**Capstone Project - Part 1**

**MYSQL**

We will upload our csv file using libraries sqlalchemy, create engine , pandas on MYSQL database. WE have already created schema name ‘capstone’ and we will import our csv data ‘olympic\_data.csv’ using python and jupyter notebook.

Once we have uploaded the data, we will check if the data is being loaded correctly load or not as follows:

SELECT \* FROM capstone1.olympic;

To drop the index column from our loaded data we will use following:

ALTER TABLE `capstone1`.`olympic`

DROP COLUMN `index`;

As we have successfully imported csv file we will explore our dataset using following problem statements:

* **Find the average number of medals won by each country**

To find avg number of medals won by each country we will create a view with groupby clause which will include only two columns count, sum(total\_medal)

* 1. CREATE OR REPLACE VIEW total

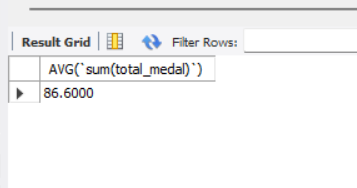
AS (SELECT country, sum(total\_medal)

FROM capstone1.olympic GROUP BY country);

* 1. SELECT \* FROM capstone1.total;

SELECT AVG(`sum(total\_medal)`) FROM capstone1.total;

We will get output of the above as follows. The average number of medals won by each country is 86.6000.



* **Display the countries and the number of gold medals they have won in decreasing order**

We will create a view name gold having two column country, sum(gold\_medal), later we will apply orderby to get the countries and the number of gold medals they have won in decreasing order.

* 1. SELECT \* FROM capstone.olympic;

CREATE OR REPLACE VIEW gold

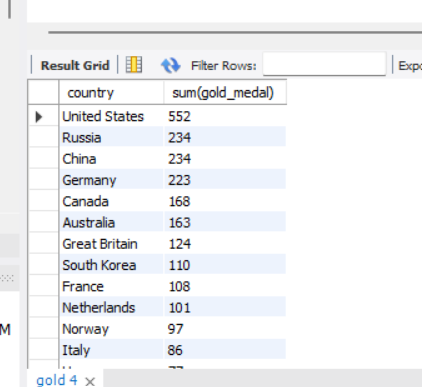
AS (SELECT country, sum(gold\_medal)

FROM capstone.olympic GROUP BY country);

* 1. SELECT \* FROM capstone.gold

ORDER BY `sum(gold\_medal)` DESC;

This will display all 110 countries in descending order as bellow:



* **Display the list of people and the medals they have won in descending order, grouped by their country**

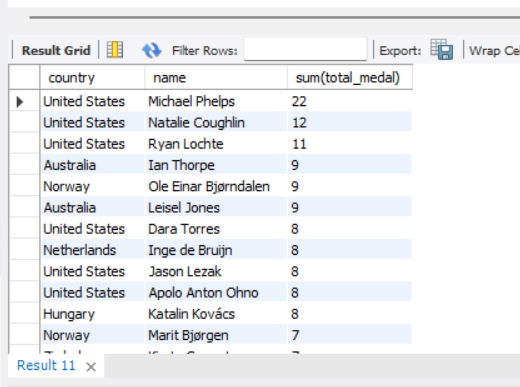
SELECT country, `name`, sum(total\_medal)

FROM capstone.olympic

GROUP BY `name`

ORDER BY sum(total\_medal) DESC;

We will get 6955 rows in output which is the list of people with the medals they have won in descending order, grouped by their country as follows:

****

* **Display the list of people with the medals they have won according to their age**

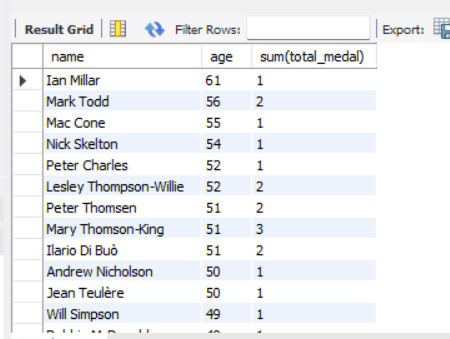
SELECT name, age, sum(total\_medal)

FROM capstone.olympic

GROUP BY name

ORDER BY age DESC;

We will get 6955 rows in output which is the list of people with the medals they have won according to their age.



* **Which country has won the most number of medals (cumulative)**

SELECT country, SUM(total\_medal)

FROM capstone.olympic

GROUP BY country

ORDER BY sum(total\_medal) DESC LIMIT 1;

We will get output as follows which means United States has won the most number of medals i.e. 1312 total medals.

