Power BI ETL Project: Sales & Employee Data Transformation Using Power Query

Project Overview

This project implements a complete ETL (Extract, Transform, Load) workflow using Power BI Power Query Editor. The objective is to extract data from multiple sources, clean and transform it systematically, and prepare reliable datasets for analysis. The project focuses entirely on data transformation logic without using DAX or visualizations.

Data Sources Used

• Monthly Sales Excel files (Jan–Apr) loaded dynamically from a folder
• Employee master data stored in Excel
• Web-based HTML dataset extracted from a public source
• Folder Path parameter for dynamic data refresh

Data Transformation Using Power Query

All data preparation steps were performed inside Power Query Editor following a structured ETL pipeline.

Web Dataset Transformation

Extracted tabular data from an HTML source, promoted headers, assigned correct data types, and validated data quality using profiling tools.

Sales Data Transformation

Standardized text fields using TRIM, CLEAN, and UPPER functions, split address fields, renamed and reordered columns, extracted date components, added conditional sales categories, and created index columns.

Employee Dataset Transformation

Promoted headers, corrected data types, calculated employee age using birth year logic, and validated data quality.

Merge & Aggregation

Appended monthly sales data using folder-based queries, merged sales with employee data, and grouped by Region to calculate total sales, average order value, and transaction count.

Folder Path Parameter

A Folder Path parameter was implemented to ensure a flexible and reusable ETL pipeline. This allows seamless ingestion of new monthly files without modifying transformation logic.

Data Modeling

The cleaned datasets were structured using Region and Employee ID as linking fields, forming a clean relational model suitable for downstream analysis.

Observations

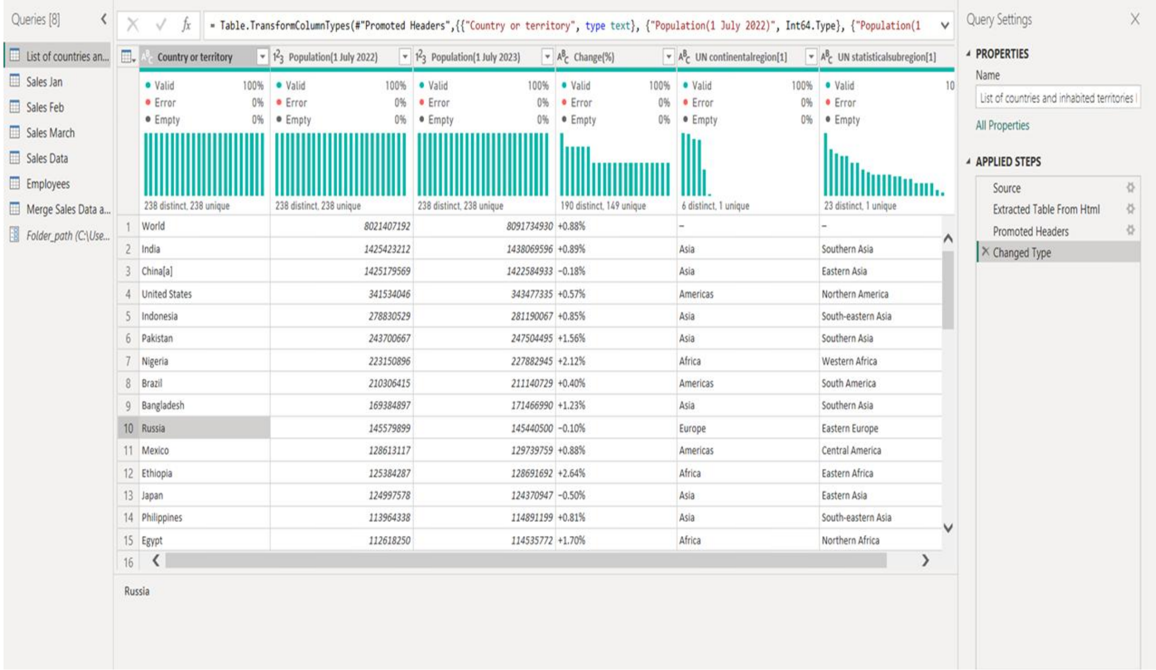
- Sales distribution varies across regions
- Some regions show higher transaction volumes
- Average order values differ region-wise

Tools & Technologies Used

- Power BI Desktop
- Power Query (M Language)
- Microsoft Excel
- GitHub

Conclusion

This project demonstrates practical knowledge of Power BI ETL workflows including data extraction, transformation, merging, parameterization, and aggregation while following industry best practices.



```
= Table.TransformColumnTypes(#"Promoted Headers",{{"Country or territory", type text}, {"Population(1 July 2022)", Int64.Type}, {"Population(1
```

	Country or territory	Population(1 July 2022)	Population(1 July 2023)	Change(%)	UN continentalregion[1]	UN statisticalsubregion[1]
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 10 ● Error ● Empty
	238 distinct, 238 unique	238 distinct, 238 unique	238 distinct, 238 unique	190 distinct, 149 unique	6 distinct, 1 unique	23 distinct, 1 unique
1	World	8021407192	8091734930	+0.88%	–	–
2	India	1425423212	1438069596	+0.89%	Asia	Southern Asia
3	China[a]	1425179569	1422584933	-0.18%	Asia	Eastern Asia
4	United States	341534046	343477335	+0.57%	Americas	Northern America
5	Indonesia	278830529	281190067	+0.85%	Asia	South-eastern Asia
6	Pakistan	243700667	247504495	+1.56%	Asia	Southern Asia
7	Nigeria	223150896	227882945	+2.12%	Africa	Western Africa
8	Brazil	210306415	211140729	+0.40%	Americas	South America
9	Bangladesh	169384897	171466990	+1.23%	Asia	Southern Asia
10	Russia	145579899	145440500	-0.10%	Europe	Eastern Europe
11	Mexico	128613117	129739759	+0.88%	Americas	Central America
12	Ethiopia	125384287	128691692	+2.64%	Africa	Eastern Africa
13	Japan	124997578	124370947	-0.50%	Asia	Eastern Asia
14	Philippines	113964338	114891199	+0.81%	Asia	South-eastern Asia
15	Egypt	112618250	114535772	+1.70%	Africa	Northern Africa
16						

Russia

Query Settings

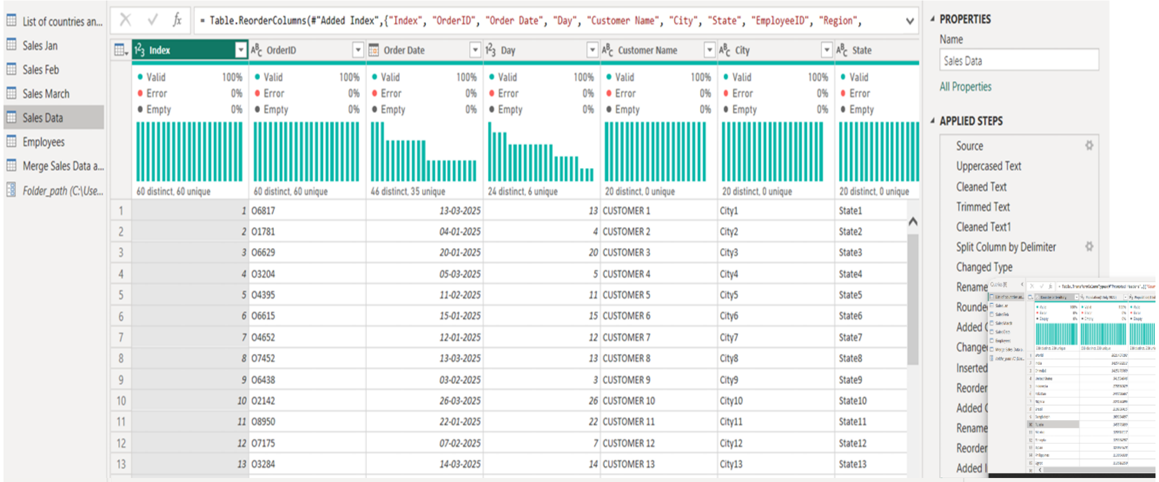
PROPERTIES

Name: List of countries and inhabited territories

All Properties

APPLIED STEPS

- Source
- Extracted Table From Html
- Promoted Headers
- Changed Type



- List of countries an...
- Sales Jan
- Sales Feb
- Sales March
- Sales Data
- Employees
- Merge Sales Data a...
- Folder_path (C:\Use...

```
= Table.ReorderColumns(#"Added Index",{"Index", "OrderID", "Order Date", "Day", "Customer Name", "City", "State", "EmployeeID", "Region",
```

	Index	OrderID	Order Date	Day	Customer Name	City	State
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid Error Empty
	60 distinct, 60 unique	60 distinct, 60 unique	46 distinct, 35 unique	24 distinct, 6 unique	20 distinct, 0 unique	20 distinct, 0 unique	20 distinct, 0 unique
1	1	O6817	13-03-2025	13	CUSTOMER 1	City1	State1
2	2	O1781	04-01-2025	4	CUSTOMER 2	City2	State2
3	3	O6629	20-01-2025	20	CUSTOMER 3	City3	State3
4	4	O3204	05-03-2025	5	CUSTOMER 4	City4	State4
5	5	O4395	11-02-2025	11	CUSTOMER 5	City5	State5
6	6	O6615	15-01-2025	15	CUSTOMER 6	City6	State6
7	7	O4652	12-01-2025	12	CUSTOMER 7	City7	State7
8	8	O7452	13-03-2025	13	CUSTOMER 8	City8	State8
9	9	O6438	03-02-2025	3	CUSTOMER 9	City9	State9
10	10	O2142	26-03-2025	26	CUSTOMER 10	City10	State10
11	11	O8950	22-01-2025	22	CUSTOMER 11	City11	State11
12	12	O7175	07-02-2025	7	CUSTOMER 12	City12	State12
13	13	O3284	14-03-2025	14	CUSTOMER 13	City13	State13

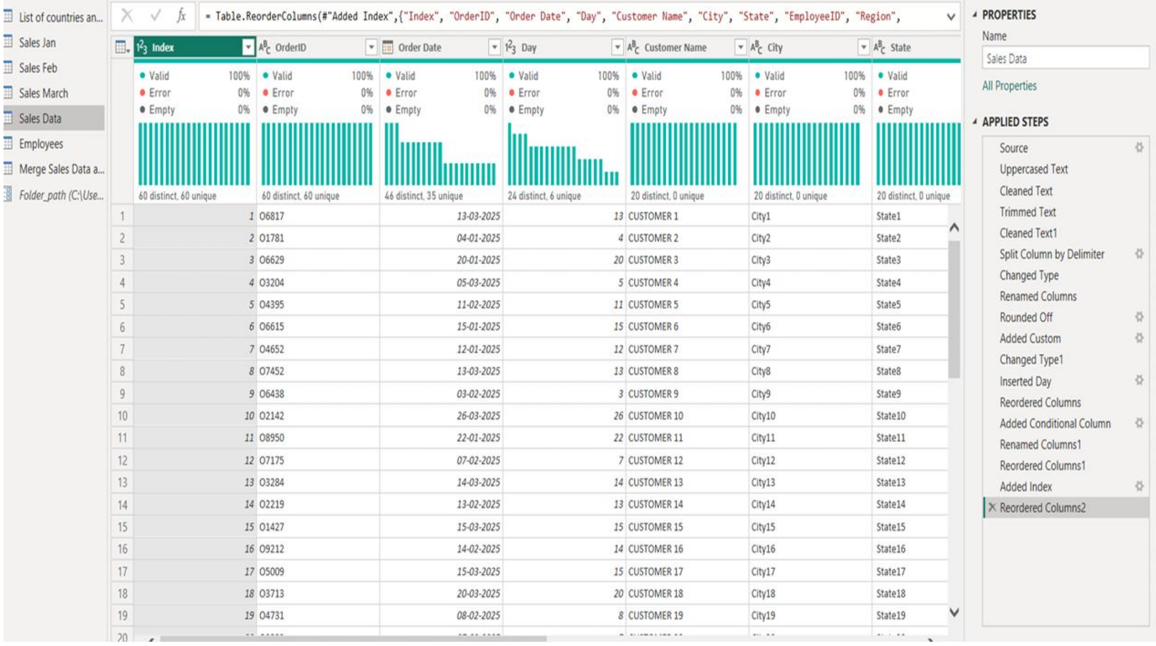
PROPERTIES

Name: Sales Data

All Properties

APPLIED STEPS

- Source
- Uppercased Text
- Cleaned Text
- Trimmed Text
- Cleaned Text1
- Split Column by Delimiter
- Changed Type
- Rename
- Rounde
- Added (
- Change
- Inserted
- Reorder
- Added (
- Rename
- Reorder
- Added I



- List of countries an...
- Sales Jan
- Sales Feb
- Sales March
- Sales Data
- Employees
- Merge Sales Data a...
- Folder_path (C:\Use...

```
= Table.ReorderColumns(#"Added Index",{"Index", "OrderID", "Order Date", "Day", "Customer Name", "City", "State", "EmployeeID", "Region",
```

	Index	OrderID	Order Date	Day	Customer Name	City	State
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid Error Empty
	60 distinct, 60 unique	60 distinct, 60 unique	46 distinct, 35 unique	24 distinct, 6 unique	20 distinct, 0 unique	20 distinct, 0 unique	20 distinct, 0 unique
1	1	O6817	13-03-2025	13	CUSTOMER 1	City1	State1
2	2	O1781	04-01-2025	4	CUSTOMER 2	City2	State2
3	3	O6629	20-01-2025	20	CUSTOMER 3	City3	State3
4	4	O3204	05-03-2025	5	CUSTOMER 4	City4	State4
5	5	O4395	11-02-2025	11	CUSTOMER 5	City5	State5
6	6	O6615	15-01-2025	15	CUSTOMER 6	City6	State6
7	7	O4652	12-01-2025	12	CUSTOMER 7	City7	State7
8	8	O7452	13-03-2025	13	CUSTOMER 8	City8	State8
9	9	O6438	03-02-2025	3	CUSTOMER 9	City9	State9
10	10	O2142	26-03-2025	26	CUSTOMER 10	City10	State10
11	11	O8950	22-01-2025	22	CUSTOMER 11	City11	State11
12	12	O7175	07-02-2025	7	CUSTOMER 12	City12	State12
13	13	O3284	14-03-2025	14	CUSTOMER 13	City13	State13
14	14	O2219	13-02-2025	13	CUSTOMER 14	City14	State14
15	15	O1427	15-03-2025	15	CUSTOMER 15	City15	State15
16	16	O9212	14-02-2025	14	CUSTOMER 16	City16	State16
17	17	O5009	15-03-2025	15	CUSTOMER 17	City17	State17
18	18	O3713	20-03-2025	20	CUSTOMER 18	City18	State18
19	19	O4731	08-02-2025	8	CUSTOMER 19	City19	State19
20							

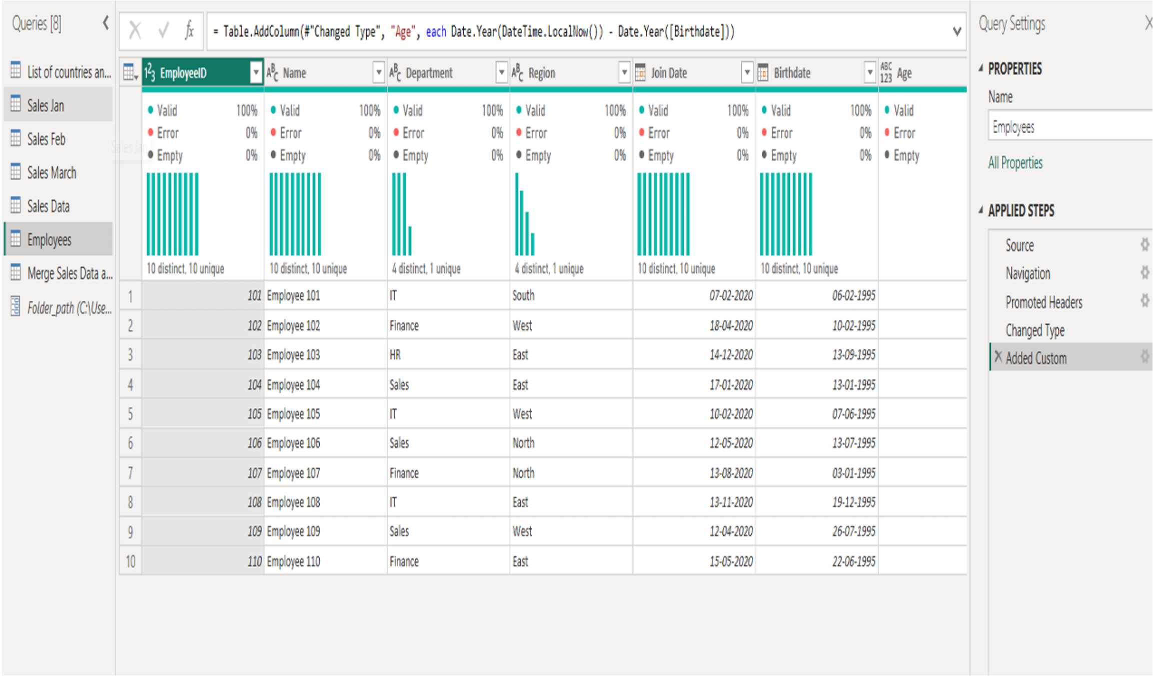
PROPERTIES

Name: Sales Data

All Properties

APPLIED STEPS

- Source
- Uppercased Text
- Cleaned Text
- Trimmed Text
- Cleaned Text1
- Split Column by Delimiter
- Changed Type
- Renamed Columns
- Rounded Off
- Added Custom
- Changed Type1
- Inserted Day
- Reordered Columns
- Added Conditional Column
- Renamed Columns1
- Reordered Columns1
- Added Index
- Reordered Columns2

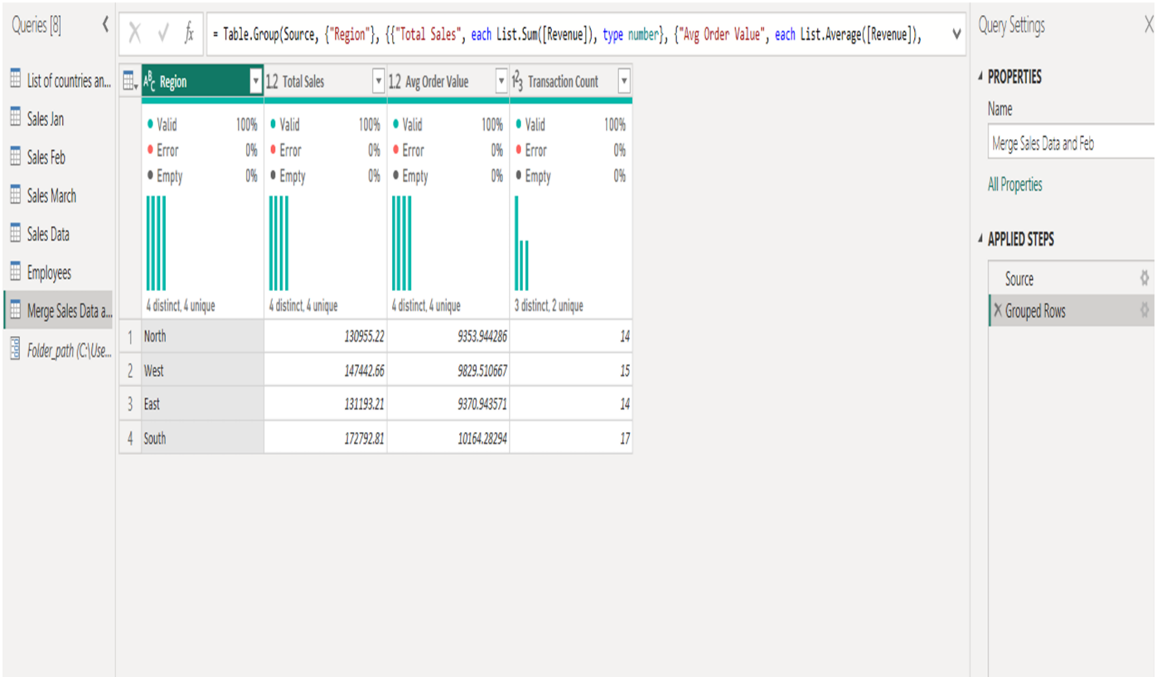


```
= Table.AddColumn(#"Changed Type", "Age", each Date.Year(DateTime.LocalNow()) - Date.Year([Birthdate]))
```

	EmployeeID	Name	Department	Region	Join Date	Birthdate	Age
1	101	Employee 101	IT	South	07-02-2020	06-02-1995	
2	102	Employee 102	Finance	West	18-04-2020	10-02-1995	
3	103	Employee 103	HR	East	14-12-2020	13-09-1995	
4	104	Employee 104	Sales	East	17-01-2020	13-01-1995	
5	105	Employee 105	IT	West	10-02-2020	07-06-1995	
6	106	Employee 106	Sales	North	12-05-2020	13-07-1995	
7	107	Employee 107	Finance	North	13-08-2020	03-01-1995	
8	108	Employee 108	IT	East	13-11-2020	19-12-1995	
9	109	Employee 109	Sales	West	12-04-2020	26-07-1995	
10	110	Employee 110	Finance	East	15-05-2020	22-06-1995	

Name: Employees

Applied steps:
- Source
- Navigation
- Promoted Headers
- Changed Type
- Added Custom

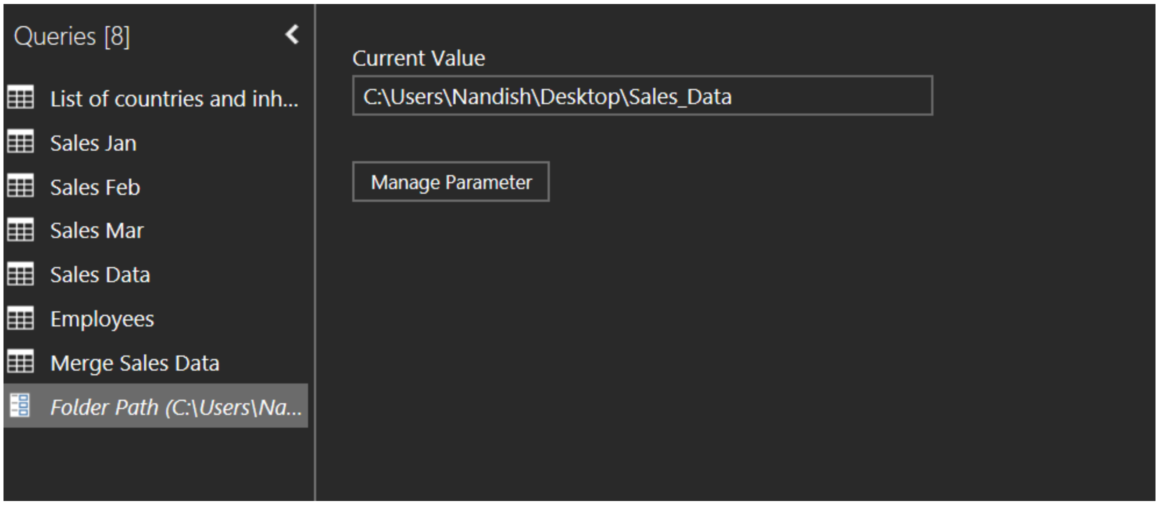


```
= Table.Group(Source, {"Region"}, {{"Total Sales", each List.Sum([Revenue]), type number}, {"Avg Order Value", each List.Average([Revenue]),
```

	Region	Total Sales	Avg Order Value	Transaction Count
1	North	130955.22	9353.944286	14
2	West	147442.66	9829.510667	15
3	East	131193.21	9370.943571	14
4	South	172792.81	10164.28294	17

Name: Merge Sales Data and Feb

Applied steps:
- Source
- Grouped Rows



Current Value

C:\Users\Nandish\Desktop\Sales_Data

Author: Nandish Patel

Course: B.Tech – Artificial Intelligence & Machine Learning

Tool Used: Power BI (Power Query)