Sports Stats Capstone

Table Of Content

Preparing Proposal

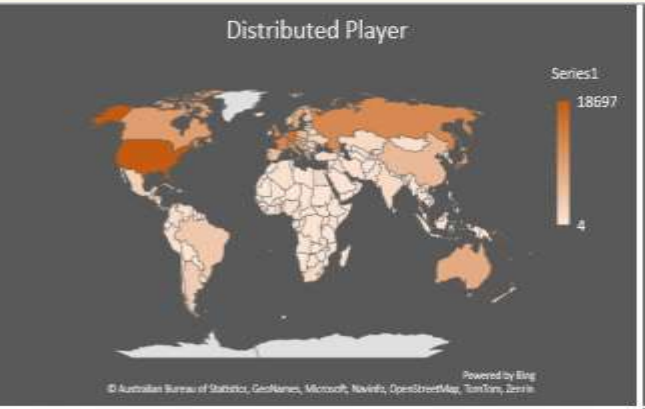
Selecting client

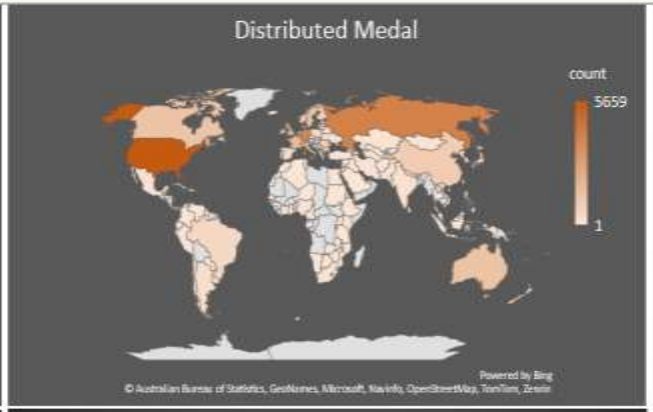
- I choose Sport Stats dataset. The reason why I choose this dataset because I wanna know what factors to winning a medal

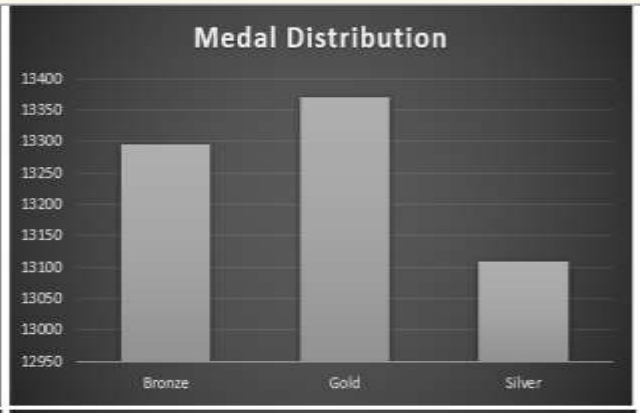
Import And Cleaning Data

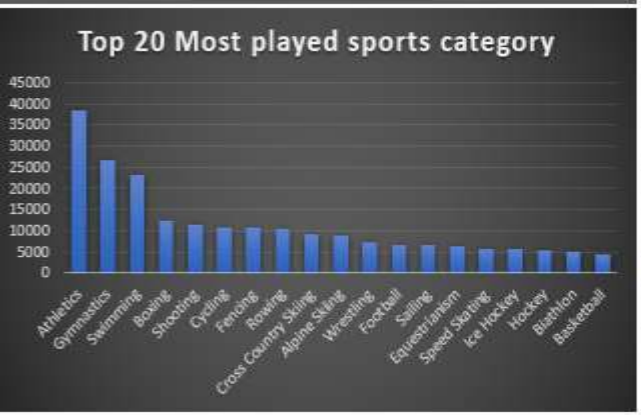
- For importing and splitting colum into different table I use gigasheet and then using online tool to convert .csv file into sql
- Before converting .csv file I clean unnecessary data also some duplicate(same data every rows) data simply using excel
- Since dataset have alot NA value that can effect the results, no need to clean it

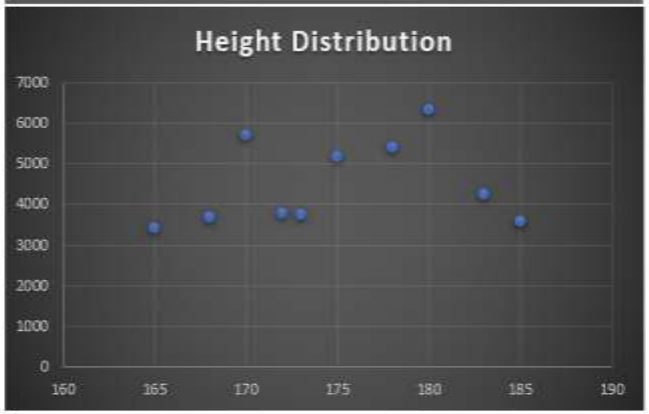
Exploring Data

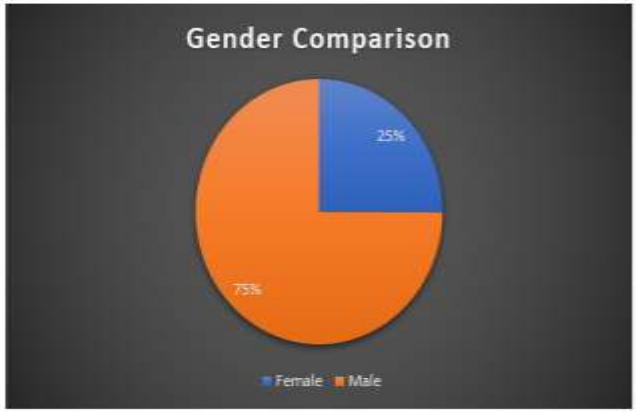




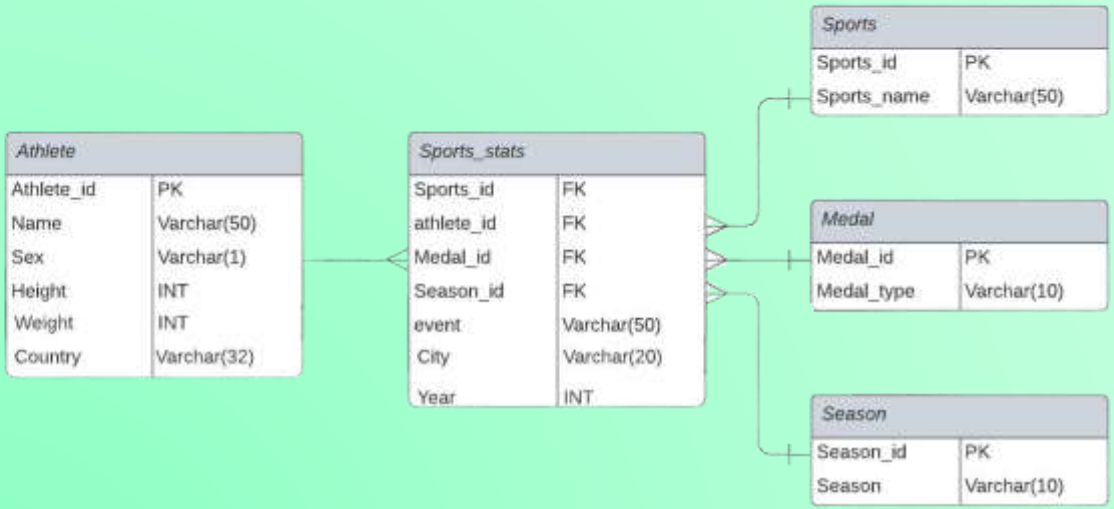








ER Diagram


Sports
Sports_id PK
Sports_name Varchar(50)
Athlete
Athlete_id PK
Name Varchar(50)
Sex Varchar(1)
Height INT
Weight INT
Country Varchar(32)
Sports_stats
Sports_id FK
athlete_id FK
Medal_id FK
Season_id FK
event Varchar(50)
City Varchar(20)
Year INT
Medal
Medal_id PK
Medal_type Varchar(10)
Season
Season_id PK
Season Varchar(10)

Develop Project Proposal

Description

- My goal was to see what factors make winning a medal
- There no limitation about my audience

Questions

- Does the more players mean the more medals?
- Does the taller player, the more medals won?
- What success rate winning medal between male and female?

Hypotesis

- The more player mean having higher chance to win medal
- Taller player having more medal
- Female player having higher success rate to win medal

Approach

- To answering first question can be done by comparing distributed player with distributed medal across country
- For second and third question need to look data in table athlete and compare them

Descriptive Statistics

summary of the different descriptive statistics

The first analyst is to count player been played and getting medal, and grouping by country

```sql
select country,
count(*)
from athlete
group by country

```

	country character varying (32)	count bigint
1	Haiti	76
2	Equatorial Guinea	26
3	Samoa	53
4	Andorra	61
5	Somalia	22
6	Bangladesh	38

```sql
select Country,
count(*)
from Athlete as a
join Sports_stats as ss
on a.athlete_id = ss.athlete_id
where medal_id <> 0
group by country
```

	country character varying (32)	count bigint
5	Argentina	275
6	Armenia	23
7	Australia	1364
8	Austria	460
9	Azerbaijan	52
10	Bahamas	40

summary of the different descriptive statistics

Next is counting medal, grouping by height and sexes

```
1  select height,
2  count(*)
3  from Athlete as a
4  join Sports_stats as ss
5  on a.athlete_id = ss.athlete_id
6  where medal_id <> 0 AND height is not null
7  group by height
8  order by height desc
```

	height integer	count bigint
1	223	7
2	220	3
3	219	1
4	218	6
5	217	6
6	216	8

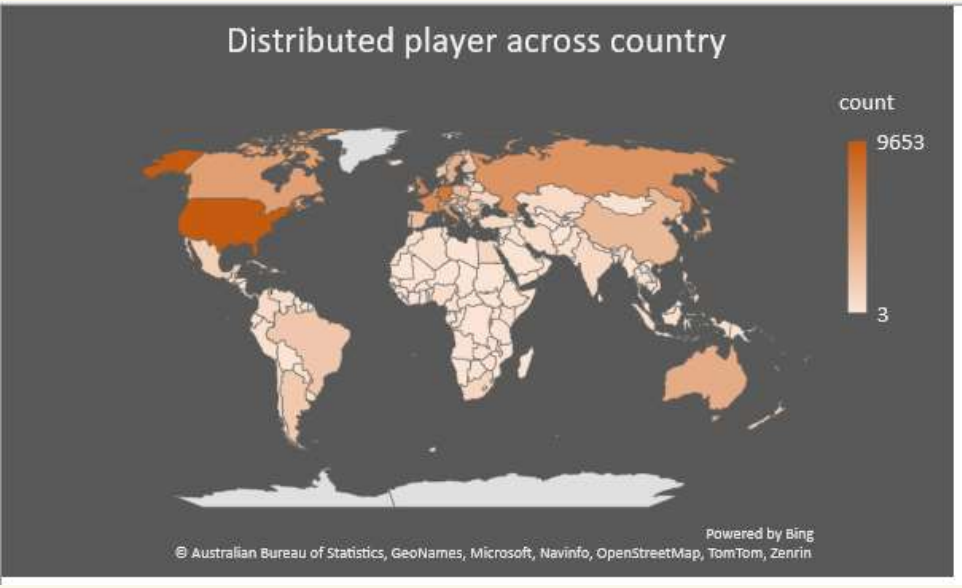
```
1  select sex,
2  count(*)
3  from Athlete as a
4  join Sports_stats as ss
5  on a.athlete_id = ss.athlete_id
6  group by sex
```

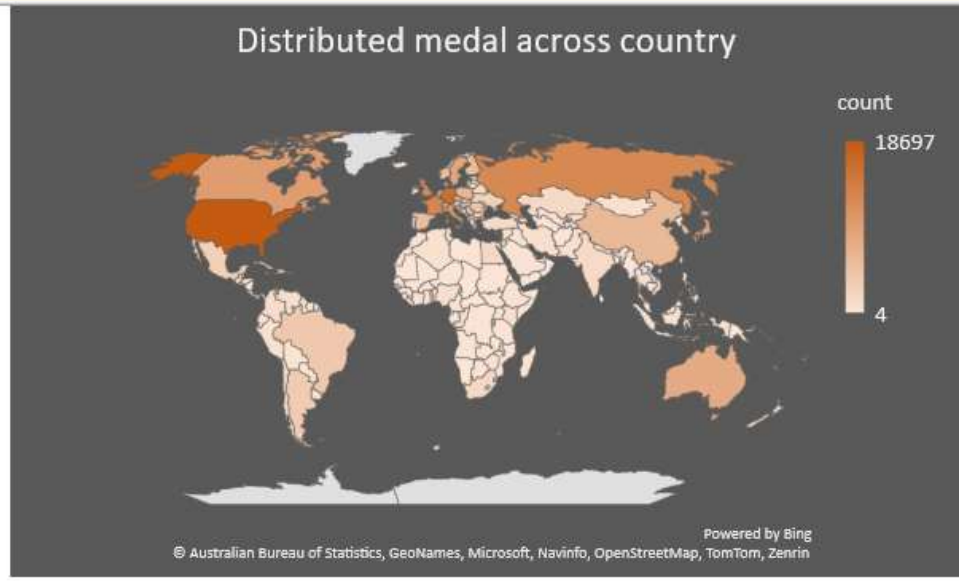
	sex character varying (1)	count bigint
1	F	75485
2	M	197412

```
1  select sex,
2  count(*)
3  from Athlete as a
4  join Sports_stats as ss
5  on a.athlete_id = ss.athlete_id
6  where medal_id <> 0
7  group by sex
```

	sex character varying (1)	count bigint
1	F	11435
2	M	28869

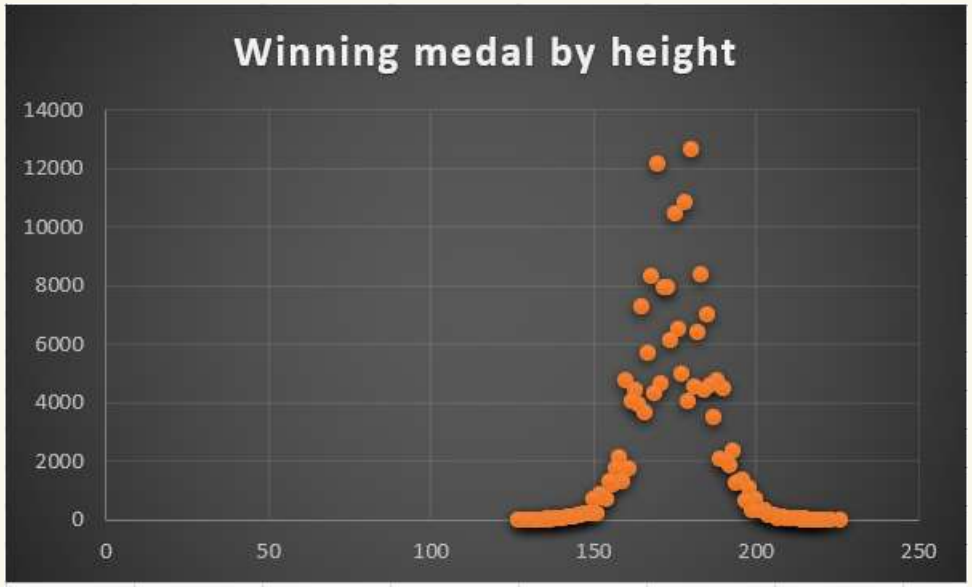
prove or disprove Hypotesis





My first assumption was The more player mean having higher chance to win medal, from map graphic above we can see that they almost indentical meaning country with more player participated has more winning medal so my assumption was right

prove or disprove Hypotesis



My second assumption was the taller player the more medal winning, from graphic above we can see that is not true, but player need to taller around 170-190 to having higher chance of winning medal

prove or disprove Hypotesis

- male player attempt playing in olympics event 197412, winning medal 28869. **Success rate 14.62%**
- female player attempt playing in olympics event 75485, winning medal 11435. **Success rate 15.15%**

My last assumption was female player having more success rate winning medal than male, from data above we can see that is slighty female have more success rate wih female player having 15.15% compare to male player having 14.62% success rate

Additional Question

- What country with higher success rate winning medal ?
- If more taller player doesn't mean more winning medal, how about player with ideal weight? Does they have more chance to winning medal ?

Thanks!