

DATA ANALYSIS AND VISUALIZATION

PROJECT

TOPIC : INDIAN PREMIER LEAGUE ANALYSIS USING PYTHON & TABLEAU

**SAHIL SANJAY LANDGE
RAHUL VIJAY KHUBCHANDANI**



Rutgers Business School – Newark and New Brunswick

Dataset Source : <https://www.kaggle.com/datasets/manasgarg/ipl>

Number of variables: 37

The quality of data is good as most rows and columns have data which is correct and can be used. Also, we encountered very few NaN values.

Objective :

The objective is to create an interactive dashboard in Tableau that focuses on analyzing the IPL data set and creating visualizations upon conclusions.

Analytics software going to be used : Python & Tableau

EDA :

Here we imported all the required libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

Read the csv file and used head command to use initial few rows with all columns

```
matches_data = pd.read_csv("matches.csv")
```

```
matches_data.head()
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2	umpire3
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Ulong	NaN
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	S Ravi	NaN
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK Nandan	NaN
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	Gl Maxwell	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin	NaN
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium	NaN	NaN	NaN

Read the csv file and used head command to use initial few rows with all columns

```
In [5]: deliveries_data = pd.read_csv("deliveries.csv")
```

```
In [6]: deliveries_data.head()
```

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs	penalty_runs	batsman_runs	extra_runs	total_runs	player_dismissed	dismissal_kind
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0	NaN	NaN
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0	NaN	NaN
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	4	0	4	NaN	NaN
3	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	0	0	NaN	NaN
4	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0	0	0	2	2	NaN	NaN

5 rows × 21 columns

Now we will merge season data

```
In [8]: season_data=matches_data[['id','season','winner']]
```

```
complete_data=deliveries_data.merge(season_data,how='inner',left_on='match_id',right_on='id')
```

```
In [9]: matches_data.columns.values
```

```
Out[9]: array(['id', 'season', 'city', 'date', 'team1', 'team2', 'toss_winner',
   'toss_decision', 'result', 'dl_applied', 'winner', 'win_by_runs',
   'win_by_wickets', 'player_of_match', 'venue', 'umpire1', 'umpire2',
   'umpire3'], dtype=object)
```

As we can see the Umpire 3 column value is not needed we will clean the data by removing the same by using .drop

```
In [11]: matches_data = matches_data.drop(columns=["umpire3"],axis=1)
matches_data.head()
```

	id	season	city	date	team1	team2	toss_winner	toss_decision	result	dl_applied	winner	win_by_runs	win_by_wickets	player_of_match	venue	umpire1	umpire2
0	1	2017	Hyderabad	2017-04-05	Sunrisers Hyderabad	Royal Challengers Bangalore	Royal Challengers Bangalore	field	normal	0	Sunrisers Hyderabad	35	0	Yuvraj Singh	Rajiv Gandhi International Stadium, Uppal	AY Dandekar	NJ Llong
1	2	2017	Pune	2017-04-06	Mumbai Indians	Rising Pune Supergiant	Rising Pune Supergiant	field	normal	0	Rising Pune Supergiant	0	7	SPD Smith	Maharashtra Cricket Association Stadium	A Nand Kishore	S Ravi
2	3	2017	Rajkot	2017-04-07	Gujarat Lions	Kolkata Knight Riders	Kolkata Knight Riders	field	normal	0	Kolkata Knight Riders	0	10	CA Lynn	Saurashtra Cricket Association Stadium	Nitin Menon	CK Nandan
3	4	2017	Indore	2017-04-08	Rising Pune Supergiant	Kings XI Punjab	Kings XI Punjab	field	normal	0	Kings XI Punjab	0	6	GJ Maxwell	Holkar Cricket Stadium	AK Chaudhary	C Shamshuddin
4	5	2017	Bangalore	2017-04-08	Royal Challengers Bangalore	Delhi Daredevils	Royal Challengers Bangalore	bat	normal	0	Royal Challengers Bangalore	15	0	KM Jadhav	M Chinnaswamy Stadium	NaN	NaN

```
In [15]: matches_data.to_csv('matchescleaned.csv')
```

```
In [16]: deliveries_data.to_csv('deliveriescleaned.csv')
```

We have exported the cleaned data to visualize the same on Tableau.

Types of analysis & visualizations on data :

Now we want to find the data related to sixes. Here we got the maximum number of six hit by a team.

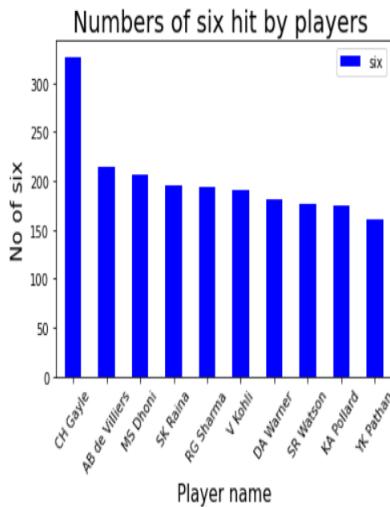
```
In [17]: six_data=complete_data[complete_data['batsman_runs']==6]
six_data.groupby('batting_team')[['batsman_runs']].agg([('runs by six','sum'),('sixes','count')])
```

```
Out[17]: runs by six  sixes
```

batting_team	runs by six	sixes
Chennai Super Kings	5838	973
Deccan Chargers	2400	400
Delhi Capitals	522	87
Delhi Daredevils	4806	801
Gujarat Lions	930	155
Kings XI Punjab	5856	976
Kochi Tuskers Kerala	318	53
Kolkata Knight Riders	5580	930
Mumbai Indians	6576	1096
Pune Warriors	1176	196
Rajasthan Royals	4086	681
Rising Pune Supergiant	534	89
Rising Pune Supergiants	408	68
Royal Challengers Bangalore	6792	1132
Sunrisers Hyderabad	3198	533

Here we can see the top 10 players who have hit the maximum number of sixes in IPL

```
In [18]: batsman_six=six_data.groupby('batsman')[['batsman_runs']].agg([('six','count')]).reset_index().sort_values('six',ascending=0)
ax=batsman_six.iloc[:10,:].plot('batsman','six',kind='bar',color='blue')
plt.title("Numbers of six hit by players ",fontsize=20)
plt.xticks(rotation=50)
plt.xlabel("Player name",fontsize=15)
plt.ylabel("No of six",fontsize=15)
plt.show()
```



Now we have to find top 5 batsman who have scored maximum runs in the IPL

```
In [24]: batsman_score=deliveries_data.groupby('batsman')[['batsman_runs']].agg(['sum']).reset_index().sort_values('sum',ascending=False).reset_index(drop=True)
batsman_score=batsman_score.rename(columns={'sum':'batsman_runs'})
print("*** Top 5 Leading Run Scorer in IPL ***")
batsman_score.iloc[:5,:]
```

*** Top 5 Leading Run Scorer in IPL ***

Out[24]:

	batsman	batsman_runs
0	V Kohli	5434
1	SK Raina	5415
2	RG Sharma	4914
3	DA Warner	4741
4	S Dhawan	4632

Here we find the highest wicket takers in IPL

```
In [26]: wicket_data=deliveries_data.dropna(subset=['dismissal_kind'])
wicket_data=wicket_data[wicket_data['dismissal_kind'].isin(['run out','retired hurt','obstructing the field'])]
```

```
In [27]: wicket_data.groupby('bowler')['dismissal_kind'].agg(['count']).reset_index().sort_values('count',ascending=False).reset_index(drop=True).iloc[:5,:]
```

```
Out[27]:   bowler  count
```

0	SL Malinga	170
1	A Mishra	156
2	Harbhajan Singh	150
3	PP Chawla	149
4	DJ Bravo	147

DASHBOARDS :

DASHBOARD 1 - BATTING ANALYSIS

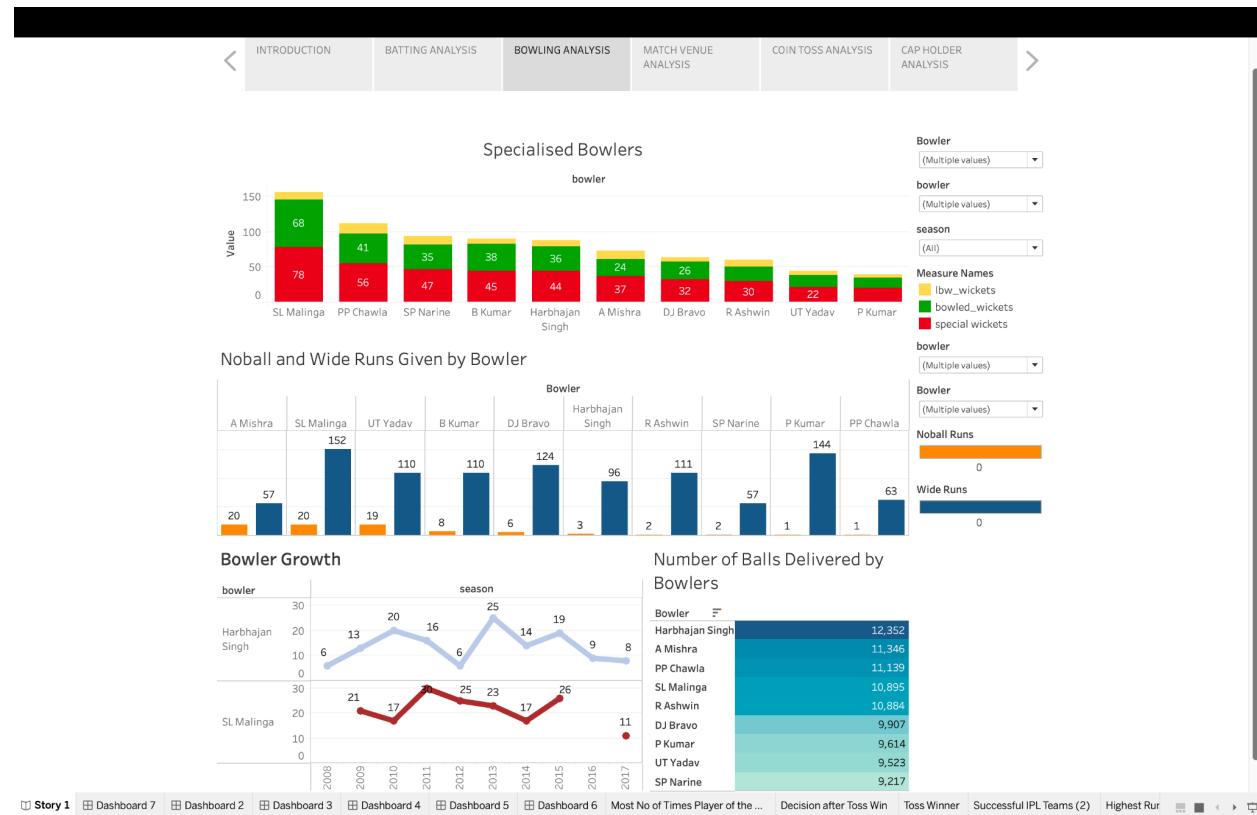


Here in the above dashboard we are analyzing batsmen on the basis of their growth, number of sixes scored, type of dismissal, highest run scored.

Insights -

- From the Batsman growth chart, we can see that DA Warner has a good increasing performance in the matches played over all in making runs. V Kohli was averaging decently over the years but his performance has drastically decreased after 2016.
- From the Max number of sixes by Batsman chart, we can see the players with the number of sixes they have scored over all during all seasons. CH Gayle has topped the list by scoring highest number scores by scoring total 344 sixes over all.
- From the Kind of Dismissal of Batsmen chart, we can see how each player was out by different forms like bowled, caught, run out & many more. V. Kohli was bowled 30 times, caught 95 times & run out 15 times during all the matches he played.
- From the Highest Run Scoring Batsman chat, we can see that on the top left most corner is the player who has scored the most number of runs (i.e V. Kohli) & on the bottom right corner is the player who has scored the least number of runs in over all matches(i.e Abdur Razzak).

DASHBOARD 2 - BOWLING ANALYSIS

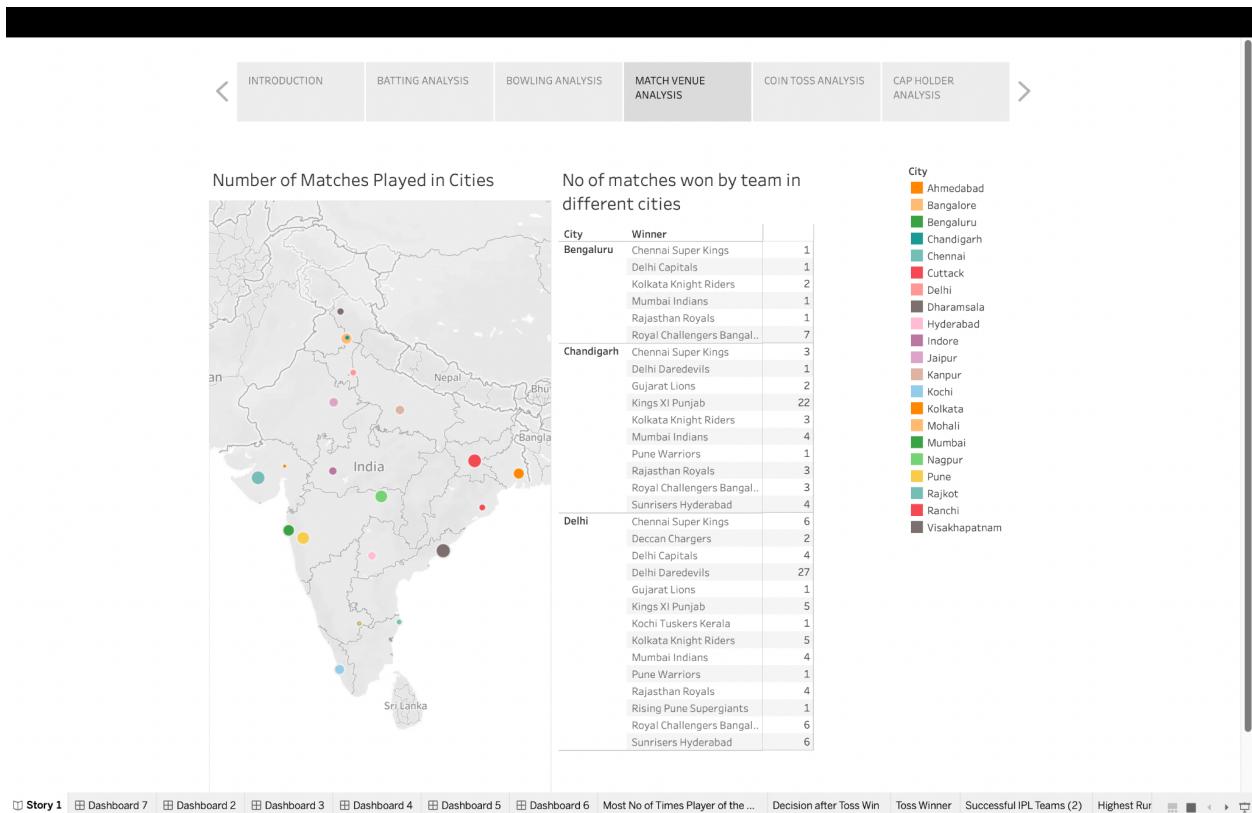


Here in the above chart we are analyzing bowlers based on number of wickets they have taken in different forms, no & wide ball runs given, highest number of balls delivered, growth.

Insights -

- From the Specialized bowlers chart, we can see analysis on how a player has taken different wickets in different forms like lbw, bowled, special. SL Malinga has topped the chart in taking the highest number of wickets and to be specific 10 by lbw and 68 bowled.
- From the No & Wide Ball Runs chart, we can see the players who gave runs in the form of no or wide ball deliveries. For instance, A Mishra has give 57 wide runs & 20 no ball runs.
- From the bowler growth chart, we can analyze performance of players overall in the matches they have bowled which tells us how good or bad they have made their game over the years. H Singh has a seasonal trend in all the matches where we can see that his performance is not constantly increasing/decreasing but is full of ups and downs. The best he had done was taking 25 wickets in some season.
- From the number of balls delivered by the bowlers chart, we have created a table which shows different player with respect to the delivers they have made over all in descending order.

DASHBOARD 3 - MATCH VENUE ANALYSIS



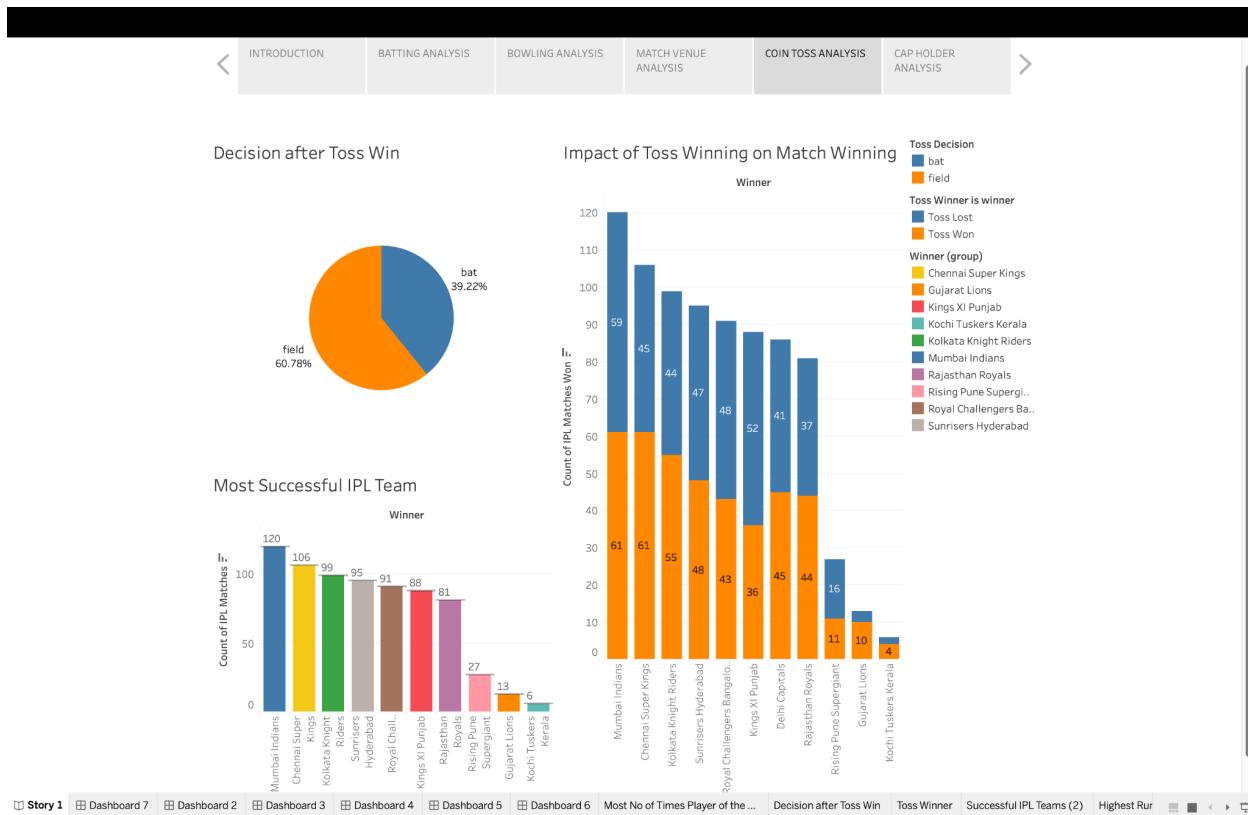
Story 1 | Dashboard 7 | Dashboard 2 | Dashboard 3 | Dashboard 4 | Dashboard 5 | Dashboard 6 | Most No of Times Player of the ... | Decision after Toss Win | Toss Winner | Successful IPL Teams (2) | Highest Rur

Here in the above chart we are analyzing how many different cities of India matches were held and in those cities where matches were held which teams have won matches on the field of those cities.

Insights -

- From the Number of Matches Played in Cities chart, Mumbai was the city with the highest number of games (101) and Kanpur was the city with the lowest number of games (4) played.
- From the number of matches won by a team in different cities chart, we can see how many times different teams have won matches in different cities. In Delhi, Delhi Daredevils have won the highest number of matches which is 27. In Chandigarh, Pune Warriors & Delhi Daredevils have won the lowest number of matches.

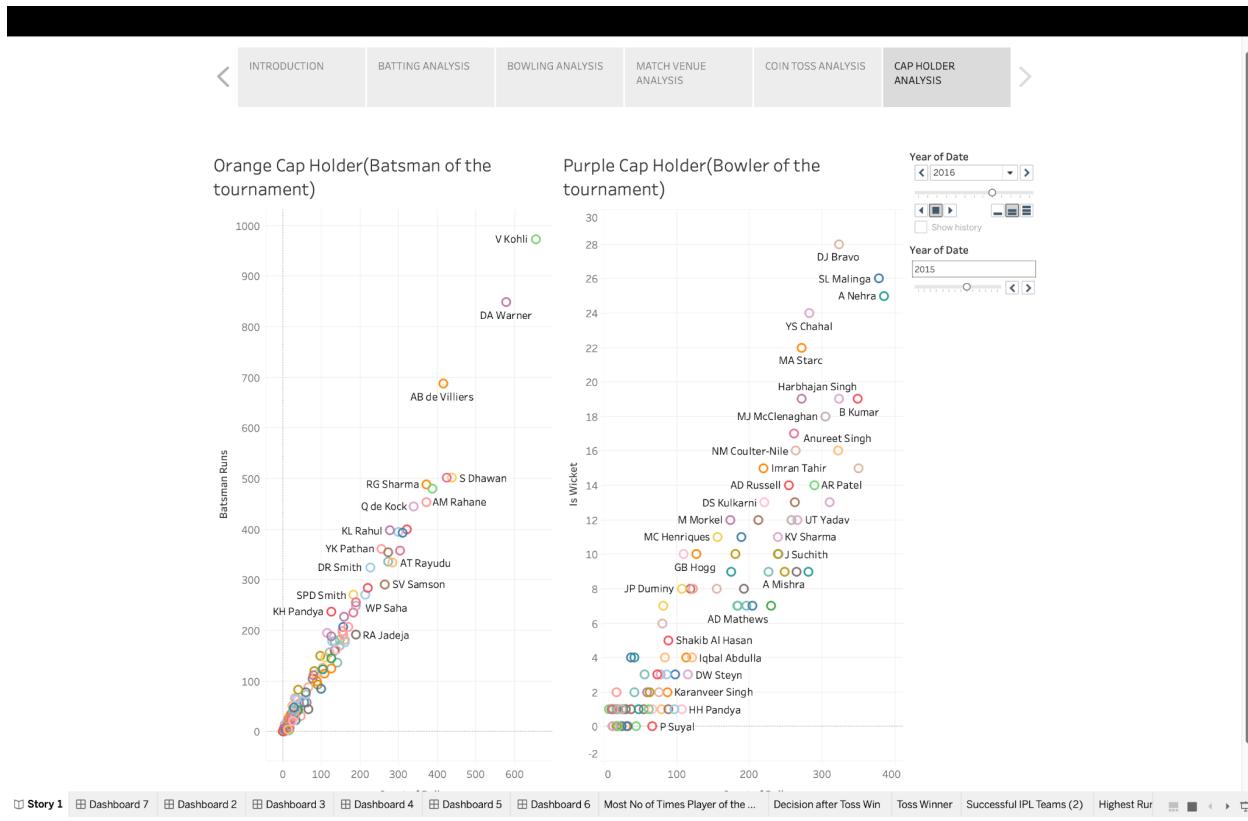
DASHBOARD 4 - COIN TOSS ANALYSIS



Here in the above chart we are analyzing the toss decision, the impact it has on different teams & the result of the matches.

- From the Decision after Toss Win chart, we can understand that the percentage of choosing the field is more than bat.
- From the Impact of Toss Winning on Match Winning chart, Mumbai has won the maximum number of tosses (61), Chennai Super Kings have also won the same number of tosses & so on for other teams. Now, from the Most Successful IPL Team chart, Mumbai Indians is leading with the number of seasons won followed by Chennai Super Kings and so on. It basically implies that the team who has won the most number of tosses is most likely to win the match.

DASHBOARD 5 - CAP HOLDER ANALYSIS



Here in the above chart we are analyzing the bowlers & batsmen who got purple & orange cap based on their performance each year.

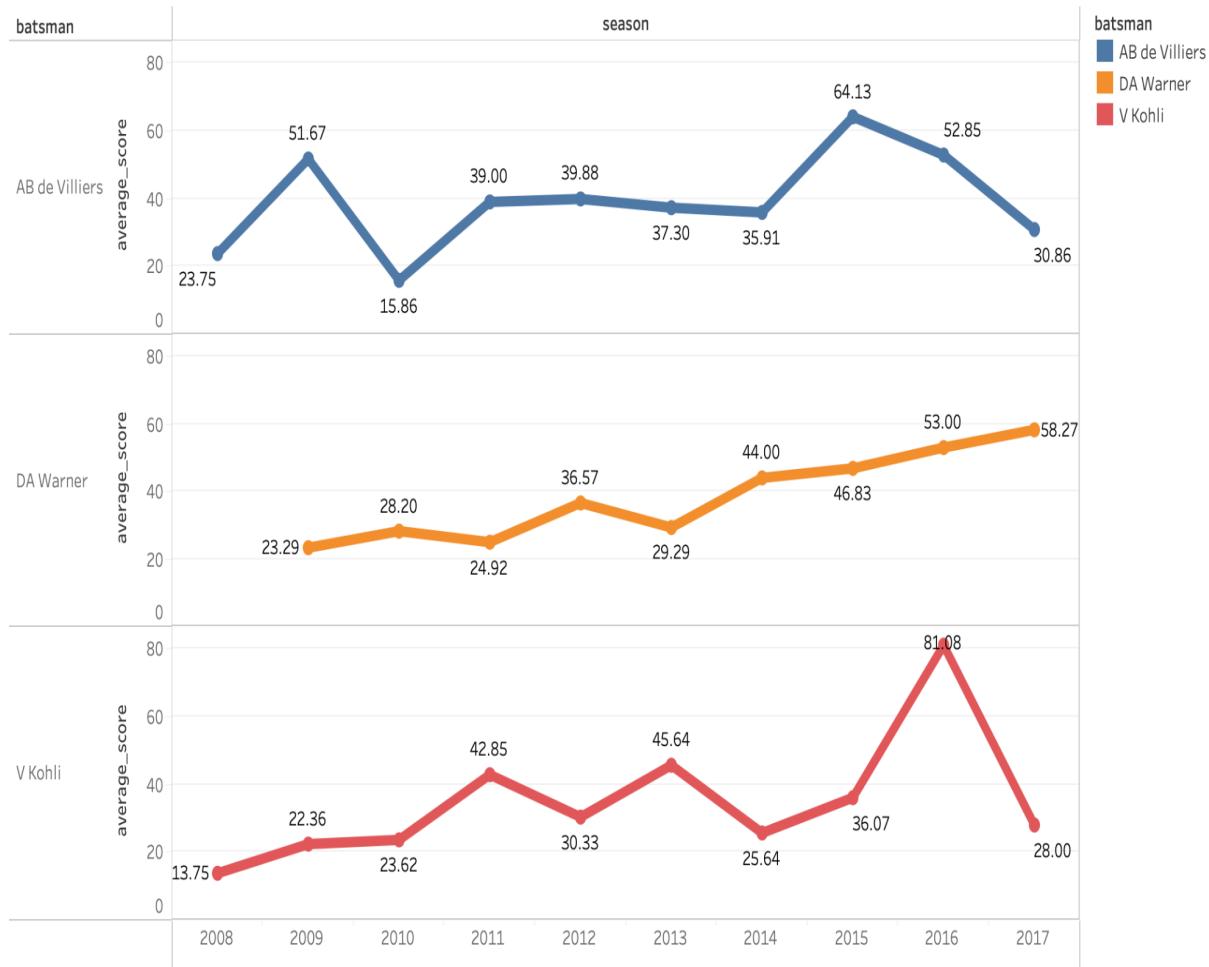
Insights -

- From the Orange Cap Holder chart, we see the player that has scored the highest numbers of runs each year and received the cap in honor. V. Kohli got the Orange Cap in 2016.
- From the Purple Cap Holder chart, we see the player that has taken the highest numbers of wickets each year and received the cap in honor. DJ Bravo got the Purple Cap in 2015..

INDIVIDUAL INSIGHTS OF EACH CHART AS FOLLOWS -

CHART 1 :

Batsman growth

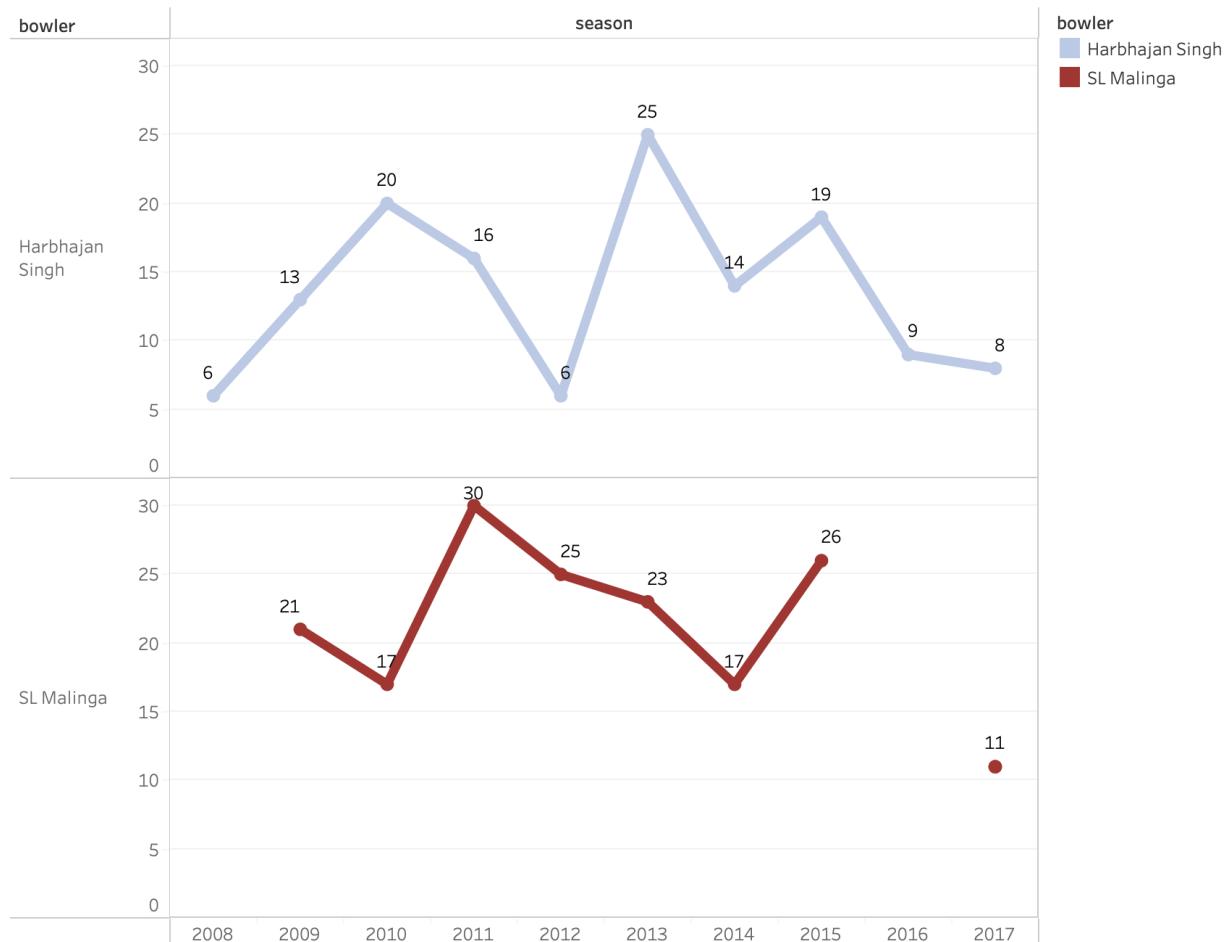


The trend of average_score for season broken down by batsman. Color shows details about batsman. The view is filtered on batsman and average_score. The batsman filter keeps AB de Villiers, DA Warner and V Kohli. The average_score filter keeps non-Null values only.

The above line graph illustrates the batting growth of three random players and their performances every year. We can see the averages of the runs scored every year by these players

CHART 2 :

Bowler Growth

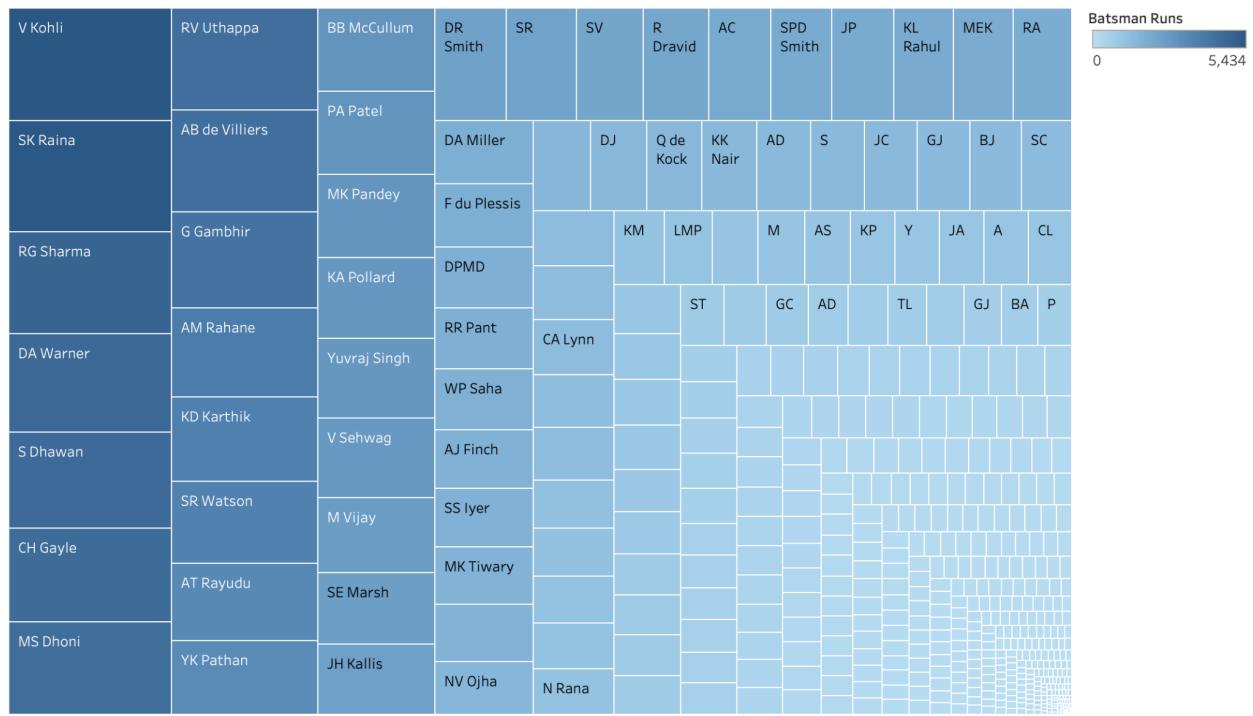


The trend of Sum_of_wickets for season broken down by bowler. Color shows details about bowler. The view is filtered on bowler, which keeps Harbhajan Singh and SL Malinga.

The above line graph illustrates the bowling growth of 2 random bowlers and their performances every year on how many wickets they have taken.

CHART 3 :

Highest Runs Scoring Batsman

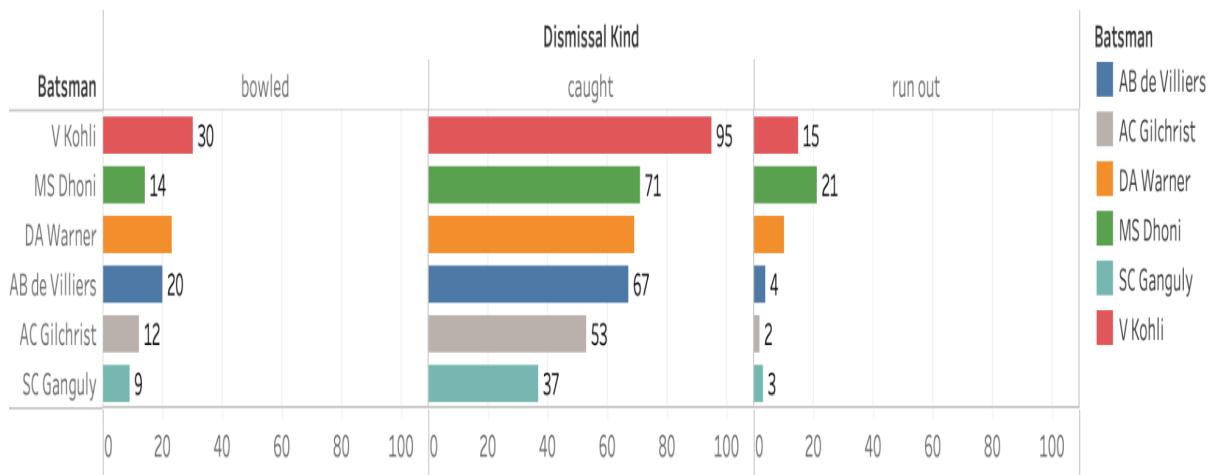


Batsman. Color shows sum of Batsman Runs. Size shows sum of Batsman Runs. The marks are labeled by Batsman.

The above chart illustrates runs scored by different players. In which we can see that the top left corner player has scored the highest number of runs whereas the bottom right corner player has scored the least number of runs.

CHART 4 :

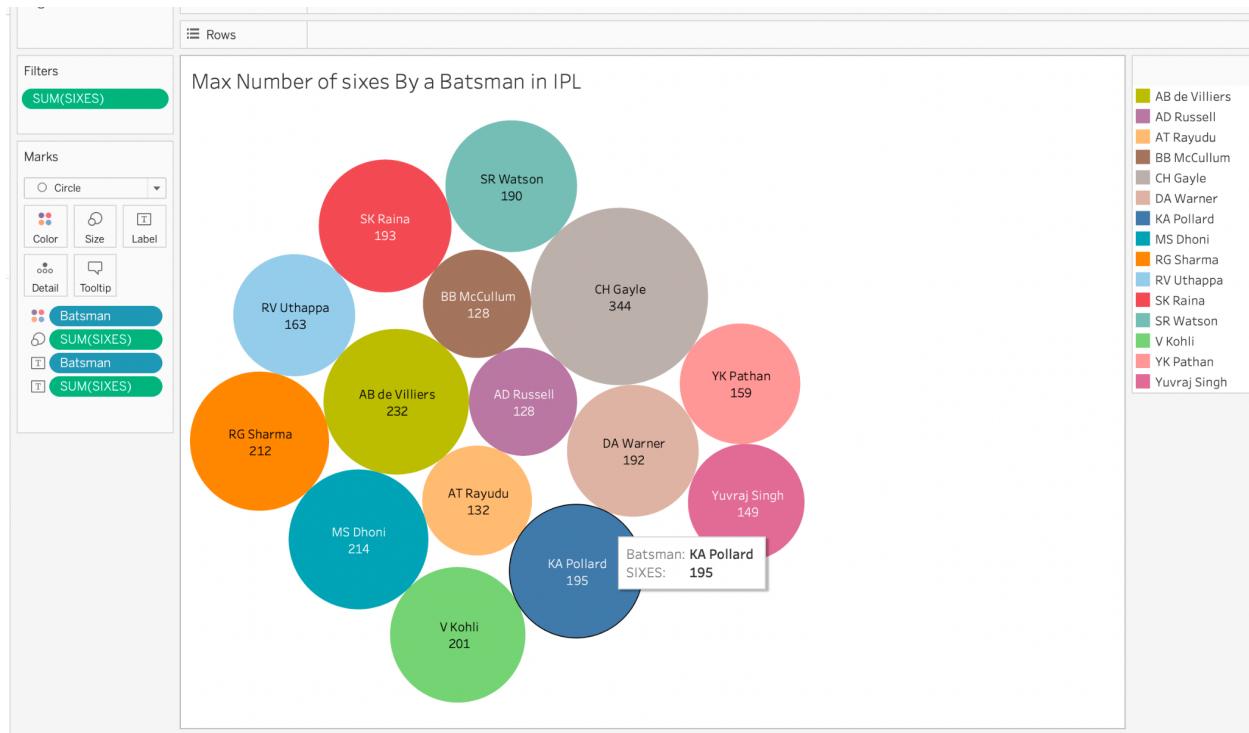
Kind of Dismissal of Batsmen



Count of Dismissal Kind for each Batsman broken down by Dismissal Kind. Color shows details about Batsman. The view is filtered on Batsman and Dismissal Kind. The Batsman filter keeps 6 of 516 members. The Dismissal Kind filter keeps bowled, caught and run out.

The above chart illustrates the kind of dismissal of the batsman. As we can see MS Dhoni lost his wicket by being bowled 14 times, being caught out 71 times and getting run-out 21 times.

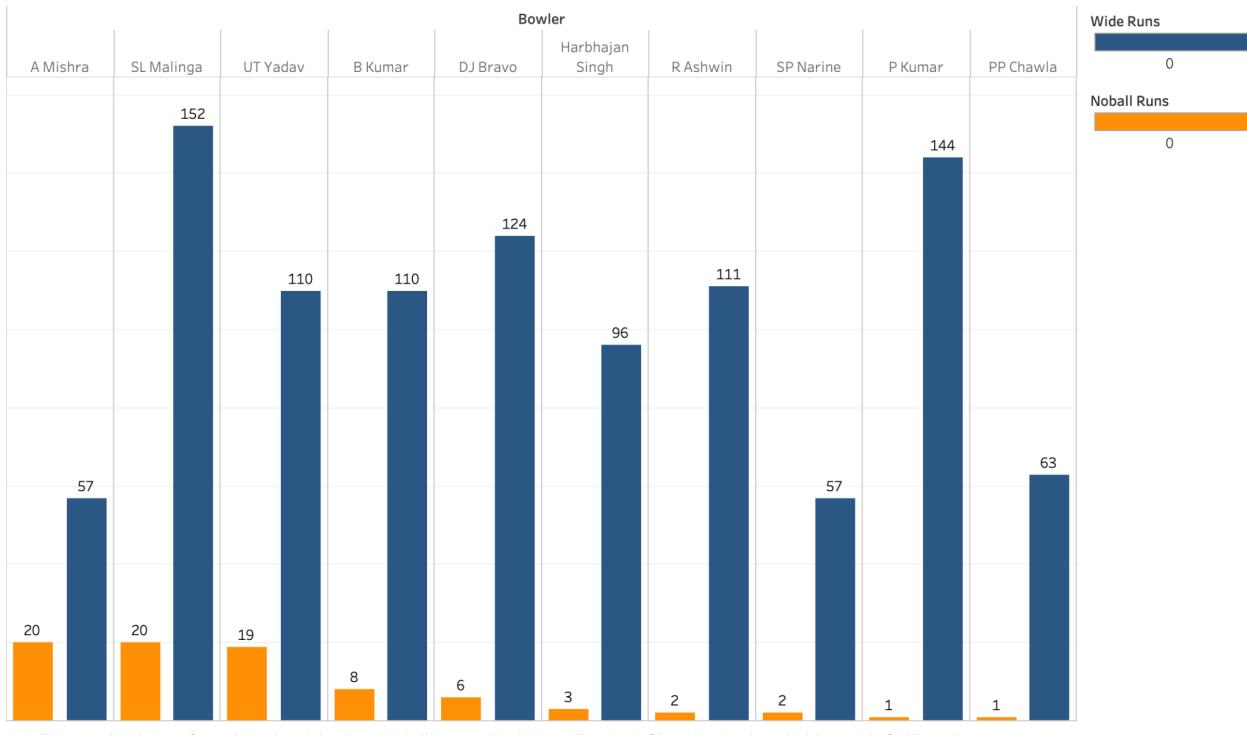
CHART 5 :



The above chart illustrates the maximum number of sixes hit by a batsman in the IPL. As we can clearly see, Chris Gayle is leading this record with 344 sixes.

CHART 6 :

Noball and Wide Runs Given by Bowler

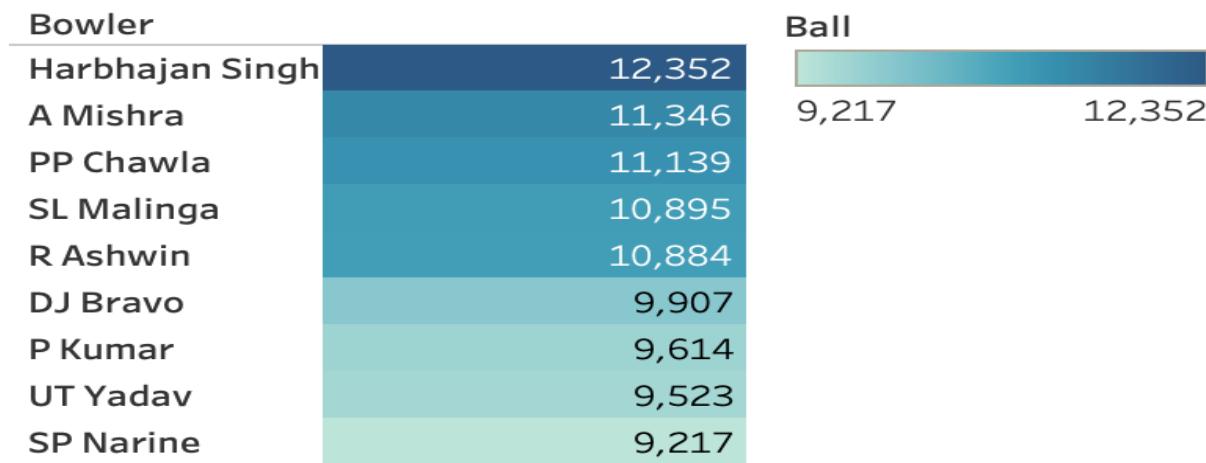


Noball Runs and Wide Runs for each Bowler. Color shows Noball Runs and Wide Runs. The view is filtered on Bowler, which keeps 10 of 405 members.

The above chart illustrates the number of no-ball and wide runs delivered by the bowler.

CHART 7 :

Number of Balls Delivered by Bowlers

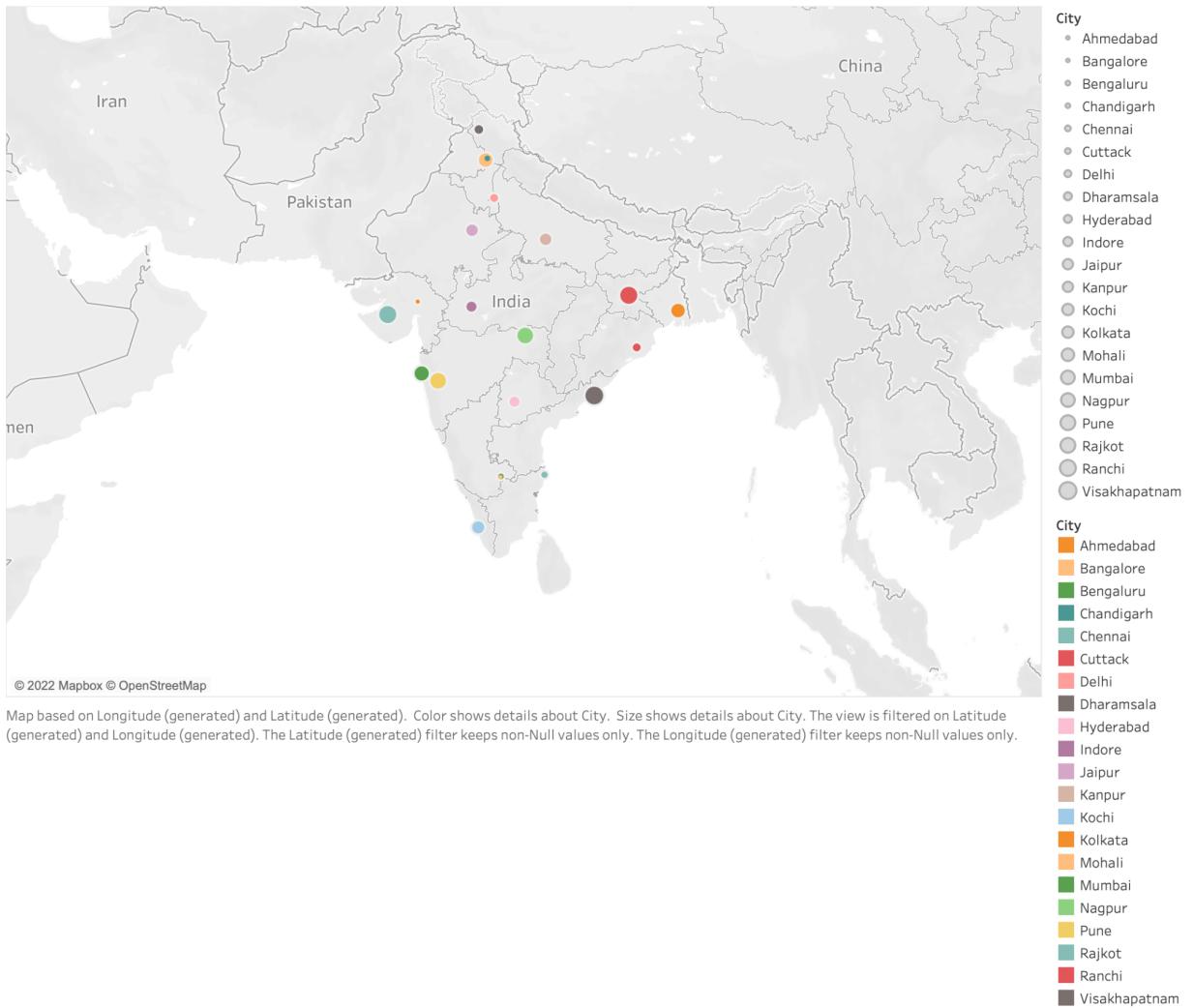


Sum of Ball broken down by Bowler.
Color shows sum of Ball. The marks are labeled by sum of Ball. The view is filtered on Bowler, which keeps 9 of 405 members.

The above chart illustrates the number of balls delivered till date in the IPL. As we can see Harbhajan Singh has delivered the most number of balls (12352).

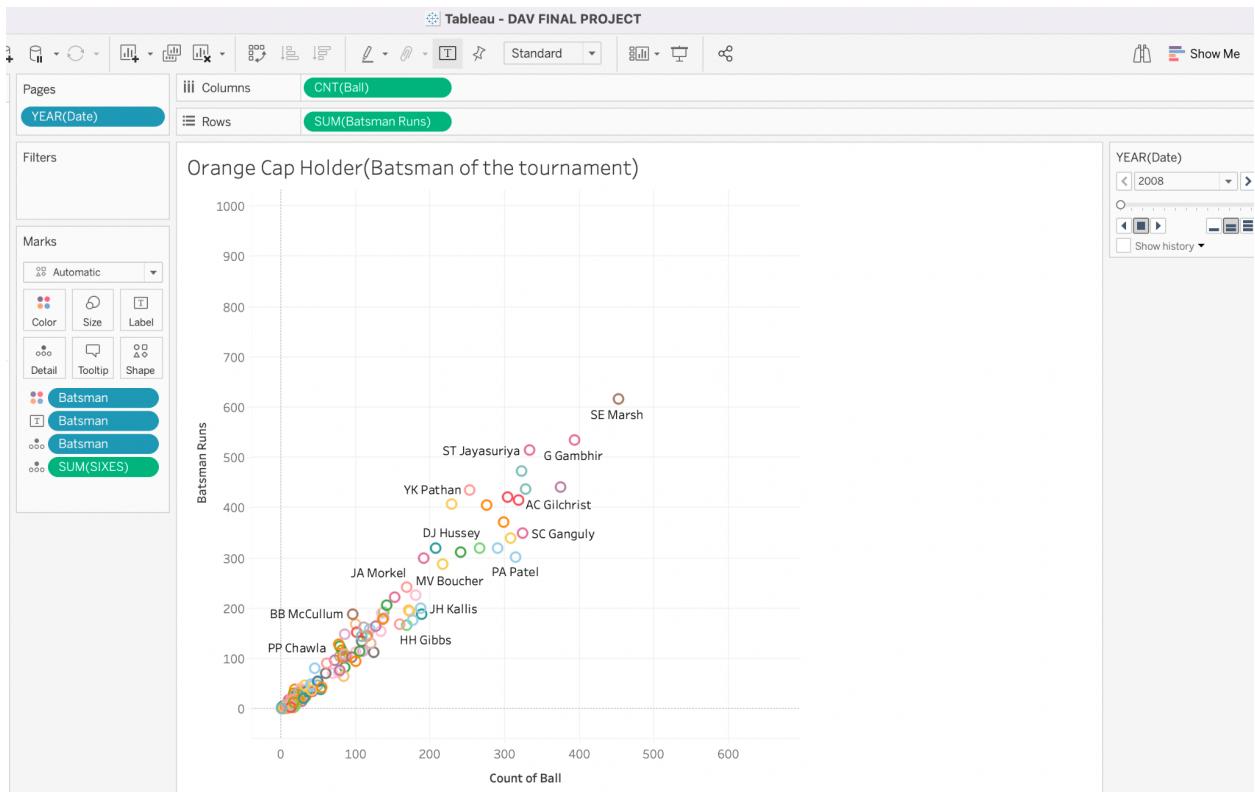
CHART 8:

Number of Matches Played in Cities



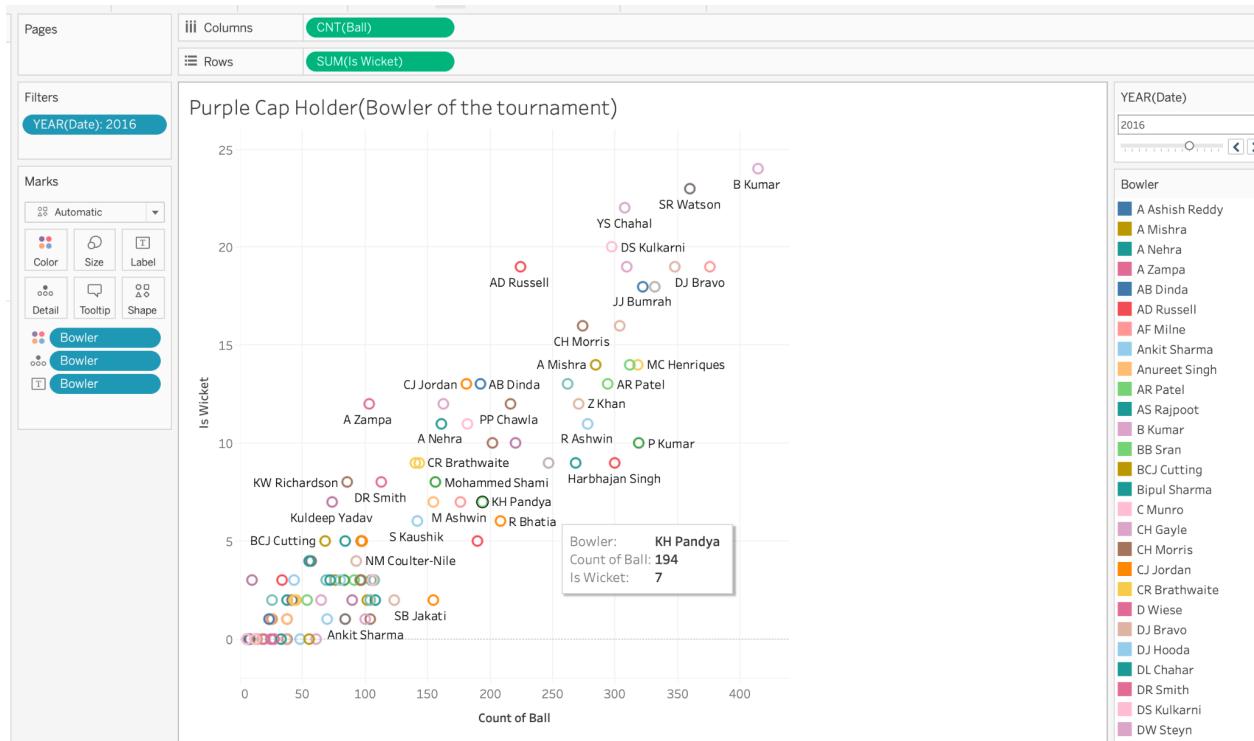
The above chart illustrates the number of matches played at each city in India.

CHART 9:



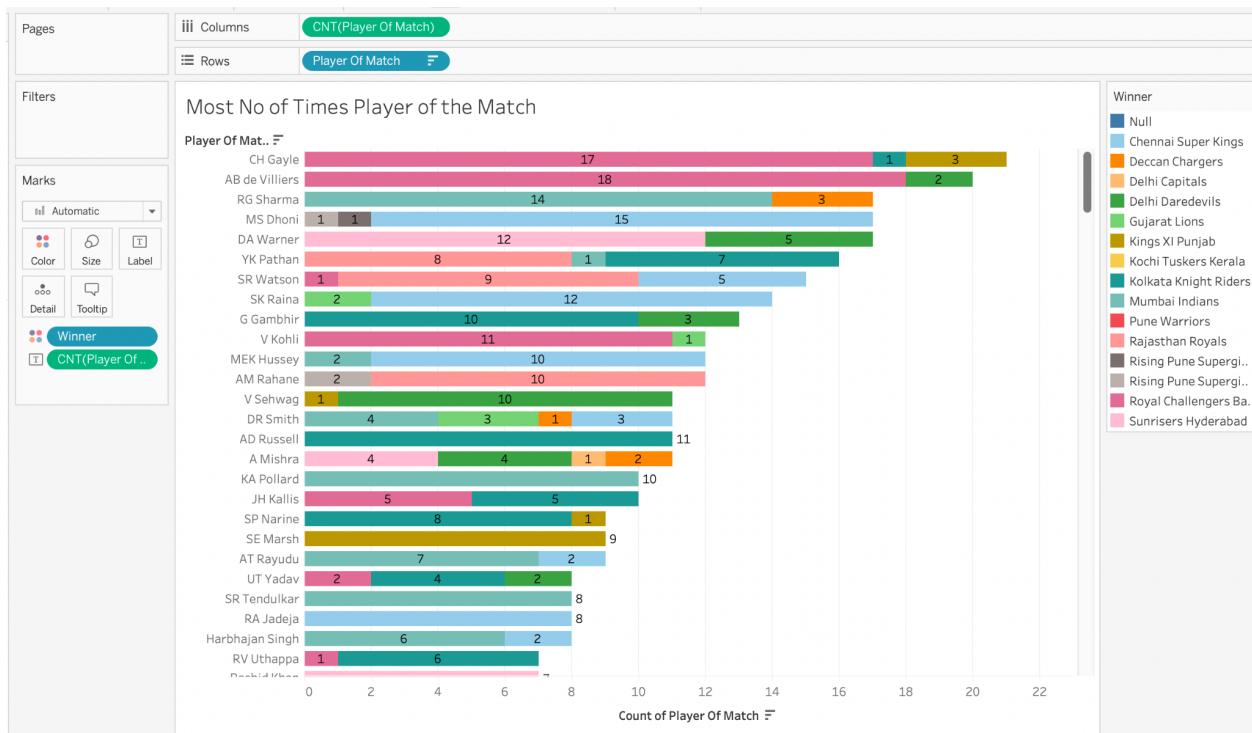
The above chart illustrates the orange cap holders of each season. As we can see above Shaun Marsh got the orange cap in the year 2008 for scoring most runs in that season.

CHART 10:



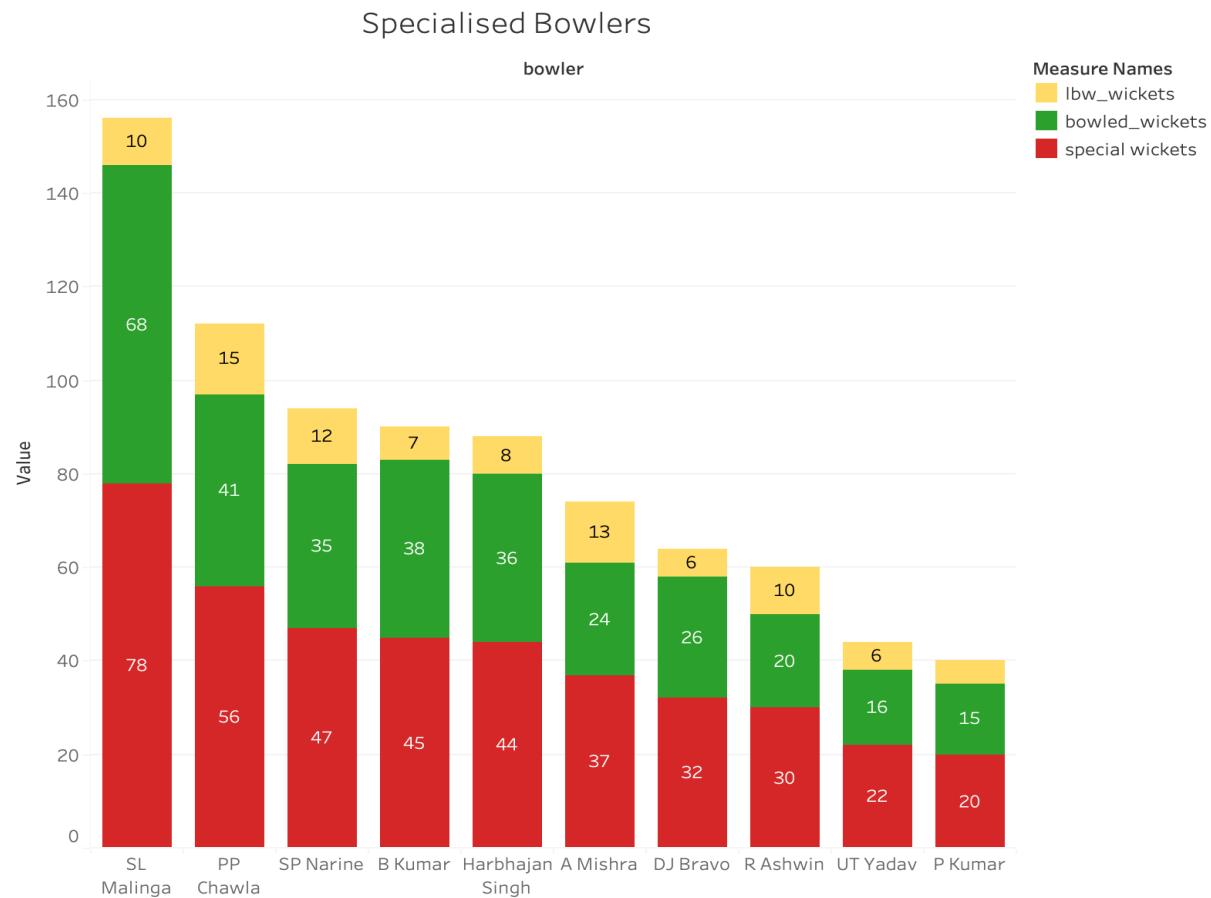
The above chart illustrates the purple cap holders of each season. As we can see above Bhuvaneshwar Kumar got the purple cap in the year 2016 for taking most wickets in that season.

CHART 11 :



The above chart illustrates the most number of times a player won the man of the match award. As we can see, Chris Gayle has won most MOM awards.

CHART 12 :

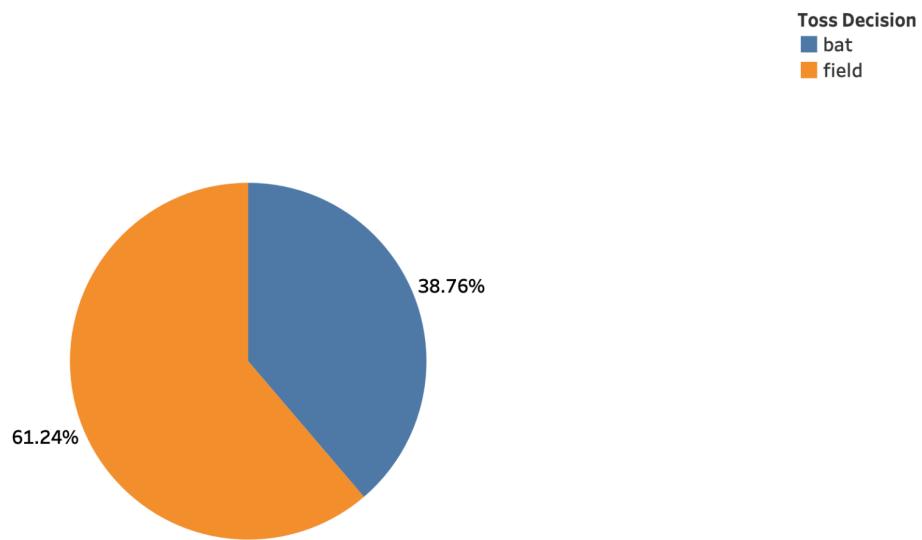


Lbw_wickets, bowled_wickets and special wickets for each bowler. Color shows details about lbw_wickets, bowled_wickets and special wickets. The data is filtered on season, which keeps 10 of 10 members. The view is filtered on special wickets and bowler. The special wickets filter includes values greater than or equal to 11. The bowler filter keeps 10 of 356 members.

The above chart illustrates bowlers and the way they have taken wickets. As we can see, SL Malinga has taken most wickets which comprises 10 LBW wickets, 68 bowled wickets, and 78 special wickets.

CHART 13:

Toss Match Result

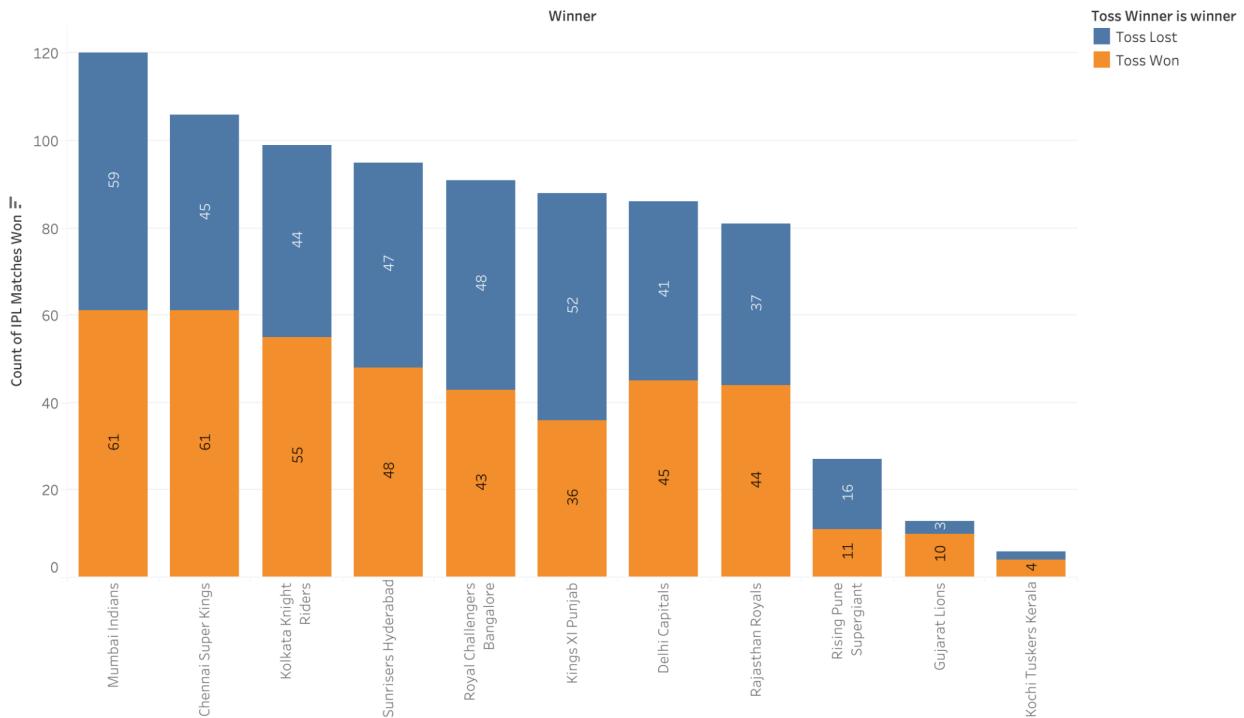


Toss Decision (color). Percents are based on each column of the table.

From the above chart, we can understand that the percentage of choosing the field is more than bat after the toss.

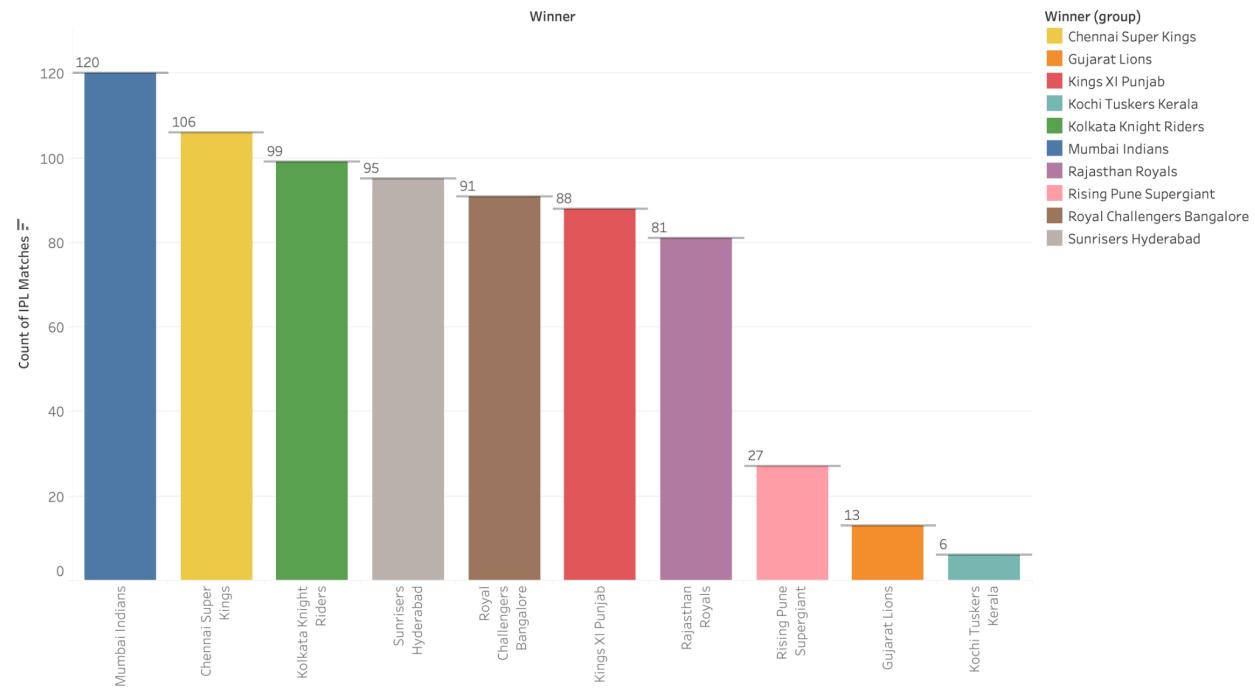
CHART 14 & 15 :

Impact of Toss Winning on Match Winning



Count of IPL Matches 2008-2020_Cleaned.csv for each Winner. Color shows details about Toss Winner is winner. The marks are labeled by count of Winner. The data is filtered on Winner (group) and Date Year. The Winner (group) filter excludes Null and Null. The Date Year filter has multiple members selected. The view is filtered on Toss Winner is winner, which keeps Toss Lost and Toss Won.

Most Successful IPL Team



Count of IPL Matches 2008-2020_Cleaned.csv for each Winner. Color shows details about Winner (group). The view is filtered on Winner and Winner (group). The Winner filter excludes Delhi Capitals, NA, Pune Warriors and Rising Pune Supergiants. The Winner (group) filter excludes Null.

From the above two charts Impact of Toss Winning on Match Winning chart & Most Successful IPL Team chart we can understand that Mumbai has won the maximum number of tosses (61), Chennai Super Kings have also won the same number of tosses & so on for other teams. Now, Mumbai Indians is leading with the number of seasons won followed by Chennai Super Kings and so on. It basically implies that the team who has won the most number of tosses is most likely to win the match.