Web Application of Olympic data analysis

Farkande Vaishnavi
Student BSIOTR
Pune, Maharashtra, 412207

Gurav Vaishnavi
Student BSIOTR
Pune, Maharashtra, 412207

Borse Tejas
Student BSIOTR
Pune, Maharashtra, 412207

Tushar Phadtare
Project Guide,BSIOTR
Pune, Maharashtra, 412207

Abstract

Olympic Games are one of the main international events and also a matter of prestige for countries and therefore each country tries to give their best performance during the event. Despite a lot of hard work, many countries or players are unable to perform well during the events and grab medals whereas there are many countries that perform very well in the event and secure many medals. An analysis needs to be done by each country to evaluate the previous statistics which will detect the mistakes which they have done previously and will also help them in future development.Visualisation of the data over various factors will provide us with the statistical view of the various factors which lead to the evolution of the Olympic Games and Improvement in the performance of various Countries/Players over time. The primary objective of this Research paper is to analyse the large Olympic dataset using Exploratory Data Analysis to evaluate the evolution of the Olympic Games over the years. An analysis can also be done by the host country to find out the mistakes in the arrangements of the event which will help them in overcoming these mistakes and host the event accurately. This analysis will provide detailed and accurate information regarding various factors which lead to the evolution of the Olympic Games and the improvement of Countries/Players over time in a visual format.

General Terms:

Exploratory Data Analysis for Olympic data of years from 1896 to 2016.

Keywords:
Metal Tally, Overall Analysis, Athlete Wise Analysis, Country Wise Analysis.

---

1. Introduction:

The modern Olympic Games or Olympics (French: Jeux Olympiques) are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are normally held every four years, alternating between the Summer and Winter Olympics every two years in the four years.

Various scenarios come to our mind when we look into the Evolution of the Olympic Games over the years. These scenarios are: Increase in the number of participating nations, Increase in the number of participating Athletes, Increase/Decrease in the number of events, Increase in the expenditure cost of the event, improvement in the performance of the particular country, improvement in the performance of a particular player, Increase in women participation, Participation Ratio of Men to Women, improvement in medication facilities during competition, the effect of pandemic (if any) on the performance of the players. Analysis of these scenarios would depict the evolution of the Olympics over the years. This analysis would help in future prediction

2. Motivation:

The main objective of this study is to analyse the various factors mentioned above which plays a vital role in the evolution of the Olympic Games over the years. The Analysis will include the visualisation and explanation of the change in trends of the various factors over the years which will help to predict the information of future Olympic Games. As the Olympic Games are one of the most important sporting events across the world, each country and each player tries to give their best performance in the event. To improve their performance, every country should perform such an Analysis which would help them in the improvement of their policies and strategies by providing current statistics to them.

3. Problem Definition:

Building Olympic Data Analysis Web Application for analyzing data over the years which helps athletes in widening their scope of winning a medal and might be useful for further predictions.

4. Literature Survey:

Performance measures for a country in the Olympics can be predicted using their past performance. By predicting their win using the maximum value scored by them in previous participation, the chance of winning gold in 2016 has been identified.If a person wins a medal in an Olympics during a year, the chance of winning a medal in the upcoming Olympics was predicted.Having sports performance data, predicting one‟s future performance has been done .Their performance can also be increased if they are not performing well in certain areas, and then placing them accordingly in the training program will provide considerable measure in their outcomes.Machine learning techniques were used for heuristics prediction of Olympic medals of a country.Estimation of Olympics success of a country can be done by efficiency analyse and importance of sports in society. When analysing the sports categories are mainly being more representative towards viewpoint-based content rather than being a viewpoint that is spatiotemporal. The video content of the analysis has the significance of providing more interior information than structured collected data .In addition to these techniques, the exploratory data analysis uses visual methods to provide a deep understanding and statistical summary of the data.

Data interpretation and analysis is one of the primary tasks in the field of big data analytics. There has been a lot of analysis on the Olympic Games like statistics visualisation, performance analysis of players, improvement in the performance of various countries, and many more. The type of analysis which is quite popular and suitable while analysing the evolution of the Olympics is Exploratory Data Analysis. In Exploratory Data Analysis, we examine large data and elucidate its various characteristics basically in the visual format(Graphs, Charts, and many more). EDA is an approach that provides a deeper understanding of the dataset. There has been a research paper that analyses the outbreak of the Novel CoronaVirus. The exploratory Data Analysis technique is used to analyze the data and find out the number of cases reported(positive, dead, discharged) inside China and Outside China. This paper took data from different datasets and apply the EDA technique to analyse various factors like the number of cases recovered during January and February inside and outside China, the number of cases confirmed in the different provinces of China, and outside China till 16 February 2020.The main aim of this analysis was to find out the growth in the performance of a country in the Olympics over the years. With the Help of such an Analysis, any player can check their progress record and can also have a look at their opponent’s progress.

5. Software Requirements Specifications:

5.1 Functional Requirements: Python

Libraries:- Numpy, Pandas, Plotly, Matplotlib, Seaborn.

5.2 External Interface Required:

API:- Streamlit.

5.3 System Requirement:

Operating Environment: - Linux, Windows.

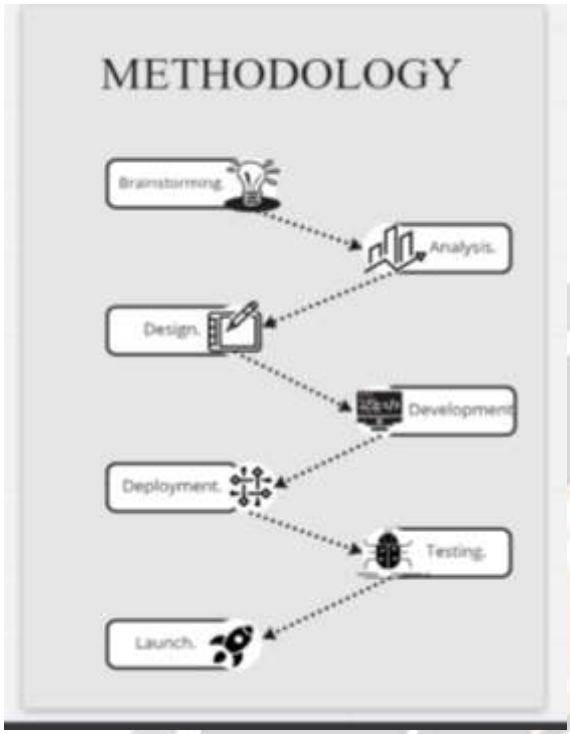
Deployment Environment: - Heroku.

Development Environment: - Pycharm, Jupyter Notebook, Google Collab.

5.4. Analysis Models: SDLC Model to be applied:

A project methodology can outline whatever you want to outline. So, you can have one for the initiation process, which would start with the brainstorming idea selection. The planning process is complex as

it initiates with planning, analysis of project charter with proof of concepts, collecting requirements, and the development of scope. User design is important to step in the methodology of a project. After designing, the user must start the deployment of the project with continuous integration and continuous deployment method. In the end, deployment testing and launching of the application should be done respectively.

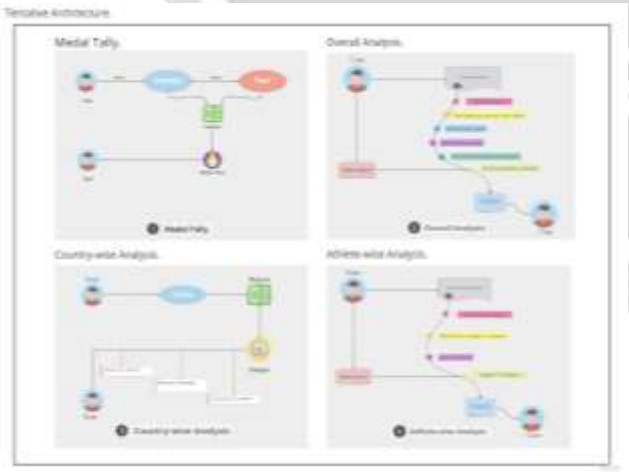


(Fig 1: SDLC Model)

6. System Design:

6.1. System Architecture:

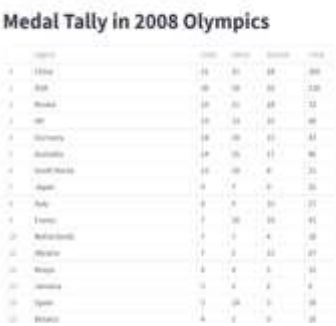
Tentative architecture decides the entire flow of execution of the project. from above we can see here are mainly four modules which are the main part of the charter as follows:



(Fig 2: System Architecture)

6.1.1 Medal Tally:

Ranking of countries is given Concerning the total number of medals won and are categorized by the number of Gold, Silver And Bronze medals respectively.



6.1.2 Overall Analysis:

6.1.2.1 Top Statistics:

Displays number of Editions, Hosts, Sports, Events, Athletes, and Nations over the years.

6.1.2.2 Participating Nations Over The Years:

Line Graph for the number of countries participating over the editions.

6.1.2.3 Events over years:

Line Graph for number of Events organised over the editions.

6.1.2.4 Athletes over the years:

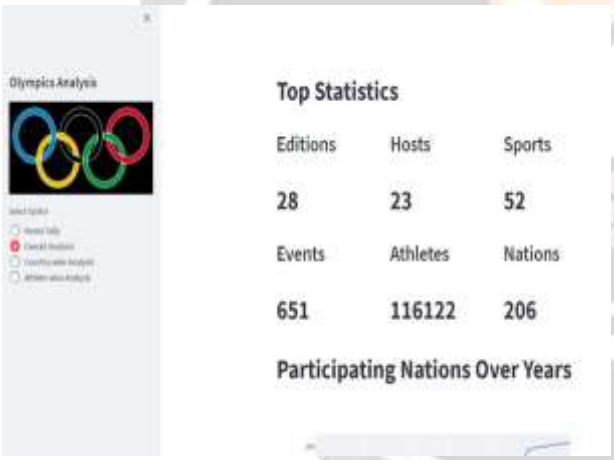
Line Graph for number of Athletes participating over the editions.

6.1.2.5 Number of Events in each Sport:

HeatMap, which describes the number of events in all sports over the editions.

6.1.2.6 Most Successful Athletes:

By selecting sports, it displays details of the top 15 record-breaker athletes.



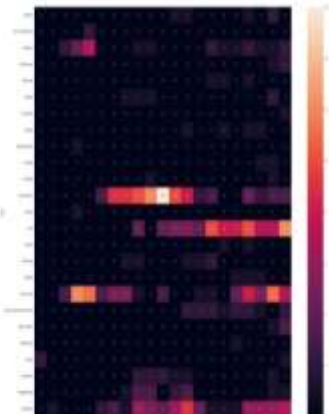
6.1.3 CountryWise Analysis:

In this module, by selecting country it displays the following 3 parameters:

6.1.3.1 Country medal tally

6.1.3.2 Country's all-inclusive achievements.

6.1.3.3 Top 10 athletes of the country.



6.1.4 Athlete Wise Analysis:

6.1.4.1 Distribution of age:

Analyzes data of athletes for age distribution w.r.t medals using a distplot which shows the curve of the probability distribution function.

6.1.4.2 Distribution of age w.r.t sport:

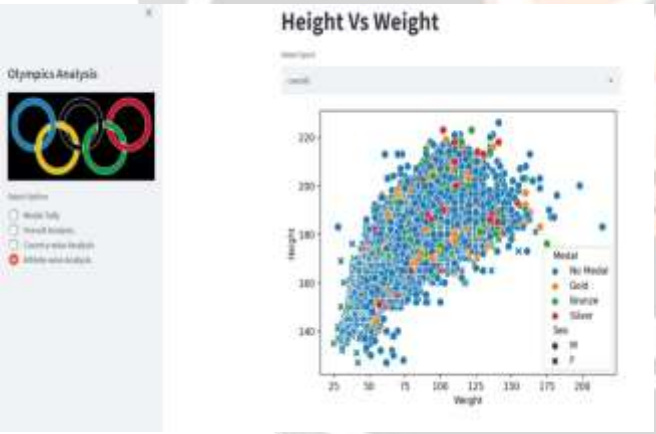
Analyzes data of athletes for age distribution w.r.t sports using a distplot which shows the curve of the probability distribution function.

6.1.4.3 Height VS Weight:

By selecting the sport, it will show the scatter plot for height vs weight w.r.t medal and gender.

6.1.4.4 Men VS Women:

Line graph for participation of men and women over the editions.



7 Other Specifications:

7.1.Applications:

7.1.1 By using Data analysis, the number of medals received by countries can be classified and can be displayed.

7.1.2 Can determine which country hosted the Olympics in which year.

7.1.3 Athletes can study which age group is best to excel in their game and is more likely to win the medal.

7.1.4 Can determine which country is best in which sport.

7.1.5 Can determine the number of games a country is participating in and participated in past years.

7.1.6 Can determine which sport event started in which year.

7.1.7 Can determine the number of events increasing or decreasing over the years.

7.2 Advantages:

1. Deeper insight into the performance of countries in the Olympics over the years and helps athletes to quickly analyze their own and competitors' performance.

7.3 Disadvantages:

Due to former geographical or historical changes analysis may vary.

8. Conclusion and Future work:

8.1 Conclusion:

The main objective of this study was to analyse and visualise the various factors which have contributed to the Evolution of the Olympic Games over the years. This type of analysis is very helpful as this type of analysis can be performed by any Country or Player which can help them in analysing their performance so that they can improve their performance by changing their strategies.

We have used a technique named Exploratory Data Analysis which enables you to encapsulate the primary factors of a dataset into a visual format. We selected Python language to implement our work because It is one of the best languages suitable for Data Analysis and is the platform where we have performed this Analysis. As a result of the Analysis, we can conclude that It is true that Olympic Games have evolved considerably over time from the 1896 Olympic Games till the 2016 Rio Olympics. Various factors provide valid evidence that the Olympics
have changed a lot. some of these factors are the launch of the Winter Olympic Games apart from the Summer Olympic Games in 1924, an increase in the number of participating countries in both Summer and Winter Olympics, the Average age of players in the Olympic Games, the increase in the participation of the females in both Summer and Winter Olympics over the time, Total number of medals won by various participating countries over the years, Average height and the weight of Players who contributes to victory of Games in the event. Apart from these, there are many more factors that depict the Evolution of the Olympic Games over time. Visualisation of these factors has been done to explain and validate the Analysis in various Graphical formats like a Line graph, Scatter Plots, Bar, Graphs, Dist Plots, etc

8.2Future Scope:

We all know that any Analysis is not perfect and it consists of some limitations which define the Future scope of the Research Work. This project work also contains some limitations which we are considering as the Future Scope of the Project. These are ASCI-2020 IOP Conf. Series: Materials Science and Engineering

We have visualised our data only in Graphical format. We can also describe the data in other formats like Geographical format where we can depict the countries on the World map.

Till now we have only performed Data Analysis using Exploratory Data Analysis. We can also apply various Machine Learning Algorithms to the data set after Analysis and can create a Predictive Model which can predict the statistics of the Future Olympic Games.

We can also perform Correlation Analysis on the data set and analyse the relationship between two continuous variables.

9.References:

01. Wikipedia contributors: https://en.m.wikipedia.org/wiki/Olympic Games, last accessed 2020/11/02.

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

03. Antarlina Sen and Gaurang Margaj, “A prediction model for which country will win the highest number of Gold“, 2016.

04. Leonardo De Marchi, “Data mining of Sports performance data”, 2011.

05. 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.

06. Chandra Segar Thirumalai and Monica Sankar, “Heuristic Prediction of Olympics using Machine Learning”, International Conference on Electronics, Communication and Aerospace Technology, April 2017.

07. Alexander Rathke and Ulrich Woitek, “Economics and Olympics: An Efficiency Analysis”, January 2007.

08. Summer Olympics Dataset, Available: https://www.kaggle.com/theguardian/olympic-games/data