LENDING CLUB CASE STUDY

EXPLORATORY DATA ANALYSIS

Group Members:
Pratik Patil
Manish Kumar

Problem Statement

Company Profile:

The Lending Club is one of the largest online marketplace for the loans. Consumer can easily apply for the loans in various categories. The loans are funded by Investors based on the Credit ratings of the Customers.

Problem:

Lots of loans are Defaulted and Charged Off. The company wants to understand the driving factors behind loan default. The company can utilise this knowledge for its portfolio and risk assessment to make better decisions.

Objective:

As a Data Analyst, perform EDA to understand how consumer attributes and loan attributes influence the tendency of loans default.

Data Availability

The data available contains the information about past loan applicants and whether they 'defaulted' or not. The dataset does not contain any information for the loans which were 'NOT Approved'.

The provided dataset can be analyzed using Exploratory Data Analysis (EDA) technique.

Datasets provided:

- loan.csv
  The historical data for the approved loans which have either been fully-paid or Charged Off. It is in comma separated value (csv) file format

- Data_Dictionary.xlsx
  Contains information about the various columns. It is required for Data Understanding. It is in excel format.

Analysis Approach

Data Cleaning

- Drop columns with all null values/ single categorical value / random values
- Impute null values for important columns
- Convert values into proper data.

Derived Metrices

- Create derived Metrices for year and month for date.
- Convert Quantitative values into categories.
- Calculate ratios and percent and other derived metrices.

Univariate Analysis

- Check distribution and frequencies for numerical and categorical data
- Analyze the variable against segments of other variables.

Bivariate Analysis

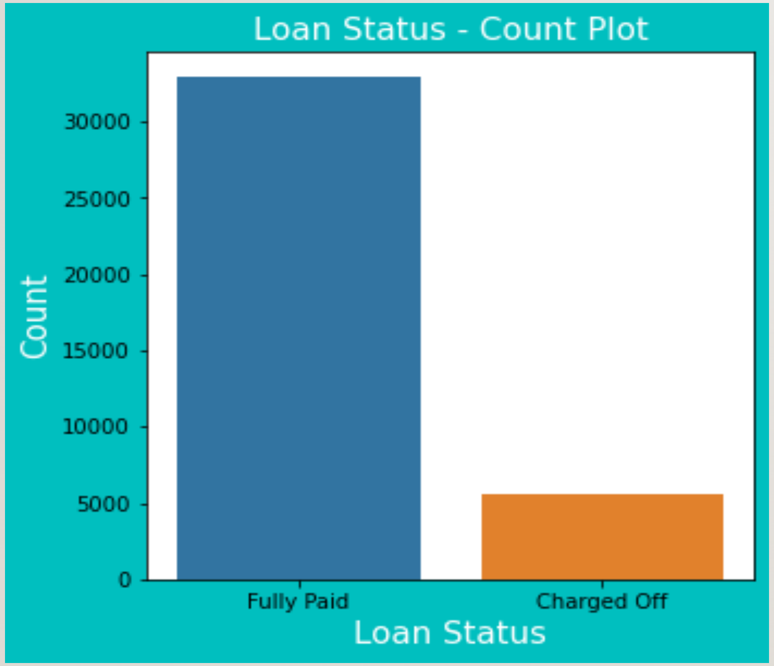
- Analyze how 2 variables influence each other.
- Check for the correlations.

Conclusion

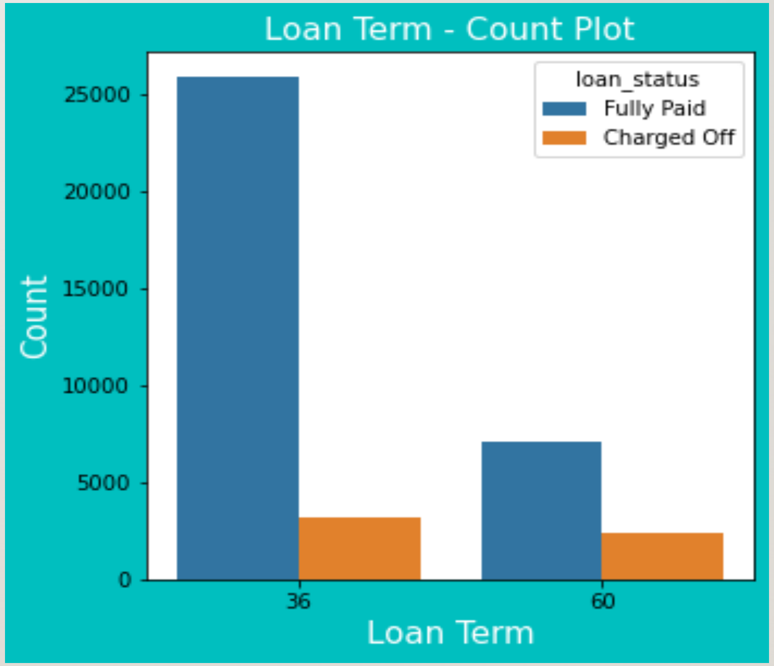
- Publish insights and observations from the analysis.

Analysis

Out of all closed loans, approx. 15% of loans are Charged Off.

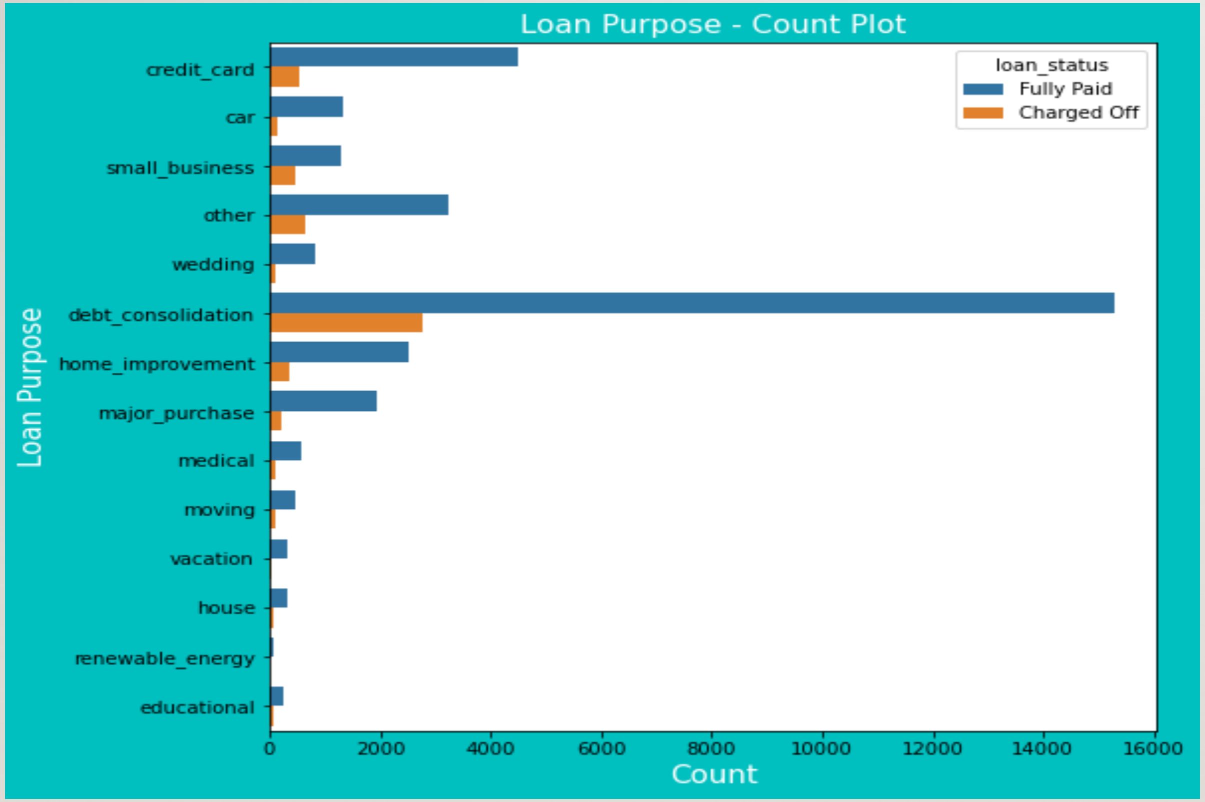


Approx. 25% of the total loan applied is for 36 months

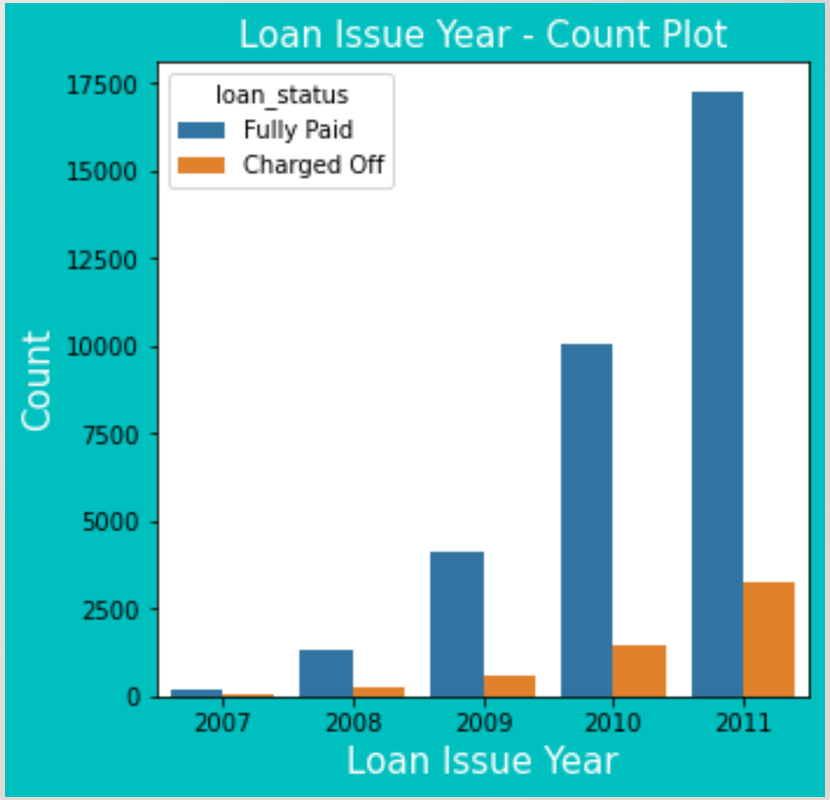


Analysis Contd.

Bulk of the loan is applied for "Debt Consolidation", followed by 'Credit Card'.

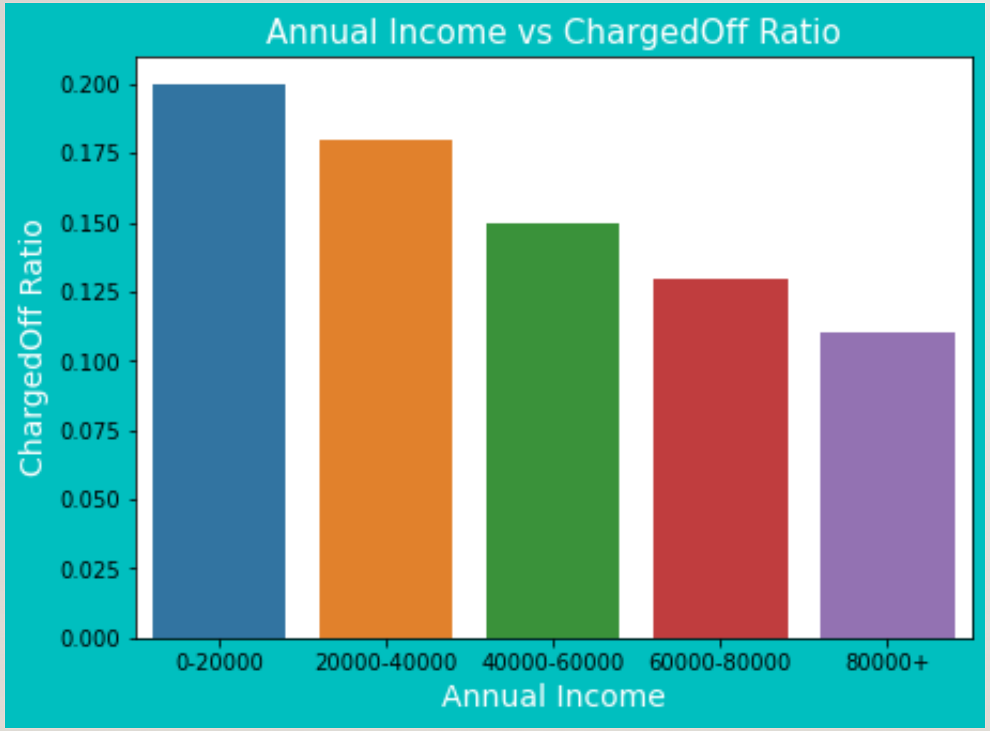


Exponential year on year growth from 2007 – 2011.

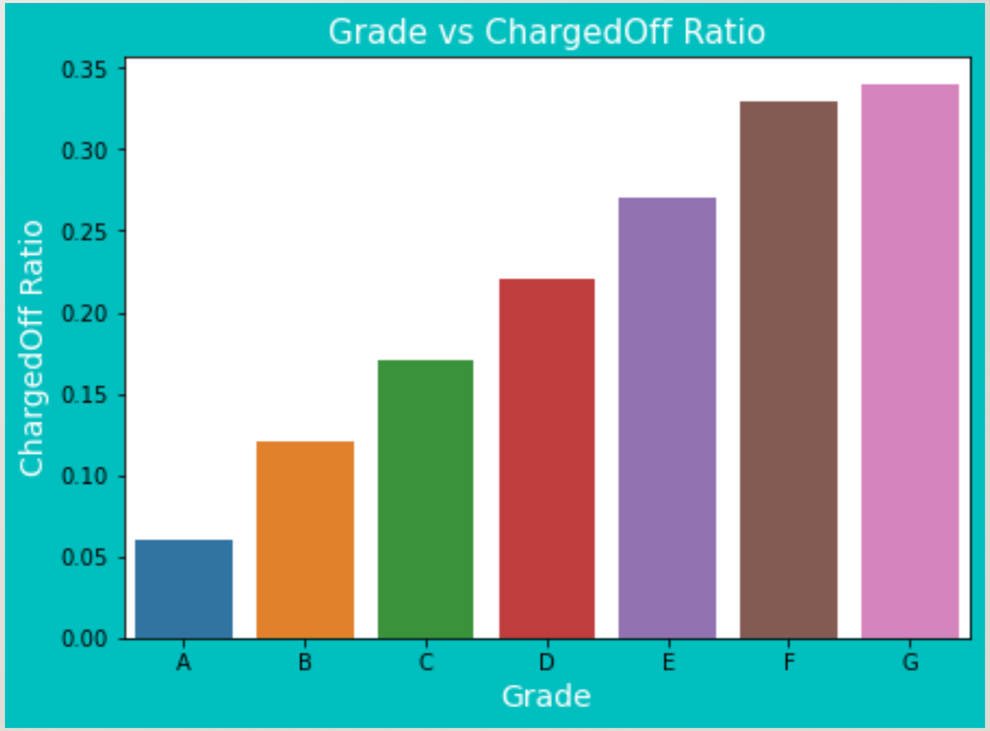


Analysis Contd.

Lower annual income group (< $20,000) has highest risk of defaulting loan. The risk reduces with increase in annual income.

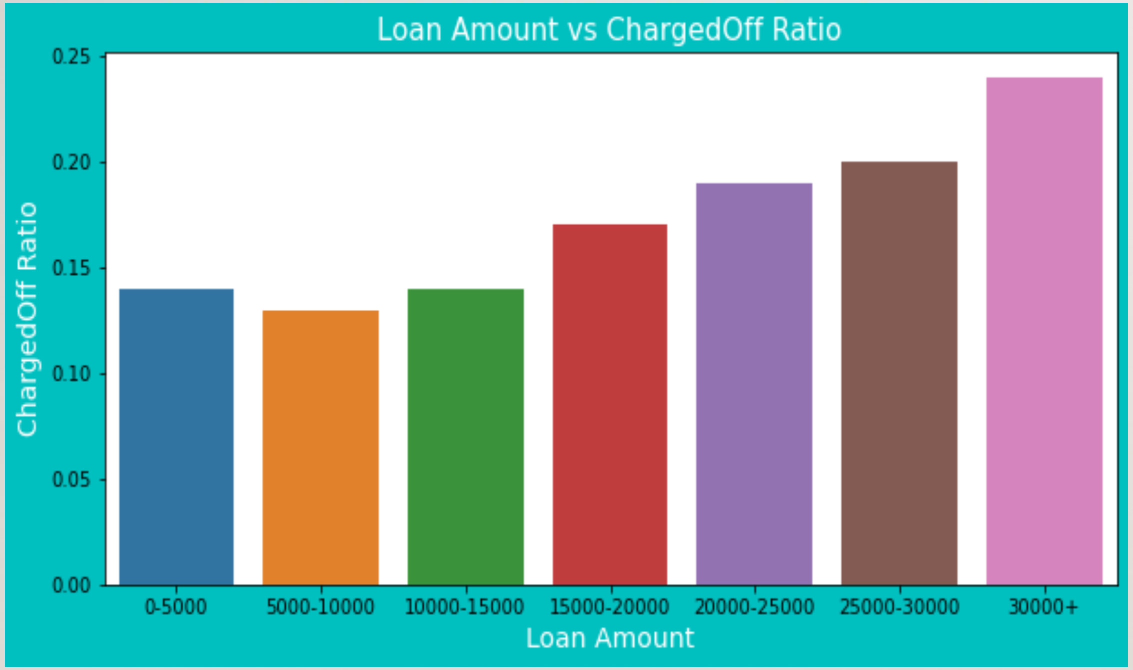


Loans issues for Grades F and G has highest risk of defaulting. Loans approved for worse grades such as E,F and G should be scrutinized thoroughly.

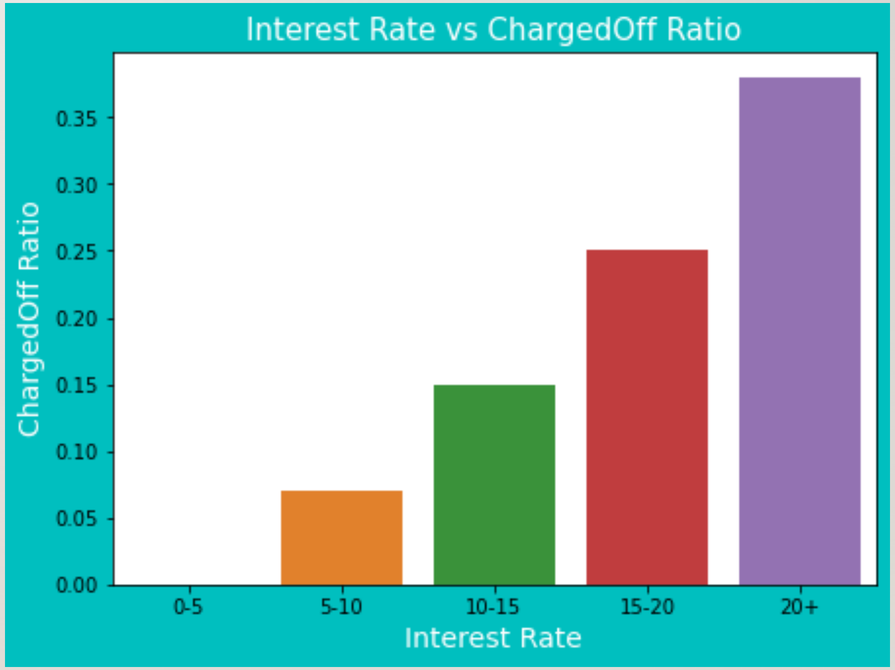


Analysis Contd.

Higher loan amount has highest risk of defaulting loan. The risk increase with increase in loan amount especially for amount > $30,000, which is highest risk category.

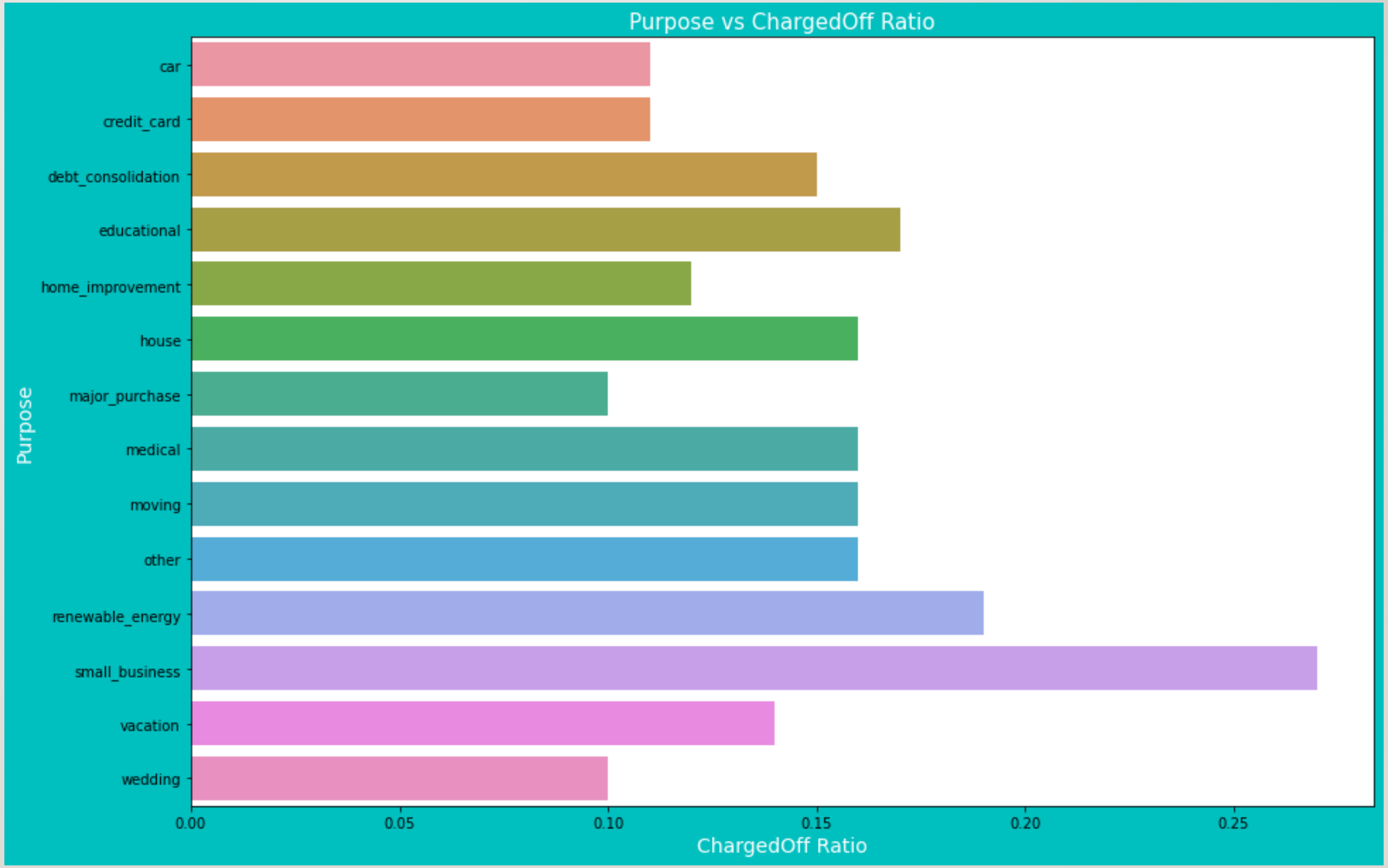


High interest rates has higher risk of loan defaults. Interest rate of 20+ percent has highest percent of loan defaults.



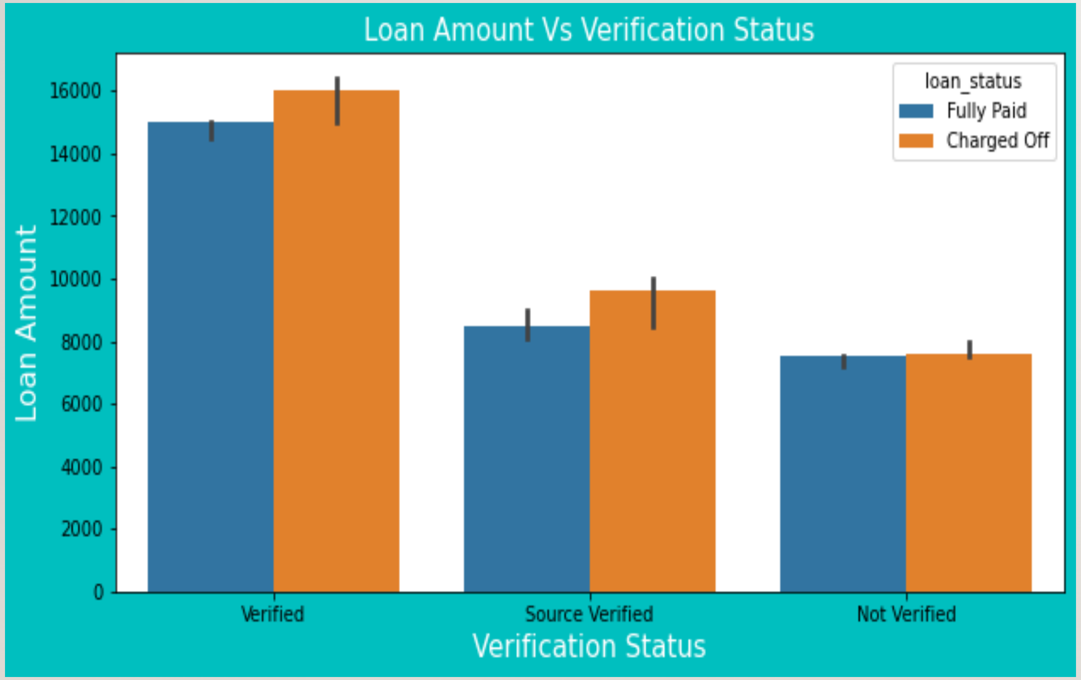
Analysis Contd.

- Loans issued for Small Business purpose carries a significant amount of high Charge Off risk.
- This is followed by renewable energy and education, which also have high defaulting risk.
- Other loan purpose as debt consolidation, house, medical, moving and other purpose have similar high risk
- The lowest risk loans are ones taken for wedding and major purchase.

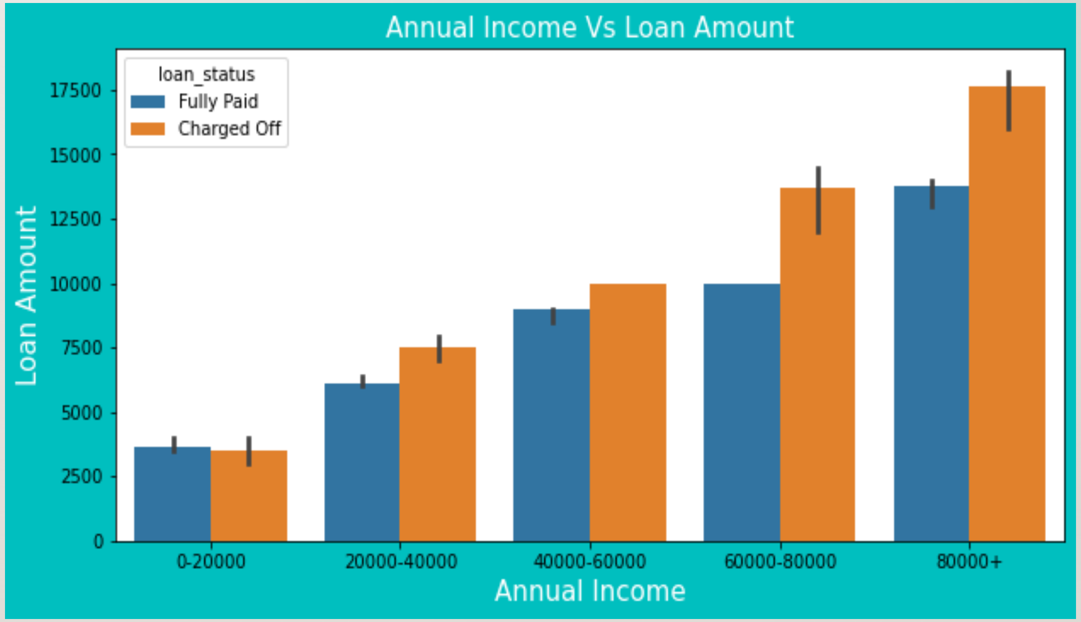


Analysis Contd.

Individuals who have not been verified have NOT been offered higher loan amount. Verified Individuals have been issued loans with greater loan amount.



Individuals having higher income have been approved greater loan amount. However, loans with higher loan amount in each income group have been defaulted.

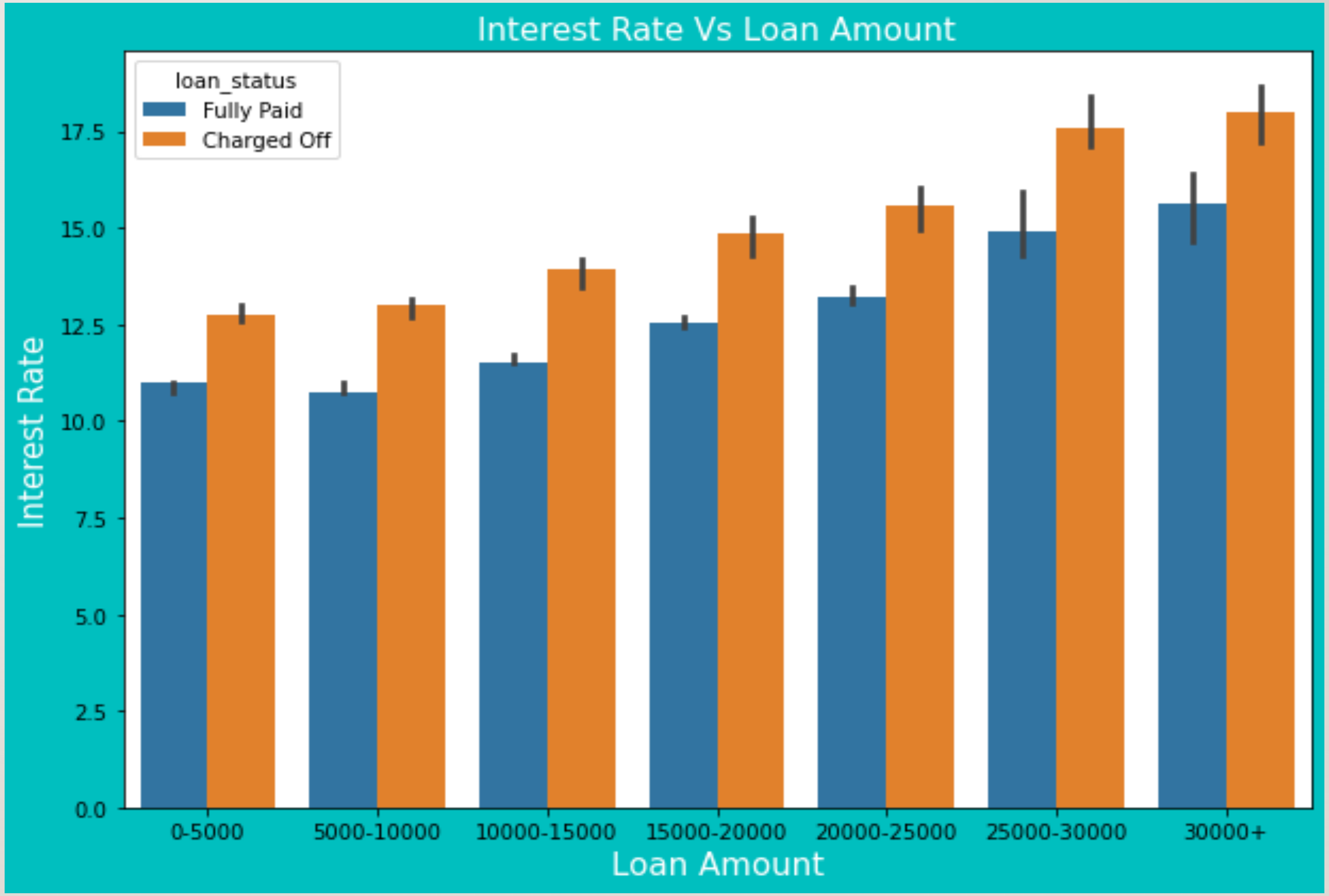


Analysis Contd.

- Higher loan amount attracts higher interest rate.

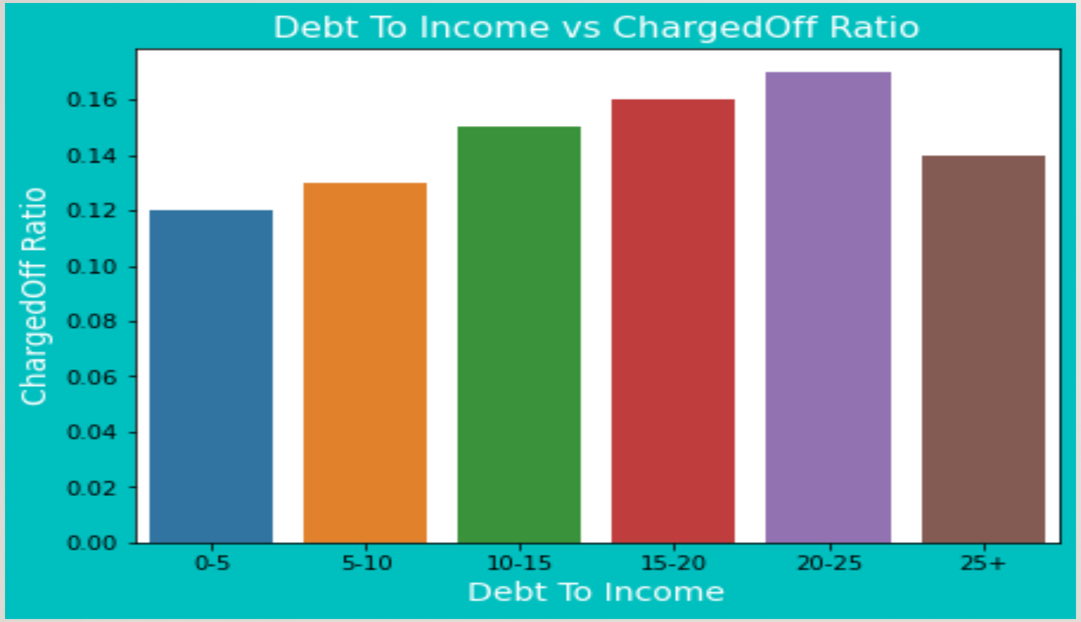
- Charged Off loans for each loan amount groups have higher interest rate, which means, if the interest rate is higher for the same loan loan amount, the risk of getting loan default is greater.

- Loans having higher loan amount and higher interest rates have greater risks of getting default.

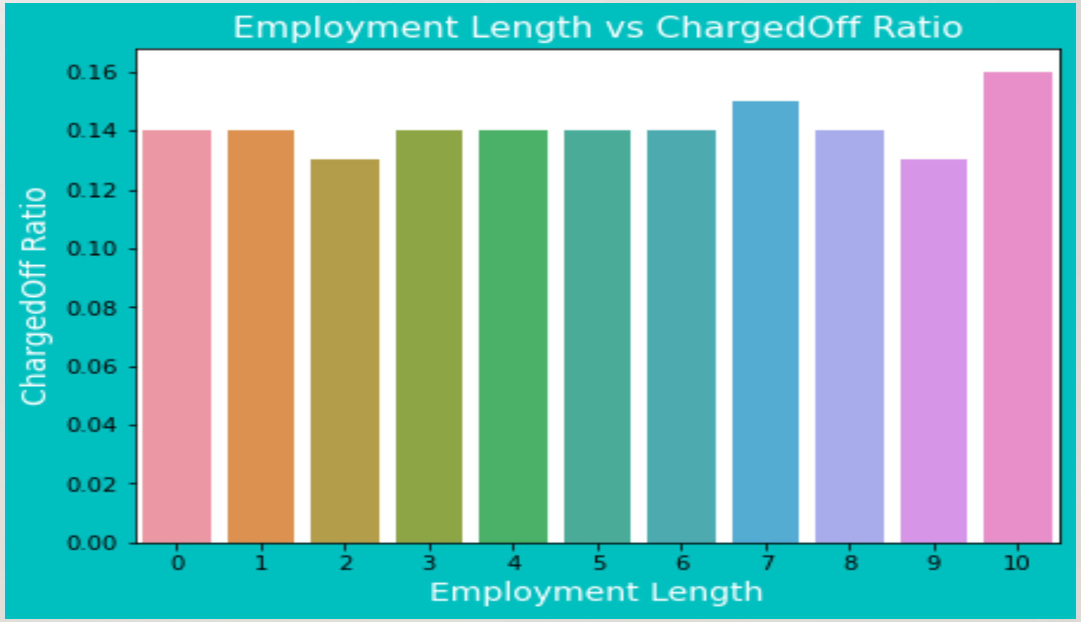


Analysis Contd.

The individuals having debt to income ration of 20 – 25 have defaulted greatest % of loan.

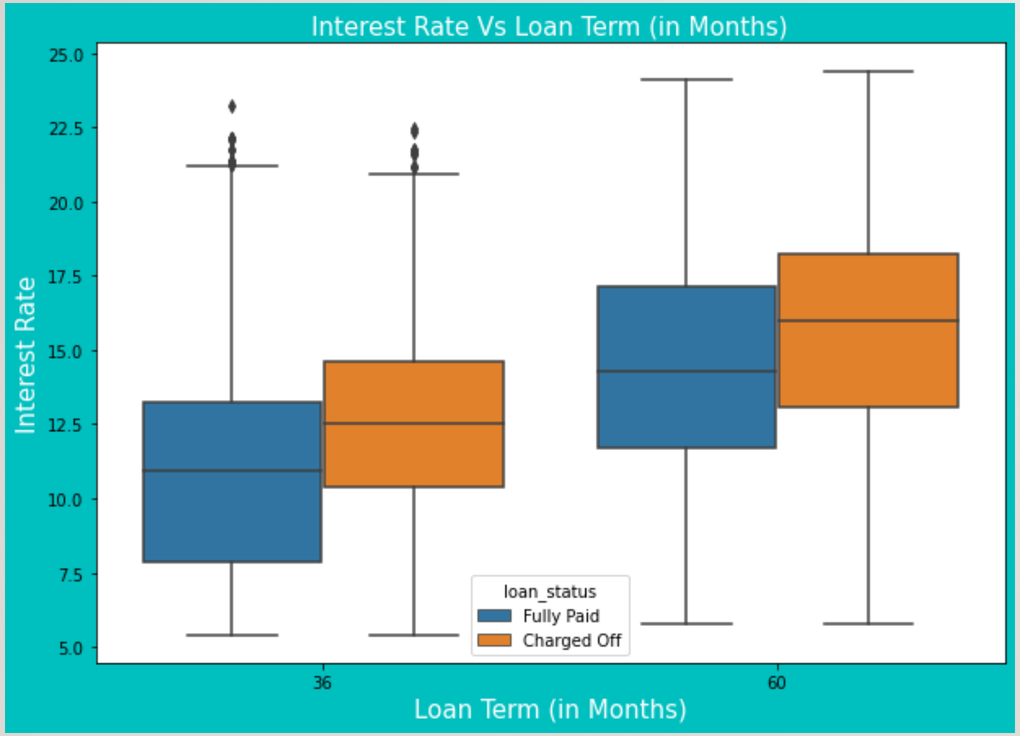


Loan Charge off ratio is similar for all the individuals irrespective of their employment lengths.

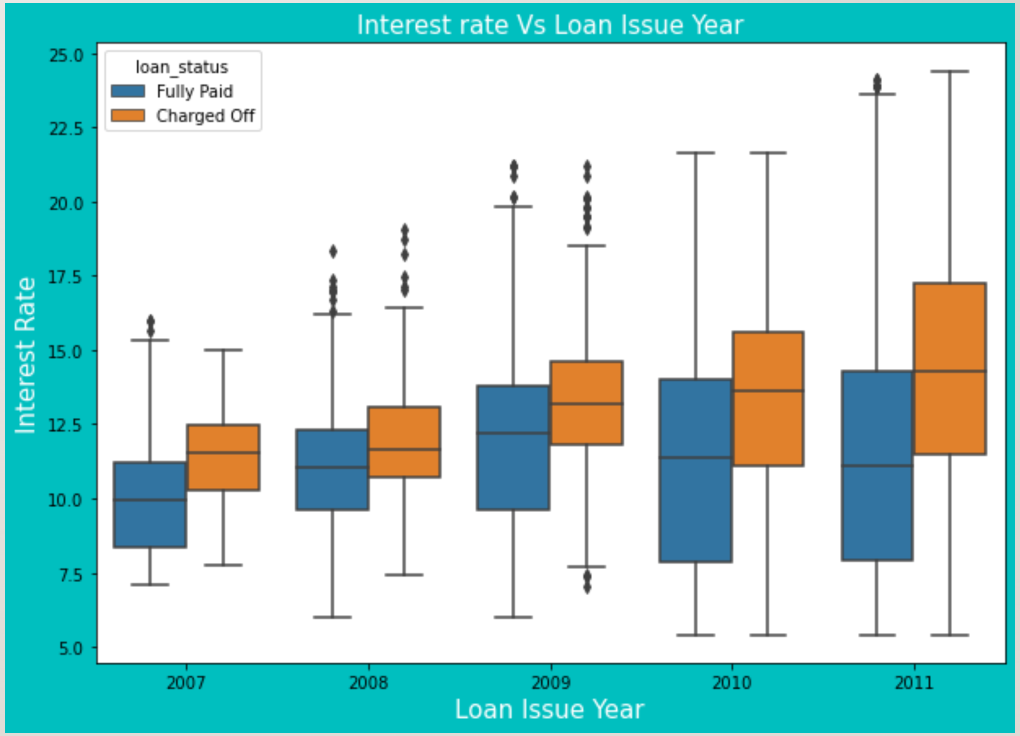


Analysis Contd.

Higher loan term have attracted higher interest rate.

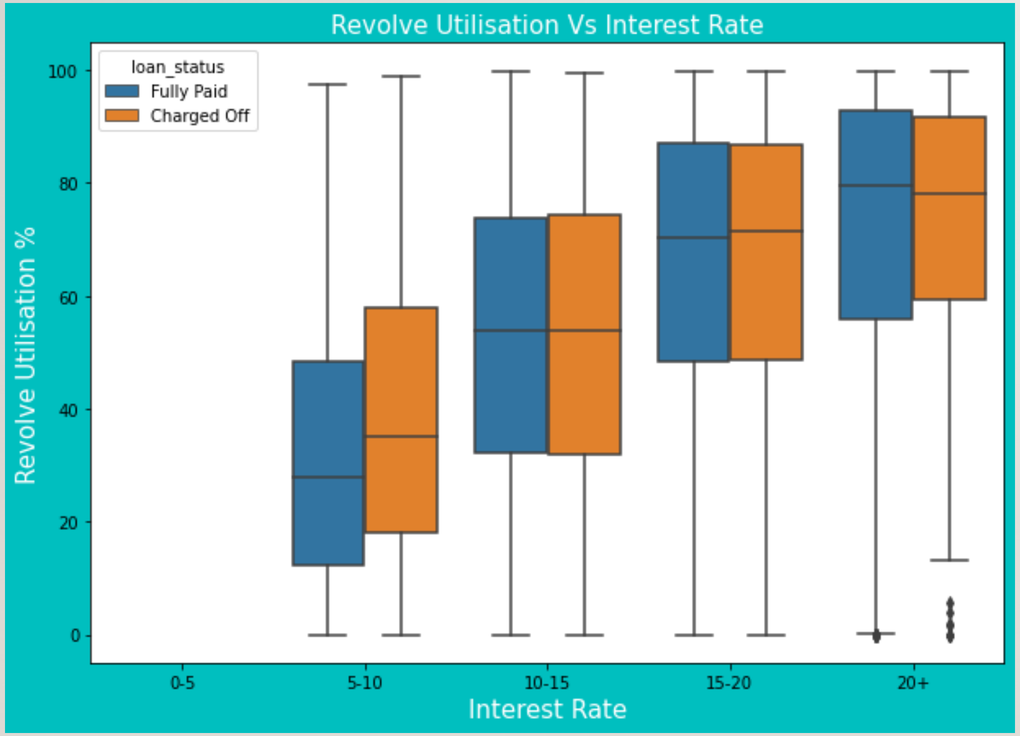


The average interest rate have gradually risen over the years.

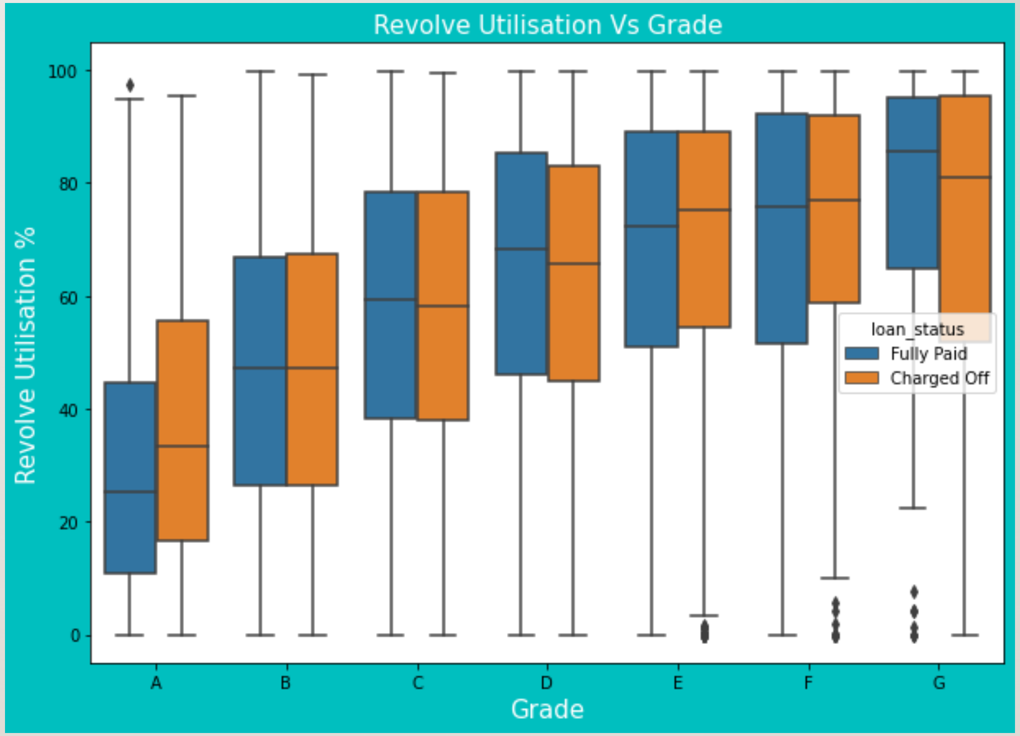


Analysis Contd.

Individuals having loans with higher interest rates have high revolve amount utilization.



Individuals having loans with worse grade have high revolve amount utilization.



Recommendations

To make better decision and bring down the risk of loan default, below are recommended;

- Reduce the number of loans having loan amount > $30,000. Approve only for high income individual.
- Avoid approving Loans having interest rate > 20%, instead, make a counter loan offer with lower loan amount and lower interest rate, wherever possible.
- Reduce number of loan approvals for Loans for Small Business, Renewable Energy & Education.
- Avoid approving high value loans to individuals with Revolve Utilization rate > 75%.
- Avoid approving Loans with Bad Loan Grades F, G.
- Avoid approving high value Loans to individuals with Annual Income < $20000, instead, make a counter loan offer with lower loan amount.
- Avoid approving high value loans to individuals having Debt To Income ratio > 20% - 25%

THANK YOU !!!