Analysis of IPL data using NoSql

Karthik Kasarapu
B00817333
kkasara1@binghamton.edu
State University of New York at Binghamton
Binghamton, New York

Aravind Reddy Yenugula
B00800219
ayenugu1@binghamton.edu
State University of New York at Binghamton
Binghamton, New York

I. PROBLEM

We worked on IPL(Indian Primier League) data set which helps board council to decide best player at the end of specific season. In addition, it helps team to decide whether to opt for batting or bowling when they win toss. We retrieved results based on multiple attributes like wins, toss, team name, player name etc. Furthermore, we also visually represented the data using graphs which helps cricket council to understand the information easily and accurately.

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. *Software Design and NoSQL-Database and Tools Used*

We used MongoDB as a NoSQL database and Py-Mongo(Python distribution tool) for working with MongoDB using python. MongoDB can contain many databases and each database in turn contain many documents.These documents hold data in the form of JSON objects. Environment setup of this project,

- Database_name: ipl
- Collections :
  - batsman_score: This collection contains batsman name, season in which he played, total number of runs scored, number of balls played.
  - ipl_venue: This collection contains information of the venue name, city of the venue, team names and winner of the match.
  - details: This collection contains personal information of all the players.

B. *Supported Queries*

We have implemented the following queries:

- Retrieving names and details of all the players who played in a given season from the collection batsman score.
- Calculating the strike rate of a player during particular season.
- To decide the best player based on the average of the runs scored in a particular season.
- To retrieve additional information of the player by performing join operation with another collection (details).
- Analyse the pitch conditions based on the previous matches and determine whether it's a first batting favourable or first bowling favourable.
- Plotting the graph for overall number of runs scored by a batsman.
- Plotting the graph of player average for a given season.

III. PROJECT OUTCOME

Github link to the source code:
https://github.com/YAravindReddy/Database-project3

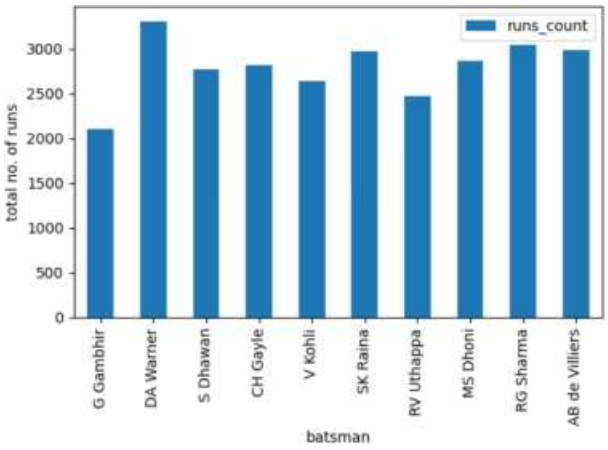

Fig. 1. Overall runs scored by a batsman.

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/manasgarg/ipl
[4]https://docs.mongodb.com/manual/aggregation/