

Analysis of Indian Premier League (Cricket) dataset

Cricket is a bat-and-ball game played between two teams of eleven players each on a cricket field, at the centre of which is a rectangular 20-metre (22-yard) pitch with a target at each end called the wicket (a set of three wooden stumps upon which two bails sit). Each phase of play is called an innings, during which one team bats, attempting to score as many runs as possible, whilst their opponents bowl and field, attempting to minimise the number of runs scored. When each innings ends, the teams usually swap roles for the next innings (i.e. the team that previously batted will bowl/field, and vice versa). The teams each bat for one or two innings, depending on the type of match. The winning team is the one that scores the most runs, including any extras gained (except when the result is not a win/loss result).

Dataset

All Indian Premier League Cricket matches between 2008 and 2016.

This is the ball by ball data of all the IPL cricket matches till season 9.

The dataset contains 1 file: matches.csv.

Matches.csv contains details related to the match such as location, contesting teams, Umpires, results, etc.

Tools & Libraries

Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks

Data Description

The dataset contains the following Columns:

- Id: It's a sequence of number.
- Season: The league held in year.
- City :Contains match held in city
- Team1: Data of teams playing against team2.
- Team2: Data of teams playing against team1
- Toss Winner: Contains record of team1 and team2 toss.

- Toss decision: Contain data of after winning toss.
- Result: Result of the match.
- DI_applied: Was match interrupted by any reason (1,0)
- Winner: Winner of match.
- Wins_by_runs: Contain team win by runs data
- Wins_by_wicket: Contain team win by no of wickets.
- Player_of_match: Best player performance in match.
- Venue: match played at stadium name.
- Umpire1: referee of the match.
- Umpire2: referee of the match
- Umpire3: 3rd umpire of the match.

EDA

I looked at the different-different trends of the data and Below is a few highlights of the analysis.

- Top 10 player of the match.
- Plotting bar chart of top10 palyer of the match.
- Number of toss wins wrt to teams.
- Finding unique values.
- Finding null values.
- Extracting the records where team that won did batting first.
- Extracting the records where team that won did fielding first.
- Making a plot for distribution of runs.
- Plotting distribution of win by wickets for team who won and choose to do field first.
- Which team won match most doing fielding first?
- Plotting with top 3 teams who batted second with most number of wins.
- Number of matches played each season.
- Number of matches played in each city.
- Has toss winning helped in match winning?
- Which season had most number of matches?
- Most IPL Matches played in a Venue.
- IPL Matches Played by Each Team.
- The most successful IPL Team.
- Pie chart for successfully winner in season.