

**Department of Data Science**  
**II B.Tech II Semester MiniProject-2024**  
**Submitted by**  
**Batch -03**

# **Olympic Data Analysis**

**Supervisor:**

**Mr.A.Kiran Kumar Reddy,**  
Assistant Professor,  
Department of Data Science and Cyber  
Security.

**Names**

**Roll no**

S.Vaishali	22R21A67J7
A.Tanvi Reddy	22R21A67D1
M.Sanjana	22R21A67G1
Ch.Kavisha	22R21A67E4

# Contents

- ❖ Problem Statement
- ❖ Abstract
- ❖ Introduction
- ❖ Literature Survey
- ❖ Existing system
- ❖ Disadvantages
- ❖ Proposed System
- ❖ Advantages
- ❖ System Requirements
- ❖ System Design(Architecture)
- ❖ Implementation
- ❖ Source Code
- ❖ Results
- ❖ Conclusion
- ❖ References

# Problem Statement

Analyzing Olympic Games data with traditional methods is difficult and time-consuming. There is a need for an easy-to-use platform that makes this data accessible and understandable for researchers and enthusiasts. This platform should provide clear, visual insights to simplify complex data analysis.



# Abstract

This project involves the creation of a web application for analyzing Olympic data using Python. The application allows users to select various data points from Olympic Games, including countries, sports, events, and medal counts, and visualize them in a variety of ways, such as graphs, tables, and maps. The data is obtained from a publicly available Olympic database, and the application uses various Python libraries, such as Pandas, Matplotlib, and seaborn, for data analysis and visualization. The project provides a user-friendly interface that allows users to interact with the data and gain insights into Olympic trends and statistics. Overall, this project demonstrates the potential of Python for creating data analysis applications that can be used for various purposes, including sports analysis, business intelligence, and scientific research. In this project, we are trying to create Olympic data analysis web application python for that we are using 120 years of Olympic history dataset.

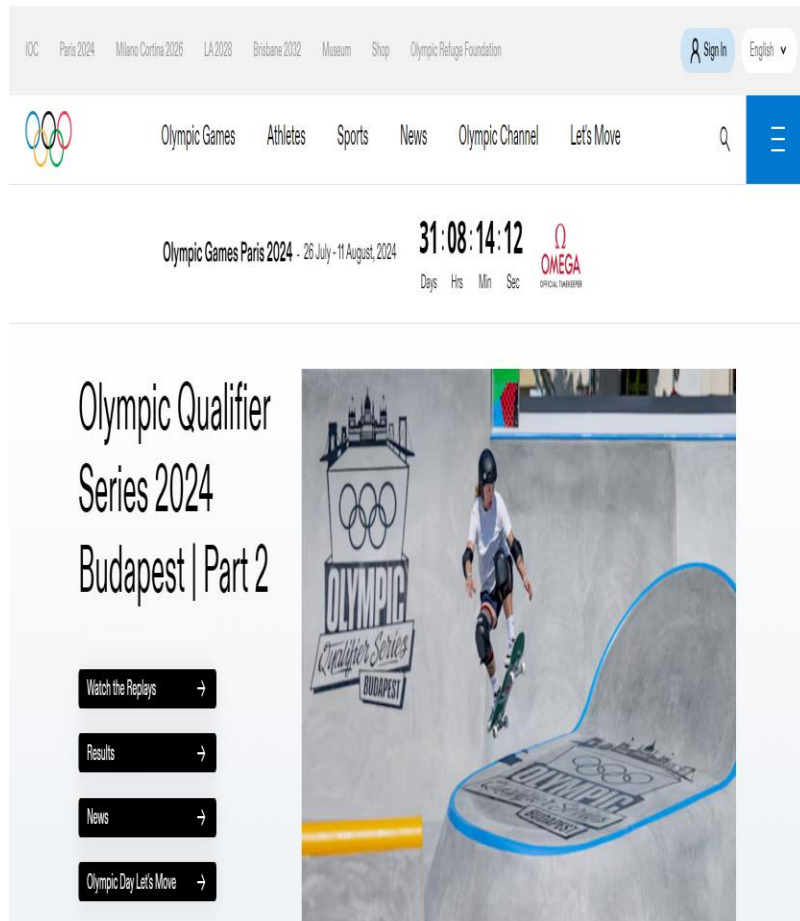
# Introduction

The main objective of this study is to analyze the various factors mentioned above which plays a vital role in the evolution of the Olympic Games over the years. The Analysis will include the visualization and explanation of the change in trends of the various factors over the years which will help to understand the information of Olympic Games. As the Olympic Games are one of the most important sporting events across the world, each country and each player tries to give their best performance in the event. To improve their performance, every country should perform such an Analysis which would help them in the improvement of their policies and strategies by providing accurate statistics to them.

# Literature survey

- ❖ D. Yamunathangam, G. Kirthicka and S. Parveen, "Performance analysis in project games using exploratory data analysis techniques", International Journal of Recent Technology and Engineering.
- ❖ Vaishnavi Laxman Farkande, Vaishnavi Rajshekhar Gurav, Tejas Pradiprao Borse and Tushar Phadtare, "Web Application of Olympic data analysis", International Journal Of Advance Research And Innovative ideas In Education.

# Existing System



- ❖ The existing model is the Official Olympic website, which consists of all the player details and game details.
- ❖ It has some additional news and latest updates portal, which gives us the latest information.
- ❖ Our project has all the data and details which the official Olympic website has, along with analysis of that data.
- ❖ Our project contains various sections like Overall medal tally, specific country year tally, overall analysis, specific country analysis, overall game-wise analysis, specific game wise analysis, athlete-wise analysis, and other comparison graphs and charts over the period of time.
- ❖ All these features are not provided by official Olympic website and availability of these features in our portal, makes our project more useful and dynamic.

# Disadvantages

- ❖ **Limited Scope:** The first research paper focuses on the 2016 Rio Olympics, limiting the generalizability of its findings to other Olympic Games or different contexts. Relying solely on a performance indicator from the public sector might overlook other crucial factors affecting Olympic Games' functioning.
- ❖ **Data Limitation:** Analyzing only historical volunteering activities might not account for potential changes or developments in recent Olympic Games, rendering the analysis incomplete.
- ❖ **Sampling Bias:** Depending on the willingness of volunteers to participate in surveys can introduce sampling bias, affecting the representativeness of the data. Official reports might lack in-depth insights into the volunteer experience, possibly missing nuances and qualitative aspects of their contributions.
- ❖ **Lack of Analysis:** While the official website provides vast amounts of data, it lacks analytical depth. It doesn't offer in-depth analysis of the data, limiting its usability for researchers and analysts.



# Proposed System

- ❖ In order to provide a user-friendly, all-in-one platform for Olympic data, we use python libraries like matplotlib, seaborn to generate plots and heatmaps. Here is the outline of the proposed Solution:
- ❖ **Medal Tally** gives all medals won by every country in the past 120 years.
- ❖ **Overall Analysis** gives the plots and heatmaps of no of participating nations over the years, most successful athletes in tabular form, and events, etc.
- ❖ **Country-Wise Analysis** gives the top athletes, medal streak, and area of sport excellence over the years in plots and heatmaps.
- ❖ **Athlete-Wise Analysis** presents distribution of age and by sport played, height v/s weight of men and women by sport and participation over the years with neat charts and suitable legends.

# Advantages

- ❖ The accurate and effective results are provided through analysis of past olympic data and insights are given in the form of graphs, tables, heatmaps and plots.
- ❖ it will be very helpful for the research scholars and other enthusiasts, if they find all that data in an analyzed form in a visually appealing way on a single platform.
- ❖ This web application is a user-friendly and interactive platform that enables users to explore Olympic data in a meaningful way.

# System Requirements

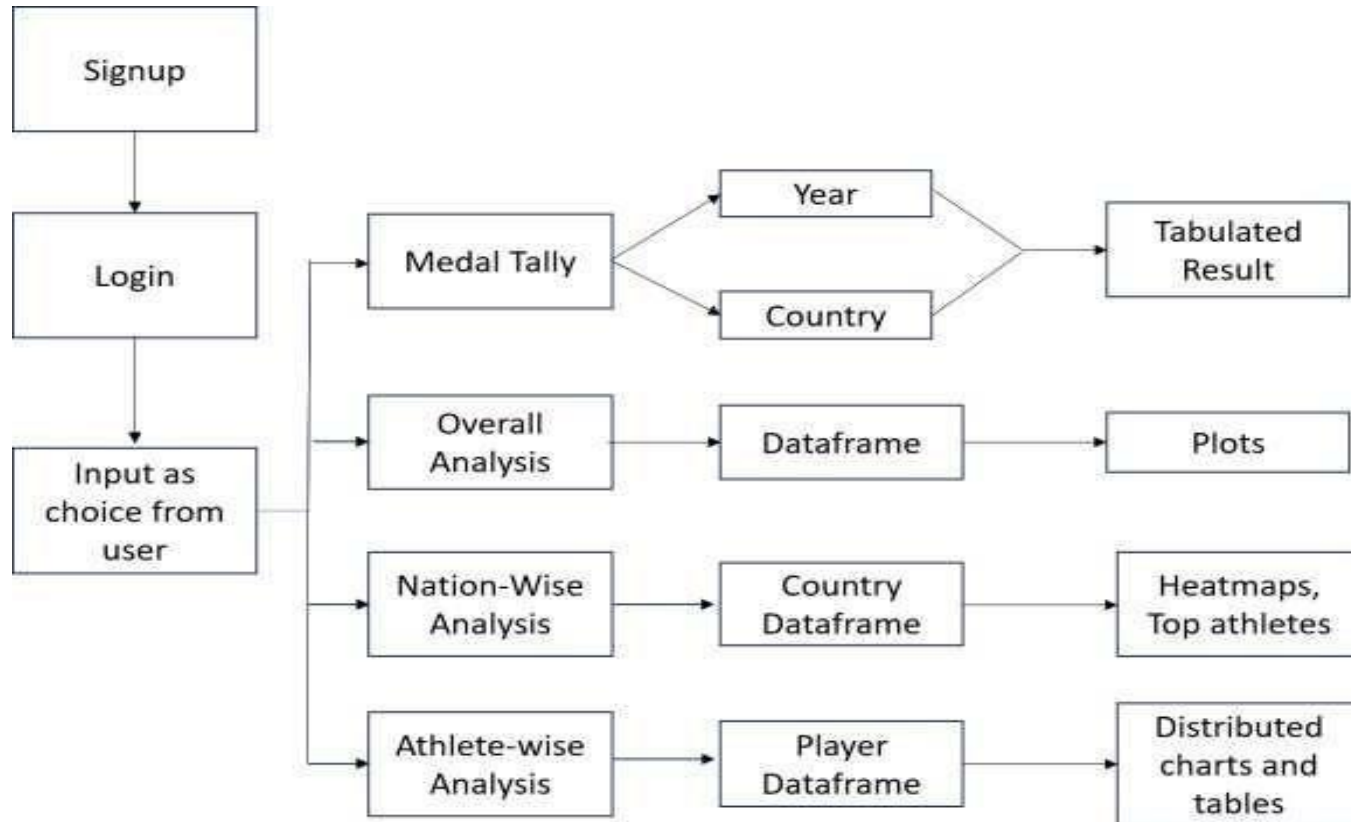
## ➤ Hardware Requirements:

- ❖ Processor: Intel Core i5 or higher
- ❖ RAM: 8 GB or higher
- ❖ Storage: 256 GB SSD or higher
- ❖ Display: 1920x1080 resolution or higher
- ❖ Internet Connection: Required

## ➤ Software Requirements:

- ❖ Frontend Streamlit: A popular open-source Python library that makes it easy to create and share custom web apps for machine learning and data science.
- ❖ Backend Python: It is a language frequently used to construct sites and programming, robotize undertakings, and direct information examination.
- ❖ IDE Visual Studio Code: It is a program that upholds numerous parts of programming advancement.

# System Design (Architecture)



# Implementation

- ❖ **Sign up page:** This module is to get the user register in the website by adding details like, E-mail, Username, Password, Confirm Password.
- ❖ **Login Page:** This module is used for user login . It includes fields like username, password.
- ❖ **Side Bar:** This page contains a logout button, olympic analysis in which user can select and search to get the desired analysis of medal tally, overall analysis, country-wise analysis and athlete-wise analysis in particular year and particular country.

# Source Code

## ❖ Setup and Data Loading

```
import streamlit as st
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.figure_factory as ff
import preprocessor, helper

st.set_page_config(page_title='Olympic Data Analytics', page_icon='img/favicon.png',
layout='wide')

# Load and preprocess data
df = pd.read_csv('data/athlete_events.csv')
region_df = pd.read_csv('data/noc_regions.csv')
df = preprocessor.preprocess(df, region_df)
```

## ❖ Sidebar and User Menu

```
st.sidebar.title('Olympics Data Analysis')
```

```
st.sidebar.image('img/olympics.png')
```

```
user_menu = st.sidebar.radio(
```

```
    'Select an Option',
```

```
    ('Medal Tally', 'Overall Analysis', 'Country-wise Analysis', 'Athlete-wise Analysis')
```

```
)
```

## ❖ Medal Tally and Overall Analysis

```
if user_menu == 'Medal Tally':
    years, country = helper.country_year_list(df)
    selected_year = st.sidebar.selectbox('Select Year', years)
    selected_country = st.sidebar.selectbox('Select Country', country)
    medal_tally = helper.fetch_medal_tally(df, selected_year, selected_country)
    st.title(f'{selected_country} Performance in {selected_year} Olympics' if selected_year !=
'Overall' and selected_country != 'Overall' else 'Overall Tally')
    st.table(medal_tally)

if user_menu == 'Overall Analysis':
    st.title('Top Statistics')
    col1, col2, col3 = st.columns(3)
    with col1: st.header('Editions'); st.title(df['Year'].unique().shape[0] - 1)
    with col2: st.header('Hosts'); st.title(df['City'].unique().shape[0])
    with col3: st.header('Sports'); st.title(df['Sport'].unique().shape[0])

    st.title('Events and Athletes over the years')
    st.plotly_chart(px.line(helper.data_over_time(df, 'Event'), x="Edition", y="Event"))
    st.plotly_chart(px.line(helper.data_over_time(df, 'Name'), x="Edition", y="Name"))
```



## ❖ Country-wise and Athlete-wise Analysis

```
if user_menu == 'Country-wise Analysis':  
    country_list = sorted(df['region'].dropna().unique().tolist())  
    selected_country = st.sidebar.selectbox('Select a Country', country_list)  
    st.title(f'{selected_country} Medal Tally over the years')  
    st.plotly_chart(px.line(helper.yearwise_medal_tally(df, selected_country), x="Year",  
y="Medal"))  
  
    st.title(f'{selected_country} excels in the following sports')  
    pt = helper.country_event_heatmap(df, selected_country)  
    if pt is not None and not empty: sns.heatmap(pt, annot=True)  
  
    st.title(f"Top 10 athletes of {selected_country}")  
    st.table(helper.most_successful_countrywise(df, selected_country))
```

```
if user_menu == 'Athlete-wise Analysis':  
    athlete_df = df.drop_duplicates(subset=['Name', 'region'])  
    st.title("Distribution of Age for winning medals")  
    st.plotly_chart(ff.create_distplot([athlete_df['Age'].dropna(),  
    athlete_df[athlete_df['Medal'] == 'Gold']['Age'].dropna()],  
    ['Overall Age', 'Gold Medalist'], show_hist=False, show_rug=False))  
  
st.title('Height Vs Weight')  
    selected_sport = st.selectbox('Select a Sport', sorted(df['Sport'].unique().tolist()))  
    sns.scatterplot(x=helper.weight_v_height(df, selected_sport)['Weight'],  
    y=helper.weight_v_height(df, selected_sport)['Height'],  
    hue=helper.weight_v_height(df, selected_sport)['Medal'])
```

# Results

## Olympics Data Analysis



Select an Option

- ☒ Medal Tally
- ☐ Overall Analysis
- ☐ Country-wise Analysis
- ☐ Athlete-wise Analysis

### Medal Tally

Select Year

Overall



Select Country

Overall



## Overall Tally

	region	Gold	Silver	Bronze	total
0	USA	1035	802	708	2545
1	Russia	592	498	487	1577
2	Germany	444	457	491	1392
3	UK	278	317	300	895
4	France	234	256	287	777
5	China	228	163	154	545
6	Italy	219	191	198	608
7	Hungary	178	154	172	504
8	Sweden	150	175	188	513
9	Australia	150	171	197	518
10	Japan	142	134	161	437
11	Finland	104	86	120	310
12	South Korea	90	85	89	264
13	Netherlands	88	97	114	299
14	Romania	88	95	120	303
15	Cuba	77	67	70	214
16	Poland	69	87	134	290

## Olympics Data Analysis



Select an Option

- ☐ Medal Tally
- ☒ Overall Analysis
- ☐ Country-wise Analysis
- ☐ Athlete-wise Analysis

## Top Statistics

Editions

**28**

Events

**651**

Hosts

**23**

Nations

**206**

Sports

**52**

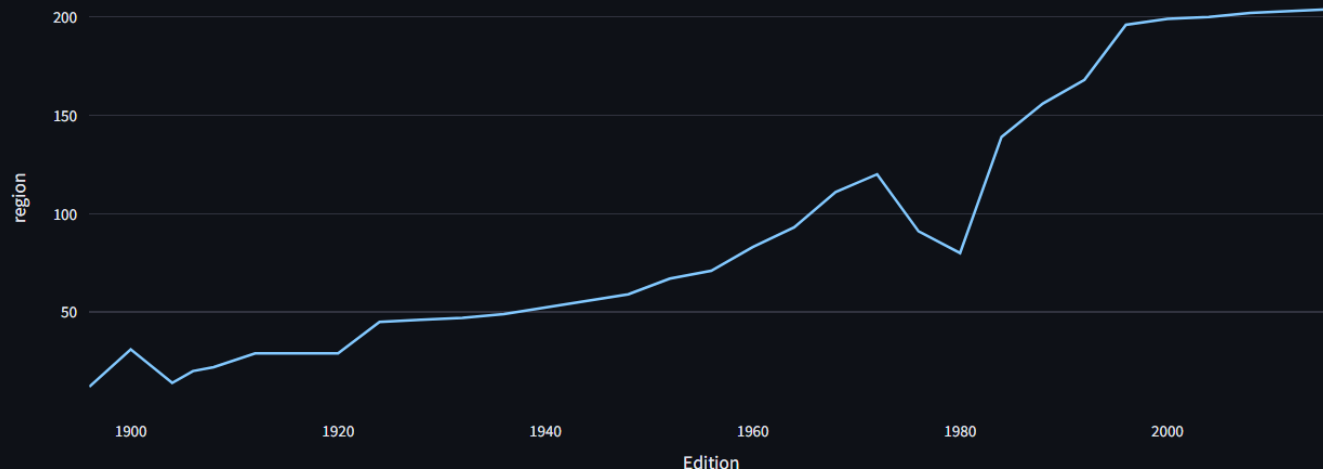
Athletes

**116122**

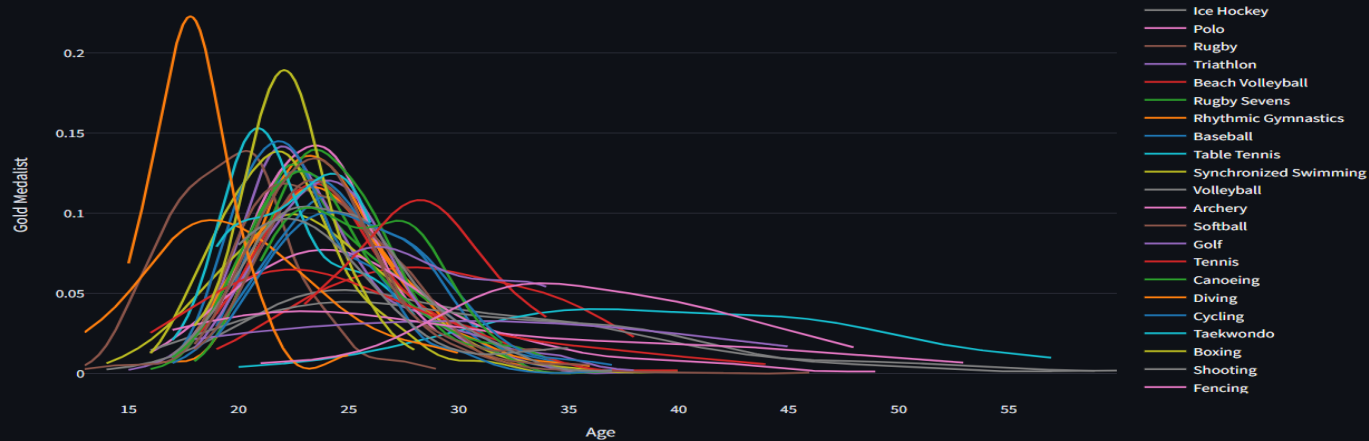
## Participating Nations over the years

200

## Participating Nations over the years ↗



## Distribution of Age w.r.t Sports(Gold Medalist)



## Top 10 athletes of Italy ⇄

	Name	Medals	Sport
0	Edoardo Mangiarotti	13	Fencing
14	Giulio Gaudini	9	Fencing
25	Maria Valentina Vezzali	9	Fencing
34	Giovanna Trillini	8	Fencing
43	Gustavo Marzi	7	Fencing
51	Piero D'Inzeo	6	Equestrianism
67	Raimondo D'Inzeo	6	Equestrianism
82	Giuseppe Delfino	6	Fencing
89	Aldo Montano	6	Fencing
98	Nedo Nadi	6	Fencing

# Conclusion

- ❖ Data analytics is a powerful approach to extract valuable information and insights from data, making it an indispensable tool in understanding and improving various aspects of the Olympic Games' performance and other fields of study.
- ❖ In conclusion, creating a website that provides all the overall analysis of 120 years olympic data which helps researchers, sports enthusiasts and beginner athletes to get their desired analysis accurately and efficiently.

# References

- ❖ Official Olympic Websites and Databases - (Olympics.co)](<https://olympics.com>)
- ❖ Academic Papers and Articles - Olympic Games: The impact of participation and hosting on national human rights practices" (Author: Jacquelyn K. Beauchamp, Year: 2020)
- ❖ Books - A Social Science Perspective" by Alan Tomlinson (Year: 2012) - Provides an in-depth analysis of the socio-economic and cultural impacts of the Olympics. - "Goldblatt, D. (2016).
- ❖ Statistical Analysis Reports - The Economic Impact of the Olympic Games" by PricewaterhouseCoopers (PwC) - A report detailing the economic outcomes of hosting the Olympics.
- ❖ Online Data Repositories and Tools - [KaggleDatasets](<https://www.kaggle.com/datasets>) - Search for Olympic datasets for statistical analysis and machine learning projects. (<https://www.olympicanalytics.com>) - A website offering various tools and visualizations for Olympic data analysis.

Thank You