

## **Introduction**

The Olympic Games are the pinnacle of international sports, showcasing the best athletes worldwide every four years. This project aims to analyze the performance trends in the Olympic Games, focusing on the 2020 Tokyo Olympics and comparing it to previous summer and winter games as well as any data available for the current 2024 Paris Olympics. By leveraging datasets from Kaggle, we will explore various aspects such as athlete demographics, medal distributions, and performance trends over time.

## **Datasets**

The analysis will be conducted using the following datasets:

1. **124 Years of Summer Olympics Summary**
  - Index: Index #
  - Name: Name of the athletes
  - Sex: Gender of the athletes
  - Age: Age of the athletes
  - Team: Country of the athletes
  - NOC: National Olympic Committee
  - Games: Years with season
  - Year: Year in which Olympics played
  - Season: Season in which Olympics played
  - City: City where Olympics played
2. **124 Years of Winter Olympics Summary:**
  - Index: Index #
  - Name: Name of the athletes
  - Sex: Gender of the athletes
  - Age: Age of the athletes
  - Team: Country of the athletes
  - NOC: National olympics Committee
  - Games: Years with season
  - Year: Year in which olympics played
  - Season: Season in which olympics played
  - City: City where olympics played
3. **2024 Paris Olympics**

**Specific files and variables to be determined**

4. **The 2022 Winter Olympics in Beijing**

**Specific files and variables to be determined**

## Scope of Analysis

As a part of the analysis, the following will be investigated:

1. How do the medal counts compare between the Summer and Winter Olympics?Sandy
2. Can predictions be made on medals won for each country based on past performance?Sandy
3. Insights into gender participation and its evolution over time.-Keli  
Gender by Country and sport- Keli
4. Medal distribution patterns and trends in the 2020 Olympics-MH  
such as: Number of athletes per country, Medal counts per country and Gender distribution among athletes-MH
5. Medals won in this year's 2024 current Olympics by country. -MH
6. Which countries dominated specific sports or events.-MH

## Potential Methods of Analysis

Common analysis techniques will be utilized for the project. These methods include but are not limited to:

- ❖ Data manipulation tools: Numpy and pandas
- ❖ Data Visualization: Bar charts, histograms, pie charts, scatter plots, heat maps.
- ❖ Descriptive Statistics: Mean, median, mode, standard deviation, variance, frequency distribution.
- ❖ Machine Learning: Regression, clustering, decision trees.

**Visualization:** Matplotlib or Seaborn to create some basic visualizations.

- ❖ Number of athletes per country
- ❖ Medal counts per country
- ❖ Gender distribution among athletes

## Tools and Technologies:

- ❖ Pandas: For data manipulation and analysis.
- ❖ NumPy: For numerical operations.
- ❖ Matplotlib and Seaborn: For data visualization.
- ❖ Jupyter Notebook: For interactive analysis and reporting.

## Deliverables and Milestone Schedule

- ❖ Required Deliverables:
  - Project proposal
  - Project report with Jupyter Notebook and program output
  - Project Presentation
- ❖ Proposed Milestone Schedule:
  - Select topic and dataset - July 28
  - Project proposal draft and group review - August 4
  - Project proposal - August 16
  - Project draft and group review - September 1
  - Project report (final draft) - September 8

- Project report - September 13
- Project presentation (final draft) - September 13
- Project presentation (live session) - September 16

### **Potential Roadblocks**

- ❖ NAs or null spaces.
- ❖ Learning to collaborate in Google Colab
- ❖ Error messages using the tools and technologies.

### **Resources**

- ❖ Required Resources:
  - Dataset source from Kaggle
  - Python programming training and references