Lending Club Case Study Assignment

Submission

Group Members

1. Prince Grover

2. Vasanth Kumar

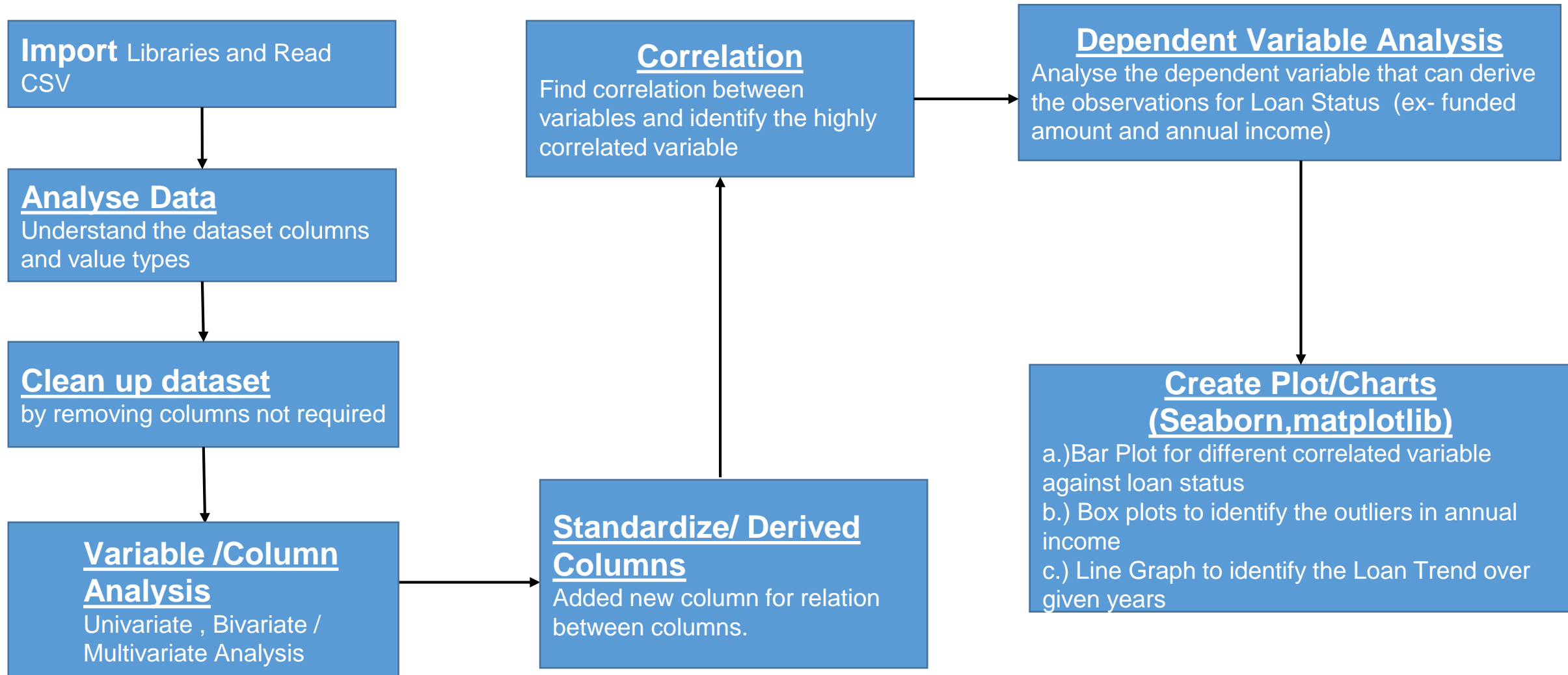# Business Object and Strategy

Business Objective :

- Consumer Finance company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

- In other words, the 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 utilise this knowledge for its portfolio and risk assessment.

Strategy:

Need to analyse different variable provided in the dataset to identify the driving variables for Loan Defaults.ie, Charged-Off

1. Identify the type of variables
2. Find and remove the variables not needed for Driving factor identification
3. Create Derived variable if needed and analyse the trend with other variable
4. Provide the Chart/ Plot View of the dataset for the above defined objectives

Exploratory Data Analysis - Flow

Given Files : (i) loan.csv (ii) Data_Dictionary.xlsx

Import Libraries and Read CSV

Analyse Data
Understand the dataset columns and value types

Clean up dataset
by removing columns not required

Variable /Column Analysis
Univariate , Bivariate / Multivariate Analysis

Standardize/ Derived Columns
Added new column for relation between columns.

Correlation
Find correlation between variables and identify the highly correlated variable

Dependent Variable Analysis
Analyse the dependent variable that can derive the observations for Loan Status (ex- funded amount and annual income)

Create Plot/Charts (Seaborn,matplotlib)
a.)Bar Plot for different correlated variable against loan status
b.) Box plots to identify the outliers in annual income
c.) Line Graph to identify the Loan Trend over given years

Analysis

The essence of the whole project is to analyse and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

We performed **data cleaning and preparation** on the Loan dataset:

- Removed the identified not required columns

- Imputed the NA values for the identified variables

- Formatting the data is done

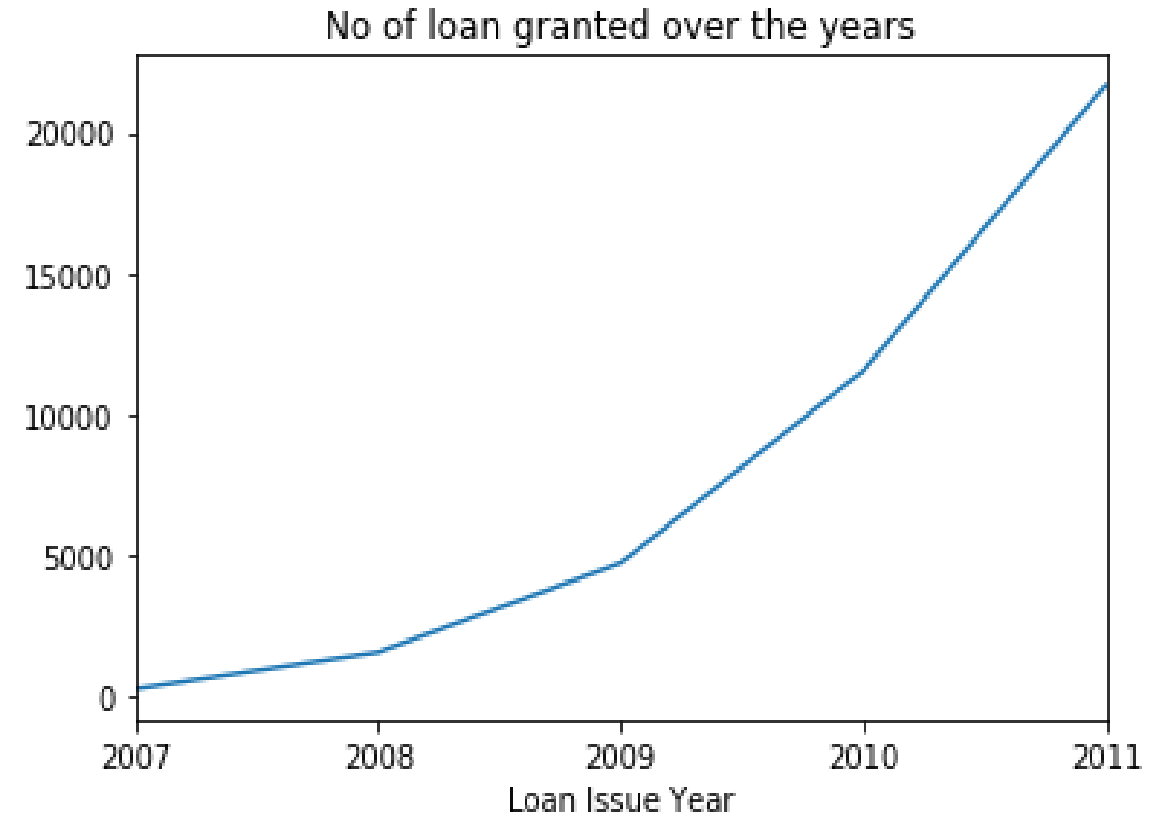During **Data analysis** we have created:

- Bar charts to check out the distribution of all the driver variables

- Box plots to detect the Outliers

Univariate , Bi/Multivariate Analysis

- To understand how different variables interact with each other.

Loan Status Analysis
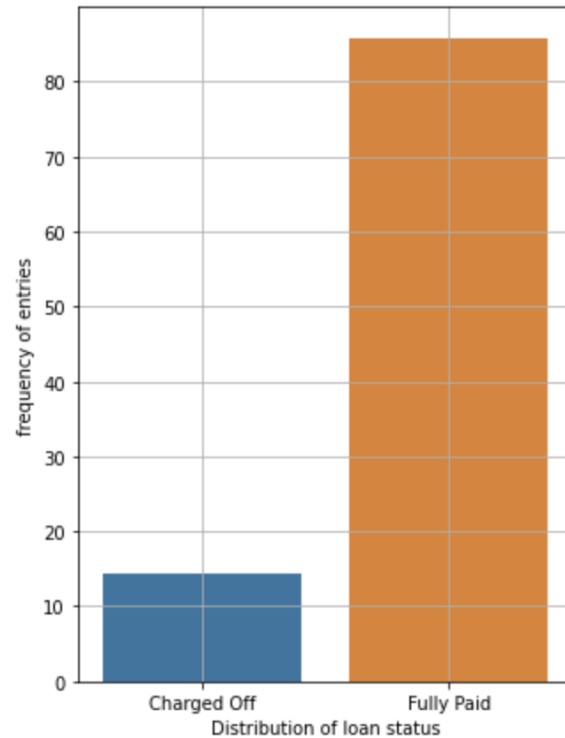
|	loan_status	count
0	Charged Off	14.33
1	Fully Paid	85.67

No of loan granted over the years

Observations:

1. Most of the loans are Fully Paid.

2. About 14% of loan are having status as Charged Off.

3. The loan has been increasing exponentially over the years.

# **Income Analysis** for the loan applicants
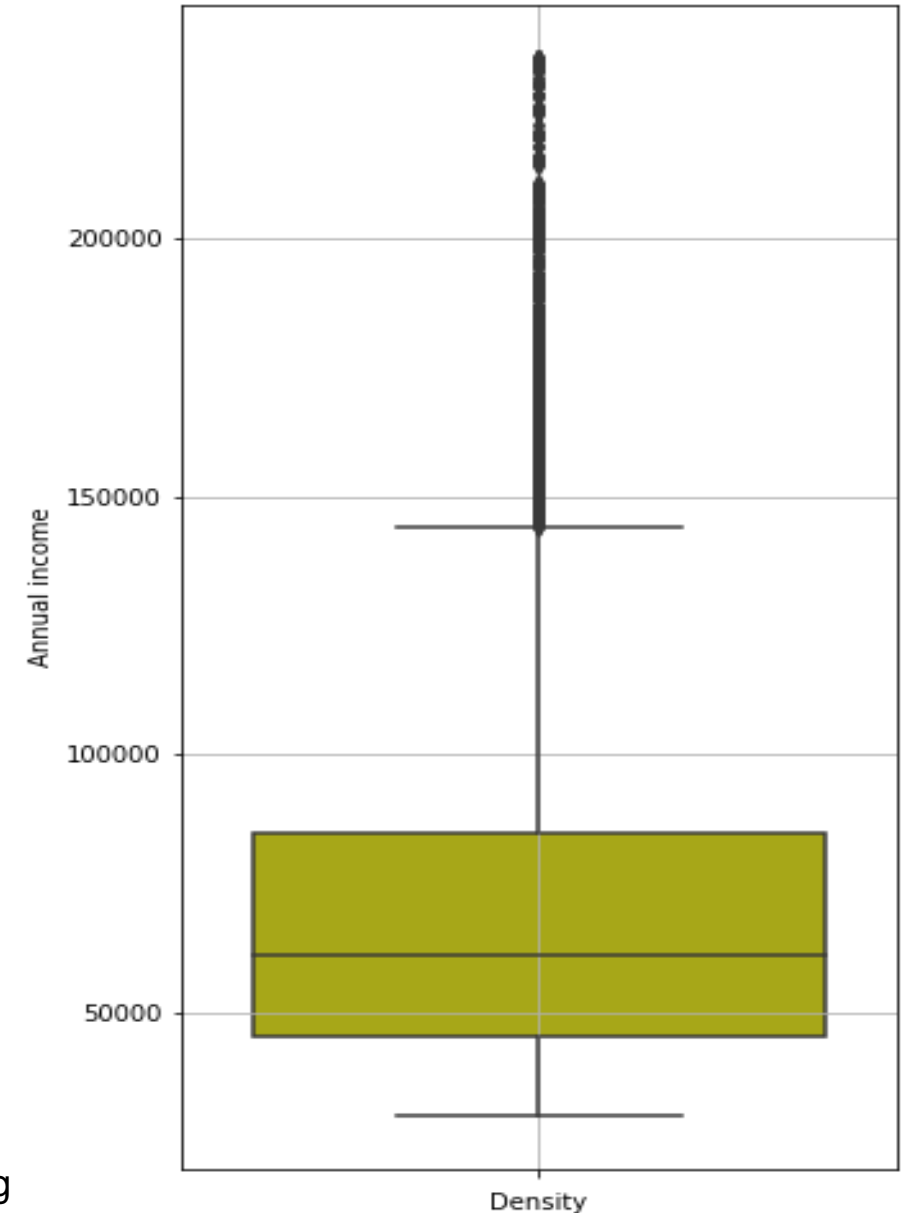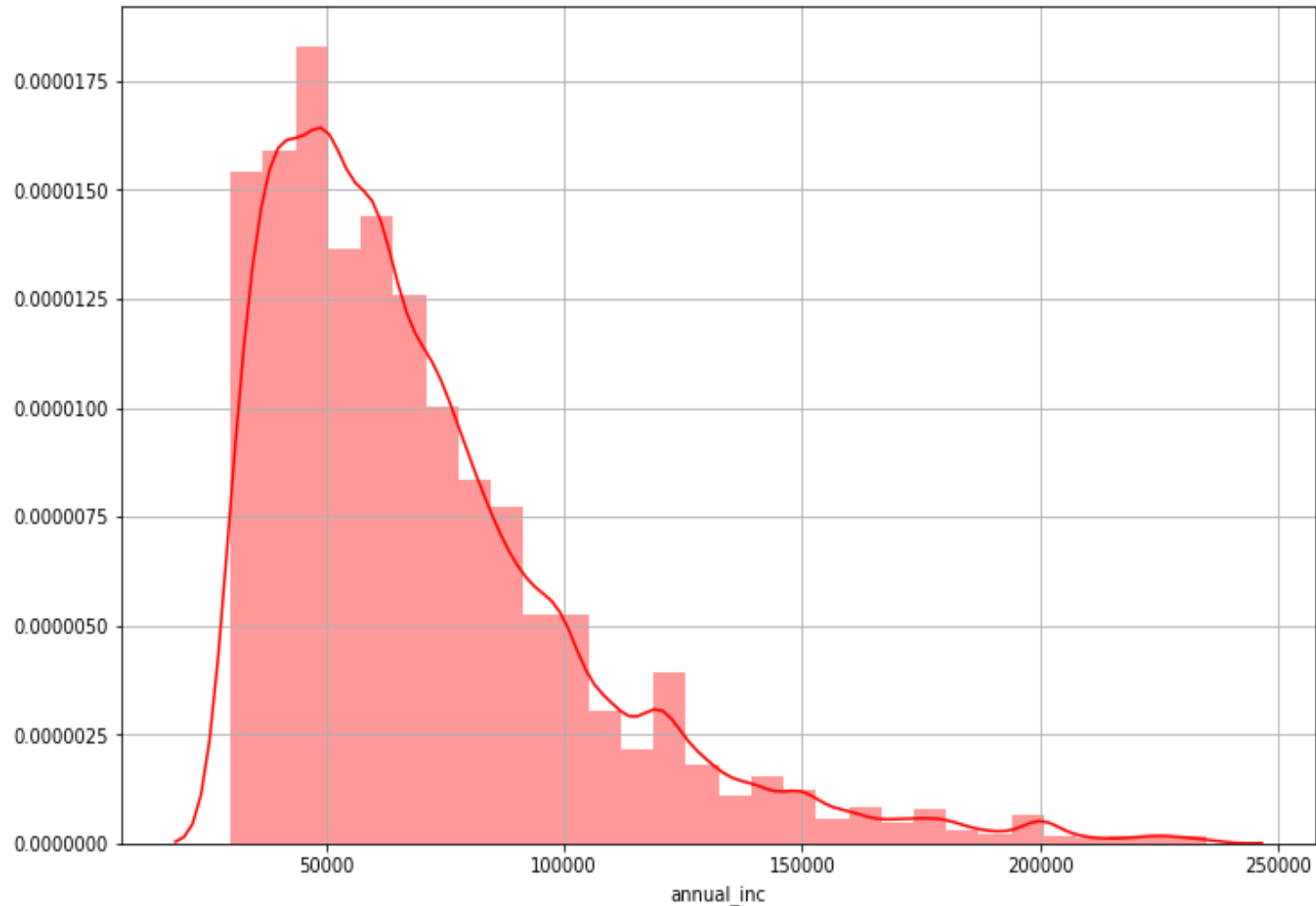
Observations:
1. There are two outliers having the income more than 3000000 (30 lakhs)
2. The annual income is less than 500000 (5 lakhs) with the exception of two outliers having more than 30 lakhs
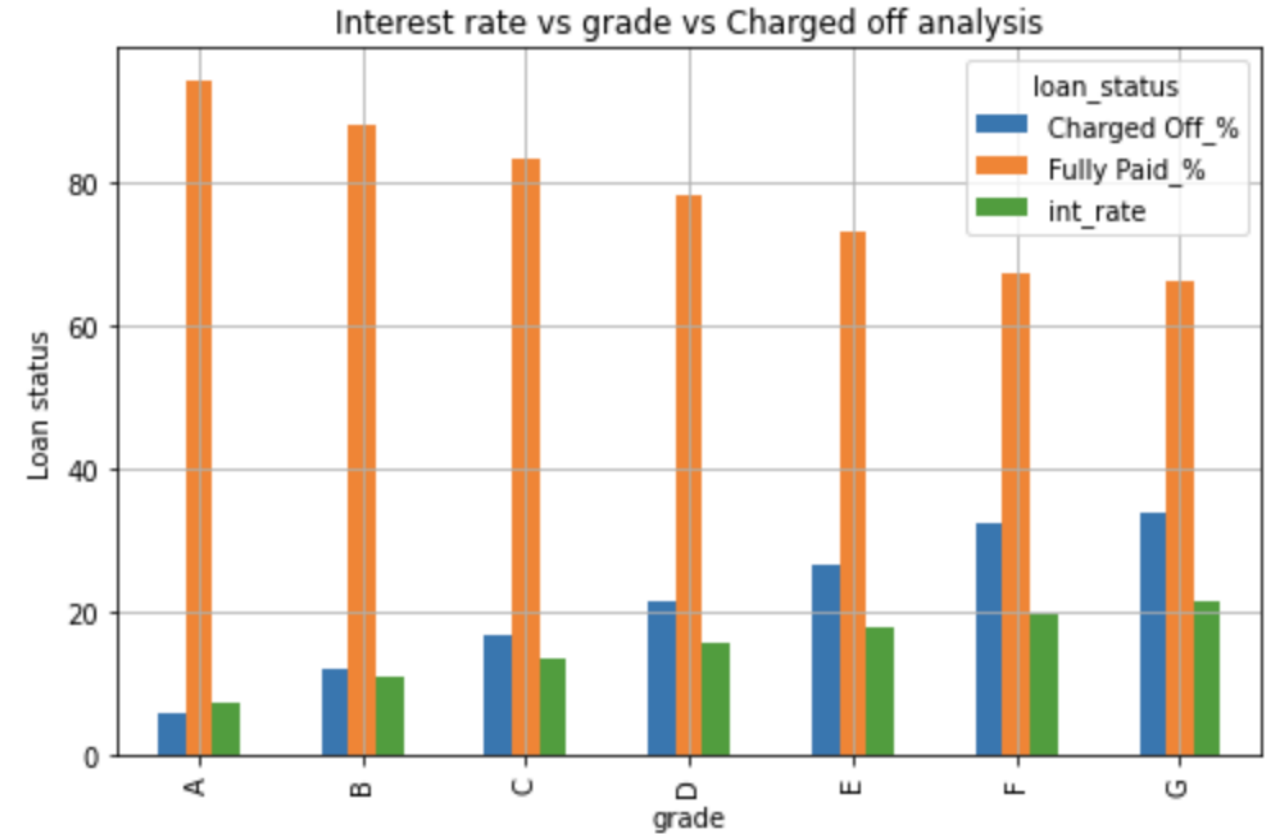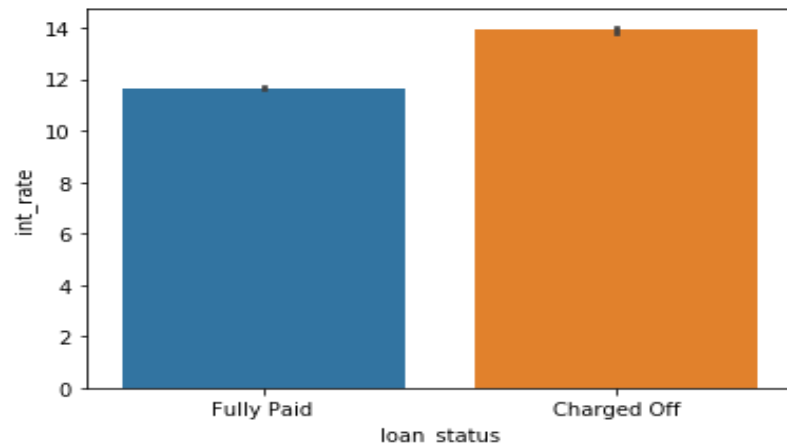
Loan Term Analysis with Purpose
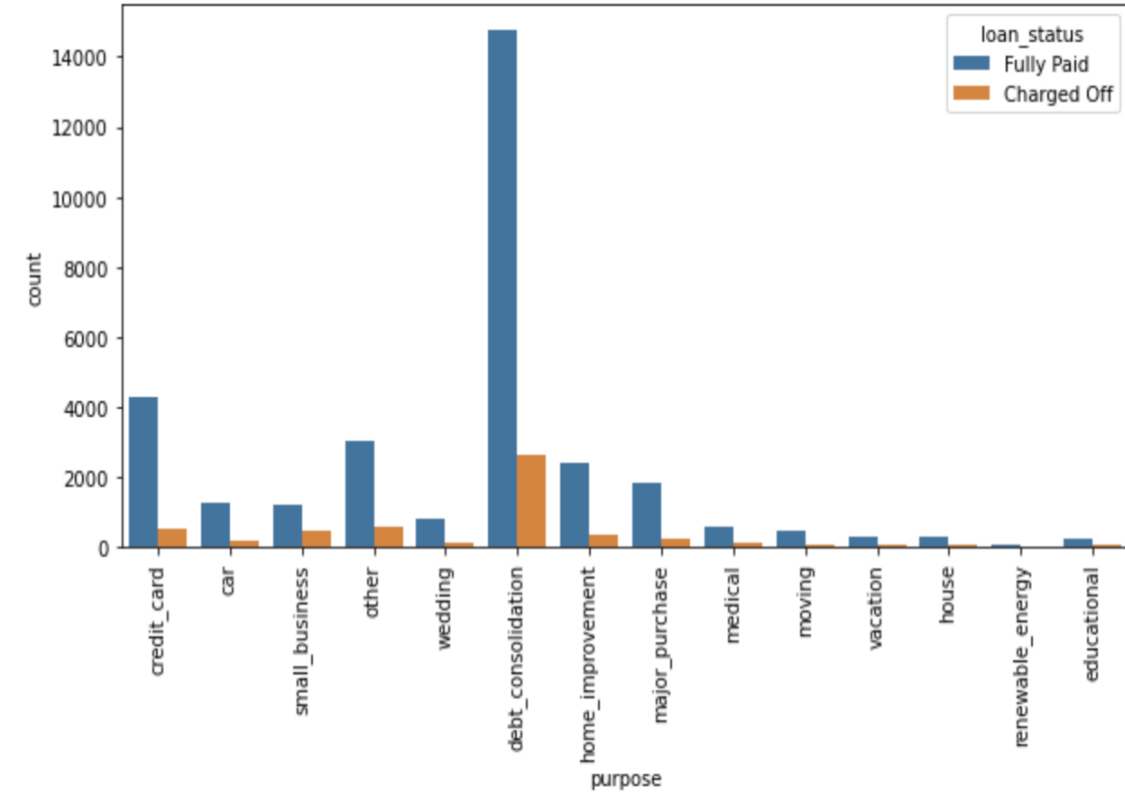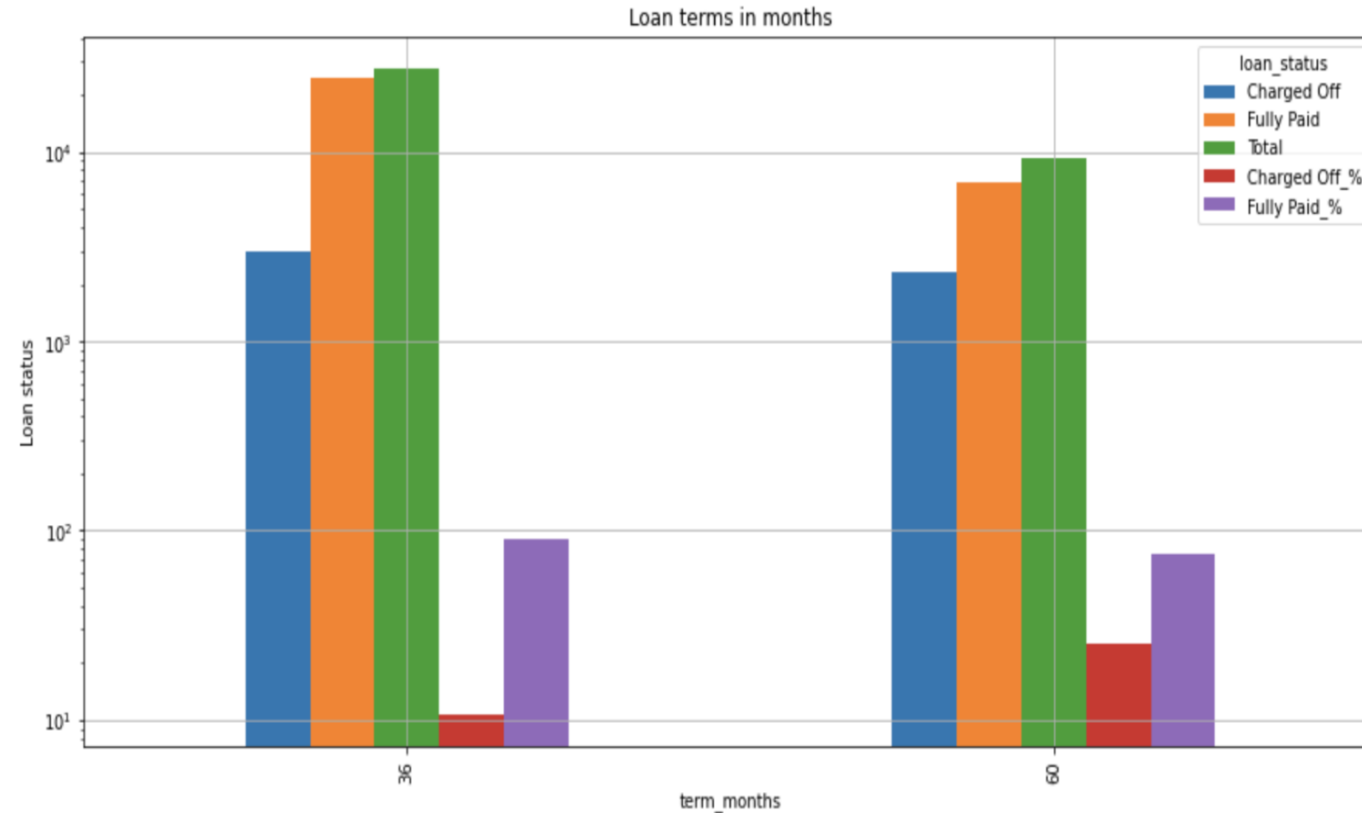
Observations:

1. Most of the loans are granted were for debt consolidation.
2. The Debt Consolidation cases have highest Charged Off compared to other purposes.
3. The total num of loans are higher for 36 months term
4. Charge Off % is more in 60 months term than 36 terms loan.

loan_status	Charged Off	Fully Paid	Total	Charged Off_%	Fully Paid_%
term_months					
36	2955	24635	27590	10.71	89.29
60	2326	6931	9257	25.13	74.87

Interest Rate and Grade Analysis

Loan garde
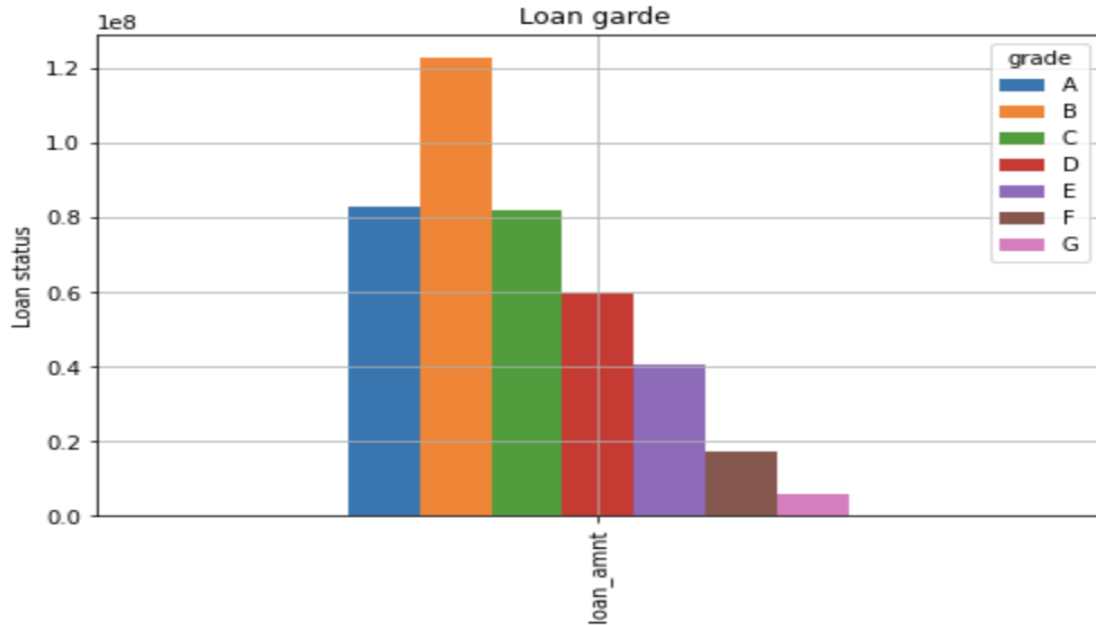
grade	A	B	C	D	E	F	G
loan_amnt	83009475	122866050	81677600	59377575	40508375	17093400	5992025

Interest rate vs grade vs Charged off analysis

Observations:
1. The Interest Rate for Charged Off Loans is Higher
2. With Increasing Grade , the charge Off % and respective Interest Rates are increasing respectively
3. Charged Off % is directly proportional to interest rate.
4. Loans given to grades E, F, G needs **more scrutiny** from Lending Club as charged off % is higher compared to other grades.

UpGrad

Bankrupcy rate vs Charged off Rate Analysis
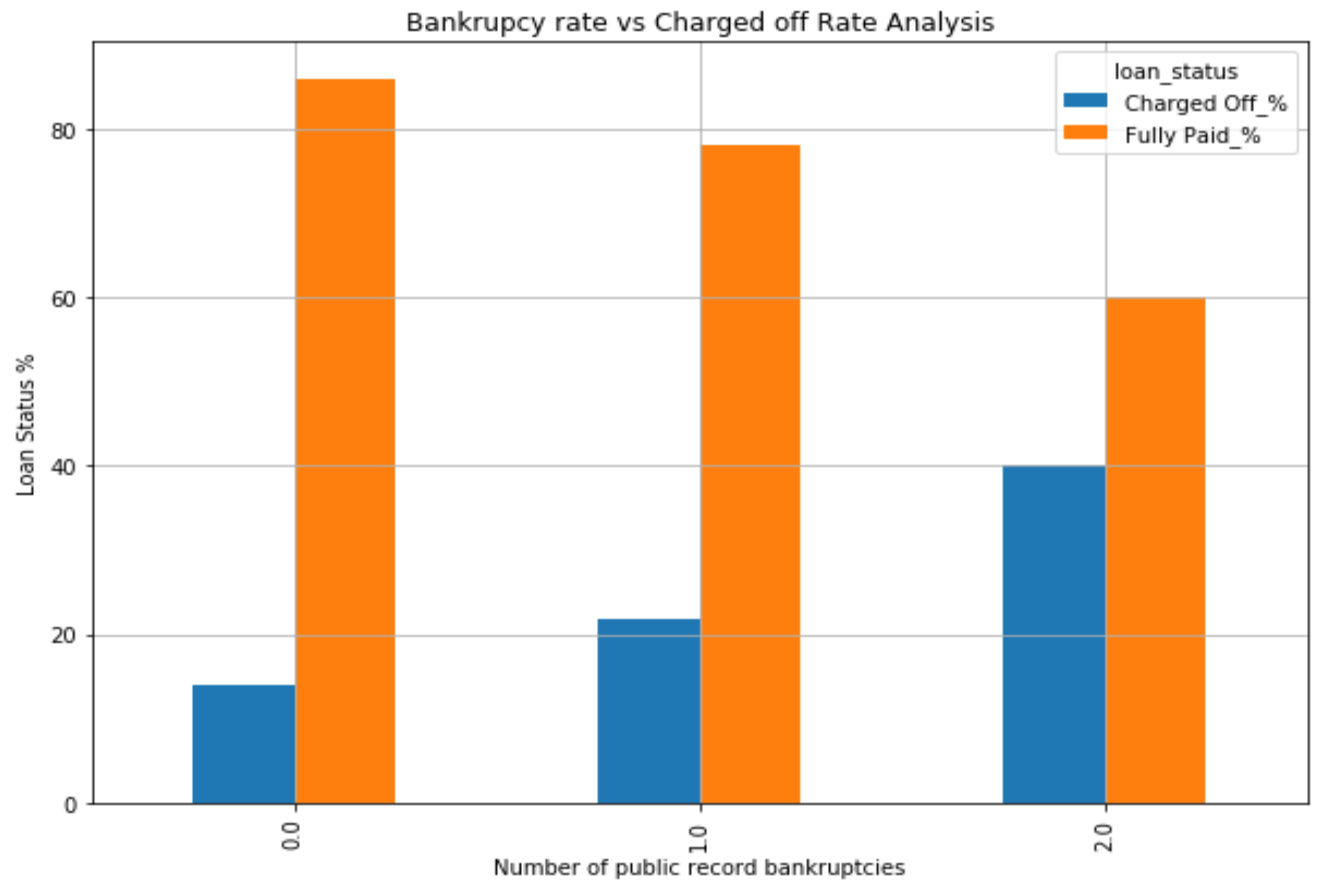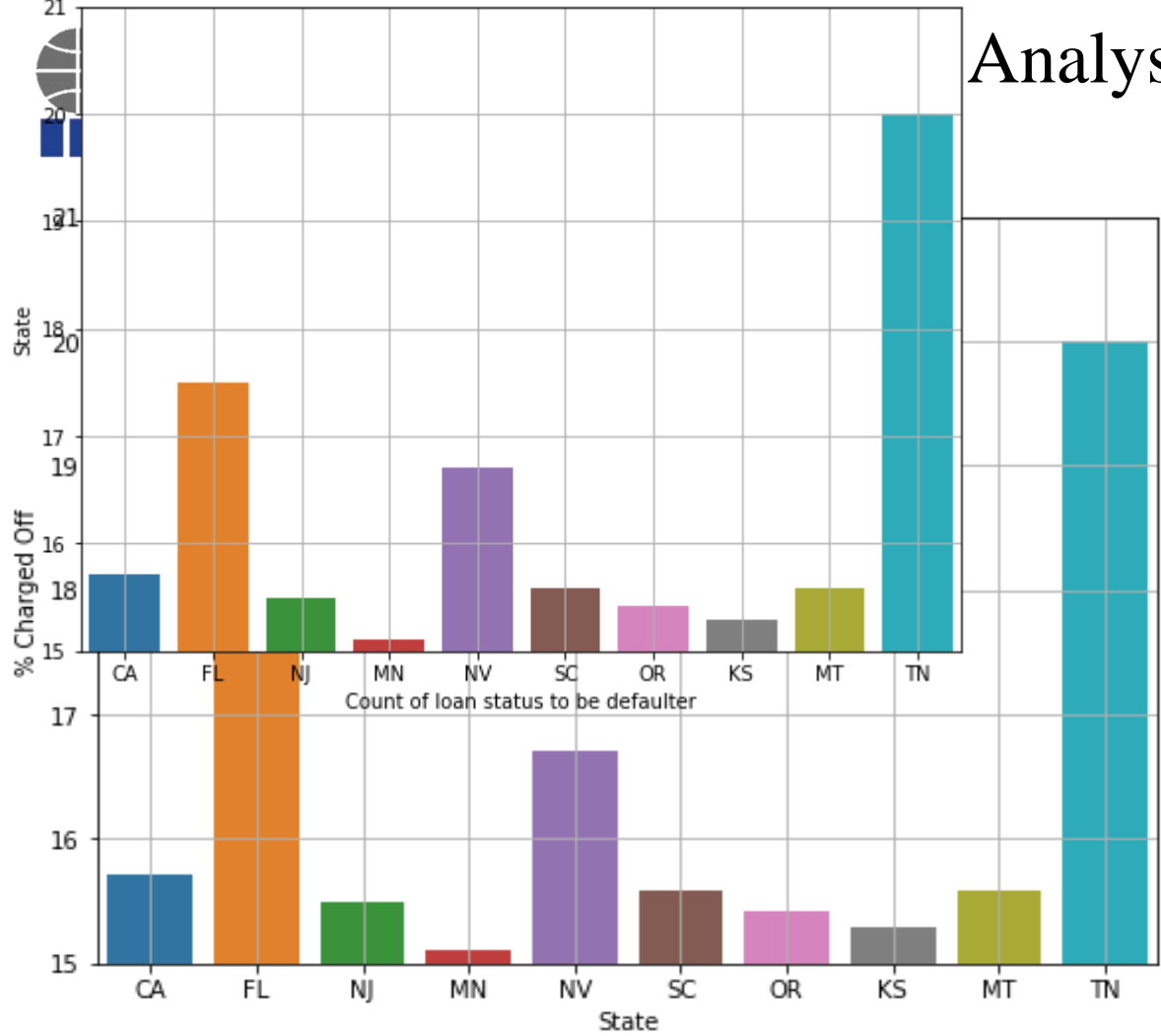
Observations:

1. Average Charged Off % is approx. between 13 to 14.
2. States : **'CA', 'FL', 'NJ', 'MN', 'NV', 'SC', 'OR', 'KS', 'TN'** have more than 15% Charged Off Loans. Economy condition may not be good in these states. More Scrutiny is needed here by Lending Club.
3. Approx. 8% difference between 0 and 1 bankruptcies.
4. Number of bankruptcy has enough impact on Charged Off %.

| index	TN	FL	NV	CA	SC	MT	NJ	OR	KS	MN
addr_state	10.00	2,642.00	461.00	6,685.00	443.00	77.00	1,735.00	415.00	242.00	576.00
charge_off	2.00	462.00	77.00	1,051.00	69.00	12.00	269.00	64.00	37.00	87.00
percent	20.00	17.49	16.70	15.72	15.58	15.58	15.50	15.42	15.29	15.10

DTI and Open credit lines Analysis
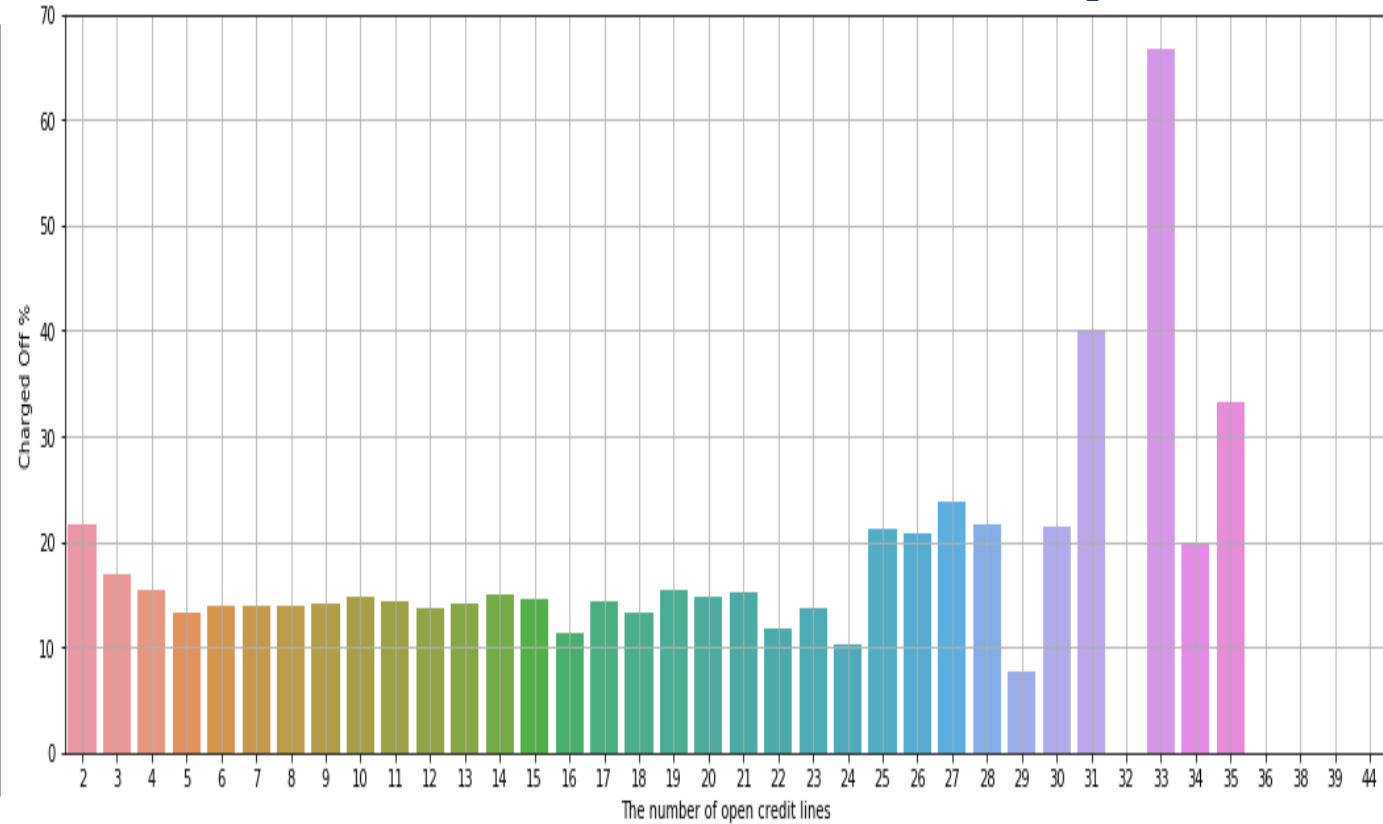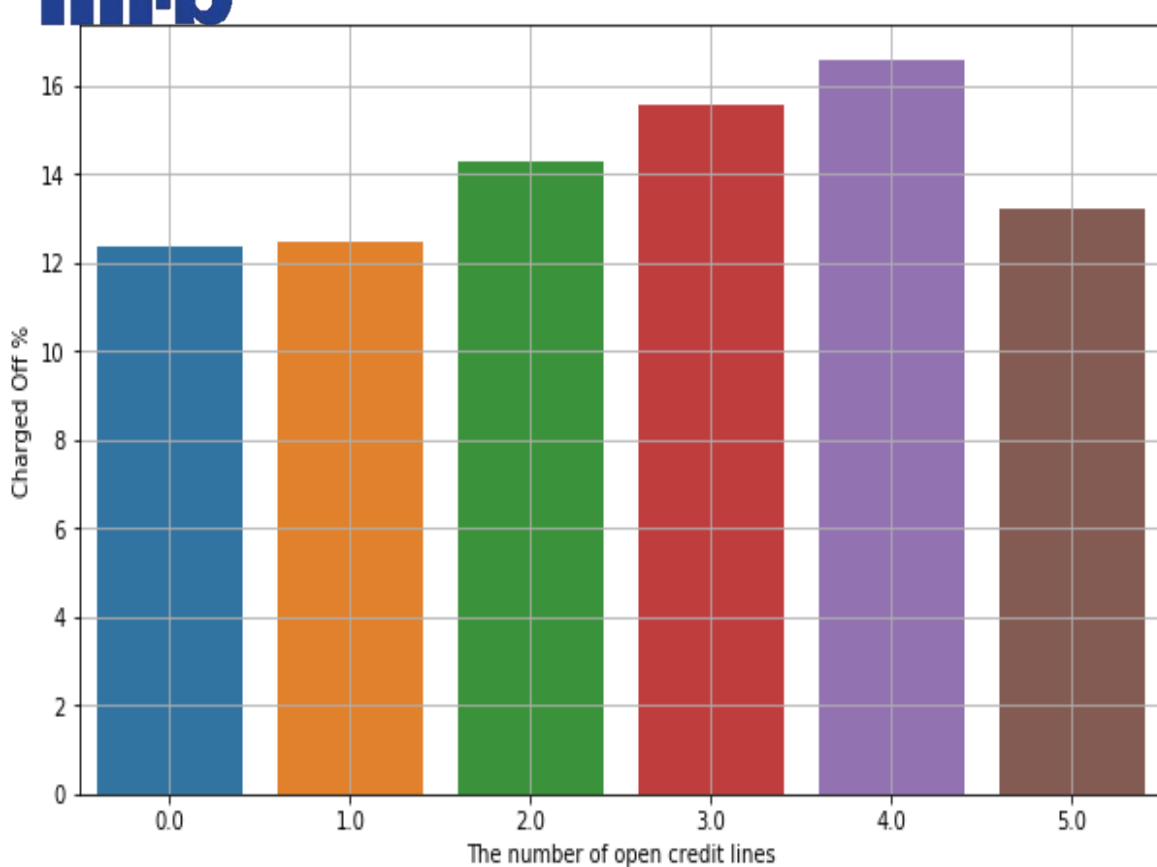
Observations:

1. % Charged Off is highest for DTI range of 15-19 and 20-24 (i.e., DTI Category = 3 & 4) The Customers with the given DTI ranges need more scrutiny from Lending Club

2. Customers having Credit Lines >= 25 have higher Charged Off % thus need more scrutiny.

Funded Amount vs Annual Income Analysis

Observation:

If Annual income is between:
1. 4k to 38 k : Loan amount greater than 8k has higher risk of "Charged-Off"
2. 38 k to 51 k , Loan amount greater than 12k has higher risk of "Charged-Off"
3. 51 k to 66 k, Loan amount greater than 13k has higher risk of "Charged-Off"
4. 66k to 90k, Loan amount greater than 16k has higher risk of "Charged-Off"
5. >90k , Loan amount greater than 20k has higher risk of "Charged-Off"

Summary
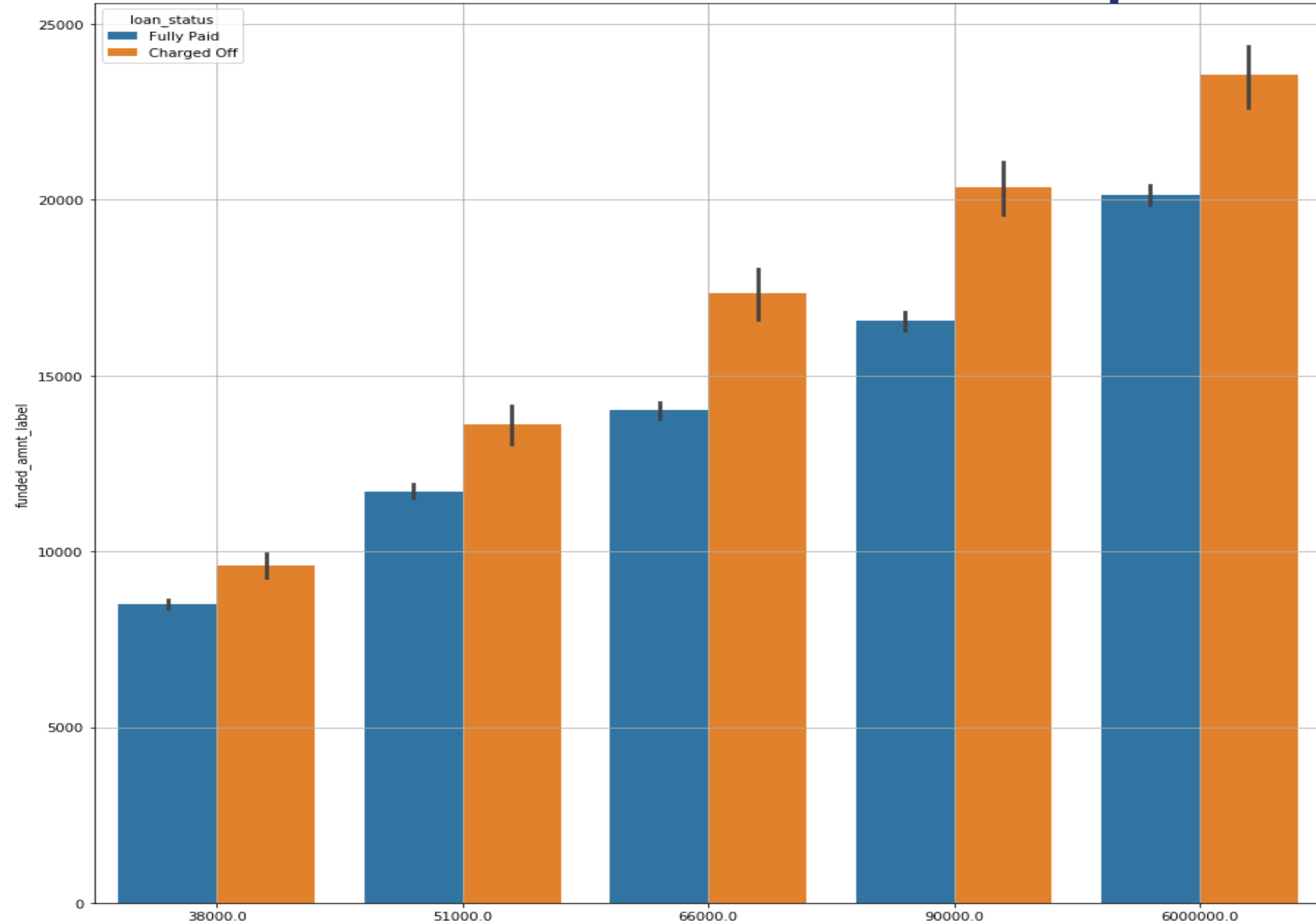
The overall Analysis based on the requirements, below is the high level conclusion :

1. Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.

2. When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.

3. Customers need more scrutiny if :
   - **DTI Range is between 15-24**
   - **Credit Lines>25**
   - **Customers belong to State TN**
   - **Grade is E, F, G.**

4. Follow below matrix for annual income and funded amount to identify defaults :

Annual Income Range	Loan Funded Amount (Greater than)	Charged Off Risk
4k-38k	8k	High
38k – 51k	12k	High
51k-66k	13k	High
66k-90k	16k	High
>90k	20k	High