Tennis Leaderboard & Insights: Detailed Report

1. Workflow Overview

Objective

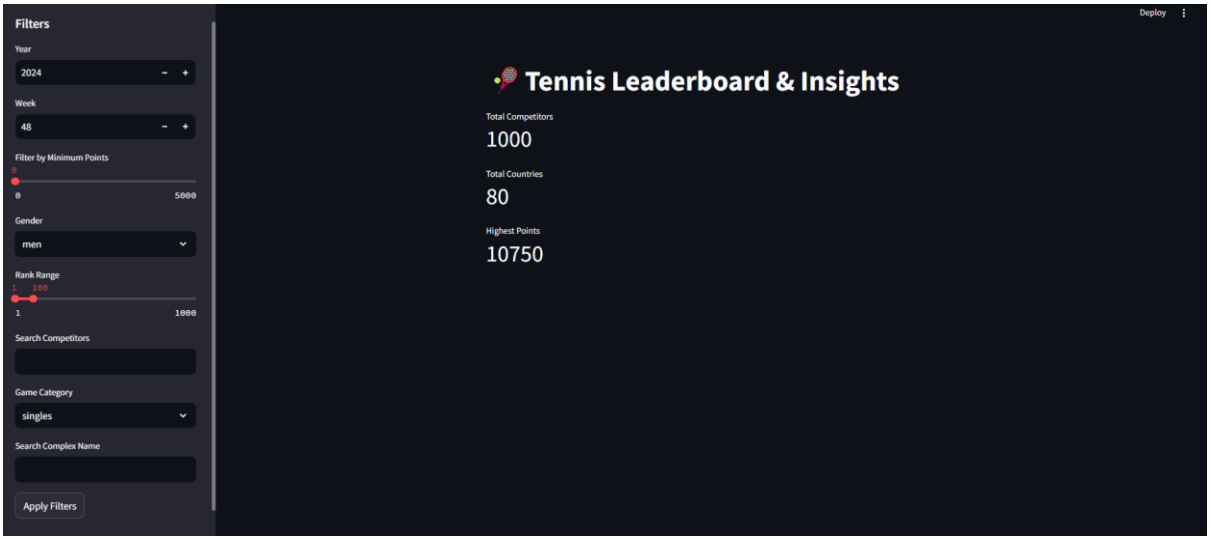
The Tennis leaderboard & insights Dashboard is designed to provide users with real-time insights into various sports metrics, including competitor rankings, tournament schedules, and venue popularity. The application allows users to filter data based on multiple criteria such as year, week, gender, ranking range, and competition category.

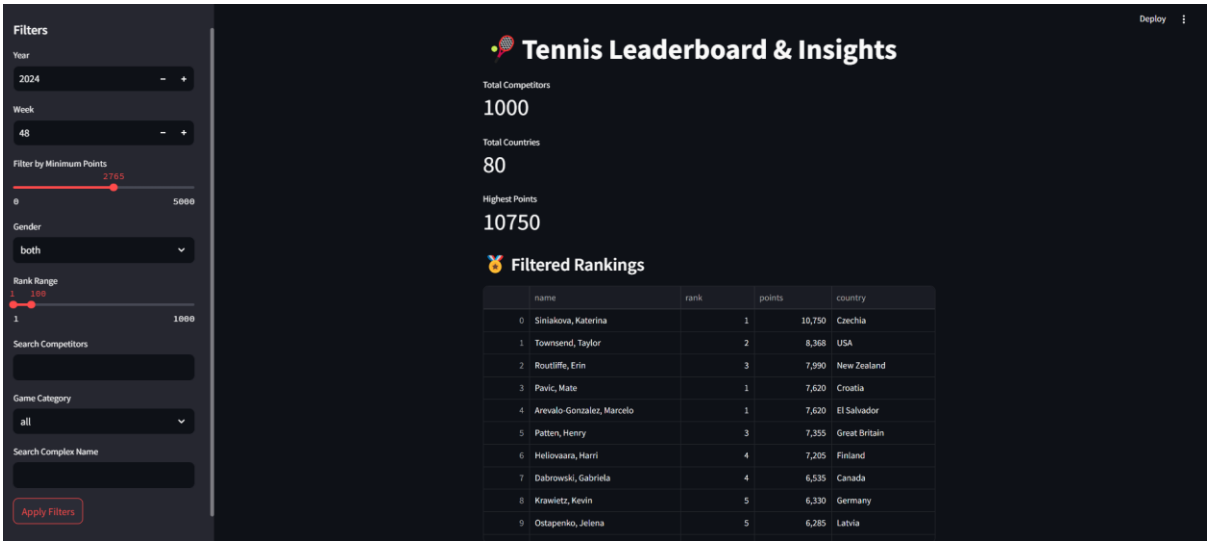
Technology Stack

- **Frontend:** Streamlit
- **Backend:** MySQL Database
- **Data Processing:** Pandas
- **Deployment:** Localhost

Data Flow

1. **User Interaction:** Users apply filters using the sidebar.
2. **Query Execution:** Based on the filters, SQL queries are dynamically constructed and executed against the MySQL database.
3. **Data Processing:** The fetched data is processed using Pandas to ensure it meets the UI requirements.
4. **Visualization & Presentation:** Processed data is displayed in tables for easy interpretation.





Filtered Rankings

	name	rank	points	country
0	Siniakova, Katerina	1	10,750	Czechia
1	Townsend, Taylor	2	8,368	USA
2	Routliffe, Erin	3	7,990	New Zealand
3	Pavic, Mate	1	7,620	Croatia
4	Arevalo-Gonzalez, Marcelo	1	7,620	El Salvador
5	Patten, Henry	3	7,355	Great Britain
6	Heliovaara, Harri	4	7,205	Finland
7	Dabrowski, Gabriela	4	6,535	Canada
8	Krawietz, Kevin	5	6,330	Germany
9	Ostapenko, Jelena	5	6,285	Latvia

Top 10 Competitors

	name	rank	points	country
0	Pavic, Mate	1	7,620	Croatia
1	Arevalo-Gonzalez, Marcelo	1	7,620	El Salvador
2	Siniakova, Katerina	1	10,750	Czechia
3	Townsend, Taylor	2	8,368	USA
4	Patten, Henry	3	7,355	Great Britain
5	Routliffe, Erin	3	7,990	New Zealand
6	Heliovaara, Harri	4	7,205	Finland
7	Dabrowski, Gabriela	4	6,535	Canada
8	Krawietz, Kevin	5	6,330	Germany
9	Ostapenko, Jelena	5	6,285	Latvia

🌎 Most Popular Venues by Countries

	country_name	total_venues
0	USA	455
1	France	236
2	Italy	236
3	China	207
4	Spain	195
5	Germany	129
6	Mexico	112
7	England	110
8	Brazil	100
9	Australia	100

📈 Biggest Ranking Movements

	name	rank	movement
0	Popyrin, Alexei	148	716
1	Struff, Jan-Lennard	392	-261
2	Bianchi, Juan Jose	427	259
3	Zink, Tyler	311	190
4	De Jong, Jesper	370	182
5	Somani, Parikshit	466	180
6	Kessler, McCartney	210	120
7	Droguet, Titouan	480	-118
8	Rodriguez, Victoria	278	117
9	Sanchez, Ana Sofia	277	115

📊 Most Popular Categories (By Competitions)

	category_name	total_competitions
0	ITF Men	2,198
1	ITF Women	2,032
2	Challenger	892
3	WTA	253
4	ATP	222
5	WTA 125K	162
6	UTR Men	50
7	UTR Women	46
8	Exhibition	32
9	Juniors	16

🗓️ Upcoming Competitions by Category

	competition_name	category_name	type	gender
0	2020 Trieste, Italy Doubles	Challenger	doubles	men
1	2022 ITF Australia F12, Men Doubles	ITF Men	doubles	men
2	2025 ATP Challenger Abidjan 2, Cote d Ivoire, Men Singles	Challenger	singles	men
3	Abidjan, Cote d Ivoire	Challenger	singles	men
4	Abidjan, Cote d Ivoire Men Doubles	Challenger	doubles	men
5	Algeria F2 Men Doubles	ITF Men	doubles	men
6	Algeria F2 Men Singles	ITF Men	singles	men
7	AO Asia-Pacific Wildcard Playoff Men Doubles	ATP	doubles	men
8	AO Asia-Pacific Wildcard Playoff Men Singles	ATP	singles	men
9	AO Asia-Pacific Wildcard Playoff Women Doubles	WTA	doubles	women

👨 👧 Number of Competitions by Gender

	gender	total_competitions
0	men	3,403
1	women	2,526
2	mixed	12

2. Schema Design

Tables & Relationships

1. Complexes Table

Column Name	Data Type	Constraints	Description
complex_id	VARCHAR(50)	PRIMARY KEY	Unique ID for the complex
complex_name	VARCHAR(100)	NOT NULL	Name of the sports complex

2. Venues Table

Column Name	Data Type	Constraints	Description
venue_id	VARCHAR(50)	PRIMARY KEY	Unique ID for the venue
venue_name	VARCHAR(100)	NOT NULL	Name of the venue
city_name	VARCHAR(100)	NOT NULL	Name of the city
country_name	VARCHAR(100)	NOT NULL	Name of the country
country_code	CHAR(3)	NOT NULL	ISO country code
timezone	VARCHAR(100)	NOT NULL	Timezone of the venue
complex_id	VARCHAR(50)	FOREIGN KEY (Complexes)	Links to the Complexes table

3. Categories Table

Column Name	Data Type	Constraints	Description
category_id	VARCHAR(50)	PRIMARY KEY	Unique ID for the category
category_name	VARCHAR(100)	NOT NULL	Name of the category

4. Competitions Table

Column Name	Data Type	Constraints	Description
competition_id	VARCHAR(50)	PRIMARY KEY	Unique ID for the competition
competition_name	VARCHAR(100)	NOT NULL	Name of the competition
parent_id	VARCHAR(50)	NULLABLE	Parent competition ID
type	VARCHAR(20)	NOT NULL	Type of competition (e.g., doubles)

Column Name	Data Type	Constraints	Description
gender	VARCHAR(10)	NOT NULL	Gender of participants (e.g., men)
category_id	VARCHAR(50)	FOREIGN KEY REFERENCES Categories(category_id)	Links to the category table

---

5. Competitors Table

Column Name	Data Type	Constraints	Description
competitor_id	VARCHAR(50)	PRIMARY KEY	Unique ID for the competitor
name	VARCHAR(100)	NOT NULL	Name of the competitor
country	VARCHAR(100)	NOT NULL	Country of the competitor
age	INT	NOT NULL	Age of the competitor

6. Competitors_Ranking Table

Column Name	Data Type	Constraints	Description
competitor_id	VARCHAR(50)	FOREIGN KEY REFERENCES Competitors(competitor_id)	Links to the Competitors table
rank	INT	NOT NULL	Current ranking of the competitor
points	INT	NOT NULL	Ranking points earned
movement	INT	NOT NULL	Change in ranking position
category	VARCHAR(50)	FOREIGN KEY REFERENCES Categories(category_id)	Competition category

---

3. Challenges & Solutions

Challenge 1: Database Connection Errors

- **Issue**: The MySQL connection was not always established successfully, causing an "always rendering" error.
- **Solution**: Implemented a **connection timeout** and used Streamlit's `@st.cache_resource` to cache the database connection.

Challenge 2: Debugging Errors

- **Issue**: It was difficult to debug errors directly in Streamlit.

- **Solution**: Added a **debugging log file** that records SQL queries and errors for analysis.

---

4. Insights & Key Takeaways

1. **Country-Based Venue Popularity:**
    - The most active sports venues are concentrated in countries with strong sporting infrastructure.
    - Filtering by country allows quick insights into where major tournaments are hosted.
2. **Competitor Trends:**
    - Analysis of ranking movements highlights emerging players and declining stars.
    - Gender-based filters provide insights into competition trends in men's vs. women's tournaments.
3. **Venue Optimization:**
    - Understanding venue popularity can help event organizers optimize scheduling and facilities management.
    - Certain cities repeatedly host high-profile tournaments, indicating strong local infrastructure and audience engagement.
4. **User Interaction & Performance:**
    - Introducing dynamic filters improves user engagement.
    - Optimizing query execution time ensures a smoother user experience.

---

5. Future Enhancements

1. **Integration with External APIs**
    - Fetch live match results and statistics from sports APIs.
    - Provide real-time updates on player rankings and tournament outcomes.
2. **Advanced Data Visualization**
    - Implement interactive charts for ranking trends.
    - Geospatial mapping of venues to visually explore location-based statistics.

By implementing these improvements, the dashboard can evolve into a comprehensive sports analytics tool, benefiting organizers, analysts, and enthusiasts alike.