Initial Plots : Applicant Profile

December 9, 2023

1. Initial Setup and Function Definition:

- **Libraries Used**: Pandas, NumPy, Matplotlib, Seaborn, SciPy.
- **Visual Style**: Plots are designed to be clear and easy to understand.
- **Memory Optimization Function**: Implemented a function to save memory and improve performance when working with the data.

2. Data Loading and Preparation:

- **Data Loading**: Loaded the data for analysis.
- **Data Cleaning**: Cleaned the data by handling missing values and correcting errors. For missing values, we employed various strategies, such as replacing null values with the median for numerical features. This helps ensure a robust analysis by mitigating the impact of missing data on our conclusions.

3. Exploratory Data Analysis (EDA):

- **Descriptive Statistics**: Used simple statistics to summarize the main characteristics of the data.
- **Data Visualization**: Created easy-to-understand charts and graphs to explore patterns and trends.
- **Outlier Detection**: Identified unusual data points that may need special attention.

Applicant Profile

In this section, we explored information about loan applicants to gain insights into their characteristics. The visualizations and summaries provided valuable information:

- **Histograms of Applicant Attributes**: These charts give a visual overview of key features such as income, credit amount, age, car ownership, and employment days. This helps us understand the diversity among applicants.
- **Pairwise Plots of Key Features**: By examining relationships between income, credit amount, annuity, and age, we aim to identify patterns. For instance, we observed that higher-income applicants tend to apply for higher credit amounts.
- **Gender Distribution**: The count plot shows the distribution of applicants by gender, helping us understand the demographic breakdown. This information is crucial for detecting any gender-based patterns in the application process.

The data suggests a diverse range of applicant incomes, potential correlations between certain attributes, and insights into gender distribution. Notably, our data cleaning procedures, such as handling missing values by replacing them with medians, contribute to the reliability of our findings. These insights will guide further analysis and decision-making in the project.