LENDING CLUB CASE STUDY

Anshul Khadse

Problem Statement

Lending Club

Lending Club 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.

Objective

Lending Club wants to understand the driving factors behind loan default, i.e., the driver variables which are strong indicators of default.

The company can utilize this knowledge for its portfolio and risk assessment.

Problem Statement

As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Analytical Approach

Data Cleaning

- Columns with 80% or more missing values, all random values, or a single value column should be removed.
- Convert values to the appropriate int, datetime and float representations

1

Univariate Analysis

- Check distributions , frequencies and count of various numerical and categorical variables
- Create derived variables through binning

2

Segmented Univariate Analysis

- Analyze variables against segments of other variables
- Create derived variables

3

Bivariate Analysis

- Do correlation analysis
- Check how two variables affect each other
- Analyze joint distributions

4

Summarize Results

- Summaries the key factors helps in detection of Defaulters
- Publish the results

5

Data Provide By Lending Club

Lending Club data is broadly classified into three categories, as shown in the table below, and columns under each category are also shown below:

Applicant Information

- Grade
- Sub-Grade
- Employment Title
- Employment Length
- Home Ownership
- Annual Income
- Purpose
- Address

Loan Amount & Interest Rate

- Loan Amount
- Funded Loan Amount
- Term
- Interest Rate
- Installment
- Loan Status
- Open Account
- Revol Balance

Applicant Behaviors

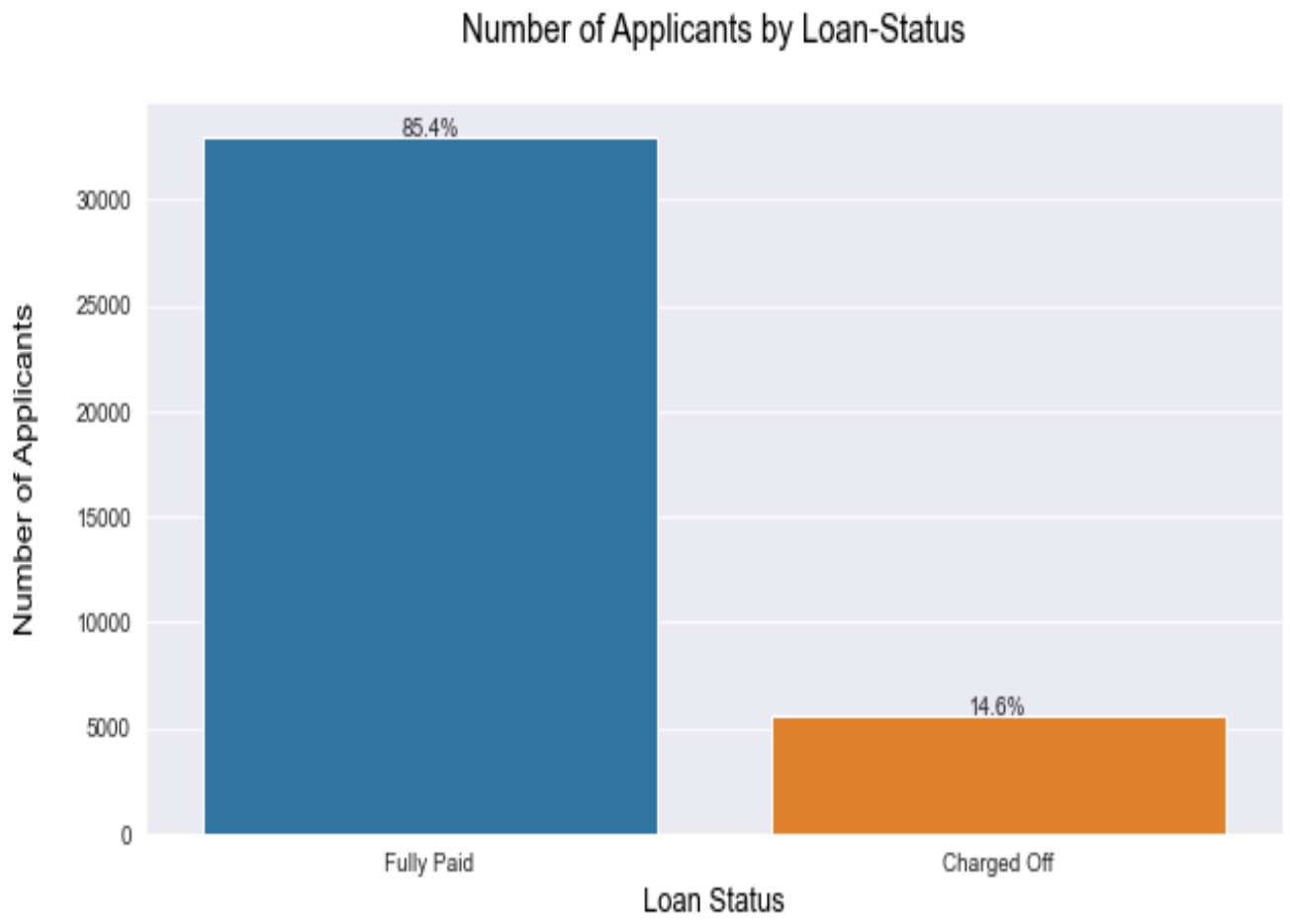
- Deling_2yrs
- earliest_cr_line
- inq_last_6mths
- total_acc
- total_pymnt_inv
- last_pymnt_d
- last_pymnt_amnt
- last_credit_pull_d

DATA CLEANING

- Lending Club provided us with a dataset that included **39717 records and 111 columns**.
- We discovered **56 columns with more than 80% missing** values after analyzing the missing values percentage from each column.
- Columns with 80% or more missing values are excluded from further analysis.
- Columns containing information about **customer behaviors**, such *delinq_2, earliest_cr_line, recoveries, out_prncp, total_acc, last_pumnt_amnt* and so on, are also **excluded from analysis**.
- The goal of this study is to identify the factors that assist clients in determining defaulters because these above-mentioned characteristics are only known to clients after loan approval and thus cannot be useful in our analysis.
- We also found out that there are some column hold **single value for all records**. There were total 9 columns which show this pattern; hence we **exclude them from our analysis**.
- Column such as “int_rate” and “issue_d” are in string formate. We convert “int_rate” to float and derived “issue_d_month” and “issue_d_year” column from “issue_d” column.
- Rows with **loan status “Current”** were also **removed** as they were irrelevant to the purpose .

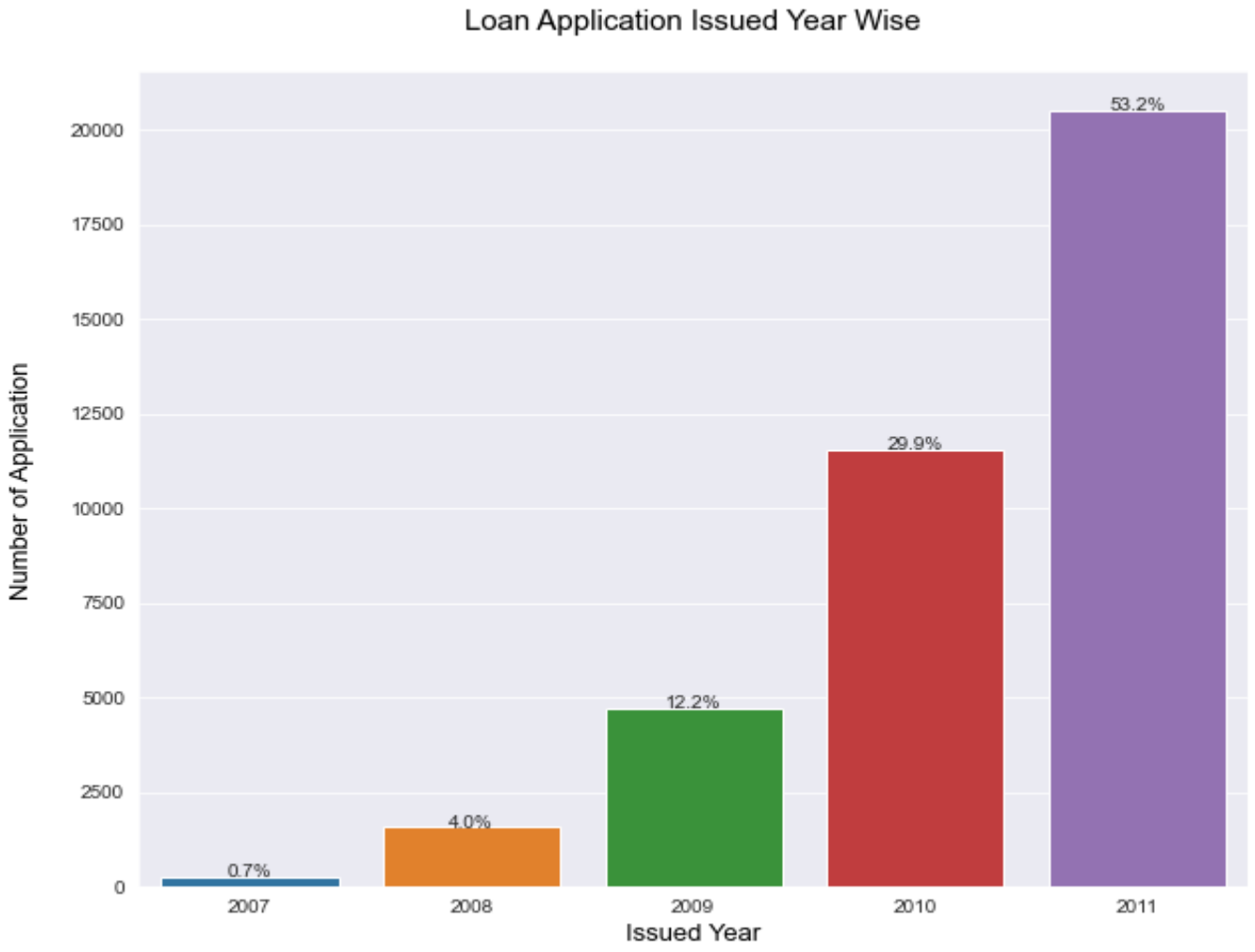
Data Understanding

- There are 38,577 applicants in all.
- The number of defaulter applications is 5827.
- The default rate is 14.6%.



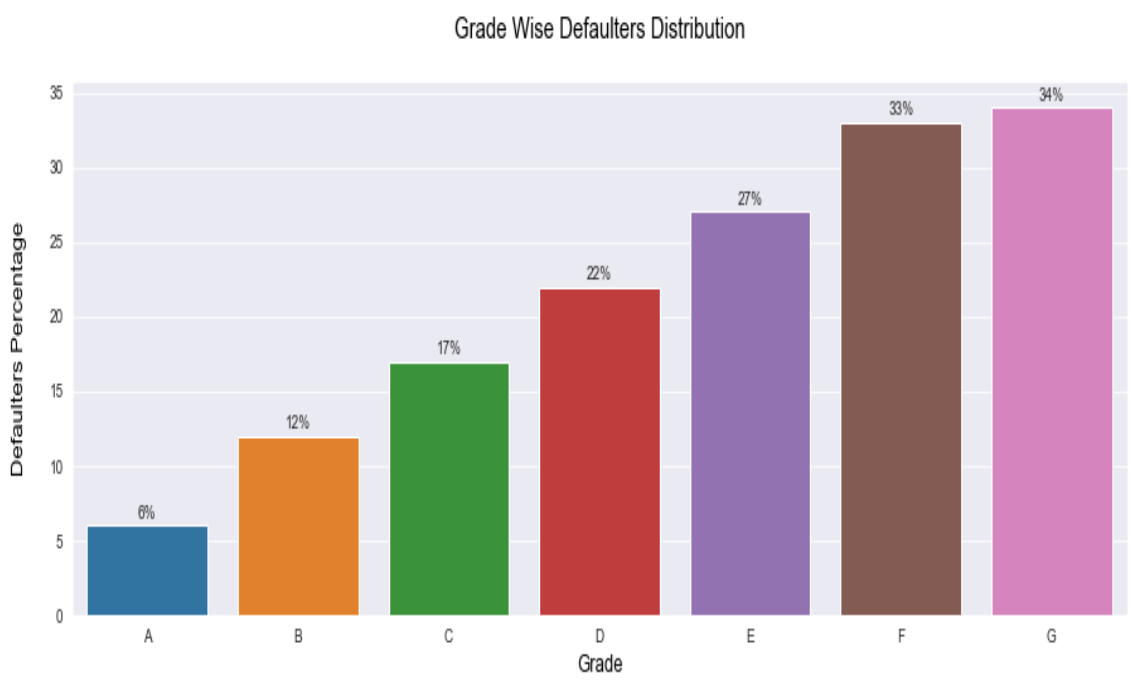
Year wise Application Distribution

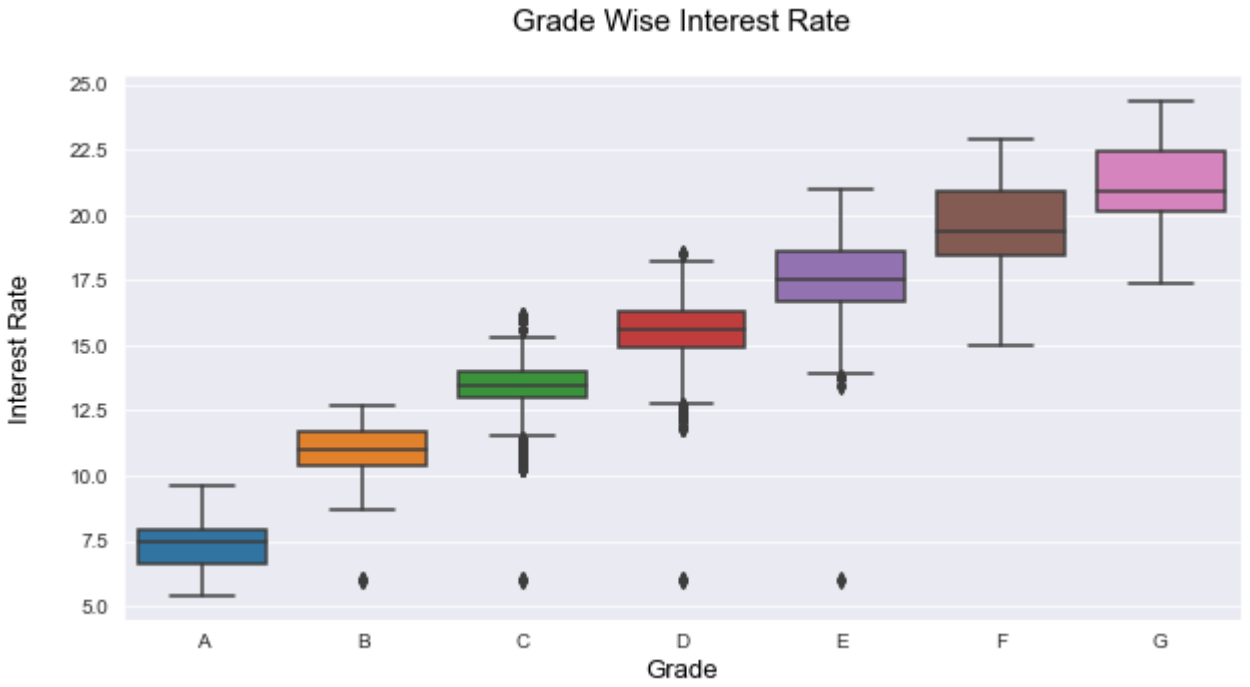
- The number of application for loan are drastically increasing every year
- In 2011 there is 78% increase in application compared to 2010
- Similarly, there is 145% increase in application from 2009 to 2010



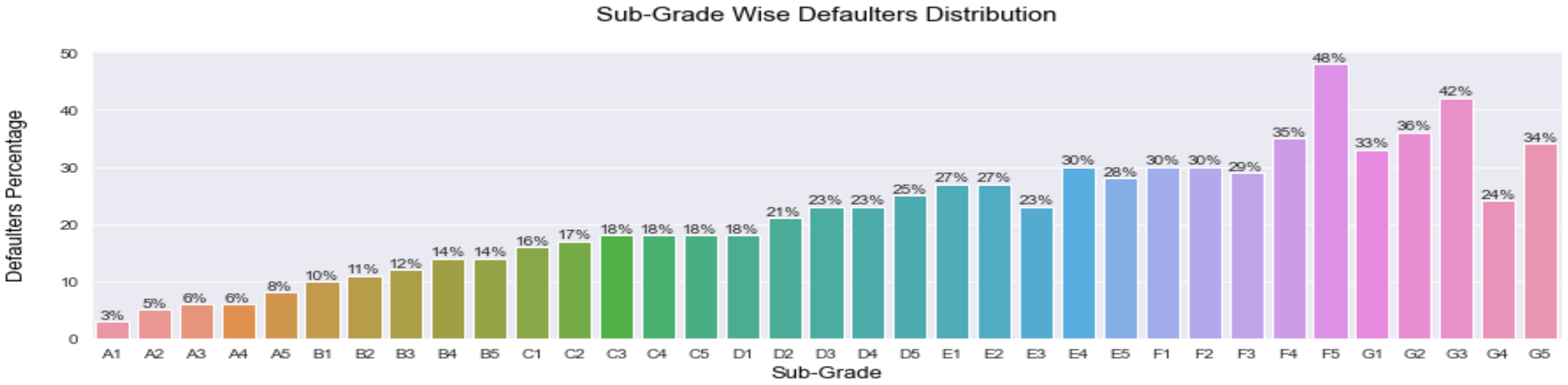
Effect of Grade on Defaulters

- For grade G and F has high percentage of defaulter i.e 34% and 32% respectively
- Lower grades have high chance of defaulter
- Lending Club charges higher interest rates as the grade of loan becomes worse



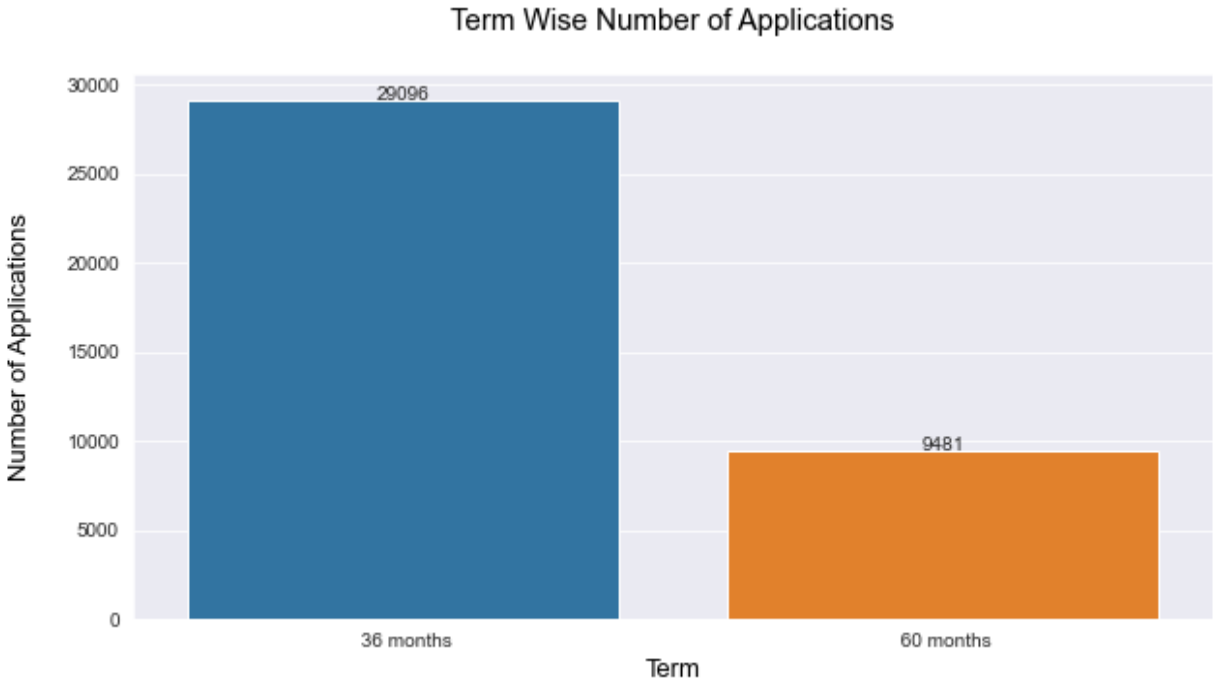


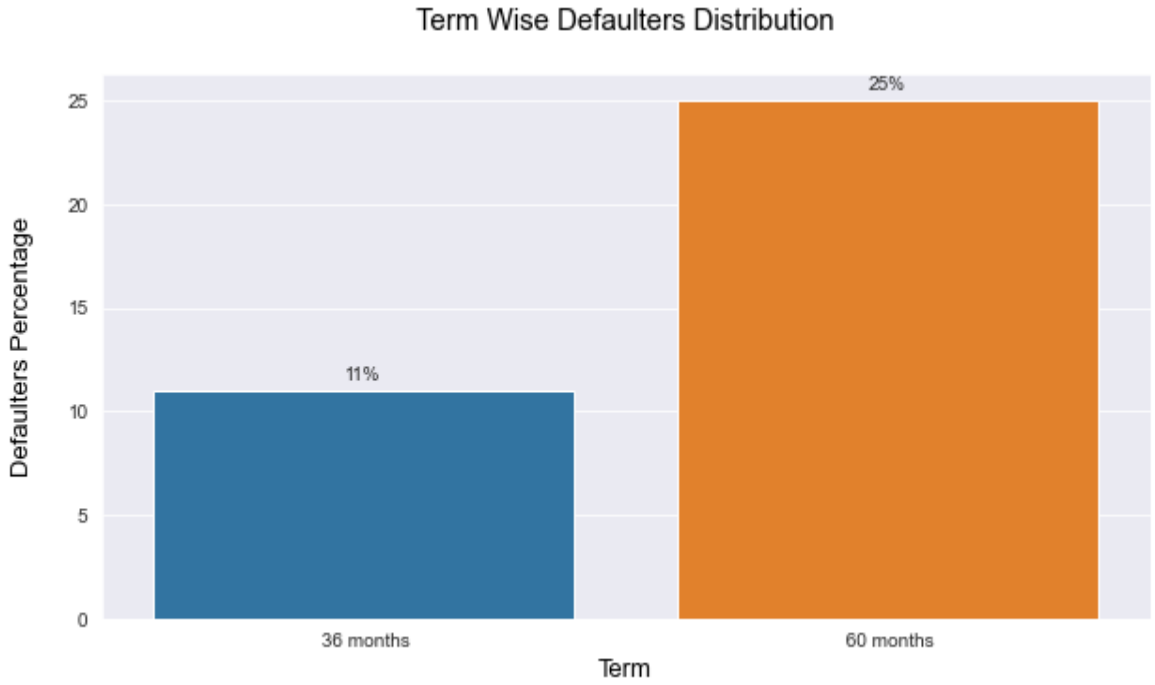
Effect of Sub Grade on Defaulters



- Sub Grade F5, G3 and G5 has high number of defaulters
- As the Sub Grade increases from A to G the percentage of defaulters are also increasing this is due to high interest rate on lower grades

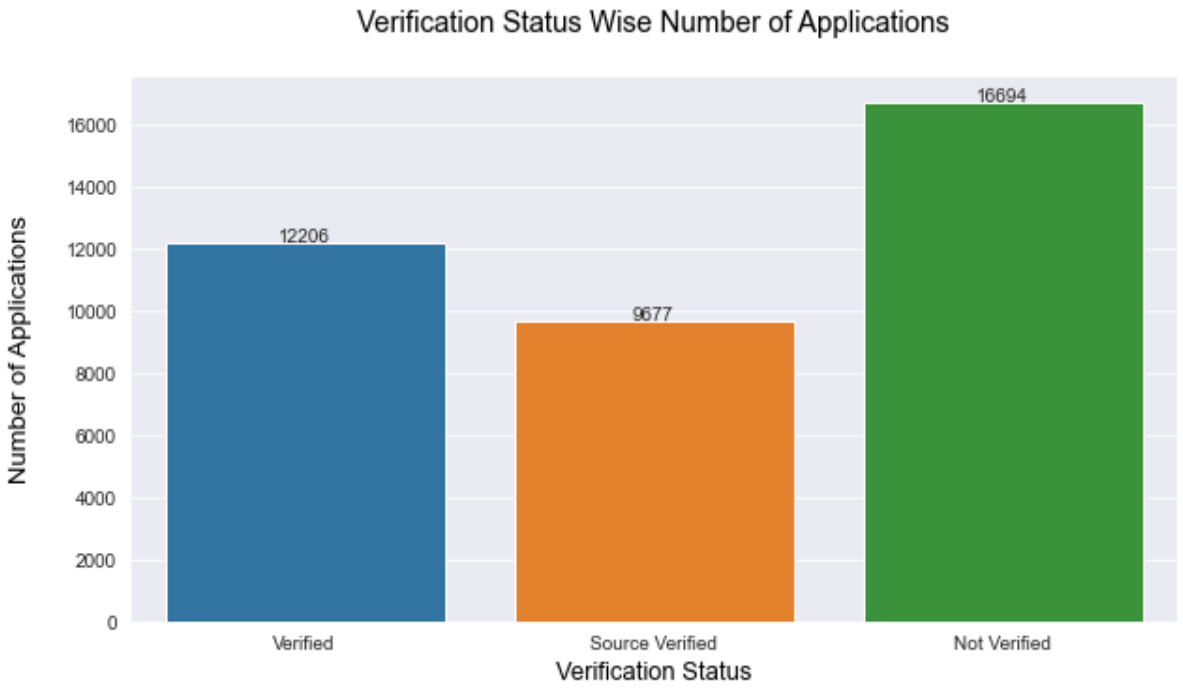
Effect of Term on Defaulters

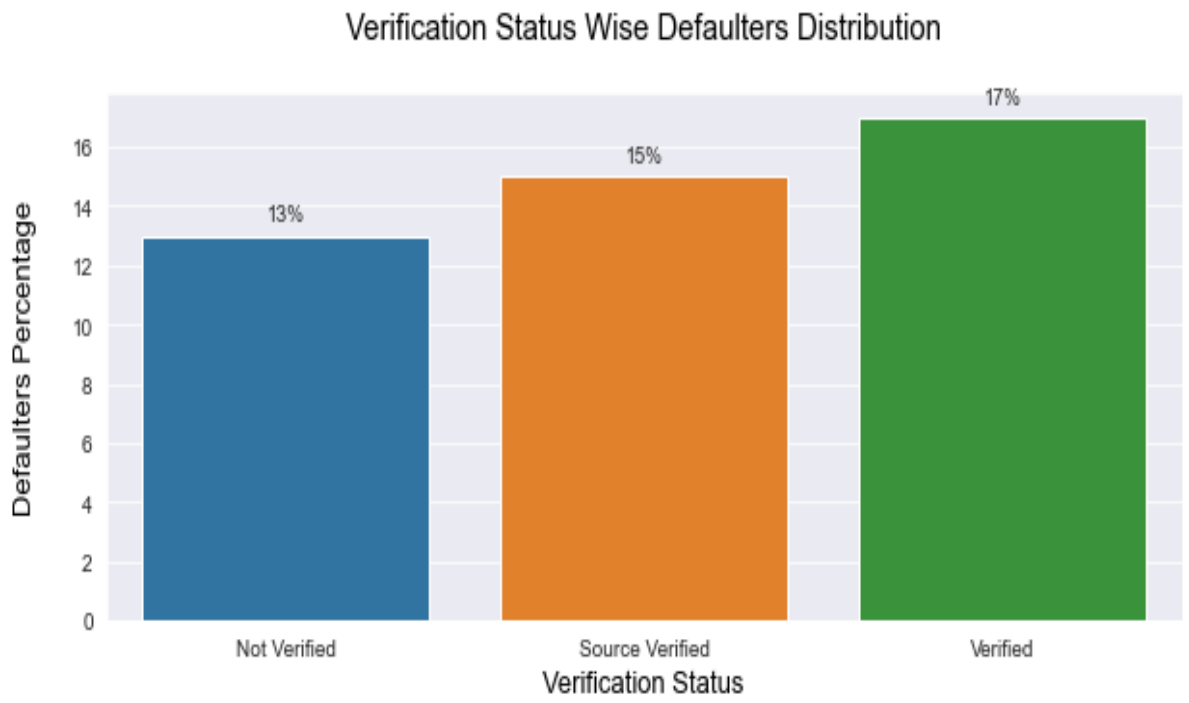




- According to the left graph, more loan applications are for Short-Term loans, and 29,096 of total applications are for Short-Term loans.
- However, the rate of defaulters is higher in the Long-Term. As compared to the Short-Term, it is more than double. As a result, Long-Term applicants are more likely to default.

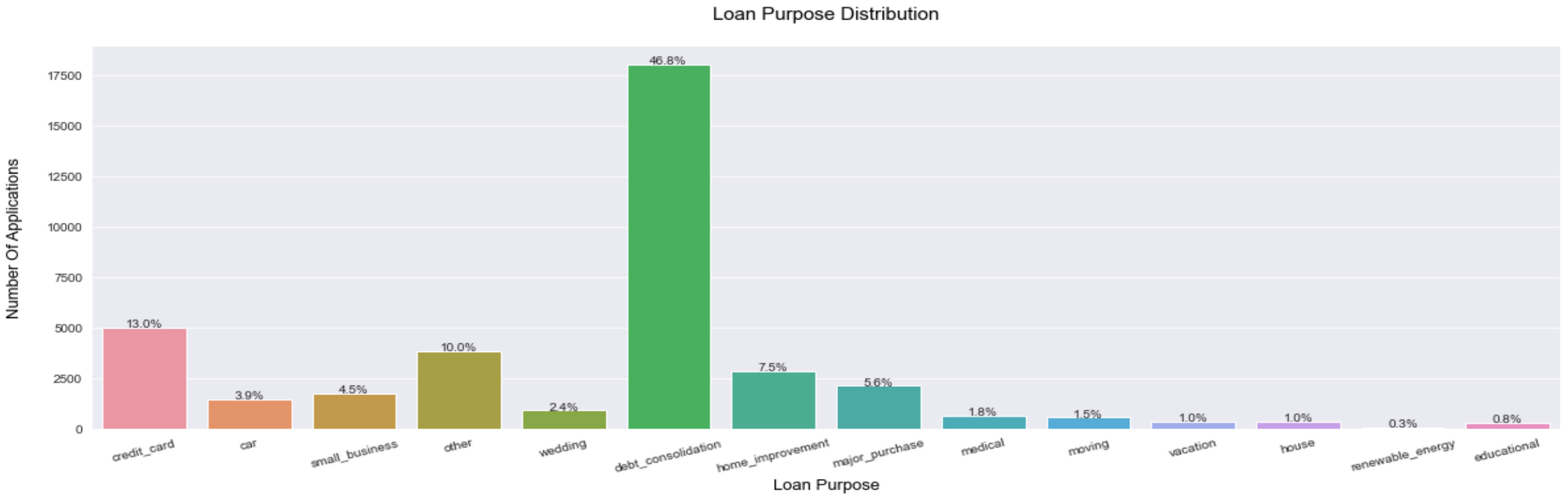
Effect of Verification Status on Defaulters





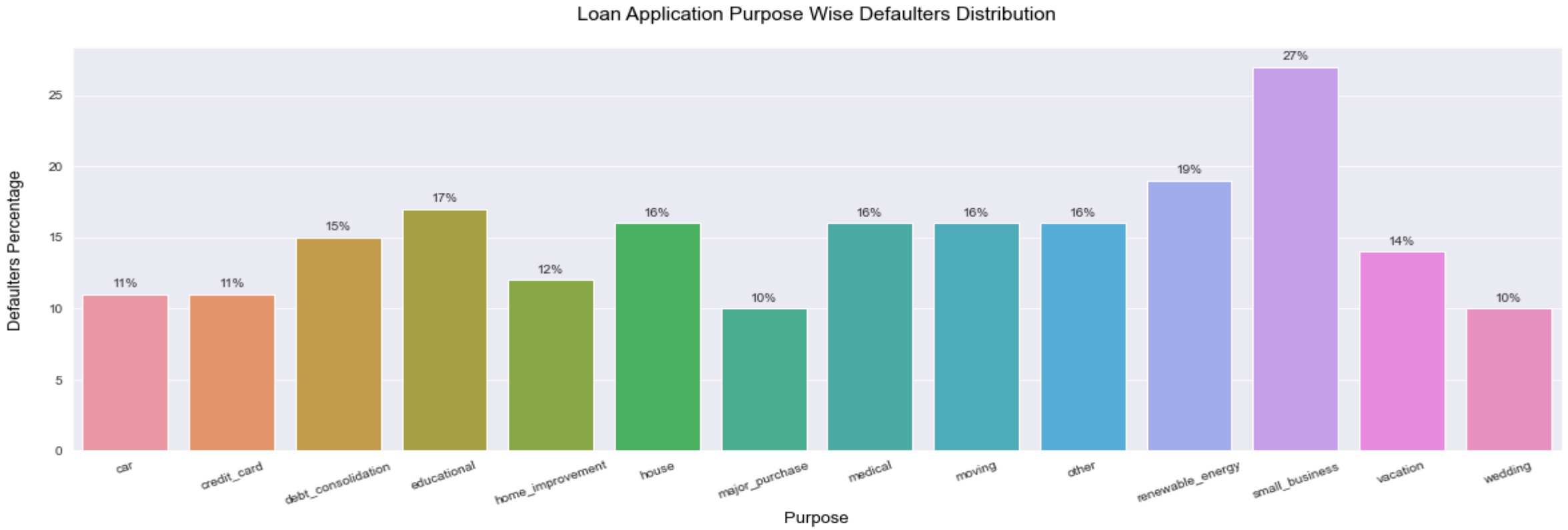
- From left chat we can clearly see that number on “Not Verified” applications are more than “Verified” and “Source Verified”
- Surprisingly, right chart show that the percentage of defaulter is more in “Verified” category

Purpose for Loan Application



- The 40.8% of loan application is for “debt_consolidation” purpose
- Other than that, 13% are for “credit_card” and 7.5% are for “home_improvement” are some major purpose

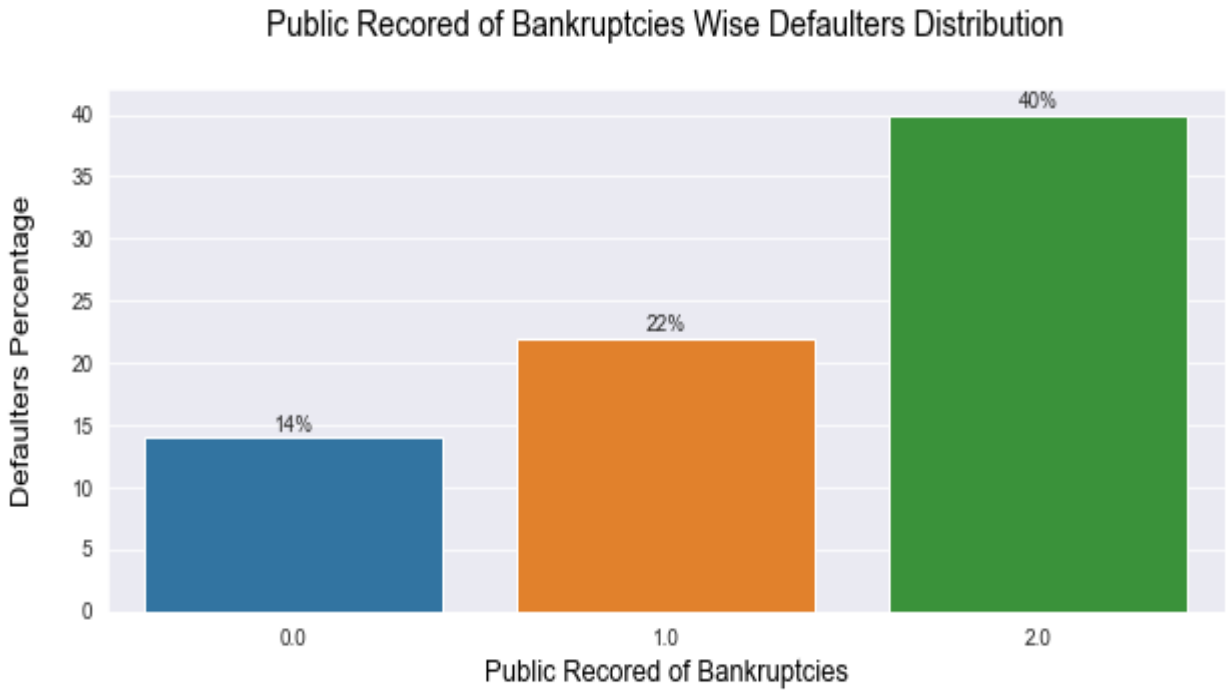
Effect of Loan Purpose on Defaulters



- “small_business” has around 27% of defaulter which is heights among all other categories
- “education” has 17% of defaulter which is 2nd height among the list
- More than a quarter of loans taken for the purpose of running a “small_business” see defaulters.

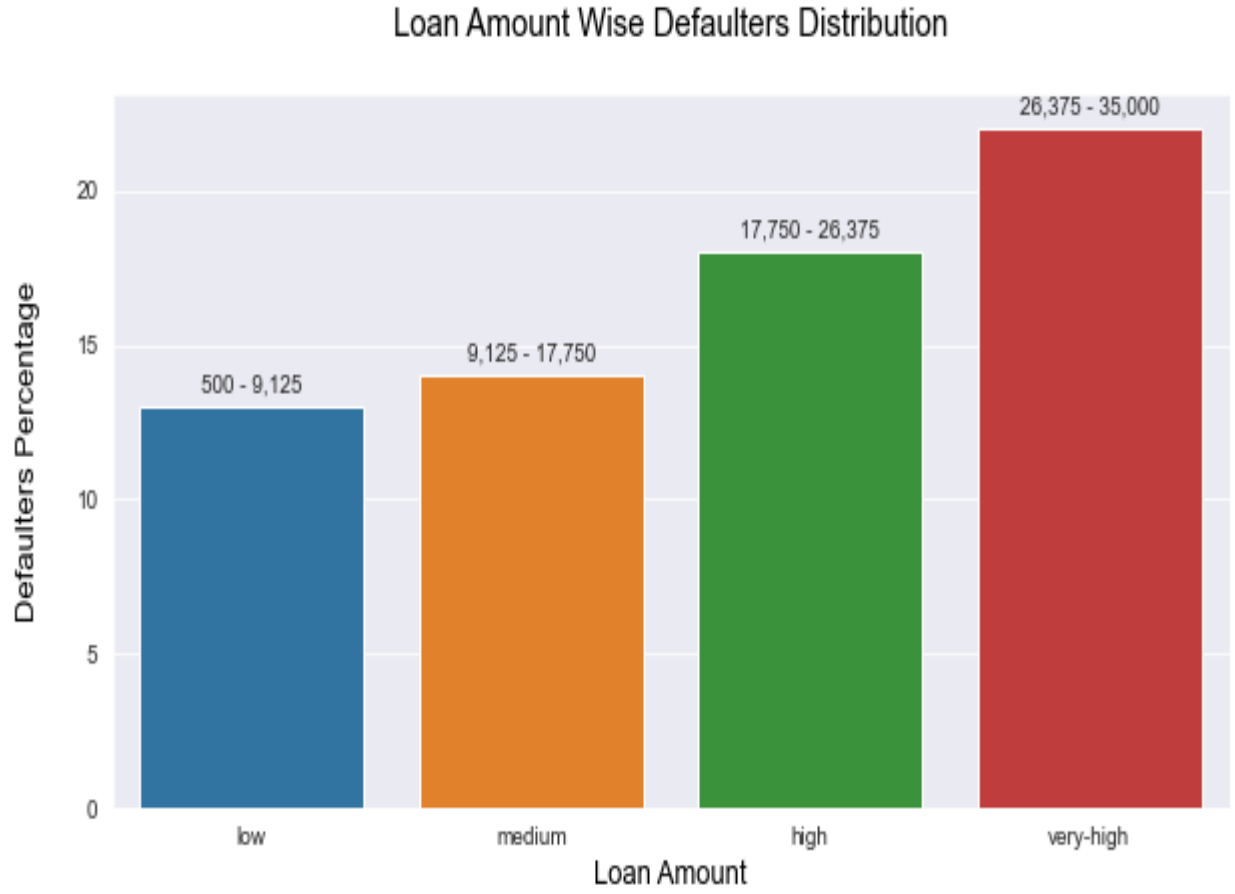
Public Record of Bankruptcies

- Applicant having prior record of bankruptcies are more likely to be defaulters
- Hence extra care needs to taken while approving loan for such applicants



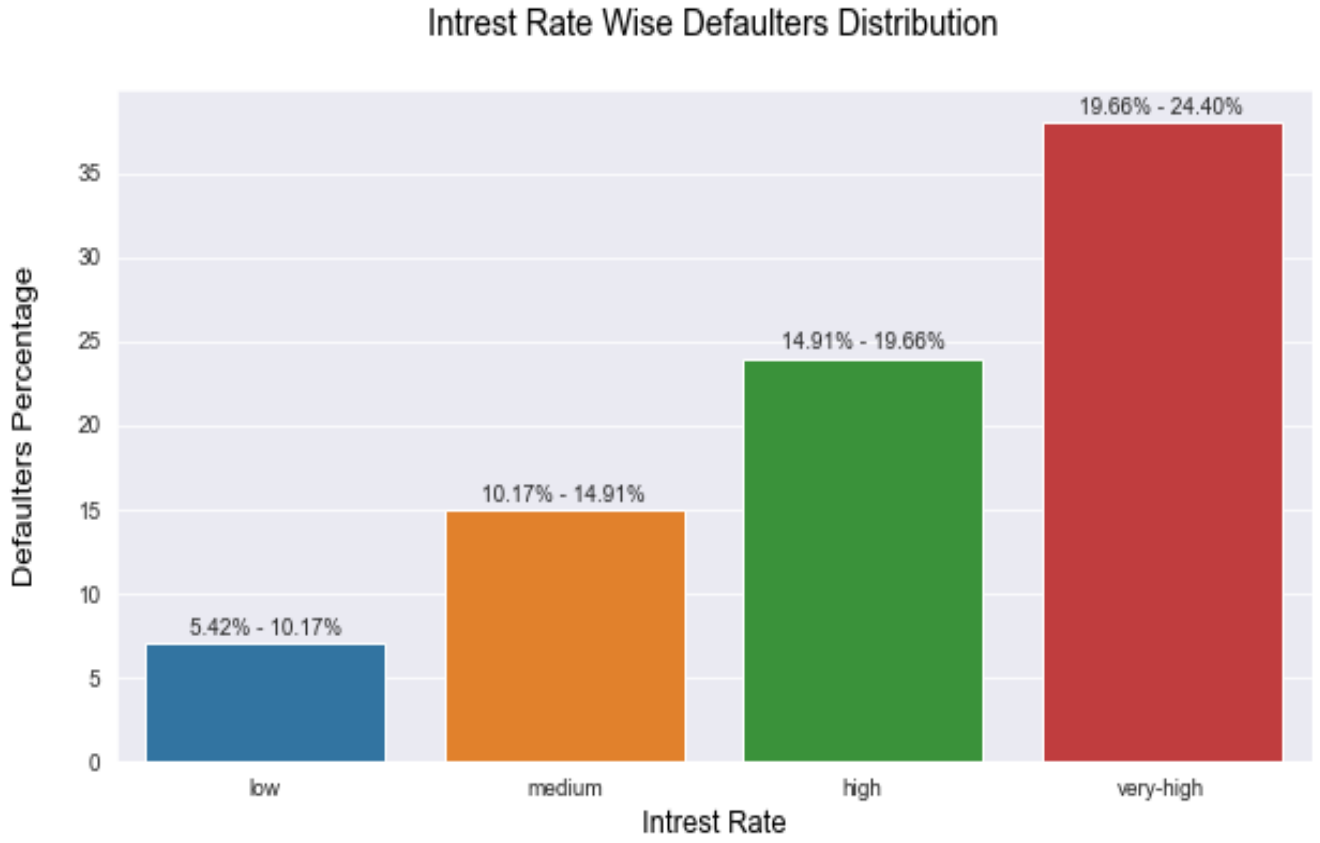
Effect of Loan Amount on Defaulters

- Value above bar indicating the range of value info different category
- Loan Amount 500 to 9,125 belong to “low” bins
- Clearly, we can see the trend that as the loam amount increases the percentage of defaulter is also increase

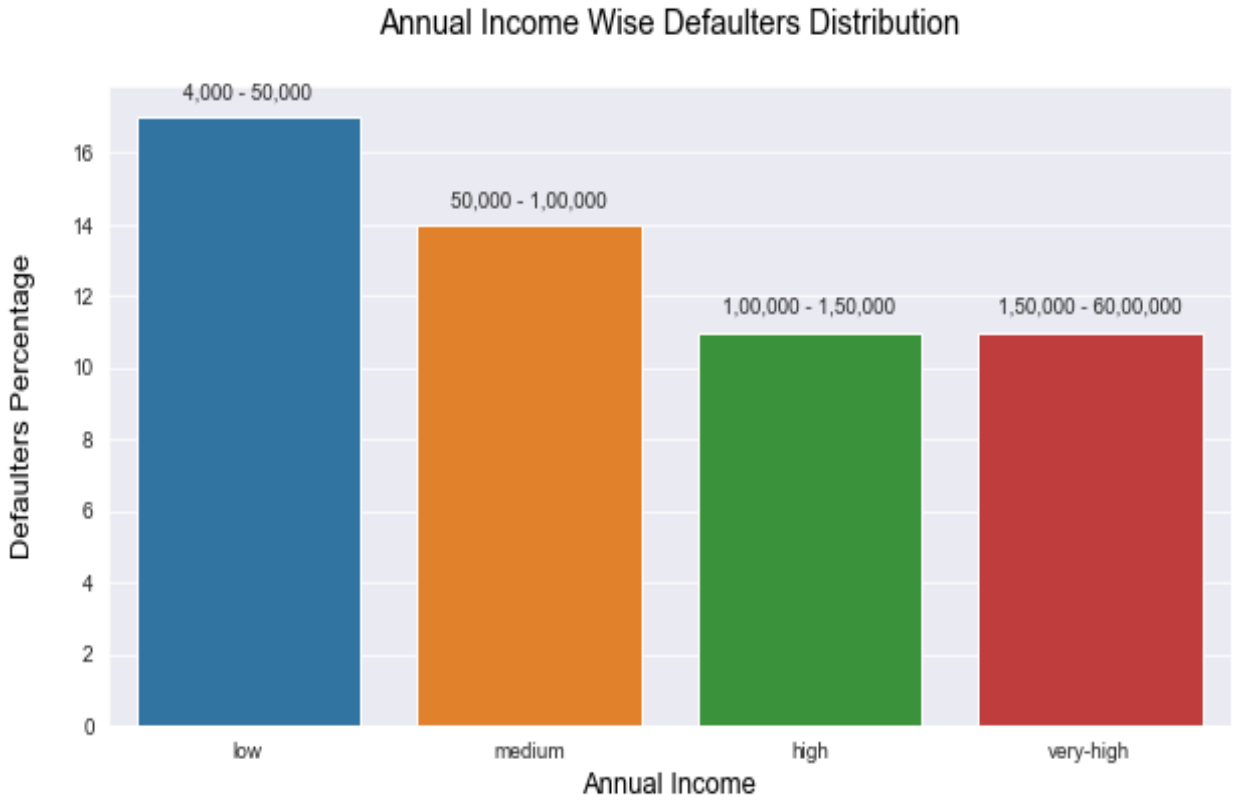


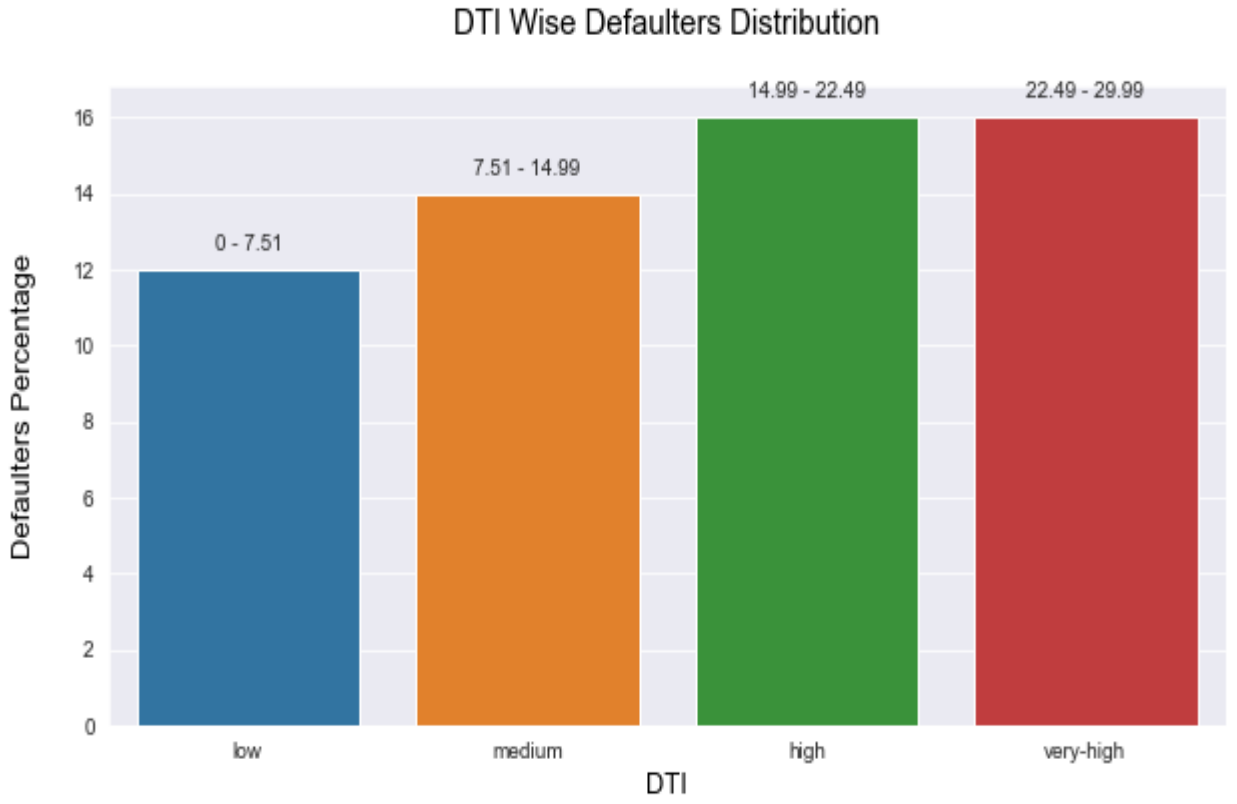
Effect of Interest Rate on Defaulters

- Value above bar indicating the range of interest into different category
- Interest rate 10.17% to 14.91% are represented by “medium” category
- As the rate of interest increases the chance of applicant being defaulter is also increase



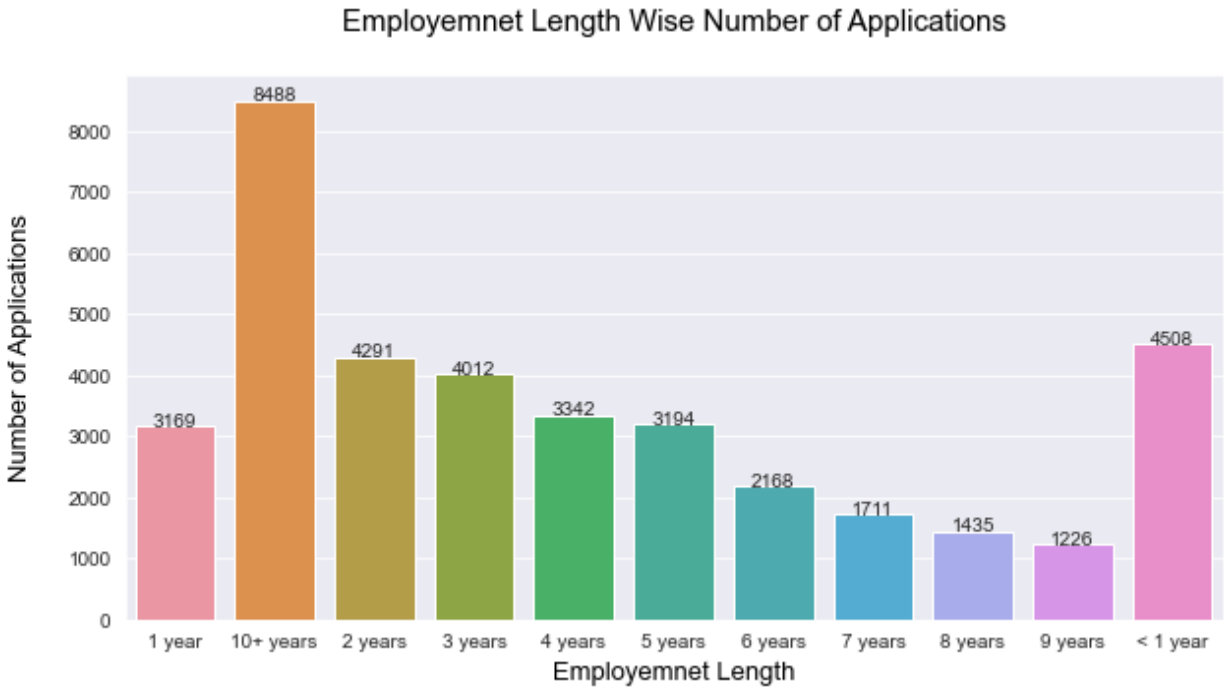
Effect of Annual Income & DTI on Defaulters

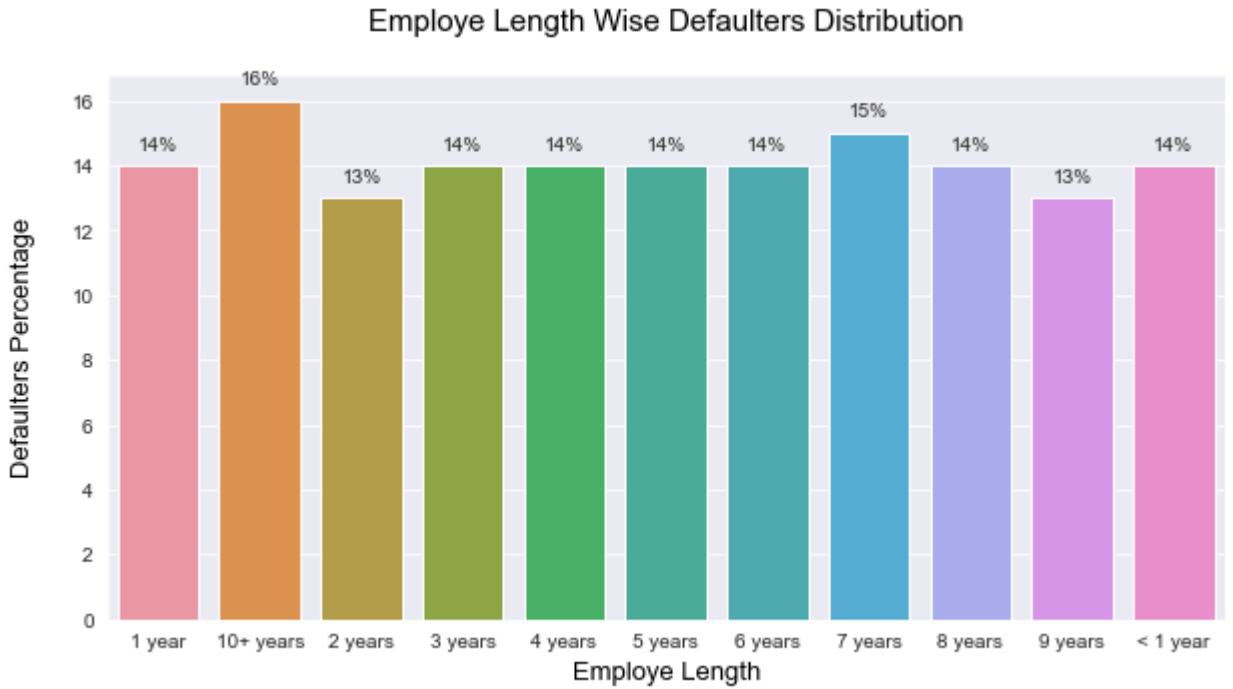




- Lower the annual income higher the percentages of the defaulter
- Higher the DTI higher the chance of applicant is defaulter

Effect of Employment Length





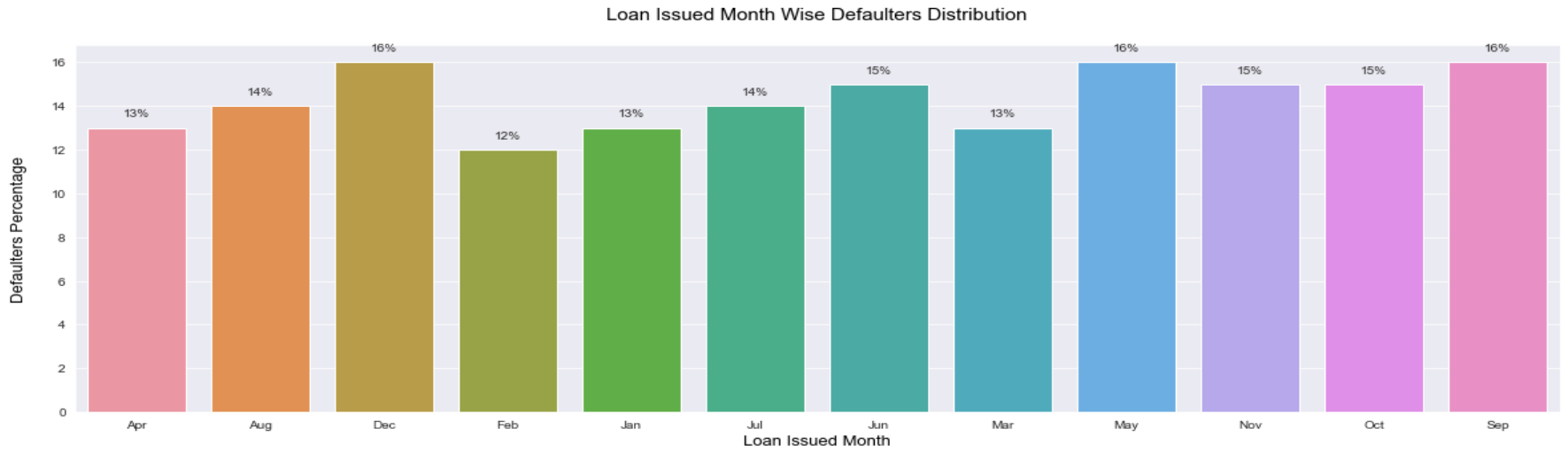
- Majority of employees applying for the loan have more than 10 years of experience
- Tendency of person to default the loan with 10 years of experience is also high. So, company need to be careful when granting loan.

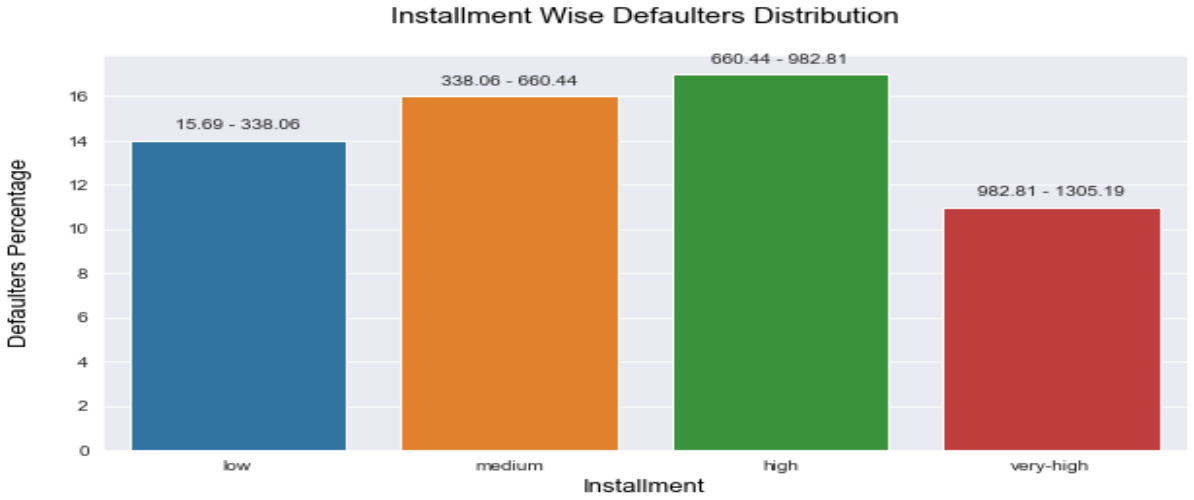
Loan Amount vs EMPLOYEE EXP

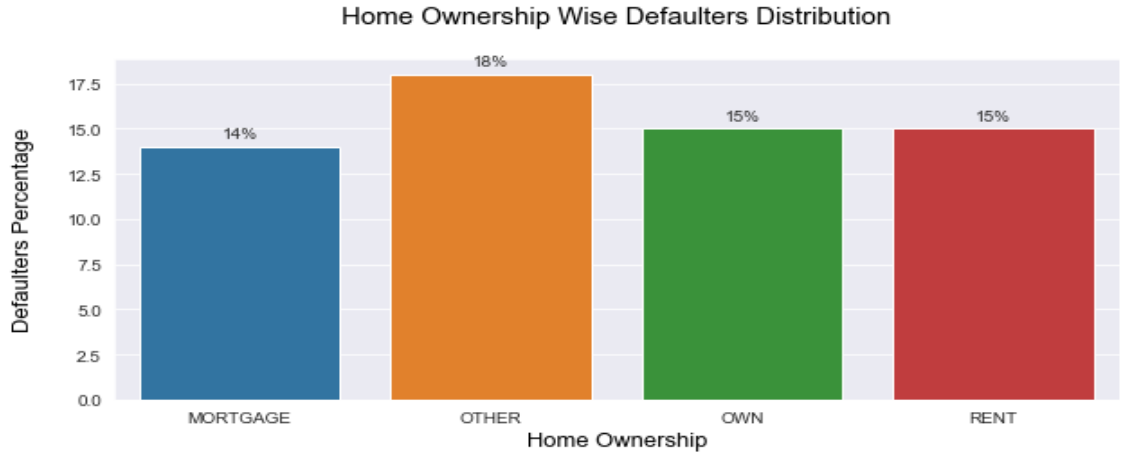
- By doing bivariate analysis it has been observed that the employee exp. With 10+ years is taking more amount of loan
- Those who are less than 1 year is taken less amount of loan

Feature Which has No major Impact on Defaulter

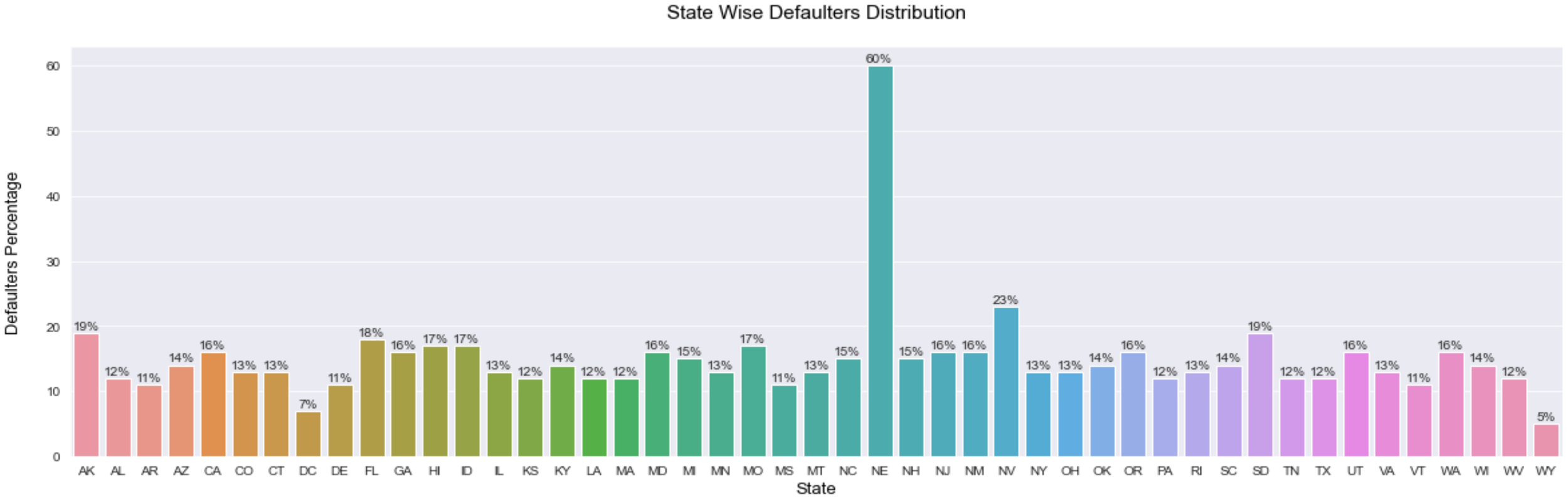
- Home Ownership
- Loan Issued Month
- Installments





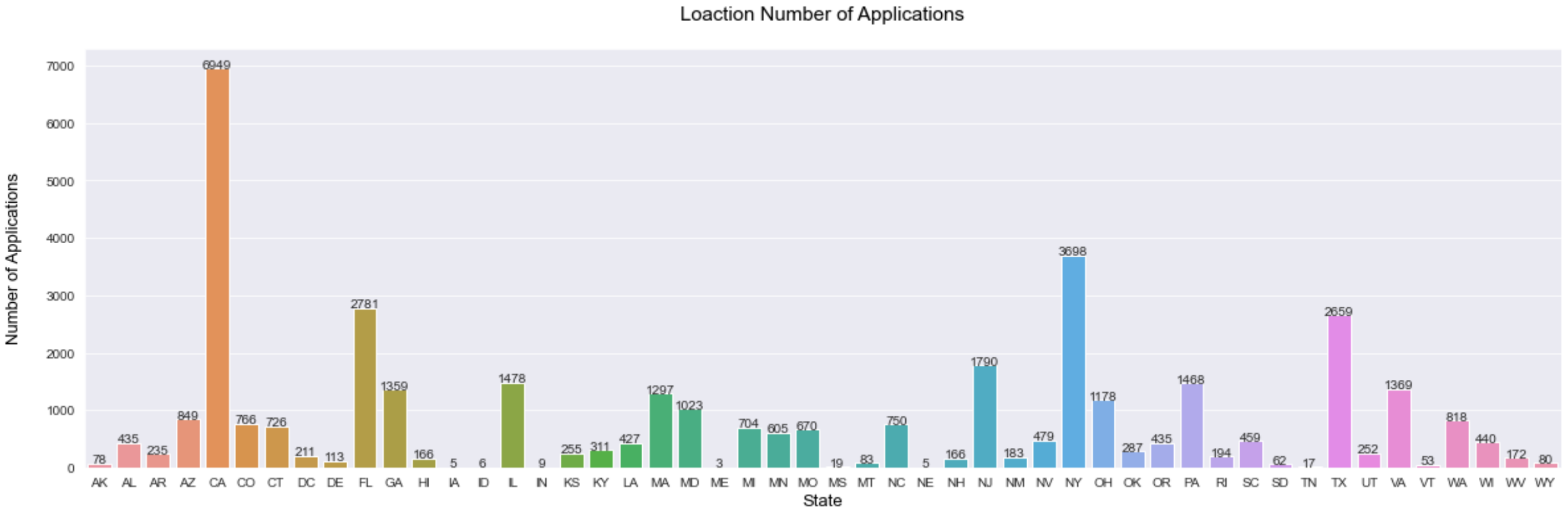


Effect of Location on Defaulters



- NE state has highest numbers of defaulters which is around 60%
- Followed by NV which has 23% and SD and AK has 19% of defaulters

Loan Given State Wise



- The loans which are given for CA, NY, FL,TX state are more than any other state.

Loan amount vs interest rate for Loan Status

- Even if the loan amount is low but the interest rate is too high then the count of charged off customers are more than fully paid.
- For loan amounts above 25000, approved applications for loans have decreased significantly.
- For loan amounts greater than 25000 and interest rate above 15% people mostly charged off than fully paid .
- To reduce defaulter high interest rate on lower loan amount should be avoid

Correlation Plot

- pub_rec_bankruptcies and pub_rec are highly correlated
- annual_inc and loan amount are correlated

Conclusions

- Grades and Sub-Grade are one of the good metrics for finding loan defaulters. Lending club has to look into the information from borrowers before issuing loans to low grade like grade F & G.
- Lending club should reduce the high interest rate for 60 months tenure, there are more chances for loan to get default.
- Lending Club before approving loan should take a look at the DTI (Debt to Income ratio) of the applicants. As DTI of borrowers increases their tendency of defaults also increases.
- Lending club should be more conscious when approving loans for applicants in the lower annual income segment, borrowers having annual income in range of 4k-50k has more of defaults.
- When the loan amount higher the chance of applicants is defaulters. Lending club should properly analyze the applicants demanding higher loan.
- Lending club should be more cautious when approving loan for the purpose of Small business, Debt consolidation ,Credit card and Education.
- Loans having higher interest rate have more defaulters. Lending Club should check the background of applicant thoroughly if interest rate is high.
- Lending Club should take extra security while approving applicants belonging to NE, NV, SD and AK state, as tendency to default is high.
- If applicant employment length is >10 years Lending Club should thoroughly check applicants background as there are many defaulters in >10 years as an employment length.

THANK YOU