

---

# INDIAN PREMIER LEAGUE (IPL) DATA VISUALIZATIONS

---

**Muhammad Murtaza Mir**  
MS - DS Graduate Student  
Luddy School  
[mmir@iu.edu](mailto:mmir@iu.edu)

**Raunav Sharma**  
BS - DS Undergraduate  
Luddy School  
[rashar@iu.edu](mailto:rashar@iu.edu)

**Umesh Kumar Gattam**  
MS - DS Graduate Student  
Luddy School  
[ugattem@iu.edu](mailto:ugattem@iu.edu)

## Abstract

Cricket is played in almost all the erstwhile colonies of the British Empire, has a devoted fan following and base that is mind boggling.

Cricket may have started as a leisurely pastime, played in the quaint English countryside. But it was soon taken by the English colonists as the marauding British Empire took giant strides across most parts of the globe. For little over a century the game, and its proponents who were once mocked as “flannelled fools”, have strode confidently and tirelessly across grounds and maidans in almost all the erstwhile colonies of the British Empire. One of the most popular sports on the planet, cricket has a devoted fan following and a mind boggling support base. A market research survey conducted by the International Cricket Council (ICC) in 2018 indicated that cricket has over a billion fans globally, with the Indian sub-continent alone accounting for more than 90 percent of them.

Cricket was introduced to North America by the English colonists as early as the 17th century, and by the 18th century it had arrived in other parts of the world. It was taken to the sunny West Indies by the colonists, who still fondly refer to it as “crickoot maan” and to India by the expanding British East India Company. Cricket arrived in Australia almost as soon as colonization began in 1788, the sport reaching New Zealand and South Africa in the early years of the 19th century.

Having dramatically evolved from those laidback times, cricket today is played in three formats – test cricket, one day Internationals (ODIs for short), and what is rapidly becoming the most popular version of the game, Twenty20 (T20). Tests are played over five consecutive days and consist of each team being allowed to play two innings. ODIs, as the name suggests, are played over a day, each side being allowed a maximum of 50 overs each, each over, as it is in all forms of the game, consisting of six balls. T20 is cricket's power-packed, latest innovation. First spun off in the early 2000s, T20 consists of 20 overs per side, the team making more runs or taking more wickets being adjudged the winner. The team batting second must chase down the score given by the team batting first. If the team batting second overtakes the score made by the team batting first, then the team batting second is declared the winner. On the other hand if the team batting first manages to defend their score or wraps up all the wickets of the team batting second, it will be judged the winner.

## **Introduction**

The Indian Premier League (IPL) was an innovation that had to happen. The vacuum that existed between domestic cricket and cricket played between the ICC member countries, was waiting to be taken and occupied. T20, already the glitzy, fastest, and shortest version of the game, was given an added fillip in the franchise-based IPL. The immensely popular IPL with its 'obscene' payouts to contracted players, reflects in many ways the growing consumerism, demand for instant entertainment and gratification, and the relatively fast-paced life that has permeated most Indian cities and larger towns.

A professional Twenty20 cricket league, the IPL, which was founded by the Board of Control for Cricket in India (BCCI) in 2007, is contested between eight teams. Each of the franchise teams is based out of eight different Indian cities, the name of the city or the region, emblazoned in the name of the team. The IPL league's format lends itself to intensive competition as teams find themselves going up and down the points ladder, a literal, snakes, and ladders contest. Conducted annually between March and May, the IPL has been designed to unearth young talent across all countries, and to increase the sports' fan base. It certainly has succeeded, with many more children and women getting engrossed in the game thanks to the instant appeal of IPL. The first day-night league match of the IPL got off to a rain interrupted but rousing start, in 2008.

By far the most-attended cricket league in the world, the IPL even as far back as 2014, was ranked as high as sixth by way of average attendance among all sports leagues. By 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube, and according to

Duff & Phelps, the leagues' brand value was a whopping US\$6.7 billion in 2019. The IPL's 2020 edition smashed all viewership records with a whopping 31.57 million average impressions

One of the reasons why we have opted to analyze the IPL data is because of the popularity and bonhomie that this league has spun off in India, radicalizing the way the sport is played, and in the process bringing the sport into the drawing rooms of millions. The IPL has also been a money spinner for the Indian cricket board – the BCCI – further enhancing its stature as the richest and most powerful cricket board in the world. The IPL is also one of the biggest sports leagues on the planet.

### **Tournament Format**

The IPL is currently played with eight teams, each team playing the other twice in a home-and-away round-robin, league format. At the conclusion of the league stage, the top four teams qualify for the playoffs. The top two teams from the league phase will play against each other in the first qualifying match, with the winner going straight through to the IPL final, the loser getting another chance to qualify for the IPL final by playing the second qualifying match. Meanwhile, the third and fourth placed teams from the league phase play against each other in an eliminator match and the winner from that match playing the loser from the first qualifying match. The winner of the second qualifying match will move into the final to play the winner of the first qualifying match. The winner of the IPL final, being crowned the Indian Premier League champions.

### **About Teams**

IPL's most successful team has been the star studded and powerful Mumbai Indians. Ably led by Rohit Sharma they have pocketed five titles. Chennai Super Kings led by the unflappable Mahendra Singh Dhoni are the second most successful team in the IPL, having won the finals thrice.

Team	[hide]	City	Home ground	Debut	Owner <small>[1]</small>	Captain	Coach
Chennai Super Kings		Chennai, Tamil Nadu	M. A. Chidambaram Stadium	2008	N. Srinivasan	MS Dhoni	Stephen Fleming
Delhi Capitals		New Delhi, NCT of Delhi	Arun Jaitley Stadium		Grandhi Mallikarjuna Rao Sajjan Jindal	Rishabh Pant	Ricky Ponting
Kolkata Knight Riders		Kolkata, West Bengal	Eden Gardens		Shah Rukh Khan Jay Mehta	+ TBD	Brendon McCullum
Mumbai Indians		Mumbai, Maharashtra	Wankhede Stadium		Mukesh Ambani	Rohit Sharma	Mahela Jayawardene
Punjab Kings		Mohali, Punjab	Inderjit Singh Bindra Stadium		Mohit Burman Ness Wadia Preity Zinta Karan Paul	TBD	Anil Kumble
Rajasthan Royals		Jaipur, Rajasthan	Sawai Mansingh Stadium		Manoj Badale Lachlan Murdoch Gerry Cardinale	Sanju Samson	Kumar Sangakkara
Royal Challengers Bangalore		Bengaluru, Karnataka	M. Chinnaswamy Stadium		United Spirits	TBD	Sanjay Bangar
Sunrisers Hyderabad		Hyderabad, Telangana	Rajiv Gandhi International Cricket Stadium	2013	Kalanithi Maran	Kane Williamson	Trevor Bayliss

BCCI has planned to introduce more teams namely Ahmedabad and Lucknow in 2022.

## Awards

Players will be awarded individual prizes after each IPL match and on the completion of the league stage.

Some of the awards are as follows:

### Orange Cap

The Orange Cap is awarded to the top run-scorer in the IPL across the entire season. It is an ongoing competition with the leader wearing the cap throughout the tournament until the completion of the final game, with the eventual winner keeping the cap for the season.

### Purple Cap

The Purple Cap is awarded to the top wicket-taker in the IPL during an entire season. It is an ongoing competition with the leader wearing the cap throughout the tournament until the final game; the eventual winner getting to keep the cap.

### Most valuable Player

The award was called the ‘man of the tournament’ until the 2012 season. The IPL introduced the ‘Most Valuable Player’ rating system in 2013, the leader of which would be named the ‘Most Valuable Player’ at the end of the season.

### **Fair Play Award**

The Fair Play Award is given at the end of each season to the team with the best record of fair play. The winner is decided based on the points the umpires give to the various teams. After each match, the two on field umpires, and the third umpire, rate and score the sense of fairness exhibited by both the teams.

### **Emerging Player Award**

It has been an award that has evolved over the years. In 2008, the IPL governing council christened and presented an award to the ‘Best under-19 Player’. In 2009 and 2010, the leagues’ best under-23 player was awarded the ‘Under-23 Success of the Tournament’ award. In 2011 and 2012, the award was known as the ‘Rising Star of the Year’, while in 2013, it was called the ‘Best Young Player of the Season’. Since 2014, the award has been called the ‘Emerging Player of the Year’.

### **Most Sixes Award**

The player who sends the ball sailing over the boundary ropes the maximum number of times during the entire season, is awarded the ‘Unacademy Let’s Crack It Sixes Award’, the name of the award a reflection of the sponsoring entity.

### **Motivation**

As mentioned above, the IPL is one of the most popular and widely loved sports leagues in the world’s cricketing regions. Due to this, a lot of commercial value is involved for everyone who participates in it, especially for the players and the team owners. From the perspective of the team owners, it can be an important case study to visualize the performance of the teams and the players in order to help the owners make the right selections in the future. Furthermore, the stats which we obtain can be used to determine the various strengths and weaknesses of each individual player, thus leading to decisions about the future competitors and matches being played.

Our main motivation is to understand the utility and applications of the plots in the best way possible. We want to gain expertise on which plot is appropriate for a given scenario. We have seen that many types of plots can be used to portray the same scenarios. Many times, different plots can convey different information regardless of the original intent. Therefore, we want to understand, through this project, how to choose plots to depict the information that does not distort the perception and also conveys the information that can be well perceived and understood by others with as little effort as possible. Our aim would be to convey a story with

our visualization to the viewers. They should be able to easily understand the message of the visualization.

One example in our research that we found about not using the plot appropriately is shown below:

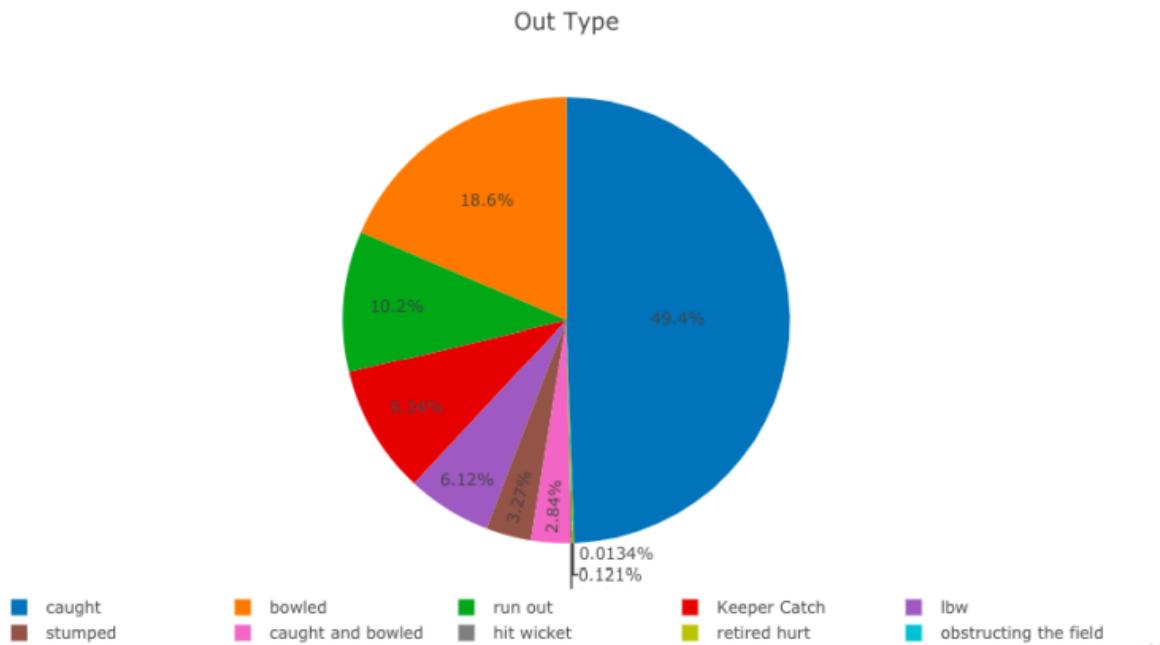


Source: <https://techtrunk.in/ipl-data-analysis/>

The plot wants to convey the wins of the teams while chasing the score (batting second) put up by the other team on the scorecard (batting first) at different venues with the intent to identify the best venue for chase and also the number of wickets by which the team chasing the score wins by. But the people who know about the basics of cricket are well aware of the fact that while chasing, the least number of wickets with which a team can win is one and not zero! But the plot shows that there have been teams winning by zero wickets. Therefore, the plot has not been used appropriately.

Let's take another example.

**Dismissal:** Dismissal occurs when the batsman is out (also known as the fielding side taking a wicket and/or the batting side losing a wicket). At this point, a batsman must discontinue batting and leave the field permanently for the innings. Let's see how the players are getting dismissed across all seasons.



There are few issues with this plot. First thing is that it is a pie chart. We know that Pie charts are one of the most overused graphs and in most cases are not the best way to present data especially when we are trying to categorize more variables. They often distort the information and make it more difficult for decision-makers to understand the messages they contain. As it can be observed that there are multiple categories involved and hence the entire information is not being conveyed properly. Alsom there are two specific categories the values for which are less than one percent. They are not visible enough for comprehension.

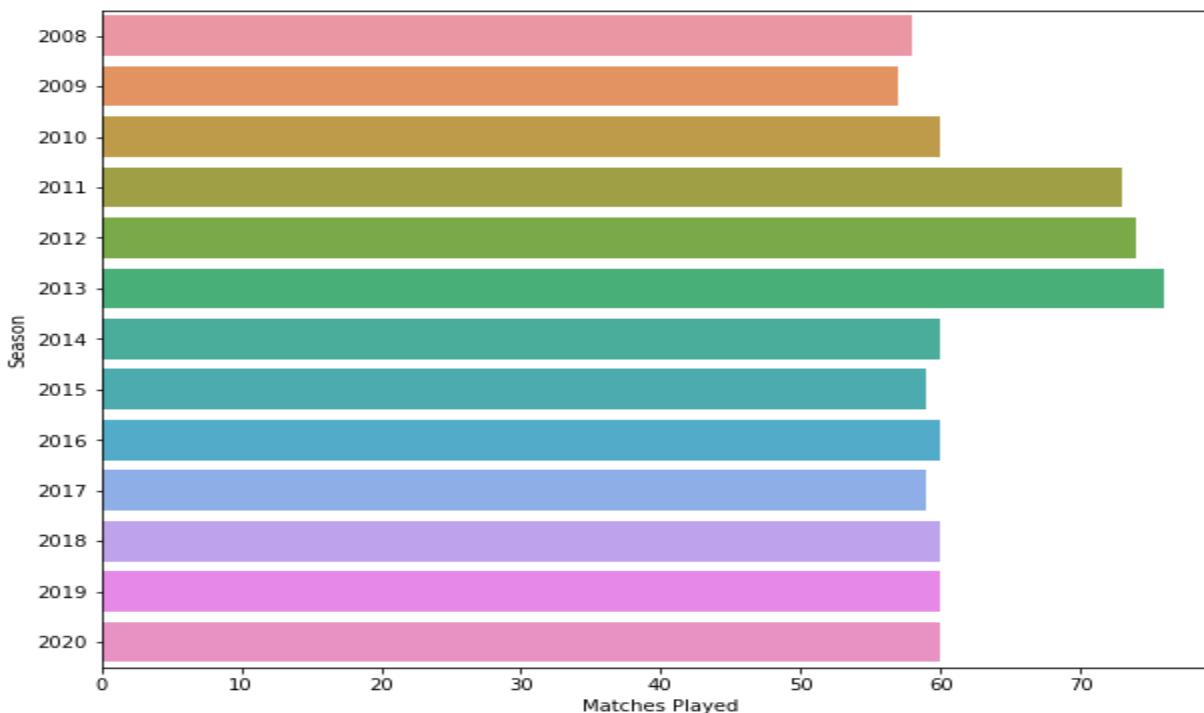
The second problem to notice is that it is using the Red and Green color combination. People with strong CVD - Color Vision Deficiency (strong meaning a more severe condition of CVD) would see both red and green as brown. People with weak CVD can see strong red and green colors as red and green. It is also an issue with color blind people since ten percent of men are colorblind and mostly about the red and green combination.

## Existing Work

Cricket visual analysis is becoming a very big area of interest especially in India specifically because of the sheer amount of popularity and the capital involved. The Indian Cricket Board known as the Board of Control for Cricket in India abbreviated as BCCI is the richest cricket board in the world with an estimated net worth of around 2 Billion USD. Therefore, cricket analysis especially in the field of data visualization has been gaining a steady pace in the recent past. As a consequence of that, there have been some visualizations in the recent past that have tried to capture and visualize the individual performances of the players as well as the overall team performances.

Some of the visualizations include:

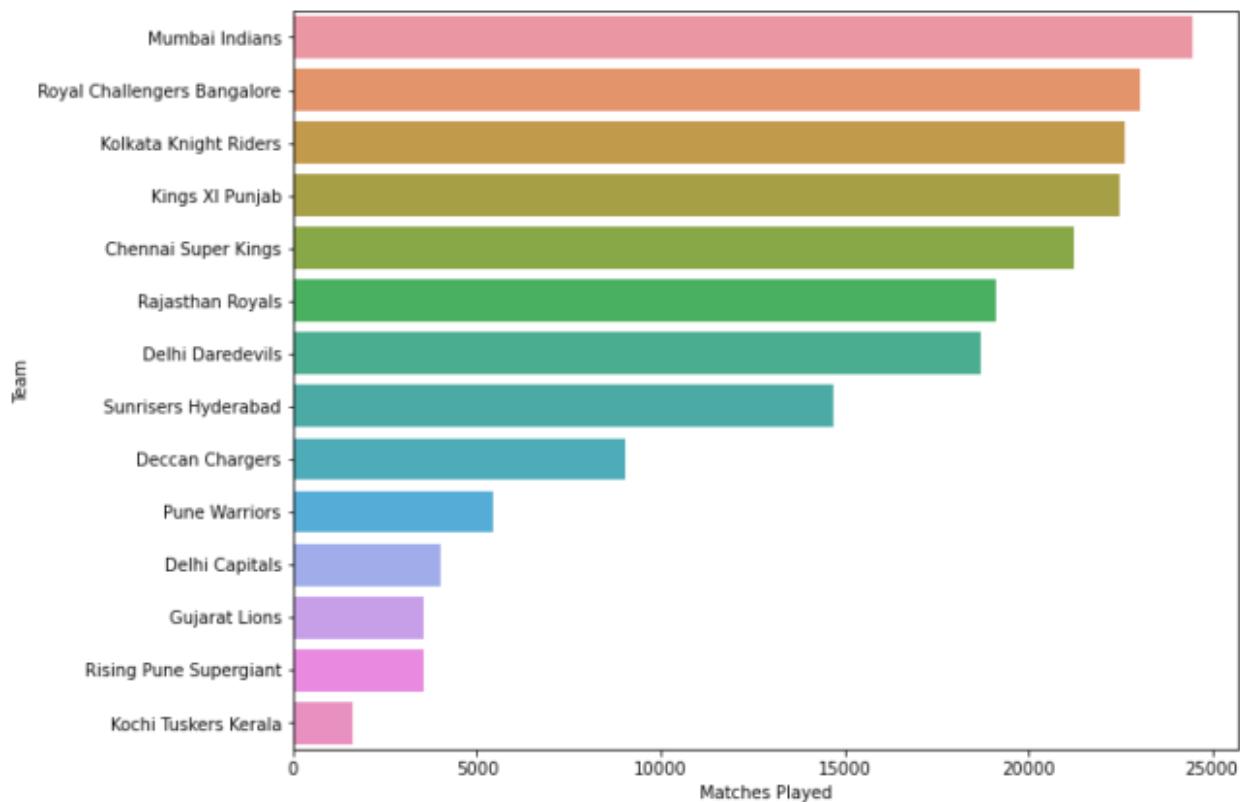
**SEASON WISE IPL MATCHES**



Source: <https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/>

The purpose of the visualization is to convey the number of matches played in each season. Bar plot has been used to convey this information. In the context of this visualization, the plotting technique is appropriate. The number of matches played in each season are represented by using length. Length is very effective in terms of conveying quantities especially for representing the differences in the quantities. As it can be observed from the plot that the seasons of 2011, 2012 and 2013 had considerably more number of matches as compared to other seasons. For the rest of the seasons, it is quite observable that there have been some differences in the number of matches played and the difference gets highlighted well in a bar plot.

#### **IPL MATCHES PLAYED BY EACH TEAM**



Source: <https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/>

This plot intends to convey the total number of matches played by all the teams across all seasons. However, the information conveyed is not correct. The X-Label is supposed to represent the total matches played. In this plot, Mumbai Indians have played the most number of matches (which is factual) but are nowhere near the value of 25000. The maximum number of matches a team can play in each season is 17. The total number of seasons played till 2020 is 13. Even if we assume that Mumbai Indians played 17 matches in all the seasons (which is not true), then the total number of matches will not be more than 221. Except for the values on the x-axis, bar plot is appropriate for conveying this information as it uses length.

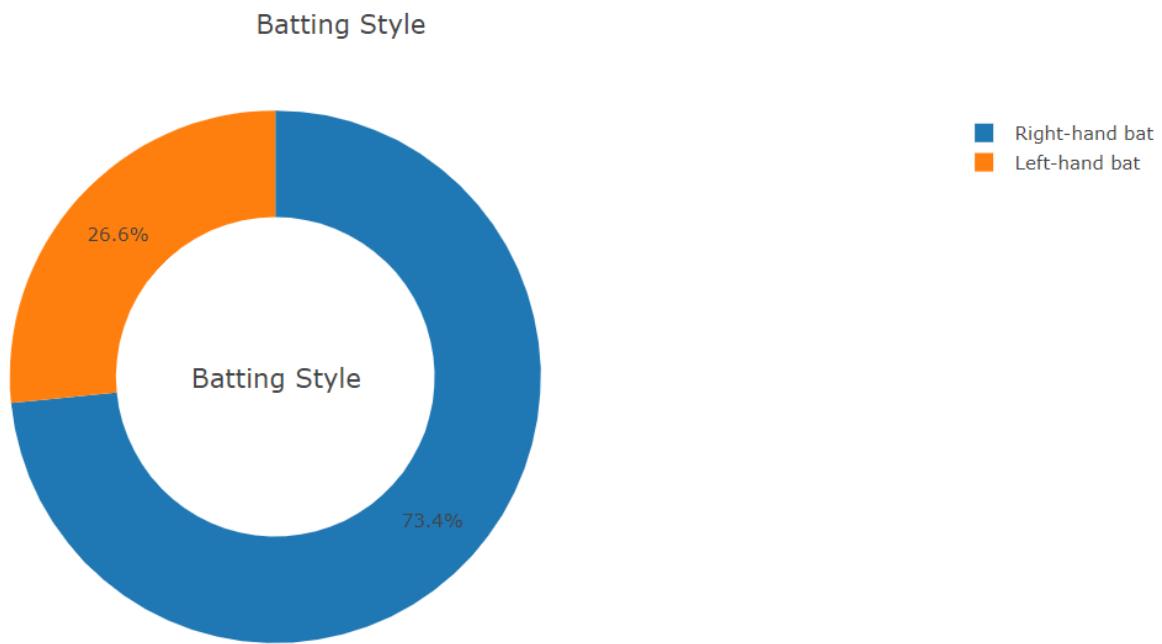
## TOP ALL-ROUNDERS



Source: <https://www.sciencedirect.com/science/article/pii/S1877050917327023>

The above visualization depicts the total runs scored by cricketers who are considered as “all-rounders”. All-rounders are those cricketers who can bat as well as bowl with a good degree of consistency in both. The author here has used a combination of points to depict the exact total runs scored by each batsman and lines between them (interpolation) to make the visualization. In terms of technicality, there is no relation (as depicted by lines) between the runs scored by each batsman. However, the visualization is able to convey the basic message of total runs scored by these batsmen.

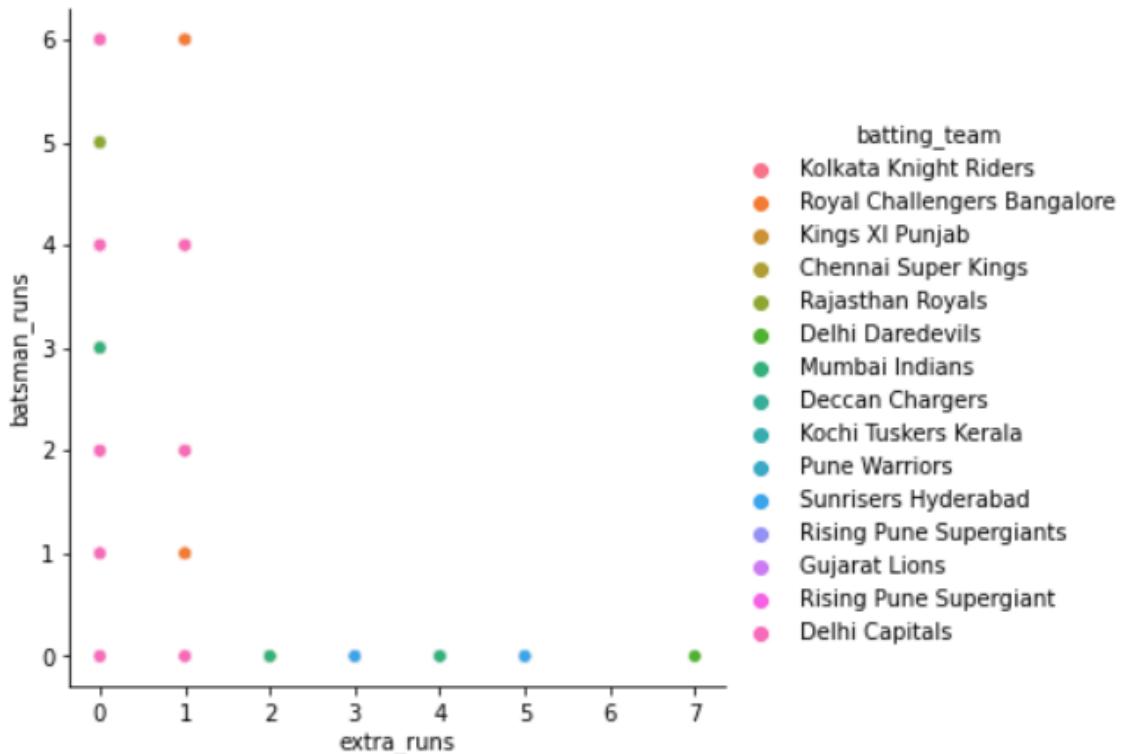
## BATTING STYLE



Source: <https://www.kaggle.com/rishpande/indian-premier-league-ipl-data-visualization>

In terms of Batting style, a batsman can be a right-handed batsman or a left-handed batsman depending upon whether the batsman is right-handed or left-handed respectively. This visualization depicts the percentage of right-handed and left-handed batsmen playing in the IPL. The author here uses a hollow pie chart to make the visualization. In the context of this particular visualization, using a pie chart is good as the number of categories is only two and also the difference in their percentages is significant. We believe that perhaps the annotation could have been done in a better way especially for the chuck that represents the right-handed batsmen. There is not enough contrast between the blue color and the annotated percentage thereby making it slightly difficult to read. Annotation could have been done outside of the colored chunks utilising the white background.

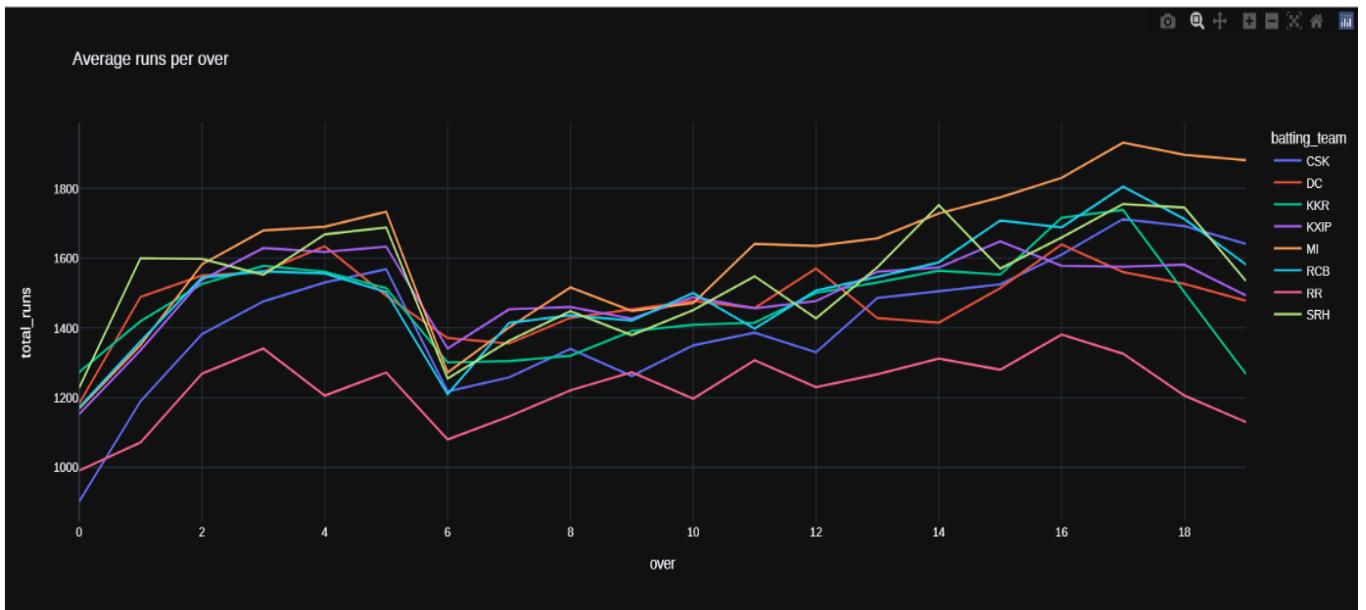
## **EXTRA RUNS vs BATSMEN RUNS**



Source: <https://www.kaggle.com/shashankpatel0111/eda-on-ipl-data>

Extra run, also known as “sundry”, is a run awarded to the batting side which is not scored by any individual batsman of the batting side. These runs are not credited to any batsman. There are five types of Extra namely *No-Ball*, *Wide*, *Bye*, *Leg Bye* and *Penalty run*. There are multiple problems with this visualization. Firstly, a lot of categories are represented by colors. Colors are not a proper way of representing data when there are a significant number of categories as it becomes difficult to differentiate between them. Apart from that, the purpose of this visualization is unclear. The relationship between *Extra runs* and *Batsman runs* that the author here is trying to convey (there is no direct relationship between these two variables) is not explained in the plot making the plot very difficult to understand.

## AVERAGE RUNS PER OVER SCORED BY EACH TEAM

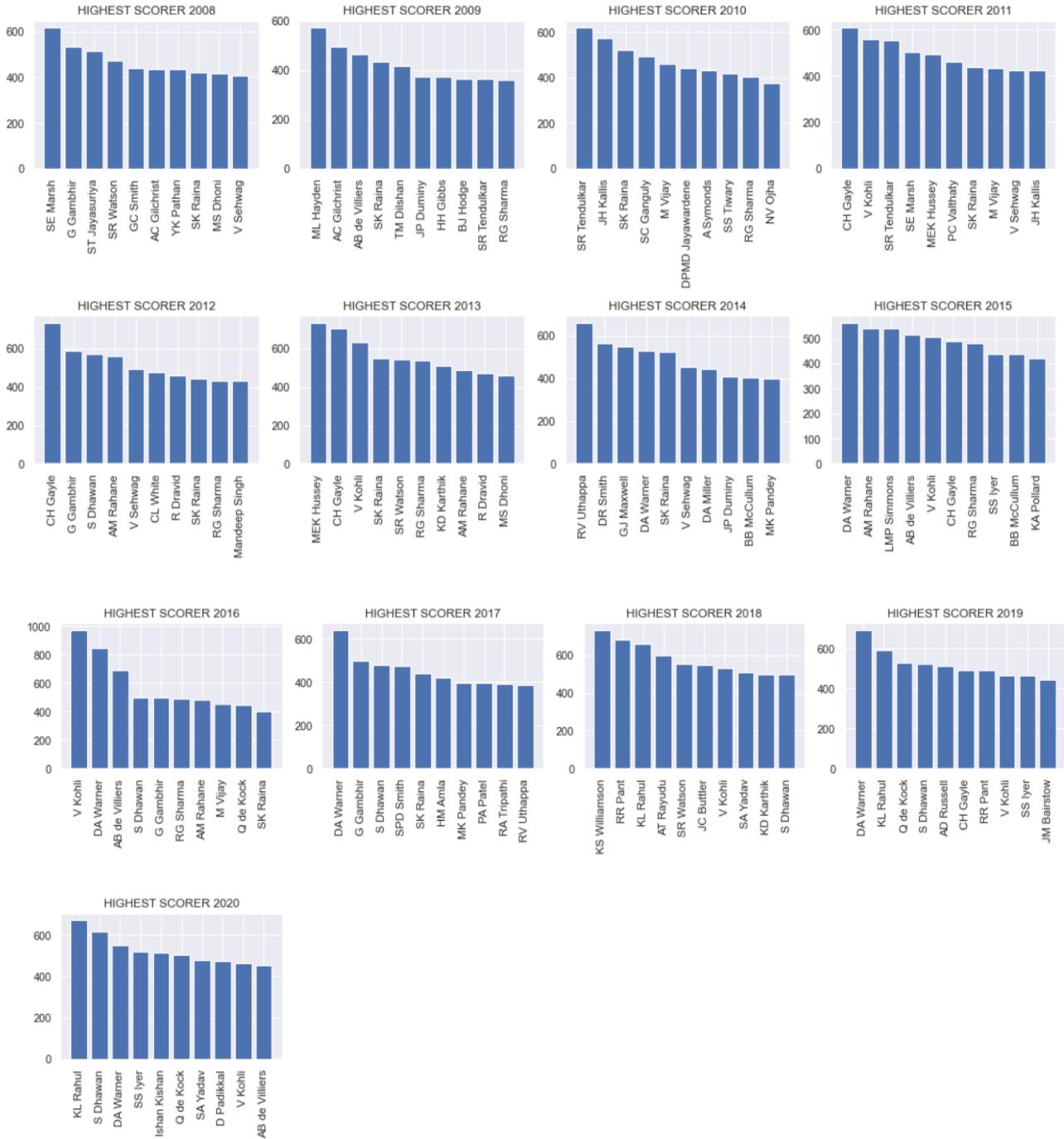


Source: <https://github.com/Sankethsp/IPL-data-visualization-using-Python/blob/main/IPL%20Data%20Visualization.pdf>

This purpose of the plot is to depict the average runs scored by all teams in each and every over of the game. The author here uses lines for visualization. Even though the visualization appears to be quite appealing, there are some mistakes with the visualization. The stated purpose of this plot is to show the average runs per over by all teams. However, the author has actually plotted the total runs scored per over by all teams. As far as the visualization is concerned, it can get a bit difficult to comprehend especially where the lines are at close proximity to each other. A better way to visualize this plot could have been using points/scatters with some opacity instead of lines. That would have conveyed the information in a better manner.

## Our Visualizations

### TOP 10 HIGHEST SCORING BATSMEN IN EACH SEASON



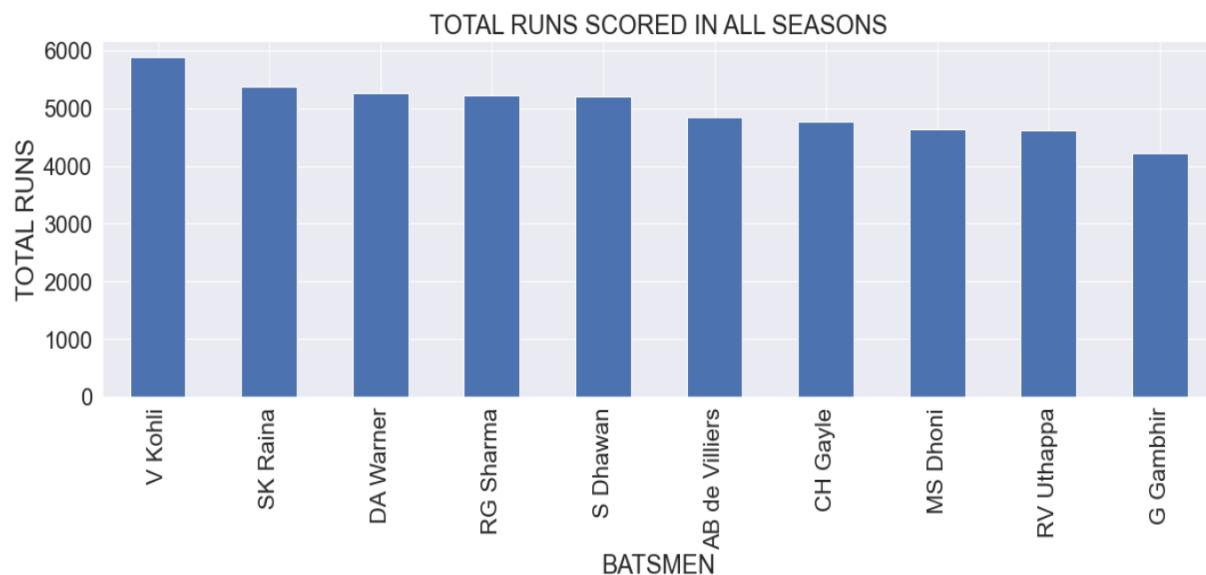
This plot visualises the Top 10 highest run scorers in each season of IPL. We have used a bar plot to create this particular visualization. This is because barplot uses length to represent the

quantities and differences among them. It is an effective way to create this kind of visualization to depict quantities as it leverages the findings of Steven's power law that fundamentally states that length is the most effective way of depicting quantities and differences among them. The purpose of this visualization is to determine which players have been consistently performing well in each and every season of the league and also their rank in terms of scoring in each season. This information is highly valuable for the team owners to decide whether they want to purchase the services of a batsman based on their consistent performance or if they want to retain a particular batsman for the upcoming seasons.

#### *Key Insights:*

1. In most of the seasons, the top scorer scored in the range of 500 to 700.
2. *V Kohli* scored close to 1000 runs (973) in a single season of 2016 which is the highest ever for a single season. It is a record that is very difficult to eclipse.
3. *SK Raina* has been the most consistent batsman in the history of IPL getting featured regularly in the Top 10 with a count of 9 times out of the 13 seasons played. He has never been the top scorer in any of the seasons but despite that he has been the most consistent batsman.
4. *DA Warner* has been the top scorer in three seasons (2015, 2017, 2019) and has featured in the Top 3 (excluding the first position) in two seasons (2016, 2020).
5. *RG Sharma* has featured six times in the list of Top 10 (2009, 2010, 2012, 2013, 2015, 2016).
6. *CH Gayle* has been the Top scorer in two seasons (2011, 2012) and has been in the Top 10 (excluding the first position) in three seasons (2013, 2015, 2019).

#### **TOP 10 HIGHEST SCORING BATSMEN ACROSS ALL SEASONS**

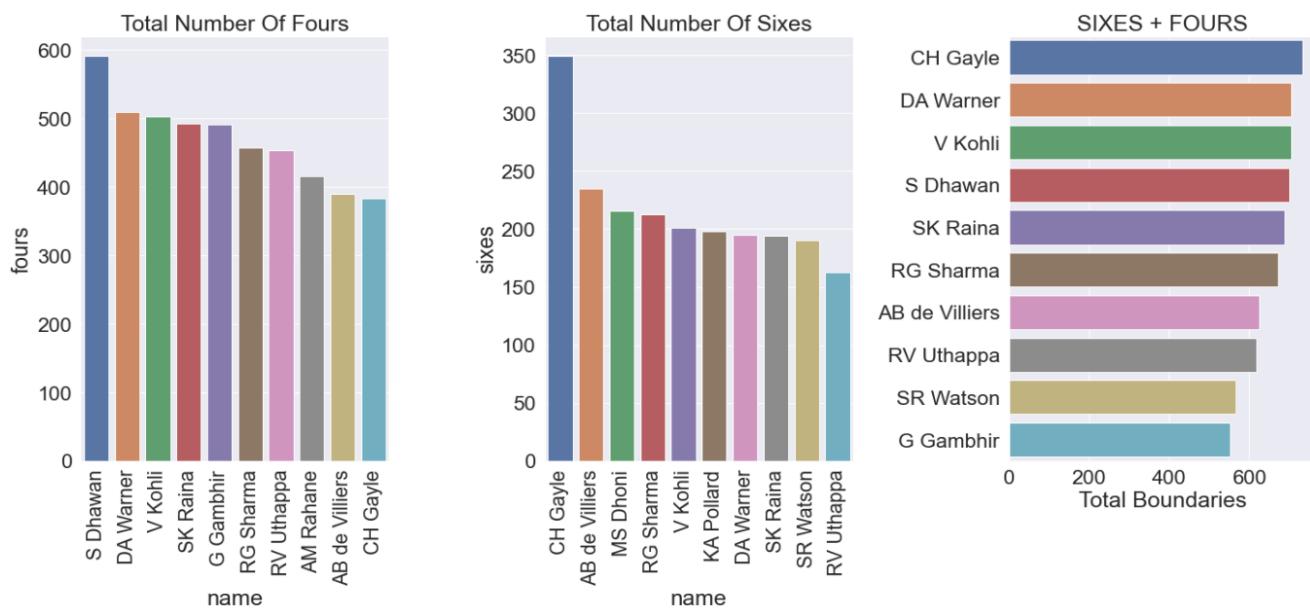


This visualization represents the Top 10 highest scoring batsmen across all seasons of IPL. We have used the barplot to create this visualization for the same purpose as stated above.

#### *Key Insights:*

1. *V. Kohli* is the highest run scorer across all seasons in the history of IPL.
2. *SK Raina* comes at the second spot.
3. Even though *SK Raina* has been a more consistent player than *V Kohli*, the performance of *V Kohli* in the season of 2016 in which he scored close to 1000 runs managed to put him ahead of *SK Raina* in terms of total runs.

#### **TOP 10 TOTAL BOUNDARIES HIT BY BATSMEN ACROSS ALL SEASONS**

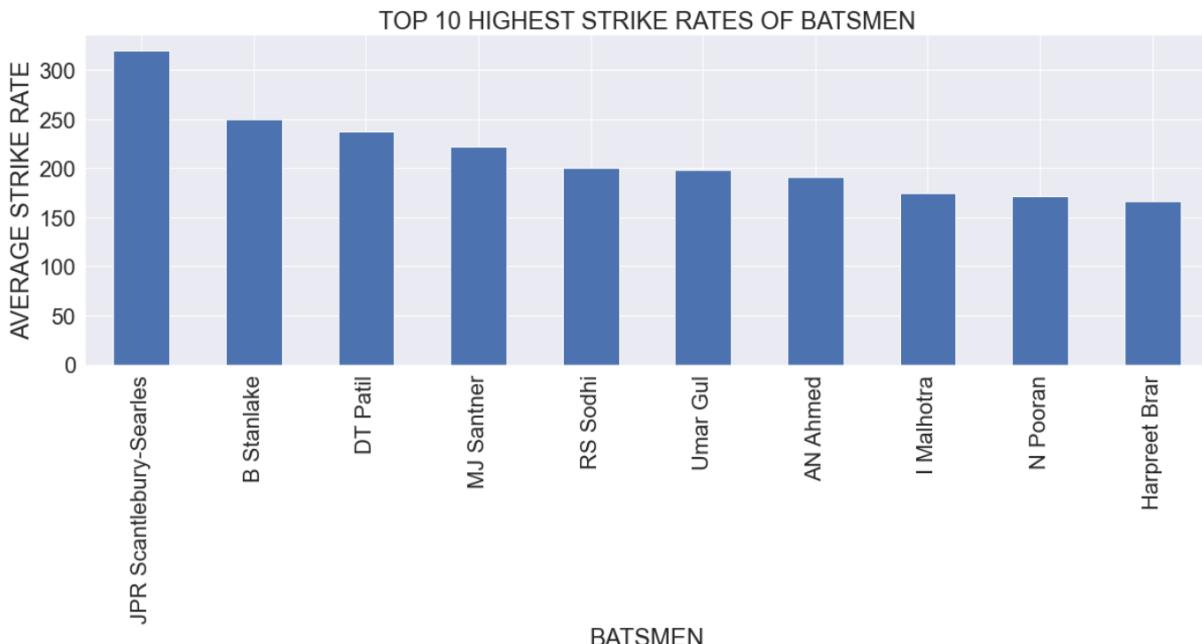


Boundaries consist of *fours* as well as *sixes*. We have used a barplot for this visualization for the purpose stated above.

#### *Key Insights:*

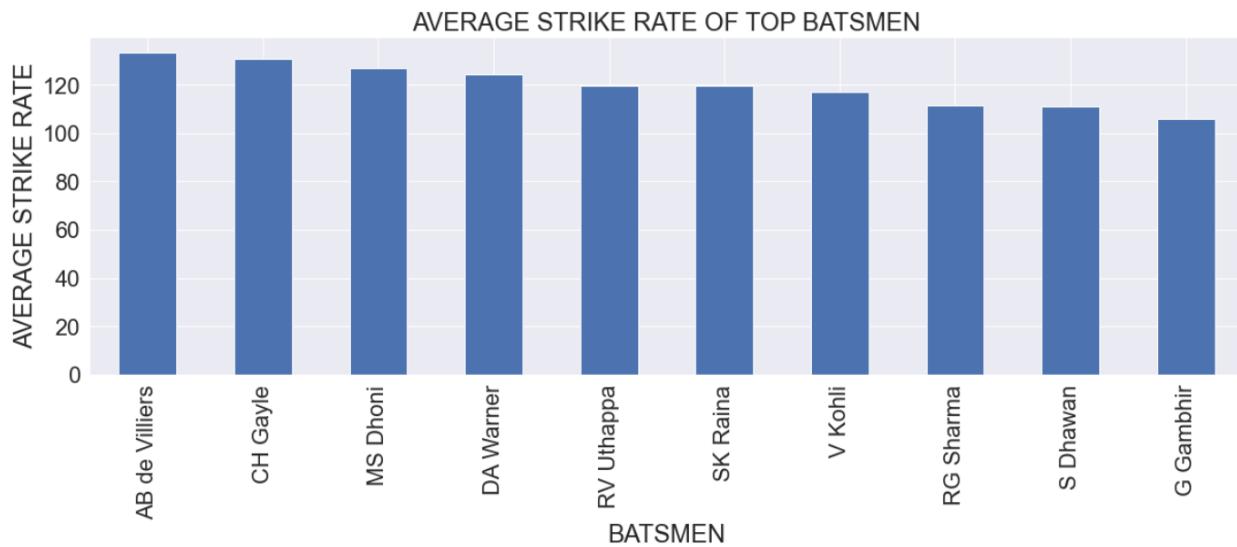
1. In terms of total *fours* scored across all seasons, *SA Dhawan* comes out on top and *CH Gayle* is at the lowest in the Top 10.
2. In terms of total *sixes* scored across all seasons, *CH Gayle* comes out on top and leads by a significant margin.
3. The difference in *sixes* is so huge that it manages to place *CH Gayle* on top in terms of the total boundaries.
4. *CH Gayle* is the biggest hitter of the ball!
5. IPL is based on the T20 format of the game and this quality of scoring high number of boundaries is one of the most valuable assets for a batsman in the T20 format.

## TOP 10 HIGHEST STRIKE RATES FOR BATSMEN



Batting strike rate measures how quickly a batsman scores runs given the number of deliveries (balls) the batsman faces. It is calculated by dividing the total runs the batsman has scored by total deliveries the batsman has faced. This ratio is then multiplied by 100 to compute the strike rate of the batsman. This visualization intends to determine the Top 10 highest strike rate of batsmen across all seasons. However, there is a fundamental problem with this visualization. Strike rate is a function of the number of balls faced and not the overall score of the batsman. Even if a batsman faces one delivery and manages to score one run and then gets dismissed will have a strike rate of 100. If the same batsman manages to score two runs on this one delivery then the strike rate will be 200. Therefore, the plot in this form does not convey the intended information of highlighting the strike rates of the regular batsmen. We say that because the players highlighted in this plot are actually regular bowlers and not regular batsmen. In fact, with this visualization, we were able to identify this discrepancy. Hence, we calculated the strike rates of regular batsmen who have scored highest total runs in the IPL.

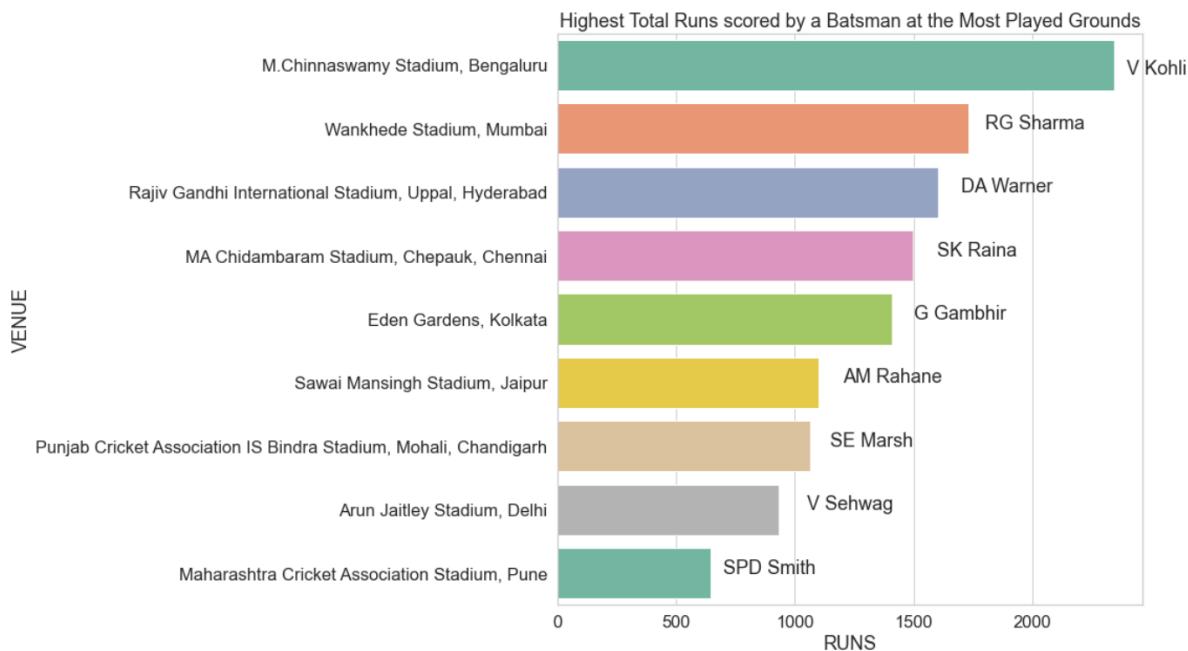
The correct visualization is as follows:



*Key insights:*

1. *AB de Villiers* has the highest strike rate among all batsmen followed by *CH Gayle*.
2. Even though *CH Gayle* has significantly more number of boundaries scored to his credit as compared to *AB de Villiers*, this plot conveys that *AB de Villiers* is more consistent in terms of scoring runs quickly in terms of running between the wickets as compared to *CH Gayle*.
3. Overall, all these batsmen have comparable strike rates.

#### HIGHEST TOTAL RUNS SCORED BY BATSMEN AT THE MAIN VENUES

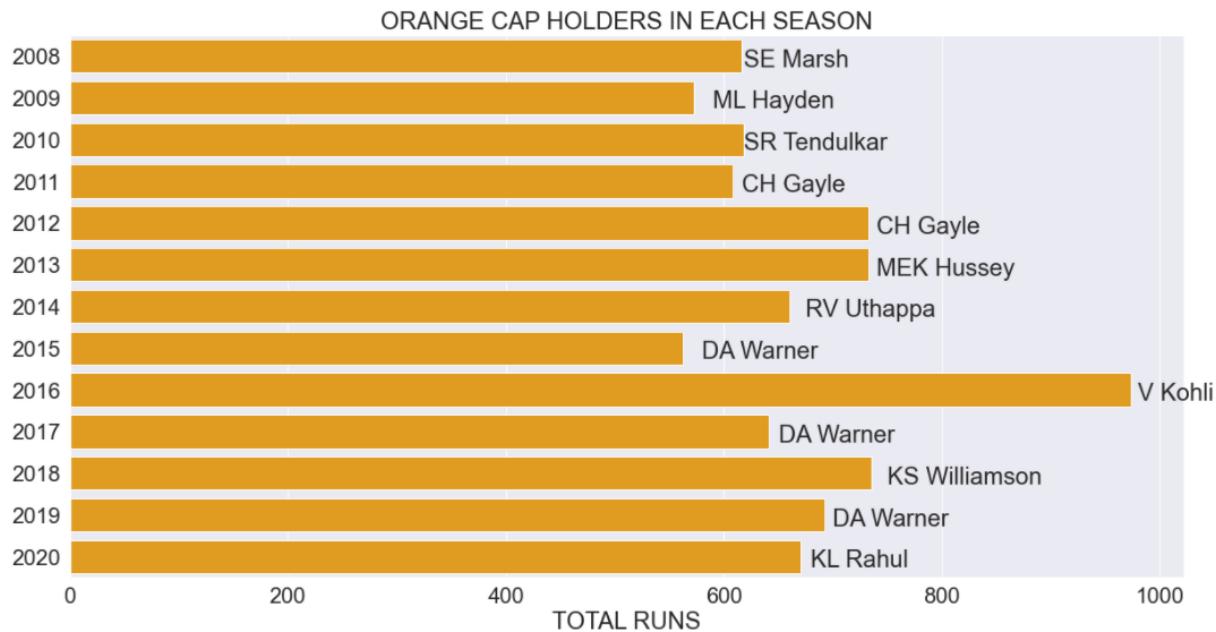


The aim of this plot is to highlight the highest scoring batsmen in terms of total runs at the main venues for IPL.

#### *Key Insights:*

1. Almost all the players (*V Kohli, RG Sharma, DA Warner, SK Raina, G Gambhir, AM Rahane, SE Marsh*) highlighted in this plot scored maximum runs at their home grounds.
2. *V Sehwag and SPD Smith* initially belonged to the teams of Delhi and Pune respectively and hence the same inference can be drawn for them as well.
3. This plot indicates that players tend to perform very highly at their respective home grounds.

#### **ORANGE CAP HOLDERS**

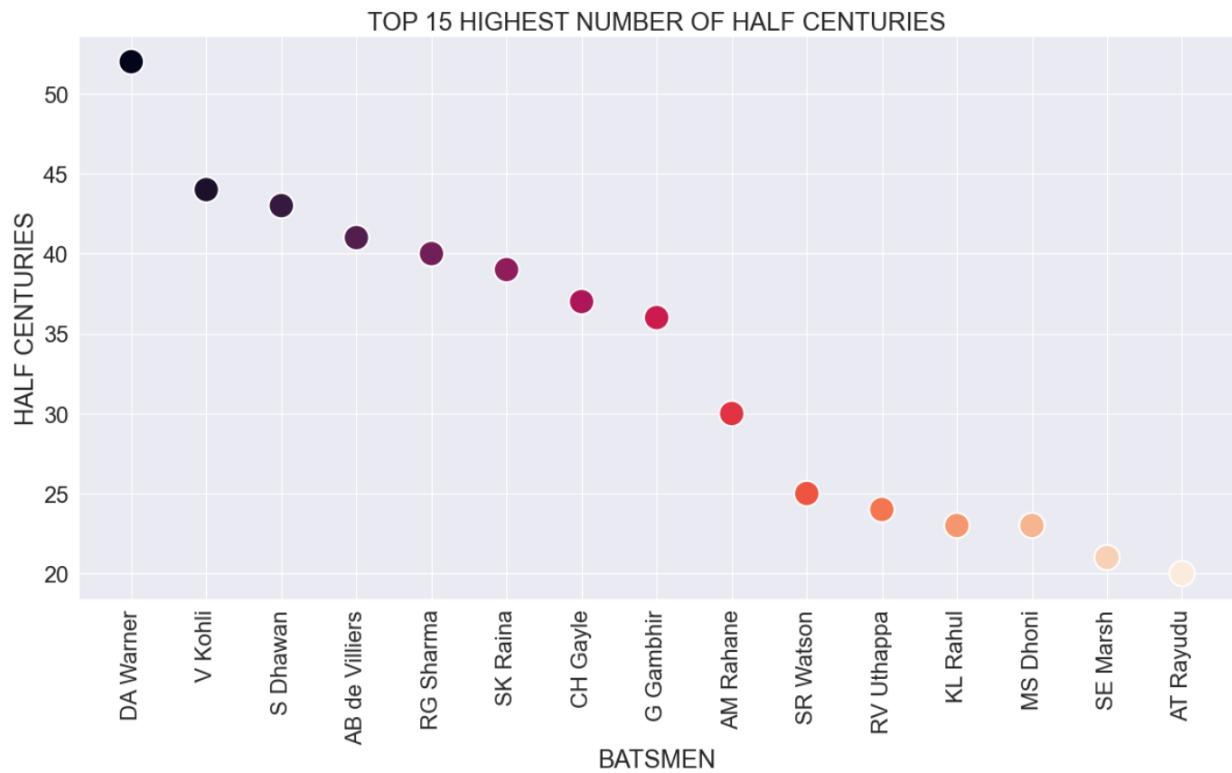


This visualization aims to highlight the orange cap holders (top scorers) in each and every season.

#### *Key Insights:*

1. *DA Warner* has received the orange cap title the maximum number of times, that is three times, in the history of IPL. He received the title in 2015, 2017 and 2019.
2. *CH Gayle* comes at second with the title credited to his name a total of two times in 2011 and 2012.
3. Rest of the players have received the title only once.

## TOP 15 HIGHEST NUMBER OF HALF CENTURIES

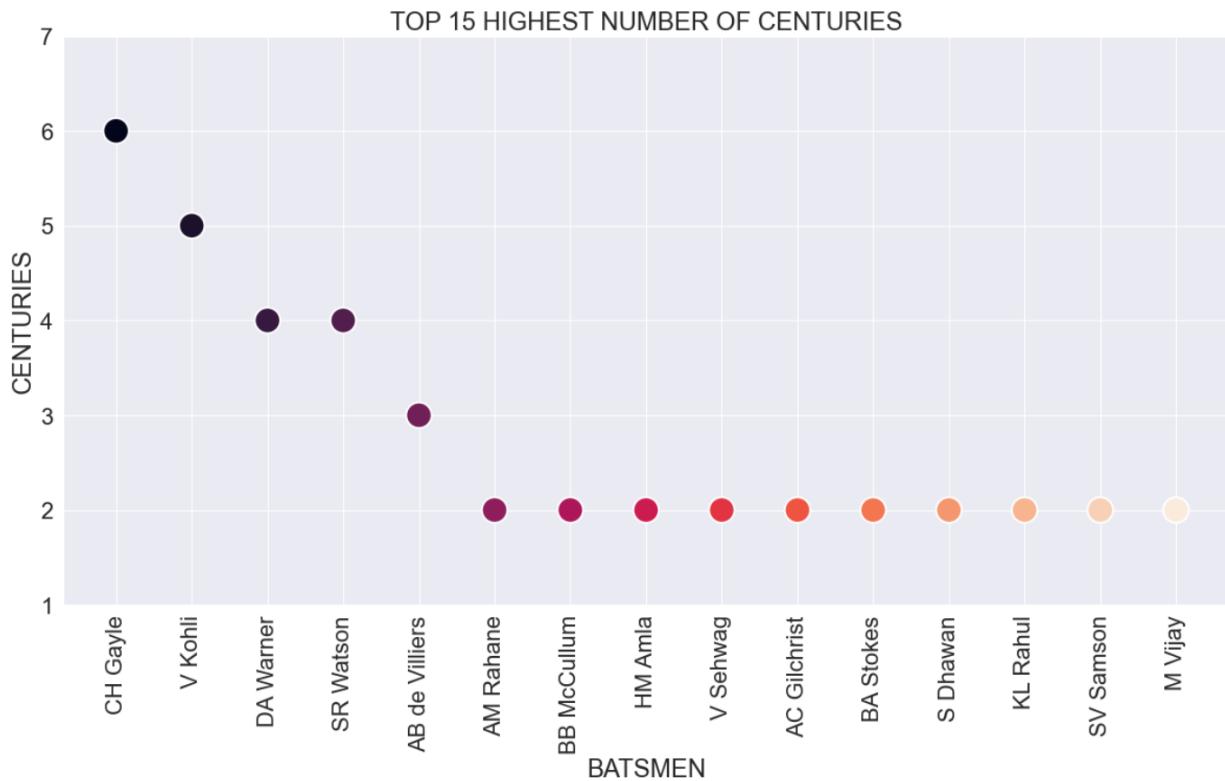


When a batsman scores at least fifty runs in a single innings, then the score is considered as a half century. A half century is considered a good metric for evaluating a batsman because scoring a half century is more difficult in the format of T20 as compared to other formats of cricket. This is because the total number of overs the batting side can play are considerably lesser. To make this visualization, we used points (scatterplot) to depict the number of half centuries but essentially it uses length to convey the difference in the number of half centuries scored by different batsmen. On top of that, we have also used color gradients to make the visualization more appealing.

### Key Insights:

1. *DA Warner* has the highest number of half centuries credited to his name.
2. *DA Warner* is the only player in IPL history with more than fifty half centuries.
3. *V Kohli* is at the second spot with forty four half centuries. The difference between him and *DA Warner* is relatively significant.

## TOP 15 HIGHEST NUMBER OF CENTURIES



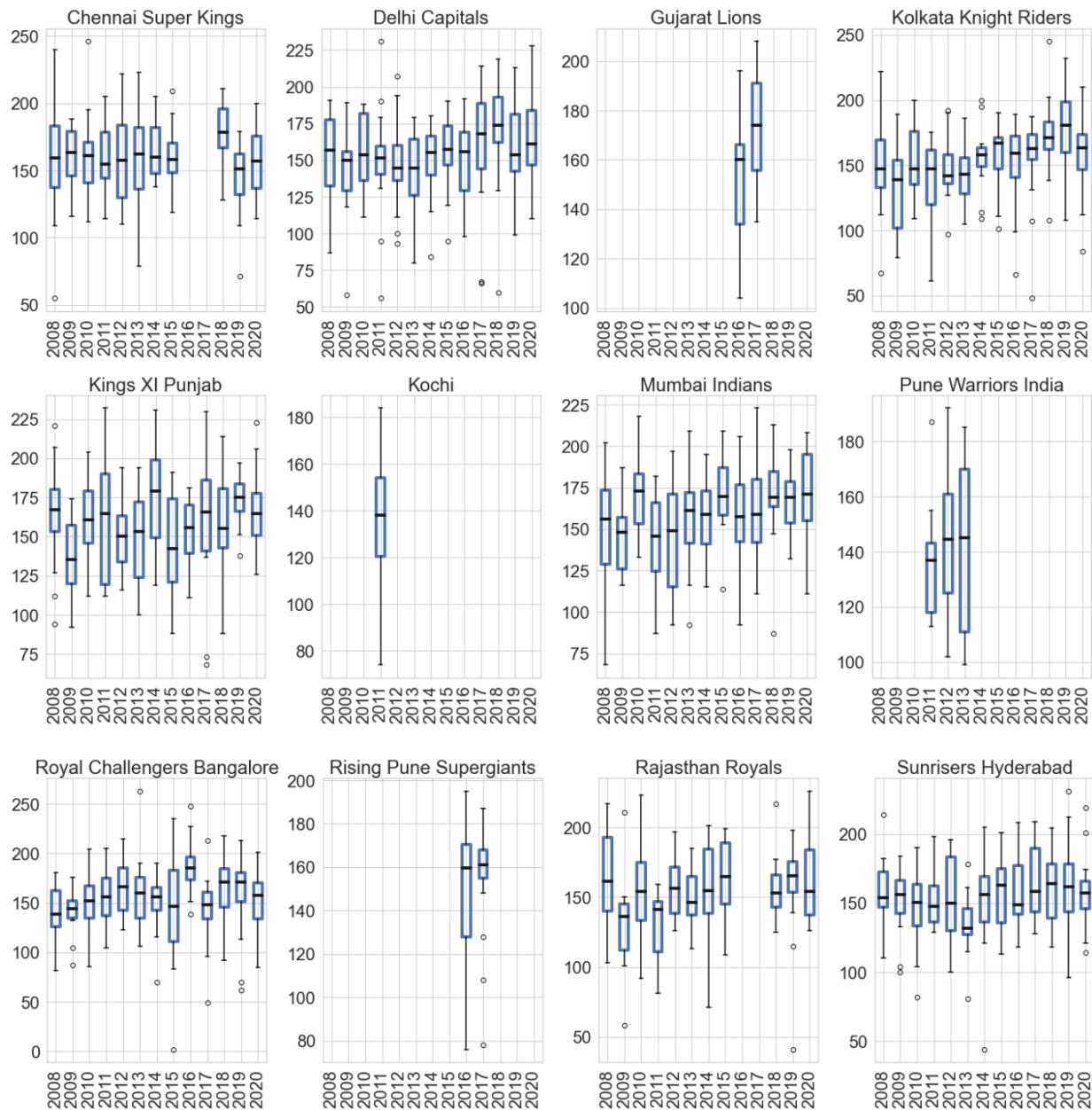
When a batsman scores at least hundred runs, then the score is considered as a century. Scoring a century in any format of cricket is generally considered an uphill task. Especially for T20 format, this task gets even more difficult given the less number of overs a batting side plays. Therefore, the number of centuries is a good indication of a batsman's quality.

### Key Insights:

1. *CH Gayle* has scored the maximum number of centuries totalling six.
2. *V Kohli* comes at the second spot with five centuries.
3. *DA Warner* and *SR Watson* both share an equal number of centuries, that is, four and *AB de Villiers* has scored a total of three centuries.
4. Rest of the batsmen in the list have scored two centuries each.

This visualization might give the wrong impression that there is a difference between batsmen who have scored an equal number of centuries since the plot appears to be a decreasing curve. However there is no difference among batsmen with an equal number of centuries. Since this visualization uses length, therefore, we believe that it is still a very effective way of conveying this piece of information.

## SUMMARY STATISTICS OF RUNS SCORED BY EACH TEAM ACROSS ALL SEASONS

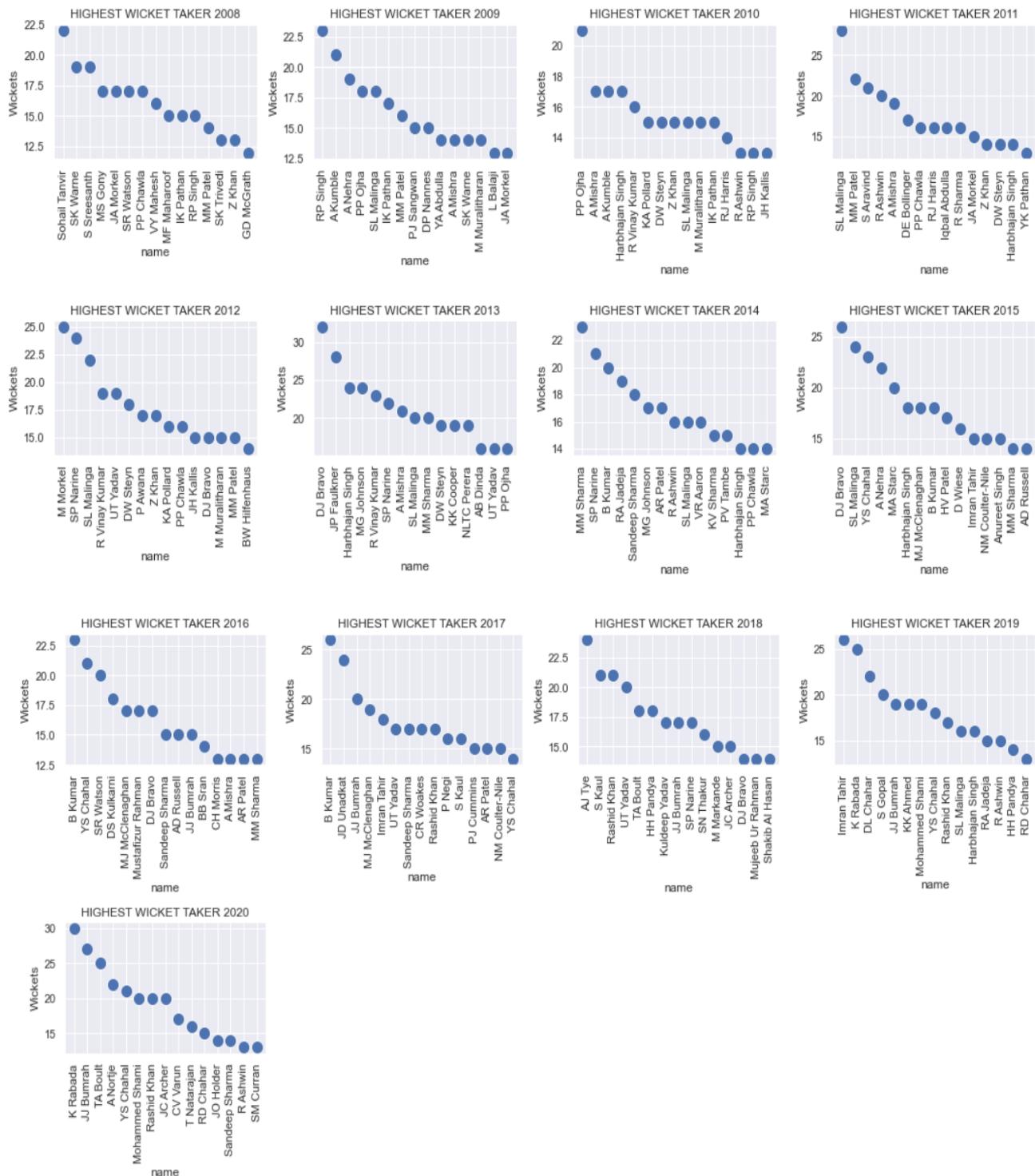


The purpose of this plot is to highlight the summary statistics (such as median and quartiles) of the runs scored by each team across all seasons. We have used Boxplot for the purpose of this visualization. Boxplot is very effective in terms of communicating the summary statistics of quantities and hence we are using it for this particular visualization.

*Key Insights:*

1. The median for all the teams across all seasons mostly ranges from 140 runs to 160 runs.
2. *Royal Challengers Bangalore* is the only team in IPL history to have scored more than 250 runs. They achieved this feat in the season of 2013.
3. Other teams that have come to the feat of 250 runs include *Chennai Super Kings*, *Kolkata Knight Riders*. They achieved this in the seasons of 2010 and 2018 respectively.
4. *Chennai Super Kings* did not play in the seasons of 2016 and 2017 as observed in the plot.
5. Teams such as *Gujrat Lions*, *Kochi*, *Rune Warriors India* and *Rising Pune Supergiants* played only for a limited number of seasons. Among these teams, only *Gujrat Lions* managed to score 200 runs at least once.
6. All the other regular playing teams have managed to score 200 runs in a single game.

## TOP 15 HIGHEST WICKET TAKERS IN EACH SEASON



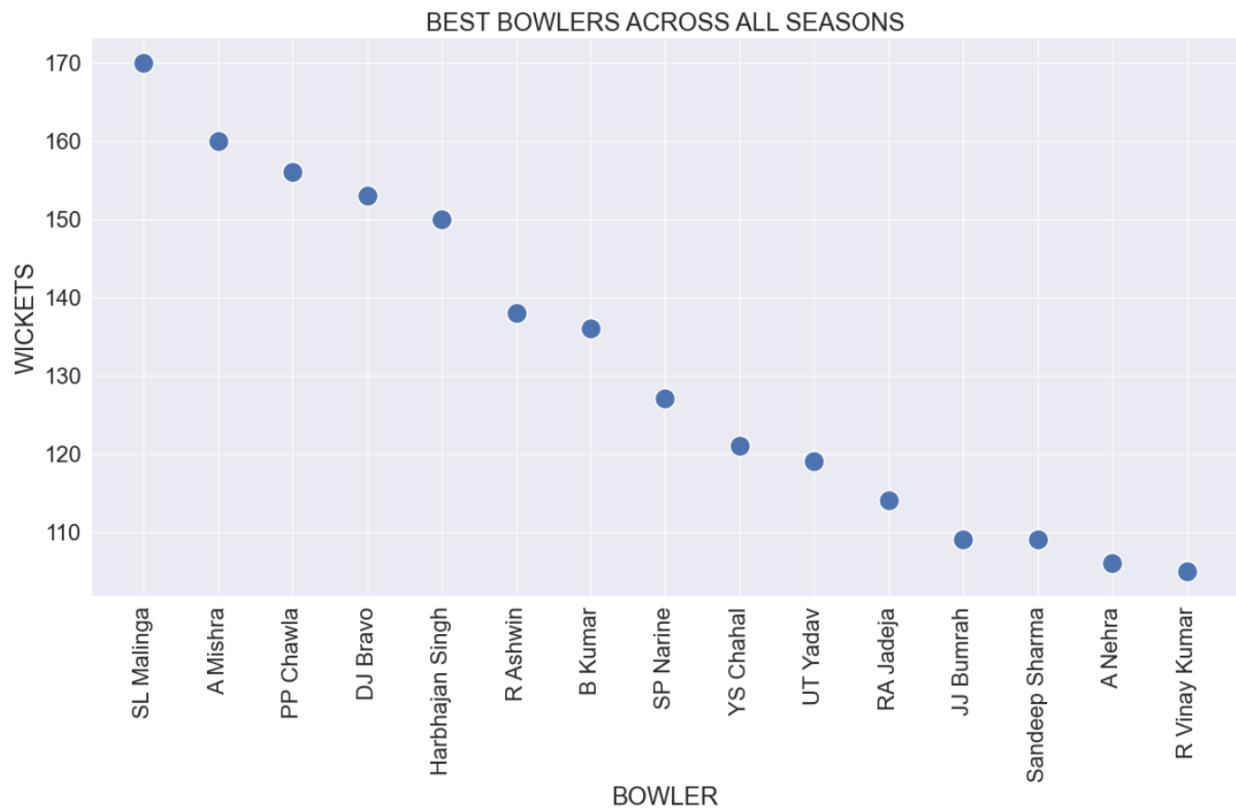
The aim of this plot is to visualize the highest wicket takers in each season. We have used points (scatter) to represent the wickets and not barplot because we thought that since we are

visualising multiple plots in the same figure therefore using points instead of bar plots would be more effective at readability. But essentially, this plotting technique also used length to visualize the number of wickets.

#### *Key Insights:*

1. Bowlers typically manage to pick up wickets in the range of 20 to 26 in each season.
2. The best bowling performance ever recorded in a single season in IPL history is credited to *DJ Bravo* in the season of 2013. He managed to pick up 32 wickets.
3. *DJ Bravo* is the only bowler in IPL history to have managed to pick up the highest number of wickets in a single season on two separate occasions, that is, in the season of 2013 and 2015.

#### **TOP 15 HIGHEST WICKET TAKERS ACROSS ALL SEASONS**



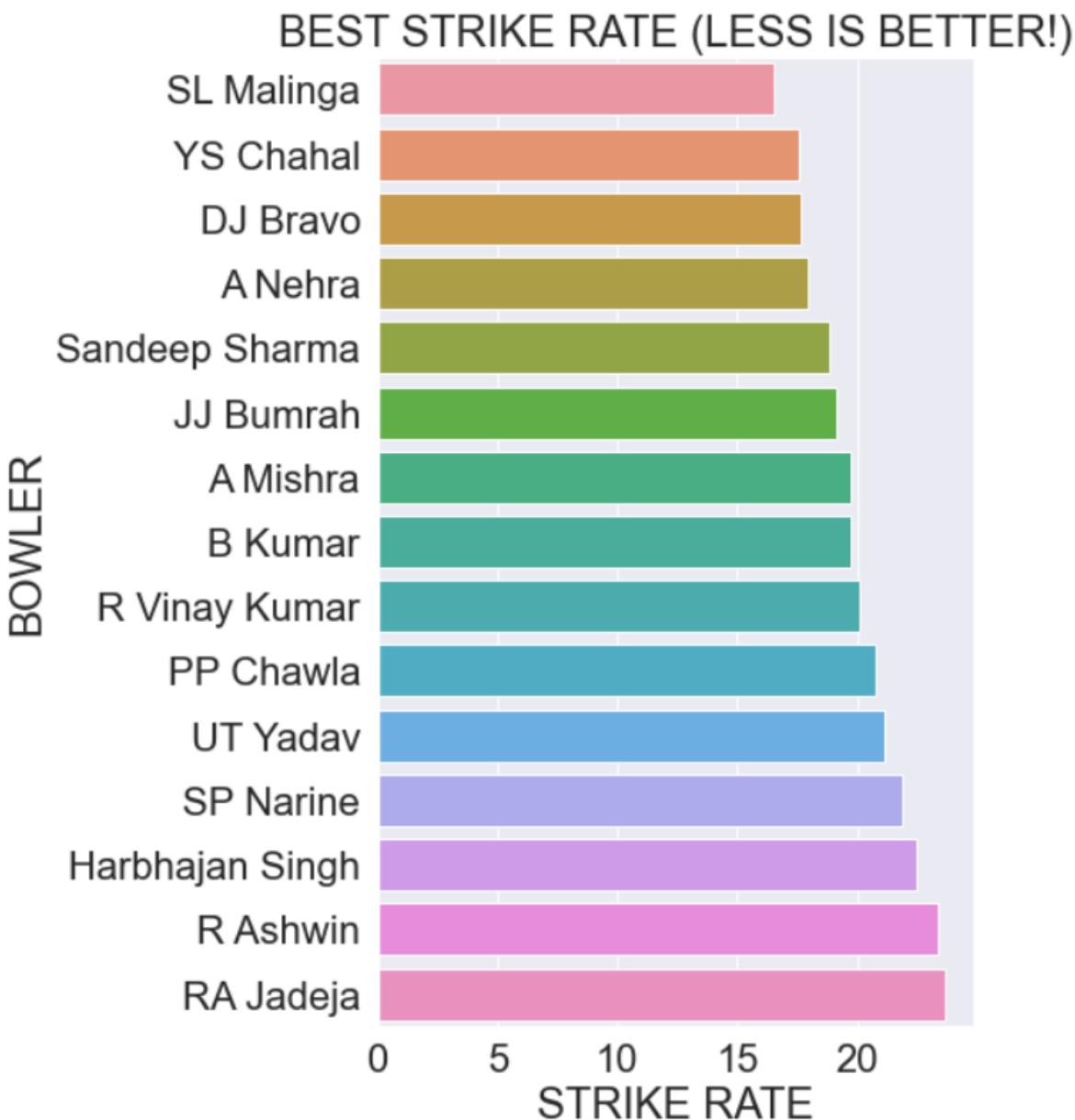
The plot here aims to highlight the best bowlers in terms of total wickets taken across all seasons. This is one of the metrics of evaluation for gauging the overall performances of bowlers along with strike rate and economy rate.

#### *Key Insights:*

1. *SL Malinga* comfortably rests on the first position with a record of 170 total wickets.

2. Even though *DJ Bravo* has been the highest wicket taker in a single season on two separate occasions (courtesy to previous plot), there are three bowlers who have been more consistent than him in terms of total wickets taken namely *SL Malinga*, *A Mishra* and *PP Chawla*.

**TOP 15 STRIKE RATE OF BOWLERS**

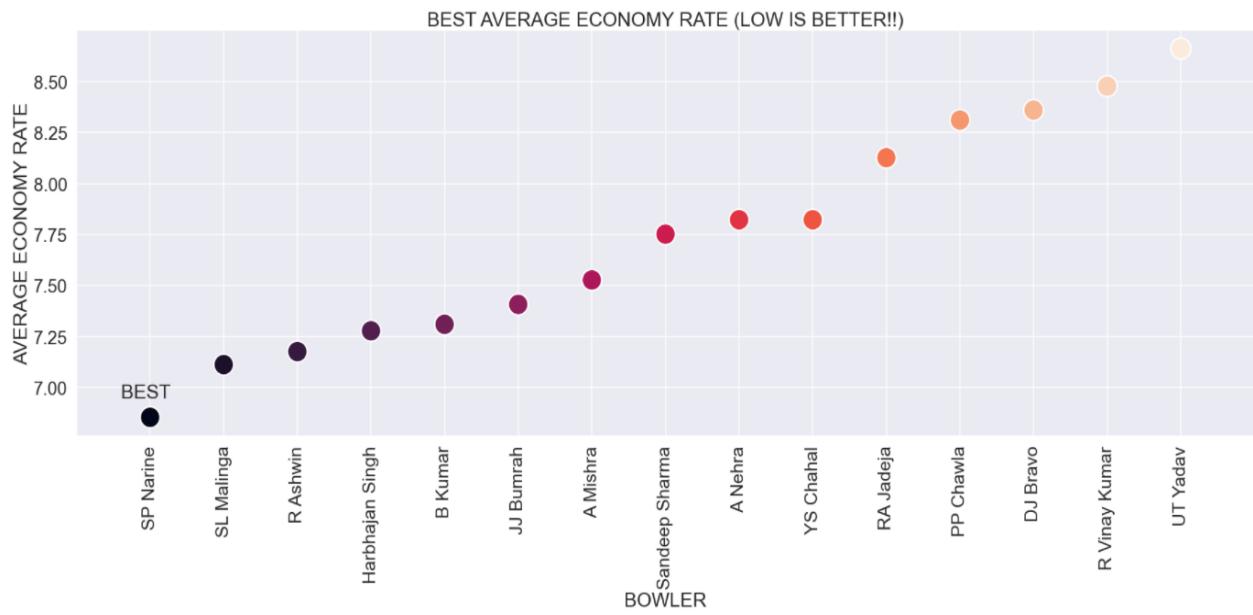


Strike rate for a bowler is defined as the average number of bowls bowled per wicket taken. The lower the strike rate, the more effective a bowler is at taking wickets quickly. This plot aims at visualizing the best bowlers in terms of strike rates.

#### *Key Insights:*

1. *SL Malinga* has the lowest strike rate (around 16) and hence is the best bowler as far as strike rate is concerned.
2. The difference in the Top 15 strike rates is not significant enough. However, the small differences are identified because of the usage of lengths.

#### **BEST AVERAGE ECONOMY RATE**

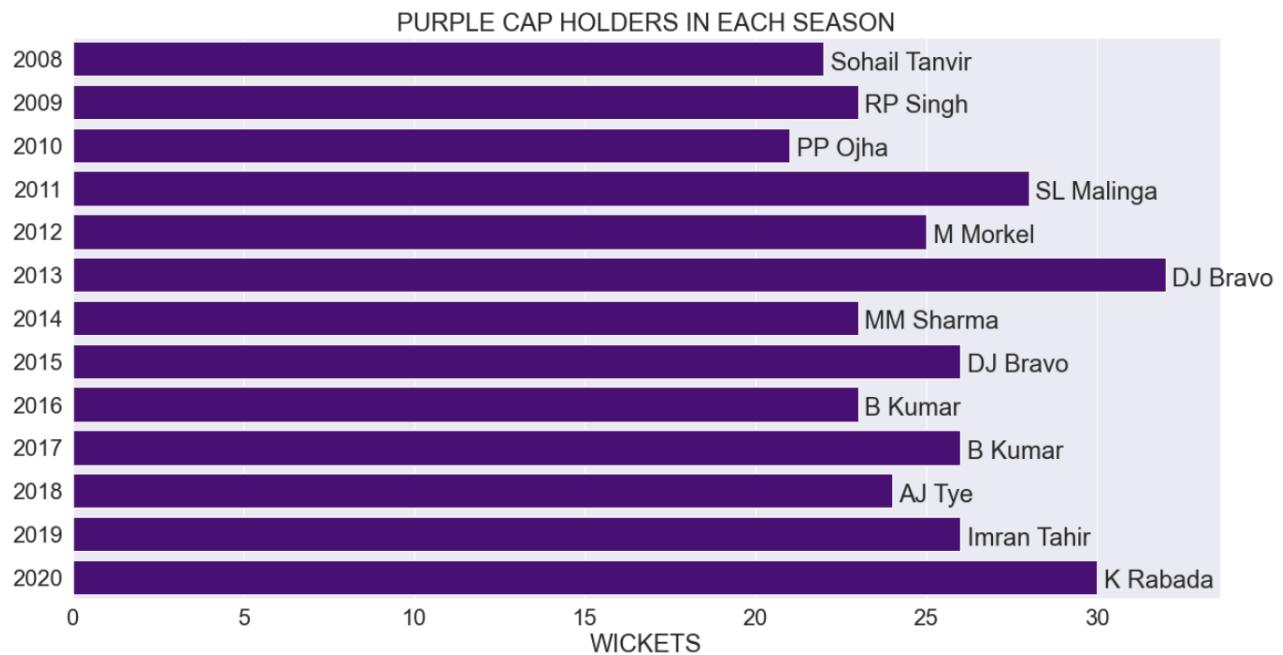


Economy rate of a bowler is defined as the average number of runs that they have conceded per over bowled. The lower the economy rate, the better is the performance of the bowler. This is specifically a very valued metric in T20 format. In this shorter version of the game, the primary aim of the batsmen is to score runs as quickly as possible. Therefore, the economy rate of a bowler in such a format of cricket, that is T20, is highly valuable.

#### *Key Insights:*

1. *SP Narine* has the best economy rate compared to all other bowlers.
2. *SL Malinga* comes at the second spot. The difference between *SP Narine* and *SL Malinga* is not very significant.
3. Taking into account the total number of wickets, strike rate as well as economy rate, it can be concluded that *SL Malinga* is the best bowler in IPL history.

## PURPLE CAP HOLDERS

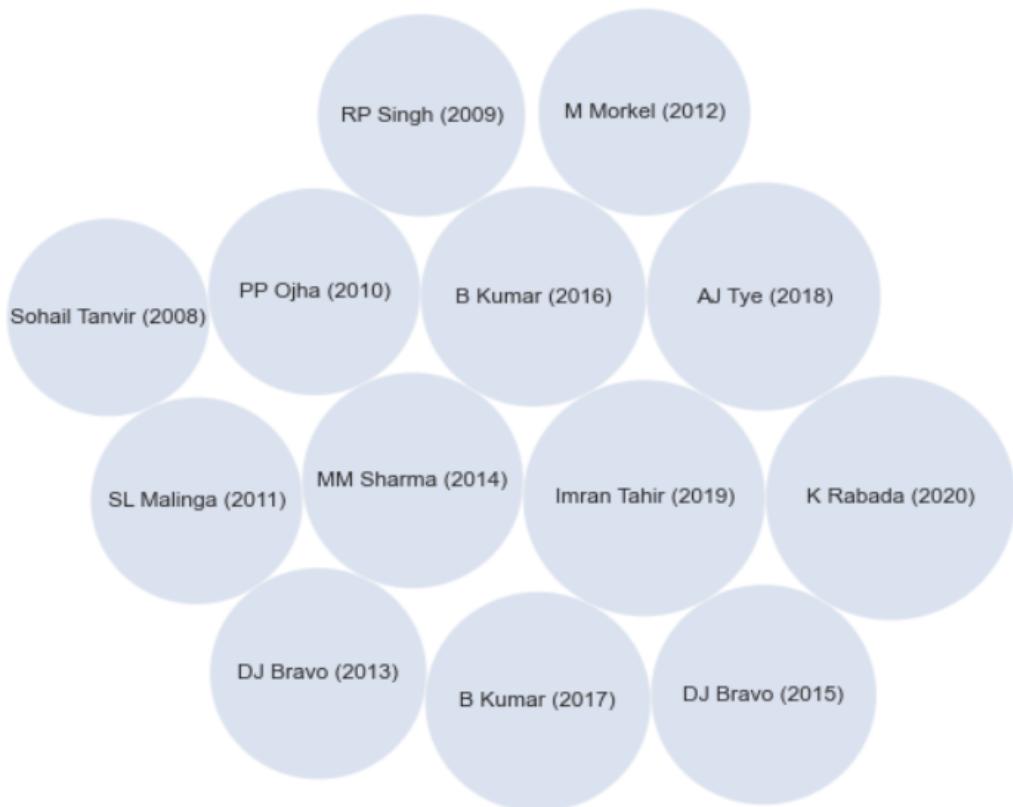


The aim of this plot is to visualize the highest wicket takers (Purple Cap holders) in each and every season.

### Key Insights:

1. As highlighted before as well, *DJ Bravo* is the only player in IPL history to be crowned with Purple Cap two times.
2. *DJ Bravo* is the only player exceeding 30 wickets in a single season which is clearly visible in this plot.

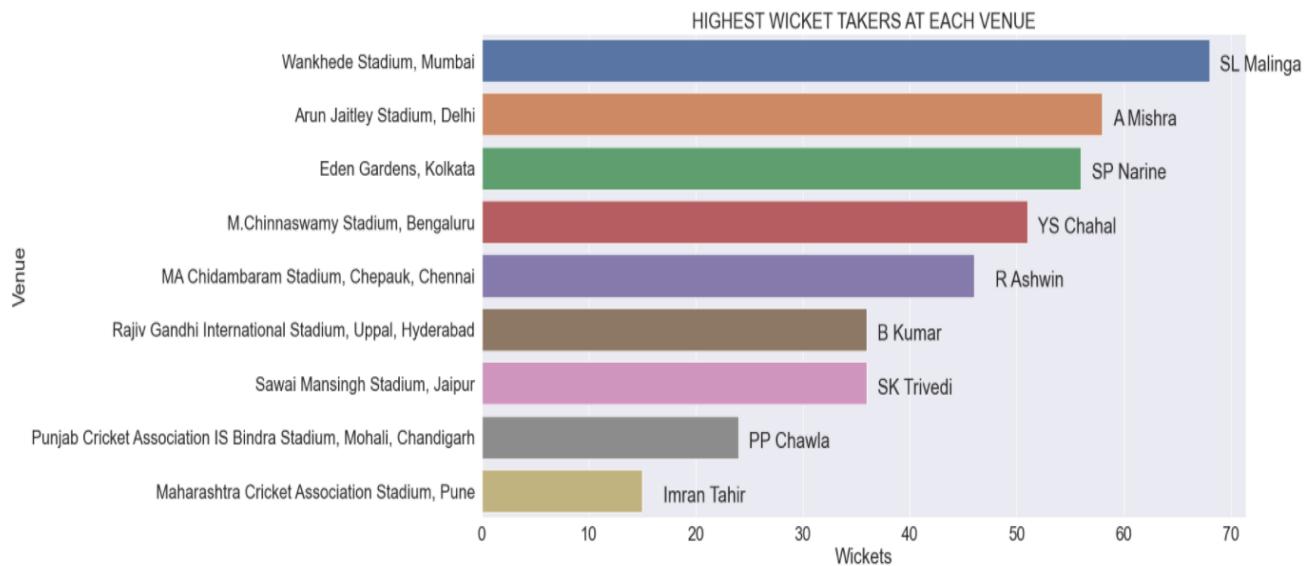
**AN ALTERNATE WAY OF VISUALIZING THE SAME INFORMATION (USING AREA INSTEAD OF LENGTH)**



*Area of the circle represents the total wickets taken by each Purple Cap holder in each season.*

We want to highlight the fact that lengths are better at visualizing quantities as compared to areas. That is what motivated us to plot the same piece of information of Purple Cap holders in each season using circular areas. As it can be observed that even though the purple cap holders in each season are quite apparent in this visualization, the difference in the total number of wickets (which is represented by the area) taken by each bowler is not easily perceptible. Hence, it proves that lengths are much better for representing quantities as compared to areas.

## HIGHEST WICKET TAKERS AT THE MAIN VENUES

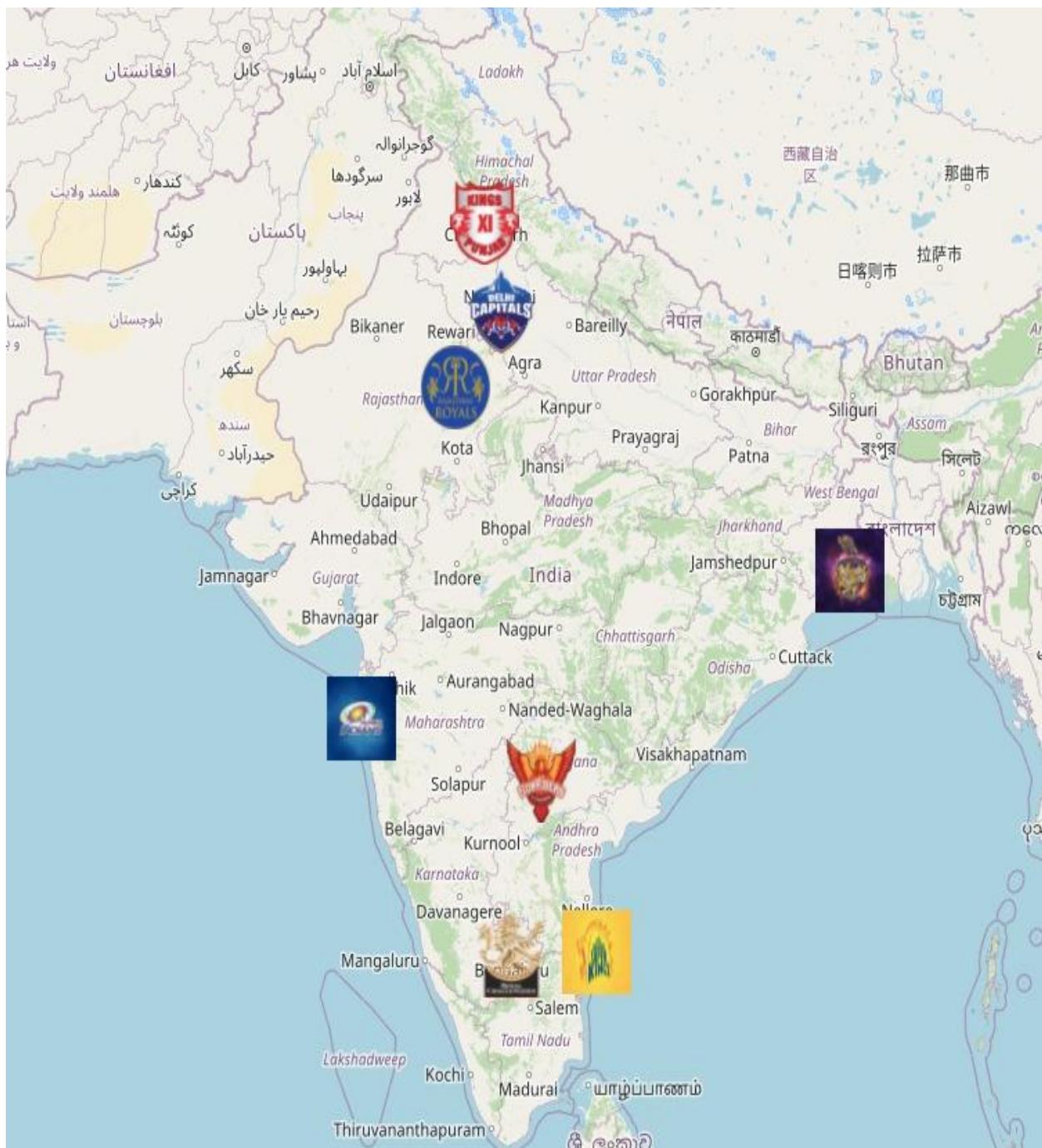


This plot aims at visualizing the best wicket taking performances in terms of total wickets taken by bowlers at the main venues of IPL.

### Key Insights:

1. Almost all the bowlers (except *Imran Tahir*) present in the plot are the highest wicket takers at their home grounds.
2. This information again establishes that players tend to be better performers at their home grounds as compared to other venues.

## HOME GROUNDS OF EACH TEAM



This visualization aims to represent the location of each team's home ground.

## VARIOUS LOCATIONS OF WHERE MATCHES WERE PLAYED THROUGHOUT THE ENTIRE IPL



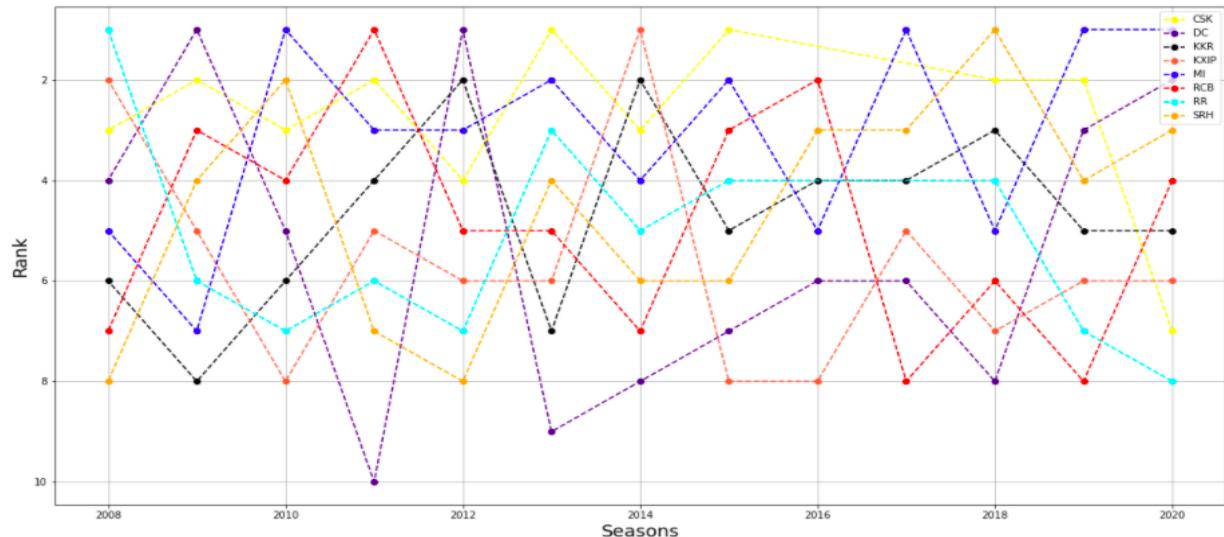
This visualization represents the locations of each cricket ground where various matches were played

### *Key Insights:*

1. We can see that the IPL has ensured that matches are played in various locations, thus giving a variety of outcomes to each game.
2. This ensures that each team has a fair and equal chance of doing well in accordance with their strengths and weaknesses.
3. This relates to each ground's pitch surface, size, etc.

## TEAM POSITION ACROSS SEASONS USING A BUMP CHART

A bump chart is a special form of a line plot designed for exploring changes in rank over time. Here, we've represented each team's position across all seasons using this technique.



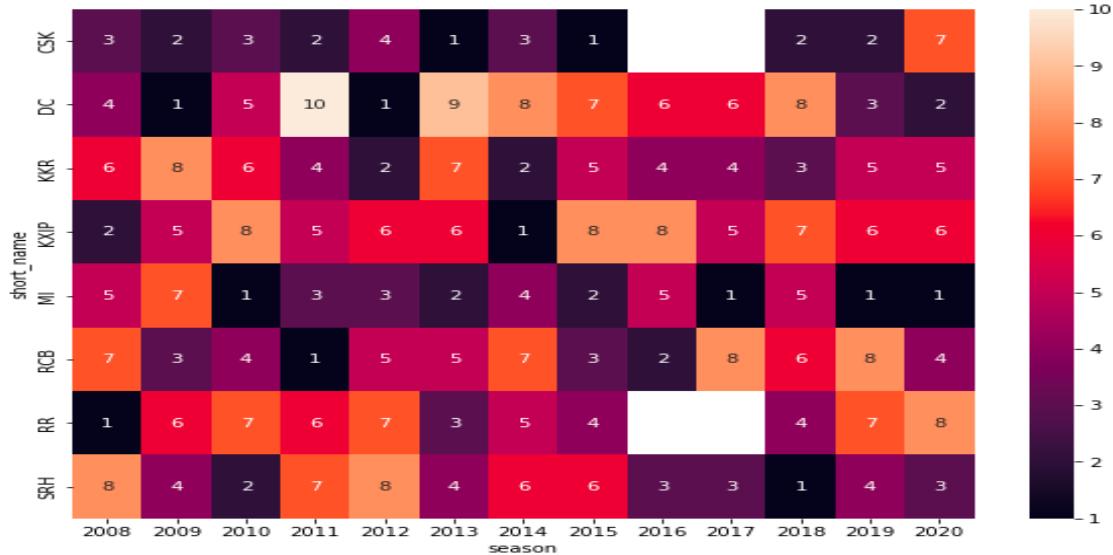
This visualization represents each team's performance across all the seasons of the IPL. Each team has been represented in accordance with their color.

#### *Key Insights:*

1. We can see that CSK has been fairly consistent by being in the top four across almost all the seasons.
2. The other teams however, have not been static and have been inconsistently playing either good, bad or even worse.
3. An inconsistent performance throughout the seasons shows bad team performance and coordination.

#### **TEAM POSITION ACROSS SEASONS USING A HEAT MAP**

A heat map shows the magnitude of a phenomenon as color in two dimensions. Here, we've represented the color densities as each team's performance across all seasons, ranging from 1-10.



This visualization is just like the previous one, it represents each team's performance across all the seasons of the IPL but, by using a heat map.

#### *Key Insights:*

1. By using a heat map, it gives us a direct overview of each team's performance across the seasons.
2. Here, a darker shade represents a stronger position, for example, the first position (1) is the best and thus, is represented by the darkest shade.
3. As we move to the lighter shades, it represents the weaker positions, for example, the tenth position (10) is represented as the lightest shade.
4. Again, we see that CSK has been fairly consistent throughout the seasons and is thus being represented by darker shades throughout.
5. The other teams, however, are inconsistent throughout and thus, the colors represented are highly volatile.
6. By using a heat map, we can easily differentiate the colors and there isn't any overlapping and therefore, easily understandable.

**TEAMS WITH THE MOST WINS USING A WORD CLOUD** - Word clouds are visual representations of words that give greater prominence to the words which appear more frequently. Here, the teams with more wins are represented by bigger words as compared to the teams with lesser wins.



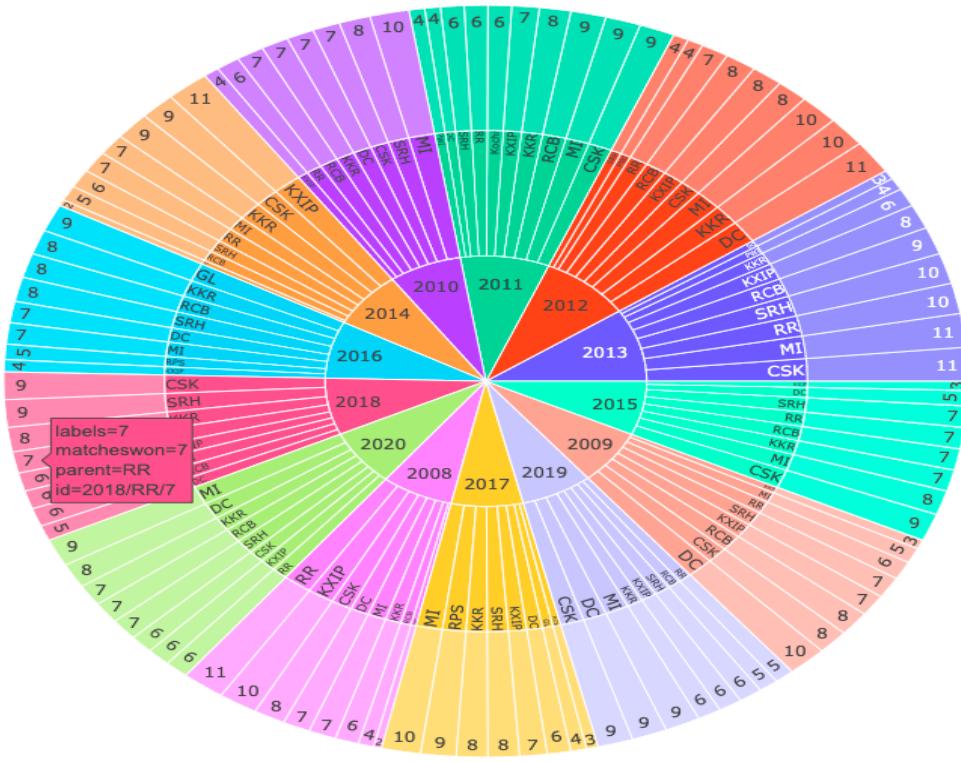
This visualization is a representation of the teams with the most wins by using a word cloud.

*Key Insights:*

1. In this word cloud, the size of the word is correlated to the number of wins by each team. Therefore, the teams with the most wins are represented with the larger sized words.
2. We can see that MI and CSK are the largest words which mean that they have won the IPL several times.
3. It also relates to the consistency of each team in the previous visualization.

## TEAMS WITH THE MOST WINS USING A SUNBURST CHART

Sunburst charts are used to visualize hierarchical datasets. The hierarchy is shown through a series of concentric rings where each ring corresponds to a level in the hierarchy. Here, since we have a multi level hierarchy, a sunburst chart is ideal.

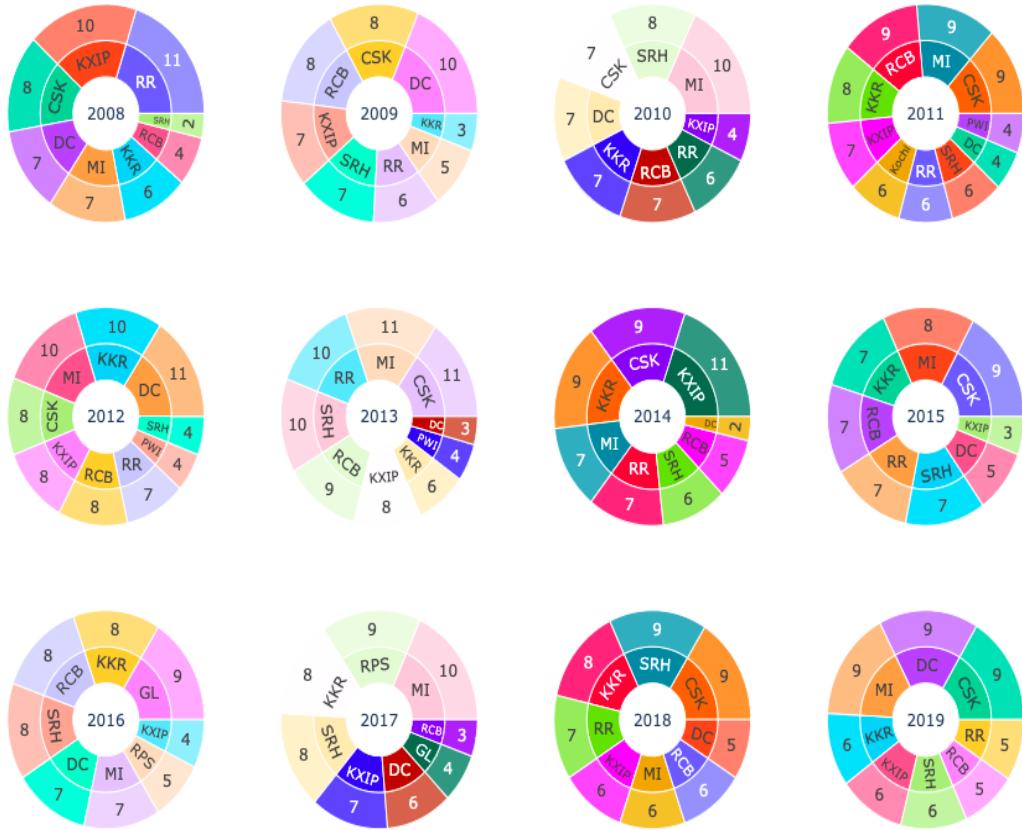


This visualization represents the number of wins by each team across all the seasons by using a sunburst chart. The colors represent each team.

### Key Insights:

1. Again, we see the number of times each team has won - with respect to each season.
2. We can understand the data by hovering over each team.
3. However, as there's so much information, it can be hard to gauge the information accurately.
4. One drawback is that as there's so much data in a single plot, it can be hard to visualize and understand well.

## TEAMS WITH THE MOST WINS USING MULTIPLE SUNBURST CHARTS AS SUBPLOTS

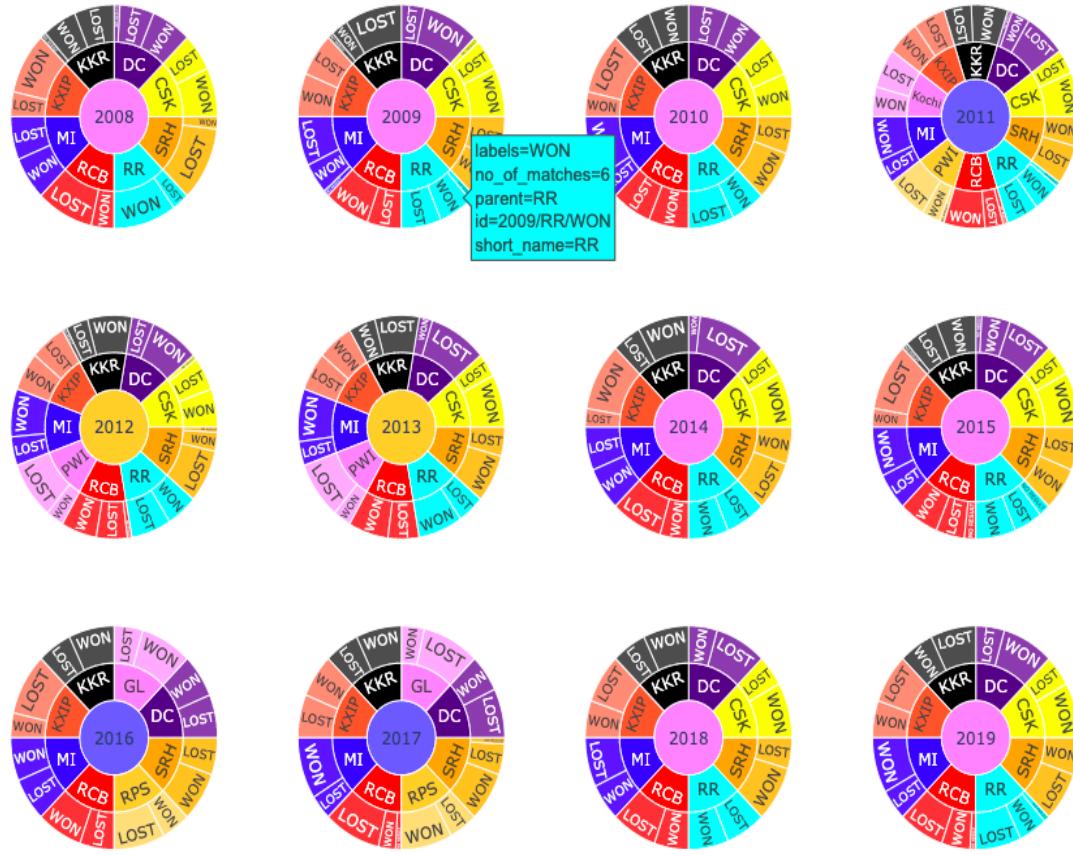


This visualization illustrates multiple subplots of the previous visualization so as to easily understand each team's performance across the seasons. Each subplot represents a particular season.

### Key Insights:

1. By using multiple subplots to represent each season, we can get a better understanding of each team's performance.
2. The numbers clearly represent the number of wins by each team.
3. However, the color is not constant for each team and the team positions are not constant.

**TEAMS WITH THE MOST WINS USING MULTIPLE SUNBURST CHARTS AS SUBPLOTS & MAKING THE COLOR CONSTANT**

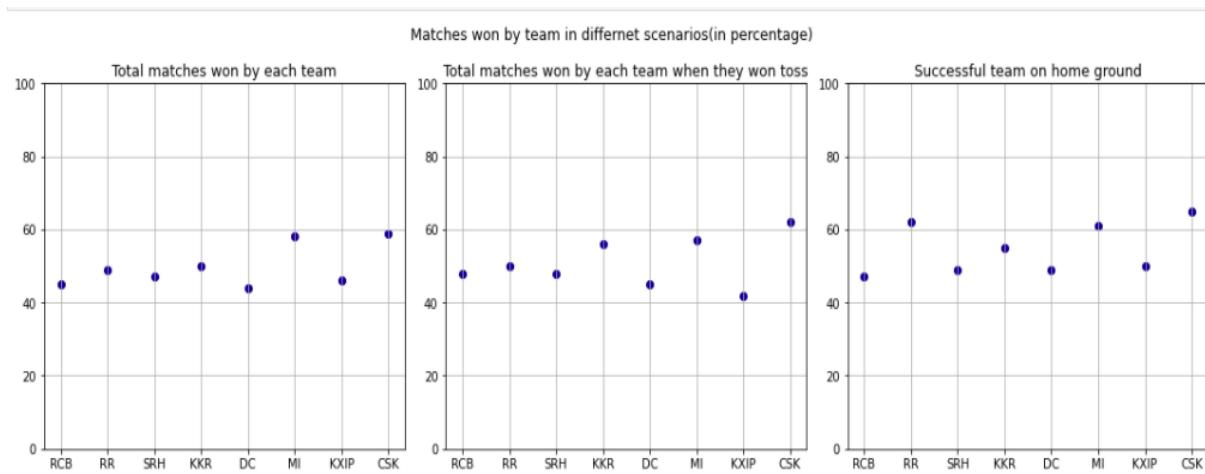


**Key Insights:**

1. This chart maintains the constant color for each team and in each position.
2. Another advantage is that each subplot shows wins/losses, both.

## HIGHEST PERCENTAGE OF WINS BY A TEAM

A scatterplot uses cartesian coordinates to display values for two variables of a dataset. Here, the x axis represents the different teams and the y axis represents the total percentage of matches won.



This visualization represents the percentage total of matches won with respect to three scenarios - the total percentage of individual matches won, the total percentage of matches won when a team loses the toss, and the total percentage of matches won by a team playing on their home ground.

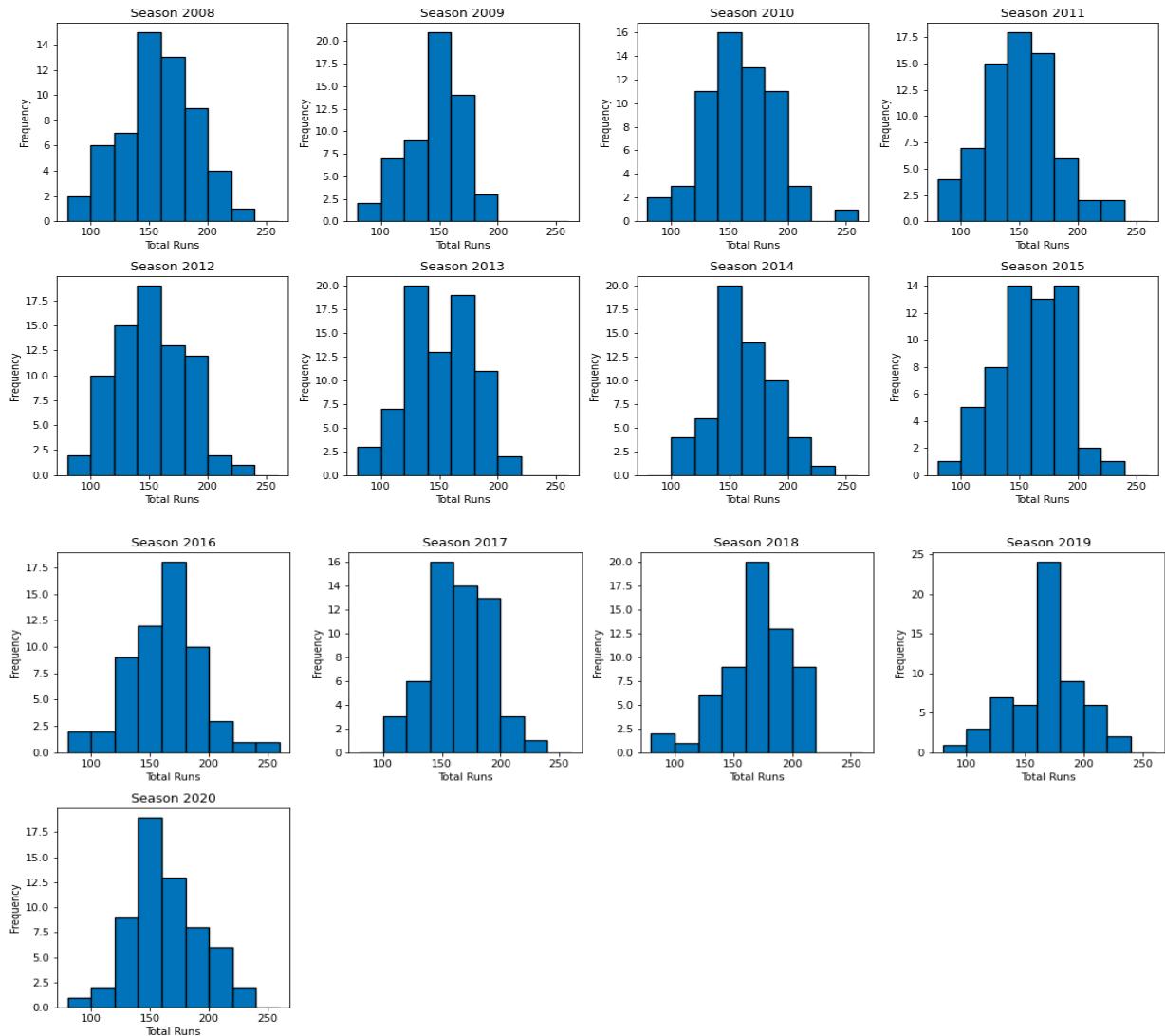
### Key Insights:

1. Once again, MI and CSK seem to be the highest, winning numerous matches.
2. With respect to winning the toss, we can see that, once again, MI and CSK have the higher percentage number of match wins after winning the toss.
3. Apart from those two teams, KKR also stands out by winning more matches by winning the toss.
4. KXIP seem to do the most poorly after winning the toss.
5. With respect to winning matches on each team's home ground, most teams are fairly consistent apart from three - RR, MI and CSK.

## RUNS SCORED BY ALL TEAMS COMBINED IN EACH SEASON

A histogram is a graphical representation that organizes a group of data points into specific ranges. It is similar to a bar graph in terms of condensing a data series into an easily interpreted

*visual. However, the bins in a histogram are grouped together as opposed to the spread out bins in a bar graph.*



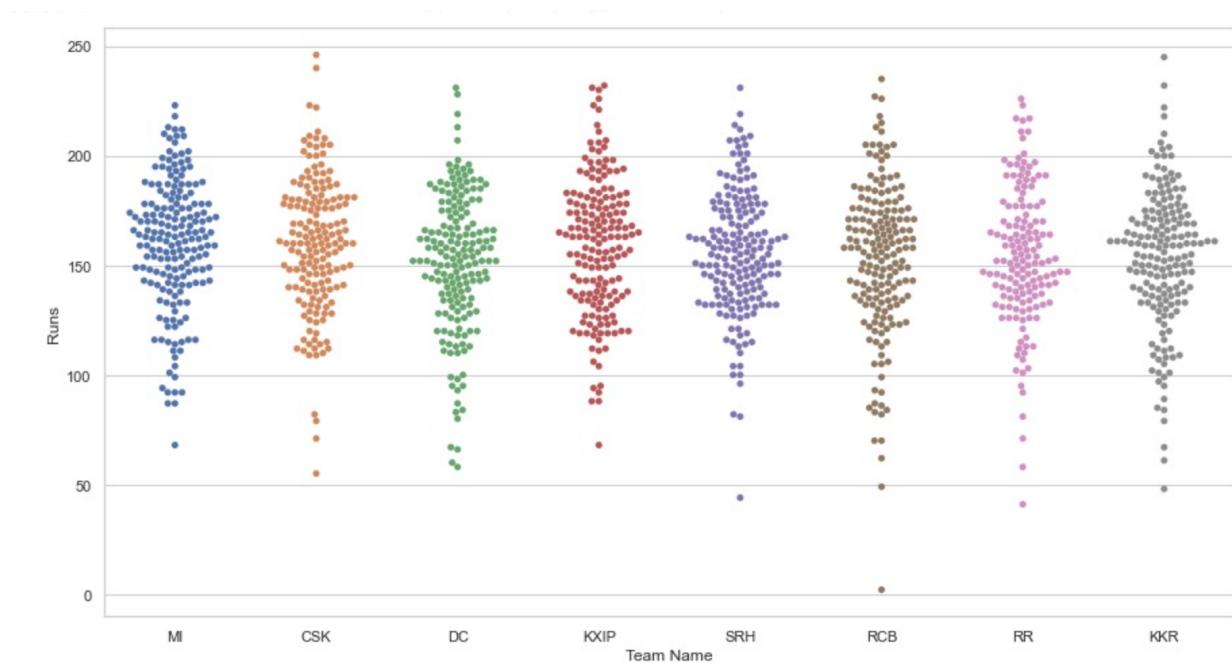
This visualization represents the gross total runs scored by all teams across various seasons in the form of histograms.

#### *Key Insights:*

1. Across most seasons, the number of runs scored by each team is in the range of 150-180 runs.
2. However, in a couple of seasons, the number of runs scored were considerably higher and lower.

## AVERAGE RUNS SCORED BY EACH TEAM IN ALL SEASONS USING A BEE SWARM CHART

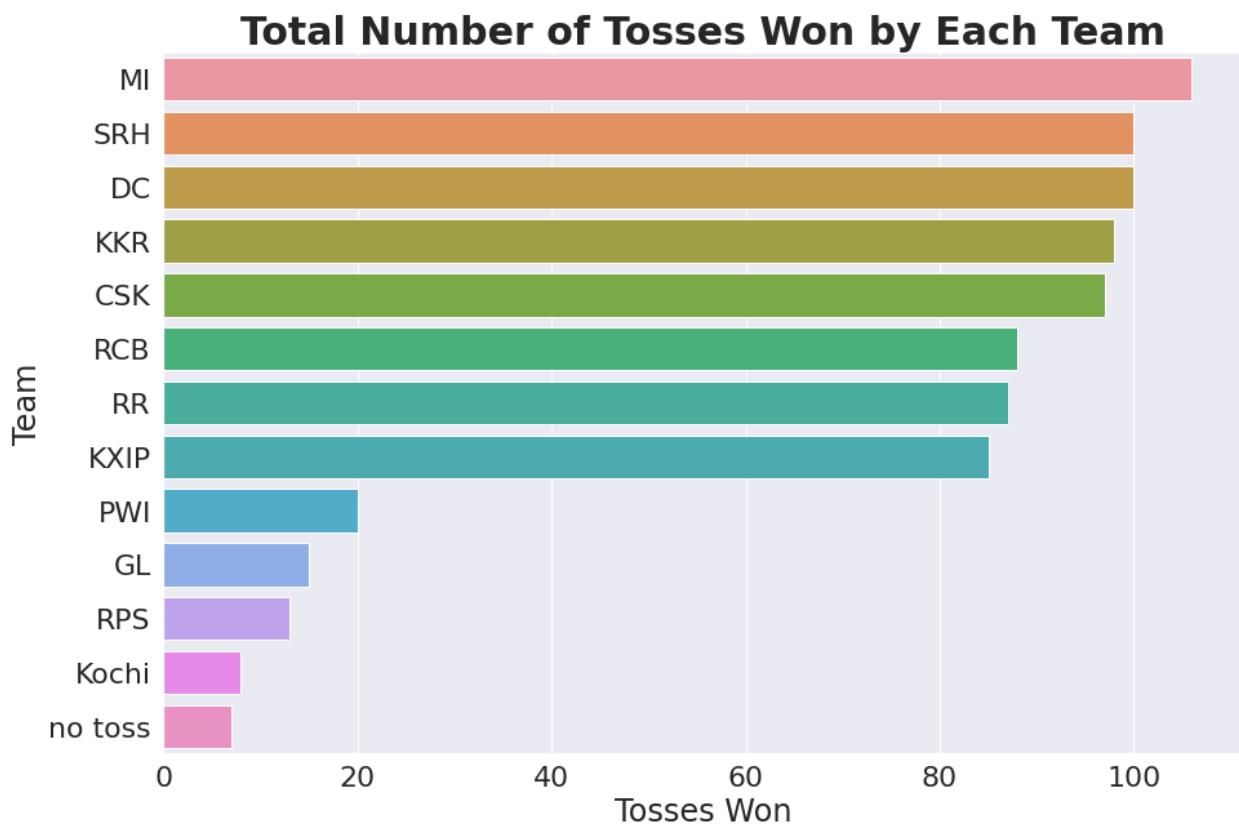
A bee swarm chart is similar to a stripchart but there aren't any overlapping points. This makes us understand each data point well. Here, we've done a bee swarm chart for each team's average runs scored throughout all the seasons.



### Key Insights:

1. We can see that all the teams have their runs per game to be in the range of 150-200 runs.
2. WE can also see that there are some extreme values as well. There are lesser values of 200 plus runs scored by all the teams.

**TOTAL NUMBER OF TOSSES WON BY EACH TEAM**

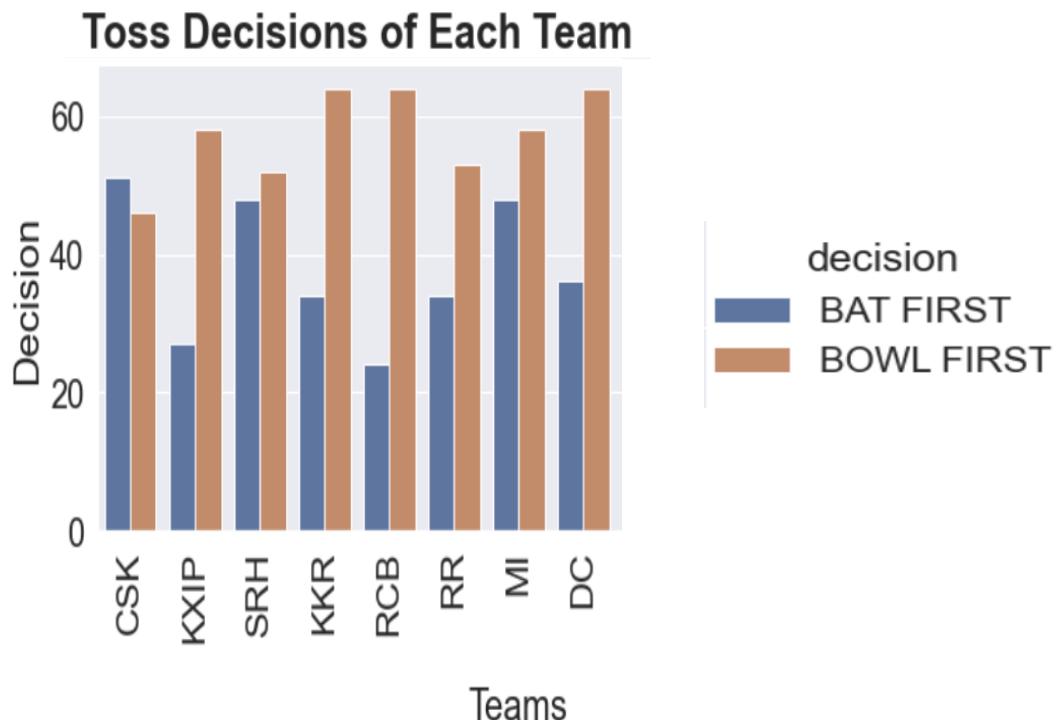


*Key Insights:*

1. We can see that MI have won the most number of tosses with 110 wins under their belt.
2. SRH and DC come in second with 100 toss wins.
3. The tosses are directly related with the total number of matches played by each team and that is why the outlier teams PWI, GL, RPS and Kochi haven't won many tosses

### **TOSS DECISIONS OF EACH TEAM AFTER WINNING THE TOSS**

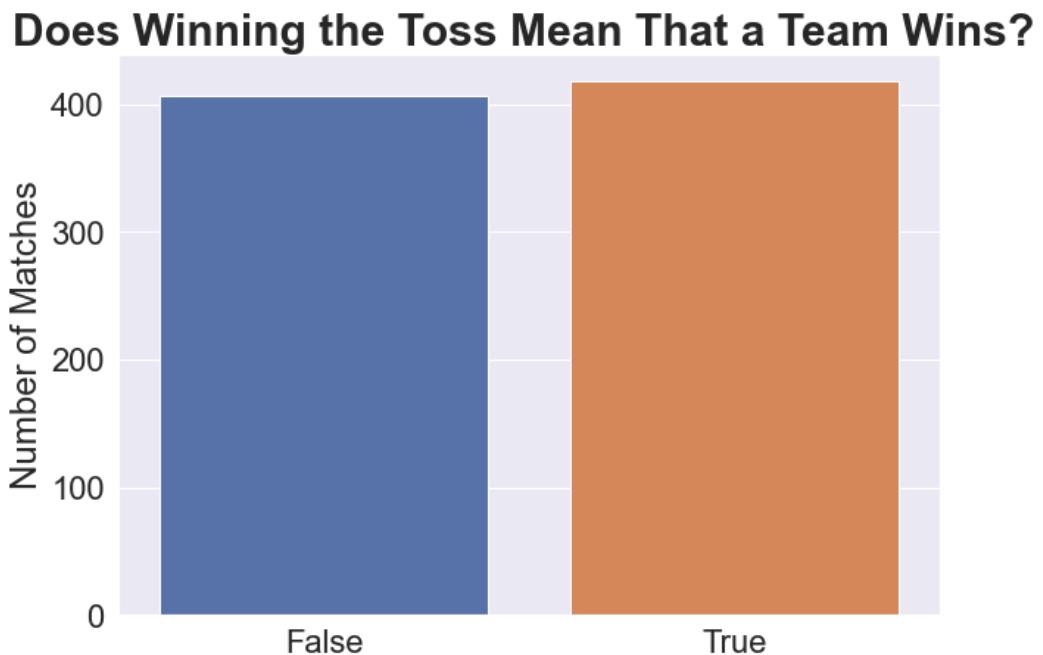
A Bar graph represents categorical data with rectangular bars/heights proportional to the values that they represent. Here, we've represented the bars as the number of times each team has decided to bat or bowl first after winning the toss.



#### *Key Insights:*

1. We see that most teams want to bowl first if they win the toss. This just goes to show that in T20 cricket such as the IPL, teams are more confident in chasing down scores rather than batting first.
2. However, CSK is the only team who have chosen to bat first after winning the toss as compared to bowling first.

**DOES WINNING THE TOSS MEAN THAT A TEAM WINS?**

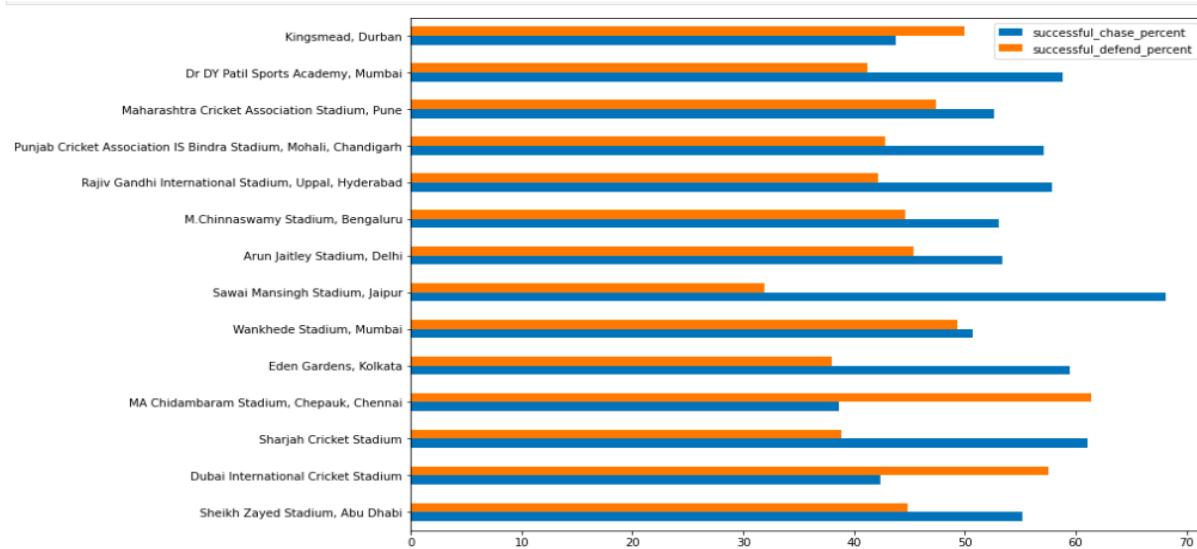


*Key Insights:*

1. As we can see in the above visualization, there is only a marginal difference in the number of wins when a team wins the toss.
2. From this, we can draw the conclusion that winning the toss doesn't cause much difference in the match outcome.

## SUCCESSFUL VENUES FOR DEFENDING AND CHASING RUNS

This barchart represents the successful chase and defend percentage in each cricket ground.

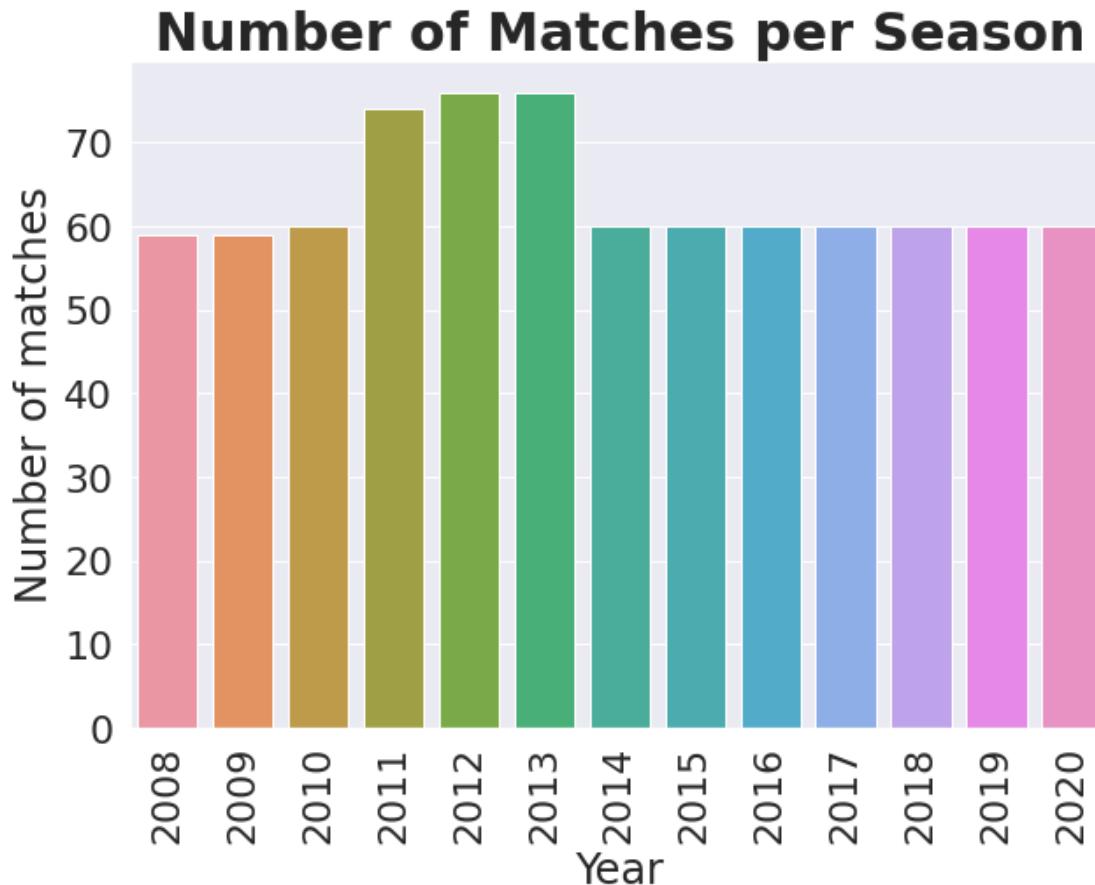


### Key Insights:

1. We see that the successful chase percent is far greater than the successful defend percent.
2. This correlates to most team's decisions to bowl first in the previous visualization.
3. We can also see that CSK's home ground in Chennai has a far higher successful defense percentage and that's why CSK was one of the few teams to bat first after winning the toss.
4. Apart from the Chennai stadium, the Dubai stadium and the Durban stadium are the only other two stadiums which have a higher defend percent as opposed to a chase percent.

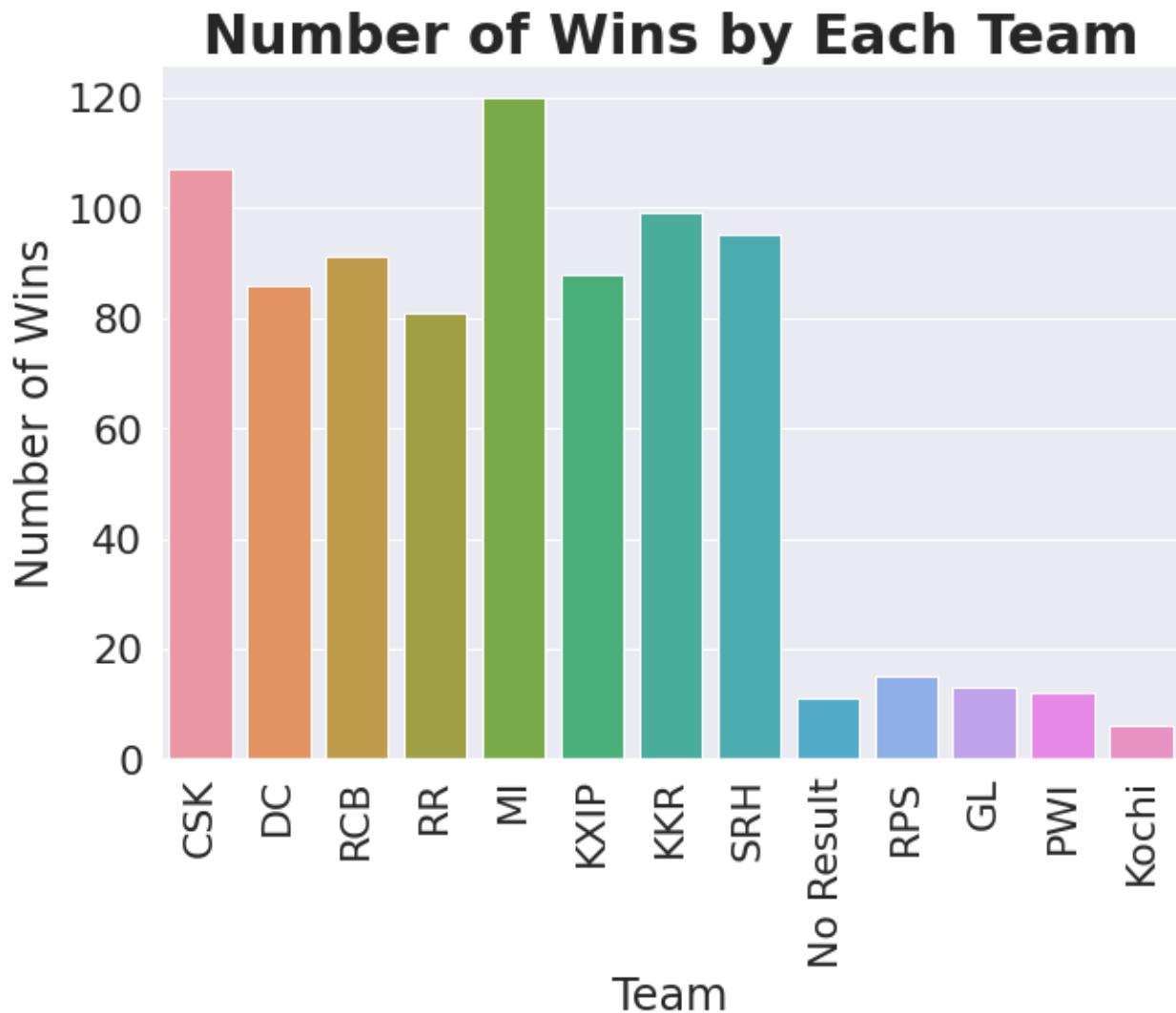
### **NUMBER OF MATCHES PLAYED PER SEASON**

*Key Insights:*



1. The most matches were played in 2011, 2012, and 2013. The rest of the years, the matches were consistently around 60.

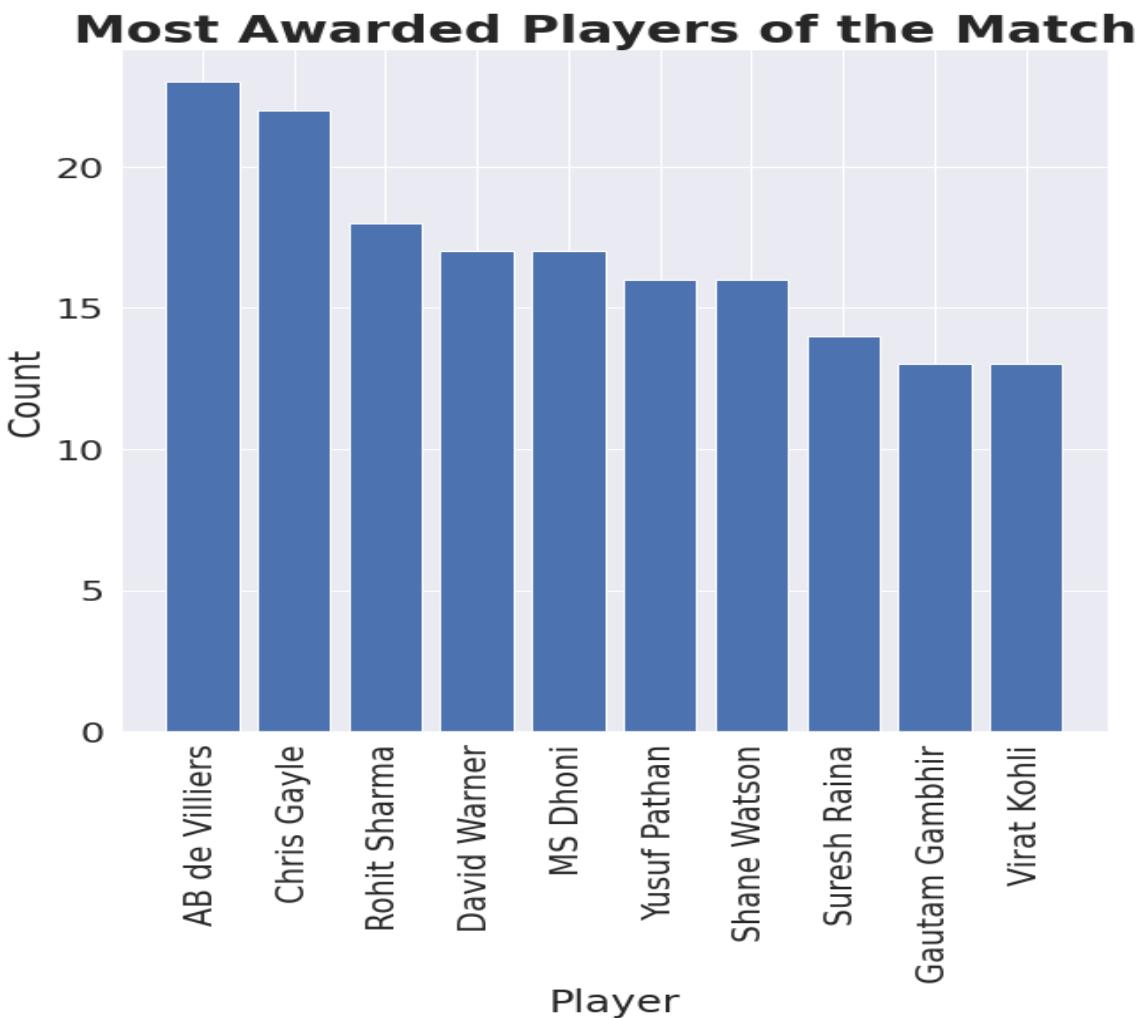
**TOTAL NUMBER OF MATCH WINS BY EACH TEAM**



*Key Insights:*

1. We can see that MI has the most number of wins as compared to any other team.
2. Following close behind is CSK, KKR and SRH.

**TOP 10 PLAYERS WHO HAVE BEEN AWARDED THE MOST 'PLAYER OF THE MATCH' AWARD ACROSS ALL SEASONS**

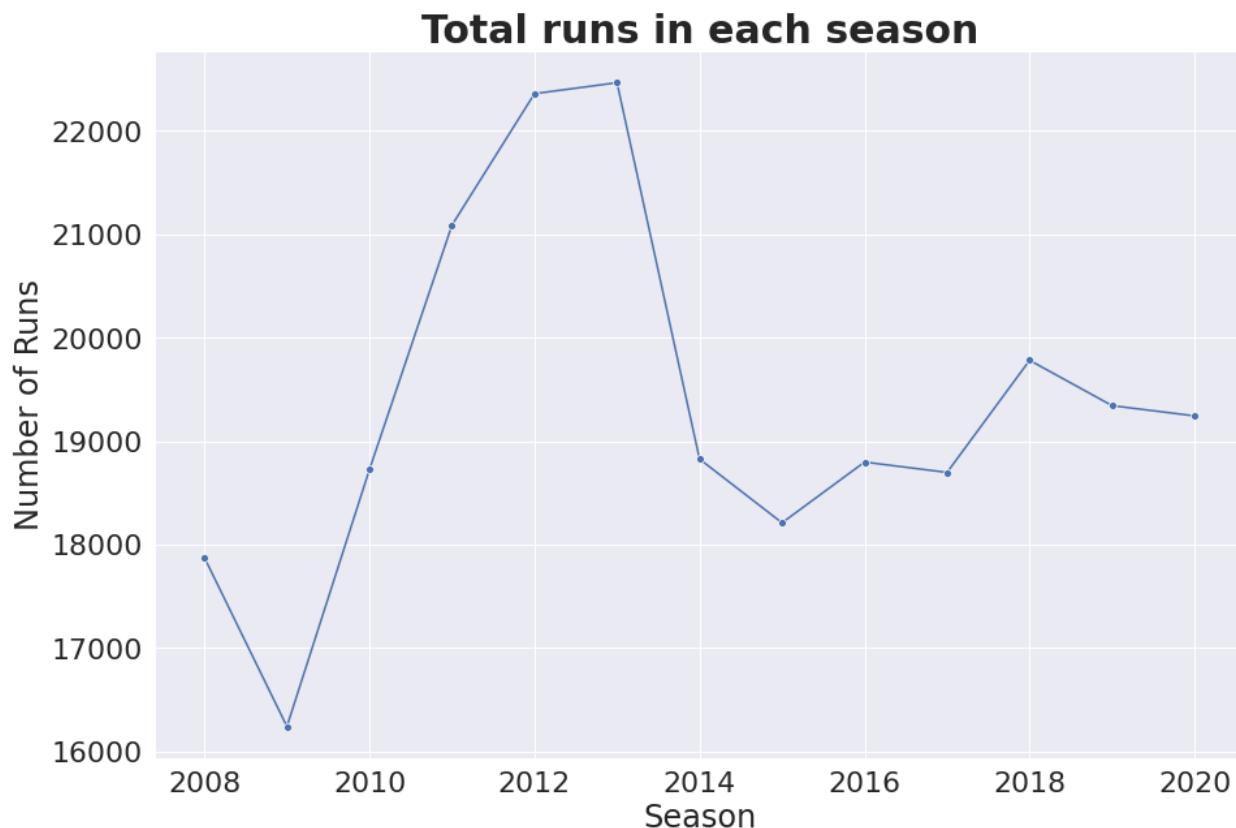


*Key Insights:*

1. We can see that AB de Villiers has been awarded the most 'player of the match' awards in the history of the IPL. Around 23 times.
2. Chris Gayle follows closely behind with being awarded it about 22 times.

### **TOTAL RUNS SCORED BY ALL TEAMS IN EACH SEASON**

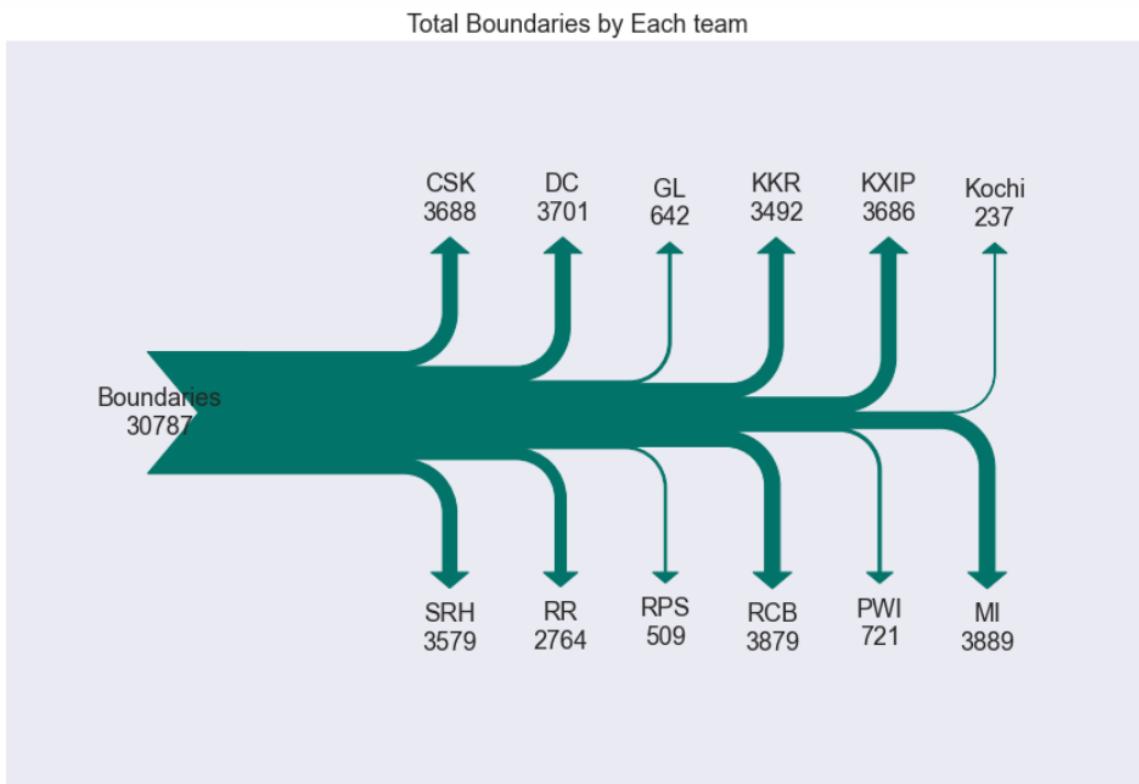
A line chart displays information as a series of data points called markers which are then connected by straight line segments.



#### *Key Insights:*

1. We can see that the total runs peaked in 2012 and 2013, where 22000 plus runs were scored.
2. From 2014 onwards, it dropped to 19000 runs and since then, it has been somewhat stable till now.

## TOTAL BOUNDARIES BY EACH TEAM



The visualization technique used here is called the “Sankey” plot. Sankey plot is a type of flow diagram in which the width of the arrows is proportional to the flow rate. In the plot above, the total flow is representing the total Boundaries scored by all teams in IPL history and the arrows represent the portion of total boundaries scored by the respective teams.

### Key Insights:

1. *Mumbai Indians* have scored the highest number of boundaries in IPL history among all the major teams.
2. Teams such as *Gujrat Lions*, *Kochi*, *Rising Supergiants Pune* and *Pune Warriors India* have scored very less number of boundaries. That is because they are not regular teams and this information can be inferred from the plot.

3. Among the regular teams, *Rajasthan Royals* have scored the least number of boundaries.

#### **Links :**

We can get the code and data from the github link <https://github.com/umesh-gattem/IPL-D-VIZ>

#### **Libraries used:**

Above visualizations are plotted using Python programming using different visualization modules like **Matplotlib, Pandas, Numpy, Plotly, Seaborn, Word cloud, Circlify, Geopy ,Folium**.

#### **Conclusion :**

We have worked on various statistics from the IPL dataset and have gathered a vast amount of information and actionable insights as to the why, how, etc of the data. The team statistics gave us a good insight on each team in the IPL, ranging from the best ones to the worst ones.

The player statistics gave us insights on different players who have done well for themselves and for the team. We can also figure out the correlation between good teams and good players and whether they're only playing for themselves or for their entire team.

The venue statistics gives us information about each venue and their impact on the performance of the players. Each cricket ground holds various challenges and truly tests each team's strengths and weaknesses. At the end, only the best team will brave through and emerge victorious at the end of each IPL journey.

The season statistics give us various insights with respect to each season and how each season changes - how the teams and players must change as well.

With all the data we've gathered and worked on, we wouldn't have been able to understand and interpret anything if not for creating the visualizations. By visualizing the data, we were able to get various insights and interpretations. We've learnt that by visualizing data, it tells us a story - a story that can never be found by just looking at the numbers.

## **Future Work:**

The same code can be used for later seasons of the IPL data also since code is written in such a way that we just need to provide new data and it will render the visualizations upto the date. So we are planning to maintain the project in the following years and work on any new insights or new problem statements and add more visualizations.

## **References :**

- [https://www.kaggle.com/rajsengo/indian-premier-league-ipl-all-seasons?select=all\\_seasons\\_on\\_details.csv](https://www.kaggle.com/rajsengo/indian-premier-league-ipl-all-seasons?select=all_seasons_on_details.csv)
- <https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/>
- <https://techtrunk.in/ipl-data-analysis/>
- <https://www.kaggle.com/rishpande/indian-premier-league-ipl-data-visualization>
- <https://www.sciencedirect.com/science/article/pii/S1877050917327023>
- <https://livecodestream.dev/post/how-to-plot-your-data-on-maps-using-python-and-folium/>
- <https://www.geeksforgeeks.org/how-to-get-geolocation-in-python/>
- <https://colab.research.google.com/drive/1MnZepW9RLJjXKnbgxtlI0zcGAmwO4R9J?usp=sharing#scrollTo=KQcZdM3kQvyY>
- <https://plotly.com/python/sunburst-charts/>
- <https://github.com/holtzy/The-Python-Graph-Gallery/blob/master/src/notebooks/circular-packing-1-level-hierarchy.ipynb>