Initial Plots: Loan Analysis

December 9, 2023

1. Data Aggregation for Insightful Analysis:

- Data Aggregation for Insightful Analysis: Instead of merely loading and cleaning the data, we strategically aggregated relevant information to gain deeper insights. Aggregations such as sum, max, and min were applied to specific columns, providing a more comprehensive understanding of the credit data. This process not only simplifies the analysis but also uncovers patterns and trends that may not be immediately apparent in raw data.
- Handling Missing Values and Error Correction: During the data cleaning phase, various strategies were employed to ensure a robust analysis. Null values were replaced with medians, and errors were corrected to maintain data integrity.

2. 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.

Loan Analysis

In this section, we delve into the distribution analysis of key numerical features related to loan applications, providing insights into the financial behaviors of applicants:

- Loan Amount Distribution: The histogram reveals that most loans are concentrated in the lower to mid-range of the credit spectrum, with a steep drop-off as the credit amounts increase. This suggests that higher loan amounts are less common, representing a smaller, more targeted segment of the applicant pool.
- Loan Amount by Contract Status: The bar plot illustrates the average loan amount by contract status, showing that approved loans have a lower average amount compared to unused offers. This observation implies that while applicants may be approved for lower amounts, they are offered higher amounts that they ultimately do not use.
- Total Payment by Loan Type: The bar plot displays the total payment made by loan type, indicating that cash loans have the highest total payments followed by consumer loans, with revolving loans having the least. This reflects the differing natures of the loan types, with cash loans typically involving larger amounts and consumer loans being more common but involving smaller amounts.

• Outlier Detection: Applying the z-score for detecting outliers in loan amounts reveals the presence of outliers in the dataset. The table of outliers indicates that some loans have significantly higher amounts than typical, which could be crucial for risk analysis and credit policy adjustments.

• Insights and Conclusions:

- The diverse financial profiles of applicants, with most seeking smaller loan amounts, suggest a broad range of economic backgrounds and needs within the customer base.
- The concentration of loans in the lower to mid-range indicates a market preference for these credit levels. Understanding this trend is vital for product development and marketing strategies.
- The discrepancy in average loan amounts between approved and unused offers could be attributed to varying financial needs or preferences among applicants. Further investigation into this behavior may offer opportunities for product customization.
- The dominance of cash loans in total payments reflects the larger sums associated with this loan type. This information can guide the allocation of resources and marketing efforts.
- The identification of outliers can lead to enhanced credit risk models, allowing the institution to identify and manage potentially high-risk applications more effectively.
- These distribution patterns and payment behaviors across different loan types provide valuable insights for refining existing products and developing new ones tailored to customer preferences and market demands.

The strategic data aggregation process has not only enabled a more insightful exploration of the loan data but has also provided actionable insights for decision-making and strategic planning.