

Tennis Match Prediction

The data

The website <http://tennis-data.co.uk/alldata.php> gathers outcomes of both WTA (*Women Tennis Association*) and ATP (*Association of Tennis Professionals - men only*) tennis games over several years.

A short description of each variable can be found here :

<http://www.tennis-data.co.uk/notes.txt>

What is expected from you

First, we would like you to answer the questions on page 2.

Then, we want you to predict the outcome for each ATP tennis game during the year 2017.

To do so, we would like you to send us a self-contained Python notebook that handles:

- the data preparation process
- the modeling phase
- the evaluation of your model(s)
- any observation or explanation you deem interesting

Indications

How much time you spend on this problem is entirely up to you (but it should not take you more than 4 hours). Feel free to answer in French or in English.

This test comes with a Jupyter Notebook “starter-kit.ipynb”. Executing all the cells should provide you with :

- one pandas dataframe for all ATP matches (**df_atp**)
- one pandas dataframe for all WTA matches (**df_wta**)

Bonus points if :

- your notebook contains graphics that are both interesting and pretty
- we can go through your entire notebook without frowning
- you teach us something cool

Questions

Please answer the following questions about the dataset with the appropriate line(s) of code.

Example

Question ⇒ *How many ATP matches are there in the dataset ?*

Answer ⇒ `len(df_atp)`

1. Who are the three ATP players with the most wins ?
2. How many sets did the player "Federer R." win in total ?
3. How many sets did the player "Federer R." win during the years 2016 and 2017 ?
4. For each match, what is the percentage of victories of the winner in the past ?