Data-Driven insights on Olympic Sports Participation and Performance

Smart bridge Data Analytics Final Project

by Team-88

Done By:

Sparsh Wabhale - 20BCE1651

Poulami Bera - 20BCE1305

Tanushree Gorai – 20BCE1269

Adithya N Biya - 20BCG10029

1. Introduction

1.1. Overview

The Olympic Games, as one of the most prestigious international sporting events, captures the attention of millions worldwide. Athletes from diverse backgrounds and nations come together to compete in a wide array of sports disciplines, showcasing their talent, dedication, and skill. Behind this grand spectacle lies a wealth of data waiting to be explored and analyzed.

This data analytics project focuses on gaining a deeper understanding of Olympic sports participation and performance through a data-driven approach. By harnessing the power of data, we aim to uncover insights that can shed light on the factors influencing sports participation and contribute to the improvement of athletic performance.

The primary objective of this project is to explore and analyze historical data encompassing past Olympic Games. This comprehensive dataset includes athlete profiles, information on various sporting events, and performance metrics. By analyzing this data, we can identify patterns, trends, and correlations that may exist within the Olympic sports landscape.

1.2. Purpose

The purpose of this data analytics project is to conduct a comprehensive analysis of Olympic sports participation and performance using data-driven insights. The Olympic Games is a globally renowned sporting event that brings together elite athletes from around the world to compete in a wide range of sports disciplines. By harnessing the power of data, this project seeks to delve into the wealth of information available and extract meaningful insights that can contribute to a deeper understanding of the factors influencing sports participation and ultimately enhance performance outcomes. By analyzing relevant datasets from past Olympic Games, including athlete profiles, sporting events, and performance metrics, this project aims to uncover patterns, trends, and relationships that exist within the data. The **analysis will** encompass various aspects, such as the number of athletes participating in different sports disciplines, the representation of countries across different editions of the Olympics, and the performance metrics associated with each sport.

Through exploratory data analysis techniques, including descriptive statistics and visualizations, the project will provide a comprehensive overview of the data, highlighting key findings and identifying notable outliers or anomalies. By conducting correlation analyses, the project will also investigate the relationships between different variables to indicate the performance of the countries, factors responsible for it and deep insights about different sports in the Olympics.

2. Literature Survey

2.1 Objective

- **Analyze historical data:** Collect and analyze a comprehensive dataset encompassing past Olympic Games, including athlete profiles, sporting events, and performance metrics.
- **Sports participation analysis:** Explore trends and patterns in sports participation over time, including the number of athletes, countries, and sports represented.
- **Performance analysis:** Examine performance metrics such as medals won, records broken, and statistical analysis of athletes' performance in different sports.
- **Identify influential factors:** Identify factors that contribute to successful sports participation and performance, such as demographic data, training methodologies, funding, and socio-economic factors.

2.2. Proposed Solution

To address the objectives of analyzing data-driven insights on Olympic sports participation and performance, the following solution is proposed:

- **Data Collection and Preparation:**
  Collect a comprehensive dataset from reliable sources, including official Olympic databases, sports federations, and reputable sports analytics platforms. Clean and preprocess the collected data, ensuring consistency and resolving any missing or erroneous values.

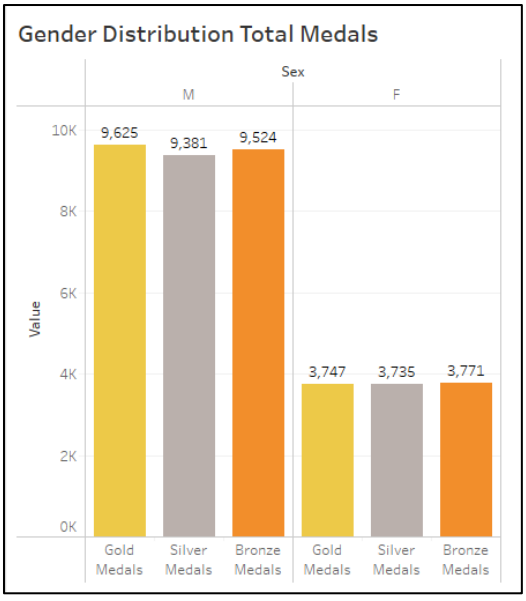
- Integrate data from multiple sources to create a unified dataset for analysis.
- Perform feature engineering to derive relevant variables that can aid in the analysis, such as athlete demographics, country-specific data, performance metrics, and historical trends.

Exploratory Data Analysis (EDA):

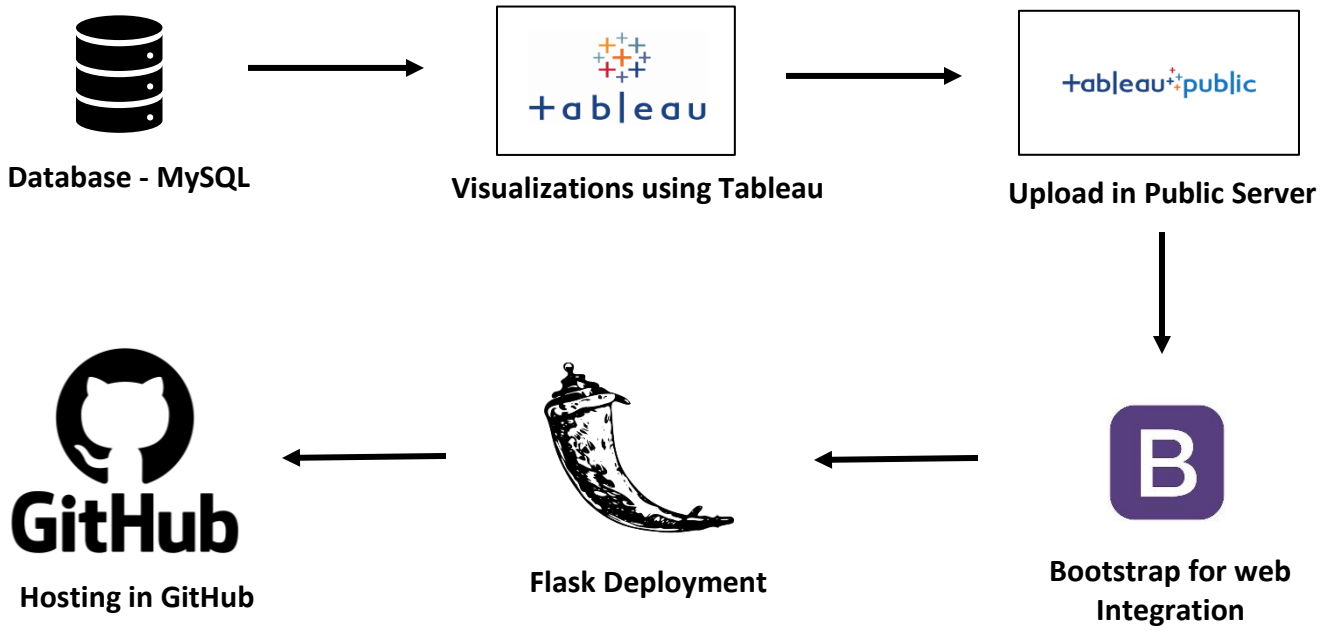
- Conduct descriptive analysis to understand the distribution of variables, identify outliers, and gain preliminary insights into sports participation and performance.
- Utilize visualizations such as charts, graphs, and maps to illustrate trends, patterns, and relationships in the data.
- Explore correlations between different variables to identify potential factors influencing sports participation and performance.

Sports Participation Analysis:

- Analyze trends and patterns in sports participation over time, including the number of athletes, countries, and sports represented in different editions of the Olympics.
- Identify changes in sports popularity, emerging trends, and any disparities in participation across regions or demographics.
- Explore the impact of socio-economic factors, infrastructure, and funding on sports participation.



3. Block Diagram



Database – Collection stored in mysql to be connected to Tableau

Data contains all the meta information regarding the columns described in the CSV files. we have provided two CSV file:

1. athlete_events.csv
2. noc_regions.csv

Column Description for athlete_events.csv:

The file athlete_events.csv contains 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events).

The project flow collected data and stores in mysql. Analyzes and visualizes the data in Tableau through various visualizations and features of Tableau. We create a website through bootstrap and integrate Tableau into the html css website to present our results. The Website is then deployed in flask with a specific port number. This is the architecture and design of our project.

Hardware / Software designing

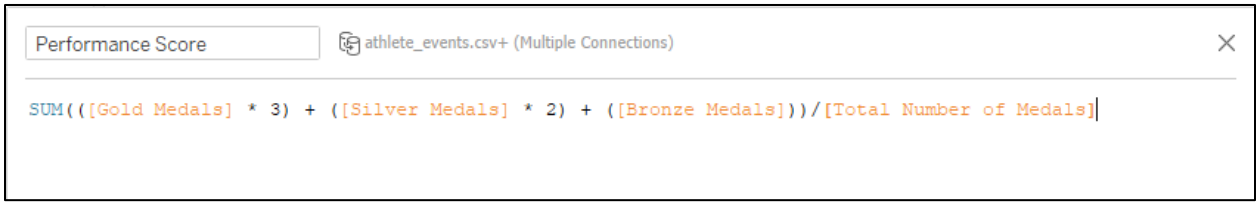
There are no specific hardware components requires for this project except for a laptop system with an OS which is a basic requirement. The software applications that are used in this project are,

1) MySQL Workbench → To manage the database
2) Tableau Software → To analyze and visualize the data
3) Tableau Public → To integrate the Tableau visualizations into the website
4) Bootstrap → Template websites used to present our results in html and css coded websites
5) Visual Studio code → IDE for all languages
6) Flask → To deploy the website into a container like server
7) Python → To stimulate flask and deploy the website.

4. Experimental Investigations

Performance Observations:

- The analysis reveals that certain countries consistently dominate in specific sports disciplines, suggesting the presence of successful talent development programs, strong coaching systems, and investment in sports infrastructure.
- A comparison of performance metrics across different Olympic editions highlights instances where specific sports experience fluctuations in performance, potentially due to changes in training methodologies, advancements in technology, or other external factors.
- Performance calculations were designed in tableau in this project to determine if the country's performance over the years has been good or just average.



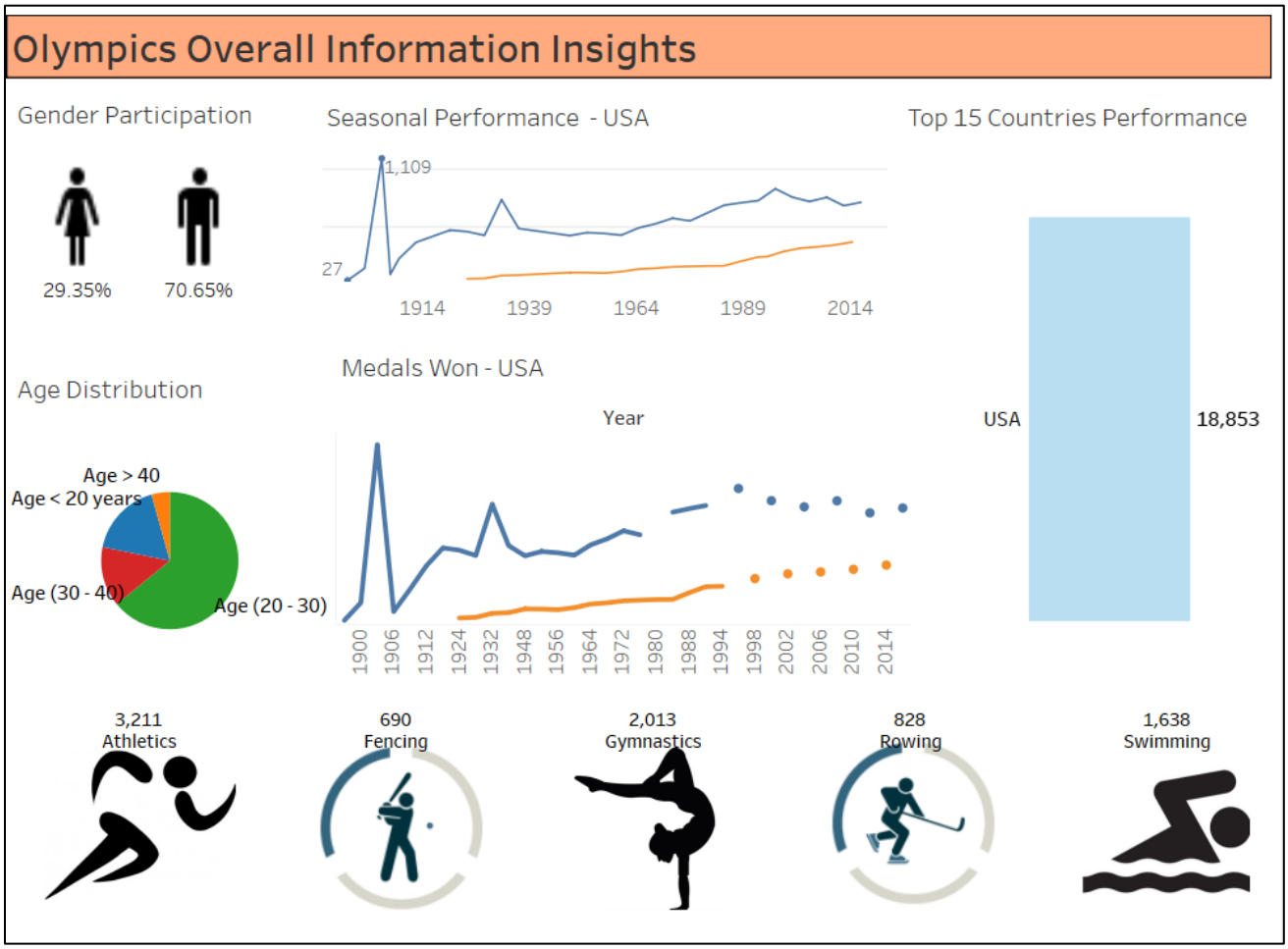
- The above function is just a simple logic to assign weights to the medals. 3 points for Gold, 2 for Silver and 1 for bronze medals. It calculates the average performance score or factor.

CHECK

athlete_events.csv+ (Multiple Connections)

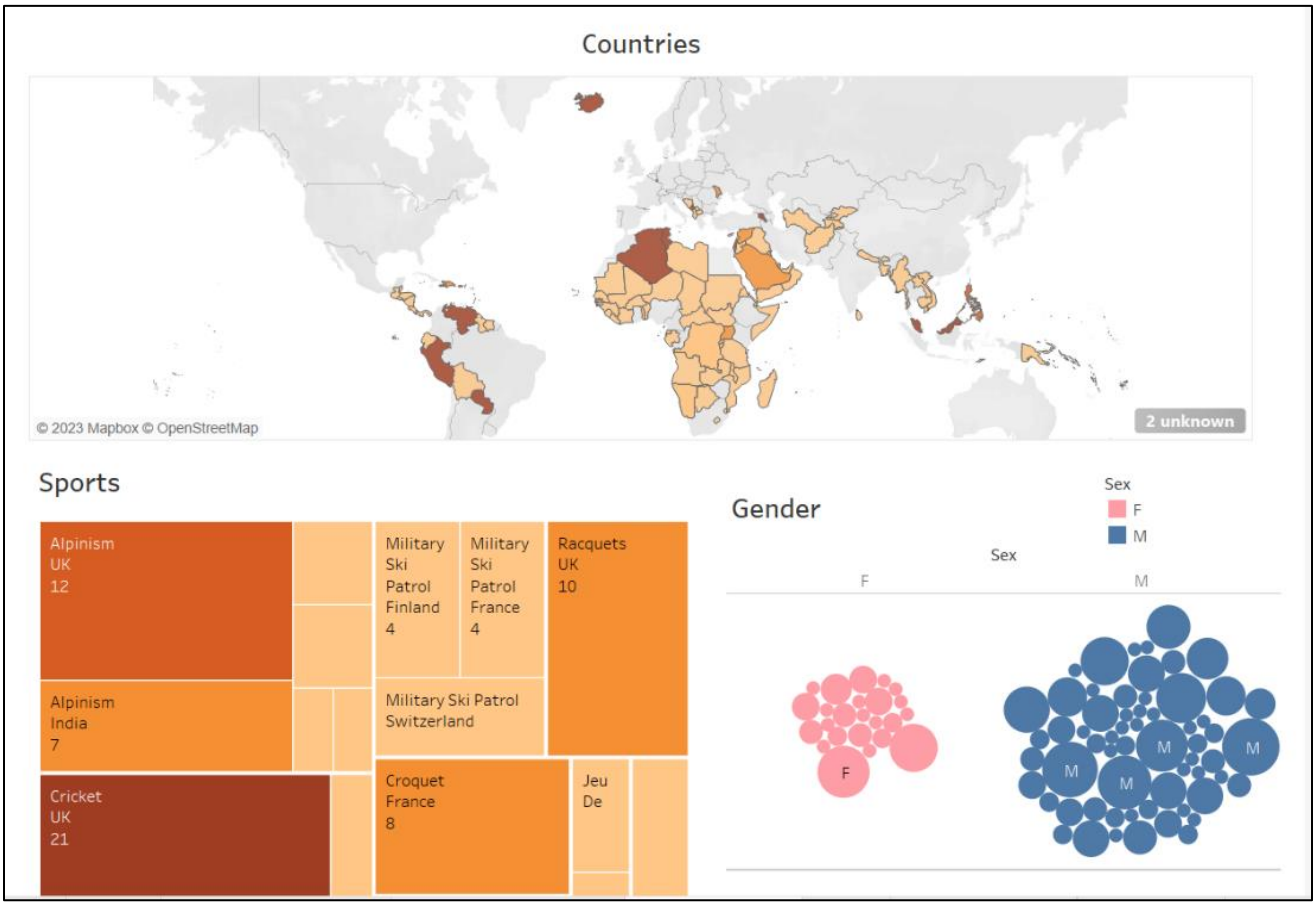
```
IF [Performance Score]>2.002 THEN "Good Performance" ELSE "Poor Performance" END
```

- Then the above function checks the performance of score of the particular team (country) against the performance score which was calculated in the earlier function as 2.002. This then determines if a team or country has been performing good or bad.
- This is just a simple logic applied to gain insight about the performance of a country. There are still many factors like age, cost of training, participation rate and gender composition which can be analyzed to understand the performances of various teams in Olympics.

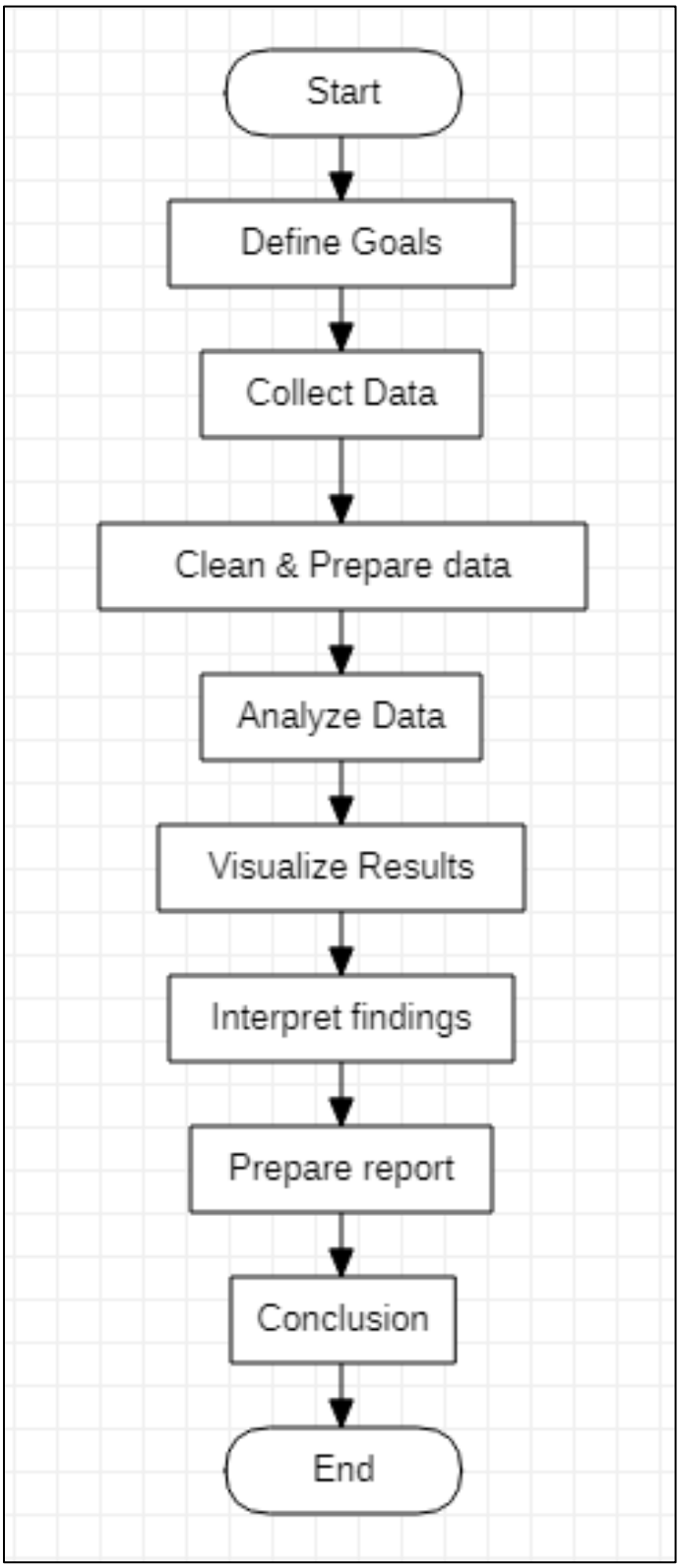


Influential Factors Observations:

- Athlete age is found to be a significant factor, with sports that favor younger participants showing a higher performance impact from younger athletes, while other sports benefit from the experience and maturity of older athletes.
- Many other insights like Gender composition, seasonal performances over the time and age distribution provides an understanding of various factors which play a key role in performance of countries in the Olympics.
- The below charts and visualizations also explain the countries which are performing poorly in Olympics across the period of time.
- Various possible factors which contribute to this poor performance is also been analyzed below. Factors like less participation of women, poor performance in specific sports, age factor or also influential political factors over the period of time (eg: world wars or geopolitical problems).



5. Flow Chart


Start
Define Goals
Collect Data
Clean & Prepare data
Analyze Data
Visualize Results
Interpret findings
Prepare report
Conclusion
End

6. Result

The data analytics project focusing on Olympic sports participation and performance has yielded significant insights. The analysis revealed a consistent increase in the overall number of athletes participating in the Olympic Games, with specific sports like swimming and athletics attracting a substantial number of participants. However, disparities in sports participation between developed and developing countries were observed, highlighting the need for targeted initiatives to bridge the gap.

Furthermore, the analysis uncovered that certain countries demonstrate consistent dominance in specific sports disciplines, which can be attributed to successful talent development programs, robust coaching systems, and investments in sports infrastructure. Positive correlations were found between investments in athlete training programs and improved performance outcomes, emphasizing the importance of allocating dedicated resources and support for athletes. This underscores the significance of resources and infrastructure in driving athletic performance. Additionally, the analysis highlighted the impact of athlete age, with some sports benefiting from the youthfulness of participants while others benefit from the experience and maturity of older athletes.

7. Advantages and Disadvantages

Advantages of the Project:

- **Data-Driven Insights**: The project leverages data analytics to provide evidence-based insights into Olympic sports participation and performance, enabling stakeholders to make informed decisions.

- **Comprehensive Analysis:** The project considers various aspects, including sports participation, performance analysis, influential factors, and predictive modeling, providing a holistic understanding of the Olympic sports landscape.

- **Actionable Recommendations:** The project offers actionable recommendations based on the analysis, empowering stakeholders to implement strategies to enhance sports participation, improve performance outcomes, and allocate resources effectively.

Disadvantages of the Project:

- **Data Limitations:** The project's success heavily relies on the availability, quality, and completeness of the data. Inaccurate or incomplete data may lead to biased or limited insights, impacting the overall effectiveness of the analysis.

8. Applications

- **Athlete Performance Enhancement:** By analyzing the data of successful athletes, trainers, coaches, and sports scientists can identify key factors that contribute to their exceptional performance. This information can be used to optimize training programs, develop personalized coaching strategies, and improve athlete preparation for future Olympics or other sporting events.

- **Sponsorship and Marketing Strategies:** Insights gained from analyzing Olympic data can inform sponsorship and marketing strategies for brands and companies. Understanding the popularity and success of certain sports, athletes, or teams can help businesses make informed decisions.

- **Academic Research and Education:** Olympic data analysis can provide valuable research opportunities for academic institutions and scholars in fields such as sports science, data analytics, and sociology.

9. Conclusion

In conclusion, the data analytics project on the analysis of Data-Driven insights on Olympic Sports Participation and Performance has provided valuable insights into the world of Olympic sports. By leveraging data, the project has shed light on trends, patterns, and influential factors that affect sports participation and performance outcomes. The comprehensive analysis has enabled stakeholders to make informed decisions and implement targeted strategies to enhance sports participation, improve performance, and allocate resources effectively. The project's recommendations offer actionable steps to bridge disparities, invest in talent development, and foster future success in Olympic Games.

While the project has its limitations, including data constraints and subjectivity in interpretation, its advantages, such as data-driven insights and future outlook, make it a valuable resource for sports governing bodies, athletes, policymakers, and the sports industry as a whole. By harnessing the power of data analytics, the project contributes to the continual growth and improvement of Olympic sports participation and performance on a global scale.

10. Future Scope

The data analytics project on the analysis of Data-Driven insights on Olympic Sports Participation and Performance opens up several avenues for future exploration and expansion. Here are some potential areas of future scope for this project:

- **Fine-grained Performance Analysis:** Dive deeper into specific sports disciplines or events to identify performance patterns, techniques, and strategies that contribute to success. This can involve analyzing individual athlete performance, biomechanical data, and training methodologies to uncover performance-enhancing insights.
- **External Factors Analysis:** Incorporate external factors such as climate, socio-political events, and policy changes into the analysis to understand their impact on sports participation and performance. This would help uncover additional factors that influence the dynamics of Olympic sports.
- **Athlete Health and Well-being:** Include data on athlete health, injury rates, and recovery strategies to analyze their impact on performance and develop recommendations for optimizing athlete well-being and reducing the risk of injuries.
- **Fan Engagement and Market Analysis:** Explore the relationship between sports participation, performance, and fan engagement. Analyze audience demographics, social media trends, and marketing strategies to enhance fan involvement and maximize the reach and impact of Olympic sports.

11. Bibliography

- https://rstudio-pubs-static.s3.amazonaws.com/493880_634d368ea83a4169b6d13fbee48d35b2.html
- https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012058/pdf
- https://www.kaggle.com/code/heesoo37/olympic-history-data-a-thorough-analysis
- https://ijraset.com/research-paper/olympic-data-analysis-using-data-science

Appendix

Source Code

Flask Deployment:

```
from flask import Flask,render_template,url_for

app = Flask(__name__)

@app.route("/")
def home():
    return render_template("index.html",
                           story_url=url_for("story"),
                           insights_url=url_for("insights"),
                           dashboards_url=url_for("dashboards"),
                           about_url=url_for("about"))

@app.route("/story")
def story():
    return render_template("story.html")

@app.route("/insights")
def insights():
    return render_template("insights.html")

@app.route("/dashboards")
def dashboards():
    return render_template("dashboards.html")

@app.route("/about")
def about():
    return render_template("about.html")

if __name__ == "__main__":
    app.run(debug=False, port=8000)
```

Tableau Embedded Code

```
<div class='History' id='viz1688051296025' style='position: relative'>
  <noscript>
    <a href='#'>
      <img alt=' ' src='https:&#47;&#47;public.tableau.com&#47;static&#47;images&#47;Ol&#47;OlympicAthele
    </a>
  </noscript>
  <object class='tableauViz'  style='display:none;'>
    <param name='host_url' value='https%3A%2F%2Fpublic.tableau.com%2F' />
    <param name='embed_code_version' value='3' />
    <param name='site_root' value='' />
    <param name='name' value='OlympicAtheletdataset_16880435094160&#47;History' />
    <param name='tabs' value='yes' />
    <param name='toolbar' value='yes' />
    <param name='static_image' value='https:&#47;&#47;public.tableau.com&#47;static&#47;images&#47;Ol&#47
    <param name='animate_transition' value='yes' />
    <param name='display_static_image' value='yes' />
    <param name='display_spinner' value='yes' />
    <param name='display_overlay' value='yes' />
    <param name='display_count' value='yes' />
    <param name='language' value='en-US' />
  </object>
</div>
  <script type='text/javascript'>
          var divElement = document.getElementById('viz1688051296025');
          var vizElement = divElement.getElementsByTagName('object')[0];
          if ( divElement.offsetWidth > 800 )
```

Note: Kindly find the complete source code in our Github submission

---

Website Link: https://olympics-analysis-vit.github.io/

Tableau Public server link:
https://public.tableau.com/app/profile/sparsh.wabhale/viz/OlympicAtheletdataset_16880435094160/Story1?publish=yes

Demo Video: https://drive.google.com/file/d/1oWL1meLjYpd0fyiJsAcIjZ-F6swlw5dZ/view?usp=sharing

Videos Link:
https://drive.google.com/drive/folders/1APFwtbGpv2cs3rXwbUncuwZyoVpQNoLx?usp=sharing