**TECHNICAL REPORT ON POLLING UNIT OUTLIER DETECTION USING GEOGRAPHIC PROXIMITY IN LAGOS STATE**

**Introduction**

The last election conducted in Nigeria was alleged to be full of disparities and inaccuracies. This technical report aims to analyze these disparities and identify outliers in the voting patterns across different polling units, focusing on four major political parties: APC, PDP, NNPP, and Labour Party in Lagos State. By leveraging geographic proximity and voting data, we can pinpoint anomalies that warrant further investigation.

DATASET -

**Methodology**

The methodology consists of several key steps: geocoding, data cleaning, data grouping, neighbor identification, outlier score calculation, and result analysis.

1. **Geocoding**:
   * Geocoding was performed using OpenCageAPI and Geocode Earth to obtain the latitude and longitude of each polling unit based on its address.
   * This step ensured that each polling unit had accurate geographic coordinates necessary for distance calculations.
2. **Data Cleaning**:
   * Any polling units that could not be geocoded accurately were dropped from the dataset.
   * This step ensured that the analysis was based on reliable geographic data.
3. **Data Grouping by Wards**:
   * The dataset is grouped by wards to isolate polling units within each ward.
   * This allows for localized analysis, ensuring that comparisons are made within the same administrative boundaries.
4. **Neighbor Identification**:
   * For each polling unit within a ward, neighboring polling units within a 1 km radius are identified using the Haversine formula. This formula calculates the great-circle distance between two points on the Earth's surface.
   * The neighbors are stored in a dictionary where each key is a polling unit code, and the value is a list of neighboring polling units.
5. **Outlier Score Calculation**:
   * For each polling unit, the mean votes for each party (APC, LP, PDP, NNPP) are calculated from its neighboring units.
   * The deviation of the polling unit's votes from this mean is computed to derive the outlier score. This score quantifies how different the votes of a polling unit are compared to its neighbors.
   * The outlier score is normalized by dividing the absolute deviation by the mean neighbor votes, avoiding division by zero by checking if the mean is greater than zero.
6. **Sorting and Reporting**:
   * The outlier scores are sorted to identify the most significant outliers.
   * A detailed report is generated, highlighting the top 3 outliers and their neighboring units.

Detailed Steps and Code Explanation

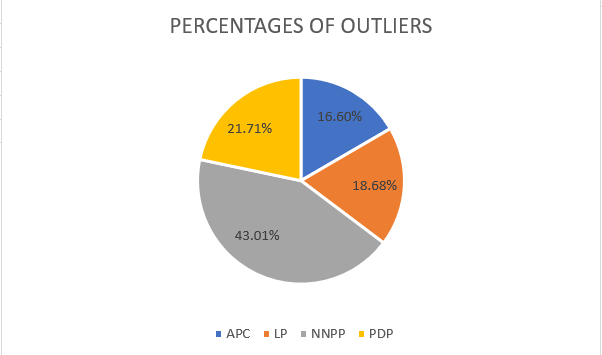
The links to the code can found

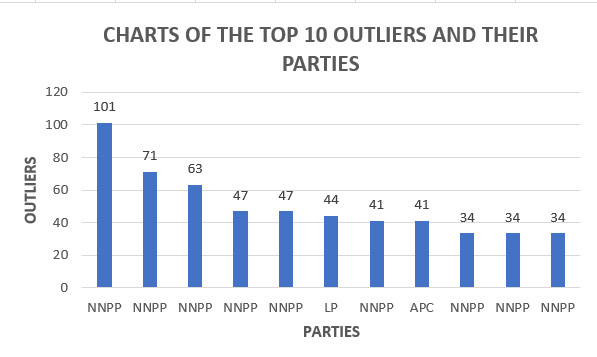
Geocoding -

Data cleaning -

Outlier detection -

Results





From the analysis, it is evident that the NNPP exhibits the most significant outliers, followed by PDP, APC, and LP. This suggests notable disparities in the voting patterns of these parties compared to their neighboring polling units. These disparities might be due to various factors such as localized political influence, voter behavior anomalies, or potential data inaccuracies

Top 3 Outliers:

1. Polling Unit: 24-20-05-019 in Ward: YABA/OJUELEGBA

- Party: NNPP

- Outlier Score: 101.45

- Explanation: This polling unit has a significantly different number of votes for NNPP compared to its neighboring units.

2.Polling Unit: 24-08-02-082 in Ward: VICTORIA ISLAND II

- Party: NNPP

- Outlier Score: 71.0

3. Polling Unit: 24-10-08-125 in Ward: IJAIYE/OJOKORO

- Party: NNPP

- Outlier Score: 63.0

**Conclusion**

The analysis reveals significant voting disparities across several polling units, particularly for the NNPP. These outliers indicate irregularities that warrant further investigation to determine their underlying causes. Potential factors contributing to these anomalies could include localized political dynamics, specific voter behavior, or data reporting errors. Addressing these issues is crucial for enhancing the integrity and accuracy of future elections.