**Analyzing Olympic Performance Data**

**INTRODUCTION**

The Olympics is an international event which is conducted every four years. In the Olympics, people from different countries represent their nations by participating in various sports. The modern Olympic Games began in Athens, Greece in 1896. The International Olympic Committee (IOC) is in charge for Olympics Games, it selects the host city for each edition of the Games. Generally, there are 2 editions which are Summer and Winter. Each edition features different sports respectively. The summer edition of the Olympics includes athletics, swimming, cycling etc. The Winter edition of the Olympics includes events related to snow or ice like skiing, skating etc.

The International Olympic Committee (IOC) maintains the records of Olympic data. The access to this data is restricted to public. It stores the information of athletes, events held and the countries participating in respective years of Olympics. There are other organizations which provide the statistics of these Olympic data that can be collected to analyze. The Olympics also have significant impact in promoting international relations by paying respect towards country’s national sports by including diverse range of sports.

As the Olympics is the biggest event which is held internationally, I am interested in exploring the data and analyzing to draw conclusions. I believe that analyzing these Olympics data would provide the pattern in athletic performance by gender and economic factors that impacts the countries which are participating and the countries that hosts.

The main objective of this project is to analyze the performance of different countries in various sports and the athletes participation at the Olympics over the years. I have collected the Olympics data over the years to draw the conclusions by visualizing them using Tableau.

From visualizing these data, we can get the information on which country has best performance in Olympics over the years by winning the most medals and is there any relation with participation count for the country winning more medals. I believe that analyzing and visualizing these data would provide better decision making in promoting sports and improve athletic performance over different countries. Let’s visualize the data and explore if there are any relationships or patterns for countries performing best in the Olympics.

**Methodology**

For this project I have gathered the datasets on Olympic games from the below data sources:

<https://data.world/msantolini/individual-performance-in-team/workspace/file?filename=Olympics.tsv>

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

<https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/f3d559220372d99a2d5c13a213cf9c2bceb8c6d9/all/all.csv>

The data.world provides the data in the form of tsv, which I have changed to csv for accessing the data for visualization. It gives information of countries, athletes with their age, sports conducted and participated in, and medals won by category over the years from 1896.

The Kaggle provides data on 120 years of Olympic history of athletes including the participants, country and city, sport conduction in each season/edition and the result.

I have used the other dataset from GitHub which gives information on continents to country data. This dataset provides information on total countries which have participated in the Olympics consisting of country code along with the continent and subregion to which it belongs.

Firstly, I have explored the datasets and removed the column (athelete\_unique\_url) which I believe is not required. I have used the columns country names as the related field for building the relationships between the datasets.

I have created calculated fields that calculate the ranking of countries by gold, silver, and bronze medals respectively. I have also created a calculated field that calculates no medals won by countries.

**Calculated fields with calculation are:**

**RANK (SUM (IF [Medal] = "Gold" THEN 1 ELSE 0 END))**

**RANK (SUM (IF [Medal] = "Silver" THEN 1 ELSE 0 END))**

**RANK (SUM (IF [Medal] = "Bronze" THEN 1 ELSE 0 END))**

I have also created a hierarchy for Regions-Subregions which I have used for drill-down in my project. The regions provide the continent information of the country belongs that has participated at the Olympics. Hence it shows continent wide data.

**Analysis**

-> Before getting to actual research questions, I have just explored the data using simple visualization. They are:

* **Athletes Participated by Country:**

The below Visualization provides the information of no. of athletes participating in Olympics by country. From this I could see that the United States of America (USA) has the highest participants and South Yemen has least number of participants in Olympics.

Chart, bar chart

Description automatically generated

* **Count Of Participation by Gender:**

The below visualization provides information of total number of male and female participants participated in each event. From this I could see that majority of participations in events are from male category.

Chart, bar chart

Description automatically generated

* **Number of Medals by Country:**

Chart

Description automatically generated

The above visualization provides the information on the number of medals won by each country. I could see United States of America (USA) has won most no. of medals by any other country. There are many countries which have no medals and I have created other visualizations for that.

* **No. of Medals by Country and Gender:**

The below Visualization provides the information on no. of medals won by countries with respect to gender (i.e male and female). I have considered checking the ratio of medals won for USA as this has most no. of medals won, I could see that male category won more medals than female.

Chart, bar chart

Description automatically generated

**Research question -1:**

* **How old are people on average, by gender, in each sport they participated in?**

The below visualization provides the information on the average age groups of both male and female athletes who participated in sports in Olympics. I could see that for both categories (male and female) the age group remains same which is between 20-25yrs.

Chart, bar chart

Description automatically generated

**Research Question -2:**

* **What is the total number of gold, silver, and bronze medals won by each country?**

The below visualization provides information on the total number of medals won by each country in each category. For visualizing this I have used Scatter plot, I could see that United States of America (USA) has won most medals by medal category. There are many countries that are below the margin of a hundred medals won.

Scatter chart

Description automatically generated with medium confidence

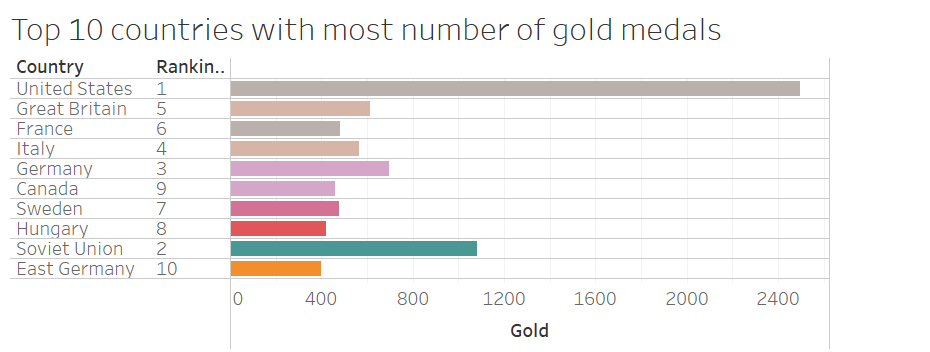
**Research Question – 3:**

* **Which are the top 10 countries that won gold, silver, and bronze medals?**

The below visualization provides information on a list of the top 10 countries that have won gold, silver, and bronze medals respectively. For this visualization I have created a calculated field for calculating the rank for each medal.

I could observe that United States of America (USA) has most medals in Gold, Silver, and Bronze categories. For the 2nd and 3rd position there are Soviet Union and Germany countries in all the categories by having changed their positions for bronze medals won.

I could also observe that most of the top 10 countries were the same in all three categories of medals. When compared between the medals won in three categories between the countries, USA has most number of wins than any other country in the world.



Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

**Research Question – 4:**

* **Which countries have not won any medals over the years?**

The below visualization shows the list of countries that have not won any medals but participated in the Olympics over the years. I could see that there are many countries from the list.

Chart

Description automatically generated with medium confidence

**Research Question -5:**

* **How many Olympic games did the USA participate in each season over the years?**

Chart, line chart

Description automatically generated

As we can see that the USA is dominating over all the countries, I have focused on USA to find the pattern regarding this country.

The above visualization provides the information on no. of Olympics games in which USA participate in each season over years. I have used animation to view this visualization and I could see that the USA has representation in almost all the seasons in all years. I could see that the USA has more participations in SUMMER than WINTER games.

**Research Question - 6:**

* **How many medals in total did the USA win over the years?**

The below visualization provides the information on total medals won by USA in each year. I could see that USA has won the greatest number of medals in the year 1904 which are 394.

Chart, histogram

Description automatically generated

**Research Question – 7:**

* **What is the total number of participations from countries across seasons?**

The below visualizations provide the information on the total number of participations that have been listed from each country across different seasons (winter and summer).

Map

Description automatically generated

Map

Description automatically generated

**Dashboard 1:**

Dashboard is generally used to visualize the combination of charts that have similar activity or related to the same business issue. It interactive and consists of filters which users can access easily to find the pattern or relationship when focused on specific business requirements.

The below dashboard provides the information on total number of participations by each country in each season (winter and summer) respectively.

Map

Description automatically generated

**Dashboard 2:**

The below dashboard provides the information on the top 10 countries winning the medals in gold, silver and bronze categories. This dashboard consists of filter which consists of countries when selected on particular country it shows the contribution of it at Olympics.

Chart, bar chart

Description automatically generated

* **Proportion of Total Medals won by Region:**

I have used Pie Chart to visualize the proportion of total medals won by each region. For this visualization I have used Hierarchy as (Region-Subregion) to know the percentage of medals won by them.

Chart, pie chart

Description automatically generated

**Conclusion**

Finally, after analyzing the Olympic data I could conclude that United States of America (USA) has performed best in Olympics over the years. USA is the country which has won most medals in total. It also stands first in winning the most of number of medals in each category (i.e., gold, silver, and bronze).

To find the relation for USA performing best in each Olympics, I could see that USA has most number of participants every year than other countries. I could also see that it has most number of athletes participating in events in both male and female category. I believe that the USA promotes strong sports culture by having more sports events from school level. It might also have the highest investment from government funding towards the sports.

Additional research questions I would work on:

* What factors contribute to a country's success in winning Olympic medals?
* How has the popularity of different Olympic sports evolved over time?
* Does a country's investment in sports have an impact on encouraging people to participate?