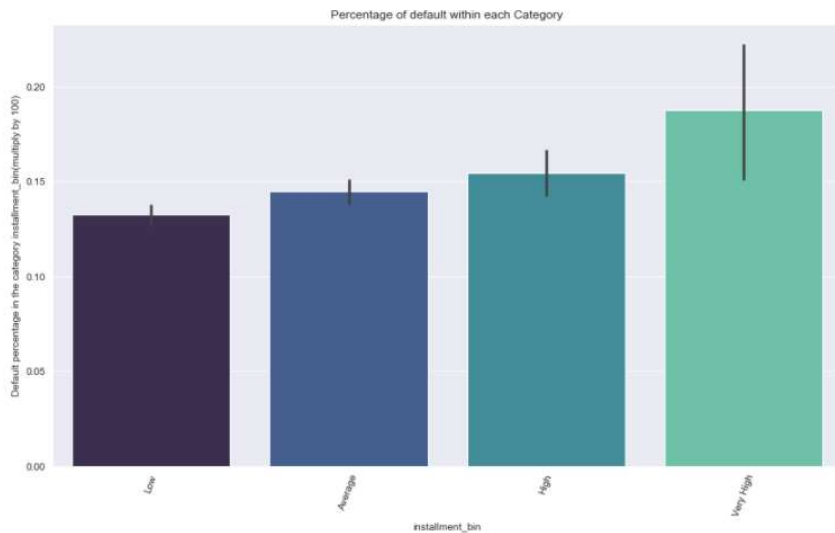
- Default rate decreases as annual income of the borrower increases.
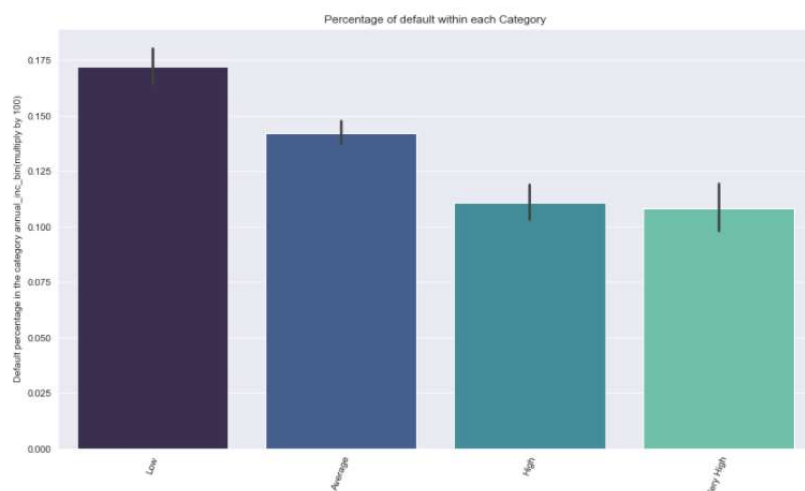
- Default tends to increase with decrease in grade and subgrade

grade	A	B	C	D	E	F	G
is_default	0.056127	0.116906	0.16293	0.210165	0.261185	0.317073	0.329787

- There is significant higher rate of default if the purpose of loan is small business.

Analysis

- Loans issued in 2007 has significant higher rate of default.

- Higher the debt to income ratio, higher the rate of default

- The default rate increases with increase in delinquency records.
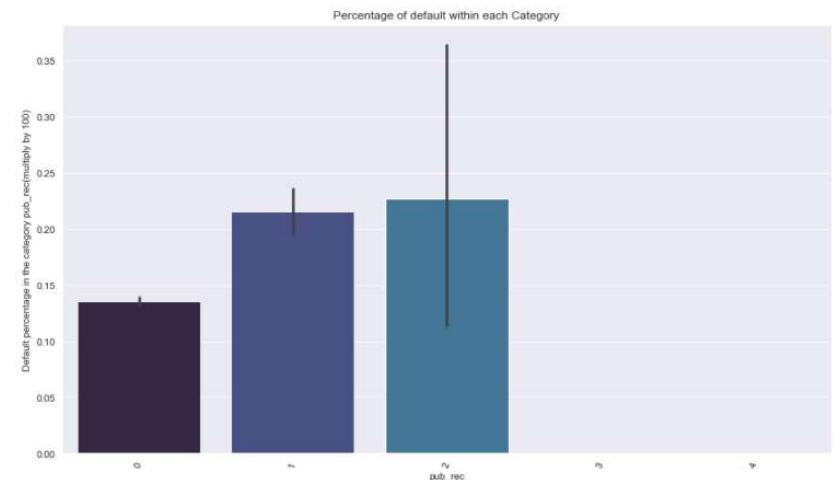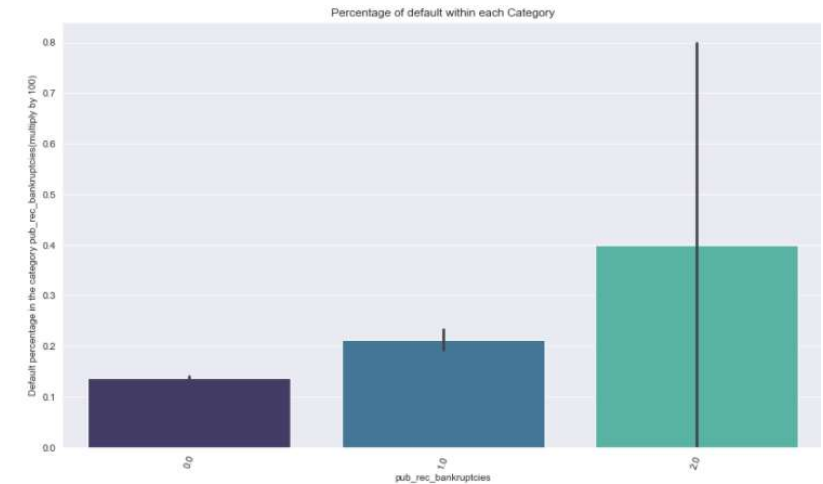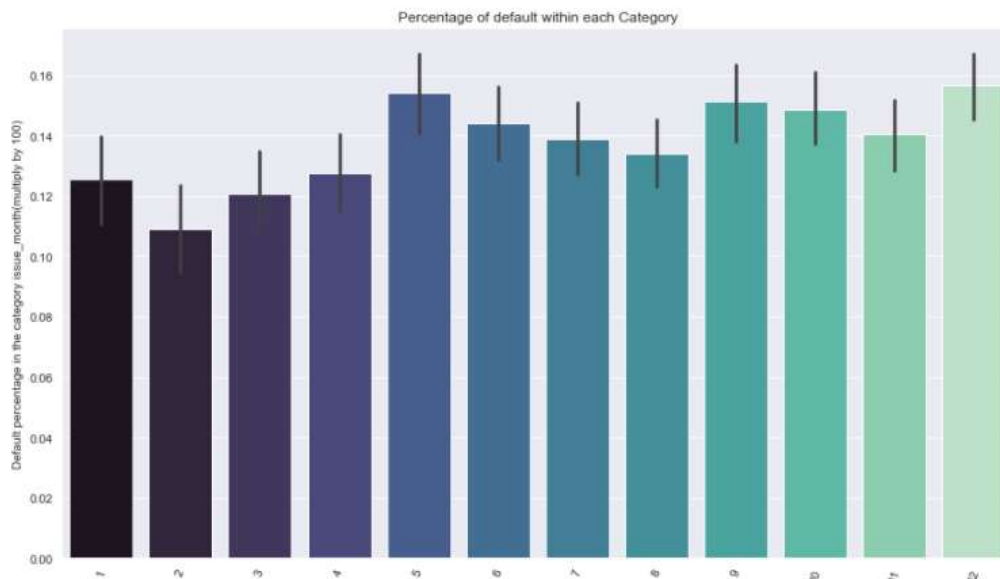
Analysis
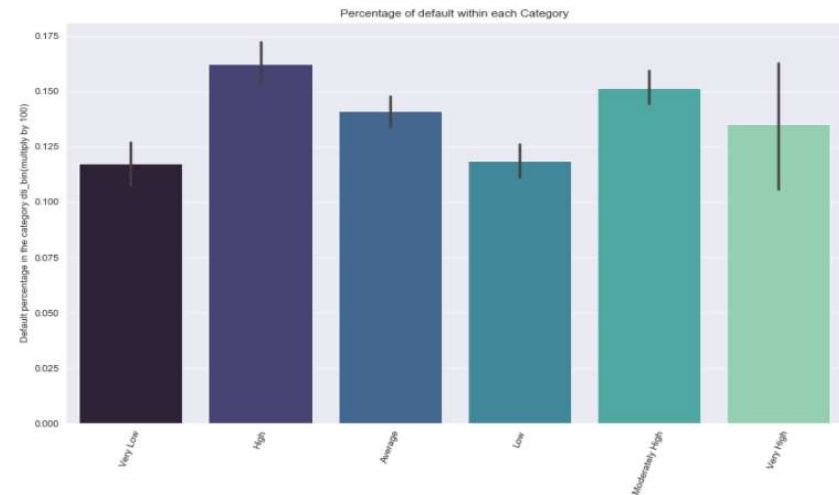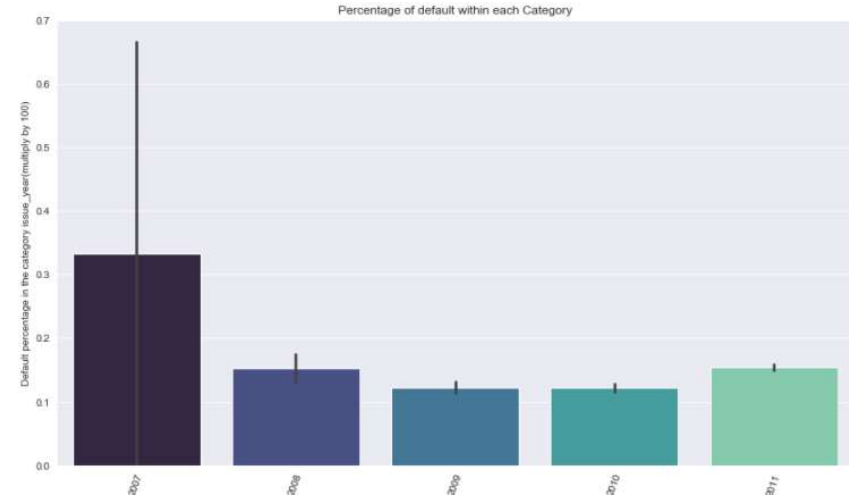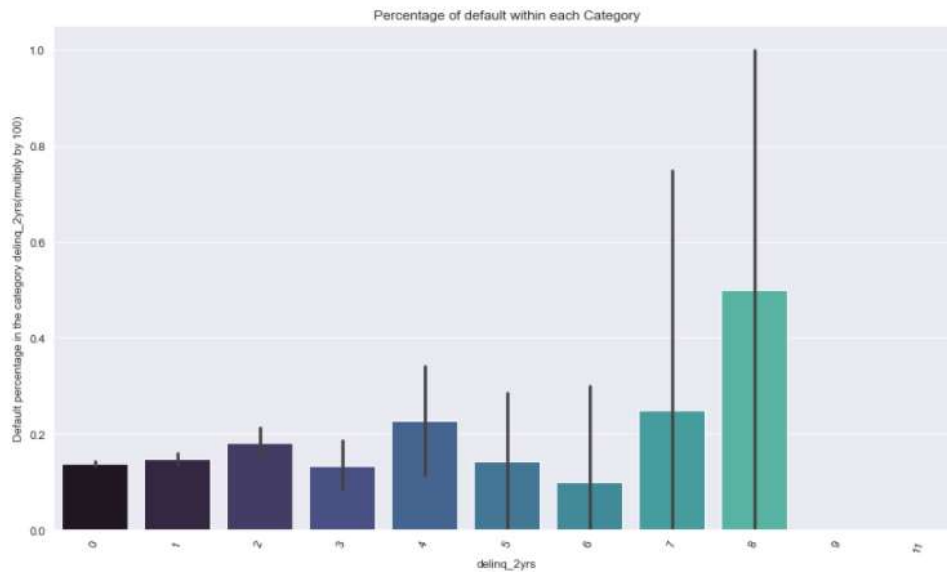
The default rate increases with increasing utilization of revolving line

Most of the borrowers do not make any inquiry. People more than 2 inquiries default at higher rates.

Analysis
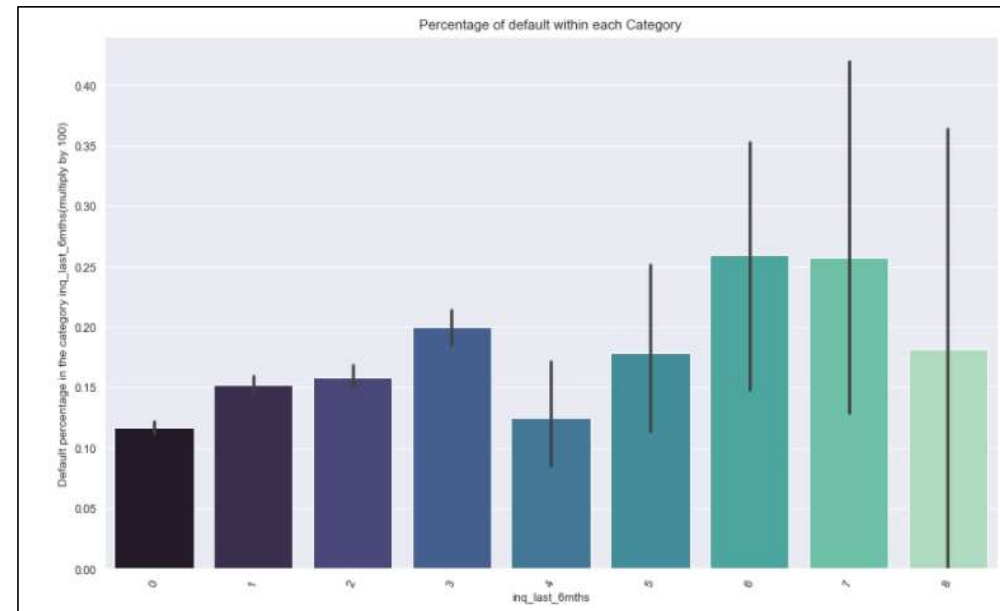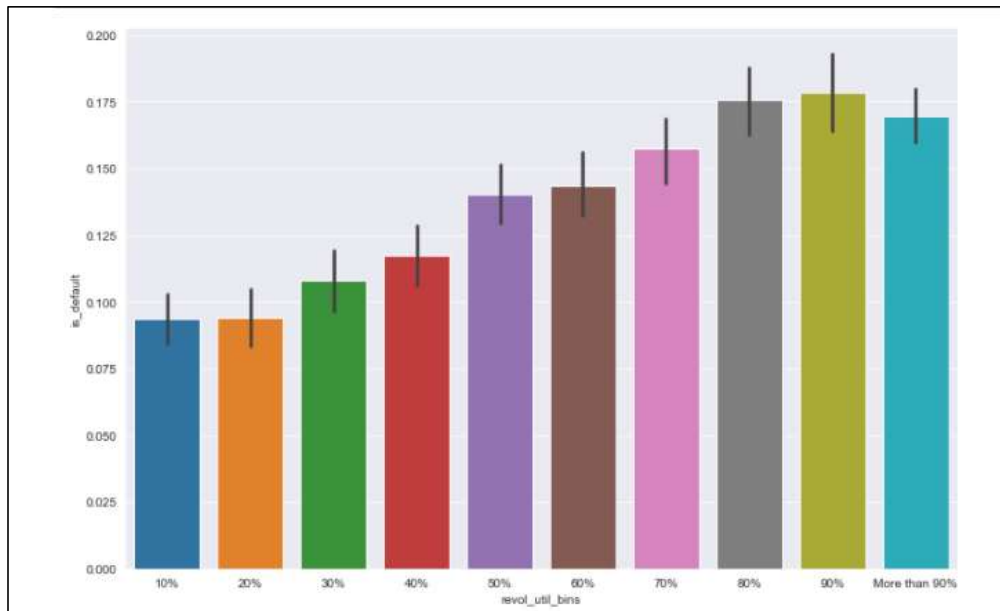
- It can be seen that there is a increase of loans approved as the year end approaches near. December sees most number of loans issued. Also loans issued in December has highest default rates.

- The default rate increases if there are 1 or more records of derogatory public records or records of bankruptcy.

Conclusion

From the analysis, the below can be concluded:

- Higher the loan amount, higher the default rate.

- Percentage of defaults increases with the increase of interest rates in loan

- Default rate decreases as annual income of the borrower increases.

- Loans having term of 60 months has higher default rate compared to loans with 36 months tenure.

- Default tends to increase with decrease in grade and subgrade. There is significant higher rate of default if the purpose of loan is small business.

- The default rate of loans increase with the increase of installment amount.

- The default rate increases with increasing utilization of revolving line and debt to income ratio

- Loans issued in 2007 has higher percentage of default compared to other years.

- Most of the borrowers do not make any inquiry. People more than 2 inquiries default at higher rates.

- The default rate increases if there are 1 or more records of derogatory public records, delinquency records, records of bankruptcy.