

Bank Loan Case Study



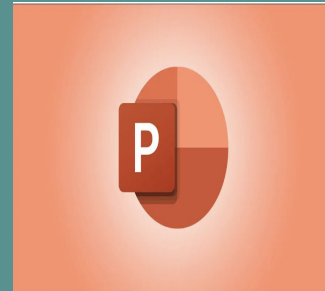
Problem Statement

- Present the overall approach of the analysis in a presentation. Mention the problem statement and the analysis approach briefly.
- Identify the missing data and use appropriate method to deal with it.
- the missing value, what should be the approach. Clearly mention the approach.
- Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier. Again, remember that for this exercise, it is not necessary to remove any data points.
- Identify if there is data imbalance in the data. Find the ratio of data imbalance.



Technology Used

- I have used Microsoft Excel to do the analysis and Jupyter for cleaning the Application data as it is having large number of rows.
- I used Microsoft presentation to make ppt for the project



Clean Data

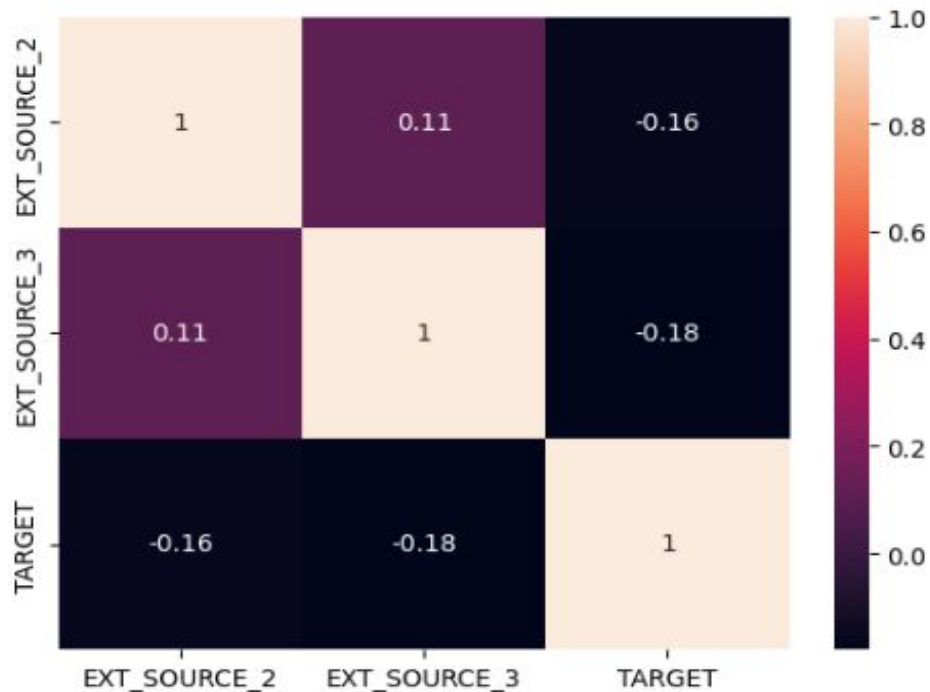
- Data cleaning is a crucial step in any data analysis process as it ensures the data is accurate, reliable, and consistent. Without proper data cleaning, the analysis can become inaccurate, incomplete, and inconsistent, potentially leading to serious consequences in decision-making. Ensuring data quality is essential for making informed and effective decisions.



Data Cleaning

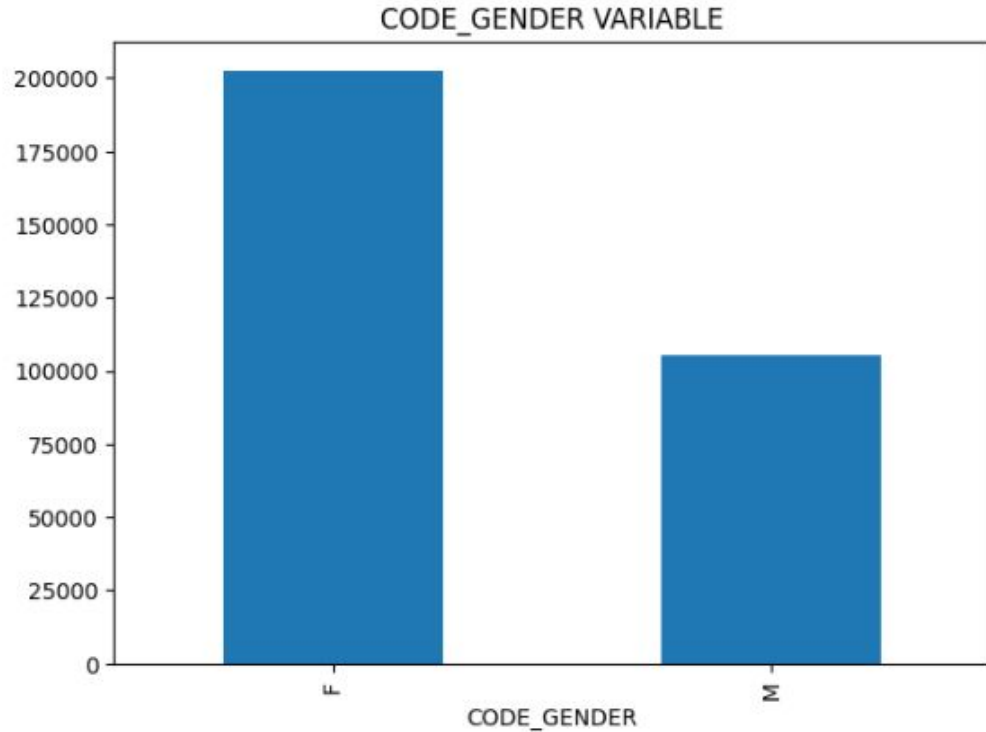
Data Analysis

- EXT_SOURCE and target



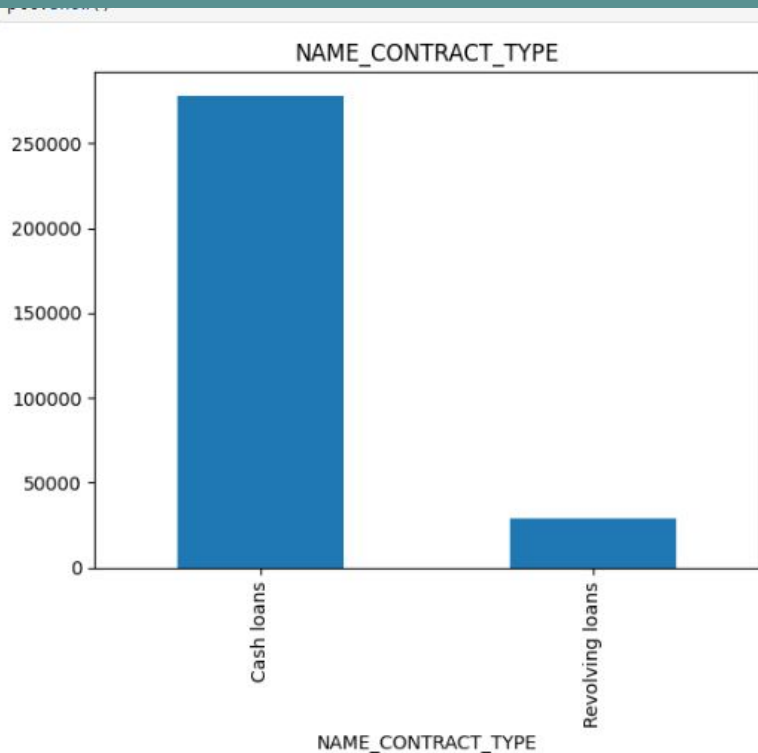
Data Analysis

Observation of male female and XNA, so we replace xna to female



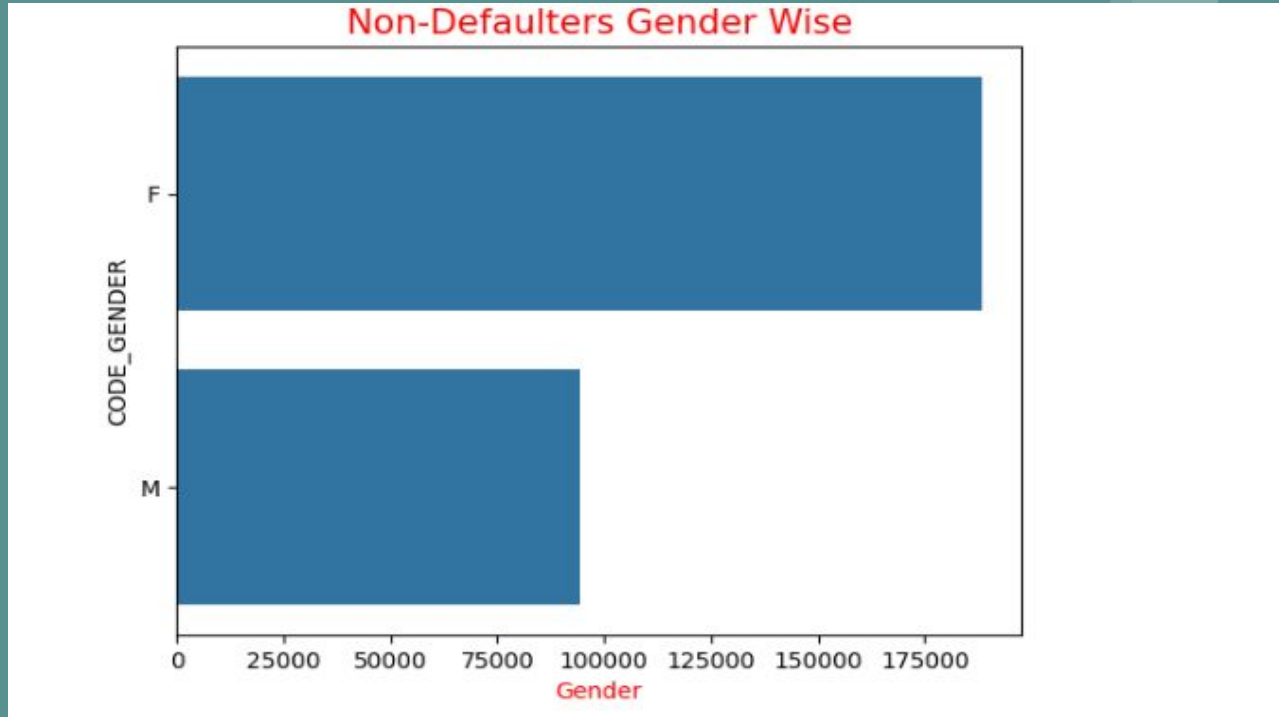
Data Analysis

Contract type cash loans and revolving loans



Data Analysis

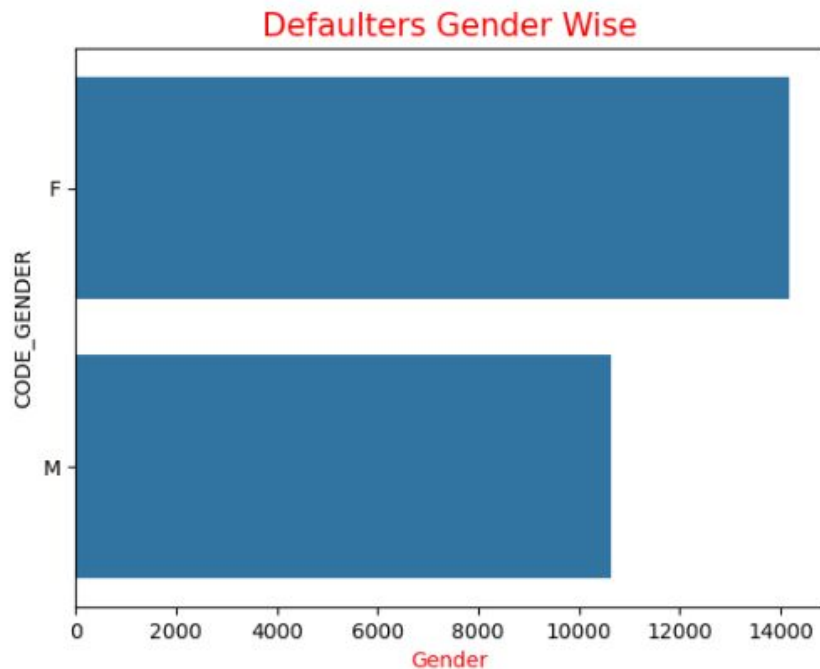
- Non defaulter Gender Wise
- It's clearly visible now that the more no of females are paying their bills correctly compare to males.
- Same goes for Defaulters as well. Females nos are more compare male here.



Data Analysis

Defaulter Gender Wise

```
pct.show()
```

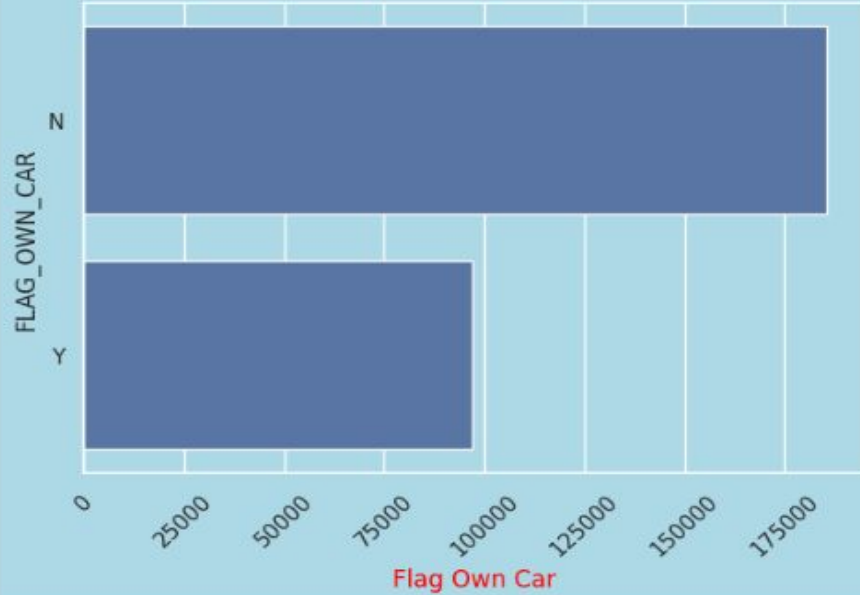


```
[1]: # Summary:
```

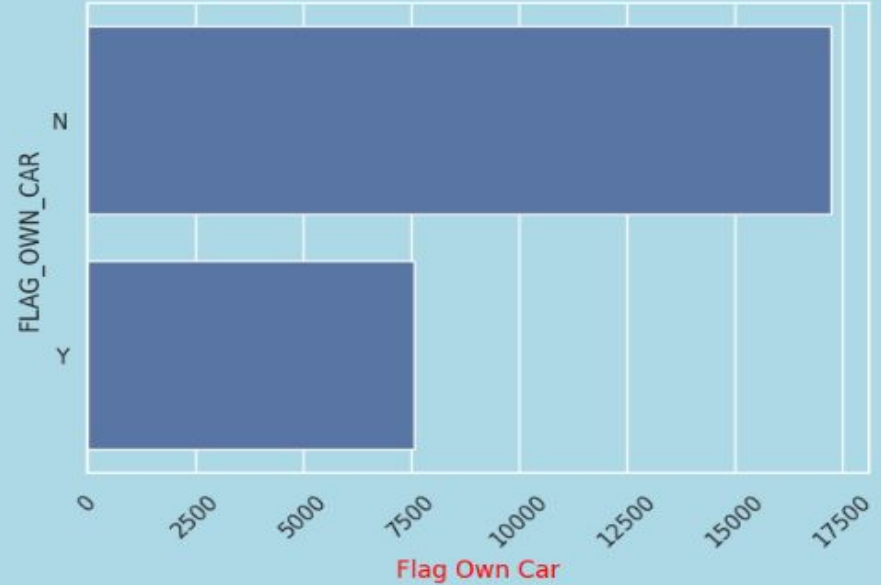
Data Analysis

Maximum no of Non-defaulters are not having their own car.
Maximum Defaulters also not having their own car.

Non-Defaulters With/Without Car

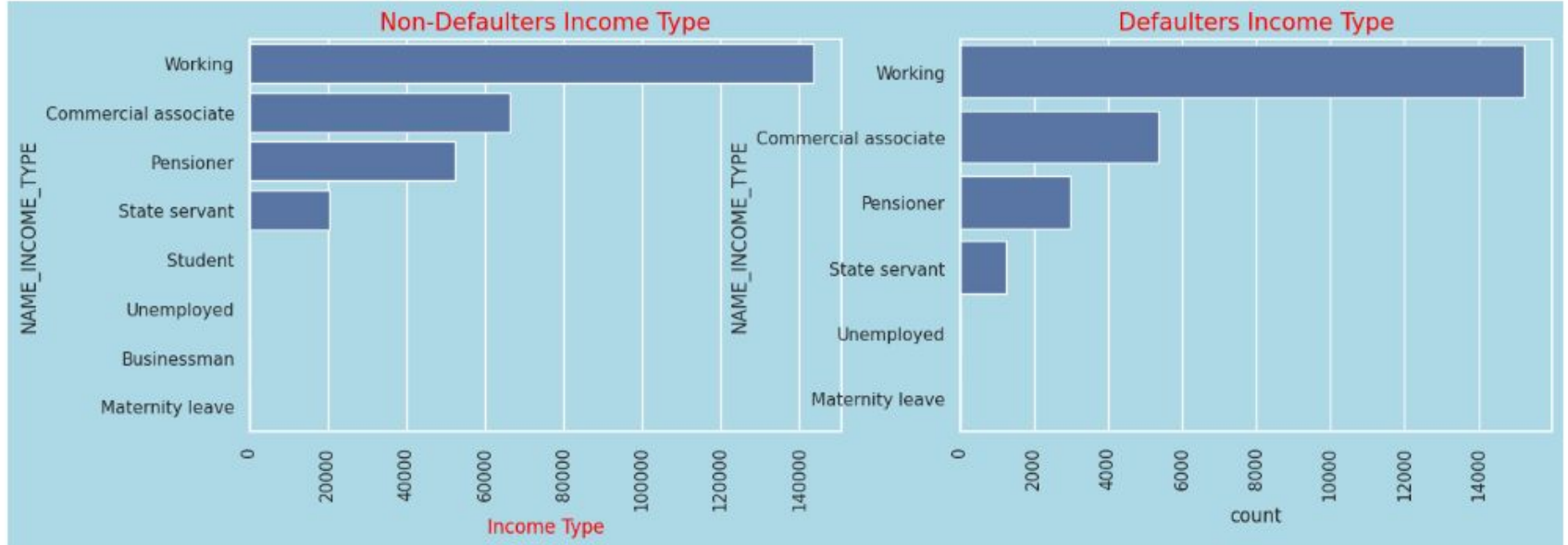


Defaulters With/Without Car



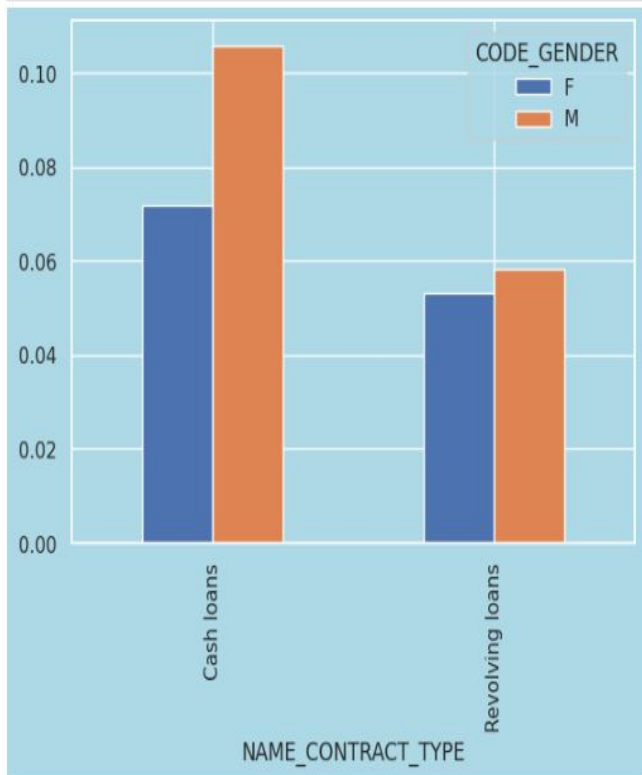
Data Analysis

The Defaulters are mostly working professionals following by commercial Associates.

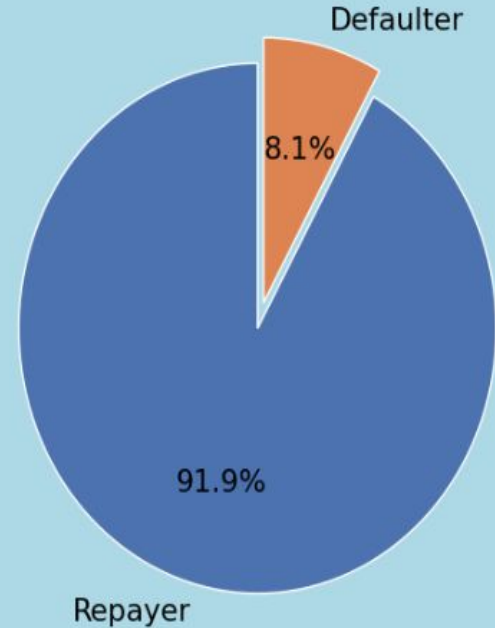


Data Analysis

Cash Loans and Revolving loans

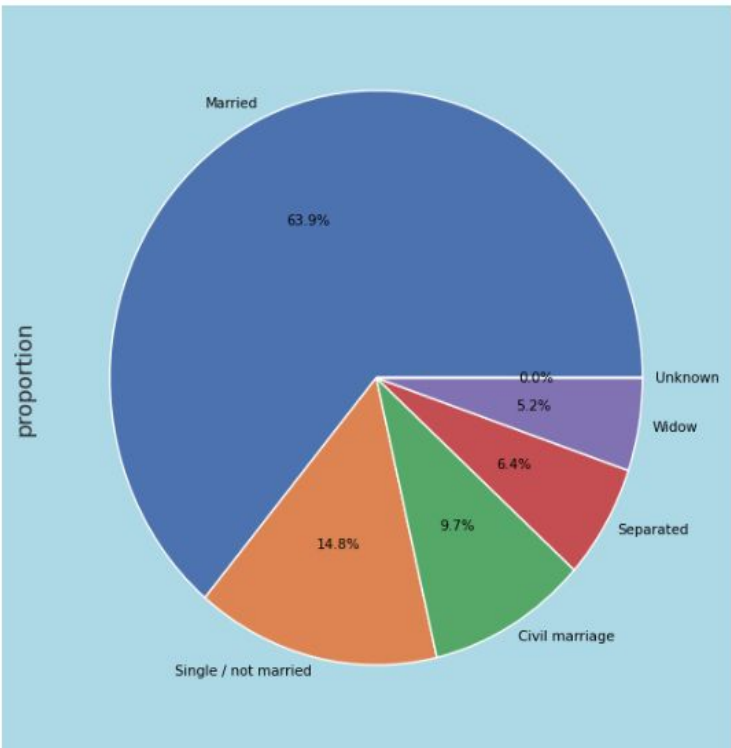


Target Variable data Imbalance



Data Analysis

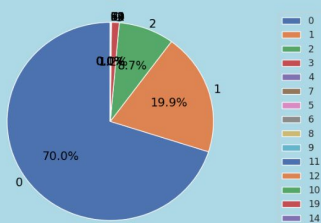
Most of people take loan is married



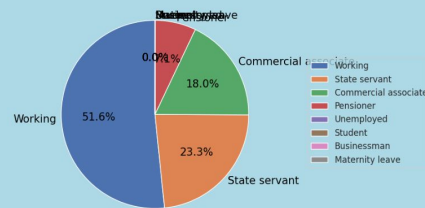
Data Analysis

Other different charts:

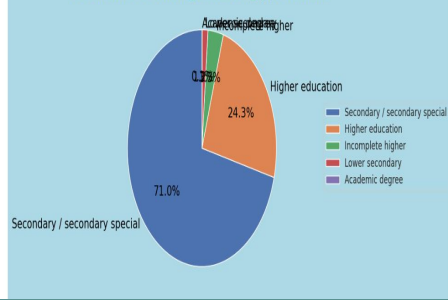
Distribution of Cnt_Children Variable



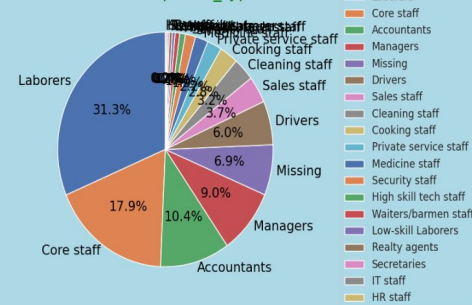
Distribution of Name_Income_Type Variable



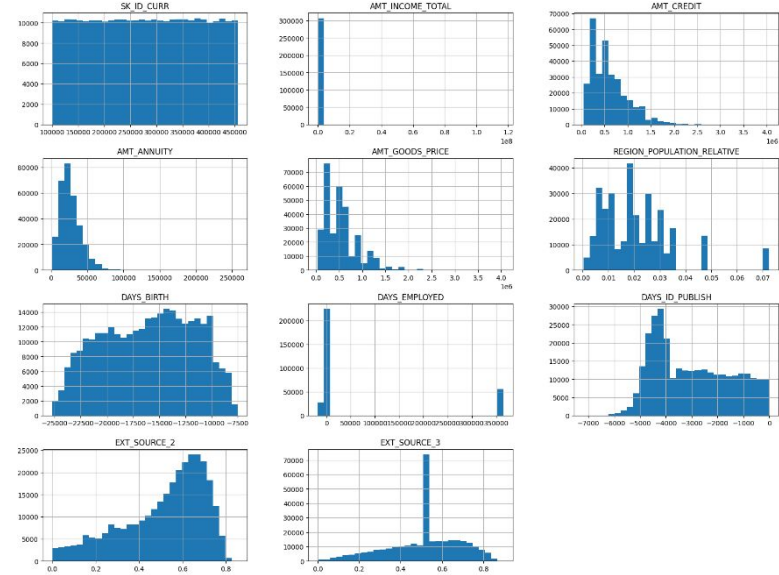
Distribution of Name_Education_Type Variable



Distribution of Occupation_Type Variable



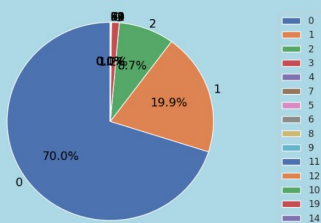
Histograms of Numerical Features



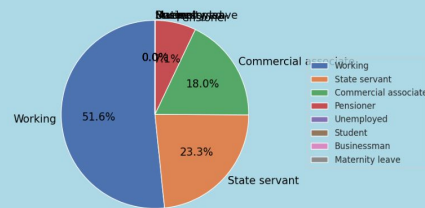
Data Analysis

Other different charts:

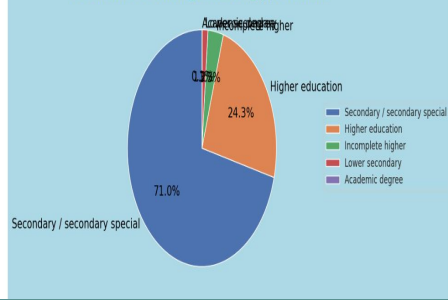
Distribution of Cnt_Children Variable



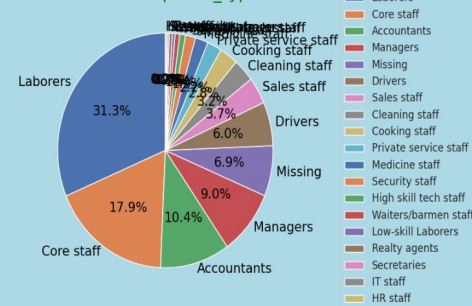
Distribution of Name_Income_Type Variable



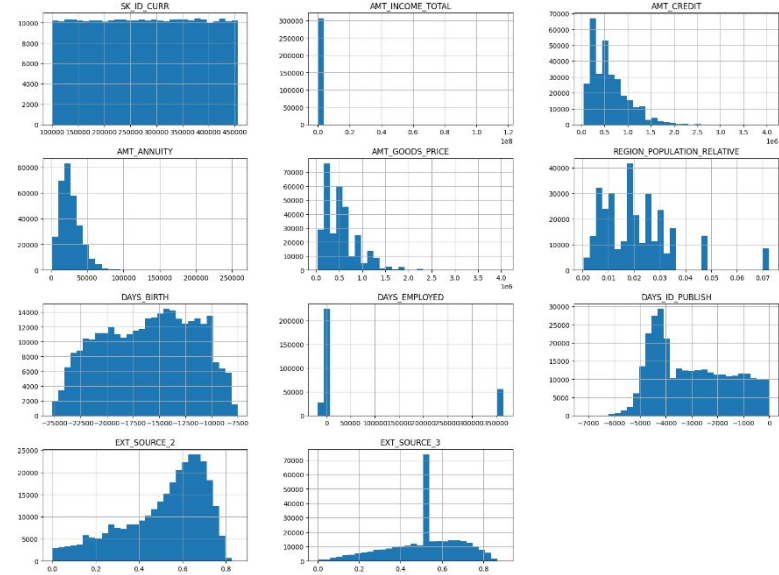
Distribution of Name_Education_Type Variable



Distribution of Occupation_Type Variable



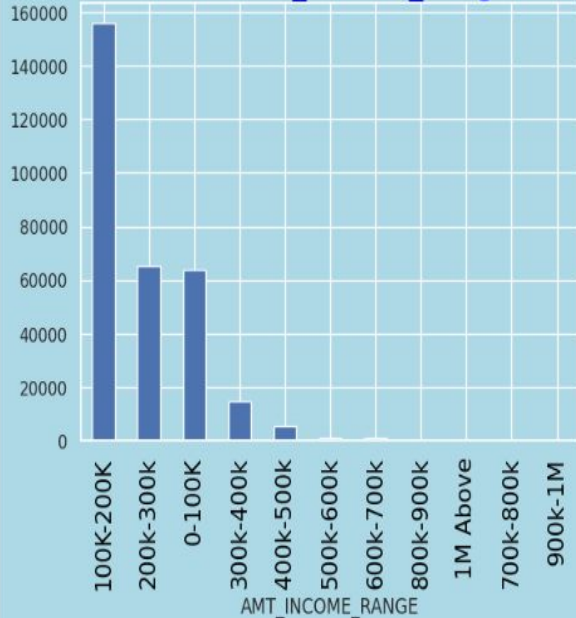
Histograms of Numerical Features



Data Analysis

Amt income range variable and heat map

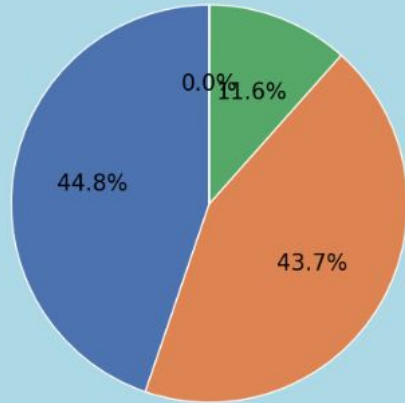
Distribution of Amt_Income_Range Variable



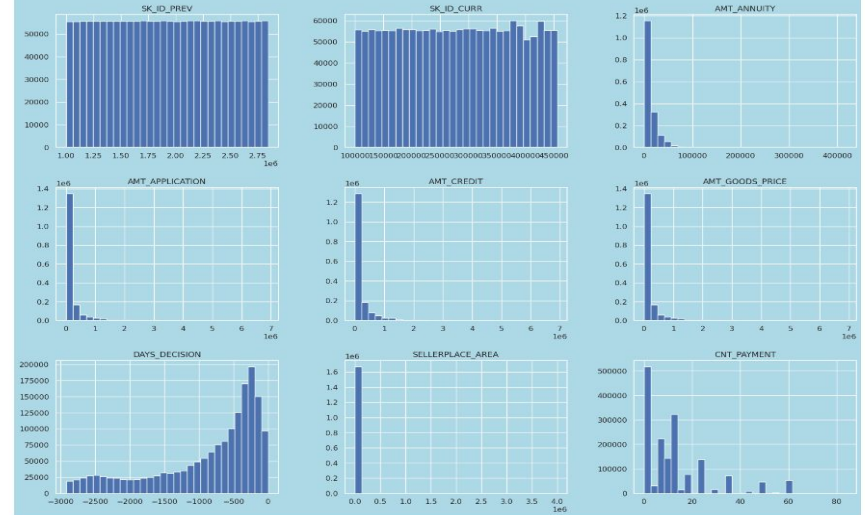
Data Analysis

Specific previous application data charta

Distribution of Name_Contract_Type Variable

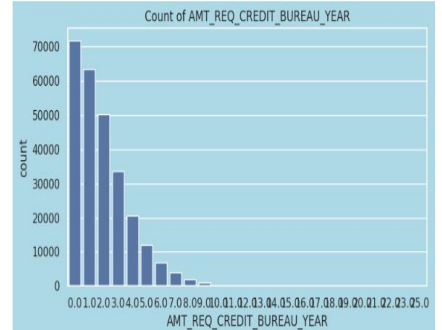
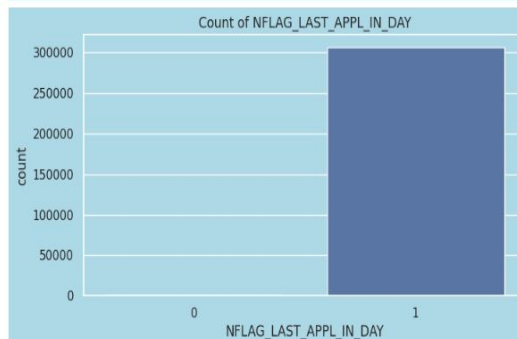
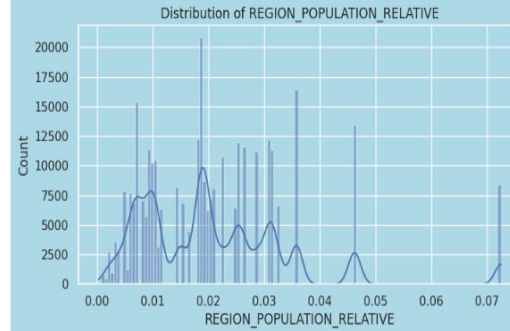
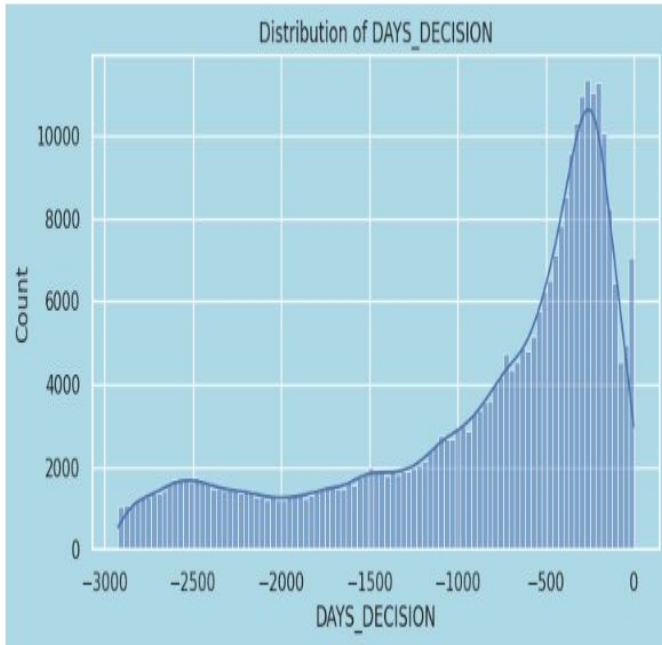


Histograms of Numerical Features



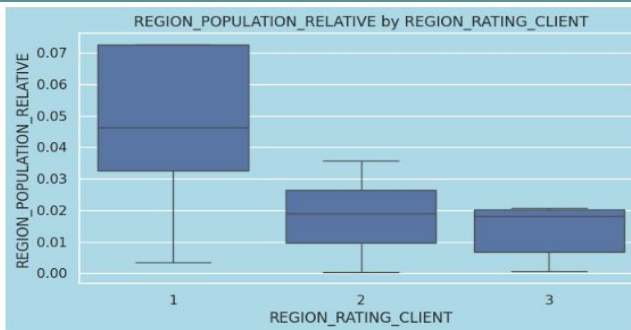
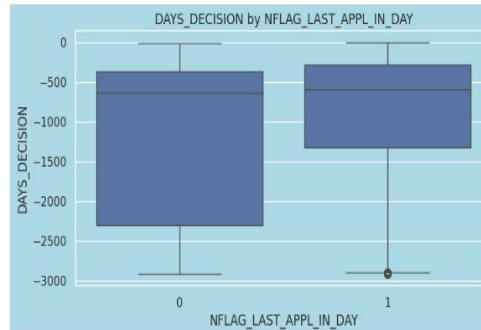
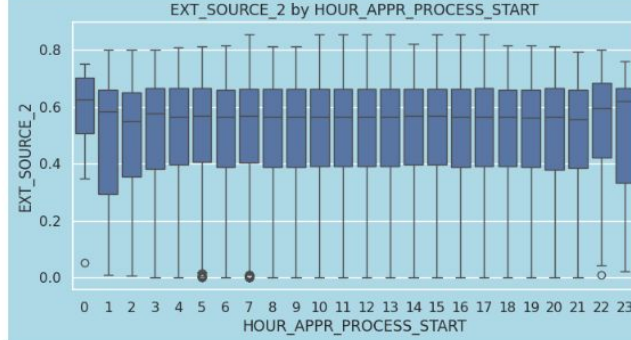
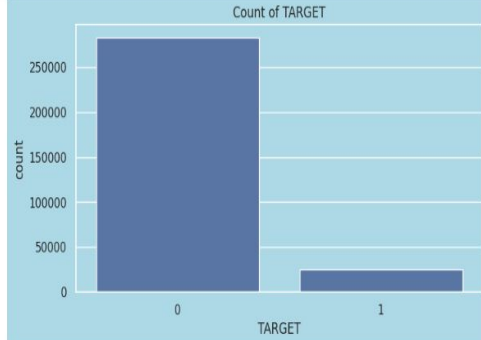
Data Analysis

After merge the previous day sheets and appdata sheets get the output



Data Analysis

After merge the previous day sheets and appdata sheets get the output



Summary :

Highly Recommended:

Previous Application Approved Clients: These clients have a history of approved loans, indicating reliability and trustworthiness.

Married Clients: Marital status often correlates with financial stability and dual income potential.

Senior Clients: Older clients typically have more established financial histories and assets.

More Educated Clients: Higher education levels often lead to better job opportunities and higher incomes.

Customers with a High Income: These clients have the means to repay loans more comfortably.

Clients with a Greater External Source: Additional sources of income or financial support reduce the risk of default.

Females: Statistically, women have shown to be reliable in repaying loans.

Customers with Strong Work Experience: Extensive work experience indicates job stability and a steady income.



Summary:

High-Risk Groups :

Clients that are Unemployed: Lack of employment makes loan repayment uncertain.

Youth Clients: Younger clients often have less financial history and stability.

Low-Income Clientele: Limited income can make it challenging to meet loan repayment obligations.

Clients with Insufficient External Sources: Lack of additional financial support increases the risk of default.

Customers with Little Work Experience: Inexperienced workers might have unstable incomes.

Customers on Maternity Leave: Temporary leave from work can impact income and repayment capacity.

Clients with a Larger Number of Family Members: More dependents can strain financial resources, making loan repayment more difficult.



Thank You

- Nayan Bharada

