**IPL Data Analytics**

12010504-Ganesh Karode, 12011084-Omkar Karpe, 12010914-Kartik Rupauliha, 12010397-KartikRajput, 12011421-Nakul Kasar

Department of Engineering, Sciences and Humanities (DESH)

*Abstract*— *In today’s date data analysis is need for every data analytics to examine the sets of data to extract the useful information from it and to draw conclusion according to the information. Data analytics techniques and algorithms are more used by the commercial industries which enables them to take precise business decisions. It is also used by the analysts and the experts to authenticate or negate experimental layouts, assumptions and conclusions. In recent years the analytics is being used in the field of sports to predict and draw various insights. Due to the involvement of money, team spirit, city loyalty and a massive fan following, the outcome of matches is very important for all stake holders. In this paper, the past thirteen years data of IPL containing the players details, match venue details, teams, ball to ball details, is taken and analyzed to draw various conclusions which help in the improvement of a player’s and team’s performance. Various other features are also predicted. The advantage of this tool is that is implemented using simple python modules such as pandas, matplotlib, numpy which are used to create data frames, plot different graphs, and are easy to understand.*

*.*

*Keywords — Python, Cricket, Data Analytics, IPL, Pandas, Matplotlib.*

# Introduction

Cricket is the most popular sports in India by far, and is played almost everywhere in India. The Indian cricket team has won the Cricket World Cup in 1983. This is followed by the 2007 ICC World Twenty20, the 2011 Cricket World Cup, 2013 ICC Champions Trophy. In 2002 India shared the 2002 ICC Champions Trophy with Sri Lanka. The Ranji Trophy, the Deodhar Trophy, the Irani Trophy, NKP Salve Challenger Trophy, the Vijay Hazare Trophy and the Duleep Trophy are some of the domestic competitions in India. In addition, the BCCI conducts the Indian Premier League known as IPL which is a limited overs competition. The Twenty20 competition as it is better known, is one of the biggest sports leagues in the world and the biggest cricket league in the world. While cricket is by far the most popular sport in the country it does not feature as the country’s national sport. Cricket features as a prominent part of the Indian culture. The limited overs matches have become more popular and have seen a rapid grown in the years 2016-18. The IPL as a whole was valued by financial experts at $4.16 Billion US Dollars in 2016, but that number grew to $5.3 Billion in 2017 and $6.13 Billion in 2018.

Matches are scheduled in late afternoon or evening so as to ensure that a part of them at least are played under floodlights at night. This is done to increase the viewership of the matches for television audience worldwide. Initial, league matches are played on a home-and-away basis across all teams. With the planned expansion to 10 clubs in 2011(2 groups of five each), that format has been revised. Currently matches between some teams would be limited to a single encounter. The top four teams contest three play-off matches, with one losing team being given a second chance to reach the final, which has been introduced with an aim of maximizing potential television revenue. The play-off portion of the tournament involves the four teams that finished at the top of the tables in a series of knockout games that allows one team which had lost its first-round game a second opportunity to advance to the final match. In this paper we perform the analysis of the IPL cricket teams from 2008 to 2020. Factors such as most wins by a team, number of wickets obtained by bowlers, decisions on toss are analyzed graphically. Using this analysis, which team has a strong batting power and strong bowling power. We also analyze that which team and player have the most records in various domains of cricket.

# Implementation details

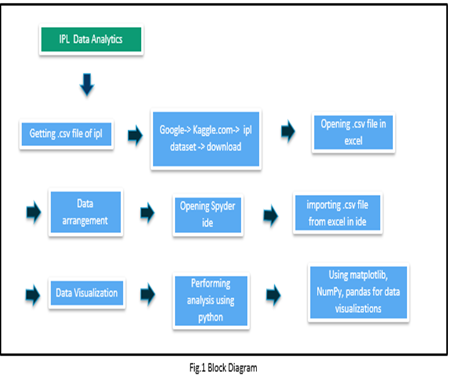
Python 3.9 is an interpreted, high-level general-purpose programming language that we used. Guido Van Rossum created the film, which was initially released in 1991. Its language elements and object-oriented approach are intended to assist programmers in writing clear, logical code for both small and large-scale tools. Python is used for server-side web development, software development, and mathematics. It can be used in conjunction with other software to create workflows, connect to database systems, read and modify files, handle big data, and perform complex mathematics. It can also be used for rapid prototyping or production-ready software development.

This project is a statistical analysis tool for IPL data from all seasons beginning in 2008 and running through 2020. Working with IPL data from all seasons and creating various statistical plots, graphs, and charts with the data. The data for the project is stored in CSV files. The CSV files contain all of the information on the matches that took place between the IPL teams. Data pre-processing: the first step in the project is to learn how to use and implement various Python built-in modules. These diverse modules help to improve code representation and user comprehension. The libraries pandas, matplotlib, numpy, plotly, and others are used. Pandas is a data manipulation and analysis software library for the Python programming language. It includes data structures and methods for manipulating numerical tables and time series, in particular. It's open-source software with a three-clause BSD license. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. Plotly and seaborn are used to create the graphical visualizations for better understanding of the data.

The proposed system is developed by using Spyder and JupyterLab software. These software are free and open-source scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. They feature a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

Algorithm:-

* We start by downloading the IPL dataset csv file(.csv)
* Further the data from this file is processed and split into required format in excel.
* Data-frame is created from this sorted output using pandas in Spyder IDE.
* Once the file is loaded in IDE using Pandas various operations can be performed over it using Python.
* Data is visualized from this data-frame using matplotlib, pyplot, NumPy, seaborn, etc.



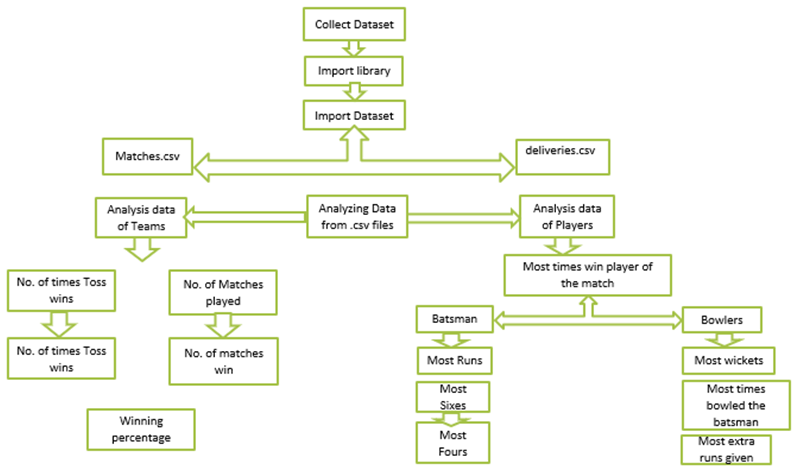
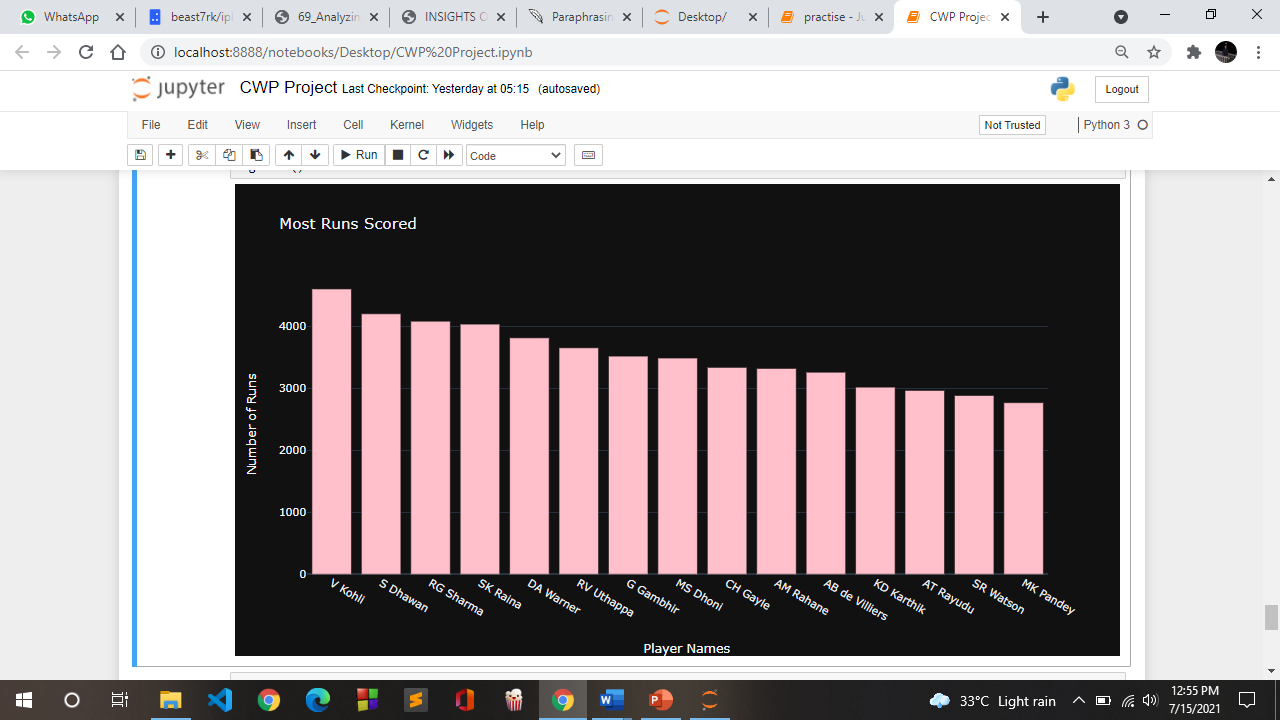


Fig.2 Flowchart

# Results and discussion

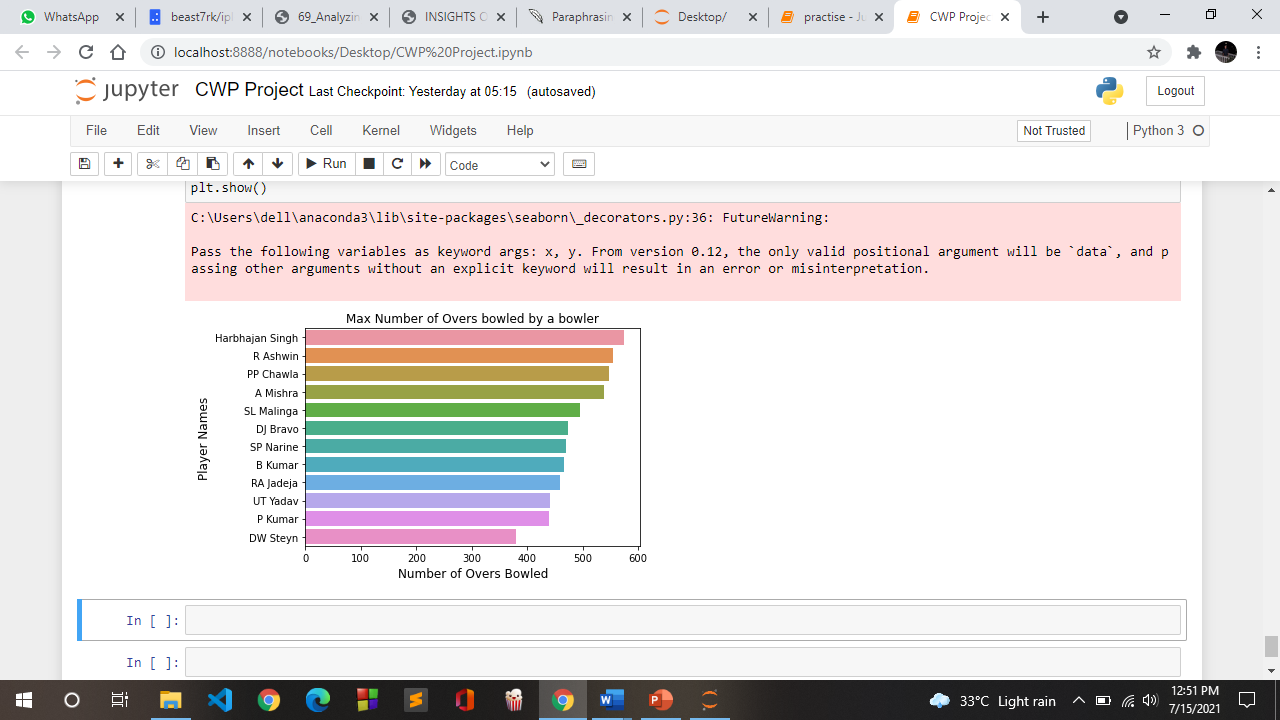
The following represents the output of the result of the analysis done with Python on the ipl dataset.

**Visualization:**



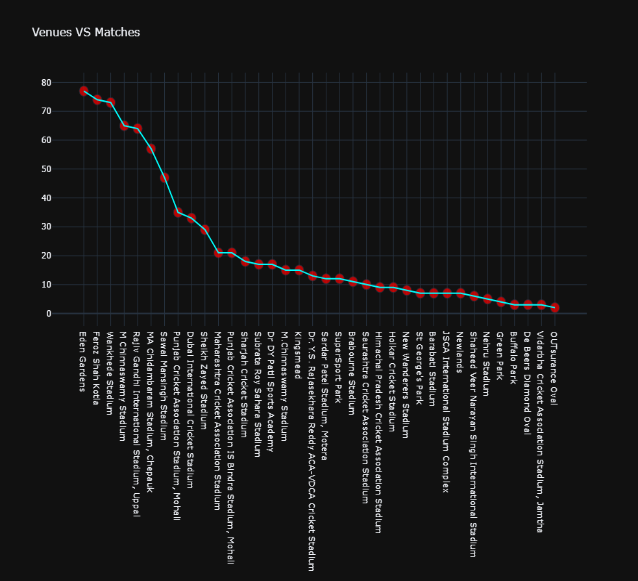
**Discussion:** This screenshot shows the top 15 maximum run scorers in The IPL.

**Visualization:**



**Discussion:** This screenshot shows the top 12 bowlers with maximum number of overs bowled.

**Visualization:**



# Conclusion and Future Scope

In the paper we focus on IPL team data from matches played from 2008 to 2020. The results of the study could be used to make future judgments based on visualization. The observations may also be used for articles in other media in order to gain a better knowledge of the IPL teams' performance.

Here we analyze the performance of the IPL players with respect to the following: Most successful team by runs, most successful team by wickets, Overall performance of the team, Man of the match winners, toss winners by runs, Toss winner by wickets. It can be observed that the toss plays a vital role in deciding the winning team. The significant beneficiaries are the sponsor (to choose the effective team and team members), players (to improve their self-efficiency in their upcoming tournament) and also used for fans (they will be curious enough to guess the winners of the next tournament).

Moreover, we can further combine the project with machine learning and predict the winner of matches aur next seasons according to the past performances of the players and the teams.

##### References

1. YouTube tutorial Data Analytics with python by freecodecamp.
2. machinelearningknowledge.ai
3. towardsdatascience.com
4. Kaggle.com
5. GeeksForGeeks.com
6. w3schools.com