Faleti Olusola Aanuoluwapo
STC Data Analyst Project – Week 2 Report

INTRODUCTION

During Week 2, This report documents the data cleaning steps applied to the Bank of America consumer complaint dataset to ensure accuracy, and consistency for analysis.

1. Tool Used:

Power BI

For the data cleaning process, I imported the dataset from Excel into Power Query Editor in Power BI. Power Query is a built-in tool for data preparation and transformation. It allowed me to clean, organize, and structure the dataset to ensuring it was ready for accurate analysis

2. What issues did you find in the dataset?

During the data cleaning session, I identified the following issues

- Inconsistent column names.

- Missing values in column.

- Blank cells in multiple columns

3. What specific steps did you take to clean the data?

Although the dataset wasn't very messy, the following steps were taken to clean the data.

- **Checked column headers** in Power Query to ensure they aligned correctly with the data in their respective fields.

- **Checked for duplicates** using *Remove Duplicates* based on Complaint_ID in Power Query but no duplicate records were found.

- **Dealing with missing values:** Replaced Columns such as

- Sub-issue with nulls as not specified,

- company public response with pending the reason for this decision is after using filter for the own data I notice that all the null is links to the company response to customer which is in progress

- Timely response with pending the reason for this decision is after using filter for the own data I notice that all the null is links to the company response to customer which is in progress

- Sub-products that accounted for less than 1% were replaced with "Checking Account." This decision was made because, after filtering the column, I noticed that all values representing less than 1% were linked to savings or checking accounts — and all sub-products under that category were "Checking Account.

4. What changes did you make to column names or structure? Mention any renaming or formatting decisions.

Renamed columns for clarity and standardization

Submitted via was changed to Complaint_media using the *Transform - Rename Columns* feature.

5. **Is the dataset now clean and ready for analysis?**
   Yes.

   The dataset has been standardized, missing values have been handled appropriately, and column headers are now consistent and descriptive.

6. **Additional Notes (Optional):**

Were there any assumptions you had to make?

Yes
Sub-products that accounted for less than 1% were replaced with "Checking Account." This decision was made because, after filtering the column, I noticed that all values representing less than 1% were linked to savings or checking accounts — and all sub-products under that category were "Checking Account.