

Outlier Detection in Election Data Using Geospatial Analysis

Case Study: Ensuring Election Integrity

In the latest election, the Independent National Electoral Commission (INEC) faced numerous legal challenges questioning the integrity and accuracy of the results. Allegations of vote manipulation and irregularities were widespread, necessitating a thorough investigation into these claims.

Objective

The goal was to help uncover potential voting irregularities and ensure the election results were transparent. This was achieved by identifying polling units where the voting results significantly differed from those of neighboring units, which could indicate possible influences or manipulations.

Approach

1. **Data Preparation:** Cleaned and preprocessed the election dataset, converting necessary columns and handling duplicates.
2. **Geospatial Analysis:** Used geospatial techniques to examine voting patterns and detect anomalies in polling units.
3. **Outlier Detection:** Applied statistical methods to identify polling units with voting results that deviated significantly from neighboring units.
4. **Visualization:** Highlighted and analyzed significant outliers to ensure transparency and address potential irregularities.

Methodology

Stage 1: Data Preparation

I used Geocode by Awesome Table on Google Sheets to obtain the longitude and latitude of the polling units. This geospatial data is essential for analyzing the geographic distribution of polling units and identifying potential anomalies in voting results.

Stage 2: Neighbor Identification

I loaded the CSV file into Python (Jupyter Notebook) and used the geodesic method to identify neighboring polling units based on geographical proximity. This step involved grouping neighboring polling units into lists represented by their PU codes, providing a structured way to analyze proximity-based relationships.

Stage 3: Outlier Detection

I employed the Interquartile Range (IQR) proximity rule to identify outliers for each polling unit by comparing it with its neighboring units. This statistical method helped pinpoint polling units where voting results deviated significantly from the norm, highlighting potential anomalies in the data.

Stage 4: Analysis and Visualization

I sorted the data to identify the top five polling units with the most significant outliers for each party. A scatter plot was used to visualize these outliers, providing a clear representation of the most anomalous results. The analysis highlighted the following significant outliers:

No APC LP PDP NNPP

1	562	521	392	60
2	518	508	272	54
3	500	499	243	44
4	482	490	239	31
5	453	460	224	28

Stage 5: Highlighting Top Outliers

For the top three outliers and their polling units from the different parties, the outliers are identified based on their deviation from the norm, either as exceptionally high or low values compared to neighboring polling units:

APC (All Progressives Congress)

1. PU-Code: 32-01-07-003, Outlier Value: 562.0

- **Reason:** This polling unit's vote count of 562 is significantly higher compared to its closest polling units. The discrepancy suggests this unit might have experienced an unusual voting pattern or error, warranting further investigation.

2. PU-Code: 32-06-11-020, Outlier Value: 518.0

- **Reason:** With 518 votes, this polling unit stands out because the vote count is unusually high compared to its neighbors. This significant deviation raises questions about the accuracy of the reported votes or possible anomalies.

3. PU-Code: 32-21-10-007, Outlier Value: 500.0

- **Reason:** This polling unit's vote count of 500 is considerably higher than the majority of surrounding units. The high vote count suggests it may be an outlier, potentially due to reporting errors or discrepancies.

LP (Labour Party)

1. PU-Code: 32-15-15-023, Outlier Value: 521.0

- **Reason:** This unit's 521 votes are notably high relative to nearby units. Such a deviation could indicate reporting errors, unusual voter turnout, or other anomalies.

2. PU-Code: 32-15-02-001, Outlier Value: 508.0

- **Reason:** A vote count of 508 is unusually high compared to surrounding polling units. This significant difference might be due to errors in data entry or an irregular voting pattern.

3. PU-Code: 32-11-03-008, Outlier Value: 499.0

- **Reason:** With 499 votes, this polling unit's count is abnormally high compared to its neighbors. This could suggest a reporting error or an unusual increase in votes.

PDP (People's Democratic Party)

1. PU-Code: 32-05-07-003, Outlier Value: 392.0

- **Reason:** This unit's vote count of 392 is unusually high compared to surrounding units. Such a high count may indicate data entry issues or an irregular voting pattern.

2. PU-Code: 32-04-08-005, Outlier Value: 272.0

- **Reason:** A vote count of 272 is quite high compared to neighboring units. This significant difference can suggest potential errors in data reporting or an abnormal voting trend.

3. PU-Code: 32-15-06-025, Outlier Value: 243.0

- **Reason:** The 243 votes in this unit are notably higher than vote counts in surrounding units, indicating possible data anomalies or unusual voting behavior.

NNPP (New Nigeria Peoples Party)

1. PU-Code: 32-21-10-007, Outlier Value: 60.0

- **Reason:** With only 60 votes, this unit's count is significantly lower compared to its nearby units. Such a low value could indicate missing data, reporting errors, or unusual voter turnout.

2. PU-Code: 32-03-12-007, Outlier Value: 54.0

- **Reason:** This polling unit's count of 54 is unusually low compared to surrounding units, which could suggest data reporting issues or an unusual pattern of voter turnout.

3. PU-Code: 32-03-07-005, Outlier Value: 44.0

- **Reason:** A vote count of 44 is significantly lower than nearby polling units. This low count might indicate potential issues with data accuracy or unusual voting patterns.

Conclusion

This analysis provides a systematic approach to identifying potential voting irregularities by leveraging geospatial data and statistical methods. By pinpointing outlier polling units, it highlights areas where further investigation is warranted to ensure the integrity and accuracy of election results. This approach not only aids in uncovering potential anomalies but also enhances transparency in the electoral process.