

DATA ANALYTICS

Short-Term Internship

Data-Driven Insights On Olympic Sports Performance And Participation

TEAM MEMBERS

1. A Pavan Kumar - LEAD
2. B Upendra
3. B Hari Prasad
4. V Akash
5. B Joshna Rani

(1) INTRODUCTION

a) OVERVIEW:

→ The project "Data-Driven Insights on Olympic Sports participation and performance" delves into the intricate world of Olympic sports by harnessing the power of data analysis. This study aimed to uncover patterns, trends, and correlations within the vast realm of Olympic sports, focusing on both participation rates and athletic performance. By meticulously analysing historical data spanning various Olympic events, disciplines, and countries, the research team sought to gain valuable insights into the factors influencing sports participation and the determinants of exceptional athletic performance on the grandest international stage.

• Objectives:

Data collection: Gather comprehensive data sets encompassing Olympic sports, athletes, countries, and performance metrics.

Exploratory Data Analysis (EDA): conduct in-depth exploratory analysis to identify patterns and trends

in sports participation across different regions and demographics.

Performance Analysis: Analyze athletes' performance data, examining factors such as training regimen, age, and nationality to understand the drivers of exceptional Olympic achievements.

Predictive modeling: Develop predictive models to forecast sports participation trends and predict potential medal winners based on historical performance data.

Policy Implications: Derive insights to inform sports policies, training programs, and investment strategies for enhancing both sports participation and performance at the Olympic level.

b) PURPOSE:

- The primary purpose of this project was to leverage data-driven methodologies to gain profound insights into the world of Olympic sports. The project aimed to achieve the following objectives.

Informed decision Making: Provide relevant data and

analysis to empower sports organizations, policymakers, and stakeholders with valuable information for making informed decisions regarding sports policies, investments, and resource allocation.

Understanding Participation Dynamics: Explore the intricate factors influencing sports participation rates at the Olympic level, including social, economic, cultural, and geographical variables. By understanding these dynamics, the project aimed to contribute to the development of strategies that encourage broader and more diverse participation in sports.

Enhancing Athletic Performances: Investigate the underlying factors contributing to exceptional athletic performance in Olympic sports. By analyzing training routines, historical data, and performance metrics, the project sought to identify patterns and best practices that could be used to enhance athletes' training programs and improve overall performance outcomes.

Predictive Insights: Develop predictive models to forecast sports participation trends and identify potential

Medal winners. By predicting future trends, sports organizations and athletes can adapt their strategies, training and participation to align with the evolving landscape of Olympic sports.

Policy Recommendations: Derive actionable policy recommendations based on data analysis. These recommendations aimed to guide policymakers in creating policies that promote sports participation, talent development, and the overall growth of the sports industry at both grassroots and elite levels.

In Summary, the purpose of the project was to utilize data analysis techniques, to uncover valuable insights, inform decision-making processes, enhance sports participation, improve athletic performance, and contribute to overall advancement of the Olympic sports domain.

(2)

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 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 Olympic success of a country can be done by efficiency analysis 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 spati-temporal. The video content of the analysis has the

the content rather than the significance of providing more interim information than structured collected data. In addition to these techniques, the exploratory data analysis used visual methods to provide a deep understanding and statistical summary of the data. The main aim of this analysis was to find out the growth in the performance of a country in the Olympics over the years.

a) EXISTING PROBLEM:

Some of the existing problems of Olympic Games were:

Infrastructure challenges: Hosting a mega event always involves urban renewal and regeneration. Yet developing sporting stadia, accommodation and transportation networks to cope with increased number of tourists and athletes is anything but straight forward.

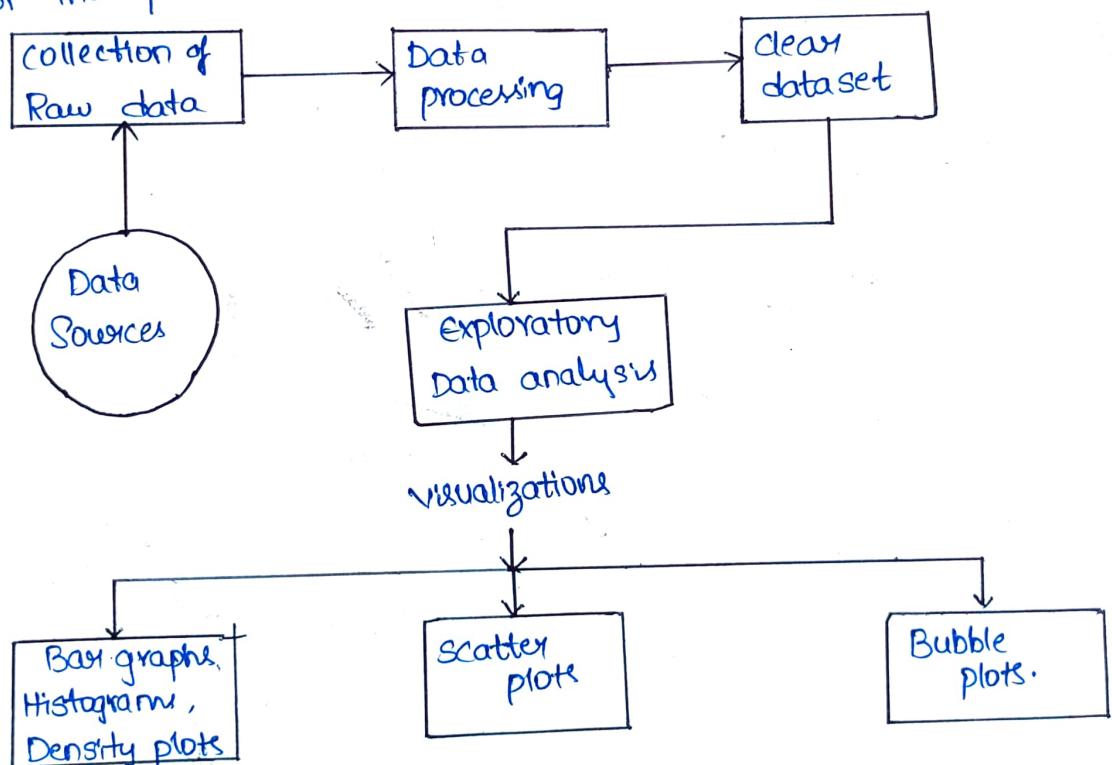
Olympics are becoming less sustainable and the winter games in Beijing was no exception. The winter Olympics require snow on the ground from sports such as skiing and snowboarding but the three cities Beijing, Yanqing and Zhangjiakou have arid climates. While snow reliability is declining globally because of planetary warming, Beijing is particularly unsuited for the winter participation in the first place.

To compensate, Beijing is largely relying on artificial snowmaking a process that consumes large quantities of water and electricity.

Many economists suggest that more radical changes may be needed to guarantee the long-term survival of the games without imposing massive burdens on host cities. Olympics could simply encourage hosting by region rather than by city.

b) Proposed Solution:

An approach is referred as a systematic path to reach a solution. Given below is the flow chart of the proposed approach for the problem.



Data collection: Data directly obtained from a data source such as dataset is known as Raw data. It is advised to take a abundance data because larger the volume

of data for analysis, the greater should be accuracy in the result.

Data pre-processing: It is the process of translating the raw data into useful data by conscientiously checking for errors and eliminating redundant incomplete, or incorrect data.

Exploratory data analysis: Analysis is done on data using various techniques like Text analysis, Diagnostic ~~vector~~ analysis, Exploratory data analysis.

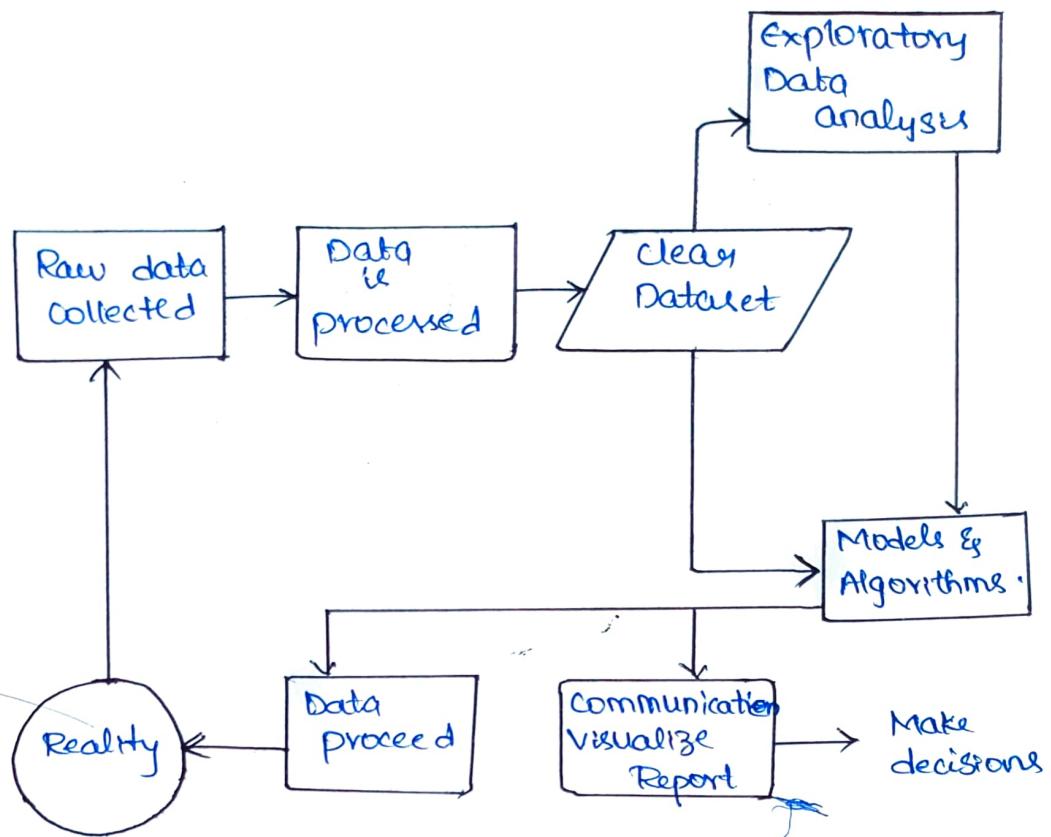
Various types of plots are

- Histogram
- Bar graph
- Box plot
- Scatter plot and many more. We can view the data in the visual format and also perform a comparative study between different plots.

(3) THEORETICAL ANALYSIS:

a) Block diagram:

This is an exploratory data analysis in which firstly raw data is collected and the data is processed then we create a data module and we clean the dataset with unwanted matter and create relationship to the tables dataset and apply models and algorithms and then we make the visualizations like bar graph, bubble plot, scatter plot, tree map etc. by using the data set modified and filtered and then the final data product will be obtained.



b) Hardware / Software designing:

Recommended to use a windows-enabled computer.

This recommendation is based on feedback from students who found many activities are easier to complete using windows 10 to 11.

Hard Drive SSD is preferred.

CPU - intel i5 minimum & intel i7 & i9 are preferred.

Some of the hardware not critical for success in our programs are like external keyboard, headphones second monitor for Remote students.

Development Environment Requirements: Jupyter Notebook, Google.

Operating Environment Requirements: Linux, Windows.

API: Streamlit an external Interface Required.

Functional Requirement: Python, Spider, Anaconda, Vscode, Bootstrapmade.

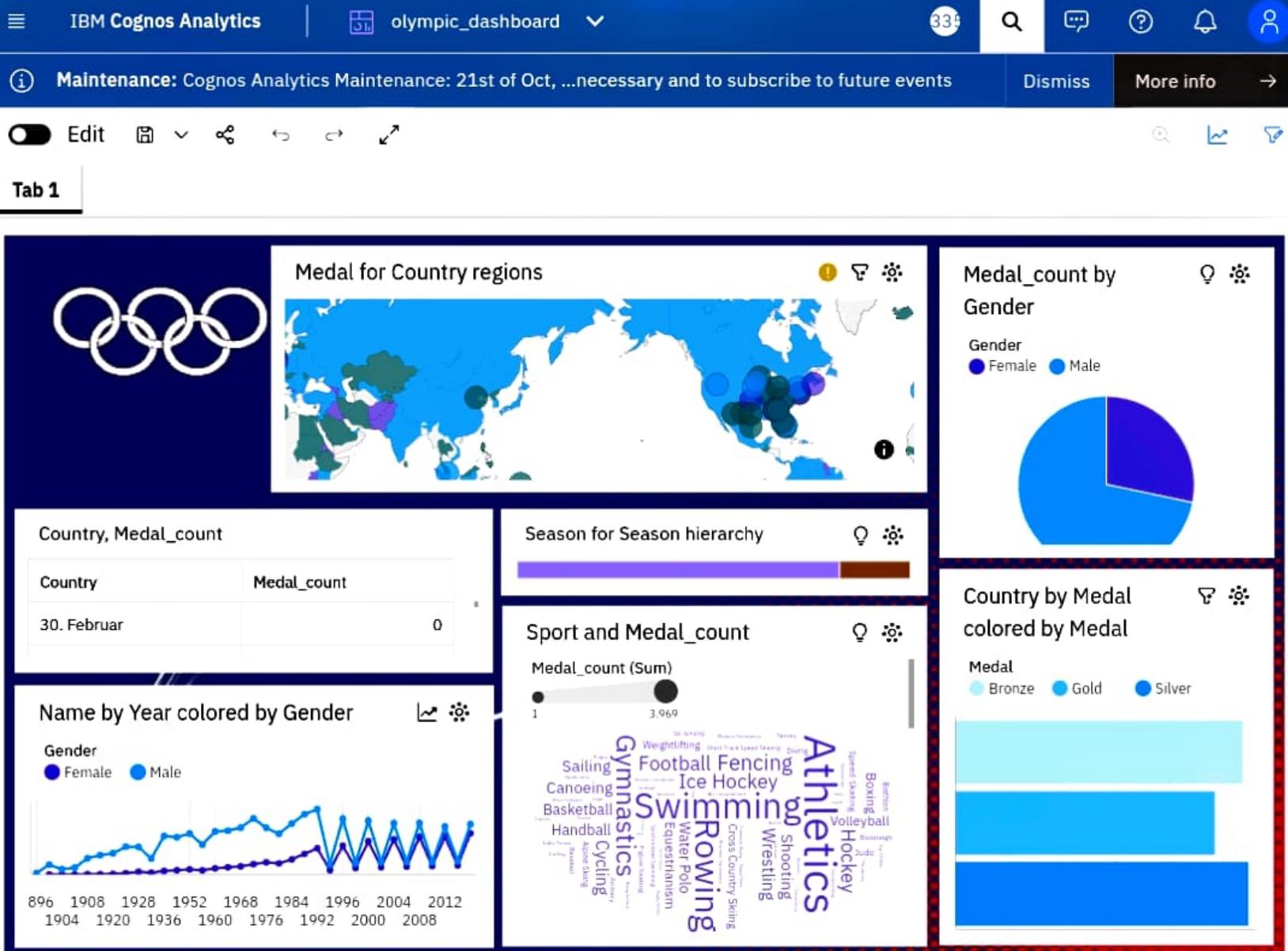
4) ADVANTAGES:

1. Performance improvement: Data analysis can identify areas for improvement in athletes' performance. Helping them train more effectively.
2. Talent Identification: It aids in the identification of emerging talents and potential future champions.
3. Strategic planning: Coaches and teams can develop better strategies based on historical data and opponent analysis.
4. Injury prevention: By tracking athletes physical condition. Injuries can be prevented & managed more effectively.
5. Fan engagement: Data insights can enhance fan engagement through statistics, visualizations, and interesting narratives.

5) DISADVANTAGES:

1. Privacy concerns: collecting personal and performance data can raise privacy concerns for athletes.

2. Data quality: The accuracy and reliability of data can be a challenge, especially for subjective metrics like judge scores.
3. Data overload: Handling a large volume of data can be overwhelming and it's essential to filter and interpret it correctly.
4. cost and resources: Data analysis requires significant resources which might not be available to all Olympic organizations.
5. Bias and Fairness: Analyzing data may inadvertently introduce bias if not done carefully, and fairness in competition can be compromised.
The effectiveness of any proposed data analysis solution would depend on addressing these advantages to enhance Olympic sport performance.



Maintenance: Cognos Analytics Maintenance: 21st of Oct, ...necessary and to subscribe to future events

Dismiss

More info

→

Edit

Filters

Sporting Triumphs

- Athletics emerges as the frontrunner with a total of 1,339 medals, combining gold, silver, and bronze, showcasing its enduring significance in the Olympic arena and the exceptional athleticism of participants.
 - Swimming follows closely with 1,099 medals, highlighting the sport's universal appeal and the remarkable achievements of swimmers from various nations, making it one of the most celebrated disciplines in the Olympic Games.



Country and Medals

- Analyzing the specific medal counts for individual countries
- The comprehensive data on gold, silver, and bronze medals for each country underscores the international diversity and competitive spirit in sports, reflecting the world's collective excellence in various disciplines.

| | | | | |
|--|--|--|--|--|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |



Olympic Games in 1924, Increase in the number of participating countries in both Summer and winter Olympics, the Average age of players in Olympic games, the Increase in the participation of the females in both Summer and Winter Olympics over the time, the 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 which depicts the evolution of Olympic Analysis Games over the time. Visualization of these factors has been done to explain and validate the Analysis in various Graphical formats like a Line graph, Scatter plots, Bar graphs, Density Plots etc.

(8) FUTURE SCOPE:-

We all know that any analysis is not perfect and it consists of some limitations which defines the future scope of the Research work. This project work also contains some limitations which we are considering as Future Scope of the project. These are :-

- We have visualized 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 on the data set after Analysis and can create a Predictive-Model which can predict the statistics of future Olympic games.
- We can also perform Correlation Analysis on the data set and analyze the relation between two continuous variables.

2:19

88.0 KB/S Lte 5G 22%



```
from flask import Flask  
  
app = Flask(__name__)  
  
@app.route('/')  
def hello_world():  
    return 'Hello, World!'  
  
if __name__ == '__main__':  
    app.run()
```