

Modern Data Warehouse Final Project

**Tejas Padavalamane
A20543726**

Introduction to Data Warehousing in Microsoft Fabric

The Medallion architecture encapsulates core principles such as scalability, modularity, and performance enhancement. It achieves this through a structured layering of data processing tasks, namely Bronze, Silver, and Gold. Each layer serves a distinct purpose in the data pipeline:

- The Bronze layer acts as the initial repository for raw data, aggregating it from diverse sources in its unprocessed state.
- Moving forward, the Silver layer refines and cleanses the data, standardizing it for consistency and reliability.
- Finally, the Gold layer houses curated and processed data, optimized for analytical queries and real-time insights generation.

By adhering to this layered approach, the Medallion architecture provides a systematic framework for ingesting, transforming, and analyzing data, thereby facilitating robust data management and analytics workflows.

Tools Used:

Microsoft fabric services:

- Lakehouse
- Synapse Data Engineering
- Notebook
- Schematic model
- Power BI
- Data Pipeline

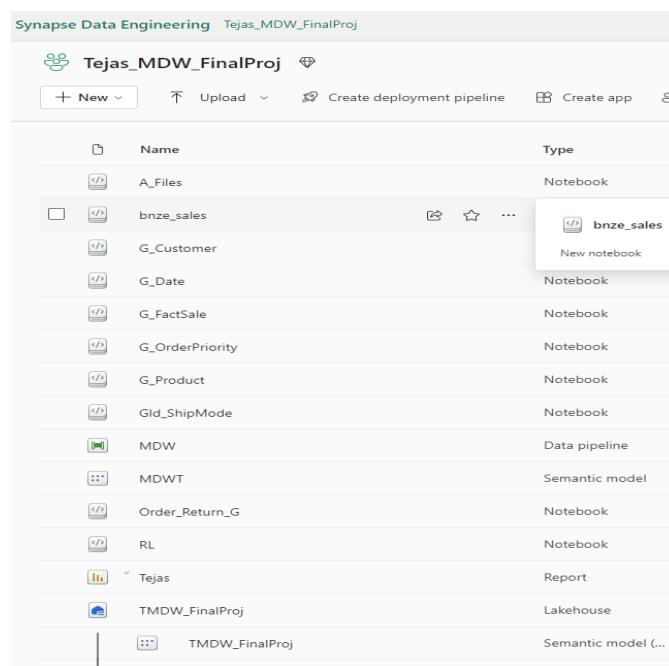
Data Source:

For this project, we utilized data from multiple sources:

1. Sales Data: This includes information about sales transactions such as order details, customer information, product details, and shipping details.
2. Order Priority Data: Information about the priority of orders.
3. Order Return Data: Information about returned orders.
4. Product Data: Information about products.
5. Ship Mode Data: Information about shipping modes.
6. Customer Data: Contains details about customer



| ARK_OID | Order_Date | Ship_Date | Aging | Ship_Mode | Product_Cat... | Product | Sales | Quantity | Discount | Profit | Ship_Cost | Order_Pric |
|---------------|--------------------|--------------------|-------|----------------|----------------|---------------|-------|----------|----------|--------|-----------|------------|
| FA-2015-1015 | 1/13/2023 12:00... | 1/18/2023 12:00... | 5 | Standard Class | Fashion | Jeans | 218 | 1 | 0.03 | 131.46 | 13.146 | Medium |
| FA-2015-2012 | 1/30/2023 12:00... | 2/5/2023 12:00... | 6 | Standard Class | Fashion | Titak watch | 228 | 1 | 0.02 | 143.44 | 14.344 | Medium |
| FA-2015-417 | 1/12/2023 12:00... | 1/20/2023 12:00... | 8 | Standard Class | Fashion | Titak watch | 228 | 1 | 0.01 | 145.72 | 14.572 | Medium |
| FA-2015-688 | 1/28/2023 12:00... | 1/30/2023 12:00... | 2 | Standard Class | Fashion | Casula Shoes | 122 | 1 | 0.03 | 38.34 | 3.834 | Medium |
| FA-2015-3096 | 1/6/2023 12:00... | 1/9/2023 12:00... | 3 | Standard Class | Fashion | Sports Wear | 85 | 1 | 0.01 | 4.15 | 0.415 | Medium |
| FA-2015-1138 | 1/2/2023 12:00... | 1/9/2023 12:00... | 7 | Standard Class | Fashion | Sports Wear | 85 | 1 | 0.02 | 3.3 | 0.33 | Medium |
| FA-2015-373 | 1/18/2023 12:00... | 1/24/2023 12:00... | 6 | Standard Class | Fashion | Titak watch | 228 | 1 | 0.05 | 136.6 | 13.66 | Medium |
| FA-2015-3080 | 1/6/2023 12:00... | 1/10/2023 12:00... | 4 | Standard Class | Fashion | Fossil Watch | 159 | 1 | 0.05 | 71.05 | 7.105 | Medium |
| FA-2015-5527 | 1/12/2023 12:00... | 1/14/2023 12:00... | 2 | Standard Class | Fashion | Sports Wear | 85 | 1 | 0.01 | 4.15 | 0.415 | Medium |
| FA-2015-8471 | 1/21/2023 12:00... | 1/25/2023 12:00... | 4 | Standard Class | Fashion | T - Shirts | 248 | 1 | 0.04 | 158.08 | 15.808 | Medium |
| FA-2015-9751 | 1/30/2023 12:00... | 2/5/2023 12:00... | 6 | Standard Class | Fashion | Sports Wear | 85 | 1 | 0.04 | 1.6 | 0.16 | Medium |
| FA-2015-9018 | 1/24/2023 12:00... | 1/30/2023 12:00... | 6 | Standard Class | Fashion | Sneakers | 62 | 1 | 0.05 | 62 | 6.2 | Medium |
| FA-2015-8500 | 1/12/2023 12:00... | 1/13/2023 12:00... | 1 | Standard Class | Fashion | Formal Shoes | 213 | 1 | 0.01 | 130.87 | 13.087 | Medium |
| FA-2015-6319 | 1/11/2023 12:00... | 1/13/2023 12:00... | 2 | Standard Class | Fashion | Sports Wear | 85 | 1 | 0.02 | 3.3 | 0.33 | Medium |
| FA-2015-9934 | 1/3/2023 12:00... | 1/5/2023 12:00... | 2 | Standard Class | Fashion | T - Shirts | 248 | 1 | 0.04 | 158.08 | 15.808 | Medium |
| FA-2015-13835 | 1/10/2023 12:00... | 1/19/2023 12:00... | 9 | Standard Class | Fashion | Formal Shoes | 213 | 1 | 0.03 | 126.61 | 12.661 | Medium |
| FA-2015-13277 | 1/3/2023 12:00... | 1/6/2023 12:00... | 3 | Standard Class | Fashion | Fossil Watch | 159 | 1 | 0.03 | 74.23 | 7.423 | Medium |
| FA-2015-12471 | 1/11/2023 12:00... | 1/14/2023 12:00... | 3 | Standard Class | Fashion | Formal Shoes | 213 | 1 | 0.03 | 126.61 | 12.661 | Medium |
| FA-2015-17028 | 1/18/2023 12:00... | 1/27/2023 12:00... | 9 | Standard Class | Fashion | Fossil Watch | 159 | 1 | 0.02 | 75.82 | 7.582 | Medium |
| FA-2015-16678 | 1/26/2023 12:00... | 1/28/2023 12:00... | 2 | Standard Class | Fashion | Shirts | 196 | 1 | 0.02 | 112.08 | 11.208 | Medium |
| FA-2015-12581 | 1/25/2023 12:00... | 2/1/2023 12:00... | 7 | Standard Class | Fashion | Formal Shoes | 213 | 1 | 0.02 | 128.74 | 12.874 | Medium |
| FA-2015-16918 | 1/22/2023 12:00... | 1/30/2023 12:00... | 8 | Standard Class | Fashion | Fossil Watch | 159 | 1 | 0.04 | 72.64 | 7.264 | Medium |
| FA-2015-22603 | 1/30/2023 12:00... | 1/31/2023 12:00... | 1 | Standard Class | Fashion | Formal Shoes | 213 | 1 | 0.03 | 126.61 | 12.661 | Medium |
| FA-2015-23531 | 1/20/2023 12:00... | 1/28/2023 12:00... | 8 | Standard Class | Fashion | T - Shirts | 248 | 1 | 0.01 | 165.52 | 16.552 | Medium |
| FA-2015-24170 | 1/15/2023 12:00... | 1/17/2023 12:00... | 2 | Standard Class | Fashion | Shirts | 196 | 1 | 0.01 | 114.04 | 11.404 | Medium |
| FA-2015-18344 | 1/28/2023 12:00... | 2/1/2023 12:00... | 4 | Standard Class | Fashion | Running Shoes | 224 | 1 | 0.04 | 135.04 | 13.504 | Medium |
| FA-2015-20918 | 1/17/2023 12:00... | 1/24/2023 12:00... | 7 | Standard Class | Fashion | Running Shoes | 224 | 1 | 0.04 | 135.04 | 13.504 | Medium |

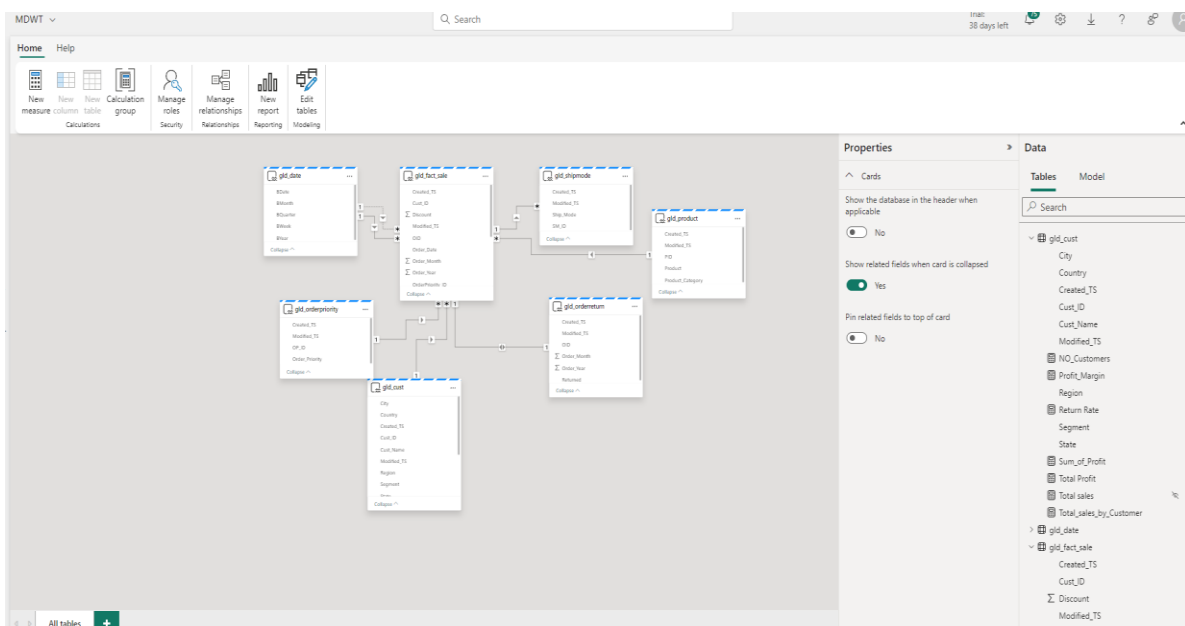


| Name | Type |
|-----------------|----------------------|
| A_Files | Notebook |
| bnze_sales | Notebook |
| G_Customer | Notebook |
| G_Date | Notebook |
| G_FactSale | Notebook |
| G_OrderPriority | Notebook |
| G_Product | Notebook |
| Gld_ShipMode | Notebook |
| MDW | Data pipeline |
| MDWT | Semantic model |
| Order_Return_G | Notebook |
| RL | Notebook |
| Tejas | Report |
| TMDW_FinalProj | Lakehouse |
| TMDW_FinalProj | Semantic model (...) |

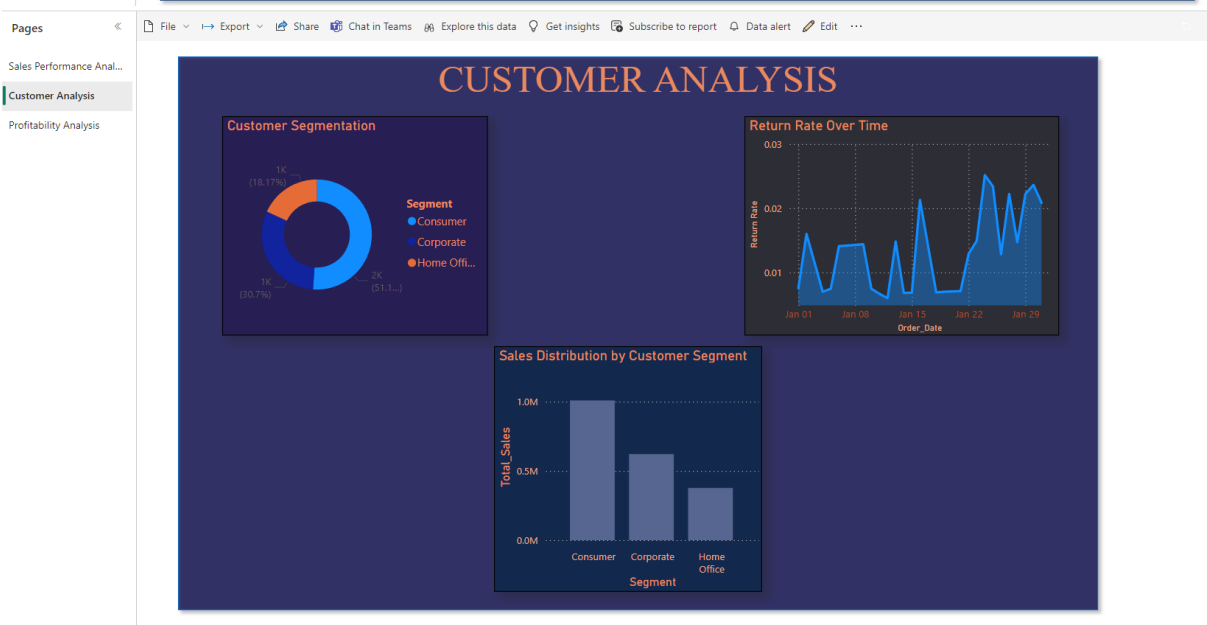
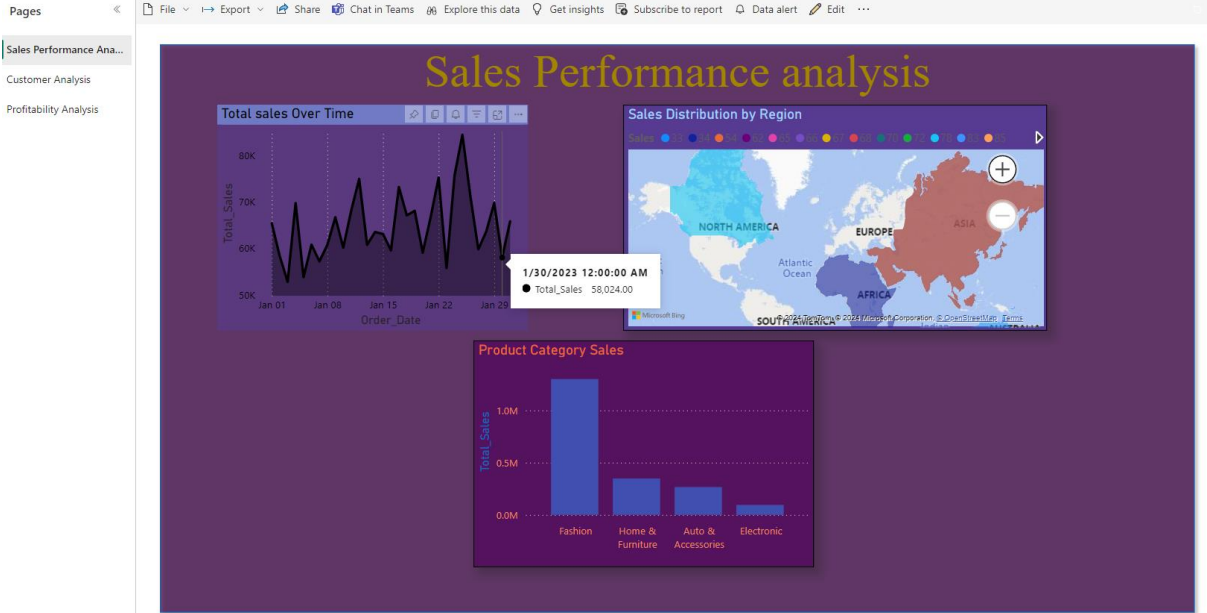
| Synapse Data Engineering Tejas_MDW_FinalProj | | |
|---|----------------------|--|
| <div> <div>Tejas_MDW_FinalProj</div> <div> <div>+ New</div> <div>Upload</div> <div>Create deployment pipeline</div> <div>Create app</div> </div> </div> | | |
| Name | Type | |
| bnze_sales | Notebook | |
| G_Customer | Notebook | |
| G_Date | Notebook | |
| G_FactSale | Notebook | |
| G_OrderPriority | Notebook | |
| G_Product | Notebook | |
| Gld_ShipMode | Notebook | |
| MDW | Data pipeline | |
| MDWT | Semantic model | |
| Order_Return_G | Notebook | |
| RL | Notebook | |
| Tejas | Report | |
| TMDW_FinalProj | Lakehouse | |
| TMDW_FinalProj | Semantic model (...) | |
| TMDW_FinalProj | SQL analytics end... | |

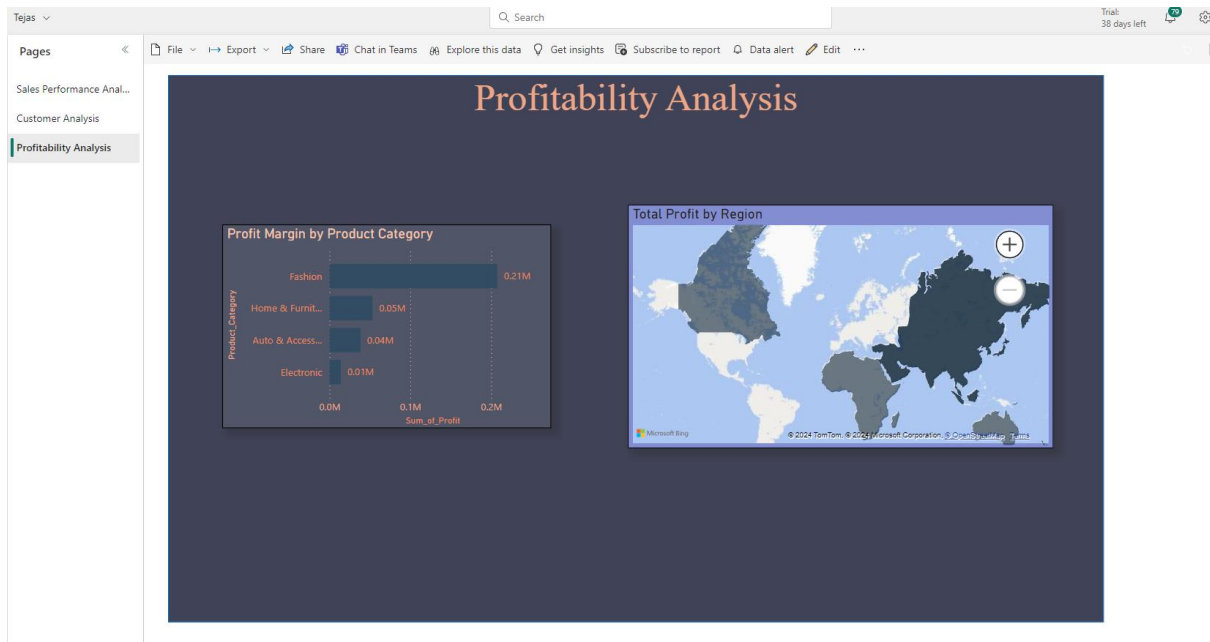
- The above screenshot will display the creation of the Data Lake, tables, and the architecture for all three levels: Bronze, Silver, and Gold.

Schematic Model:



My Dashboard





Code:

```
1 from pyspark.sql import *
2 from pyspark.sql.functions import *
3 from pyspark.sql.types import *
4 import pandas as pd
```

```
(Variable) dframe_sales: Any
dframe_sales = pd.read_excel("abfss://Tejas_MDW_FinalProj@onelake.dfs.fabric.microsoft.com/TMDW_FinalProj.Lakehouse/Files/C/Sales*.xlsx")
```

```
#creating Spark data frame for sales
dframe1 = spark.createDataFrame(dframe_sales)
```

```
# Displaying the Data Frame
display(dframe1.head(16))
```

Code zip files are attached.