

Data Description

1. The data set we've selected comes from <http://www.tennis-data.co.uk/alldata.php>. We selected the women's 2023 WTA tour tennis data because we are interested in performing cluster analysis. We want to know more about the relationship between stats of players and their corresponding betting odds.
2. There are 32 variables and 480 observations.
3. Description of all variables:

WTA = Tournament number (women)

Location = Venue of tournament

Tournament = Name of tournament (including sponsor if relevant)

Data = Date of match (note: prior to 2003 the date shown for all matches played in a single tournament is the start date)

Series = Name of ATP tennis series (Grand Slam, Masters, International or International Gold)

Tier = Tier (tournament ranking) of WTA tennis series.

Court = Type of court (outdoors or indoors)

Surface = Type of surface (clay, hard, carpet or grass)

Round = Round of match

Best of = Maximum number of sets playable in match

Winner = Match winner

Loser = Match loser

WRank = ATP Entry ranking of the match winner as of the start of the tournament

LRank = ATP Entry ranking of the match loser as of the start of the tournament

WPts = ATP Entry points of the match winner as of the start of the tournament

LPts = ATP Entry points of the match loser as of the start of the tournament

W1 = Number of games won in 1st set by match winner

L1 = Number of games won in 1st set by match loser

W2 = Number of games won in 2nd set by match winner

L2 = Number of games won in 2nd set by match loser

W3 = Number of games won in 3rd set by match winner

L3 = Number of games won in 3rd set by match loser

W4 = Number of games won in 4th set by match winner

L4 = Number of games won in 4th set by match loser

W5 = Number of games won in 5th set by match winner

L5 = Number of games won in 5th set by match loser

Wsets = Number of sets won by match winner

Lsets = Number of sets won by match loser

Comment = Comment on the match (Completed, won through retirement of loser, or via Walkover)

Key to match betting odds data:

B365W = Bet365 odds of match winner

B365L = Bet365 odds of match loser
B&WW = Bet&Win odds of match winner
B&WL = Bet&Win odds of match loser
CBW = Centrebet odds of match winner
CBL = Centrebet odds of match loser
EXW = Expekt odds of match winner
EXL = Expekt odds of match loser
LBW = Ladbrokes odds of match winner
LBL = Ladbrokes odds of match loser
GBW = Gamebookers odds of match winner
GBL = Gamebookers odds of match loser
IWW = Interwetten odds of match winner
IWL = Interwetten odds of match loser
PSW = Pinnacles Sports odds of match winner
PSL = Pinnacles Sports odds of match loser
SBW = Sportingbet odds of match winner
SBL = Sportingbet odds of match loser
SJW = Stan James odds of match winner
SJL = Stan James odds of match loser
UBW = Unibet odds of match winner
UBL = Unibet odds of match loser

MaxW= Maximum odds of match winner (as shown by Oddsportal.com)
MaxL= Maximum odds of match loser (as shown by Oddsportal.com)
AvgW= Average odds of match winner (as shown by Oddsportal.com)
AvgL= Average odds of match loser (as shown by Oddsportal.com)