

# Tennis Analytics: Game Data Analysis and Visualization

## 1. Objective

The **Tennis Analytics** project focuses on analyzing player and competition data from international tennis tournaments. The main objective is to build a **complete data pipeline** — from database design and SQL querying to visualization — providing valuable insights into tennis performance analytics.

This project establishes a structured **MySQL database** containing information about categories, competitions, venues, complexes, competitors, and their rankings. Analytical SQL queries are used to extract insights such as top players, venue distributions, and ranking trends.

The goal is to present a **data-driven view of tennis performance** that supports better understanding of player success and tournament organization.

## 2. Methodology

### ◆ 2.1 Database Creation

The project begins by creating a new database named **tennis\_analytics** in MySQL.

The database follows a **normalized relational schema** ensuring data consistency and efficient querying.

#### Tables Created:

1. **Categories** – Defines tennis competition categories like “ITF Men”, “ITF Women”, etc.
2. **Competitions** – Stores details about tournaments, their types, gender, and links to categories.
3. **Complexes** – Represents major sports complexes or stadium groups.
4. **Venues** – Lists locations where matches are played, along with city, country, and timezone.
5. **Competitors** – Contains information about players, their country, and abbreviations.
6. **Competitor\_Rankings** – Tracks rankings, movement, and total points for each competitor.

Each table is linked using **primary and foreign keys** for integrity.

### 3. Database Schema Design

Table	Primary Key	Foreign Keys	Description
Categories	category_id	—	Stores types of competitions
Competitions	competition_id	category_id → Categories	Details of tournaments
Complexes	complex_id	—	Contains sports complex information
Venues	venue_id	complex_id → Complexes	Venue details including country and timezone
Competitors	competitor_id	—	Player information
Competitor_Rankings	rank_id	competitor_id → Competitors	Ranking and points for each player

### 4. SQL Analysis and Queries

#### 4.1 Competitions Analysis

- **List all competitions with their category name:**

```
SELECT c.competition_name, cat.category_name
FROM Competitions c
JOIN Categories cat ON c.category_id = cat.category_id;
```

- **Count competitions in each category:**

```
SELECT category_id, COUNT(*) AS total_competitions
FROM Competitions
GROUP BY category_id;
```

- **List all competitions of type 'doubles':**

```
SELECT * FROM Competitions WHERE type = 'doubles';
```

- **Find competitions without a parent tournament:**

```
SELECT * FROM Competitions WHERE parent_id IS NULL;
```

- **Group competitions by category and type:**

```
SELECT category_id, type, COUNT(*) AS total
FROM Competitions
GROUP BY category_id, type;
```

#### 4.2 Venues Analysis

- **List venues with their complex names:**

```
SELECT v.venue_name, c.complex_name
FROM Venues v
JOIN Complexes c ON v.complex_id = c.complex_id;
```

- **Count venues by complex:**

```
SELECT complex_id, COUNT(*) AS total_venues
FROM Venues
GROUP BY complex_id;
```

- **Find venues in specific countries (e.g., Chile):**  

```
SELECT * FROM Venues WHERE country_name = 'Chile';
```
- **List all venue names with timezones:**  

```
SELECT venue_name, timezone FROM Venues;
```
- **Count venues by country:**  

```
SELECT country_name, COUNT(*) AS total_venues
FROM Venues
GROUP BY country_name;
```

#### 4.3 Competitor Rankings Analysis

- **Show competitor name, rank, and points:**  

```
SELECT c.name, r.rank, r.points
FROM Competitors c
JOIN Competitor_Rankings r ON c.competitor_id = r.competitor_id;
```
- **Top 5 ranked competitors:**  

```
SELECT * FROM Competitor_Rankings ORDER BY rank ASC LIMIT 5;
```
- **Players with no change in ranking:**  

```
SELECT * FROM Competitor_Rankings WHERE movement = 0;
```
- **Total points by country (e.g., Croatia):**  

```
SELECT c.country, SUM(r.points) AS total_points
FROM Competitors c
JOIN Competitor_Rankings r ON c.competitor_id = r.competitor_id
WHERE c.country = 'Croatia'
GROUP BY c.country;
```
- **Competitor with the highest points:**  

```
SELECT c.name, r.points
FROM Competitors c
JOIN Competitor_Rankings r ON c.competitor_id = r.competitor_id
ORDER BY r.points DESC LIMIT 1;
```

## 5. Workflow Summary

1. **Database Design:** Created six normalized tables for efficient relational data handling.
2. **Data Population:** Inserted real or sample data through Python scripts and CSV imports.
3. **Query Execution:** Ran analytical queries using MySQL Workbench to validate relationships and extract insights.
4. **Dashboard Integration (optional):** Designed a Streamlit front-end for interactive exploration of competitor and venue data.

## 6. Tools and Technologies Used

Component	Tool / Technology
Database	MySQL
Language	Python (for data insertion)
Libraries	pandas, mysql-connector
Visualization	Streamlit
Environment	Virtual Environment (.venv)
Version Control	Git + GitHub

## 7. Challenges Faced

- Ensuring **referential integrity** while inserting interdependent data (foreign keys).
- Managing **NULL parent relationships** in the Competitions table.
- Cleaning inconsistent **country codes and abbreviations** during data import.
- Handling **auto-increment conflicts** in the Competitor\_Rankings table.

## 8. Results & Insights

- Database successfully created with **six interrelated tables**.
- SQL queries efficiently retrieved insights such as:
  - Total competitors and their country distribution.
  - Category-wise and venue-wise tournament breakdowns.
  - Top 5 players based on ranking points.
- Enables clear understanding of global tennis performance trends.

## 9. Screenshots

1. Dashboard Overview – Showing all KPI metrics and navigation panels

The screenshot shows the Game Analytics Dashboard. On the left, there's a sidebar titled "Controls" with a "Navigate" dropdown set to "Dashboard" and a "Filter by Country (optional)" dropdown set to "All". The main area has a title "Game Analytics — Dashboard" with three summary metrics: "Total Competitors" (1004), "Countries Represented" (76), and "Highest Points" (8375). Below these are two tables. The first table, "Competitions by Category", lists categories and their total competitions. The second table, "Top 10 Competitors (by points)", lists players and their points.

category_name	total_competitions
ITF Men	2198
ITF Women	2032
Challenger	952
WTA	255
ATP	225
UTR Men	210
WTA 125K	206
UTR Women	175
Exhibition	32
Wheelchairs	16

2. Leaderboard Section – Top 10 players sorted by points and rank

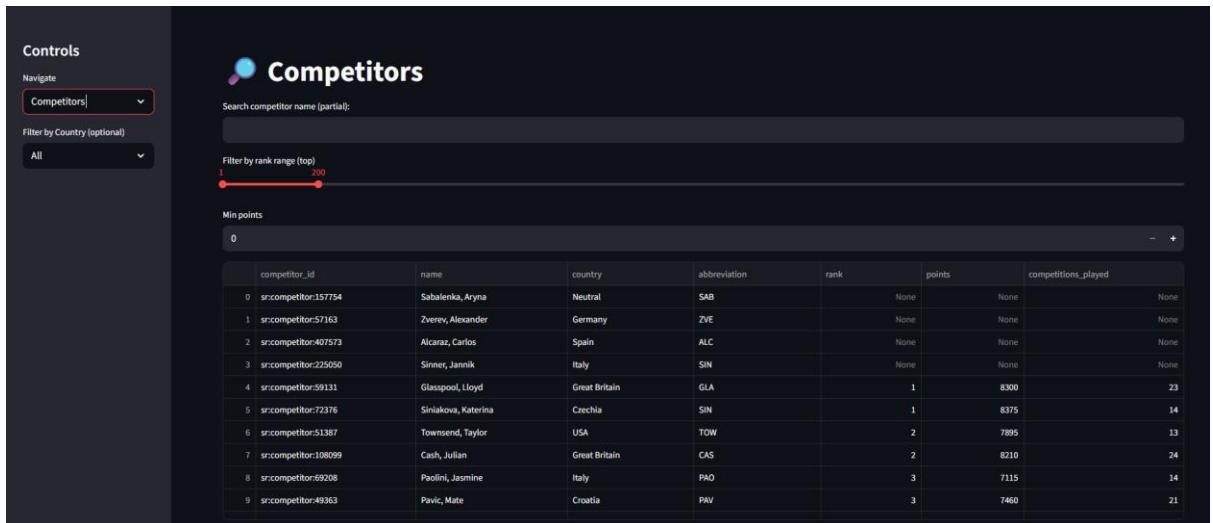
The screenshot shows the Leaderboards section. The sidebar is identical to the dashboard. The main area has a title "Leaderboards" and a subtitle "Top Ranked (by rank)". It displays a table of top-ranked players with columns for name, rank, and points.

name	rank	points
Glasspool, Lloyd	1	8300
Siniakova, Katerina	1	8375
Townsend, Taylor	2	7895
Cash, Julian	2	8210
Arevalo-Gonzalez, Marcelo	3	7460
Paolini, Jasmine	3	7115
Errani, Sara	3	7115
Pavic, Mate	3	7460
Patten, Henry	5	7280
Routliffe, Erin	5	6250

The screenshot shows the "Top 10 Competitors (by points)" section. The sidebar is identical. The main area displays a table of top competitors with columns for name, rank, and points.

name	rank	points
Siniakova, Katerina	1	8375
Glasspool, Lloyd	1	8300
Cash, Julian	2	8210
Townsend, Taylor	2	7895
Pavic, Mate	3	7460
Arevalo-Gonzalez, Marcelo	3	7460
Patten, Henry	5	7280
Heliövaara, Harri	5	7280
Errani, Sara	3	7115
Paolini, Jasmine	3	7115

3. Competitor Search Feature – Screenshot of user searching for a player and viewing detailed information



## 10. Future Enhancements

- Integrate real-time data from APIs (e.g., Sportradar).
- Extend dashboard for **live ranking updates**.
- Apply **machine learning models** for performance prediction.
- Deploy dashboard on **Streamlit Cloud** for public access.

## 11. Conclusion

The **Tennis Analytics Project** demonstrates how **data engineering and data analytics** can be integrated into a single system to generate meaningful insights. From extracting and cleaning real-time sports data to designing a robust database and visualizing results through a modern dashboard, the workflow showcases the **complete lifecycle of data**.

The platform can be further extended by adding:

- Match-level performance analytics
- Time-series trend visualizations
- Player comparison tools
- Predictive modeling using machine learning for ranking forecasts

Overall, the Tennis Game Analytics project provides a strong foundation for **sports data intelligence**, bridging technology, data, and decision-making in a unified solution.

## 12. GitHub Repository

All source code, SQL scripts, and documentation are available at:

[https://github.com/Omkar2807/Tennis-Game-Analytics/tree/main/Game\\_analytics](https://github.com/Omkar2807/Tennis-Game-Analytics/tree/main/Game_analytics)