

IPL Data Analysis

Joshi Atharva, Sadamwar Akash, Wani Tejas

I. PROBLEM

Our project centers on deriving actionable insights from an extensive analysis of Indian Premier League (IPL). The goal is to unveil key trends, patterns, and performance metrics related to teams, players, and match outcomes, thereby providing valuable information for strategic decision-making by IPL teams, sponsors, and stakeholders.

However, we encountered several challenges throughout the project. Firstly, dealing with the vast and unstructured nature of the IPL dataset proved non-trivial, requiring intricate handling of data to extract meaningful insights. Implementing a query with NoSQL from scratch was particularly challenging in this context.

Additionally, the task of creating a web interface posed its own set of challenges. One notable complexity was determining the preferable strategy between bowling and batting based on pitch conditions through the analysis of historical data. Developing a robust framework for such analysis and integrating it into a user-friendly web interface required a meticulous approach.

These challenges, though demanding, provided valuable learning experiences, pushing us to develop innovative solutions and refine our skills in data analysis and interface design. The detailed insights obtained from our analysis contribute significantly to the overarching objective of supporting informed decision-making in the context of the IPL.

The N+1 data/analysis queries are given below:

1. Is it advantageous for the toss winner to be the match winner?
2. Overall performance of various teams by analyzing their past matches.
3. Predict match outcomes by assessing the number of wickets lost during the powerplay (first 6 overs).
4. Determining if bowling is preferable or batting based on the pitch conditions by analyzing the previous data.

We used public data set from Kaggle [[link](#)].

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Software Design and NoSQL-Database and Tools Used

In this project we have used MongoDB as our NoSQL database and to integrate MongoDB with Python to implement queries and plot the outcomes on the User Interface we have used the tool PyMongo. Libraries such as Matplotlib and plotly have also been utilized to illustrate the statistical results. Along with this, we have taken help of the Flask framework to display result on webpages using HTML and CSS.

B. Parts that we have implemented

List of four tasks which we have implemented in our project.

1. In this task, we have displayed output in the form of pie-chart where we have shown percentage of teams which have won the toss wins the match and percentage toss losing team wins the team. So by this data we can analyze that whether it is advantageous for the toss winner to be the match winner. We presented the output using a pie chart to illustrate the percentage of teams that won the toss and subsequently won the match, as well as the percentage of teams that lost the toss but still emerged victorious. This analysis helps us determine the advantage, if any, for the toss-winning team in becoming the overall match winner. The visual representation of this data provides a clear perspective on the correlation between winning the toss and securing a match victory.
2. In the second task, we have presented two graphs illustrating key performance indicators. The first graph depicts the Net Run Rate (NRR) of the team over the period (2008-2023) while the second one showcases the trend in team bowling economy rates over the years. Through the analysis of these metrics, we gain valuable insights into the overall performance of the team. Evaluating both NRR and economy rate provides a comprehensive view of the team's effectiveness and consistency in the IPL from 2008 to 2023.
3. In Task three, we have illustrated the correlation between the number of wickets lost in the power play

(X-axis) and total wins (Y-axis) using a bar graph. The data analysis reveals a strong association between a higher number of power play wickets and increased winning probability. This insight, derived from comprehensive historical match data, underscores the strategic importance of early wickets for teams.

4. In task four, we displayed output in the form of a pie-chart. Here we have shown two pie charts: 1st one shows the percentage of teams selected to field first after winning the toss and winning the match also the percentage of teams selected to field first after winning the toss and losing the match in the stadium. 2nd one shows the percentage of teams selected to bat first after winning the toss and winning the match as the percentage of teams selected to bat first after winning the toss and losing the match in a particular stadium.

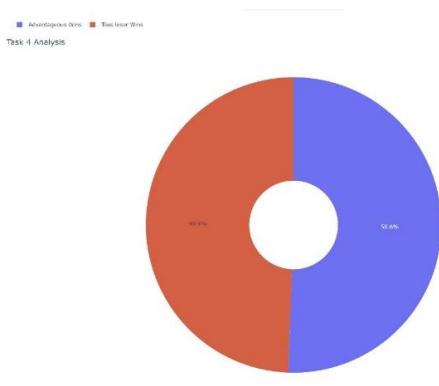
So, by analyzing this data, we can easily determine if bowling is preferable or batting.

III. PROJECT OUTCOME

Below are the outcomes of the tasks which we have performed in this project.

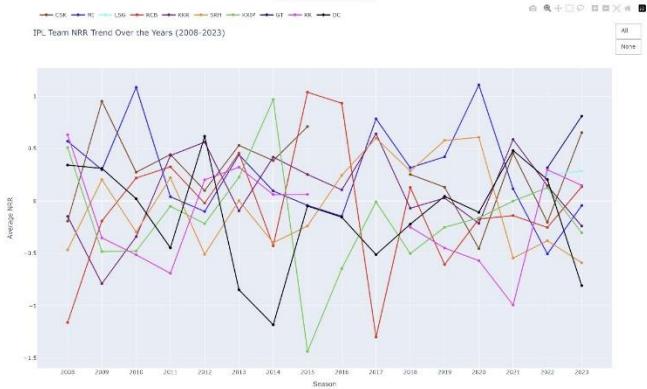
Outcome of Task 1:

A pie chart representation that depicts the chances of winning of a team which wins the toss and further it decides to field or bat.



Outcome of Task 2:

The trend of Net Run Rate over the years 2008 to 2023 for all the IPL teams.

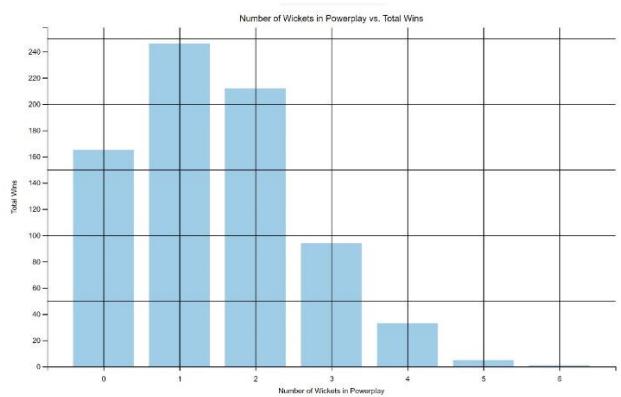


The trend of bowling economy over the years 2008 to 2023 for all the IPL teams.



Outcome of Task 3:

A bar chart visualization of the chances of winning of IPL teams depending upon the number of wickets lost in the Powerplay of the innings (first six overs).



Outcome of Task 4:

This illustration shows the probability of IPL team's chances of winning based upon their decision of choosing to Bat/Bowl. Two pie charts depicting two different statistics of decision of field/bat and further the result of match.

REFERENCES

- [1] What is MongoDB and how does it work?
<https://www.techtarget.com/searchdatamanagement/definition/MongoDB>
- [2] Flask Framework
<https://flask.palletsprojects.com/en/3.0.x/>

- [3] Matplotlib Guide:
https://matplotlib.org/stable/plot_types/index.html

