**By Guneet Kaur**

**ID 300352178**

**Data Visualization CSIS 3860 Section - 070**

**Professor – Rupa Manabala**

*Term Project – Final Report*

**

Fall 2022

Grizli777

Fall

**Dataset Information:**

***Analysis on Olympics Medalists using Tableau***

* *Dataset used for Assignment#1 (Previous Dataset)* :

My previous dataset was collected on the most famous over-the-top (OTT) platform of the era, **Netflix**. The data source was Kaggle.com and the link source is annexed in reference section at the end of the document.

The dataset included information on the *titles* of two *types* of streams (Movies or TV shows) over different *countries*. It had the data of the stream *duration* and *release year* of the movies or shows moreover the *directors* name. Each show had its unique *show id*. Shows were given distinct *ratings* depending upon the access for the type of audience and were *listed* in categorical genres like comedy, action, adventure, horror, dramas and so on.

* *Why I changed the dataset?*

The Netflix dataset **did not include sufficient integral values**. For this project, to calculate trend lines, table calculations and create calculative field, we are required to have at least two or three integer valued variables such as population, number of plants, number of health cases and likewise. Unique Id is not enough to carry out valuable research on a data. This is the reason why I decided to change my dataset.

**Olympics Dataset Introduction:**

* *Current Dataset Description :*

My dataset is on **Olympics Medalists** who won medals for various countries in the world among different years (1896-2008). The data source is *public.tableau.com* and the link source is annexed in the reference section at the end of the document.

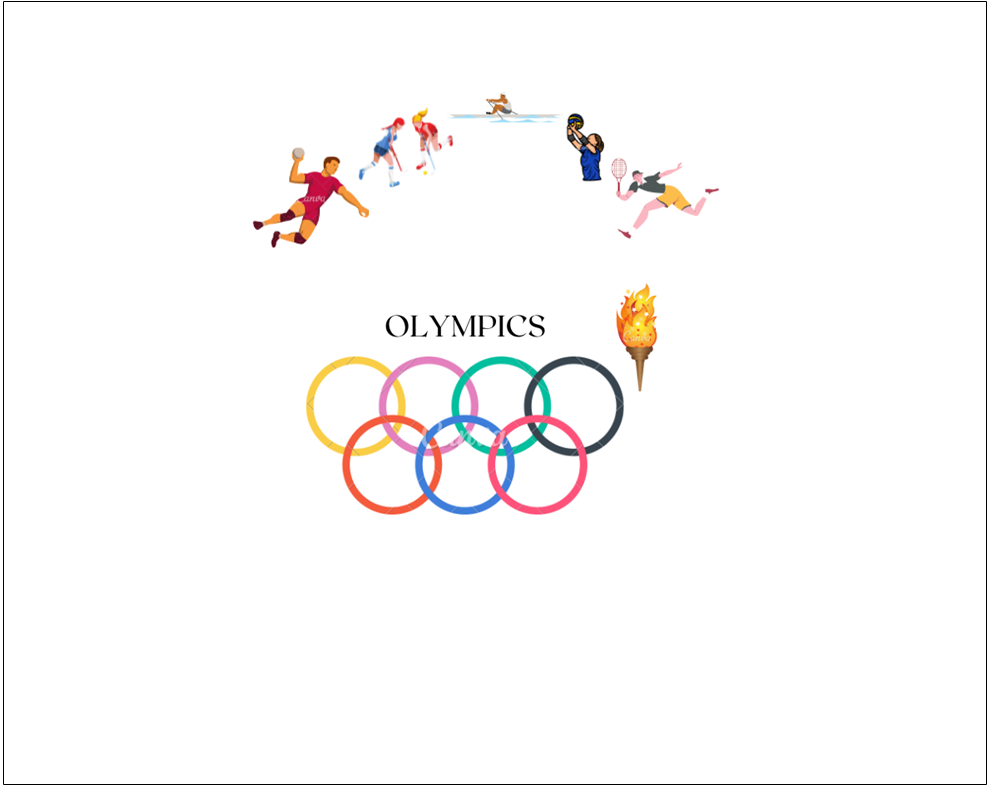
In this dataset I have two excel files, one has the complete dataset and other connects with *country codes* for each *city* with respective *countries*. In the main dataset, we have information of *Edition* (year in which the Olympic game held). Moreover, there is information of *Discipline* that is the game played for example, Swimming, Table Tennis, Diving, Basketball, Hockey and so on. For each Discipline we have *Sport* justified for it, like, for disciplines such as Swimming, Water polo and Diving there is a category of “Aquatics”. Similarly Rhythmic Gymnastics, Trampoline and Artistic Gymnastics are categorized under “Gymnastics”.

Furthermore, we have *Gender* to distinguish players’ gender (Men or Women). Games are associated with separate *Events*, for instance; Cycling can have events of 10km, 100km, 12-hour race, etc. Lastly, there is a column representing the type of *Medal* won by the country in corresponding sport (Gold, Silver, and Bronze).

**Dashboards:**

* *Introductory Dashboard*:

A dashboard representing the dataset topic – Olympics.



* *Dashboard#1: Countries and Athletes who performed the best (1998-2008):*

In my visualization, there is a Tree Map that has the names of top 10 athletes who won the maximum medals in Olympics throughout. This chart also shows the country they represented, the number of medals they won and the sports that they played.

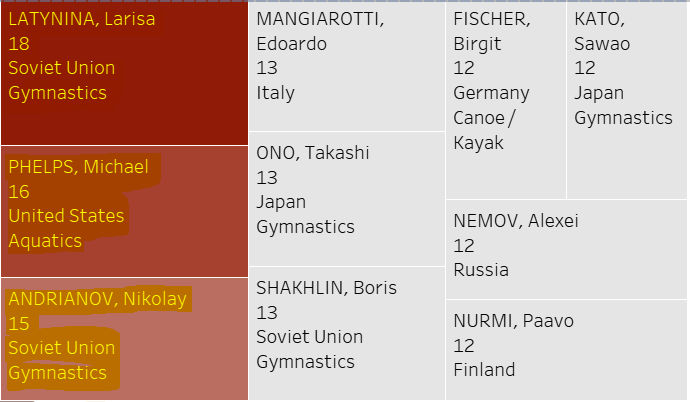
On the other hand, second block is a bubble chart showing the top 10 countries who won the most number of medals.

For instance, according to the bubble chart we see United States has won the maximum number of medals, and tree map shows that Phelps Michael won 16 medals for the United States in the sport Aquatics.

*Question addressed*:

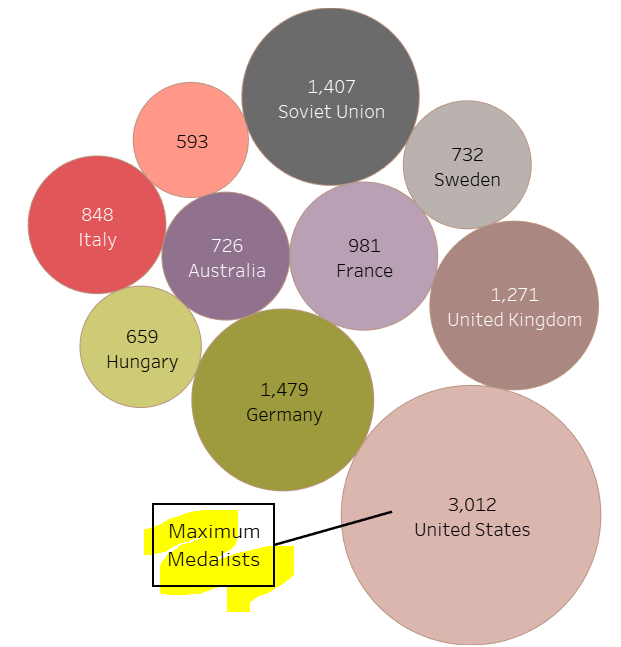
1. Name the medalist who won maximum medals through the Olympics.

* Latynina Larisa won maximum medals in Olympics.



1. Which country outperformed in the Olympics throughout?

* The United States outperformed in Olympics till date.



* *Dashboard#2 - Count of Medals won by Top Athletes*

The second dashboard depicts three tables for all types of medals (gold, silver and bronze) won by top 10 athletes for each category. Whereas, the fourth table represents the athletes who performed best in the most recent Olympics game as per the dataset (Edition 2008). Phelps Michael from United States and Fischer Birgit from Germany were the top winners from recent Olympics game.

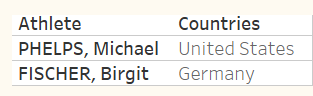
*Question addressed*:

1. Who won the average number of gold medals?

* Nurmi Paavo from Finland and Latynina Larisa from Soviet Union with 9 medals.

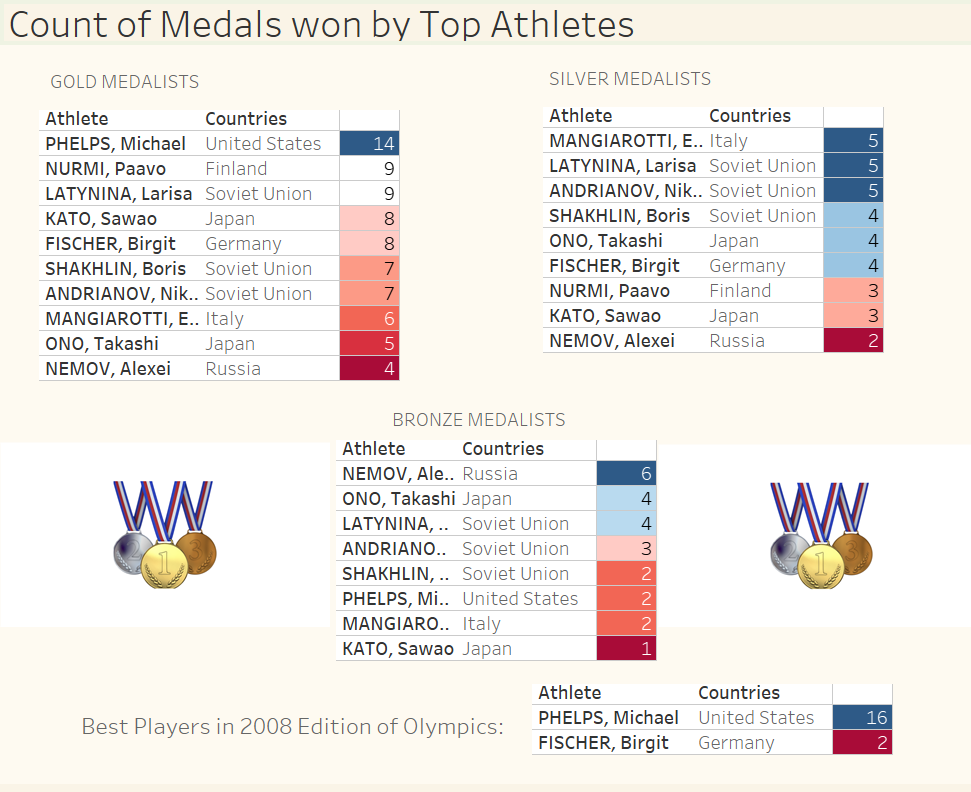
1. Who won the latest Olympics game with highest medals?

* Phelps Michael and Frischer Brigit won the latest Olympics game.



1. What is the highest number of Silver medals won by an athlete?

* 5 medals are the highest of Silver won by an athlete.



**Story:**

* *Which sport was played the most?:*

Box and Whisker Plot is used to study the variation of data set, mostly used with Time Series data.

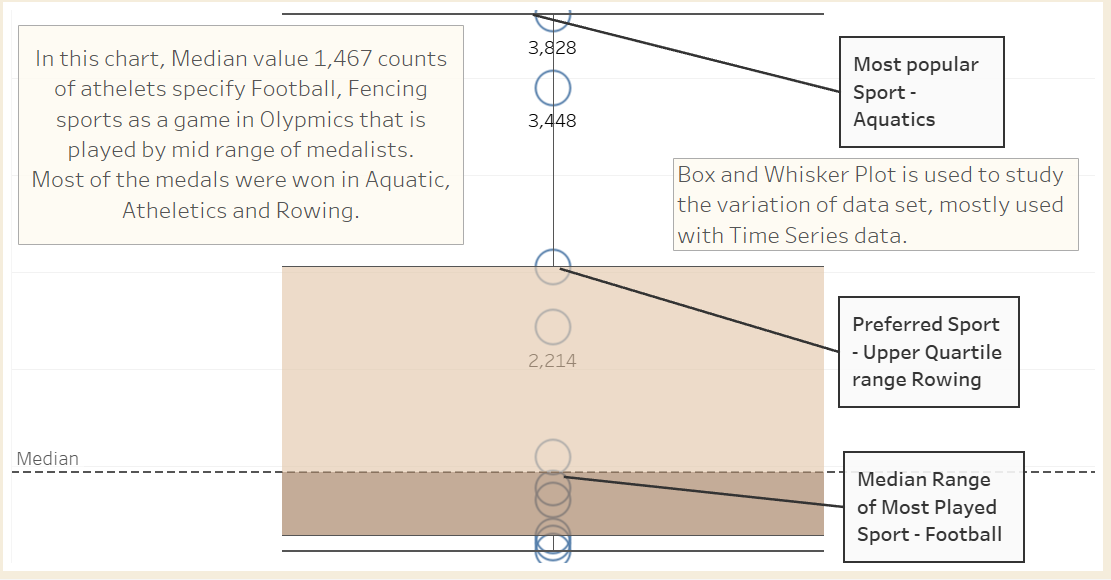
But here, this chart has been used to find out the sports that were played the most, or the sports that were not played much.

For example, people like playing Football, Fencing and Hockey at approximately median (1467) level of all sports preferences.

*Question addressed:*

1. What are the sports that are like even above the level of upper quartile range?

* Aquatics and Athletics are two sports that are liked beyond range.



* *What was the performance of players in each Discipline?:*

Calculated Field - Performance Reviews of Total Medals, Performance of Each Discipline with Review Field. This helps us rank the performance on the following basis.

In the first block, the table shows a column of Performance Reviews based on the number of medals earned per Discipline. A calculated field is made with the count of medals using the script below:

IF COUNT([Medal])<100 THEN 'Highly Unsatisfied'

ELSEIF 100<COUNT([Medal]) AND COUNT([Medal])<500 THEN 'Unsatisfied'

ELSEIF 500<COUNT([Medal]) AND COUNT([Medal])<1000 THEN 'Neutral'

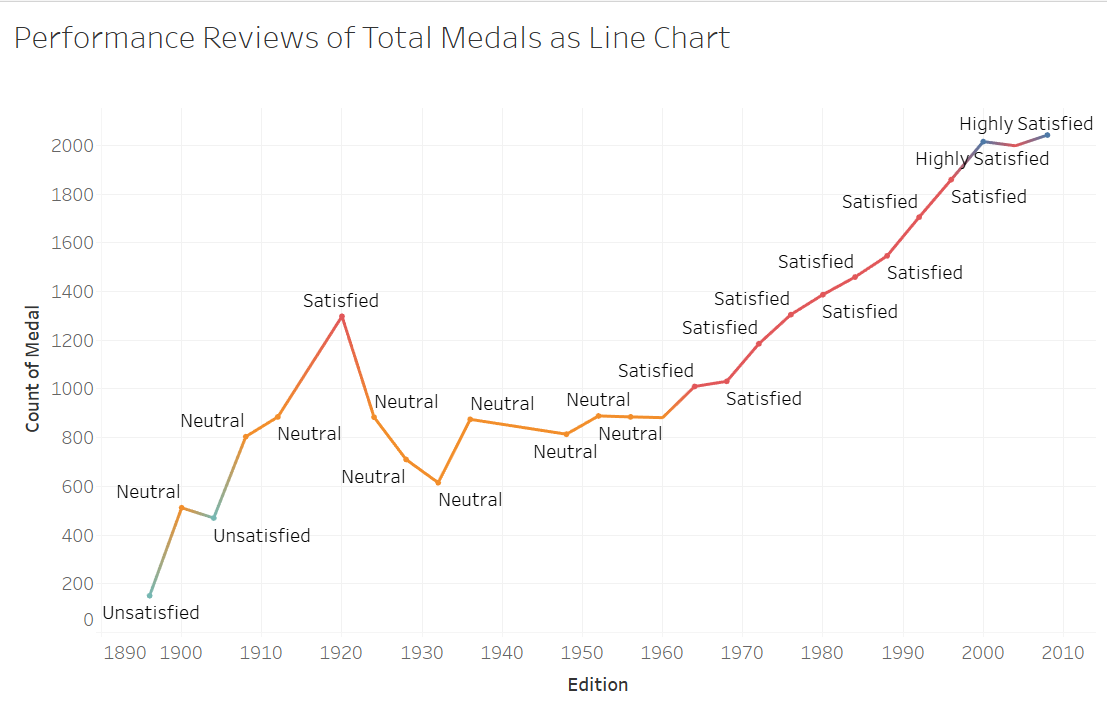
ELSEIF 1000<COUNT([Medal]) AND COUNT([Medal])<2000 THEN 'Satisfied'

ELSE 'Highly Satisfied'

END

From the table, we can observe that Athletics is the Discipline (sport) that received a Highly Satisfied performance with maximum number of medals been distributed, 3,448. In other words, the calculated field represents a performance review in different disciplines with the help of count of medals. Where, the medal count above 2000 in a discipline represents a highly satisfied performance in the discipline. At the end of the table is the grand total that displays the total number of medals being won in all the sports.

*Figure: The line graph shows the density of medals won in all Editions. So to know in which edition the players were the best. The variation from 1900 to 1930 shows a mountain shaped performance among the players.*



Further from the second graph we can say that over the years number of discipline in Olympics have risen from 10 to 41 games.

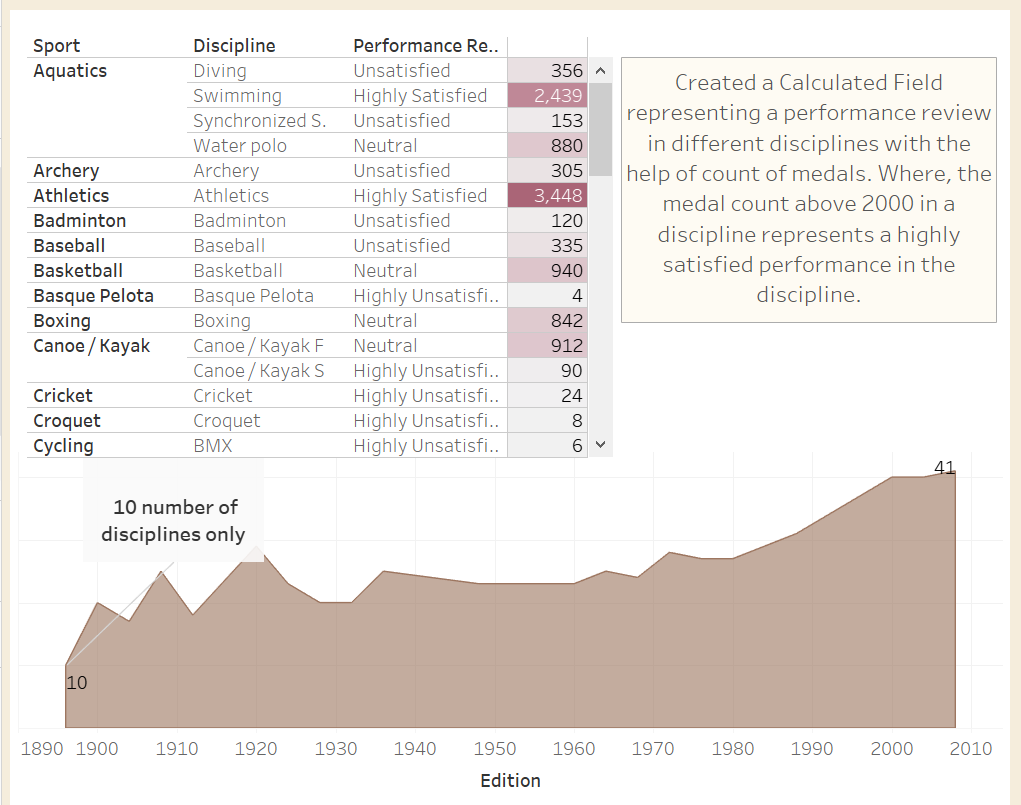
*Question addressed:*

1. What Disciplines are the worst to play?

* Roque and Jeu de Paume are the two disciplines with Highly Unsatisfactory performance in both with only 3 medals.

1. How many Disciplines were initially being played in Olympics?

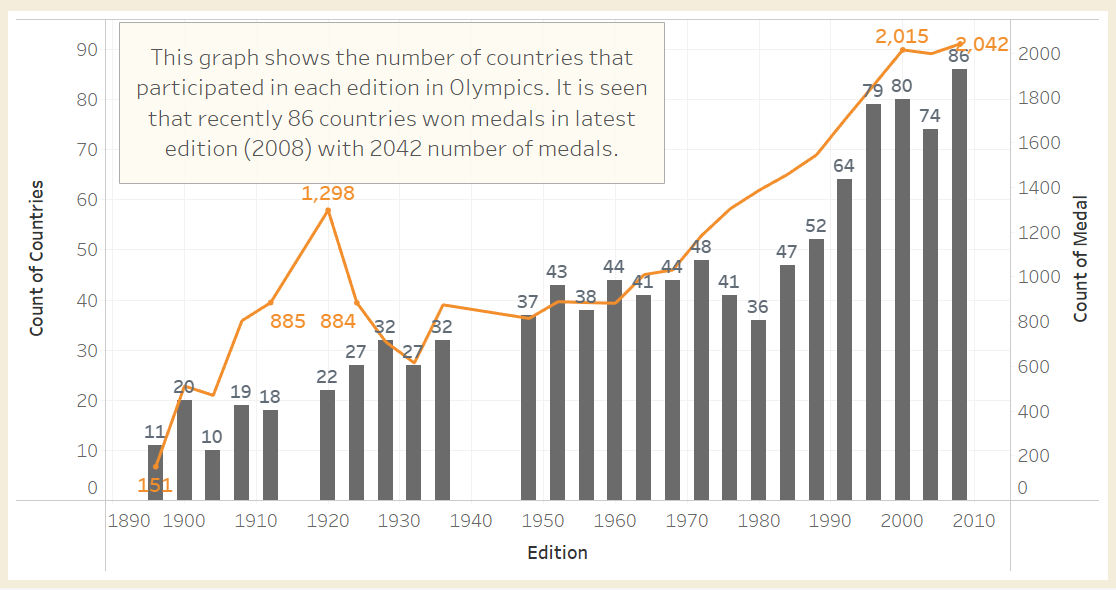
* In 1896, only 10 games were played in Olympics.



* *What was the number of countries participated over the years and count of medals earned?:*

This graph shows the number of countries that participated in each edition in Olympics. It is seen that recently 86 countries won medals in latest edition (2008) with 2042 number of medals.

It is a dual axis bar chart for number of countries participated and counts of medals earned by them over the Editions.

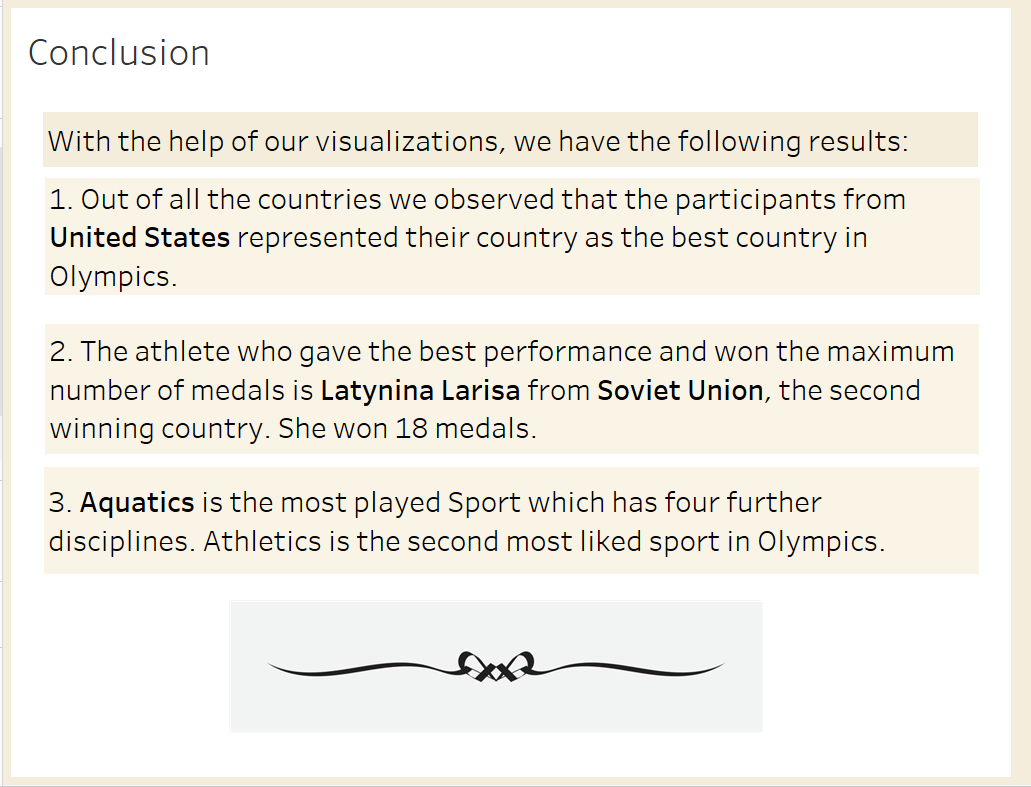


* *What is the conclusion from the analysis?:*

Final Note:

Yes, with the help of our visualizations we were able to address our questions. We have the following results:

1. Out of all the countries we observed that the participants from United States represented their country as the best country in Olympics.
2. The athlete who gave the best performance and won the maximum number of medals is Latynina Larisa from Soviet Union, the second winning country. She won 18 medals.
3. Aquatics is the most played Sport which has four further disciplines. Athletics is the second most liked sport in Olympics.



# Bibliography:

* Previous Dataset from Kaggle *(Data Source) - Netflix data in form of netflix1.csv file* [*https://www.kaggle.com/datasets*](https://www.kaggle.com/datasets)
* Tableau Public *(Data Source) - Olympics data in form of .xlsx file* [*https://public.tableau.com/app/resources/sample-data*](https://public.tableau.com/app/resources/sample-data)