LENDING CLUB CASE STUDY

SUBMISSION

Team:

Name: Yogesh Kolhe

Name: Fahad Baigh

Objectives

Strategy: Borrowers can easily access lower interest rate loans through a fast-online interface.

Business Objective:

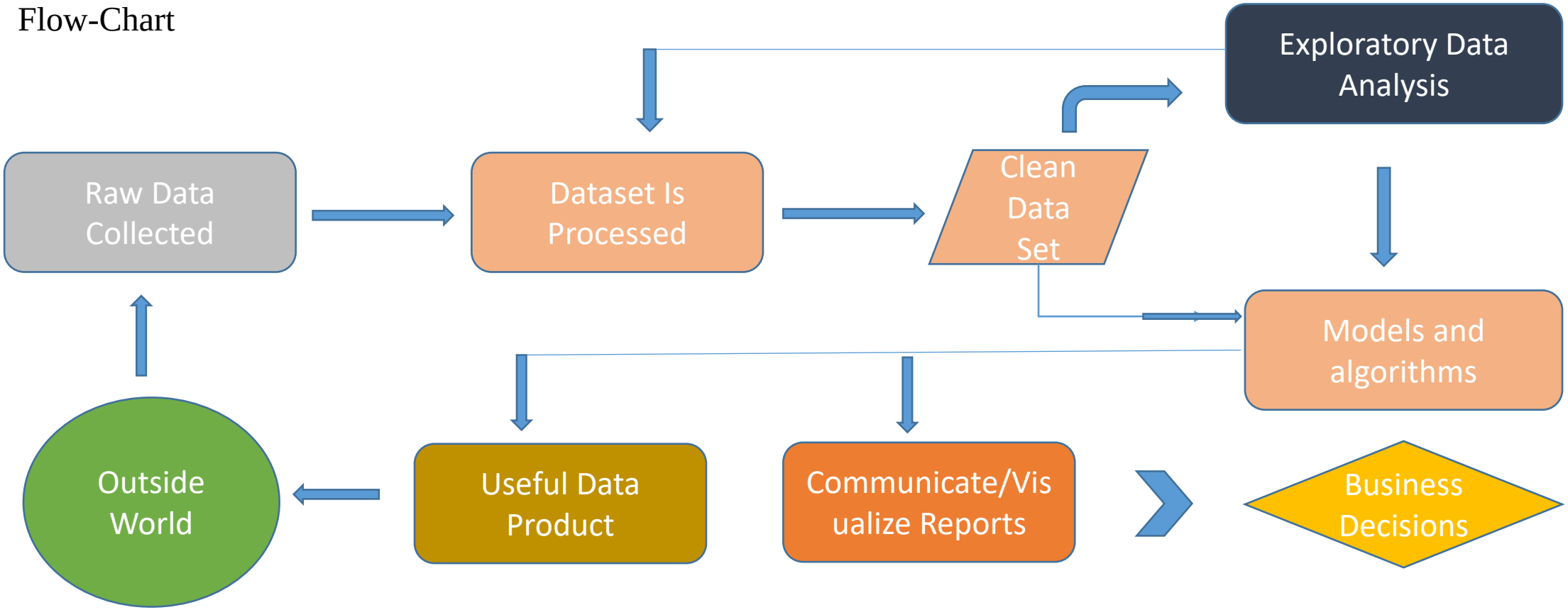
- Lending loans to 'risky' applicants is the largest source of financial loss or credit loss
- To identify the borrowers who default and cause the largest amount of loss to the lenders

Goals of Data Analysis:

- **Identify the risky loan applications:** Identify risky loan applicants using EDA (Exploratory Data Analysis), then such loans can be reduced thereby cutting down the amount of credit loss to the company.
- **Identify driving factors behind loan default:** Analyze the driving factors (driver variables) behind loan default, i.e. the variables which are strong indicators of default like 'Annual Income', 'Term' of loan, 'Interest Rate' etc. The company can utilize this knowledge for its portfolio and risk assessment.
- **Risk Analytics:** Understanding the types of variables and their significance to check the distribution of default rate across various driving variables.

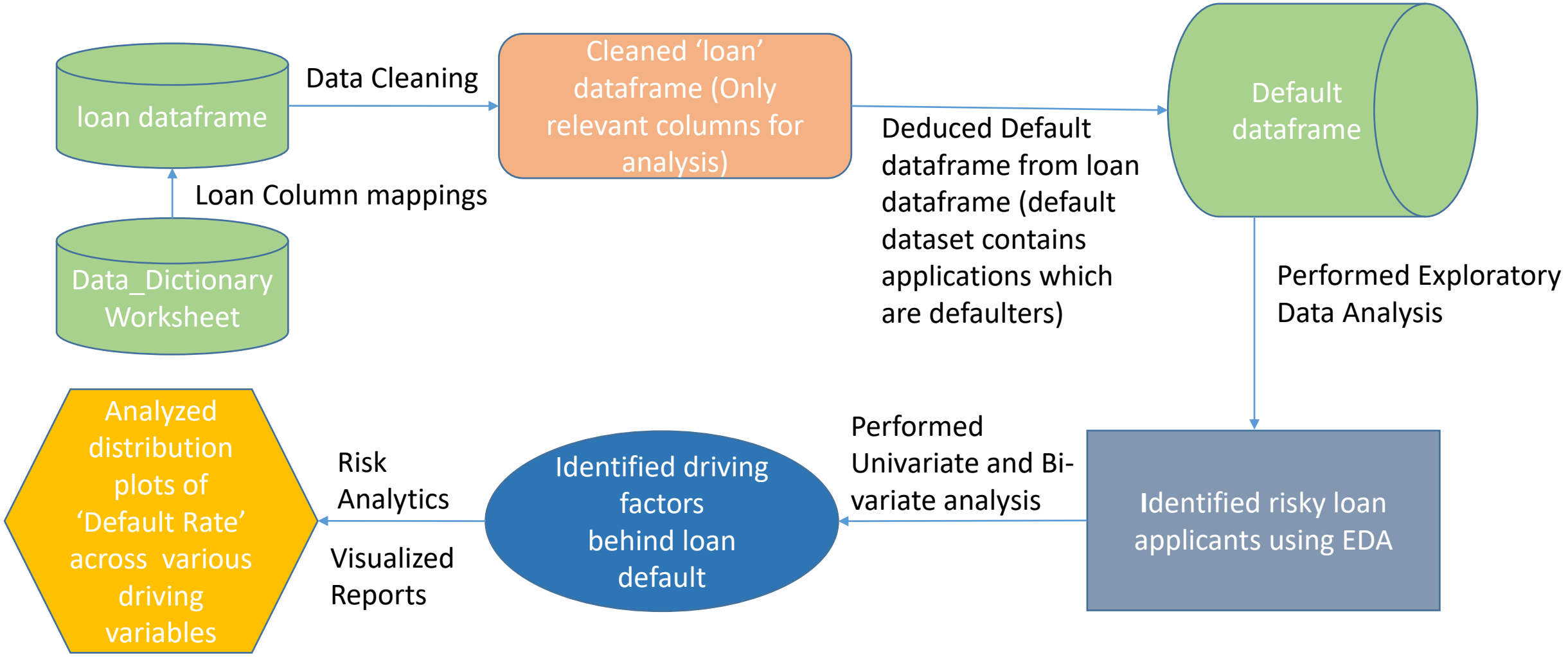
Problem solving methodology: General Approach

Flow-Chart



Problem solving methodology: Lending Club Case Study

Flow-Chart



Analysis

Understanding the Data Set

- Identified and removed redundant and unnecessary columns from the ‘loan’ data set
- Performed data cleaning on ‘loan’ data sets to remove the junk data (Columns having 100% null values and rows where the entire row is having null values) and the data which is not required for analysis
- Analyzed the variables and their different data types, also performed null value checks for the variables.

Data Cleaning and Manipulation

- Data quality issues are addressed in the right way (missing value imputation, outlier treatment and other kinds of data redundancies
- Wherever applicable, data is converted to a suitable and convenient format to work with using the right methods for e.g. converting date type columns to date type format and converting numeric column to int or float
- Extracted numeric values from the columns like ‘Interest rate’
- Manipulation of strings and dates is done correctly wherever required

Analysis

Categorical values Analysis

- Categorized ‘loan status’ to default and non –default (categorized the target variable into 1 and 0 respectively)
- Performed binning on continuous variables to categories the values for e.g. column ‘loan amount’ have been binned into low, medium and high

Data Analysis

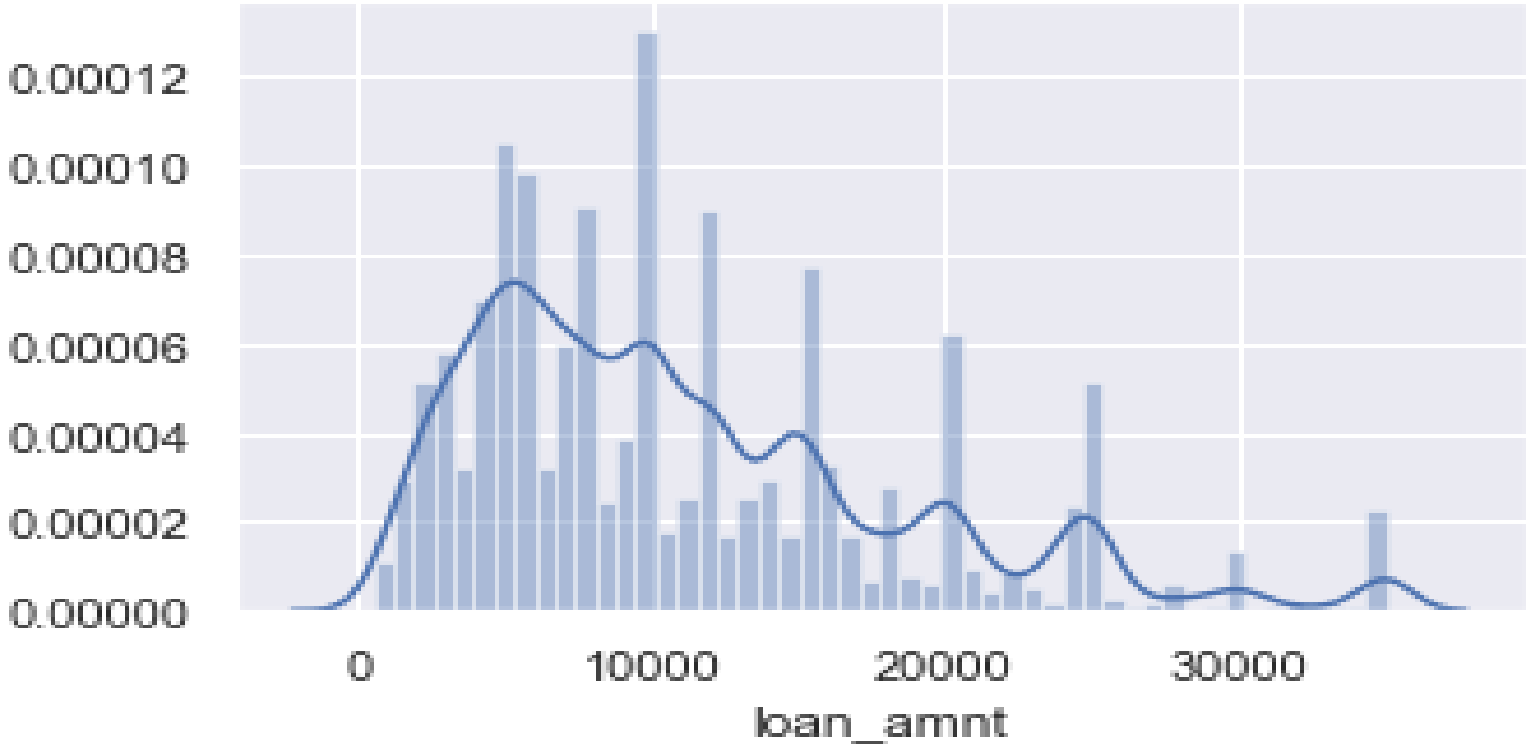
- Univariate and segmented univariate analysis is done correctly, and appropriate realistic assumptions are made wherever required
- The analyses successfully identify the important driver variables (i.e. variables which are strong indicators of default)
- Business-driven, type-driven and data-driven metrics are created for the important variables and utilized for analysis
- Bivariate analysis is performed correctly and is able to identify the important combinations of driver variables
- Appropriate plots are created to present the results of the analysis
- The plots clearly present the relevant insights and shows the distributions of ‘default rate’ among various diver variables

Univariate Analysis : Identified top variables affecting the 'Default Rate'

- A plot showing the loan 'Term' and 'Rate of Default'. This chart represents clearly that Loan Term 60 months has more 'Rate of Default'
- A plot showing the 'Loan Amount' and 'Rate of Default', This chart represents that there is a significant increase in the 'Default Rate' as the loan amount increases5
- A plot showing the 'Annual Income' and 'Rate of Default', This chart represents that there is a direct correlation between the annual income and 'Rate of Default. As the Annual income decrease, rate of default increases. Applicants having low annual income have more rate of default.
- A plot showing the 'Grade' and 'Rate of Default', This chart represents that there is a significant increase in the 'Rate of Default' as we go from Grade A towards Grade G.
- A plot showing the 'Sub-Grade' and 'Rate of Default', This chart represents clearly that subgrade follows the same trend as Grad for 'Rate of Default'
- A plot showing the 'Interest Rate' and 'Rate of Default', This chart represents that there is a direct correlation between the interest rate and 'Rate of Default. As the interest rate increases, default rate also increases. Applicants having low annual income have more rate of default.
- A plot showing the 'Verification Status' and 'Rate of Default', This plot represents that 'Verified Status' have more 'Default Rate'

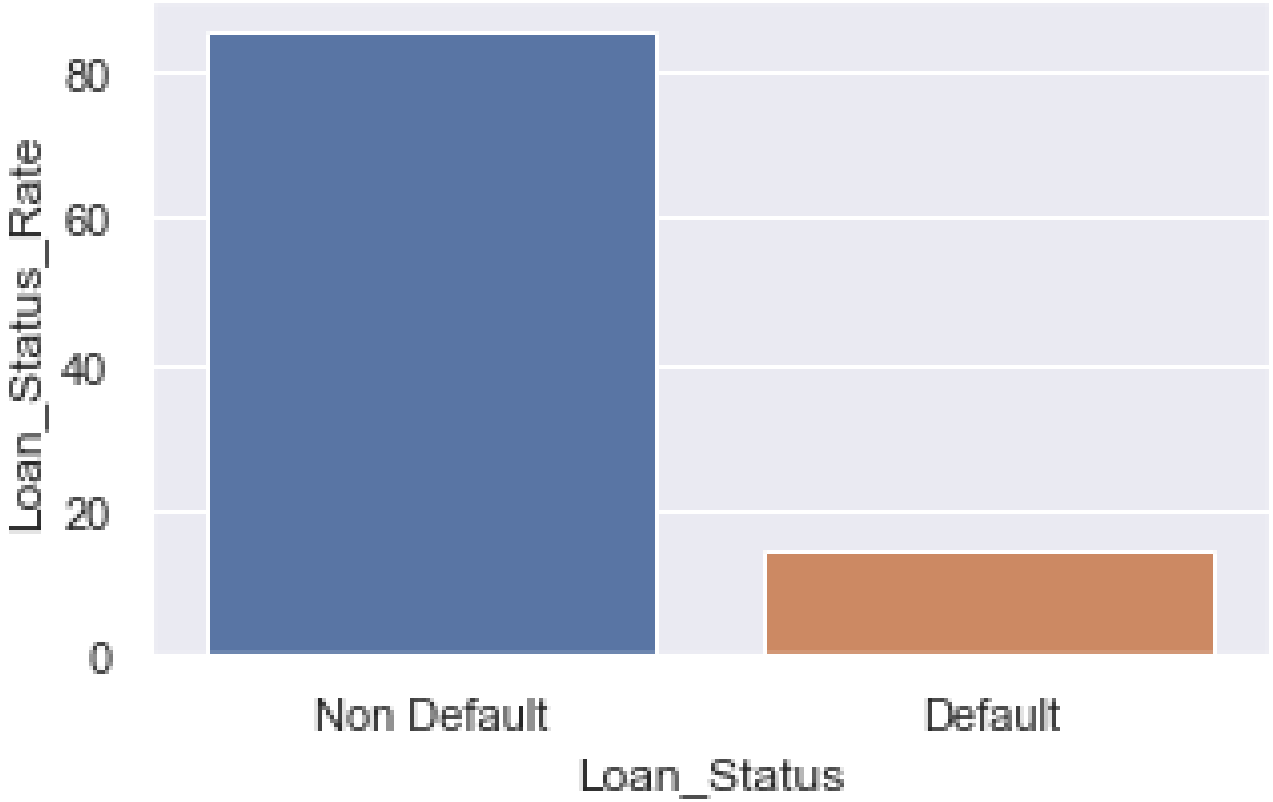
Visual Representation of the Univariate Analysis

Loan Amount Distribution Plot



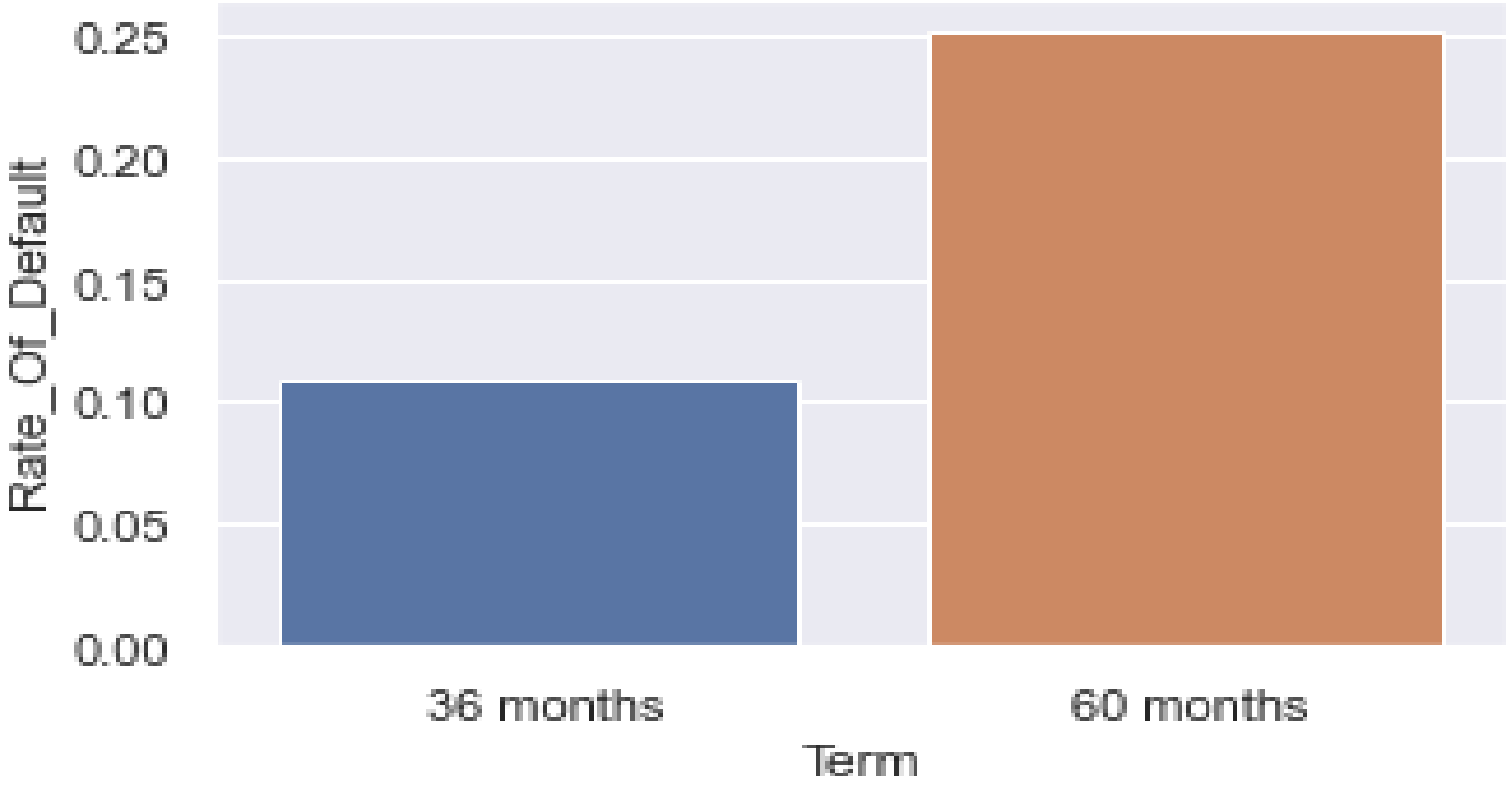
Visual Representation of the Univariate Analysis

Bar-Plot representing the percentage of ‘Default’ and ‘Non Default’ applicants in the dataset



Visual Representation of the Univariate Analysis

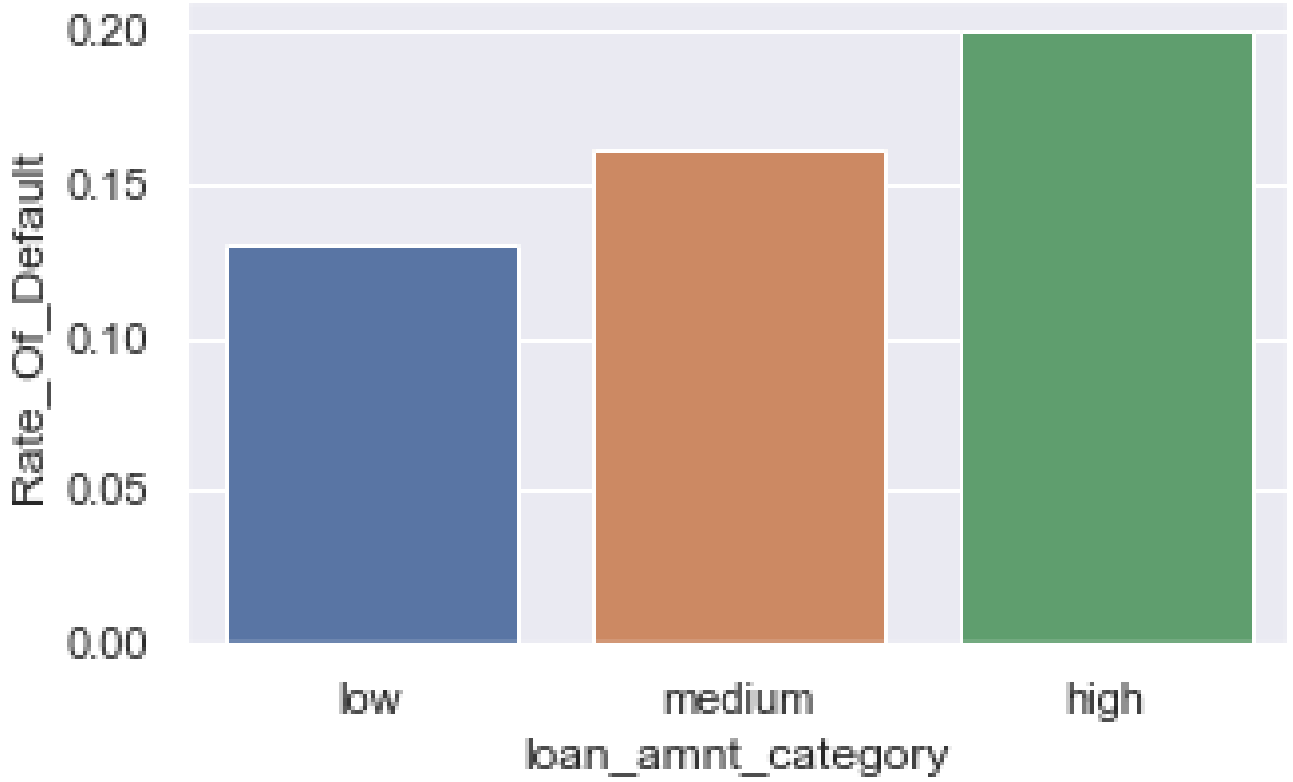
A plot showing the loan ‘Term’ and ‘Rate of Default’. This chart represents clearly that Loan Term 60 months has more ‘Rate of Default’



Visual Representation of the Univariate Analysis

A plot showing the ‘Loan Amount’ and ‘Rate of Default’, This chart represents that there is a significant increase in the ‘Default Rate’ as the loan amount increases

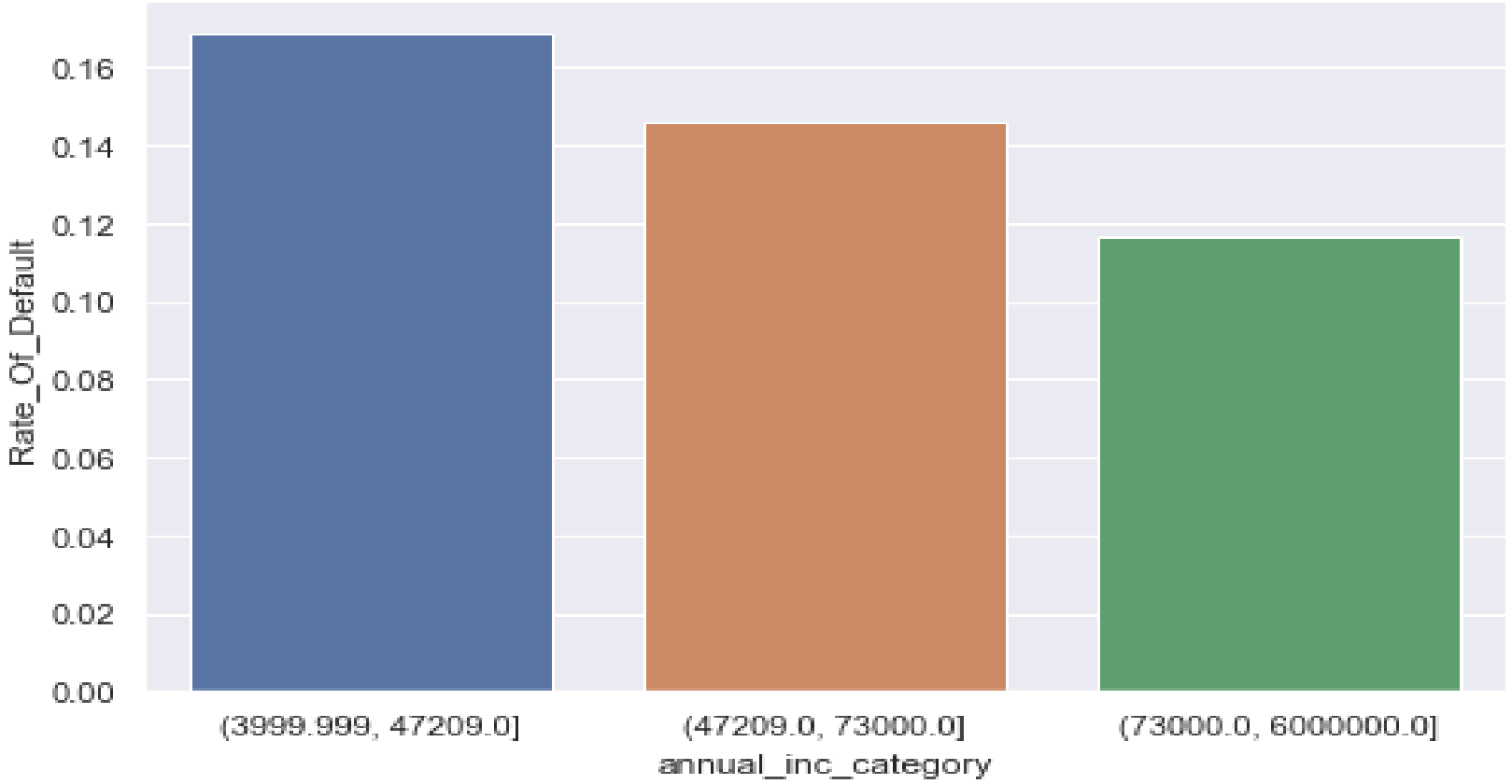
Note: Loan Amount values have been binned into 3 categories (low, medium and high)



Visual Representation of the Univariate Analysis

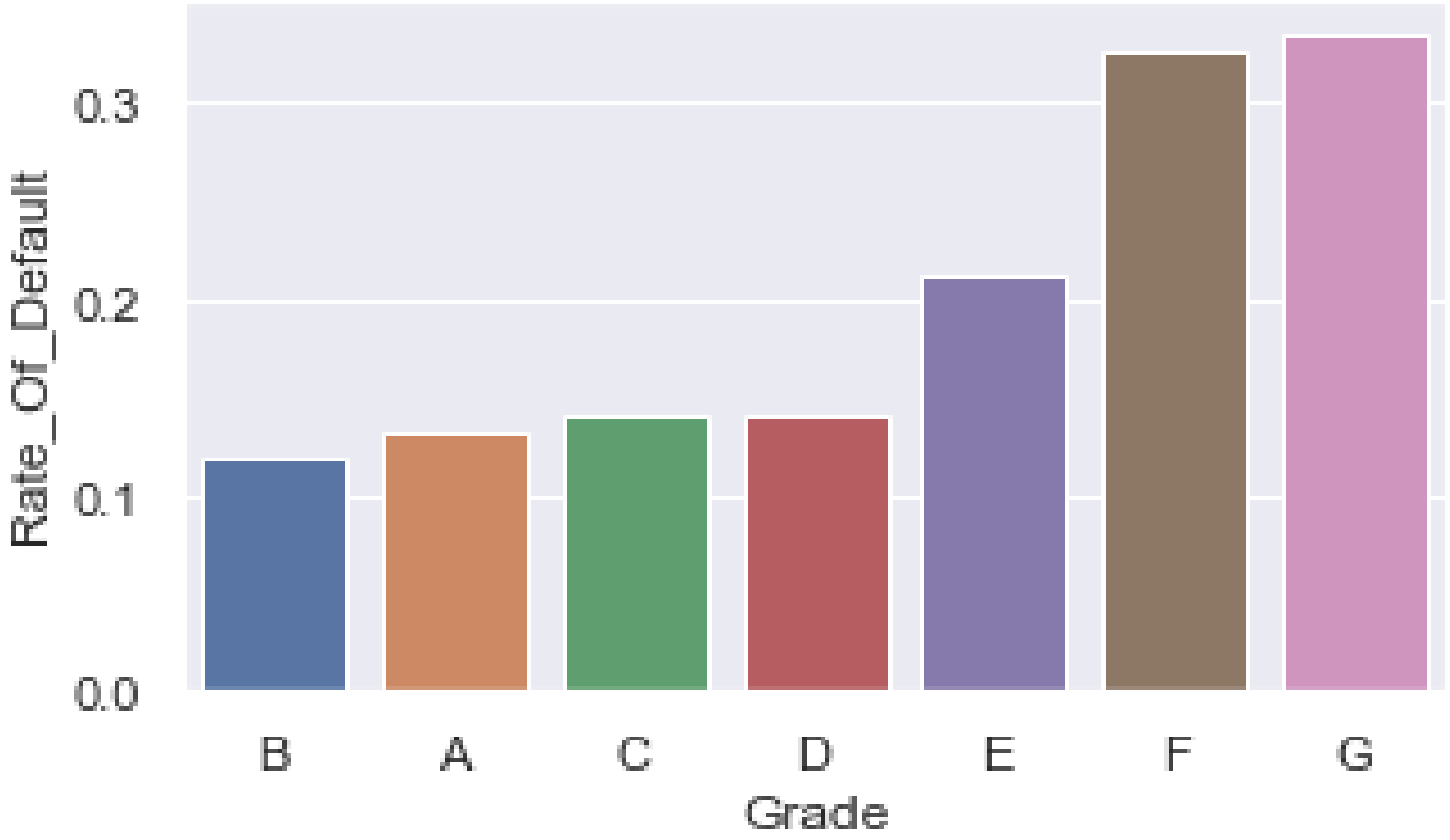
A plot showing the ‘Annual Income’ and ‘Rate of Default’, This chart represents that there is a direct correlation between the annual income and ‘Rate of Default. As the Annual income decrease, rate of default increases. Applicants having low annual income have more rate of default.

Note: Annual income values have been binned into 3 categories



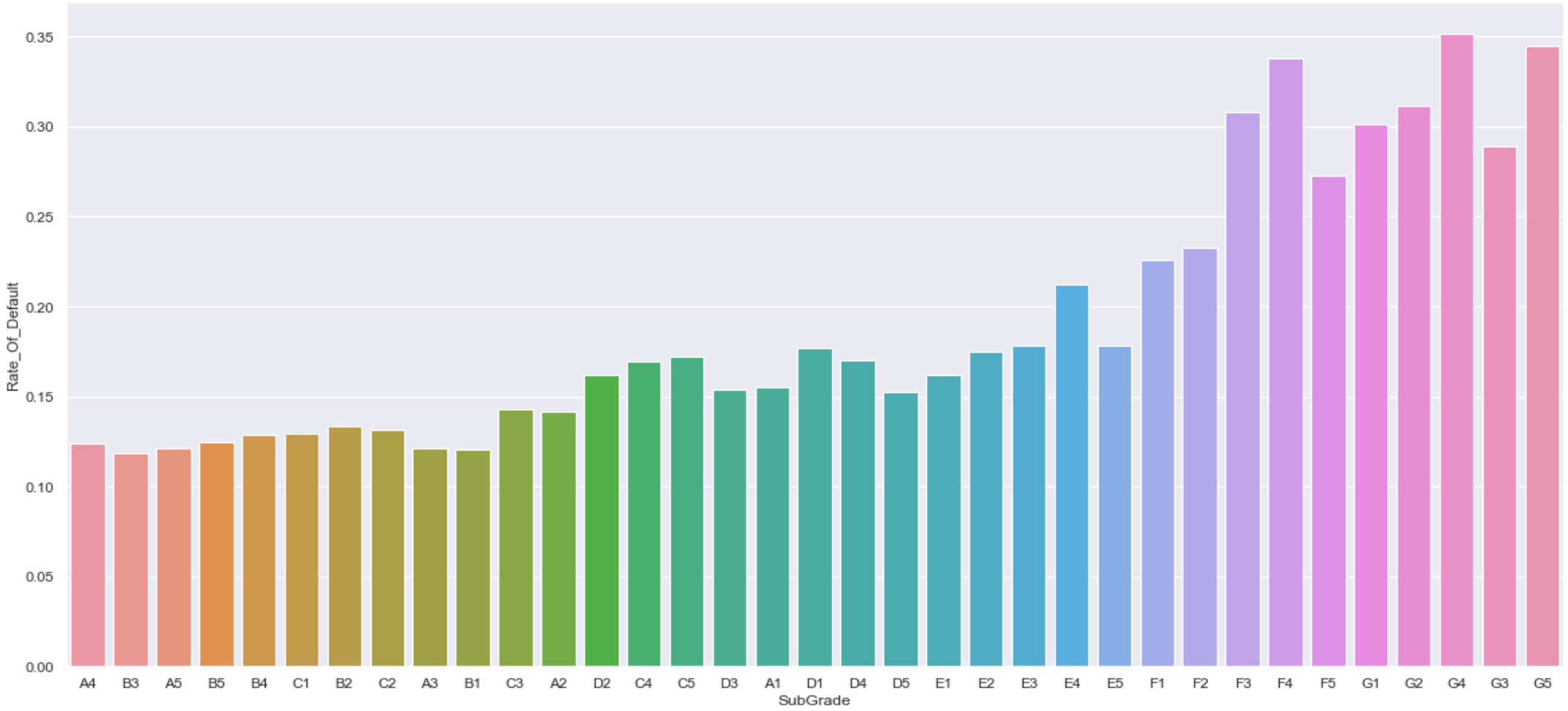
Visual Representation of the Univariate Analysis

A plot showing the ‘Grade’ and ‘Rate of Default’, This chart represents that there is a significant increase in the ‘Rate of Default’ as we go from Grade A towards Grade G.



Visual Representation of the Univariate Analysis

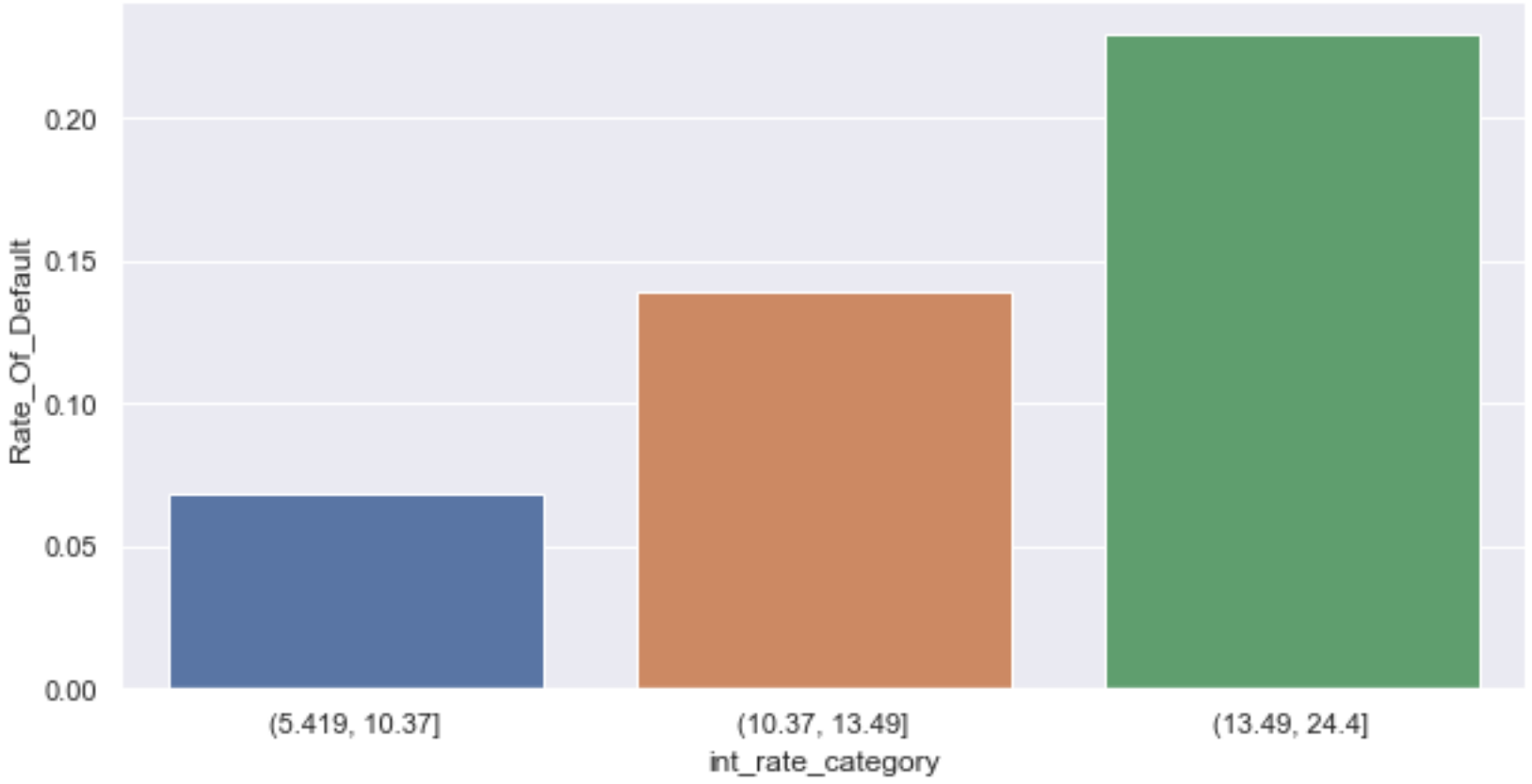
A plot showing the ‘Sub-Grade’ and ‘Rate of Default’, This chart represents clearly that subgrade follows the same trend as Grad for ‘Rate of Default’



Visual Representation of the Univariate Analysis

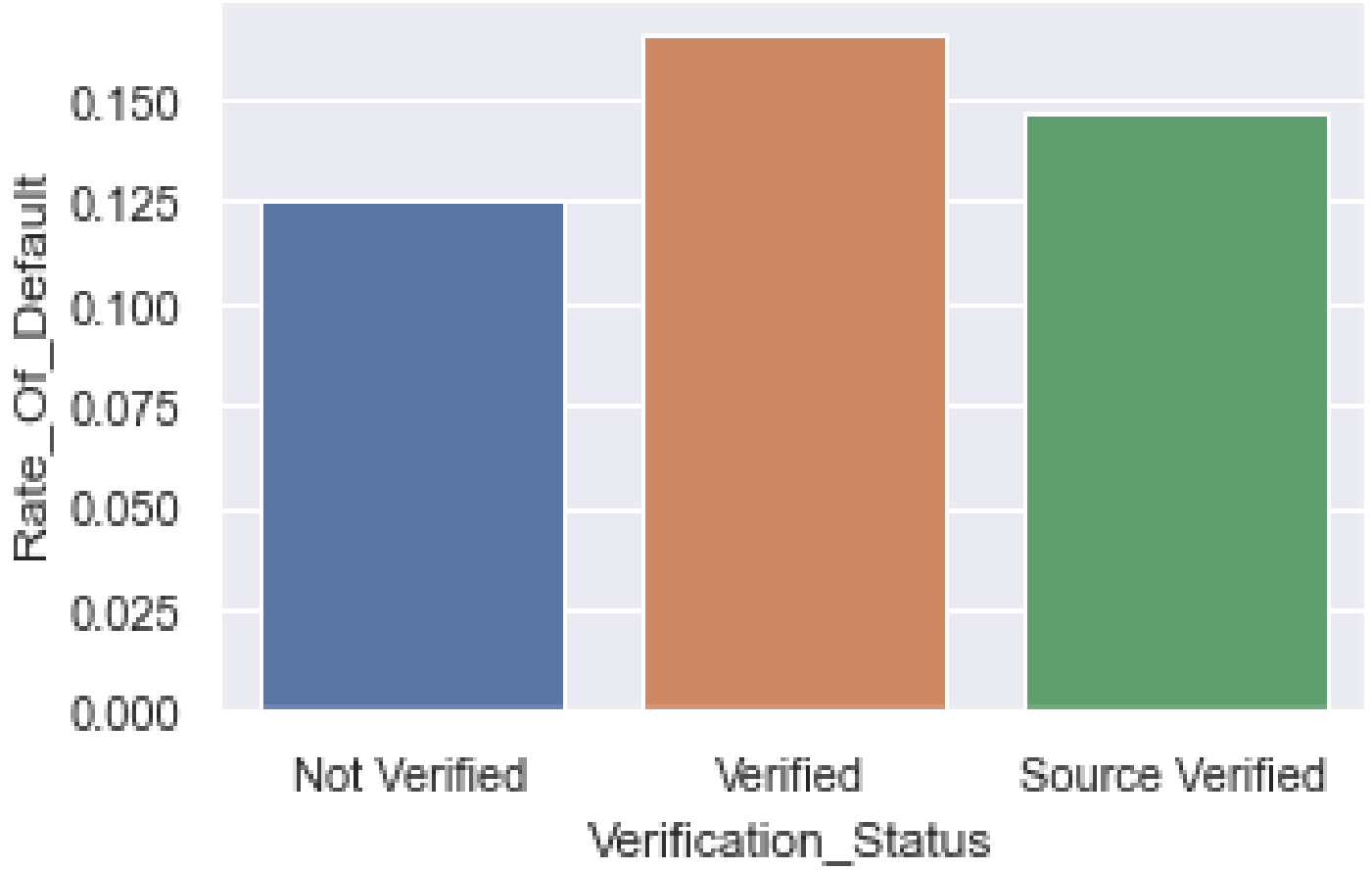
A plot showing the ‘Interest Rate’ and ‘Rate of Default’, This chart represents that there is a direct correlation between the interest rate and ‘Rate of Default. As the interest rate increases, rate of default also increases. Applicants having low annual income have more rate of default.

Note: Interest Rate values have been binned into 3 categories



Visual Representation of the Univariate Analysis

A plot showing the ‘Verification Status’ and ‘Rate of Default’, This plot represents that ‘Verified Status’ have more ‘Rate of Default’



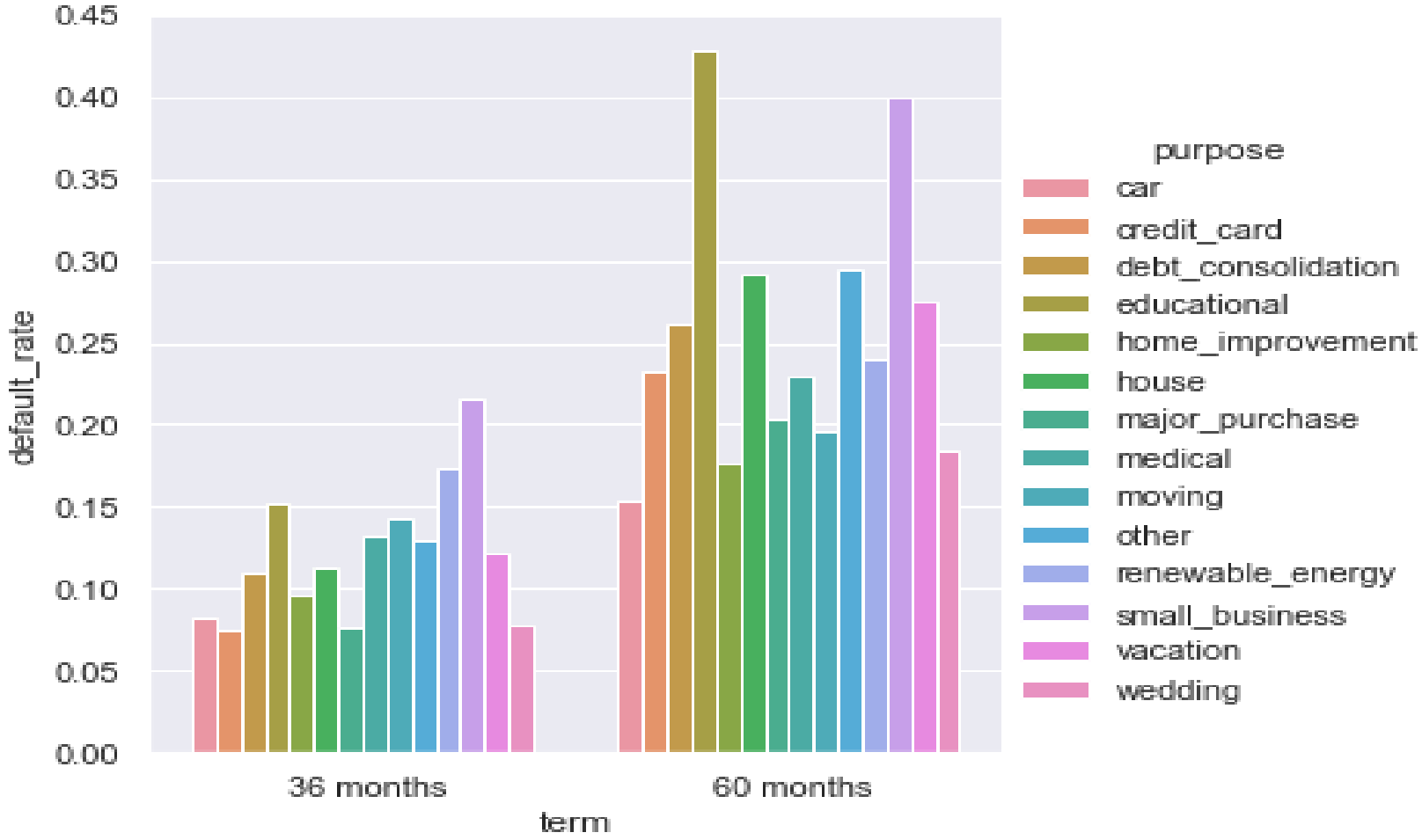
Bi-variate Analysis

Performed Bi-variate analysis for below driver variables

- ‘Purpose of Loan’ for each ‘Term’ against ‘Default Rate’ . This chart should represents the distribution of ‘Purpose of Loan’ for term 36 months and 60 months.
- A scatter distribution showing the distribution of ‘Emp Length’ against the ‘Loan Amount’

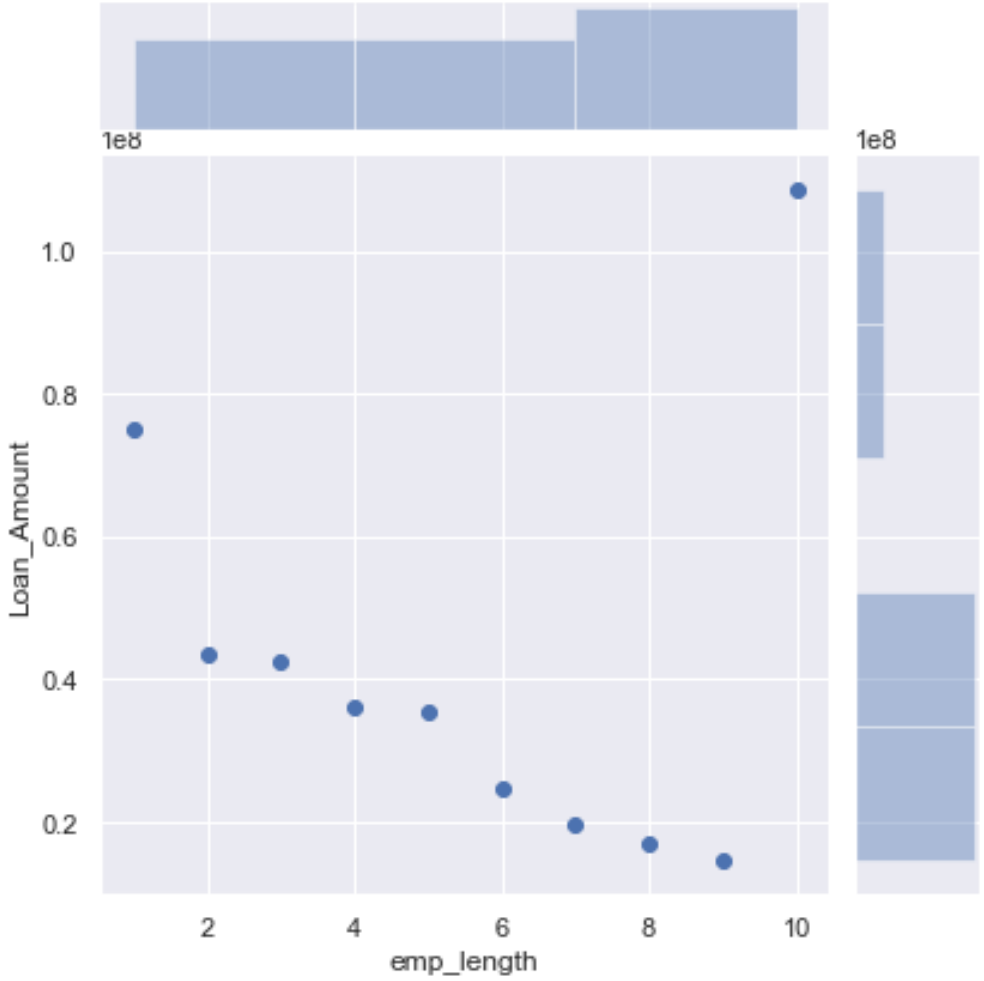
Visual Representation of the Bi-variate Analysis

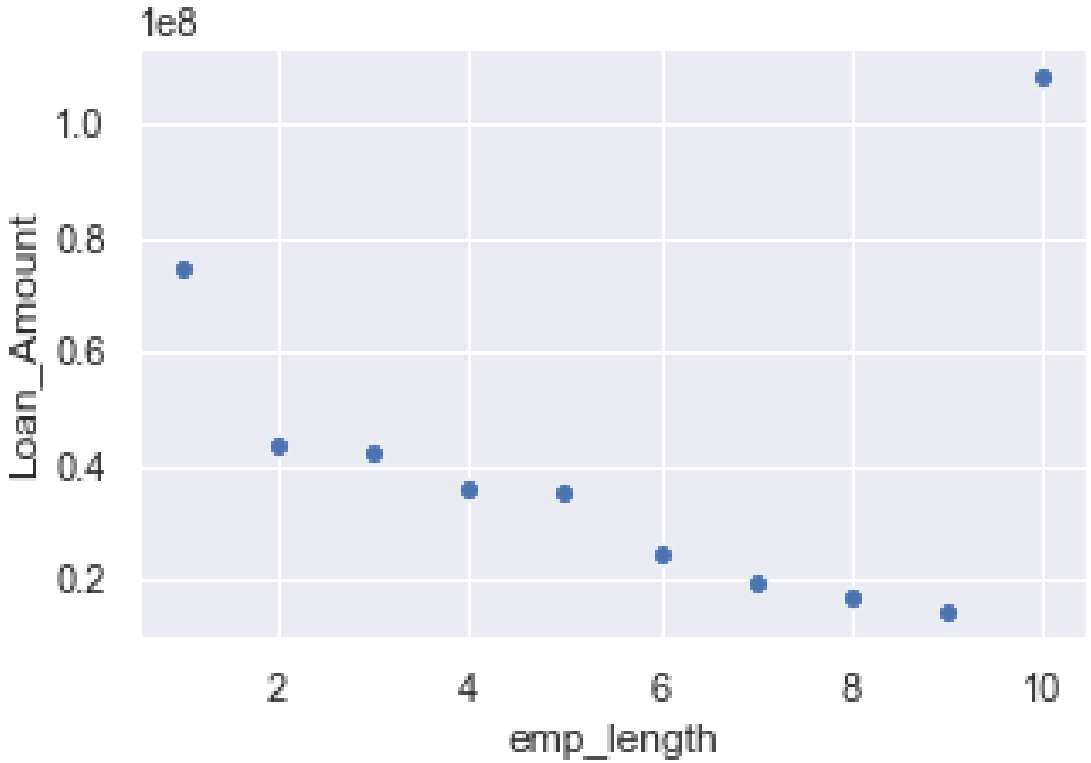
A plot showing the ‘Purpose of Loan’ for each ‘Term’ against ‘Default Rate’ . This chart represents the distribution of ‘Purpose of Loan’ for term 36 months and 60 months, which has more ‘Rate of Default’



Visual Representation of the Bi-variate Analysis

A scatter plot showing the distribution of ‘Emp Length’ against the ‘Loan Amount’





Conclusions and Recommendations

Based on the univariate analysis of driving factors for the 'Default Rate' the top important variables are as below:

- **Term :** As the Loan term increases default rate also increases. Loan Term 60 months has more 'Rate of Default'
- **Loan Amount :** There is a significant increase in the 'Default Rate' as the loan amount increases
- **Annual Income :** There is a direct correlation between the annual income and 'Rate of Default. As the Annual income decrease, rate of default increases. Applicants having low annual income have more rate of default.
- **Grade :** There is a significant increase in the 'Default Rate' as we go from Grade A towards Grade G.
- **Sub-Grade :** Subgrade follows the same trend as Grad for 'Default Rate'
- **Interest Rate :** There is a direct correlation between the interest rate and 'Rate of Default. As the interest rate increases, rate of default also increases.
- **Verification Status:** 'Verified Status' have more 'Default Rate'

Conclusions and Recommendations

Recommendations:

✓ As a loan funding company it should analyze below important driver variables behind loan default

- **Term**
- **Loan Amount**
- **Annual Income**
- **Grade**
- **Sub-Grade**
- **Interest Rate**
- **Verification Status**

✓ These variables are the strong indicators of default

✓ The company can utilize this analysis for its portfolio and risk assessment and thereby cutting down the credit loss

Conclusions and Recommendations

Based on the Bi-variate analysis of driving factors for the 'Default Rate', below are the conclusions:

- **'Purpose of Loan' for each 'Term' against 'Default Rate' :** The distribution represents that as the term increases the 'Default Rate' also increases for each type of 'Purpose of Loan'. Term 60 months, has higher 'Rate of Default'
- **Employment length in years against the 'Loan Amount :** A scatter plot distribution of 'Emp Length' against the 'Loan Amount' represents that there is a decrease in the Loam Amount as the employment length or the experience of an employee increases from 0 to 10 years. (10 years being an outlier in the given distribution)

Recommendations:

✓ As a loan funding company : Applications received for smaller term period and having greater employment length are the suitable ones to grant the loan application.