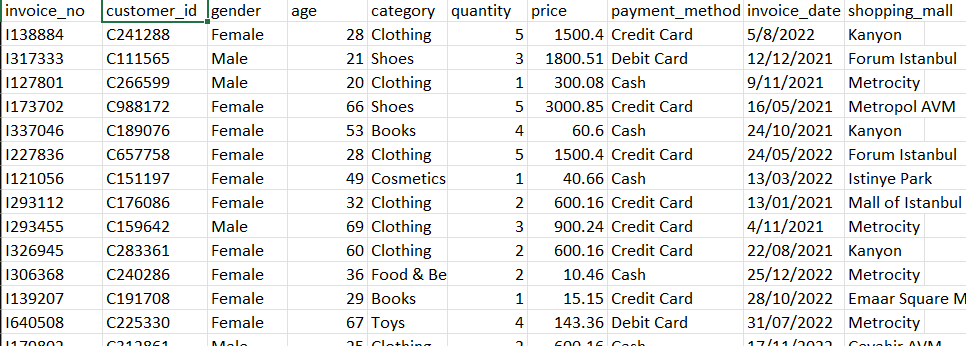
**Project 1: Customer Shopping Data Integration and Analytics**

**Objective:** Design and implement an end-to-end data integration and analytics solution for customer shopping data of a brand. This involves extracting, cleaning, transforming, and analyzing customer shopping data to gain actionable insights into retail sales trends, customer behavior, and product performance.

**Source:** [Customer Shopping Dataset - Retail Sales Data](https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset)



**Tools Used and Deliverables:**

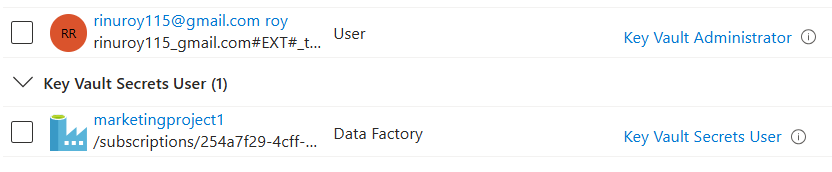
|  |  |
| --- | --- |
| **Tools Used:** | **Deliverables:** |
| Azure Data Factory | Data Factory pipelines for data extraction |
| Azure Data Lake Storage Gen2 | To store the raw and processed data |
| Azure Databricks | Databricks notebooks for data Cleaning |
| Azure Synapse Analytics | Synapse Analytics serverless pool for data analysis |
| Power BI | To visualize customer data |
| Azure Key Vault | To store blob storage account key to ensure data security |

**Implementation Steps:**

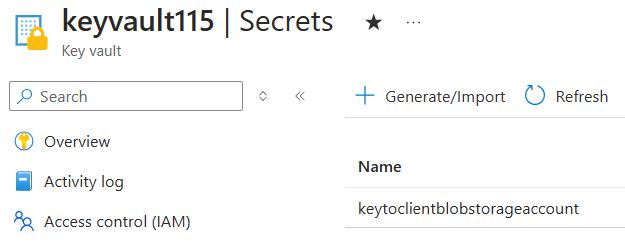
**BRONZE LAYER: Extracting raw data using ADF**

1. Set up the Azure key vault to store the storage account key of clients’s blob storage:

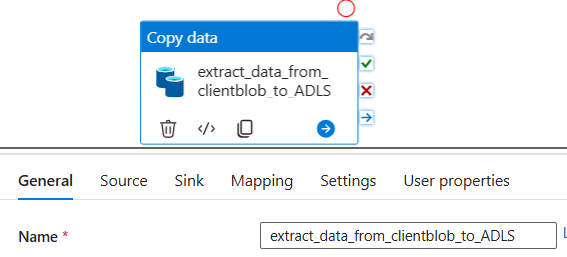
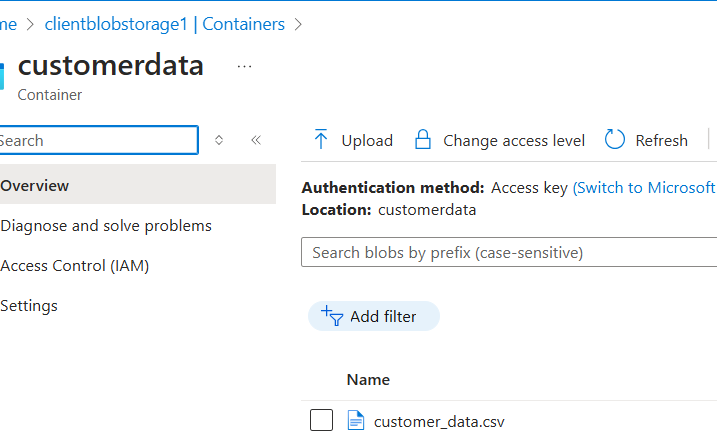
* In IAM, give necessary permissions(add role assignments) to user and ADF to add and access secret respectively:

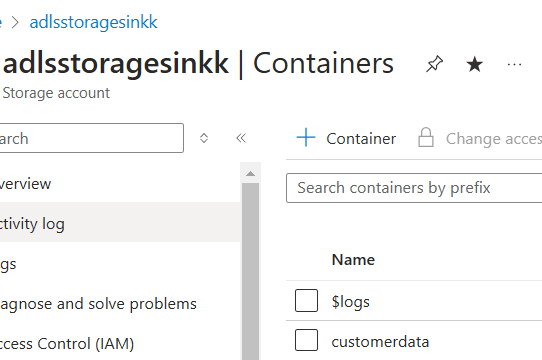


* Create a secret in key vault that store account key of client blob storage:

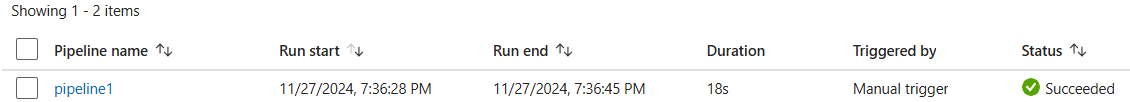


1. Set up the ADF pipeline with a copy activity to ingest the data from client blob storage to the ADLS account:

* Pipeline with Copy activity: 
* Source: Client Blob Storage Account
* Sink: ADLS

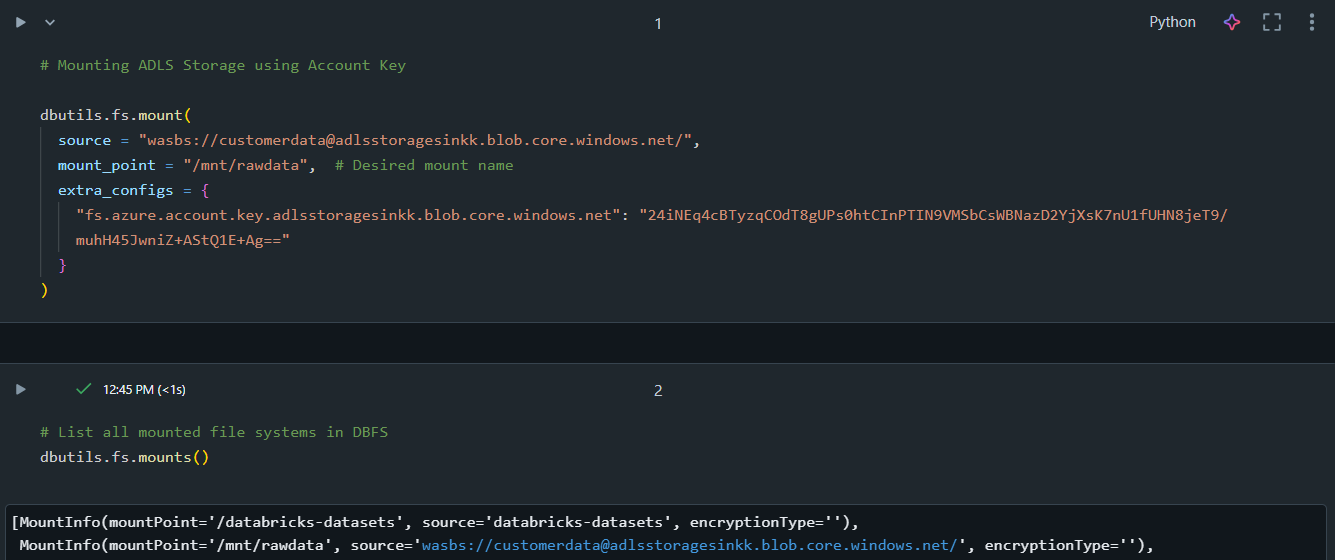


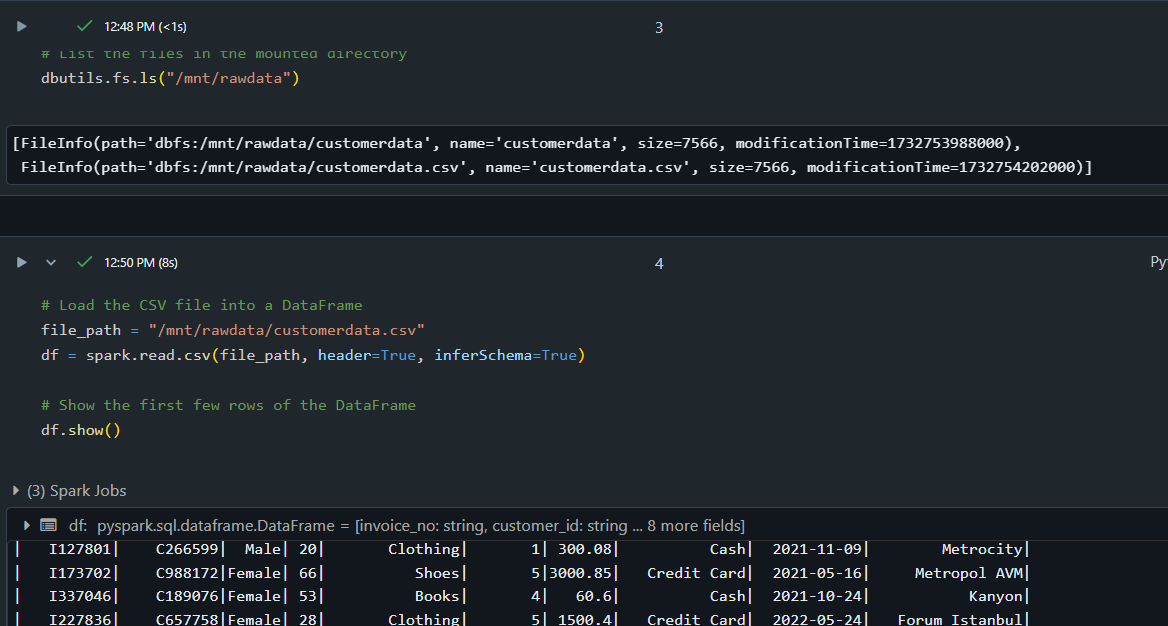
* Pipeline got triggered and completed successfully loading raw customer data to ADLS



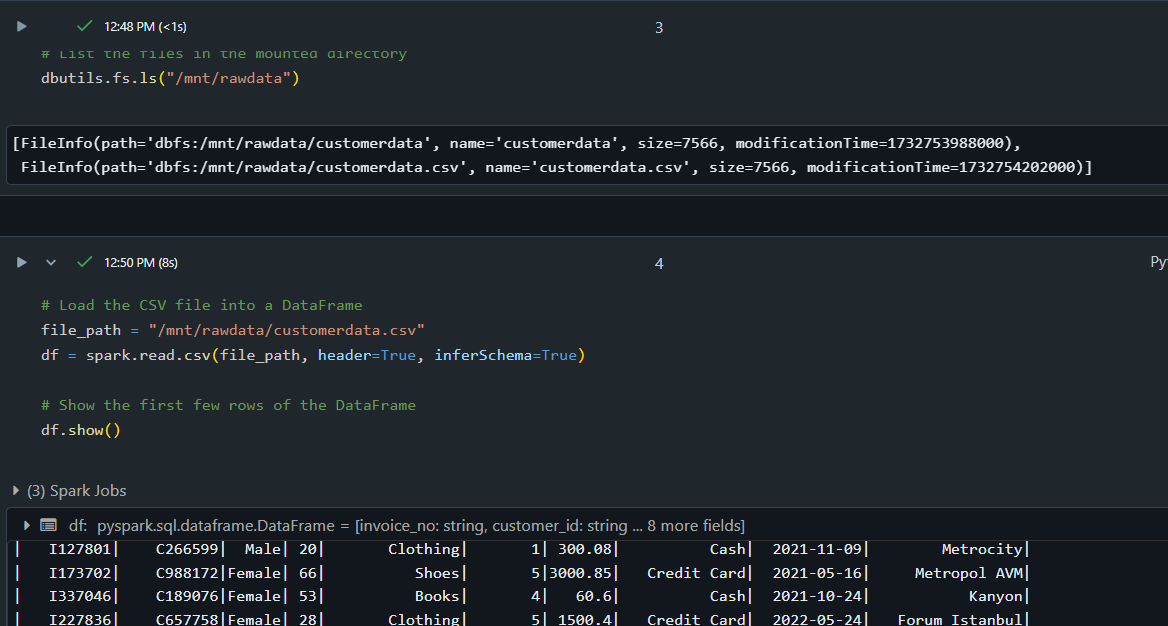
**SILVER LAYER: Data cleaning using Databricks**

1. Mount the ADLS to databricks using Account key:



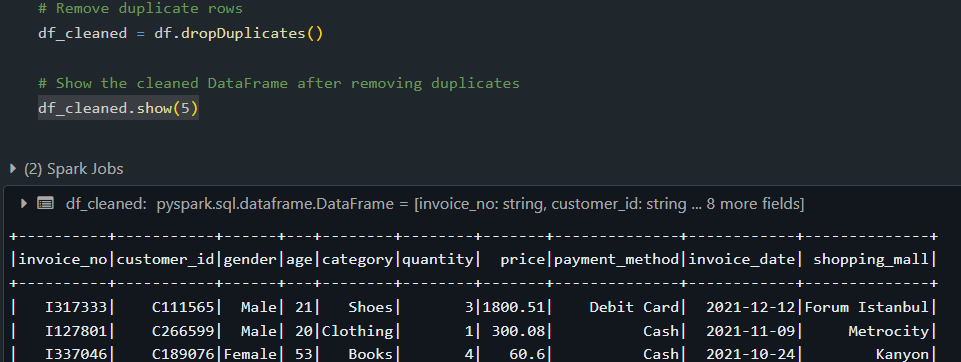


1. Load the csv file in to a data frame:

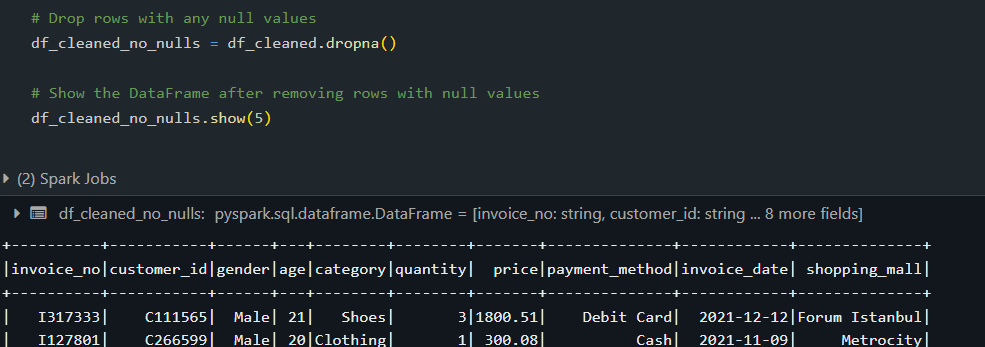


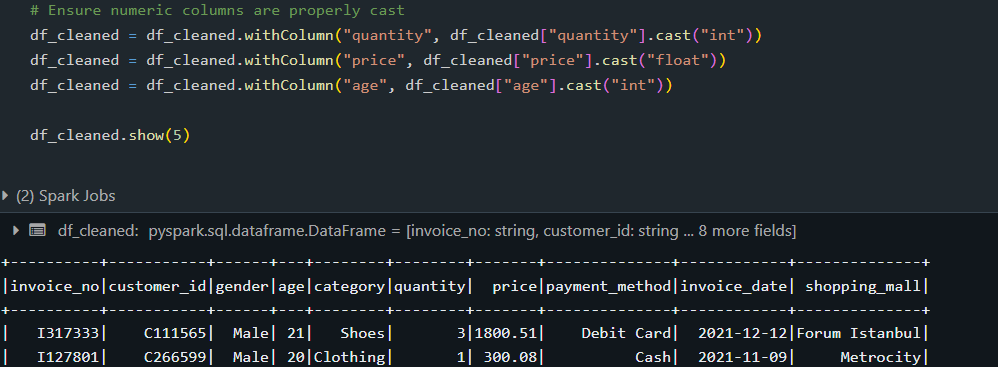
1. Perform some data cleaning on the data frame:

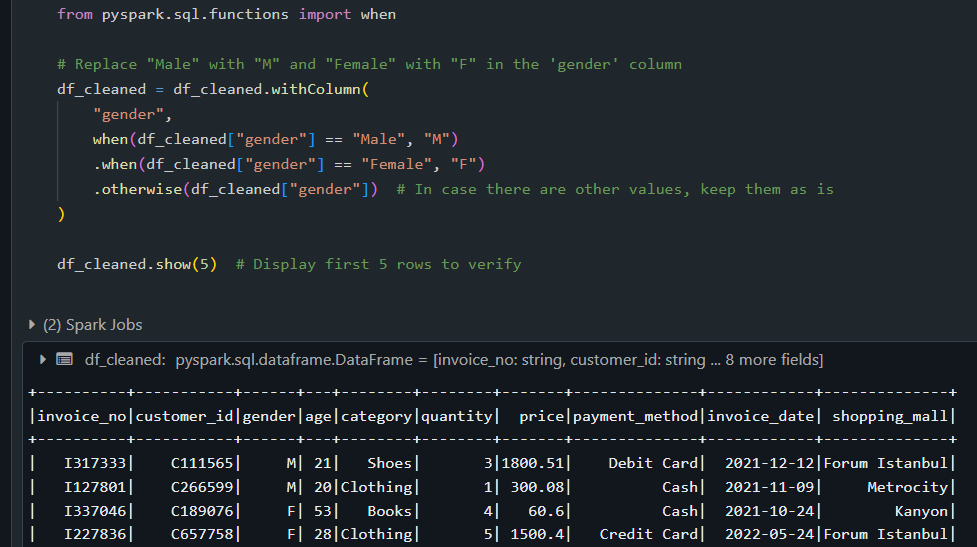
* Remove the duplicates:



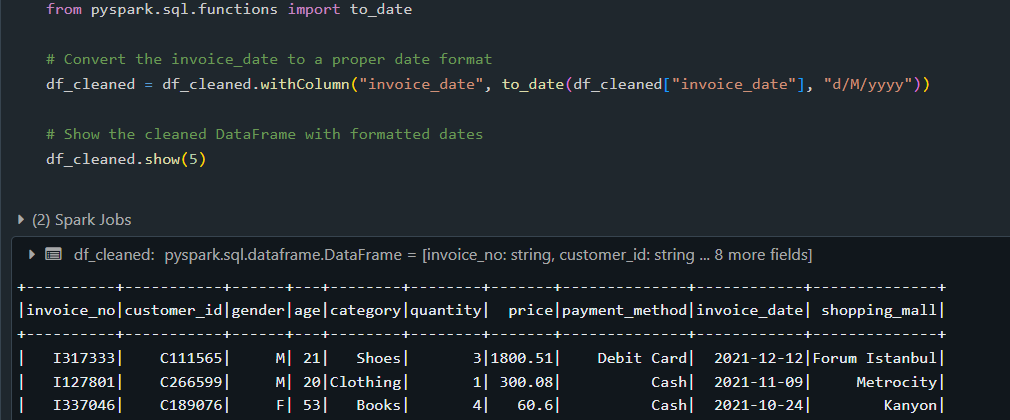
* Remove null values:



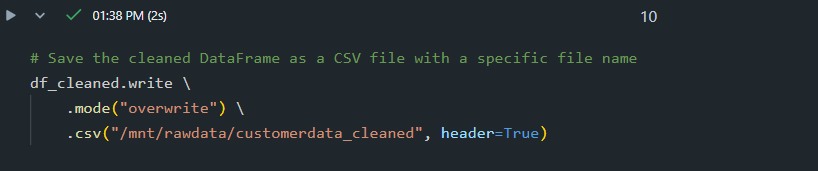
* Cast correct data types: 
* Standardize column values:



* Ensure proper date format:

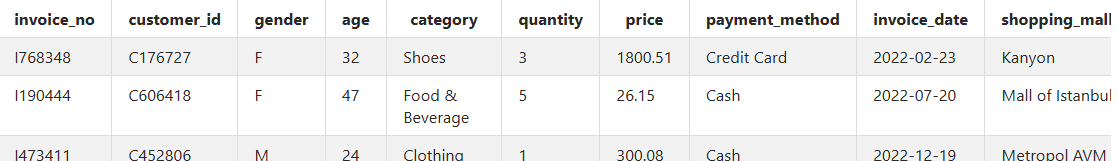


1. Save the cleaned data frame into ADLS:



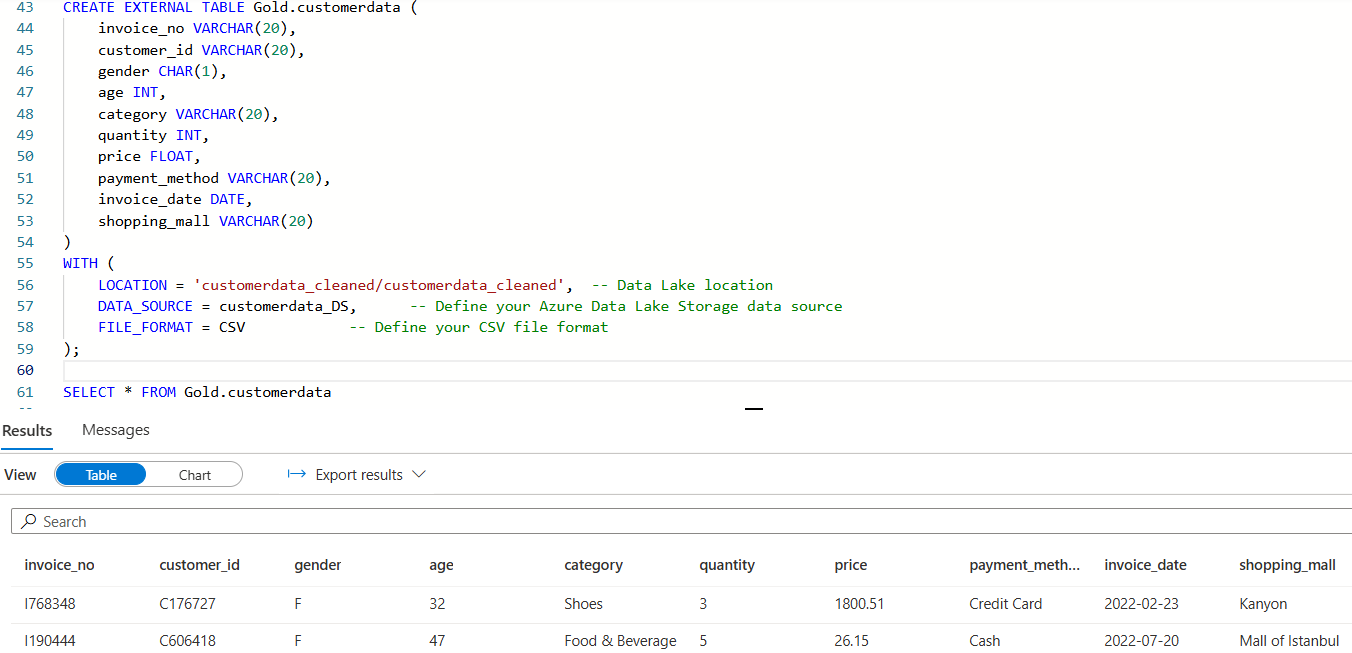
ADLS location and cleaned data preview:





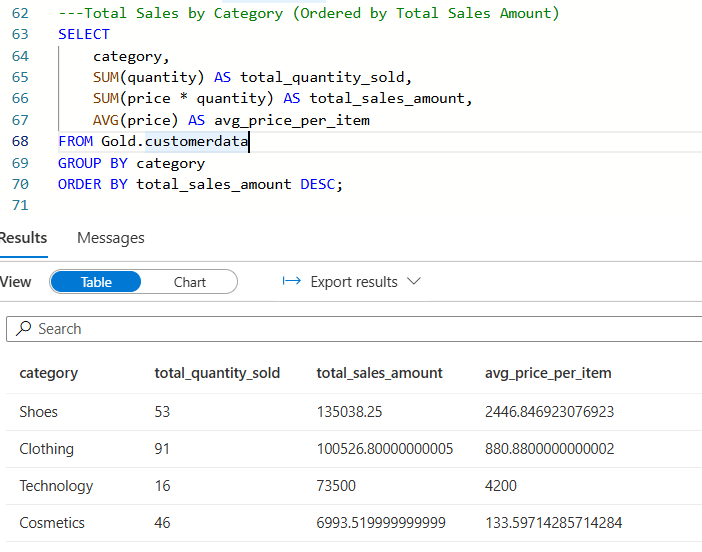
**GOLD LAYER: Data Aggregations using synapse analytics**

1. Create external data source, file format and table to read the cleaned data from ADLS using the serverless pool of Azure synapse analytics:

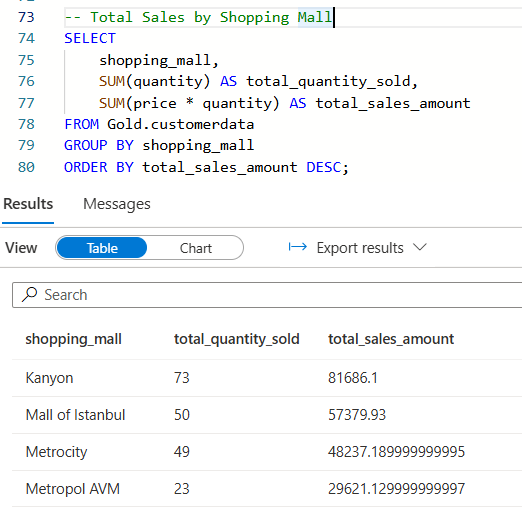


1. Perform some aggregations on the data:

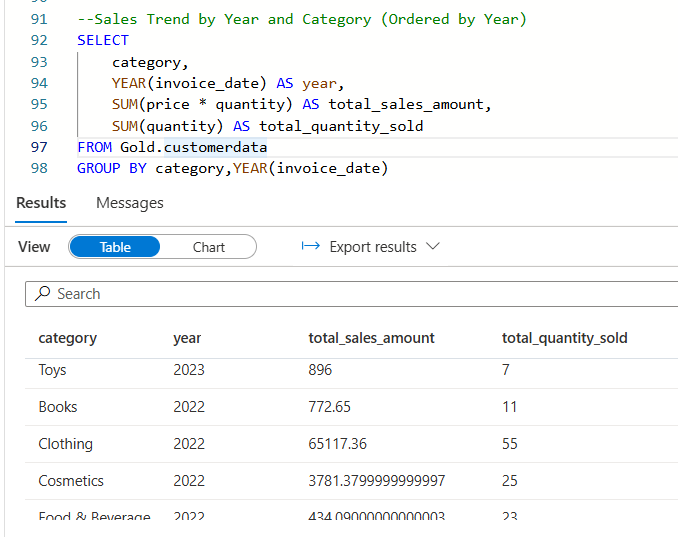
* Total sales by category:



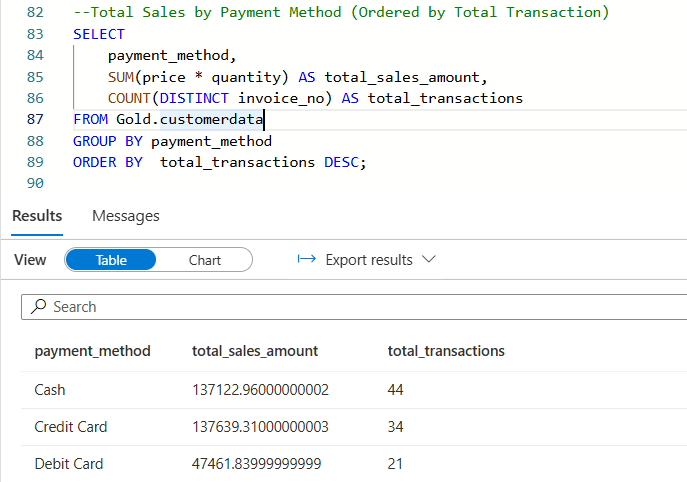
* Total sales by shopping mall:



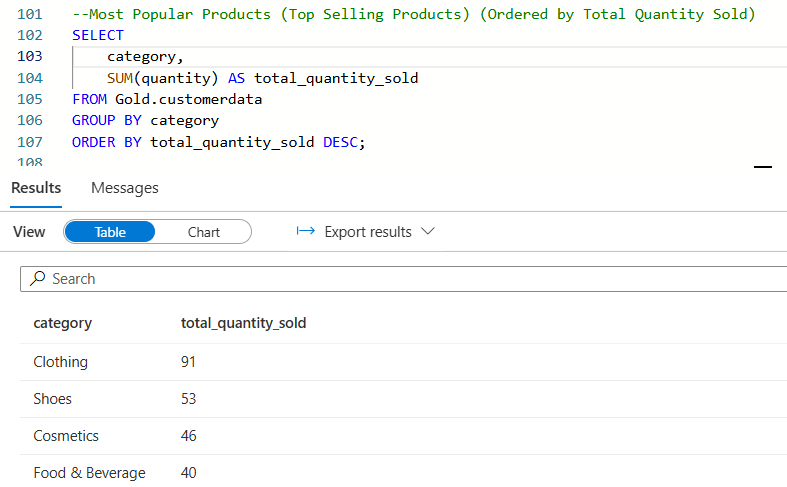
* Sales trend by year and category:



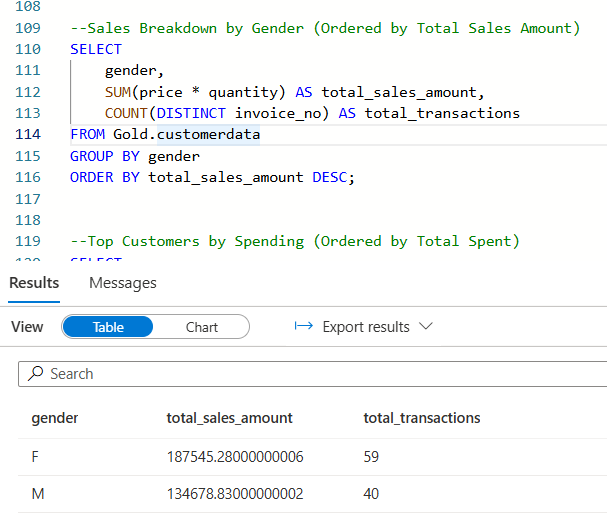
* Total sales by payment method:



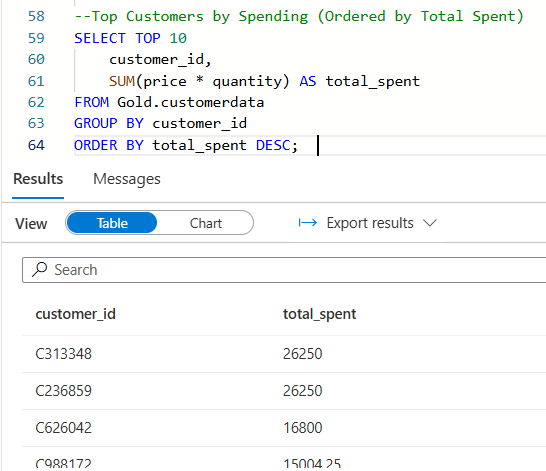
* Top selling products:



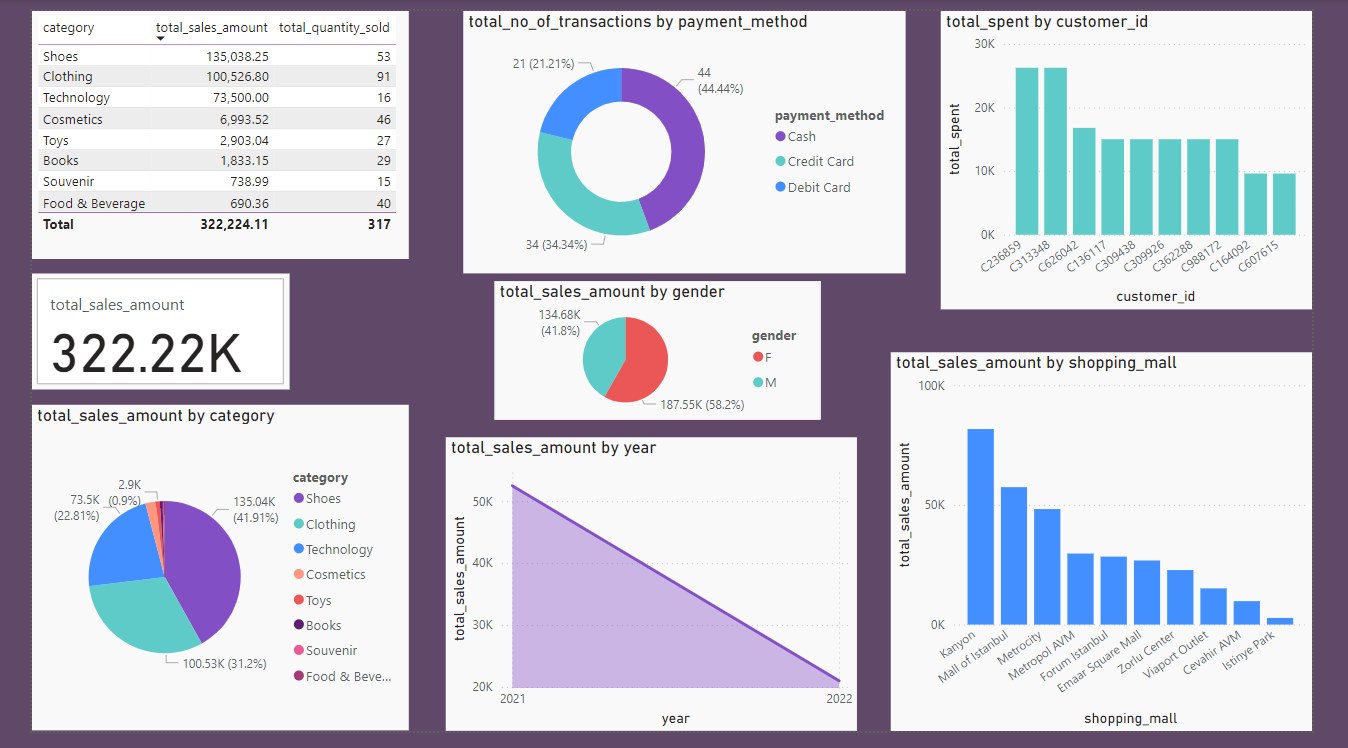
* Total sales by gender:



* Top customers by total spent



1. Data visulization using Power BI:



**Conclusion:**

This project successfully created an end-to-end data integration and analytics solution for customer shopping data . By leveraging Azure services, the data was efficiently stored, transformed, and analyzed, enabling meaningful insights into customer behavior and sales trends.

Key highlights include:

* Storing raw and processed data securely in **Azure Data Lake Storage Gen2**.
* Transforming and standardizing data using **Azure Databricks**.
* Performing efficient queries and analysis with **Azure Synapse Analytics**.
* Visualizing insights through interactive dashboards in **Power BI**.
* Securing sensitive information using **Azure Key Vault**.