Data Leverager – Power Query Transformation Project

Student Name: Arya Patel
Subject: Power BI
Submission Date: 27 February 2026

1. Project Objective

The objective of this project is to perform complete ETL (Extract, Transform, Load) operations using Power BI Power Query Editor. The focus is on data cleaning, transformation, integration, and preparation without using DAX or visualizations.

2. Data Sources Used

- Excel Folder Source: Sales_Jan.xlsx, Sales_Feb.xlsx, Sales_Mar.xlsx
- Employee Dataset (Excel File)
- Web HTML Table (GDP data from Wikipedia)

3. Step-by-Step Transformations Applied

- Loaded monthly sales files using Folder Source and combined them.
- Removed blank rows and duplicate records.
- Promoted first row as headers.
- Corrected data types (Date, Decimal Number, Whole Number).
- Created Profit column using custom formula (Revenue - Cost).
- Created Sales Category using Conditional Column (High, Medium, Low).
- Extracted Year, Month, and Quarter from Order Date.
- Loaded and cleaned Employee dataset.
- Merged Sales and Employee tables using EmployeeID (Left Outer Join).
- Expanded employee details (Name, Department, Region, Birthdate).
- Created Age column from Birthdate.
- Performed Group By on Region (Total Sales, Average Sales, Transactions).
- Added Index column for row numbering.
- Loaded Web GDP data and cleaned numeric values.
- Applied Column Quality, Column Distribution, and Column Profile tools.

4. Challenges Faced and Solutions

- Data type mismatch – Fixed by proper data type conversion.
- Duplicate and blank rows – Removed using Power Query tools.
- Combining multiple monthly files – Used Folder source and automatic append.
- Text inconsistencies – Cleaned using Trim and Clean transformations.
- Numeric formatting issues in Web data – Removed commas and converted to Decimal.

5. Conclusion

This project successfully demonstrates a complete ETL workflow using Power BI Power Query. Data was extracted from multiple sources, cleaned, transformed, merged, and summarized. The project ensures data accuracy, consistency, and automation while maintaining data quality standards.