

Portfolio Project:

Cleaning, Transforming & Loading Data in Power BI

Project Overview

This project focused on the essential processes of data cleaning, transformation, and loading using Power BI. The objective was to prepare a dataset for analysis by applying various data transformation techniques, ensuring data integrity, and creating a structured data model. The project involved working with multiple queries, each representing different aspects of a sales analysis dataset, and aimed to enhance proficiency in Power BI's Power Query Editor.

Objectives:

1. **Data Preparation:**
 - Download and set up the necessary files for the project exercise.
 - Familiarize myself with the Power BI interface and the Power Query Editor.
2. **Data Transformation:**
 - Apply various data transformation techniques, including filtering, merging, and renaming columns.
 - Create new calculated columns to handle missing values and improve data usability.
 - Ensure that the data model is concise, user-friendly, and ready for analysis.
3. **Data Loading:**
 - Load the transformed queries into the Power BI data model.
 - Verify the integrity and accuracy of the loaded data.
 - Prepare the dataset for visualization and further analysis in Power BI.
4. **Hands-on Experience:**
 - Gain practical experience in using Power BI for data cleansing and transformation.
 - Develop skills in managing queries and understanding the implications of data types and structures.
5. **Final Review:**
 - Conduct a thorough review of the final data model to ensure all queries are correctly configured and named.
 - Understand the importance of disabling unnecessary loads to optimize performance in Power BI.

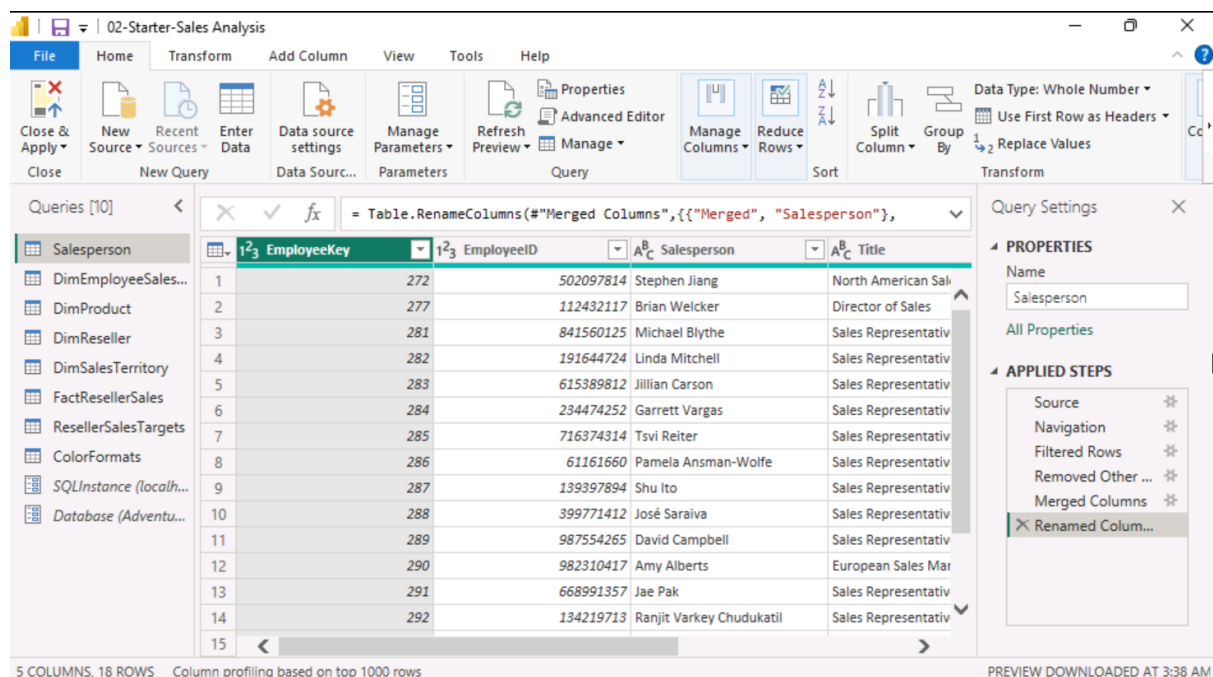
Data Preparation

- Downloaded the necessary project files from the provided GitHub link.
- Extracted the files to the designated folder on my local machine.
- Opened the 02-Starter-Sales Analysis.pbix file in Power BI Desktop.

Data Transformation

Configuring the Salesperson Query

- Opened Power Query Editor and selected the DimEmployee query.
- Renamed the query to Salesperson.
- Filtered the SalesPersonFlag column to include only TRUE values.
- Removed unnecessary columns and retained:
 - EmployeeKey
 - EmployeeNationalIDAlternateKey
 - FirstName
 - LastName
 - Title
 - EmailAddress
- Merged FirstName and LastName into a new column named Salesperson.
- Renamed EmployeeNationalIDAlternateKey to EmployeeID and EmailAddress to UPN.



02-Starter-Sales Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Columns Reduce Rows Sort Split Column Group By Data Type: Whole Number Use First Row as Headers Replace Values Transform

Queries [10]

Salesperson

123 EmployeeKey 123 EmployeeID A^BC Salesperson A^BC Title

1	272	502097814	Stephen Jiang	North American Sales Representative
2	277	112432117	Brian Welcker	Director of Sales
3	281	841560125	Michael Blythe	Sales Representative
4	282	191644724	Linda Mitchell	Sales Representative
5	283	615389812	Jillian Carson	Sales Representative
6	284	234474252	Garrett Vargas	Sales Representative
7	285	716374314	Tsvi Reiter	Sales Representative
8	286	61161660	Pamela Ansman-Wolfe	Sales Representative
9	287	139397894	Shu Ito	Sales Representative
10	288	399771412	José Saraiva	Sales Representative
11	289	987554265	David Campbell	Sales Representative
12	290	982310417	Amy Alberts	European Sales Representative
13	291	668991357	Jae Pak	Sales Representative
14	292	134219713	Ranjit Varkey Chudukatil	Sales Representative
15				

5 COLUMNS, 18 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 3:38 AM

Query Settings

PROPERTIES

Name

Salesperson

All Properties

APPLIED STEPS

Source

Navigation

Filtered Rows

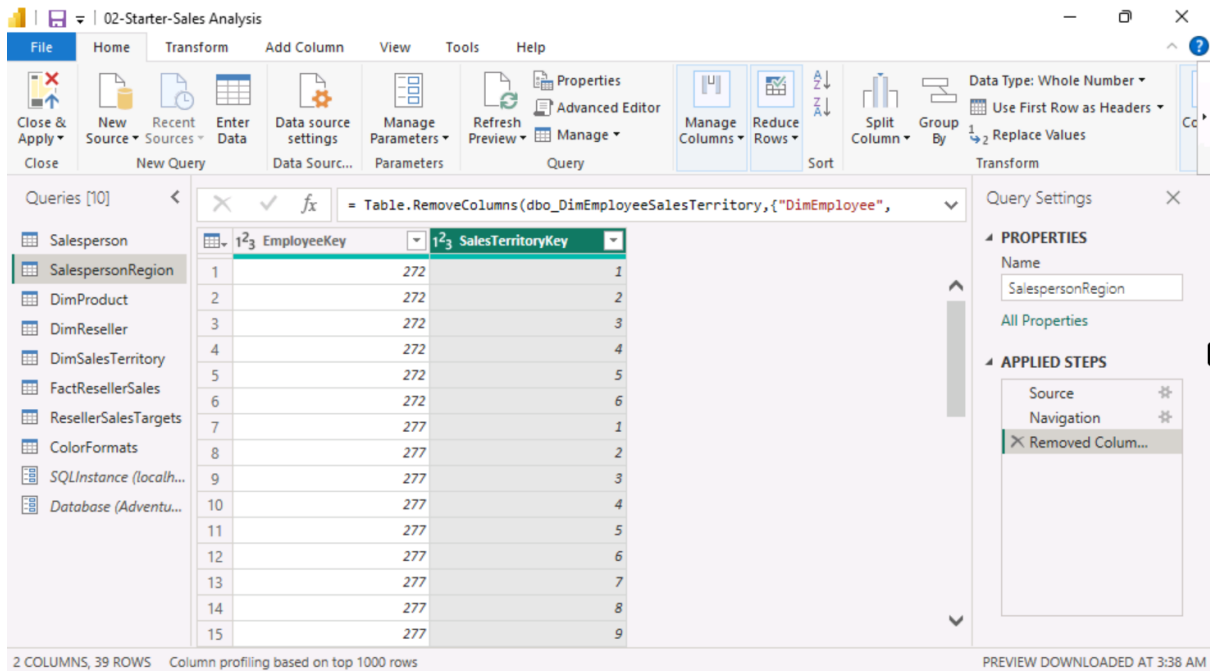
Removed Other ...

Merged Columns

Renamed Column...

Configuring the SalespersonRegion Query

- Selected the DimEmployeeSalesTerritory query and renamed it to SalespersonRegion.
- Removed the last two columns, retaining only the necessary data.
- Verified the query had 2 columns and 39 rows.



Configuring the Product Query

- Renamed the DimProduct query to Product.
- Filtered the FinishedGoodsFlag column to include only TRUE values.
- Removed all columns except:
 - ProductKey
 - EnglishProductName
 - StandardCost
 - Color
 - DimProductSubcategory
- Expanded the DimProductSubcategory column to include EnglishProductSubcategoryName and DimProductCategory.
- Renamed columns to:
 - EnglishProductName to Product
 - StandardCost to Standard Cost
 - EnglishProductSubcategoryName to Subcategory
 - EnglishProductCategoryName to Category.

02-Starter-Sales Analysis

File Home Transform Add Column View Tools Help

Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Query

Manage Columns Reduce Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Replace Values Transform

Combine Text Analytics Vision Azure Machine Learning AI Insights

Queries [10]

Salesperson SalespersonRegion Product Reseller Region Sales Targets ColorFormats SQLInstance (localh... Database (Adventu...

ProductKey Product Standard Cost Color

1	210	HL Road Frame - Black, 58	868.63	Black
2	215	Sport-100 Helmet, Black	12.03	Black
3	216	Sport-100 Helmet, Black	13.88	Black
4	217	Sport-100 Helmet, Black	13.09	Black
5	211	HL Road Frame - Red, 58	868.63	Red
6	212	Sport-100 Helmet, Red	12.03	Red
7	213	Sport-100 Helmet, Red	13.88	Red
8	214	Sport-100 Helmet, Red	13.09	Red
9	220	Sport-100 Helmet, Blue	12.03	Blue
10	218	Mountain Bike Socks, M	3.40	White
11	219	Mountain Bike Socks, L	3.40	White
12	221	Sport-100 Helmet, Blue	13.88	Blue
13	222	Sport-100 Helmet, Blue	13.09	Blue
14	223	AWC Logo Cap	5.71	Multi
15				

8 COLUMNS, 397 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 4:10 AM

Query Settings

PROPERTIES

Name Product

All Properties

APPLIED STEPS

- Source
- Navigation
- Filtered Rows
- Removed Other ...
- Expanded DimPr...
- Expanded DimPr...
- Renamed Colum...
- Merged Queries
- Expanded Color...

Configuring the Reseller Query

- Renamed the DimReseller query to Reseller.
- Retained only the necessary columns:
 - ResellerKey
 - BusinessType
 - ResellerName
 - DimGeography
- Expanded the DimGeography column to include:
 - City
 - StateProvinceName
 - EnglishCountryRegionName
- Replaced values in the BusinessType column to standardize naming.
- Renamed columns to improve clarity.

02-Starter-Sales Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Columns Reduce Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Replace Values

Queries [10]

- Salesperson
- SalespersonRegion
- Product
- Reseller**
- DimSalesTerritory
- FactResellerSales
- ResellerSalesTargets
- ColorFormats
- SQLInstance (localh...
- Database (Adventu...

123 ResellerKey Business Type Reseller City

1	2	3	4
1	105	Warehouse	Volume Bike Sellers
2	658	Value Added Reseller	Mass Market Bikes
3	555	Value Added Reseller	Twin Cycles
4	87	Warehouse	Rich Department Store
5	141	Specialty Bike Shop	Rental Gallery
6	15	Warehouse	Budget Toy Store
7	33	Warehouse	Global Sports Outlet
8	69	Warehouse	Online Bike Catalog
9	231	Value Added Reseller	Helmets and Cycles
10	501	Specialty Bike Shop	Jumbo Bikes
11	321	Value Added Reseller	Second Bike Shop
12	339	Warehouse	Major Bicycle Store
13	616	Specialty Bike Shop	Eastward Bike Accessories
14	393	Warehouse	Gears and Parts Company
15			

6 COLUMNS, 701 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 3:49 AM

Query Settings

PROPERTIES

Name: Reseller

APPLIED STEPS

- Source
- Navigation
- Removed Other ...
- Expanded DimG...
- Replaced Value
- Renamed Column...**

Configuring the Region Query

- Renamed the DimSalesTerritory query to Region.
- Filtered the SalesTerritoryAlternateKey column to remove zero values.
- Retained only essential columns and renamed them for clarity.

02-Starter-Sales Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Columns Reduce Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Replace Values

Queries [10]

- Salesperson
- SalespersonRegion
- Product
- Reseller
- Region**
- FactResellerSales
- ResellerSalesTargets
- ColorFormats
- SQLInstance (localh...
- Database (Adventu...

123 SalesTerritoryKey Region Country Group

1	2	3	4
1	1	Northwest	United States
2	2	Northeast	United States
3	3	Central	United States
4	4	Southwest	United States
5	5	Southeast	United States
6	6	Canada	Canada
7	7	France	France
8	8	Germany	Germany
9	9	Australia	Australia
10	10	United Kingdom	United Kingdom

4 COLUMNS, 10 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 3:51 AM

Query Settings

PROPERTIES

Name: Region

APPLIED STEPS

- Source
- Navigation
- Filtered Rows
- Removed Other ...
- Renamed Column...**

Configuring the Sales Query

- Renamed the FactResellerSales query to Sales.
- Removed unnecessary columns, retaining key sales data.
- Expanded the DimProduct column to include StandardCost.
- Created a custom column Cost to handle missing TotalProductCost values.
- Renamed columns for better understanding and modified data types for accuracy.

The screenshot shows the Microsoft Power BI Desktop interface. The main window displays a table with the following data:

	AB SalesOrderNumber	OrderDate	1 ² 3 ProductKey	1 ² 3 ResellerKey
1	SO43659	7/1/2017		349
2	SO43659	7/1/2017		350
3	SO43659	7/1/2017		351
4	SO43659	7/1/2017		344
5	SO43659	7/1/2017		345
6	SO43659	7/1/2017		346
7	SO43659	7/1/2017		347
8	SO43659	7/1/2017		229
9	SO43659	7/1/2017		235
10	SO43659	7/1/2017		218
11	SO43659	7/1/2017		223
12	SO43659	7/1/2017		220
13	SO43660	7/1/2017		326
14	SO43660	7/1/2017		319
15				

The bottom status bar indicates: 10 COLUMNS, 999+ ROWS. Column profiling based on top 1000 rows. PREVIEW DOWNLOADED AT 3:59 AM.

Configuring the Targets Query

- Renamed the ResellerSalesTargets query to Targets.
- Unpivoted the 12-month columns to create a more usable format.
- Cleaned the Value column to remove hyphen values.
- Created a new date column TargetMonth from Year and MonthNumber.
- Adjusted data types for accuracy and multiplied target values by 1000.

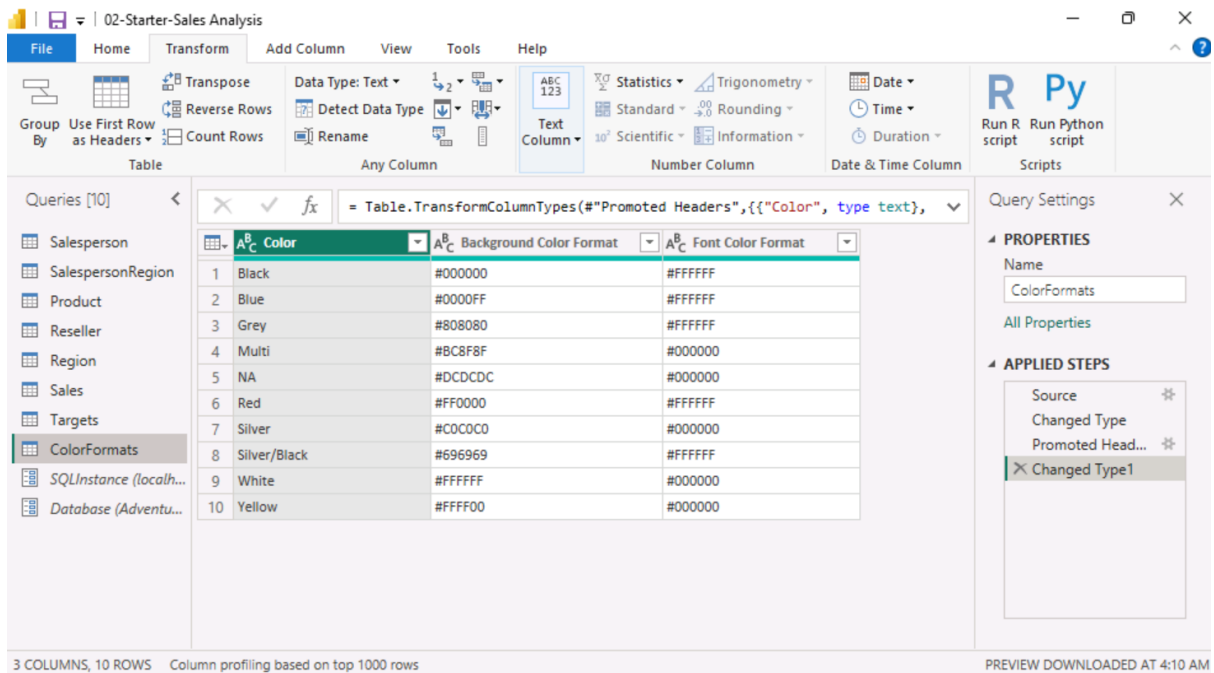
The screenshot shows the Power BI Desktop interface with the 'Targets' query selected. The main view displays a table with the following data:

	EmployeeID	Target	TargetMonth
1	61161660	200,000.00	7/1/2017
2	61161660	400,000.00	8/1/2017
3	61161660	600,000.00	9/1/2017
4	61161660	400,000.00	10/1/2017
5	61161660	800,000.00	11/1/2017
6	61161660	800,000.00	12/1/2017
7	90836195	100,000.00	7/1/2017
8	90836195	200,000.00	8/1/2017
9	90836195	300,000.00	9/1/2017
10	90836195	400,000.00	10/1/2017
11	90836195	400,000.00	11/1/2017
12	90836195	500,000.00	12/1/2017
13	112432117	500,000.00	7/1/2017
14	112432117	1,500,000.00	8/1/2017
15	112432117	1,000,000.00	9/1/2017

The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list, which includes: Changed Type, Unpivoted Ot..., Filtered Rows, Renamed Col..., Replaced Val..., Changed Typ..., Inserted Mer..., Removed Col..., Changed Typ..., and Multiplied Co... The status bar at the bottom indicates '3 COLUMNS, 809 ROWS' and 'Column profiling based on top 1000 rows'.

Configuring the ColorFormats Query

- Selected the ColorFormats query and set the first row as headers.
- Updated the Product query by merging with the ColorFormats query.
- Expanded the merged column to include color format details.



Data Loading & Final Review

Final Review and Load

- Verified that all queries were correctly named and configured.
- Disabled load for the ColorFormats query as it was merged with the Product query.
- Closed and applied changes to load the data model in Power BI.

Results

- ✓ Successfully transformed and loaded 7 tables into the Power BI data model.
- ✓ Gained hands-on experience in data cleansing, transformation, and loading processes using Power BI.

Conclusion

This project provided valuable insights into the data preparation process, which is crucial for effective data analysis and reporting.

Resources

Source file: <https://github.com/MicrosoftLearning/PL-300-Microsoft-Power-BI-Data-Analyst/blob/Main/Allfiles/Labs/02-transform-data-power-bi/02-Starter-Sales%20Analysis.pbix>

GitHub: <https://github.com/ThatoMTNG/Microsoft-Power-BI-Data-Analyst-PL-300>

MENTIONS

Project Author: Thato Metsing (<https://www.linkedin.com/in/thatometsing/>)

Project Mentor: Maureen Direro (<https://www.linkedin.com/in/maureen-direro-46a6b220/>)