First Payment Default Prediction

Name: Vijendar G ***Mail id:*** ***vijendarreddy.g@gmail.com***

Introduction:

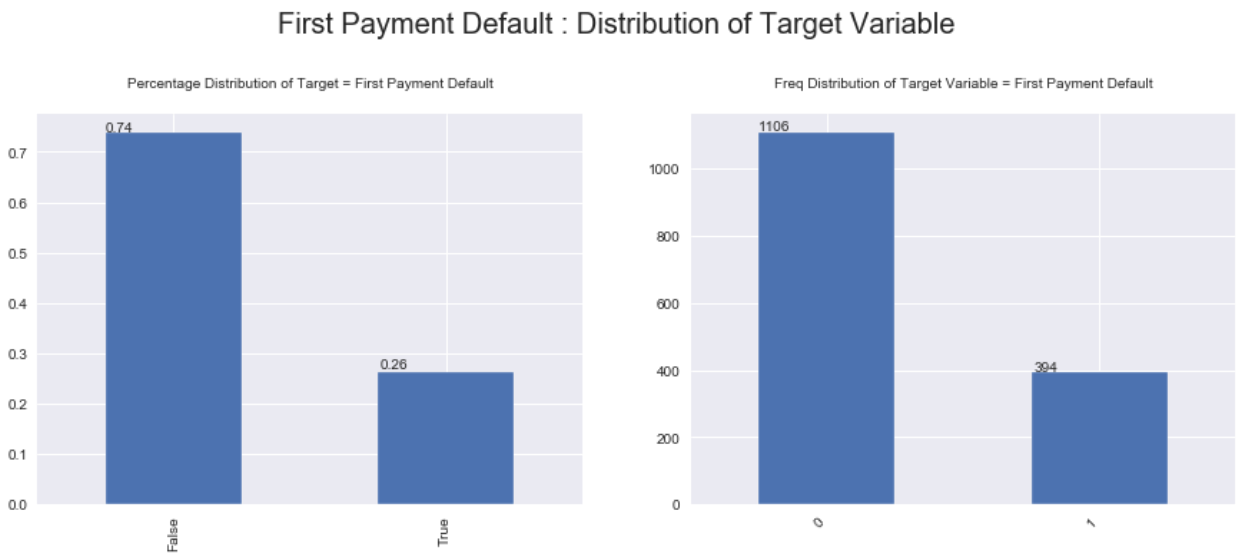
Loan payment default prediction is classic case of classification technique, where we will be predicting the Default/Non-default (Dichotomous Response) records based on additional information is provided on other characteristics which are believed to affect the occurrence of the event of interest.

The given data is having 12 columns, with financial demographic information like Time of Application, State, Monthly Net Income, Paycheck Net Income, Rent or Own, Months at Residence, Bank Account Months, Pay Cycle, Loan Amount, Loan Funded Date and Loan Due Date, and was asked to predict whether the customer default first payment(variable of interest).

Descriptive statistics:

From basic statistical analysis, we can observe

- The data is having more non default records (74%).



- The data collected from two states CA (72.4%) and TX (27.6%).

Value	Count	Frequency (%)
CA	1159	72.4%
TX	441	27.6%

- Majority of Loan seekers were residing in rented home (it was 87.9% in the given data).

Value	Count	Frequency (%)
R	1406	87.9%
O	194	12.1%

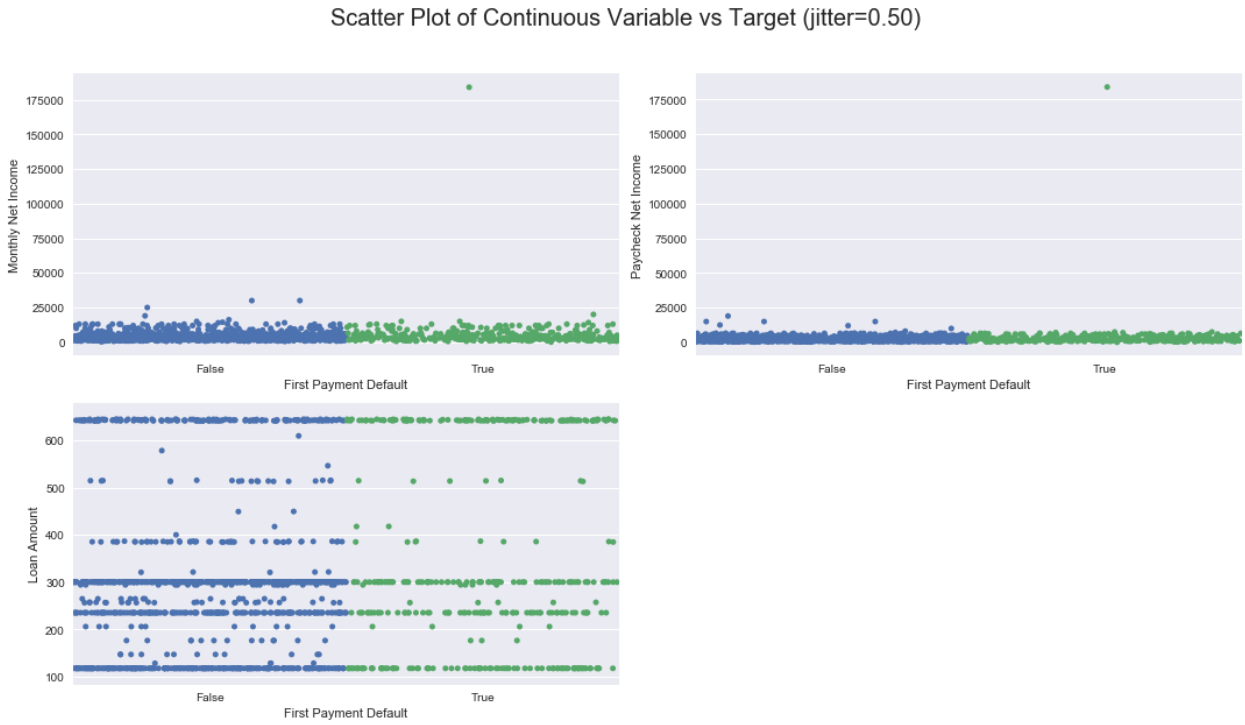
- Bank is offering 4 different Pay cycle options. As majority of firms do salary payments Bi-Weekly the data (with 45% customers Bi-Weekly pay cycle) shows customers are preferring to pay on salary credits.

Value	Count	Frequency (%)
BiWeekly	721	45.1%
Monthly	572	35.8%
BiMonthly	204	12.8%
Weekly	103	6.4%

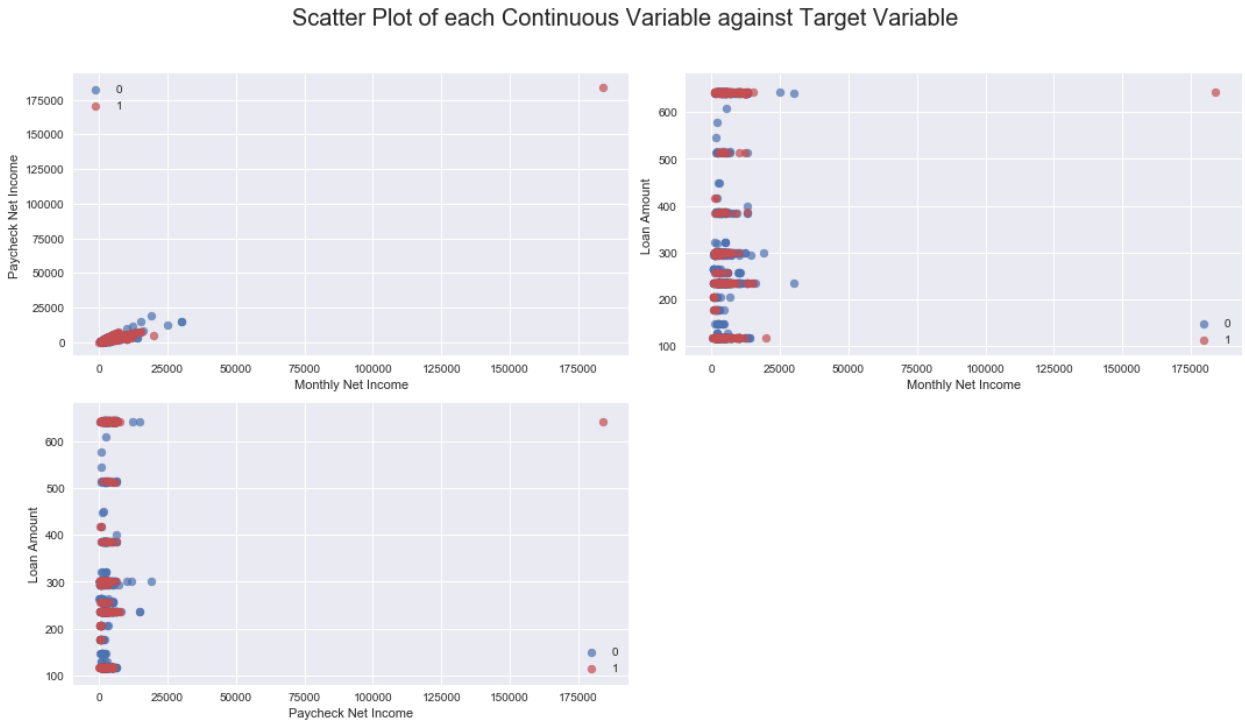
- Monthly Net Income and Paycheck Net Income were distributed with high rightly skewness. And log versions were also non normal (Right skewed). Majority of loan seekers were less income group people (95% of loan seekers earning, Monthly Net Income, less than 12k).
- Loan amount is varying from 117.65 to 644.24. Common loan amounts are 300, 117.65 and 235.3 with 29.8%, 21.4% and 15.6% respectively, accounting for 2 third of overall population. Please follow below most common values.

Value	Count	Frequency (%)
300	476	29.8%
117.65	342	21.4%
235.3	249	15.6%
642.05	29	1.8%

This will reduce expandability of payment defaults. PFB scatter plots evidence.



- As the majority loan amounts were belongs very few numbers and Monthly incomes were right skewed there is no direct correlation against payment defaults. PFB scatter plots.



Please find Data Cleansing techniques, and model development and validation comments in code file.

First_Payment_Default.ipynb