Super Predictor of Indian Premier League (IPL)

Team : THRESHERS

1. Introduction

1.1. Overview

The project concentrates on finding the solution to the problem statement by creating a dashboard that visualizes the eleven capabilities and also forecast the future result using the ipl data sets given by IBM .

1.2. Purpose

The purpose of Super Predictor of IPL is to utilize the data and provide the best players and strategy for winning the game by team management as IPL is a game for players but business for team franchise or management , so winning the match with there best strategy can be achieved by this super predictor , the Super Predictor of IPL can also prove beneficial for the team managements in the player auctions for selecting the right team and by visualizing the data and showing the probability of the team winning or predicting the targeted score in the sports channel for gaining the interests of the audience and creating a spice of entertainment and knowledge for the vast fans of IPL.

2. Literature survey

2.1. Existing problem

The existing approaches to solve this problem include various softwares to prepare , visualize , explore the data and then make predictions using several machine learning algorithms . The existing approaches use softwares like Pandas library to prepare the data in the process of data collection and then using parsing to extract data from the HTML page, and once after the data is prepared , the collected data will be visualized through graphical and statistical analysis with the help of another software named Tableau , then the prediction using KNN algorithm ,which will be compared with other machine learning algorithms like decision tree, logistic regression ,random forest etc for better accuracy .Thus by using many softwares the data might get corrupt or duplicate values and the Business Intelligence used in any existing approaches doesn't have cloud benefits and doesn't contain enough graphical and statistical analyzing options to explore the data to its maximum ability which makes the predictions less accurate.

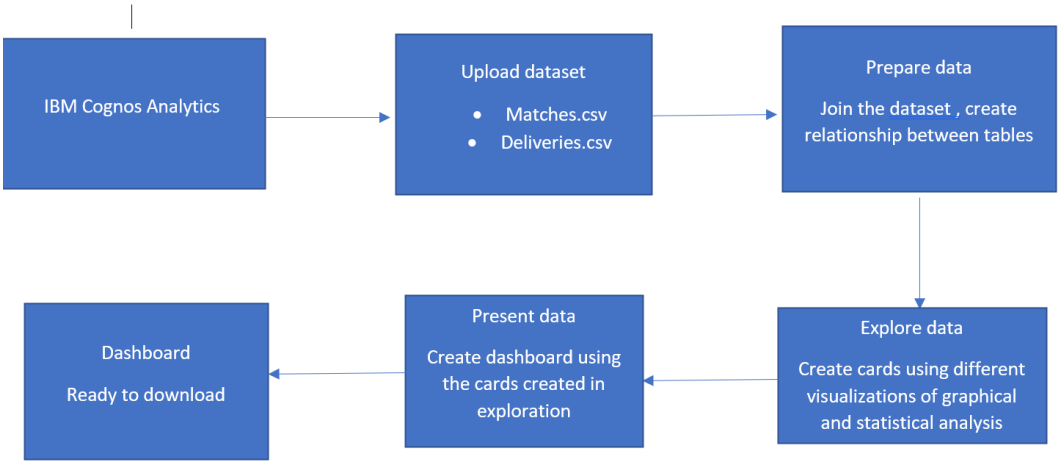
2.2. Proposed solution

The solution suggested for the problem is by using the Business Intelligence software offered by IBM namely IBM Cognos Analytics a cloud based application which converts the raw data into meaningful data through some process or architecture . Business Intelligence is

a method of collecting , sorting , analyzing business options , the ultimate goal is to drive business decisions , increase revenue , improve operational efficiency and gain competitive advantages over other teams or business rivals . So through IBM Cognos Analytics application that is using a single application we can get the proposed solution by first uploading the data set then by preparing the data by removing null values etc ,then by exploring the data through visualizations of several graphical and statistical analysis , in this process we will explore the data to find the solutions to the eleven capabilities by identifying trends in the data . Then we present the data through multi-query dashboard presenting the solutions to the eleven queries.

3. Theoritical Analysis

3.1 Block diagram



3.2 Hardware / Software designing

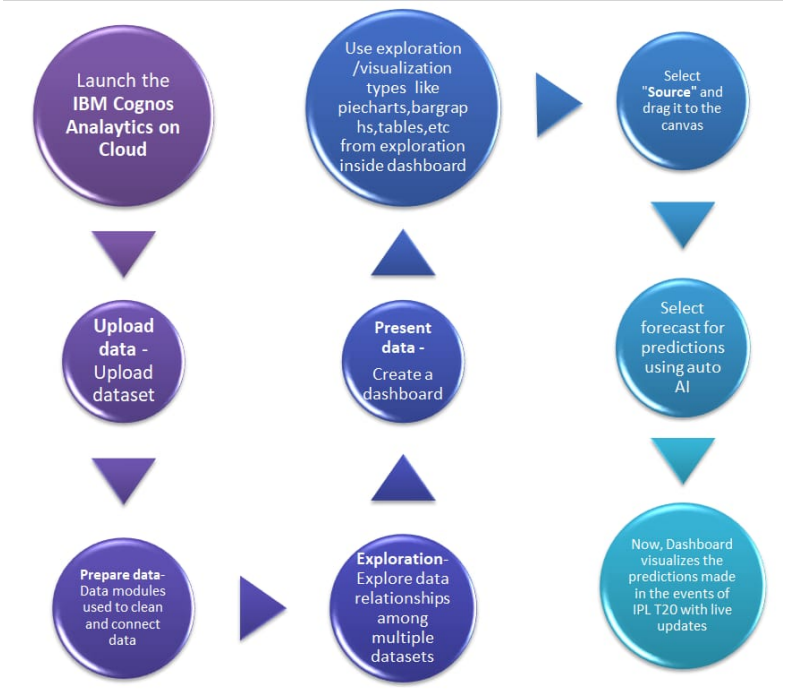
The only software required to this project is IBM Cognos Analytics which is a cloud based application which provides the required services to prepare , analyse , explore and present the data . The reason for using a single software is because of the features of the IBM Cognos Analytics offering in-memory streaming analytics , provides real time events ,alerts and notifications , user can edit existing data , drill through capability , potential Image Documentation Integration , it is also platform independent ,scalable and reliable .The components in this software are cognos connection , cognos business insight , cognos query stuio , cognos analysis studio , cognos report studio , cognos event studio , workspace , workspace advanced .The important types of cognos reporting tool are content store to store app data , content manager which helps to manage storage and retrieval of data , cognos content database that is server with all files , datasets , information ,everything stored in

content database .

4. Experimental Investigations

The analysis and the investigations made while working on the solution are going through the IBM offered courses such as Getting started on Enterprise AI , Cloud , Data Science then attending the Bootcamp offered by SmartInternz in hands on training experience on the softwares provided by the IBM.The experimental investigations were made during collecting data , as while preparing the data, null values were been investigated with other resources in order to find cancelled matches due to weather .The graphical and statistical formulas were being analysed while selecting the type of visual for a particular query , we cant use same type of visualization method for all the eleven capabilities , we have to analyse the query and select the perfect visualization for the particular query .

5. Flowchart



6.Result

The final findings of the project are :

1. **To find the team that won the most number of matches in the entire IPL.**

   Mumbai Indians is the most frequently occurring category of winner with a count of 109 items (20.1 % of the total).

2. **To find the team that lost the most number of matches in the entire IPL.**

Royal Challengers Bangalore has lost number of matches with the count of 96 lost matches followed by KXIP and Dehli Dardevils with count of 94 matches

3. **Does winning a toss increase the chances of victory?**

   The chances of victory increases for certain teams with a varying percentage of chance of victory Mumbai Indians (13 %), Kolkata Knight Riders (12.2 %), Chennai Super Kings (11.8 %), Kings XI Punjab (10.8 %), and Royal Challengers Bangalore (10.6 %) are the most frequently occurring categories of toss_winner with a combined count of 440 items (58.5 % of the total).

4. **To find the player with the most player of the match awards.**

   AB de Villiers is the player who is awarded the most player of the match with a count of 18 awards around 20.5% then followed by CH Gayle with the count of 17 awards around 19.3 % and then followed by MS Dhoni with a count of 15 award .

5. **To find the city that hosted the maximum number of IPL matches.**

   The Mumbai city has hosted the maximum number of IPL matches to be precise exactly 101 matches followed by Kolkata with 77 matches and then followed by Delhi with 74 matches.

6. **To find the most winning team for each season.**

   Kings XI Punjab and Chennai Super Kings are the most winning team for the 2008 season with the count of winning 10 matches by KXIP and 9 matches by CSK.
   Royal Challengers Bangalore is the most winning team with the count of 9 matches for the 2009 season.
   Mumbai Indians and Chennai Super Kings are the most winning team for the 2010 season with the count of winning 11 matches by MI and 9 matches by CSK.
   Chennai Super Kings , Mumbai Indians and Royal Challengers Bangalore are the most winning team for the 2011 season with the count of winning 11 matches by CSK and 10 matches by MI and RCB.

   Kolkata Knight Riders , Chennai Super Kings and Mumbai Indians are the most winning team for the 2012 season with the count of winning 12 matches by KKR and 10 matches by CSK and MI.

   Mumbai Indians, Chennai Super Kings and Royal Challengers Bangalore are the most winning team for the 2013 season with the count of winning 13 matches by MI and 12 matches by CSK and 9 matches by RCB.
   Kings XI Punjab , Kolkata Knight Riders , Chennai Super Kings are the most winning team for the 2014 season with the count of winning 12 matches by KXIP, 11 matches by KKR and 10 matches by CSK.
   Mumbai Indians and Chennai Super Kings are the most winning team for the 2015 season with the count of winning 10 matches by MI and CSK.
   Royal Challengers Bangalore is the most winning team with the count of 9 matches for the 2016 season.
   Mumbai Indians and Kolkata Knight Riders is the most winning team with the count of 12

matches by MI and 9 matches by KKR for the 2017 season.
Chennai Super Kings and Kolkata Knight Riders is the most winning team with the count of 11 matches by CSK and 9 matches by KKR for the 2018 season.
Mumbai Indians and Chennai Super Kings are the most winning team for the 2019 season with the count of winning 11 matches by MI and 10 matches by CSK.

7. **To find the on-field umpire with the maximum number of IPL matches.**
S Ravi is the on-field umpire with the maximum number of IPL matches with the count of 106 matches followed by HDPK Dharmasena with count of 73 matches followed byC Shamshuddin with the count of 57 matches.
8. **To find the biggest victories in IPL while defending a total and while chasing a total.**
The biggest victories while defending a total is by Mumbai Indians in 2017 by winning by 146 runs, then followed by RCB in 2016 winning by 144 runs , then KKR in 2008 winning by 140 runs and the biggest victories while chasing a total are Deccan Chargers , Rajasthan Royals , KXIP winning by 10 wickets , 9 wickets , 8 wickets .

9. **Which team won the most matches while batting first?**
Mumbai Indians won the most matches while batting first with the count of 58 matches that is 26.9% .

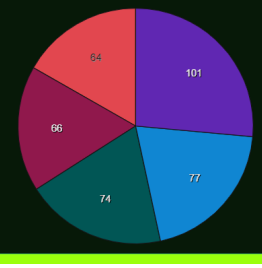
10. **Which team won the most matches while batting second.?**
Kolkata Knight Riders has won most of the matches while batting second with a count of 58 matches around 19.5% of the matches.
11. **List of teams which have won matches by most runs cumulatively.**

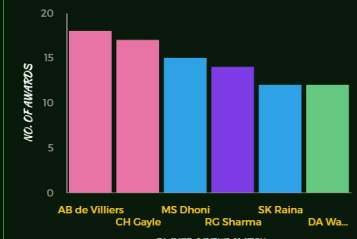
Rank	Winner	Cumulative runs
1	Mumbai Indians	1,866
2	Chennai Super Kings	1,778
3	Royal Challengers Bangalore	1,252
4	Kolkata Knight Riders	1,086
5	Kings XI Punjab	925

Tab 1 | Tab 1-2 | Tab 2 | Tab 3-1 | Tab 4 | +

81	756	247K	223K
Total Super Overs	Total No.of IPL matches	Total Runs	Total Batsman Runs

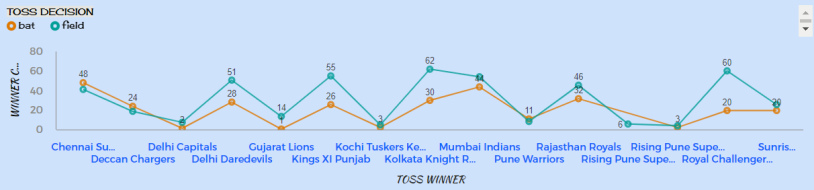
749	CITY THAT HOSTED MAXIMUM NO. OF IPL MATCHES	MOST PLAYER OF THE MATCH AWARDS
No Ball Runs	Mumbai, Kolkata, Delhi, Bangalore, Hyderabad	Chennai Super Kings, Mumbai Indians, Royal Challengers Bangal..., Sunrisers Hyderabad
6.58K		
Wide Runs		
3.79K		
Leg bye Runs		



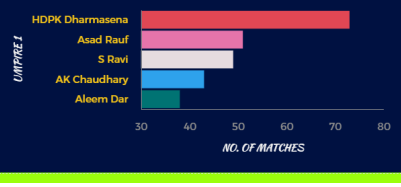


Tab 1 | Tab 1-2 | Tab 2 | Tab 3-1 | Tab 4 | +

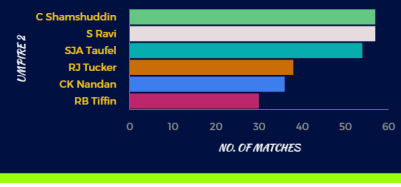
CHANCES OF VICTORY BY WINNING A TOSS



TOP 5 ON-FIELD UMPIRE 1 WITH MAXIMUM NO. OF IPL MATCHES

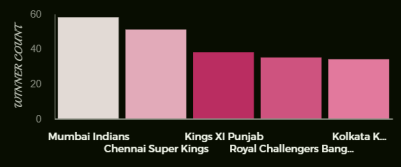


TOP 5 ON-FIELD UMPIRE 2 WITH MAXIMUM NO. OF IPL MATCHES

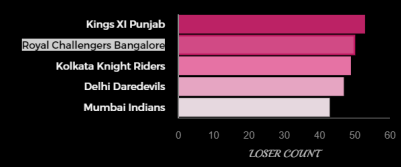


Tab 1 | Tab 1-2 | Tab 2 | Tab 3-1 | Tab 4 | +

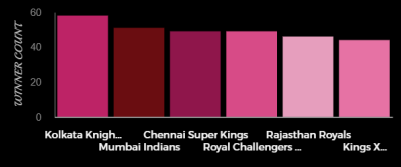
TEAM WON THE MOST MATCHES WHILE BATTING FIRST



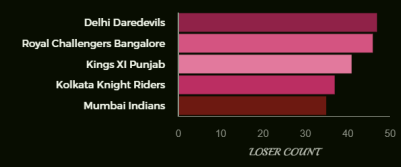
TEAM LOST THE MOST MATCHES WHILE BATTING FIRST

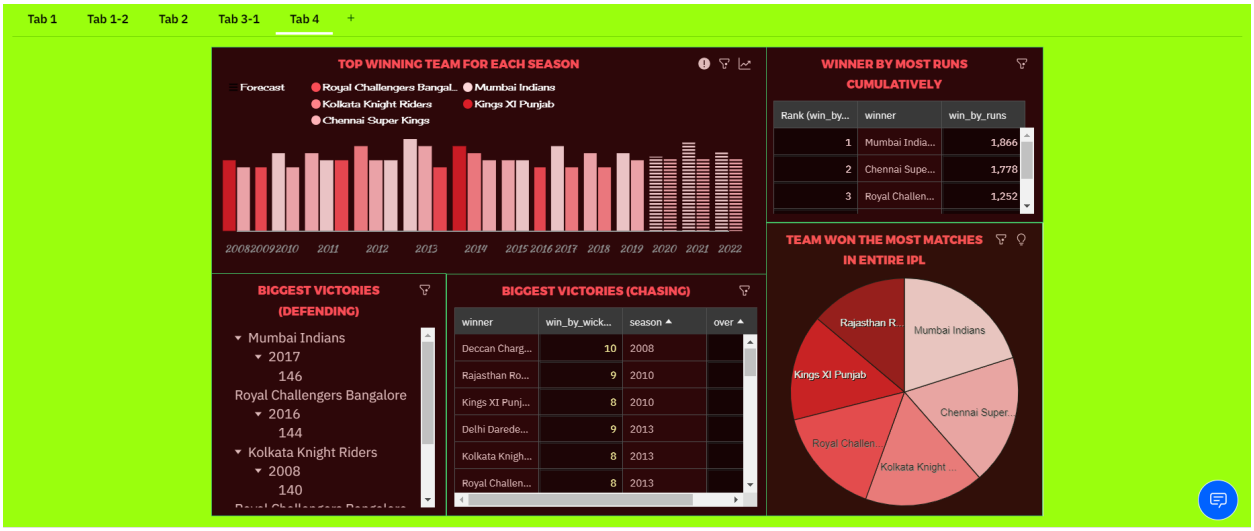


TEAM WON THE MOST MATCHES WHILE BATTING SECOND



TEAM LOST THE MOST MATCHES WHILE BATTING SECOND





7. Advantages

The advantages of the proposed solution is that the predictions result are the most accurate and the dashboard contains processed data which are so valuable for the team and team management in preparing strategies and to provide the statistical analysis of players based on different characteristics ,to predict the performance of a team depending on individual player statistics ,to successfully predict the outcome of IPL matches.

8. Applications

The areas where this solution can be applied is in IPL auctions to construct a team covering all the positions of batsman , bowler, wicket-keeper, all-rounder , captaincy etc and according to the analysis they can choose certain players as options to select in auctions for the positions needed in the team . This solution can also be used in discussing strategies according to the opponent team , venue , toss winner , etc can determine the winner in the match which can be used by teams to plan a strategy and execute it with the help of the visualization given by our solution for example a bowler has a good record in taking wickets against a team or a batsman , the team can use that bowling option instead of any other bowlers . This solution can be used by broadcasting tv channels to increase the interest of the viewers by visualizing the data by studying the pattern of dismissals by a batsman , by knowing the win or lost count in a particular venue , by seeing a record of a bowler by his number of wickets in the season , by the number of runs of the batsman , by number of sixes , with all these data shared with the viewers to increase the spice of the game by sharing knowledge of the game to common viewers for better understanding to the viewers hence increasing the viewers by there interests thus the channels success.

9.Conclusion

The project has been successfully executed by creating a dashboard for the eleven capabilities and by forecasting the future results using the help of IBM Cognos Analytics a

Business Intelligence Software. This dashboard can be used by the team and team management to prepare strategies and get a clear understanding of the skills and performance of there team with the help of the dashboard which is so easy to understand and creative with the visualizations.

10.Future Scope

The enhancements that can be made in the future are to add more data as the IPL season continues , explore more capabilities and study more trends from the data using different graphical and statistical analysis . The IBM cloud based software can enhance more real time solutions , by uploading live real time data to predict the match for more accurate results .

11.Bibilography

- https://us1.ca.analytics.ibm.com/bi/?perspective=home
- https://www.researchgate.net/publication/339681163_Prediction_of_Indian_Premier_League-IPL_2020_using_Data_Mining_Algorithms
- https://towardsdatascience.com/predicting-ipl-match-winner-fc9e89f583ce
- https://www.ibm.com/academic/technology/ibm-watson
- https://www.ibm.com/academic/technology/data-science
- https://www.ibm.com/academic/technology/cloud
- https://www.kaggle.com/nowke9/ipldata?select=matches.csv