## 

```
(https://databricks.com)
                                                                   1
   from pyspark.sql.functions import col
   from pyspark.sql.types import IntegerType, DoubleType, BooleanType, DateType
                                                                   2
   # Define Storage Account details
   storage_account_name = "tokyoolympicdatajuan" # Ensure this name is correct and follows Azure rules
   storage_account_key = "j61vrb31EbTiz1sk/s1KxvyOddPtQQKfnRPhqJxddcGqg5EvEauEep6XJnuta3W5/c+T9fGB4D02+AStj0sWEQ==" # Ensure no extra
   container_name = "tokyo-olympic-data" # Ensure this name is correct and follows Azure rules
   \# Set up the configuration for your storage account
   spark.conf.set(
       f"fs.azure.account.key.{storage_account_name}.blob.core.windows.net",
       storage_account_key
   # Mount the Blob storage to DBFS
   dbutils.fs.mount(
       source = f"wasbs://{container_name}@{storage_account_name}.blob.core.windows.net",
       mount_point = "/mnt/tokyoolympic", # You can change this mount point as needed
       extra_configs = {f"fs.azure.account.key.{storage_account_name}.blob.core.windows.net": storage_account_key}
   # List the files to check if the mounting worked
   dbutils.fs.ls("/mnt/tokyoolympic")
 Out[2]: [FileInfo(path='dbfs:/mnt/tokyoolympic/raw data/', name='raw data/', size=0, modificationTime=0),
  FileInfo(path='dbfs:/mnt/tokyoolympic/transformed data/', name='transformed data/', size=0, modificationTime=0)]
                                                                   3
    print(storage_account_name)
   print(container_name)
 tokyoolympicdatajuan
 tokyo-olympic-data
                                                                   4
    spark
   SparkSession - hive
   SparkContext
   Spark UI
   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")

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")
```

```
dbutils.fs.ls("/mnt/tokyoolympic")

Out[7]: [FileInfo(path='dbfs:/mnt/tokyoolympic/raw data/', name='raw data/', size=0, modificationTime=0),
FileInfo(path='dbfs:/mnt/tokyoolympic/transformed data/', name='transformed data/', size=0, modificationTime=0)]
```

```
%fs
ls "/mnt/tokyoolympic"

Table
```

```
8
 athletes.show()
   ABAGNALE Giovanni
                                                     Rowing|
                                   Italy|
      ABALDE Alberto|
                                   Spain|
                                                 Basketball|
      ABALDE Tamara
                                   Spain|
                                                 Basketball|
                                                   Handball|
          ABALO Luc
                                  France
       ABAROA Cesar
                                  Chile|
                                                     Rowing
      ABASS Abobakrl
                                   Sudan
                                                   Swimming|
   ABBASALI Hamideh|Islamic Republic ...|
                                                     Karate|
      ABBASOV Islam
                                                  Wrestling
                            Azerbaijan|
       ABBINGH Lois
                             Netherlands
                                                   Handball|
        ABBOT Emily
                             Australia|Rhythmic Gymnastics|
      ABBOTT Monica | United States of ... | Baseball/Softball |
|ABDALLA Abubaker ...|
                                                  Athletics|
                                  Qatar|
```

```
| ARDFFATT ADDGTT9|
                                  Egypt|
                                                    Karatei
+-----
only showing top 20 rows
                                                            9
  athletes.printSchema()
|-- PersonName: string (nullable = true)
 |-- Country: string (nullable = true)
 |-- Discipline: string (nullable = true)
                                                            10
  coaches.show()
                                              Basketball|
       ABE Katsuhiko
                                  Japan
                                                            null
        ADAMA Cherif
                           C�te d'Ivoire|
                                                 Football|
                                                            null|
          AGEBA Yuya|
                                              Volleyball|
                                  Japan
                                                            null|
|AIKMAN Siegfried ...|
                                  Japan|
                                                  Hockey
                                                             Men|
       AL SAADI Kais|
                                Germany
                                                  Hockey|
                                                             Men|
                                 Canada|Baseball/Softball|Softball|
       ALAMEDA Lonni
     ALEKNO Vladimir|Islamic Republic ...|
                                              Volleyball|
                                                             Men
     ALEKSEEV Alexey
                                               Handball|
                                  ROC |
                                                           Women
|ALLER CARBALLO Ma...|
                                  Spain|
                                              Basketball|
                                                           null|
                           Saudi Arabia|
       ALSHEHRI Saad
                                               Football|
                                                            Men
           ALY Kamal
                                 Egypt|
                                                Football|
                                                            null|
| AMAYA GAITAN Fabian|
                            Puerto Rico|
                                              Basketball|
                                                            null|
    AMO AGUADO Pablo
                                                Football|
                                                            null|
                                Spain
   ANDONOVSKI Vlatko United States of ...
                                                Football|
                                                           Women
                                                 Hockey
        ANNAN Alyson|
                            Netherlands|
                                                           Women
  ARNAU CREUS Xavier
                              Japan
                                                 Hockey
                                                           Women
       ARNOLD Graham
                              Australia
                                                Football|
                                                             Men
                                                Handball|
        AXNER Tomas
                               Sweden
                                                           Womenl
only showing top 20 rows
                                                            11
  coaches.printSchema()
root
|-- Name: string (nullable = true)
 |-- Country: string (nullable = true)
 |-- Discipline: string (nullable = true)
 |-- Event: string (nullable = true)
                                                            12
  entriesgender.show()
| Artistic Gymnastics|
                        98 | 98 | 196 |
```

```
Reacu volleAball
                     48| 48|
                               961
            Boxing| 102| 187|
                               289
       Canoe Slalom | 41 | 41 |
                               82 l
       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 |
                      72 | 71 | 143 |
            Diving|
         Equestrian|
                      73 | 125 | 198 |
           Fencing | 107 | 108 | 215 |
+----+
only showing top 20 rows
```

```
entriesgender.printSchema()

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

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

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

```
16
 medals.show()
                  Japan | 27|
                                14|
                                      17|
                                            58|
                                                          5|
           Great Britain| 22|
   4|
                                21
                                            65
                                                          4|
                                      22
   5|
                    ROC| 20|
                                28|
                                      23|
                                            71|
                                                          3|
   6
              Australia| 17|
                                7|
                                      22
                                                         6|
                                            46
  7|
             Netherlands | 10|
                                                         9|
                                12
                                      14
                                            36
   8
                 France| 10|
                                12
                                      11
                                           33
                                                         10
| 9|
                Germany| 10|
                                11|
                                      16|
                                           37|
                                                         8|
```

```
151
                 Hungary
                           61
                                  71
                                        71
                                             201
                                                           13|
  16
        Republic of Korea
                            6
                                  4
                                        10
                                             20
                                                           13
                                                           19|
171
                  Poland
                           41
                                  5 l
                                         5 l
                                             14
| 18|
           Czech Republic
                                         3|
                                             11|
                                                           23|
| 19|
                   Kenya|
                           4
                                  4
                                         2
                                             10
                                                           25
                                                           29|
20
                  Norway|
                           4|
                                  2|
                                         2|
                                              8|
only showing top 20 rows
```

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

```
medals = medals.withColumn("Rank",col("Rank").cast(IntegerType()))\
    .withColumn("Gold",col("Gold").cast(IntegerType()))\
    .withColumn("Silver",col("Silver").cast(IntegerType()))\
    .withColumn("Bronze",col("Bronze").cast(IntegerType()))\
    .withColumn("Total",col("Total").cast(IntegerType()))\
    .withColumn("Rank by Total",col("Rank by Total").cast(IntegerType()))
```

```
20
teams.show()
       China|3x3 Basketball|People's Republic...|
                                                        Women|
     France|3x3 Basketball|
                                          France
                                                        Women I
       Italy|3x3 Basketball|
                                           Italy|
                                                        Women|
       Japan|3x3 Basketball|
                                           Japan
                                                          Men|
      Japan|3x3 Basketball|
                                          Japan
                                                        Women
     Latvia|3x3 Basketball|
                                          Latvia|
                                                          Men|
```

```
komania|3x3 Basketball|
                                 котапıа
                                             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|
  Bangladesh|
                Archery|
                              Bangladesh| Mixed Team|
     Belarus
                Archery|
                               Belarus|Women's Team|
+-----
only showing top 20 rows
```

```
teams.printSchema()

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

```
22
  # Find the top countries with the highest number of gold medals
  top_gold_medal_countries = medals.orderBy("Gold", ascending=False).select("TeamCountry","Gold").show()
               Japan | 27 |
       Great Britain | 22|
                 ROC | 20 |
           Australia | 17|
         Netherlands | 10|
              France | 10|
             Germany| 10|
               Italy | 10
              Canada
              Brazil
         New Zealand | 7|
                Cuba | 7|
             Hungary|
                       61
   Republic of Korea
              Poland
      Czech Republic
               Kenya
                       4
              Norway| 4|
only showing top 20 rows
```

```
# 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()
```

	Raseball/Sottball	90  144	234 0.	.38461538461538464	0.6153846153846154
	Basketball	144   144	288	0.5	0.5
	Beach Volleyball	48   48	96	0.5	0.5
	Boxing	102  187	289 0.	35294117647058826	0.6470588235294118
	Canoe Slalom	41   41	82	0.5	0.5
	Canoe Sprint	123  126	249   6	0.4939759036144578	0.5060240963855421
	Cycling BMX Frees	10  9	19  0	0.5263157894736842	0.47368421052631576
	Cycling BMX Racing	24   24	48	0.5	0.5
	Cycling Mountain	38  38	76	0.5	0.5
	Cycling Road	70  131	201   0	3.3482587064676617	0.6517412935323383
	Cycling Track	90  99	189 0.	47619047619047616	0.5238095238095238
	Diving	72  71	143   6	0.5034965034965035	0.4965034965034965
	Equestrian	73  125	198  0	3.3686868686868687	0.6313131313131313
	Fencing	107  108	215 0.	49767441860465117	0.5023255813953489
+	+-	++		+	+
(	only showing top 20 rows				
L					

24

athletes.repartition(1).write.mode("overwrite").option("header",'true').csv("/mnt/tokyoolympic/transformed data/athletes")

25