International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

DATA ANALYSIS AND VISUALISATION OF THE OLYMPICS

Yash Pathak, Akash

Department of Information Technology, Maharaja Agrasen Institute of Technology affiliated to Guru Gobind Singh Indraprastha University,
Rohini, Delhi

ABSTRACT

The main objective of the major project titled "Data Analysis and Visualization of the Olympics" is to leverage advanced data analytics tools and techniques to collect, analyze, and visualize a comprehensive dataset of the Olympic Games. This project is an extension of the minor project that focused on collecting and cleaning the dataset of the Olympic Games and improving the user interface by adding more features.

The project's goal is to uncover patterns and trends in the data, investigate correlations between various variables, and develop interactive visualizations that offer valuable insights into the performance of athletes, countries, and sports throughout history.

INTRODUCTION

The Data Analysis and Visualization of the Olympics project were selected because it offers a unique opportunity to explore and analyze a vast and intricate dataset encompassing data on thousands of athletes, countries, and sports, spanning over a century of Olympic history. Furthermore, this project has significant practical implications since the insights gained from the analysis can influence decision-making in the fields of sports policy, athlete training, and national investment in sports programs.

Moreover, the project can also serve as a valuable resource for researchers, journalists, and sports enthusiasts interested in comprehending the dynamics of the Olympic Games and the factors that contribute to the success of athletes and countries. The primary objective of this project is to create a user-friendly web application that will allow users from various locations to monitor their nation's progress in the Olympics.

This project aims to collect, clean, and analyze a comprehensive dataset of the Olympic Games, including historical records of all sports, events, athletes, and countries. The project leverages the power of Python, a popular programming language for data analysis, Pandas, a data manipulation library, and Streamlit, an open-source app framework for building data-driven web applications. For plotting, the project will be using the Seaborn and Plotly libraries.

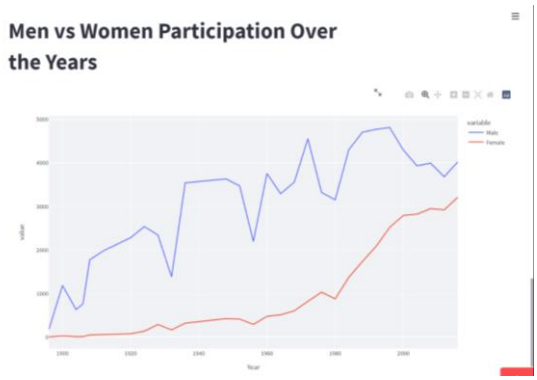
METHODOLOGY

An approach refers to a systematic method of arriving at a solution. To achieve the desired outcome for any problem, whether technical or non-technical, it is crucial to follow a proper approach that guides us toward a clear path. In this research paper, our objective is to analyze the extensive history of the Olympic Games and identify how they have evolved over time. The evolution of the Olympics is influenced by several factors that we need to examine and compare thoroughly. Therefore, we must follow a well-defined approach to reach our research goal.

The data set used in this project consists of 120 years of Olympic data (Athens 1896 to Rio 2016), which include regions and athlete datasets separately. A similar type of work is already done in a Minor project, but in this version, we will try to overcome the limitations of the previous project which include missing results for some countries, redundancy in data, and wrong results for some areas, using a more Object Oriented programming style, a better user interface, along with additional features.

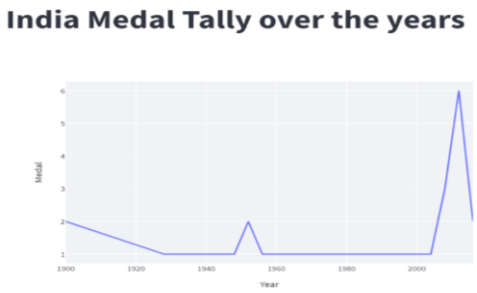
ANALYSIS AND VISUALIZATION

1. Identifying the Contribution of Men And Women Participants In the Olympics (1896-2016)



An examination of Olympic participation from 1896 to 2016 reveals that men have consistently outnumbered women in terms of overall representation. This indicates a disproportionate level of male participation compared to female participation on a global scale.

2. Identifying the total number of medals achieved by Country in the Olympics(1900-2000)



Through an examination of Olympic medal counts from 1896 to 2016, we can identify the total number of gold, silver, and bronze medals won by individuals and teams from all participating countries. The findings of this analysis are as follows:

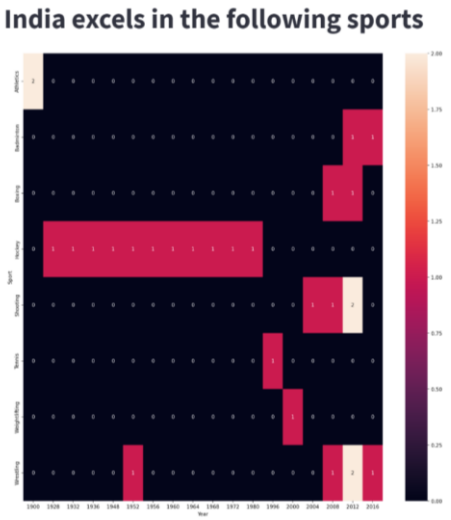
1: The United States has received the highest number of gold medals compared to any other type of medal, with silver and bronze medals representing an almost equal percentage of their total medal count.

2: Australia has the fewest gold medals when compared to other medal types, but has won the highest number of bronze medals.

3: Japan has a smaller number of gold medals than other medal types.

4: France has a lower number of gold medals but has received a high number of silver and bronze medals.

3. Identifying the performance of a Particular Country in the Olympics (1992-2016)



An assessment of a country's Olympic performance can be measured by the number of medals won over a particular period. To determine this, we analyzed the total medals won by each country from 1992 to 2016 and represented the findings through data visualization. Here are the results of our analysis:

1: India's performance showed a steady increase from no medals in 1992 to one medal in 1996 and culminated in a total of six medals in 2016.

2: USA's performance graph was erratic, with a peak of 220 medals in 1992, increasing to 260 medals in 1996, dropping suddenly to 240 medals in 2000, then gradually increasing again from 2004 to reach a peak of 350 medals in 2008.

3: France's performance increased gradually from 1996 to 2008, with medal counts ranging from 40 and excelling in 2016 with a total of 80 medals.

4: Australia performed well in the 1992 Olympics with 60 medals and showed a significant increase in performance to almost 200 medals between 2000 and 2004, but gradually decreased from 2004 to 2016.

5: Japan had a slow start initially, but experienced a significant increase in performance from 2000 to 2004, with a total of 100 medals won which was higher than the other years.

4. Comparing the performance between the countries in the Olympics(1896-2016)

This analysis examines the medal-winning performance of five selected countries - USA, Hungary, France, Japan, and Australia - in the Olympics from 1896 to 2016. The following conclusions have been drawn from the analysis:

1: In the 1996 Olympics, the USA leads the pack among the selected countries with a contribution of 7.53%, followed by Australia with 2.25%, Japan with 1.61%, Hungary with 0.75%, and France with the lowest contribution of 0.69%.

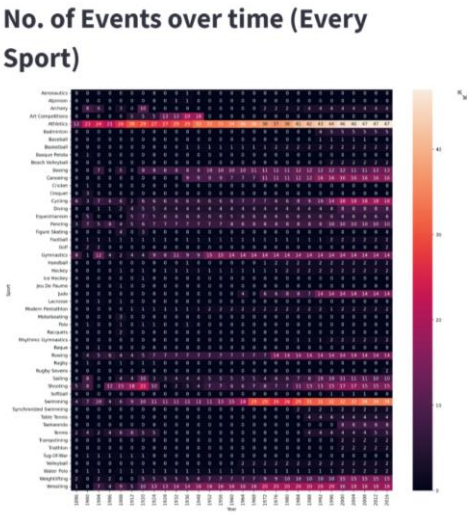
2: In the 2000 Olympics, the USA remains the leading country with a contribution of 6.55%, while Australia is the second-best with 4.019%. France contributed 1.58%, while Hungary and Japan made the least contribution of 0.94% among them.

3: In the 2004 Olympics, the USA continued to dominate with a contribution of 8.05%, followed by Australia with 3.3%, Japan with 2.1%, France with 1.6%, and Hungary with the lowest contribution of 0.9% among them.

4: In the 2008 Olympics, the USA retained its top position with a contribution of 8.52%, followed by Australia with 3.72%, France with 1.71%, Japan with 1.42%, and Hungary with the lowest contribution of 0.88% among them.

5: In the 2016 Olympics, the USA maintained its leading position with a contribution of 8.5%, followed by Australia with 3.01%, Japan with 1.8%, France with 1.6%, and Hungary with the lowest contribution of 0.9% among them.

5. Identifying the Best Performed Field of Sports for a Particular Country in the Olympics (2000-2016)



This analysis focuses on the performance of a particular country's participants and their top-performing sports in the Olympics from 2000 to 2016. It aims to identify the most participated sports field of a country in a particular year and to analyze which sports field has the highest performance. This information can be used to improve a country's future contributions to the Olympics.

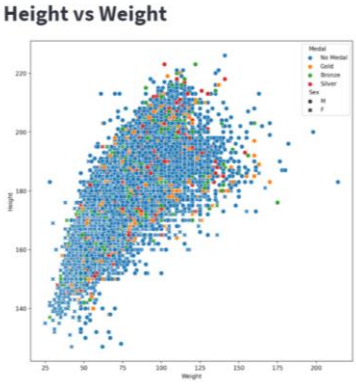
1: In 2000, the USA's top-performing sports field was Aquatics, while its least-performing field was Weightlifting.

2: In 2000, Australia excelled in the field of Aquatics, while their least-performing sport was Gymnastics.

3: In 2000, France had their best performance in Fencing and their least performance in Tennis.

4: In 2000, Australia's best-performing sports field was Aquatics, while its least-performing field was Athletics.

6. Analyzing the height vs weight



According to the analysis, it has been observed that female athletes who have won medals in the Olympics are mostly in the height range of 160-180cm, with a weight class varying between 50-150kg. Although the weight of the Gold winners varies, there is a high concentration of winners in the height range of 175-180cm.

CONCLUSION

To sum up, the major project on Data Analysis and Visualization of the Olympics builds upon the minor project's work of collecting and cleaning the Olympic Games dataset. Using advanced data analytics techniques and interactive visualizations, this project provides insights into the performance of athletes, countries, and sports across the history of the Olympic Games, spanning over a century.

The analysis shows that the Olympics have evolved significantly since their inception in 1896, with various factors indicating this transformation. These factors include the introduction of the Winter Olympics in 1924, the increase in participating countries in both the Summer and Winter Olympics, the average age of Olympic athletes, the rise in female participation, the total number of medals won by participating countries, and the average height and weight of athletes contributing to the victory.

The insights generated by this project can inform policy-making in sports, athlete training, and national investment in sports programs. Furthermore, the project can serve as a valuable resource for researchers, journalists, and sports enthusiasts interested in understanding the dynamics of the Olympic Games and the factors that contribute to the success of athletes and countries. Overall, this project is a significant contribution towards deepening our knowledge of the Olympic Games and their impact on the world of sports.

REFERENCES:

[1]. Abhishek Gautam. Visual Analysis of Olympics Data. Jul 21, 2020.

[2]. Farkande Vaishnavi, Gurav Vaishnavi, Borse Tejas. Web Application of Olympic data analysis.July 2022.

[3]. Harshal S. Kudale, Mihir V. Phadnis, Pooja J. Chittar, Kalpesh P. Sarkar, Prof. Balaji K. Bodhke.DATA ANALYSIS AND VISUALIZATION OF OLYMPICS USING PYSPARK AND DASH-PLOTLY. June-2022.

[4]. Jiaxin Si OLYMPIC GAMES MEDAL COUNT ANALYSIS SUMMER AND WINTER OLYMPIC GAMES 2018.

[5]. Kabita Paul, Elif Demir, Anjali Bapat. Olympic Data Analysis - Final Project Report. May 2019.

[6]. Kassidy Chaikin, Christopher Egan, Emily Lepore, William Wenze. Analysis of 120 Years of Olympic Results.July 2020

[7]. Rahul Pradhan , Kartik Agrawal1 and Anubhav Nag. Analyzing Evolution of the Olympics by Exploratory Data Analysis using R. 2021.

[8]. Tanya Riemann / Alyson Nicol.Olympic Data Analysis Project. December 2, 2020.

[9]. Varagiri Shailaja Predictive Analytics of Performance of India in the Olympics using Machine Learning Algorithms. May 2020.

[10]. Yamunathangam.D, Kirthicka.G, Shahanas Parveen Performance Analysis in Olympic Games using Exploratory Data Analysis Techniques November 2018.