International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1425-1429 www.ijaem.net ISSN: 2395-5252

Olympics Data Analyzer with Prediction

Hitanshi Shah, Jay Sheth, Hetvi Savla, Jyoti Bansode, Bijal Patel, Aruna Yewale

Information Technology, Shah and Anchor Kutchhi Engineering College, Mumbai, India
Information Technology, Shah & Anchor Kutchhi Engineering College, Mumbai,
Information Technology, Shah and Anchor Kutchhi Engineering College, Mumbai, India
Information Technology, Shah & Anchor Kutchhi Engineering College, Mumbai, India
Information Technology, Shah and Anchor Kutchhi Engineering College, Mumbai, India
Information Technology, Shah and Anchor Kutchhi Engineering College, Mumbai, India

Date of Submission: 15-04-2023 Date of Acceptance: 26-04-2023

ABSTRACT—Olympic Games are a major international event and a source of pride for nations. More than 200 countries compete in a variety of sports at the Olympic games, which are an international sporting event. Sportspeople from different nations compete and represent their nations with pride thanks to their superior athletic ability. The main goal of this work is to use Python to analyze the Olympic dataset and evaluate overall performance of countries and to assess what each one brought to the Olympics. These studies enable athletes to immediately assess their own and their opponents' performances while providing more insight into how various nations have performed at the Olympics over time. In order to compare the performances of different nations and their contributions to the Olympics, exploratory data analysis approaches are employed in this research. The status of countries in the Olympics is provided by the visualization of the Olympic dataset in many ways, and it assists nations with bad performances in the Olympics by producing quality players.

Despite putting in a lot of effort, several countries or players are unable to perform well during the events and win medals, while many other countries perform exceptionally well and win numerous medals. Each nation should do a study of its previous data to identify any errors they may have made in the past and to aid in their future development. The statistical picture of the many elements that contribute to the evolution of the Olympic Games and Improvement in the performance of various Countries/Players over time will be given to us by the visualization of the data over various factors. This research paper's main goal is to examine the extensive Olympic dataset using exploratory data analysis to assess how the Olympic Games have changed throughout time.

Keywords—Olympics, Sports, Nations, Machine learning, dataset

I. INTRODUCTION

The Olympics is considered as the most important event worldwide, which provides a common platform to players from various nations to show their talents. The Olympics is a global sporting event that brings together athletes from around the world to compete in various sports. Held four years, with both summer and winter games celebrated as one of the largest and most prestigious sporting events globally in the world. Analyzing the data can provide valuable insights into athlete performance, medal counts, historical trends, and other factors that influence the outcome of the games. Exploratory data analysis is a widely used approach in analyzing Olympic datasets, which involves statistical and visual representations of the performance of nations and individual athletes in the games, researchers can also gain the deeper understanding of the performance of player and nation in the Olympics. [1] When we investigate the Evolution of Olympics Games over the years, various scenarios come to mind. These scenarios include an increase in the number of participating nations and athletes, changes in the number of events, changes in the expenditure cost of the event, improvements in the performance of specific countries or players, an increase in women's participation, the participation ratio of men to women, improvements in medication facilities during competition, and the effect of pandemics on player performance. It helps to understand the evolution of the Olympics over time and make future predictions. Analyzing these scenarios provides us with valuable insights into the past, present, and future of the Olympic Games [2] The effective use of past performances data. However, data collection and cleaning present

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1425-1429 www.ijaem.net ISSN: 2395-5252

significant challenges in any Data Mining Project. However, data collection and cleaning present significant challenges in any Data Mining Project. Once data is collected and sorted, model building becomes relatively straightforward. Having a reliable source for that purpose is always an added advantage but the process can be tedious.

For this project, we collected the datasets from Kaggle. To ensure a comprehensive analysis, we focused on the top ten most popular sports in the United States, as determined by their average viewership ranking. By prioritizing these sports, we aimed to provide a thorough and insightful analysis of past Olympic performances in the most widely followed sports. By comparing the performance of each sport with others, we can gain a better understanding of the strengths and weaknesses of each country and sport. This information can be used to identify the fields of sports that require more participation and the necessary actions that players and nations can take to improve their future contributions to the Olympics.

II. LITERATURE SURVEY

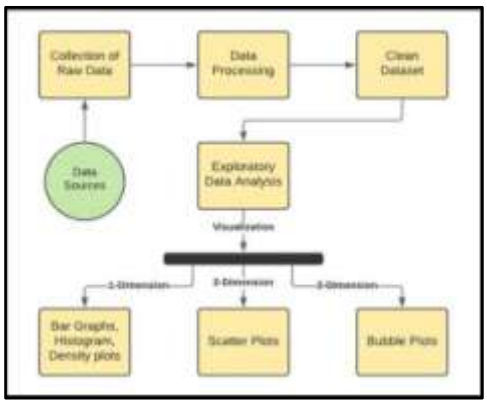
Inspecting and analysing numerous study papers allowed us to learn about several fresh methods and procedures. We discovered Heuristics, which employed machine learning algorithms to forecast a country's overall number of Olympic medals. We also learned that it is possible to gauge a country's level of success through effective research and the importance of sports in society. One can predict a country's performance at the Olympics based on its historical results. The likelihood of someone getting a medal at the following Olympics is calculated if they win one this year.If they need to improve in any areas, they can do so, and joining the right training program will have a significant impact on their results. In addition to these approaches, a procedure known as exploratory data analysis was utilized to break down the data statistically and give it a full understanding. The interpretation and analysis of data are one of primary duties in the field of big data analytics. The Olympic Games were subjected to a wide range of investigations, including statistical visualization, player performance evaluation, changes in the performances of various nations, and many more. Therefore, we concluded that Exploratory Data Analysis is a common and useful technique for assessing how the Olympics have changed through time.

III. PROPOSED SYSTEM

Our proposed system for Olympics data analysis will involve collecting and cleaning large volumes of datasets from Kaggle. Once we have collected and organized the data, we will leverage exploratory data analysis techniques to extract meaningful insights into the performance of athletes, countries, and sports. We will use various visualization techniques to present the data in a visually appealing and user-friendly format that is easy to comprehend. These visualizations will enable us to identify patterns, trends, and outliers that may not be evident in raw data. Our aim is to provide a comprehensive understanding of the Olympic data that can be used to inform decisions and strategies for athletes, countries, and sports. To ensure reliable results, our project will gather data from previous Olympic seasons for model training and testing. The outcomes of our analysis can be used to prepare a report that motivates countries to improve their performance in prestigious sporting events. Our project aims to achieve high accuracy and specific insights through the concept of correlation. This approach will help us understand whether the features that directly impact performance are the only ones to consider, or if there are other factors that analysts may have overlooked. Ultimately, our project seeks to provide actionable insights that will help countries improve their performance in the Olympics and other sports events and make informed decisions about the future of the game.

IV. METHODOLOGY

An approach is a methodical approach to finding a solution. Every issue, whether technical or not, needs to be approached properly in order to determine the best course of action for achieving the desired outcome. This research paper tries to examine the extensive history of the Olympic Games and evaluate how they have changed throughout time. The evolution of the Olympics is influenced by several variables. We used the following methodologies to create the Olympics data analysis:



International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1425-1429 www.ijaem.net ISSN: 2395-5252

1) Data Collection: Data collection is the initial step in the analysis process. We need a lot of data to undertake analysis, and then we use different approaches and algorithms to get our desired conclusions from the data. For our analysis on the evolution of the Olympics across time, we used data from Kaggle. Three datasets have been collected. The first dataset contains facts about the players, including their gender, height, weight, country of origin, medals earned (Gold, Silver, and Bronze), and many other things. This information can be used to evaluate a specific player's performance and to compare the performances of two or more players. The second dataset contains statistics on the nations that have so far competed in the Olympics, together with a list of the total amount of medals (Gold, Silver, and Bronze) they have earned. The third dataset includes a list of nations together with their country codes, which serve as the nations' identifiers.

2) Data Pre-procession: Data processing comes next after data collection. By constantly inspecting for flaws and removing superfluous, incomplete, or wrong data, data preprocessing transforms raw data into useful data. The dataset includes several fields, including Age, Gender, and others, some of which include null values, which leads to problems in the final product, which is the graphical visualization of the data. To finish this assignment, we employed a process called Deterministic Imputation. In deterministic imputation, the null values (NA or Nan) are computed using the data from the other values in the same column. There are many models available for this purpose, including the Simple Numeric Imputation Model, in which the null value is changed to the Mean or Median of other values in the same column of the dataset. Another strategy, called "Hot Deck Imputation," substitutes a dataset record with a similar value for the null value.

3) Exploratory Data Analysis: At this step, data is analyzed to obtain at a certain conclusion using a variety of techniques, including text analysis, diagnostic analysis, exploratory data analysis, etc. Exploratory Data Analysis (EDA) is a method for thoroughly analyzing data and essentially visualizing its key characteristics. By using various types of graphs and plots that can be created using EDA, we are able to understand the structure and content of the dataset. We can view the data in a visual format, explain the analysis using that information, and conduct a comparison study between several plots. Plots come in a variety of forms and are utilized in EDA. Some of them are mentioned below:

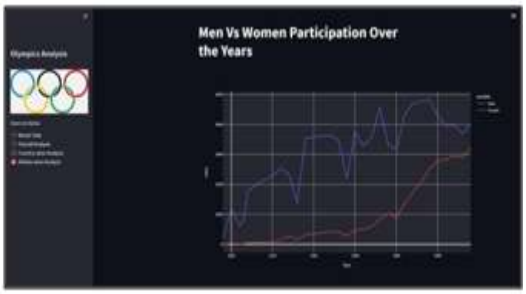
a) Histogram
b) Line Graph
c) Box Plot
d) Scatter Plot and many more.

V. ANALYSIS AND VISUALIZATION

Analysis has been done on the Summer Olympics dataset, which includes data collections from 1896 to 2016[6]. There are approximately 30,000 rows and 9 columns in this dataset. Year, Sport, Discipline, Medal, Gender, Nation, City, Event, and Athlete are some of the fields. [6]

A. Identifying Contribution of Men and Women Participants in Olympics (1896-2016):

Analysis of the total number of male and female Olympians from 1896 to 2016 allows for the calculation of the male to female participation ratio. The data demonstrates that men contribute more globally than women do. Figure 1 depicts the Olympic players' contributions by gender.



B. Identifying total number of gold, silver and bronze medals won by participants in a Country in Olympics (1896-2012):

The number of medals a country wins during the Olympics can be used to judge its level of excellence. This report details a certain nation's Olympic performance from 1992 to 2016. The outcome of a specific country can be represented through data visualization. The outcomes are(i) India's performance improved throughout time, going from 1992 with no medals to1996 with one medal to 2016 with six medals. (ii) The USA's performance was characterized by a zig-zag pattern, peaking in 2008 with 350 medals after contributing with220 medals and in 1996 with 260. (iii) From 1996 to 2008 France's performance steadily improved with medal totals of around 40, and it did well in 2016 with 80 medals. (iv) Japan's performance was not great at first, but between 2000 and 2004 there was a dramatic improvement

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1425-1429 www.ijaem.net ISSN: 2395-5252

and it acquired 100 medals, which was more than the rest.

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

The research analyzes the performances of the nations based on the medals that athletes from chosen nations took home from the Olympics between 1896 and 2016. The USA, Hungary, France, Japan, and Australia are among the nations chosen for analysis. The following conclusions have been drawn as a result of this analysis. (i)Among the five countries chosen for the1996 Olympics, the United States led with a contribution of7.53%, Australia came in second with 2.25%, Japan came in third with 1.61%, Hungary came in fourth with 0.75%, and France came in last with 0.69%. (ii)In the 2004 Olympics, the USA took first place with a contribution of 8.05%, followed by Australia with 3.3%, Japan with 2.1%, France with 1.6%, and Hungary with 0.9%. (iii)In the 2016 Olympics, the USA took the lead with a contribution of 8.5%, followed by Australia insecondplacewith3.01%, Japan in third place with1.8%, France in fourth place with 1.6%, and Hungary in last place with 0.9%

D. Analyzing the height vs weight

According to this data, most female medal winners are between 160 and 180cm height, with weight classes ranging from 50 to 150kg .No matter how much they weighed, the number of goldmeda lists did well, but the density was highest at heights between 175and180cm.

VI. RESULT AND DISCUSSION

Since 1896, the modern Olympics have grown in significance as they strive to be quicker, higher, and stronger while always maintaining a moral standard. The major goal of this research project is to graphically display and examine the numerous elements that have contributed to the evolution of the Olympics over time. It also compares these various factors. The prediction of medals in the Olympic Games is a common practice that involves using statistical models and algorithms to forecast which countries are likely to win the most medals in the upcoming games. We have used linear and logistic regression techniques used to gain insights into the relationship between various factors and the outcomes of Olympic events. Linear regression can be used to analyze the relationship between an athlete's age and their performance in an event. By analyzing the age of Olympic medalists, we can determine if there is a correlation between age and performance. Logistic regression can be used to predict the probability of an athlete winning a medal based on their previous performances, world ranking, and other factors. This can help countries and teams make strategic decisions about which athletes to select for the Olympic Games. The results of predictions of medals in the Olympic Games can vary in their accuracy.

VII. CONCLUSION AND FUTURE SCOPE

In conclusion, we have used a technique named Exploratory Data Analysis that provides a comprehensive statistical and visual representation of the performance of nations and players from the 1896 Olympic Games to the 2016 Rio Olympics. The use of visualization techniques helped us to understand the data more clearly and draw meaningful conclusions from it. This analysis of Olympic data is performed by countries and players to evaluate their performance, identifying areas for improvements by changing and making informed decisions about future strategies, increasing their chances of success in future Olympics Games. Several factors were analyzed in this project are the launch of the Summer Olympics games, the increasing number of participating countries in

both Summer and Winter Olympics, the average age of players in the Olympic Games, the participation of females in both Summer and Winter Olympics over time, the total number of medals won by various participating countries, as well as the average height and weight of players who contributed to the victory of their respective teams.

For future scope, we are integrating the impact of new technologies and training methods on athletic performance. To represent the Olympic data in a geographical format, we can use various mapping techniques to plot the performance of countries in different Olympic events on a world map. Correlation analysis can help to identify the relationship between two continuous variables. There are many potential areas for future research and analysis in Olympic data is vast. Some of these areas are predictive modeling, social media analysis, real-time data analysis, injury analysis and can provide valuable insights for athletes, coaches and countries participating in the games.

REFERENCES

[1]. Huang-Chih Shih," Survey on content-aware Video Analysis for Sports", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 99, No. 9, January 2017

[2]. "The Modern Olympic Games" (PDF). The Olympic Museum. Archived from the original (PDF) on 6 September 2008. Retrieved 29 August 2008.

[3]. Antarlina Sen and Gaurang Margaj, "A prediction model for which country will win the highest number of Gold", 2016.

[4]. Leonardo De Marchi, "Data mining of Sports performance data", 2011.

[5]. Chandra Segar Thirumalai and Monica Sankar, "Heuristic Prediction of Olympics using Machine Learning", International Conference on Electronics, Communication and Aerospace Technology, April 2017.

[6]. Summer Olympic Dataset Available: https://www.kaggle.com/the guardian/olympic-games/data

[7]. Wikipedia:https://en.m.wikipedia.org/wiki /Olympic Games, last accessed 2020/11/02.

[8]. Dey S K, Rahman M M, Siddiqi U R and Howlader A 2020 Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach J. Med. Virol. 92 632–8

[9]. Bondu R, Cloutier V, Rosa E and Roy M 2020 An exploratory data analysis approach for assessing the sources and distribution of naturally occurring contaminants (F, Ba, Mn, As) in groundwater from southern Quebec (Canada) Appl. Geochem. 114 104500

[10]. Cutait M: Management performance of the Rio 2016 Summer Olympic Gaes. Research paper submitted and approved to obtain the Master degree in Sports Administration at AISTS in Lausanne, Switzerland.

[11]. Moreno A, Moraga's M and Paniagua R 1999 The evolution of volunteers at the Olympic games' proceedings of symposium on volunteers (Lausanne, Switzerland: Global Society and the Olympic movement) pp 1-18.