

IPL Matches Data Analysis Using Spark (Core)

Let's mine the data of IPL and derive some important primitives from it like which stadium is most suitable for batting first and which stadium is most suitable for bowling first.

Here is the data set (**matches.csv**) description:

id,season,city,date,team1,team2,toss_winner,toss_decision,result,dl_applied,winner,win_by_runs,win_by_wickets,player_of_match,venue,umpire1,umpire2,umpire3

1. Which stadium is best suitable for first batting (%)

Here we evaluate that which stadium is most suitable for first batting. Here are the details how can we do that.

win_by_runs means – Team batted first and won the Match by margin of some runs.

win_by_wickets means – Team batted second and chased the target successfully.

Take out the columns **toss_decision**, **won_by_runs**, **won_by_wickets**, **venue** and perform the analysis.

2. Which stadium is best suitable for first bowling (%)

Here we evaluate that which stadium is most suitable for first batting. Here are the details how can we do that.

win_by_runs means – first bat won or second bowl

win_by_wickets means – second bat won or first bowl

Take out the columns **toss_decision**, **won_by_runs**, **won_by_wickets**, **venue** and perform the analysis.