Name: Aryan Surve

Student ID: AF0434042

Batch code: ANP-C8256

Project Name: Olympics Data Analysis

Project Guide: Ms. Maseera Jamal Shaikh

# Objectives:

The objective of Olympics data analysis is to extract meaningful insights from historical and current data related to the Olympic Games. This analysis seeks to understand trends in athlete performance across various sports and events, focusing on factors such as age, gender, nationality, and experience that contribute to success. It examines the evolution of country participation, highlighting changes over time and exploring the relationship between a nation’s economic, social, and political context and its performance in the games. Medal distribution is another key focus, aiming to uncover patterns in dominance across nations, regions, and specific sports, while also assessing the influence of host-nation advantages. Additionally, the analysis investigates the popularity and evolution of events, examining the introduction and removal of sports, audience engagement trends, and progress in gender equity within the competition. It also evaluates the broader impact of the Olympics on global sports development, technological advancements, and societal trends, including sustainability and inclusivity.

# Problem statements:

The analysis of Olympics data poses several problem statements that guide the exploration of trends and insights. One key issue is identifying factors that influence athlete performance, such as age, gender, nationality, and training, to uncover patterns that contribute to success in different sports and events. Another challenge lies in understanding disparities in country participation and performance, which may be influenced by economic, social, or political factors, as well as access to sports infrastructure and resources. Exploring the dominance of certain countries or regions in specific sports and understanding the advantages of hosting the Olympics also form critical areas of investigation. A further challenge is analyzing the evolution of events within the Olympics, such as changes in the inclusion of sports and shifts in audience engagement over time. This includes assessing progress in gender equity and identifying areas where inclusivity can be improved. The environmental and societal impacts of the Olympics, including sustainability practices and their integration into the games, also require careful evaluation.

# Data Description:

* **Source**: The dataset was retrieved from Kaggle and contains information about athletes, events, countries, medal counts, historical records and participation trends

# Data Models:

## Entities and Attributes

* + **Attributes:**
* **ID**: A unique identifier assigned to each athlete.
* **Name**: The full name of the athlete.
* **Sex**: The gender of the athlete (e.g., Male or Female).
* **Age**: The age of the athlete during the specific Olympic event.
* **Height**: The height of the athlete, typically measured in centimeters.
* **Weight**: The weight of the athlete, typically measured in kilograms.
* **Team**: The name of the team or country the athlete represents.
* **NOC**: The three-letter National Olympic Committee code for the athlete's country.
* **Games**: The edition of the Olympics, combining the year and season (e.g., "2016 Summer").
* **Year**: The year in which the Olympics took place.
* **Season**: The season of the Olympics, either Summer or Winter.
* **City**: The host city of the Olympics for that event.
* **Sport**: The general category or discipline of the athlete's competition (e.g., Athletics, Swimming).
* **Event**: The specific event in which the athlete competed (e.g., 100m Freestyle, Long Jump).
* **Medal**: The type of medal won by the athlete (Gold, Silver, Bronze, or NA if no medal was awarded).

# Approach:

### Data Import and Libraries:

* + **Libraries Used**: Pandas, Matplotlib, Seaborn, NumPy

### Process:

* + - Import relevant libraries.
    - Load the dataset and inspect its structure for initial insights.

### Data Cleaning:

* + checks missing values, duplicates, or any inconsistencies in the dataset.
  + Standardize values for clarity.

### Exploratory Data Analysis (EDA):

* + Use visualizations (e.g., bar charts, line plots) to identify trends over time and geographical patterns.

# Project Results:

1. **Top Participating Countries**: USA, France, and Great Britain are historically among the most participating countries in the Olympics.
2. **Age Distribution of Athletes**: Athletes between the ages of 20 and 30 are typically the most common participants in the Olympics.
3. **Gender Distribution of Athletes**: The data shows that there are significantly more male athletes (196,594) than female athletes (74,522) in the Olympics, with a male-to-female ratio of about 2.6:1.
4. **Participation of female athletes over time**: The percentage of female athletes has steadily increased, and the gap between male and female participation has narrowed.
5. **Top Countries by Gold medals**: This data shows the USA as the dominant country in terms of gold medals, followed by Russia and Germany.
6. **Best Performing Countries in Rio Olympics**: The United States led by a significant margin, followed by Great Britain, Russia, and Germany. These countries showcased strong performances across various sports.

# Conclusion:

In conclusion, Olympics data analysis provides valuable insights into the dynamics of the world’s most prestigious sporting event. By examining trends in athlete performance, country participation, medal distribution, and the evolution of events, it uncovers patterns and factors that contribute to success and inclusivity. The analysis also sheds light on disparities, highlights progress in gender equity, and evaluates the impact of technological advancements and sustainability practices within the Olympics. These findings not only enhance our understanding of the games but also support informed decision-making for stakeholders, including policymakers, sports organizations, and sponsors. Ultimately, leveraging data-driven insights fosters a more inclusive, competitive, and sustainable Olympic movement, ensuring its continued relevance and global appeal.