Olympic Data Analysis

Project Synopsis

Shubh Sharma

21013570096

Introduction

As we know, The Olympic Games are considered the world's foremost, largest, and ancient sports competition in which sportspersons from all over the world participate and strive hard to create a footprint in the Olympic Games. Presently more than 200 countries participate in this competition. The Olympic Games are normally held every four years.

In our Olympics dataset we are provided with the names of participants, their country name, year and date of participation in the Olympics, which sport they participated in, and Olympic Medals they have won to make their country proud. The given data set only includes the participants who have won at least one medal in any sport in the years 2000 to 2012. We will use Python to perform our analysis. The column names of the dataset are self-explanatory in order to provide information.

Objectives

The project aims to achieve the following objectives through a comprehensive analysis of the dataset:

1. Conduct a basic exploration of the data frame using pandas and Numpy libraries.
2. Address missing values in a manner suitable for the dataset.
3. Manage and mitigate biases introduced by duplicate values in the study.
4. Utilize data visualization techniques on relevant columns, providing clear explanations.
5. Apply appropriate numerical measures to each column and present detailed reports.
6. Identify and list the top 10 countries with gold, silver, bronze, and total medals, offering insights into Olympic success.

Data Collection, Analysis, and Visualization

The dataset for this analysis was collected from [this source] https://app.gigasheet.com/spreadsheet/olympicwinners-export-csv/ef65340c_0cf1_4ae0_be65_

The dataset comprises several key columns, each offering valuable insights into the accomplishments of Olympic participants from 2000 to 2012. Here's a brief elaboration on the notable columns:

1. *Name:* - Represents the names of the athletes who have achieved Olympic success during the specified years.
2. *Country:* - Indicates the countries to which the athletes belong, providing a basis for exploring national performance.
3. *Year:* - Specifies the year in which the participants competed, allowing for a chronological analysis of Olympic achievements.
4. *Sport:* - Identifies the specific sports in which the athletes participated, offering granularity in understanding their diverse achievements.
5. *Gold Medal, Silver Medal, Bronze Medal, Total Medal:* - Quantifies the number of gold, silver, bronze, and total medals won by each participant, enabling a comprehensive assessment of their overall performance.

For this data analysis project, we utilize Python as the main programming language, leveraging libraries like pandas and NumPy for efficient data manipulation. Matplotlib and Seaborn aid in visualization, while Jupyter Notebooks provide an interactive environment for transparent and reproducible analysis. These technologies collectively enable a streamlined process for exploring, cleaning, and deriving insights from the Olympic dataset.

Results and Insights

Our analysis of the Olympic Games from 2000 to 2012 showed some interesting things -

1. Athletes were usually around 26 years old on average.
2. There were a total of 49 different sports during this time.
3. The United States did really well, winning the most medals.
4. Russia and Germany also did great in the Olympics.
5. India won 11 medals in different years and sports.
6. Abhinav Bindra was the only Indian to win a gold medal.

The findings align with our initial objectives -

1. *Data Exploration:* - Average athlete age and the diversity of 49 sports align with exploring the dataset.

2. *Handling Missing Values and Duplicates:* - The comprehensive analysis inherently addresses data quality concerns.
3. *Data Visualization:* - Insights on top-performing countries and individual achievements fulfill the visualization objective.
4. *Numerical Measures:* - Analysis of India's medal count and Abhinav Bindra's gold align with performing numerical measures.

The findings not only meet but surpass the set objectives, providing a holistic understanding of Olympic achievements.

Conclusion

What it means:

1. Knowing the average age helps understand when athletes perform best.
2. The variety of sports shows how diverse the Olympics are.
3. The success of countries like the United States can guide others in sports.
4. India's achievements suggest where to focus for future success.

Why it matters:

1. It shows how people from all over the world come together in sports.
2. The information helps make smart decisions for sports planning.
3. It inspires new athletes to dream big and work hard.