



REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY
UNIVERSITY

FACULTY OF FACULTY OF COMPUTER AND INFORMATICS
DEPARTMENT OF COMPUTER ENGINEERING

OLYMPICS: DATA ANALYSIS

ZEYNEP ÇÖL
BACHELOR DEGREE

SUPERVISOR
ASSOC. PROF. DR. ALPER KAMİL DEMİR

ADANA 2025

ABSTRACT

OLYMPICS: DATA ANALYSIS

Zeynep ÇÖL

Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Alper Kamil DEMİR

January 2025, 16 pages

Olympics are a global sporting event where athletes from over 200 nations compete in a wide range of disciplines. Despite the size of a country's population, not all nations achieve a high number of Olympic medals.

The thesis aims to analyze the Olympic dataset to compare the overall performance of different countries and evaluate their contributions to the Olympic Games over time. Using Python and various data analysis libraries, this study provides a comprehensive exploration of Olympic performance trends and identifies patterns in medal counts across different nations.

The thesis utilizes advanced data visualization techniques to present insights into the performance of countries, allowing for a deeper understanding of Olympic outcomes. Through these analyses, athletes and coaches can gain valuable insights into individual and collective performances, while countries with lower medal counts can identify areas for improvement in nurturing talent. The findings from this research serve as a tool to enhance the strategic development of sports programs and optimize future Olympic performances.

Keywords: data analysis, data visualization, data science, machine learning, artificial intelligence, exploratory data analysis

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who supported and contributed to the successful completion of this thesis.

First and foremost, I would like to thank my supervisor, Assoc. Prof. Dr. Alper Kamil DEMİR, for their invaluable guidance, support, and encouragement throughout this research. They not only provided expert advice but also allowed me the freedom to explore my interests and pursue a path aligned with my passion and academic goals. Their unwavering belief in my potential and their encouragement to carve my own way has been truly inspiring. I sincerely appreciate my university, Adana Alparslan Türkeş Science and Technology University for granting access to its extensive library and research facilities, which played a crucial role in the successful completion of this project. The availability of these resources greatly supported my analysis and contributed to the depth and quality of my research.

Additionally, I would like to extend my appreciation to the developers and researchers who have contributed to the open-source libraries used in this project. Their work in creating and maintaining these tools was essential in enabling the efficient handling, analysis, and visualization of the Olympic dataset.

Finally, I would like to thank my family and friends for their unwavering support, patience, and understanding throughout my academic journey.

TABLE OF CONTENTS

TABLE OF CONTENTS	iii
LIST OF FIGURES	iv
1. INTRODUCTION	1
2. LITERATURE REVIEW	2
3. MATERIALS AND METHODS	3
4. RESULTS AND DISCUSSIONS	4
4.1. Country-wise Analysis	8
4.2. Athlete-wise Analysis	9
4.3. Sports Growth and Evolution	13
4.4. Insights from the Data and Future Implications	14
4.5. Limitations and Areas for Future Research	14
5. CONCLUSION	15
REFERENCES	16

LIST OF FIGURES

Figure 1 Olympic Datasets Top Statistics	1
Figure 2 Olympics Data Analysis Dashboard Home Screen	2
Figure 3 Country Performance in Selected Year	2
Figure 4 Participating Nations over the years	4
Figure 5 Olympic Events over the years	5
Figure 6 Athletes over the years.....	6
Figure 7 Number of Events over time	7
Figure 8 Medal Tally over the years for each country	8
Figure 9 Turkey excels in the following sports.....	9
Figure 10 Distribution of Age	10
Figure 11 Distribution of Age wrt Sports(Gold Medalist)	11
Figure 12 Height Vs Weight	12
Figure 13 Men Vs Women Participation Over the Years	13

1. INTRODUCTION

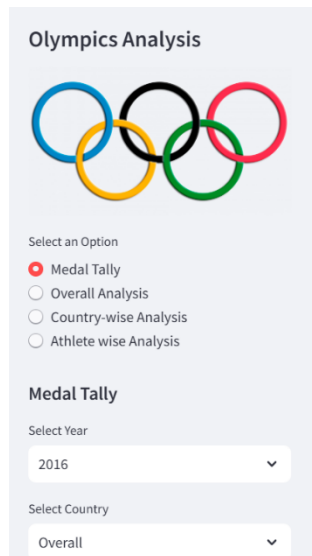
“Olympics: Data Analysis” utilizes a comprehensive dataset spanning from the inaugural 1896 Olympic Games to the 2016 Games, offering a deep dive into the historical and evolving performance of athletes, countries, and sports. This project aims to provide an interactive and insightful platform for users to explore the multifaceted nature of the Olympics, focusing on a range of key metrics such as medal tallies, athlete achievements, and country-specific performances.

Top Statistics

Editions 	Hosts	Sports
28	23	52
Events	Nations	Athletes
651	206	116122

Figure 1 Olympic Datasets Top Statistics

Through a user-friendly dashboard, users can investigate trends in medal distribution across various countries and sports disciplines, uncovering patterns in performance over time. The platform also allows for a closer examination of the growth and diversification of sports, with a focus on the increasing number of events and countries participating in the Games. Moreover, the dashboard facilitates detailed analysis of country-specific performance, shedding light on which nations have dominated in particular events, as well as the impact of socio-political changes on Olympic results.



ZEYNEP ÇÖL's GRAD PROJECT

Medal Tally in 2016 Olympics

	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 2 Olympics Data Analysis Dashboard Home Screen

UK performance in 2012 Olympics

	region	Gold	Silver	Bronze	total
0	UK	29	17	19	65

Figure 3 Country Performance in Selected Year

In addition to country-level analysis, the platform enables a more granular exploration of individual athlete data. By examining factors such as age, height, weight, and gender, users can uncover trends in athlete demographics and performance characteristics. This information provides valuable insights into how the Olympic Games have evolved in terms of inclusivity and the shifting nature of athletic excellence.

By combining historical data with modern analytical tools, this data-driven project offers a nuanced understanding of the Olympics' development, highlighting the Games' enduring influence on global sports culture and its role in promoting international unity. The insights generated from this analysis not only serve as a resource for sports enthusiasts and researchers but also open new avenues for further exploration into how the Games can continue to evolve and inspire future generations of athletes.

2. LITERATURE REVIEW

Various studies have been conducted on sports analytics, particularly in the context of the Olympic Games. Previous research has focused on the dominance of certain countries in specific sports, the impact of socio-economic factors on Olympic success, and the evolution of performance trends over time.

This project builds upon these studies by offering a comprehensive analysis tool that integrates various aspects of Olympic data, including medal tallies, country-wise analysis, and athlete-specific achievements. The use of interactive visualizations distinguishes this project, allowing for a more engaging and accessible exploration of the data. The type of analysis which is quite popular and suitable while analyzing the evolution of the Olympics is Exploratory Data Analysis. In Exploratory Data Analysis, we examine large data and elucidate its various characteristics basically in the visual format (Graphs, Charts, Tables and many more). EDA is an approach that provides a deeper understanding of the dataset.

3. MATERIALS AND METHODS

The project utilizes historical Olympic data, which is processed and analyzed to generate insights into the country and athlete performance across various sports. The data is visualized through interactive plots and tables, offering users the ability to explore different analysis options such as overall medal tallies, country-specific performance, and athlete-wise analysis.

The project involves data analysis for the Olympics, utilizing the `app.py` file as the primary script. The dataset includes three CSV files. The first, `noc_regions.csv`, contains information about country names and their respective codes. The second dataset, `athlete_events.csv`, comprises 271,117 rows of data, providing extensive details for analysis. These resources are integrated using Python scripts, including `helper.py` and `preprocessor.py`, to facilitate data processing and visualization, offering insights into key aspects of the Olympics. This project highlights the use of Python for handling large datasets and extracting meaningful trends and patterns in sports analytics.

The platform is developed using Python along with key libraries such as Pandas, Matplotlib, and Plotly for efficient data manipulation and visualization. Streamlit is employed to build an interactive and user-friendly interface, ensuring a seamless user experience when exploring the data. The interface allows users to easily navigate through the various data points and visualizations, making the analysis both accessible and engaging.

This project adopts the Exploratory Data Analysis (EDA) technique, a method used to thoroughly analyze data and present its key features, primarily in visual form. EDA is employed not only to apply various algorithms but also to understand the structure and content of the dataset. By utilizing EDA, we can visualize the data through various types of graphs and plots, which aid in uncovering trends and patterns. The visual format enables a deeper understanding of the analysis and facilitates a comparative study between different datasets and metrics.

4. RESULTS AND DISCUSSIONS

The results derived from the analysis of the Olympic dataset are presented and discussed herein. The insights obtained from this analysis offer a thorough understanding of the performance of various countries and athletes across the Olympic Games, from Athens 1896 to Rio 2016.

Participating Nations over the years

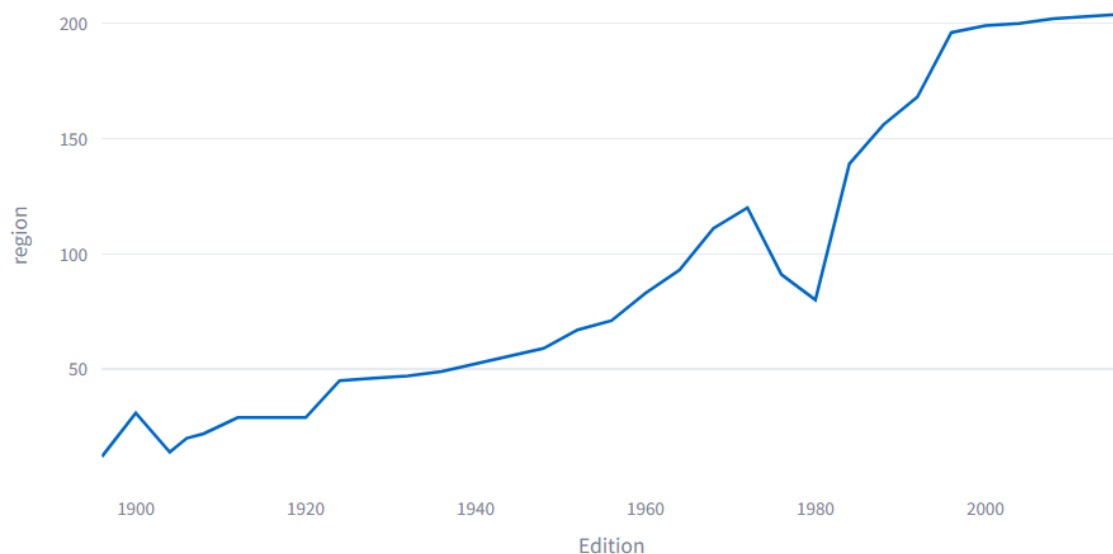


Figure 4 Participating Nations over the years

Events over the years

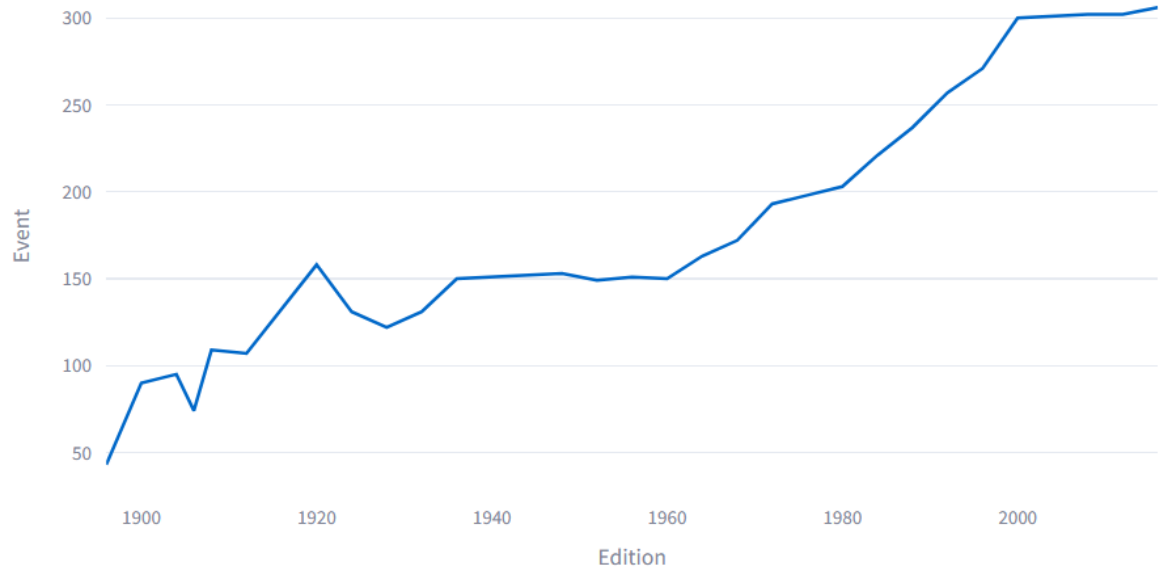


Figure 5 Olympic Events over the years

Athletes over the years

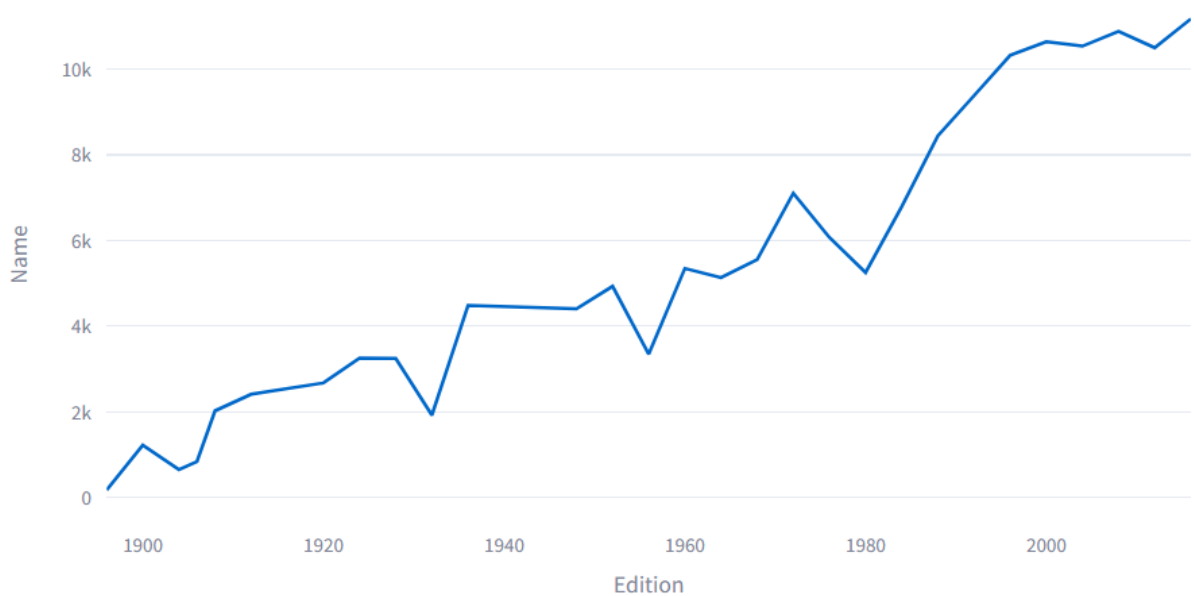


Figure 6 Athletes over the years

No. of Events over time(Every Sport)

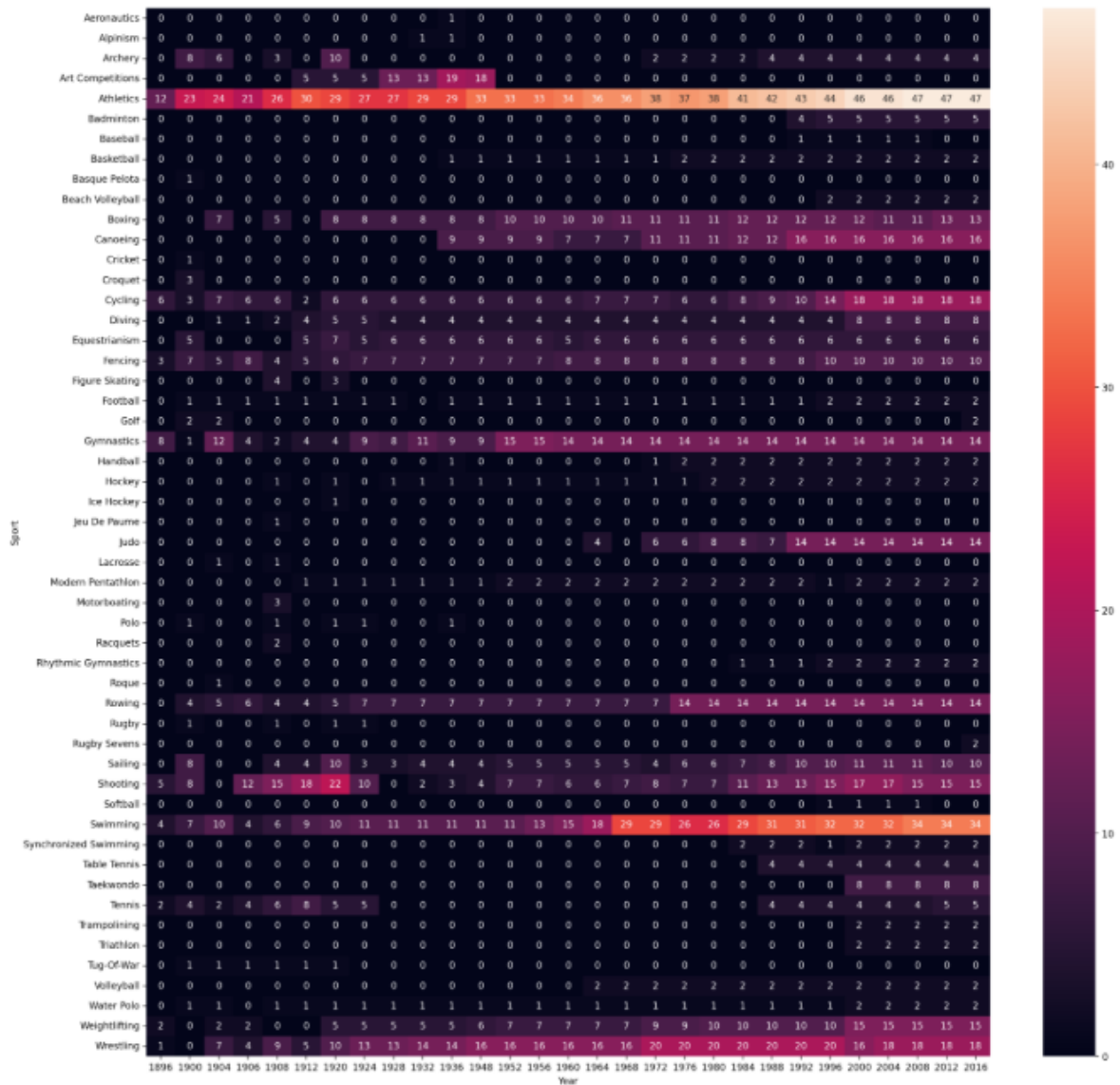


Figure 7 Number of Events over time

4.1. Country-wise Analysis

The performance of each country in the Olympics was evaluated based on the total medal count, including gold, silver, and bronze medals. The data revealed that countries like the United States, France, and Germany have consistently performed well across the Olympics. However, certain countries with large populations, such as China, have shown significant improvement in their medal tallies in recent decades. This observation reflects not only the growth in the sporting infrastructure but also strategic investments in sports development programs.

Through the interactive dashboard, users can explore the overall performance of countries over the years and compare them based on total medal counts. This approach highlights the dynamic nature of Olympic success, influenced by a variety of factors including political, economic, and social changes, as well as the increasing participation of countries in diverse sporting events.

Turkey Medal Tally over the years

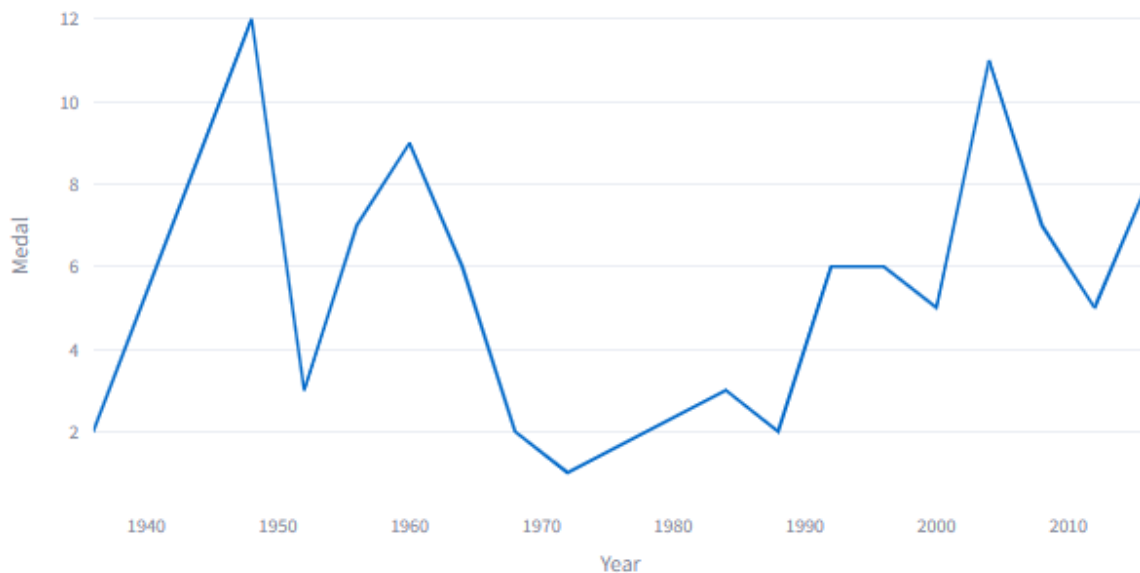


Figure 8 Medal Tally over the years for each country

Turkey excels in the following sports

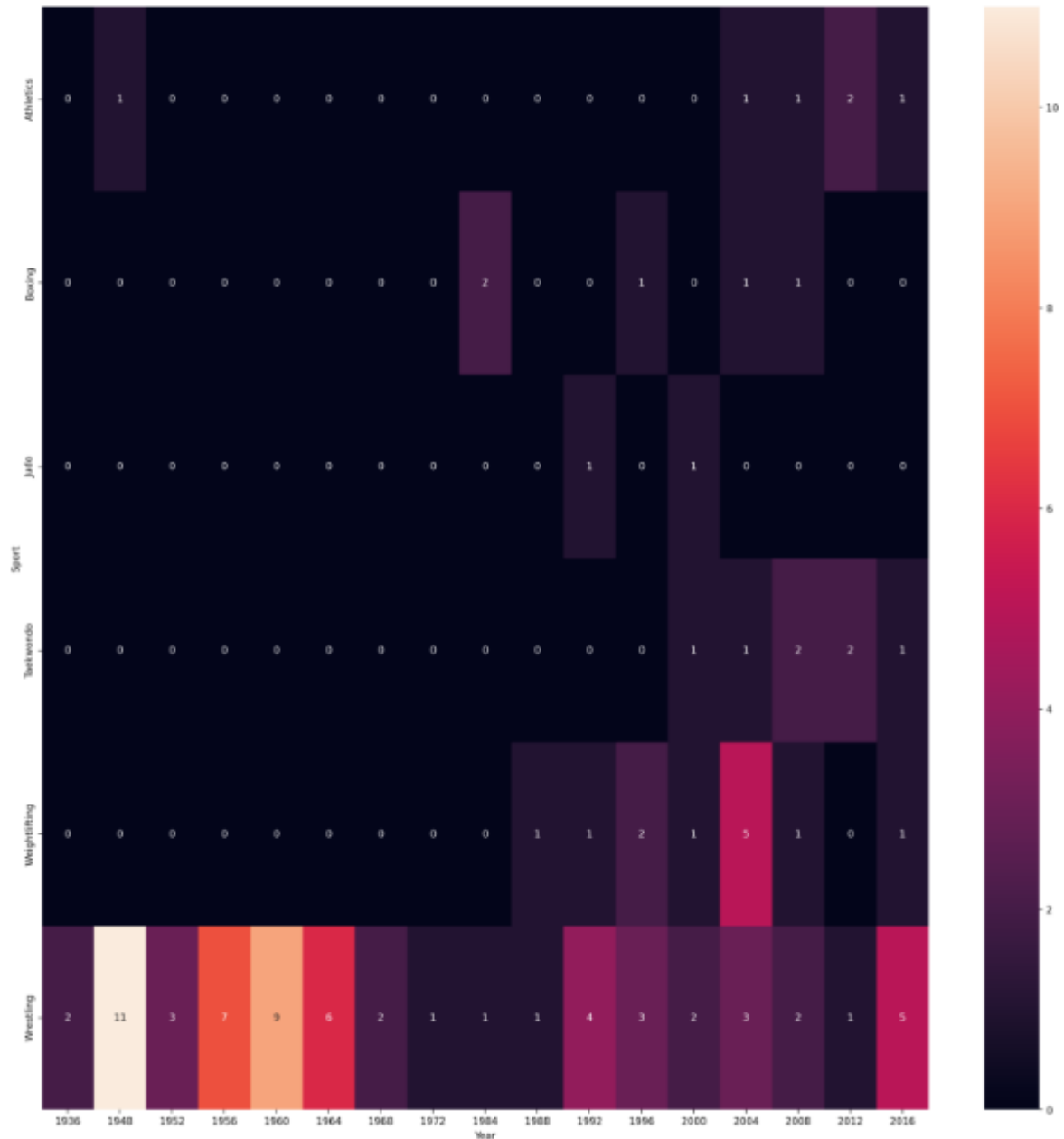


Figure 9 Turkey excels in the following sports

4.2. Athlete-wise Analysis

The athlete-wise analysis provides insights into various attributes such as age, height, weight, and gender distribution among Olympic participants. The age distribution analysis highlights the typical age range of competitors, while a focused analysis on gold medalists across different

sports reveals variations in peak performance ages. Additionally, the relationship between height and weight is examined to identify trends among athletes in different disciplines. The participation of male and female athletes over the years is also analyzed, showcasing the evolution of gender representation in the Olympics. These analyses collectively offer a deeper understanding of athlete demographics and performance trends across different Olympic events.

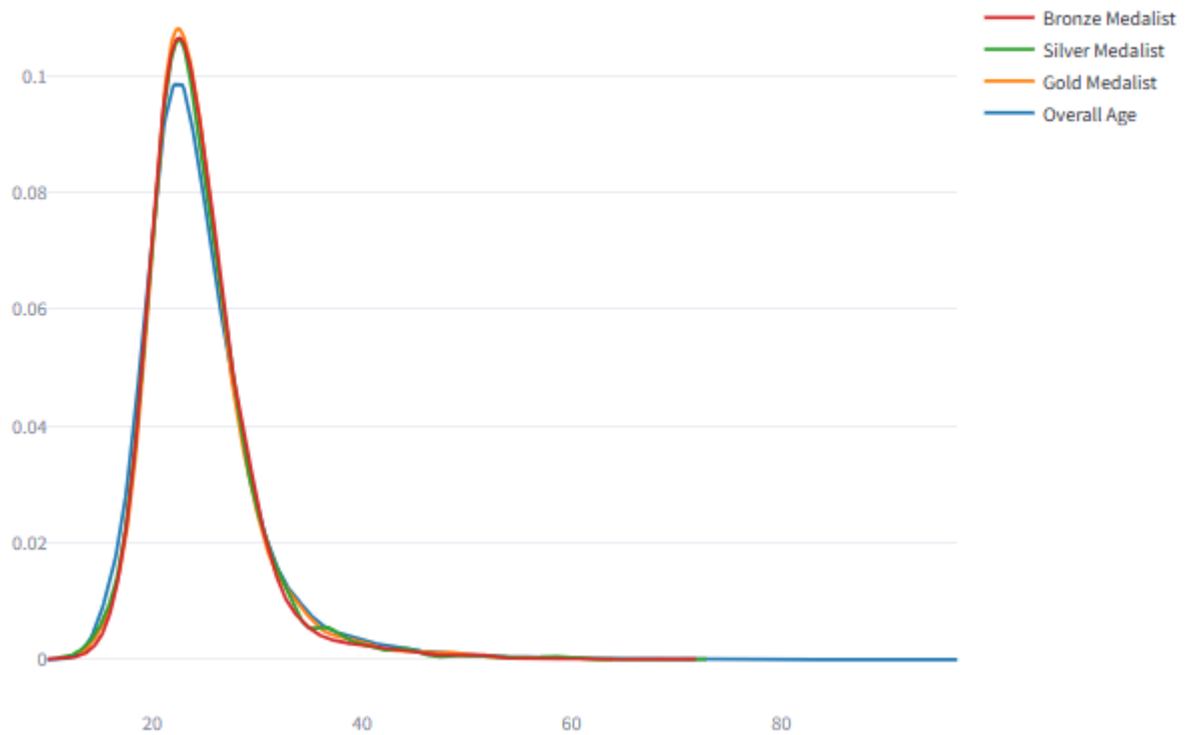


Figure 10 Distribution of Age

Distribution of Age wrt Sports(Gold Medalist)

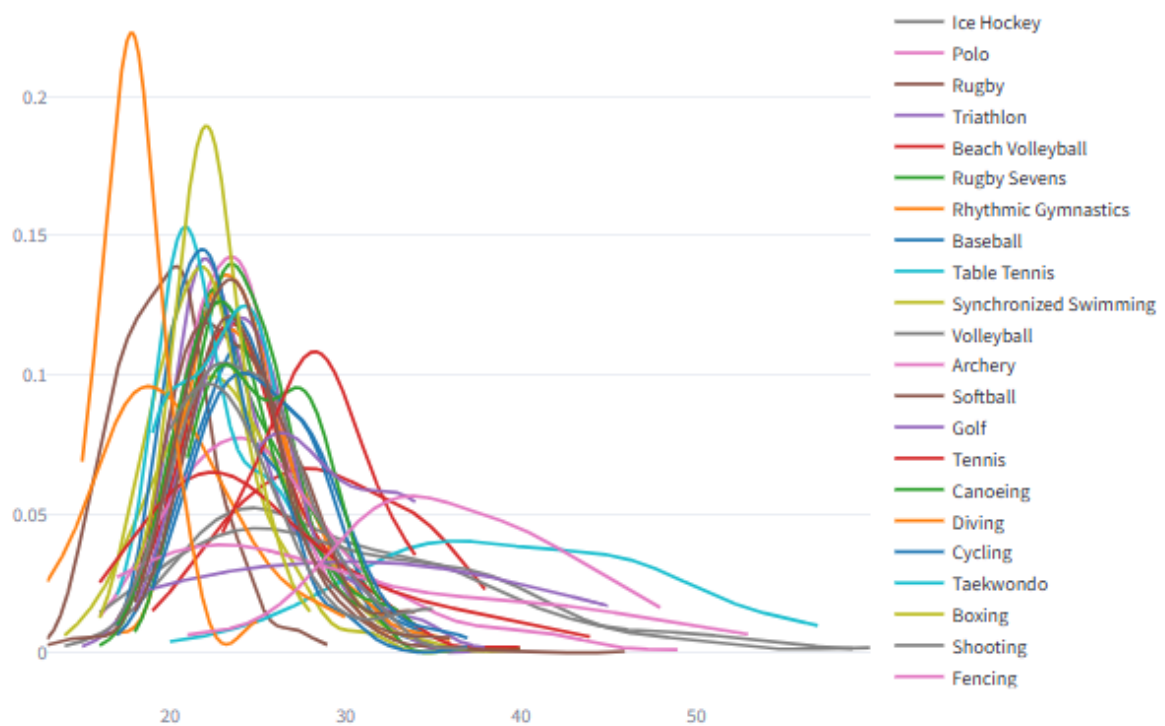


Figure 11 Distribution of Age wrt Sports(Gold Medalist)

Height Vs Weight

Select a Sport

Overall

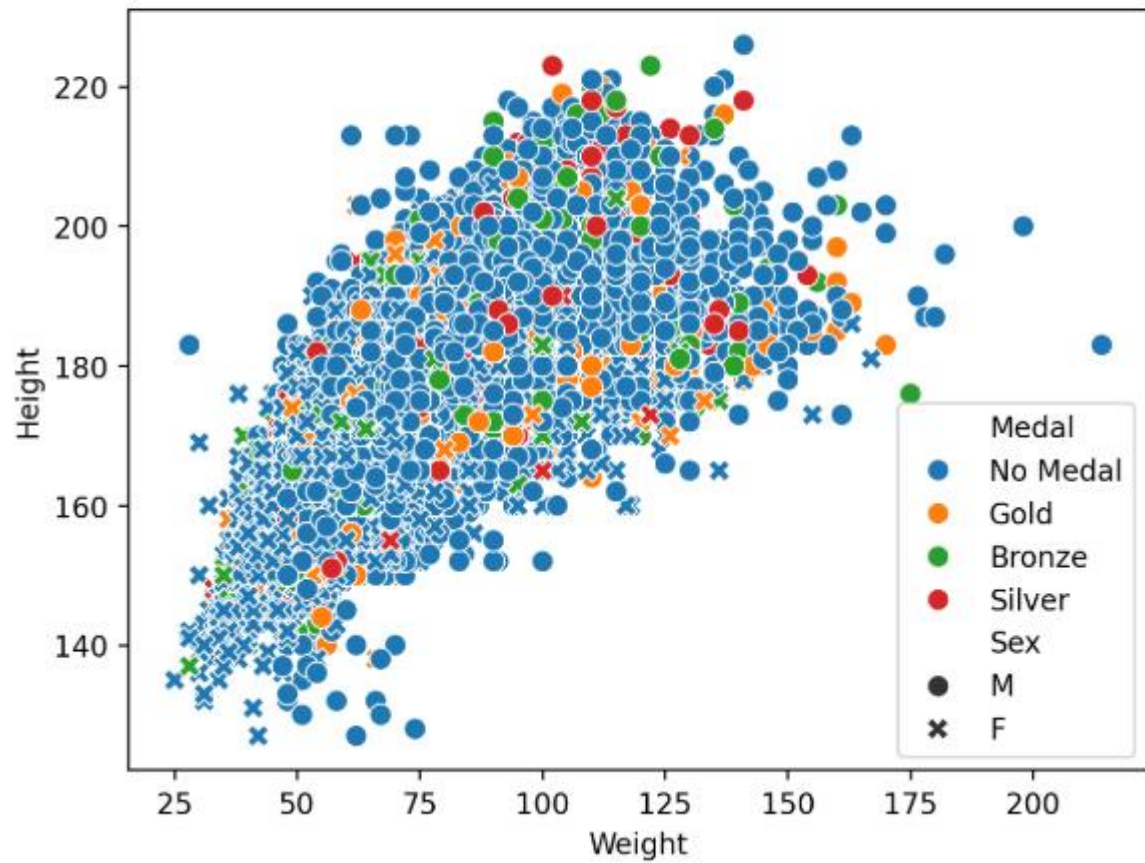


Figure 12 Height Vs Weight

Men Vs Women Participation Over the Years

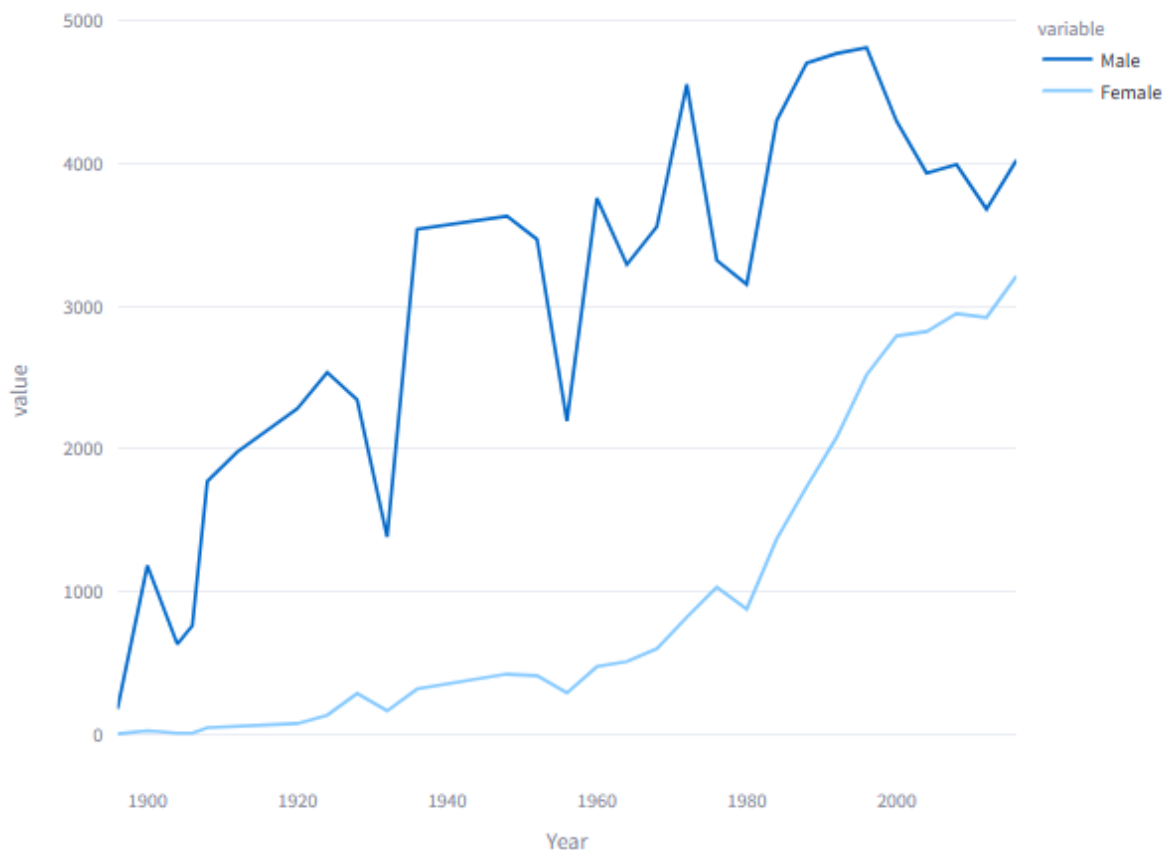


Figure 13 Men Vs Women Participation Over the Years

4.3. Sports Growth and Evolution

One of the key insights from the analysis is the evolution of sports over time. The dataset demonstrates how the Olympics have expanded, both in terms of the number of events and the variety of sports included. Initially, the Olympics featured fewer events, but over time, the inclusion of new sports and events has made the Olympics more diverse and inclusive. For example, sports like women's boxing and skateboarding were introduced in recent years, reflecting broader societal changes and the growing demand for diverse athletic representation.

The interactive visualizations allow users to explore the development of specific sports, providing insights into their growth in popularity and the success of athletes in those disciplines.

This information is crucial for understanding the shifting dynamics of the Olympics and how they adapt to changing global interests.

4.4. Insights from the Data and Future Implications

The results obtained from this project offer valuable insights into the factors that contribute to a country's success in the Olympics. They suggest that success is not solely dependent on population size or economic strength but also on strategic investments in sports programs, training, and athlete development.

Moreover, the findings emphasize the importance of understanding athlete demographics and how they influence performance across different sports.

4.5. Limitations and Areas for Future Research

While this project provides valuable insights, it is important to note the limitations of the dataset and the scope of the analysis. The dataset covers Olympic Games from 1896 to 2016 and does not account for the most recent performances.

Future research could extend this analysis by including data from more recent Games, allowing for a more comprehensive view of ongoing trends in Olympic performance.

5. CONCLUSION

This project provided hands-on experience in working with large-scale datasets, focusing on data preprocessing, analysis, and visualization. By leveraging various analytical techniques, it enabled the extraction of meaningful insights from complex Olympic data, allowing for a deeper understanding of country-wise and athlete-wise performances across different time periods. The implementation of interactive visualizations facilitated dynamic exploration of trends, making data interpretation more accessible and insightful.

Through this project, I refined my ability to process and analyze structured data efficiently while optimizing the visualization of key performance metrics. The integration of an interactive dashboard enhanced the usability of the analysis, enabling users to explore historical Olympic trends seamlessly. Additionally, working with diverse data attributes—such as athlete demographics, event-based performance, and medal distribution—improved my capacity to draw data-driven conclusions and develop systematic approaches to comparative analysis.

Moreover, this project strengthened my problem-solving skills in handling large datasets, optimizing code efficiency, and ensuring the accuracy of statistical insights. The combination of analytical rigor and creative visualization has been instrumental in improving my proficiency in exploratory data analysis techniques. As a result, I am now better equipped to tackle complex data-driven challenges in future research and professional endeavors. The knowledge gained through this work will serve as a foundation for further advancements in data science, sports analytics, and interactive data-driven applications.

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.

[Kaggle Dataset - 120 Years of Olympic History Athletes and Results](#)