**Point analysis with boxplot– Indian premier league (IPL)**

*Arun Sekar*

[*arunsworld2200@gmail.com*](http://arunsworld2200@gmail.com)

**Introduction:**

Sport of Cricket is not significantly famous across globe, but some Cricket loving nations have a huge fan base. Formats for this game are evolving since 19th century with the most recent being the 10 over per team game called T10. At the international level there are 12 nations allowed to play by the International Cricket Council (ICC) in the longest format of the game called Test cricket. The other famous formats are One Day international (ODIs) and T20 internationals. Teams play 50 overs each in ODIs and 20 overs each in T20s. In addition to that, there are bunch of rules which are common across formats and some which are specific to formats. These are governed by ICC and MCC (Marylebone Cricket Club) based on which the games are played across continents.

In additional to the international games which are played between Nations, there are some major leagues (typically in T20 format) which are played between teams which are owned by corporate entities or rich individuals across countries. These leagues are organized by the cricket boards of different countries. Some of such big names are as below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cricket leagues** | **Organizer** | **Country** | **Total teams** |
| Indian Premier League (IPL) | Board of Control for Cricket in India (BCCI) | India | 8 |
| Big Bash League (BBL) | Cricket Australia (CA) | Australia | 8 |
| Pakistan Super League (PSL) | Pakistan Cricket Board (PCB) | Pakistan | 6 |
| Caribbean Premier League (CPL) | Cricket West Indies (CWI) | Caribbean Islands nations (West Indies) | 6 |
| T20 Blast | English and Wales Cricket Board (ECB) | England & Wales | 18 |
| Bangladesh Super League (BPL) | Bangladesh Cricket Board (BCB) | Bangladesh | 7 |

***Source:*** *Cricket league websites*

These leagues are played for huge prize money. Players across countries are auctioned to the team management where they can bid for the players of their choice. IPL is currently the richest league of all with huge prize money (winner gets Rs. 10 crore which is equivalent to $1.35 mn as on the date of this write up) and team managements have huge purchasing power. The most a player has been purchased ever is for $2.4 mn (Gautam Gambhir in 2011). The league is being played since 2008 with 8 teams competing in the 2020 edition. The league is a commercial success for corporate entities, the owners and all the other associated entities.

In this publication, you will find the distribution of the points secured by each team ever since 2008 and the significance of the tournament in 2020 along with other distinguishing factors across years.

**More background on IPL:**

This is the longest cricket league of strong repute conducted in round-robin method. In its current format (with latest being 2020), each team gets to compete with the other team twice. If 8 teams are competing, then each team gets to play 14 games in the league stage. In total 56 matches (8P2) are played in the league stage among all the teams. The tournament was initially designed to consist 8 teams (each named and associated with different major cities in India). But two were added to the league (one from Kochi and one from Pune) in the 4th edition of IPL (2011). So, 10 teams competed in the edition. A total of (In the next edition (2012), the team from Kochi (**Kochi Tuskers Kerela**) pulled off from the league. The 5th (2012) and 6th (2013) edition of the IPL was played with 9 teams with just the team from Pune (**Pune Warrior Indians** – PWI) added to the initial list of teams from 2008-2010.

The winner of each game gets 2 points while the looser gets nothing. If a match is called off or abandoned due to geographical events (like rain) or due to any other unplayable conditions, then each team gets a point. If a game is tied (in the scenario where both the teams are left at the same score at the end of the 20th over or if the team playing second lost all their wickets for the same score as that of the first team) then the winner is decided by “**Super Over**”. Until 2019 (IPL 12), each team will have to play 1 over (with 2 wickets in hand) and the team which scores the most in with their respective resources wins. If the teams are tied, the team which had scored more boundaries in their initial quota of 20 overs will be declared winner.

But due to the incidents and the controversies around the 2019 Cricket world cup final played between England and New Zealand, the format for Super over was changed in 2020. As per the new rule, if the teams are tied in the super over, then another iteration of super over is played between the teams. The iteration is repeated until a clear winner (with no tie) arises. Ironically, the winner of the match 36 of IPL 2020 (on 18th October 2020) between Kings XI Punjab (KXIP) and Mumbai Indians (MI) was decided with two super overs being played. This was the first such game at a globally recognized level to be decided with more than 1 super over.

Top 4 teams at the end of the league stage will qualify for the play offs. The teams to play the final will be decided based on **page playoff system** where the top 2 teams play in **Qualifier 1** while the teams in 3rd and the 4th positions will play **Eliminator**. The winner of the qualifier 1 will directly qualify for the final. The looser of the Qualifier 1 will play with the winner of the Eliminator. This match is called the **Qualifier 2**. Looser of the Eliminator is out of the league. The winner of Qualifier 2 will play against the winner of the Qualifier 1 in the Final game of the tournament.

**Distribution of team points - Objective:**

The objective of this document is to present the distribution of total team points (for all teams) across all the IPL seasons (from 2008 – 2020) and to draw inferences on the same. This is expected to compare the level of competence across teams for each season and to identify the most closely fought season. The csv file “**Points summary\_IPL\_2008-2020.csv**” has the data set for the team names, team rank, the points scored by each team and the year of the event. Dataset is uploaded in the [Github link](https://github.com/arunscode-data/Public-Repo/blob/master/IPL-Point-analysis_boxplot/Dataset/Points%20summary_IPL_2008-2020.csv).

**Approach:**

To distribute the points across different years, the attributes in the dataset are presented on a boxplot.

1. Boxplot is plotted between the total team points and the year
2. Total number of Teams and their ranks will vary based on the participation of different teams in different IPL seasons
3. Boxplot is plotted using below tools/software
   1. Tableau Public

* Year in Columns
* Points in Rows
* Team and Rank in Details
  1. R program
     + Year in X-axis
     + Points in Y-axis
     + Whiskers are extended to all the outliers

1. Script for the R program is available on the [Github link](https://github.com/arunscode-data/Public-Repo/blob/master/IPL-Point-analysis_boxplot/IPLPointDistribution_Boxplot.R)
2. Boxplot whiskers are extended up to the Outliers even if they are beyond 1.5 x IQR

**Observation:**

The difference between the two boxplot whiskers for each year is the indicator for the degree of closeness with which a particular IPL series was contested. It directly measures the difference in the points scored between the top ranked team and the bottom ranked team.

* Amongst all the years, it was in 2008 (the first IPL season) is when the points are distributed the farthest (the difference in the points scored by the first ranked team and the points scored by the last ranked team is maximum)

In 2008,

* + the first ranked team (Rajasthan Royals) scored 22 points
  + the last ranked team (Sunrisers Hyderabad, formerly Deccan Chargers) scored 4 points
  + difference in the points for both the teams is 18
* IPL season for 2020 is the instance when the point distribution is the narrowest. The difference in the points scored by the first ranked team and the points scored by the last ranked team is the least.

In 2020,

* + the first ranked team (Mumbai Indians) scored 18 points
  + the last ranked team (Rajasthan Royals) scored 12 points
  + difference in the points for both the teams is 6

Chart, box and whisker chart

Description automatically generated

*Tableau image of the boxplot*

**Inference:**

* From the above observation, we can clearly state that the IPL season for 2008 is the least closely fought season. The better teams won many games while the bottom most team (in the points table) won the fewest (just 2 games). Ironically, this (4 points scored by Deccan Chargers) is the joint lowest scored by Delhi Daredevils (lately called Delhi Capitals) in 2014 ever in an IPL season.
* 2008 being the first season, teams were fresh, and the knowledge of a franchise based T20 games were less mature. Teams went by the big names amongst international teams and the team balance was challenging. This could be the potential reasons for the scenario.
* 2020 is the most closely fought season where the difference between the top ranked team and the bottom ranked team is the least (6). Teams from rank 2 to rank 5 scored 14 points and teams from rank 6 to 8 scored 12 points. Every game ended up so crucial towards the end.
* Teams were more competitive and games several games were too close. The 36th league game of the season where Mumbai Indians and Kings XI Punjab played in Dubai on 18th October 2020 had to be decided with two super overs. This is the first such incident across the recognized cricket played at all levels anywhere.
* On the same day, the 35th game of the season contested between Sunrisers Hyderabad and Kolkota Knight Riders was also decided with a super over.
* This significance can be credited to the teams possessing better understanding of the game and the conditions they play on. The team composition was rightly chosen to have a better balance to ensure they stay in the contest for the playoff.
* IPL 2020 (the 13th edition) had several such closely fought contests in 2020 which were held in United Arab Emirates (UAE) across Dubai, Abu Dhabi and Sharjah. The event was shifted to UAE due to challenges in conducting it in India owing to COVID-19 pandemic.

**Conclusion:**

Boxplot can be used as an effective measure to quantify the difference between the outliers. In the IPL point distribution boxplot for 2008-2020, the year 2020 is the most closely contested season, while the year 2008 is the least closely contested season.

**Reference:**

**Dataset source:**

* [www.cricinfo.com](http://www.cricinfo.com)
* [www.cricbuzz.com](http://www.cricbuzz.com)

**R programming for boxplot:**

* [www.r-bloggers.com](http://www.r-bloggers.com)