Dataset presentation - Tennis prediction

Hi! PARIS DataBootcamp 2024

1. About the dataset

The `tennis_matches.csv` contains 28000 tennis player match results from 2018 to 2022.

The dataset contains characteristics for each player (height, age, ranking, hand…) as well as match statistics (number of aces, who won the first set, whether the player won…).

2. About the variables

Name	Description	Values (if categorical)
Hand	Hand the player plays with	R: Right, L: Left
Age	Age of the player	
Ace	Number of aces made by the player	
df	Number of double faults by the player	
svpt	Total serve points	
1stIn	Percent of first serves of the player that were in	
1stWon	Percent of exchanges won on a first serve by the player	
2ndWon	Percent of exchanges won on a second serve by the player	
SvGms	Total number of service games a player has played during a match.	
bpSaved	Number of breakpoints saved by the player	
bpFaced	Number of breakpoints faced by the player	
rank	Rank of the player	
rank_point	Rank of the player in terms of points	
won	Whether the player won or not the match	0: Lost, 1: Won
surface	Surface in which the match is played	
draw_size	Number of players or teams participating in the match's tournament	
tourney_level	Tournament level	G: Grand Slams M: Masters 1000s A: Other tour-level events C: Challengers S: Satellites/ITFs F: Tour finals and other season-ending events D: Davis Cup
tourney_date	Year of the tournament	
Best_of	Maximum number of sets played	
round	Round in the tournament a match belongs to	
minutes	Duration of the match in minutes	