# Retail Store Analytics

What I did in this project?

* Designed and built a **Retail Store Analytics pipeline** using Azure services including **Azure SQL DB**, **Azure Data Factory**, **Azure Data Lake Storage Gen2 (ADLS)**, **Azure Databricks**, and **Power BI Service**, structured following the **medallion architecture** (Bronze, Silver, Gold layers).
* Created raw data sources (transactions, stores, products, customers) in Azure SQL and JSON, ingested into ADLS via Data Factory pipelines, and stored as **Parquet files** in Bronze layer directories.
* Processed and cleaned data using **PySpark in Databricks**, producing cleaned datasets (Silver layer) and aggregated business KPIs (Gold layer), including total sales, units sold, transaction count, and average transaction value.
* Stored the refined data as **Delta files** for optimized querying and analysis.
* Exported Gold layer data as CSV files and developed business reports in **Power BI Service**, visualizing key operational insights.
* Managed version control and project files using **GitHub**, ensuring security of sensitive keys.

What were my issues and workarounds?

**Issue 1**: SKU Not Available While Creating Databricks Compute Cluster

* Problem:

Faced a SKUNotAvailable error when trying to create a Databricks compute cluster. This was due to the selected VM SKU not being available in the Azure region linked to the Databricks workspace.

* Workaround:

Queried the available VM resources using Azure CLI based on location (eastus), ResourceType, Name, Zones, and Restrictions. Selected a compatible VM SKU without restrictions within the existing zone and successfully created the cluster.

### **Issue 2**: Accessing ADLS Data via Mounting in Databricks

* Problem:

Required the ADLS Gen2 storage account access key to configure the mount point and access ADLS directories from Databricks notebooks.

* Workaround:

Retrieved the key from Azure Portal via:

Storage Account → Security + Networking → Access Keys.

Copied the key and used it in the mounting script (dbutils.fs.mount()) in Databricks to establish access to ADLS files for reading and writing operations.

### **Issue 3**: Unable to Create Delta Table in Unity Catalog

* Problem:

While attempting to register Delta files as a Delta Table in Unity Catalog (using CREATE TABLE USING DELTA LOCATION), encountered the error:

[NO\_PARENT\_EXTERNAL\_LOCATION\_FOR\_PATH]

The system required a registered External Location for the ABFSS path due to Unity Catalog’s governance policies, blocking table creation.

* Workaround:

Since External Locations were not configured (admin-level restriction), opted for a practical workaround:

* + Exported Gold layer data as CSV files from Databricks instead of creating a Delta Table.
  + Used these CSV files directly in Power BI Service for building reports, bypassing Unity Catalog restrictions and allowing the project to proceed without delays.