**Factors Affecting Olympic Medal Performance**

**Introduction**

The Olympic Games are enormously important on many levels, including cultural, social, economic, and sporting dimensions. The medal performance of athletes, countries and continents depend on various factors like the sports infrastructure available, economic strength of the country, culture of sports, location of the Games along with the demographic factors of age, gender and the country of the athlete.

In this project, we sought to understand the how the dimensions like gross domestic product (GDP) of the country, age of the athlete, gender of the athlete, location of the Games interplay impact the medal performance.

**Problem Statement**

Through data-driven visuals, we tried to arrive at conclusions for following hypotheses questions:

1. How does GDP of the country affect the total medal count across continents and countries?
2. How does GDP impact the distribution of different types of Olympic medals (Gold, Silver, Bronze)?
3. Does host country has advantage in percentage share of total medals won?
4. Does any particular continent dominate in medal performance across all sports categories?
5. How is the medal distribution varying across continents when gender is added as a factor in the analysis?(that is, if North American men have won higher number of medals compared to men of all continents, have North American women also won higher number of medals compared to women of all continents?
6. Do women gold medal winners have lower average age than men gold medal winners across all sport categories?

**Data Description**

Two datasets, historical data of Olympics from 1960 to 2016 and GDP data from 1960 to 2016 were extracted from Kaggle. The Olympic data contained following major columns:

ID - Unique number for each athlete

Name - Athlete's name

Sex - M or F

Age - Integer

Height - In centimeters

Weight - In kilograms

Team - Team name

NOC - National Olympic Committee 3-letter code

Games - Year and season

Year - Integer

Season - Summer or Winter

City - Host city

Sport - Sport

Event - Event

Medal - Gold, Silver, Bronze

The GDP data contained following columns:

Year – Integer

Country: Name of the Country

GDP: GDP of the Country for that year

**Methodology**

Datasets for analyzing Olympic athletes' performance and GDP data were merged. The missing GDP values with the mean GDP. The updated dataset, which now includes economic data for analysis, was saved as 'updated\_athlete\_events\_with\_gdp\_imputed.csv’. The countries were categorized by GDP. The final dataset, including these categorizations and visualizations, was saved as 'olympic\_data\_with\_gdp\_categories\_v3.csv'. The dataset from 'updated\_athlete\_events\_with\_gdp\_imputed.csv' was analyzed for Olympic performances, grouped by event year, country, and continent, with a focus on medal counts and economic data integration. The consolidated data, aimed at revealing trends across time, geography, and economic contexts, was saved'olympic\_medal\_data\_with\_gdp\_and\_medal\_counts\_v2.csv'

Sports were clubbed into different sports categories like Track Events, Water Sports, etc and the data was saved in ‘olympic event with age.csv’, which was used in analysing gender continental gender disparities among medal winners and age disparity among genders of medal winners. A column of host country status was added to denote whether country was host or non-host for that particular games. Then average percentage share of medals won by country when it was host and average percentage share of medals won by country when it was not host was calculated for all the countries that hosted the games and was added in a separate column. This dataset was saved in the file ‘host\_final\_2.csv’ and was used to analyse whether host country has advantage in share of medals.

**Design Process**

* The KPI part of the dashboard shows the host country, Total number of participants, Number of participants and the number of events for the selected season (Single select) and year (Single select slider). The indicators are prepared by using the marks and filters.
* Scatter Plot of Total GDP vs Total Medals was created using Season, Year, Total medals, Participant continent. Participant Country was added to label mark to display country of each individual points.
* Stacked Bar Chart of Olympic medal distribution (Gold, Silver, Bronze) across different GDP categories was created by using Season, Year, Medal as filter and Medal was put in Color mark to display different colors for different medals.
* Group Bar Chart for Host Country Advantage: To analyze host country advantage in terms of medals won we have created group bar chart with host and non-host country as grouped columns for 7 different countries each belong to one continent we have created grouped bar chart with percent count of medals the country has won when it hosted Olympics with that of percent count of medals when the country was not host. For this visual we have created a filter on country column by creating a context variable to restrict only seven countries to be visible in the graph. We have created different color for host and non-host columns to improve visual effect.
* Tree Map for Continent Dominance: To analyze continent dominance in terms of sports category we have created tree map visual with count of medals as measurable quantity for different sport category and sports category as a radio button filter in the visual. The visual has been assigned a tri color pallet to differentiate between top and lest performing country
* Scatter Plot for Continent Gender Disparities: To analyze continent gender disparities, we have created a scatter plot with male and female filter for gender. The visual has count of medals as measurable quantity for different sport category with sports category as a radio button filter in the visual the visual has continents in X axis and count of medals in Y axis with gender added to color to improve visual effect.
* Bar Graph for Gender Disparities in Age: To analyze average age of men and woman across different sports subcategories we have created a bar graph to analyze average age of men and women across different sports sub category we have created avg age field as measure and a radio button to switch between different sports categories. We have differentiated male and female with different color to improve visual effect

**Key Insights**

* There is a positive correlation between a country's GDP and Olympic medal count, indicating higher GDP countries often achieve more Olympic success. However, exceptions exist where some high GDP countries don't win many medals, and some lower GDP countries excel, suggesting factors like targeted sports investment and national sports culture also play significant roles.
* Nations like the USA and China, classified under 'High-Income Economies', often lead the Olympic medal tally, showcasing their economic strength's role in sports success.
* The countries have won significantly higher average number of medals when it has hosted Olympics as compared with that when country did not host the Olympics. This trend holds true for most of the countries but for Russia and United States the number of medal comparison when it hosted Olympics with that of not hosting is significantly lower.
* Europe was dominating continent for most of sports category followed by Asia while Africa and North America won significantly had less number of medal share.
* European men have won higher number of medals compared to men of all continents and also European women also won higher number of medals compared to men of all continents across all sport categories. It was followed Asia, North America, Africa and South America among men but in women this trend was observed except African women have the least number of medals across all sports categories. This indicates gender barriers.
* Most of the sports subcategory had woman medal winners having less average compared with men. However women medal winners in athletics have higher age than men medal winners in athletics.