LENDING CLUB CASE STUDY

SUBMISSION

Anirban Chakraborty
Raghul Ranganathan J

A **Consumer Finance Company,** who specialises in lending various types of loans to urban customers has to make decision for loan approval based on applicant's profile.

Business Objective

The main goal is to **identify the risky loan applicants** and thereby cutting down the amount of credit loss. This can be achieved by finding the **driving factors** (strong indicators) behind the loan default.

By achieving this objective the company can **use this knowledge for its portfolio and risk assessment.**

Types of Risks Associated

- If the applicant is **likely to repay the loan**, then not approving the loan results in a loss of business to the company.
- If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Deciding Factors

- **On Loan Accepted -** Use the data for analysis
  - **Fully Paid -** Applicant has fully paid the loan.
  - **Charged-off:** Applicant has not paid the instalments in due time for a long period of time (Defaulted on Loan)

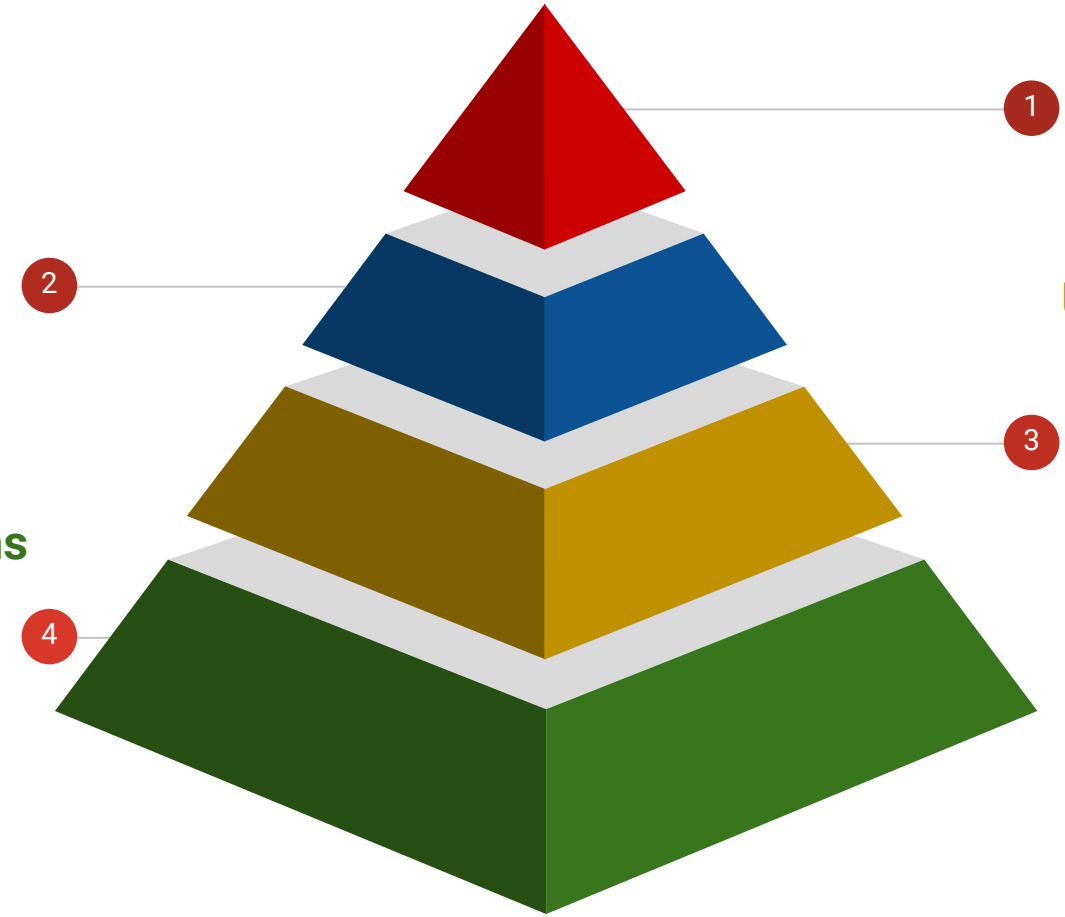
Problem Solving Process

Understanding the Dataset

- Comprehend about the columns
- Develop basic idea of dataset
- Get an insights between the columns
- Decide on the required columns that we are about to focus for our analysis

Data Cleaning

- Remove unwanted columns
- Drop all the columns which have more null values.
- Impute the missing data in required columns.
- Distinguish the columns based on its numerical and categorical value.
- Formulate new required numerical columns from categorical columns.



Univariate & Bivariate Analysis

- Analyze a data set columns independently
- Visualize and plot the desired data
- Explore the relationship between the columns
- Visualize and plot the relation of the desired data.

Summary & Recommendations

- List down the driving factors behind the loan defaults
- Recommendations to be done to avoid the financial loss for the company

Data Cleaning

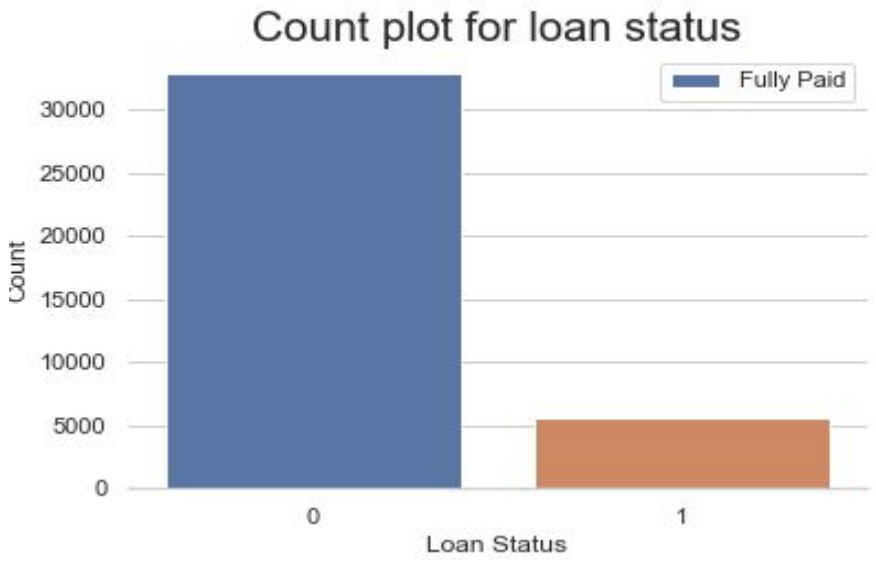
From the available data, various analysis is done and the quality of data is maintained without making any deviation to goal and business objective

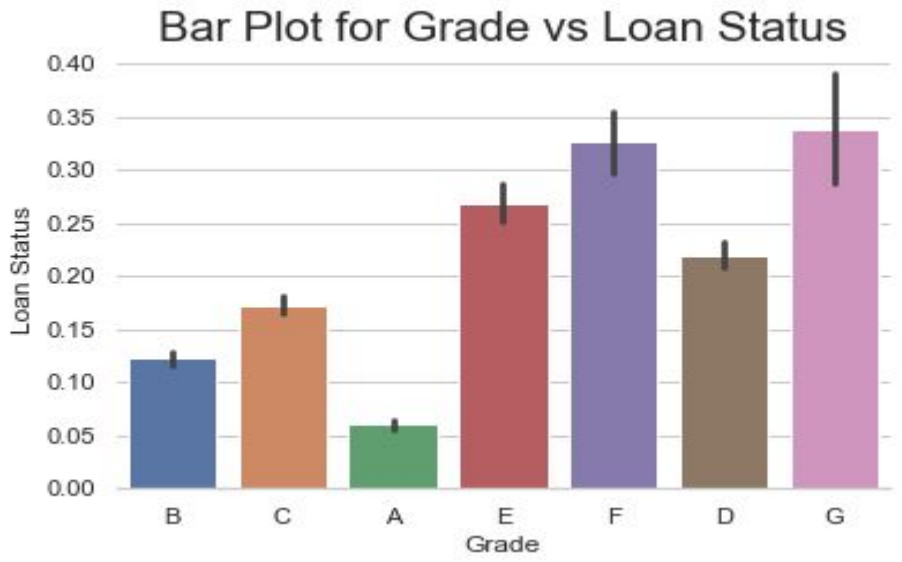
Steps Involved

- **Remove Columns** - Containing more null values or not required for analysis
- **Remove Rows** - Containing most number of empty values
- **String Manipulations** - Replacing cell values, case changes, fill partial values, rename columns to have a proper naming.
- **Data Correction** - Correcting incorrect data, data types, impute missing data, etc.

Data Cleaning needs to be done for all data columns and data can be merged and maintained as a master data frame for upcoming process.

Univariate Analysis

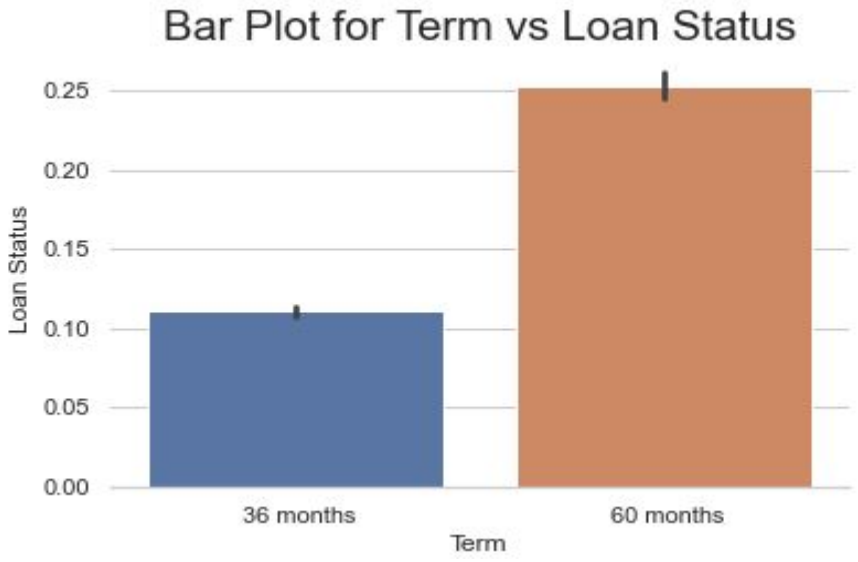




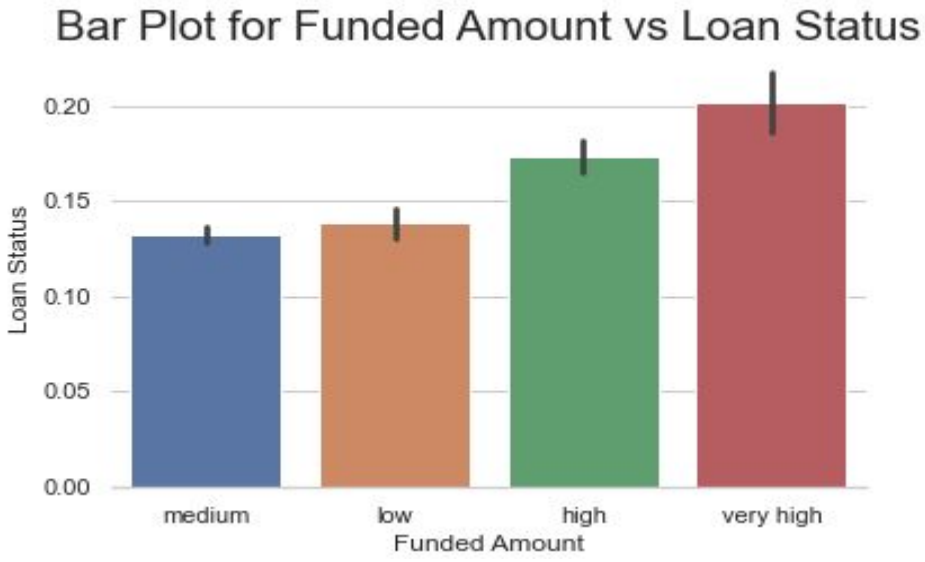
- Count plot is done against number of loans and the loan status only with “fully paid” and “charged off”.
- Out of the **total number of granted loans, 16% are of charged off** which meant to be high in number

- We categorized the Default rates based on the Loan Grade
- We can clearly infer that Loans with Grade **G, F has very high rate of defaulters**
- We can also see Loans with Grade **A, B has low rate of defaulters**

Univariate Analysis

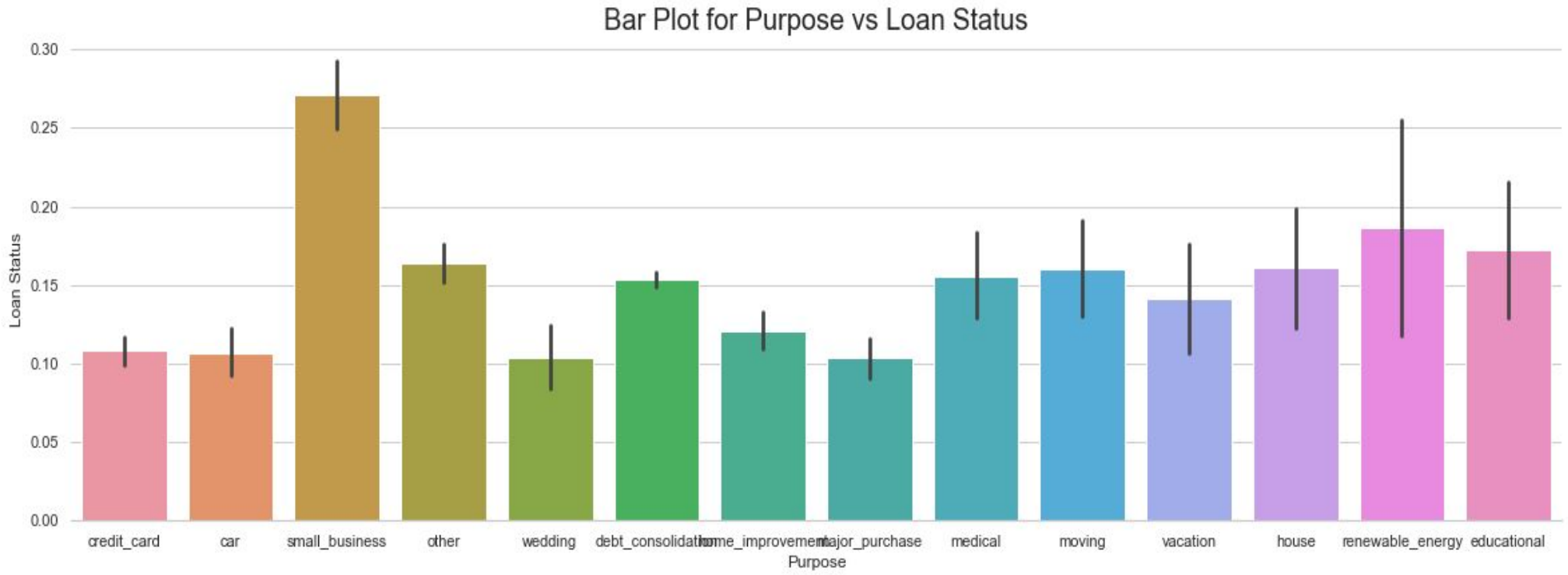


- Visualization of Loan defaulters based the Loan tenure was made
- We can infer that loans with **high terms (60 months) has high chances for loan defaults** than the lower term (36 months)



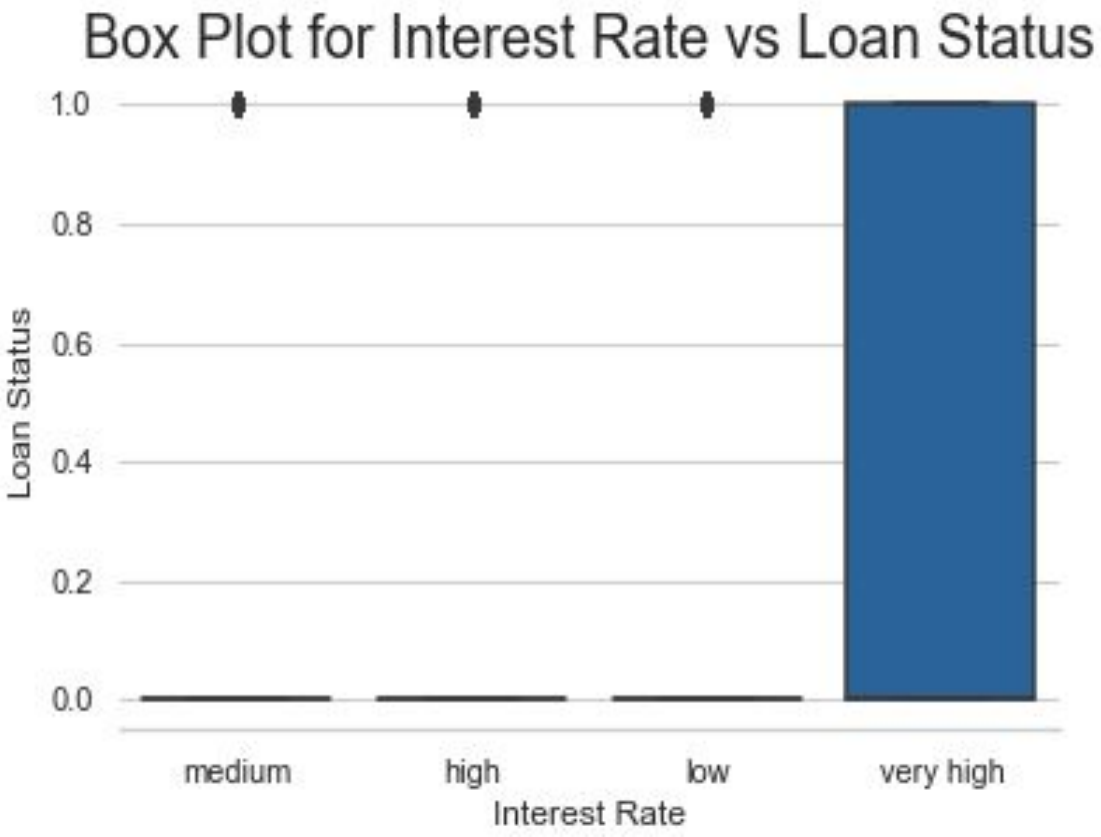
- We categorized the Default rates based on the Amount committed to the loan
- Higher funded amount has higher chances of getting defaulted
- Loan Amount committed **less than 25000 has lowest chance** of getting defaulted
- Loan Amount committed **more than 25000 has high chance** of getting defaulted

Univariate Analysis



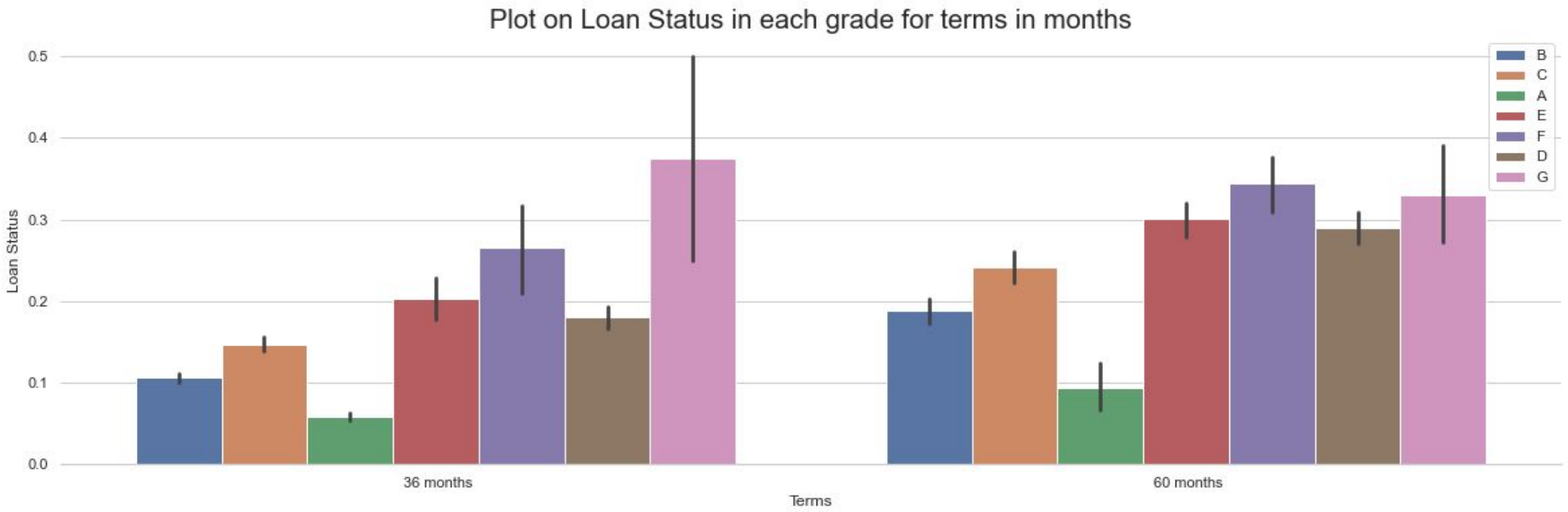
- We categorized the Default rates based on the Purpose(Category) of the loan
- Loan with the purpose of **Small Business has the high risk and most likely to get default.**
- Loan with the purposes like Credit card, Car, Wedding, Purchases has the low risk and low probability to get default.

Univariate Analysis



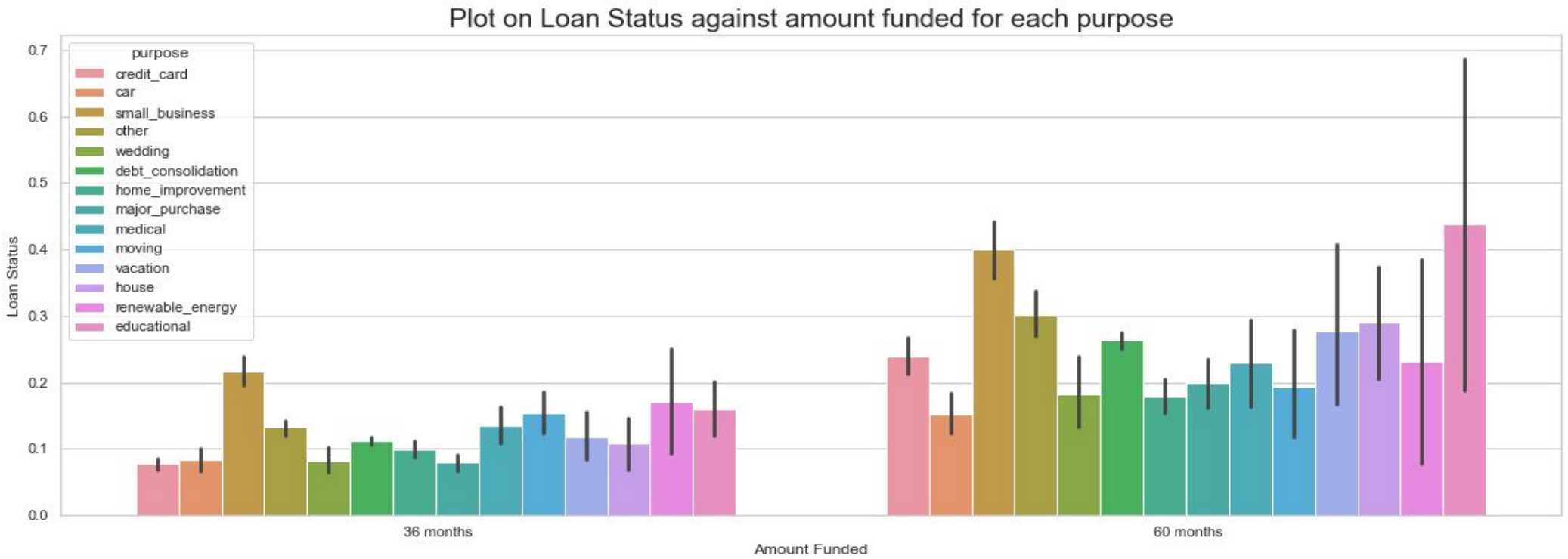
- Visualization of Loan defaulters based on the Interest Rate provided to the loan was made.
- We can infer that loans with **high interest rates has high chances for loan defaults**.
- Here Interest Rates of value more than 20% has maximum chances of default

Bivariate Analysis



- We categorized all grades of loan in each terms against the loan status
- **Grade G** loans for interval 36 months are mostly likely to default
- **Grade F** loans for interval 60 months are mostly likely to default
- **Grade A** loans are safest irrespective of their terms

Bivariate Analysis



- We categorized all purpose of loan in each terms against the loan status
- Loan with **purpose of Small Business has high default** in the term with **36 months**
- Loan with **purpose of Education has high default** in the term with **60 months**

Correlation Matrix

	annual_inc	delinq_2yrs	dti	emp_length	open_acc	pub_rec	pub_rec_bankruptcies	total_acc	loan_status
annual_inc	1.00	0.02	-0.12	0.11	0.16	-0.02	-0.02	0.23	-0.04
delinq_2yrs	0.02	1.00	-0.03	0.02	0.01	0.01	0.00	0.07	0.02
dti	-0.12	-0.03	1.00	0.05	0.29	-0.00	0.01	0.23	0.05
emp_length	0.11	0.02	0.05	1.00	0.10	0.06	0.06	0.21	0.02
open_acc	0.16	0.01	0.29	0.10	1.00	0.00	0.01	0.69	-0.01
pub_rec	-0.02	0.01	-0.00	0.06	0.00	1.00	0.84	-0.02	0.05
pub_rec_bankruptcies	-0.02	0.00	0.01	0.06	0.01	0.84	1.00	-0.01	0.05
total_acc	0.23	0.07	0.23	0.21	0.69	-0.02	-0.01	1.00	-0.02
loan_status	-0.04	0.02	0.05	0.02	-0.01	0.05	0.05	-0.02	1.00

From the above matrix we can infer few details as follows:

- The **DTI** value is **high then the borrower has very high obligation to pay other debts** or has more expenditure than income which is the clear indicator that is **more likely to be a defaulter**
- We can see that **Annual Income of the borrower** is **inversely proportional to the Public Record of Bankruptcies** for the applicant of any type of loan

Conclusions

- The driving factors like **Annual Income, Interest Rate, Loan Terms, DTI, Loan Status, Grade, Purpose and Interest rate** should be monitored and verified before approving the loan.
- The above driving factors will help the company to avoid the financial loss by rejecting the loans to the defaulters.

Recommendations:

- Around **16% of borrower has defaulted** which can be considered as a **high number** and investors needs to take measures to **collect the amount back** that was lended .
- **Higher graded loans (G, F)** has high percentage of defaulters, investors should not give a **high grade loan** without looking at details of borrowers.
- A **high loan term has higher chances of default** and so Investors should stop lending money for higher terms
- Loan given for **Small Business or Education Loan debts** involves high risk and most likely to default and thus Investors should judiciously invest with small business ventures.
- A **high interest rate often leads to higher chances of loan getting default**.
- Company should look into the **DTI score** carefully before lending money. Company should **lend money to applicants** having very **low DTI** score .