LENDING CLUB CASE STUDY

Assignment
Name:

Pruthvi Raju Vegesna

PROBLEM STATEMENT

- A financial company which specialises in providing various types of loans to urban customers. When the company receives several loan application and has to make a decision for loan approval based on available data about applicant.

- Risks associated with the bank's decision are:
  - Applicant will repay loan completely and not giving loan will result in business.
  - Applicant may not repay loan and proving loan will lead to financial loss to company

- **Aim of this project is to identify patterns which indicate if a applicant is likely to default loan or not by doing EDA on loan and applicant attributes**

- 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
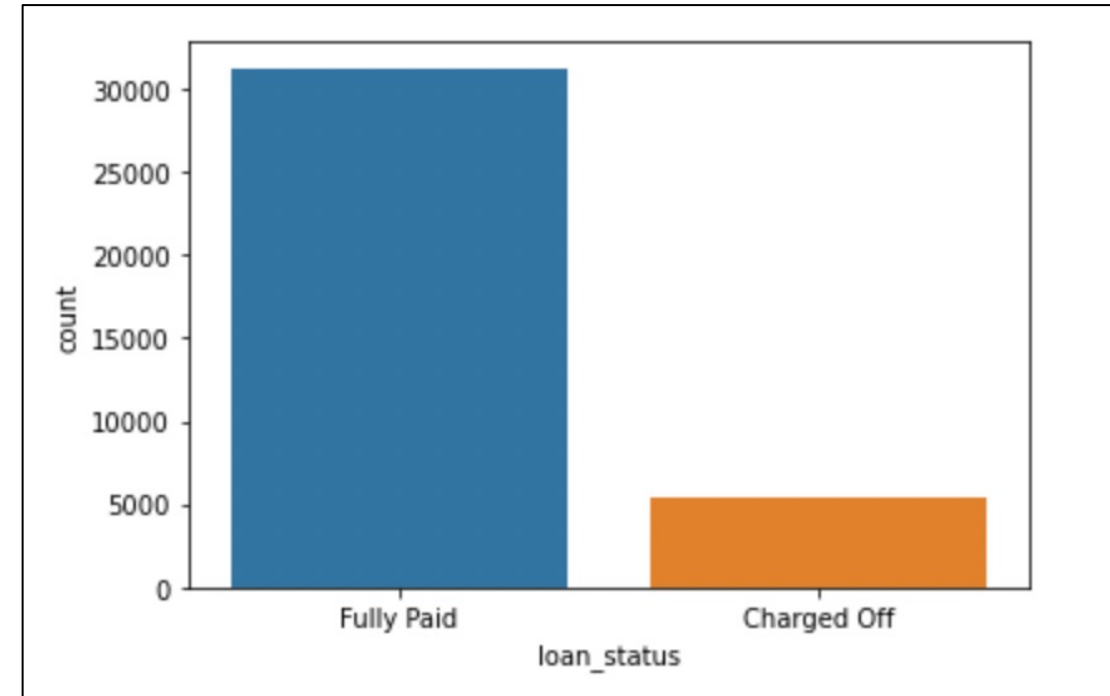
PROBLEM SOLVING WORKFLOW

- Understanding data & Identifying target columns

- Data cleaning and manipulation
  - Fixing rows and columns
  - Handling missing data
  - Convert data to important format
  - Removed outliers

- Data analysis
  - Univariate analysis : identify top variables
  - Segmented univariate analysis : done on target column
  - Derived metrics
 - Bi-Variate analysis :to understand combined impact

- Results and Conclusions

UNDERSTANDING DATA

- Target column is loan_status
  - %15 defaulted loans

- There are 3 types of data provide
  - Loan Attributes
  - Applicant attributes
  - Post loan attributes

- For our analysis we are interested in:
  - Loan & Applicant attributes
  - Post loan attributes don't come into picture while approving loan
  - Data with loan status as current are ignored

# DATA CLEANING AND MANIPULATION

- Fixing rows and columns
  - Dropped columns with all null & all same values
  - Removing data will loan status as current
  - Removed post loan attributes

- Imputing & Handling missing data
  - Replaced data with most appropriate data

- Convert data to important format(ex: remove %, date format etc.)
  - Converted obj to numeric (ex interest rate was 9%, removed %, term from 36 months to 3years(numeric)etc)
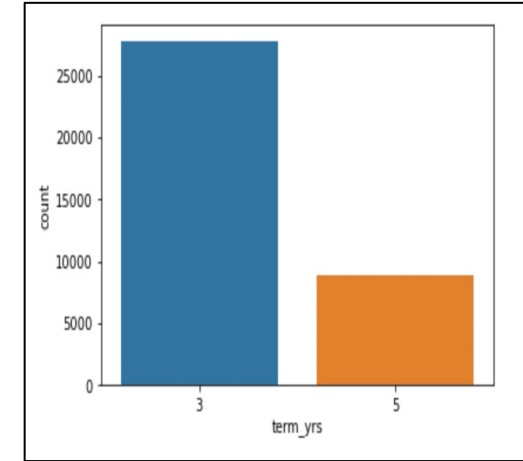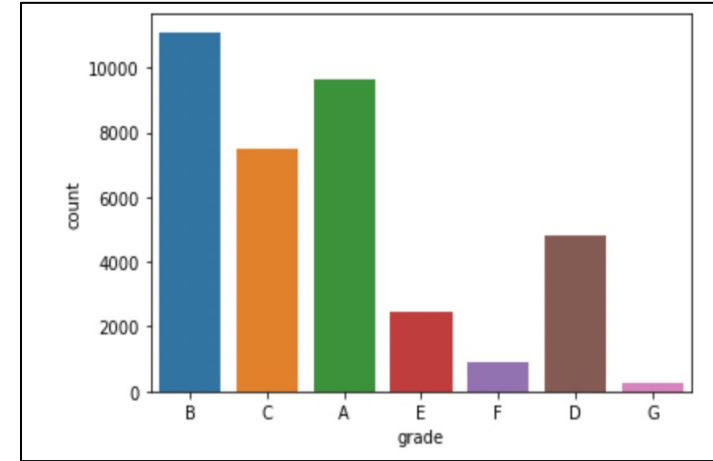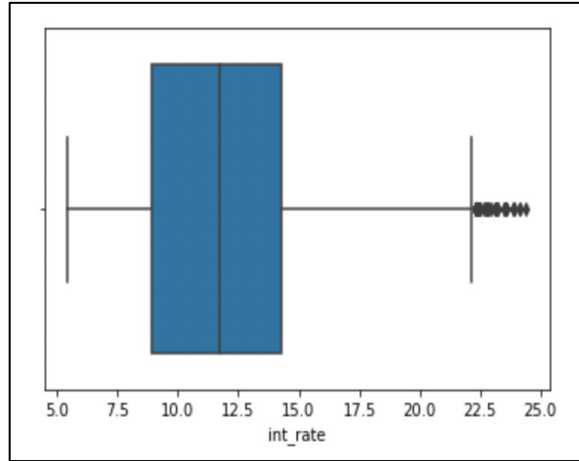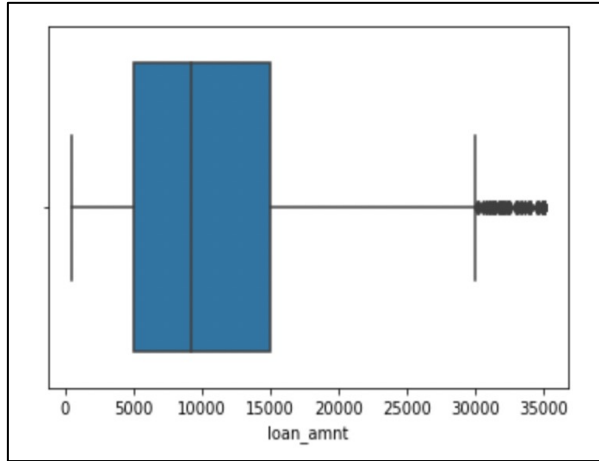 - Set date format

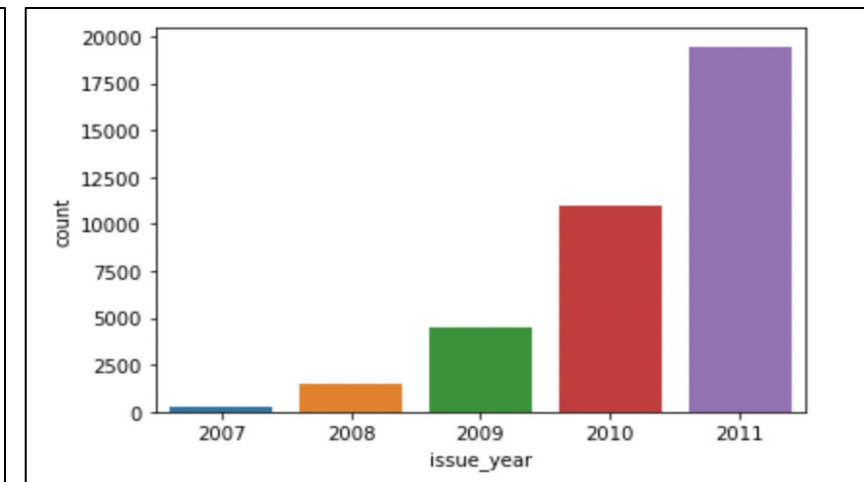- Removed Outliers
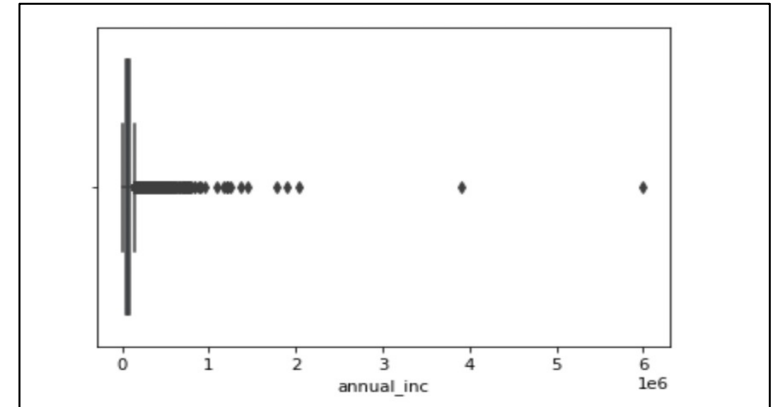
Annual Income
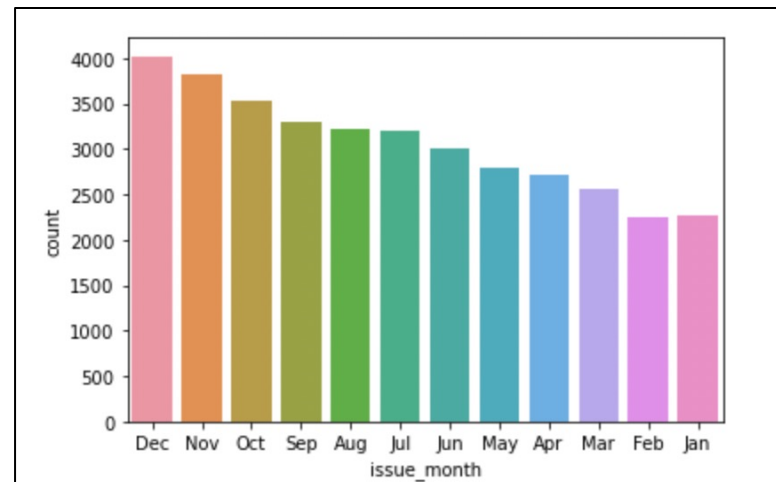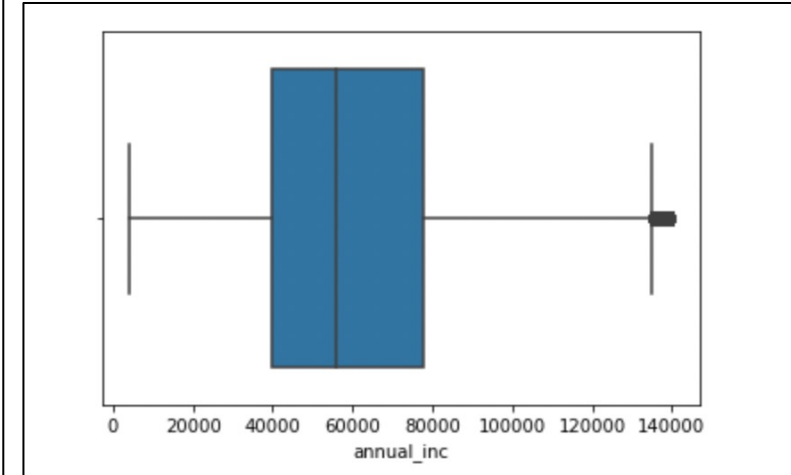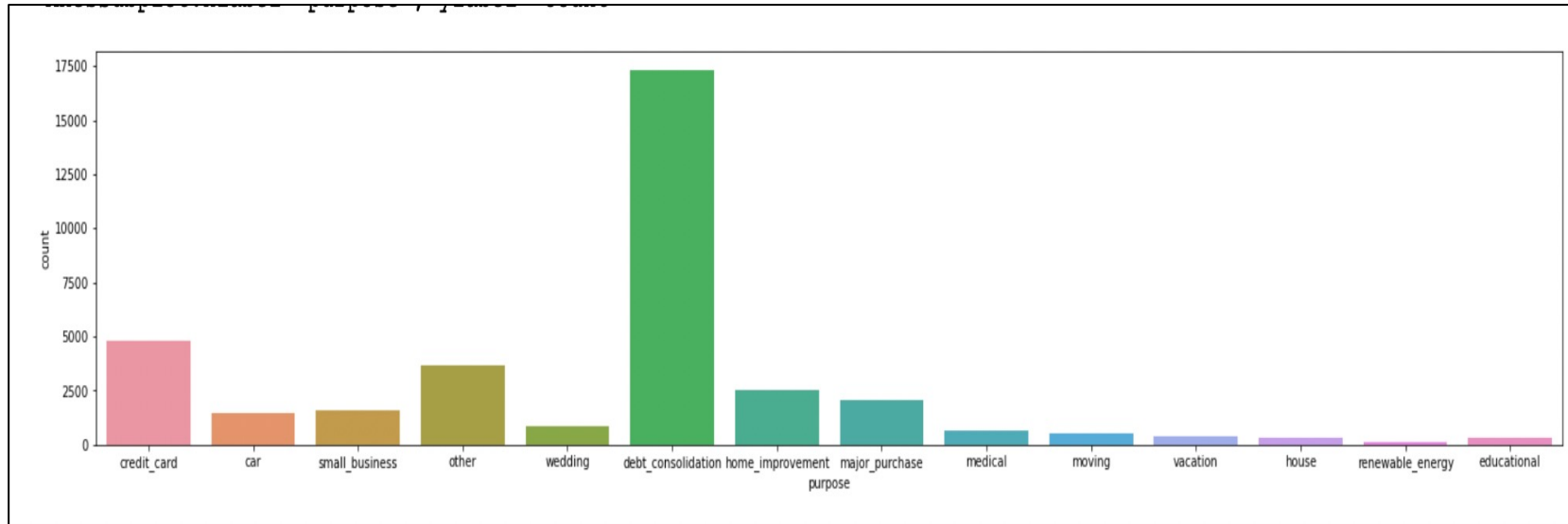
Removed Outliers
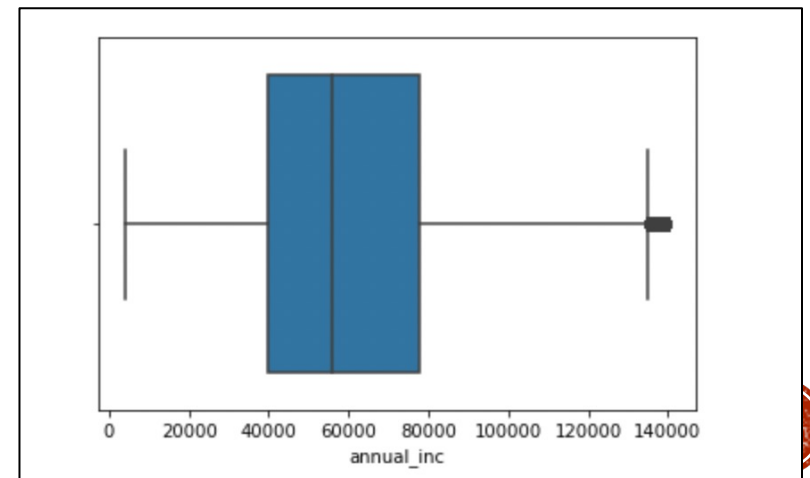
UNIVARIATE ANALYSIS ON LOAN

- **Loan Attributes(Loan):**
  - Average loan Amount is 9000
  - Avg Interest % is 11%
  - More loans are given for 3 year terms
  - Loan given grade decreases from A to G number of loans decreases
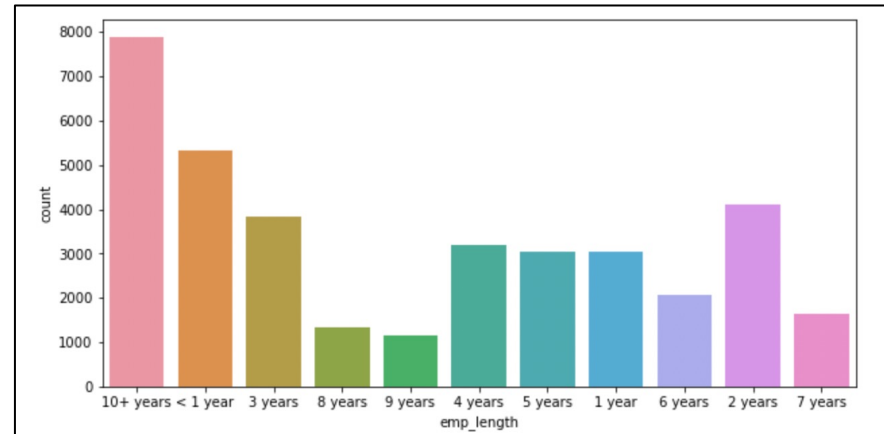 - Every year number of loans increased
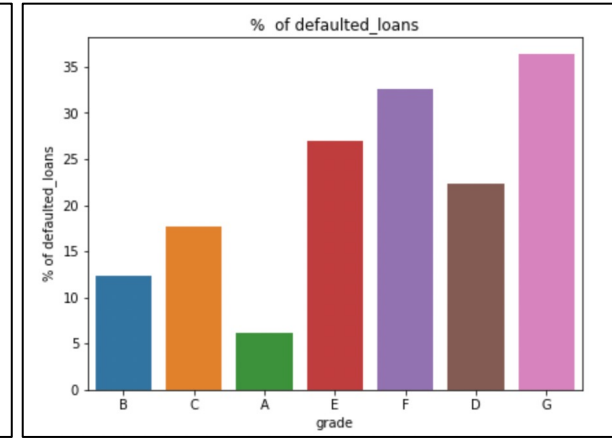 - Loans are given more during Dec (year end)
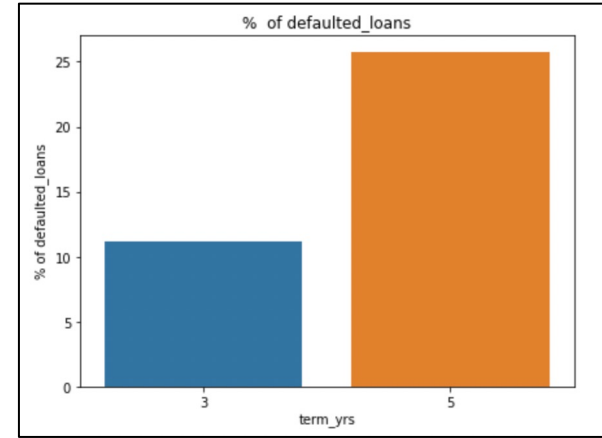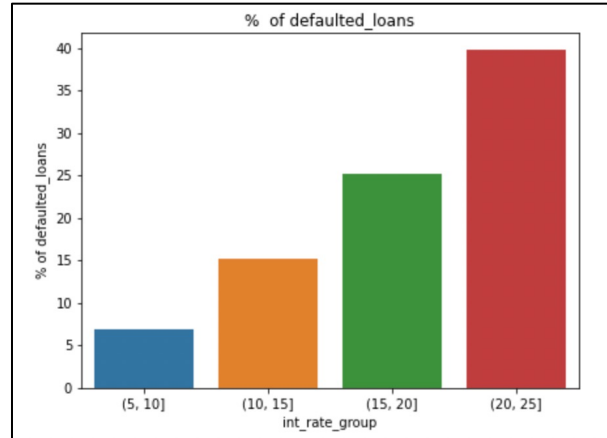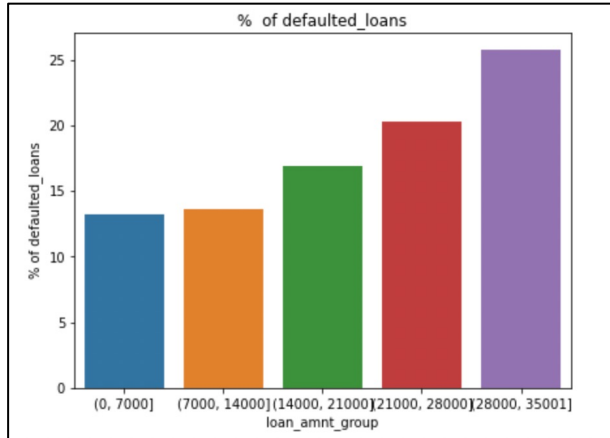
UNIVARIATE ANALYSIS ON LOAN

Applicant Attributes(Loan):

- Average annual income of applicants is 55000
- Applicants with <1 years and 10 yrs+ have taken more loans
- Reason for taking loan is mostly debt_consolidation
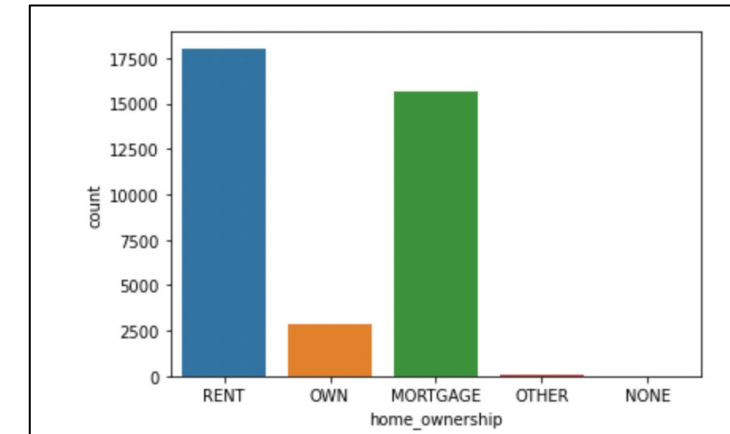
UNIVARIATE & SEGMENTED UNIVARIATE ANALYSIS

- **Loan Attributes(Defaulting):**
  - 1.) higher the loan amount higher chance of defaulting
  - 2.) higher the interest higher chance of defaulting
  - 3.) longer the tenure of loan higher chance of defaulting
  - 4.) As Loan grade increases from A to F defaulting increase
  - 4.) defaulting was low in 2008-2009-2010 might be due to effect of recession in 2008-09 where bank was more careful
 - 5.) generally defaulting is higher in month of Dec
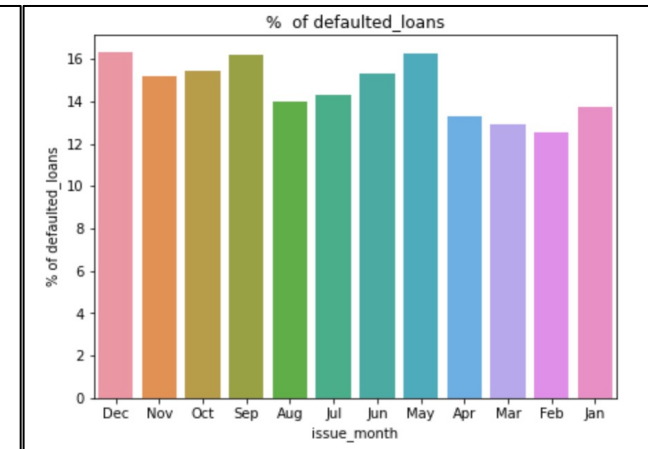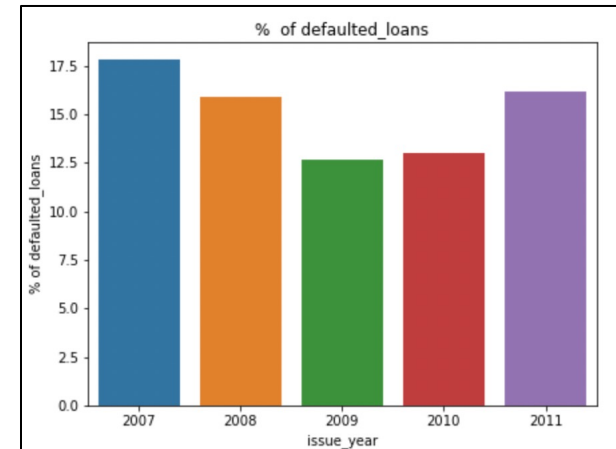
# UNIVARIATE & SEGMENTED UNIVARIATE ANALYSIS
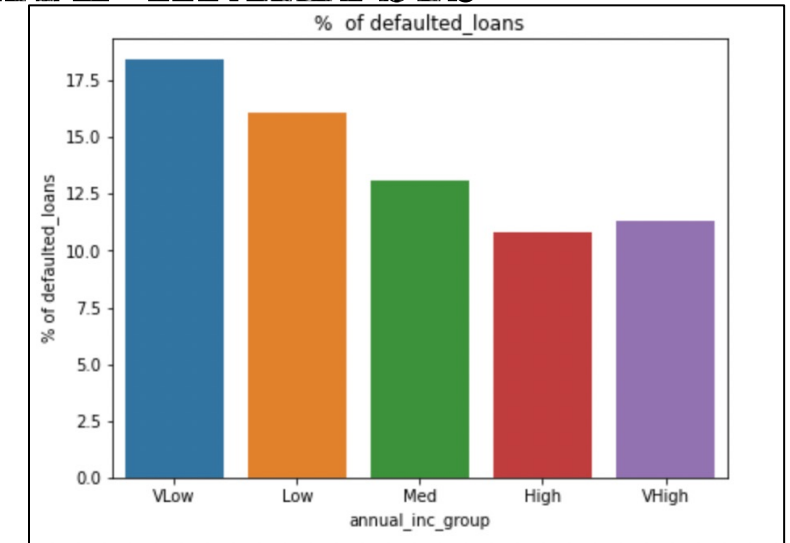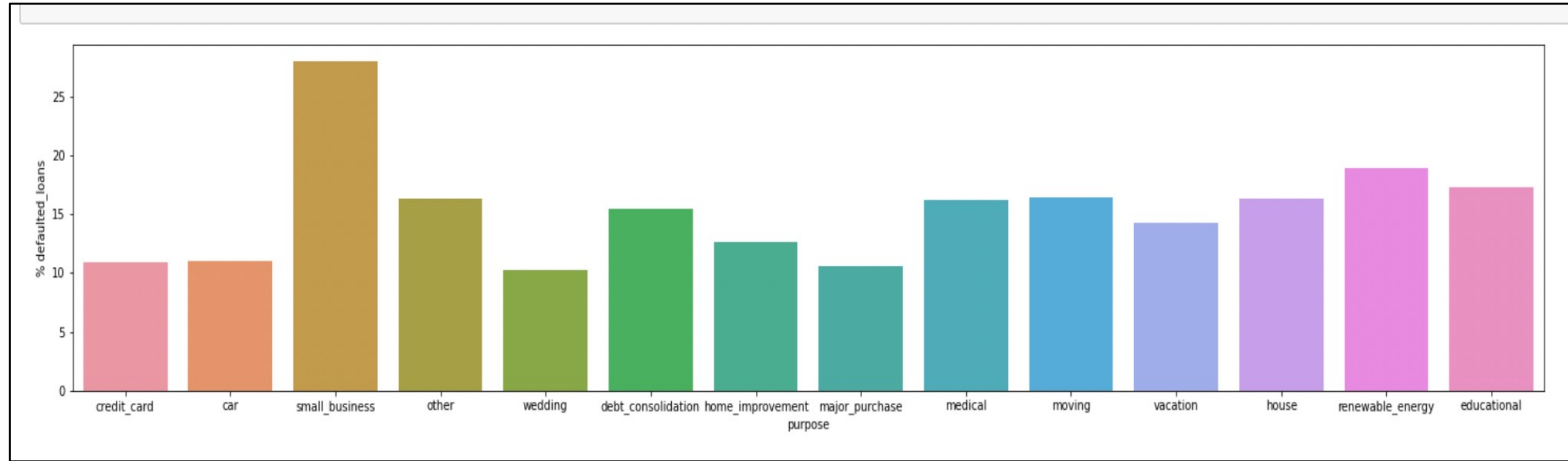
- **Applicant Attributes(Defaulting):**
  - 1.) Applicant with low annual income are likely to default
  - 2.) Person with reasons as small business & Renewable energy generally tend to default higher
  - 3.) person with record of bankruptcies tend to default loan
  - 4.) generally as 'dti' increases defaulting increased
  - 5.) Applicant address doesn't matter most of times expect for state NE can be ignored as no of loans are very low for that state and is outlier
  - 6.) Applicant with more no of open accounts tends to default higher except last bin where
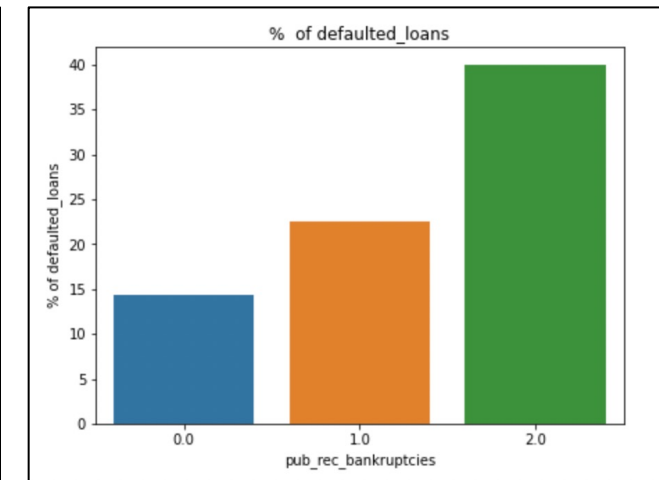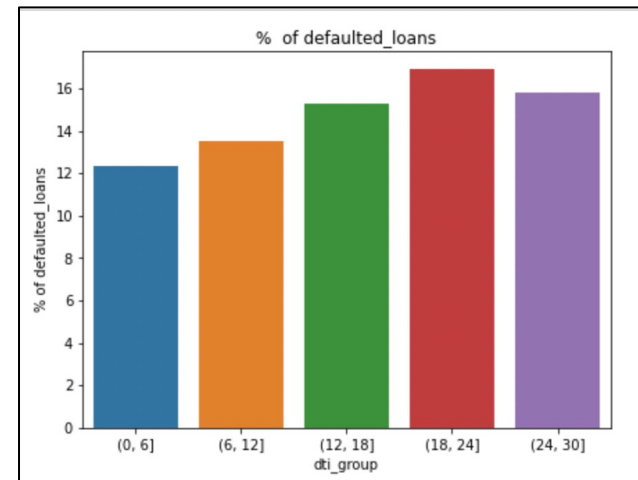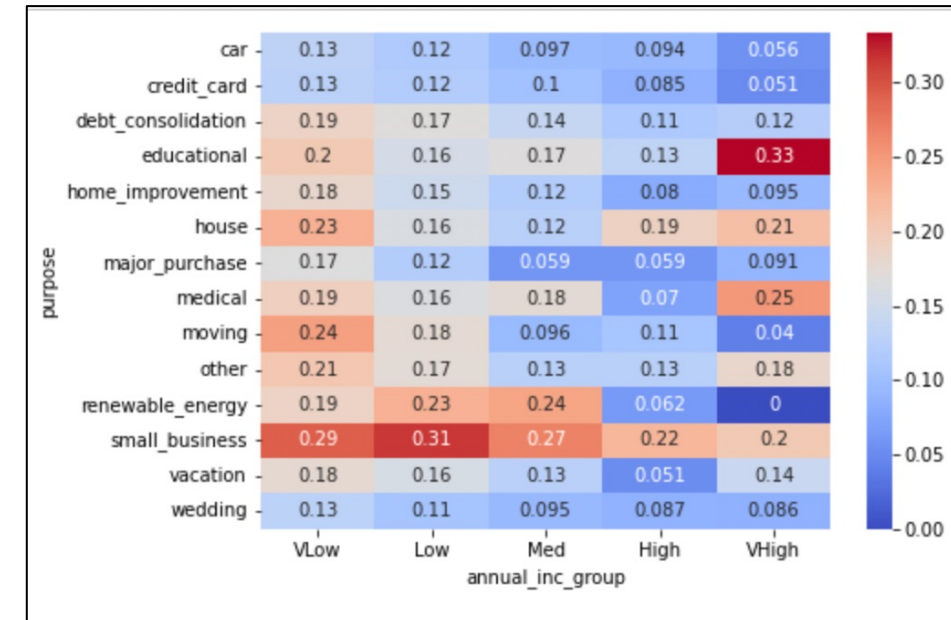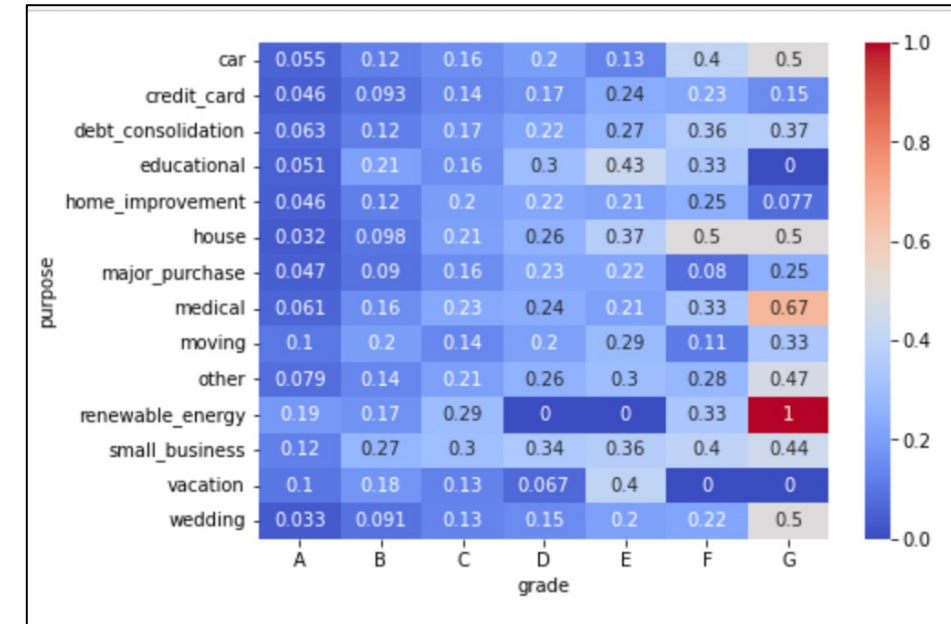
BI-VARIATE ANALYSIS
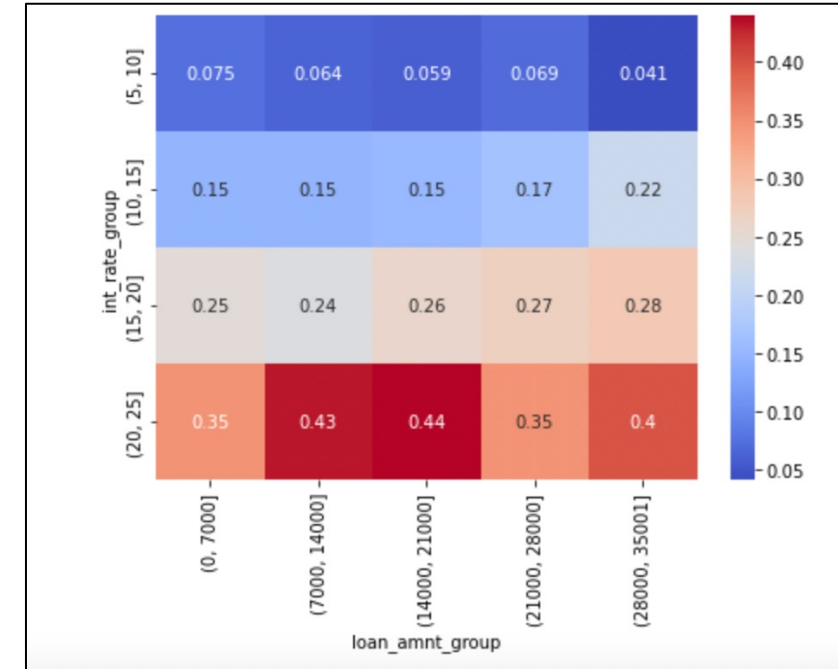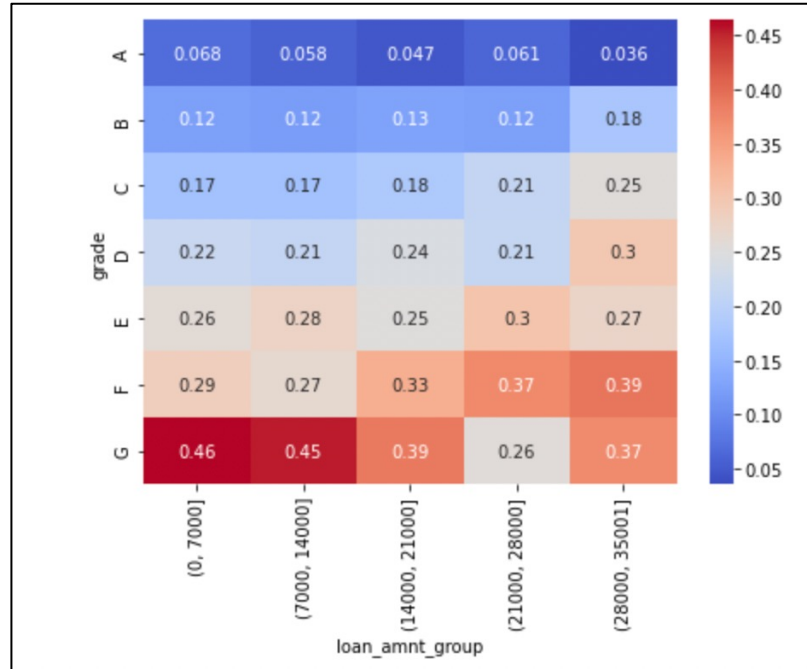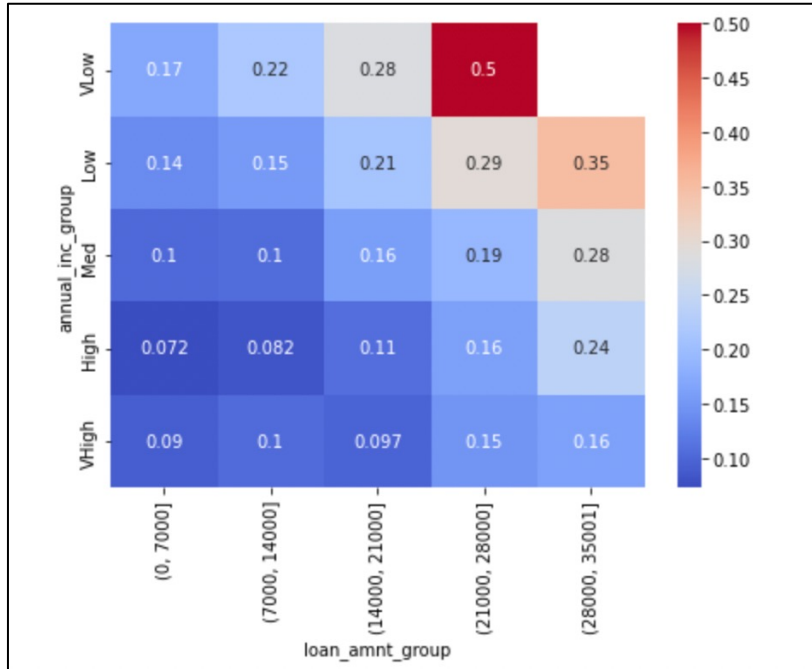
- Loan amount and annual income tend to have combined impact
- Loan grade G and loan amount lower tend to default more
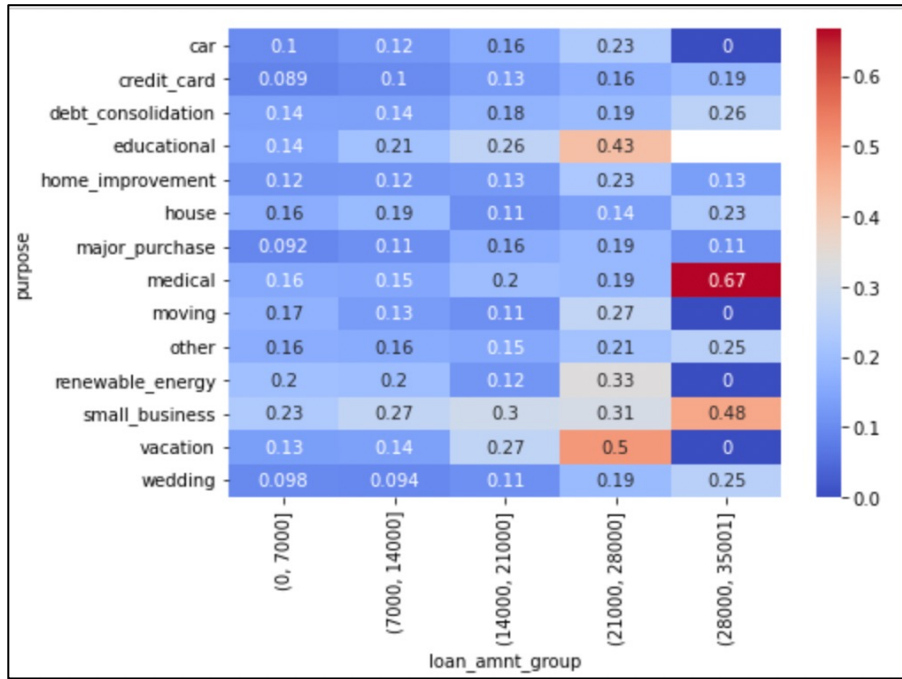
- Applicant who take loan for (medical, small business, vacation, educational, renewable) purpose and higher amount tend to default higher
- Applicant with grade G and medical or renewable as purpose mostly default loan
- Applicant with small business as purpose default irrespective of annual Income

SUMMARY

- **Loan attributes** leading to defaulting :
  - Higher the interest higher the chance of defaulting
  - Higher the loan amount chance of defaulting is higher
  - Longer the tenure of loan higher chance of defaulting
  - Lower the loan grade (Assuming G is lower compared to A), defaulting increased

- **Applicant attributes** leading to defaulting :
  - Lower the annual income of applicant higher the chance of defaulting
  - Applicant with reasons as 'small business' or 'renewable energy' tend to default loan
  - Applicant with higher 'dti' tend to default loan
  - If applicant having record of bankruptcies will most likely default loan

  We see some combined impact of few variables also

THANK YOU