Atm Transactions

Mithila Arman

DataSet Description

This dataset contains 30,156 records with 12 mixed-type features, including six numerical and six categorical. It covers various timestamps.

Statistical Summary

- Total Balance Range: from 19,900 to 57,050 with a mean of 32,371.
- Transactions:
    - Income Transactions: Vary from 0 to 100 with an average of about 48.
    - Outcome Transactions: Range from 0 to 100 with a mean of approximately 55.
- Transaction Amounts:
    - Total Income: Varies from 0 to 8,440 with a mean of 3,089.
    - Total Outcome: Ranges from 0 to 9,710 with an average of 3,534.
- Transaction Days: Data for all seven days of the week, with Wednesday having the highest frequency of recorded transactions.

Detailed Atm Transactions Data Analysis and Insights:

1. Data Cleaning and Exploration
    - Initial Setup: Libraries, matplotlib.pyplot and pandas are imported for data manipulation and visualization.
    - Data Loading: Data is loaded from a CSV file and the initial structure columns and data types are checked using data_new.info().
    - Data Review: Displays the DataFrame for an initial look at the data.
    - Handling Missing Values and Data Types: It checks for missing values and confirms the data types of each column.
    - Duplicate Records: Duplicate entries are removed to ensure the quality of analysis.
2. Distribution Analysis
    - Visualization: Used seaborn for plotting a histogram of the totalBalance to understand the distribution of balances across ATMs, overlaying a kernel density estimate (KDE) for a smoother distribution curve.
3. Correlation Analysis
    - Numeric Focus: Selects numeric data types for correlation analysis to explore relationships between variables.

    - Correlation Matrix and Heatmap: Calculates and displays a correlation matrix using a heatmap to visualize how different numerical attributes correlate with each other.
4. Descriptive Statistics
    - Time Conversion: Converted transactionTime from string to datetime format.
    - Summary Statistics: Generated descriptive statistics of the data to provide an overview of central tendencies, dispersion, and shape of the dataset's distribution.
5. Transaction Volume Analysis
    - Time Features Extraction: Extracted hour, date, day of the week, and week number from transactionTime for more detailed temporal analysis.
    - Aggregation by Time Units: Grouped data by hour, day, and week to analyze transaction volumes over different time frames.
    - Visualizations: Created bar and line charts to visually represent transaction volumes by hour, day, and week, which could help in identifying patterns like peak transaction times or busy days.
6. ATM Performance Analysis
    - Performance by ATM: Aggregated transaction data by ATM ID to summarize performance metrics like total transactions, income, and outcomes for each machine.
    - Summary Output: Presented a summary table of aggregated values which can be crucial for identifying high or low performing ATMs.