

# Credit EDA Case Study



Submitted by : Neha Patil (DS C39)





# Problem Statement

There are two types of risks are associated with loan application in the bank's decision:

- **T0** : If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- **T1** : If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

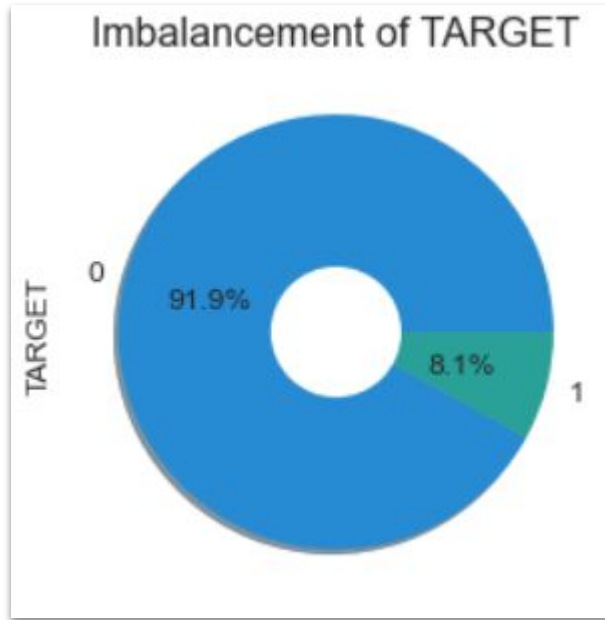
Analysis of the data has been done in Python on Jupyter Notebook. Plots displayed in the presentation are based on analysis.



# Analysis Steps

- Missing value checking ; Handling of Missing values.
- Check for Outliers ; Handling of outliers.
- Data imbalance , Ratio of imbalance.
- Top 10 correlation in both cases; Non-Default and Default.
- Which correlation is most relevant.
- Previous Application Data analysis.

# Imbalanced proportion of TARGET variable



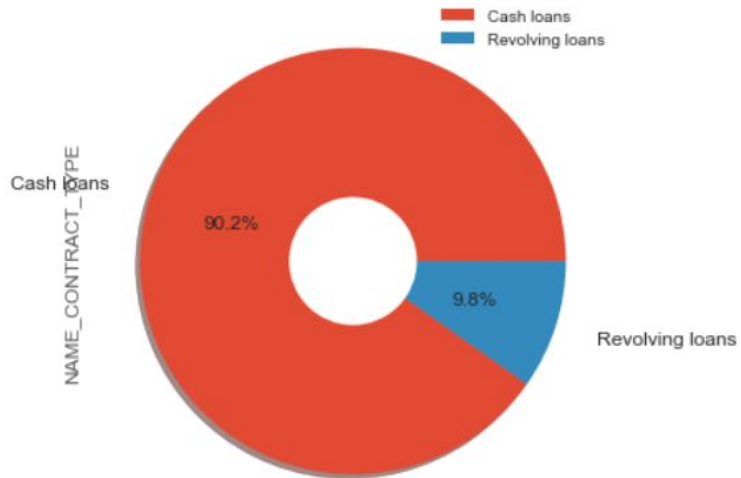
TARGET 1 = defaulters, proportion = 8.1 %

TARGET 0 = non-defaulters, proportion = 91.9%

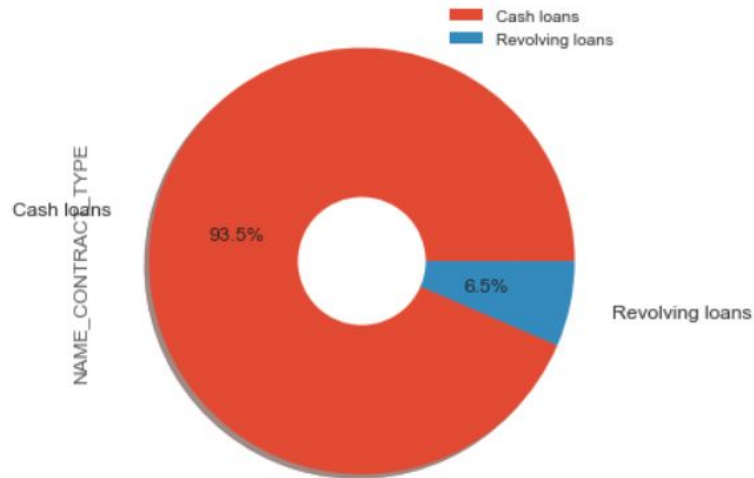


# Proportion of defaulters by type of loan

Distribution of NAME\_CONTRACT\_TYPE by Non-Default Applicants



Distribution of NAME\_CONTRACT\_TYPE by Default Applicants

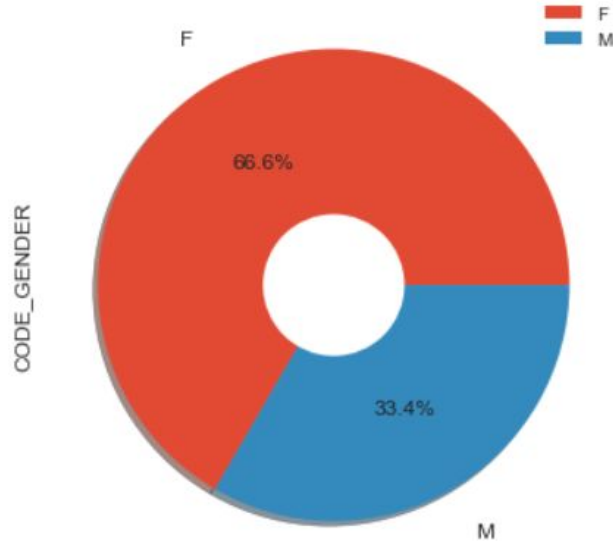


More default proportion for cash loan and less for revolving loan

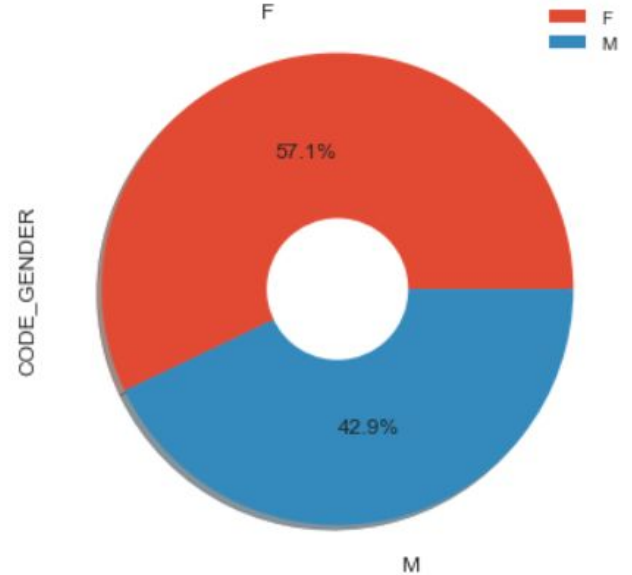


# Proportion of defaulters by gender

Distribution of CODE\_GENDER by Non-Default Applicants

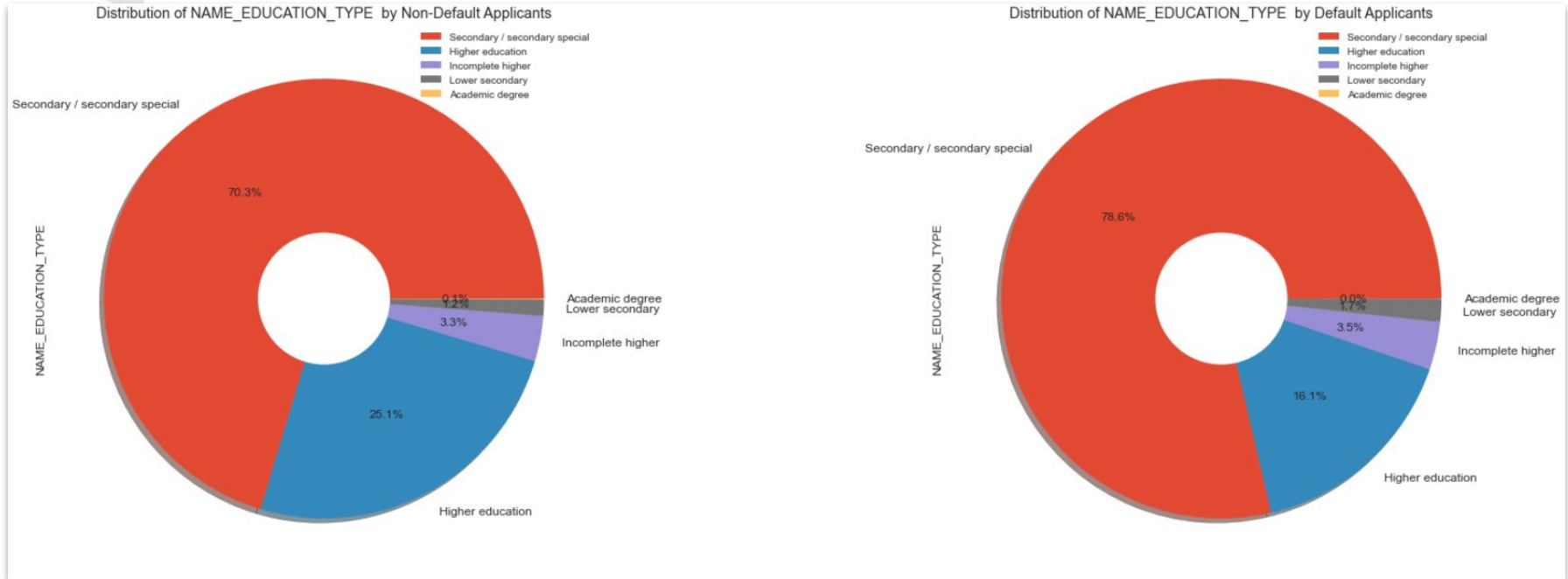


Distribution of CODE\_GENDER by Default Applicants



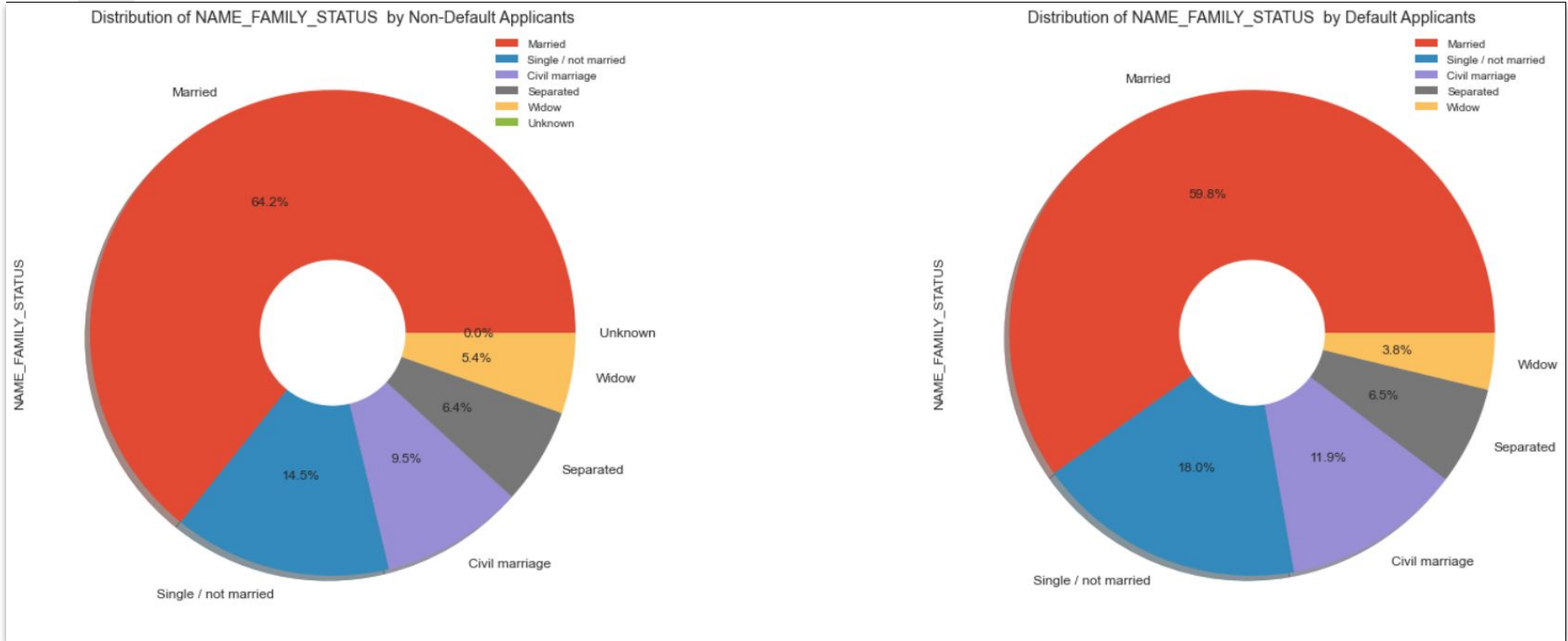
Female applicants are more in both default as well as non default applicants while proportion of male applicants increased for defaulters

# Proportion of defaulters by education type



Secondary/Secondary special type of education default more

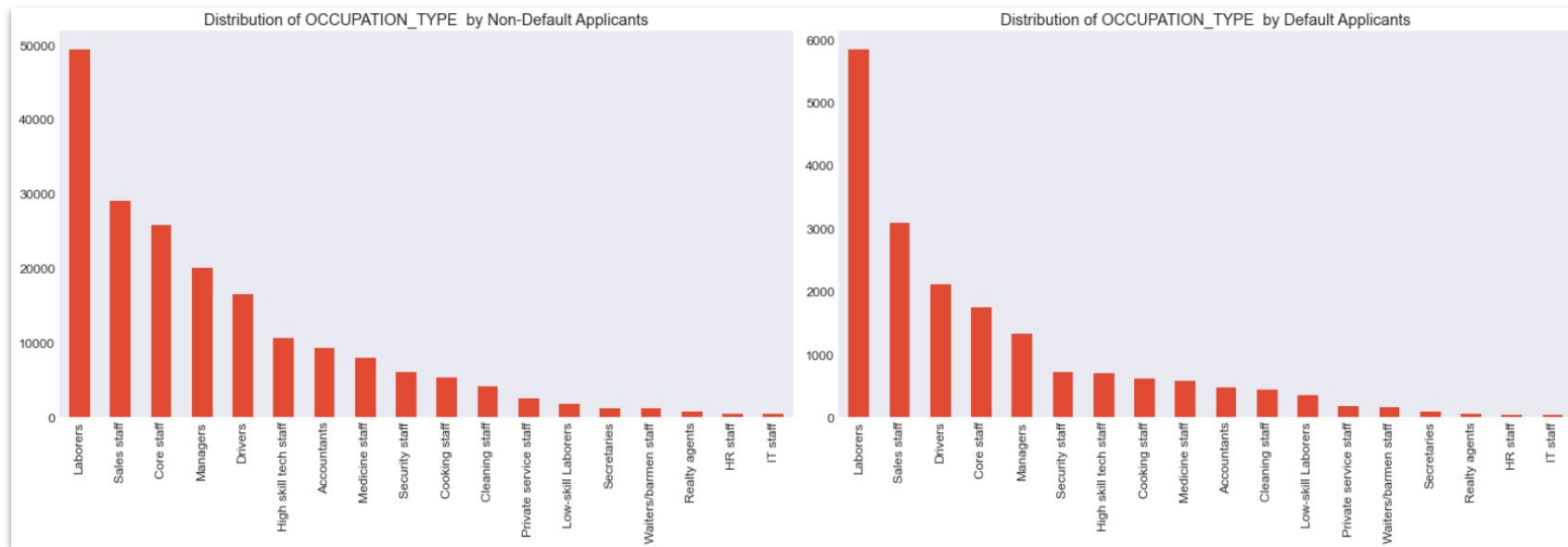
# Proportion of defaulters by family status



Single/not-married, Civil marriages default more



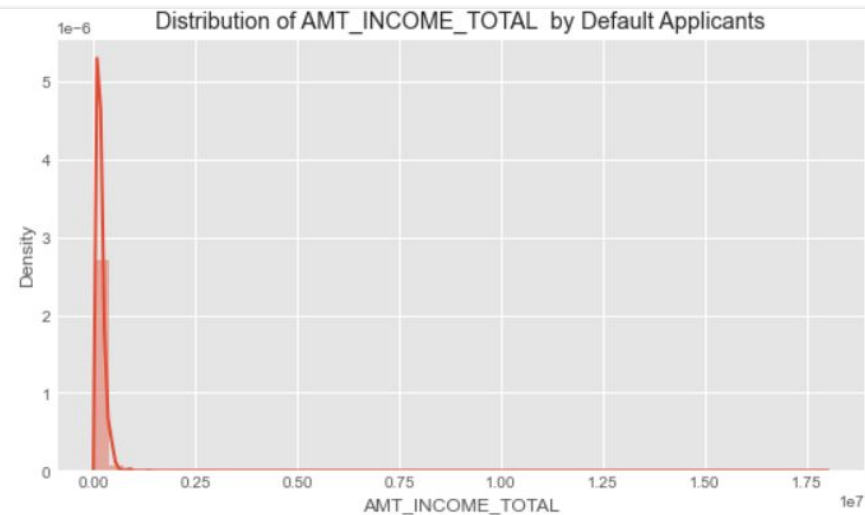
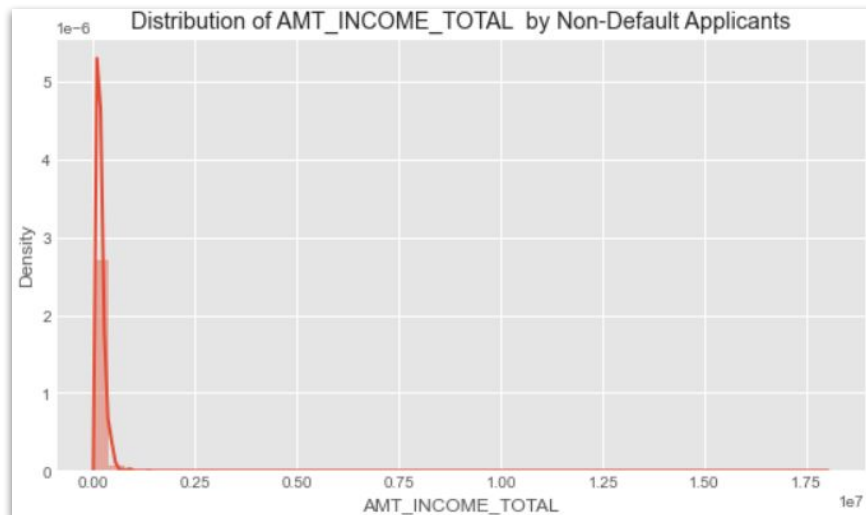
# Proportion of defaulters by occupation



Laborers, drivers, security staff, Cooking staff, low skill staff tend more to default



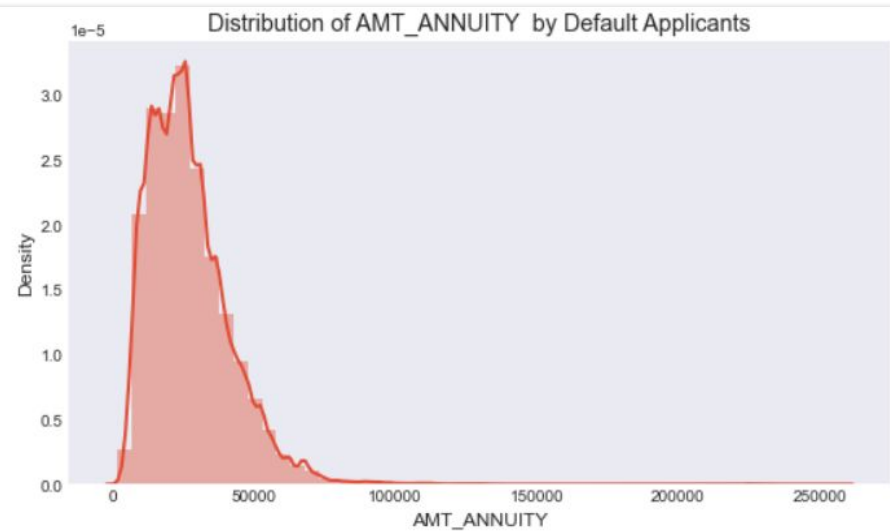
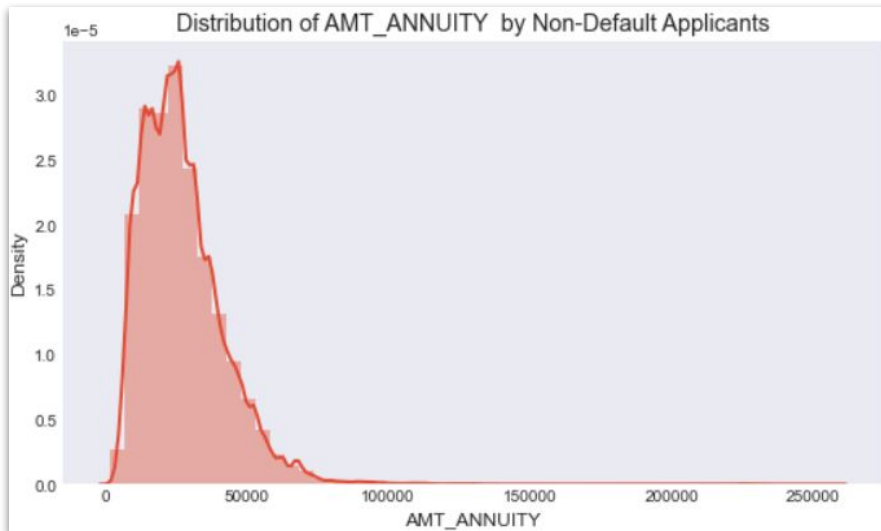
# Proportion of defaulters by Total Income



People with low income in total are more likely to default

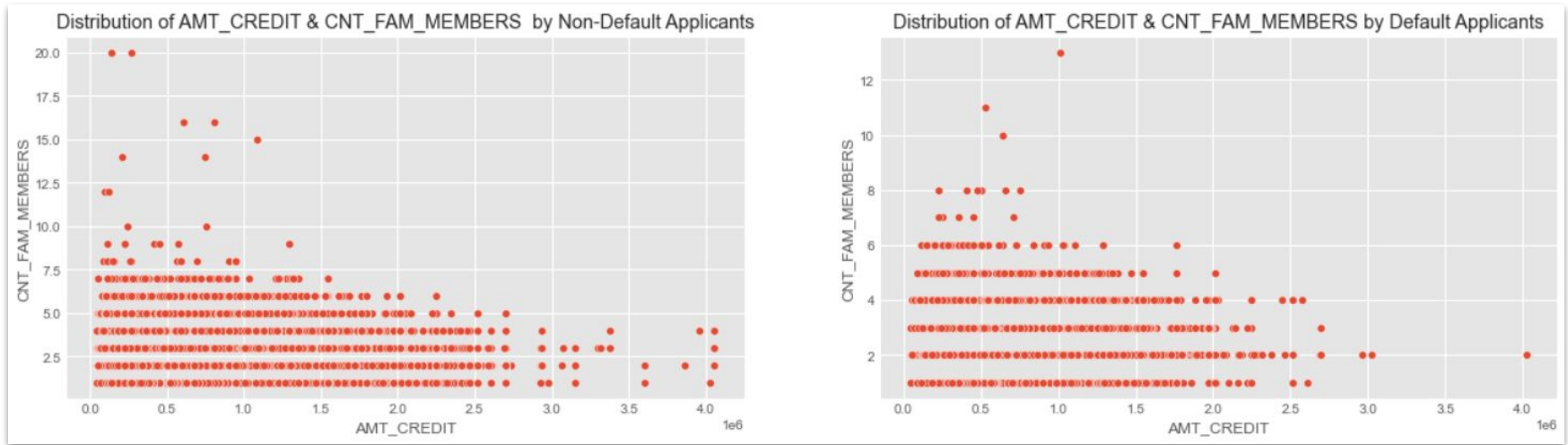


# Proportion of defaulter by Annuity



Defaultness decreases as amount of annuity increases

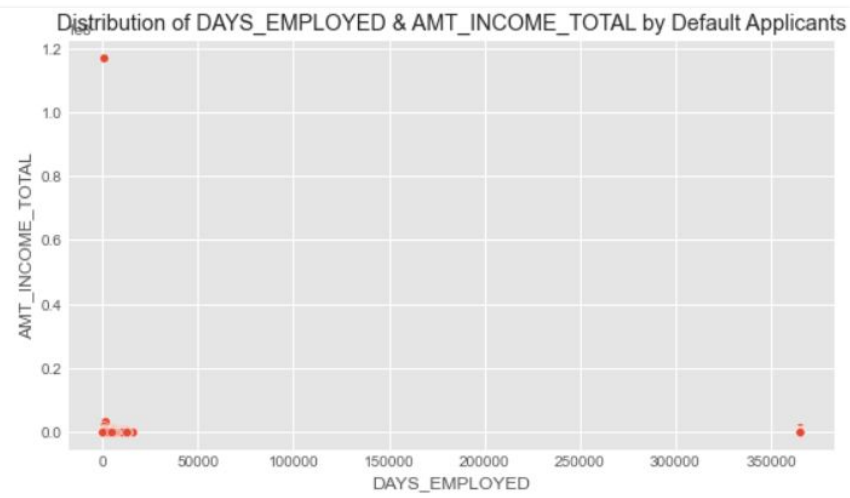
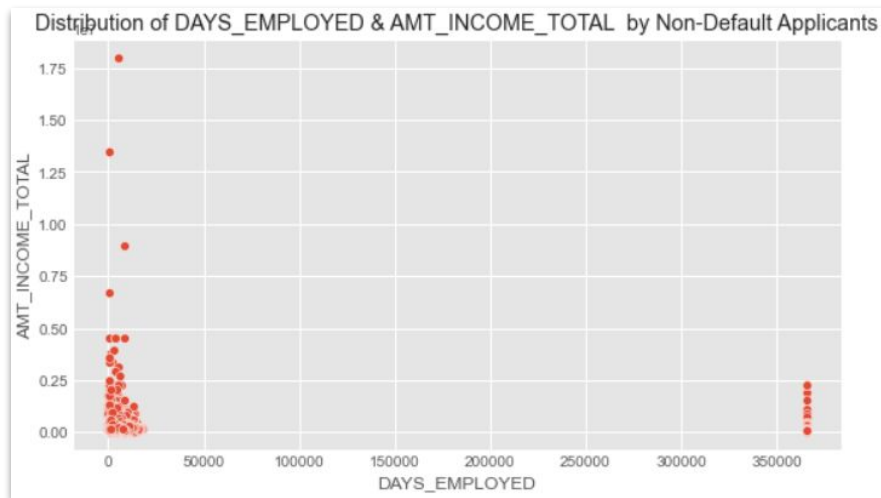
# AMT\_CREDIT Vs CNT\_FAM\_MEMBERS



People with less family members and low credit amount default more

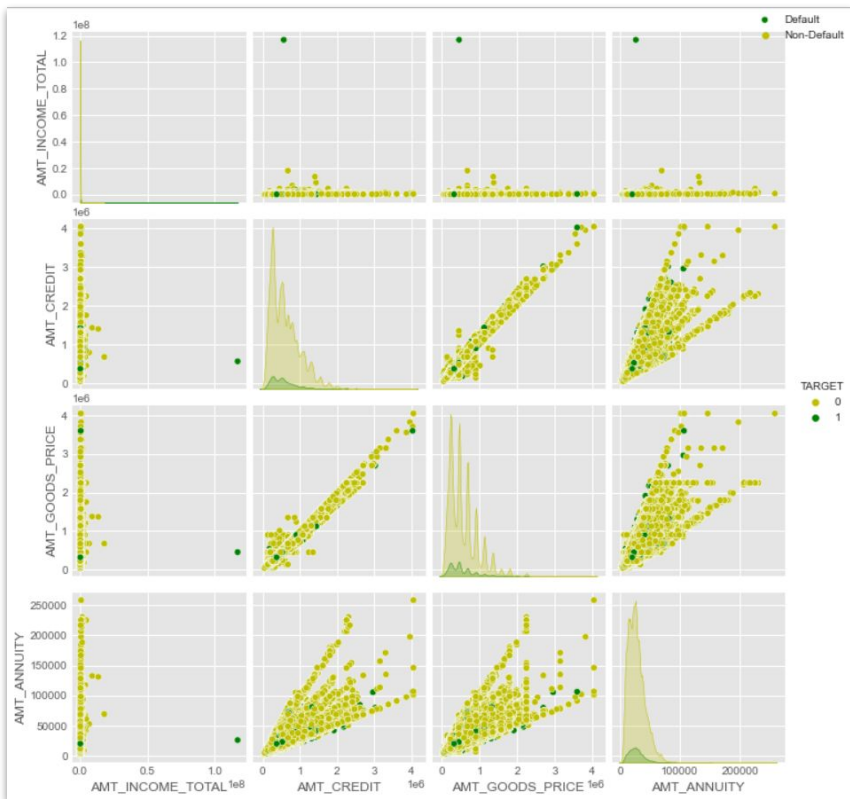


# DAYS\_EMPLOYED Vs AMT\_INCOME\_TOTAL



People with less days of employment and less income tend to default more

# Correlation between various variables



AMT\_GOODS\_PRICE, AMT\_ANNUITY have high correlation for default applications

AMT\_CREDIT and AMT\_GOODS\_PRICE have relation

# Correlation between variables for non-default and default applications

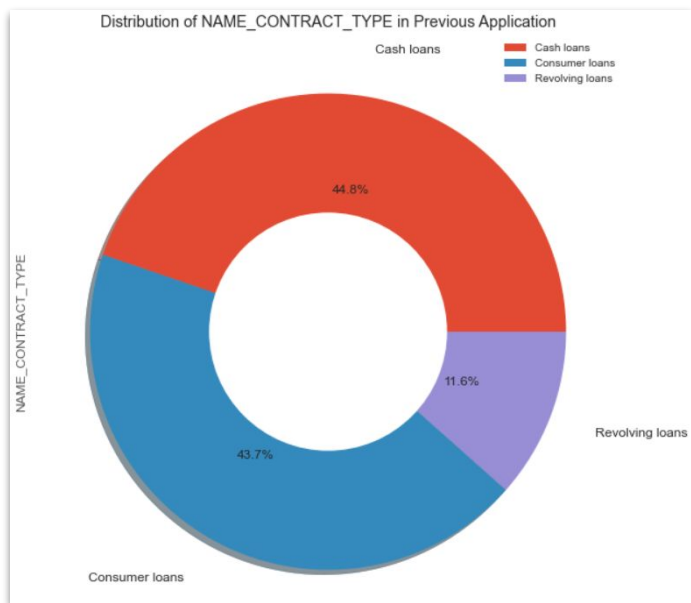


1. Non default - AMT\_GOODS\_PRICE & AMT\_CREDIT, AMT\_ANNUITY & AMT\_CREDIT are highly correlated. CNT\_FAM members and CNT\_CHILDREN are moderately related
2. Default - AMT\_GOODS\_PRICE and AMT\_CREDIT are highly correlated but less than that of in non default applications.



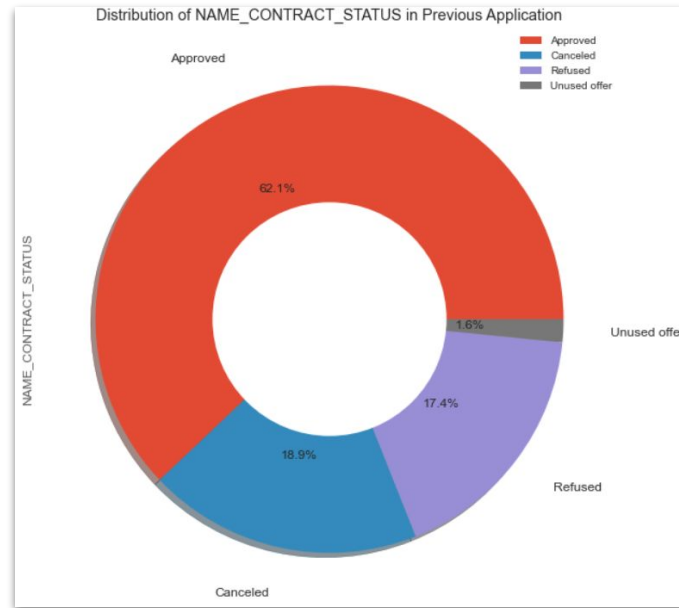
# Previous application data analysis

## 1. NAME\_CONTRACT\_TYPE



Consumer loan have more percentage

## 2. NAME\_CONTRACT\_STATUS



17.4% loan applications are refused. Percentage of unused offers is less





# Insights of Application Data

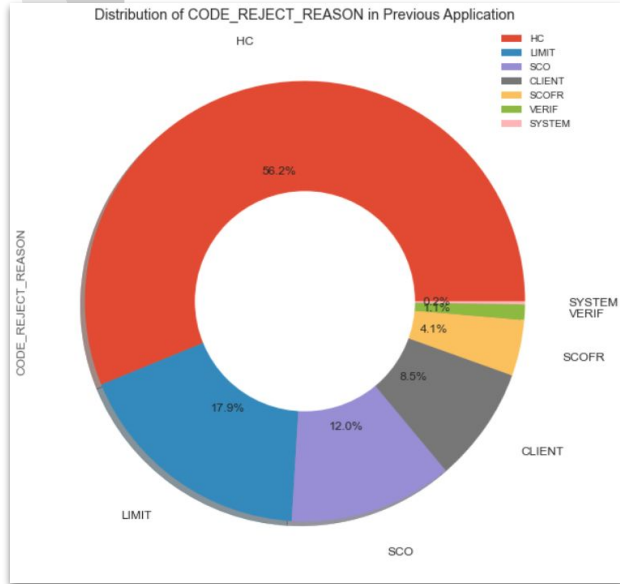
## Non - Default

- Applicants with no/low income i.e. students and business persons have no tendency to default.
- Applicants with more age or above 50 have low tendency to default.
- Academic degree applications are not likely to default.
- applicants having experience of 40+ are less likely to default.
- High income , about > 70,000 have low tendency to default.
- people with moderate family members are no likely to default.

## Default

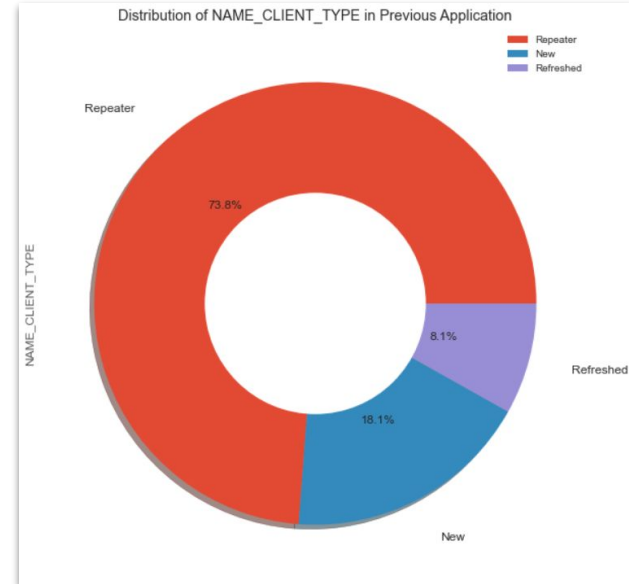
- Secondary and lower secondary education type has seen to be more inclined towards defaulted.
- Male applicants have more tendency of Defaultness.
- Single or Civil marriage type of family are more likely to default.
- Drivers, Low skilled laborers, Security staff, Laborers, Cooking Staff type of occupations have greater default rate as compared to others.
- Organizations such as Transport, Industry (type 13, type 8) have more tendency to defaultness.
- age between 20-35 are likely to default more.

### 3. CODE\_REJECT\_REASON



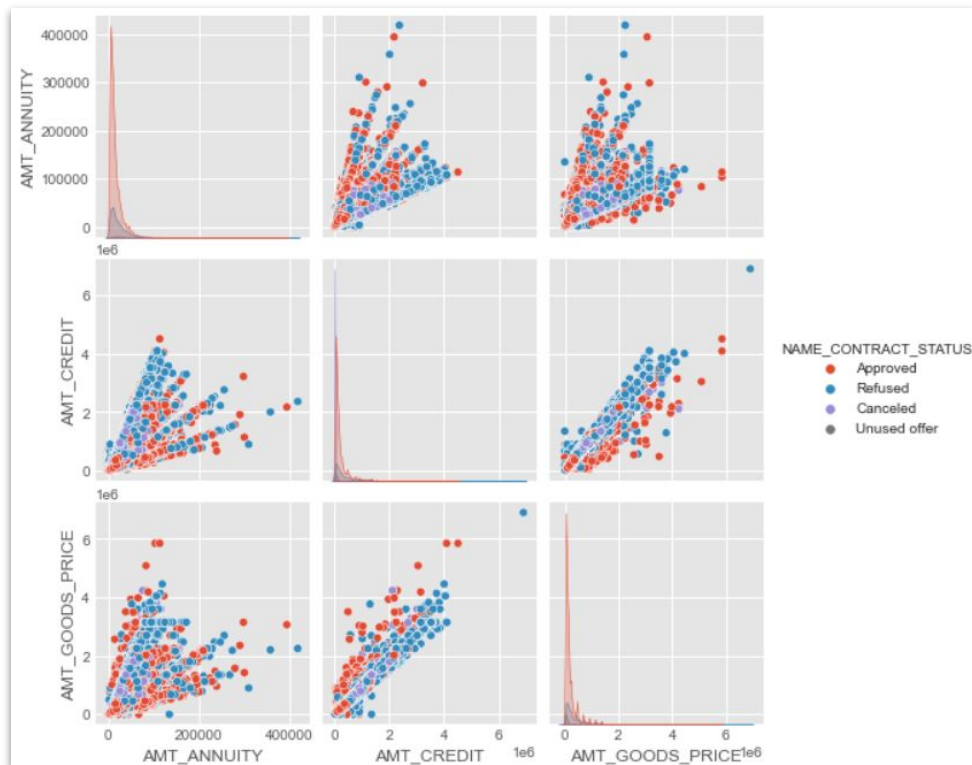
Major percentage of rejection reason is HC

### 4. NAME\_CLIENT\_TYPE



Most of the clients are repeaters

# Correlation between various variables



AMT\_ANNUIITY & AMT\_CREDIT are highly correlated.



# Insights of Previous Application data

1. The percentage of loan type 'Revolving loan' in NAME\_CONTRACT\_TYPE is less in previous applications but it has maximum percentage have difficulties in payment and more likely to default.
2. We can observe from graphs that, applications with 'Revolving loan' and 'Refused' previous application are more tend to default. These can be driving variables of Defaultness.