

# Lending Club Case Study

Submitted by:-

- Amir Khan
- Shwetha Pujari

# Problem Statement



LENDING CLUB IS SEEING A SIGNIFICANT JUMP IN THE LOAN DEFAULT RATE AND THIS IS CAUSING BIG REVENUE LOSS TO THE COMPANY.



REDUCING NUMBER OF LOANS WOULD AFFECT THE GROWTH, HENCE THEY WANT TO UNDERSTAND THE DRIVER VARIABLES FOR DEFAULT LOANS TO STRENGTHEN THEIR RISK ASSESMENT.

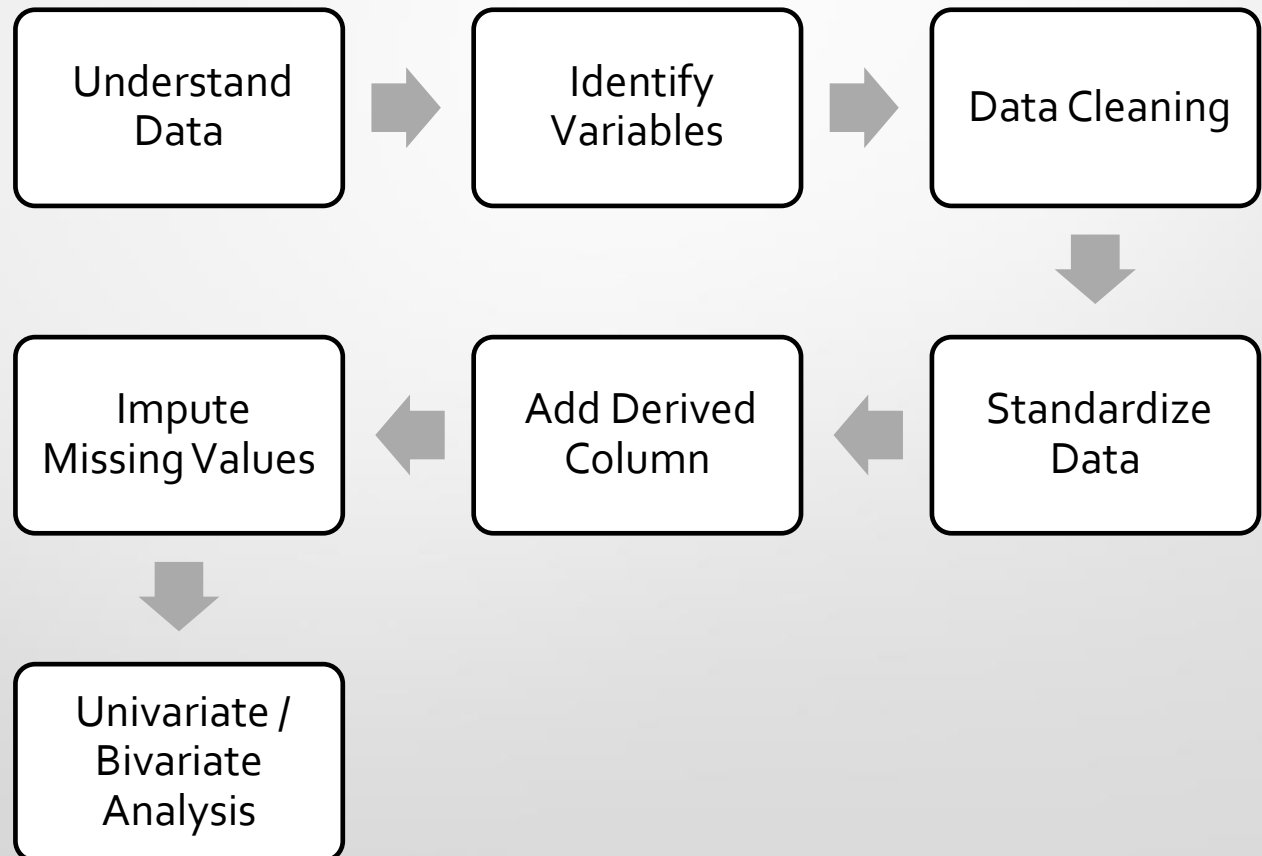


USE THE PROVIDED DATASET OF LOANS ISSUED IN LAST FEW YEARS TO IDENTIFY THESE DRIVER VARIABLES.

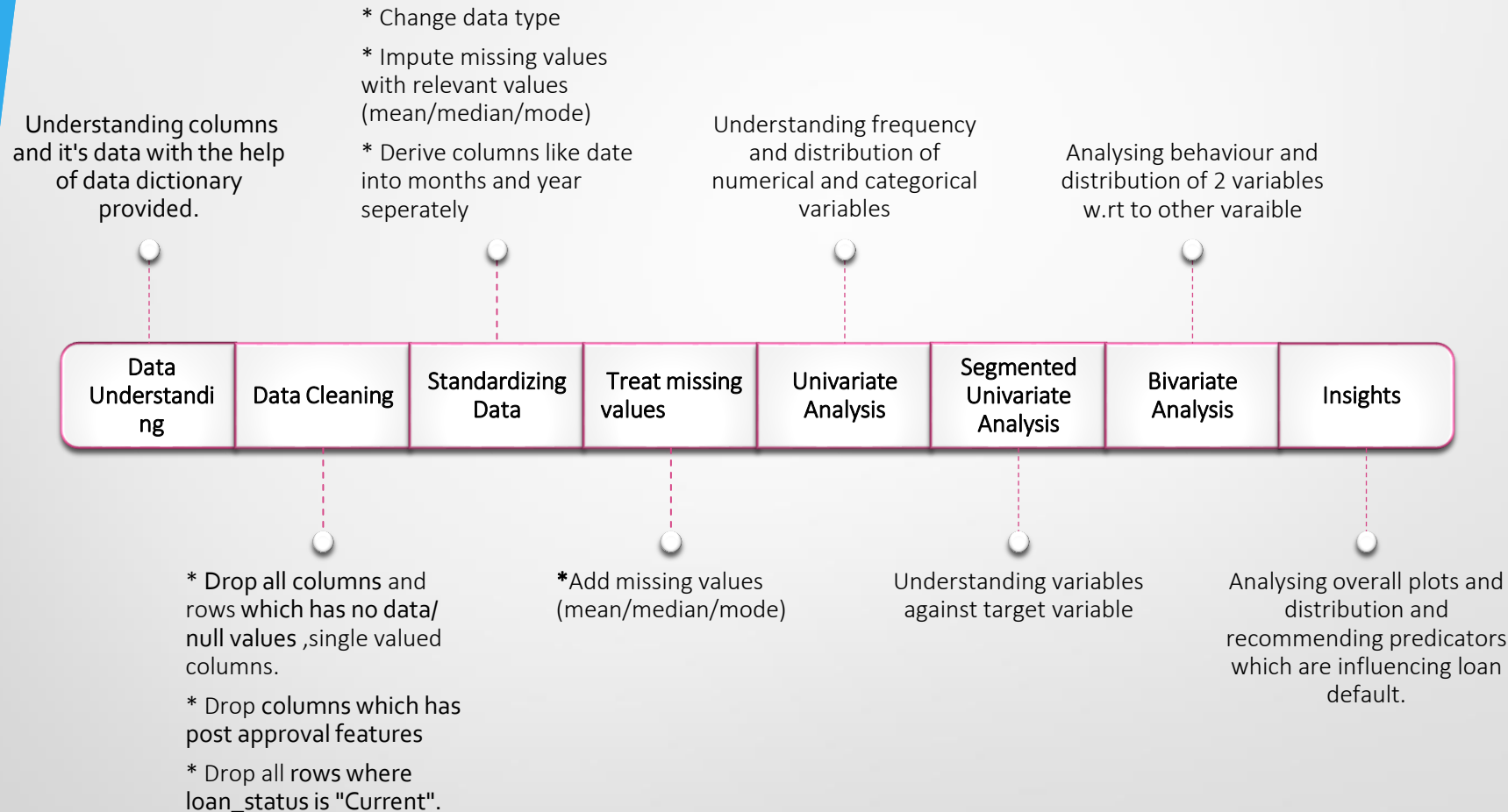


# Problem Solving Methodology

EDA - Exploratory Data Analysis



# Analysis Approach



# Understanding Data

```
In [5]: 1 ## Looking at the dataset to understand about it
        2 bank_data.head()
```

Out[5]:

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_12m
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	NaN
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	NaN
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	NaN
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	NaN
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	NaN

11 rows x 111 columns

## Data Source

We used Lending Club's data for this analysis. The data set is for the period from 2007 to 2011. There are approximately ~~39717~~ **39717** loan applicant's details with **111 variables**.

It is often very hard to run analysis with too many variables, hence broke down the dataset in three categories

- Loan variables that are available at the time of loan application
- Consumer demographic variables
- Loan variables that are available after the loan is processed

# Identify variables

## Hypothesis to select variables

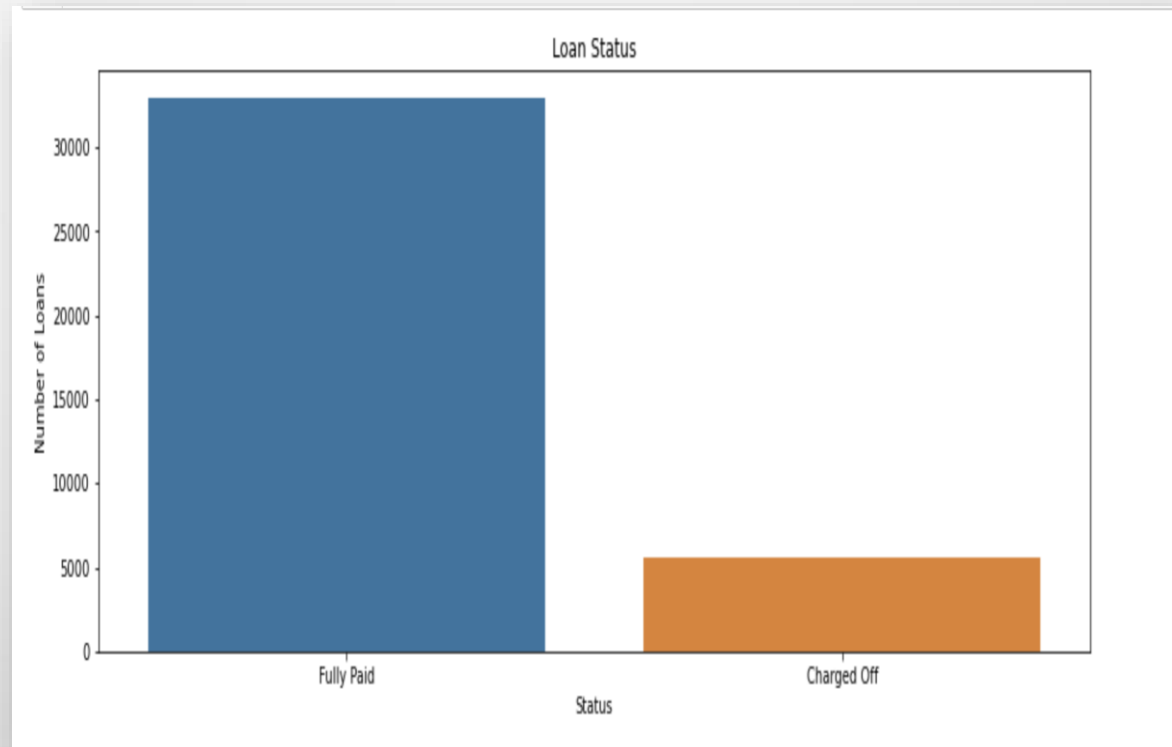
- Since the problem statement is to identify the variables that can help reduce default rate, we should focus on variables that are available at the time of loan application
- We have divided the variables in two categories
  - Loan Characteristics ( loan amount, int rate etc)
  - Customer behaviour characteristics (state, emp\_length etc)

Loan Characterisitics	Customer Behaviour
• loan_amnt	• emp_length
• term	• home_ownership
• int_rate	• annual_inc
• installment	• addr_state
• grade	• dti
• sub_grade	• emp_title
• purpose	
• pub_rec_bankruptcies	

# Loan status analysis

## Observations

- Loan Default rate is 14.5%, i.e out of every 100 loans issued 14 are defaulting
- Default loans has costed the company net loss of **~426 million** dollars which is equal to 16% of total loan amount issued

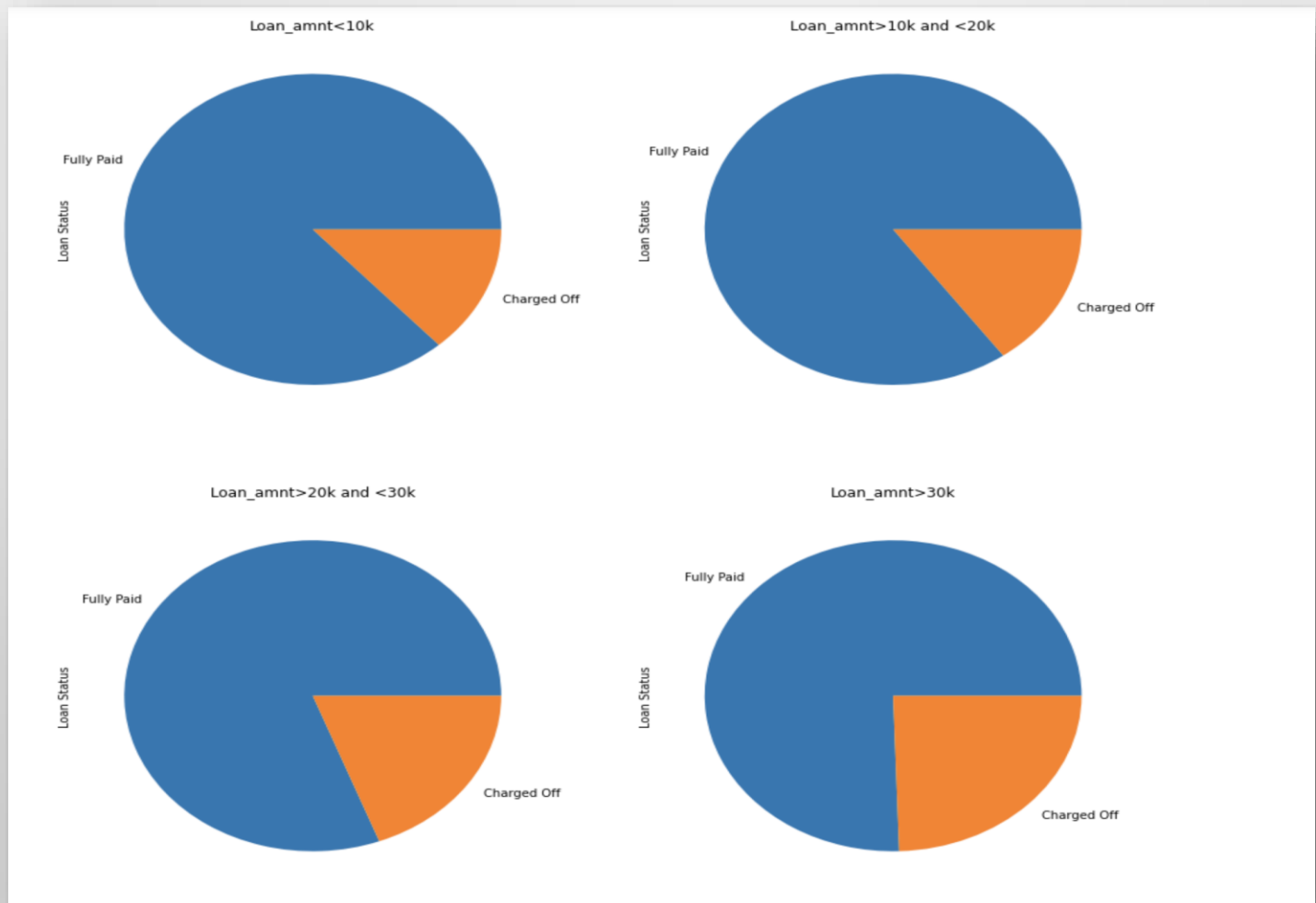


# Loan Amount Analysis

Analysis of Loan amount w.r.t distribution of Loan Status

## Observations

- We can see that loan default rate increases as the amount increases
- For loan amount less than 20K, loan default rate is around ~15%
- As the loan amount increases above 20K the default rate increases till 25%



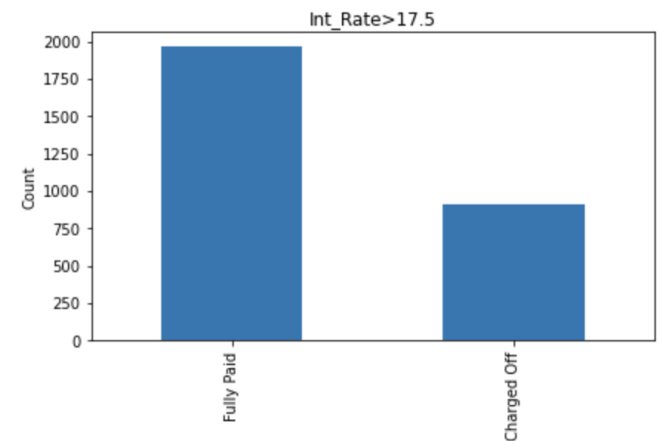
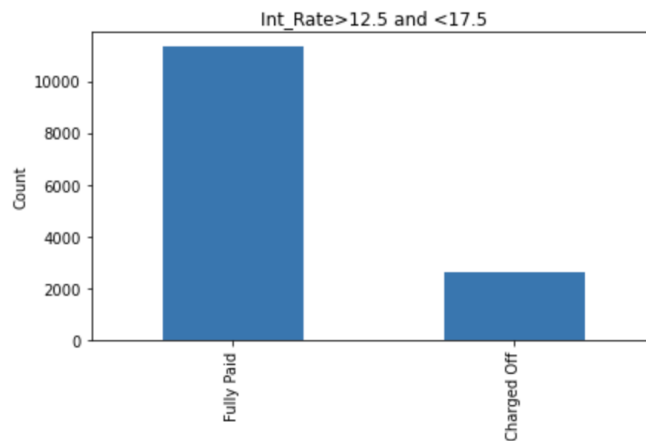
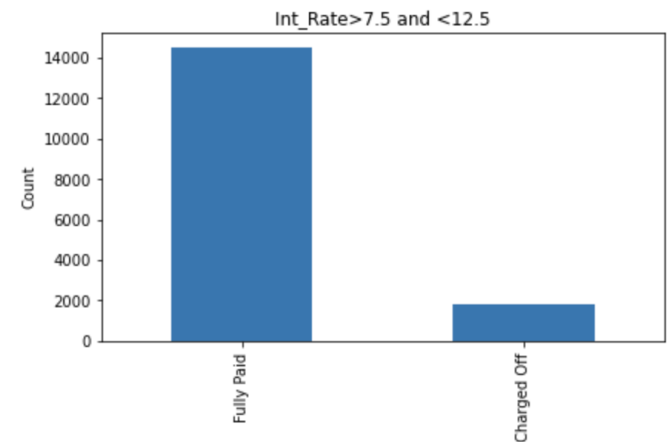
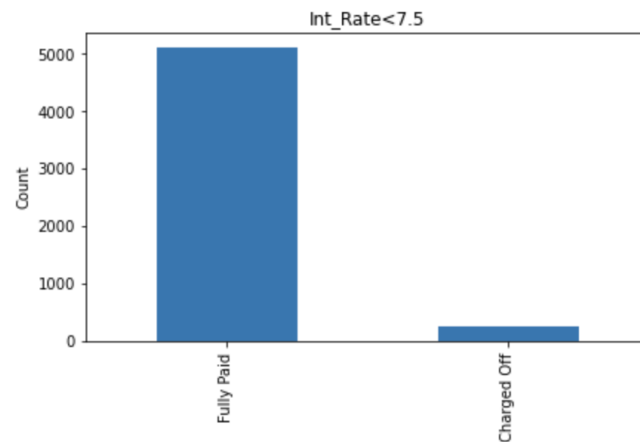


# Interest Rate Analysis

## Analysis of Interest rate w.r.t distribution of Loan Status

### Observations

- We can see that loan default rate increases as the interest rate increases
- For interest rate less than 7.5% the default rate is just 5%
- As the interest rate increases above 12.5% the default rate increases and goes till 31%

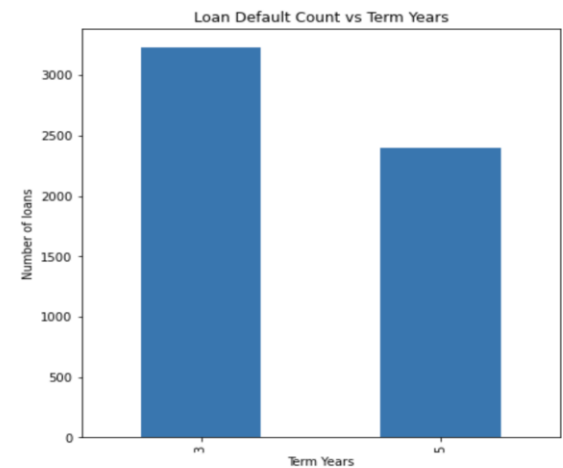
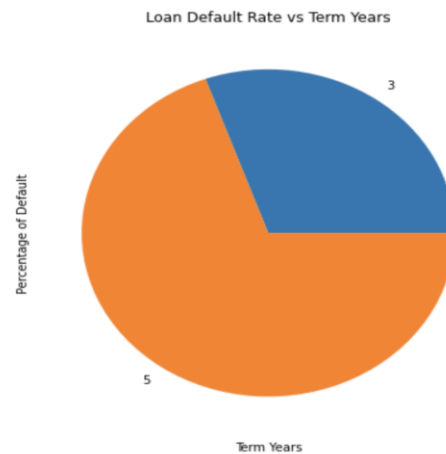
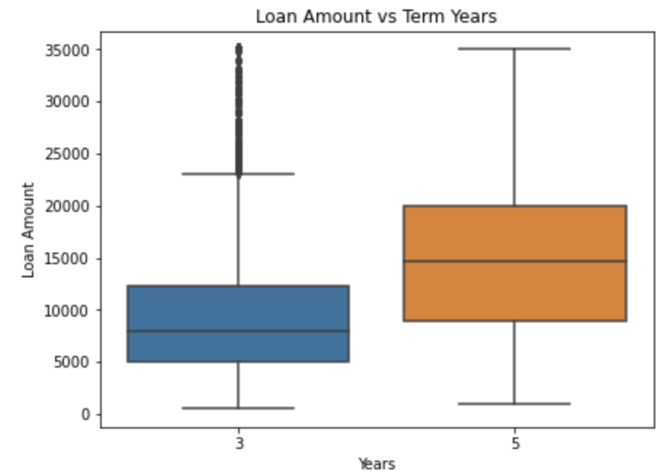
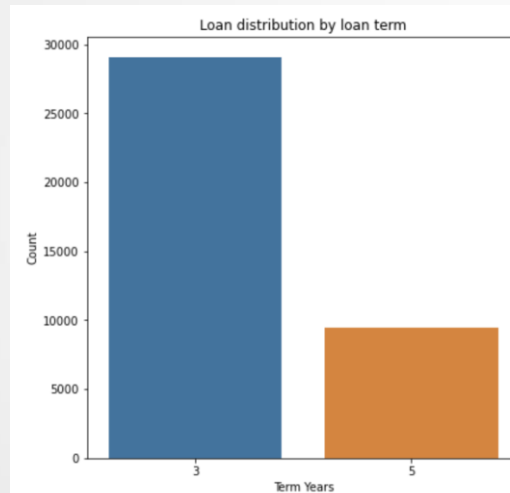


# Term Years Analysis

## Analysis of Term Years w.r.t distribution of Loan Status

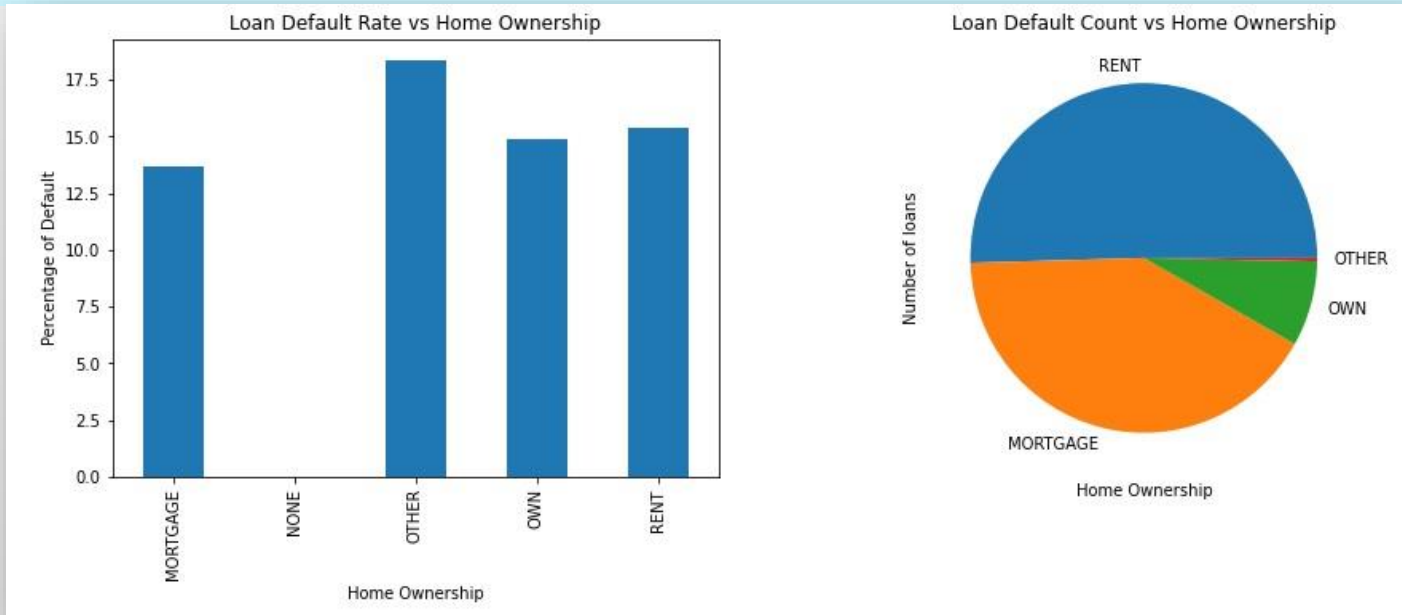
### Observations

- Percentage of loans getting charged-off for 5year term is higher as compared to 3year term.
- People asking for more money are opting for more time to pay the loans.
- Number of loans being processed is more for 3year term



# Home Ownership Analysis

## Analysis of Home Ownership w.r.t distribution of Loan Status

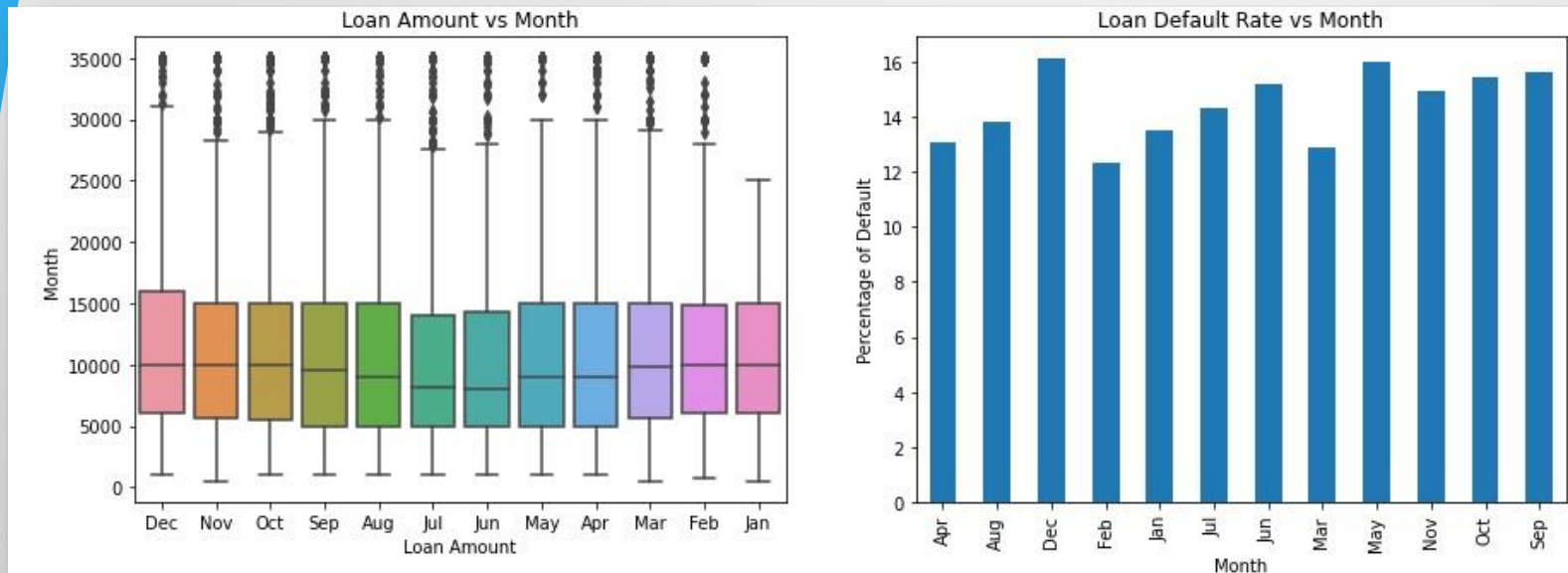


### Observations

- Most people applying for loans have home ownership as mortgage or Rent
- When looking at the graph the highest count of the borrowers are living in a rented housing. Rented house category have higher number of default count compared to people who have mortgaged their homes.
- Though the category 'Other' has the highest loan default rate. No. of applicants in this category are relatively less.

# Loan Month Analysis

## Analysis of Loan Month w.r.t distribution of Loan Status



### Observations

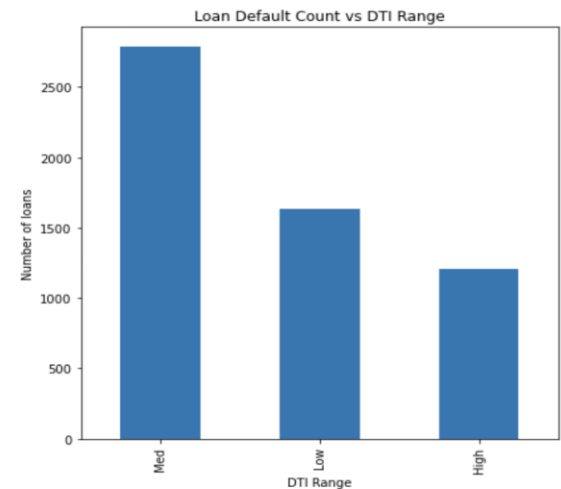
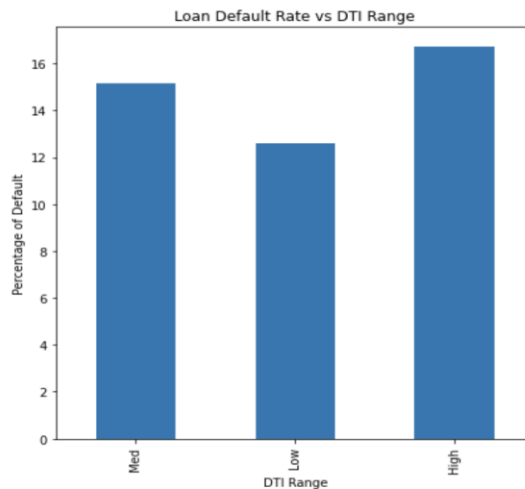
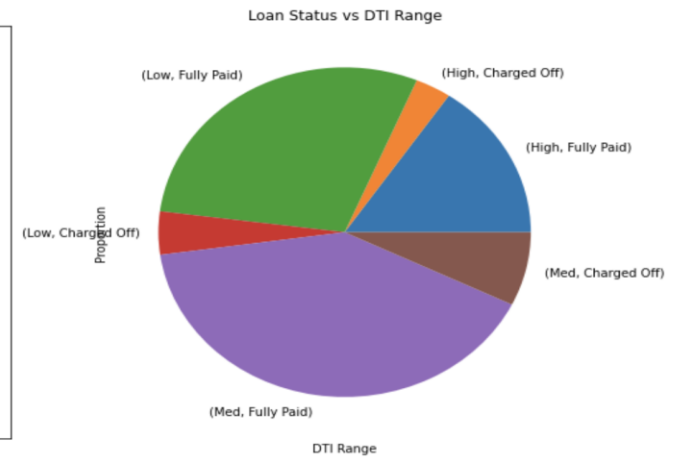
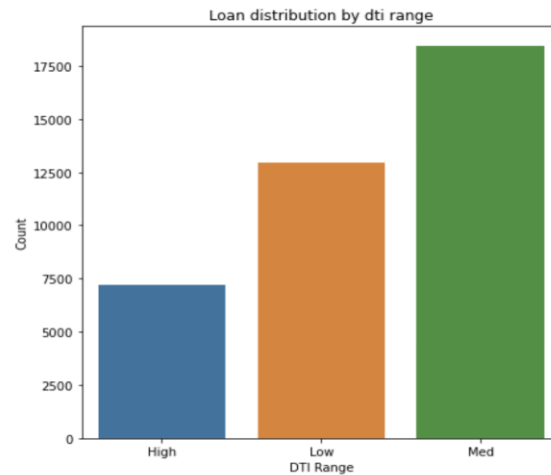
- The default rate is higher in December and the amount of loan being requested in the month of December seems to be on the higher side, Hence we can derive the inference that people are asking for more loans during the holiday season

# DTI Range Analysis

## Analysis of DTI w.r.t distribution of Loan Status

### Observations

- Most people asking for loans are in the medium DTI range (i.e dti is in between 10 and 20)
- Default count is higher in the medium DTI range
- We can see that default rate increases as dti increases, hence we can conclude that higher the dti, higher the chances of loan being charged off.

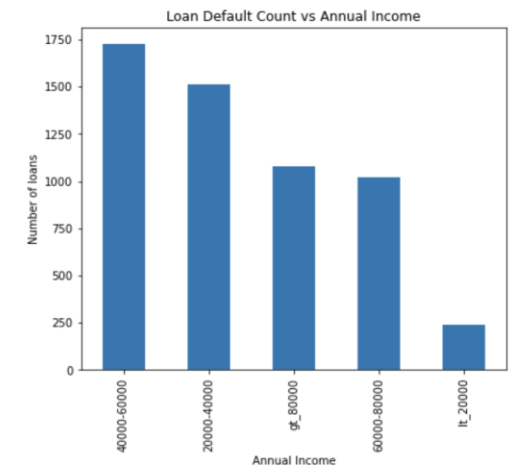
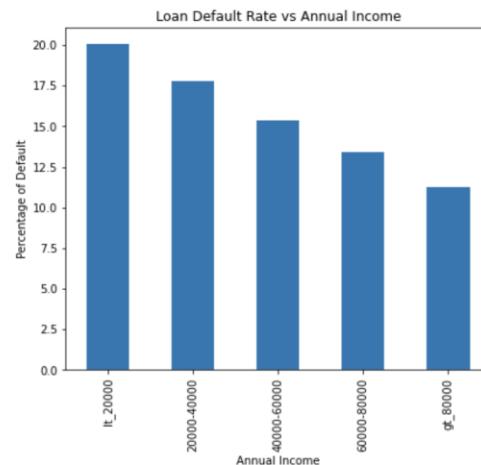
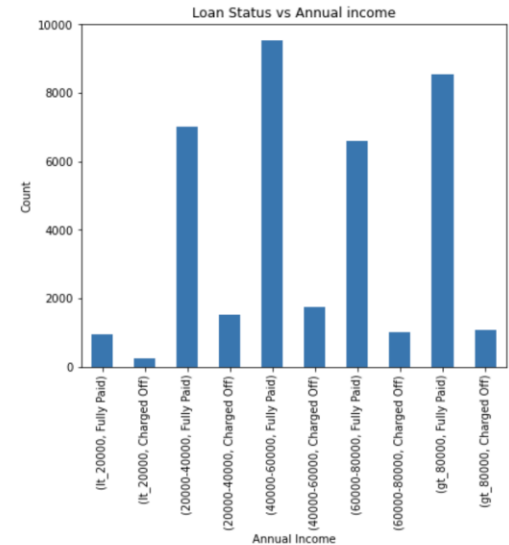
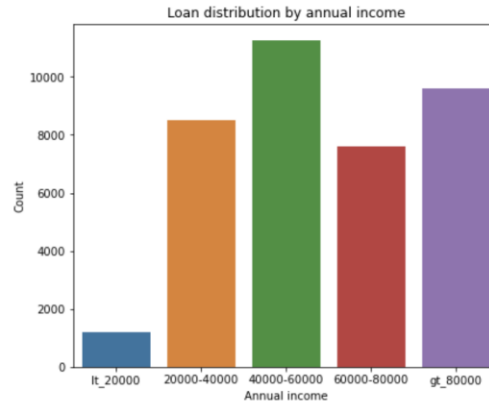


# Annual Income Analysis

## Analysis of Annual Income w.r.t distribution of Loan Status

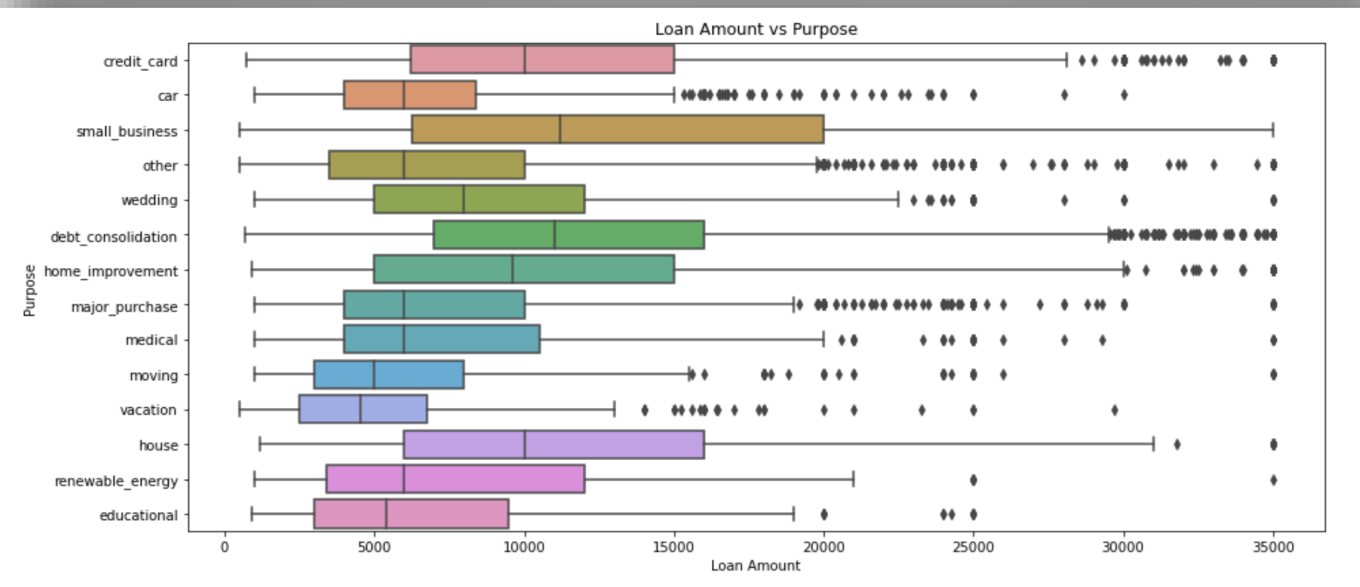
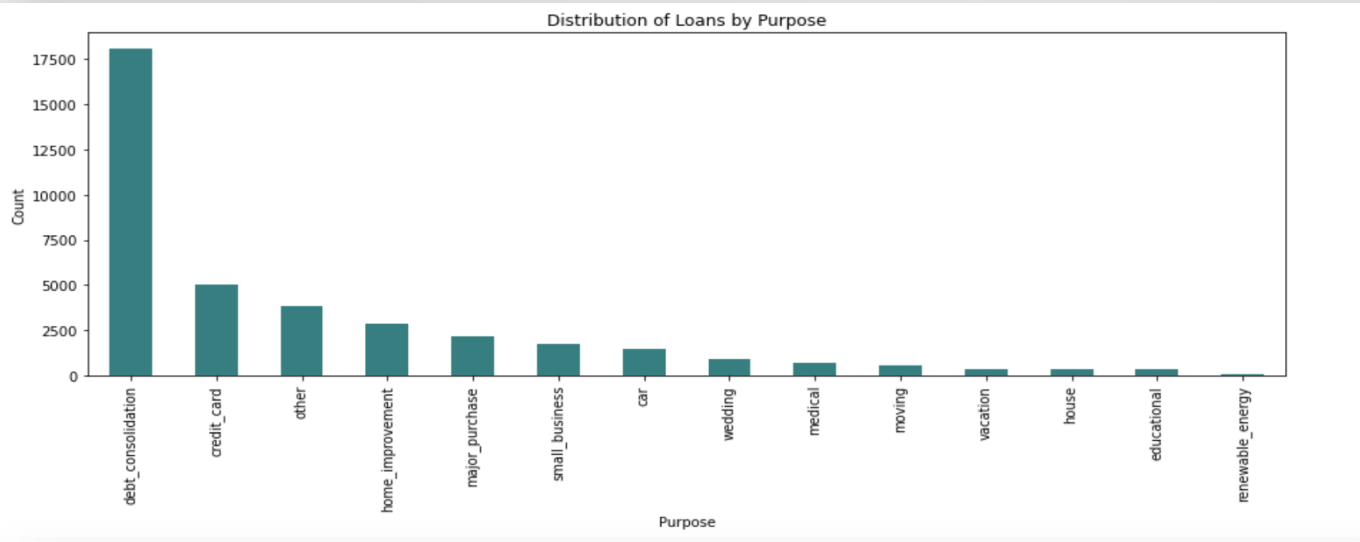
### Observations

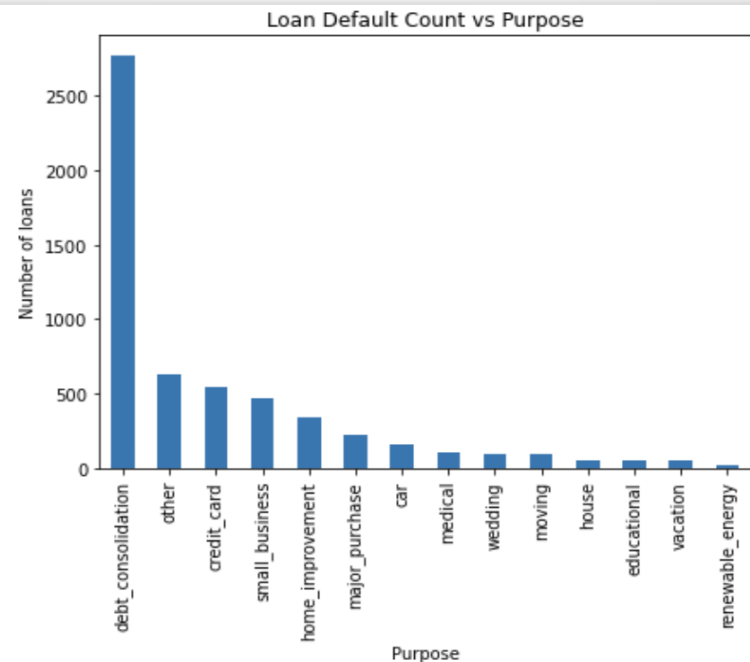
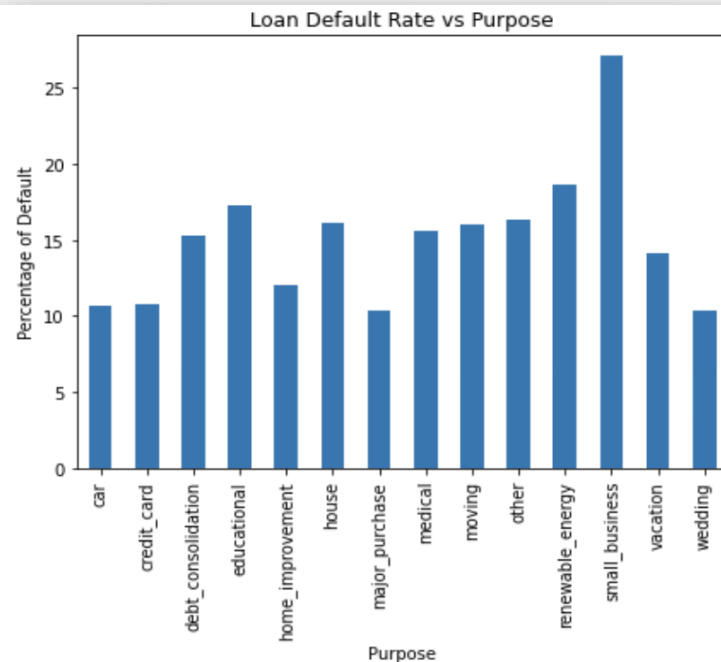
- People asking for most loans are in the income category of 40K to 60K
- Number of loan increases in the order (40K-60K>gt\_80K>20K-40K)
- Loan default rate is higher in income group where the income is less than 20k
- Count of default loans are in order (40K-60K>20K-40K>gt\_80K), hence focus should be more for income group 20K to 40K ( as they are asking for less loans but defaulting more)
- So, higher the annual income, higher the likelihood of loan being fully paid. However, if annual income is low, there are high chances of loan to be charged off.



# Loan Purpose Analysis

## Analysis of Purpose w.r.t distribution of Loan Amount





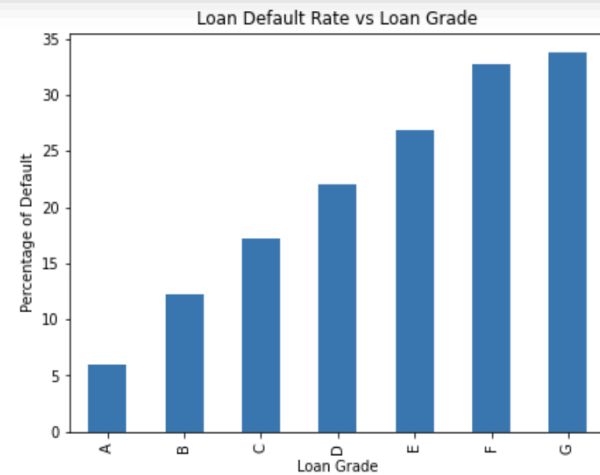
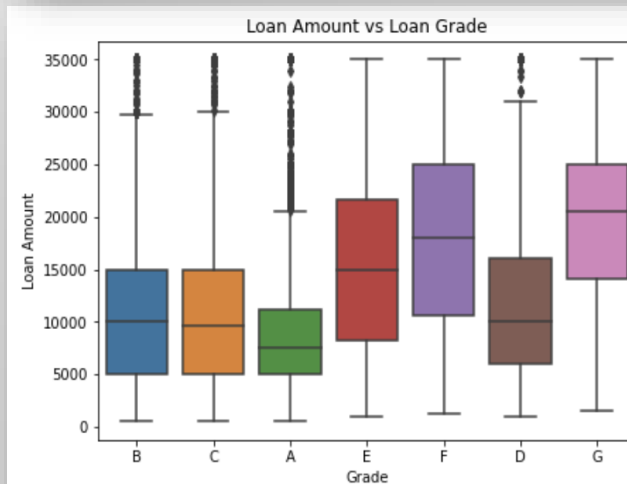
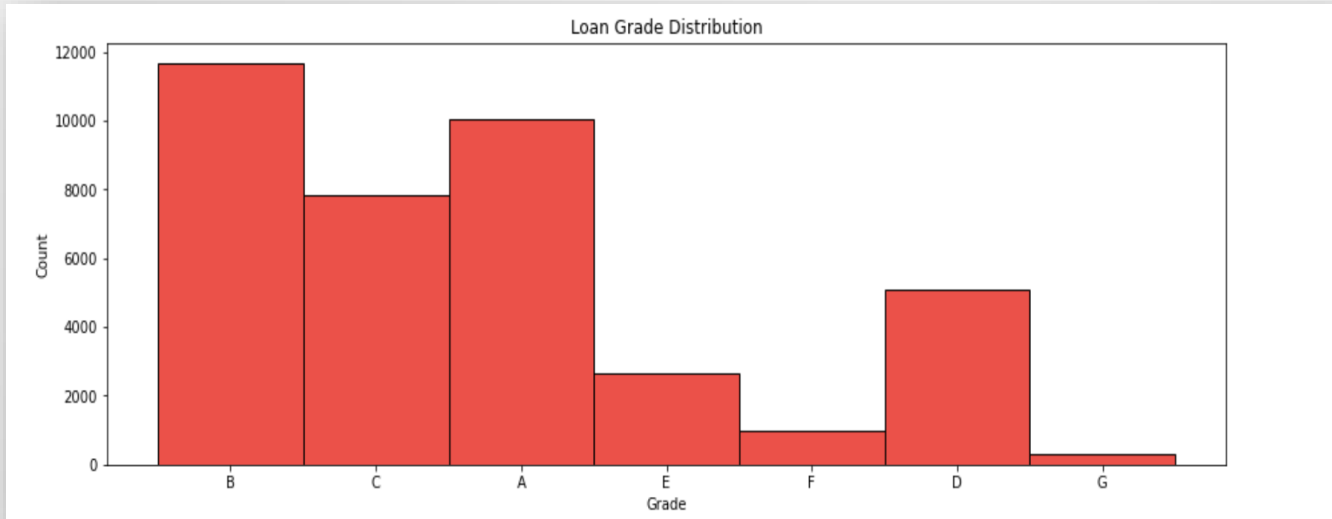
## Observations

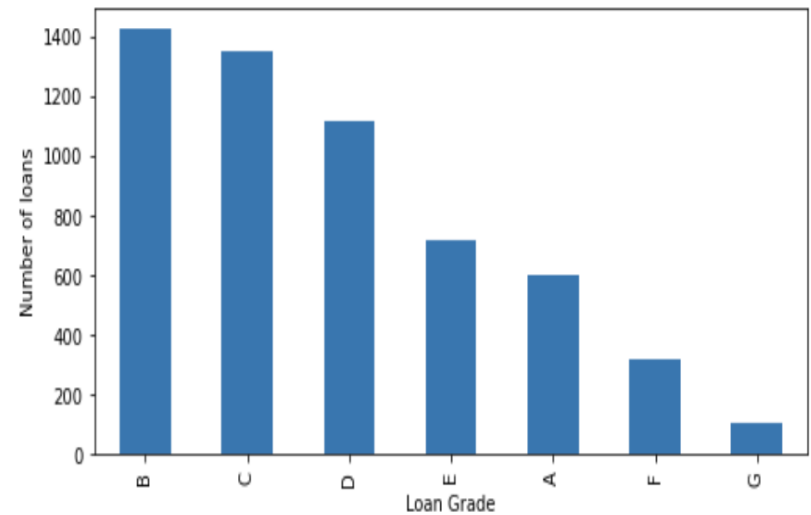
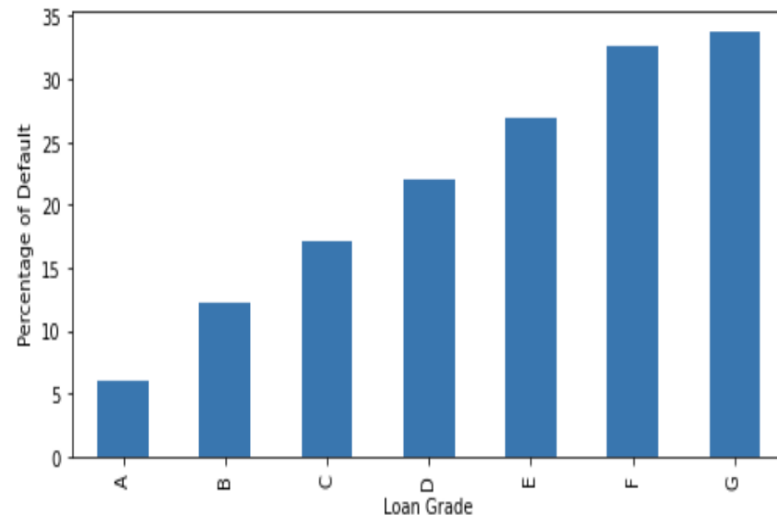
- Small business category is asking for more loan amount and their default rate is also high
- House and debt\_consolidation is another category where people are asking for more loan amount and their default rate is also higher.
- We see that the maximum number of loans are under the category of debt consolidation and max count of defaults are also in the same category



# Loan Grade Analysis

Analysis of Loan Grade w.r.t distribution of Loan Status





## Observations

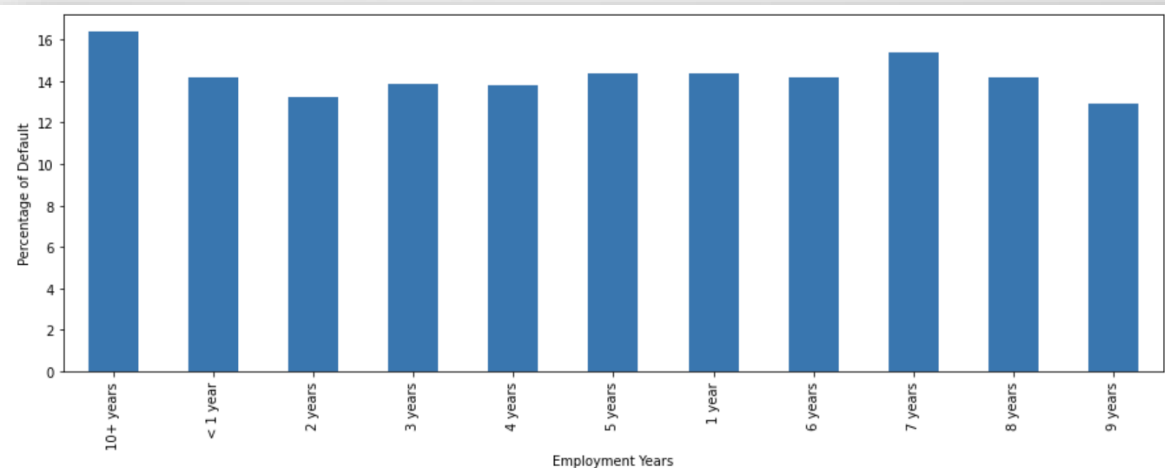
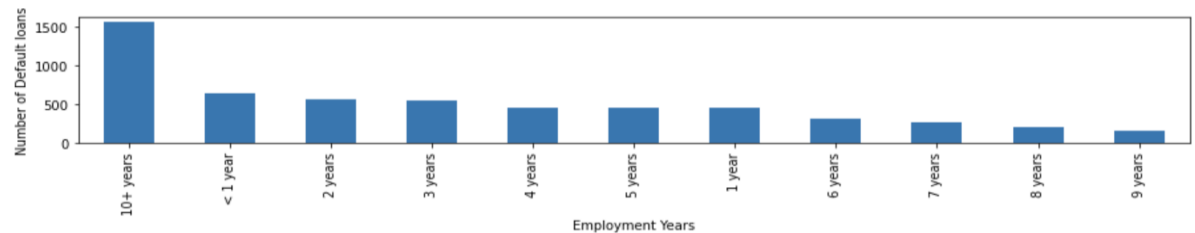
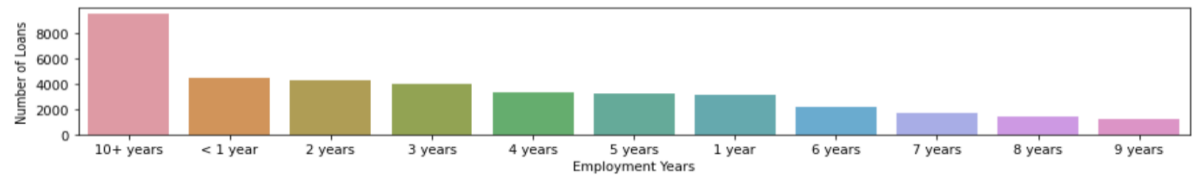
- Maximum number of loans disbursed are of high grade
- We can see that loan grade **E,F,G** are asking for more loan amounts compared to other grades
- **E,F,G** Loan grade has the highest default rate, and **B,C,D** category has the highest number of loan defaults

# Emp Length Analysis

## Analysis of Experience w.r.t distribution of Loan Status

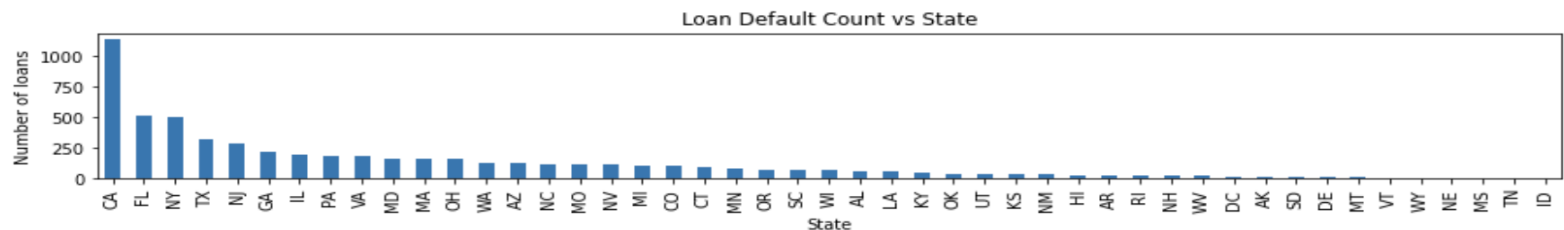
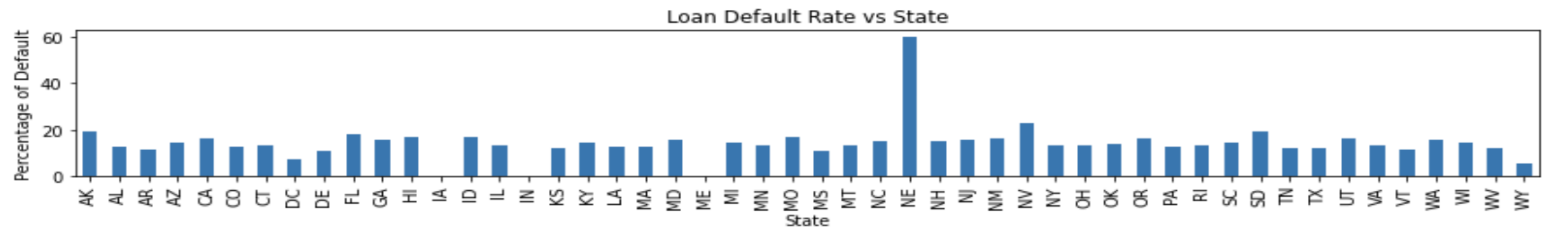
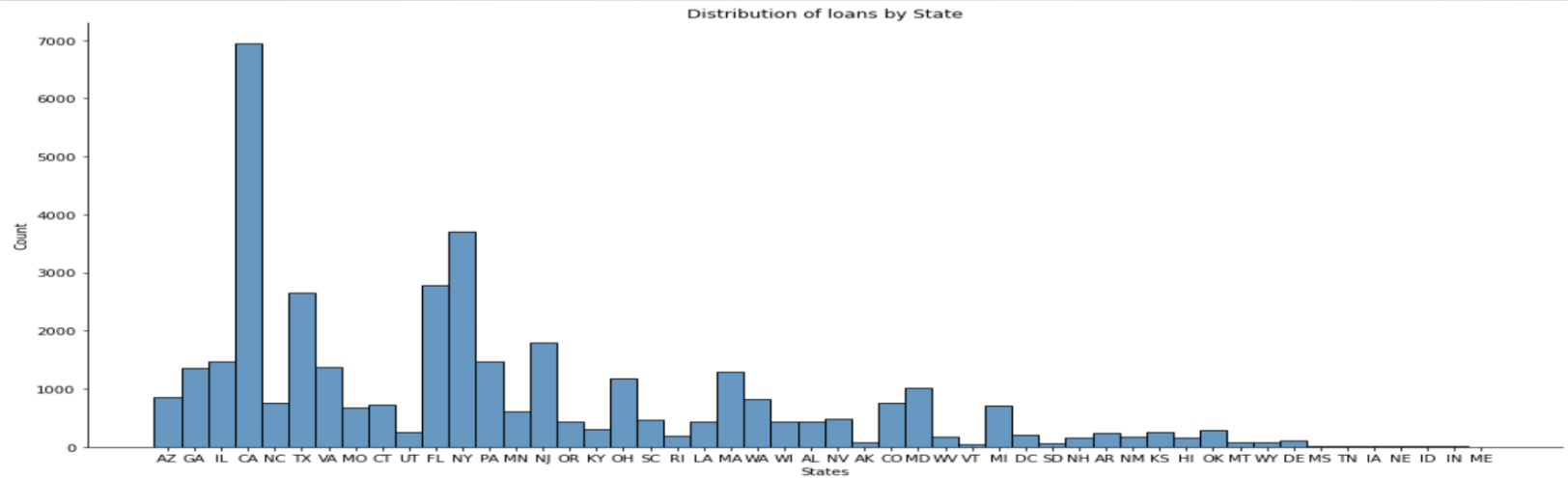
### Observations

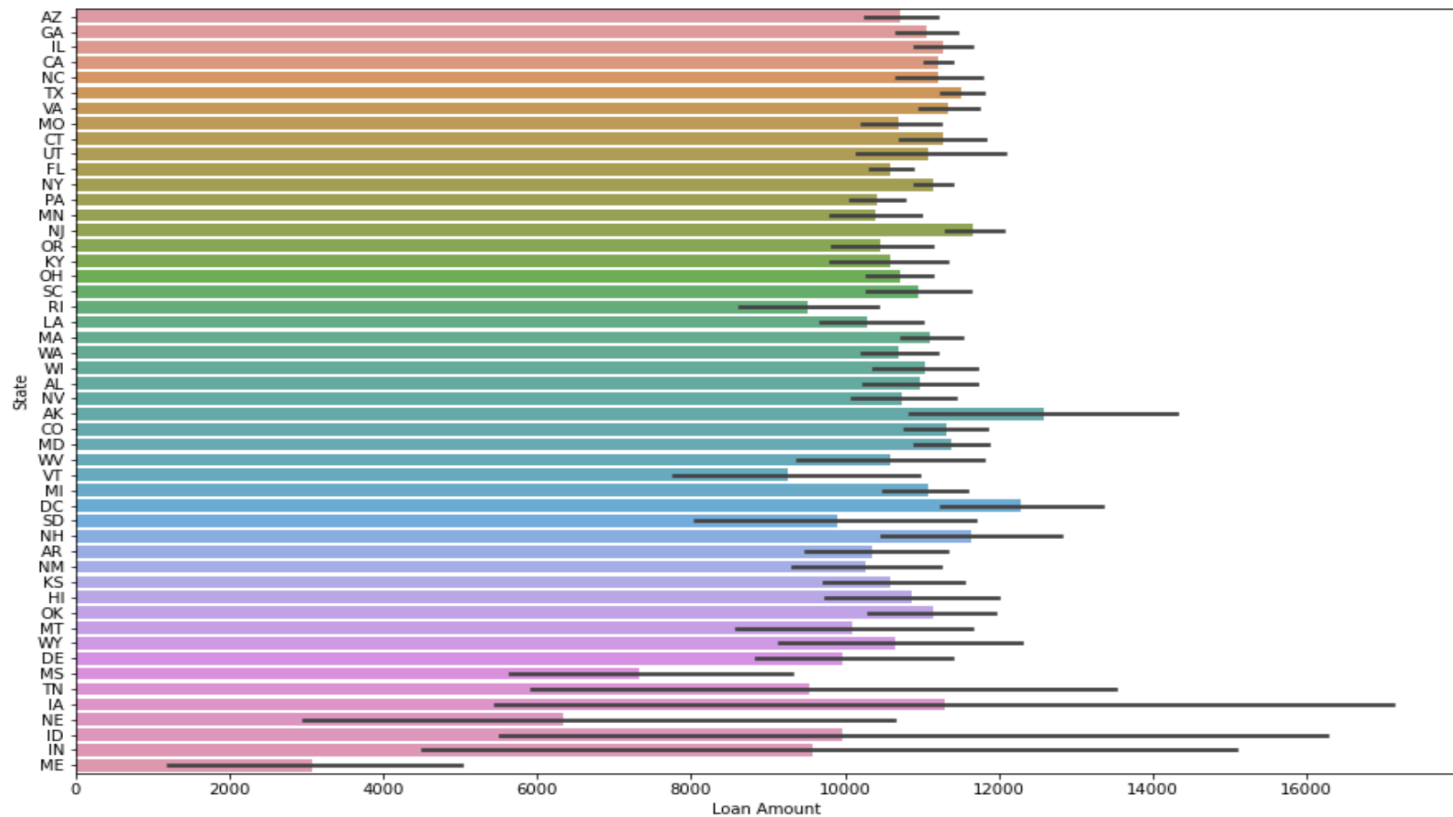
- We see that maximum number of loans are being asked by people who have more than 10 years of experience.
- People who are early in their professional career are asking for more loans compared to other category
- Default rate is higher for people with high experience while the default count is higher for people who are new to the corporate world.



# Address State Analysis

## Analysis of Loan Defaulters by State w.r.t distribution of Loan Status





## Observations

- California, New York, Florida and Texas are the states where most of the applicants are coming from
- People from state Nebraska and Alaska are asking for more loan amount as compared to other states, and people from these states are defaulting more.
- Count of default is of the order (California > Florida > New York > Texas)

## Recommendations

- Loan applications should be scrutinized Where loan amount requested is higher than 20K, and interest rate charged is greater than 15%.
- Loan applications should be scrutinized if the grading/subgrading of loans falls under (B5,B3,C1,B4,C2) or (F5,G3,G5,G2,F4) category.
- Lending club should do additional check for dti that is higher than 20 as more the dti the chances of default is higher. Lower the dti higher the income, lesser the chance of default
- Employees with experience greater than 7 years are asking for big loan amount and defaulting more hence LC should check such applications with additional guardrails.
- Loan applications coming during holiday season are defaulting more, hence LC should check these loan applications more rigorously.
- Loan applications asking for more time to pay are defaulting more hence when term is set to 5 years, LC should check those applications.
- LC should loan applications from California, Florida, Nebraska, Alaska, Newyork and Texas more thoroughly.
- If the loan purpose is debt consolidation or small business LC should check those loans more as the default rate and number of defaults are higher around these categories.