CS432/532: Final Project Report

DATA ANALYSIS ON IPL TOURNAMENT (2008-2019)

Sahith Chowdary Dacharla
B00869531
sdachar1@binghamton.edu

Sai Phanindra Mutya
B00859323
smutya1@binghamton.edu

I. PROBLEM

We took a data set and worked on IPL – Biggest cricket league in the world. Using this data set we worked on deciding few key factors that strongly effect the tournament's outcome. The result can help a team to decide on areas like toss, players, result of a match, etc. We also represented the data visually using graphs for a better understanding. The results have been retrieved based on many attributes like toss, team, player, wins and losses, etc.

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Software Design and NoSQL-Databse and Tools Used

We used MongoDB for the NoSQL database. Because MongoDB can contain many databases and each database can contain many documents. These documents will hold the data in form of JSON objects. To work with MongoDB, we used the python distribution tool – PY-Mongo.
Project's environment setup: -
- Database name – test
- Collections made:
  - IPL matches – We took data like team name, match venue, toss decision, winner, won scenario, etc. From this we filter the required data out. We will be having stats of all the players and the teams.

B. Parts that you have implemented

I have implemented the following queries:
- Most Successful Team: This analysis is done for each season. Team with most wins in any season selected is termed as most successful team.
- Toss impact: Depending upon the toss winner data and match winner data, we get the past data of the impact of toss. Depending upon this data we determine whether toss played an important role in any season given.
- Best win in runs and wickets: Using the win_by_runs and win_by_wickets data, we determined the highest win in terms of runs and wickets in each season.

- Most experienced umpire: Using the umpire1 and umpire2 data, we found the number of matches umpired by each umpire, by which we determined the most experienced umpire.
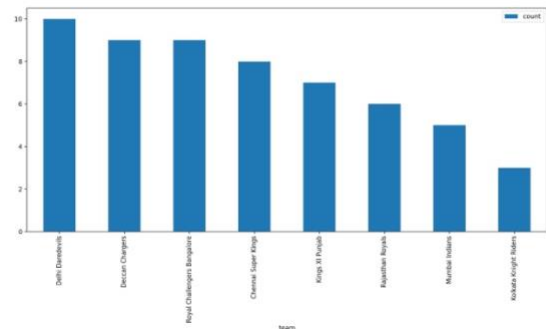
III. PROJECT OUTCOME
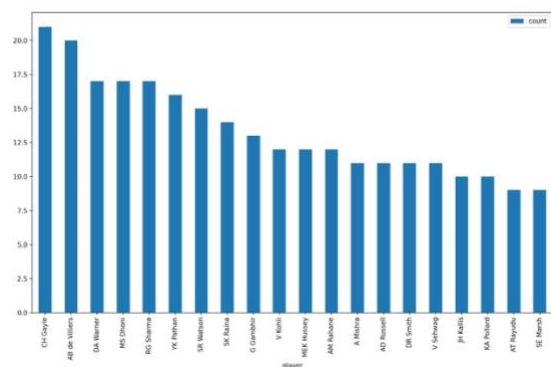

Fig. 1. Wins in a particular season
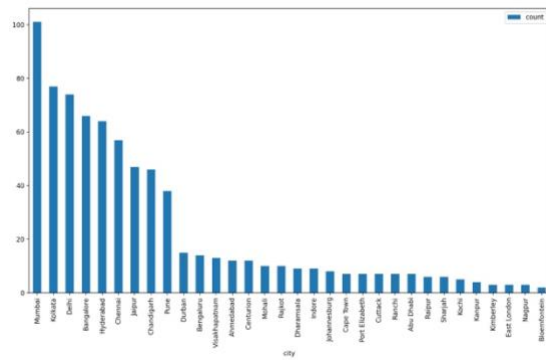

Fig. 2. Most Valuable Player

Fig. 3. Most Popular City

Above graphs show the information we obtained by implementing queries like Most Valuable Player, most popular venue, number of wins a particular season. This outcome will further provide additional information that can be used to get a detailed report about the tournament.

References

[1] https://pymongo.readthedocs.io/en/stable/tutorial.html

[2] https://docs.mongodb.com/manual/tutorial/iterate-a-cursor/

[3] https://www.kaggle.com/nowke9/ipldata?select=matches.csv

[4] https://docs.mongodb.com/manual/reference/method/db.collection.distinct/