

ATP Matches (2015-2016-2017) Classification MVP

➤ Goal the project:

The goal of this project is to better understand who's the player has been steadily improving over the 3 years and looking at How does height, age and player hand effect his level?

➤ The Data:

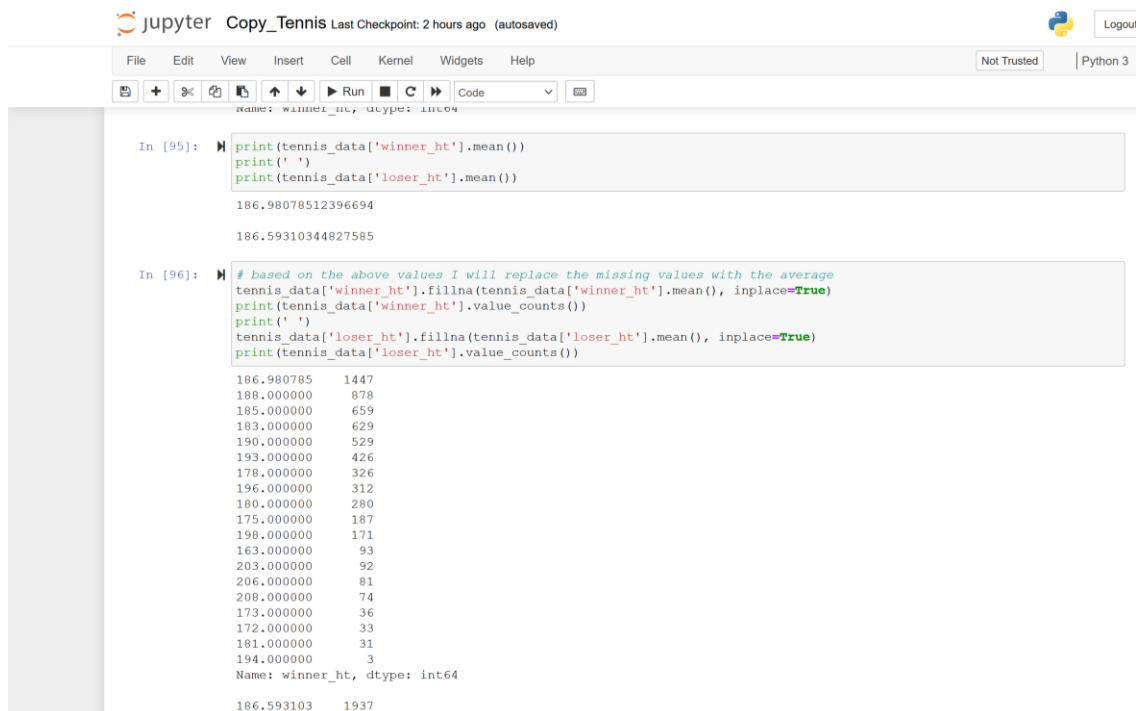
A dataset of ATP matches including individual statistics. These datasets are csv file for ATP tournament from 2015 to 2017.

List of all columns:

```
tourney_name (Tournament Name)
surface (Types of tennis courts - grass, clay and hard)
draw_size (The size of each playoff draw is dependent on the number of players)
tourney_level (Level tournaments were rated as A, B or C ....)
tourney_date (Tournaments Date)
match_num (Match Number)
winner_id (Winner ID)
winner_name (Winner Name)
winner_hand (Winner Hand "RIGHT - Left - Both")
winner_ht (Winner height)
winner_ioc (Winner Nationality)
winner_age (Winner Age)
winner_rank (winner Ranking)
winner_rank_points (Ranking Point)
loser_id (loser ID)
loser_name (loser name)
loser_hand (loser Hand "RIGHT - Left - Both")
loser_ht (loser height)
loser_ioc (loser Nationality)
loser_age (Loser Age)
loser_rank_points (Ranking Point)
score (Match points )
round (128- 64- 32- 16-Quarter finals- Semi finals - Final)
```

➤ The MVP notebook summarizes the following:

- Explore The Features
- Clean The Data
- Identify Null Values in Each Columns



```
jupyter Copy_Tennis Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

name: winner_ht, dtype: int64

In [95]: print(tennis_data['winner_ht'].mean())
print(' ')
print(tennis_data['loser_ht'].mean())

186.98078512396694

186.59310344827585

In [96]: # based on the above values I will replace the missing values with the average
tennis_data['winner_ht'].fillna(tennis_data['winner_ht'].mean(), inplace=True)
print(tennis_data['winner_ht'].value_counts())
print(' ')
tennis_data['loser_ht'].fillna(tennis_data['loser_ht'].mean(), inplace=True)
print(tennis_data['loser_ht'].value_counts())

186.980785    1447
188.000000     878
185.000000     659
183.000000     629
190.000000     529
193.000000     426
178.000000     326
196.000000     312
180.000000     280
175.000000     187
198.000000     171
163.000000     93
203.000000     92
206.000000     81
208.000000     74
173.000000     36
172.000000     33
181.000000     31
194.000000      3
Name: winner_ht, dtype: int64

186.593103    1937
```

➤ Future Work:

- Complete the cleaning.
- Working on the model.
- Get the results.
- Visualization the result.