

3/5/2021

2019 Lok SABHA ELECTION ANALYSIS

Submitted to:

Suman Gangopadhyay

(linuxgurusuman@gmail.com)

Vayuj Kandoi

(vayuj.kandoi@chubb.com)



ACKNOWLEDGEMENT

I would like to thank Mr. Suman Gangopadhyay for his immense support and guidance. This report would not have been possible without the dataset provided by him.

ABSTRACT

The 2019 Indian general election was the 17th Lok Sabha election. Members of Lok Sabha are elected by the citizens of India who are 18 or above years of age, from the candidates participating in election who stand in their respective constituencies. In this report we will be seeing an analysis of 2019 GE. The analysis is done on parameters like parties win-loss count, candidates age distribution, candidates' assets and liabilities, candidates' gender, criminal cases count against candidates, reservation wise, candidates' education profile, etc. We will also perform ANOVA test to determine whether few of the above parameters play an important role in winning.

DATASET DESCRIPTION

The dataset I have considered in this report is lok-sabha-candidate-details-2019.csv.

The data contains the following information:

- CONSTITUENCY: Name of the constituency.
- NAME: Candidate name.
- WINNER: 0-1 value representing loss or win.
- PARTY: Candidate party name.
- SYMBOL: Party symbol.
- GENDER: Candidate gender.
- CRIMINALCASES: Criminal cases count against the respective candidate.
- AGE: Age of the candidate.

- CATEGORY: Reservation category of the candidate.
- EDUCATION: Education qualification of the candidate.
- ASSETS: Assets worth of the candidate.
- LIABILITIES: Total liability of candidate.
- GENERALVOTES: Vote count captured in EVM for a candidate.
- POSTALVOTES: Vote count captured on ballot paper for a candidate.
- TOTALVOTES: (GENERALVOTES + POSTALVOTES) Total count of vote a candidate achieved.
- OVERTOTALELECTORSINCONSTITUENCY: Percentage of people who voted for the candidate in respective constituency.
- OVERTOTALVOTESPOLLEDINCONSTITUENCY: Percentage of votes in favour of the candidate.
- TOTAL ELECTORS: Number of people eligible to vote in respective constituency.

DATASET CLEANING

Before proceeding with analysis, non-processable values like 'na', 'nan', 'empty blanks' or 'Not Available' values were replaced with suitable values. Assets and Liabilities data were converted to integer figures.

ANOVA

An ANOVA test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis

The null hypothesis is a characteristic arithmetic theory suggesting that no statistical relationship and significance exists in a set of given, single, observed variables between two sets of observed data and measured phenomena.

Null hypothesis->The parameters are not related

If p value > level of significance -> accept the null hypothesis

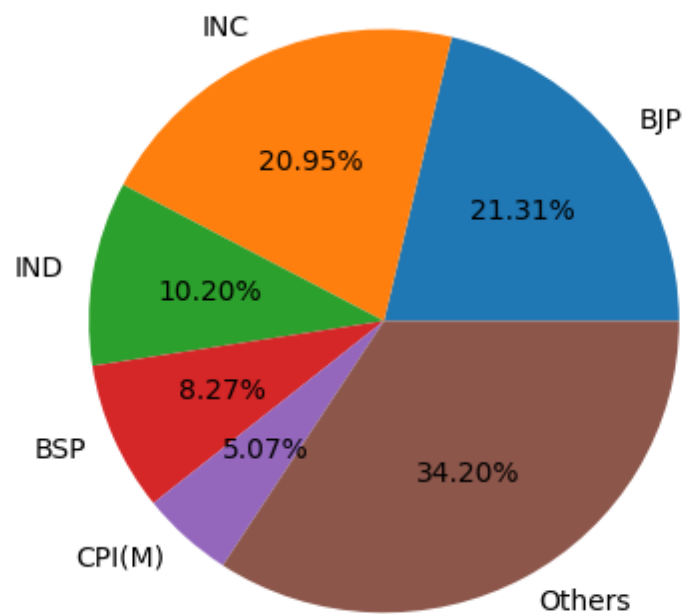
if p value < level of significance -> reject the null hypothesis

level of significance -> 0.05

ANALYSIS

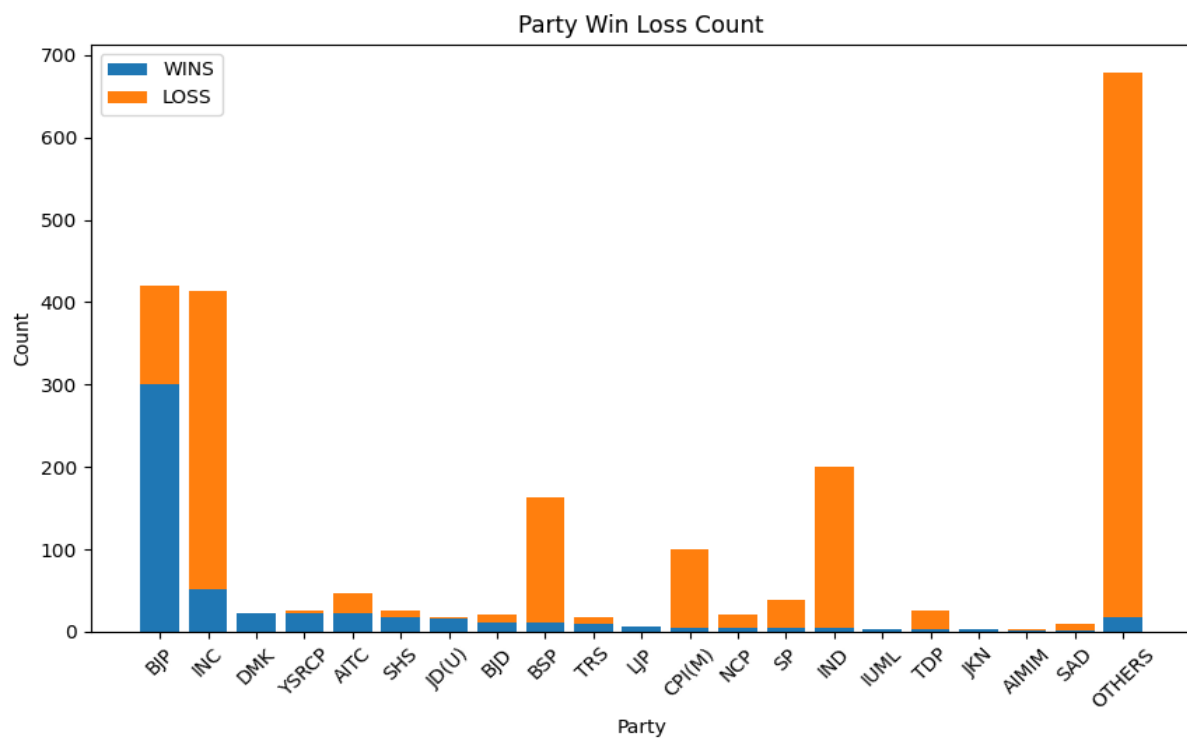
- Based on participation percentage Party wise

Election 2019 participation proportion party wise



BJP was the biggest party in the 2019 GE, followed by INC. Other large parties were IND, BSP and CPI.

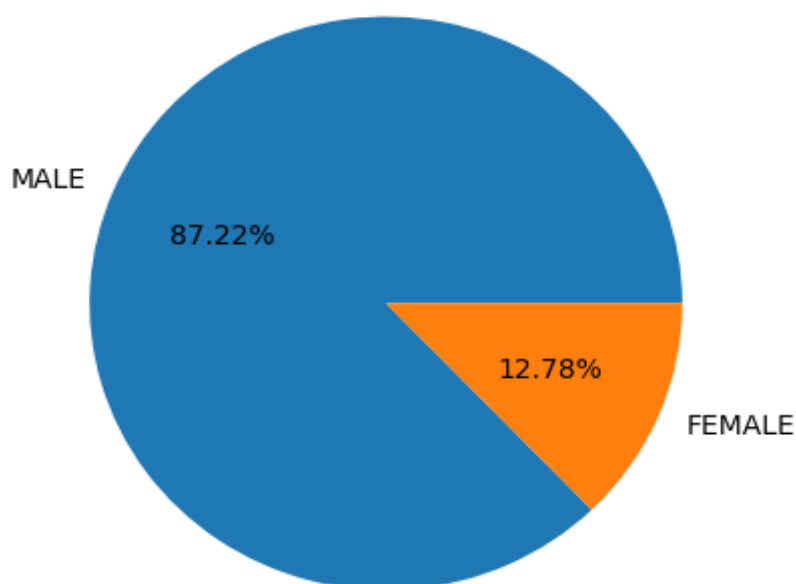
- **Party Win-Loss count**



Winning rate of parties like BJP, DMK, YSRCP are very high, unlike INC, BSP, IND who have lost more constituencies than they have won.

- **Based on Gender**

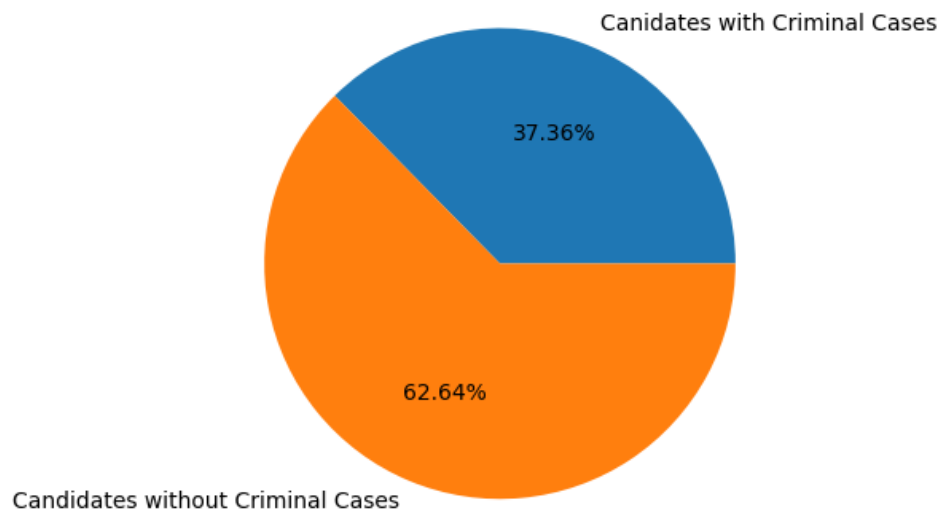
Gender proportion in Election 2019



The percentages of male candidates are way high (approx. 6.8 times) than female candidates.

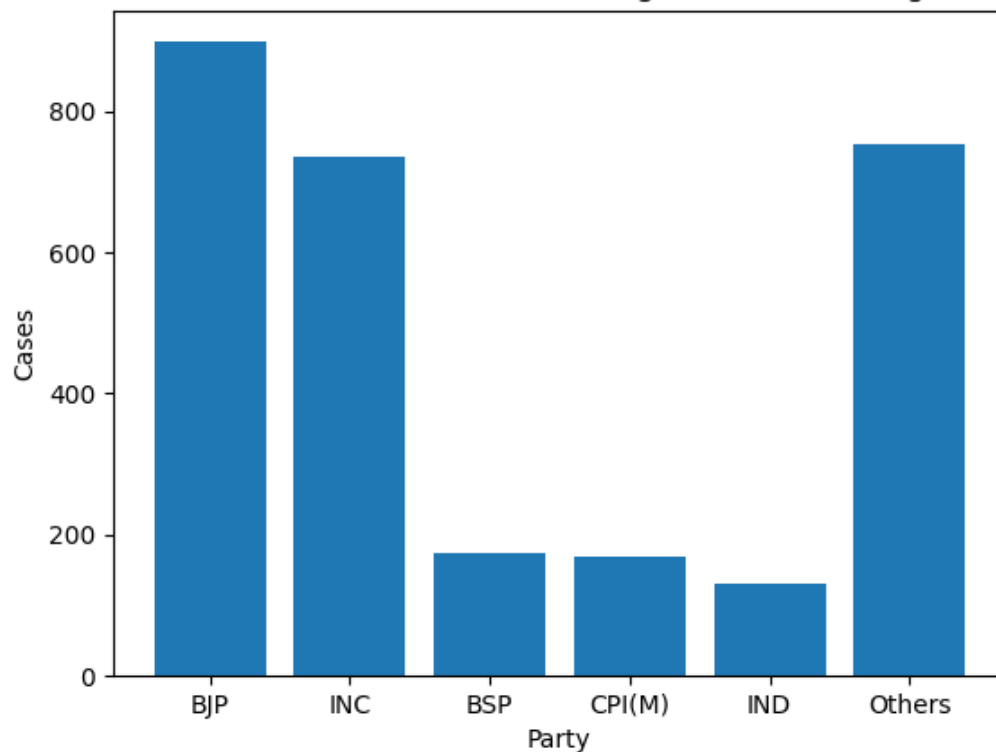
- Based on Criminal Cases

Proportion of Candidates having criminal charges against them



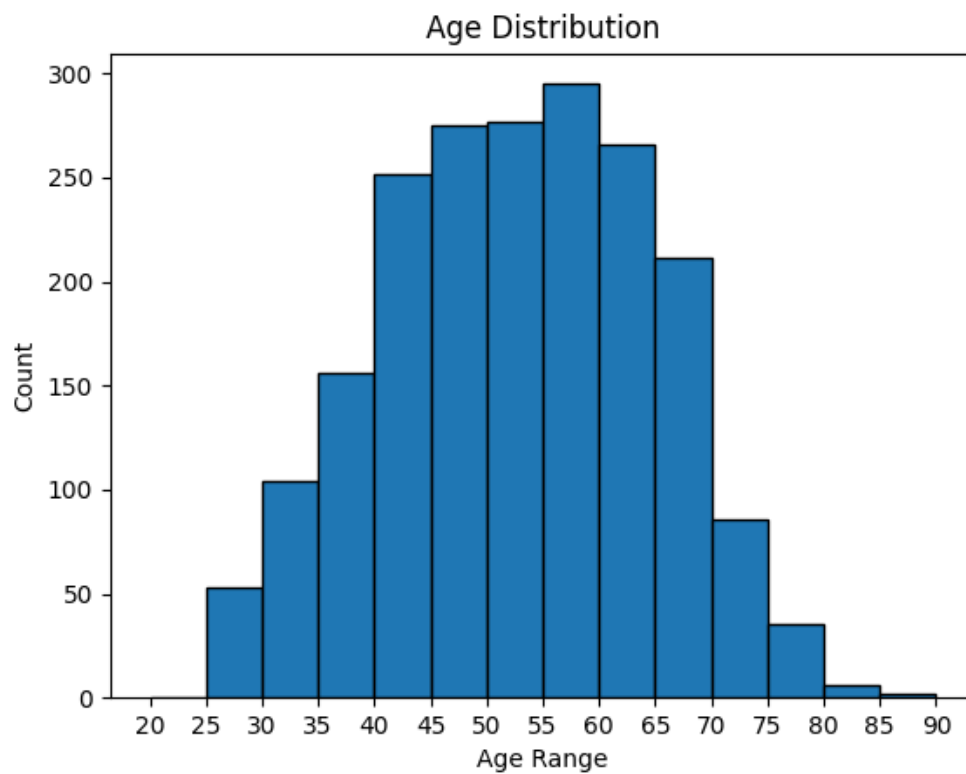
More than one-third of the candidates have pending criminal cases against them. This implies that parties do not do a thorough background check of candidates before issuing election ticket.

Parties with number of candidates having criminal cases against them



BJP has the highest number of candidates with pending criminal charges against them. INC is not too far behind.

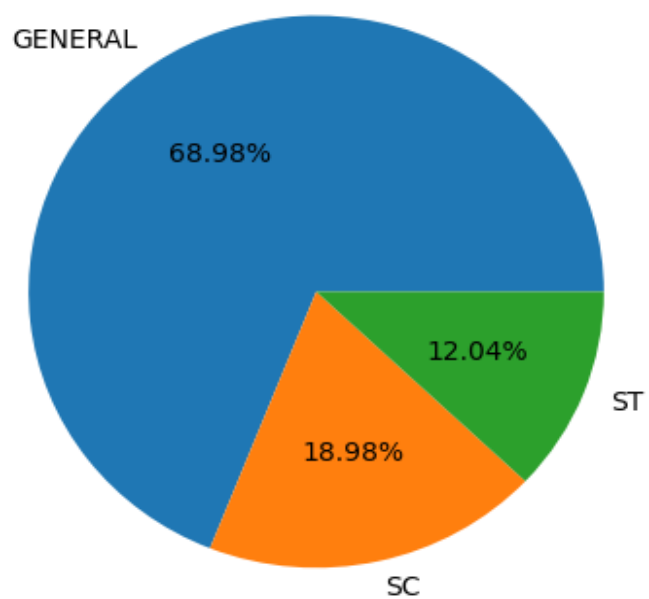
- **Candidates Age distribution**



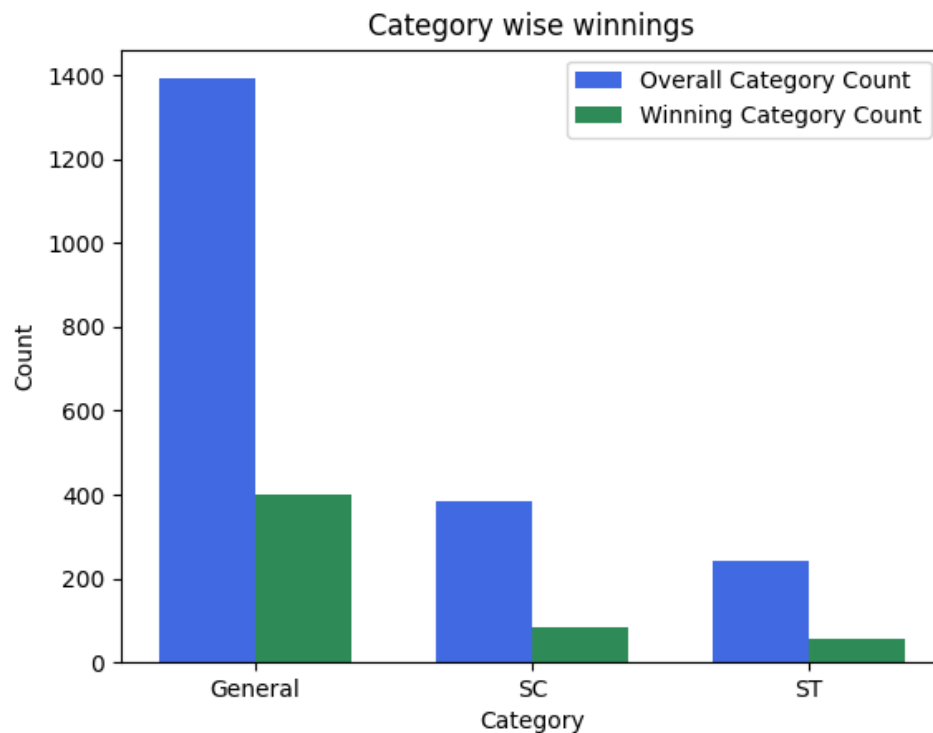
Maximum candidates are in the age group of 55-60. Youth age group (generally considered to be below 45 years in politics) has a good proportion too.

- **Category Wise**

Percentages of candidates category wise

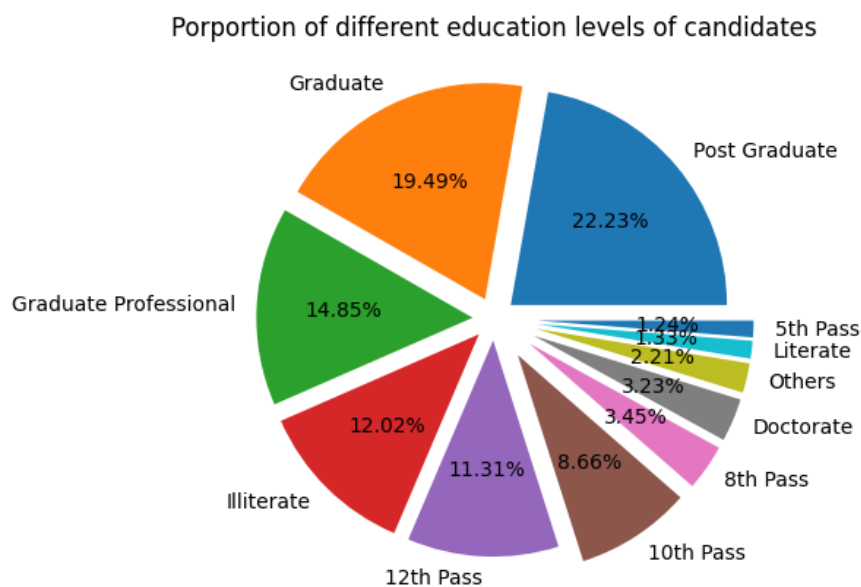


Almost 69% of the candidates belonged to General category whereas almost 19% belonged to SC category and 12% to ST category.



The winner's percentage in General category is around 28.7% whereas in SC and ST category it is around 22.2% and 22.6% respectively.

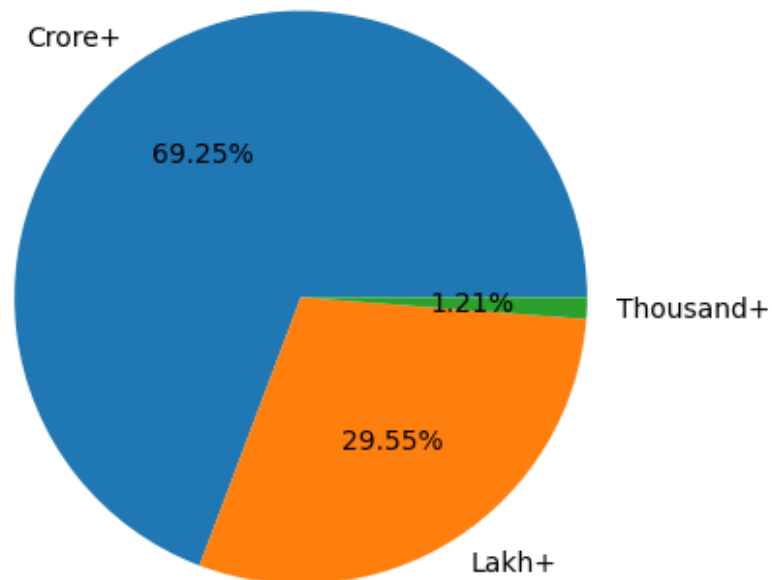
- **Education qualifications of the candidate**



Around 22% of the candidates have Post Graduate degree, 14.85% have a Graduate degree, almost 15% are Graduate professionals where 12% of the candidates are Illiterate.

- **Assets Wise**

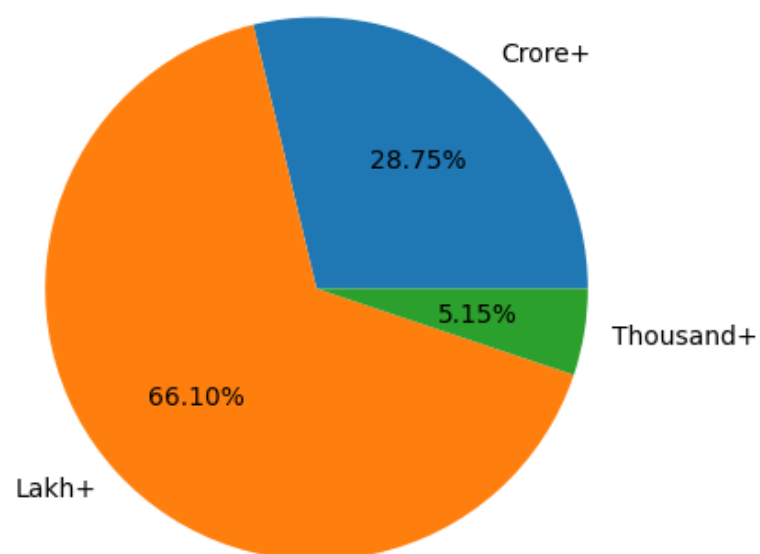
Candidates percentage Assets wise



About 69% of the candidates have assets worth crore plus. 29.6% have assets worth in lakhs and 1.21 % of candidate have worth thousands.

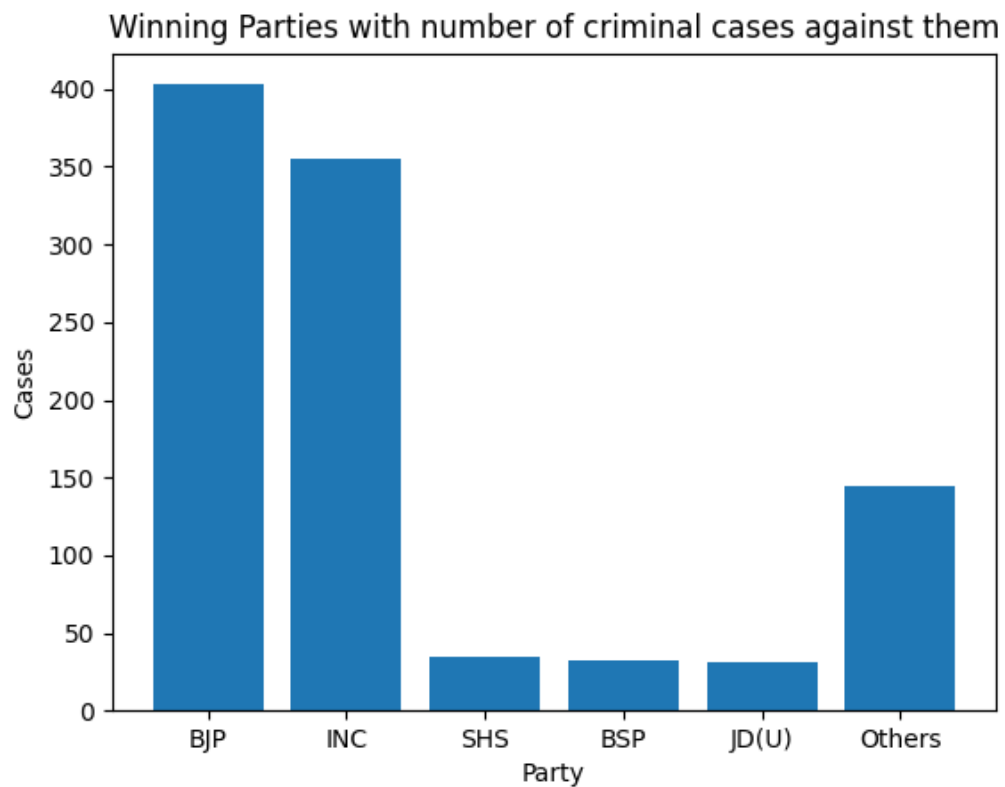
- **Liabilities Wise**

Candidates percentage Liabilites wise

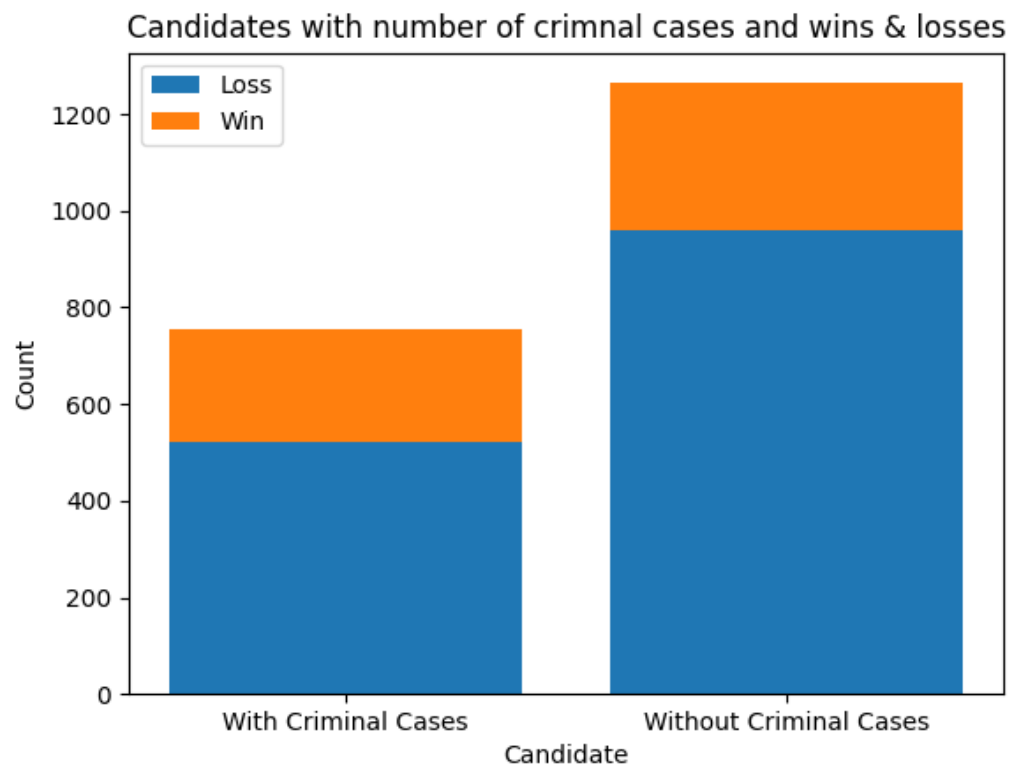


Almost 29% of candidates had liabilities worth crore plus, 66% had liabilities in lakhs and 5.15% had in thousands.

- Does the criminal cases parameter play a significant role in winning an election?



BJP has the highest number of winner candidates who had pending criminal charges against them. INC is not too far behind.



Percentage of winners in no criminal case category: almost 24%

Percentages of winners in criminal case category: almost 31%

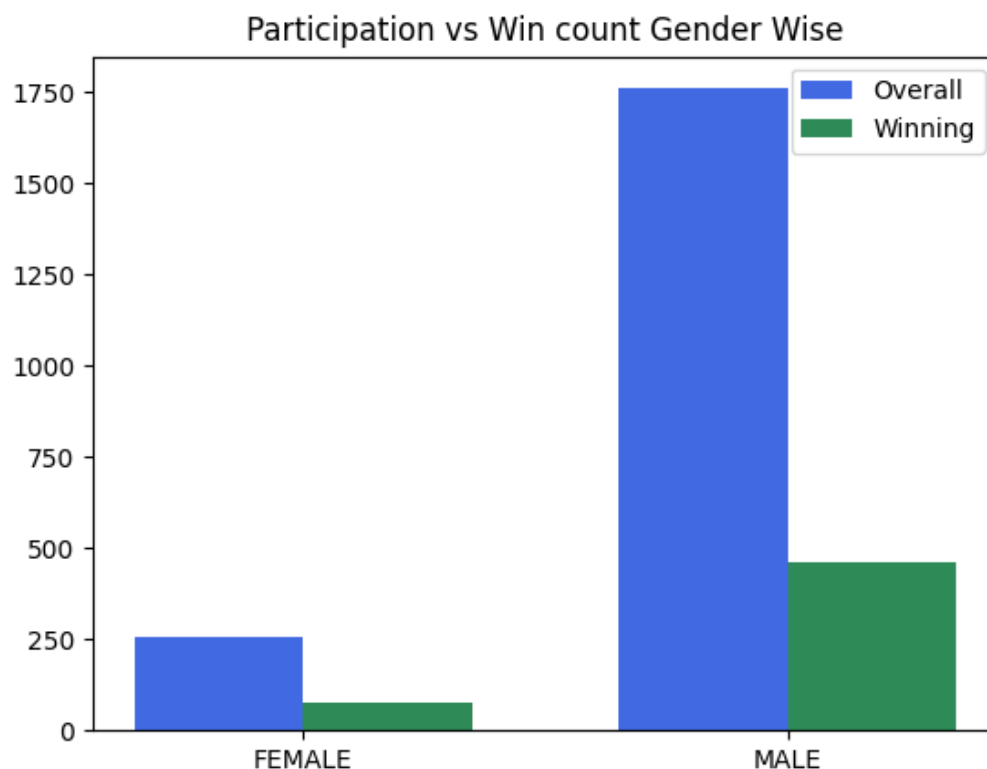
So, it can be inferred that having a criminal case pending does give an upper hand in election. This can be verified through ANOVA test.

- ➔ null hypothesis -> Criminals does not play an important role in election
- ➔ alternate hypothesis -> Criminals does play an important role in election

	df	sum_sq	mean_sq	F	PR(>F)
CRIMINALCASES	1.0	2.115624	2.115624	10.854886	0.001002
Residual	2016.0	392.919560	0.194901	NaN	NaN

Since the P value < level of significance, we reject the null hypothesis.

- Does the presence of Female candidates irrespective of party somehow influences the outcome of the result?



Percentage of winners in the female category: almost 29.5%

Percentage of winners in male category: almost 26.3%

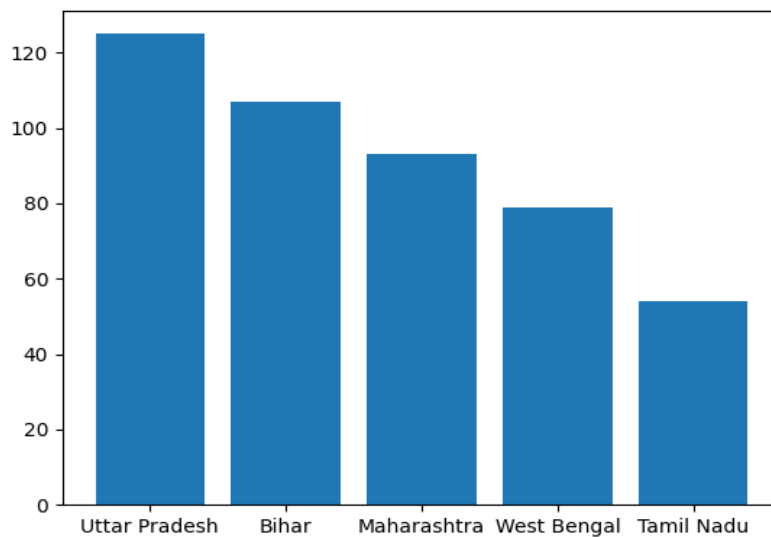
So, it can be inferred that being a female candidate does not give an upper hand. This can be verified through ANOVA test.

- ➔ Null hypothesis: Gender is not related to Winnings
- ➔ Alternate hypothesis: Gender is related to Winnings

	df	sum_sq	mean_sq	F	PR(>F)
GENDER	1.0	0.223348	0.223348	1.140468	0.285681
Residual	2016.0	394.811835	0.195839	NaN	NaN

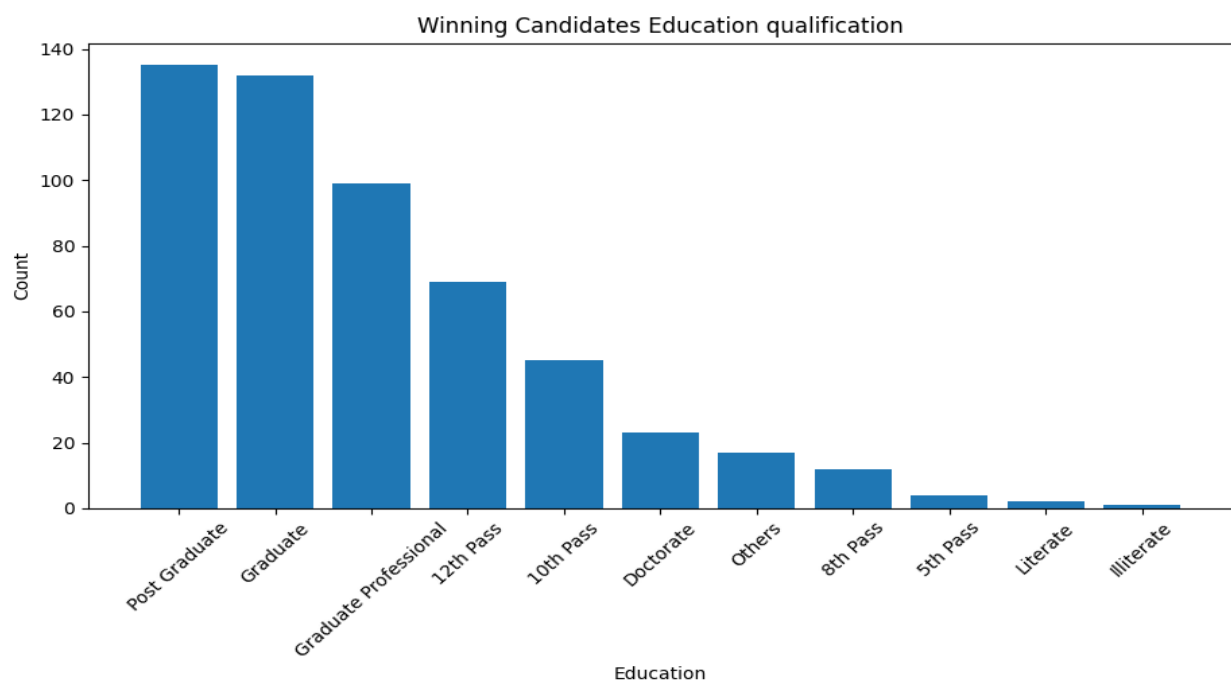
As the p value > level of significance, we accept the null hypothesis.

- **State with highest numbers of criminal Elected**



The above figures show top 5 states where number of criminal elected are highest. Uttar Pradesh leads the chart with Bihar following up.

- **Elected members Education qualifications and the parameter's importance**



Most of the members elected are educated which is a good sign. Here educated refers to a candidate having a Post Graduate degree or Graduate or Graduate Professional or 12th Pass or a Doctorate.

Number of candidates considered as educated: 1609

It can be inferred that education does play an important role in winning elections. This can be verified through ANOVA test.

- ➔ Null hypothesis: Education does not play an important role in winning elections
- ➔ Alternate hypothesis: Education plays an important role in winning elections.

	df	sum_sq	mean_sq	F	PR(>F)
EDUCATION	1.0	2.445922	2.445922	12.560144	0.000403
Residual	2016.0	392.589262	0.194737	NaN	NaN

As the P value < level of significance, we reject the null hypothesis.

CONCLUSION

We have done a general analysis of 2019 India Lok Sabha Election on parameters like party, age, gender, criminal cases, category, education qualification, assets and liabilities and tested the significance of few parameters in winning election. We have also performed the ANOVA test to verify the same.