Project Report on

Super Predictor Of Indian Premier League

Submitted for

IBM Hack Challenge 2021

Under the Guidance of: Shivam Shivare

Submitted by: Team Graviton (SPS_PRO_1986)

Team Members: Kumar Nilind

Table of Contents

Sr No.	Topic	Pg. No.
1.	Introduction	2
2.	Literature Survey	2
3.	Theoritical Analysis	3
4.	Experimental Investigations	4
5.	Flowchart	4
6.	Result	6
7.	Advantages	12
8.	Applications	13
9.	Conclusion	13
10.	Future Scope	14
11.	Bibiliography	14

1. Introduction

1.1 Overview

Indian Premiere League (IPL) is a major sports event in the field of Cricket, this event is often watched live by millions of spectators across India. This brings us a lot of data about the event observed by thousand of people, created by hundreds of cricketers and a few dozen teams collected over years. Such data is very useful is predicting various events that might happen in the future events and to analyze the performance of certain players and other important factors. (Umpires, Location, Teams etc.) This project aims to provide such analysis and prediction on the data available on the internet to predict performance of Teams, and players in the IPL.

1.2 Purpose

IPL is a major annual event that occurs with the collaboration of thousand of cricketers and investors. Therefore, it is useful to provide some data analysis and predictions on the teams and the players in the IPL to find the best match to watch various other things. The purpose of this Project is to provide predictions and data analysis of the past teams, players, and various other factors in the IPL using IBM Cognos Analytics provided by IBM Cloud.

2. Literature Survery

2.1 Exisiting Problem

The objective of this solution is to create a dashboard that visualizes the following capabilities and also forecast the future results, the problem statements of the solution are given below:

1. To find the team that won the most number of matches in the entire IPL.
2. To find the team that lost the most number of matches in the entire IPL.
3. Does winning a toss increase the chances of victory.
4. To find the player with the most player of the match awards.
5. To find the city that hosted the maximum number of IPL matches.
6. To find the most winning team for each season.
7. To find the on-field umpire with the maximum number of IPL matches.
8. To find the biggest victories in IPL while defending a total and while chasing a total.

9. Which team won the most matches while batting first
10. Which team won the most matches while batting second
11. List of teams which have won matches by most runs cumulatively

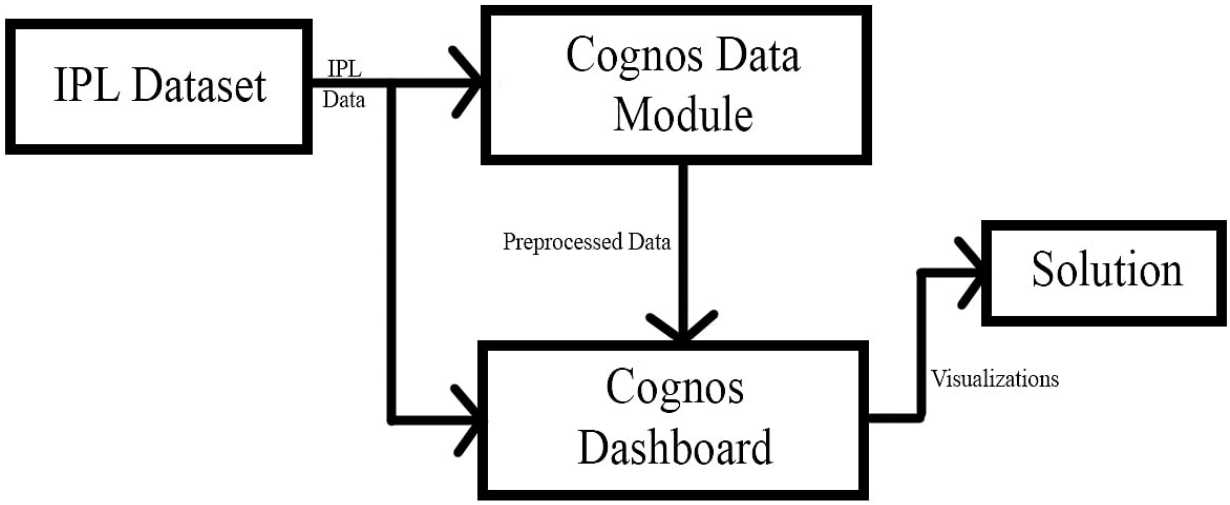
2.2 Proposed Solution

The Solution is built on IBM Cognos Analytics using various preprocessing of data in order to produce visualizations and predictions of IPL Data provided.
The Solution first produces the data required from the given dataset like losiing teams, total umpires and defending and chasing teams as required by the problem statements. The produced data is then used in Dashboard to create visualizations using various graphs. Mostly Bar Graphs, Histograms and Line graphs are used in this solution.

3. Theoritical Analysis

3.1 Block Diagram

The Block Diagram of the solution is given as:



3.2 Hardware/Software designing

We will use the following for producing our Solution:

1. IBM Cloud
2. IBM Cognos Analytics
3. IBM Watson Studio
4. Kaggle (for Datasets)

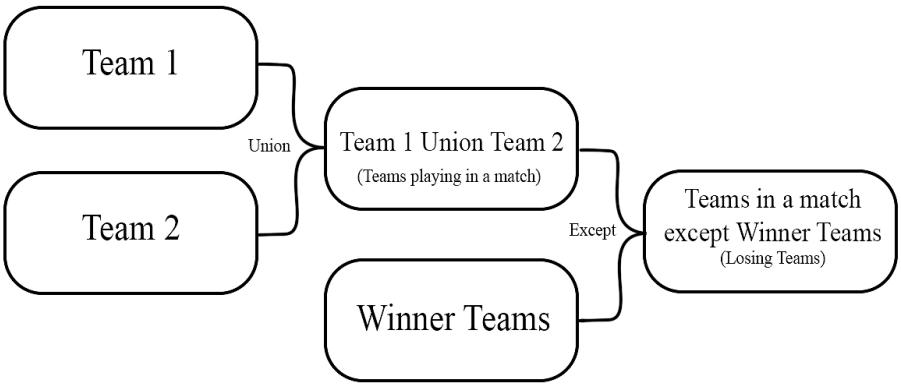
4. Experimental Investigations

The Project was created through experimental work with Cognos Analytics.
The Data is preprocessed using Cognos Analytics Data Module, Data module contains various features to preprocess tabel and create relationships between tables. The Solution uses features like Union, Intersection, Joining, Except of various columns in the dataset to produce required data. The resulting tables are then provided relationship with the main table in order to be able to be used with other required columns. Before doing this, data is also cleaned for better working like Teams with same location but different names are merged, unnecessary NULL values are removed.
After the data is preprocessed, we use the Dashboard functionality of Cognos which contains various features for Creating Visualizations like Widgets, Graphs, FIlters, Data groups, etc. The columns are then added to the required visualizations.

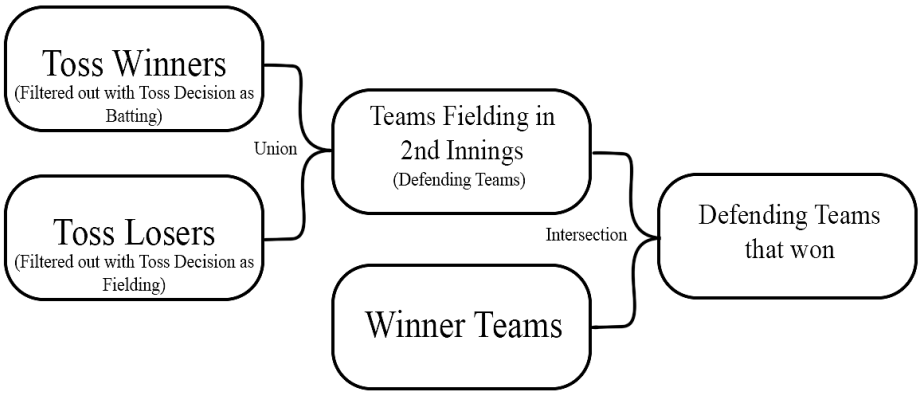
5. Flowchart

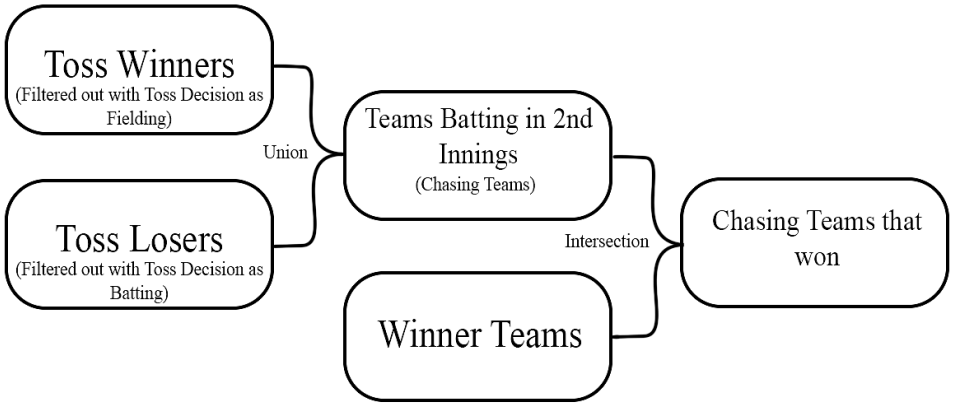
The Solution is divided into various problem staements, these statements are divided into various workflows, most of them are single step, so they are not mentioned, however any more complex workflow diagrams are given.

Flowchart to Determine Losing Teams,

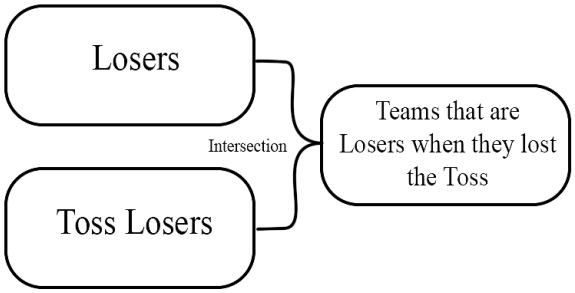


Using the Above chart we can determine losing teams from the given dataset. This flowchart is also necessary to determine Toss Losers which is later used to determine Defending and Chasing team Winners.

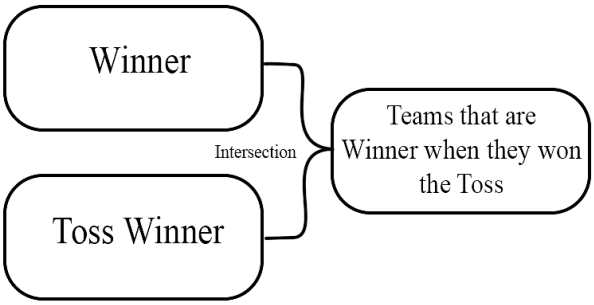




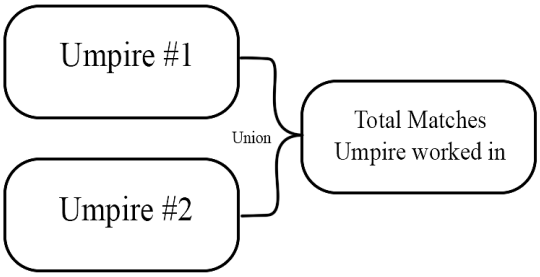
Flowchart for determining the Teams that Lost when they Lost the Toss



Flowchart for teams that Won when they won the Toss,



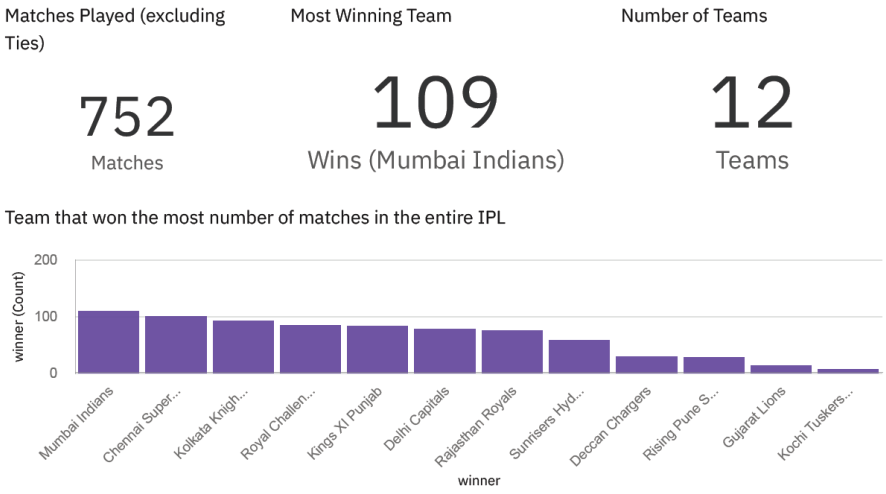
Flowchart for finding the Matches an Umpire worked in,



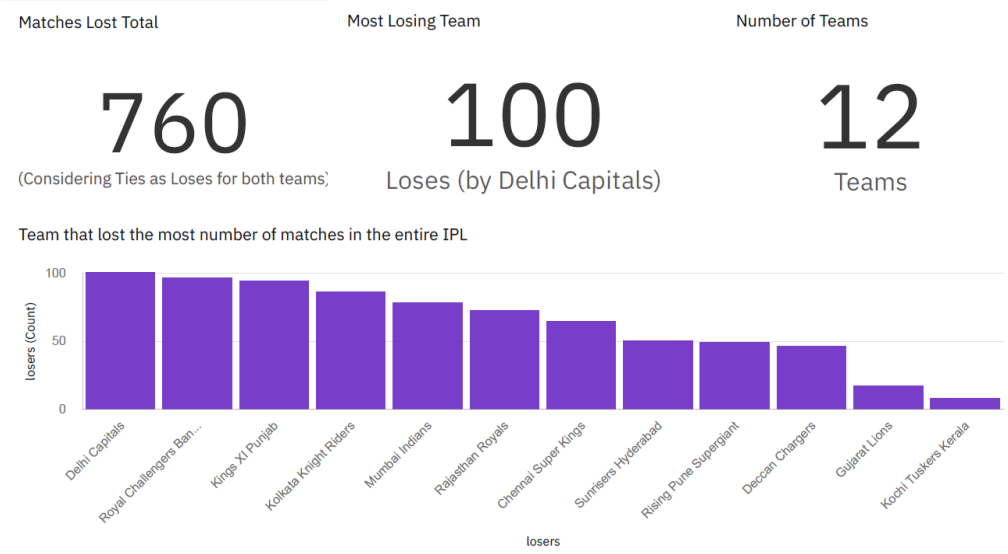
6. Result

Output Screenshots are given below,

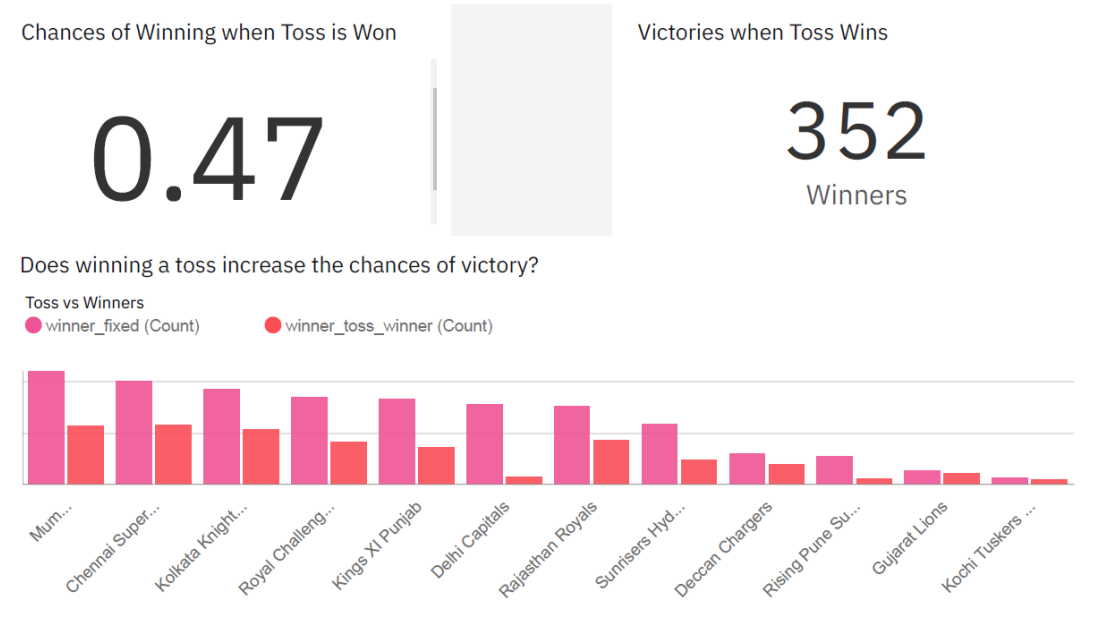
Statement #1: To find the team that won the most number of matches in the entire IPL.



Statement #2: To find the team that lost the most number of matches in the entire IPL.



Statement #3: Does winning a toss increase the chances of victory.



Statement #4:To find the player with the most player of the match awards.

Total Player of the Match Awards

752

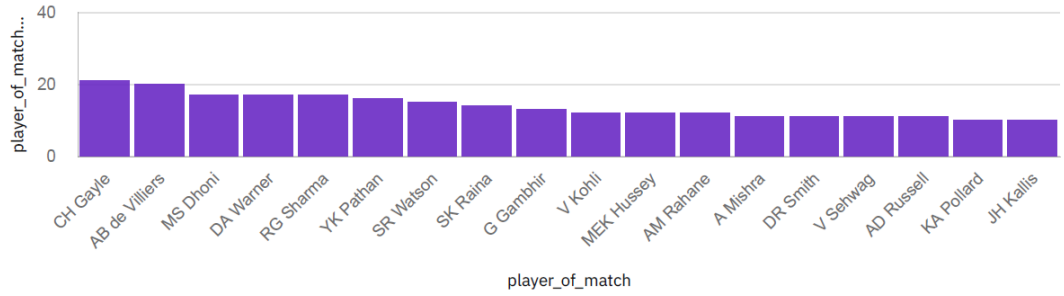
Awards

Most Player of the Match Awards

21

Awards (to CH Gayle)

The players with the most Player of the Match awards



Statement #5:To find the city that hosted the maximum number of IPL matches.

Cities in which matches held

32

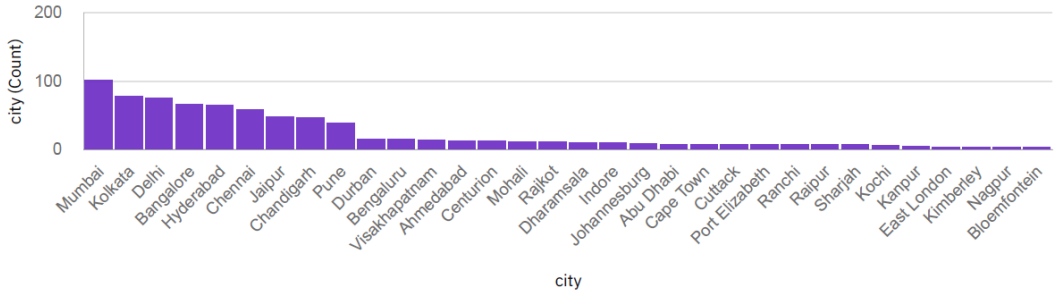
city

Most matches held

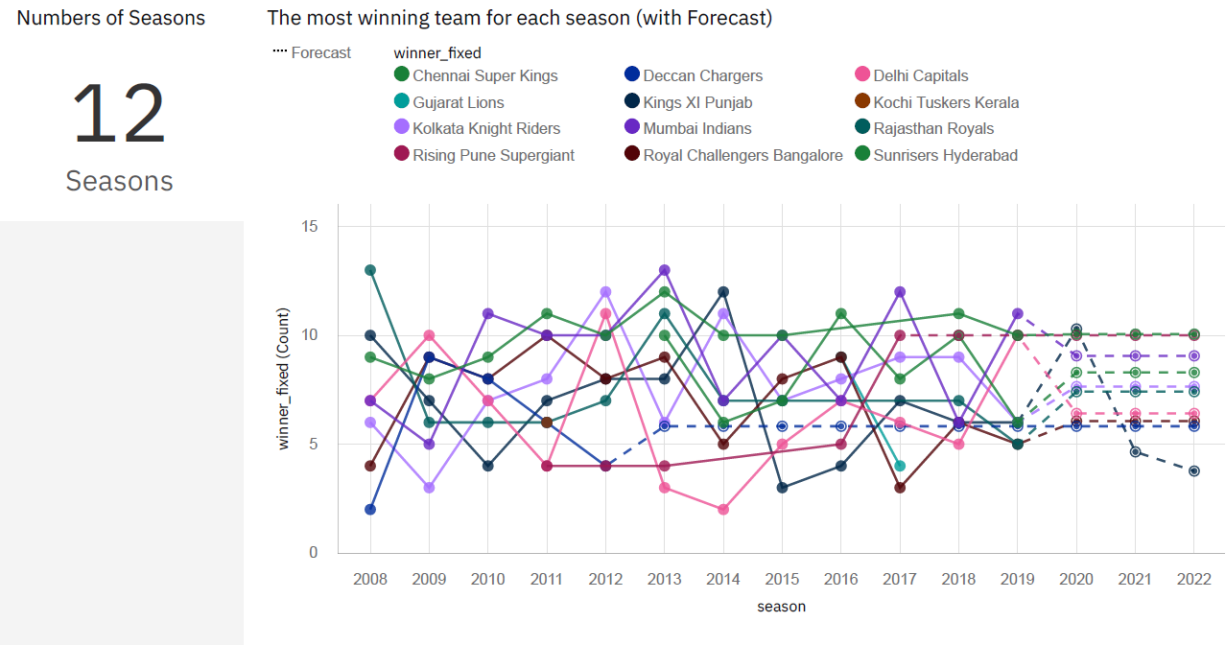
101

Matches (in Mumbai)

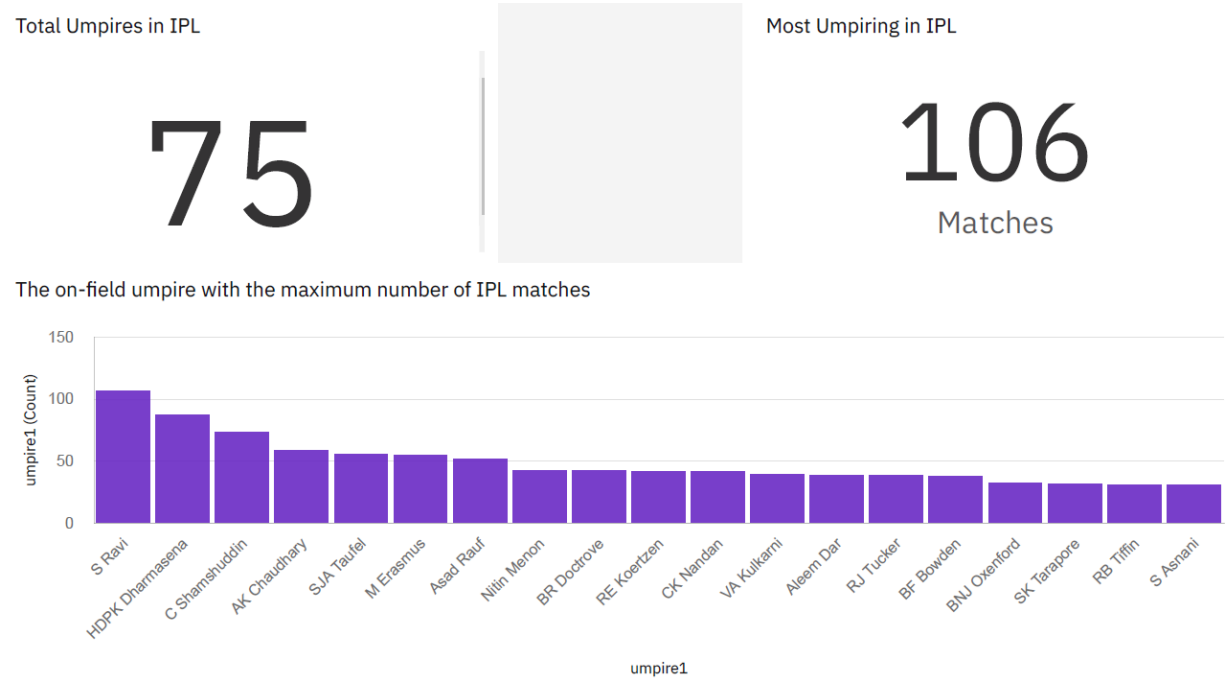
The city that hosted the maximum number of IPL matches



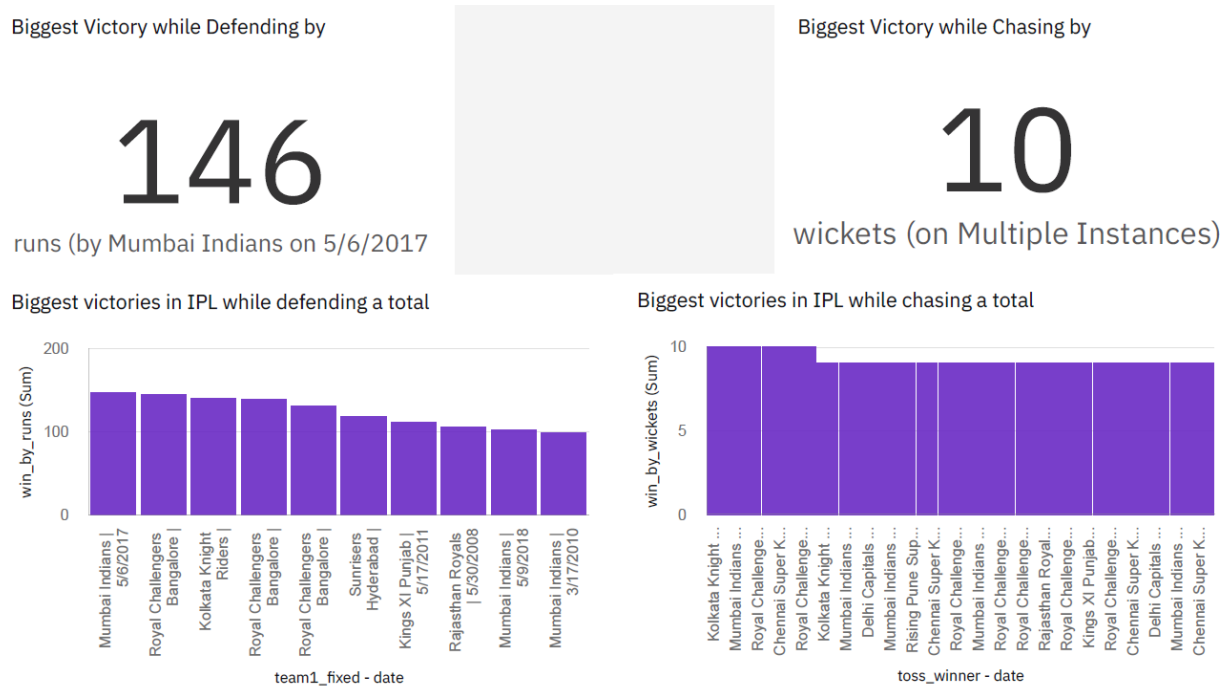
Statement #6:To find the most winning team for each season.



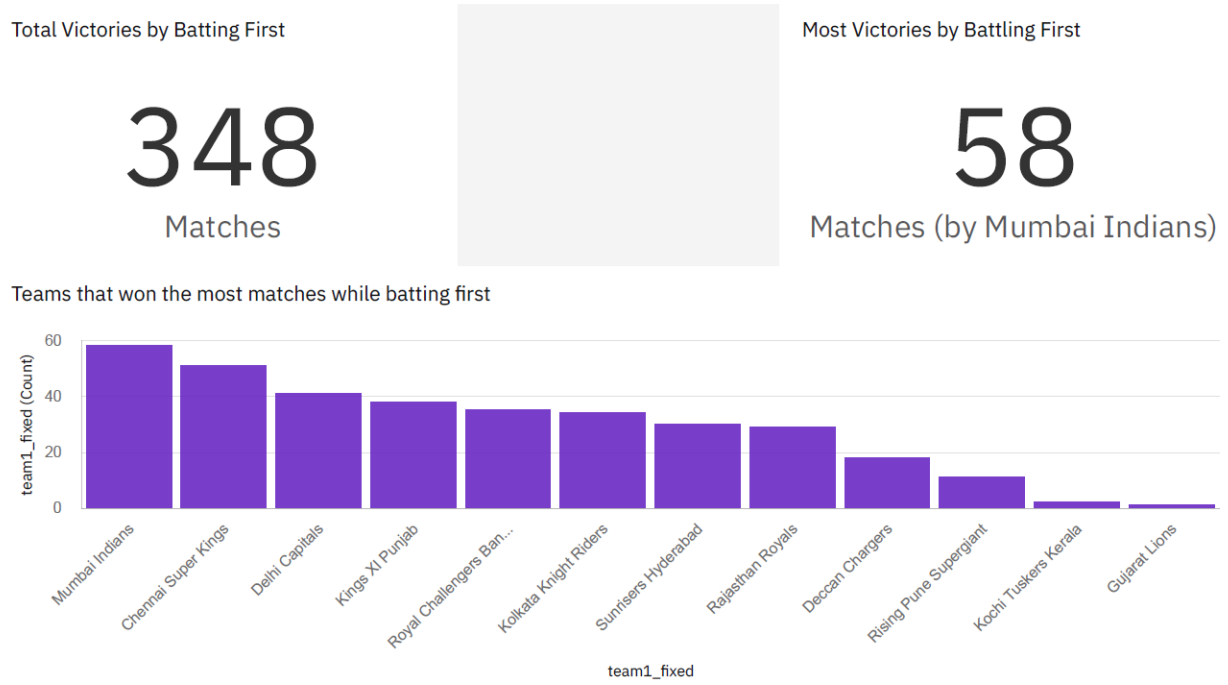
Statement #7:To find the on-field umpire with the maximum number of IPL matches.



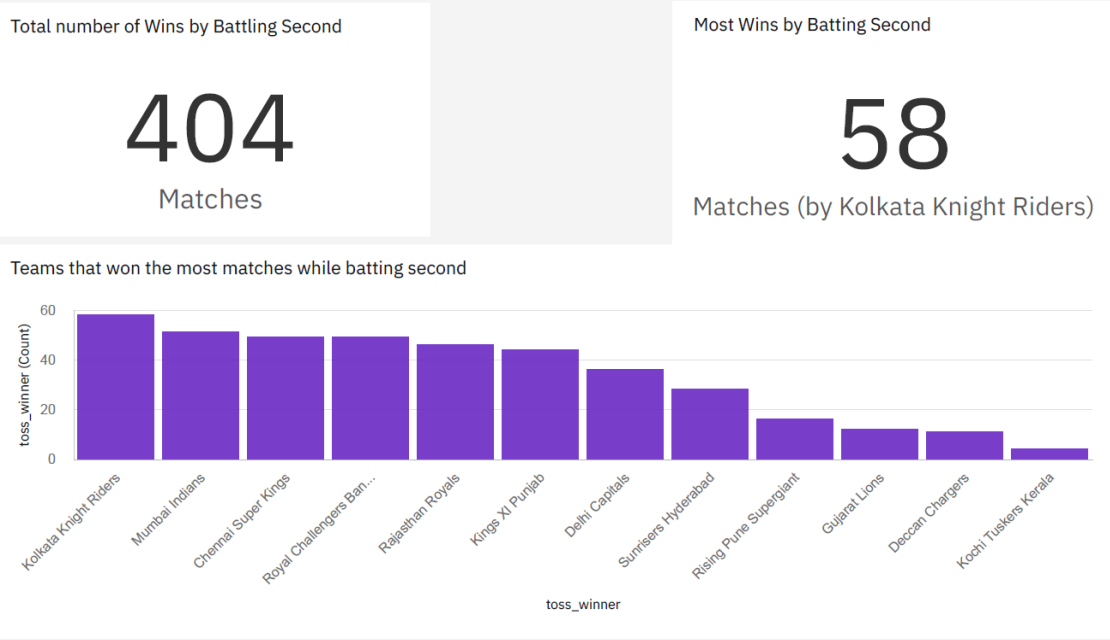
Statement #8:To find the biggest victories in IPL while defending a total and while chasing a total.



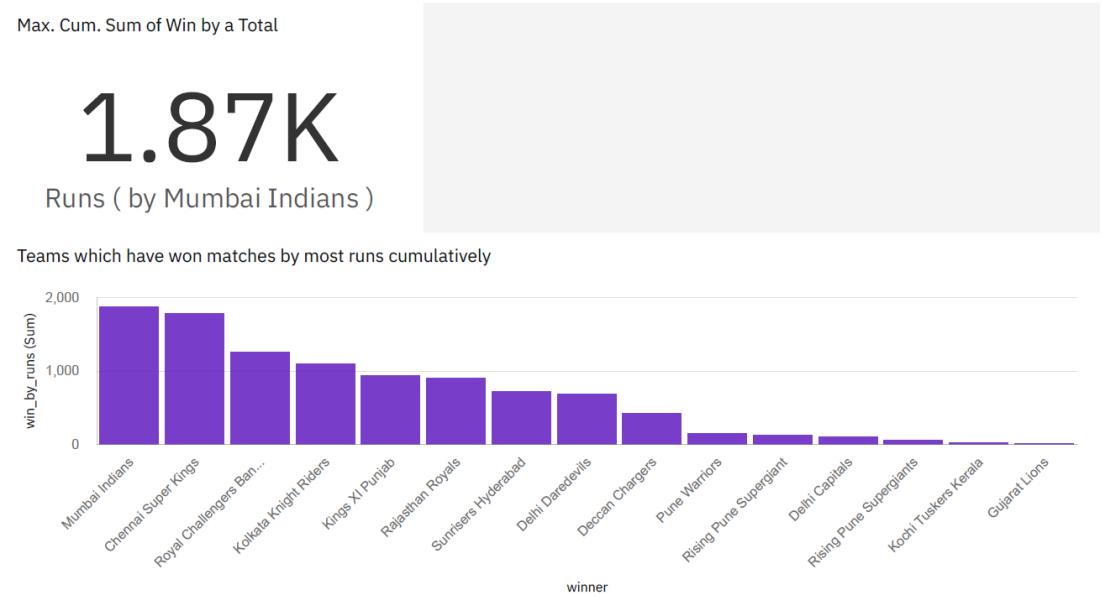
Statement #9:Which team won the most matches while batting first.



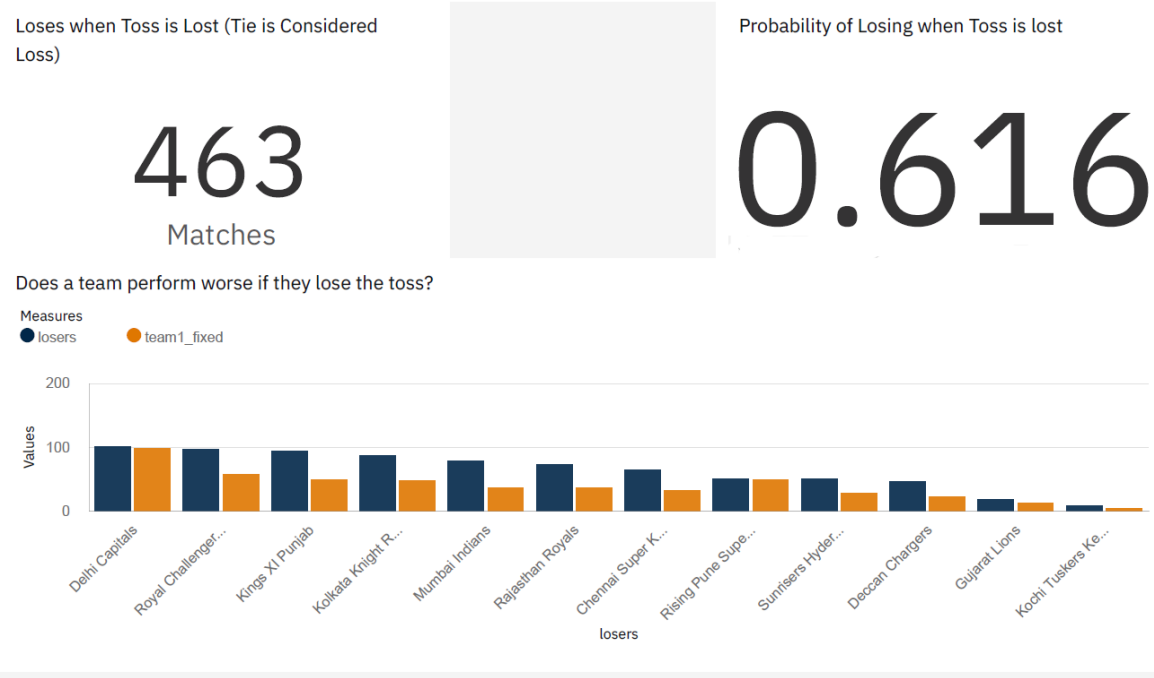
Statement #10:Which team won the most matches while batting second.



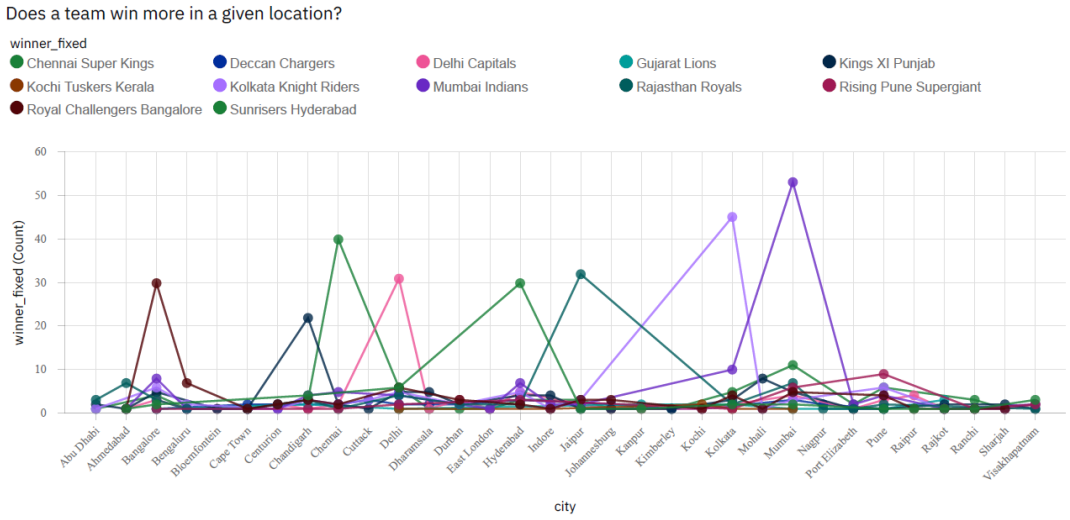
Statement #11:List of teams which have won matches by most runs cumulatively.



Statement #12:Does Losing a Toss decreases chances of Victory?



Statement #13:Does a Team win more in a given Location?



7. Advantages

The Advantages of the proposed Solution includes:

- It can be used to provide insights into the IPL teams an their performance based on their previous performance and location of the match.
- It can be used to provide insights into the Top IPL Players.
- We can use the data to perform various visualizations and this data can be integrated with newer dataset for better predictions.

8. Applications

This work solves various problem statements, the data gathered from the can be of much use for future analysis. The Inferences accquired from these statements is also very useful for Future use. This Solution can also be applied in the fields Data Analysis, Data Visualizations and Data Mining.

9. Conclusions

Statement-wise Conclusions are given below,

1. *Statement #1: To find the team that won the most number of matches in the entire IPL.*
   Conclusion: The team that won the most number of matches is **Mumbai Indians** with **109 Victories**.
2. *Statement #2: To find the team that lost the most number of matches in the entire IPL.*
   Conclusion: The team that lost the most number of matches is **Delhi Capitals** with **100 Loses**.
3. *Statement #3: Does winning a toss increase the chances of victory.*
   Conclusion: The chances are about equal with the **probability of winning after winning a toss being slightly lower 0.47**.
4. *Statement #4:To find the player with the most player of the match awards.*
   Conclusion: **C.H. Gayle** is the Player with most Player of the Match Awards with **21 Awards**.
5. *Statement #5:To find the city that hosted the maximum number of IPL matches.*
   Conclusion: **Mumbai** has hosted **101 Matches** overall.
6. *Statement #6:To find the most winning team for each season.*
   Conclusion: Result for this is given for each season from 2008 to 2019 in the Result section, we can see which team has won which season and by how many matches.
7. *Statement #7:To find the on-field umpire with the maximum number of IPL matches.*
   Conclusion: **S. Ravi** is the On-field umpire with **106 matches**.
8. *Statement #8:To find the biggest victories in IPL while defending a total and while chasing a total.*
   Conclusion: **Mumbai Indians** (on **5/6/2017**) has scored multiple Victories while

defending a total by the highest score difference of **146**. However, **multiple teams** have scored highest victories while chasing a total, by getting **all 10 wickets secured**.

9. *Statement #9:Which team won the most matches while batting first.*
   Conclusion: **Mumbai Indians** has won the maximum (**58 matches)** while batting first.
10. *Statement #10:Which team won the most matches while batting second.*
    Conclusion: **Kolkata Knight Riders** has won the maximum (**58 matches)** while batting first.
11. *Statement #11:List of teams which have won matches by most runs cumulatively.*
    Conclusion: **Mumbai Indians** has scored the most score difference (**1.87K Runs**) cumulatively.
12. *Statement #12:Does Losing a Toss decreases chances of Victory?*
    Conclusion: **Not Necessarily** as the probability of loss when toss is lost turned out to be **0.61** it seems like there is not much chances to lose when a toss is lost.
13. *Statement #13:Does a Team win more in a given Location?*
    Conclusion: **Yes,** Teams win more in the grounds they are more accustomed on playing so the respective teams win significantly more on their own states.

10. Future Scope

There is much future scope in the proposed solution as the data for the IPL can be integrated into the solution indefinitely to get more and more accurate predictions.
It is also possible to get information about teams and their performance based on their players and location of the match.

11. Bibiliography

- IBM Cognos Analytics Documentation, (https://www.ibm.com/docs/en/cognos-analytics/11.0.0?topic=manuals)
- SmartInternz Bootcamp Recordings available on YouTube.
- Dataset for IPL from 2008 to 2019 from Kaggle (https://www.kaggle.com/nowke9/ipldata?select=matches.csv)