# EXCEL DATA CLEANING BASED PROJECT

Jimoh, Sekinat Eniola

March 6, 2025

# **Tables of Content**

- 1. Introduction
- 2. Data Cleaning Process
  - Initial Assessment
  - Handling Duplicates and Missing Values
  - o Column-Specific Cleaning
    - Date Column
    - Vehicle Registration Number
    - Make/Model Column
    - Item Description
    - Assigned to Column
    - Service KM
    - Workshop Name
    - Invoice Number
    - Amount (NGN)
    - Remarks
  - o Final Verification Using Power Query
- 3. Conclusion
- 4. Pictures
- 5. Link to the Spreadsheet

#### INTRODUCTION

The Project involved cleaning a structured dataset using Microsoft Excel to improve its accuracy and usability. The dataset contained 3,833 records with various columns, including Date, Vehicle Registration Number, Make/Model, Item Description, Assigned to, Service KM, Workshop Name, Invoice Number, Amount (NGN), and Remarks. The primary objective was to identify and resolve inconsistencies, missing values, and redundant information to ensure a well-structured dataset ready for further analysis and decision-making.

## **Data Cleaning Process**

#### **4** Initial Assessment

- The dataset contained 3,833 records with various structured fields.
- An image was embedded in the file, which was removed.
- Unnecessary and empty rows were deleted, including section titles and project details, ensuring the dataset started with actual column headers. A total of 7 empty rows were removed.

### **4** Handling Duplicates and Missing Values

- The dataset was reduced to 3,830 records after removing 6 duplicate entries using the **Remove Duplicates** feature in the **Data** tab.
- Missing values in numerical columns were replaced with appropriate values (e.g., 0 in Service KM where data was absent).

## **♣** Column-Specific Cleaning

#### a. Date Column

- The original date format displayed as numbers (e.g., "42064").
- Converted to proper date format using the **Format Cells** option (e.g., "01-03-15").
- Identified and manually corrected two inconsistencies in date formatting.

#### b. Vehicle Registration Number

- Inconsistencies such as values enclosed in brackets were identified.
- Used the **Find and Replace** function to remove unnecessary brackets, affecting 67 values.

#### c. Make/Model Column

- The column was split using the **Text to Columns** function, resulting in four columns.
- The Make column remained unchanged, while Columns 2, 3, and 4 contained fragmented model names.
- Used the **Concatenate** function to merge them back into a single column.
- Applied **Find and Replace** to remove brackets and correct typographical errors.

#### d. Item Description

- Standardized the item descriptions into five categories using **IF**, **OR**, **ISNUMBER**, and **SEARCH** functions:
  - Maintenance
  - > Body Repair
  - ➤ Electrical Repairs
  - > Tyres and Wheels
  - > Other Repairs
- Example in using the function "= (IF (OR (ISNUMBER (SEARCH ("OIL", E2)), ISNUMBER (SEARCH ("FILTER", E2)), ISNUMBER (SEARCH ("BELT", E2))), "MAINTENANCE", ..." Simply meaning if either oil, filter or plugs are found in column E2, group it as maintenance.
- Replaced the original description column with a newly created **Item Category** column.

#### e. Assigned to Column

- Standardized name formats by sorting alphabetically and applying **Find and Replace**.
- Example: "A. A. Warsi," "A. Warsi," and "Ahfad Warsi" were unified to maintain consistency.

#### f. Service(km)

- Changed data type to numerical format using **Format Cells**.
- Missing values replaced with "0" for consistency.
- Applied number formatting to include commas for readability.

#### g. Workshop Name

- Some entries contained state names; these were separated using **Text to Columns**, and the redundant state column was deleted.
- Sorted alphabetically and corrected typographical errors using **Find and Replace**.
- Applied TRIM function to remove unnecessary spaces and PROPER function for uniform casing.

#### h. Invoice Number

- Contained numeric, alphanumeric, and alphabetic values with some missing data.
- Changed data type to text since invoice numbers do not require calculations.
- Standardized values by replacing "Petty Cash" entries with "PC001".
- Missing values replaced with "NA" (Not Available) for clarity.

#### i. Amount (NGN)

- Changed data type to currency.
- Customized formatting to display the Naira (₦) symbol.
- No missing values were detected in this column.

#### j. Remarks

• The column was removed due to being empty.

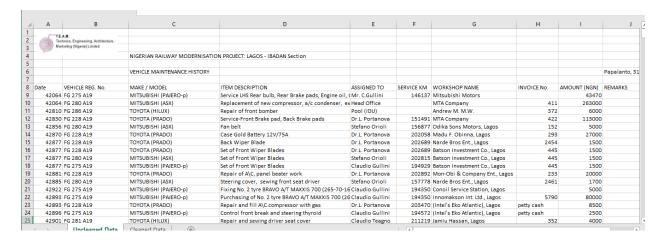
## **4** Final Verification Using Power Query

- Loaded the cleaned dataset into **Power Query** for additional validation.
- Identified and corrected minor typographical errors.
- Reloaded the refined dataset back into Microsoft Excel for final use.

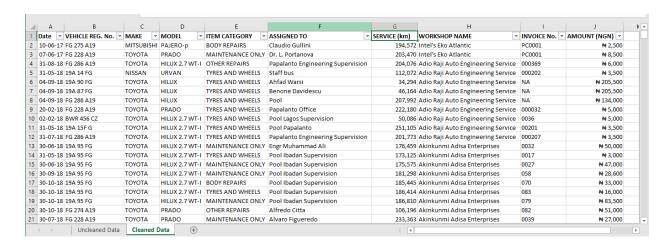
## **Conclusion**

The data cleaning process significantly enhanced the dataset's accuracy, structure, and usability. By systematically addressing inconsistencies, removing redundant information, and applying appropriate formatting techniques, the dataset is now well-structured for further analysis. The refinements improve data integrity, ensuring reliable insights for decision-making. The final dataset is free from duplicates, contains correctly formatted values, and adheres to standardized naming conventions, making it more efficient for reporting and analysis

#### **PICTURES**



#### **BEFORE CLEANING**



#### **AFTER CLEANING**

# LINK

https://docs.google.com/spreadsheets/d/1TGoD1g7KLcqjrqMIFXpBs7-ibfA3cl9W/edit?usp=sharing&ouid=106812219724605440864&rtpof=true&sd=true