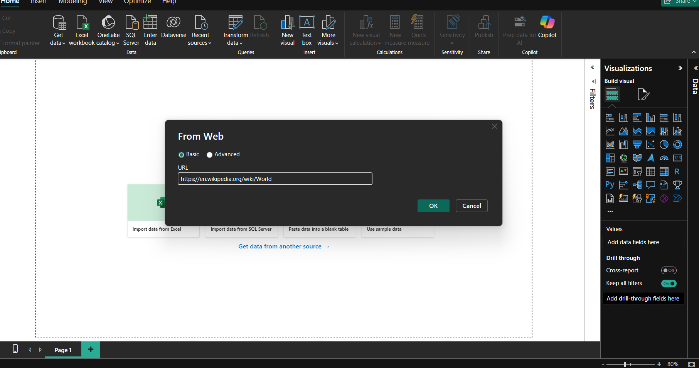
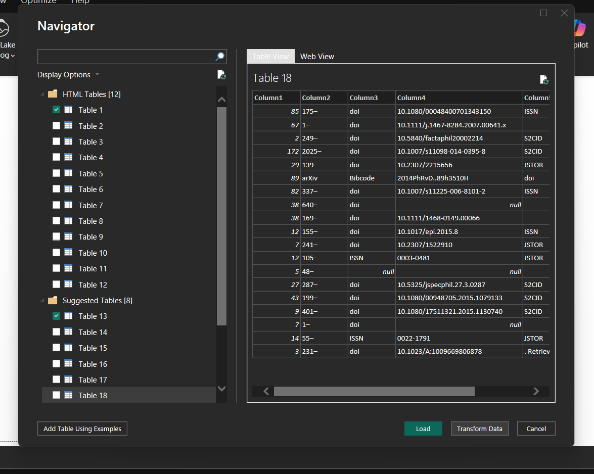
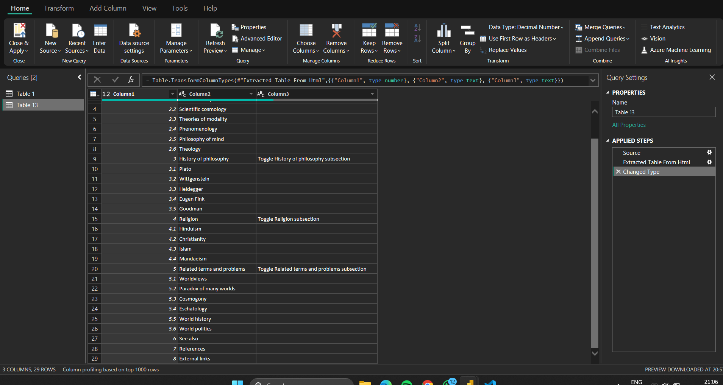
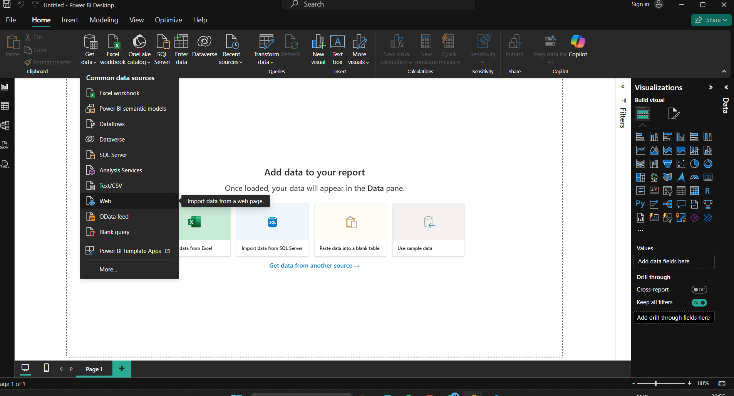
**Project Title:**  
*Data Leverage – Power Query Transformation Project*

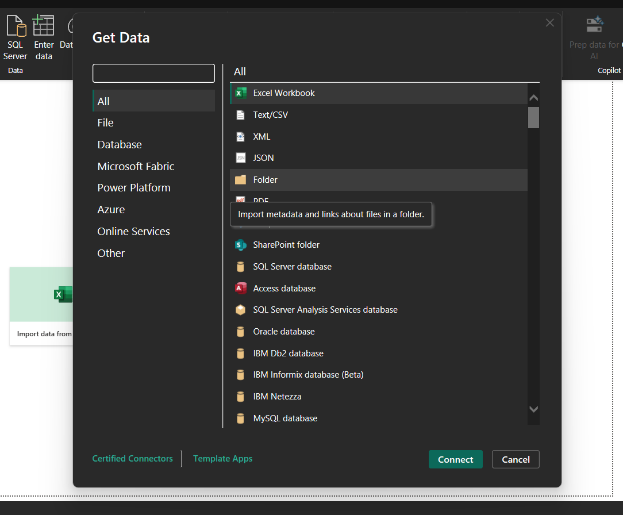
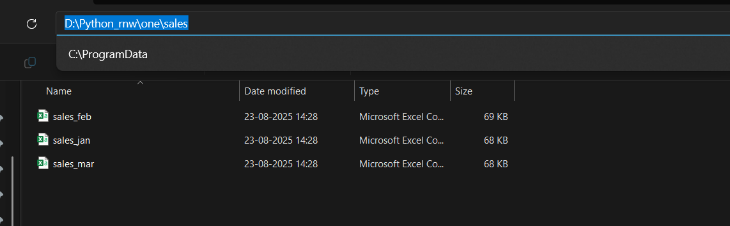
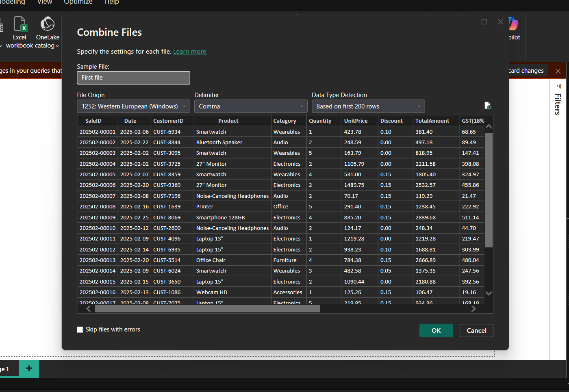
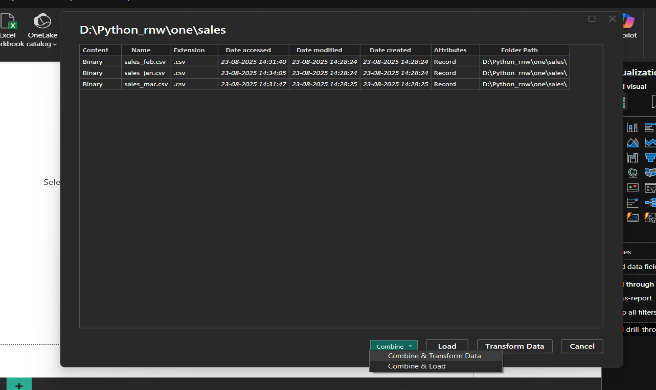
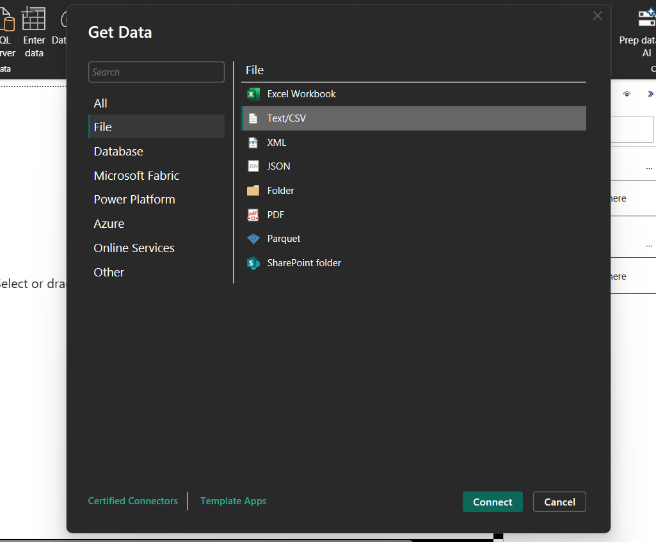
**Objective:**  
The purpose of this project was to simulate a **data engineering scenario** in Power BI using the **Power Query Editor**. The focus was exclusively on **data extraction, cleaning, transformation, and integration** without the use of DAX or visualizations.

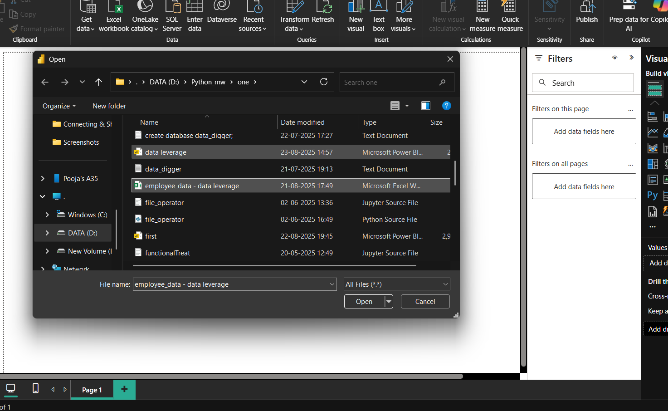
**Data Sources**

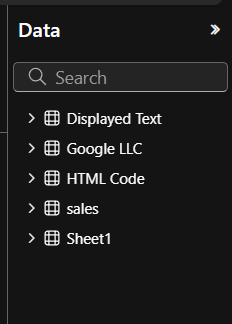
1. **Web Source:** Wikipedia table (Country-wise GDP).

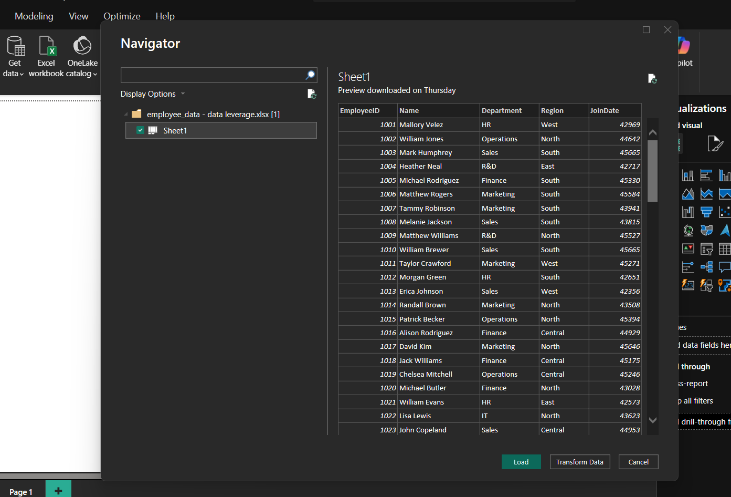




1. **Folder Source:** Three Excel files (*Sales\_Jan, Sales\_Feb, Sales\_Mar*).
2. **Excel File:** Employee dataset (EmpID, Name, Department, Region, Join Date).

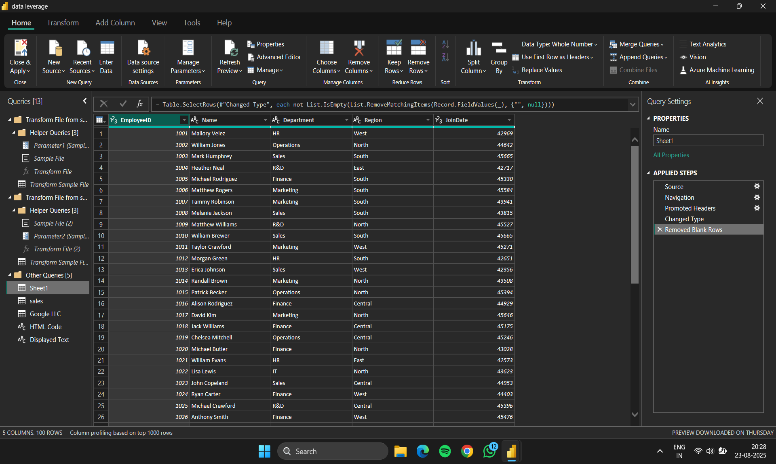


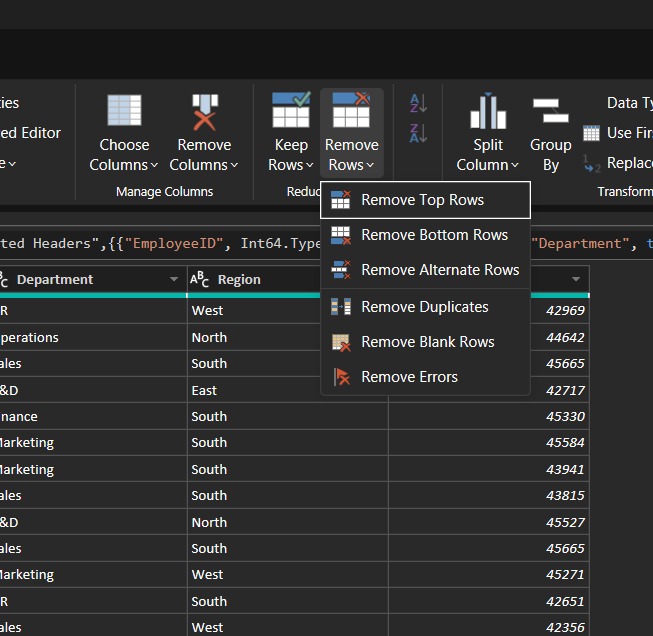




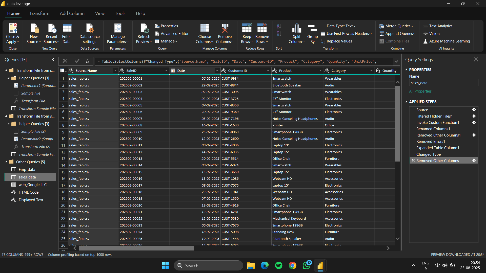
**Transformations Applied**

**1. Data Cleaning**

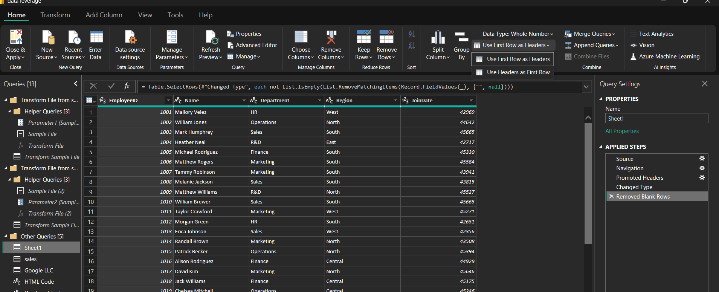
* Removed blank rows and null values.



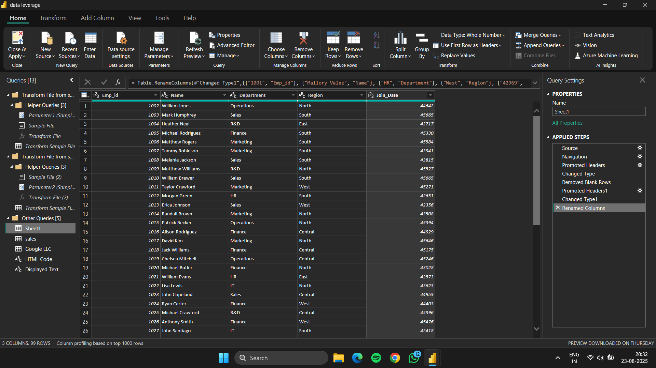
* Removed unnecessary columns (including fully blank columns).



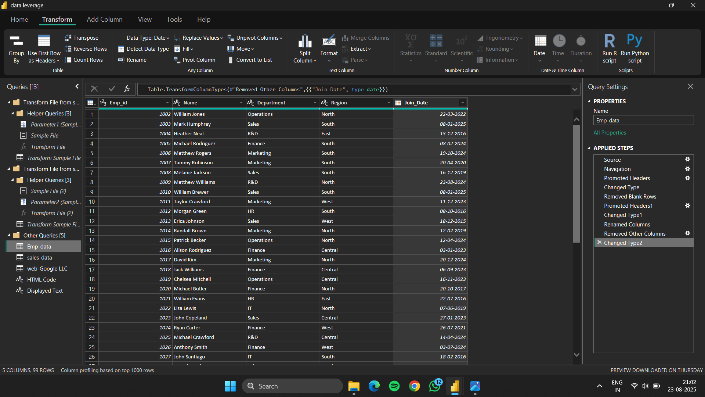
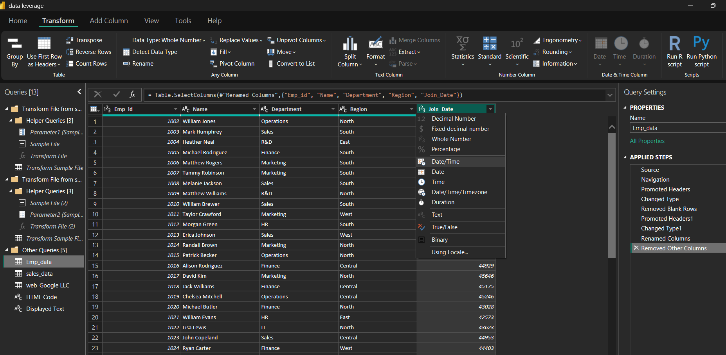
* Promoted headers.



* Renamed columns with meaningful names.

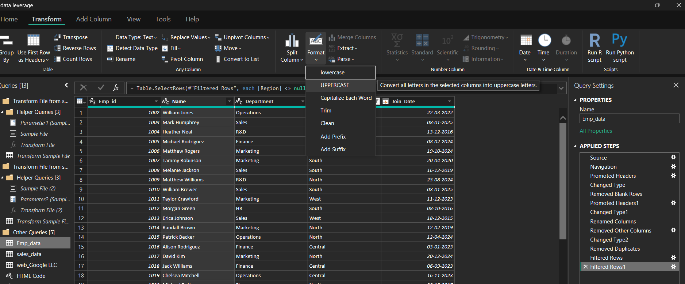


* Applied correct data types

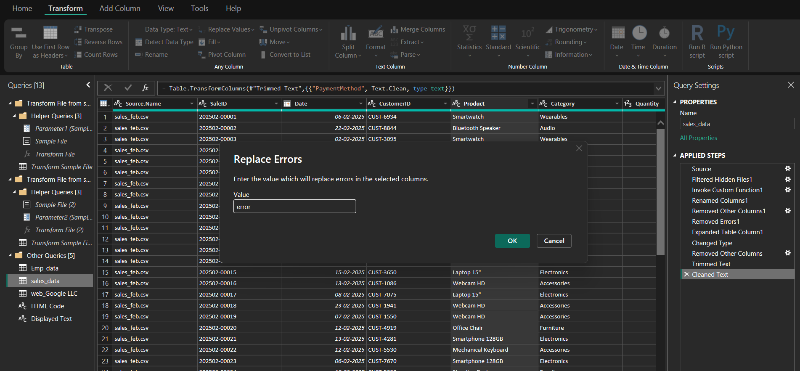


**2. Text Transformations**

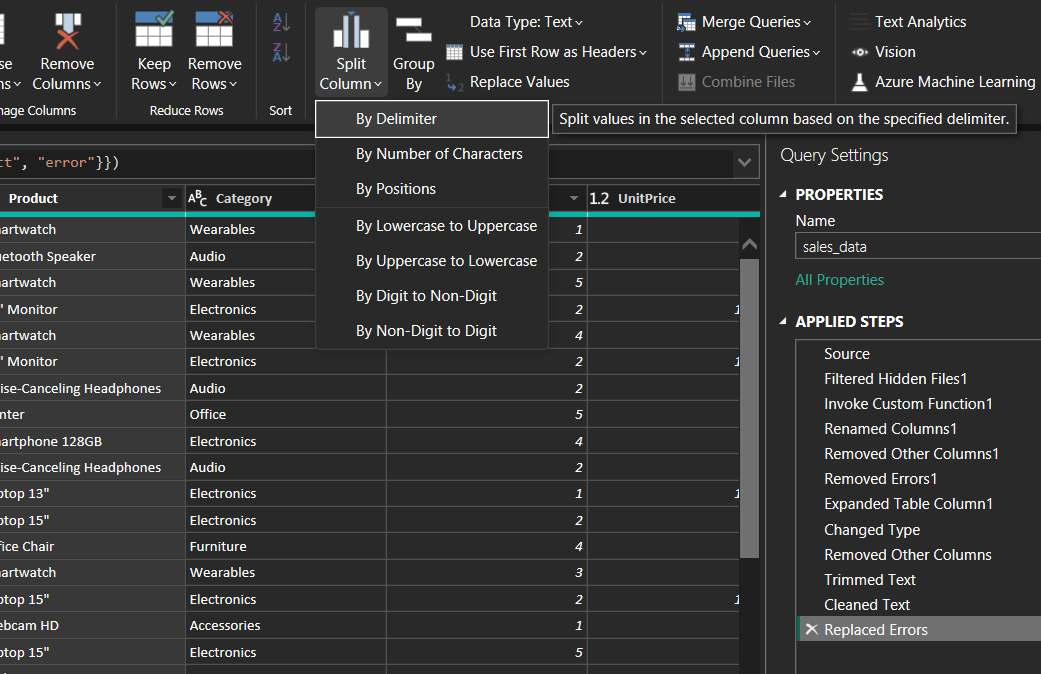
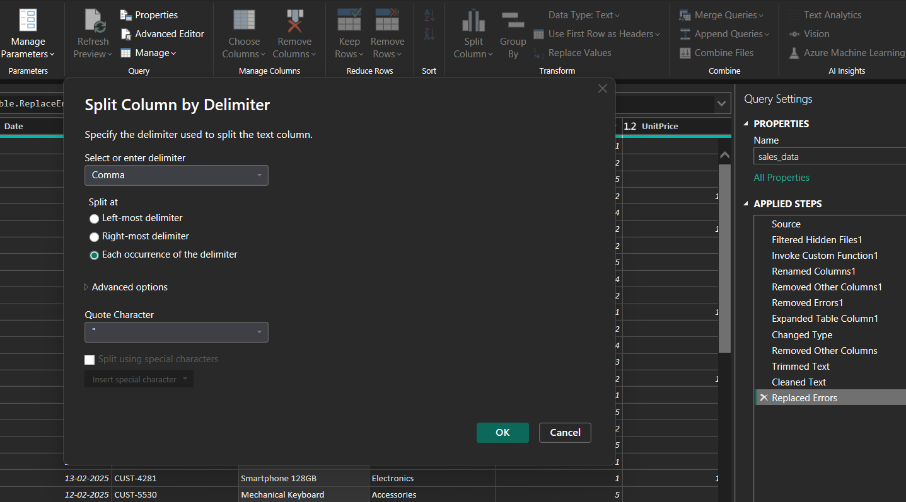
* Standardized names and addresses with UPPER(), LOWER(), TRIM(), CLEAN().



* Used REPLACE() to correct typos.

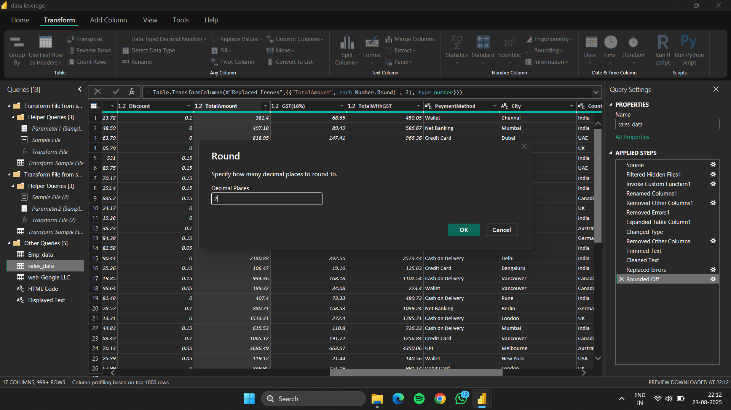


* Split address column by delimiter (e.g., city/state).

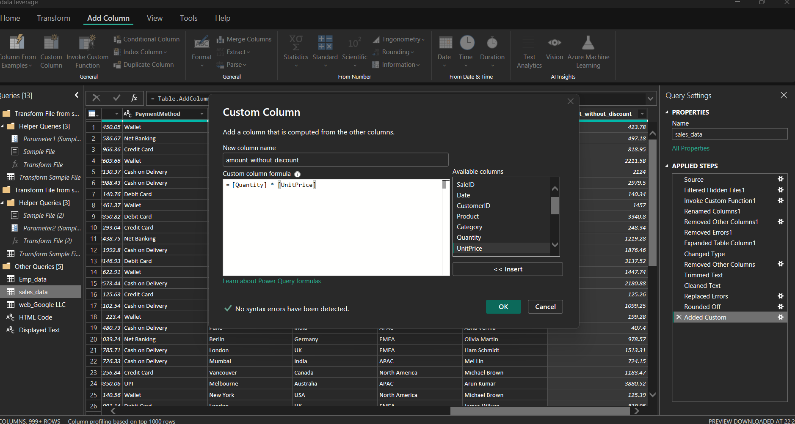


**3. Numeric Transformations**

* Rounded Revenue to 2 decimal places.

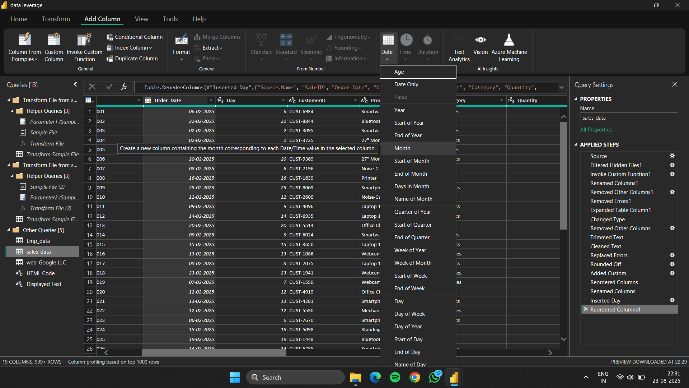


* Created new column **Amount\_without\_discount = Unit Price \* Quantity**.

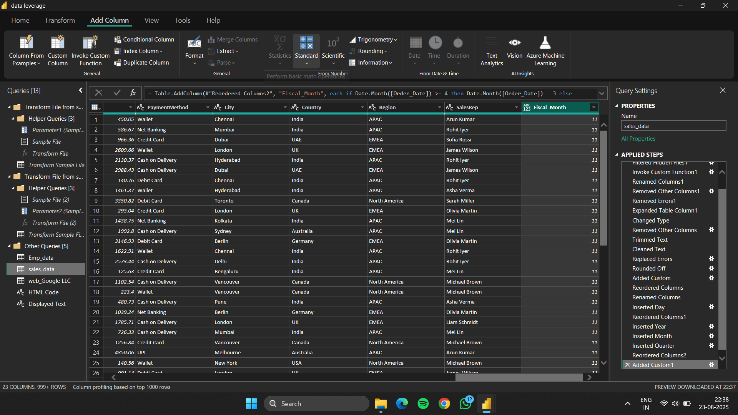


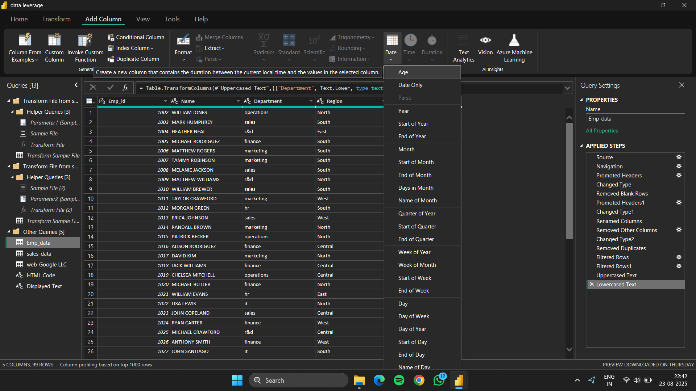
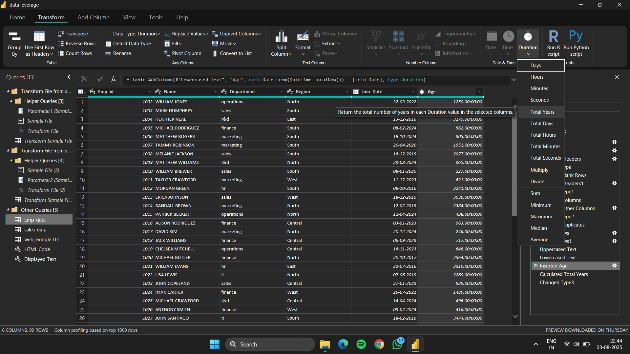
**4. Date & Time Transformations**

* Extracted Day, Month, Year, Quarter from Order Date.



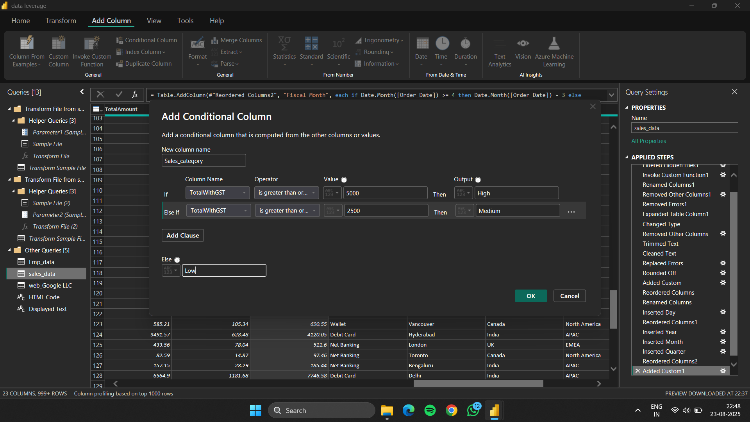
* Created **Fiscal Month** column with custom logic.

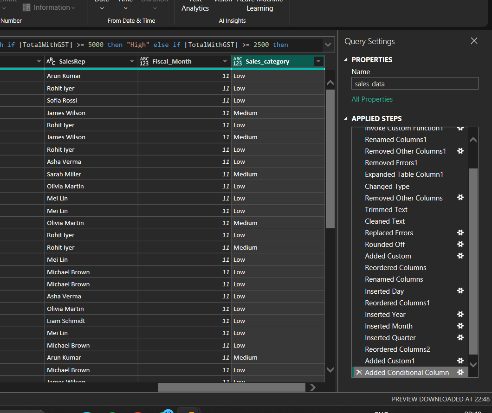


* Calculated **Joining** **Year** from **Join\_date**.

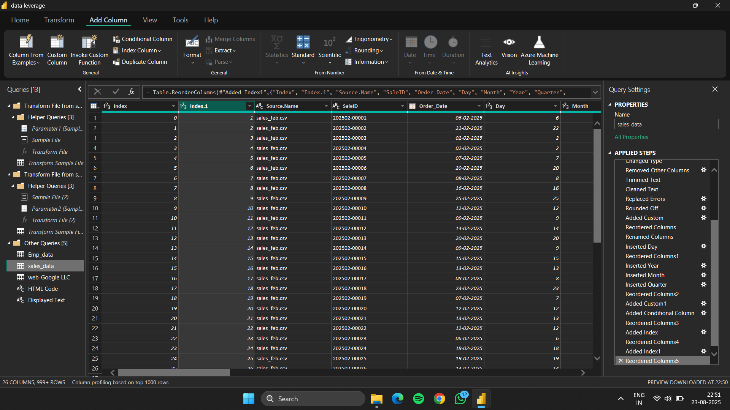
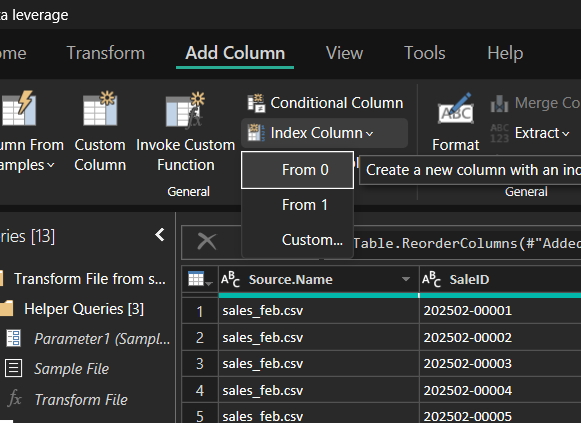
**5. Conditional & Index Columns**

* Added **Total\_amount** column: High, Medium, Low (based on revenue).

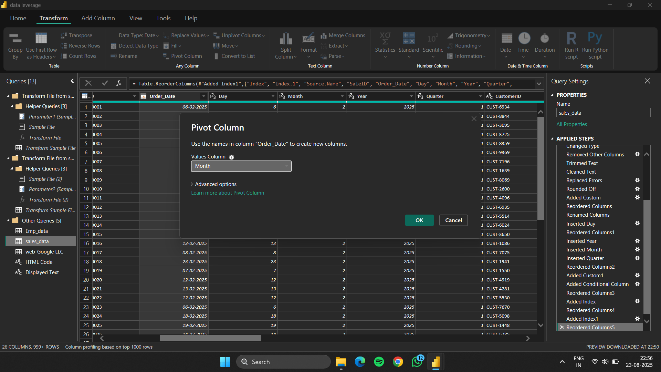
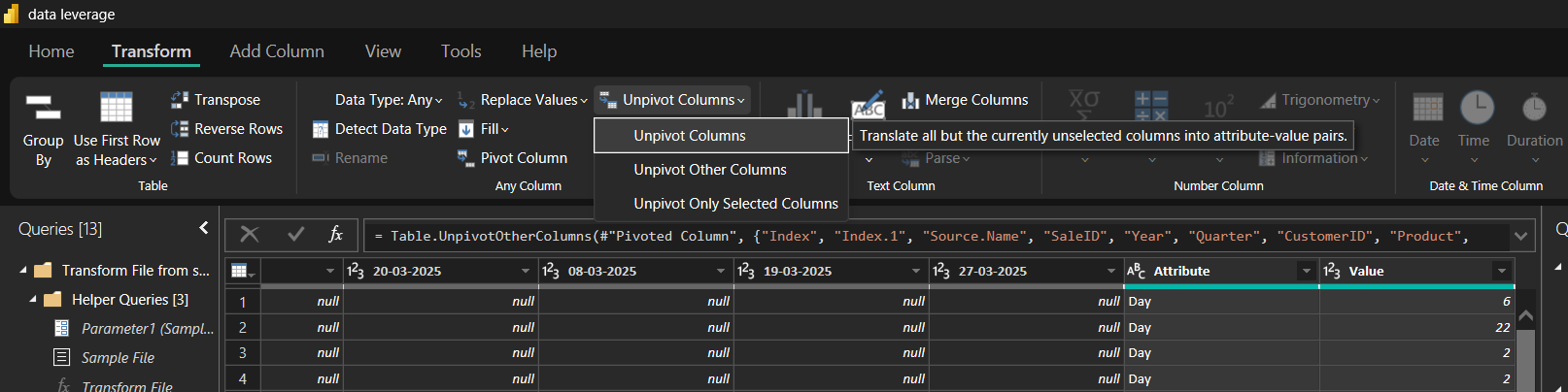




* Added Index columns (0-based and 1-based).

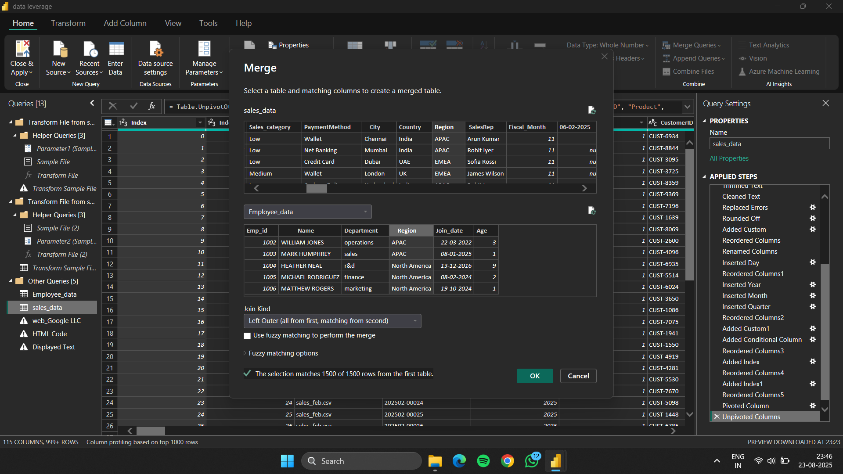


**6. Pivoting & Unpivoting**

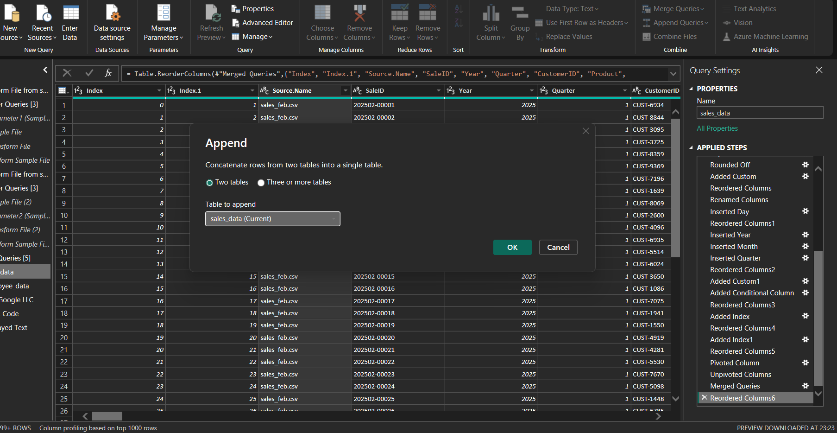
* Pivoted sales data into wide format.
* Unpivoted data back into normalized long format.

**7. Merging & Appending**

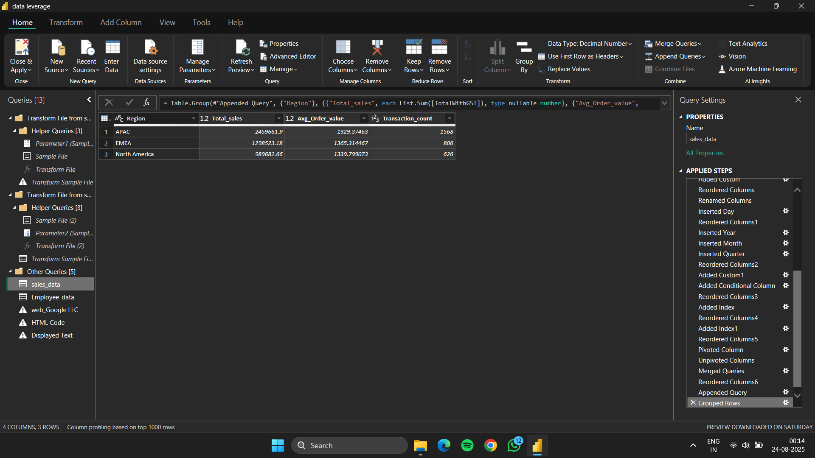
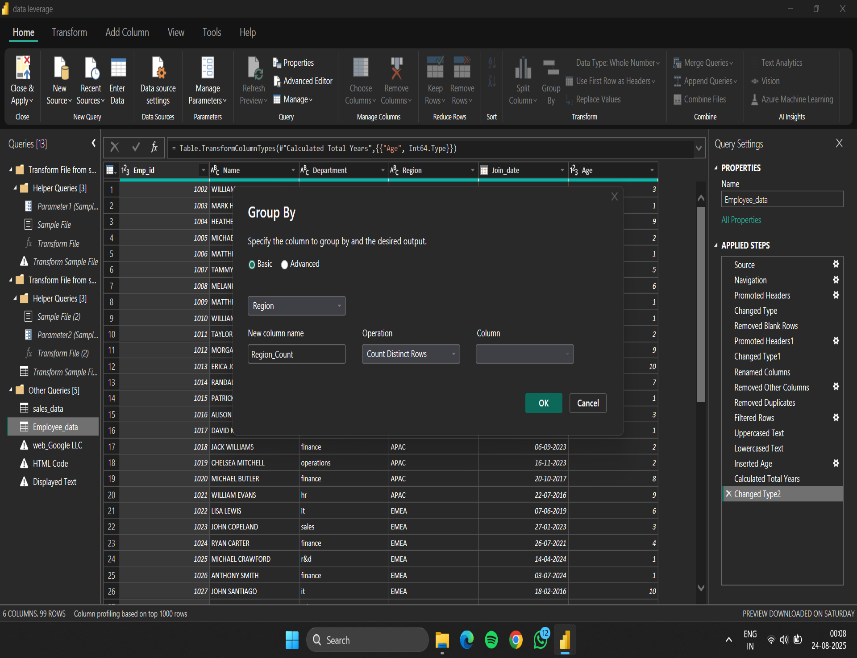
* Merged Sales data with Employee data using Region/EmployeeID.



* Appended Sales\_Jan, Sales\_Feb, and Sales\_Mar queries.

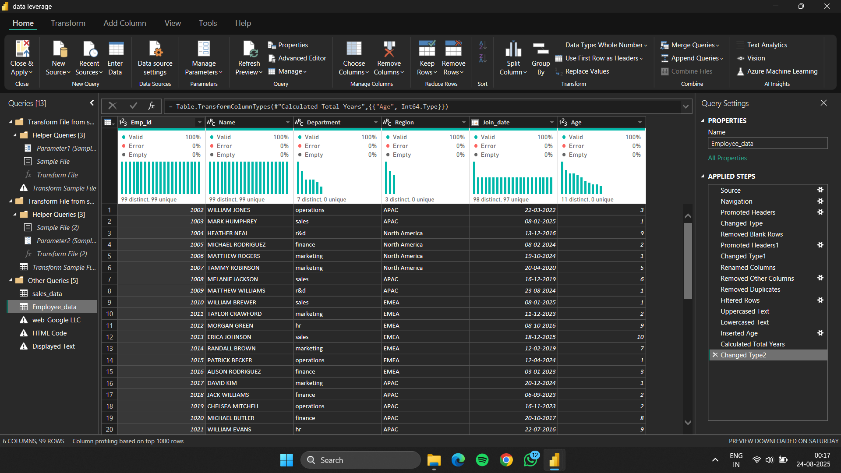


**8. Grouping & Aggregation**

* Grouped by Region to compute:
  + Total Sales
  + Average Order Value
  + Transaction Count

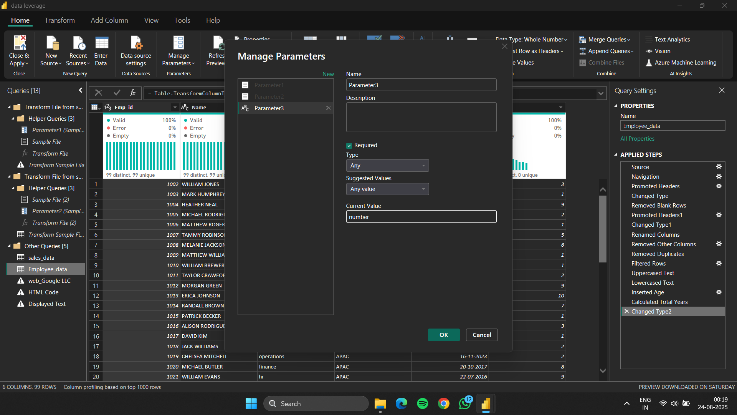
**9. Data Profiling & Quality**

* Used Column Profile, Column Quality, and Column Distribution to detect:
  + Missing values
  + Data type mismatches
  + Distinct/unique values

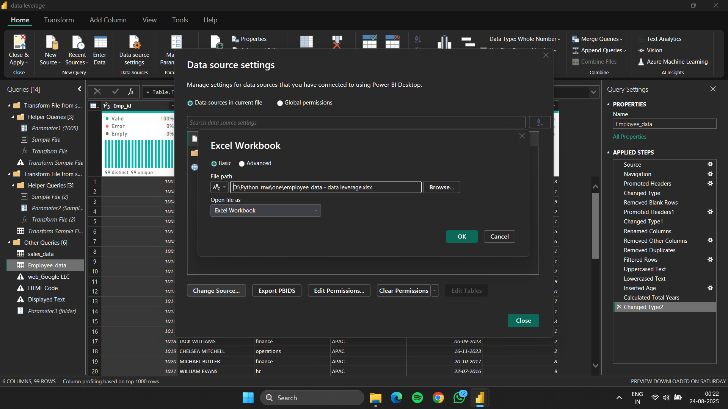


**10. Source Settings & Parameters**

* Created **parameterized folder path** for dynamic file loading.

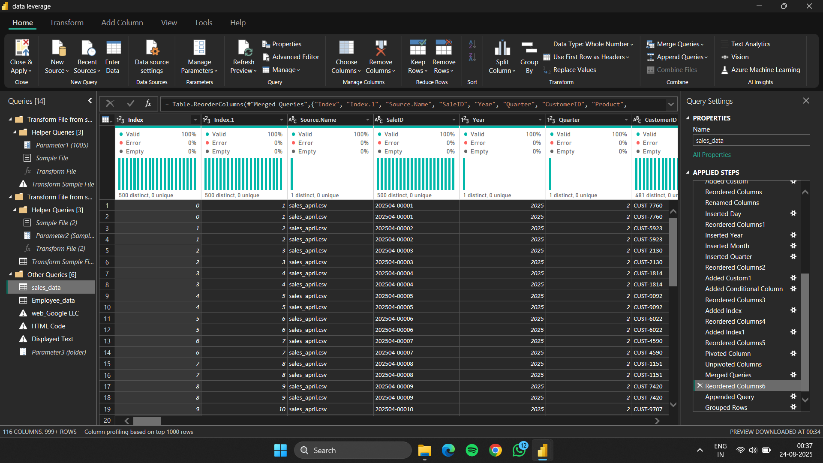


* Adjusted **Data Source Settings** for credentials.



**11. Refresh Simulation**

* Added **Sales\_Apr.xlsx** to the folder.
* Refreshed queries → data auto-loaded with transformations intact.



**Challenges Faced & Solutions**

1. **Blank Columns:** No direct “Remove Blank Columns” button in Power Query.
   * *Solution:* Removed manually and applied M-code to exclude fully null columns.
2. **Null Values:** Sales and Employee datasets contained missing IDs and dates.
   * *Solution:* Applied filters to remove (null) values from critical columns.
3. **Locale Issues in Currency/Date:** Different formats in source files.
   * *Solution:* Used **Change Type with Locale** for proper standardization.
4. **Refresh Simulation:** Initial refresh did not include new file.
   * *Solution:* Corrected parameterized folder path to auto-include new files.

**Conclusion**

This project demonstrated the **end-to-end ETL process** in Power Query:

* Extracting multiple data sources
* Cleaning and standardizing data
* Transforming text, numeric, and date fields
* Creating derived columns
* Merging and appending datasets
* Profiling and validating data quality
* Automating refresh with parameterized folder paths

The final dataset is now **clean, structured, and ready for analysis**.