

Olympic History: Athletes and Results

Lakshmi Pratyusha Bavanam, Abdul Khayyam sahib Shaik,
Kousalya Dasari, Murali Krishna Kethavath, and Manoj Kumar
Barugu

April 15, 2023

1 Introduction

The "Olympic History: Athletes and Results" dataset contains information on athletes and their performances at the Olympic games from 1896 to 2016. The data was collected from the official website of the Olympic Games, and it includes information on more than 270,000 athletes from over 200 countries. The dataset includes information on each athlete's name, gender, age, height, weight, team, and country of origin, as well as their performance in each Olympic event, including the type of event, their rank, and the medals they won.

2 Project Idea

The project aims to analyze the athlete's performance in the Olympics over 120 years. Loading the data using Jupyter Lab and analyzing the goals using PySpark and visualizing these goals using Tableau.

3 Technology Summary

- **Pyspark** for Data Processing and Analysis
- **Jupyter Lab** for Coding and Analysis
- **Tableau** for Visualization

4 Architecture Diagram and Summary

Import the libraries like pyspark and load the data set in jupyter in the CSV form. Write the SQL queries in Pyspark. To represent the goals in visualisation use tableau.

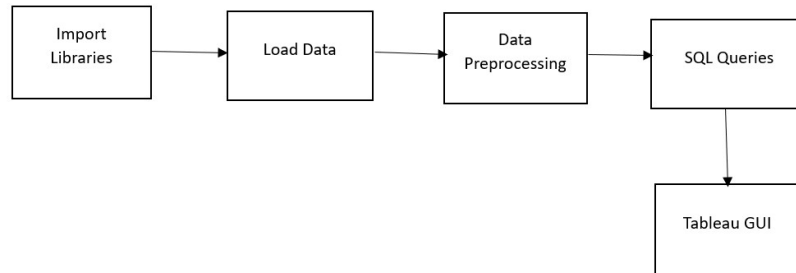


Figure 1: Architecture Diagram

5 Project Goals

- Which country has the highest medals (gold, silver, bronze)?
- What is the most common Olympic event for male athletes?
- What is the most common Olympic event for female athletes?
- Which Olympic games had the highest number of participating athletes?
- Average age of male athletes in each country with respective game
- Average age of female athletes in each country with respective game
- which owns the most medals according to the respective game
- what is the average height and weight of female and male athletes in each sport
- what are the different sports and year held in Paris?
- Which countries won the maximum gold Medals in the last held olympic competitions?

6 Data Set

Link: [Athletes and Results Data set](#)