LENDING CASE STUDY

SUBMISSION

Sharukh Shaikh
Jawaad Shaikh

Analysis Overview

- **Contents**
  - Analysis Anecdote.
  - About the Data
  - Cleaning Chronology
  - Univariate Analysis
  - Bivariate Analysis
  - Driving Factors on Defaulting on a Loan
  - Conclusion Pointers

Analysis Anecdote

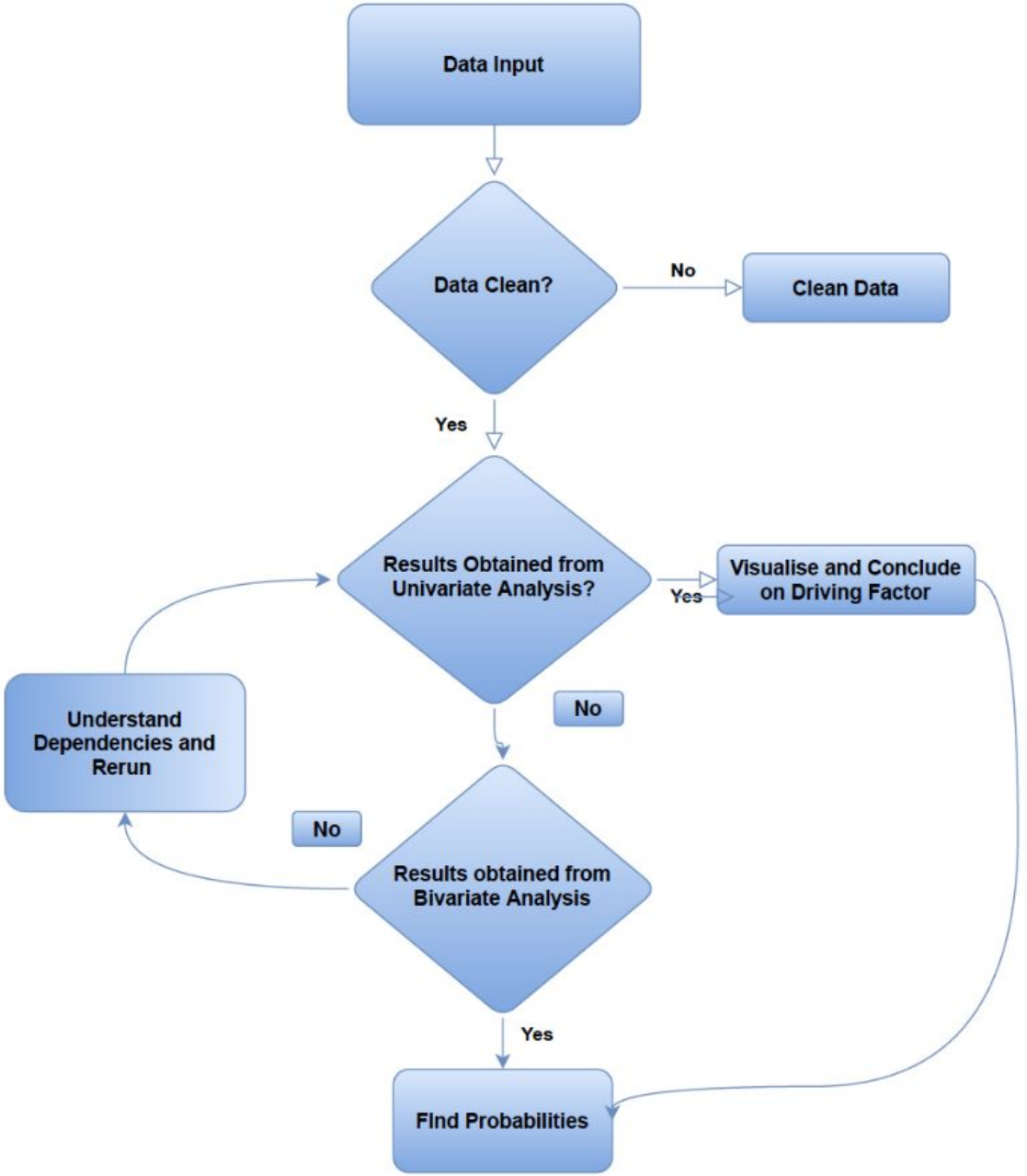
- **Goals**
  - A consumer finance company intends to identify risks pertaining to loan defaulters
  - Analyzing applicants who are likely to repay the loan
  - Analyzing applicants who are likely to default on loans
  - Identify Driver Variables that cause an applicant to default
  - Quantify the magnitude of probability for an applicant to default

- **Assumptions**
  - Historical dataset shall be used to infer on a singular application
  - The dataset contains accepted applications and the respective status of the loan

About The Data

- ***loans***: Table including various lending details
- ***Data-dictionary***: *Metadata* about the column names
- ***Loans*** dataset contains applicants' loan status, under three categories
  - ***Fully Paid:*** Applicant has fully paid the loan
  - ***Current:*** Applicant is in the process of paying the instalments
  - ***Charged Off:*** Applicant has not paid the instalments in due time for a long period of time
- Based on the loan status, there are different associated variables such as:
  - Annual Income of Applicants
  - Reason for Loan and Others
- Referring to the loan status and the associated variables, the anlaysis is conducted to understand the "driver variables" to understand different factors that cause an individual to default on loans.

Process

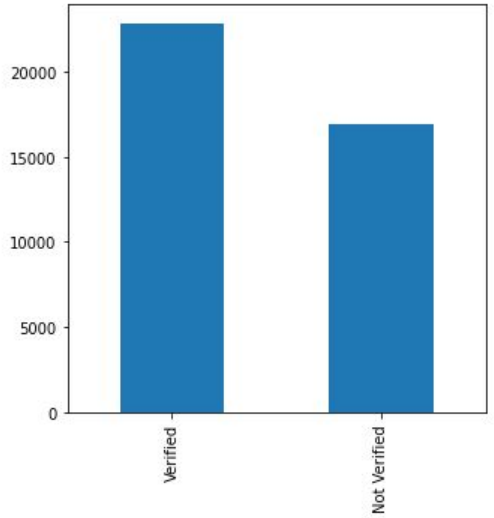

Data Input
Data Clean?
No
Clean Data
Yes
Results Obtained from Univariate Analysis?
Yes
Visualise and Conclude on Driving Factor
No
Understand Dependencies and Rerun
No
Results obtained from Bivariate Analysis
Yes
Find Probabilities

Cleaning Chronology

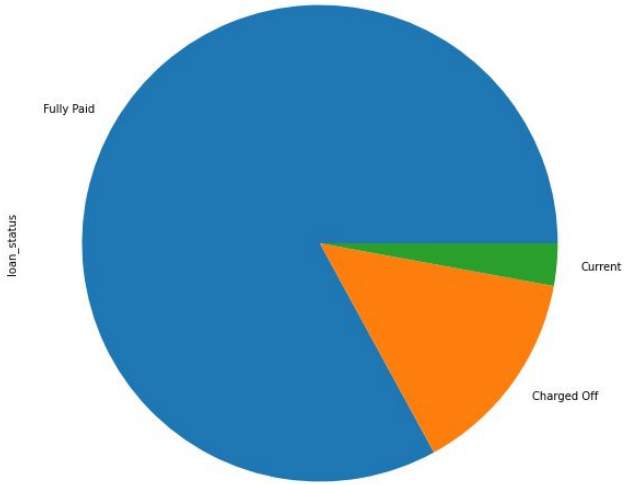
- Removed columns / dimensions which have **real values less than 2**;
  - ***63 columns removed, 47 columns remain***
- Checking columns for **junk values greater than 90%;**
  - ***2 columns removed, 45 columns remain***

- Removing Columns having **% missing data greater than 30%;**
  - ***2 columns removed, 43 columns remain***
- Corrected Value errors such as:
  - Object / String values to integer and float (***converting to proper dtypes***)
  - ***Date – Time Corrections***
  - Eliminated String Errors such as; ***Interest Rate is 5%; removed %***

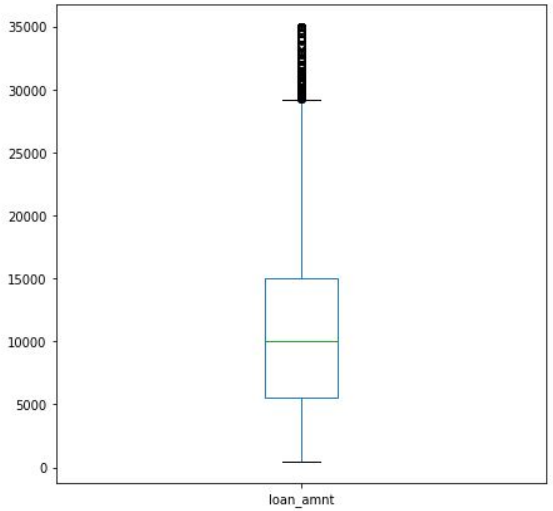
Univariate Analysis - 1



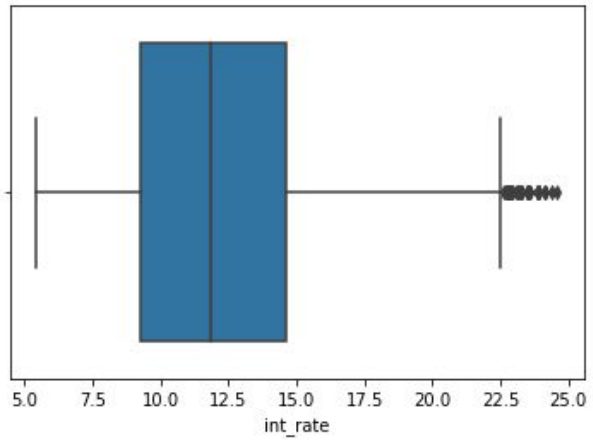
Most Accepted Applicants have their source verified for a specific loan



Most of the loan applicants have fully paid their loans

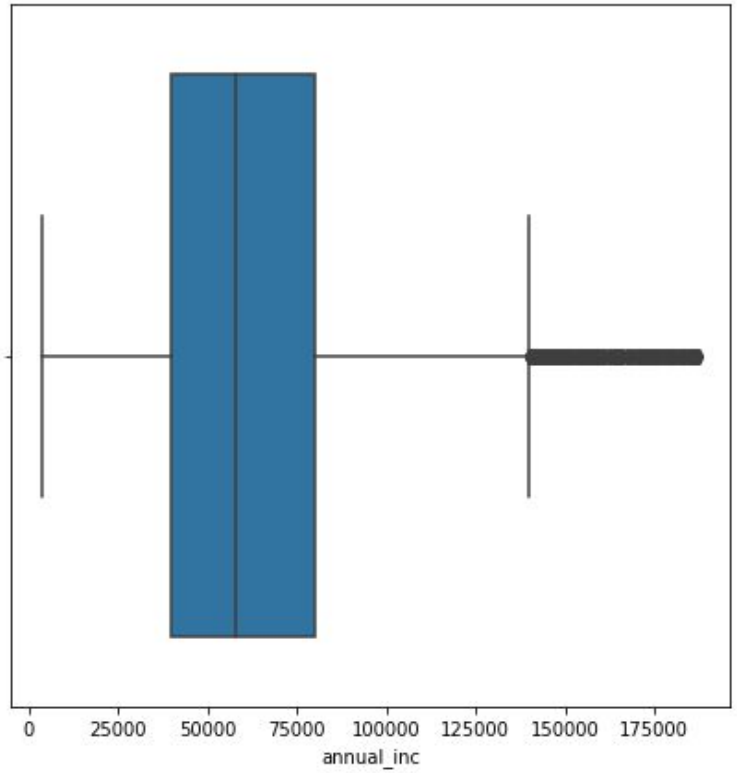


The ***loan amount is heavily varying between 5k to 15k USD;*** with some outliers

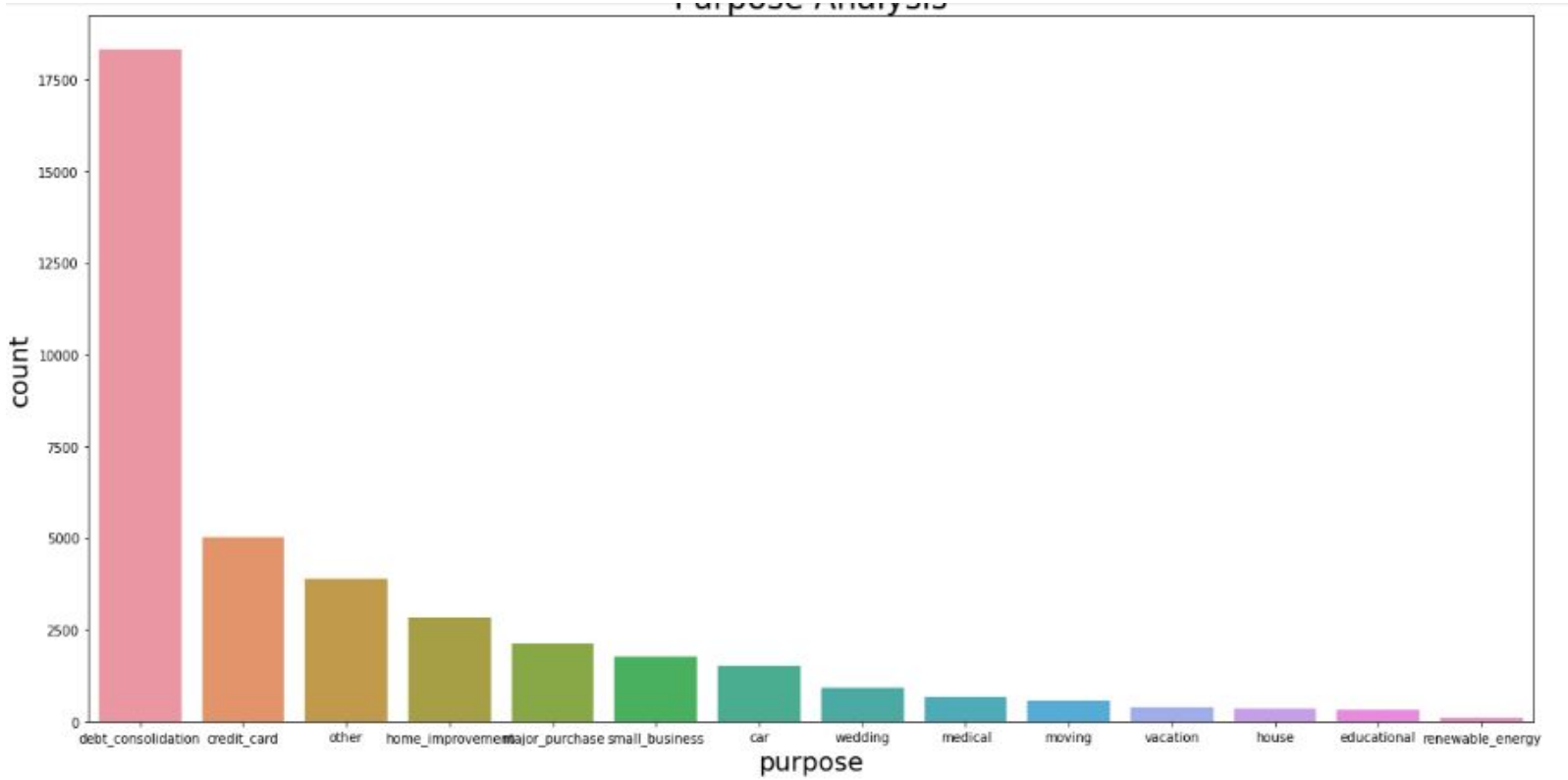


The ***interest rate varies from 9% to 15%;*** however, there are cases where the interest rate is shot up to 25%.

Univariate Analysis - 2

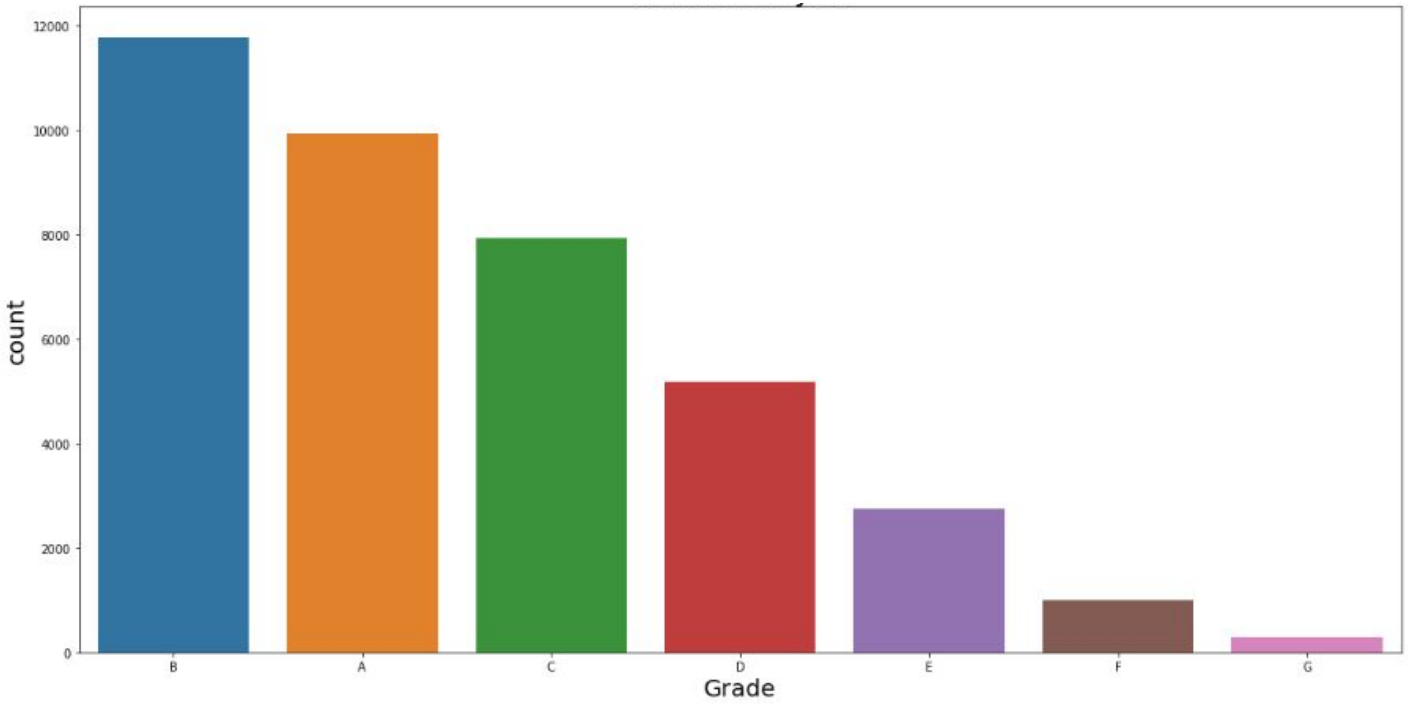


A most alarming driver is noted to be annual income onto the loan applicants. The ***nominal income ranges between 30k to 80k USD***.

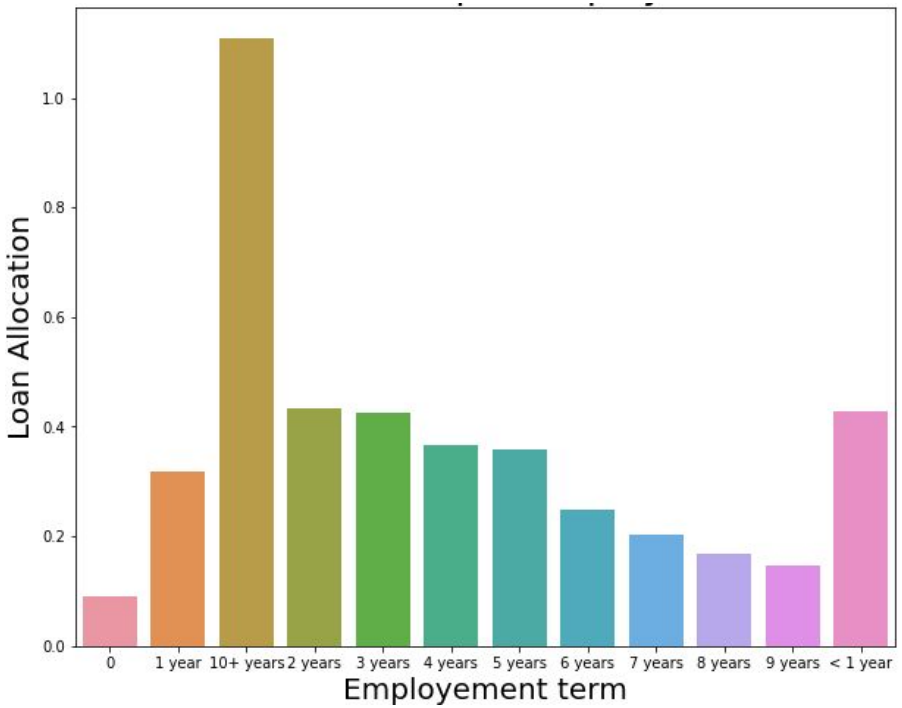


Another dominating factor is the reason for applying for loan. Majority of the ***applicants are debt – ridden***. This factor should be carefully accounted in clearing an individual's application.

Univariate Analysis - 3



Based on the assigned grades on an individual's application. It is noted that ***most applications are grade as "B"***, which is closely followed b A and C

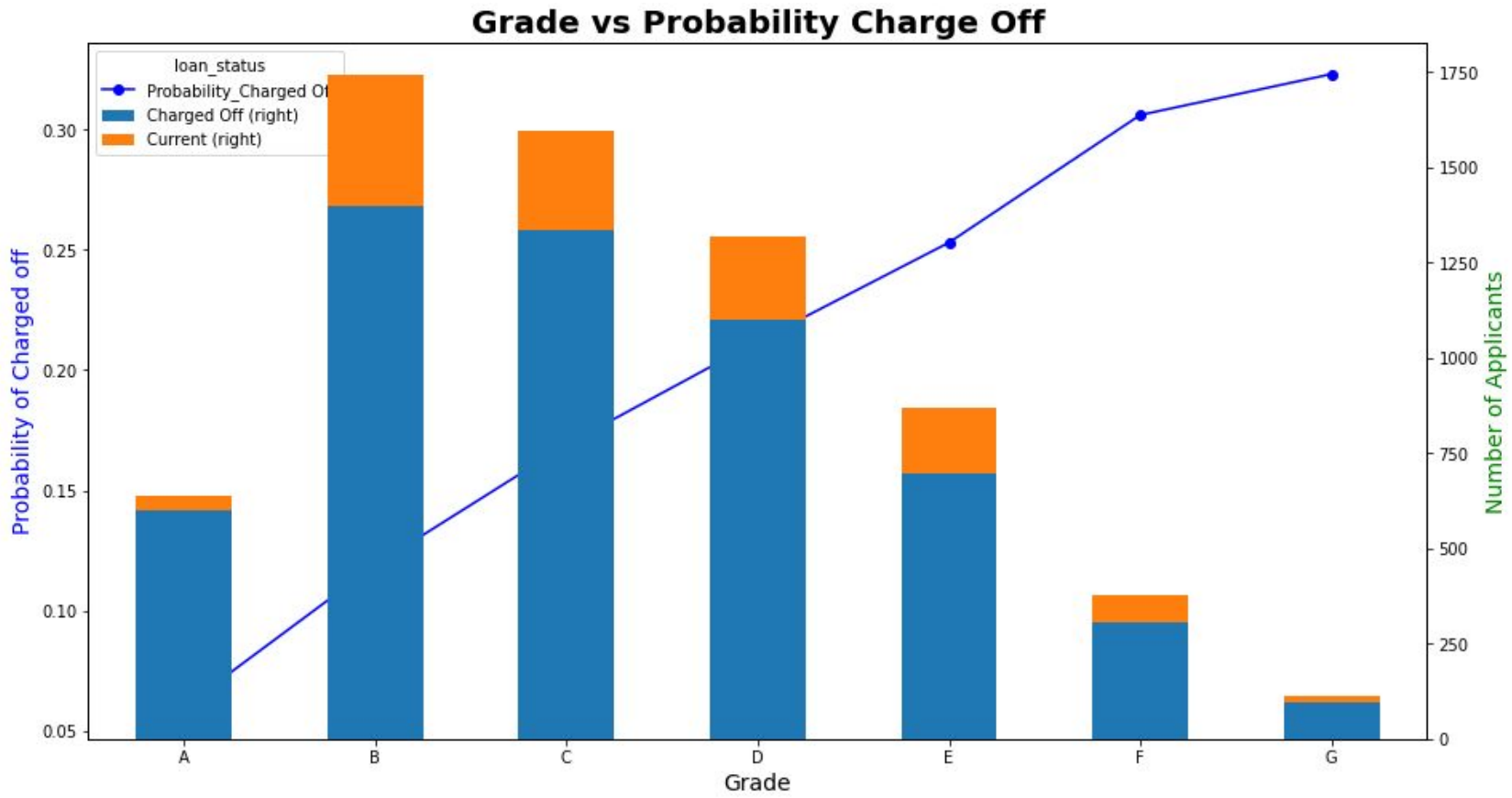


The highest number of applications are noted from those individuals whose working ***experience is greater than 10 years***

Bivariate Analysis – Correlation Matrix

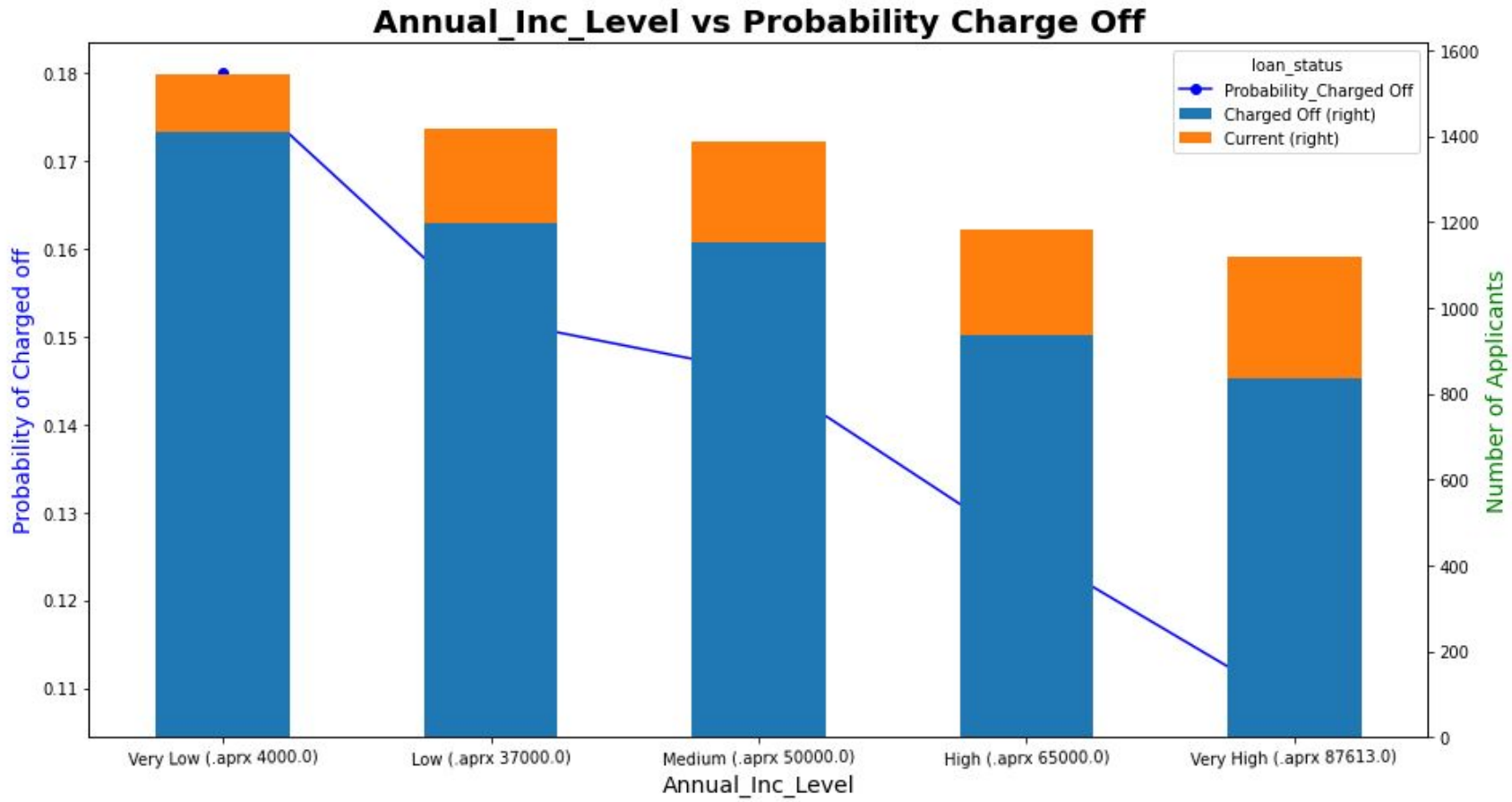
- The interest rate and the amount is highly correlated, indicating, ***more the amount greater the interest.***
- The installments are greater if the loan amount is greater.

Probabilities – Grade v/s Loan Status



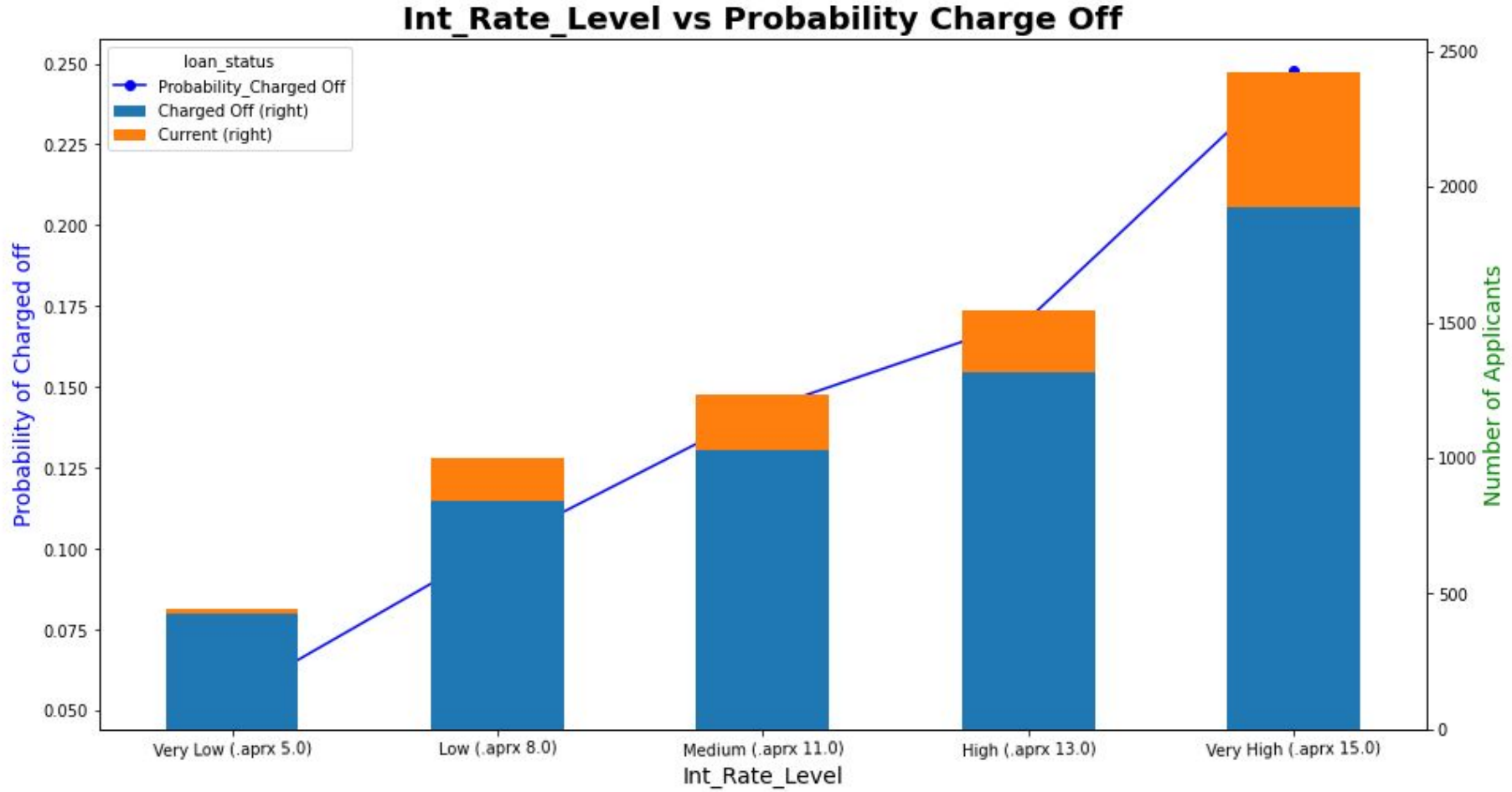
The probablity of an individual to default on loans increases for grades after D

Probabilities – Annual Income v/s Loan Defaults



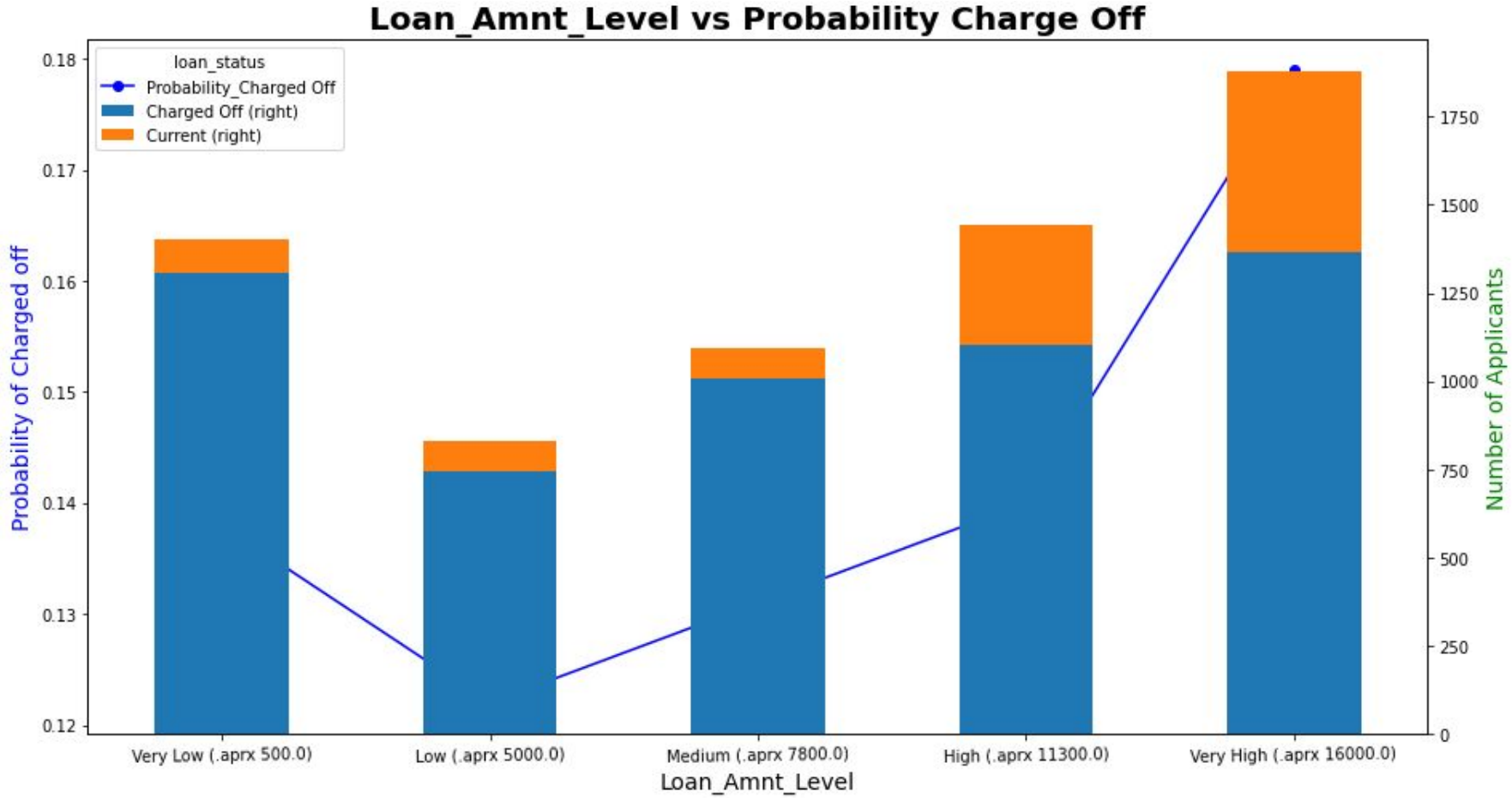
The probability of an individual to default on loans increases for individuals with low annual incomes; i.e below 4k USD

Probabilities – Interest Rates v/s Loan Defaults



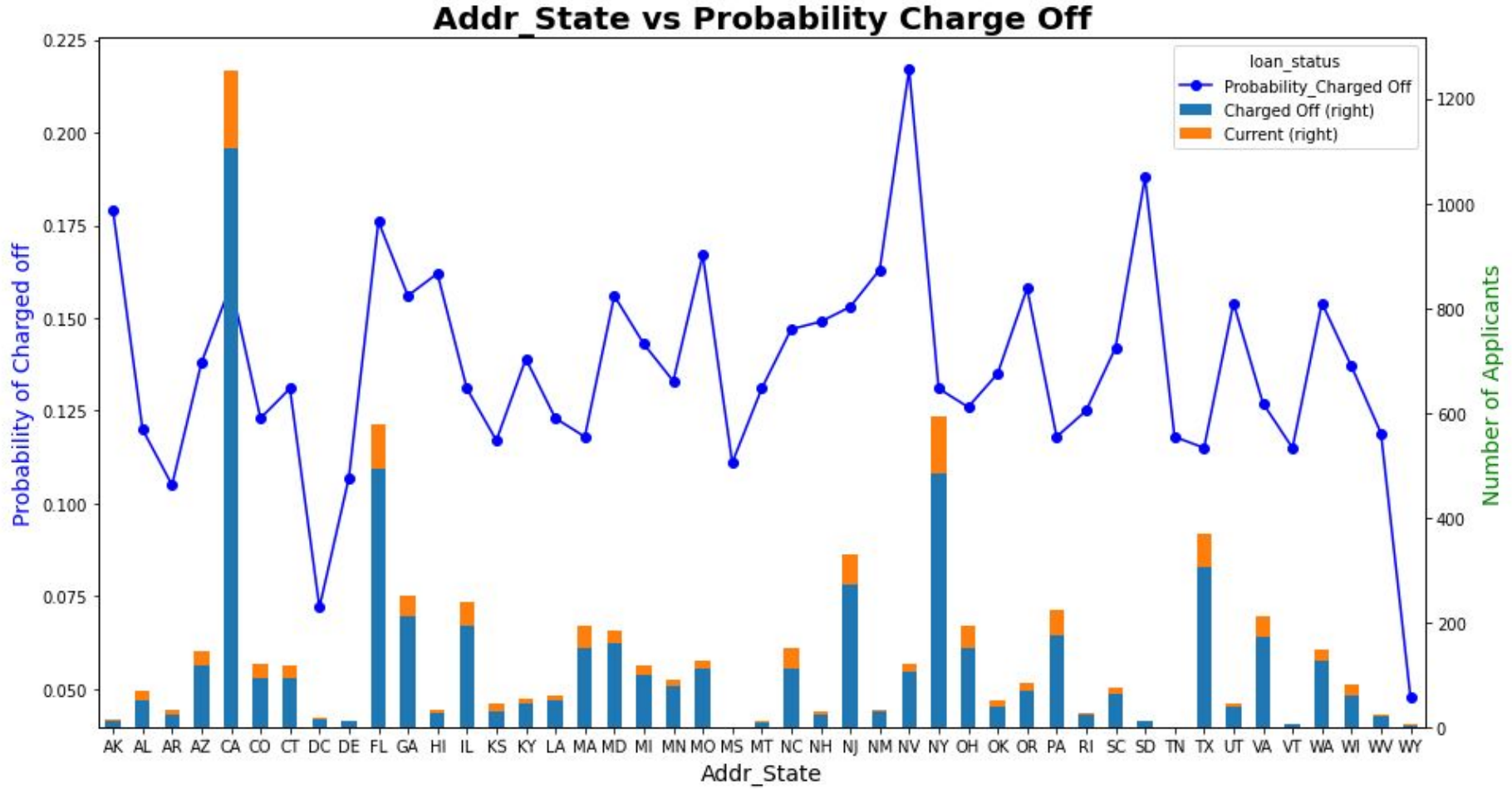
The probability of an individual to default on loans increases for individuals with high interest rates; i.e interest rates greater than 13%

Probabilities – Loan Amount v/s Loan Defaults



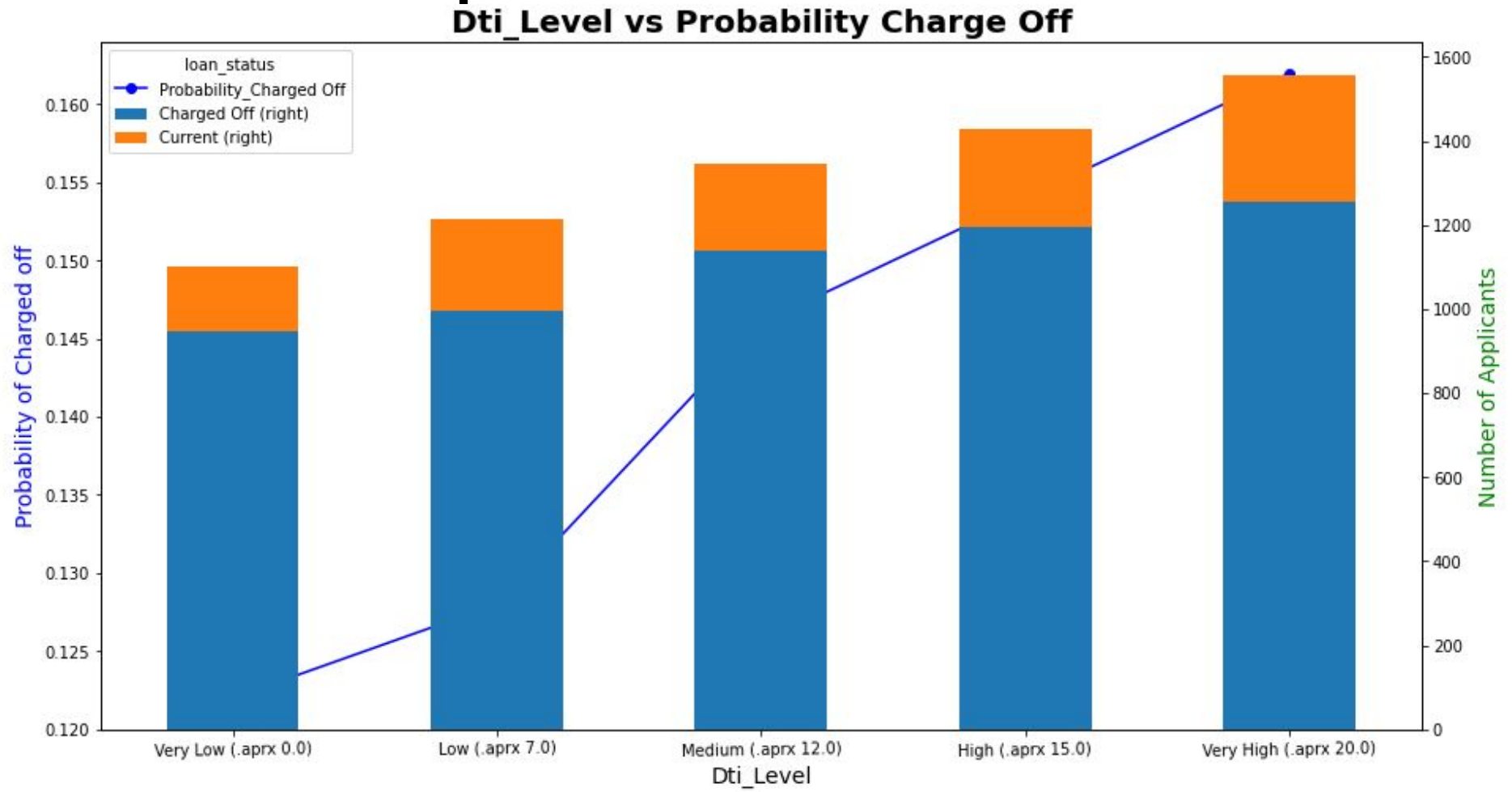
The probability of default is increased with higher loan amounts; however, it is observed even for loan amounts lesser than 500USD, there are significant defaulters

Probabilities – Region v/s Loan Defaults



Nevada (NV) has the highest defaulters, followed by South Carolina (SC) and Florida (FL)

Probabilities – Impact of DTI



As the DTI (Debt to Income Ratio) increases, the probability of defaulting increases

Conclusions

We have performed different types of analysis to find few of the important factors/variables that needs to be considered to determine is an applicant will default or not. Those variables are as follows:-

1. Grades (grade)

From A to G including the subgrades the trend is the probability of being a loan default increases.

2. Annual Income (annual_inc)

It has a negative co-relation, the more the income, the lesser chance to default the

3. Inerest Rate (int_rate)

It is positively co-related, the more the interest rate the more chance of an applicante to be a loan defaulter

4. Loan Ammount (loan_amnt)

It is positively co-related, the more the loan amount, the more there is a chance of an applicant to be a loan defaulter

5. Debt To Income ratio (dti)

It is positively co-related, the more the DTI, the more there is a chance of an applicant to be a loan defaulter