PROJECT DOCUMENTATION

Olympic sports Analysis

1. Introduction :

1.1 Project Overview:

The project focuses on leveraging data analytics and visualization techniques to delve into Olympic sports participation and performance. It entails exploring data from the Olympic Games, encompassing various dimensions such as athlete demographics, events, medals, and country-wise representations. The primary goal is to extract meaningful insights from this extensive dataset and translate them into actionable information for stakeholders within the sports domain.

1.2 Purpose

The purpose of this project is to understand and analyze the dynamics of Olympic sports, utilizing data-driven approaches. By investigating athlete performance, country-wise achievements, and historical trends within the Olympic Games, the aim is to derive insights beneficial for diverse stakeholders, including sports organizations, athletes, sponsors, and governing bodies. This analysis serves to enhance decision-making, athlete development strategies, marketing endeavors, and overall strategic planning within the Olympic sports landscape.

2. Literature survey :

2.1 Existing Problem:

Within the realm of Olympic sports, several challenges and issues persist, influencing various facets of the Games:

- Athlete Development: Challenges in identifying and nurturing talent, providing adequate training facilities, and supporting athletes across different nations.

- Equitable Representation: Disparities in opportunities and resources among participating nations, impact their performance and representation.
- Technological Integration: Adapting to and leveraging technology for fair play, performance analysis, and enhancing the overall sports experience.
- Commercialization and Ethical Concerns: Balancing the commercial interests and ethical considerations to maintain the integrity of the Games.
- Sustainability and Global Reach: Ensuring the sustainability of the Games and extending its global reach to foster participation and inclusivity.

2.2 References:

References for the literature survey could include academic articles, books, and research papers on various aspects of the Olympic Games, sports economics, athlete development, technological advancements, governance, and societal impacts of the Olympics. Some key references might include scholarly articles on the evolution of the Games, sports governance, athlete performance analysis, and the economic impact of the Olympics on host cities and countries.

2.3 Problem Statement Definition:

Problem Statement: How might we enhance our understanding of the evolving patterns of Olympic sports participation and performance, considering factors such as diverse global representation, the impact of professional athletes, geopolitical influences like boycotts, the role of media and commercialization, inclusivity in specialized games, and the long-term effects of historical disruptions like wars and cancellations?

Scope: Analyzing global trends in Olympic sports participation, considering diverse representation, the rise of professional athletes, geopolitical influences, media and commercialization impact, inclusivity in specialized games, and the long-term effects of historical disruptions.

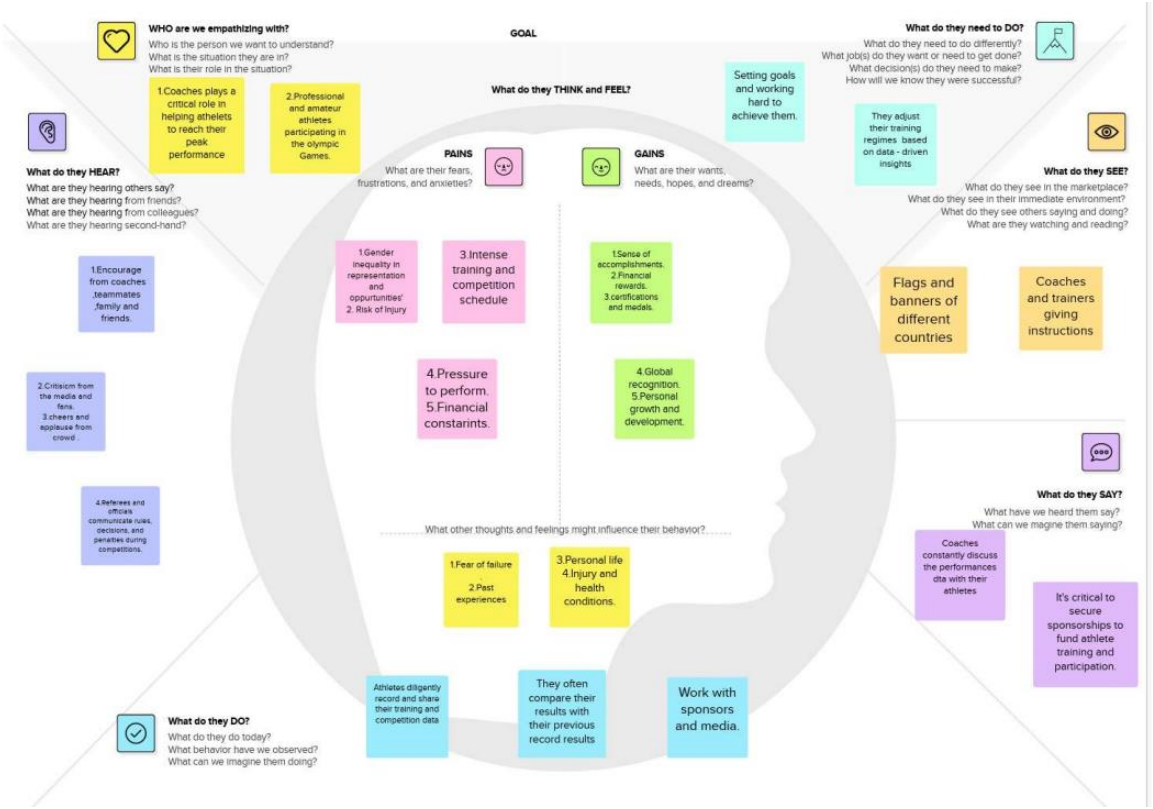
Impact: Informed decision-making for resource allocation, policy formulation for inclusivity, enhanced athlete development strategies, improved international relations through understanding geopolitical influences, and historical insights for future planning, ensuring the resilience and sustainability of the Olympic Games.

3.Ideation & Proposed Solution

3.1 Empathy Map Canvas:

In the context of our project, "Data-Driven Insights on Olympic Sports Participation and Performance," an empathy map is a strategic tool used to gain a deeper understanding of the athletes and participants involved in Olympic sports. It helps you develop empathy by putting yourself in their shoes, allowing you to comprehend their needs,

desires, frustrations, and motivations. By understanding the perspective of these athletes, you can better design data driven solutions and insights tailored to their unique challenges and goals.

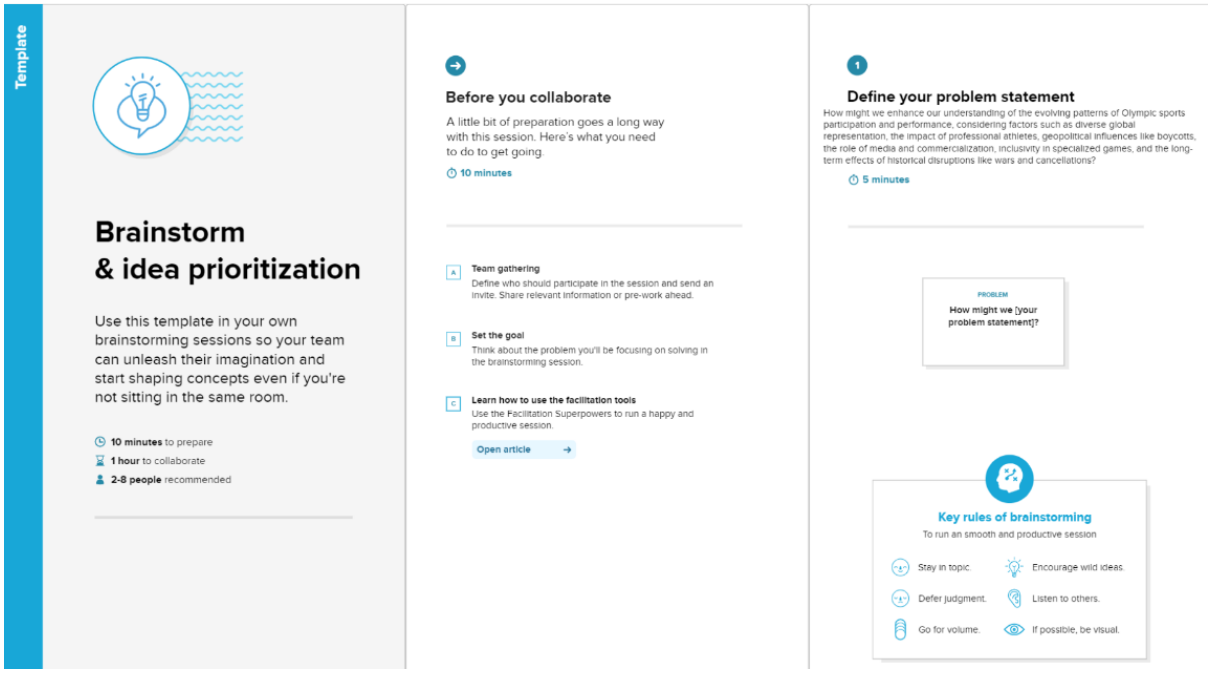


Empathy Map Link:
https://app.mural.co/t/ideation8747/m/ideation8747/1697009014908/c3060d037455682de5abd10a

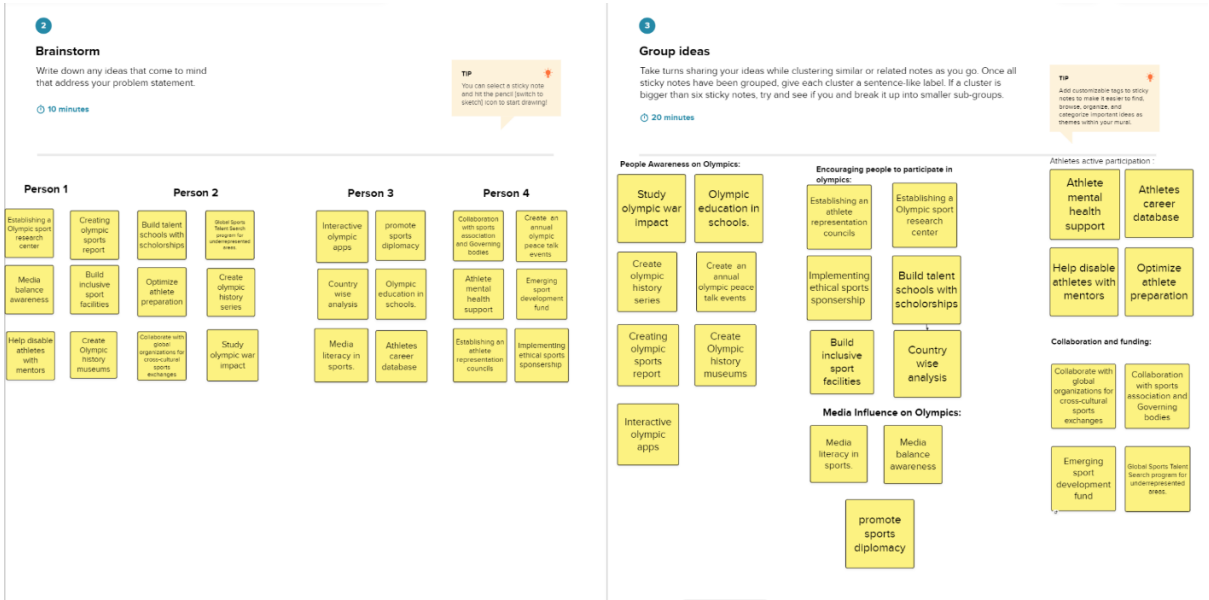
3.2 Ideation & Brainstorming:

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

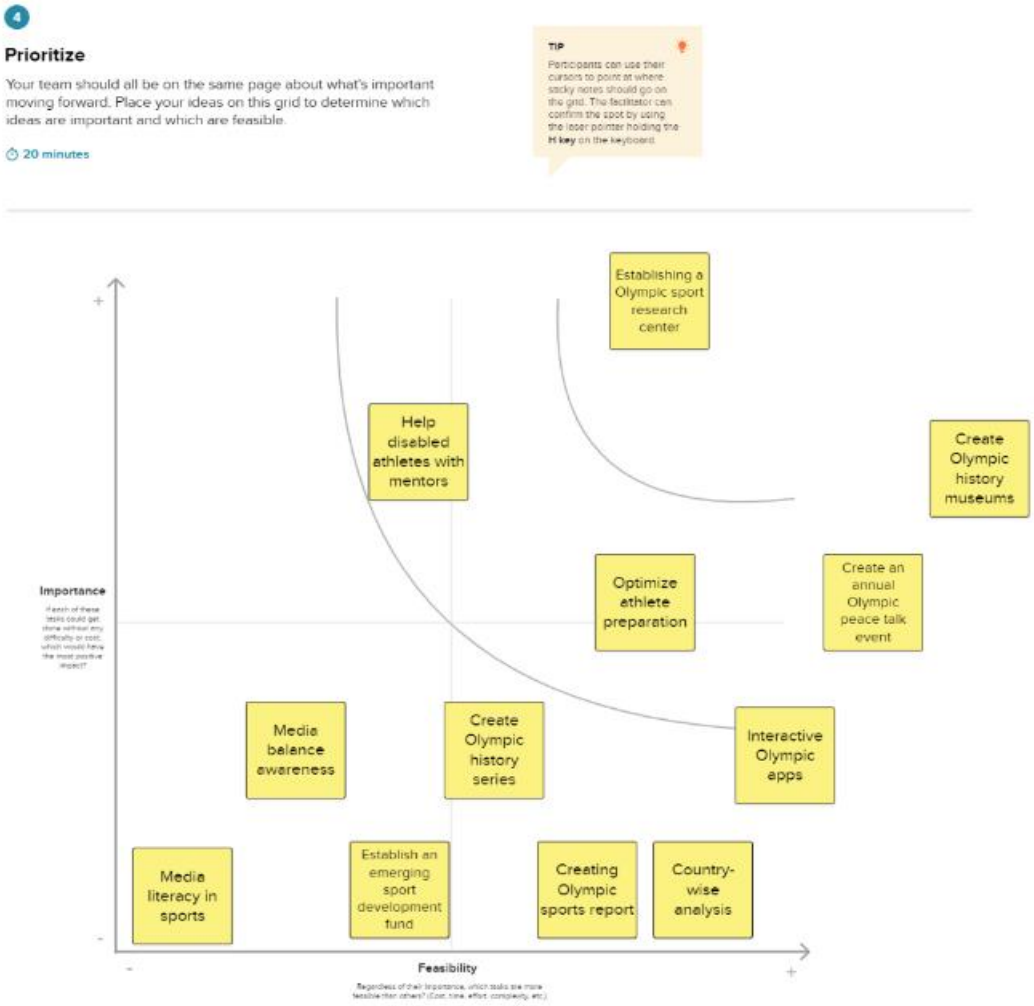
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization:



Template link :

https://app.mural.co/t/ideation8747/m/ideation8747/1697625690884/43e80f2e25f17263c40698781c3d55c0d7f52de5?sender=u57fa52d393bb8ccfb25a9343

4.Required Analysis :

Data collection: Gathering information from a data source relevant to artificial intelligence is the first prerequisite.

Data preparation and cleaning: To make sure the gathered data is appropriate for analysis, it needs to be processed and cleaned. This could entail purging unneeded information, fixing errors and missing values, and formatting the data so that it works with the analysis software.

Data analysis: To find important insights, the data must be analyzed. To better understand the data, this may entail applying methods like regression analysis, data visualization, and descriptive statistics.

4.1 Functional Requirements:

Functional requirements are those that directly specify what the system or project should do. In the context of this project:

Data Collection and Integration: The system should effectively gather, clean, and integrate data from various sources related to Olympic sports, such as athlete details, event information, and medal statistics.

Visualization and Analysis: It should provide a platform for robust data visualization, allowing for the creation of diverse charts, graphs, and visual representations that aid in understanding athlete performance, medal counts, and trends over time.

Dashboard Creation: The system must facilitate the creation of dashboards that present key insights, trends, and performance metrics in an accessible and easily interpretable manner.

Reporting and Documentation: Capability to generate reports and documentation summarizing the findings and insights from the data analysis.

4.2 Non-Functional Requirements:

Non-functional requirements address how the system should perform, rather than what it should do. In this context:

Performance: The system should deliver efficient response times, ensuring smooth data processing, visualization, and analysis even with large datasets.

User Interface: A user-friendly and intuitive interface for ease of use, ensuring stakeholders can interact with and understand the presented information easily.

Security: The platform should maintain data security and integrity, especially given the sensitive nature of athlete information and competition results.

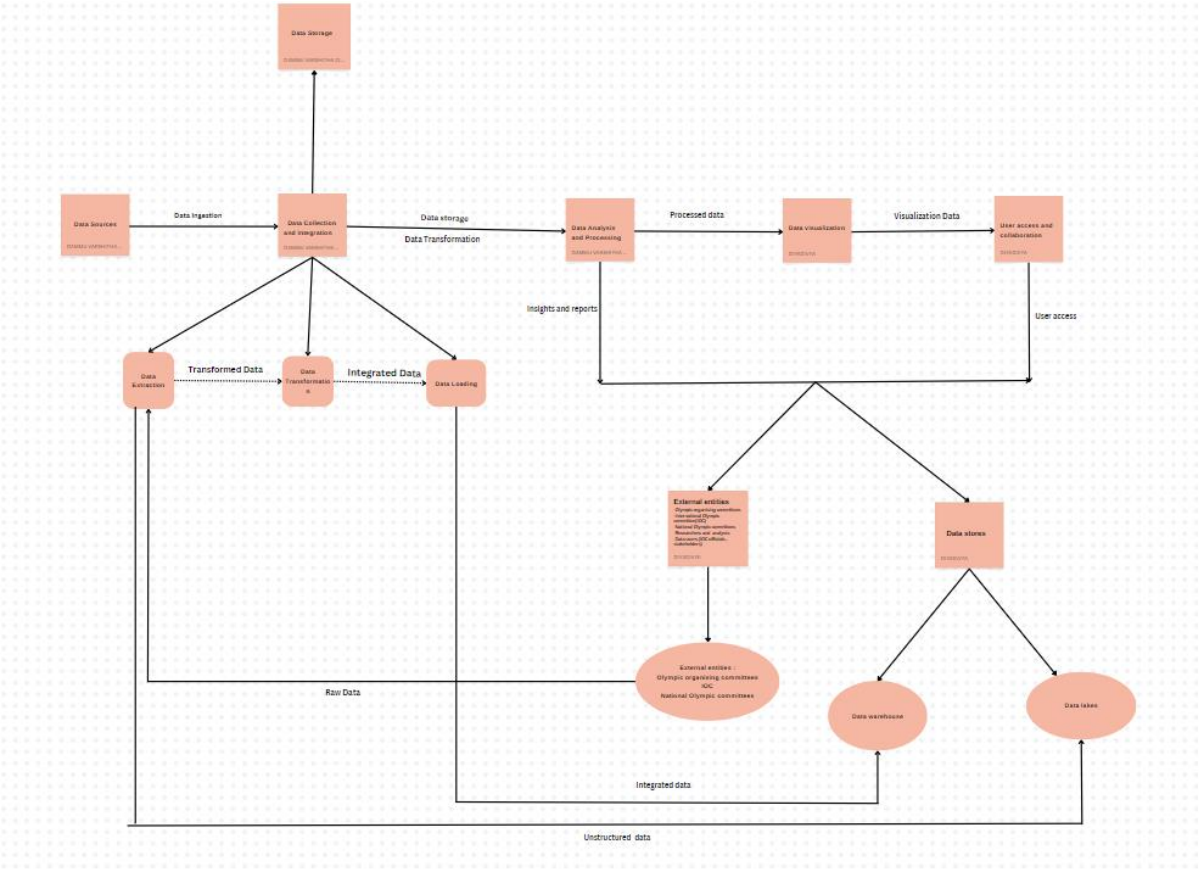
Scalability: The system should be scalable to accommodate potential increases in data volume and user base, ensuring it remains effective in the long term.

These requirements collectively aim to ensure the system's functionality, usability, performance, and security, aligning with the project's goals of data-driven insights in Olympic sports.

5.Project Design

5.1 Data Flow Diagrams :

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the rightamount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



https://www.canva.com/design/DAFy73EfAw/rllhHCtJOpWMaCnX1eTtAA/edit?utm_content=DAFy73EfA-w&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Here's an overview of the key components of the DFD :

Level-0:

Context Diagram This DFD level focuses on high-level system processes or functions and the data sources that flow to or from them. Level 0 diagrams are designed to be simple, straightforward overviews of a process or system.

Level-1:

This Level 1 Data Flow Diagram focuses on the data collection and integration process. The data is initially extracted from external entities, transformed and integrated, and then loaded into the data warehouse and data lakes. Some unstructured data is directed to the data lakes.

5.2 User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I want the data flow within the system to be efficient and responsive, ensuring quick access to the information I need.	I can access my account/dashboard	Low	Sprint-2
As an Olympic Organizing Committee member	Data sets	USN-2	I want to upload historical athlete and event data to the system so that it can be analyzed and used for future planning.	Ensure that all uploaded data compiles with data privacy regulations and guidelines.	High	Sprint-1
As a Researcher	Access to previous data	USN-3	I want to access athlete performance data from previous Olympic Games so that I can analyze trends and identify factors contributing to success.	The system should provide easy access to athlete performance data from past olympic games through a user friendly interface.	High	Sprint-1
As an Analyst	Access to real-time data websites	USN-4	I want to have access to real-time data feeds from ongoing Olympic events to provide up-to-date insights and reports.	The system should rovide continuous ,low latency access to eveny data,athlete updates, and results with clear documentation for data sources.	Medium	Sprint-1
As an IOC official	Dashboard	USN-5	I want to be able to view interactive visualizations and dashboards that display medal counts, athlete profiles, and historical performance data for decision-making.	User friendly , interactive dashboards for viewing medal counts, athlete profiles and historical performance data across devices.	High	Sprint-1
As a Data engineer	Warehouse access	USN-6	I want to ensure that data is collected, transformed, and loaded accurately into the data warehouse, with validation checks in place.	Ensure data collection , transformation, and loading with validation checks.	Medium	Sprint-1
As a project manager		USN-7	I want to ensure that the data flow diagram accurately represents the system's functionality, and that user requirements are met in the system design and development.	Confirm that the data flow diagram matces user requirements in system design and development.	High	Sprint 1
As a developver	APIs	USN-8	I want to create APIs for external systems to retrieve specific data, such as real-time event results or athlete profiles, for integration into their applications.		Medium	Sprint-1

5.3 Solution Architecture

Data Source and Collection:

* **Olympic Organizing Committees:** Historical data from previous Olympic Games, including participant information, results, and event details.

* **National Olympic Committees:** Data on athletes and their backgrounds, participation, and performance.

* ETL (Extract, Transform, Load) processes to collect and consolidate data from various sources.

Data Warehousing:

* **Data Warehouses:** Centralized storage for structured data, including athlete profiles, medal counts, and event histories.

* **Data Lakes**: Storage for unstructured and semi-structured data, such as media coverage, social media sentiment, and historical documents.

Data Analysis and Processing:

*** Data Analysis Tools:** Implement tools and frameworks like Python, R, and Jupyter notebooks for exploratory data analysis.

*** Big Data Technologies:** Use platforms like Hadoop and Spark for processing large volumes of data.

*** Machine Learning and AI:** Implement algorithms for predictive analytics, sentiment analysis, and performance modelling.

Data Visualization:

*** Dashboard Creation:** Use visualization tools like Tableau, Power BI, or custom web-based dashboards to present insights in a user-friendly manner.

*** Interactive Charts and Maps:** Display trends, patterns, and geographical insights for a comprehensive view of the data.

Data Encryption:

* Implement data encryption, access control, and data anonymization techniques to protect sensitive information.

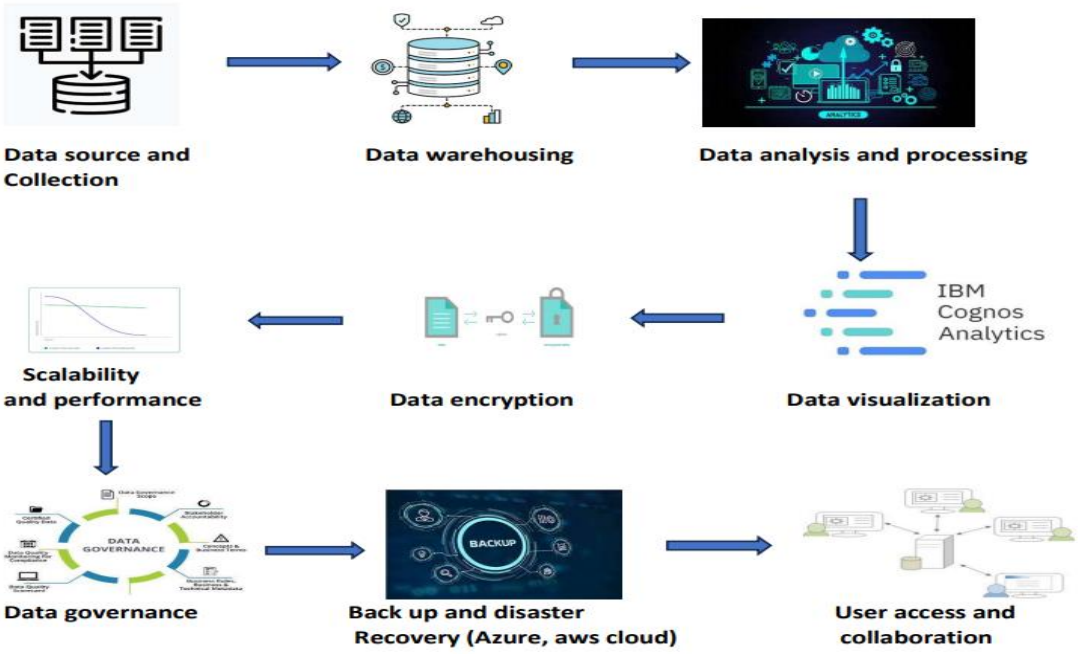
* Ensure compliance with data privacy regulations, especially when dealing with athlete profiles and personal data.

Scalability and Performance:

* Choose a scalable infrastructure that can handle data growth over time and during peak usage (e.g., during Olympic events).

* Implement load balancing and data replication for performance optimization.

Solution Architecture Diagram:

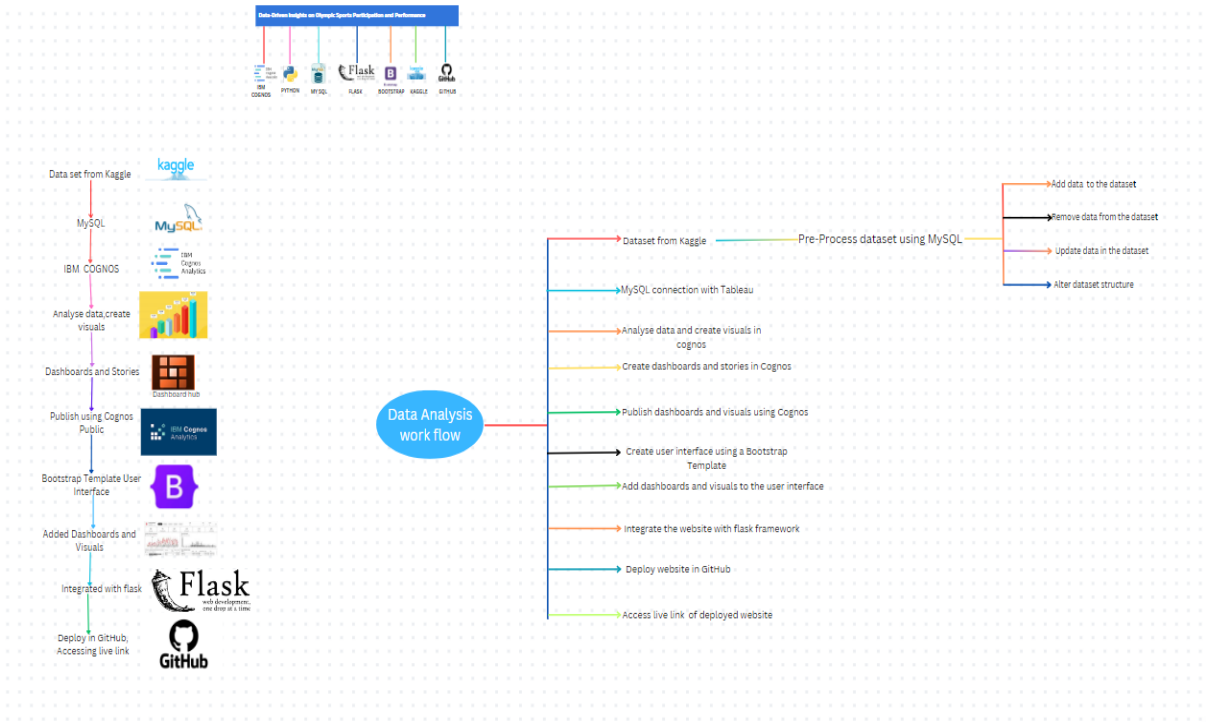


6. Projecct planning and scheduling :

6.1 Technical Architecture:

The technical architecture for this project encompasses a series of stages involving data collection, storage, preparation, visualization, and dissemination. It includes the use of Cognos Analytics for data handling and visualization. The architecture would involve processes such as:

S. No	Component	Description	Technology
1.	Dataset from Kaggle	Data source for analysis	Kaggle dataset or other data sources
2.	MySQL Database	Data storage, cleaning, preprocessing	MySQL database management
3.	MySQL Connection with Cognos	Connect MySQL with Cognos for data extraction	IBM Cognos Desktop forMySQL connection
4.	Data Analysis and Visualization (Cognos)	Analyse data, create visualizations, dashboards	IBM Cognos Desktop for dataanalysis and visualization
5.	Publishing with Cognos Public	Publish dashboards and stories	Cognos Public for sharing
6.	User Interface with Bootstrap	Create a user-friendly webinterface	Bootstrap framework forweb development
7.	Integration with Flask	Develop a web applicationfor hosting visuals	Flask for web applicationdevelopment
8.	Website Integration	Embed Cognos visuals into the website	HTML, CSS, and JavaScript for web development
9.	Version Control and Deployment (Git/GitHub)	Manage version control and deployment	Git for version control, GitHub for hosting
10.	Deployment and Hosting	Deploy and host the web application	GitHub Pages and Hostinger web hosting services



Wanna see my work :

https://www.canva.com/design/DAFzGYV0Wvs/ZR9altUZH7Ors75vTc2Ryw/edit?utm_content=DAFzGYV0Wvs&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

6.2 Sprint Planning & Estimation:

Sprint planning involves breaking down the project tasks into manageable units and estimating the effort required for each task. Sprint cycles could involve:

- **Data Collection & Preparation:** Estimate time for downloading, cleaning, and organizing the athlete event and NOC region data.

- **Data Visualization:** Allocating time for creating specific visualizations as outlined in the project milestones.

- **Dashboard and Story Development:** Planning for the creation and refinement of dashboards, stories, and reports.

6.3 Sprint Delivery Schedule:

The sprint delivery schedule outlines the timeline for each phase of the project, including specific milestones and expected completion dates for each task. For instance:

- Sprint 1 (Week 1-2): Data Collection & Preparation

- Sprint 2 (Week 3-4): Data Visualization

- Sprint 3 (Week 5-6): Dashboard and Story Development

- Sprint 4 (Week 7): Finalization, Testing, and Delivery

This structure allows for an agile approach to manage the project tasks, ensuring each phase is completed within a defined time frame and contributing to the overall project

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	6 Days	18-10-2023	24-10-2023	10	24-10-2023
Sprint-2	10	6 Days	23-10-2023	29-10-2023	10	01-11-2023
Sprint-3	5	7 Days	27-10-2023	02-11-2023	5	03-11-2023
Sprint-4	15	7 Days	06-11-2023	12-11-2023	15	12-11-2023
Sprint-5	15	7 Days	09-11-2023	15-11-2023	15	15-11-2023

Average Velocity Calculation:

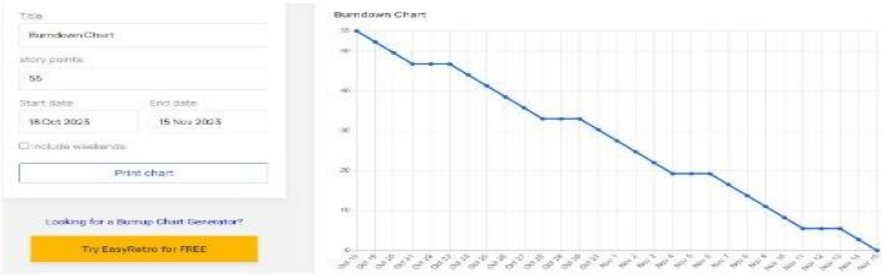
Average Velocity (AV) per iteration unit (story points per day) can be calculated by dividing the total story points completed by the total duration across all sprints.

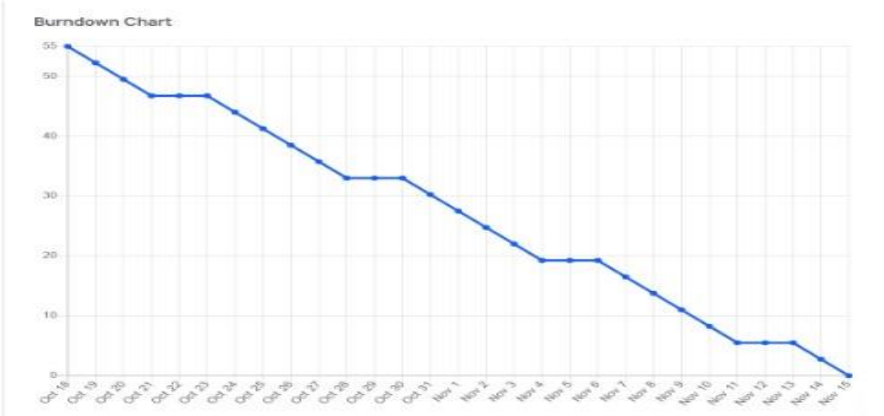
Total Story Points Completed: 10 + 10 + 5 + 15 + 15 = 55
Total Duration: 6 + 6+7+ 7 + 7 = 33 days

Average Velocity (AV) = Total Story Points Completed / Total Duration

Average Velocity (AV) = 55 / 33 ≈ 1.67 story points per day

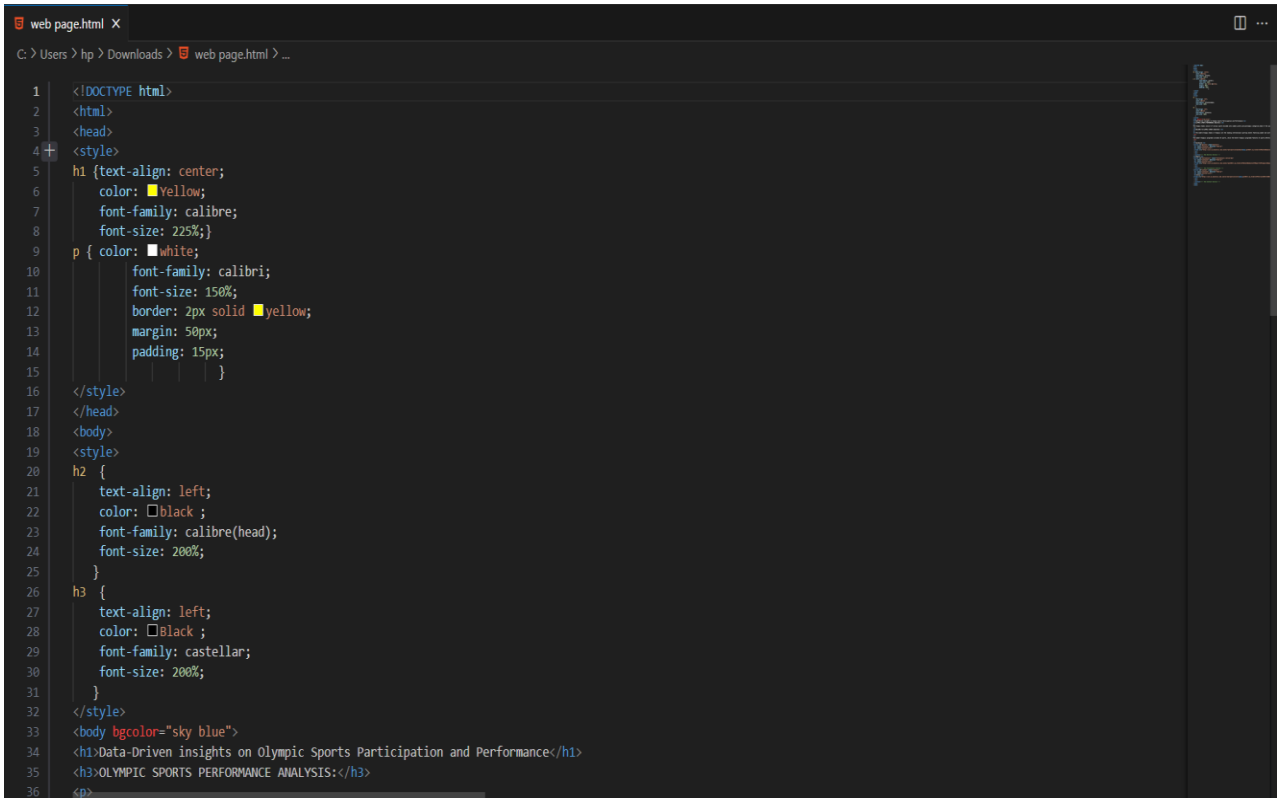
BURNDOWN CHART: A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.





7. Coding & Solutioning (Explain the features added in the project along with code)

7.1 Feature 1



Drive link :

https://drive.google.com/drive/folders/1DuJL7uZ8jXpbWZsv-XwmVdGauQQsSL8f?usp=sharing

8. PERFORMANCE TESTING

S.no	Parameter	Screenshots/Values
1.	Dashboard design	No .of Visualization/Graphs -12 Graphs,3 Dashboards
2.	Data responsiveness	

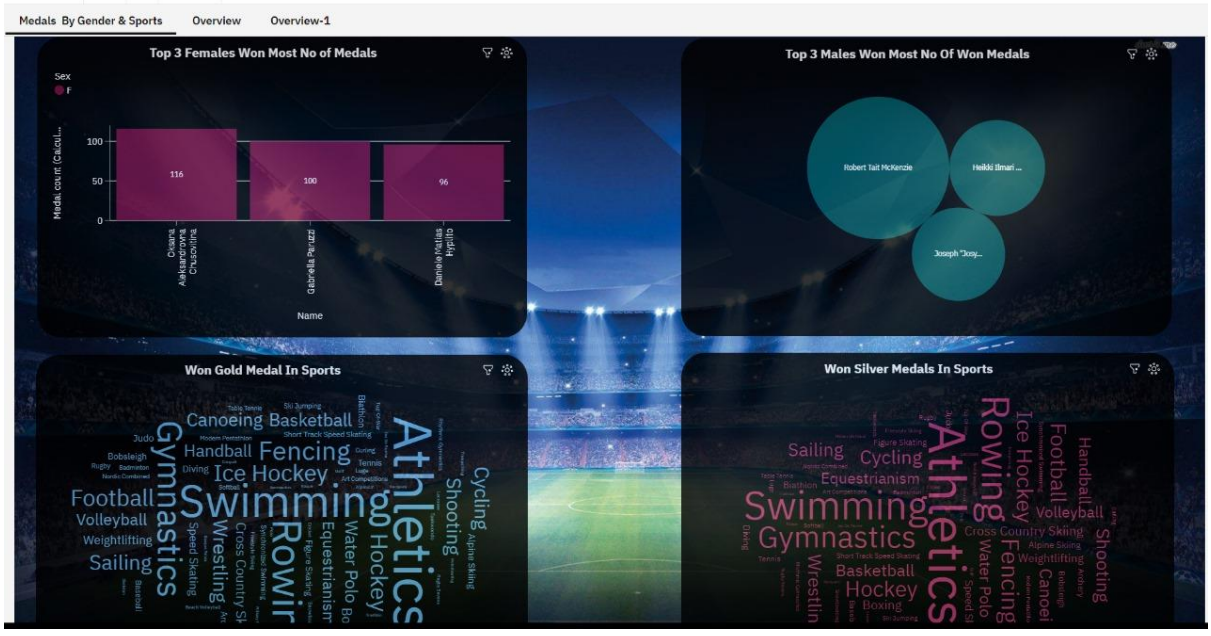
		<table><tr><th>Sprint</th><th>Total Story Points</th><th>Duration</th><th>Sprint Start Date</th><th>Sprint End Date (Planned)</th><th>Story Points Completed (as on Planned End Date)</th><th>Sprint Release Date (Actual)</th></tr><tr><td>Sprint-1</td><td>10</td><td>6 Days</td><td>18-10-2023</td><td>24-10-2023</td><td>10</td><td>24-10-2023</td></tr><tr><td>Sprint-2</td><td>10</td><td>6 Days</td><td>23-10-2023</td><td>29-10-2023</td><td>10</td><td>01-11-2023</td></tr><tr><td>Sprint-3</td><td>5</td><td>7 Days</td><td>27-10-2023</td><td>02-11-2023</td><td>5</td><td>03-11-2023</td></tr><tr><td>Sprint-4</td><td>15</td><td>7 Days</td><td>06-11-2023</td><td>12-11-2023</td><td>15</td><td>12-11-2023</td></tr><tr><td>Sprint-5</td><td>15</td><td>7 Days</td><td>09-11-2023</td><td>15-11-2023</td><td>15</td><td>15-11-2023</td></tr></table> Average Velocity Calculation: Average Velocity (AV) per iteration unit (story points per day) can be calculated by dividing the total story points completed by the total duration across all sprints. Total Story Points Completed: 10 + 10 + 5 + 15 + 15 = 55 Total Duration: 6 + 6+7+ 7 + 7 = 33 days Average Velocity (AV) = Total Story Points Completed / Total Duration Average Velocity (AV) = 55 / 33 ≈ 1.67 story points per day
3.	Utilization of Data Filters	Medals Count=Gold,Silver,Bronze
4.	Effective user story	No of Scenes Added -8
5.	Descriptive Reports	No of visualizations/Graphs -1 https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2FData%2Bmodules%2FReport%2FOlympic%2Bsport%2Breport&action=run&format=HTML&prompt=false

9. Results

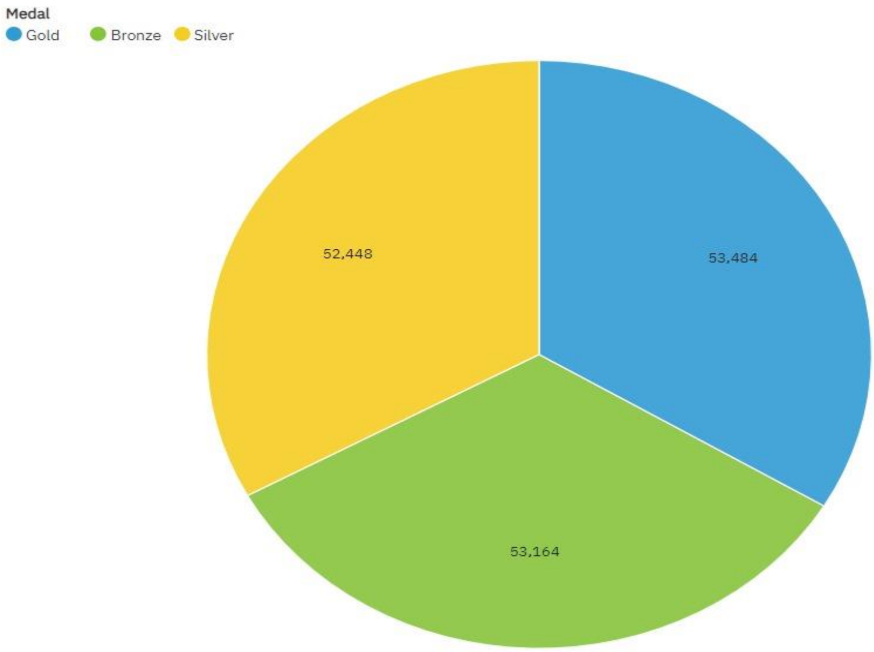
9.1 Output Screenshots

Dashboard :

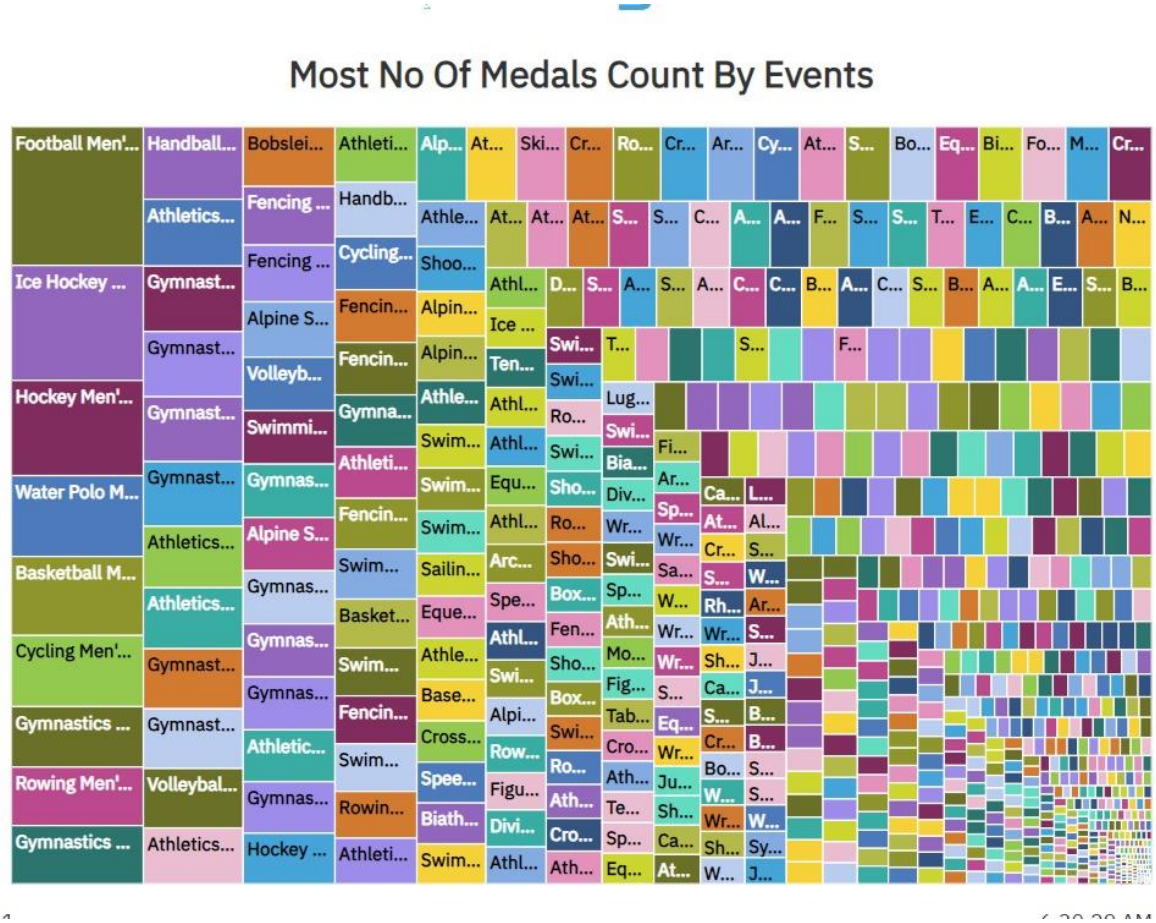
https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FData%2Bmodules%2FDashboard%2B1&action=view&mode=dashboard&subView=model0000018ba3256d71_000000000



Report :
https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2FData%2Bmodules%2FReport%2FOlympic%2Bsport%2Breport&action=run&format=HTML&prompt=false



Most No Of Gold Medals Won By Team



Story :

https://us3.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my_folders%2FStories%2Bfor%2BOlympics%2FOlympic%2BStory_1&action=view&sceneId=model0000018ba34d8ee5_00000001&sceneTime=0

10. Advantages & Disadvantages :

Advantages:

1. Insight Generation: The project enables the extraction of meaningful insights from a vast dataset, potentially offering valuable information for stakeholders in the Olympic sports domain.

2. Decision Support: The derived insights aid in decision-making processes for athlete development, strategic planning, and resource allocation within the Olympic sports arena.

3. Visual Representation: Using Cognos Analytics for visualization offers a user-friendly and visually intuitive representation of complex data, aiding in easy comprehension of trends and patterns.

4. Accessibility: By embedding dashboards and reports in a web interface, the project ensures broader accessibility and easier sharing of information with relevant stakeholders.

5. Structured Approach: The project follows a systematic methodology, from data collection to visualization, ensuring a structured and organized process.

Disadvantages:

1. Data Limitations: The insights derived heavily depend on the quality and depth of the available dataset. Incomplete or biased data might impact the accuracy of conclusions drawn.

2. Complexity: Handling a vast amount of data and utilizing tools like Cognos Analytics might pose a challenge in terms of complexity and resource requirements.

3. Subjectivity in Interpretation: Interpretation of the insights might be subjective and could lead to varying conclusions, impacting the accuracy of decision-making.

4.Maintenance and Scalability: As the project scales or evolves, maintaining and scaling the system might demand additional resources and expertise.

5. Security Concerns: Managing the security of sensitive athlete-related data and competition results is crucial and might pose challenges in terms of compliance and data protection.

11. Conclusion

In conclusion, the project to extract insights from Olympic sports data through a structured process of data collection, preparation, visualization, and dissemination holds significant promise for the domain of sports analytics. By leveraging Cognos Analytics and a methodical approach, the project aims to provide valuable insights for stakeholders in the Olympic sports arena.

The structured approach ensures a systematic handling of the extensive dataset, offering potential advantages in decision support, visual representation, and improved accessibility. The generation of insights from this data can significantly impact athlete development strategies, resource allocation, and overall strategic planning within the Olympic sports landscape.

However, the project faces challenges, including potential limitations in data quality, complexity in data analysis, subjectivity in interpretation, and the need for stringent data security measures.

Nonetheless, by embracing the advantages and addressing the challenges, this project is poised to contribute significantly to the understanding and enhancement of Olympic sports,

providing a foundation for informed decision-making and strategy development within the realm of international sports competitions

12. Future Scope

The project's completion opens avenues for several future endeavors within the realm of Olympic sports and data analytics:

1. Enhanced Insights: Further exploration of advanced analytics techniques, including machine learning and predictive modeling, can deepen the insights extracted from the dataset. These techniques can provide predictive analytics for athlete performance, aiding in talent identification and performance forecasting.

2. Longitudinal Studies: Conducting longitudinal studies by analyzing data from multiple Olympic Games can offer comprehensive insights into trends and changes over time, facilitating a deeper understanding of sports evolution.

3. Expanded Dataset Integration: Integrating additional datasets, such as training regimes, health records, and socio-economic factors, can provide a holistic view of an athlete's journey, further enriching the analysis.

4. Real-time Analytics: Developing real-time analytics capabilities during live Olympic events can offer immediate insights, aiding broadcasters, commentators, and sports analysts in providing engaging and informed coverage.

5. Stakeholder Collaboration: Collaborating with sports organizations, federations, and academic institutions to expand data collection and analysis efforts can create a more comprehensive and diverse dataset for broader insights.

6. User-Focused Interfaces: Designing interfaces tailored to specific user groups, such as coaches, analysts, or policymakers, can ensure the delivery of customized and actionable insights relevant to their needs.

7. Ethical Considerations & AI: Exploring the ethical implications of leveraging AI and data analytics in sports, ensuring fairness, privacy, and the ethical use of athlete-related data.

The future scope encompasses the continual evolution and refinement of data analytics processes in Olympic sports, fostering a deeper understanding of athletic performance and facilitating more informed decisions for athletes, sports organizations, and stakeholders.

13. APPENDIX (Source Code GitHub & Project Demo Link completion.)

Project demo link :

https://drive.google.com/drive/folders/1DuJL7uZ8jXpbWZsv-XwmVdGauQQsSL8f?usp=sharing