--Client 3: SportsStats (Olympics Dataset - 120 years of data)
SportsStats is a sports analysis firm partnering with local news and elite personal trainers to provide “interesting” insights to help their partners. Insights could be patterns/trends highlighting certain groups/events/countries, etc. for the purpose of developing a news story or discovering key health insights.--

PROJECT PROPOSAL AND DATA SELECTION/PREPARATION

--I choose the dataset "athlete events," and the American Olympic Investment Committee is the customer. They requested our opinion on how to bring more business and Opportunities . We can effectively address that issue with the help of this dataset.

--I love sports and enjoy analysing sports data to find maximum business oportunities--

--I am using postgres not **data bricks**--

--Did the learner describe the steps they took to import and clean the data?--

.Step1: Create a Table in Postgres

Step 2: Copy the data to the table.

Step 3: Delete the N/A

Step 4: Check data type of columns if needed alther the data type.--

.--Perform initial exploration of data and provide some screenshots or display some stats of the data you are looking at

I have put multiple screen shots--

2. Develop Project Proposal

   My project focuses on analyzing historical data on individuals age and probability of winning a Medal.. My target audience is representatives of Olympic Investment committees and individual athletes. The project will give them an overview so they can invest in facilities effectively for sports.

Hypotheses

Questions - Is there any correlation between the age of the participants and the probability of winning a medal. Which country has the most medals

```
CREATE TABLE public.athlete_events
(ID integer,
Name varchar (200),
Sex varchar (100),
Age varchar (100),
Height varchar (10),
Weight varchar (100),
Team varchar (100),
NOC varchar (5),
Games varchar (100),
Year varchar (100),
 Season varchar(100),
 City varchar (100),
 Sport varchar (100),
 Event varchar (100),
 Medal varchar (100));
```

– Type this code to check the data--

```
SELECT * FROM athlete_events;
```

–

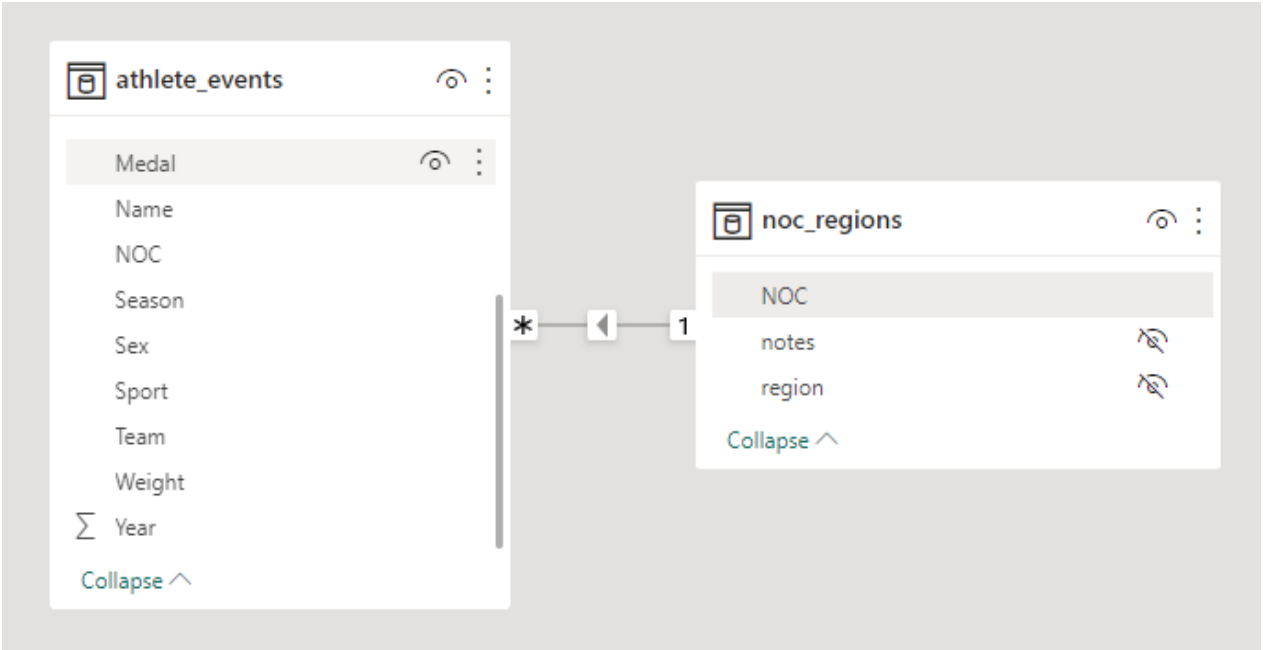
--Import data--

```
COPY athlete_events
FROM 'C:\Users\laxmi\Downloads\athlete_events.csv'
WITH (FORMAT CSV, HEADER);
```

--check all the rows have been uploaded--

--COPY 271116 Query returned successfully in 1 secs 878 msec. –

Did the learner create an ERD to show the relationships of the data they are exploring?



--Primary key is NOC--

SELECT* From public.athlete_events;

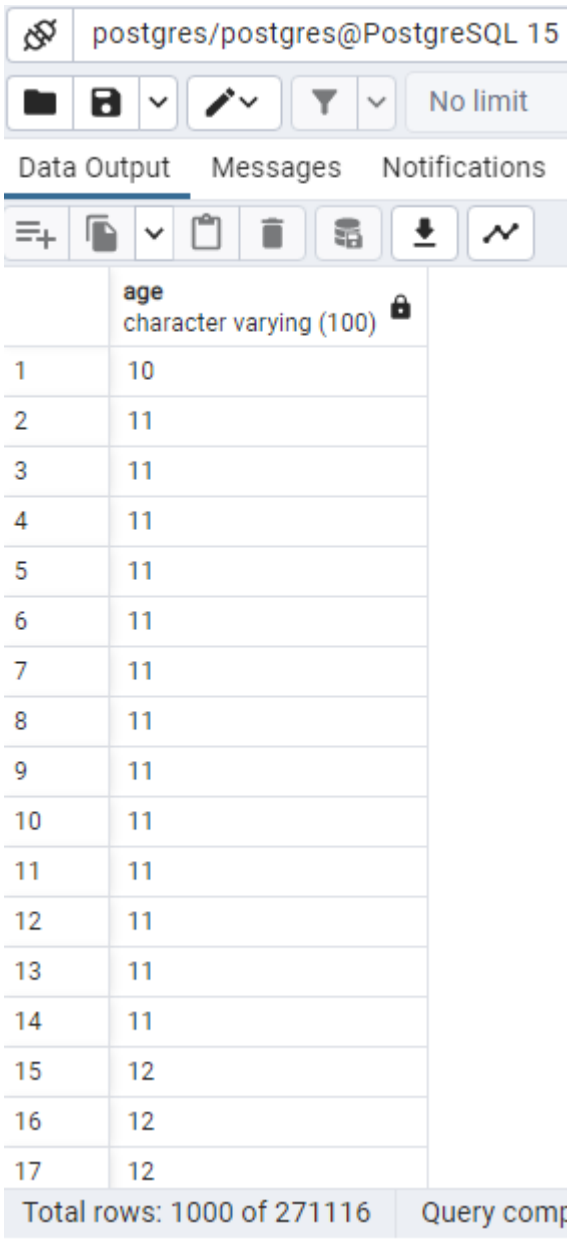
No limit

Data Output | Messages | Notifications

	id integer	name character varying (200)	sex character varying (100)	age character varying (100)	height character varying (10)	weigl chara
1	1	A Dijiang	M	24	180	80
2	2	A Lamusi	M	23	170	60
3	3	Gunnar Nielsen Aaby	M	24	NA	NA
4	4	Edgar Lindenau Aabye	M	34	NA	NA
5	5	Christine Jacoba Aaftink	F	21	185	82
6	5	Christine Jacoba Aaftink	F	21	185	82
7	5	Christine Jacoba Aaftink	F	25	185	82
8	5	Christine Jacoba Aaftink	F	25	185	82
9	5	Christine Jacoba Aaftink	F	27	185	82
10	5	Christine Jacoba Aaftink	F	27	185	82
11	6	Per Knut Aaland	M	31	188	75
12	6	Per Knut Aaland	M	31	188	75
13	6	Per Knut Aaland	M	31	188	75
14	6	Per Knut Aaland	M	31	188	75
15	6	Per Knut Aaland	M	33	188	75
16	6	Per Knut Aaland	M	33	188	75

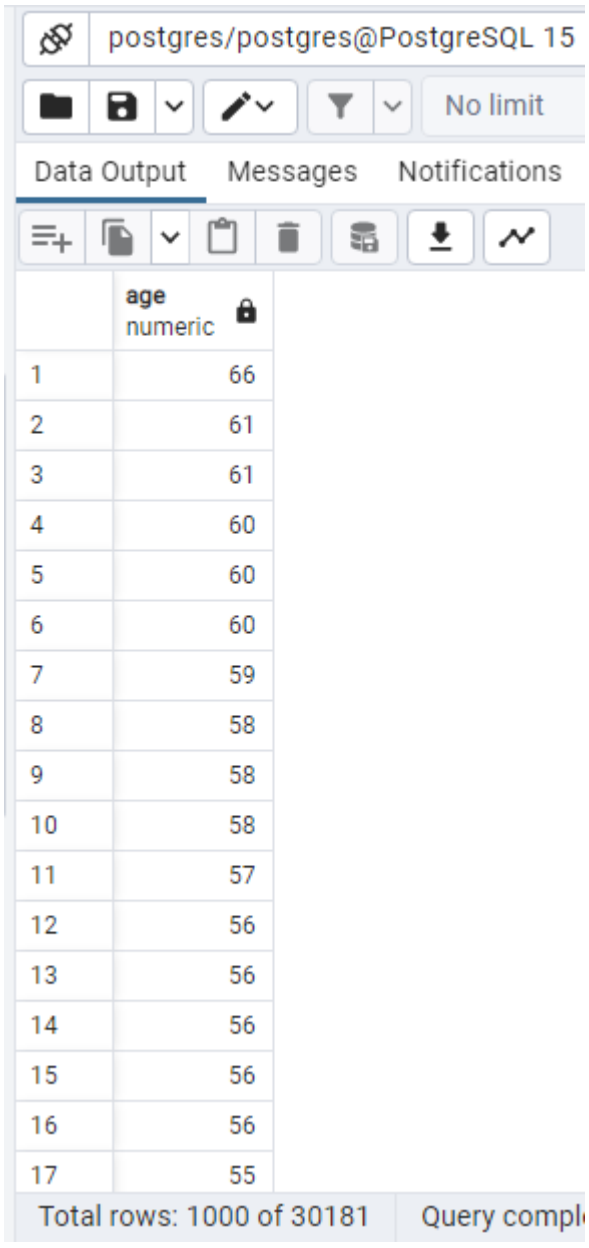
--Total rows: 1000 of 271116 –

SELECT Age FROM public.athlete_events
ORDER BY Age;



--Minumum age is 10 years--

SELECT Age FROM public.athlete_events
ORDER BY Age DESC;



– Maximum age is 66--
--My client should think of building facilities suitable to people in age range 10-66 age--

--minimum height--

SELECT Height FROM public.athlete_events
ORDER BY Height;

Data Output | Messages | Notificatio

	height character varying (10)
1	127
2	127
3	127
4	127
5	127
6	127
7	127
8	128
9	130
10	130
11	131
12	131
13	132
14	132
15	132
16	132
17	132

Total rows: 1000 of 271116 | Query

--minimum height is 127--

SELECT Height FROM public.athlete_events
ORDER BY Height DESC;

postgres/postgres@Postgr

Data Output | Messages | Notifi

	height numeric (8,4)
1	223.0000
2	223.0000
3	223.0000
4	223.0000
5	220.0000
6	220.0000
7	220.0000
8	219.0000
9	218.0000
10	218.0000
11	218.0000
12	218.0000
13	218.0000
14	218.0000
15	217.0000
16	217.0000
17	217.0000

--Maximum height is 223.00--
--My client should build facilities , Cloth and other staff etc. which can accomadate people with height range 127- 223--

SELECT Year FROM public.athlete_events
ORDER BY Year;

Data Output | Messages | Notifications

	year character varying (100)
1	1896
2	1896
3	1896
4	1896
5	1896
6	1896
7	1896
8	1896
9	1896
10	1896
11	1896
12	1896
13	1896
14	1896
15	1896
16	1896
17	1896

Total rows: 1000 of 271116 | Query comp

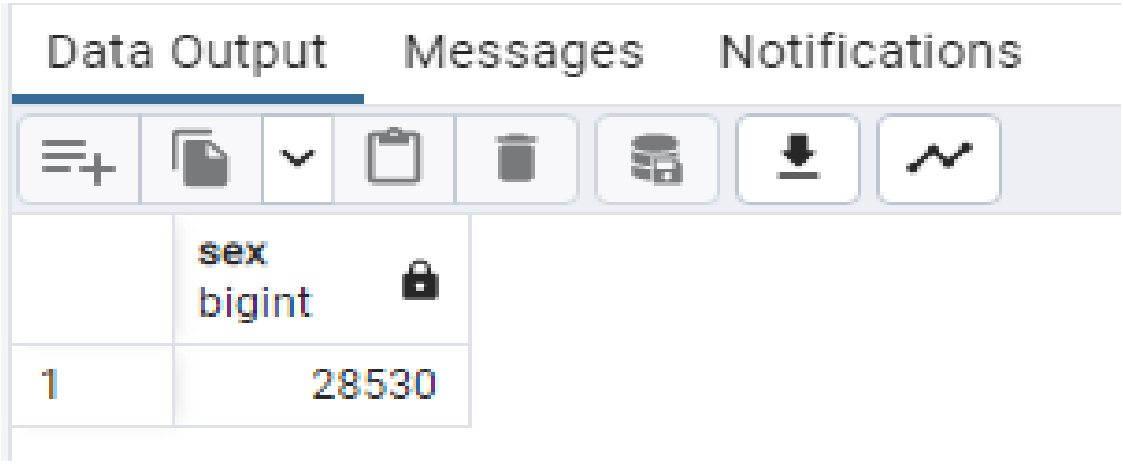
--Minimum year is 1896--
SELECT Year FROM public.athlete_events
ORDER BY Year DESC;
--Max year 2016--
– my client can explore new business opportunities by comparing staff from 1896 to 2016--

--Medal query tool--

DELETE FROM athlete_events
WHERE medal = 'NA';

--DELETE 231333 Query returned successfully in 1 secs 696 msec. –

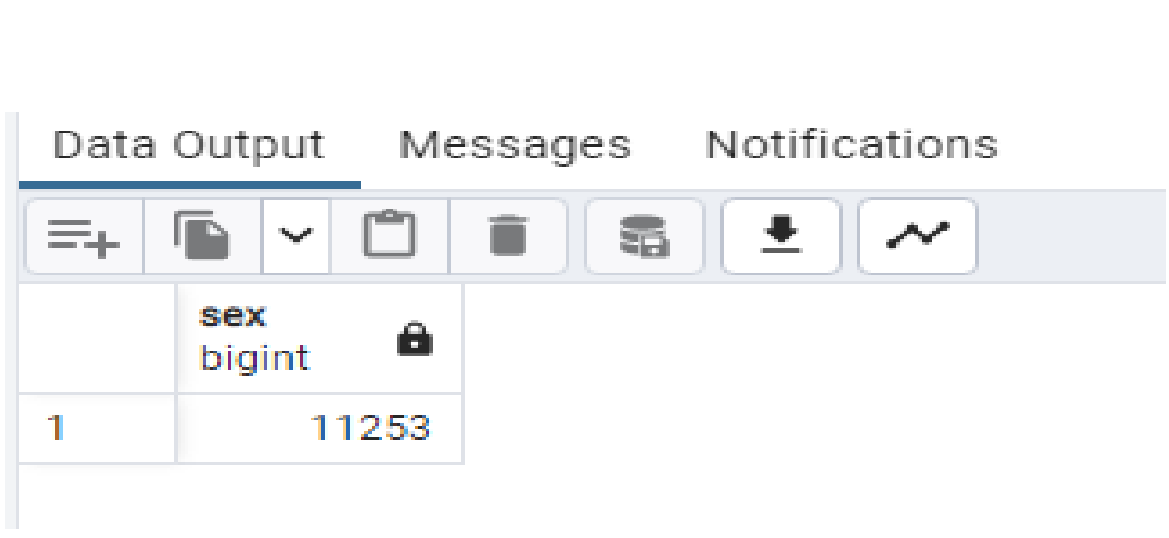
```
SELECT COUNT (*)Sex
  FROM  public.athlete_events
  WHERE Sex= 'M'
  ORDER BY Sex;
```



--28530--

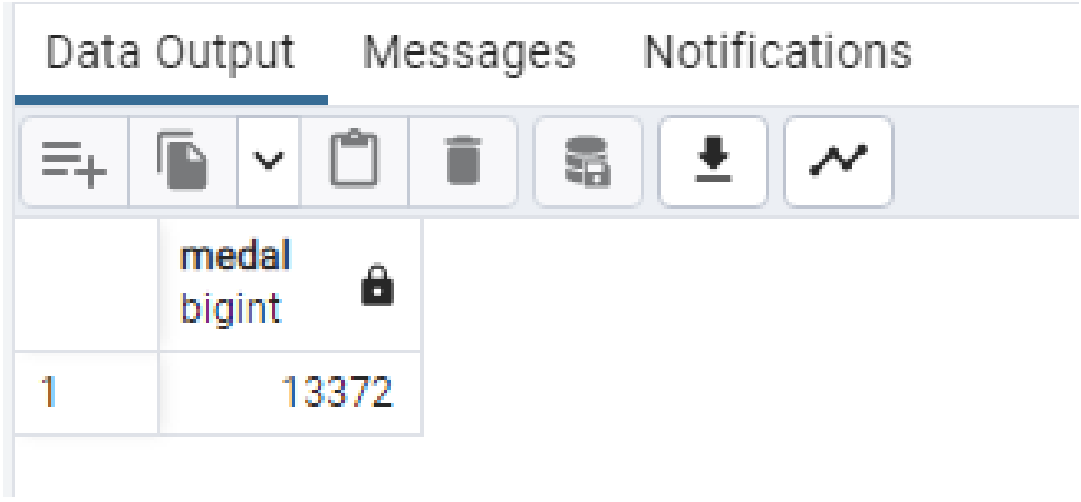
```
SELECT COUNT (*)Sex
  FROM  public.athlete_events
  WHERE Sex= 'F'
  ORDER BY Sex;
```

--11253 Female--

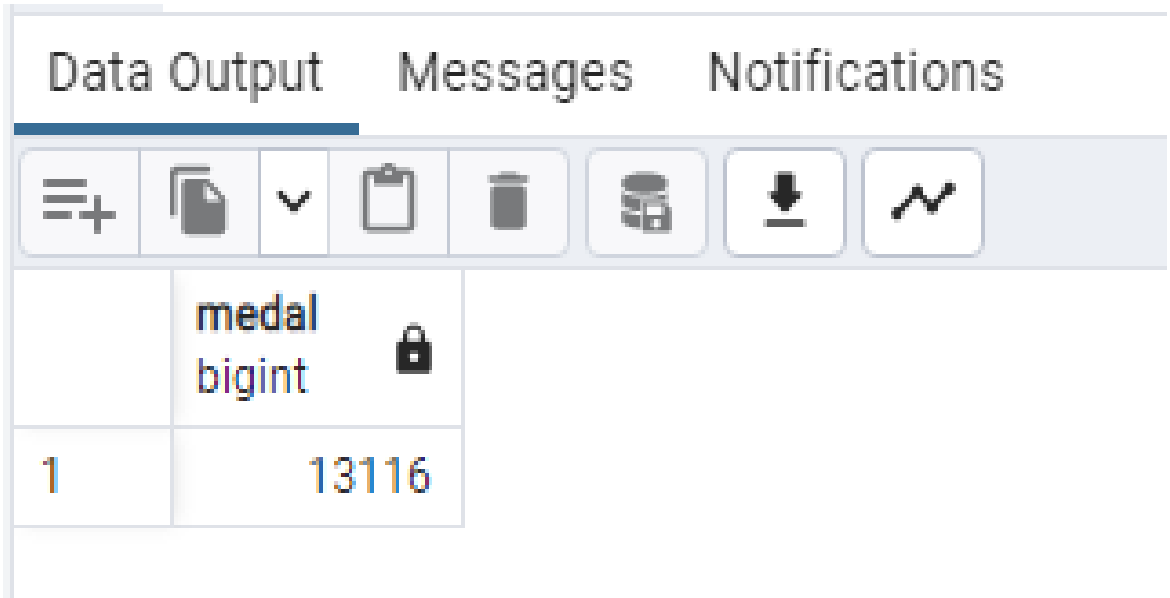



```
SELECT COUNT (*)medal
  FROM  public.athlete_events
  WHERE medal= 'Gold';
```

--Gold medal is 13372--



SELECT COUNT (*)medal
FROM public.athlete_events
WHERE medal= 'Silver';



-13116 Silver--

SELECT COUNT (*)medal
FROM public.athlete_events
WHERE medal= 'Bronze';

Data Output | Messages | Notifications

	medal bigint
1	13295

--13295 Bronze--

```
SELECT COUNT (*)medal,Sex
 FROM  public.athlete_events
 GROUP BY Sex;
```

--Did the learner create 2-3 questions that they want to answer with the data?--

I want to compary Male and Female Participants

Data Output | Messages | Notifications

	medal bigint	sex character varying (100)
1	28530	M
2	11253	F

```
SELECT * FROM athlete_events
WHERE medal = (
   SELECT MAX (medal)
   FROM athlete_events);
```

No limit

Data Output | Messages | Notifications

	id integer	name character varying (200)	sex character varying (100)	age character varying (100)	height character varying (10)	weight charac
1	20	Kjetil Andr Aamodt	M	22	176	85
2	20	Kjetil Andr Aamodt	M	22	176	85
3	25	Alf Lied Aanning	M	24	NA	NA
4	30	Pepijn Aardewijn	M	26	189	72
5	38	Karl Jan Aas	M	20	NA	NA
6	67	Mariya Vasilyevna Abakumova (-Tarabina)	F	22	179	80
7	73	Luc Abalo	M	31	182	86
8	84	Stephen Anthony Abas	M	26	165	55
9	90	Tamila Rashidovna Abasova	F	21	163	60
10	99	Pter Abay	M	30	181	79
11	107	Carmine Abbagnale	M	30	182	90
12	108	Giuseppe Abbagnale	M	33	187	97
13	153	Monica Cecilia Abbott	F	23	191	88
14	165	Nia Nicole Abdallah	F	20	175	56
15	298	Muhammad Abdul Rashid	M	23	181	72
16	339	Namiq Yadulla Abdullayev	M	25	167	55

Total rows: 1000 of 13116 | Query complete 00:00:00.207 | Ln 66, Col 1

SELECT COUNT (*) Name
FROM athlete_events
WHERE Name = ('Kjetil Andr Aamodt');

--8--

Data Output | Messages | Notifications

	name bigint
1	8

SELECT *
FROM athlete_events
WHERE Name = ('Kjetil Andr Aamodt');

--Review--

DELETE FROM athlete_events
WHERE Age = 'NA';

--Deleted 732 rows--

```
DELETE 732 Query returned successfully in 157 msec
```

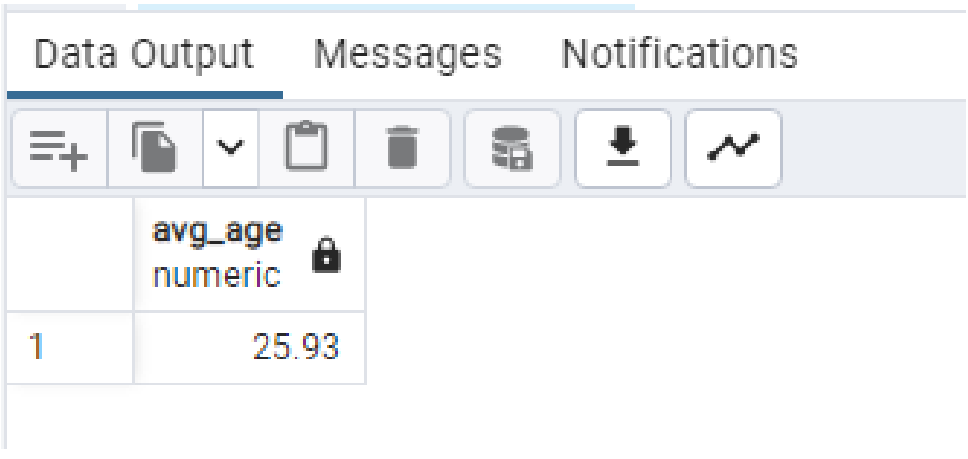
ALTER TABLE athlete_events
ALTER COLUMN Age TYPE INT
USING Age::INT;

```
ALTER TABLE Query returned successfully in 279 msec.
```

--Changed data type of Age to Integer--

SELECT ROUND (AVG(Age), 2) avg_Age
FROM athlete_events;

--Avg 25.93--



--Go to Height query tool and type this code. If you type this code in athlete_events pgadmin will crash--

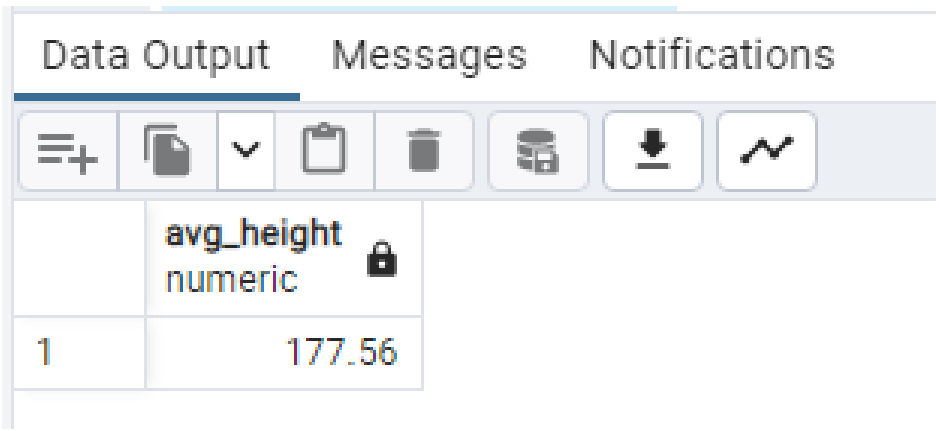
DELETE FROM athlete_events
WHERE Height = 'NA';

--Deleted NA rows—8002--

ALTER TABLE athlete_events
ALTER COLUMN Height TYPE INT
USING Height::INT;

SELECT ROUND (AVG(Height), 2) avg_Height
FROM athlete_events;

--AVG Height is 177.56--

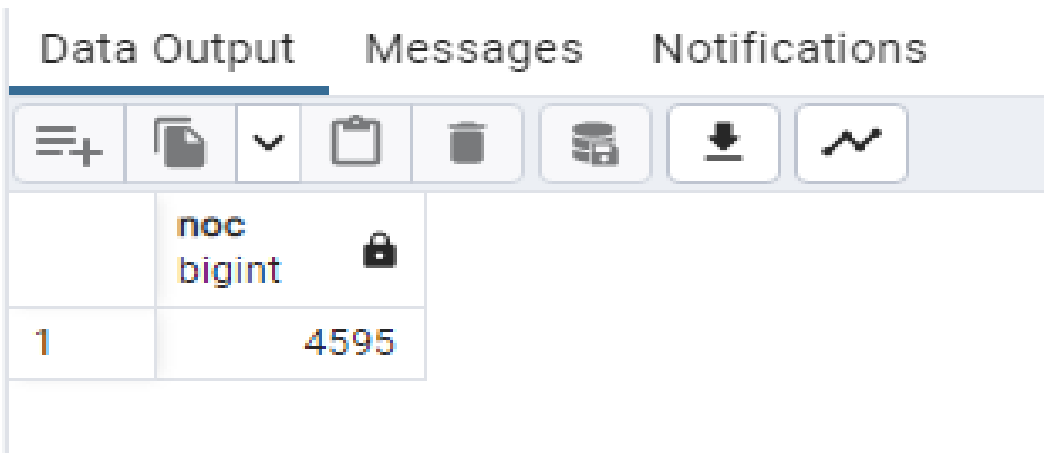


Did the learner create 2-3 questions that they want to answer with the data?

--I want to check NOC of different countries--

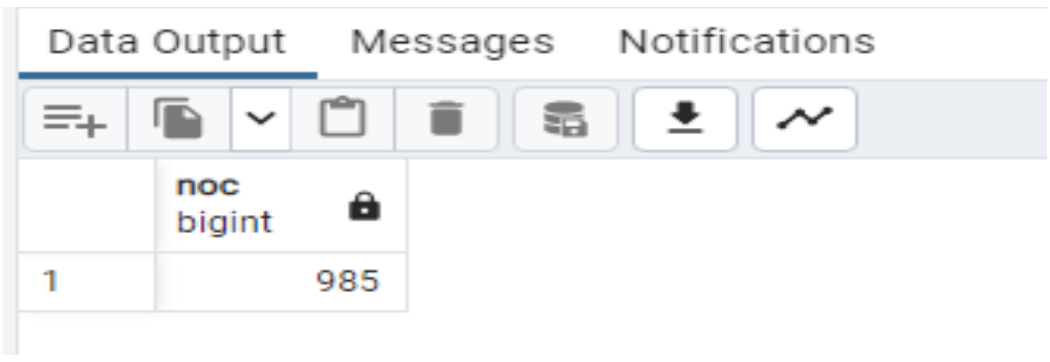
```
SELECT COUNT (*)NOC
   FROM
athlete_events
WHERE NOC='USA'
   GROUP BY NOC;
```

--4595--



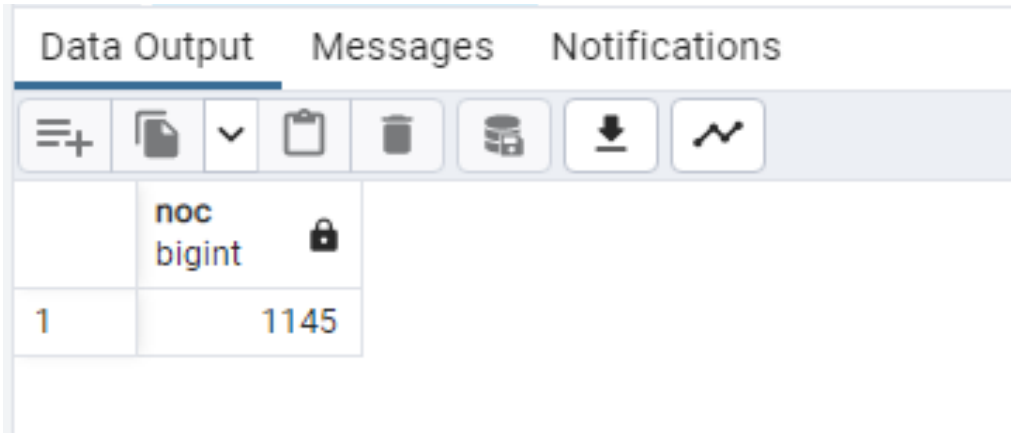

```
SELECT COUNT (*)NOC
   FROM
athlete_events
WHERE NOC='CHN'
   GROUP BY NOC;
```

--985--



```
SELECT COUNT (*)NOC
   FROM
athlete_events
WHERE NOC='RUS'
   GROUP BY NOC;
```

--1145--



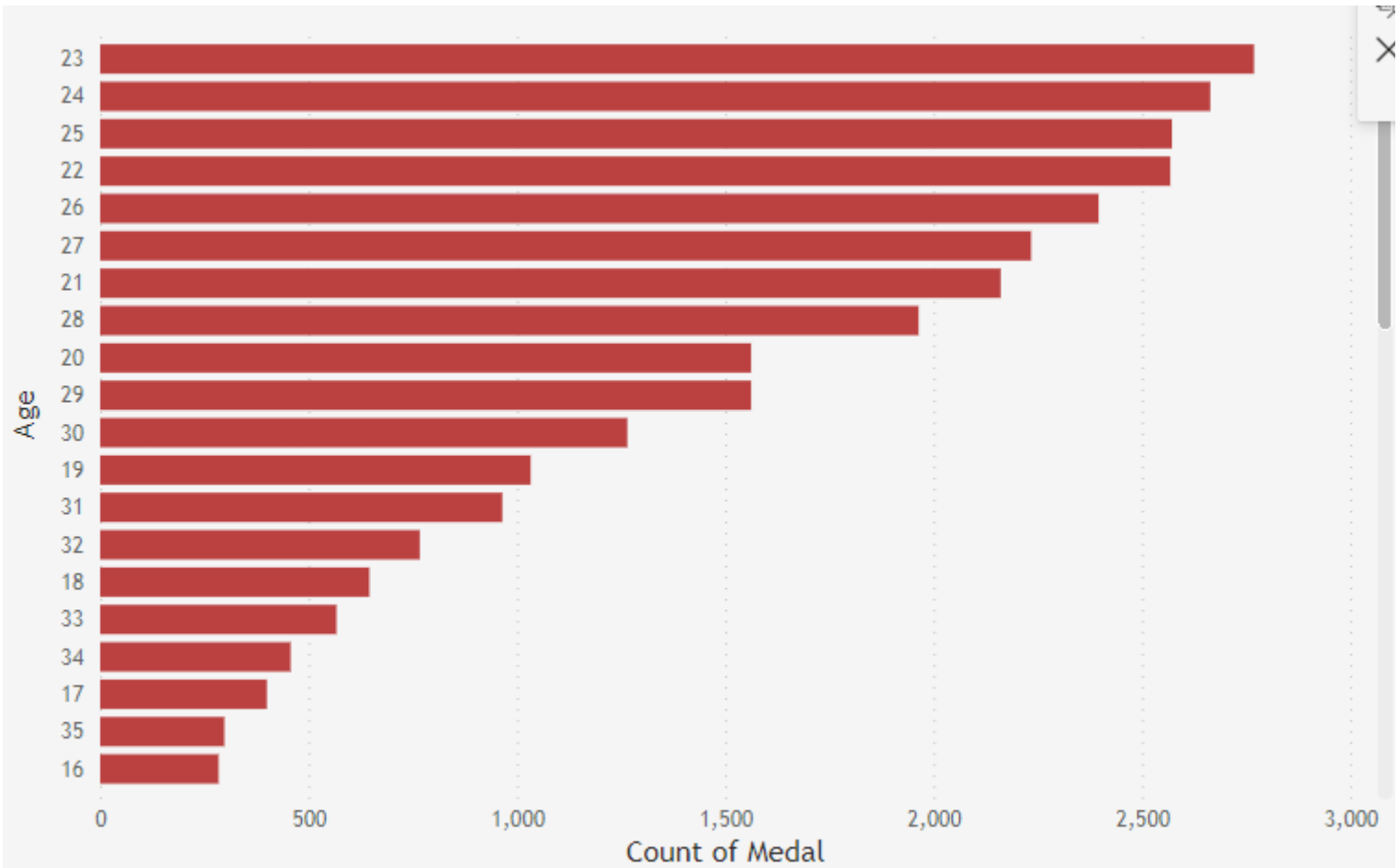

```
ALTER TABLE athlete_events
ALTER COLUMN Weight TYPE NUMERIC(8, 4)
USING Weight::NUMERIC(8, 4);
```

--Link to my you tube working with this data using postgres SQL--

--https://youtu.be/-eggmz6Iiv8--

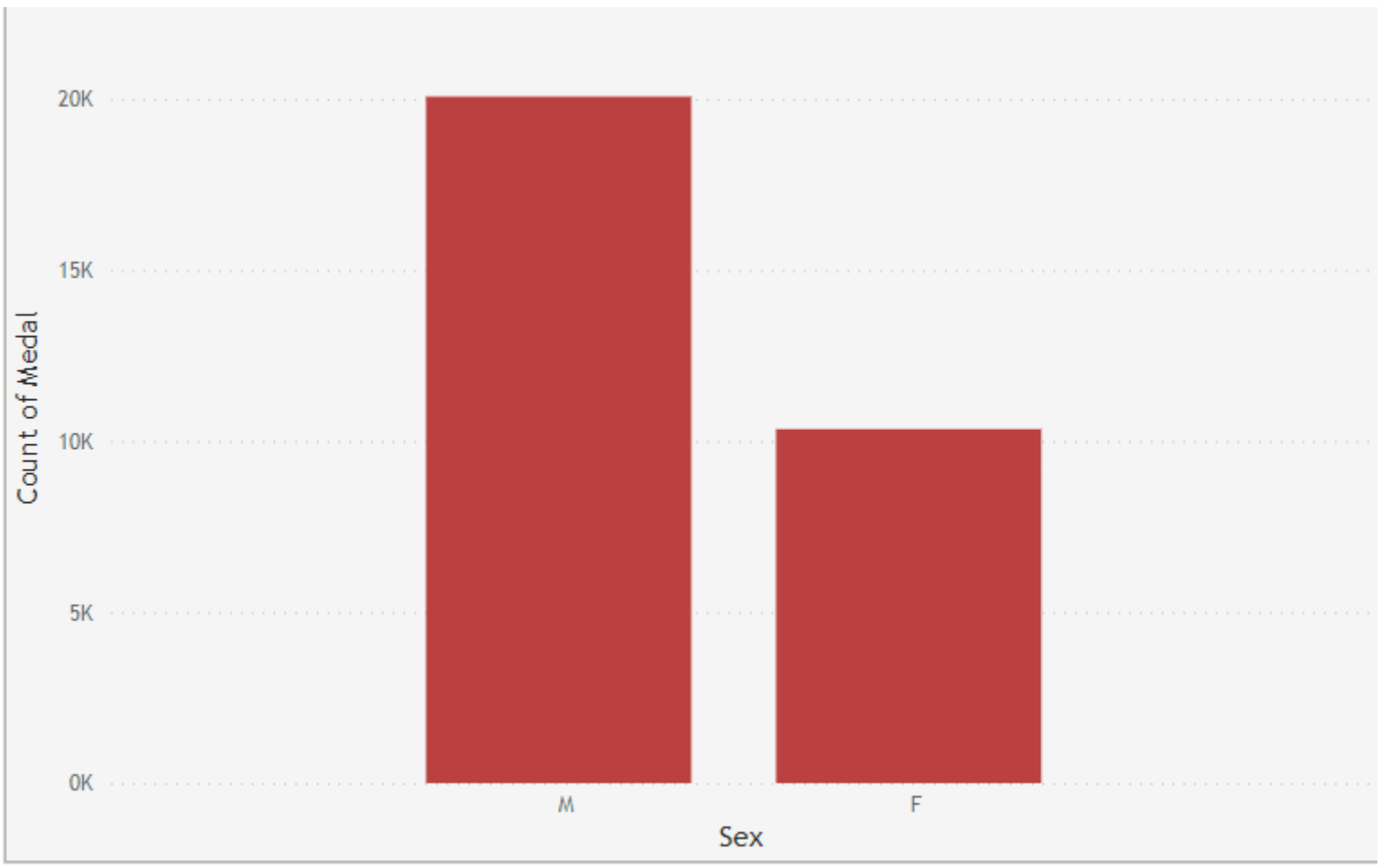
--Link to my you tube working with this data using Microsoft Power bi--
--https://youtu.be/vaMNs0nsmGI--

--Note here I used **Microsoft power Bi** to analyse and create dashboard--



--Did the learner create 2-3 questions that they want to answer with the data?--

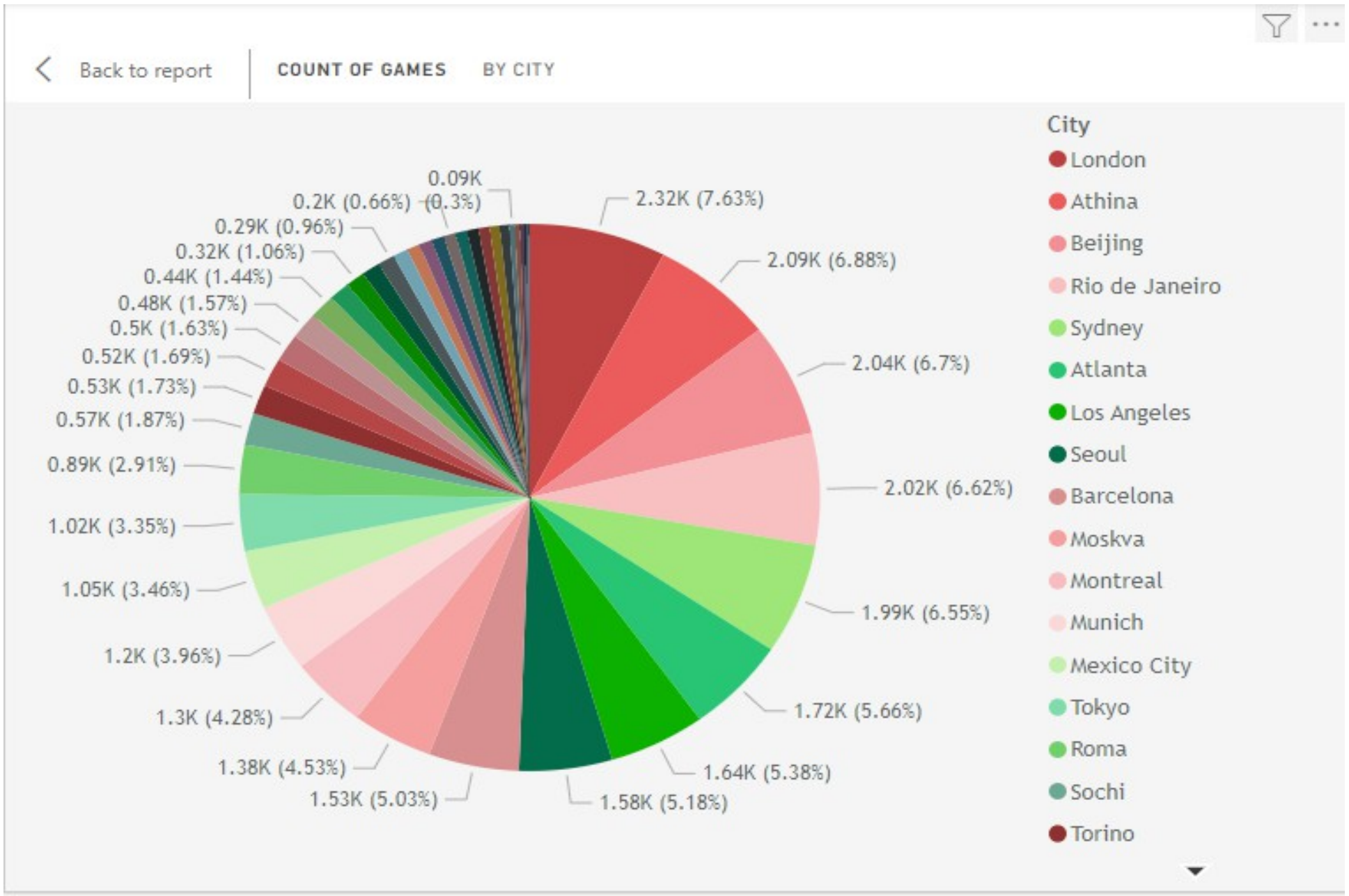
--The above image indicates that most of the participants who was in the age group of 23 won maximum medals--



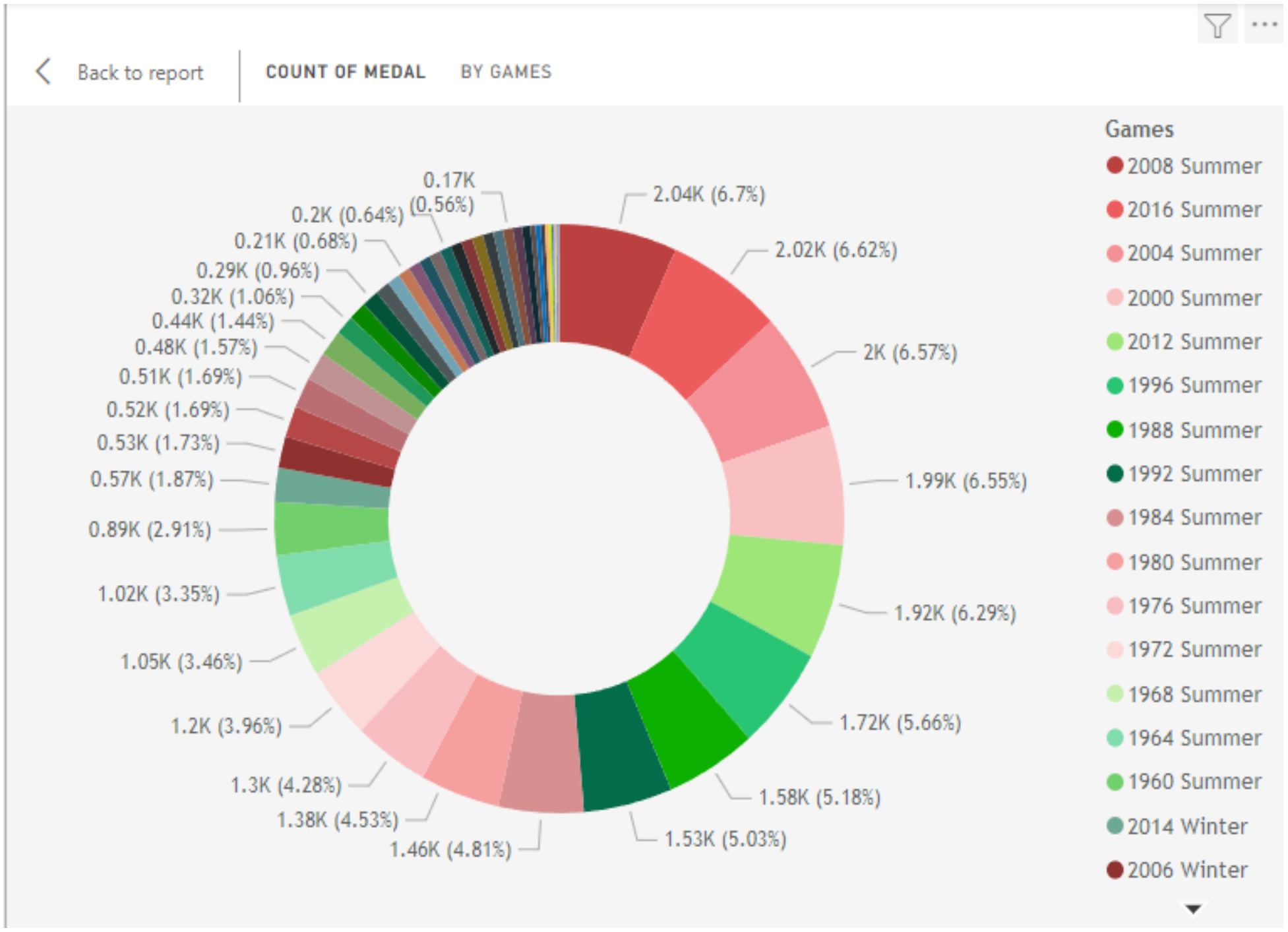
--The above image indicates that male participants earned more medals then female--

Did the learner provide a 5-6 sentence summary describing their project and include who is a potential audience of their findings?

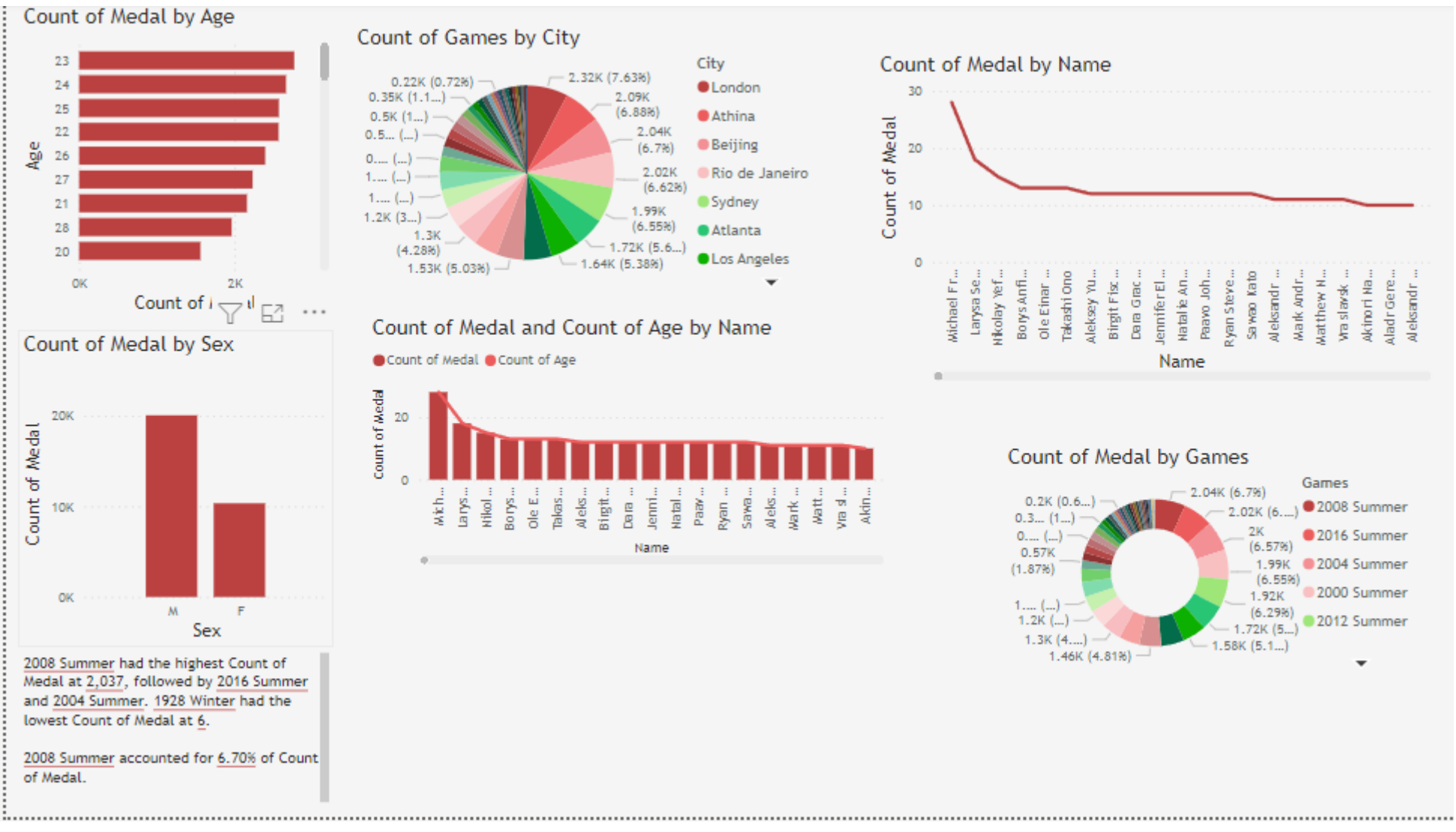
--I created this anlalysis for the people who are interested in sports and love to find the insight for further improvementin their performance individually, or as a team or as a country.--



--The above image indicates that maximum games were in London city more games more income and investment opportunities--



-The above image indicates maximum number of medals were won in 2008 Summer Olympics--



--My power bi Dash Board--

-

- Conclusion the age of the participants who won maximum medals was 23 years. London city had maximum number of games. 2008 Summer olympics had maximum number of games.so more business opportunities--