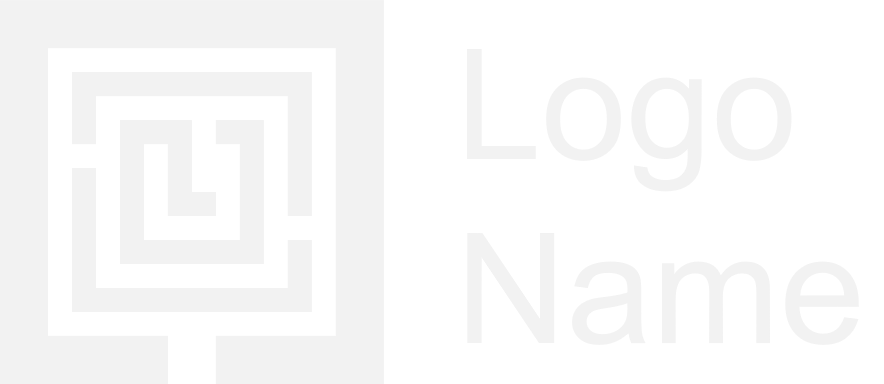


|  |
| --- |
| Real-Life Fraud Detection Analysis |
|  |
| May 3  Advance Programming Skills  Authored by: Hebatallah Saleh AL-jabri |



|  |
| --- |
| Introduction: Fraud detection is very important for banks and financial companies. It helps protect money and build trust with customers. Today, many people use mobile money apps to send and receive money quickly. Because these transactions happened fast and in large numbers, it can be hard to catch fraud in time. If fraud is not committed, it can cause big financial losses and hurt the company’s reputation.  Fraud can include fake transfers, stolen accounts, or moving money in illegal ways. If these actions are not stopped, the company could lose a lot of money and customers might stop using their services. To stop this, banks use data analysis to find patterns that look suspicious.  By looking at the data from past transactions, we can find out what fraud looks like. We can check how much money was sent, what kind of transaction ii was, and which accounts were used. This helps us understand what is normal and what might be fraud. With the help of data and some simple tools, we make better decisions and improve how fraud is caught. **What the Dataset Can Tell Us:** In this assignment, we are studying a dataset of mobile money transactions. The data includes information such as how much money was sent, type of transaction, and the account details before and after the transfer. This helps us see if anything unusual happened during a transaction.  By exploring this data, the aim to find patterns That can help us understand fraud better. For example, we can look at when fraud usually happens, what transaction types are most affected, and how the amount of money relates to fraud. We will also give some useful suggestions based on what we find, so that the company can improve it fraud detection system and reduce risk in the future. Data Preparation: Getting the Data Ready Before we can find fraud, we need to clean and understand the data. The dataset has mobile money transactions, where each row is one transaction. The columns show details like time, amount, type of transaction, sender and receiver IDs, and account balance before and after.   **Here is what each column means:**  * step – The time of the transaction (like hour or day). * type – The type of transaction (like transfer, cash out, or payment). * amount – The amount of money sent in the transaction. * nameOrig – The ID of the person who sent the money. * oldbalanceOrg – The sender’s balance before the transaction. * newbalanceOrig – The sender’s balance after the transaction. * nameDest – The ID of the person who received the money. * oldbalanceDest – The receiver’s balance before the transaction. * newbalanceDest – The receiver’s balance after the transaction. * isFraud – This is 1 if the transaction was fraud, or 0 if it was normal. * isFlaggedFraud – This is 1 if the system flagged it as a possible fraud (for example, if the amount was very high), or 0 if not.   We check for missing values, fix errors, and remove columns that aren’t useful. After cleaning, the data is ready for analysis t help us spot fraud patterns. **Removing Less Useful Balance Columns to Simplify Analysis:**  In the dataset, some columns are not very useful for detecting fraud. One example is:   * newbalanceOrig and newbalanceDest – These might seem useful, but often the fraud patterns can already be seen using the original balance and amount. In some cases, I might drop one if it adds no new information or is hard to trust.   Why Remove These?  Removing such columns:   * Makes the dataset smaller and easier to work with * Reduces noise (unhelpful data that confuses the model) * Focuses only on the features that truly affect fraud detection, like amount, type, step, and oldbalanceOrg.    Phase 3: Real-Life Fraud Detection Analysis**Critical Thinking Task:** Sometimes, a normal transaction might look like fraud just because the amount is very high, or it happens at a strong time. For example, a business might send a large payment late at night, which cloud make the system think it’s fraud. To avoid these mistakes, we can look at each customer’s past behaviour and build a profile of what’s normal for them. This helps the system tell the difference between real fraud and regular activity more accurately. Phase 4: Insights and Recommendations**What we Found** After looking at the data using Python, we found some clear signs of fraud:   * Most fraud happens during TRANSFER and CASH\_OUT transactions. These are the types used to send money to others or take money out, and fraudsters seem to use them the most. * Big amounts of money are more common in fraud cases. When we marked transactions over 200,000, many of them were either fraud or marked as suspicious. This tells us that large transfers should be watched more closely. * Some users (senders or receivers) are involved in many fraud cases. These could be fake accounts, stolen accounts, or accounts used again and again for illegal activities. * Fraud tends to happen at certain times. This means that fraud may happen in planned groups or at times when fewer security checks are working.   These patterns can help us focus on the riskiest areas and make better rules to stop fraud. **What I suggest:**  Based on what I found, here are some things that can help reduce fraud:   * Check risky transaction types like TRANSFER and CASH\_OUT more carefully. Use extra checks or limits on these, especially for large amounts. * Flag high-value transactions (like over 200,000) for review. These are more likely to be fraud. * Watch repeats users involved in fraud. If someone appears in fraud cases many times (either sending or receiving), they should be flagged and reviewed. * Use customer behaviour patterns to find fraud. If a customer usually sends small payments but suddenly sends a large amount at a strange time, it could be fraud. * Focus on common fraud times. Add more checks during hours when fraud happens most often, based on the step values.   By using these ideas, companies can catch fraud faster, protect customers, and reduce losses. Visualization: *This graph highlights the relationship between overall transaction activity and fraudulent transactions, showing how fraud correlates with the total transaction volume at each time step. It simplifies the analysis by focusing only on the total and fraudulent transaction counts.*  *The histogram displays the distribution of the receiver's new balance after a transaction. The x-axis represents the new balance amounts, while the y-axis shows the frequency or count of occurrences for each balance range.This visualization helps to understand how receiver balances are spread out across transactions, highlighting trends and potential outliers.*  *This pie chart illustrates the distribution of different transaction types in the dataset. Each segment represents a specific transaction type, such as transfer, cash out, or payment, with the chart showing the percentage share of each type. This chart helps to quickly understand the relative frequency of each transaction type within the dataset.* |

x



*This box plot shows the distribution of transaction amounts for fraudulent and non-fraudulent transactions. The x-axis represents fraud status, and the y-axis shows the transaction amounts on a logarithmic scale, helping to highlight differences and outliers between the two groups.*