INTRODUCTION TO DATA MANAGEMENT
PROJECT REPORT

(Project Semester August-December 2021)

Tokyo Olympic 2021 Dashboard

Submitted by

Akash Shiladitya

Registration No: 11901807

Roll No: B36

Section: KM039

Subject Code : INT217

Under the Guidance of

Sameeksha Khare , UId 26806

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

DECLARATION

I, Akash Shiladitya, student of Lovely Professional University (Bachelor of Technology) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 14 December 2021

Name : Akash Shiladitya

Registration No :11901807

Acknowledgement

Table Of Content

In this project I have tried to show the statistics of Tokyo Olympic based on different parameters, by creating a dashboard. A dashboard is a visual display of the most important information needed to achieve one or more objectives i.e., crux of whole raw data.

Objectives

The objectives of my project are as follows:

Objective1. Visualization of medals won by countries.

Objective2. Which sports consists of highest female and male players.

Objective3. Counting number of athletes in each individual sports.

Objective4. Counting number of athletes from their respective country.

Objective5. Counting number of coaches from their respective country.

Source of Dataset

I downloaded this data set from the website named: **Kaggle**

https://www.kaggle.com/arjunprasadsarkhel/2021-olympics-in-tokyo

ETL Process

The full form of ETL is Extract Transform Load

Extraction:

The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML, and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult. Therefore, this is one of the most important steps of ETL process.

Transformation:

The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

Filtering – loading only certain attributes into the data warehouse.

Joining – joining multiple attributes into one.

Splitting – splitting a single attribute into multiple attributes.

Sorting – sorting tuples based on some attribute (generally key-attribute).

Loading:

The third and final step of the ETL process is loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

In my project I have used ETL Process to hide some unwanted rows where data were not clear and was filled with 'NA'.

Data Before ETL Process

	A	B	C	D	E	F	G	H
1	Rank	Team/NO(	Gold	Silver	Bronze	Total	Rank by Total	
2	1	United Sta	39	41	33	113	1	
3	2	People's R	38	32	18	88	2	
4	3	Japan	27	14	17	58	5	
5	4	Great Brit;	22	21	22	65	4	
6	5	ROC	20	28	23	71	3	
7	6	Australia	17	7	22	46	6	
8	7	Netherlan(	10	12	14	36	9	
9	8	France	10	12	11	33	10	
10	9	Germany	10	11	16	37	8	
11	10	Italy	10	10	20	40	7	
12	11	Canada	7	6	11	24	11	
13	12	Brazil	7	6	8	21	12	
14	13	New Zeala	7	6	7	20	13	
15	14	Cuba	7	3	5	15	18	
16	15	Hungary	6	7	7	20	13	
17	16	Republic o	6	4	10	20	13	
18	17	Poland	4	5	5	14	19	
19	18	Czech Rep	4	4	3	11	23	
20	19	Kenya	4	4	2	10	25	
21	20	Norway	4	2	2	8	29	
22	21	Jamaica	4	1	4	9	26	
23	22	Spain	3	8	6	17	17	
24	23	Sweden	3	6	0	9	26	
25	24	Switzerlan	3	4	6	13	20	
26	25	Denmark	3	4	4	11	23	
27	26	Croatia	3	3	2	8	29	
28	27	Islamic Re|	3	2	2	7	33	
29	28	Serbia	3	1	5	9	26	
30	29	Belgium	3	1	3	7	33	
31	30	Bulgaria	3	1	2	6	39	

Details

Data After ETL Process.

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

Analysis

Objective 1: Visualization of medals won by countries.
General Description:
In this objective I have displayed number of medals won by the countries as gold, silver and bronze so that we can be able to distinguish which country won most medals.

Specific Requirement:
 I have added all the individual medals in their separate category as gold, silver and bronze and after that I have performed sorting to get the appropriate data to show the result and I have showed only top ten countries whose's won most medals in Olympic.

Row Labels	Sum of Gold	Sum of Silver	Sum of Bronze
United States of America	39	41	33
People's Republic of China	38	32	18
Japan	27	14	17
Great Britain	22	21	22
ROC	20	28	23
Australia	17	7	22
Netherlands	10	12	14
France	10	12	11
Germany	10	11	16
Italy	10	10	20
Grand Total	203	188	196

This table shows number of medals won by top 10 country.

The fields of this table are as follows.

The pivot chart based on this table was
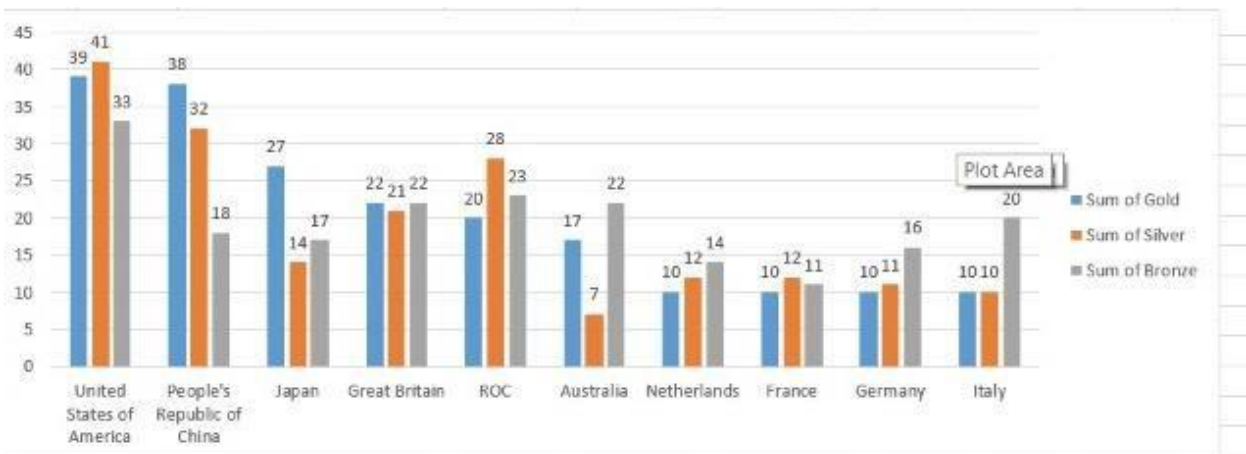
Which is showing the top 10 countries whose's won most medals.

Objective 2 : Which sports consists of highest female and male players.
General Description :
In this objective I have displayed which sports have highest number of male and female players.

Specific Requirement :

To achieve this objective I have used this formula
 =INDEX (A2:D47,MATCH(MAX(B2:B47),B2:B47,0),1) to find the which sport have the maximum number of female players.

And I have used this formula **=INDEX (A2:D47,MATCH(MAX(C2:C47),C2:C47,0),1)** to find the which sport have the maximum number of male players.

Here I have used this formula =MAX(B2:B47) to find out the maximum number of female players from the gender table.

f_x =MAX(B2:B47)
   C   D   E
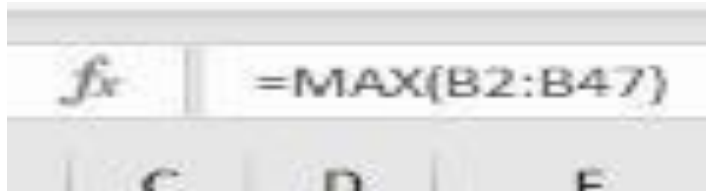
Here I have used this formula =MAX(C2:C47) to find out the maximum number of male players from the gender table.

f_x =MAX(C2:C47)
   C   D   E

After all of that we finally get which sport have the maximum number of male and female players from the gender table.

Question:- Which sports consists of highest female and male players?		
Athletics	969	
Athletics	1072	

Objective3: Counting number of athletes in each individual sports.
General Description:
In this objective I have counted number of athletes from each single sports.

Specific Requirement:
To achieve this objective I have used this formula =UNIQUE(C2:C11086) to find the unique sports from the athlete's table.

Artistic Gymnastics
Rowing
Basketball
Handball
Swimming
Karate
Wrestling
Rhythmic Gymnastics
Baseball/Softball
Athletics
Artistic Swimming
Judo
Shooting
Table Tennis
Football
Taekwondo
Fencing
Badminton
Boxing

61°F Haze ∧
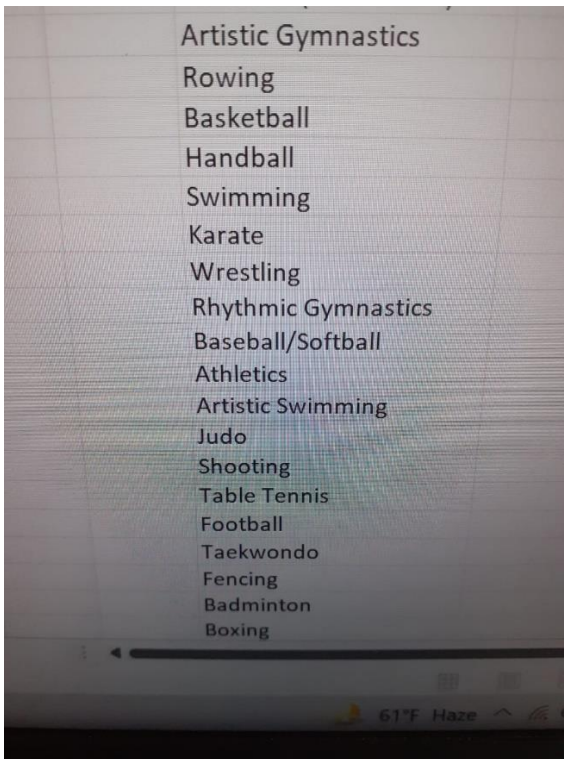
Here I have used this formula =COUNTIF (C2: C11086,F21) to find out the number of athletes from the each sports.

Cycling Road	190
Artistic Gymnastics	187
Rowing	496
Basketball	280
Handball	343
Swimming	743
Karate	77
Wrestling	279
Rhythmic Gymnastics	95
Baseball/Softball	220
Athletics	2068
Artistic Swimming	98
Judo	373
Shooting	342
Table Tennis	164
Football	567
Taekwondo	123
Fencing	249
Badminton	164
Boxing	270
Weightlifting	187

Here this chart shows top 10 sports which have maximum number of athletes.
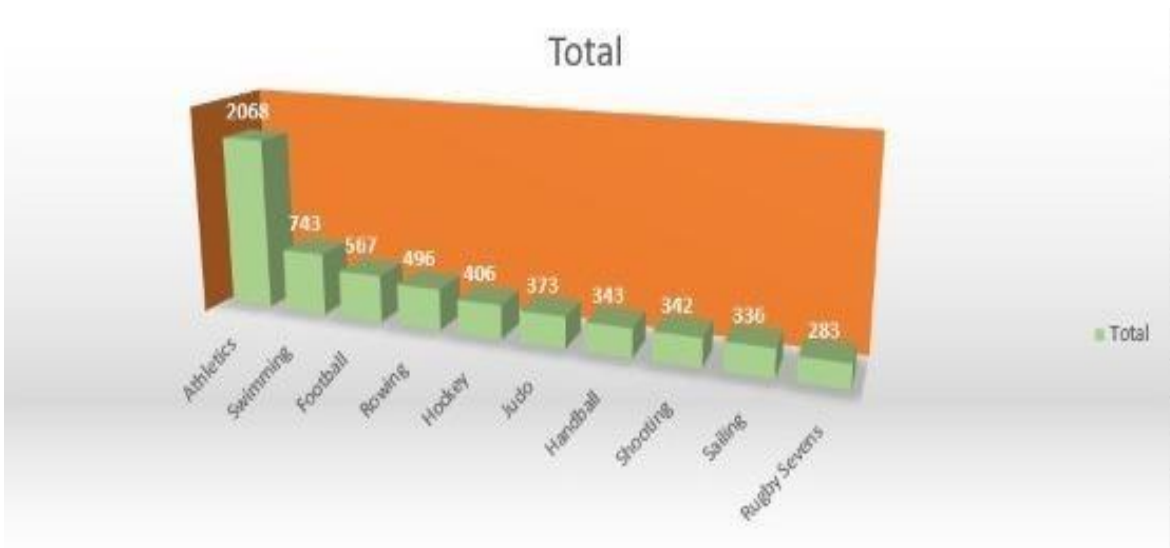
Objective4: Counting number of athletes from their respective country.
General Description:
In this objective I have counted number of athletes from their respective country so that we can know how many athletes represent their country in the Olympics.

Specific Requirement:
I have counted NOC in values field by the pivot chart and pivot table.

Here is the list of top 25 countries which is have the greatest number of athletes represent their country.

Row Labels	Count of NOC
United States of America	615
Japan	586
Australia	470
People's Republic of China	401
Germany	400
France	377
Canada	368
Great Britain	366
Italy	356
Spain	324
ROC	318
Brazil	291
Netherlands	274
Republic of Korea	223
New Zealand	202
Poland	195
Argentina	180
South Africa	171
Mexico	155
Hungary	155
Ukraine	152
Egypt	133
Sweden	129
Belgium	125
Czech Republic	117
India	117
Grand Total	**7200**

Here is the field setting for the pivot table.

This chart shows how many numbers of athletes represents their country in the Olympics.
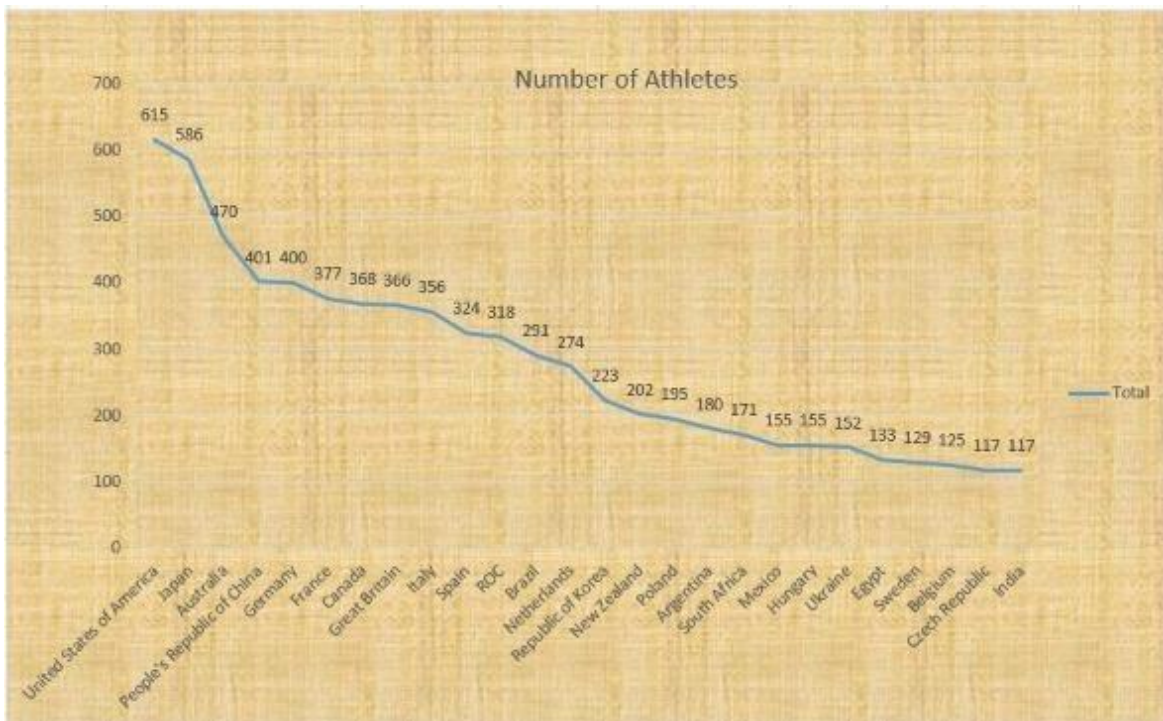
Objective5: Counting number of coaches from their respective country.
General Description:
In this objective I have counted number of coaches from their respective country so that we can know how many coaches a country carried for their athletes in the Olympics.

Specific Requirement:
I have counted NOC in values field by the pivot chart and pivot table.

Here we are showing no of coaches for athletes.

Row Labels	Count of NOC
Japan	35
United States of America	28
Spain	28
Australia	22
Canada	16
Italy	14
South Africa	12
Egypt	12
ROC	12
People's Republic of China	12
Grand Total	191

Here is the field setting for the pivot table.

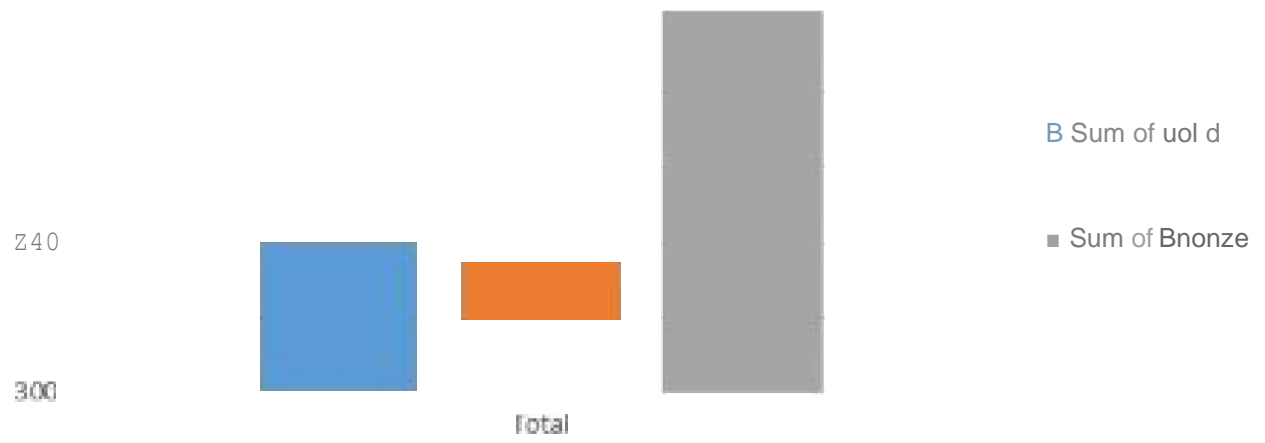This chart shows how many number of coaches a country carried For their athletes in the Olympics

Making of Dashboard

First, I created a new sheet named as All data Set in which I combined all the CSV's files so that I can understand the data easily and then I changed the CSV's file into a table format so that we can perform several tasks in it and then I created pivot chart of all tables with the help of slicer.

THESE ARE THE PIVOT CHARTS GIVEN BELOW....

Z40

300

Total

B Sum of uol d

■ Sum of Bnonze

5900

5‹500

5•tDO

■ Sum of Female

a SulTi oT k1ale

Total

4A

Iafal

Total

50

ZU
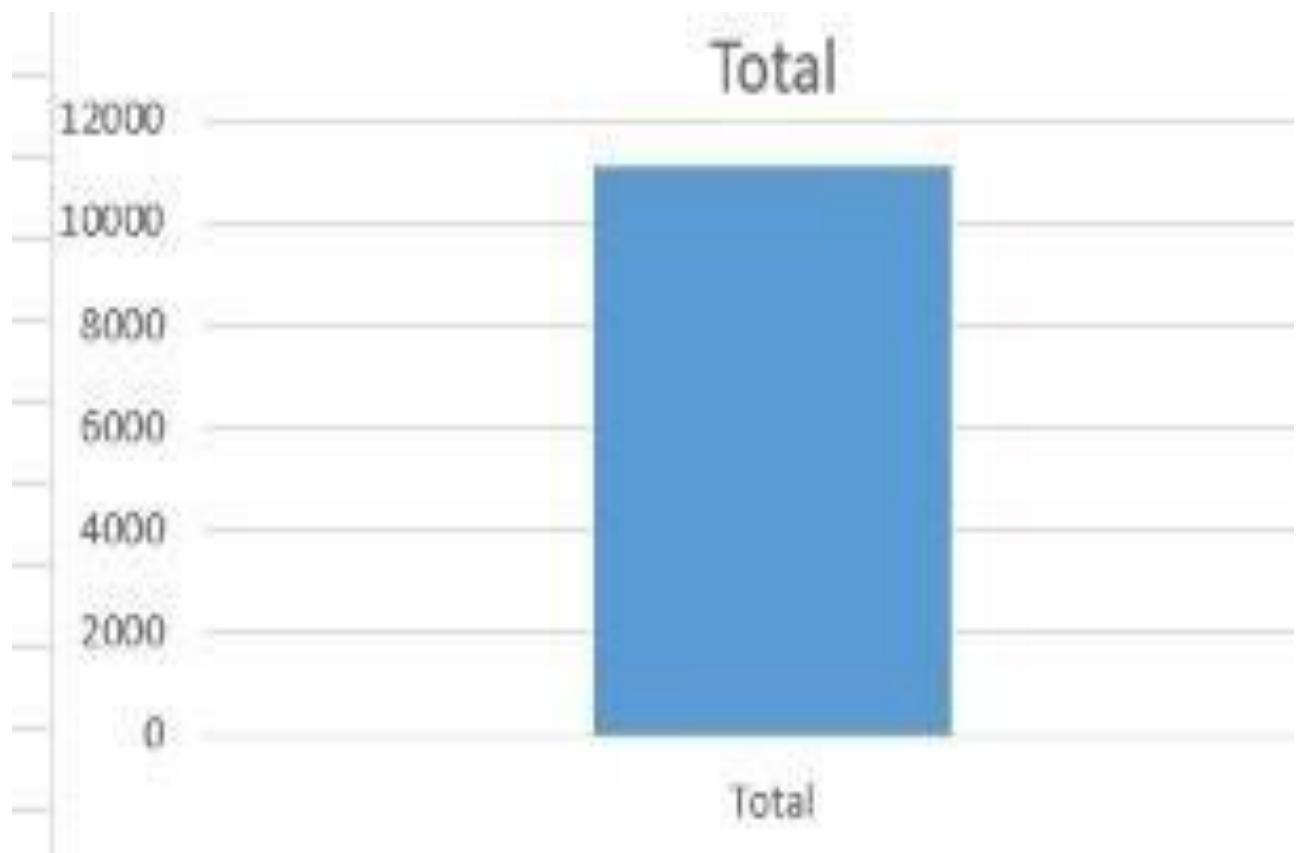
10

0

Swimming · ball/Softball · Basketball · Handball · Hockey · gby Sevens · Volleyball · Water

■TMa

Plot Area

Chart Area

SLICERS

In this project I have created three Slicer named as Team /NOC, Discipline and Events. Slicers provide buttons that you can click to filter tables, or PivotTables. In addition to quick filtering, slicers also indicate the current filtering state, which makes it easy to understand what exactly is currently displayed.
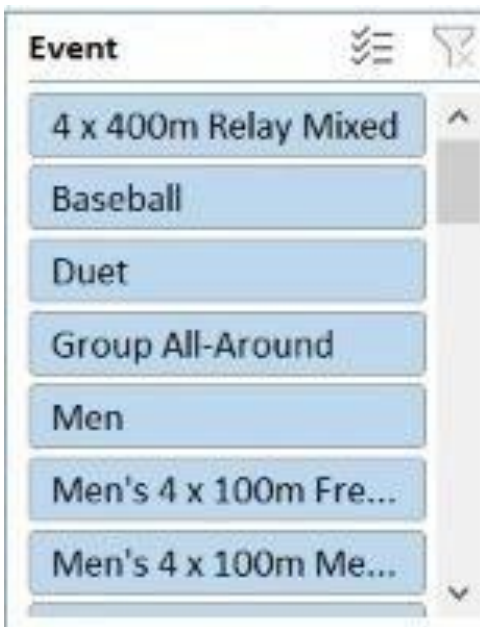
In this slicer I have showed team and country names.

In this slicer I have showed Discipline and sports names.
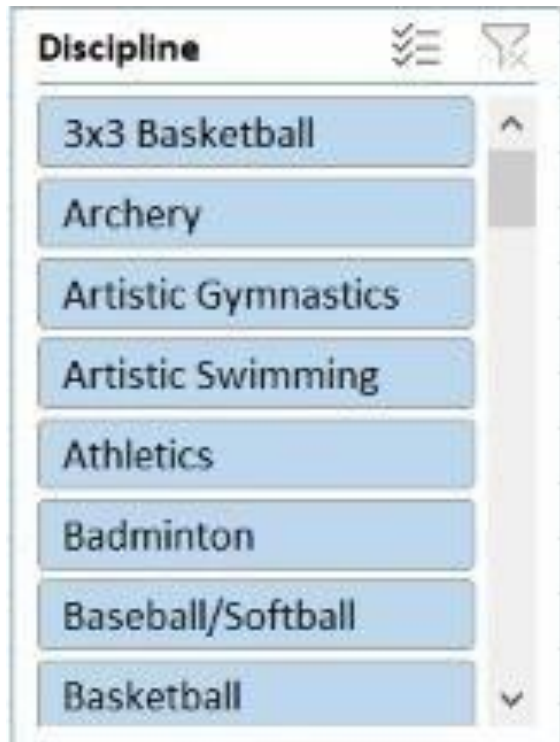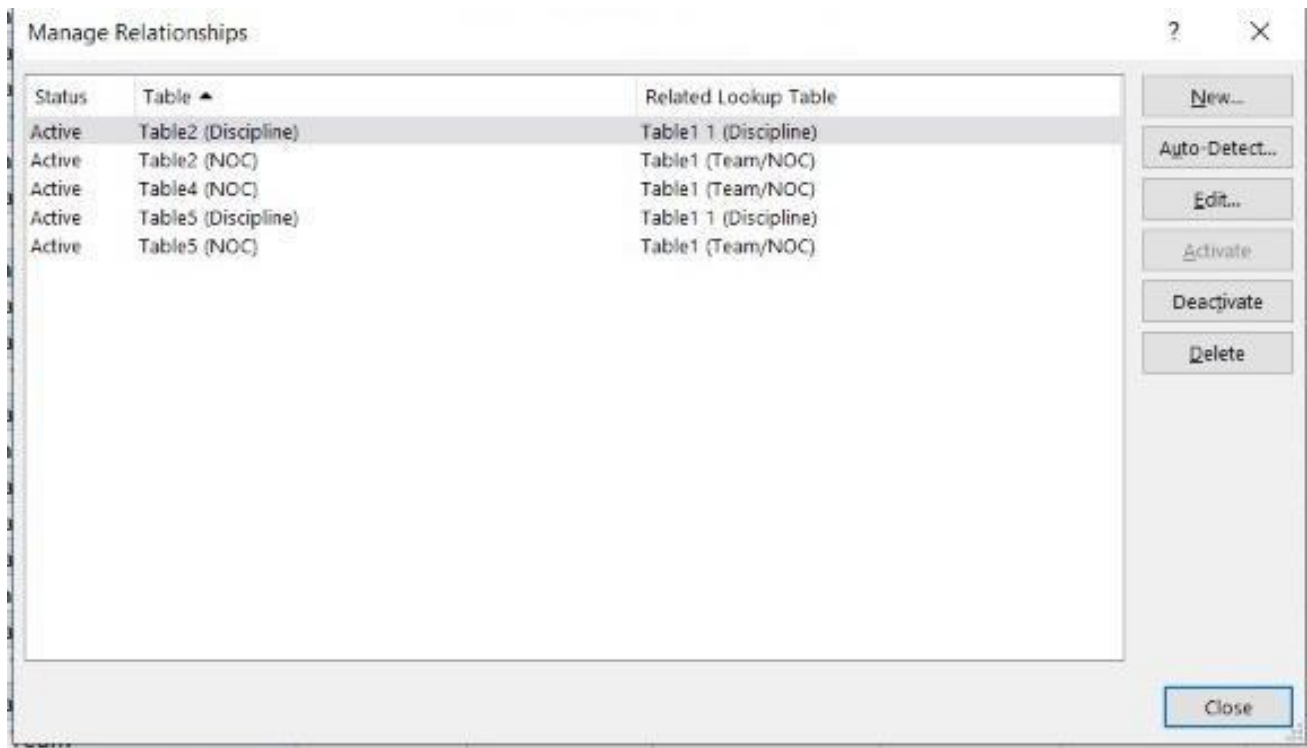

In this slicer I have showed Events names.

Relationship between two pivot table

A table relationship works by matching data in key fields — often a field with the same name in both tables. In most cases, these matching fields are the primary key from one table, which provides a unique identifier for each record, and a foreign key in the other table.

For creating relationships between tables first I have clicked on pivot table and the I went into a pivot table analyze and then went into the relationships options and from there I have created relationship between tables.

IN GIVEN BELOW THIS IS MY RELATIONSHIPS BETWEEN TABLES AS FOLLOWS.

Status	Table ▲	Related Lookup Table
Active	Table2 (Discipline)	Table1 1 (Discipline)
Active	Table2 (NOC)	Table1 (Team/NOC)
Active	Table4 (NOC)	Table1 (Team/NOC)
Active	Table5 (Discipline)	Table1 1 (Discipline)
Active	Table5 (NOC)	Table1 (Team/NOC)

Manage Relationships

New... Auto-Detect... Edit... Activate Deactivate Delete Close
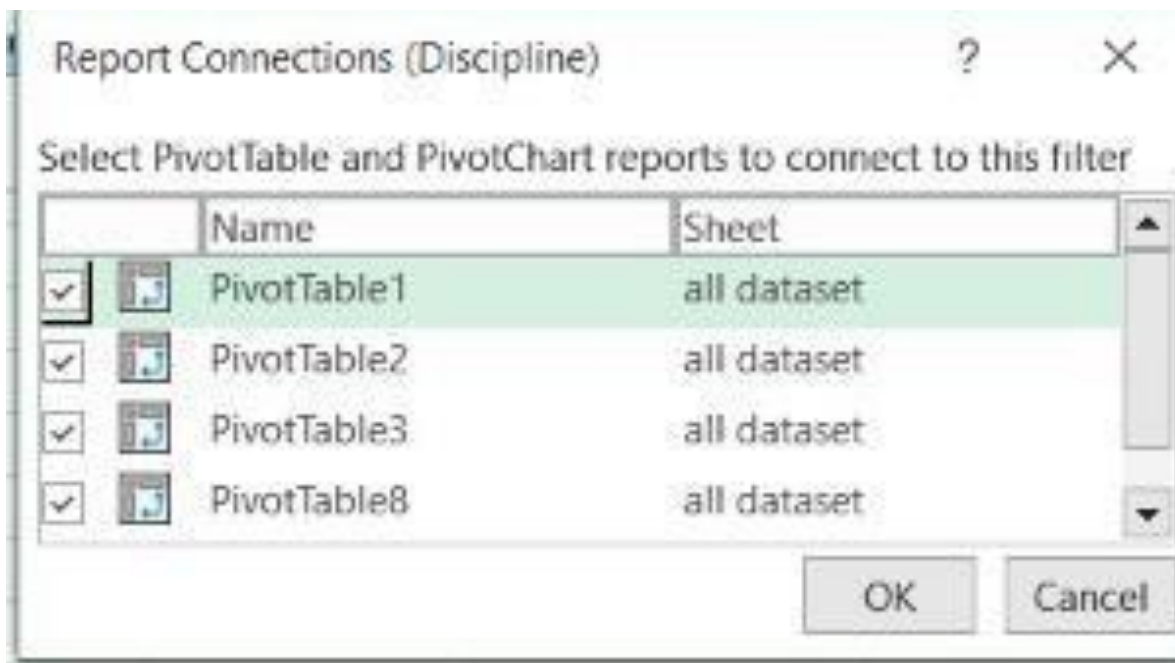
REPORT CONNECTION

Report Connections icon, you can have a slicer control multiple pivot tables on a dashboard—but only if all the pivot tables are from the same data set.
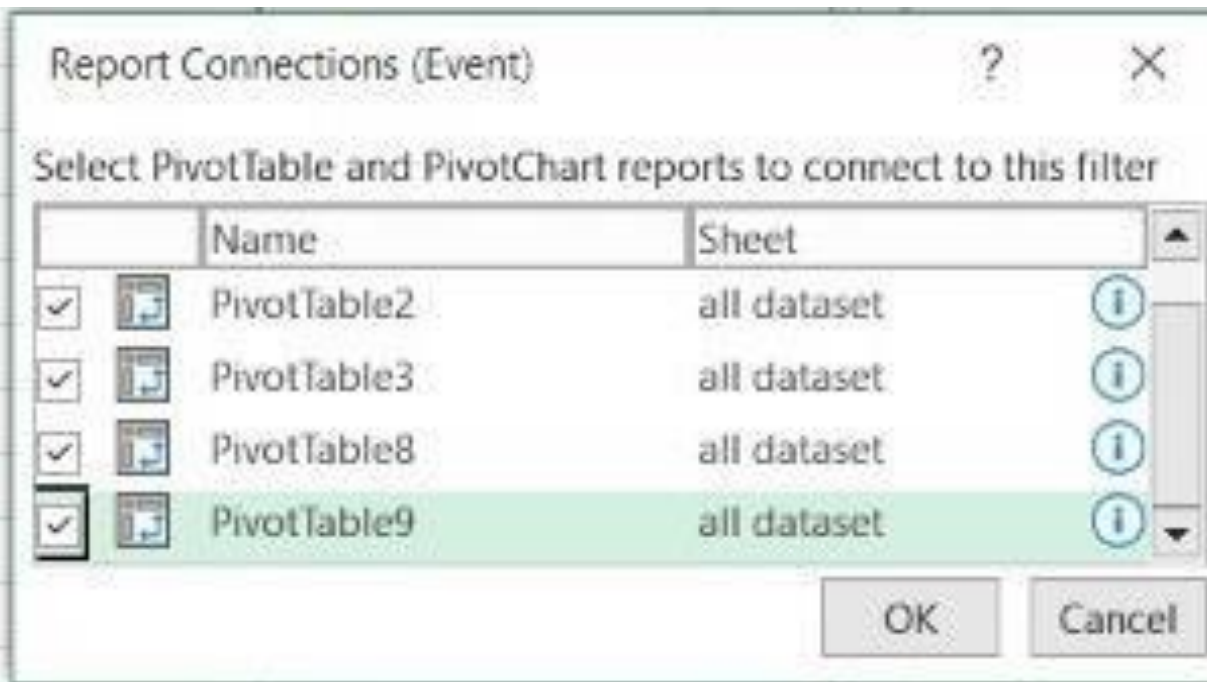
THERE I HAVE CREATED CONNECTION BETWEEN SLICER AND MULTIPLE PIVOT CHARTS AND THIS REPORT CONNCETION IS FOR TEAM/NOC.
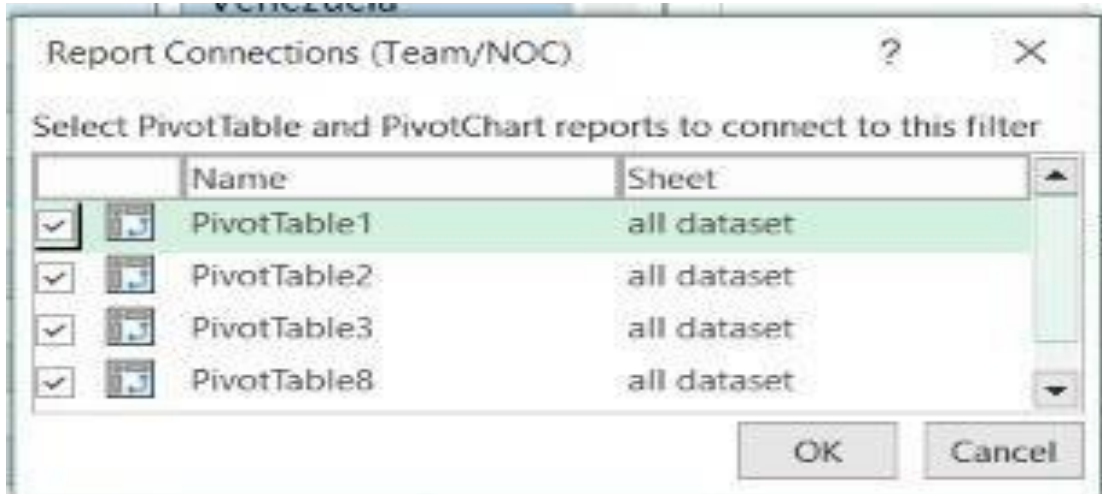
THERE I HAVE CREATED CONNECTION BETWEEN SLICER AND MULTIPLE PIVOT CHARTS AND THIS REPORT CONNCETION IS FOR DISCIPLINE.

THERE I HAVE CREATED CONNECTION BETWEEN SLICER AND MULTIPLE PIVOT CHARTS AND THIS REPORT CONNCETION IS FOR EVENT.
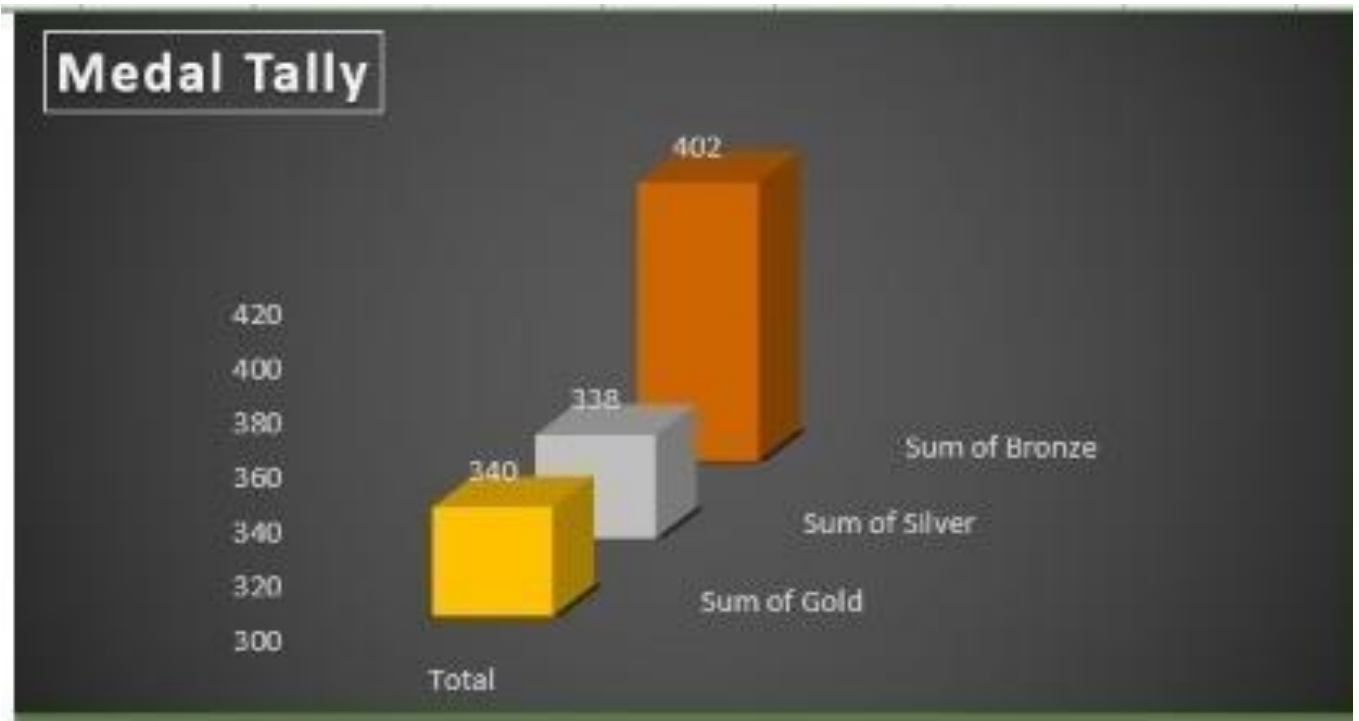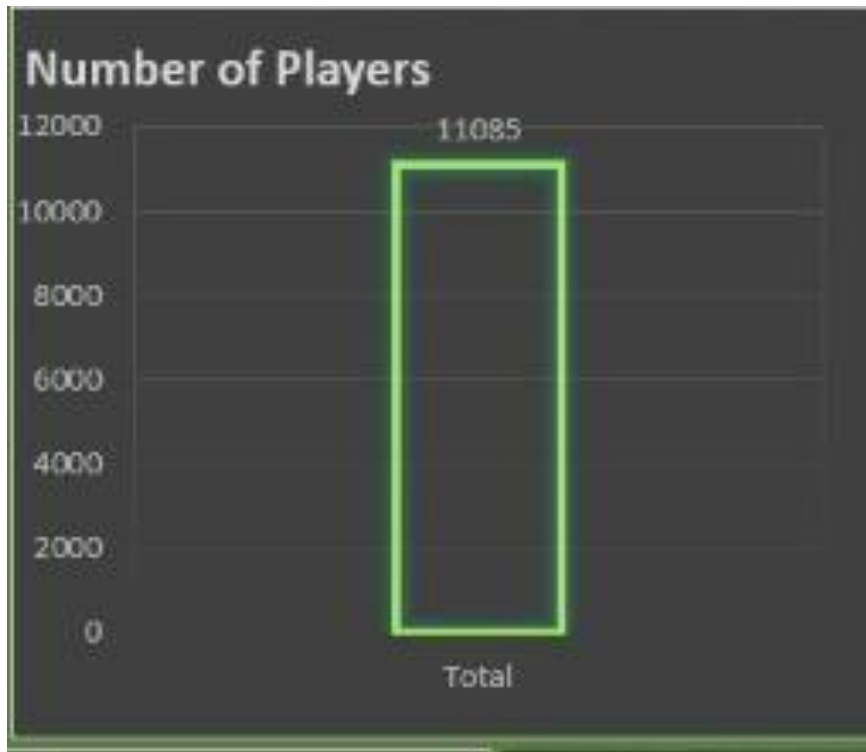
DASHBOARD

A dashboard is a visual representation of key metrics that allow you to quickly view and analyze your data in one place.

IN THIS DASHBOARD I HAVE CLUB ALL THE PIVOT CHARTS AND SLICER IN A SINGLE DASHBOARD TO SHOW ALL THE DATA AT A TIME.
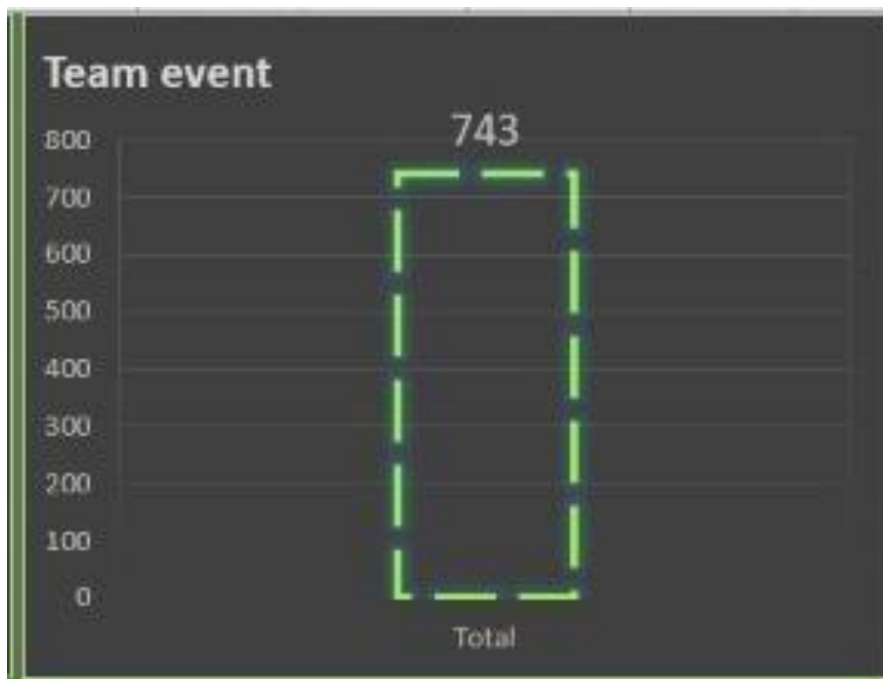
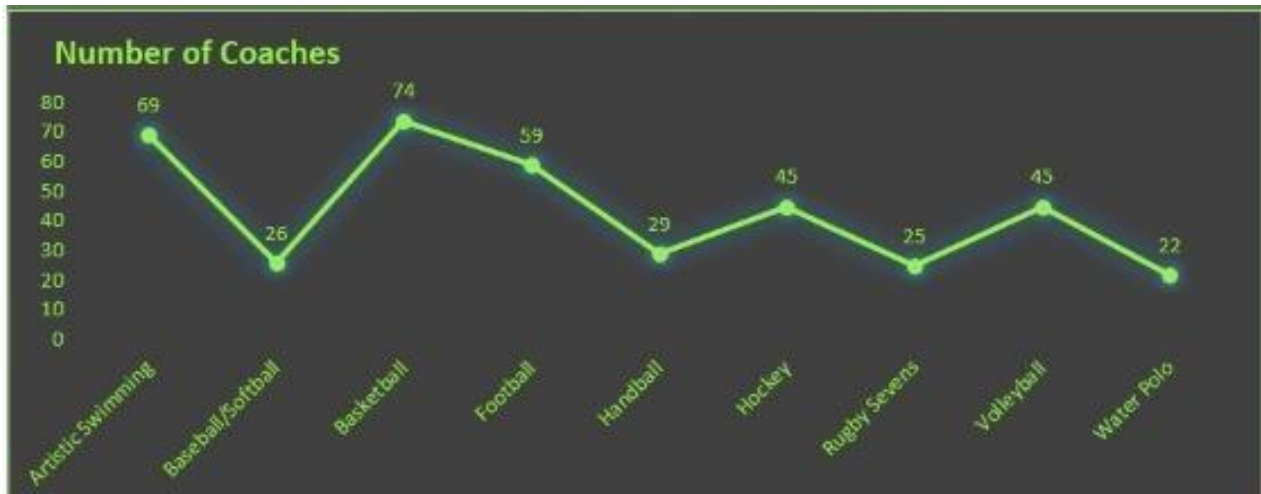IN THIS DASHBOARD I HAVE SHOWED MEDALS TALLY OF OLYMPICS.

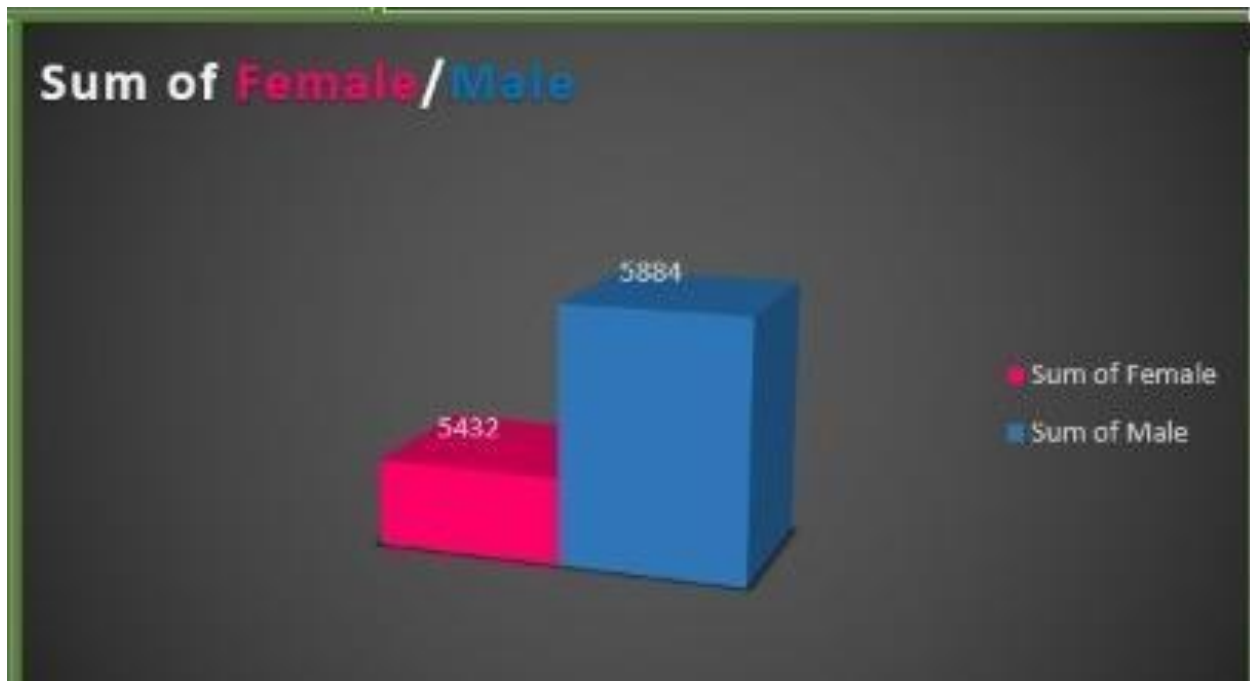IN THIS DASHBOARD I HAVE SHOWED TOTAL NUMBER OF PLAYERS.

IN THIS DASHBOARD I HAVE SHOWED NUMBER OF TEAM EVENTS.

IN THIS DASHBOARD I HAVE SHOWED NUMBER OF COACHES.

IN THIS DASHBOARD I HAVE SHOWED NUMBER OF MALE AND FEMALE ATHLETES.

THESE ARE THE THREE SLICERS OF DASHBOARD

Team/NOC

- Slovakia
- Slovenia
- South Africa
- Spain
- Sweden
- Switzerland
- Syrian Arab Rep...
- Thailand
- Tunisia
- Turkey
- Turkmenistan
- Uganda
- Ukraine
- United States of...
- Uzbekistan
- Venezuela
- (blank)

Discipline

- 3x3 Basketball
- Archery
- Artistic Gymnast...
- Artistic Swimming
- Athletics
- Badminton
- Baseball/Softball
- Basketball
- Beach Volleyball

Event

- Women's 4 x 40...
- Women's Épée T...
- Women's Foil Te...
- Women's Madis...
- Women's Sabre ...
- Women's Team
- Women's Team ...

References

[1] https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

[2] https://www.youtube.com/watch?v=OW5OgsLpDCQ

Bibliography

[1] LPU Report Format

[2] https://www.youtube.com/watch?v=OW5OgsLpDCQ

[3] https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

[4] https://www.kaggle.com/arjunprasadsarkhel/2021-olympics-in-tokyo

[5] Microsoft Excel