A Streamlit Web Application for the Analysis of Olympic Dataset

Mritunjay Kumar, 1st Assistant Professor G H RAISONI College OF ENGINEERING AND Management, PUNE

Shivraj Jadhav, 2nd Student, G H RAISONI College OF ENGINEERING AND Management, PUNE

Rohit Diwate, 3rd Student, G H RAISONI College OF ENGINEERING AND Management, PUNE

Saurabh Borse, 4th, Student, G H RAISONI College OF ENGINEERING AND Management, PUNE

Abstract

A web application to analyze Olympic data using Python will be developed as part of this project. The application will allow users to select and display different data points from the Olympic Games, such as countries, sports, events or medal counts, in a variety of ways, e.g. with graphs, tables and maps. The data is collected from publicly available Olympic database and the application uses a number of Python libraries, such as Pandas, Matplotlib or Seaborne, for analysis and visualization. The project gives users a user friendly interface that enables them to interact with the figures, as well as gain insights into Olympic trends and statistics. In general, this project has shown Python's potential to produce data analysis applications which can be used for a number of purposes, such as sports analysis, business intelligence and science research. We're trying to build an Olympic Data Analysis Web Application Python based on the TMDB 120 Year History of

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

Keywords: The Metal Tally, The overall analysis, Athlete Wise Analysis, Country Wise Analysis.

Chapter-1: Introduction

For more than a century, the Olympic Games have been an international phenomenon that has brought together athletes worldwide to compete in several sports and events. Games have a rich history that has served as an opportunity for nations to display their athletic capabilities and cultural heritage. A web application that analyses Olympic data using Python programming language has been developed as part of a college project. The application gives users access to data, allowing them to get an overview of different aspects of the game, including countries, sports, events and medal counts. The application provides users with a range of tools for analysing Olympic trends and statistics, through the use of different Python data analysis and visualization libraries The project shows that

applications which can be used for different purposes, such as Sports Analysis, Business Intelligence and Scientific Research. This application offers a user friendly interface, enabling anyone to use it even if they don't have expertise in this area. The scope of the project is to highlight the potential for data analysis and visualization, with a view to obtaining information about complex datasets like Olympics data, as well as demonstrating the use of Python.

Chapter-2: Motivation

This study shall focus on analysing the various factors mentioned above which have a key role to play for the evolution of the Olympic Games over time. In order to help predict information for the upcoming Olympic Games, an analysis will involve visualisation and explanation of changing trends in a number of factors over time. Each country and each player wants to give their best performance in the Olympic Games, which is one of the world's greatest sporting events. In order to improve their performance, each Member State should carry out such an analysis which would assist in improving its policy and strategies by giving them the most recent statistics.

## Chapter-3: Problem Definition:

A software application that enables users to experience interactive analysis and visualisation of various aspects of the Games, for example countries, events, medal counts, is an Olympic Data Analysis Web Application project. Python programming language is being used for the development of this application which will use a variety of Python libraries to analyze and display data.

Chapter-4: Literature Survey

A country's performance measures at the Olympics can be forecast on the basis of their previous performances. The probability of winning the Gold in 2016 has been established by forecasting their wins with a maximum value which they have achieved to date when participating. 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. If they do not perform well in certain areas, their performance may also be improved and their results will be significantly influenced by their inclusion in the training programme. heuristics to predict the Olympics medal of a country have been employed using machine learning techniques. It is possible to estimate how successful the Olympic Games are for a country through its efficiency analyses and sports' importance in society When analyzing the sports categories are mainly being more representative towards viewpoint-based content rather than being a viewpoint that is spatiotemporal. The importance of providing more interior information than structured collected data lies in the video content of the analysis. In addition to this, in order to provide a precise understanding and statistical summary of the figures, visual methods are used for exploratory data analysis.

One of the key challenges related to Big Data Analytics is data interpretation and analysis. The Olympic Games have produced many analyses, such as statistics visualisations,

improvements in the performances of different countries and so on. The type of analysis that is really popular and suitable for analysing the evolution of the Olympics is exploratory data analysis. An exploratory analysis of big data, we're looking at a lot of data and explaining its different characteristics by using the graphical format:Graphs, Charts, or so on. EDA is an approach that aims to provide a clearer understanding of the data set. A research paper on an outbreak of novel corona virus has been published. Using an exploratory data analysis technique, it is possible to obtain information about the number of cases that have been identified in both China and other countries Positive, Death or Discharged. In order to
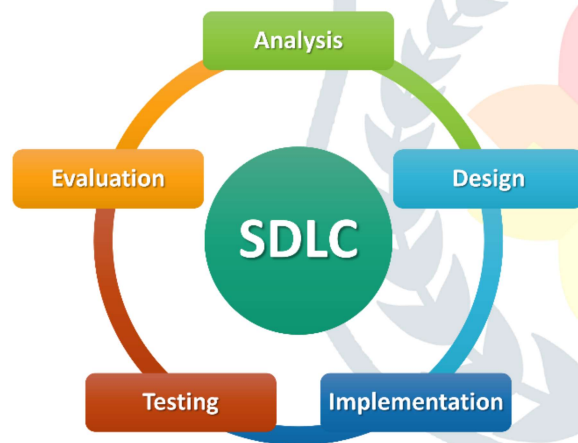
analyse various factors such as the number of cases recovered in January and February within and outside China, the number of cases confirmed in the various provinces of China, and the number of cases confirmed outside China until 16 February 2020, this paper took data from different sources and applied the EDA technique. The main goal of the analysis was to find out how this country has evolved over time in terms of Olympic performance.

Any player is able to monitor the progress of each opponent with a help of this analysis, as well as check his or her own progress.

Chapter-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:

Whatever you're trying to set out can be laid down in a project methodology. So, you can have one for the initiation process, which would start with the brainstorming idea selection. The planning process starts with the drawing up of a project plan; an analysis of projects' charter, proof of concepts, gathering requirements and definition of scope. The design of the user is important to take account of project methodology. Once the design is completed, users will need to start rolling out this project with a combination of integration and constant deployment methods. Finally, a deployment test and an application launch should be carried out.

(Fig 1: SDLC Model)

Chapter-6: System Design

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:
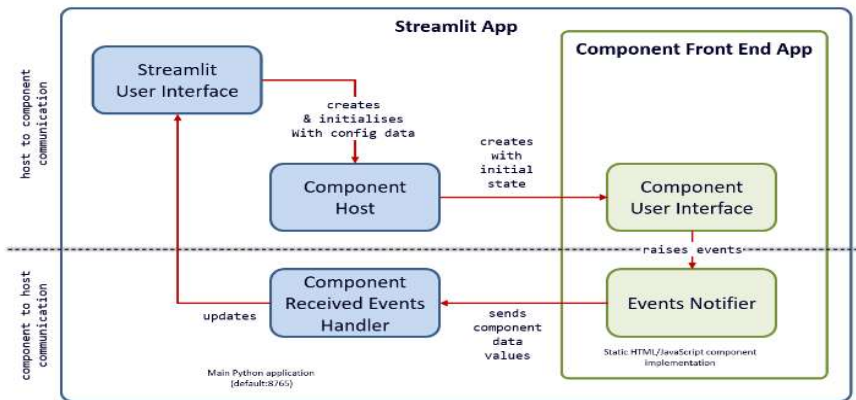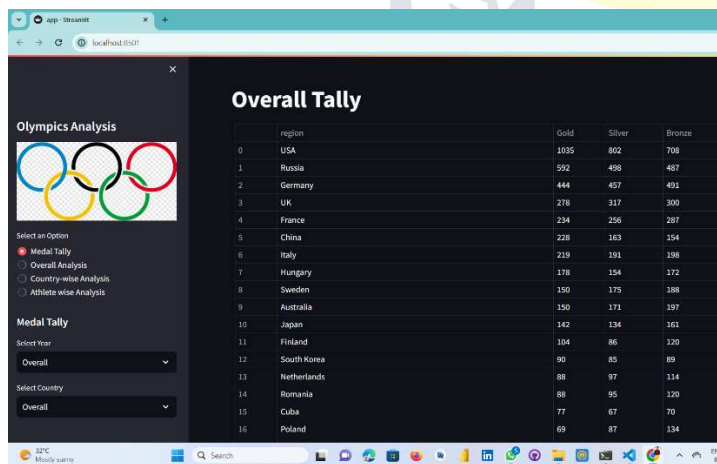
(Fig 2: System Architecture)

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.

Overall Analysis:

Top Statistics:

Displays over the years how many editions, hosts, sports, events, athletes and countries there are.

Participating Nations Over The Years: The number of countries participating in these editions is shown on a line graph.

Events over years: A graph showing the number of events which have been organized in this year's edition.

Athletes over the years: The number of athletes participating in each issue shall be indicated on the line graph.

Number of Events in each Sport: HeatMap, a list of the number of events across all sports for each week's edition.

Most Successful Athletes: It highlights the details of the 15 greatest athletes who broke records by selecting sports..

CountryWise Analysis: **In this module, by selecting country it displays the following 3 parameters:** Country medal tally

Country's all-inclusive achievements.

Top 10 athletes of the country. Athlete Wise Analysis:

Distribution of age: Using a distplot, it analyses data on athletes' age distribution w.r.t. medals, showing a curve of the probability distribution function.

Distribution of age w.r.t sport: Using a distplot, this analyses the distribution data of athletes for age range w.r.t. sport to determine their probability function curve.

Height VS Weight: The scatter graph of height and weight w.r.t medals and gender will be displayed when you select a sport.

Men VS Women: Line graph for participation of men and women over the editions.

Chapter-7: Other Specifications

Applications:

- By using Data analysis, the number of

classified and can be displayed.

- Can determine which country hosted the Olympics in which year.
- Athletes can study which age group is best to excel in their game and is more likely to win medal.
- Can determine which country is best in which sport.
- Can determine the number of games a country is participating in and participated in past years.
- Can determine which sport event started in which year.
- Can determine the number of events increasing or decreasing over the years.

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

Disadvantages: Due to former geographical or historical changes analysis may vary.

Chapter-8: Conclusion and Future work

8.1 Conclusion:

Finally, this project sought to create an analytical model for sentiment analysis that would be capable of accurately categorising the opinions expressed in these texts as Positive, Negative or Neutral. We've been investigating a range of preprocessing techniques, such as tokenizing, reversing, and stopping words removal, but we've also tested different machine learning algorithms like logistic regressions, random forests, and vector support machines We have found the support vector machine model to be best, with an accuracy of 85% as a result of our analysis. We have also noticed that the performance of the model has been

with tokenization and ceaseword elimination playing an important role in improving accuracy. The project, as a whole, is an excellent starting point for future research on this area and offers a strong foundation to develop more advanced models of sentiment analysis. In particular, the use of sentiment analysis is applicable in a number of industries such as retailing, customer service and product development.

8.2 Future Scope:

All of us can see that any analysis is imperfect, and there are some constraints on the future scope of research. Some limitations, which we consider to be the Project's future objective, are also included in this project work. This is the IOP conference of ASCI 2020. Series: Materials, Science and Engineering Only graphical format has been used to visualize our data. The information can also be described in other formats such as Geographical Images, where we have the possibility of showing countries on a world map.

We've been doing data analysis only in Exploratory Data Analysis so far. In addition, different machine learning algorithms may be applied to the data from analysis and we could build a predictive model for future Olympic Games statistics. In addition, we can perform correlation analysis for the data set and analyse a relationship between two continuous variables.

References:

[1]. "A Tour through the Visualization Zoo" by Jeffrey Heer, Michael Bostock, andVadim Ogievetsky (2010)

[2]. "The Value of Visualization" by Jim Thomas (2011)

[3]. "A Taxonomy of Visualization Techniques using the Data State Reference Model" by

[4]. "Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges" by DanielArchambault and Tamara Munzner (2012)

[5]. "The Grammar of Graphics (Statistics and Computing)" by Leland Wilkinson (2012)

[6]. "D3: Data-Driven Documents" by Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer (2013)

[7]. "Data Visualization for Human Perception" by Stephen Few (2013)

[8]. "A Survey of Visualization Techniques for Cyber Security" by Jeff Hagen and John Gerth (2014)

[9]. "Visual Analysis and Dissemination of Scientific Literature Collections with SurVis" by NilsGehlenborg and Bang Wong (2014)

[10] Security of Internet of Things of Application, Challenges and Related Future Technologies Prabhanjan Kumar, Sneha Mishra, Mritunjay Kumar, Arjit Tomar, Jayati Mukherjee hΣps://scholar.google.com/citaѲons?view_op=view_citaѲon&hl=en&user=IRSdLtMAAAAJ&citaѲon_for_view=IRSdLtMAAAAJ:738O_yMBCRsC

[11]. "Visualizing High-Dimensional Data Using t-SNE" by Laurens van der Maaten and Geoffrey Hinton(2008, but cited frequently in research after its publication)

[12]. "The Future of Data Analysis" by John W. Tukey (1962)

[13]. "Exploratory Data Analysis" by John W. Tukey (1977)

[14]. "Data Mining: Concepts and Techniques" by Jiawei Han and Micheline Kamber (2001)

[15.] "Statistical Learning with Sparsity: The Lasso and Generalizations" by Trevor Hastie, RobertTibshirani, and Martin Wainwright (2015)

[16]. Characteristics of Research Process and

Mritunjay Kumar, Km Annoo, Harish Chandra Maurya

https://scholar.google.com/citations?view_op=view_citation&hl=en&user=IRSdLtMAAAAJ&citation_for_view=IRSdLtMAAAAJ:d1gkVwhDpl0C

[17.] "Data Analysis Using Regression and Multilevel/Hierarchical Models" by Andrew Gelman andJennifer Hill (2007)

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

[18] The Internet of Things Applications for Challenges and Related Future Technologies & Development Mritunjay Kumar1, Km Annoo2, Raman Kumar Mandal3 https://scholar.google.com/citations?view_op=view_citation&hl=en&user=IRSdLtMAAAAJ&citation_for_view=IRSdLtMAAAAJ:u5HHmVD_uO8C