



# **SANDIP** FOUNDATION

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**A PROJECT REPORT ON**

## **Data Analysis on IPL**

SUBMITTED TOWARDS THE  
PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

**BACHELOR OF ENGINEERING**

**(Computer Engineering)  
BY**

- |                                |                   |
|--------------------------------|-------------------|
| <b>1. TEJAS RAJENDRA PATIL</b> | <b>B150614293</b> |
| <b>2. SWAGAT DADAJI AHIRE</b>  | <b>B150614205</b> |
| <b>3. SATYAM CHAUDHARI</b>     | <b>B150614220</b> |

**Under the Guidance of**

**Prof. Vivek Waghmare**

**DEPARTMENT OF COMPUTER ENGINEERING**

**Sandip Foundations**

**Sandip Institute of Technology and Research Centre Mahiravani, Trimbak  
Road, Nashik, (M.S.), India**



# **SANDIP**

**FOUNDATION**

**Sandip Foundations**  
**Sandip Institute of Technology and Research Centre**

## **DEPARTMENT OF COMPUTER ENGINEERING**

### **CERTIFICATE**

This is to certify that the Project Entitled

**Data Analysis on IPL**  
Submitted by

- 1. TEJAS RAJENDRA PATIL**
- 2. SWAGAT DADAJI AHIRE**
- 3. SATYAM CHAUDHARI**

is a bonafide work carried out by Students under the supervision of Prof. Vivek Waghmare and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

**Prof. Vivek Waghmare**  
Internal Guide  
Dept. of Computer Engg.

**Prof. A.D.Potgantwar**  
H.O.D  
Dept. of Computer Engg.

**Dr.S.T.Gandhe**  
Principal  
Sandip Institute of Technology and Research Centre

Signature of Internal Examiner

Signature of External Examiner

## **PROJECT APPROVAL SHEET**

A Project Title:

Data Analysis on IPL

Is successfully completed by

1. TEJAS RAJENDRA PATIL
2. SWAGAT DADAJI AHIRE
3. SATYAM CHAUDHARI

At

DEPARTMENT OF COMPUTER ENGINEERING

**Sandip Institute of Technology and Research Centre**

**SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE ACADEMIC YEAR 2019-2020**

Prof. Vivek Waghmare  
Internal Guide  
Dept. of Computer Engg.

Prof.(Dr)A.D.Potgantwar  
H.O.D  
Dept. of Computer Engg.

❖ **Contents:**

- **Abstract**
- **Introduction**
- **Related Work**
- **Methodology**
- **Factors**
  - **Pitch**
  - **Team Strength**
  - **Home Ground Advantage**
  - **Past Record**
  - **Toss**
- **Conclusion**

# **Python Mini Project**

## **Title:**

Data Analysis on IPL.

## **Aim:**

To perform data analysis on the given data set through data analysis and data visualization.

### **• Abstract**

Cricket is one the most watched sport now-a-days. Winning in cricket depends on various factors like home ground advantage, performances in the past matches, experience of the players, performance at the specific venue, performance against the specific team and the current form of the team and the player. In the recent past, a lot of research has been done which measures the player's performance and predicts the winning percentage.

This article briefs about the factors that cricket game depends on and discusses various researches which predicted the winning of a team with an advent of statistical modelling in sports. Cricket is one of the most popular team games in the world. With this article, we embark on predicting the outcome of Indian Premier League (IPL) cricket match using a supervised learning approach from a team composition perspective. Our work suggests that the relative team strength between the competing teams forms a distinctive feature for predicting the winner.

Modeling the team strength boils down to modeling individual player's batting and bowling performances, forming the basis of our approach. We use statistics and recent performance of a player to model him. Player independent factors have also been considered in order to predict the outcome of a match. Machine learning is used in predicting the outcome of a cricket match before and during a match.

## • INTRODUCTION

Sports analytics play a major role in various problems associated with sport. Some of these problems are the ranking of individual players and their specialized skills, the composition of teams with an optimal balance of specialized skills, the ranking of teams, the negotiation of contracts, their potential revenue streams, the planning of both physical and mental training, the development of strategies for winning games and tournaments, assessing the effectiveness of coaches and referees, the medical aspects of sports injuries (health and insurance), the analysis and improvisation of rules, the quality of equipment and technology, the determination of awards, historical records and the generation of odds for gambling activities. Related to above information, the coherent statistical presentation of both raw data and its inference to the decision makers is to facilitate successful planning and implementation. Further more, the media and the public have a great appetite for well visualized statistics.

New opportunities for sports analytics have arisen due to the advent and availability of detailed and high quality data. For example, in Major League Baseball (MLB), the system shave provided comprehensive data on pitching and fielding. These systems record every play while also tracking the exact movements of all players on the field. Using these data sources, we can make very useful prediction, and various Statistics for improvement purposes.

Today's level of sports analytics has evolved where both the technology which provides data, and the statistical methodologies which provide the tools for analysing data, improved very rapidly. Though sports analytics has been rapidly developing, it has not been the case with cricket. Due to historical reasons where cricket was perceived as a leisurely gentleman's game played without remuneration to players (until recently), cricket was not subject to large financial transactions.

## • RELATEDWORK

In this paper, a methodology for identifying promising batting orders in one-day cricket was presented. In particular, they suggested some batting orders that have never been tried by the Indian team and contradict prevailing wisdom.

As a byproduct of investigation, a simulation procedure was developed for generating first innings runs against an average opponent. The simulation procedure was based on estimates from a Bayesian log-linear model. Finally, methods were developed with the intention of finding optimal or nearly optimal batting orders at the start of a team's innings.

In this paper, two methodologies have been used. MySQL database is used for storing data whereas Python for the GUI. The algorithm used is Clustering Algorithm for prediction.

The steps followed are as

1. Begin with a decision on the value of k being the number of clusters.
2. Put any initial partition that classifies the data into clusters.
3. Take every sample in the sequence compute its distance from centroid of each of the clusters. If sample is not in the cluster with the closest centroid currently, switch this sample to that cluster and update the centroid of the cluster accepting the new sample and the cluster losing the sample.
4. Repeat above step until convergence is achieved, that is until a pass through the training sample causes no new assignments.

This paper introduces a model that has three segments which focuses on diverse contemplations developing out of a more profound examination of T20 cricket. The models are made utilizing Data Analytics strategies from machine learning area. In this work 5 highlights of IPL vocation and 5 highlights of International T20 Career have been thought about for both batsmen and bowlers yet in future work more highlights can be made and considered.

- **METHODOLOGY:**

The work of our project focuses on two models. The two models are:

1. Descriptive model
2. Predictive model

### **DESCRIPTIVE MODEL:**

The descriptive model focuses mainly on two aspects: It describes the data and statistics of the previous information i.e. batting, balling or all-rounder. It gives the past information of the matches played by the IPL teams.

### **PREDICTIVE MODEL:**

The predictive model focuses on predicting the winning percentage of the team. The ranking of the players is displayed as well. The user has the liability to choose the two teams playing against each other. The selection of the teams works on the criteria as

1. If the players are batsmen then, sorting is done according to the strike rate of the batsmen.
2. If the players are bowlers then, sorting is done according to the average rate of the bowler.
3. If the players are all-rounder then, sorting is done considering both strike rate as well as average rate. The algorithm used for this model is Decision Tree Classifier. A decision tree is built using top-down approach. In this algorithm the root node i.e. the prior factor considered is the city where the match is being played. The tree is built according to the prominent factors (city, venue, teams, toss decision) considered in the match.

## ➤ Factors

Cricket first innings score can be predicted like all other games. We need to find the best attributes or factors that influence the match outcome. The result of a cricket match depends on more of in-game and pre-game attributes. Pre-game attributes like Pitch, Team Strength, Weather, Venue etc. and in-game attributed like run rate, total run, strike rate, wickets in hand etc. influence a match score result predominantly. Below are the attributes that decide outcome of the score in first innings cricket match.

### ➤ Pitch:

Not fixed except the dimensions of the inner circle and pitch which are 30 yards and 22 yards respectively. Outfield variations and pitch can have a substantiate effect on bowling and batting. The spin of the ball, seam movement and the bounce depends on the nature of the pitch. It depends on how wet is the pitch. The more wet the pitch, the slower it will play. On the off chance that it is drying out, those balls will change significantly, yet all it will get less difficult those drier it gets Green pitches tend to get easier to bat on. Wickets can get significantly more dry or wet (on the off chance that its downpours). They might start to break up if they are soft (which would help the bowlers).

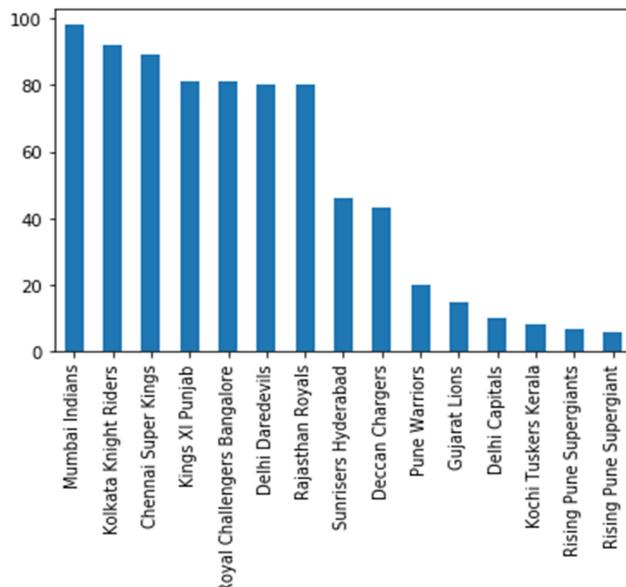
### ➤ Team strength

The team strength should be balanced for winning a match. Captainship in a team is also a deciding factor.

## ➤ Toss:

According to cricket analysts, there is sure measure of advantage for a team if it wins the toss. This might not be deciding factor in a match but it would give the team advantages.

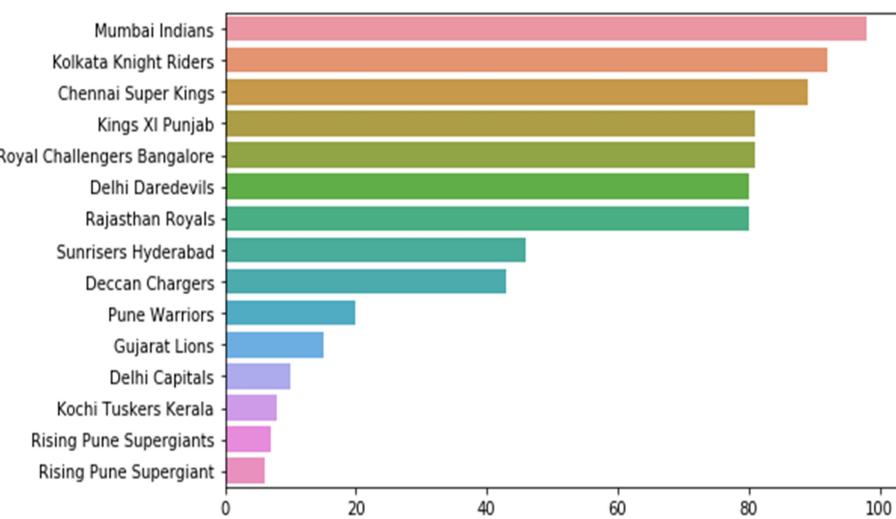
```
In [9]: #Normal plot  
data.toss_winner.value_counts().plot(kind='bar')  
  
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0xce210fc4a8>
```



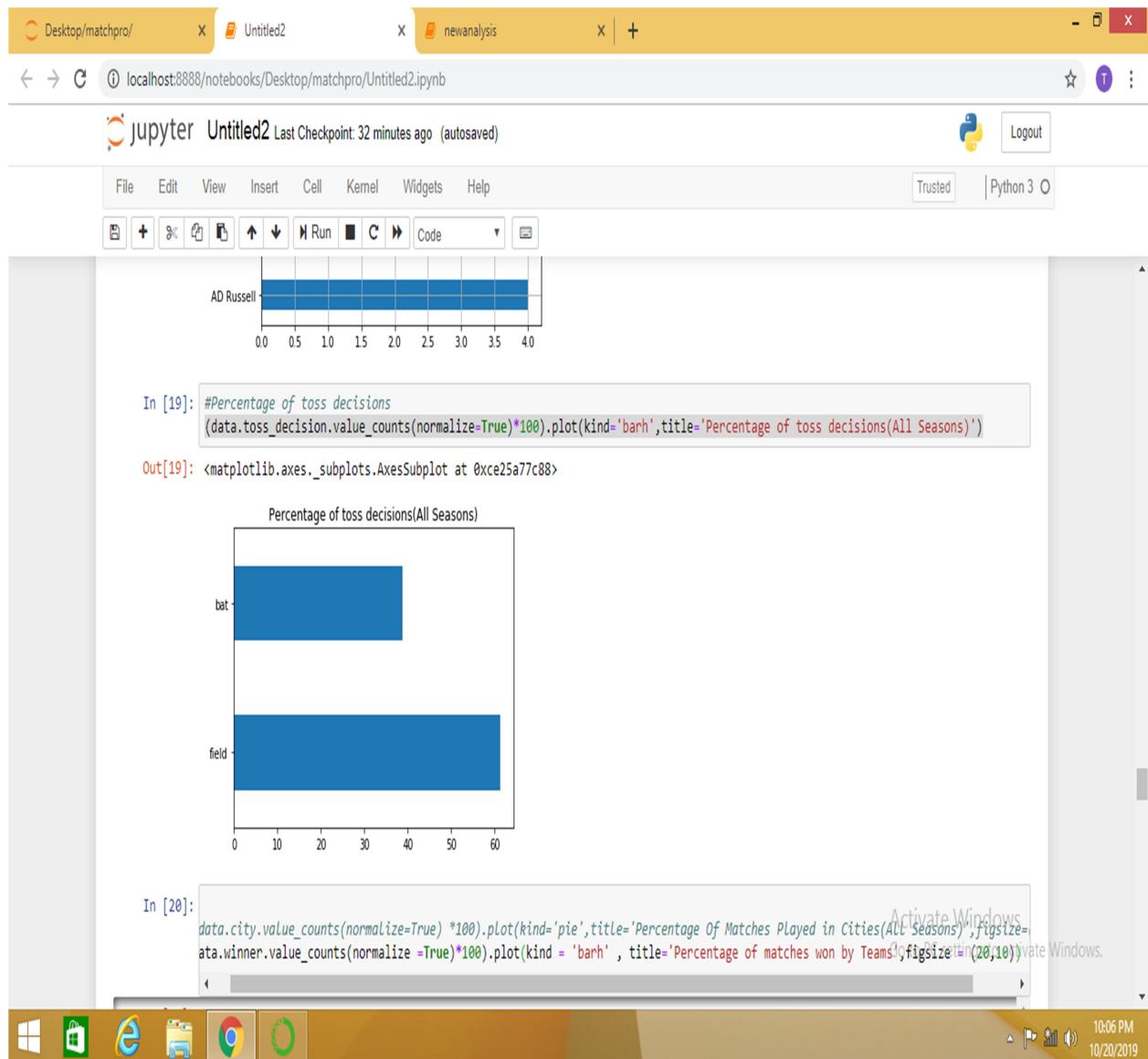
**Fig. Toss analysis**

```
In [10]: #By seaborn  
plt.subplots(figsize=(8,5))  
sns.barplot(x=data.toss_winner.value_counts().values,y=data.toss_winner.value_counts().index)
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0xce21041eb8>
```



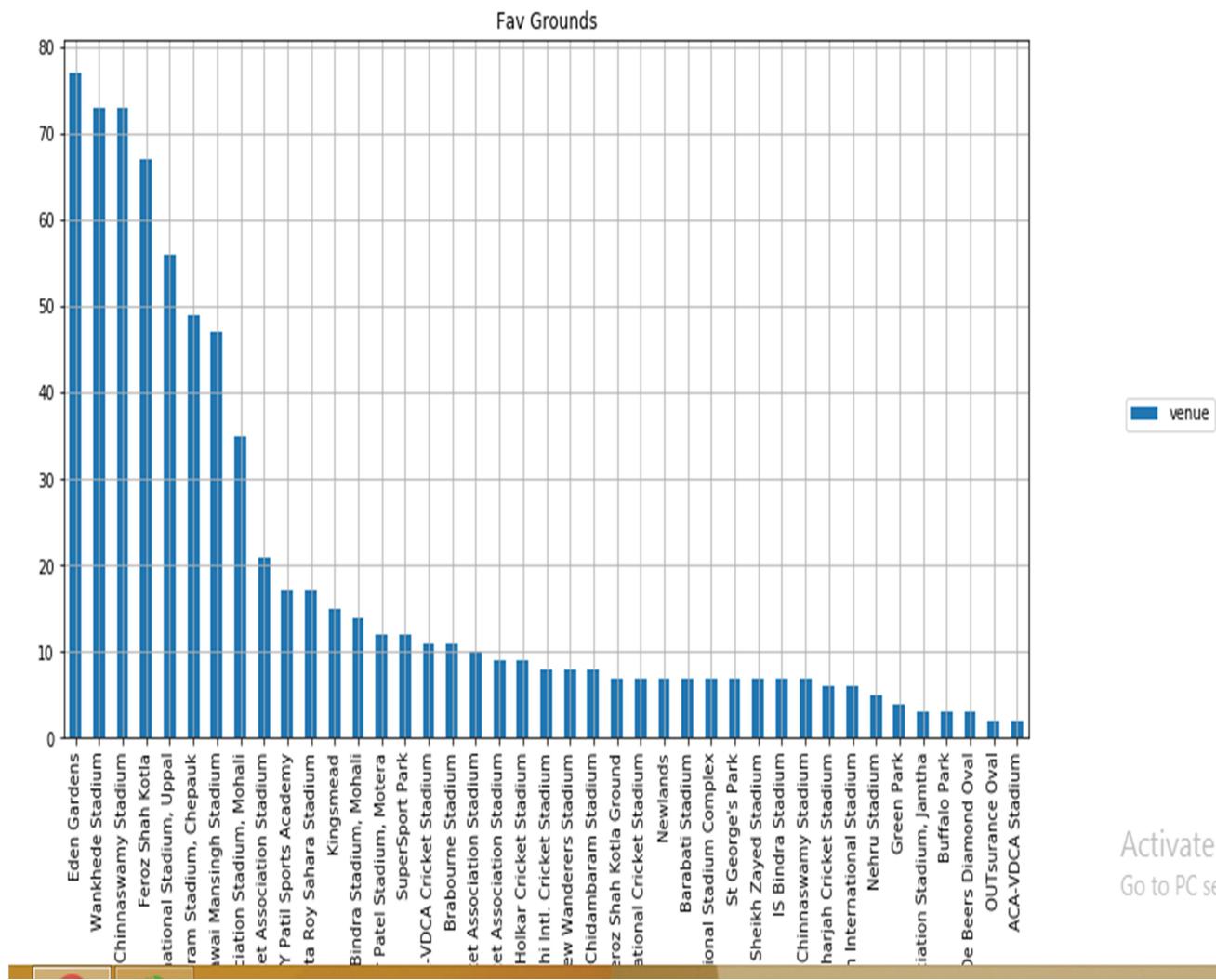
**Fig. Toss Winner**



**Fig. Percentage of Toss decision**

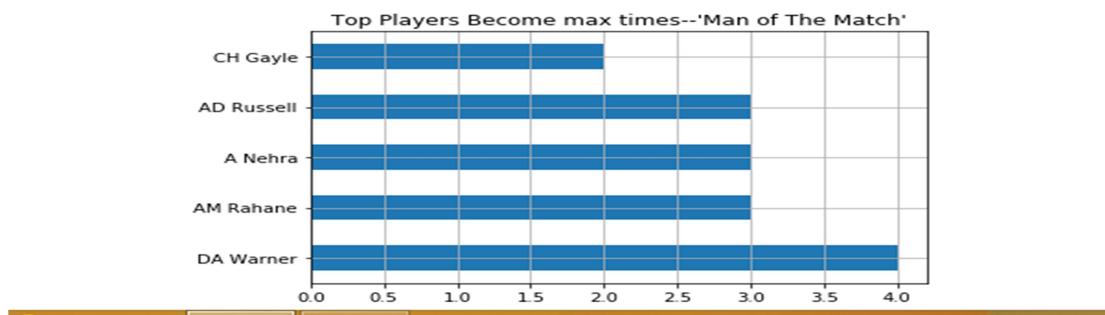
## ➤ Home Ground Advantage

This is another attribute which determines the score in the match. If you are playing in the home ground conditions everything would be in your hands like climatic factors, pitch nature and major role is played by the home crowd. Home Team gets better motivation.

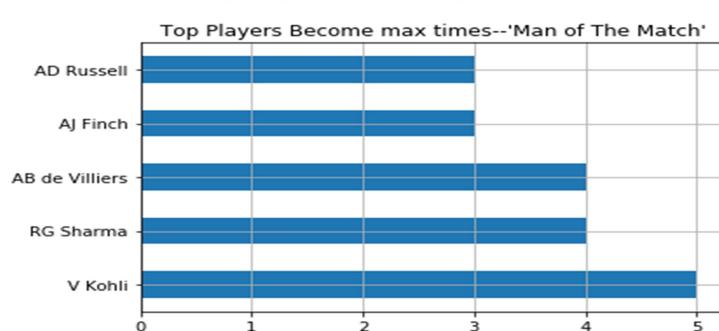


**Fig. Fav ground**

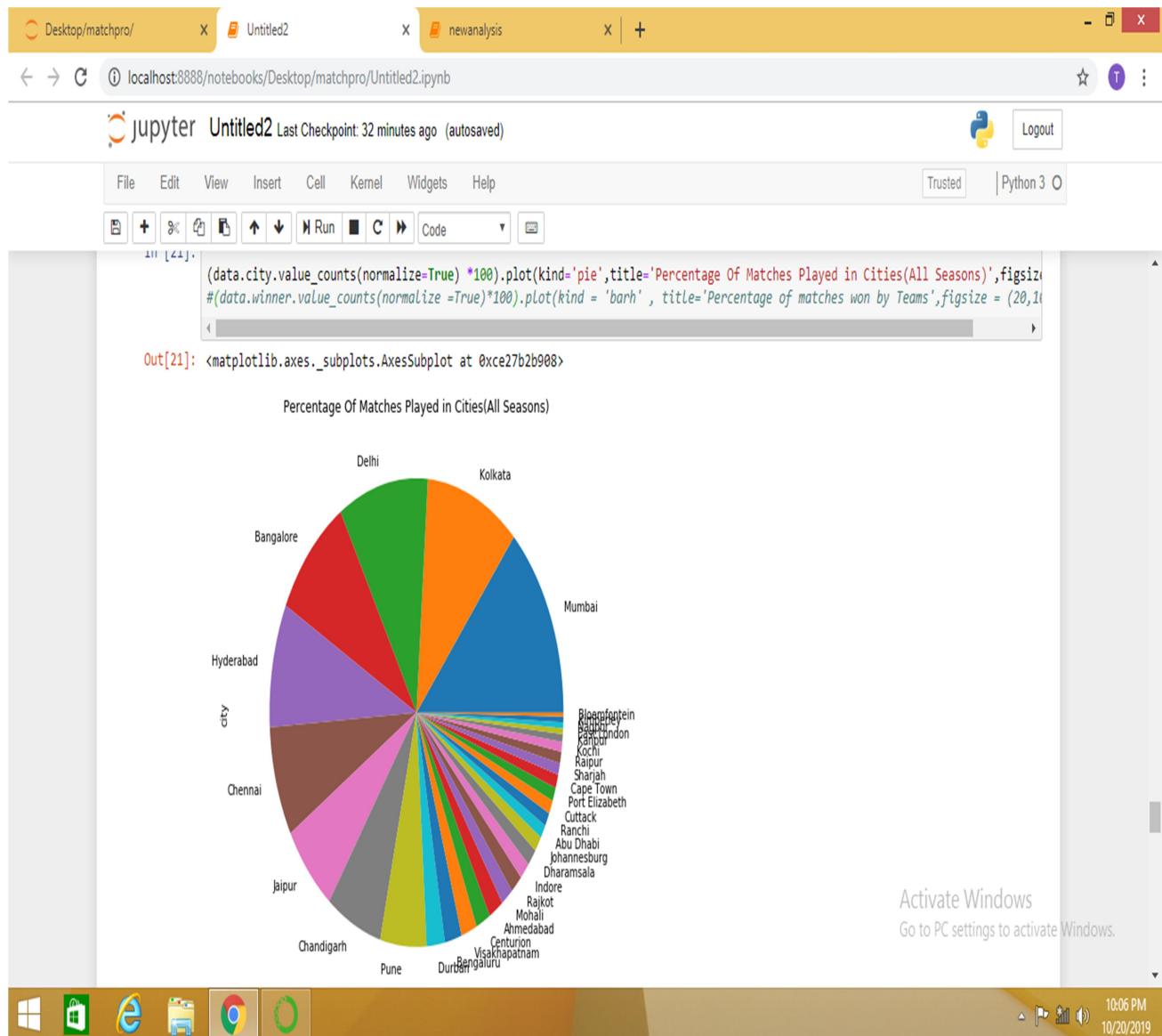
```
In [14]: d().plot(kind='barh',title="Top Players Become max times--\nMan of\nad().plot(kind='barh',title="Top Players Become max times--\nMan of\n\nOut[14]: <matplotlib.axes._subplots.AxesSubplot at 0xce23d6edd8>
```



```
In [15]: #data2015.player_of_match.value_counts().head().plot(kind='barh',title=\n#data2016.player_of_match.value_counts().head().plot(kind='barh',title=\n\nOut[15]: <matplotlib.axes._subplots.AxesSubplot at 0xce26f3b0b8>
```



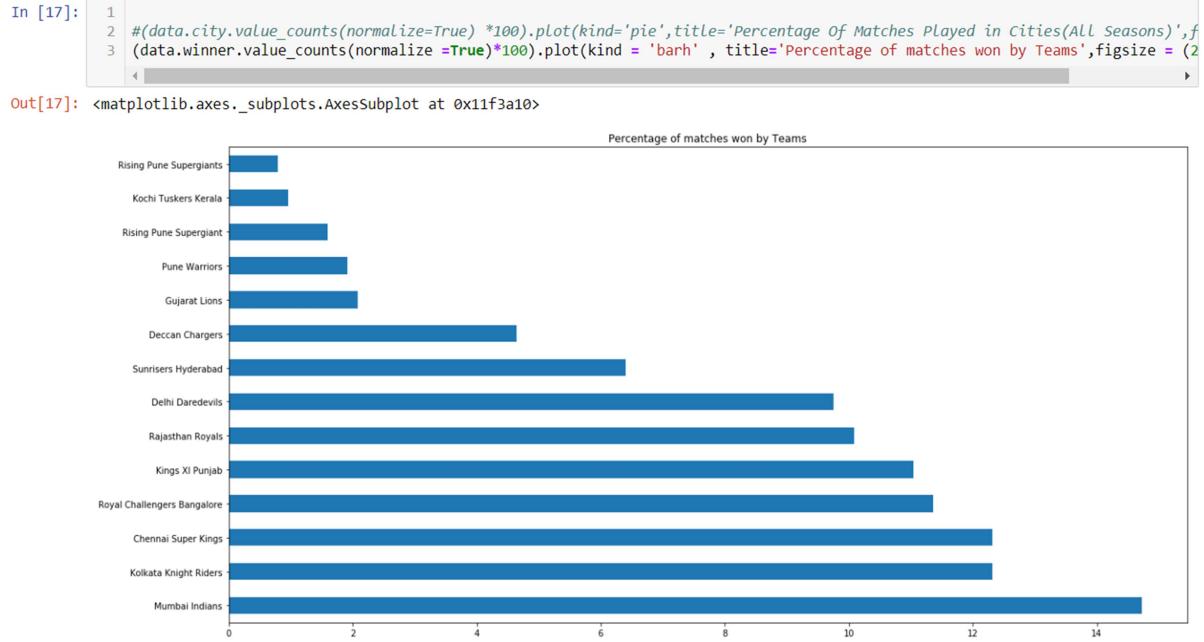
**Fig. Man of the Match**



**Fig. Percentage of Matches In Cities**

## ➤ Past Records

The past team performances can be considered to predict the outcome of a match. History of games at that venue how did the teams perform, performance at that specific venue, Performance against the specific opposition and experience at the specific venue.



**Fig. Matches Won by Team**

- **CONCLUSION**

The Fantasy Cricket has the problem of churning of customers within the duration of the season. This problem is therefore, addressed by using interactive models of Predictions where a user predicts the results of each game in order to be rewarded which would further help him strengthen his Fantasy squad. The project thus, aims not only to attract more users to this game that is Fantasy Cricket, but also aims at improving the general attraction to the Premier League. This happens because in a predictive model, a user makes a prediction on every game, and ends up watching that game to check if his prediction is going right Thus our project will not only improve the existing system of Fantasy Cricket, but will also augment the reach of Cricket in India.