

A Project Report

*On*

**OLYMPICS DATA ANALYSIS**

*By*

Rutuja Shete(TC264)

Sanika Paste (TC270)

*Under the guidance of*

Ms. Kirti Satpute



**Department of Computer Engineering**

**Marathwada Mitra Mandal's College of Engineering**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2022-2023**

Marathwada Mitra Mandal's College of  
Engineering  
Department of Computer Engineering,Pune\_51



---

Date:

## **CERTIFICATE**

This is to certify that,

Rutuja Shete(TC264)  
Sanika Paste (TC270)

of class T.E Computer have successfully completed their project work on “OLYMPICS DATA ANALYSIS” at MARATHWADA MITRA MANDALS COLLEGE OF ENGINEERING in the partial fulfillment of the Graduate Degree course in T.E Data Science Big Data Analytics Subject at the Department of Computer Engineering, in the academic Year 2023-2024 Semester – II as prescribed by the Savitribai Phule Pune University.

Ms. Kirti Satpute  
Guide

Prof KS Thakre  
Head of the Department  
(Department of Computer Engineering)

## ACKNOWLEDGEMENT

I feel great pleasure in expressing my deepest sense of gratitude and sincere thanks to my guide **Ms. Kirti Satpute** for their valuable guidance during the Project work, without which it would have been very difficult task. I have no words to express my sincere thanks for valuable guidance, extreme assistance and cooperation extended to all the **Staff Members** of my Department.

This acknowledgement would be incomplete without expressing my special thanks to Prof. KS Thakre, Head of the Department (Computer Engineering) for their support during the work.

I would also like to extend my heartfelt gratitude to my **Principal, Dr. V N Gohokar** who provided a lot of valuable support, mostly being behind the veils of college bureaucracy.

Finally, I would like to thank all the Teaching, Non- Teaching staff members of my department, my parents and my colleagues those who helped me directly or indirectly for completing of this Project successfully.

Name of Students

Rutuja Shete(TC264)

Sanika Paste (TC270)

# **CONTENTS**

**1. TITLE OF THE PROJECT**

**2. ABSTRACT**

**3. INTRODUCTION**

**4. PROBLEM STATEMENT**

**5. TECHNICAL**

- SOFTWARE

- DATASETS

- LIBRARY

- FUNCTIONS

**6. ALGORITHM**

**7. PERFORMANCE METRICS**

**8. RESULTS/ VISUALIZATIONS**

**9. REFERENCES**

# TITLE OF PROJECT

Title: **"Olympics Data Analysis: Unveiling Insights Through Data Science and Visualization"**

Description: Delve into 120 Years of Olympic History

The Olympics, a global spectacle of athletic prowess and unity, have captivated audiences for over a century. In this mini-project, we embark on a journey of comprehensive analysis, leveraging the power of data science techniques and visualization tools to uncover insights and trends spanning the vast landscape of Olympic history. Using Jupyter for data processing and PyCharm for front-end development, our dynamic dashboard offers a multifaceted exploration of Olympic evolution.

Our analysis transcends individual events and eras, encompassing a variety of graphical representations such as heat maps, distribution plots, and line graphs..

## ABSTRACT

The Olympics, a global spectacle of athletic prowess and unity, have captivated audiences for over a century. In this mini-project, we delve into a comprehensive analysis of 120 years of Olympics data, leveraging the power of data science techniques and visualization tools to uncover insights and trends. Utilizing Jupyter for data processing and PyCharm for front-end development, we present a dynamic dashboard that offers a multifaceted exploration of Olympic history.

Our analysis encompasses a variety of graphical representations, including heat maps, distribution plots, and line graphs, to elucidate patterns in Olympic performance over time. By meticulously examining data spanning numerous Olympic Games, we aim to identify key factors influencing success, uncovering nuanced narratives that transcend individual events and eras.

Furthermore, our dashboard showcases a curated list of the most successful athletes, shedding light on the individuals who have left an indelible mark on Olympic history through their remarkable achievements. Through this project, we offer a compelling narrative of the evolution of the Olympics, blending data-driven insights with captivating visualizations to provide a comprehensive understanding of this enduring global phenomenon

# INTRODUCTION

The Olympics stand as a beacon of athletic excellence and international unity, captivating audiences worldwide for over a century. Rooted in ancient tradition yet continually evolving, the Games represent a tapestry of human achievement, resilience, and camaraderie. In this mini-project, we embark on a comprehensive analysis of 120 years of Olympics data, delving deep into the rich history of this global phenomenon.

Harnessing the power of data science techniques and visualization tools, we aim to unravel hidden insights and uncover trends that have shaped the trajectory of the Olympics. From the inaugural Games of the modern era in Athens 1896 to the present day, each edition of the Olympics has left its mark on history, reflecting the evolving landscape of sports, culture, and society.

By leveraging Jupyter for data processing and PyCharm for front-end development, we present a dynamic dashboard that serves as a gateway to exploring the intricacies of Olympic history. Through a myriad of graphical representations, including heat maps, distribution plots, and line graphs, we seek to illuminate patterns and trajectories that define the essence of the Games.

Our analysis transcends mere statistics, delving into the stories behind the numbers to uncover the driving forces behind Olympic success. From the rise of dominant nations to the triumphs of individual athletes, each data point tells a story of determination, perseverance, and triumph in the face of adversity.

As we embark on this journey of exploration, our goal is not only to understand the past but also to glean insights that can inform the future of the Olympics. By shedding light on the key factors influencing success and showcasing the achievements of iconic athletes, we hope to offer a comprehensive understanding of the enduring legacy of the Games and the indelible impact they have on the world stage.

# PROBLEM STATEMENT

Develop Olympic data analysis system using machine learning.

## TECHNICAL

### SOFTWARE USED

- Windows 10
- Jupyter
- Pycharm
- Python ○ GUI

### DATASET

- IPL Olympics dataset [here](#)

### LIBRARIES USED

- NumPy
- Pandas
- Scikit-learn
- Plotly
- Matplotlib
- Streamlit
- Preprocessor
- Pandas
- Seaborn

## FUNCTIONS

- load\_data()
- kmeans\_clustering()
- main()
- preprocess()
- fetch\_medal\_tally()
- medal\_tally()
- country\_year\_list()
- most\_successful()
- yearwise\_medal\_tally()
- unique()
- merge()
- drop\_duplicates()
- fit\_predict()

## ALGORITHM

### STEP 1

Importing all the required libraries.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import sklearn
```

### STEP 2

Import the dataset :

```
df = pd.read_csv('athlete_events.csv')
region_df = pd.read_csv('noc_regions.csv')
```



### STEP 3

Analyzing the first 5 rows of the dataset

```
df.head(5)
```

### STEP 4

Brief summary of the olympic dataset.

```
df.describe()
```

### STEP 5

Checking whether there are any null values present in the dataset.

```
df.isnull().sum()
```

### STEP 6

Selecting only summer Olympic events.

```
df = df[df['Season'] == 'Summer']
```

### STEP 7

Applying some logics to analyse the data and visualizing the data.

```
medal_tally=df.drop_duplicates(subset=['Team','NOC','Games','Year','City','Sport','Event','Medal'])
```

### STEP 8

Making clusters using Kmeans clustering algorithm

```
Kmeans=KMeans(n_clusters=3,random_state=42)  
clusters = kmeans.fit_predict(scaled_data)
```

# PERFORMANCE METRICS

The Silhouette Score is a metric used to evaluate the quality of clusters formed by a clustering algorithm, such as K-means. It provides a measure of how similar an object is to its own cluster compared to other clusters. The Silhouette Score ranges from -1 to 1, where:


- A score close to +1 indicates that the data point is well-clustered and lies within the correct cluster.
- A score around 0 indicates that the data point is close to the decision boundary between two clusters.
- A score close to -1 indicates that the data point may have been assigned to the wrong cluster.

# RESULTS/ VISUALIZATIONS

← → ↻ 🌐 localhost:8501 🔍 ☆ 🗄️ 🌐 ⋮

Deploy ⋮

Olympics Analysis



Select an Option

☒ Medal Tally

☐ Overall Analysis

☐ Country-wise Analysis

☐ Athlete wise Analysis

Medal Tally

select Year

2016 ▼

Select Country

India ▼

Performance of India in 2016 Olympics

	region	Gold	Silver	Bronze	total
0	India	0	1	1	2

Activate Windows  
Go to Settings to activate Windows.



localhost:8501

Olympics Analysis

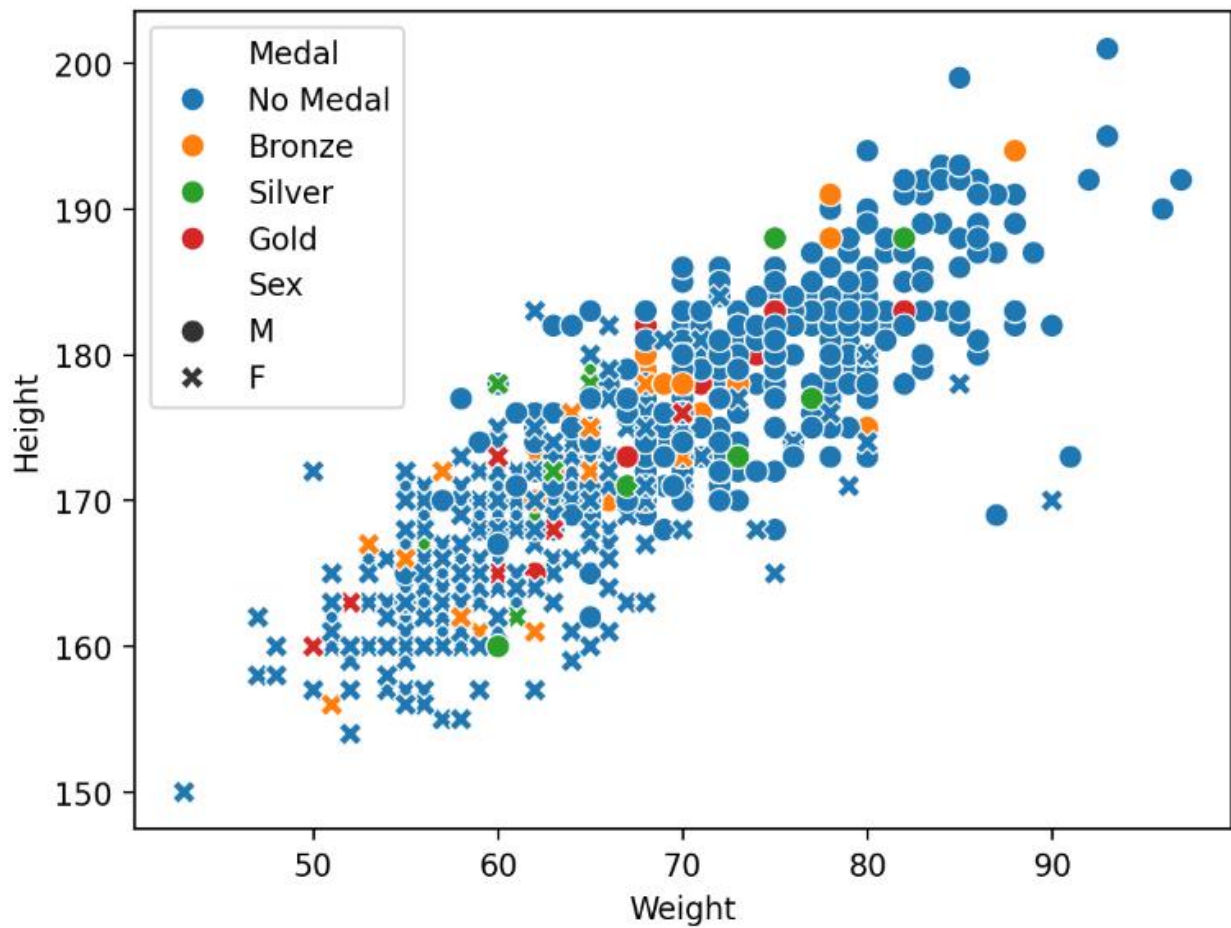
Select an Option

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

## Top 10 Atheletes of India

	Name	MedalCount	Sport
0	Udham Singh Kular	4	Hockey
4	Leslie Walter Claudius	4	Hockey
8	Shankar Pillay Laxman	3	Hockey
11	Ranganathan Francis	3	Hockey
14	Richard James Allen	3	Hockey
17	Balbir Singh	3	Hockey
20	Harbinder Singh Chimni	3	Hockey
23	Balbir Singh Dosanjh, Sr.	3	Hockey
26	Victor John "V. J." Peter	3	Hockey
29	Dhyan Chand Bais	3	Hockey

Activate Windows  
Go to Settings to activate Windows.



## REFERENCES

### ➤ Youtube links:

- Streamlit :  
<https://www.youtube.com/live/YzvMpvXyUfs?si=pWWYLCVTC8ZQWLEh>
- Jupyter:  
[https://youtu.be/q1FttL\\_G1G4?si=ts19AhGpvTuukVD3](https://youtu.be/q1FttL_G1G4?si=ts19AhGpvTuukVD3)

### ➤ Wikipedia:

[https://en.wikipedia.org/wiki/Olympic\\_Games](https://en.wikipedia.org/wiki/Olympic_Games)