

Data Leverager – Power Query Transformation Project

Tool Used: Power BI Desktop (Power Query Editor)

Project Type: ETL (Extract, Transform, Load)

Data Sources Used

This project uses multiple data sources to demonstrate real-world data integration and transformation.

1. Employee Dataset (Excel File)

- **Source Type:** Excel (.xlsx)
 - **Description:**
Contains employee-related information such as:
 - EmployeeID
 - Employee Name
 - Department
 - Region
 - Join Date
 - **Purpose:**
Used to enrich sales data with department and employee details.
-

2. Sales Dataset (Excel File)

- **Source Type:** Excel (.xlsx)
 - **Description:**
Contains transactional sales data including:
 - Date
 - Product
 - Region
 - Salesperson
 - Units Sold
 - Unit Price
 - Total Sales
 - **Purpose:**
Used to analyze sales performance across regions, products, and time.
-

3. GDP Dataset (Web Source – Wikipedia)

- **Source Type:** Web (HTML Table)
 - **URL Used:**
[https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal))
 - **Description:**
Country-wise GDP information including:
 - GDP (Full Value)
 - GDP Growth (%)
 - GDP per Capita
 - Share of World GDP
 - **Purpose:**
Demonstrates web data extraction and advanced data cleaning.
-

Transformations Applied

All transformations were performed using **Power Query Editor**.

◆ Common Transformations

- Removed unnecessary columns
 - Removed blank rows and duplicate records
 - Promoted first row as headers
 - Renamed columns for clarity
 - Set correct data types (Text, Date, Whole Number, Decimal)
-

◆ Employee Data Transformations

- Converted **Excel serial Join Date** into proper Date format using locale settings
 - Cleaned text columns using **Trim** and **Clean**
 - Added **Index columns** (from 0 and from 1)
 - Created a calculated column for **Employee Tenure (Years)**
-

◆ Sales Data Transformations

- Converted **Excel serial Date** into standard Date format
- Cleaned Product, Region, and Salesperson text fields
- Validated **Total Sales** using:

- Units Sold × Unit Price
 - Created **Sales Category** using Conditional Column:
 - High (≥ 30000)
 - Medium (10000–29999)
 - Low (< 10000)
 - Extracted Date features:
 - Year
 - Month
 - Month Name
 - Quarter
 - Added Index columns
 - Performed **Group By** to calculate:
 - Total Revenue
 - Average Sales
 - Transaction Count
-

◆ GDP Data Transformations

- Removed Rank (#) column
 - Cleaned GDP Growth by removing % and fixing negative symbols
 - Converted GDP Growth to Decimal Number
 - Cleaned GDP (Full Value) by removing \$ and ,
 - Converted GDP per Capita to numeric format
 - Added GDP Category using Conditional Column
 - Sorted countries by GDP value
 - Added Index column for ranking
-

◆ Merging & Integration

- Merged **Sales data with Employee data** using:
 - Salesperson \leftrightarrow Employee Name
 - Join Type: Left Outer
- Ensured all sales records were preserved after merge

Challenges Faced & Solutions

Challenge 1: Excel Serial Dates

- **Problem:**
Dates appeared as numbers like 45292, 44699
 - **Solution:**
Converted using **Change Data Type → Using Locale → Date**
 - **Outcome:**
Dates became readable and usable for time analysis
-

Challenge 2: Currency & Percentage Values

- **Problem:**
GDP values contained \$, , and growth values had %
 - **Solution:**
Used **Replace Values** to remove symbols before changing data type
 - **Outcome:**
Numeric columns became calculation-ready
-

Challenge 3: Data Consistency for Merge

- **Problem:**
Employee and Sales data used names instead of IDs
- **Solution:**
Performed merge using **Salesperson ↔ Employee Name** with Left Join
- **Outcome:**
Sales data was successfully enriched with employee details