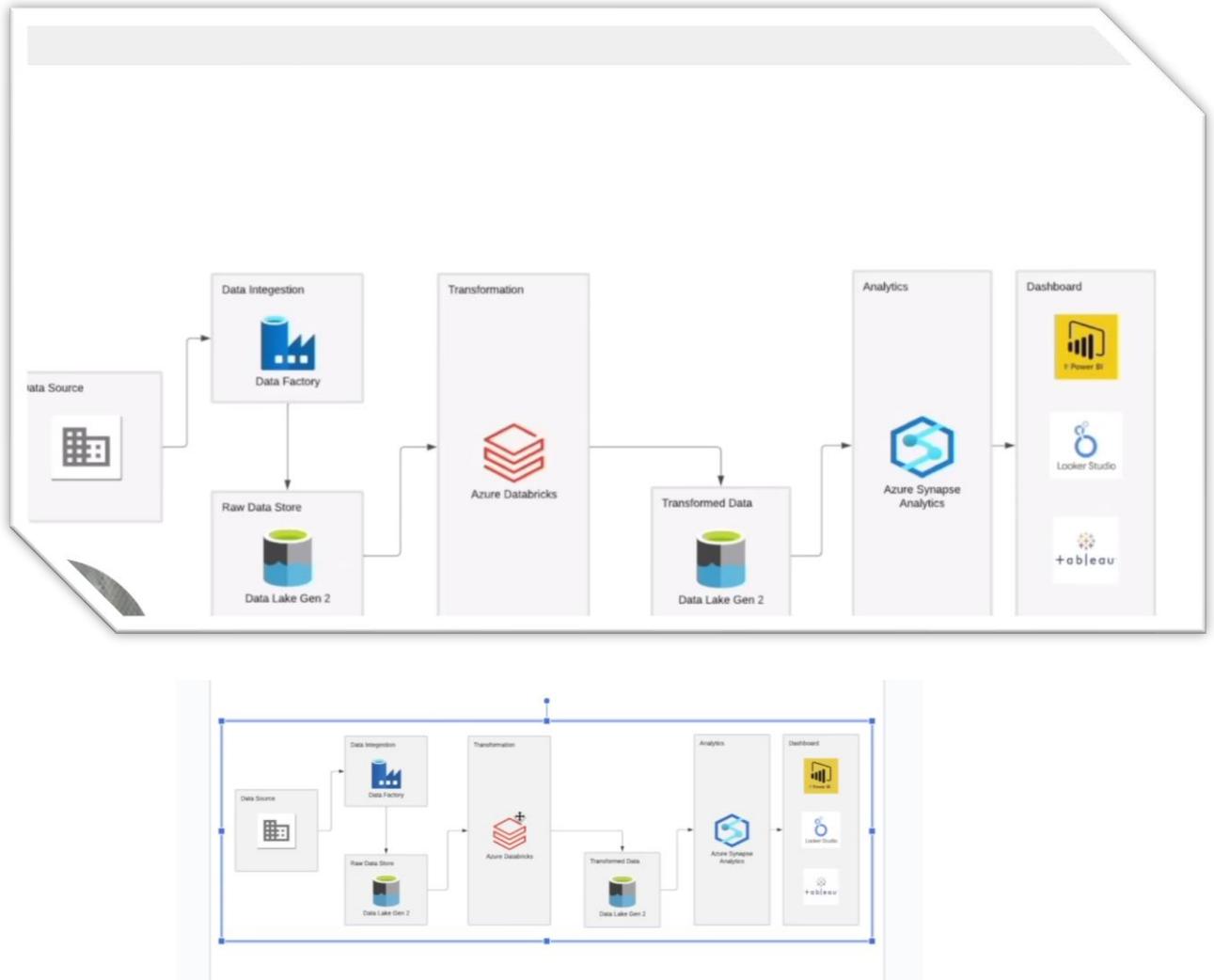


Tokyo Olympic Data Analytics | Azure End-To-End Data Engineering Project

Pipeline Design Architecture:

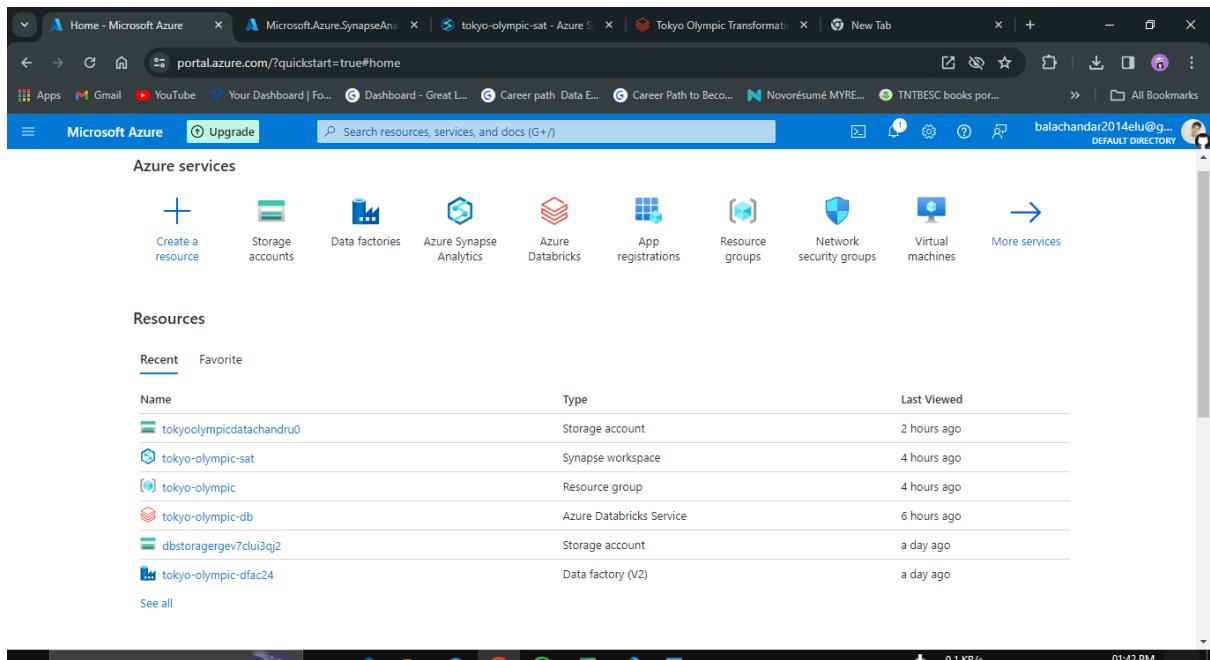


Analyse Tokyo Olympic data using various tools and technologies, including Azure Data Factory, Data Lake Gen 2, Synapse Analytics, and Azure Databricks.

Dataset Used

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 contains the details of the Athletes, Coaches, Teams participating as well as the Entries by gender. It contains their names, countries represented, discipline, gender of competitors, name of the coaches.

Source(Kaggle): [2021 Olympics in Tokyo](#)



The screenshot shows the Microsoft Azure portal homepage. At the top, there are several tabs: "Home - Microsoft Azure", "Microsoft.Azure.SynapseAnalytics", "tokyo-olympic-sat - Azure Synapse Analytics", "Tokyo Olympic Transformation", and "New Tab". Below the tabs, the search bar contains "portal.azure.com/?quickstart=true#home". The main navigation bar includes "Microsoft Azure" and "Upgrade" buttons, along with a search bar for "Search resources, services, and docs (G+/)". On the right, the user's profile is shown with the email "balachandar2014elu@gmail.com" and "DEFAULT DIRECTORY".

Azure services

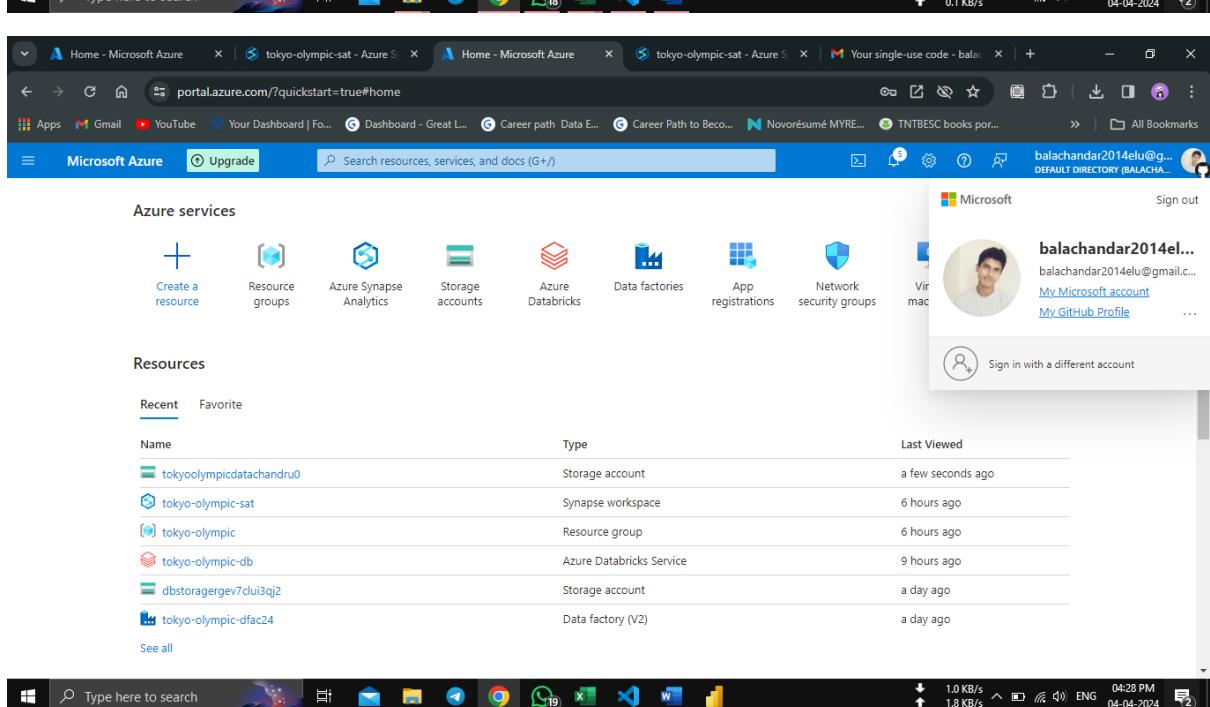
- Create a resource
- Storage accounts
- Data factories
- Azure Synapse Analytics
- Azure Databricks
- App registrations
- Resource groups
- Network security groups
- Virtual machines
- More services

Resources

Recent Favorite

Name	Type	Last Viewed
tokyoolympicdatachandru0	Storage account	2 hours ago
tokyo-olympic-sat	Synapse workspace	4 hours ago
tokyo-olympic	Resource group	4 hours ago
tokyo-olympic-db	Azure Databricks Service	6 hours ago
dbstoragegev7clui3qj2	Storage account	a day ago
tokyo-olympic-dfac24	Data factory (V2)	a day ago

See all



The second screenshot shows the same Azure portal interface, but with a different user profile. The user's name is "balachandar2014elu..." and their email is "balachandar2014elu@gmail.com". There are links for "My Microsoft account" and "My GitHub Profile". A "Sign out" button is also visible.

Azure services

- Create a resource
- Resource groups
- Azure Synapse Analytics
- Storage accounts
- Azure Databricks
- Data factories
- App registrations
- Network security groups
- Virtual machines

Resources

Recent Favorite

Name	Type	Last Viewed
tokyoolympicdatachandru0	Storage account	a few seconds ago
tokyo-olympic-sat	Synapse workspace	6 hours ago
tokyo-olympic	Resource group	6 hours ago
tokyo-olympic-db	Azure Databricks Service	9 hours ago
dbstoragegev7clui3qj2	Storage account	a day ago
tokyo-olympic-dfac24	Data factory (V2)	a day ago

See all

List of Resources used for this project.. **Azure Services Used**

- Azure Data Factory:** For data ingestion from GitHub.
- Azure Data Lake Storage Gen2:** As the primary data storage solution.
- Azure Databricks:** For data transformation tasks.
- Azure Synapse Analytics:** To perform in-depth data analytics.

The screenshot shows the Azure portal homepage with the 'tokyo-olympic' resource group selected. The 'Resource details' section indicates it is a Resource group located in Southeast Asia, using a Free Trial subscription. The 'Recent' resources list includes 'tokyo-olympic-c', 'tokyo-olympic-c', 'tokyoolympicdata', 'tokyo-olympic-s', 'tokyo-olympic', and 'dbstoragergev7'. The 'Resource details' table shows the last viewed items:

Type	Last Viewed
Azure Databricks Service	5 minutes ago
Data factory (V2)	28 minutes ago
Storage account	32 minutes ago
Synapse workspace	4 hours ago
Resource group	4 hours ago
Storage account	a day ago

The screenshot shows the 'Resource groups' page. It lists four resource groups: 'databricks-rg-tokyo-olympic-db-3kcdqgb62qsu', 'NetworkWatcherRG', 'synapseworkspace-managedrg-4fbf0656-148e-422d-b647-2a47c9653641', and 'tokyo-olympic'. Each entry includes a checkbox, the name, subscription information ('Free Trial'), location ('Southeast Asia'), and a three-dot menu icon.

Resource group created.

The screenshot shows the Microsoft Azure portal interface. The user is navigating through the 'Data factories' section under the 'Microsoft Azure' category. A search bar at the top contains the query 'Data factories'. Below the search bar, there are several filter options: 'Subscription equals all', 'Type equals all', 'Resource group equals all', and 'Location equals all'. The results table displays one record: 'tokyo-olympic-dfac24', which is a 'Data factory (V2)' located in the 'Free Trial' subscription, resource group 'tokyo-olympic', and location 'Southeast Asia'. The table has columns for Name, Type, Subscription, Resource group, and Location.

Showing 1 to 1 of 1 records.

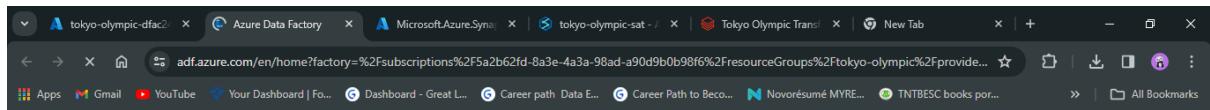
Name	Type	Subscription	Resource group	Location
tokyo-olympic-dfac24	Data factory (V2)	Free Trial	tokyo-olympic	Southeast Asia

< Previous Page 1 of 1 Next >

Give feedback

The screenshot shows the 'Azure Data Factory Studio' interface for the data factory 'tokyo-olympic-dfac24'. The left sidebar lists navigation options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Networking, Managed identities, Properties, Locks), Getting started (Quick start), and Monitoring. The main area features a large blue icon of a factory building. Below it, the text 'Azure Data Factory Studio' is displayed. A prominent blue button labeled 'Launch studio' is centered. To the right of the button are four cards: 'Quick Starts' (cloud icon), 'Tutorials' (book icon), 'Template Gallery' (document icon), and 'Training Modules' (certificate icon).

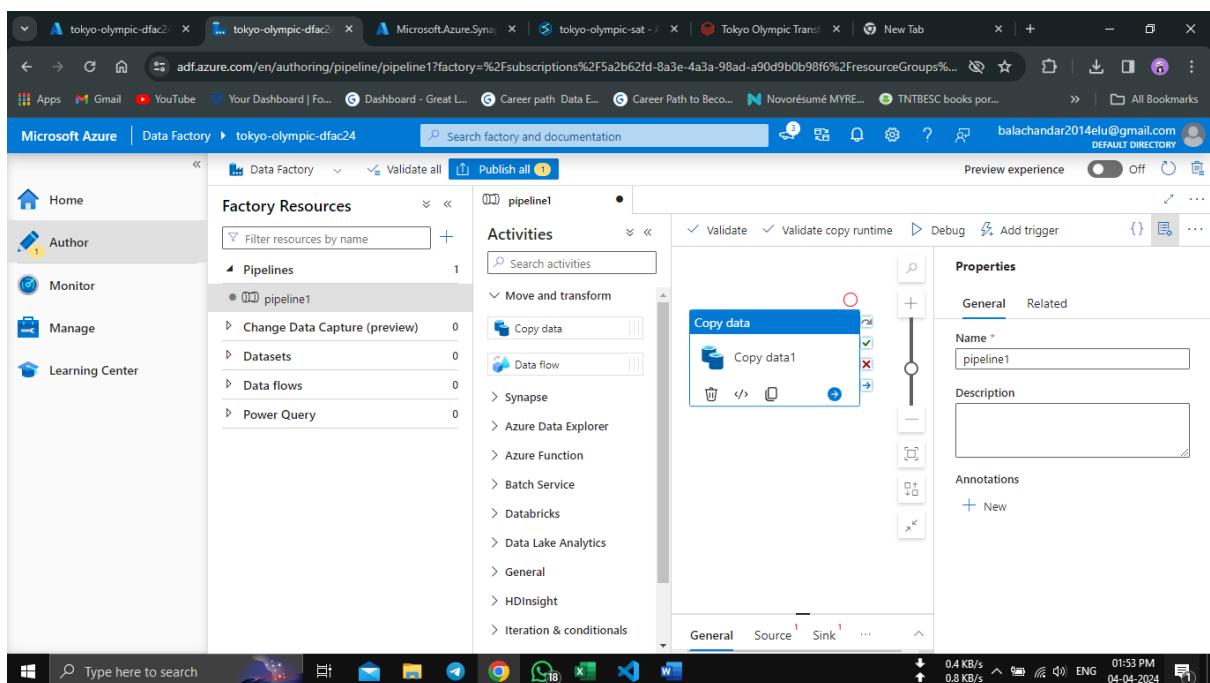
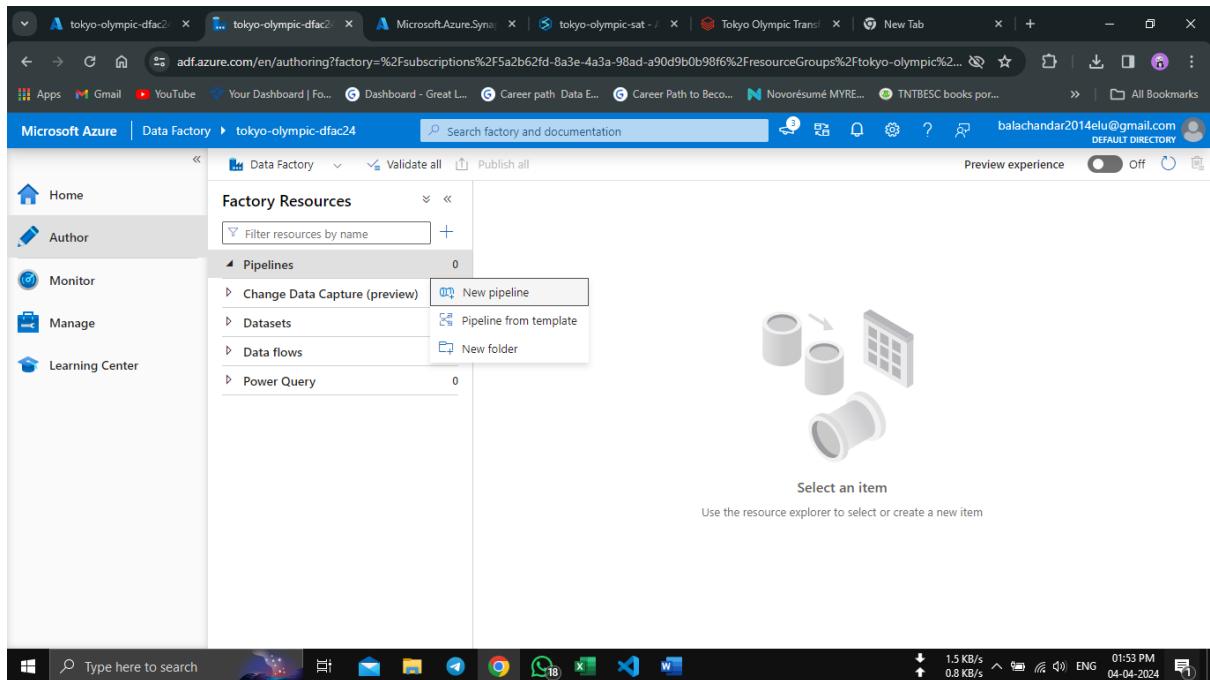
Created the Data Factory Studio for Data injection from GitHub(HTML) Source.



The screenshot shows the Microsoft Azure Data Factory studio interface. At the top, there's a navigation bar with 'Microsoft Azure' and 'Data Factory'. Below it, the workspace name 'tokyo-olympic-dfac24' is displayed. On the left, there's a sidebar with icons for Home, New, Ingest, Orchestrate, Transform data, and Configure SSIS. The main area features a large diagram illustrating data flow from various sources (represented by buildings) through a central processing stage (represented by a factory building) to sinks (represented by storage tanks). Below the diagram, four cards provide quick access to 'Ingest', 'Orchestrate', 'Transform data', and 'Configure SSIS'. At the bottom, there's a section for 'Recent resources'.

1. Data Ingestion using Azure Data Factory
2. Begin by creating an Azure Data Factory workspace within the previously established resource group.
3. After setting up the workspace, launch the Azure Data Factory Studio.
4. Upload the Tokyo Olympics dataset from kaggle to GitHub.
5. Within the studio, initialize a new data integration pipeline. Now use the task Copy Data to move data efficiently between various supported sources and destinations.
6. Configuring the Data Source with HTTP template as we are using http request to get the data from Github repo.
7. Establishing the Linked Service for source.
8. Configuring the File Format for and setting up the Linked Service Sink.
9. Repeat above steps to load all the datasets.

You can connect all the copy data activity together and run them all at once.



Created the pipeline → Used the Copy Data Activity → For Data Ingestion & pipeline creation..

The screenshot shows the Microsoft Azure Data Factory pipeline editor. A pipeline named 'Athletes' is displayed. The 'Source' tab is selected, showing a dataset named 'Athletes'. The pipeline has one activity, 'Copy data', which is configured to copy data from the 'Athletes' dataset.

Pipeline Details:

- Name:** Athletes
- Description:** (empty)
- Activity state:** Activated

Source Tab:

- Dataset:** Athletes
- Format:** (dropdown menu)
- Filter:** (checkbox)
- Row limit:** (checkbox)
- Max row count:** (text input)
- Max error count:** (text input)
- Sampling method:** (dropdown menu)
- Sampling percentage:** (text input)
- Sampling seed:** (text input)

Source Dataset Selection:

Select a data store

HTTP

Continue Cancel

Microsoft Azure | Data Factory > tokyo-olympic-dfac24 | Search factory and documentation | Publish all

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

Azure Cosmos DB (NoSQL)	Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2	Azure Database for MariaDB	Azure Database for MySQL
Azure Databricks	Azure Machine Learning	Azure Synapse Analytics

Continue Cancel

Microsoft Azure | Data Factory > tokyo-olympic-dfac24 | Search factory and documentation | Publish all

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

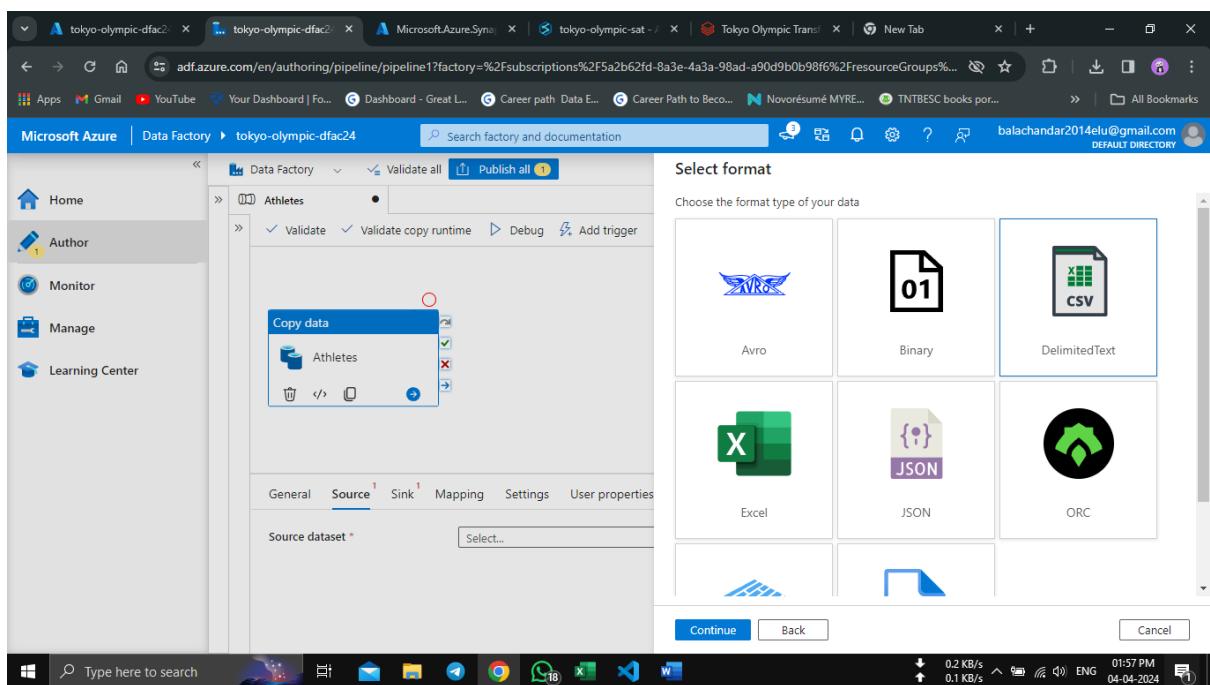
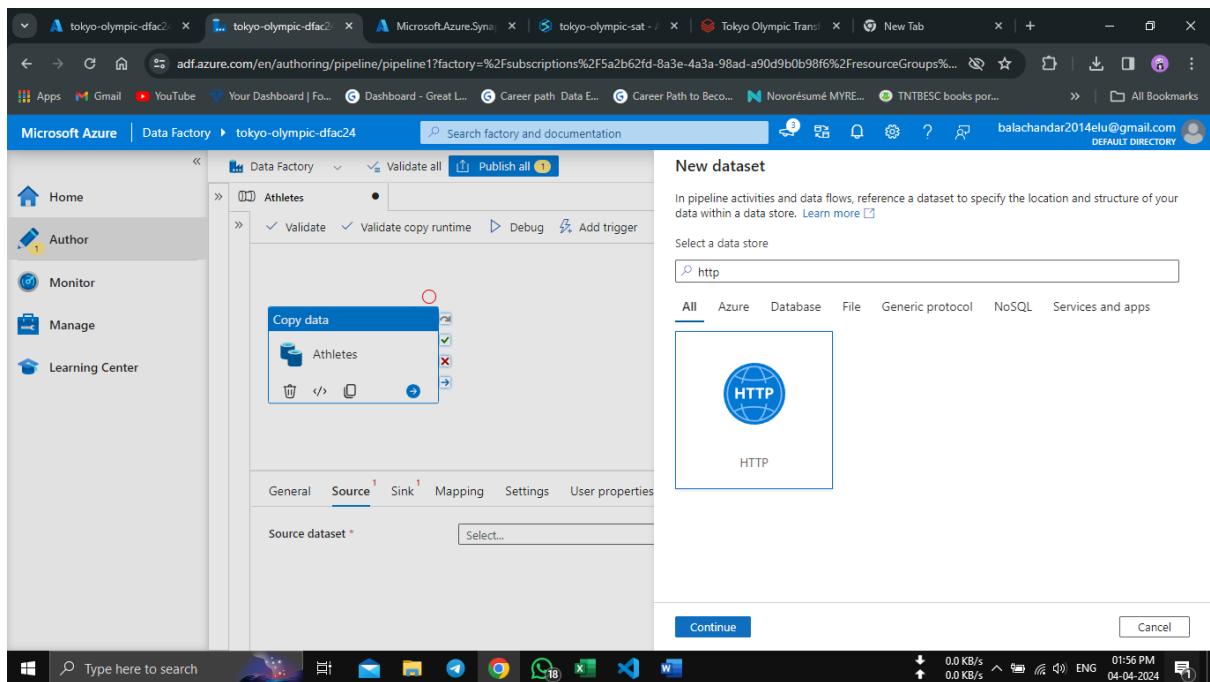
Select a data store

Search

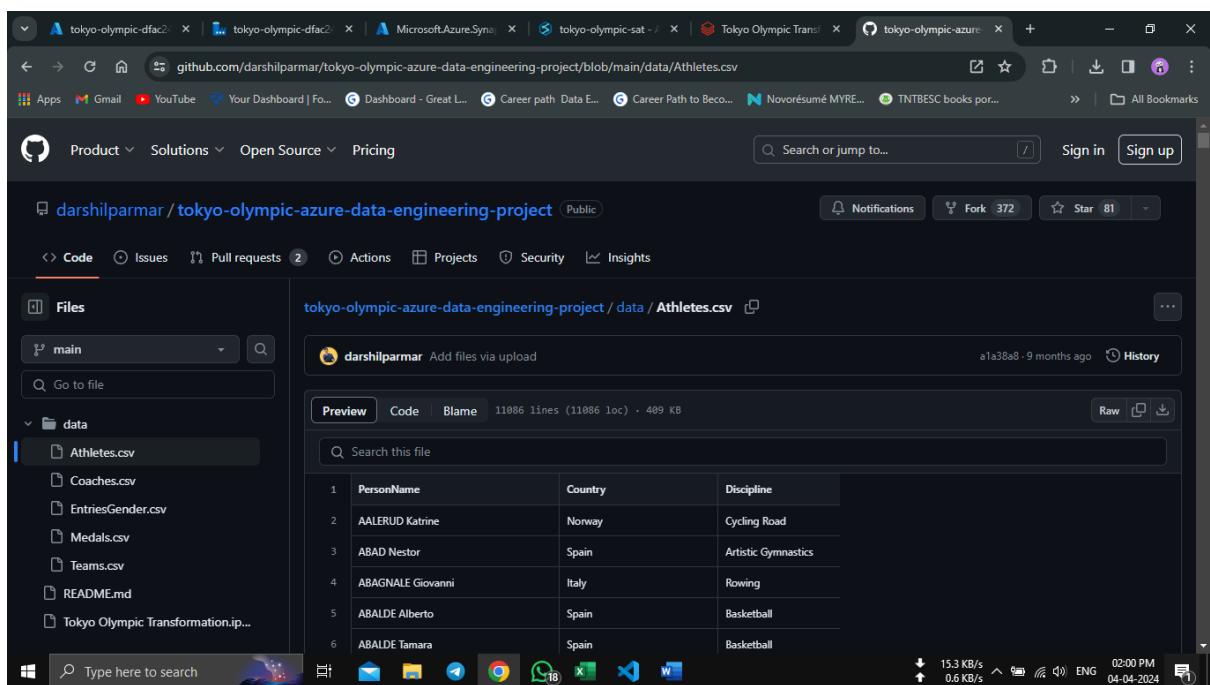
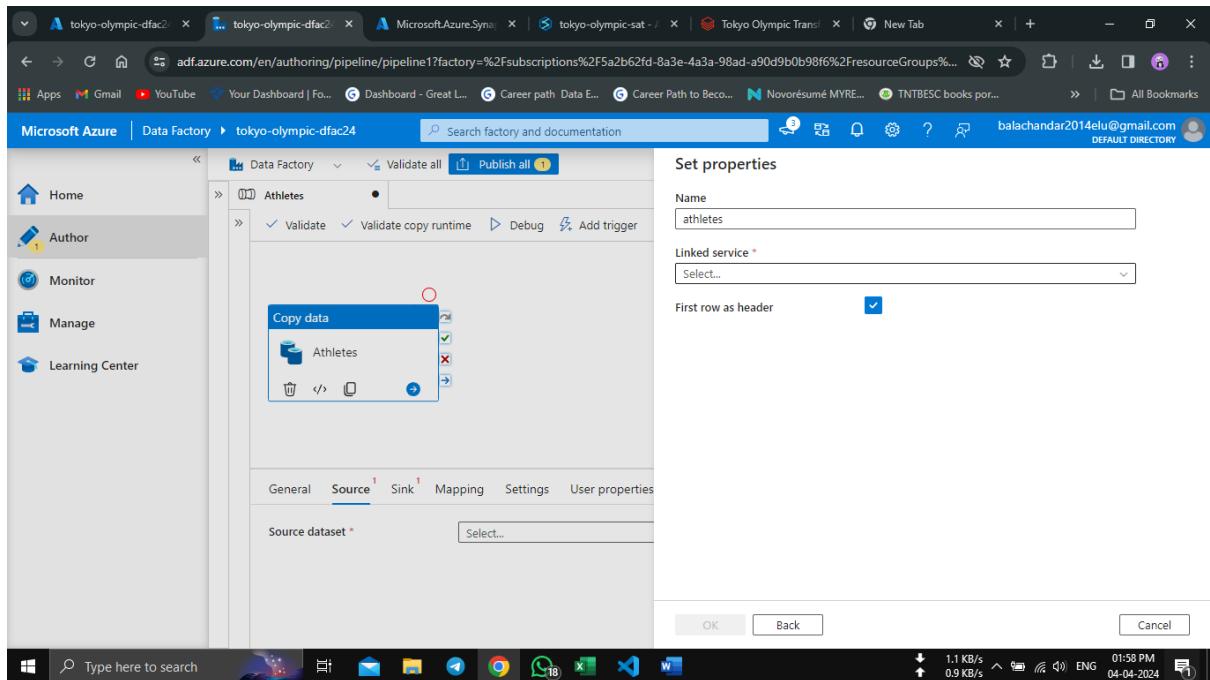
All Azure Database File Generic protocol NoSQL Services and apps

Amazon Marketplace Web Service	Amazon RDS for Oracle	Amazon RDS for SQL Server
Amazon Simple Queue Service (SQS)	Amazon Simple Table Storage (SimpleDB)	Amazon Simple Storage Service (Amazon S3)

Continue Cancel



Source file is in CSV. So, I selected the Delimited Text



GitHub Data Source Link : <https://raw.githubusercontent.com/darshilparmar/tokyo-olympic-azure-data-engineering-project/main/data/Athletes.csv>

PersonName	Country	Discipline
AALERUD Katrine	Norway	Cycling Road
ABAD Nestor	Spain	Artistic Gymnastics
ABAGNALE Giovanni	Italy	Rowing
ABALDE Alberto	Spain	Basketball
ABALDE Tamara	Spain	Basketball
ABALI Luc	France	Handball
ABAROA Cesar	Chile	Rowing
ABAS Abubakr	Sudan	Swimming
ABBASAI LI Hamdah	Islamic Republic of Iran	Karate
ABBENSONS Sander	Netherlands	Handball
ABBOTT Emily	Australia	Rhythmic Gymnastics
ABBOTT Monica	United States of America	Baseball/Softball
ABDALLA Abubaker	Haydar	Qatar,Athletics
ABDALLA Maryam	Egypt	Artistic Swimming
ABDALLAH Shahid	Egypt	Artistic Swimming
ABDALRASOOL Mohamed	Sudan	Judo
ABDEL LATIF Radwa	Egypt	Shooting
ABDEL RAZEK Samy	Egypt	Shooting
ABDELLAZIZ Abdalla	Egypt	Karate
ABDELLAZIZ Farah	Egypt	Table Tennis
ABDELLAZIZ Feryal	Egypt	Karate
ABDELMAGGOUD Mohamed	Egypt	Judo
ABDELMOTALEB Diaeladin Kamal	Gouda,Egypt	Wrestling
ABDELRHAMAN Thab	Egypt	Athletics
ABDELSALAM Mohamed	Egypt	Football
ABDELSALAM Tarek	Egypt	Handball
ABDELMWAHED Ahmed	Egypt	Athletics
ABDI Bashir	Belgium	Athletics
ABDURAHMAN Abdi	United States of America	Athletics
ABDUL HAID Farah Ann	Malaysia	Artistic Gymnastics
ABDUL RAHMAN Kiria Tilkanah	Singapore	Fencing
ABDUL RAZZAQ Fathimath	Nabaaha,Maldives	Badminton
ABDULRAHIMID Saud	Saudi Arabia	Football
ABDULJABBAR Ammar Riad	Germany	Boxing
ABDULLAEV Gulomjon	Uzbekistan	Wrestling
ABDULLAEV Muninjon	Uzbekistan	Wrestling
ABDULLAH Rahmat Erwin	Indonesia	Weightlifting
ABDULLAH TM Tifat	Kazakhstan	Archery

The screenshot shows the Microsoft Azure Data Factory pipeline editor. A 'Copy data' activity is selected within a pipeline named 'Athletes'. On the right side, a 'New linked service' dialog is open, titled 'AthletesHTTP'. It includes fields for 'Name' (set to 'AthletesHTTP'), 'Description' (empty), 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Base URL' (set to 'https://raw.githubusercontent.com/darshilparmar/tokyo-olympic-azure-data-engineering-project/main/data/Athletes.csv'), 'Server Certificate Validation' (set to 'Enable'), and 'Authentication type' (set to 'Anonymous'). At the bottom of the dialog are 'Create' and 'Cancel' buttons, along with a 'Test connection' link.

Copy and paste the data source URL and to create the Linked service between the Azure Data factory and GitHub api and ingest the Data from Github HTML API with the help of Linked service..

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

Azure AI Search Azure Blob Storage Azure Cosmos DB for MongoDB

Azure Data Lake Storage

Continue Cancel

Set properties

Name

Linked service *

File path / /

First row as header

Import schema From connection/store From sample file None

> Advanced

OK Back Cancel

Configure the SINK folder in the Azure Data Lake Gen2 Storage → create containers as Raw DATA, Transformed data.

The screenshot shows the Microsoft Azure Data Factory pipeline authoring interface. A 'Copy data' activity is selected in the main pane, configured to copy data from an 'Athletes' source dataset to a 'Sink dataset'. The 'Sink' tab is active, showing a dropdown menu where 'Select...' is chosen. A modal dialog titled 'Browse' is open, prompting the user to 'Select a file or folder'. The 'Root folder' is set to 'tokyo-olympic-data'. Inside this folder, there is one item named 'tokyo-olympic-data'. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

The second screenshot shows the same interface after the user has selected a different root folder. The 'Root folder' is now set to 'raw-data', which contains a subfolder named 'transformed-data'. Inside 'transformed-data', there is one item named 'raw-data'. The 'OK' and 'Cancel' buttons are visible at the bottom of the dialog.

create containers as Raw DATA, Transformed data .

The screenshot shows the Microsoft Azure Data Factory pipeline authoring interface. A 'Copy data' step is selected in the 'Athletes' dataset's 'Sink' tab. A modal dialog titled 'Copy data' is open, showing the 'Athletes' dataset as the sink. To the right, a 'Browse' window displays a list of CSV files in the 'raw-data' folder under 'Root folder > tokyo-olympic-data'. The files listed are: athletes.csv, coaches.csv, entriesgender.csv, medals.csv, and teams.csv. The 'OK' button is visible at the bottom of the modal.

Copy data

Athletes

athletes.csv
coaches.csv
entriesgender.csv
medals.csv
teams.csv

Showing 1 - 5 of 5 items

OK Cancel

Set properties

Name: Athletesdsf
Linked service: AzureDataLakeStorage1
File path: tokyo-olympic-data / raw-data / Athletes-r1
First row as header:
Import schema: From connection/store From sample file None
Advanced

OK Back Cancel

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Olympicdatipeline'. The main area displays a 'Copy data' step with two datasets: 'Athletes' (source) and 'Coachess' (sink). Below the step, tabs for 'General', 'Source', 'Sink', 'Mapping', 'Settings', and 'User properties' are visible. The 'Source' tab is selected. At the bottom, there's a 'Source dataset' dropdown with 'Select...' and 'New' options.

The screenshot shows the Microsoft Azure Storage account settings page for 'tokyoolympicdatachandru0'. The left sidebar includes sections for 'Data migration', 'Events', 'Storage browser', 'Data storage' (Containers, File shares, Queues, Tables), and 'Security + networking' (Networking, Access keys, Shared access signature, Encryption). The main pane displays the storage account's configuration, including resource group (move to 'tokyo-olympic'), location ('southeastasia'), primary/secondary location ('Primary: Southeast Asia, Secondary: East Asia'), subscription ('Free Trial'), and disk state ('Primary: Available, Secondary: Available'). It also shows tags ('Add tags') and a 'Properties' tab with sections for 'Data Lake Storage' (Hierarchical namespace: Enabled, Default access tier: Hot, Blob anonymous access: Disabled) and 'Security' (Require secure transfer for REST API operations: Enabled, Storage account key access: Enabled, Minimum TLS version: Version 1.2).

The screenshot shows two consecutive views of the Azure Storage Account interface.

Top Screenshot: The user is viewing the 'Containers' section of the 'tokyoolympicdatachandru0' storage account. The left sidebar shows navigation options like Data migration, Events, Storage browser, Data storage, and Containers (which is selected). The main area lists two containers: '\$logs' and 'tokyo-olympic-data'. Both are private and available.

Name	Last modified	Anonymous access level	Lease state
\$logs	1/4/2024, 1:16:50 pm	Private	Available
tokyo-olympic-data	1/4/2024, 1:20:23 pm	Private	Available

Bottom Screenshot: The user has navigated to the 'tokyo-olympic-data' container details page. The left sidebar shows Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Manage ACL, Access policy, Properties, Metadata), and a blob list. The main area shows the container's authentication method (Access key) and location (tokyo-olympic-data). It lists blobs: 'raw-data' and 'transformed-data', both of which are empty (Size: -).

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
raw-data					-	---
transformed-data					-	---

10. After the pipeline completes its execution, navigate to your Azure Data Lake Storage Gen2. Dive into the "raw_data" folder and validate that the files, like "athletes.csv", "medals.csv", etc., are present and populated with the expected data.

tokyo-olympic-data Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview Diagnose and solve problems Access Control (IAM)

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / raw-data

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[-]						***
athletes.csv	2/4/2024, 8:38:27 am	Hot (Inferred)		Block blob	408.68 KiB	Available
coaches.csv	2/4/2024, 8:38:40 am	Hot (Inferred)		Block blob	16.49 KiB	Available
entriesgender.csv	2/4/2024, 8:39:02 am	Hot (Inferred)		Block blob	1.1 KiB	Available
medals.csv	2/4/2024, 8:39:19 am	Hot (Inferred)		Block blob	2.36 KiB	Available
teams.csv	2/4/2024, 8:39:35 am	Hot (Inferred)		Block blob	34.44 KiB	Available

tokyo-olympic-data Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview Diagnose and solve problems Access Control (IAM)

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
raw-data					-	***
transformed-data					-	***

Check all the data available in the container.

The screenshot shows the Microsoft Azure Storage Container Overview page for the 'tokyo-olympic-data' container. The container has a single blob named 'athletes'. The table below lists the blob details:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
athletes	4/4/2024, 9:02:57 am	Hot (inferred)	Not yet archived	Block blob	112 B	Available

The screenshot shows the Microsoft Azure Storage Container Overview page for the 'athletes' blob within the 'tokyo-olympic-data' container. There are four log files: '_committed_475657...', '_started_4756574684...', '_SUCCESS', and 'part-00000-tid-4756...'. The table below lists these log file details:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
_committed_475657...	4/4/2024, 9:02:57 am	Hot (inferred)	Not yet archived	Block blob	112 B	Available
_started_4756574684...	4/4/2024, 9:02:57 am	Hot (inferred)	Not yet archived	Block blob	0 B	Available
_SUCCESS	4/4/2024, 9:02:57 am	Hot (inferred)	Not yet archived	Block blob	0 B	Available
part-00000-tid-4756...	4/4/2024, 9:02:57 am	Hot (inferred)	Not yet archived	Block blob	397.91 kB	Available



The screenshot shows the Microsoft Azure Storage Container Overview page for the 'tokyo-olympic-data' container. The container name is displayed at the top left. On the left, a sidebar menu includes 'Overview', 'Diagnose and solve problems', and 'Access Control (IAM)'. The main content area displays blob details with a table:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						***
_committed_202135...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	112 B	Available
_started_2021355756...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	0 B	Available
part-00000-tid-2021...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	16.12 KiB	Available



The screenshot shows the Microsoft Azure Storage Container Overview page for the 'tokyo-olympic-data' container. The container name is displayed at the top left. On the left, a sidebar menu includes 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', 'Properties', and 'Metadata'. The main content area displays blob details with columns: Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. A search bar at the top allows filtering by prefix. The table lists several blobs, including '_committed_338725...', '_started_338725566...', '_SUCCESS', and 'part-00000-tid-3387...'. Each blob entry includes a checkbox for selecting multiple items.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	112 B	Available
_committed_338725...	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	0 B	Available
_started_338725566...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	1.05 KiB	Available
part-00000-tid-3387...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	1.05 KiB	Available

This screenshot is identical to the one above, showing the Microsoft Azure Storage Container Overview page for the 'tokyo-olympic-data' container. It displays the same sidebar menu, blob list, and table structure. The table rows are identical, showing blobs like '_committed_338725...', '_started_338725566...', '_SUCCESS', and 'part-00000-tid-3387...' with their respective details.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	112 B	Available
_committed_338725...	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	0 B	Available
_started_338725566...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	4/4/2024, 9:02:59 am	Hot (Inferred)		Block blob	1.05 KiB	Available
part-00000-tid-3387...	4/4/2024, 9:02:58 am	Hot (Inferred)		Block blob	1.05 KiB	Available



The screenshot shows the Microsoft Azure Storage Container page for the 'tokyo-olympic-data' container. The left sidebar includes links for Overview, Diagnose and solve problems, Access Control (IAM), Shared access tokens, Manage ACL, Access policy, Properties, and Metadata. The main content area displays a list of blobs with columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. A search bar at the top allows filtering by prefix. The top navigation bar shows other tabs like Microsoft Azure Synapse Analytics and Tokyo Olympic Transformation.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]	4/4/2024, 9:03:00 am	Hot (Inferred)		Block blob	112 B	Available
_committed_217174...	4/4/2024, 9:03:00 am	Hot (Inferred)		Block blob	0 B	Available
_started_2171749806...	4/4/2024, 9:03:00 am	Hot (Inferred)		Block blob	0 B	Available
_SUCCESS	4/4/2024, 9:03:00 am	Hot (Inferred)		Block blob	33.79 KiB	Available
part-00000-tid-2171...	4/4/2024, 9:03:00 am	Hot (Inferred)		Block blob		

A screenshot of a Microsoft Edge browser window showing the Azure Storage Container overview for 'tokyo-olympic-data'. The container contains blobs named 'athletes', 'coaches', 'entriesgender', 'medals', and 'teams'.

Key details from the screenshot:

- Authentication method:** Access key (Switch to Microsoft Entra user account)
- Location:** tokyo-olympic-data / transformed-data
- Search:** Search blobs by prefix (case-sensitive)
- Show deleted objects:** Enabled
- Table Headers:** Name, Modified, Access tier, Archive status, Blob type, Size, Lease state
- Blob List:**

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]					-	***
athletes					-	***
coaches					-	***
entriesgender					-	***
medals					-	***
teams					-	***



Recent - Microsoft Azure | Microsoft.Azure.SynapseAnalytics | tokyo-olympic-sat - Azure | Tokyo Olympic Transformation | New Tab

portal.azure.com/?quickstart=true#view/HubsExtension/BrowseRecentResources.ReactView

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > Recent

Manage view Refresh Export to CSV Clear Assign tags

Group by none

Filter for any field... Subscription equals all Resource Group equals all Type equals all Location equals all Add filter

Name	Type	Location	Resource Group	Subscription	Last accessed
tokyoolympicdatachandru0	Storage account	Southeast Asia	tokyo-olympic	Free Trial	2 hours ago
tokyo-olympic-sat	Synapse workspace	Southeast Asia	tokyo-olympic	Free Trial	3 hours ago
tokyo-olympic	Resource group	Southeast Asia	tokyo-olympic	Free Trial	3 hours ago
tokyo-olympic-db	Azure Databricks Service	Southeast Asia	tokyo-olympic	Free Trial	6 hours ago
dbstoragegev7clui3qj2	Storage account	Southeast Asia	databricks-rg-tokyo-olympic	Free Trial	1 day ago
tokyo-olympic-dfac24	Data factory (V2)	Southeast Asia	tokyo-olympic	Free Trial	1 day ago

Showing 1 - 6 of 6 results.

[Give feedback](#)

tokyo-olympic-db - Microsoft Azure | Microsoft.Azure.SynapseAnalytics | tokyo-olympic-sat - Azure | Tokyo Olympic Transformation | New Tab | raw.githubusercontent.com | portal.azure.com/?quickstart=true#balachandar2014elugmail.onmicrosoft.com/resource/subscriptions/5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6#resource

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > tokyo-olympic-db

Azure Databricks Service

Search Delete

Overview

Activity log Access control (IAM) Tags Diagnose and solve problems

Settings

Virtual Network Peerings Encryption Networking Security & compliance Properties Locks Monitoring

Essentials

Status : Active Managed Resource Group : databricks-rg-tokyo-olympic-db-3kcdqgbg62qsu

Resource group : tokyo-olympic URL : https://adb-76715826667401844.azure.databricks.net

Location : Southeast Asia Pricing Tier : Premium (+ Role-based access controls) (Click to change)

Subscription : Free Trial

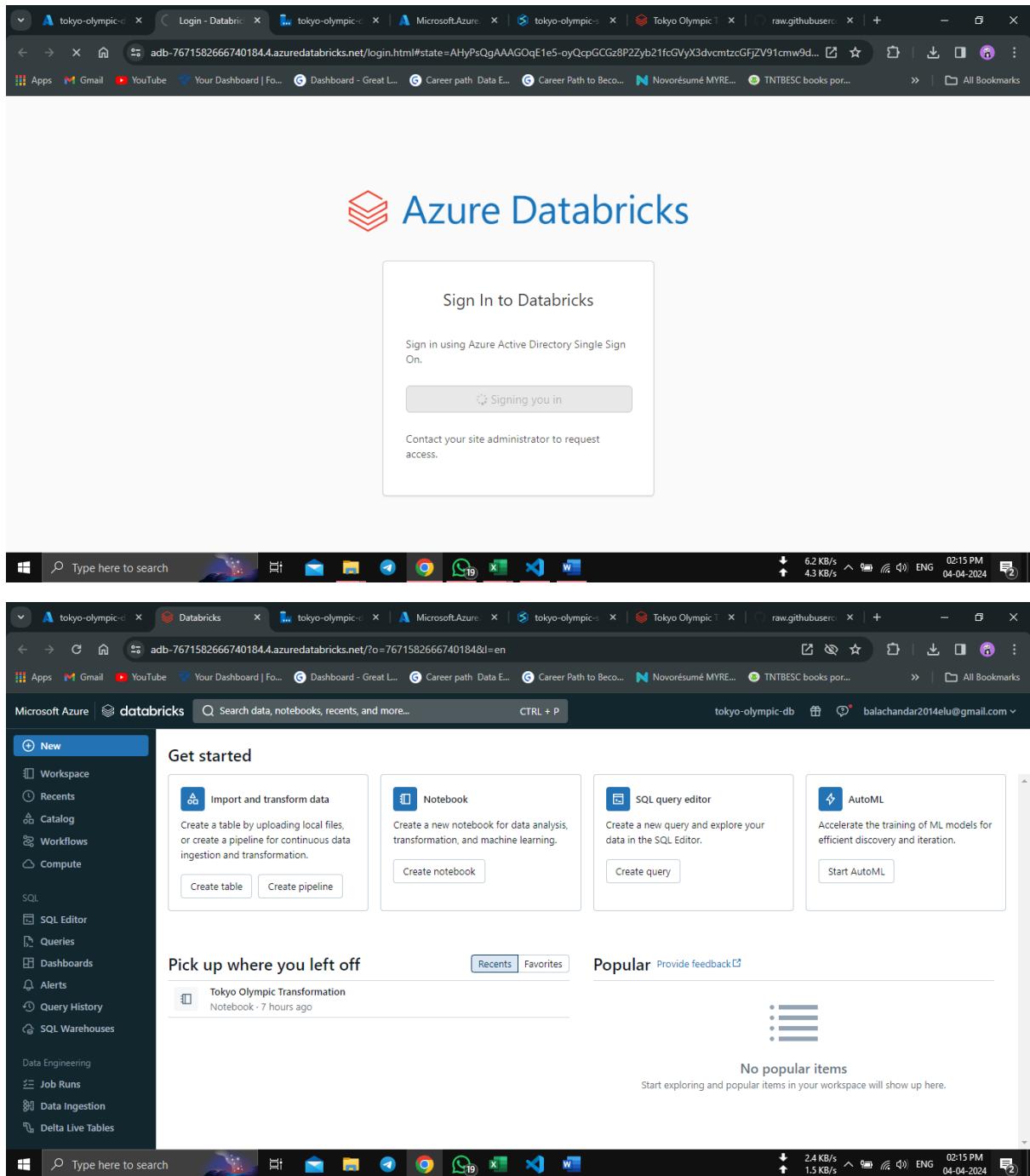
Subscription ID : 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6

Tags (edit) : Add tags

JSON View

Launch Workspace





Do this transformation by using Data Bricks.

Data Transformation using Azure Databricks

1. Navigate to Azure Databricks within the Azure portal and create a workspace within the previously established resource group and launch it.
2. Configuring Compute in Databricks
3. Create a new notebook within Databricks and rename it appropriately, reflecting its purpose or the dataset it pertains to.
4. Establishing a Connection to Azure Data Lake Storage (ADLS)

5. Using the credentials (Client ID, Tenant ID, Secret), write the appropriate code in the Databricks notebook to mount ADLS.
6. Writing Data Transformations mount ADLS Gen2 to Databricks.
7. Writing Transformed Data to ADLS Gen2.

```

from pyspark.sql.functions import col
from pyspark.sql.types import IntegerType, DoubleType, BooleanType, DataType

# Last execution failed
1 config = {"fs.azure.account.auth.type": "OAuth",
2 "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",
3 "fs.azure.account.oauth2.client.id": "5f5ff3d8-f05b-4829-bca3-f3b743300ace",
4 "fs.azure.account.oauth2.client.secret": "pu48Q~oqjTg2xsI9V7JHDHmwpTAOzVfSvkv0da9",
5 "fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/bfcf83e-1bb2-4d03-8384-8466413990a2/oauth2/token"
6
7 dbutils.fs.mount(
8 source = "abfss://tokyoolympic-data@tokyoolympicdatachandru0.dfs.core.windows.net", # contrainer@storageacc
9 mount_point = "/mnt/tokyoolympic",
10 extra_configs = config)
11
12 > java.rmi.RemoteException: java.lang.IllegalArgumentException: requirement failed: Directory already mounted: /mnt/tokyoolympic
13   mpir: nested exception is:

```

Create the Compute Resource & Mount the data into Data-Bricks.

path	name	size	modificationTime
dbfs:/mnt/tokyoolympic/raw-data/	raw-data/	0	1711957877000
dbfs:/mnt/tokyoolympic/transformed-data/	transformed-data/	0	1711957894000

spark

SparkSession - hive

SparkContext

Spark UI

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 1 hour ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share

spark

```
SparkSession - hive
SparkContext
SparkUI
Version
v3.3.2
Master
local[*, 4]
AppName
Databricks Shell
```

athletes = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/athletes.csv")
coaches = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/coaches.csv")
entriesgender = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/entriesgender.csv")

01:00 PM (9s) 5

01:06 PM 04-04-2024

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 1 hour ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share

```
athletes = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/athletes.csv")
coaches = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/coaches.csv")
entriesgender = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/entriesgender.csv")
medals = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/medals.csv")
teams = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-data/teams.csv")
```

(10) Spark Jobs

```
athletes: pyspark.sql.dataframe.DataFrame
    PersonName: string
    Country: string
    Discipline: string

coaches: pyspark.sql.dataframe.DataFrame = [Name: string, Country: string ... 2 more fields]

entriesgender: pyspark.sql.dataframe.DataFrame = [Discipline: string, Female: integer ... 2 more fields]

medals: pyspark.sql.dataframe.DataFrame = [Rank: integer, Team_Country: string ... 5 more fields]

teams: pyspark.sql.dataframe.DataFrame = [TeamName: string, Discipline: string ... 2 more fields]
```

01:00 PM (9s) 5

01:06 PM 04-04-2024

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 1 hour ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share |

athletes.show()

(1) Spark Jobs

PersonName	Country	Discipline
AALERUD Katrine	Norway	Cycling Road
ABAD Nestor	Spain	Artistic Gymnastics
ABAGNALE Giovanni	Italy	Rowing
ABALDE Alberto	Spain	Basketball
ABALDE Tamara	Spain	Basketball
ABALO Luc	France	Handball
ABAROA Cesar	Chile	Rowing
ABASS Abobakr	Sudan	Swimming
ABBASALI Hamideh	Islamic Republic ...	Karate
ABBASOV Islam	Azerbaijan	Wrestling
ABBINGH Lois	Netherlands	Handball
ABBOTT Emily	Australia	Rhythmic Gymnastics
ABBOTT Monica	United States of ...	Baseball/Softball
ABDALLA Abubaker ...	Qatar	Athletics
ABDALLA Maryam	Egypt	Artistic Swimming
ABDALLAH Shahd	Egypt	Artistic Swimming

0.5 KB/s 1.5 KB/s ENG 01:07 PM 04-04-2024

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 1 hour ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share |

athletes.printSchema()

```
root
 |-- PersonName: string (nullable = true)
 |-- Country: string (nullable = true)
 |-- Discipline: string (nullable = true)
```

coaches.printSchema()

```
root
 |-- Name: string (nullable = true)
 |-- Country: string (nullable = true)
 |-- Discipline: string (nullable = true)
 |-- Event: string (nullable = true)
```

0.5 KB/s 1.3 KB/s ENG 01:07 PM 04-04-2024

tokyo-olympic-data - Microsoft Edge | tokyo-olympic-sat - Microsoft Edge | tokyo-olympic-sat - Azure Databricks | Tokyo Olympic Transformation | New Tab

adb-7671582666740184.4.azuredatabricks.net/?o=7671582666740184#notebook/493846747091998/command/493846747092006

Apps Gmail YouTube Your Dashboard | Foo... Dashboard - Great L... Career path Data E... Career Path to Beco... Novorésumé MYRE... TNTBESC books por... All Bookmarks

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P tokyo-olympic-db Balachandar's Cluster Share

Tokyo Olympic Transformation Python 01:00 PM (1s) coaches.show()

(1) Spark Jobs

Name	Country	Discipline	Event
ABDELMAGID Wael	Egypt	Football	null
ABE Junya	Japan	Volleyball	null
ABE Katsuhiko	Japan	Basketball	null
ADAMA Cherif	Côte d'Ivoire	Football	null
AGEBA Yuya	Japan	Volleyball	null
AIKMAN Siegfried	Japan	Hockey	Men
AL SAIDI Kais	Germany	Hockey	Men
ALAHEDA Loni	Canada	Baseball/Softball	Softball
ALEKNO Vladimir	Islamic Republic of Iran	Volleyball	Men
ALEKSEEV Alexey	Russia	Handball	Women
ALLER CARBALLO Ma...	Spain	Basketball	null
ALSHEHRI Saad	Saudi Arabia	Football	Men
ALY Kamal	Egypt	Football	null
AMAYA GAITAN Fabian	Puerto Rico	Basketball	null
AND AGUADO Pablo	Spain	Football	null
ANDONOVSKI Vlatko	United States of America	Football	Women
ANNAN Alyson	Netherlands	Hockey	Women
ARNAU CREUS Xavier	Japan	Hockey	Women

0.7 KB/s 4.5 KB/s ENG 01:00 PM 04-04-2024

Type here to search

tokyo-olympic-data - Microsoft Edge | tokyo-olympic-sat - Microsoft Edge | tokyo-olympic-sat - Azure Databricks | Tokyo Olympic Transformation | New Tab

adb-7671582666740184.4.azuredatabricks.net/?o=7671582666740184#notebook/493846747091998/command/493846747092006

Apps Gmail YouTube Your Dashboard | Foo... Dashboard - Great L... Career path Data E... Career Path to Beco... Novorésumé MYRE... TNTBESC books por... All Bookmarks

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P tokyo-olympic-db Balachandar's Cluster Share

Tokyo Olympic Transformation Python 01:00 PM (1s) entriesgender.show()

(1) Spark Jobs

Discipline	Female	Male	Total
3x3 Basketball	32	32	64
Archery	64	64	128
Artistic Gymnastics	98	98	196
Artistic Swimming	105	0	105
Athletics	969	1072	2041
Badminton	86	87	173
Baseball/Softball	90	144	234
Basketball	144	144	288
Beach Volleyball	48	48	96
Boxing	102	187	289
Canoe Slalom	41	41	82
Canoe Sprint	123	126	249
Cycling BMX Freestyle	10	9	19
Cycling BMX Racing	24	24	48
Cycling Mountain Bike	38	38	76
Cycling Road	70	131	201
Cycling Track	90	99	189
Diving	72	71	143

0.9 KB/s 4.3 KB/s ENG 01:09 PM 04-04-2024

Type here to search

```
root
 |-- Discipline: string (nullable = true)
 |-- Female: integer (nullable = true)
 |-- Male: integer (nullable = true)
 |-- Total: integer (nullable = true)
```

```
entriesgender = entriesgender.withColumn("Female", col("Female").cast(IntegerType()))
 .withColumn("Male", col("Male").cast(IntegerType()))
 .withColumn("Total", col("Total").cast(IntegerType()))
```

```
entriesgender: pyspark.sql.dataframe.DataFrame
Discipline: string
Female: integer
Male: integer
Total: integer
```

```
root
 |-- Discipline: string (nullable = true)
 |-- Female: integer (nullable = true)
 |-- Male: integer (nullable = true)
 |-- Total: integer (nullable = true)
```

```
entriesgender.show()
```

Discipline	Female	Male	Total
Artistic Gymnastics	98	98	196
Artistic Swimming	105	0	105
Athletics	969	1072	2041
Badminton	86	87	173
Baseball/Softball	90	144	234
Basketball	144	144	288
Beach Volleyball	48	48	96
Boxing	102	187	289
Canoe Sprint	41	41	82

Data Transformation part

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar's Cluster | Share |

Tokyo Olympic Transformation | Python | 14

```
01:01 PM (<1s)
entriesgender.show()

+-----+-----+-----+
| Discipline|Female|Male|Total|
+-----+-----+-----+
| 3x3 Basketball| 32| 32| 64|
| Archery| 64| 64| 128|
| Artistic Gymnastics| 98| 98| 196|
| Artistic Swimming| 105| 0| 105|
| Athletics| 969| 1872| 2841|
| Badminton| 86| 87| 173|
| Baseball/Softball| 90| 144| 234|
| Basketball| 144| 144| 288|
| Beach Volleyball| 48| 48| 96|
| Boxing| 102| 187| 289|
| Canoe Slalom| 41| 41| 82|
| Canoe Sprint| 123| 126| 249|
| Cycling BMX Frees...| 10| 9| 19|
| Cycling BMX Racing| 24| 24| 48|
| Cycling Mountain ...| 38| 38| 76|
| Cycling Road| 70| 131| 201|
| Cycling Track| 90| 99| 189|
```

3.0 KB/s
28.1 KB/s
01:10 PM
04-04-2024

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar's Cluster | Share |

Tokyo Olympic Transformation | Python | 15

```
01:01 PM (<1s)
medals.printSchema()

root
 |-- Rank: integer (nullable = true)
 |-- Team_Country: string (nullable = true)
 |-- Gold: integer (nullable = true)
 |-- Silver: integer (nullable = true)
 |-- Bronze: integer (nullable = true)
 |-- Total: integer (nullable = true)
 |-- Rank by Total: integer (nullable = true)
```

1.5 KB/s
6.1 KB/s
01:10 PM
04-04-2024

The screenshot shows a Microsoft Edge browser window with the following details:

- Address Bar:** adb-7671582666740184.4.azuredatabricks.net/?o=7671582666740184#notebook/493846747091998/command/493846747092012
- Tab Bar:** Multiple tabs are open, including "tokyo-olympic-data - Microsoft Edge", "tokyo-olympic-sat - Microsoft Edge", "tokyo-olympic-sat - Azure Storage", "Tokyo Olympic Transformation - Microsoft Edge", and "New Tab".
- Header:** Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | Balachandar's Cluster | Schedule | Share | All Bookmarks | balachandar2014elu@gmail.com
- Left Sidebar (Workspace):**
 - + New
 - Workspace (selected)
 - Recents
 - Catalog
 - Workflows
 - Compute
 - SQL
 - SQL Editor
 - Queries
 - Dashboards
 - Alerts
 - Query History
 - SQL Warehouses
 - Data Engineering
 - Job Runs
 - Data Ingestion
 - Delta Live Tables
- Main Area:** Tokyo Olympic Transformation - Python notebook. The code cell content is:

```
medals.printSchema()
```

root
|-- Rank: integer (nullable = true)
|-- Team_Country: string (nullable = true)
|-- Gold: integer (nullable = true)
|-- Silver: integer (nullable = true)
|-- Bronze: integer (nullable = true)
|-- Total: integer (nullable = true)
|-- Rank by Total: integer (nullable = true)

medals.printSchema()

root
|-- Rank: integer (nullable = true)
|-- Team_Country: string (nullable = true)
|-- Gold: integer (nullable = true)
|-- Silver: integer (nullable = true)
|-- Bronze: integer (nullable = true)
- Status Bar:** 0.3 KB/s 1.0 KB/s ENG 04-04-2024 01:12 PM

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com | New Tab

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 2 hours ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share

01:01 PM (1s) | 19 | medals.printSchema()

```
root
|-- Rank: integer (nullable = true)
|-- Team_Country: string (nullable = true)
|-- Gold: integer (nullable = true)
|-- Silver: integer (nullable = true)
|-- Bronze: integer (nullable = true)
|-- Total: integer (nullable = true)
|-- Rank by Total: integer (nullable = true)
```

01:02 PM (1s) | 20 | medals.show()

(1) Spark Jobs

		Japan	27	14	17	58	5
3	Great Britain	22	21	22	65	4	
4	ROC	20	28	23	71	3	
5	Australia	17	7	22	46	6	
6	Netherlands	10	12	14	36	9	
7	France	10	12	11	33	10	

01:02 PM (1s) | 20 | medals.show()

(1) Spark Jobs

Rank	Team_Country	Gold	Silver	Bronze	Total	Rank by Total
1	United States of ...	39	41	33	113	1
2	People's Republic...	38	32	18	88	2
3	Japan	27	14	17	58	5
4	Great Britain	22	21	22	65	4
5	ROC	20	28	23	71	3
6	Australia	17	7	22	46	6
7	Netherlands	10	12	14	36	9
8	France	10	12	11	33	10
9	Germany	10	11	16	37	8
10	Italy	10	10	20	40	7
11	Canada	7	6	11	24	11
12	Brazil	7	6	8	21	12
13	New Zealand	7	6	7	20	13
14	Cuba	7	3	5	15	18
15	Hungary	6	7	7	20	13
16	Republic of Korea	6	4	10	20	13
17	Poland	4	5	5	14	19
18	Costa Rica	4	1	1	6	22

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 2 hours ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share

01:02 PM (1s) 21
medals.printSchema()
root
|-- Rank: integer (nullable = true)
|-- Team_Country: string (nullable = true)
|-- Gold: integer (nullable = true)
|-- Silver: integer (nullable = true)
|-- Bronze: integer (nullable = true)
|-- Total: integer (nullable = true)
|-- Rank by Total: integer (nullable = true)

01:02 PM (1s) 22
teams.show()
(1) Spark Jobs
+-----+-----+-----+
| TeamName| Discipline| Country| Event|
+-----+-----+-----+
Belgium	3x3 Basketball	Belgium	Men
China	3x3 Basketball	People's Republic...	Men
China	3x3 Basketball	People's Republic...	Women
France	3x3 Basketball	France	Women
Italy	3x3 Basketball	Italy	Women
Japan	3x3 Basketball	Japan	Men
Japan	3x3 Basketball	Japan	Women

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar2014elu@gmail.com

Tokyo Olympic Transformation | Python | File Edit View Run Help Last edit was 2 hours ago | New cell UI: ON | Run all | Balachandar's Cluster | Schedule | Share

01:02 PM (1s) 22
teams.show()
(1) Spark Jobs
+-----+-----+-----+
| TeamName| Discipline| Country| Event|
+-----+-----+-----+
Belgium	3x3 Basketball	Belgium	Men
China	3x3 Basketball	People's Republic...	Men
China	3x3 Basketball	People's Republic...	Women
France	3x3 Basketball	France	Women
Italy	3x3 Basketball	Italy	Women
Japan	3x3 Basketball	Japan	Men
Japan	3x3 Basketball	Japan	Women
Latvia	3x3 Basketball	Latvia	Men
Mongolia	3x3 Basketball	Mongolia	Women
Netherlands	3x3 Basketball	Netherlands	Men
Poland	3x3 Basketball	Poland	Men
ROC	3x3 Basketball	ROC	Men
ROC	3x3 Basketball	ROC	Women
Romania	3x3 Basketball	Romania	Women
Serbia	3x3 Basketball	Serbia	Men
United States	3x3 Basketball	United States of ...	Women
Australia	Archery	Australia	Men's Team
Australia	Archery	Australia	Mixed Team

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar's Cluster | Schedule | Share |

Tokyo Olympic Transformation | Python | 23

```
teams.printSchema()
```

```
root
 |-- TeamName: string (nullable = true)
 |-- Discipline: string (nullable = true)
 |-- Country: string (nullable = true)
 |-- Event: string (nullable = true)
```

01:02 PM (1s) | 24

```
# Find the top countries with the highest number of gold medals
top_gold_medal_countries = medals.orderBy("Gold", ascending=False).show()
```

(1) Spark Jobs

		Japan	27	14	17	58	5
3	Great Britain	22	21	22	65	4	
4	ROC	20	28	23	71	3	
5	Australia	17	7	22	46	6	
6	Netherlands	10	12	14	36	9	
7	France	10	12	11	33	10	
8	Germany	10	11	16	37	8	
9	Italy	10	10	20	40	7	
10	Canada	7	6	11	24	11	
11	Brazil	7	6	8	21	12	
12	New Zealand	7	6	7	20	13	
13	Cuba	7	3	5	15	18	
14	Hungary	6	7	7	20	13	
15	Republic of Korea	6	4	10	20	13	
16	Poland	4	5	5	14	19	
17	Czech Republic	4	4	3	11	23	
18	Kenya	4	4	2	10	25	
19	Norway	4	2	2	8	29	
20							

01:02 PM (1s) | 24

```
# Find the top countries with the highest number of gold medals
top_gold_medal_countries = medals.orderBy("Gold", ascending=False).show()
```

(1) Spark Jobs

		Japan	27	14	17	58	5
3	Great Britain	22	21	22	65	4	
4	ROC	20	28	23	71	3	
5	Australia	17	7	22	46	6	
6	Netherlands	10	12	14	36	9	
7	France	10	12	11	33	10	
8	Germany	10	11	16	37	8	
9	Italy	10	10	20	40	7	
10	Canada	7	6	11	24	11	
11	Brazil	7	6	8	21	12	
12	New Zealand	7	6	7	20	13	
13	Cuba	7	3	5	15	18	
14	Hungary	6	7	7	20	13	
15	Republic of Korea	6	4	10	20	13	
16	Poland	4	5	5	14	19	
17	Czech Republic	4	4	3	11	23	
18	Kenya	4	4	2	10	25	
19	Norway	4	2	2	8	29	
20							

tokyo-olympic-data - Microsoft Edge | tokyo-olympic-sat - Microsoft Edge | tokyo-olympic-sat - Azure SQL | Tokyo Olympic Transformation | New Tab

adb-7671582666740184.4.azuredatabricks.net/?o=7671582666740184#notebook/493846747091998/command/1000253047504383

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar's Cluster | Schedule | Share | tokyo-olympic-db | balachandar@gmail.com

Tokyo Olympic Transformation | Python | 01:02 PM (1s) | 25 | Run all | Balachandar's Cluster | Schedule | Share |

Find the top countries with the highest number of gold medals
top_gold_medal_countries = medals.orderBy("Gold", ascending=False).select("Team_country", "Gold").show()

(1) Spark Jobs

Team_country	Gold
United States of America	39
People's Republic of China	38
Japan	27
Great Britain	22
ROC	20
Australia	17
Netherlands	10
France	10
Germany	10
Italy	10
Canada	7
Brazil	7
New Zealand	7
Cuba	7
Hungary	6
Republic of Korea	6
Poland	4

4.4 KB/s 3.0 KB/s ENG 01:19 PM 04-04-2024

tokyo-olympic-data - Microsoft Edge | tokyo-olympic-sat - Microsoft Edge | tokyo-olympic-sat - Azure SQL | Tokyo Olympic Transformation | New Tab

adb-7671582666740184.4.azuredatabricks.net/?o=7671582666740184#notebook/493846747091998/command/1000253047504387

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | tokyo-olympic-db | balachandar's Cluster | Schedule | Share |

Tokyo Olympic Transformation | Python | 01:02 PM (1s) | 26 | Run all | Balachandar's Cluster | Schedule | Share |

Calculate the average number of entries by gender for each discipline
average_entries_by_gender = entriesgender.withColumn("Avg_Female", entriesgender["Female"] / entriesgender["Total"]).withColumn("Avg_Male", entriesgender["Male"] / entriesgender["Total"])
average_entries_by_gender.show()

(1) Spark Jobs

Discipline	Female	Total	Avg_Female	Avg_Male
3x3 Basketball	32	32	0.5	0.5
Archery	64	64	0.5	0.5
Artistic Gymnastics	98	98	0.5	0.5
Artistic Swimming	105	105	1.0	0.0
Athletics	969	1072	0.4747672709456149	0.5252327290543851
Badminton	86	87	0.49710982658959535	0.5028901734104047
Baseball/Softball	90	144	0.38461538461538464	0.6153846153846154
Basketball	144	144	0.5	0.5
Beach Volleyball	48	48	0.5	0.5
Boxing	102	187	0.35294117647058826	0.6470588235294118

0.5 KB/s 0.4 KB/s ENG 01:20 PM 04-04-2024

Do some transformation → Calculate the Average Number of Entries for each discipline.

The screenshot shows a Microsoft Azure Databricks workspace. The left sidebar includes options like New, Workspace, Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, and SQL Warehouses. The main area displays a notebook titled "Tokyo Olympic Transformation" in Python. The code cell contains the command `entriesgender.show()`. The resulting DataFrame is as follows:

Discipline	Female	Male	Total
3x3 Basketball	32	32	64
Archery	64	64	128
Artistic Gymnastics	98	98	196
Artistic Swimming	105	0	105
Athletics	969	1072	2041
Badminton	86	87	173
Baseball/Softball	90	144	234
Basketball	144	144	288
Beach Volleyball	48	48	96
Boxing	102	187	289
Canoe Slalom	41	41	82
Canoe Sprint	123	126	249
Cycling BMX Freestyle	10	9	19
Cycling BMX Racing	24	24	48
Cycling Mountain Bike	38	38	76
Cycling Road	76	131	201

Tokyo Olympic Transformation

```
athletes.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/athletes")
coaches.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/coaches")
entriesgender.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/entriesgender")
medals.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/medals")
teams.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/teams")
```

medals.show()

Rank	Team_Country	Gold	Silver	Bronze	Total	Rank by Total
1	United States of ...	39	41	33	113	1
2	People's Republic...	38	32	18	88	2
3	Japan	27	14	17	58	5
4	Great Britain	22	21	22	65	4

Tokyo Olympic Transformation | Python | 01:03 PM (1s) | 29

```
medals.show()
```

(1) Spark Jobs

Rank	Team_Country	Gold	Silver	Bronze	Total	Rank by Total
1	United States of ...	39	41	33	113	1
2	People's Republic...	38	32	18	88	2
3	Japan	27	14	17	58	3
4	Great Britain	22	21	22	65	4
5	ROC	20	28	23	71	5
6	Australia	17	7	22	46	6
7	Netherlands	10	12	14	36	9
8	France	10	12	11	33	10
9	Germany	10	11	16	37	8
10	Italy	10	10	20	40	7
11	Canada	7	6	11	24	11
12	Brazil	7	6	8	21	12
13	New Zealand	7	6	7	20	13
14	Cuba	7	3	5	15	18
15	Hungary	6	7	7	20	13
16	Republic of Korea	6	4	10	20	13
17	Poland	4	5	5	14	19
18	Czech Republic	41	41	31	111	23

Microsoft Azure Synapse Analytics - Overview | Overview | 01:21 PM 04-04-2024

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20240404094923 Start time : 4/4/2024, 10:00:15 am
Subscription : Free Trial Correlation ID : 7aa29a7b-7adc-4de2-b431-0120837...

Resource group : tokyo-olympic

Deployment details

Next steps

Go to resource group

Give feedback

Tell us about your experience with deployment

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.

Set up cost alerts >

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

The screenshot shows the Microsoft Azure portal interface. The URL in the address bar is <https://portal.azure.com/?quickstart=true#@balachandar2014elugmail.onmicrosoft.com/resource/subscriptions/5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6/resourceGroups/tokyo-olympic/providers/Microsoft.Synapse/workspaces/tokyo-olympic-sat/>. The page displays the 'Essentials' section of the 'tokyo-olympic-sat' Synapse workspace. Key details shown include:

- Resource group: tokyo-olympic
- Status: Succeeded
- Location: Southeast Asia
- Subscription: Free Trial
- Subscription ID: 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6
- Managed virtual network: No
- Managed identity object ID: b167b0c7-15d2-4321-9eb8-3b5da271f540
- Workspace web URL: <https://web.azuresynapse.net?workspace=%2fsubs...>
- Tags: Add tags
- Networking: Show firewall settings
- Primary ADLS Gen2 account: https://tokyolympicdatachandru0.dfs.core.windows.net
- Primary ADLS Gen2 file system: tokyo-olympic-data
- SQL admin username: sqldadminuser
- SQL Microsoft Entra admin: live.com#balachandar2014elu@gmail.com
- Dedicated SQL endpoint: tokyo-olympic-sat.sql.azuresynapse.net
- Serverless SQL endpoint: tokyo-olympic-sat-on-demand.sql.azuresynapse.net
- Development endpoint: https://tokyo-olympic-sat.dev.azuresynapse.net

Data Loading & Analytics by using Azure Synapse Analytics Studio..

The screenshot shows a browser window with the URL <https://web.azuresynapse.net/en/?workspace=%2fsubscriptions%2f5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6%2fresourceGroups%2ftokyo-olympic%2fproviders%2fMicrosoft.Synapse%2fworkspaces%2ftokyo-olympic-sat/>. The page is currently loading, displaying a large blue 'Loading' screen.

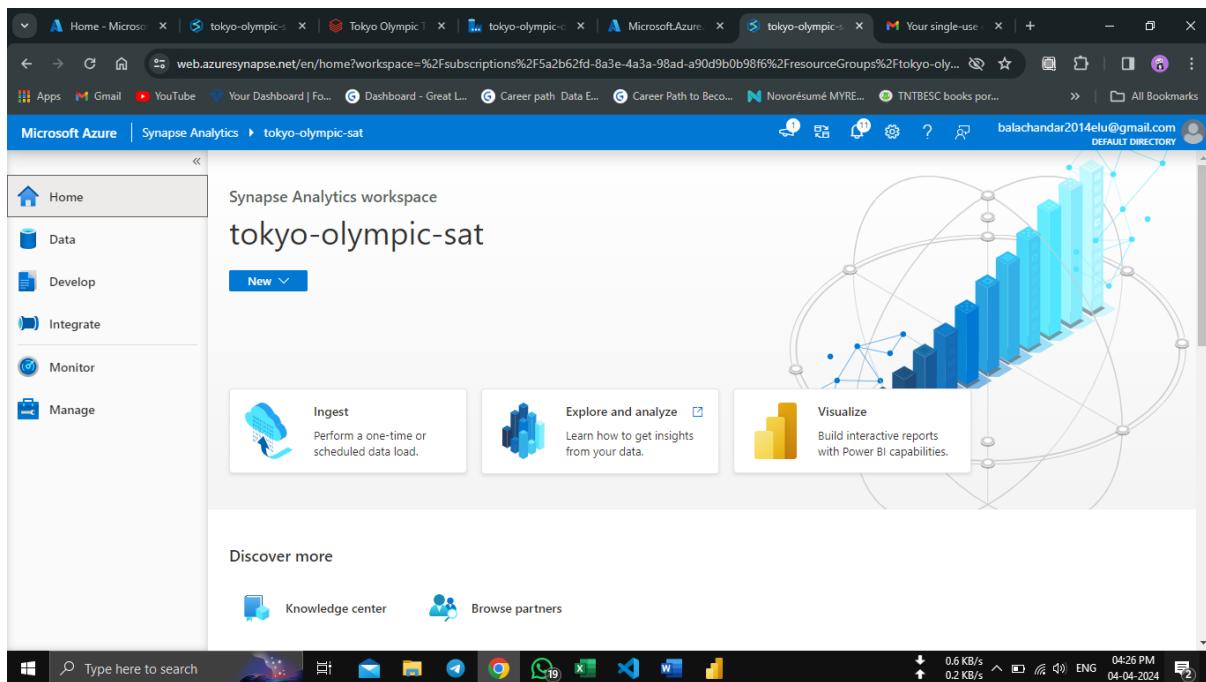
Setting Up and Using Azure Synapse Analytics

1. Creating a Synapse Analytics Workspace.
2. Within Workspace navigate to the "Data" section, choose "Lake Database" and create a Database "TokyoOlympicDB"
3. Creating Table from Data Lake from the Transformed Data folder within your ADLS Gen2 storage.

Screenshot of Microsoft Azure Synapse Analytics workspace titled "tokyo-olympic-sat".

The workspace interface includes:

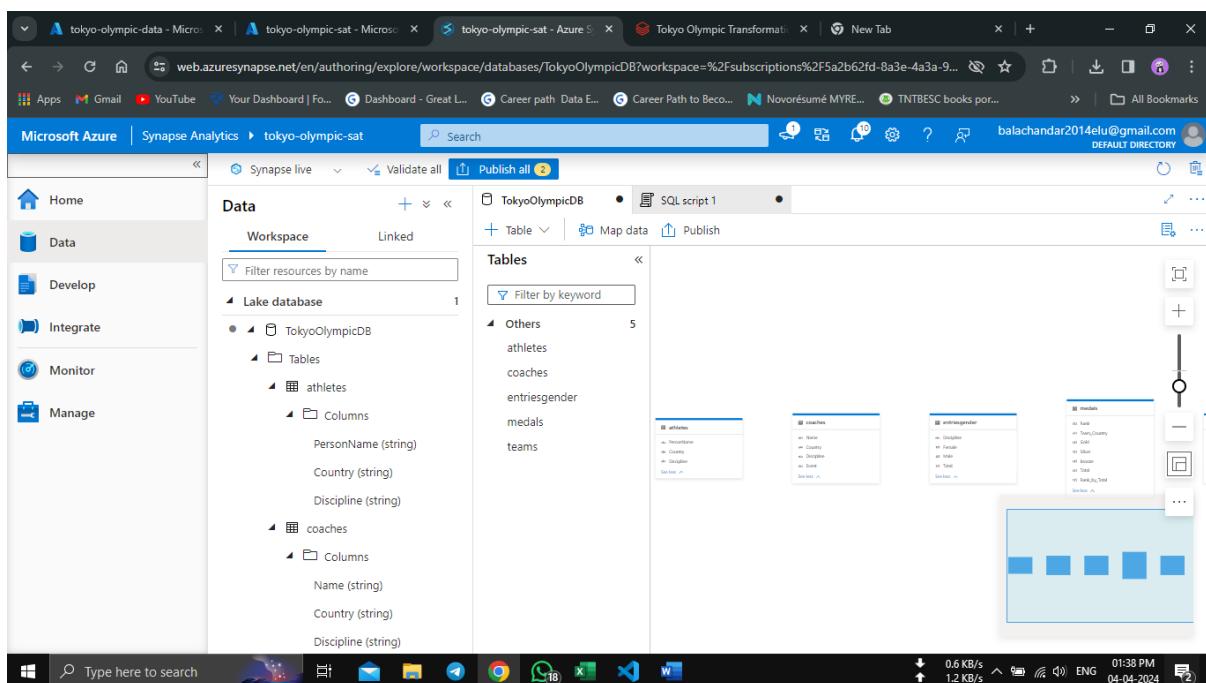
- Left sidebar:** Home, Data, Develop, Integrate, Monitor, Manage.
- Main area:** "Synapse Analytics workspace" and "tokyo-olympic-sat".
- Buttons:** Ingest, Explore and analyze, Visualize.
- Discover more:** Knowledge center, Browse partners.



Screenshot of Microsoft Azure Synapse Analytics workspace titled "tokyo-olympic-sat".

The workspace interface includes:

- Left sidebar:** Home, Data, Develop, Integrate, Monitor, Manage.
- Main area:** "Data" section showing "Lake database" named "TokyoOlympicDB".
- Tables:** athletes, coaches, entriesgender, medals, teams.
- Visualizations:** A small preview of a bar chart.



This screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The left sidebar includes links for Home, Data, Develop, Integrate, Monitor, and Manage. The main area displays five tables: athletes, coaches, entriesgender, medals, and teams. The medals table is currently selected, showing columns for Rank, Team_Country, Gold, Silver, Bronze, Total, and Rank_by_Total. Below the tables is a small chart consisting of blue squares. At the top, there are tabs for Synapse live, Validate all, and Publish all, along with a search bar and various navigation icons.

This screenshot shows the Microsoft Azure Synapse Analytics workspace interface, specifically the Data section. The left sidebar shows the same navigation links as the previous screenshot. The main area features a query editor titled "SQL script 1" under the "TokyoOlympicDB" workspace. The editor contains the following SQL code:

```
2 SELECT Country, COUNT(*) AS TotalAthletes
3 FROM athletes
4 GROUP BY Country
5 ORDER BY TotalAthletes DESC;
6
7 --Calculate the total medals won by each country:
8 SELECT
9 Team_Country,
10 SUM(GOLD) Total_Gold,
11 SUM(Silver) Total_Silver,
12 SUM(Bronze) Total_Bronze
13 FROM medals
14 GROUP BY Team_Country
15 ORDER BY Total_Gold DESC;
16
17 --Calculate the average number of entries by gender for each discipline:
18 SELECT Discipline,
19 Avg(Female) Avg_Female,
20 Avg(Male) Avg_Male
21 FROM entriesgender
22 GROUP BY Discipline;
```

The results of the query execution are shown below the code, indicating a successful execution time of 0.00007 seconds.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. On the left, there's a navigation sidebar with options like Home, Data, Develop, Integrate, Monitor, and Manage. The main area is titled "Data" and shows a "Workspace" section with a "Lake database" named "TokyoOlympicDB". Inside this database, there are two tables: "athletes" and "coaches". The "athletes" table has columns for PersonName (string), Country (string), and Discipline (string). The "coaches" table has columns for Name (string), Country (string), and Discipline (string). A SQL script editor window is open, titled "TokyoOlympicDB" and "SQL script 1". The script contains the following code:

```
2 SELECT Country, COUNT(*) AS TotalAthletes
3 FROM athletes
4 GROUP BY Country
5 ORDER BY TotalAthletes DESC;
6
7 --Calculate the total medals won by each country:
8 SELECT
9 Team_Country,
10 SUM(GOLD) Total_Gold,
11 SUM(Silver)Total_Silver,
12 SUM(Bronze)Total_Bronze
13 FROM medals
14 GROUP BY Team_Country
15 ORDER BY Total_Gold DESC;
16
17 --Calculate the average number of entries by gender for each discipline:
18 SELECT Discipline,
19 Avg(Female) Avg_Female,
20 Avg(Male) Avg_Male
21 FROM entriesgender
22 GROUP BY Discipline;
```

The "Results" tab at the bottom shows a message: "00:00:07 Query executed successfully." The status bar at the bottom right indicates the date and time as "04-04-2024 01:36 PM".

This screenshot is nearly identical to the one above, showing the Microsoft Azure Synapse Analytics workspace. The main difference is in the SQL script editor. The "Use database" dropdown menu at the top right is now set to "TokyoOlympicDB". The rest of the interface, including the workspace structure, the tables, and the SQL code, remains the same.

Calculate the total medals won by each country,

Calculate the average number of entries by gender for discipline by Using sql queries.

Microsoft Azure | Synapse Analytics > tokyo-olympic-sat

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

```
6 --Calculate the total medals won by each country:
7 SELECT
8 Team_Country,
9 SUM(GOLD) Total_Gold,
10 SUM(Silver) Total_Silver,
11 SUM(Bronze) Total_Bronze
12 FROM medals
13 GROUP BY Team_Country
14 ORDER BY Total_Gold DESC;
```

Results Messages

View Table Chart Export results

Team_Country	Total_Gold	Total_Silver	Total_Bronze
United States of America	39	41	33
People's Republic of China	38	32	18

00:00:03 Query executed successfully.

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 94 bytes

Results settings per query

First 5000 rows (default)

All rows

0.9 KB/s 1.5 KB/s 12:01 PM 04-04-2024

Microsoft Azure | Synapse Analytics > tokyo-olympic-sat

SQL script 1

1 --Count the number of Athletes from each Country:

```
2 SELECT Country, COUNT(*) AS TotalAthletes
3 FROM athletes
4 GROUP BY Country
5 ORDER BY TotalAthletes DESC;
6
```

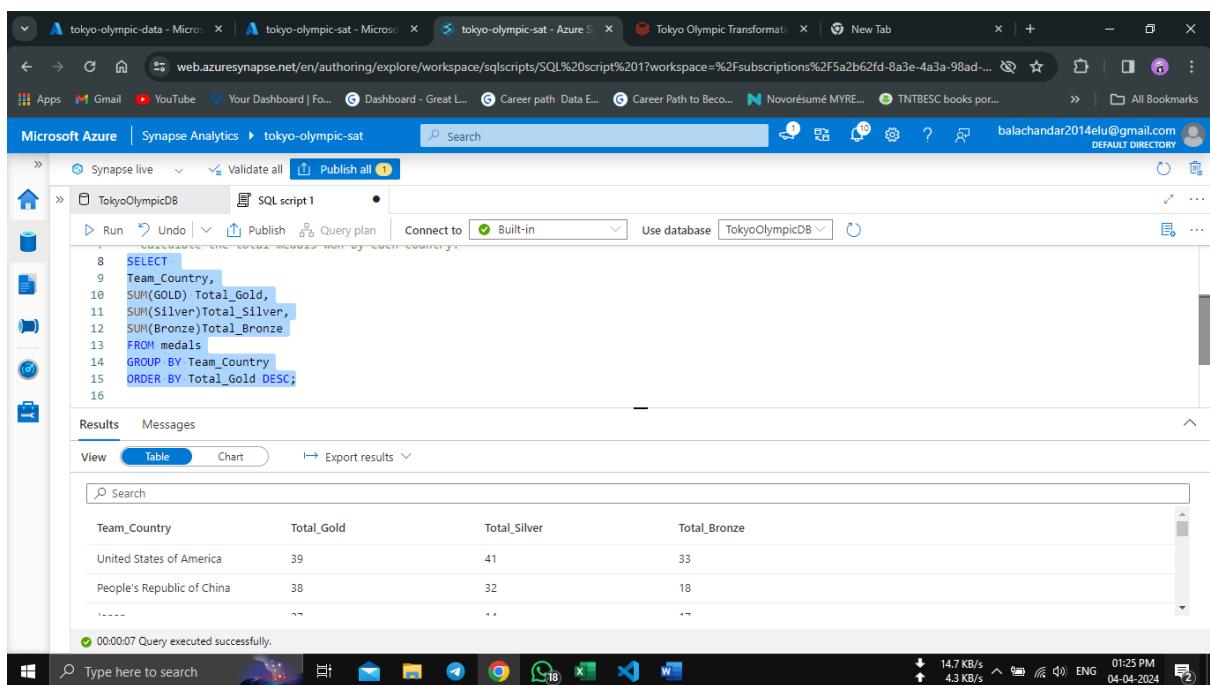
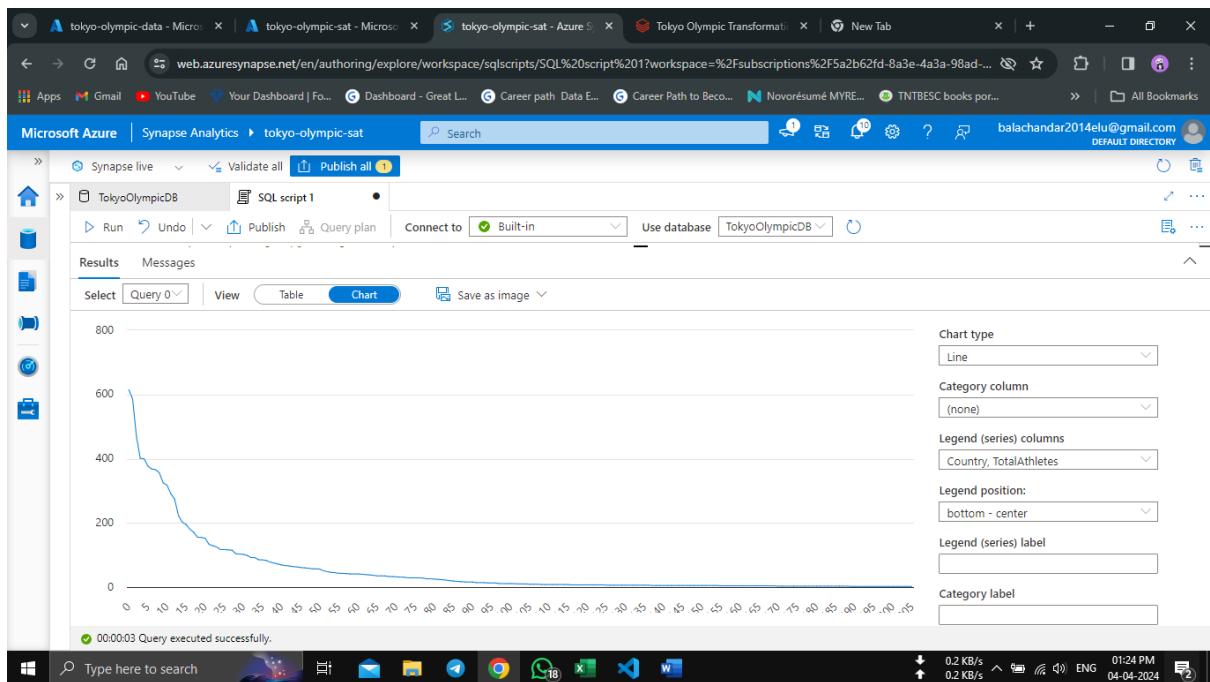
Results Messages

Select Query 0 View Table Chart Export results

Country	TotalAthletes
United States of America	615
Japan	586
Australia	470
People's Republic of China	401
Germany	400

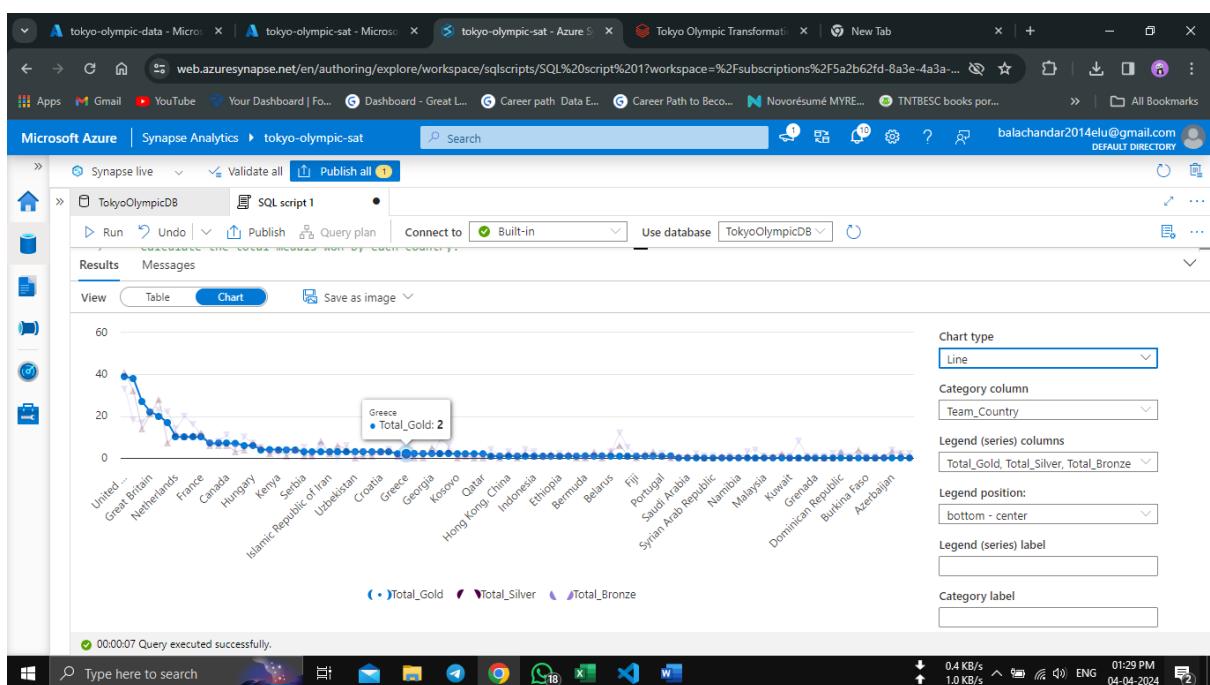
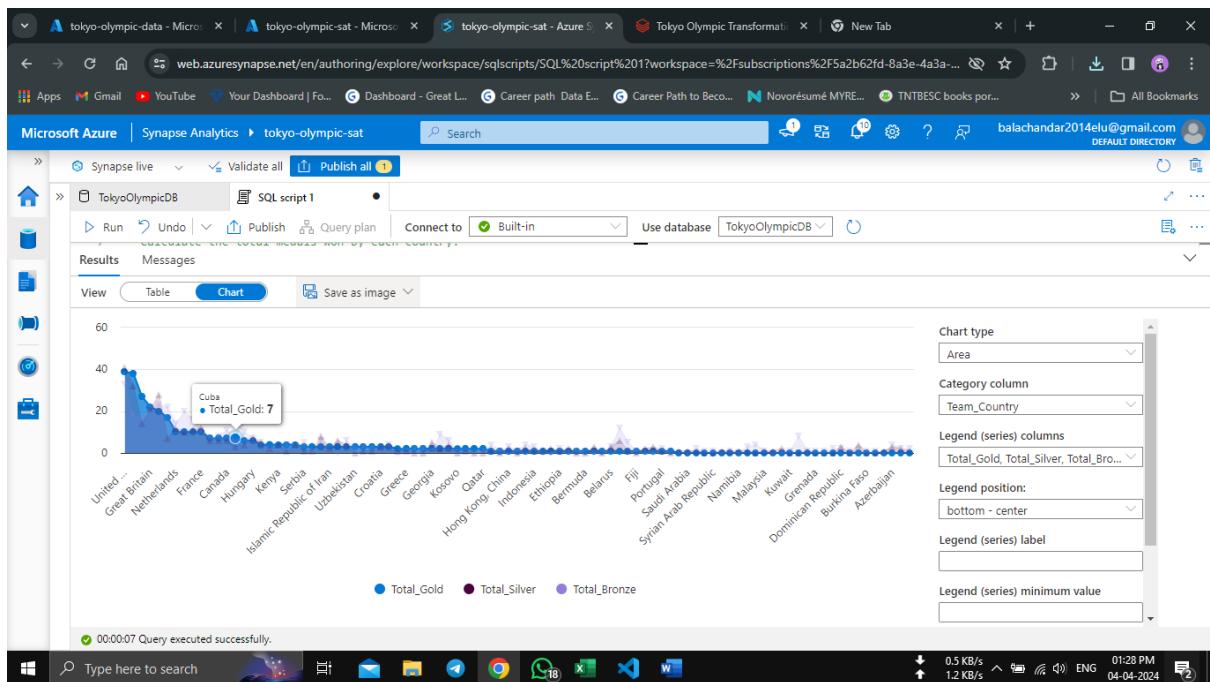
00:00:03 Query executed successfully.

1.2 KB/s 1.5 KB/s 01:23 PM 04-04-2024

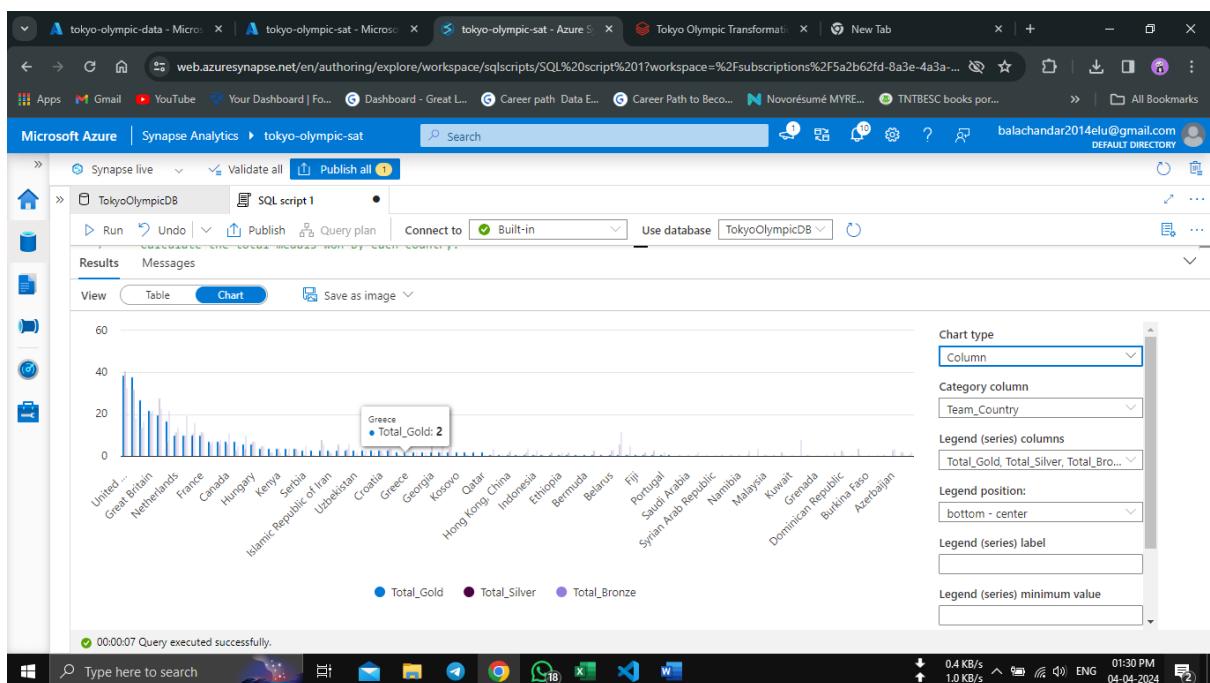
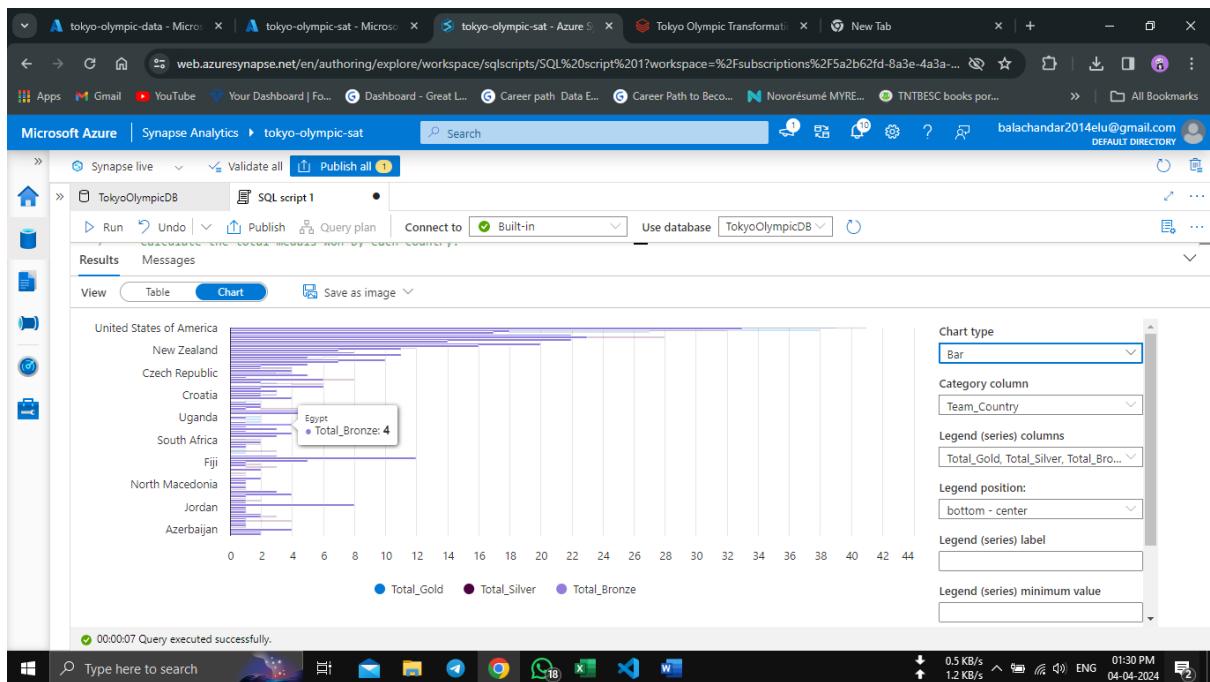


Performing Data Analysis on the Data

Create SQL script to Perform Exploratory data analysis using SQL. You can also use PowerBI to generate your analysis reports.



Visualize the data according to the SQL query in Azure Synapse Analytics.



Microsoft Azure | Synapse Analytics > tokyo-olympic-sat

Synapse live | Validate all | Publish all | SQL script 1

TokyoOlympicDB | SQL script 1

Run | Undo | Publish | Query plan | Connect to: Built-in | Use database: TokyoOlympicDB

```

15 ORDER BY Total_Gold DESC;
16
17 --Calculate the average number of entries by gender for each discipline:
18 SELECT Discipline,
19 Avg(Female) Avg_Female,
20 Avg(Male) Avg_Male
21 FROM entriesgender
22 GROUP BY Discipline;

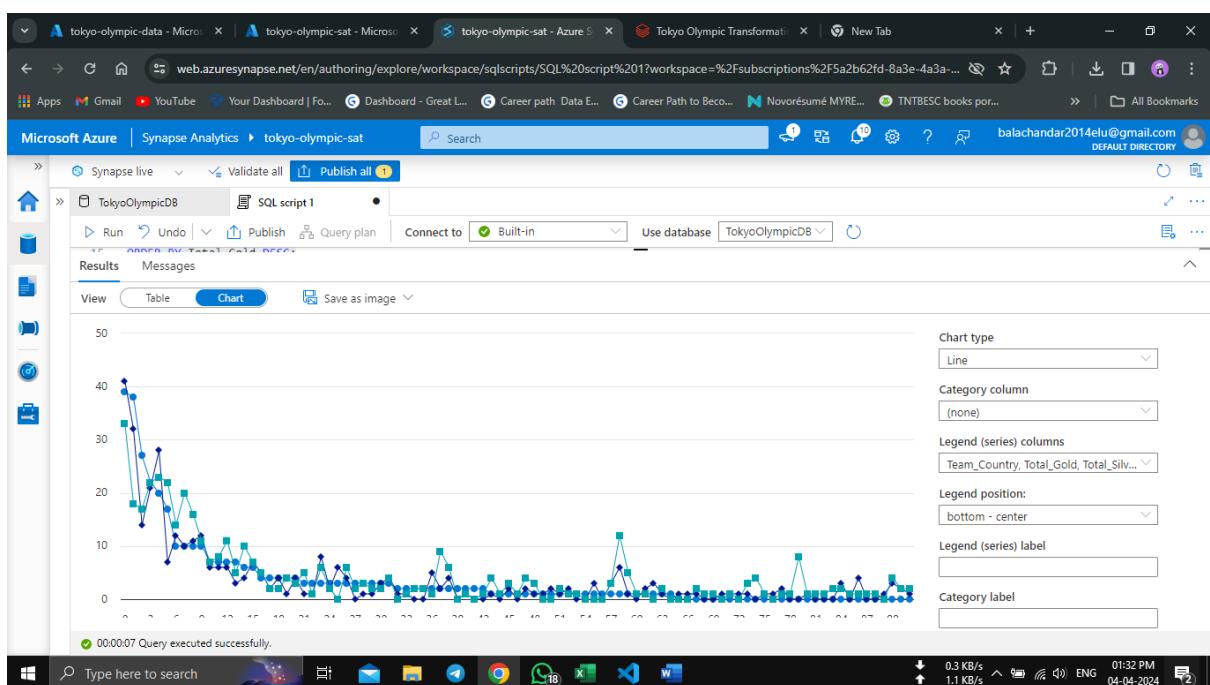
```

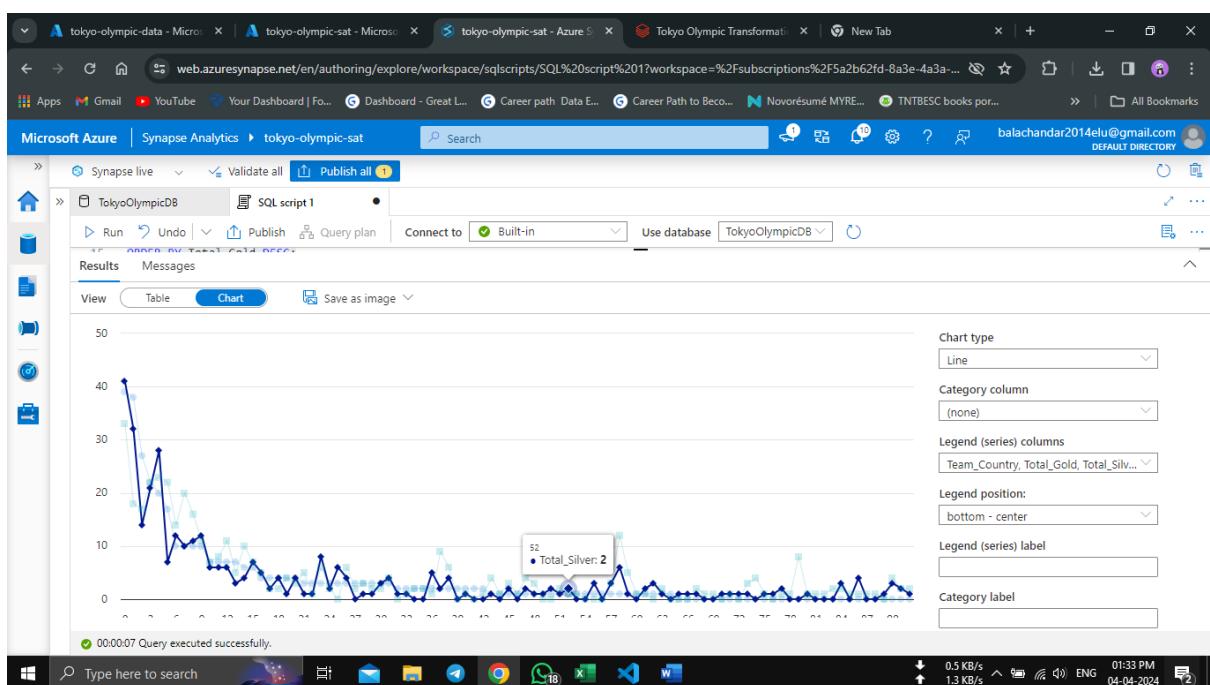
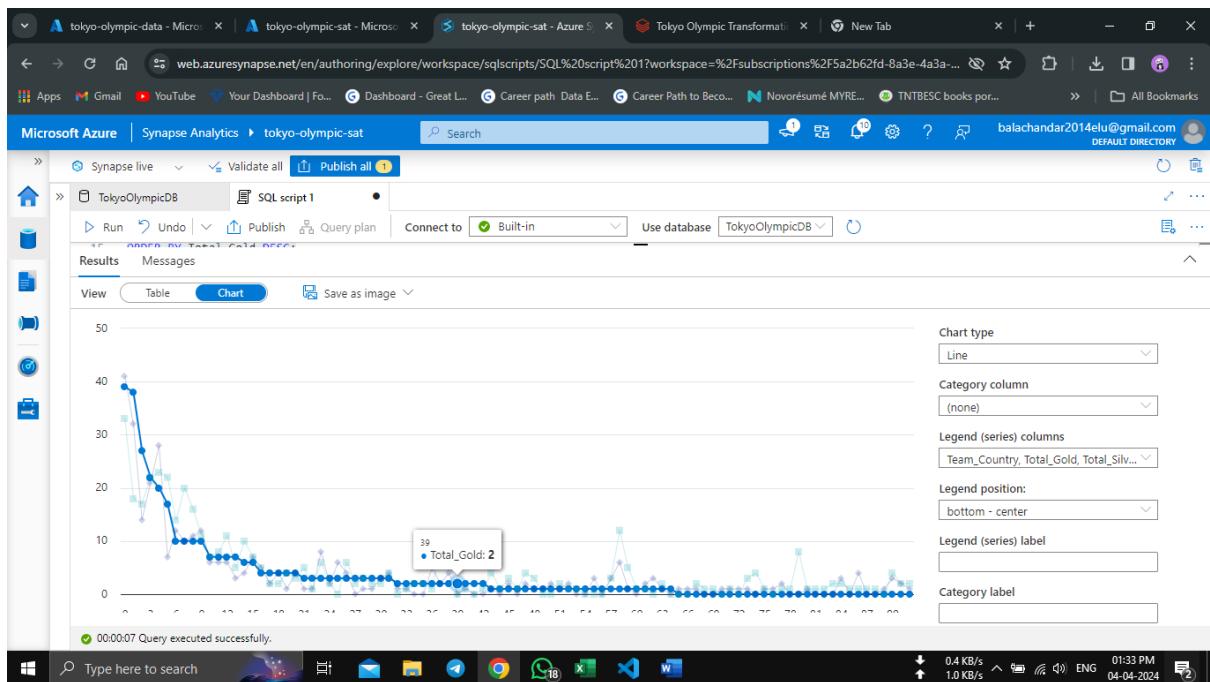
Results | Messages

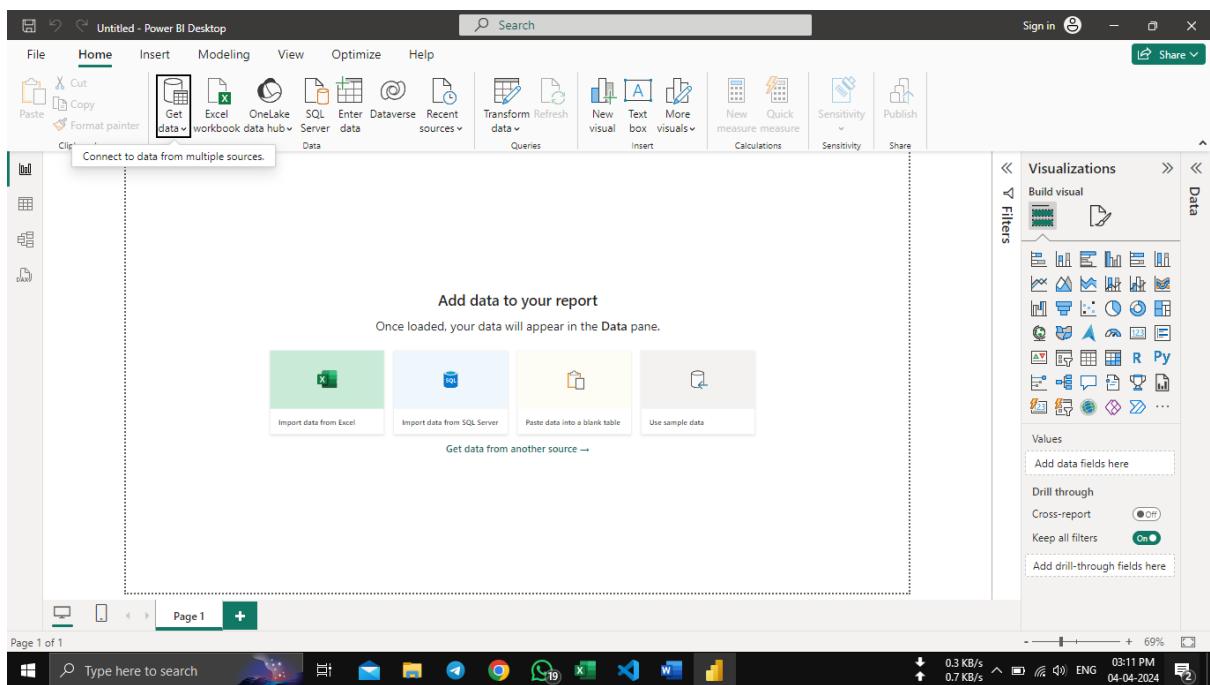
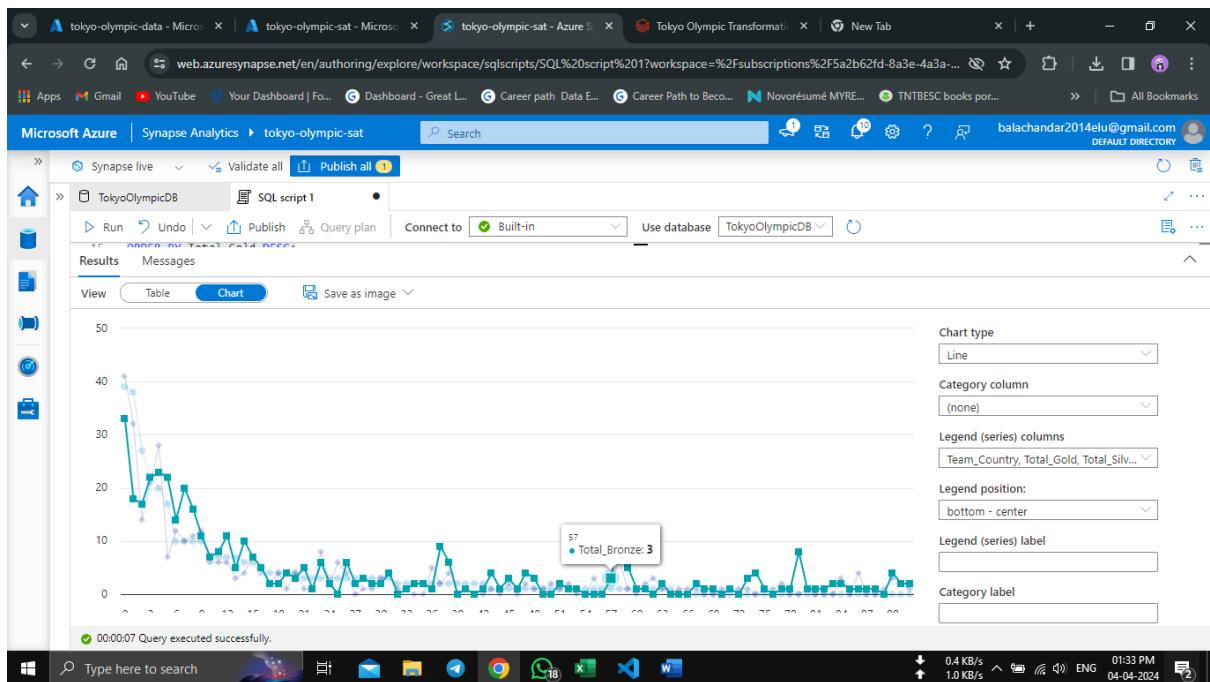
View: Table | Chart | Export results

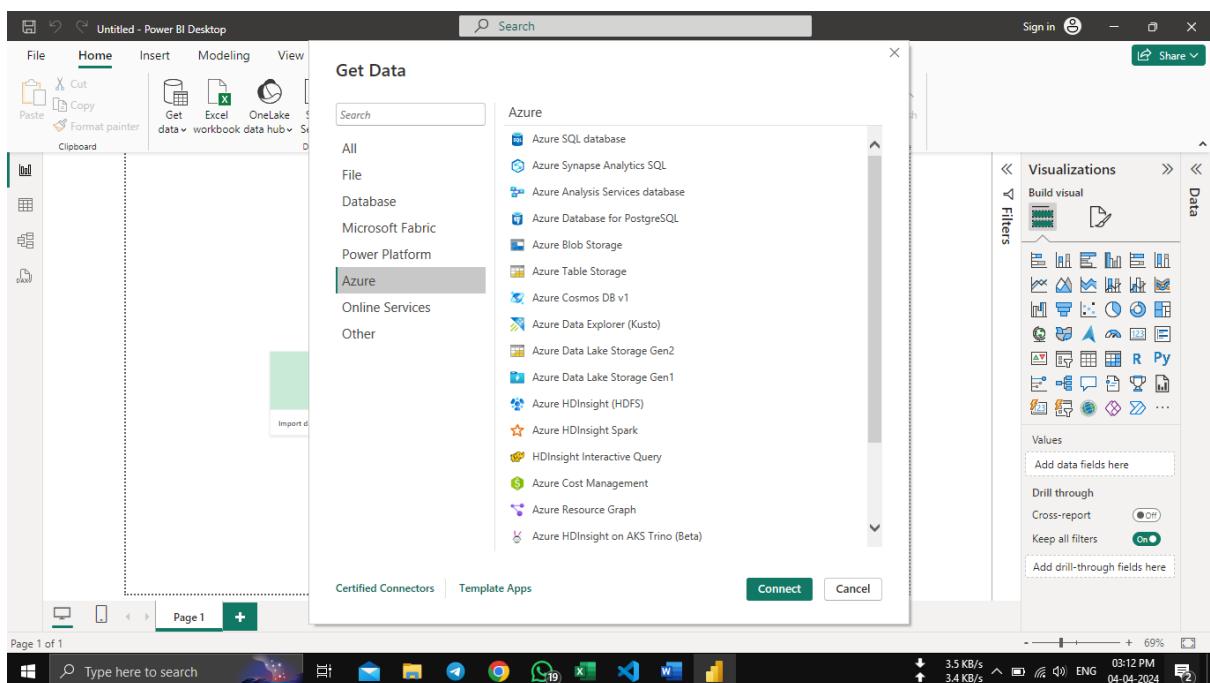
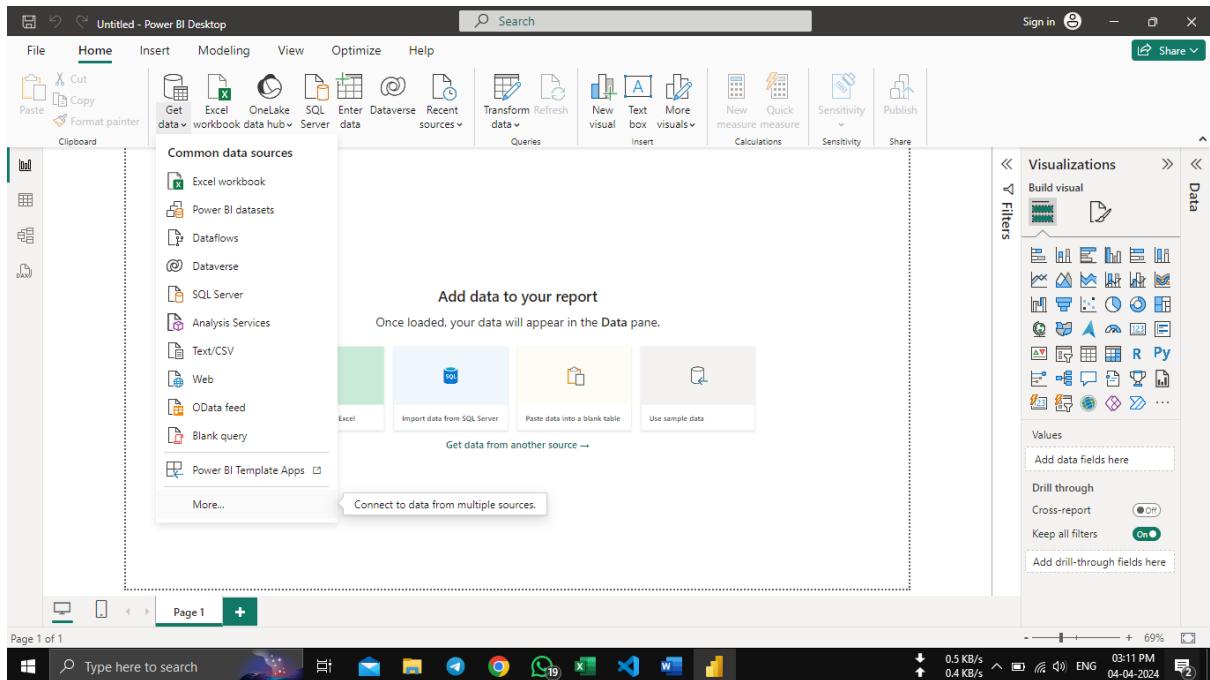
Team_Country	Total_Gold	Total_Silver	Total_Bronze
United States of America	39	41	33
People's Republic of China	38	32	18
Japan	27	14	17

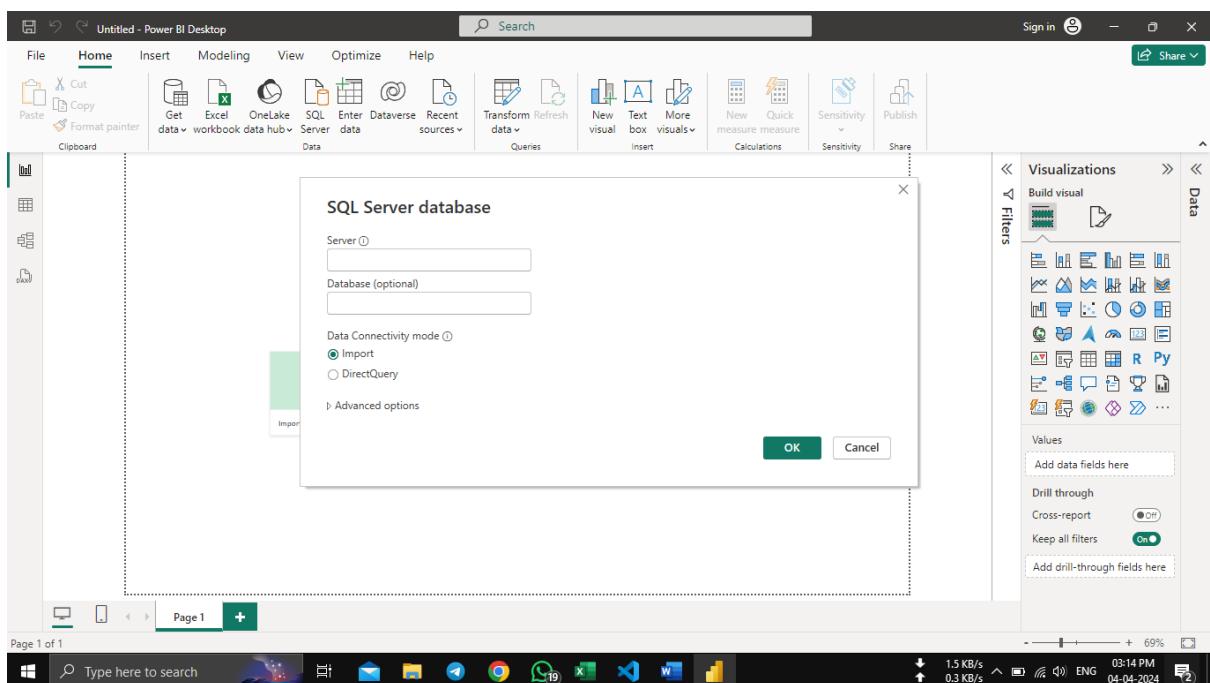
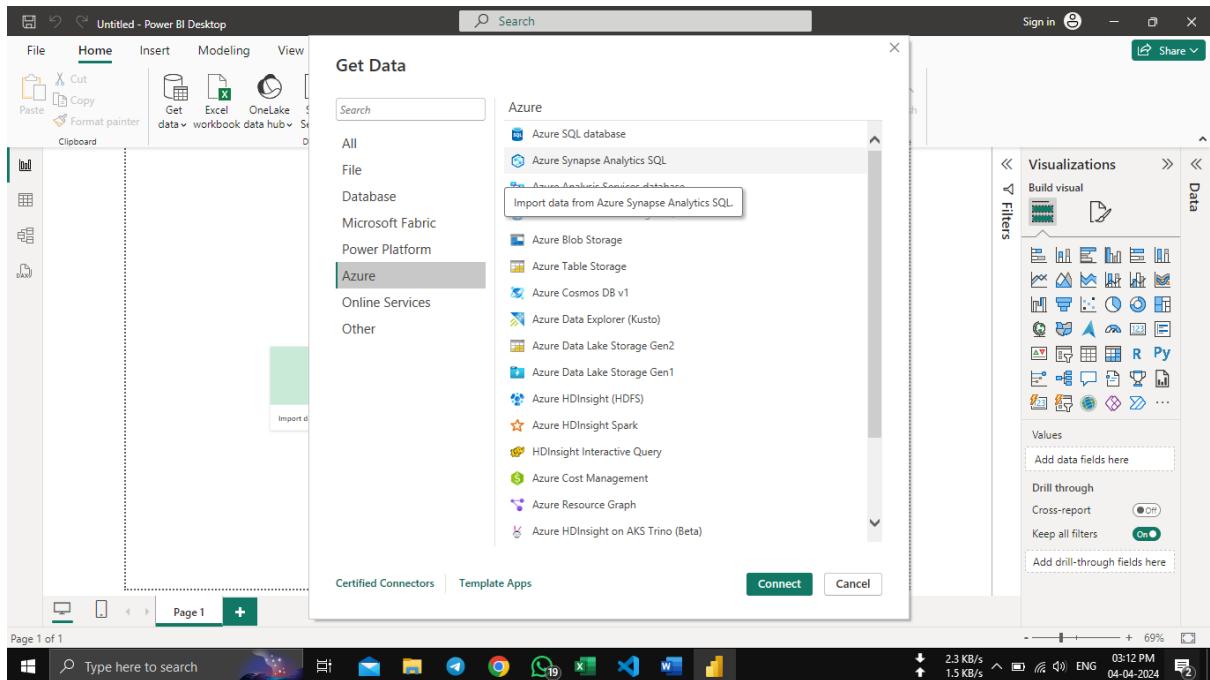
00:00:07 Query executed successfully.











Recent

Name ↑

Name	Type	Location	Resource Group	Subscription	Last accessed
tokyo-olympic-db	Azure Databricks Service	Southeast Asia	tokyo-olympic	Free Trial	1 hour ago
tokyo-olympic-dfac24	Data factory (V2)	Southeast Asia	tokyo-olympic	Free Trial	1 hour ago
tokyoolympicdatachandru0	Storage account	Southeast Asia	tokyo-olympic	Free Trial	1 hour ago
tokyo-olympic-sat	Synapse workspace	Southeast Asia	tokyo-olympic	Free Trial	5 hours ago
tokyo-olympic	Resource group	Southeast Asia	tokyo-olympic	Free Trial	5 hours ago
dbstoragegev7clui3qj2	Storage account	Southeast Asia	databricks-rg-tokyo-olympic	Free Trial	1 day ago

Showing 1 - 6 of 6 results.

https://portal.azure.com/#resource/subscriptions/5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6/resourceGroups/tokyo-olympic/providers/Microsoft.Synapse/workspaces/tokyo-olympic-sat

tokyo-olympic-sat | Properties

Primary ADLS Gen2 account URL: https://tokyoolympicdatachandru0.dfs.core.windows.net

Primary ADLS Gen2 file system: tokyo-olympic-data

Dedicated SQL endpoint: tokyo-olympic-sat.sql.azuresynapse.net

Serverless SQL endpoint: tokyo-olympic-sat-on-demand.sql.azuresynapse.net

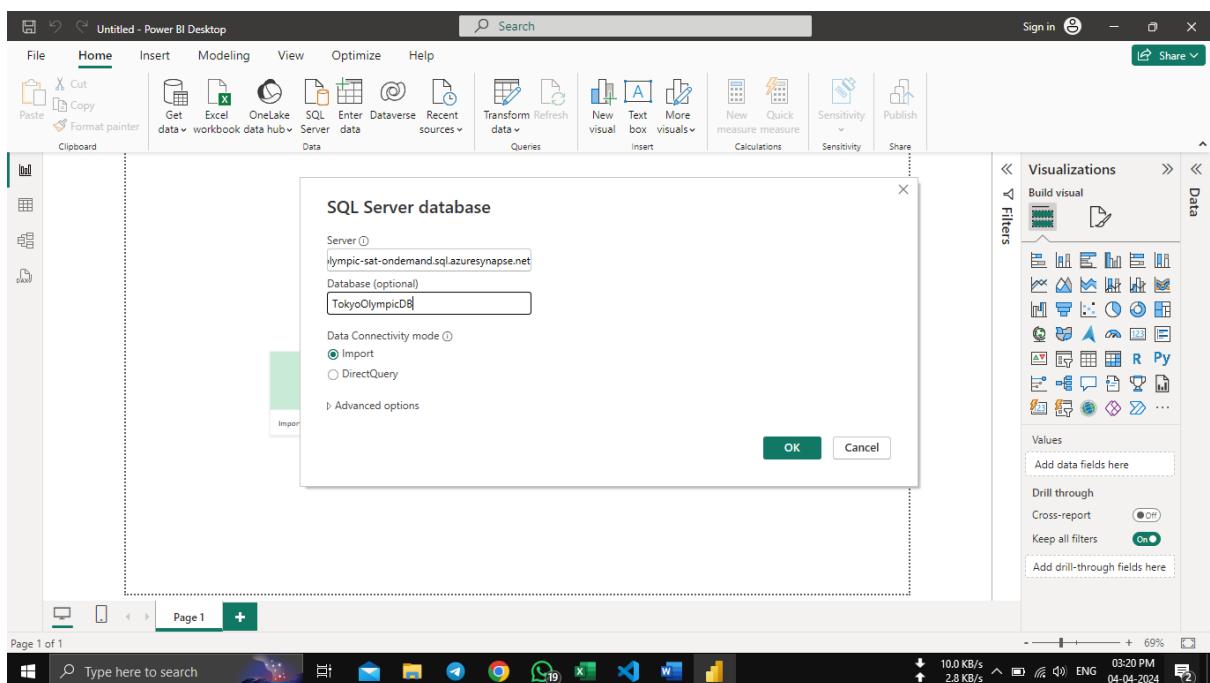
Development endpoint: tokyo-olympic-sat-on-demand.dev.azuresynapse.net

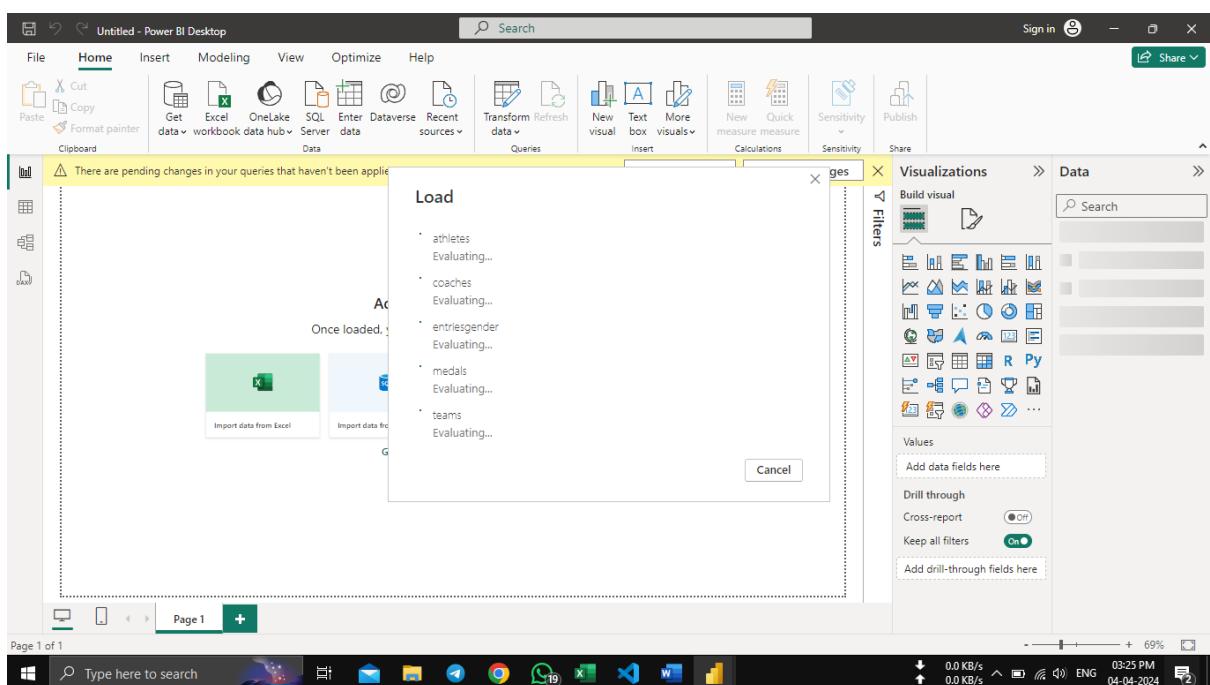
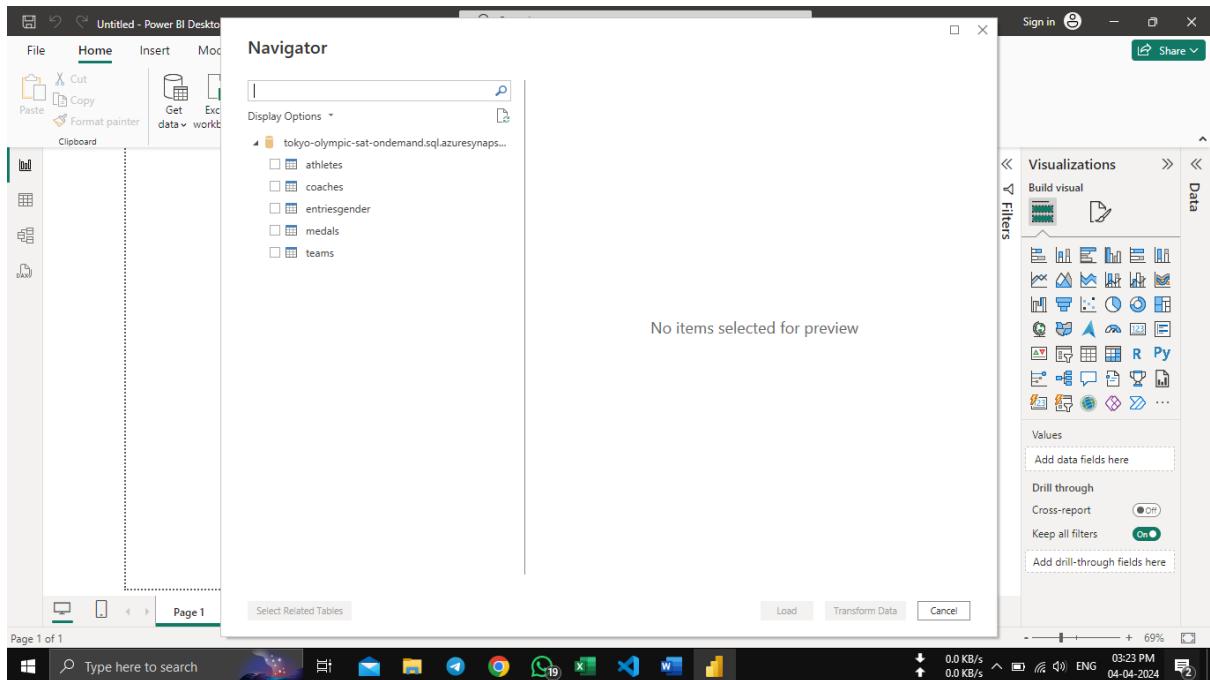
Managed identity object ID: b167b0c7-15d2-4321-9eb8-3b5da271f540

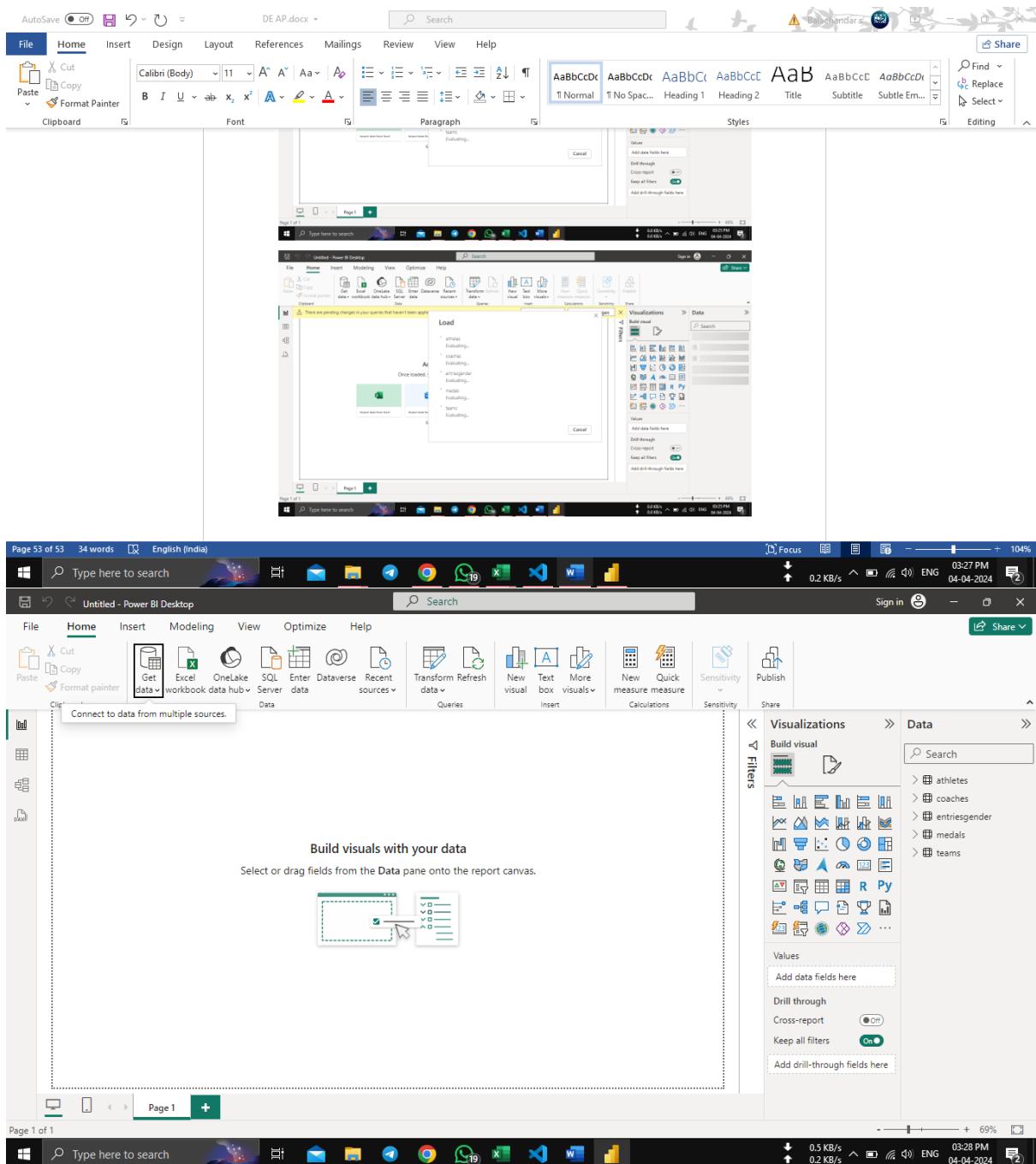
The screenshot shows the Microsoft Azure Synapse Analytics workspace. On the left, the Data blade is open, displaying the 'Lake database' section with 'TokyoOlympicDB' selected. Under 'Tables', there are two tables: 'athletes' and 'coaches'. The 'athletes' table has columns 'PersonName' (string), 'Country' (string), and 'Discipline' (string). The 'coaches' table has columns 'Name' (string), 'Country' (string), and 'Discipline' (string). In the center, a 'SQL script 1' tab is active, containing the following SQL code:

```
1 --Count the number of Athletes from each Country:
2 SELECT Country, COUNT(*) AS totalAthletes
3 FROM athletes
4 GROUP BY Country
5 ORDER BY TotalAthletes DESC;
6
7 --Calculate the total medals won by each country:
8 SELECT
9 Team_Country,
10 SUM(GOLD) Total_Gold,
11 SUM(Silver)Total_Silver,
12 SUM(Bronze)Total_Bronze
13 FROM medals
14 GROUP BY Team_Country
15 ORDER BY Total_Gold DESC;
16
17 --Calculate the average number of entries by gender for each discipline:
18 SELECT Discipline,
19 Avg(Female) Avg_Female,
20 Avg(Male) Avg_Male
21 FROM entriesgender
```

The 'Results' tab shows a message: "00:00:07 Query executed successfully." At the bottom right of the browser window, the status bar indicates "71.9 KB/s 1.4 KB/s" and the date "04-04-2024".







Untitled - Power BI Desktop

File Home Help Table tools

Name athletes

Mark as date table Calendars Manage relationships New Quick measure New measure column New table Calculations

Structure

PersonName Country Discipline

PersonName	Country	Discipline
ABDIKHAMAN Abdi	United States of America	Athletics
ALLEN Devon	United States of America	Athletics
ALLMAN Valerie	United States of America	Athletics
ANDERSEN Brooke	United States of America	Athletics
ANDERSON Shae	United States of America	Athletics
AQUILLA Adelaide	United States of America	Athletics
ASHLEY Whitney	United States of America	Athletics
BAILEY JR Aldrich	United States of America	Athletics
BAKER Ronnie	United States of America	Athletics
BASTIEN Steven	United States of America	Athletics
BATTLE Anavia	United States of America	Athletics
BEDNAREK Kenneth	United States of America	Athletics
BENARD Chris	United States of America	Athletics
BENJAMIN Rai	United States of America	Athletics
BERRY Gwen	United States of America	Athletics
BOUGARD Erica	United States of America	Athletics
BROMELL Trayvon	United States of America	Athletics
BURKS Quanesha	United States of America	Athletics
BUTTS-TOWNSEND Tynita	United States of America	Athletics
CARD Kelsey	United States of America	Athletics
CARTER Chris	United States of America	Athletics
CENTROWITZ Matthew	United States of America	Athletics
CHERRY Michael	United States of America	Athletics

Table: athletes (11,085 rows)

Search

Sign in Share

Data

athletes coaches entriesgender medals teams

Untitled - Power BI Desktop

File Home Help Table tools

Name: medals

Structure

Mark as date table Calendars Manage relationships Relationships New measure Quick New measure column New table Calculations

Rank Team_Country Gold Silver Bronze Total Rank_by_Total

Rank	Team_Country	Gold	Silver	Bronze	Total	Rank_by_Total
1	United States of America	39	41	33	113	1
2	People's Republic of China	38	32	18	88	2
3	Japan	27	14	17	58	5
4	Great Britain	22	21	22	65	4
5	ROC	20	28	23	71	3
6	Australia	17	7	22	46	6
7	Netherlands	10	12	14	36	9
8	France	10	12	11	33	10
9	Germany	10	11	16	37	8
10	Italy	10	10	20	40	7
11	Canada	7	6	11	24	11
12	Brazil	7	6	8	21	12
13	New Zealand	7	6	7	20	13
14	Cuba	7	3	5	15	18
15	Hungary	6	7	7	20	13
16	Republic of Korea	6	4	10	20	13
17	Poland	4	5	5	14	19
18	Czech Republic	4	4	3	11	23
19	Kenya	4	4	2	10	25
20	Norway	4	2	2	8	29
21	Jamaica	4	1	4	9	26
22	Spain	3	8	6	17	17
23	Sweden	3	6	0	9	26

Data

Name: medals
Storage mode: Import
Data refreshed: 4/4/2024, 3:26:23 PM

Untitled - Power BI Desktop

File Home Help Table tools

Name: teams

Structure

Mark as date table Calendars Manage relationships Relationships New measure Quick New measure column New table Calculations

TeamName Discipline Country Event

TeamName	Discipline	Country	Event
Belgium	3x3 Basketball	Belgium	Men
China	3x3 Basketball	People's Republic of China	Men
Japan	3x3 Basketball	Japan	Men
Latvia	3x3 Basketball	Latvia	Men
Netherlands	3x3 Basketball	Netherlands	Men
Poland	3x3 Basketball	Poland	Men
ROC	3x3 Basketball	ROC	Men
Serbia	3x3 Basketball	Serbia	Men
Argentina	Basketball	Argentina	Men
Australia	Basketball	Australia	Men
Czech Republic	Basketball	Czech Republic	Men
France	Basketball	France	Men
Germany	Basketball	Germany	Men
Islamic Rep. of Iran	Basketball	Islamic Republic of Iran	Men
Italy	Basketball	Italy	Men
Japan	Basketball	Japan	Men
Nigeria	Basketball	Nigeria	Men
Slovenia	Basketball	Slovenia	Men
Spain	Basketball	Spain	Men
United States	Basketball	United States of America	Men
Azaad/Capogrosso	Beach Volleyball	Argentina	Men
McHugh/Schumann	Beach Volleyball	Australia	Men
Alison/Alvaro Filho	Beach Volleyball	Brazil	Men

Data

Name: teams
Storage mode: Import
Data refreshed: 4/4/2024, 3:26:27 PM



Untitled - Power BI Desktop

File Home Help Table tools

Name teams

Structure

Mark as date table Relationships Calculations

Calendars

New measure New measure column New table

TeamName Discipline Country Event

TeamName	Discipline	Country	Event
Belgium	3x3 Basketball	Belgium	Men
C	3x3 Basketball	People's Republic of China	Men
Japan	3x3 Basketball	Japan	Men
Latvia	3x3 Basketball	Latvia	Men
Netherlands	3x3 Basketball	Netherlands	Men
Poland	3x3 Basketball	Poland	Men
ROC	3x3 Basketball	ROC	Men
Serbia	3x3 Basketball	Serbia	Men
Argentina	Basketball	Argentina	Men
Australia	Basketball	Australia	Men
Czech Republic	Basketball	Czech Republic	Men
France	Basketball	France	Men
Germany	Basketball	Germany	Men
Islamic Rep. of Iran	Basketball	Islamic Republic of Iran	Men
Italy	Basketball	Italy	Men
Japan	Basketball	Japan	Men
Nigeria	Basketball	Nigeria	Men
Slovenia	Basketball	Slovenia	Men
Spain	Basketball	Spain	Men
United States	Basketball	United States of America	Men
Azaad/Capogrosso	Beach Volleyball	Argentina	Men
McHugh/Schumann	Beach Volleyball	Australia	Men
Alison/Alvaro Filho	Beach Volleyball	Brazil	Men

Table: teams (743 rows)

Clipboard Data Queries Relationships Calculations Security Q&A Sensitivity Share

0.3 KB/s 1.0 KB/s ENG 03:31 PM 04-04-2024

Untitled - Power BI Desktop

File Home Help

Paste Cut Copy Get data Excel OneLake Data Server Enter Dataverse Recent sources Transform Refresh data Manage relationships New table New measure column View roles as Security Q&A Language schema Sensitivity Share

Clipboard Data Properties

teams

- Country
- Discipline
- Event
- TeamName

athletes

- Country
- Discipline
- PersonName

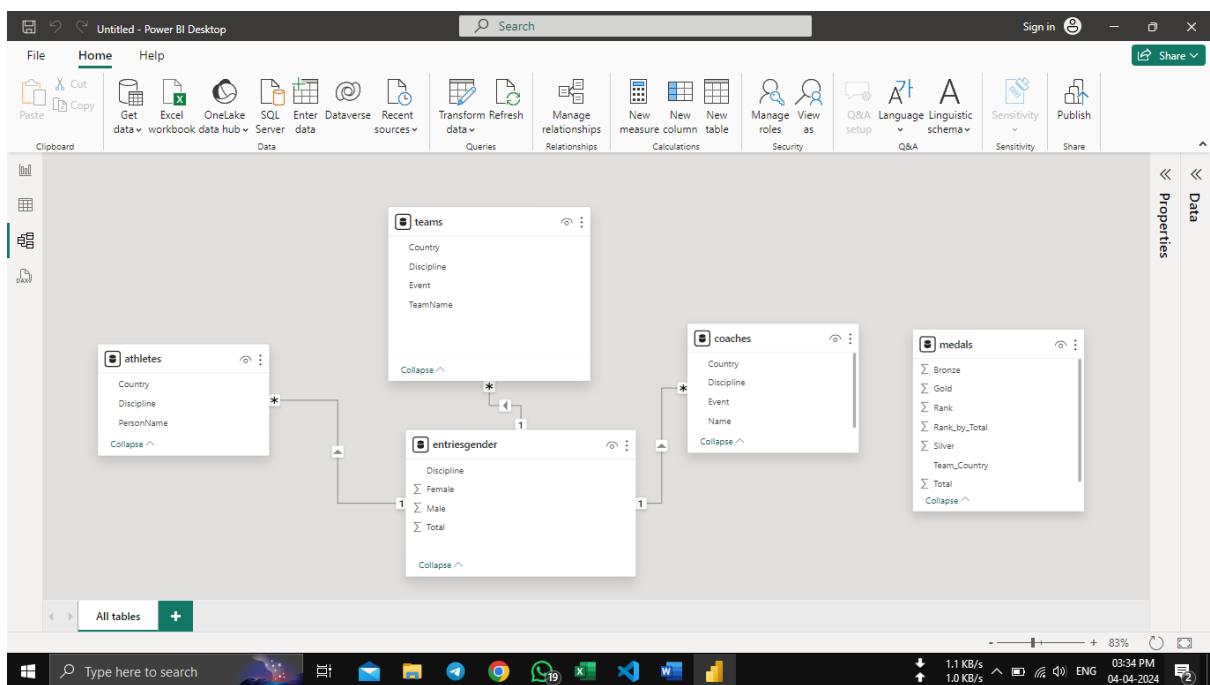
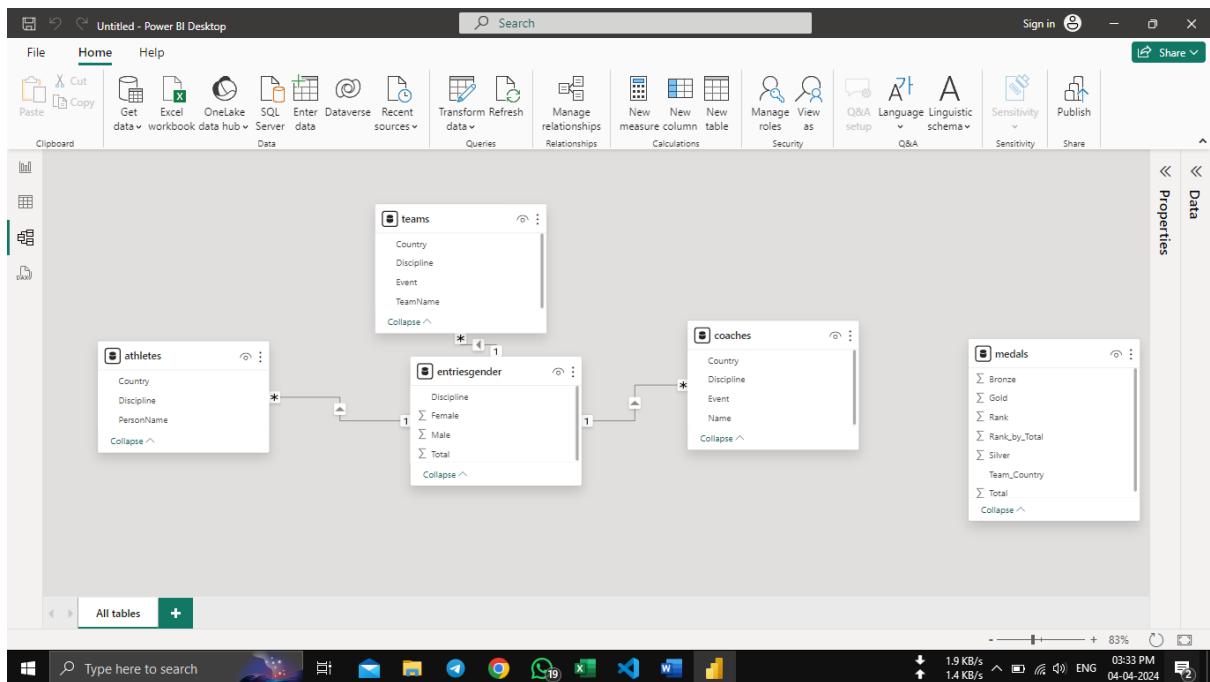
entriesgender

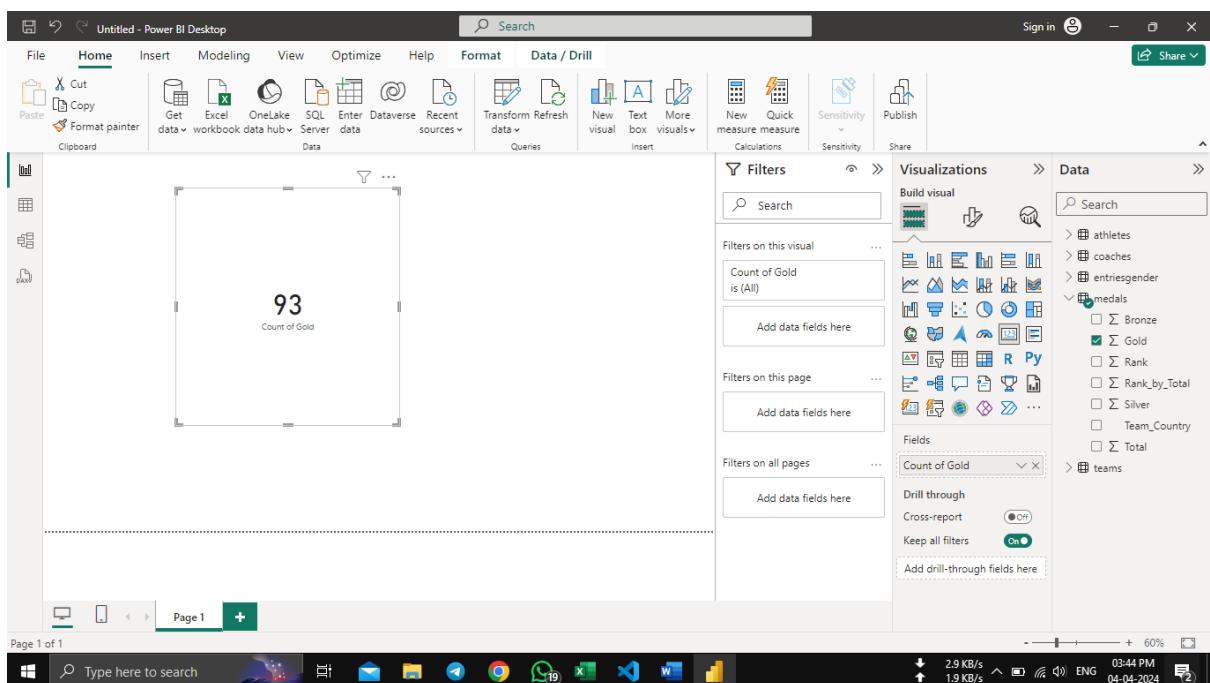
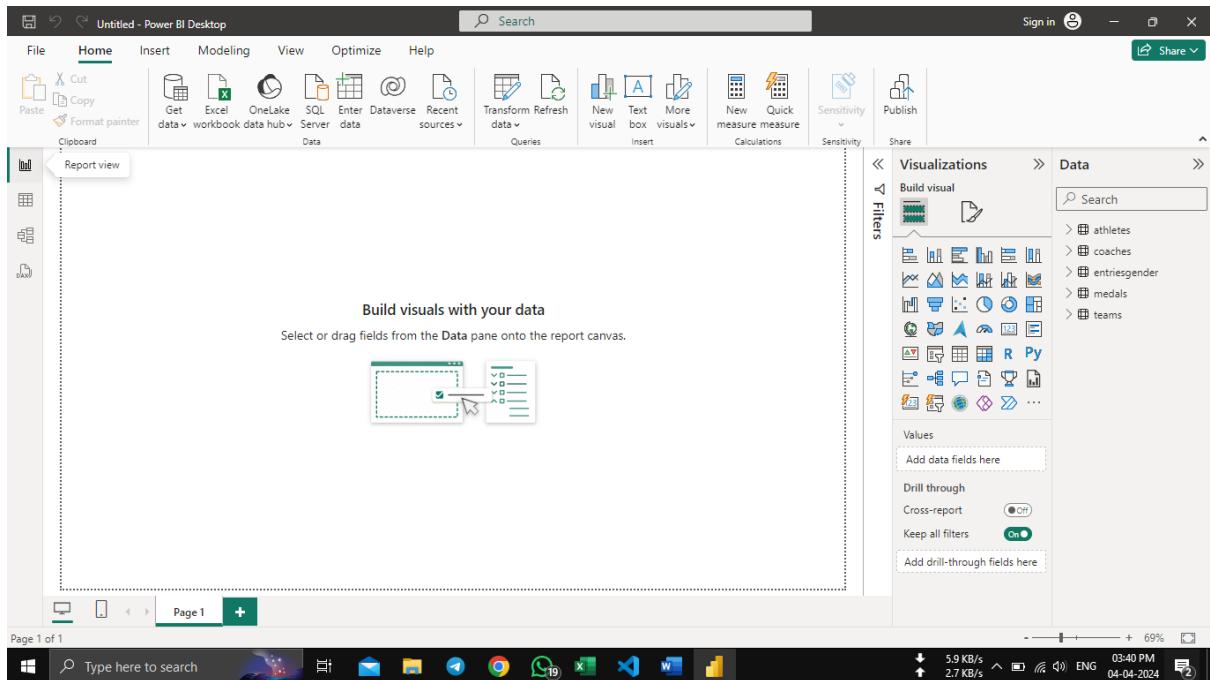
- Discipline
- Female
- Male
- Total

All tables +

100% 1.5 KB/s 1.2 KB/s ENG 03:31 PM 04-04-2024

Data Modelling...





Untitled - Power BI Desktop

Home

Clipboard

Total No. Gold Medals
93

Country

Filters

Visualizations

Data

Page 1

Page 1 of 1

Type here to search

0.4 KB/s 0.1 KB/s ENG 04:08 PM 04-04-2024

Untitled - Power BI Desktop

Home

Clipboard

Total No. Gold Medals
93

Country and Discipline

Team_Country

Country	Sum of Gold	Sum of Rank	Sum of Rank by Event
United States of America	38	1	
People's Republic of China	38	2	
Japan	27	3	
United Kingdom	23	4	
ROC	20	5	
Australia	17	6	
France	10	8	
Germany	10	9	
Italy	10	10	
Netherlands	10	12	
Great Britain	7	13	
Canada	7	11	
Cuba	7	14	
New Zealand	7	15	
Hungary	6	15	
Republic of Korea	6	16	
Czech Republic	4	18	
Ukraine	4	21	
Kenya	4	19	
Norway	4	20	
Peru	4	17	
Belgium	3	28	
Bulgaria	3	30	
Total	340	4309	

Filters

Visualizations

Data

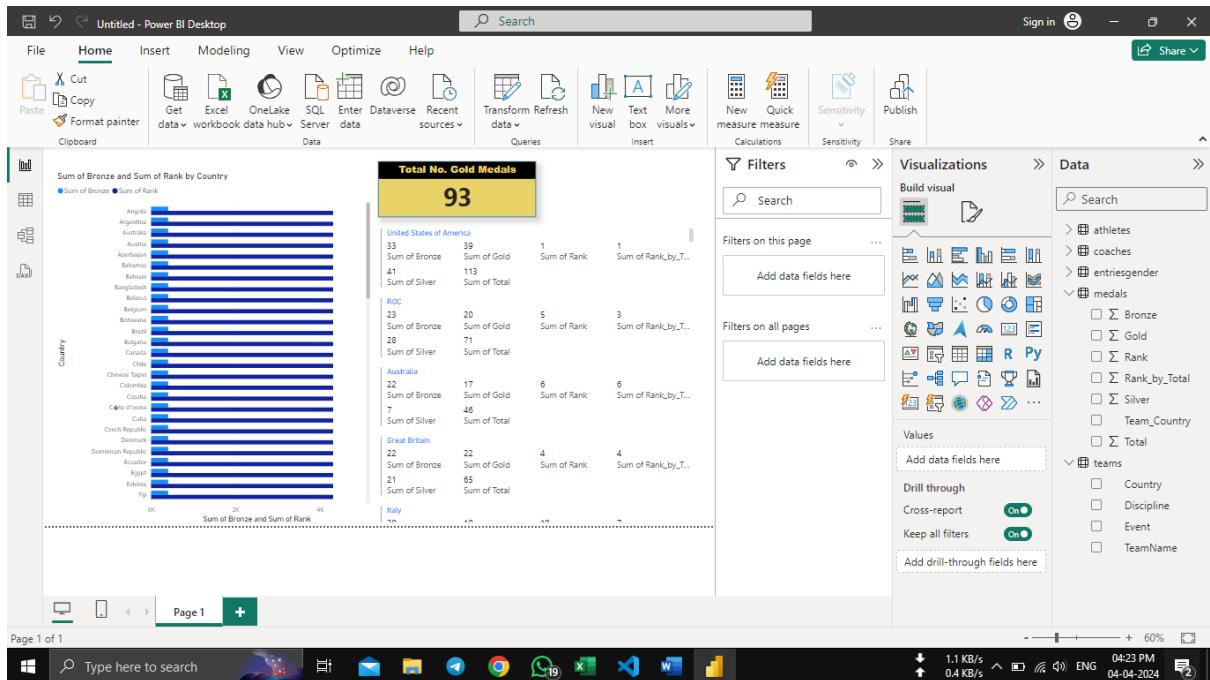
Page 1

Page 1 of 1

Type here to search

1.3 KB/s 0.9 KB/s ENG 04:21 PM 04-04-2024

Dashboard Creation by using Microsoft Power BI.



Project Over....

Delete the Project resource and avoid the Over Billing..

Delete storage account ... Deleting storage account...
Deleting storage account 'tokyoolympicdatachandru0'...

The following storage account and its contents will be deleted.

Resource to be deleted

- tokyoolympicdatachandru0

Dependent resources to be deleted

The data provided is regularly updated about 2-4 times a day and published hourly. If your account has extremely large objects, it may be over a day between updates.

Resource	Number of instances	Total data stored
Containers	2	2.86 MiB
File shares	0	0
Tables	14	2.26 MiB
Queues	0	0

Enter storage account name to confirm deletion *
tokyoolympicdatachandru0

Delete **Cancel** **Give feedback**

Delete Resources

The selected resources along with their related resources and contents will be permanently deleted. If you are unsure of the selected resource dependencies, navigate to the individual resource page to perform the delete operation. More details of the resource dependencies are available in the manage experience.

Resources to be deleted (1)

Name	Resource type
tokyo-olympic-db	Azure Databricks Service

Enter "delete" to confirm deletion *

Delete **Cancel**

The screenshot shows the Microsoft Azure portal interface. On the left, the 'Resource groups' blade is open, displaying a list of resource groups, including 'NetworkWatcherRG' and 'tokyo-olympic'. On the right, a modal window titled 'Delete a resource group' is displayed, showing the 'tokyo-olympic' resource group selected for deletion. The modal lists one dependent resource, 'tokyo-olympic-db', which is an Azure Databricks Service. A confirmation input field contains the name 'tokyo-olympic'. The bottom of the screen shows the Windows taskbar with various pinned icons.

Resource groups

Default Directory (balachandar2014elugmail.onmicrosoft.com)

Name ↑

NetworkWatcherRG

tokyo-olympic

Page 1 of 1

tokyo-olympic

Delete a resource group

The following resource group and all its dependent resources will be permanently deleted.

Resource group to be deleted

tokyo-olympic

Dependent resources to be deleted (1)

All dependent resources, including hidden types, are shown

Name	Resource type
tokyo-olympic-db	Azure Databricks Service

Enter resource group name to confirm deletion *

tokyo-olympic

Delete Cancel

0.8 KB/s
1.9 KB/s
ENG 04-04-2024

Thank you!