

Shivnagar Vidya Prasarak Mandal’s

COLLAGE OF ENGINEERING

Malegaon (Bk.), Tal: Baramati, Dist: Pune, Pin: 413115.

**DEPARTMENT OF COMPUTER ENGINEERING**

A

**Project Report On**

**“Exploratory Data Analysis-Sports (Indian Premier League)”**

**Submitted By**

**Amruta Sambhaji khalate Jaydeep Hanumant Natakale**

**Vaishnavi Sandip Kulkurni Poonam Changdeo Bhosale**

**Under the Guidance**

**Prof. Mr. S.S.Nimbalkar**

***In the partial fulfillment of the fifth semester of Degree in Computer Engineering***

**Academic Year 2021-2022**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Title** | **Page No** |
| 1. | Abstract | 3 |
| 2. | Introduction | 4 |
| 3. | Objectives | 5 |
| 4. | Technology | 6 |
| 5. | Output | 7 |
| 6. | Conclusion | 12 |

**ABSTRACT**

Cricket is one of the famous outdoor sports that contain a large set of statistical data in real world. As IPL games rise in popularity, it is necessary to examine the possible predictors that affect the outcome of the matches.Several 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 performance. It focuses on measuring the outcome of Indian Premier League (IPL) matches by applying the existing data mining algorithms to the balanced as well as imbalanced dataset. This model is very much popular in predictive modelling. Currently, in Twenty-Twenty (T20) cricket matches first innings score is predicted on the basis of current run-rate which can be calculated as the amount of runs scored per the number of over’s bowled. It includes factors like number of wickets fallen, venue of the match, toss and predicts the score in each of the innings and finally the winner of the match using Random Forest algorithm.Prediction of IPL2020 are done on the basis of survey, and analysis.

**INTRODUCTION**

The use of analytical methods in various aspects of cricket including IPL results prediction is very important. There is a huge demand for the algorithm that best predicts the result of cricket because of its popularity and huge amount of money involved in the game. Thus the analysis of IPL results becomes more important. Prediction of outcome of a match using machine learning algorithms is an important aspect in cricket. Records of the past performance of players and other related data can be analysed to create models that predicts the winning team. This model can be created using the machine learning algorithms such as random forest algorithm. Data mining tools predict the future trends and behaviours, which gives an opportunity to predict the outcome of an IPL (Indian Premier League) match using algorithms. Algorithms have been applied to the IPL dataset and the knowledge from each algorithm has been obtained and analyzed thoroughly as the results are obtained with good accuracy performance. Cricket is one of the most popular sports.

The International Cricket Council (ICC) listed out 106 cricket playing nations representing 10 belongs to the full members, 37 of them are associates, and the remaining 59 are considered to be affiliate members. The game of cricket is played in various formats, i.e., One Day International, T20 and Test Matches. The Indian Premier League (IPL) is a Twenty-20 cricket tournament league established with the objective of promoting cricket in India and thereby nurturing young and talented players. The teams for IPL are selected by means of an auction. Players’ auctions are not a new phenomenon in the sports world. However, in India, selection of a team from a pool of available players by means of auctioning of players was done in Indian Premier League (IPL) for the first time. This in turn, is dependent on the complex rules governing the game, luck of the team (Toss), the ability of players and their performances on a given day. A way of predicting the outcome of the matches between various teams can aid in the team selection process. The tool presented can be used to evaluate in the performance of players. This tool provides a visualization of players’ performance. The result has been predicted using the algorithm approaches and have analyzed the results of the IPL match using the above approaches. Some of the popular variables considered in cricket literature are home-field advantage, coin-toss result, bat-first or second. Thus we measure the outcome of the IPL matches using the random forest algorithm algorithms.

**OBJECTIVES**

* To find the team that won the most number of matches in a season.
* To find the team that lost the most number of matches in a season.
* Does winning toss increases the chances of victory.
* To find the player with the most player of the match awards.
* To find the city that hosted the maximum number of IPL matches.
* To find the most winning team for each season.
* To find the on-field umpire with the maximum number of IPL matches.
* To find the biggest victories in IPL while defending a total and while chasing a total.

**TECHNOLOGY**

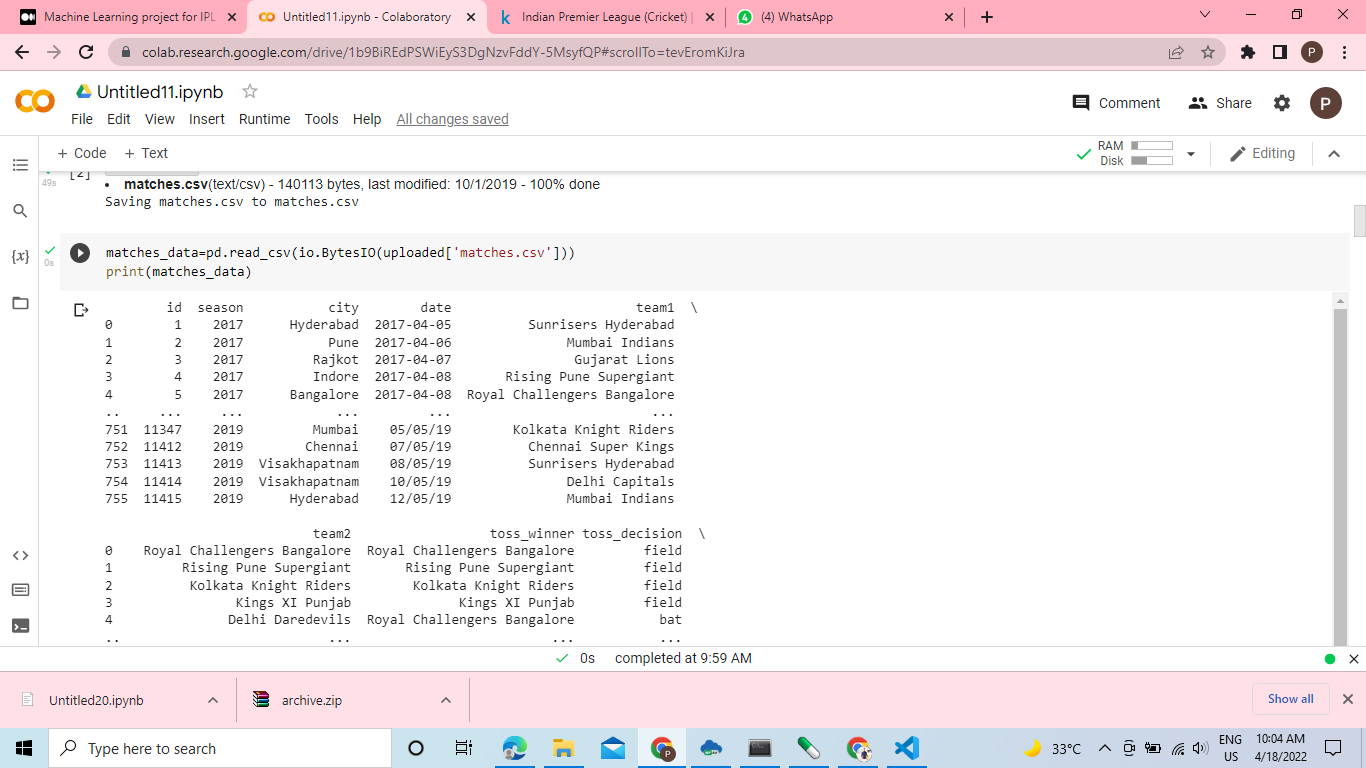
Python

Random Forest Algorithm

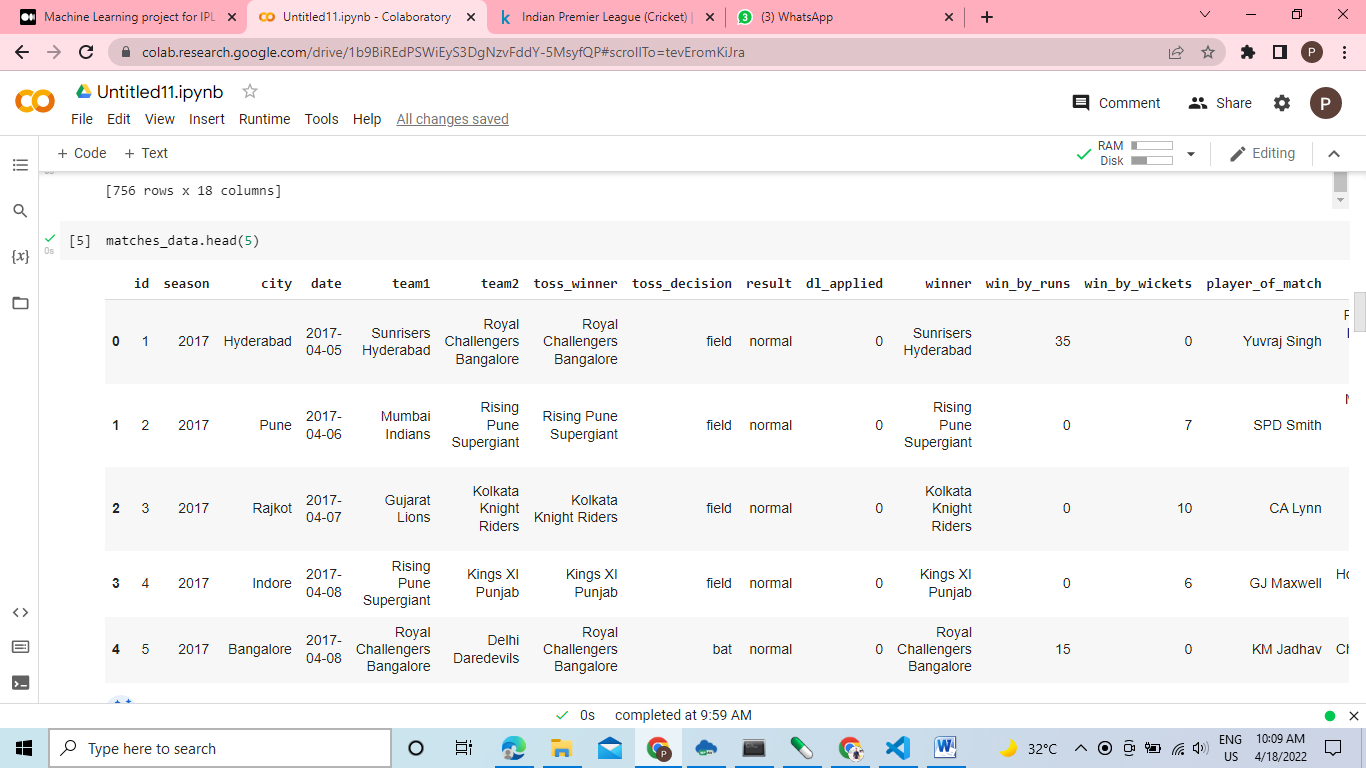
Logistic Regression

**OUTPUT**

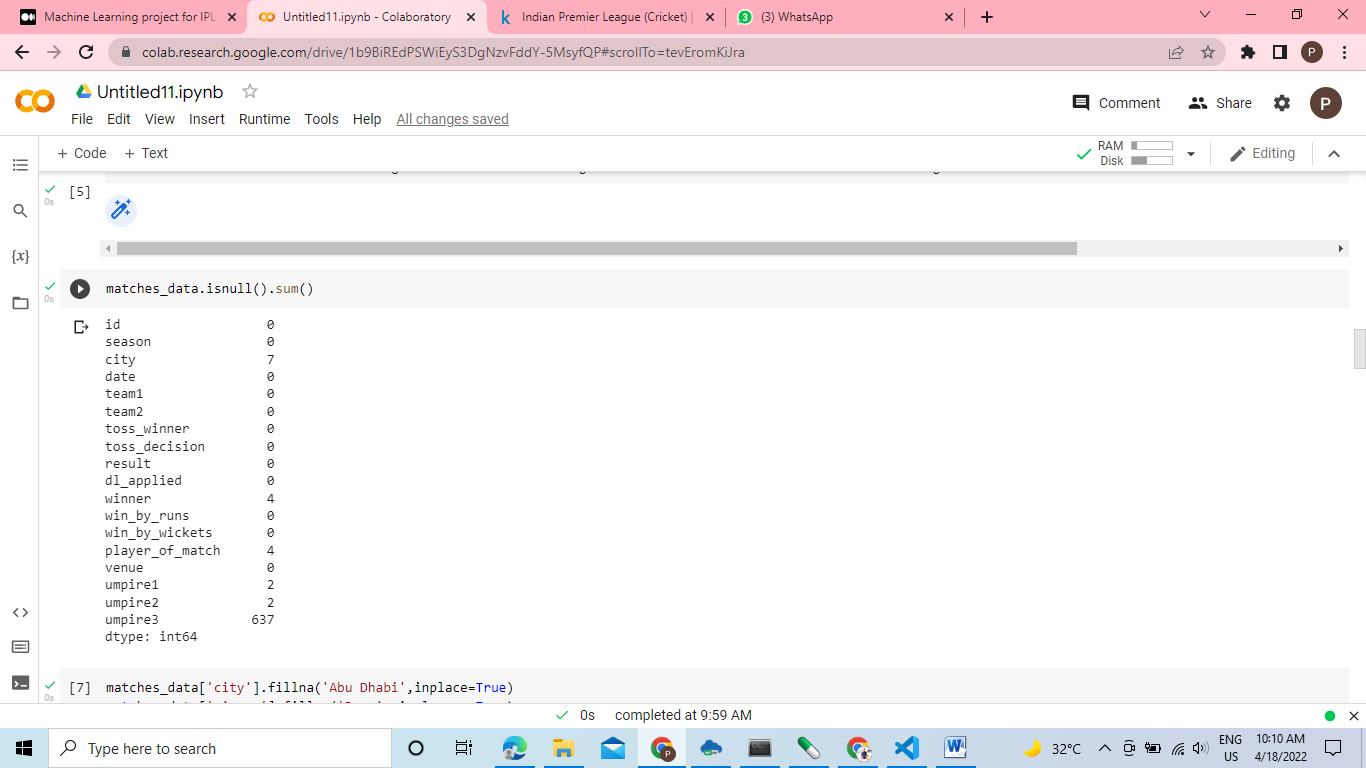
**Print Data:**



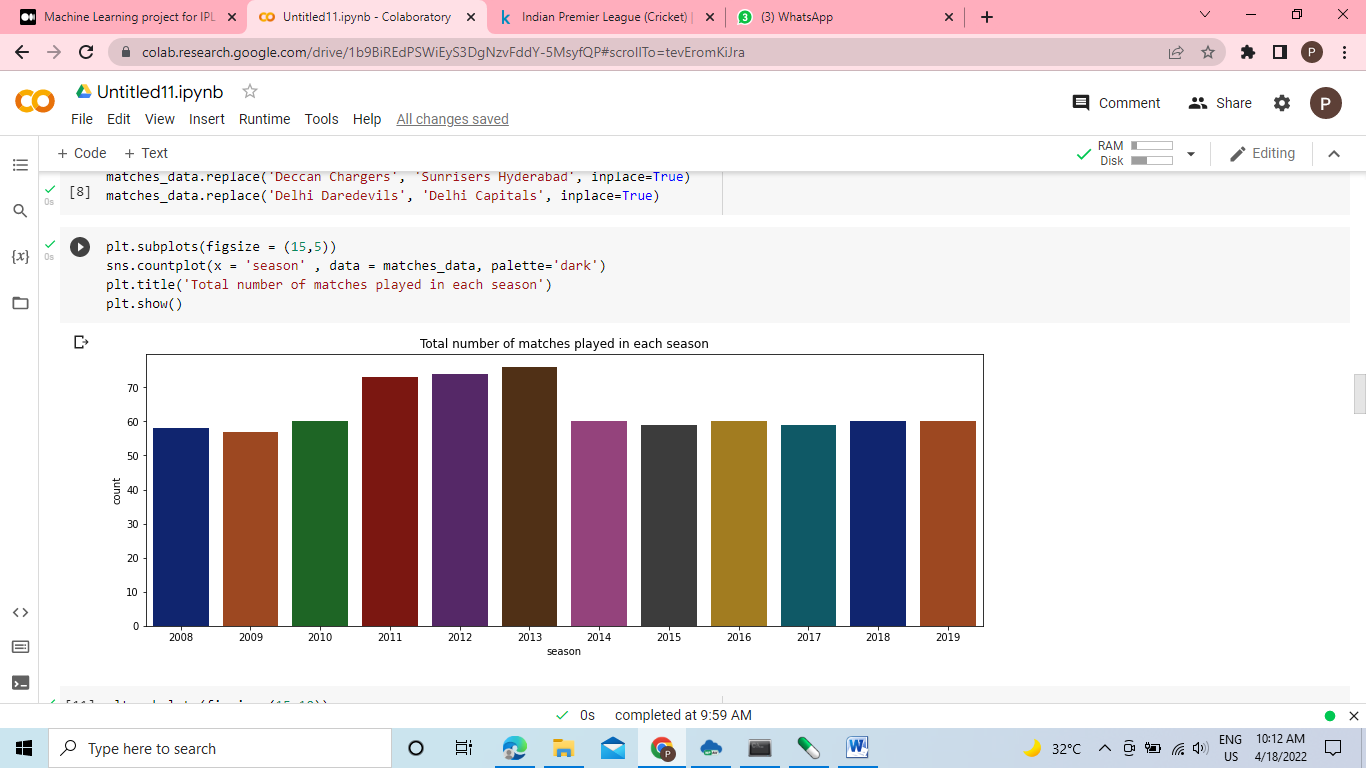
**Print head:**



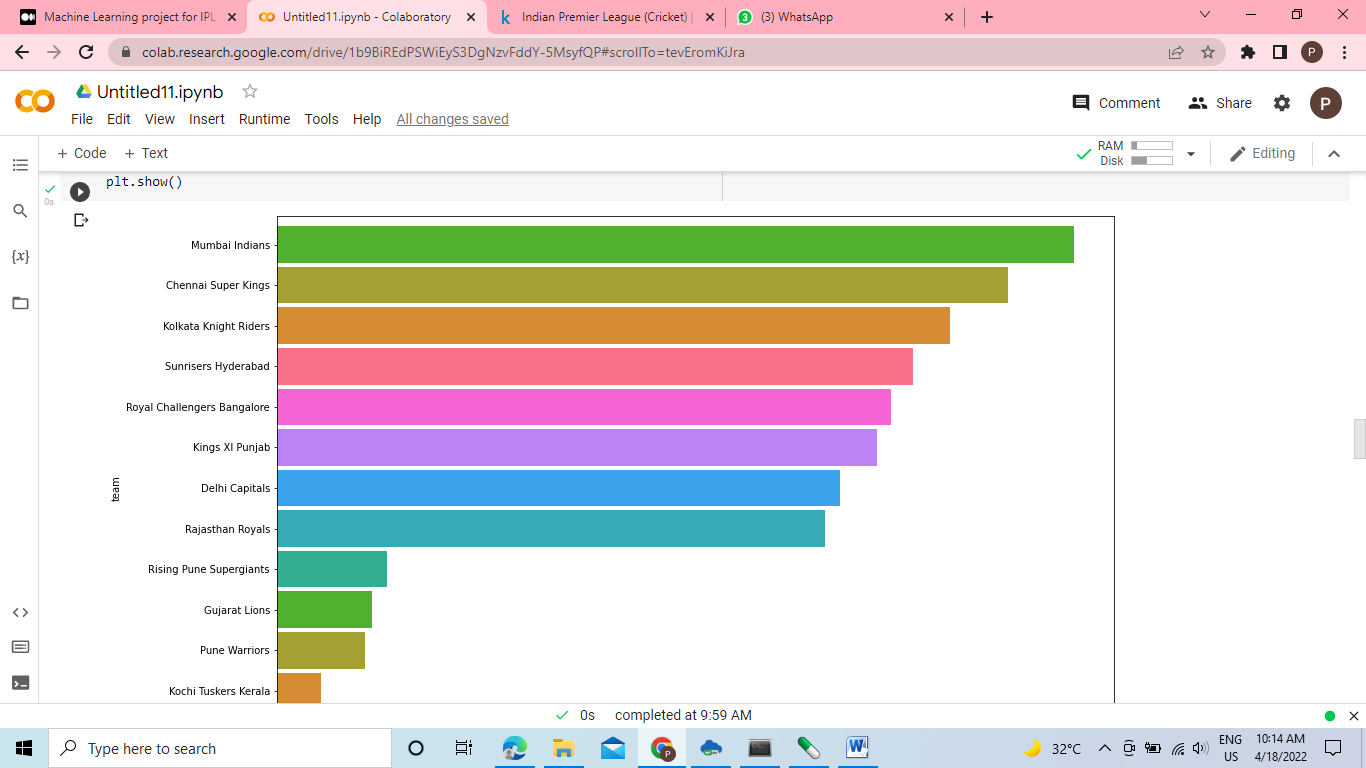
**Print Null Sum:**



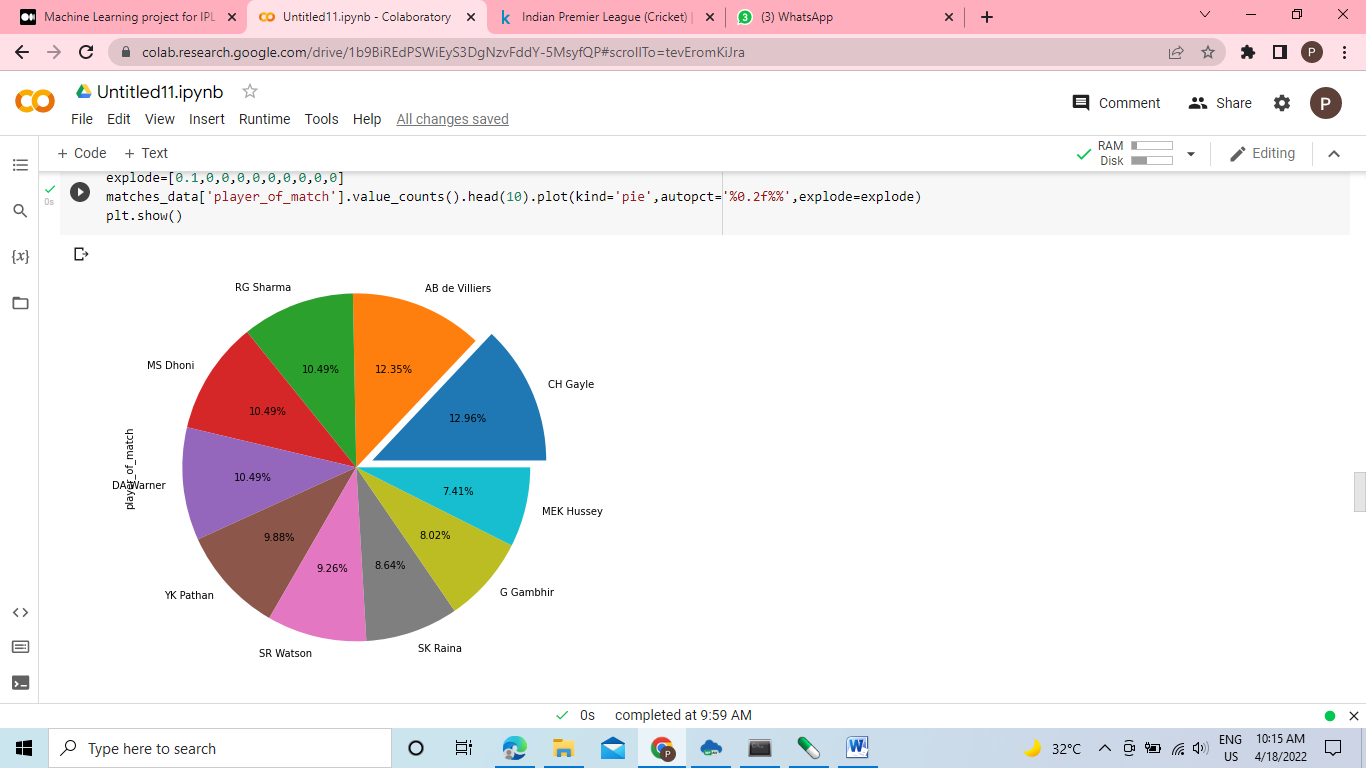
**Print Total Number Of Matches Played in each Season:**



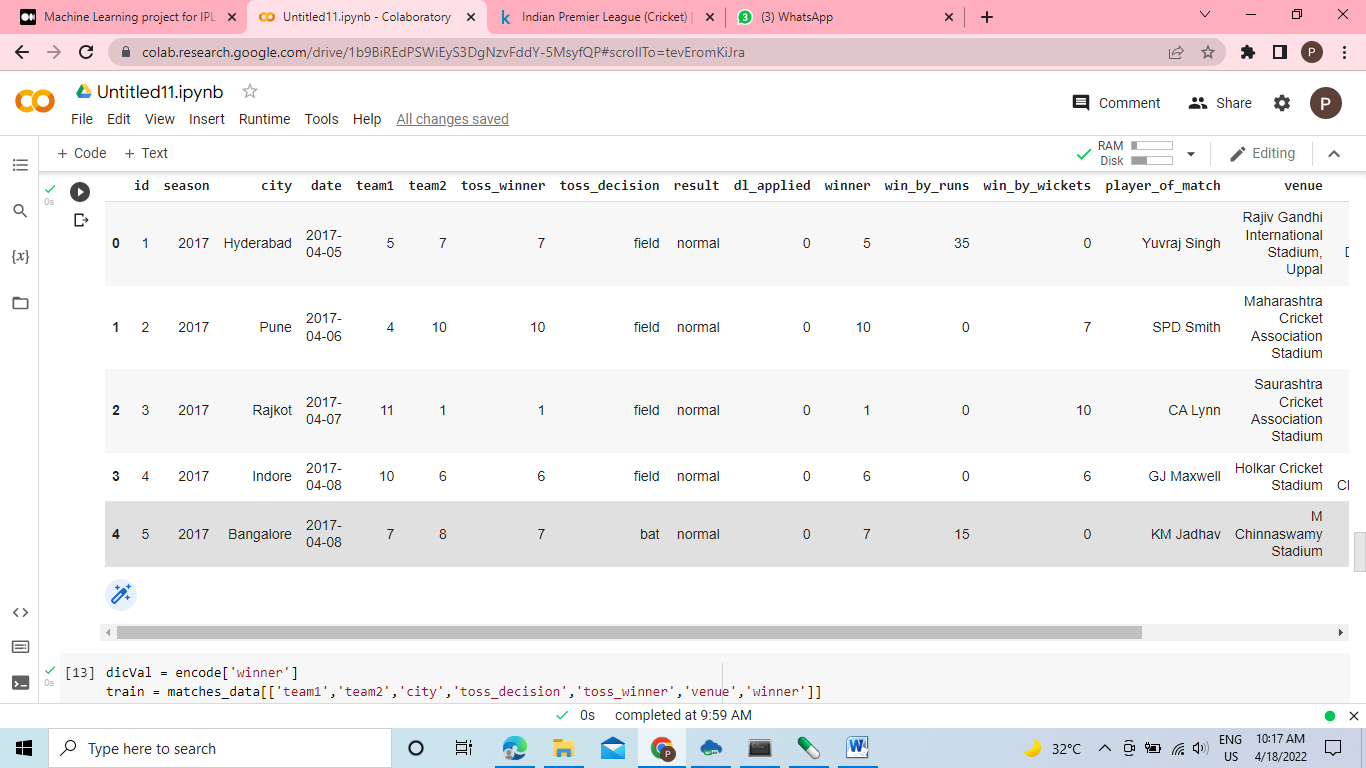
**Analysis How Much Time Team Winner:**



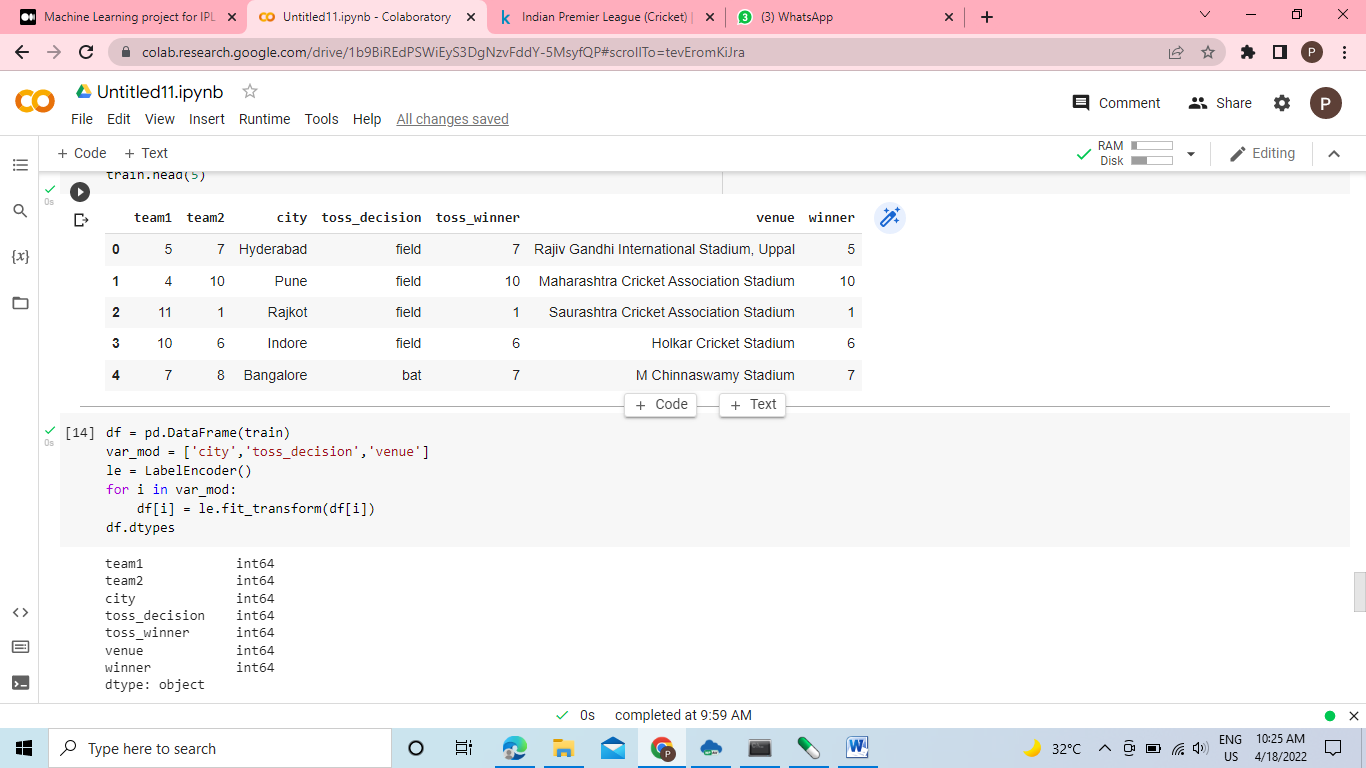
**Print Top Player Of Matches Winner:**



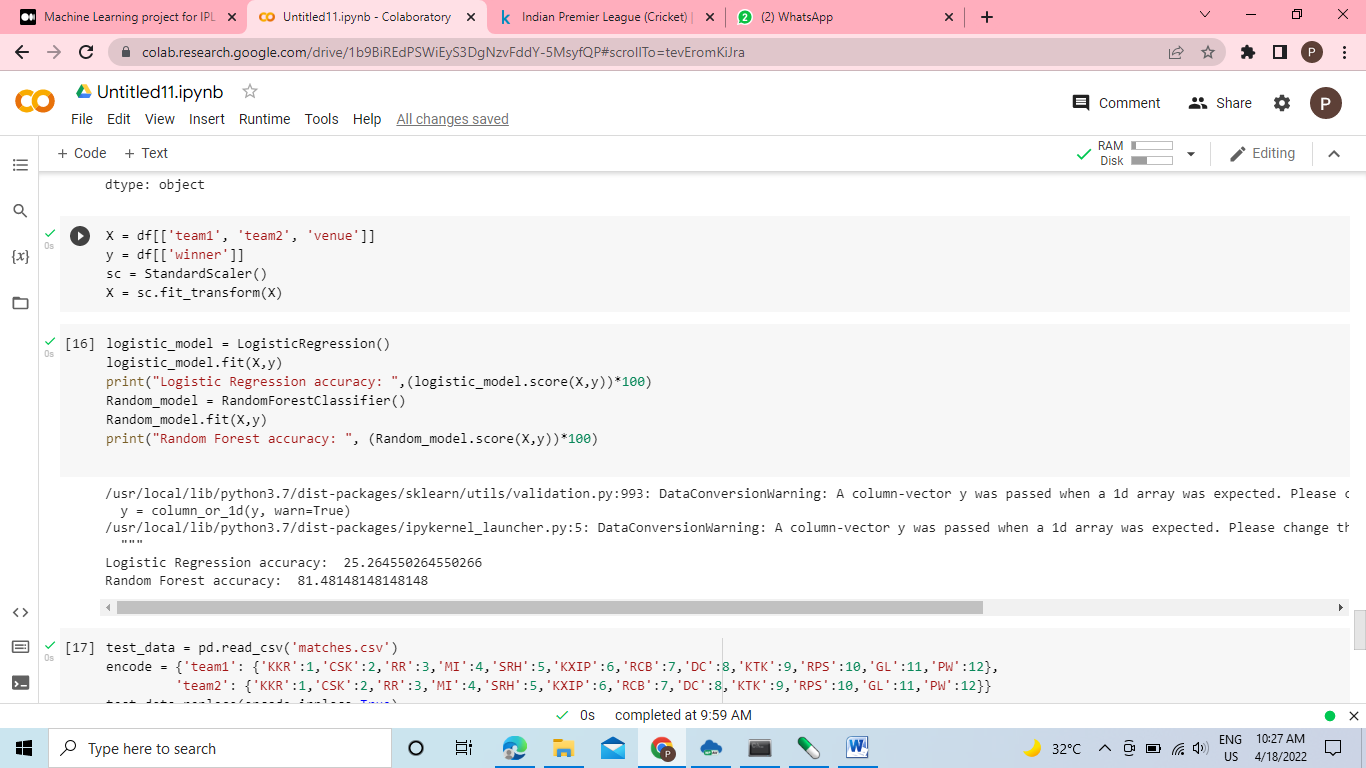
**Print Replace Data :**

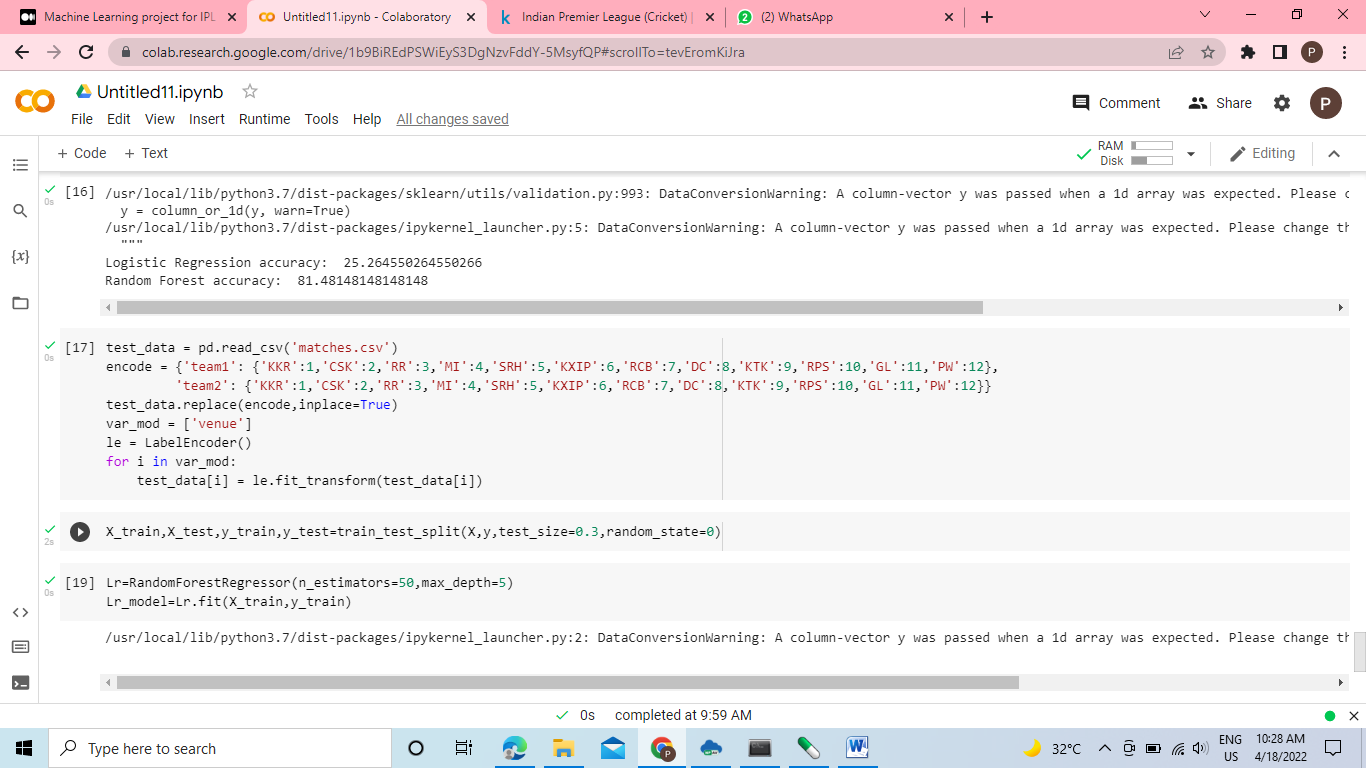


**Encode and LabelEncoder Data:**



**Fit data and Logistic Regression Algorithm:**





**CONCLUSION**

Predicting a winner in a sport such as cricket is especially challenging and involves very complex processes. But with the introduction of random forest algorithm, this can be made much easier and simpler. In this various factors have been identified that contribute to the results of the Indian Premier League matches. Factors that have a major impact on the outcome of an IPL match include the teams playing, the venue, the city, the toss winner and the toss decision. We have analyzed IPL datasets and predicted game results based on player performance. The methods used in the work to obtain the final test are Logistic regression, Support Vector Machine (SVM), Decision tree, Random Forest classifier and K-nearest neighborhood. Random Forest classification (RFC) outperforms the other algorithm. As for the scope of the future, the focus can be on each player’s performance and evaluate that on a regular basis for the season. His ratings for bowling and batting can also be predicted. There can be a chance to predict the man of the match for the two teams.