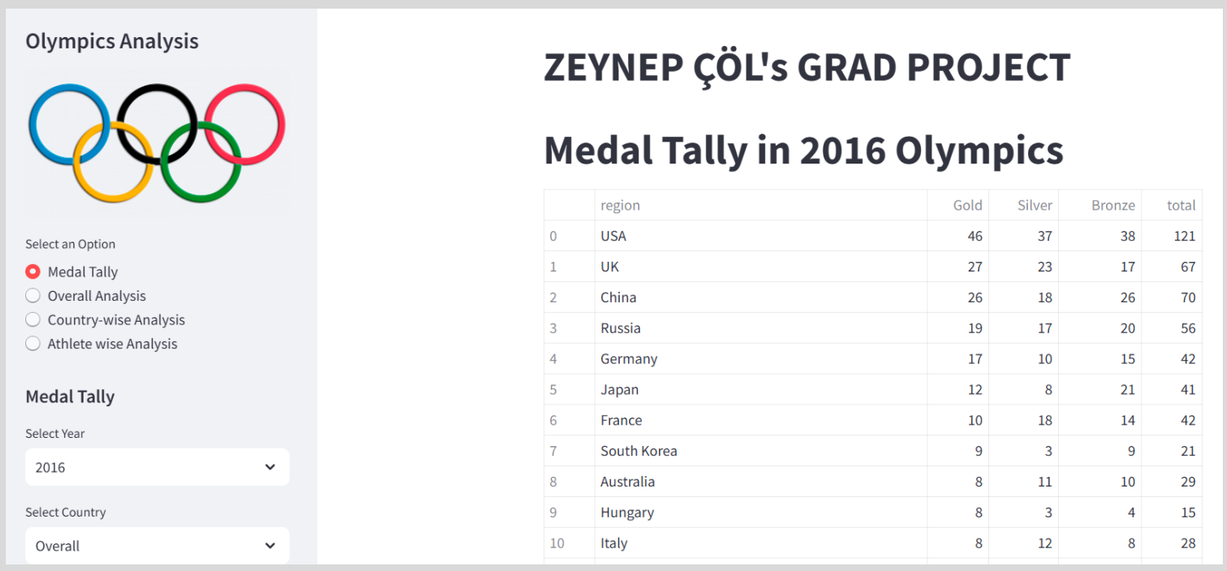
Olympics: Data Analysis

Zeynep ÇÖL
Department of Computer Engineering
Adana, Turkey
210101094@ogr.atu.edu.tr

Introduction

"Olympics: Data Analysis" utilizes a comprehensive dataset from the 1896 to 2016 Olympic Games to provide an interactive analysis of Olympic performance. The dashboard offers insights into medal tallies, athlete achievements, country performances, and the evolution of sports. Users can explore trends in medals, sports growth, and athlete participation. It also allows for detailed analysis of country-specific performances and individual athlete data, including age, height, weight, and gender trends. This data-driven approach offers a deeper understanding of the Olympics' evolution and its ongoing impact on the world of sports.



	region	Gold	Silver	Bronze	total
0	USA	46	37	38	121
1	UK	27	23	17	67
2	China	26	18	26	70
3	Russia	19	17	20	56
4	Germany	17	10	15	42
5	Japan	12	8	21	41
6	France	10	18	14	42
7	South Korea	9	3	9	21
8	Australia	8	11	10	29
9	Hungary	8	3	4	15
10	Italy	8	12	8	28

Figure 1. Olympics Data Analysis Dashboard Home Screen

Evolution of the Olympics: A Data Perspective

Olympics have evolved significantly over the past century, becoming the pinnacle of global sports competition. This project provides an in-depth data-driven analysis of the modern Olympics, covering 28 editions across 23 host cities, with over 52 sports and 651 events. With the participation of 206 nations and more than 116.000 athletes, highlighting the increasing scale and diversity of the Games. Using historical data from Athens 1896 to Rio 2016, this dashboard enables users to explore key trends such as medal distributions, athlete demographics, and nation-specific performances.

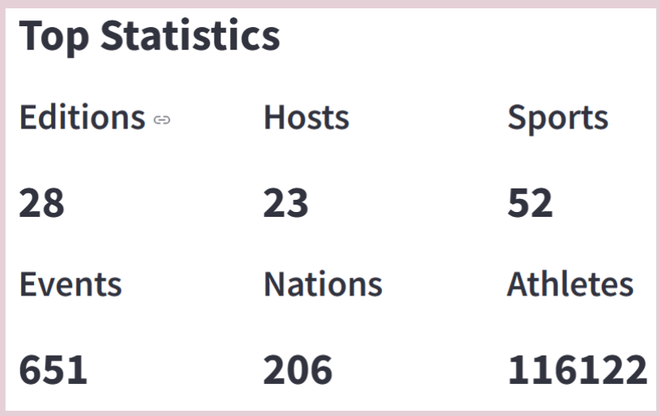


Figure 2. Olympic Datasets Top Statistics

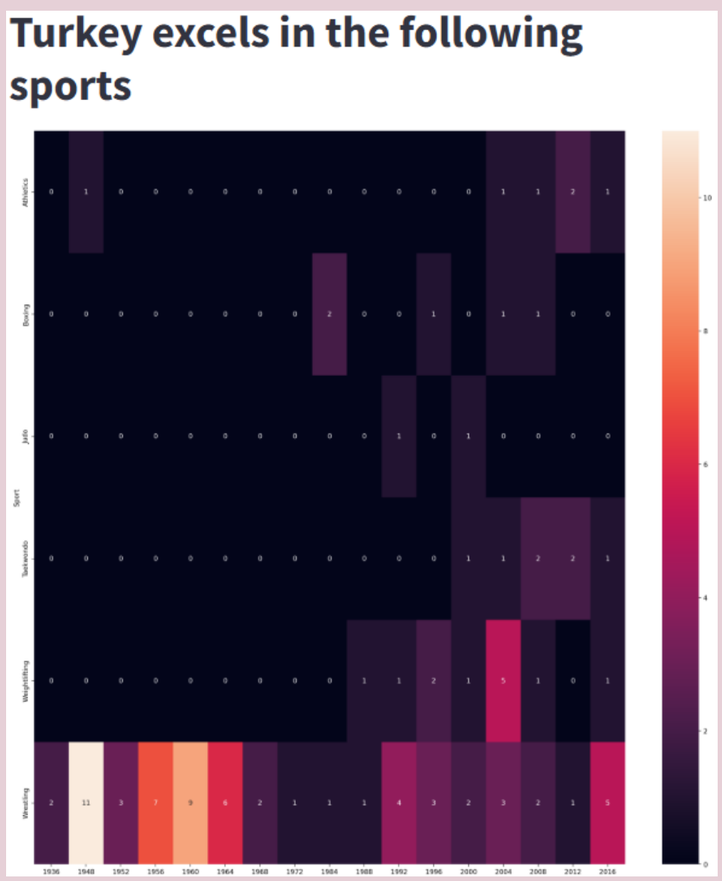


Figure 3. The country excels in the following sports

Features

This project offers a comprehensive and interactive analysis of Olympic history, allowing users to explore key aspects of the Olympics through dynamic visualizations and data-driven insights.

The Medal Tally section provides insights into the distribution of gold, silver, and bronze medals, allowing users to filter results by year and country. The Country-wise Analysis explores a nation's performance over the years, identifying the sports in which it has excelled. The Athlete-wise Analysis offers a deeper look into athlete demographics, including age distribution, height-to-weight comparisons across different sports, and trends in male and female participation over time. The Overall Analysis further visualizes the historical progression of the Olympics, illustrating the increasing number of participating nations, events, and athletes. Each dataset-driven visualization provides a unique perspective, transforming raw data into meaningful insights that showcase the evolution of the Olympics.

By integrating interactive selections and dynamic visualizations, this project enables users to uncover patterns and trends in Olympic history, making data exploration both intuitive and insightful.

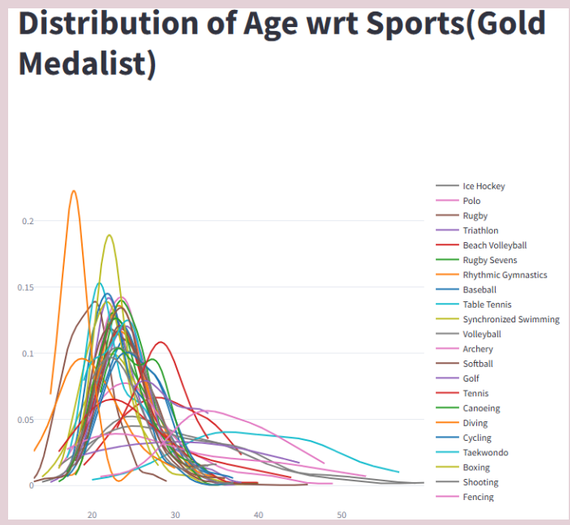


Figure 4. Distribution of Age wrt Sports(Gold Medalist)

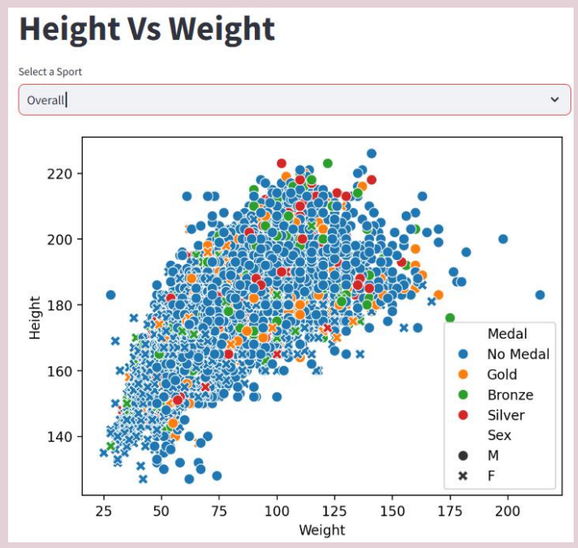


Figure 5. Height vs. weight of athletes by sport, medal, and gender.

Tools and Technologies

The project uses Python for data cleaning and manipulation, with Pandas, NumPy, and Matplotlib for analysis and visualization. Streamlit is integrated to build an interactive interface, while Plotly and Seaborn provide advanced visualizations. Scipy is used for scientific computations. The datasets is sourced from Kaggle. These tools work together to deliver an efficient solution for processing and visualizing data interactively.

Conclusion

This project provided practical experience with large datasets, emphasizing data analysis and visualization. It enhanced my ability to extract actionable insights and develop efficient solutions for interactive data exploration. Throughout, I refined skills in data visualization, insight extraction, and code optimization, combining analytical rigor with creative problem-solving. This experience has deepened my understanding of data-driven challenges, equipping me for future work in data-centric fields.

References

D. Yamunathangam, G. Kirthicka and S. Parveen, "Performance analysis in olympic games using exploratory data analysis techniques", *International Journal of Recent Technology and Engineering.*, vol. 7, pp. 251-253, 2019.

www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results