



<h1>Lending Club Case study</h1>	
----------------------------------	--

Problem Statement

- Lending company wants to understand the **driving factors (or driver variables)** behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.
- To achieve the goal loan dataset, need to be analysed. Cleanup need to be performed and univariate, bivariate analysis must be done on data to identify the patterns and derive the variables contributing to loan default.

Data Cleanup and Variable segregation

1

Data consists of more than 30% null values in rows and columns are cleaned up

2

New column created to get loan Income ratio (loan amount/annual Income)

3

Created new integer columns for Charged off and Fully paid from Loan status column

4

Converted date columns to date datatype from object datatype

5

Converted interest rate column to float by removing %

6

Segregated all the values into categorical and numerical

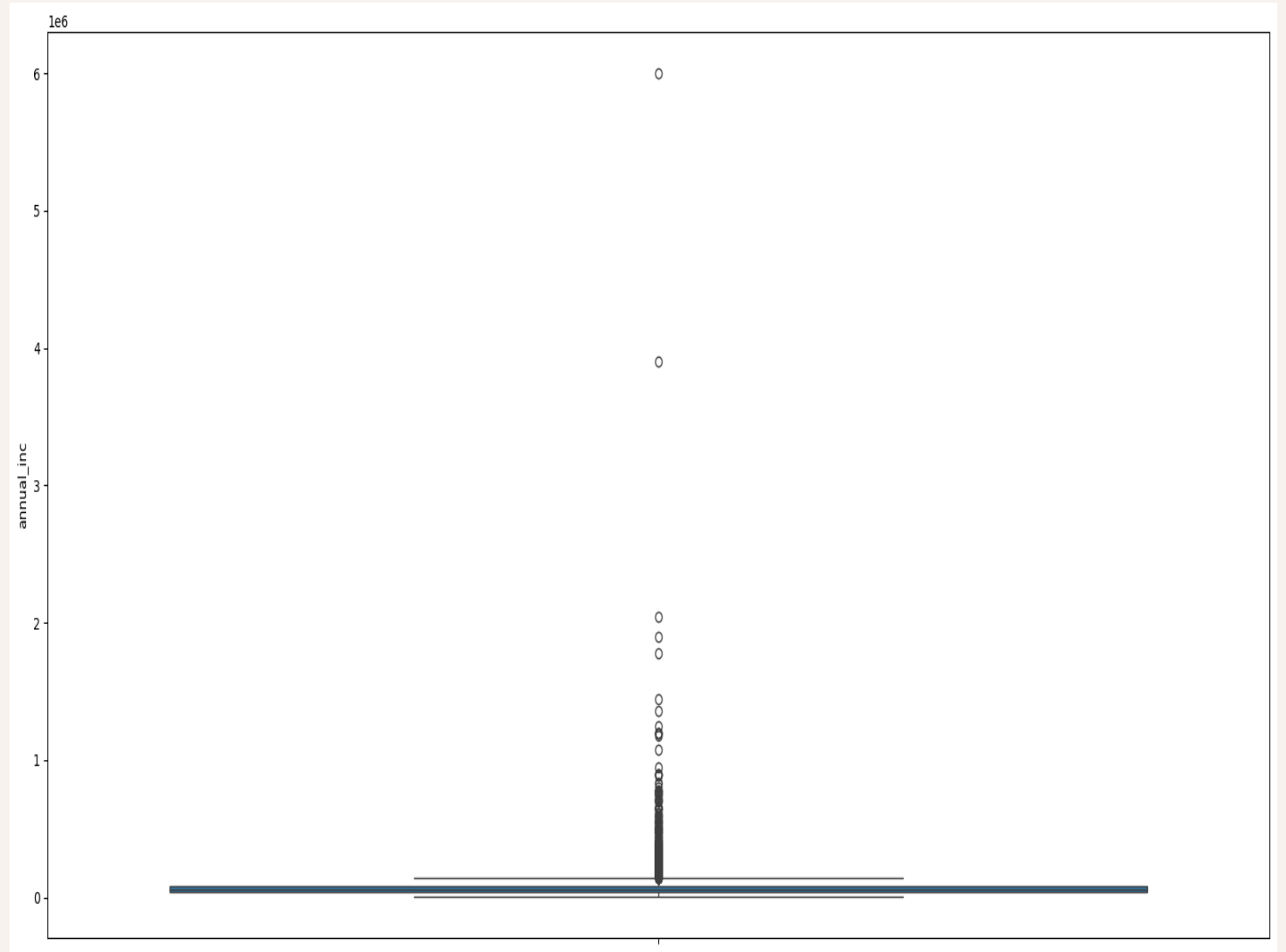
7

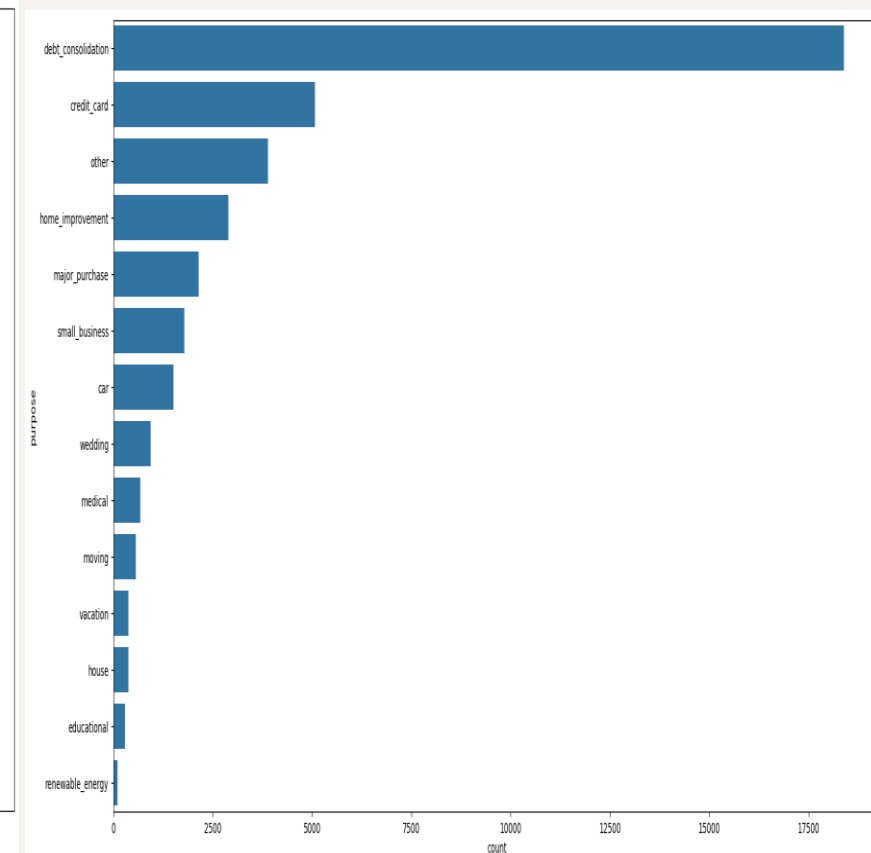
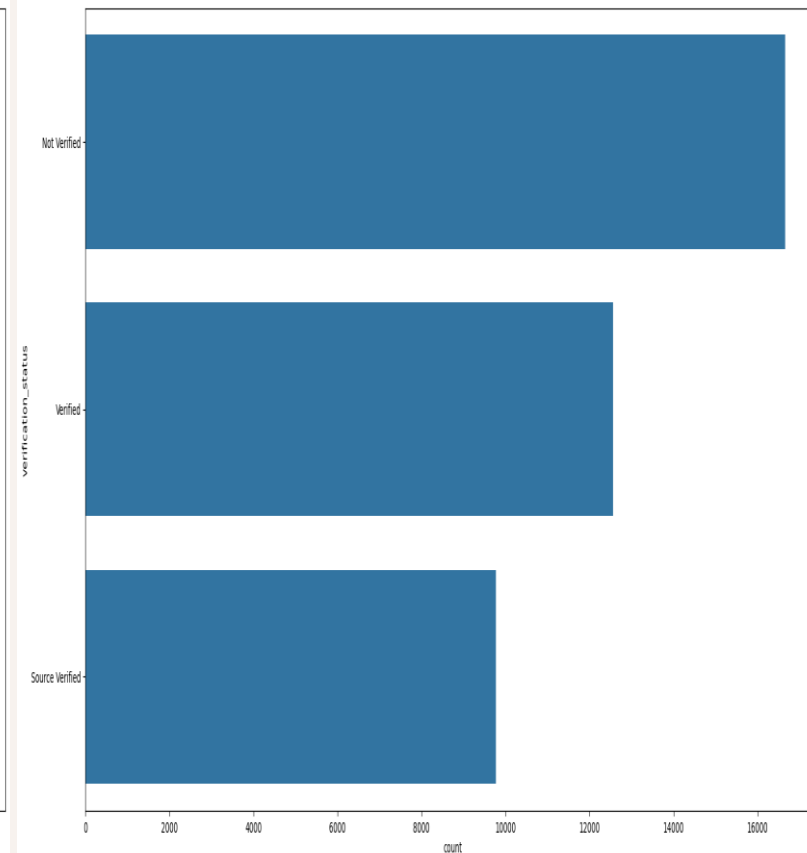
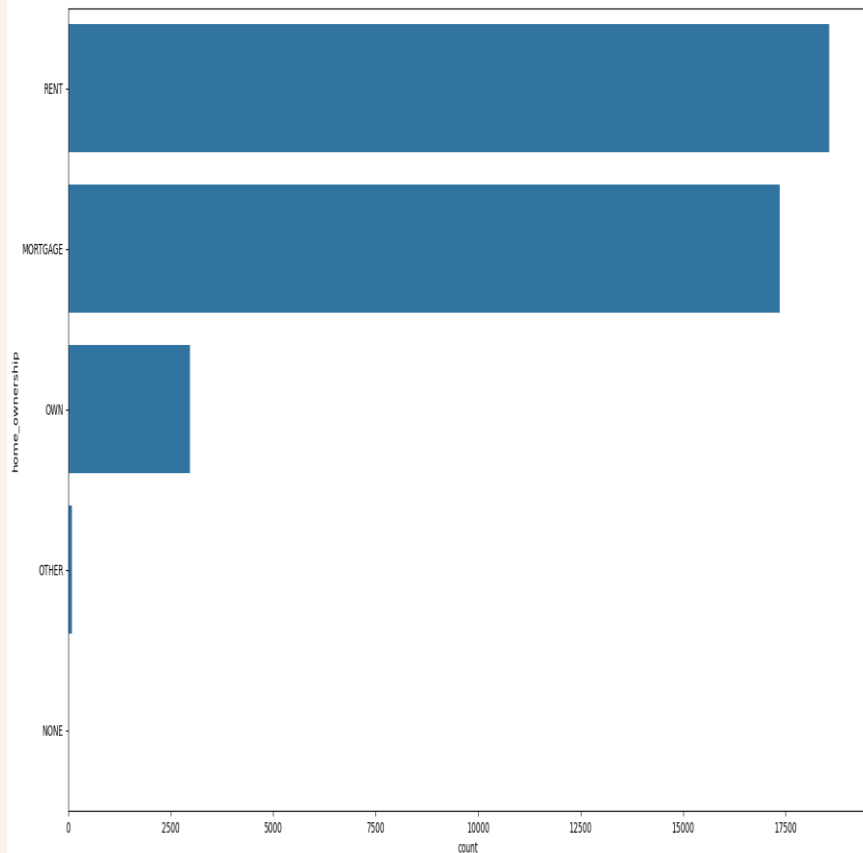
Removed Categorical columns with only one unique value

Univariate Analysis

Box plot was used to see how numerical variables are distributed

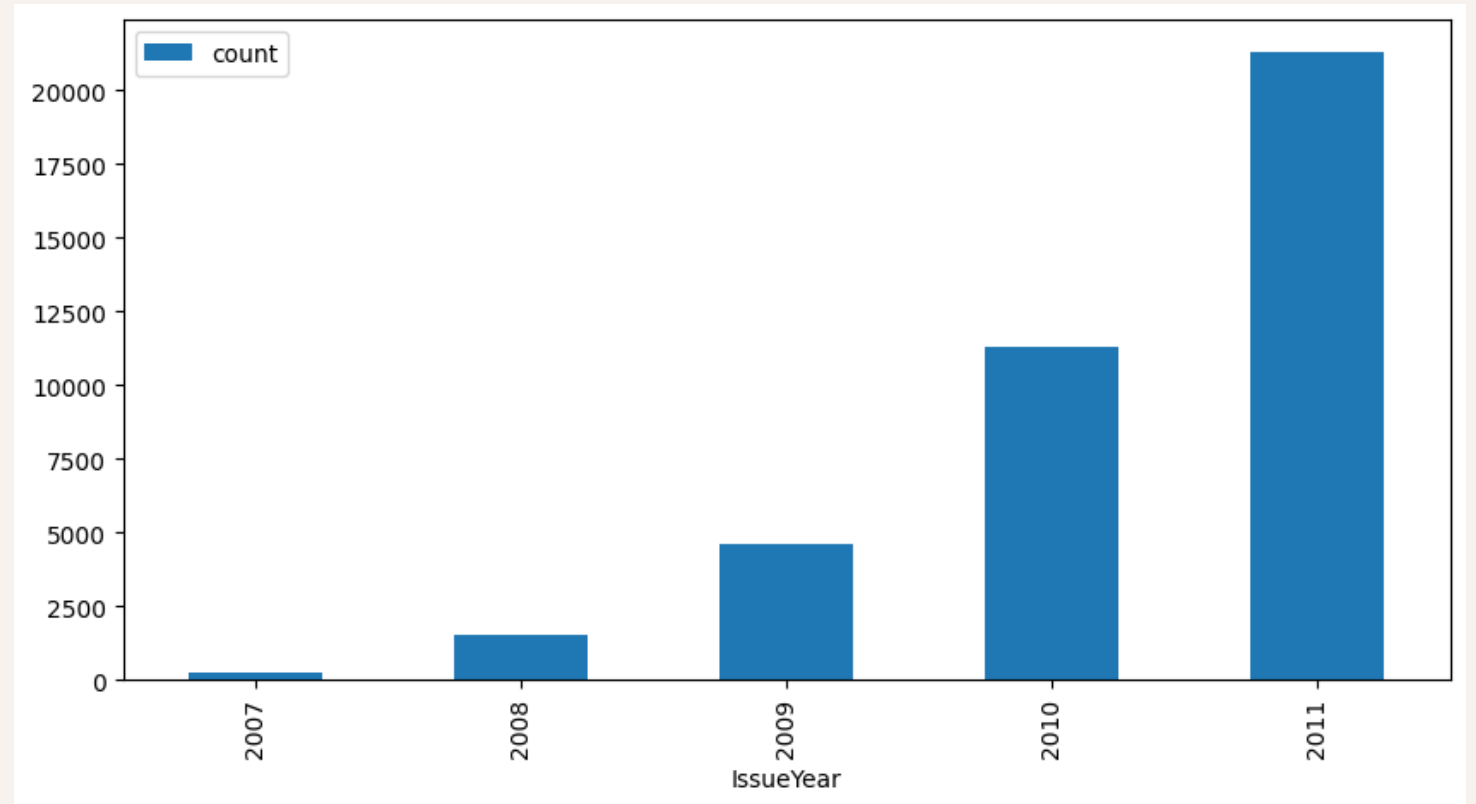
Outliers found on annual income column ; they have been removed by taking values from 1% to 99% of distribution



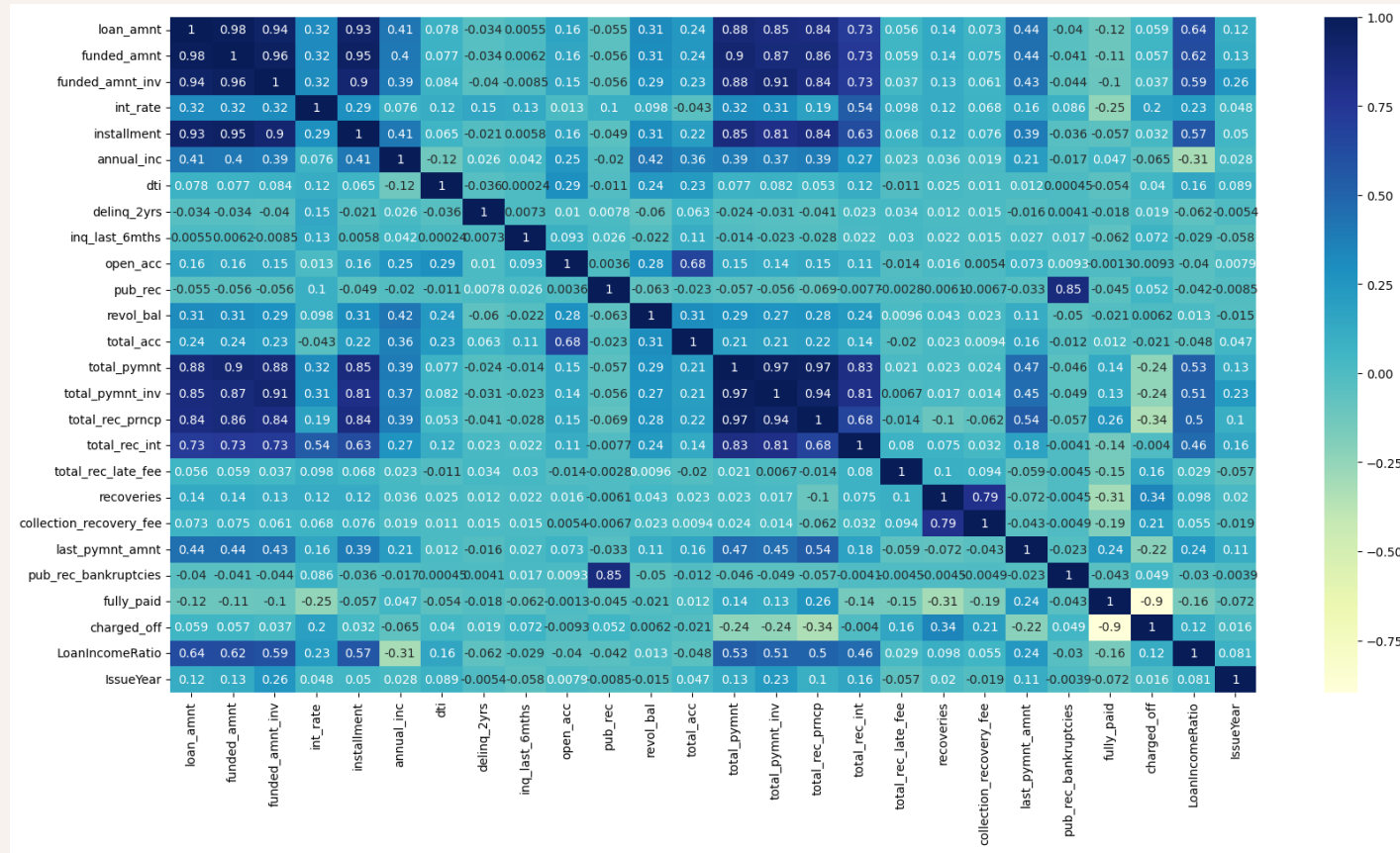


- Most of the loans are taken for debt consolidation to close existing loans
- Combined verified and source verified values to verified for verification status column
- Other, None value rows removed on purpose and home ownership columns

As the number of year increases persons applying for loans also increasing company can anticipate more loans for future years

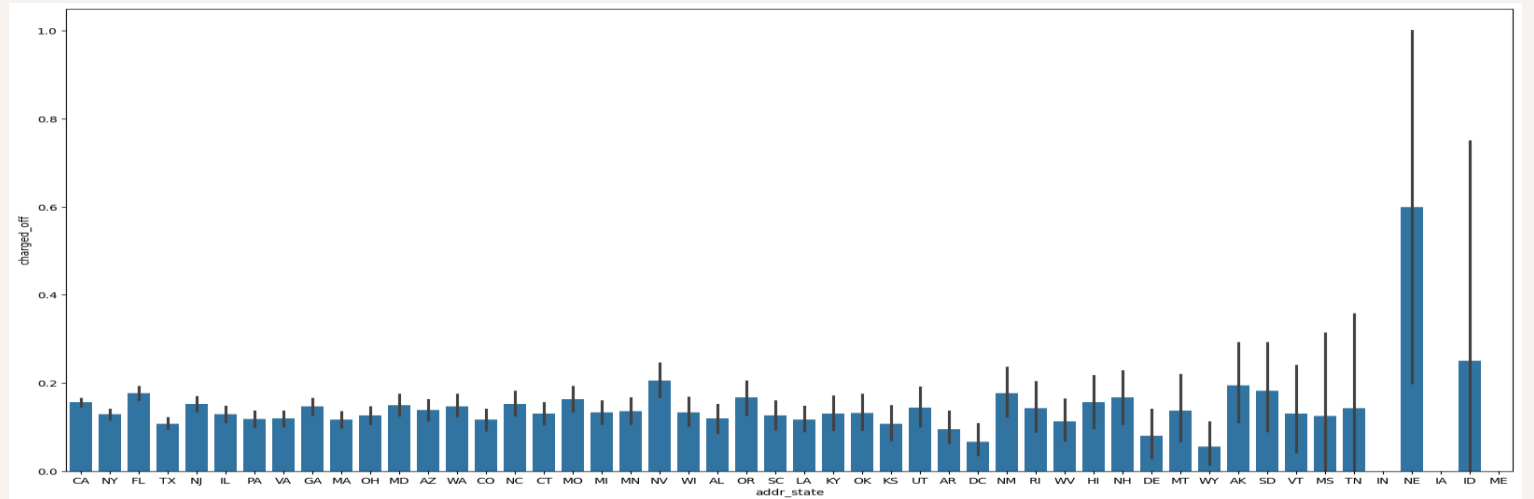
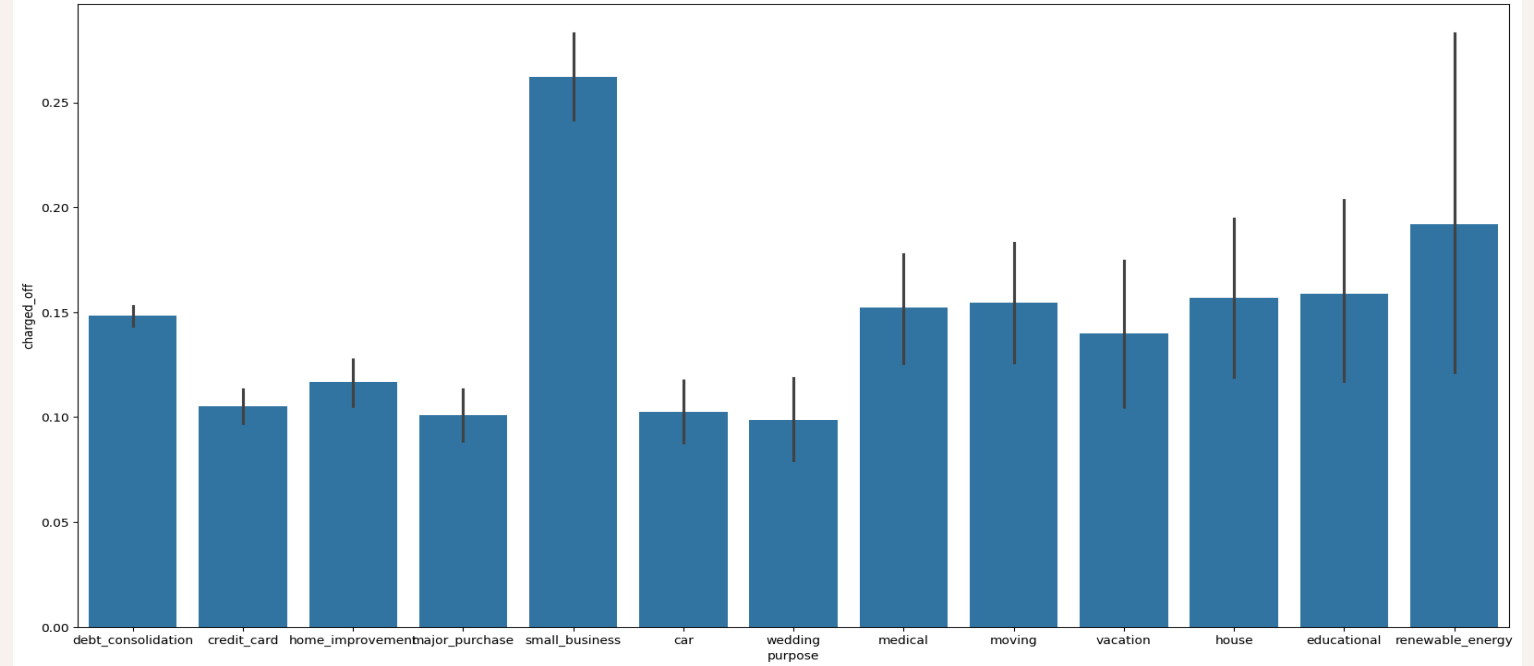


Bivariate/Multivariate Analysis

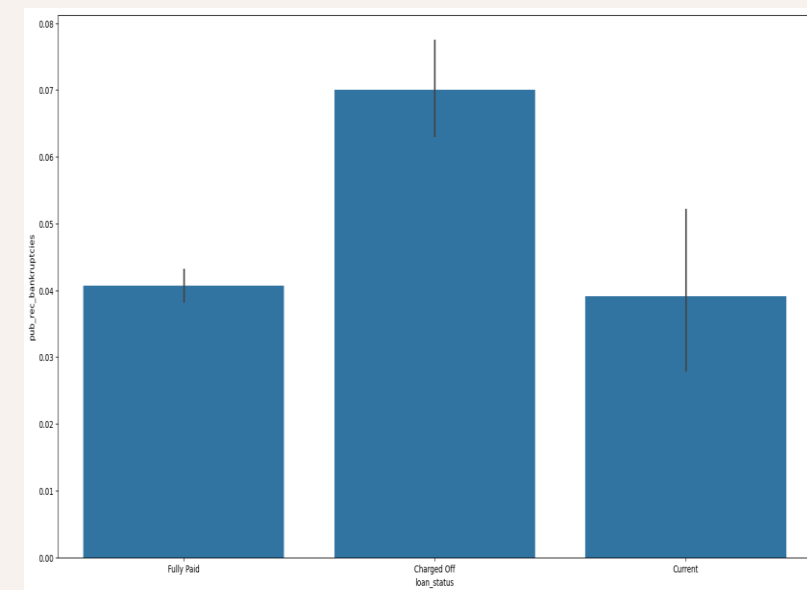
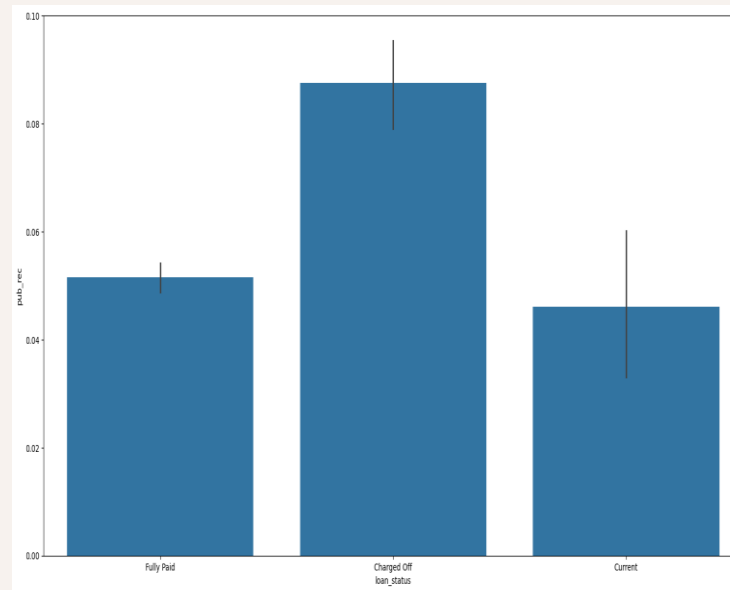
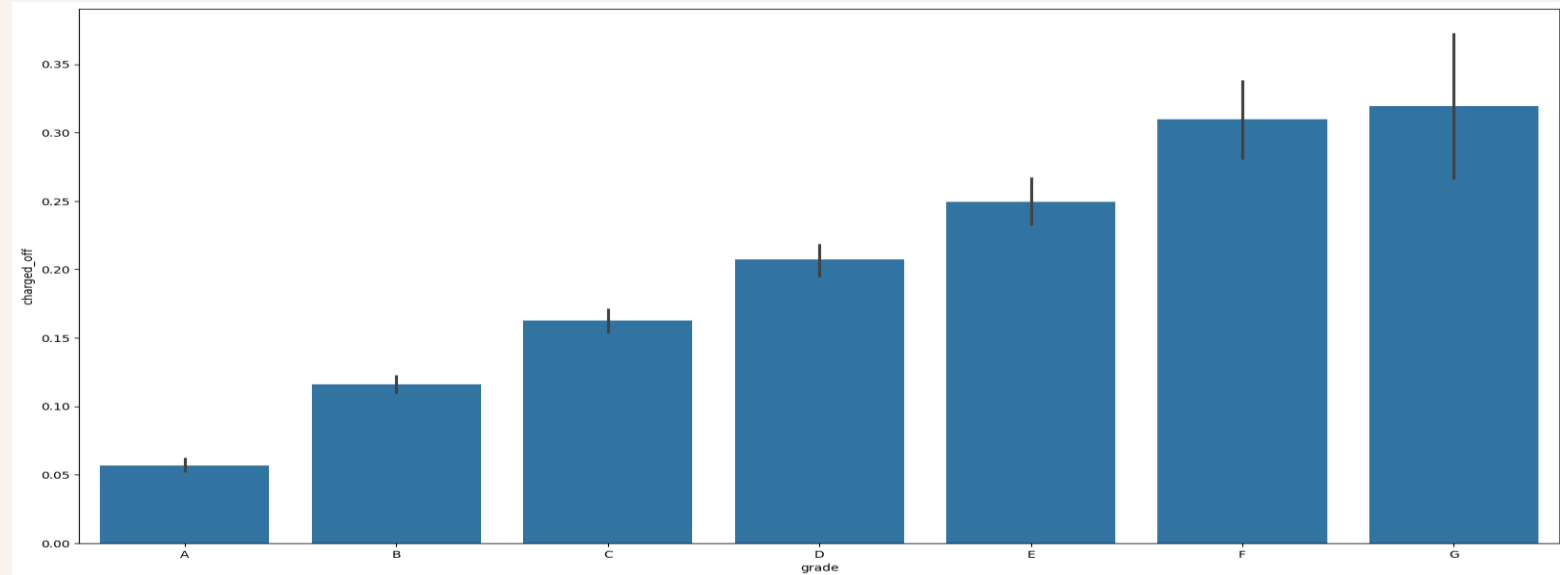


- High correlation between loan amount, funded amount, funded amount inv out of three any one can be used for analysis they are equally correlated to total payment, total rec principle, total payment inv

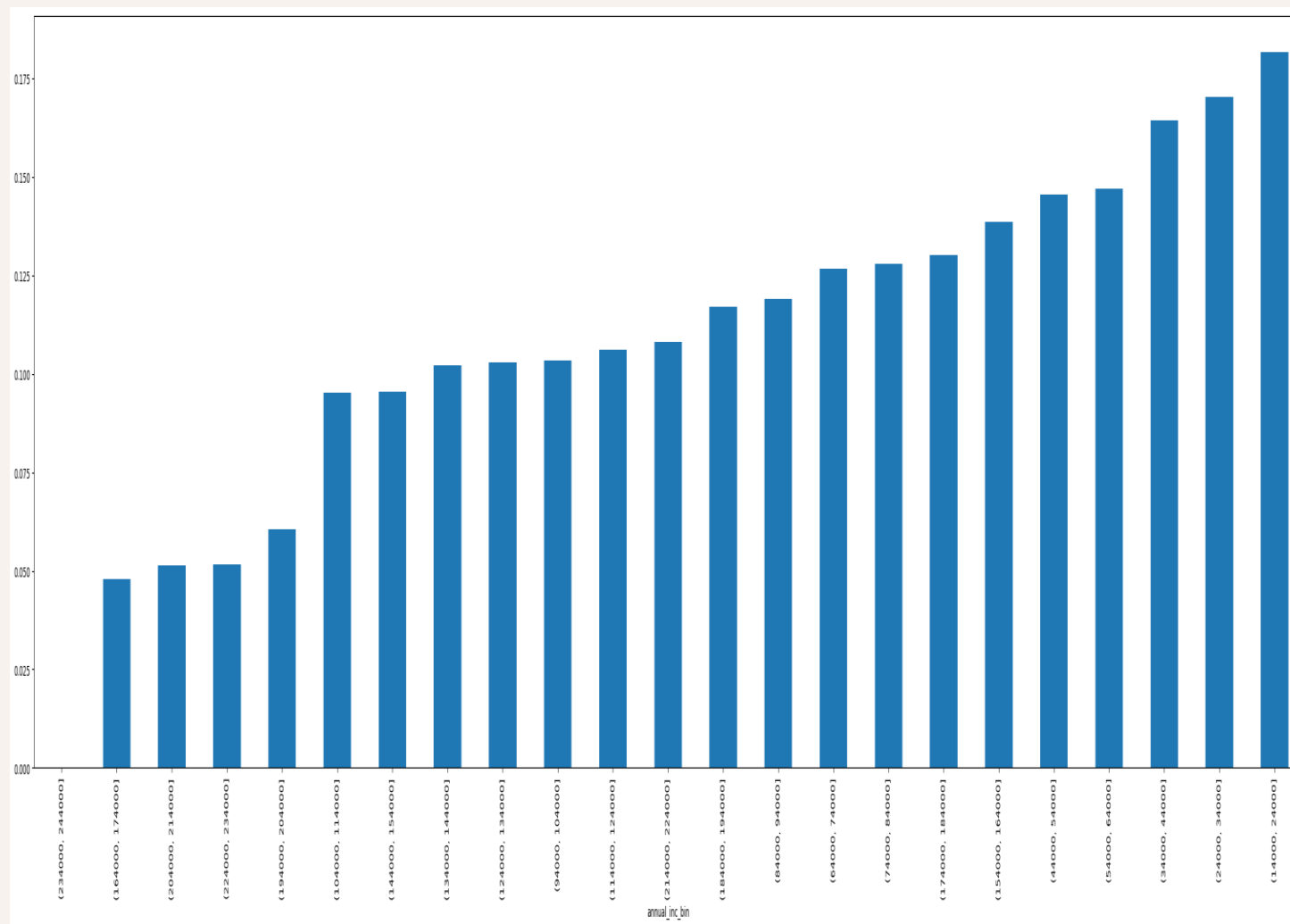
- Need to be cautious while giving loan to Small business purpose as this value has highest number of defaulters
- High defaults are happening in address state NE

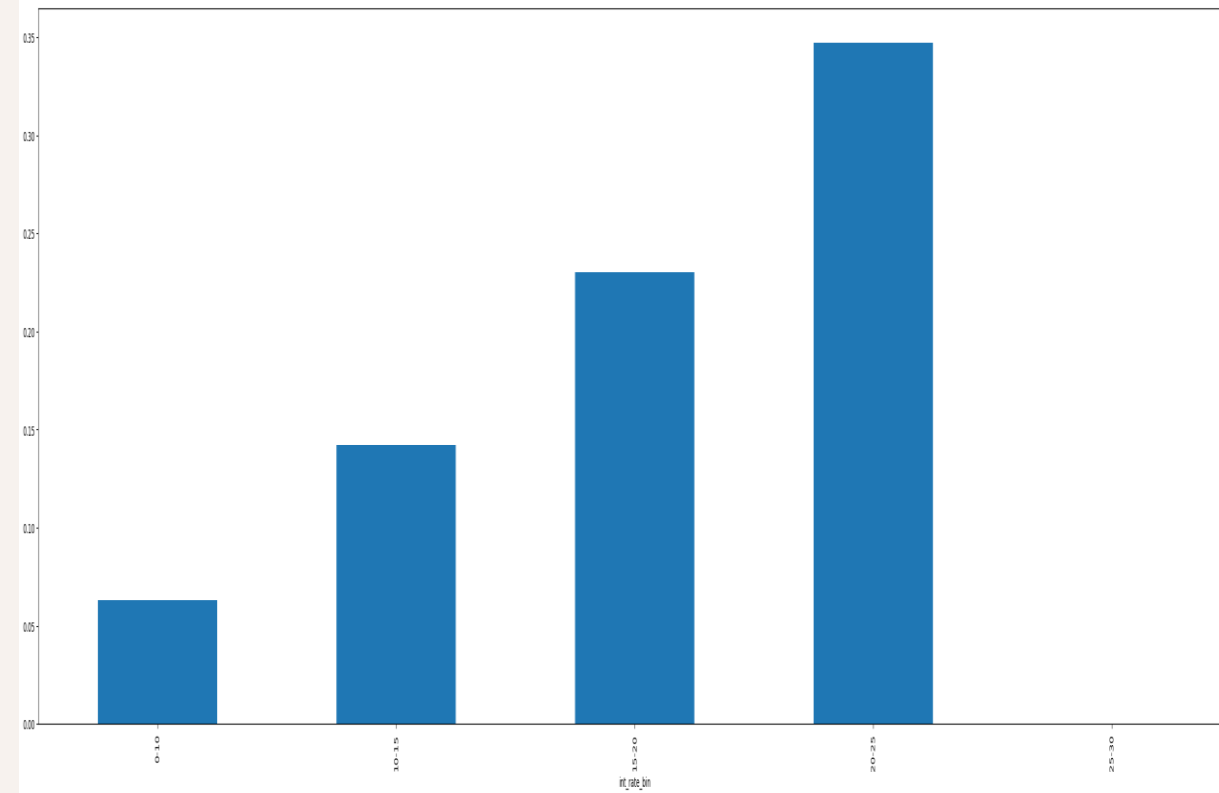
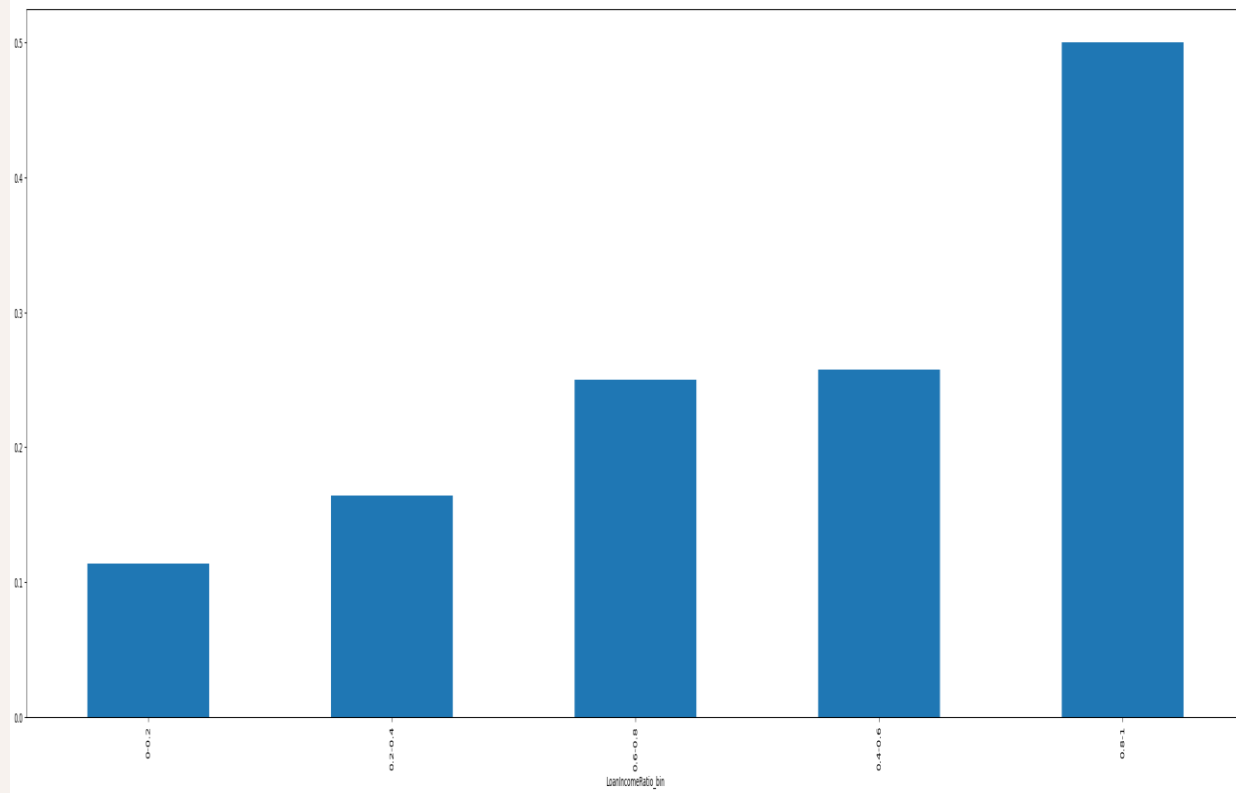


- As the Grade increases the person is more probable to default
- Company also need to check derogatory public records and public record bankruptcies as they are highly likely to default



- People with high annual salary are less likely to default





- As interest rate increases probability of person to default also increasing
- As per loan income ratio if a high loan given to less income person, the person is likely to default