

Initial Plots : Financial Variables Distribution

December 9, 2023

1. Data Aggregation for Insightful Analysis:

- **Data Aggregation for Insightful Analysis:** The sheer volume of data posed challenges in processing time and computational resources. To address this, 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. Truncating tables played a crucial role in mitigating the computational burden, allowing for more efficient analysis.
- **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.

Financial Variables Distribution

In this section, we analyze the distribution of various financial variables associated with the bank's credit and loan data, providing insights into the bank's loan operations:

- **Credit Installments and Balances:** The distributions of `months_balance`, `cnt_installment`, `cnt_installment_future`, and `amt_balance` show the frequency of different counts and amounts for point of sale (POS) cash loans and credit card usage. These histograms indicate varied engagement levels with the bank's products, reflecting the diversity in customers' credit behavior and the bank's installment-based financial products.
- **Future Installment Distribution in POS Cash Loans:** The distribution plot for the count of future installments in POS cash loans suggests a skew toward fewer future installments, with a majority of loans having a lower count of installments remaining. This could imply that most of these loans are close to completion or are structured with fewer installments over a shorter period.
- **Average Balance by Contract Status in Credit Cards:** The bar plot shows the average balance held on credit cards by contract status, providing insights into the typical credit card debt customers carry in each category. Notably, 'Completed' contracts have the highest average balances, potentially indicating that once customers fulfill their contractual obligations, they tend to maintain higher balances.

- **Total POS Cash Loan Count by Contract Status:** The bar plot illustrates the total count of POS cash loans by contract status, with 'Active' loans dominating the count. This dominance of active loans indicates a robust engagement with the bank's POS loan products.
- **Outlier Detection in Credit Card Balances:** The table presents outliers identified using the z-score method, which are credit card balances that deviate significantly from the mean. The presence of these outliers, with balances far exceeding the credit limit, could be indicative of atypical customer behavior or potential issues in credit card usage that warrant further investigation.
- **Detailed Analysis and Report Summary:**
 - The analysis provides a comprehensive look at the bank's credit and loan product usage, focusing on payment behaviors and outstanding balances.
 - The skew in the distribution of future installments could imply a loan structure favoring shorter-term loans or a customer preference for quicker repayment.
 - The average credit card balance by contract status provides valuable insights into how customers manage their credit card debt throughout the lifecycle of their contracts.
 - The total count of POS cash loans underscores the popularity and active engagement with this loan product.
 - The identification of outliers in credit card balances points to exceptional cases that may require further risk assessment and possibly intervention by the bank.

The strategic data aggregation process, coupled with table truncation to address computational challenges, has not only enabled a more insightful exploration of the financial variables but has also provided actionable insights for decision-making and strategic planning.