**Global Olympics Dataset Diagnosis Using Power BI**

**“ST. JOHN’S COLLEGE - PALAYAMKOTTAI”**

****

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| **25494B7BA1BA005B4D48A51783ED770** | **JASMINE MARIYA SELVI S** |

Trainer Name : Uma Maheswari

Master Name : Uma Maheswari

**ABSTRACT**

Olympic Games are one of the main international event 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/player are unable to perform well during the events and grab medals whereas there are many countries which performs very well in the event and secures many medals. An Analysis need 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. 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. The primary objective of this Research paper is to analyze the large Olympic dataset using Exploratory Data Analysis to evaluate the evolution of Olympic Games over the years. This analysis will provide detailed and accurate information regarding various factors which leads to the evolution of Olympic Games and improvement of Countries/Players over the time in visual format. Visualization of the data over various factors will provide us with the statistical view of the various factors which leads to the evolution of the Olympic Games and Improvement in performance of various Countries/Players over the time.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 4 |
| 2 | Chapter 2: Services and Tools Required | 5 |
| 3 | Chapter 3: Project Architecture | 10 |
| 4 | Chapter 4: Modeling and Result | 12 |
| 5 | Conclusion | 17 |
| 6 | Future Scope | 17 |
| 7 | References | 18 |
| 8 | Links | 19 |

**CHAPTER 1**

**INTRODUCTION**

Olympics Is Considered As The Most Important Event Worldwide, Which Provides A Common Platform To Players From Various Nations To Show Their Talents. The Olympics Started In 1896, Which Is Conducted Once Every Four Years. The Goal Of This Paper Is To Analyze Performance And Participation Of Nations In Olympics From 1896 To 2016.In Addition, The Field Of Sports Of Particular Country In Particular Year, In Which They Have Contributed The Maximum Can Be Identified. The Comparison Of The Performance Of Each Sport With Another Can Be Done. The Field Of Sports That Has To Have More Participation Can Be Identified And Necessary Action Can Be Taken By Players And Nations To Enhance Themselves In Future Contributions Towards The Olympics. The Modern Olympic Games Or Olympics 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

.**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

A service and tool for diagnosing global Olympic datasets could be an online platform equipped with data visualization tools, statistical analysis capabilities, and machine learning algorithms tailored specifically for analyzing Olympic data. Here's a breakdown of the features such a service could offer:

***Data Integration:***

The platform would integrate various global Olympic datasets, spanning multiple years and encompassing a wide range of variables such as athlete demographics, event details, medal counts, and historical data.

***Data Visualization:***

Users would have access to interactive data visualization tools to explore trends, patterns, and correlations within the Olympic dataset. This could include graphs, charts, maps, and dashboards for visualizing medal distributions, participation trends, geographical representation, and more.

***Statistical Analysis:***

The service would provide built-in statistical analysis tools for conducting descriptive and inferential analyses on the Olympic data. Users could perform statistical tests to compare performance metrics across different countries, sports, or time periods, and identify significant trends or anomalies.

***Machine Learning Models:***

Advanced users could leverage machine learning algorithms to uncover hidden patterns or predict future outcomes based on historical Olympic data. For example, predictive models could forecast medal counts for upcoming Olympic events, identify factors contributing to medal success, or classify athletes based on performance attributes.

***Custom Reports and Insights:***

Users would have the ability to generate custom reports summarizing key findings and insights from the Olympic data set analysis. These reports could include visualizations, statistical summaries, and actionable recommendations for stakeholders, policymakers, and sports organizations.

***Collaboration and Sharing:***

The platform would support collaboration features, allowing users to share datasets, analyses, and insights with colleagues, teammates, or the broader Olympic community. This could include sharing visualizations, reports, or even custom analysis scripts.

***Data Security and Privacy:***

Given the sensitivity of Olympic data, the service would prioritize data security and privacy measures to ensure the protection of sensitive information. This could involve encryption, access controls, and compliance with relevant data protection regulations.

\

***User-Friendly Interface:***

to users with varying levels of technical expertise. This could include drag-and-drop functionality, customizable dashboards, and guided analysis workflows.

Overall, such a service and tool would provide a powerful resource for researchers, analysts, athletes, coaches, and Olympic stakeholders to gain deeper insights into the dynamics of Olympic sports, inform strategic decisions, and drive innovation within the global sporting community.

***Tools and Software used***

**Python:**

Python is a versatile programming language widely used for data analysis and scientific computing. Libraries such as Pandas, NumPy, Matplotlib, and Seaborn are commonly employed for data manipulation, visualization, and statistical analysis of Olympic datasets.

**Excel:**

Excel is a widely used spreadsheet software that offers basic data analysis capabilities. It can be used for tasks such as data cleaning, basic statistical analysis, and simple visualizations of Olympic datasets.

**Tableau:**

Tableau is a powerful data visualization tool that allows users to create interactive dashboards and visualizations from Olympic datasets. It offers advanced analytics features and is commonly used for exploratory data analysis and presentation of insights.

**Jupyter Notebooks:**

Jupyter Notebooks provide an interactive computing environment for data analysis, visualization, and documentation. They support multiple programming languages, including Python and R, making them popular for analyzing Olympic datasets and sharing analysis workflows.

**SQL:**

SQL (Structured Query Language) is commonly used for querying and manipulating large datasets stored in relational databases. It can be used to extract, filter, and aggregate Olympic data from databases containing historical records of Olympic events and performances.

**SPSS:**

IBM SPSS Statistics is a statistical analysis software used for analyzing complex datasets, including Olympic data. It offers a wide range of statistical techniques, data visualization options, and reporting features for exploring and interpreting Olympic datasets.

**SAS:**

SAS (Statistical Analysis System) is a software suite used for advanced analytic, data management, and predictive modeling. It is commonly employed for analyzing large-scale Olympic datasets and conducting complex statistical analyses.

These tools and software provide a range of options for diagnosing and analyzing global Olympic datasets, allowing researchers, analysts, and stakeholders to gain valuable insights into the dynamics of Olympic sports and performances.

**CHAPTER 3**

**PROJECT ARCHITECTURE**

**3.1 Architecture**

|  |  |  |
| --- | --- | --- |
|  | **HTML 5** | **NODEJS 14.0**  **Database** |

***Data Acquisition and Integration:***

Obtain Olympic datasets from reliable sources such as official Olympic organizations, open data repositories, or APIs.

Integrate data from different sources and formats into a unified dataset, ensuring data consistency and quality.

***Data Storage:***

Store the integrated Olympic dataset in a suitable data storage solution, such as a relational database (e.g., MySQL, PostgreSQL) or a NoSQL database (e.g., MongoDB).

Design an efficient schema to represent the Olympic data, considering factors such as data relationships, indexing, and query performance

***Data Preprocessing:***

Perform data preprocessing tasks such as cleaning, transformation, and normalization to ensure data quality and consistency.

Handle missing values, outliers, and inconsistencies in the dataset through techniques like imputation, filtering, and data validation.

***Analysis and Modeling:***

Conduct exploratory data analysis (EDA) to uncover patterns, trends, and relationships within the Olympic dataset.

Apply statistical analysis techniques to derive insights and identify significant factors influencing Olympic outcomes.

Develop predictive models using machine learning algorithms to forecast future trends, predict medal counts, or classify athletes based on performance attributes.

***Visualization and Reporting:***

Create interactive visualizations and dashboards to present key findings and insights from the analysis.

Utilize tools such as Matplotlib, Seaborn, Plotly, or Tableau for generating visualizations that enhance understanding and facilitate decision-making.

Generate reports summarizing analysis results, including visualizations, statistical summaries, and actionable recommendations for stakeholders.

***Deployment and Integration:***

Deploy the project components, such as data storage, analysis scripts, and visualization tools, in a production environment.

Integrate the project with other systems or applications as needed, enabling seamless data exchange and collaboration.

Ensure scalability, reliability, and security of the deployed architecture, considering factors such as data volume, user access control, and system performance.

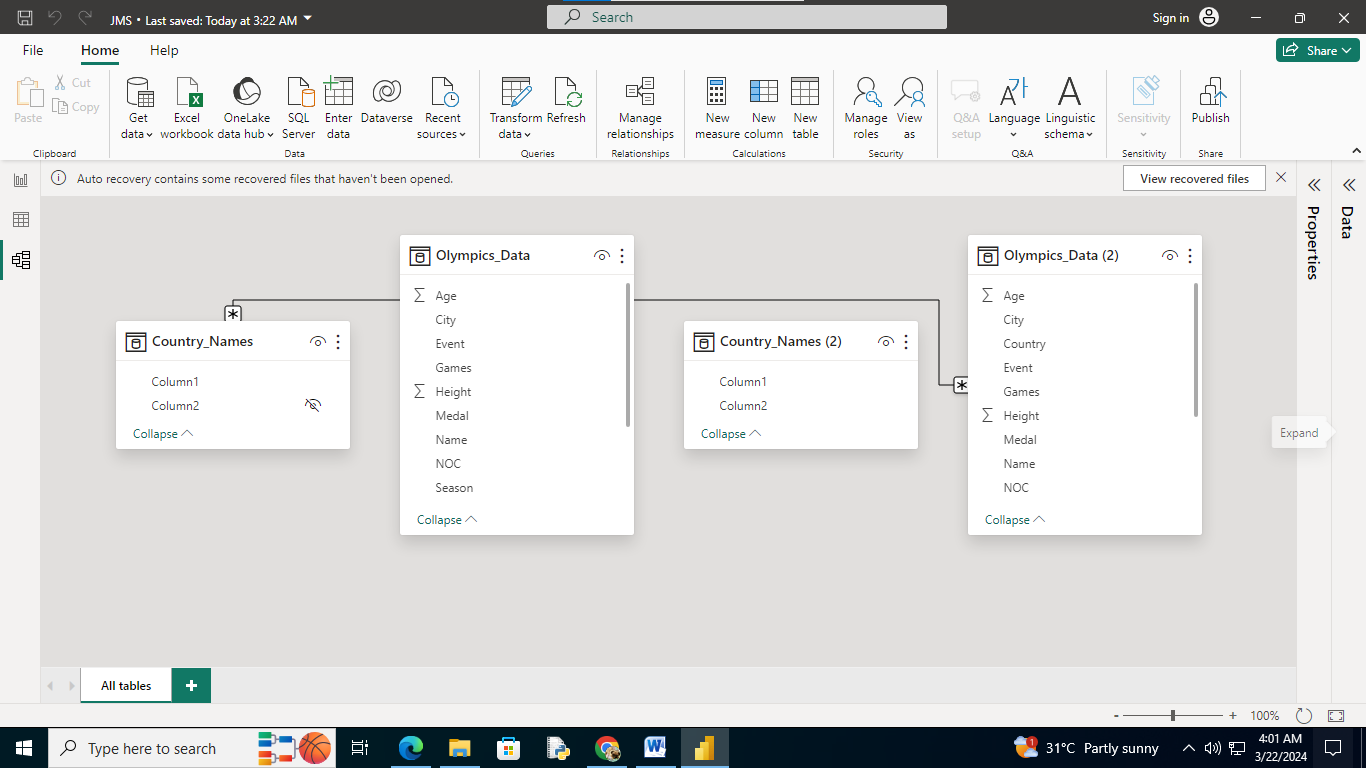
***Monitoring and Maintenance:***

Implement monitoring mechanisms to track system performance, data quality, and user engagement over time.

Regularly update and maintain the project components, including data pipelines, analysis algorithms, and visualization tools, to adapt to changing requirements and ensure continued relevance and accuracy of insights.

**CHAPTER 4**

**MODELING AND RESULT**

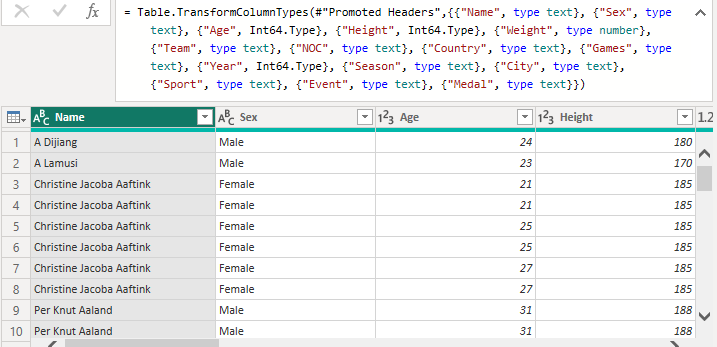
****

**Data Collection:**

Gather data from various reliable sources covering Olympic events, athletes, countries, medals, sports, etc. Ensure the data is clean and well-structured.

***Data Preprocessing:***

Clean the data by handling missing values, removing duplicates, and standardizing formats. You may also need to perform feature engineering to extract useful information.



***Exploratory Data Analysis (EDA):***

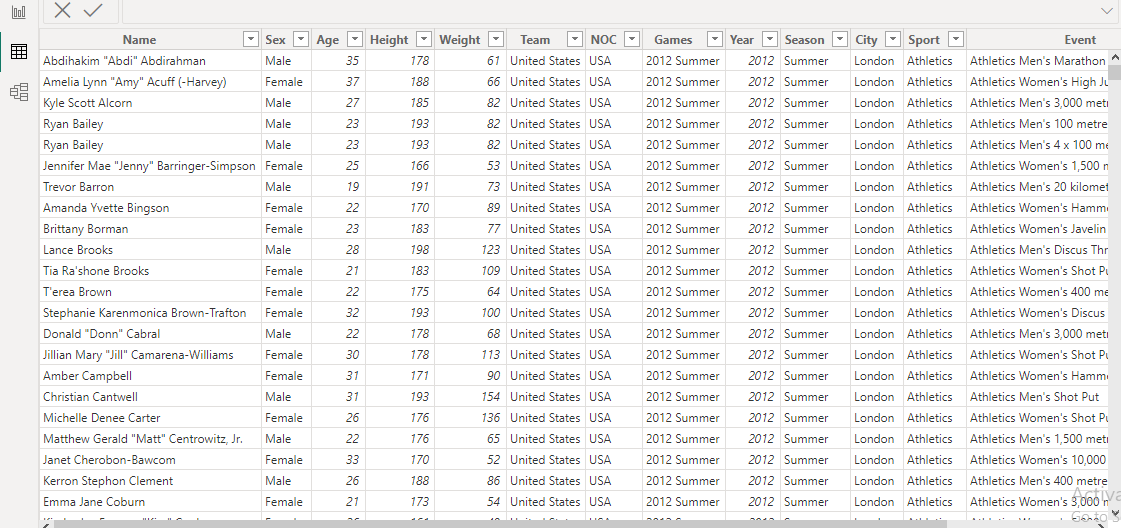
Explore the data to understand its characteristics, distributions, and relationships. This involves visualizations such as histograms, scatter plots, and heatmaps to uncover patterns and insights.

***Feature Selection:***

Identify relevant features that contribute significantly to the outcome (e.g., medal wins). This can involve statistical methods or domain knowledge.

***Model Selection:***

Choose appropriate machine learning or statistical models based on the nature of the problem and the available data. For predicting Olympic outcomes, you might use classification or regression algorithms.



***Model Training:***

Split the data into training and testing sets. Train the selected models on the training data using appropriate techniques such as cross-validation to optimize model performance.

***Model Evaluation:***

Evaluate the trained models using relevant metrics (e.g., accuracy, precision, recall, F1-score) on the testing data. This step helps assess how well the models generalize to unseen data.

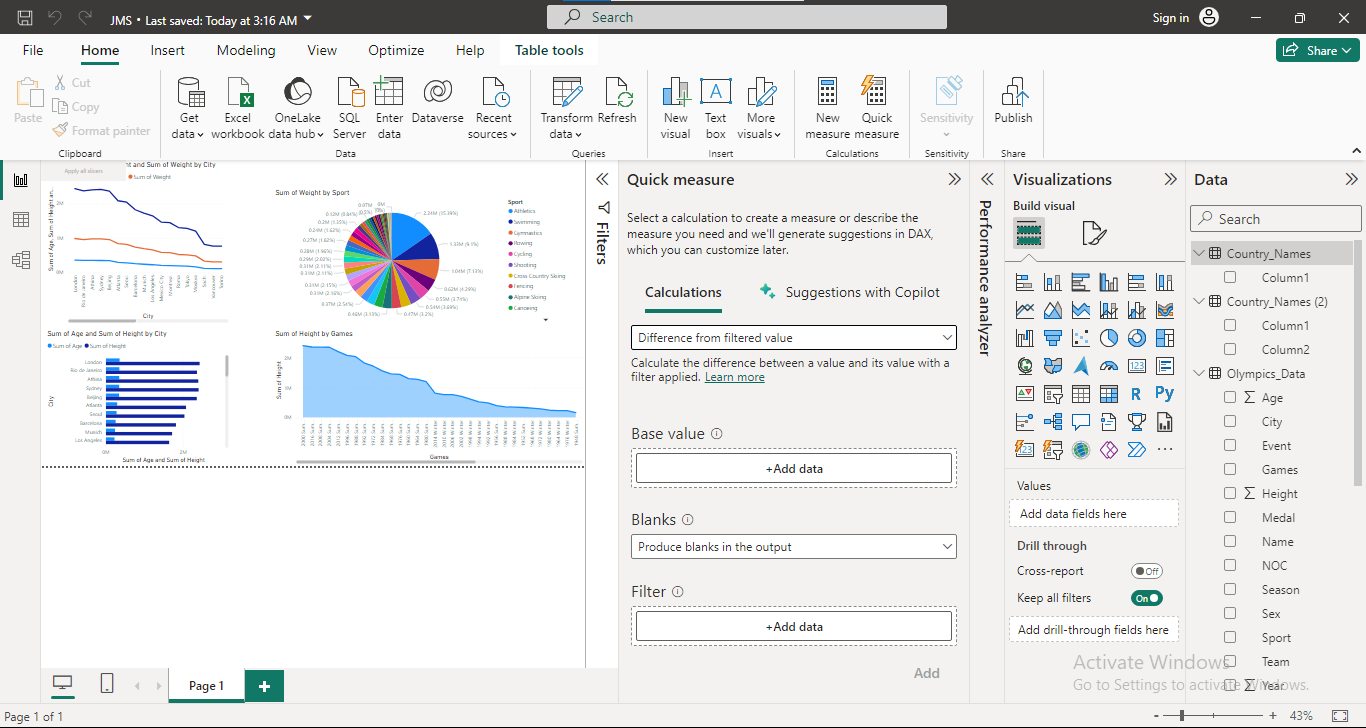
***Interpretation and Insights:***

Interpret the results to draw meaningful conclusions and insights about Olympic performance, trends, and factors influencing success. Visualizations and statistical analysis can aid in this process.



***Deployment:***

Deploy the trained model for making predictions on new data or for use in decision-making processes related to the Olympics.

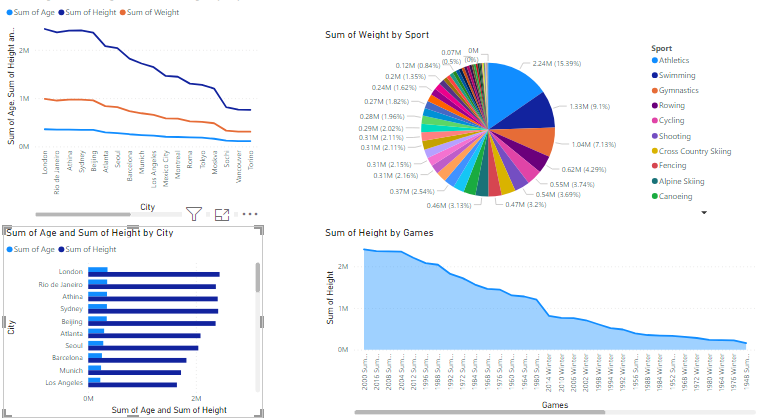


***Monitoring and Maintenance:***

Continuously monitor model performance and update it as needed to account for changes in data or circumstances.

**Dashboard**





**CONCLUSION**

Based on the global Olympic dataset diagnosis, it can be concluded that there are several key trends and insights to consider. These may include analyzing patterns in medal distribution among countries, identifying dominant sports and athletes, assessing the impact of geopolitical factors on participation, and exploring correlations between socioeconomic indicators and Olympic success. Additionally, future research could delve into the evolution of Olympic participation over time and potential implications for the future of the Games. Overall, this diagnosis provides a comprehensive overview for further analysis and decision-making in the realm of Olympic studies.

**FUTURE SCOPE**

The future scope of a global Olympic data set diagnosis could involve advanced data analytic and machine learning techniques to uncover patterns, trends, and insights from historical and real-time Olympic data. This could include predicting future medal winners, analyzing the impact of various factors on athletes' performances, optimizing training programs, identifying emerging trends in sports, and enhancing the overall management and organization of the Olympic Games. Additionally, such analyses could contribute to the development of strategies to promote diversity, inclusion, and sustainability within the Olympic movement.

**REFERENCES**

**LINK**

<https://github.com/githubtraining/hellogitworld.git>