**EXECUTABLE INFO DOCUMENT**

**DATA WAREHOUSING INTEGRATION**

**1. INTRODUCTION**

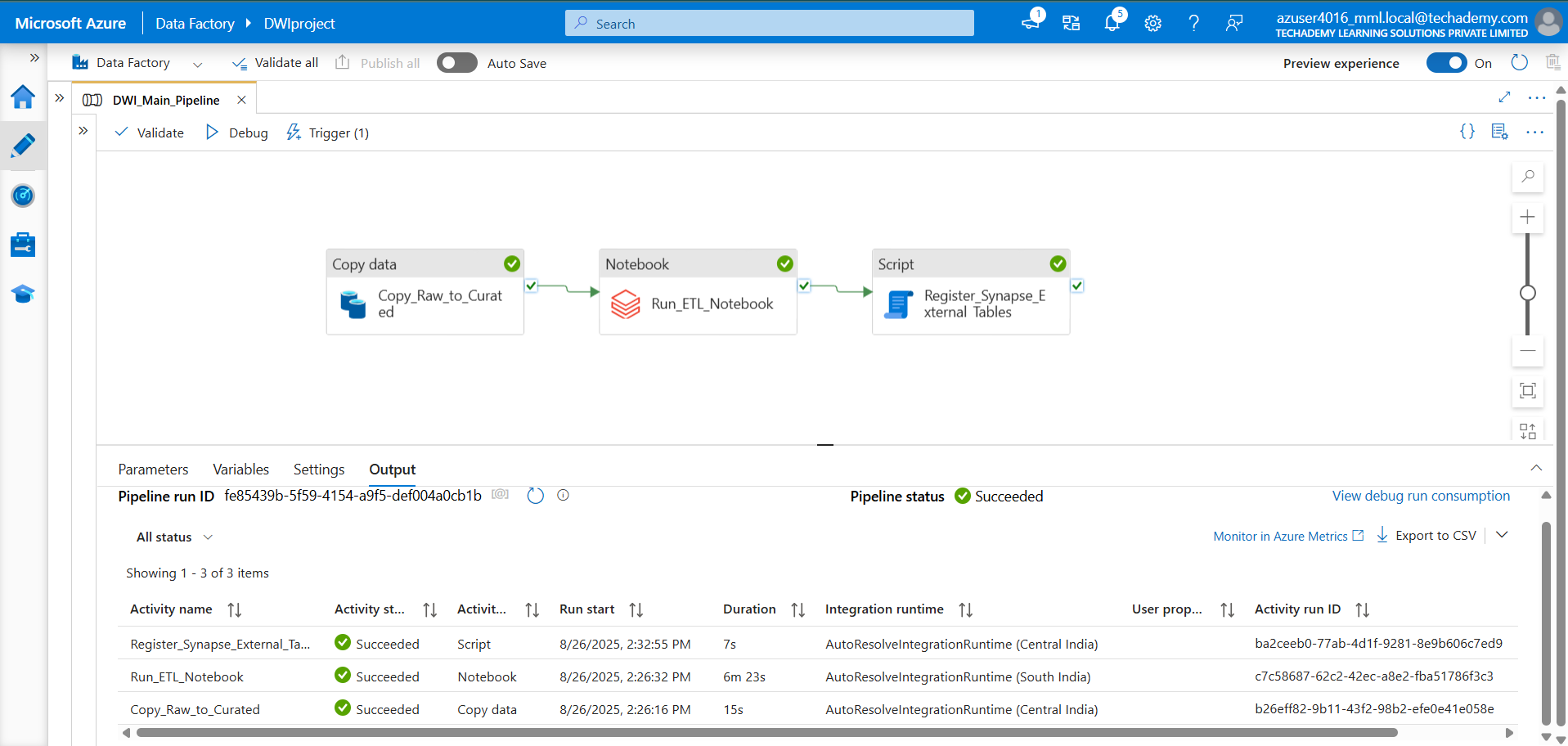
This document provides execution evidence and project results for the **Data Warehousing Integration** solution built using Azure Data Factory (ADF), Azure Databricks, Azure Data Lake Storage (ADLS), and Azure Synapse Analytics.

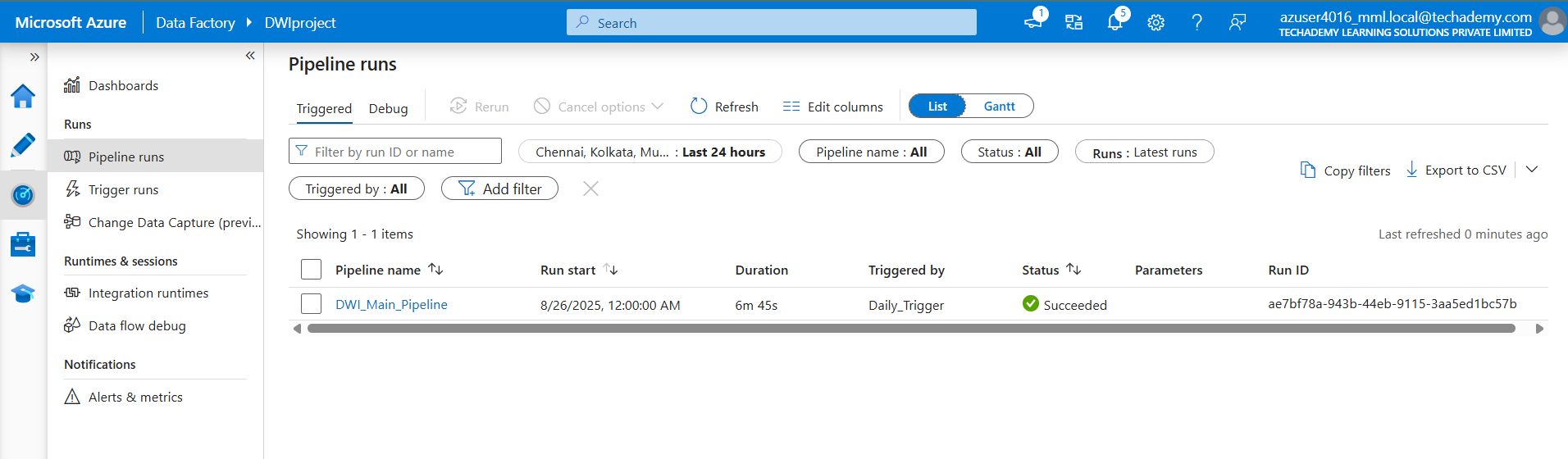
The purpose of this document is to:

* Demonstrate the successful end-to-end execution of the ETL pipeline.
* Showcase validation outputs and query results proving schema consistency and data correctness.
* Present analytical insights derived from curated warehouse tables.

## **2. PIPELINE EXECUTION PROOF**

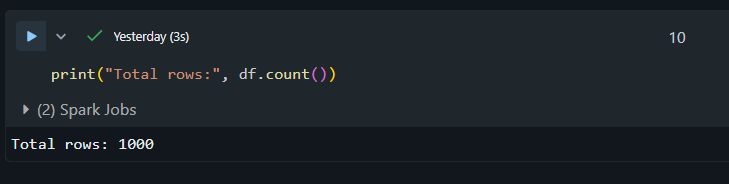
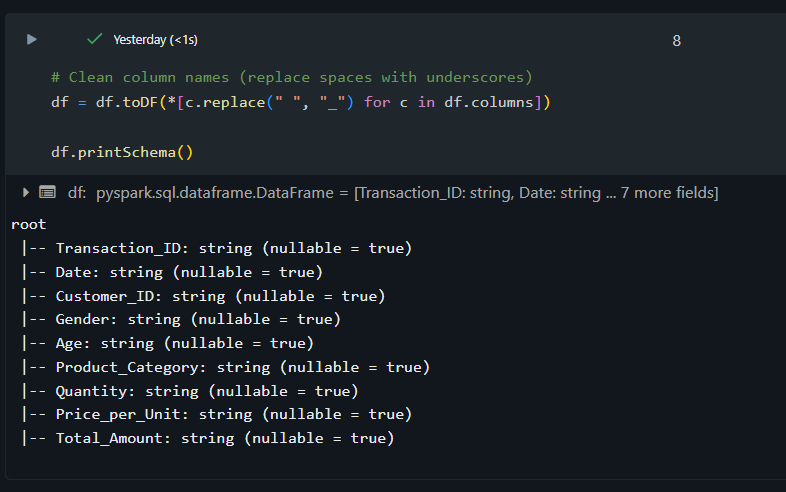
### **2.1 AZURE DATA FACTORY (ADF)**

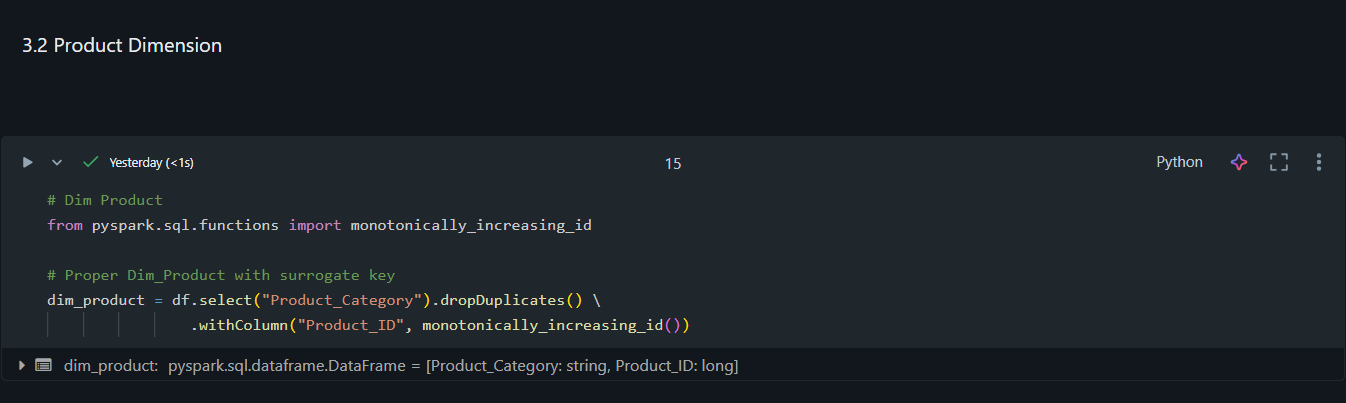
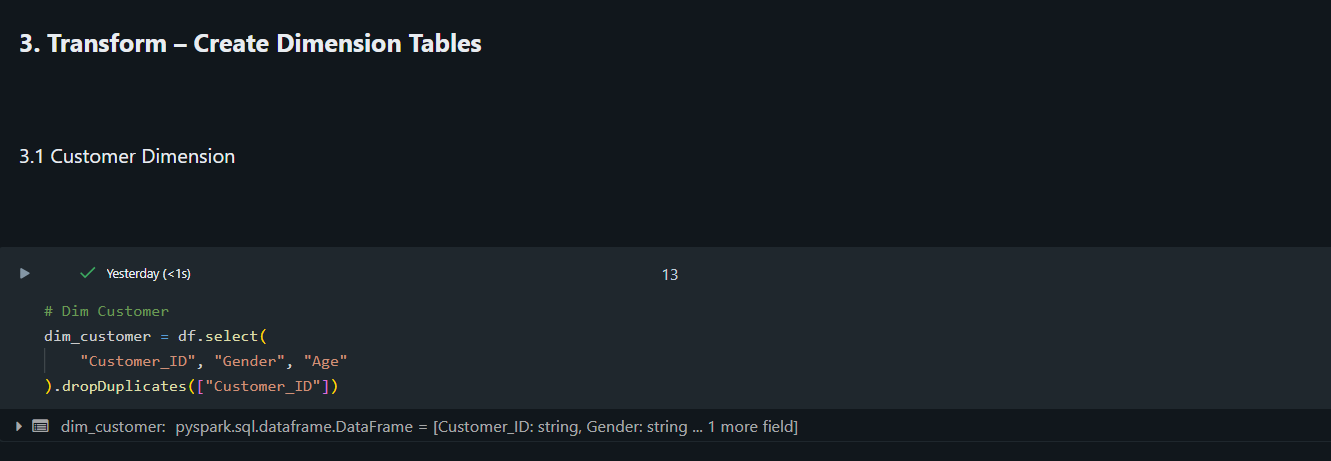
* **Pipeline Design View**
* **Pipeline Monitoring**

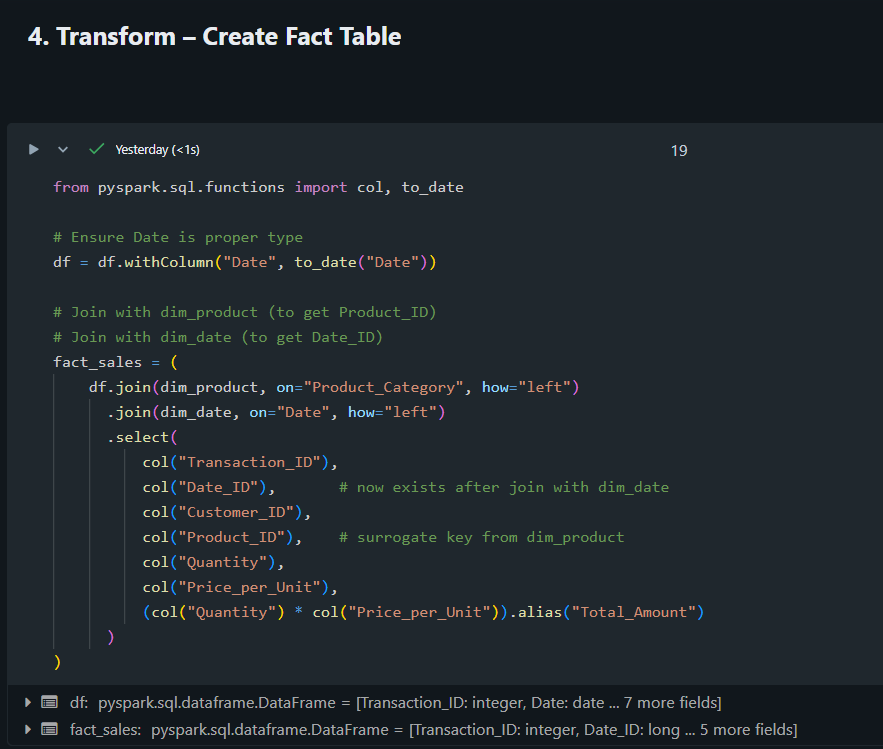


### **2.2 AZURE DATABRICKS**

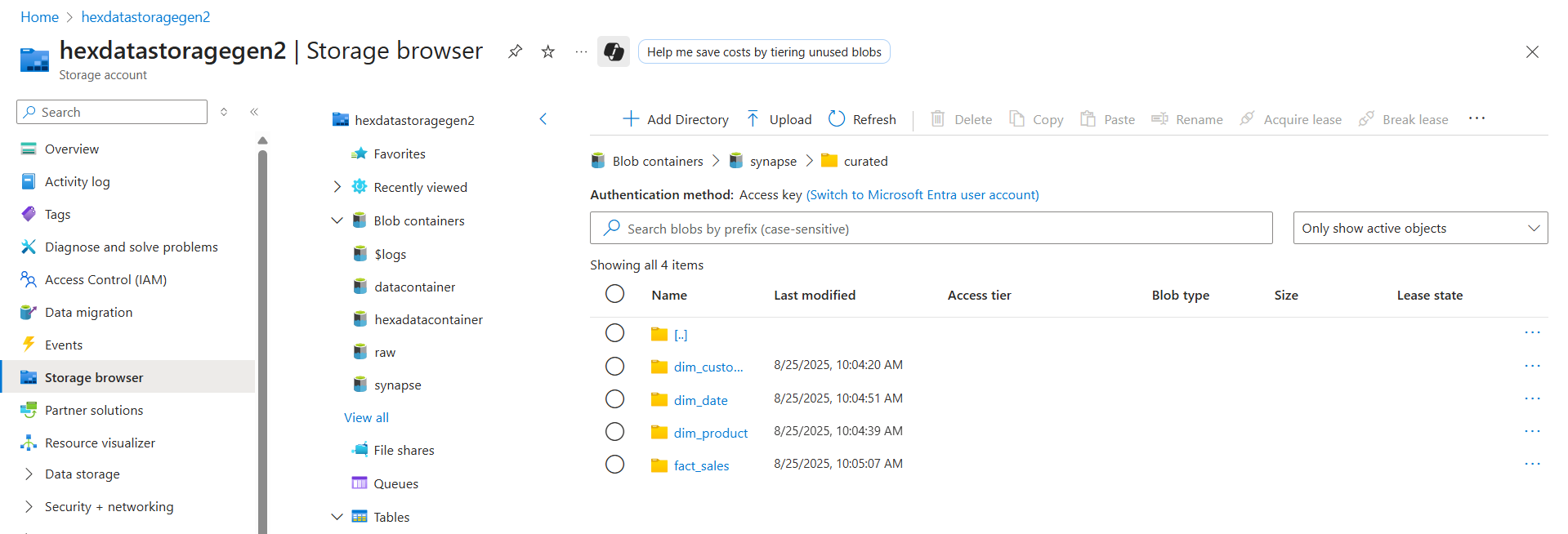
* **Databricks Notebook Transformation Output**

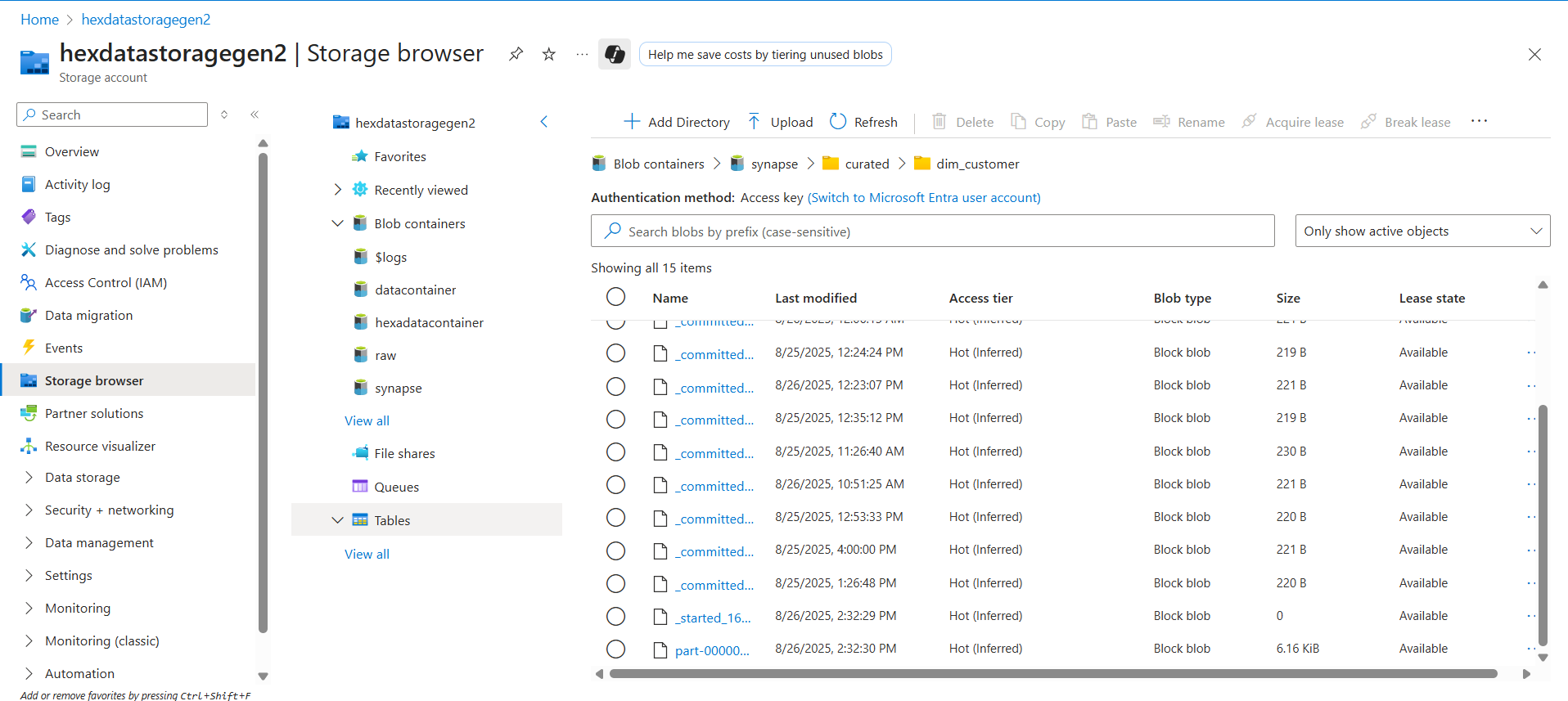


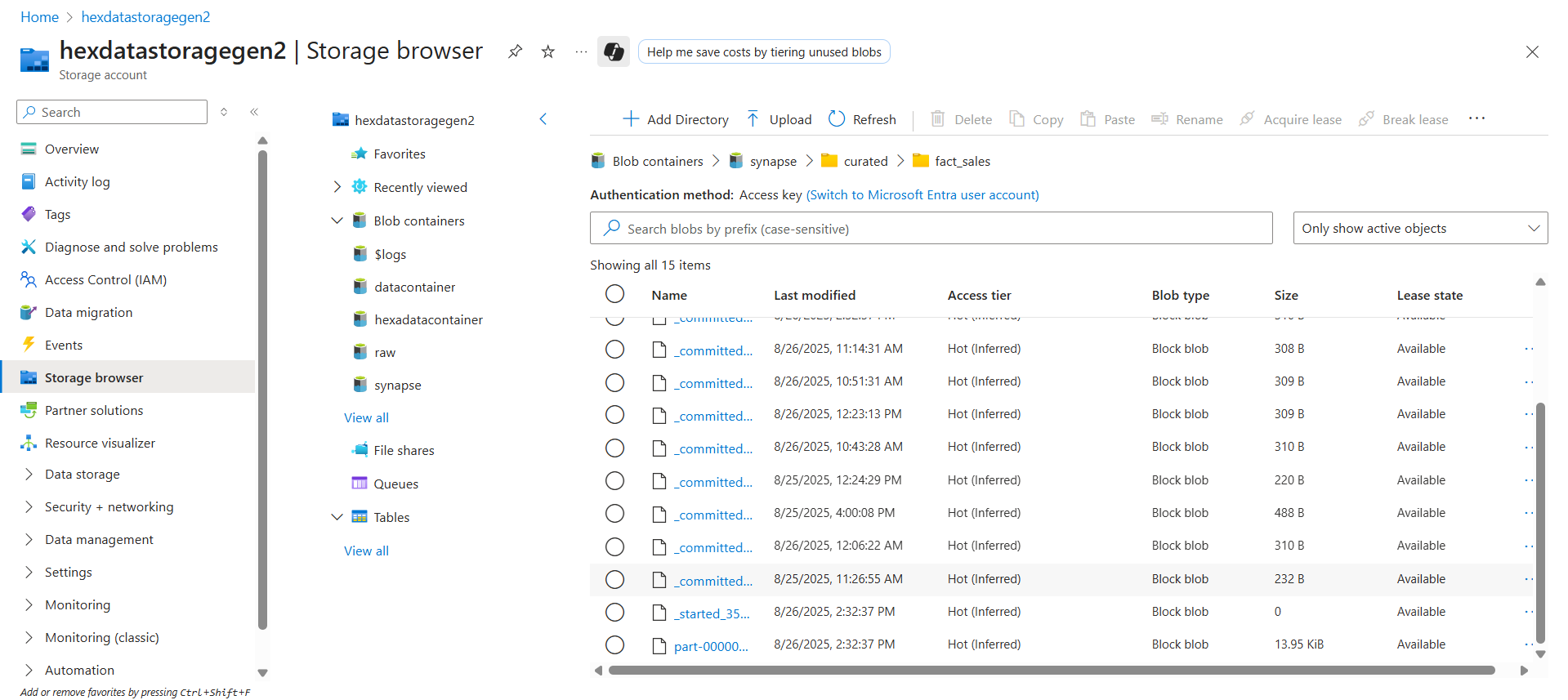




* **Curated Data Stored in ADLS**

**





### 

### 

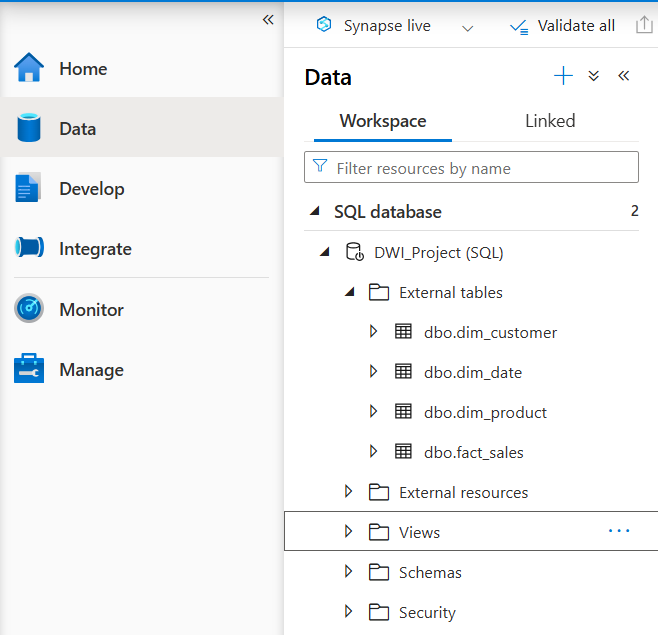
### 

### 

### 

### 

### **2.3 AZURE SYNAPSE ANALYTICS**

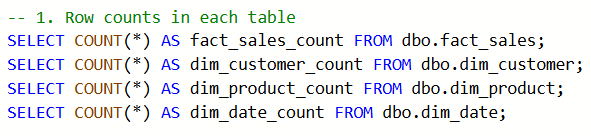
* **External Tables Registered  
  **
* **Data Source & Credential Setup**

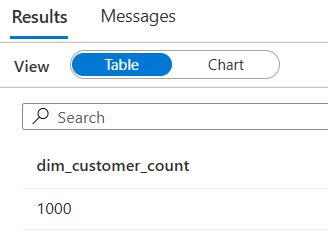
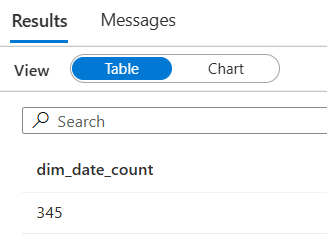
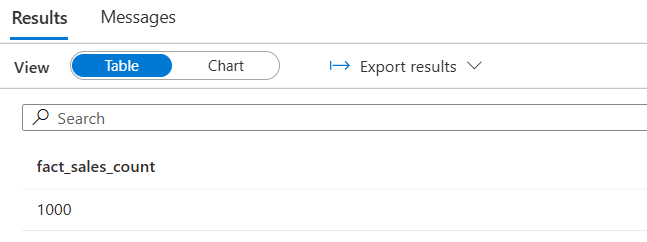
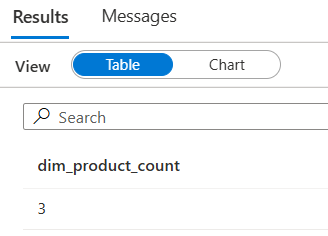
## 

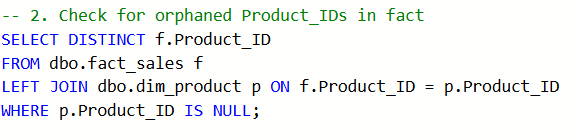
## **3. VALIDATION OUTPUTS**

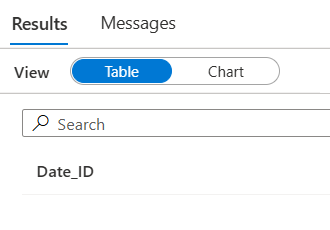
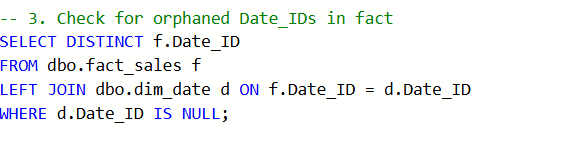
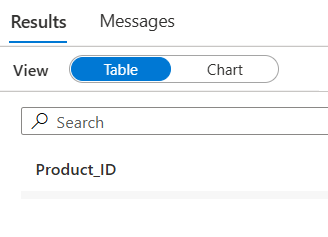
**3. VALIDATION OUTPUTS**

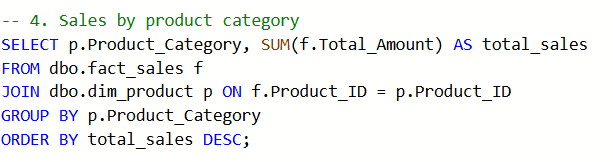
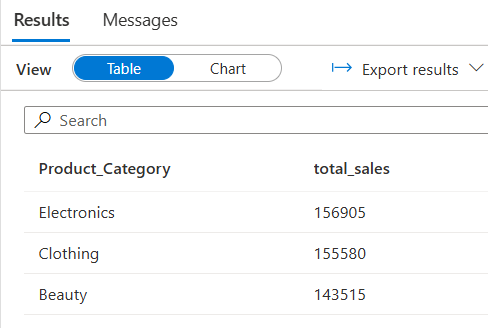
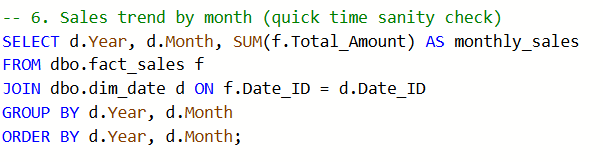
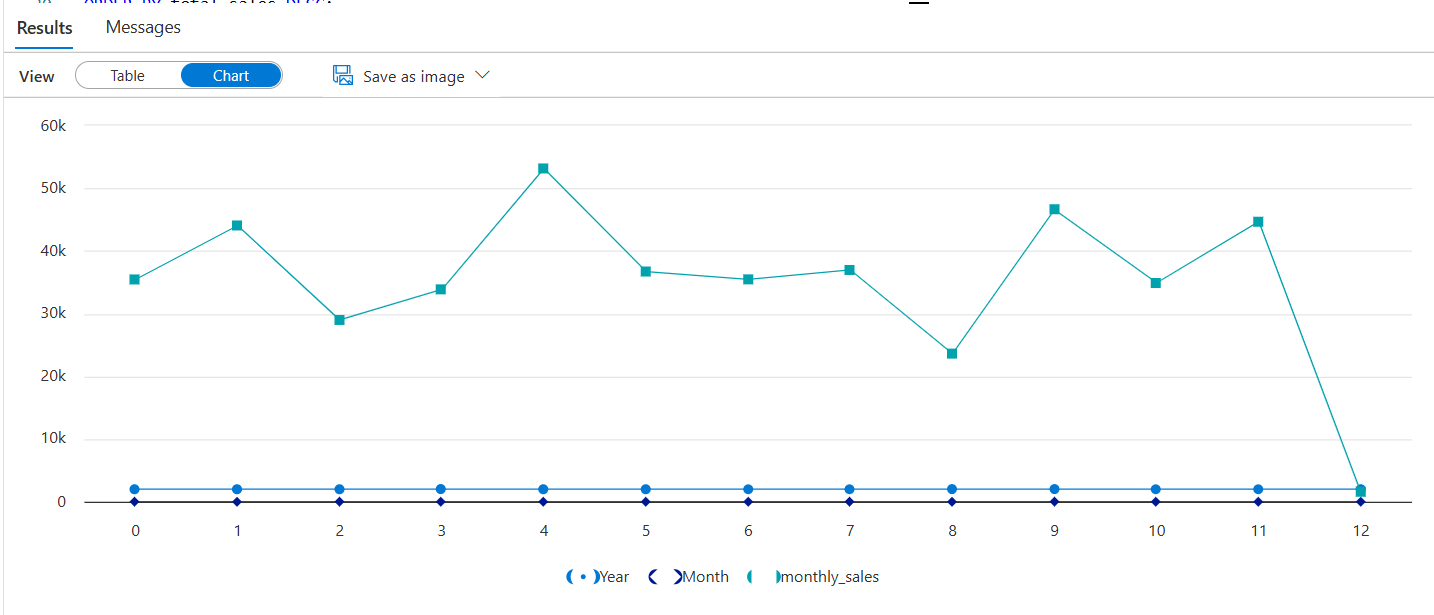
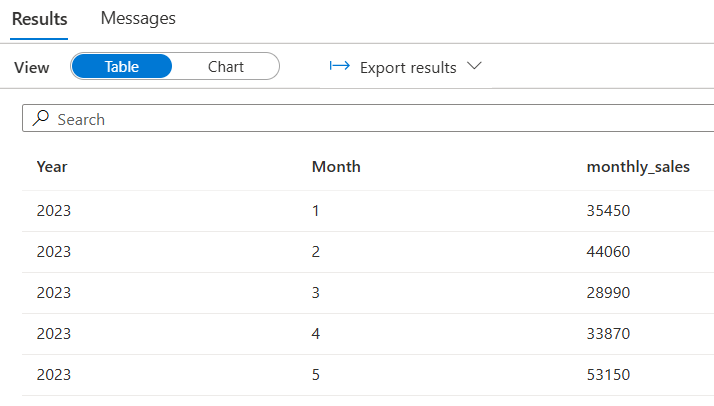
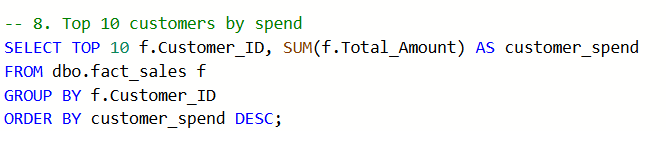
Validation was carried out in Synapse Studio using SQL queries to confirm row counts, schema correctness, and referential integrity.

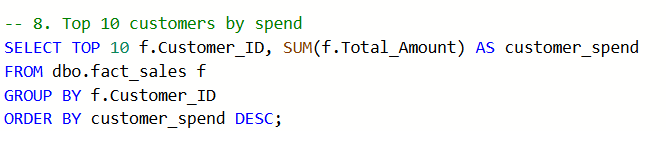
* **Row Count Validation**



* **Referential Integrity Checks**



* **Sales by Product Category**
* **Monthly Sales Trends**
* **Top 10 Customers by Spend**

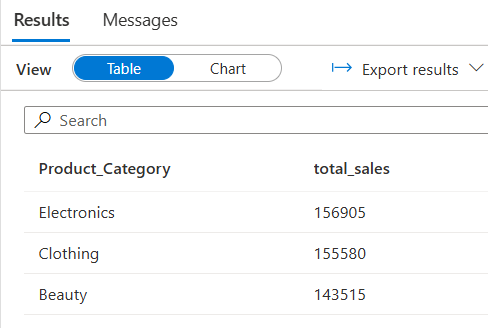
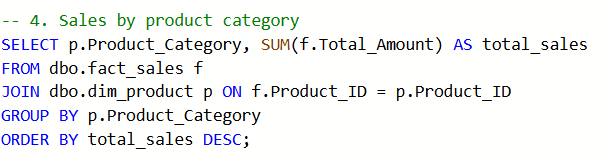
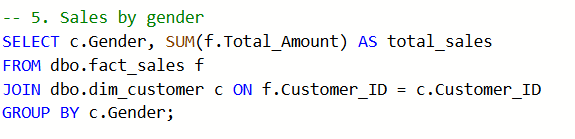


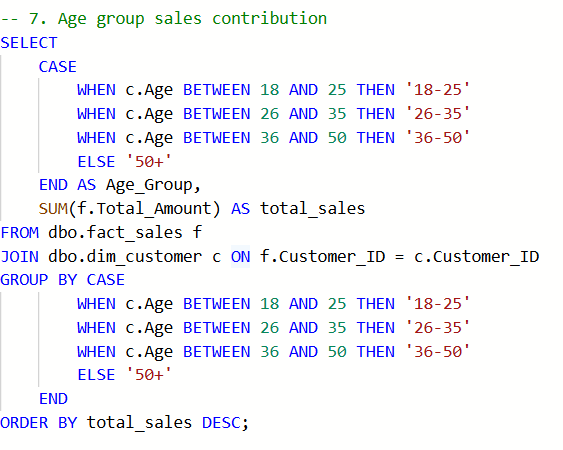
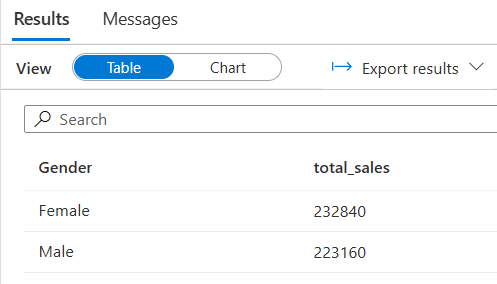
## 

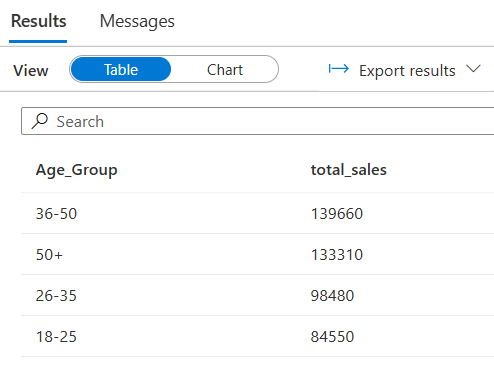
## 

## **4. ANALYSIS & INSIGHTS**

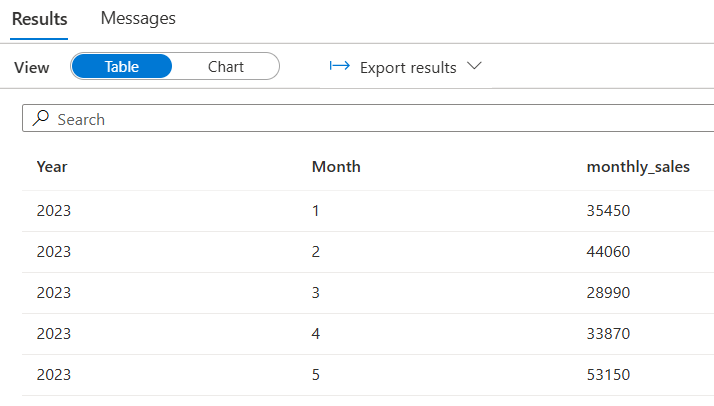
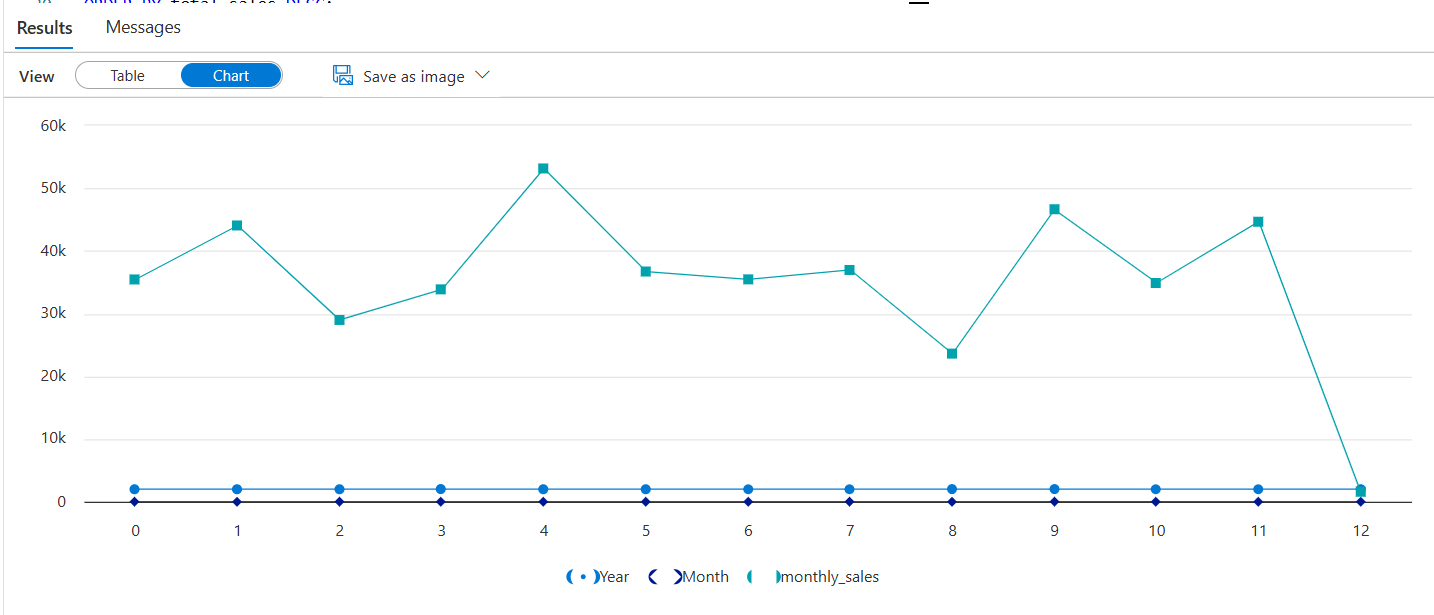
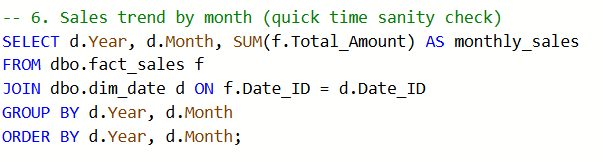
The validated star schema enables rich analytical queries. Some business insights observed:

* **Product Category Contribution**: Certain categories dominate sales.
* **Customer Segmentation**: Spending patterns vary across gender and age groups. (insert score.





* **Revenue Trends Over Time**: Sales volume shows monthly variations, enabling time-series analysis.



## **5. CONCLUSION**

The project successfully delivered:

* Automated ETL pipelines (ADF orchestrated).
* Scalable data transformations (Databricks PySpark).
* Optimized loading into a **Synapse star schema** (fact + dimensions).
* Validation queries confirming data integrity and usability for analytics.

This system is now **analytics-ready**, with potential for Power BI integration to deliver dashboards and business reporting.