Data Cleaning Process for Dashboard: "A Case Study of the Consumption Rate of Hard Substances in Lagos"

## Step 1: Data Importation

Opened Excel and imported raw data from a CSV, Excel, or flat file. Loaded data into Power Query Editor for cleaning and transformation.

## Step 2: Initial Assessment

Inspected the dataset for:  
- Missing values  
- Inconsistent formats (text, number, date)  
- Irrelevant columns  
- Duplicates

## Step 3: Data Cleaning in Power Query

### 1. Renamed Columns

Renamed ambiguous column names like:  
- Col1 → ID  
- Col2 → Age  
- Col3 → Gender  
- Col4 → Substance  
- Col5 → Location  
- Col6 → First Time Usage  
- Col7 → Monthly Income  
- Col8 → Financial Status

### 2. Removed Duplicates

Removed duplicate rows based on ID or a combination of ID, Age, Gender, and Substance.

### 3. Replaced Nulls/Missing Values

For Gender, Financial Status, and Substance Type, used 'Replace Values' to fill missing with 'Unknown' or 'Not Specified'.  
For numerical columns like Monthly Income or First Time Usage, replaced nulls with 0 or used mean/median if required.

### 4. Data Type Conversion

Converted:  
- Age → Whole Number  
- Monthly Income → Currency  
- First Time Usage → Whole Number  
- Gender and Substance → Text  
- Date fields (if present) → Date/Time

### 5. Filtered Out Invalid Records

Removed:  
- Rows with empty Substance or Gender  
- Unrealistic Age values (e.g., under 10 or over 80)  
- Locations not in Lagos

### 6. Categorization/Grouping

Created age groups using 'Add Column → Conditional Column':  
- ≤18 = '≤18'  
- 19–25 = '19-25'  
- 26–35 = '26-35'  
- 36–45 = '36-45'  
- >45 = '>45'  
  
Grouped financial status into:  
- Low Income  
- Middle Income  
- High Income

### 7. Standardized Values

Ensured consistency in Substance names:  
- e.g., 'Codein', 'Codeine', 'Codine' → 'Codeine'  
Capitalized values using Transform → Format → Capitalize Each Word

## Step 4: Load to Excel / Power BI

Loaded the clean data back into Excel as a table. Exported table to Power BI for building the interactive dashboard.