

Lending Club Case Study

Suprasad Mutalik Desai

Content

- Problem Statement and Goal of this Case Study
- Overview of Data Analysis approach
- Univariate Analysis and Observations
- Segmented Univariate Analysis and Observations
- Bivariate Analysis and Observations
- Summary of Analysis

Problem Statement and Goal of this Case Study

- Problem Statement

- Online lending company wants to access the “Risky” applicants before lending loans
- Lending loans to such “Risky” applicants can lead to huge Credit loss

- Goal

- Using Data Analysis, identify strong indicators of default which can be used for Risk Management
- Basically identify driving factors behind loan default

Overview of Data Analysis approach

- Dataset used : “loan.csv” file
- Data Analysis approach
 - Load the Dataset
 - Perform Data Cleaning
 - Fixing rows and columns
 - Remove columns which are not useful for Data analysis ex : url, desc
 - Remove columns which have same value throughout the column .i.e. nunique() is 1.
 - Remove columns which has not much characteristics to derive any meaningful data ex: emp_title
 - Missing Value Handling
 - Remove all the columns which has missing values > 60% of the overall elements in that column
 - Data Standardization
 - Had to be done on multiple columns for ex: Columns term remove "months" , int_rate remove "%", emp_length remove "years", "+", "<" , zip_code remove "xx" , revol_util remove "%",
 - Perform EDA – Univariate, Segmented Univariate and Bivariate Analysis

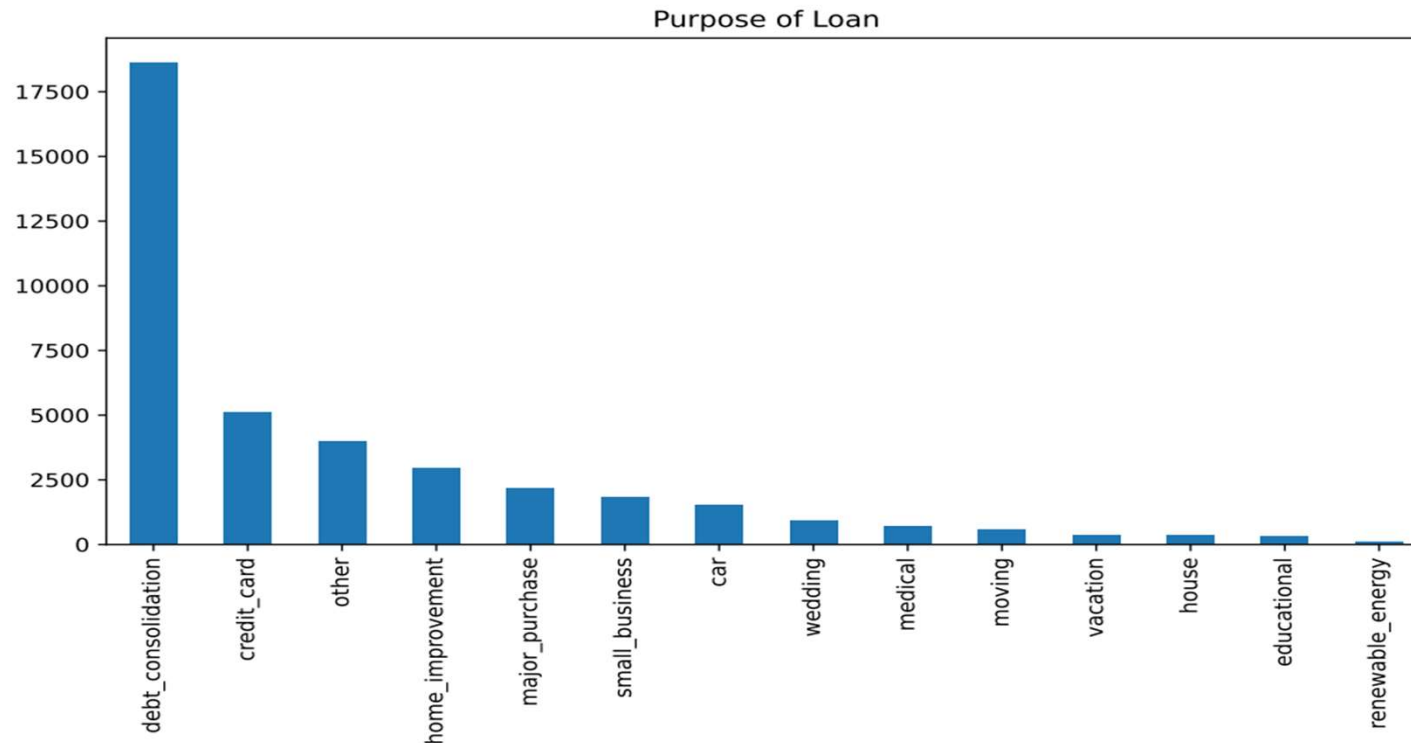
Univariate Analysis and Observations

Overall Loan Status Distribution b/w 2007 - 2011



- Observations
 - 83% of applicants have Fully Paid the Loan
 - 14.2% have defaulted on their loan which is a Credit loss
 - 2.9% have ongoing Loan

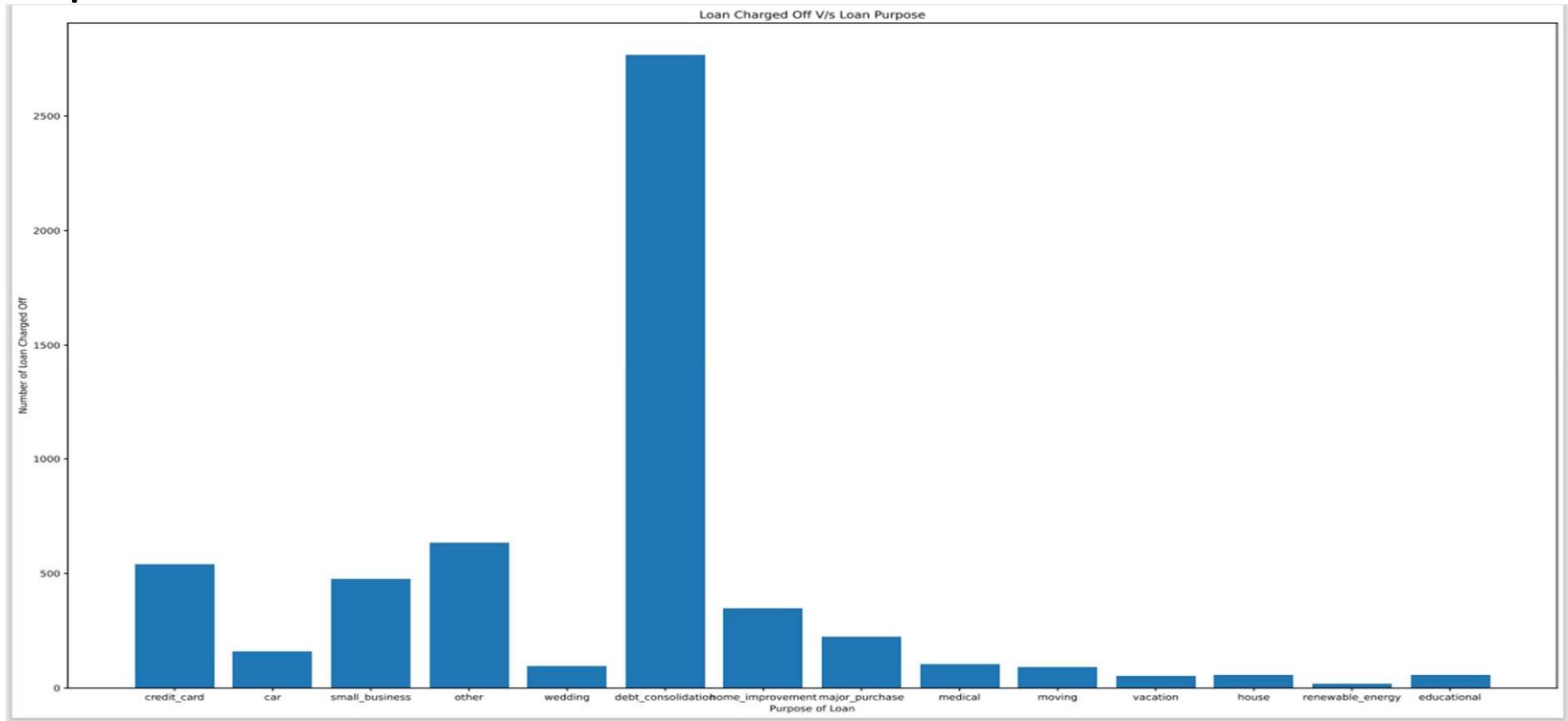
Purpose of Overall Loan Distribution



- Observations
 - TOP 3 reasons of taking loans are Debt Consolidation, Credit card and other
 - Question is, Does the dataset also show same pattern w.r.t defaulters V/s Purpose of availing loan ?
 - “Other” Category needs to be further detailed out as this doesn’t give more details . This data is missing in the dataset

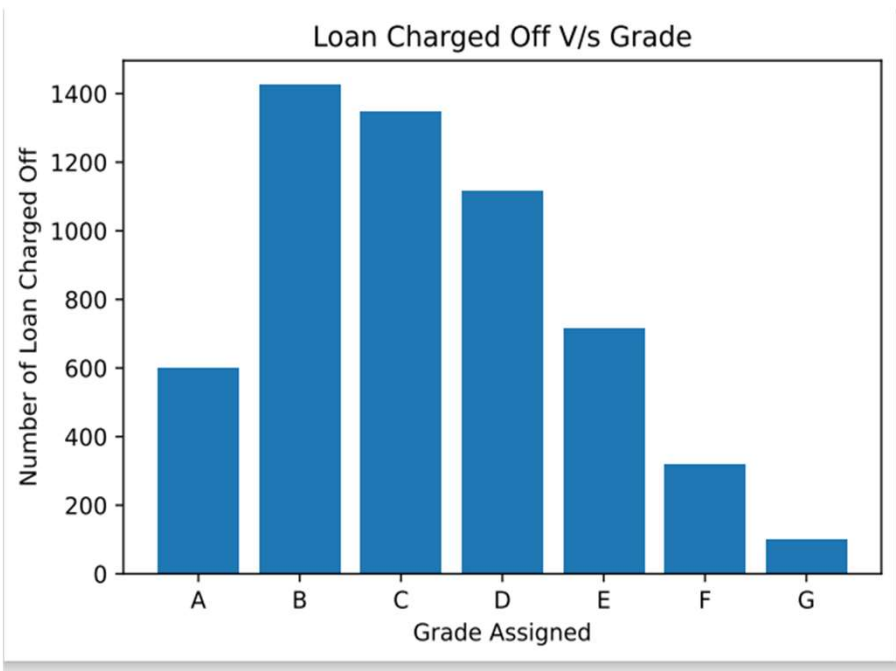
Segmented Univariate Analysis and Observations w.r.t Defaulters

Purpose of Loan distribution w.r.t Loan Defaults



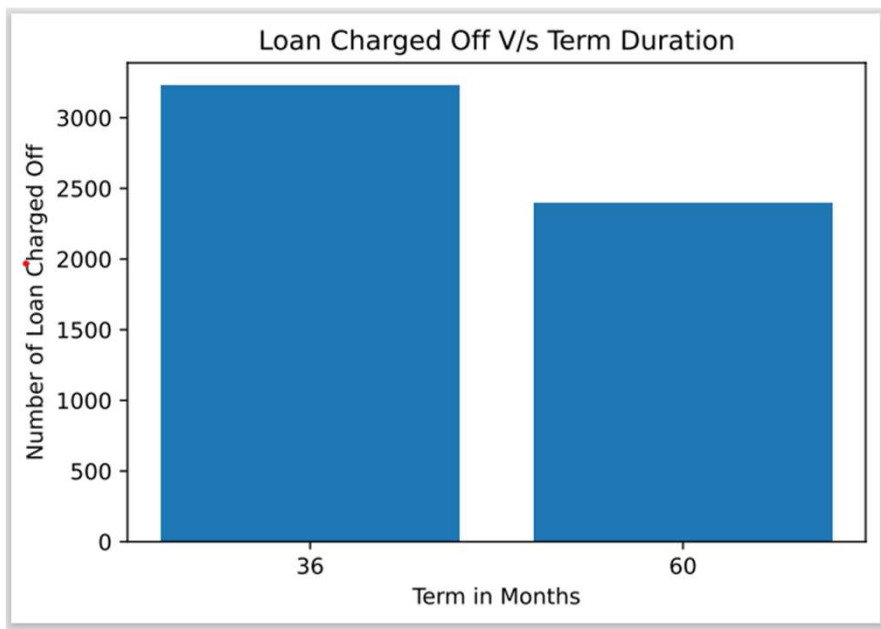
- Observations → Purpose of the loan can be a good indicator of Defaults
 - Top 3 High Risk purpose of Loan Defaulters are Debt Consolidation, Credit card payments and Others
 - Point to Note is Debt Consolidation is the biggest contributor
 - Secondly MUST need to define "Others" Purpose as it's not clear.
 - Low Risk 3 Purposes are Vacation, Education , House and Renewable Energy

Grade distribution w.r.t Loan Defaults



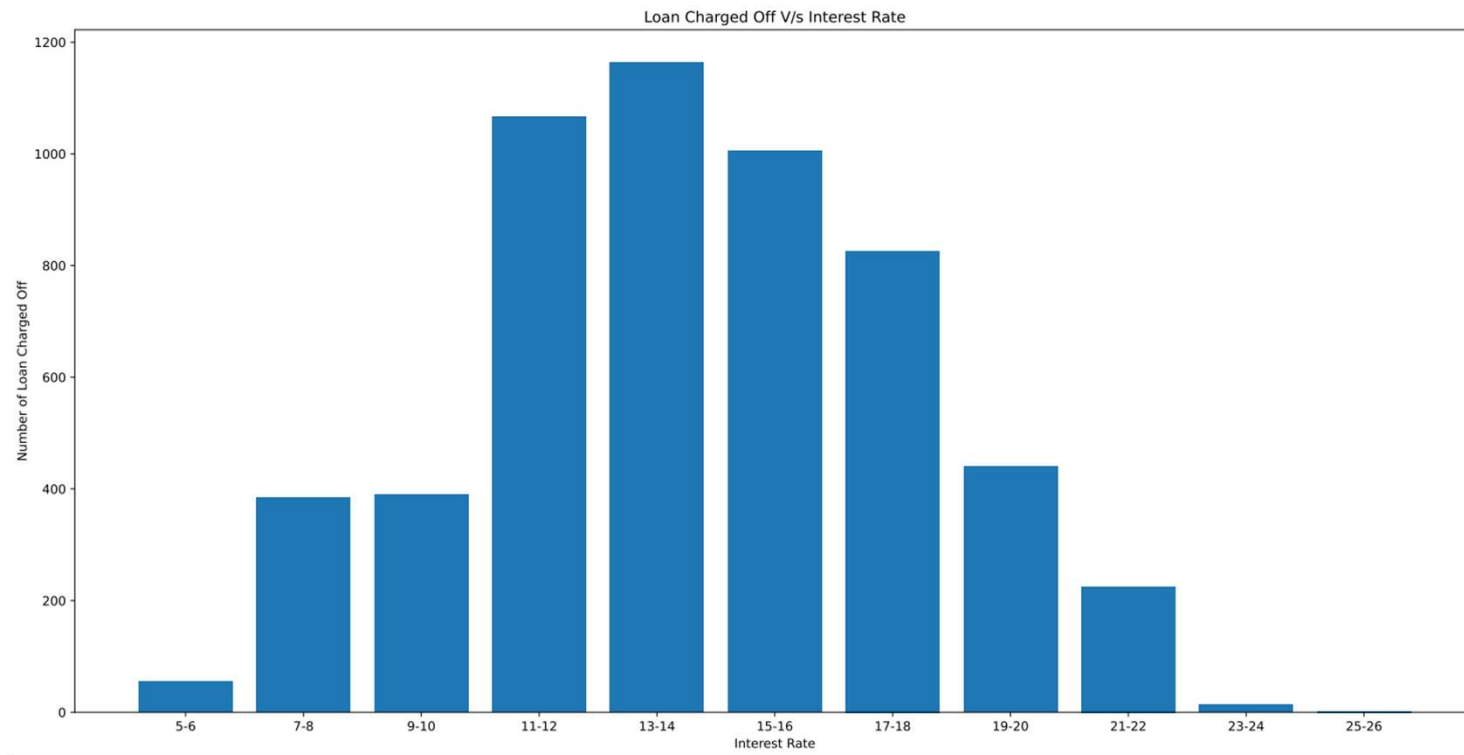
- Observations →
 - Although Data shows the Number of Defaulters decrease from B -> G grade with highest number of Defaults are in Grade B and C .
 - Data may not be conclusive as Grade B and C are theoretically lesser risk but seems to be having higher defaults so there may be other factors are contributing

Term distribution w.r.t Loan Defaults



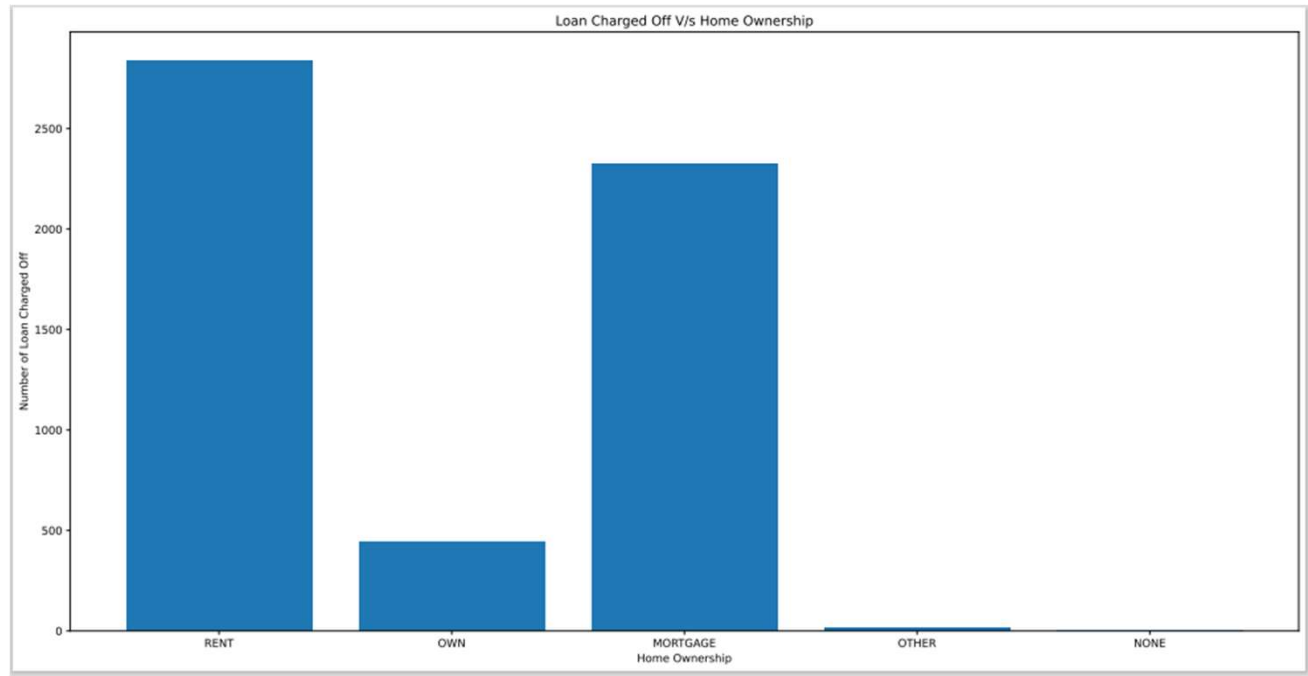
- Observations → loan term can be a good indicator of Defaults
 - More Number of Loan Defaults are observed with Lesser Term Durations .
 - So offering a higher term duration is better to avoid Loan Defaults

Interest Rate distribution w.r.t Loan Defaults



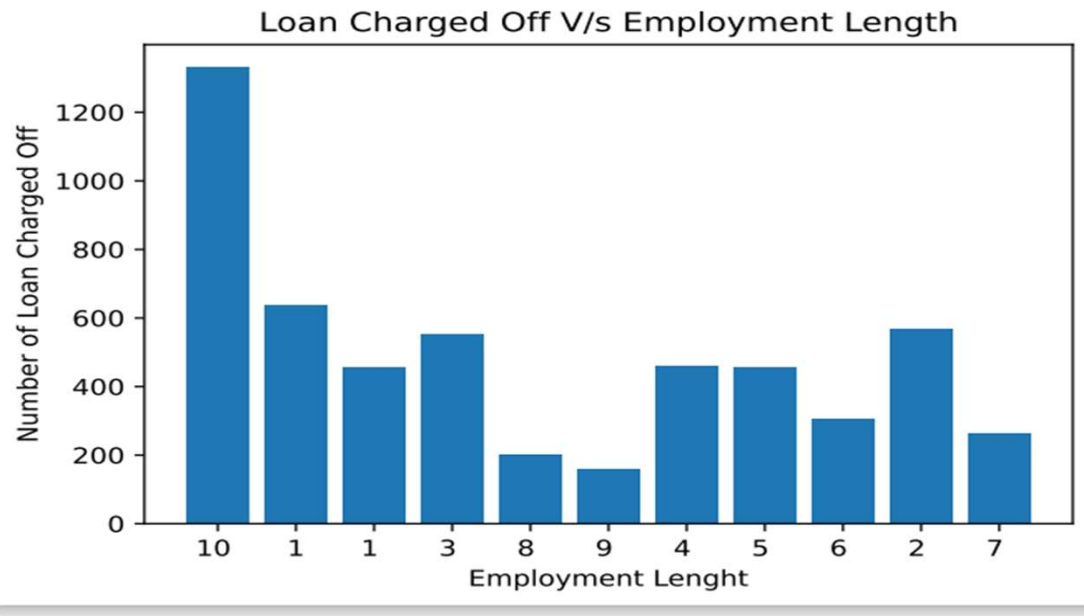
- Observations → Interest Rates of the loan can be a good indicator of Defaults
 - For clear understanding have grouped the interest rates in buckets of 2% each
 - More Number of Defaulter are observed with Interest Rates greater than 10% till 18% .

Home Ownership distribution w.r.t Loan Defaults



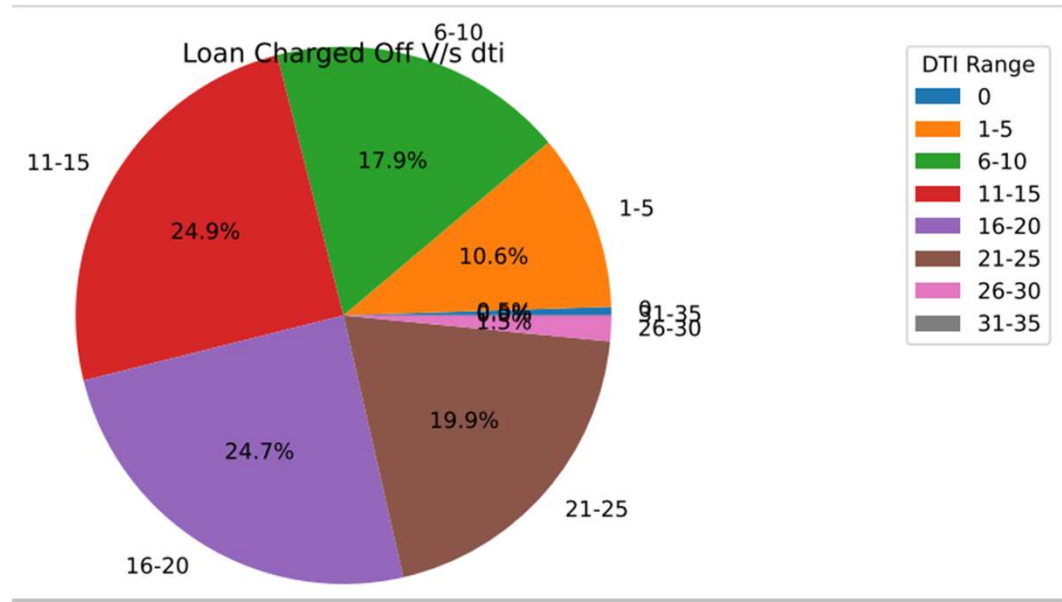
- Observations → Home Ownership can be a good indicator of Defaults
 - Data shows More defaulters are with RENT and MORTGAGE House ownership - So Higher Risk
 - Applicants with Own house have lesser tendency of defaulting.

Employment length distribution w.r.t Loan Defaults



- Observations → **Employment Length can be a good indicator of Defaults**
 - Interesting to see more number of Defaulters are observed with 10+ years and < 1years of Employment length
 - Highest defaulters are with 10+ years of experience
 - Lesser risk on defaults for employment range b/w 1 - 9 years

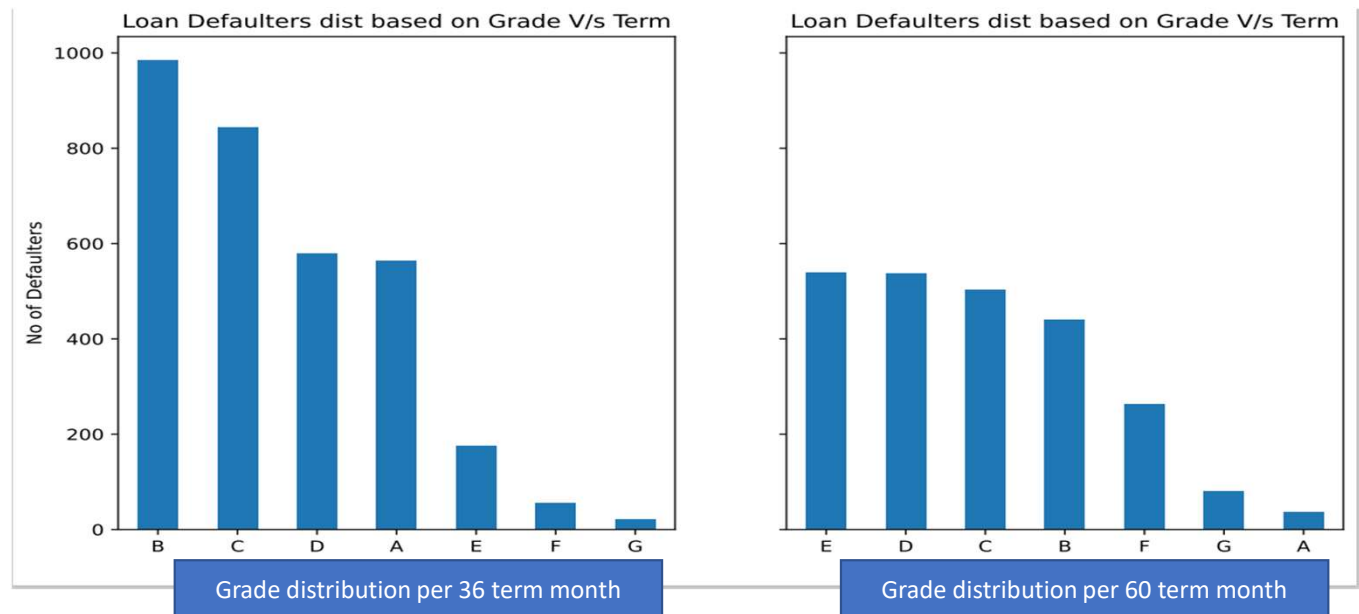
DTI distribution w.r.t Loan Defaults



- Observations → DTI can be a good indicator of Defaults
 - High probability of loan defaults with higher DTI range b/w 1-25 .
 - Have considered DTI over Income column as it gives overall debt to Income ratio which is more realistic variable to consider

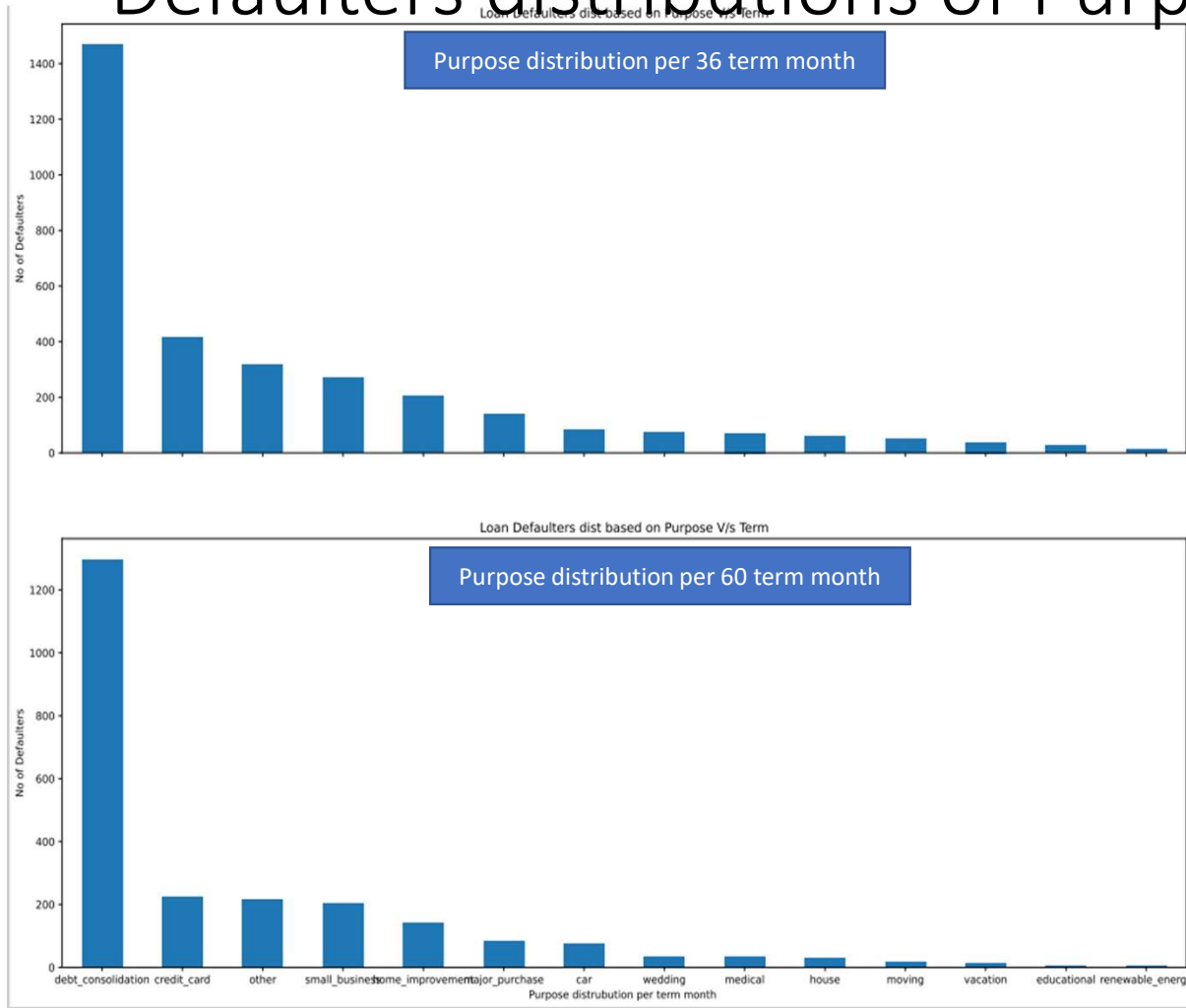
Bivariate Analysis and Observations

Defaulters distributions of Grade w.r.t Loan Term



- Observations →
 - Graph 1- Shows the Grade distribution of defaulters w.r.t Term = 36 months
 - Graph 2- Shows the Grade distribution of defaulters w.r.t Term = 60 months
 - Grade B, C, D are the TOP 3 defaulters in both Terms - 36 and 60 months

Defaulters distributions of Purpose w.r.t Loan Term



- Observations →
 - Graph 1- Shows the Purpose distribution of defaulters w.r.t Term = 36 months
 - Graph 2- Shows the Purpose distribution of defaulters w.r.t Term = 60 months
 - Debt Consolidation, Credit Card and Others Purposes are the TOP 3 defaulters in both Terms - 36 and 60 months

Summary of Analysis

- Top Driving factors for Loan defaults are ,
 - Purpose
 - Home Ownership
 - DTI
 - Employment length
 - Loan term
- Inclusive data which may need multivariate analysis
 - Grade
- Summary on loan defaults
 - Debt consolidation and Credit payment are the main reasons for default
 - Applicants in Rented and Mortgage house have higher risk of loan default
 - High probability of loan defaults with higher DTI range b/w 1-25 .
 - Applicants having >10+ years and < 1years have high probability of loan default
 - Short Repayment Term is a risk so better advice applicants for higher term