EDA Case Study

Bank Loan Application Data

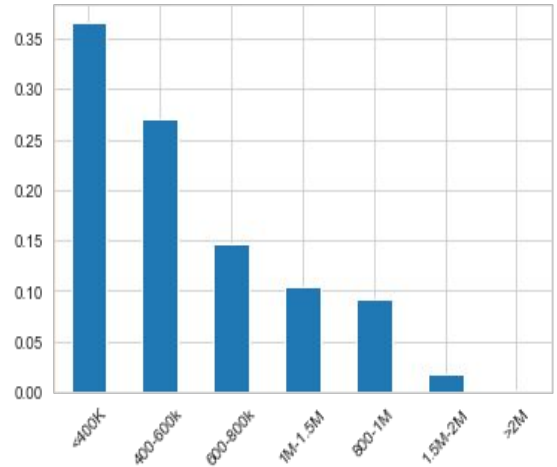
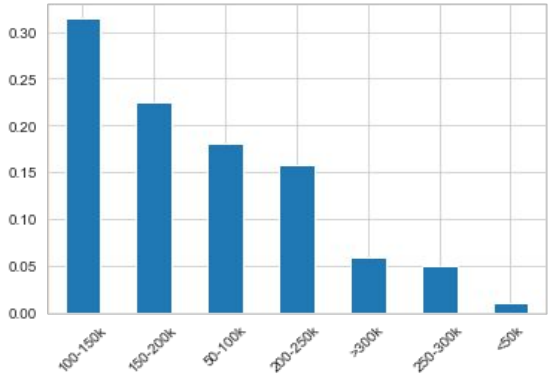
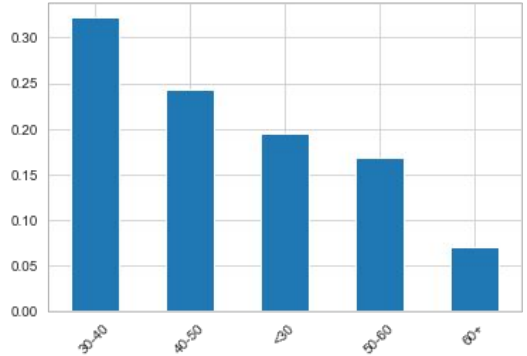
Analysis of Application data file

- Load Application data file and check the shape of the file
- Check the column length
- Print the data of the file and understand the used and unused data.

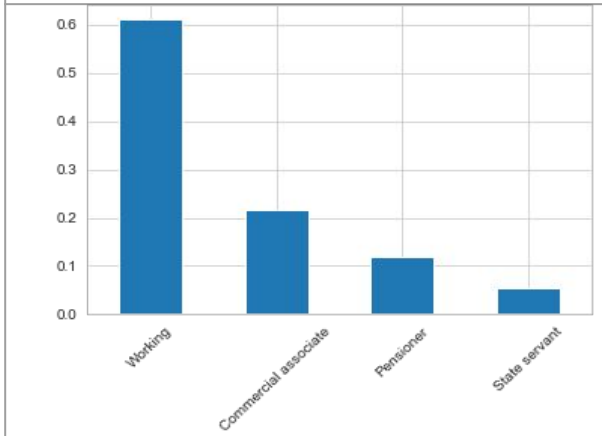
Data Cleaning

- Remove all the column which are null more than 50%
- Also remove all the unused column which are not required in analysis.
- Treat the null values with column ‘mode’ value
- Calculate the null % of each columns
- Change the Data type of the column which are wrong datatype.
- Rename the column name and correct the data from negative to positive.
- Create categorical data of continuous columns to analysis in batter ways.
- Calculate the defaulters and non-defaulter %.
- Create two data frame based on ‘TARGET’ column.

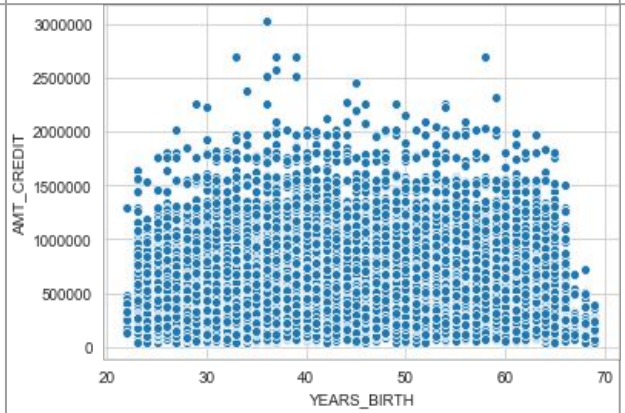
Default(Target=1) analysis and Visualize Univariate and bivariate

Credit amount Group	Total Income Group	Age Group
		

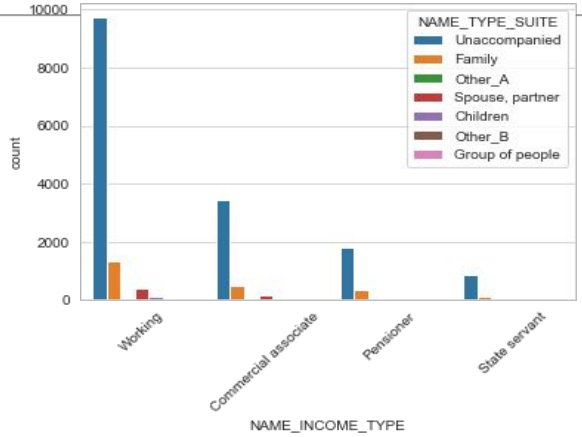
Income Type



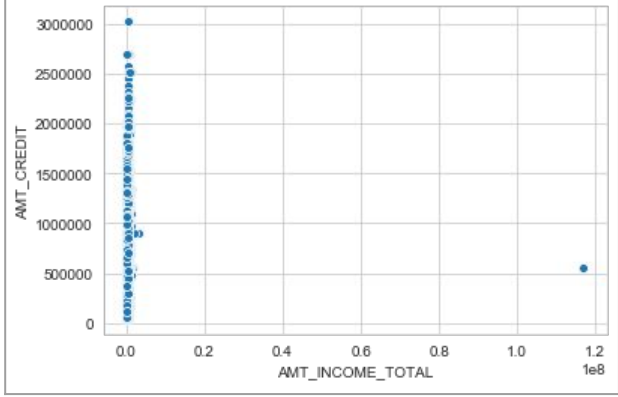
Age and Credit amount



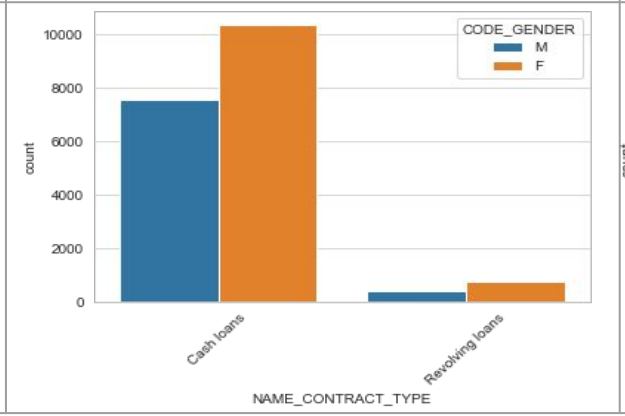
Income Type & Suite



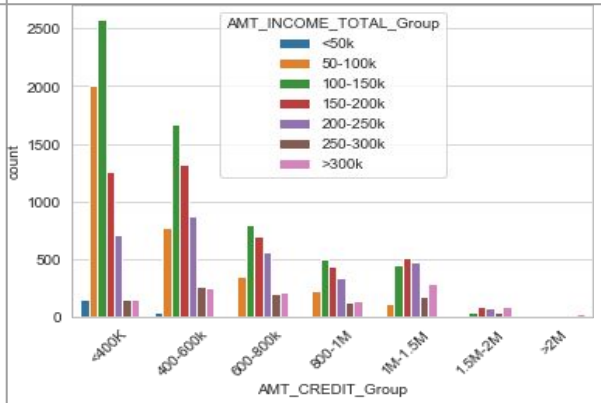
Total Income & Credit



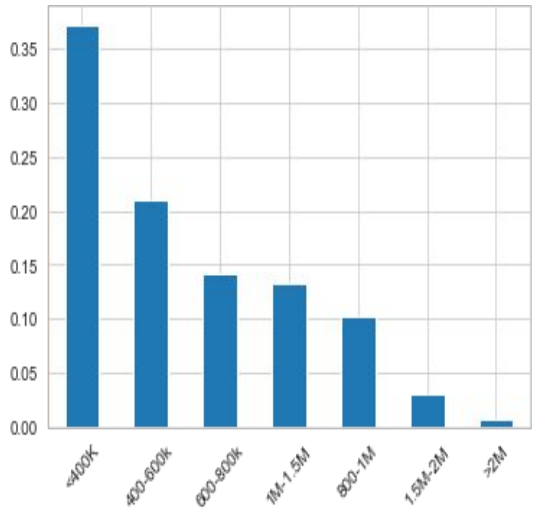
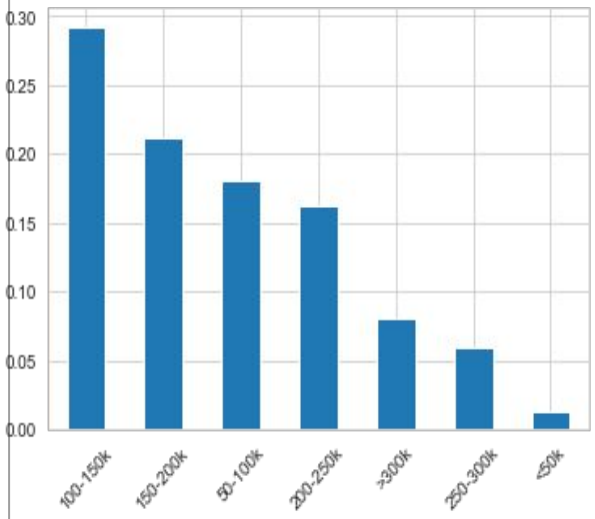
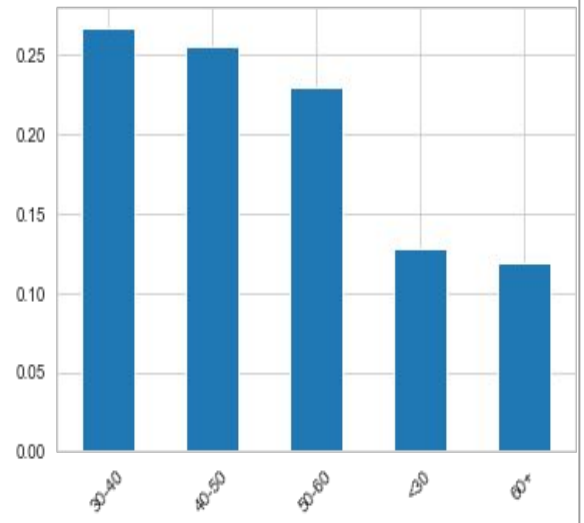
Genter & Loan More Defaulter



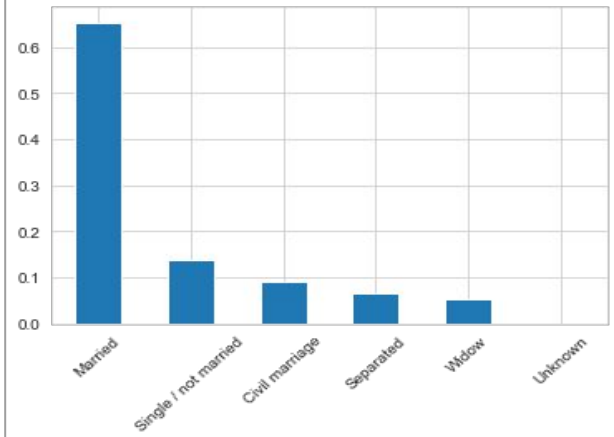
Credit & Income Groupwise



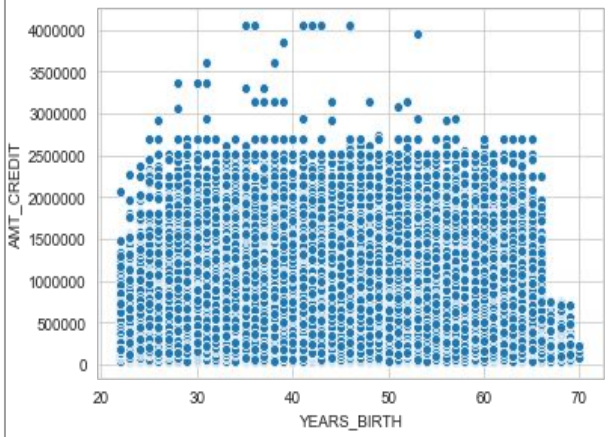
Non Default(Target=0) analysis & Visualize Univariate and bivariate

Credit amount Group	Total Income Group	Age Group
		

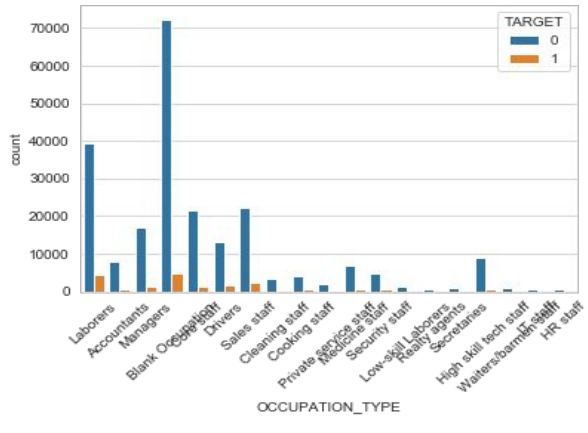
Non default Based Family Type



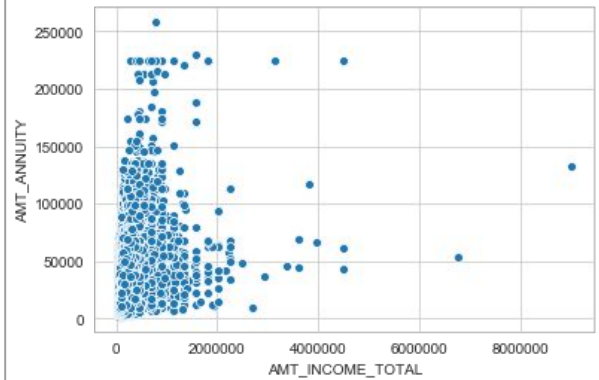
Age and Credit amount



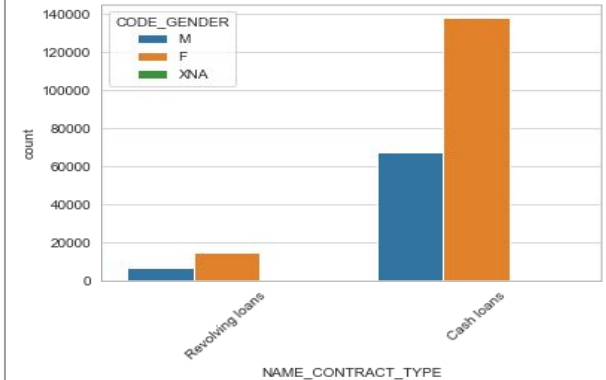
Occupation and Target Column



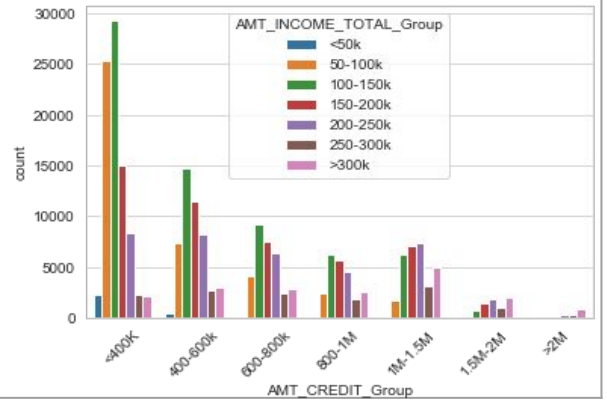
Total Income & Annuity



Genter & Loan More Defaulter



Credit & Income Groupwise



Merge Application and Previous Application

- Merge Both file , create df3 dataframe from both
- Drop Column who have more than 40 % missing value
- Calculate mean of messing Value
- Droup more column that look like not useful for this analysis

Data Cleaning:

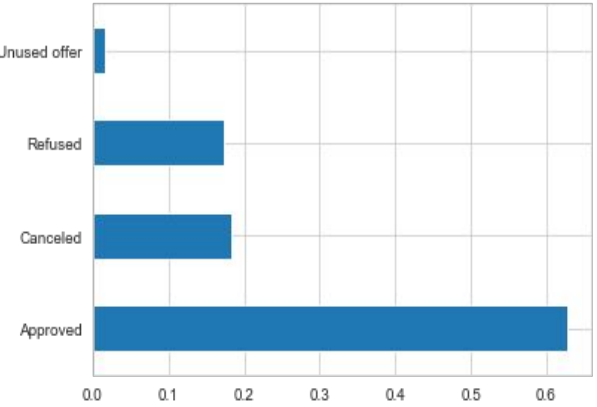
- CODE_GENDER fillNa value with ‘F’ because its having percentage
- FLAG_OWN_REALTY fillNa value with ‘Y’ because its having percentage
- AMT_ANNUITY_x and AMT_ANNUITY_y updated with mean value (Both mean and median have almost same value(minor diff))
- CNT_FAM_MEMBERS ‘Na’ updated with median (Because mean and median almost same.

- NAME_CONTRACT_STATUS filter NaN value because very less row having NaN

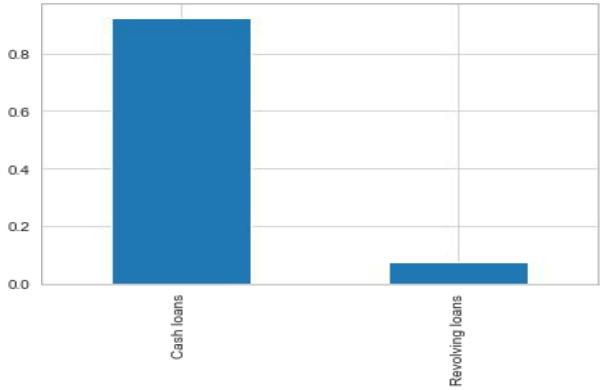
Univariant

- Converted DAYS_BIRTH into Years
- Perform Univaraint on NAME_CONTRACT_STATUS
- Perform on NAME_CONTRACT_TYPE_x
- Perform on Code Gender
- Perform on CODE_GENDER
- Perform on NAME_INCOME_TYPE
- Perform on NAME_EDUCATION_TYPE
- Created Age group for DAYS_BIRTH
-

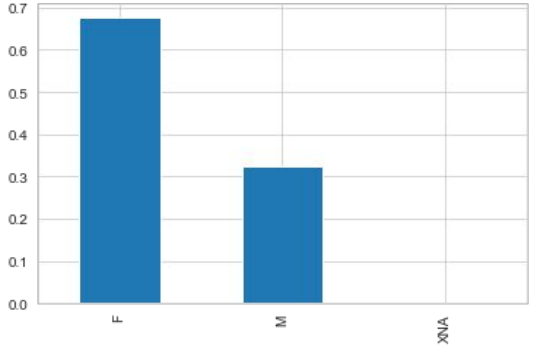
NAME_CONTRACT_STATUS



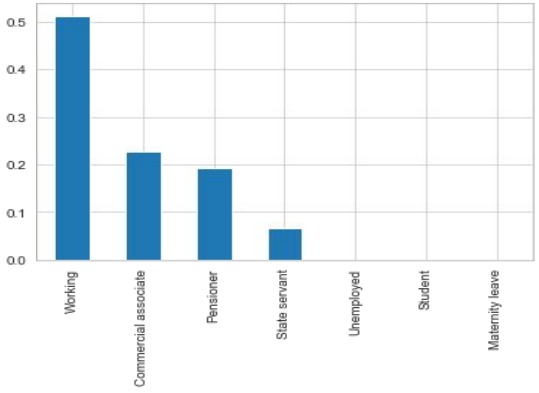
NAME_CONTRACT_TYPE_x



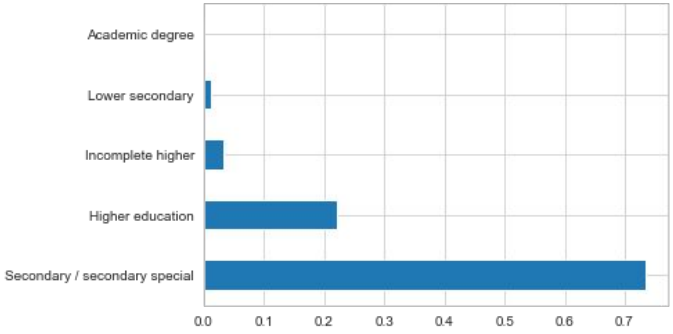
CODE_GENDER



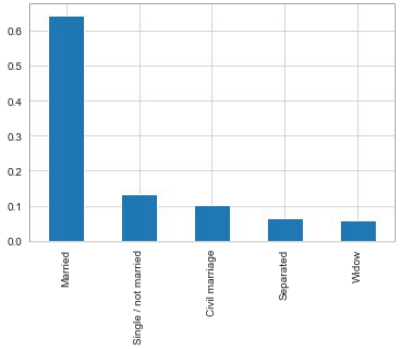
NAME_INCOME_TYPE



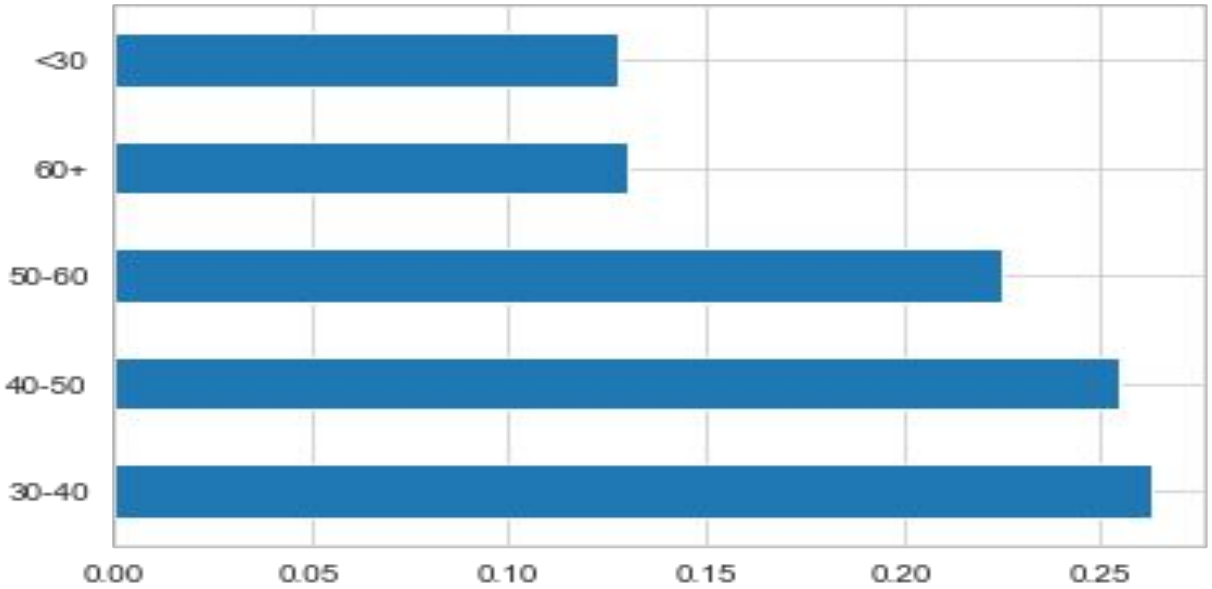
NAME_EDUCATION_TYPE



NAME_FAMILY_STATUS



YEARS_BIRTH

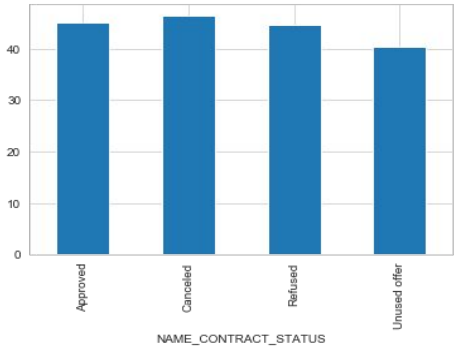


Bivariante(With NAME_CONTRACT_STATUS)

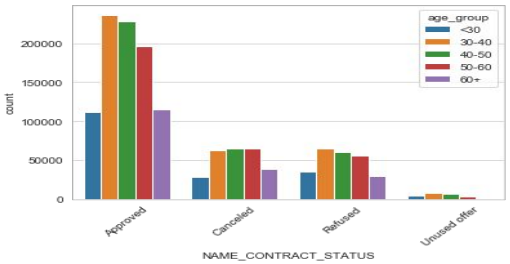
Perform Bivariate analysis bellow Attributes

- YEARS_BIRTH
- Age_group
- NAME_CONTRACT_TYPE_x
- CODE_GENDER
- AMT_INCOME_TOTAL
- AMT_ANNUITY_y
- NAME_INCOME_TYPE
- NAME_FAMILY_STATUS
- CNT_PAYMENT
- AMT_APPLICATION
-

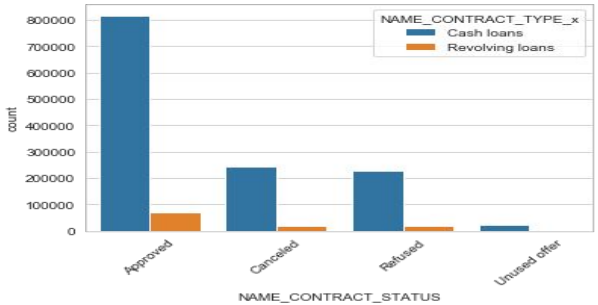
YEARS_BIRTH



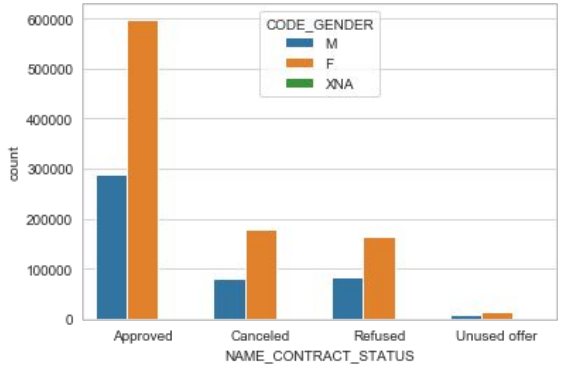
age_group



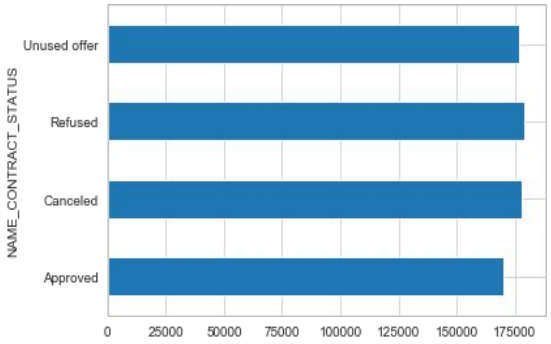
NAME_CONTRACT_TYPE_x



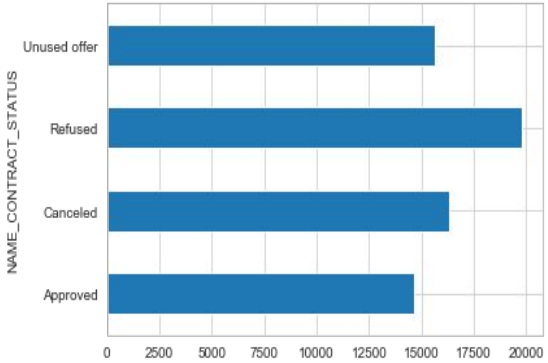
CODE_GENDER



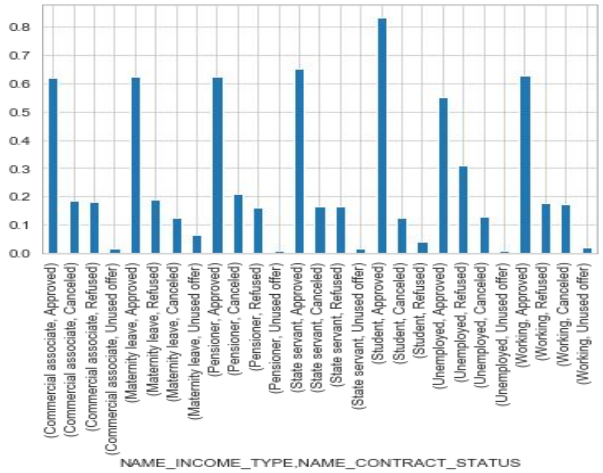
AMT_INCOME_TOTAL



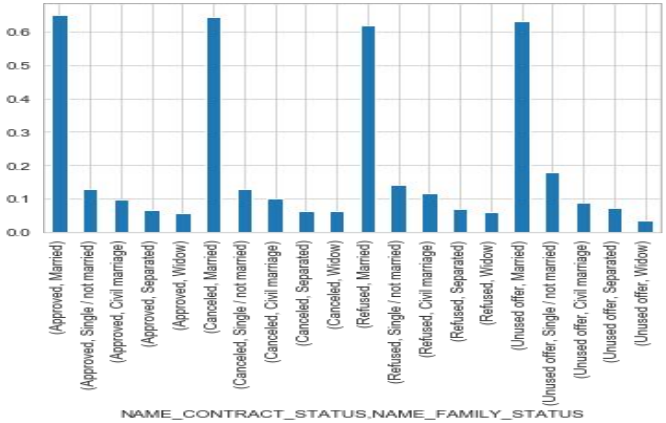
AMT_ANNUITY_y



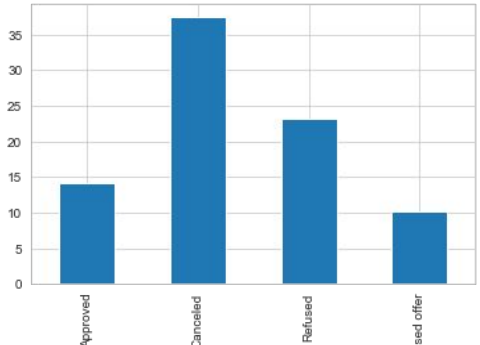
NAME_INCOME_TYPE



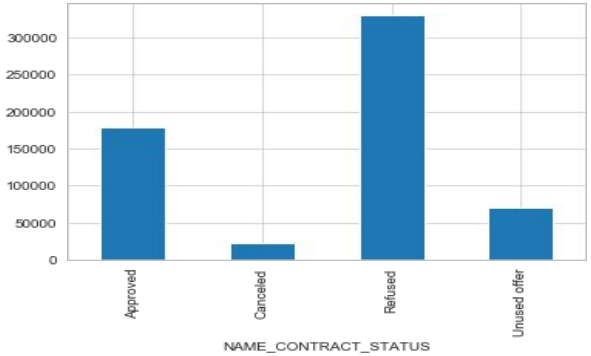
NAME_FAMILY_STATUS



CNT_PAYMENT



AMT_APPLICATION



Conclusion Of Analysis

- We try to show many reason of rejection or approval of loan
- Try to find Outliers form Univariant and bivariate
- Perform Comparison for Target with many Column in application to see the reason for any action(Rejection or Approval)
- Merge the both application and perform Univariate to see who applied the loan (Category)
- Perform bivariante on merge dataframe from ‘NAME_CONTRACT_STATUS’ with many column to see the reason of approval ,Canceled, Refused