EDA_report_22000245_Moonbyeori

April 7, 2024

1 Exploratory Data Analysis

1.0.1 - StudentID: 22000245

1.0.2 - Name: Moon ByeoRi

1.0.3 - 1st Major: ACE

1.0.4 - 2nd Major: Data Science

1. Data overview Data review -BT_data 5000000 obs, 5 variable Date, Description, De[psots, Withdrawls, Balance

Date 2020-Aug-21 ~ 2155-Nov-15.

Description type is char, each ATM, Bill, Cash, Cheque, Commission, Debit Card, IMPS, Interest, Miscellaneous, NEFT, Purchase, Reversal, RTGS, Tax, Transfer.

Deposits type is char, but convert to numeric data and use it.

Withdrawls also char type, necessary convert to nymeric data.

Balance type is char, It is Results after the transaction ends.

2. Univariate analysis

I started the analysis first with the date variable.

1.0.5 2.1 Variable 1: Date

~~ First, we examine the first variable, Date, from the BT_Data. The Date variable is of char type and is formatted like 21-Aug-2020. To check the volume of transactions on a daily basis, as well as monthly and yearly, we separated the Date variable.

After that, we calculated the average daily trading volume for each month and visualized each to make them easy to see. # 2020 Annual Graph Image

The monthly average trading volume for the year 2020 was visualized in a graph. August appears slightly higher due to fewer samples compared to other months, while the volumes for the other months are almost similar.

2 2021 Annual Graph Image

This is the graph image for the year 2021. It's not sorted by month, which makes it a bit harder to read. Notably, there was a sharp increase in trading volume in June compared to other months, and again in December, the trading volume showed a significant spike in average volume.

3 2020~ 2024 Annual Graph Image

We visualized the changes in trading volume for five years, from 2020 to 2024, in a line graph to observe the trading volume of all 12 months at once. The most noticeable feature is the graph for 2024, which shows very rapid changes. Additionally, significant fluctuations are observed once in other years as well.

The variance between the graphs is quite severe, and since we only checked the frequency of trades, it seems difficult to gain meaningful insights from this alone.

3. Multivariate analysis

We were unable to obtain meaningful results with just the daily transaction frequency. This time, we will use another variable, "Description," to explore the frequency of each transaction method, the annual monthly average, and so on.

4 ATM Transaction Frequency Image

This time, in addition to the date, we will also use the description variable for a more diversified analysis. We have gathered the daily average ATM usage frequency by month for the entire year. On average, the daily number of ATM uses exceeds 200, but only February is observed to be slightly lower than that.

5 Monthly Average Transactions of the Entire Description Image

Feeling it was insufficient to look at just one transaction method, we filled the graph for all descriptions. Overall, the results were very uniform, with only February noticeably lower. Apart from that peculiarity, there are no other significant outliers.

6 ATM Deposit/Withdrawal Image

Then, let's check the frequency of deposits and withdrawals. We visualized how much the deposit and withdrawal records differ within the day's transaction volume. In December, the most active trading month, deposits and withdrawals occurred at similar frequencies. For most months, as the average daily transaction volume increases, both deposits and withdrawals increase proportionally, but overall, the frequency of withdrawals is higher than that of deposits.

4. Suggestion

Upon summarizing the above graph, we observe that a single transaction method, for example, financial transactions using an ATM, occurs about 6 times a day on average. When checking all

transaction methods, on average, nearly 100 transactions occur. Moreover, among these, each transaction method averages around 200 transactions per month, and it was difficult to find significant differences among the financial transaction methods.

Within this, deposit and withdrawal records show slightly more deposits than withdrawals. It's challenging to identify cases where transactions of a specific method significantly increase on certain months or days. Instead of conducting promotions for specific transaction methods, it might be more beneficial to offer fee discounts during December to January when average transactions are higher, to attract more users.