Historical Voting Behaviour and Geospatial Analysis of Electoral Trends in Osun State

Table of Contents

---

1. Introduction

Following widespread allegations of electoral irregularities in the recently concluded election in Osun State, Nigeria, this analysis aims to identify outlier polling units where voting patterns significantly deviate from expected norms or neighbouring units, potentially indicating manipulation or irregularities. Leveraging advanced geospatial clustering, statistical methods, machine learning validation, and demographic integration, this report presents a rigorous methodology, identifies the top 5 outlier polling units, and provides hypotheses and recommendations to ensure election integrity.

1.1 Objectives

- To identify and justify the top 5 outlier polling units.
- To utilize geospatial clustering and statistical methodologies to detect anomalies.
- To integrate demographic data for a comprehensive understanding of voting irregularities.
- To provide recommendations for election authorities based on findings.

2. Overview of the Data

The dataset consists of 2,249 rows and includes the following columns:

- State
- LGA (Local Government Area)
- Ward
- PU-Code (Polling Unit Code)
- PU-Name (Polling Unit Name)
- Accredited_Voters
- Registered_Voters
- Results_Found
- Transcription_Count
- Result_Sheet_Stamped
- Result_Sheet_Corrected
- Result_Sheet_Invalid
- Result_Sheet_Unclear

- Result_Sheet_Unsigned

- APC (All Progressives Congress Votes)

- LP (Labour Party Votes)

- PDP (People's Democratic Party Votes)

- NNPP (New Nigeria Peoples Party Votes)

- Results_File

Additionally, Latitude and Longitude were sourced to enable geospatial analysis. Sociodemographic data was obtained from the Nigerian Bureau of Statistics and census data, while voting history was gathered from internet research.

2.1 Data Validation

Prior to analysis, the following data validation steps were conducted:

- **Missing Values Check**: No missing values were found.

- **Duplicate Entries Check**: No duplicate records were detected.

- **Standardization**: Data formats were standardized for consistency.

2.2 Data Sourcing & Geocoding

To enhance the dataset, geographic coordinates were sourced using Google Cloud Console's Geocoding API. The API key was restricted to geocoding requests within Nigeria to ensure data security. Latitude and longitude were retrieved for each polling unit, enabling geospatial clustering and analysis.

3. Methodology

3.1 Enhanced Dataset Preparation

- Dataset: The analysis began with the `osun_geocoded.csv` dataset, containing polling unit data for Osun State, including vote counts for APC, PDP, LP, and NNPP, and geospatial coordinates (latitude and longitude).

- Tools: Python libraries (`pandas`, `geopandas`, `utm`) were used to load and preprocess the data.

3.2 Advanced Neighbor Identification (Geospatial Clustering)

- Method: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was applied to identify polling unit clusters based on geographic proximity. Coordinates were converted from latitude/longitude to UTM (Universal Transverse Mercator) for distance-based clustering. Sensitivity analysis was conducted with radii of 500m, 1km, and 2km.

- Implementation:

- Loaded `osun_geocoded.csv` and converted coordinates to UTM Easting and Northing.

- Applied DBSCAN with `eps=500m`, `eps=1km`, and `eps=2km` (min_samples=5), saving results to `osungeocoded_with_clusters.csv`.

- Findings:

- 500m Radius: Dense clusters in urban centers (Osogbo, Ilesa, Ife) indicated high accessibility, while rural areas showed sparse or no clusters, suggesting potential voter access challenges.

file:///C:/Users/user/osun_polling_units_map_500m.html

- 1km Radius: More clusters emerged in semi-urban and rural areas, with urban centers remaining highly connected. Border regions (e.g., near Oyo/Ondo States) lacked clusters, highlighting accessibility issues.

file:///C:/Users/user/osun_polling_units_map_1000m.html

- 2km Radius: Broad regional connectivity was observed, with large clusters in urban/semi-urban areas, though rural border areas remained underserved. The 500m radius suited hyper-local analysis, 1km practical walking distances, and 2km regional planning.

file:///C:/Users/user/osun_polling_units_map_2000m.html

3.3 Spatial Weights Calculation

- Method: Spatial weights were calculated using a 1km radius to define neighbor relationships for subsequent statistical analysis.

- Implementation:

- Converted the dataset to a GeoDataFrame using `geopandas`.

- Used `libpysal`'s `DistanceBand` (threshold=1000m, binary=True), set self-weights to 1.0 with `fill_diagonal`, and row-standardized weights (`w.transform='r'`).

- Saved results to `temp_gdf_with_weightsosun.csv`.

- Purpose: Established a foundation for spatial autocorrelation and hot spot analysis.

3.4 Local Moran's I (Spatial Autocorrelation)

- Method: Local Moran's I was calculated to identify localized spatial autocorrelation in vote counts for each party (APC, PDP, LP, NNPP).

- Implementation:

- Loaded `temp_gdf_with_weightsosun.csv` and reconstructed spatial weights.

- Computed Moran’s I using `esda.Moran_Local` with 99 permutations, adding `moran_I`, `moran_p`, and `moran_outlier` (p<0.05) columns for each party.

- Saved results to `temp_gdf_with_moran.csv`.

- Purpose: Detected polling units with vote patterns significantly different from their neighbors.

- Results: For the top 5 units (e.g., PU 29-11-05-002), LP showed negative Moran’s I (-0.017682), indicating dissimilarity from neighbors, while APC and PDP had near-zero values (e.g., -0.002183, -0.009761), suggesting subtle deviations.

- Insights: LP votes exhibited the strongest local anomalies, potentially signaling irregular patterns in urban clusters.

3.5 Getis-Ord Gi* (Hot Spot Analysis)

- Method: Getis-Ord Gi* was used to identify significant vote concentrations (hot spots) for each party.

- Implementation:

- Loaded `temp_gdf_with_moranosun.csv` and reconstructed 1km spatial weights.

- Applied `esda.G_Local` (star=True, permutations=99), adding `getis_gi`, `getis_p`, and `hot_spot` (p<0.05, Z>0) columns for each party.

- Saved results to `temp_gdf_with_getisosun.csv`.

- Purpose: Highlighted areas of unusually high vote concentrations.

- Results: Top units showed positive Gi* Z-scores (e.g., APC: 0.002873, PDP: 0.006986 at PU 29-11-05-002), with LP negative (-0.006088), indicating no hot spots but potential cold spots.

- Insights: APC and PDP had concentrated vote patterns in urban areas, while LP’s negative Gi* suggests isolated spikes.

3.6 Isolation Forest (Machine Learning Validation)

- Method: Isolation Forest was employed to cross-validate anomalies using vote counts and geospatial coordinates.

- Implementation:

- Loaded `temp_gdf_with_getisosun.csv`.

- Applied `sklearn.IsolationForest` (contamination=0.1, n_estimators=50) on features `[APC, PDP, LP, NNPP, Latitude, Longitude]`, adding an `iso_outlier` column (1 for outliers).

- Saved results to `temp_gdf_with_isolationosun.csv`.

- Purpose: Provided a non-spatial anomaly detection method to complement geospatial findings.

- Results: All top 5 units were flagged as outliers (`iso_outlier=1`), indicating significant deviation in vote and spatial data.

- Insights: Consistent outlier detection across urban units suggests robust anomalies beyond spatial clustering.

3.7 Composite Score Calculation

- Method: A composite score was calculated to integrate spatial and machine learning results, identifying overall outliers.

- Implementation:

- Loaded `temp_gdf_with_isolationosun.csv`.

- For each party, computed `composite_score = (abs(moran_I) + abs(getis_gi)) / 2`.

- Calculated `overall_composite_score = (mean(party_composite_scores) + iso_outlier) / 2`.

- Sorted by `overall_composite_score` and saved to `osun_state_outlier_scores1.csv`.

- Results Interpretation: The top 5 overall outliers had high composite scores (0.501933–0.502867), all flagged by Isolation Forest (`iso_outlier=1`), indicating significant deviation across methods. Party-specific scores showed LP with the highest anomalies (e.g., 0.011885–0.012020), followed by PDP, APC, and NNPP.

3.8 Temporal Comparative Analysis (Voting Trends)

- Method: Historical voting trends were visualized to assess changes over time.

- Implementation:

- Plotted vote counts for APC, PDP, LP, and NNPP from 2015, 2019, and 2023 using `matplotlib`.

- Data: APC (383603, 347634, 342941), PDP (249929, 337377, 353860), LP (0, 33, 23283), NNPP (0, 129, 713).

- Findings: APC votes declined slightly, PDP votes increased, LP surged in 2023, and NNPP showed modest growth, suggesting shifting voter preferences.

3.9 Demographic Integration (Socioeconomic Analysis)

- Method: Socioeconomic data (population, literate population, unemployment) was analyzed to contextualize anomalies.

- Implementation:

- Loaded data for 30 LGAs in Osun State.

- Identified top 5 LGAs by unemployment: Iwo (12,470), Ife East (12,294), Ife Central (10,898), Osogbo (10,135), Ife North (9,990).

- Identified bottom 5 LGAs by literate population: Ifedayo (44,108), Ila (72,999), Atakunmosa West (80,336), Boluwaduro (83,415), Egbedore (86,947).

- Findings: High unemployment in urban LGAs (e.g., Ife Central) and low literacy in rural LGAs (e.g., Ifedayo) may influence voting patterns.

4. Identification and Justification of Top 5 Outlier Polling Units

4.1 Summary of Top 5 Outliers

The top five outlier polling units, identified based on the highest composite scores, are all located in Ife Central LGA. Below is a summary table:

PU-Code	PU-Name	Score	APC Votes	LP Votes	PDP Votes	NNPP Votes
29-11-05-002	New Garage (Open Space Motor Park)	0.502867	57	53	24	0
29-11-05-008	Blessing Nursery and Pry. School, Faola Layout	0.502623	72	47	38	0
29-11-05-010	L.A. Pry. School, Ajebamidele	0.502389	135	54	101	2
29-11-05-001	New Garage (Open Space Motor Park)	0.501979	106	28	65	4
29-11-02-013	Open Space Ibukunolu Line I	0.501933	89	20	45	3

4.2 Justification of Top 5 Outlier Polling Units

1. PU-Code: 29-11-05-002, PU-Name: New Garage (Open Space Motor Park)

- Justification: High LP (0.011885) and PDP (0.008374) composite scores, flagged by Isolation Forest. DBSCAN (500m/1km) maps show it in a dense urban cluster, suggesting accessibility or manipulation potential. Historically, LP's 2023 surge (23,283 votes) contrasts with its 2015–2019 absence, raising irregularity flags. Located in an urban center with high unemployment (10,898), suggesting potential vote manipulation.

2. PU-Code: 29-11-05-008, PU-Name: Blessing Nursery and Pry. School, Faola Layout

- Justification: Elevated LP (0.011038) and PDP (0.007581) scores, consistent across methods. Gi* hot spot analysis and 1km cluster maps indicate vote concentration. Historical trends show PDP's steady rise (249929 to 353860), possibly amplified here. Proximity to urban clusters (500m/1km) and socioeconomic pressures may indicate irregularities.

3. PU-Code: 29-11-05-010, PU-Name: L.A. Pry. School, Ajebamidele

- Justification: Highest LP score (0.012020) and significant APC votes, flagged as an outlier. Moran's I and Gi* charts highlight spatial deviation and concentration. Ife Central's unemployment (10,898) and LP's 2023 spike suggest targeted influence.

4. PU-Code: 29-11-05-001, PU-Name: New Garage (Open Space Motor Park)

- Justification: Moderate LP (0.007936) and PDP (0.005964) scores, reinforced by Isolation Forest. Cluster maps (500m) and voting trend charts (APC decline: 383603 to 342941) suggest anomalies amid shifting preferences.

5. PU-Code: 29-11-02-013, PU-Name: Open Space Ibukunolu Line I

- Justification: LP (0.006439) and PDP (0.007173) anomalies, with urban clustering (1km maps). Historical LP growth and Ife Central's socioeconomic context (high unemployment) support its outlier status.

5. Hypotheses on Potential Reasons for Anomalies

1. Vote Buying or Coercion: High unemployment in Ife Central (10,898) may make voters susceptible to financial incentives, inflating votes (e.g., APC: 135, LP: 54 at PU 29-11-05-010).

2. Logistical Manipulation: Dense urban clusters (500m/1km) could facilitate ballot stuffing or result tampering due to concentrated polling units (e.g., New Garage PUs).

3**.** Emerging Party Influence and Socioeconomic Shifts: LP's sudden rise (from 0 to 23,283 votes by 2023) in urban areas like Ife Central may reflect targeted mobilization or irregularities. Additionally, the high unemployment rate in Ife Central (10,898 unemployed) could have driven voters to seek an alternative to traditional parties like PDP and APC, prompting them to support LP as a new option in hopes of addressing socioeconomic challenges.

4. Data Reporting Issue: The identical coordinates (7.55548, 4.531503) across the top 5 outlier polling units likely result from errors or limitations in the original dataset provided by INEC or another source, such as reused coordinates for co-located units (e.g., New Garage) or low-resolution mapping, rather than geocoding inaccuracies during analysis. In rare cases, identical coordinates could reflect intentional data tampering or a placeholder value used during reporting (e.g., if actual locations were unavailable or obscured). This aligns with the broader context of electoral irregularity allegations.

5. Socioeconomic Disparity: Urban-rural divides (e.g., low literacy in Ifedayo vs. high unemployment in Ife Central) may drive irregular voting patterns.

6. Power BI Dashboard Documentation

6.1 Page 1: Election Voting & Socioeconomic Analysis

Overview

The first page of the Power BI dashboard provides a high-level summary of election voting and socioeconomic metrics for Osun State, alongside geospatial visualizations of polling unit clusters and outliers. The dashboard is designed to support the identification of anomalies by integrating voting data, socioeconomic factors, and geospatial analysis.

Metrics Observed

1. Total Registered Voters:

- Value: 1 million (1M)

- Interpretation: Represents the total number of registered voters in Osun State, serving as a baseline for voter turnout analysis.

2. Total Accredited Voters:

- Value: 498,000 (498K)

- Interpretation: Indicates a 49.8% turnout rate, providing a benchmark for assessing vote counts.

3. Total Population:

- Value: 4 million (4M)

- Interpretation: Shows a 25% voter registration rate (1M out of 4M), useful for evaluating registration distribution.

4. Literacy Rate (%):

- Value: 90.57%

- Interpretation: Suggests most voters can engage with election materials, though disparities exist in certain LGAs.

5. Unemployment Rate (%):

- Value: 5.02%

- Interpretation: Indicates socioeconomic pressures that may influence voting behavior, particularly in high-unemployment LGAs like Ife Central (10,898 unemployed).

Visualizations

1. Clusters by Polling Unit (Map):

- Description: A map with orange dots representing polling units, using a 1km radius (Distance filter: 1km).

- Observation: Dense clustering in urban areas (e.g., Ife Central, Osogbo, Ilesa) contrasts with sparse distribution near borders (e.g., Oyo/Ondo), highlighting accessibility disparities.

2. Clusters by Political Party (Map):

- Description: A map visualizing vote clusters by party (APC, PDP, LP, NNPP), with the "Outlier Type" filter set to "APC Getis-Ord Gi."

- Observation: APC vote concentrations are prominent in urban centers, indicating potential hot spots for vote anomalies.

3. Top Outliers by Composite Score (Map):

- Description: A map highlighting top outliers, with the "PU-Name" filter set to "New Garage (Open Space Motor Park)" and "Open Space Ibukunolu Line I."

- Observation: Outliers are tightly clustered in Ife Central, confirming the urban focus of anomalies.

Filters Applied

- Distance

- PU-Name

- Outlier Type

Insights

- Urban clustering in Ife Central aligns with high voter turnout and potential anomalies.

- Socioeconomic metrics (5.02% unemployment, 90.57% literacy) provide context for voter behavior influences.

- The 49.8% turnout rate offers a benchmark for evaluating vote counts against accredited voters.

6.2 Page 2: Election Voting & Socioeconomic Analysis

Overview

The second page provides a detailed breakdown of socioeconomic and voting metrics across Osun State, focusing on unemployment, polling unit distribution, voting trends by political party, and literacy rates by LGA. This page enhances the understanding of demographic factors influencing voting patterns and anomalies.

Metrics Observed

1. Total LGA:

- Value: 30

- Interpretation: Confirms coverage of all 30 Local Government Areas in Osun State.

2. Total Unemployed Population by LGA:

- Value: 212,700 (212K)

- Interpretation: Aligns with the 5.02% unemployment rate (212,700 out of 4M population). Top 5 LGAs: Iwo (12,470), Ife East (12,294), Ife Central (10,898), Osogbo (10,135), Ife North (9,990).

3. Total Polling Units by LGA:

- Value: 15,000 (15K)

- Interpretation: Provides context for the density and distribution of voting locations.

4. Total PDP Votes by Year:

- Value: 221,000 (221K) in 2023 (2015: 249,929; 2019: 337,377; 2023: 353,860)

- Interpretation: Shows PDP's steady increase, influencing voting patterns in urban areas.

5. Total APC Votes by Year:

- Value: 219,000 (219K) in 2023 (2015: 383,603; 2019: 347,634; 2023: 342,941)

- Interpretation: Indicates APC's slight decline, providing context for vote concentration anomalies.

6. Literacy Rate (%):

- Value: 90.57%

- Interpretation: High literacy rate, but disparities exist in certain LGAs, affecting voter engagement.

Visualizations

1. Top 5 Unemployed Population by LGA (Bar Chart):

- Description: A bar chart showing the top 5 LGAs by unemployed population: Iwo, Ife East, Ife Central, Osogbo, and Ife North.

- Observation: Ife Central (10,898 unemployed) is a hotspot, correlating with the location of top outliers.

2. Total PDP and APC Votes by Year (Line Chart):

- Description: A line chart comparing PDP, APC, LP, and NNPP votes from 2015 to 2023.

- Observation: PDP votes increased from 249,929 to 353,860, while APC votes decreased from 383,603 to 342,941. LP (23,283 in 2023) and NNPP (713 in 2023) trends align with historical data.

3. Top 5 Literate Population by LGA (Bar Chart):

- Description: A bar chart showing the top 5 LGAs by literate population: Olorunda, Obokun, Ilesha West, Ilesha East, and Irepodun.

- Observation: Contrasts with the bottom 5 (e.g., Ifedayo: 44,108), highlighting literacy disparities.

Filters Applied

- LGA: Options include Atakunmosa East, Atakunmosa West, Aiyedaade, and Aiyedire, but no specific filter is applied.

- Year: Slider set from 2015 to 2023, covering the full historical range.

Insights

- High unemployment in Ife Central (10,898) aligns with outlier locations, suggesting socioeconomic influences on voting anomalies.

- Voting trends (PDP growth, APC decline) provide context for party-specific anomalies in urban areas.

- Literacy disparities highlight varying voter engagement levels, with rural LGAs potentially more vulnerable to irregularities.

---

7. Conclusion

This study provides a comprehensive analysis of historical voting behavior in Osun State, leveraging geospatial clustering, voting pattern evaluation, and sociodemographic correlations. Through the use of DBSCAN clustering, spatial-temporal anomalies were detected, shedding light on potential irregularities in voter turnout and result patterns across different polling units. The integration of socioeconomic and demographic data from official sources further contextualized these findings, allowing for a deeper understanding of how factors such as literacy rates, employment levels, and population density influence electoral outcomes. The development of an interactive Power BI dashboard enhances accessibility to these insights, enabling stakeholders to explore key trends, detect anomalies, and compare political party performances across different locations and election cycles. This analytical approach not only aids in identifying potential areas of concern in the electoral process but also serves as a valuable tool for policymakers, election monitors, and researchers aiming to improve electoral transparency and integrity in Osun State. Future research can extend this work by incorporating more granular demographic data, refining clustering techniques, and applying machine learning models to predict voting behavior. Ultimately, this analysis contributes to a data-driven understanding of electoral dynamics, supporting informed decision-making in governance and election monitoring.