

Tech Saksham

Case Study Report

Data Analytics with Power BI

“ IPL Analysis using Power BI ”

“GOVERNMENT ARTS AND SCIENCE COLLEGE FOR WOMEN,PULIAKULAM-45”

NM ID	NAME
2AD85129AEEE96F81F54460780854633	G CHITHRA

Trainer Name :UMAMAHESWARI.R

Master Trainer: UMAMAHESWARI.R

ABSTRACT

The Indian Premier League (IPL) stands as a pinnacle of Twenty20 cricket, captivating audiences worldwide since its inception in 2008. This project delves into the rich dataset encompassing IPL matches from 2008 to 2022, focusing on key variables such as team compositions, player statistics, match outcomes, venues, and more. The primary objective is to extract valuable insights through data cleaning, transformation, and analysis using Power BI.

The project commences with meticulous data cleaning procedures, addressing null values, duplicates, and inconsistencies to ensure data integrity. Subsequently, columns are split, merged, and formatted to facilitate efficient analysis. Filtering mechanisms are employed to isolate specific subsets of data, such as matches within certain date ranges or involving particular teams.

Power BI's robust data transformation capabilities are harnessed to add custom columns, enabling the creation of calculated measures for deeper analytical depth. Grouping and aggregation techniques are applied to distill trends and patterns from the dataset, unveiling performance metrics of teams, players, and venues across IPL seasons.

Furthermore, the project explores the dynamic world of IPL through pivot and unpivot operations, providing versatile perspectives on match statistics and player performances. The visualization prowess of Power BI is leveraged to craft intuitive dashboards and reports, offering stakeholders a comprehensive view of IPL insights.

Through this project, a thorough analysis of IPL matches emerges, shedding light on team dynamics, player contributions, winning strategies, and venue influences. The findings not only serve to enhance predictive capabilities for future IPL seasons but also pave the way for informed recommendations and conclusive insights into one of cricket's most captivating spectacles, the Indian Premier League.

INDEX

Sr. No.	Table of Contents	Page No.
1	Chapter 1: Introduction	1
2	Chapter 2: Services and Tools Required	4
3	Chapter 3: Project Architecture	6
4	Chapter 4: Modeling and Result	8
5	Conclusion	11
6	Future Scope	12
7	References	13
8	Links	14

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The Indian Premier League (IPL) stands as a cornerstone in the world of cricket, captivating audiences with its blend of talent, strategy, and entertainment since its inception in 2008. To delve deeper into the dynamics of this premier T20 league, this project aims to analyze and derive insights from a comprehensive dataset encompassing IPL matches from 2008 to 2022.

1.2 Proposed Solution

The project aims to utilize Power BI's robust data transformation and visualization capabilities to analyze and derive insights from a comprehensive dataset of Indian Premier League (IPL) matches spanning from 2008 to 2022. The IPL, a premier Twenty20 cricket league in India, has garnered immense popularity globally since its inception, making it a rich ground for in-depth analysis.

1.3 Features

- **Team Performance Comparison:** Comparative analysis of teams based on wins, losses, net run rates, and batting/bowling performances.
- **Player Insights:** Individual player statistics, including highest run-scorers, leading wicket-takers, strike rates, and player-of-the-match awards.
- **Match Trends:** Examination of trends such as toss-winning impact, home ground advantage, and performance in knockout matches.
- **Venue Analysis:** Evaluation of venue-wise performance, average scores, and winning patterns.
- **Predictive Modeling (Optional):** For advanced users, the project can include predictive modeling using DAX expressions to forecast match outcomes based on historical data.

1.4 Advantages

1. **Grouping and Aggregation:** Power BI simplifies grouping data based on specific columns and performing aggregations like sum, average, count, etc. This is crucial for data analysis. **Efficient Data Cleaning:** Power BI's Query Editor allows you to clean your data efficiently. You can easily remove null values, duplicates, or inconsistent data that might skew your analysis. This ensures that your insights are based on accurate and reliable information.
2. **Flexible Data Manipulation:** With Power BI, you have the flexibility to split and merge columns as needed. This is particularly useful when dealing with complex datasets like IPL matches, where you may want to separate player names, match outcomes, or team statistics into distinct columns for better analysis.
3. **Data Formatting:** Power BI enables easy formatting of various data types, such as dates, times, or text. This ensures uniformity in your dataset, making it easier to perform calculations and comparisons across different matches, teams, or time periods.
4. **Custom Column Creation:** Power BI allows you to create custom columns based on calculated expressions. This means you can derive new insights by adding columns that represent specific metrics or calculations relevant to IPL analysis. For example, you could create a column for "winning streaks" or "average runs per match" to delve deeper into team performances.
5. **Advanced Data Filtering:** Power BI's filtering capabilities are robust, allowing you to filter data based on specific criteria. Whether you want to analyze matches within a certain date range, compare performance between specific teams, or focus on particular venues, Power BI makes it easy to filter the data accordingly.

1.5 Scope

The project aims to leverage the capabilities of Power BI to conduct a comprehensive analysis of the Indian Premier League (IPL) matches spanning from 2008 to 2022. The dataset for this analysis encompasses crucial variables including IPL teams, team players, toss outcomes, batting performances, bowling statistics, venues (stadiums), head-to-head matches, and various winning scenarios.

The primary focus of this project is to clean the dataset by addressing null values, duplicates, and inconsistencies. This will ensure that the data is accurate and reliable for subsequent analysis. Cleaning operations may involve removing incomplete entries, replacing missing values, and standardizing data formats.

Following data cleaning, the project will involve splitting and merging columns as needed to facilitate better analysis. This step could involve breaking down composite fields into individual attributes for easier processing and understanding. Additionally, the formatting of data types such as dates, times, and text will be standardized for consistency.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

Query Editor: This powerful tool within Power BI allows you to clean, transform, and manipulate your data before loading it into your data model. You can use Query Editor to remove null values, duplicates, or inconsistent data. This is crucial for ensuring the data quality of your IPL dataset.

DAX Expressions: Data Analysis Expressions (DAX) is a formula language that allows you to create custom calculations in Power BI. You can use DAX expressions to add custom columns based on calculated expressions. For example, you might want to create a column calculating the average runs per match for each team, or the win percentage of a team based on matches played.

Power Query: Power Query is another essential tool in Power BI for data transformation. You can split and merge columns using Power Query, which might be useful if you need to separate player names into first and last names or merge columns containing match details. Power Query also allows you to format data, such as changing the format of date and time columns.

Filtering Data: Power BI provides easy-to-use filtering options, allowing you to filter data based on specific criteria. For instance, you can filter matches based on a particular season, venue, or team. This helps in focusing your analysis on specific subsets of the IPL data.

2.2 Tools and Software used

Tools:

- **PowerBI:** The main tool for this project is PowerBI, which will be used to create interactive dashboards for real-time data visualization.
- **Power Query:** This is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources.

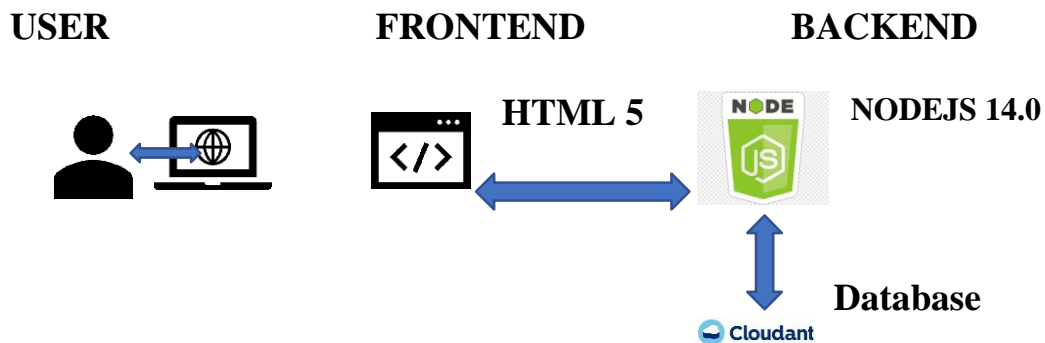
Software Requirements:

- **PowerBI Desktop:** This is a Windows application that you can use to create reports and publish them to PowerBI.
- **PowerBI Service:** This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.
- **PowerBI Mobile:** This is a mobile application that you can use to access your reports and dashboards on the go.

CHAPTER 3

PROJECT ARCHITECTURE

3.1 Architecture



Here's a high-level architecture for the project:

1. **Data Collection:** For my IPL project analysis, I collected a comprehensive dataset spanning from 2008 to 2022, focusing on various crucial variables essential for insights into IPL teams' performance. Through Power BI's versatile data transformation options such as Query Editor, DAX expressions, and Power Query. This meticulous data preparation sets the foundation for meaningful IPL team analysis, aiding in predictions, recommendations, and conclusive insights into the league's trends and outcomes over the years.
2. **Data Storage:** For your IPL data project, it's recommended to use Power BI for data storage and analysis. Power BI provides a robust platform with its Query Editor, DAX expressions, and Power Query for data transformation and cleaning tasks like removing null values, splitting columns, formatting data types, filtering based on criteria, adding custom columns, grouping, aggregating, and pivoting data.
3. **Data Processing:** The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
4. **Machine Learning:** Predictive models are built based on processed data using Azure Machine Learning or AWS SageMaker. These models can help in predicting player stats, performance, etc.

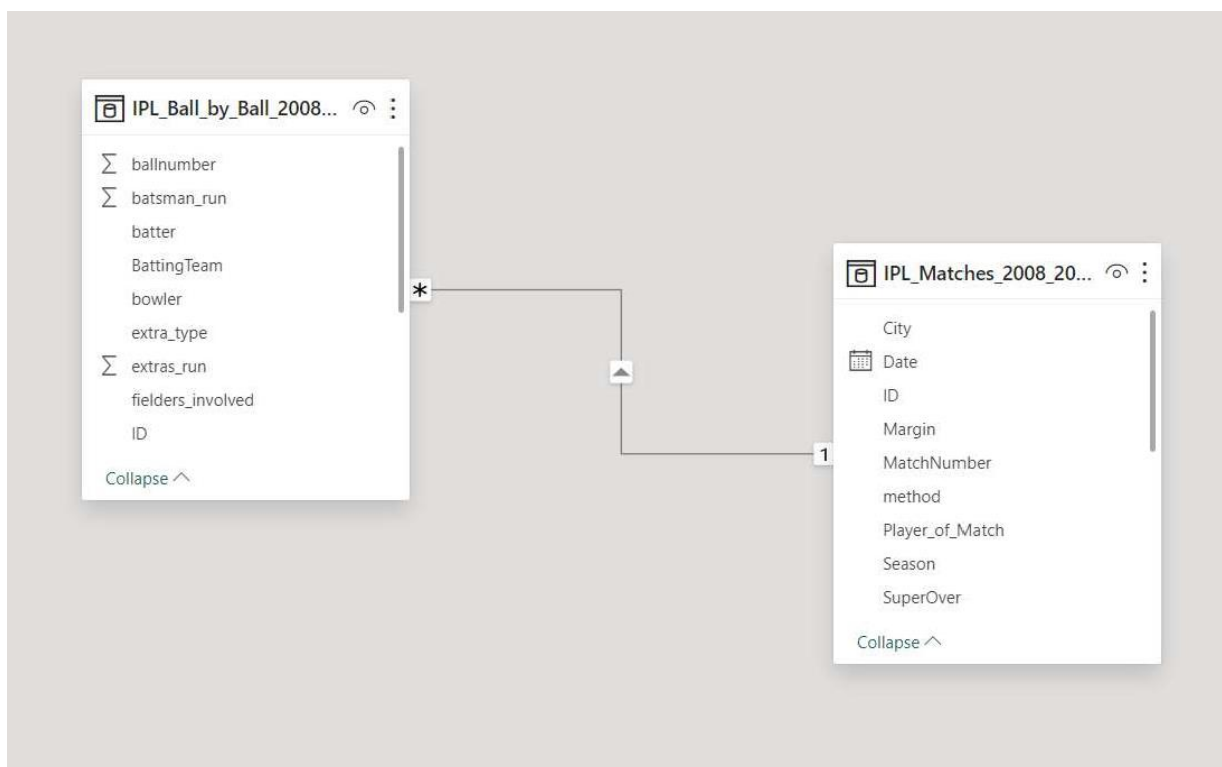
5. **Data Visualization:** The processed data and the results from the predictive models are visualized in real-time using PowerBI. PowerBI allows you to create interactive stats that can provide valuable insights into the data.
6. **Data Access:** The dashboards created in PowerBI can be accessed through PowerBI Desktop, PowerBI Service (online), and PowerBI Mobile.

CHAPTER 4

MODELING AND RESULT

Manage relationship

The “IPL Ball by Ball 2008-2022” file will be used as the main connector as it contains most key identifier (ballnumber, batsman_run, batter) which can be use to relates the 8 data files together.



DAX Functions

Create table:

```
calender1 table = CALENDAR(MIN(OUT[Date]),MAX(OUT[Date]))
```

Measure:

AVERAGE by bowler =

```
DIVIDE(SUMX(FILTER('IN','IN'[extra_type]<>"legbyes"&&'IN'[extra_type]<>"byes"),'IN'[total_run]),SUM('IN'[isWicketDelivery]))
```

```
Batter Runs = CONCATENATE(SUM('IN'[batsman_run] ), " Runs")
```

```
Bowler SR = COUNT('IN'[bowler])/SUM('IN'[isWicketDelivery])
```

```
Bowler Wicket = CONCATENATE(SUM('IN'[isWicketDelivery]),"Wicket")
```

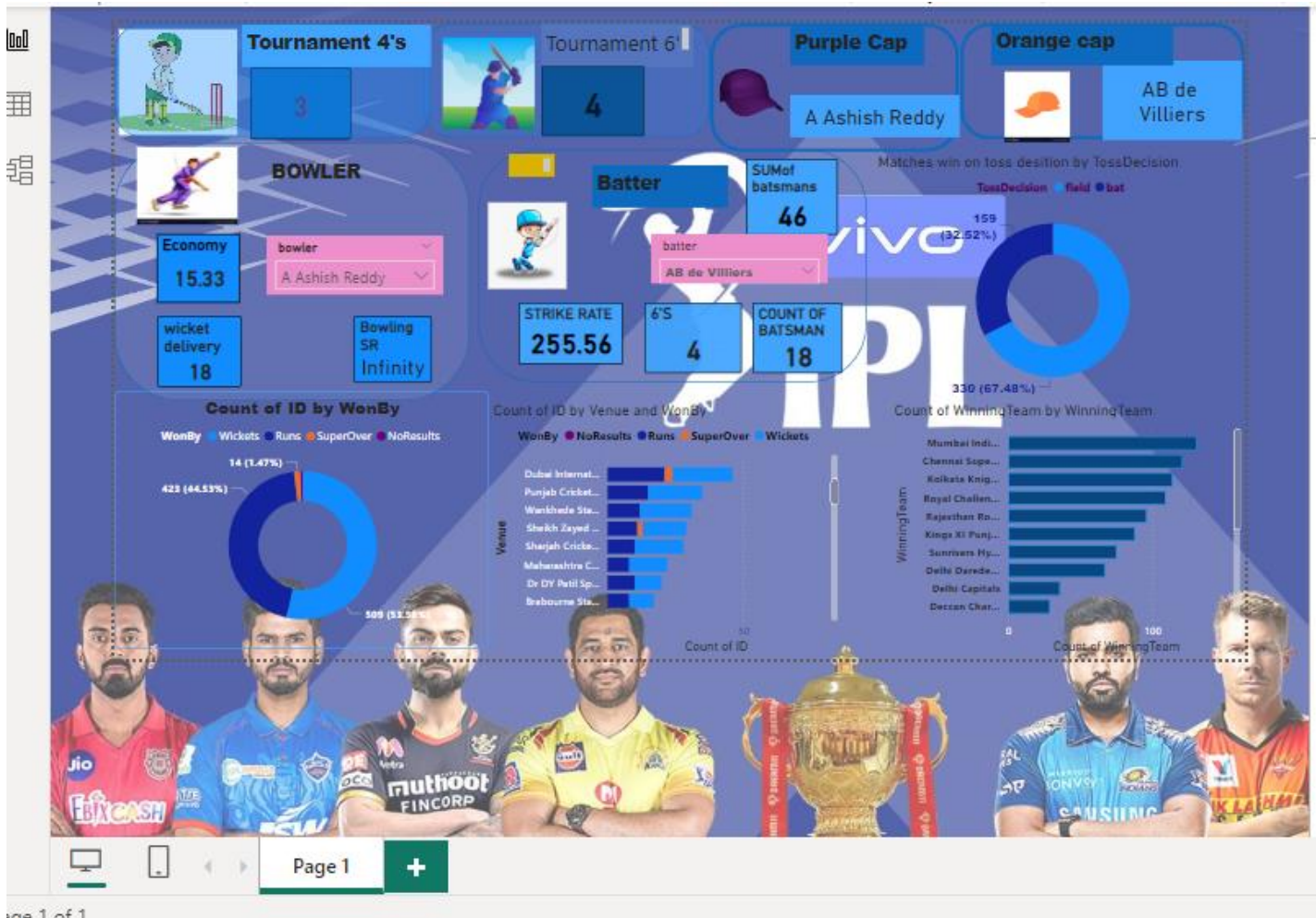
```
Economy = DIVIDE(SUMX(FILTER('IN','IN'[extra_type]<>"legbyes"&&'IN'[extra_type]<>"byes"),'IN'[total_run]),COUNT('IN'[overs])/6)
```

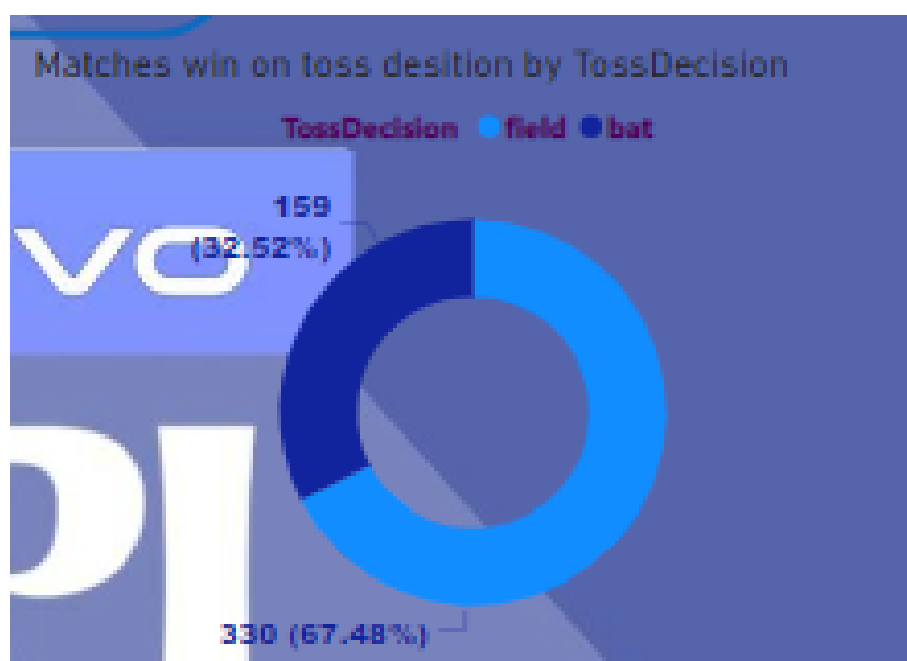
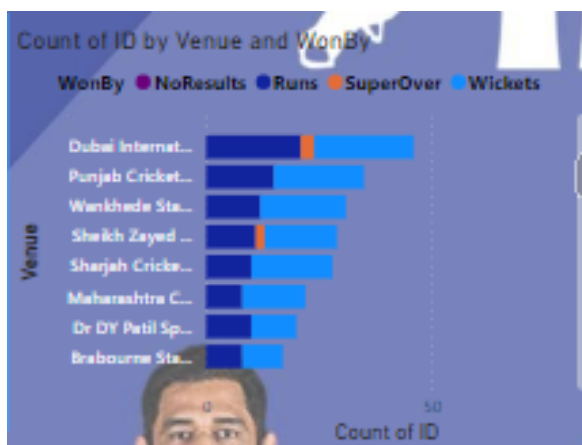
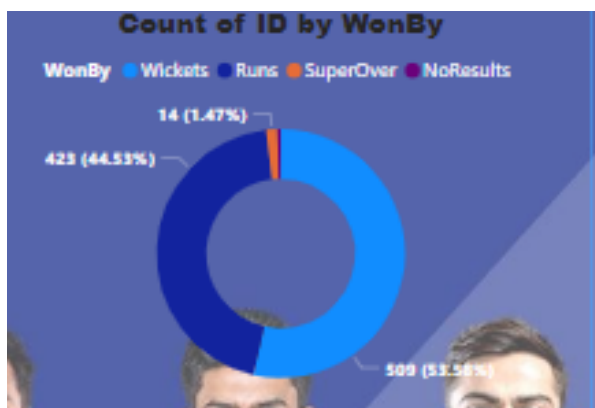
```
Strike rate = SUM('IN'[batsman_run])/COUNT('IN'[ballnumber])*100
```

```
Matches win on toss desition = CALCULATE(COUNTROWS(OUT),OUT[TossWinner]=OUT[WinningTeam])
```

```
Title winner = VAR max_date=CALCULATE(MAX('calender1 table'[Date]),ALLSELECTED(OUT),VALUES(OUT))  
var_tile_winner  
=CALCULATE(SELECTEDVALUE(OUT[WinningTeam]),'calender1 table'[Date]=max_date)  
RETURN var_title_winner
```

Dashboard





CONCLUSION

The IPL dataset analysis through Power BI has provided valuable insights into team performances, player contributions, and match dynamics over the tournament's history. These insights can serve as a foundation for informed decision-making by teams, coaches, and stakeholders. As the league continues to evolve, leveraging data analytics will be crucial for staying competitive and engaging fans in the exciting world of IPL cricket.

FUTURE SCOPE

The future scope of this project is vast. With the advent of advanced analytics and machine learning, PowerBI can be leveraged to predict future trends based on historical data. Integrating these predictive analytics into the project could enable the bank to anticipate customer needs and proactively offer solutions. Furthermore, PowerBI's capability to integrate with various data sources opens up the possibility of incorporating more diverse datasets for a more holistic view of customers. As data privacy and security become increasingly important, future iterations of this project should focus on implementing robust data governance strategies. This would ensure the secure handling of sensitive customer data while complying with data protection regulations. Additionally, the project could explore the integration of real-time data streams to provide even more timely and relevant insights. This could potentially transform the way banks interact with their customers, leading to improved customer satisfaction and loyalty.

REFERENCES

https://youtu.be/kfTVbJb9yIE?si=UiR47QFpm35B_Pw7

LINK

- <https://learn.microsoft.com/en-us/power-bi/report-server/quickstart-create-powerbi-report>
- <https://learn.microsoft.com/en-us/power-bi/create-reports/>