**IPL ANALYSIS AND DASHBOARD CREATION**

**A PROJECT REPORT**

**Submitted by**

**DEEPAN RAJ. S (21ADR010)**

**ASWATH. M (21ADR005)**

**PAVITHRA (21ADL063)**

**VEENAS KUMAR. S (21ADR058)**

***for***

# 20ADC33 DATA ANALYSIS

# DEPARTMENT OF ARTIFICIAL INTELLIGENCE



**KONGU ENGINEERING COLLEGE**

# (Autonomous)

**PERUNDURAI ERODE– 638 060**

DECEMBER 2022

**KONGU ENGINEERING COLLEGE(Autonomous)**

**PERUNDURAI ERODE – 638 656**

**DECEMBER 2022**

Department of Artificial Intelligence

20ADC33 – Data Analysis Project Report

Signature of course –– in-charge Signature of the HOD

Submitted for the continuous Assessment viva voice examination held on \_\_ \_

**EXAMINER I EXAMINER II**

# ABSTRACT

Following India's success in the 2007 T20 World Cup, the BCCI launched the Indian Premier League, a franchise-based Twenty20 cricket league, on September 13. The first season was supposed to premiere in April 2008 in New Delhi during a "High-profile event." The IPL effort's driving force, BCCI vice-president Lalit Modi, laid out all the specifics of the competition, including its format, prize money, franchise income scheme, and squad makeup guidelines. There is no set amount of time that teams must finish their innings during IPL games since television timeouts are used instead. However, if the umpires determine that teams are abusing this advantage, a fine may be assessed.

In this project, analyzing the IPL data using data analysis techniques and build a dashboard that makes it simple for users to navigate and view the data. The dashboard will contain a range of interactive tools including filters and drill-down options in addition to visualisations like bar charts, line graphs, and heat maps.The important performance indicators, including runs scored, wickets taken, win-loss ratios, individual and team statistics, will be the focus of the data analysis. Additionally, examine current data and evaluate the effectiveness of various teams and players.

The data analysis and dashboard application also offers insights into how certain teams and players have performed. The programme offers data on players, teams, and matches as well as visualisations of important variables like run rate, strike rate, and wicket-taking ability. Teams may utilise these insights to identify areas for development, and supporters can understand better the performance of the team they support for by using insights.

In conclusion, teams, analysts, and fans looking to better understand the Indian Premier League may surely benefit from our data analysis and dashboard tool. Users may filter the data and go further into it to understand the success of the league because it is interactive and user-friendly. It is a helpful tool for evaluating player and team performance and finding areas that require development. Anyone who is interested in the Indian Premier League must have this tool.

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** |  |
| 1. | **INTRODUCTION** |  |
|  | 1.1 INTRODUCTION | 5 |
|  | 1.1.1 DATA COLLECTION | 6 |
|  | 1.2 PROBLEM STATEMENT | 9 |
|  | 1.3 BUSINESS OBJECTIVE | 9 |
| **2.** | **DATA PREPARATION AND MODELING** |  |
|  | 2.1 DATA CLEANING | 10 |
|  | 2.2 DATA TRANSFORMATION | 10 |
|  | 2.3 DATA MODELLING | 20 |
| **3.** | **DATA ANALYSIS AND INTERPRETATION** |  |
|  | 3.1 DATA ANALYSIS | 21 |
|  | 3.2 PUBLISHING DASBOARDS | 32 |
|  | 3.3 INFERENCE | 34 |
| **4.** | **CONCLUSION** |  |
|  | 4.1 RECOMMENDATIONS | 36 |
| **5.** | **REFERENCES** | 37 |
|  |  |  |

# CHAPTER 1

# INTRODUCTION

## INTRODUCTION

Following India's success in the 2007 T20 World Cup, the BCCI launched the Indian Premier League, a franchise-based Twenty20 cricket league, on September 13. The first season was supposed to premiere in April 2008 in New Delhi during a "high-profile event." The IPL initiative was driven by BCCI vice-president Lalit Modi, who laid out the specifics of the competition, including its structure, prize money, franchise income scheme, and squad makeup guidelines.

A data analysis dashboard may be created to evaluate the performance of the teams and players in the IPL. A multitude of data and insights, such as team and player statistics, performance measures, and current data, may be found on this dashboard.

The dashboard may be made to be interactive and user-friendly, enabling users to quickly filter and sort data by several criteria including team, player, season, and kind of match. To make the data more comprehensible and usable, the dashboard might also incorporate visuals like charts and graphs.

On the dashboard, one can evaluate a number of important parameters, such as:

* Averages for batting and bowling.
* Rates of strikes and the economy.
* Runs scored and wickets taken.
* Fielding data.
* Rankings of teams and players.

The dashboard may assist viewers, the media, and commentators in understanding the performance of the teams and players in the IPL and in making more educated forecasts and analyses by offering a comprehensive and simple-to-use data analysis tool.

Overall, the IPL data analysis dashboard is an effective tool for comprehending the performance of the teams and players in the league. It can also serve to improve media and analysts' analysis and commentary, as well as the viewing experience for fans.

## 1.2 DATA COLLECTION

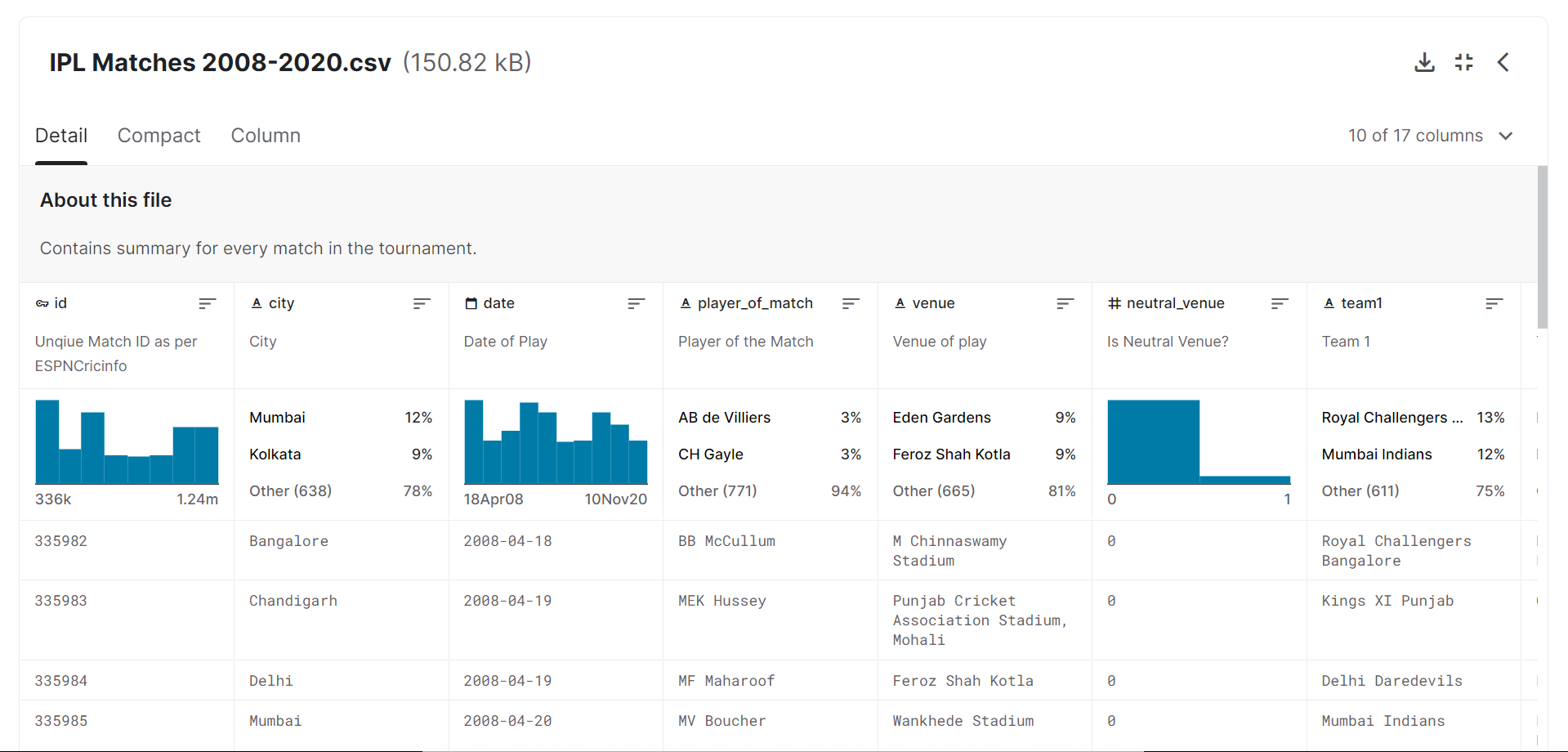
Data collection is a process which is the main part of Data analysis. Because through the dataset one can analyze and create some dashboards. The downloaded dataset is **“IPL COMPLETE DATA SET (2008 – 2020)”**. The dataset is collected from Kaggle and it contains mostly all the attributes essential for creating visuals and dash boards. The dataset is available in Excel format. This dataset consists of two separate CSV files: matches and deliveries. These files contain the information of two data sets. The first data-set is Ball by ball data set. This data set contains Each ball data. By using dataset, analyze the player and ball data. The second data set is IPL matches data set. This data set contains each match data and also analyze the team data and match data. The first data set contains 18 columns. Like match id, inning, over, ball-batsman, bowler, extra runs. The second data set contains 17 columns. Like Id, city, date, player of match, team1, team2, venue, toss winner and connect the tables by using their id. Because the match id is the primary key.

Figure-1.1- Dataset from Kaggle

DATA SET: <https://www.kaggle.com/datasets/vora1011/ipl-2008-to-2021-all-match-dataset?select=IPL_Ball_by_Ball_2008_2022.csv>

**DESCRIPTION OF DATASET**

CONTENT OF DATASET:

The dataset contains a detailed information about IPL. Because it contains 2 tables. One contains the whole ball by ball data. And another table contains team description in clear manner. Mainly the dataset are gathered form 2008 to 2022. This dataset contains the data about whole IPL history.

**Indian Premiere League Dataset**

*The variable in IPL matches 2008-2022 data-set are:*

* ID int: unique id for each record
* City str35: City where matches held
* Date date: Matches held Date
* Season int: Season Year
* Match Number varchar (20) : matches team played
* Team1 str35: first team
* Team2 str35: Second Team
* Venue str35: Match held place
* TossWinner str35; Toss Won Team name
* TossDecision str5: Toss won Team’s Decision
* SuperOver str: the match has Super Over no
* WinningTeam str35: Winning Team
* WonBy str20: Won type
* Margin int: Margin by runs/wickets
* Player\_of\_Match str35: Man of the match
* Team1Players str90: Team1 player list
* Team2Players str90: Team2 player list
* Umpire1 str35: first Umpire
* Umpire2 str35: Second Umpire

*The variables of IPL\_Ball\_by\_Ball\_2008\_2022 dataset are:*

* ID int: unique value for each record
* Innings int: innings value
* Overs int: Each over count
* ball number int: ball count
* batter str35: batter name
* bowler str35: bowler name
* non-striker str35: non-striker name
* extra\_type str20: any other type to get run
* batsman\_run int: bats man run
* extras\_run int: extra run
* total\_run int: total run for the ball
* non-boundary int: value for boundary balls
* isWicketDelivery int: count of wicket
* Player\_out str35: Out player name
* Kind str35: how the player out
* Fielders\_involved str35: fielders involved to out the batsman
* BattingTeam str35: Batting team name

Contains data for each ball for the entire tournament from 2008 to 2022. The data will be updated every day until the season's final match. The match ID is given by ESPN Cricinfo and is also used by other datasheets and acting as the primary key.

This IPL analysis is used to analyze data about players in detail to understand who is good at what aspect. In IPL a metric that they use, is MVPI or The Most Valuable Player Index, which is a weighted composite score of the different attributes of a player.

This whole data helps us to understand the weak and the strong area of different players, whether a player is good at hitting boundaries or at the running between the wickets, whether a bowler performs better against left-handed batsman or right-handed batsman, whether a batsman perform better against spinner or fast bowlers.

**1.3 PROBLEM STATEMENT**

Using python programs like pandas, matplotlib, and seaborn, IPL matches from 2008 to 2017 were analyzed in this study. This exploratory data analysis will assist us in identifying correlations and patterns in the data. By gathering data on each team's success rate, identifying the team that has won the most seasons, best defensive and chasing teams, and toss choices, among other things, try to determine which team has the greatest probability of winning the forthcoming seasons.

## 1.4 BUSINESS OBJECTIVE

* A professional Twenty20 cricket competition in India is called the Indian Premier League (IPL). The league's financial goal is to make money from a variety of sources, including broadcast rights, sponsorship agreements, ticket sales, and merchandising.
* The IPL has been successful in attaining this goal by creating a league that draws a sizable viewership and is extremely engaging and exciting. Strong fan bases for the league exist both in India and overseas, which has resulted in lucrative broadcast rights agreements with top television networks. The league and its franchisees have benefited financially as a consequence.
* The league has also benefited from sponsorship agreements with well-known companies like Vivo, Pepsi, and Dream11. The selling of goods and match tickets has also allowed the league to make a sizable profit.
* Additionally, the IPL has been effective in forging a distinctive brand identity and advancing the league through a number of marketing initiatives. This has contributed to the league's growing popularity and ability to draw in a wider audience, which has generated income.
* In general, the IPL's business goal is to make money via selling broadcast rights, sponsorships, tickets, and merchandise in addition to other things. The league has succeeded in attaining this goal by developing a very exciting and engaging league that draws a sizable viewership.

# CHAPTER 2

**DATA PREPARATION AND MODELING**

## DATA CLEANING

* The data may be inaccurate, include duplicates, or be incomplete after processing and organisation.
* Problems with the way the data is input and stored in tables will lead to the requirement for data cleansing.
* The act of avoiding and fixing these problems is known as data cleaning, and DAX may be used in this process to clean the data.
* Columns and measurements can be created using the computer language Data Analysis Expressions [DAX].

## DATA TRANSFORMATION

This method is used to change the data into formats that are appropriate for the mining process. This contains the following process:

* NORMALIZATION
* ATTRIBUTE SELECTION
* DISCREATION
* CONCEPT HEIRACHY LEVEL
* Data sources for IPLs from 2008 to 2022 that are freely downloadable from the internet. Get started on how to create efficient dashboard in Power BI with IPL dataset.
* The IPL, also known as the T20 Professional Cricket League of India, is played in April and May amongst teams representing various Indian states and towns. Therefore, this information includes information on clubs, players, etc. as well as ball-by-ball specifics of games played up to the previous year.
* Here,two CSV files with different data from the IPL match are downloaded from the kaggle.

**Data preparing**

When data is first acquired, it has to be organised or processed before analysis. These, for instance, can entail organising data into rows and columns in a table style (known as structured data) for later analysis, sometimes with the use of statistical or spreadsheet software .

STEP1: The dataset is first loaded from to POWER BI file for further processing and visualization.

STEP2: Then go to DATA view and then in the home tab select TRANSFORM DATA and load all the four data tables into int for pre-processing and modelling.

****

Figure-2.1- Transform Data tab

### STEP3: To make the first row as column headers, select Home Tab > Use First Row as Headers. To make column as headers to the first row, select Home tab, select the arrow next to Use First Row as Headers, select Use Headers as First Row.

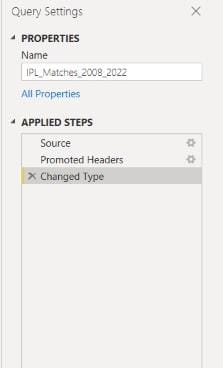
****

Figure-2.2-Applied Steps

Step4: Replace the value which is not appropriate.

**Graphical user interface, application, table, Excel

Description automatically generated**

Figure-2.3-Select replace values

Step5: Replace the team name from rising pune supergiants to rising pune supergiant

****

Figure-2.4-Replacing values

### Step 6: Again replace the 2009/10 to 2010.

**Graphical user interface, table, Excel

Description automatically generated**

Figure-2.5-Replacing Year

### Step 7 : Similarly replaced the year 2020/21 to 2021.

**Graphical user interface, table, Excel

Description automatically generated**

Figure-2.6-Replacing another year.

**Graphical user interface, table, Excel

Description automatically generated**

Figure-2.7-select Seasons column

### Step 8 : Right click query settings . By right-clicking a step in the list to organize, it shows replace, remove, and rearrange steps so that one can track the changes they made on the data.

**Graphical user interface, text, application

Description automatically generated**

Figure-2.8-Applied Queries

### Step 9: Finally, choose close & Apply option to load the Dataset to power-BI

**Graphical user interface, application

Description automatically generated**

Figure-2.9 select close & apply

The descriptions make it very easy to interpret the  content.

* Let's start with Power BI Desktop. Upload these 2 Excel files to Power BI Desktop and start building your reports as described in my previous post.
* Graphical user interface, application

  Description automatically generatedOnce the data from these Excels into your spreadsheet, one can start modelling the data. Click Modelling as shown in the screenshot below.

Figure-2.10- After applied all transformations

**COLUMN CATEGORIES**

Graphical user interface, application

Description automatically generated

Figure-2.11-Catogorize the data

Consider the scenario where one can want to show the number of matches played at each venue. This sort of data is included in the match table, but they 've altered the data categories to make it possible to identify them as geographic places.

Click the Modelling menu or the Data icon in the left sidebar to accomplish this, as seen in the picture above. After choosing a city, one will see an icon close to a column that one may use to display geographic imagery. To use additional location fields in geographic visualisations,and also do same with other location fields by changing their category.

If would like to show the number of matches played at each venue. The match table has this sort of data and changed the data categories so they can be recognised as geographic places.

To do this, select the Data icon on the left sidebar or the Modelling menu, as seen in the picture above. Once a city has been chosen, a column with an icon next to it may be used to display geographic imagery. Similar to this, one can utilise additional location parameters and modify their category in order to use them in geographic graphics..

**Relationship in data tables**

For better visualisations, one should need to link some of the many tables that they have. Based on the identical columns, Power BI desktop automatically detects the relationship in the tables. One can view the relationships as indicated in the following sample if they click Manage relationships or the symbol in the side bar.

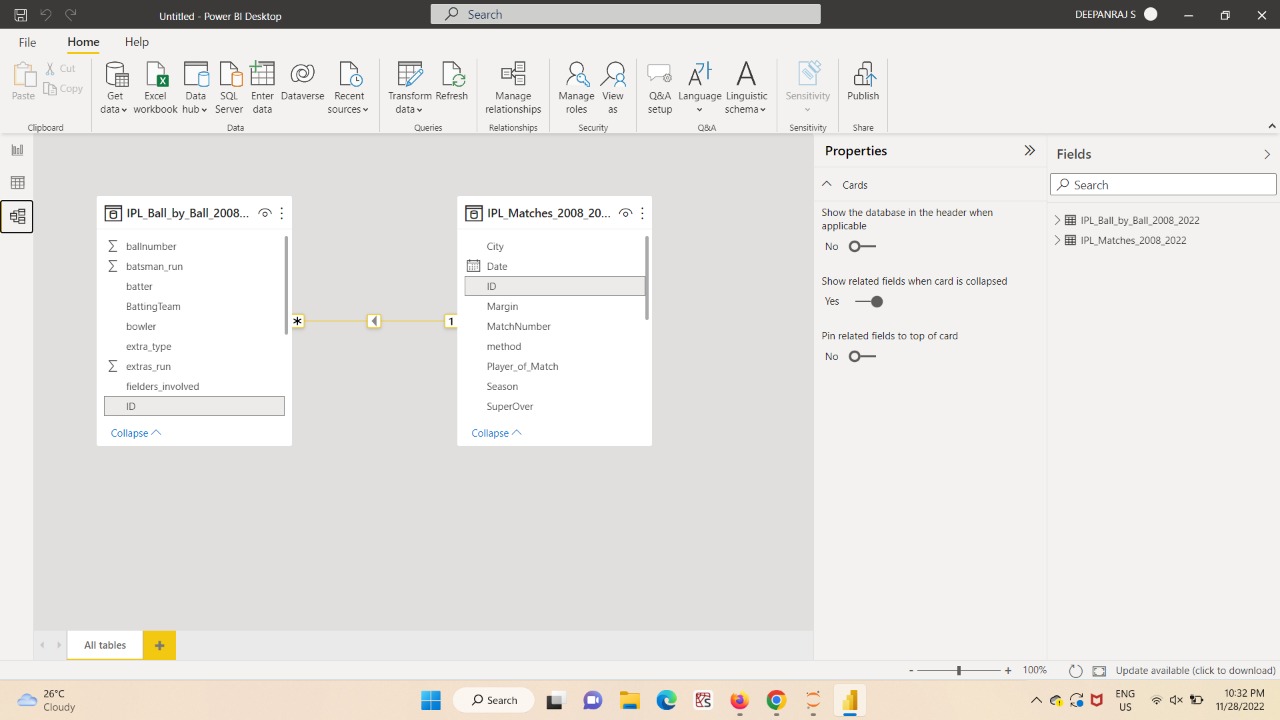


Figure -2.12-Model view

The match ID is given by ESPN Cric-info and is also used by other datasheets and acting as the primary key.

Match table that contains all match-related information, including the winning side, the player of the match, whether the match was decided by runs or wickets, and more. How can one display the statistics for "Win by Wickets" and "Win By Runs" based on the winning type? Open the table now. Two columns (Win Type and Win Margin) can give this information; all that is required is to display it in the rows.

**ADD COLUMNS**

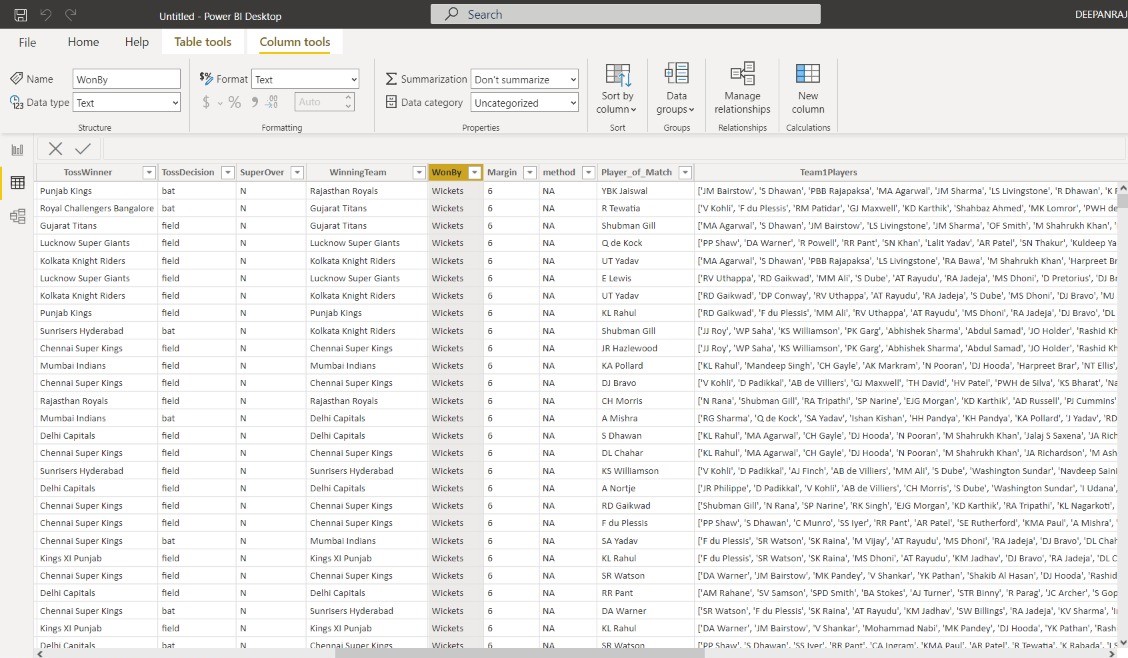
******

Figure-2.13-WonBy column

In order to display the victory margin in the event that the match is won by wickets, let's add a new column named Win\_By\_Wickets and add the algorithm. Let's perform the actions detailed in the samples below.

Step1: Touch the Add column tab which is placed in right of the navigation tab.

Graphical user interface, text, application

Description automatically generated

Figure-2.14-Add column

Step 2: Through a DAX ,one can fill with needed values which is processed in later purpose.

Graphical user interface, application, Word

Description automatically generated

Figure-2.14- Win By wickets column

The complete fromula:

Win\_By\_Wicketes=

IF(IPL\_Matches\_2008\_2022[WonBy]“Wickets”,IPL\_Matches\_2008\_2022[Margin],”0”)

Similarly for Win by runs , it would be

Win\_By\_Runs= IF(IPL\_Matches\_2008\_2022[WonBy]="Runs",IPL\_Matches\_2008\_2022[Margin],"0")

Now ,these two new columns added in the table for further visualization purpose.

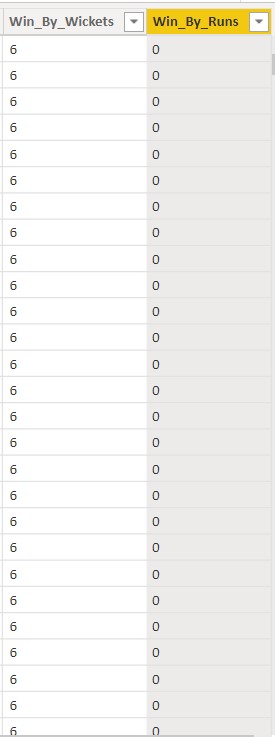


Figure-2.15-Win\_By\_Wickets and Win\_By\_Runs column

## DATA MODELLING

* Using solid data, organisations may establish baselines, benchmarks, and goals in order to advance. Information must be organised through data description, data semantics, and data consistency requirements in order for this measuring to be possible. A data model is this abstract model that enables the establishment of links between data items and further development of conceptual models.
* Even if an organisation has a sizable data repository, it is useless if there is no standard to guarantee the fundamental accuracy and interpretability of that data. The best tools to access the data and knowledge of best practises for the data are all certified by a proper data model.
* The Indian Premier League (IPL) uses data modelling to gather, analyse, and interpret information on the clubs, players, matches, and statistics in the league. To comprehend patterns, forecast consequences, and make strategic decisions, the data is employed.
* The official match statistics, player statistics, and team statistics are the primary data sources for IPL data modelling. Models that can forecast game results, individual performance, and team performance are developed using these data sources.
* The development of a player performance model is one instance of data modelling in the IPL. To forecast a player's performance in upcoming games, this model combines data on player statistics including runs scored, wickets taken, and fielding statistics. The technique may also be used to pinpoint important players and forecast who will do well in upcoming games.
* Overall, data modelling in the IPL is a crucial instrument for comprehending the functioning of the league and formulating strategic plans. Teams and organisations may enhance their performance by using data analysis and interpretation to understand league patterns, forecast results, and make wise decisions.

# CHAPTER 3

**DATA ANALYSIS AND INTERPRETATION**

## DATA ANALYSIS

**Team analysis**

1. What is the average run score by the teams in first innings?
2. What is the average run score by the teams in second innings?
3. What is the total runs score by team?
4. What is the total balls bowled by the team?
5. What is the total number of matches played in 2009?
6. How many number of matches win by Mumbai-Indians?
7. Which team conceded more extra runs?
8. How toss decision impact the match winning?
9. Which type of extra runs is more conceded?
10. Compare CSK and MI based on runs, extra type on first and second innings?

**Player Analysis**

1. Total number of players played in between 2008 to 2022.
2. Total number of bowlers played in 2022
3. Total number of batsmen played for CSK.
4. Which player has got highest man of the match
5. Which player takes more number of catches.
6. Total runs scored by Virat Kohli in the year 2008
7. Total wickets taken by the DJ bravo in IPL history.
8. Which player give more extra runs?
9. Total number of fours scored by SK. Raina?
10. Total number of sixers scored by CH. Gail?

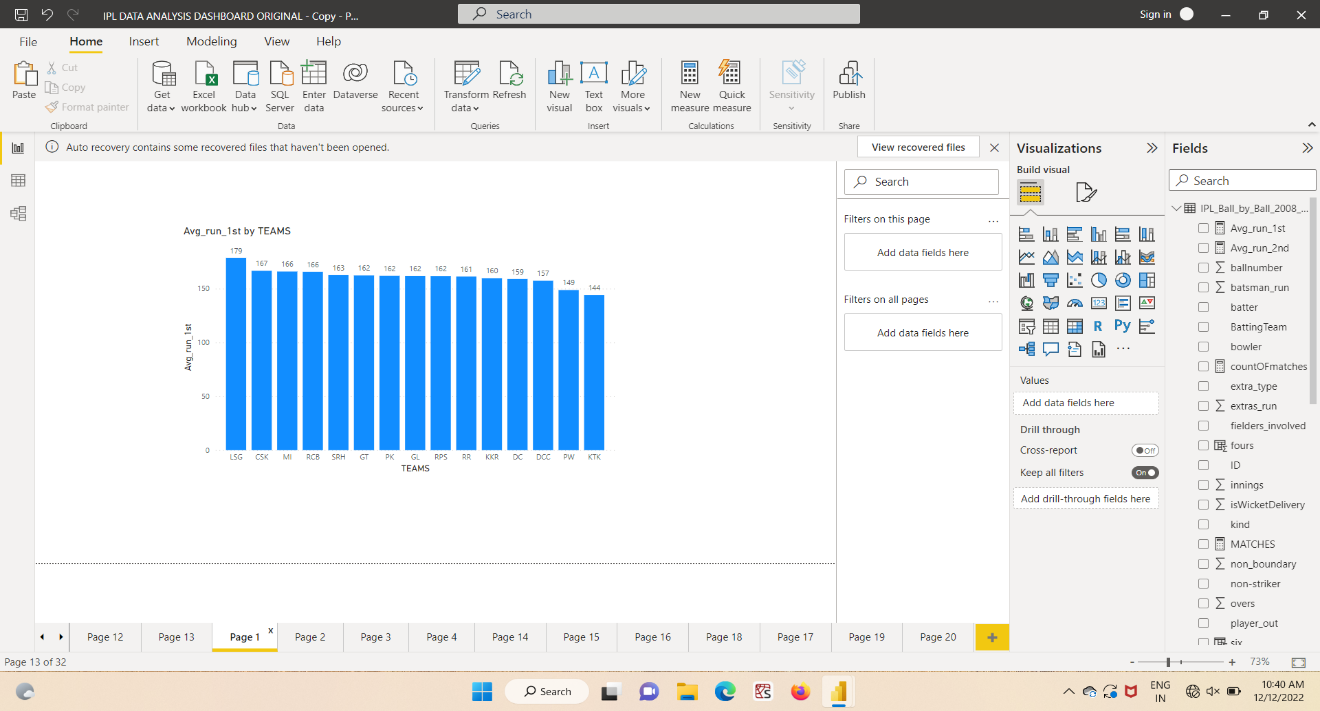


Figure : 3.1 Average runs by teams in first innings

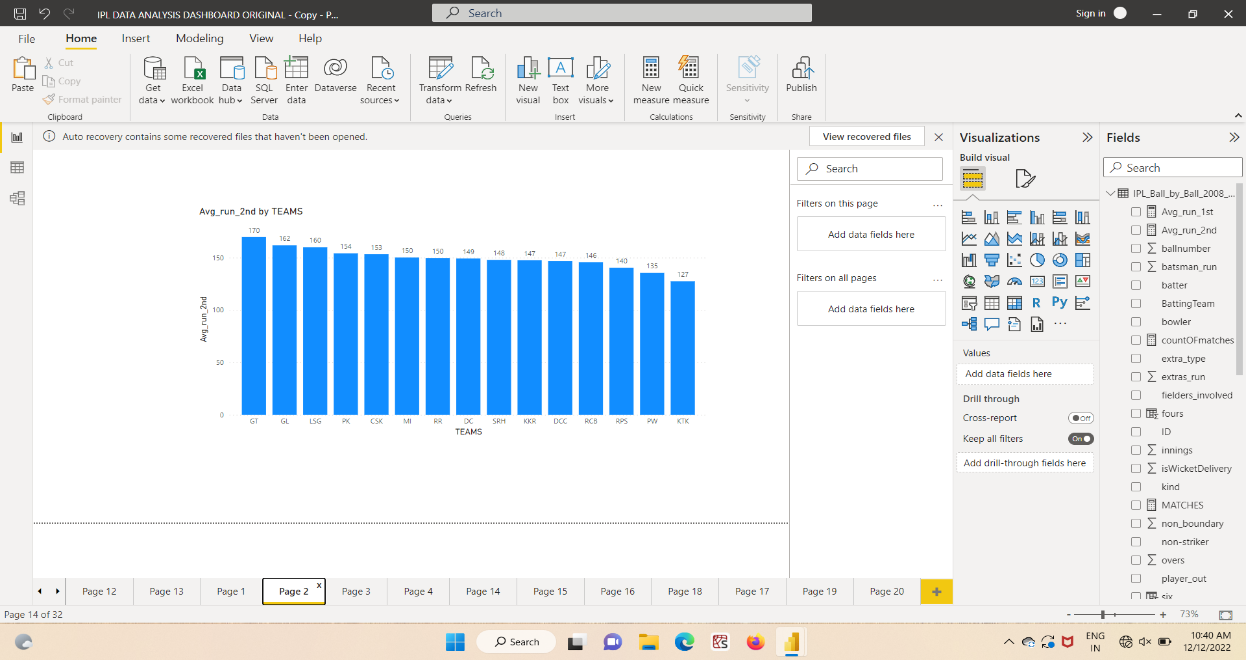


Figure : 3.2 Average runs by teams in second innings

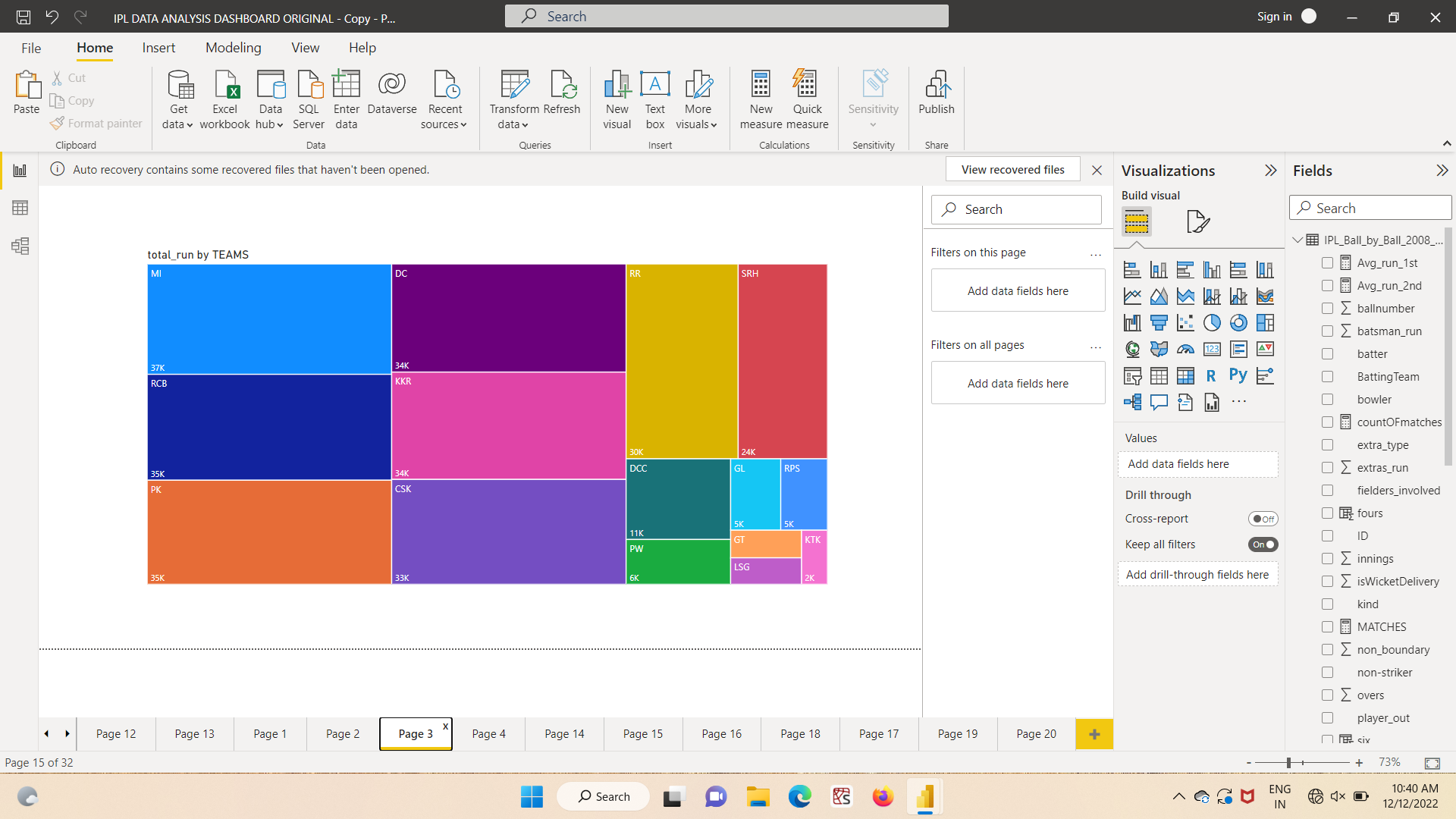


Figure : 3.3 Total runs scored by teams

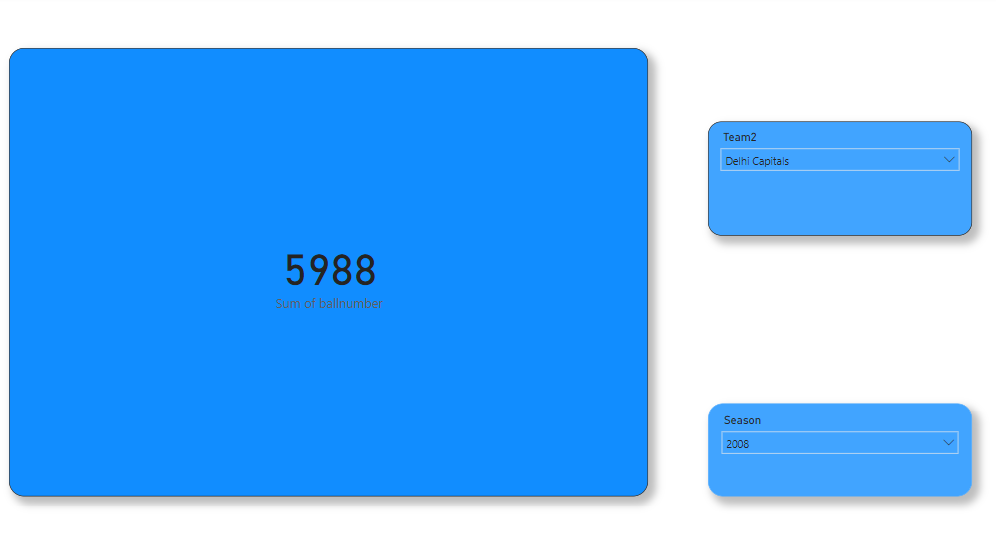


Figure : 3.4 Total balls bowled by the teams



Figure : 3.5 total number of matches played by all team in 2009

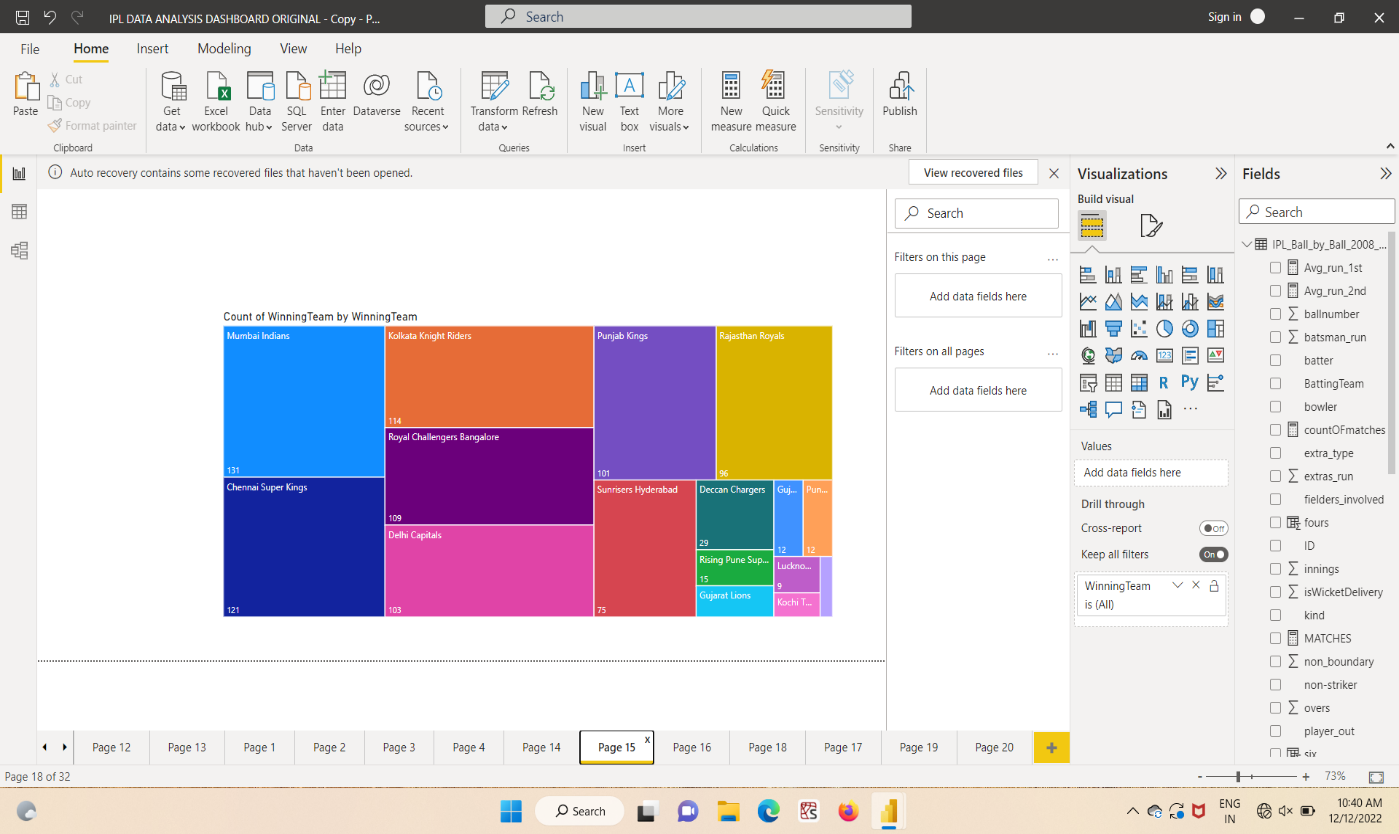


Figure : 3.6 total number of matches played by all team in 2009

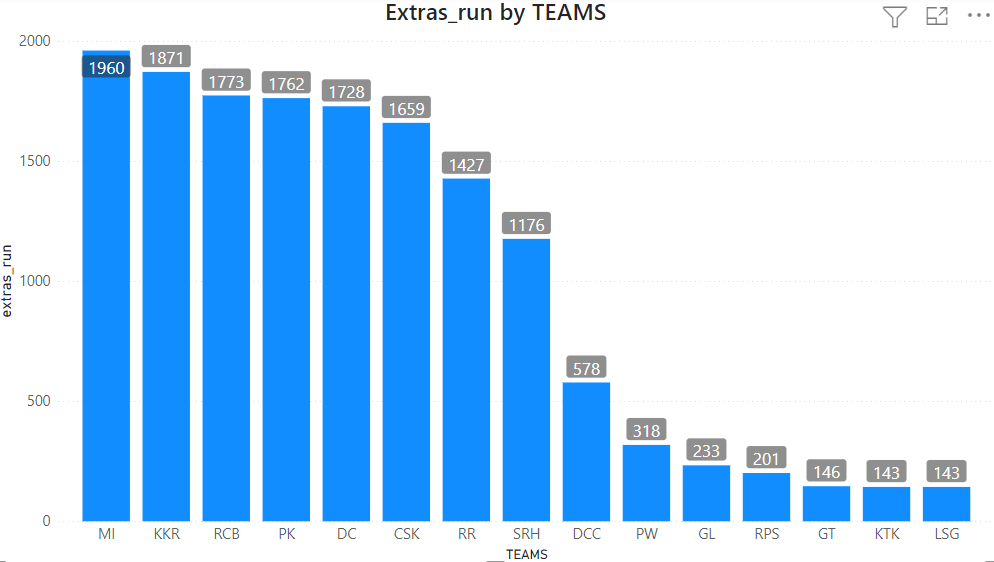


Figure : 3.7 extra runs conceded by each teams

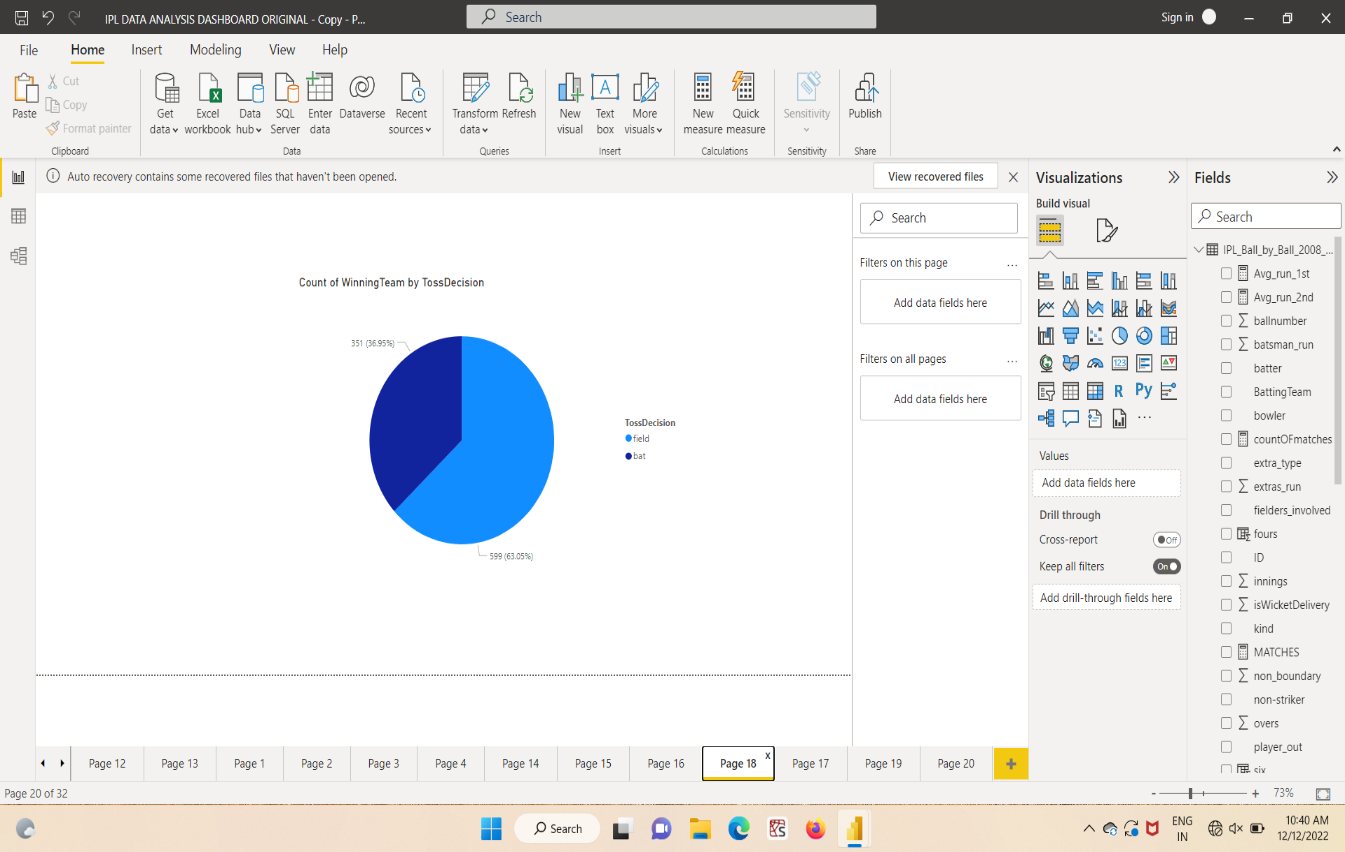


Figure : 3.8 impact on match winning by toss decision

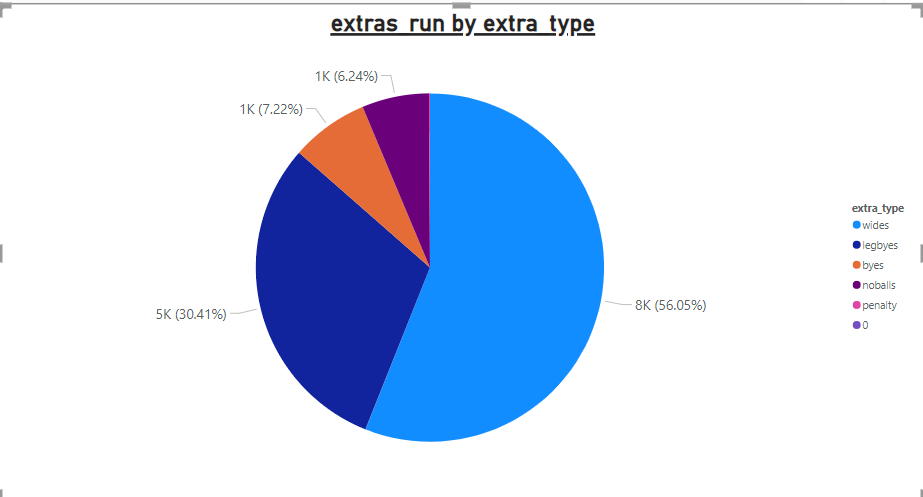


Figure : 3.9 Types of Extra runs

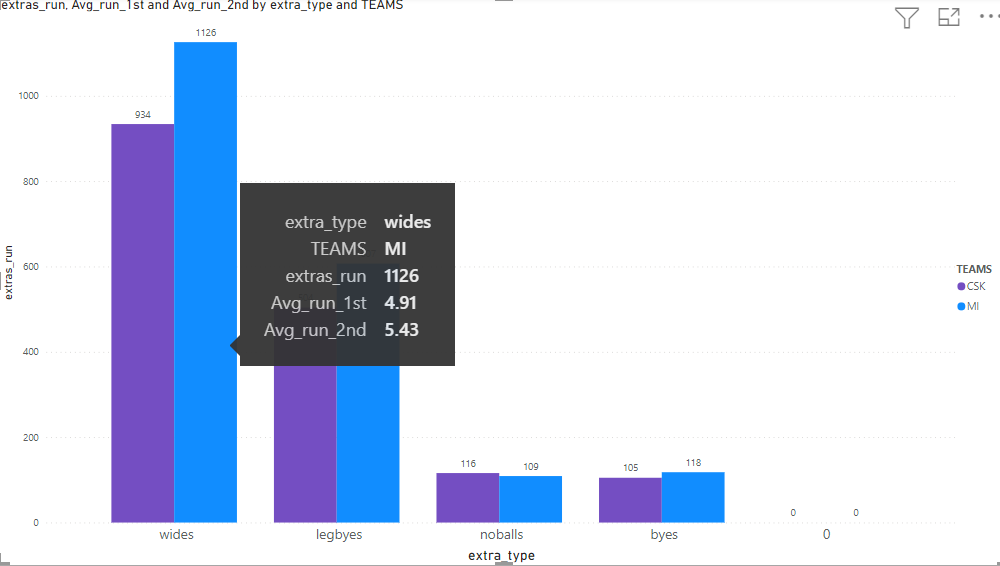


Figure : 3.10 Extra runs conceded by CSK and MI

**Player analysis**

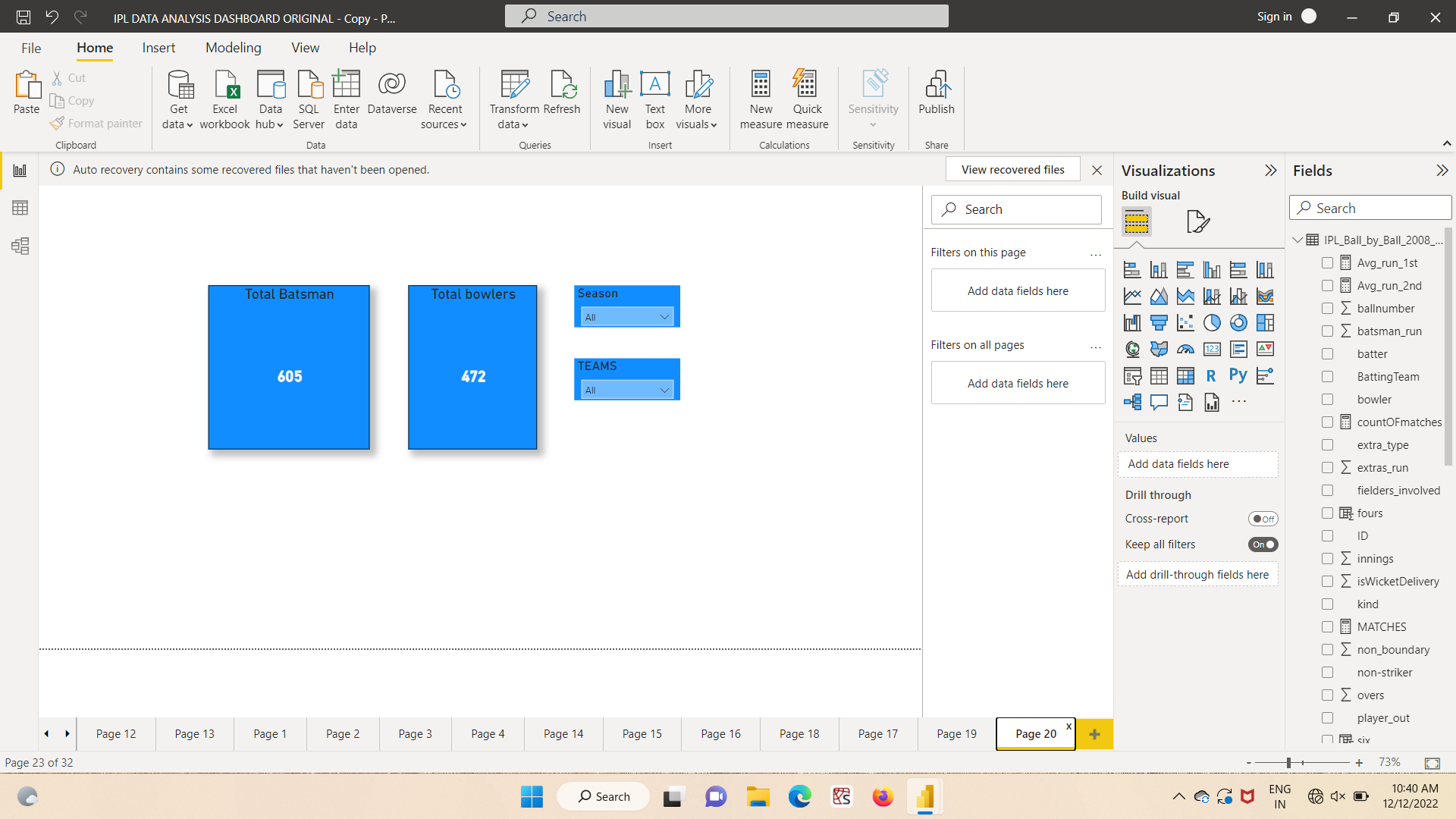


Figure : 3.11 Total no.of players

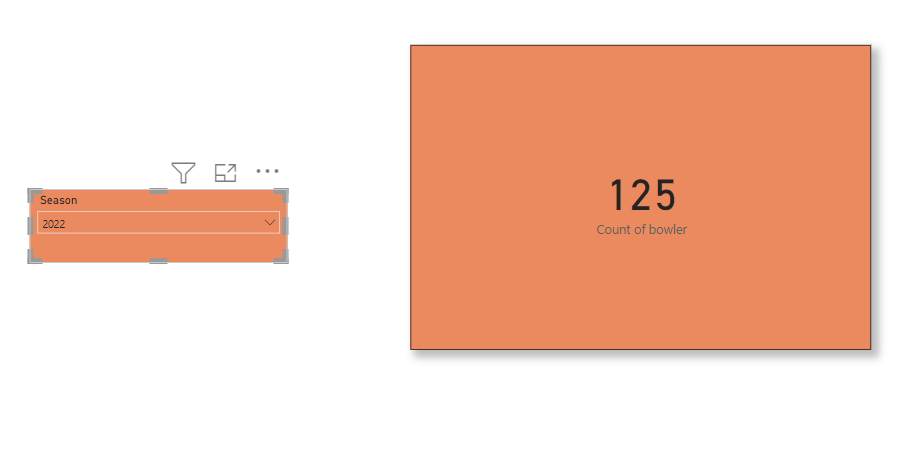


Figure : 3.12 Total no.of Bowlers in 2022

Graphical user interface, application, PowerPoint

Description automatically generated

Figure : 3.13 Total no. of batsmen in CSK

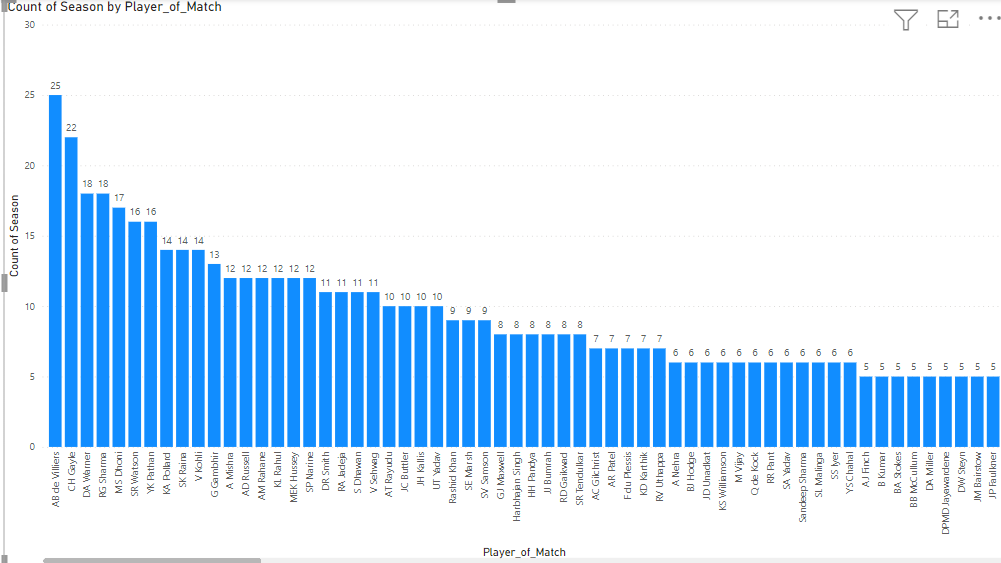


Figure : 3.14 Man of match by players

Graphical user interface, application

Description automatically generated

Figure : 3.15 number of catches taken by players

Graphical user interface, application, PowerPoint

Description automatically generated

Figure : 3.16 Total runs by Virat Kohli in 2008

Graphical user interface, application, PowerPoint

Description automatically generated

Figure : 3.17 Total wickets by DJ Bravo



Figure : 3.18 Extra Runs conceded by bowlers

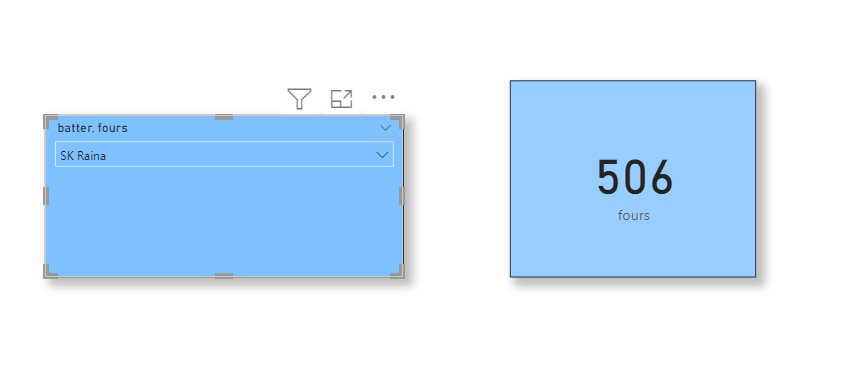


Figure : 3.19 Total number of boundaries by Raina

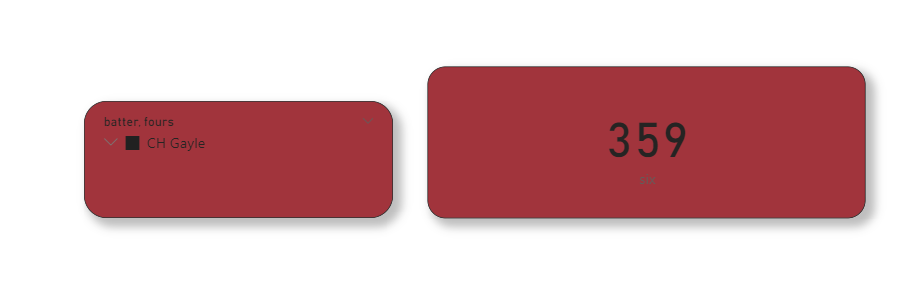


Figure : 3.19 number of sixers by CH Gail

## 3.2 PUBLISHING DASHBOARD

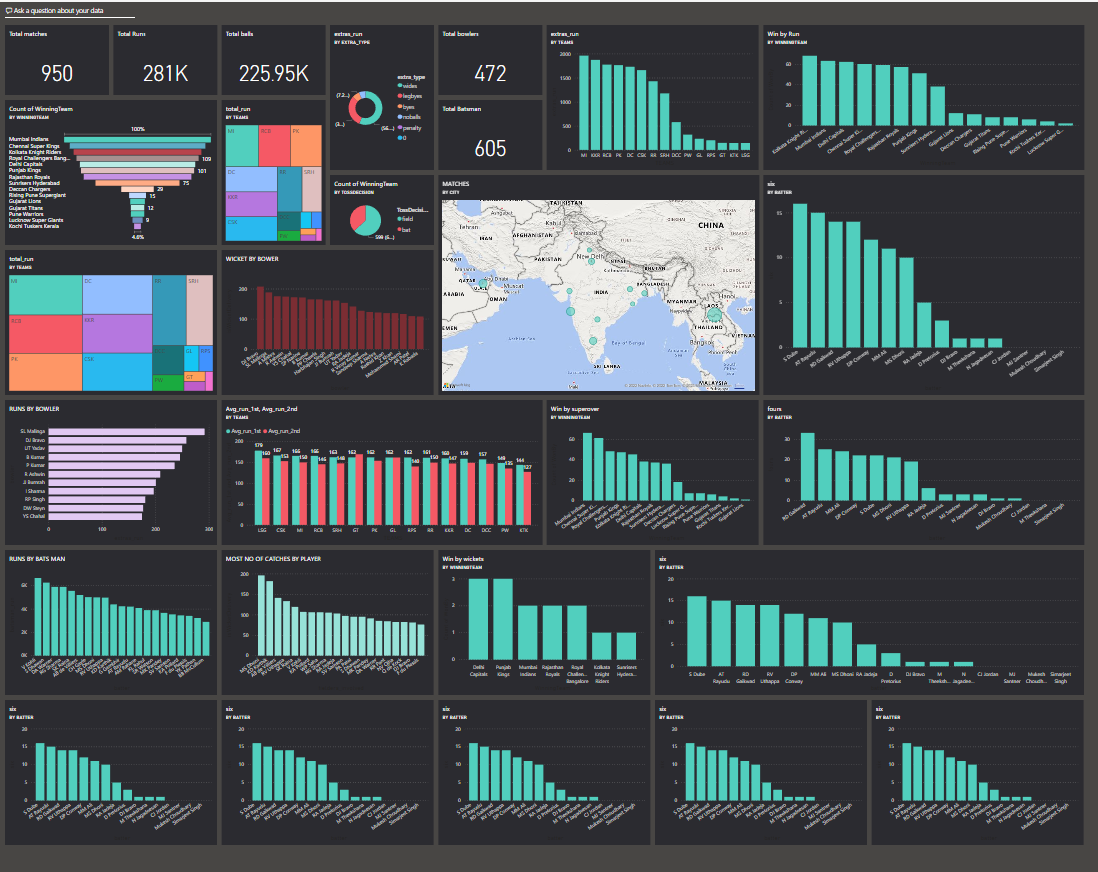


Figure 3.21 IPL Analysis Dashboard

## 3.3 INFERENCES :

## Team analysis

* The average run score by the teams in first innings is Lucknow super giants (LSG) is 179.
* The average run score by the teams in second innings is Gujarat titans(GT) 170.
* In total runs by teams Mumbai Indians (MI) scored the most runs is approx 37000.
* The total balls bowled by Delhi capitals in 2008 is 5988 balls.
* Total number of matches played by all team in between 2008 to 2022 is 56 matches.
* Number of matches won by Mumbai-Indians are 131 matches.
* The team conceded more extra runs is Mumbai Indians(MI).
* The toss decision impact the match winning in fielding.
* Wide Type of extra runs is more conceded is also about 56.05%
* Comparing CSK ,MI conceed more extra runs, extra type and average score on first and second innings.

**player analysis**

* Total number of players played in between 2008 to 2018 is 1077 players.
* Total number of bowlers played in 2022 is 125.
* The total number of batsman played for CSK in all seasons are 69 batsman.
* AB devilliers has got highest man of the match is of about 25 man of maches.
* Total runs scored by Virat Kohli in the year 2008 is 165 runs.
* The player takes more number of catches in ipl history is MS .Dhoni.
* The total wickets taken by the DJ bravo in 2022 is 17 wickets.
* The player who gives more extra runs is Malinga.
* The total number of fours scored by SK. Raina is 506 fours
* The total number of sixers scored by CH. Gayle is 359. %

The above dashboard represents the IPL data from the IPL beginning season to till now this is mostly helpful for the analysis of an individual player performance and the team’s performance and this analysis is mostly used for sponsors during IPL auction.

# CHAPTER 4

# CONCLUSION AND FUTURE WORK

**4.1 RECOMMENDATIONS**

The IPL teams with limited resources will undoubtedly want to know whether the player they are purchasing is worth the money they paid for their team during the player auction. Because it frequently occurs that the player who sells for the highest price at the IPL auction does not necessarily have the best performance in the league.

The clearest illustration of this would be the inaugural IPL season in 2008, when Rajasthan Royals won the championship despite being one of the least expensive teams. It indicates that their expenditure on players was significantly lower than that of other teams. The IPL was won by one of the least expensive teams.

In fact, the SAP Auction Analysis received a lot of praise when Kolkata Knight Riders won the title in 2014. In order to thoroughly explain "What sort of team should be constructed, which player should be sent, where, when, and what should the tactics be?," Kolkata Knight Riders had employed the SAP Data Analysis Company. Based on this study, KKR ultimately took home the prize.

But in the end, I'll just state that everything here is a game of probability; while considering all those factors will undoubtedly boost your odds of winning the game, no strategy will ever ensure victory. Because the IPL players are, after all, human people,

# REFERENCES

1. A. S. Bhat, V. S. Amit, N. S. Prasad, and D. M. Mohan, “An efficient classification algorithm for music mood detection in western and Hindi music using audio feature extraction,” in 2014 Fifth International Conference on Signal and Image Processing, Juju Island, 2014, pp. 359 –364
2. A. Lopes, E. De Aguiar, A. De Souza, et al, “Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and The Training Sample Order,” Pattern Recognition, vol.61, no.3, pp.610-628, 2017.

[3] Loh, C. S. and Sheng, “Measuring expert-performance for Serious Games Analytics: From data to insights,” Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement, [Chapter 5], New

York, NY: Springer, 2015.

[4] David H.S. Chung, et al., “Knowledge-Assisted Ranking: A Visual Analytic Application for Sports Event Data,” IEEE Computer Graphics and Applications, vol. 36, no. 3, 2016.

[5] Sricharan Shah, et al., “A Study on Performance of Cricket Players using Factor Analysis Approach,” International Journal of Advanced Re-search in Computer Science, vol. 8, no. 3, 2017.

[6] Van Staden, P. J., “Comparison of cricketers’ bowling and batting performances using graphical displays,” Current Science, vol. 96, no. 6, 2009.