INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2021)

120 YEARS OF OLYMPIC HISTORY: ATHLETES AND RESULTS

Submitted by

Loveleen Verma

Registration No.: 11912122

Programme: B. Tech. (CSE)

Section: K19RC

Course Code: INT217

Under the Guidance of

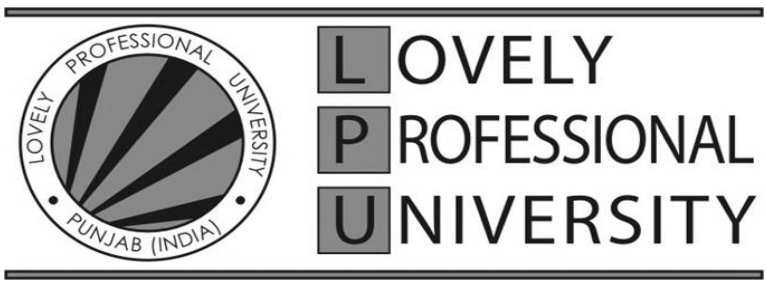
Ms. Harneet Kaur

U. ID: 19371

Discipline of CSE/IT

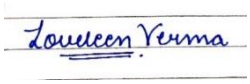
Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara



DECLARATION

I, Loveleen Verma, student of B. Tech. under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 1 December, 2021 **Signature:** 

Registration No. 11912122 **Name of the student:** Loveleen Verma

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher Ms. Harneet Kaur who gave me the golden opportunity to do this project on the topic 120 YEARS OF OLYMPIC HISTORY: ATHLETES AND RESULTS. It helped me in doing a lot of Research and i came to know about a lot of things related to this topic.

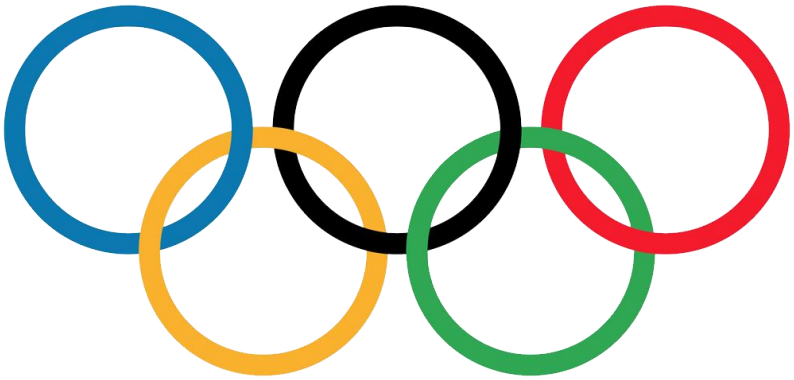
Finally, I would like to thank my parents who have helped me with their valuable suggestions and guidance and have been very helpful in various stages of project completion.

Date: 1 December, 2021

- Loveleen Verma
- 11912122
- B. Tech. (CSE)

TABLE OF CONTENT

INTRODUCTION



The modern **Olympic Games** or **Olympics** are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are normally held every four years, alternating between the summer and Winter Olympics every two years in the four-year period.

The Olympic Games began in ancient Greece. The original Greek games were staged every fourth year for several hundred years, until they were abolished in the early Christian era. The revival of the Olympic Games took place in 1896, and since then they have been staged every fourth year, except during World War I and World War II (1916, 1940, 1944).

Perhaps the basic difference between the ancient and modern Olympics is that the former was the ancient Greeks' way of saluting their gods, whereas the modern Games are a manner of saluting the athletic talents of citizens of all nations. The original Olympics featured competition in music, oratory, and theatre performances as well. The modern Games have a more expansive athletic agenda, and for 2 and a half weeks they are supposed to replace the rancour of international conflict with friendly competition. In recent times, however, that lofty ideal has not always been attained.

HOW HAVE THE OLYMPICS CHANGED OVER TIME?

The evolution of the Olympic Movement during the 20th and 21st centuries has resulted in several changes to the Olympic Games.

1. Number of countries participating

In the first modern Olympic Games in Athens, just 14 countries were represented. At the 2012 London Olympics, 204 countries were represented.

2. The Olympic Torch

The Olympic flame represents the theft of fire from Zeus by Prometheus. The modern flame is now lit by the sun's rays (using a parabolic mirror) at the ruins of the Temple of Hera in Olympia, Greece. But while the Olympic flame was a prominent symbol for the ancient Olympics, it wasn't introduced to the modern games until 1928. And the torch relay, now an important event in the build-up to the games, was an even later addition, introduced at the 1936 Berlin Olympic Games.

3. Olympic Oath

The Olympic Oath is integral to the opening ceremonies of the Olympic Games. One athlete and one official make the oaths as representatives of their fellow competitors and officials, while holding a corner of the Olympic flag.

The athletes' oath is as follows: *In the name of all the competitors I promise that we shall take part in these Olympic Games, respecting and abiding by the rules which govern them, committing ourselves to a sport without doping and without drugs, in the true spirit of sportsmanship, for the glory of sport and the honour of our teams.*

And the officials' oath is as follows: *In the name of all the judges and officials, I promise that we shall officiate in these Olympic Games with complete impartiality, respecting and abiding by the rules which govern them in the true spirit of sportsmanship.*

The father of the modern Olympics, Pierre de Coubertin, suggested the introduction of an oath in 1906, to promote the ideas of fairness and impartiality. However, one was only introduced in 1920 and this was just for the athletes. The officials' version was not introduced until 1972.

4. Number of events

In 1896 there were just 43 events encompassing 9 sports. At London 2012 there were 302 events and in Rio this summer, 306 events will be contested, representing 28 sports (and 41 disciplines) – the newest additions being rugby sevens and golf.

5. Female athletes

The ancient Olympics were open only to male athletes and women were also not allowed to participate in the first modern Olympic Games. Pierre de Coubertin, the father of the modern Olympic Games thought it would be inappropriate to invite female athletes to take part. You might be surprised to learn that it actually wasn't until London 2012 that women competed in all sports included in the Olympic programme. It was also the very first Games in which every single country fielded female athletes.

ABOUT PROJECT:

As I have already discussed above that how Olympics evolved with time. Hence, the major motive of my project is to analyse some historic data regarding Olympics in order to extract the useful insights from the data which will help us to understand the evolution and some facts about Olympics.

Gathering this information can further help us to predict about the results of upcoming Olympic Games. More information about the data on which I have performed the analysis, about the website from where I have taken this data and about the important steps involved in the analysis is as follows:

- This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016.
- The main dataset is from Kaggle - 120 years of Olympic history: athletes and results
- Also for more information about the host city and countries, I had collected data from Wikipedia.
- This dataset provides an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of women, different nations, and different sports and events.
- I did **Exploratory Data Analysis (EDA)** and **Feature Selection** to get better results.
- I did **Data Analysis** to get the important insights from the data to answer my objectives.
- Then built a **Dashboard** for this dataset using **MS Excel**.

Note: - The Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four year cycle starting with 1994, then summer in 1996, then winter in 1998, and so on. A common mistake people make when analyzing this data is to assume that the summer and Winter Games have always been staggered.

DATA

The basic information about the data and its fields are as follows:

ATHLETE_EVENTS.CSV

- **ID** - Unique number for each athlete
- **Name** - Athlete's name
- **Sex** - M or F
- **Age** - Integer
- **Height** - In centimetres
- **Weight** - In kilograms
- **Team** - Team name
- **NOC** - National Olympic Committee 3-letter code
- **Games** - Year and season
- **Year** – Integer (Year in which Olympic happens)
- **Season** - Summer or Winter
- **City** - Host city
- **Sport** - Sport
- **Event** - Event
- **Medal** - Gold, Silver, Bronze, or NA

NOC_REGIONS.CSV

- **NOC-** National Olympic Committee 3-letter code
- **Region-** Name of Country for corresponding NOC code

HOST CITY DATA

For the analysis, I need information about the host countries also which is not there in the Kaggle dataset. Therefore, I extracted that data from an article on Wikipedia.

- **City-** Host city
- **Country-** Host Country
- **Year-** Integer (Year in which Olympic happens)
- **Continent-** Name Of Continent of host country

- **Summer-** Contain a value if Summer Olympic happen in corresponding year
- **Winter-** Contain a value if Winter Olympic happen in corresponding year
- **Opening Ceremony-** Date of opening ceremony
- **Closing Ceremony-** Date of closing ceremony

OBJECTIVES AND SCOPE OF ANALYSIS

SCOPE OF ANALYSIS:

This dataset provides me an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of women, different nations, and different sports and events. This data also needed to be analyses effectively using various features, formulas and fields present in excel. Hence, it helps me to enhance and use my skills productively. I have to perform some cleaning of data after extracting it. Therefore, I learn how to perform Exploratory Data Analysis on a real world data problem. I need to work with pivot charts and tables, slicers, complex functions and formulas etc. in order to extract the insights and visualize them through a beautiful dashboard.

Moreover, gathering information and extricating useful facts from this historic data of Olympic Games (from 1896 to 2016) not only provides the details about the evolution of Olympic Games but it also is really effective and helpful in the prediction about upcoming Olympic Games. We can predict that which countries are more likely to win medals in which sports by analyzing the trend in the historic data where the countries performance get affected by seasons, economy etc. We can also get to know whether there is any effect of a player's age, height on his performance or not? And then make prediction in the future depending on these facts only.

Even though these predictions are only approximations about the game, a player or a country can behave unpredictably also in future. But in more than an average of the cases, the trends keeps on repeating.

However, this analysis is really helpful for the countries to know about their players' strengths and weaknesses. They will be able to know that from past many years in which sport their country is dominating in Olympics and also in which they have poor performance constantly. So, as a result they can invest in the dominating sports more and encourage or put in focus the sports in which their country is having poor performances.

OBJECTIVES:

The objectives on which I have worked on this project majorly are as follows:

- How Seasons affect the countries' performance in Olympic Games.
- Which Country shows domination in which sport by winning gold medal.
- What are the total number and percentage of female and male won a different medal in an Olympics according to season, year, sports, events or countries.
- Performance of a country according to year.
- Effect of hosting an Olympic on host country's performance?
- Does age, height and weight plays an important role in winning a medal?
- Top countries and athletes who won medal according to year, sex (female or male), sports, events or seasons.
- What are the total number and percentage of female and male participated in an Olympics according to season, year, sports, events or countries.
- Individual Player analysis- Total participation, no. of Gold, Silver and Bronze medal.

- Which athlete won maximum and which won minimum medals in Olympics according to Year, Season, Sports and gender.
- Which athlete appeared in Olympics for most of the time?
- How many countries won at least one medal or never won any medal in Olympics?
- How many countries won at least once or never won a Gold medal in Olympics?
- How many countries won at least once or never won a silver medal in Olympics?
- How many countries won at least once or never won a bronze medal in Olympics?

SOURCE OF DATASET

This dataset that I have taken is from KAGGLE. Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

Kaggle got its start in 2010 by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. Its key personnel were Anthony Gold bloom and Jeremy Howard. Nicholas Gruen was founding chair succeeded by Max Levchin. Equity was raised in 2011 valuing the company at $25 million. On 8 March 2017, Google announced that they were acquiring Kaggle.

You can view the dataset at:

DATASET

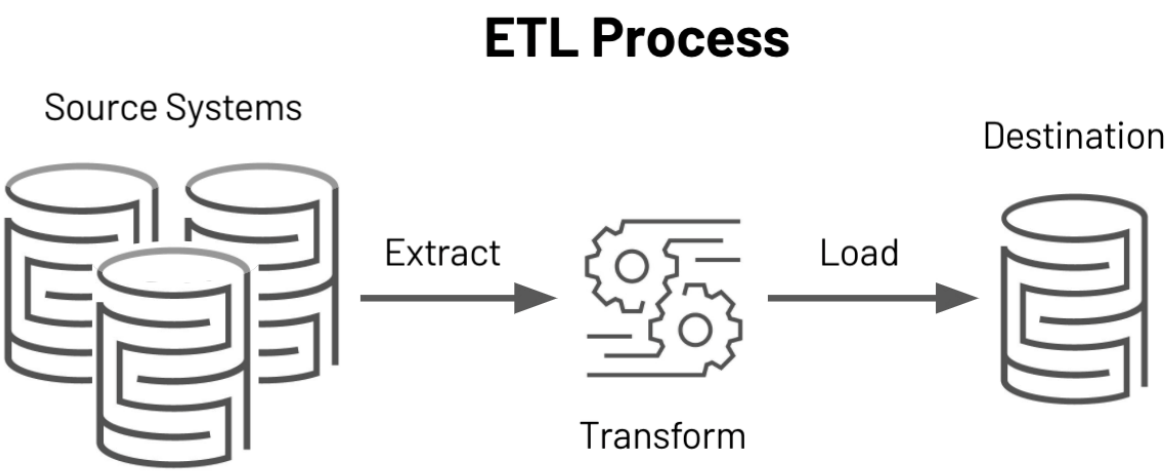
Also, the data for Host Countries is taken from an article about Olympics on Wikipedia. The link for that Data is here-

DATA ON WIKIPEDIA

After extracting the raw data and storing it in a excel file, I have uploaded that data on Google Drive. The link for the same is as follows-

HOST CITY DATA ON DRIVE

ETL PROCESS



EXTRACT:

The main data is taken from Kaggle website. From there I downloaded that data and it is in the form of .CSV files. There are two files that I have downloaded-

- Athlete_events.csv
- Noc_regions.csv

Also the host city and country data is taken from Wikipedia. And then I have stored that data in an excel file names as Host city data.xlsx

After extracting this data I have transform this data according to my needs.

TRANSFORM:

- First, I have opened the .CSV files on my MS EXCEL in two different sheets and named them ATHLETES AND EVENTS or NOC_REGIONS.
- Then I converted the data in those files by splitting the data about the semicolon into columns.
- After that I find out that some columns that is age, height and weight contains NA values which are of no use for analysis. Therefore, I decided to fill average of the respective columns in the cell containing NA in a new column namely age*, height* and weight* respectively.
- Also I made a new column for **country** according to the NOC values as it will be easy to recognize country name rather than NOC values. This I have done by using the NOC_REGIONS sheet values.
- In the Host city data, the opening ceremony, continent and closing ceremony is of no use to me. So, I have removed that columns from the data.
- There are some rows that contain the data about the Olympic Games which had been cancelled that time, so I removed that rows.
- There is no information about the 1906 Olympic Games, so I have added that manually.
- Combine the Column named summer, winter and make a single column Season which contains the values as summer, winter.

LOAD:

- After performing some sort of EDA and cleaning the data I load that data in the form of a table in a new sheet named as DATA connected to the ATHLETES AND EVENTS sheet.
- I have also removed some columns while loading the data like ID, which is of no use for me and the original columns of age, weight, and height that contains the NA values.
- The host city data is loaded as a table to DATA2 sheet connected to HOST CITY DATA sheet, after that summer and winter columns are being removed.
- This DATA and DATA2 sheet is now further be used for analysis.

ANALYSIS ON DATASET

ANALYSIS - 1

- **INTRODUCTION:**

  What are the total number and percentage of female and male participated in an Olympics according to season, year, sports, events or countries.

- **GENERAL DESCRIPTION:**

  In this the main motive is to find out the total participants in Olympics according to season, year, sports, events or countries. Moreover, in this analysis we will also find out the No. of male and female participants respectively to know the ratio of male VS female participants. Also to show the ratio or percentage of male VS female participants a visualization will be there for better understanding about the same.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart, slicers, and two functions named as GETPIVOTDATA and IFERROR.

- **ANALYSIS RESULTS:**

  TOTAL PARTICIPANTS: 271116
  FEMALE PARTICIPANTS: 74522 (27%)
  MALE PARTICIPANTS: 196594 (73%)

  We can use slicers also to filter out the results on the basis of season, country, year, sports and events.

  FOR EXAMPLE:
  By using season slicer we get:

  IN SUMMER OLYMPICS:
  TOTAL PARTICIPANTS: 222552
  FEMALE PARTICIPANTS: 59443 (27%)
  MALE PARTICIPANTS: 163109 (73%)

IN WINTER OLYMPICS:

TOTAL PARTICIPANTS: 48564

FEMALE PARTICIPANTS: 15079 (31%)

MALE PARTICIPANTS: 33485 (69%)

By using year slicer, let's say we want to find out the statistics for year 1972

TOTAL PARTICIPANTS: 11959

FEMALE PARTICIPANTS: 2608 (22%)

MALE PARTICIPANTS: 9351 (78%)

By using country slicer we can find out the total participation by a particular country in Olympics: Let's see for AFGANISTAN

TOTAL PARTICIPANTS: 126

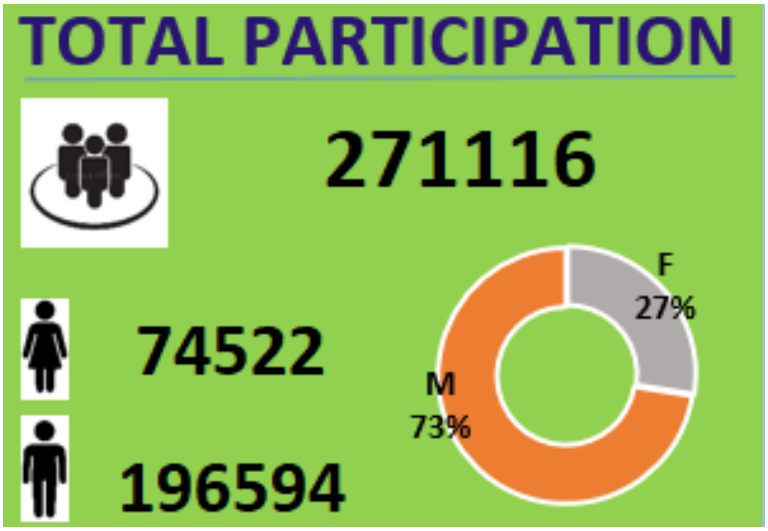
FEMALE PARTICIPANTS: 5 (4%)

MALE PARTICIPANTS: 121 (96%)

Similarly by using the slicers for sports and events also I analyse the total participation and ratio between male and female. We can also find out the results by applying multiple filters.

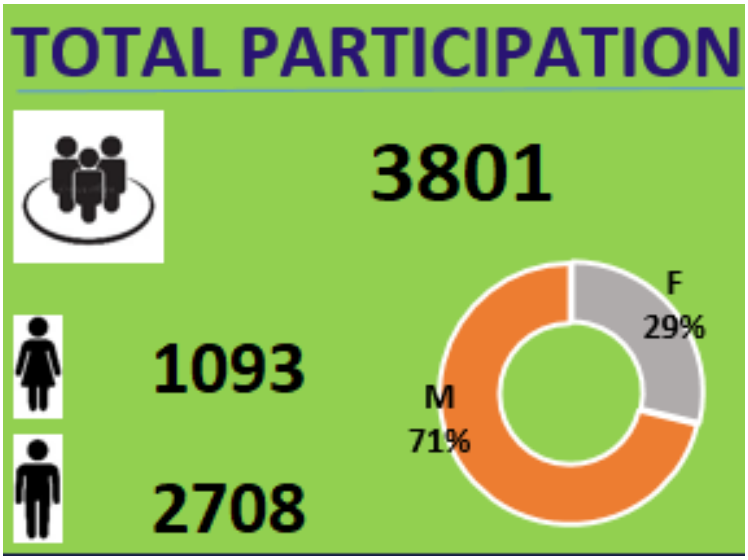
HERE ONE THING THAT NEED TO BE FOCUSSED IS THAT IN NONE OF THE CASES THE PERCENTAGE OF FEMALE ARISE UP TO OR EQUALS TO 50%. THAT MEANS WE NEED TO FOCUS ON OUR FEMALE ATHLETES.

- **VISUALIZATION:**



This is the visualization that I have created in order to represent the results of the analysis which is currently representing the total no of players participating in both summer and winter Olympics. After applying the filters through slicers the same visualization show the respective results.

LIKE BELOW IS THE VISUALIZATION FOR THE PARTICIPATION OF PLAYERS IN SPORTS JUDO.



ANALYSIS -2

- **INTRODUCTION:**

  What are the total number and percentage of female and male won a different medal in an Olympics according to season, year, sports, events or countries.

- **GENERAL DESCRIPTION:**

  In this the main motive is to find out the total number of athletes that won gold medal, silver medal, bronze medal in Olympics and also the total number of players that never won any medal in Olympics. Moreover in the analysis for different medals and never winning any medal, the number and percentage of male and females is also being calculated so as to find out the ratio between them. Also I have made a visualization to understand the same in a better

way. The options to filter out the information through slicers of year, country, season, sports and events is also available.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart, slicers, and two functions named as GETPIVOTDATA and IFERROR.

- **ANALYSIS RESULTS:**

  During analysis, I have find out the following statistics-

  GOLD MEDAL:
  TOTAL NUMBER OF ATHLETES WINNING: 13372
  NUMBER OF FEMALES WINNING GOLD MEDAL: 3747 (28%)
  NUMBER OF MALES WINNING GOLD MEDAL: 9625 (72%)

  SILVER MEDAL:
  TOTAL NUMBER OF ATHLETES WINNING: 13116
  NUMBER OF FEMALES WINNING SILVER MEDAL: 3735 (28%)
  NUMBER OF MALES WINNING SILVER MEDAL: 9381 (72%)

  BRONZE MEDAL:
  TOTAL NUMBER OF ATHLETES WINNING: 13295
  NUMBER OF FEMALES WINNING BRONZE MEDAL: 3771 (28%)
  NUMBER OF MALES WINNING BRONZE MEDAL: 9524 (72%)

  NEVER WON:
  TOTAL NUMBER OF ATHLETES: 231333
  NUMBER OF FEMALES: 63269 (27%)
  NUMBER OF MALES: 168064 (73%)

  We can also analyse the medals statistics by applying year, country, season, sports and events filters through slicers.

FOR EXAMPLE: We have to find the number of medals winning athletes of Austria in winter Olympics in the year of 2002 in Alpine Skiing Men's Super G.

GOLD MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 0

NUMBER OF FEMALES WINNING GOLD MEDAL: 0 (0%)

NUMBER OF MALES WINNING GOLD MEDAL: 0(0%)

SILVER MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 1

NUMBER OF FEMALES WINNING SILVER MEDAL: 0 (0%)

NUMBER OF MALES WINNING SILVER MEDAL: 1 (100%)

BRONZE MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 1

NUMBER OF FEMALES WINNING BRONZE MEDAL: 0 (0%)

NUMBER OF MALES WINNING BRONZE MEDAL: 1 (100%)

NEVER WON:

TOTAL NUMBER OF ATHLETES: 2

NUMBER OF FEMALES: 0 (0%)

NUMBER OF MALES: 2 (100%)

THIS ANALYSIS HELP US TO UNDERSTAND THAT IN A PARTCULAR SPORTS WHICH COUNTRY HAS WON MAXIMUM MEDALS WHICH IS VERY USEFUL. ALSO WE CAN FIND OUT THE MEDALS WON BY DIFFERENT COUNTRIES IN DIFFERENT YEARS IN EITHER SUMMER OR WINTER OLYMPIC OR BOTH. WE CAN FIND OUT AND THEN COMPARE THE RESULTS.

Like for USA we want to find the number of medals won by them in shooting in 1896, 1972 and 2008, in order to see that whether there is improvement in their performance or decrement. We can fetch the results for every year and then compare in order to find out the results.

For 1896

GOLD MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 2

NUMBER OF FEMALES WINNING GOLD MEDAL: 0 (0%)

NUMBER OF MALES WINNING GOLD MEDAL: 2 (100%)

SILVER MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 1

NUMBER OF FEMALES WINNING SILVER MEDAL: 0 (0%)

NUMBER OF MALES WINNING SILVER MEDAL: 1 (100%)

BRONZE MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 0

NUMBER OF FEMALES WINNING BRONZE MEDAL: 0 (0%)

NUMBER OF MALES WINNING BRONZE MEDAL: 0 (0%)

NEVER WON:

TOTAL NUMBER OF ATHLETES: 1

NUMBER OF FEMALES: 0 (0%)

NUMBER OF MALES: 1 (100%)

For 1972

GOLD MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 2

NUMBER OF FEMALES WINNING GOLD MEDAL: 0 (0%)

NUMBER OF MALES WINNING GOLD MEDAL: 2 (100%)

SILVER MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 2

NUMBER OF FEMALES WINNING SILVER MEDAL: 0 (0%)

NUMBER OF MALES WINNING SILVER MEDAL: 2 (100%)

BRONZE MEDAL:

TOTAL NUMBER OF ATHLETES WINNING: 0
NUMBER OF FEMALES WINNING BRONZE MEDAL: 0 (0%)
NUMBER OF MALES WINNING BRONZE MEDAL: 0 (0%)

NEVER WON:
TOTAL NUMBER OF ATHLETES: 12
NUMBER OF FEMALES: 0 (0%)
NUMBER OF MALES: 12 (100%)

For 2008
GOLD MEDAL:
TOTAL NUMBER OF ATHLETES WINNING: 2
NUMBER OF FEMALES WINNING GOLD MEDAL: 0 (0%)
NUMBER OF MALES WINNING GOLD MEDAL: 2 (100%)

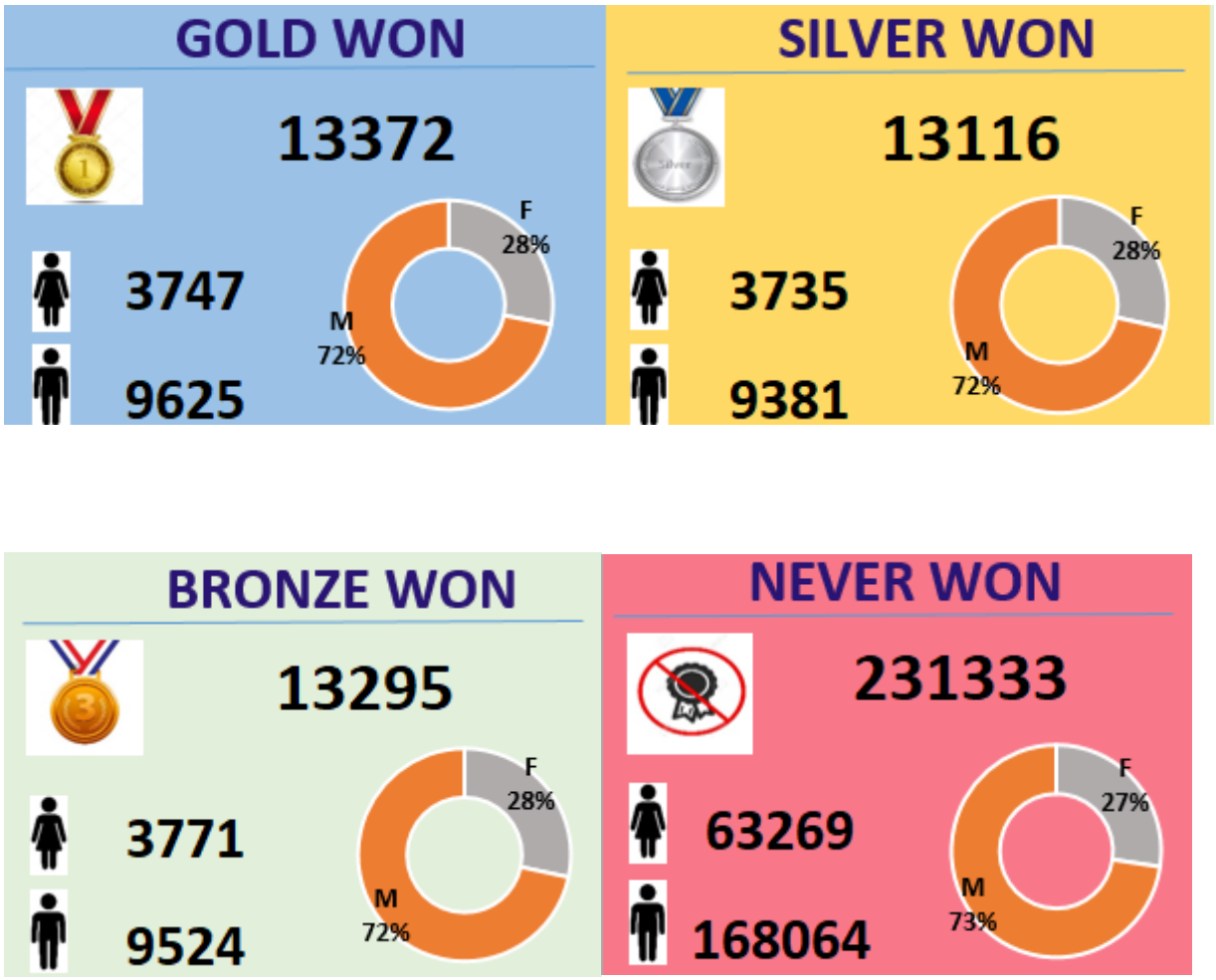
SILVER MEDAL:
TOTAL NUMBER OF ATHLETES WINNING: 2
NUMBER OF FEMALES WINNING SILVER MEDAL: 1 (50%)
NUMBER OF MALES WINNING SILVER MEDAL: 1 (50%)

BRONZE MEDAL:
TOTAL NUMBER OF ATHLETES WINNING: 2
NUMBER OF FEMALES WINNING BRONZE MEDAL: 1 (50%)
NUMBER OF MALES WINNING BRONZE MEDAL: 1 (50%)

NEVER WON:
TOTAL NUMBER OF ATHLETES: 21
NUMBER OF FEMALES: 8 (38%)
NUMBER OF MALES: 13 (62%)

Comparing these results we can find out that the total participation of athletes in Olympics have been increased from 1896 to 2008, and also the medals won by USA from 1896 to 2008 have increased which is a good thing showing that their performance is becoming good over years.

- **VISUALIZATION:**



ANALYSIS- 3

- **INTRODUCTION:**

  Individual Player analysis- Total participation, no. of Gold, Silver and Bronze medal.

- **GENERAL DESCRIPTION:**

  In this the main motive is to check out the athletes' performance individually. This is very helpful if we want to know any particular player performance statistics. In this I have analyse and visualize the Athletes' performances in different year. I have used slicers with name and year, where we can select name and check out the statistics and also we can apply year filters through slicer to check out that player performance in that specific year.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

For this analysis, I require pivot table, slicers, and two functions named as GETPIVOTDATA and IFERROR.

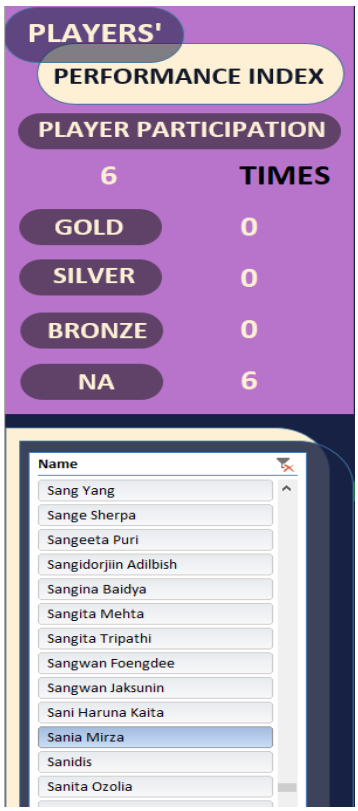
- **ANALYSIS RESULTS:**

  I have analyse the individual players' performance through this analysis. Let us see the performance of Athlete named Sania Mirza.

  TOTAL PARTICIPATION: 6
  GOLD MEDAL: 0
  SILVER MEDAL: 0
  BRONZE MEDAL: 0
  NA: 6

  Thus through this we are able to get the information that Sania Mirza has participated six times but never able to won any medal in Olympics.

- **VISUALIZATION:**



ANALYSIS- 4

- **INTRODUCTION:**

  Top countries and athletes who won medal according to year, sex (female or male), sports, events or seasons.

- **GENERAL DESCRIPTION:**

  In this the main motive is to find out the top athletes and top countries in Olympics. Also we can find out the top athletes and countries in different years, seasons, sports, events and sex.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart and slicers.

- **ANALYSIS RESULTS:**
- **TOP COUNTRIES:**

  THE OVERALL TOP PERFORMING COUNTRY IS:

  USA with 5637 medals

  THE TOP PERFORMING COUNTRY IN SUMMER OLYMPICS IS:

  USA with 5002 medals

  THE TOP PERFORMING COUNTRY IN WINTER OLYMPICS IS:

  RUSSIA with 759 medals.

  THE OVERALL TOP PERFORMING COUNTRY IN TERMS OF FEMALE ATHLETES IS:

  USA with 1805 medals

  THE TOP PERFORMING COUNTRY IN SUMMER OLYMPICS IN TERMS OF FEMALE ATHLETES:

  USA with 1564 medals

THE TOP PERFORMING COUNTRY IN WINTER OLYMPICS IN TERMS OF FEMALE ATHLETES:

RUSSIA with 242 medals

THE OVERALL TOP PERFORMING COUNTRY IN TERMS OF MALE ATHLETES IS:

USA with 3832 medals

THE TOP PERFORMING COUNTRY IN SUMMER OLYMPICS IN TERMS OF MALE ATHLETES:

USA with 3438 medals

THE TOP PERFORMING COUNTRY IN WINTER OLYMPICS IN TERMS OF MALE ATHLETES:

RUSSIA with 517 medals

Thus in conclusion we can say that USA performs extremely well in Summer Olympics while Russia done great in Winter Olympics.

- **TOP ATHLETES:**

  THE OVERALL TOP PERFORMING ATHLETE IS:

  Michael Fred Phelps, II with 28 medals

  THE TOP PERFORMING ATHLETE IN SUMMER OLYMPICS IS:

  Michael Fred Phelps, II with 28 medals

  THE TOP PERFORMING ATHLETE IN WINTER OLYMPICS IS:

  Ole Einar Bjrndalen with 13 medals

  THE OVERALL TOP PERFORMING ATHLETE IN TERMS OF FEMALE ATHLETES IS:

  Larysa Semenivna Latynina (Diriy-) with 18 medals

  THE TOP PERFORMING ATHLETE IN WINTER OLYMPICS IN TERMS OF FEMALE ATHLETES:

  Stefania Belmondo with 10 medals

THE OVERALL TOP PERFORMING ATHLETE IN TERMS OF MALE ATHLETES IS:

Michael Fred Phelps, II with 28 medals

THE TOP PERFORMING ATHLETE IN SUMMER OLYMPICS IN TERMS OF MALE ATHLETES:

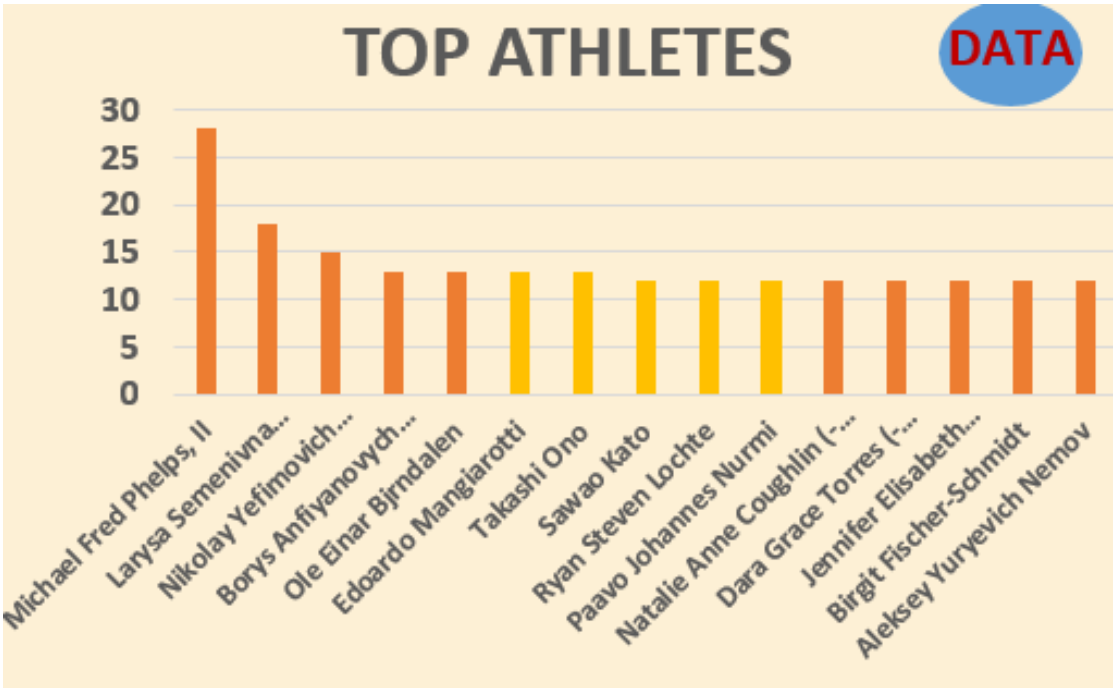
Michael Fred Phelps, II with 28 medals

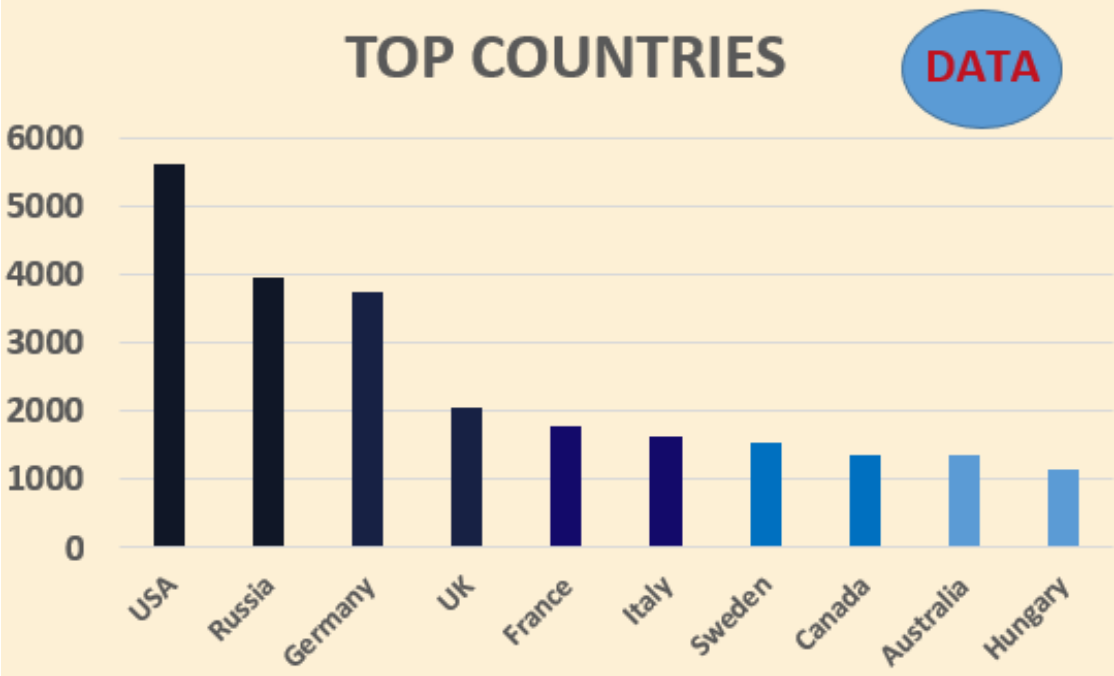
THE TOP PERFORMING ATHLETE IN WINTER OLYMPICS IN TERMS OF MALE ATHLETES:

Ole Einar Bjrndalen with 13 medals

We can even find out more specific information by being particular about year, sports and events too through slicers.

- **VISUALIZATION:**





These are visulizations that comes out for overall top countries and athletes. Similarly, other results are also being showed like these.

ANALYSIS- 5

- **INTRODUCTION:**

  Does age, height and weight plays an important role in winning a medal?

- **GENERAL DESCRIPTION:**

  In this the main motive is to find whether there is any relation between the athletes' age, height and weight with the number of medals he won.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart and slicers.

- **ANALYSIS RESULTS:**
- **HEIGHT:**

After analysing the data I came to know that there is a strong relation between an athletes' height with the no. of medals he won.

The analysis shows that the athletes with 175 cm of height have won the 46.22% of medals. While after this the maximum medals are being won by athletes with 180 cm height with is only 8.20% of medals.

Specific to Males also shows that male athletes with 175 cm of height won 50.39% of medals, the next is 180 cm with 7.92% of medals.

In case of female athletes also the height 175 cm shows high percentage of winning medals as female athletes with 175 cm of height win 22.02% of medals, while next is 170 cm with 13.41% of medals.

So, this shows that there is a huge impact of player's height (175 cm) with its certainty to win a medal.

- **WEIGHT:**

The statistics came out by analysing the weight of athletes with the medals they won is also very useful.

The analysis shows that the athletes with 71 kg of weight is more likely to win a medal as they have won the 53.22% of medals.

The next one after this is the 70 kg category but they have won only 6.79 % of medals.

Specific to Males also shows that male athletes with 71 kg of weight won 58.38% of medals, the next is 75 kg with 6.11% of medals.

In case of female athletes also the weight 71 kg shows high percentage of winning medals as female athletes with 71kg of weight win 20.68% of medals, while next is 60 kg with 10.63% of medals.

In conclusion, there is a huge impact of player's weight (71 kg) with its certainty to win a medal.

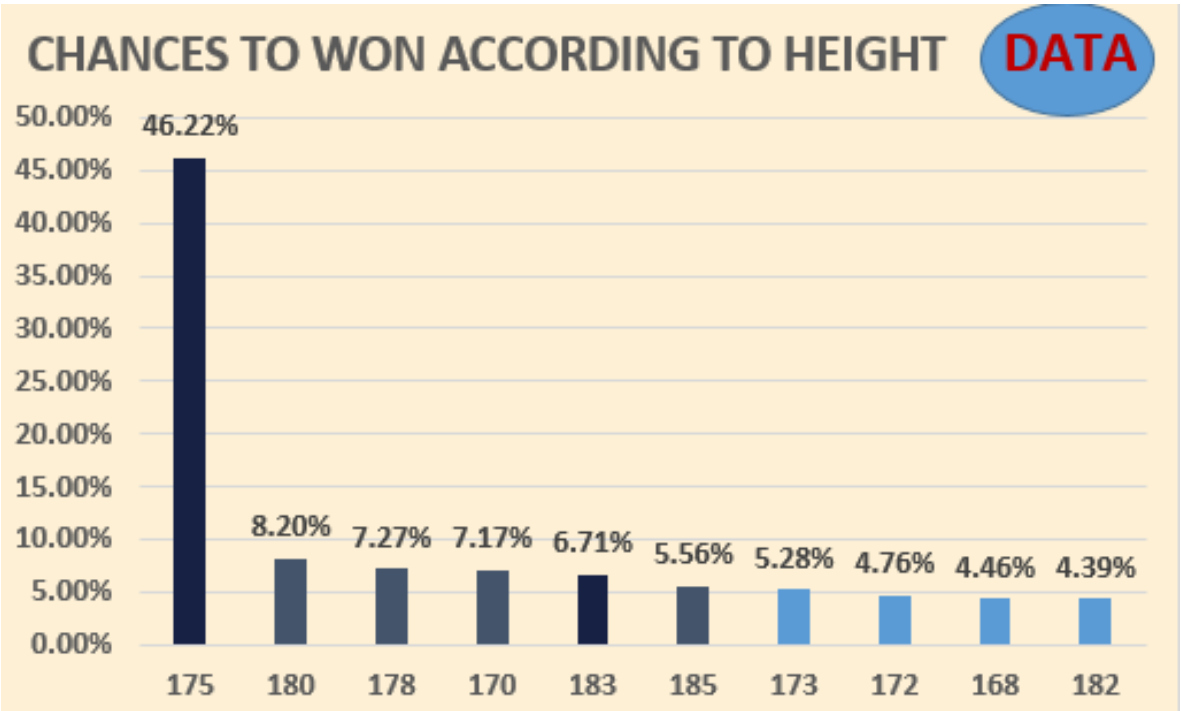
- **AGE:**

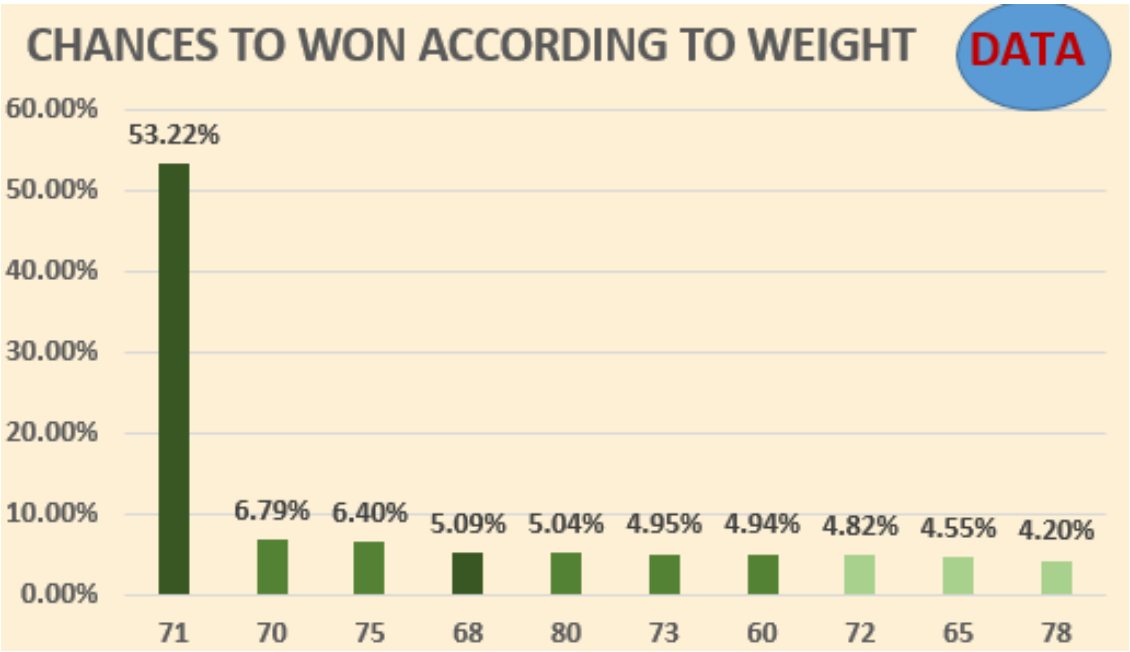
  Analysing about the age and no of medals they won, I came to know that **the athletes within the age from 20 – 28 won maximum number of medals as compared to the athletes with ages above and below them.**

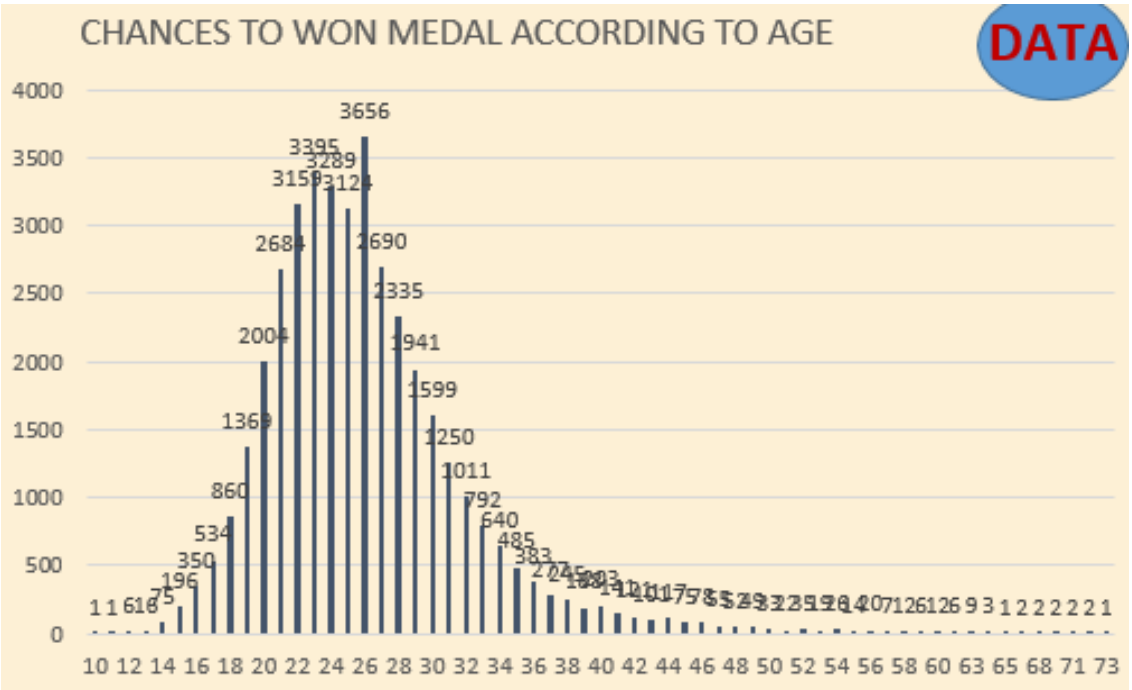
  Within 20-28, Athletes with age 26 have won the maximum no of medals (3656 medals). Whereas the athletes within the age of 34 to 40 won very less medals as compared to the age group 20-28 and this graph falls down even more worse for higher age athletes that is above 40.

  In conclusion, Age have an important impact on athlete's performance and athletes with age within 20-28 especially 26 years are have more certainty to win a medal.

- **<u>VISUALIZATION:</u>**







However we can visualize the results in respect to specific sports, season, sex and event too. But the below mentioned tables of result is for the overall analysis. We can filter out the results using slicers in the dashboard and then see the desired outcome.

ANALYSIS- 6

- **INTRODUCTION:**

  Which athlete won maximum and which won minimum medals in Olympics according to Year, Season, Sports and gender.

- **GENERAL DESCRIPTION:**

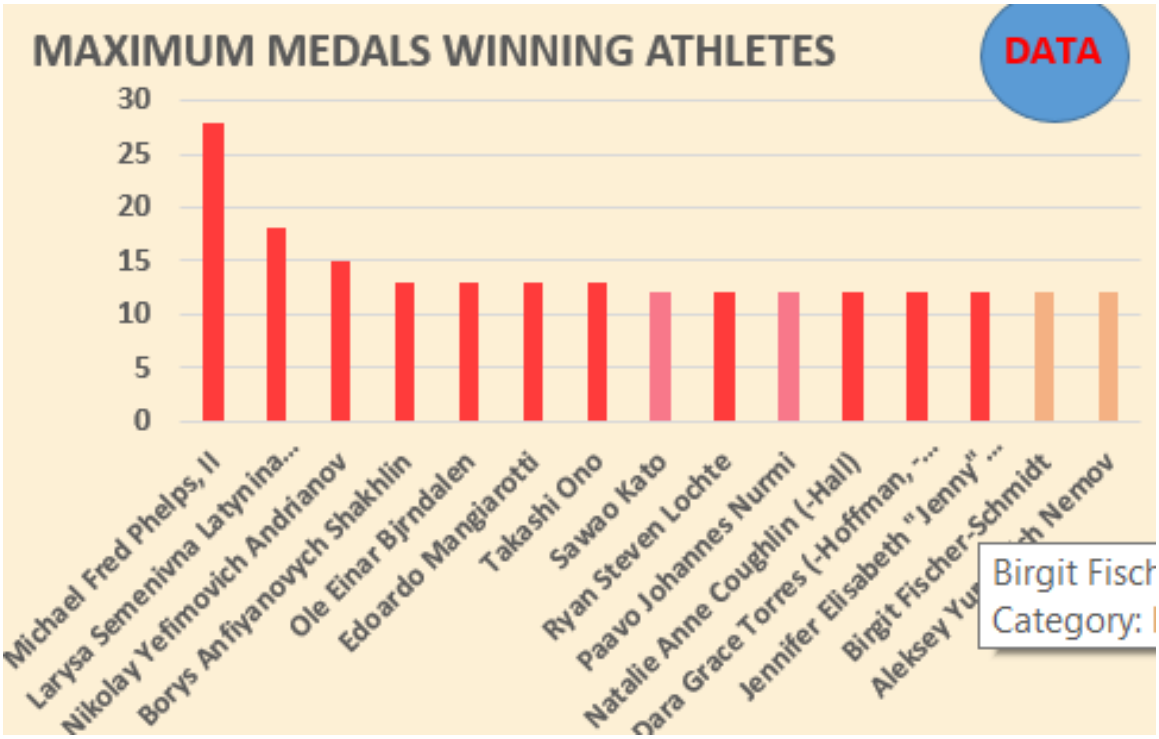
  In this the main motive is to find out the Athletes that have won the maximum and minimum number of medals in different year, sports, season, and gender.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

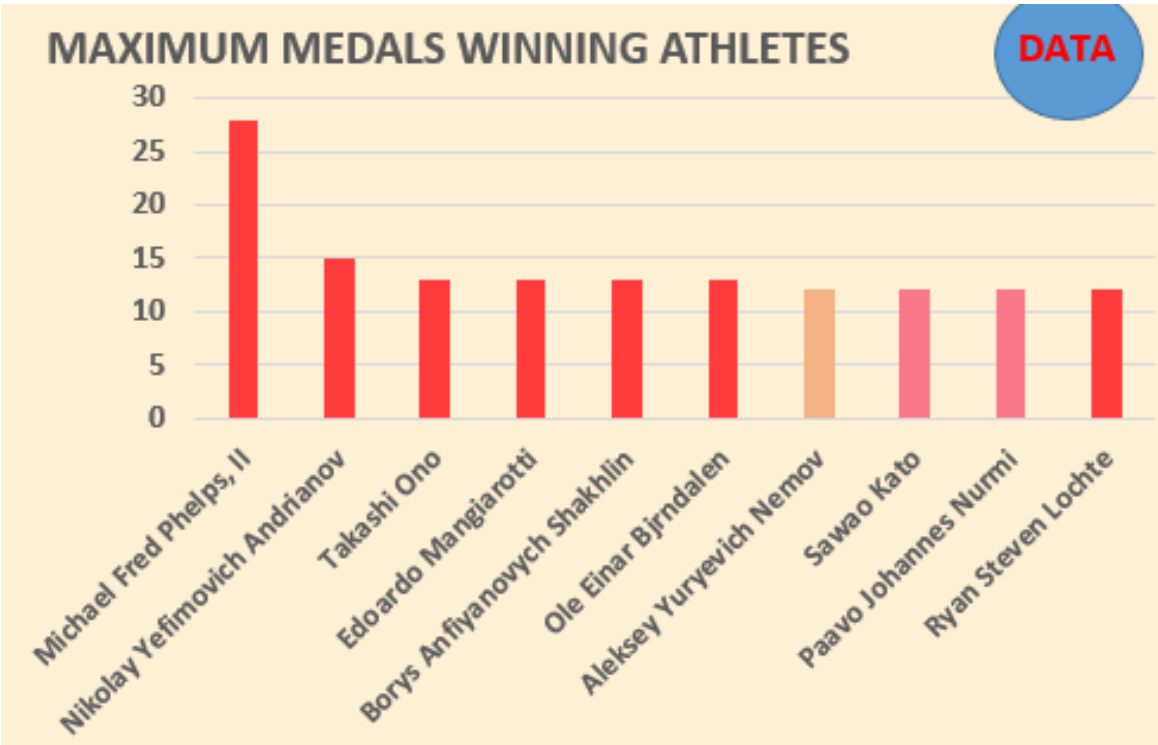
  For this analysis, I require pivot table, pivot chart and slicers.

- **ANALYSIS RESULTS AND VISUALIZATION:**

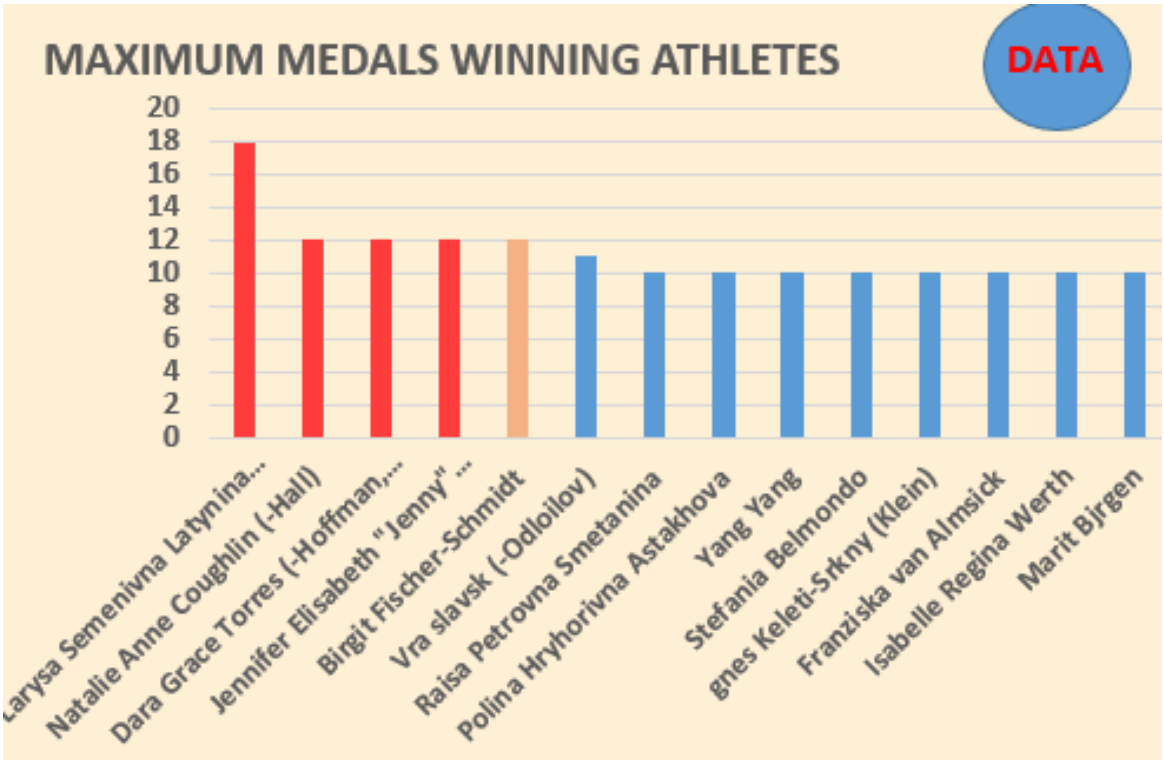
  Overall maximum medal winning athletes



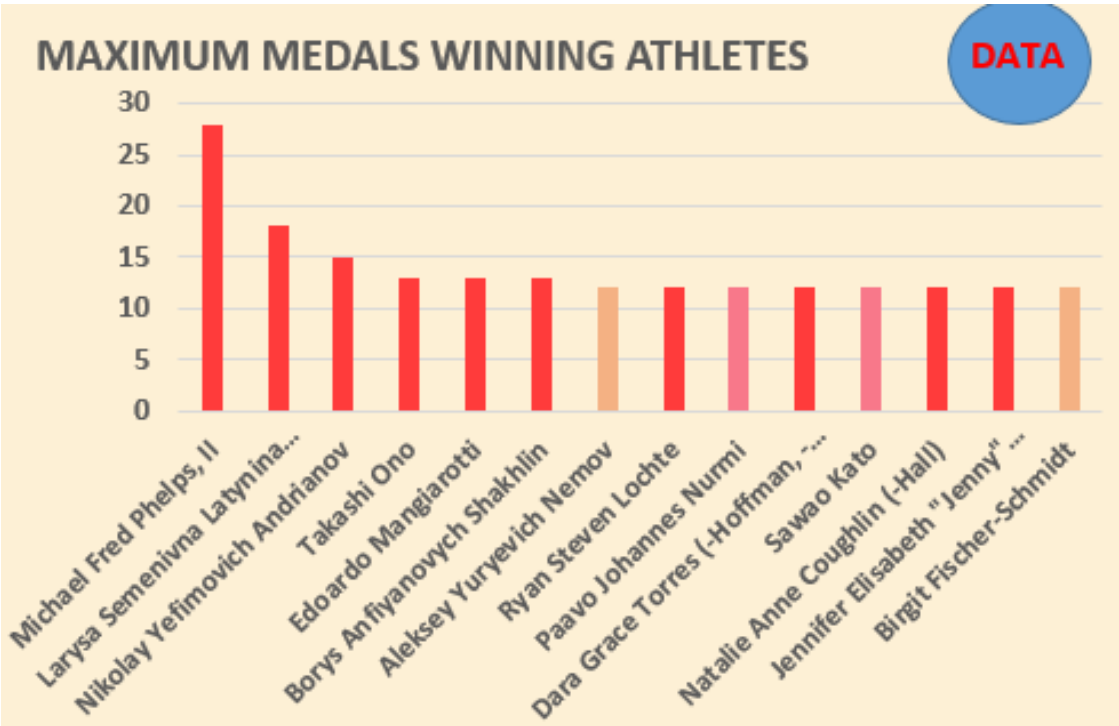
Maximum medals winning athletes in terms of male only



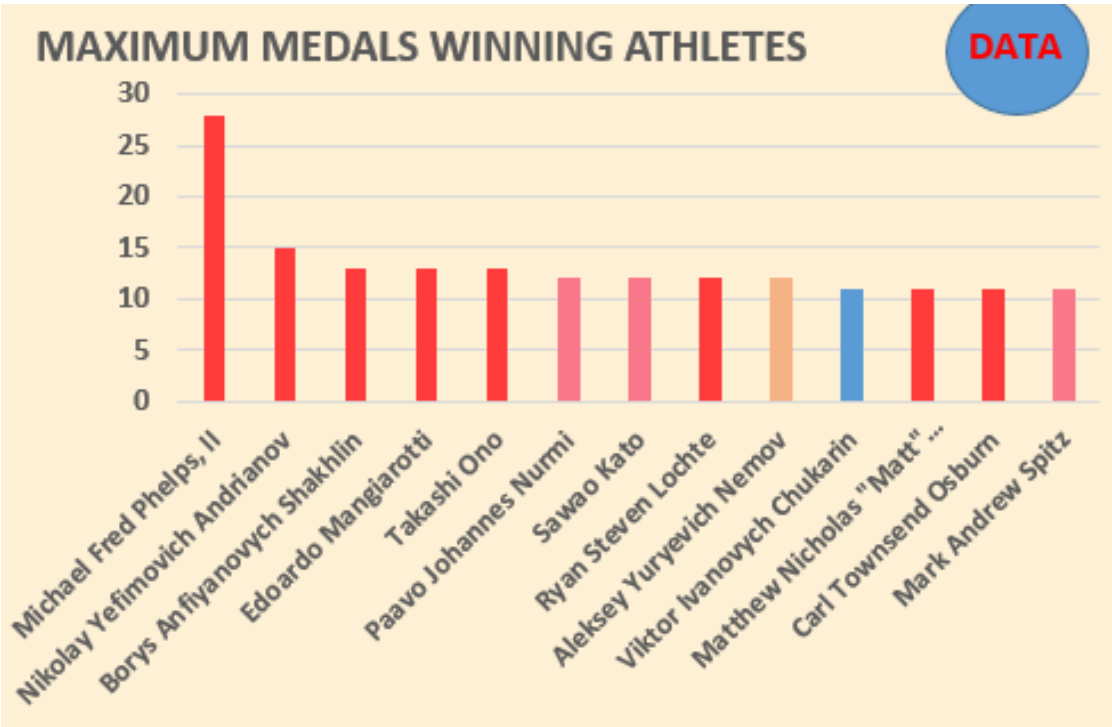
Maximum medals winning athletes in terms of female athletes



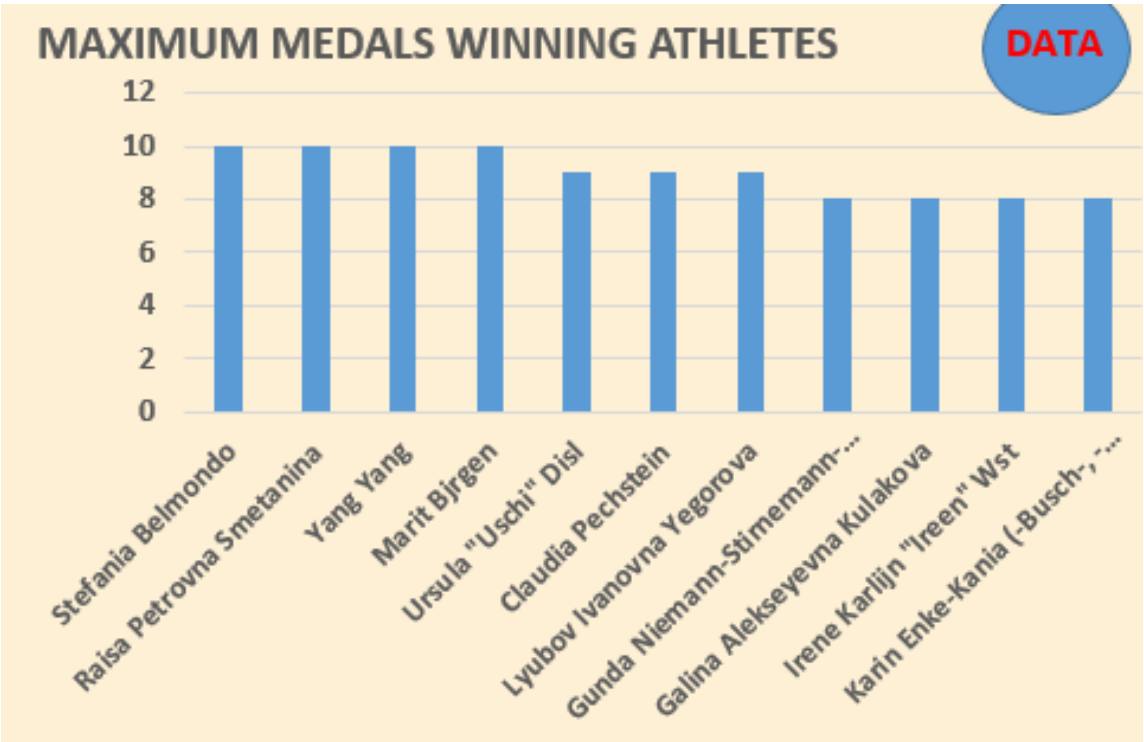
Overall Maximum medals winning athletes in Summer Olympics



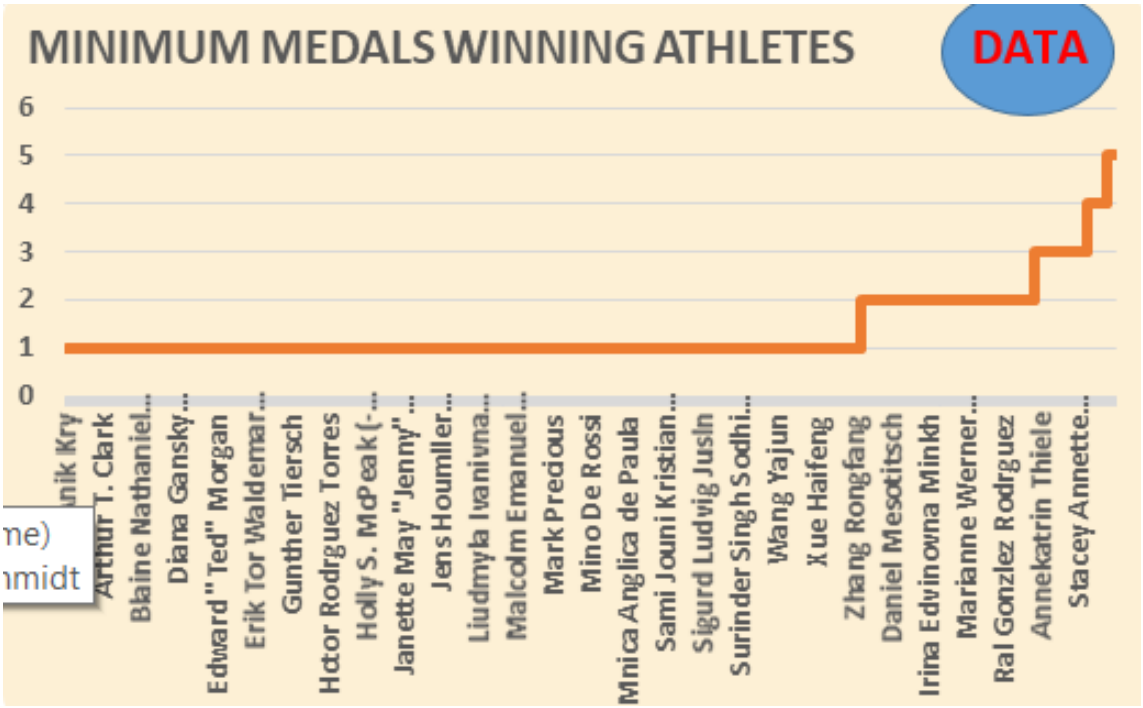
Maximum medals winning Athletes in Summer Olympics in terms of male only



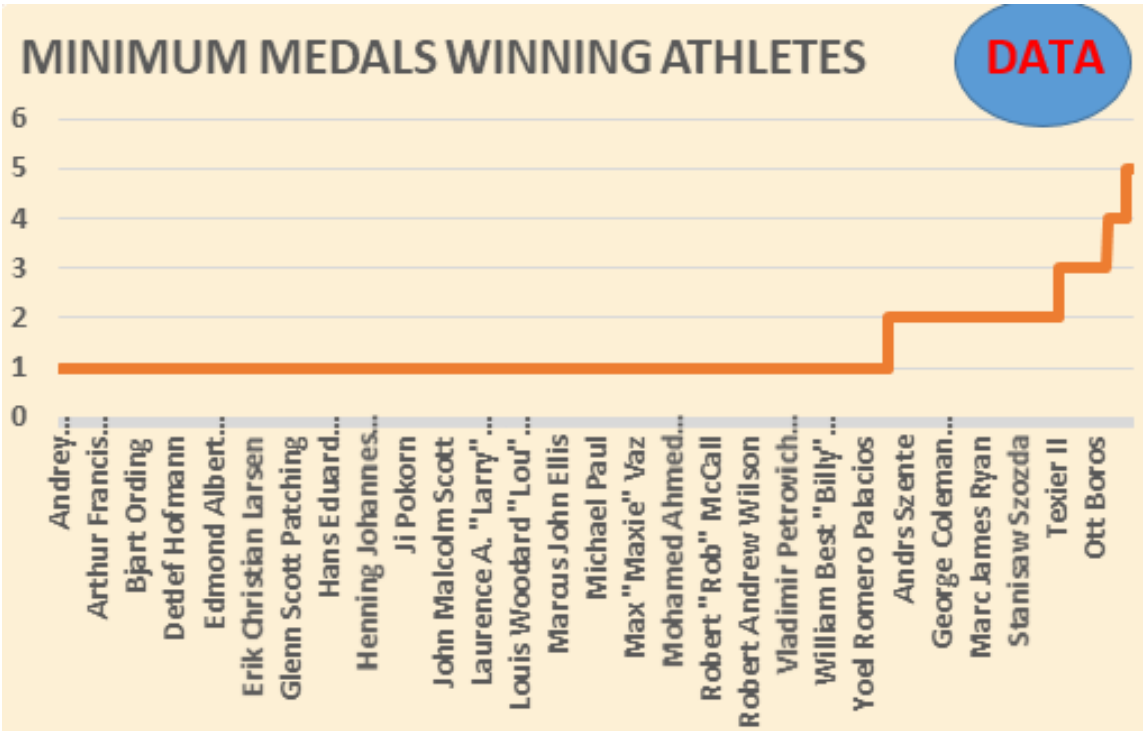
Maximum medals winning Athletes in Winter Olympics in terms of female only



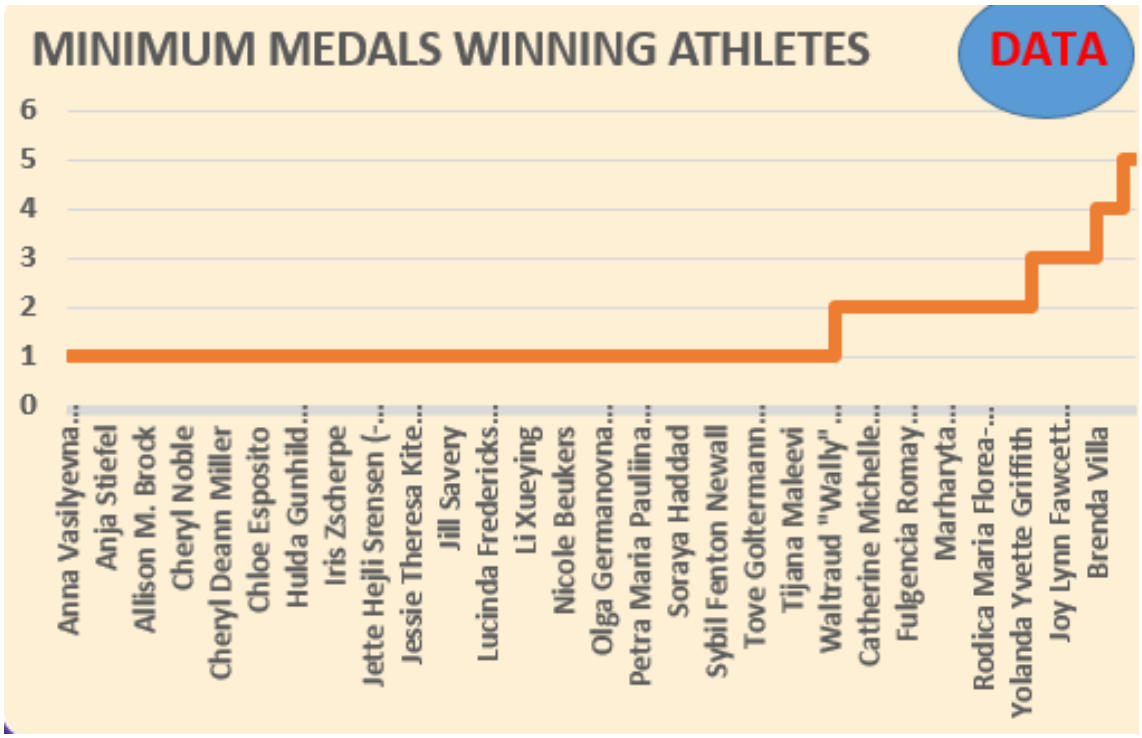
Overall minimum medal winning athletes



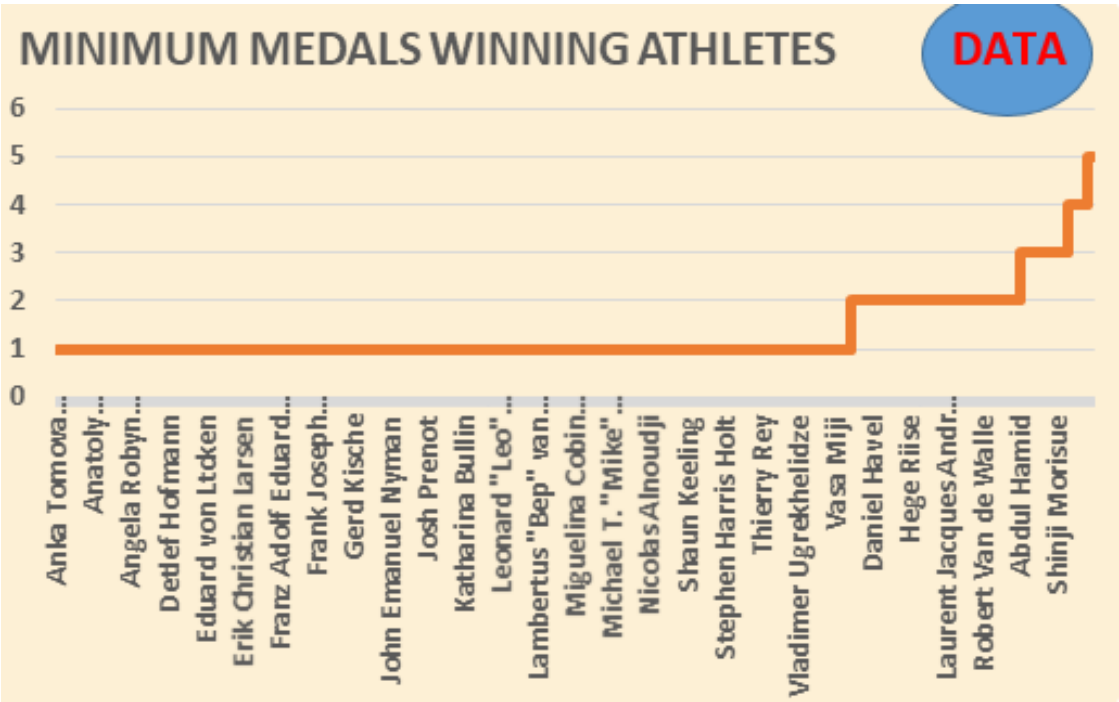
Minimum medals winning athletes in terms of male only



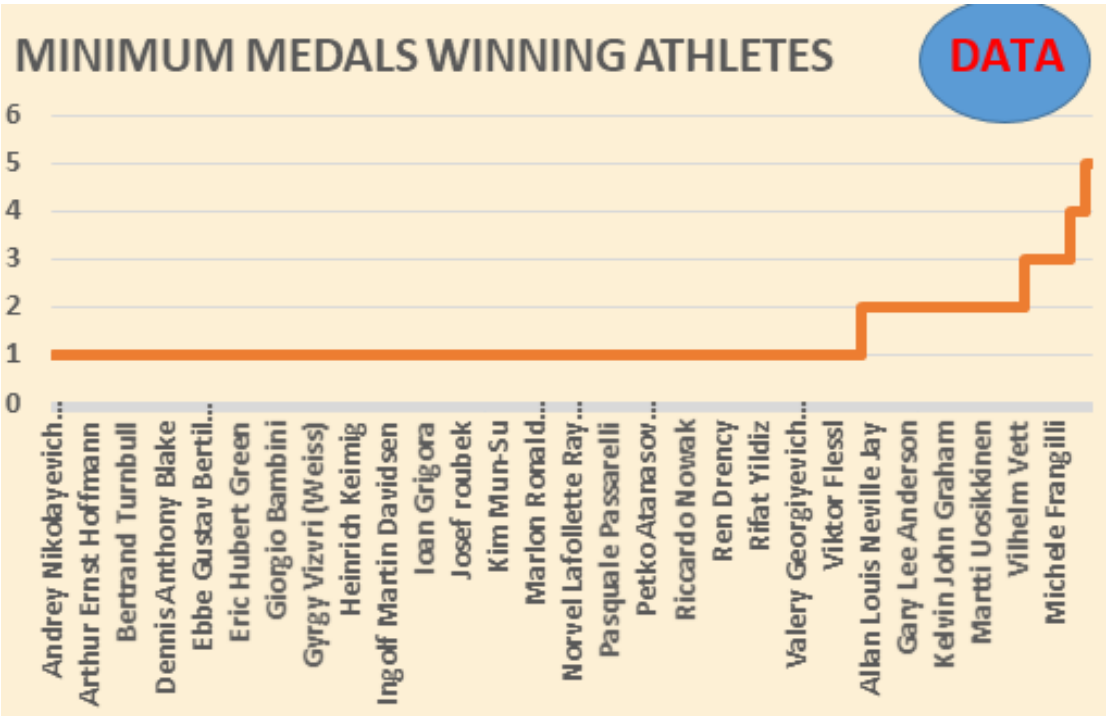
Minimum medals winning athletes in terms of female athletes



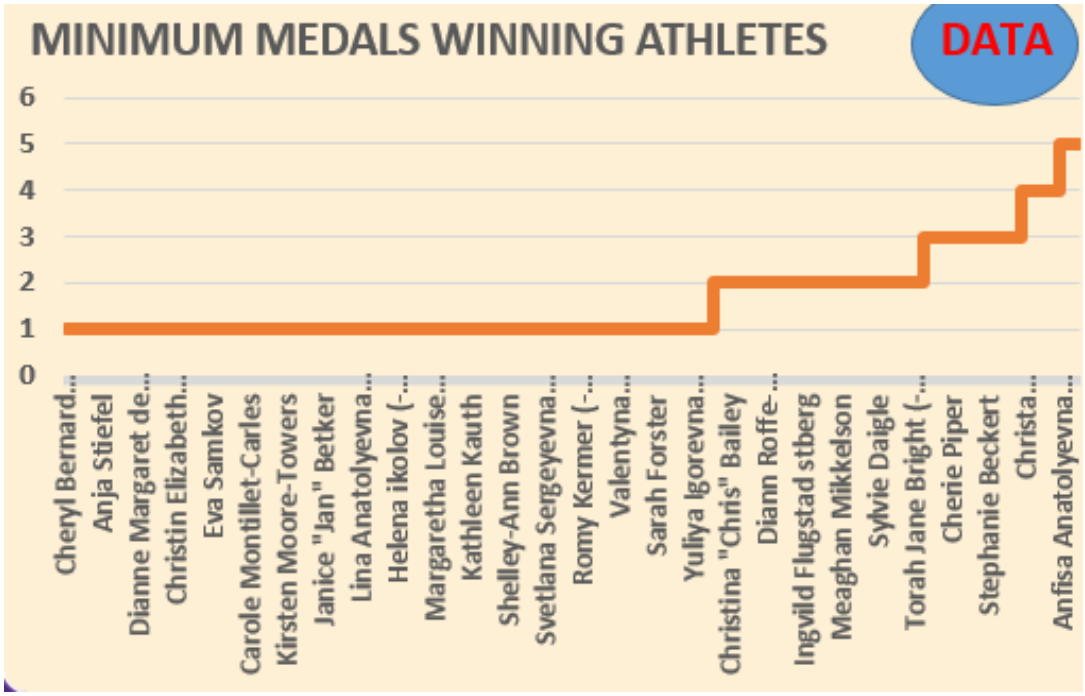
Overall Minimum medals winning athletes in Summer Olympics



Minimum medals winning Athletes in Summer Olympics in terms of male only



Maximum medals winning Athletes in Winter Olympics in terms of female only



Similarly, we can do a lot on dashboard to find out the particular results.

ANALYSIS- 7

- **INTRODUCTION:**

  Which athlete appeared in Olympics for most of the time?

- **GENERAL DESCRIPTION:**

  In this the main motive is to find out the maximum participating athlete. Also to analyse that who is the maximum participating athlete with respect to gender and season.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, GETPIVOTDATA function, IFERROR function and slicers.

- **ANALYSIS RESULTS :**

  OVERALL MAXIMUM PARTICIPATING ATHLETE:

  Robert Tait McKenzie

  MAXIMUM PARTICIPATING ATHLETE AMONG MALES:

Robert Tait McKenzie

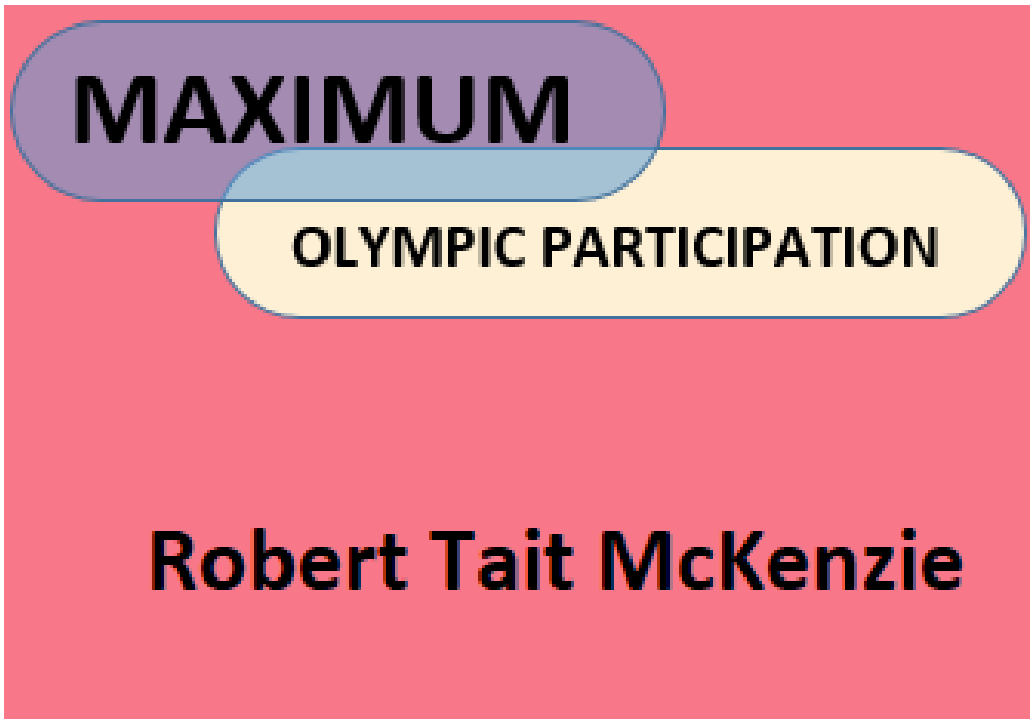
MAXIMUM PARTICIPATING ATHLETE AMONG FEMALES:
Oksana Aleksandrovna Chusovitina

MAXIMUM PARTICIPATING ATHETE IN SUMMER OLYMPICS:
Robert Tait McKenzie

MAXIMUM PARTICIPATING ATHLETE IN WINTER OLYMPICS:
Ole Einar Bjrndalen

MAXIMUM PARTICIPATING ATHLETE IN WINTER OLYMPICS IN FEMALES:
Gabriella Paruzzi

- **VISUALIZATION :**



The above visualization shows the overall maximum participating athlete.

ANALYSIS- 8

- **INTRODUCTION:**
- How many countries won at least one medal or never won any medal in Olympics?
- How many countries won at least once or never won a Gold medal in Olympics?
- How many countries won at least once or never won a silver medal in Olympics?
- How many countries won at least once or never won a bronze medal in Olympics?

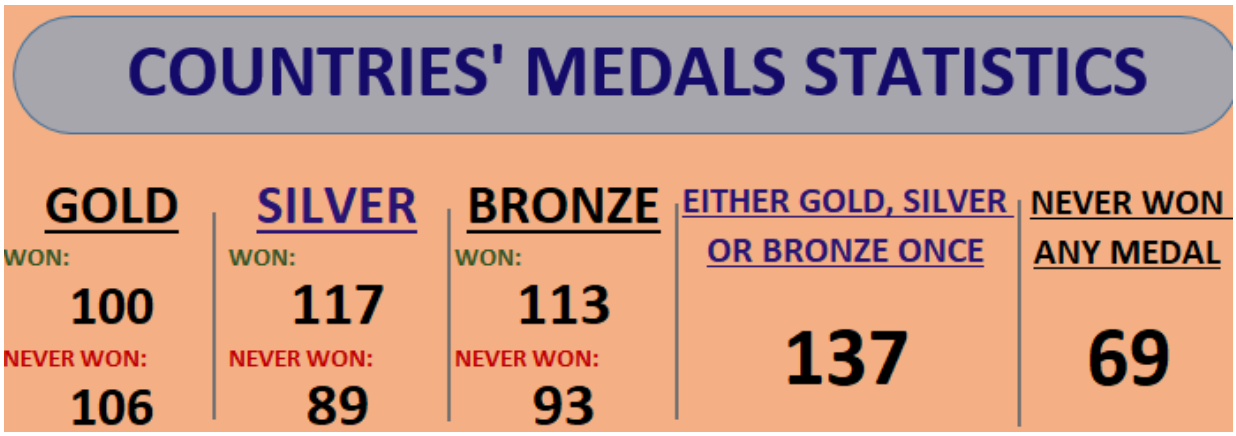
- **GENERAL DESCRIPTION:**

  In this the main motive is to find out the medal statistics of countries and knowing about how many different countries have won or never won any medal.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, GETPIVOTDATA function, and IFERROR function. Also I need to make a data model to perform this analysis.

- **ANALYSIS RESULTS AND VISUALIZATION :**



ANALYSIS- 9

- **INTRODUCTION:**

  How Seasons affect the countries' performance in Olympic Games.

- **GENERAL DESCRIPTION:**

  In this the main motive is to analyse the impact of seasons (summer or winter) on Countries' performance.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart, slicers.

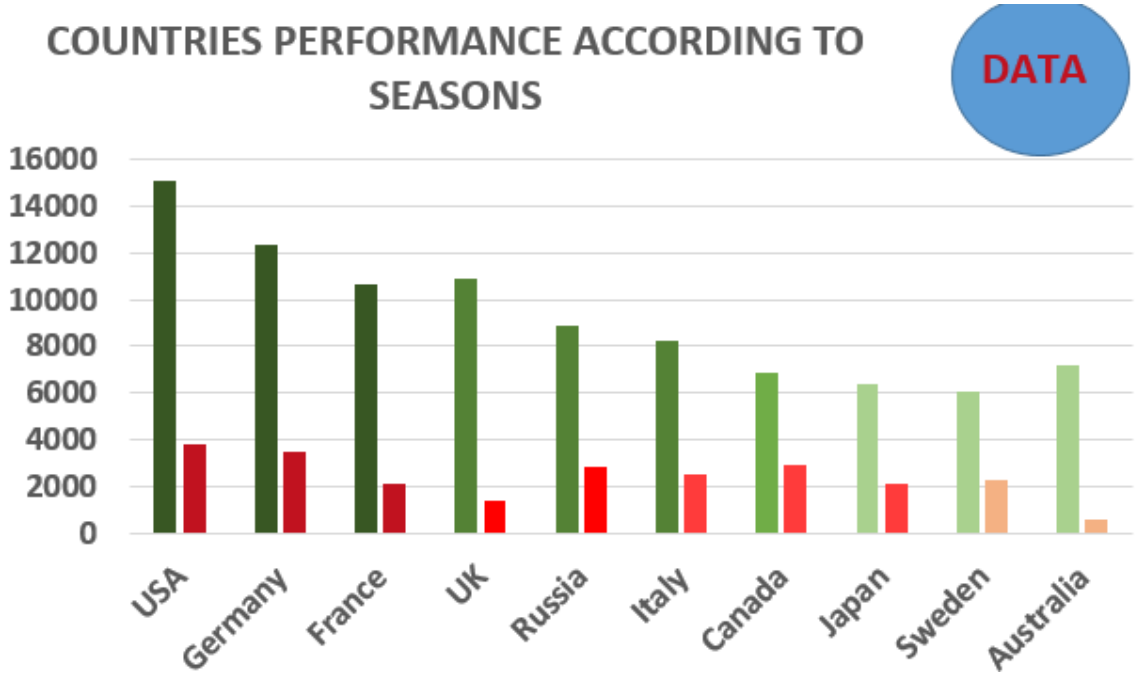
- **ANALYSIS RESULTS:**

  We will talk about the top 10 countries' performance analysis.

COUNTRY	SUMMER	WINTER
USA	15064	3789
Germany	12377	3506
France	10633	2125
UK	10917	1339
Russia	8855	2837
Italy	8217	2498
Canada	6861	2873
Japan	6336	2108
Sweden	6076	2263
Australia	7178	546

  The difference in the medals in winter and summer is because of the number of sports and events as they are way more in summer than in winter. But the interesting fact here what we can observe is that the countries which perform great in summer Olympics are also very well in winter Olympics except for some. Like here for Italy and Japan as their performance in winter Olympics as their performance in winter Olympics is somewhat less satisfying than the following country performance. But these exceptions we can neglect as we are only here looking for approximations.

- **VISUALIZATION:**



We can also check out the impact for specific gender, sports and events.

ANALYSIS- 10

- **INTRODUCTION:**

  Performance of a country according to year.

- **GENERAL DESCRIPTION:**

  In this the main motive is to analyse the performance of countries from 1896 to 2016. This will give us the idea about the growth of a country. This analysis is very helpful as knowing the countries growth and performance we can predict future performance analytics of the country and also taken into consideration if there is any special care to be taken in a particular sports or gender to enhance their performance.

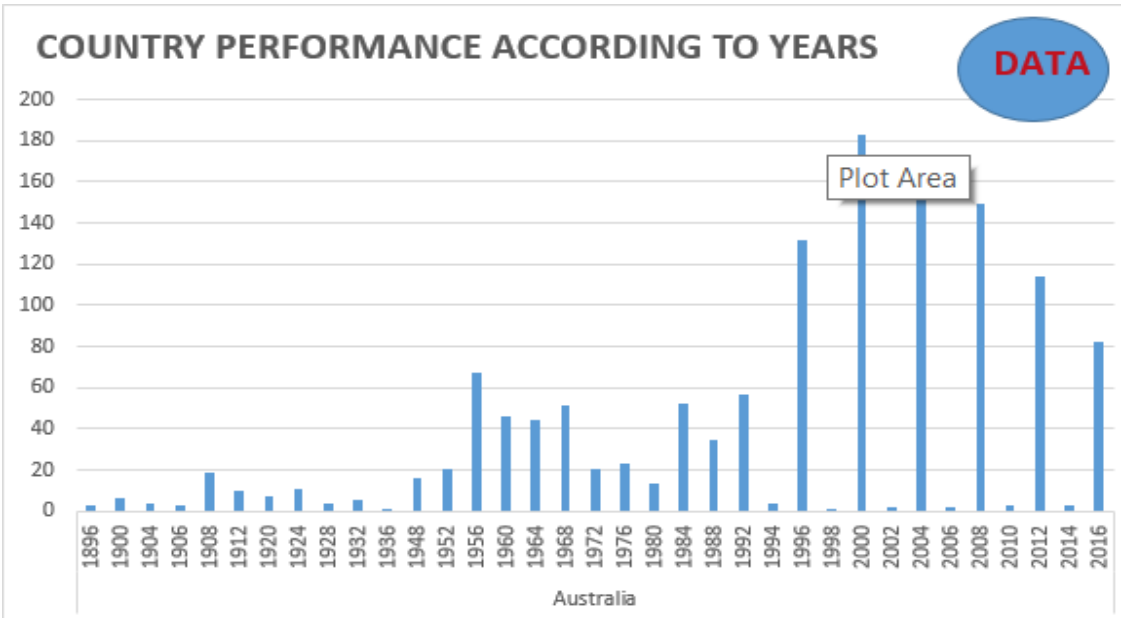
- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart, and slicers.

- **ANALYSIS RESULTS AND VISUALIZATION:**

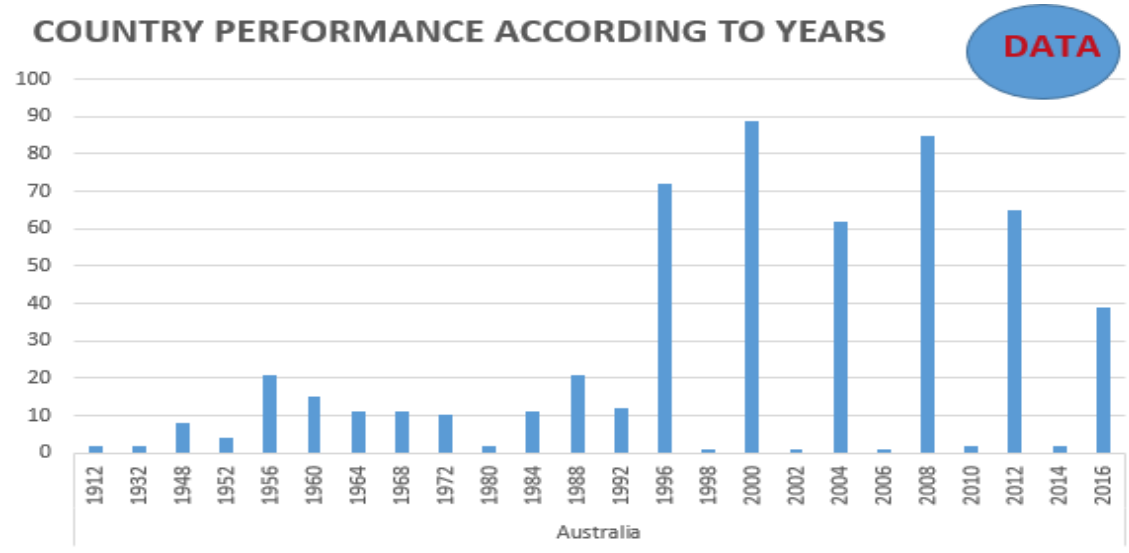
We can chose any country from slicer and analyse its performance throughout the years with additional filters of gender and sports too.

Let us here analyse performance of country Australia.



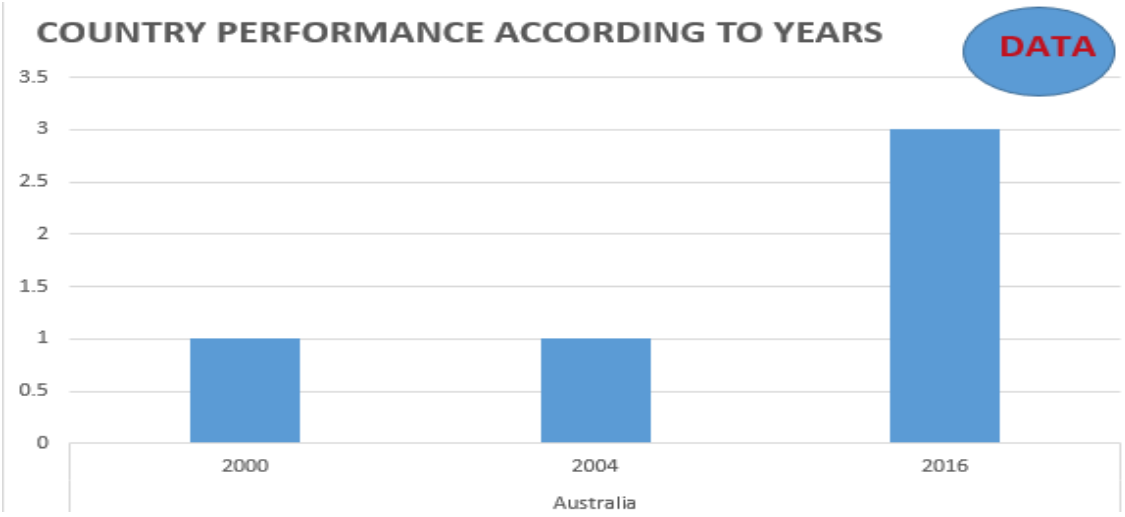
As it is clear from the above visualization that the performance of Australia is not well in the recent years of Olympics, It done somewhat well in 1956 but again the rise and fall in their performance starts but after 1992, when the Summer and Winter Olympics starts to being held in different years, the performance of Australia becomes extremely well in Summer Olympics but the low bars is continuous for them in Winter Olympics. This shows that if the next Olympic is summer then Australia has high chances to win medals but for winters is very less certain.

Talking about the performance of Australia in terms of Female Athletes.



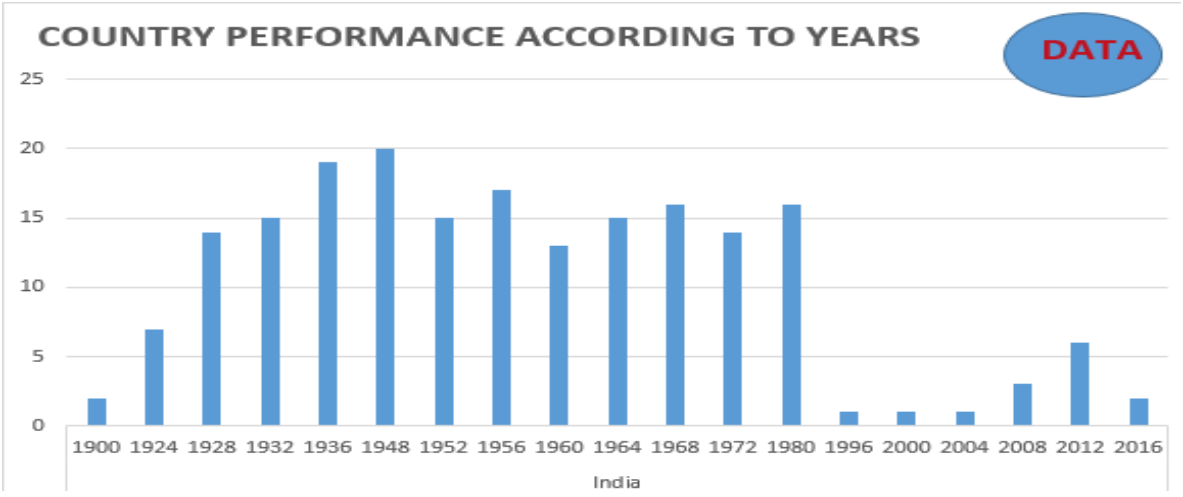
As it is clear from the above visualization that the performance of Female athletes improves from year 1996. One of the reasons that the performance index is very low in recent Olympics is due to the least participation in Olympics because there is not that much awareness and freedom being given to them those years. Also as we see previously that the performance of Australia is not very well in Winter Olympics, same we can see here from the tiny bars in above visualization.

We can also visualize the performance of a country in a particular sports and events. Here lets we see the performance of Australia in Archery.



It is understandable from this that Australia won 1 medal in 2000, 1 in 2004 and 3 medals in 2016 in Archery. This shows that Australia need to work on their Archery players as it is clearly interpreted that the players of Archery not even able to qualify in the rest of the years.

Now let's see the performance of India in Olympics.



The big bars in the starting years shows that the performance of India is quite good than now. But we also need to focus that the maximum medals it has won is only 20 that is in year 1948. From 1900 to 1980 the performance of Indian players is still better but after 1980 it is not satisfying. Here we also need to take into consideration the fact the there are many factors that affect a countries' performance. Also the main point here is that India not even qualify in Winter Olympics. So, from these statistics we can't expect that India would perform extremely well in Winter Olympics in future too. However exceptions will be always there.

ANALYSIS- 11

- **INTRODUCTION:**

  Which Country shows domination in which sport by winning gold medal.

- **GENERAL DESCRIPTION:**

  In this the main motive is to analyse that which country dominate overall in Olympics by winning maximum Gold medals. Also to analyse the dominated country in different sports.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

  For this analysis, I require pivot table, pivot chart, and slicers.

- **ANALYSIS RESULTS:**

  THE DOMINATED TOP 10 COUNTRIES IN OLYMPICS IS:
  USA>Russia>Germany>UK>Italy>France>Sweden>Canada>Hungary>Norway

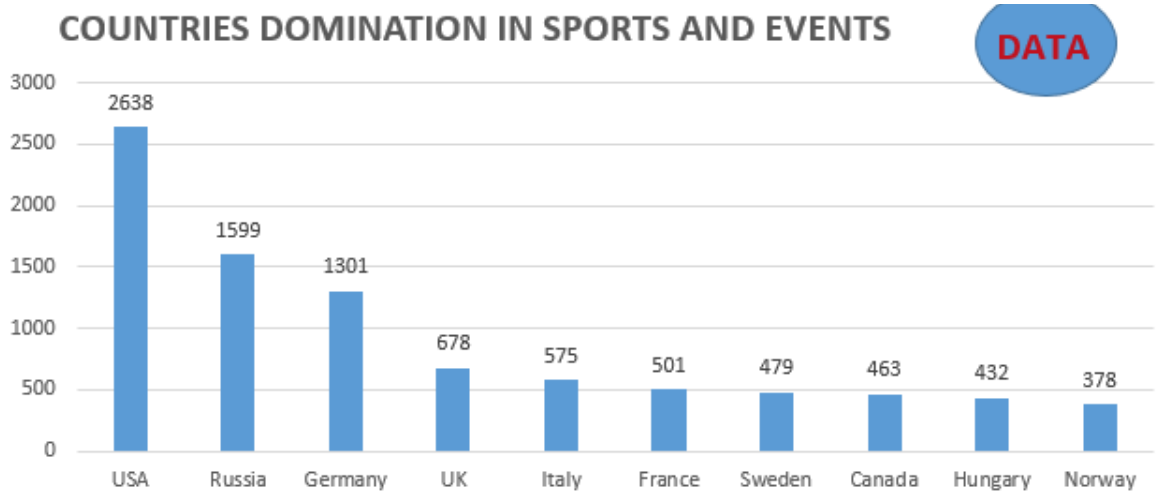
  THE DOMINATED COUNTRIES IN SUMMER OLYMPICS IS:
  USA>Russia>Germany>UK>Italy>France>Hungary>Australia>Sweden>China

  THE DOMINATED COUNTRIES IN WINTER OLYMPICS:
  Russia>Canada>Germany>USA>Norway>Sweden>Austria>Switzerland>Finland>Italy

  THE DOMINATED COUNTRIES IN OLYMPICS IN FOOTBALL:
  USA>Hungary>UK>Russia>Germany>Argentine>Uruguay>Cameroon>Brazil>Poland>Spain>Norway

Similarly we can find out and analyse this for any sport and event. Knowing this will help us prediction about country winning a medal in any sport in future. Like if USA is participating in football in Summer Olympics then after the above analysis we will predict that there are high chances that USA will won Gold medal and if not Gold then either Silver or Bronze with high chances because of its past performance.

- **VISUALIZATION:**



ANALYSIS- 12

- **INTRODUCTION:**

Effect of hosting an Olympic on host country's performance?

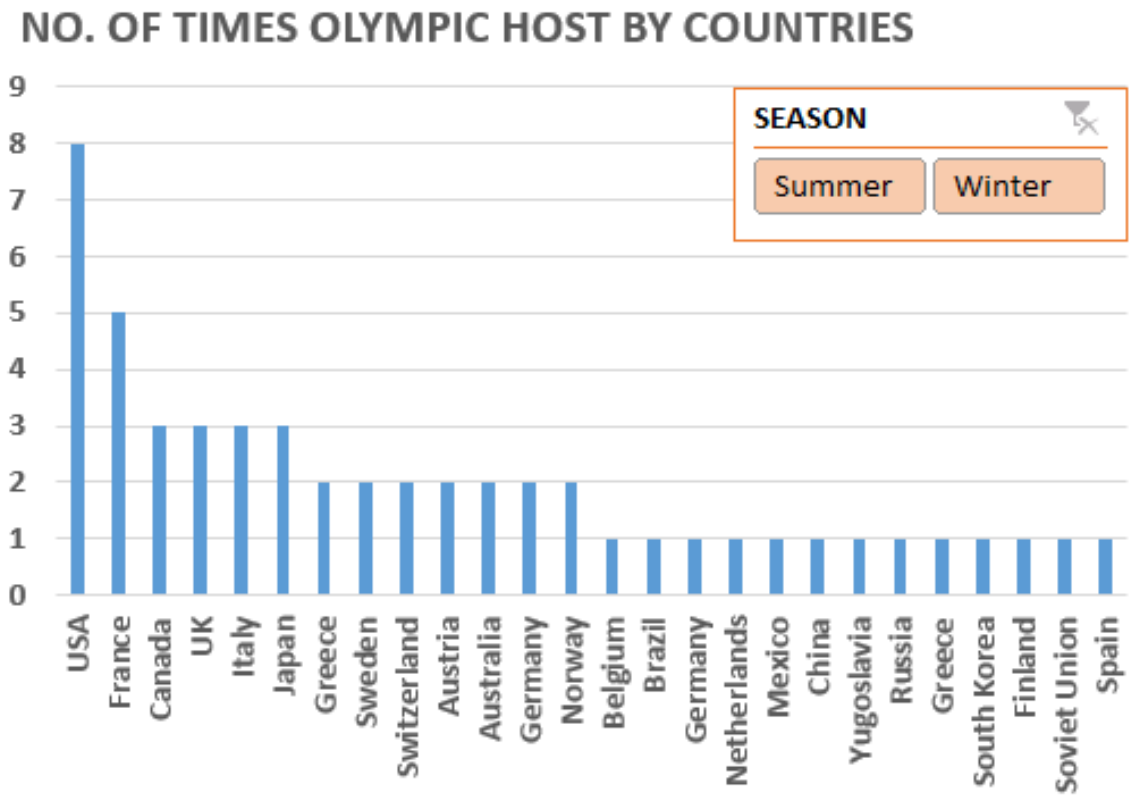
- **GENERAL DESCRIPTION:**

As Olympics happen in different countries and there are many countries that successfully host Olympics. So, here in this analysis we will came to know that if hosting the Olympics impact the performance of host Country. Do they become better after hosting that or do better while hosting it.

- **SPECIFIC REQUIREMENTS, FUNCTIONS AND FORMULAS:**

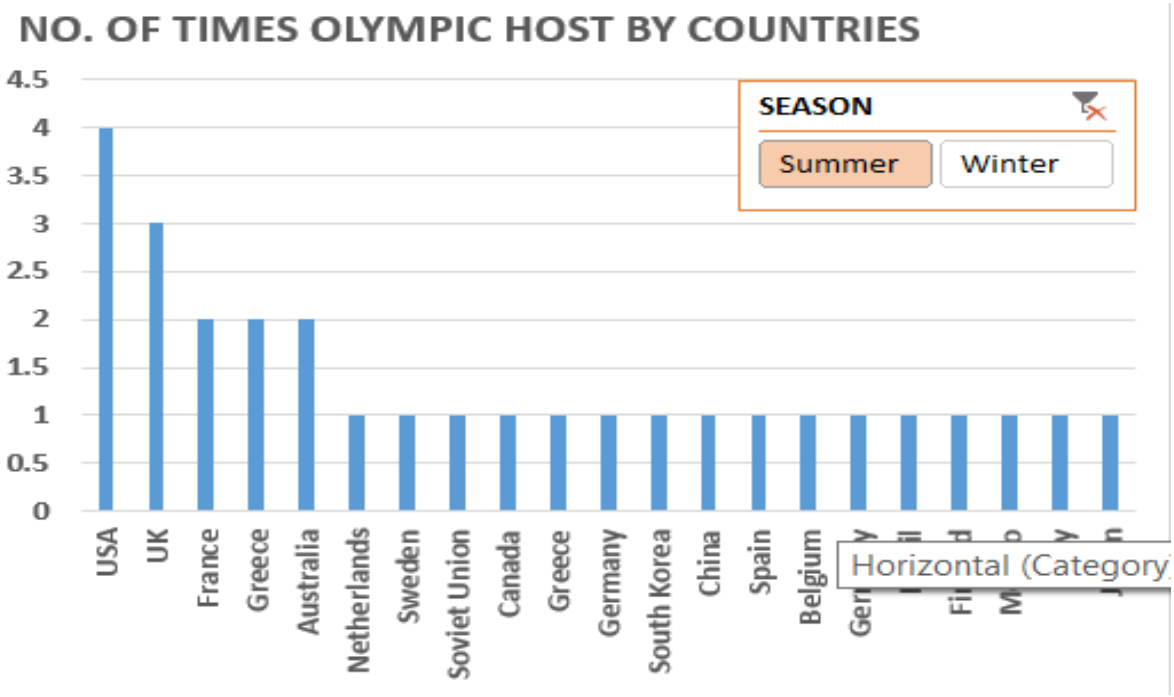
For this analysis, I require pivot table, pivot chart, and slicers.

- **ANALYSIS RESULTS AND VISUALIZATION:**

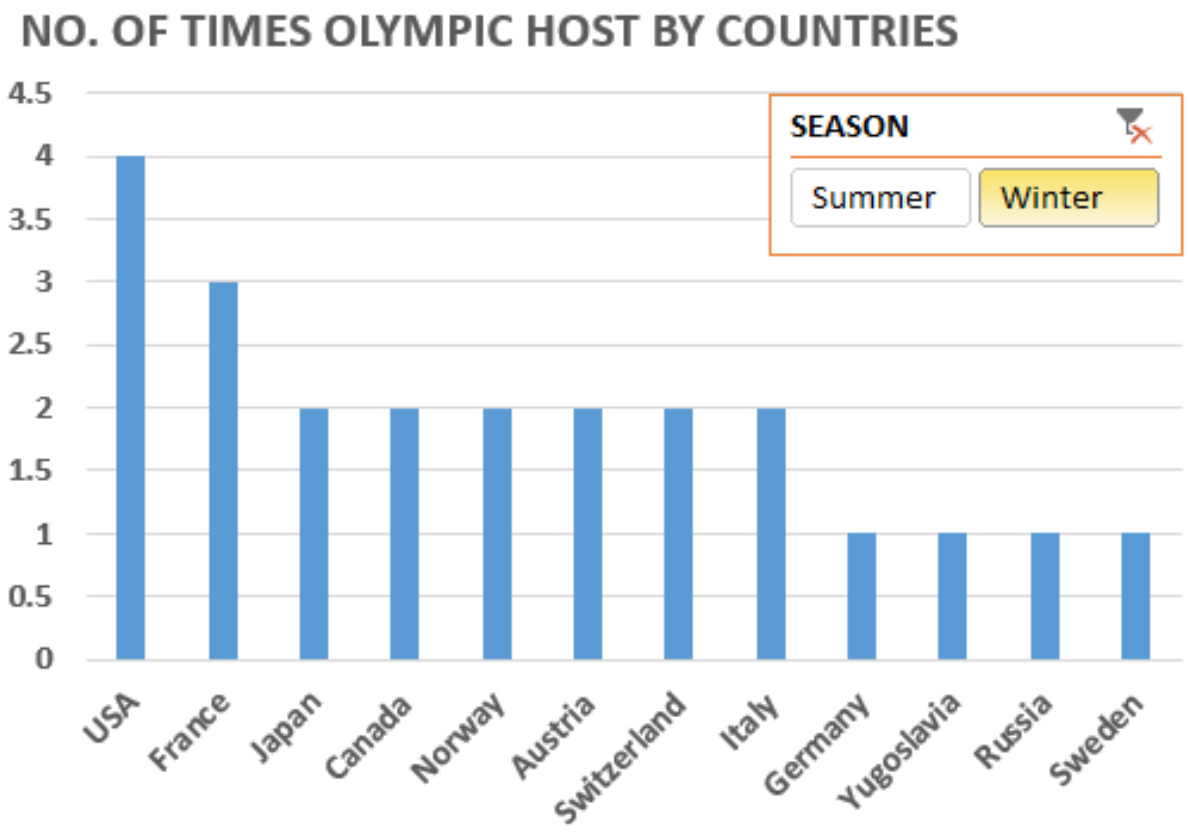


This visualization clearly indicates that USA hosted Olympics maximum number of times. After that it is France, Canada, UK, and Italy and so on.

The country hosted Summer Olympics maximum times is again USA then UK, France.

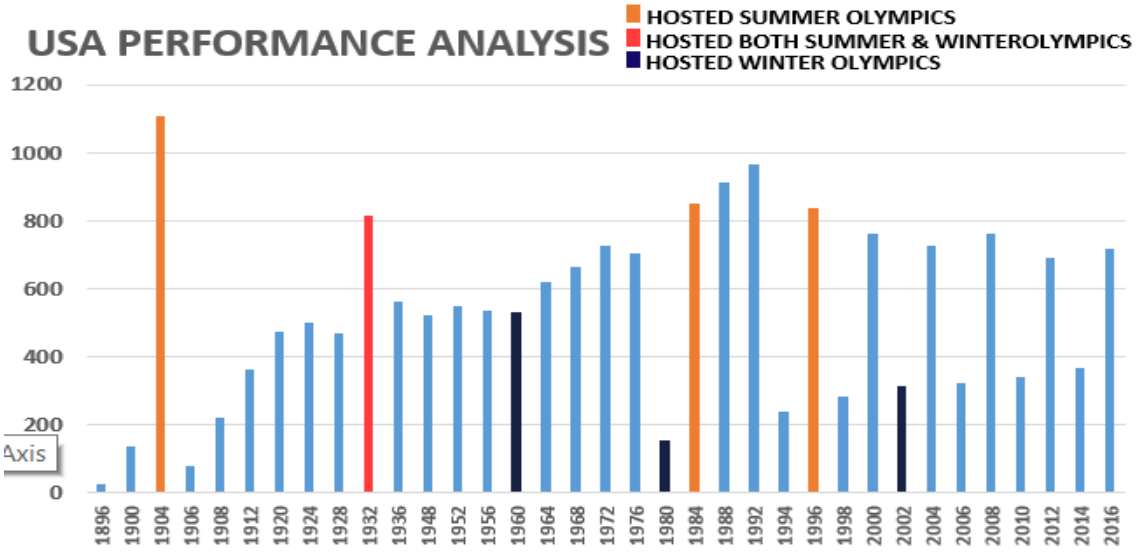


The country hosted winter Olympics maximum number of times is again USA. Then France and Japan.



So as we came to know that maximum times USA hosted the Olympics and then France. So, we will analyse their performance and focus on the difference in their performances during hosting Olympics, after and before hosting Olympics.

USA:

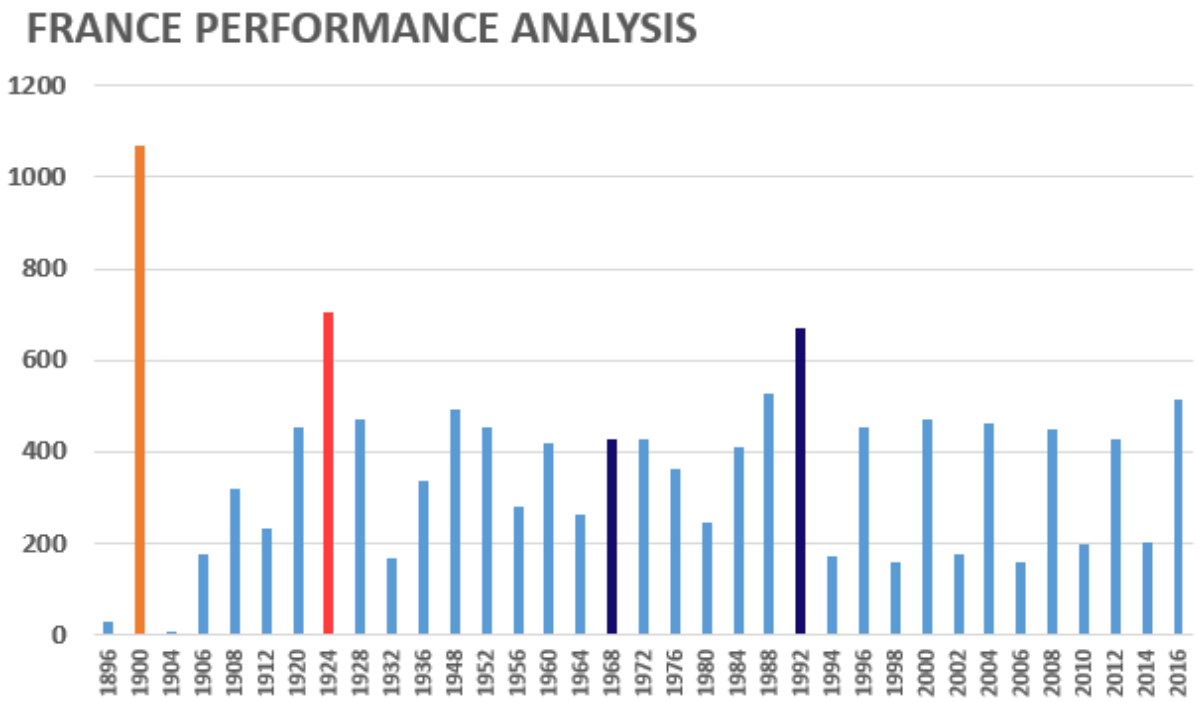


From above the big orange bars clearly showcase that whenever USA hosted Summer Olympics its performance is extremely well and in that year USA has won above 800 medals every time.

The small navy blue bars is for hosting winter Olympics. Although they are not that much giant but still if we compare them to the performance of USA only in winter Olympics then they are also pretty well.

Also one thing that needed to be focus is that in 1904 USA hosted the Olympics for the very first time and before that its performance in Olympics is very poor. But after it hosted the graph of USA arises in every upcoming Olympic event.

France:



Like USA, France being the second country who has hosted Olympics for maximum number of times. Here also we can see the big bars in orange, pink and navy blue colours wonderfully depicting that how well France have performed when it hosted the Olympics.

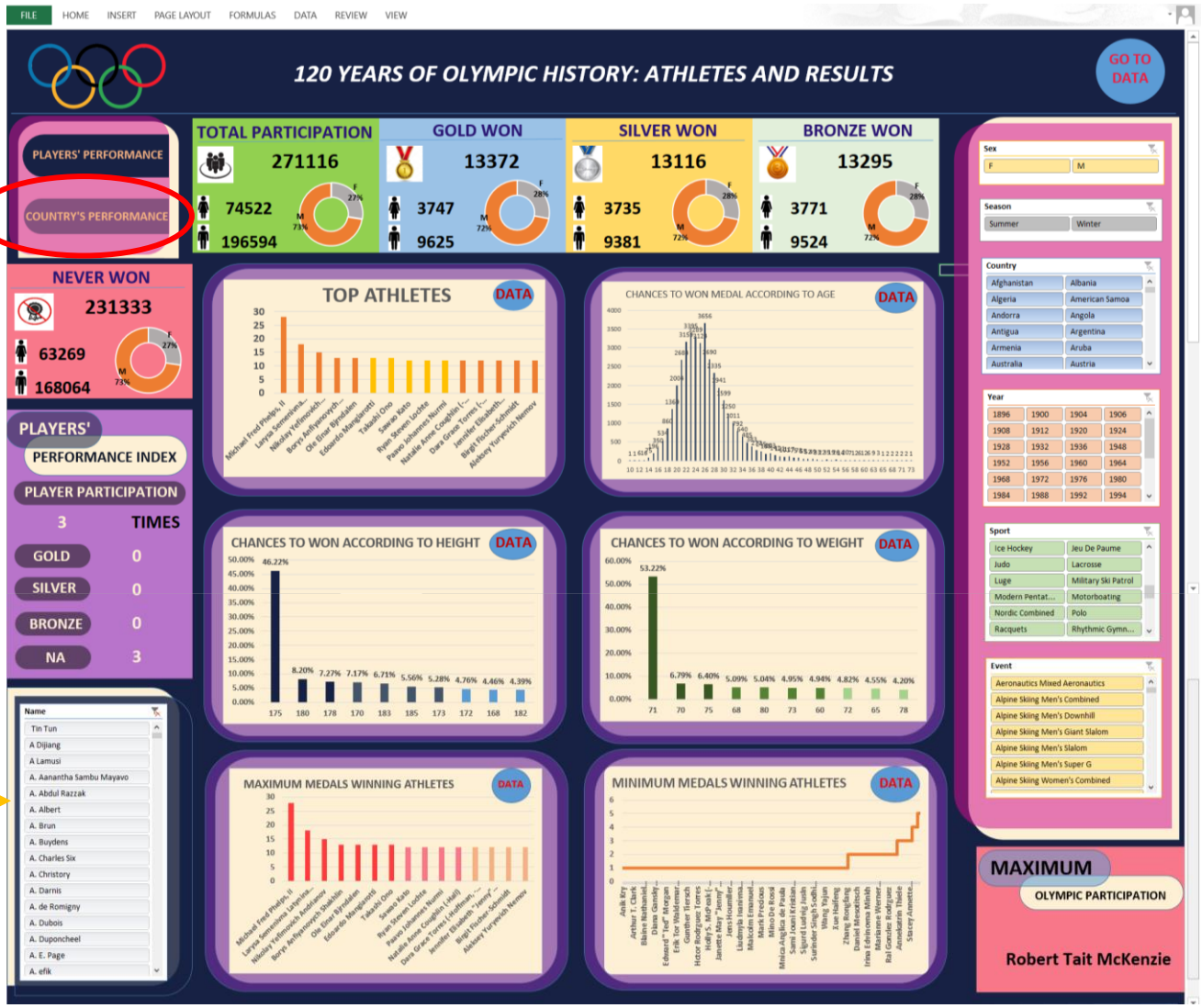
Moreover the performance of France has also being improving in every upcoming Olympic event after it hosted the Olympics.

From all these information we can interpret that **whenever any country host the Olympics in that particular year, their players perform extremely well**. This can be because of the fact

that the players are performing on their home ground. They are comfortable with the environment and familiar with the conditions.

Besides this, **hosting an Olympic impacted the performance of the country in a positive manner.** USA and France example showcase this very well from their performance before hosting Olympics and after hosting that.

LIST OF ANALYSISWITH RESULTS (DASHBOARD)



REFERENCES or BIBLIOGRAPHY

- Learn many things to make the dashboard interactive through surfing on Google.
- YouTube is also beneficial to nourish the concepts even more.