LENDING CLUB CASE STUDY

SUBMISSION

Group Members:

1. Krishna Kumar
2. Anjana Nair

Abstract

- When a **consumer finance company** which specialises in lending various types of loans to urban customers receives a loan application, the company has to make a decision for loan approval based on the applicant’s profile.
- The data given contains the information about past loan applicants and whether they ‘defaulted’ or not.
- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Business Objective :

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Problem solving methodology

Data Understanding

Understanding the dataset and gaining statistical insights from dataset at a high level.

Data Cleaning and Manipulation

Cleaning Rows and Columns of Dataset.

Identifying and fixing missing values.

Fixing Strings and Dates.

Exploratory Data Analysis

Identifying predictors of default.

Performed Univariate and Bivariate Analysis.

Results and Recommend ations

Summarizing results.

Coming up with key recommendations based on the result of analysis

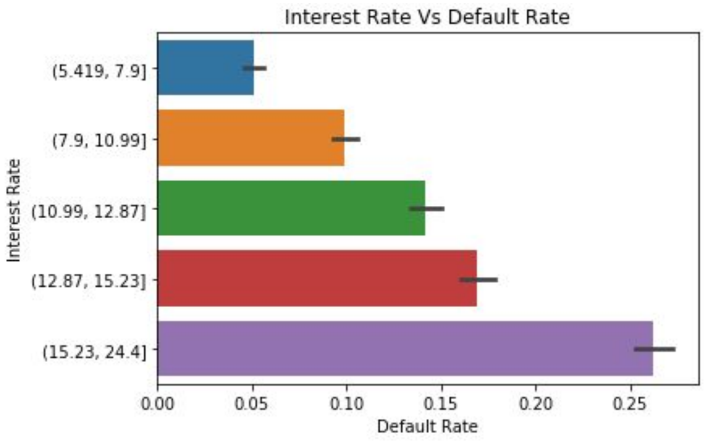
Analysis

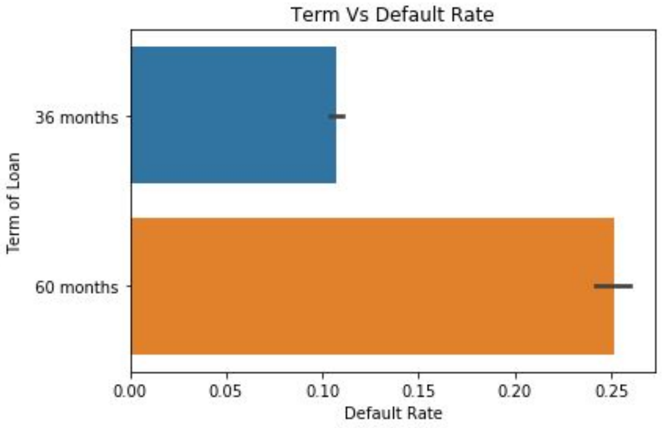
- Got a high level understanding of data. For example : number of rows, number of columns type of data in columns etc. We have a total of **39717** rows and **111** columns
- Checked the percentage of missing values. Removed 56 columns since they have more than 80% of missing values.
- Reviewed and dropped more columns like url, desc etc. which are of no use to the analysis.
- Reviewed and dropped columns which has same values in all rows since they are of no use.
- Dropped customer behaviour variables (those generated after loan is approved such as next payment date, revolving balance etc.) since they are of no use to this analysis.
- Dropped rows with missing values in atleast one of the columns since we have only 7.23% such rows.
- Removed rows which are marked as "Current" since they are neither "Fully Paid" nor "Charged Off". Also, tagged them as 0 and 1 to make analysis simpler and cleaner.
- Cleaned data more by fixing Strings (example - interest rate string to float) and splitting dates into separate features of month and year.
- We now have **36847 rows and 25 columns** (including target variable) for our analysis.
- Performed Univariate and Bivariate analysis on these columns to draw inferences like predictors of default cases.

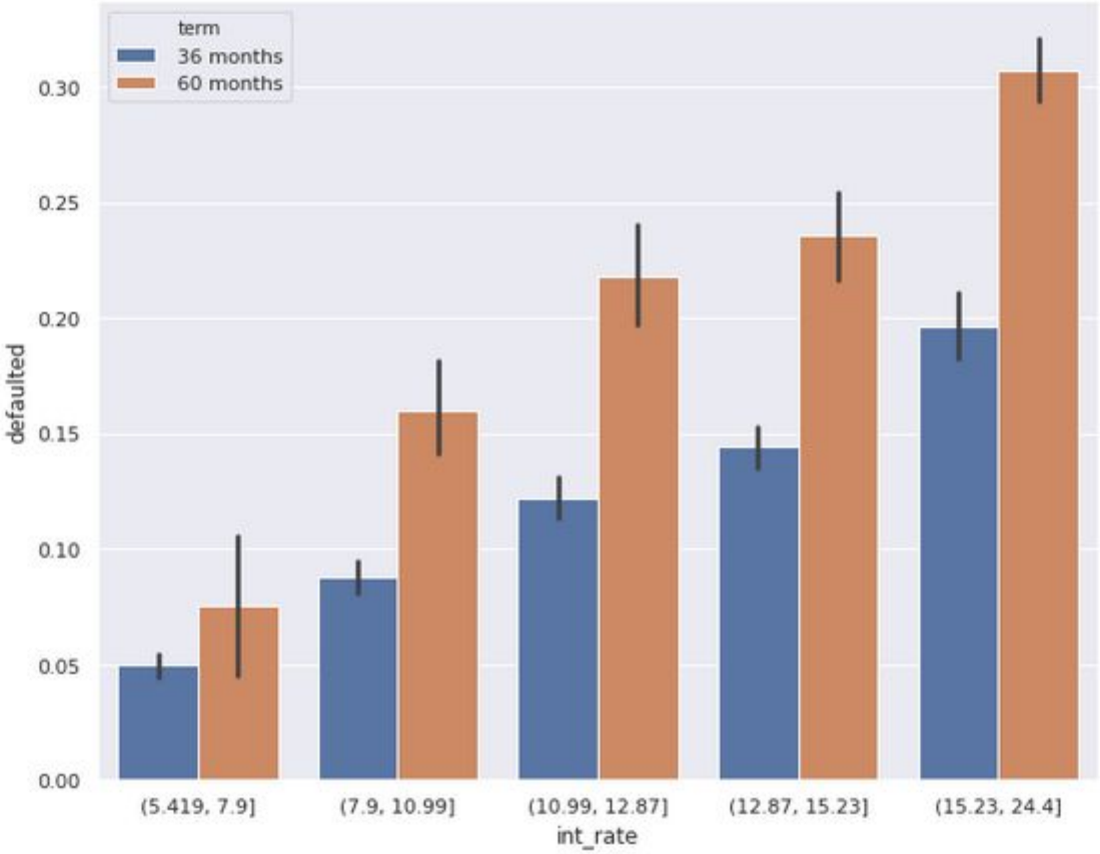
Results

Interest Rate Vs Term Vs Defaults

- As interest rate increases, the number of defaults increase
- The number of default cases are higher in 60 months term for most of the scenarios.



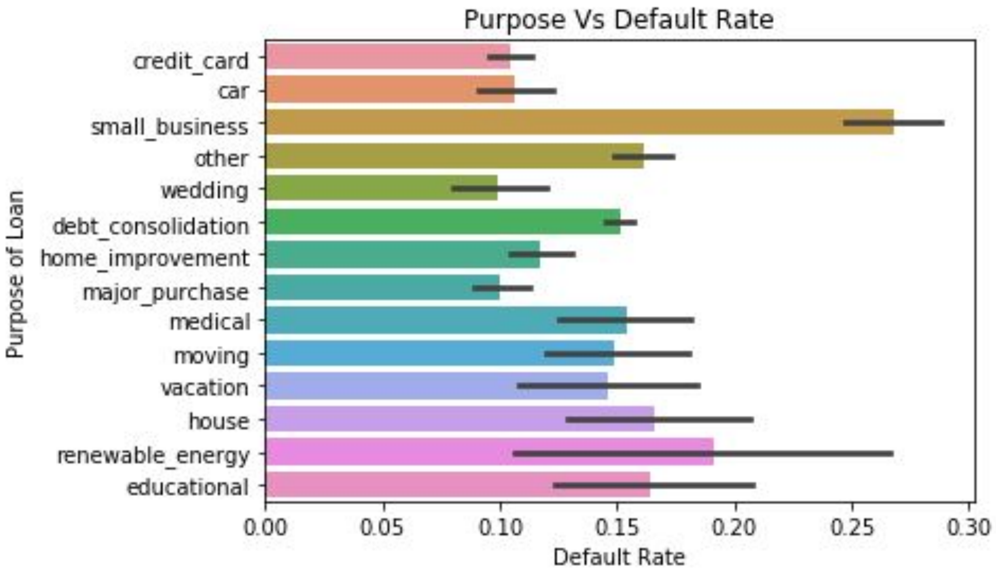


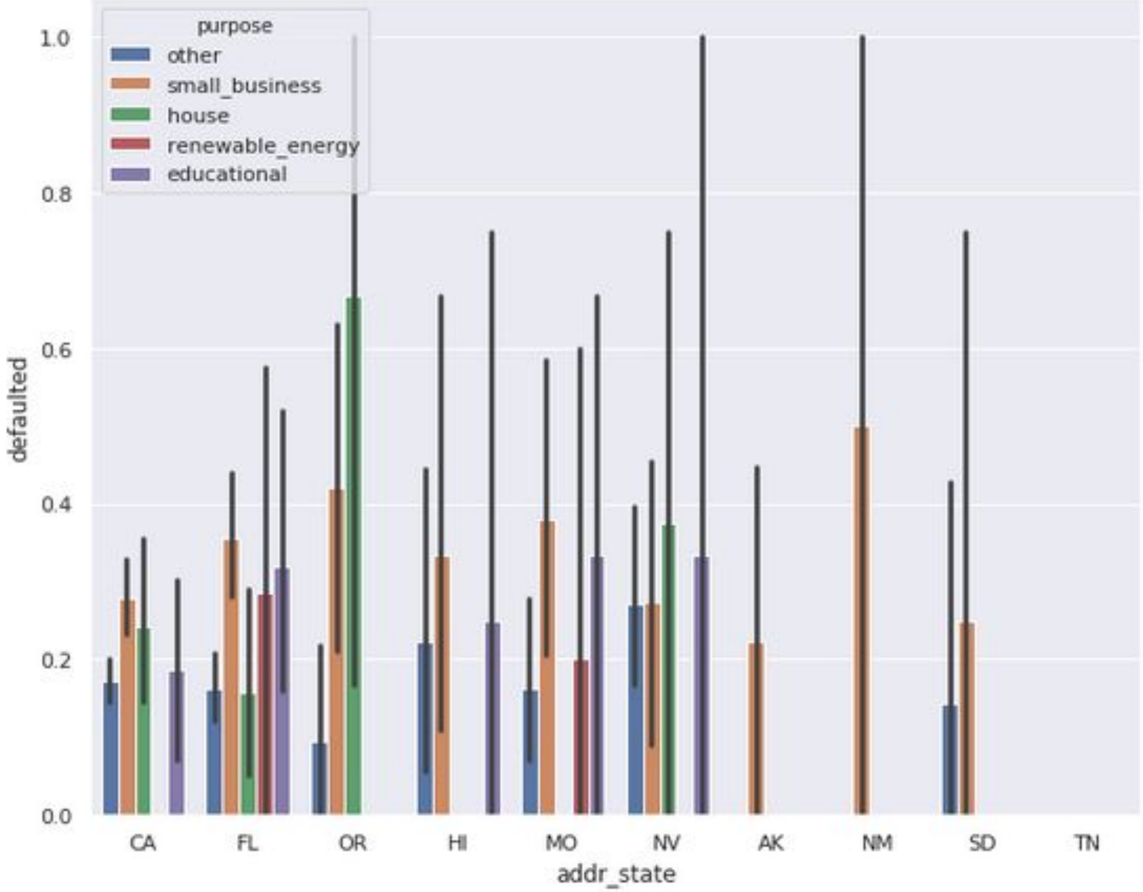


Results

Purpose Vs State Vs Defaults

- Purposes like small business, renewable energy, credit cards, etc. show spikes in defaults.
- States like NE, IN, ID, IA have no defaults whereas states like NV, AK, TN, SD, FL show spikes in defaults.

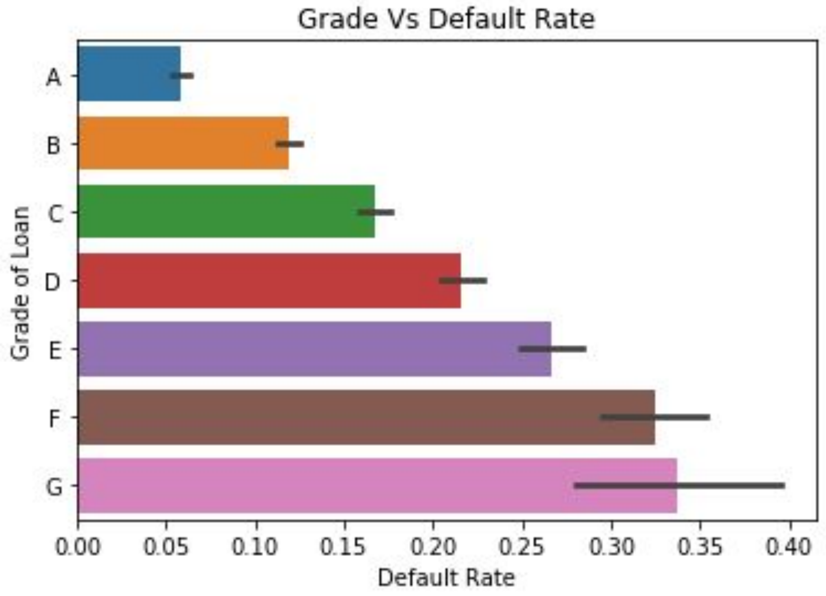


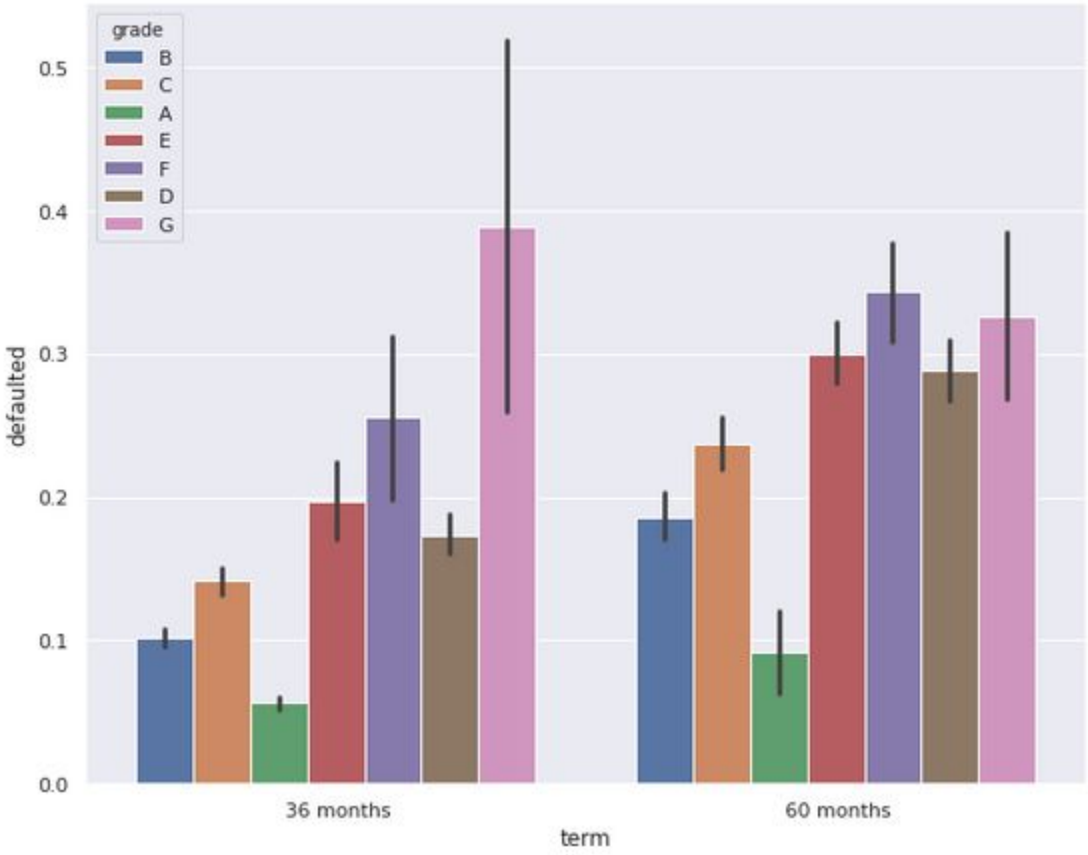


Results

Grade Vs Term Vs Defaults

- As the grade increases from A to G the risk of default increases.
- This trend is observed in both the terms of loans.





Recommendations

Predictors and Trends

- **sub_grade** - lower the better
- **grade** - lower the better
- **pub_rec** - Lower numbers to 0 are better
- **addr_state** - better if they are NE, IN, ID, IA or other states with lesser defaults
- **int_rate** - better if less than 15%
- **purpose** - Go for purposes like wedding, major purchase or other less risky ones.
- **term** - Go for 36 months since that seems less risky in many scenarios