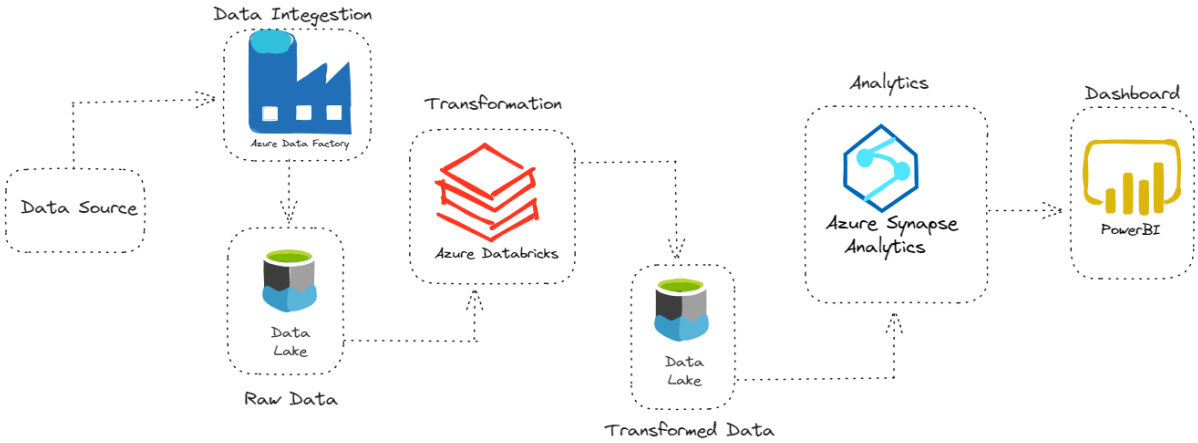
Tokyo 2021 Olympic Data Analytics Project

Pipeline Design Architecture:

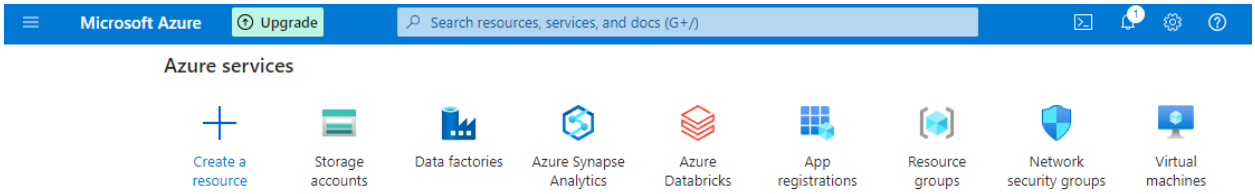


To idea of this project is to analyse Tokyo 2021 Olympics data using data engineering tools and technologies, including Microsoft Azure services such as Azure Data Factory, Data Lake Gen 2, Azure Synapse Analytics and Azure Databricks.

Dataset Used

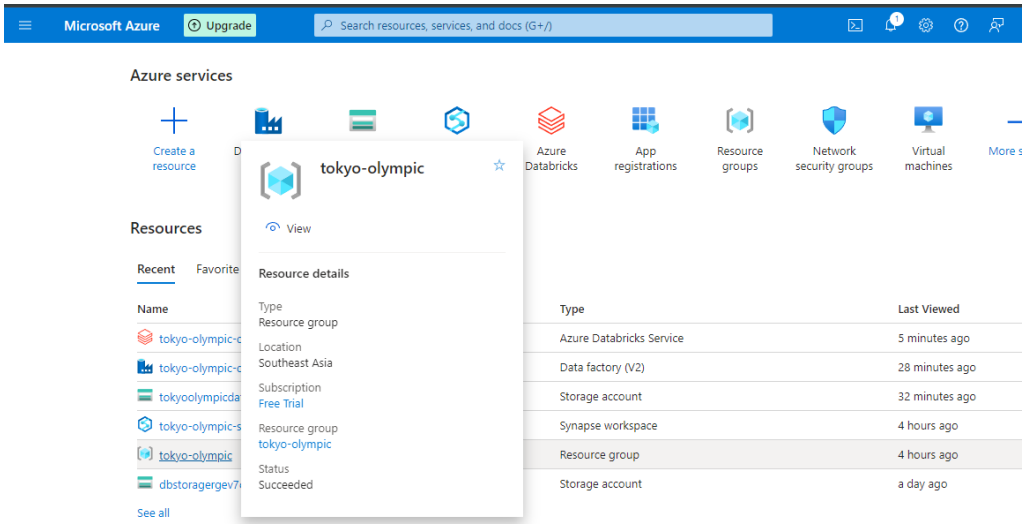
This dataset contains the details of over 11,000 athletes, with 47 disciplines, along with 743 Teams taking part in the 2021(2020) Tokyo Olympics. This dataset contains the details of the Athletes, Coaches, Teams participating as well as the Entries by gender. It contains their names, countries represented,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

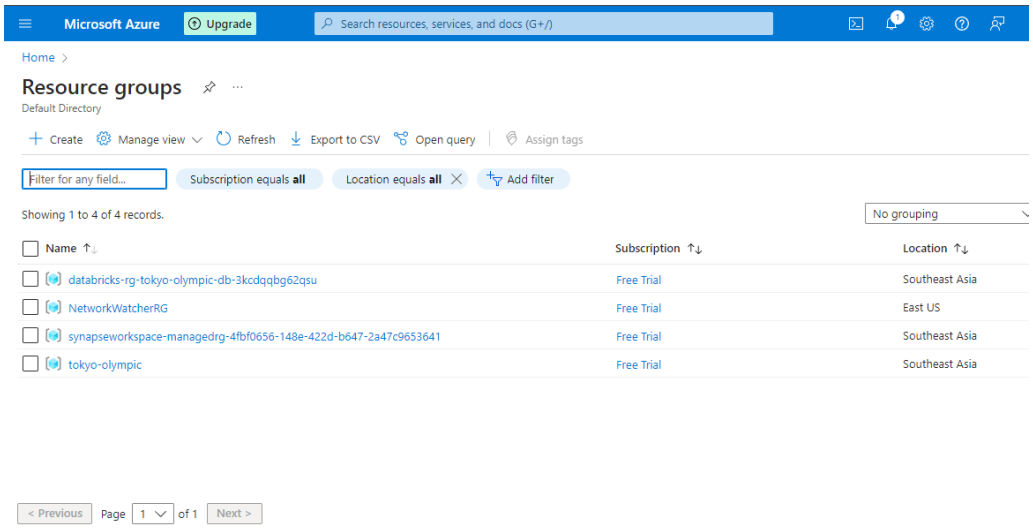
Source(Kaggle): https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo



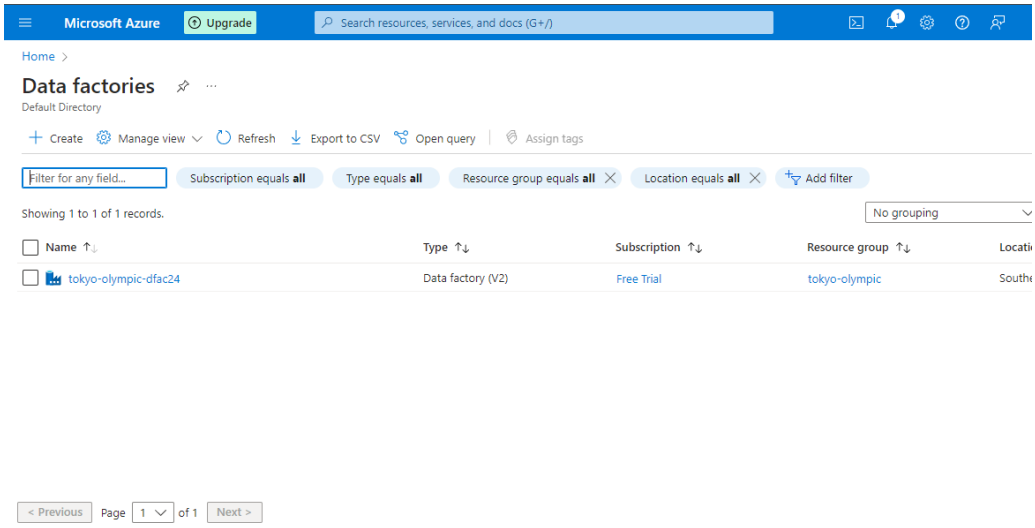
List of Azure Services used for this project

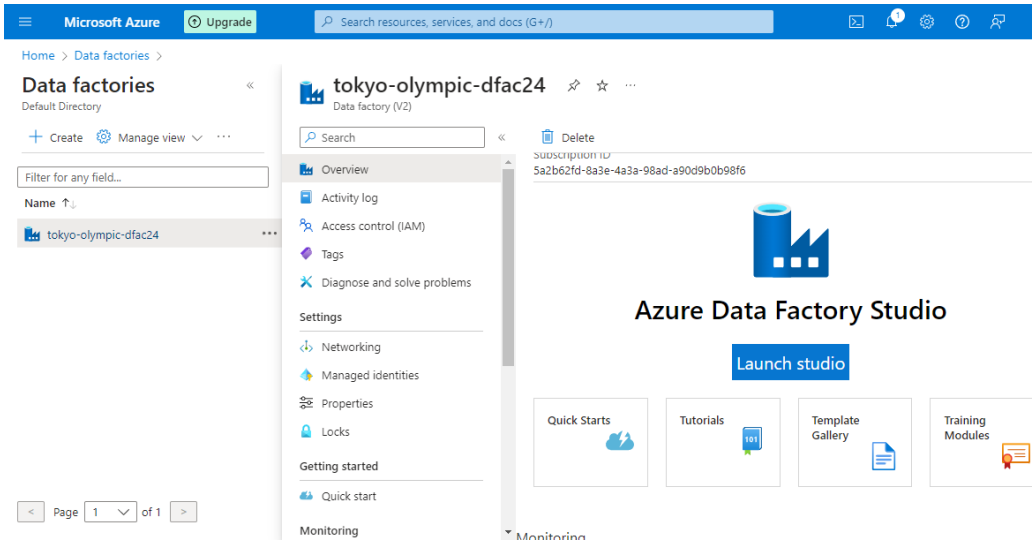
1. **Azure Data Factory:** For data ingestion from GitHub.
2. **Azure Data Lake Storage Gen2**: As the primary data storage solution.
3. **Azure Databricks:** For data transformation tasks.
4. **Azure Synapse Analytics:** To perform in-depth data analytics.



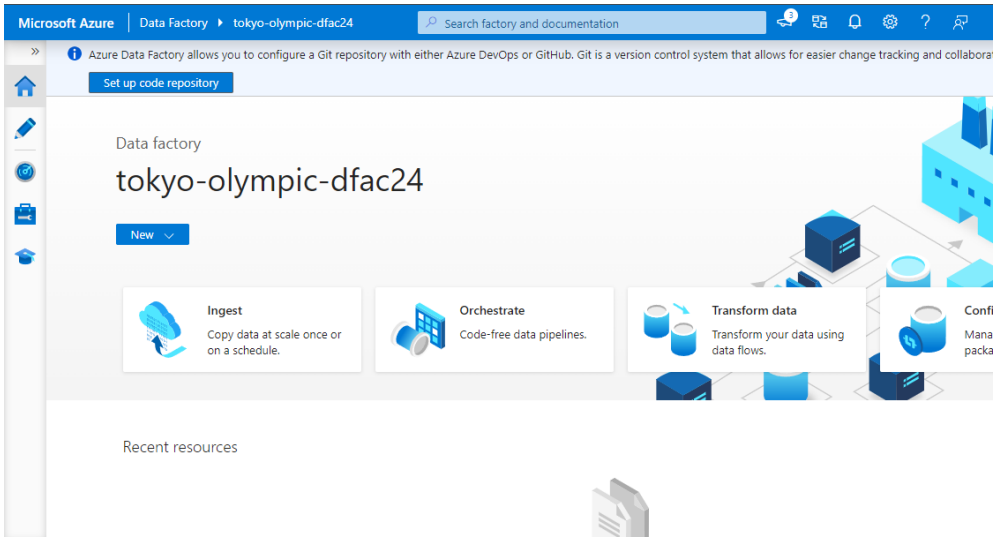


Resource group created.



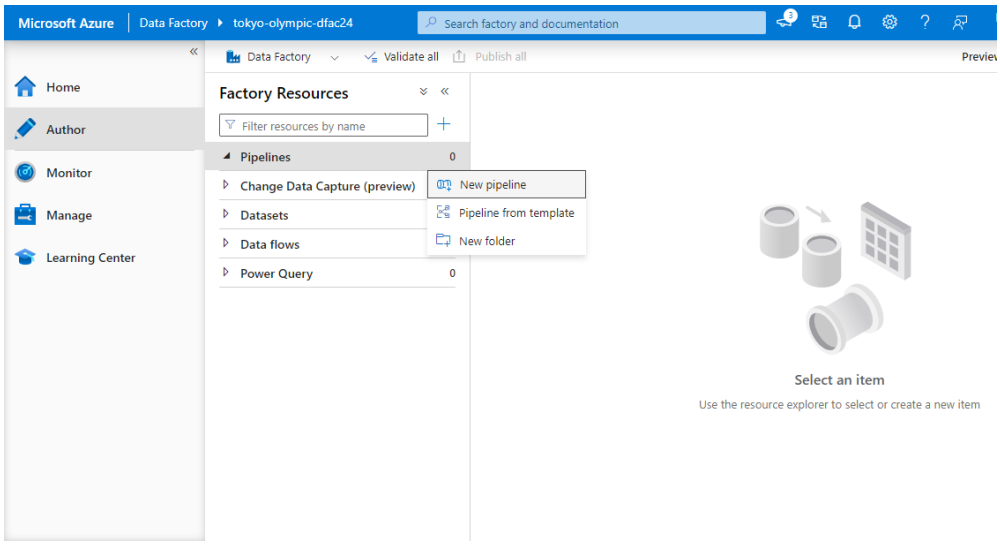


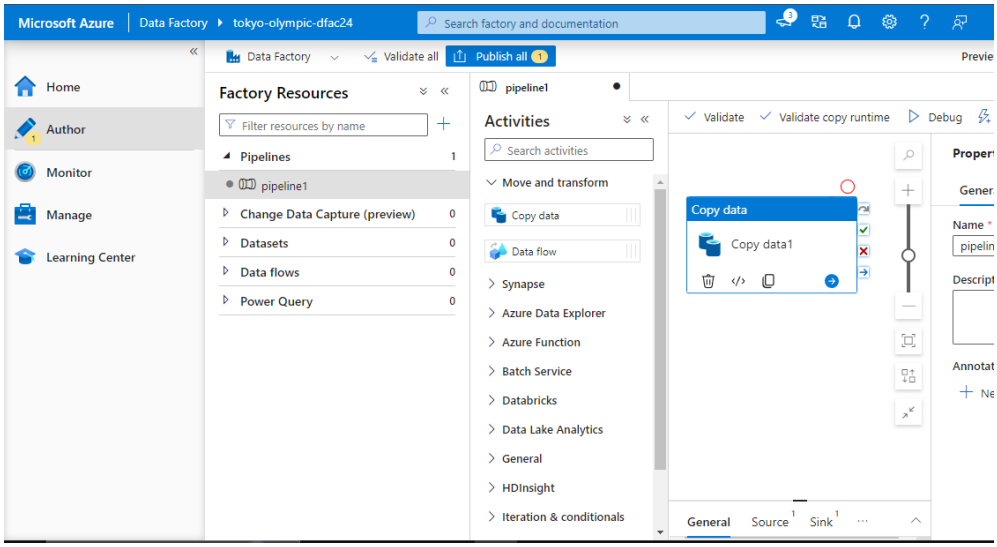
Created the Data Factory Studio for Data injection from GitHub(HTML) Source.



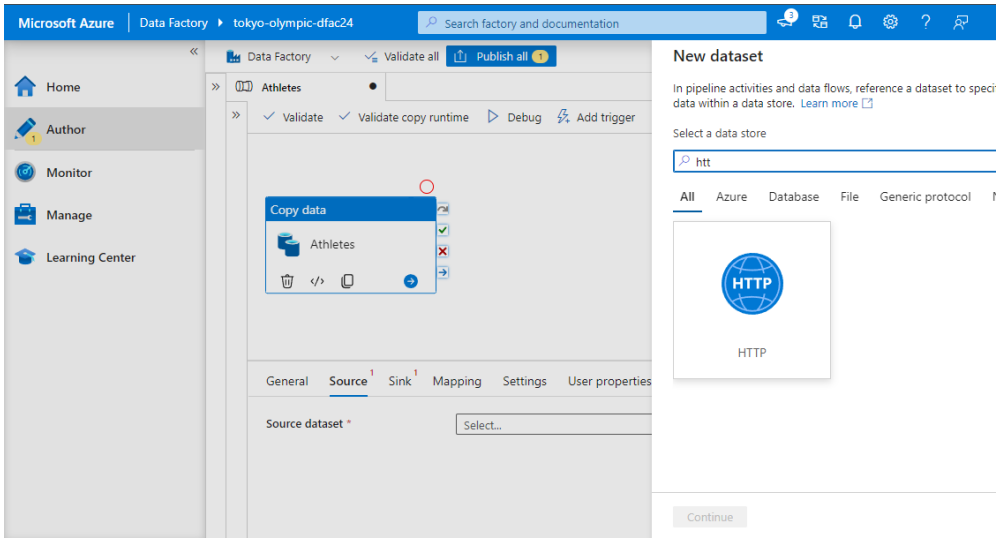
1. Data Ingestion using Azure Data Factory
2. Begin by creating an Azure Data Factory workspace within the previously established resource group.
3. After setting up the workspace, launch the Azure Data Factory Studio.
4. Upload the Tokyo Olympics dataset from kaggle to GitHub.
5. Within the studio, 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.

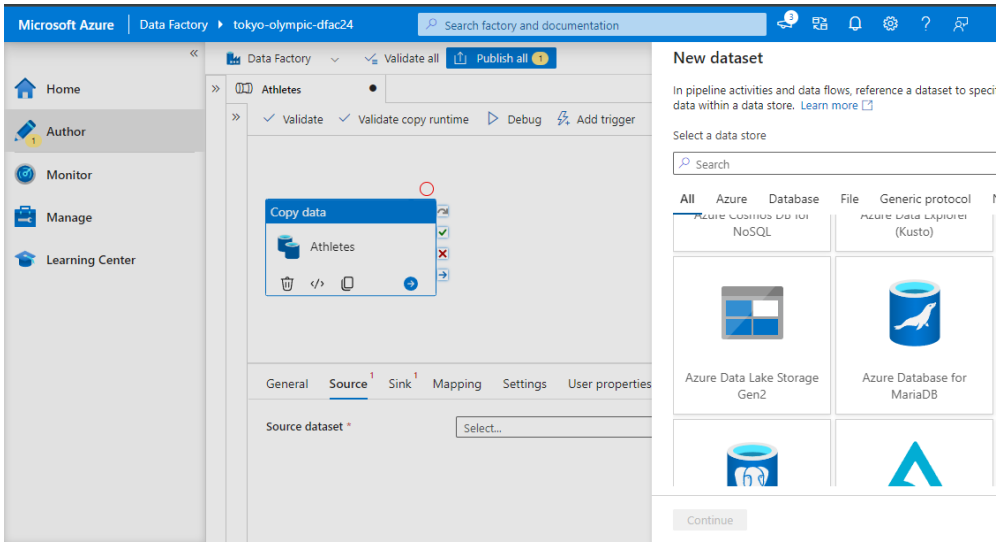
You can connect all the copy data activity together and run them all at once.

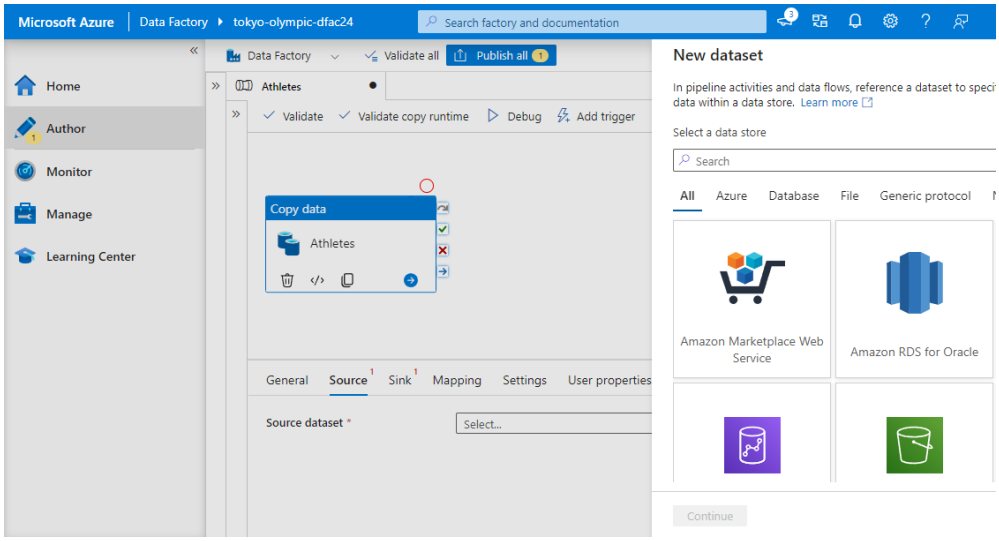


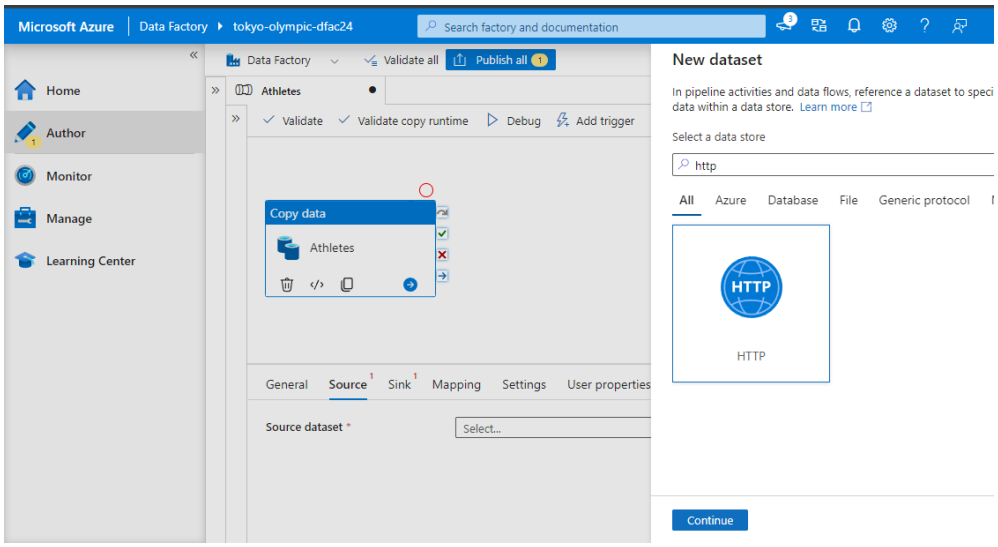


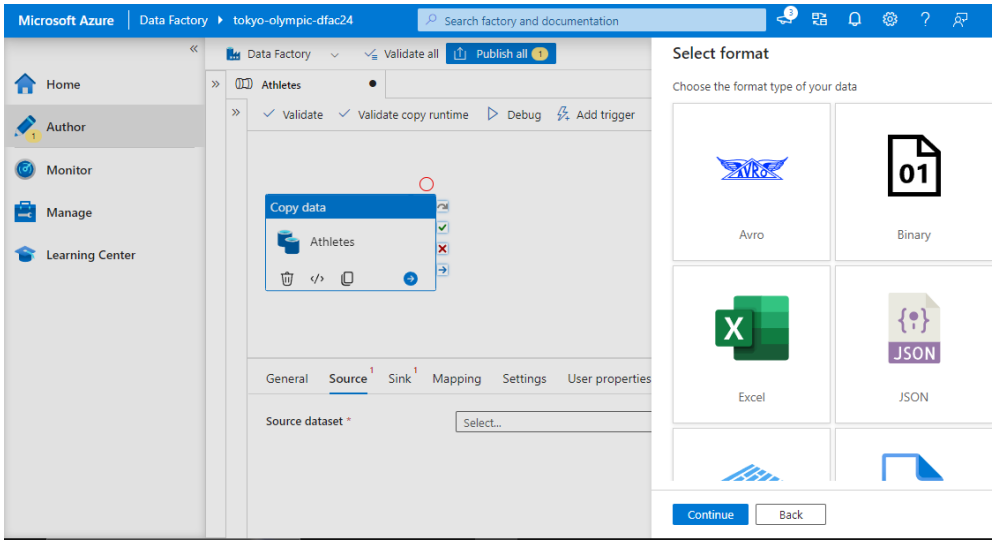
Created the pipeline → Used the Copy Data Activity → For Data Ingestion & pipeline creation.



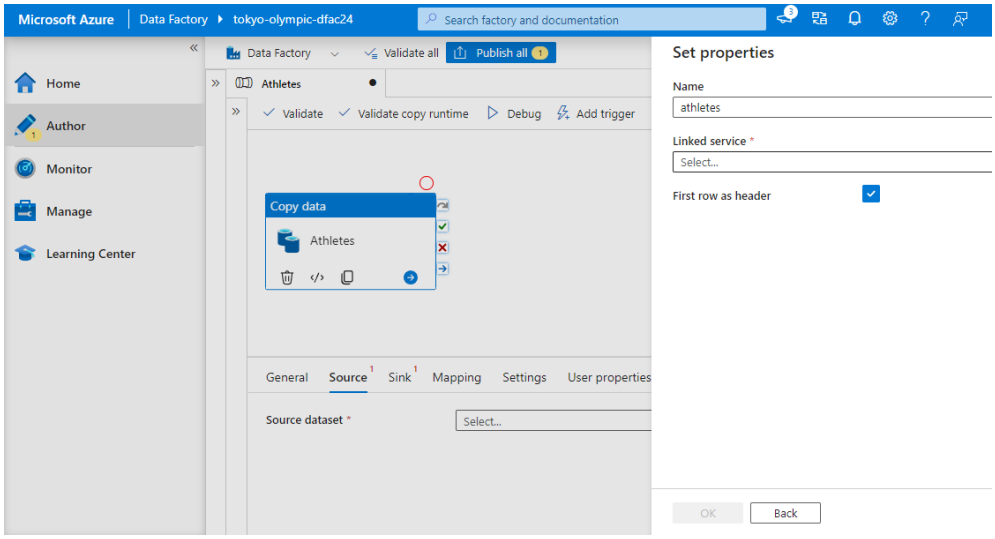


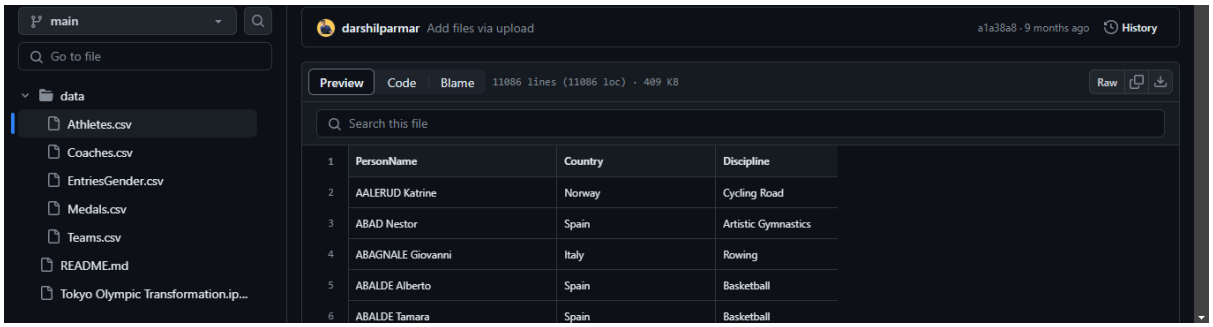




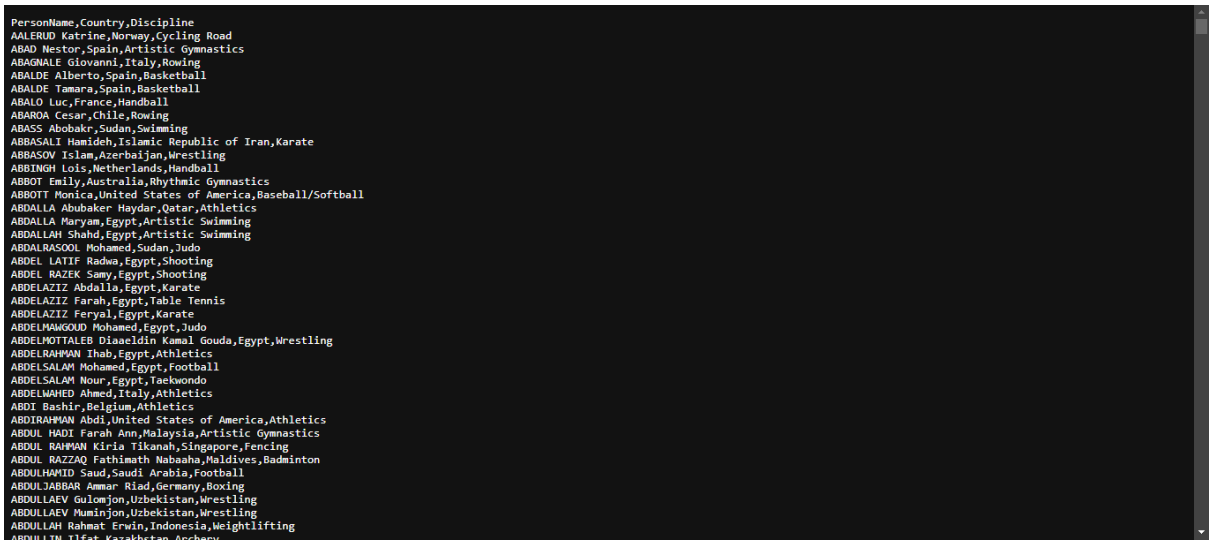


Source file is in CSV. So, I selected the Delimited Text

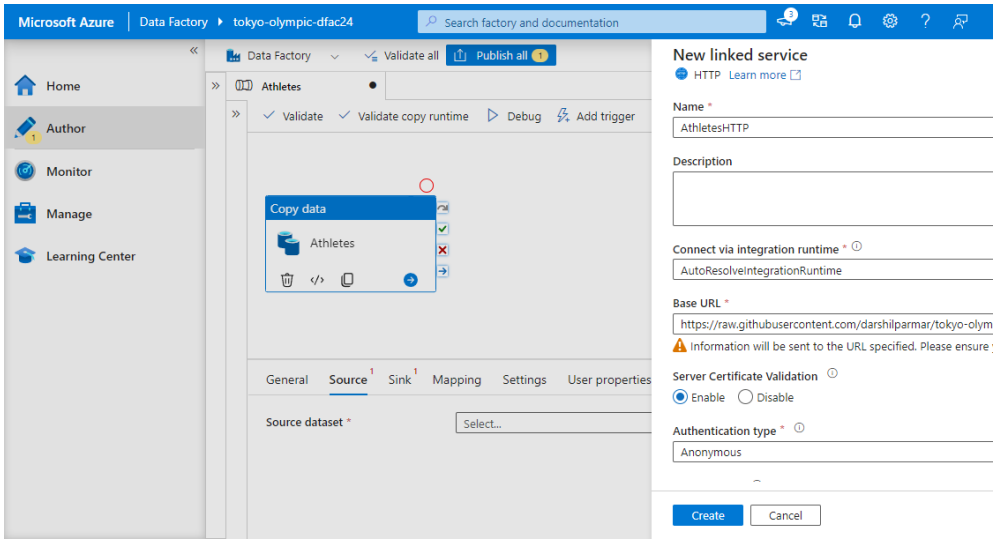




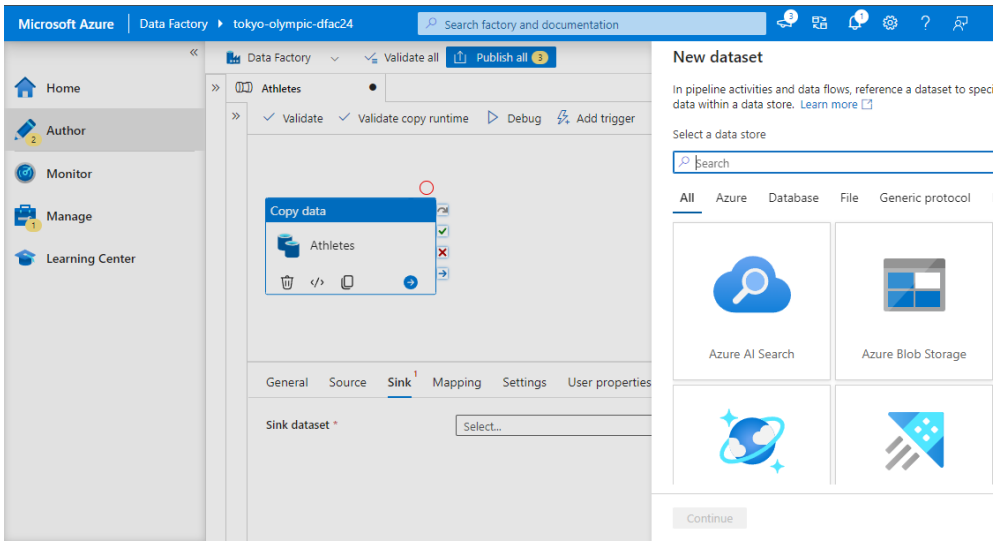
Github Data Source Link : https://raw.githubusercontent.com/darshilparmar/tokyo-olympic-azure-data-engineering-project/main/data/Athletes.csv

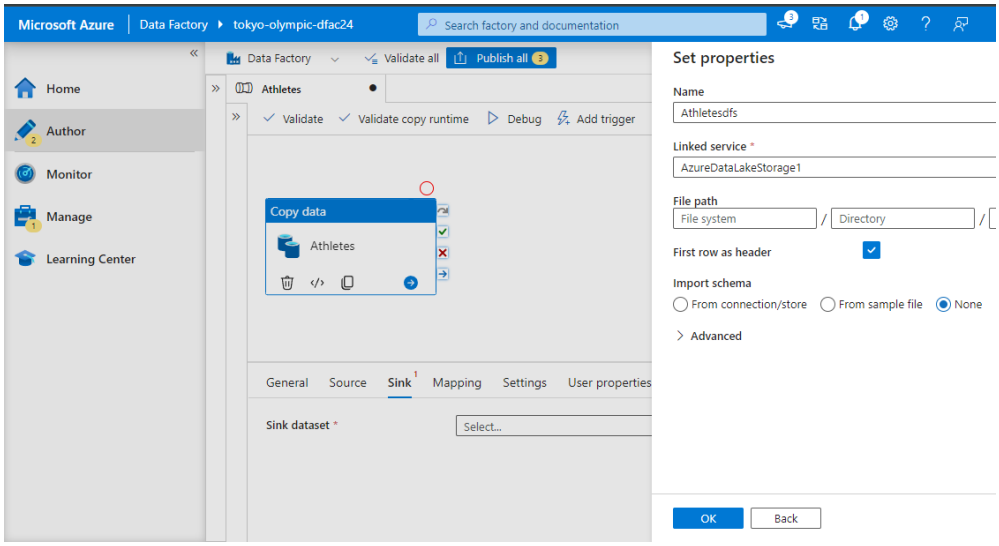


```
PersonName,Country,Discipline
AALERUD Katrine,Norway,Cycling Road
ABAD Nestor,Spain,Artistic Gymnastics
ABAGNALE Giovanni,Italy,Rowing
ABALDE Alberto,Spain,Basketball
ABALDE Tamara,Spain,Basketball
ABALO Luc,France,Handball
ABAROA Cesar,Chile,Rowing
ABASS Abobakr,Sudan,Swimming
ABBASALI Hamideh,Islamic Republic of Iran,Karate
ABBASOV Islam,Azerbaijan,Wrestling
ABBINGH Lois,Netherlands,Handball
ABBOT Emily,Australia,Rhythmic Gymnastics
ABBOTT Monica,United States of America,Baseball/Softball
ABDALLA Abubaker Haydar,Qatar,Athletics
ABDALLA Maryam,Egypt,Artistic Swimming
ABDALLAH Shahd,Egypt,Artistic Swimming
ABDALRASOOL Mohamed,Sudan,Judo
ABDEL LATIF Radwa,Egypt,Shooting
ABDEL RAZEK Samy,Egypt,Shooting
ABDELAZIZ Abdalla,Egypt,Karate
ABDELAZIZ Farah,Egypt,Table Tennis
ABDELAZIZ Feryal,Egypt,Karate
ABDELMAWGOUD Mohamed,Egypt,Judo
ABDELMOTTALEB Diaaeldin Kamal Gouda,Egypt,Wrestling
ABDELRAHMAN Ihab,Egypt,Athletics
ABDELSALAM Mohamed,Egypt,Football
ABDELSALAM Nour,Egypt,Taekwondo
ABDELWAHED Ahmed,Italy,Athletics
ABDI Bashir,Belgium,Athletics
ABDIRAHMAN Abdi,United States of America,Athletics
ABDUL HADI Farah Ann,Malaysia,Artistic Gymnastics
ABDUL RAHMAN Kiria Tikanah,Singapore,Fencing
ABDUL RAZZAQ Fathimath Nabaaha,Maldives,Badminton
ABDULHAMID Saud,Saudi Arabia,Football
ABDULJABBAR Ammar Riad,Germany,Boxing
ABDULLAEV Gulomjon,Uzbekistan,Wrestling
ABDULLAEV Muminjon,Uzbekistan,Wrestling
ABDULLAH Rahmat Erwin,Indonesia,Weightlifting
ABDULLIN Ilfat,Kazakhstan,Archery
```

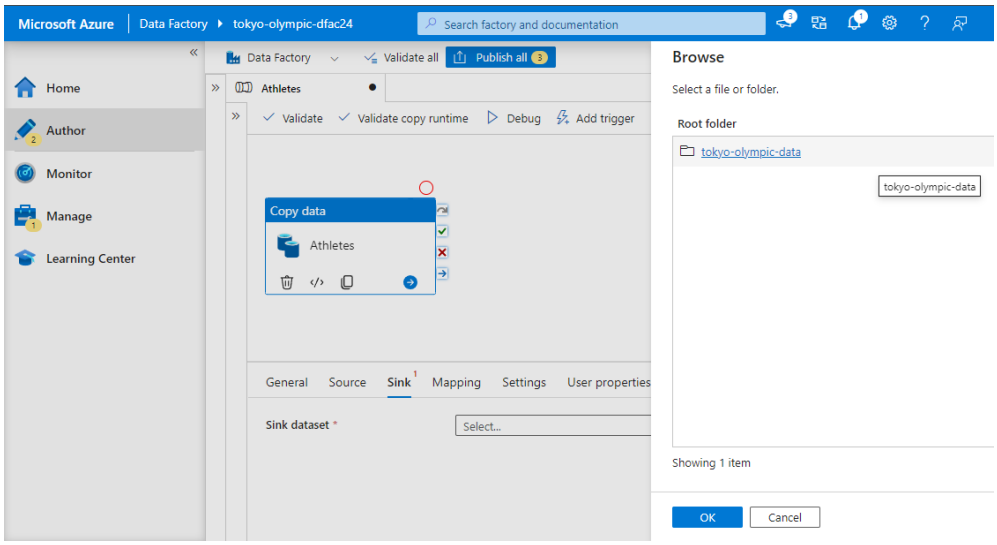


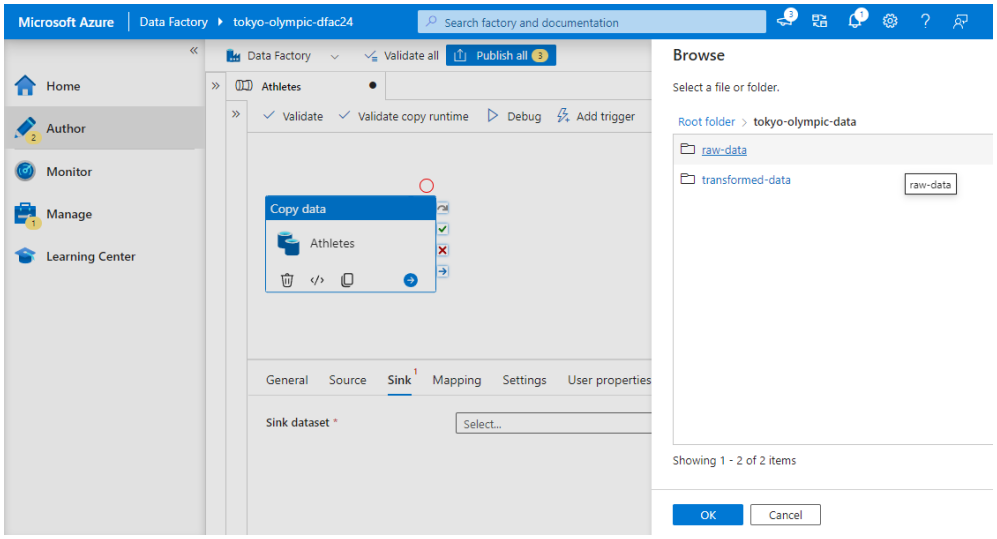
Copy and paste the data source URL and to create the Linked service between the Azure Data factory and GitHub api and ingest the Data from Github HTML API with the help of Linked service.



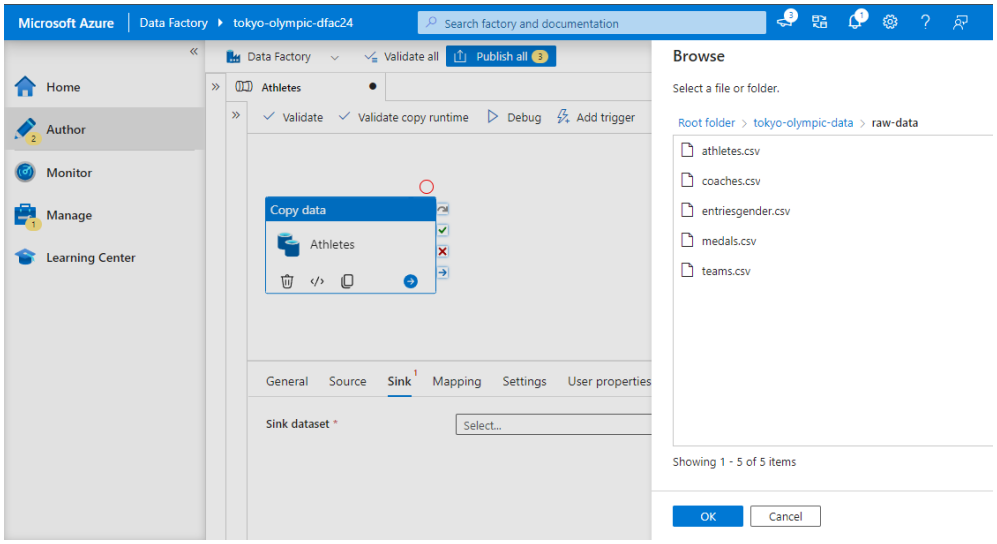


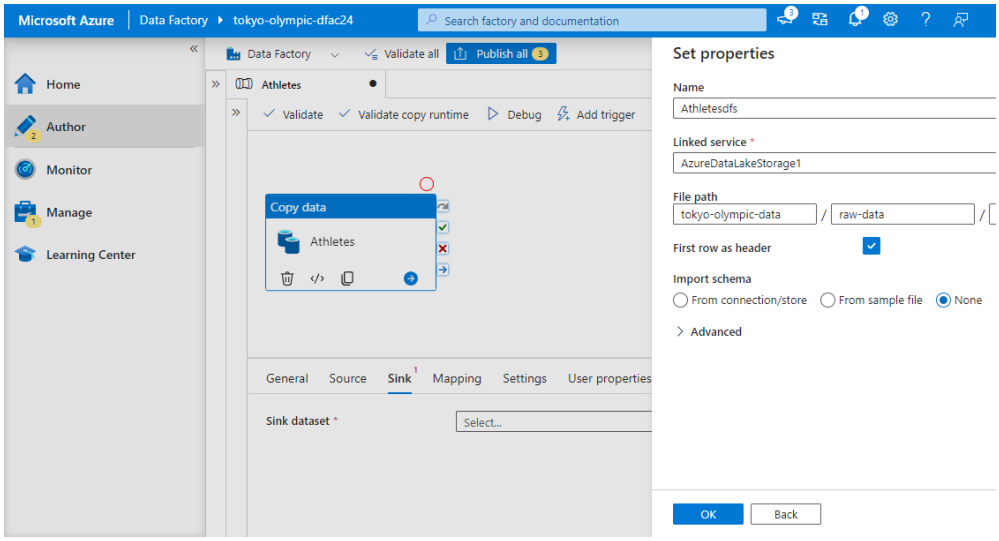
Configure the SINK folder in the Azure Data Lake Gen2 Storage → create containers as Raw DATA, Transformed data.

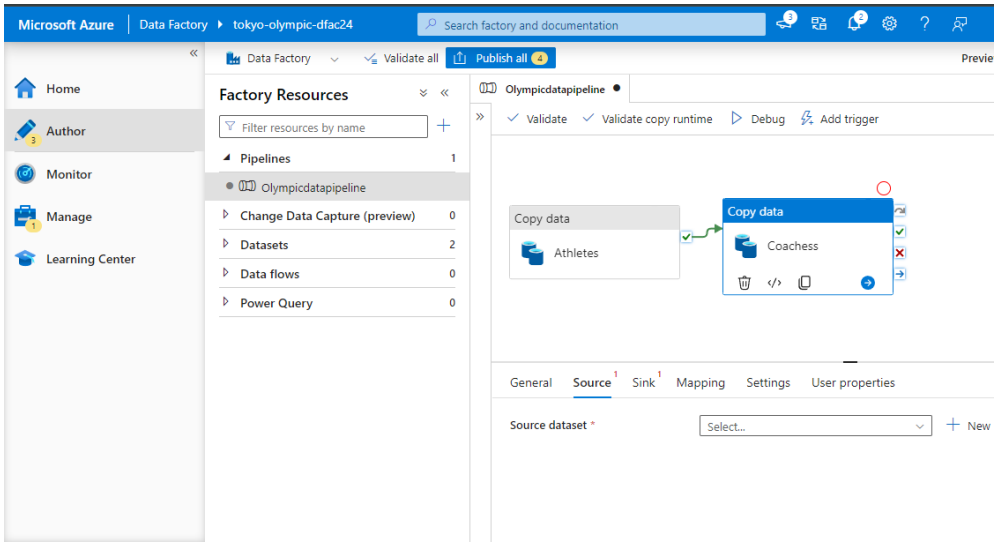


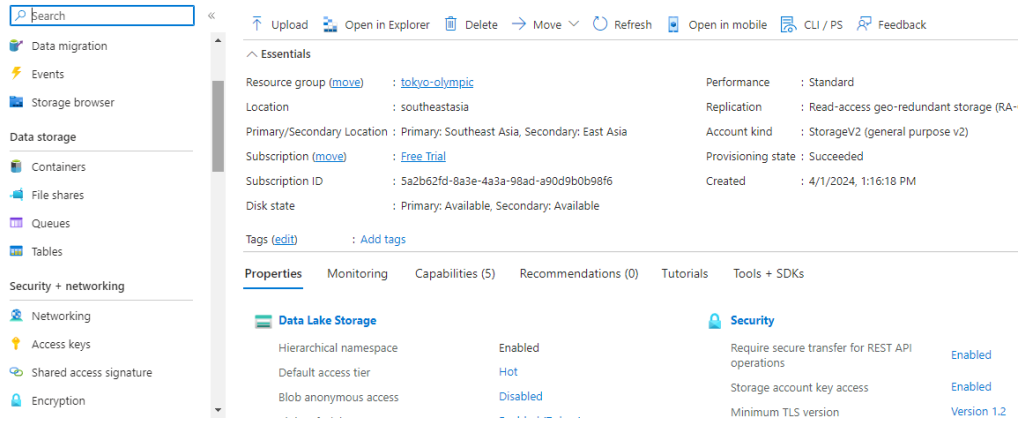


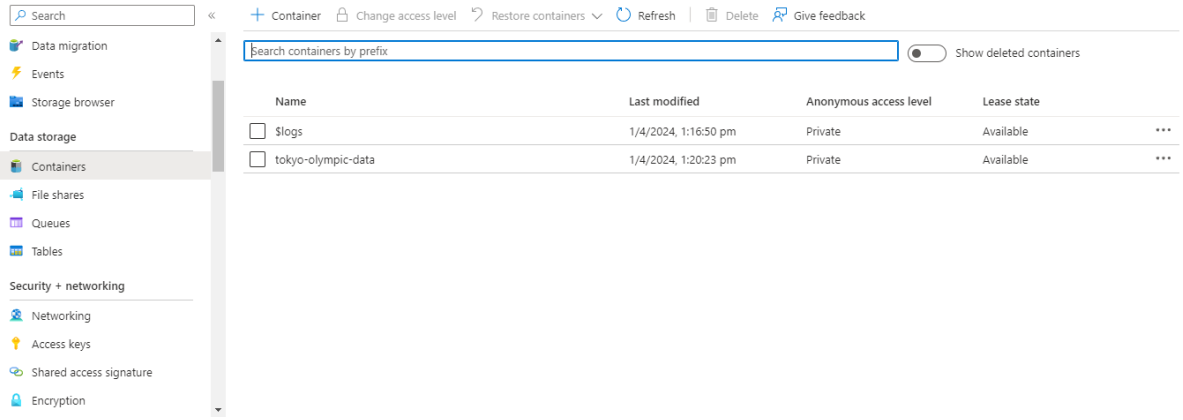
create containers as Raw DATA, Transformed data.

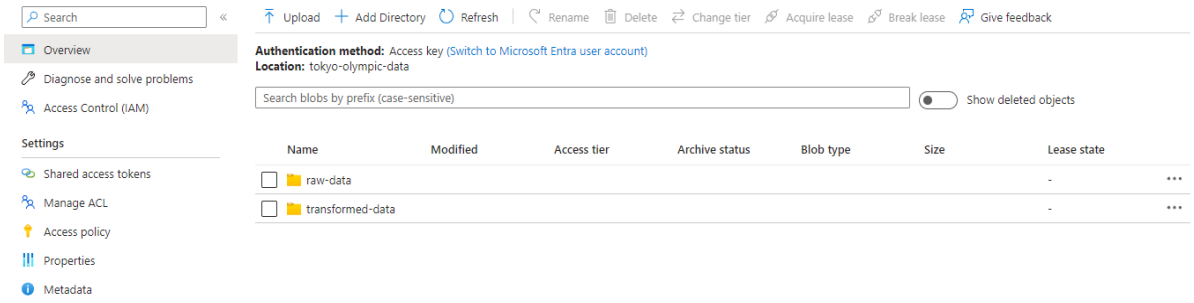




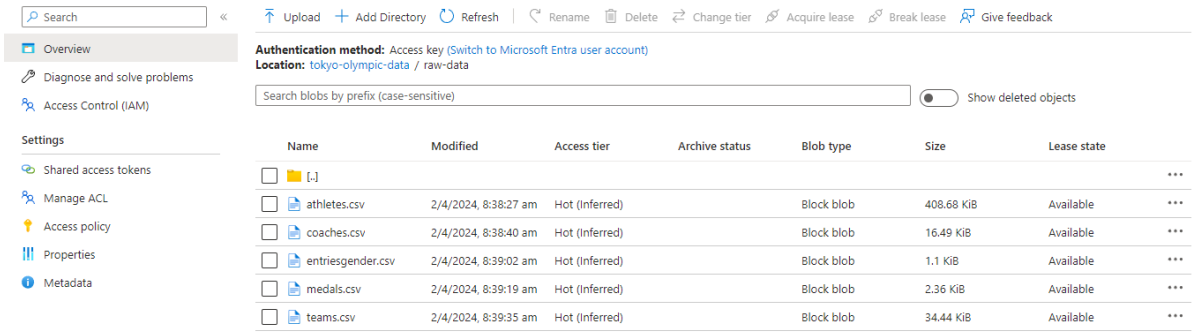

Microsoft Azure
Data Factory
tokyo-olympic-dfac24
Search factory and documentation
Data Factory
Validate all
Publish all
Home
Author
Monitor
Manage
Learning Center
Factory Resources
Filter resources by name
Pipelines
Olympicdatapipeline
Change Data Capture (preview)
Datasets
Data flows
Power Query
Olympicdatapipeline
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Athletes
Copy data
Coachess
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
Select...
New

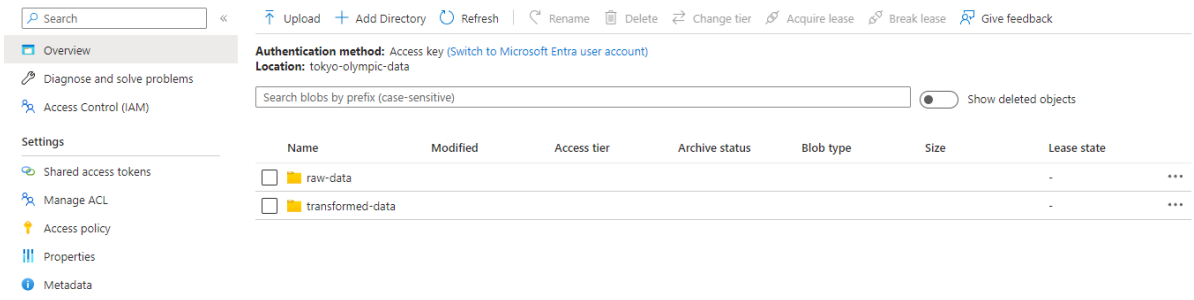

Search
Data migration
Events
Storage browser
Data storage
Containers
File shares
Queues
Tables
Security + networking
Networking
Access keys
Shared access signature
Encryption
Upload
Open in Explorer
Delete
Move
Refresh
Open in mobile
CLI / PS
Feedback
Essentials
Resource group (move) : tokyo-olympic
Location : southeastasia
Primary/Secondary Location : Primary: Southeast Asia, Secondary: East Asia
Subscription (move) : Free Trial
Subscription ID : 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6
Disk state : Primary: Available, Secondary: Available
Performance : Standard
Replication : Read-access geo-redundant storage (RA-
Account kind : StorageV2 (general purpose v2)
Provisioning state : Succeeded
Created : 4/1/2024, 1:16:18 PM
Tags (edit) : Add tags
Properties
Monitoring
Capabilities (5)
Recommendations (0)
Tutorials
Tools + SDKs
Data Lake Storage
Hierarchical namespace Enabled
Default access tier Hot
Blob anonymous access Disabled
Security
Require secure transfer for REST API operations Enabled
Storage account key access Enabled
Minimum TLS version Version 1.2


Search
Data migration
Events
Storage browser
Data storage
Containers
File shares
Queues
Tables
Security + networking
Networking
Access keys
Shared access signature
Encryption
Container
Change access level
Restore containers
Refresh
Delete
Give feedback
Search containers by prefix
Show deleted containers
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

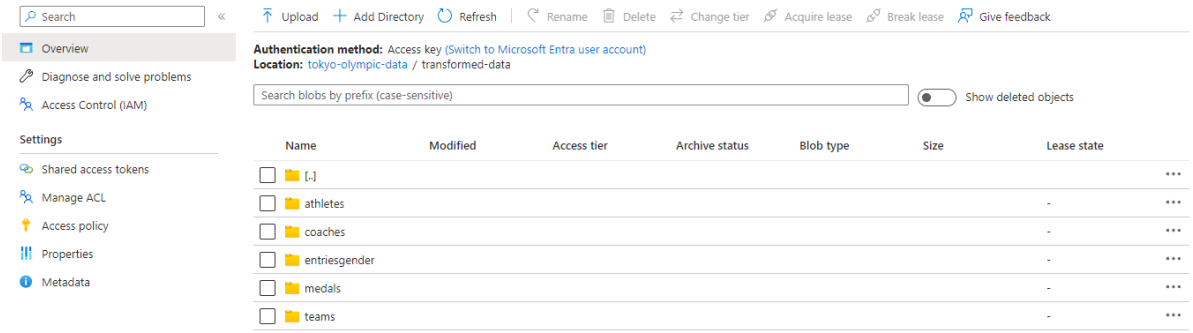


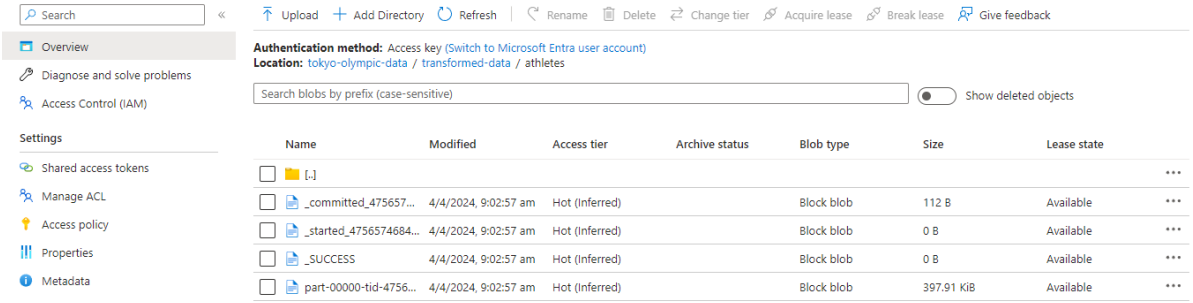
10. After the pipeline completes its execution, navigate to your Azure Data Lake Storage Gen2. Dive into the "raw_data" folder and validate that the files, like "athletes.csv", "medals.csv", etc., are present and populated with the expected data.

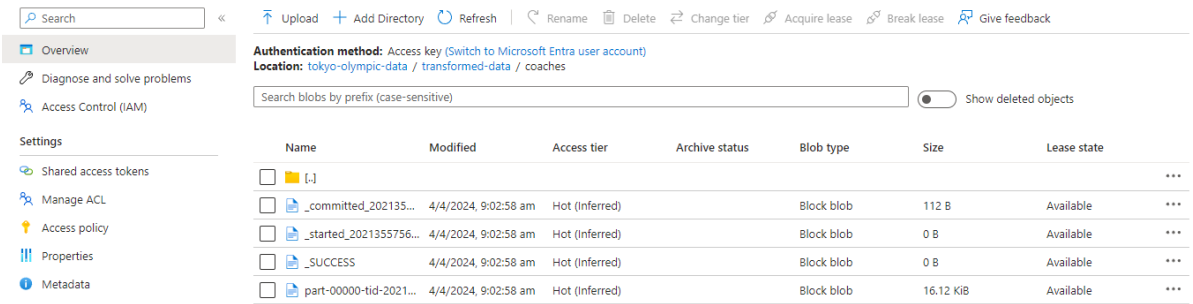


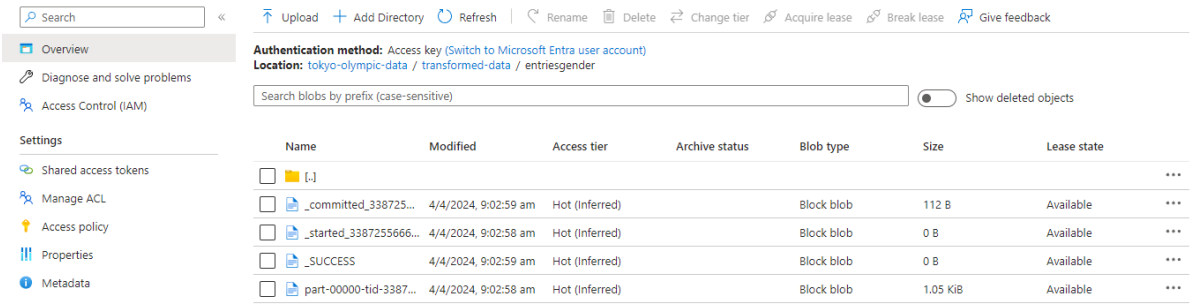


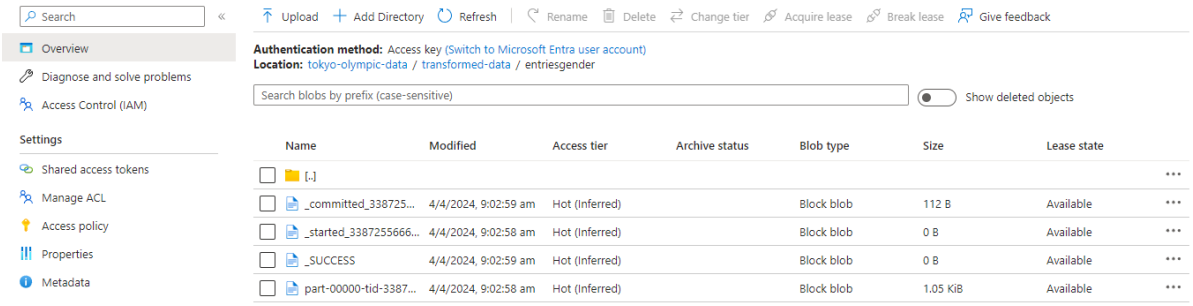
Check all the data available in the container.

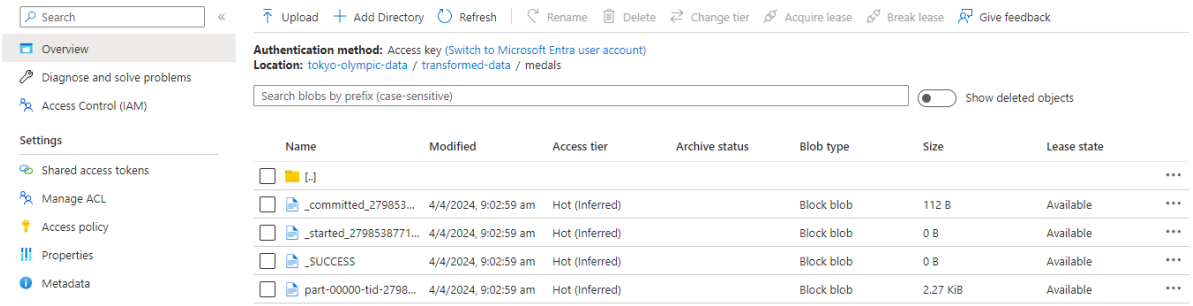

Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
athletes
coaches
entriesgender
medals
teams

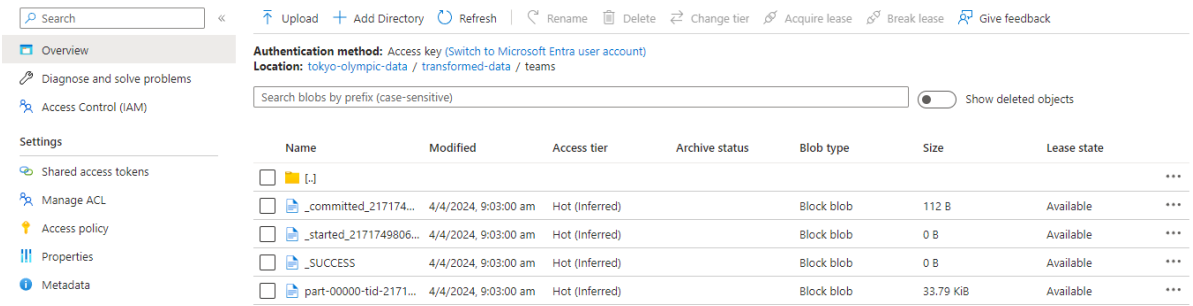

Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / athletes
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_475657...
_started_4756574684...
_SUCCESS
part-00000-tid-4756...
4/4/2024, 9:02:57 am
Hot (Inferred)
Block blob
112 B
0 B
397.91 KiB
Available

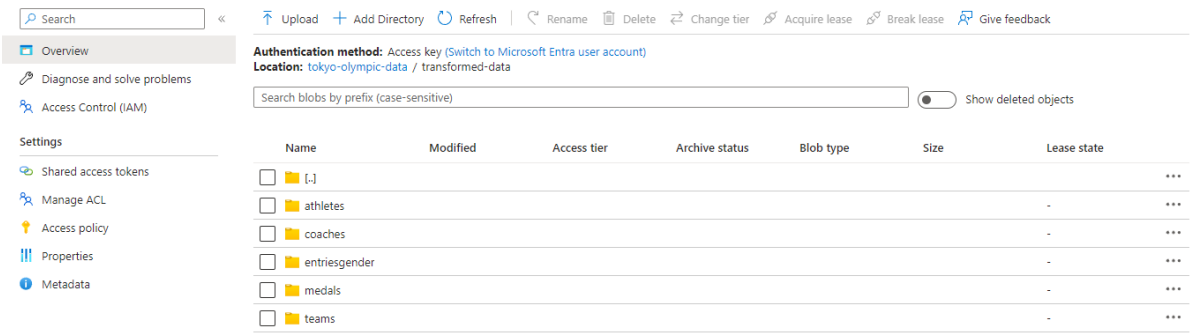

Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / coaches
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_202135...
_started_2021355756...
_SUCCESS
part-00000-tid-2021...
4/4/2024, 9:02:58 am
Hot (Inferred)
Block blob
112 B
0 B
16.12 KiB
Available

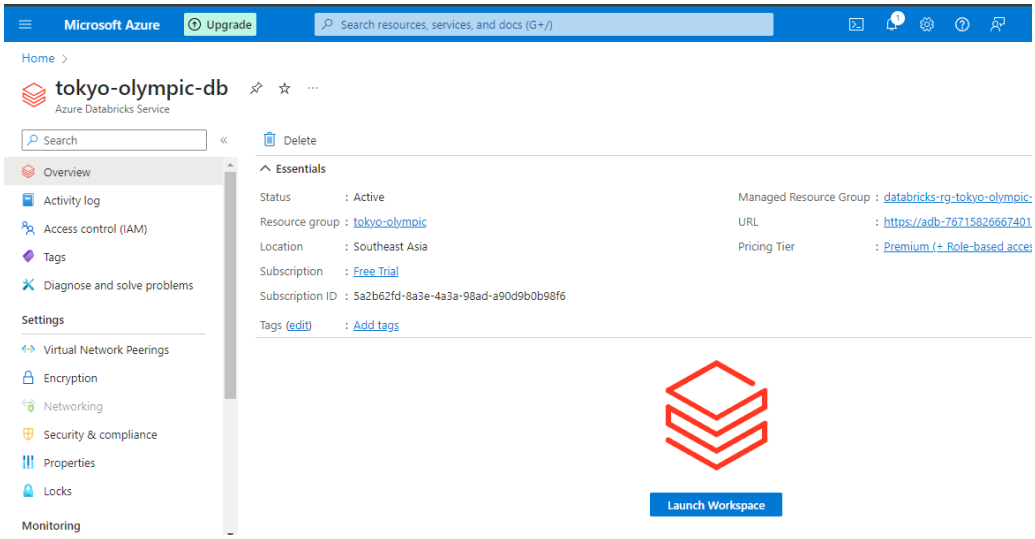

Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / entriesgender
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_338725...
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
112 B
Available
_started_3387255666...
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
0 B
Available
part-00000-tid-3387...
4/4/2024, 9:02:58 am
Hot (Inferred)
Block blob
1.05 KiB
Available

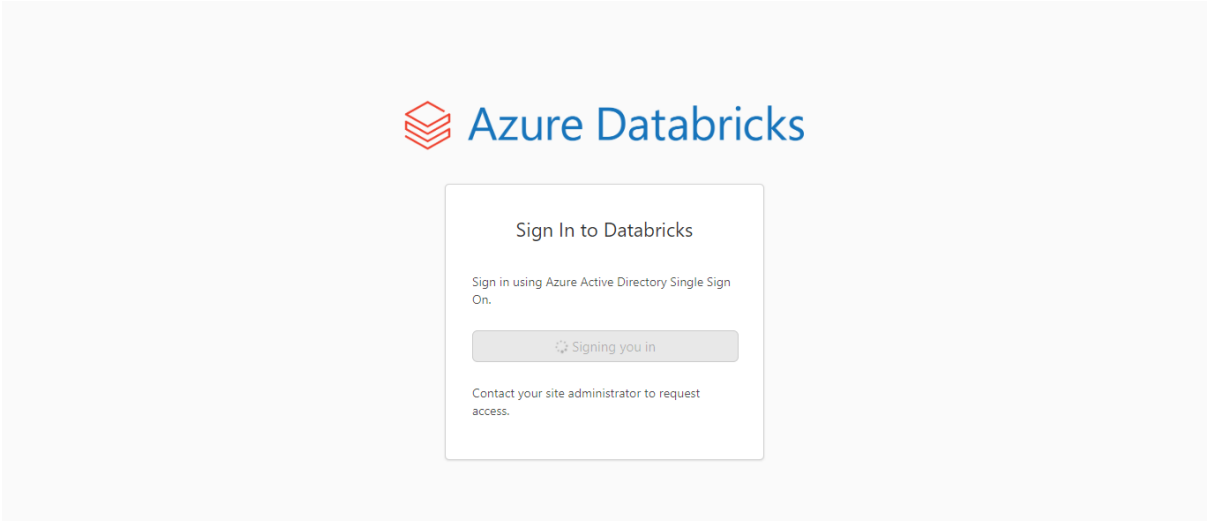

Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / entriesgender
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_338725...
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
112 B
Available
_started_3387255666...
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
0 B
Available
part-00000-tid-3387...
4/4/2024, 9:02:58 am
Hot (Inferred)
Block blob
1.05 KiB
Available

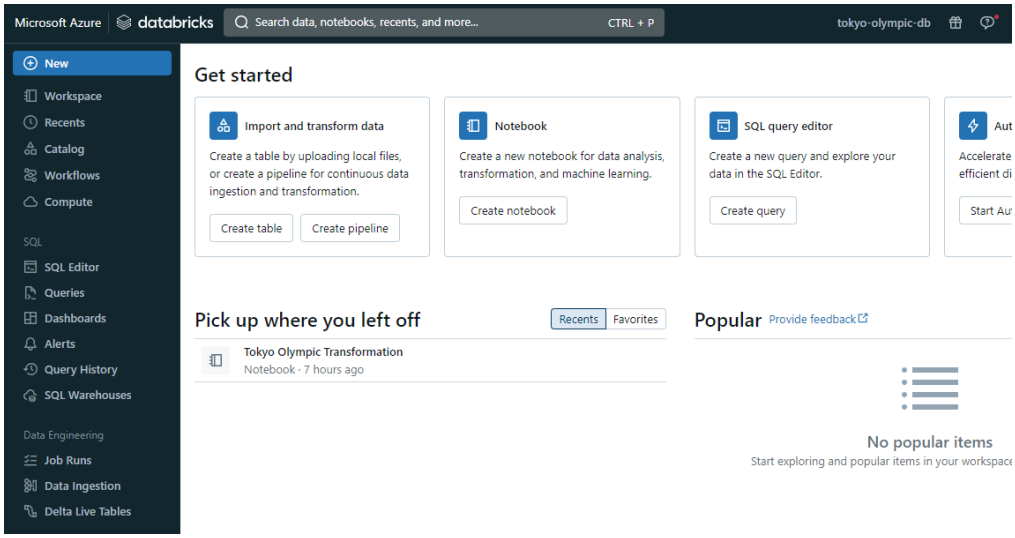

Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / medals
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_279853...
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
112 B
Available
_started_2798538771...
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
0 B
Available
_SUCCESS
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
0 B
Available
part-00000-tid-2798...
4/4/2024, 9:02:59 am
Hot (Inferred)
Block blob
2.27 KiB
Available


Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data / teams
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
_committed_217174...
4/4/2024, 9:03:00 am
Hot (Inferred)
Block blob
112 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
0 B
Available
part-00000-tid-2171...
4/4/2024, 9:03:00 am
Hot (Inferred)
Block blob
33.79 KiB
Available


Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: tokyo-olympic-data / transformed-data
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
Lease state
[..]
athletes
coaches
entriesgender
medals
teams


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Home
tokyo-olympic-db
Azure Databricks Service
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Settings
Virtual Network Peerings
Encryption
Networking
Security & compliance
Properties
Locks
Monitoring
Delete
Essentials
Status : Active
Resource group : tokyo-olympic
Location : Southeast Asia
Subscription : Free Trial
Subscription ID : 5a2b62fd-8a3e-4a3a-98ad-a90d9b0b98f6
Tags (edit) : Add tags
Managed Resource Group : databricks-rg-tokyo-olympic-
URL : https://adb-76715826667401
Pricing Tier : Premium (+ Role-based acces
Launch Workspace

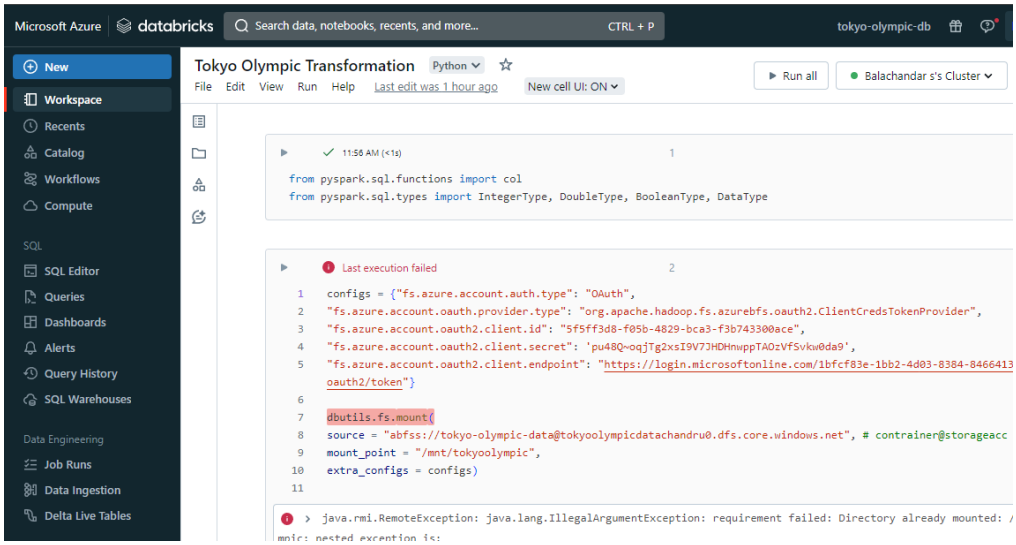

Azure Databricks
Sign In to Databricks
Sign in using Azure Active Directory Single Sign On.
Signing you in
Contact your site administrator to request access.



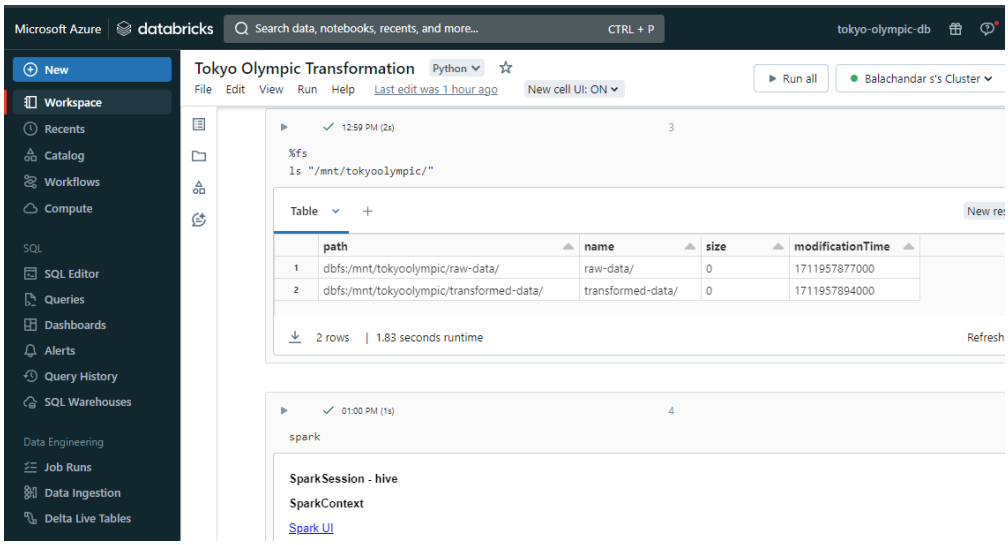
Do this transformation by using Data Bricks.

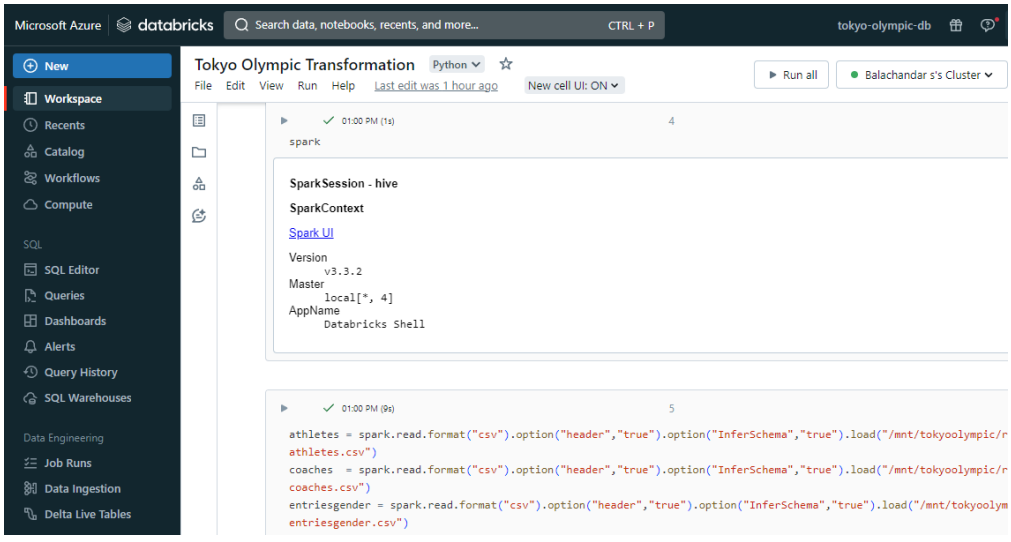
Data Transformation using Azure Databricks

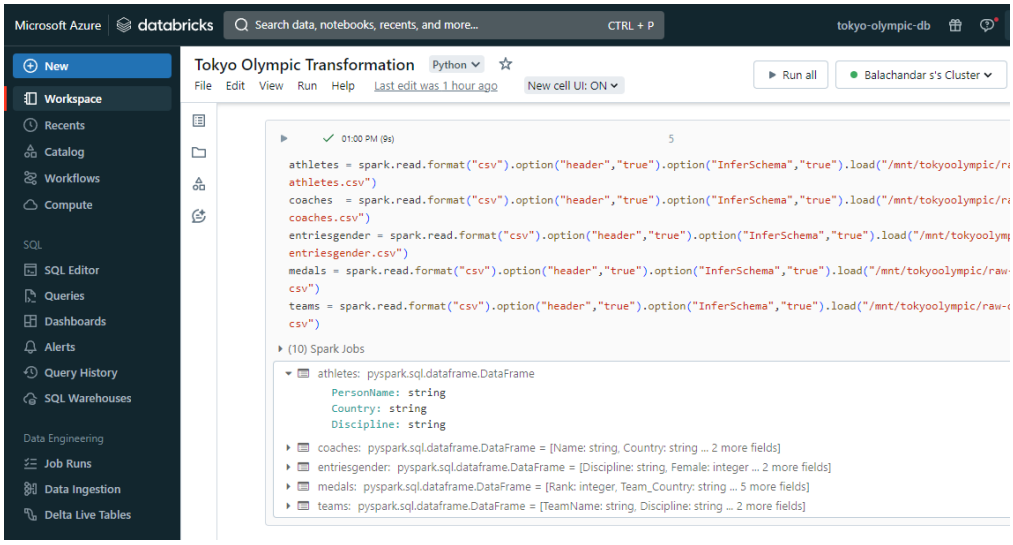
1. Navigate to Azure Databricks within the Azure portal and create a workspace within the previously established resource group and launch it.
2. Configuring Compute in Databricks
3. Create a new notebook within Databricks and rename it appropriately, reflecting its purpose or the dataset it pertains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

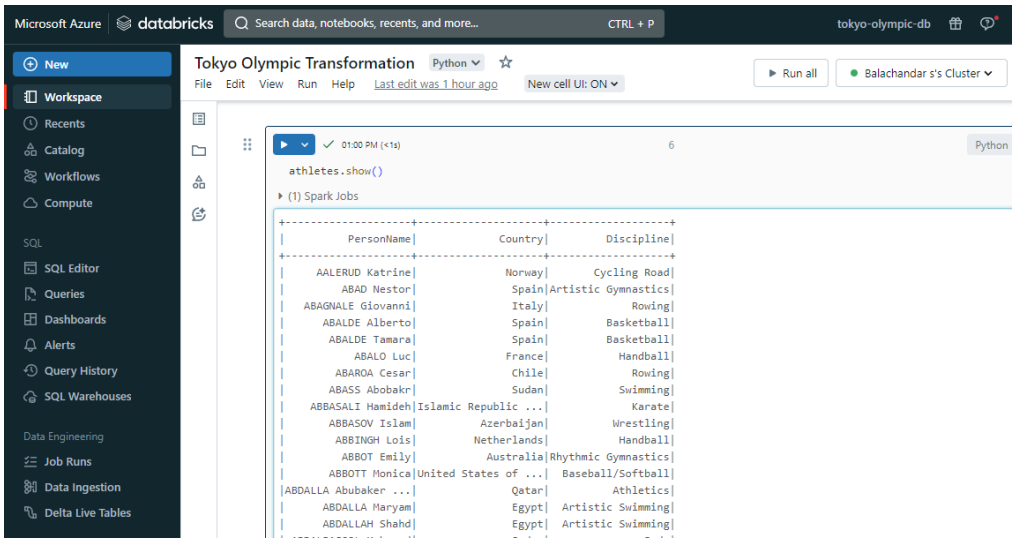


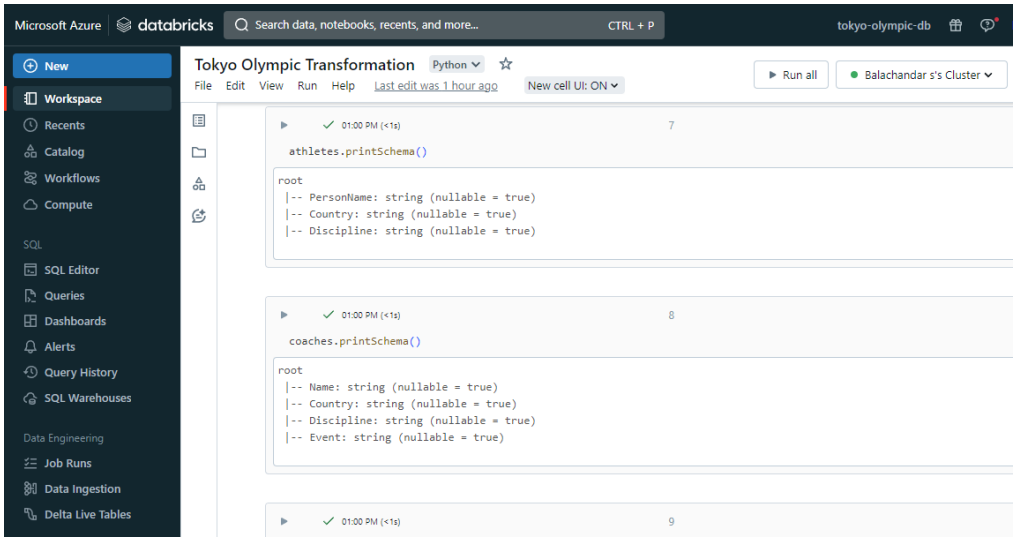
Create the Compute Resource & Mount the data into Data-Bricks.

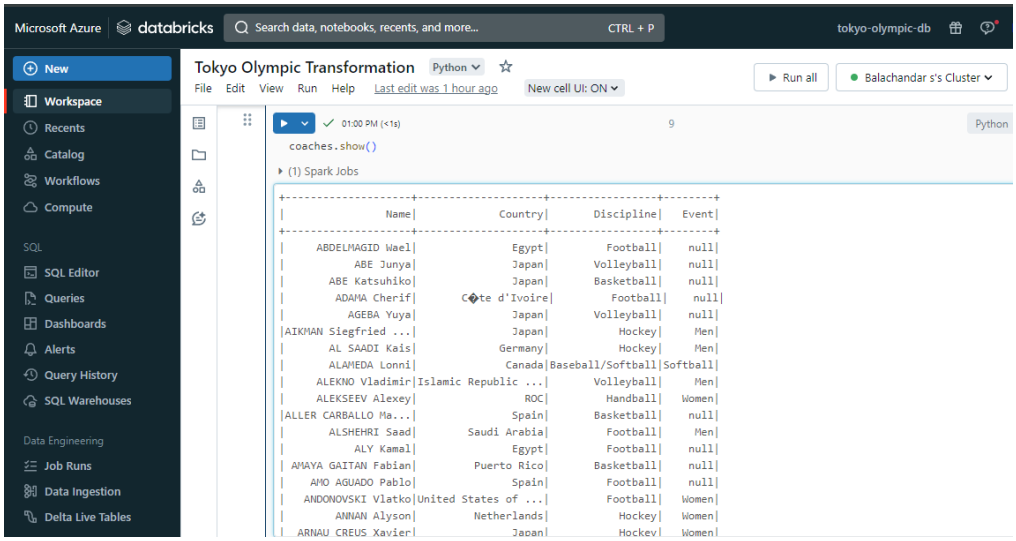


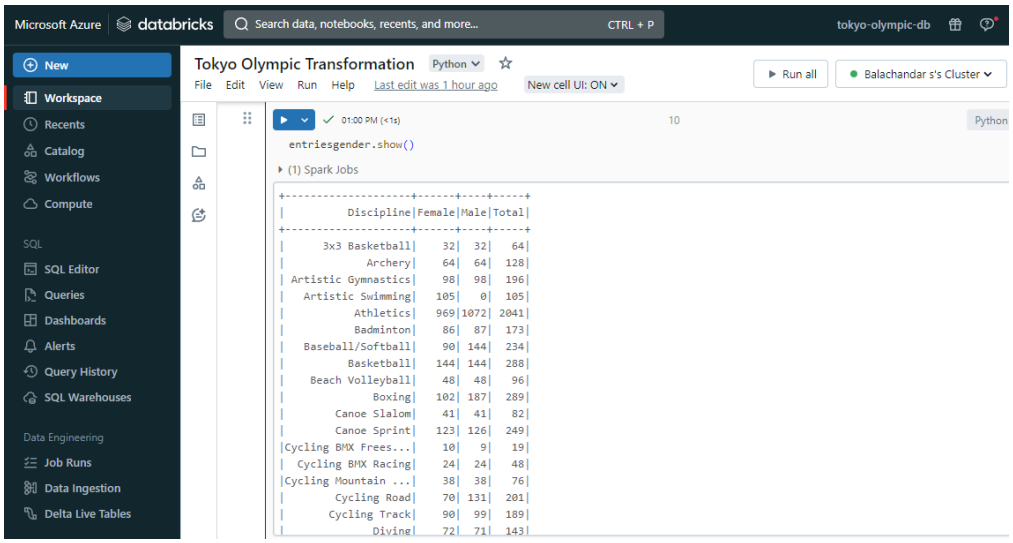


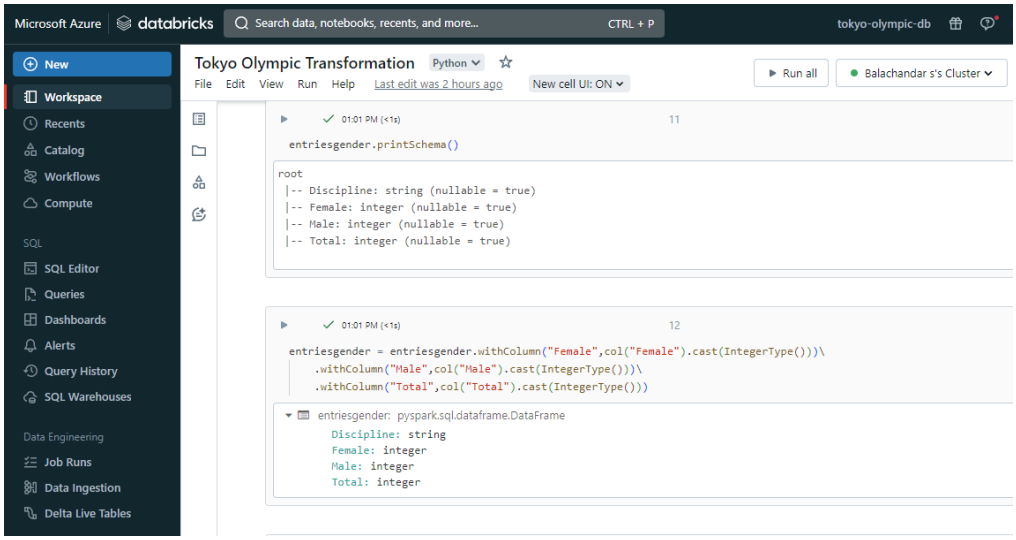

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 1 hour ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:00 PM (9s)
5
athletes = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/r
athletes.csv")
coaches = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/r
coaches.csv")
entriesgender = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolym
entriesgender.csv")
medals = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw
csv")
teams = spark.read.format("csv").option("header","true").option("InferSchema","true").load("/mnt/tokyoolympic/raw-
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]

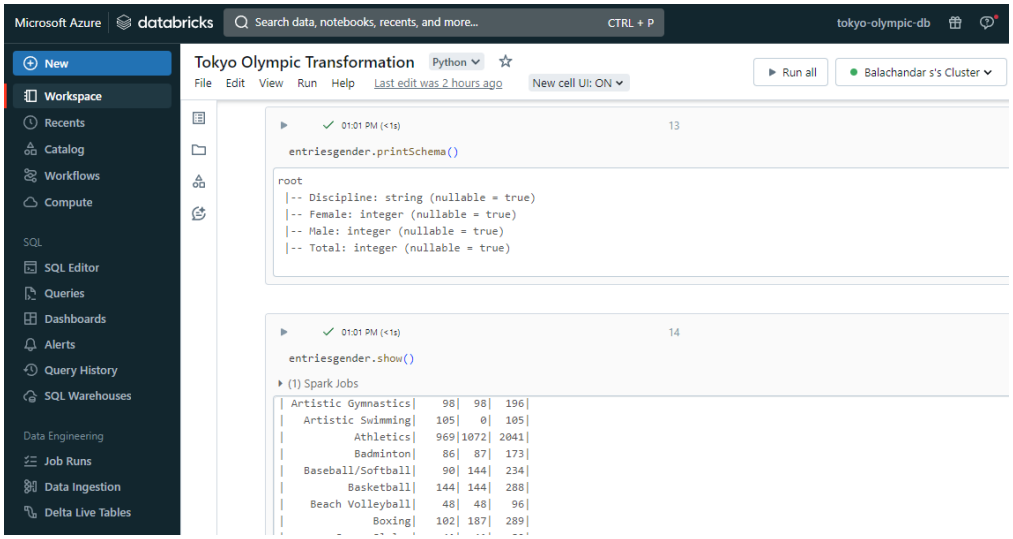

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 1 hour ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:00 PM (<1s)
6
Python
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
ABBOT 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


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 1 hour ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:00 PM (<1s)
7
athletes.printSchema()
root
|-- PersonName: string (nullable = true)
|-- Country: string (nullable = true)
|-- Discipline: string (nullable = true)
01:00 PM (<1s)
8
coaches.printSchema()
root
|-- Name: string (nullable = true)
|-- Country: string (nullable = true)
|-- Discipline: string (nullable = true)
|-- Event: string (nullable = true)
01:00 PM (<1s)
9

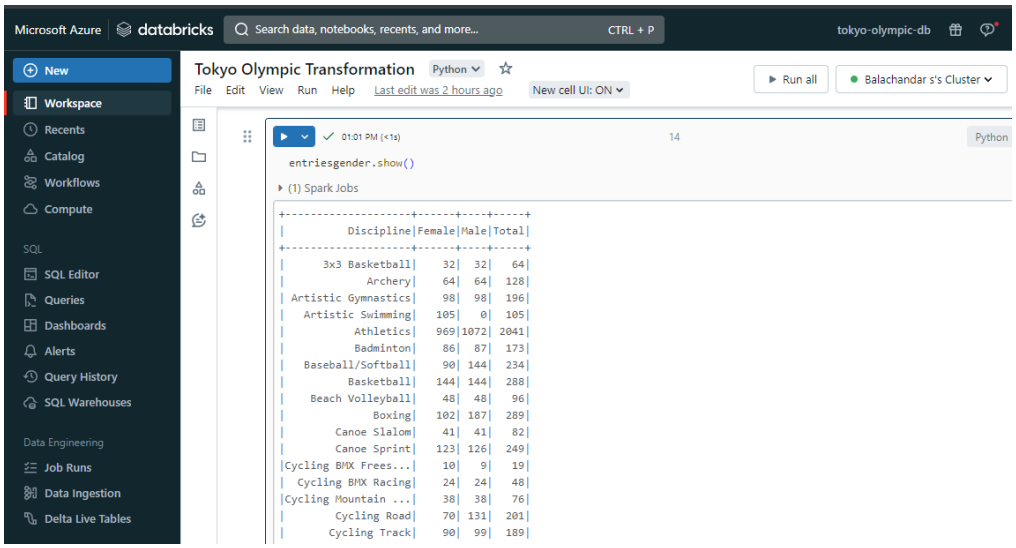

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 1 hour ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:00 PM (<1s)
9
Python
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	C�te d'Ivoire	Football	null
AGEBA Yuya	Japan	Volleyball	null
AIKMAN Siegfried ...	Japan	Hockey	Men
AL SAADI Kais	Germany	Hockey	Men
ALAMEDA Lonni	Canada	Baseball/Softball	Softball
ALEKNO Vladimir	Islamic Republic ...	Volleyball	Men
ALEKSEEV Alexey	ROC	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
AMO AGUADO Pablo	Spain	Football	null
ANDONOVSKI Vlatko	United States of ...	Football	Women
ANNAN Alyson	Netherlands	Hockey	Women
ARNAU CREUS Xavier	Japan	Hockey	Women

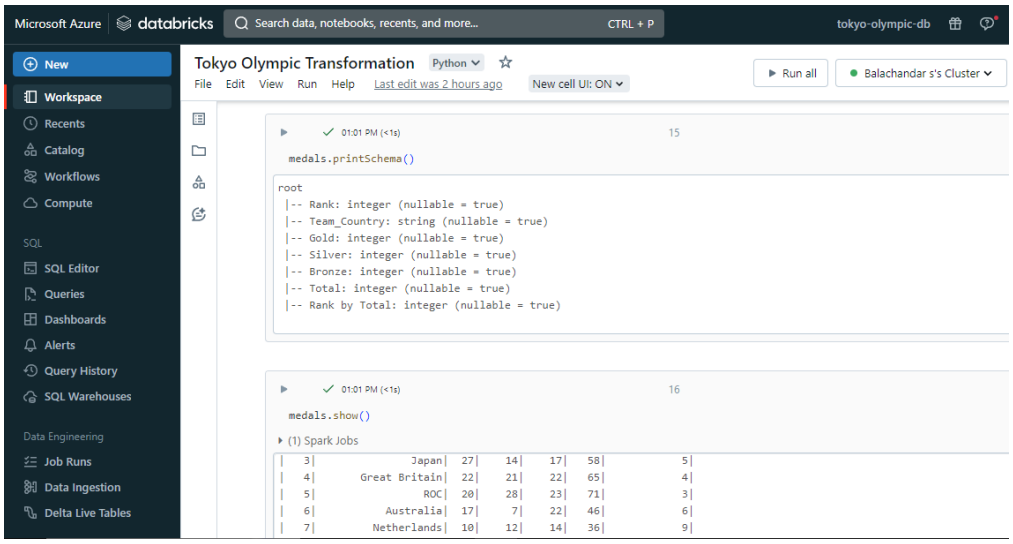

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 1 hour ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:00 PM (<1s)
10
Python
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 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
Diving	72	71	143

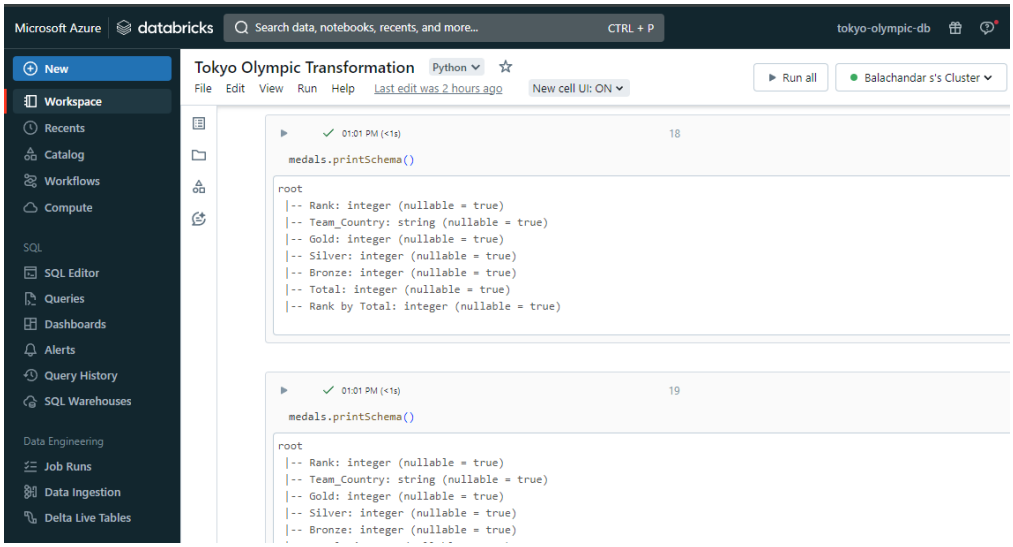

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:01 PM (<1s)
11
entriesgender.printSchema()
root
|-- Discipline: string (nullable = true)
|-- Female: integer (nullable = true)
|-- Male: integer (nullable = true)
|-- Total: integer (nullable = true)
01:01 PM (<1s)
12
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

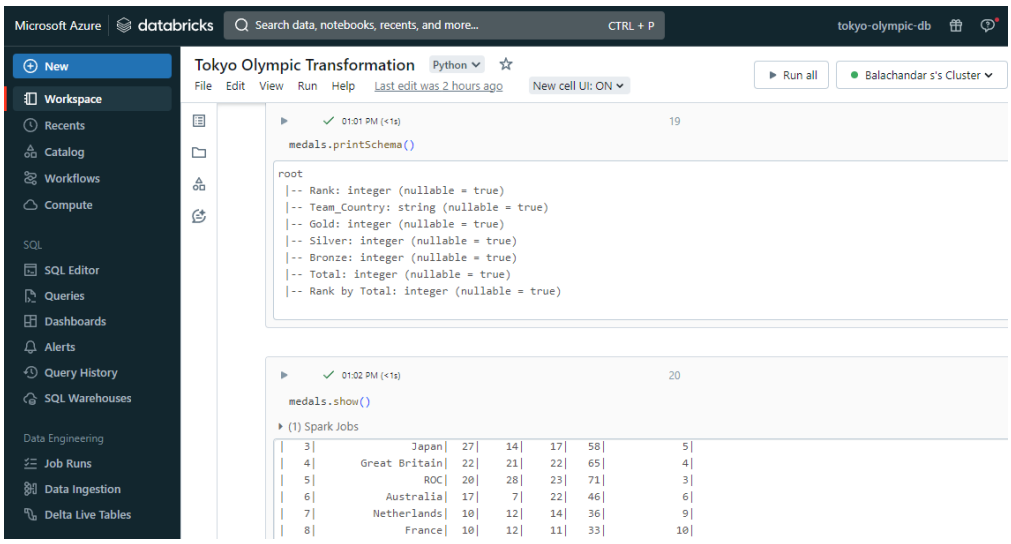


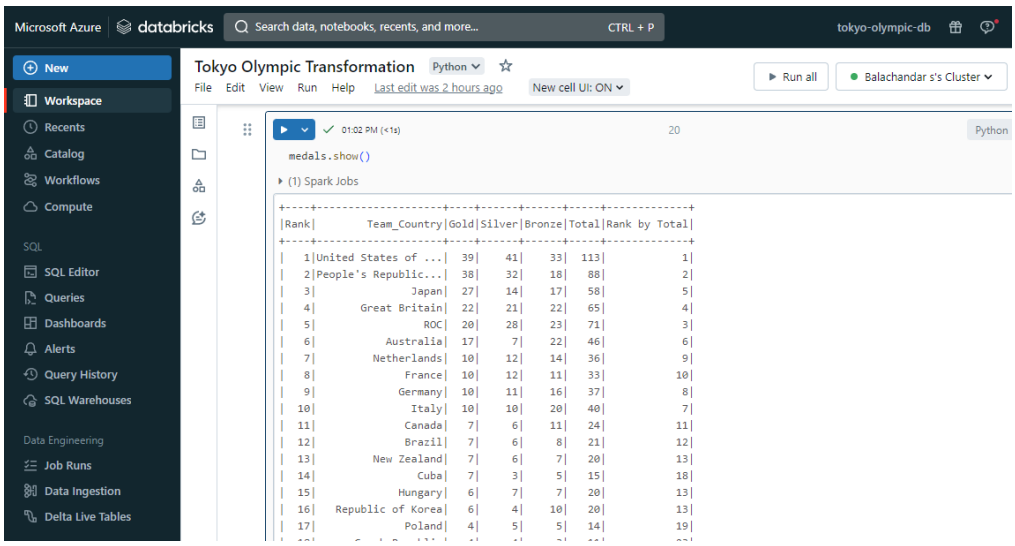
Data Transformation part

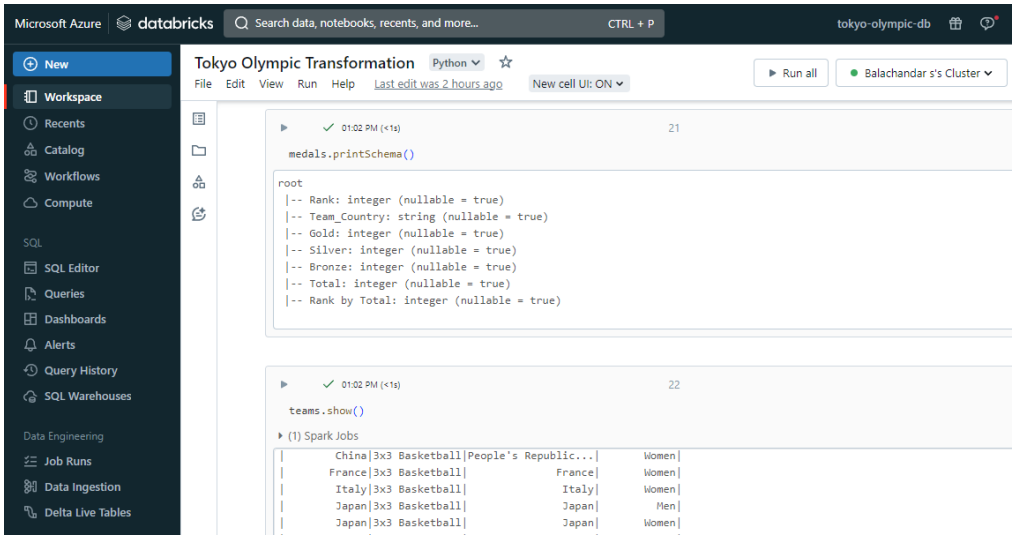


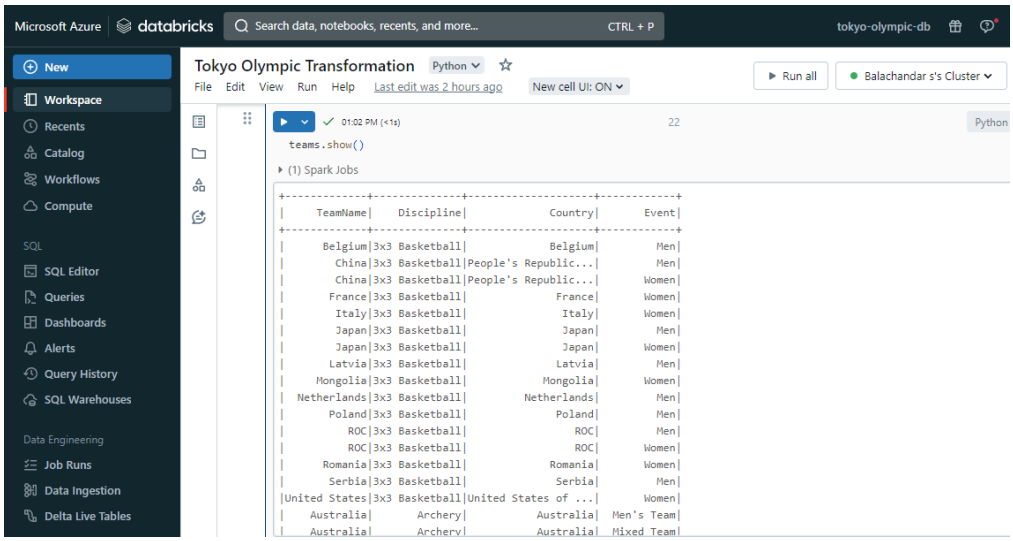


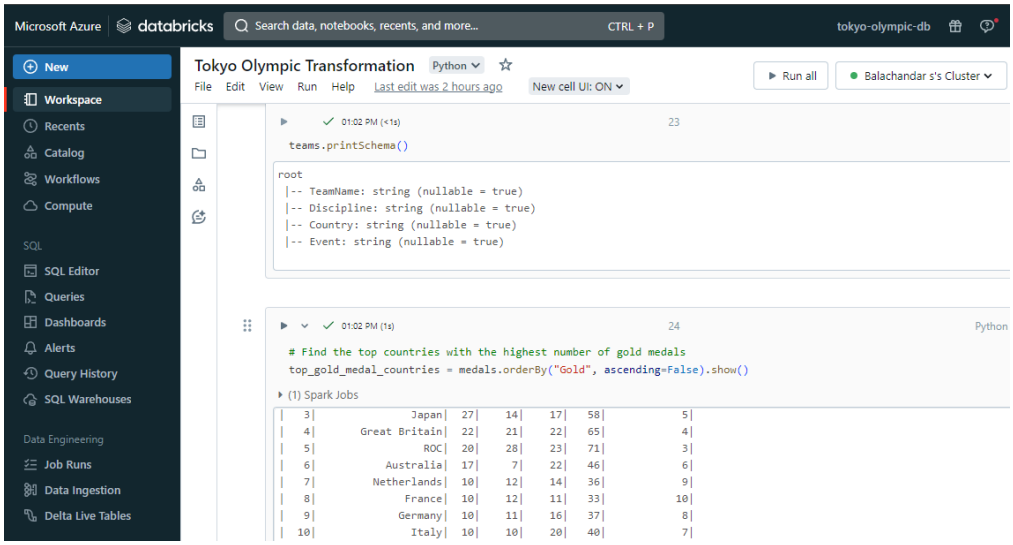

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:01 PM (<1s)
18
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)
01:01 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)

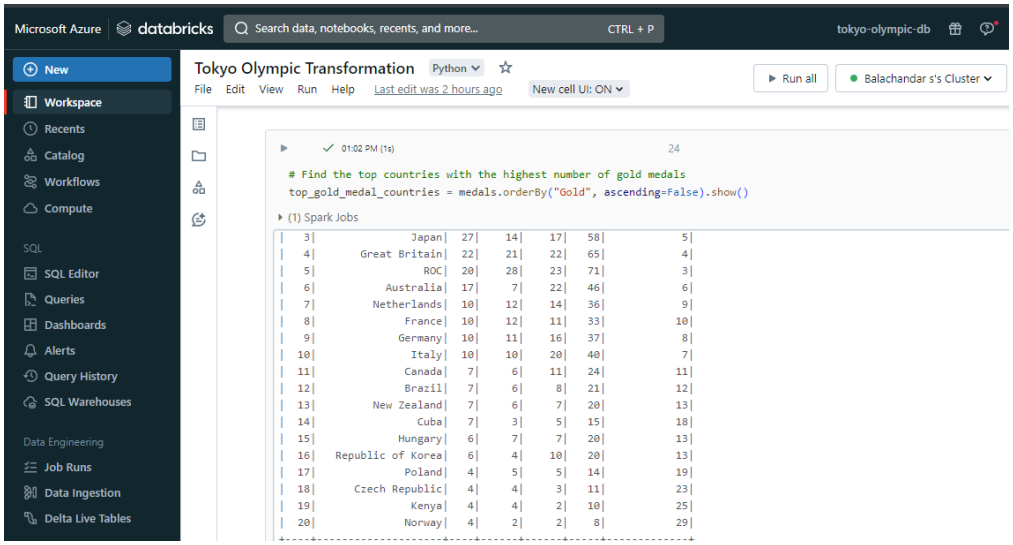

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
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
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

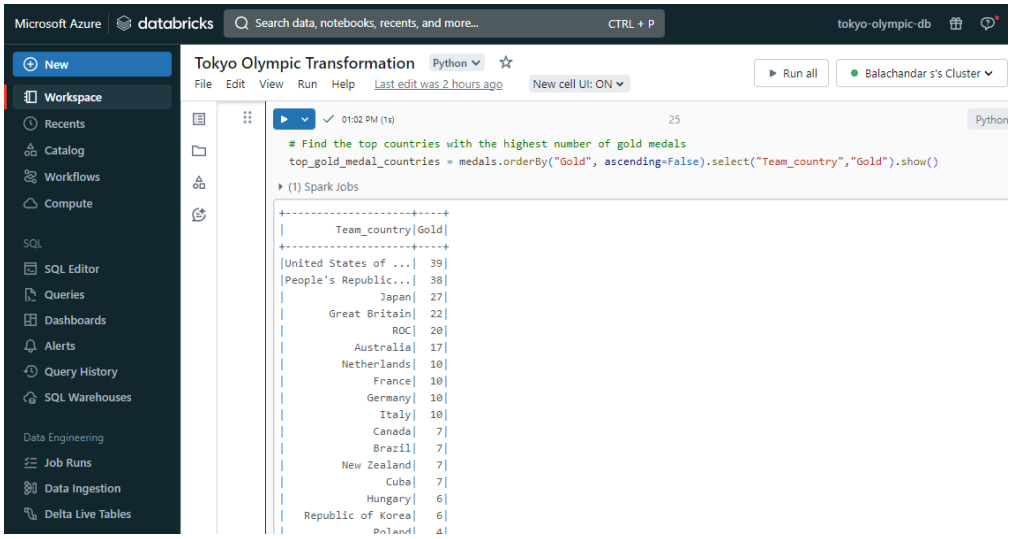

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:02 PM (<1s)
20
Python
medals.show()
(1) Spark Jobs
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

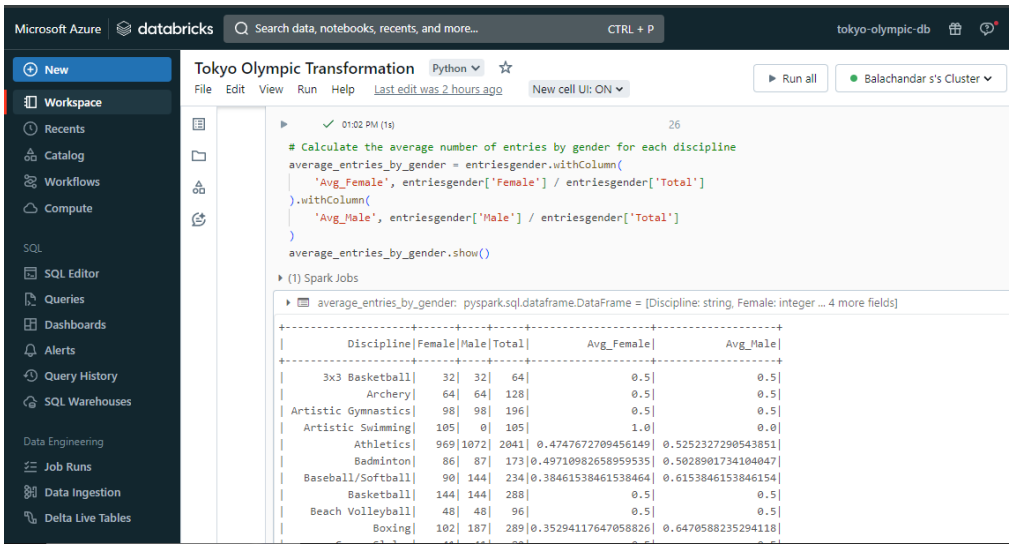

Microsoft Azure
databricks
Search data, notebooks, recents, and more... CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
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
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
Tokyo Olympic Transformation
Python
File Edit View Run Help Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:02 PM (<1s)
22
Python
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

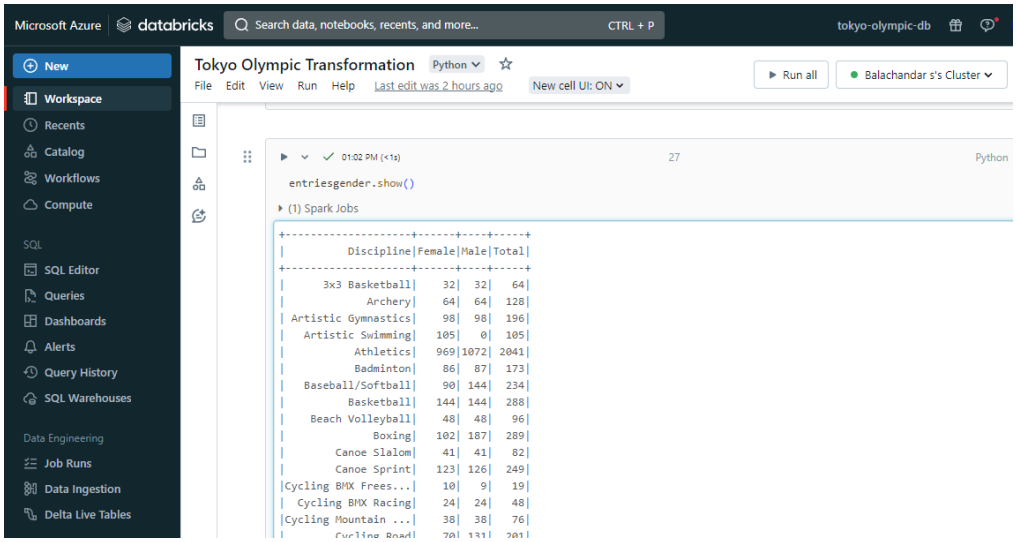

Microsoft Azure
databricks
Search data, notebooks, recents, and more... CTRL + P
tokyo-olympic-db
Tokyo Olympic Transformation
Python
File Edit View Run Help Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:02 PM (<1s)
23
teams.printSchema()
root
|-- TeamName: string (nullable = true)
|-- Discipline: string (nullable = true)
|-- Country: string (nullable = true)
|-- Event: string (nullable = true)
01:02 PM (1s)
24
Python
# Find the top countries with the highest number of gold medals
top_gold_medal_countries = medals.orderBy("Gold", ascending=False).show()
(1) Spark Jobs
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

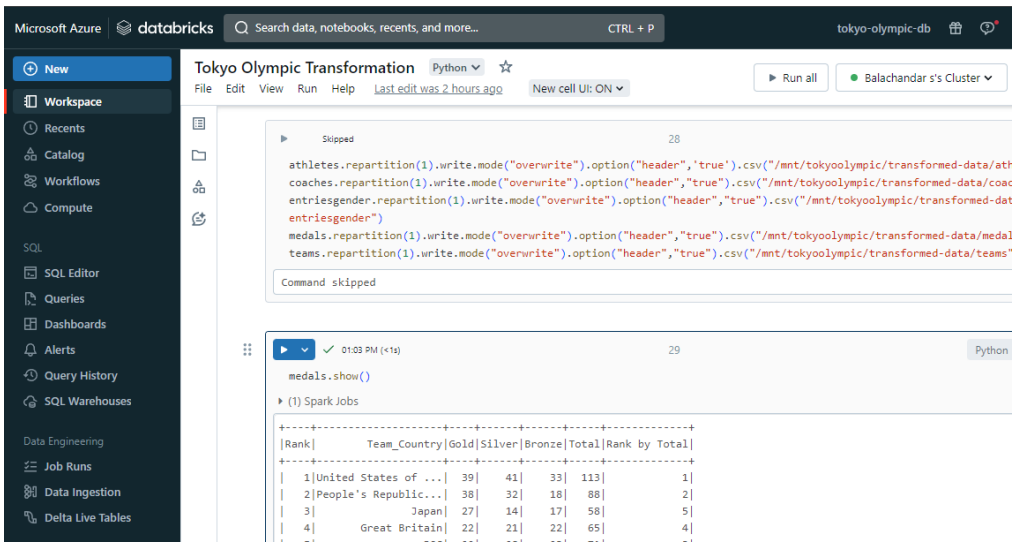


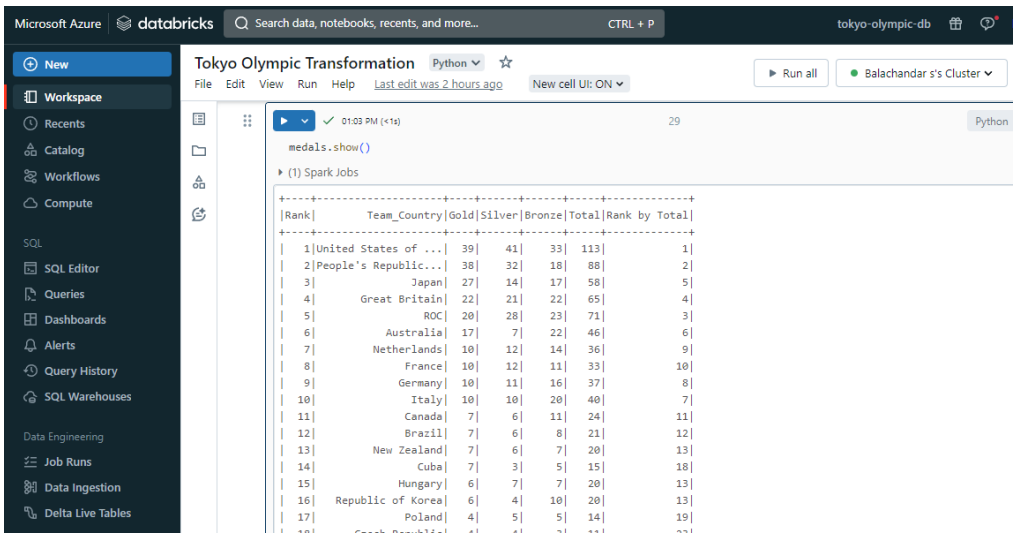


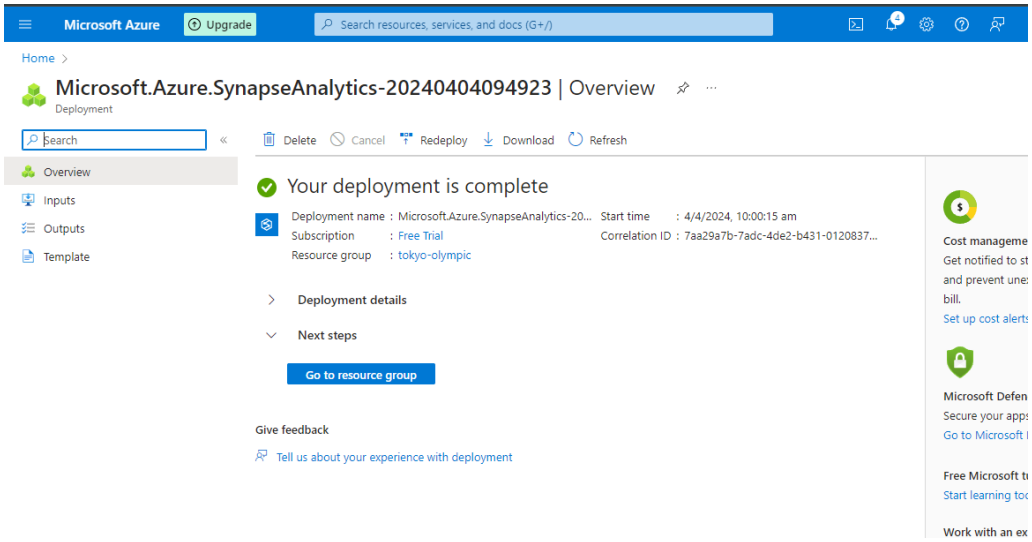


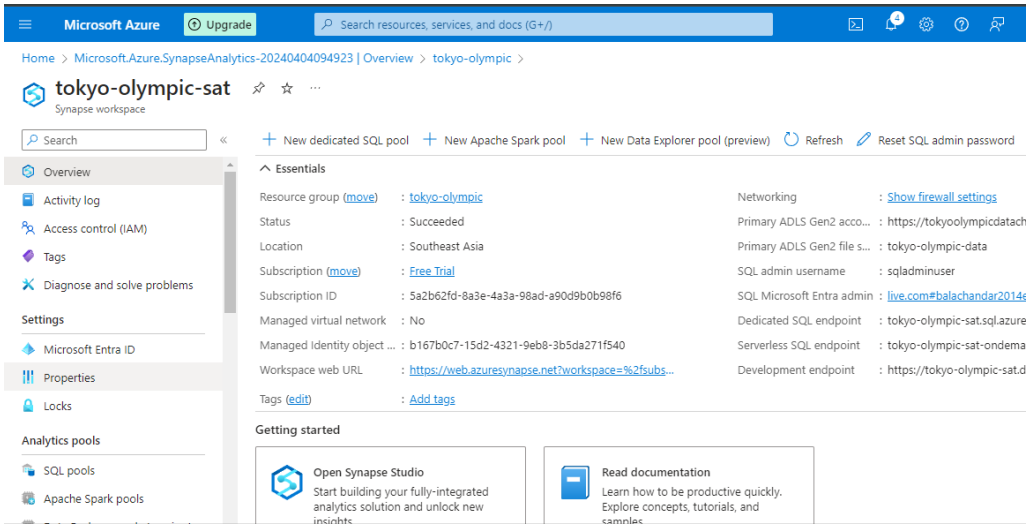
Do some transformation →Calculate the Average Number of Entries for each discipline.


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:02 PM (<1s)
27
Python
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 Frees...	10	9	19
Cycling BMX Racing	24	24	48
Cycling Mountain ...	38	38	76

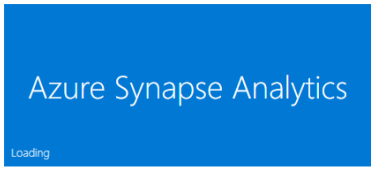

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
Skipped
28
athletes.repartition(1).write.mode("overwrite").option("header",'true').csv("/mnt/tokyoolympic/transformed-data/ath
coaches.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/coac
entriesgender.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-dat
entriesgender")
medals.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/medal
teams.repartition(1).write.mode("overwrite").option("header","true").csv("/mnt/tokyoolympic/transformed-data/teams"
Command skipped
01:03 PM (<1s)
29
Python
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


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
tokyo-olympic-db
Tokyo Olympic Transformation
Python
File Edit View Run Help
Last edit was 2 hours ago
New cell UI: ON
Run all
Balachandar s's Cluster
01:03 PM (<1s)
29
Python
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



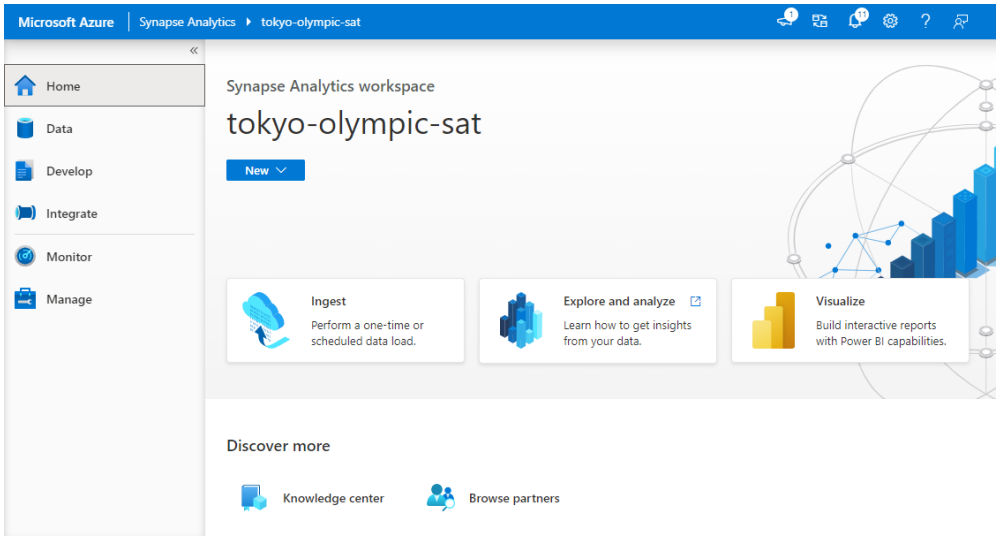


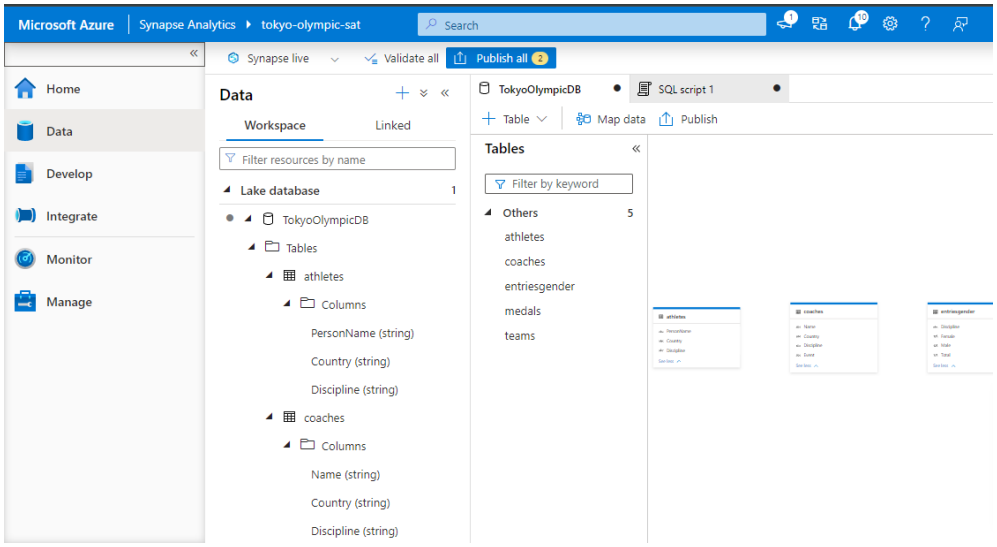
Data Loading & Analytics by using Azure Synapse Analytics Studio.

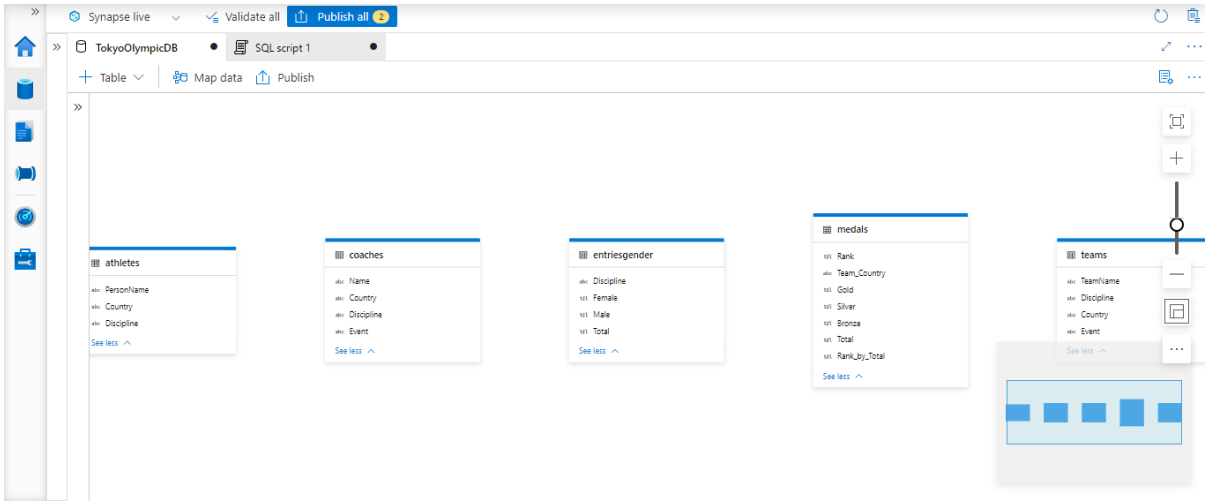


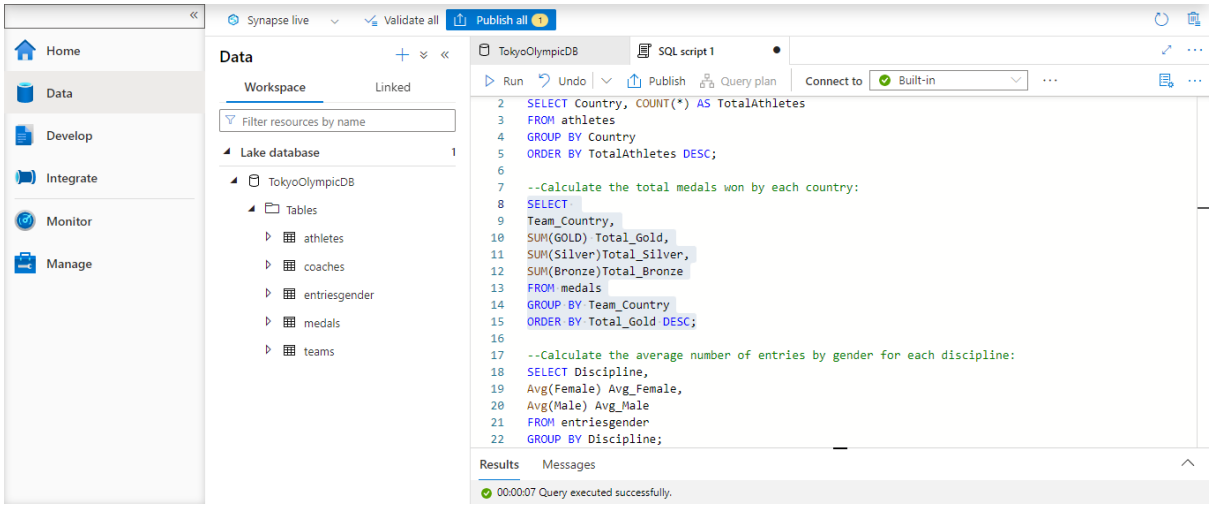
Setting Up and Using Azure Synapse Analytics

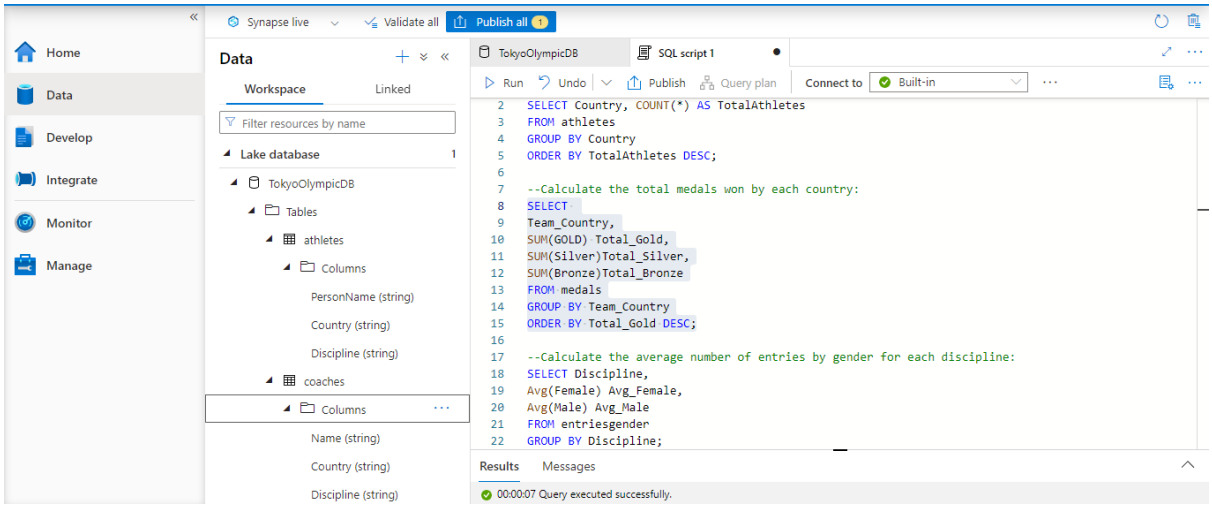
1. Creating a Synapse Analytics Workspace.
2. Within Workspace navigate to the "Data" section, choose "Lake Database" and create a Database "TokyoOlympicDB".
3. Creating Table from Data Lake from the Transformed Data folder within your ADLS Gen2 storage.

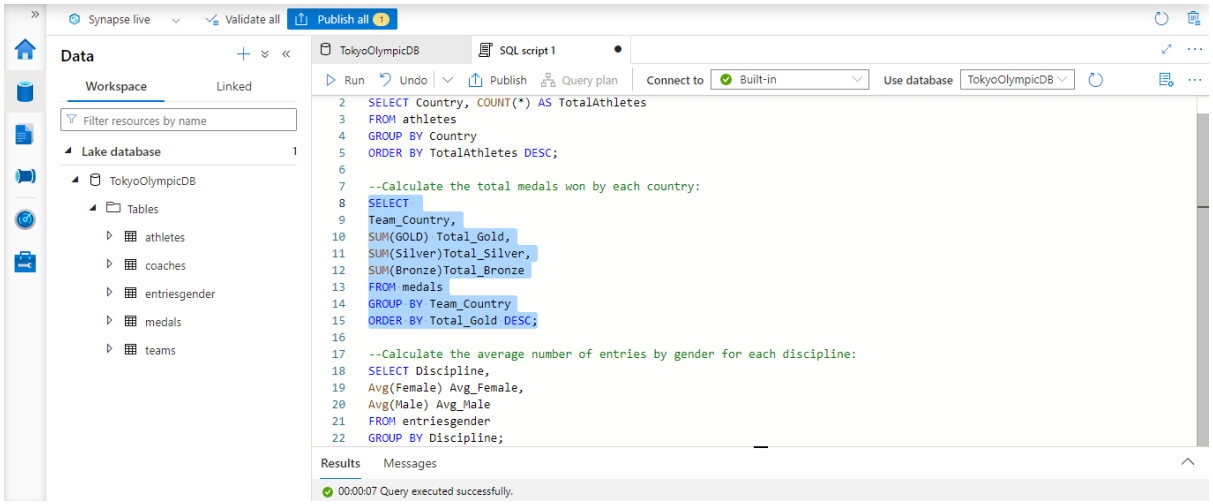





Synapse live
Validate all
Publish all 2
TokyoOlympicDB
SQL script 1
Table
Map data
Publish
athletes
PersonName
Country
Discipline
See less
coaches
Name
Country
Discipline
Event
See less
entriesgender
Discipline
Female
Male
Total
See less
medals
Rank
Team_Country
Gold
Silver
Bronze
Total
Rank_by_Total
See less
teams
TeamName
Discipline
Country
Event
See less

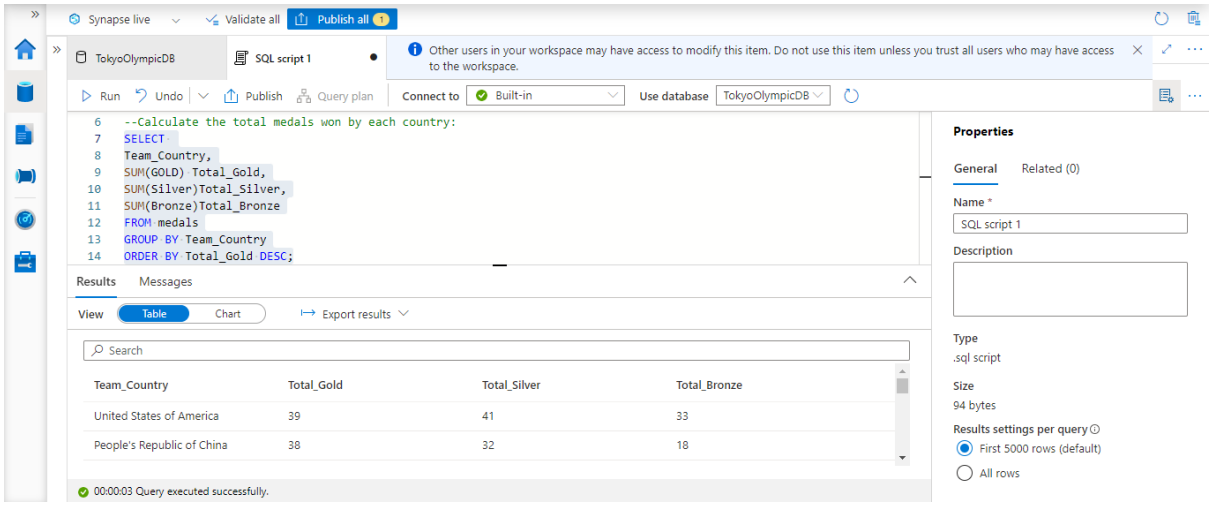

Synapse live
Validate all
Publish all 1
Home
Data
Develop
Integrate
Monitor
Manage
Data
Workspace
Linked
Filter resources by name
Lake database 1
TokyoOlympicDB
Tables
athletes
coaches
entriesgender
medals
teams
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
SELECT Country, COUNT(*) AS TotalAthletes
FROM athletes
GROUP BY Country
ORDER BY TotalAthletes DESC;
--Calculate the total medals won by each country:
SELECT
Team_Country,
SUM(GOLD) Total_Gold,
SUM(Silver)Total_Silver,
SUM(Bronze)Total_Bronze
FROM medals
GROUP BY Team_Country
ORDER BY Total_Gold DESC;
--Calculate the average number of entries by gender for each discipline:
SELECT Discipline,
Avg(Female) Avg_Female,
Avg(Male) Avg_Male
FROM entriesgender
GROUP BY Discipline;
Results
Messages
00:00:07 Query executed successfully.

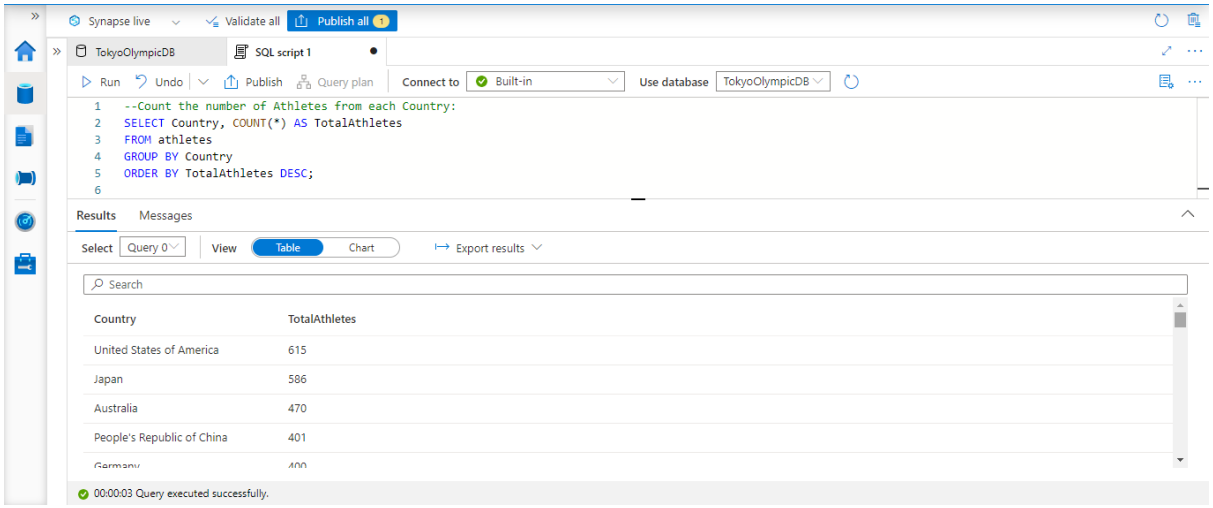

Synapse live
Validate all
Publish all 1
Home
Data
Develop
Integrate
Monitor
Manage
Data
Workspace
Linked
Filter resources by name
Lake database 1
TokyoOlympicDB
Tables
athletes
Columns
PersonName (string)
Country (string)
Discipline (string)
coaches
Columns
Name (string)
Country (string)
Discipline (string)
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
SELECT Country, COUNT(*) AS TotalAthletes
FROM athletes
GROUP BY Country
ORDER BY TotalAthletes DESC;
--Calculate the total medals won by each country:
SELECT
Team_Country,
SUM(GOLD) Total_Gold,
SUM(Silver)Total_Silver,
SUM(Bronze)Total_Bronze
FROM medals
GROUP BY Team_Country
ORDER BY Total_Gold DESC;
--Calculate the average number of entries by gender for each discipline:
SELECT Discipline,
Avg(Female) Avg_Female,
Avg(Male) Avg_Male
FROM entriesgender
GROUP BY Discipline;
Results
Messages
00:00:07 Query executed successfully.

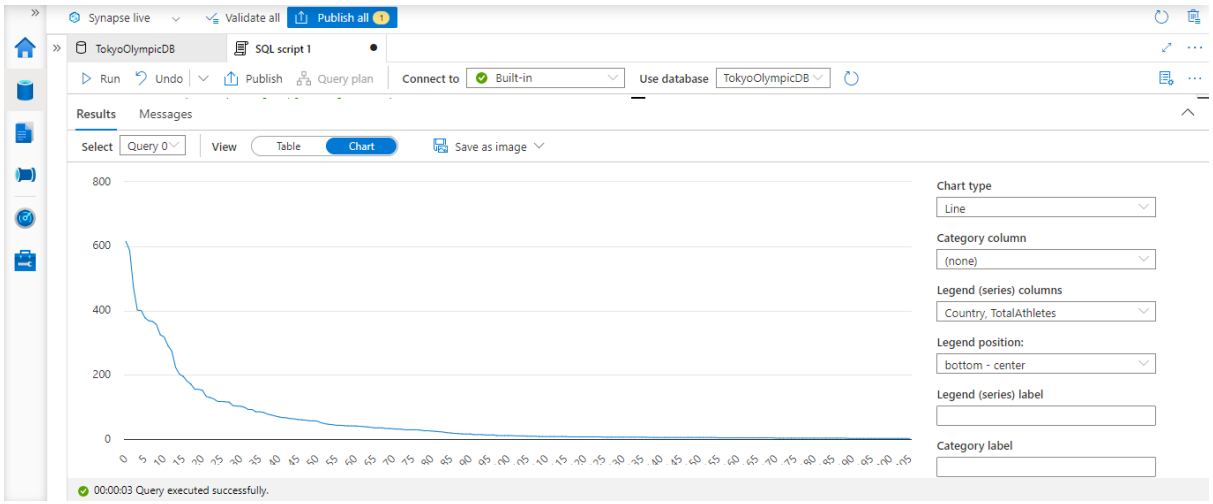


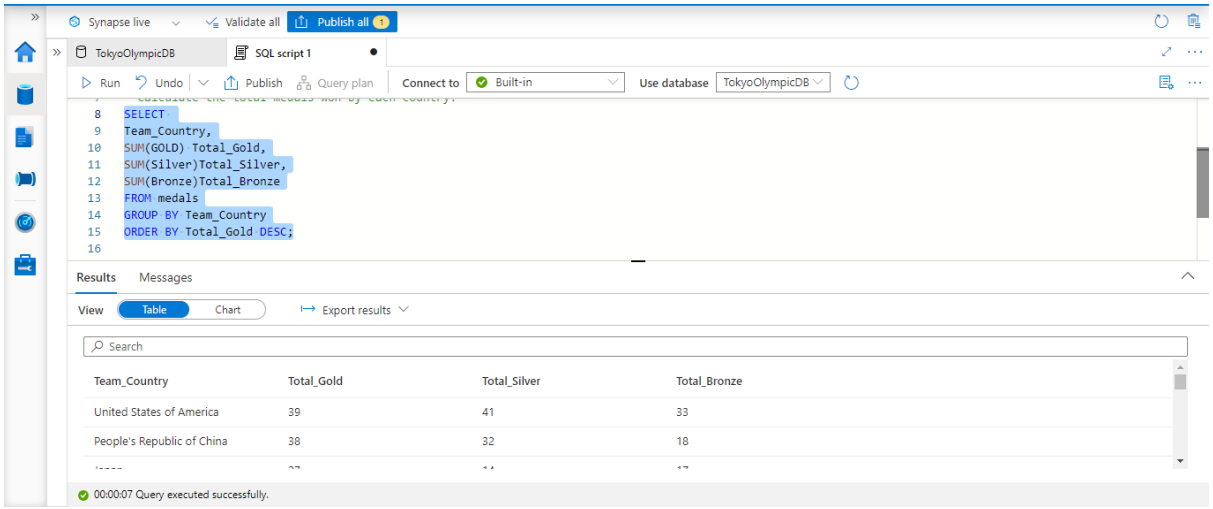
Calculate the total medals won by each country,

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



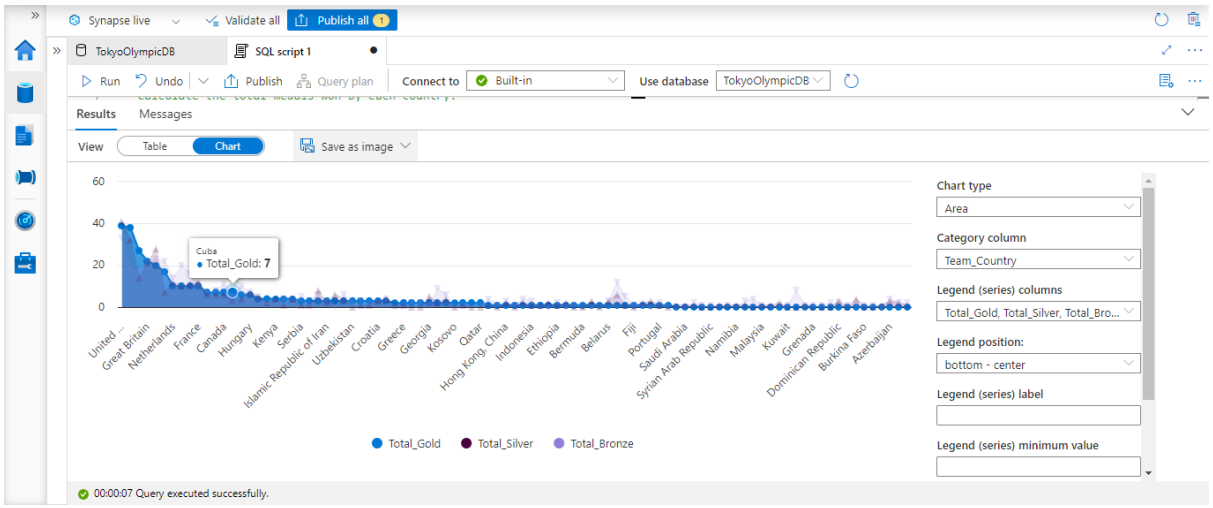

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
--Count the number of Athletes from each Country:
SELECT Country, COUNT(*) AS TotalAthletes
FROM athletes
GROUP BY Country
ORDER BY TotalAthletes DESC;
Results
Messages
Select
Query 0
View
Table
Chart
Export results
Search
Country
TotalAthletes
United States of America
615
Japan
586
Australia
470
People's Republic of China
401
00:00:03 Query executed successfully.

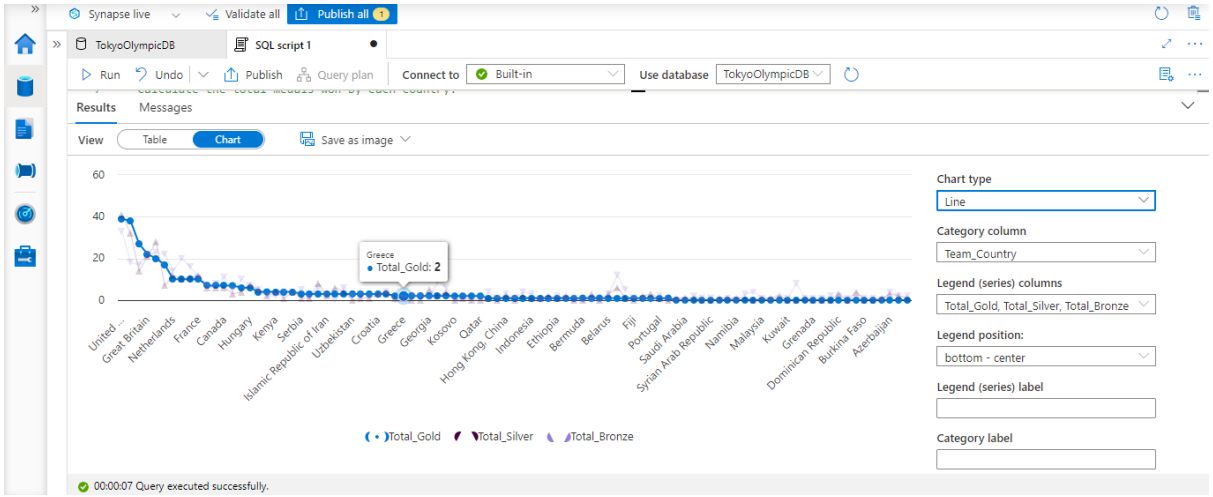

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
Select
Query 0
View
Table
Chart
Save as image
800
600
400
200
0
Chart type
Line
Category column
(none)
Legend (series) columns
Country, TotalAthletes
Legend position:
bottom - center
Legend (series) label
Category label
00:00:03 Query executed successfully.


Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
SELECT
Team_Country,
SUM(GOLD) Total_Gold,
SUM(Silver)Total_Silver,
SUM(Bronze)Total_Bronze
FROM medals
GROUP BY Team_Country
ORDER BY Total_Gold DESC;
Results
Messages
View
Table
Chart
Export results
Search
Team_Country
Total_Gold
Total_Silver
Total_Bronze
United States of America
39
41
33
People's Republic of China
38
32
18
00:00:07 Query executed successfully.

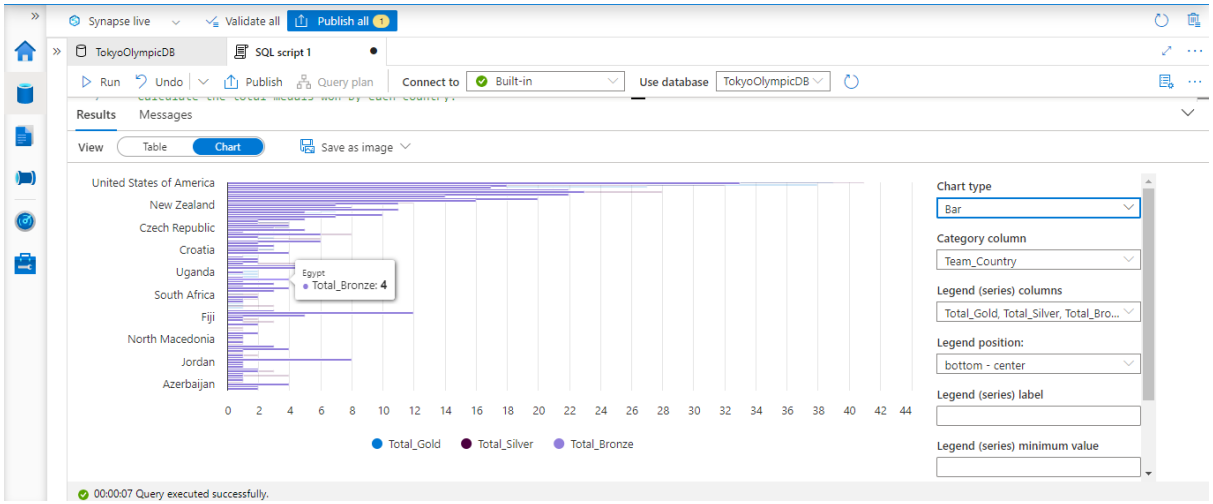
Performing Data Analysis on the Data

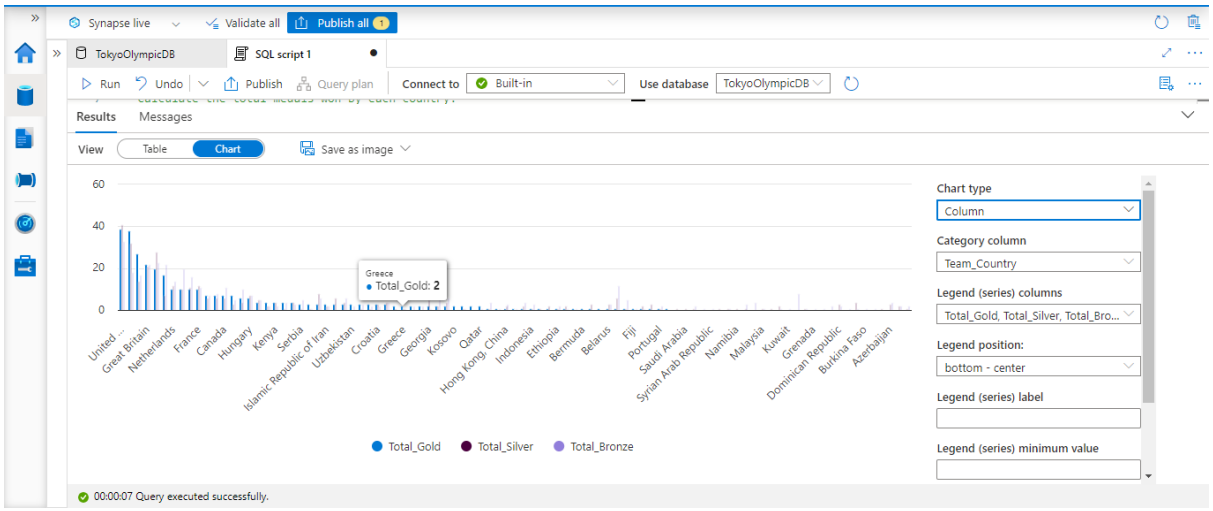
Create SQL script to Perform Exploratory data analysis using SQL. You can also use PowerBI to generate your analysis reports.

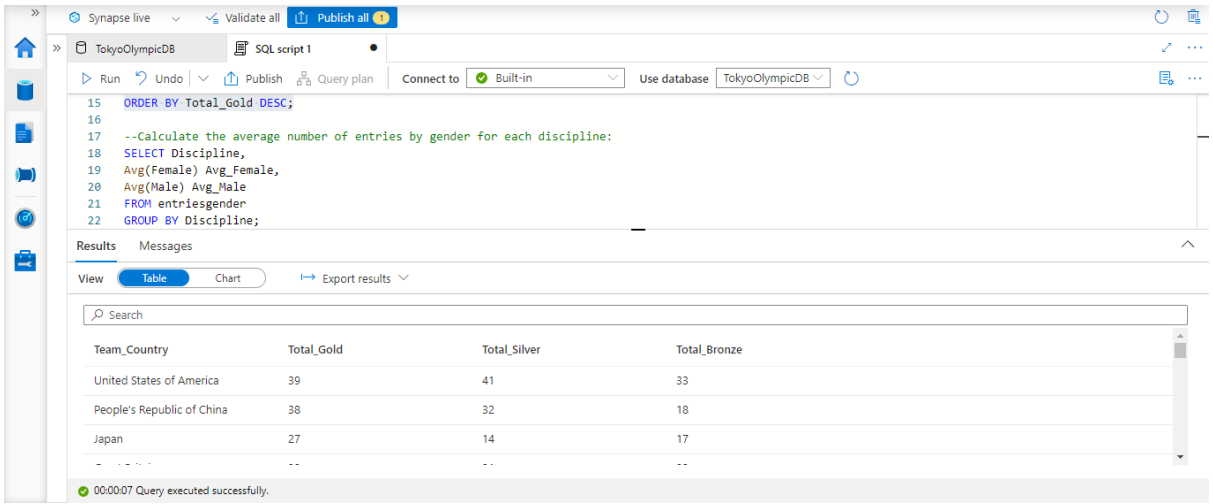


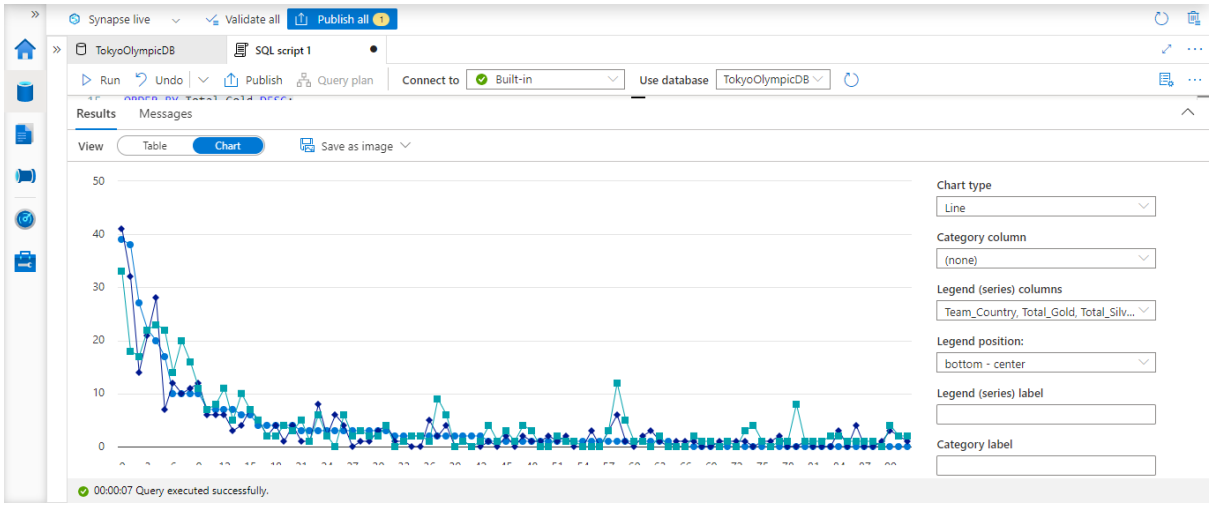


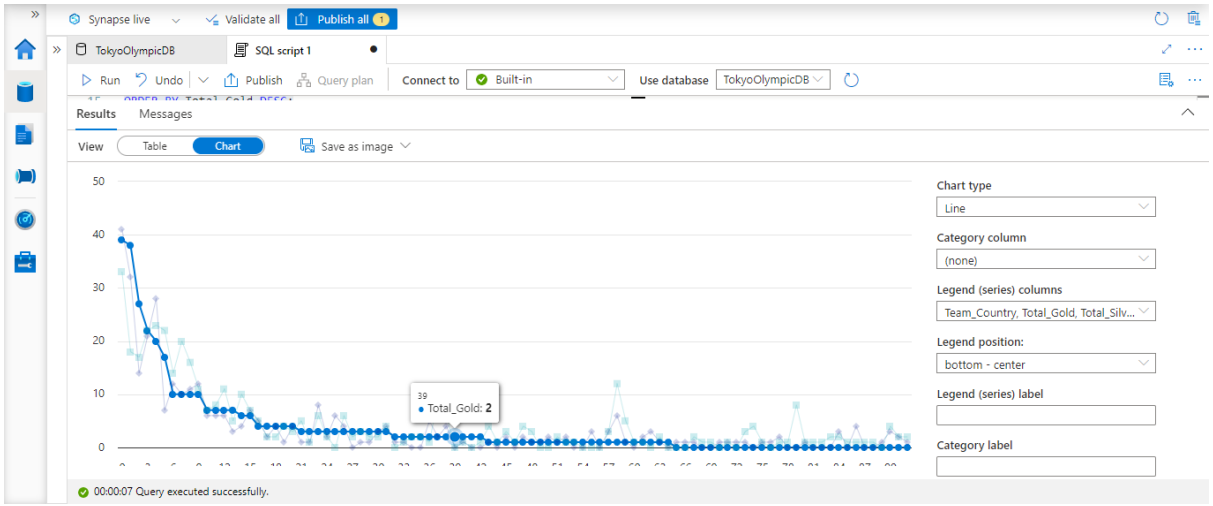
Visualize the data according to the SQL query in Azure Synapse Analytics.

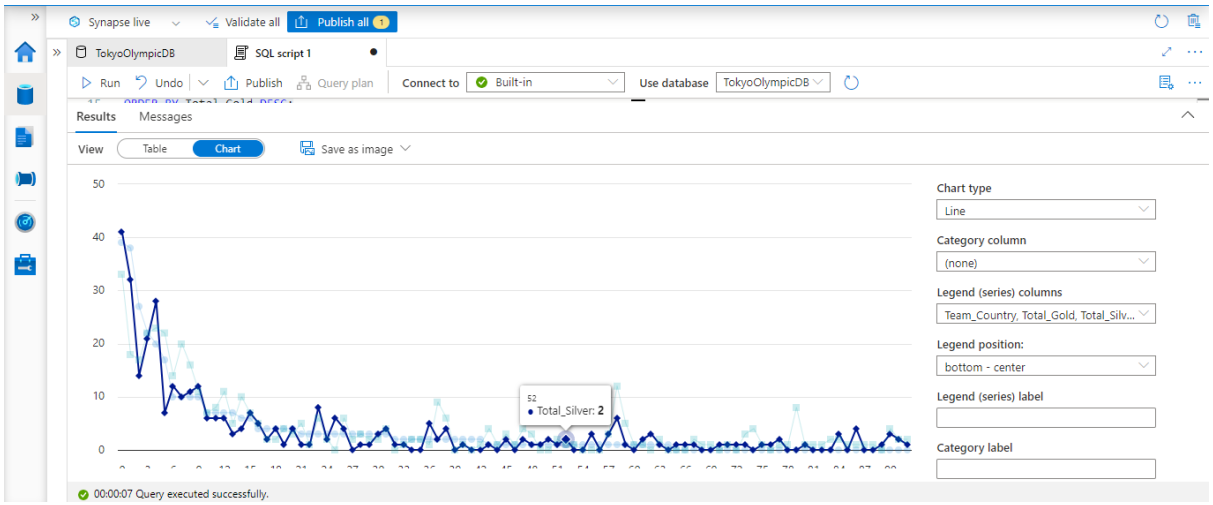

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
United States of America
New Zealand
Czech Republic
Croatia
Uganda
South Africa
Fiji
North Macedonia
Jordan
Azerbaijan
Egypt
Total_Bronze: 4
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44
Total_Gold
Total_Silver
Total_Bronze
Chart type
Bar
Category column
Team_Country
Legend (series) columns
Total_Gold, Total_Silver, Total_Bro...
Legend position:
bottom - center
Legend (series) label
Legend (series) minimum value
00:00:07 Query executed successfully.

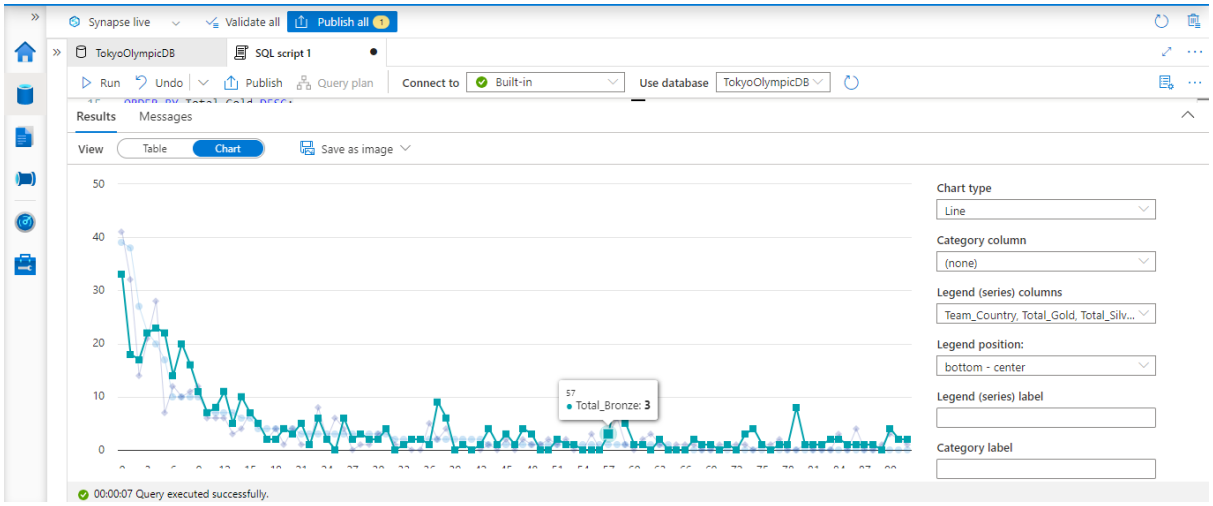

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
60
40
20
0
Greece
Total_Gold: 2
United ...
Great Britain
Netherlands
France
Canada
Hungary
Kenya
Serbia
Islamic Republic of Iran
Uzbekistan
Croatia
Greece
Georgia
Kosovo
Qatar
Hong Kong, China
Indonesia
Ethiopia
Bermuda
Belarus
Fiji
Portugal
Saudi Arabia
Syrian Arab Republic
Namibia
Malaysia
Kuwait
Grenada
Dominican Republic
Burkina Faso
Azerbaijan
Total_Gold
Total_Silver
Total_Bronze
Chart type
Column
Category column
Team_Country
Legend (series) columns
Total_Gold, Total_Silver, Total_Bro...
Legend position:
bottom - center
Legend (series) label
Legend (series) minimum value
00:00:07 Query executed successfully.

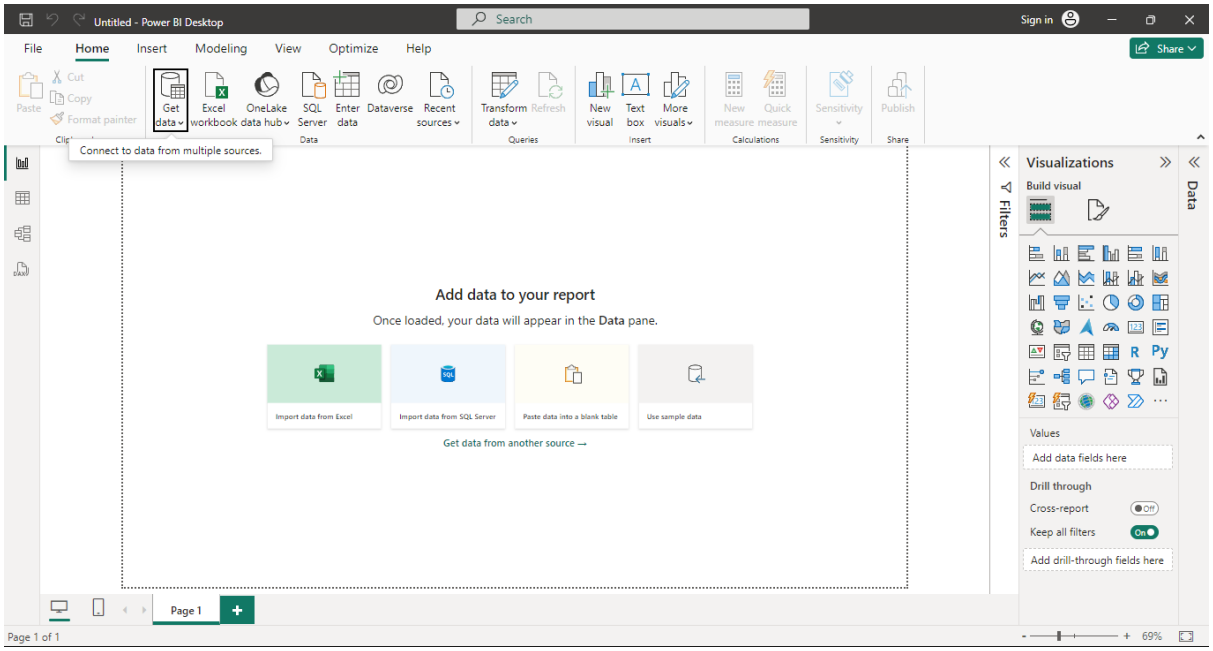

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
ORDER BY Total_Gold DESC;
--Calculate the average number of entries by gender for each discipline:
SELECT Discipline,
Avg(Female) Avg_Female,
Avg(Male) Avg_Male
FROM entriesgender
GROUP BY Discipline;
Results
Messages
View
Table
Chart
Export results
Search
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.

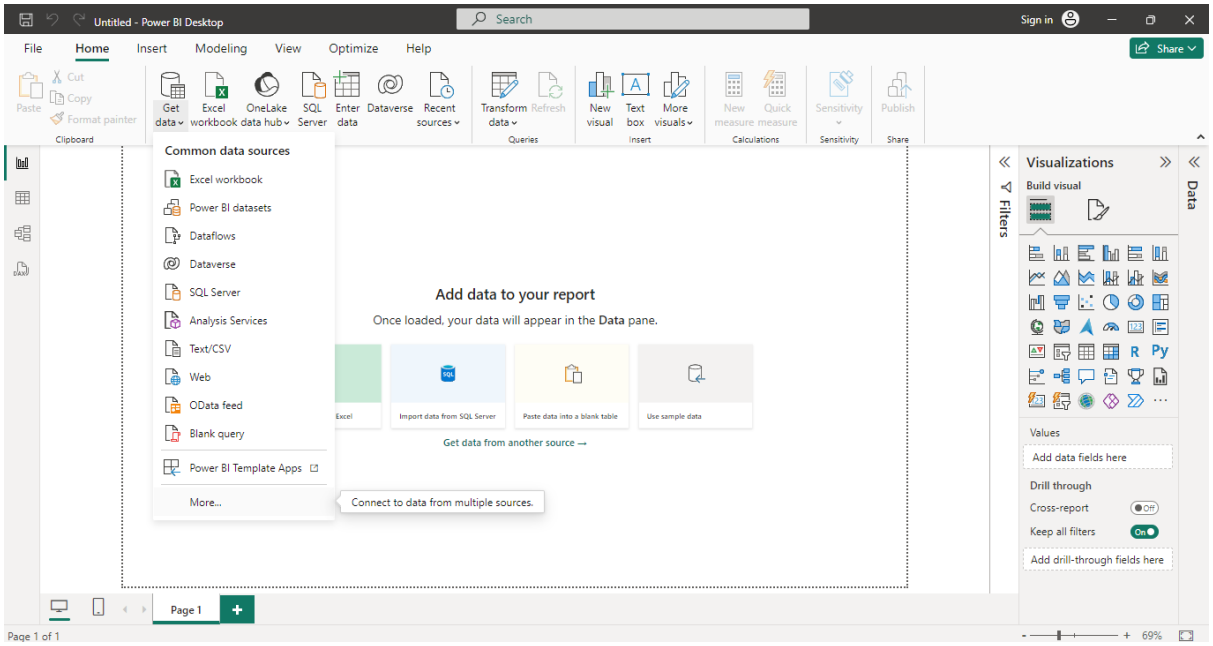

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
50
40
30
20
10
0
Chart type
Line
Category column
(none)
Legend (series) columns
Team_Country, Total_Gold, Total_Silv...
Legend position:
bottom - center
Legend (series) label
Category label
00:00:07 Query executed successfully.

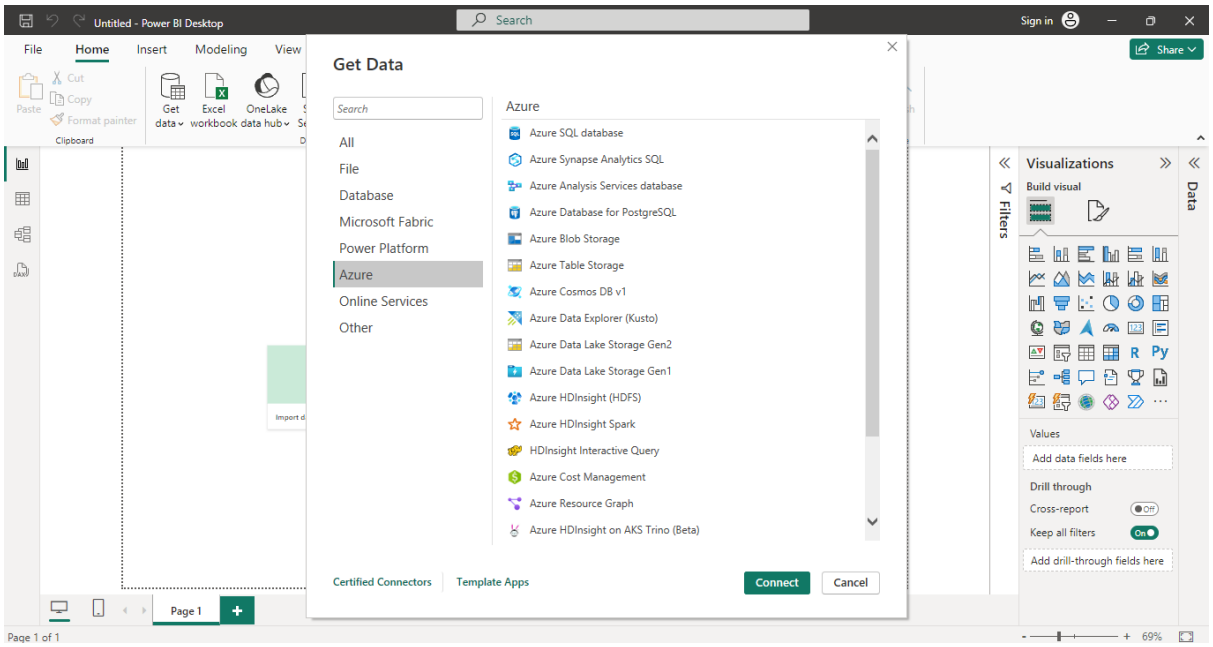

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
50
40
30
20
10
0
39
Total_Gold: 2
Chart type
Line
Category column
(none)
Legend (series) columns
Team_Country, Total_Gold, Total_Silv...
Legend position:
bottom - center
Legend (series) label
Category label
00:00:07 Query executed successfully.

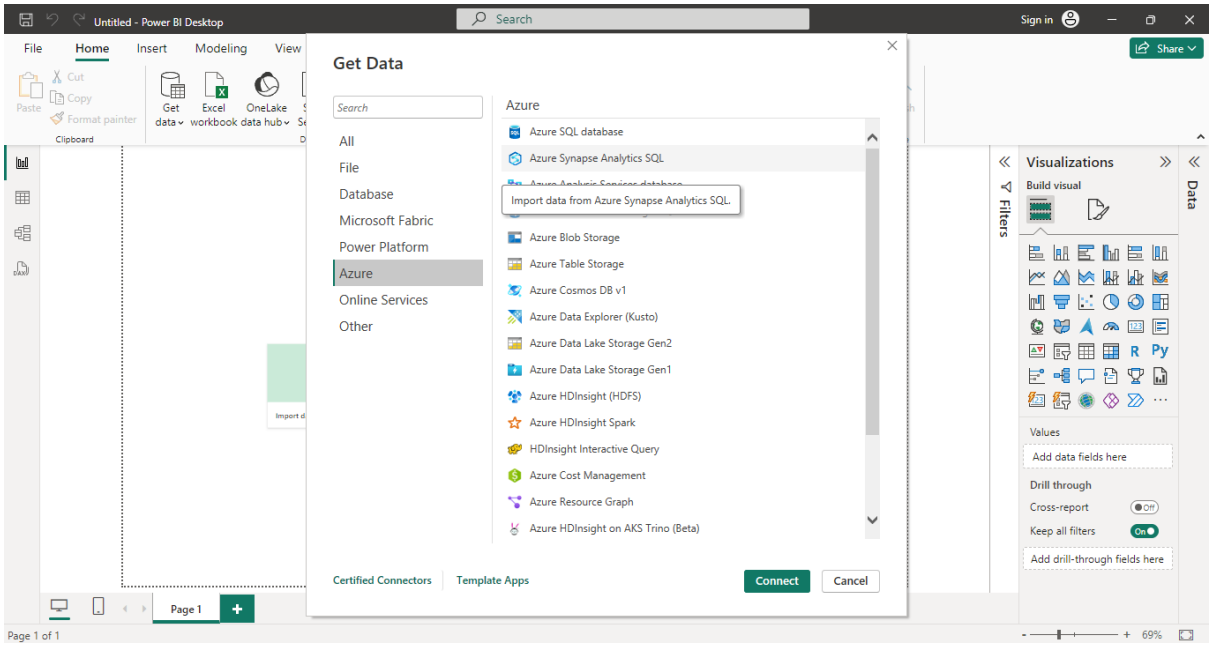

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
50
40
30
20
10
0
52
Total_Silver: 2
Chart type
Line
Category column
(none)
Legend (series) columns
Team_Country, Total_Gold, Total_Silv...
Legend position:
bottom - center
Legend (series) label
Category label
00:00:07 Query executed successfully.

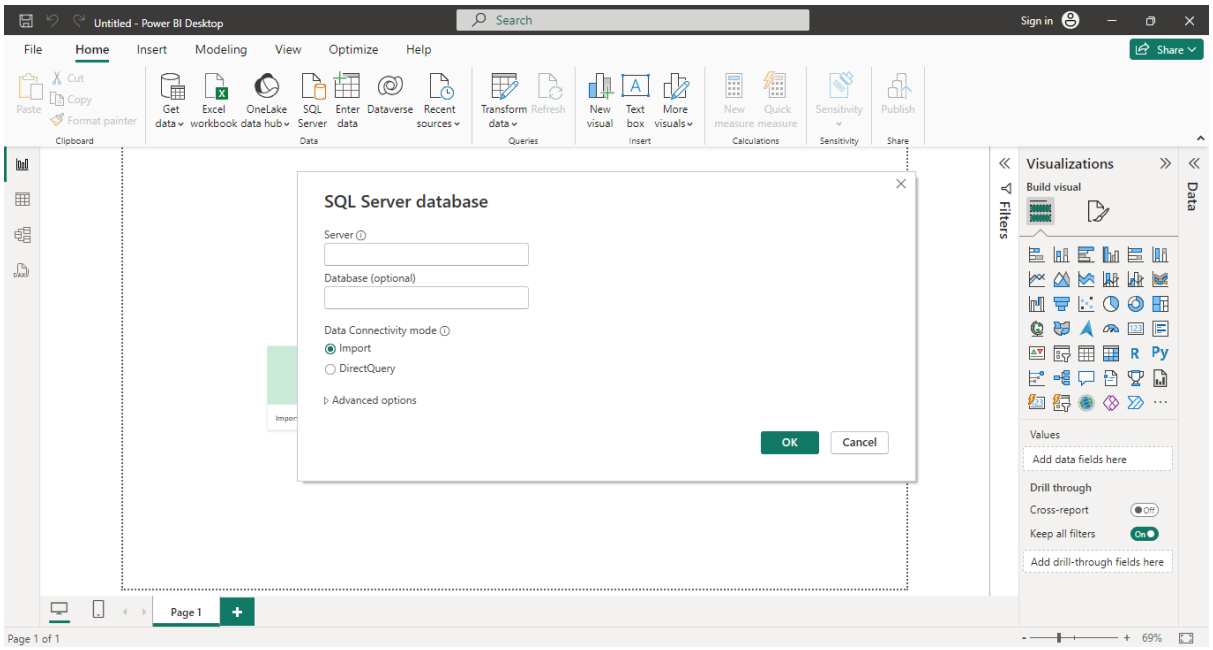

Synapse live
Validate all
Publish all 1
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
Use database
TokyoOlympicDB
Results
Messages
View
Table
Chart
Save as image
50
40
30
20
10
0
57
Total_Bronze: 3
Chart type
Line
Category column
(none)
Legend (series) columns
Team_Country, Total_Gold, Total_Silv...
Legend position:
bottom - center
Legend (series) label
Category label
00:00:07 Query executed successfully.

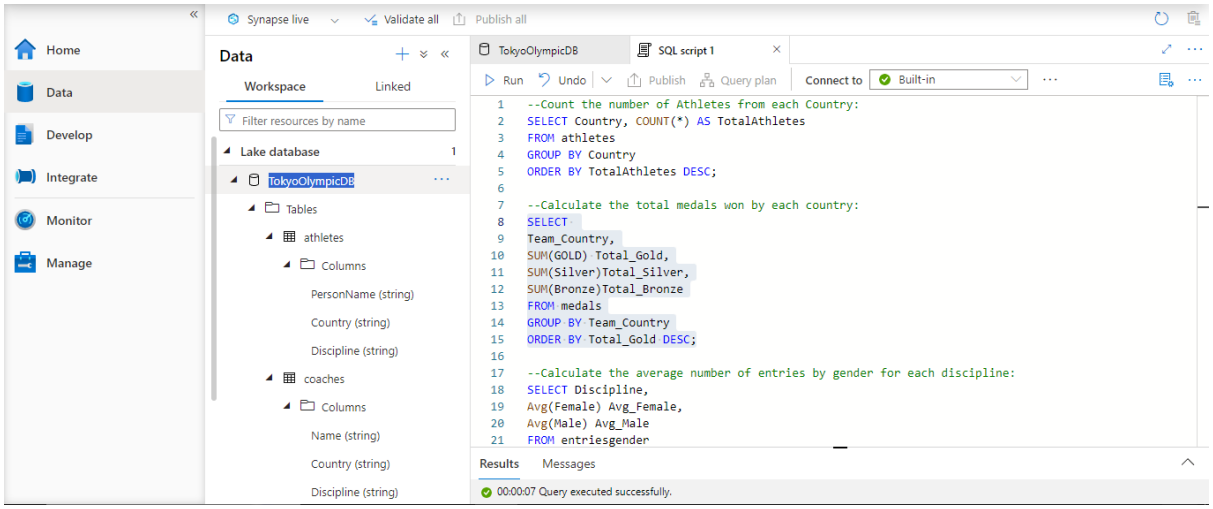

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Paste
Cut
Copy
Format painter
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New measure
Quick measure
Sensitivity
Publish
Connect to data from multiple sources.
Add data to your report
Once loaded, your data will appear in the Data pane.
Import data from Excel
Import data from SQL Server
Paste data into a blank table
Use sample data
Get data from another source →
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

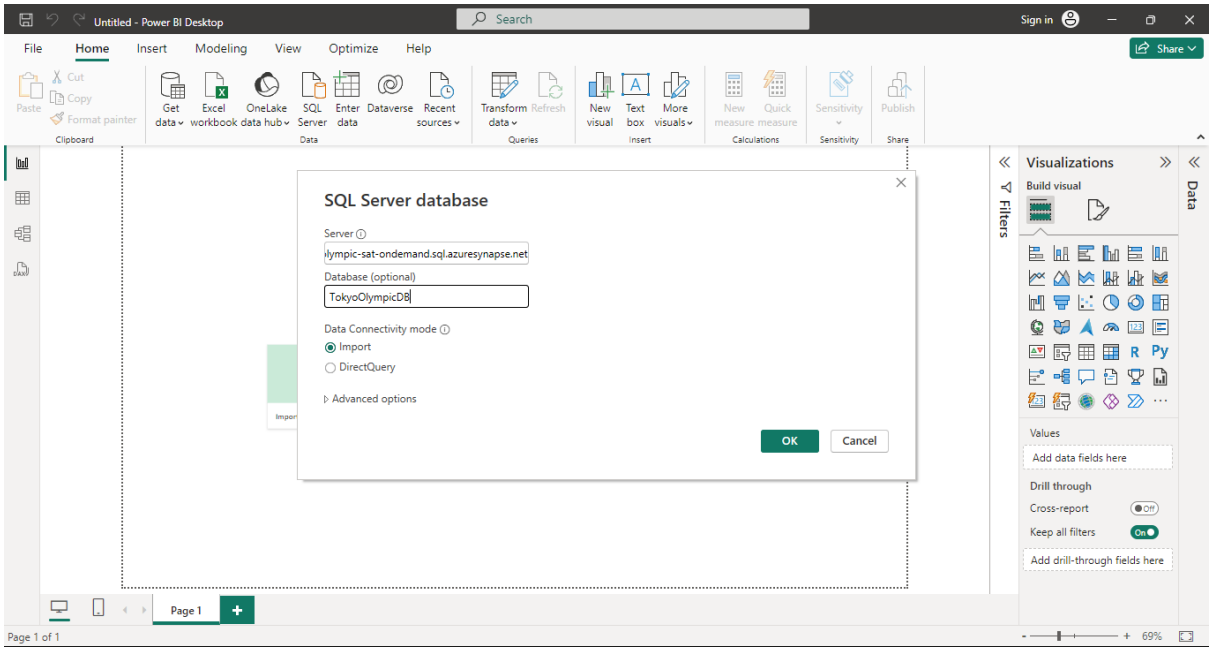

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New measure
Quick measure
Sensitivity
Publish
Common data sources
Excel workbook
Power BI datasets
Dataflows
Dataverse
SQL Server
Analysis Services
Text/CSV
Web
OData feed
Blank query
Power BI Template Apps
More...
Connect to data from multiple sources.
Add data to your report
Once loaded, your data will appear in the Data pane.
Import data from SQL Server
Paste data into a blank table
Use sample data
Get data from another source →
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

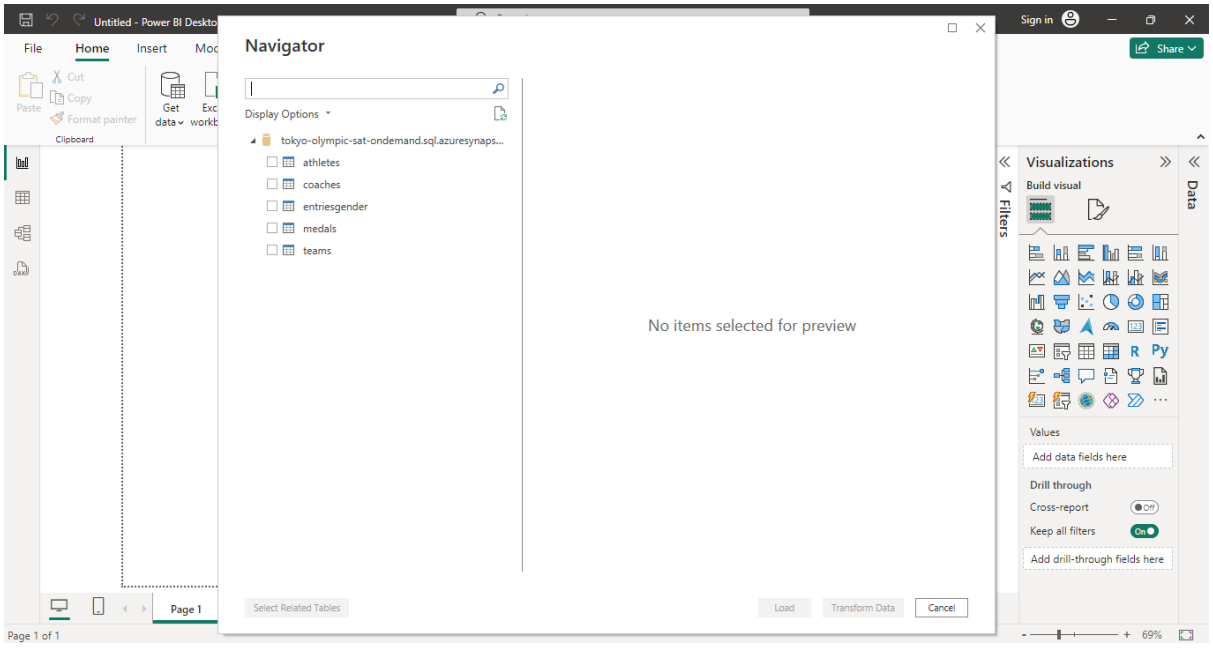

Untitled - Power BI Desktop
Search
Sign in
Share
Get Data
Search
All
File
Database
Microsoft Fabric
Power Platform
Azure
Online Services
Other
Azure
Azure SQL database
Azure Synapse Analytics SQL
Azure Analysis Services database
Azure Database for PostgreSQL
Azure Blob Storage
Azure Table Storage
Azure Cosmos DB v1
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen2
Azure Data Lake Storage Gen1
Azure HDInsight (HDFS)
Azure HDInsight Spark
HDInsight Interactive Query
Azure Cost Management
Azure Resource Graph
Azure HDInsight on AKS Trino (Beta)
Certified Connectors
Template Apps
Connect
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

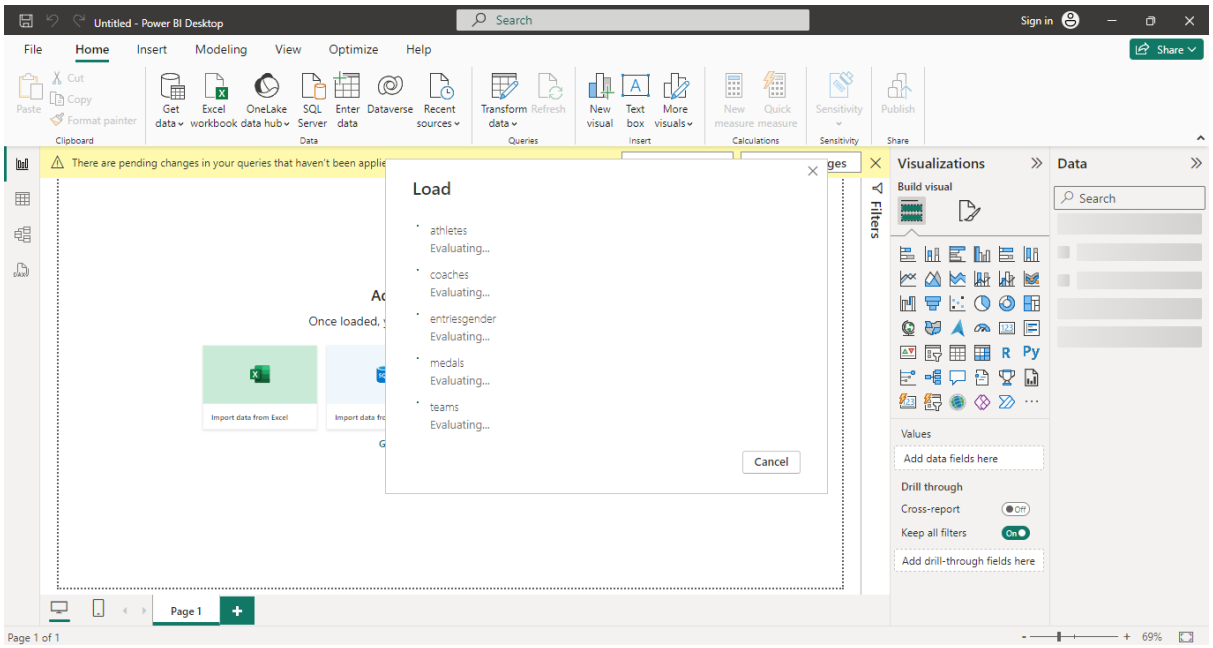

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Share
Get Data
Search
All
File
Database
Microsoft Fabric
Power Platform
Azure
Online Services
Other
Azure
Azure SQL database
Azure Synapse Analytics SQL
Import data from Azure Synapse Analytics SQL.
Azure Blob Storage
Azure Table Storage
Azure Cosmos DB v1
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen2
Azure Data Lake Storage Gen1
Azure HDInsight (HDFS)
Azure HDInsight Spark
HDInsight Interactive Query
Azure Cost Management
Azure Resource Graph
Azure HDInsight on AKS Trino (Beta)
Certified Connectors
Template Apps
Connect
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

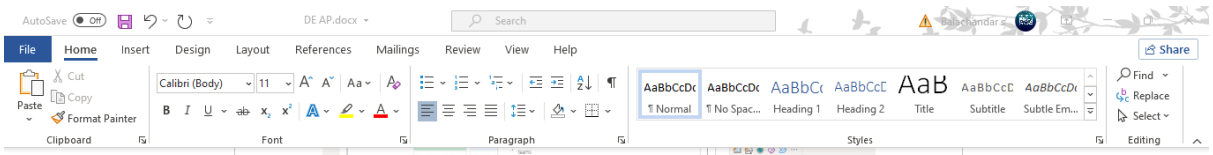

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
SQL Server database
Server
Database (optional)
Data Connectivity mode
Import
DirectQuery
Advanced options
OK
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

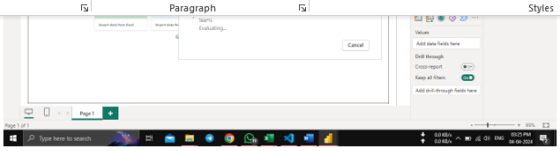

Synapse live
Validate all
Publish all
Home
Data
Develop
Integrate
Monitor
Manage
Data
Workspace
Linked
Filter resources by name
Lake database
TokyoOlympicDB
Tables
athletes
Columns
PersonName (string)
Country (string)
Discipline (string)
coaches
Columns
Name (string)
Country (string)
Discipline (string)
TokyoOlympicDB
SQL script 1
Run
Undo
Publish
Query plan
Connect to
Built-in
--Count the number of Athletes from each Country:
SELECT Country, COUNT(*) AS TotalAthletes
FROM athletes
GROUP BY Country
ORDER BY TotalAthletes DESC;
--Calculate the total medals won by each country:
SELECT
Team_Country,
SUM(GOLD) Total_Gold,
SUM(Silver)Total_Silver,
SUM(Bronze)Total_Bronze
FROM medals
GROUP BY Team_Country
ORDER BY Total_Gold DESC;
--Calculate the average number of entries by gender for each discipline:
SELECT Discipline,
Avg(Female) Avg_Female,
Avg(Male) Avg_Male
FROM entriesgender
Results
Messages
00:00:07 Query executed successfully.

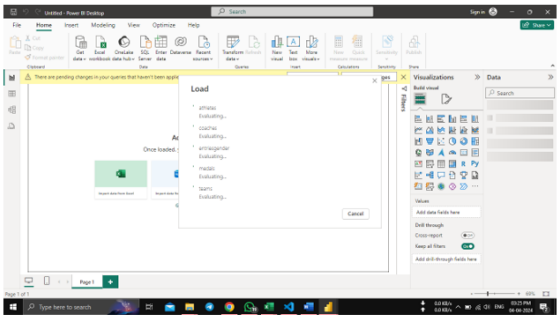

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
SQL Server database
Server
lympic-sat-ondemand.sql.azuresynapse.net
Database (optional)
TokyoOlympicDB
Data Connectivity mode
Import
DirectQuery
Advanced options
OK
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Filters
Data
Page 1
Page 1 of 1
69%

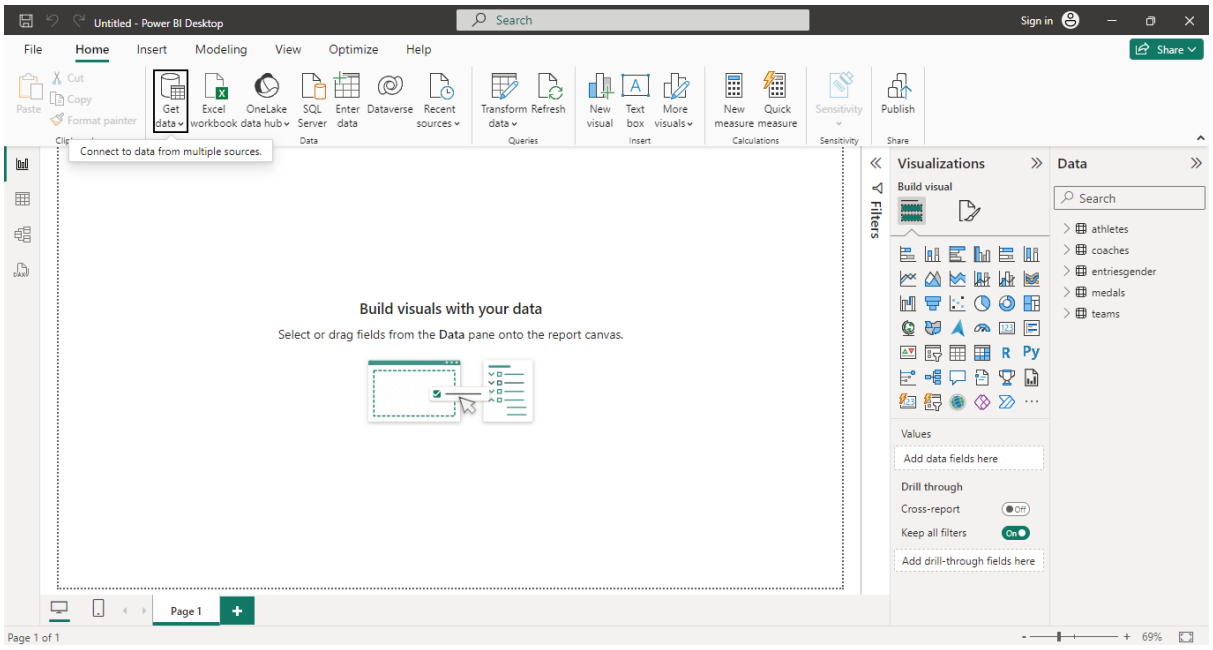

Untitled - Power BI Desktop
Navigator
Display Options
tokyo-olympic-sat-ondemand.sql.azuresynaps...
athletes
coaches
entriesgender
medals
teams
No items selected for preview
Select Related Tables
Load
Transform Data
Cancel
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1

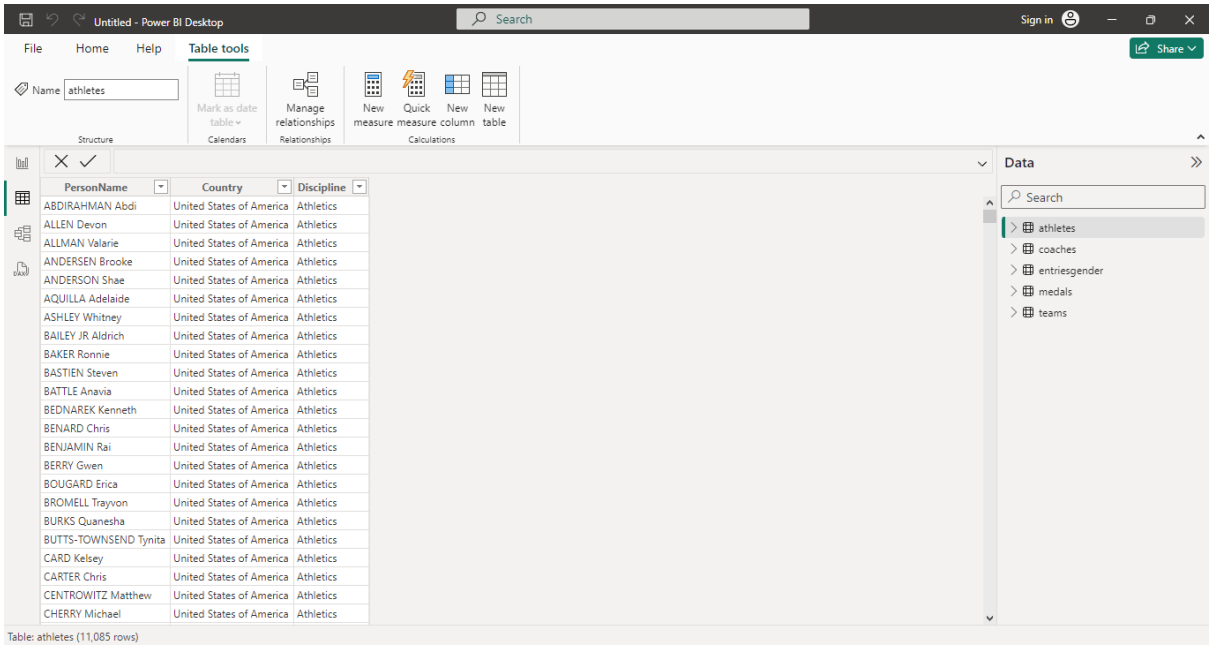

Untitled - Power BI Desktop
Search
Load
athletes
Evaluating...
coaches
Evaluating...
entriesgender
Evaluating...
medals
Evaluating...
teams
Evaluating...
Cancel
There are pending changes in your queries that haven't been applie
Visualizations
Data
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 1 of 1



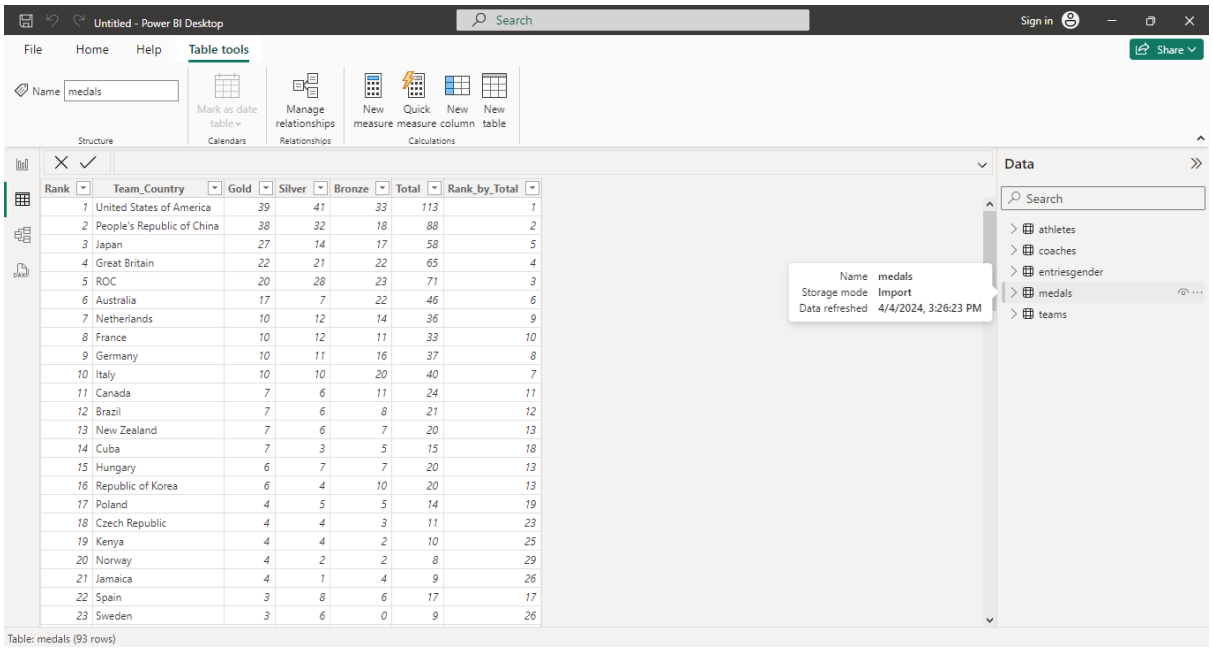




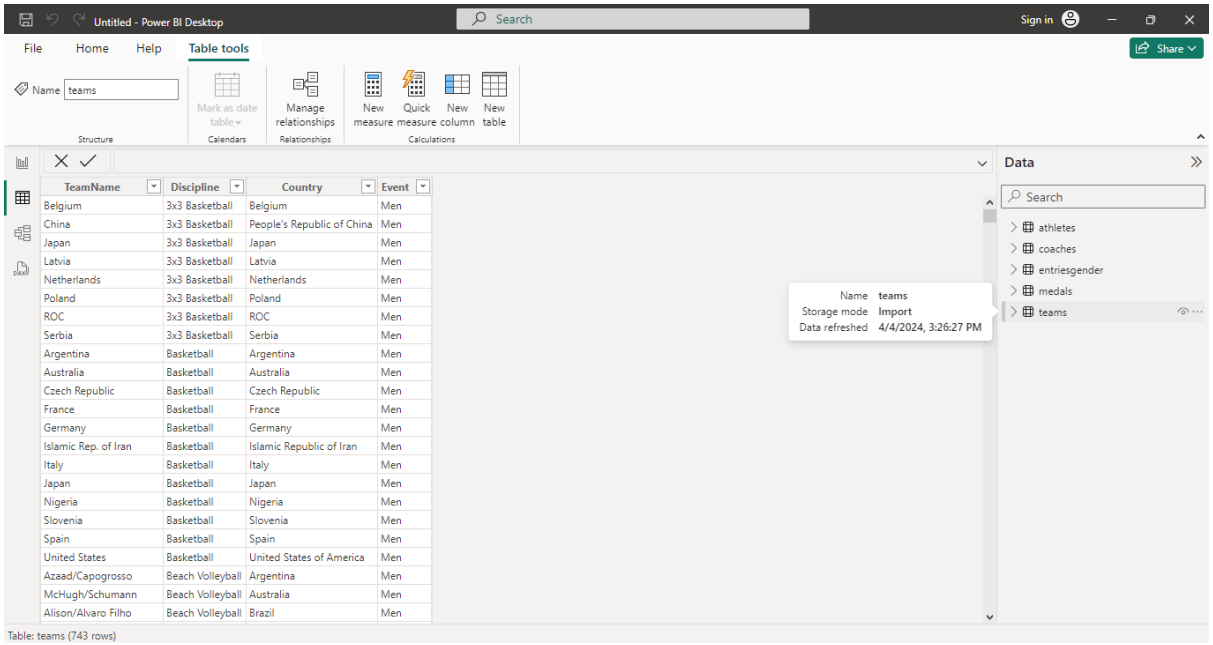

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New measure
Quick measure
Sensitivity
Publish
Connect to data from multiple sources.
Build visuals with your data
Select or drag fields from the Data pane onto the report canvas.
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Data
athletes
coaches
entriesgender
medals
teams
Filters
Page 1
Page 1 of 1
69%



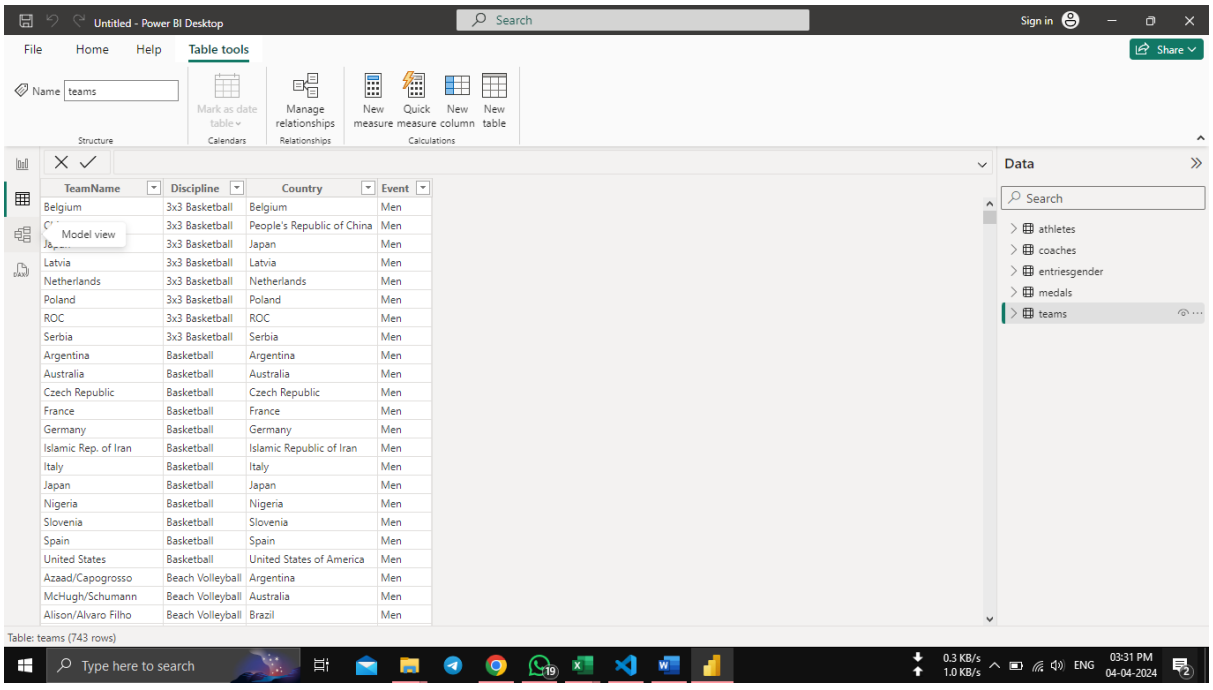
PersonName	Country	Discipline
ABDIRAHMAN Abdi	United States of America	Athletics
ALLEN Devon	United States of America	Athletics
ALLMAN Valarie	United States of America	Athletics
ANDERSEN Brooke	United States of America	Athletics
ANDERSON Shae	United States of America	Athletics
AQUILLA Adelaide	United 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



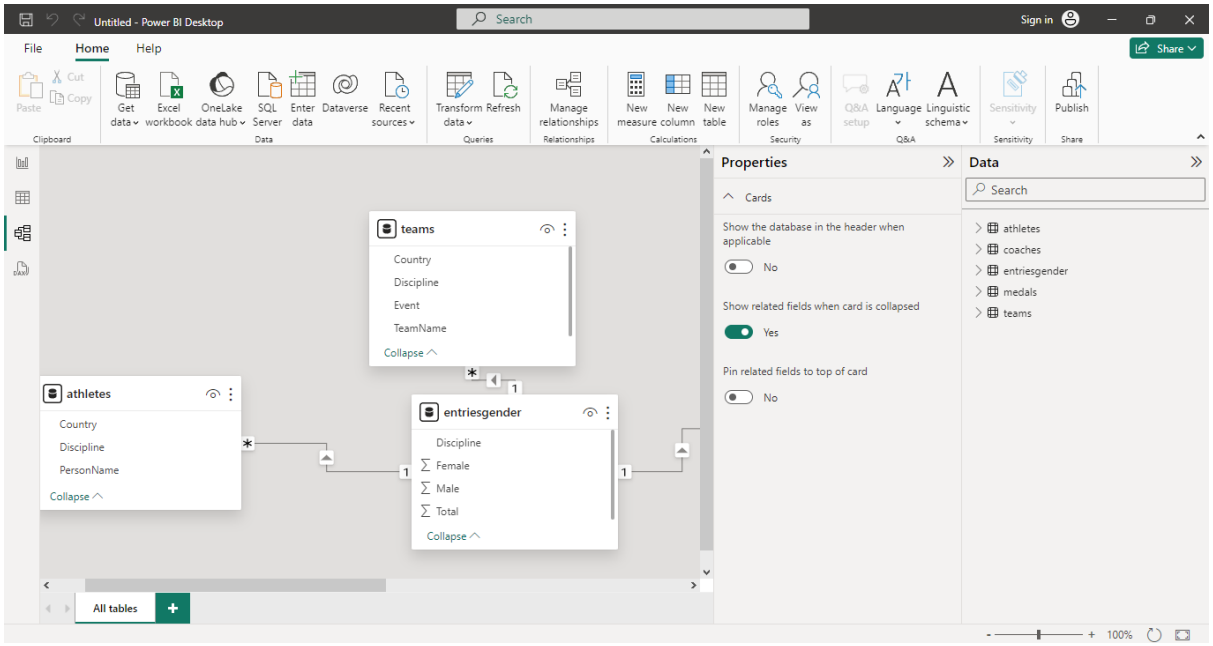
Rank	Team_Country	Gold	Silver	Bronze	Total	Rank_by_Total
1	United States of America	39	41	33	113	1
2	People's Republic of China	38	32	18	88	2
3	Japan	27	14	17	58	5
4	Great Britain	22	21	22	65	4
5	ROC	20	28	23	71	3
6	Australia	17	7	22	46	6
7	Netherlands	10	12	14	36	9
8	France	10	12	11	33	10
9	Germany	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



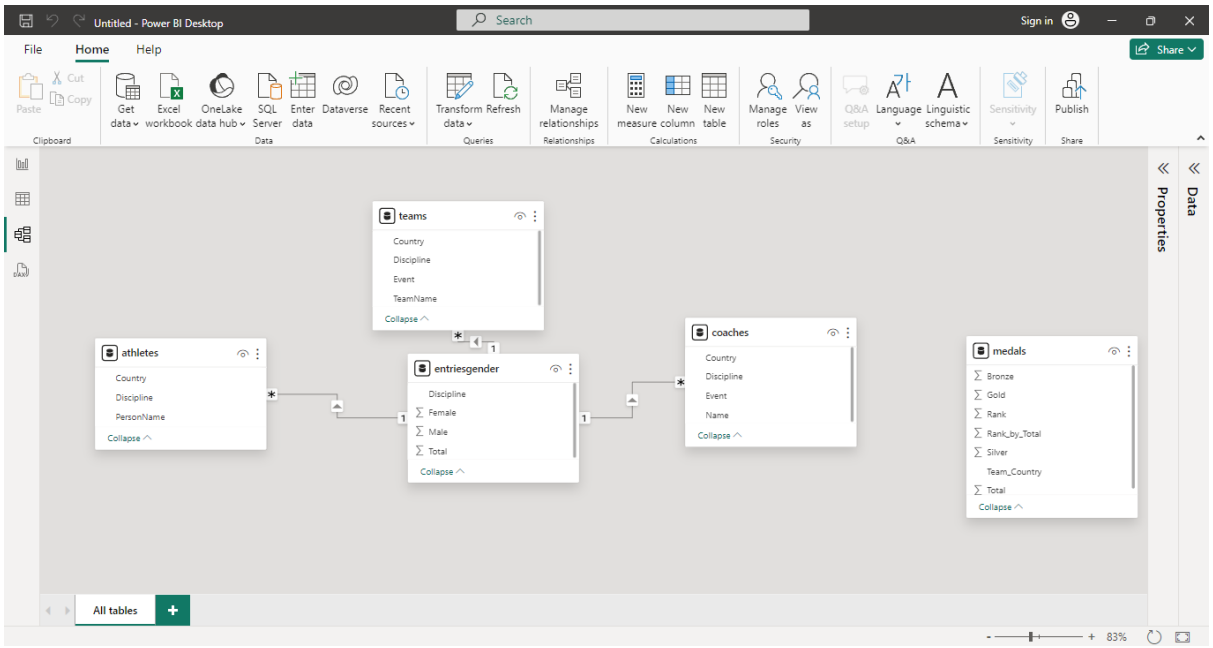
TeamName	Discipline	Country	Event
Belgium	3x3 Basketball	Belgium	Men
China	3x3 Basketball	People's Republic of China	Men
Japan	3x3 Basketball	Japan	Men
Latvia	3x3 Basketball	Latvia	Men
Netherlands	3x3 Basketball	Netherlands	Men
Poland	3x3 Basketball	Poland	Men
ROC	3x3 Basketball	ROC	Men
Serbia	3x3 Basketball	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

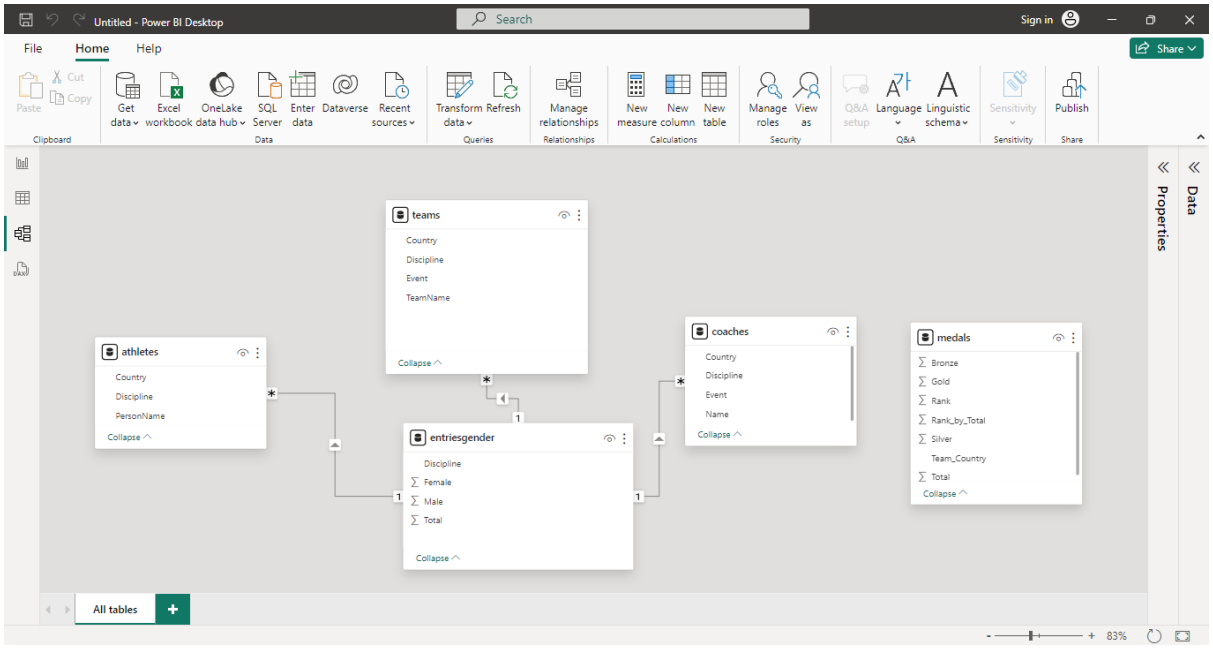


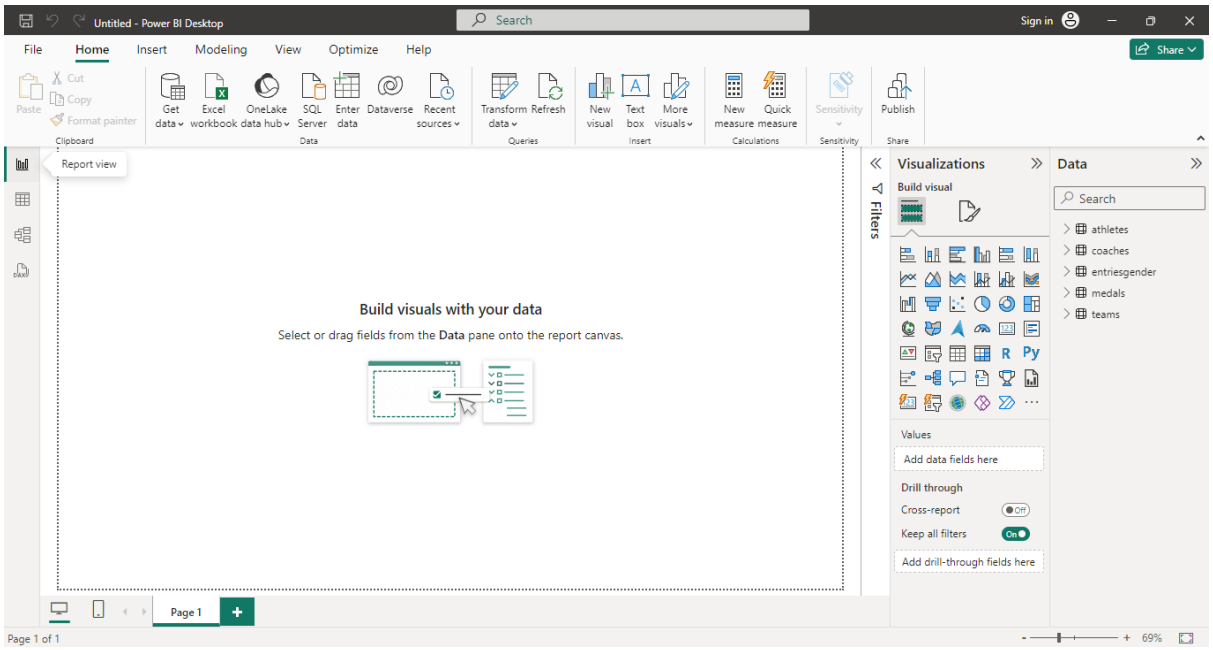
TeamName	Discipline	Country	Event
Belgium	3x3 Basketball	Belgium	Men
[illegible]	3x3 Basketball	People's Republic of China	Men
[illegible]	3x3 Basketball	Japan	Men
Latvia	3x3 Basketball	Latvia	Men
Netherlands	3x3 Basketball	Netherlands	Men
Poland	3x3 Basketball	Poland	Men
ROC	3x3 Basketball	ROC	Men
Serbia	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

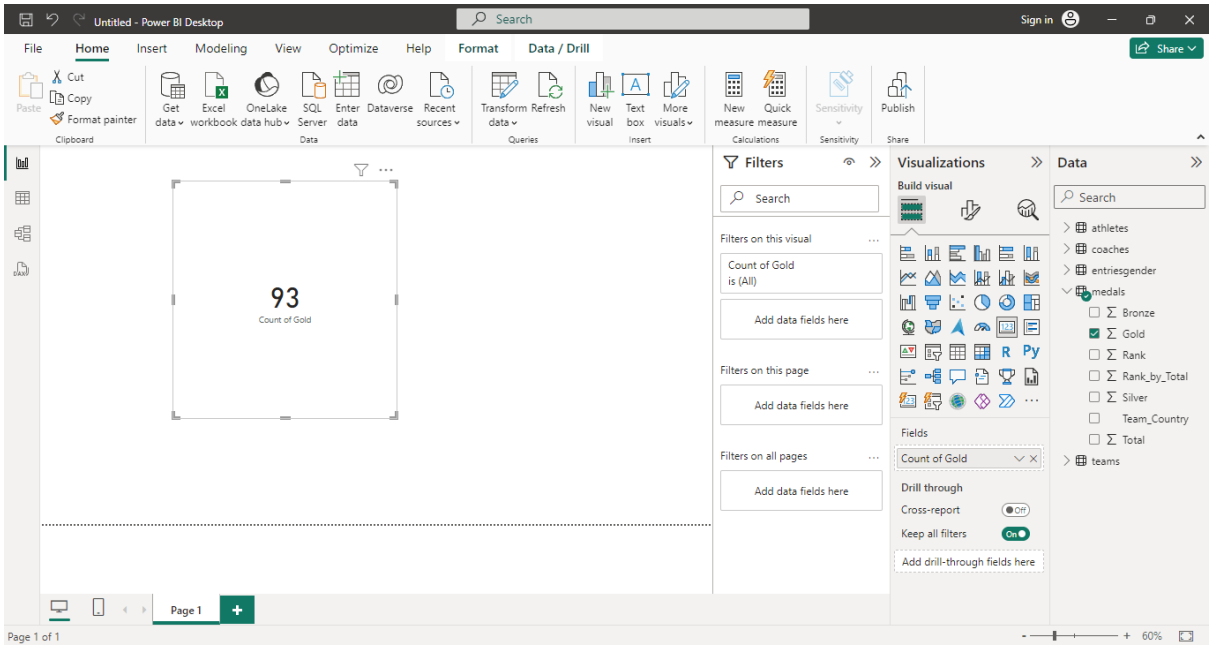


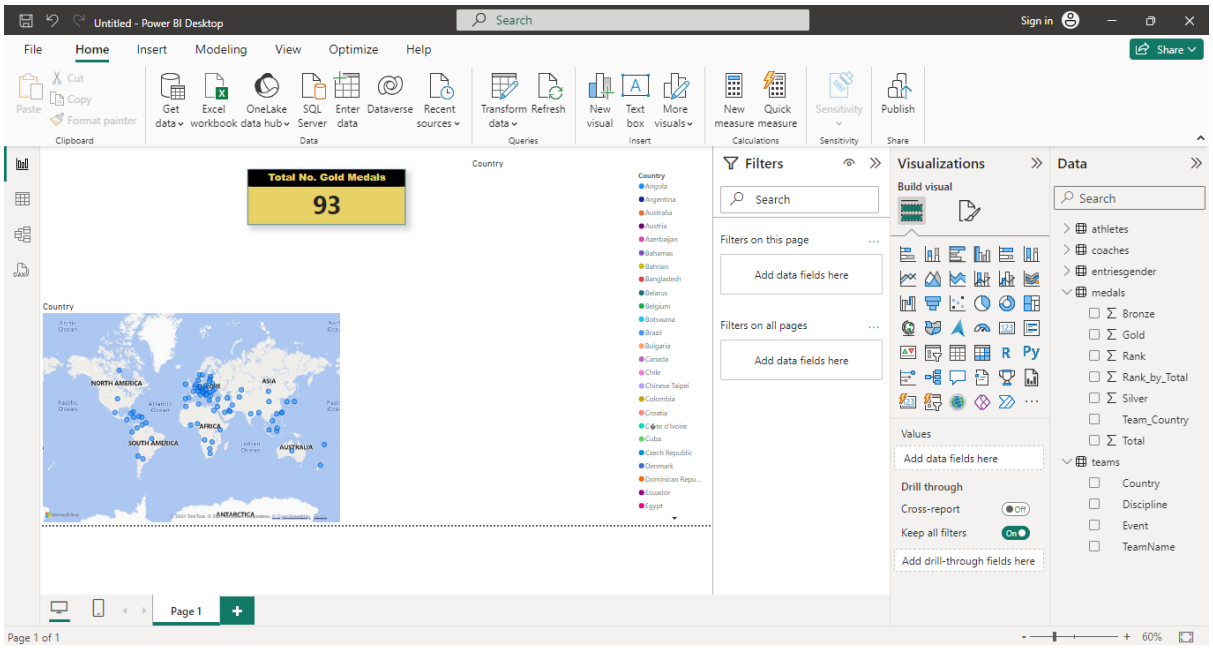
Data Modelling…

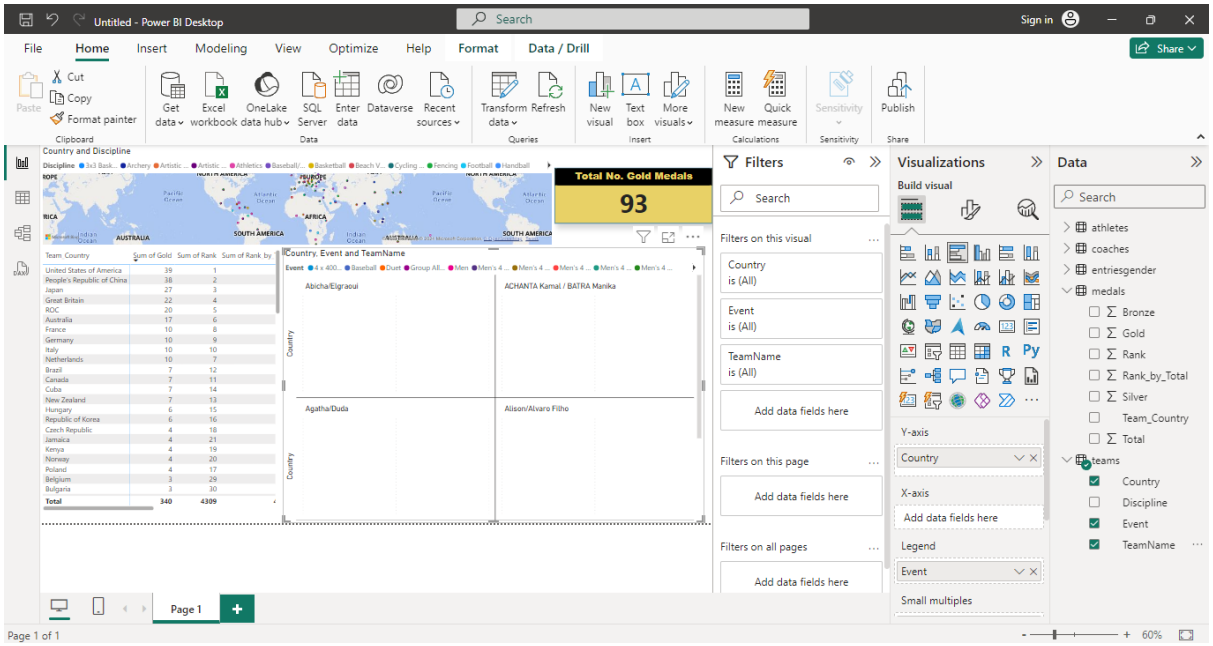



Untitled - Power BI Desktop
Search
Sign in
File
Home
Help
Share
Cut
Copy
Paste
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
Manage relationships
New measure
New column
New table
Manage roles
View as
Q&A setup
Language
Linguistic schema
Sensitivity
Publish
Clipboard
Data
Queries
Relationships
Calculations
Security
Q&A
Sensitivity
Share
Properties
Data
teams
Country
Discipline
Event
TeamName
Collapse
athletes
Country
Discipline
PersonName
Collapse
entriesgender
Discipline
Female
Male
Total
Collapse
coaches
Country
Discipline
Event
Name
Collapse
medals
Bronze
Gold
Rank
Rank_by_Total
Silver
Team_Country
Total
Collapse
All tables
83%

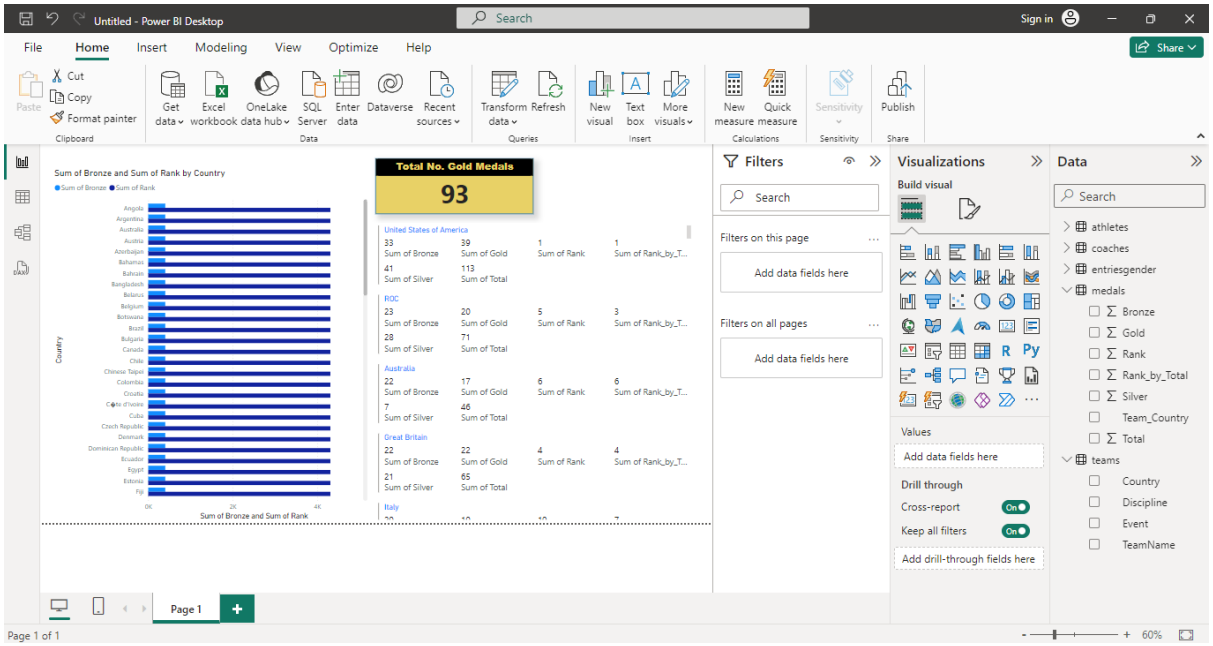

Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Cut
Copy
Format painter
Paste
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New measure
Quick measure
Sensitivity
Publish
Clipboard
Data
Queries
Insert
Calculations
Sensitivity
Share
Report view
Build visuals with your data
Select or drag fields from the Data pane onto the report canvas.
Filters
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Data
Search
athletes
coaches
entriesgender
medals
teams
Page 1
Page 1 of 1
69%


Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Share
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
93
Count of Gold
Filters
Search
Filters on this visual
Count of Gold is (All)
Add data fields here
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
Visualizations
Build visual
Fields
Count of Gold
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Data
Search
athletes
coaches
entriesgender
medals
Bronze
Gold
Rank
Rank_by_Total
Silver
Team_Country
Total
teams
Page 1
Page 1 of 1
60%


Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Total No. Gold Medals
93
Country
NORTH AMERICA
ASIA
AFRICA
SOUTH AMERICA
AUSTRALIA
ANTARCTICA
Filters
Search
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Data
Search
athletes
coaches
entriesgender
medals
Bronze
Gold
Rank
Rank_by_Total
Silver
Team_Country
Total
teams
Country
Discipline
Event
TeamName
Page 1
Page 1 of 1
60%

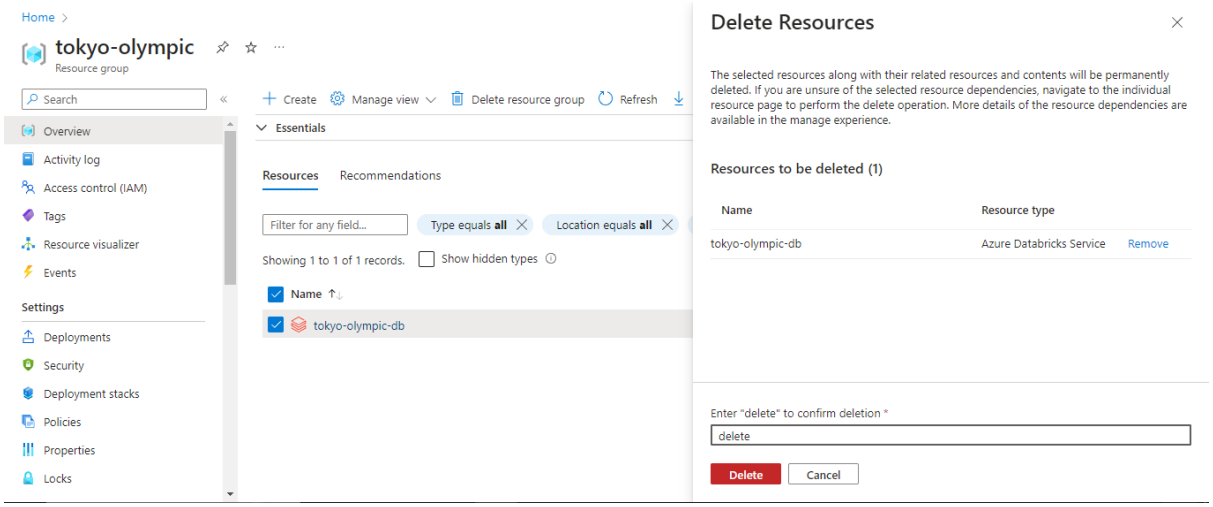


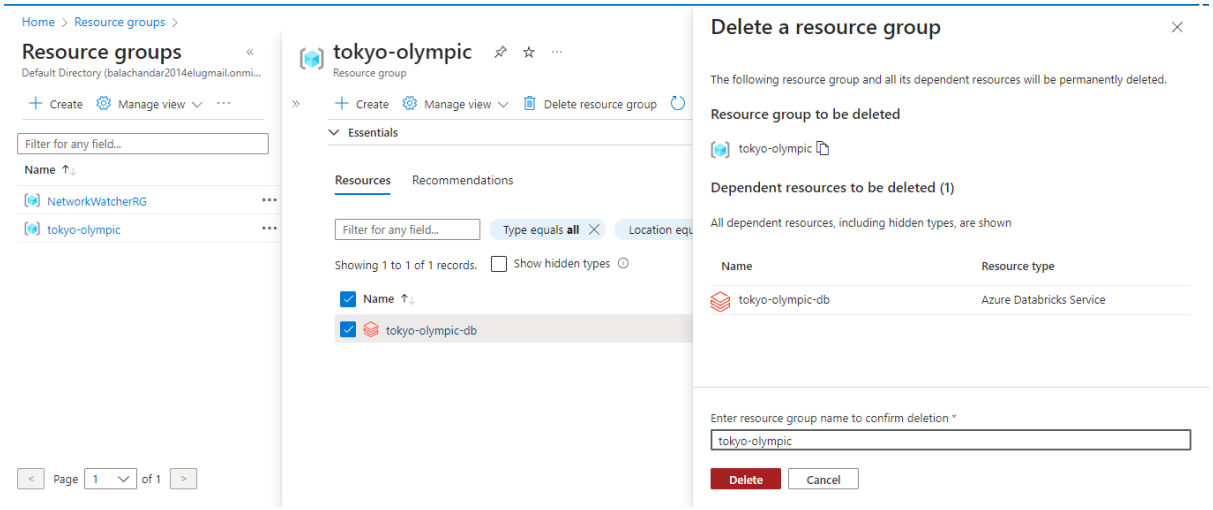
Dashboard Creation by using Microsoft Power BI.



End of Project....

Delete the Project resource and avoid the Over Billing.


Home
tokyo-olympic
Resource group
Search
Overview
Activity log
Access control (IAM)
Tags
Resource visualizer
Events
Settings
Deployments
Security
Deployment stacks
Policies
Properties
Locks
Create
Manage view
Delete resource group
Refresh
Essentials
Resources
Recommendations
Filter for any field...
Type equals all
Location equals all
Showing 1 to 1 of 1 records.
Show hidden types
Name
tokyo-olympic-db
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
Remove
Enter "delete" to confirm deletion *
delete
Delete
Cancel


Home
Resource groups
Resource groups
Default Directory (balachandar2014elugmail.onmi...
Create
Manage view
Filter for any field...
Name
NetworkWatcherRG
tokyo-olympic
Page
1
of 1
tokyo-olympic
Resource group
Delete resource group
Essentials
Resources
Recommendations
Type equals all
Showing 1 to 1 of 1 records.
Show hidden types
tokyo-olympic-db
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

Thank you!