Rajalakshmi Engineering College(An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project : Summer Olympics Medal Analysis (1896–2024)

Report submitted by

REG. NO. : 231501074

STUDENT NAME : KARTICKEEYAAN M

YEAR : 2023-2027

SUBJECT CODE : AD23632

INDEX

Chapter 1: Abstract

This project presents a comprehensive visual analytics study of the Summer Olympics Medal Dataset (1896–2024). The primary objective is to analyze patterns and trends in medal distributions across different countries, sports, genders, and years using Python (Matplotlib, Seaborn, Plotly) , Power BI, and Tableau. The project involves data preprocessing, static and interactive visualizations, and dashboard creation to highlight global dominance in sports, athlete achievements, and the evolution of the Olympic Games over time. By leveraging multiple frameworks, this study offers an integrated comparison of visualization tools for sports analytics.

Chapter 2: Introduction

The **Olympic Games** are the world's foremost sporting event, showcasing global athletic excellence since **1896**. Analyzing over a century of Olympic data provides valuable insights into sports trends, gender participation, and national dominance.
This project applies **data visualization and analytics** techniques to uncover these insights and present them through interactive dashboards.

Tools used:

- **Python (Matplotlib, Plotly):** For statistical and exploratory data visualization.
- **Power BI:** For dynamic, business-style dashboards with filters and drill-downs.
- **Tableau:** For aesthetically rich, interactive data exploration.

The outcome provides a data-driven understanding of Olympic history and comparative performance among nations.

Chapter 3: Dataset Description

Dataset Name: Summer Olympics Medals (1896–2024)

Source: Kaggle / Official Olympic data archive

Data Size: ~271,000 records

Key Columns:

- *Year* — Olympic year
- *City* — Host city
- *Sport* — Category of sport
- *Event* — Specific competition
- *Name* — Athlete name
- *Sex* — Gender (M/F)

- *Team* — Country/Team name
- *NOC* — National Olympic Committee code
- *Medal* — Gold/Silver/Bronze
- *Season* — Summer/Winter (filtered for Summer)

	A	B	C	D	E	F	G	H	I	J	K
1	player_id	Name	Sex	Team	NOC	Year	Season	City	Sport	Event	Medal
2	0	A Dijiang	M	China	CHN	1992	Summer	Barcelona	Basketball	Basketball	No medal
3	1	A Lamusi	M	China	CHN	2012	Summer	London	Judo	Judo Men'	No medal
4	2	Gunnar Aa	M	Denmark	DEN	1920	Summer	Antwerper	Football	Football M	No medal
5	3	Edgar Aaby	M	Denmark/S	DEN	1900	Summer	Paris	Tug-Of-Wa	Tug-Of-Wa	Gold
6	26	Cornelia (-	F	Netherlan	NED	1932	Summer	Los Angele	Athletics	Athletics W	No medal
7	27	Cornelia (-	F	Netherlan	NED	1932	Summer	Los Angele	Athletics	Athletics W	No medal
8	29	Einar Aalto	M	Finland	FIN	1952	Summer	Helsinki	Swimming	Swimming	No medal
9	31	Jyri Aalto	M	Finland	FIN	2000	Summer	Sydney	Badmintor	Badmintor	No medal
10	32	Minna Aalt	F	Finland	FIN	1996	Summer	Atlanta	Sailing	Sailing Wo	No medal
11	33	Minna Aalt	F	Finland	FIN	2000	Summer	Sydney	Sailing	Sailing Wo	No medal
12	35	Arvo Aalto	M	Finland	FIN	1912	Summer	Stockholm	Swimming	Swimming	No medal
13	36	Arvo Aalto	M	Finland	FIN	1912	Summer	Stockholm	Swimming	Swimming	No medal
14	37	Arvo Aalto	M	Finland	FIN	1920	Summer	Antwerper	Swimming	Swimming	Bronze
15	38	Arvo Aalto	M	Finland	FIN	1920	Summer	Antwerper	Swimming	Swimming	Bronze
16	39	Arvo Aalto	M	Finland	FIN	1924	Summer	Paris	Swimming	Swimming	No medal
17	41	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	Bronze
18	42	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	Gold
19	43	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	No medal
20	44	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	Gold
21	45	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	No medal
22	46	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	No medal
23	47	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	No medal
24	48	Paavo Aalt	M	Finland	FIN	1948	Summer	London	Gymnastic	Gymnastic	Gold
25	49	Paavo Aalt	M	Finland	FIN	1952	Summer	Helsinki	Gymnastic	Gymnastic	No medal
26	50	Paavo Aalt	M	Finland	FIN	1952	Summer	Helsinki	Gymnastic	Gymnastic	Bronze

Chapter 4: Objectives

To **preprocess and clean** the Olympics dataset for consistency and accuracy.

1. To perform **Exploratory Data Analysis (EDA)** using Python libraries.
2. To create **static visualizations** using Matplotlib and Seaborn.
3. To build **interactive visualizations** using Plotly for in-depth insights.
4. To design **interactive dashboards** in Power BI and Tableau.
5. To derive **analytical insights** about medal trends, athlete dominance, and country performance.
6. To compare visualization frameworks based on interactivity, usability, and insight clarity.

Chapter 5: Methodology

1. **Data Preprocessing:**
    - Loaded CSV data into Pandas.
    - Filtered *Season = "Summer"*.
    - Cleaned missing values in *Medal* and *NOC*.
    - Standardized country codes and gender labels.
    - Created derived features like *Medal Count* and *Decade*.
2. **Data Visualization Frameworks:**
    - **Python:** Used Matplotlib, Seaborn, and Plotly for EDA and charts.
    - **Power BI:** Imported dataset, created calculated fields (e.g., Total Medals, Gold Ratio), and developed slicers for Year, Country, and Sport.
    - **Tableau:** Built worksheets for Medal trends, Top Countries, Gender Distribution, and Athlete Rankings.
3. **Comparative Analysis:**
    - Evaluated dashboards for interactivity, flexibility, and presentation clarity.

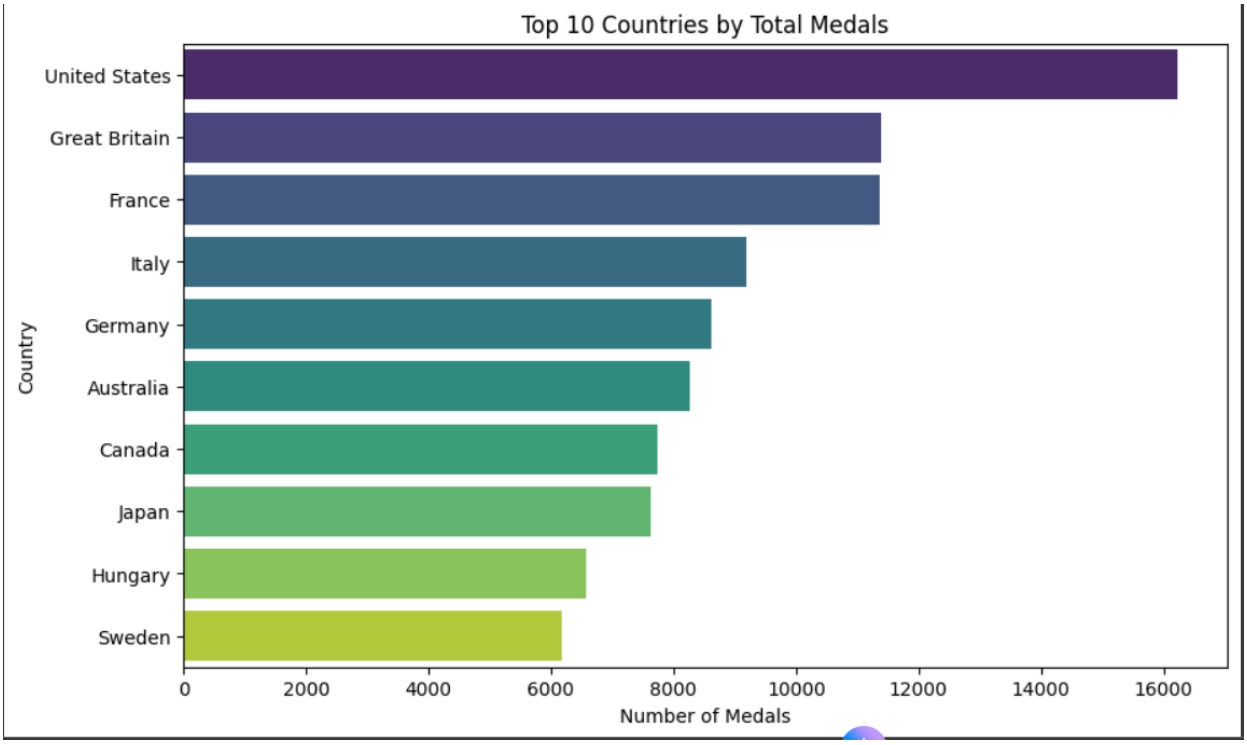
Chapter 6: Python Implementation (Matplotlib & Streamlit)

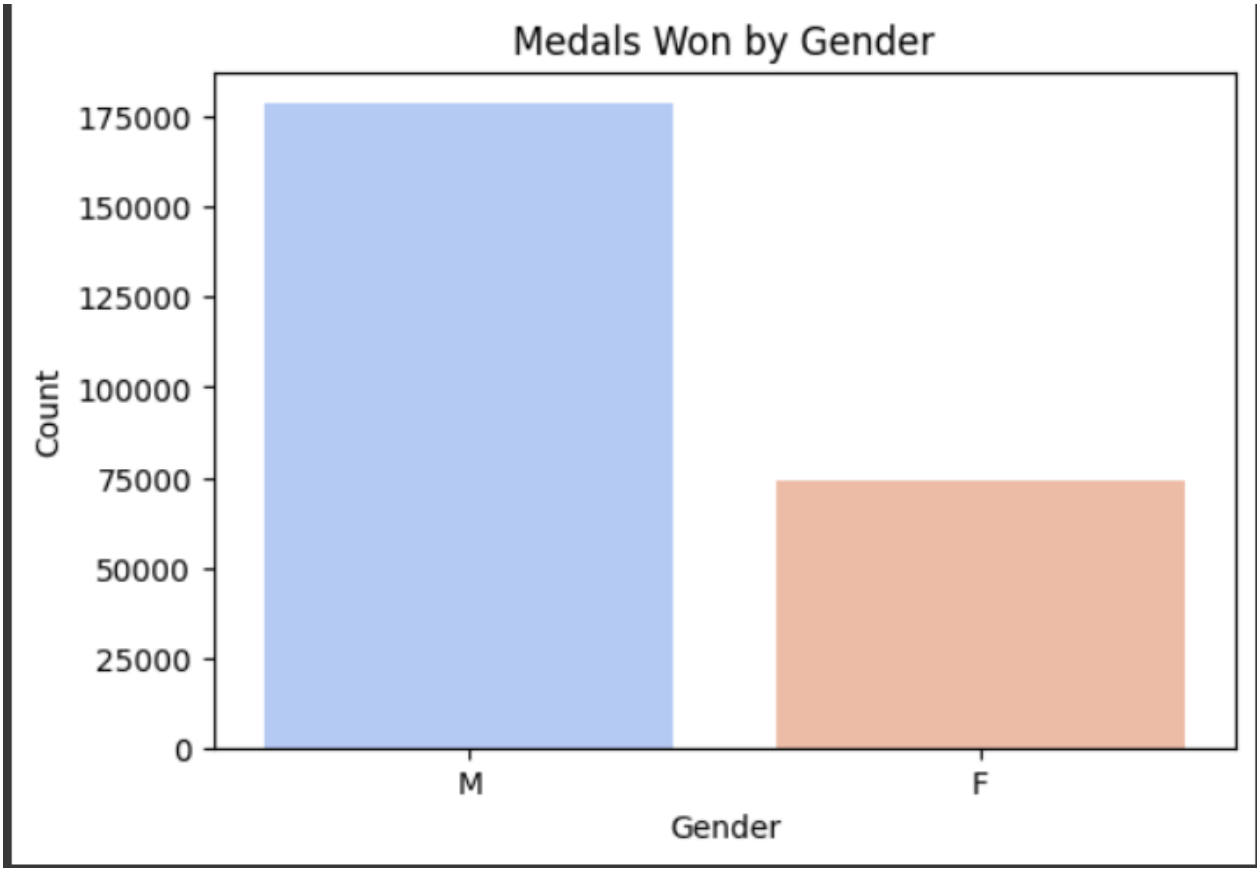
Tools Used:

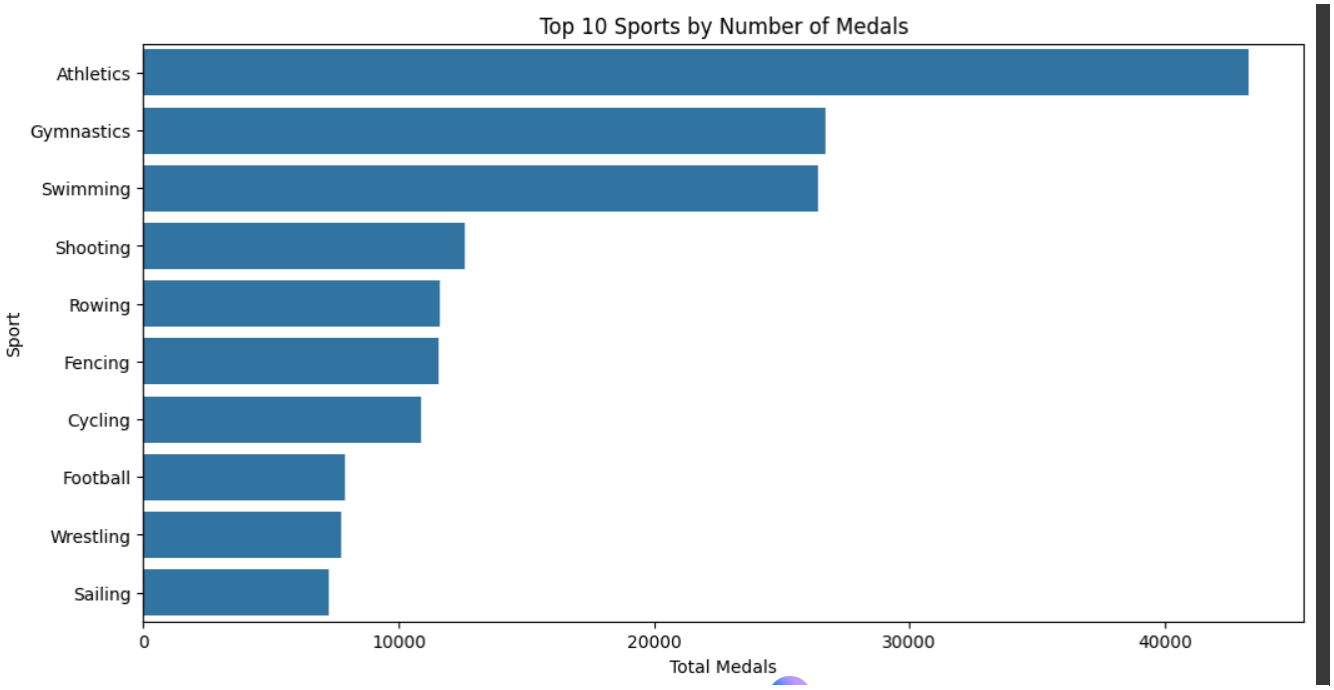
- Pandas for data cleaning
- Matplotlib & Seaborn for static visualization
- Plotly for interactive visualization

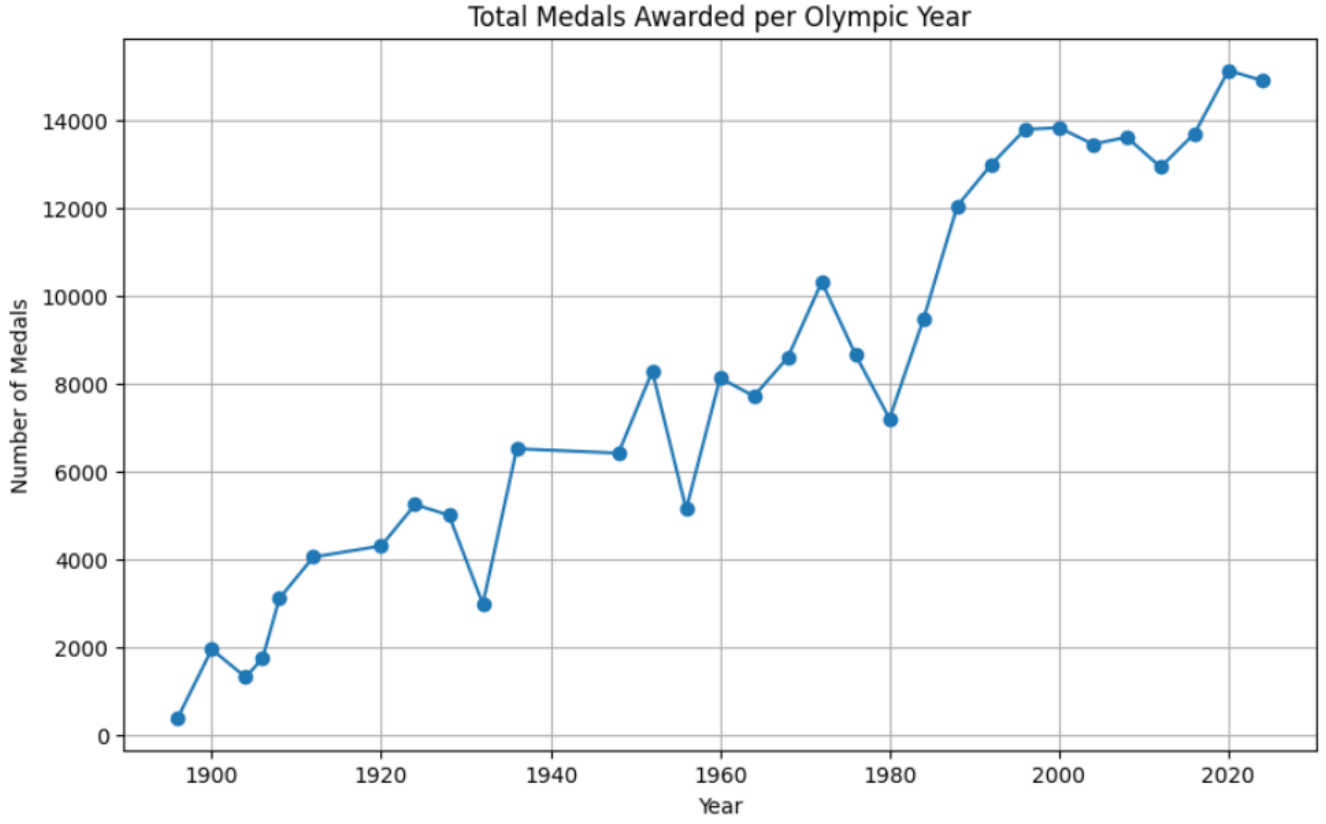
Key Visualizations:

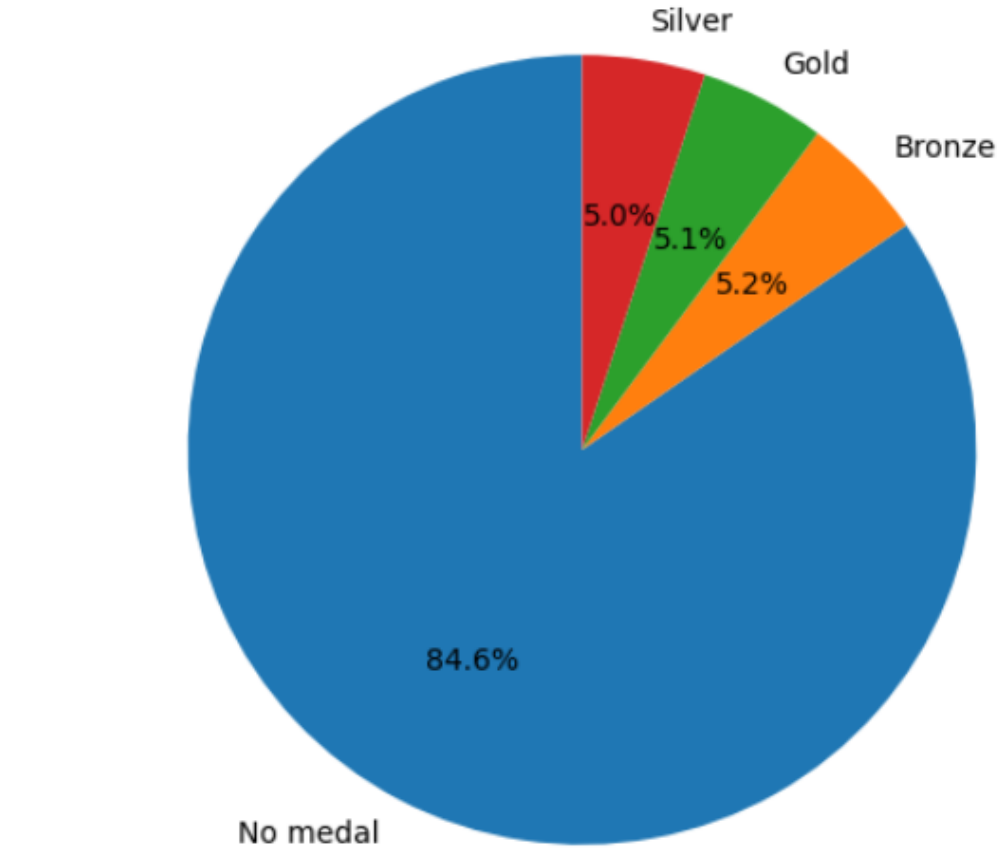
- Medal Distribution by Country (Bar Chart)
- Top 10 Athletes (Horizontal Bar Chart)
- Year-wise Total Medals (Line Chart)
- Gender Participation Growth (Area Chart)
- Sport-wise Medal Trends (Heatmap)

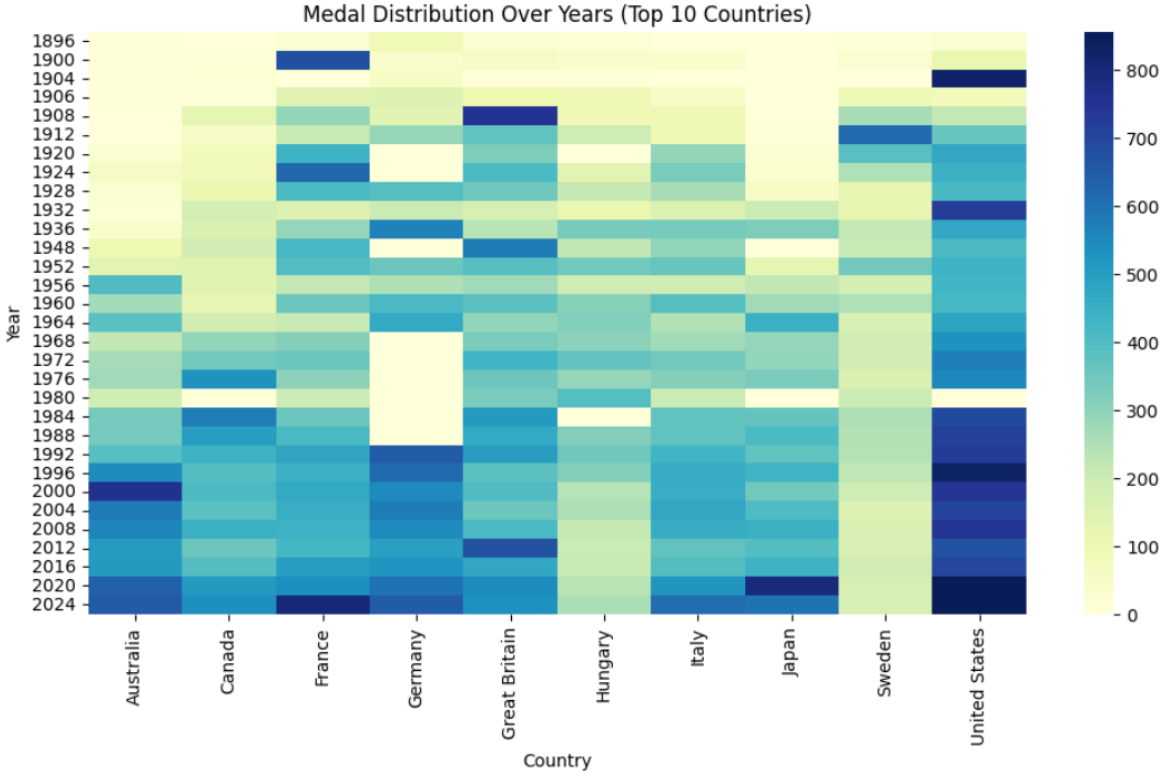

Top 10 Countries by Total Medals
United States
Great Britain
France
Italy
Germany
Australia
Canada
Japan
Hungary
Sweden
Country
0
2000
4000
6000
8000
10000
12000
14000
16000
Number of Medals


Medals Won by Gender
175000
150000
125000
100000
75000
50000
25000
0
Count
M
F
Gender


Top 10 Sports by Number of Medals
Athletics
Gymnastics
Swimming
Shooting
Rowing
Fencing
Cycling
Football
Wrestling
Sailing
Sport
0
10000
20000
30000
40000
Total Medals


Total Medals Awarded per Olympic Year
Number of Medals
0
2000
4000
6000
8000
10000
12000
14000
1900
1920
1940
1960
1980
2000
2020
Year

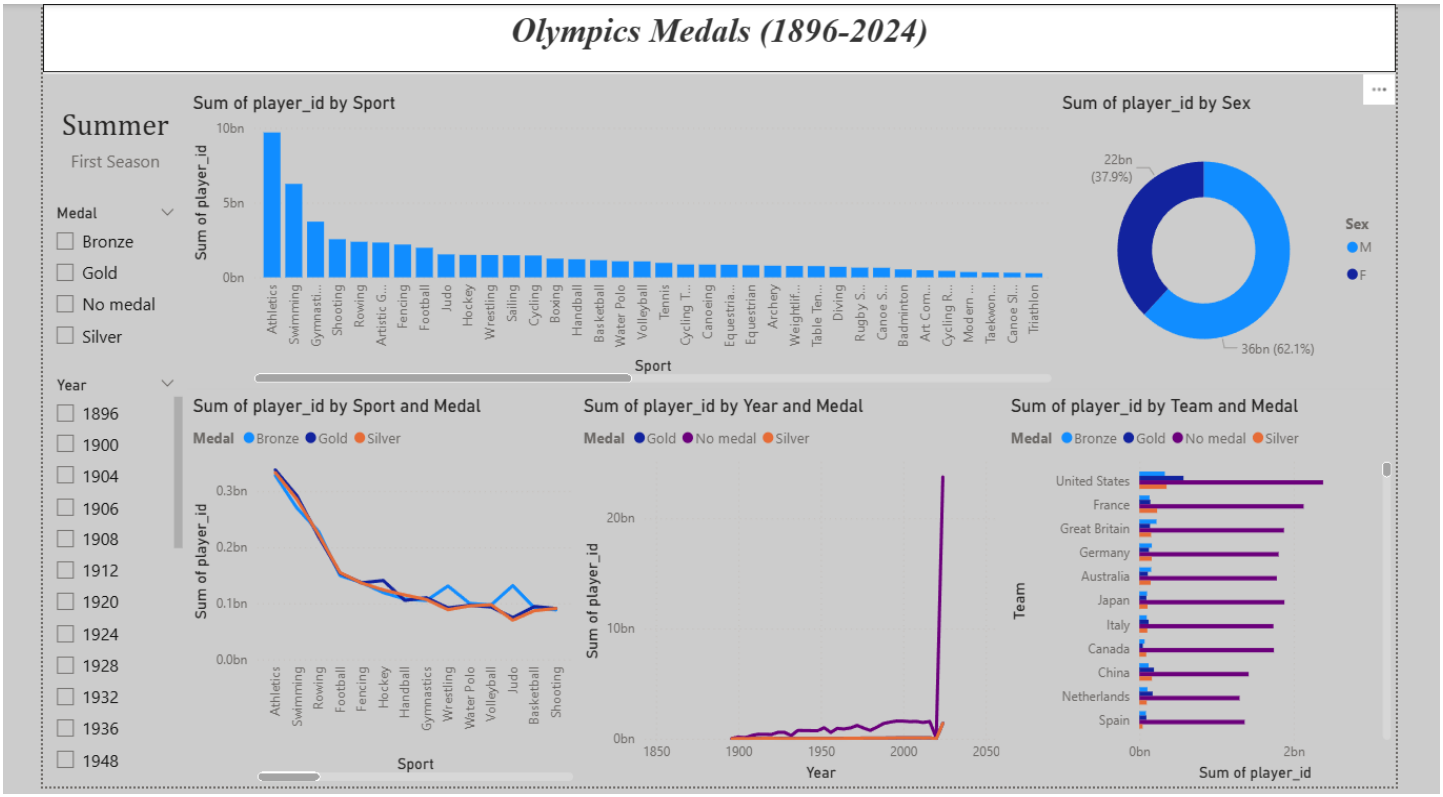

Overall Medal Type Distribution (1994–2024 Summer Olympics)
Silver
Gold
Bronze
5.0%
5.1%
5.2%
84.6%
No medal


Medal Distribution Over Years (Top 10 Countries)
Year
1896
1900
1904
1906
1908
1912
1920
1924
1928
1932
1936
1948
1952
1956
1960
1964
1968
1972
1976
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016
2020
2024
Australia
Canada
France
Germany
Great Britain
Hungary
Italy
Japan
Sweden
United States
Country
800
700
600
500
400
300
200
100
0

Chapter 7: Power BI Dashboard

Features:

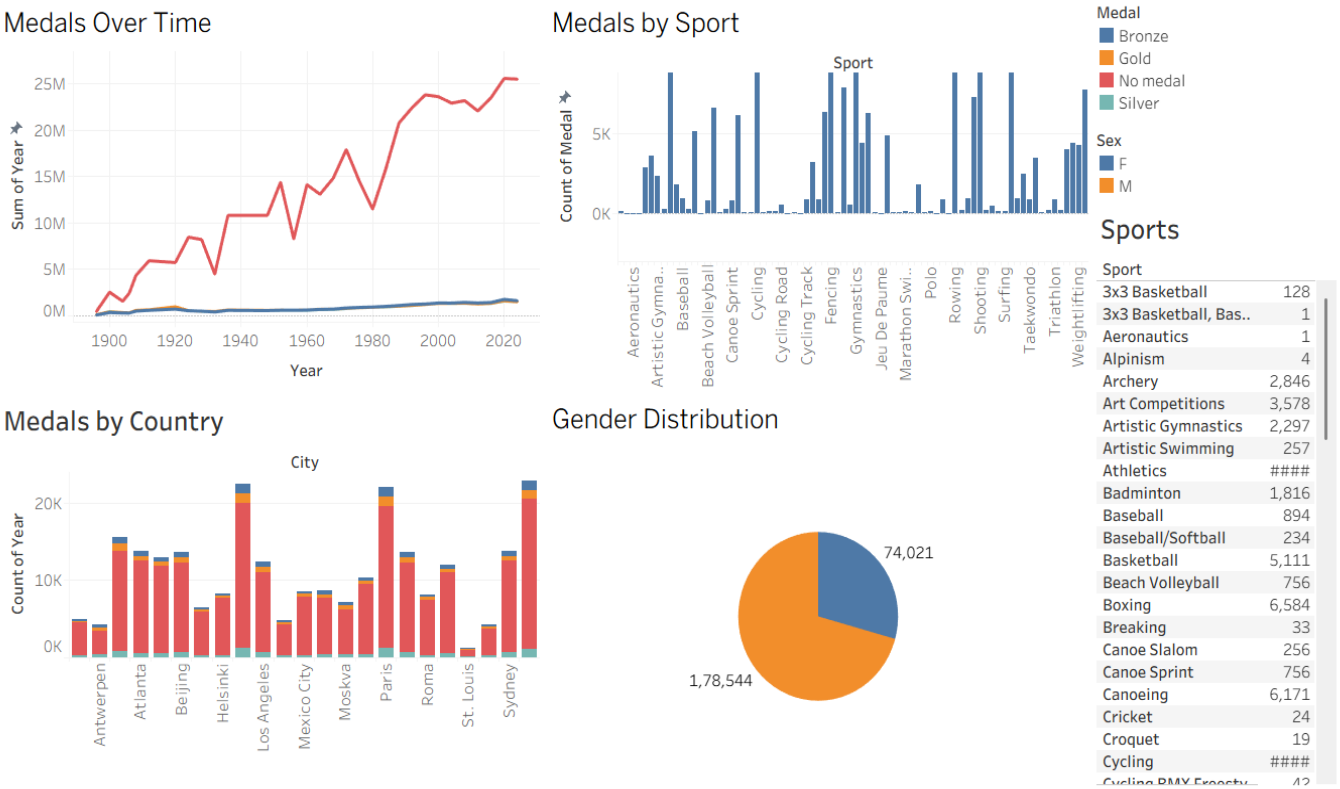
- Filters for *Year*, *Sport*, and *Country*.
- KPI Cards: Total Medals, Gold Medals, Total Athletes.
- Visuals:
  - Medal count by country (Bar Chart)
  - Medal trend by year (Line Chart)
  - Gender participation (Donut Chart)
  - Sport-wise medals (Tree Map)
- Interactive cross-filtering and drill-down to specific games or sports.



Chapter 8: Tableau Dashboard

Features:

- Clean pastel theme with interactive filters.
- Visuals:
  - Country vs Medal Count (Horizontal Bar Chart)
  - Gender Distribution over Time (Line Chart)
  - Top 5 Athletes by Medals (Bar Chart)
  - Map showing Global Medal Spread by Country.
- Hover tooltips and region filters enhance interactivity.



Chapter 9: Analysis & Findings

1. USA and Soviet Union are the top-performing nations across all years.
2. Athletics and Swimming dominate the overall medal count.
3. Female participation has increased steadily since the 1970s.
4. Hosting countries often see a spike in total medals won.
5. Interactive dashboards make exploring Olympic data more engaging and insightful.

Chapter 10: Conclusion

This project successfully demonstrates how data visualization can transform large historical sports data into actionable insights. Using Python, Power BI, and Tableau, the analysis revealed long-term patterns and allowed intuitive exploration of Olympic history.
Each tool had distinct advantages — Python for customization, Power BI for business-style dashboards, and Tableau for visual storytelling.

Chapter 11: Future Scope

1. Integrate machine learning models to predict medal counts for upcoming Olympics.
2. Add real-time updates from sports APIs.
3. Include Winter Olympics for broader analysis.
4. Explore country-level socio-economic factors influencing performance.

Chapter 12: Appendix (Code Snippets)

Key Python code snippets used for data transformation and visualization.

1. Data Loading and Preparation

```
# Import required libraries
import pandas as pd
import numpy as np

# Load the Olympics dataset
df = pd.read_csv("olympics_dataset.csv")

# Display basic info
print(df.head())
print(df.info())

# Filter for Summer Olympics only
df = df[df['Season'] == 'Summer']
```

```
# Drop missing medal values
df.dropna(subset=['Medal'], inplace=True)

# Clean and standardize column names
df.columns = df.columns.str.strip().str.replace(' ', '_')

# Create additional columns for analysis
df['Decade'] = (df['Year'] // 10) * 10
df['Medal_Count'] = 1
```

2. Exploratory Data Analysis

```
# Check missing values
print(df.isnull().sum())

# Summary statistics
print(df.describe())

# Top 10 countries by medal count
top_countries = df['NOC'].value_counts().head(10)
print(top_countries)
```

3. Visualization using Matplotlib & Seaborn

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.style.use('seaborn-v0_8-muted')

# 1. Top 10 Countries by Medal Count
plt.figure(figsize=(10,6))
sns.barplot(x=top_countries.index, y=top_countries.values, palette='coolwarm')
plt.title("Top 10 Countries by Total Medals (1896–2024)")
plt.xlabel("Country Code")
plt.ylabel("Total Medals")
plt.show()

# 2. Medal Distribution by Type
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='Medal', order=['Gold','Silver','Bronze'], palette='Set2')
plt.title("Medal Distribution by Type")
plt.show()

# 3. Gender Participation Trend
gender_trend = df.groupby(['Year','Sex']).size().reset_index(name='Count')
plt.figure(figsize=(10,5))
sns.lineplot(data=gender_trend, x='Year', y='Count', hue='Sex', marker='o')
plt.title("Male vs Female Participation Over Time")
plt.show()
```

4. Interactive Visualization using Plotly

```
import plotly.express as px

# 1. Interactive Medal Trend Over Years
trend = df.groupby('Year')['Medal_Count'].sum().reset_index()
fig = px.line(trend, x='Year', y='Medal_Count', title='Total Medals Awarded Over the Years')
fig.show()

# 2. Country Medal Count
country_medals = df['NOC'].value_counts().reset_index()
country_medals.columns = ['Country', 'Total Medals']
fig = px.bar(country_medals.head(10), x='Country', y='Total Medals',
        title='Top 10 Countries by Total Medals',
        color='Total Medals', color_continuous_scale='Viridis')
fig.show()

# 3. Gender Distribution Pie Chart
gender_count = df['Sex'].value_counts().reset_index()
gender_count.columns = ['Gender', 'Count']
fig = px.pie(gender_count, names='Gender', values='Count', title='Gender Distribution in Olympics')
fig.show()

# 4. Medal Distribution by Sport
sport_medals = df['Sport'].value_counts().reset_index().head(15)
sport_medals.columns = ['Sport', 'Total Medals']
fig = px.bar(sport_medals, x='Sport', y='Total Medals',
        title='Top 15 Sports by Medal Count', color='Total Medals')
fig.update_layout(xaxis_tickangle=-45)
fig.show()
```

5. Saving Cleaned Dataset

```
# Save the cleaned dataset for Power BI and Tableau
df.to_csv("olympics_cleaned.csv", index=False)
print("Cleaned dataset saved successfully!")
```
