

# **Indian Premier League(IPL) VISUALIZATION**

## ***Project Proposal***

**Umesh Kumar Gatterm** - MS-DS Graduate Student from Luddy School, Indiana University, Bloomington (email: ugatterm@iu.edu)

**Muhammad Murtaza Mir** - MS-DS Graduate Student from Luddy School, Indiana University, Bloomington (email: mmir@iu.edu)

**Raunav Sharma** - Undergraduate Junior from Luddy School, Indiana University, Bloomington(email: rashar@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.

2021 is witnessing the 14<sup>th</sup> edition of the IPL league, the previous 13 editions having been runaway successes. 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.

The main reason 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	City	Home ground	Debut	Owner	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	2008	<ul style="list-style-type: none"><li>Grandhi Mallikarjuna Rao</li><li>Sajjan Jindal</li></ul>	Rishabh Pant	Ricky Ponting
Kolkata Knight Riders	Kolkata, West Bengal	Eden Gardens	2008	<ul style="list-style-type: none"><li>Shah Rukh Khan</li><li>Jay Mehta</li></ul>	Eoin Morgan	Brendon McCullum
Mumbai Indians	Mumbai, Maharashtra	Wankhede Stadium	2008	Mukesh Ambani	Rohit Sharma	Mahela Jayawardene
Punjab Kings	Mohali, Punjab	PCA Stadium, Mohali	2008	<ul style="list-style-type: none"><li>Mohit Burman</li><li>Ness Wadia</li><li>Preity Zinta</li><li>Karan Paul</li></ul>	KL Rahul	Anil Kumble
Rajasthan Royals	Jaipur, Rajasthan	Sawai Mansingh Stadium	2008	<ul style="list-style-type: none"><li>Manoj Badale</li><li>Gerry Cardinale</li><li>Lachlan Murdoch</li></ul>	Sanju Samson	Trevor Penney
Royal Challengers Bangalore	Bengaluru, Karnataka	M. Chinnaswamy Stadium	2008	Vijay Mallya	Virat Kohli	Simon Katich
Sunrisers Hyderabad	Hyderabad, Telangana	Rajiv Gandhi International Cricket Stadium	2013	Kalanithi Maran	Kane Williamson	Trevor Bayliss

## **Awards**

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

Few of the awards are

### **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.

### **Dataset:**

IPL Dataset contains two csv files, one containing information about all the matches played from 2008-2020 and second one having the Ball by Ball data of all the matches in all leagues played from 2008-2020.

### **IPL Matches Dataset:**

IPL Matches Dataset contains the information about all the matches played from 2008 - 2020 which includes nearly 20 columns. A few columns are described below with their data types.

- **Match id(string)** - defines the unique id for each match
- **Season(categorical)** - defines the season of the league
- **Date (Date)**- Date of the match
- **Toss decision(Categorical)** - Result of the toss
- **Place/City (String)**- Geo location of the place
- **Venue(String)** - Stadium name
- **Team batting first(Categorical)** - Name of the Team batting first
- **Team chasing(Categorical)** - Name of the team chasing
- **Result(Strings)** - Match Result. It can be contained in multiple columns which tells about the result of the match and by how many runs does the team batting first win or by how many wickets the team won the match.

The following is the sample dataset of 5 rows.

1	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_margin	el
2	335982	Bangalore	2008-04-18	BB McCullum	M Chinnaswamy Stadium		Royal Challengers Bangalore	Kolkata Knight Riders	Royal Challengers Bangalore	field	Kolkata Knight Riders	runs	140	N
3	335983	Chandigarh	2008-04-19	MEK Hussey	Punjab Cricket Association Stadium, Mohali		Kings XI Punjab	Chennai Super Kings	Chennai Super Kings	bat	Chennai Super Kings	runs	33	N
4	335984	Delhi	2008-04-19	MF Maharoof	Feroz Shah Kotla		Delhi Daredevils	Rajasthan Royals	Rajasthan Royals	bat	Delhi Daredevils	wickets	9	N
5	335985	Mumbai	2008-04-20	MV Boucher	Wankhede Stadium		Mumbai Indians	Royal Challengers Bangalore	Mumbai Indians	bat	Royal Challengers Bangalore	wickets	5	N
6	335986	Kolkata	2008-04-20	DJ Hussey	Eden Gardens		Kolkata Knight Riders	Deccan Chargers	Deccan Chargers	bat	Kolkata Knight Riders	wickets	5	N

## IPL Ball by Ball Dataset :

IPL Ball by Ball Dataset contains the information of all the balls bowled in all the matches from in all leagues from 2008-2020. This dataset contains nearly 45 columns. The few of them are described below with their data types.

- **Ball id (String):** Unique Id of the ball.
- **Innings(Categorical):** Whether first innings or second innings
- **Over No(Categorical) :** The over of the innings
- **Ball No (Categorical):** Ball of the over
- **Batsman playing (String):** Batsman who is striking the ball
- **Non Striker (String):** Batsman who is non-striker
- **Bowler(String):** Player who is bowling that over/ball
- **Runs Scores (Int):** Runs scored on that ball
- **Extra Runs (Int):** Extra runs given on that ball.
- **Boundary (Boolean):** Whether that went to boundary or not on that ball
- **Is Wicket(Boolean) :** Whether wicket is taken or not on that ball

The following is the sample dataset of 5 rows.

1	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs	non_boundary	is_wicket	dismissal_kind	pl
2	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1	0	0	NA	N
3	335982	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	1	0	0	NA	N
4	335982	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	0	0	0	NA	N
5	335982	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	1	0	0	NA	N
6	335982	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	1	0	0	NA	N

This is the information of the complete IPL dataset.

## Why this Dataset??

We have chosen this dataset because we have ample information on the various stats of all matches and all the players in any specified season. It also provides the ball by ball statistic which gives us information about each player. Potentially, by using this dataset we can build the complete information of each batsman playing against all other bowlers across various teams. Similarly, we can obtain ample information from the bowler's perspective - where we get information of each bowler bowling to other players across all other teams. The IPL is the most-attended cricket league in the world and in 2014, it was ranked sixth by average attendance among all sports leagues. Therefore, by working on this dataset, it gives us a vast learning experience and also teaches us about the various methods and technologies of different data visualizations.

## Objective

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.

## Expected Outcomes

We expect to gain "actionable" insights from the visualizations. Since the data set has records of all matches - both at a macro level as well as at a granular level, we can leverage that data to understand the performance of the players through various perspectives such as *'performance at home ground'*, *'performance away from home ground'*, *'best performance at a non-home ground'*, *'strike rates of batsmen'*, *'averages of batsmen'*, *'strike rates of bowlers'*, *'averages of bowlers'*, *'performance of a batsman with respect to specific bowlers'*, *'performance of a bowler with respect to particular batsmen'* and potentially many more. All this information helps in evaluating the true skills of the players. This analysis, as mentioned above, is also extremely important from the owners' perspectives as well. This is because the specified information can help the owners to make the right selection and purchase the right players that suit their teams' interests the best. Cricket fans around the world can also use this dashboard to analyse and determine the performance of their favourite teams or players.

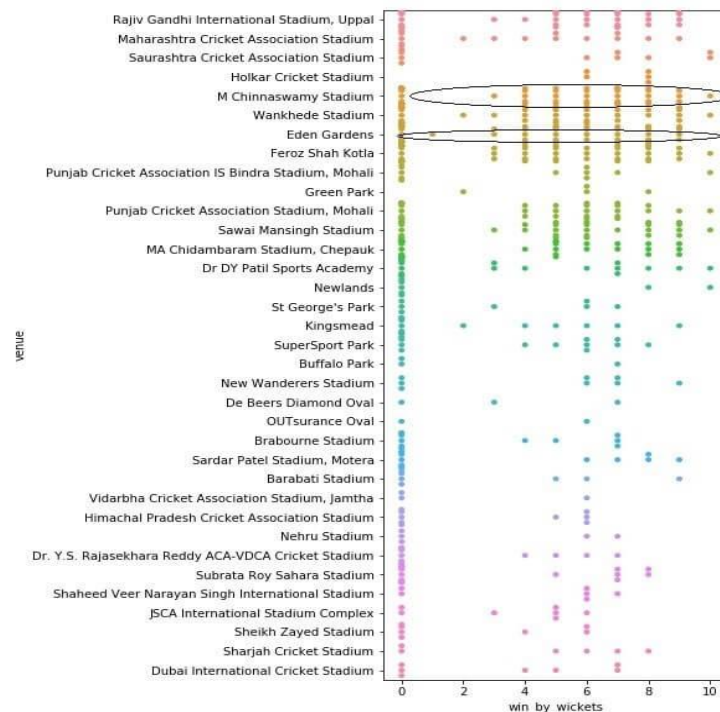
We are also trying to visualize the various awards that we mentioned earlier. For example - the number of players with the orange cap, most valuable player of each season, etc.

For the above scenarios we are trying to plot different visualizations that we have been taught in the course like **scatterplots**, **bar graphs**, **boxplots**, **histograms**, **PCA**, **MDS etc.**, and also getting to know about these applications using various python libraries such as **matplotlib**, **seaborn**, **bokeh** and **datashader**.

## Intent of the project:

Apart from the motivation that we have discussed above, our main intent 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.

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

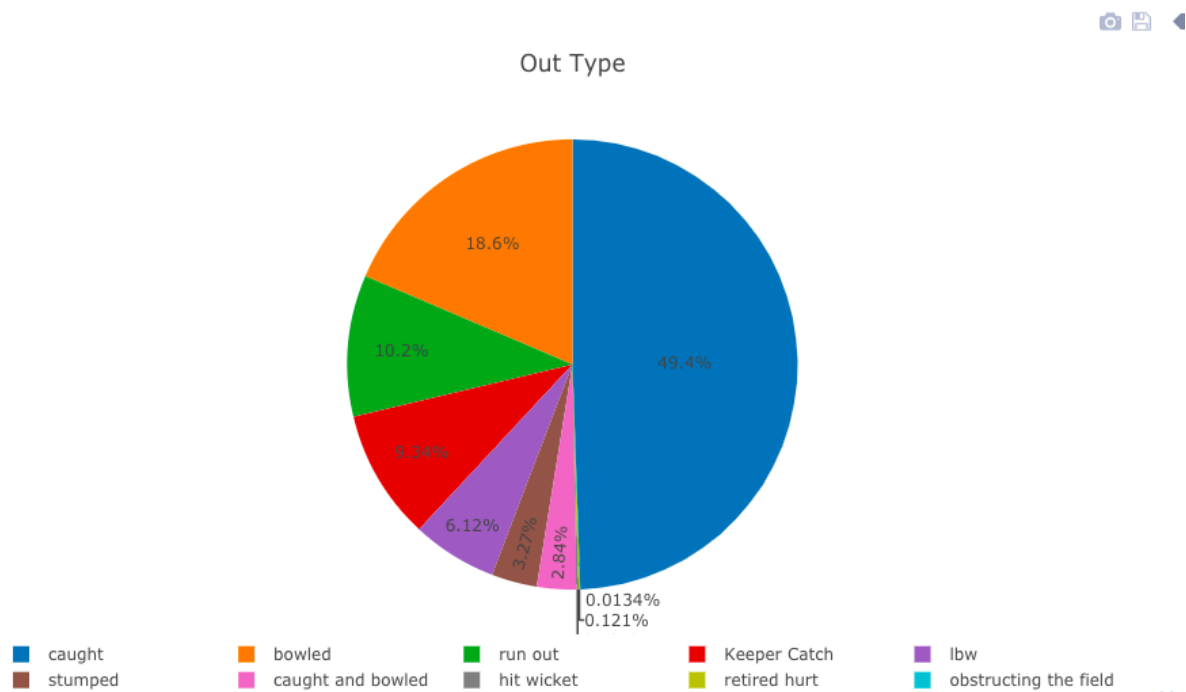




The plot wants to convey the wins of the teams while chasing the score (batting second) put up by the other team on the scorecard (batting first) at different venues with the intent to identify the best venue for chase and also the number of wickets by which the team chasing the score wins by. Anyone who knows about the basics of cricket knows 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 in this context according to our judgement.

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.



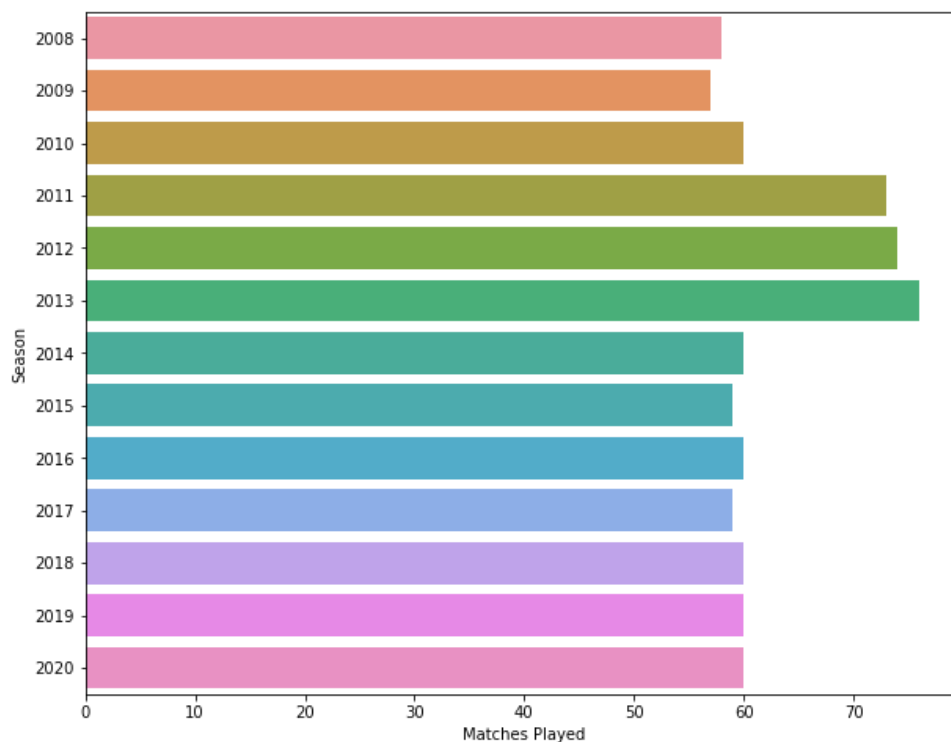
There are few issues with this plot. First thing is that it is a pie chart. We know that Pie charts are one of the most overused graphs and in most cases are not the best way to present data especially when we are trying to categorize more variables. They often distort the information and make it more difficult for decision-makers to understand the messages they contain. Second thing is that it is using the Red and Green 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.

### Previous Relevant work:

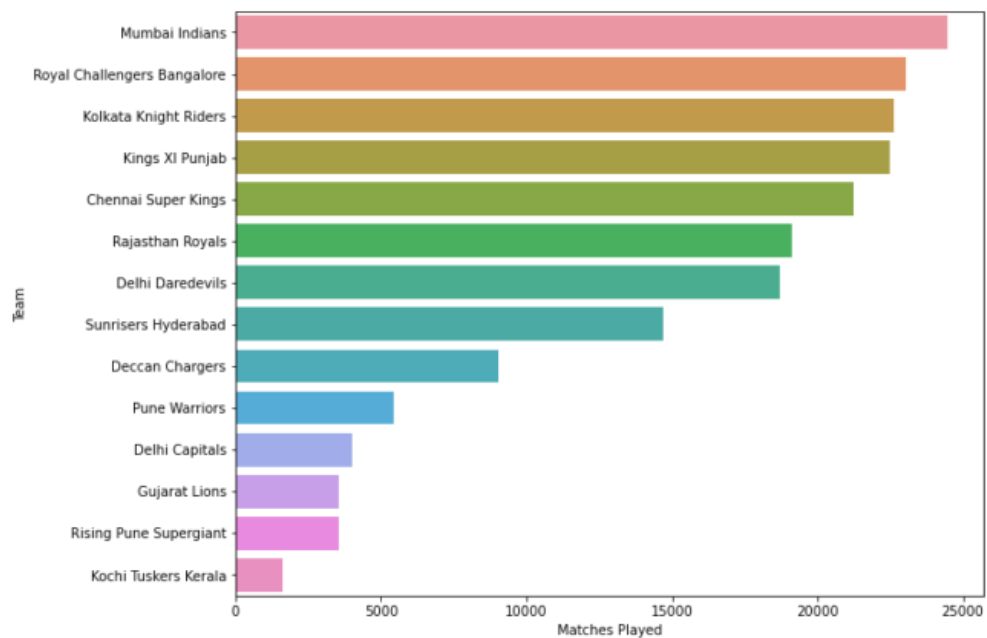
Cricket visual analysis is becoming a very big area of interest especially in India specifically because of the sheer amount of popularity and the capital involved. The Indian Cricket Board known as the **Board of Control for Cricket in India** abbreviated as **BCCI** is the richest cricket board in 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*



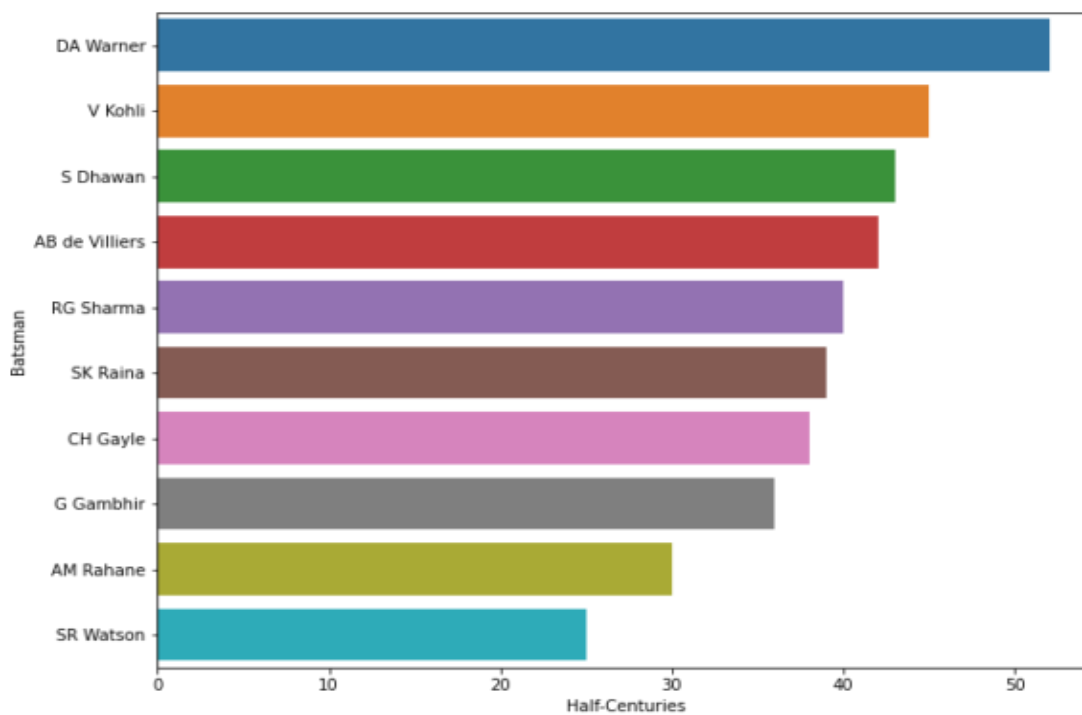
*This plot is visualizing the number of total matches played each season.*

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



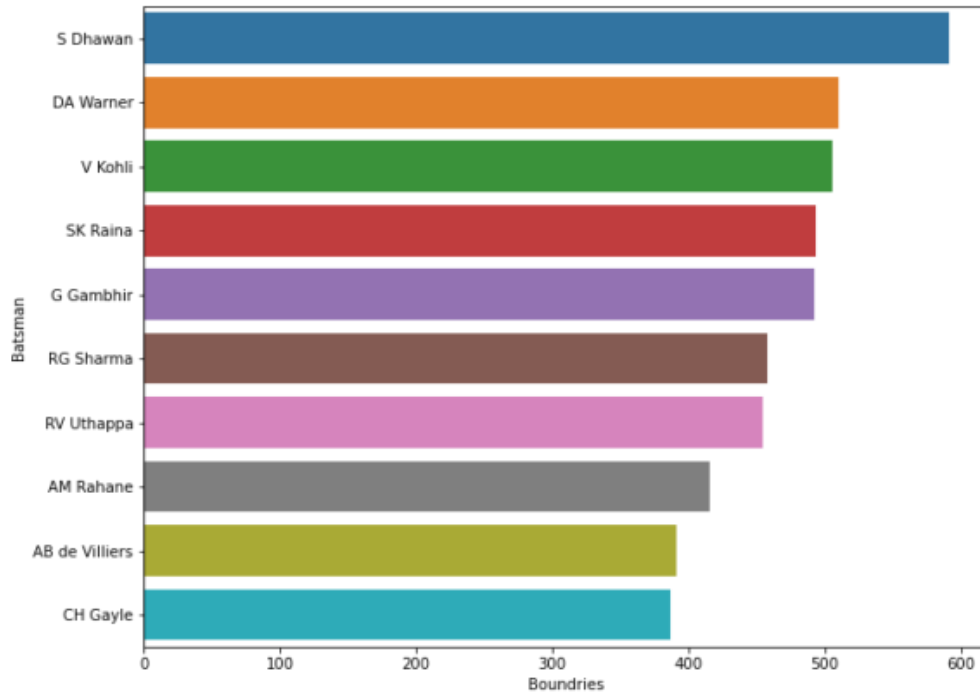
*This plot is visualizing the total matches played by each team combining all the seasons.*

### ***MOST IPL FIFTY BY A PLAYER***



*This plot is visualizing the highest number of at least 50 scores by batsmen.*

### ***MOST BOUNDARY (4s) HIT BY A BATSMAN***



This plot is visualizing the highest total number of 4s scored by batsmen.

As seen above, most of the visualizations are in the form of bar graphs. We are trying to visualize them in a different way using boxplots, scatterplots, histograms, etc.

### **Conclusion**

All in all, based on the above proposal, we would like to visualize the current IPL Dataset using various data visualization techniques while understanding the applications of those visualizations.

### **References**

- [https://www.kaggle.com/rajsengo/indian-premier-league-ipl-all-seasons?select=all\\_season\\_details.csv](https://www.kaggle.com/rajsengo/indian-premier-league-ipl-all-seasons?select=all_season_details.csv)
- <https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020>
- <https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/>

- <https://community.powerbi.com/t5/Data-Stories-Gallery/IPL-Data-Analysis-2008-2016/m-p/141219>
- <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>