



## Tennis Professionals Dataset

# ATP Tennis Professionals Dataset

<https://www.atptour.com/en>

Data collected from <https://www.atptour.com/en> regarding all single matches of 10.361 male players (top 500 players who played between and 28/03/1973 and 14/02/2022)

File *sitesplayerAll.txt*: links to all players data

File *atpplayers.json*: json data extracted (mongoexport --db atp --collection atpplayers --out atpplayers.json)

## JSON Data

```
"_id" : ObjectId("624ab34913b144c54b3c9acd"),
"PlayerName" : "Novak Djokovic",
"Born" : "Belgrade, Serbia",
"Height" : 188,
"Hand" : "Right-Handed, Two-Handed Backhand",
"LinkPlayer" : "https://www.atptour.com/en/players/novak-djo

" Tournament" : "Tokyo Olympics",
"Location" : "Tokyo, Japan",
"Date" : "2021.07.26 - 2021.08.01",
"Ground" : "Hard",
"Prize" : "",
"GameRound" : "Semi-Finals",
"GameRank" : 5,
"Oponent" : "Alexander Zverev",
"WL" : "L",
"Score" : "61 36 16"
```

Useful link: <https://datahub.io/core/country-list#data>



## Tennis Professionals Dataset

**Goal:** to build a relational database in Mysql, ensuring all the integrity restrictions and present the result of the following queries:

- 1) For each country (only one select)
  - a) Number of players
  - b) Number of tournaments
  - c) Number of rounds

CountryName	TotalPlayers	TotalTournaments	TotalGames
Afghanistan	0	0	0
Zambia	1	0	0
Zimbabwe	11	41	666

- 2) List of top 10 players name with their % of winning games, order by %
- 3) List of top 10 left-handed players name with their % of winning games in Grand Slam games
- 4) List of top 5 players name with their winning games in Hard Ground

### Additional Information

All data cleaning and preparation must be done in mongo DB. MySQL is only used to execute the Select commands regarding the four queries.

### Deliverables (in a zip file)

1. pdf file explaining all the steps, easy to read (should use graphics), highlighting the decisions, lost rows, etc. Don't forget the relational diagram with the tables and keys that connect them (use phpMyAdmin designer).
2. txt file with the mongoDB complete script (ready to copy & paste & test against the original file atpplayers.json). The script must generate one collection for each table to be exported/created to Mysql. Also, the script may not have any text in between, i.e., **if we are not able to run it directly without any change, it will not be accepted.**
3. txt file with the four select commands and output (with query duration time).
4. SQL script file (dump file, i.e., all SQL instructions to create and populate the database) with the MySQL database (include the create database command) in order to test the four select commands. Like for the MongoDB commands, **if we are not able to run it directly without any change, it will not be accepted.**

*Note: The mongo script must assume that the file countries\_net.csv is located in c:/data/atp/countries\_net.csv. the name of the file is the concatenation of the students names (first\_last). Example AnnaFerreira\_LuísCastro.zip.*