



# UPGRAD ASSIGNMENT

EDA AND STATISTICS CASE STUDY  
By: Arsh Dinesh Vijayvargiya

# PROBLEM INSPECTION

The given data is of a loan providing company whose purpose for sharing the data was to predict whether the loan given is going to be paid off or being 'default'.

The main objective of the analysis was to determine the conditions and situations that leads to an applicant being charged off or default.

The data set had initially 111 columns with 39716 entries.

# DATA CLEANING AND MANIPULATION

The first step for data analysis had been removing the insignificant columns and rows from the dataset.

Criterion of Insignificance:

- If a column had more than 10% of Null Values than it had to be removed as it is not feasible to impute this much of information.
- If a row had more than 30% of Null Values it had to be removed for the same reason.
- If a column has only single value then it can't be used in our analysis thus it had to be removed.

With these measures the dataset was lowered to 44 columns.

Further the columns going to be used in our analysis had to be imputed or modified such columns were employment length, interest rate etc.

New Columns were derived for better analysis.

# ANALYSIS PROCEDURE

Initially we carried out univariate analysis on our continuous variables to check the spread of values and estimating the major constituent of that variable.

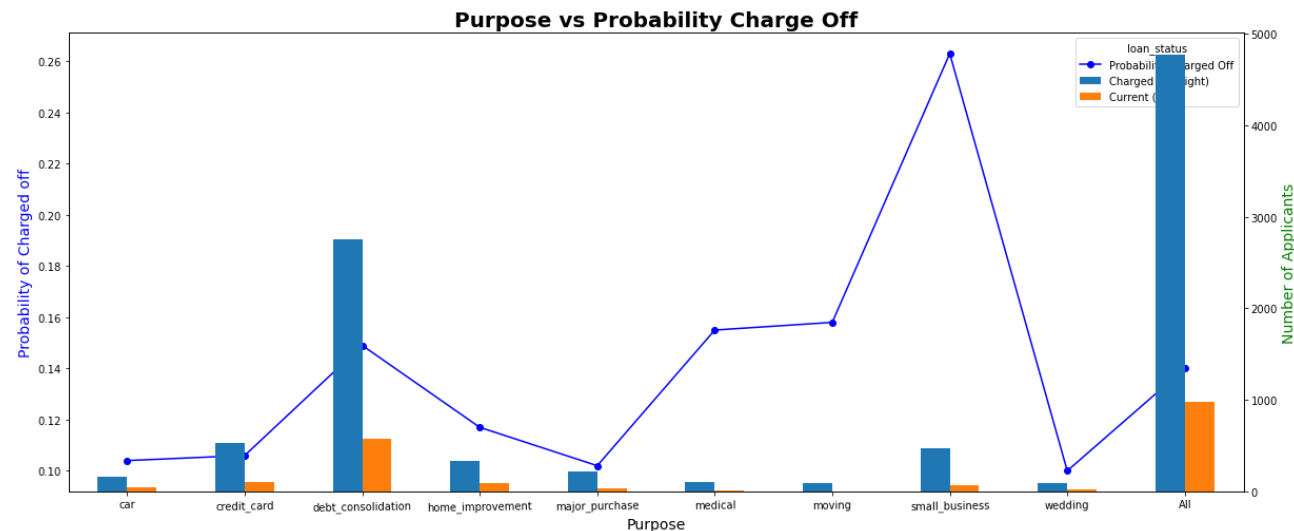
Then we moved on to categorical variables to check the defaulters in each category.

After finished with univariate analysis we proceeded with bivariate analysis.

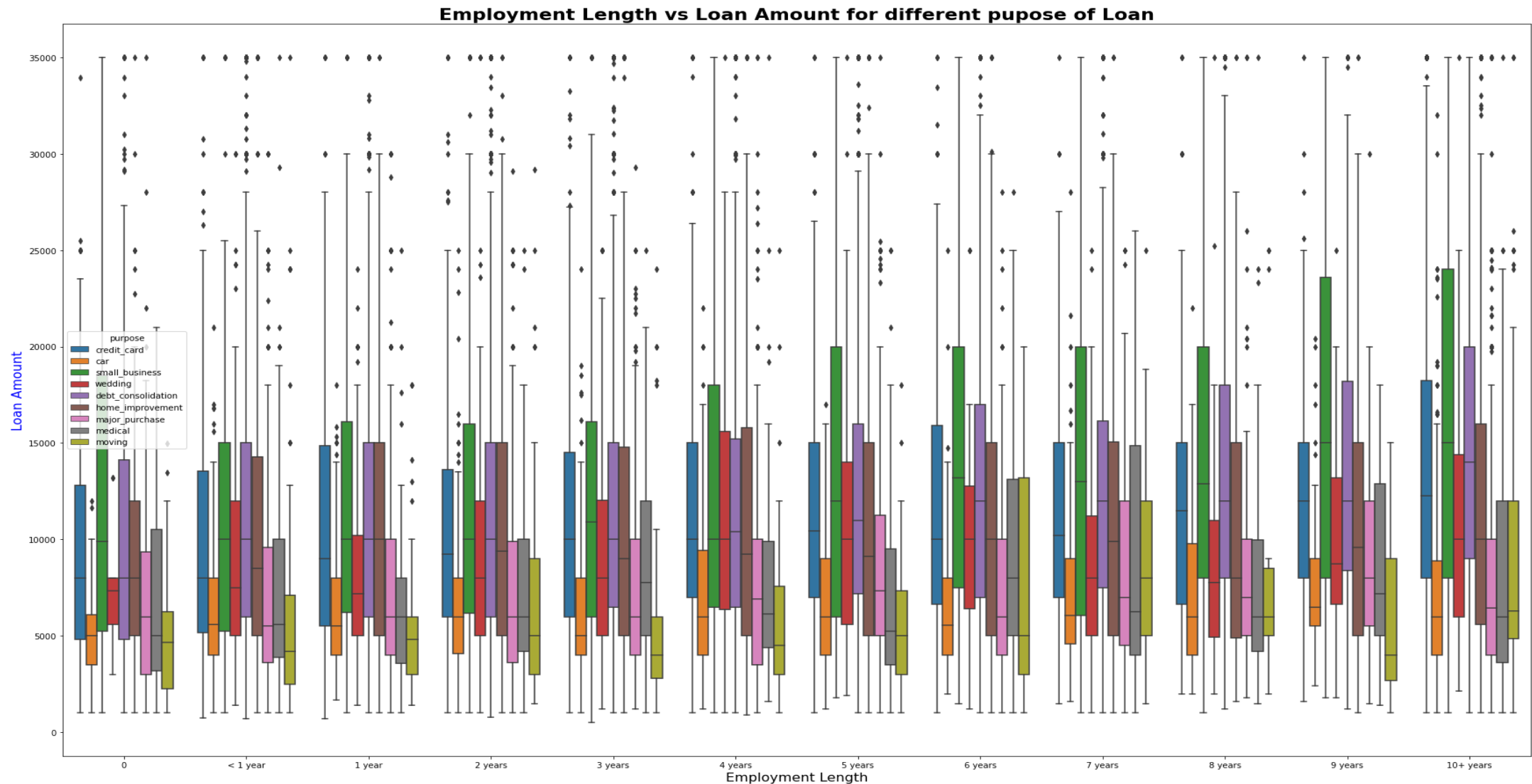
And Finally finished off our analysis by finding out the probability of applicant being charged off.

# OBSERVATIONS FROM ANALYSIS

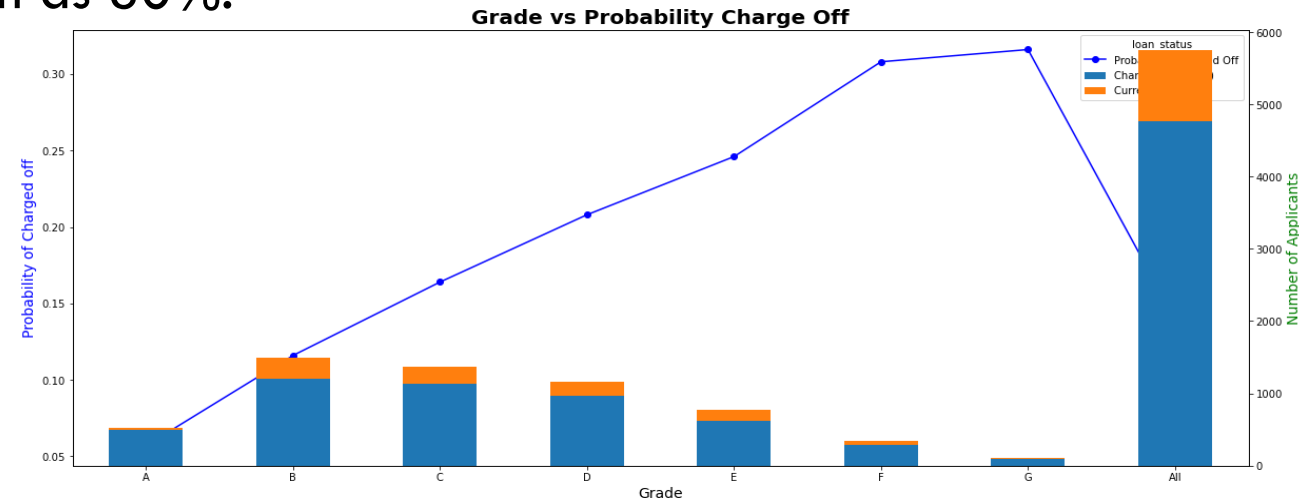
- Loans taken for the purpose of 'small business' are more riskier and even though the most significant purpose was 'debt consolidation' its probability of being charged off was quite less than others. Thus special cautions should be taken for applicants applying for loan under 'small business' purpose.



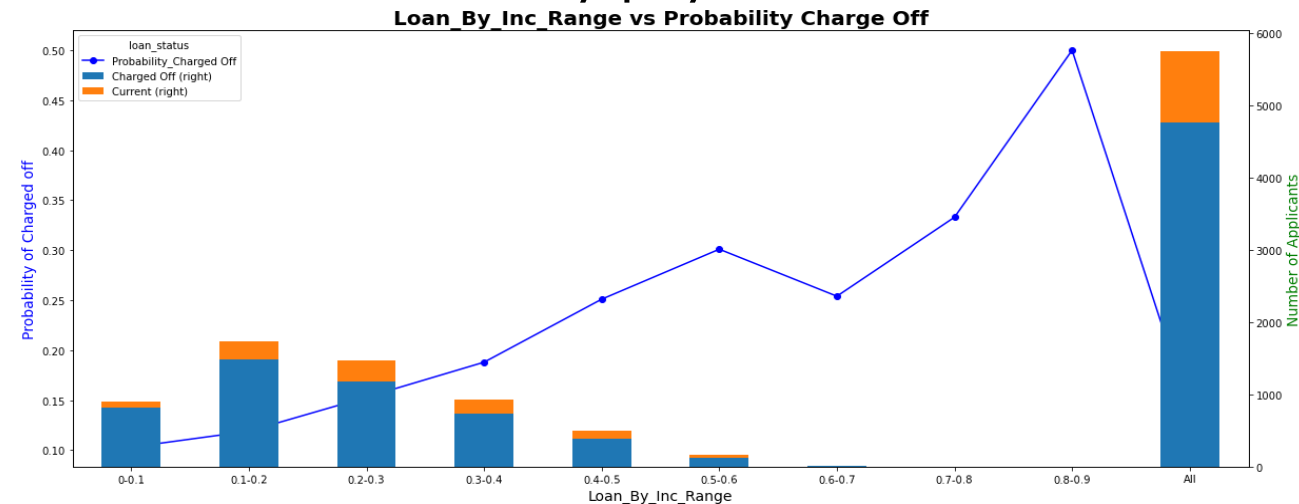
- The graph shows priority of interest of various applicants throughout their employment term. It is quite visible that as employment length increases the loan amount has also increased. Loans for small business has also increased with seniority level with exception of '0' label as it includes self-employed.



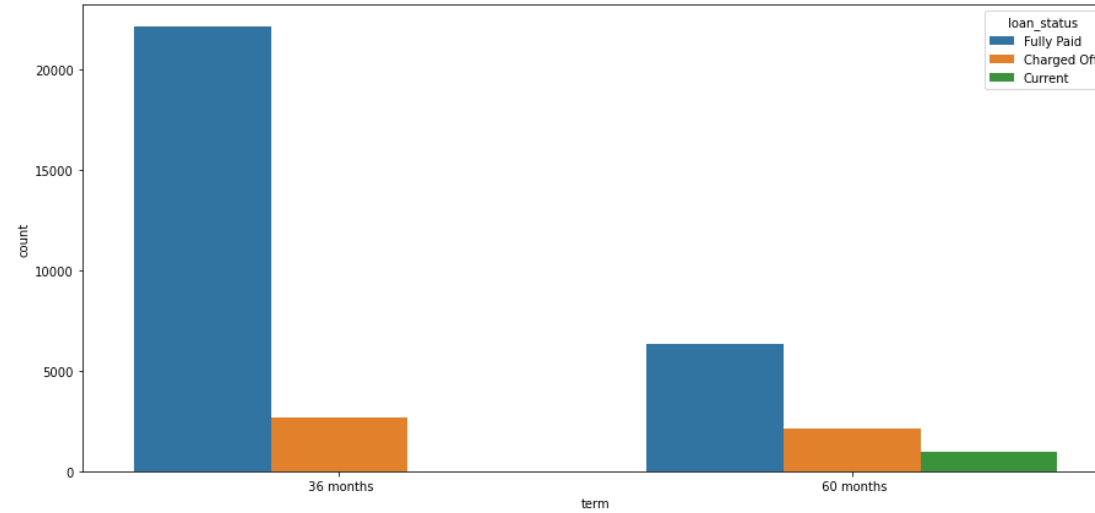
- Furthermore as we see drop in grade the probability of it being charged off increases as high as 30%.



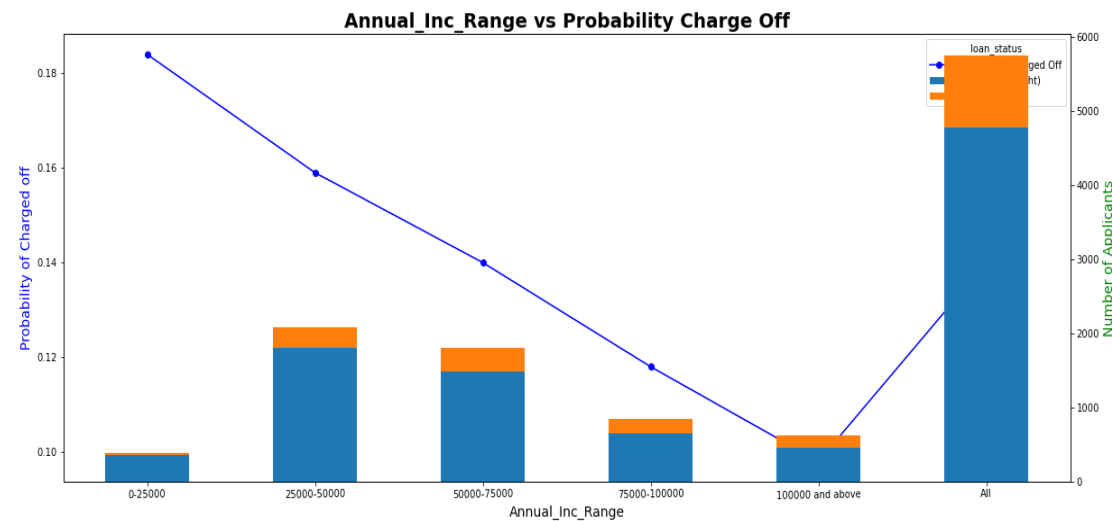
- It is observed that when loan amount is in range of 1%-30% of annual income of the applicant, he/she is more efficient to fully pay off the loan



- This shows that loans of longer term are more tend to be charged off than short term

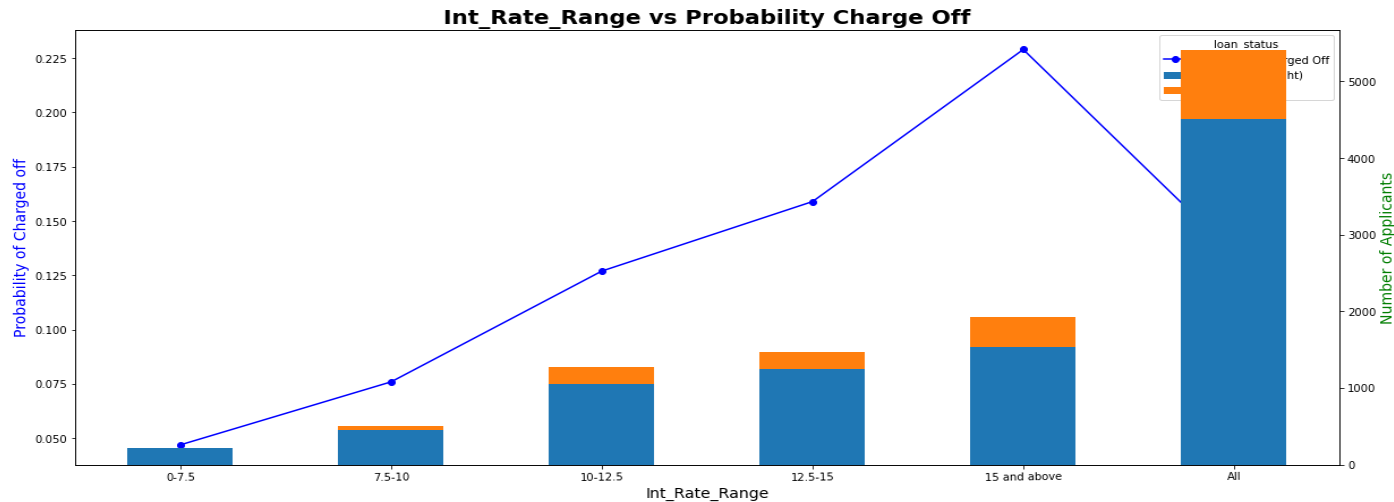


- As expected, lower income applicants are more probable to be charged off.





- The Interest Rate also affects the probability of applicant being discharged, lower the interest easier for applicant to fully pay the loan amount thus reducing the probability of being charged off



The major deciding factors for an loan to be approved are :

- Annual Income
- Loan By Income Ratio (Derived)
- Grade
- Interest Rate
- Purpose