**Analysis of IPL complete dataset (2008-2020)**

**The Indian Premier League (IPL) is a professional**[**Twenty20 cricket**](https://en.wikipedia.org/wiki/Twenty20_cricket)**league, contested by various teams based out of different Indian cities. The league was founded by the**[**Board of Control for Cricket in India**](https://en.wikipedia.org/wiki/Board_of_Control_for_Cricket_in_India)**(BCCI) in 2007. It is usually held between March and May of every year and has an exclusive window in the ICC Future Tours Program.**



**Dataset**

The dataset used for this project was found on Kaggle. This data contains record of all IPL matches which had been played since season 2008 till season 2020. The basic idea of analyzing the IPL dataset is to come up with some sort of basic analysis for cricket lovers, some simple facts which can be unfolded with the help of exploratory data analysis and data visualization.

Following analyses were done using the data:

* Team which has won maximum matches from all seasons
* Players which have bagged POM award for the maximum times
* Stadium on which maximum matches had been played
* Teams which had a good luck with winning toss
* Head to Head comparison between all teams from all seasons
* Analysis of relation between winning toss and winning matches
* Analysis of possibility of winning matches based on home ground factor

**Tools & Libraries**

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

**Data Description**

This data contains record of all IPL matches which had been played since season 2008 till season 2020 as mentioned above. The dataset contains total 33 columns and I am going to list down only those columns which are used in the EDA. Following Columns have been used:

|  |  |
| --- | --- |
| **Column** | **Description** |
| season | season of IPL |
| short\_name | Both playing team names |
| home\_team | Team playing on home ground |
| toss\_won | Team which has won toss |
| decision | Decision after wininnig toss(Bat/Bowl) |
| winner | Winning team |
| venue\_name | Stadium name with its city location |
| pom | Player of the match (Man of the match) |

**Data Cleaning**

I made the following changes with data:

* Deleted rows with data having NULL values as count of NULLS was negligible compared to the whole size of the dataset
* Deleted unwanted columns of data as it was not required for my analysis purpose e.g. umpires details, commentary details, etc.
* Duplicate short names were reduced by replacing and keeping a single short name (e.g. MI vs CSK and CSK vs MI, both are same. Kept only a single entry as both mean the same).

**EDA**

I looked at the different trends of the data and below are few key highlights of the analysis.

* In all IPL seasons, MI has won highest matches followed by CSK & KKR
* In all IPL seasons, Chris Gayle was player of the match for the maximum number of times followed by AB de Villiers and Rohit Sharma
* In all IPL seasons, MI has won toss maximum times followed by DC and KKR
* MI has won toss and matches together maximum times followed by CSK & KKR
* Being a home team, CSK has won maximum matches followed by MI and KKR
* Bowling first was the maximum chosen option by all toss won teams from all seasons. It shows teams prefer to chase score than defending Maximum IPL matches are played on M Chidambaram Stadium Bangalore