OLYMPIC DATA ANALYSIS

Submitted in the partial fulfillment of the requirements

for the degree of B.Tech in Computer Engineering

by

Sanika Sawale 1 (21CE1109)

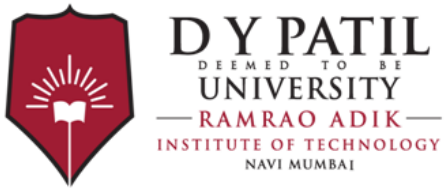
Neha Memane 2(21CE1166)

Mohit Bagri 3 (21CE1141)

Siddhesh Salunkhe 4 (21CE1143)

Supervisor

Mrs. Sanjivani C. Chakote



Department of Computer Engineering

Ramrao Adik Institute of Technology

Sector 7, Nerul, Navi Mumbai

(Under the ambit of D. Y. Patil Deemed to be University)

October 2023

Ramrao Adik Institute of Technology

(Under the ambit of D. Y. Patil Deemed to be University)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706

CERTIFICATE

This is to certify that, the Mini Project-III report entitled

OLYMPIC DATA ANALYSIS

is a bonafide work done by

Sanika Sawale 1 (21CE1109)

Neha Memane 2(21CE1166)

Mohit Bagri 3 (21CE1141)

Siddhesh Salunkhe 4 (21CE1143)

and is submitted in the partial fulfillment of the requirement for the degree of

B.Tech in Computer Engineering

to the

D. Y. Patil Deemed to be University

Supervisor	Project Co-ordinator
(Mrs. Sanjivani C. Chakote)	**(Mrs. sumithra T.V)**
Head of Department	Principal
(Dr. Amarsinh V. Vidhate)	**(Dr. Mukesh D. Patil)**

Mini Project Report - III Approval

This is to certify that the Mini Project - III entitled ***" Olympic Data Analysis "*** is a bonafide work done by ***Sanika Sawale 1 (21CE1109)***, ***Neha Memane 2 (21CE1166)***, ***Mohit Bagri 3 (21CE1141)***, and ***Siddhesh Salunkhe 4 (21CE1143)*** under the supervision of ***Mrs. Sanjivani C. Chakote***. This Mini Project is approved in the partial fulfillment of the requirement for the degree of ***B.tech in Computer Engineering***

Internal Examiner :

1. .....

2. .....

External Examiners :

1. .....

2. .....

Date : .../.../.....

Place : .....

DECLARATION

I declare that this written submission represents my ideas and does not invovle plagiarism. I have adequately cited and referenced the original sources wherever others' ideas or words have been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action against me by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: _____

Sanika Sawale 1 (21CE1109)
Neha Memane 2(21CE1166)
Mohit Bagri 3 (21CE1141)
Siddhesh Salunkhe 4 (21CE1143)

Abstract

There are many countries that have achieved good results and won medals in this event. Many countries' medals need to be analyzed to evaluate previous statistics It identifies previous mistakes and also helps in future development. Host countries can also conduct analysis to find mistakes. Event Arrangements will help you overcome these mistakes and host your event. Detailed Events The main purpose of this article is to analyze the Olympic Games on a large scale. A dataset that uses exploratory data analysis to assess the evolution of the Olympic Games Year. This analysis provides detailed and detailed information on various factors. This will lead to the evolution of the Olympics and the development of countries and athletes. Displays the time visually. By visualizing data in different elements, A statistical look at the various factors leading to the development of the Olympic Games and Over time, performance improved for different countries or players.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Big data Analytics is becoming more and more popular in our daily life. Every day, tons data is generated and analyzed using big data tools to extract useful and meaningful information. This study aims to analyze and visualize the modern Olympic Games from 1896 to the present. This project shows an Analysis of the medals and athletes who participated in the modern Olympic Games for the 120th years and gives a preview of the same. Apache Spark is probably the most popular Data Analytics framework with multiple supporting libraries like Spark SQL, Spark Streaming, Graph X, ML lib etc. this experiment uses the Python API for Spark i.e. p y Spark. This application uses specific examples of analysis and claims that they provide effective, efficient, functional and convenient model for users. Discussion, conclusion, future work are summarized together with the application recommendations for improvement of restrictions, which were identified in the assessment as next section. The primary goal of a dashboard is to explain what the user can benefit from from an evolved system experiencing visual representation. The assessment should have min consider whether the product meets specific requirements, effective.

1.1 Overview

1.1.1 AI & ML

Artificial intelligence (AI) and machine learning (ML) play a significant role in the field. Olympian data analysis that offers advanced techniques and tools for extracting insights and patterns from complex datasets. Here's an overview of how AI and ML are being applied to Olympic data analysis:

1. **Data processing and cleaning:** Artificial intelligence algorithms such as natural language processing (NLP) and computer vision can help process and clean unstructured data various resources, including text and images. This is especially useful for extraction information from news articles, social media and historical documents related to Olympics.

2. **Data integration:** ML models can be used to integrate disparate and fragmented datasets, facilitating the analysis and visualization of complex Olympic data.

3. **Time Series Analysis:** Predictive time series models, a subset of ML, can predict trends and patterns in Olympic data over time. Researchers can use these models to analyze his toric trends, predict future participation and identify seasonality in athlete performance.

4. **Medal count prediction:** ML algorithms can be used to build prediction models the number of medals for various countries in future Olympic Games. These models take into account factors such as historical performance, investment in the sport and the economy and demographic conditions of the participating countries.

5. **Analysis of athlete performance:** Machine learning models can analyze influencing factors athlete performance such as age, gender, training regimes and other individual honors. This knowledge can be used to optimize training programs and improve the athlete performance.

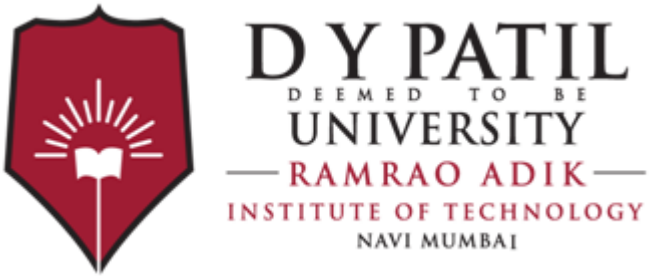


Figure 1.1: DY Patil Deemed University logo

Table 1.1: Name of Students

Name	**Roll No.**
Ms. Sanika Sawale	(21CE1109)
Ms. Neha Memane	(21CE1166)
Mr. Mohit Bagri	(21CE1141)
Mr. Siddhesh Salunkhe	(21CE1143)

1.2 Motivation

Motivation is a key factor in conducting Olympic data analysis research because it can be challenging and rewarding field of study. Here are some motivational factors that can inspire researchers to explore this topic:

1. **Impact on society:** Research in Olympic data analysis can help shed light on the wider environment the impact of the Games on society, from inspiring athletes to fostering international relations and promote a healthy lifestyle..

2. **Historical context:**The Olympics have a rich history spanning over a century. Research-Olympic data allows you to immerse yourself in historical, cultural and social contexts different time periods and regions..

3. **Advancing Sports Science:** Understanding the data behind athlete performance, sports trends, and sports development can contribute to advancements in sports science and training methods.

4. **Performance Prediction:** Developing predictive models for Olympic outcomes can be intellectually stimulating and has practical applications in sports betting, sports governance, and even athlete training.

5. **Data Innovation:** Olympic data provides a vast and diverse dataset for analysis. Researchers can explore innovative data analysis techniques and visualization methods, which can be applied to other domains as well.

6. **Public Engagement:** Research in Olympic data analysis often attracts public interest, allowing researchers to communicate their findings to a broader audience and inspire interest in data science and sports analytics.

1.3 Problem Statement and Objectives

The main objective of this study is to analyze the various factors mentioned above that play a role essential role in the development of the Olympic Games over the years. The analysis will include visualizing and explaining the change in trends of various factors over the years which will help predict information about future Olympic Games.

The Olympic Games are one of the most important sporting events in the world country and every player tries to give the best possible performance in action. For improvement each country should carry out such an analysis that would help it in improving their policies and strategies by providing up-to-date statistics

1.4 Organization of the report

In our report, we examine, research and analyze in Literature Survey from chapter 2, trying to understand the features, algorithms and working of each model used. We covered brief description of data analysis which we can understand on the basis of visual representation and proceeded to see the limitations of that system. In Chapter 3, the main proposed system, design and the techniques that we planned to implement are described with relevant diagrams like system design diagram, use case diagram and class diagram for a clear understanding of the working. We go into greater detail with each step involved in the process. We explain the architecture with the data flow of the system. Finally, concluding chapter 3, we delivered the concept of the hardware and software requirements used in project in detail. In chapter 4 we look at the results of the implemented project and understand, how proposed system is compared to existing system with graphically represented outputs. We move on to the final part of the project in chapter 5 while concluding it, the overall impact of the project and what modifications the future of this project might hold.

Chapter 2

Literature Survey

1. Using machine learning technology to predict Olympic medal wins from a specific country.

2. Estimating the success of a country can be done by analyzing the efficiency and importance of sports in society..

3. When analyzing sports categories, the following points are mainly more representative. Perspective-based content rather than spatio-temporal perspective.

4. Interpretation and analysis of data is one of the main tasks in the field of big data Analysis. Many statistical and other analyzes have been done about the Olympics. Visualization, analysis of player performance, performance improvement Different countries and more countries.

5. One type of analysis that is very popular and suitable for analyzing the evolution of the Olympic Games is exploratory data analysis. Exploratory data analysis This technology is used to analyze the data and discover the number of reported cases.

2.1 Survey of Existing System

A review of existing systems in the field of Olympic data analysis includes a review and summary of approaches and tools used by researchers, organizations and bodies. A platform that has worked on similar projects. Below is an overview of some quick research. Examples of main approaches and systems applied to Olympic data analysis:

1. **Official Olympic Websites:** The official Olympic websites, such as the International Olympic Committee's (IOC) website, provide a wealth of historical data, results, and statistics. Researchers and enthusiasts often refer to these sources for official and up-to-date information.

2. **Data Analysis Tools:** Various data analysis and visualization tools are commonly used for Olympic data analysis. This includes statistical software like R and Python (with libraries such as Pandas, NumPy, and Matplotlib), as well as specialized analytics and visualization tools.

3. **Custom databases:** Some researchers create custom databases of Olympic data.A place where data is collected, cleaned and stored in a structured format. These databases are designed for specific research needs and often contain detailed information on athletes.

4. **Visualization platform:** panel or Power BI is used to create dynamic dashboards and exploratory visualizations. Olympic data trends and insights

5. **Academic Research:** Numerous academic articles and studies focus on performance analysis, historical trends, and various aspects of the Olympics. Socio-Cultural Influences These research papers often use statistics and specialized data analysis method.

2.2 Limitations of Existing System or Research Gap

Our dashboards provide users with attractive graphical representations of information and help them understand content more easily. This Olympic data analytics website represents a unique platform never seen before in the world of sports analytics. We look forward to providing superior data coverage, predictive analytics capabilities, and an immersive user experience.

Chapter 3

Proposed System

3.1 Problem Statement

1. The website should be designed to be scalable and easy to maintain, with a robust back-end that can handle multiple analysis simultaneously.

2. Overall, the website should be a valuable tool for users who would like to analyse the data for their understanding or knowledge Broadcasters can use data to enhance coverage, provide in-depth statistics, and engage viewers with insightful commentary.

3. Evaluating the return on investment for sponsors and advertisers and helping them make informed marketing decisions.

4. Analysis of the impact of rule changes and innovations in sports equipment and techniques on sports performance. Providing comprehensive profiles of athletes, including their achievements, backgrounds and career trajectories..

3.2 Proposed Methodology/Techniques

This approach is called a systematic path to a solution. There are different.Effective factors in the evolution of the Olympic Games to determine these factors.To conduct a comparative study of these factors, the following approach should be followed. It takes us to our destination. The flowchart of the proposed method is shown below,Discuss each step in detail.

1. **Data Collection:**

   The first step in any type of analysis, whether technical or non-technical, is data collection. We have used data from various data sources to analyze the long-term evolution of the Olympics. The dataset includes information about players and details such as gender, height, weight, country they play for and medals won. This data can be used to analyze the performance of a particular player and is also useful for comparative studies between two or more players.

2. **Data Pre-Processing:**

   After data collection, the next step is data processing. Data obtained directly from a data source such as a dataset is called raw data. Various machine learning techniques and algorithms such as linear regression, decision trees, and SVM cannot be directly applied to raw data. This data must be processed and converted into useful data. Data preprocessing is the process of converting raw data into useful data by carefully checking for errors and removing redundant, incomplete or incorrect data.

3. **Exploratory Data Analysis:**

   The next step after data pre-processing is data analysis. In this step, analysis is done on datausing various techniques. Exploratory Data Analysis (EDA) is an approach to analyze data thoroughly and encapsulate its primary attributes basically in visual format. With the help of EDA, we can understand the structure and content of the dataset by various types of graphs and plots which can be drawn with the help of EDA.There are various types of plots which used in EDA. Some of them are mentioned below: •Histogram •Bar Graph •Box Plot •Scatter Plot

3.3 System Design

Designing a system for Olympic data analysis involves planning the architecture, data pipeline, and tools necessary to collect, process, analyse, and visualize the data effectively. Here's a high-level overview of a system design for Olympic data analysis:

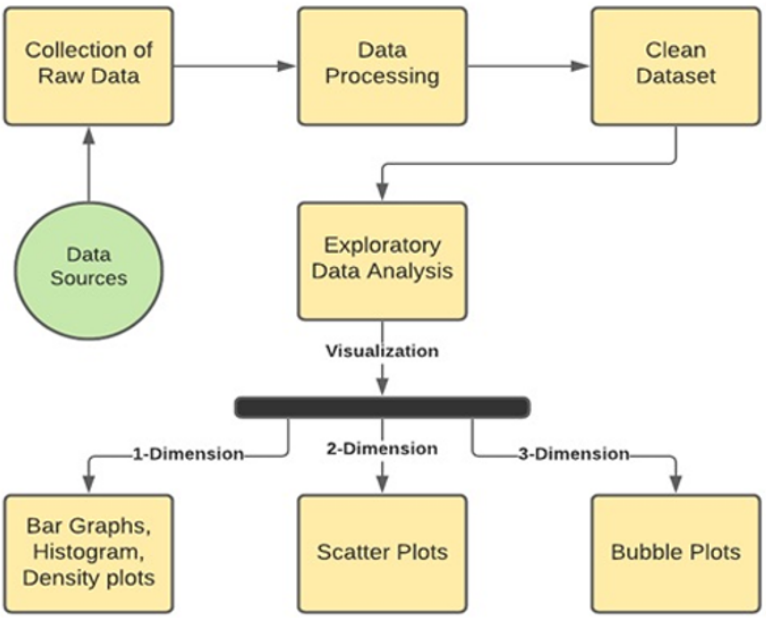


Figure 3.1: flow chart

1. **Data Collection and Storage:**We collect and store Olympic data from reliable sources.
2. **Data Cleaning and Preprocessing:** clean and prepare your data for analysis.
3. Data Analysis Engine: Use appropriate tools and libraries for data analysis.
4. **Time Series Analysis:** Analyze historical trends in Olympic data.
5. **Sports Analysis:** Analyze the popularity and trends of various sports.
6. **Geographic Analysis:** examining the distribution of cities and countries hosting the Olympics.
7. **Performance Analysis:** Evaluate the factors affecting athletic success.

3.4 Details of Hardware/Software Requirement

1. **Python Libraries:** Numpy, Pandas, Plotly, Matplotlib, Seaborn were used to obtain useful visualizations for exploratory data analysis in just a few lines of code.

2. **External Interface Required:** Streamlit allows us to display descriptive text and model outputs, visualize data and model performance and modify model inputs through the UI using sidebars.

3. **NumPy:** NumPy is a basic library for numerical calculations in Python. It supports large multidimensional arrays and matrices and a set of mathematical functions to manipulate these arrays.

4. **Pandas:** Pandas is a powerful library for data manipulation and analysis. It provides data structures like DataFrames for handling structured data and tools for data cleaning, transformation, and exploration.

5. **Operating Environment:-** Windows.

6. **Dataset :** Platforms like Kaggle provide a treasure trove of diverse datasets against which you can explore real-world scenarios, apply analytical techniques, and extract valuable insights.

7. **Streamlit:**is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.

Chapter 4

Results and Discussion

1. The main objective of this study was to analyze and visualize the various factors that have contributed to the evolution of the Olympic Games over the years.

2. This type of analysis can be done by any country or player and is very useful because it helps you analyze your performance to change your strategy and improve your performance.

3. We used a technique called exploratory data analysis, which allows us to capture the key elements of a data set in a visual format.

4.1 Implementation Details

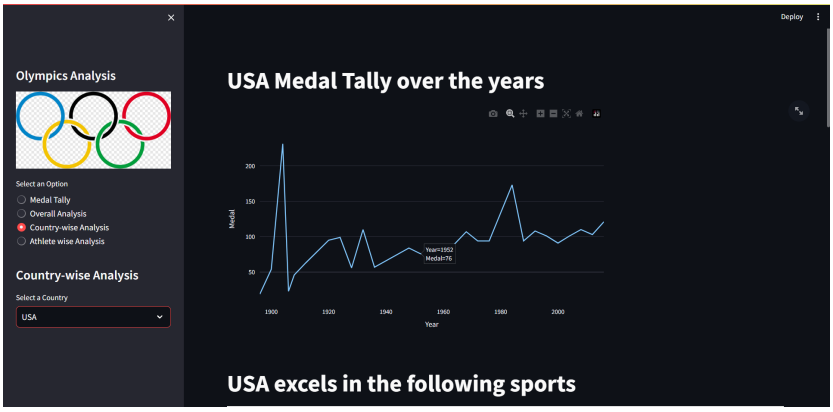


Figure 4.1: Frontend

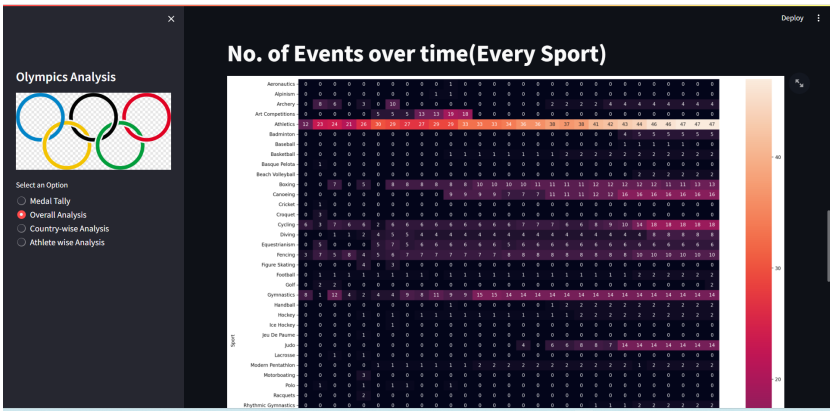


Figure 4.2: Frontend

4.2 Result Analysis

Implementation of the results of the Olympic data analysis project is not only beneficial to all stakeholders involved in the Olympic Games, but also essential. By leveraging the insights gained from this analysis, we can celebrate the rich history and tradition of the Olympics while embracing new sports and athletes, ensuring that the Olympic movement continues to evolve and progress. continues This in turn contributes to the continuing legacy of the Olympic Games as a symbol of global unity, excellence and cooperation.

Chapter 5

Conclusion and Further Work

Conclusion

In this report, Data analysis and visualization of Modern Olympics using Big Data tools is presented. 120 years of Olympics gives you brief information about the Sports, history, Athletes, events, Event Locations, etc.
The Web application gives you insight of these topics in Statistical and Interactive way all under few clicks.

Further Work

We all know that every analysis is not perfect and contains limitations that define the scope of future research work. The work on this project also includes limitations that we consider as the next work of the project. Data can also be written in other formats, such as geographic formats that can depict countries on a world map. You can also apply various machine learning algorithms to the analyzed data set to create predictive models that can predict future Olympic Games statistics.

Appendices

Appendix A

Weekly Progress Report

Figure A.1: Weekly Progress Report

Appendix B

Plagiarism Report

Appendix C

Publication Details / Copyright / Project Competitions

1. S. Gaud, S. Kale, R. Sarambale, P. Gunjgur, "A Recommendation System for Integrated Agriculture Using Convolutional Neural Networks with Random Forest Algorithm," Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), July 2022. [1] Cutait M.: Management performance of the Rio 2016 Summer Olympic Games. Research Paper submitted and approved to obtain the Master's degree in Sports Administration at AISTS in Lausanne, Switzerland.

2 Moreno A, Moragas M and Paningua R 1999 The evolution of volunteers at the Olympic Games Proceedings of Symposium on Volunteers.

3 Switzerland: Global Society and the Olympic Movement) pp 1–18.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.

4 https://www.ijraset.com/research-paper/olympic-data-analysis-using-data-science

5 https://www.kaggle.com/

6 https://olympics.com/en/

Acknowledgments

I thank the many people who have done lots of nice things for me.

Date: _____