s Data Analysis on Olympics Dataset

The dataset used to analyze the data is from following

<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results?resource=download>.

The data set contains following data:

**ID** - Unique number for each athlete

**Name** - Athlete's name

**Sex** - M or F

**Age** - Integer

**Height** - In centimeters

**Weight** - In kilograms

**Team** - Team name

**NOC** - National Olympic Committee 3-letter code

**Games** - Year and season

**Year** – Integer

**Season** - Summer or Winter

**City** - Host city

**Sport** - Sport

**Event** - Event

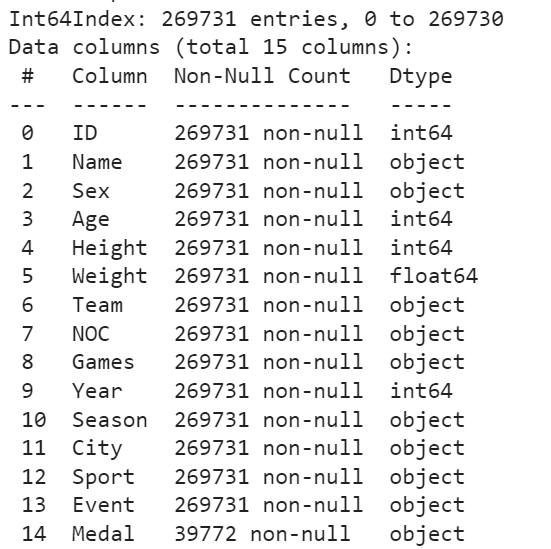
**Medal** - Gold, Silver, Bronze, or NA



This data set is not in a clean form . so, that first i import this data set into the google colab and I removed all duplicate rows that are contained in the data set and some necessary null values are replaced with my data cleaning skills.

Before cleaning it contains

After cleaning it contains



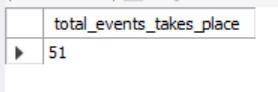
Now the data ready to do queries on the data set

**SQL Queries:**

1. **How many olympics games have been held?**

**Ans)**

**select count(distinct Games) as "total\_events\_takes\_place" from athelate\_events\_cleaned;**

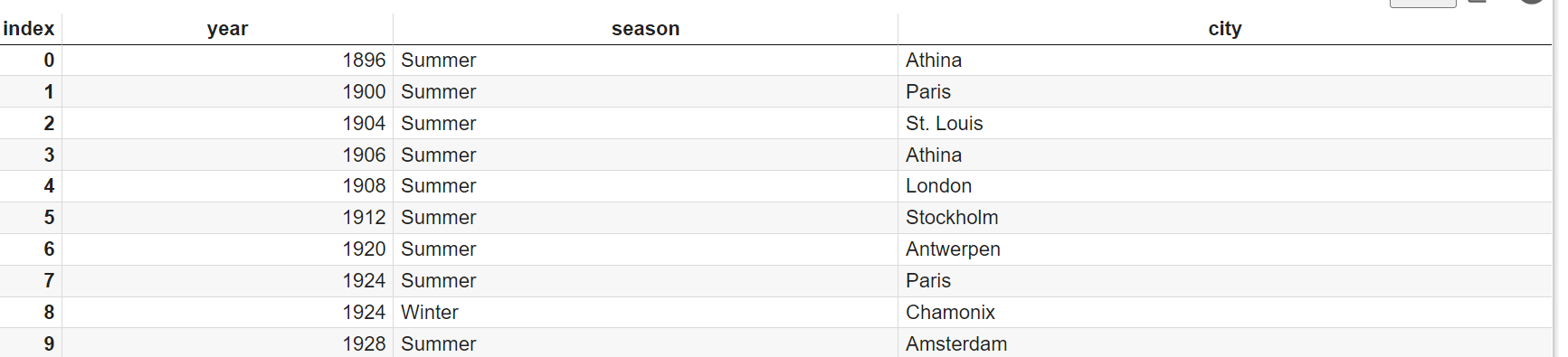
**OUTPUT:** ****

1. **List down all Olympics games held so far.**

**Ans)**

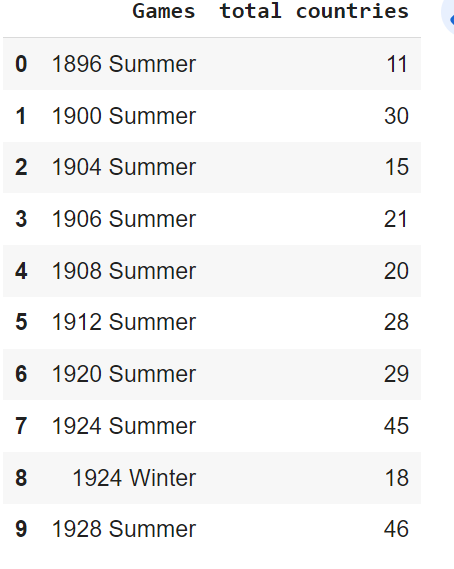
**select distinct year,season,city from athelate\_events\_cleaned order by year limit 269731;**

**output:**

****

1. **Mention the total no of nations who participated in each olympics game?**

**Ans) select Games,count(distinct NOC) as "total countries" from athelate\_events\_cleaned group by Games order by Games;**

****

1. **Which year saw the highest and lowest no of**

**countries participating in Olympics**

**Ans)**

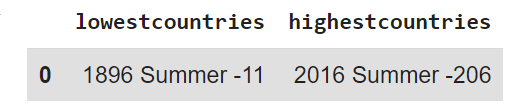
**with total\_countries as(**

**select Games,count(distinct NOC) as "totalcountries" from athelate\_events\_cleaned group by Games**

**order by Games) select concat(first\_value(Games)over(order by totalcountries asc),' -',first\_value(totalcountries)over(order by totalcountries asc)) as "lowestcountries" ,**

**concat(first\_value(Games)over(order by totalcountries desc),' -',first\_value(totalcountries)over(order by totalcountries desc)) as "highestcountries" from total\_countries limit 1**

**output:**



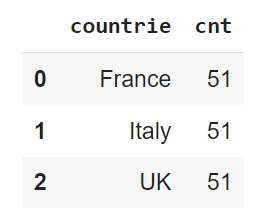
1. **Which nation has participated in all of the olympic games**

**Ans)** **with toatal\_countytries\_participted as (select count(distinct Games) as tot\_countries from athelate\_events\_cleaned),**

**countries\_s as (select distinct Games,region as countrie from athelate\_events\_cleaned ah inner join noc\_regions\_cleaned noc on ah.NOC=noc.NOC group by Games,region),**

**conuntries as (select countrie,count(\*) as countc from countries\_s group by countrie)**

**select co.countrie,co.countc as cnt from conuntries co inner join toatal\_countytries\_participted tc on tc.tot\_countries=co.countc;**



1. **Identify the sport which was played in all summer olympics.**

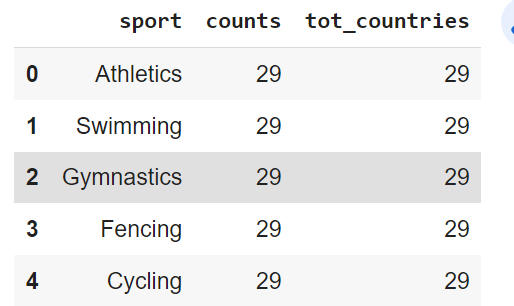
**Ans):**

**with toatal\_countytries\_participted as (select count(distinct Games) as tot\_countries from athelate\_events\_cleaned where Season="Summer")**

**,sports\_s as (select distinct Games,Sport as sport from athelate\_events\_cleaned where Season="Summer" group by Games,Sport),**

**sports as (select sport,count(\*) as counts from sports\_s group by sport)**

**select \* from sports inner join toatal\_countytries\_participted tc on tc.tot\_countries=sports.counts ;**

****

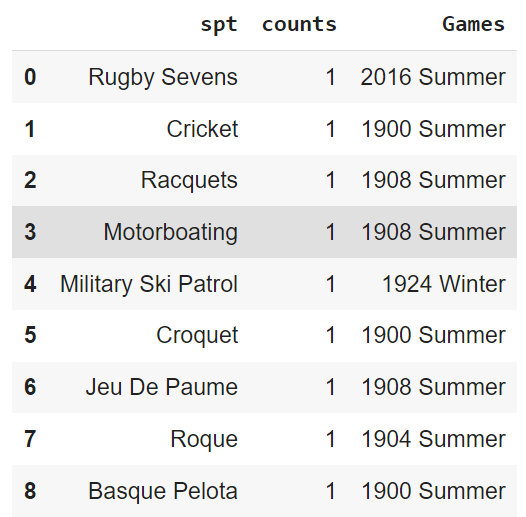
1. **Which Sports were just played only once in the olympics.**

**Ans.**

**with sports\_s as (select distinct Games,Sport as sportss from athelate\_events\_cleaned group by Games,sportss),**

**sports as (select sportss as spt,count(\*) as counts from sports\_s group by spt having count(\*)=1)**

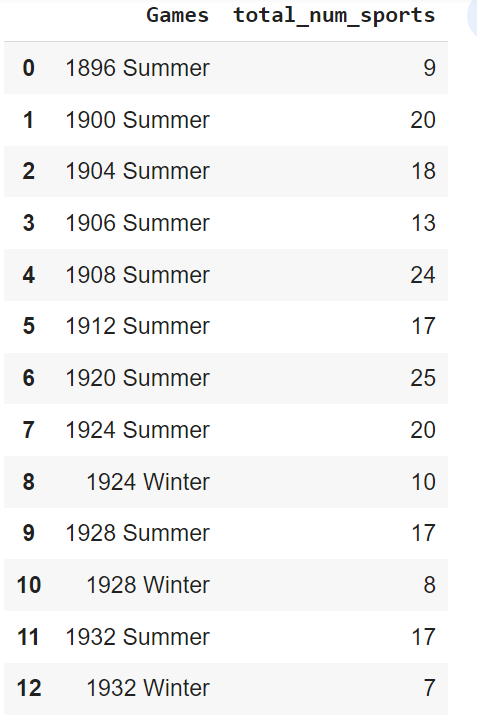
**select spt,counts,Games from sports inner join sports\_s on sports.spt=sports\_s.sportss;**

****

1. **Fetch the total no of sports played in each olympic games.**

**Ans)**

**select Games,count(distinct Sport) as total\_num\_sports from athelate\_events\_cleaned group by Games order by Games;**

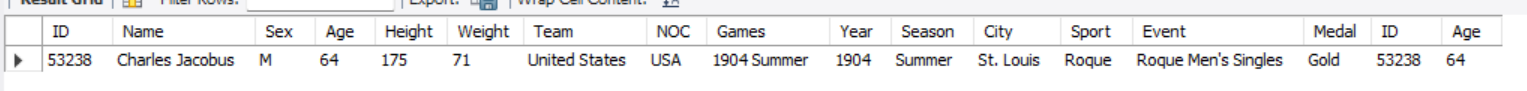


**9.Fetch oldest athletes to win a gold medal**

**Ans) with maxage as (select ID,Age from athelate\_events\_cleaned group by ID,Age,Medal having Medal="Gold" order by Age desc ),**

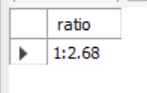
**maxagedata as (select \* from maxage where Age in (select max(Age) from maxage))**

**select \* from athelate\_events\_cleaned ae inner join maxagedata md on ae.ID=md.ID;**

****

1. **. Find the Ratio of male and female athletes participated in all olympic games.**

**Ans) with malstrngrh as ( select count(Sex) as "malesstrength" from athelate\_events\_cleaned where Sex="M"),femalestrngth as (select count(Sex) as "femalesstrength" from athelate\_events\_cleaned where Sex="F")selectconcat(1,':',round((ms.malesstrength/fs.femalesstrength),2)) as "ratio" from malstrngrh ms,femalestrngth fs;**

****

**11. Fetch the top 5 athletes who have won the most gold medals.**

**Ans) select Name,team,count(Medal) "medalsgained" from athelate\_events\_cleaned group by Name,team,Medal having Medal="Gold" order by medalsgained desc limit 5;**

****

**12. Fetch the top 5 athletes who have won the most medals (gold/silver/bronze).**

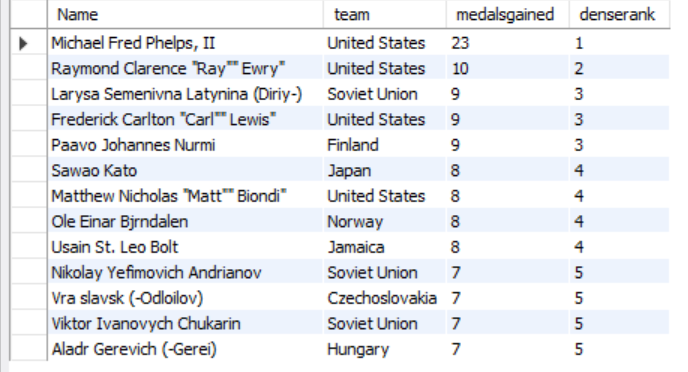
**Ans)**

**with**

**dataall as (select Name,team,count(1) as "medalsgained" from athelate\_events\_cleaned where Medal in ('Gold', 'Silver', 'Bronze') group by Name,team order by medalsgained desc),**

**ranki as (select \*,dense\_rank() over(order by medalsgained desc) as denserank from dataall)**

**select \* from ranki where denserank<=5;**

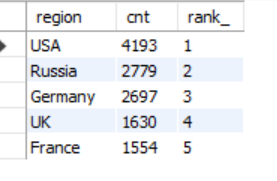


**13. Fetch the top 5 most successful countries in olympics. Success is defined by no of medals won.**

**Ans) with selected\_data as (select region,count(\*) as cnt from athelate\_events\_cleaned ae inner join noc\_regions\_cleaned nr on ae.NOC=nr.NOC where medal in ('Gold', 'Silver', 'Bronze') group by region order by cnt desc**

**),ranked\_data as (select \* ,dense\_rank()over(order by cnt desc) as rank\_ from selected\_data)**

**select \* from ranked\_data where rank\_<=5;**

****

**14. List down total gold, silver and bronze medals won by each country.**

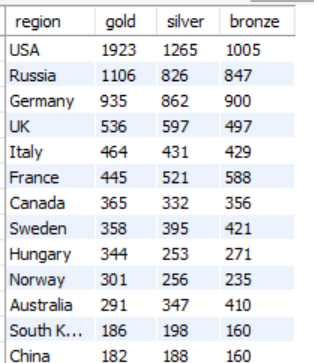
**Ans)select region,sum(case when medal="Gold" then 1 else 0 end) as gold,**

**sum(case when medal="Silver" then 1 else 0 end) as silver,**

**sum(case when medal="Bronze" then 1 else 0 end) as bronze**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC where medal in ('Gold','Silver','Bronze')**

**group by region order by gold desc,silver desc,bronze desc;**

****

**15. List down total gold, silver and bronze medals won by each country corresponding to each olympic games.**

**Ans)**

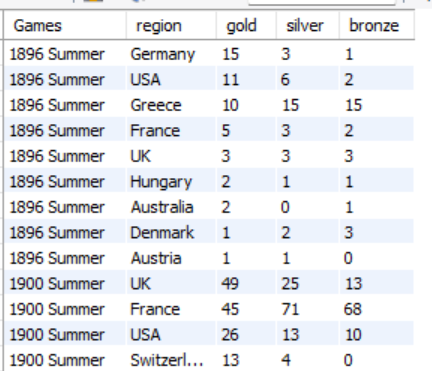
**select Games,region,sum(case when medal="Gold" then 1 else 0 end) as gold,**

**sum(case when medal="Silver" then 1 else 0 end) as silver,**

**sum(case when medal="Bronze" then 1 else 0 end) as bronze**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC where medal in ('Gold','Silver','Bronze')**

**group by Games,region order by Games,gold desc,silver desc,bronze desc;**

****

**16. Identify which country won the most gold, most silver and most bronze medals in each olympic games.**

**Ans) with resdata as(select Games,region,sum(case when medal="Gold" then 1 else 0 end) as gold,**

**sum(case when medal="Silver" then 1 else 0 end) as silver,**

**sum(case when medal="Bronze" then 1 else 0 end) as bronze**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC where medal in ('Gold','Silver','Bronze')**

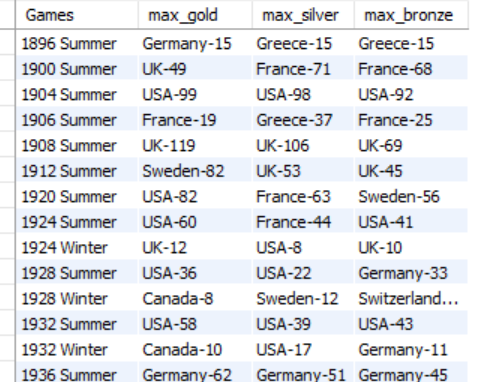
**group by Games,region order by Games,gold desc,silver desc,bronze desc)**

**select distinct Games,concat(first\_value(region)over(partition by Games order by gold desc) ,'-',first\_value(gold)over(partition by Games order by gold desc)) as max\_gold,**

**concat(first\_value(region)over(partition by Games order by silver desc) ,'-',first\_value(silver)over(partition by Games order by silver desc)) as max\_silver,**

**concat(first\_value(region)over(partition by Games order by bronze desc) ,'-',first\_value(bronze)over(partition by Games order by bronze desc)) as max\_bronze**

**from resdata group by Games,region ;**

****

**17. Identify which country won the most gold, most silver, most bronze medals and the most medals in each olympic games.**

**Ans)** **with resdata as(select Games,region,sum(case when medal="Gold" then 1 else 0 end) as gold,**

**sum(case when medal="Silver" then 1 else 0 end) as silver,**

**sum(case when medal="Bronze" then 1 else 0 end) as bronze,count(medal) as tot\_medals**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC where medal in ('Gold','Silver','Bronze')**

**group by Games,region order by Games,gold desc,silver desc,bronze desc)**

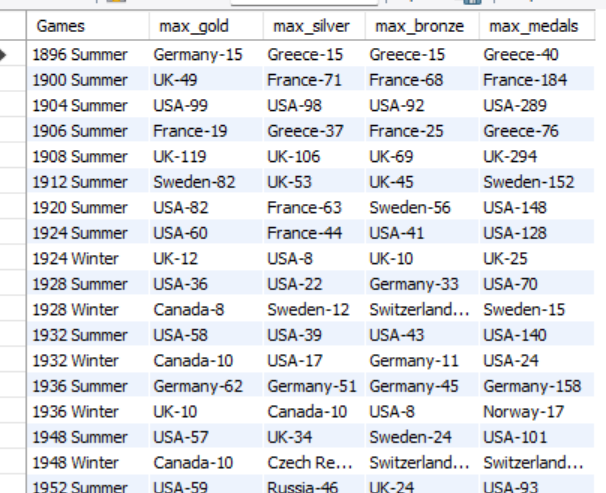
**select distinct Games,concat(first\_value(region)over(partition by Games order by gold desc) ,'-',first\_value(gold)over(partition by Games order by gold desc)) as max\_gold,**

**concat(first\_value(region)over(partition by Games order by silver desc) ,'-',first\_value(silver)over(partition by Games order by silver desc)) as max\_silver,**

**concat(first\_value(region)over(partition by Games order by bronze desc) ,'-',first\_value(bronze)over(partition by Games order by bronze desc)) as max\_bronze,**

**concat(first\_value(region)over(partition by Games order by tot\_medals desc) ,'-',first\_value(tot\_medals)over(partition by Games order by tot\_medals desc)) as max\_medals**

**from resdata group by Games,region ;**

****

**18. Which countries have never won gold medal but have won silver/bronze medals?**

**Ans)**

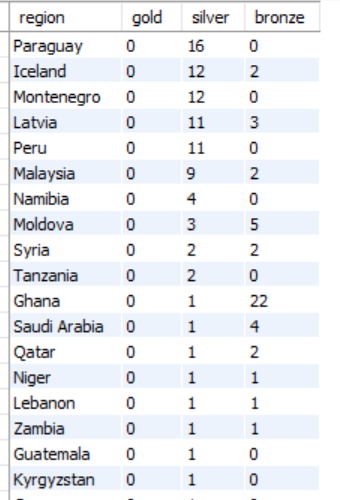
**select region,sum(case when medal="Gold" then 1 else 0 end) as gold,**

**sum(case when medal="Silver" then 1 else 0 end) as silver,**

**sum(case when medal="Bronze" then 1 else 0 end) as bronze**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC where medal in ('Gold','Silver','Bronze')**

**group by region having gold=0 and (silver>0 or bronze>0) order by silver desc,bronze desc;**

****

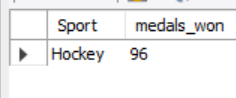
**19. In which Sport/event, India has won highest medals.**

**Ans)**

**select Sport,count(\*) as medals\_won**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC**

**where medal in ('Gold','Silver','Bronze') and region="India" group by sport order by medals\_won desc limit 1;**

****

**20. Break down all olympic games where India won medal for Hockey and how many medals in each olympic games**

**Ans) select Team,Sport,Games,count(1) as medals\_won**

**from athelate\_events\_cleaned ac inner join noc\_regions\_cleaned nc on ac.NOC=nc.NOC**

**where medal in ('Gold','Silver','Bronze') and Team="India" and Sport="Hockey" group by Team,Sport,Games order by medals\_won desc;**

****