Lending Club Case Study

PREPARED BY

SANJEEV SURENDRAN

Case Study Objective

- Understand and Analyse the data set to understand why an loan applicant was charged-off.

Data Analysis

- Using Exploratory Data Analysis (EDA), find out the factors that strongly indicate whether an loan applicant will default or not.

Driving Factors

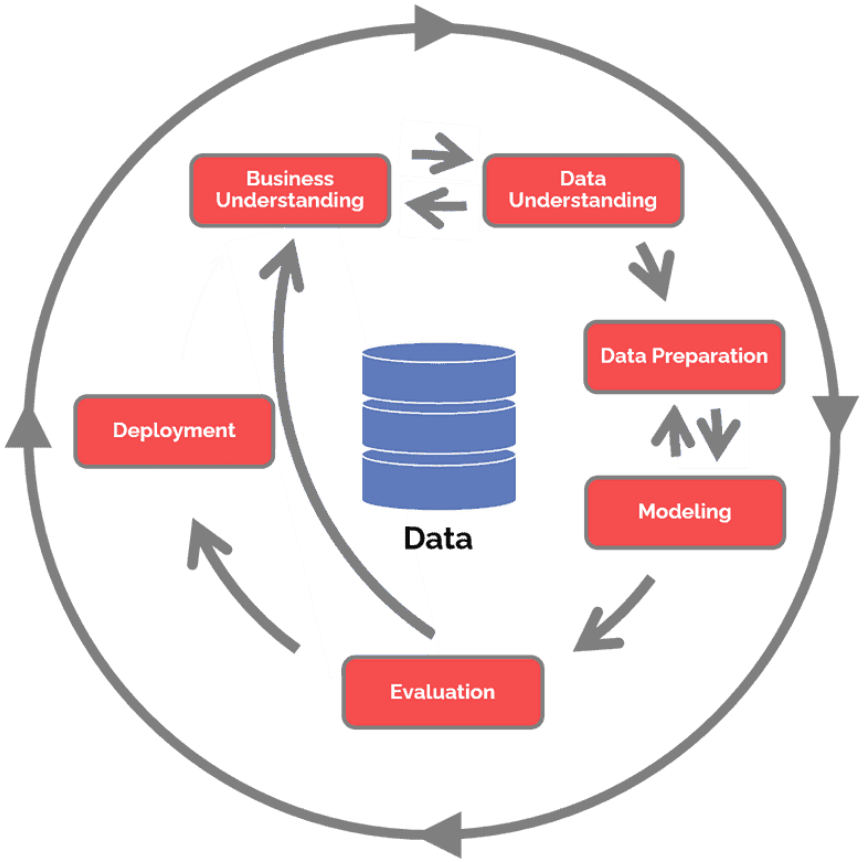
- Present a solution using driving factors, which can be used by Lending Club to assess the risk of loan applicants and minimize financial and business loss.

Business Solution

Solving Business Problem using CRISP-DM

The **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (*CRISP-DM*) is a process model to solve analytics problem. It has six sequential phases:

- **Business understanding** – Identify Business Objective
- **Data understanding** – Analyze data set to summarize and verify its quality
- **Data preparation** – Clean, Construct and Format the data to perform EDA
- **Modeling** – Generate a mathematical model and solve it by using clustering
- **Evaluation** – Evaluate model that best meets the business objectives
- **Deployment** – Deploy the model to Stakeholders and continue the cycle



Business Understanding

Create a Business Model for a Consumer Finance Company to predict the risk of loan approval to applicants.

Deployed model should be able to avoid two types of risks:

- **Avoid business loss** by not allowing loan approval to applicants, who are likely to repay the loan.
- **Avoid financial loss** by clearly identifying the loan applicants, who are likely to Charge-Off (Default) – doesn't repay the loan.

Understand the driving factors (or driver variables) behind loan default. Consumer Finance Company can utilize this knowledge for its portfolio and risk assessment.

Data Understanding

Analysis

- Acquired and loaded the comma separated data (CSV) data set to analyze its structure.
- The data set have 39717 data entries.
- There are 111 columns in data set and description of each column is briefly explained in the available data dictionary.
- There are 74 float type variables (Floating point numbers), 13 integer type variables (Integer numbers) and 24 object type variables (Text or String data).

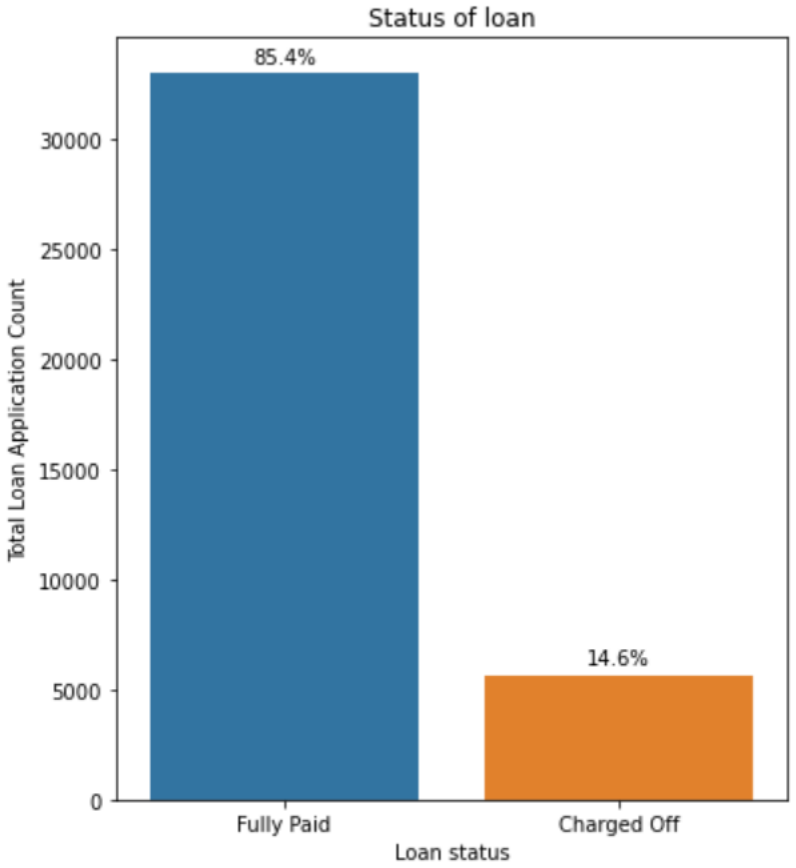
Observation

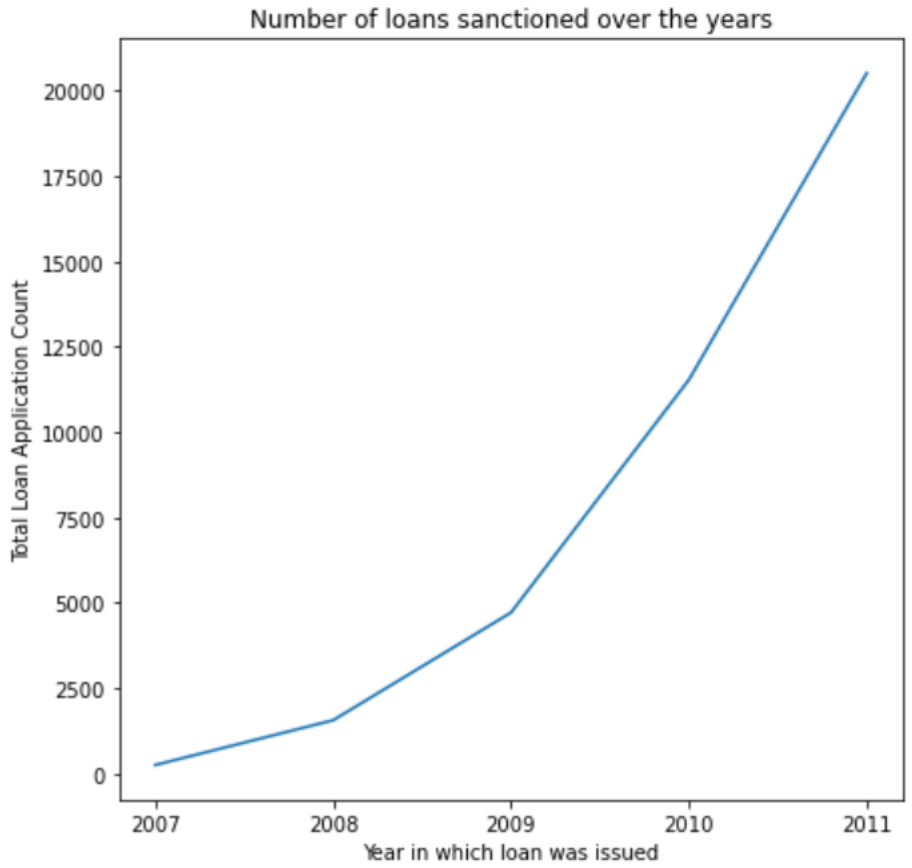
- Data set is complete and there are no columns without name.
- No merged columns.
- Date is represented in Month-Year
- Unique id given to each loan application
- Some numerical columns have texts and symbols
- Some columns have NULL values in it and some have entire index range as NULL values
- Mismatch in data type

Data Preparation

- **Data Deletion** – Deleted the columns in data set, that have entire index range as NULL values. After deletion, there are only 57 columns.
- **Data Filtering** – Analyzed the column entries using data dictionary and deleted the columns that have either same value in entire index range or it is not important for the outcome in finding solution to Business Objective. After deletion, there are only 35 columns.
- **Data Cleaning** – Removed symbols and texts in numerical columns.
- **Data Formatting** – Changed data types of columns.
- **Data Entry Removal** – Removed all data in data sets that have Loan Status as 'Current' because we can't predict whether these loan applicants will default or not.
- **Data Imputation** – Percentage of missing data is less. Imputed NULL values in Public Record of Bankruptcies as 'Status Unknown'.
- **Data Derivation** – New data was derived from existing columns
  1) Created bins of Interest Rate, Debt to Income (DTI), Loan Amount and Annual Income.
  2) Separated month and year from date columns
  3) Created a column that indicates Non-Default (0 – Fully Paid) Default (1 – Charged-Off) from Loan State

EDA – Loan Status

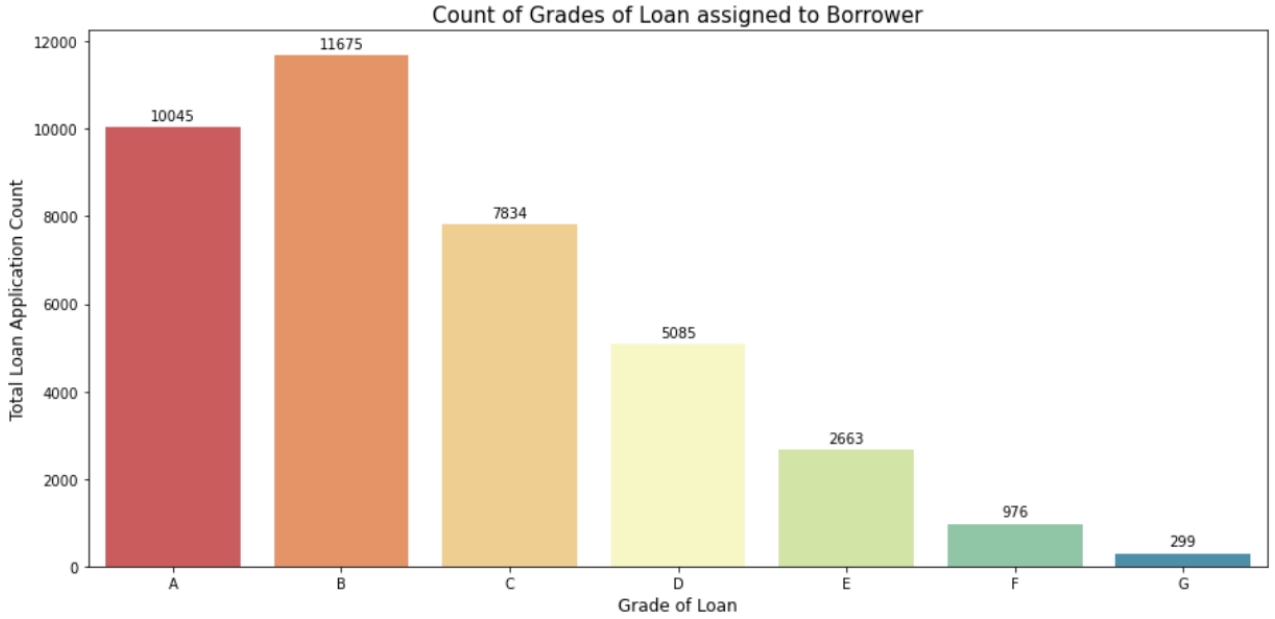


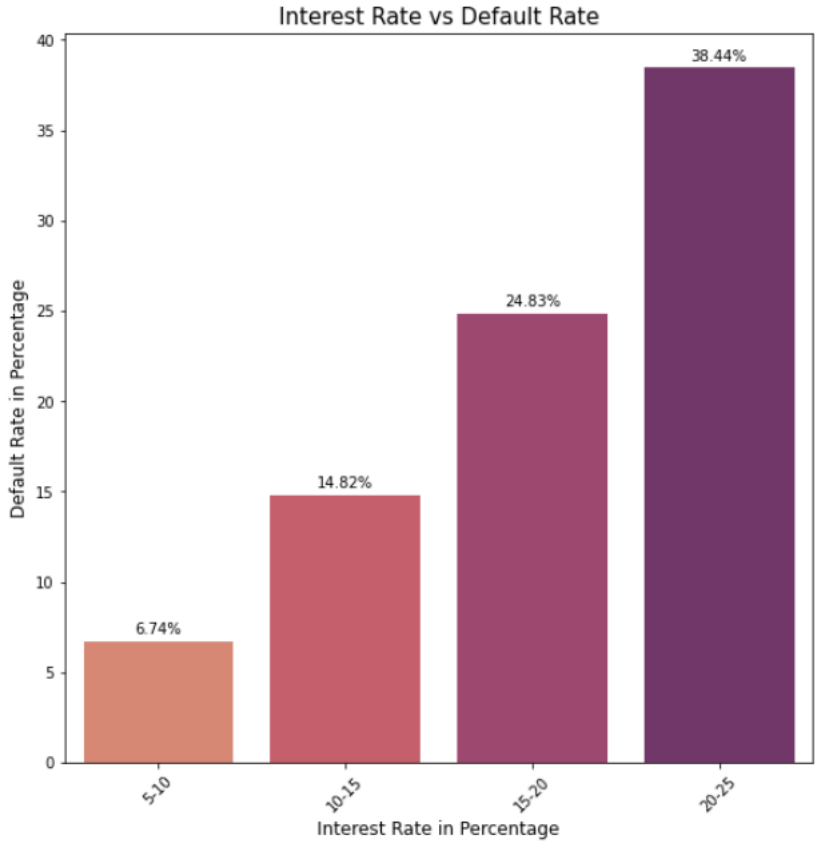


Observation

- Most of the loans applied by applicants are fully paid.
- About 14.6% of applicants were defaulted.
- Number of loans approved increases exponentially over the years.

EDA – Loan Grade and Interest Rate Charge-Off





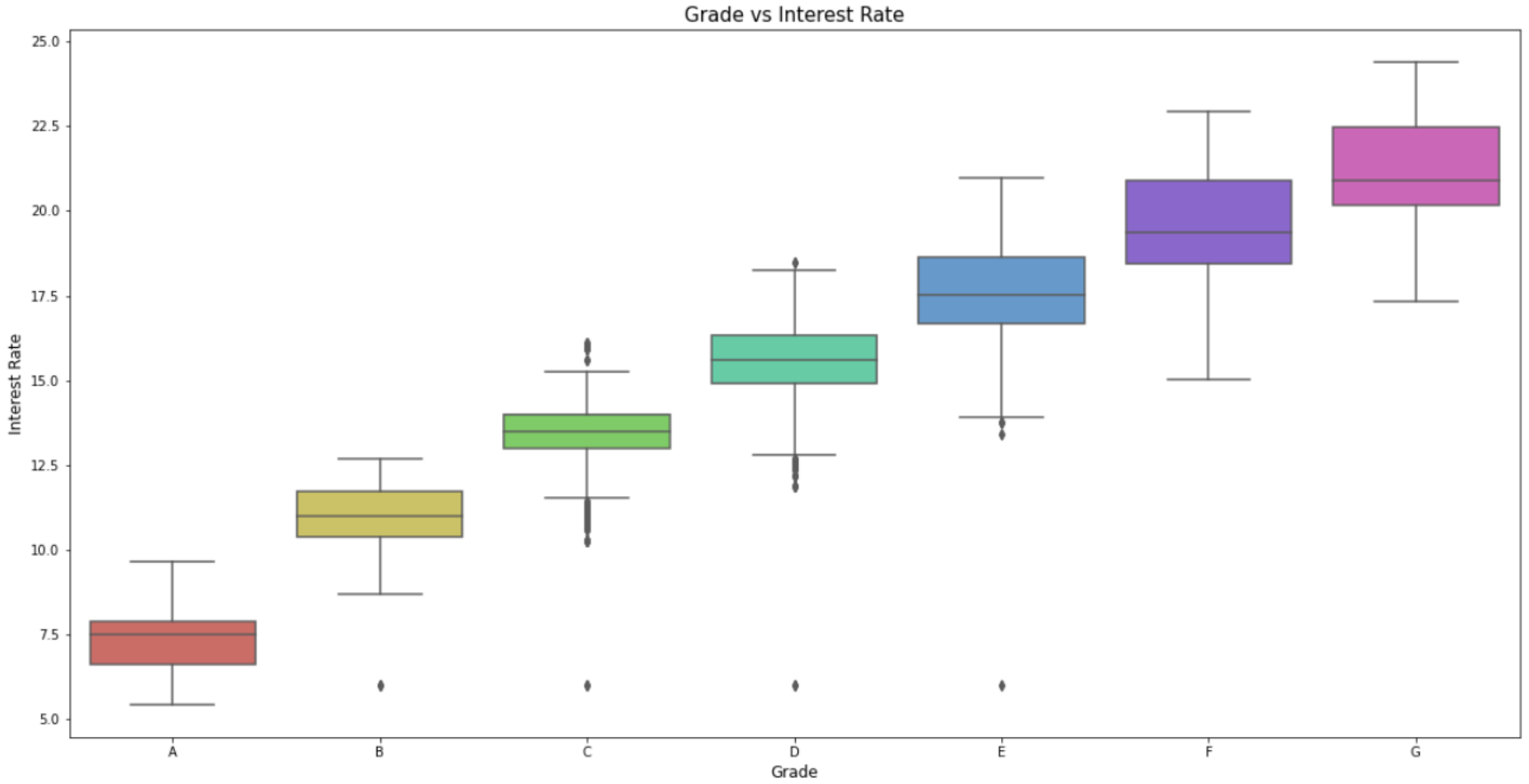
Observation

- Majority of loans belongs to grades A and B.
- When interest rate increases, there is more chance that applicants will default.

EDA – Grade vs Interest Rate

Observation

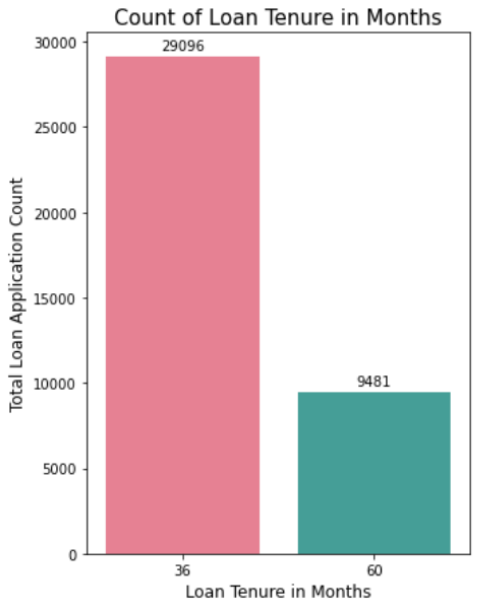
- Interest Rate increases from grades A to G.
- This implies that the interest rate is more for grade G loan.

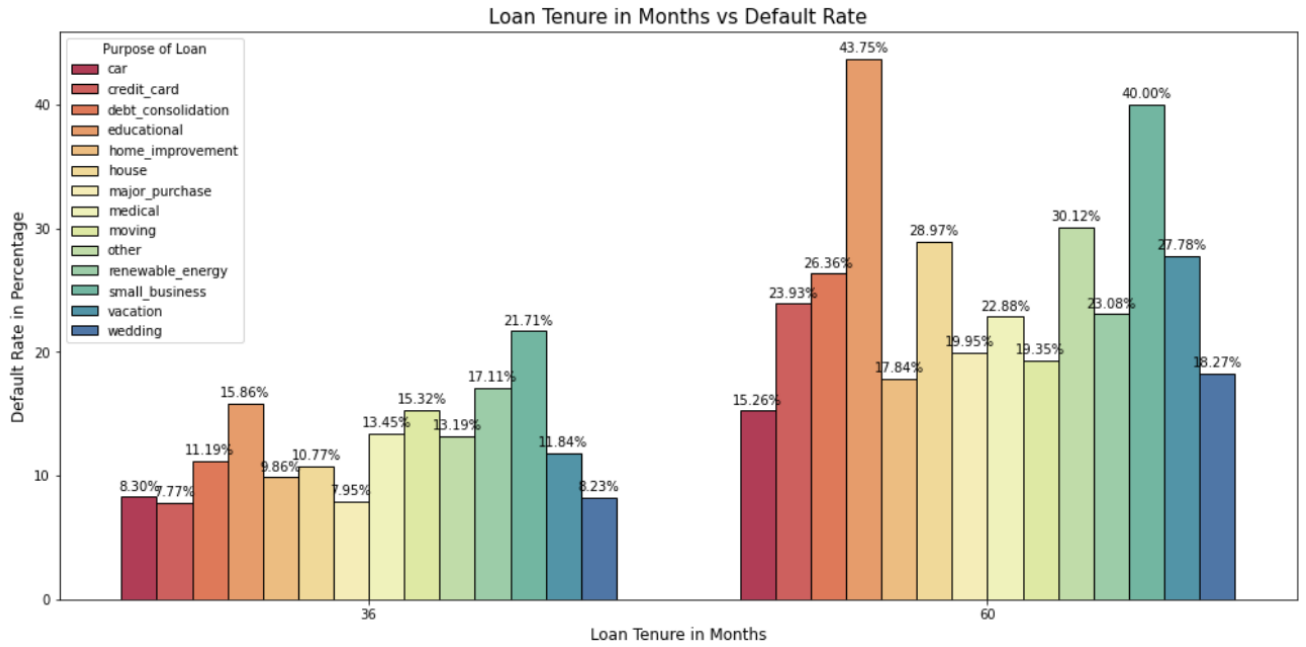


EDA – Loan Tenure and Purpose

Observation

- Majority of loans have tenure of 36 months.
- Loans having tenure of 60 months have high tendency to default.
- Loans applied for Small business having either 36 or 60 months of tenure have high chances to default.

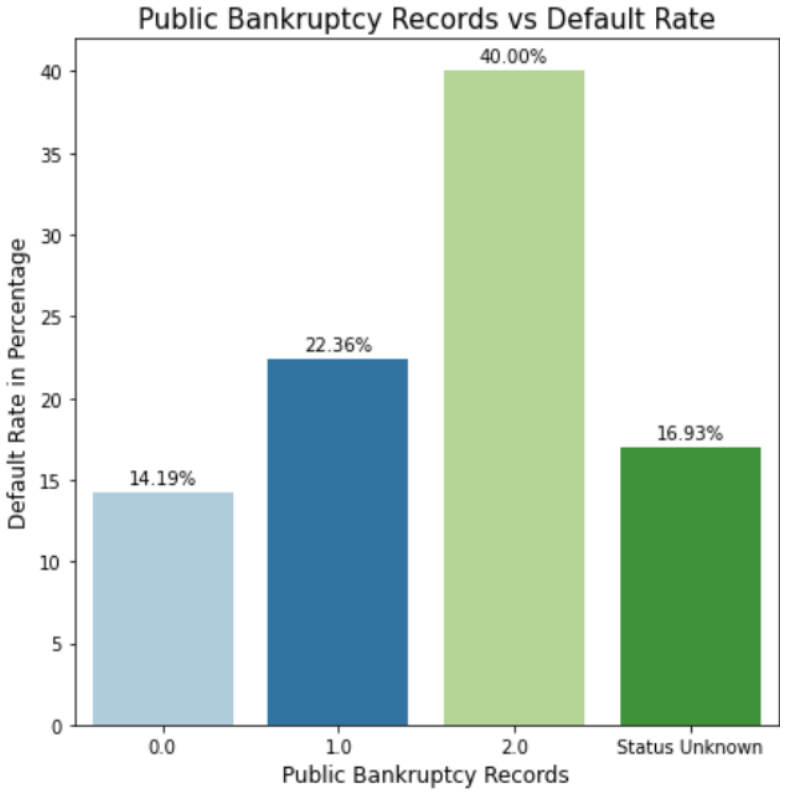


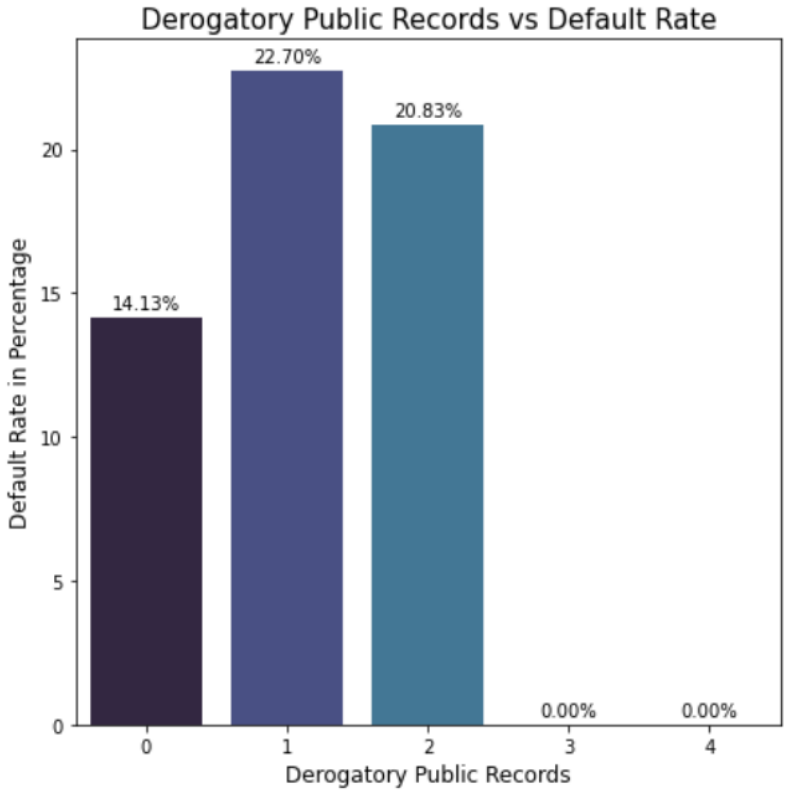


EDA – Bankruptcies and Derogatory Records

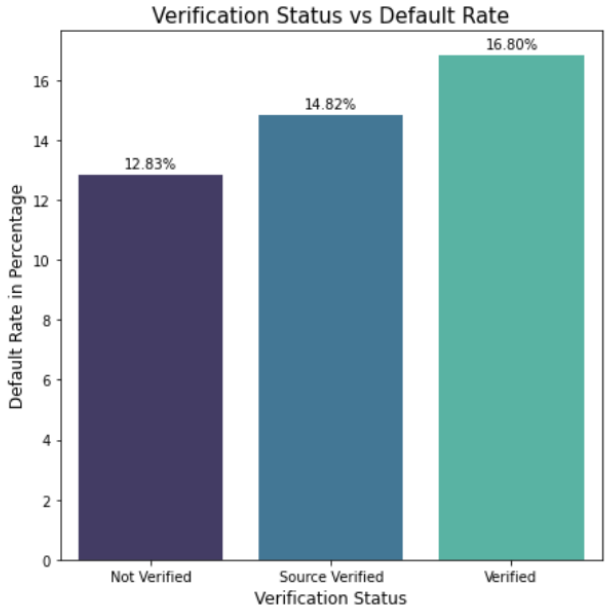
Observation

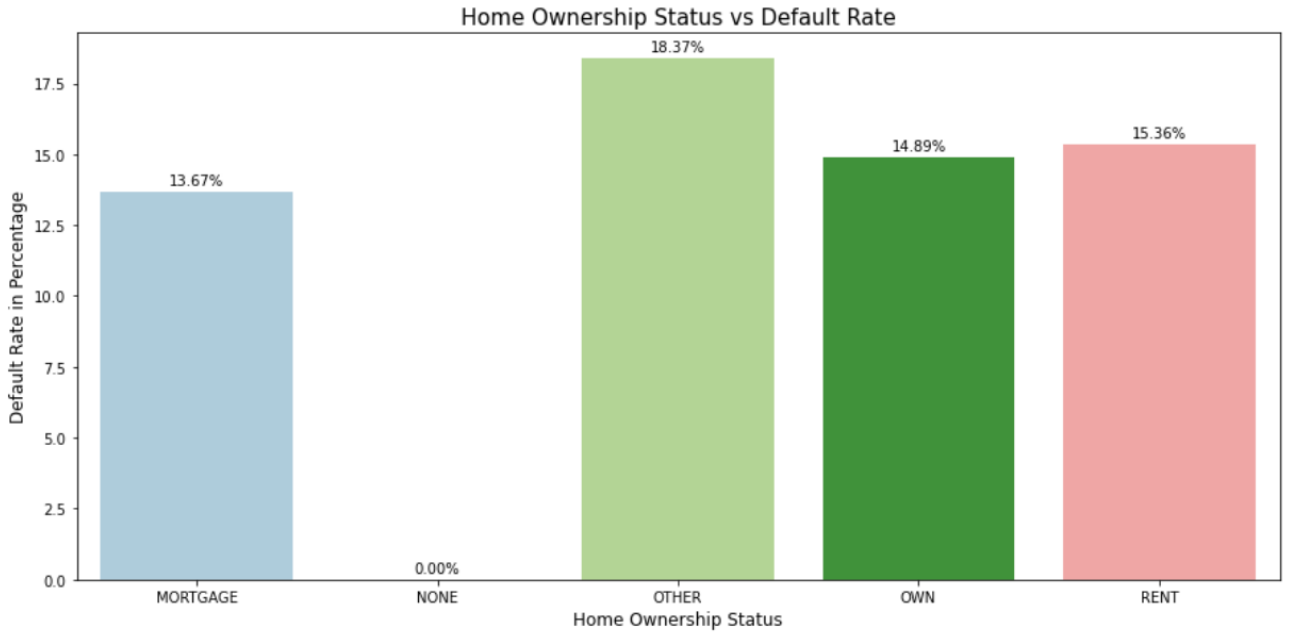
- As per Bankruptcy Public Record, majority of the loans are defaulted by applicants have past history of default
- Majority of the loans are defaulted by applicants have 1 or 2 Derogatory Public Record





EDA – Verified Source and Home Ownership





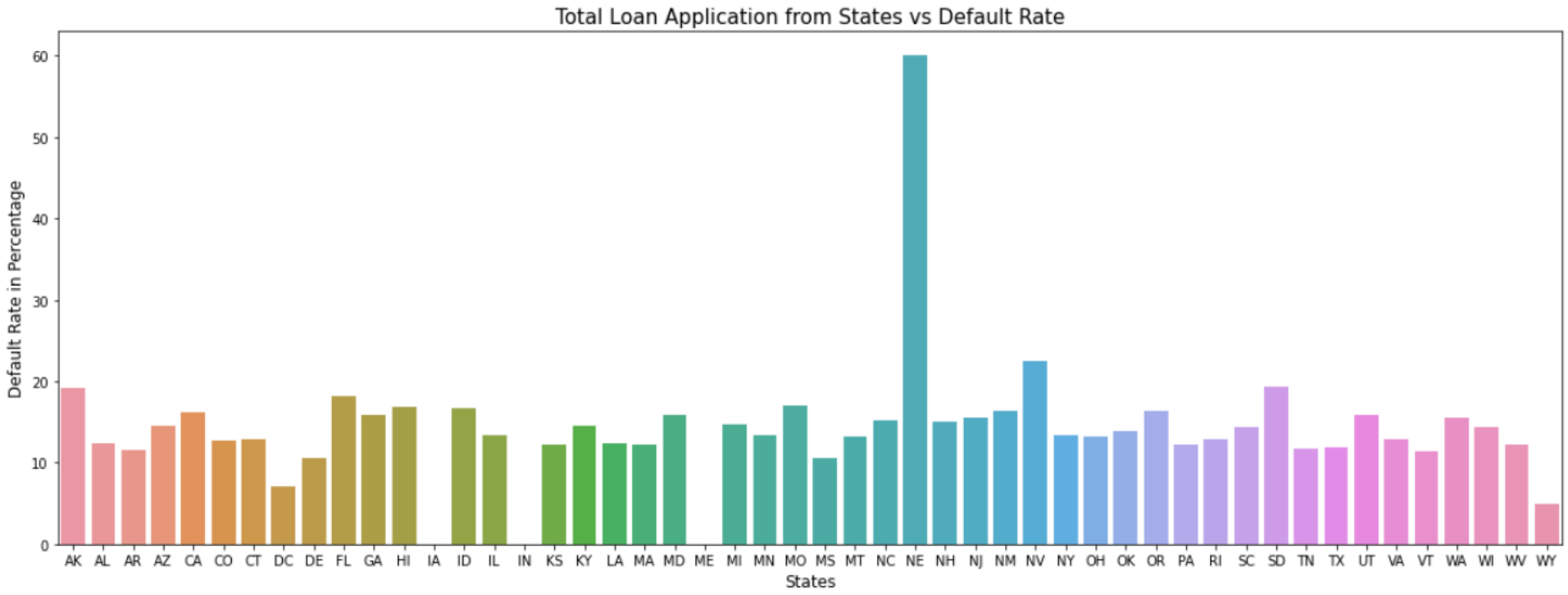
Observation

More defaulters are observed for applicant's who have their source of income verified and home ownership status is unknown.

EDA – Defaulters rate vs State

Observation

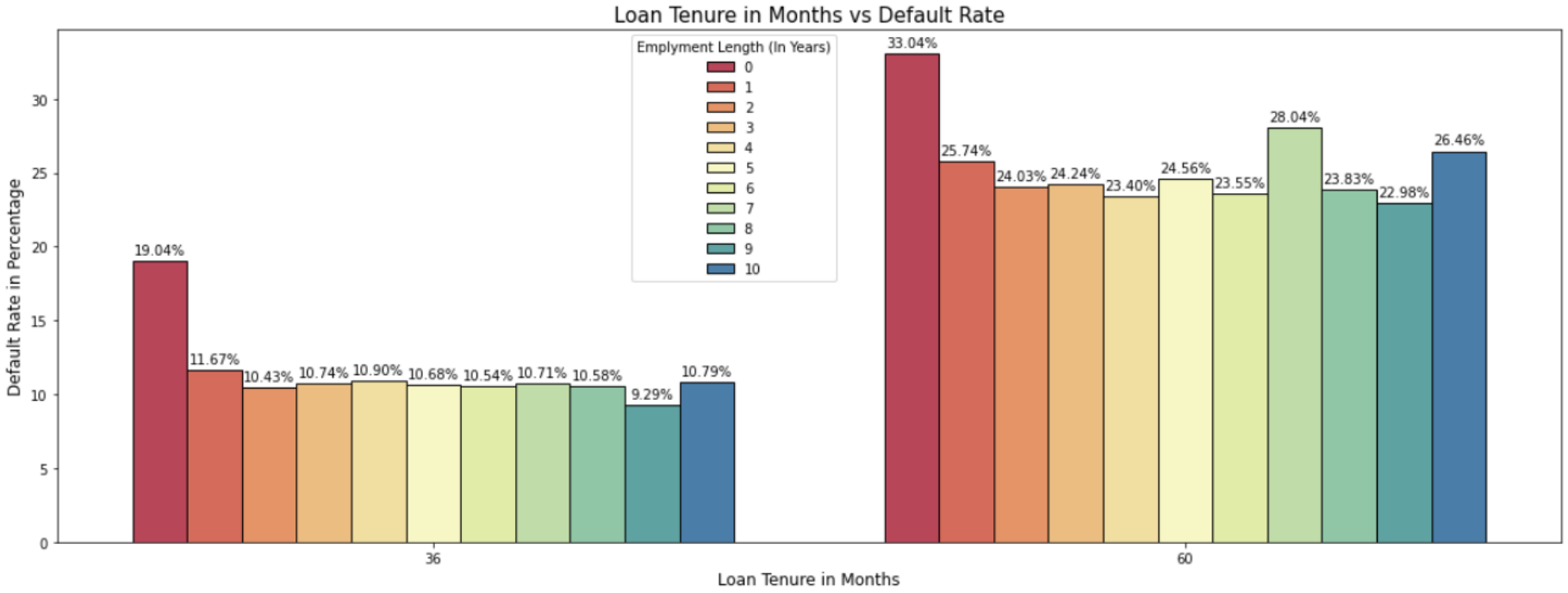
- Default Rate is higher in State NE because of low loan application count.
- States NV and CA have good number of loan applications and default rate is also high.



EDA – Employment Experience

Observation

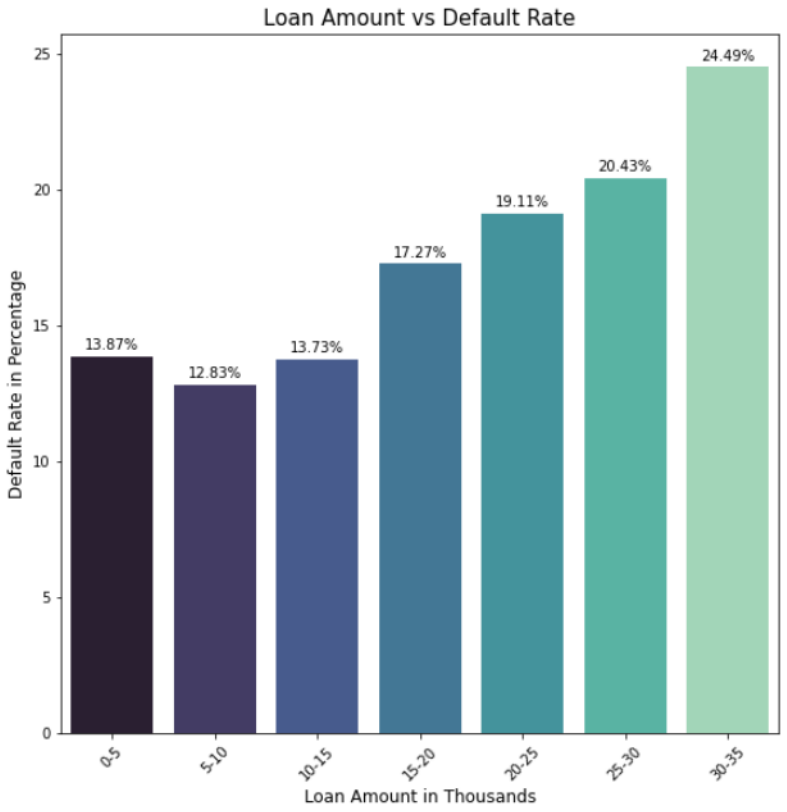
- Applicant's who have no job and less than one year experience have high tendency to default.
- It implies that they don't have proper source of income or annual income issue.

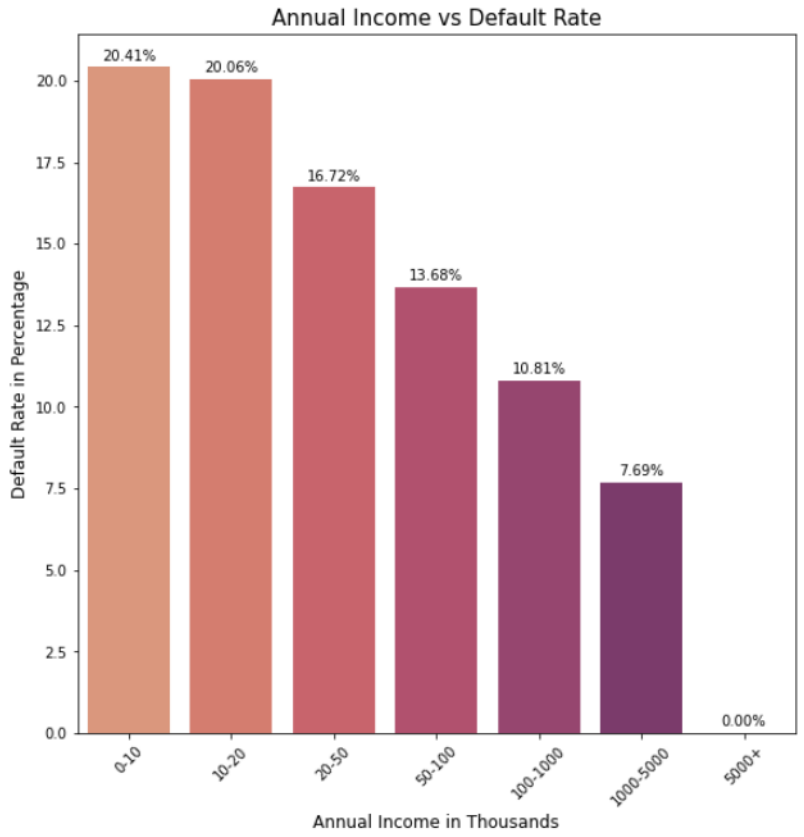


EDA - Loan Amount and Annual Income

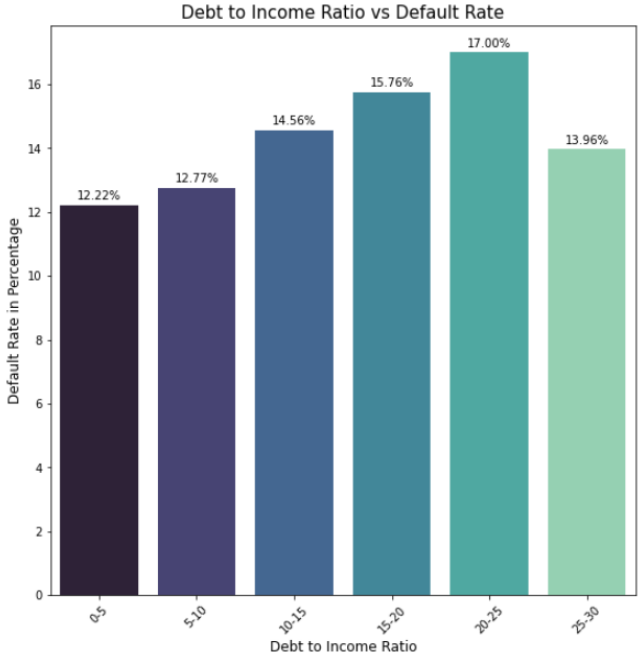
Observation

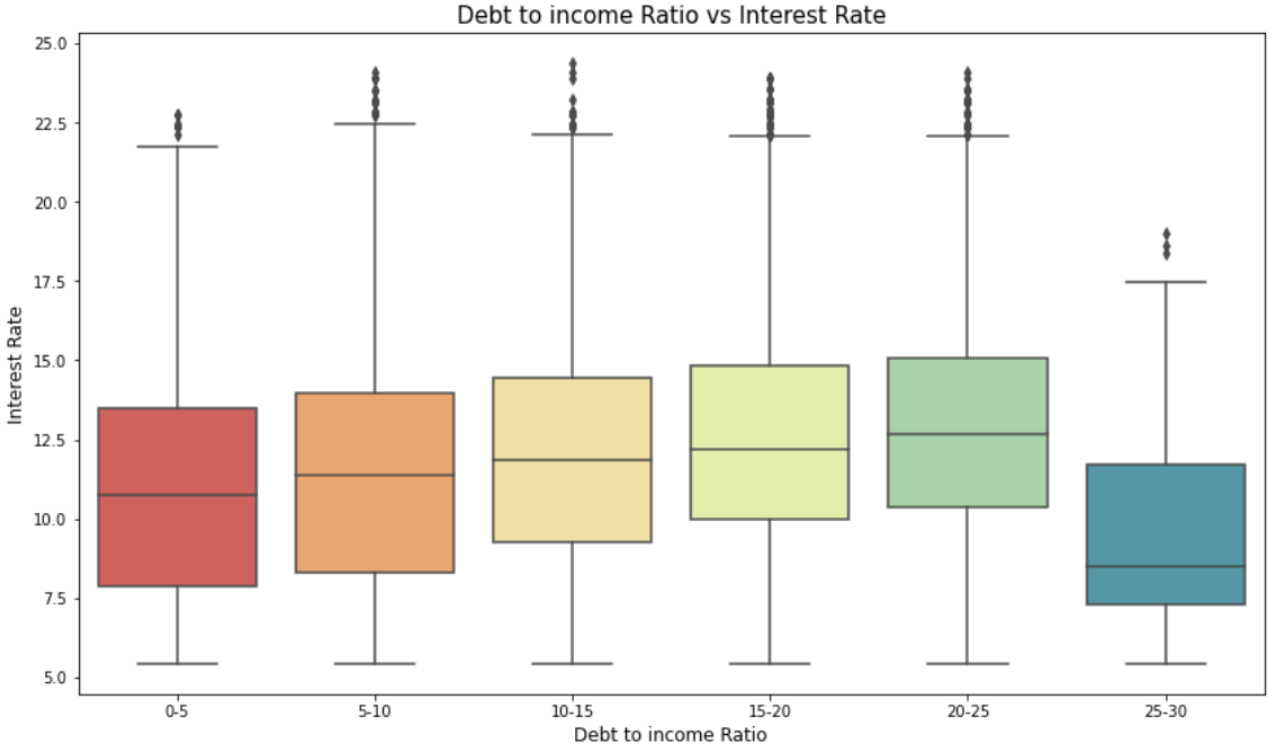
- Applicant's who have for large loan amount have high tendency to default.
- As the annual income increases, there is less chance of default. This clearly relates to employment length, since annual income is very less for employment length less than 1 year or no stable job.





EDA – Debt to Income (DTI) and Interest Rate





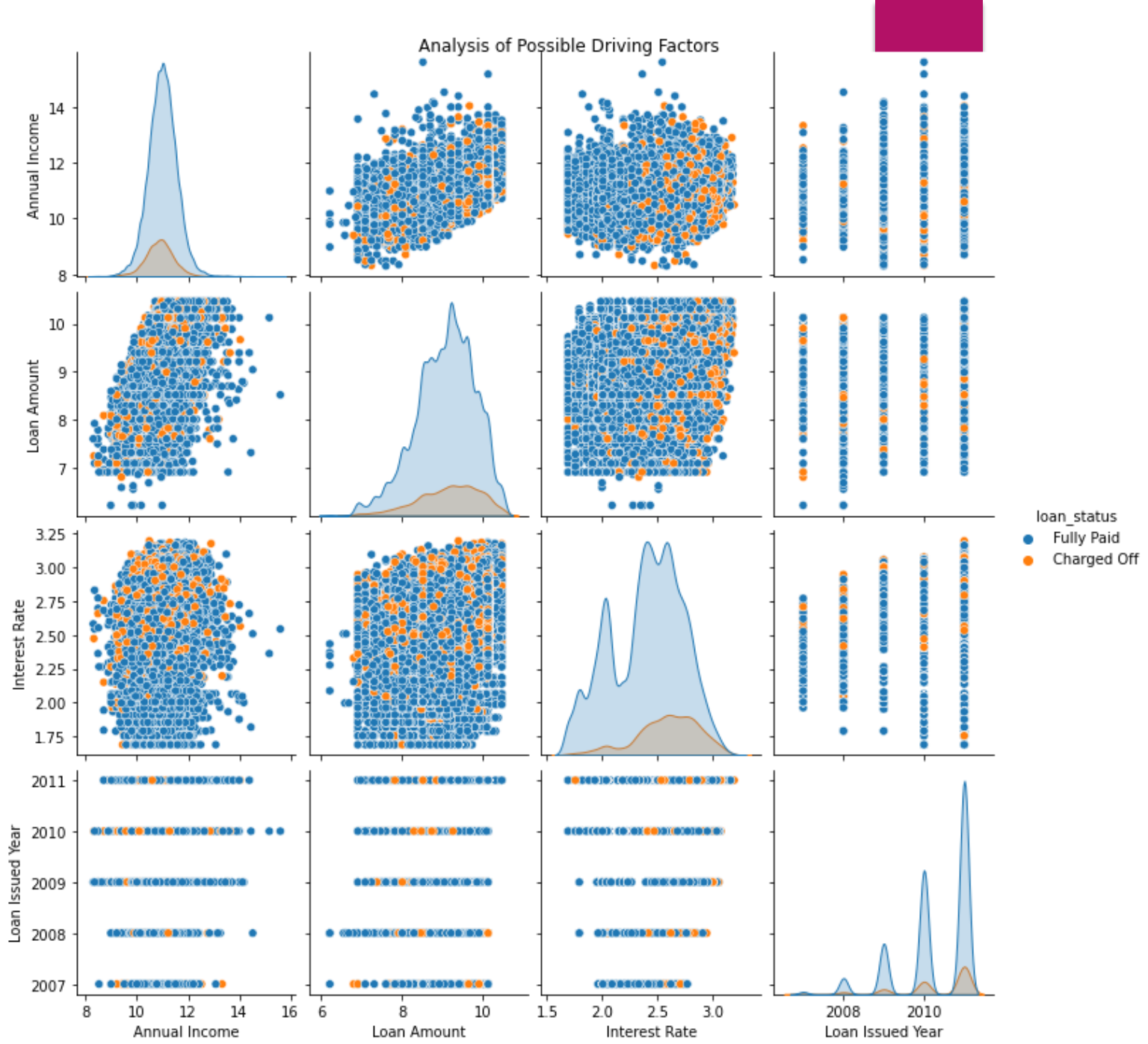
Observation

- More defaulters in DTI 20 – 25 % bucket
- Interest Rate is low for DTI 25 to 30 % bucket, which explains why defaulters are less in this range.

EDA – Multivariate Analysis

Observation

- Loan amount applied is higher with high annual income.
- Loan amount applied increases with high interest rate.
- More loans are charged off with increasing year.
- More loans were charged off when interest rate was high.



Conclusion and Recommendation

Based on Demography of Applicant

- Applicants who previously defaulted will default again.
- If Annual Income is more, there is very less chance to default.
- Extra scrutiny is required for Applicants belonging States NV and CA, since they have high tendency to default.
- Extra scrutiny is required if Applicants have employment length less than 1 year or no job history.

Based on Loan Characteristics

- Loans with high interest rate have high tendency to default.
- High tendency to default, if the loan amount is high and tenure is 60 months.
- Extra scrutiny is required if the purpose of loan applied is for small business.