**Term Project: Fraudulent Activities in Bank Transactions**

Sreenivasulu Somu

Department of Data Science, Bellevue University

DSC530-T303 Data Exploration and Analysis (2251-1)

Mr. Matthew Metzger

November 16, 2024

# Term Project: Fraudulent Activities in Bank Transactions

# Statistical/Hypothetical Question:

The primary question addressed in this analysis is, Are there significant patterns or indicators in the transaction characteristics that could help detect potentially fraudulent activities in the bank transactions ?

# Outcome of EDA:

* **Transaction Amount:** The distribution of transaction amounts is right-skewed, with most transactions being of lower value. There are outliers with very high transaction amounts, which could potentially indicate the fraudulent activities or legitimate large transactions.
* **Customer Age:** The age distribution of customers appeared to be roughly normal, with most customers between 20 to 60 years old, indicating a diverse customer base across different age groups.
* **Transaction Duration:** Most transactions took less than 200 seconds, but there are some outliers with very long durations. These could represent complex transactions or potential technical issues.
* **Login Attempts:** Most of the transactions have only one login attempt, but there are cases with multiple attempts. This could be a potential indicator of fraudulent activities.
* **Account Balance:** The distribution of account balances is right-skewed, indicating a wide range of customer wealth levels.

# What was missed during the analysis:

* **Time-based patterns:** The analysis could have benefited from exploring time-based patterns, such as transaction frequencies at different times of day or days of the week.
* **Geographic analysis:** While the location data was available, the analysis did not search into the geographical patterns of transactions, which might reveal important insights.
* **Device and IP analysis:** More in-depth analysis of Device IDs (mobile, laptops, etc.) and IP addresses could have uncovered patterns related to fraudulent activities.

# Variables that could have helped in the analysis:

* **Transaction history:** Information about a customer's previous transaction patterns could have
* **Card present/not present:** Information about whether the transaction was made in person or online could be valuable for detecting a fraud transaction.

# Potentially incorrect assumptions:

* **Outlier handling:** The analysis assumed that outliers should be kept in the dataset, but some of the extreme values might be due to data errors rather than the actual transactions.
* **Linear relationships:** The use of Pearson's correlation assumes linear relationships between variables, which may not be the case always.

# Challenges faced and areas of uncertainty:

* **Defining fraud:** Without labeled data indicating which transactions were fraudulent, it was a challenge to identify fraud indicators.
* **Balancing privacy and analysis:** Working with sensitive financial data raises questions about how to balance thorough analysis with the customer privacy concerns.
* **Interpreting complex interactions:** Understanding the interplay between multiple variables in determining fraud risk was challenging and may require more advanced modeling techniques.