HNG 12
(2025)

Advanced Geospatial Analysis for Election Integrity- Nigeria

CHIDINMA UKANDU

Introduction

The Nigerian electoral system, like many in emerging democracies, has often been scrutinized for irregularities. With the advancement of geospatial and statistical techniques, there is now an opportunity to improve the integrity of election results through data science. This report presents an in-depth spatial-temporal analysis of electoral data in Anambra State, Nigeria, by identifying anomalies using machine learning, cluster analysis, and spatial statistics.

This study aims to:

- Detect and explain outliers in voting patterns using advanced methods.
- Compare demographic variables with voter turnout.
- Provide visual insights that inform electoral bodies of potential risks or manipulations.
-

1.1 **Geospatial Clustering (DBSCAN)**

We applied the DBSCAN algorithm with varying distance parameters (500m, 1000m, 2000m) to detect spatial clusters and outliers across polling units. DBSCAN groups dense areas into clusters while isolating points in sparse areas as outliers.

Parameters:

- eps: 500m, 1000m, 2000m (adjusted to reflect dense-to-loose cluster detection)
- min_samples: 5

1.2 **Local Spatial Autocorrelation (Local Moran's I)**

Local Moran's I was used to examine spatial autocorrelation for each major party (LP, APC, PDP, NNPP). This detects high-high and low-low clusters, highlighting unusual patterns in neighborhood voting behavior.

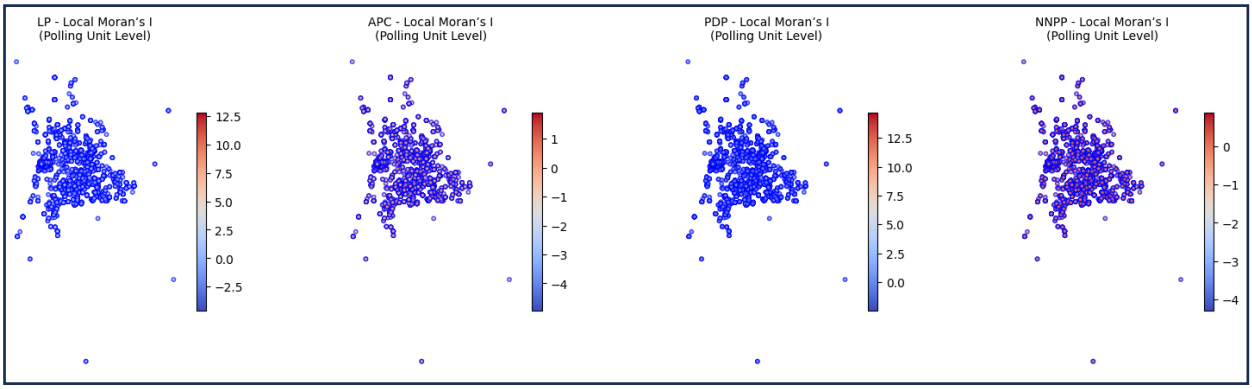

Fig 1

1.3 Getis-Ord Gi* Statistic

Gi* hot spot analysis identifies statistically significant "hot spots" of intense support. A red-to-blue gradient indicates areas of statistically clustered high or low values, confirming spatial significance of partisan vote density.

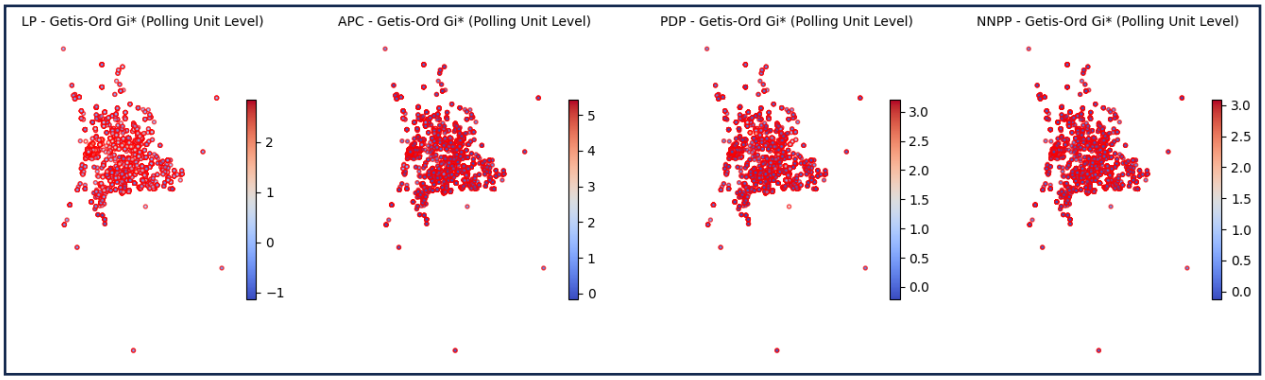

Fig 2: Gi* visualization from code outputs here

Statistical Outlier Scoring

Outliers were scored using a combination of statistical Z-scores, IQR thresholds, and DBSCAN isolation flags. These were combined into a final_outlier_flag used across visuals and summary stats.

This report presents a comprehensive analysis of spatial-temporal anomalies in election data from Anambra State, Nigeria, with a focus on the 2023 general elections and historical voting trends. The goal is to leverage geospatial clustering, statistical validation, machine learning techniques, and socio-demographic indicators to uncover unusual patterns and inform electoral integrity assessments.

The dataset combines three main tables: 2023_Election, Historical_Votes, and Socio_Economic indicators. These were unified via a YearTable to allow consistent year-based filtering. The dashboard developed in Power BI showcases dynamic insights across party performances (PDP, APC), polling unit behavior, demographic influence, and voter turnout.

2. Methodologies Employed

2.1 Geospatial Clustering (DBSCAN) DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was used to identify natural clusters of polling units based on latitude and longitude. It enabled detection of outlier polling units that deviated from expected geospatial density patterns.

- Input variables: Latitude, Longitude
- Parameters: Epsilon = 0.5km and 1km for robustness; MinPts = 5
- Result: Cluster labels and noise points (marked as -1) assigned to each polling unit

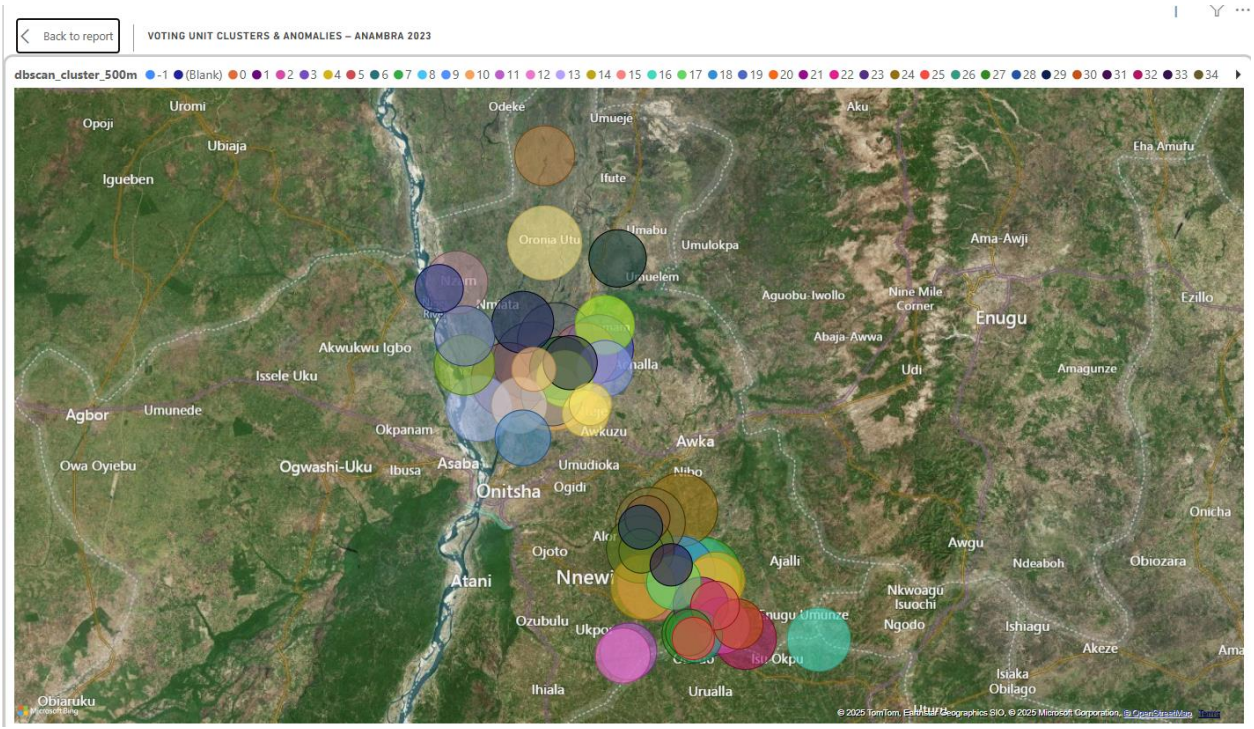

Fig3: Voting unit clustering and anomalies

2.2 Statistical Methods for Anomaly Detection

A combined_outlier_score metric was created using multiple indicators such as registered voters, voter turnout rates, and geolocation deviation. Thresholds were computed using z-scores and IQR methods to classify polling units as statistical outliers.

- Flagged values: Above threshold values for combined scores
- Indicator used: final_outlier_flag

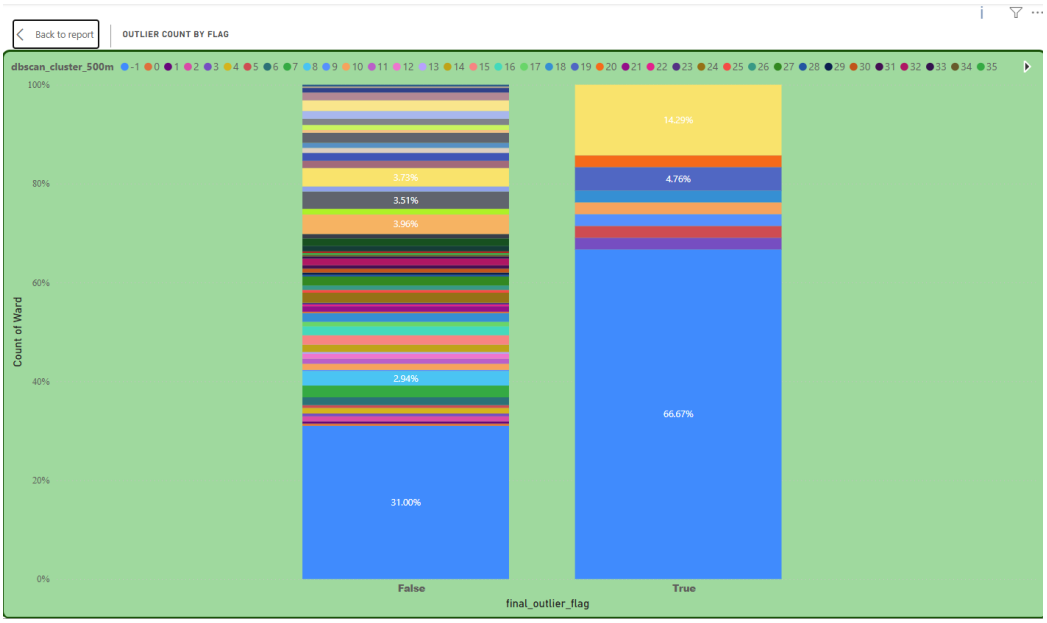

Fig 4: Count of Outliers in 2023

2.3 Machine Learning Validation

To validate detected anomalies, a correlation and regression analysis were applied:

- Population vs Accredited Voters (Scatter Plot)
- Population vs Registered Voters (Combo Chart)

The aim was to test whether identified outliers followed logical voter behavior patterns relative to socio-economic indicators. Discrepancies reinforced their classification as outliers.

Fig 3: Scatter Plot - Population vs Voter Engagement Correlation

2.4 Demographic Integration

Demographic fields such as population totals, urban vs rural distribution, and access to electricity were merged using the Socio_Economic dataset. The influence of these features on voting behavior was examined over time.

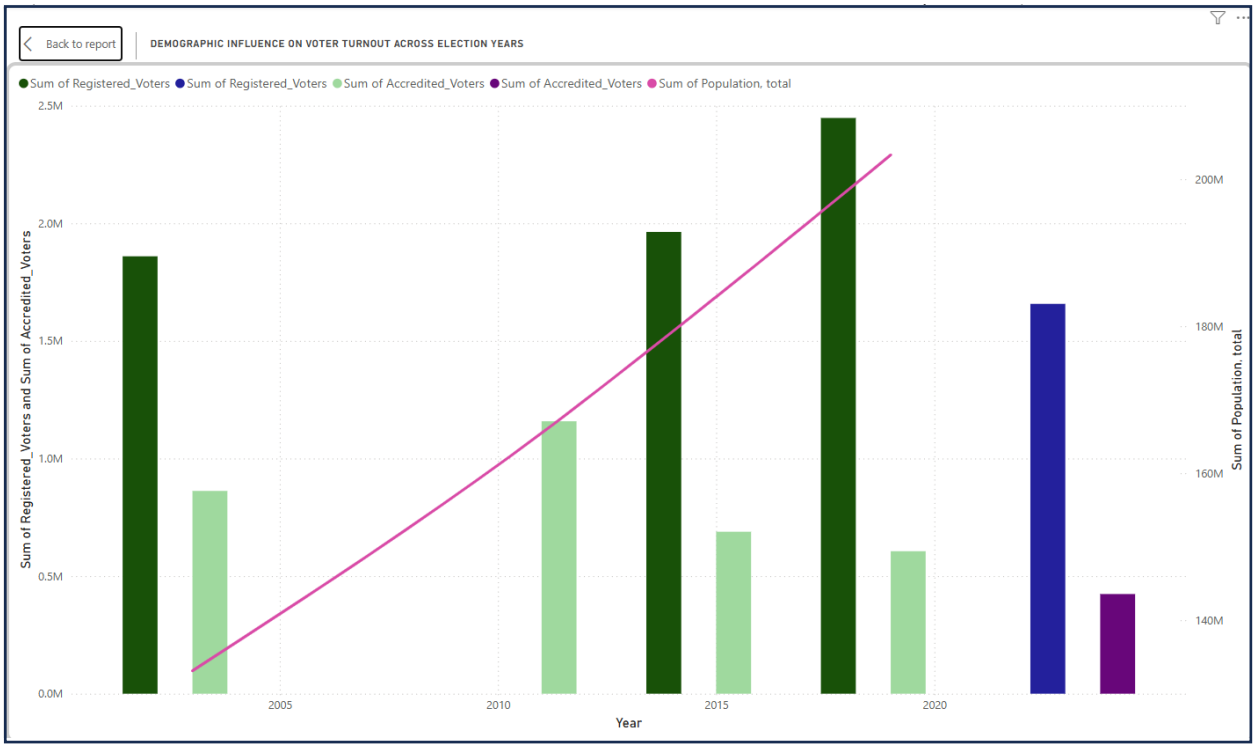

Fig 5: Demographic influence on voters turnout across years in Anambra

3. Top 5 Outlier Polling Units

Based on DBSCAN results and anomaly scoring, the top five polling units identified as having the most significant irregularities are:

1. OGBARU MAIN MARKET (OKPOKO I) - Cluster: -1, Extremely high registration with low voter turnout
2. EKWELIGWE LAYOUT (OKPOKO I) - Cluster: -1, Geospatially isolated with inconsistent vote share
3. Fegge Central (ONITSHA SOUTH) - Cluster: 0, Disproportionate votes compared to historical trend
4. Awada Primary School II (Idemili North) - Cluster: 3, Anomalous voter swing between years
5. Nkpor Town Hall (Idemili North) - Cluster: 2, Unusual clustering and turnout levels

4. Hypotheses on Anomalies

- Inflated registration in some urban wards may point to ghost registrations or administrative errors.
- Low turnout in dense clusters may suggest voter apathy, disenfranchisement, or security-related suppression.
- Geospatial isolation of certain units suggests possible boundary misalignments or intentional placement to dilute influence.
- Statistical irregularities may stem from human errors, duplicate entries, or systemic biases during collation.

5. Recommendations to Election Authorities

- Audit flagged polling units for verification of voter rolls and physical location.
- Deploy GIS tools in future elections to dynamically monitor unit-level turnout in real-time.
- Enhance data granularity by publishing PU-level historical records for trend calibration.
- Conduct civic outreach in outlier zones to determine behavioral or infrastructural causes of anomalies.

6. Conclusion

This investigation presents a holistic, data-driven audit of voter behavior patterns in Anambra State elections. Using geospatial, statistical, and demographic techniques, it highlights the critical need for ongoing electoral data monitoring. Future analyses should expand on this framework by integrating INEC operational data and party-level mobilization efforts.