**2023 Election Analysis Report**

In this project, I analyzed the 2023 Nigerian election dataset. I broke down my analysis into regional, state and party analysis and create chart and discovered patterns. My process of analysis began with the data cleaning and standardization process, I then proceeded to exploratory data analysis, uncovering patterns and outliers.

**The Dataset**

This dataset was gotten online and is saved as dataset.csv. I overviewed the dataset using excel to get a picture of what the data looks like. The dataset contained 16 columns of strings, integers and floats and 37 rows. The data listed the 37 states that participated in the election, pairing them with the votes the parties got, total votes and highest votes.

**The original columns**

|  |
| --- |
| STATE |
| REGION |
| APC |
| APC % |
| PDP |
| PDP % |
| LP |
| LP % |
| NNPP |
| NNPP % |
| OTHERS |
| OTHERS % |
| HIGHEST VOTES |
| PARTY WITH HIGHEST VOTES |
| HIGHEST PERCENTAGE |
| TOTAL VOTES COUNTED |

**Data Cleaning and Standardization**

The process of data cleaning and standardization was done mostly in python using **pandas** and **numpy** libraries. This process was done in **dcs.ipynb**

1. Merged the party columns to one column, creating a new column, **Party** and their values as **Votes.** This process multiplied the rows by 5 as there were 5 parties represented in the dataset, 37 x 5 = 185.
2. Created a new column Votes100, which represents that percentage of votes in the states. The votes 100 is a representation of the APC%, LP%, NNPP%, PDP% and OTHERS% columns which were originally dropped.
3. Sorted the values by states and reset the index
4. Reduced that float decimal places to 2
5. Rearranged and reordered the columns to soothe the analysis
6. Changed the columns to lower pipe case which is best suited for sql
7. Save the dataset as **dataset\_clean.csv**

This process created a cleaner and more standard dataset with 185 rows and 7 columns.

**Exploratory Data Analysis**

This process involved understanding the dataset and breaking it down to further analyze it. I applied series of mathematical and statistical steps to the dataset. The dataset I worked with was the **dataset\_clean.csv** which I had already cleaned and standardized in python. This process was done in 2 phases. The python phase and the SQL phase.

In the python phase, the data was explored using the **numpy**, **pandas** and **matplotlib** libraries in python. This process was done in **eda.ipynb.** In the SQL phase, the dataset was imported to a mysql database and further explored.

1. **The overview**

I confirmed the shape and the size of the dataset. The shape was 185 rows and 7 columns as produced in the data cleaning phase

1. **Global variables**

The regions, parties and states were selected and confirmed to be 6 regions, 5 parties (Others is a combination of other parties) and the states where confirmed to be 37 states

1. **Voting Power by region**

The regional voting power was confirmed by producing a dataset of the regions and sum of the votes in that region and the percentage of votes in a region over the total number of votes in the region. **North West** with 7 states had the highest voting power of **27.83%** while **South east** with 5 states had the least voting power of **9.29%.** This was represented in a bar chart and pie chart respectively

1. **Party performance by region**

The regions were further analyzed to observe how each party performed across the regions. The dataframe was broken down to regions, parties, sum of votes and the percentage of votes each party gathered from a region. This reveals party dominance in the regions, like the instance the LP dominance in the southeast, APC dominance in the South West and so on.

1. **Voting power by states**

This analysis produced a dataframe of states, sum of votes and percentage of votes. The rows are individual states and the votes the gathered and the percentage of votes which is their voting power across the country. This analysis showed that Kano has the highest voting power of 7.08% across the nation and Bayelsa had the lowest voting power of 0.69% across the nation. A bar chart was plotted to show this analysis

1. **Party dominance across the nation**

This analysis involved pairing party, votes and votes100. This showed the APC more dominant on individual votes than LP who came 2nd and PDP who came 3rd

1. **Party dominance across states**

This analysis involved pairing states by party, votes and votes100. This showed the LP was very dominant in Anambra, PDP very dominant in Adamawa and APC very dominant in most south western state.

**Final Report**

The above analysis produced the following reports

1. On regional dominance, APC party was more dominant across the regions, dominating 3 regions, while LP dominated 2 and PDP 1
2. On individual votes, APC gathered more individual votes than any other party, PDP came 2nd and LP 3rd.
3. State dominance was equal among APC, PDP and LP but NNPP dominated only 1 state

All files for this project can be found in the containing folder of this report.