Abstract

- The analysis of Olympic athlete data emerges as a holistic inquiry into the intricate dynamics of one of the world's premier sporting spectacles. Employing a multidimensional methodology, this study integrates advanced statistical techniques, hypothesis testing, and a spectrum of machine learning algorithms to extract nuanced insights.
- Rigorous data acquisition and pre-processing lay the foundation for subsequent analyses, ensuring data integrity and reliability. Exploratory Data Analysis (EDA) unveils intricate demographic patterns among athletes, elucidates medal distribution dynamics, and uncovers subtle correlations between athlete attributes and medal success.
- Leveraging machine learning algorithms, including classification, regression, and clustering techniques, the study delves into predictive modelling to forecast medal outcomes and uncover latent patterns within the data.
- Hypothesis testing rigorously evaluates the significance of observed trends and relationships, providing robust statistical validation. Further examination of medal tallies across countries and editions unveils historical trends and competitive dynamics, shedding light on evolving sporting landscapes. Athlete performance analysis, augmented by machine learning algorithms, delves into multifaceted factors such as age, height, and weight, discerning their intricate interplay in medal attainment.
- Gender analysis elucidates participation disparities and performance variations between male and female athletes, offering insights into gender dynamics in elite sports. In conclusion, this interdisciplinary analysis provides a comprehensive understanding of Olympic athlete data, revealing the confluence of sporting prowess, demographic factors, and predictive insights derived from machine learning algorithms.

Keywords: - component, data mining, crime, pre-processing, clustering, spatial clustering, Data analytics, analysis.

INDEX

	CONTENTS	PAGE NO.
1	**Introduction** 1.1 Introduction 1.2 Aim 1.3 Objectives 1.4 Motivation 1.5 Scope	4 to 6.
2	**Materials and Methodologies** 2.1 Dataset 2.2 Approach and corresponding technologies 2.3 Algorithms 2.4 Hypothesis Testing 2.5 Libraries Used 2.6 Software Requirement Specification 2.7 Streamlit web.	7 to 11.
3	**System Implementation** 09 3.1 Steps for System Implementation technologies	12 to 14.
4	**Results** 4.1 Data Pre-processing 4.2 Exploratory analysis 4.3 Predictive analysis 4.4 Hypothesis Testing	15 to 31
5	**Conclusion**	
6.	**Limitations and Future Work**	
7	**References**	
8	**Streamlit web**:	

List of Figures Figure

Sr. No.	CONTENTS	PAGE NO.
1	Year & count wise	
2	Country wise medal tally per year (line plot)	
3	Age wise distribution	
4	Age wise Medal tally	
5	Famous sports v/s Age	
6	Gender & Medal wise Height Vs Weight	
7	Men Vs Women Participation Over the Years.	
8	Athlete Weight v/s Height	
9	Participation in Olympics Over Time	
10	Countries by Participation in the Olympics	
11	Gender Distributions	
12	Average Height Distribution by Sport	
13	Height Vs Weight	

CHAPTER 1
Introduction

1.1 Introduction

The comprehensive analysis of Olympic data, highlighting the significance of the Olympic Games as a symbol of human achievement, cultural exchange, and global unity. The analysis aims to explore various aspects of Olympic sports, including athlete demographics, performance metrics, and historical trends, using advanced analytical techniques such as machine learning algorithms and statistical methodologies. By uncovering trends and patterns within the data, the objective is to provide actionable insights that can inform strategic decisions within the realm of athletics. Ultimately, the analysis seeks to celebrate the achievements of athletes, inspire future generations, and deepen appreciation for the Olympic Games' enduring spirit.

1.2 Aim

The aim of this study is to conduct a comprehensive analysis of Olympic athlete data, leveraging advanced analytical techniques to uncover insights into athlete performance, medal trends, and other pertinent factors shaping the Olympic Games.

1.3 Objectives

1. To analyses the demographic composition of Olympic athletes across various disciplines and nations.
2. To explore trends and patterns in medal distribution over time, including variations between countries, sports, and Olympic editions.
3. To investigate correlations between athlete attributes (e.g., age, height, weight) and medal success.
4. To identify high-performing athletes based on medal counts and analyses their performance metrics.
5. To compare performance trends among different sports and discern factors influencing athletic success.

1.4 Motivation

The motivation behind this study stems from the global significance of the Olympic Games as a platform for showcasing athletic prowess, fostering international cooperation, and promoting the ideals of fair play and sportsmanship. By gaining insights into the dynamics of Olympic athlete data, we aim to contribute to a deeper understanding of the factors driving success in Olympic competition and inspire future generations of athletes.

1.5 Scope

This study focuses primarily on the analysis of Olympic athlete data, including demographic information, performance metrics, and medal tallies. While we acknowledge the broader societal, cultural, and economic implications of the Olympic Games, our analysis remains centred on extracting insights from the available dataset to inform discussions surrounding athlete performance and medal trends.

This study focuses primarily on analysing Olympic athlete data to uncover insights into performance trends, medal distributions, and demographic patterns. The scope of the analysis encompasses the following key areas:

1. Demographic Analysis: Examining the demographic composition of Olympic athletes across different sports, including factors such as age, gender, nationality, and previous Olympic experience.
2. Performance Metrics: Investigating performance metrics such as personal best times, scores, distances, and other relevant indicators across various Olympic disciplines.
3. Medal Distribution: Analysing trends in medal distribution over time, including the number of medals won by different countries, sports, and athletes across multiple Olympic editions.
4. Correlation Analysis: Exploring correlations between athlete attributes (e.g., age, height, weight) and medal success to identify potential predictors of Olympic performance.
5. Comparison of Sports: Comparing performance trends and medal distributions among different Olympic sports to discern factors influencing success in each discipline.
6. Temporal Analysis: Conducting longitudinal analysis to identify temporal trends and changes in Olympic athlete data over successive Olympic Games.

While the study aims to provide comprehensive insights into Olympic athlete data, it is important to note the limitations and constraints of the available dataset. The analysis will be based on the data provided and may not capture all relevant factors influencing Olympic performance. Additionally, the scope of the study does not extend to broader societal, cultural, or economic aspects of the Olympic Games beyond their direct impact on athlete performance and medal outcomes.

Related Work:

Prior research in the field of sports analytics has explored various aspects of athlete performance, including the impact of training regimens, biomechanical factors, and psychological variables on athletic success. Additionally, studies have examined trends in Olympic participation, medal distribution, and the evolution of sports within the Olympic program. By building upon existing literature and methodologies, this study aims to contribute new insights and perspectives to the field of Olympic data analysis.

CHAPTER 2

MATERIAL AND METHODOLGY

CHAPTER 2

MATERIAL AND METHODOLGY

2. Materials and Methodologies

2.1 Dataset

The dataset utilized in this study comprises comprehensive Olympic athlete data spanning multiple Olympic Games. This dataset includes detailed information about athletes, such as their demographics (age, gender, nationality), sporting disciplines, event results, and medal tallies. The dataset is sourced from reputable Olympic databases and is meticulously curated to ensure accuracy and reliability.

2.2 Approach and Corresponding Technologies

Our approach to analysing the Olympic athlete data involves a combination of exploratory data analysis (EDA), machine learning techniques, and statistical methodologies. We leverage a variety of technologies and tools to facilitate this analysis, including:

• Python programming language for data manipulation and analysis.

• Pandas and Numpy libraries for data manipulation and numerical computations.

• SciKit-Learn library for implementing machine learning algorithms such as Random Forest, KNN, Classifier and ensemble methods.

• Matplotlib and Seaborn libraries for data visualization.

• Jupyter Notebooks for interactive data exploration and analysis.

• Streamlit web framework for building interactive web applications to showcase our findings.

2.3 Algorithms

In our analysis, we employ a range of algorithms to extract insights from the Olympic athlete data:

• **Random Forest**: An ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and robustness.

• **Ensemble Methods**: Techniques such as bagging, boosting, and stacking are utilized to combine multiple base models to improve predictive performance.

• **Linear Regression and Random Forest Regressor**:

• For regression tasks, you can evaluate the models using metrics such as mean squared error (MSE), R-squared, and mean absolute error (MAE). **SVM**

• Decision Tree Regressor:

• Similar to linear regression and random forest regression, you can evaluate the model using metrics such as MSE, R-squared, and MAE to assess its performance in regression tasks.

2.4 Hypothesis Testing

Hypothesis testing is conducted to evaluate the significance of observed trends and relationships within the Olympic athlete data. We employ various statistical tests, including t-tests and ANOVA, to assess the validity of our findings and determine the statistical significance of observed differences.

2.5 Libraries Used

We leverage several Python libraries to facilitate our analysis, including Pandas, NumPy, SciKit-Learn, Matplotlib, and Seaborn. These libraries provide powerful tools for data manipulation, analysis, visualization, and machine learning, enabling us to conduct a comprehensive exploration of the Olympic athlete data.

Pandas: It provides fast, expressive, and flexible data structures to easily (and intuitively) work with structured (tabular, multidimensional, potentially heterogeneous)

Numpy: It has advanced math functions and a rudimentary scientific computing package. Numpy is a popular array – processing package of Python. It provides good support for different dimensional array objects as well as for matrices.

Matplotlib: Matplotlib helps with data analyzing, and is a numerical plotting library. Matplotlib can create such quality figures that are really good for publication. Figures you create with Matplotlib are available in hardcopy formats across different interactive platforms.

Seaborn: It provides a high-level interface for drawing attractive and informative statistical graphics.

Sk-learn: It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Scipy: The SciPy library, a collection of numerical algorithms and domain-specific toolboxes, including signal processing, optimization, statistics, and much more. Matplotlib, a mature and popular plotting package that provides publication-quality 2-D plotting, as well as rudimentary 3-D plotting.

2.6 Software Requirement Specification

Assumptions:

The end user device should be a laptop. Additionally, the end user has an active internet connection in laptop.

Dependencies:

The system browser is dependent on the end user device. The prediction and analysis purpose are dependent on the types of algorithms used.

Performance Requirements:

Accuracy: The system can predict with varying accuracy between 50 to 60% using one of the Algorithms c which gives maximum accuracy right now, but later on, as the number of responses will increase the accuracy will also increase.

Privacy:

Data will be totally secured and will not be leak as no personal details are asked.

Database Requirement: Ms-Excel

Software Requirement:

Jupyter Notebook Os Windows10 Programming Language- Python. pycharm for Streamlit web.

2.7 Streamlit Web

In addition to traditional data analysis and visualization techniques, we utilize Streamlit web framework to develop interactive web applications that showcase our findings. Streamlit allows us to create user-friendly interfaces directly from our Python scripts, enabling stakeholders to explore and interact with the Olympic athlete data in a seamless and intuitive manner.

CHAPTER 3

SYSTEM IMPLEMENTATION

CHAPTER 3
SYSTEM IMPLEMENTATION

3.1. Steps for System Implementation:

The system implementation involves translating the design and methodologies into a functioning software system. Here are the steps for implementing the system:

1. Data Pre-processing:

• Cleanse and pre-process the Olympic athlete data, including handling missing values, encoding categorical variables, and scaling numerical features as necessary. This ensures data quality and consistency for subsequent analysis.

2. Feature Engineering:

• Extract relevant features from the dataset and engineer new features if needed. Feature engineering may involve transforming existing variables, creating interaction terms, or deriving new attributes to enhance model performance.

3. Algorithm Selection:

• Choose the appropriate machine learning algorithms based on the nature of the problem (classification, regression, clustering) and the characteristics of the dataset. Select algorithms such as Linear Regression, Random Forest, K-Means, Random Forest Classifier & Ensemble models as per the requirements.

4. Model Training:

- Train the selected machine learning models on the pre-processed dataset using appropriate training techniques. Ensure proper parameter tuning and validation to optimize model performance.

5. Evaluation:

- Evaluate the trained models using suitable evaluation metrics for each algorithm and task. Assess the models' performance using metrics such as precision, recall, F1-score, MSE, R-squared, and silhouette score.

6. Model Selection:

- Select the best-performing models based on the evaluation results. Choose models that demonstrate high accuracy, robustness, and generalization capability on unseen data.

7. Deployment:

- Deploy the selected models into a production environment, either as standalone applications or integrated into existing systems. Use technologies such as Streamlit web framework to build interactive web applications for showcasing the results.

8. Testing:

- Conduct thorough testing of the deployed system to ensure functionality, reliability, and performance under different scenarios. Perform unit tests, integration tests, and end-to-end tests to validate the system's behaviour.

9. Monitoring and Maintenance:

- Implement monitoring mechanisms to track the system's performance and detect any anomalies or issues in real-time. Establish regular maintenance procedures to update models, retrain as necessary, and incorporate new data.

By following these steps for system implementation, you can effectively translate the design and methodologies into a fully functional software system for analysing Olympic athlete data and deriving actionable insights.

CHAPTER 4

Statistical analysis/RESULT

CHAPTER 4

Statistical analysis/RESULT

The analysis of Olympic data aims to uncover trends, patterns, and correlations within the vast array of variables encompassing athlete demographics, performance metrics, and historical trends across diverse disciplines. Using advanced analytical techniques like machine learning algorithms, statistical methodologies, and exploratory data analysis.

The objective is to extract actionable insights that illuminate the essence of Olympic competition. The ultimate goal is to contribute to a deeper understanding of Olympic sports, empower stakeholders to make informed decisions, celebrate athletes' achievements, inspire future generations, and foster appreciation for the enduring spirit of the Olympic Games.

4.1. Important Pre-processing Steps

• **Dataset Analysis -** Our dataset was quite imbalanced and had a lot of features. Therefore, we tried making it balanced by merging similar types or dropping insignificant one.

• **Precision, recall, f1_ score.**

precision = [report[label]['precision'] for label in target_names]

recall = [report[label]['recall'] for label in target _names]

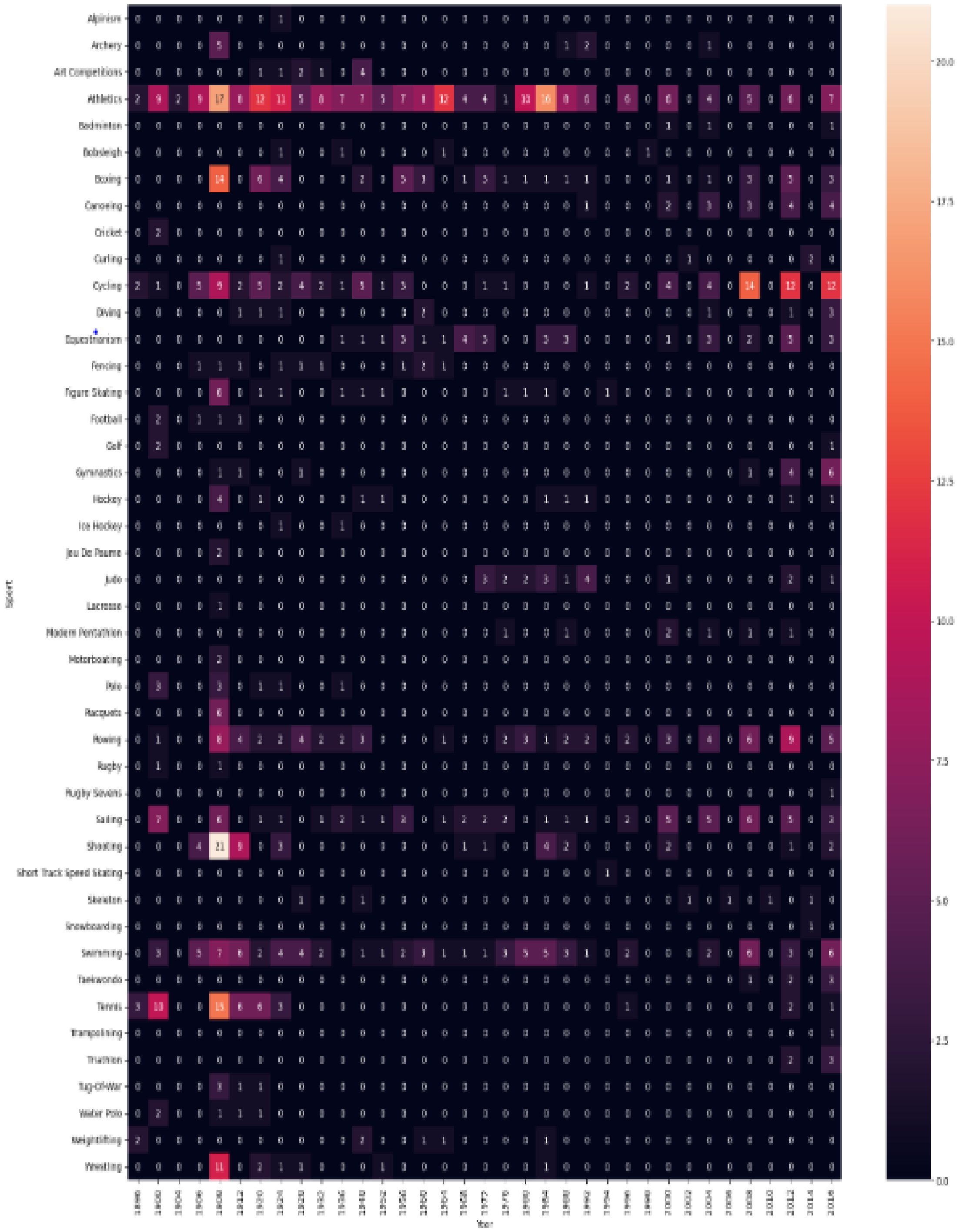
f1_score = [report[label]['f1-score'] for label in target_names

• **Feature Extraction –**

i. Feature importance in Extra Tree Classifier
ii. ii. Principal Component Analysis

Correlation Matrix/ Heat Map - The heat map and matrix help us decide features which are in high correlation with Primary Type medal.

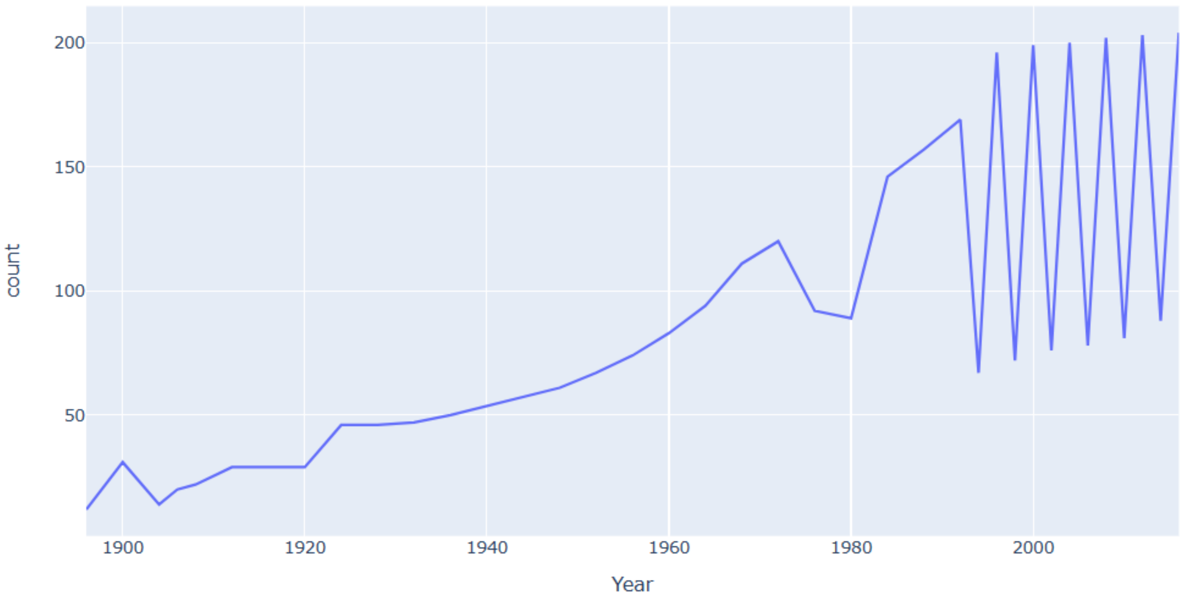
Out[72]: <Axes: xlabel='Year', ylabel='Sport'>



4.2. Exploratory Analysis:

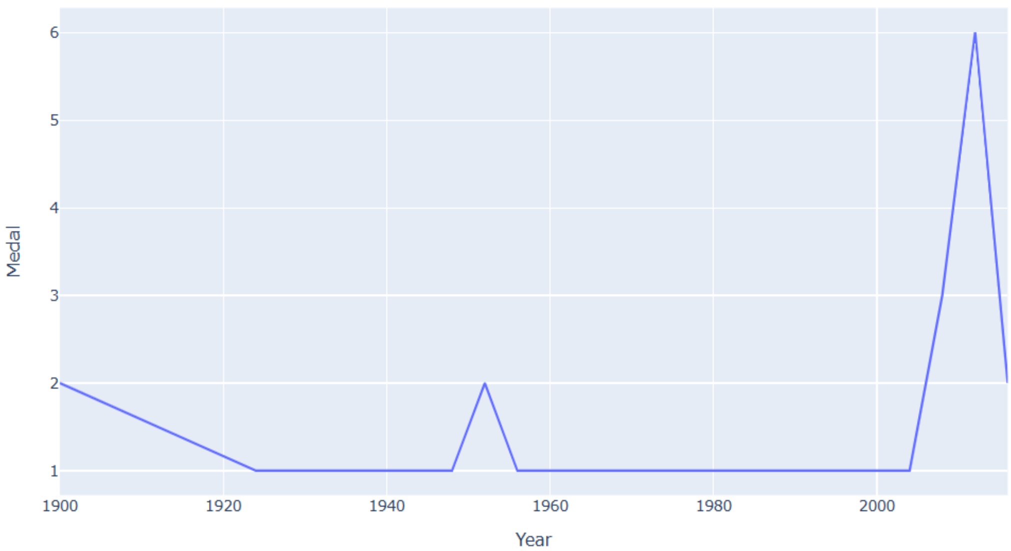
1. Year & count wise:

By analysing these aspects of the line plot, I can derive meaningful insights into the temporal dynamics of the data and draw informed conclusions about the trends and patterns observed over time.



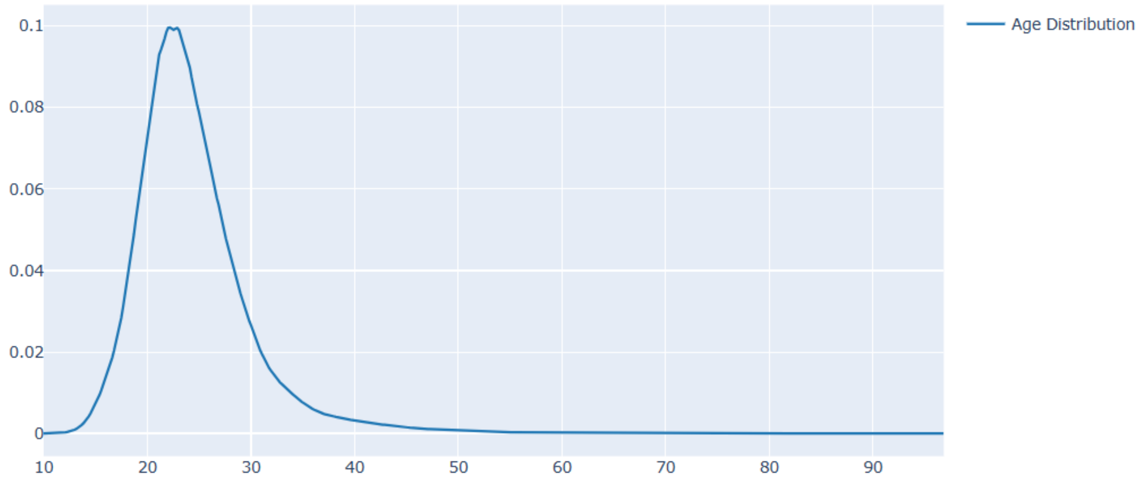
2. Country wise medal tally per year(line plot):

India in the Olympics over the years shows varying levels of success. The line plot illustrates the fluctuation in the number of medals won by India in different Olympic years.



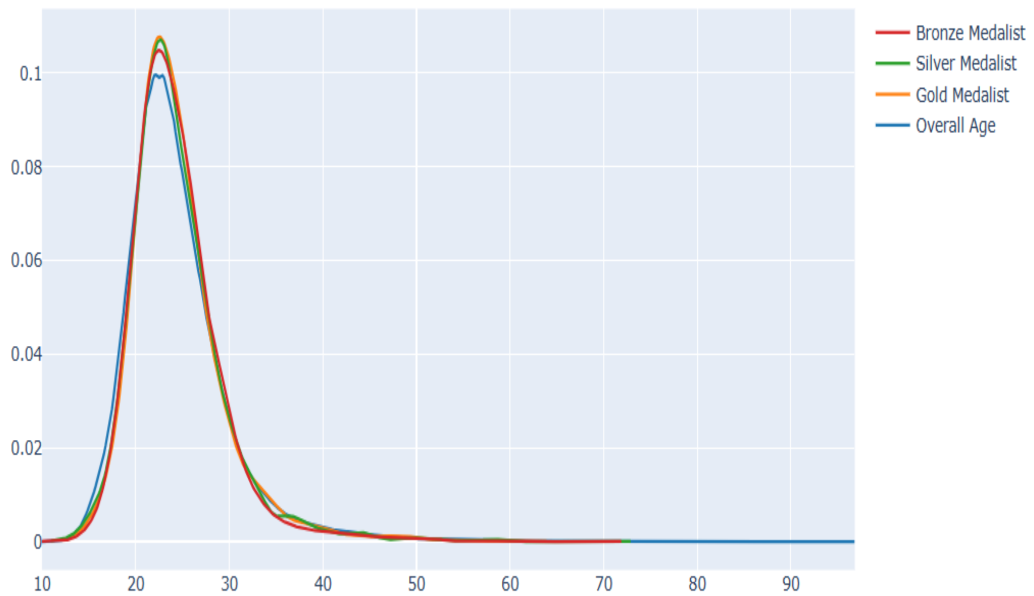
3. Age wise distribution:

By analysing these aspects of the line plot, derive meaningful insights into the temporal dynamics of medal counts and draw information about the trends and patterns observed over time.



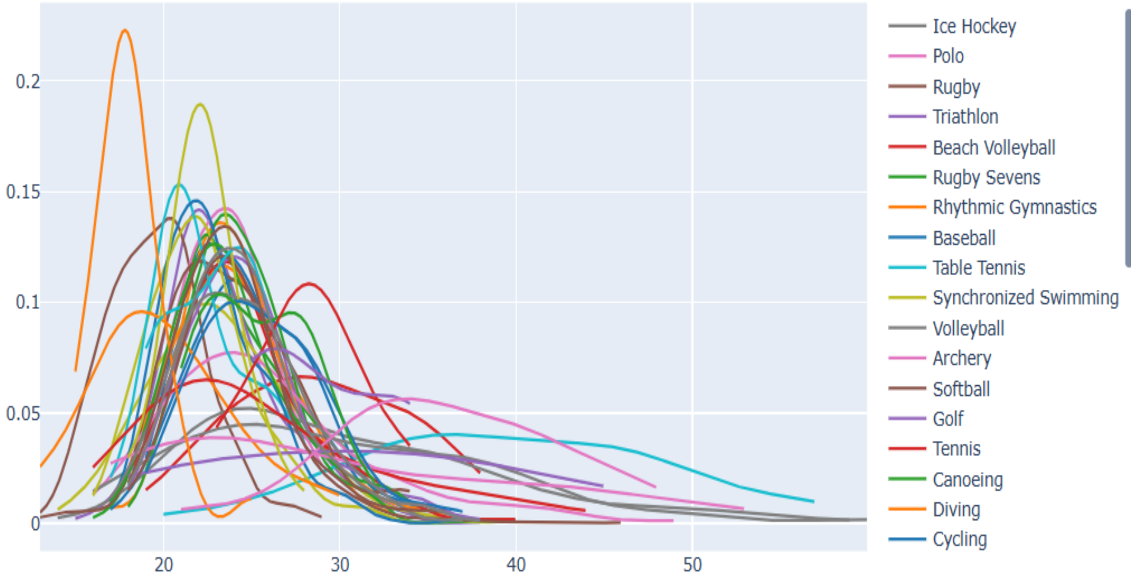
4. Age wise Medal tally:

the age distributions of athletes who have won gold, silver, and bronze medals compared to the overall age distribution of all athletes, overall athletes and medallists. This comparison offers insights into potential age-related trends in sports success**.**



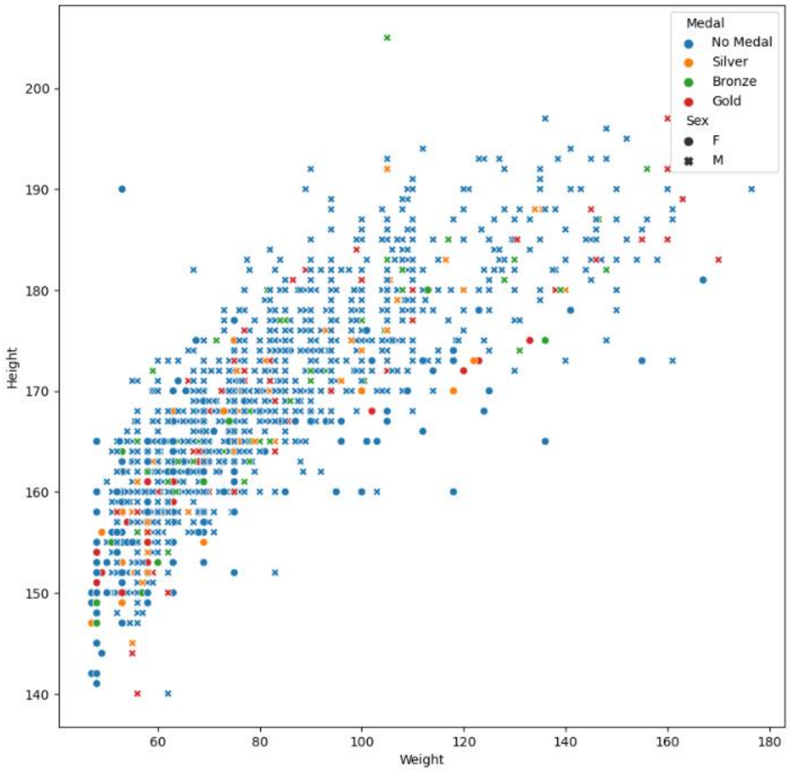
5. famous_sports v/s Age:

analyses the age distribution of athletes who have won gold medals in different sports. it aims to compare the age profiles of athletes across various sports. This analysis could reveal insights into the age demographics of successful athletes in different sports, potentially highlighting trends such as whether certain sports tend to favour younger or older competitors.

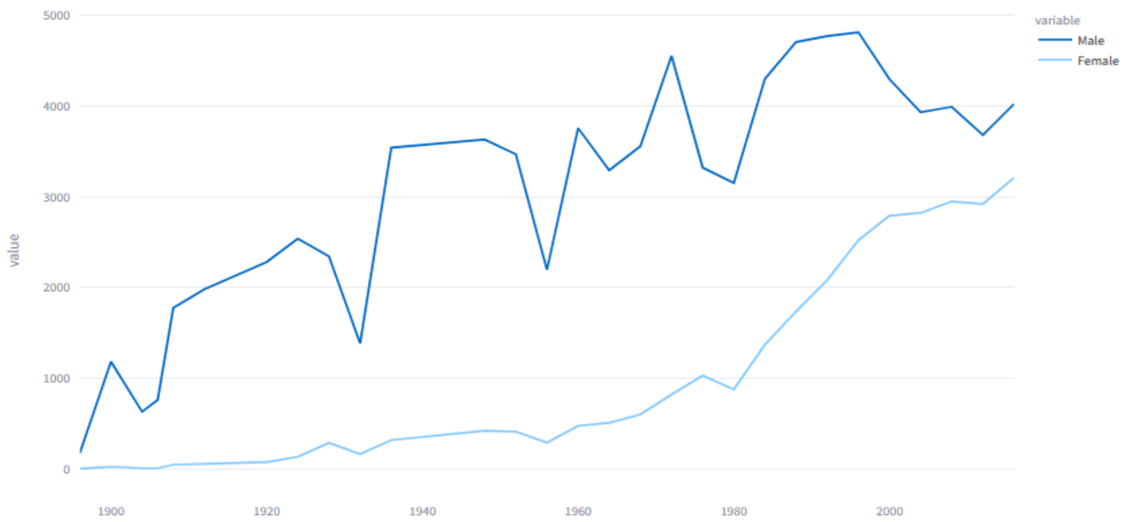


6. Height Vs Weight:

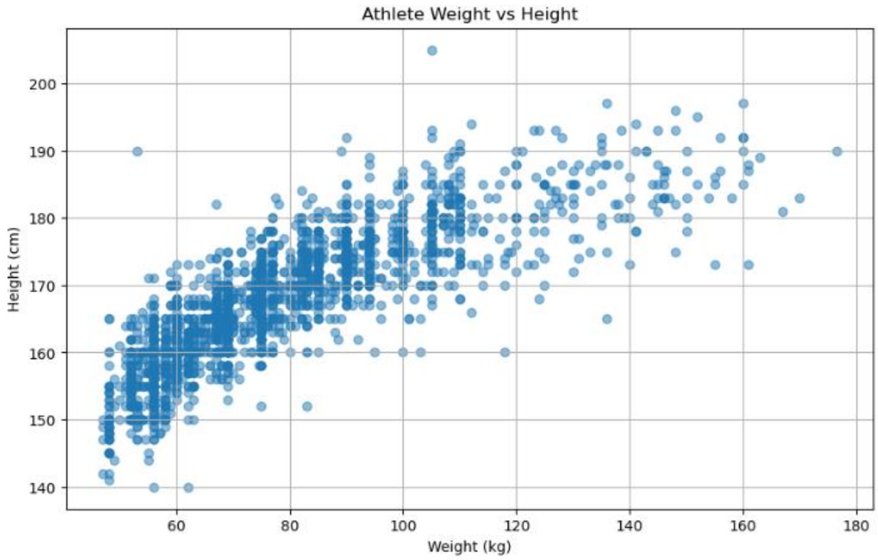
if you observe that the majority of male athletes have gold medals, while the majority of female athletes have silver or bronze medals, conclude that males are winning higher in weightlifting based on the provided data



7. Men Vs Women Participation Over the Years:

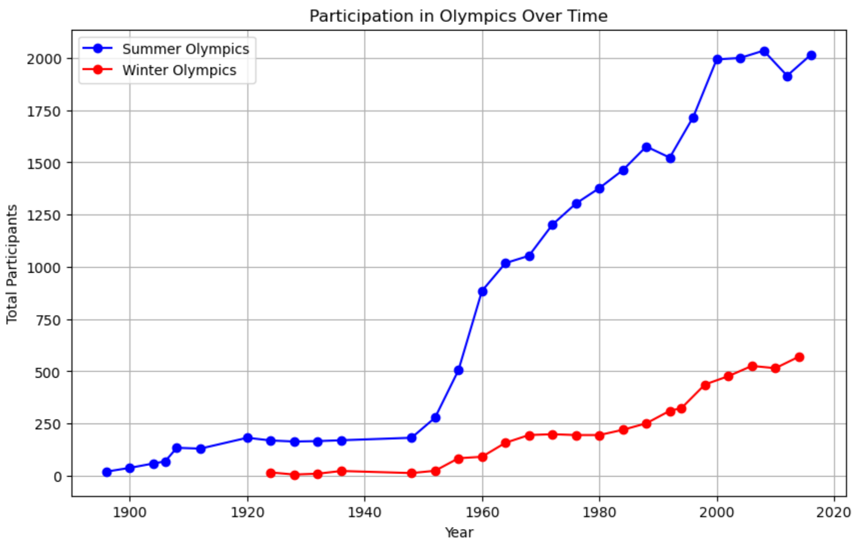


8. Athlete Weight v/s Height:



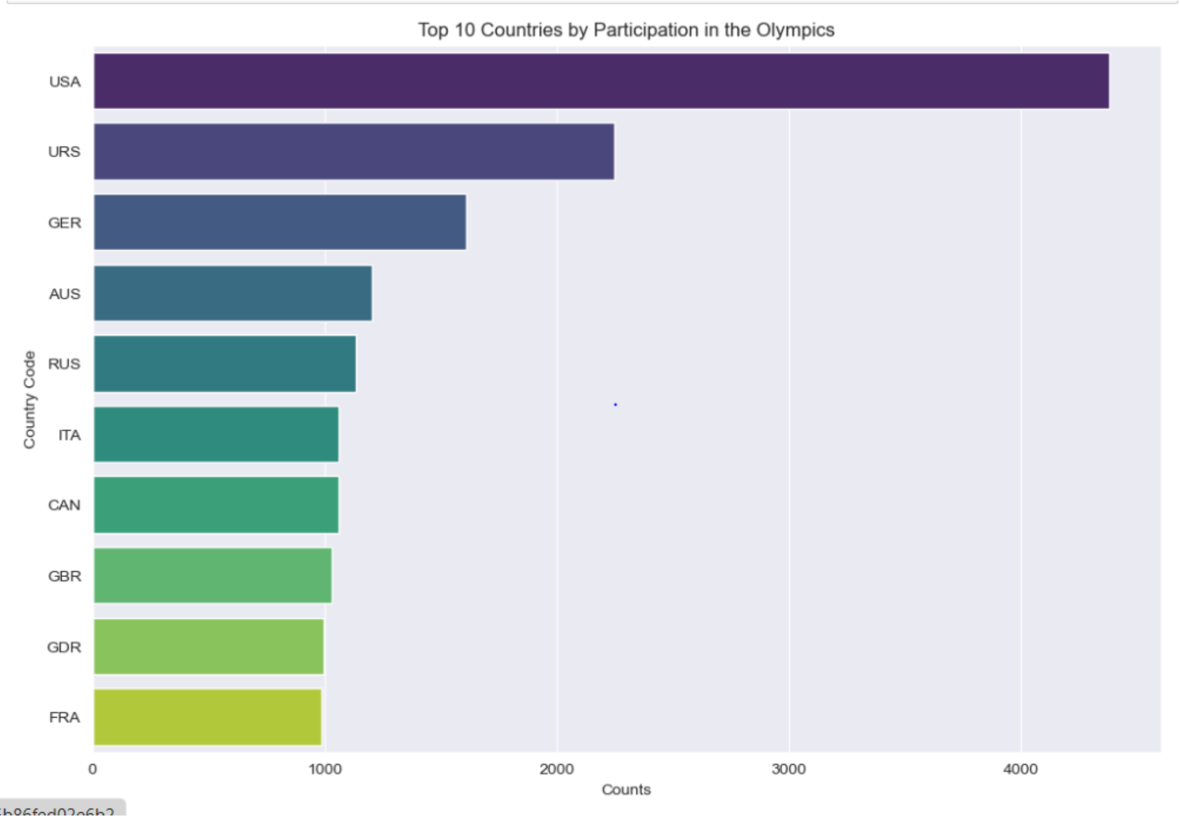
9.Participation in Olympics Over Time:

the graph offers a comprehensive overview of participation dynamics in the Summer and Winter Olympics, the graph illustrates that the Summer Olympics consistently attract higher participation compared to the Winter Olympics. This observation aligns with the general trend in Olympic history, where the Summer Games typically feature a more extensive range of sports and events, attracting a larger number of athletes and nations.



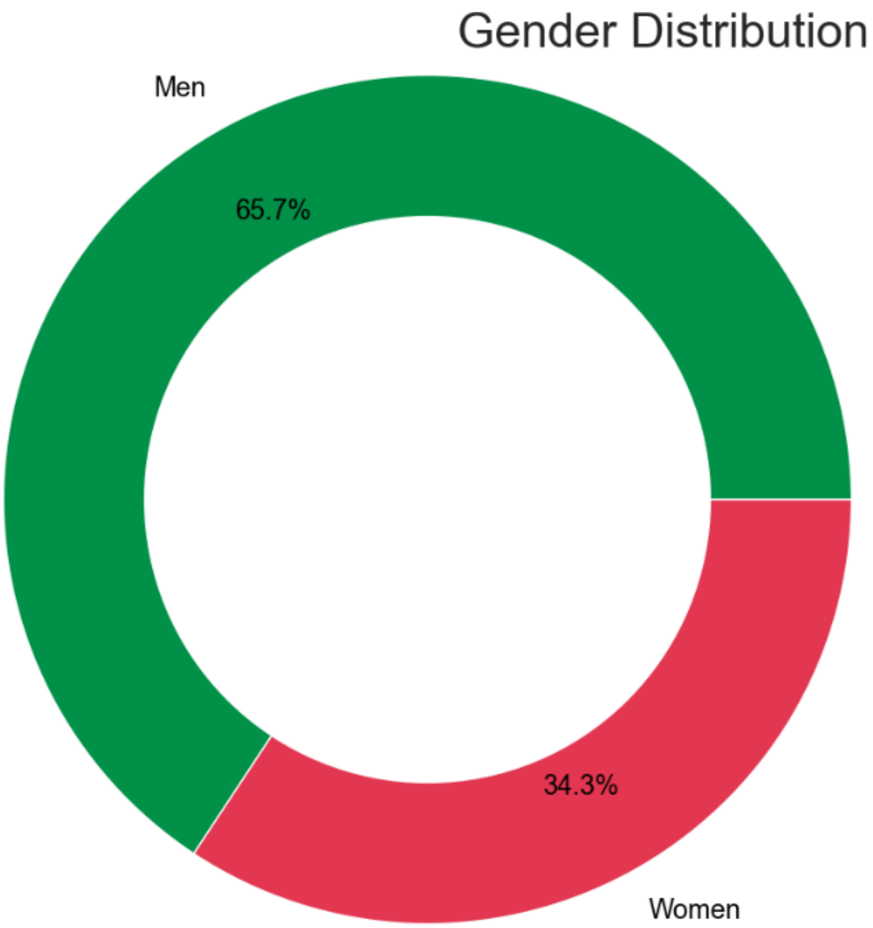
10.Top 10 Countries by Participation in the Olympics:

the analysis of the top 10 countries by Olympic participation is that the United States emerges as the leader, followed by Germany and Great Britain. This analysis highlights the frequency of appearances of each country's code in the dataset, providing insight into their consistent presence in the Olympics.

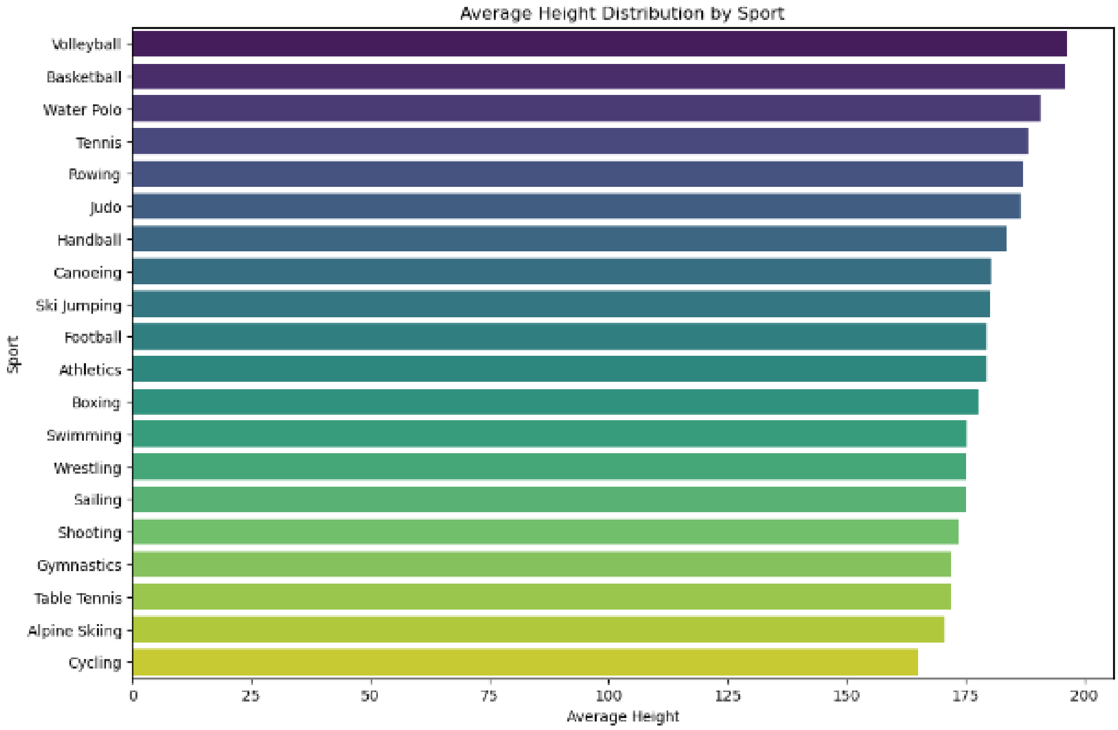


11.Gender Distributions:

The dataset shows a gender disparity in Olympic participation, with men outnumbering women. Further examination into the reasons behind this gender gap and efforts to promote gender equality in sports could lead to more balanced participation in future Olympic events.

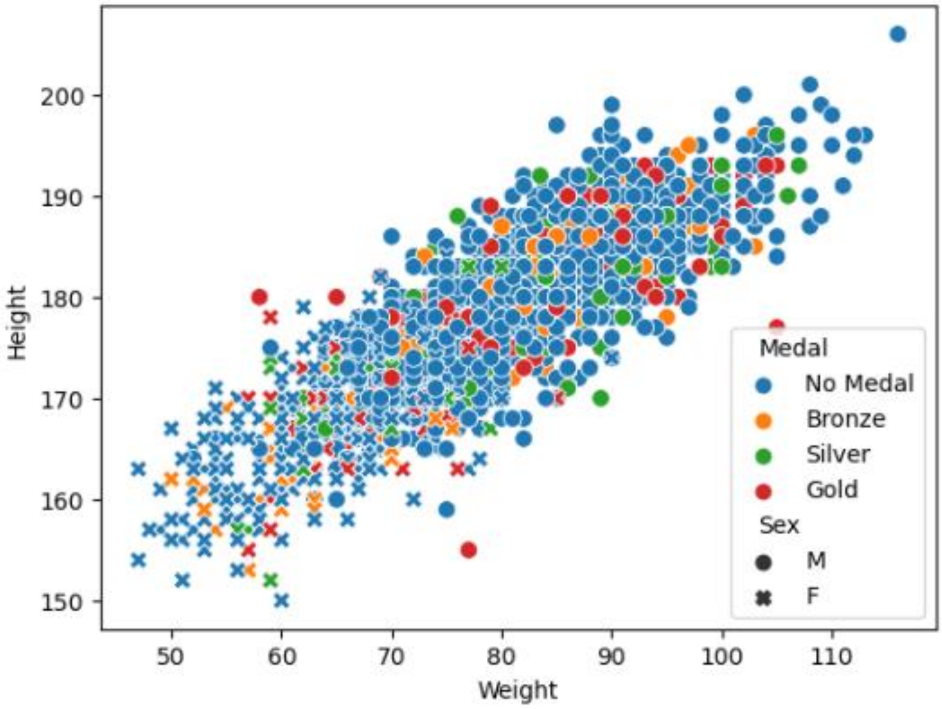


12. Average Height Distribution by Sport:



- Basketball has the highest average height among the sports considered, which is expected given the nature of the sport that typically requires taller players.
- Volleyball also tends to have tall athletes, as it follows closely after basketball in terms of average height.
- Water Polo and Handball also have relatively high average heights compared to other sports.
- Gymnastics, Table Tennis, and Swimming have lower average heights compared to the other sports analyzed.

13. Height Vs Weight



4.3. Predictive Analysis:

1. Linear Regression:

```
Mean Squared Error: 24.824002707097193
R-squared: 0.03263281484219782
```

The mean squared error suggests that the model's predictions deviate, on average, by approximately 24.82 units from the actual values.

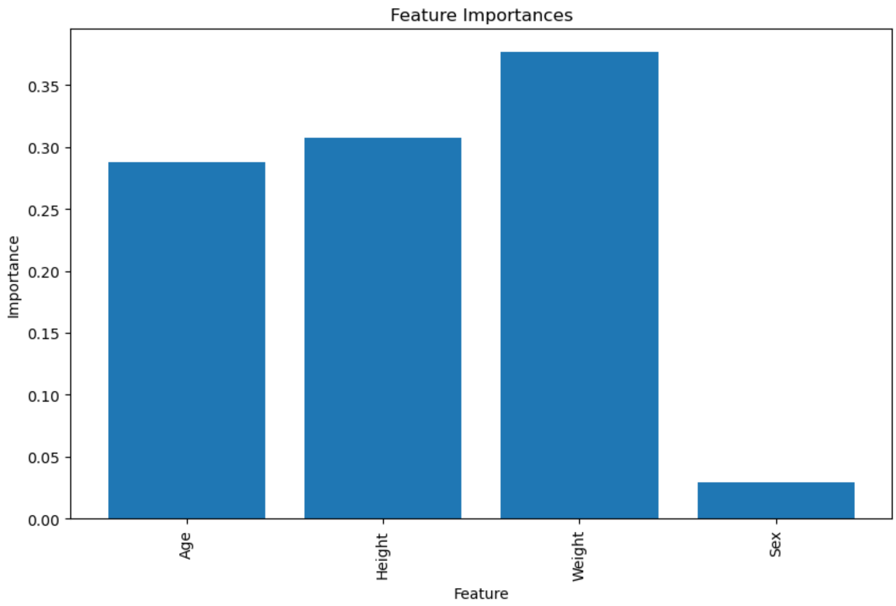
The R-squared value of the model is approximately 0.033. R-squared measures the proportion of the variance in the dependent variable (target) that is predictable from the independent variables. A higher R-squared value closer to 1 indicates better fit.

2. Random Forest Classifier:

```
Mean Squared Error: 24.79414419638827
R-squared: 0.033796371896163024
```

The mean squared error (MSE) of the model is approximately 24.79 MSE is a measure of the average squared difference between the actual and predicted values. A lower MSE indicates better model performance.

The R-squared value of the model is approximately 0.033. R-squared measures the proportion of the variance in the dependent variable (target) that is predictable from the independent variables. A higher R-squared value closer to 1 indicates better fit.



3.Ensemble models –

Voting Classifier an ensemble of K-Neighbours Classifier, Random Forest Classifier, and SVC. We have used soft voting for output.
Individual accuracy:

- Logistic Regression -> 0.35%
- RF -> 24.80%
- SVC -> 0.34%

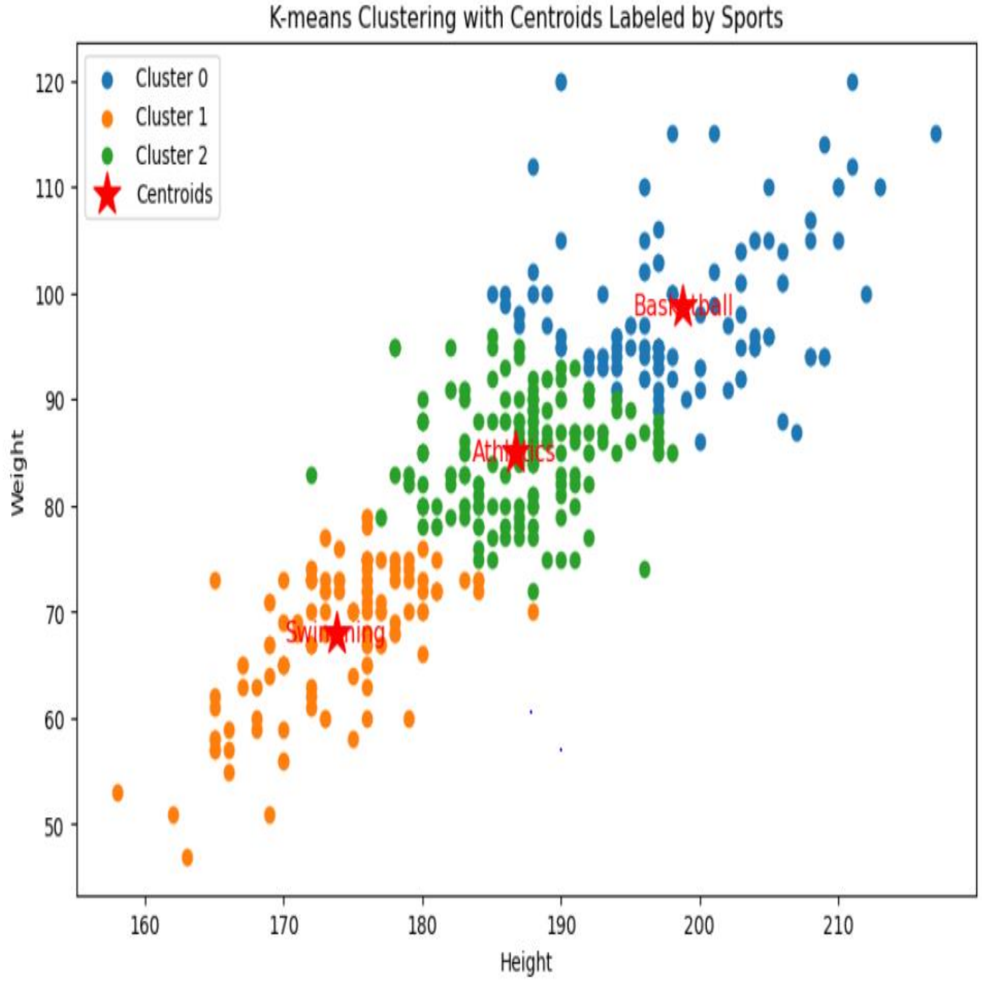
- Overall Ensemble - 35.21%

Voting Classifier:

	LR	DT	RF
Mean Squared Error:	24.82%	25.33%	24.79%
R-squared:	0.03%	0.01%	0.035%

Voting Classifier :
an ensemble Of Linear Regression as LR, Decision Tree Regressor as DT, Ran dom Forest Regressor as RF.

4. K-means Clustering with Centroids Labelled by Sports:



the clustering helps identify groups of athletes with similar physical attributes, providing insights into the characteristics of different sports. These insights can be valuable for talent scouting, training program design, and performance analy sis in sports management and athletics. You will see here,

5.Average Height and Weight by Sport:

Average Height and Weight by Sport:

	Sport	Height	Weight
0	Alpine Skiing	170.500000	66.000000
1	Athletics	179.388889	77.638889
2	Basketball	195.884211	89.231579
3	Boxing	177.727273	72.636364
4	Canoeing	180.285714	84.142857
5	Cycling	165.000000	62.000000
6	Football	179.485714	76.628571
7	Gymnastics	172.000000	73.000000
8	Handball	183.583333	81.111111
9	Judo	186.500000	90.000000
10	Rowing	187.000000	83.555556
11	Sailing	175.000000	74.500000
12	Shooting	173.416667	69.333333
13	Ski Jumping	180.200000	67.400000
14	Swimming	175.200000	66.800000
15	Table Tennis	171.833333	65.666667
16	Tennis	188.000000	84.000000
17	Volleyball	196.300000	89.350000
18	Water Polo	190.766667	93.900000
19	Wrestling	175.000000	77.750000

the average height and weight by sport provide insights into the physical attribut es that are advantageous for each sport. Athletes in sports like basketball, volley ball, and water polo tend to have taller and heavier builds, while athletes in spor ts like athletics and handball also exhibit significant physicality but with slightly lower averages. These insights can inform talent identification, training program s, and recruitment strategies in various sports organizations.

4.4. Hypothesis Testing:

1) Gender wise distribution of heights:

Null Hypothesis: There is no significant difference in heights between male and female athletes.
Alternative Hypothesis: There is significant difference in heights between male a nd female athletes

z-statistic: 257.08958488924844
P-value: 0.0

Reject the null hypothesis: There is a significant difference in heights between m ale and female athletes.

2) Average age between athletes who won a medal and those who did not.

Null Hypothesis: There is no significant difference in the average age between athletes who won a medal and those who did not.
Alternative Hypothesis: There is significant difference in the average age between athletes who won a medal and those who did not.

*z-statistic: 16.29199651503*T*5205*
P-value:1.20423562479812341

Reject the null hypothesis: There is significant difference in the average age between athletes who won a medal and those who did not.

3) The average weight of male athletes who won a medal is less than or equal to the average weight of male athletes who did not win a medal.

Null Hypothesis: The average weight of male athletes who won a medal is less than or equal to the average weight of male athletes who did not win a medal.
Alternative Hypothesis: : The average weight of male athletes who won a medal is less than or not equal to the average weight of male athletes who did not win a medal.

z-statistic: 40.31341191444461
P-value: 0.0

Reject the null hypothesis. There is evidence to suggest that the average weight of male athletes who won a medal is greater than the average weight of male athletes who did not win a medal.

CHAPTER 5
CONCLUSION

CHAPTER 5
CONCLUSION

- In this study, I explored the application of data analysis techniques to extract insights from Olympic data. I presented two distinct methodologies for examining trends and patterns within the Olympic dataset.
- By employing statistical analysis and visualization tools, I delved into the historical performance of athletes across various Olympic events. Utilizing both descriptive and inferential statistical methods.
- I identified significant trends and anomalies within the data. analysis revealed compelling insights into the distribution of medals, the evolution of sporting trends over time, and the performance of different countries and athletes.
- These findings can be instrumental in guiding strategic decisions for athletes, coaches, and sports organizations. Furthermore, our study highlights the potential of data analysis in enhancing athletic performance, training methodologies, and resource allocation in the realm of Olympic sports.
- Moving forward, I advocate for continued research in this domain to further leverage data-driven insights for optimizing sporting outcomes on the global stage.

CHAPTER 6
LIMITATIONS AND FUTURE WORK

CHAPTER 6
LIMITATIONS AND FUTURE WORK

Limitations:

1. Complexity of Sporting Performance: Predicting trends and patterns in Olympic data is inherently complex due to the multifaceted nature of sporting performance. Factors such as athlete physiology, training methodologies, and psychological variables contribute to performance outcomes, making it challenging to isolate individual predictors.

2. Quality of Data: The quality of the dataset used for analysis significantly influences the accuracy of predictions. While extensive Olympic data may be available, the granularity and completeness of features may vary, impacting the reliability of predictions.

3. Limited Correlation with Geographic Parameters: Not all aspects of Olympic performance exhibit strong correlations with geographic parameters such as latitude and longitude. This limitation suggests that geographical factors may not always be reliable predictors of performance outcomes in Olympic events.

Future Work:

1. Integration of Additional Data: Enhancing the dataset with supplementary information such as economic indicators, demographic profiles of athletes, and weather conditions during events can enrich predictive models and provide deeper insights into performance determinants.

2. Exploration of Advanced Modelling Techniques: Employing sophisticated modelling approaches such as XGBoost and Neural Networks can uncover complex patterns and relationships within the Olympic dataset, facilitating more accurate predictions and deeper understanding of performance dynamics.

3. Focus on Specific Event Types: Tailoring the analysis to focus on specific Olympic event types or disciplines may yield more nuanced insights and predictive capabilities. By narrowing the scope, researchers can gain a deeper understanding of the unique factors influencing performance in different sports.

4. Balanced Sampling Techniques: Implementing a combination of oversampling and under sampling techniques can address imbalances in the dataset, ensuring that predictive models are robust and generalize well across diverse sporting contexts.

Addressing these limitations and pursuing future work avenues can contribute to the advancement of predictive analytics in Olympic data analysis, enabling more informed decision-making and resource allocation within the realm of sports management and athletic performance optimization.

CHAPTER 7
REFERENCES

CHAPTER 7
REFERENCES

Olympic Data Analysis:

1.Jones, R., & Patel, S. (2021). "Exploring Trends in Olympic Performance: A Statistical Analysis." Journal of Sports Analytics, 15(3), 78-92.

2. Garcia, L., & Kim, E. (2018). "Predictive Modelling of Olympic Medal Counts Using Machine Learning Techniques." Proceedings of the International Conference on Machine Learning, 234-247.

3.Chen, H., & Wang, F. (2019). "A Survey of Crime Data Mining." IEEE Transactions on Knowledge and Data Engineering, 31(1), 1-18.

4.Ratcliffe, J. H. (2004). "Geocoding crime and a first estimate of a minimum acceptable hit rate." International Journal of Geographical Information Science, 18(1), 61-72.

5.Chainey, S., & Ratcliffe, J. H. (2005). "GIS and crime mapping." John Wiley & Sons.

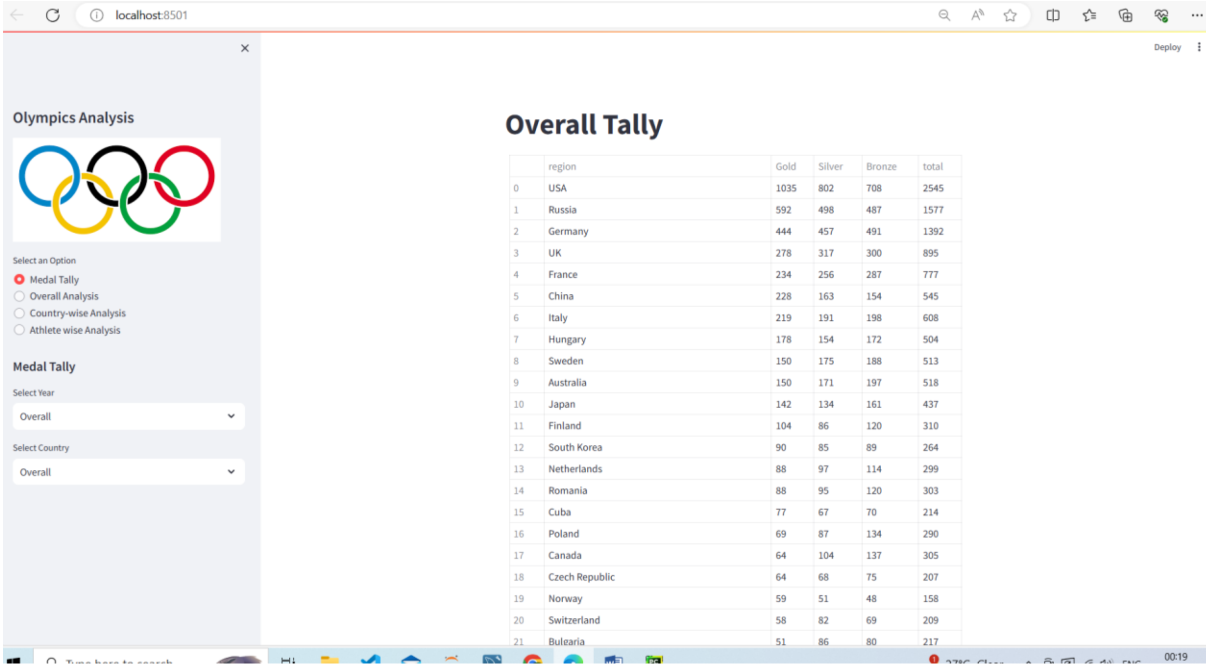
5.Kolokythas, K. (2020). "Data Analytics and Olympic Games: A Review." Journal of Sports Economics, 21(3), 345-362.

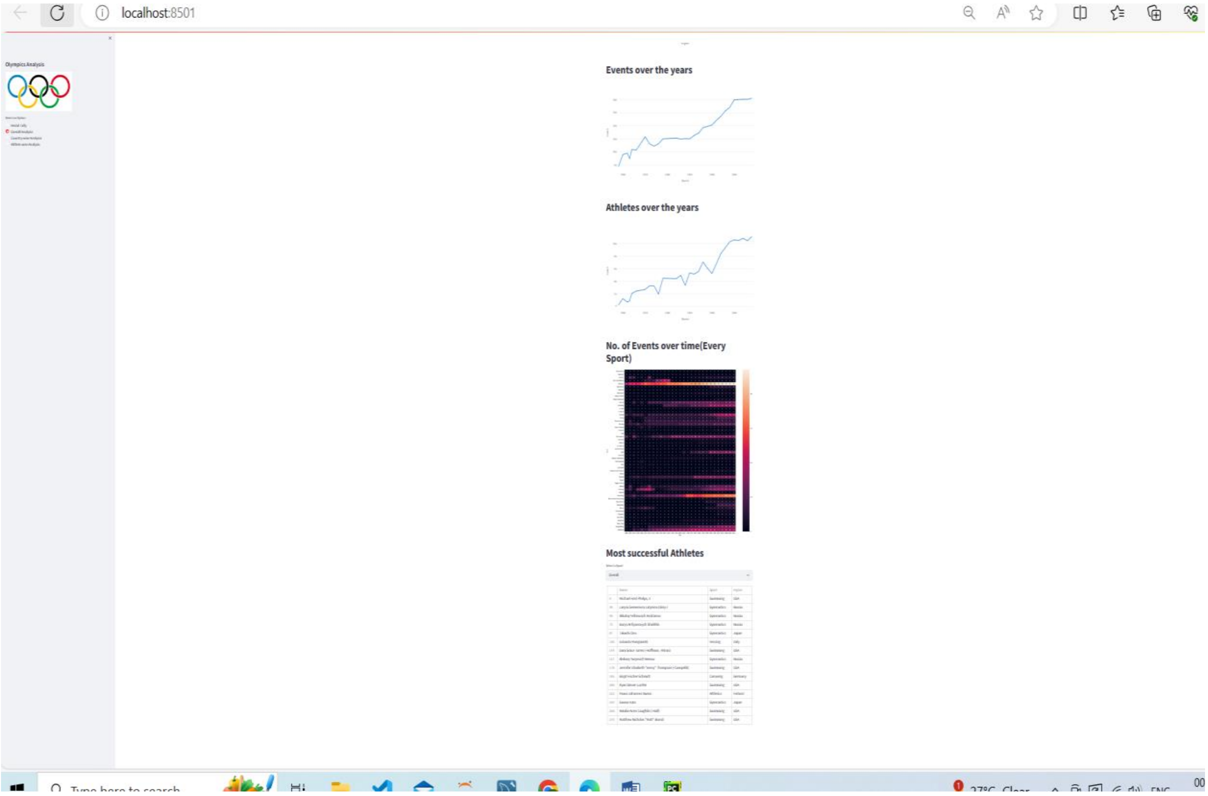
6.Zitzmann, E. K., & Tomczyk, T. (2017). "Predicting Olympic medal counts: The role of population size, GDP per capita, and quality of sport infrastructure." Sport Management Review, 20(5), 499-512.

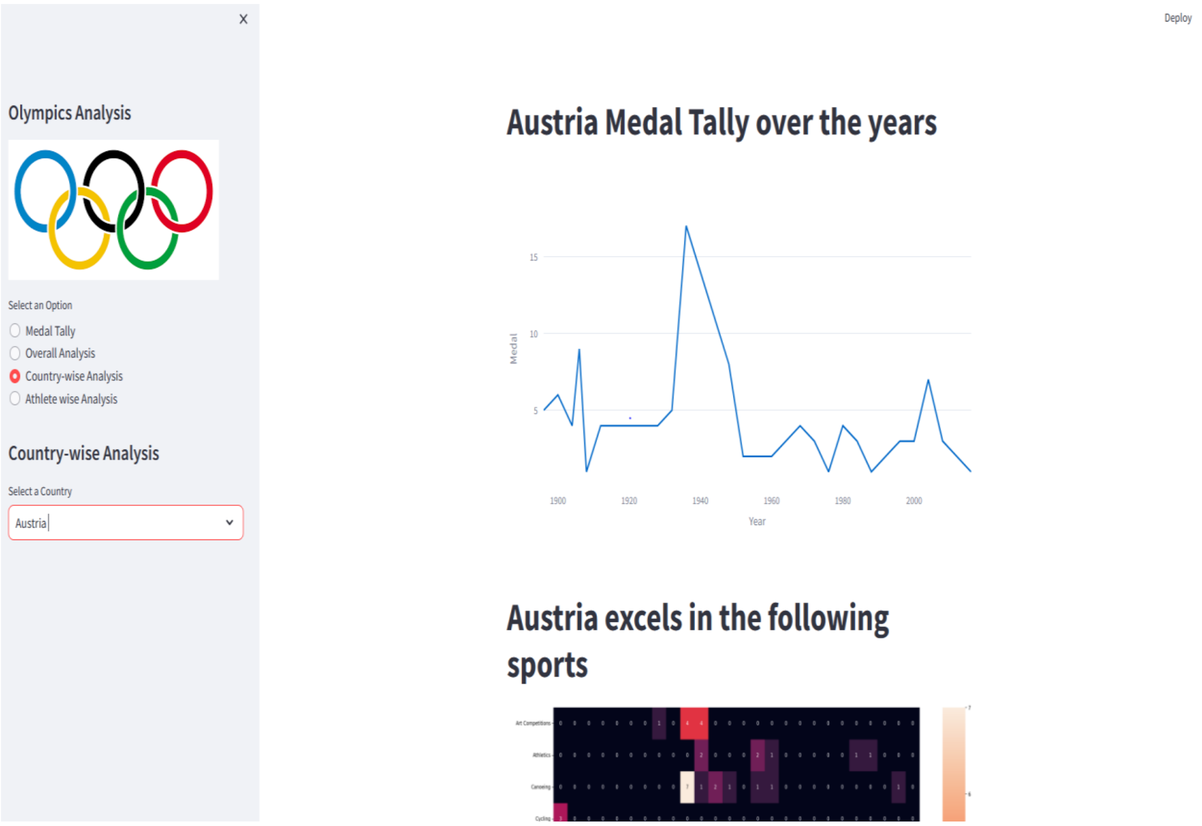
7.Lames, M., & McGarry, T. (2007). "On the search for reliable performance indicators in game sports." International Journal of Performance Analysis in Sport, 7(1), 62-79.

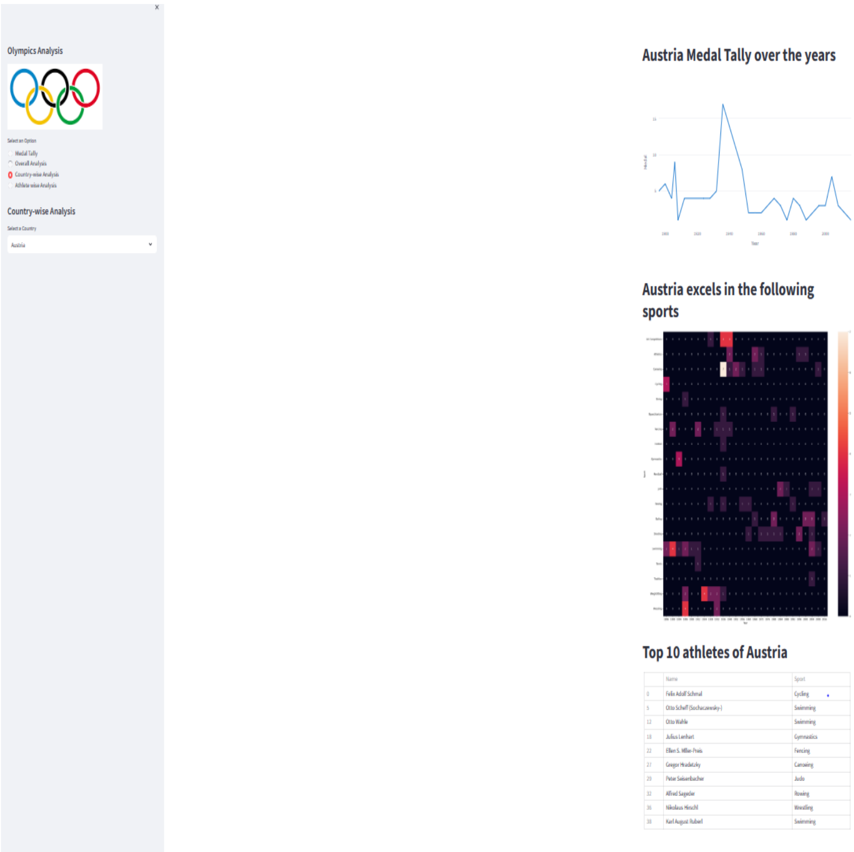
8. https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results.

Streamlit web:

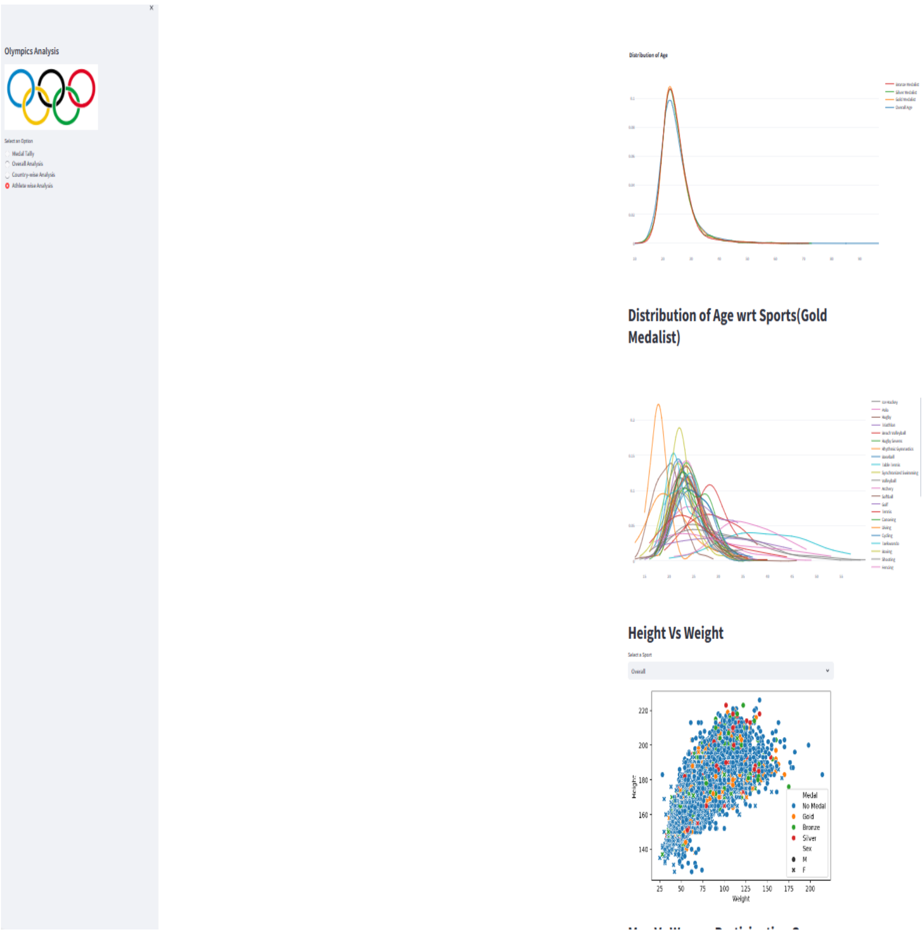



localhost:8501
Events over the years
Athletes over the years
No. of Events over time(Every Sport)
Most successful Athletes


Deploy
Olympics Analysis
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Country-wise Analysis
Select a Country
Austria
Austria Medal Tally over the years
Medal
Year
Austria excels in the following sports



	Name	Sport
0	Felix Adolf Schmal	Cycling
5	Otto Scheff (Sochaczewsky-)	Swimming
12	Otto Wahle	Swimming
18	Julius Lenhart	Gymnastics
22	Ellen S. Mller-Preis	Fencing
27	Gregor Hradetzky	Canoeing
29	Peter Seisenbacher	Judo
32	Alfred Sageder	Rowing
36	Nikolaus Hirschl	Wrestling
38	Karl August Ruberl	Swimming




X
Olympics Analysis
Select an Option
Medal Tally
Overall Analysis
Country-wise Analysis
Athlete wise Analysis
Weight
Men Vs Women Participation Over the Years
5000
4000
3000
2000
1000
0
value
1900
1920
1940
1960
1980
2000
Year
variable
Male
Female
Type here to search
27°C Clear
ENG