"Exploring Risk in Banking and Financial Services through Data Analysis"

Pravalika Bariki
DS-C56

PROBLEM STATEMENT

- In day-to-day life many finance company are giving various types of loans to customers.
- These makes hard for the companies to give loan to the people e due to their insufficient or non-existent credit history.
- Because of that, some consumers use it to their advantage by becoming a defaulter

2 types of risk before approving the loan

1. Able to pay - not approving - company loss in business
2. Able to not repay - approving - company loss in financial

ANALYSIS APPROACH

- Given three datasets with current and previous application data of clients, along with column descriptions.
- Task involves performing **Exploratory Data Analysis** (EDA) and utilizing **data visualization** techniques.
- Objective is to identify clients with high likelihood of loan repayment and those with low likelihood.
- EDA and visualization will help discern patterns and insights for effective loan repayment prediction.

→My approach

- If data is missing, we're taking out columns with a lot of missing data. It's better not to guess for these.
- For data with just a bit missing, we're guessing the missing parts.
- For numbers, we're guessing with the middle value of the rest of the numbers (median).
- For categories, we're guessing with the most common value (mode) from the rest of the categories.

Methodology

Mainly my analysis involves 5 steps

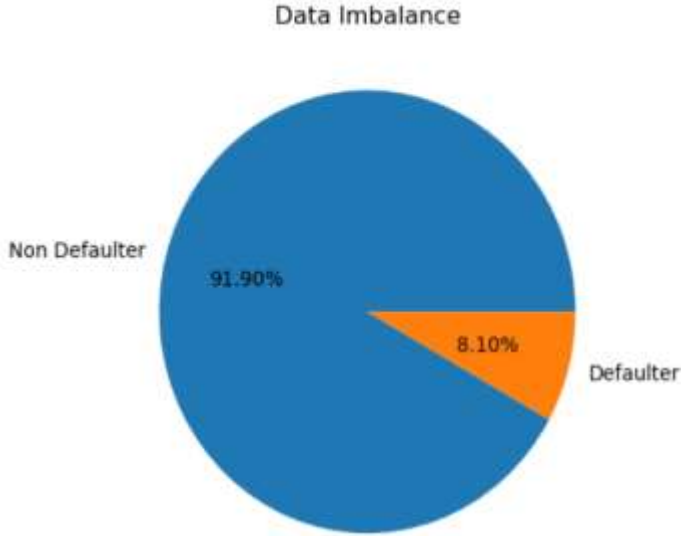
1. The first step is to **Understand the Data**, where we're given a dataset with column descriptions to understand what the data means and how important it is.
2. **Filtering** the data into the platform to see its size, types, and other details.
3. **Delete** the columns with over 30% missing data (adjusting based on total columns), and for columns with less missing data, we put the middle value for numbers and the most common value for categories.
4. After dealing with the missing data, we either use the describe method or **boxplot** to spot possible unusual values, which we later double-check to ensure the data is accurate. Where we are Analyzing the **OUTLIERS**
5. We use advanced **Visual tools** to spot important features, patterns, and connections in the data, which help us make useful predictions for the target goal, as shown in upcoming examples.

Three Types of Analysis:-

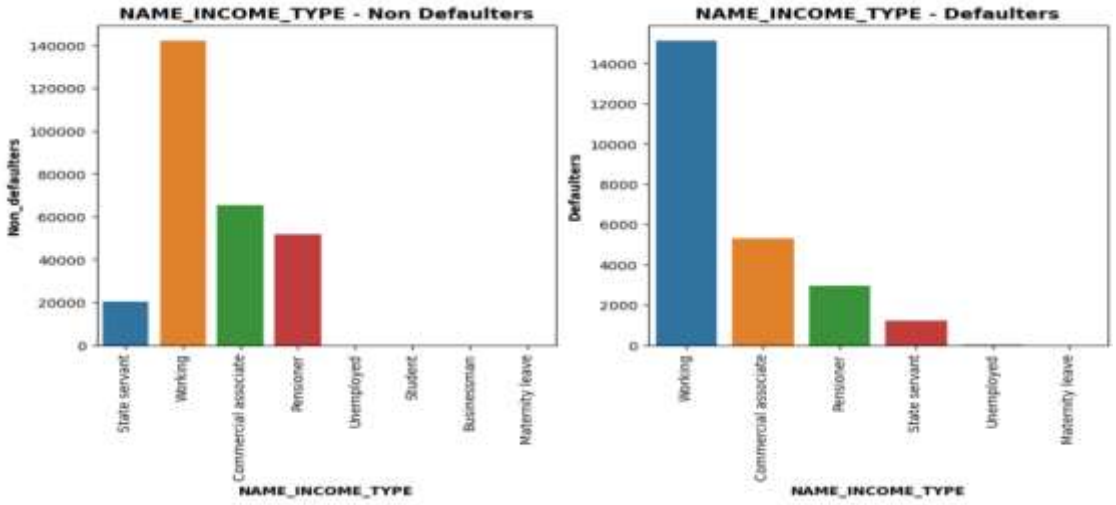
1. UNIVARIATE ANALYSIS
2. BIVARIATE ANALYSIS
3. MULTIVARIATE ANALYSIS

DATA IMBALANCE ANALYSIS

- The observed data imbalance ratio of 11.35 is indicative of a favorable situation for the bank.



UNIVARIATE ANALYSIS



- **Univariate analysis** involves examining and understanding a single variable in a dataset to gain insights about its distribution, patterns, and characteristics.
- In the second graph we can see that the working-class people are applying more for the loan.

BIVARIATE ANALYSIS

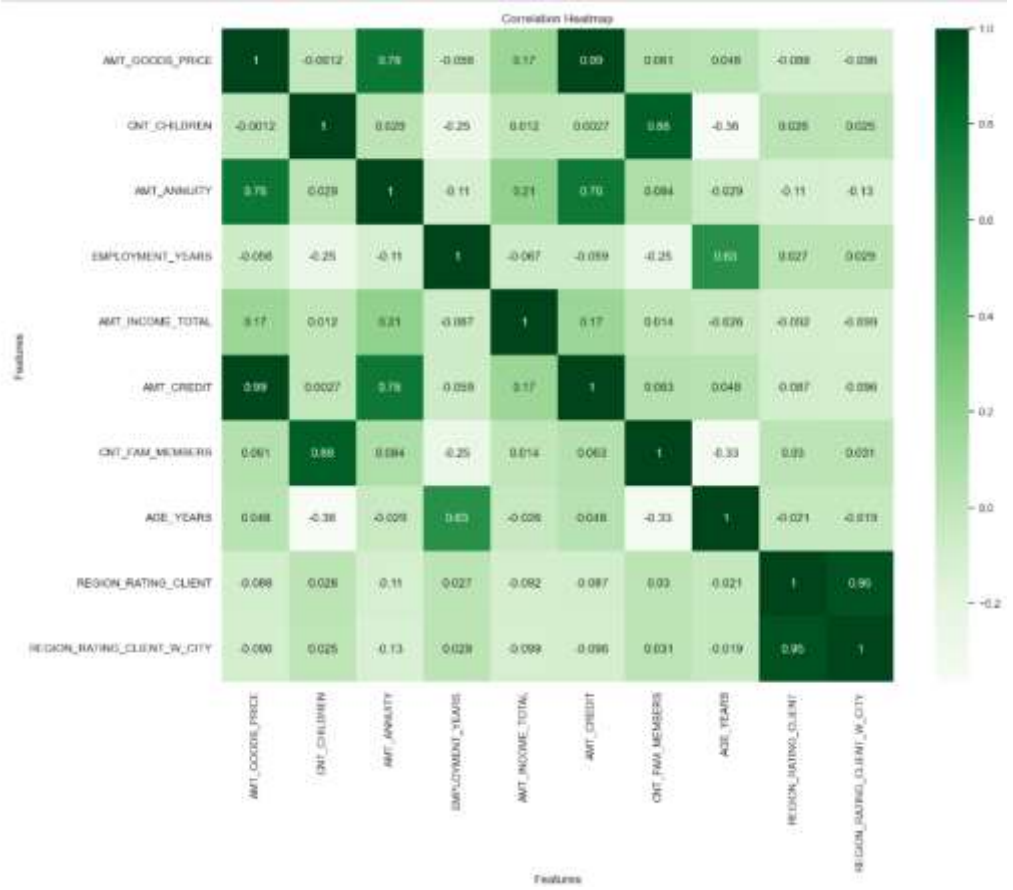
- Relationship between two different variables in a data set to uncover patterns, correlation, or associations between them.

Repeaters have higher approvals

Females have higher approvals

MULTIVARIATE ANALYSIS

- The process of analyzing and understanding the relationships between three or more variables in a dataset simultaneously, uncovering complex interactions and patterns among them is **Multivariate analysis**



A. The price of goods someone wants to buy (AMT_GOODS_PRICE) and the amount of credit they're applying for (AMT_CREDIT) have a strong connection. When one increases, the other tends to increase as well.

B. The amount of credit (AMT_CREDIT) and the annual payment (AMT_ANNUITY) are also closely linked. If someone is asking for a higher credit amount, they will usually have a higher annual payment.

CONCLUSION

- People with secondary education tend to repay loans well, so the bank can consider lending to them.
- It's a good idea for banks to provide loans to repeat borrowers because they have a positive credit history and are likely to repay.
- Students and business owners have very few cases of not repaying loans, making them suitable candidates for loans.
- Banks might want to be cautious with clients having an academic degree and moderate income, as there are more cases of loan default in this group.

THANK YOU