



Credit EDA Case Study

- Nikhil Chandra
- Pratap Ramgiri

Nov 23, 2020

upGrad

PG Diploma in Data Science

Case Study - Objectives



- This case study aims to identify patterns which indicate if a client has difficulty paying their installments
- The patterns may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- This will ensure that the consumers capable of repaying the loan are not rejected
- Identification of such applicants using EDA is the aim of this case study

Case Study - Approach



- The approach for this case study is to adhere to data science industry standard of the steps for Exploratory Data Analysis (EDA)
 - Data Sourcing
 - Data Cleansing
 - Univariate Analysis
 - Bivariate and Multivariate Analysis
 - Presentation

Case Study – Data Sourcing



- The Source for this case study is provided in three excel files:
 1. 'application_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
 2. 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
 3. 'columns_description.csv' is data dictionary which describes the meaning of the variables.

The data is read into PD data frames in python

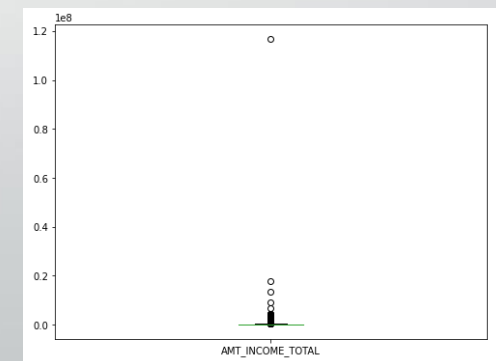
```
appl_data = pd.read_csv("application_data.csv")
```

```
pre_appl_data = pd.read_csv("previous_application.csv")
```

Case Study – Data Cleansing



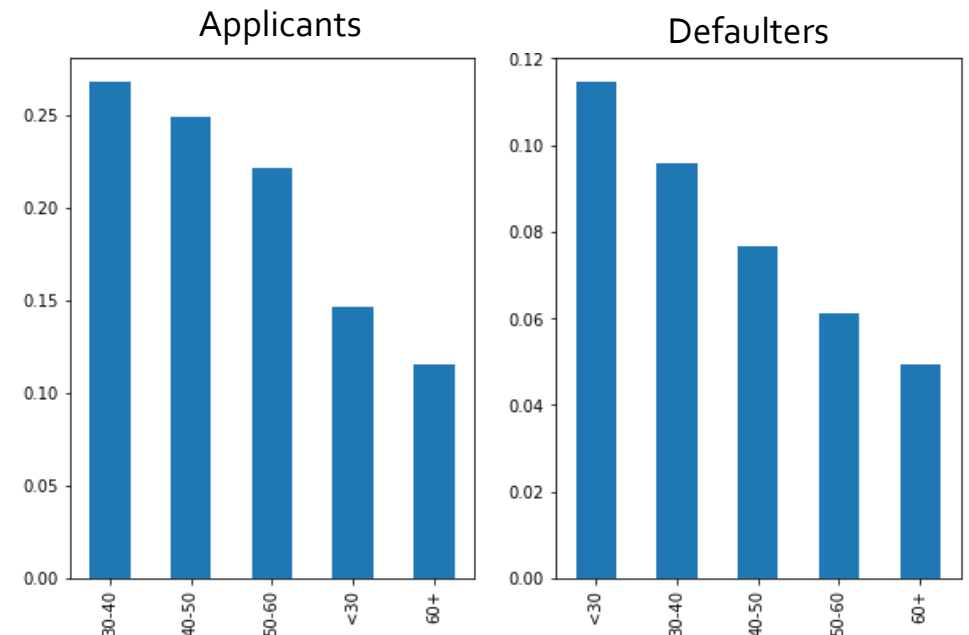
- The Source application data has total 307511 entries and 122 columns
- After analyzing the data many columns found to be not relevant for our analysis
- Many columns found to have more than 50% of null values, we can safely drop them
- After dropping the columns the data shrink to 53 columns
- There are also certain fields which have impractical values, the data is filtered, taking out those entries for analysis
- There are few outliers which normally taken out but our analysis we didn't make any changes.
 - Ex: Income Total has clearly showed outliers in boxplot



Case Study – Analysis



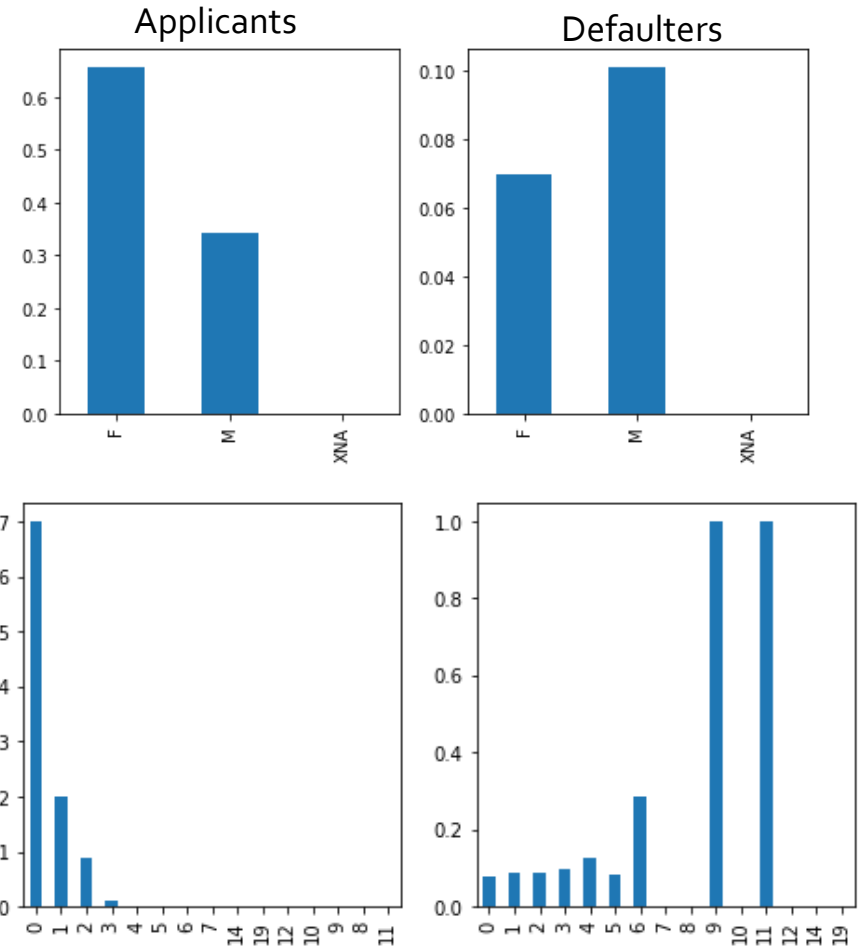
- Most of the Analysis is performed against Target variable with few important columns which will provide in depth analysis on the characteristics of the defaulters which will in turn provide the bankers on decision making
- The below bar plot is providing the percentage on what age group are applying for loans and percentage of defaulters
- Based on the first we can say that the most applications are received by bank are in the age group of 30 to 40
- Most of the defaulters are in the below the age of 30



Case Study – Analysis



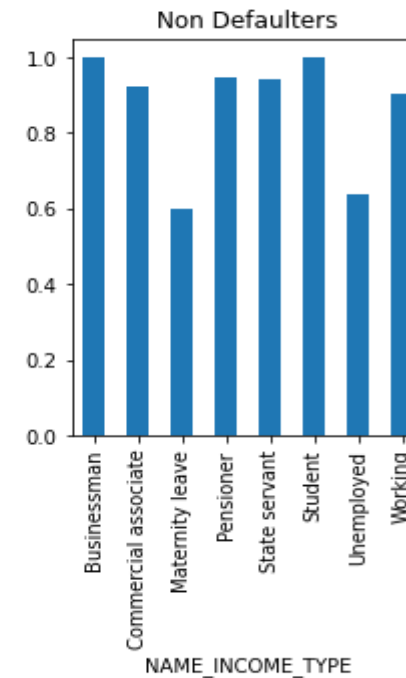
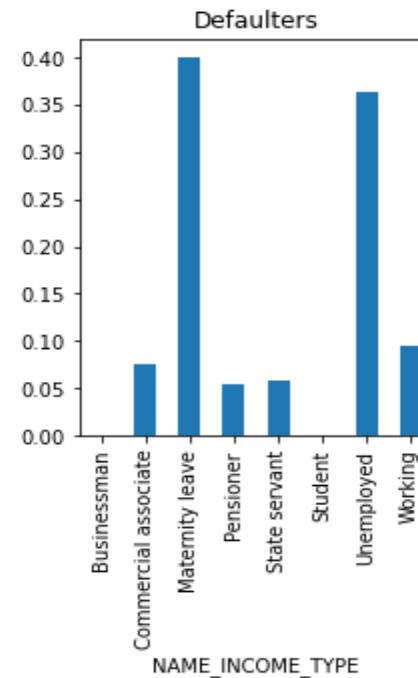
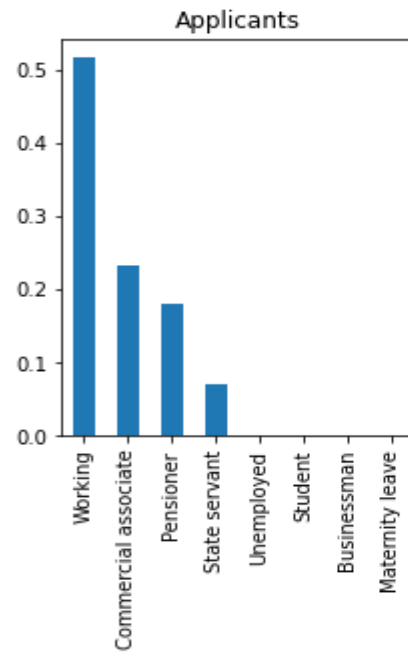
- One of the interesting outcome of the analysis is there are more female applicants, but more percentage of defaulters are male
- Female customers are less risky
- There are more applicants with no children
- The higher the number of children the probability of defaulting is more
- The count of children 9 and 11 are kind of outliers who can be ignored in our analysis



Case Study – Analysis – Income Type



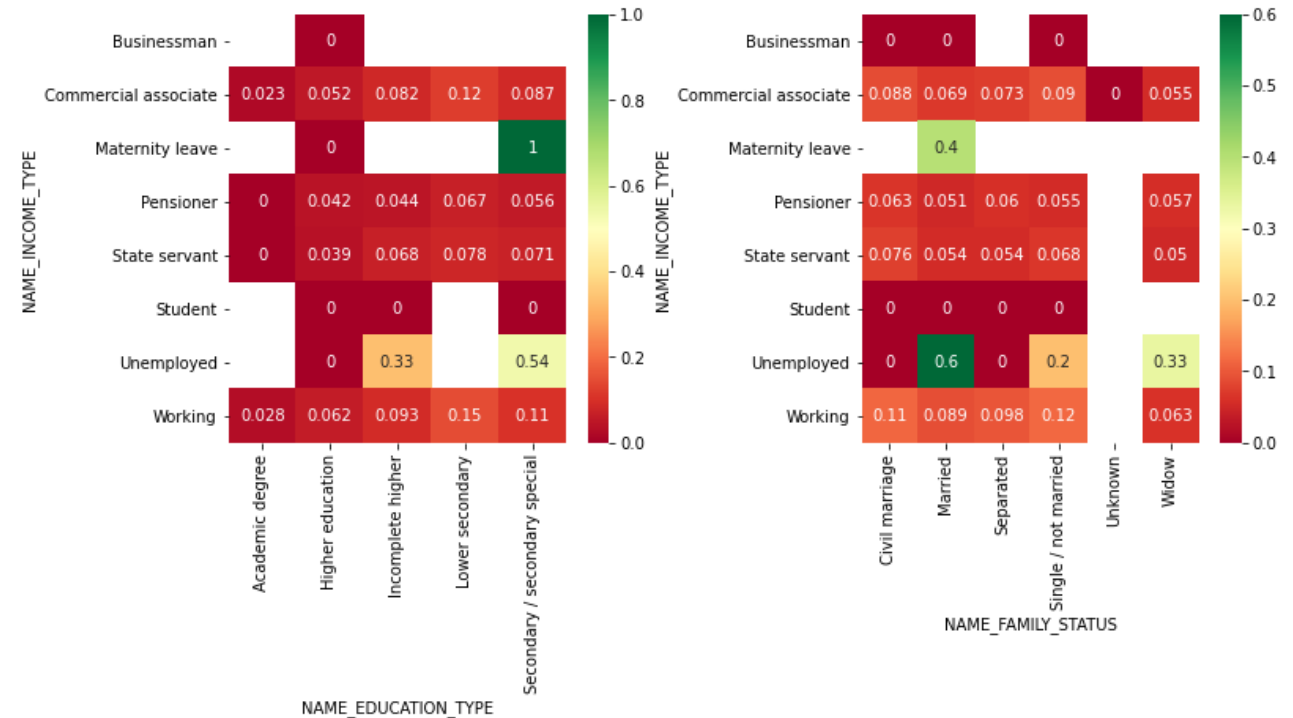
- The working class are the most number of applicants for the loans
- Applicants on Maternity Leave and Unemployed are most probable defaulters
- Businessman and students are 100 % non defaulters



Case Study – Multivariate Analysis



- Left heatmap shows correlation between the income type and education and Target variable
- Right heatmap shows correlation between the income type and family status and Target variable
- Sample Inferences:
 - 54% unemployed with secondary education are defaulters
 - 60% unemployed who are married are defaulters
 - 33% of unemployed widows are defaulters



Case Study – Bivariate Analysis Channel



- Most of the applications coming thru Credit and Cash Offices
- Applicants came thru AP+ (Cash Loans) are more probable defaulters, next probable defaulters are from Contact Center
- Applicants came thru Car dealers are less risky for loan offerings

