INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1780)
Available online at: https://www.ijariit.com

Insights of IPL: 2008 to 2020 and why it is interesting

Rakshit Patel
rakshit17p@gmail.com
G H Patel College of Engineering and Technology, Anand, Gujarat

Mihir Brahmbhatt
mihirbrahmbhatt23@gmail.com
Navrachana University, Vadodara, Gujarat

ABSTRACT

Cricket runs through every Indian's blood; without Cricket one cannot imagine India. Cricket is also the most famous outdoor sport in India as there is a huge craze of it in every age group of people. Also, India has been very much the pioneer of the modern day cricket that is the T20, it can be said that because of the Indian Premier League the advent of T20 started in the world. In the modern era with such high technologies available in every sport, one of the most important things is statistical data. This research paper mainly focuses on the results of the matches taking various factors into consideration such as the toss analysis, season wise number of matches, team performances and also the venues where the matches were played.

Keywords: *Statistical Data, Factors, Analysis.*

1. INTRODUCTION

Indian Premier League is one of the most celebrated events in the history of Cricket. Fast paced games and glamour are the two of the standout features of IPL, a lot of money is involved in this event. We thought of using the data set of IPL from 2008 to 2020 and predicting the end results of the matches taking various factors into consideration such as the toss analysis, season wise number of matches, team performances and also the venues where the matches were played.

We have taken 786 matches into consideration and till date a total of 14 teams have taken part in the IPL. In this era where the competition between bat and ball has increased to a greater extent, the teams should be prepared from the outside before going into the field.

Sports Analytics plays an important role for the team's preparation, using this tool the analyst informs the players, coaches and the other staff for decision making on the field, such as whether to choose to bat or ball taking the previous matches into consideration. In Sports Analytics two of the most important aspects are as follows: - On-Field and Off-Field analysis. On-field analytics becomes important for the players to make key decisions on the field. Data plays a key role in analytics.

2. PROBLEM STATEMENT

Fast-paced cricket demands on-field analysis for better performance of the players. Analysts work on the data and give the best possible idea to the players to make a decision on the field for better outcome.

3. OBJECTIVE

The main objective of this project is to help the players to make a good decision for better results, based on the previous data. Toss analysis, venue analysis helps in deciding a variety of factors for the best possible outcome. We have used the data set from 2008 to 2020, the more the data the better it is for the analyst to help the players make a decision. Also we have tried to find out the best player, best team on the basis of the data available and also we have shown why IPL is interesting to watch as a viewer.

4. TOOLS AND METHODOLOGY

For visualizing and analyzing the data of IPL we have used a jupyter notebook. We have used packages such as NumPy, pandas, Matplotlib and seaborn to get deeper insights of the data. We have used Python as the programming language. To clean the data,

we have used NumPy and pandas. To create basic visuals, we have used matplotlib and for better eye-catching visuals we have used the Seaborn package.

5. DATA COLLECTION

The data has been collected from www.kaggle.com. It is published by Prateek Bharadwaj. The dataset contains data of every match that has been played from 2008 to 2020. It has 17 columns and 816 rows. Each row represents details of a unique match, and each column represents a specific attribute. The dataset gives information about the city where the match was played, on which date the match was played, player of the match, venue of match (stadium), teams which played the match, toss winner, toss decision, match winner, match result, margin of win, super over, was D/L applied, names of the umpires of the match.

Dataset Description

Attributes	Description
Id	Unique Match ID as per ESPN cricinfo
City	City in which match was played
Date of Play	Date on which match was played
Player of match	Player who played best in the match
Venue of play	Name of the stadium in which match was played
Neutral venue	Was venue neutral?
Team1	Name of team 1
Team2	Name of team 2
Toss_winner	Which team won the toss
Toss_decision	Team which won the toss and decided to bat or field
Winner	Name of team which won the match
Result	team which won won by chasing the target or opposite
Result margin	By how many runs team won or by how many wickets team won.
Eliminator	Was there a super over?
Method	was D/L method used or not?
Umpire 1	Name of umpire 1
Umpire 2	Name of umpire 2

Data Preprocessing

To get better insights from data, it needs to be preprocessed and should be ready for analysis. Data can be noisy, missing, inconsistent and biased. There may be duplicate records. If raw data is analyzed, it can lead to wrong conclusions. To get better conclusions data should be cleaned, transformed, and reduced as per the requirements. There are 816 rows and 17 columns in the dataset but for few matches the city where the match was played, player of the match, result and by what margin the team won is missing. So, we dropped the record of such matches and cleaned the data. Also, few columns are not useful for our analysis and visualization has been dropped. Umpire1, umpire2, venue, neutral venue columns have been dropped. The data type of Date column data type has been changed from int to date data type and that column is replaced by year instead of date which shows in which year particular match was played. After preprocessing the data, it has resulted in 13 columns and 786 match records.

6. DATA ANALYSIS AND VISUALIZATION

Analyzing and visualizing the data gives us insights about the data and helps us to make future decisions based on the data. Once the data is preprocessed it is ready for visualizing and finding some interesting facts and hidden insights from the data. From 2008 until 2020 786 matches were played. During the TOSS of the match, out of 786 matches, the team had decided to field first in 477 (60.69%) matches and in 309 (39.31%) matches. Out of 786 matches in 408 (51.9%) matches the team which won the toss also won the match too. In 477 matches the team decided to field first and in 266 (55.7%) matches they have won whereas in 309 matches the team has decided to bat first and in 142 (45.9%) matches they have won.

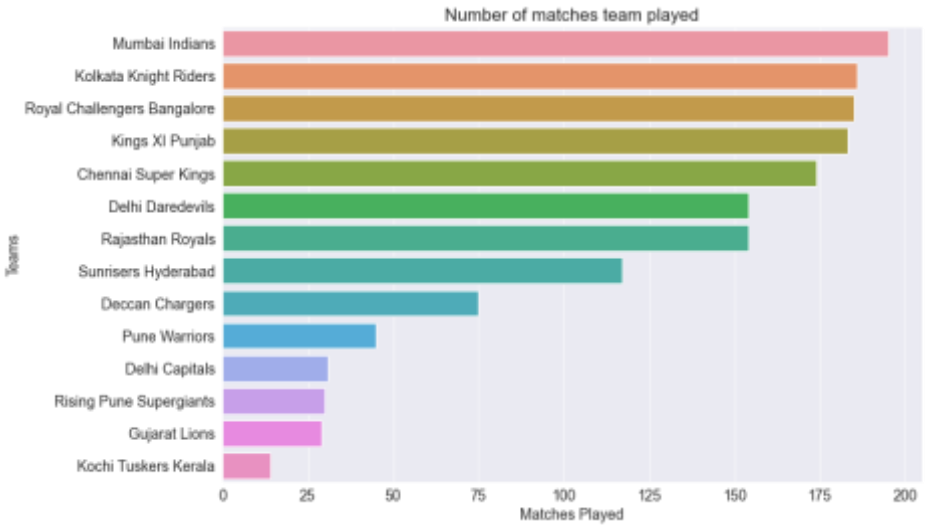

Fig.1 Number of matches teams played

From 2008 until 2020, 14 teams have participated in the IPL. From 786 matches in 19 matches Duckworth Lewis method (D/L) method was used to declare the winner and for the rest 767 results were declared normally.

From 2008 until 2020 maximum number of matches played by Mumbai Indians and least by Kochi Tuskers Kerala. Mumbai Indians played nearly 200 matches till 2020. Also, KKR, RCB and Kings XI Punjab have played more than 175 matches. It seems that Pune Warriors, Delhi Capitals, Gujarat Lions, Rising Pune Supergiants, Kochi Tuskers Kerala have played less than 50 matches until 2020. It seems the former teams have not participated in all seasons of IPL. More insights can be obtained from the visual given below.

During 2008 to 2010 there were less than 60 matches played in every season. From 2012-2013 there were about 74 matches played and in 2014 it reduced to 46 and from 2015 to 2020 the number of matches played were less than 60 as it was in the initial seasons. It can be easily observed from the visual below.

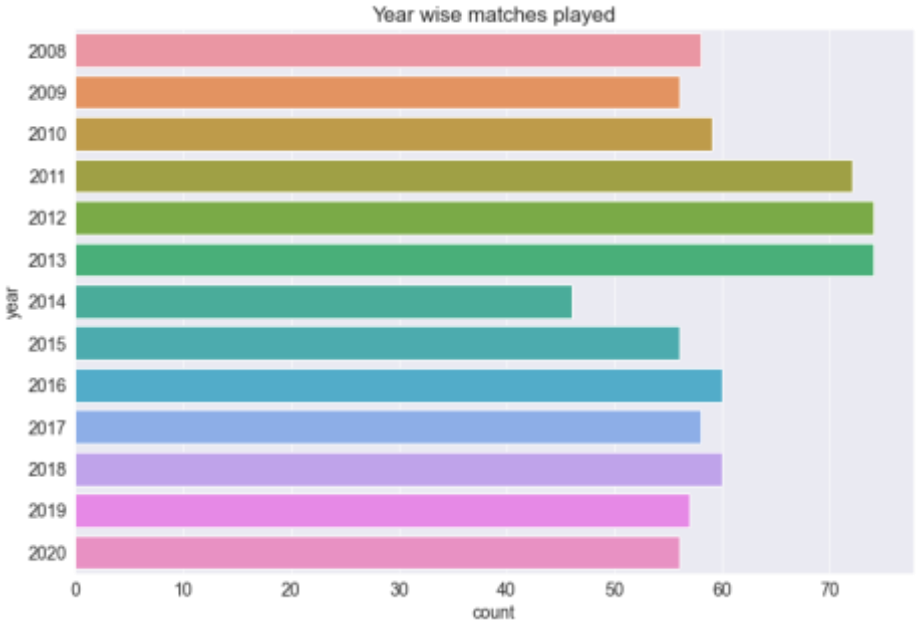

Fig.2 Year wise matches played

IPL has been played in 32 cities across the globe till 2020 and majority of matches has been played in India. Around 100 matches have been played in Mumbai. More than 60 matches have been played in Kolkata, Delhi, Hyderabad. 40+ matches have been played in Chennai, Chandigarh, Jaipur. IPL has been played in Abu Dhabi, Dubai, Sharjah and at many more places. The visual below gives us a clear picture.

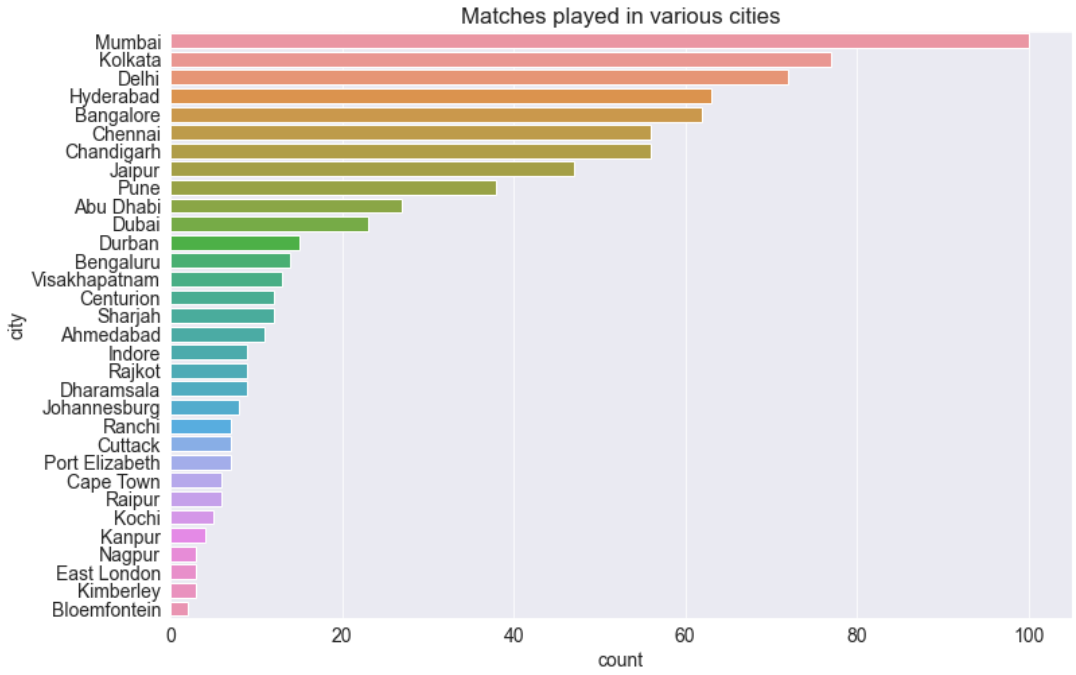

Fig.3 Matches played in various cities

From 2008 to 2020, it would be exciting to know which team won maximum matches. The data shows that Mumbai Indians has won in about 120 matches followed by Chennai Super Kings winning more than 100 matches. Kolkata Knight Riders won more than 90 matches. The graph gives us a detailed picture regarding the matches won by various teams.

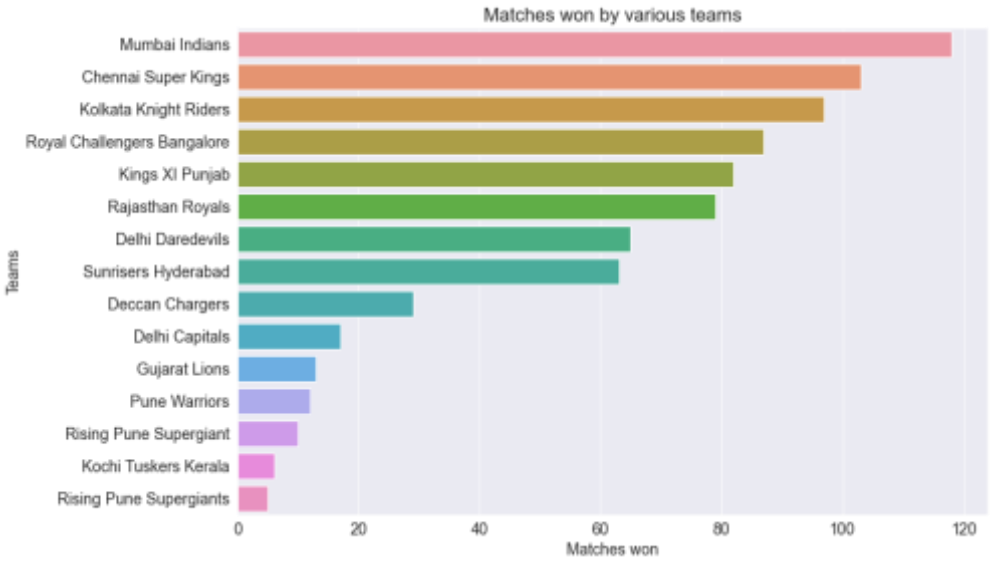

Fig.4 Matches won by various teams

We got a clear picture regarding which team won the match and how many times. Data also shows the number of times the team won the matches. From the below visual we get the clear picture that Mumbai Indians won close to 100 tosses followed by KKR, CSK, and RR. It seems that Pune Warriors, Delhi Capitals, Gujarat Lions, Rising Pune Supergiants, Kochi Tuskers Kerala have won less than 20 tosses till 2020.

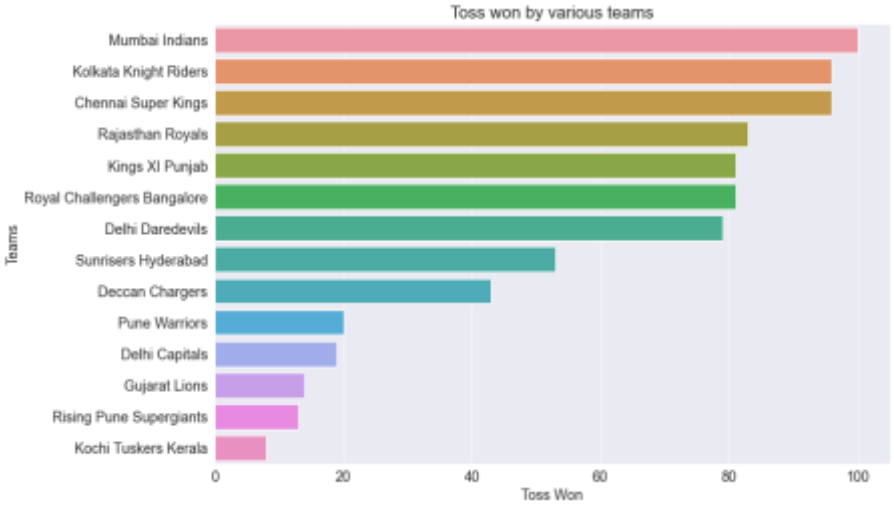

Fig.5 Toss won by various teams

We know that MI won maximum matches, but it would be super exciting to know which player became man of the match maximum number of times. The data shows that C H Gayle and AB de Villiers have become man of the match for 20+ matches. Rohit Sharma, M S Dhoni and DA Warner and S R Watson have been awarded man of the match for 15+ matches. The below visual showcases top 10 players who have been awarded man of the match in IPL from 2008 until 2020.

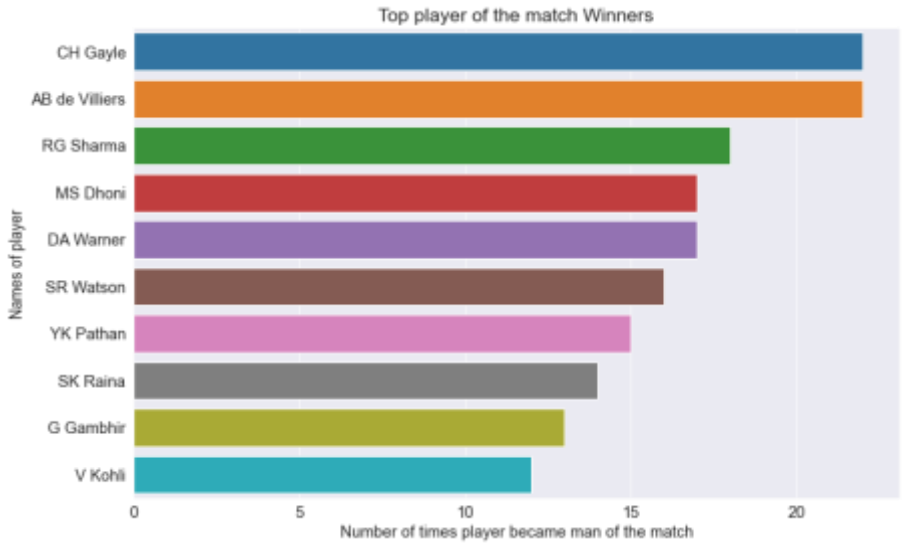

Fig.6 Top players

Why is IPL interesting?

A. The margin of win by wickets is average around 4 to 8 wickets, so as the graph shows most of the matches are nail-biting.

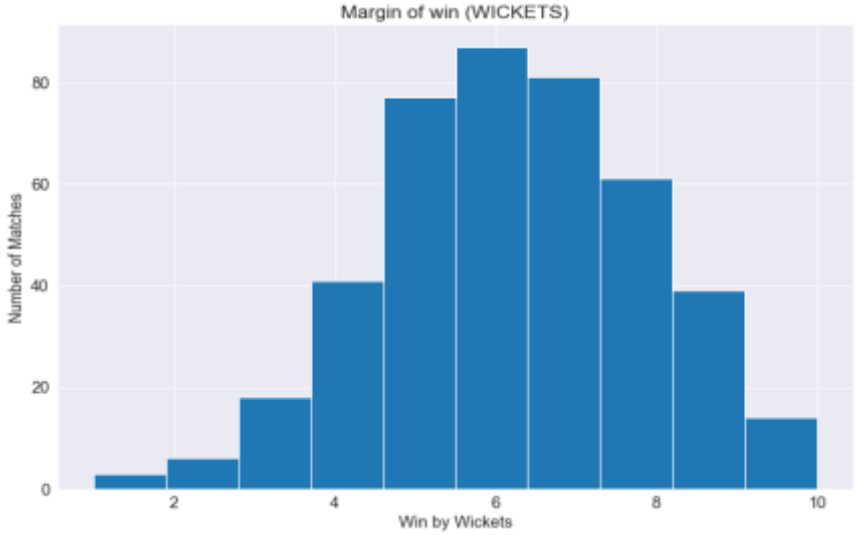

Fig.7 Margin by wickets

B. Also as the graph shown below suggests that the margin of win by runs is also quite low, which in turn suggests the matches that have a result is quite nerve wrecking.

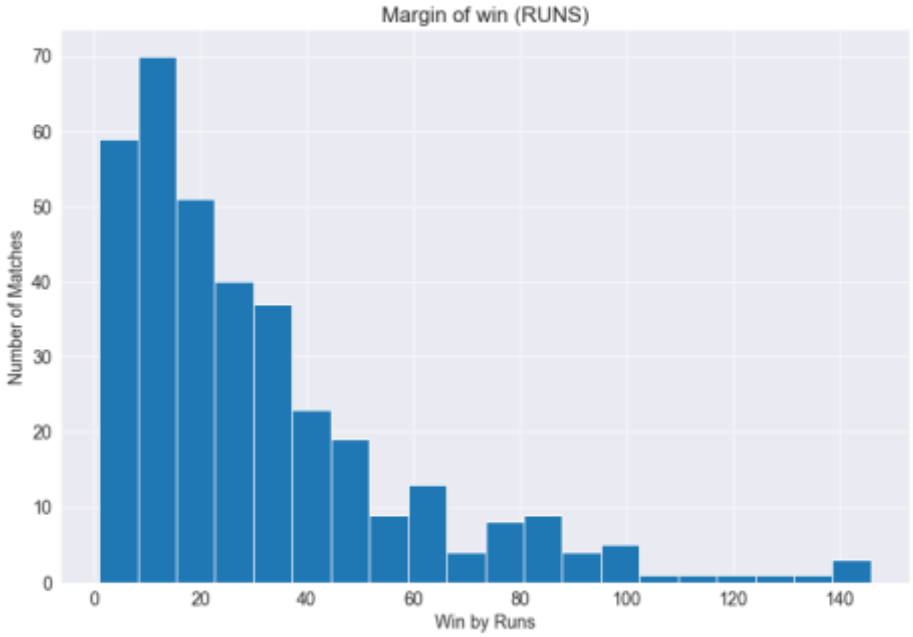

Fig.8 Margin by runs

7. CONCLUSION

This paper has intended on analyzing and visualizing the results of the Indian Premier League matches during the year 2008-2020, also on the basis of the data we were successful in showing that IPL as a tournament is quite interesting from user point of view, as most of the matches that have a result are quite nail-biting.

8. REFERENCES

[1] Shimona.S, Yuvarani.P, Nivetha.S, "Analyzing IPL match results using Data Mining Algorithms", International Journal of Scientific & Engineering Research Volume 9, Issue 3, March-2018.

[2] Vidit Kanungo, Tulasi Bomatpalli, "Data visualization and toss related analysis of IPL teams and batsmen performances, International Journal of Electrical and Computer Engineering, October 2019.

[3] https://www.kaggle.com/patrickb1912/ipl-complete-dataset-20082020?select=IPL+Matches+2008-2020.csv

[4] https://www.iplt20.com/

[5] Gupta, Amit. (2009). India and the IPL: Cricket's Globalized Empire. The Round Table. 98. 201-211. 10.1080/00358530902757875.

