**ANALYSIS OF INDIAN LOK SABHA ELECTION 2019**

**AND PREDICTION OF WINNERS**

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED SPECIALIST**

**IN**

**DATA SCIENCE & ANALYTICS**

submitted by

**ICT\_DS Batch 1, Group 3**

Sandra Anna Shaji

Bineeth Mathew

Basil Chacko Mathew

Neenu Sebastian



**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**Feb 2022**

**List of Figures**

Figure 1: Graph of crime count in different states

Figure2: Graph of state Vs criminal case

Figure3: Graph of educational qualification of the candidates

Figure4: Bar graph education Vs crime cases

Figure5: Pie chart of male Vs female candidates

Figure6: State wise candidate with crime cases

Figure7: Bar graph of category growth

Figure8: Bar graph of candidate allocation in Lok Sabha election

Figure9: Bar graph of party Vs candidates with crime cases

Figure10: Bar graph of age Vs crime cases

Figure11: Bar graph of gender Vs crime

Figure12: Bar graph of state Vs total votes

**List of Abbreviations**

**Table of Contents**

1. **Problem Definition**
   1. Overview
   2. Problem Statement
2. **Introduction**
3. **Literature Survey**
4. **Approach and Methodology**

4.1 Exploratory Data Analysis

**Abstract**

We have picked a dataset related to our ELECTION: Indian Election 2019 for our data science project. In this dataset, we have 2018 candidates who are participated in 2019 Lok Sabha election from 543 parliamentary constituencies of India. Data set comprises the details of each candidate from 29 states the country. And it also includes there corresponding part’s symbol, candidates gender, age, category, education, assets, number of votes.

Here our target variable is “winner”. The information was gathered in order to create prediction models that would predict the winner from each constituency. For this prediction we use regression models.

The following is how our inquiry will proceed: gain a thorough understanding of the datasets, prepare and ranked data, analysis the data, model the data and provide conclusion. To better understand the data’s patterns, trends we want to employ descriptive analytics. We can also find the party which has won the most constituencies. We indent to apply a verity of machine learning techniques to discover the best model for our dataset.

**1. Problem Definition**

**1.1 Overview**

Everything about Indian general elections is colossal - the Economist magazine once compared it to a "lumbering elephant embarking on an epic trek". The number of voters is bigger than the population of Europe and Australia combined. India's Centre for Media Studies estimated parties and candidates spent some $5bn (£3.8bn) for the 2014 elections. A total of 10 lakh polling stations were set up in 2019 as compared to around nine lakh in 2014. Looking at all these stats it is only going to get bigger in future. Now let's just jump into the analysis of 2019 Lok Sabha election data.

The purpose of using this specific language is due to its versatility, vast libraries (Pandas, Numpy , Matplotlib, etc.), speed limitations, and ease of learning. We will be analyzing large election data sets in this project which can not be easily analyzed in other tools as compared to python. Python does not have it’s limitation to only data analytics but also in many other fields such as Artificial intelligence, Machine learning, and many more.

**1.2 Problem Statement**

* The Dataset is based on Indian Lok Sabha Election.
* 539 Constituency participation data in this dataset.
* There were 2263 candidates who contested 2019 Lok Sabha Election.
* Minimum age of the candidates was 25 whereas maximum age was 86.
* Average age of all the candidates who contested election was 52.
* 19367 postal votes were casted in the election.

**2 . Introduction**

Lok Sabha elections are held every five years in order to elect the next Prime Minister of India. People vote for their preferred party and the candidate of the winning party goes on to represent the Prime Minister of our country. The Lok Sabha elections are also called general elections. The Lok Sabha has a maximum strength of 552 members, with two nominations by the President of India from the Anglo-Indian community. Twenty of the members are from union territories while 530 members are represented from the state.  
  
Any party which manages at least 273 seats is considered victorious and is eligible to form the central government. The parties are at liberty to announce their prime ministerial candidate either before the elections or after the victory. The Prime Ministerial candidate is elected by the party officials.

The **Lok Sabha**, or **House of the People**, is the [lower house](https://en.wikipedia.org/wiki/Lower_house)  of [India](https://en.wikipedia.org/wiki/India)’s [bicameral](https://en.wikipedia.org/wiki/Bicameralism) [Parliament](https://en.wikipedia.org/wiki/Parliament_of_India), with the [upper house](https://en.wikipedia.org/wiki/Upper_house) being the [Rajya Sabha](https://en.wikipedia.org/wiki/Rajya_Sabha). [Members of the Lok Sabha](https://en.wikipedia.org/wiki/Member_of_Parliament,_Lok_Sabha) are elected by an adult [universal suffrage](https://en.wikipedia.org/wiki/Universal_suffrage) and a [first-past-the-post](https://en.wikipedia.org/wiki/First-past-the-post) system to represent their respective [constituencies](https://en.wikipedia.org/wiki/List_of_constituencies_of_the_Lok_Sabha), and they hold their seats for five years or until the body is dissolved by the [President](https://en.wikipedia.org/wiki/President_of_India) on the advice of the [council of ministers](https://en.wikipedia.org/wiki/Union_Council_of_Ministers). The house meets in the Lok Sabha Chambers of the  [Sansad Bhavan](https://en.wikipedia.org/wiki/Parliament_House_(India)), [New Delhi](https://en.wikipedia.org/wiki/New_Delhi).

The maximum membership of the House allotted by the [Constitution of India](https://en.wikipedia.org/wiki/Constitution_of_India) is 552 (Initially, in 1950, it was 500). Currently, the house has 543 seats which are made up by the election of up to 543 elected members and at a maximum. Between 1952 and 2020, [2 additional members](https://en.wikipedia.org/wiki/Anglo-Indian_reserved_seats_in_the_Lok_Sabha) of the [Anglo-Indian](https://en.wikipedia.org/wiki/Anglo-Indian) community were also nominated by the President of India on the advice of the [Government of India](https://en.wikipedia.org/wiki/Government_of_India), which was abolished in January 2020 by the [104th Constitutional Amendment Act, 2019](https://en.wikipedia.org/wiki/One_Hundred_and_Fourth_Amendment_of_the_Constitution_of_India). The Lok Sabha has a seating capacity of 550.

A total of 131 seats (24.03%) are reserved for representatives of [Scheduled Castes (84) and Scheduled Tribes (47)](https://en.wikipedia.org/wiki/Scheduled_Castes_and_Scheduled_Tribes). The quorum for the House is 10% of the total membership. The Lok Sabha, unless sooner dissolved, continues to operate for five years for time being from the date appointed for its first meeting. However, while a [proclamation of emergency](https://en.wikipedia.org/wiki/State_of_Emergency_in_India) is in operation, this period may be extended by [Parliament](https://en.wikipedia.org/wiki/Parliament_of_India) by law or decree.

One of the most critical ways that individuals can influence governmental decision-making is through voting. We know that everyone has the right to vote in our country. But many people are not aware of politics. So by this project, we can learn about the different political parties, their background history as well as their recent success and failure in the Lok Sabha election 2019 in India. Unfortunately, we have found few candidates with criminal history also. So through this data analysis, we can aware of the candidate's history as well as the nature of the political party. We can learn about the winning party and their success in 2019.

We specifically want to mention that it was an unbiased analysis. Here we have not supported any specific party.

# Dataset:

This Dataset is based on the Lok Sabha 2019 in India. There are a total of 2263 rows and 21 columns in this dataset. By using this dataset this data analysis project is created.

Here we use google COLAB or Jupyter to run these codes and analysis the dataset but you can use other platforms also to run the code.

**3. Literature Survey**

**4. Approach and Methodology**

4.1 **Exploratory data analysis:**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns , to spot anomalies , to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

# Bar Graph of crime Count in different states:

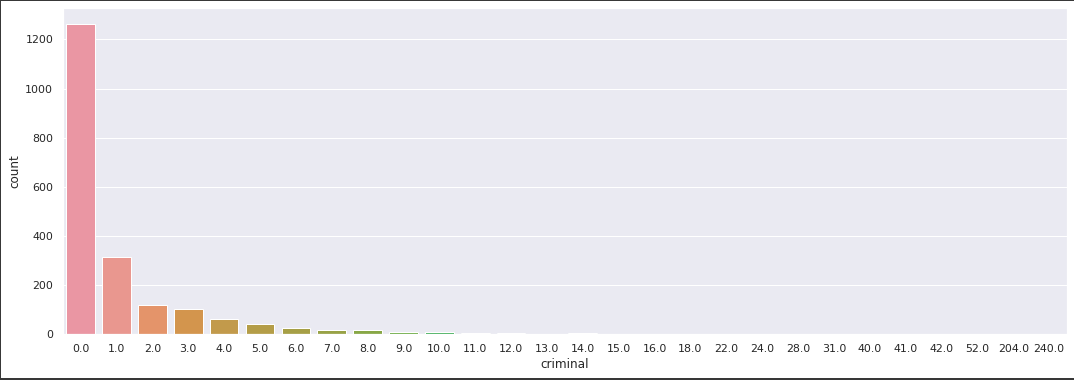


Fig 1: Graph of crime Count in different states

From the description given below, we can see that the mean of the crime among contestants is 1.45 where for the minimum crime , 25% and 50% of contestants did not make any crime but sadly in 75 % of total candidates the crime rate became 1.0. More surprisingly the maximum crime conceived by a person is 240, which’s huge.

# Line Graph of State vs Criminal Case:

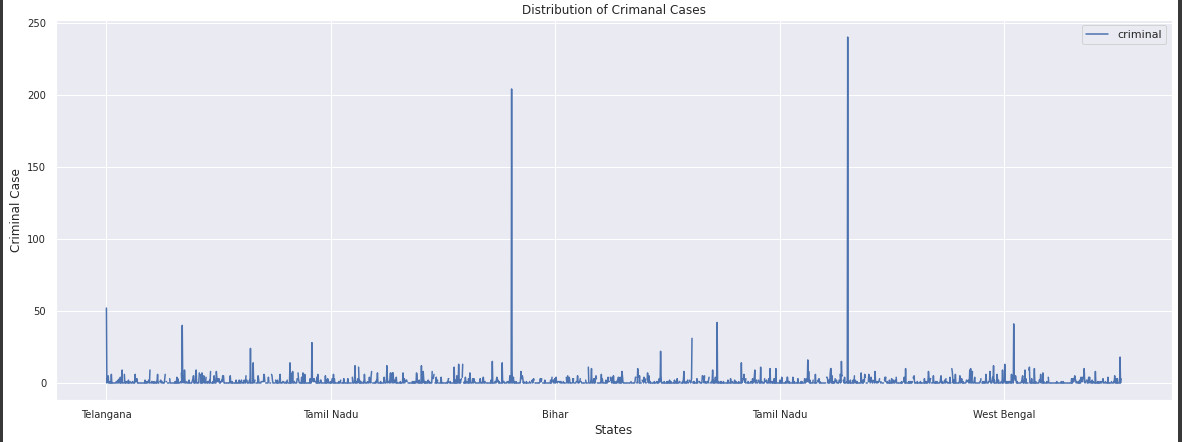


Fig 2: Graph of State vs Criminal Case

From this graph and the below description, we can see that the maximum no of criminal cases done by a single person is 240.

# The Educational Qualification of the Candidates:

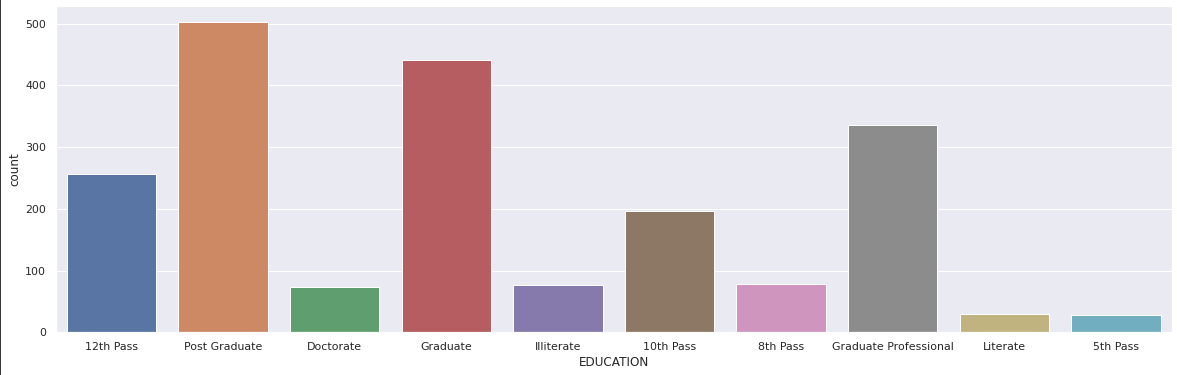


Fig 3: Graph of Educational Qualification of the Candidates

After analysing the graph, we can see that there are two columns of class VIII pass and class V pass. But we believe the minimum qualification to be called literate is X pass. So we convert all V pass and VIII candidates as illiterate.

We can see that the number of postgraduate candidates in India is maximum(officially). So this is a positive site from the educational point of view.

# ****Education vs Crime Cases Bar graph:****

This graph represents the candidate's educational qualification vs criminal cases they have. Now we are aware of their previous criminal background with their educational qualification.

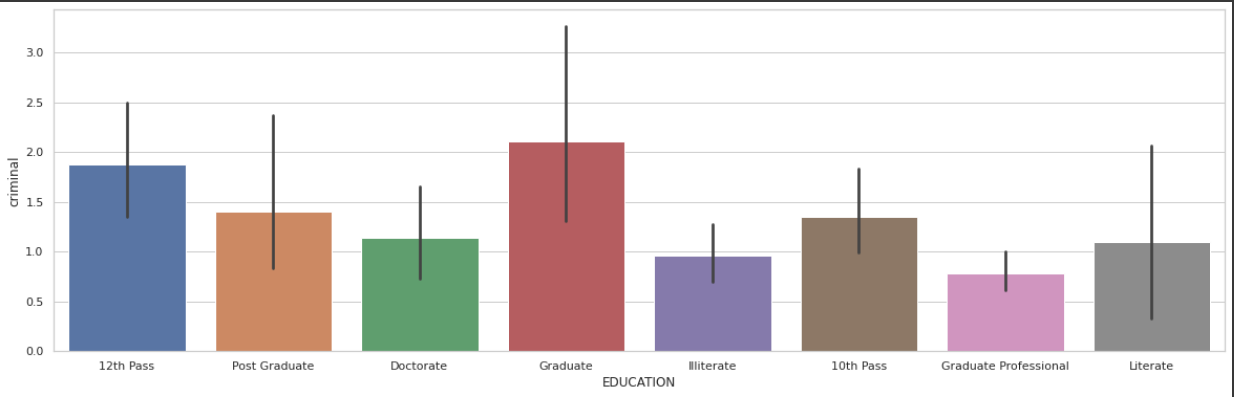


Fig 4: **Education vs Crime Cases Bar graph**

We can analyze from the graph that Graduate and 12th Pass criminal candidates are maximum. Especially we want to mention that a single graduate person has done 240 crimes.

# Pie chart of Male vs Female candidates:

This graph represents the male and female candidates who participated in Lok Sabha 2019.

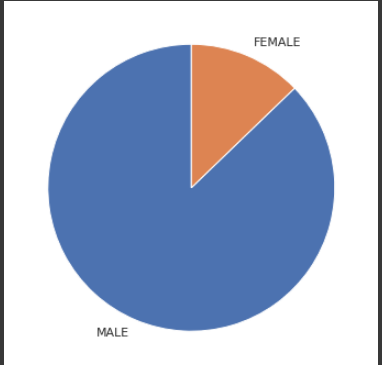


Fig 5: Pie chart of Male vs Female candidates

From this pie chart, we can see that the number of male candidates is much greater than the number of female candidates.

# State-wise Candidates with Crime Cases:

This is the bar graph of state-wise criminal case contestants and state-wise criminal case winners. The number of candidates with criminal cases is maximum in Bihar, Kerala, Maharashtra, West Bengal, Uttar Pradesh states.

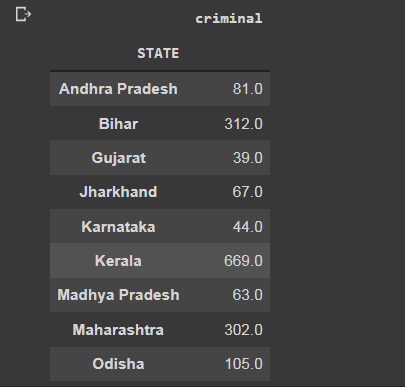
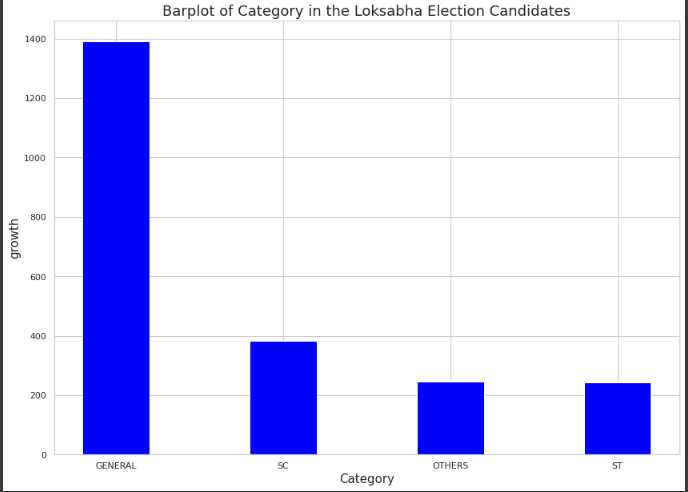


Fig 6: State-wise Candidates with Crime Cases

Here we can see the crime case across the states of candidates and winners. Here the maximum height of the bar graph is showing in the state of Kerala but West- Bengal, Uttar Pradesh, and Telangana are not far behind.

# Bar Graph of category Growth:

Here we calculating the number of SC, ST, and GENERAL candidates in the Lok Sabha election 2019.

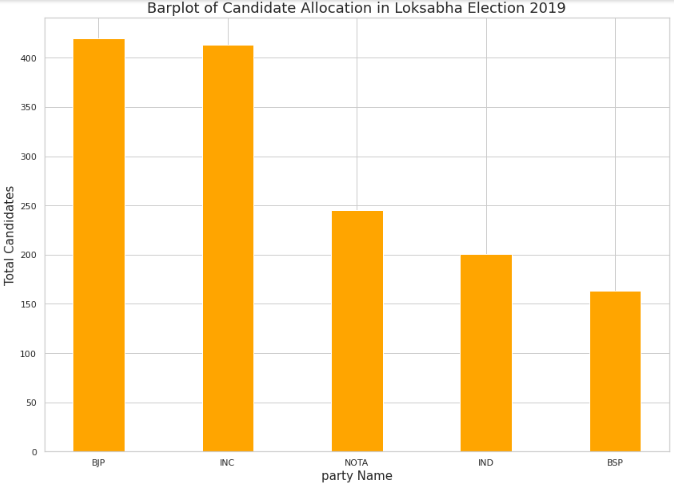
****

**Fig 7:** Bar Graph of category Growth

From the graph, we can see that the number of general candidates is maximum in India. The difference between general and other categories is very high.

# Bar Graph of Candidate Allocation in Lok Sabha Election 2019:

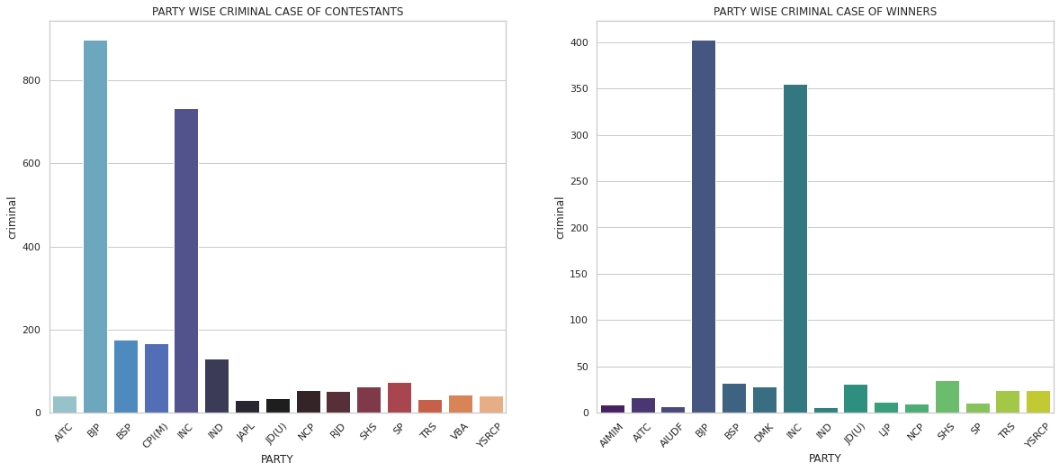
Here we are counting the total number of allocation of candidates for different parties in different constituencies in India.

****

**FIG 8:** Bar Graph of Candidate Allocation in Lok Sabha Election.

# Bar Graph of Party vs Candidates with Crime Case:

Here we are calculating the criminal case candidates in different parties. From that knowledge, we can aware of the criminal cases of the different parties.

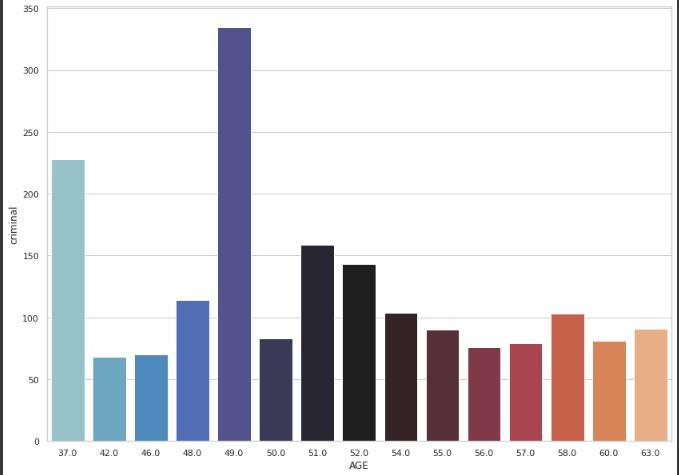


**Fig 9:** Bar Graph of Party vs Candidates with Crime Case

From the above diagram, we can see that the BJP and Congress parties have the maximum number of criminal cases in India. This is because that, these two parties are all India-based whereas most of the other parties are regional parties.

# Bar Graph of Age vs Crime Cases:

From this bar graph can know about the criminal cases of the candidates of different age group.

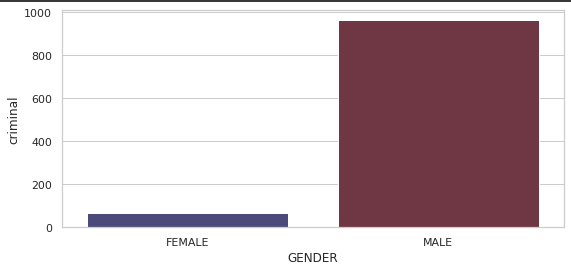
****

**Fig 10:** Bar Graph of Age vs Crime Cases

From the graph, we can notice that the criminal cases history is maximum at the age of 49,37, and 51.

# Bar Graph of Gender vs Crime:

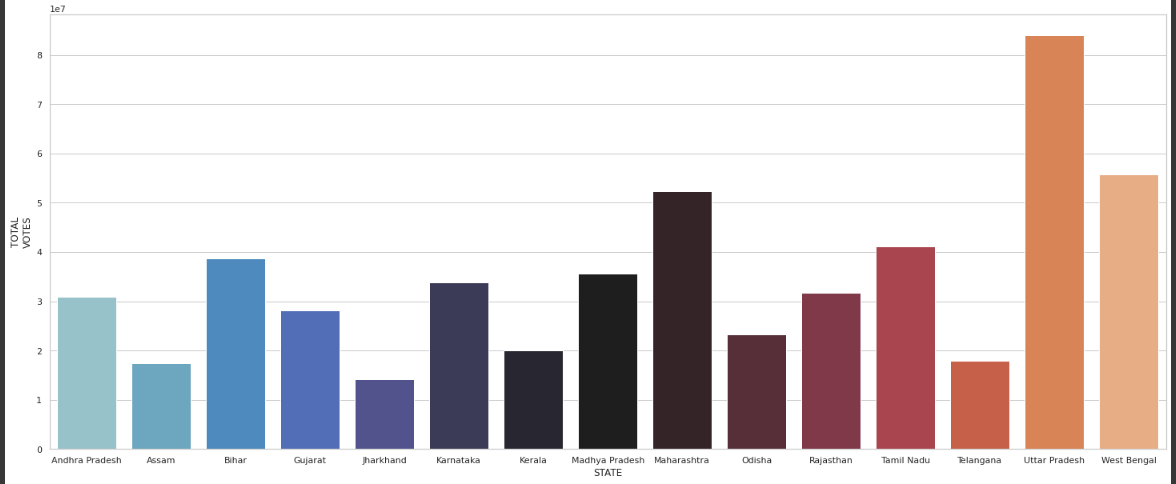
This is the bar plot of gender vs Crime from which we can know that the number of female candidates is maximum or the number of male candidates is maximum in India.

****

**Fig 11:** Bar Graph of Gender vs Crime

# Bar Graph of State vs Total Votes:

From this bar graph, we can get the knowledge about the no of votes in different states.

****

**Fig 12:** Bar Graph of State vs Total Votes

From the bar plot, it is transparent that the total number of votes in Maharashtra, Uttar Pradesh, and West Bengal are very much higher than the remaining states in India and Uttar Pradesh holds first place in the total number of votes.

**7. Result**

**8. Conclusion**

**References**