

Data Visualization On IPL

Venkata Krishna Uppuluri, Medical Informatics, and Telemedicine, Carleton University, Ottawa
Canada

Abstract-- One of the hardest and most time-consuming tasks during IPL is selecting the teams. As a result, it is necessary to analyze and forecast sports data in order to determine which players are going to be the best and in order to identify which players are going to perform well. The aim of this research paper is to utilize data analytics and visualization techniques with regard to the data. Using vast collections of statistics and data sets, it can be determined how each player performs. The application is designed to bridge the gap between selecting teams, coaches, and managers through a better understanding of player steadiness, scoring, and other capabilities.

Throughout this paper, the tableau tool was used to analyze the data. Tableau provides Business Intelligence and analytics software that is easy to use for analysing data. An explanation of how Tableau's trend line models work can be found in this presentation. By comparing the results of both regressions, we can recommend which trend model is the most appropriate.

Index Terms- Business Intelligence, Data integration & processing, Design implementation, Tableau tool, Marketing Implication, Statistics, Facts & Features.



1. INTRODUCTION

Data visualization and sports analytics have given team management and players a robust platform to improve their performance on the field. In decision-making and analysis, algorithms are applied to data to produce insights into future predictions. Several algorithms, tools, and visualization approaches are used to process this data in order to allow player suggestions while building teams.

IPL (Indian Premier League) was founded in 2008. Round-robin groups and knockouts are the formats of the league, which includes clubs from several major Indian cities. The current starting 11 can only have four foreign players and the total number of foreign players on the team can be no more than eight when team management makes a bid for as many as 25 players. There is a lot of work to be done when it comes to finding the best team for the upcoming seasons. Performance assessment for players is introduced in this paper. Players may visualize and predict their performances using this tool. Decision-makers can use the proposed model to compare two teams' strengths during IPL matches.

The organization of the paper is as follows:

1. To conduct accessible raw data analysis and interpretation.
2. To forecast each person's performance.
3. To extract player performance information from the massive dataset and visually display it as graphs for more effective analysis.

2. A SURVEY

There is often a use of player data analysis in sports. Sports analysis is replete with statistics. It is a fact of life that every professional in the sports industry lives in the present as well as the future. Besides facilitating player and team analysis, the stadium's inauguration provides accurate predictions. The challenge at [1] is attempting to predict the result of an IPL cricket match. There are several flaws in this study, including its use of a dynamic non-relationship database, the HBase application firmness. Among the data examined at [2] were run totals of IPL players, wicket totals of the most successful team, general performance of the team, MVP runs and wickets, throwers' runs and wickets, Duckworth law winners' run and wicket totals, and

throwing winners' pitching totals. Using a tableau, the whole analysis is performed on the basis of the presentation. Numerous IPL teams predict the results of nearly every game scenario using extreme analysis so as to anticipate the match winner. Athletic performance of athletes, particularly batters, can be seen by examining Man of the Matches, Top Batsmen, and Top 10 Players with the Most Runs. Through these processes, data is modified, consolidated, and refined. An analysis of cricket players' performances was conducted by the authors of [6], and it concluded that battering power was greater than bowling power. Study results suggest that the throwers' performance is a key component of changing the status quo. The player ranking concept for the IPL auction was discussed in [7].

3. DESIGN & IMPLEMENTATION

3.1 Tools & Methodologies

Indian Premier League is followed by millions of people worldwide. There have been many big leagues played but this is one of the biggest. During the period from 2008 to 2020, 816 games were played. Data about every match is available on the internet in a vast amount of detail.

A variety of data exploration and extraction activities are carried out in Excel, as well as feature selection. Data can be manipulated and analyzed using these tools. Player performance is predicted by using Tableau, which analyzes and displays the analyzed data.

3.2 Data Collection

We describe the datasets that we selected for our study in this section. Downloaded datasets were retrieved from www.kaggle.com. Our coverage includes all teams from 2008 to 2019. Matches.csv and deliveries.csv are the two datasets that are used. The Matches.csv database contains details on match IDs, cities where matches are played, dates, locations, players, teams, toss results, match winners, results, and match umpires. A Ball-by-Ball dataset includes information such as match ID, innings, overs scored by a specific bowler, strike rates by bowlers, total runs scored, and wickets were taken, as well as the names of the bowling side and the batting side. As shown in fig 01.

Content
<ul style="list-style-type: none"> Data till Season 11 (2008 - 2019) matches.csv - Match by match data deliveries.csv - Ball by ball data
Acknowledgements
<ul style="list-style-type: none"> Data source from 2008-2017 - CricSheet.org and Manas - Kaggle Data source for 2018-2019 - IPL T20 - Official website

Fig:01 showing content & references (ref 14)

3.3 processing Data Before Use

Pre-processing data is the most crucial step in any data science project. In terms of project time, it is a significant factor. As part of pre-processing data, inaccurate and inconsistent information must be removed, already-existing data can be formatted, and missing numbers should be added. In addition to eliminating undesirable information, redundant observations are also removed. The purpose of this software is mostly to transform, standardize, and correct data. Results are guaranteed to be accurate by performing this step.

3.3.1 Inter-Connected Data

The below screenshot describes that I have concorded the two raw data sheets (Match Data & Innings Data). So that I can co-relate which are in different sheets. And perform various visualization. As shown in fig 02 & 03.

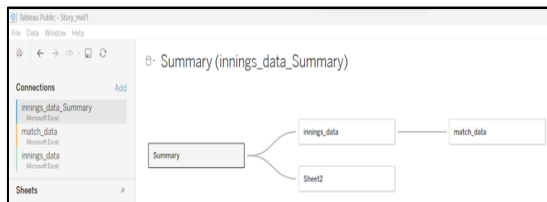


Fig:02 Created interconnected data in tableau

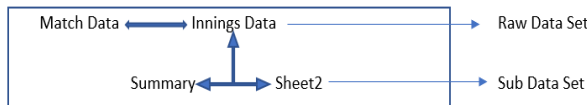


Fig:03 displays the type of data

Secondly, I manipulated the data (didn't change or add any external data) and just made a summary of it. So that I can visualize in an effective manner. As shown in fig 03. Like the raw data was about 176574 rows & 20 Columns. Which is huge and very difficult to project. So, I need to simplify the source.

3.4 Choice and Extraction of Features

A crucial part of the analysis process is determining which features are going to be used to analyze a cricketer's performance. It is important to consider factors such as the location, the opponent, the bowler the batsman was dismissed by, how many runs the batsman scored during a powerplay, how many runs he scored during a death over, how many runs he scored during the first versus second innings for a batsman. Factors such as location and the other team are taken into consideration by bowlers. These features are taken out of a cleaned-up but sizable dataset. Two new datasets were developed in accordance with the project requirements. Sheet 2 and Summary are both attached. In order to form a CSV dataset, we combine the matches.csv and deliveries.csv datasets. As part of Summary.csv, you will find the match id, total runs scored in each match, powerplay runs, death overruns, the location of the match, and the bowling team

information. All information about bowling teams, wickets, match IDs, locations, and pitches is shown in sheet 2. CSV.

In the below fig 01., I mentioned three different data sets which were created from "Innings Data". And I marked the creation links (In Innings Data there is a column named "batsmen Team" and I created a union to a column named "Teams" in sheet 2). And similar to Summary (Union to season named columns in both Data).

	Summary	innings Data	Sheet2
Season	id	Teams	
Row Labels	season	Row Labels	
Sum of batsman_runs	batsman	bowled	
Sum of extras_noballs	bowler	caught	
Sum of extras_wides	innings	caught and bowled	
Sum of extras_legbyes	non_striker	hit wicket	
Sum of extras_byes	replacements	lbw	
Sum of extras_penalty	bowled_over	obstructing the field	
Sum of total_extras_runs	batsman_team	run out	
	player_out	stumped	
	fielder_caught_out	Total Wickets	
	type_out		
	extras_wides		
	extras_legbyes		
	extras_noballs		
	extras_byes		
	extras_penalty		
	total_extras_runs		
	batsman_runs		
	total_runs		

Fig:04 Creating relations between data sets

4. EXAMINING AND INTERPRETING DESIGN

The aim of a data analysis effort is to make the data visually appealing so that consumers can understand the meaning behind it. The visual representation of data simplifies the interpretation process. The graphs contain all the information concerning the features culled from the large datasets. We also utilized a variety of tools in order to obtain precise analyses and visualizations of teams and players. Each participant's data was processed using Excel. To display graphs, Tableau data formats are created by applying calculations to data frames produced from the dataset.

In this visualization, we have a total of three dashboards. Which represents management and player performance perspectives.

4.1 Dashboard I

This is the view of management who is going to buy either the whole team or an individual. And here I have clubbed four different data visualizations into one story, so that management may be able to connect and understand about team's performance to date. We shall discuss this further in detail.

4.1.1 Scores Comparison by Team

This graph displays a run analysis of teams. It displays each team's performance in terms of runs scored regardless of the location, innings, or bowler. As shown in fig 05.

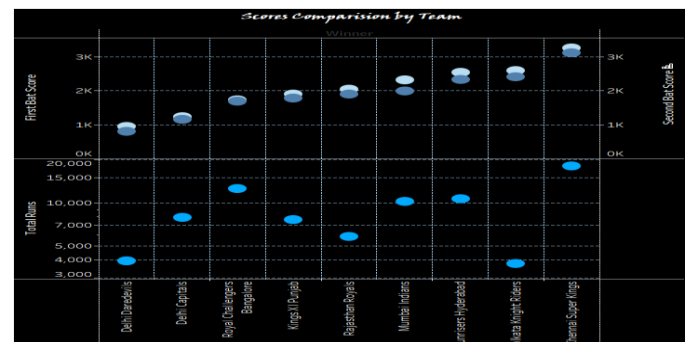


Fig:05 Scores comparison by team

4.1.2 Analyzing Teams Performance Based on Runs

The below graph displays overall performance in two types. As shown in fig 06.

- The First one shows how many runs were given to the opposite team in terms of Extras.
- The second, displays runs scored by the team apart from extras provided by the opposite team.



Fig:06 Analyzing the team's performance based on runs & extras

4.1.3 Analyzing trend on runs per year wise

The point to implement this line chart is to understand how the performance of teams is since the start of the tournament. As shown in fig 07.

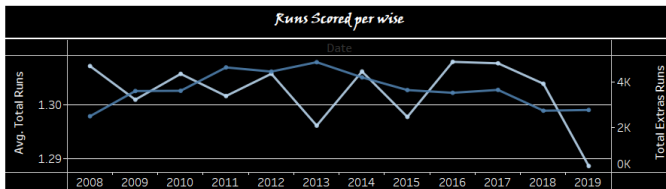


Fig:07 Analyzing trend on runs per year wise

4.1.4 Analyzing Top & Bottom Performance Teams

As this is a management view, so I have installed this to display the team's performance in terms of runs, so here we can see the change in performance over the years. As shown in fig 08.

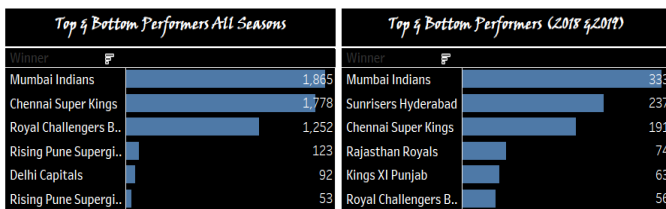


Fig:08 Analyzing top & bottom performance teams

4.2 Dashboard II

This is the view of a player's performance perspective, as we know there will be tons of players around the globe playing for a team. So, the sponsor will have an internal review for a player before he purchases one.

4.2.1 Analyzing Player's Performance

This is to basically display about player's performance in terms of bowling side, like how many wickets he has taken so far over a decade of tournaments. And also, how many runs has he given to the opposition team? As shown in fig 09.

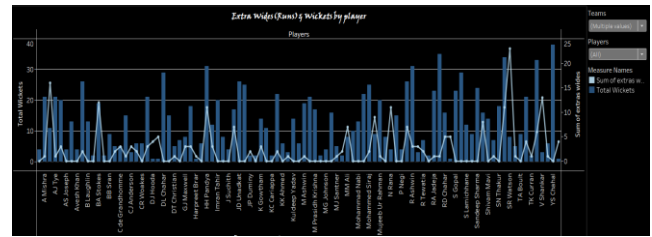


Fig:09 Analyzing player's performance

4.2.2 Analyzing Player's Performance on types of Out's.

As a part of the player's performance, this is a unique view of representing four categories in a single line and bar chart. It helps us understand where his key role in playing out of those four categories. As shown in fig 10.

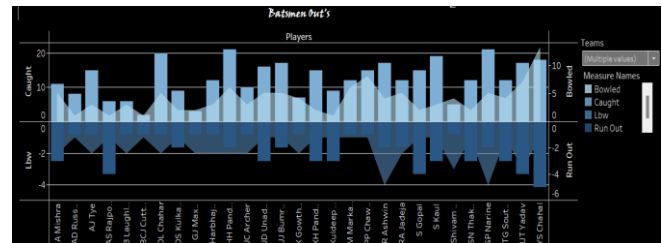


Fig:10 Analyzing player's performance on types of Out's

4.3 Dashboard III

This is the last dashboard of my data visualization, in here I wanted the user to have a look at the team's performance again. Where are we losing a match exactly, we have seen many incidents even with a good batting performance, the team lost the match, and this is because of extra runs provided by the team. By looking at this display coach will be able to understand and know areas of improvement.

4.3.1 Analyzing Extras by Categories

This is again a management view performance-based visualization. The below figure shows how many extra runs have been given by individual teams, In each category. The best way to visualize is dribbled pie chart because it is animated and can display multiple options in a single view. As shown in fig 11.

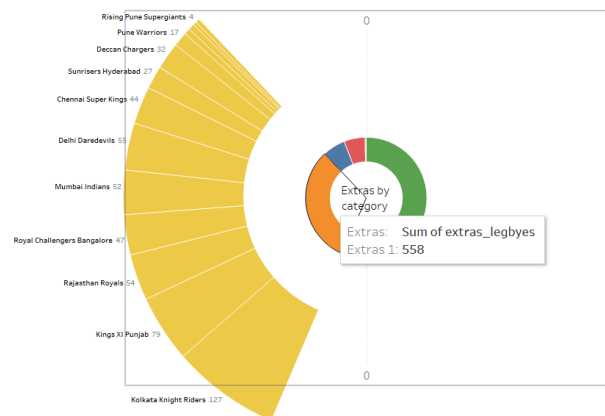


Fig:11 Analyzing extras by categories

5. RESULTS & DISCUSSION

Figures following display the web application's user interface. There are three dashboards in all. From Figures “5-11”.

5.1 The first dashboard illustrates:

This is the viewpoint of the management, which will purchase either the entire team or a single member. And in this case, I've combined four distinct data visualizations into a single narrative so that management can make connections and comprehend the team's current performance. We will go into further detail about this. As shown in figures “5-8”.

- Usually, the sponsor buys players in every auction, but there are times the whole team will be out to be sold. So, the sponsor will act accordingly. Pie charts were implemented to help understand the user, in terms of the size of the difference in either “run scored or extras given” from one team to another. The colors were taken to help the visualization look good.

5.2 The Second dashboard illustrates:

As there will be many players playing for a team around the world, this is the perspective of a player's performance. Therefore, before buying a player, the sponsor will do an internal review of that player. As shown in figures “9-10”.

- Here I have displayed both “Teams & Players” in filters. Because you can either filter on team-based or player based. And I have updated the filtering options now it will work properly for both charts.

5.3 The third dashboard illustrates:

This is the final dashboard I created for my data visualization, and I included it so the coach could review the team's performance once more. Where exactly are we losing games? In many cases, despite strong batting performances, the team fell short and this was due to extra runs contributed by the team. The coach will be able to comprehend and identify areas for improvement by looking at this display. As shown in Figure “11”.

- This is an advancement in prototype visualization, I want to show the user in which category the team is lagging in the game in terms of restricting additional(free) runs to the opposition team.

6. FUTURE WORK

Eventually, the scope of the project will expand to include more elements such as the batsman's posture as part of future development. As a result of projecting a 360-degree analysis of an individual batsman, the project can be extended to predict the performance of the bowler on the basis of the area of the ground where they are scoring the most runs. We also hope to expand the project's dataset to include statistics from other cricket competitions, such as the test & limited overs in international cricket and as well as league cricket.

7. CONCLUSION

In the proposed work, we have visualized the performance analysis of IPL cricket players from seasons 2008 to 2020 from the perspective of performance analysis. It has been identified that players' performances can be evaluated based on factors such as playing time,

innings, death overs, powerplay overs, and bowler type in the project. By having an accurate projection of batsman runs prior to the start of the match against the opponent team and at the venue, the management of the team will be able to select the best players for each match against that opponent team. Using the facts and features that have been provided by them, we have produced models for both batting and bowling data sets.

The link to Visualization : [Krish_Assignment#6 | Tableau Public](#)

8. REFERENCES

- [1] Vedit Kanungo, Tulasi B., “Data visualization and toss-related analysis of IPL teams and batsmen performances”, International Journal of Electrical and Computer Engineering, Vol. 9, No. 5, October 2019.
- [2] Shubhra Singh, Parmeet Kaur, “IPL Visualization and Prediction Using Hbase”, Information Technology and Quantitative Management, 2017.
- [3] S. Sharuka, R.Vani,” Insights on IPL Team Performance using Visual Analytics”, International Journal of Engineering Sciences & Research Technology, November 2019.
- [4] Kasukruti Raviteja, Ganesh Kumar Macha, Dr. GR Anantharaman, “Predicting and Analyzing the Performance of the IPL Cricket Using Regression Models”, Complexity International Journal, Volume 23, Issue 03, Dec 2019.
- [5] Kalpdram Passi and Niravkumar Pandey, “Predicting players performance in one-day international cricket matches using Machine Learning”, 8th International Conference on Computer Science, Engineering and Applications, February 2018.
- [6] Sricharan Shah, et al., “A Study on Performance of Cricket Players using Factor Analysis Approach,” International Journal of Advanced Research in Computer Science, vol. 8, no. 3, 2017.
- [7] D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League", Frontier Economics, vol. 116, October 2008.
- [8] C. D. Prakash, C. Patvardhan and C. V. Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL-9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October 2016.
- [9] M. Ovens and B. Bukiet, "A Mathematical Modelling Approach to One-Day Cricket Batting Orders," Journal of Sports Science and Medicine, vol. 5, pp. 49-5- 502, 15 December 2006.
- [10] Tijender Singh, et al., “Score and Winning Prediction in Cricket through Data Mining,” International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Oct 2015
- [11] Hemanta Saikia and Dibyojyoti Bhattacharjee, “An application of multilayer perceptron neural network to predict the performance of batsmen in Indian premier league,” International Journal of Research in Science and Technology, vol. 1, no. 1, 2014.
- [12] M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), 2016.
- [13] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp 37-44, Jan 2002.
- [14] Navaneesh Kumar, “Indian Premier League 2008-2019 Ball-by-ball Indian Premier League (IPL) cricket dataset”