

Importing Libraries

```
In [2]: import pandas as pd
import numpy as np

In [3]: Data=pd.read_csv(r"C:\Users\harsh\OneDrive\Desktop\CSV_files\generic_ballot_polls.csv")
```

DATA CLEANING

1. Replace NA and NaN values with 0.

```
In [4]: Data_clean=Data.replace(np.NaN,0,inplace=True)

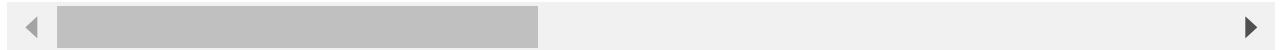
In [5]: Data_clean=Data.replace(pd.NA,0,inplace=True)

In [6]: Data
```

Out[6]:	question_id	poll_id	race_id	cycle	state	pollster_id	pollster	sponsor_ids	sponsors	partisan
0	148863	77170	8990	2022	0.0	568	YouGov	352	Economist	0
1	148835	77166	8990	2022	0.0	1189	Morning Consult	0	0	0
2	148899	77168	8990	2022	0.0	1508	Harris Insights & Analytics	763	Harvard	0
3	148624	77050	8990	2022	0.0	568	YouGov	352	Economist	0
4	148793	77149	8990	2022	0.0	736	NBC News/The Wall Street Journal	0	0	0
...
107	140128	74469	8990	2022	0.0	927	GQR Research (GQRR)	0	0	0
108	140129	74469	8990	2022	0.0	927	GQR Research (GQRR)	0	0	0
109	140351	74400	8990	2022	0.0	1302	Echelon Insights	0	0	0
110	139356	73755	8990	2022	0.0	294	McLaughlin & Associates	0	0	0

	question_id	poll_id	race_id	cycle	state	pollster_id	pollster	sponsor_ids	sponsors	partisan
111	139355	73339	8990	2022	0.0	294	McLaughlin & Associates	49	Newsmax	0

112 rows × 32 columns



In [7]:

```
Data.isna().sum
```

Out[7]:

```
<bound method NDFrame._add_numeric_operations.<locals>.sum of
race_id cycle state pollster_id pollster \
0      False  False  False  False  False  False  False  False
1      False  False  False  False  False  False  False  False
2      False  False  False  False  False  False  False  False
3      False  False  False  False  False  False  False  False
4      False  False  False  False  False  False  False  False
..      ...      ...      ...      ...      ...      ...      ...
107     False  False  False  False  False  False  False  False
108     False  False  False  False  False  False  False  False
109     False  False  False  False  False  False  False  False
110     False  False  False  False  False  False  False  False
111     False  False  False  False  False  False  False  False

sponsor_ids sponsors partisan ... sponsor_candidate tracking \
0      False  False  False  ...      False  False
1      False  False  False  ...      False  False
2      False  False  False  ...      False  False
3      False  False  False  ...      False  False
4      False  False  False  ...      False  False
..      ...      ...      ...      ...      ...      ...
107     False  False  False  ...      False  False
108     False  False  False  ...      False  False
109     False  False  False  ...      False  False
110     False  False  False  ...      False  False
111     False  False  False  ...      False  False

nationwide_batch created_at notes url stage dem rep ind
0      False      False  False  False  False  False  False  False
1      False      False  False  False  False  False  False  False
2      False      False  False  False  False  False  False  False
3      False      False  False  False  False  False  False  False
4      False      False  False  False  False  False  False  False
..      ...      ...      ...      ...      ...      ...      ...
107     False      False  False  False  False  False  False  False
108     False      False  False  False  False  False  False  False
109     False      False  False  False  False  False  False  False
110     False      False  False  False  False  False  False  False
111     False      False  False  False  False  False  False  False
```

[112 rows x 32 columns]>

1. Drop Empty cells

In []:

