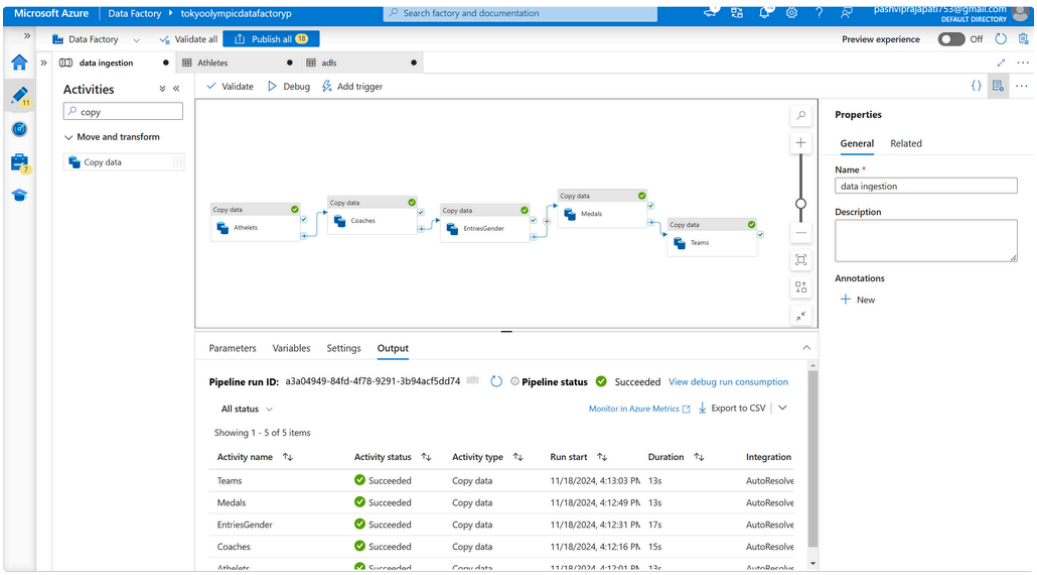
Tokyo Olympics Data End-to-End

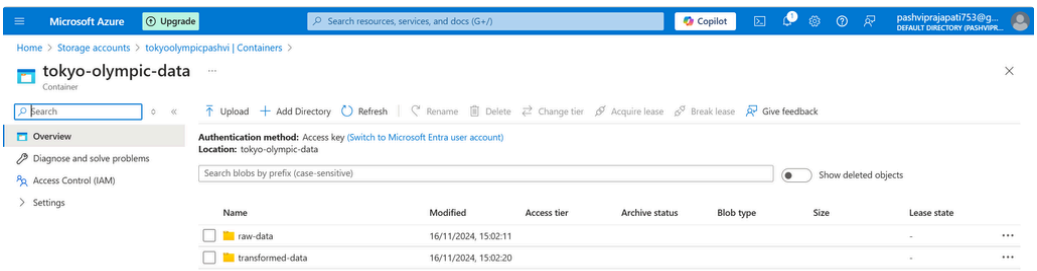
I have used the Tokyo Olympics as a Dataset.

Objective:

This contains the details of over 11,000 athletes, with 47 disciplines, along with 743 Teams taking part in the 2021(2020) Tokyo Olympics. This dataset includes the details of participating Athletes, Coaches, and Teams and the Entries by gender. It contains their names, countries represented discipline, gender of competitors, and names of the coaches.

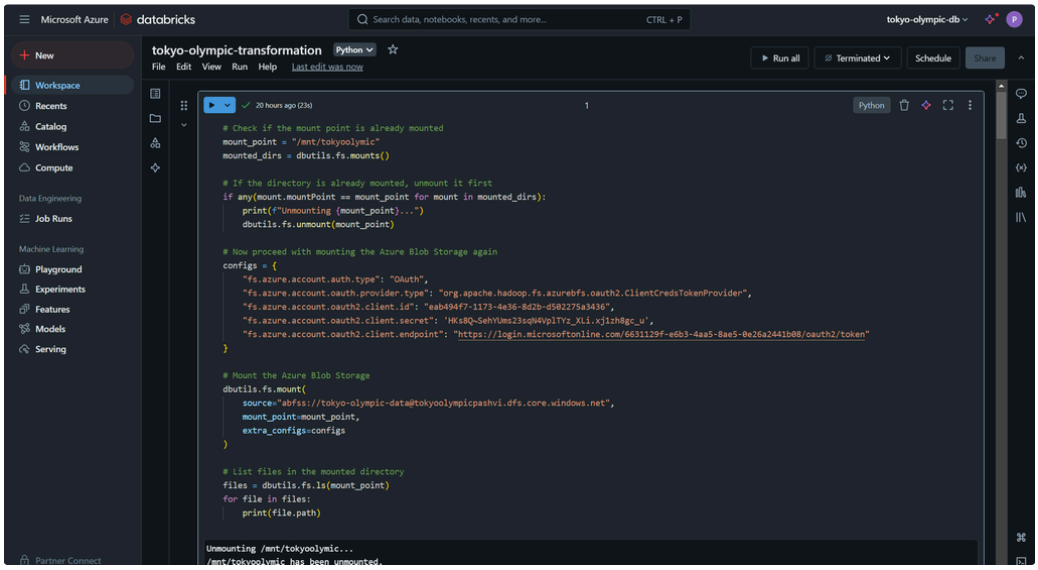
Data Ingestion from Data Factory to Data lake

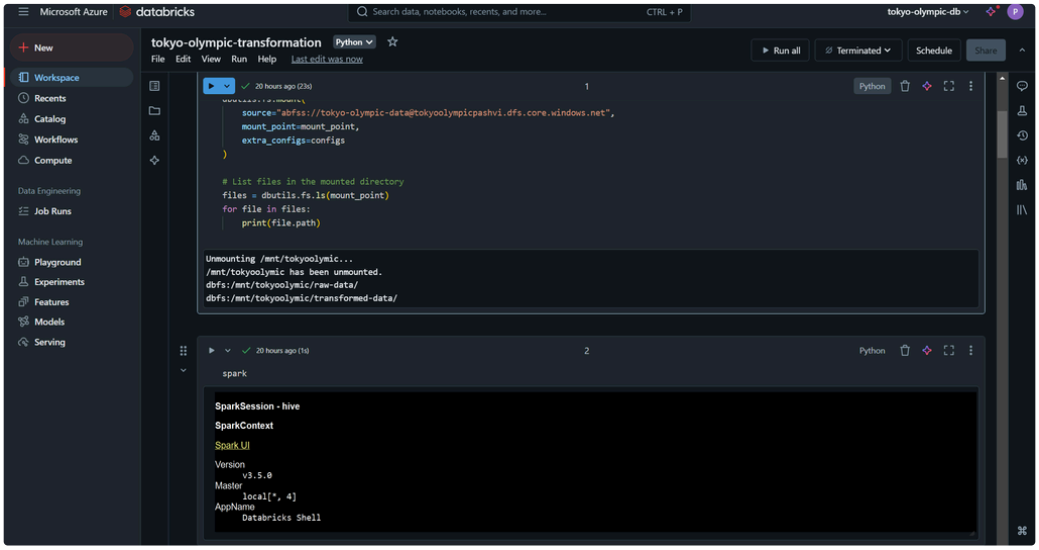


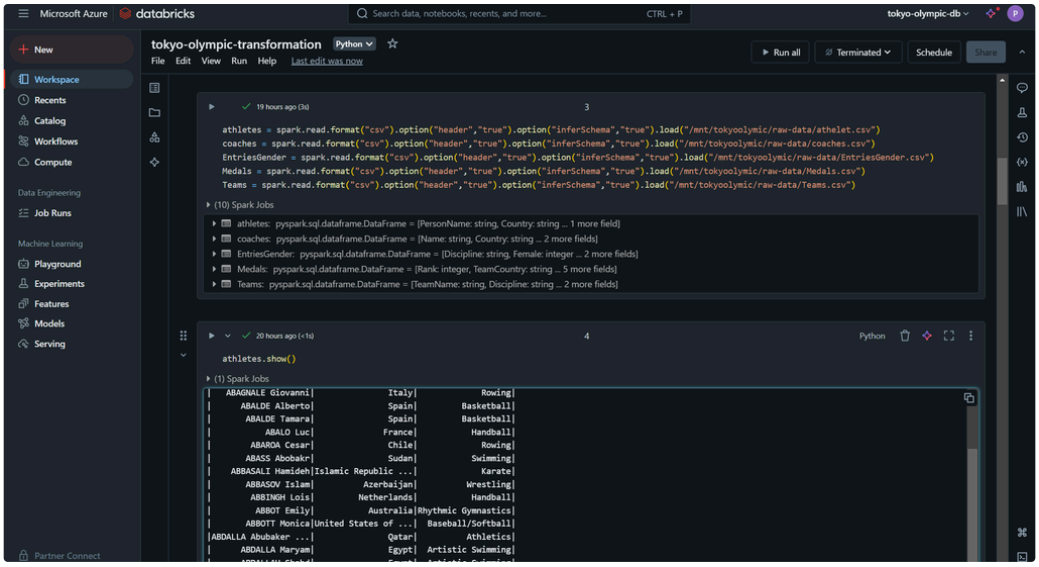


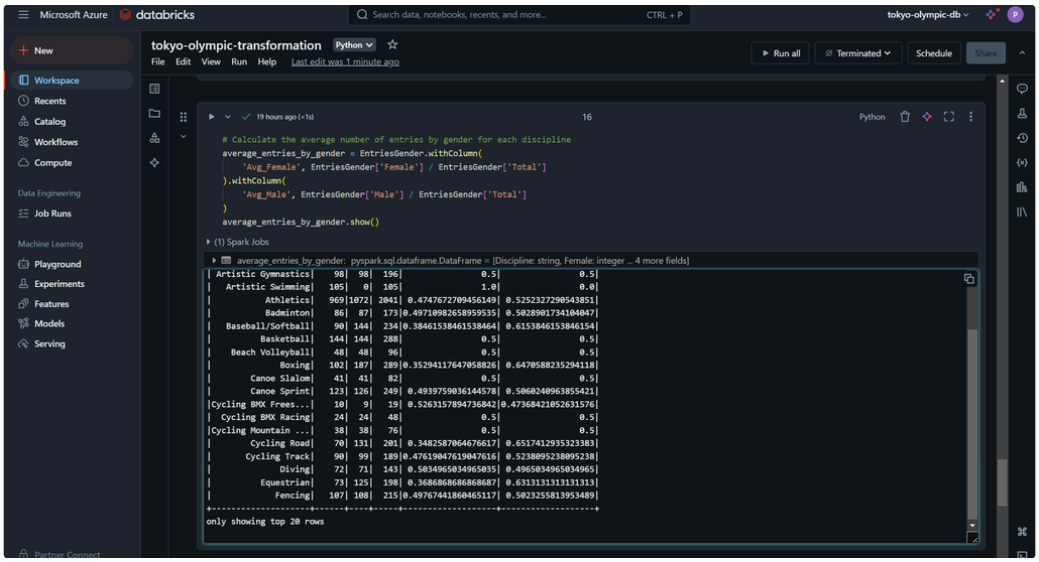
The raw dataset from the Tokyo Olympics, which includes details of over 11,000 athletes, 47 disciplines, and 743 teams, is ingested into the data lake from the data factory. This process ensures that all relevant information about participating athletes, coaches, and teams, along with entries by gender, is stored efficiently for further analysis and processing.

Data Transformation using Databricks



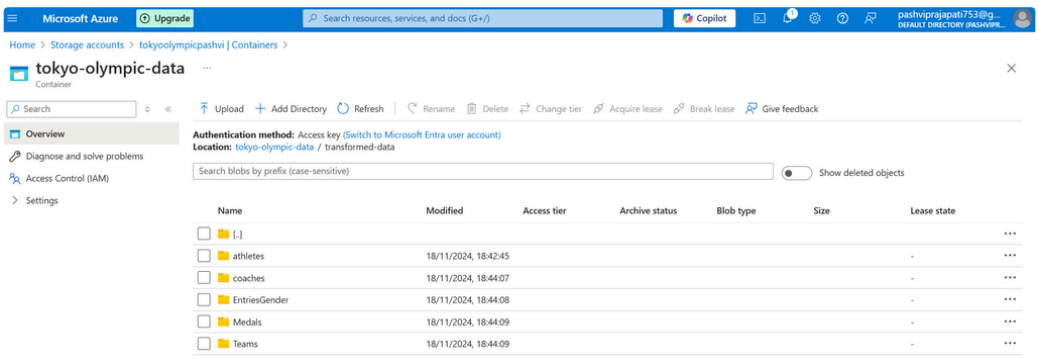






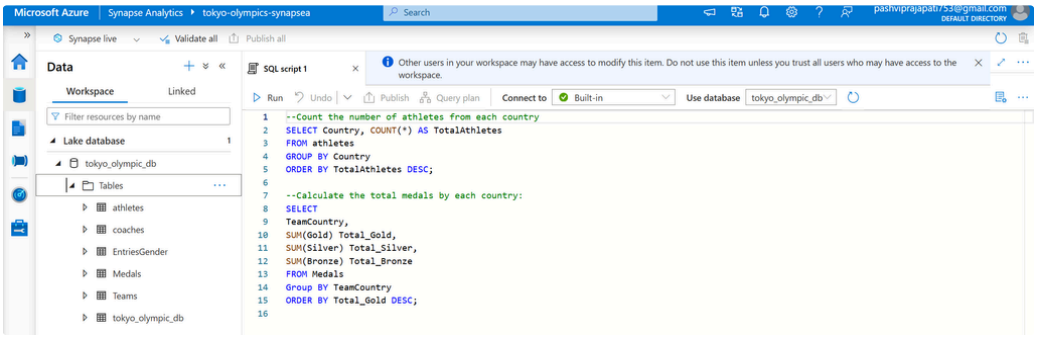
The raw data from the Tokyo Olympics has been transformed using Databricks. This process involves cleaning, structuring, and preparing the data for analysis. By leveraging the capabilities of Databricks, we can efficiently handle large datasets, apply various transformations, and ensure that the data is in a suitable format for further analysis and insights.

Loading transformed data into Date lake Gen2



The transformed data from the Tokyo Olympics has been successfully loaded into Data Lake Gen2. This step ensures that the cleaned and structured data is stored in a scalable and efficient manner, ready for further analysis and insights. The use of Data Lake Gen2 allows for enhanced performance and accessibility of the dataset, facilitating better data management and retrieval for future projects.

Alternative to using databricks, the raw data is also being transformed directly into synapse analysis



The raw data from the Tokyo Olympics can also be transformed directly using Synapse Analytics. This approach allows for efficient data transformation without the need for an intermediary step through Databricks. By leveraging Synapse, users can perform data cleaning, structuring, and preparation for analysis directly within the platform, streamlining the workflow and potentially enhancing performance for specific use cases.

PowerBI Dashboard