ANALYSIS OF LOK SHABA ELECTION IN INDIA BASED ON CANDIDATE DETAILS FROM 2004 TO 2019

Yashwanthra Seenivasan
yashwanthra171@gmail.com
Chubb Trainee
Data Analytics Track

ABSTRACT

The power of Democracy is flourished by elections conducted at regular intervals. It is indispensable for a developing country like India to elect a better leader for the better future. The results of the election are often suspected based on the elected candidate's educational qualification, gender, criminal cases and assets possessed by him. So, this analysis primely focuses on the candidate details of Lok Shaba election in India and how it influences the election results. This analysis can also act as an eye-opener for the public to look for the candidate details before electing one.

DATA COLLECTION

This analysis is based on five datasets collected from www.kaggle.com. These datasets provide information about the Lok Shaba election candidates from the year 2004 to 2019. The datasets revealing the results of the Lok Shaba elections in 2014 and 2019 are also collected for the purpose of analyzing how the candidate details influences the results.

APPROACH

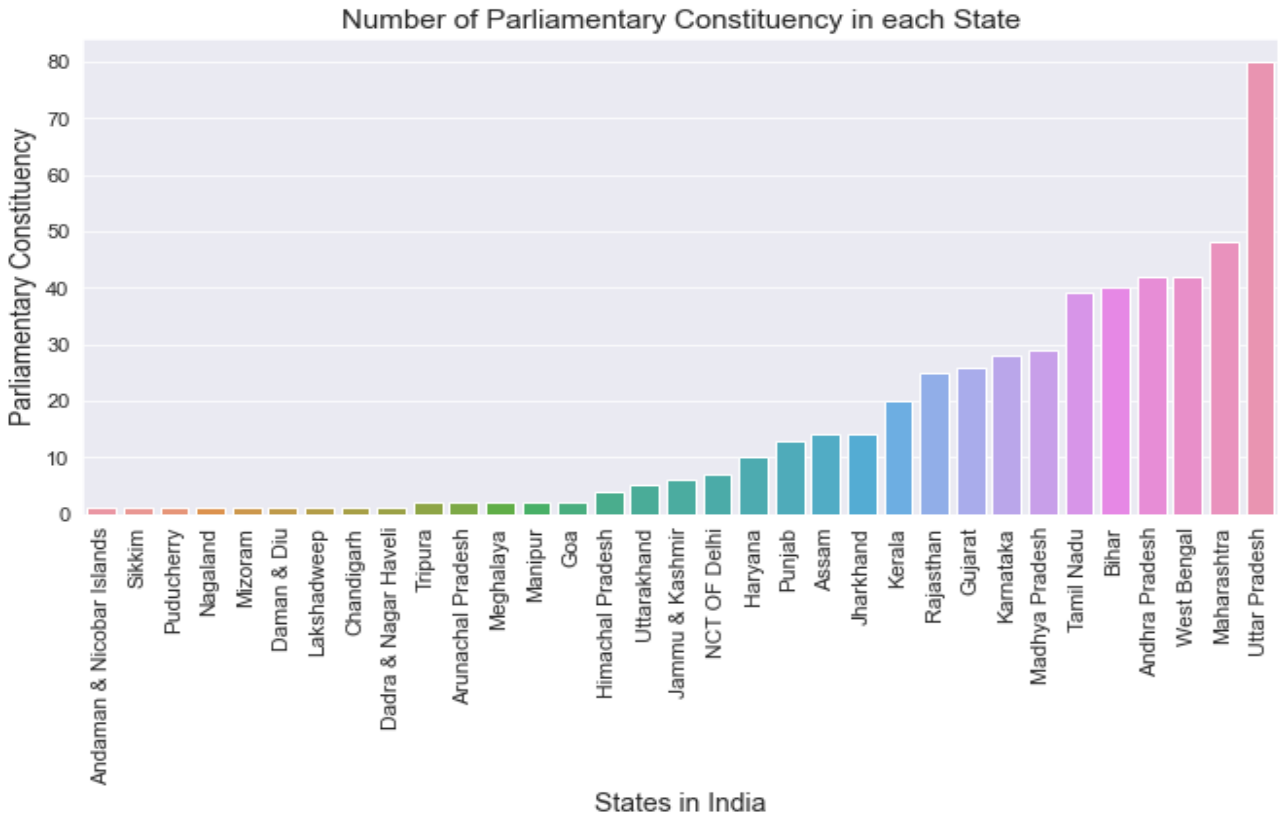
- The various python libraries such as *Numpy*, *pandas*, *Matplotlib* are used for the purpose of mathematical calculations, extraction of data and visualization respectively.
- The extracted dataset was found to have *null values* which are then effectively handled to achieve accurate analysis rather than removing them from the dataset.
- The datasets to be used for analysis was found to be highly skewed. Although *skewness* will not have much effect on descriptive analysis, it is handled for the purpose of predictive analysis to accomplish better decisions from the model. The *log transformation* was used to handle skewness in the datasets.
- Different features measured in different scales will be perplexing while comparison. Hence, it is scale down using *StandardScaler()* function in *sklearn* library, which can convert all the values to a common format and makes analysis easier.

ANALYSIS AND VISUALIZATION

1. Number of Parliamentary Constituency (PC) in each state

PURPOSE

- In Lok Shaba election every candidate is about to compete in one of the PC announced by the election commission. So, it is essential for a government to know about the number of PC in each state and plan accordingly.



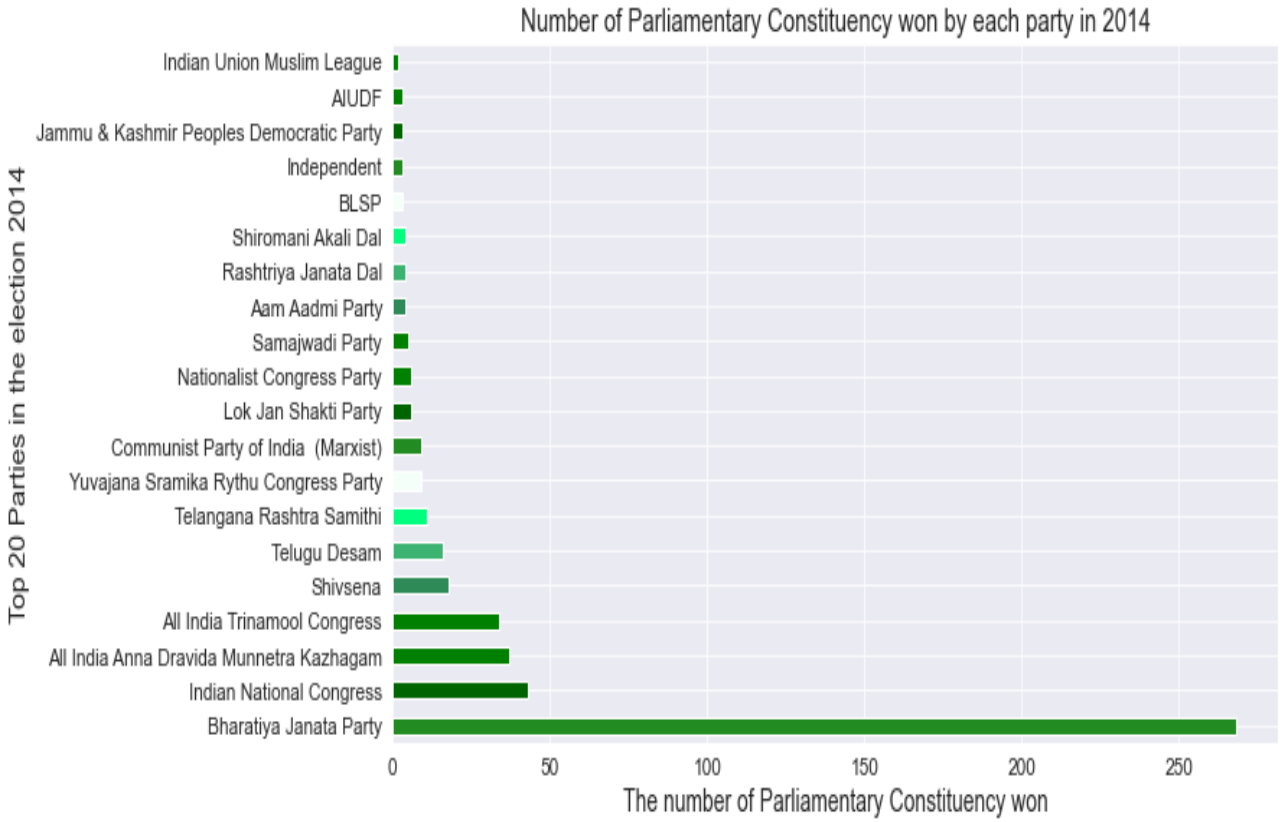
INFERENCE

- From the above analysis it is evident that Uttar Pradesh, Maharashtra and West Bengal has a greater number of PC where the government must ensure the integrity of election with high significance.

2. Total number of Parliamentary Constituency (PC) won by each party in 2014

PURPOSE

- The leader of the Lok Shaba is selected based on the party that have won in at least 272 Parliamentary Constituencies in the Lok Shaba election.
- The party with this majority will select a leader, who then becomes the Leader of the Lok Shaba house.



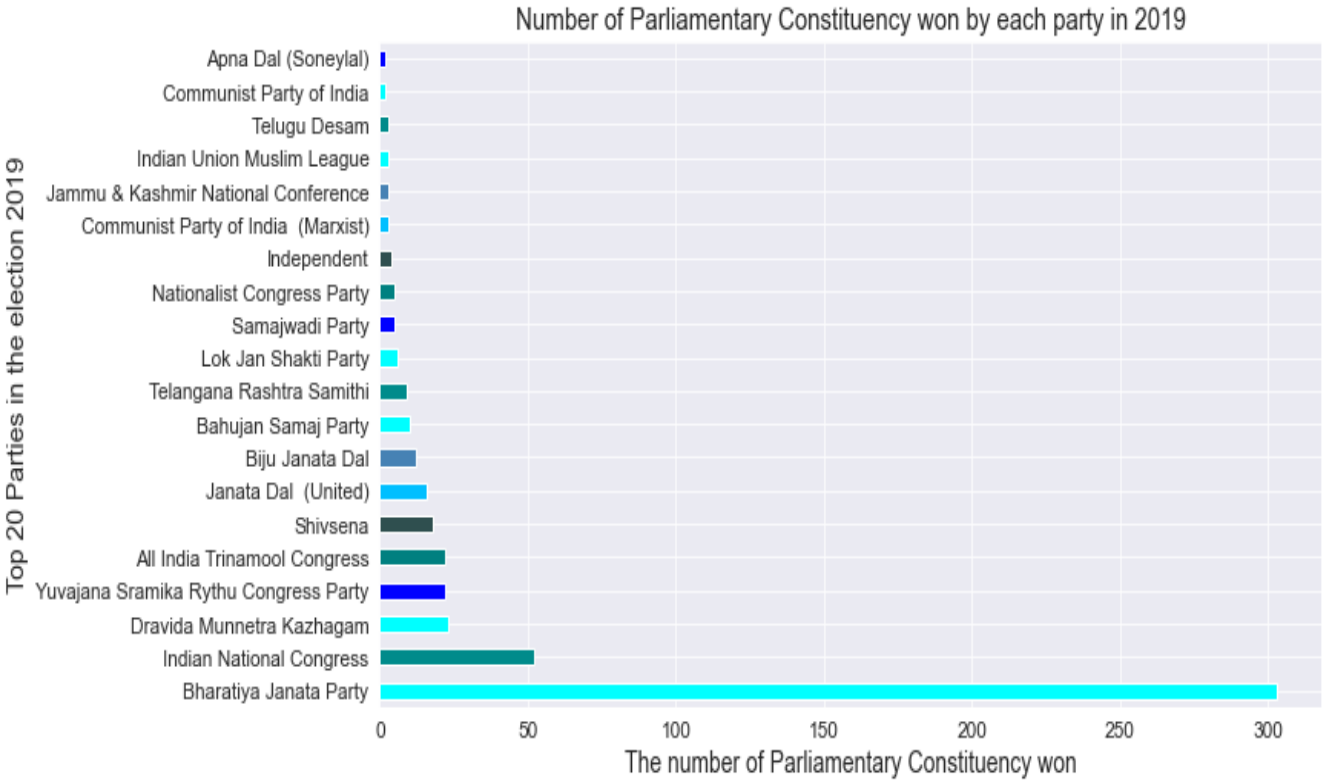
INFERENCE

- From the above plot, it is obvious that Bharatiya Janata Party (BJP) was the ruling part in 2014 with Indian National Congress (INC) being the opposition party.

3. Total number of Parliamentary Constituency (PC) won by each party in 2019

PURPOSE

- To obtain an insight regarding the ruling and opposition party in 2019



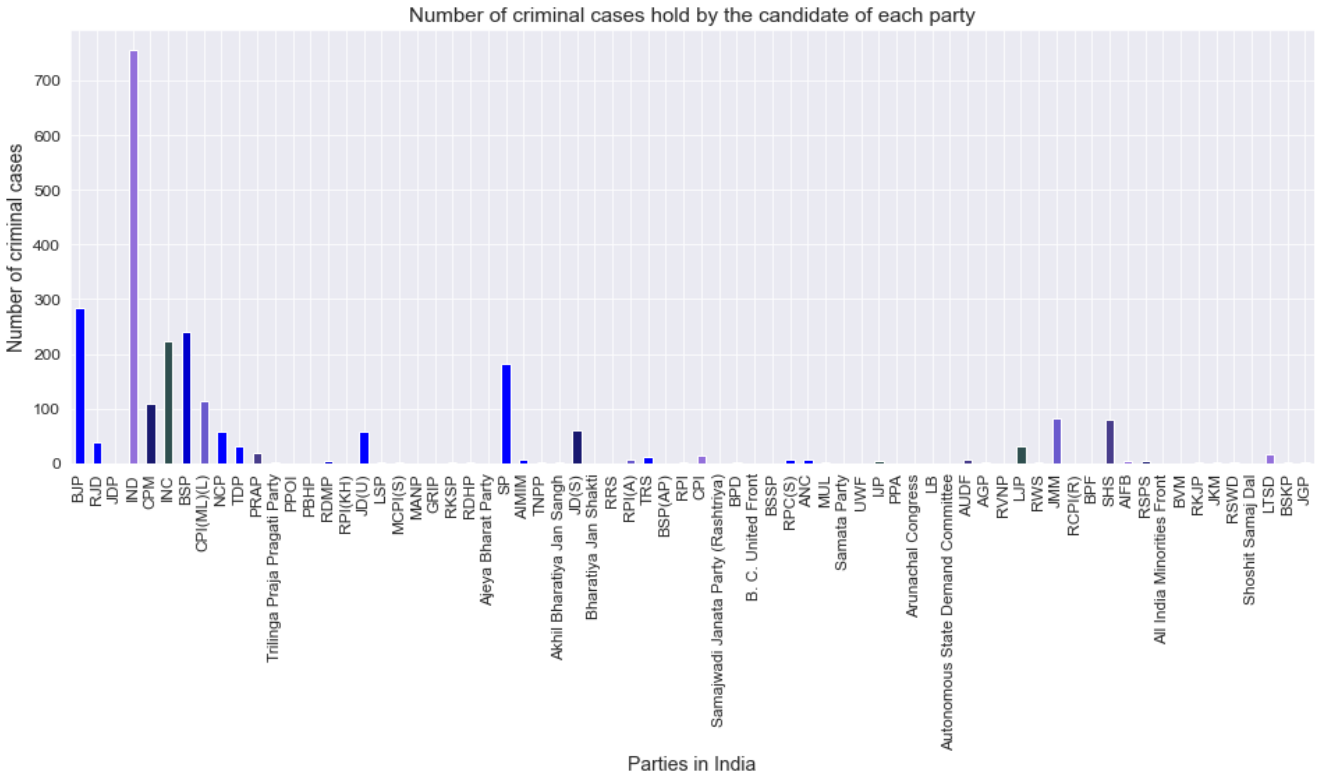
INFERENCE

- It is obvious that Bharatiya Janata Party has become the ruling party consequently, but this time with a greater number of seats.

4. Parties that withhold a greater number of criminal cases

PURPOSE

- From the above election results, BJP has conquered huge number of PC consequently. So, it is essential to analyze whether the winning candidates are criminals (a highly suspected candidate parameter in Indian elections).



INFERENCE

- It is commendable that BJP is not the party with highest criminals, but still it holds 270 and above criminal cases in its candidates.

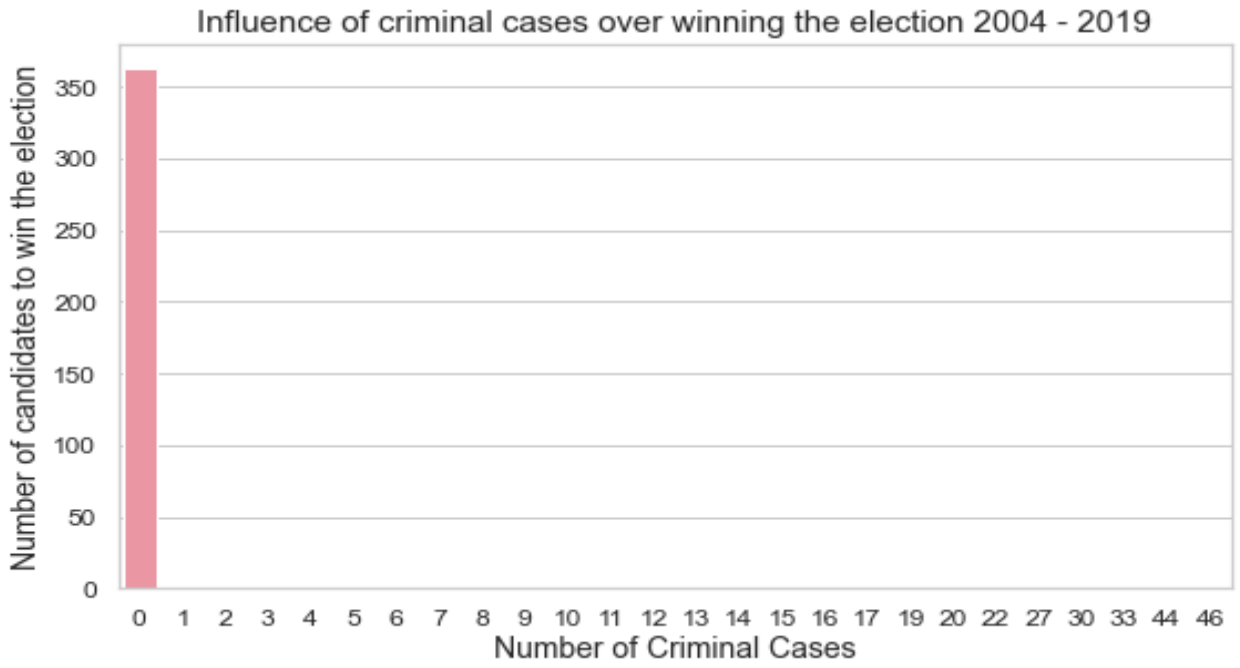
5. Influence of criminals over winning PC's in Lok Shaba elections

ANOVA Test on candidates between the year 2004 – 2019

- Null hypothesis: The criminal cases of candidates has no significance in winning the election
- Level of significance: 0.05

	df	sum_sq	mean_sq	F	PR(>F)
Winner	1.0	52.878503	52.878503	19.441402	0.000011
Residual	7918.0	21536.100159	2.719891	NaN	NaN

- Since the P value is less than Level of significance (0.000011 < 0.05), we reject the null hypothesis. Thus, criminal cases of candidates have significance over winning the election.



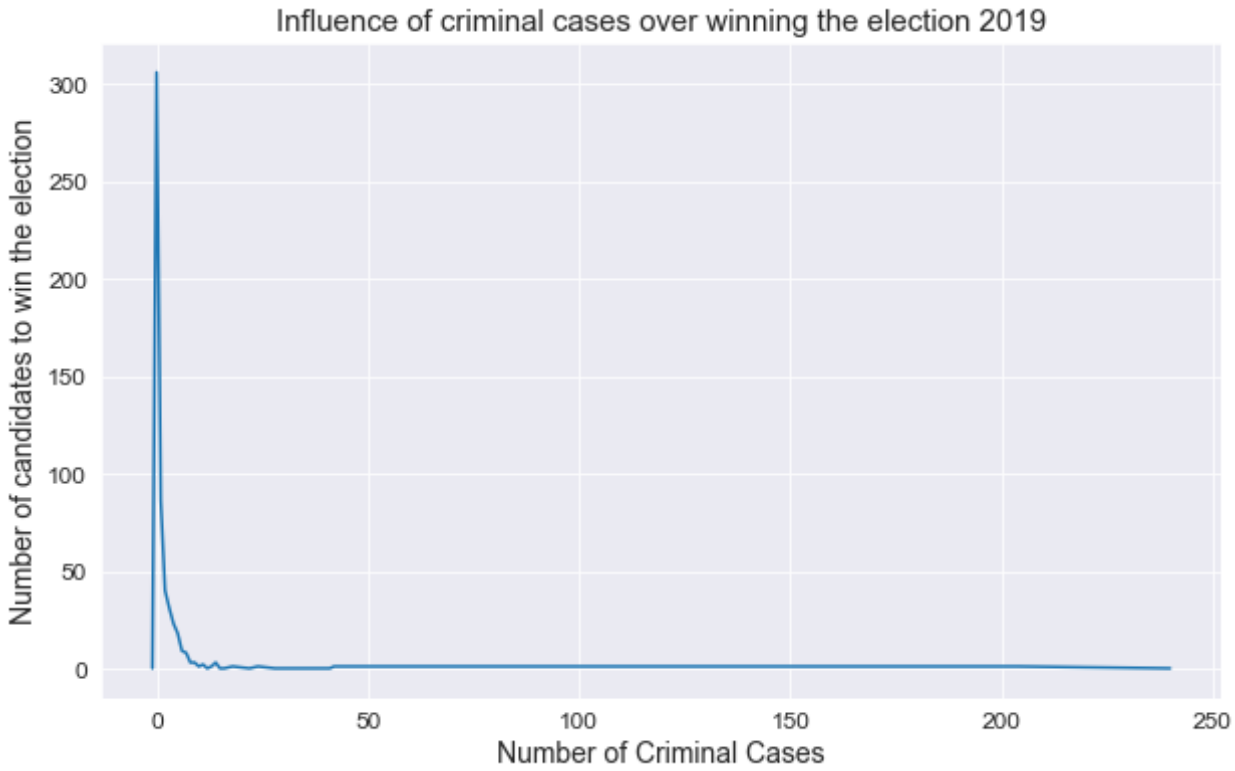
- It is commendable that, during the period from 2004 to 2019, the candidates with zero criminal cases have won the election.
- This proves the dependence of winning the election over zero criminal cases.

ANOVA Test on candidates in the year 2019

- Null hypothesis: The criminal cases of candidates has no significance in winning the election in 2019
- Level of significance: 0.05

	df	sum_sq	mean_sq	F	PR(>F)
WINNER	1.0	273.993643	273.993643	5.254232	0.021985
Residual	2261.0	117904.881301	52.147227	NaN	NaN

- Since the P value is less than Level of significance (0.021985 < 0.05), we reject the null hypothesis. Thus, criminal cases of candidates have significance over winning the election in 2019.



- From the above plot, it is evident that the winning candidates in 2019 do possess criminal cases

INFERENCE

- From the above tests and plots we can infer that the number of criminals winning the election is increasing from the year 2019 and it is in the hands of public to reduce this rate.
- So, a candidate to win an election most probably should have no criminal cases

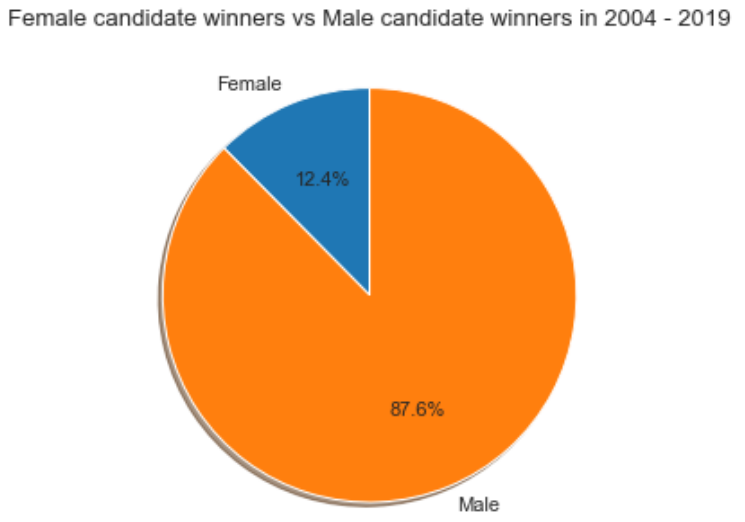
6. Influence of Female candidates in the outcome of the election result

ANOVA Test on Gender and winner of the election between 2004 – 2019

- Null hypothesis: The gender of candidates has no significance in winning the election
- Level of significance: 0.05

	df	sum_sq	mean_sq	F	PR(>F)
Gender	1.0	1.398206	1.398206	32.177911	1.456348e-08
Residual	7918.0	344.055834	0.043452	NaN	NaN

- Since the P value is less than Level of significance (1.43e-08 < 0.05), we will reject the null hypothesis. Thus, gender of candidates has significance over winning the election during 2004 - 2019.



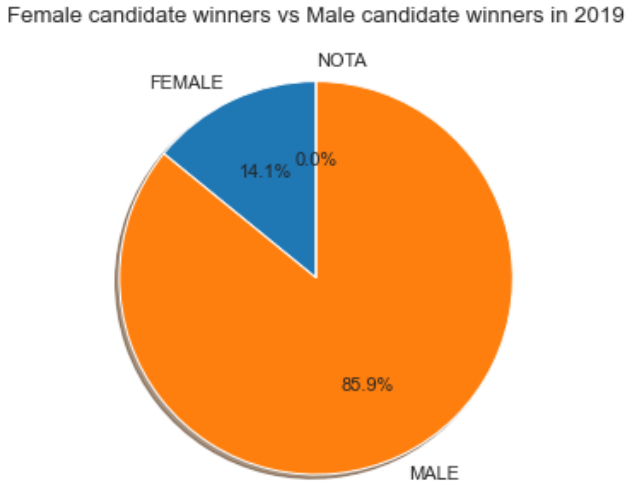
- It is evident that 45 out of 451 Female candidates have won in the election during 2004 – 2019

ANOVA Test on Gender and winner of the election in 2019

- Null hypothesis: The gender of candidates has no significance in winning the election
- Level of significance: 0.05

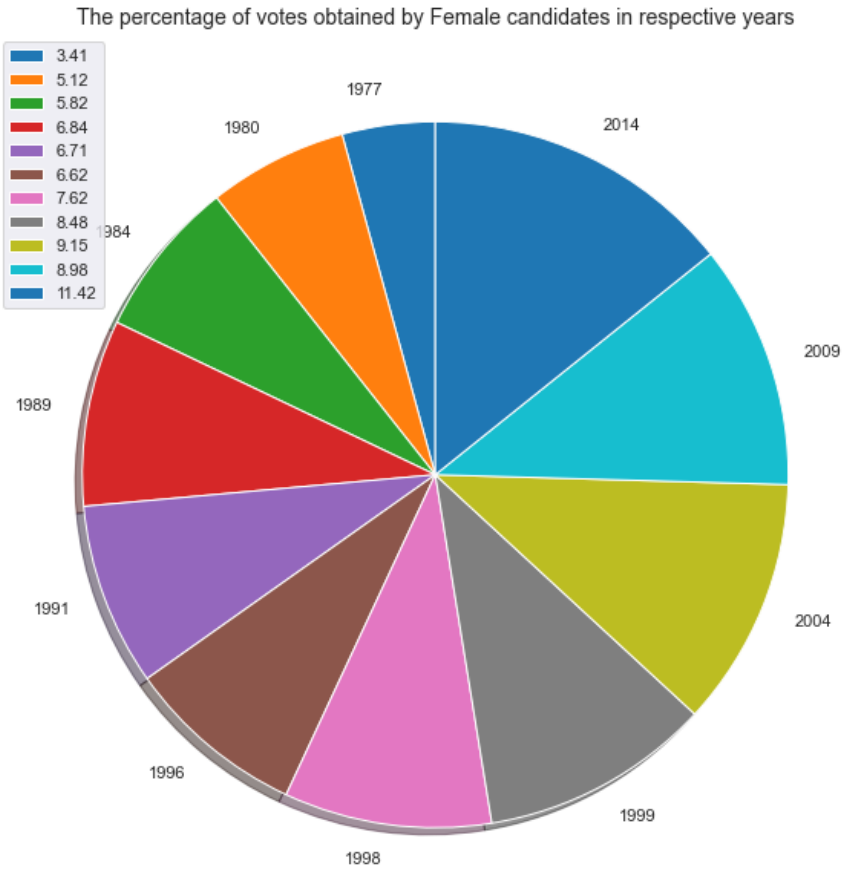
	df	sum_sq	mean_sq	F	PR(>F)
GENDER	2.0	15.809464	7.904732	45.24863	5.395350e-20
Residual	2260.0	394.811835	0.174696	NaN	NaN

- Since the P value is less than Level of significance (5.39e-20< 0.05), we reject the null hypothesis. Thus, gender of candidates has significance over winning the election in 2019.



- In 2019 election, 76 out of 258 Female candidates have won the election

The percent of votes obtained in election by the female candidate from 1977 – 2015



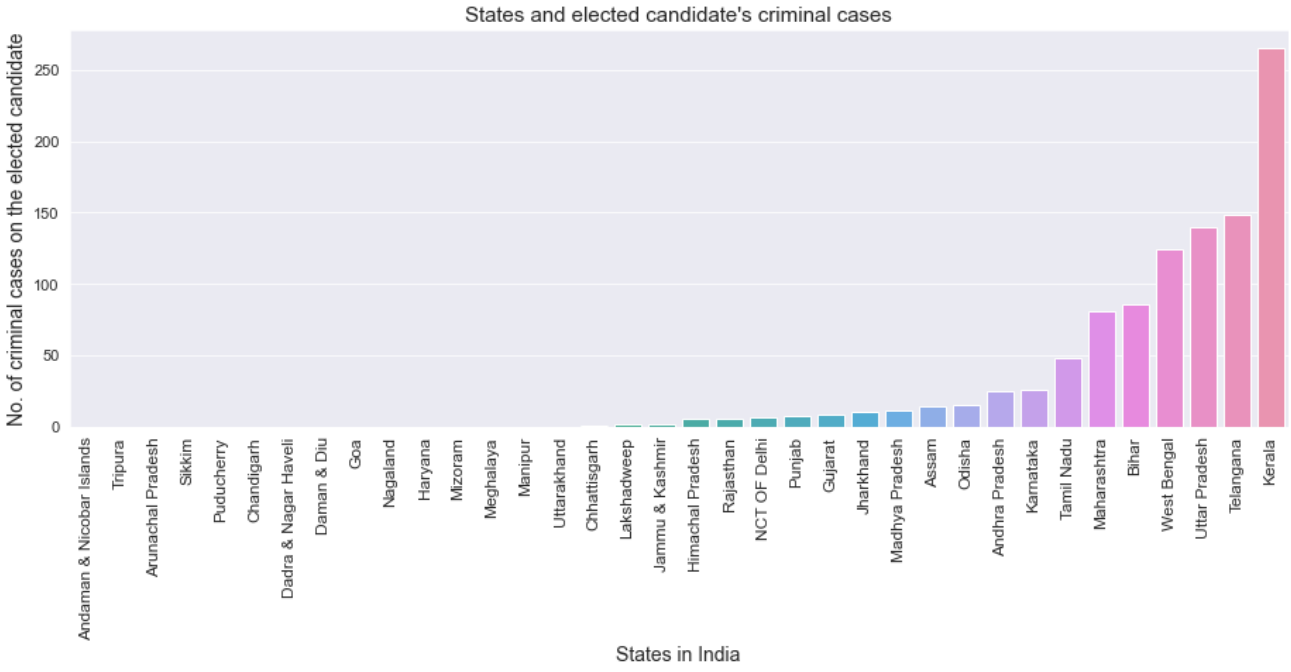
INFERENCE

- From the above tests and plots, we can infer that the number of female candidates participated in the election is comparatively lesser than that of male candidates.
- Although the female candidates are less in number, the rate of winning for a female candidate is higher than that of men.
- It is also obvious that the percentage of votes being secured by a female candidate is gradually increasing over the years from 1977 – 2014.
- Thus, the chances of winning an election by a female candidate is more than that of a male candidate.

7. The state with highest number of criminals elected

PURPOSE

- The state with highest number of criminals elected is at high risk in both development and growth.



INFERENCE

- From the above analysis, it is crystal clear that Kerala has highest number of criminals elected.
- This should be considered by the public and it can also pave a path for new candidates in the election with no criminal cases.

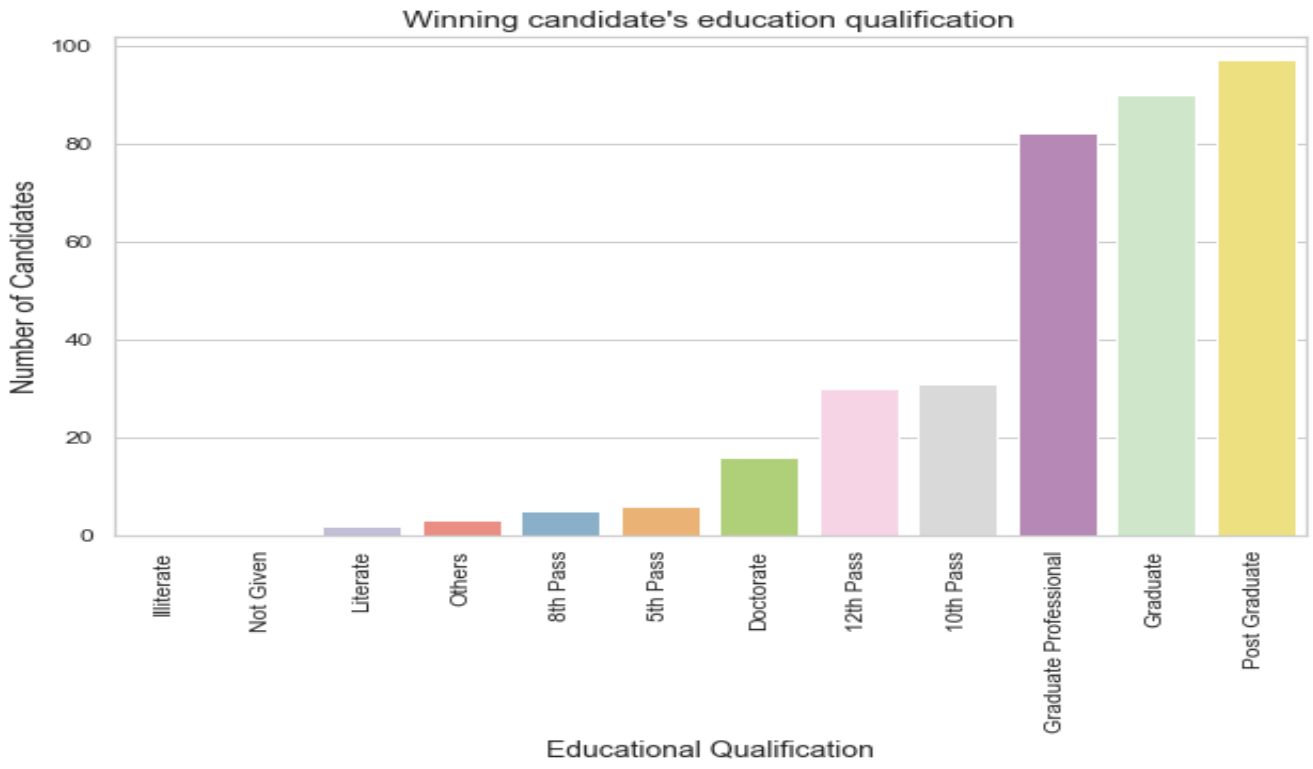
8. Influence of educational qualification over winning the election

ANOVA TEST on winner and educational qualification of a candidate between 2004-2019

- Null hypothesis: The educational qualification of candidates has no significance in winning the election
- Level of significance: 0.05

	df	sum_sq	mean_sq	F	PR(>F)
Education	11.0	11.047878	1.004353	23.750819	8.741425e-49
Residual	7908.0	334.406162	0.042287	NaN	NaN

- Since the P value is less than Level of significance (8.74e-49 < 0.05), we will reject the null hypothesis. Thus, educational qualification of candidates has significance over winning the election during 2004 - 2019.



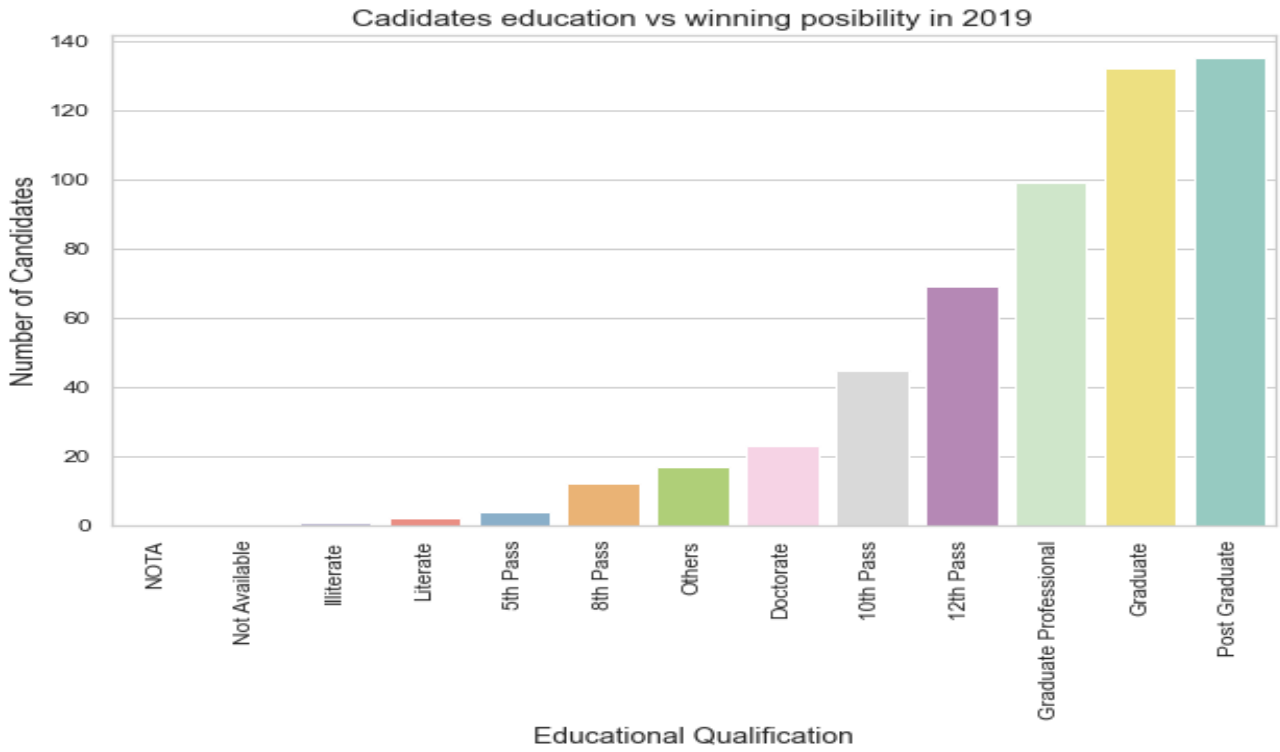
- From the above bar plot, we can deduce that illiterate doesn't win any election, higher the educational qualification higher is the chances of winning an election.

ANOVA TEST on winner and educational qualification of a candidate in 2019

- Null hypothesis: The educational qualification of candidates has no significance in winning the election in 2019
- Level of significance: 0.05

	df	sum_sq	mean_sq	F	PR(>F)
EDUCATION	12.0	21.240530	1.770044	10.228033	6.868338e-20
Residual	2250.0	389.380769	0.173058	NaN	NaN

- Since the P value is less than Level of significance (6.86e-20 < 0.05), we will reject the null hypothesis. Thus, educational qualification of candidates has significance over winning the election in 2019



INFERENCE

- From the plots it is obvious that, none of the illiterate candidates have won the election
- The probability of a post graduate candidate winning the election is higher than that of Doctorate candidate.
- 357 out of 362 MP's are educated during the period 2004 -2019
- 521 out of 539 MP's are educated in the year 2019

CONCLUSION

- From the election analysis over the years between 2004 to 2019, we can arrive at the following insights
- The probability of a candidate with criminal cases to win an election is too low
- The number of female candidates is low but the rate of winning of a female candidate is higher than that of men
- Kerala is the state with highest number of criminals elected
- The probability of an educated candidate (most probably post graduate) to win an election is high.

REFERENCE

- Dataset for Election Results of 2014 and 2019:

  https://www.kaggle.com/ankit2106/indian-general-election-2019-candidate-wise-data

- Dataset for candidate details during 2004 – 2019:

  https://www.kaggle.com/themlphdstudent/lok-sabha-election-candidate-list-2004-to-2019

- Dataset for overall election results from 1977 – 2015:

  https://www.kaggle.com/awadhi123/indian-election-dataset?select=indian-national-level-election.csv

- For code and plots

  https://github.com/Yashwanthra/election_analysis/blob/main/Election_data_analysis.ipynb