Project Report

IPL Data Analysis

—

Pooja Malage | 015294769

DATA 230 | Data Visualization | May 2021

Table of Contents

Abstract

Indian Premier League (IPL) is a sports league for cricket which is hosted by a sports body in India called BCCI.This league is similar to NBA and NFL where sports personalities from different countries participate in different club teams. Founded in the year 2007, a platform to promote new talent on such a big platform. The IPL matches are very entertaining, the balance of each team is amazing as it's a mixture of players from different nations which keeps the audience entertained and engrossed in matches.

In this report, the idea is to analyze the IPL data(2008-2020) hosted by Kaggle to come up with a visualization dashboard that would help BCCI Board of directors to understand how successful the league has been ,the top and bottom performing teams in IPL over 11 years and many other facts which can help them make decisions for upcoming IPL for 2021.

This dashboard will also show detailed statistics of each team's performance over each season, what factors contribute to the success of a team, overall team performance in every season, number of matches held in each city, number of matches won by each team, number of matches held every season, top batsmen and bowlers ,etc.

 This dashboard will summarize these events with ease to understand and visualize the data, which will help the board of directors and planner make decisions and start preparation for this year's matches.

Introduction

Cricket is a game invented in London and India was introduced to cricket during the British ruling era. Cricket is a similar version of baseball where one team bats and another team fields and balls.There are various different versions of cricket like One day cricket, test match and 20-20 version. Here the test match is played over a period of 5 days and one-day cricket consists of 50 overs per team and 20-20 consists of 20 hours per team. The 20-20 version of cricket became the most popular version of cricket as this version of cricket is more interesting and brings adrenaline to score more runs in less number of balls. This is when IPL Indian Premier League tournaments were introduced. IPL consists of various different teams where each team consists of players from various countries. Auction of players takes place at the beginning of each championship. This championship is played in various cities over India.

Data visualization brings in a new dimension of analysis which brings user friendly methods to provide analysis for various objectives. Starting from players,spectators, coaches, team owners and governing bodies everyone benefits from data analysis and various important decisions are based on it. Decisions like team selection, game plan ,etc are all based on various data analysis done on available data for previous championships and tournaments. Sports analysis is applied in various mainstream sports like football, baseball, cricket,etc.There is a lot of data for each sport, for example in cricket, we have data for each ball played in every match, what is the strike rate,how many runs are made, who took how many wickets ,etc.

Description of Dataset

Data has been retrieved from kaggle and consists of details of 816 matches from 2008 to 2020.The Dataset contains two CSV files i.e Matches 2008-2020.csv and Ball-by-Ball 2008-2020.csv .Matches file gives the details of match venue,location,season,toss winner ,toss decision,match result,winner,player of match,details of umpires etc as shown figure1. Ball-by-ball.csv is a combination of all the deliveries for all the matches from 2008-2020.It consist do different attributes Match_id,batting team,bowling team,bowler,over,total_runs,inning etc as shown in figure 2.There are around 816 rows and 17 columns in Matches.csv and 193468 rows and 1 columns in Ball_by -ball as shown in figure 3.

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_margin	eliminator	method	umpire1	umpire2
0	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium	0	Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	140.0	N	NaN	Asad Rauf	RE Koertzen
1	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali	0	Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	33.0	N	NaN	MR Benson	SL Shastri
2	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla	0	Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	9.0	N	NaN	Aleem Dar	GA Pratapkumar
3	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium	0	Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	5.0	N	NaN	SJ Davis	DJ Harper
4	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens	0	Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	5.0	N	NaN	BF Bowden	K Hariharan

Figure 1 Sample of Match Dataset

Figure 2 Sample Of Ball-by-ball data

```
print(f'Shape of Matches.csv :{match.shape}')
print(f'Shape of ball-by-ball.csv :{ball.shape}')

Shape of Matches.csv :(816, 17)
Shape of ball-by-ball.csv :(193468, 18)
```

Figure 3 Shape of Match and ball-by-ball data

Tools and Principles Used

The goal of the project is to convert the data into visual representation in the form of charts,graphs,maps etc that can be easily understood and help businesses make decisions based on the insights provided by the dashboard.

Tools Used

- Tableau Desktop 2020.4

- Microsoft Excel

- Jupyter Lab

All the visualizations and dashboards are created using tableau and cleaning of data is performed using Jupyter lab and microsoft excel.

Principles Used

Below principles were used while designing the dashboard

- Gestalt principles
- Tufte's principle

Gestalt principle

Used principles of Gestalt i.e   Law of simplicity by making use of simple charts i.e bar charts ,line charts,pie charts etc that are easy to understand and which can communicate data better and other Law of proximity was used while labelling the charts and making sure they are close to visualization.Made sure to include title,labels for axis,interactive legends and filters for all the visualizations.

Tufte's principle

We used Tufte's Design Principles to ensure high graphical integrity by removing unnecessary visual elements that are not necessary like axes grid lines that could distract the viewer and also applied the concept of Data to ink ratio ensuring the ratio of ink used to display data to the total ink is high. a showing that the labels are clear and detailed. So, all the figures have a title, labels for axis and an interactive legend if necessary.

Keeping in mind these principles, the dashboard was designed.Focused on whitespace distribution in the design process.Charts were simplified by removing axis grid lines.Moved details into the tooltip hover for cleaner design.Colors were utilized properly by keeping aware of background,foreground,luminance and consistency throughout the dashboard.

Data cleaning

Data cleaning is required to the dataset after first preview of the data.After downloading data from kaggle, it was observed that dataset consist of various NULL and duplicate values.There are various columns in dataset with Null values which need to be cleaned up as shown in figure 4.

```
match.isnull().sum()

id                   0
city                13
date                 0
player_of_match      4
venue                0
neutral_venue        0
team1                0
team2                0
toss_winner          0
toss_decision        0
winner               4
result               4
result_margin       17
eliminator           4
method             797
umpire1              0
umpire2              0
dtype: int64
```

```
ball.isnull().sum()

id                    0
inning                0
over                  0
ball                  0
batsman               0
non_striker           0
bowler                0
batsman_runs          0
extra_runs            0
total_runs            0
non_boundary          0
is_wicket             0
dismissal_kind   183973
player_dismissed 183973
fielder          186684
extras_type      183235
batting_team          0
bowling_team        191
dtype: int64
```

Figure 4 Missing value count in datasets

In the dataset, we could see the city, player_of_match, winner, result, result_margin, eliminator, method, etc . All these columns have null values. We will be replacing these Null values as shown in figure 5 for one of the columns City. Data in City column before removing Null values

```
df[df['city'].isna()]
```

	city	venue
399	NaN	Sharjah Cricket Stadium
402	NaN	Dubai International Cricket Stadium
403	NaN	Dubai International Cricket Stadium
404	NaN	Sharjah Cricket Stadium
406	NaN	Sharjah Cricket Stadium
407	NaN	Dubai International Cricket Stadium
408	NaN	Sharjah Cricket Stadium
409	NaN	Dubai International Cricket Stadium
410	NaN	Dubai International Cricket Stadium
413	NaN	Sharjah Cricket Stadium
414	NaN	Sharjah Cricket Stadium
415	NaN	Dubai International Cricket Stadium
417	NaN	Dubai International Cricket Stadium

Figure 5 Showing NaN values

Cleaning these NAN values by taking in reference the values of the venue column which did not have any NaN values as shown in figure 6.

```
df[(df.venue == 'Dubai International Cricket Stadium')].fillna('Dubai')
```

```
df[(df.venue == 'Sharjah Cricket Stadium')].fillna('Sharjah')
```

Figure 6 Code for Replacing NaN values

The data after cleaning NaN values for city it looks like as shown in figure 7

	city	venue
402	Dubai	Dubai International Cricket Stadium
403	Dubai	Dubai International Cricket Stadium
407	Dubai	Dubai International Cricket Stadium
409	Dubai	Dubai International Cricket Stadium
410	Dubai	Dubai International Cricket Stadium
415	Dubai	Dubai International Cricket Stadium
417	Dubai	Dubai International Cricket Stadium

	city	venue
399	Sharjah	Sharjah Cricket Stadium
404	Sharjah	Sharjah Cricket Stadium
406	Sharjah	Sharjah Cricket Stadium
408	Sharjah	Sharjah Cricket Stadium
413	Sharjah	Sharjah Cricket Stadium
414	Sharjah	Sharjah Cricket Stadium

Figure 7 Replaced with Dubai and Sharjah for city column

Similarly, data cleaning for the rest of the columns was performed in the above manner by referencing it to columns which do not have any NaN values.On further analysis of the data, it was found that there are multiple duplicate values in the dataset for multiple columns.For example in column for Team1 on verifying for duplicate values as shown in figure 8.

```
match['team1'].unique()

array(['Royal Challengers Bangalore', 'Kings XI Punjab',
       'Delhi Daredevils', 'Mumbai Indians', 'Kolkata Knight Riders',
       'Rajasthan Royals', 'Deccan Chargers', 'Chennai Super Kings',
       'Kochi Tuskers Kerala', 'Pune Warriors', 'Sunrisers Hyderabad',
       'Gujarat Lions', 'Rising Pune Supergiants',
       'Rising Pune Supergiant', 'Delhi Capitals'], dtype=object)
```

Figure 8 Before Cleaning Unique values for team1

For the column,team1 values like Rising Pune Supergiants, Rising Pune Supergiant were changed to Pune Warriors, Deccan Chargers to Sunrisers Hyderabad, Delhi Capitals to Delhi Daredevils.hence cleaning up discrepancies in the value names as shown in figure 9.

```python
match['team1']=match.team1.str.replace("Rising Pune Supergiants","Pune Warriors")
match['team1']=match.team1.str.replace("Rising Pune Supergiant","Pune Warriors")
match['team1']=match.team1.str.replace('Deccan Chargers','Sunrisers Hyderabad')
match['team1']=match.team1.str.replace('Delhi Capitals','Delhi Daredevils')
```

Figure 9 Code for replacing Duplicate values

After cleaning those discrepancies, data was cleaned as shown in figure 10.

```python
match['team1'].unique()

array(['Royal Challengers Bangalore', 'Kings XI Punjab',
       'Delhi Daredevils', 'Mumbai Indians', 'Kolkata Knight Riders',
       'Rajasthan Royals', 'Sunrisers Hyderabad', 'Chennai Super Kings',
       'Kochi Tuskers Kerala', 'Pune Warriors', 'Gujarat Lions'],
      dtype=object)
```

Figure 10 After Cleaning Unique values for team1

This processing was performed for all the columns which had data discrepancies.

Data Visualization

The cleaned data is imported to tableau for analysing the data. From the dataset description we can find various useful columns present in the dataset, which can help us with various analyses. This analysis will help governing body, team players, coach as well as team owners to make informed decisions.We have divided the analysis in 5 main divisions, namely;

1. Complete Summary of the IPL Matches

2. Venue Analysis

3. Batting Analysis

4. Match Analysis

5. Toss Analysis

We will cover multiple granular analysis in the above divisions of the analysis. Analysis shows the relations between various measures and attributes which are providing various meaningful outcomes.

1. Complete Summary of the IPL Matches

1.1 IPL Champions and Runner-Up team

Visualization anatomy

Columns: [Season Winner]

Rows: CNT([Winner_runner.csv]) → This is basically a generated field which gives count of winners and runner-ups for a given season.

The filters are kept discrete and with year. This is a bar plot with year by year breakdown.Added color filter for dimension season winner.

Analysis and Conclusion

This data gives an idea as to which is the strongest team so far. But as team members keep on changing every season, this analysis gives an idea as to which team retains the best players which helps this team win the majority of matches. If observed carefully it stands out-right that Mumbai Indians team has been champions five times which is the most times any team has won championship.Below figure shows the visualization for the same.
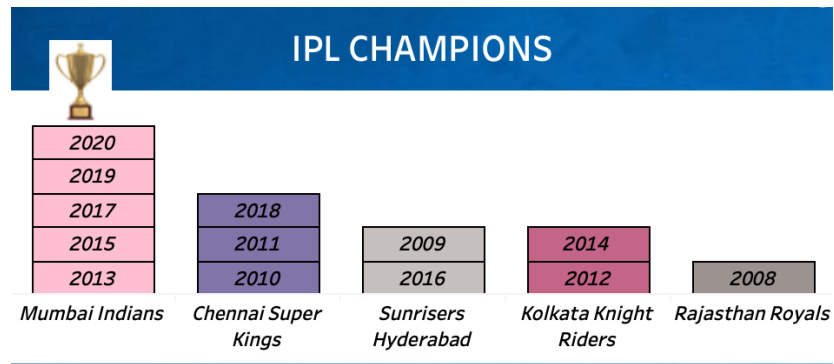
Figure 11 IPL Champion Teams year-wise

The Figure 11 shows a bar chart of top 5 teams with the most number of wins across the entire IPL season (2008-2020) and also breakdown by year wise which team won which year.

1.2 IPL Runner-up

Visualization anatomy

Columns: [Season Runner]

Rows: CNT([Winner_runner.csv]) → This is basically a generated field which gives count of winners and runner-ups for a given season.

The filters are kept discrete and with year. This is a bar plot with year by year breakdown.Added color filter for dimension season runner.

Analysis and Conclusion

On similar terms as IPL champions, using exact data we can find which team was runner-up for majority times. This helps to understand the next to top team which has potential to be

champion.Based on this data BCCI can rate teams and players during auctions of the players of the given team. Observing below chart shows Chennai superking has been runner-ups almost 5 times in total.
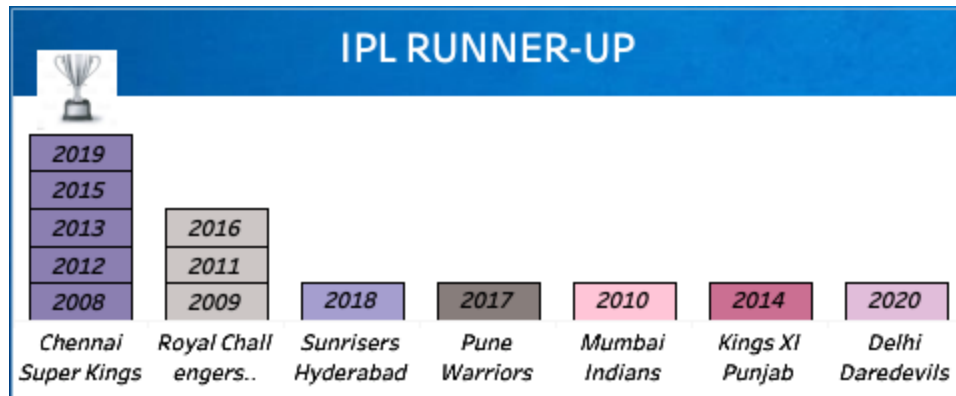
Figure 12 IPL Runner-up Teams year-wise

Teams in the IPL league who were in the finals but didn't win the championship trophy as shown in figure 12. Chennai super kings were 8 times finalists but could win only 3 times. On other hand Mumbai Indians were finalists only for 6 times, but they won the championship 5 times.

1.3 Matches played per season

Visualization anatomy:

Columns: Year

Rows: CNT([ipl matches 2008-2020.csv])CNT([ipl matches 2008-2020.csv]) → This is basically a generated field which gives count of records in above csv file.The above generated field is used twice because the goal is to create a lollipop chart using circle and bar chart.

The filters are kept discrete and with year. This is a bar plot along with a circle plot with year by year breakdown.Added color filter to the count generated field.The count can be seen in circles and count for bars are hidden.

Analysis and Conclusion:

This helps us to understand exactly how many matches are hosted each season and helps board members to make decisions accordingly.On an average there were 60 matches each season held.Figure 13 graph indicates that the highest number of matches happened in the year 2011,2012 and 2013 as more 3 teams were introduced this year.
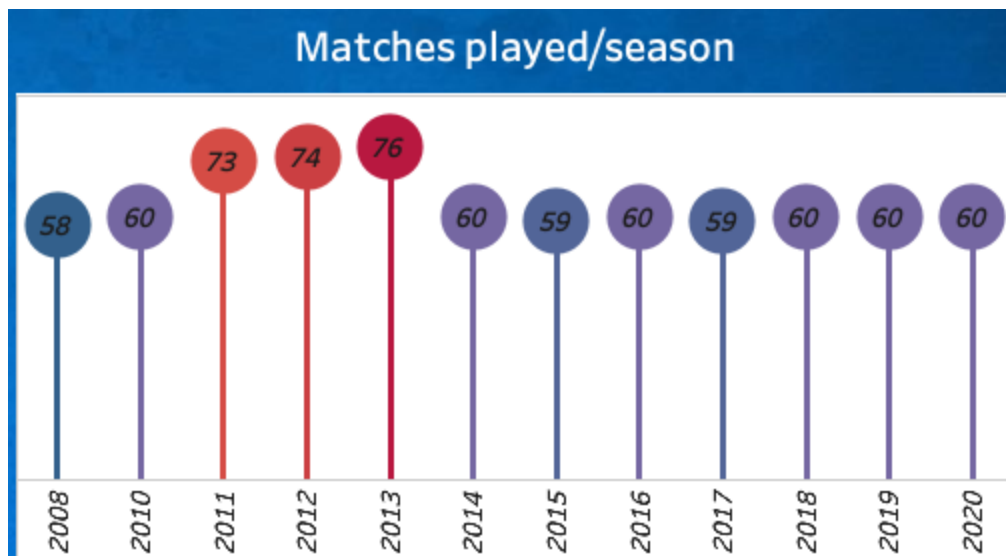
Figure 13 Matches Played each season

1.4 Player Of Match

Visualization anatomy:

Columns: Rows: Empty

Columns and Rows are kept empty as this is a packed bubble plot where the player of the match is given a color mark and CNT(IPL match 2008-2020.csv) is given a circle size mark.The filters are kept discrete and with the player of the match.

Analysis and Conclusion:

In IPL tournaments knowing which player plays the best is the critical information, as this information leads to decide base price of the player during auctions within teams.From this packed bubbles,it can be seen that AB de Villiers and CH Gayle are most promising players who won most number of IPL matches in the entire season.
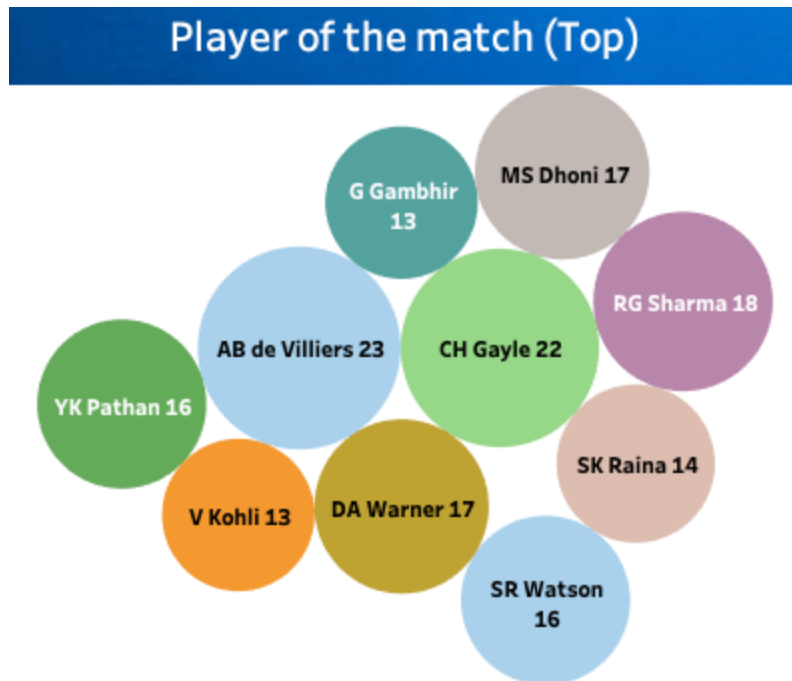
Figure 14 Top Player of match

The figure 14 also shows how many matches in total the players were players of the match i.e. this player played the match better than any players. Team owners and coaches need this data to make decisions and budgets for auctions in upcoming championships.

1.5 Toss Decision and Won by wickets/Runs Percentage

Visualization anatomy

Columns: Rows: Empty

This is a pie chart with Toss decision and CNT(IPL match 2008-2020.csv) as dimensions and color mark given to Toss decision.

Columns: result (Ascending order)

Rows: CNT(IPL match 2008-2020.csv)

This is a bar chart with a filter for result dimension.

Analysis and Conclusion

What are the decisions that teams are taking during course of tournament.Majority of the teams decided to field first rather than batting when they won the toss.After toss decision ,and based on toss decisions 54% matches are won by team which tools fielding first and 45% by teams which took batting first which means there is no clear way to say how teams are winning as shown in figure 15.There can be an chance either team can win by runs(Batting first) or wickets(Fielding first).The probability of drawing a match is 2% only in the entire season.
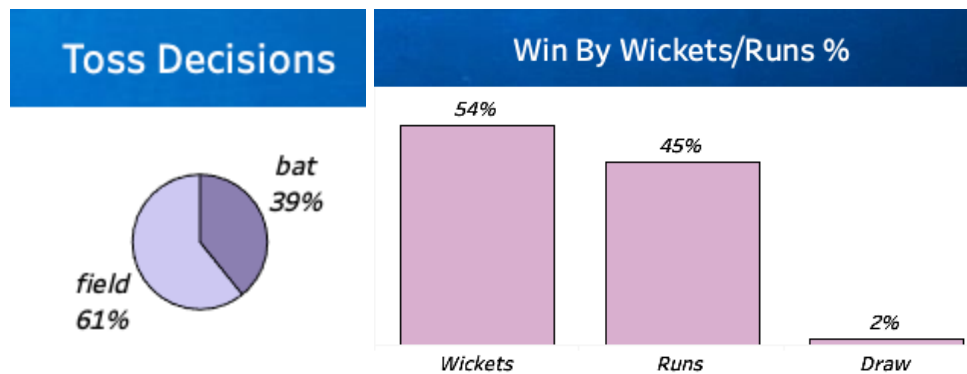
Figure 15 Toss Decision and Won by runs/wickets%

1.6 Complete Dashboard

Analysis and Conclusion:

Below Figure 16 is the complete dashboard showing IPL champions, runner-ups, total matches

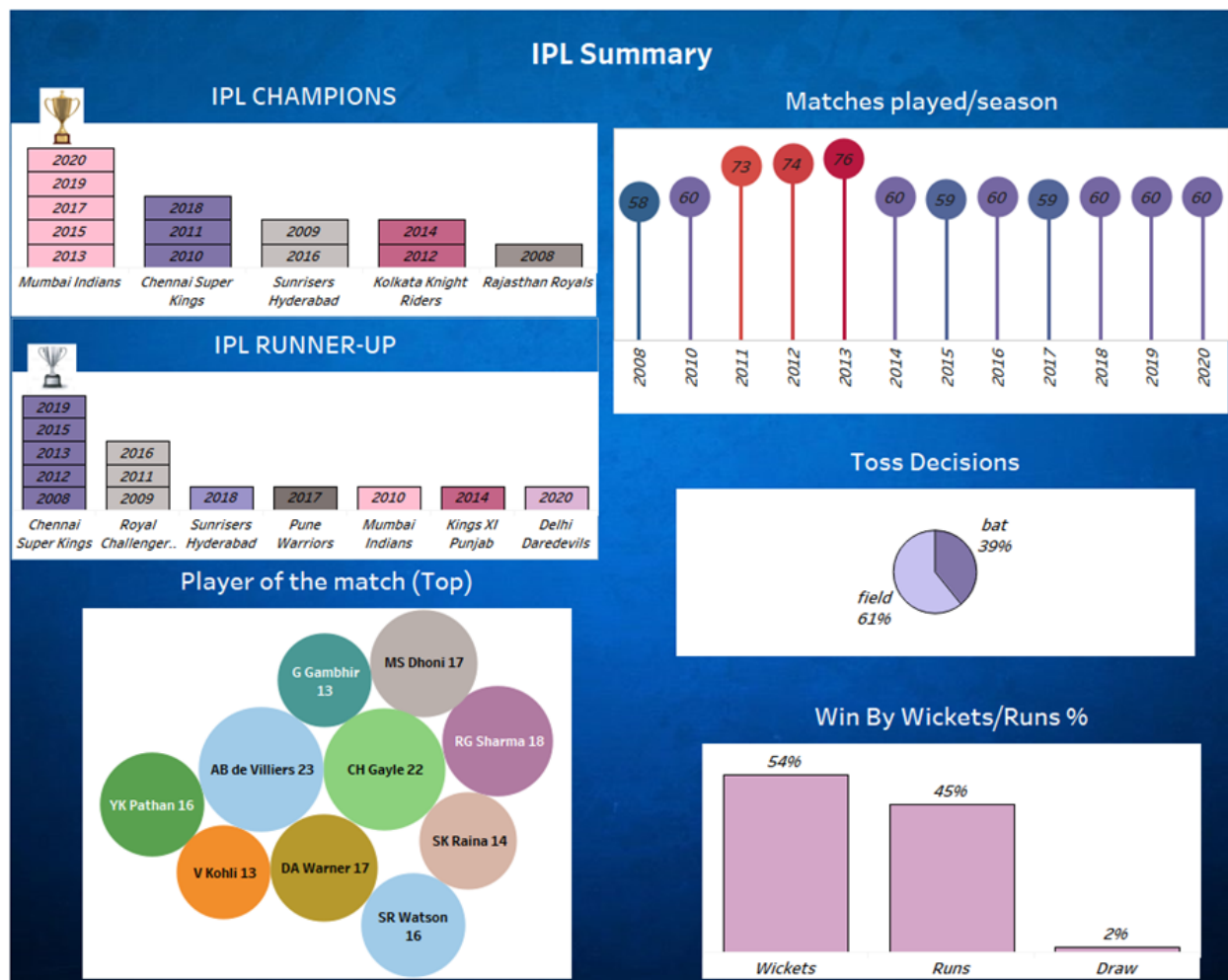played, Player of match, Toss decisions and outcome of match based on toss.

Figure 16 IPL Summary Dashboard

2. Venue Analysis

For team owners and coaches, it is also critical to make decisions on game plan based on where the match will be played. Based on the topographical changes the performance of batting and balling changes drastically. For example some playgrounds have topography favorable for batting while some have it for fielding and bowling.These factors are deciding points on what should be the game plan if a match is being played on specific ground.

2.1 Matches distribution Based on location

Visualization anatomy

Columns: Longitude → (generated field)

Rows: Latitude → (Generated field)

Color mark is given to CNT(IPL match 2008-2020.csv) Detail mark is provided to city and tag marks are provided to both city and generated field CNT. There is a filter for CNT(IPL match 2008-2020.csv)

Analysis and Conclusion:

Analysis of the matches played at various locations give an idea as to which stadium is preferred by BCCI most to host matches.It also depends on the popularity of the stadium.If analysed the figure 17 heat map, we can see majority matches are played in Mumbai, kolkata, Bangalore,

Chennai and Delhi. These are few major metro cities where matches are played alot, both due to

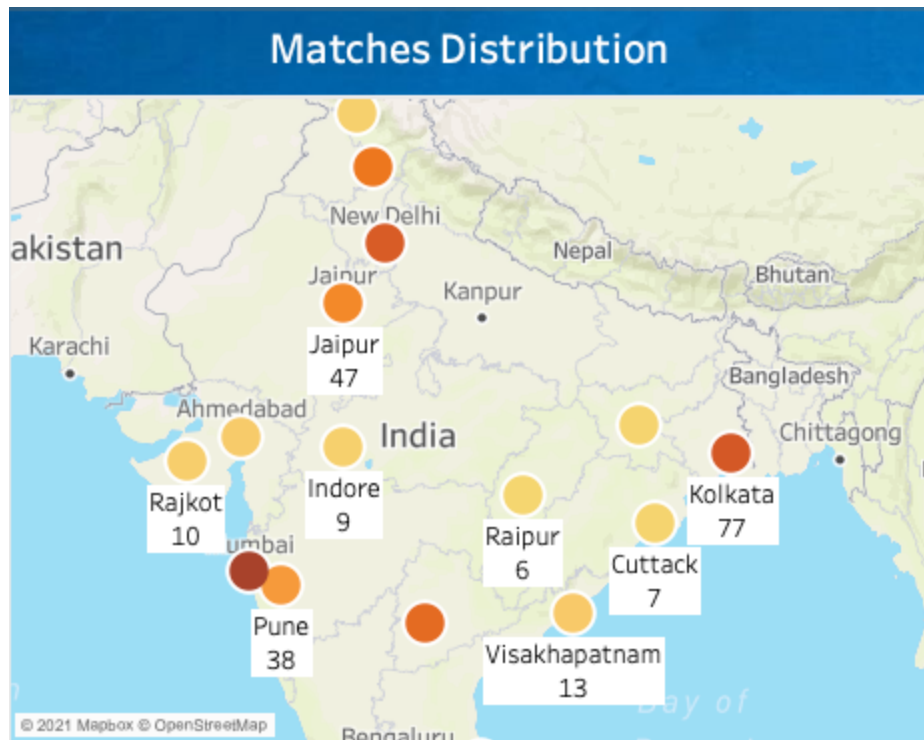popularity and due to favorable weather and topographical conditions.

Figure 17 Match Distribution based on city

2.2 Stadium popularity

Visualization anatomy

Columns: CNT(IPL match 2008-2020.csv)

Rows: Venue

Color mark is given to CNT(IPL match 2008-2020.csv) There is a filter for CNT(IPL match 2008-2020.csv), ordering the rows in descending order.

Analysis and Conclusion:

Stadium popularity is based on multiple factors, what team;s matches are played at this location, how many supporters of specific team are there in these locations, how are weather conditions at these locations, how is the ground,etc figure 18 is based of popularity based on audience and supporters along with topographical favorability. It can be seen that M Chinnaswamy Stadium based in bangalore is the most popular stadium, followed by Eden gardens stadium. These stadiums are favored both by spectators as well as players .

Check this count filter has also been added to view according to no of matches and select based on that.
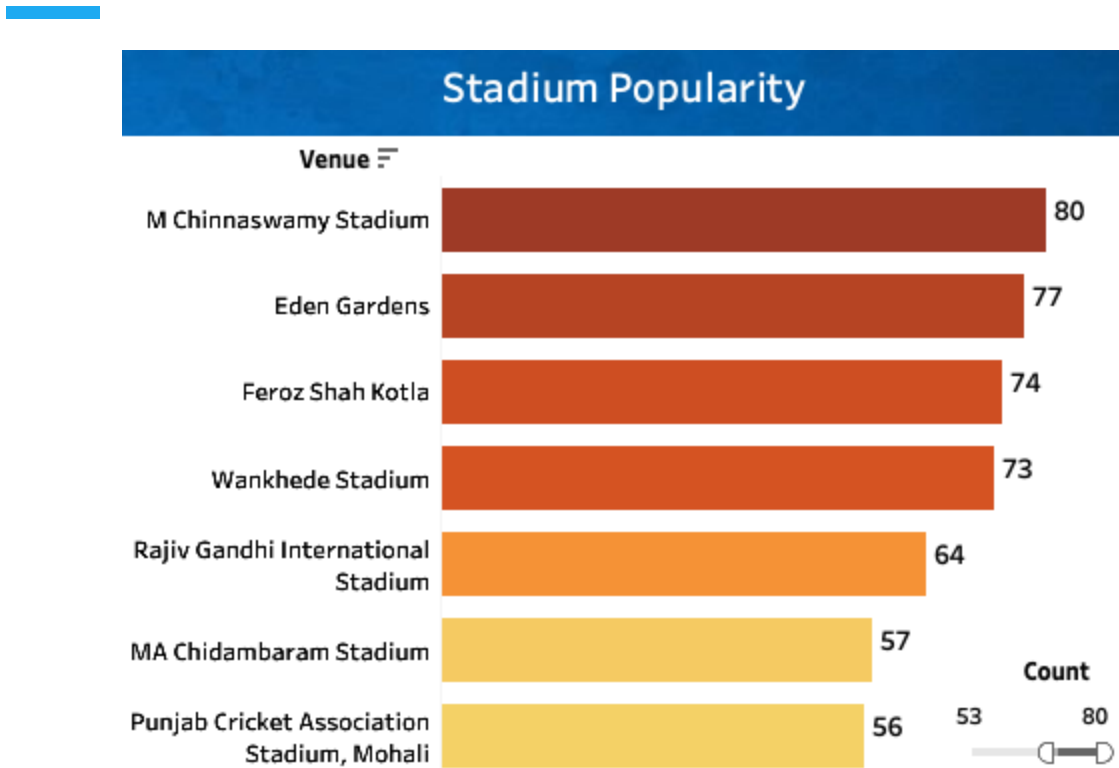
Figure 18 Stadium popularity

2.3 Match results based on type of wins

Visualization anatomy

Columns: CNT(Result)

Rows: Venue

Color mark is given to Result There is a filter for Result and Venue, ordering the rows in descending order.

Analysis and Conclusion

Theoretically, one of the major deciding factors on gameplan is stadium, toss wins and pitch condition of the ground match is being played on.Figure 19 chart data helps conclude that there is a specific trend per stadium which provides favorability of batting or feilding. If observed it can be see that in MA chidambaram stadium majority matches are won by runs, which gives us an Idea that this stadium is favorable for batting, whereas for stadium Savai Mansingh stadium it can be seen that majority matches are won based on wickets which concludes that this stadium is favorable for fielding rather than batting. Hence this analysis also helps in deciding how the game plan of the team should be based on where the match is being played.
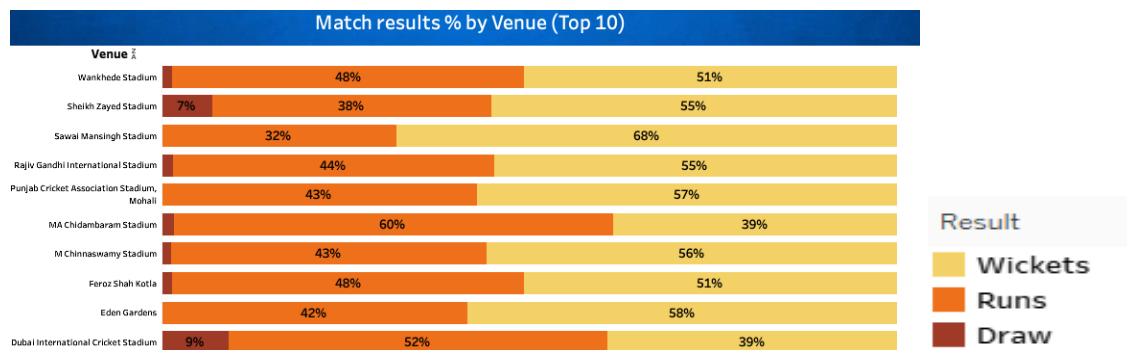
Figure 19 Match result based on type of win

2.4 Dashboard for the Venue Analysis

Analysis and Conclusion

Figure 20 dashboard consolidates all the three charts of Matches distribution, popularity of stadium and match results based on type of wins.
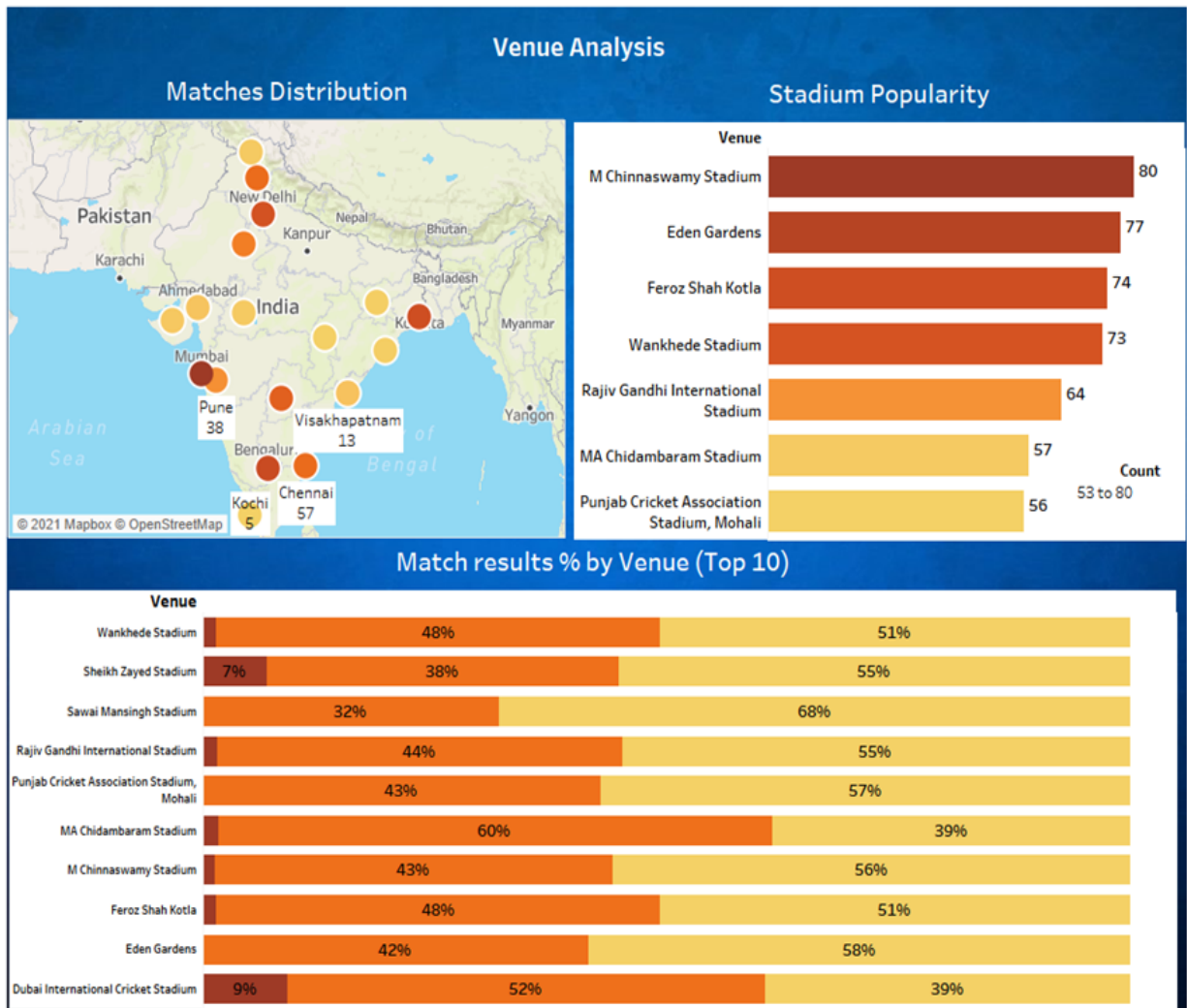
Figure 20 Dashboard for Venue analysis

3. Match Analysis

Along with venue, another major factor which helps BCCI governing body, team coaches and players make few decisions is how were matches played in a given season by specific player or team. If a team is winning how many matches in total are they winning, how aggressively a team is playing a game, what is an average score of the team it is making over seasons. This gives an idea about how the team is performing and what to expect from the opposing team.

3.1 IPL Leaderboard

Visualization anatomy

Columns: CNT(Winner)

Rows: Winner

There is a filter for Winner, ordering the rows in descending order.This is horizontal bar chart.

Analysis and Conclusion

Every tournament requires a leaderboard which provides details as to which team stands where, based on the number of matches they are winning and what is the status of the tournament. Top teams and lower teams can review their status during matches as shown in Figure 21.
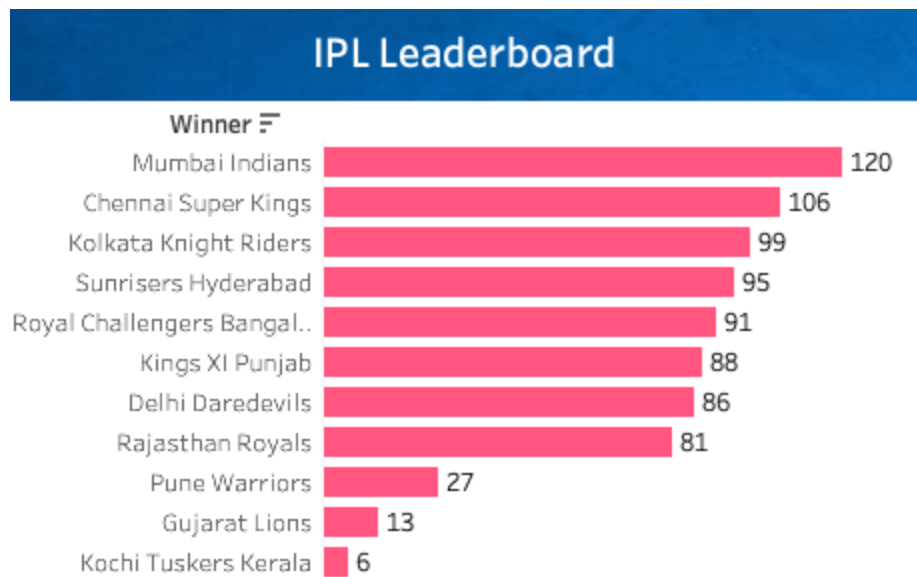
Figure 21 IPL Leaderboard

3.2 Distribution of 4s and 6s

Visualization anatomy

Columns: Year

Rows: CNT(IPL ball-by-ball 2008-2020.csv)

Color mark is given to Batsman runs.There is a filter for Batsman runs.this is discrete line graph.

Analysis and Conclusion

IPL is a shorter version of cricket which is played with only 20 overs each inning. This makes it

motivating for teams to score more in less number of balls. And the fastest way to score

maximum runs is by hitting boundaries and making 4 or 6 runs. The competitive nature of the

tournament can be seen from the total number of 4s and 6s made in a given championship. Figure 22 graph shows the distribution of 4s and 6s over all the season, it can be seen that 4s have been averaging out since the last couple of years starting 2014. But the number of 6s keep on changing drastically and it depends on new players, ofr players who are in proper form for a given season.
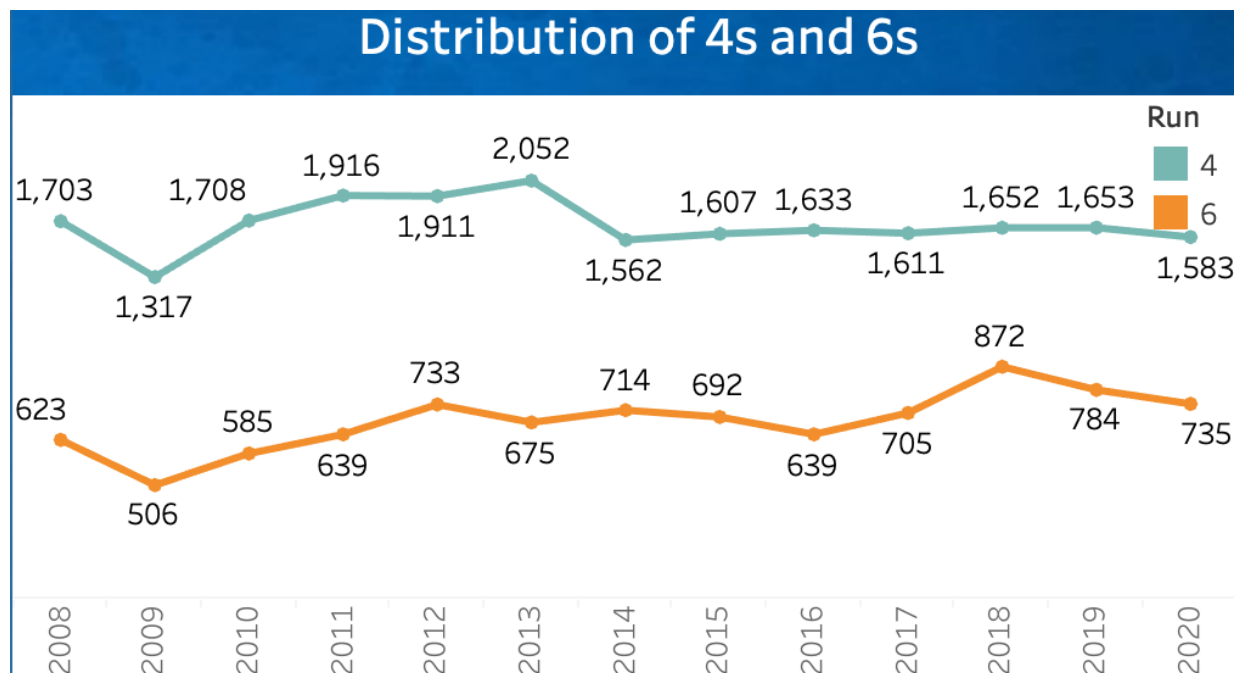

Figure 22 Distribution of 4's and 6's over entire season

3.3 Average team runs per year

Visualization anatomy

Columns: Batting team

Rows: AGG(Avg Runs) ---> Aggregated field for measure Avg runs.

Color mark is given to Years.There is a filter for years which is user selectable.This is a discrete line graph hiding the line and showing just the points.

Analysis and Conclusion

The Figure 23 graph is a very interesting and very important graph as this graph shows average runs made by a given team in every season. This analysis gives details about how a specific team is performing as regards to other teams. It also shows how a single team is performing over multiple seasons, i.e. if that team's performance is increasing or decreasing.
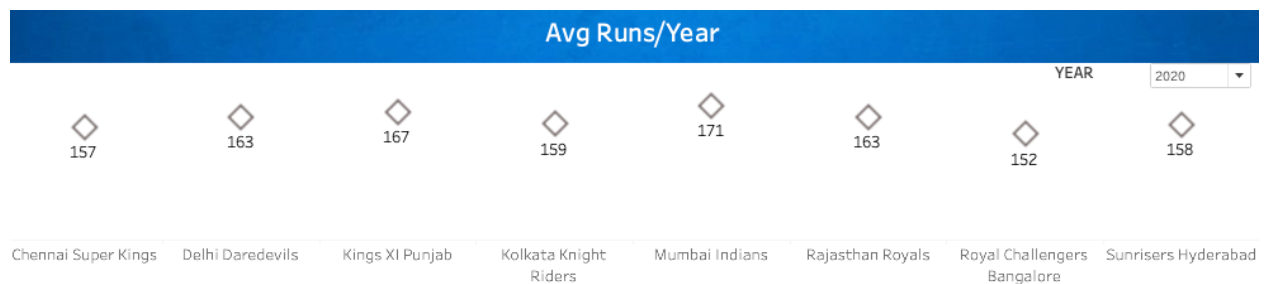
Figure 23 Average Team runs per season

3.4 Top 5 Batsmen and Bowlers

Visualization anatomy

Columns: Rows: Empty

Color mark is given to Batsman runs. Arc size mark given to SUM(Batsman Runs).There is a filter for Batsman runs.This is a pie chart.

Analysis and Conclusion

We have understood till now that to make a better team we need to have good batsmen or bowlers in team. We found a player of match in the previous chart , figure 24 chart gives top 5 batsmen from the overall tournament for a given season. We can see that Virat Kohli is the best Batsman in the tournament , the number represents the total number of runs made by that player. And best bowler being SL Malinga who took over 188 wickets in the championship.
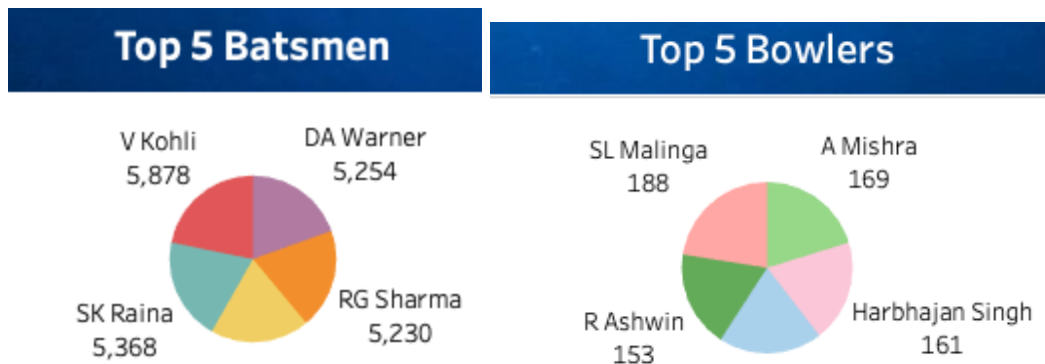
Figure 24 Top 5 batsman and bowler

3.5 Dashboard for Match Analysis

Figure 25 dashboard consolidates all the five charts of IPL Leaderboard, distribution of 4's and 6's ,Average runs per year,Top 5 batsman and top 5 bowlers .
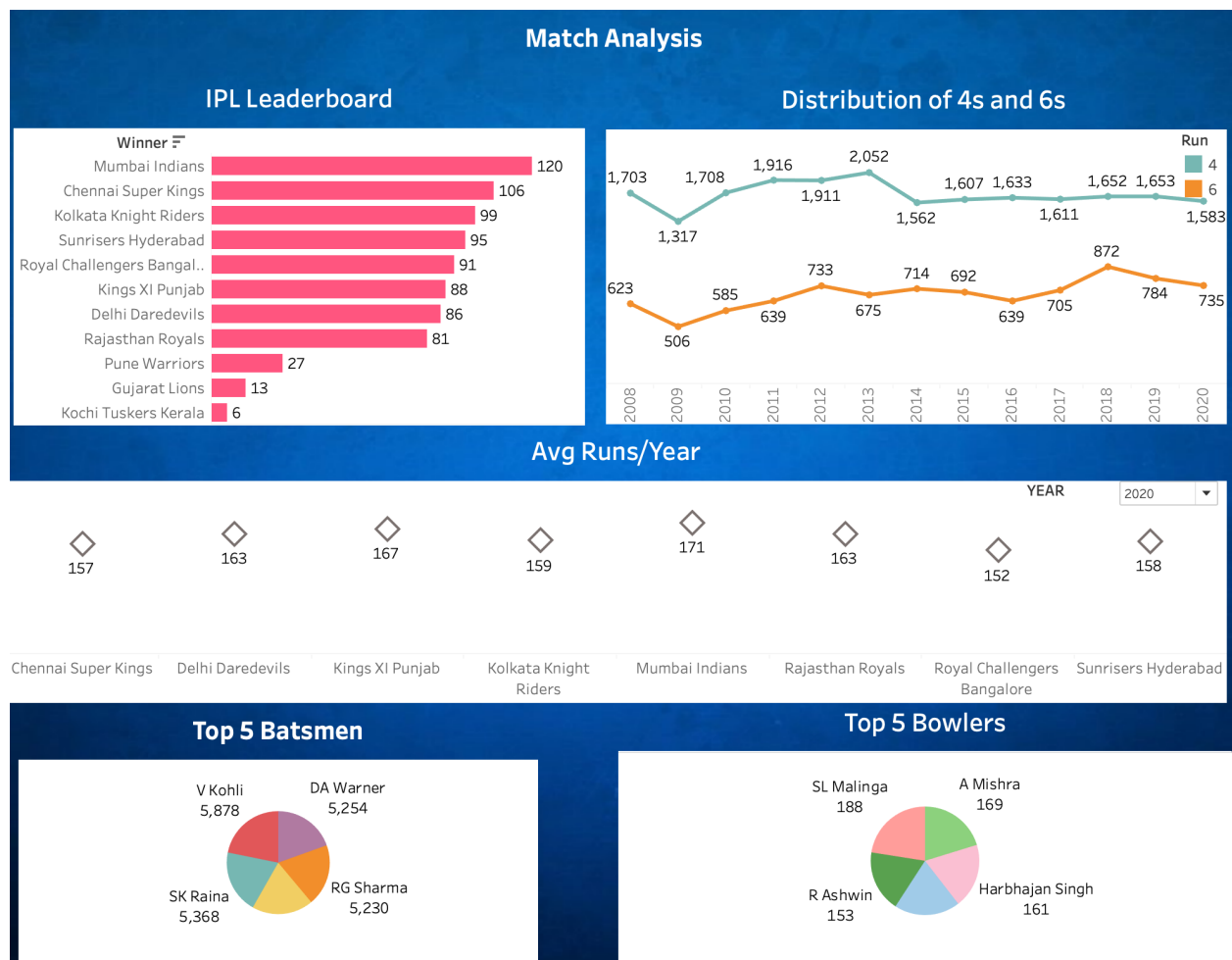

Figure 25 Dashboard for Match Analysis

4. Batting Analysis

Batting analysis is required to find best players who can bat and provide maximum performance output by batting rather than bowling or fielding.Granular analysis of batting for a specific player helps selectors to make decisions during auction of players.

4.1 Total runs per year and Runs Analysis per player

Visualization anatomy

Columns: Year

Rows: SUM(BatsmanRuns)

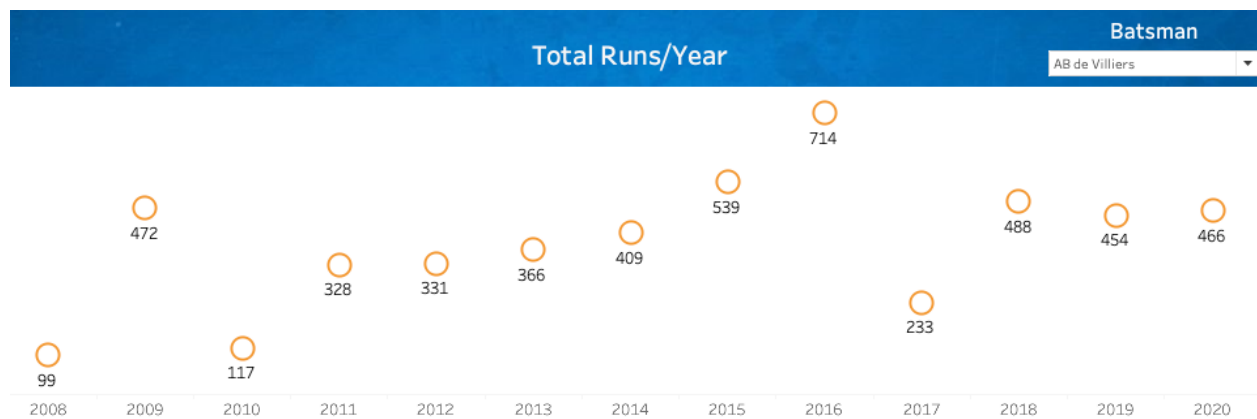Color mark is given to Batsman.There is a filter for Batsman.This is a circle views plot.

Figure 26 Total runs per year per player

Visualization anatomy

Columns: Empty

Rows: Batsman

Color mark is given to Batsman Runs.Arc size mark is given for a calculated field COUNT([IPL Ball-by-Ball 2008-2020.csv]) / TOTAL(COUNT([IPL Ball-by-Ball 2008-2020.csv]))This is a pie chart plot.
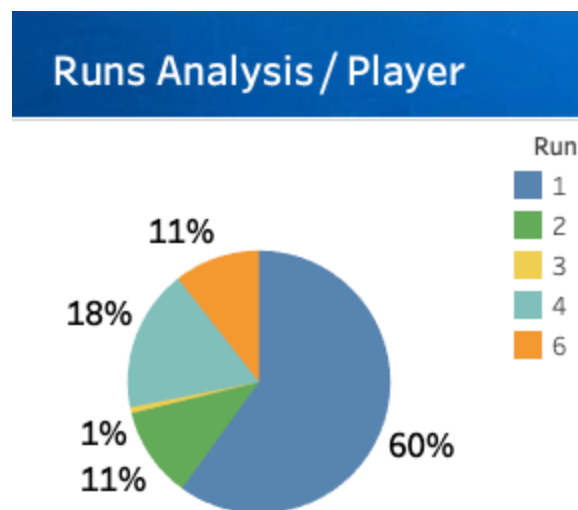

Figure 27 Run Analysis for each player

Visualization Anatomy

Columns: Year

Rows: AGG(Strike rate)

Color mark is given to an attribute for forecast of strike rate.There is a filter for Batsman.This is a continuous line plot.
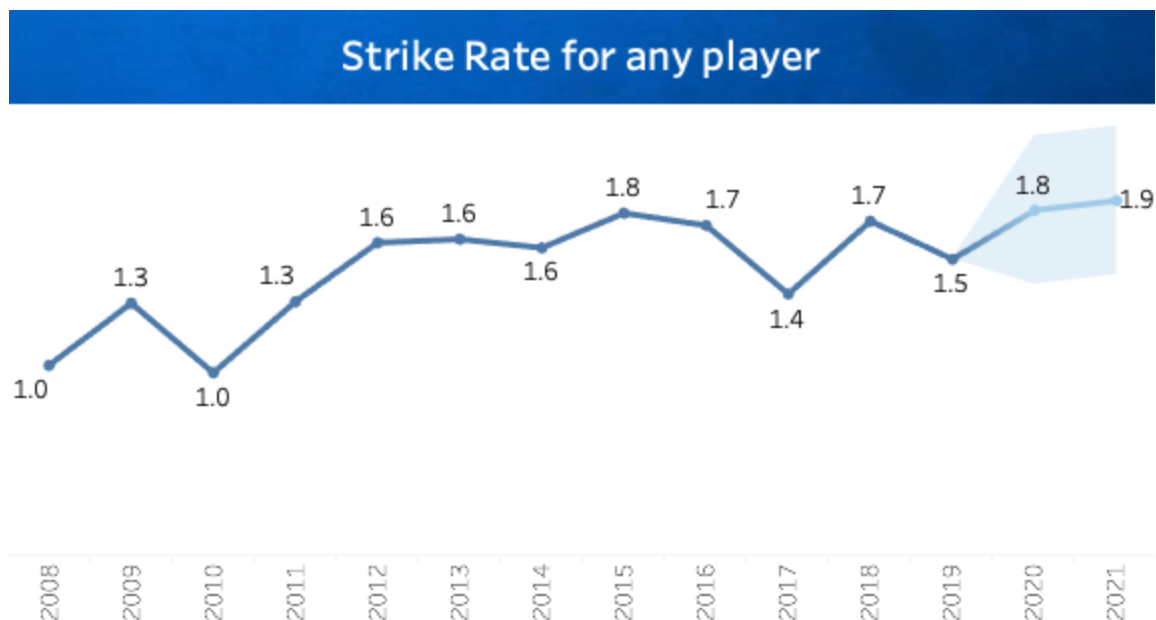

Figure 28 Batsman strike rate over years

Analysis and conclusion

Above chart in figure 26 provides details about a single player over a period of years when IPL was played and player has participated.It provides detailed input of total number of runs that a user has made combining all the matches in a given year championship.This provides analysis like if the performance of player is improving over time or not. For example, in the below chart we have selected random batsman player Ab De Villiers, his performance since 2011 till 2016 has been steadily increasing and it increased exponentially from 2014 till 2016, but from 2017 his performance has deprecated and has been average so far. Also the pie chart in figure 27,

depicts the detailed analysis of runs made by this player. For example, for 2020 how many runs out of 466 were made by hitting 6s and 4s which is 11% and 18% respectively. And the rest of the run were scored by making 1s and 2s. Also from figure 28 line chart we can see the strike rate of the player. Strike rate is nothing but runs made per ball.If we see the trend for Ab De Villiers he has a pretty consistent strike rate of 1.6 runs per ball.

This line chart does not just show the strike rates for all previous years but also does the prediction for a player's upcoming performance. For example it has been predicted that Ab De villiers' strike rate in 2021 will be 1.9.

These above analysis for each player can be done based on runs per balls and this helps in deciding the performance of a batsman.

4.2 Batting Analysis Dashboard

Figure 29 dashboard consolidates all the three charts of Total runs per year,Runs Analysis and strike rate for each individual player.
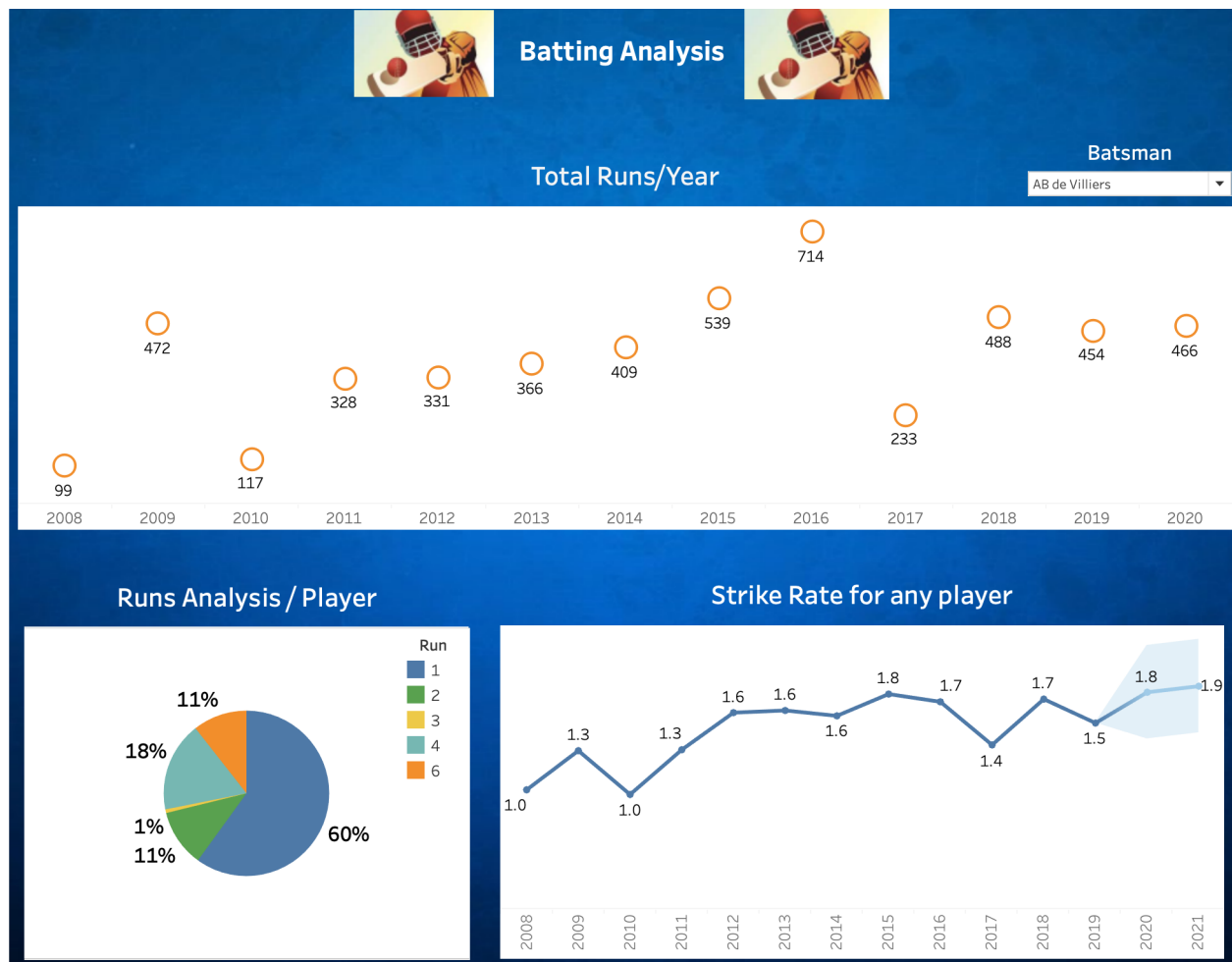
Figure 29 Dashboard for batting Analysis

5. Toss Analysis

Before the game starts, a toss is done between both opposing teams to decide who will bat or bowl first. Team captain's decision can be a game changing decision as what is decided during toss depends on multiple factors like what is opposing team, what is the venue, what are the favorable conditions of the ground,etc

5.1 Toss winner

Visualization Anatomy

Columns: Toss Winner

Rows: Toss Decision, AGG(win%)

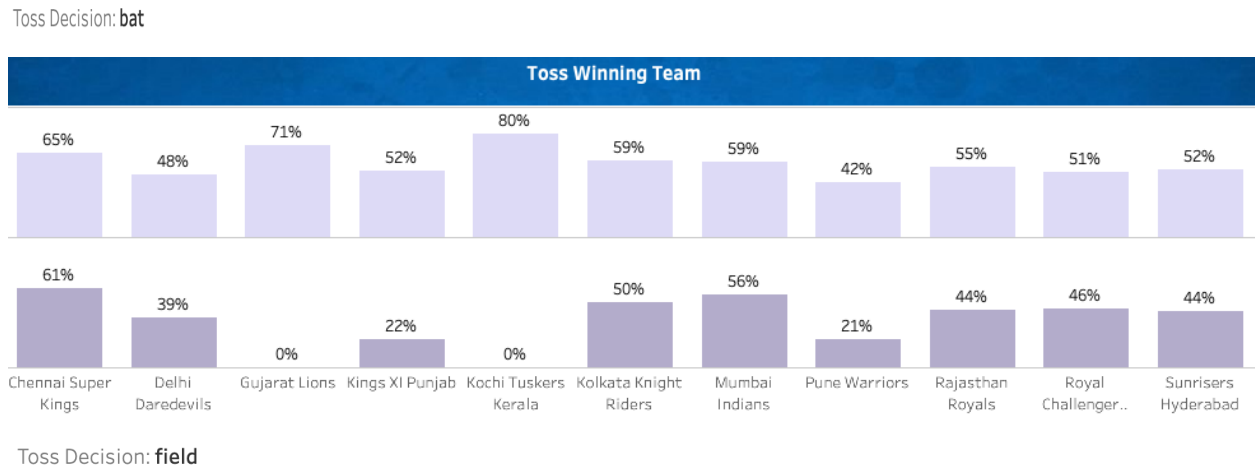Color mark is given to Toss' decision. This is a multiple bar chart plot.

Figure 30 Percentage of team winning based on toss decision wins

5.2 Toss Loser

Visualization Anatomy:

Columns: Toss Losser

Rows: Toss Decision, AGG(win%-Toss Loss)

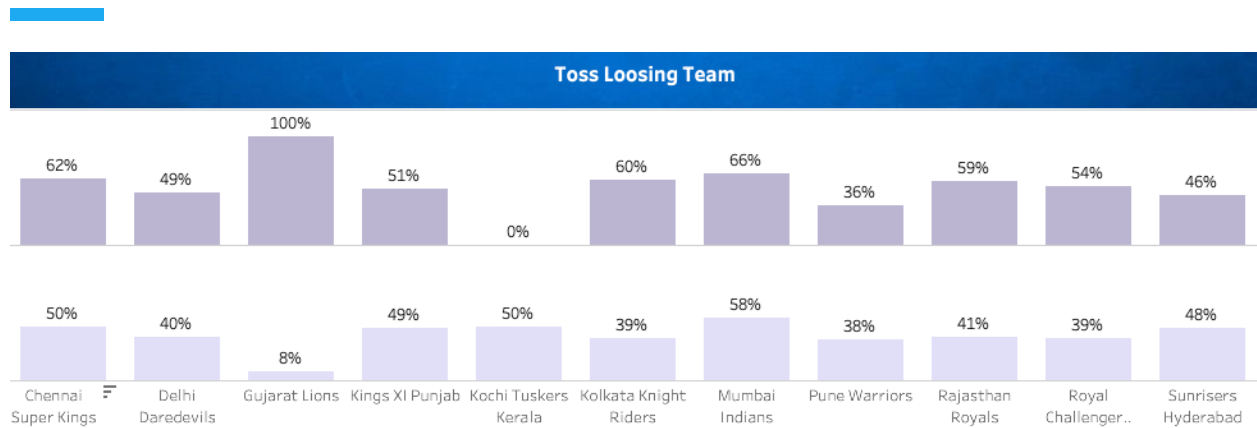Color mark is given to Toss' decision. This is a multiple bar chart plot.

Figure 31 Percentage of team winning based on toss decision loose

5.3 Distribution of Toss Decision per team

Visualization Anatomy

Columns: Empty

Rows: Toss Winner

Color mark is given to Toss decision.Arc size mark is given to CNT(IPL Matches

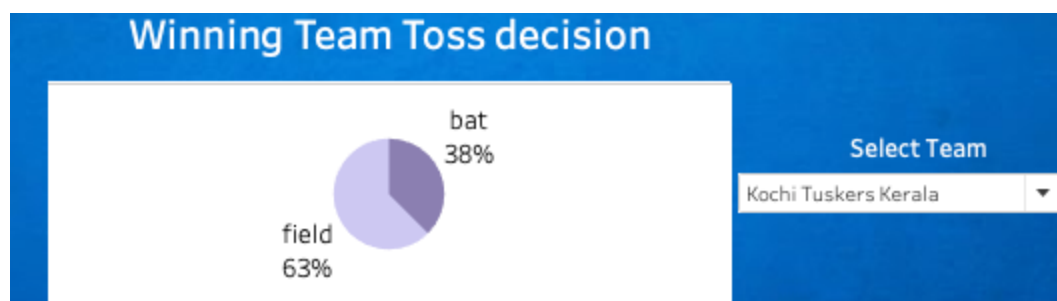2008-2020.csv).There is a filter for Toss winner.This is a pie chart plot.

Figure 32 Distribution of Toss Decision per team

The figures 30,31 and 32 provide insight for the toss analysis. From figure 30, it can be seen clearly how impactful is the toss decision for winning and losing the game.This figures shows if a team wins toss what are the possibilities that team wins the match based on what it chooses, for example if seen for team Gujarat lions, if they win toss and select fielding their probability of winning match is better. Similarly it can be seen for other teams also, they have better chances of winning a match if they choose fielding rather than batting if they win the toss. In figure 31 we can see chances of the team winning the match are almost the same even if they do not have any option of choosing what to do first as they lose the toss. So it completely depends on their performance.Figure 32 shows toss winning decisions taken by individual teams, and as seen, majority of the teams have chosen to field rather than batting. This concludes that it is easier to chase down a target rather than setting a target and win the match.

Figure 33 Dashboard for Toss Analysis

6.The main dashboard

In this dashboard,it gives overall summary of IPL match.It gives information on when IPL

started in 2008 and till now 13 seasons were completed.In total,11 teams have participated till

now.816 matches were played.420 players participated,31 cities hosted matches,Mumbai hosted most no of IPL matches i.e101 matches till now, 34 stadiums matches played ,87 M chinnaswamy stadium ,bangalore most no of matches played.

Figure 34 Statistics of complete IPL

Conclusion

In this project, the dataset provided by kaggle for IPL contained extensive data and very granular data which helped to make multiple analyses. Data visualizations have opened new doors for data analytics where data can be represented to a common man in a very simplistic manner that even he can make analysis from that data. This project focuses on creating visuals using Tableau software. Various different types of chart plots are used like vertical bar chart, horizontal bar chart, discrete line , continuous line, circle plot,scatter plot,pie chart, etc.To create these visualizations, data had to be manipulated where certain field are supposed to be calculated and generated fields are required to be used. Averaging data or summation of data is required at various plots to provide meaningful analysis. Once the worksheets are created, dashboards are created to unite various visuals related to the same topic. This results in creating multiple dashboards for multiple topics, which are then combined together using storytelling features, where all dashboards are combined and can be accessed using tabs on the same page.

References

IPL Complete Dataset (2008–2020). (2020). Kaggle.

https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020

IPL Teams and trophy image

https://indianexpress.com/article/sports/ipl/ipl-2020-full-schedule-fixtures-start-date-timings-venues-6583339/

IPL Trophy

https://www.iplt20.com/

Github link : https://github.com/poojamalage/DATA230