CASE STUDY FOR CREDIT ANALYSIS USING (EDA)

Bhaskar Mishra

INTRODUCTION

- This assignment aims to give an idea of applying EDA(exploratory data analysis) in a real business scenario.
- The basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Steps to do EDA

- PROBLEM STATEMENT UNDERSTANDING
- DATA UNDERSTANDING(APPLICATION DATA)
- DATA CLEANING: IDENTIFICATION OF MISSING VALUES AND TREATMENT OUTLIER AND TREATMENT
- SEGMENTATION (OPTIONAL)
- DATA IMBALANCE
- SANITY CHECKS
- UNIVARIATE ANALYSIS --> ANALYZING ONE VARIABLE AT A TIME :HISTOPLOT
- BIVARIATE ANALYSIS --> ANALYZING TWO VARIABLES AT A TIME: BARPLOT
- ANALYSIS OF PREVIOUS APPLICATION DATA SET WITH ABOVE STEPS

Data Understanding (Application Data)

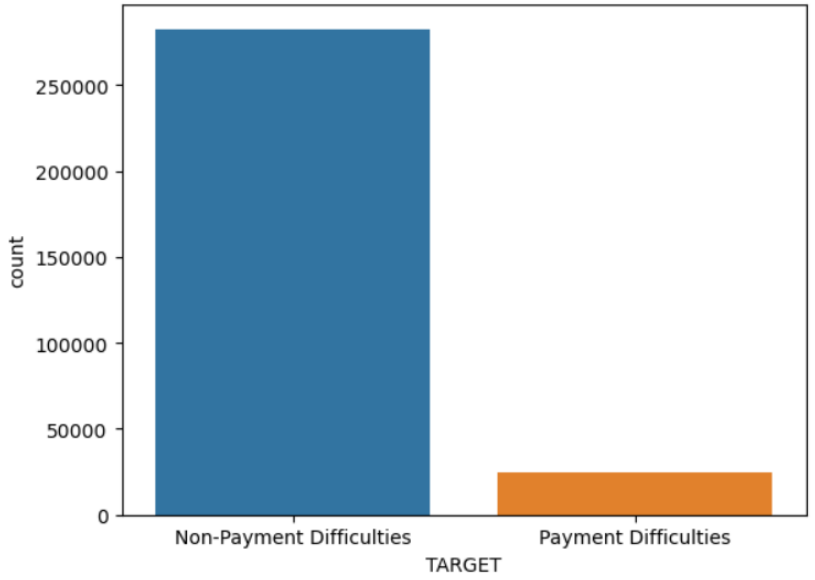
- Checking Shape, Info, D-Types , size and Describe of the application data and pervious application data to get quick understanding of the data.

Data Cleaning

- Identification of Missing Values and Treatment:
- There were several Columns with missing value percentage greater than 40% so we dropped them.
- Remaining columns with missing values we imputed them with mean/median/mode as required.
- Outliers: There were several columns which have outliers present in them.
- Depicted through boxplot in columns named as :
- AMT_ANNUITY
- AMT_GOODS_PRICE
- NAME_TYPE_SUITE
- OCCUPATION_TYPE

DATA IMBALANCE

- When we observed the data set it was highly imbalanced almost 91.9% clients were Non-Payment Difficulties(0) and about 8.1% for Payment Difficulties(1).
- 0= Non-Payment Difficulties
- 1= Payment Difficulties

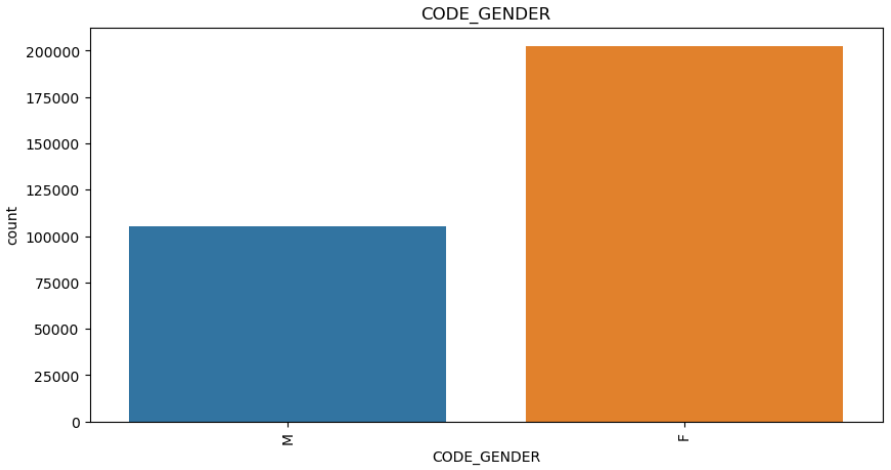


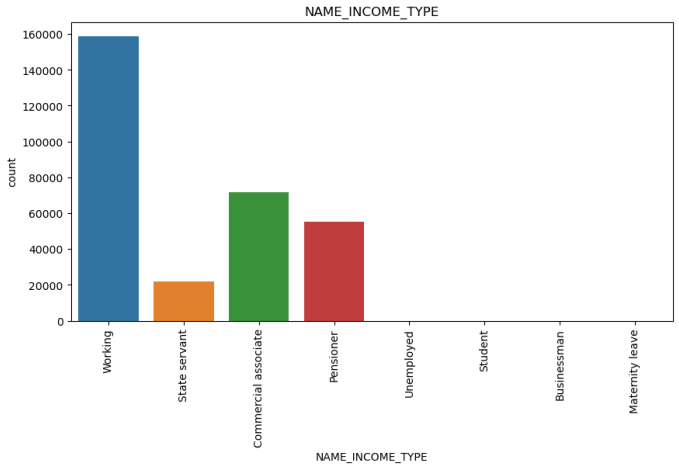
SANITY CHECKS

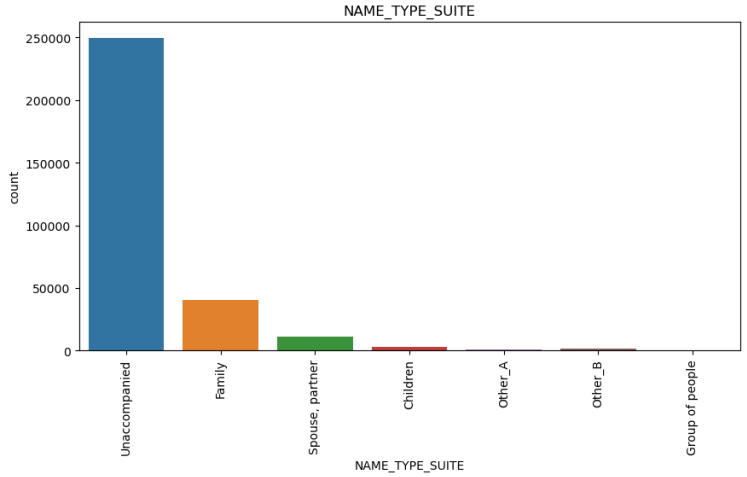
- Doing analysis through the data observed some columns have negative values. So to correct them we did sanity checks in these columns named as :
  - DAYS_BIRTH", "DAYS_EMPLOYED", "DAYS_REGISTRATION", "DAYS_ID_PUBLISH","DAYS_LAST_PHONE_CHANGE

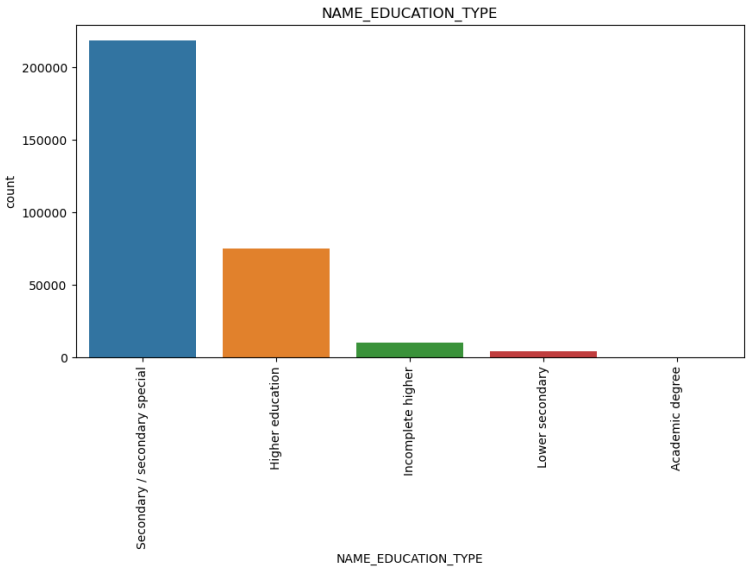
Bivariate Analysis

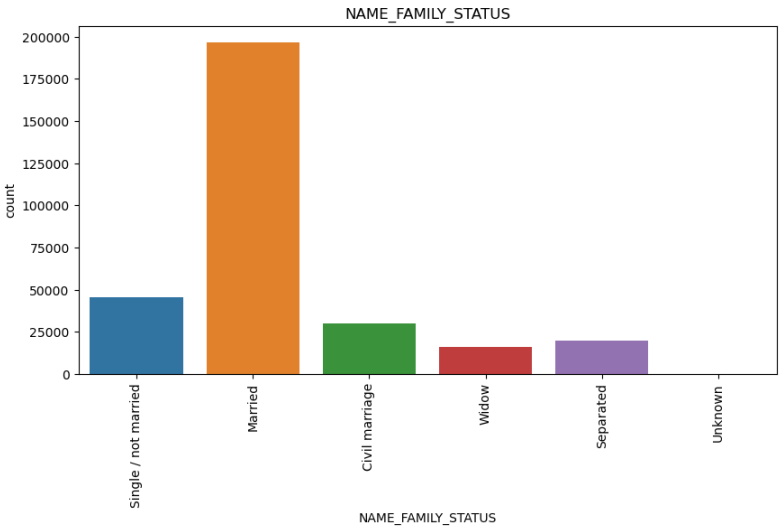
- Conclusion from analysis: We can say that most client are females, working, have secondary education and are married.
- We can say taking loan from the bank someone accompanied with them as well.







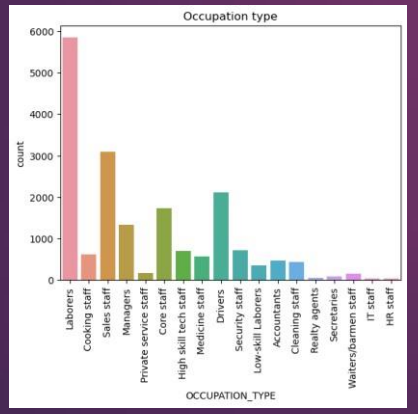

NAME_EDUCATION_TYPE
count
200000
150000
100000
50000
0
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

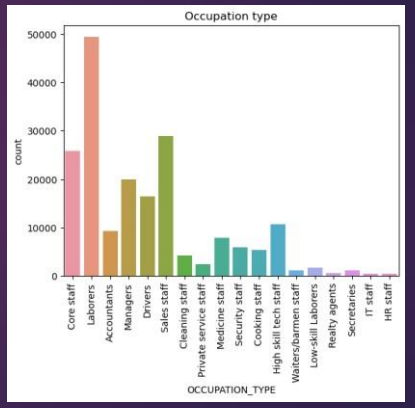

NAME_FAMILY_STATUS
count
200000
175000
150000
125000
100000
75000
50000
25000
0
Single / not married
Married
Civil marriage
Widow
Separated
Unknown
NAME_FAMILY_STATUS

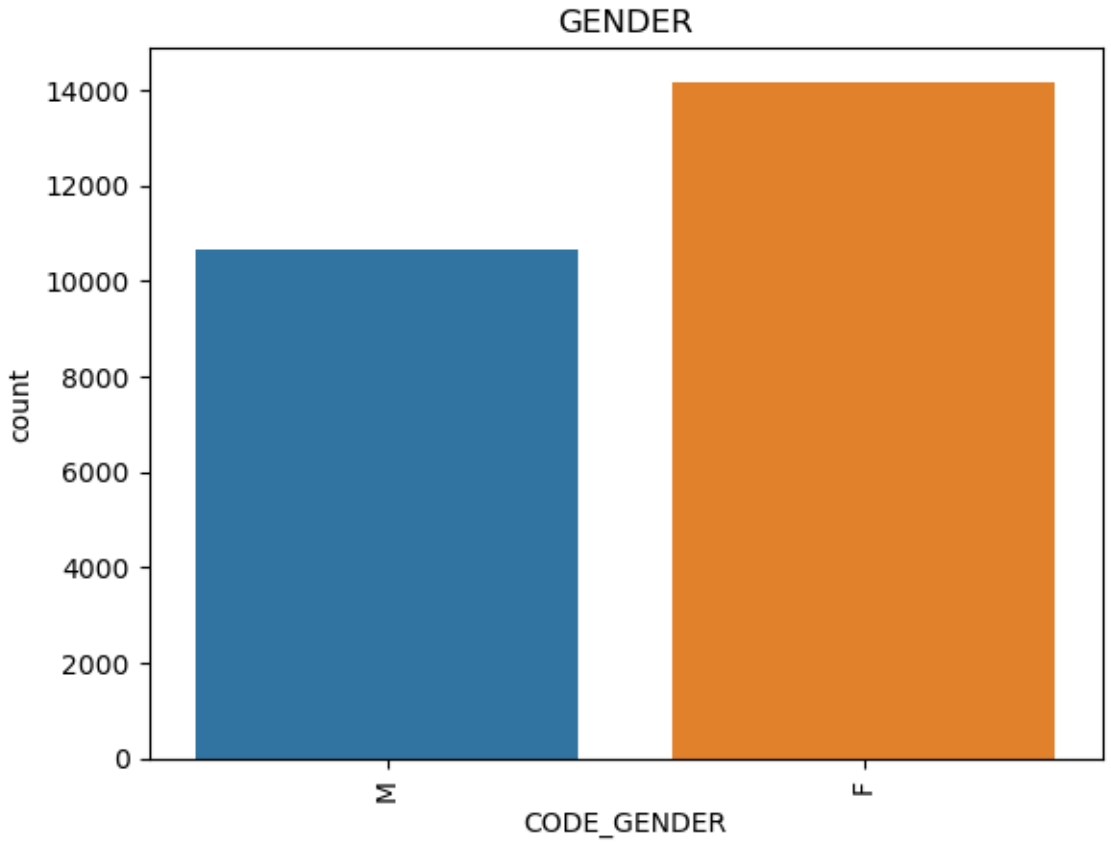
Bivariate Analysis

CONCLUSION FROM ANALYSIS:

WE OBSERVED THAT THERE IS SHARP INCREASE IN LABORERS CATEGORY IN PAYMENT DEFAULT CATEGORY AND ALSO MALE LABORER HAVE MOST DIFFICULTIES IN PAYING LOAN.






GENDER
14000
12000
10000
8000
6000
4000
2000
0
count
M
F
CODE_GENDER

CONCLUSION

- Data is Highly Imbalanced 91.9% is Client with Non-Payment Difficulties and 8.1% is client with Payment Difficulties.
- Imbalance Ratio is 11.36%.Cash Loans are more popular than Revolving loans.
- Females have higher chance of applying for a loan
- Majority of clients who applied for loan are from Working class.
- Client with Secondary Education are more likely to apply for the loan.
- Count of married people applying for a loan is higher than the rest.
- Majority of the clients have House/Apartment.
- Laborers are more frequent customer of bank.
- Most of the client who applied for loan are Unaccompanied.

- Female with the Secondary Education are the most defaulter.
- Client with married marital status and secondary education is at most risk of being a defaulter.
- Higher education client is least like to default on their loan.
- Consumer loans have the highest no of counts and also their approval rating is also the highest.
- Repeater clients are the one who have the highest frequency among the name_client_type.
- Bank takes large number of days to approve a loan in comparison with to cancelling or refusing a loan.
- Banks take most amount of time for new clients to make a decision.
- Average time taken for approval of loan application is lesser for Defaulters.
- Most of the defaulter are from the Repeater clients.
- Client with housing type as Office apartment and Co-op Apartment are the one with high Average Credit amount among defaulters category.

Recommendations

- Bank should target client with higher education as they have lesser number of non defaulters.
- Male laborers, female sales staff and male drivers have the highest payment difficulties and so the banks should put a capping on the loan amount.
- For Undisclosed goal category Banks should increase the interest rate for them.
- Non Defaulters have more cancellation and refusal as compared to the defaulters which ideally, should not be the case.
- Clients who have co-op apartment accommodation or office-apartment, Banks should put a capping on amount on their loan as they have high defaulters.
- Excise caution when giving loan to 30-40 age group people and carefully examine their loan application as they have large number of defaulters.

THANKYOU