

INTRODUCTION TO DATA MANAGEMENT (INT217)

Project Report on: -

Creation of Dashboard of Indian Primer League

Submitted By: -

Name: - Md Amjad Ansari

Reg No: - 12114768

Roll No: - 30

Section: - K21SR

Course Code: - INT217

Under the Guidance of: -

Faculty Name: - Maneet Kaur

Uid: - 15709

Discipline of CSE/IT

Lovely School of Computer Science and Engineering



LOVELY
PROFESSIONAL
UNIVERSITY

LOVELY PROFESSIONAL UNIVERSITY, PHAGWARA

CERTIFICATE

This is to certify that Md Amjad Ansari bearing Registration no. 12114768 has completed INT217 project titled, Creation of Dashboard of Indian Premier League (IPL) under my guidance and supervision of Maneet Kaur. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 06/11/2023

DECLARATION

I, Md Amjad Ansari student of Data Science under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 06/11/2023

Signature: - Amjad Ansari

Registration No: - 12114768

Name of the student: - Md Amjad Ansari

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my faculty Miss Maneet Kaur who gave me the golden opportunity to do this wonderful project of analysis of the data of a superstore namely “ Creation of Dashboard and analysis of Indian Premier Leage” which also helped me in doing a lot of research and I came to know about many new things . I am thankful to them. I would also like to thank my parents and my friends who helped me a lot in finalising the project within the limited time frame.

TABLE OF CONTENT

Sr No.	Content	Page No
1.	Introduction	
2.	Objectives / Scope of the Analysis	
3.	Source of the Dataset	
4.	ETL Process	
5.	Analysis on Dataset	
6.	List of Analysis with results	
7.	References	
8.	Bibliography	

1. INTRODUCTION

Data analysis in Excel is a powerful and widely used method for examining and making sense of data. Excel provides a range of tools and functions that can help us manipulate, visualize, and draw insights from the data. Here's an introduction to data analysis in Excel:

- **Data Input:** Start by entering data into an Excel spreadsheet. Each column should represent a variable, and each row should represent an individual data point. Ensure that your data is organized and clean, with clear headers.
- **Sorting and Filtering:** Excel allows us to sort and filter the data to easily identify patterns or specific data points. We can use the "Sort" and "Filter" functions in the "Data" tab to arrange and display the data in a desired order.
- **Data Visualization:** Excel provides various chart types, such as bar charts, line charts, scatter plots, and pie charts, which can help us visualize the data. We can select the data and then create a chart in the "Insert" tab.
- **PivotTables:** PivotTables are a powerful feature in Excel for summarizing and analysing large data sets. We can use them to create custom tables and analyse data based on different variables or criteria.
- **Data Cleaning:** Before performing any analysis, it's crucial to clean the data. This involves removing duplicates, handling missing values, and ensuring data consistency.
- **What-If Analysis:** Excel allows you to perform "What-If" analysis by changing values in your spreadsheet and observing how it affects the results. This can be done using tools like Goal Seek, Scenario Manager, and Data Tables.

- **Data Visualization and Dashboards:** Excel also allows us to create interactive dashboards using Pivot Charts and slicers. This is a great way to present our data analysis results in a user-friendly and dynamic format.

Excel is a versatile tool for data analysis, but it has its limitations, especially with very large datasets or complex statistical analyses. In such cases, you may need to use specialized software or programming languages like Python, R, or statistical packages like SPSS. However, for many basic and intermediate data analysis tasks, Excel is a valuable and accessible tool.

The analytics team of Indian Premier League (IPL) association and BCCI association anywhere in the world would love to check our through data analysis of each and every match leading to a well organised and faithful information. My analysis contains data on host teams, all stadiums, most supported teams which are most successful.

The Indian Premier League (IPL), is a professional 20 – 20- cricket league in India contested during April and May of every year by teams representing Indian cities and some states. The league was founded by the Board of control of cricket India (BCCI) in 2008, and is regarded as the brain child of Lalit Modi, the founder and former commissioner of league. IPL has an exclusive window in ICC future Tours Program.

The IPL is most-attended cricket league in the world and in 2014 ranked sixth by average attendance among all sports league. In 2010, the IPL became the first sporting event in the world to be broadcast live on YouTube.

Analysing Indian Premier League (IPL) data is an exciting and insightful endeavour for cricket enthusiasts and data analysts alike. The IPL is one of the most popular Twenty20 cricket leagues in the world, featuring top international and domestic players. Here's an introduction to IPL data analysis:

- **Data Sources:** To start your IPL data analysis, you need access to relevant data sources. IPL data is widely available on websites, APIs, and databases,

including details about matches, players, teams, scores, and more. You can also consider web scraping techniques to gather specific data.

- **Data Cleaning:** Raw data collected from various sources may be messy and inconsistent. Data cleaning involves removing duplicates, handling missing values, and ensuring data accuracy and consistency. This step is essential to ensure the quality of your analysis.
- **Visualizations:** Visualizations play a key role in understanding and presenting IPL data. Create plots and graphs to represent trends, patterns, and comparisons in the data. Common types of visualizations include bar charts for team performance, line charts for player statistics over seasons, and heatmaps to show match schedules.
- **Team Performance:** Analyse team performance by looking at factors like win-loss records, runs scored, wickets taken, and net run rate across seasons. Visualize team standings and playoff qualifications.
- **Match Results:** Analyse match results and trends, such as the impact of winning the toss, home advantage, and match outcomes on different pitches.
- **Dashboard Creation:** Create interactive dashboards using tools like Tableau or Power BI to present your analysis in a user-friendly and visually appealing way. Dashboards can provide a holistic view of IPL data and allow users to explore the data interactively.

IPL data analysis can offer valuable insights into the performance and trends within the league, which can be used by teams, broadcasters, and enthusiasts to make informed decisions or simply enjoy the game on a deeper level. Whether you're a cricket fan or a data analyst, IPL data analysis can be a fun and rewarding project.

2. OBJECTIVE / SCOPE OF ANALYSIS

The objective and scope of analysing an IPL dataset can vary depending on your specific goals and interests. This project on Indian Premier League(IPL) Statistics of India provides the overall Statistics details of the matches of IPL and teams progress in various aspects from the year 2008 to 2020. The Indian Premier League (IPL) is a professional men's Twenty20 cricket league in which 10 teams compete from ten different locations. Millions of people, especially Indians, are obsessed with the Indian Premier League (IPL), and our job involves data analysis and match prediction for IPL matches. In recent years, analytics has been used to predict and draw various insights in the field of sports. IPL Data Analysis is all about utilizing data science, machine learning to analyse data that is already existing in a data collection. This application design will be implemented for the purpose of analysing the IPL data by fetching different attributes and building a predictive model that could predict the score, batsmen run, predict the winner, overall performance of the team, and the players, head-to-head analysis.

- **Team Performance Analysis:** - Evaluate the performance of IPL teams over multiple seasons and we can use all the data in playing the game fantasy game like Dream 11, MPL, My11Circle etc.
- **Venue Analysis:** - During the IPL various teams collect data and analyse the different venues to get to know that how many teams won the match here while batting first and how many teams have won the game while bowling first.
- Analyse win-loss records, points earned, and playoff qualifications.
Identify key factors contributing to team success or failure.
- **Player performance analysis:** - Objective player performance analysis of IPL teams involves the use of statistical methods and data analysis techniques to

evaluate the performance of individual players in the Indian Premier League (IPL). This analysis helps teams make informed decisions related to team selection, strategy, and player acquisition.

1. **Data collection:** Gather data related to player statistics, match results, and other relevant information from reliable sources such as official IPL websites or APIs.
2. **Data preprocessing:** Clean and preprocess the collected data to remove any inconsistencies or errors. This may involve handling missing values, data normalization, and feature engineering.
3. **Data preprocessing:** Clean and preprocess the collected data to remove any inconsistencies or errors. This may involve handling missing values, data normalization, and feature engineering.
4. **Statistical analysis:** Apply statistical techniques to analyse the data and derive meaningful insights. This may involve calculating player averages, comparing performance metrics, identifying trends, and conducting hypothesis testing.
5. **Performance prediction:** Use machine learning algorithms or predictive models to forecast future player performance based on historical data. This can assist teams in making strategic decisions.

the actual implementation of player performance analysis involves a combination of various techniques and may require expertise in data analysis, statistics, and programming languages such as Python or R.

- **Matches Analysis:** - Analyse match outcomes, including factors like the impact of winning the toss and home-ground advantage. Investigate trends in

high-scoring matches, low-scoring matches, and close finishes. Explore the frequency of tied matches, super overs, and other exciting scenarios.

- **Stadium and Pitch Analysis:** - Investigate the performance of teams and players in different IPL venues and conditions. Analyse pitch behaviour and its influence on match outcomes and player statistics.
- **IPL Record and Milestones:** - Identify and highlight records and milestones achieved by teams, players, and the IPL itself. Celebrate extraordinary performances and significant moments in IPL history.

Objective of the projects:

- To good hand on excel.
- To use different features and get friendly with the excel.
- To learn the ETL process in the Excel prep.
- How to link one sheet to another and traverse between the different sheets.
- How to use Pivot table and pivot charts.
- Learn to make dashboard in excel.
- To make different types of graphs in excel.
- To learn how to fetch the data from other sources to excel in different formats.

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

3. SOURCE OF DATASET

IPL (Indian Premier League) datasets can be obtained from various sources, including official IPL websites, data providers, and cricket statistics databases.

I have used two different dataset one dataset is of IPL Title winner, Runner up, player of the match and player of the series. And another data is for players milestone, Group stage winner and looser which team won the toss on which venue and which team team won the maximum number of time and on which venue.

The Title winner dataset I collected from the official website of IPL, www.iplt20.com

At this website all the data related to IPL are available and we can fetch all the data easily.

And another data I collected from Kaggle www.kaggle.com Kaggle is a popular platform for sharing datasets and data science competitions. It often hosts IPL datasets, which can be freely accessed and downloaded for analysis.

Kaggle is a valuable resource for data enthusiasts and professionals, offering a platform to learn, practice, collaborate, and even earn recognition and rewards through data science competitions. Whether you're a beginner looking to learn or an experienced data scientist looking for challenges and opportunities, Kaggle has something to offer.

Apart from these many other data sources are available on the internet some of them are.

1. Sports data Aggregators.
2. GitHub.
3. Data Providers.
4. APIs.
5. Cricket Statistics Websites.

When using data from these sources, it's essential to check for data accuracy, reliability, and terms of use. Some sources may provide data under specific licensing conditions, so ensure you comply with their terms when using the data for analysis or

research. Additionally, always validate the data to confirm its completeness and correctness before conducting any analysis.

Insights od Data:

Indian Premier League (IPL) Analysis Contain the following data fields: -

- Season: - Ther year on which the following match was held.
- Date: - Contain date on which date match was held.
- Stadium: - On Which Stadium the match was held.
- City: - In which City match was held
- Toss-Winner: - Who was the Toss winner.
- Toss-decision: - What team decided after winning the toss.
- Winner: - Which team won the game.
- Win by run: - This shows by how many a team won the game.
- Win by wicket: - This shows by how many a team won by wickets.
- Player of the match: - Which Player won the player of the match award.
- Total runs: - How many total runs were scored during the fist inning.

4. ETL PROCESS

ETL stands for Extract, Transform, Load, and it refers to a process used in data integration and data warehousing. The ETL process is essential for collecting, cleaning, and preparing data from various sources for storage or analysis. ETL is a process in database usage to prepare data for analysis, especially in data warehousing.

Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purpose of analysis; Finally, data loading describe the insertion of data into the final target database such as an operational data store, a data mart or a data warehouse. A properly design ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate source can be used together, and finally delivers data in a presentation- ready format so that application developers can build application and end users can make decision.

Precisely, ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into data warehouse system.

1. Extract: -

- **Extract Data:** In this first step, data is extracted from one or more source systems. Source systems can include databases, applications, files, web services, APIs, and more.
- **Data Extraction:** Data is pulled from the source systems into a staging area, often referred to as the extraction layer. This staging area may be a temporary storage location or memory where data is held before transformation.
- **Full and Incremental Extraction:** ETL processes can perform either a full extraction, where all data is collected from the source, or an incremental extraction, where only new or modified data is collected since the last extraction. Incremental extraction is more efficient for large datasets.

2. Transform: -

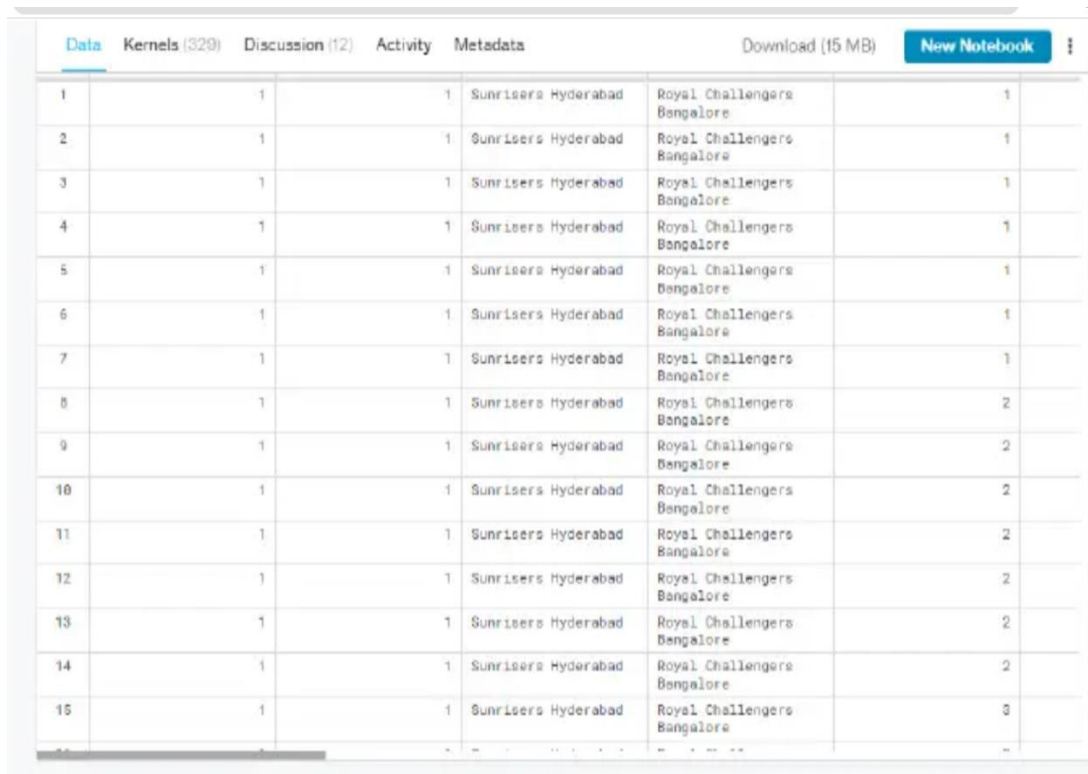
- **Data Transformation:** The extracted data is processed and transformed to meet the requirements of the target system. This step involves various operations, including.
- **Data Cleaning:** Removing or correcting errors and inconsistencies in the data.
- **Data Validation:** Checking for data quality and conformity.
- **Data Standardization:** Ensuring data is in a consistent format.
- **Data Enrichment:** Adding additional information to enhance the data.
- **Data Aggregation:** Combining and summarizing data as needed.
- **Data Filtering:** Removing unnecessary or irrelevant data.
- **Data Derivation:** Creating new data elements from existing ones.
- **Data Mapping:** Aligning data from different sources.
- **Business Rules:** Business rules and transformations are applied to ensure data is consistent and suitable for the target system.

3. Load: -

- **Data Loading:** Transformed data is loaded into the target destination, typically a data warehouse, database, or reporting system. The destination can be on-premises or in the cloud.

- **Loading Strategies:** ETL processes can employ different loading strategies, such as bulk loading, incremental loading, or real-time loading, depending on the needs of the business and the data volume.
- **Data Integrity:** Loading data with referential integrity and data consistency checks ensures that data in the target system remains accurate and reliable.

Before ETL, the dataset looked like this. **This data is taken from Kaggle.**



	Data	Kernels (329)	Discussion (12)	Activity	Metadata	Download (15 MB)	New Notebook
1		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
2		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
3		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
4		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
5		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
6		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
7		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	1
8		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
9		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
10		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
11		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
12		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
13		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
14		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	2
15		1		1	Sunrisers Hyderabad	Royal Challengers Bangalore	3

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

Step 1: Removing the Blank cells from the dataset.

For this select the whole dataset. Go to find and select in the “Home Tab” of excel. Select go to special from the drop-down menu and click tick and blank option. All the blank cells will be selected. Then go to delete option in the “Home Tab” again and

select Delete Rows from the drop-down menu. This will remove any rows with blank cells.

Select the fields to include in your flow. If you make changes to the data, the data source will be queried again.

<input checked="" type="checkbox"/>	Type	Field Name	Original Field Name	Changes	Sample Values
<input checked="" type="checkbox"/>	#	id	id		1, 2, 3
<input checked="" type="checkbox"/>	#	season	season		2,017
<input checked="" type="checkbox"/>	Abc	city	city		Hyderabad, Pune, Rajkot
<input checked="" type="checkbox"/>	Cal	date	date		05/04/2017, 06/04/2017, 07/04/2017
<input checked="" type="checkbox"/>	Abc	team1	team1		Sunrisers Hyderabad, Mumbai Indians, Gujarat Lions
<input checked="" type="checkbox"/>	Abc	team2	team2		Royal Challengers Bangalore, Rising Pune Supergiant,...
<input checked="" type="checkbox"/>	Abc	toss_winner	toss_winner		Royal Challengers Bangalore, Rising Pune Supergiant,...
<input checked="" type="checkbox"/>	Abc	toss_decision	toss_decision		field
<input checked="" type="checkbox"/>	Abc	result	result		normal
<input checked="" type="checkbox"/>	#	dl_applied	dl_applied		0
<input checked="" type="checkbox"/>	Abc	winner	winner		Sunrisers Hyderabad, Rising Pune Supergiant, Kolkat...
<input checked="" type="checkbox"/>	#	win_by_runs	win_by_runs		35, 0

Step 2: Removing Columns which are not properly defined or not crucial to our analysis.

For this we will select columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete options in the “Home tab” and then select Delete Columns from the drop-down menu.

Step 3: Giving proper and appropriate columns name.

The dataset does not have proper columns so our next step would be to give proper column names to the column wherever required.

Step 4: Excluding the NULL values from the data.

One approach is to use the BYROW function to identify all non-empty rows in the range and pass this result into the FILTER function as the include argument.

Step 5: Improving proper Data Formatting.

Without proper data formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format. For example, the date should be in date format and price and sales should be in Currency format for better results.

Step 6: Remove Duplicates values.

It might be possible that our data may be containing duplicates values which may hinder in precise analysis. So our last task in ETL will be removing duplicates values and making our data perfect for analysis.

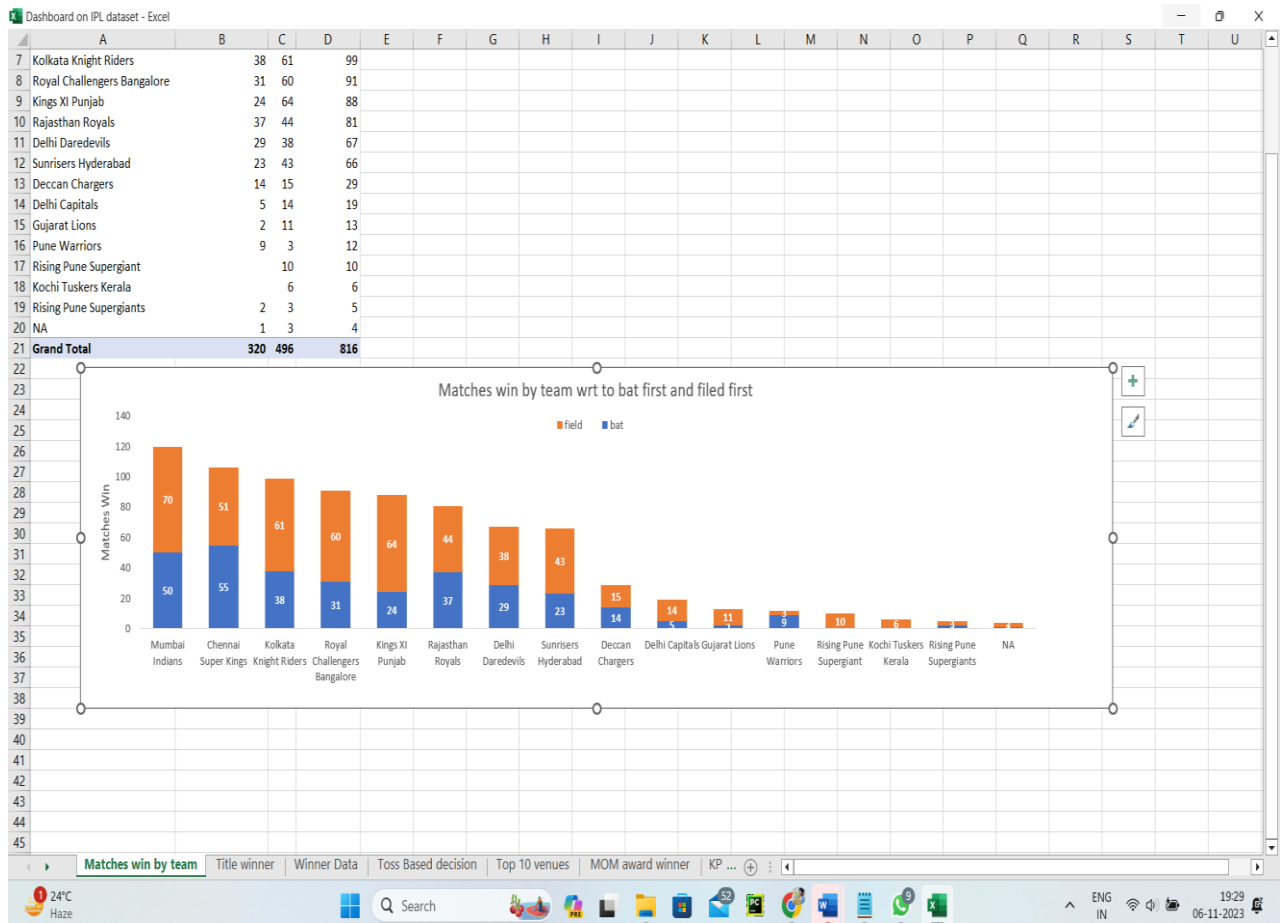
5. ANALYSIS OF DATASET:

➤ Matches win with respect to bat first and field first:

In this objective we will be finding the number of wins on the basis of Bat first and field first. Like how many teams have won the game while batting first and how many teams have won the game while chasing the scores i.e Field first.

To get these stats I have used pivot table to get the precise data on the basis of our requirements only. And with the help of pivot table I have analysed the data using Pivot charts.

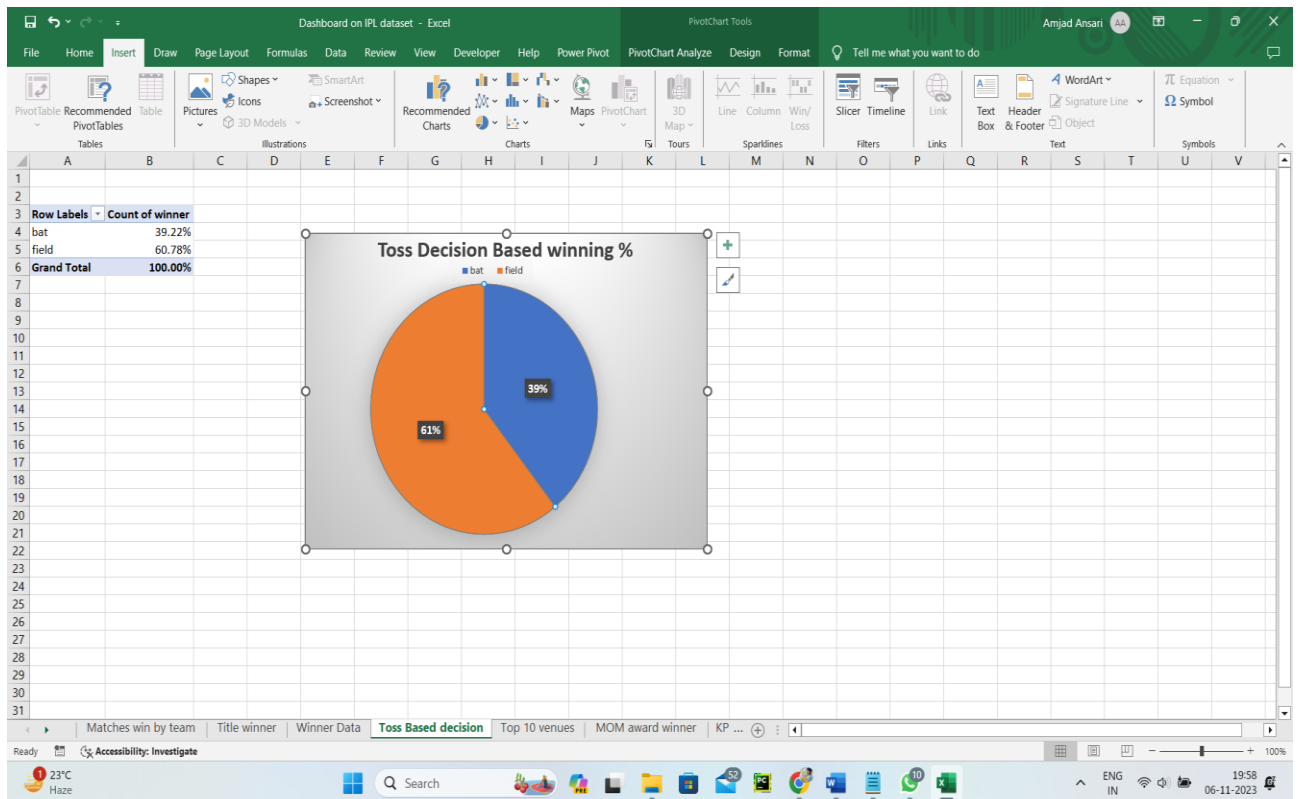
Using pivot table I have extracted all the data related to the matches win with respect to bat first and field first. And with the help of pivot table I got to know results of win and lose the game wrt to respective decision.



Using Stacked Column chart I have visualise the difference in Win and Lose percentage. In this stacked column chart Columns are denoting the number of wins and loses by any team and bar are denoting the teams.

2. Toss based decision:

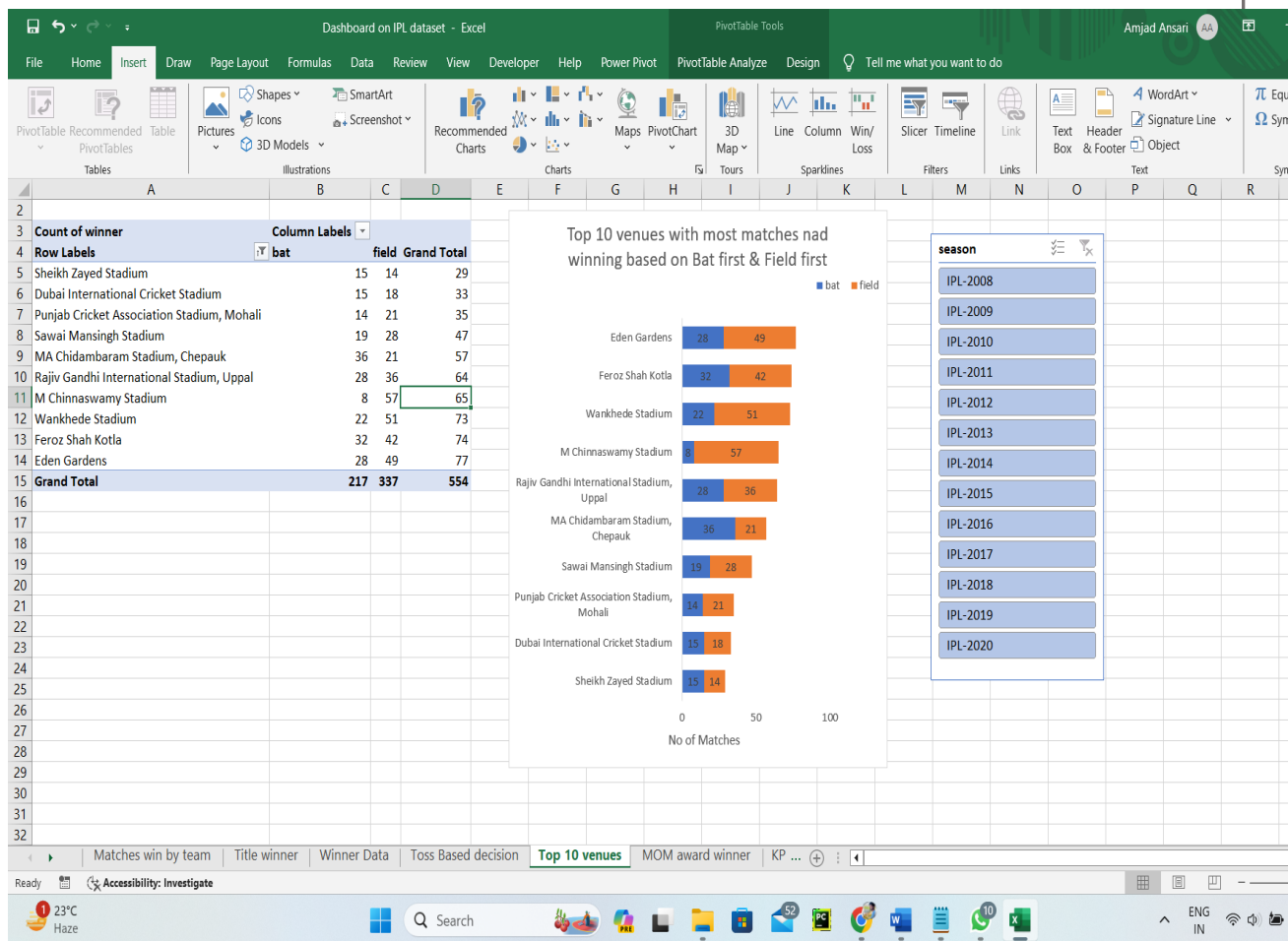
In toss-based decision we get to know about how many teams won the toss and as well as wo the game also.



In this Pie chart it is showing the Toss decision based winning percentage here till 2008 to 2020. 39% of the teams have won the toss and also won the games as well and 61% win come after loosing the toss. Means 61% of the time teams have lost the won but have won the game.

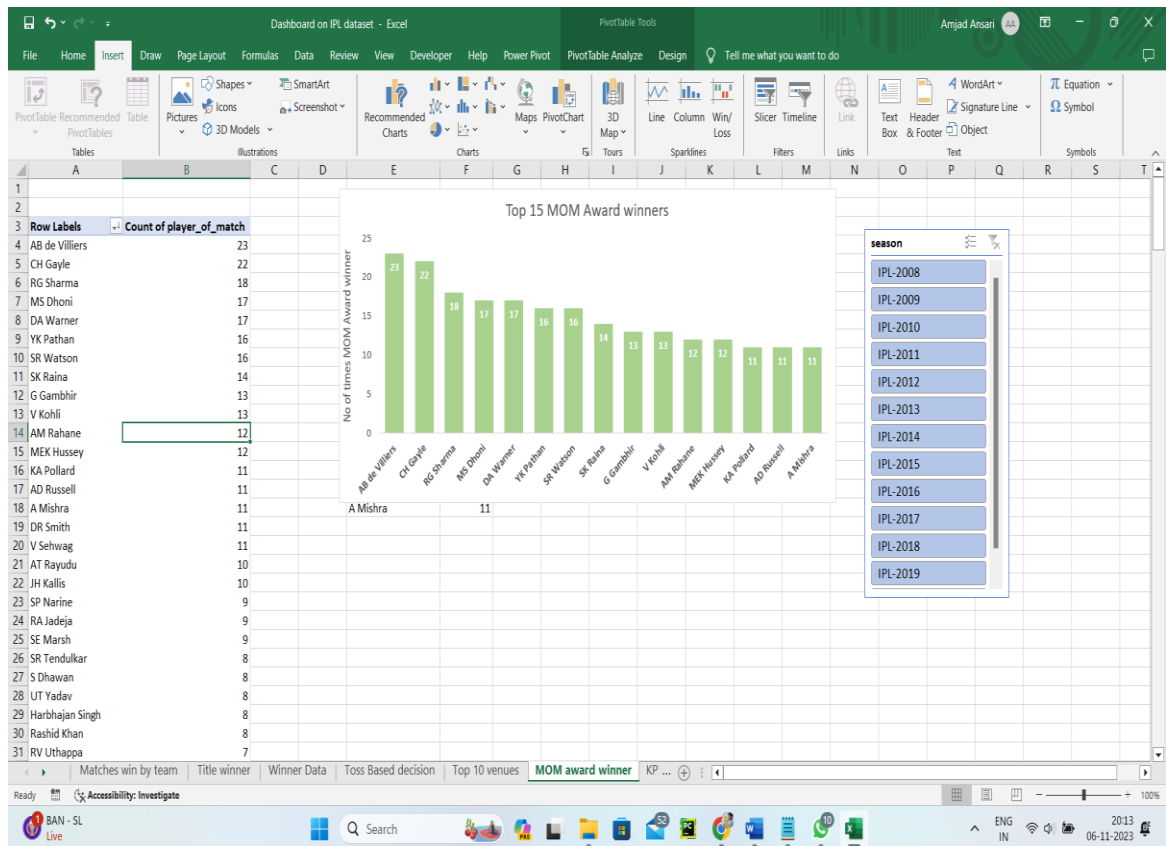
4. Top 10 venues with most matches played:

Here I have analyse the top 10 venues with most number matches have been played till 2020. And how many teams have won the game and how many teams have lost the game on that venue.

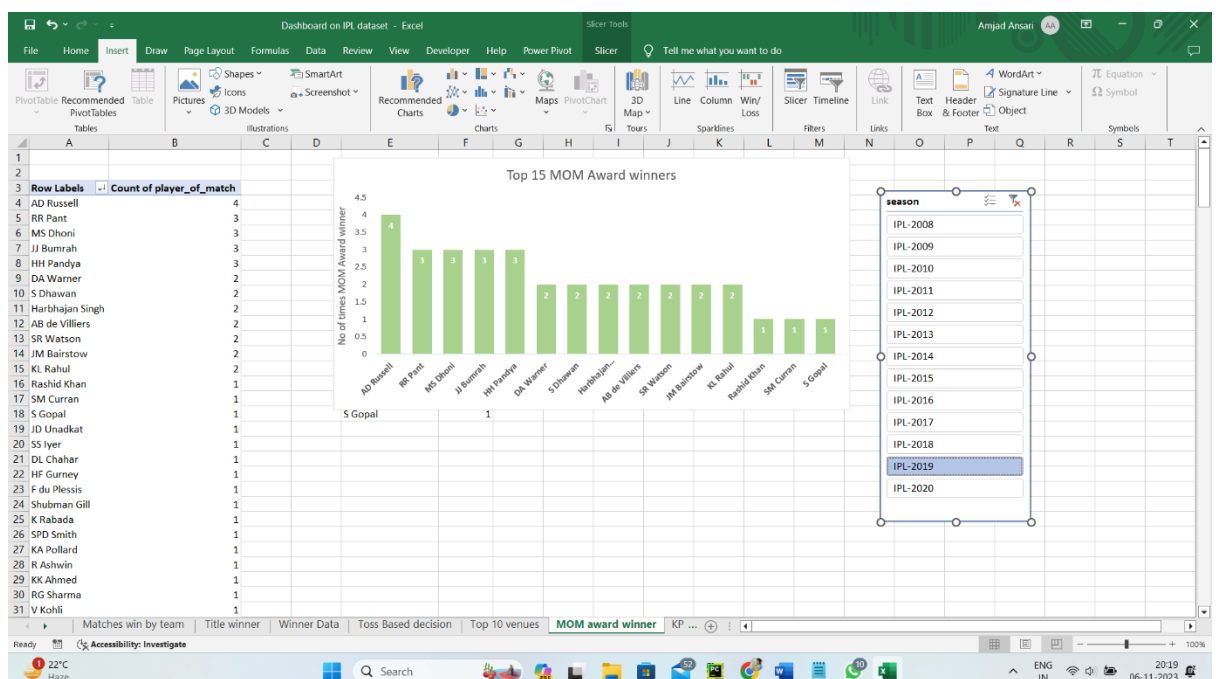


Here in this stacked bar chart I have analysed all the top 10 venue with more number of matches firstly I have created the pivot table and with the help of pivot table I have analysed the data using stacked bar chart in this analysis it clearly visible that on “Eden Garden” more number of matches have been played so far i.e 77 and out of those 28 times team have won while bating first and 49 team have won while chasing the score.

5. Player of the match award winner:



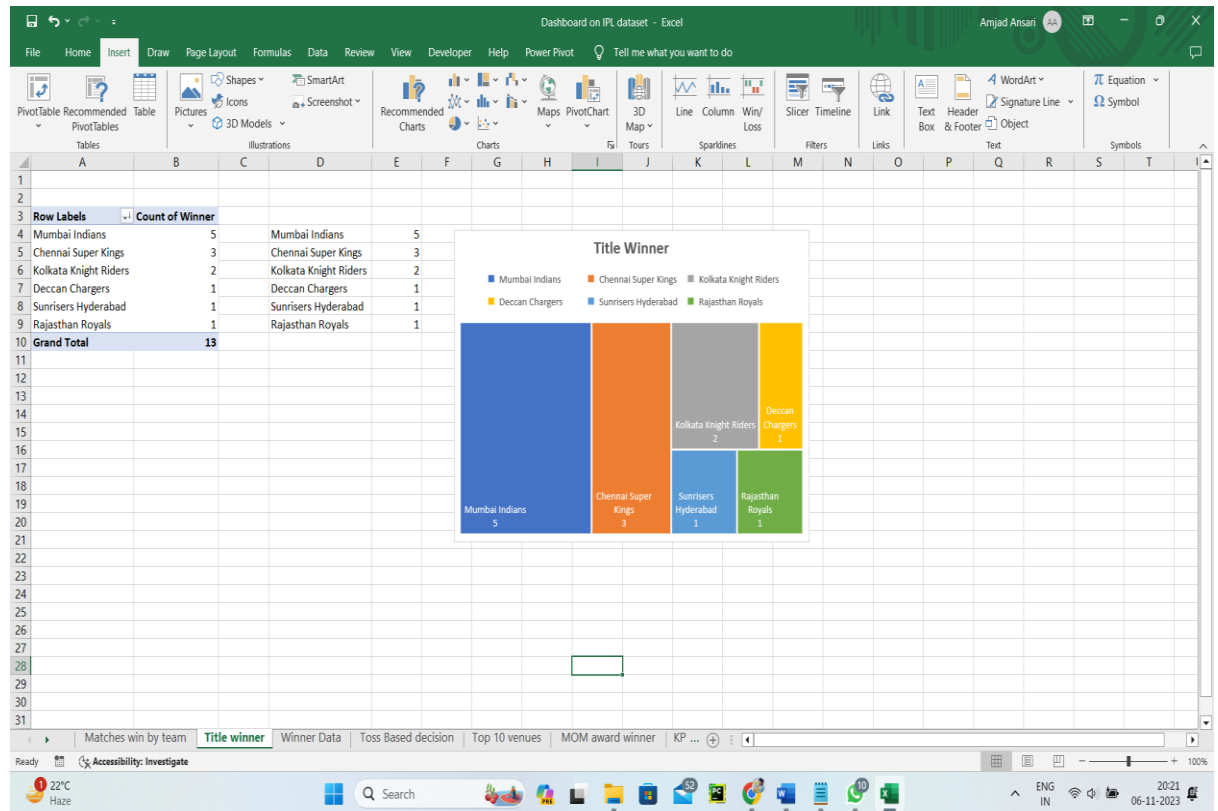
Here we can find which player have won more number of time player of the match and using slicer we can find for any specific year that which player have won the maximum number of player of the award in any specific calendar year like here we can see that from 2008 to 2020 AB de Villiers have won the maximum number of award.



Like here we can see that Anre Russel have won the more number of Player of the award in calendar year 2019. i.e 4 times.

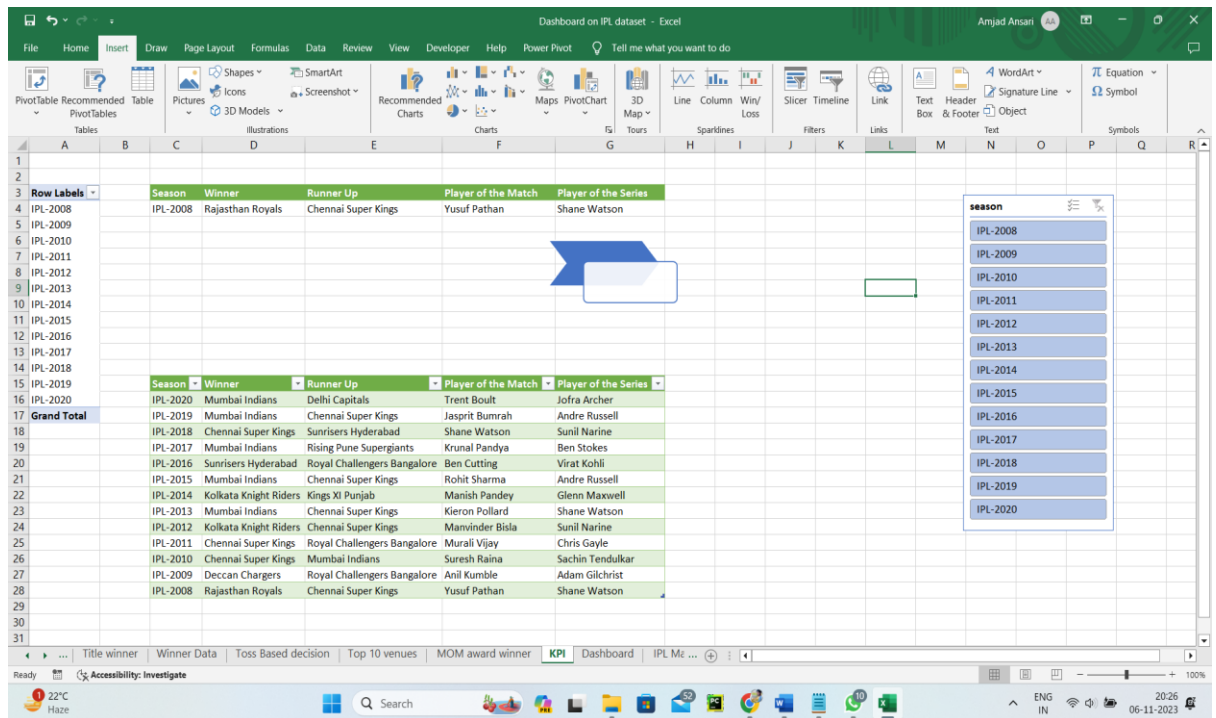
6. Title Winner:

Here we can find which team has won the more title or which team is more successful team in the IPL.



Here we can clearly find that Mumbai Indian have won more number of titles i.e 5 times followed by the Chennai super kings i.e 3 times these two teams have been consistently performing in the IPL and dominating the others team.

7. Designing of KPI:



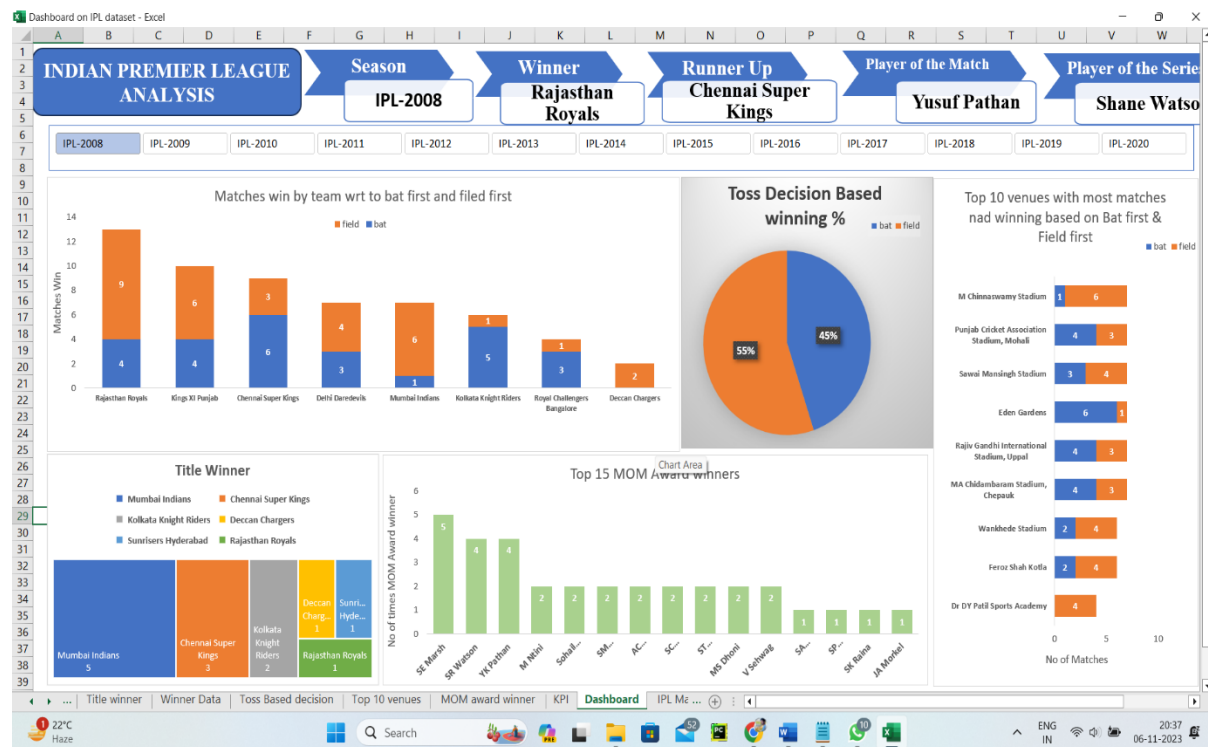
Here I have designed the KPIs and established all the relations.

8. Result Dashboard:

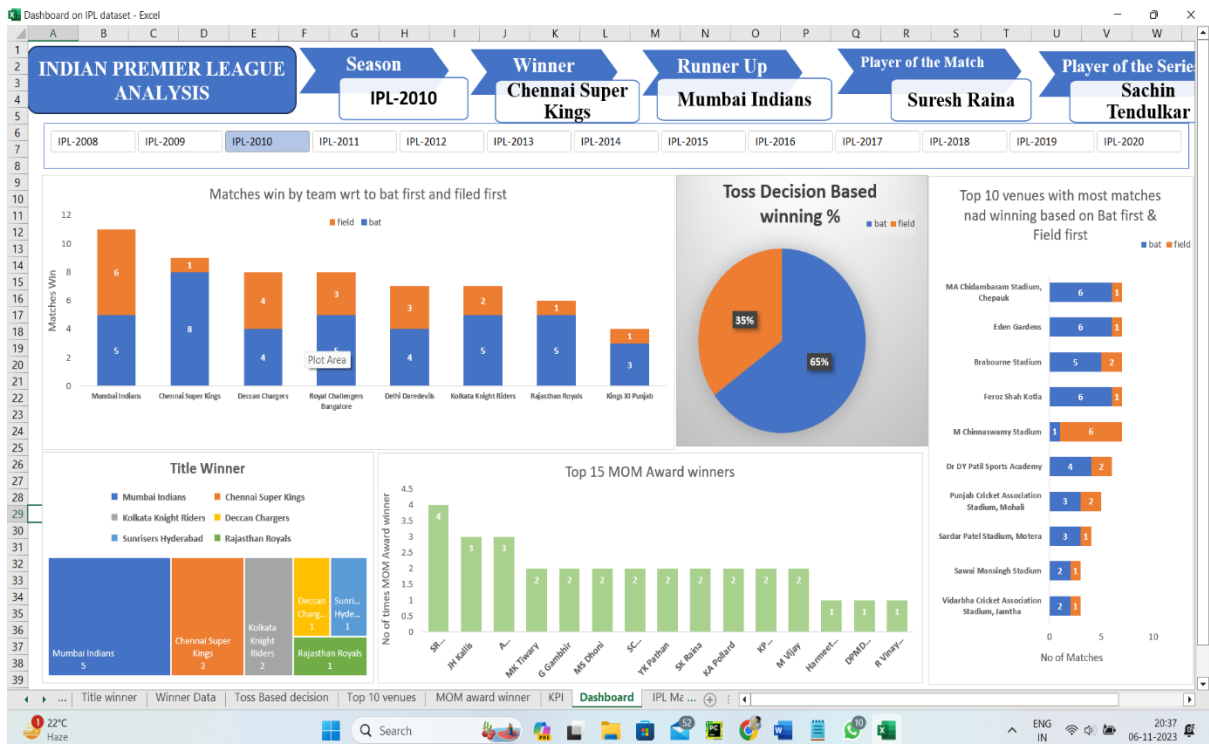
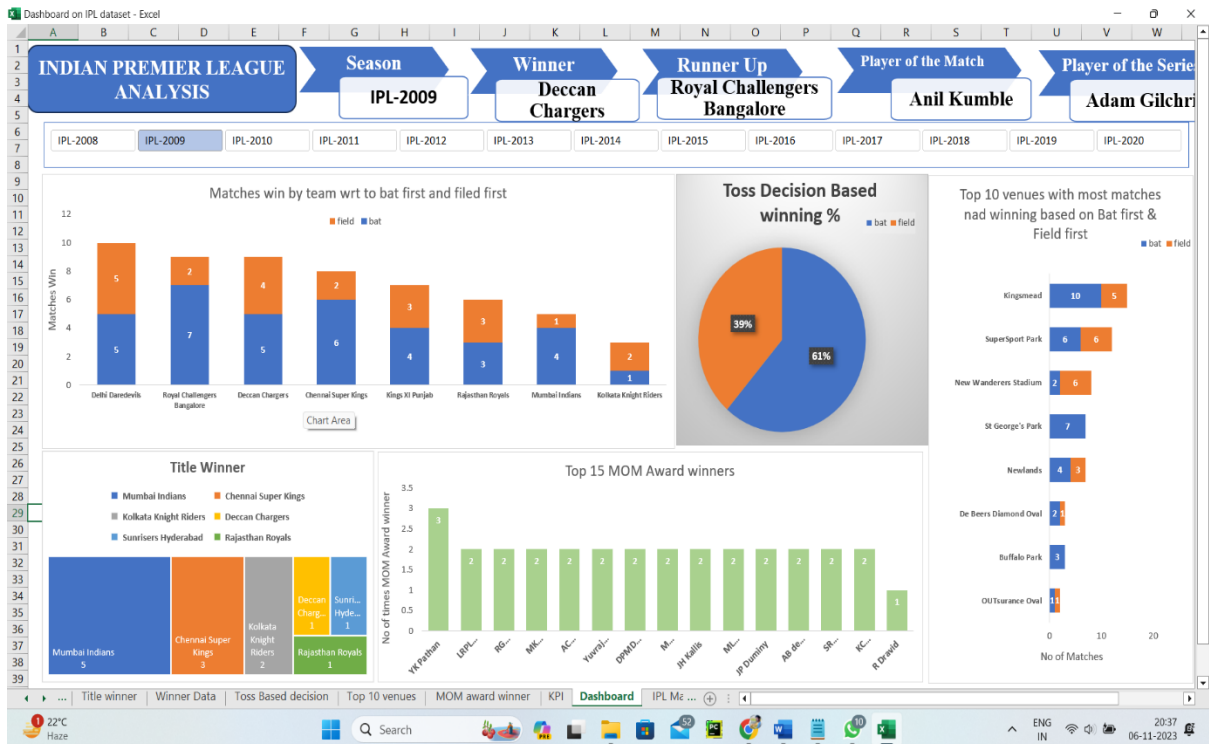


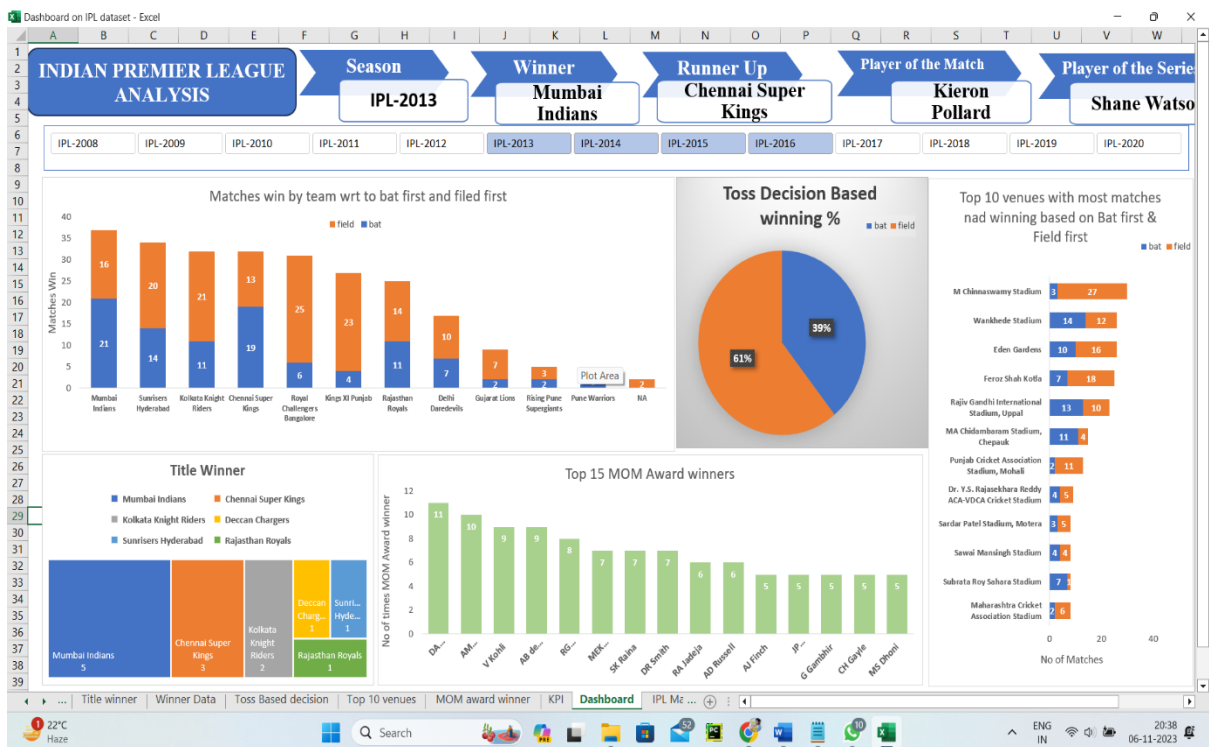
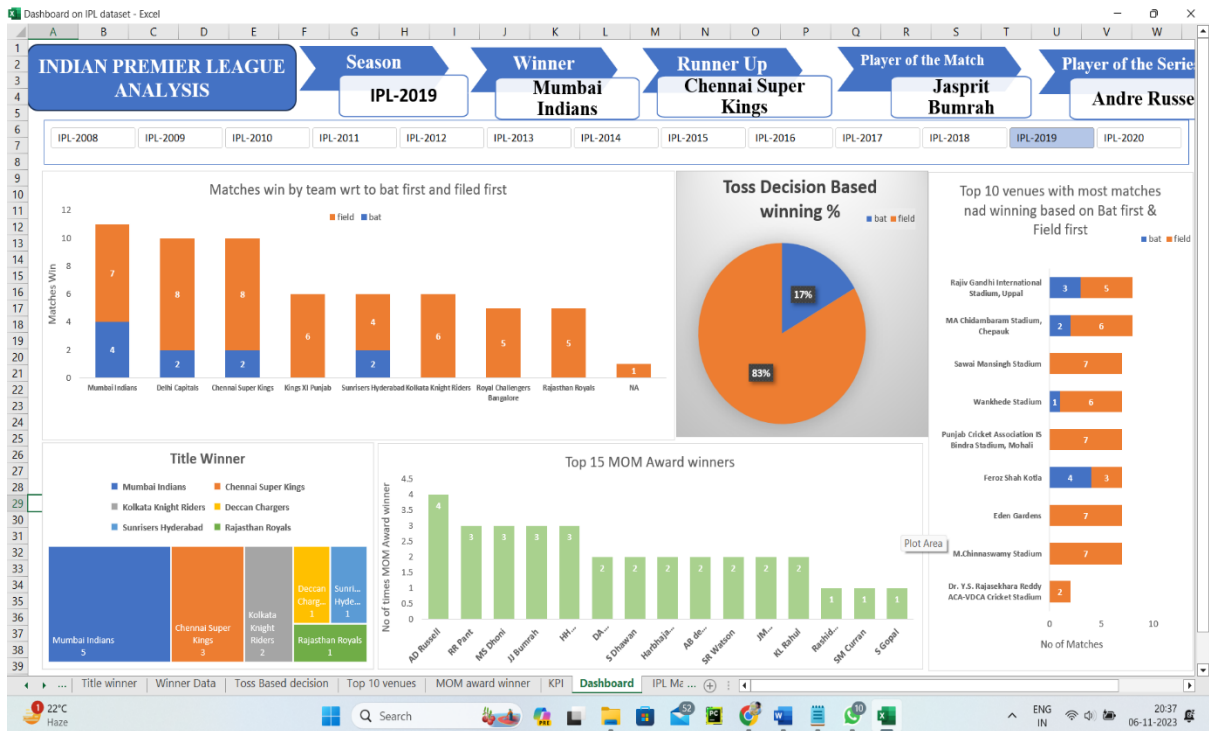
This is the final dashboard here we can manage the data of any particular year also.

We just need to change the value using slicer that I have added all the relations using slicer we can operate any particular year's data.



Here we can see only 2008 data like which team won the title in 2008 and which team was runner up and who was the player of the match and who won the player of the series in that year of campaign. Similarly we can see the data of 2009 also and so on.





Similarly if we want to see the data for multiple year that also we can see just by selecting the multiple years.

9. FUTURE SCOPE:

The Indian Premier league is one of the best entertaining tournament of the world. Every year of IPL come up with new launching style.

IPL to me looks a very good option as far as providing the exposure to the young lot is concerned having played with the biggies of the games like of Sachin, Kumble, Ponting, Virat etc. still hanging around with the team. As far as the skills is concerned, it does not dig deep into skill-testing especially for the batsmen, bowlers can still show their intelligence and skill by sticking to basics and containing the batter.

But with IPL the major issue is the crowd, since it takes place in major Indian cities which are already hosts of test matches and some one-days it provides the crowd the option to choose between matches. So this decreases the attendance in the stadium during the IPL since the crowd wants to see more or tough competitive cricket rather than friendly ventures.

And coming to the futures, it has great scope and potential and with BCCI taking care of it so well it does not seem the trophy is gonna lose any shine in near future.

10. Conclusion

Analysing IPL data is a fascinating exercise that provides valuable insights into the performance of players and teams. The data provides useful information for developing effective game strategies, evaluating player performance, engaging fans, and identifying business opportunities.

- **Key Findings:** Summarize the main findings of your analysis. Highlight significant trends, patterns, and insights that you've uncovered. This could include observations about team performance, player statistics, match outcomes, or any other areas you explored.
- **Statistical Significance:** Discuss the statistical significance of your findings. If you performed hypothesis testing or statistical analysis, mention the results and whether they are statistically significant.
- **Recommendations:** If applicable, provide recommendations or insights based on your analysis. These could be actionable suggestions for teams,

players, or IPL organizers. For example, you might suggest strategies for improving team performance or player recruitment.

- **Visualization Highlights:** Showcase key visualizations that effectively convey your findings. Visual representations, such as charts and graphs, are often the most effective way to communicate data-driven insights.
- **Limitations:** Acknowledge any limitations of your analysis. Every analysis has constraints, whether related to data quality, availability, or the analytical approach used. Being transparent about limitations is essential for a thorough and honest conclusion.
- **Future Research:** Suggest areas for future research or analysis. Identify questions or aspects of IPL data that were not covered in your analysis but could be explored in subsequent studies.
- **Contribution to Knowledge:** Summarize how your analysis contributes to the broader knowledge about the IPL, cricket analytics, or sports data analysis.

11. REFERENCES AND BIBLIOGRAPHY:

- YouTube
- Analytics Vidya
- Kaggle
- Wikkipedia.com
- Iplt20.com
- Google.com

Thank You