

EXPLORATORY DATA ANALYSIS

Data Selection:

I have chosen the IPL dataset which originally consists of two CSV files - Matches and Deliveries. The matches file describes the matches played in the IPL from 2008 - 2019 in brief. The deliveries file describes about each of the deliveries played in each of the IPL matches from 2008 - 2019. For the EDA purposes, I have chosen only the matches dataset for this assignment.

The dataset can be found at the website:

<https://www.kaggle.com/nowke9/ipldata?select=matches.csv>

Questions:

- 1) Which team has won the toss most number of times?
- 2) What was the overall win contribution percentage of the IPL teams from 2008-2019?
- 3) Which players have won the most number of "Player of Match" titles?
- 4) What was the toss decisions of teams who won the matches?
- 5) What was the total number of wins by each team from 2008-2019?
- 6) What were the number of cities hosting the matches in the IPL during the 12 years?
- 7) What were the number of match wins in the home games for the teams?
- 8) What were the head to head match win counts for the teams?

The above questions will be answered with the help of the visualisations of this dataset made using Tableau.

Dataset Description:

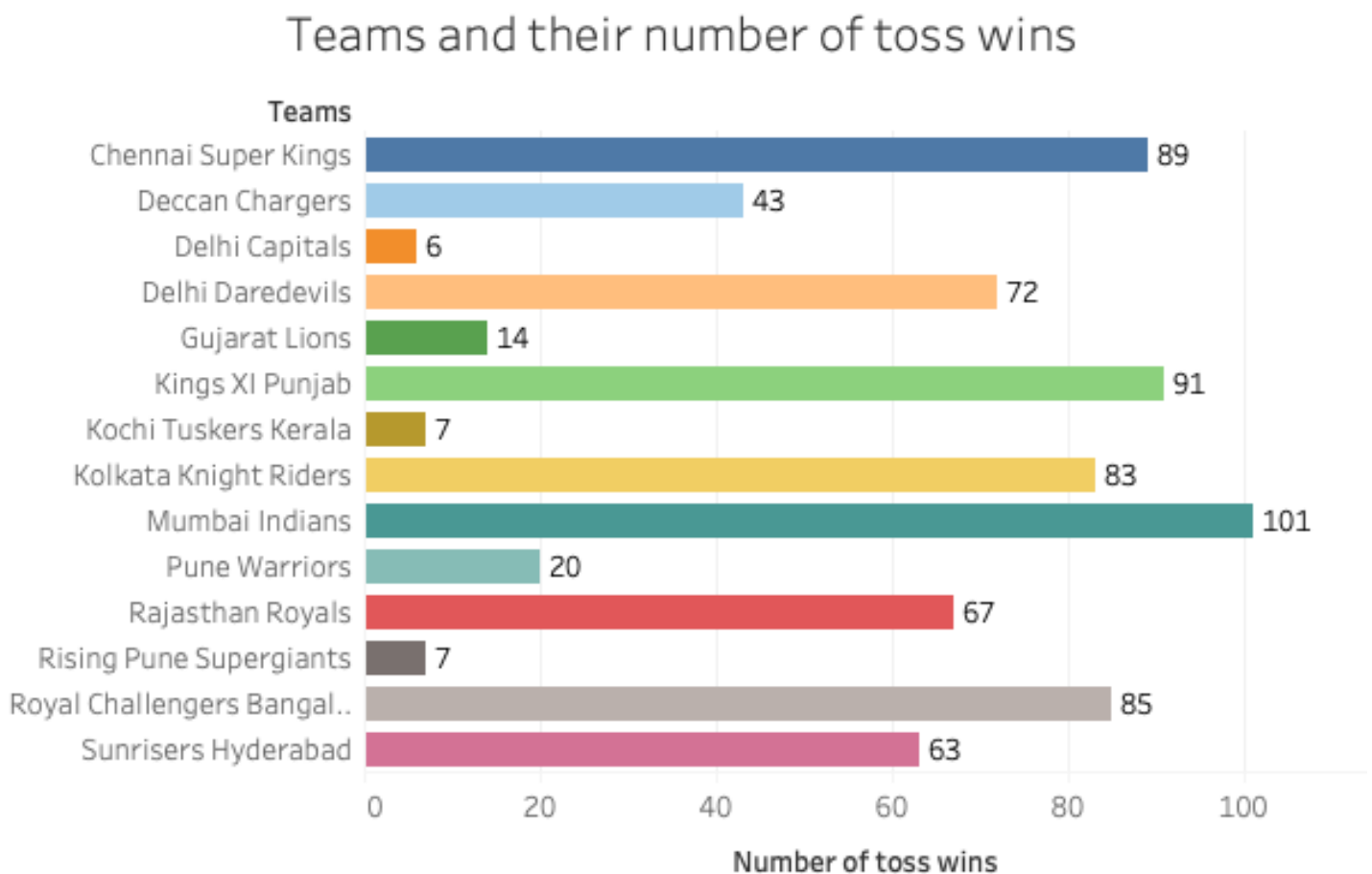
The chosen dataset is matches.csv from the IPL dataset. The matches.csv file contains the following columns.

1. Season (the year of the IPL)
2. City (the city where the match was held)
3. Date (the date when the match was played)
4. Team1 (denotes one of the teams which played that day)
5. Team2 (denotes one of the teams which played that day)
6. Toss_winner (denotes the team which won the toss)
7. Toss_decision (denotes the decision made by the toss winner)
8. Result (denotes the result of the match)
9. DL_applied (denoted whether Duckworth_Lewis_Stern method is used or not)
10. Winner (denotes the winner of the match)
11. Wins_by_runs (denotes whether the match was won by runs)
12. Wins_by_wickets (denotes whether the match was won by wickets)
13. Player_of_match (denotes the player of the match)
14. Venue (represents the venue for that match)
15. Umpire1 (denotes the first umpire of the match)
16. Umpire2 (denotes the second umpire of the match)
17. Umpire3 (denotes the third umpire of the match)

Some of the umpires columns contain missing values but they were not chosen for EDA purposes. The columns toss_decision and result contain only two categorical values which are field/bat and normal/tie respectively. The column DL_applied contains only binary value which denotes 0 for not DL and 1 for DL application.

Answers:

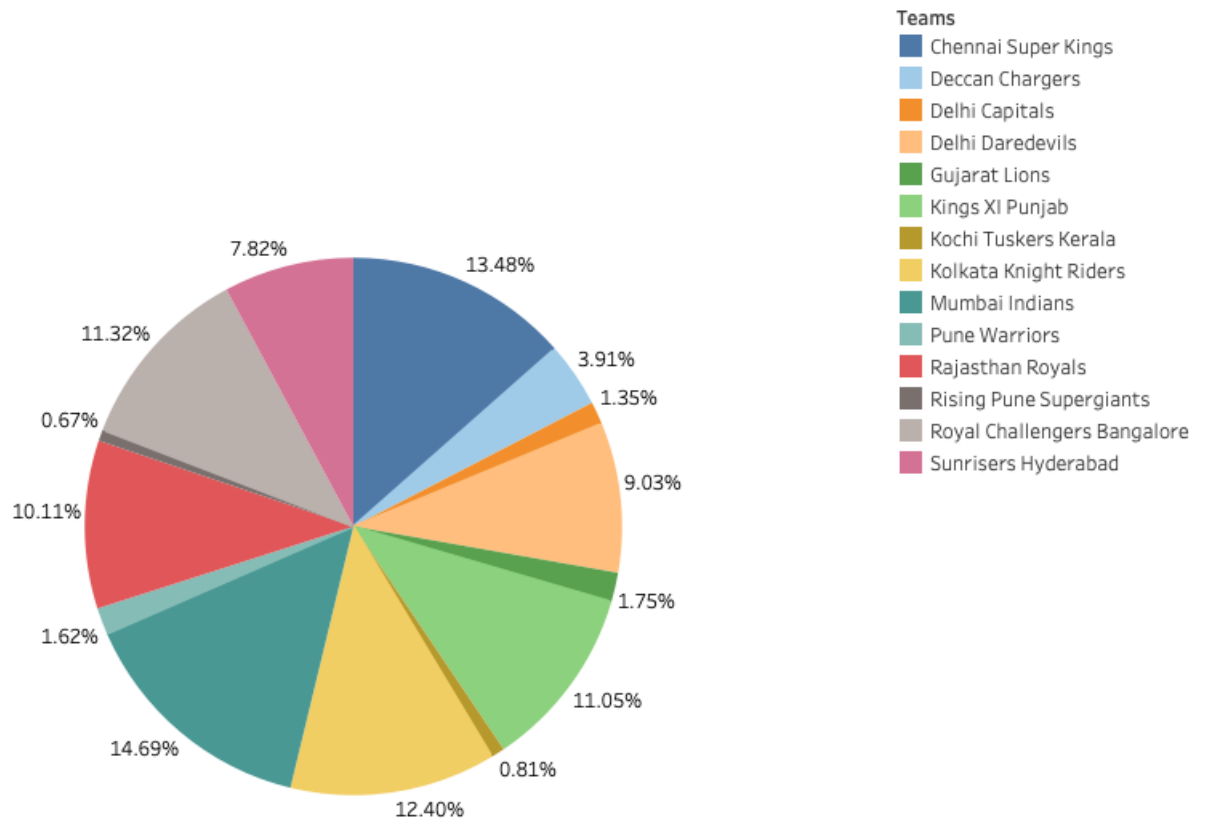
1) Which team has won the toss most number of times?



This visualization is done using bar chart which clearly explains the total number of toss won by all the teams of IPL from 2008-2019. By looking at the visualization, it can be clearly concluded that Mumbai Indians has won the toss most number of times (101) followed by Kings XI Punjab(91) and Chennai Super Kings (89).

2) What was the overall win contribution percentage of the IPL teams from 2008-2019?

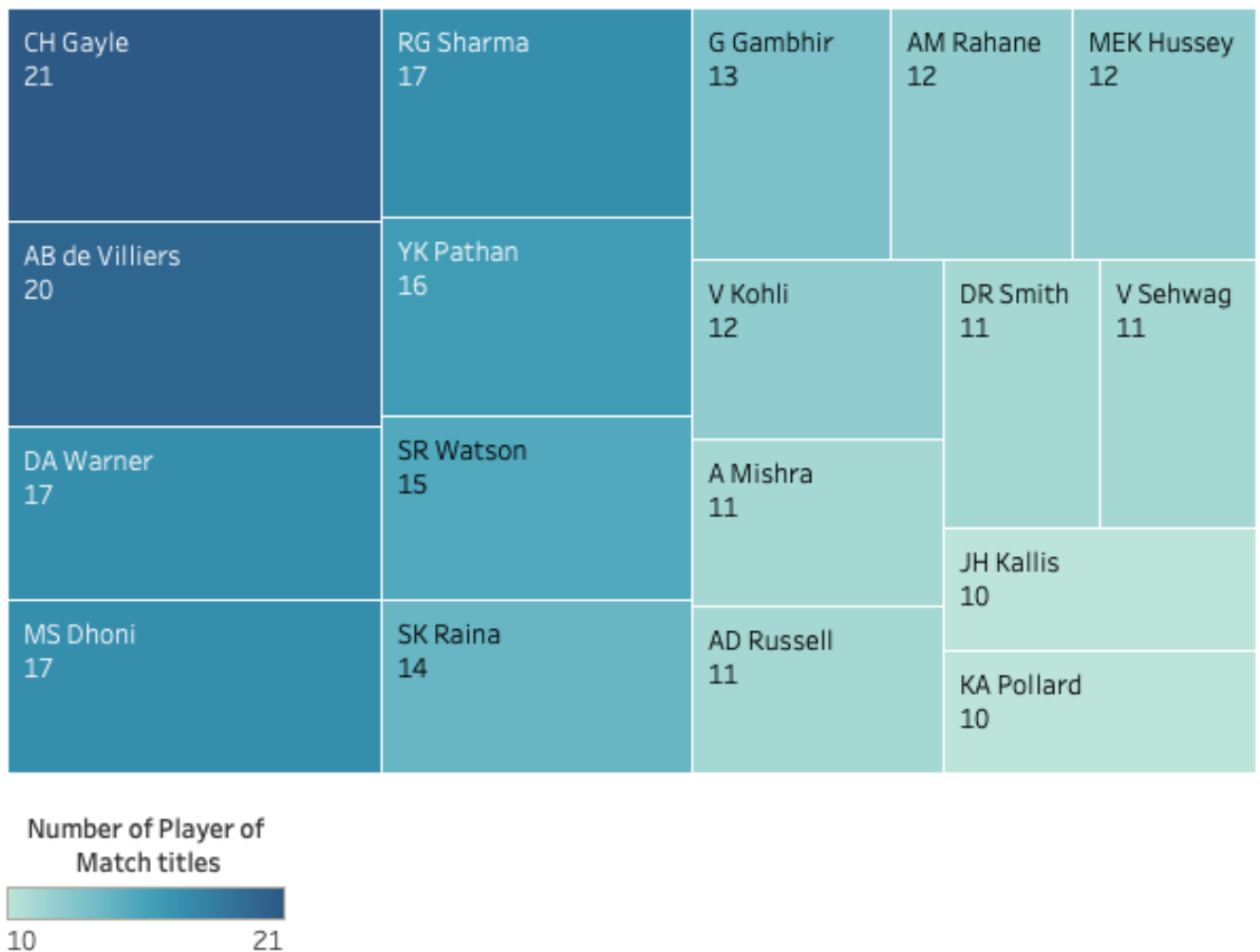
Win contribution percentage of IPL teams from 2008-2019



This visualization is done using pie chart as percentages are best expressed with pie chart as the total sum will all be equal to 100% and the teams are scaled to their percentages accordingly in the pie chart. From this pie chart, it can be concluded that Mumbai Indians won the most number of matches and Kochi Tuskers Kerala won the least number of matches.

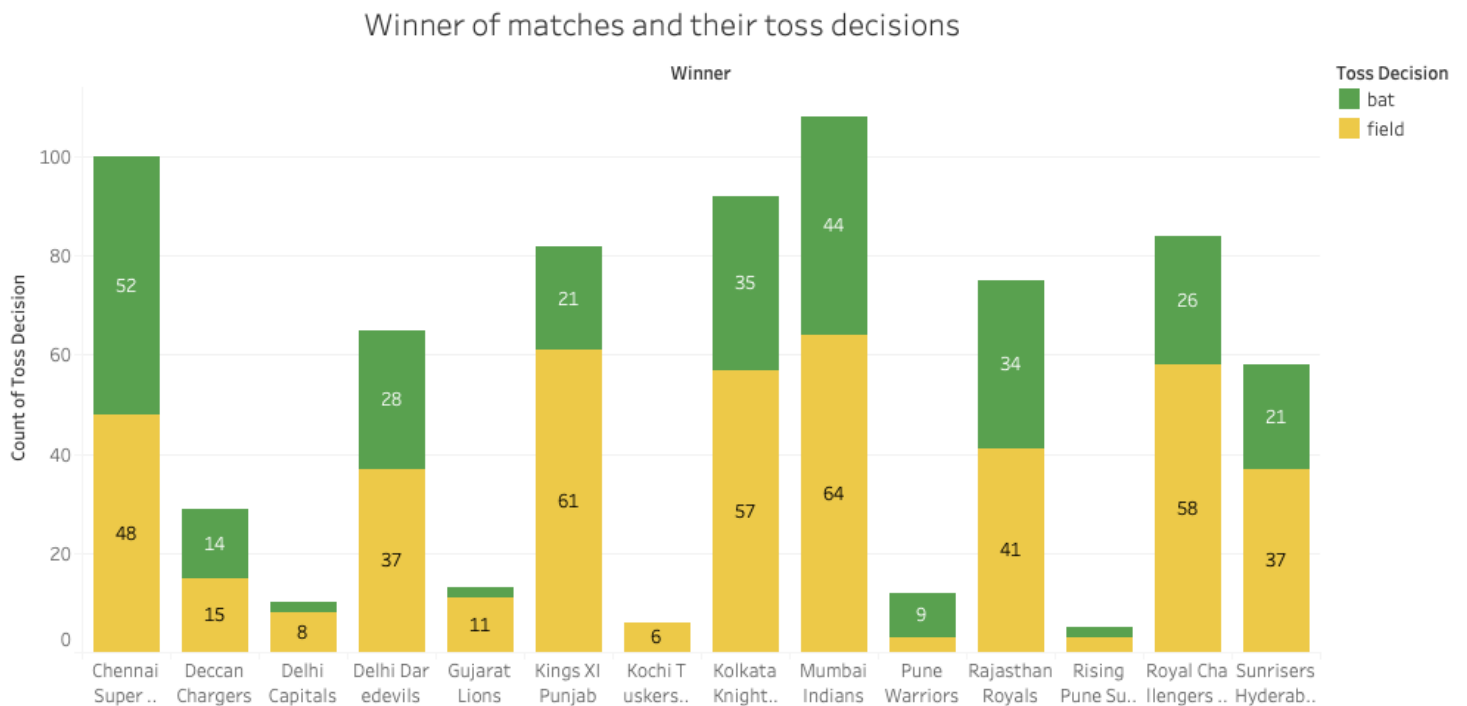
3) Which players have won the most number of “Player of Match” titles?

Players and their number of **Player of Match** titles



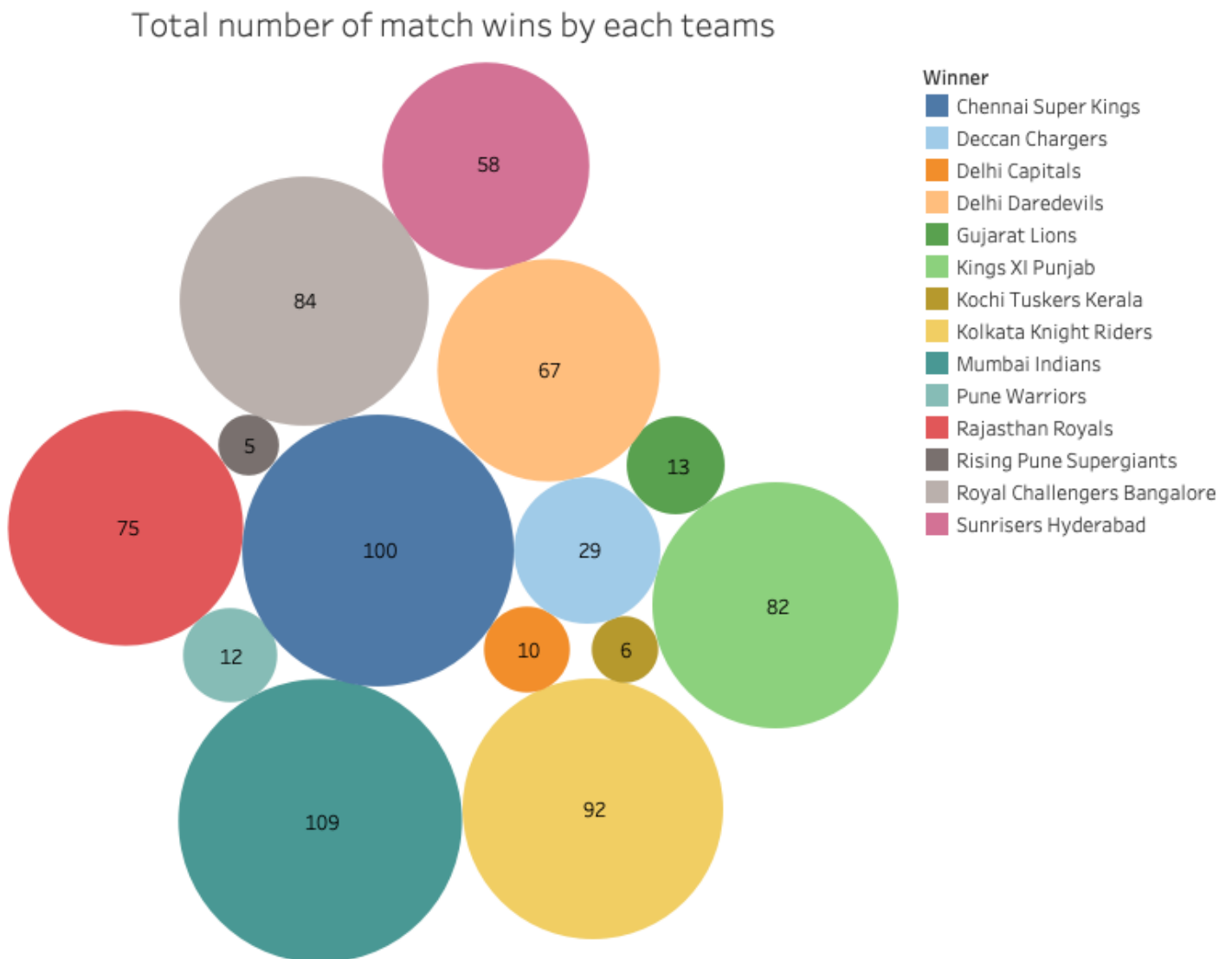
This visualization is done using Treemaps as it displays the hierarchical level of players' names with respect to their number of Player of Match titles. From this treemap, it can be said that Chris Gayle won the Player of titles the most number of times which is 21 closely followed by AB de Villiers which stand on 20.

4) What was the toss decisions of teams who won the matches?



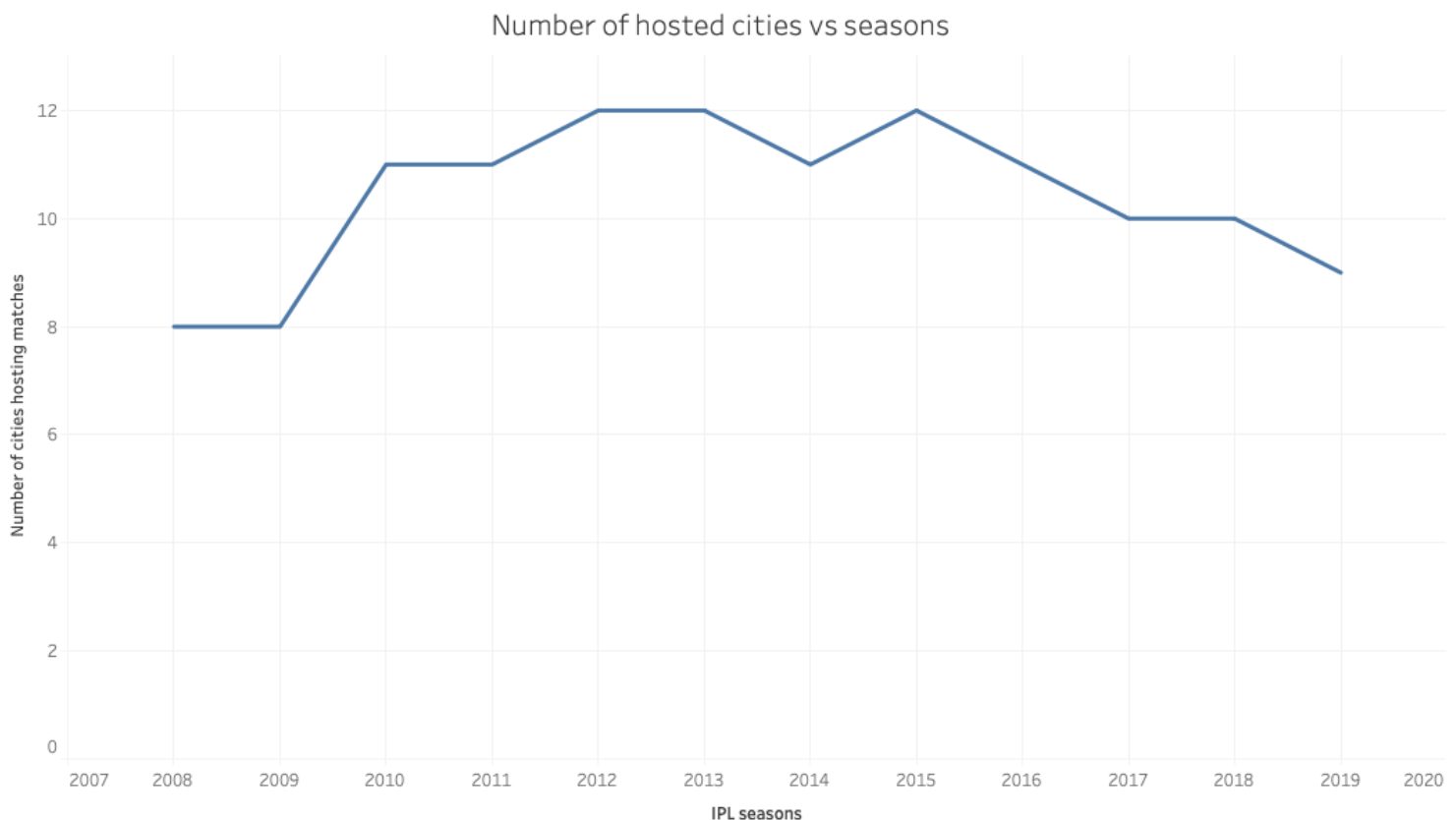
This visualisation is made using stacked bars as the toss decision has only two distinct values (bat/field), it can be easily differentiated by using stacked bars. This chart represents the toss decisions of teams who won the matches and it can be concluded that almost all the teams won most of the matches by fielding first.

5) What was the total number of wins by each team from 2008-2019?



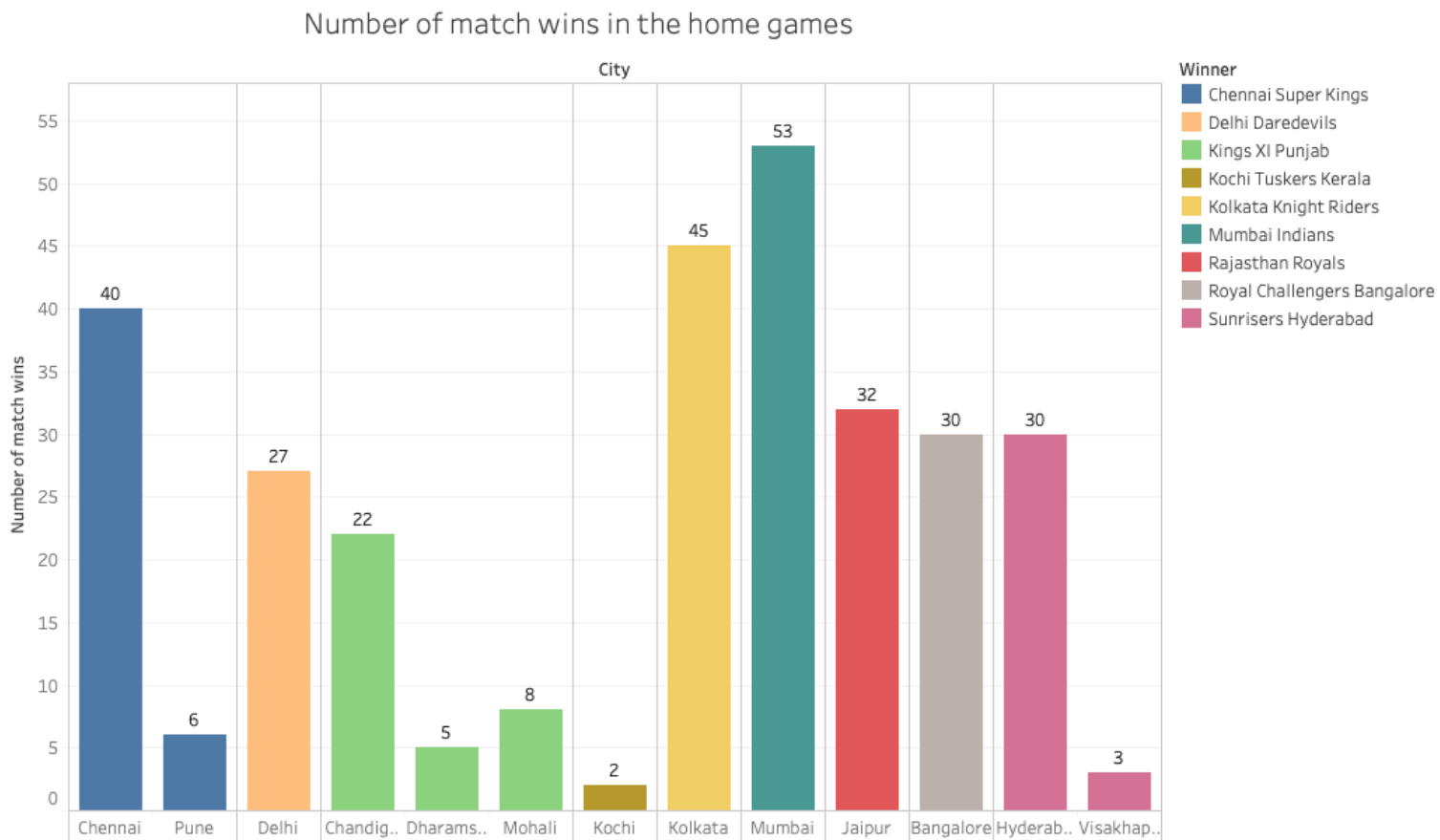
This visualisation is made using packed bubbles chart and each of the bubbles represents a team and the bubble size directly corresponds to the number of match wins in relation to the match wins of the other teams. From this bubble chart, it can be said that Mumbai Indians won the most number of matches (109) followed by Chennai Super Kings (100).

6) What were the number of cities hosting the matches in the IPL during the 12 years?



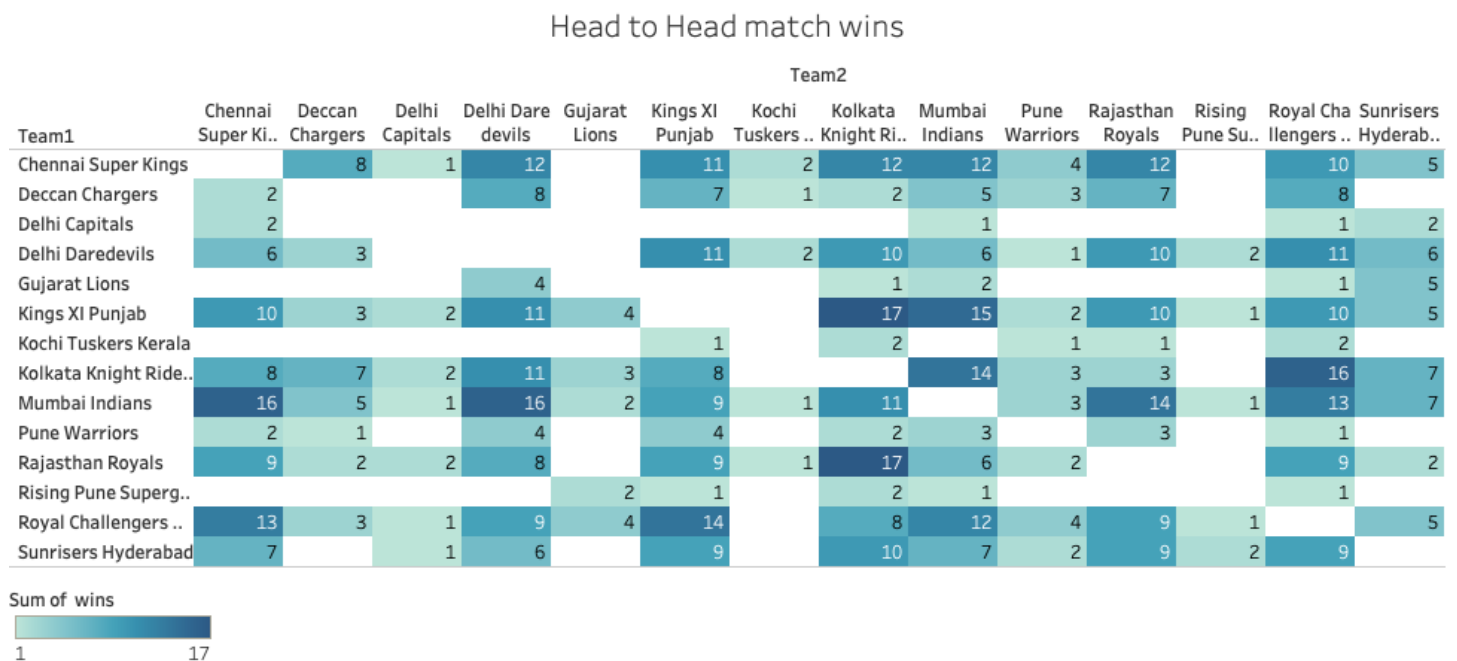
This visualization is made using simple line graph as the number of venues will be mapped against the season, it will be best to use line graph to represent them. From the line graph, it can be said that the inaugural 2008 IPL edition hosted the matches in less number of cities whereas the 2012, 2013 and 2015 IPL editions hosted the match in most number of cities.

7) What were the number of match wins in the home games for the teams?



This visualization is made using simple bar chart as it maps the teams with their home cities and the number of match wins in the home cities is mapped against them. From this bar chart, it can be clearly concluded that Mumbai Indians won the most number of home matches whereas Kochi Tuskers Kerala won the least number of home matches.

8) What were the head to head match win counts for the teams?



This visualization is done using highlight map as it clearly compares each of the match results and highlight the sum of the match wins between the two teams. From this highlight map, it can be said that Kings XI Punjab and Rajasthan Royals won the most number of matches against Kolkata Knight Riders (17).