

# Lending Club Case Study

**By Anurag Singla / Prakash D**

**8 Oct 2023**

# Introduction

- **Lending Industry** - Lending is the growing industry in India every year where Companies provide loan to borrower on attractive interest rates.
- **Problem Statement** - Lending loans to 'risky' applicants is the largest source of financial loss. We need to understand driving factors to such applicants so that companies can take better decision and minimise the loss.
- **Analysis Approach** - Utilizing Exploratory Data Analysis (EDA) to Uncover Key Insights

# Initial Observations about Data Set

## Missing Values :

- There is no row with all null values
- There are columns in data set with null values
- There are columns which very few non null values and significant null values
- There are columns which single value across all rows

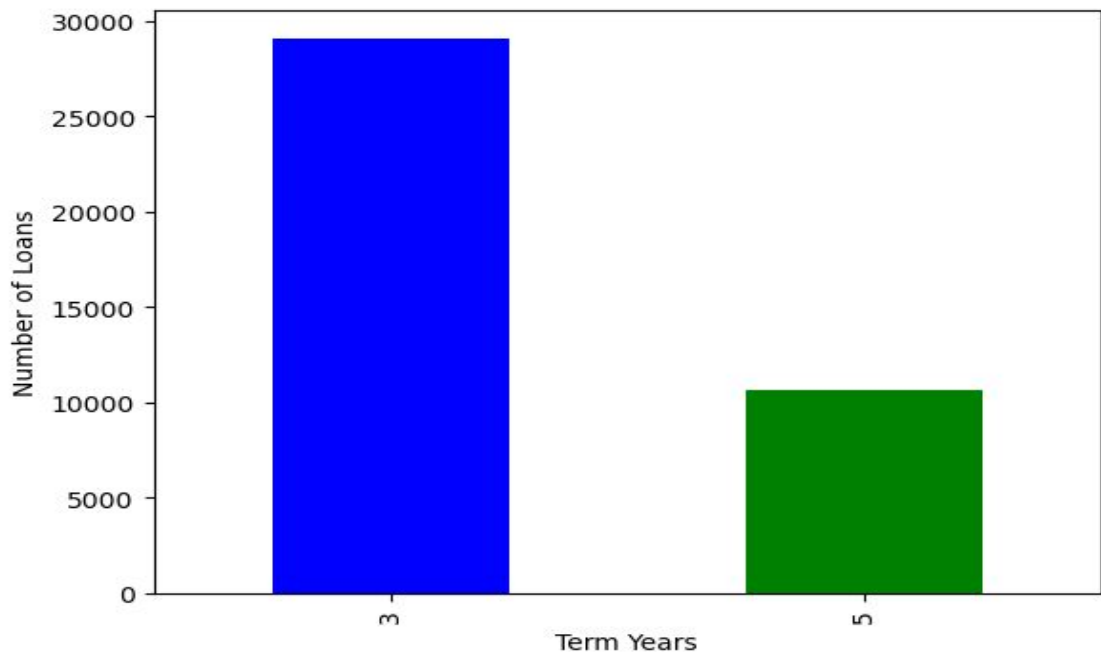
## Standard Values issues :

- prefix space should be handled on term column
- % symbol should be removed from interest rate and credit limit consumed
- Employee Length should be modified into numeric value
- Date format should be changed on issue\_d, earliest\_cr\_line, last\_pymnt\_d etc.
- round the decimal places into 2 digit for total\_pymnt column
- Zip code suffix values should be fixed

**After Data Cleaning and filtering, We are left with 24 columns and 39714 rows.**

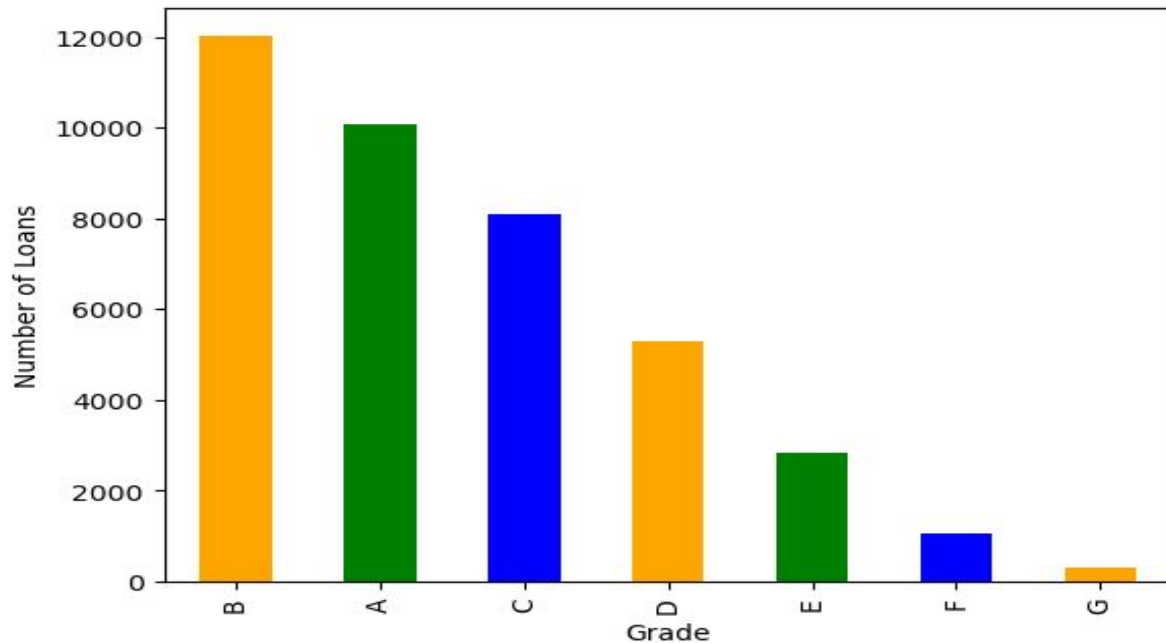
# Univariate Analysis

Number of applicants in 3 year tenure is way more than 5 years tenure



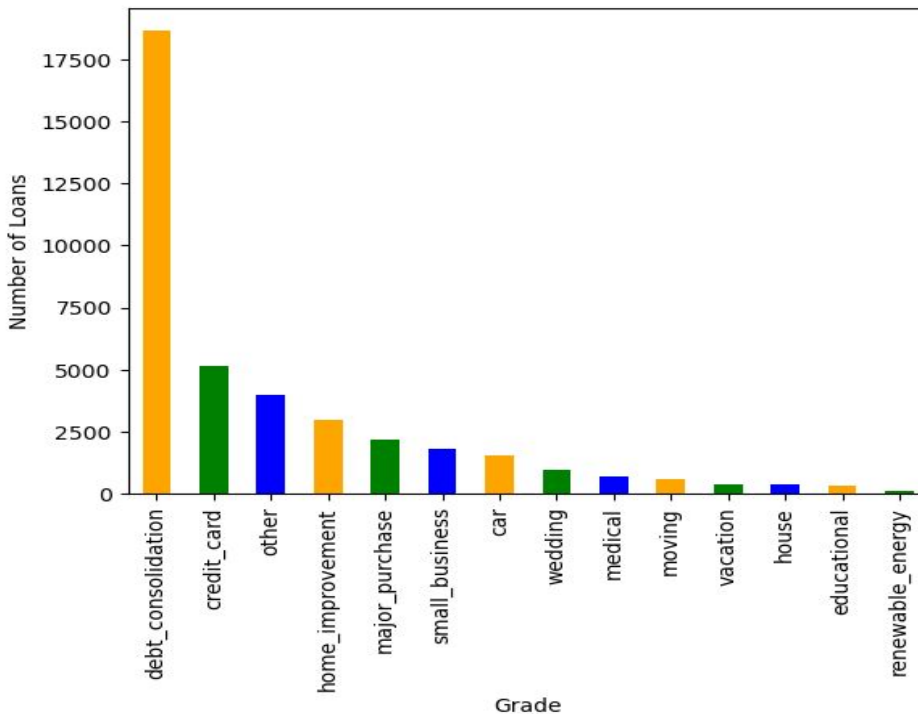
# Univariate Analysis

Surprisingly grade B has more volume than A



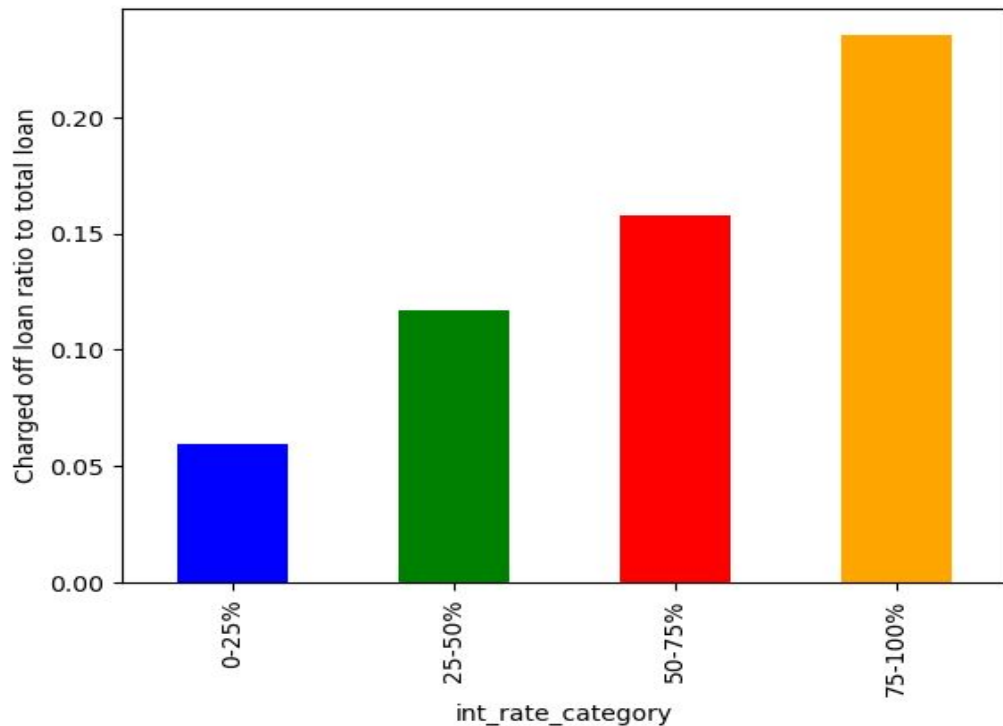
# Univariate Analysis

Most common purpose for loan is Debt Consolidation with very high volume



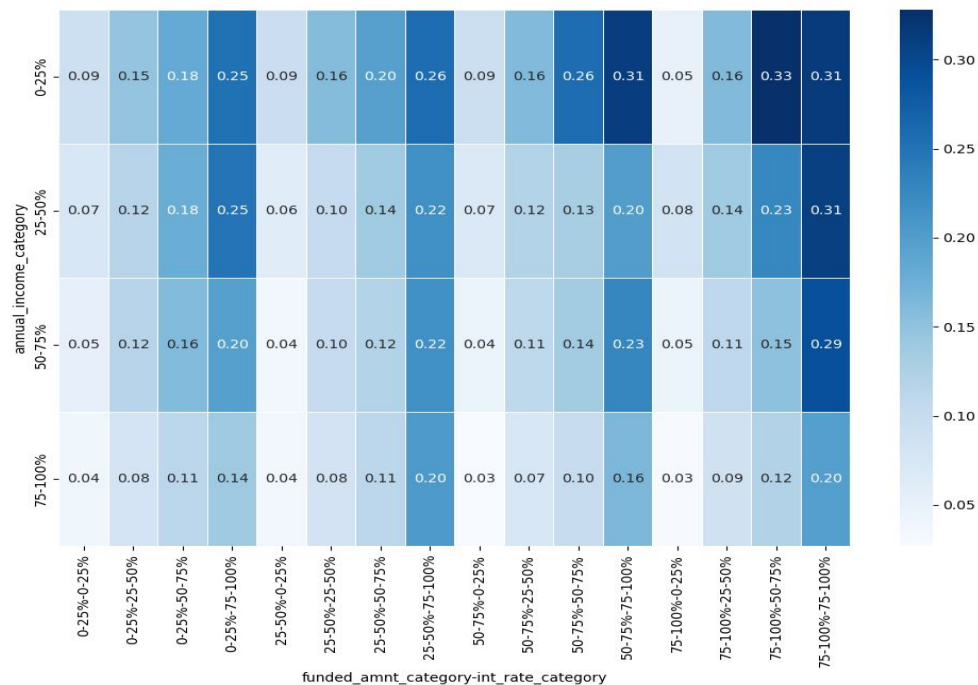
# Bivariate Analysis

Higher interest rate tends to have higher defaulters.



# Bivariate Analysis

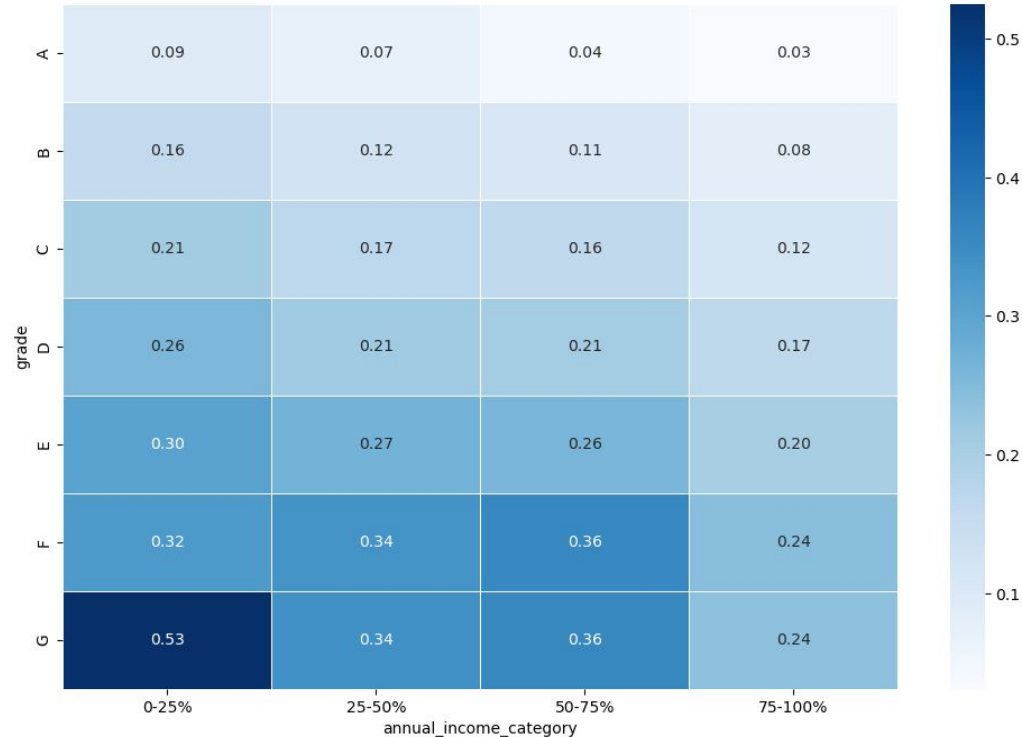
Lesser annual income with higher funded amount and interest rate tends to have higher defaulters





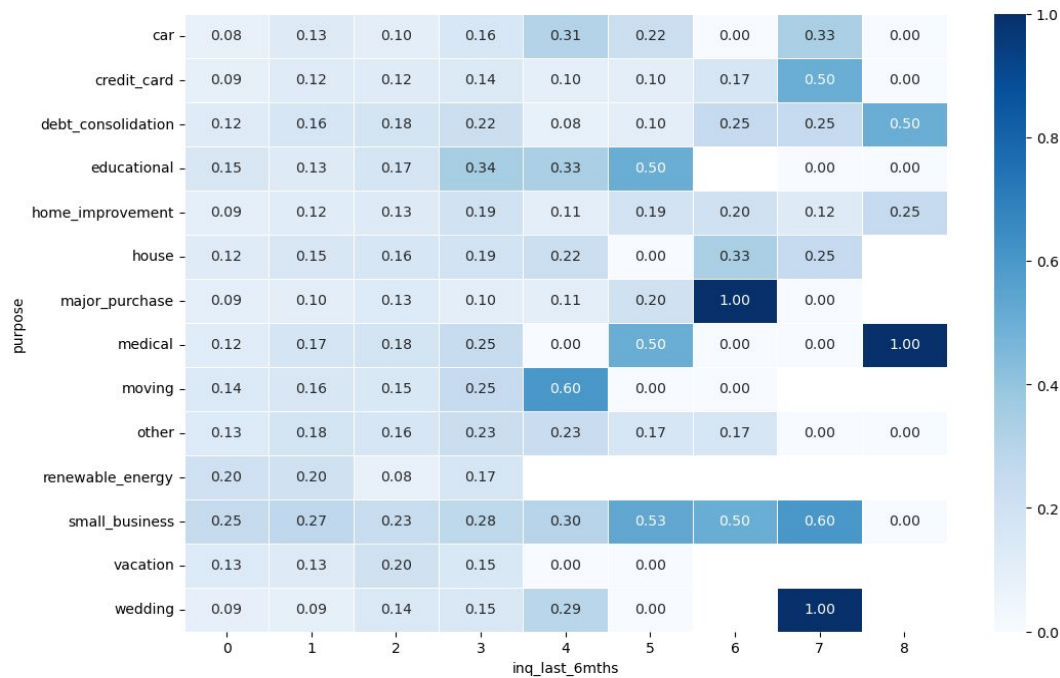
# Bivariate Analysis

**Lesser annual income and higher grade have high chances of defaulters.**



# Bivariate Analysis

Higher number of enquires in some purposes like education, small business have very high probability of defaulters.



# Results

- **Annual Income and Grade** - if the grade is higher (i.e. F,G) and annual income is less, then probability of charged off is higher
- **Number of Enquiries** - if number of enquiries are higher is number, then chances of charged off increases
- **Interest Rate, Funded Amount, Loan Amount** - Annual income is less and interest rate and funded amount is high, then probability of charged off is higher
- **Purpose** - for higher funded amount, small business and education likely to defaulter