

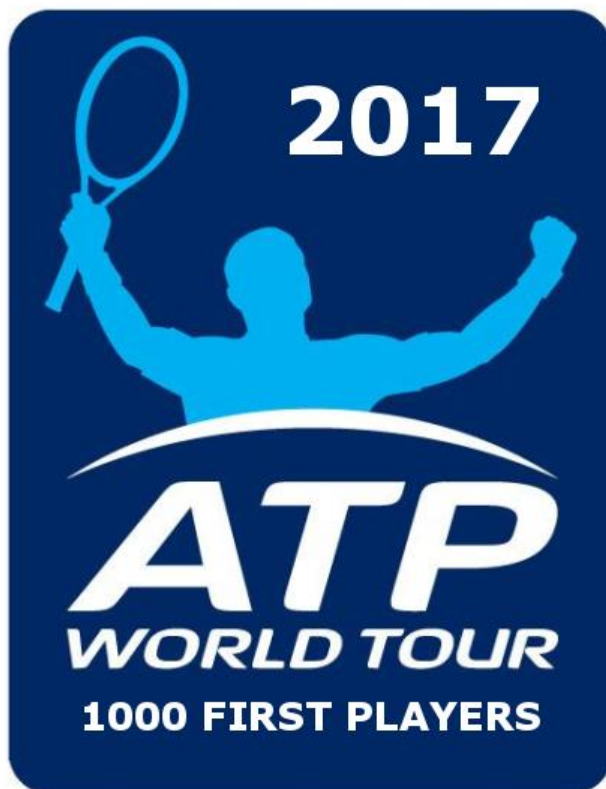
### 1. Títol:

Primers 1000 jugadors ATP 2017

### 2. Subtítol:

Llista dels primers 1000 jugadors segons els rànking individual de l'ATP, incloent els resultats esportius globals de la temporada així com els guanys obtinguts de la participació en tornejos ATP l'any 2017.

### 3. Imatge:



### 4. Context:

Aquest conjunt de dades s'ha obtingut de la pàgina web oficial de l'**Associació de Tennis Professional (ATP)** com a resultat de l'elaboració d'una pràctica de web scraping de l'assignatura Tipologia i cicle de vida de les dades, del Màster Universitari en Ciència de Dades de la Universitat Oberta de Catalunya.

El present conjunt de dades recollit conté la llista dels primers 1000 jugadors de l'ATP segons el rànking individual, amb els resultats obtinguts aquesta temporada 2017, així com els guanys provinents dels premis dels tornejos on han participat els jugadors.

## 5. Contingut:

El fitxer conté 1000 registres que corresponen als 1000 primers jugadors del rànking ATP a data 6 de novembre de 2017.

En el conjunt de dades hi podem trobar la següent informació:

RANKING: número de rànking ATP del jugador.

PAIS: abreviatura del país del jugador.

JUGADOR: nom del jugador.

EDAT: edat del jugador.

PUNTS: punts ATP del jugador.

TORNEJOS: tornejos jugats aquesta temporada.

WL: partits guanyats (W)- partits perduts (L) de categoria ATP World Tour, aquesta temporada pel jugador .

TITOLS: tornejos guanyats aquesta temporada en categoria ATP World Tour.

RANKING\_DOB: el rànking de la categoria de dobles del jugador.

GUANYES\_IND: guanyes en dollars aconseguits en tornejos individuals aquesta temporada.

GUANYES\_DOB: guanyes en dollars aconseguits en tornejos de dobles aquesta temporada.

GUANYES\_TOT: suma de guanyes individuals i dobles per jugador.

A tenir en compte per entendre les dades:

TORNEJOS: aquest camp inclou la participació del jugador en tots els tornejos puntuables pel rànking de l'ATP, no només els tornejos de màxim nivell (ATP world tour).

WL: aquest camp descriu els partits jugats en categoria individual en tornejos ATP World Tour, la màxima categoria del tennis. Per això, a mesura que anem baixant en el rànking, veiem jugadors que no han jugat cap partit. Aquesta informació significa que no han jugat cap partit a la màxima categoria. Per això no és incompatible tenir aquesta informació a zero amb haver jugat tornejos o tenir guanyes, ja que seran partits o guanyes de tornejos menors. Aquests tornejos menors també donen punts a la classificació de l'ATP, que és una classificació centralitzada, on com més puges, a més tornejos importants pots anar, o sigui, que donen més punts i diners.

TITOLS: de manera semblant a l'anterior, aquest camp reflecteix els títols a categoria ATP World Tour en categoria individual.

RANKING\_DOB: quan trobem un zero en el rànking de dobles significa que el jugador corresponent no ha jugat en aquesta categoria durant aquesta temporada.

#### Període de temps de les dades:

Aquestes dades tindran validesa a nivell esportiu fins que acabi la temporada 2018, al novembre de l'any que ve, quan s'haurien d'actualitzar amb les dades de la nova temporada. Tot i això, si es vol fer referència als guanys dels tennistes i a les relacions que se'n puguin derivar a partir d'aquesta avaluació econòmica més enllà dels resultats esportius concrets, llavors aquestes dades tindran vigència fins que els premis que atorguin els tornejos professionals de tennis canviïn substancialment.

#### Recull de dades:

Les dades s'han recollit a través del programari R i, concretament, fent servir les llibreries `rvest` i `xml2`.

## 6. Agraïments:

Agrair a l'Associació de Tennistes Professionals (ATP), perquè he pogut fer servir les seves dades publicades a la seva web per poder fer aquesta pràctica. També agrair al professorat de la UOC i concretament al del Màster Universitari en Ciència de Dades, pel suport que ens donen. I, com sempre i per acabar, moltes gràcies a la meua família per fer possible que pugui realitzar aquests treballs a través de la seva paciència, comprensió i sacrifici de temps lliure.

## 7. Inspiració:

El motiu per buscar aquest conjunt de dades és la curiositat per saber fins a quin punt els tennistes professionals es guanyen la vida més enllà dels tennistes famosos de màxim nivell. Fàcilment en podrien sorgir uns intervals, entre els quals tindríem tennistes milionaris, tennistes que es guanyen molt bé la vida, sous normals, etc., fins al punt de poder saber a quin nivell mínim s'hauria d'arribar perquè valgués la pena fer l'esforç d'intentar ser esportista d'elit, assumint el que això podria suposar, com haver de renunciar als estudis, per exemple. Aquesta balança entre nivell esportiu, ingressos i renúncies és el que penso que pot tenir interès de cara a desenvolupar-ne un estudi. La curiositat inicial, així, anava més encaminada a veure quin rendiment extreuen del tennis aquells esportistes de nivell mitjà, dins de la professionalitat, que no pas a saber el milionari que són els millors tennistes del món.

## 8. Llicència:

No deixa de ser un dataset formatiu, on l'objectiu estava més en l'aprenentatge de regles de web scraping que no pas en el valor propi de les dades obtingudes. Per aquest mateix motiu he orientat la temàtica del dataset a satisfer curiositats i no a buscar una rellevància del resultat final. Per tant, la llicència, d'acord amb aquest plantejament, crec que hauria de ser totalment lliure, renunciant a qualsevol interès d'autor ni de reconeixement. Per tant, descartaria tant les llicències d'atribució (BY), com les llicències de compartir de la mateixa manera (SA). Així, la llicència que escolliria seria:

**Released Under CC0: Public Domain License**