Data Analytics

Project Report

Forecasting the 2019 Lok Sabha Election

Teammates :

Adarsh Subhas Nayak
PES1UG20CS620

Harsh Verma
PES1UG20EC078

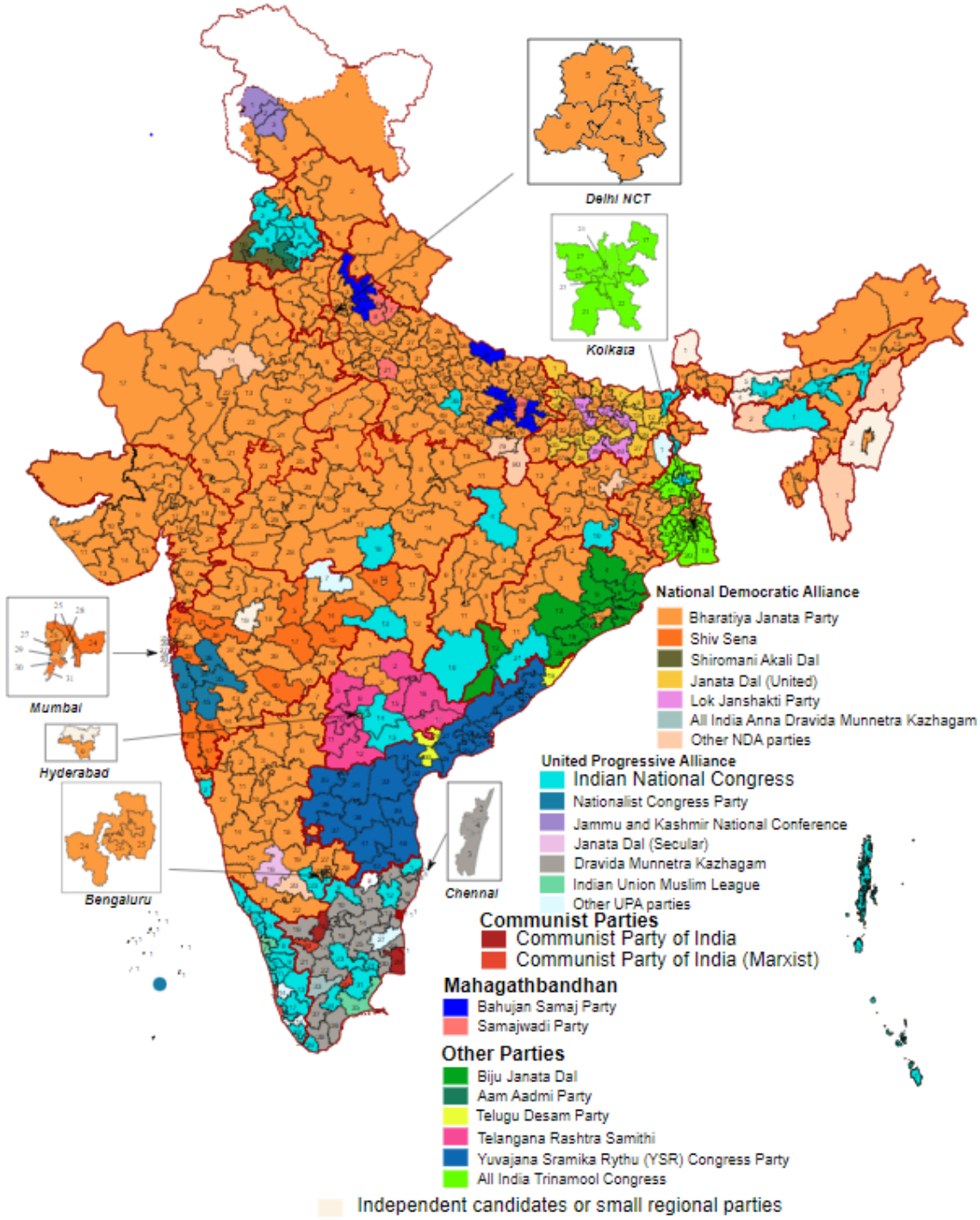
How do Indians Vote?

Introduction to the analysis

India is a country, divided into states and union territories, with a parliamentary system governed under the Constitution of India, which defines the power distribution among the federal government and the states. Election Commission is a federal body, enacted under the provisions of the Constitution, responsible for monitoring and administering all the electoral processes of India. This body is responsible for ensuring elections are free and fair, without any bias

The Analysis here is based on the election to Lok Sabha (The General Elections) for the year 2019 Members of Lok Sabha (House of the People) or the lower house of India's Parliament are elected by being voted upon by all adult citizens of India, from a set of candidates who stand in their respective constituencies. Every adult citizen of India can vote only in their constituency. Candidates who win the Lok Sabha elections are called 'Member of Parliament' and hold their seats for five years or until the body is dissolved by the President on the advice of the council of ministers. The house meets in the Lok Sabha Chambers of the Sansad Bhavan in New Delhi, on matters relating to creation of new laws, removing or improving the existing laws that affect all citizens of India.

Presenting an view before going into the Analysis



About the data

1. The data is very good- cleaned and well presented as we go on analyzing the data.
2. Some data types and columns needed basic cleaning.
3. Overall it gives us a good sense about the Indian political scenario and what factors we should consider while we choose our representative. It's our responsibility to choose the right candidates, such that they can make the difference!

Link to dataset : click here to view the dataset

Importing the dataset

In [1]:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
import os, sys
from collections import defaultdict
from urllib.request import urlopen
import json
import plotly.graph_objects as go
from plotly.subplots import make_subplots
from ipywidgets import widgets
import geopandas as gpd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import random
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
from plotly.colors import n_colors
from plotly.subplots import make_subplots
init_notebook_mode(connected=True)
import cufflinks as cf
cf.go_offline()
from wordcloud import WordCloud , ImageColorGenerator
from PIL import Image
from sklearn.utils import resample
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
```

```
/kaggle/input/india-states/Igismap/Indian_States.shp
/kaggle/input/india-states/Igismap/Indian_States.prj
/kaggle/input/india-states/Igismap/Indian_States.dbf
/kaggle/input/india-states/Igismap/Indian_States.shx
/kaggle/input/indian-candidates-for-general-election-2019/LS_2.0.csv
```

Let’s take a look at the dataset :

[2]:

```
vote=pd.read_csv('/kaggle/input/indian-candidates-for-general-election-2019/LS_2.
vote.head()
```

[2]:

	STATE	CONSTITUENCY	NAME	WINNER	PARTY	SYMBOL	GENDER	CRIMINAL\nCASES	AGE	CATEGORY
0	Telangana	ADILABAD	SOYAM BAPU RAO	1	BJP	Lotus	MALE	52	52.0	ST
1	Telangana	ADILABAD	Godam Nagesh	0	TRS	Car	MALE	0	54.0	ST
2	Telangana	ADILABAD	RATHOD RAMESH	0	INC	Hand	MALE	3	52.0	ST
3	Telangana	ADILABAD	NOTA	0	NOTA	NaN	NaN	NaN	NaN	NaN
4	Uttar Pradesh	AGRA	Satyapal Singh Baghel	1	BJP	Lotus	MALE	5	58.0	SC

To help our Analysis, let's go for some...

Data Cleaning

Identifying the Null values in the columns

In [3]:
```
vote.isnull().sum()
```

Out[3]:
```
STATE                                    0
CONSTITUENCY                             0
NAME                                     0
WINNER                                   0
PARTY                                    0
SYMBOL                                 245
GENDER                                 245
CRIMINAL\nCASES                        245
AGE                                    245
CATEGORY                               245
EDUCATION                              245
ASSETS                                 245
LIABILITIES                            245
GENERAL\nVOTES                           0
POSTAL\nVOTES                            0
TOTAL\nVOTES                             0
OVER TOTAL ELECTORS \nIN CONSTITUENCY    0
OVER TOTAL VOTES POLLED \nIN CONSTITUENCY    0
TOTAL ELECTORS                           0
dtype: int64
```

Identifying the null entries in the data

In [6]:
```
vote[vote.SYMBOL.isnull()==True]['NAME'].unique()
```

Out[6]:
```
array(['NOTA'], dtype=object)
```

Cleaning up the Assets and Liabilities columns

In [7]:
```
def value_cleaner(x):
    try:
        str_temp = (x.split('Rs')[1].split('\n')[0].strip())
        str_temp_2 = ''
        for i in str_temp.split(","):
            str_temp_2 = str_temp_2+i
        return str_temp_2
    except:
        x = 0
        return x
vote['ASSETS'] = vote['ASSETS'].apply((value_cleaner))
vote['LIABILITIES'] = vote['LIABILITIES'].apply((value_cleaner))
vote.head()
```

Out[7]:

'ARTY	SYMBOL	GENDER	CRIMINAL\nCASES	AGE	CATEGORY	EDUCATION	ASSETS	LIABILITIES	GENERAL\nVOTES	POSTAL\nVOTES	TOTAL\nVOTES
BJP	Lotus	MALE	52	52.0	ST	12th Pass	3099414	231450	376892	482	377374
TRS	Car	MALE	0	54.0	ST	Post Graduate	18477888	847000	318665	149	318814
INC	Hand	MALE	3	52.0	ST	12th Pass	36491000	15300000	314057	181	314238
NOTA	NaN	NaN	NaN	NaN	NaN	NaN	0	0	13030	6	13036
BJP	Lotus	MALE	5	58.0	SC	Doctorate	74274036	8606522	644459	2416	646875

Renaming the columns

In [8]:
```
vote.rename(columns={"CRIMINAL\nCASES": "CRIMINAL CASES", "GENERAL\nVOTES": "GENERAL VOTES", "POSTAL\nVOTES": "POSTAL VOTES",
vote.head()
```

Out[8]:

	STATE	CONSTITUENCY	NAME	WINNER	PARTY	SYMBOL	GENDER	CRIMINAL CASES	AGE	CATEGORY	EDUCATION	ASSETS	LIABILITIES	GENER VOT
0	Telangana	ADILABAD	SOYAM BAPU RAO	1	BJP	Lotus	MALE	52	52.0	ST	12th Pass	3099414	231450	3768
1	Telangana	ADILABAD	Godam Nagesh	0	TRS	Car	MALE	0	54.0	ST	Post Graduate	18477888	847000	3186
2	Telangana	ADILABAD	RATHOD RAMESH	0	INC	Hand	MALE	3	52.0	ST	12th Pass	36491000	15300000	3140
3	Telangana	ADILABAD	NOTA	0	NOTA	NaN	NaN	NaN	NaN	NaN	NaN	0	0	130
4	Uttar Pradesh	AGRA	Satyapal Singh Baghel	1	BJP	Lotus	MALE	5	58.0	SC	Doctorate	74274036	8606522	6444

Cleaning up the Educational Qualification of the election contestants

In [9]:
```
vote.EDUCATION.unique()
```

Out[9]:
```
array(['12th Pass', 'Post Graduate', nan, 'Doctorate', 'Graduate',
       'Others', '10th Pass', '8th Pass', 'Graduate Professional',
       'Literate', 'Illiterate', '5th Pass', 'Not Available',
       'Post Graduate\n'], dtype=object)
```

In [10]:
```
vote.EDUCATION.replace({'Post Graduate\n':'Post Graduate'},inplace=True)
vote.EDUCATION.unique()
```

Out[10]:
```
array(['12th Pass', 'Post Graduate', nan, 'Doctorate', 'Graduate',
       'Others', '10th Pass', '8th Pass', 'Graduate Professional',
       'Literate', 'Illiterate', '5th Pass', 'Not Available'],
      dtype=object)
```

Identifying the Data Type of the columns

In [11]:
```
vote.dtypes
```

Out[11]:
```
STATE                                  object
CONSTITUENCY                           object
NAME                                   object
WINNER                                  int64
PARTY                                  object
SYMBOL                                 object
GENDER                                 object
CRIMINAL CASES                         object
AGE                                   float64
CATEGORY                               object
EDUCATION                              object
ASSETS                                 object
LIABILITIES                            object
GENERAL VOTES                           int64
POSTAL VOTES                            int64
TOTAL VOTES                             int64
OVER TOTAL ELECTORS IN CONSTITUENCY    float64
OVER TOTAL VOTES POLLED IN CONSTITUENCY float64
TOTAL ELECTORS                          int64
dtype: object
```

Identifying Discrepancy entries in the columns

In [12]:
```
vote[vote['CRIMINAL CASES']=='Not Available'].head()
```

Out[12]:

	STATE	CONSTITUENCY	NAME	WINNER	PARTY	SYMBOL	GENDER	CRIMINAL CASES	AGE	CATEGORY	EDUCATION	ASSETS	LIABILITIES	GE
468	Bihar	BUXAR	Ramchandra Singh Yadav	0	IND	Almirah	MALE	Not Available	42.0	GENERAL	Not Available	0	0	
532	Tamil Nadu	CHIDAMBARAM	SIVAJOTHI M	0	NTK	Ganna Kisan	MALE	Not Available	35.0	SC	Not Available	0	0	
612	Uttar Pradesh	DEORIA	BINOD KUMAR JAISWAL	0	BSP	Elephant	MALE	Not Available	56.0	GENERAL	Not Available	0	0	
613	Uttar Pradesh	DEORIA	NIYAZ AHMED	0	INC	Hand	MALE	Not Available	57.0	GENERAL	Not Available	0	0	
654	Tamil Nadu	DINDIGUL	JOTHIMUTHU, K.	0	PMK	Mango	MALE	Not Available	48.0	GENERAL	Not Available	0	0	

Updating the data types for the analysis

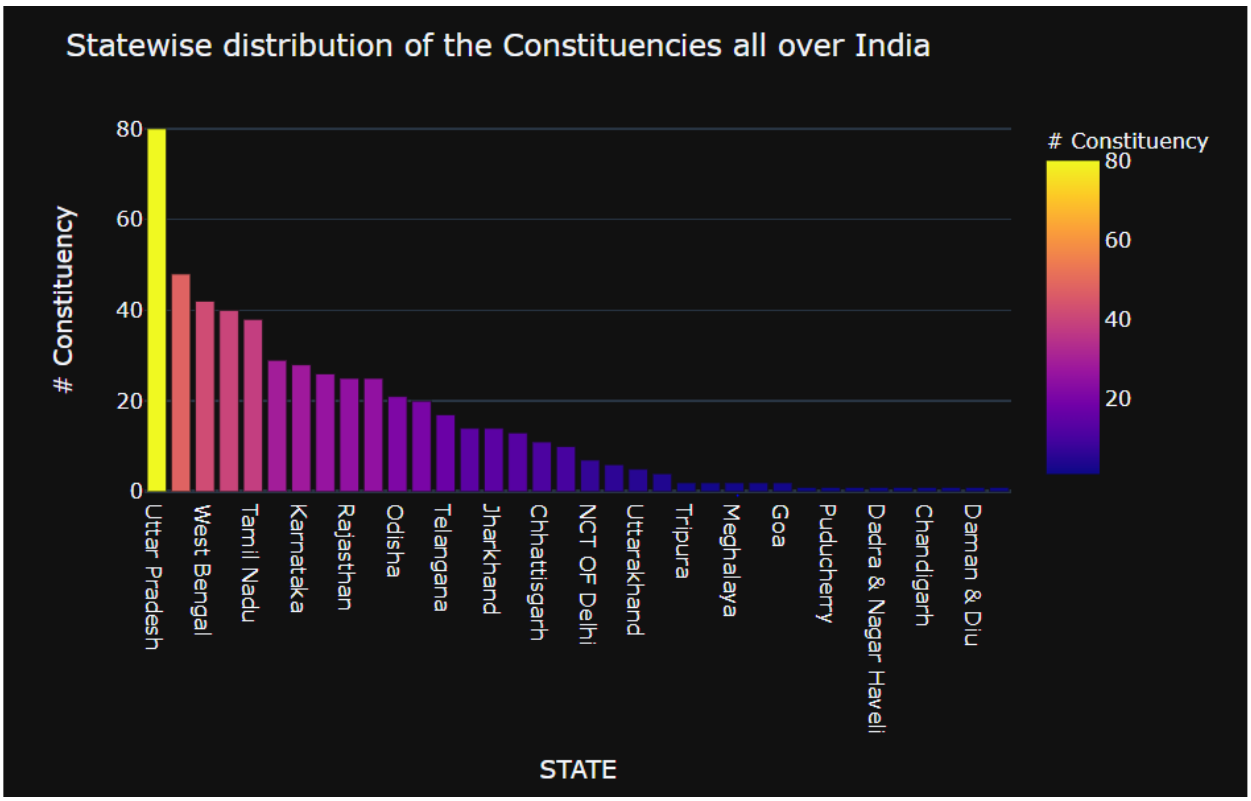
In [13]:
```
vote['ASSETS']=pd.to_numeric(vote['ASSETS'])
vote['LIABILITIES']=pd.to_numeric(vote['LIABILITIES'])
vote['CRIMINAL CASES'].replace({np.NaN:0})
vote['CRIMINAL CASES'] = pd.to_numeric(vote['CRIMINAL CASES'], errors='coerce').fillna(0).astype(np.int64)
```

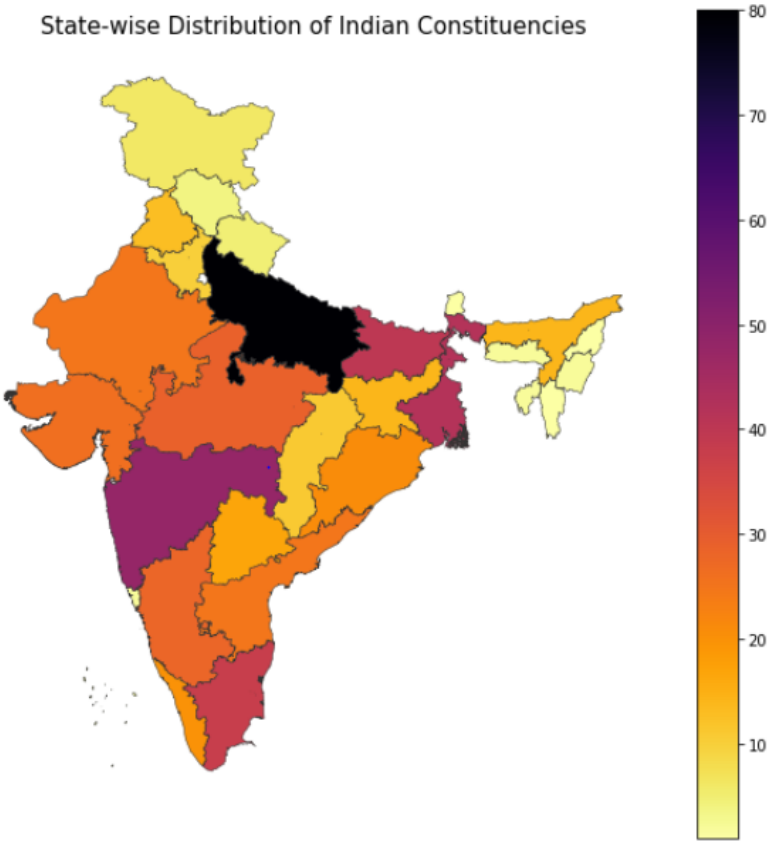
The Analysis

State and Constituency Level Analysis :

The distribution of Constituencies over all the states.

In [30]:
```
st_con=vote.groupby('STATE').apply(lambda x:x['CONSTITUENCY'].nunique()).reset_index(name='# Constituency')
shp_gdf = gpd.read_file('/kaggle/input/india-states/Igismap/Indian_States.shp')
merged = shp_gdf.set_index('st_nm').join(st_con.set_index('STATE'))
fig, ax = plt.subplots(1, figsize=(10, 10))
ax.axis('off')
ax.set_title('State-wise Distribution of Indian Constituencies',
             fontdict={'fontsize': '15', 'fontweight' : '3'})
fig = merged.plot(column='# Constituency', cmap='inferno_r',linewidth=0.5, ax=ax, edgecolor='0.2',legend=True)
st_con.sort_values(by='# Constituency',ascending=False,inplace=True)
fig2 = px.bar(st_con, x='STATE', y='# Constituency',
                    color='# Constituency',
             labels={'pop':'Constituencies of India'})
fig2.update_layout(title_text='Statewise distribution of the Constituencies all over India',template='plotly_dark')
fig2.show()
```

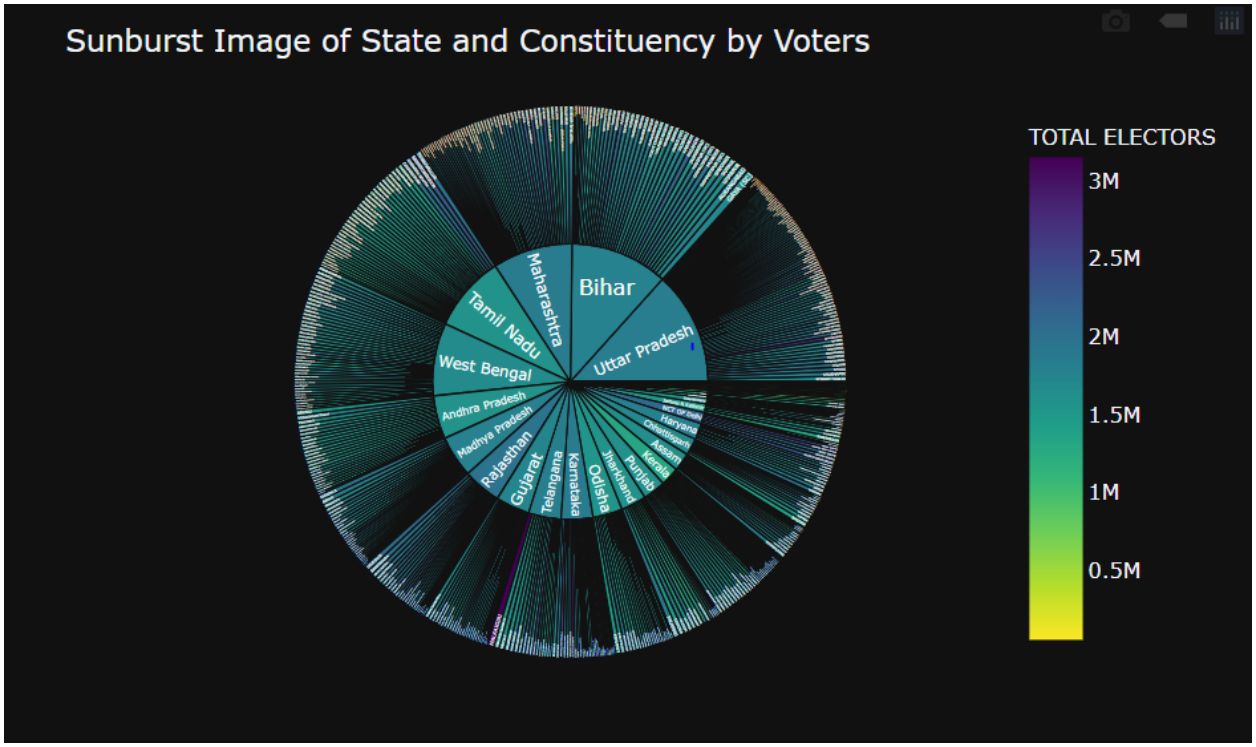




Observation : Uttar Pradesh, Maharashtra and West Bengal- The states have the most number of constituencies. There exists a direct relationship of count of constituencies and population- The constituencies are divided based on the population of 1971- and this shall remain till the year 2026. Although currently Bihar has a higher population, West Bengal has the 3rd highest constituency count based on the above fact.

Lets create a Sunburst image of all the States and Constituencies.

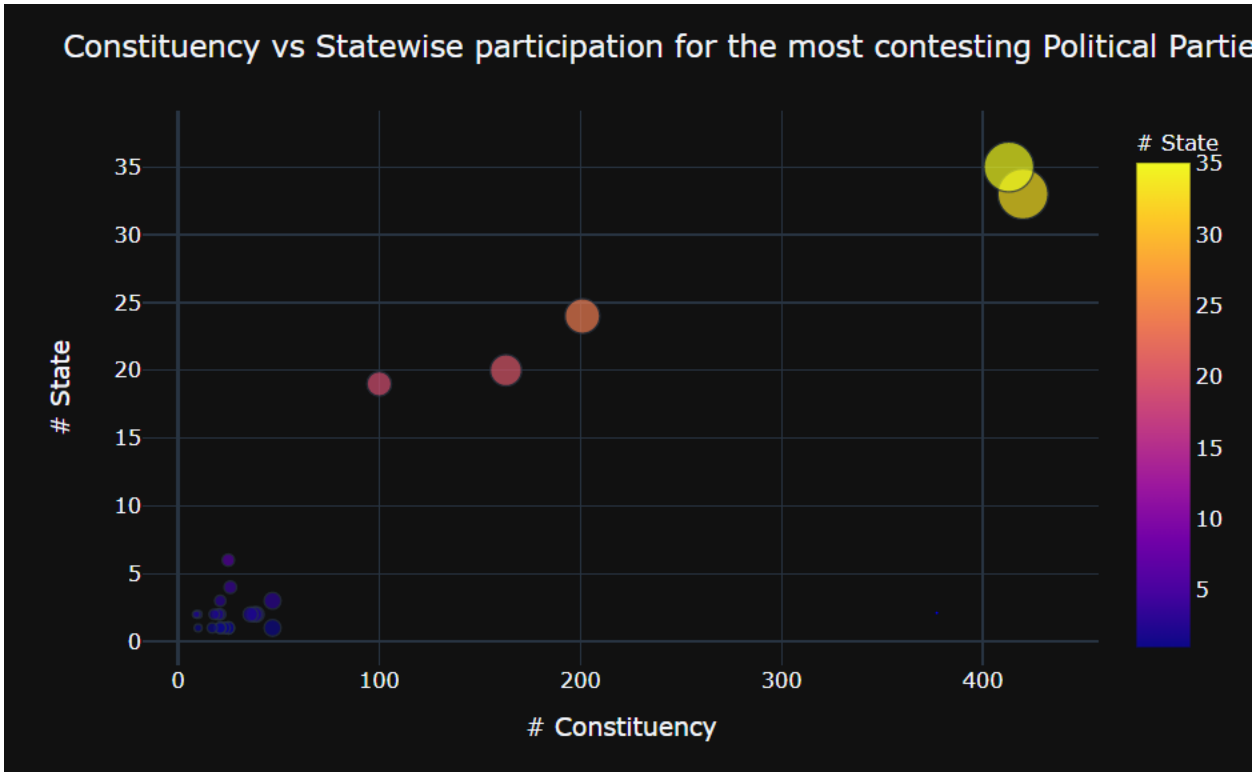
In [15]:
```
st_con_vt=vote[['STATE','CONSTITUENCY','TOTAL ELECTORS']]
fig = px.sunburst(st_con_vt, path=['STATE','CONSTITUENCY'], values='TOTAL ELECTORS',
                  color='TOTAL ELECTORS',
                  color_continuous_scale='viridis_r')
fig.update_layout(title_text='Sunburst Image of State and Constituency by Voters',template='plotly_dark')
fig.show()
```



Party Level Analysis

Which Parties have been present in most constituencies and States?

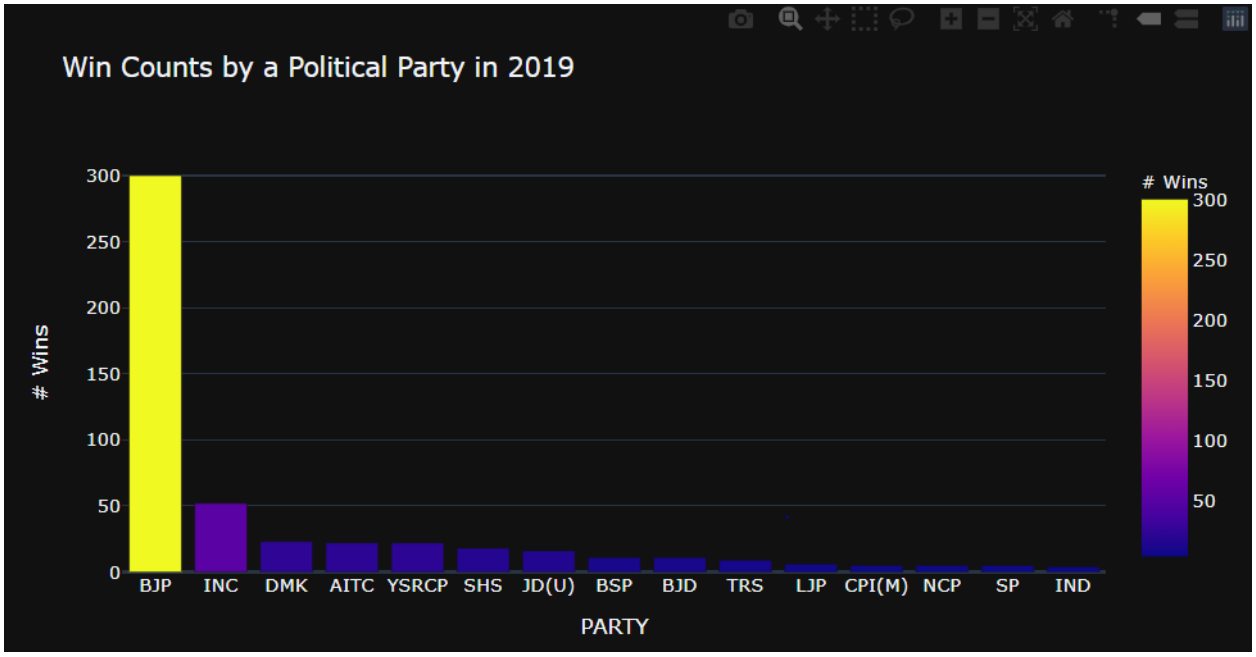
In [16]:
```
vote_prty=vote[vote['PARTY']!='NOTA']
prty_cnt=vote_prty.groupby('PARTY').apply(lambda x:x['CONSTITUENCY'].count()).reset_index(name='# Constituency')
prty_st=vote_prty.groupby('PARTY').apply(lambda x:x['STATE'].nunique()).reset_index(name='# State')
prty_cnt.sort_values(by='# Constituency',ascending=False,inplace=True)
prty_top_cn=prty_cnt[:25]
prty_top_all=pd.merge(prty_top_cn,prty_st,how='inner',left_on='PARTY',right_on='PARTY')
fig = px.scatter(prty_top_all, x='# Constituency', y='# State', color='# State',
                 size='# Constituency', hover_data=['PARTY'])
fig.update_layout(title_text='Constituency vs Statewise participation for the most contesting Political Parties',template='pl
fig.show()
```



Observation : The Bharatiya Janata Party (BJP) and Indian National Congress (INC) have participated in the most number of constituencies all over India. While BJP leads in the number of constituencies contested, INC wins in terms of the number of States. While these are the major parties contesting almost all over India, we see the rest of the parties have restricted themselves to a handful of states.

Which party has won the most constituencies?

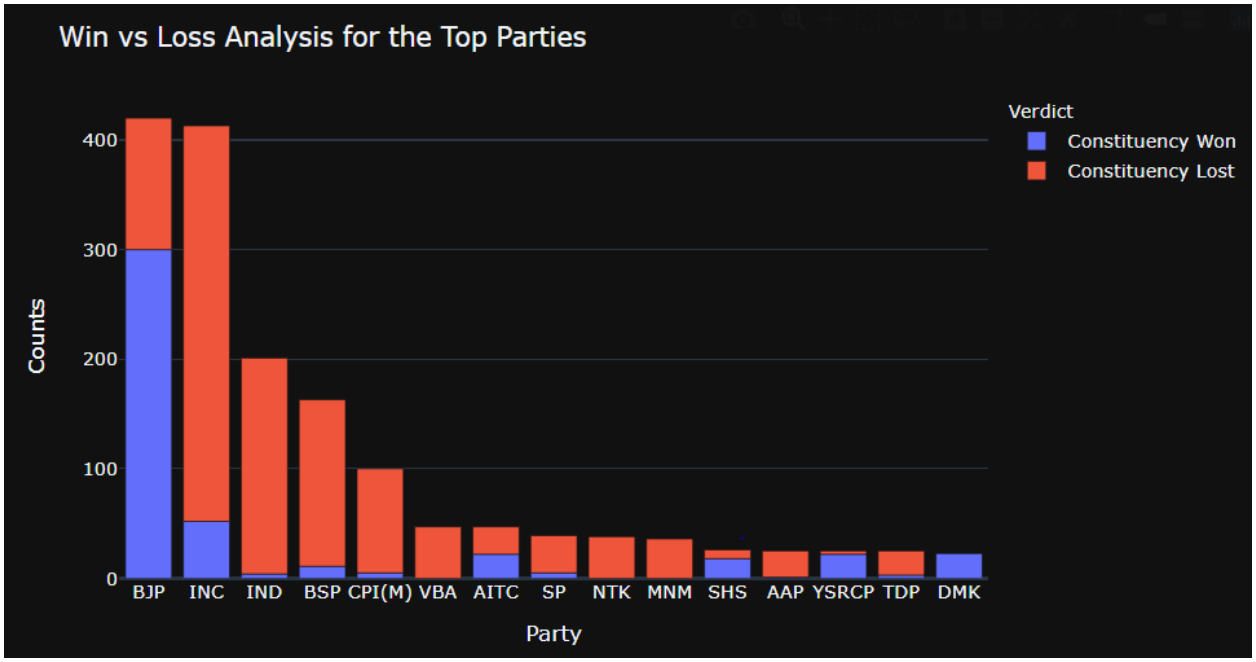
In [18]:
```
part_win=vote.groupby('PARTY').apply(lambda x:x['WINNER'].sum()).reset_index(name='# Wins')
part_win.sort_values(by='# Wins',ascending=False,inplace=True)
top_part_win=part_win[0:15]
fig = px.bar(top_part_win, x='PARTY', y='# Wins',
                    color='# Wins',title='Win Counts by a Political Party in 2019')
fig.update_layout(title_text='Win Counts by a Political Party in 2019',template='plotly_dark')
fig.show()
```



Observation : As seen from the data, In 2019, BJP has won the maximum constituencies all over India. The Image below the introduction also suggests the same. The distribution of all the parties is presented below. INC, who stood 2nd in the number of victories had only 52, which is practically 1/6th of the constituencies won by BJP.

What has been the general Win vs Loss relationship for the Parties in 2019?

In [19]:
```
prty_cnt_win=pd.merge(prty_cnt,part_win,how='inner',left_on='PARTY',right_on='PARTY')
prty_cnt_win['Lost']=prty_cnt_win['# Constituency']-prty_cnt_win['# Wins']
prty_wins_cnt=prty_cnt_win[['PARTY','# Wins']]
prty_wins_cnt['Verdict']='Constituency Won'
prty_loss_cnt=prty_cnt_win[['PARTY','Lost']]
prty_loss_cnt['Verdict']='Constituency Lost'
prty_wins_cnt.columns=['Party','Counts','Verdict']
prty_loss_cnt.columns=['Party','Counts','Verdict']
top_prty_wins_cnt=prty_wins_cnt[:15]
prty_loss_cnt_cnt=prty_loss_cnt[:15]
prt_win_loss=pd.concat([top_prty_wins_cnt,prty_loss_cnt_cnt])
fig = px.bar(prt_win_loss, x='Party', y='Counts', color='Verdict')
fig.update_layout(title_text='Win vs Loss Analysis for the Top Parties',template='plotly_dark')
fig.show()
```

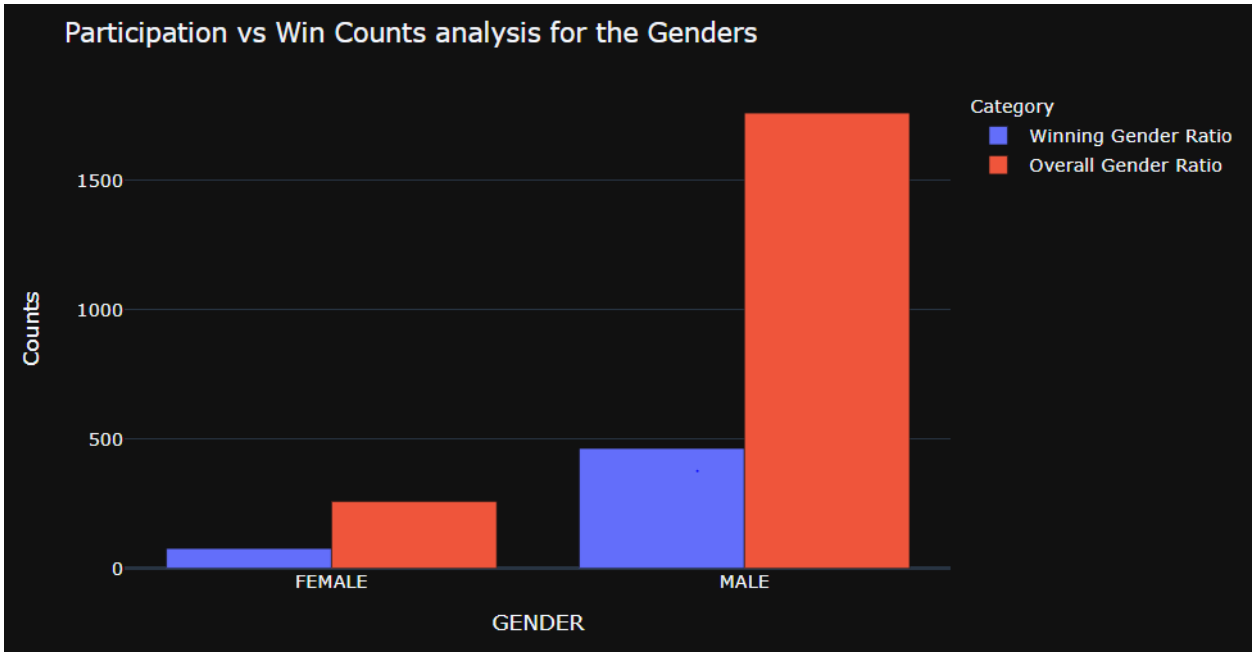


Observation As seen in the above chart, the 2019 elections have been extremely lucky for parties like BJP,SHS or DMK. But it has been a major failure for the rest of the parties, where they have lost more than they won.

Politician Level Analytics

What is the Gender Ratio of the Contestants? Also the Gender Ratio of the Winners?

In [20]:
```
vote_gndr=vote[vote['PARTY']!='NOTA']
gndr_overall=vote_gndr.groupby('GENDER').apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
gndr_overall['Category']='Overall Gender Ratio'
winners=vote_gndr[vote_gndr['WINNER']==1]
gndr_winner=winners.groupby('GENDER').apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
gndr_winner['Category']='Winning Gender Ratio'
gndr_overl_win=pd.concat([gndr_winner,gndr_overall])
fig = px.bar(gndr_overl_win, x='GENDER', y='Counts',
             color='Category', barmode='group')
fig.update_layout(title_text='Participation vs Win Counts analysis for the Genders',template='plotly_dark')
fig.show()
```

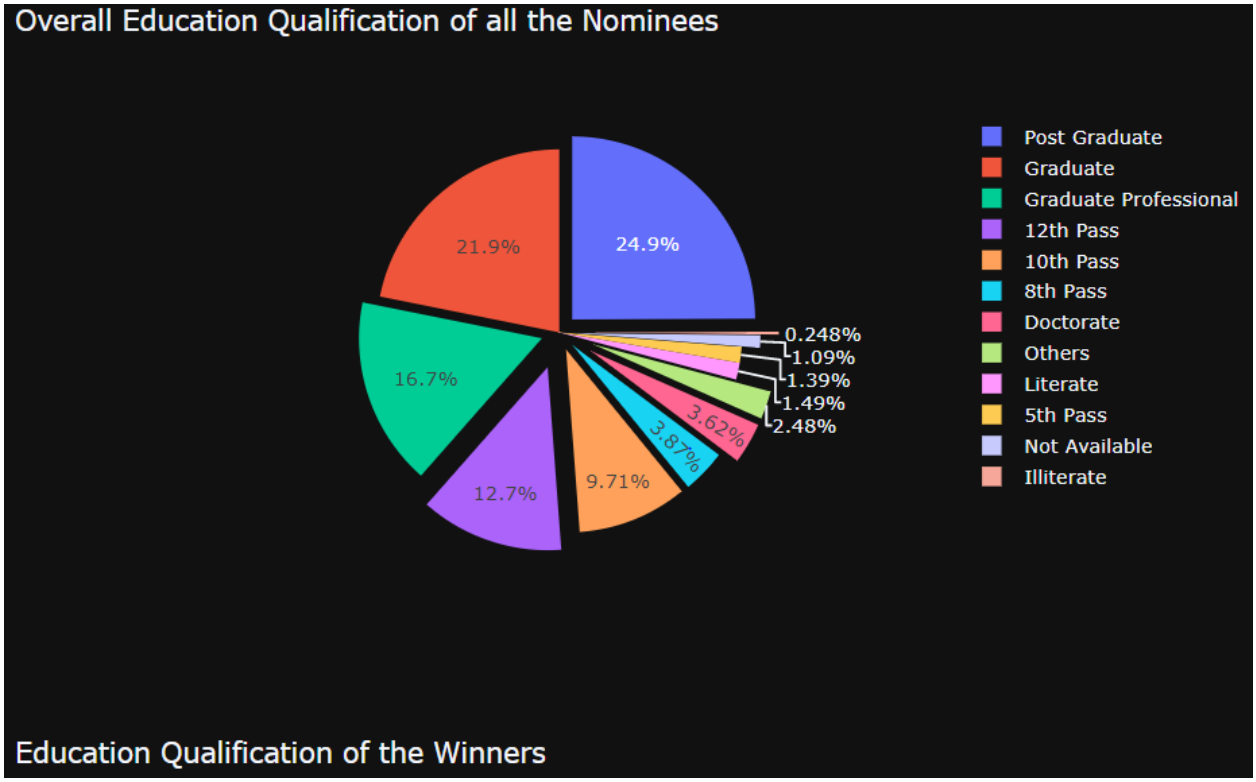


Observation: Out of the total list of participants only 12.78% (258 out of 2018) are female politicians, while 87.21% (1760 out of 2018) are male. Upon considering the winners, 14.1% (76 out of 463) are female politicians, while 85.9% are male politicians. The Gender ratio is not very well distributed as can be seen from the above presentation.

What is the Educational Qualification of our politicians?

In [21]:

```
ed_valid=vote[vote['PARTY']!="NOTA"]
ed_cnt=ed_valid.groupby('EDUCATION').apply(lambda x:x['PARTY'].count()).reset_index(name='Counts')
fig = go.Figure(data=[go.Pie(labels=ed_cnt['EDUCATION'], values=ed_cnt['Counts'], pull=[0.1, 0.2, 0, 0.1, 0.2, 0,0.1, 0.2, 0,
fig.update_layout(title_text='Overall Education Qualification of all the Nominees',template='plotly_dark')
fig.show()
ed_won=ed_valid[ed_valid['WINNER']==1]
ed_win_cnt=ed_won.groupby('EDUCATION').apply(lambda x:x['PARTY'].count()).reset_index(name='Counts')
fig2 = go.Figure(data=[go.Pie(labels=ed_win_cnt['EDUCATION'], values=ed_win_cnt['Counts'], pull=[0.1, 0.2, 0, 0.1, 0.2, 0,0.1
fig2.update_layout(title_text='Education Qualification of the Winners',template='plotly_dark')
fig2.show()
```

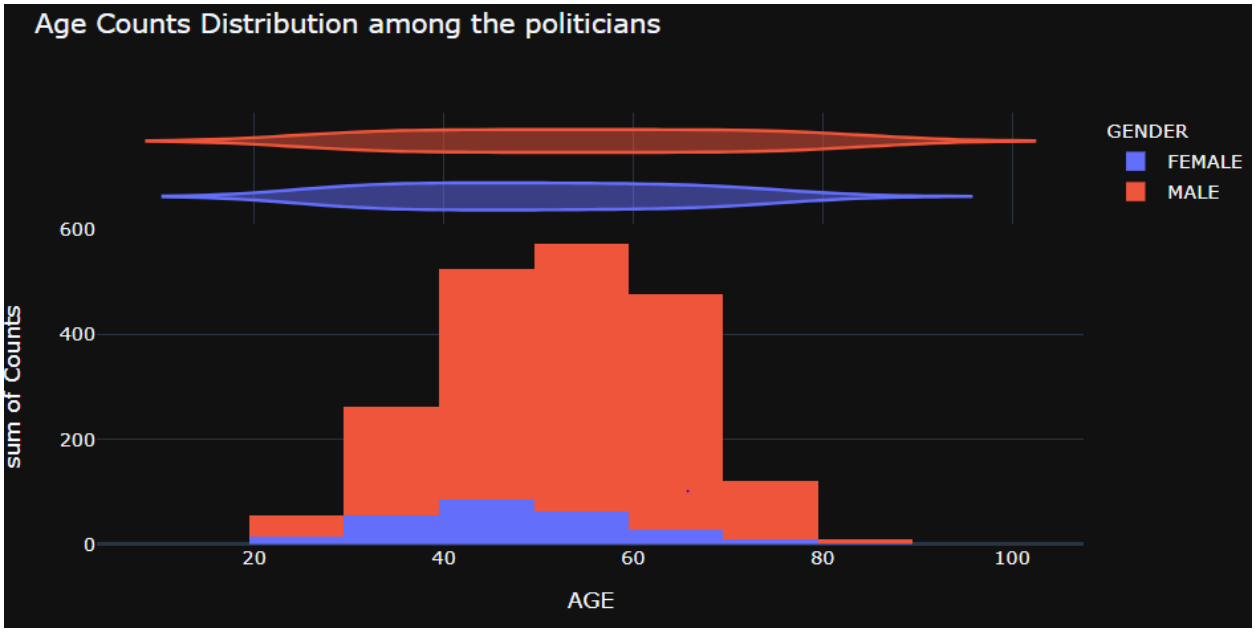


Observation: The total percentage of Graduate educated people contesting in the election is 67.12%, which has increased to 72.17% of the winners. This is actually a positive sign, as educated politicians are a very big factor towards a country's development. But still around 28% of the politicians have received no professional degree. Hope with passing time, we improve upon this factor, and consider the educational qualification as a primary requirement while voting!

What is the relationship of Age and Politics?

In [22]:

```
age_cnt=ed_valid.groupby(['AGE','GENDER']).apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
fig = px.histogram(age_cnt, x="AGE",y='Counts',color='GENDER',marginal='violin',title='Age Counts Distribution among the poli
fig.update_layout(title_text='Age Counts Distribution among the politicians',template='plotly_dark')
fig.show()
```

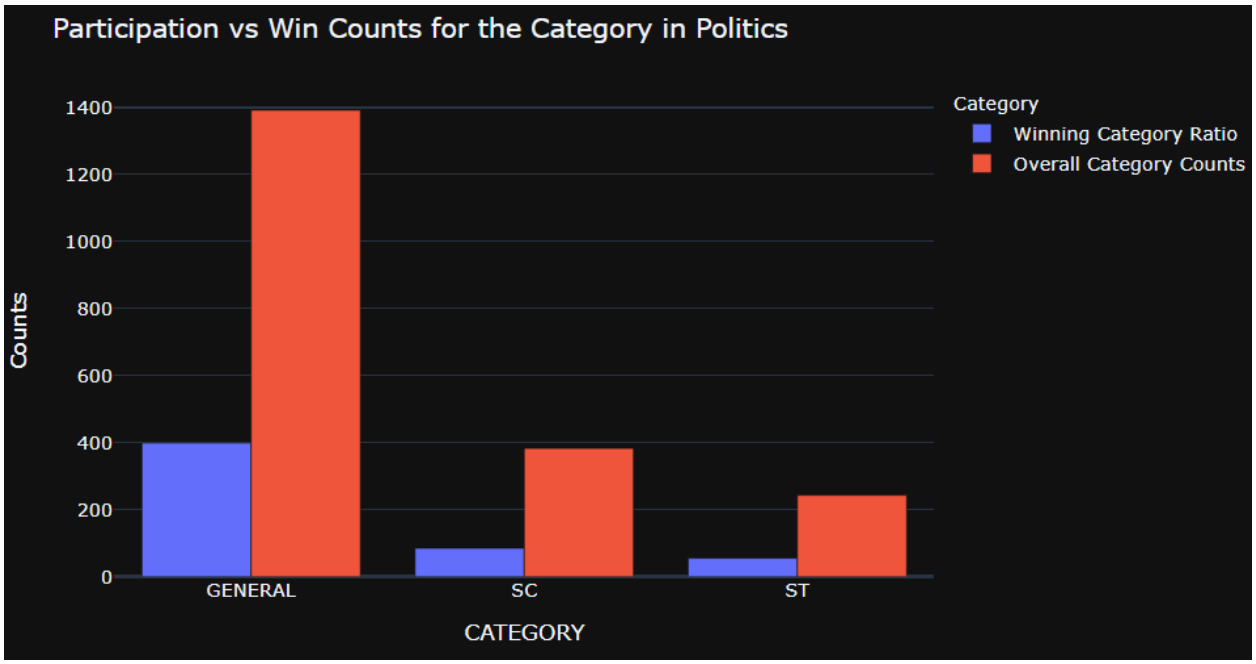


Observation : Most Number of female politicians have their average age between 45-50, while for male politicians, it ranges from 50-60 range. The average age of male politicians is more as compared to female politicians contesting for the Lok Sabha elections.

What relation does the Politician category have with the election results?

In [23]:

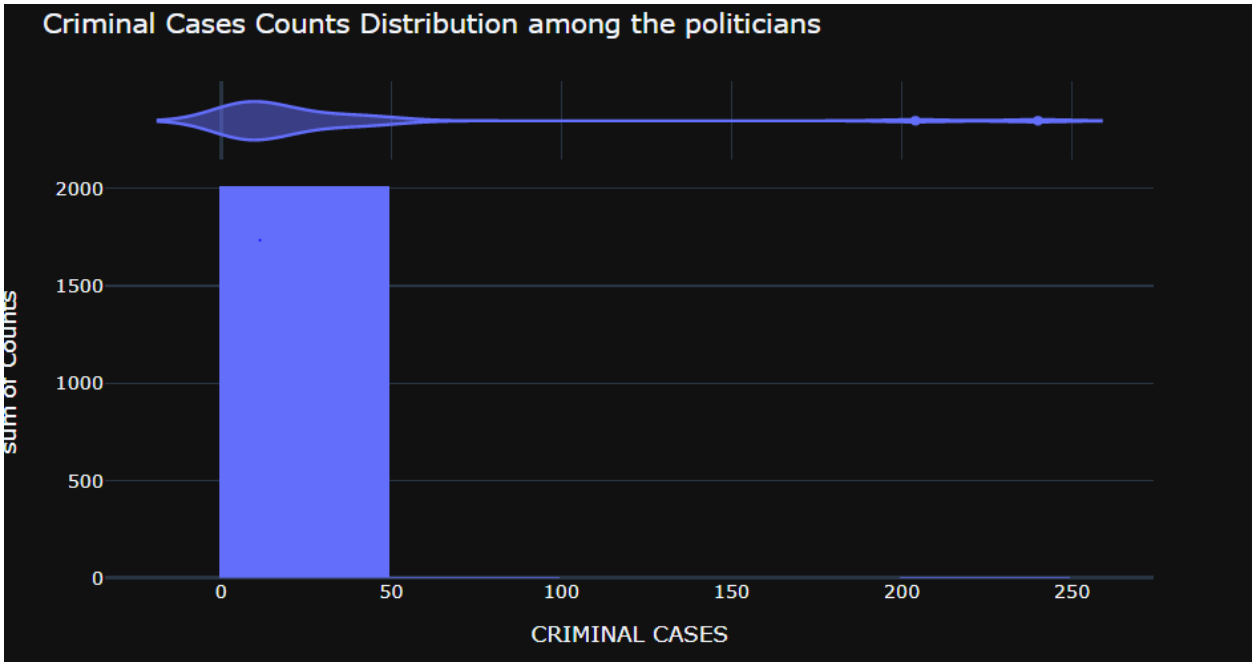
```
vote_cat=vote[vote['PARTY']!='NOTA']
cat_overall=vote_cat.groupby('CATEGORY').apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
cat_overall['Category']='Overall Category Counts'
winners_cat=vote_gndr[vote_gndr['WINNER']==1]
cat_winner=winners_cat.groupby('CATEGORY').apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
cat_winner['Category']='Winning Category Ratio'
cat_overl_win=pd.concat([cat_winner,cat_overall])
fig = px.bar(cat_overl_win, x='CATEGORY', y='Counts',
             color='Category', barmode='group')
fig.update_layout(title_text='Participation vs Win Counts for the Category in Politics',template='plotly_dark')
fig.show()
```



Observation: The Category participation of General-SC-ST have been in the ratio of 68.97:18.97:12.04- while as of the winners, the ratios have been modified to 74.02:15.76:10:20.

Have the politicians been involved with criminal activities?

In [24]:
```
crim_cnt=ed_valid.groupby('CRIMINAL CASES').apply(lambda x:x['NAME'].count()).reset_index(name='Counts')
fig = px.histogram(crim_cnt, x='CRIMINAL CASES',y='Counts',marginal='violin')
fig.update_layout(title_text='Criminal Cases Counts Distribution among the politicians',template='plotly_dark')
fig.show()
```



Observations Many politicians have been associated with criminal activities. Always these cases need not be genuine, but obviously, when it's multiple- this is a serious issue. We must take the responsibility while voting, as it's our duty to choose the right person- as a duty towards the nation.

Prediction

Note: We shall use Random Classifier to predict the results of the election.

In [26]:

```
vote_df=vote[vote['PARTY']!='NOTA']
vote_df['GENDER'].replace({'MALE':1,'FEMALE':2},inplace=True)
vote_df['CATEGORY'].replace({'GENERAL':1,'SC':2,'ST':3},inplace=True)
i=1
parties_dict={}
for j in vote_df['PARTY']:
    if j in parties_dict:
        continue
    else:
        parties_dict[j]=i
        i+=1
vote_df['PARTY'].replace(parties_dict,inplace=True)
a=1
edu_dict={}
for b in vote_df['EDUCATION']:
    if b in edu_dict:
        continue
    else:
        edu_dict[b]=a
        a+=1
vote_df['EDUCATION'].replace(edu_dict,inplace=True)
df1 = vote_df[['STATE','CONSTITUENCY','WINNER','PARTY','SYMBOL','GENDER','CRIMINAL CASES','AGE','CATEGORY','EDUCATION','TOTAL
num_cols = ['PARTY','EDUCATION','CRIMINAL CASES','AGE','TOTAL VOTES','TOTAL ELECTORS','ASSETS','CATEGORY','LIABILITIES','GEND
dataset = pd.get_dummies(df1)
from sklearn.preprocessing import StandardScaler
standardScaler = StandardScaler()
scaling_columns = num_cols
dataset[scaling_columns] = standardScaler.fit_transform(dataset[scaling_columns])
dataset.head()
```

Out[26]:

	WINNER	PARTY	GENDER	CRIMINAL CASES	AGE	CATEGORY	EDUCATION	TOTAL VOTES	TOTAL ELECTORS	ASSETS	...	SYMBOL_Tiller	SYMBOL_Tractor Chalata Kisan	S
0	1	-0.628979	-0.382872	6.620242	-0.023051	2.251127	-1.242514	0.332957	-0.538876	-0.311731	...	0	0	
1	0	-0.583879	-0.382872	-0.190426	0.145491	2.251127	-0.872637	0.103781	-0.538876	-0.274420	...	0	0	
2	0	-0.538780	-0.382872	0.202498	-0.023051	2.251127	-1.242514	0.085873	-0.538876	-0.230717	...	0	0	
4	1	-0.628979	-0.382872	0.464446	0.482577	0.816718	-0.502761	1.387657	0.876978	-0.139047	...	0	0	
5	0	-0.493681	-0.382872	-0.190426	-0.444408	0.816718	-0.872637	0.559766	0.876978	0.005336	...	0	0	

5 rows × 712 columns

Checking for accuracy

In [32]:
```
randomforest_classifier= RandomForestClassifier(n_estimators=38,random_state=0)
score=cross_val_score(randomforest_classifier,X,y,cv=10)
print('% Accuracy :', round(score.mean()*100,4))
```

```
% Accuracy : 96.2131
```

Observation : We have reached an accuracy percentage of 96.21% with our model. Will work on to improve the model further- to increase the accuracy.
This is the model that helped us to obtain the highest accuracy.

Summary

After analyzing the voting pattern of Indians, we have come upon multiple conclusions. Let us look at them in the below points:

1. In 2019, the Bharatiya Janata Party (BJP) has claimed the most number of seats all over India- dominating clearly over all the national and the state specific parties. The outreach created by them has really made them claim their position in the Lok Sabha.
2. Generally, education is regarded as a significant factor while voting. Around 72%+ of the winners are having a graduate+ degree. This is a significant factor for a developing nation like India- where we need educated politicians to lead the nation.
3. The general participation of Female politicians is much less than the male politicians. A balanced gender ratio would be good enough- as people from all sectors would be able to contribute to the progress of the society.
4. The average age of politicians is high. It might be an advantage if younger politicians contest and are able to lead the nation. Although it is never a parameter to think that the older politicians wouldn't be able to perform their tasks well, it comes with the risk of health and age related issues.
5. Criminal activities are a serious concern- as the people who are elected by us- indirectly are the face of the nation. They should be more of a role model-rather than someone who is not respected.
6. As our prediction suggests, we have reached an accuracy of 96.2% in estimating the winning participants using the Random classifier model. We can work upon to improve this model further in the future.