EXPLORATORY DATA ANALYSIS OF INDIAN PREMIER LEAGUE: AN EMPIRICAL STUDY

Punam Bhoyar[1], Pranshul Agrawal[2]
[1]Associate Professor, Indira Institute of Management Pune.
Email: punambhoyar2005@gmail.com
[2]Student, Indira Institute of Management Pune.
Email: impranshu11@gmail.com

Abstract:
The Indian Premier League (IPL) is one of the most watched cricket events in the world, and both its viewership and its finances have significantly increased over the past few years. The Indian Premier League (IPL), which debuted in 2008 and is currently the most well-known T20 league worldwide, is played in India. Cricket is a tremendously fluid sport that evolves with each ball. People enjoy and watch cricket the most, and the Twenty20 format in particular since no one can predict the winner until the final ball of the final over. India's most popular and frequently played sport, with the largest fan following, is cricket. The 20 overs formula used in the Indian Premier League is quite unexpected. Therefore, we decided to do an exploratory data analysis to research and evaluate the IPL games. The findings of the analytics applied reveal several insights from the IPL dataset that can be used to better optimize both team and individual performance.
Keywords: Sports analysis, Cricket, IPL, Exploratory Data Analysis, Data Visualization.

I. Introduction

In India, cricket has long been compared to a religion. It's considered to be more than just a game. The cricketers are worshipped and revered as divinities. Cricket's Twenty20 for-mat made history. Since the Twenty20 version of cricket was introduced to the world, everyone's ears seem to be bursting with cricket talk. The history of the Indian Premier League is fascinating. Lalit Modi, former vice president of the Board of Cricket Control in India (BCCI), had the idea for the Indian Premier League before anything else. Its design is modeled around European club football, most specifically the English Premier League. IPL was launched as a counterpunch to Zee Group's rebellious Indian Cricket League (ICL). An incredible development in cricket history is the DLF Indian Premier League. In terms of the entertainment sector, it is viewed as a really good prospect. After India won the Twenty20 World Cup, the Twenty20 cricket format became a craze, and since then, the public has been completely engrossed in it. Regarding viewership and financial gain for all parties involved, there is no question as to where the Indian Premier League is headed. IPL has teams in important Indian cities and uses Double Round Robin, group, and knockout formats. IPL, which was created by the Board of Control for Cricket in India (BCCI), has grown to be the most lucrative and well-liked cricket league in the world. Matches typically start in the late afternoon or early evening so that at least some of the games can be played at night under floodlights for international television broadcasts. To increase possible television revenue, the top four teams compete in three play-off matches, with one losing team receiving a second chance to advance to the championship match. In the tournament's play-off phase, the top four teams compete in a series of knockout matches that provide one team that lost its first-round game a second chance to proceed to the championship game. The top cricket players in the world, who had rarely

made as much money as their counterparts in other professional sports, became millionaires practically overnight with the debut of the IPL. In auctions held by the league, the owners of the IPL franchises, which included major corporations, Bollywood movie stars, and media owners, competed for the top players.

The graphic depiction of data and information is the focus of the interdisciplinary topic of data and information visualization. It is a very effective method of communication when there are lots of facts or information. In this paper, we trying to get insights from the datasets which include team as well as individual performances through visualization.

II. Literature Review

Baseball was the first sport in which analytics were used. In 1858, sportswriter Henry Chadwick created a metric known as the box score. Baseball statisticians measured individual and team performance quantitatively because the box score tabulated a baseball player's performance. [1] [5]

Many others tried vainly to demonstrate the practical application of analytics in sports until the middle of the 20th century. Baseball Abstracts, a compilation of yearly baseball statistics, was created by Bill James and garnered widespread recognition in 1977. To describe the science underlying a baseball game, he later developed the word" Sabermetrics," which quickly gained popularity. SABR, or the Society for American Baseball Research, is where the name" s a b e r m e t r i c s " originates. It is the mathematical and statistical examination of baseball records, according to his first definition of" Sabermetrics." [5]

Everyone enjoys watching sports with in-depth and perceptive commentary. Data analytics was first applied to the sports business in the early 1990s. Since then, it has been utilized by everyone, including professional athletes and amateur athletes, to improve athletic performance, audience engagement, and marketing and branding tactics. For any professional sports organization, hiring new players is a crucial task. After all, while having exceptional athletes helps them win titles and sponsorships, talent alone is not the most crucial aspect. Athletic performance is influenced by several factors. Sports organizations in the modern era have realized this and are employing data analytics technologies to find the best candidates for their team's culture. Every sport, whether it be team or individual, requires planning. These tactics are necessary for professional sportsmen and teams to compete and defeat their opponents. Big data sets are used by modern coaches to develop effective methods that will benefit both individual athletes and the squad. [7] [5] [2]

We are surrounded by technology since we live in the technology age. Given how popular cricket is around the world, technology could not have remained unaffected for very long. Over the years, cricket has increasingly incorporated technology. The innovations have continued to be adopted, from third umpires to Hawkeye and Decision Review System (DRS), to power bats and helmet cams. The most popular sport, cricket, is also utilizing AI and IoT in tandem to improve the experience for players and spectators. We'll look at a few instances of their application in cricket. [3] [8]

To solve the issue of predicting cricket match outcomes based on past match data from the IPL, the authors look at machine learning technologies. Machine learning techniques such as Naive Bayes, Random Forest, K Nearest Neighbors, and model trees (classification via regression) have been adopted to produce predictive models from distinctive feature sets derived from the filter-based methods. These feature subsets were developed, one based on home team advantage and another based on toss decision. [11]

The main determinants of a cricket match's outcome are briefly discussed in this paper, along with the regression model that best fits the data and makes the most accurate forecasts. IPL

Match Predictor is a machine learning (ML) based pre- diction approach where the data sets and previous stats are trained in all dimensions covering all significant factors such as Toss, Home Ground, Captains, Favourite Players, Opposition Battle, Previous Stats, etc., with each factor having a different strength with the assistance of KNIME Tool and with the added intelligence of Naive Bayes network and Eulers strength calculation formula. [12]

In this article, the authors outline the major issues with SportsXR, such as data gathering, in-game choice-making, and contextual sport-specific visualization design, before presenting possible user scenarios for coaching and training. In this work, they discuss the difficulties and potentials of applying IA (Immersive Analytics) to sports to close the gap between sports data, people, and tools. Additionally, they present a SportsXR vision in three distinct user scenarios that empower athletes, coaches, and fans. [13]

In this study, a Deep Mayo Predictor model is shown for predicting the outcomes of IPL (season 9) matches. The model includes three components, which implies that three prediction models have been developed based on IPL matches from past seasons and IPL IX matches that have already been played. The Mayo Predictor Model successfully predicted the outcomes of 56 games played in the league stage of 39 games during the IPL IX competition, demonstrating the excellent prediction accuracy attained. [14]

Specular BatSense the Specular Batsense sensor was developed in partnership with Intel by the Bengaluru-based sports technology business Specular. The top of the bat's handle can be used to mount the Intel Curie-powered bat sensor. Each stroke a batsman makes would result in data being produced by the sensor. It contains internal storage for keeping track of session data and Bluetooth connectivity for real-time data transfer. [4]

Power Bats the Spektacom cricket bat, developed by Anil Kumble in collaboration with Microsoft, uses an IoT-based sensor sticker and is powered by artificial intelligence. Power bats are based on the idea of employing IoT sensors to gather essential information about a batsman. To assist professionals in raising their game, Spektacom attaches a tiny sticker sensor to a cricket bat. This sensor collects data on the quality, speed, twist, and swing of the bat. [4]

Cricket is a game of numbers: the runs a batsman scores, the wickets a bowler takes, the games a cricket team wins, the frequency with which a batter reacts in a particular way to a particular bowling attack, etc. It's a huge thing because advanced analytics tools, supported by numerical computing software like NumPy, can analyze cricketing data to examine business potential, the general market, and the economics of cricket in addition to examining performance data. Cricket analytics offers fascinating perspectives on the game and foresight into potential outcomes. [6] [9]

III. Research Methodology

IPL is a well-known and esteemed cricket league that gives aspiring cricketers a great platform to display their talents and offers the ideal learning environment for new and young cricketers. It has been demonstrated to play a vital role in the success of several cricket players. The analysis is conducted using publicly accessible secondary data. This dataset contains details on the ball-by-ball information and a description of each match. This dataset has a huge amount of data to explore. We attempt to analyze that dataset and fetch some valuable information from them. This IPL dataset consists of 15 seasons from 2008 to 2022 during that period 950 matches were played between 17 teams. Eight teams from the league represent eight distinct Indian states or cities at the start but as IPL signs progress then the number of teams also increases. From this dataset, we try to explore many things which are hidden in this data which will help to analyze player performances and team performances. Sports analytics are a group of pertinent, historical statistics that might give a team or person a competitive edge. Sports

analytics gather and analyze this data to provide players, coaches, and other staff with information to help them make decisions both before and during athletic events. This dataset consists of 40 variables and the total observation in this dataset is 226905. This dataset consists of variables like match id, venue, toss winner, total runs, wickets, batting team, bowling team, ball numbers, overs, and many more which give us details of each, and every ball bowled in IPL.

IV. Data Analysis

Data analysis is the methodical implementation of logical and statistical approaches to describe and demonstrate, summarize, examine, and assess data. To convert data into interactive visualizations, we use the Microsoft PowerBI Desktop version. To convert data into interactive visualizations, it can link to more than 70 on-premises and cloud data sources. We can create reports and provide them to the Power BI service. In PowerBI we join the data and do data transformation and modeling. This Dataset is a secondary dataset that we have taken from Kaggle. This data consists of each ball record from 2008 to 2022.
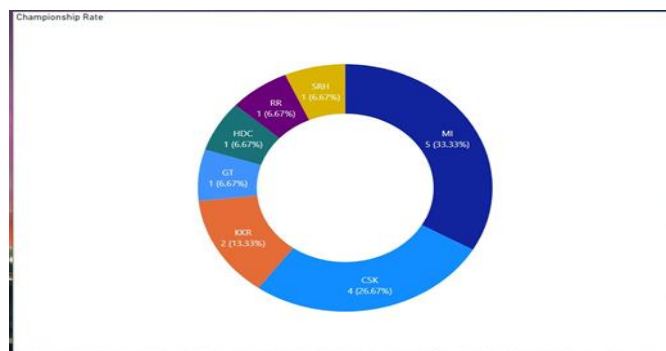

Fig. 1. Championship Rate

This visualization shows that in the percentage of teams who won the IPL trophies as we have seen Mumbai has the highest rate of titles winning percentage after that Chennai. This doughnut chart also shows how many times a particular team won the trophy.
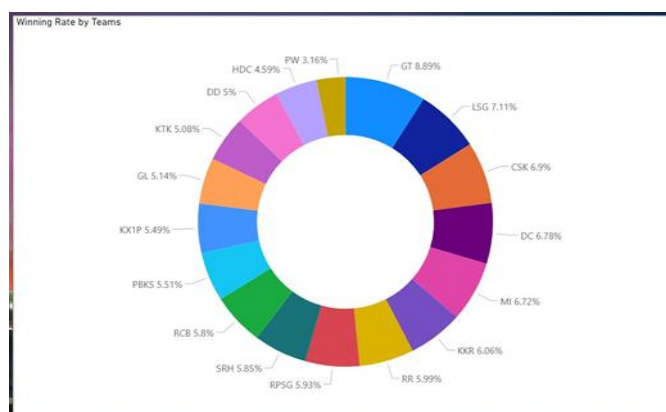

Fig. 2. Winning Rate by Teams

This visualization shows the overall winning percentage of each team. The analysis of the dataset is from 2008 to 2022 so till 2022 what is the winning percentage of each team we can find out easily from this graph.

Fig. 3. Overall Summary 1

This analysis page shows the overall summary of IPL and the number of matches that have been played, number of boundaries, 1st inning average score, 2nd inning average score, and number of centuries, and half centuries this chart covers all basic details.
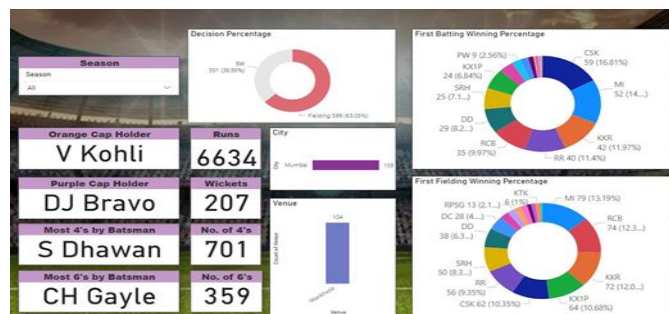

Fig. 4. Overall Summary 2

This page shows the highest run-getter and wicket-taker as well as which batsman scores the greatest number of 4's and 6's. The doughnut graph shows the percentage of decisions after toss winning; the greatest number of matches played in Mumbai at Wankhede stadium. This page also shows the winning percentage of the first batting and first bowling team as well.

Fig. 5. Overall Summary 3

This page shows the top 5 players who get the most MOM awards, the top 5 run-getters, wicket takers as well as the best fielder on the field in terms of runouts and catches taken by a fielder.

V. Conclusion

In today's sports world with the aid of analysis, we can determine the player's strengths and weaknesses which will in turn help to improve both individual player performance and team performance. This will aid in optimum decision-making. Working on player and team performance and predicting future performance is beneficial for all types of games. Other T20 leagues around the world, like the PSL, LPL, CPL, BBL, T20 Blast, BPL, and World League CLT20, benefit from the following methodology. Data Analytics have thus taken sports analytics to greater heights and thus selection of players in the teams and selection teams have become easy and accurate. Analyzing performance of individual players and teams over time have made accurate predictions over time possible.

VI. Future Scope

In today's data-driven world, data visualization is an essential tool, and the sports industry is no different. Man- agreement would need a lot of time to sift through all the data and understand the content when using raw data in tabular format, which prevents them from getting clear insights. In order for management to understand analytics given visually through graphs and plots or to discover new insights, the data must be presented in a graphical way. It is essential to comprehend the technicalities in order to fully utilize the sports industry's data analytics potential in order to get the highest player performance and increased likelihood of winning.

References

[1] https://datasportsgroup.com/news-article/74282/the-rise-of-sports- analytics/

[2] https://www.digiconasia.net/features/how-data-analytics-in-sports-is- changing-the-game

[3] https://wicky.ai/content/analytics/story-of-one-of-crickets-first-analytics- pioneers/

[4] https://www.analyticssteps.com/blogs/ai-cricket-data-analytics-cricket

[5] https://analyticsindiamag.com/a-primer-on-sports-analytics-a-new- dimension-of-sports/

[6] https://numpy.org/case-studies/cricket-analytics/

[7] https://seleritysas.com/blog/2021/03/27/how-is-big-data-analytics- changing-sports/

[8] https://sportsmatik.com/sports-corner/sports-technology/decision-review- system

[9] https://www.firstpost.com/long-reads/ipl-and-big-data-analytics-a-match- made-in-heaven-4438611.html

[10] https://skill-lync.com/blogs/how-are-ai-and-data-analytics- revolutionizing-cricket

[11] Kapadia, K., Abdel-Jaber, H., Thabtah, F., Hadi, W. (2020). Sports analytics for cricket game results using machine learning: An experimental study. Applied Computing and Informatics, (ahead-of-print).

[12] Sinha, A. (2020). Application of Machine Learning in Cricket and Predictive Analytics of IPL 2020.

[13] Lin, T., Yang, Y., Beyer, J., Pfister, H. (2020). SportsXR–Immersive Analytics in Sports. arXiv preprint arXiv:2004.08010.

[14] Prakash, C. D., Patvardhan, C., Lakshmi, C. V. (2016). Data Analytics based Deep Mayo Predictor for IPL-9. International Journal of Computer Applications, 152(6), 6-10.