Lending Club Case Study

PREPARED BY

SANJEEV SURENDRAN

Case Study Objective

Data Analysis

- Comprehend and Analyse the data set to understand why an loan applicant was charged-off (defaulted).

Driving Factors

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

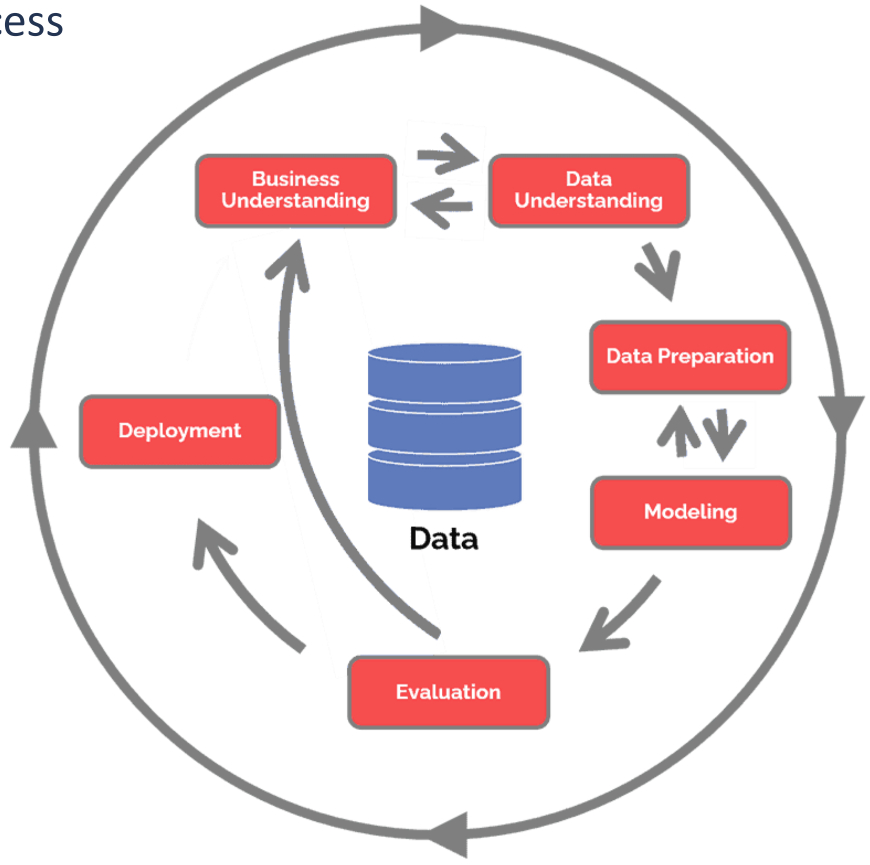
Business Solution

- 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.

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 Lending Club, 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. These Loan applicants will fall in category ***Non-Default***.

- Avoid **Financial Loss** by clearly identifying the loan applicants, who are likely to Charge-Off (Default) – doesn't repay the loan. These Loan applicants will fall in category ***Default***.

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.

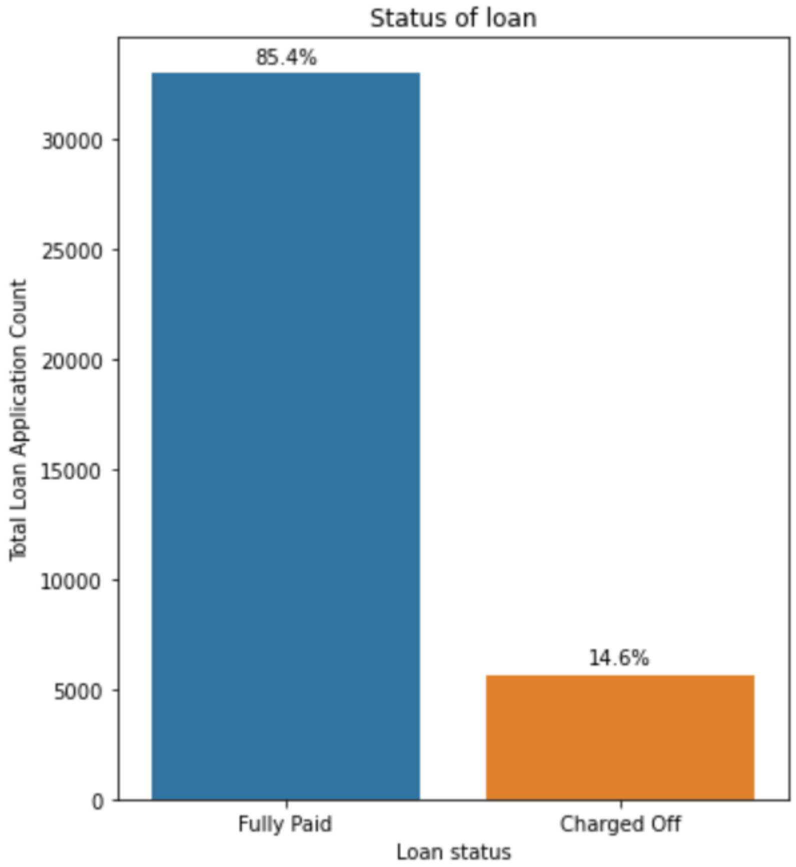
Observation

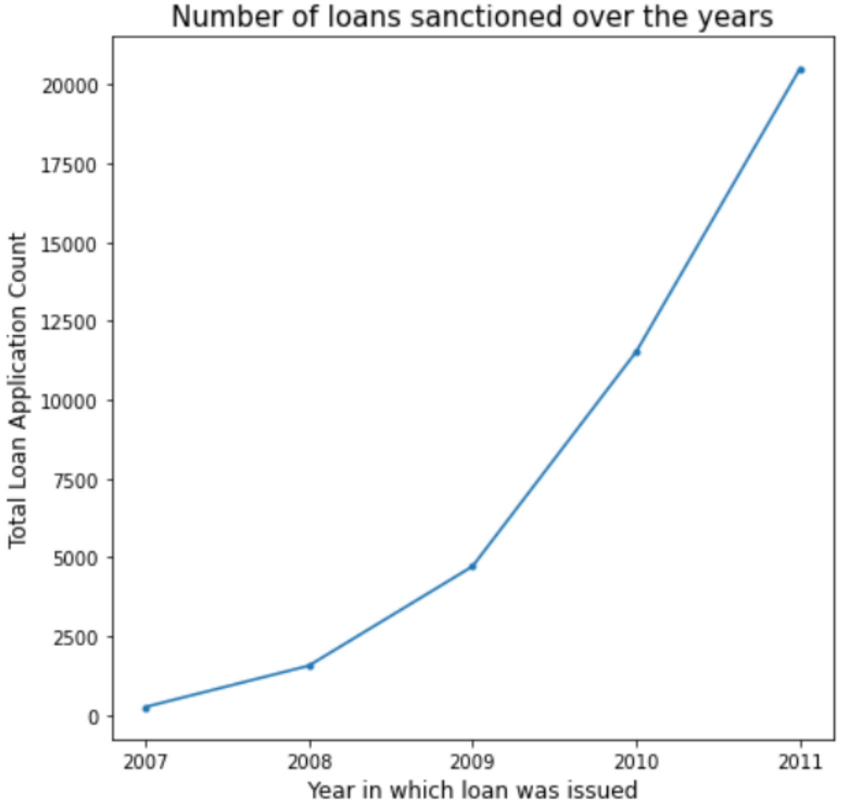
- Data set is complete and there are no columns without name.
- Date is represented in Month-Year.
- Unique id given to each loan application.
- Some numerical columns have texts and symbols.
- Some columns have NULL values in it.
- Mismatch in data type of columns.

Data Preparation

- **Data Deletion** – Deleted the columns in data set, that have entire index range as NULL values.
- **Data Filtering** – Analyzed the column entries using data dictionary and deleted the columns that is not important for finding solution for Business Objective.
- **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 (1.81 %) is very 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 loan issued date.
  3) Created a column that indicates Non-Default (0 – Fully Paid) Default (100 – Charged-Off) from Loan Status.

EDA – Loan Status

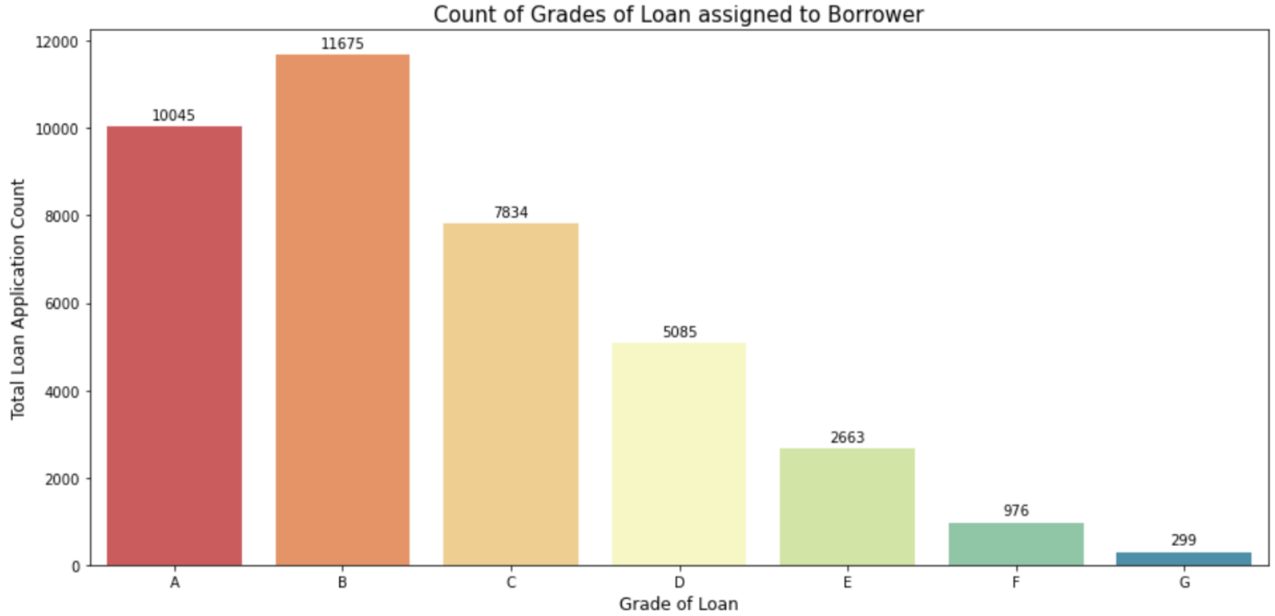


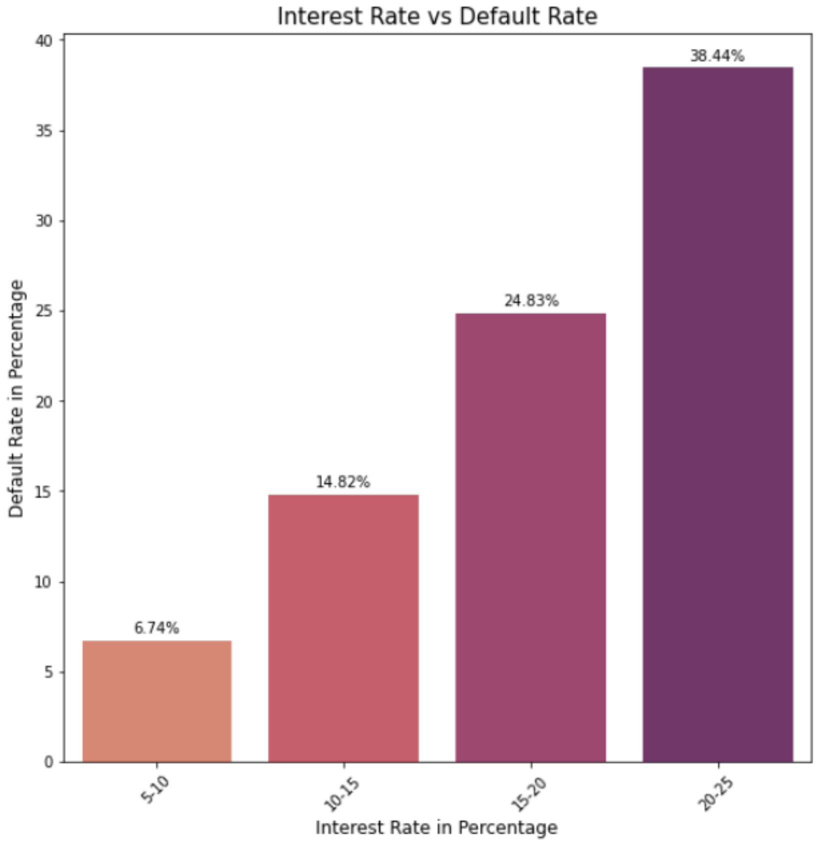


Observation

- **85.4 %** loans was fully paid by applicants.
- About **14.6 %** of applicants was defaulted (Charge Off).
- Number of loans approved, *increases exponentially* over the years.

EDA – Loan Grade and Interest Rate Charge-Off





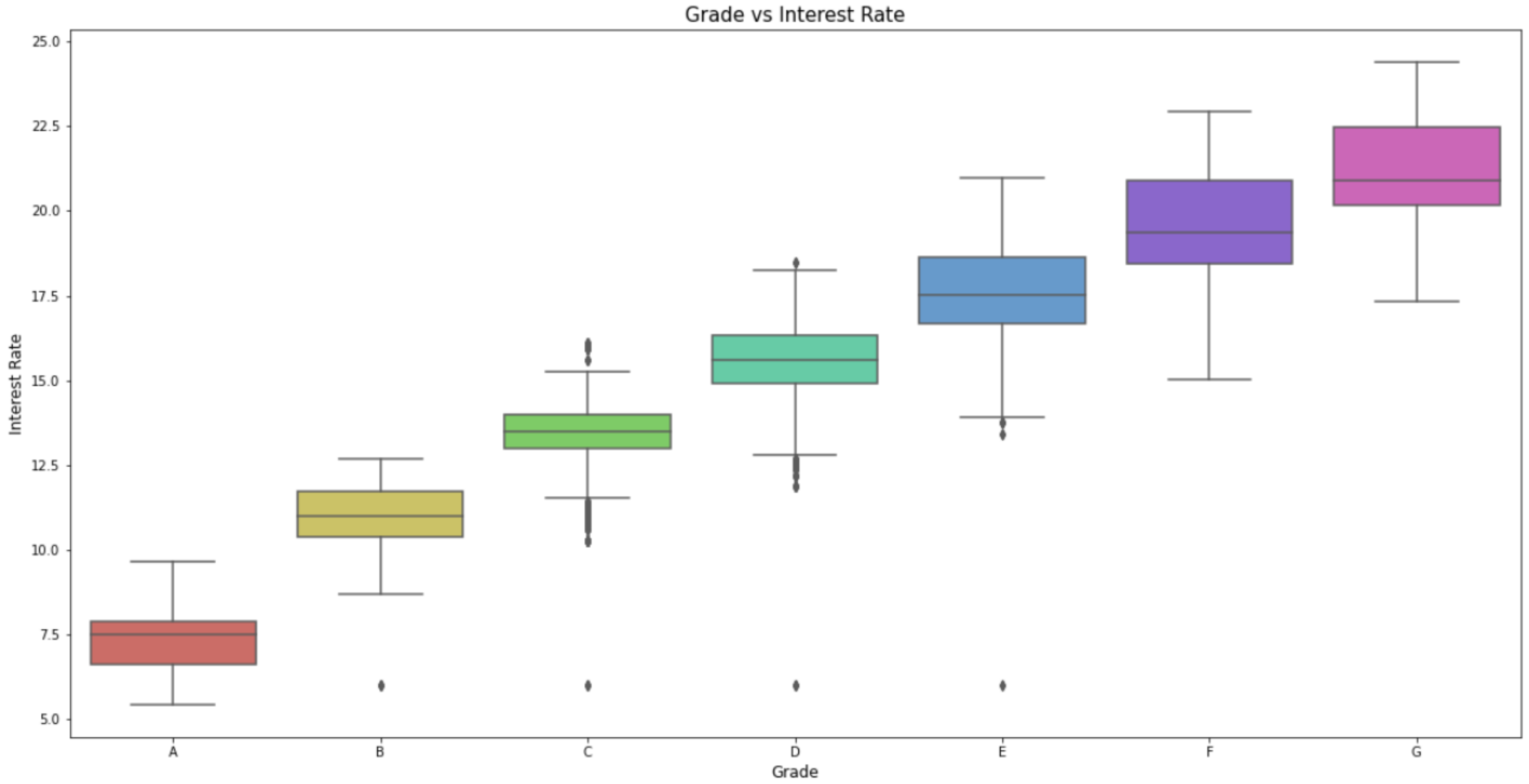
Observation

- 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

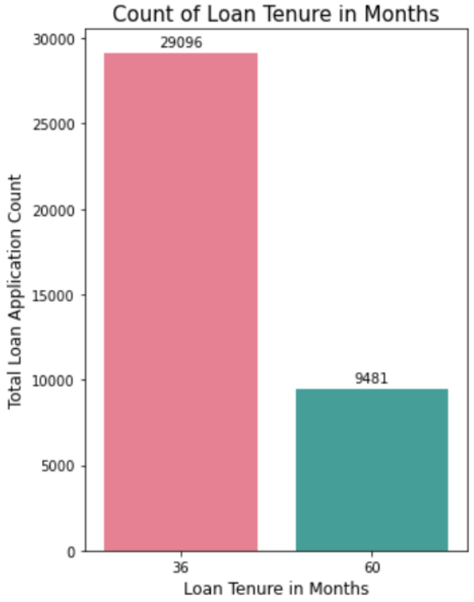
- Interest Rate increases from grades A to G.
- This implies that applicants with loan grade **A** are *less likely* to default, since interest rate is very low. And, applicants with loan grade **G** are *likely* to default.

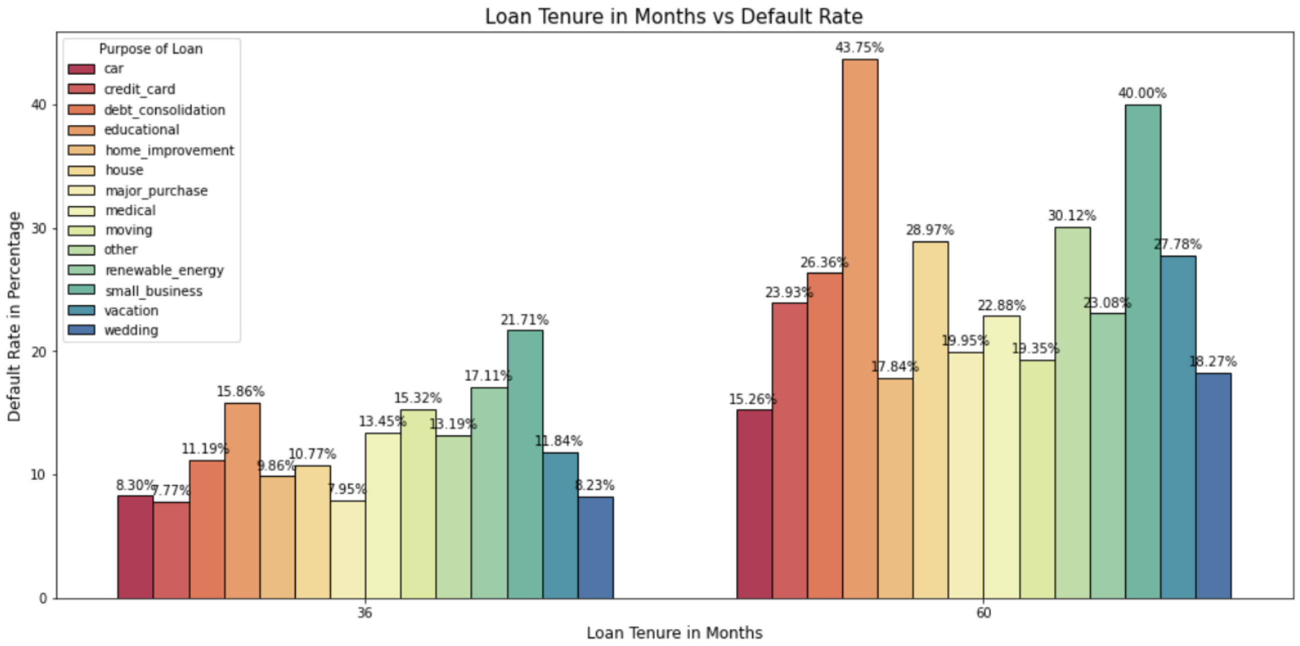


EDA – Loan Tenure and Purpose

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.





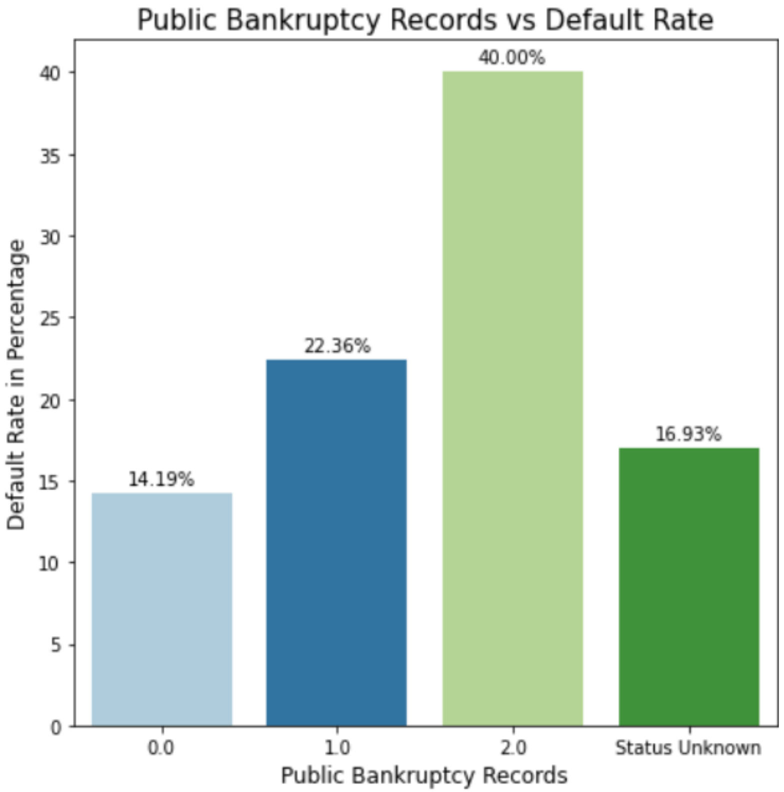
- After 'small business', applicants who applied loan for 'debt consolidation' is next in line for defaulting.
- Even though other purposes have more default percentage, loan application count is very less to conclude about defaulting.

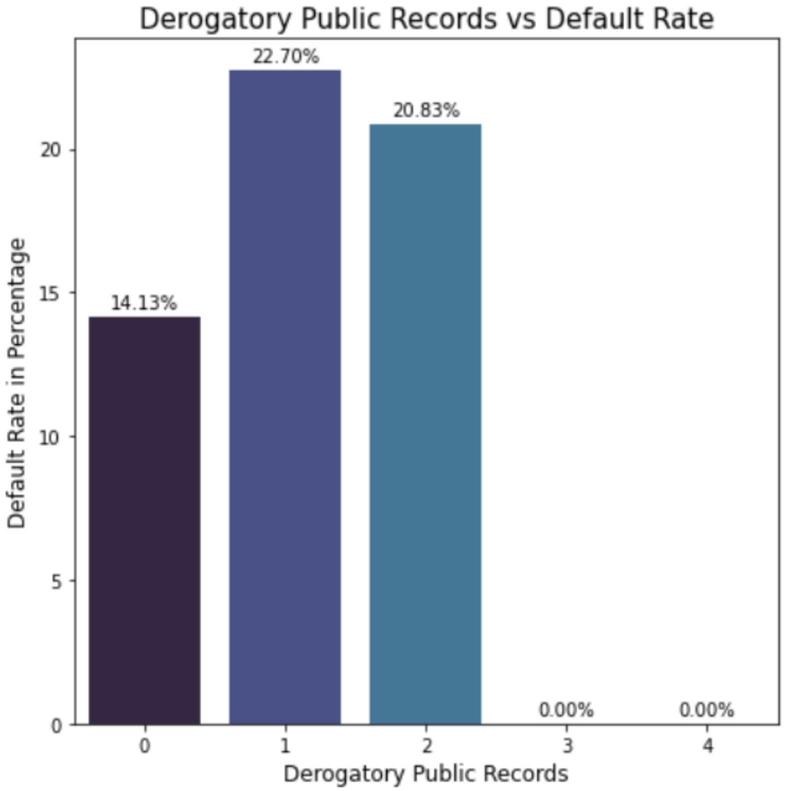
EDA – Bankruptcies and Derogatory Records

Observation

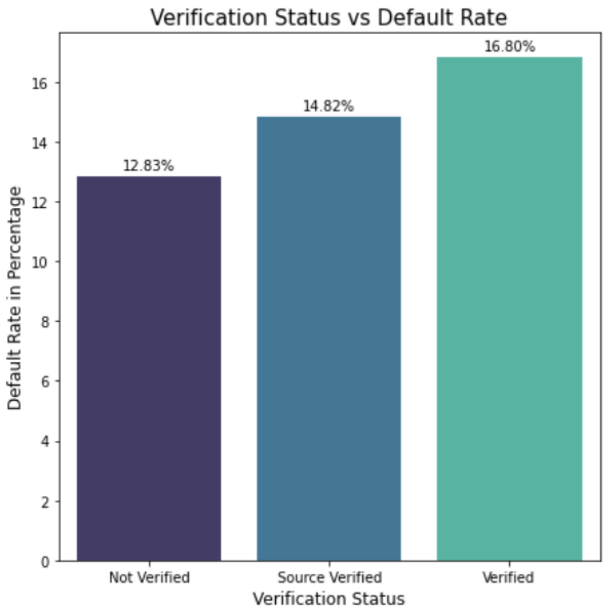
- Majority of the applicants who defaulted have past history of defaulting.
- Majority of the applicants who defaulted have 1 or 2 Derogatory Public Record. We can't predict defaulting since 3 and 4 record have no data.

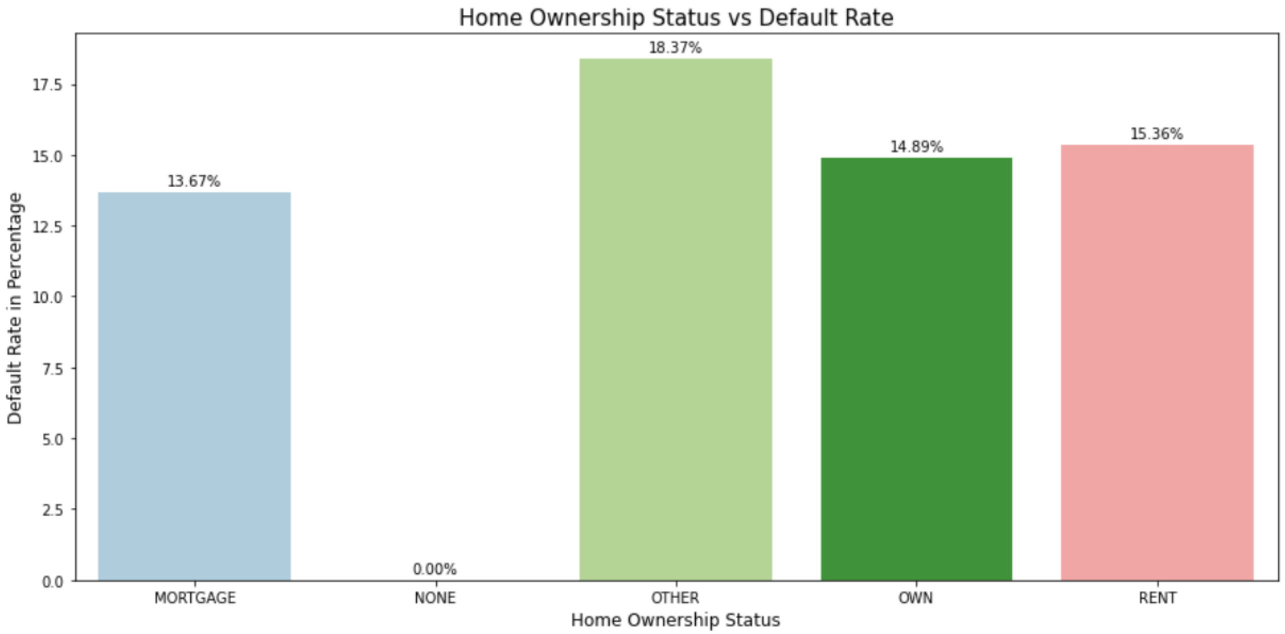
In conclusion, applicant's who have *defaulted previously*, will default in future as well.





EDA – Verified Income Source and Home Ownership



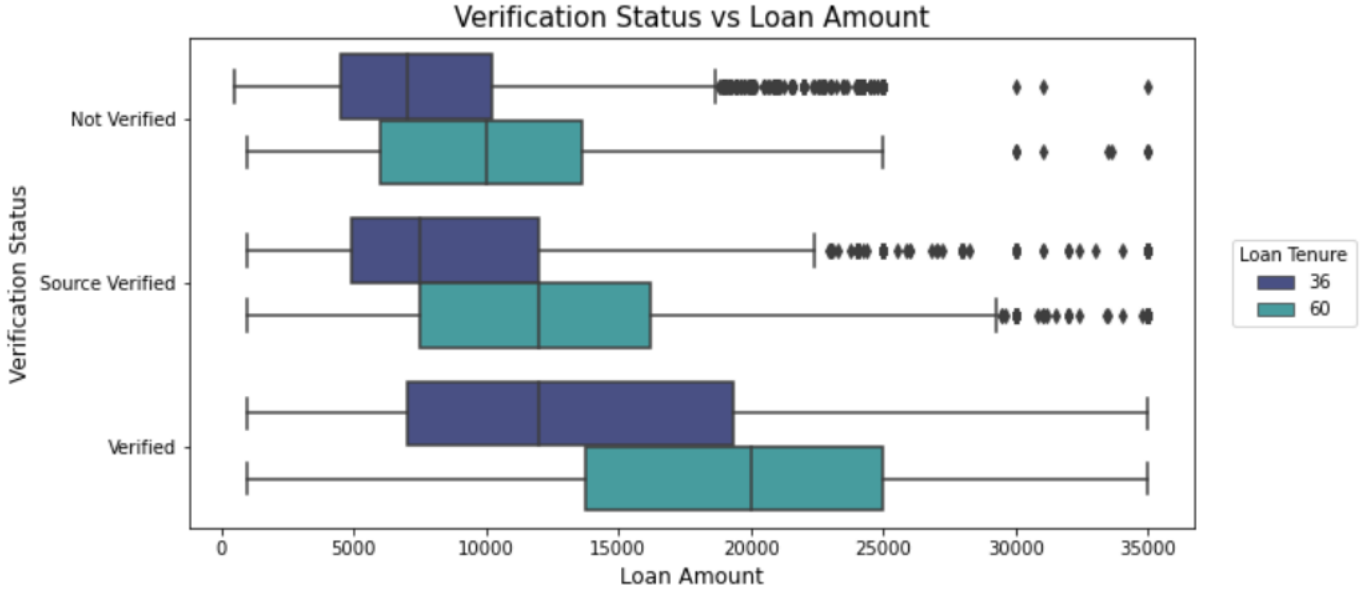


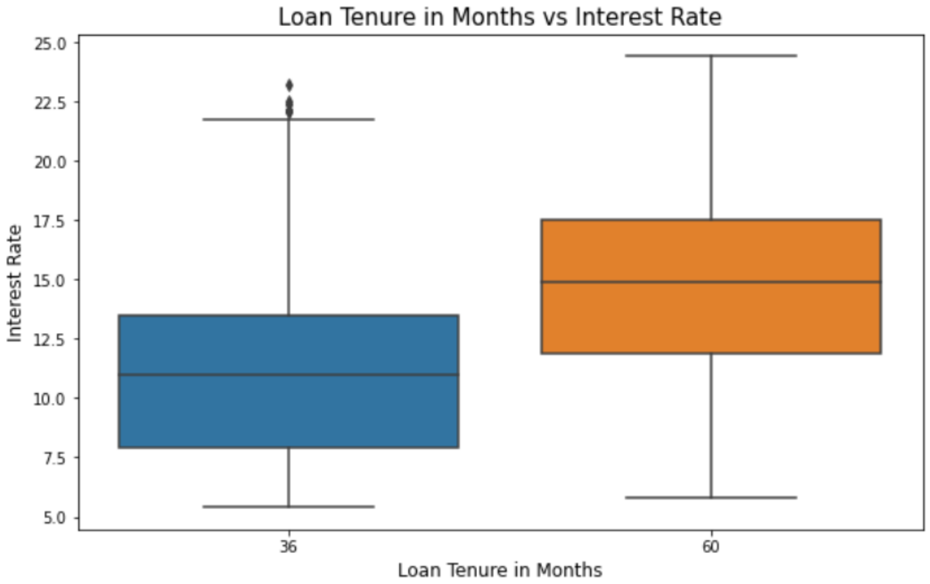
Observation

More defaulters are observed for applicant's who have their *source of income verified* by Lending Club and home ownership status is RENT.

- Home Ownership Status is not a good predictor of defaulting, since the loan application count is less. Also we can't further classify what status 'OTHER' means.

EDA - Verified Income Source and Loan Tenure





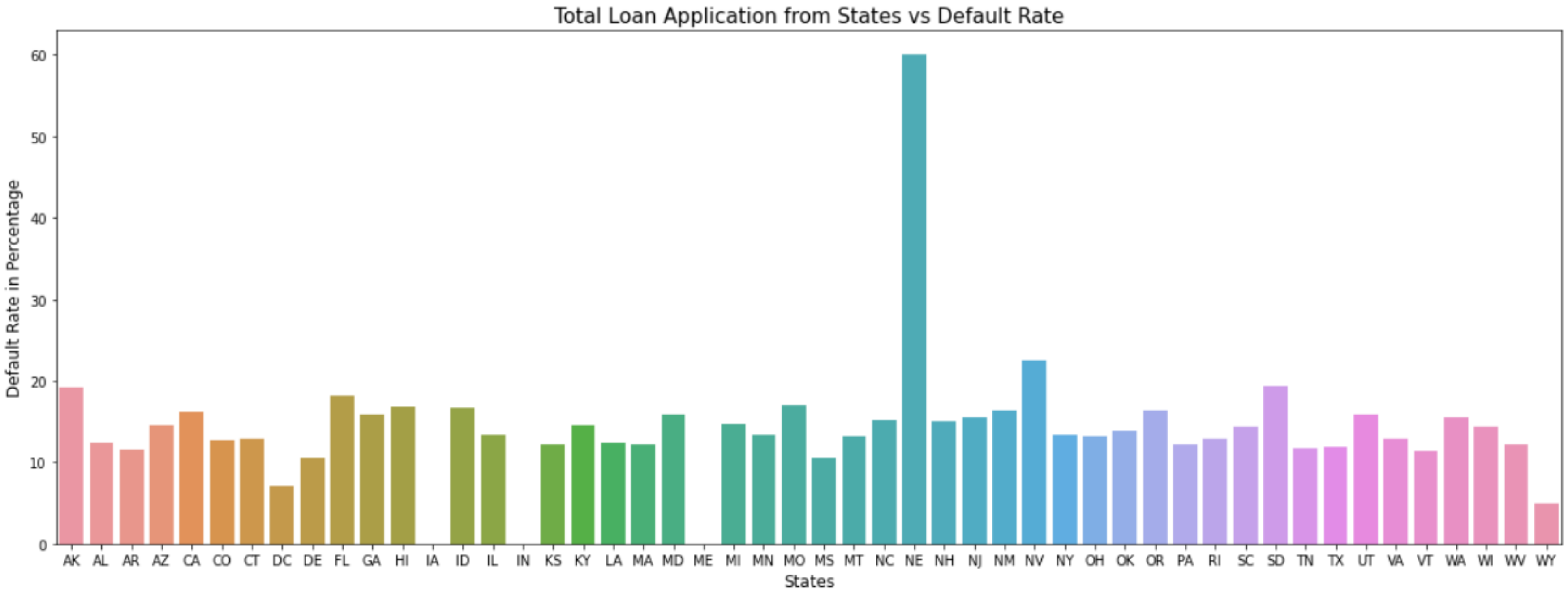
Conclusion

- Interest Rate is high for loan tenure of 60 months.
- Applicants with income verified by Lending Club applies more loan and defaults more, especially if loan tenure is 60 months.

EDA – Defaulters rate vs State

Observation

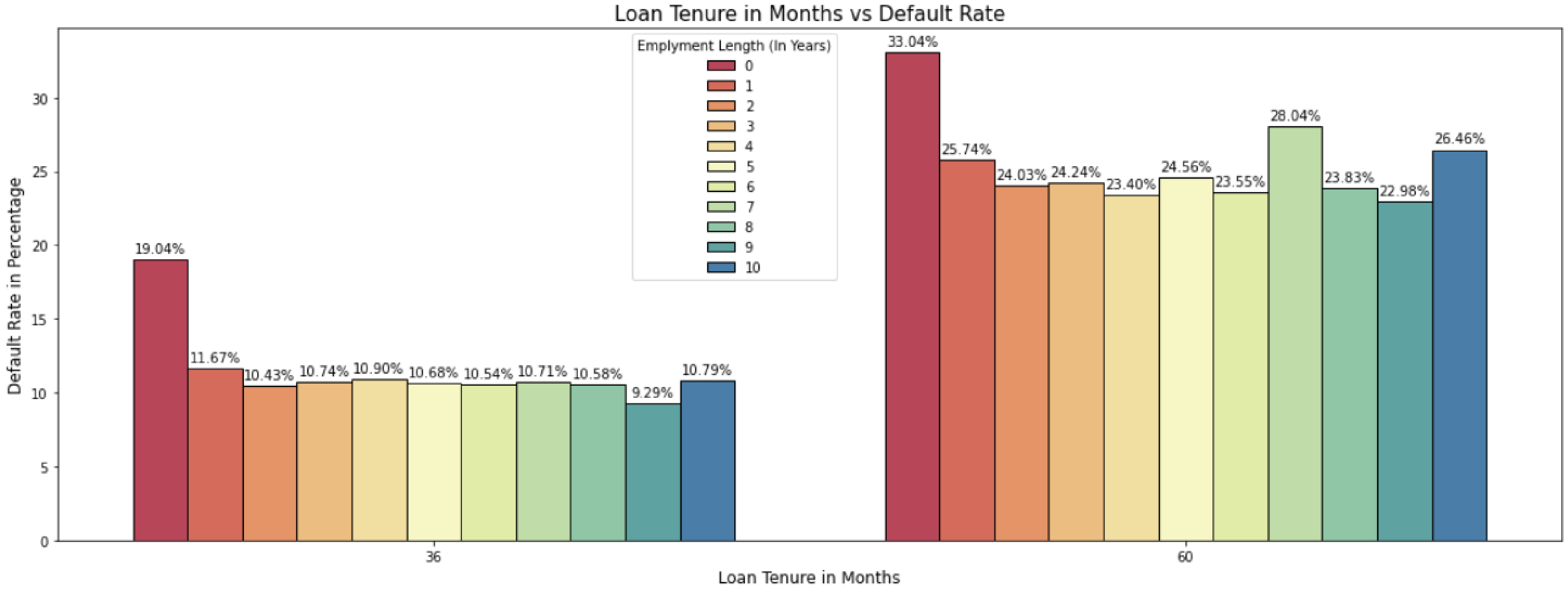
- Default Rate is higher in State NE because of low loan application count.
- States *NV* and *CA* have good number of loan applications and default rate is also high.



EDA – Employment Experience

Observation

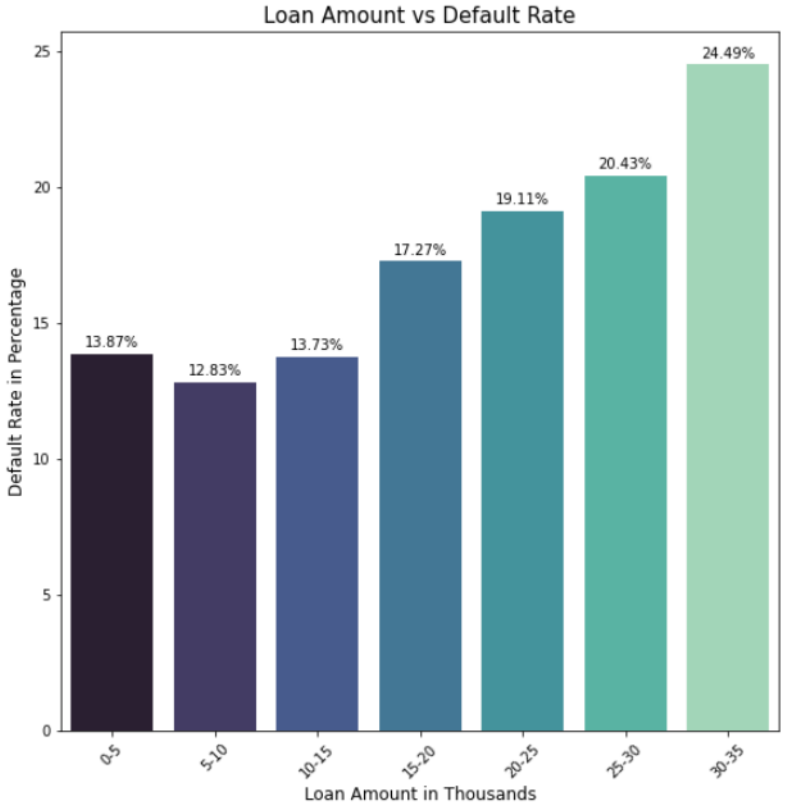
- Applicant's who have *no job* and *less than one year of experience* have high tendency to default.
- Reason for defaulting can be because of proper source of annual income.

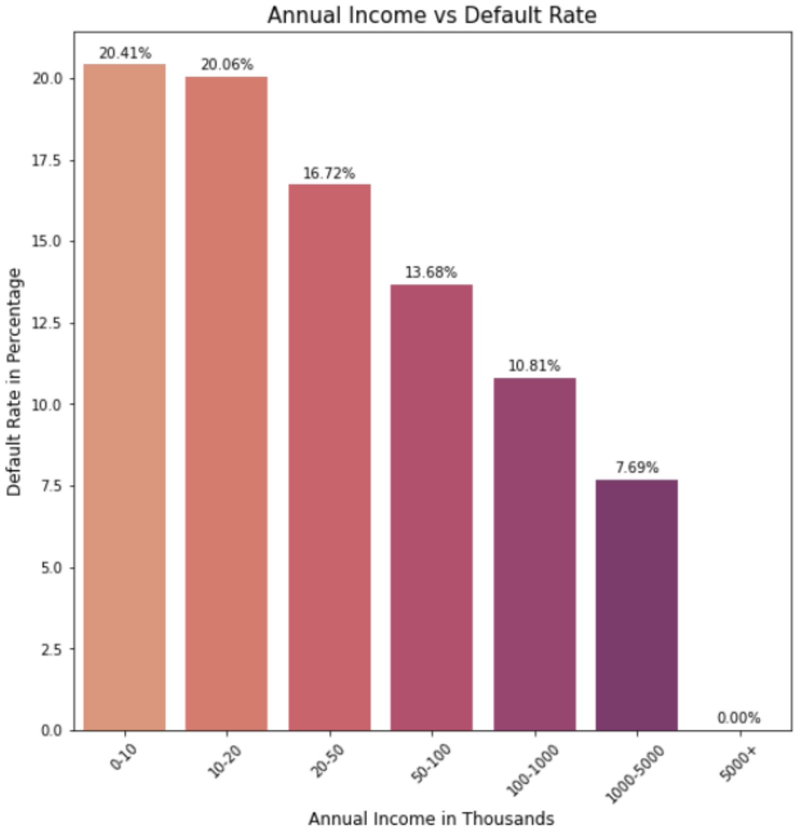


EDA - Loan Amount and Annual Income

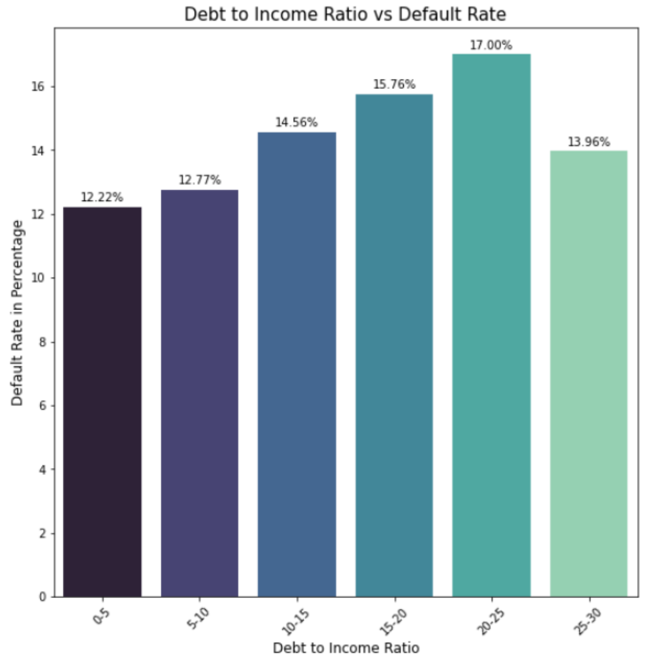
Observation

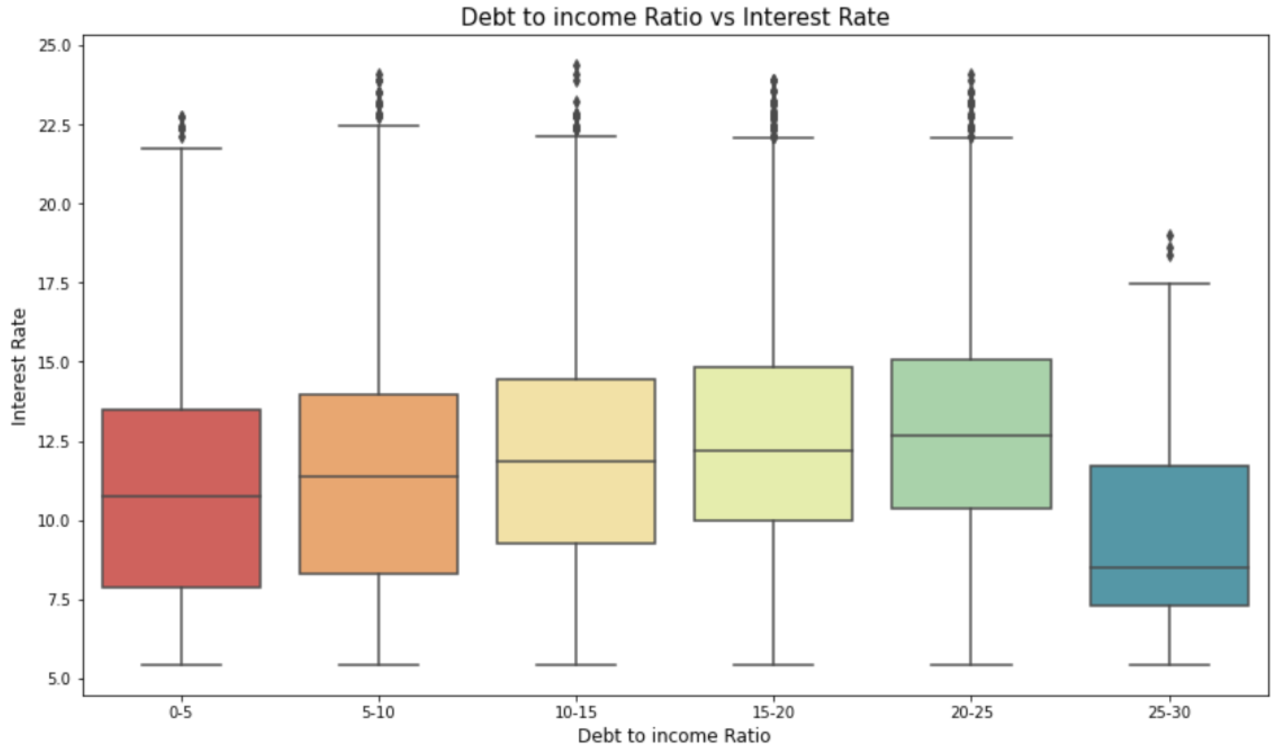
- Applicant's who have for *large loan amount* have high tendency to default.
- As *annual income increases*, there is less chance of default. This clearly relates to employment length, since annual income is very less for applicants who have 1 year of experience or no stable job.





EDA – Debt to Income (DTI) and Interest Rate



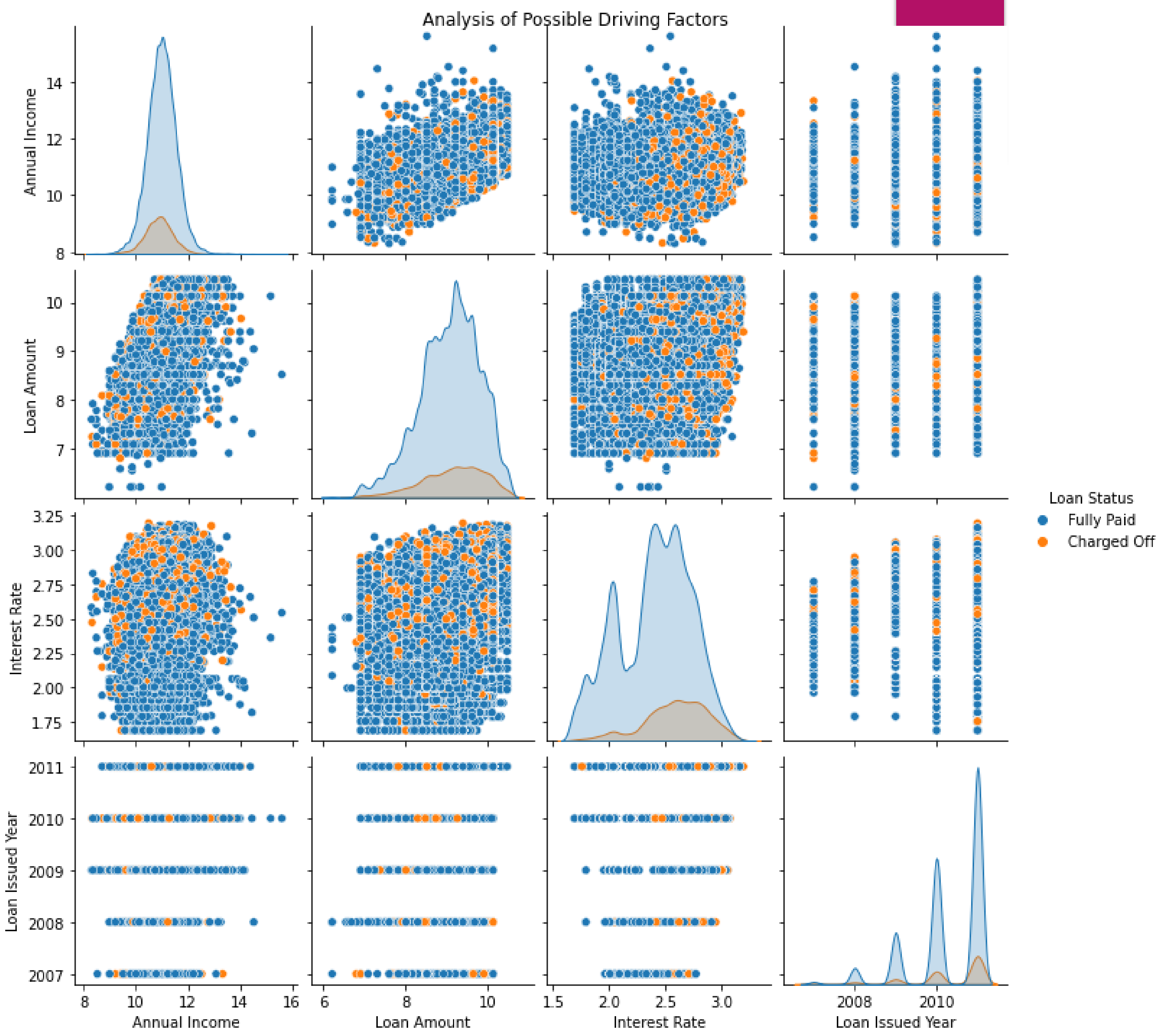


Observation

- 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.

DTI can't be good predictor, since percentage of defaulters changes as per interest rate varies.

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 – **25.81 %** prediction rate.
- If Annual Income is more, there is very less chance to default – **20.41 %** prediction rate.
- 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 applied for large amount have high tendency to default – **31.70 %** prediction rate.
- Loans with high interest rate have high tendency to default – **11.66 %** prediction rate.
- Extra precaution is required if Loan tenure is 60 months.
- Extra scrutiny is required if the purpose of loan applied is for small business.