



Power BI Data Engineering ETL- Power Query-DAX Modeling

*29 September 2025
Data Analysis*

*Frank Ouna
Data Analyst*

INTRODUCTION

This document outlines the End-to-End Data Engineering process used to build the Samba Bar & Restaurant Performance Dashboard. The process followed the standard ETL (Extract, Transform, Load) paradigm, leveraging Power BI's powerful data preparation and modelling tools.

1. Extraction (E)

The **Extraction** phase involves connecting to and pulling data from various source systems into Power BI.

File Formats & Sources Used:

This project consolidated data from multiple sources and formats, simulating a real-world data environment:

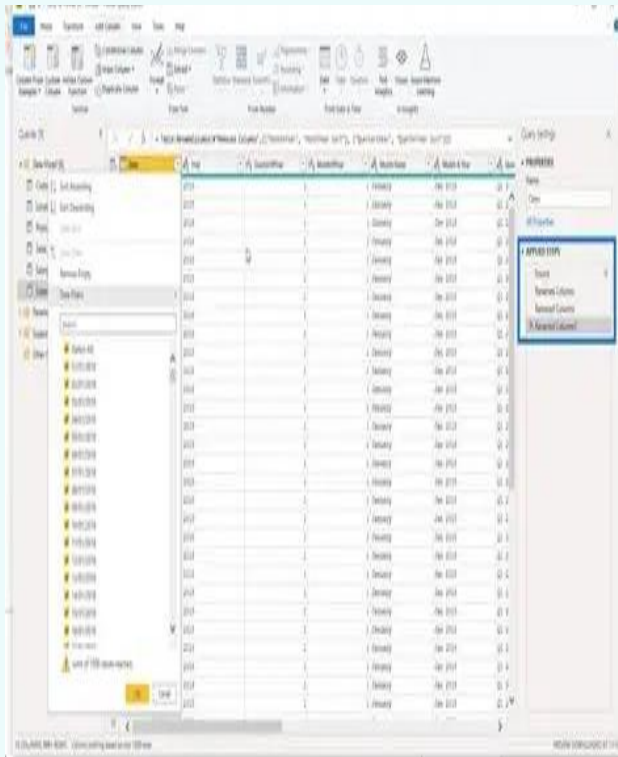
Excel Workbooks (.xlsx): Used for structured tabular data like products_with_price.xlsx and staff_branch_assignments.xlsx.

JSON Files (.json): Used for semi-structured data from APIs or database exports, such as branches_with_ids.json, staff.json, and cities_with_id.json.

Combined CSV Files: Initial sales data was provided as multiple CSV files, which were first consolidated outside Power BI into a single Excel workbook (Combined_Mombasa_Sales.xlsx) for simplicity.

Tool Used: Power BI's "Get Data" feature, which provides native connectors for dozens of data sources.





2. Transformation (T)

The Transformation phase is where raw data is cleaned, reshaped, and prepared for analysis. This is the most critical step for ensuring data quality.
Tool Used: Power Query Editor in Power BI.

Key Transformations Performed:

Data Type Conversion: **Changed the sale_date column from a general data type to a proper Date/Time type to enable time-based analysis.**

Column Derivation: **Used Power Query's built-in functions to split the sale_date into two new columns:**

date (Date Only)

time (Time Only)

Column Removal: **Removed unnecessary metadata columns like "Source_Name" that were not relevant for analysis.**

Data Structuring (for JSON): **Expanded nested JSON structures into flat, tabular formats suitable for relational modeling.**

Data Consolidation: **Initially combined multiple sales CSV files into a single table to create a unified sales fact table.**

3. Loading (L)

The **Loading** phase involves writing the transformed data into the Power BI **Data Model**, where relationships and calculations are defined.

- After all transformations were completed in Power Query, the data was loaded by clicking "**Close & Apply**".
- This action loads the cleaned tables into the in-memory engine of Power BI, ready for creating relationships and measures.

Data Modeling with DAX

Once the data is loaded, the next step is to build the data model and create advanced calculations.

Tool Used: DAX (Data Analysis Expressions), a formula language for creating custom calculations and dynamic aggregations.

Key DAX Implementation:

A custom **Calendar table** was created using DAX to enable robust time intelligence calculations (e.g., Year-to-Date, Month-over-Month comparisons).

```
Calendar =  
VAR MinDate = MIN ( Sales[Date] )  
VAR MaxDate = MAX ( Sales[Date] )  
RETURN  
    ADDCOLUMNS (  
        CALENDAR ( MinDate, MaxDate ),  
        "Year", YEAR ( [Date] ),  
        "MonthNumber", MONTH ( [Date] ),  
        "MonthName", FORMAT ( [Date], "MMMM" ),  
        "Quarter", "Q" & FORMAT ( [Date], "Q" ),  
        "YearMonth", FORMAT ( [Date], "YYYY-MM" ),  
        "Day", DAY ( [Date] ),  
        "Weekday", FORMAT ( [Date], "dddd" ),  
        "WeekdayNumber", WEEKDAY ( [Date], 2 ),  
        "day_week",  
            SWITCH(  
                MONTH ( [Date] ),  
                1, "Jan",  
                2, "Feb",  
                3, "Mar",  
                4, "Apr",  
                5, "May",  
                6, "Jun",  
                7, "Jul",  
                8, "Aug",  
                9, "Sep",  
                10, "Oct",  
                11, "Nov",  
                12, "Dec",  
                0, "Invalid" )  
            )  
    )
```

```
        WEEKDAY([Date], 2),  
        1, "Mon",  
        2, "Tues",  
        3, "Wed",  
        4, "Thu",  
        5, "Fri",  
        6, "Sat",  
        7, "Sun",  
        BLANK()  
    )  
)
```

Final Model Setup:

- The `Calendar` table was marked as an official date table.
- A relationship was created between the `Calendar[Date]` and `Sales[Date]` columns, establishing a **Single, Many-to-One** filter direction to enable accurate time-based filtering across all reports.

This ETL pipeline successfully transformed disparate raw data from **Excel and JSON files** into a clean, analysable, and related dataset. **Power Query** handled the heavy lifting of data cleaning and preparation, while **DAX** was used to build the core data model and time intelligence foundation, enabling the creation of dynamic and insightful dashboard visualizations.

-Conclusion