# Lending Club Case Study

# SUBMISSION

Name: Shiyam Adhithya VP & Chetan Wanave

# Objective

- Lending Club is the company organizing this analysis to find the driving factors leading to defaulting of loan.
- The goal of this analysis is to find the driving factors behind the loan default for the company. The company then can use these factors to determine the risk factors in any loan application.
- In this case study, we will use EDA to understand how **consumer** attributes and **loan** attributes influence the tendency of default.

# Methodology

**Data Cleansing**

- Remove all the variable with 90% and above null values.
- To understand the variable and data's structure and definition.
- Select appropriate variables for the analysis.
- As the analysis regarding defaulting loans, we are removing all the current loan types.
- Convert the variable type to appropriate type for analysis.

**Univariate Analysis**

- To check the frequencies and patterns of the variables
- To create derived metrics to further actuate the analysis.
- To determine the target variable

**Multivariate Analysis**

- Based on the univariate analysis we do the determine between the co-relation between the target and other variables.
- To determine the driving factors for the loan defaults.

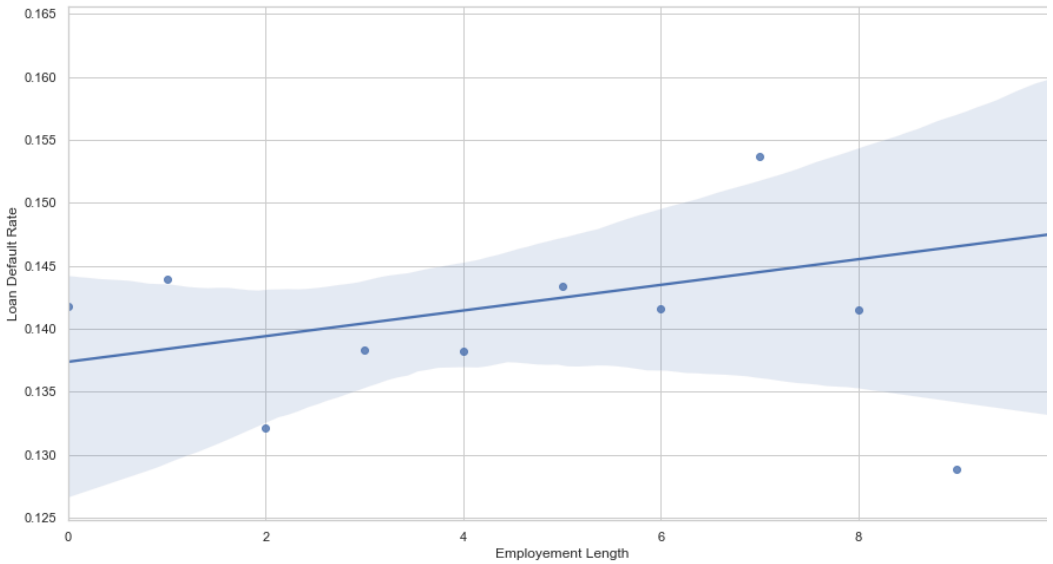
# Analysis – Data Cleansing

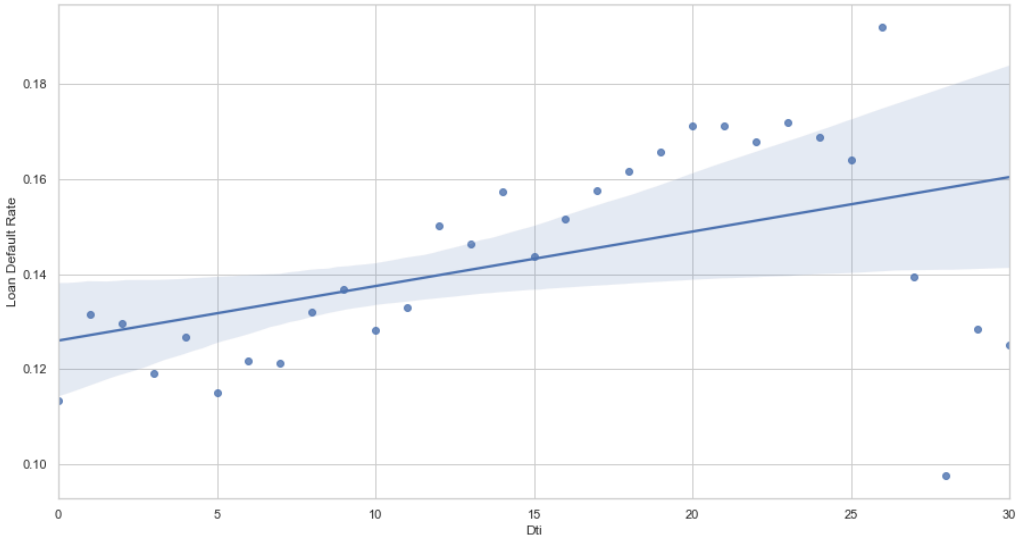
- To remove the all the variable containing more than 90% null values.
- To understand the definition and data of the remaining variables.
- To categorise the variable into Loan attributes and Applicant attributes.
- Variable not pertaining to above attributes are ignored.
- For this analysis, current loan types are not required, and thus they are removed from the dataset.
- To convert all the variable types into appropriate type for analysis.

# Analysis – Univariate Analysis

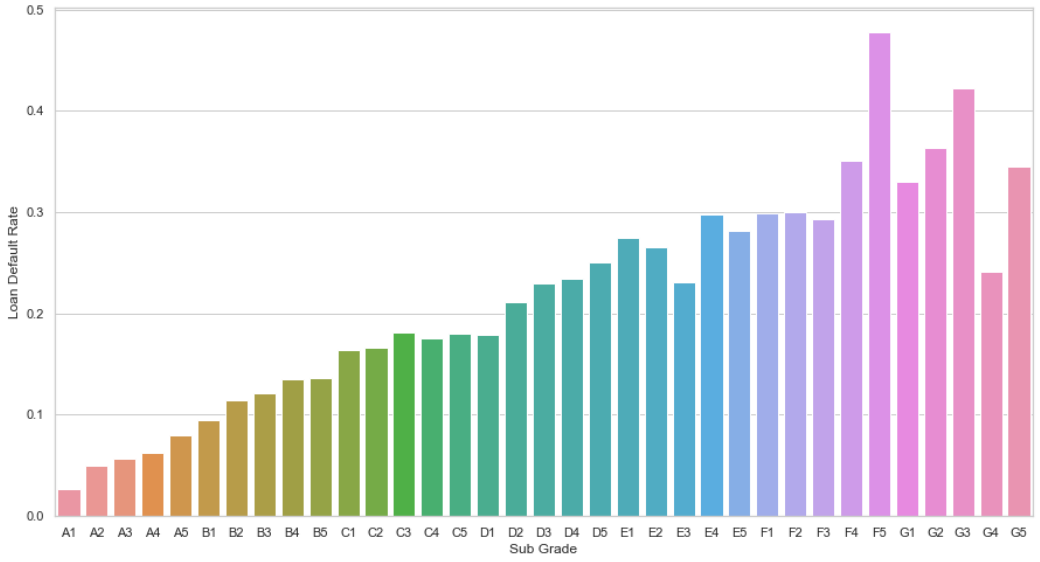
- We determined the target variable for our analysis.
- We found the different patterns and frequencies of different variable in our dataset.
- Based on analysis we formed derived variables to further help the analysis.

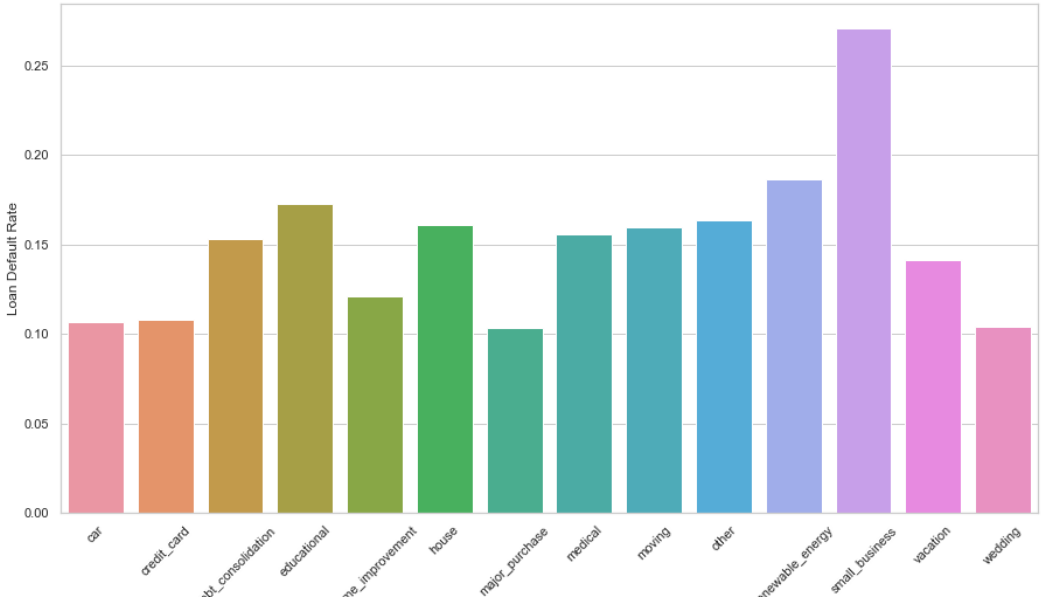
# Result – Univariate Analysis



# Result – Univariate Analysis

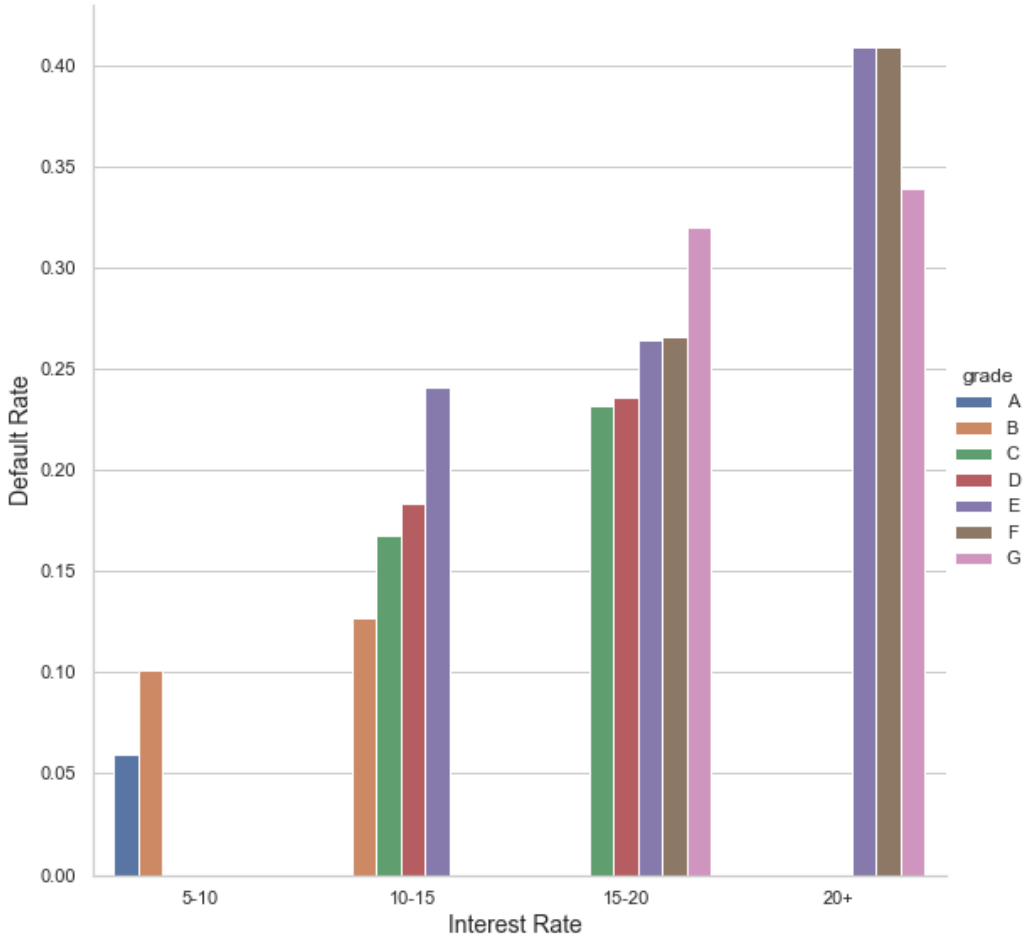


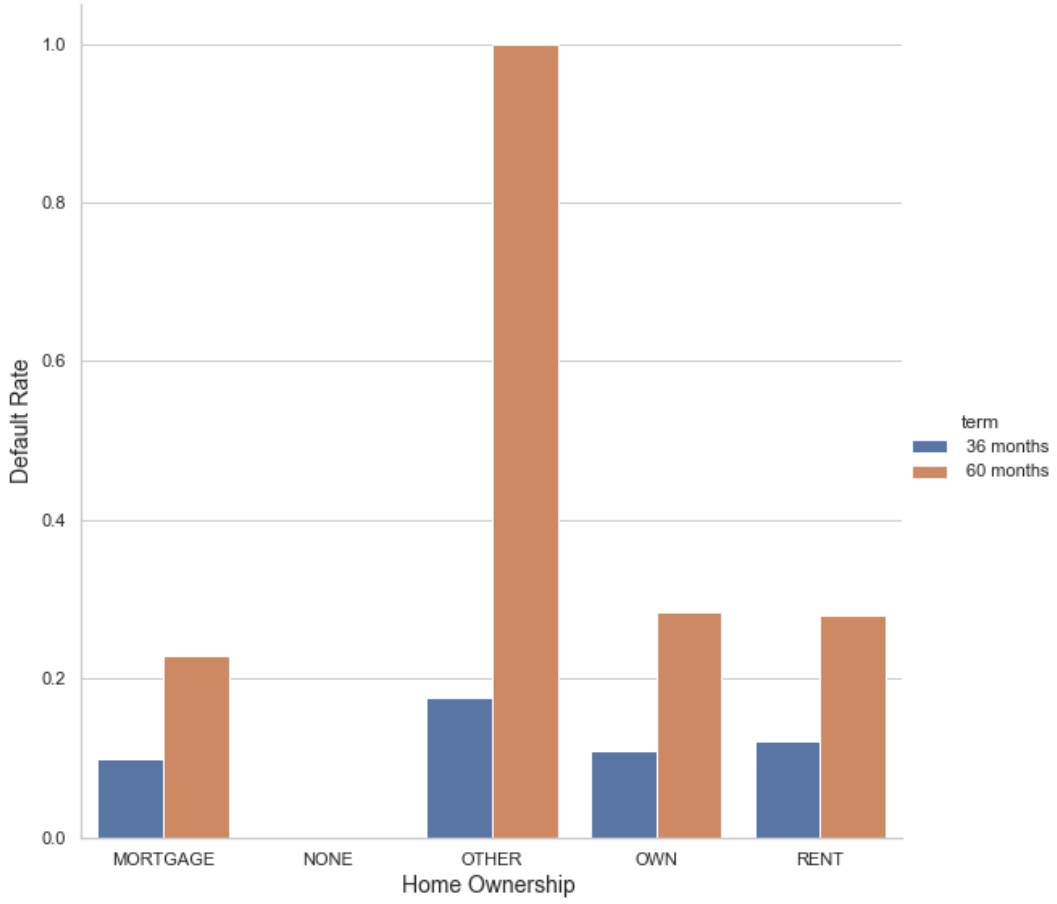



# Analysis – Multi-Variate Analysis

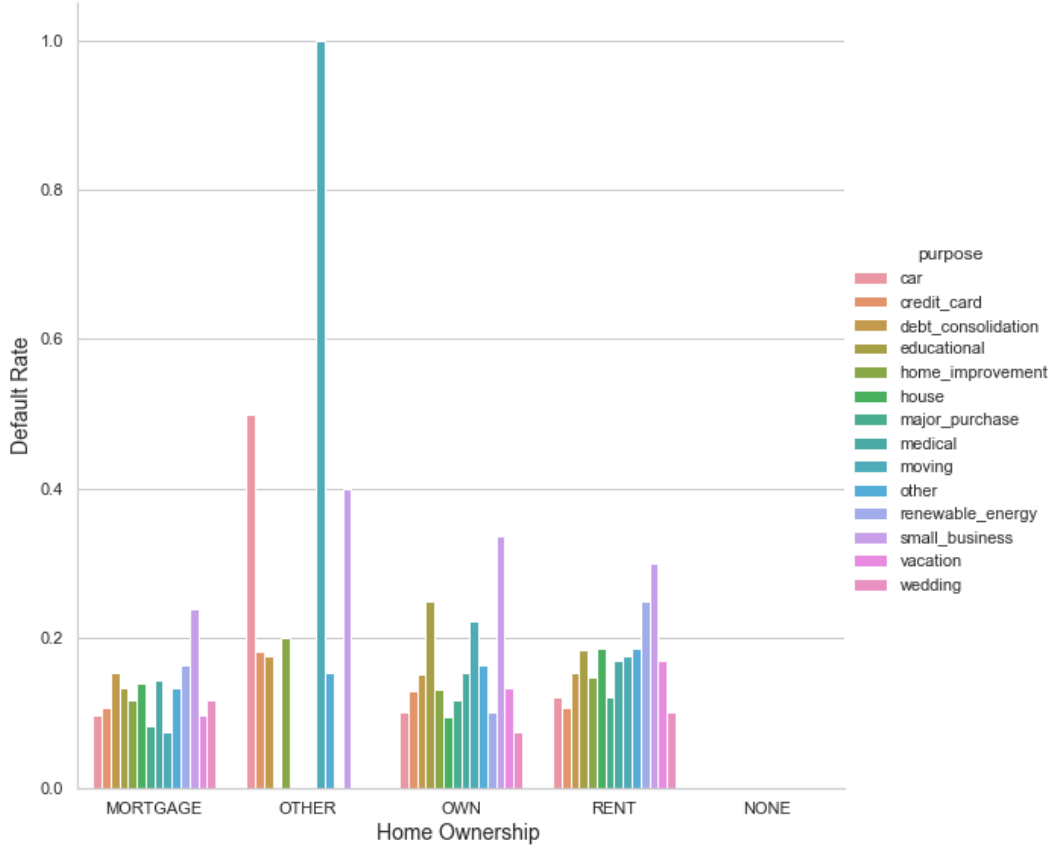
- We compared many variables with our target variable to establish the trends each variable is displaying.
- We picked up a few variables which are nominal variables, but convert them into a categorical variable to further establish pattern and for easy analysis.

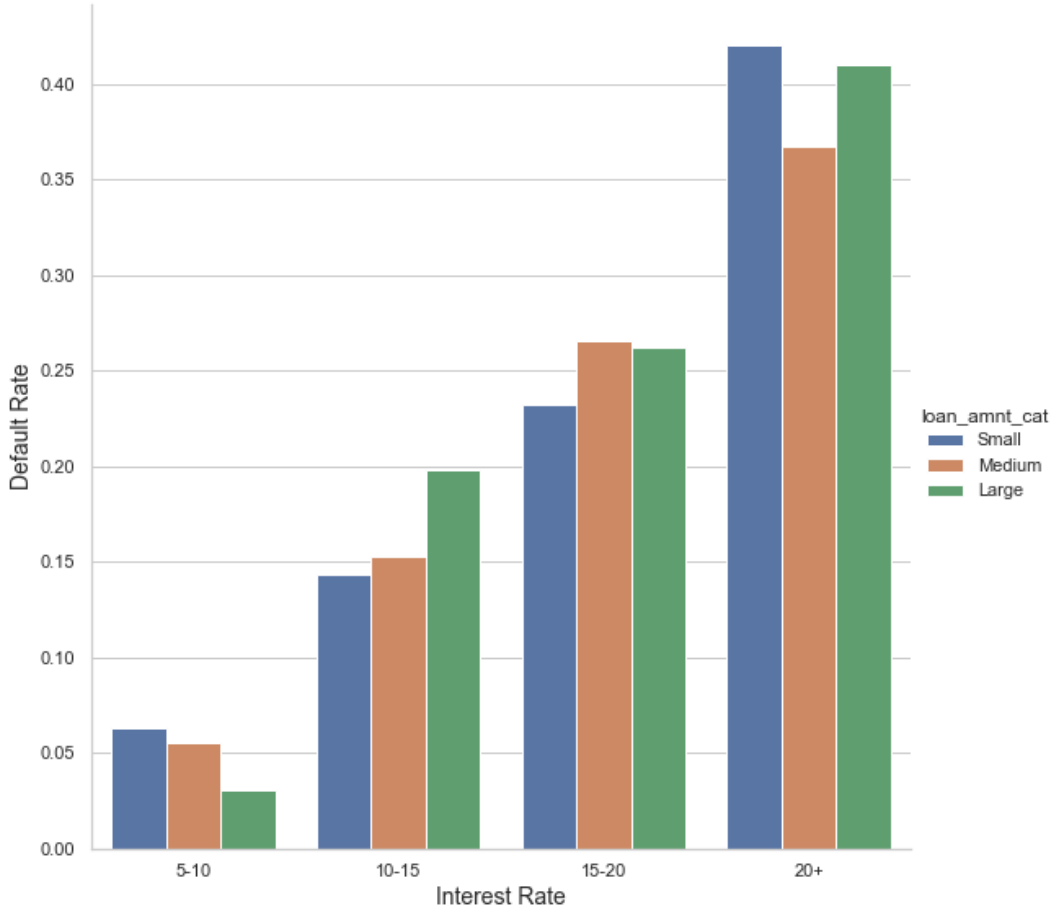
# Result – Multi-Variate Analysis



# Result – Multi-Variate Analysis



# Conclusion

- After analysis, we found various patterns and insight regarding loan defaults as well as loan and applicant attributes.
- The following are variable, we consider to be driving factors that are leading to load default:
  - Home Ownership status
  - Loan amount
  - Grade and Sub- Grade
  - Interest rate
  - Purpose
  - Loan term