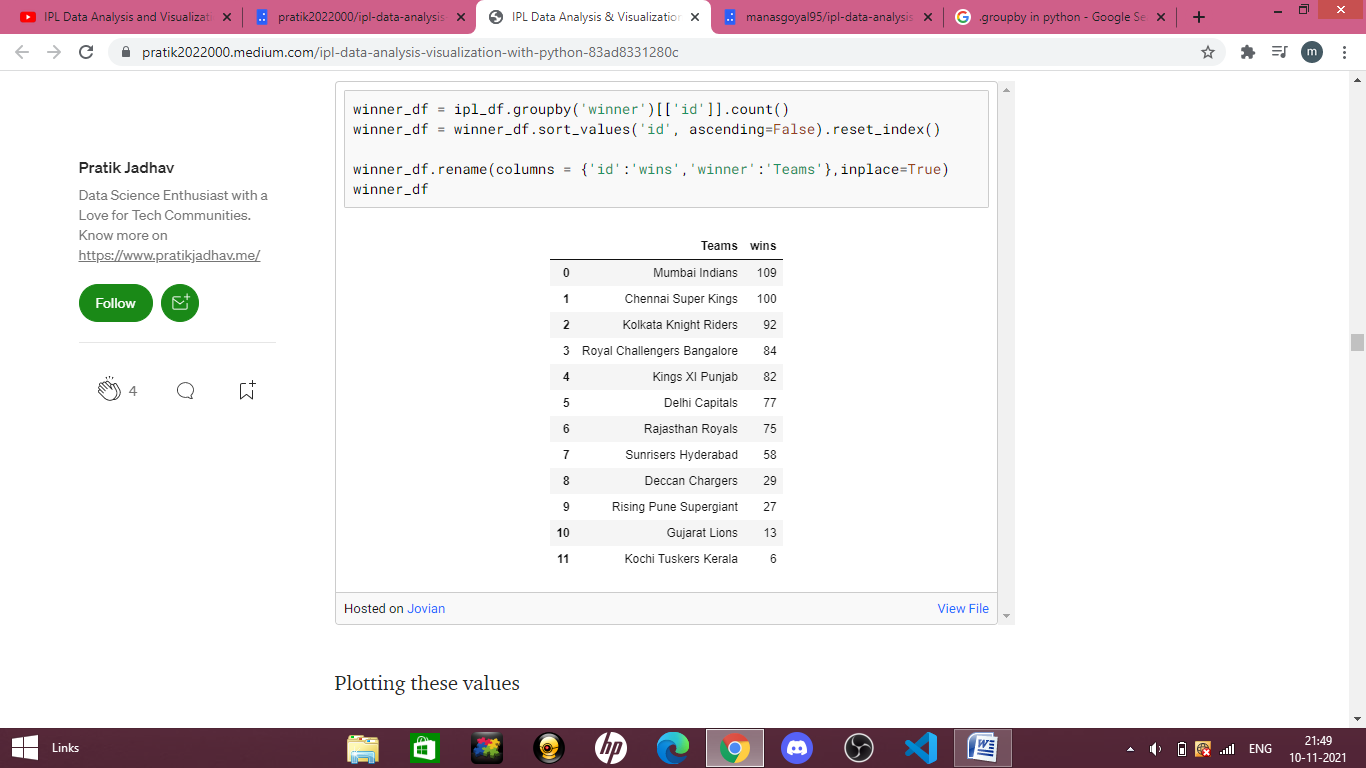
Problem Statement: IPL Data Analysis & Visualization with Python

For the Data Analysis we have used the IPL Dataset Available on Kaggle. The python libraries we used are Numpy, pandas and matplotlib. The data set comtains the information of IPL matches from 2008 till 2019. It contains various columns like (‘team1’, ‘team2’, ‘toss winner’, ‘toss decision’, ‘winner’. ‘Season’, etc.

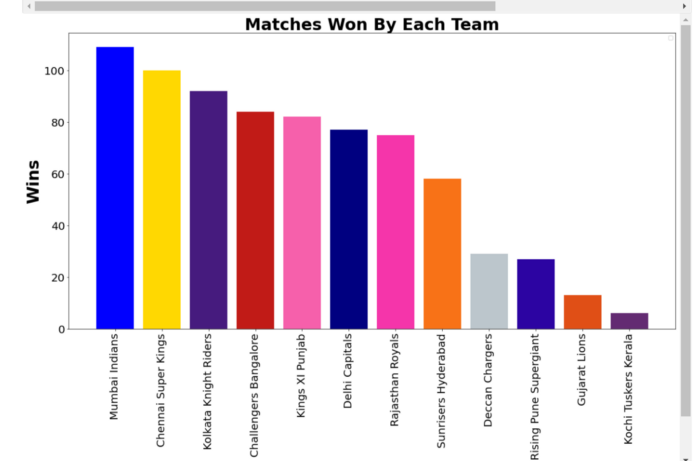
We have read the dataset using pandas and using .info() method we can see the type of values each column contains. We have used .replace() method for cleaning i.e to change the name of Delhi Daredevils with Delhi Capitals.

We can see that a total of 756 matches Played in 12 Seasons and 743 went well till end i.e they were a clear winner and not a tie.

Using .groupby function we can see which team have most number of matches from 2008 to 2019



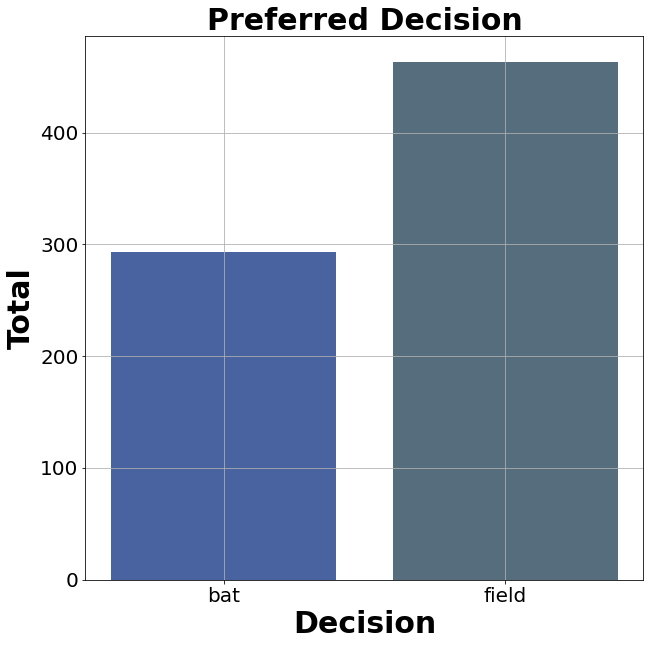
And then plotting these values



We have grouped the toss\_decision column to see the preferred decision taken by the winning team.

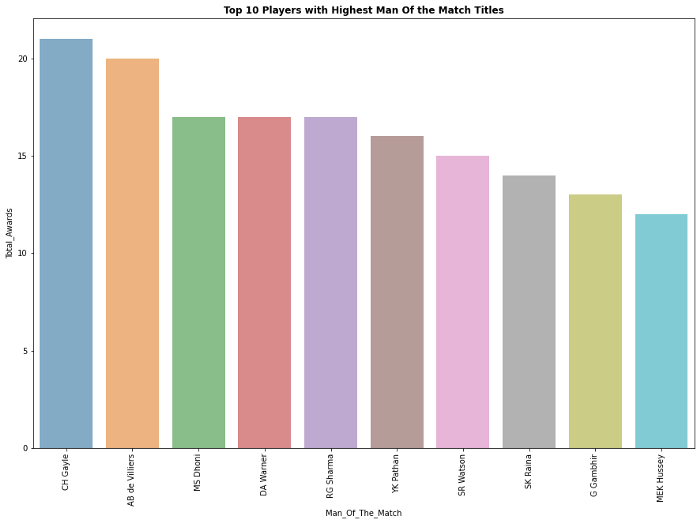
|  | **Decision** | **Total** |
| --- | --- | --- |
| **0** | bat | 293 |
| **1** | field | 463 |

This observation tells us. Almost 60% of the times the toss winner Chooses to Field First.



Players who have received most number of man of the match awards are also grouped in descending order.

|  | **Man\_Of\_The\_Match** | **Total\_Awards** |
| --- | --- | --- |
| **0** | CH Gayle | 21 |
| **1** | AB de Villiers | 20 |
| **2** | MS Dhoni | 17 |
| **3** | DA Warner | 17 |
| **4** | RG Sharma | 17 |
| **5** | YK Pathan | 16 |
| **6** | SR Watson | 15 |
| **7** | SK Raina | 14 |
| **8** | G Gambhir | 13 |
| **9** | MEK Hussey | 12 |



For finding which team holds winner title the most, will create a copy of the original data frame using .copy() so that the changes we make won’t affect the main data frame and after counting the wins

Mumbai Indians 4

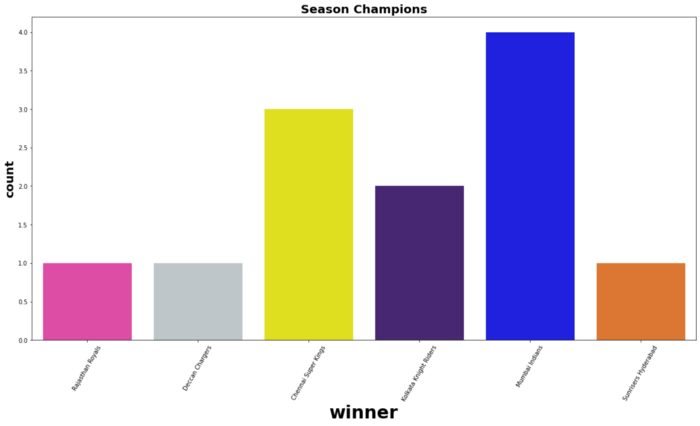
Chennai Super Kings 3

Kolkata Knight Riders 2

Rajasthan Royals 1

Deccan Chargers 1

Sunrisers Hyderabad 1



Conclusion: So With that, we’ve come to the end of this analysis on IPL Data. From the analysis it’s clear that MI and CSK are two dominant teams in IPL. Also other takeaways are the ease of python: syntax as well as the code readability. It’s a powerful language and has a wide variety of libraries like for almost everything. Numpy for numerical computing. Matplotlib for visualization and many more.