Business Intelligence and Data Visualisation

Project Report

School of Engineering and Technology

Department of

Computer Science Engineering

Parakh Singhal 18csu152

Saksham Yadav 18csu187

Sarthak Gupta 18csu193

The NorthCap University

Gurugram, Haryana

IPL Analysis

The main objective behind making these visualisations is to depict how we can analyse a dataset to show its pros and cons and plan for further decision making.

We chose this IPL dataset to express our love for cricket. Due to lockdown in India by Prime Minister Narendra Modi Ji, IPL-2020 is cancelled and we cannot experience that passion, that love to see our favourite players on the field. So we decided to take IPL as our main objective for the project and let others experience the same passion.

The Dataset includes two parts:

- Deliveries
- Matches

Link to Dataset: https://www.kaggle.com/vaishaligarg/analysing-ipl-data/data

The columns of deliveries file:

match_id – that shows the id of respective matches

inning – shows match inning

batting_team – names of batting teams

bowling_team – names of bowling teams

over – no. of overs

ball – no. of balls bowled

batsman – name of the batsman

non_striker – whether he is a non – striker or not

bowler – name of the bowler

is_super_over – it is a super over or not

wide_runs – count of wide runs

bye_runs – bye runs count

legbye_runs – legbye runs count

noball_runs – noball runs count

penalty_runs – penalty runs count

batsman_runs – runs scored by batsman

extra_runs – extra runs scored

total_runs – total runs count

player_dismissed – which player was dismissed

dismissal_kind – why he was dismissed

fielder – name of the fielder

The columns of matches file:

Id - this shows the id of respective matches

season – seasons in which matches were played

city – in which city they were played

date – on which day or month or year

team1 – names of teams

team2 - names of teams

toss_winner – which team won the toss

toss_decision – decision of toss

result – result of the match played

dl_applied – dl applied or not

winner – winner of the match

win_by_runs - by how many runs did the team won?

win_by_wickets – won by how many wickets?

player_of_match – player of the match award winner

venue – venue of match played

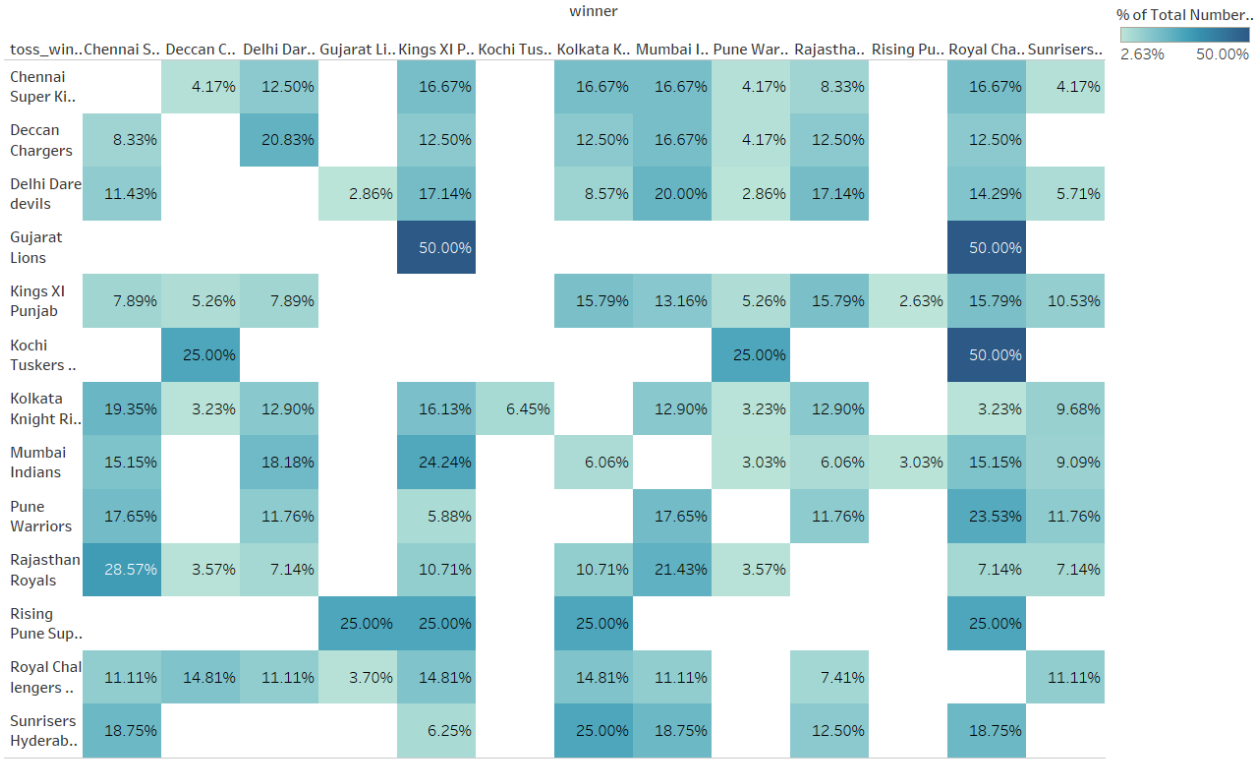
umpire1 – umpire 1

umpire2 – umpire 2

umpire 3 – umpire 3

WIN Percentage if team wins TOSS

Inference - We could predict the chances of winning/losing before starting of the game with gujrat lions able to win the game with the help of winning the toss the maximum number of times

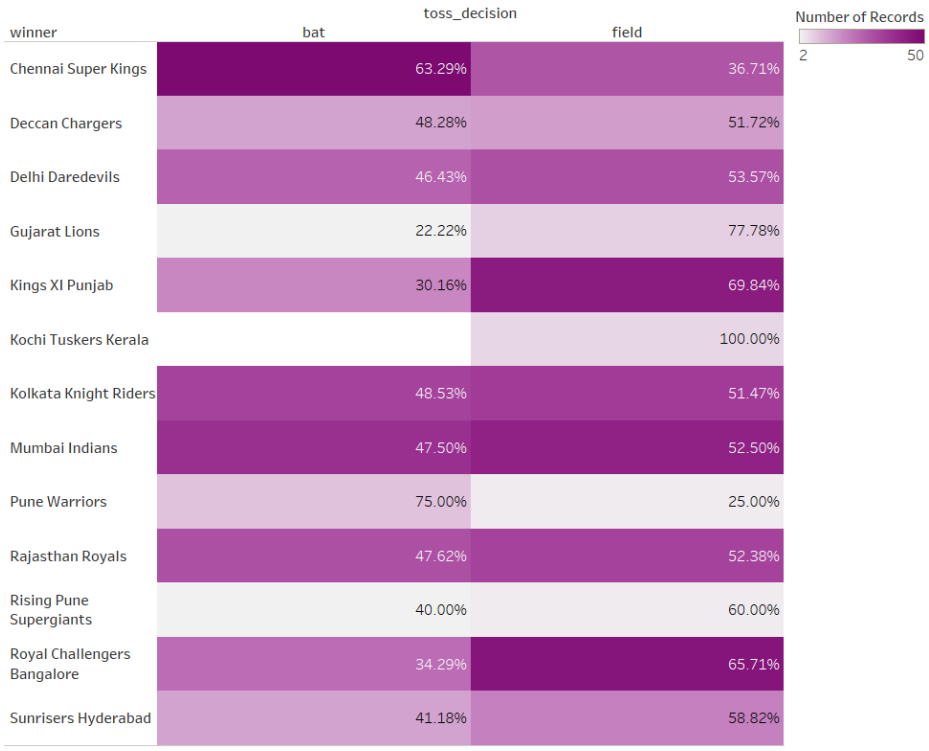


toss_win..	Chennai S..	Deccan C..	Delhi Dar..	Gujarat Li..	Kings XI P..	Kochi Tus..	Kolkata K..	Mumbai I..	Pune War..	Rajastha..	Rising Pu..	Royal Cha..	Sunrisers..
Chennai Super Ki..		4.17%	12.50%		16.67%		16.67%	16.67%	4.17%	8.33%		16.67%	4.17%
Deccan Chargers	8.33%		20.83%		12.50%		12.50%	16.67%	4.17%	12.50%		12.50%	
Delhi Dare devils	11.43%			2.86%	17.14%		8.57%	20.00%	2.86%	17.14%		14.29%	5.71%
Gujarat Lions					50.00%							50.00%	
Kings XI Punjab	7.89%	5.26%	7.89%				15.79%	13.16%	5.26%	15.79%	2.63%	15.79%	10.53%
Kochi Tuskers ..		25.00%							25.00%			50.00%	
Kolkata Knight Ri..	19.35%	3.23%	12.90%		16.13%	6.45%		12.90%	3.23%	12.90%		3.23%	9.68%
Mumbai Indians	15.15%		18.18%		24.24%		6.06%		3.03%	6.06%	3.03%	15.15%	9.09%
Pune Warriors	17.65%		11.76%		5.88%			17.65%		11.76%		23.53%	11.76%
Rajasthan Royals	28.57%	3.57%	7.14%		10.71%		10.71%	21.43%	3.57%			7.14%	7.14%
Rising Pune Sup..				25.00%	25.00%		25.00%					25.00%	
Royal Challengers ..	11.11%	14.81%	11.11%	3.70%	14.81%		14.81%	11.11%		7.41%			11.11%
Sunrisers Hyderab..	18.75%				6.25%		25.00%	18.75%		12.50%		18.75%	

% of Total Number of Records broken down by winner vs. toss_winner. Color shows % of Total Number of Records. The marks are labeled by % of Total Number of Records. The view is filtered on winner and Exclusions (toss_winner,winner). The winner filter excludes Null. The Exclusions (toss_winner,winner) filter keeps 98 members.

What if we chose to bat?

Inference - The team could predict which is their strong point batting/bowling with kochi tuskers kerela choosed to field 6 times and won all that times



winner	bat	field
Chennai Super Kings	63.29%	36.71%
Deccan Chargers	48.28%	51.72%
Delhi Daredevils	46.43%	53.57%
Gujarat Lions	22.22%	77.78%
Kings XI Punjab	30.16%	69.84%
Kochi Tuskers Kerala		100.00%
Kolkata Knight Riders	48.53%	51.47%
Mumbai Indians	47.50%	52.50%
Pune Warriors	75.00%	25.00%
Rajasthan Royals	47.62%	52.38%
Rising Pune Supergiants	40.00%	60.00%
Royal Challengers Bangalore	34.29%	65.71%
Sunrisers Hyderabad	41.18%	58.82%

% of Total Number of Records broken down by toss_decision vs. winner. Color shows sum of Number of Records. The marks are labeled by % of Total Number of Records. The view is filtered on winner, which excludes Null.

While on the field there is a lot of tension among the players of both the teams. They are worried what will the result of match, if they choose to bat first then their winning chance will increase? So we created the two crosstabs that tells us about the result of previous matches, how they won, winning chances against a particular team, if they give a big target to the opponent team at first then what will be the results? They can predict results by having a look at the previous tosses and result which will give them some self-confidence and will to cheer up the positive attitude towards the game.

1 vs 1 match result

Inference - The teams could predict to which side are they losing and by how many runs, also they need to pay more attention towards which team

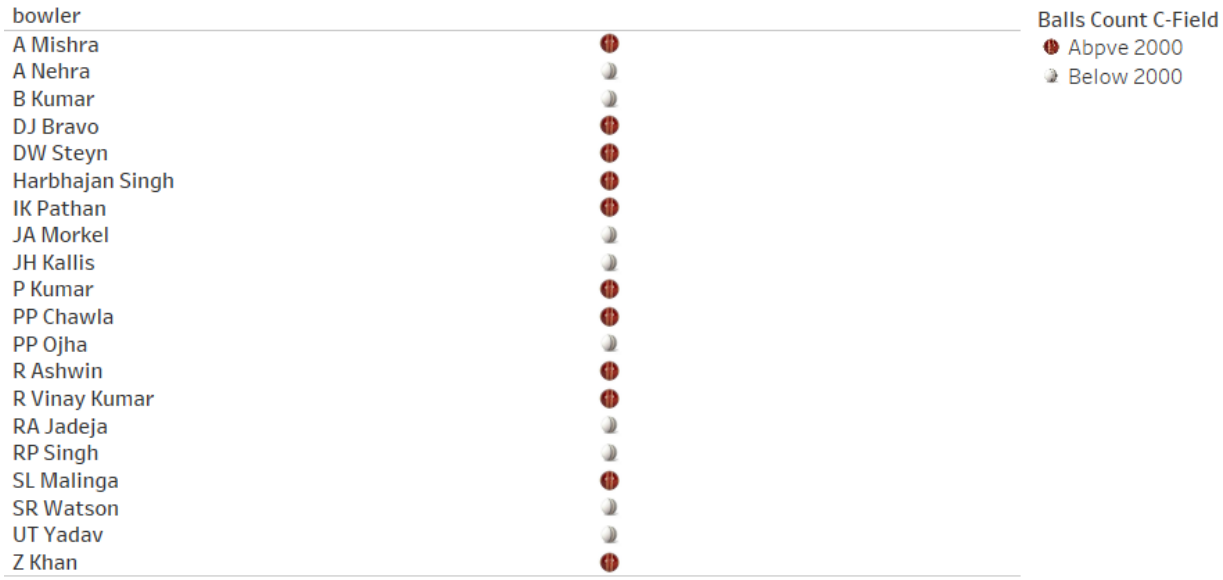
team1	Chennai ..	Delhi Dar..	Kings XI ..	Kolkata K..	Mumbai I..	Rajastha..	Royal Ch..	Sunrisers..
Chennai Super Kings		...	Chennai Super Kings	...	Mumbai Indians	...	Chennai Super Kings	Chennai Super Kings
Delhi Daredevils				Kolkata Knight Riders	...	Rajasthan Royals	Royal Challeng..	Delhi Daredevils
Kings XI Punjab	Chennai Super Kings	Delhi Daredevils		Kolkata Knight Riders	Kings XI Punjab		Kings XI Punjab	
Kolkata Knight Riders		Kolkata Knight Riders					Royal Challeng..	Kolkata Knight Riders
Mumbai Indians	...		Mumbai Indians	...		...	Mumbai Indians	Mumbai Indians
Rajasthan Royals		Rajasthan Royals	...	Rajasthan Royals			Royal Challeng..	
Royal Challengers Bangalore	Chennai Super Kings		Royal Challeng..		Royal Challeng..	Royal Challeng..		Sunrisers Hyderab..
Sunrisers Hyderabad	Sunrisers Hyderab..	Sunrisers Hyderab..	Sunrisers Hyderab..	Sunrisers Hyderab..	Mumbai Indians	...	Royal Challeng..	

Winner broken down by team2 vs. team1. Color shows sum of win_by_runs. The data is filtered on date Year, which keeps 2015.

The above crosstab included the percentage of winning over a span of time. But sometimes we want to see recent match results so that we can train according to their new tactics and new strategies. We get to know about a lot of things like how we played in the last match that insured our win, when was the time the team was at weak stage, are there any new players ready to get into the field, etc. These analysis helps a lot while planning the game strategies but the actual game depends on the players and their will and how they perform!

Top Bowlers with No. of Balls Bowled

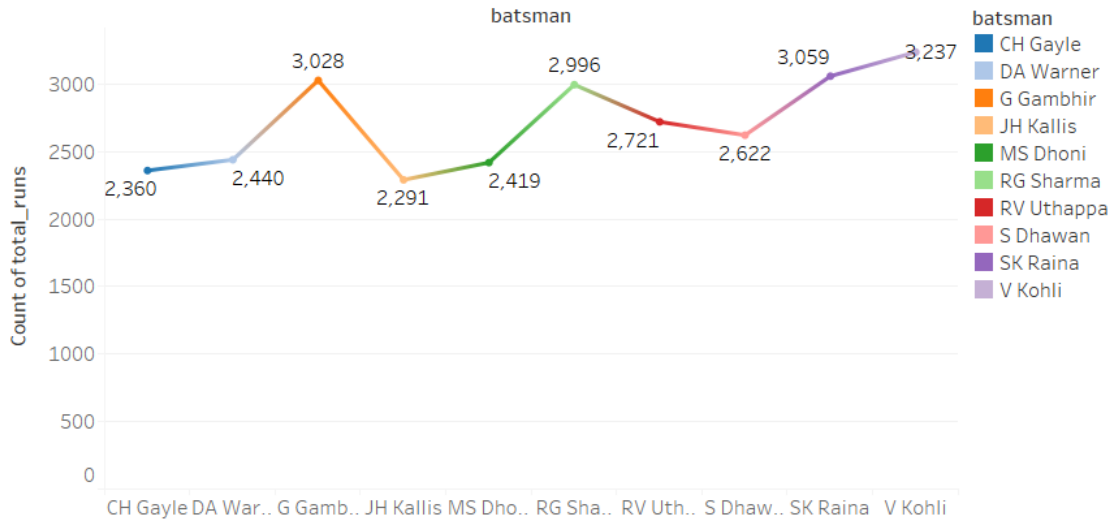
Inference - We can choose the best bowlers by having a look at this KPI view of top 20 bowlers



The view is broken down by bowler. Shape shows details about Balls Count C-Field. The view is filtered on bowler, which has multiple members selected.

Total runs scored by batsmen

Inference - The teams could analyse which batsmen is the threat towards their team, also orange cap holder could be predicted with Virat Kohli being the highest run scorer

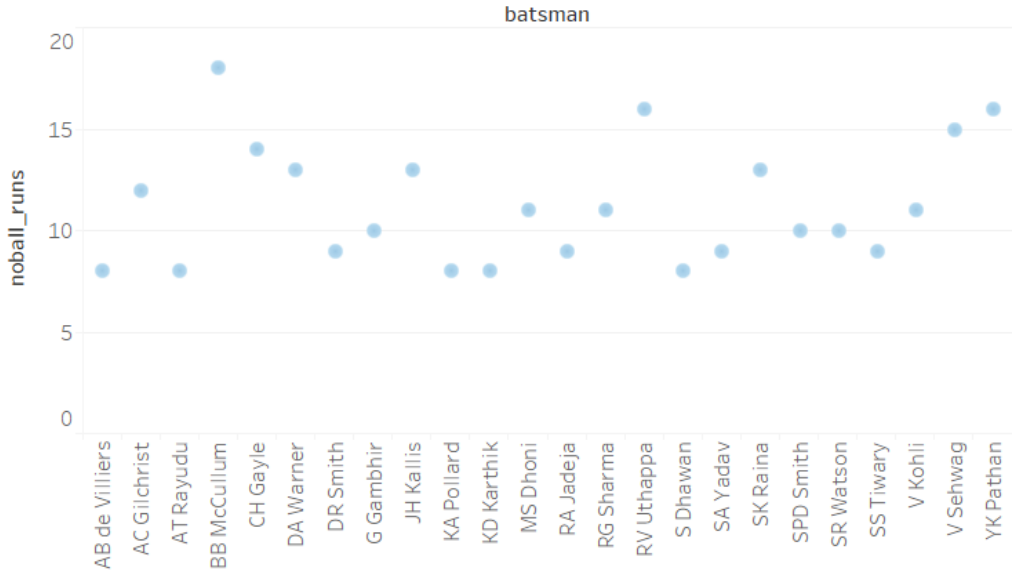


The trend of count of total_runs for batsman. Color shows details about batsman. The marks are labeled by count of total_runs. The data is filtered on Batsman Set, which keeps 10 members. The view is filtered on count of total_runs, which ranges from 1,755 to 3,237.

These types of IPL Analysis not only help while going to play on the field but also help at the time of selecting members or players for team. We need to select the best and most efficient players for our team. By looking at the previous records of the players like the number of runs they have scored, which batsmen would be the best, can this bowler be useful against these batsmen, what is their experience in field, and so on. So it proves to be best to have some of the reports like this while buying the players.

No-Ball runs scored by the batsmen

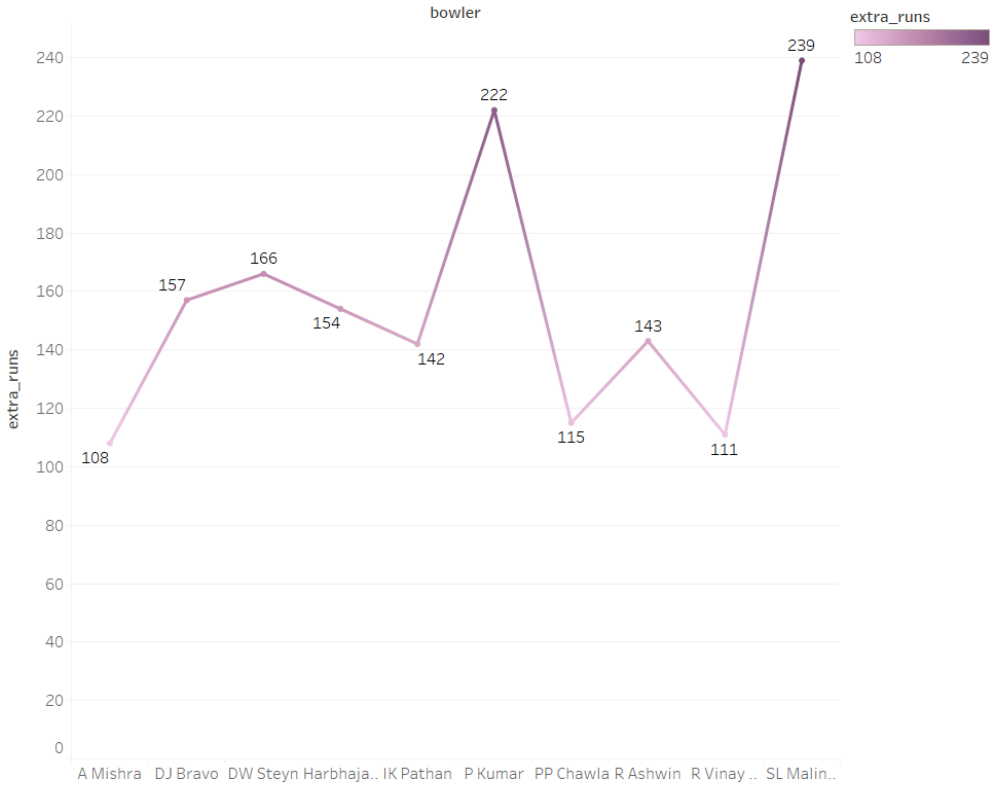
Inference - The teams could see that which player is not letting go of to an easy ball with BB McCullum scoring the maximum runs on no_ball



Sum of noball_runs for each batsman. The view is filtered on sum of noball_runs, which ranges from 8 to 18.

Extra Runs by Bowlers

Inference - Teams could analyse which bowler is giving out extra runs which could cost them the match, also whether to drop/pick the bowler in next edition with SL Maling giving out the maximum number of runs

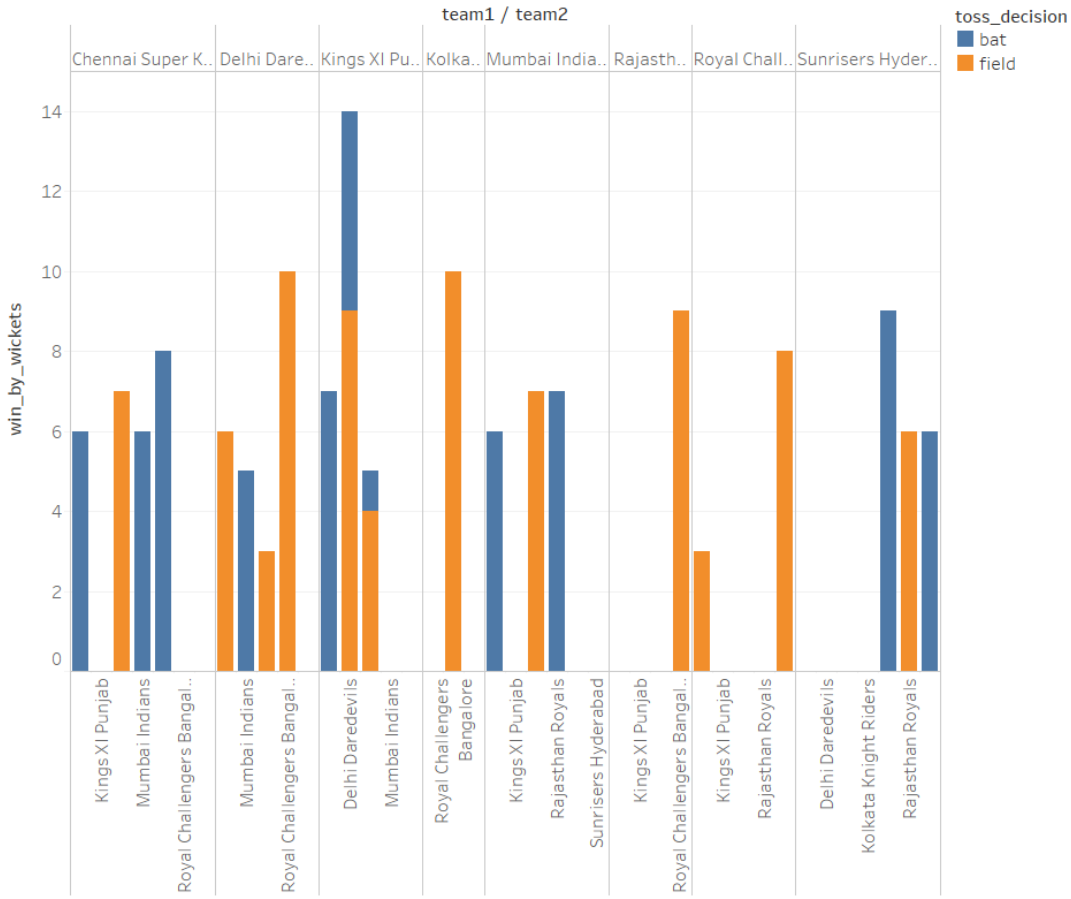


The trend of sum of extra_runs for bowler. Color shows sum of extra_runs. The marks are labeled by sum of extra_runs. The view is filtered on bowler, which has multiple members selected.

The analysis is not only for positive point of view but it can also be used to tackle some of our team limitations also. We can choose whether to drop/pick the batsmen or bowler for the coming overs and whether to choose them for other matches, as they could cost us our victory.

WON by Wickets

Inference - The teams could predict to which side are they losing and by how many wickets, also they need to pay more attention towards which team

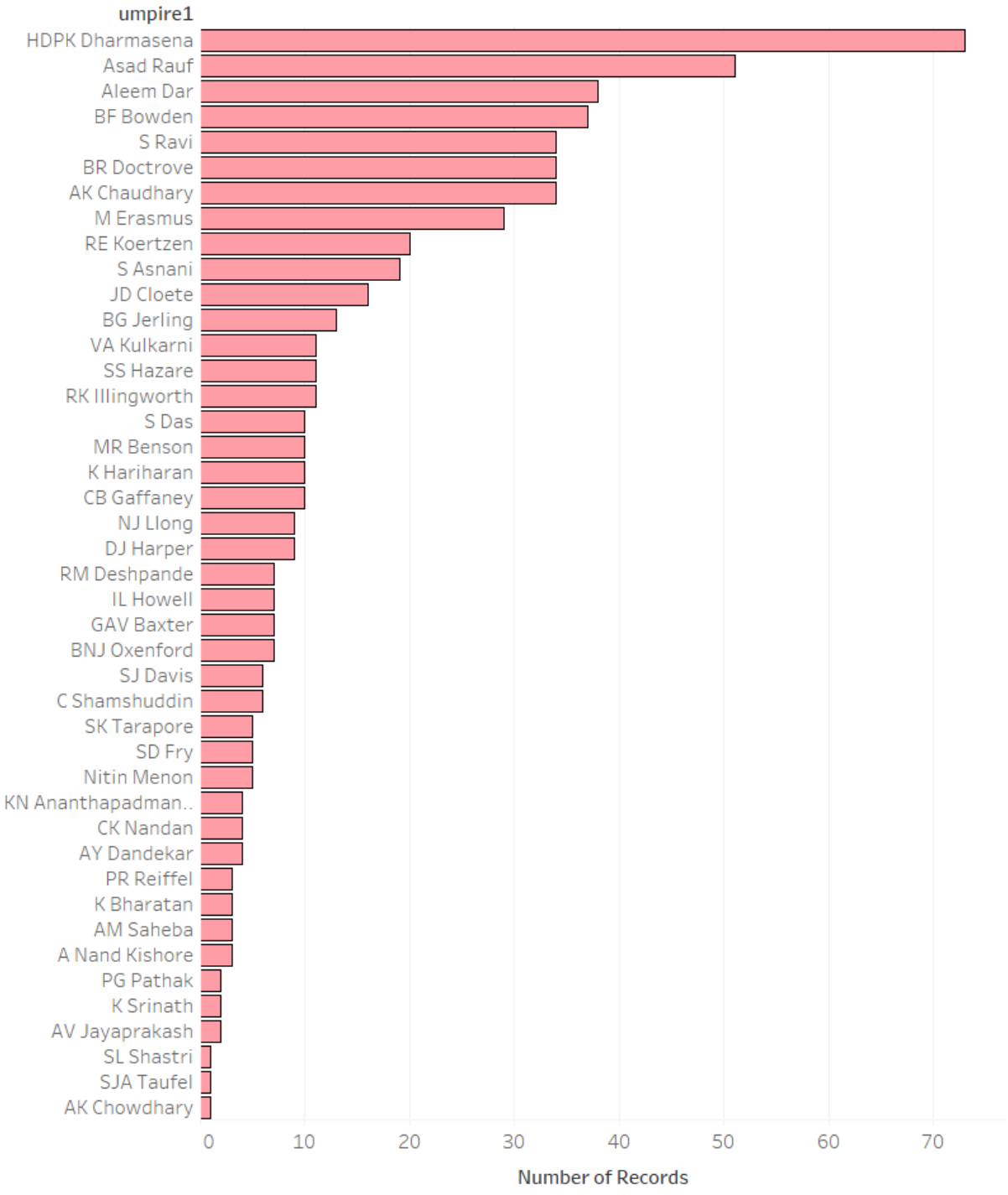


Sum of win_by_wickets for each team2 broken down by team1. Color shows details about toss_decision. Details are shown for winner and toss_winner. The data is filtered on date Year, which keeps 2015.

The above data shows match result among two teams in terms of by how many wickets did they won and what was their toss decision which is also helpful. Toss decisions also helps us in figuring the win percentage as observed in the previous visualisations.

Likely to be the Best Umpire

Inference - Which umpire has umpired how many matches with being HDPK Dharasena the highest and AK Chowdhary the lowest



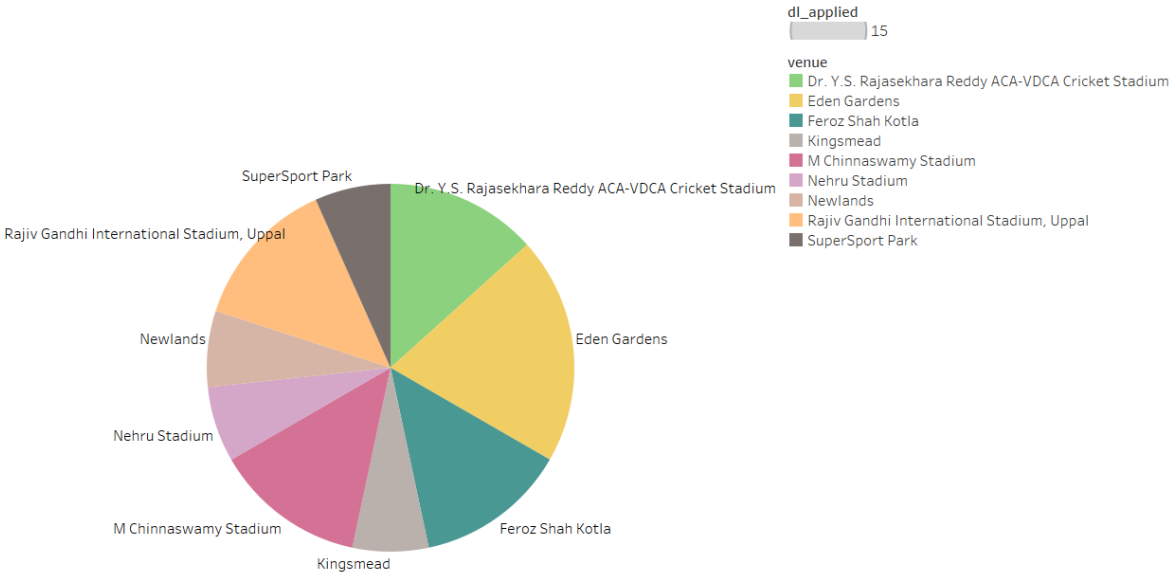
Sum of Number of Records for each umpire1. The data is filtered on date Year, which keeps 9 of 9 members.

This graph tells us about the popularity or experience of the umpire. The more the experience, the more is the accuracy of umpire. By knowing about the best umpires we can use this data in decision making whether to use our DRS at this situation or not. This can be very helpful when the match is in a critical state.

When a match is being conducted amid cloudy area or rainy area, teams need to change their strategy according to the weather. This can help them to save their energy and help them to make through that phase. For example if the weather is like to rain and normal sunny is to come after few hours, so the final decision is made by reducing the number of overs and the runs according to it. So these types of visualisations help teams to make their plans accordingly.

Places of DL applied

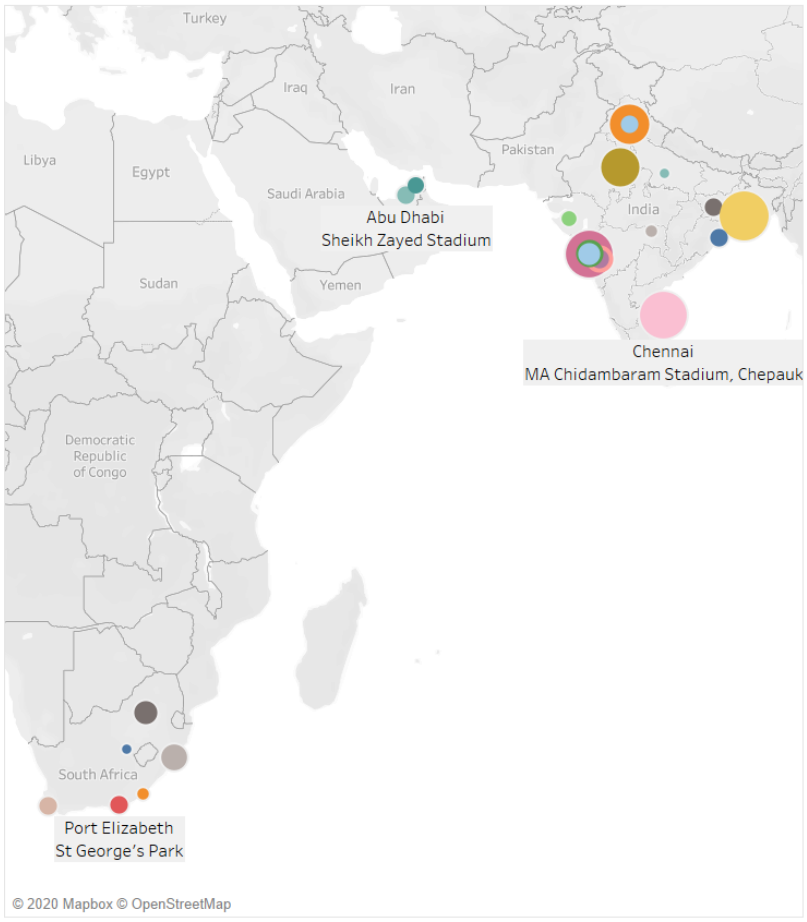
Inference - At which location the match is likely to disrupt because of rain with eden gardens,kolkata being likelist location



Venue. Color shows details about venue. Size shows sum of dl_applied. The marks are labeled by venue. Details are shown for city. The view is filtered on sum of dl_applied, which ranges from 1 to 3.

Venue where Matches were Conducted - Accross Years

Inference - We could see at which all places the ipl matches are conducted and in which all states the cricket stadium are situated



Number of Records
2
10
20
30
40
54

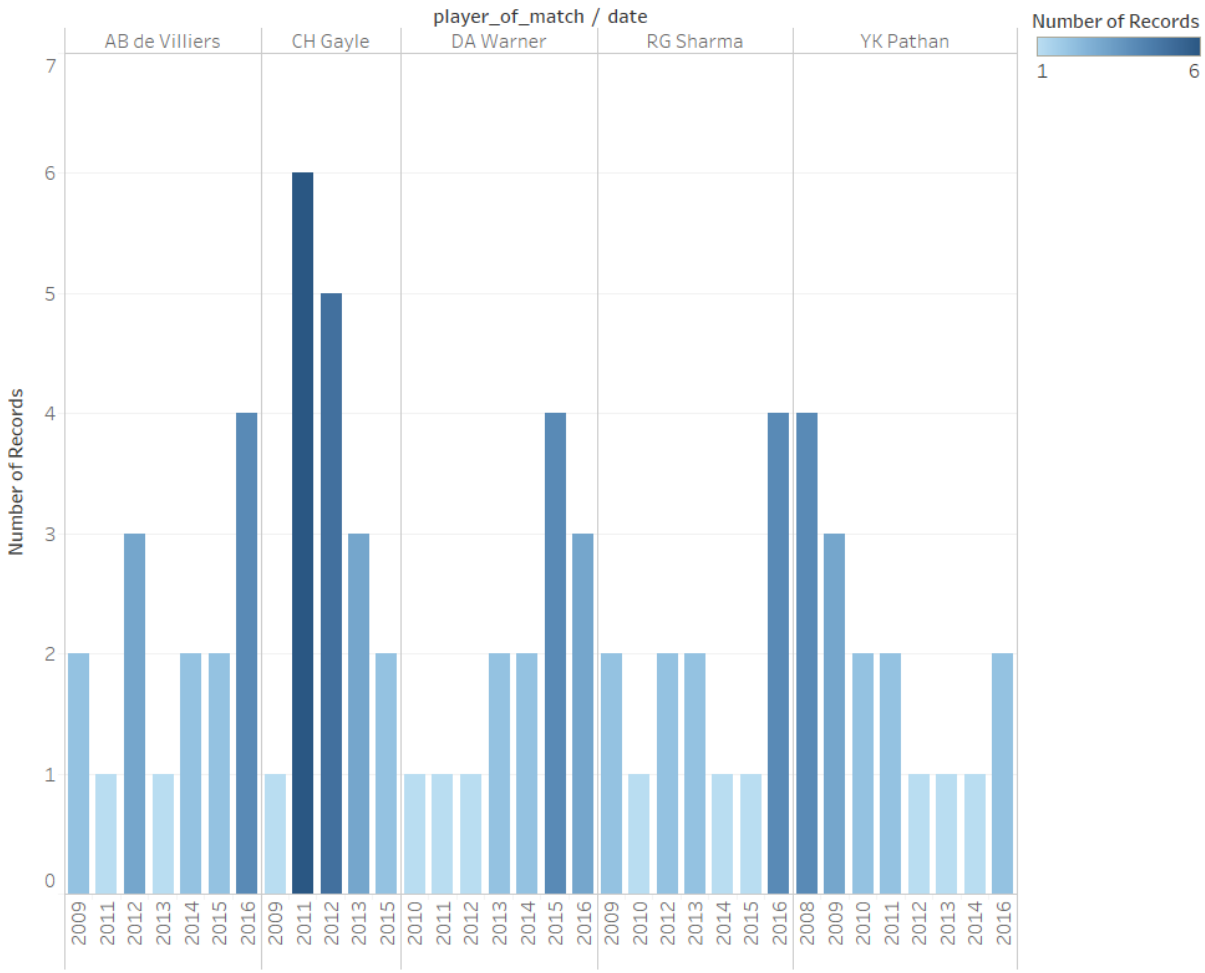
venue
- Barabati Stadium
- Brabourne Stadium
- Buffalo Park
- Dr DY Patil Sports Academy
- Eden Gardens
- Green Park
- JSCA International Stadium Complex
- Kingsmead
- MA Chidambaram Stadium, Chepauk
- Maharashtra Cricket Association Stadium
- Newlands
- OUTsurance Oval
- Punjab Cricket Association IS Bindra Stadium, Mohali
- Punjab Cricket Association Stadium, Mohali
- Saurashtra Cricket Association Stadium
- Sawai Mansingh Stadium
- Sharjah Cricket Stadium
- Sheikh Zayed Stadium
- St George's Park
- Subrata Roy Sahara Stadium
- SuperSport Park
- Vidarbha Cricket Association Stadium, Jamtha
- Wankhede Stadium

Map based on Longitude (generated) and Latitude (generated). Color shows details about venue. Size shows sum of Number of Records. The marks are labeled by city and venue. Details are shown for city. The data is filtered on date Year, which keeps 9 of 9 members. The view is filtered on city and Latitude (generated). The city filter excludes Kochi. The Latitude (generated) filter keeps non-Null values only.

These are the locations where IPL matches are conducted.

Player of Match Award Winner - Top 5

Inference - The teams could predict which player is a threat towards their team also, whether to drop the player in next season or not. CH Gayle being the player of the match 6 times(the highest) during year 2011

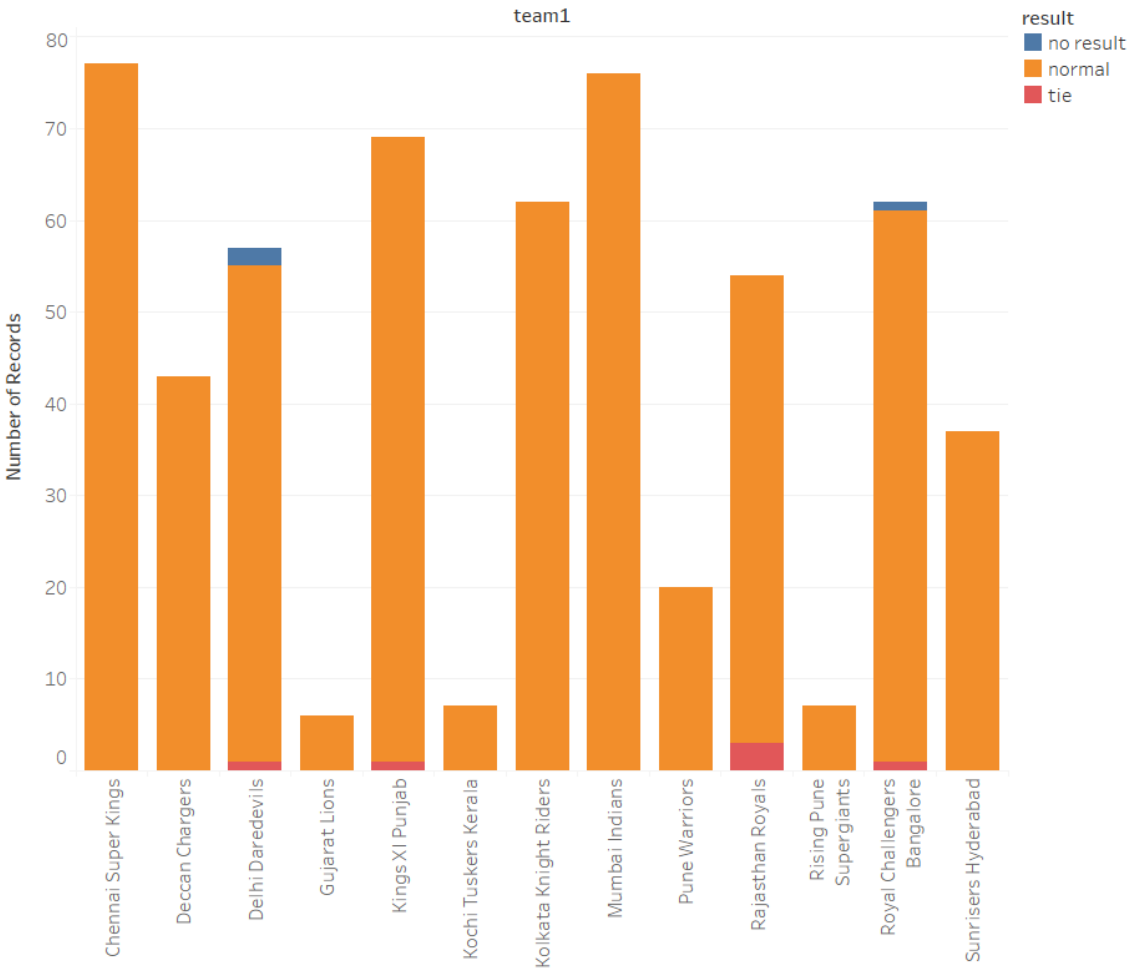


Sum of Number of Records for each date Year broken down by player_of_match. Color shows sum of Number of Records. The view is filtered on player_of_match and date Year. The player_of_match filter has multiple members selected. The date Year filter keeps 9 of 9 members.

These types of data are can be very useful when they know who they are going to face on the ground. They can train themselves and make their strategies accordingly. They can find out th weakness of the player and can make counter moves. For example for a batsmen like CH Gayle who has been the player of the match award winner must have some weak points, like he is not able to run singles and doubles but in a very rare case. Also he is known for scoring sixes on a fast balls so we can use spinners to counter attack him.

No. of matches played by each team

Inference - Which team has played the maximum number of matches with Mumbai Indians being the highest.



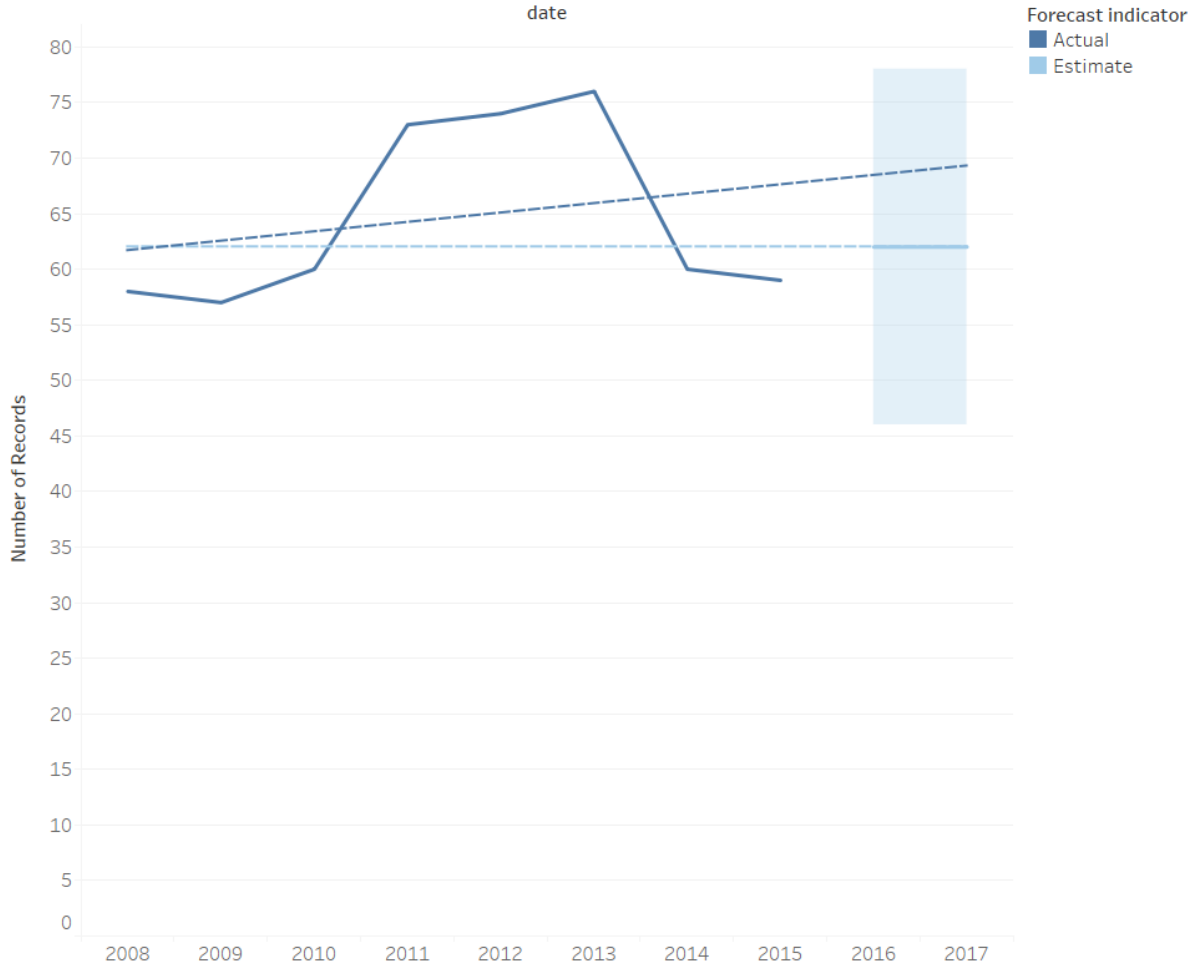
Sum of Number of Records for each team1. Color shows details about result.

Greater the number of matches played, greater is the experice. This is true when it comes to hard work because hardwork always pays off. Similar is with this situation, an experienced team is the one who has made more efforts because this make them to make decisions in every situation. So the percentage of winning is calculated accordingly.

The different colours the result of the matches with Chennai Super Kings (CSK) with highest number of matches played.

Yearwise Record of Number of Matches

Inference - We could see the increase in popularity of the matches with incresing years till 2013 and then the decrease

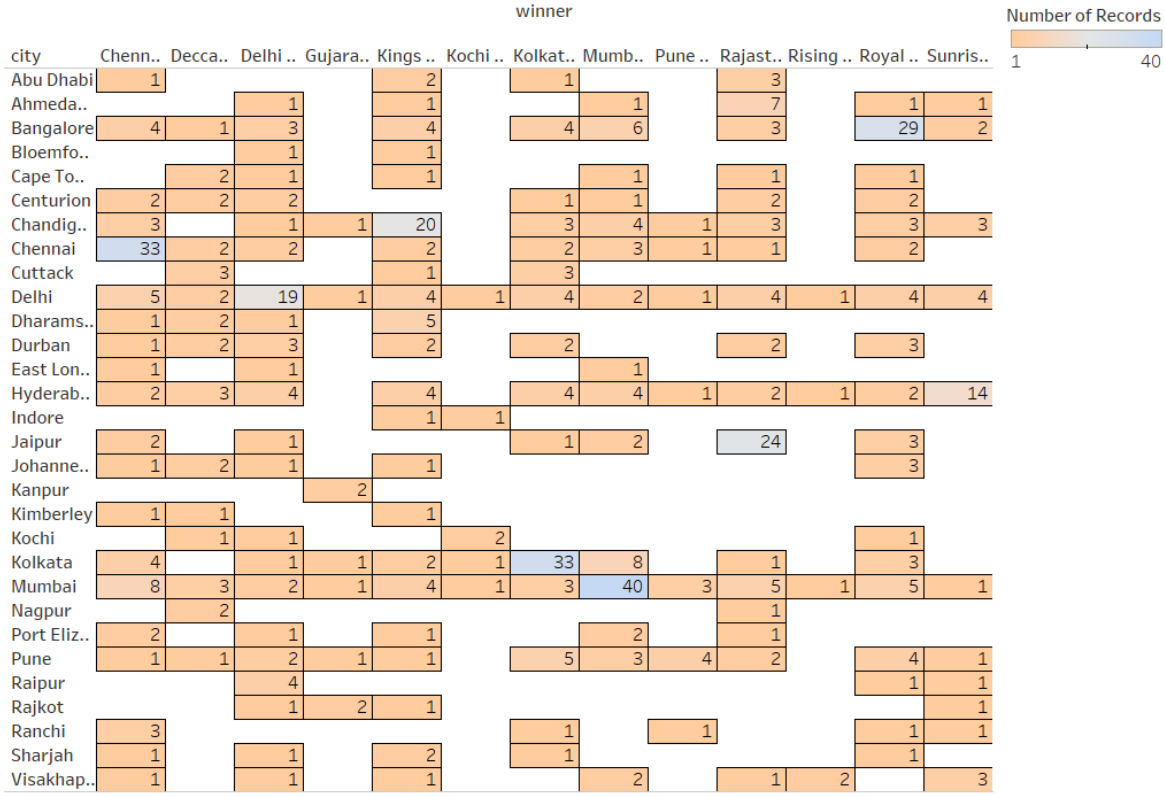


The trend of sum of Number of Records (actual & forecast) for date Year. Color shows details about Forecast indicator.

The above data shows the number of IPL matches played per year. We have created a forrcast that suggests number of record for next two years that will be around 60-65 matches.

Team wins in different city across all seasons

Inference - The teams could predict whether they will be winning/losing the match based on the location with Mumbai Indians winning in Mumbai the maximum number of times



city	Chenn..	Decca..	Delhi ..	Gujara..	Kings ..	Kochi ..	Kolkat..	Mumb..	Pune ..	Rajast..	Rising ..	Royal ..	Sunris..
Abu Dhabi	1				2		1			3			
Ahmeda..			1		1			1		7		1	1
Bangalore	4	1	3		4		4	6		3		29	2
Bloemfo..			1		1								
Cape To..		2	1		1			1		1		1	
Centurion	2	2	2				1	1		2		2	
Chandig..	3		1	1	20		3	4	1	3		3	3
Chennai	33	2	2		2		2	3	1	1		2	
Cuttack		3			1		3						
Delhi	5	2	19	1	4	1	4	2	1	4	1	4	4
Dharams..	1	2	1		5								
Durban	1	2	3		2		2			2		3	
East Lon..	1		1					1					
Hyderab..	2	3	4		4		4	4	1	2	1	2	14
Indore					1	1							
Jaipur	2		1				1	2		24		3	
Johanne..	1	2	1		1							3	
Kanpur				2									
Kimberley	1	1			1								
Kochi		1	1			2						1	
Kolkata	4		1	1	2	1	33	8		1		3	
Mumbai	8	3	2	1	4	1	3	40	3	5	1	5	1
Nagpur		2								1			
Port Eliz..	2		1		1			2		1			
Pune	1	1	2	1	1		5	3	4	2		4	1
Raipur			4									1	1
Rajkot			1	2	1								1
Ranchi	3						1		1			1	1
Sharjah	1		1		2		1					1	
Visakhap..	1		1		1			2		1	2		3

Sum of Number of Records broken down by winner vs. city. Color shows sum of Number of Records. The marks are labeled by sum of Number of Records. The view is filtered on winner and city. The winner filter excludes Null. The city filter excludes Null.

This data can be used in decision making whether to play aggressive or normal. The above data shows which city is most likely to prove a victory for the teams and this can be a plus point for them.