

# Lending Club Case Study

Project Group:

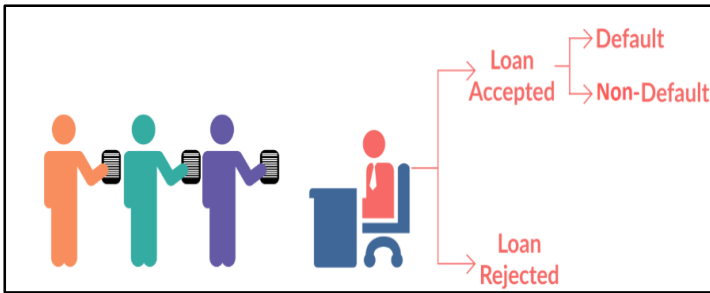
PREMAL MATALIA & GIJO PETER

Date: 03-Feb-2020



## Business Understanding:

- ❖ Lending Club is a **peer-to-peer** lending company that utilizes a group of **private investors** to fund loan requests.
- ❖ It has to make a decision for loan application approval based on the applicant's profile and does **risk assessment** by assigning borrower grades and subgrades based on their credit history
- ❖ It is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.



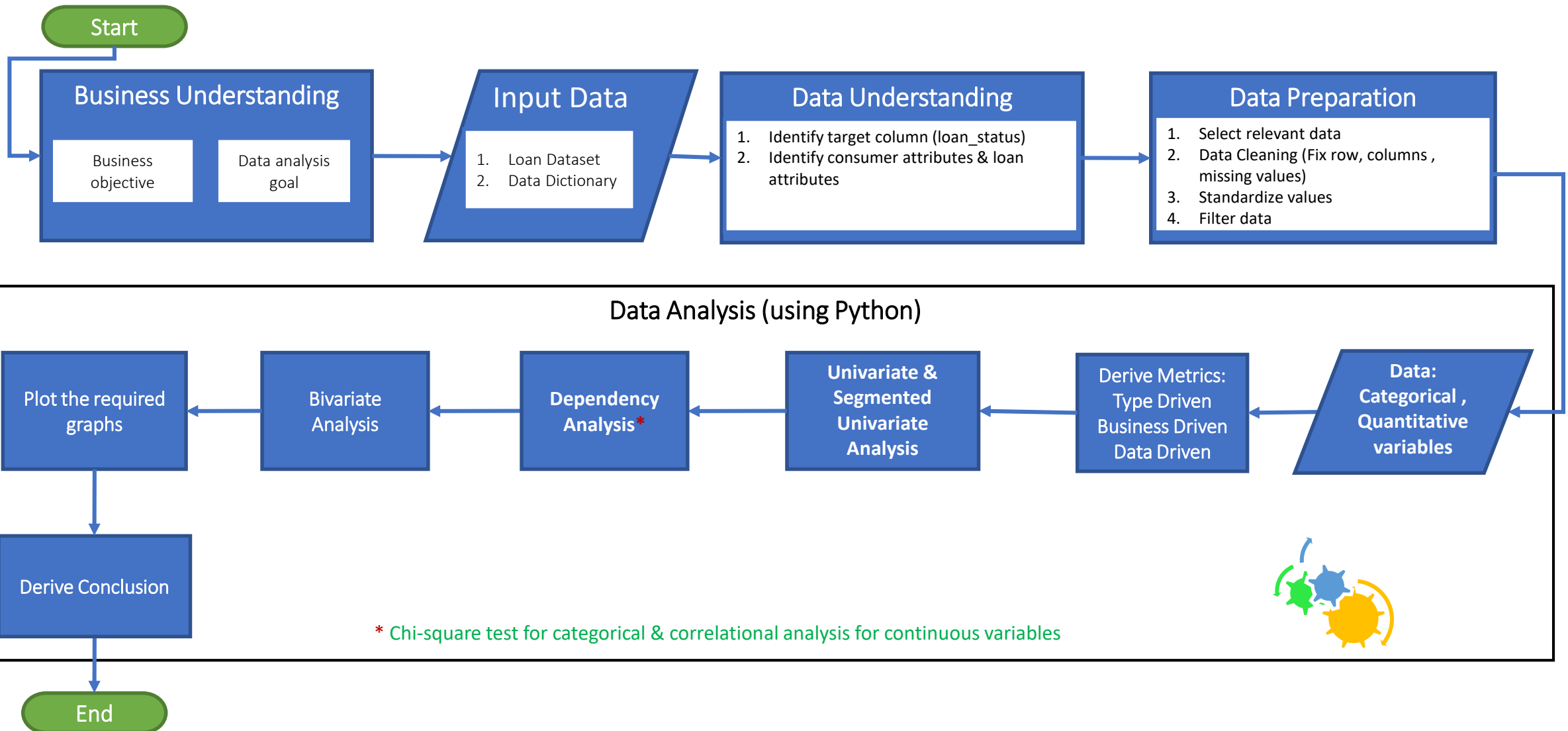
## Types of risks:

- ❖ If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- ❖ If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

## Problem Statement:

Lending club wants to understand the **driving factors (or driver variables)** behind loan default and utilize this knowledge for its portfolio and risk assessment. It can identify these risky loan applicants, then such loans can be reduced thereby **cutting down the amount of credit loss**





- ❖ Lending Club has provided complete loan data for all loans issued through the time period 2007 to 2011. This dataset contained information related to borrower's credit information along with loan information.
- ❖ *Dataset has 39717 records with 111 columns. Column description has been provided in Data Dictionary excel file.*
- ❖ *We identified **consumer attributes** and **loan attributes** which can influence the tendency of default using **Exploratory Data Analysis (EDA)***
- ❖ ***Consumer behavioural** attributes are not available at the time of loan sanctioning hence excluded from EDA*

Loan attributes	Customer Attributes	Customer behaviour attributes
Grade (grade)	Annual income (annual_inc)	Revolving balance(revol_bal)
Duration(term)	DTI	Revolving utility ratio(revol_util)
Interest %(int_rate)	Employment length(emp_length)	Incidences of delinquency(delinq_2yrs)
instalment	Address(addr_state)	Last payment date(last_pymnt_d)
Funded amount(funded_amnt)	Designation(emp_title)	
Purpose	Number of public record bankruptcies(pub_rec_bankruptcies)	

id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade
1077501	1296599	5000	5000	4975	36 months	10.65%	162.87	B	B2
1077430	1314167	2500	2500	2500	60 months	15.27%	59.83	C	C4
1077175	1313524	2400	2400	2400	36 months	15.96%	84.33	C	C5
1076863	1277178	10000	10000	10000	36 months	13.49%	339.31	C	C1
1075358	1311748	3000	3000	3000	60 months	12.69%	67.79	B	B5
1075269	1311441	5000	5000	5000	36 months	7.90%	156.46	A	A4
1069639	1304742	7000	7000	7000	60 months	15.96%	170.08	C	C5
1072053	1288686	3000	3000	3000	36 months	18.64%	109.43	E	E1
1071795	1306957	5600	5600	5600	60 months	21.28%	152.39	F	F2
1071570	1306721	5375	5375	5350	60 months	12.69%	121.45	B	B5
1070078	1305201	6500	6500	6500	60 months	14.65%	153.45	C	C3
1069908	1305008	12000	12000	12000	36 months	12.69%	402.54	B	B5
1064687	1298717	9000	9000	9000	36 months	13.49%	305.38	C	C1
1069866	1304956	3000	3000	3000	36 months	9.91%	96.68	B	B1
1069057	1303503	10000	10000	10000	36 months	10.65%	325.74	B	B2
1069759	1304871	1000	1000	1000	36 months	16.29%	35.31	D	D1

## Data cleaning:

- ❖ *Total 57 columns were having either all or majority of records were having null values so dropped them*
- ❖ *Dropped columns having no relation to the analysis and target columns*
- ❖ *Derived new columns from existing columns to facilitate further data analysis*
- ❖ *Binned continuous variable to ranges for easy interpretation and analysis*

## Loan Status:

❖ 14.6% of the total loans were defaulted in given dataset of Lending Club and remaining loans were fully paid or in current status.

## Loan Grade:

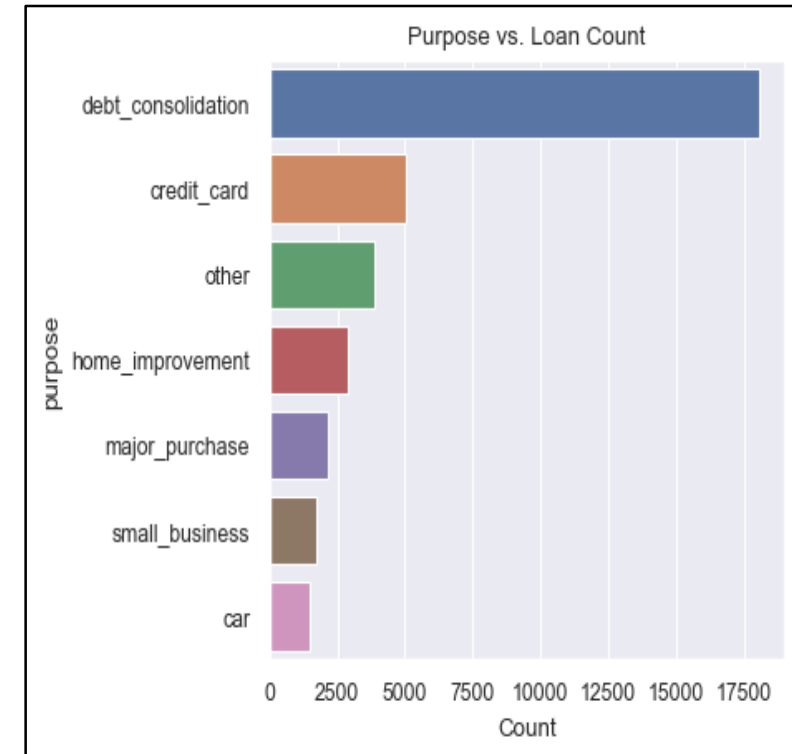
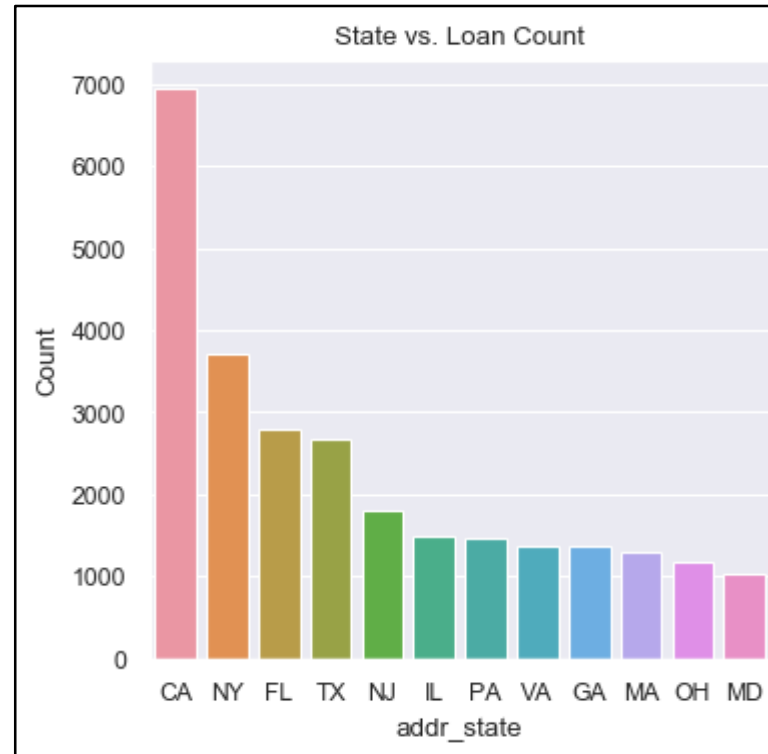
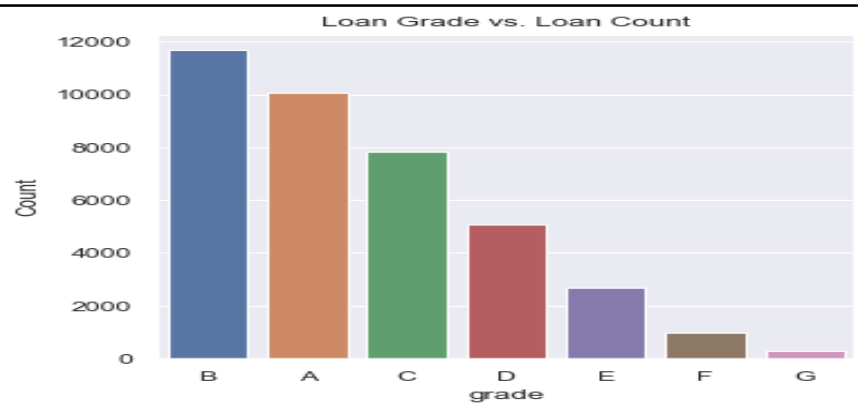
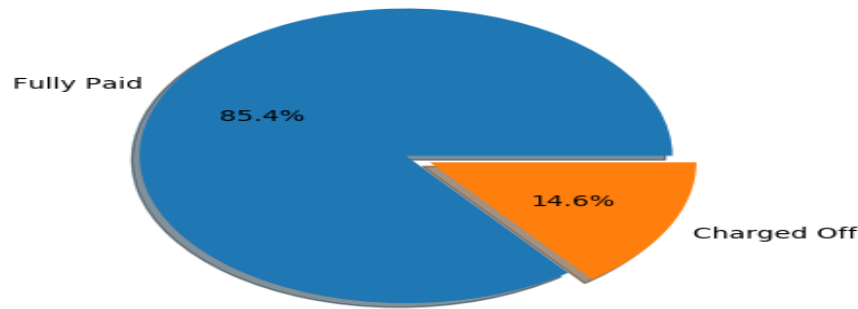
❖ Majority of loans approved by Lending Club seems to be of Grade **A to D** which are low risk and low interest rate.

## State:

❖ Most populated US States like California (**CA**), New York (**NY**), Florida (**FL**), Texas (**TX**) were having higher loan applicants than other states

## Purpose:

❖ Majority of loans were for **debt consolidation** purpose so that applicant can pay off his other liabilities and consumer debts, generally unsecured ones



## Annual Income:

- ❖ Maximum income value = 6,000,000 which is ~94 times of mean income value = 63,793 so we removed outliers values > 99% percentile. After removing outliers, majority of applicants are having annual income between 20,000 to 150,000 USD with maximum applicants are having at 60,000 or 50,000 USD

## Interest Rate:

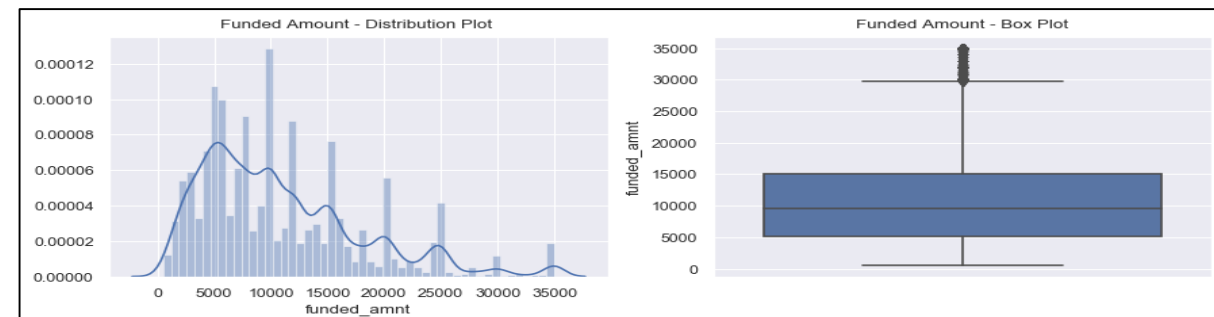
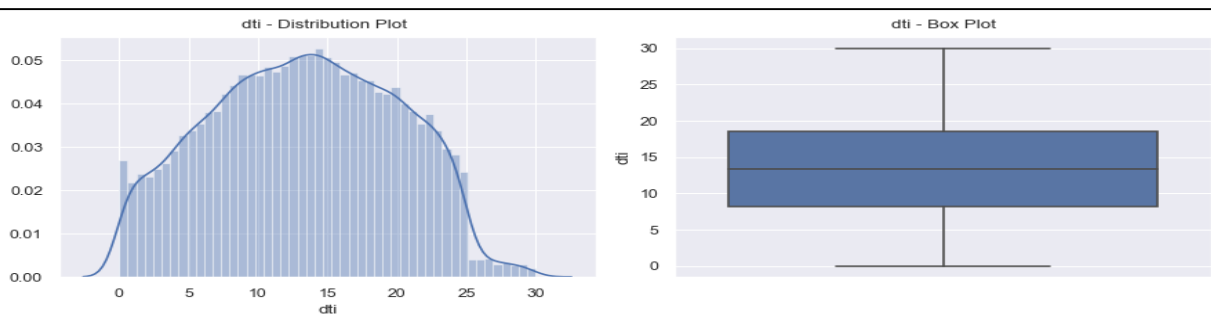
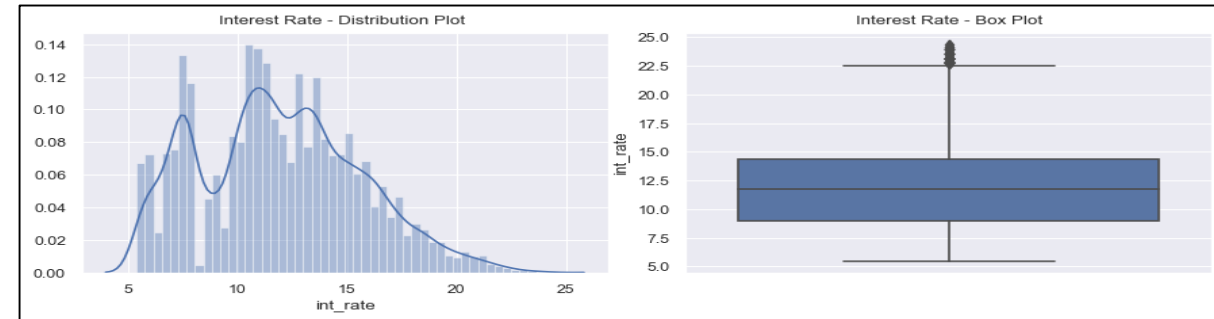
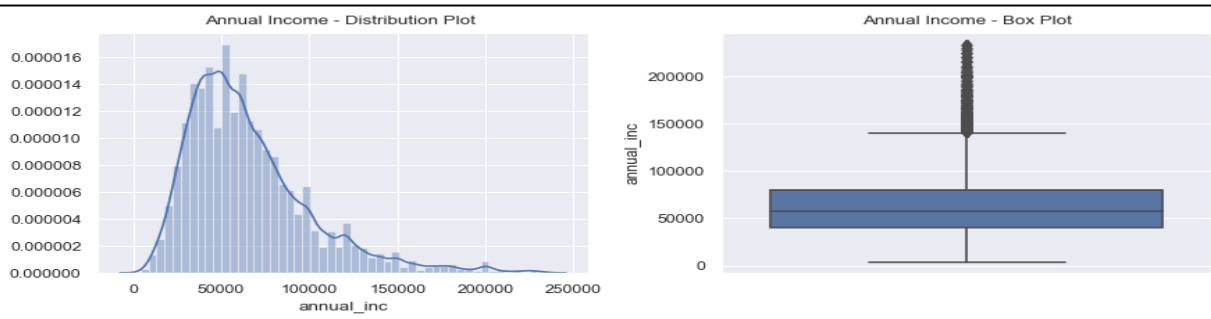
- ❖ Majority of applicants received loans at the interest rate between 5% to 15% which are low to medium risk loans

## dti:

- ❖ Majority of applicants have dti between 0 to 25%
- ❖ Lending Club seems to be approving very few loans of applicants having dti between 26 to 30 and **not** approving any loans to the applicants having **dti > 30%**

## Funded Amount:

- ❖ Lending Club provides minimum 500 USD and maximum 35,000 USD of loan and majority of applicants received loans between 5000 to 20000 USD



In this section, we have done loan default ratio/probability analysis for various segments of a categorical variable.

## Loan Grade/Interest Rate:

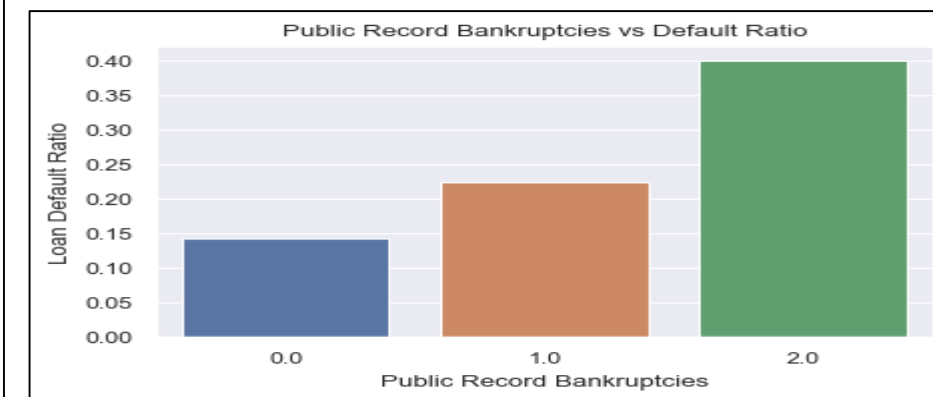
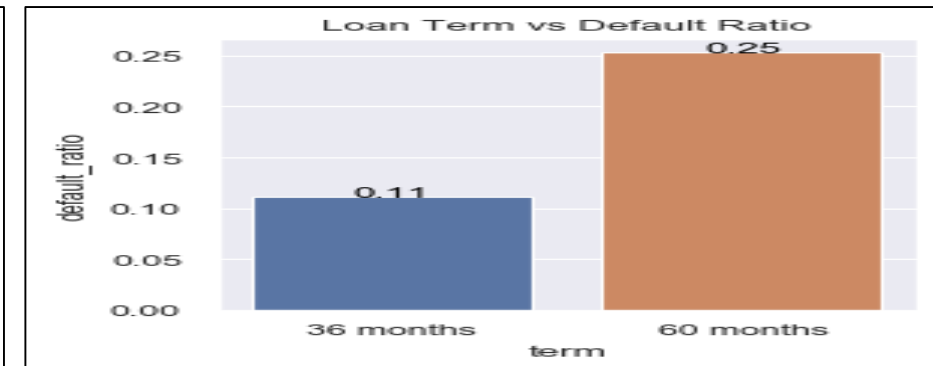
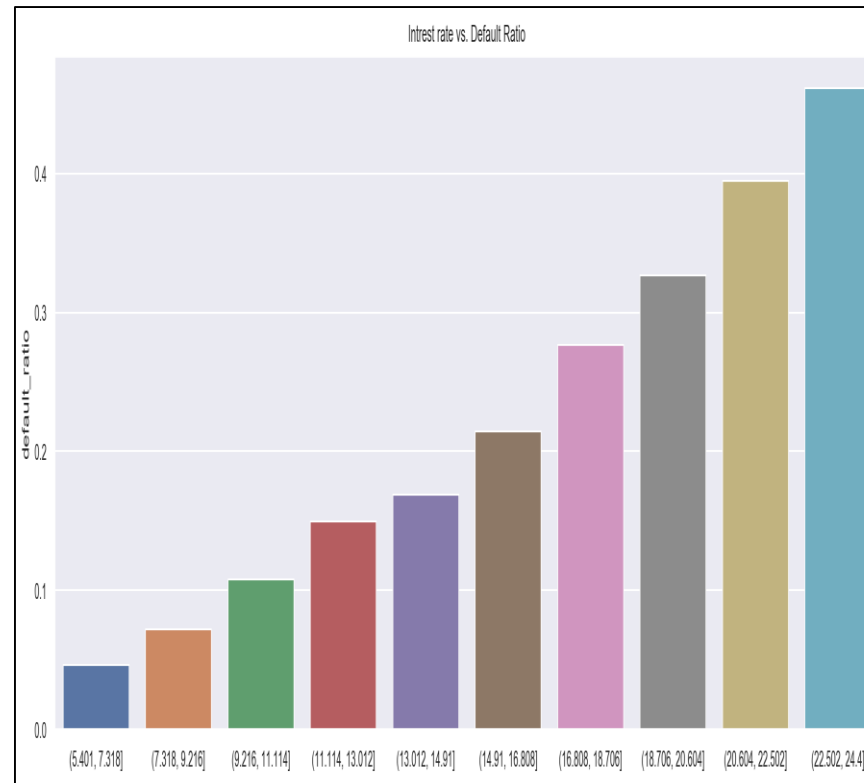
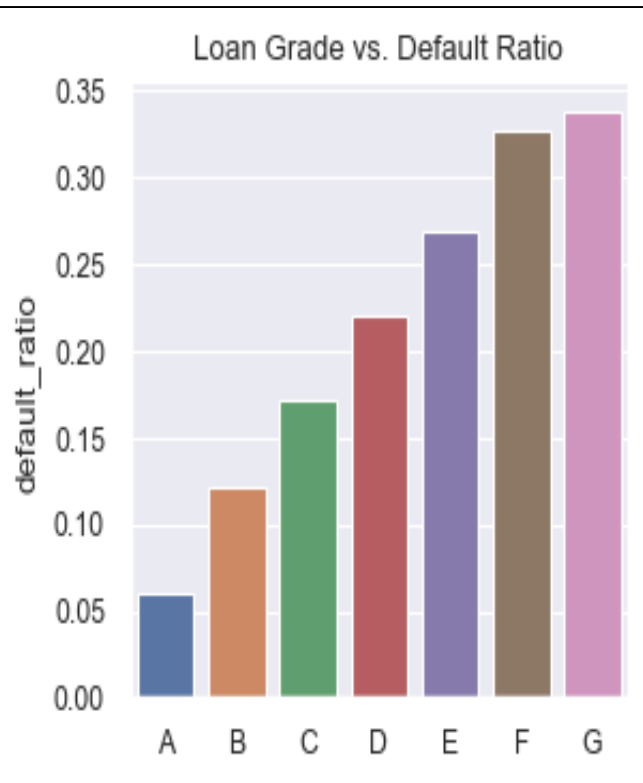
❖ *loans with lower grades (i.e. D, E, F, G) and higher interest rates are more likely to default. With loan grade D to G, it has been observed > 20% default ratio for that category*

## Loan Term:

❖ *Loan with longer term duration (60 months) are tend to 'default' more as compared to term of 36 months. It has been observed that 25% of total loans in 60 months category have been marked as Default.*

## Public Record Bankruptcies:

❖ *Loans of Consumers with more public record of bankruptcies are more likely to default*





## Annual Income:

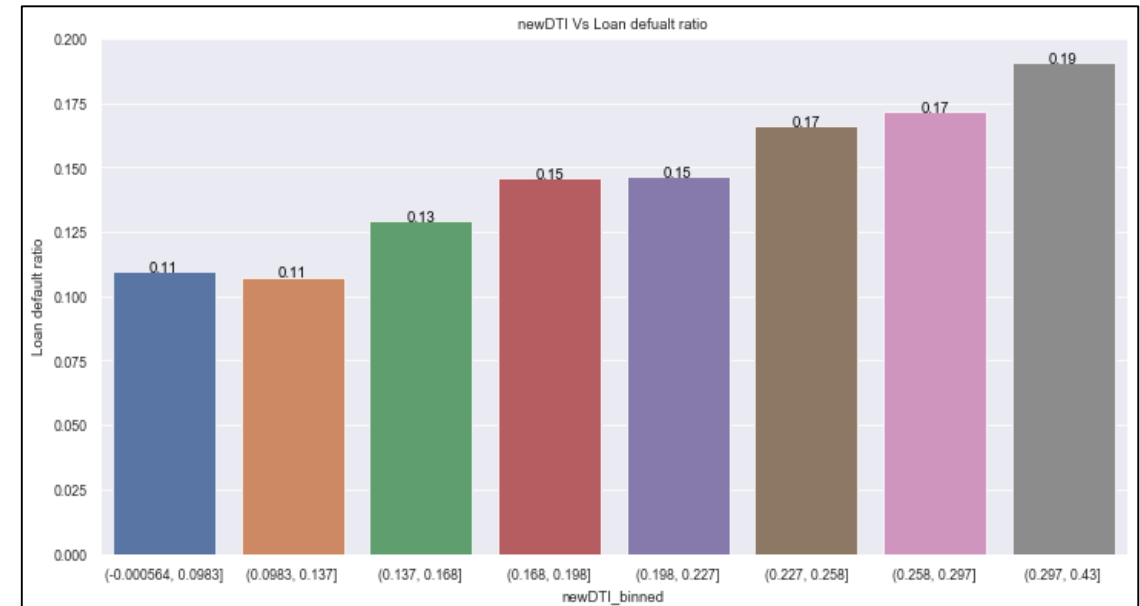
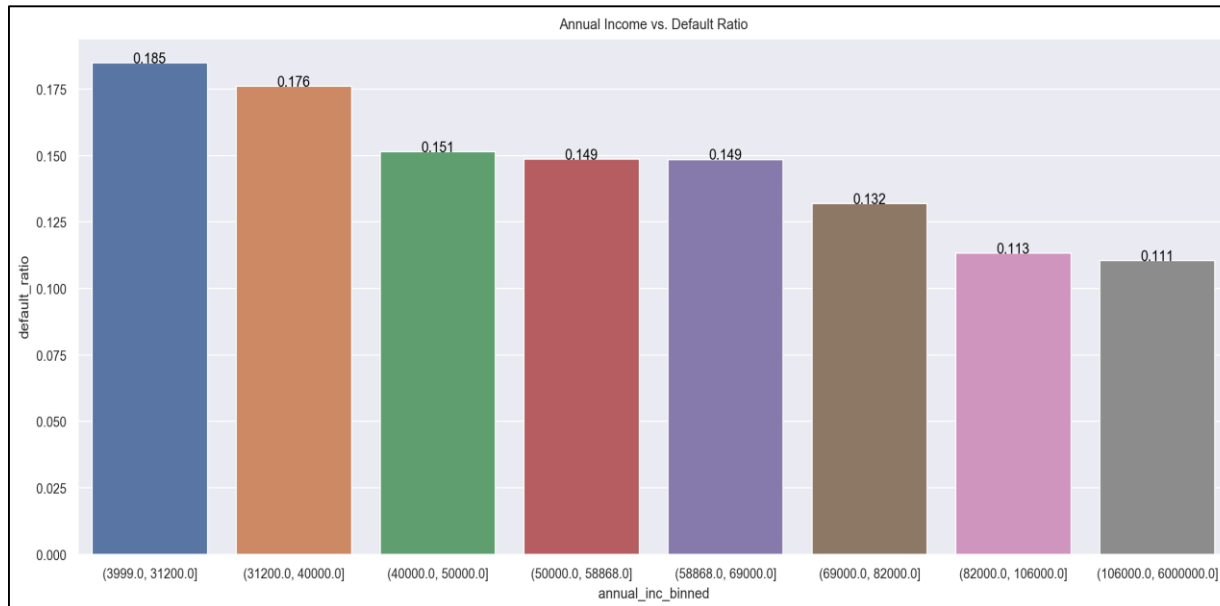
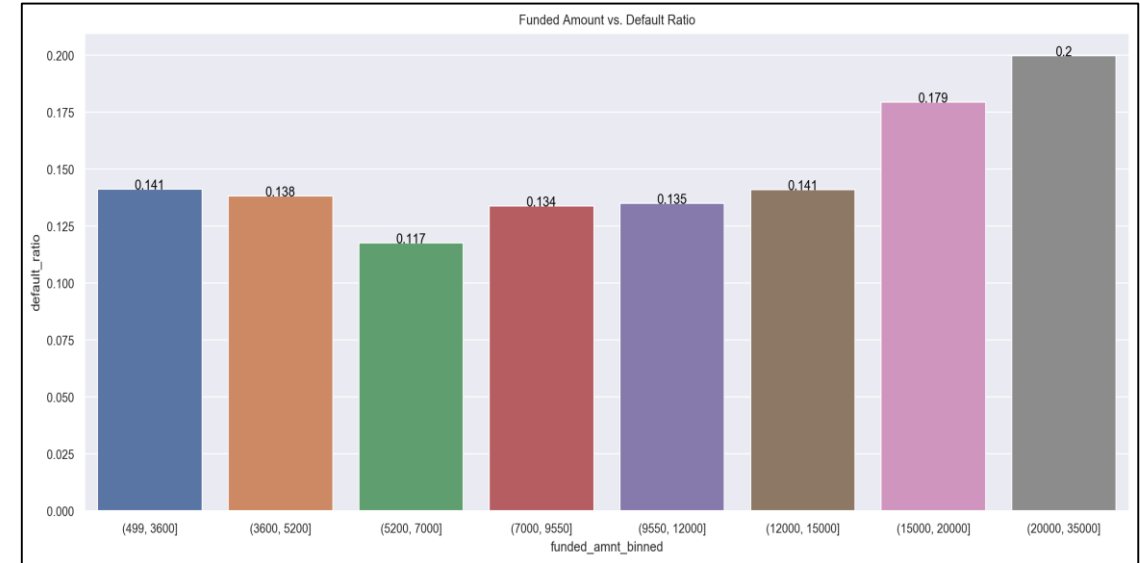
- ❖ Customer with lower income are more likely to default. Between 4000 to 31,000 USD, it has been observed 18.5% of default ratio

## Funded loan Amount:

- ❖ Higher the funded loan amount, higher the loan default rate. Between 20,000 to 35,000 USD, it is observed 20% of default ratio

## New DTI metric:

- ❖ Customers with DTI considering also current applied loan > 30 % are more likely to default hence should be avoided



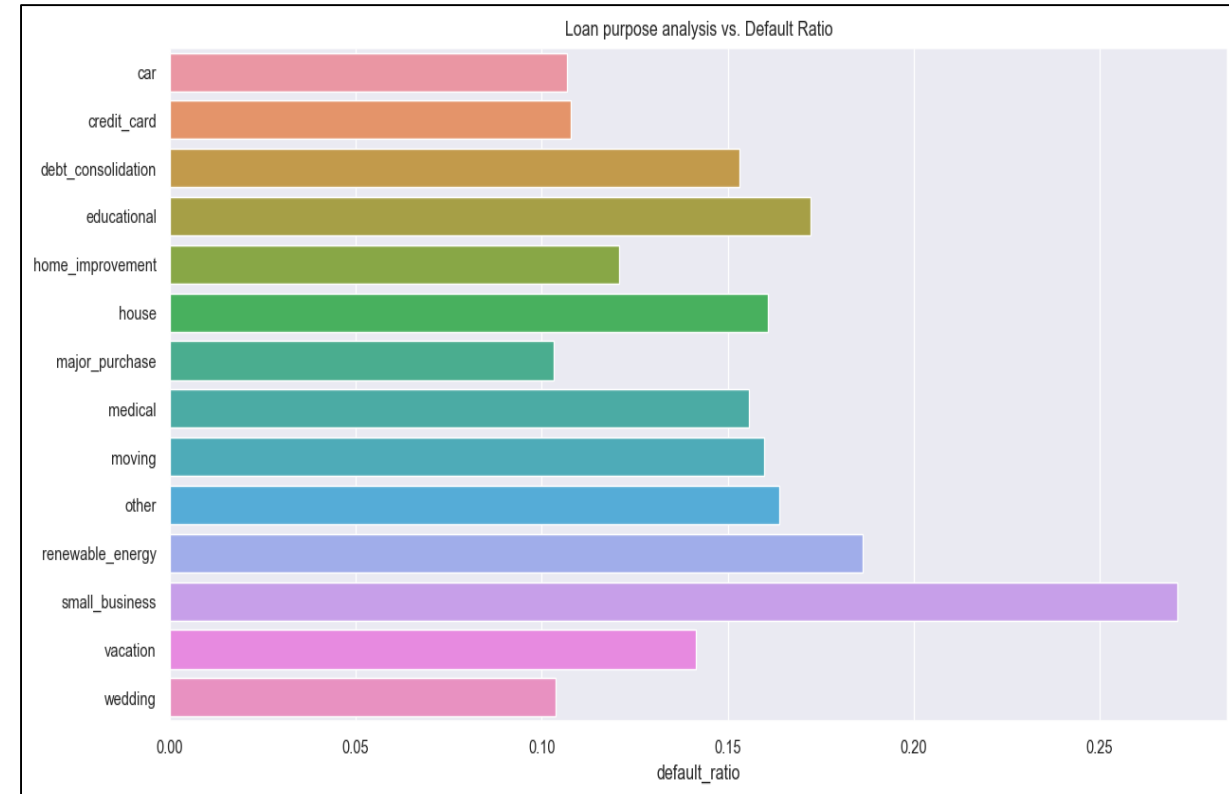
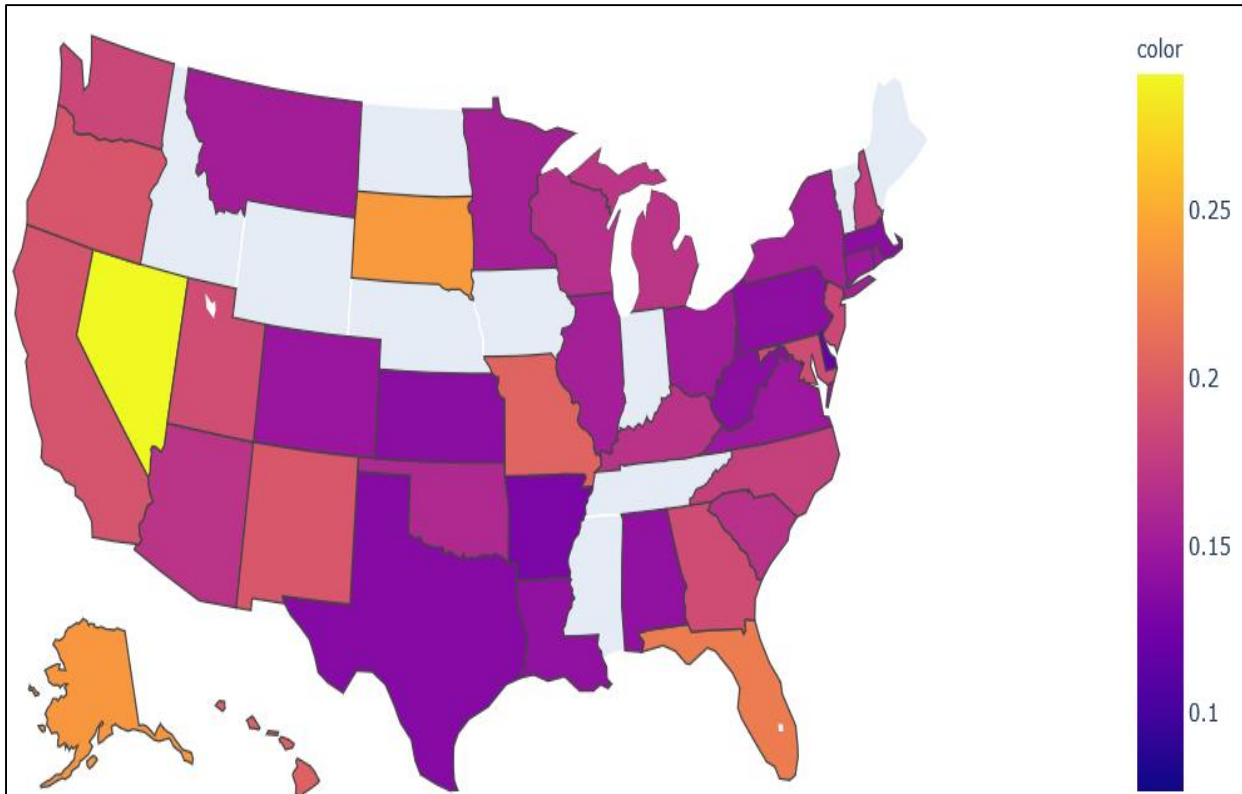


## State:

❖ Customers from Nevada(NV), South Dakota(SD) have more default ratio > 20%

## Loan Purpose:

❖ Customers who have taken loan for '**small business**' purpose are more likely to default with ratio > 25%



Dependent **categorical variables** are selected using **chi-square test** and dependant **continuous variables** are selected using **pearsons correlation** method.

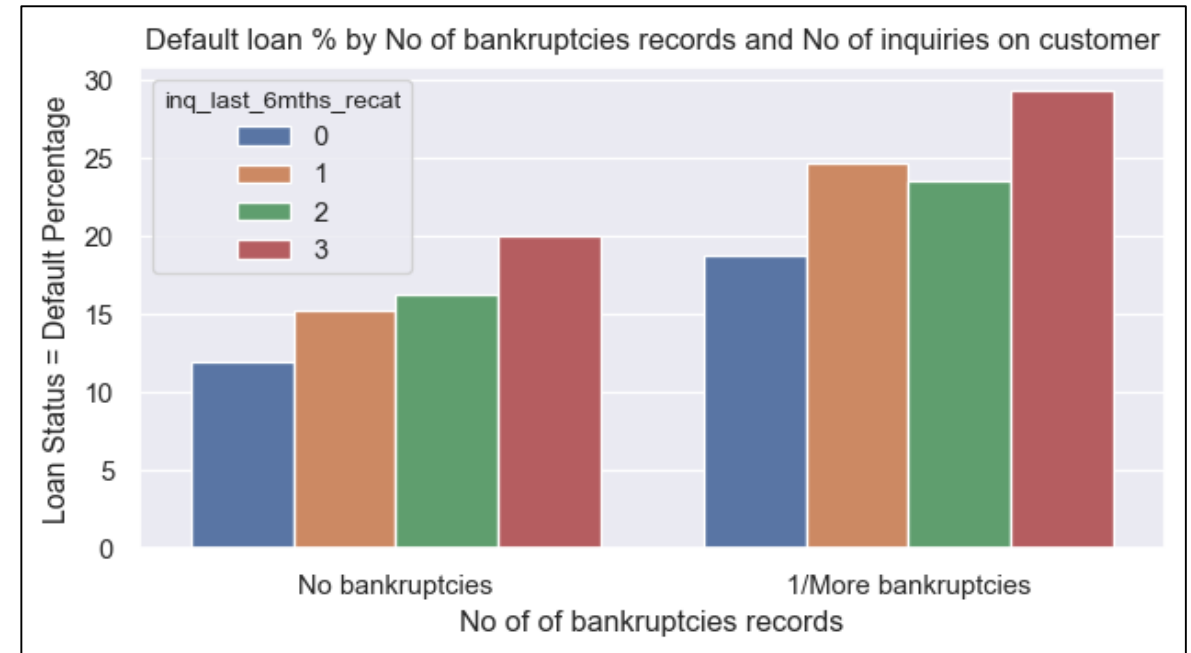
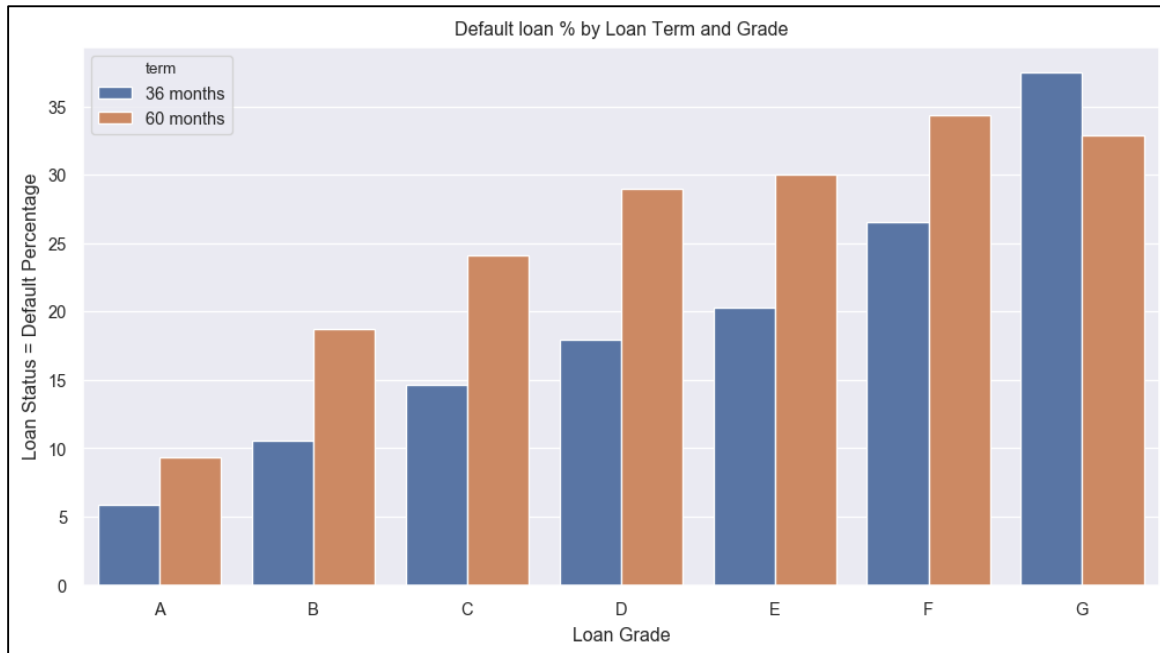
A bivariate analysis on loan default ratio is performed by identified pair of influential variables

## **Loan Grade – Loan Terms:**

*Loans with Grade A and B irrespective of loan terms and grade C and D with term of 36 months are likely to repay. Other lower grades loans with any loan terms are relatively risky.*

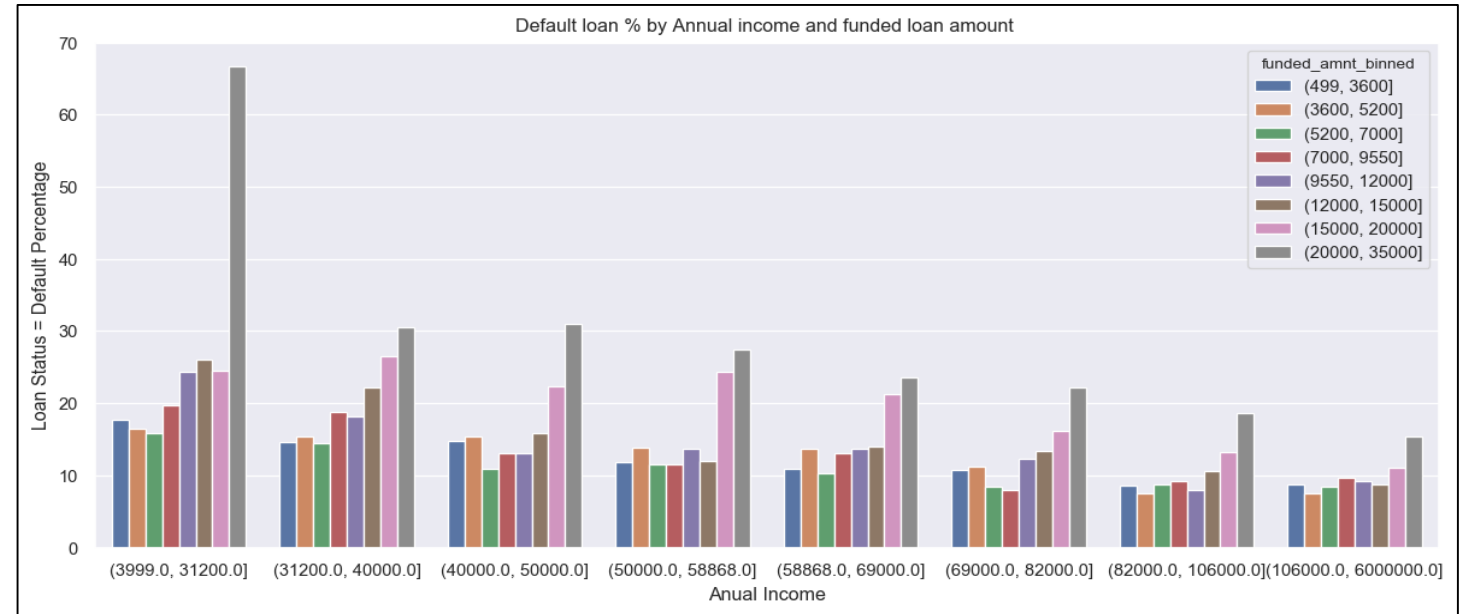
## **Public Record Bankruptcies – No of Inquiries in last 6 months:**

*Customers with Public bankruptcy record or Public derogatory records are more likely to default. In addition, customers with 3 or more inquiries in last 6 months are also risky*



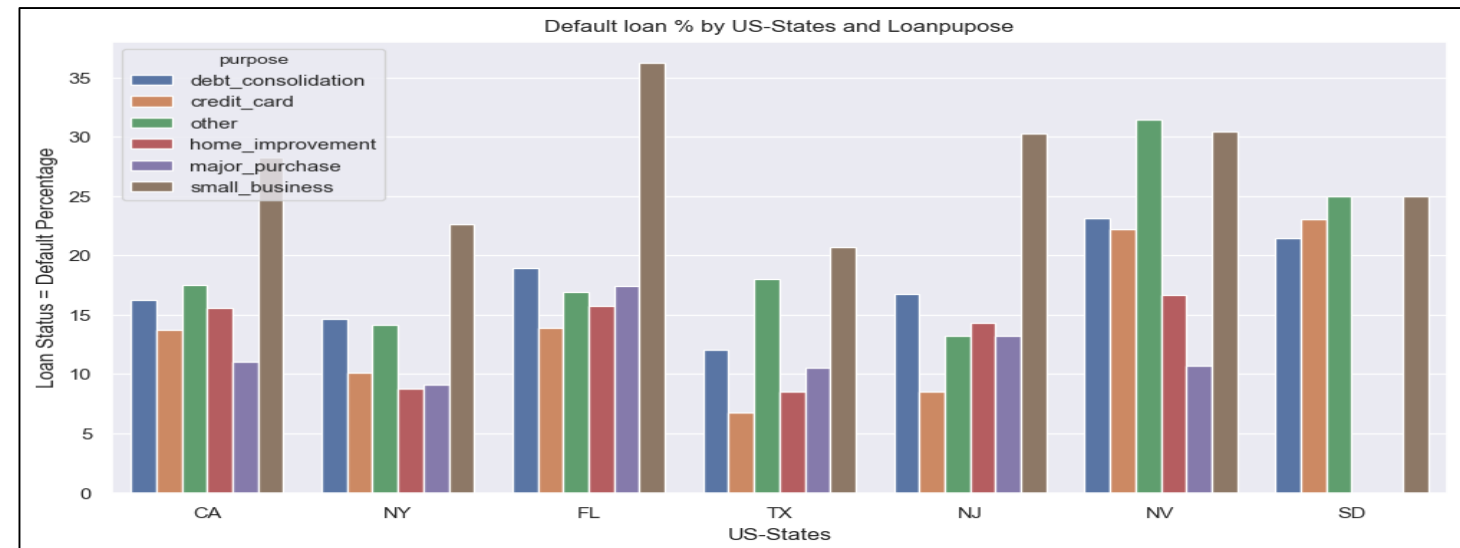
## Annual income – Funded loan amount:

Customer with lower income with higher loan amount is likely to default. Funded loan amount above 15,000 USD for customer with < 70,000 USD annual income is risky and should be avoided. Customers with more than 82,000 USD annual income are having more than 80% of probability of repayment.



## Public Record Bankruptcies – Inquiries in last 6 months:

Customers applying loan for 'small business' purpose are more likely to default with ratio > 25% so should be avoided. Same trend has been observed in states with higher loan count (California, New York, Florida..) and in states with higher default ratio (Nevada, South Dakota)



## Assumptions:

All the recommendation has been provided considering

- ❖ default ratio  $> 20\%$  is risky and approval may lead to financial losses. If Lending club wants to approve loans with higher default ratio, it should take extra caution like take some asset or guarantee
- ❖ *default ratio  $\leq 20\%$  is likely to repay the loan and rejection may lead to loss of business to the company*

## Recommendations:

1. Loans with **Grade A and B** irrespective of loan terms and grade **C and D** with term of 36 months are likely to repay. Other lower grades loans (**Grade E, F, G**) with any loan terms are relatively risky.
2. Funded loan amount above **15,000 USD** for customer with  **$< 70,000$  USD** annual income is risky and should be avoided. Customers with more than 10,000 USD annual income are having more than 85% of probability of repayment.
3. Customers applying loan for 'small business' purpose are more likely to default with ratio  $> 25\%$  so should be avoided while loan purpose of car, wedding, major purchase and credit are having 90% of probability of repayment.
4. Customers with **Public bankruptcy record** or **Public derogatory records** are more likely to default. In addition, customers with 3 or more **inquiries** in last 6 months are also risky
5. Customers with DTI considering also current applied loan  $> 30\%$  are more likely to default hence should be avoided.