Portfolio Project:

Creating a Microsoft Fabric Lakehouse

Project Overview

This project focused on creating a Microsoft Fabric lakehouse, which combines the features of data lakes and data warehouses for large-scale data analytics. The objective was to set up a lakehouse that allows for scalable file storage and querying using SQL semantics.

Objectives:

1. **Create a Workspace:**
   - Create a workspace with the Fabric trial enabled.
2. **Create a Lakehouse:**
   - Set up a data lakehouse for your data files.
3. **Upload a File:**
   - Load data into the lakehouse by uploading files from your local computer.
4. **Explore Shortcuts:**
   - Understand how to create shortcuts for externally sourced data.
5. **Load File Data into a Table:**
   - Load the uploaded file data into a table for SQL querying.
6. **Use SQL to Query Tables:**
   - Execute SQL queries to analyze the data in the lakehouse.
7. **Create a Visual Query:**
   - Utilize visual query tools to analyze data without SQL.

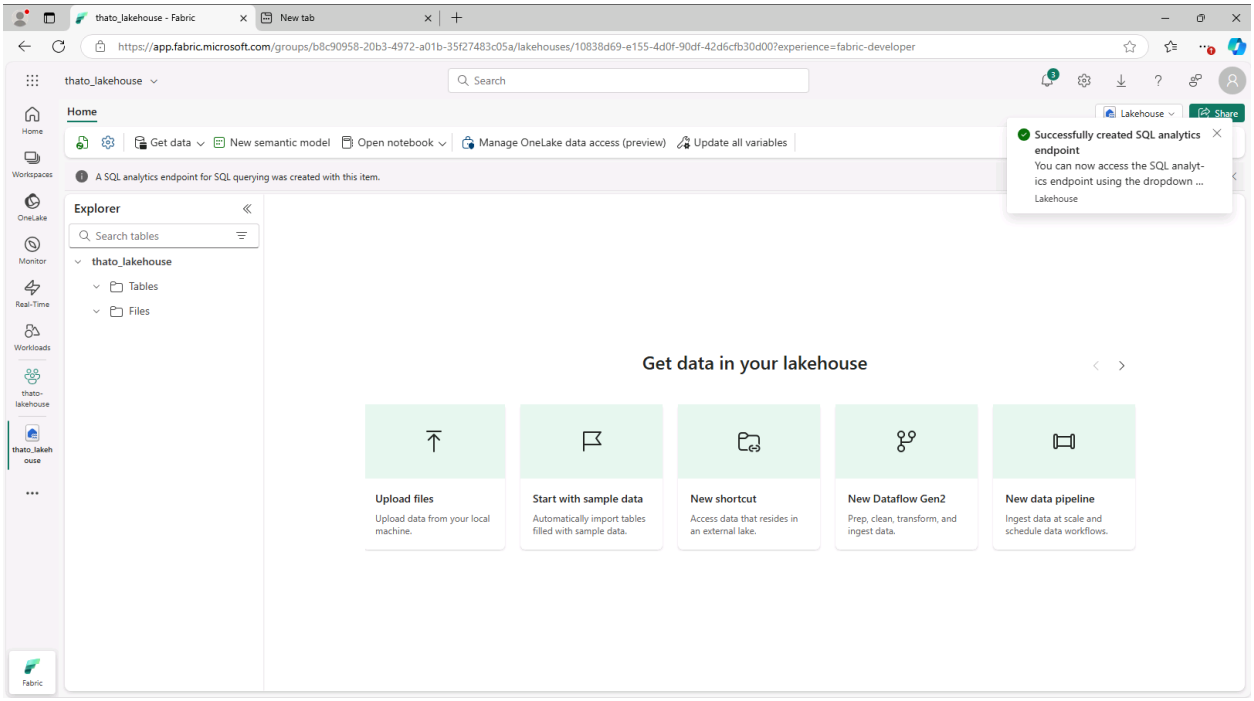
Experience

Create a Workspace

- Navigated to the Microsoft Fabric home page and signed in with credentials.
- Selected Workspaces from the menu bar.
- Created a new workspace with a name.

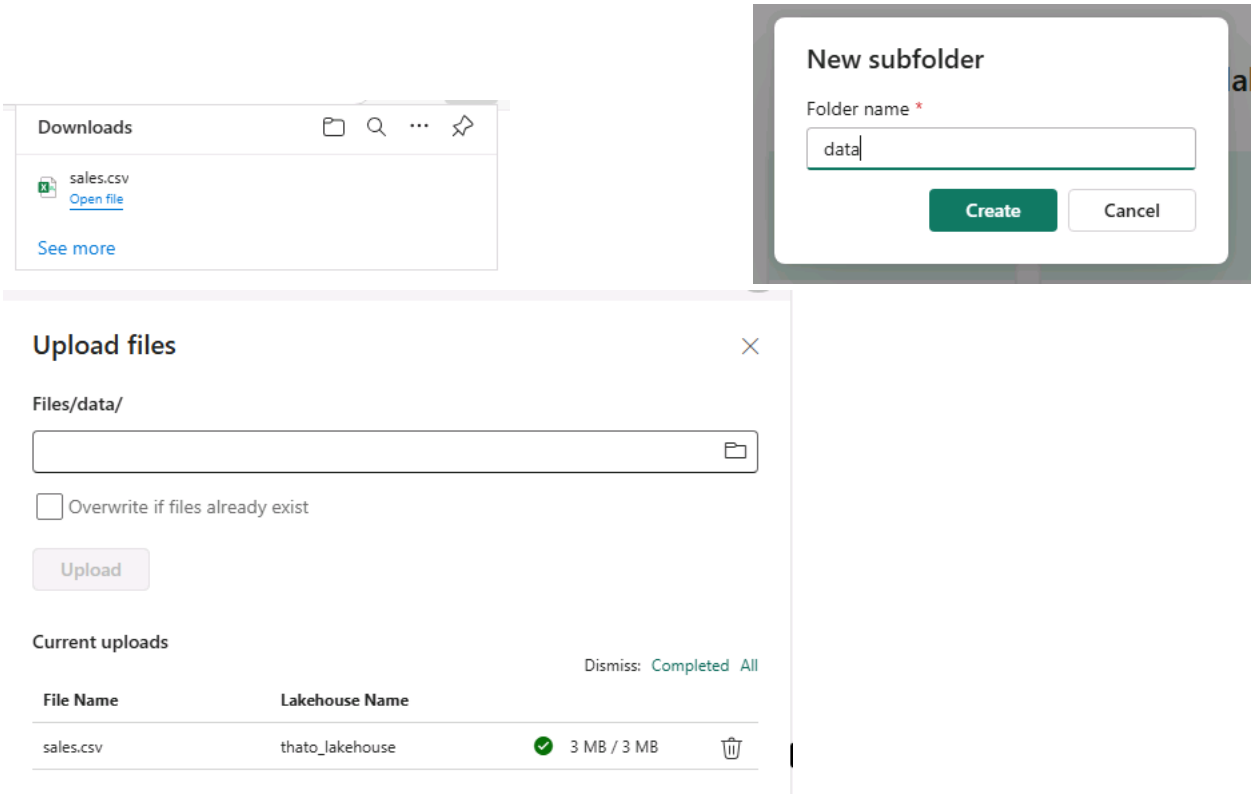
Create a Lakehouse

- In the workspace, selected Create from the menu bar.
- Under the Data Engineering section, selected Lakehouse and gave it a unique name.
- Waited for the lakehouse to be created.



Upload a File:

- Downloaded the sales.csv file from the computer.
- In the lakehouse, created a subfolder named data within the Files folder.
- Uploaded the sales.csv file into the newly created data folder.

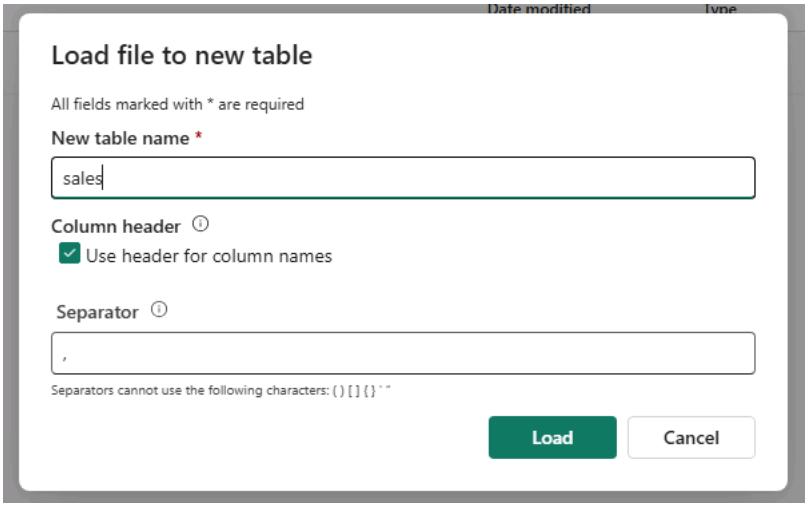


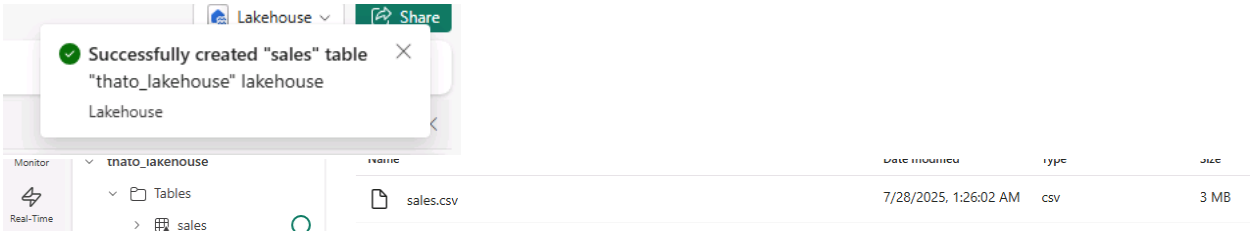
Explore Shortcuts:

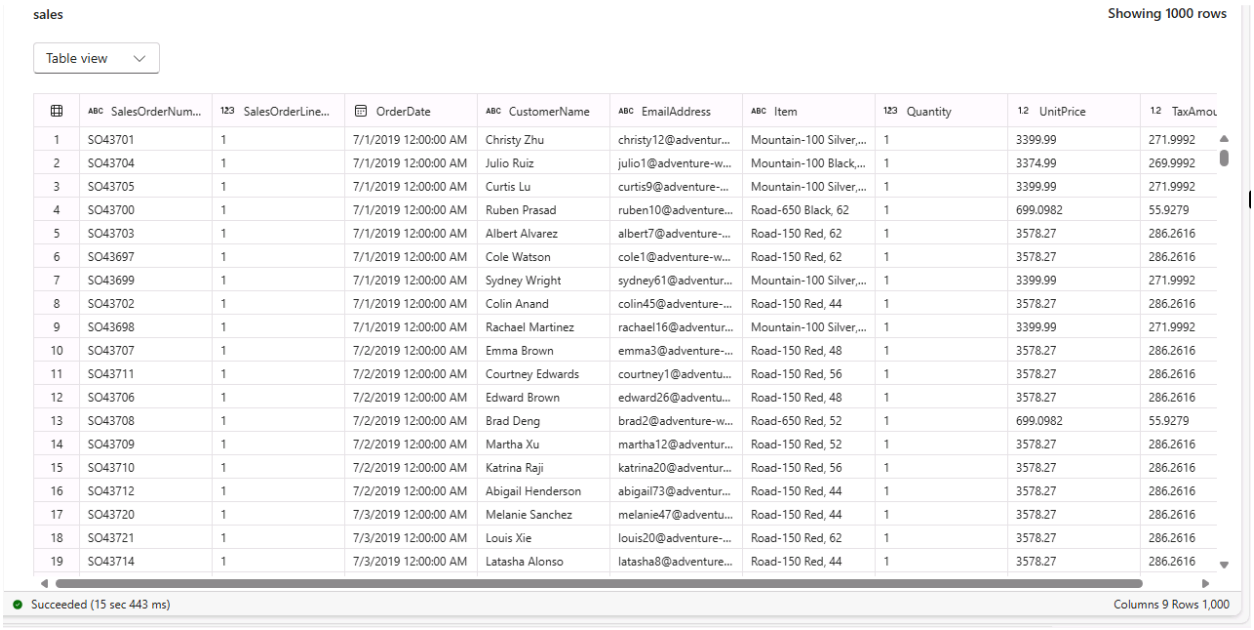
- Created a shortcut in the Files folder to include externally sourced data without copying it.

Load File Data into a Table:

- Selected the sales.csv file and chose Load to Tables > New table.
- Set the table name to sales and confirmed the load operation.
- Refreshed the Tables folder.



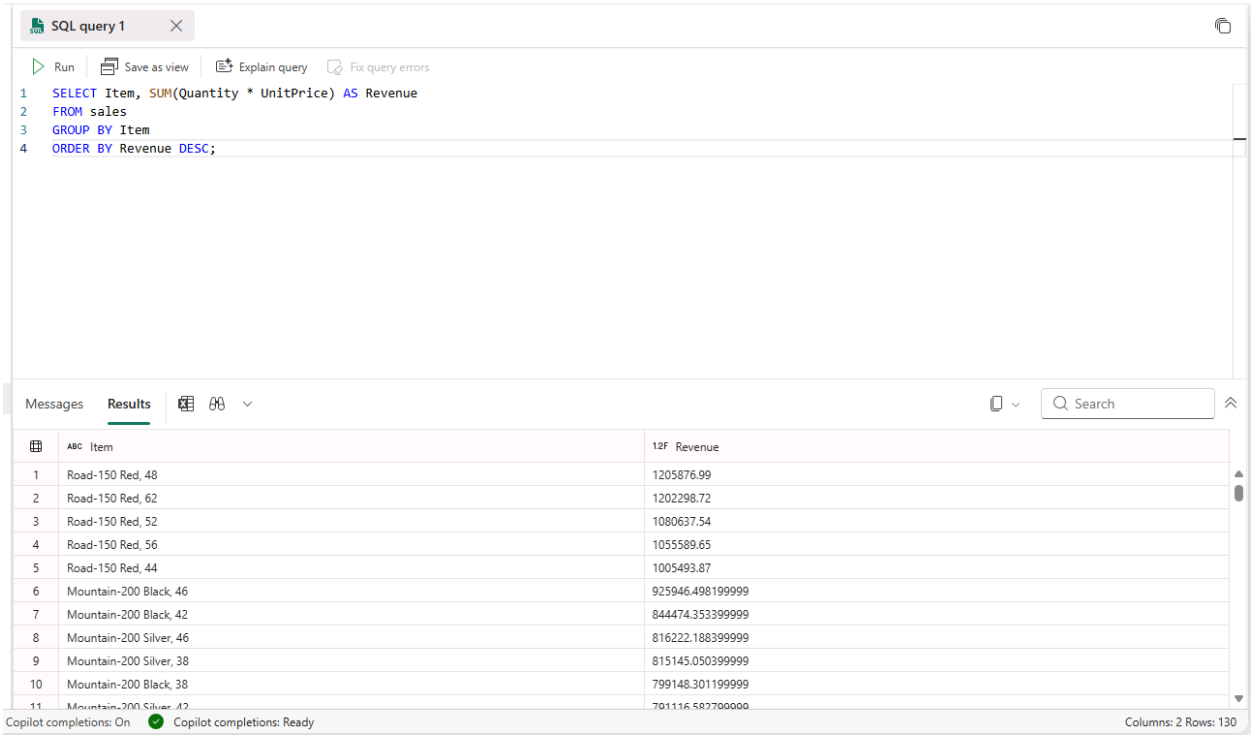




	SalesOrderNum...	SalesOrderLine...	OrderDate	CustomerName	EmailAddress	Item	Quantity	UnitPrice	TaxAmou
1	SO43701	1	7/1/2019 12:00:00 AM	Christy Zhu	christy12@adventur...	Mountain-100 Silver,...	1	3399.99	271.9992
2	SO43704	1	7/1/2019 12:00:00 AM	Julio Ruiz	julio1@adventure-w...	Mountain-100 Black,...	1	3374.99	269.9992
3	SO43705	1	7/1/2019 12:00:00 AM	Curtis Lu	curtis9@adventure-...	Mountain-100 Silver,...	1	3399.99	271.9992
4	SO43700	1	7/1/2019 12:00:00 AM	Ruben Prasad	ruben10@adventure...	Road-650 Black, 62	1	699.0982	55.9279
5	SO43703	1	7/1/2019 12:00:00 AM	Albert Alvarez	albert7@adventure-...	Road-150 Red, 62	1	3578.27	286.2616
6	SO43697	1	7/1/2019 12:00:00 AM	Cole Watson	cole1@adventure-w...	Road-150 Red, 62	1	3578.27	286.2616
7	SO43699	1	7/1/2019 12:00:00 AM	Sydney Wright	sydney61@adventur...	Mountain-100 Silver,...	1	3399.99	271.9992
8	SO43702	1	7/1/2019 12:00:00 AM	Colin Anand	colin45@adventure-...	Road-150 Red, 44	1	3578.27	286.2616
9	SO43698	1	7/1/2019 12:00:00 AM	Rachael Martinez	rachael16@adventur...	Mountain-100 Silver,...	1	3399.99	271.9992
10	SO43707	1	7/2/2019 12:00:00 AM	Emma Brown	emma3@adventure-...	Road-150 Red, 48	1	3578.27	286.2616
11	SO43711	1	7/2/2019 12:00:00 AM	Courtney Edwards	courtney1@adventu...	Road-150 Red, 56	1	3578.27	286.2616
12	SO43706	1	7/2/2019 12:00:00 AM	Edward Brown	edward26@adventu...	Road-150 Red, 48	1	3578.27	286.2616
13	SO43708	1	7/2/2019 12:00:00 AM	Brad Deng	brad2@adventure-w...	Road-650 Red, 52	1	699.0982	55.9279
14	SO43709	1	7/2/2019 12:00:00 AM	Martha Xu	martha12@adventur...	Road-150 Red, 52	1	3578.27	286.2616
15	SO43710	1	7/2/2019 12:00:00 AM	Katrina Raji	katrina20@adventur...	Road-150 Red, 56	1	3578.27	286.2616
16	SO43712	1	7/2/2019 12:00:00 AM	Abigail Henderson	abigail73@adventur...	Road-150 Red, 44	1	3578.27	286.2616
17	SO43720	1	7/3/2019 12:00:00 AM	Melanie Sanchez	melanie47@adventu...	Road-150 Red, 44	1	3578.27	286.2616
18	SO43721	1	7/3/2019 12:00:00 AM	Louis Xie	louis20@adventure-...	Road-150 Red, 62	1	3578.27	286.2616
19	SO43714	1	7/3/2019 12:00:00 AM	Latasha Alonso	latasha8@adventure...	Road-150 Red, 44	1	3578.27	286.2616

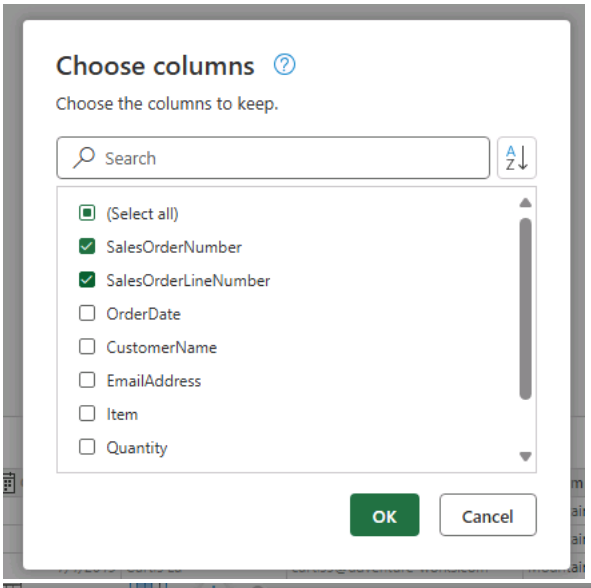
Use SQL to Query Tables:

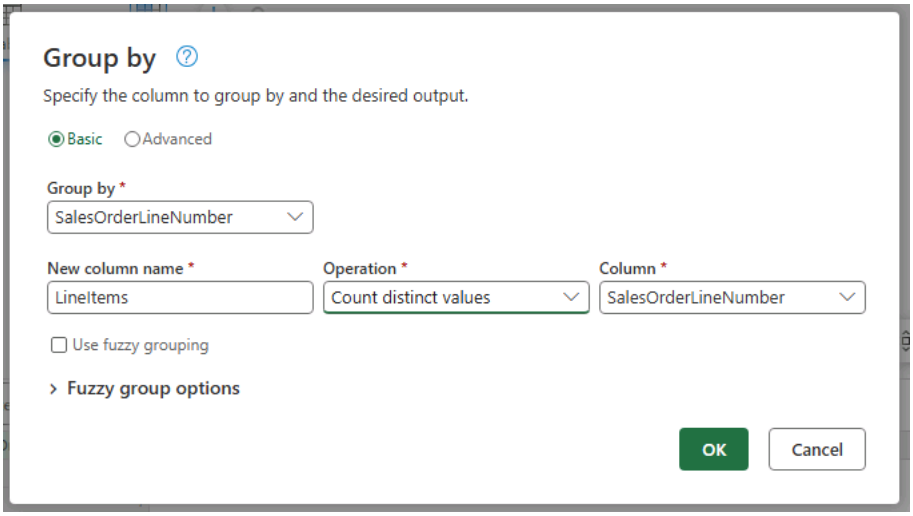
- Switched to the SQL analytics endpoint for the lakehouse.
- Opened a new SQL query editor and entered the SQL query to calculate total revenue for each product.
- Ran the query and viewed the results.

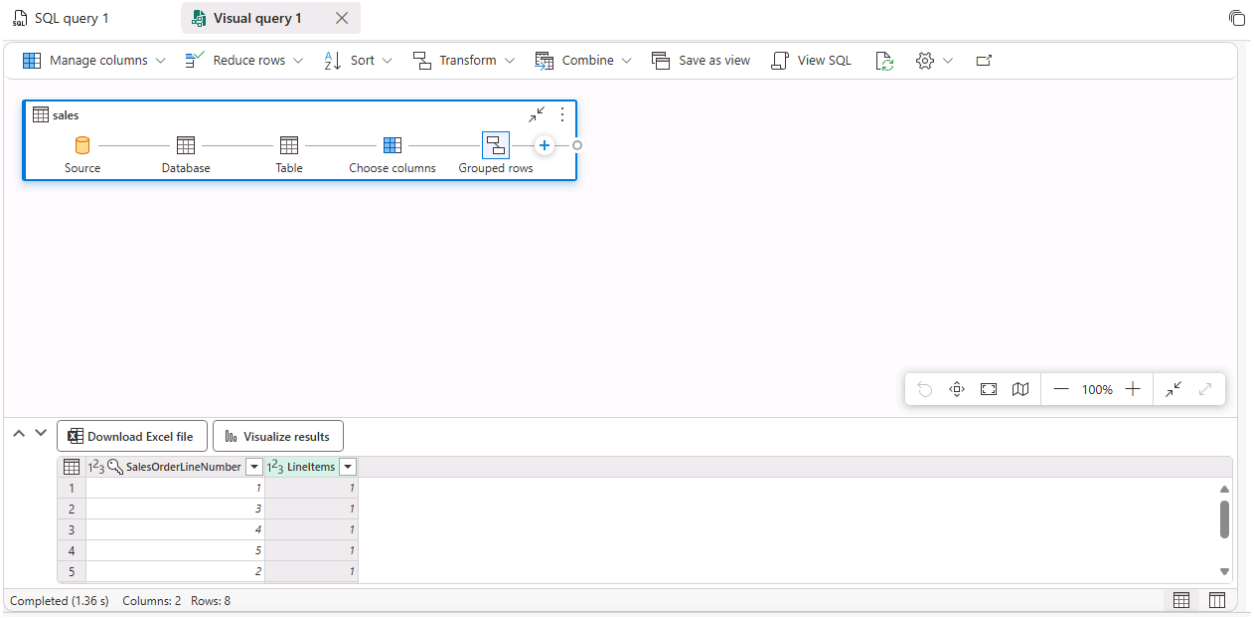


Create a Visual Query:

- Used the New visual query option to create a Power Query.
- Selected the sales table and managed columns to choose specific columns.
- Grouped the data by SalesOrderNumber to count distinct values of SalesOrderLineNumber.


Choose columns
Choose the columns to keep.
Search
(Select all)
SalesOrderNumber
SalesOrderLineNumber
OrderDate
CustomerName
EmailAddress
Item
Quantity
OK
Cancel


Group by
Specify the column to group by and the desired output.
Basic
Advanced
Group by *
SalesOrderLineNumber
New column name *
LineItems
Operation *
Count distinct values
Column *
SalesOrderLineNumber
Use fuzzy grouping
Fuzzy group options
OK
Cancel


SQL query 1
Visual query 1
Manage columns
Reduce rows
Sort
Transform
Combine
Save as view
View SQL
sales
Source
Database
Table
Choose columns
Grouped rows
100%
Download Excel file
Visualize results
SalesOrderLineNumber
LineItems
1 1 1
2 3 1
3 4 1
4 5 1
5 2 1
Completed (1.36 s) Columns: 2 Rows: 8

Results

- ✓ Successfully created a Microsoft Fabric lakehouse and uploaded the sales.csv file.
- ✓ Loaded the sales data into a table, enabling SQL queries for data analysis.
- ✓ Executed SQL queries to calculate total revenue and created visual queries to analyze sales order data.
- ✓ Gained practical experience in managing data within a lakehouse environment and utilizing Microsoft Fabric's features for data analytics.

Conclusion

This project provided hands-on experience in creating a Microsoft Fabric lakehouse, uploading data, querying it using SQL, and creating visual queries.

Resources

Source file:

https://raw.githubusercontent.com/MicrosoftLearning/dp-data/main/sales.csv

GitHub: https://github.com/ThatoMTNG/Microsoft-Fabric-Analytics-Engineer-DP-600-

Mentions

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

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