**Rajeshwar V Shinde**

**20th July 2023**

**Project Documentation:**

**Olympic History**

**PostgreSQL & Power BI**

1. Introduction

The Olympic Project aims to analyse the "120 Years of Olympic History: Athletes and Results" dataset using PostgreSQL and Power BI. The dataset contains information about Olympic athletes, their events, and medal achievements over 120 years. This documentation outlines the data analysis process, SQL queries, Power BI reports, and key insights obtained from the dataset.

2. Data Source

The dataset was obtained from Kaggle and consists of two main tables:

Athlete\_events: Contains details about athletes, their participation in events, and medal achievements.

Noc\_Region: Maps National Olympic Committee (NOC) codes to country names and regions.

3. Data Extraction & Preparation

The data was extracted from the Kaggle dataset, and basic data cleaning and preparation were performed to ensure data quality. The "Noc\_Region" table was used to map NOC codes to country names for better representation in the visualizations.

4. PostgreSQL Database Design

The PostgreSQL database was designed with the following tables:

**athlete\_events:**

* ID
* Name
* Sex
* Age
* Height
* Weight
* Team
* NOC
* Games
* Year
* Season
* City
* Sport
* Event
* Medal

**noc\_region:**

* NOC
* Country name
* Notes

SQL queries were used to create these tables in the PostgreSQL database.

5. SQL Queries and Insights

All the queries and the solutions are provided in the file ‘Solution Set.docx’.

6. Power BI Reports and Visualizations

Three Power BI reports were created to present the insights:

Report 1 - Olympic Medal History:

* Line Chart: Count of medals by region and year.
* Clustered Column Chart: Count of medals by region and medal type.
* Filters: Year, region, medal type, and season.

Report 2 - Athlete Performance Analysis:

* Line Chart: Average age and average height by sport.
* Card: Total number of sports and total number of participants.
* Filters: Sex, sports, and medal type.

Report 3 - Sports & Medals Analysis:

* Map Visualization: Count of medals by region.
* Stacked Bar Chart: Count of regions by sport.
* Tree Map: Count of medals by sport.

7. Data Storytelling

The project narrative revolves around exploring Olympic history, athlete performance, and sports analysis. Through the visualizations, we discovered the following insights:

The top countries with the most medals are dominated by established sporting nations.

Olympic medals have grown over the years, with an increasing number of events and participating athletes.

Some sports consistently outperform others in terms of medal counts.

Athletes' age and height vary significantly across different sports, with trends observed in specific sports.

8. Conclusion

The Olympic Project successfully analysed the "120 Years of Olympic History: Athletes and Results" dataset using PostgreSQL and Power BI. It provided valuable insights into Olympic medal history, athlete performance, and sports analysis. The project demonstrates the power of data analysis and visualization tools in gaining meaningful insights from complex datasets.

9. References

Kaggle Dataset: "120 Years of Olympic History: Athletes and Results" (https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results)