Outlier Detection in Kwara State Election Data Using Geospatial and Geocoding Analysis

Muhammad Rofiat Opemipo

Date: 27th of October 2025.

Required Hosted Sheet:

1. Hosted csv file with the latitude and longitude with party outliers' score - HERE
2. Hosted excel file (with different sheet representing each party in a workbook) showing sorted polling units by outliers' score - HERE

1. Methodology

1.1 Geocoding

The first step involved the process known as Geocoding: converting the names of polling units into geographical coordinates (latitude and longitude).

This was achieved with the Awesome table geocoder add-ins in Google sheet (implementing Google API) by merging the (Local Government Address, Ward Address and the Polling Unit Name) to form Full Address column for each polling unit. Each polling unit's full address was passed to the geocoder to retrieve its exact coordinates. Then the latitude and longitude were added to the dataset.

This step was essential to enable geospatial analysis of polling units and their *proximity detection*.

1.2 Neighbor Identification

Identifying neighbor polling units with coordinates obtained from Google sheet was achievable by loading the new Dataset into python to check for the neighbor polling Units.

For each polling unit, neighboring units were defined as those located within a **5km radius because 1km to 4km were not giving comparable polling unit because they were too close**.

The process was achieved by building BallTree using Haversine distance in python and it created a neighborhood structure that helped compare each polling unit to nearby units with similar geographical contexts.

1.3 Outlier Score Calculation

Since neighboring polling units were identified, an outlier score was calculated for each political party. The score represents how much a polling unit's vote count for a party deviates from the average votes of its neighboring units. Higher scores indicate that a polling unit's results differ significantly from nearby units, which could suggest possible irregularities or data anomalies.

2. Findings

2.1 Top 3 Outlier Polling Units

After calculating the outlier scores, the polling units were sorted in descending order based on their scores. The top 3 outlier polling units showed significant deviations from their neighbors.

Party	Polling Unit Name	Outlier Score	Latitude	Longitude
NNPP	OPEN SPACE ODE OPELE	133.7564162	8.4346609	4.665703
APC	WAGURU MARKET	99.70205615	9.0393189	3.62178
PDP	MOLEKE VILLAGE	47.37615434	8.8309531	4.541814

2.2 Possible Reasons for Anomalies

The observed outliers could be due to:

- Unusual voters turn out compared to neighboring units.
- Data entry errors or duplication in vote recording.
- Vote manipulation or external influence in specific locations.
- Differences in population density or polling unit size.

This needs more in-depth investigation to rule out all possibilities.

3. Visualizations and Insights

3.1 Geospatial Visualization of Outliers

The four interactive maps were created to visualize the outlier polling units for each major political party APC, LP, PDP, and NNPP.

- The red marker with a high deviation score represents the polling unit identified as the top 1 outlier.
- The red marker without high deviation score represents the polling unit identified as the remaining top 2 outliers out of Top 3 Outliers.
- The blue markers indicate nearby top 10 polling units within 5 km radius used for comparison.

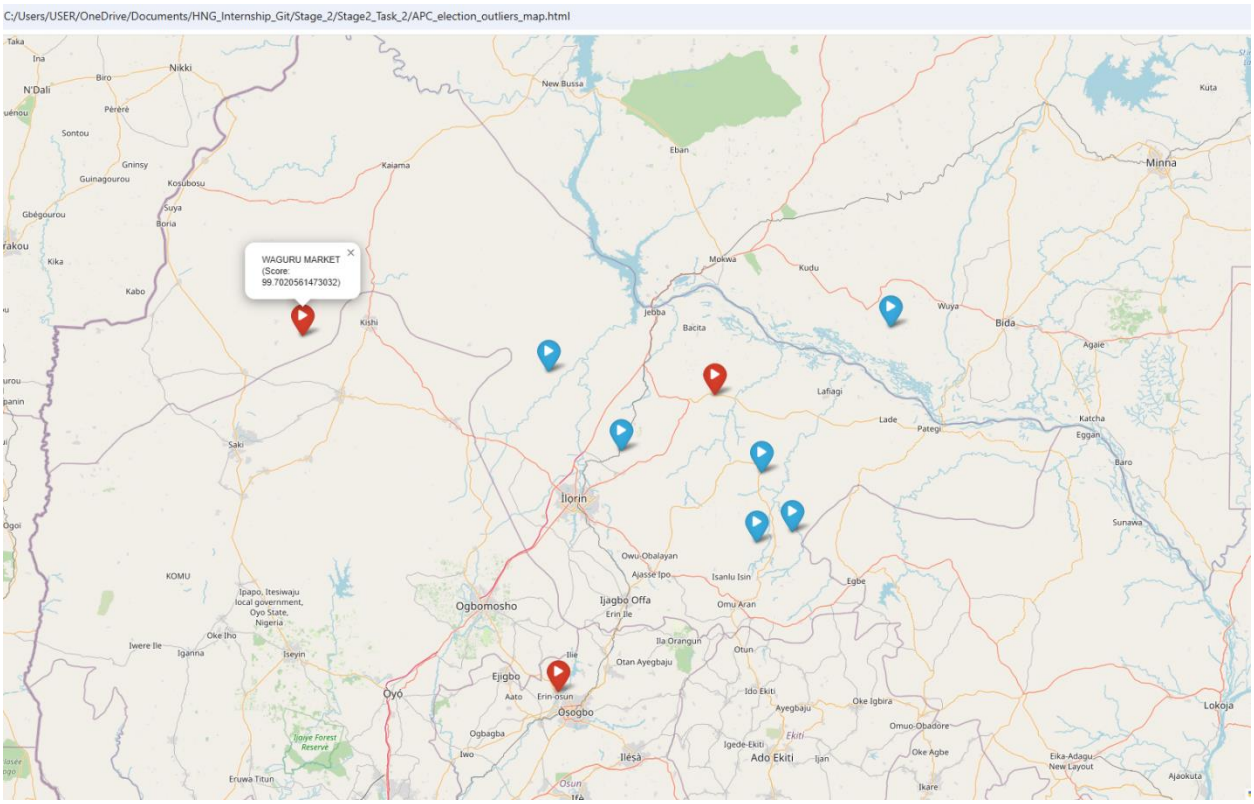

Fig 1: APC Outlier Map

Observation:

- The polling unit *Waguru Market* in Kwara State recorded a significantly higher deviation score (99.72) compared to its surrounding units.
- Most of the APC outliers are concentrated around Kwara North, close to the boundary between Kwara and Niger States.
- This could suggest unusual vote patterns or potential irregularities in these areas.

Insight:

The APC map shows localized spikes in vote differences that may point to data inconsistencies or extremely high turnout in select polling units compared to neighboring ones.

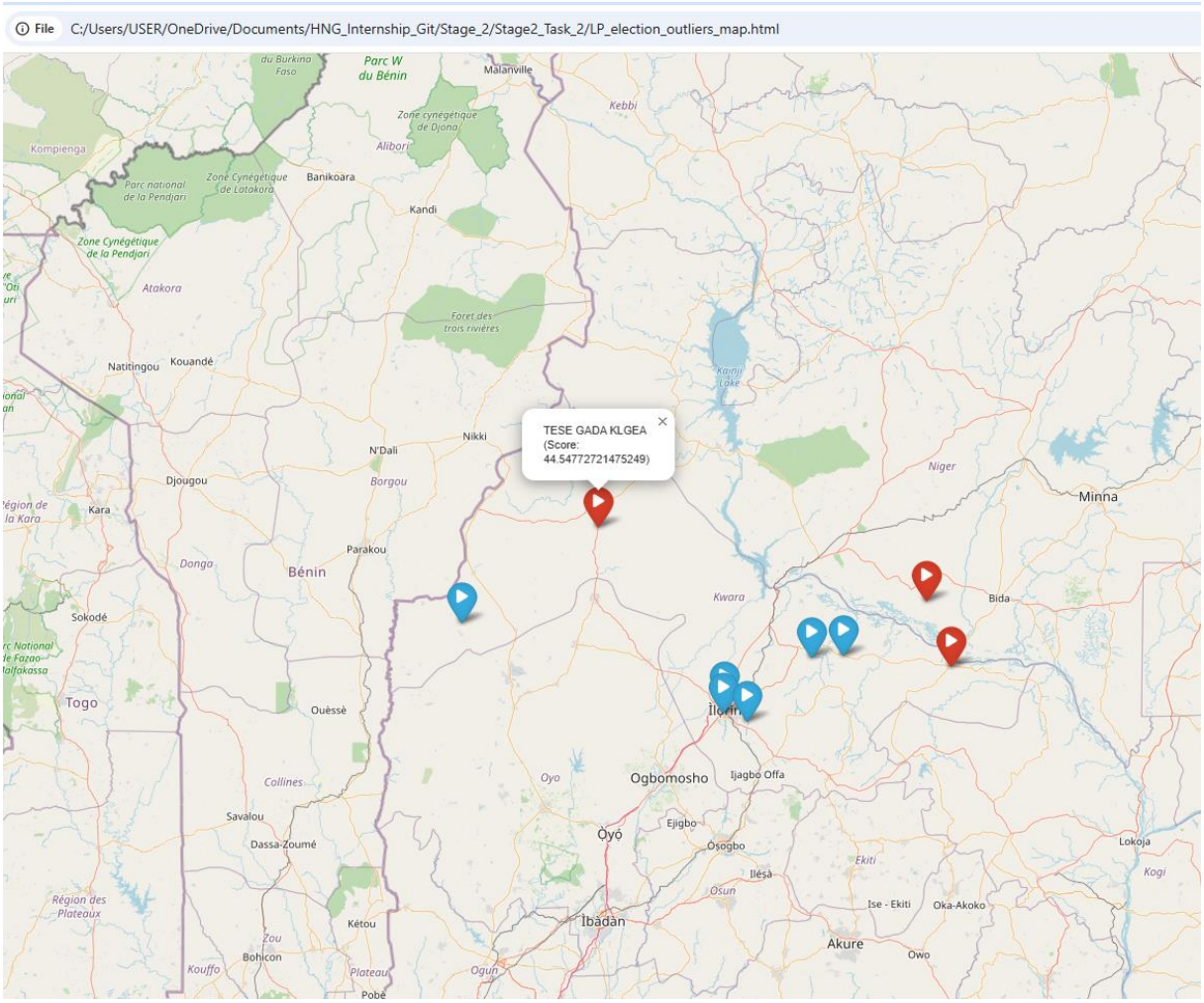

Fig 2: LP Outlier Map

Observation:

- *Tese Gada Klgea* is the most significant outlier with an outlier score of 44.55, located in the northwestern region of Kwara State.
- LP outliers are more sparsely distributed across the map, with fewer clusters.
- The spread suggests that LP's vote anomalies may not be concentrated but rather isolated incidents.

Insight:

For the LP, the scattered nature of outliers might indicate data entry errors or isolated polling units where vote counts differ sharply from neighbors, rather than region-wide irregularities.

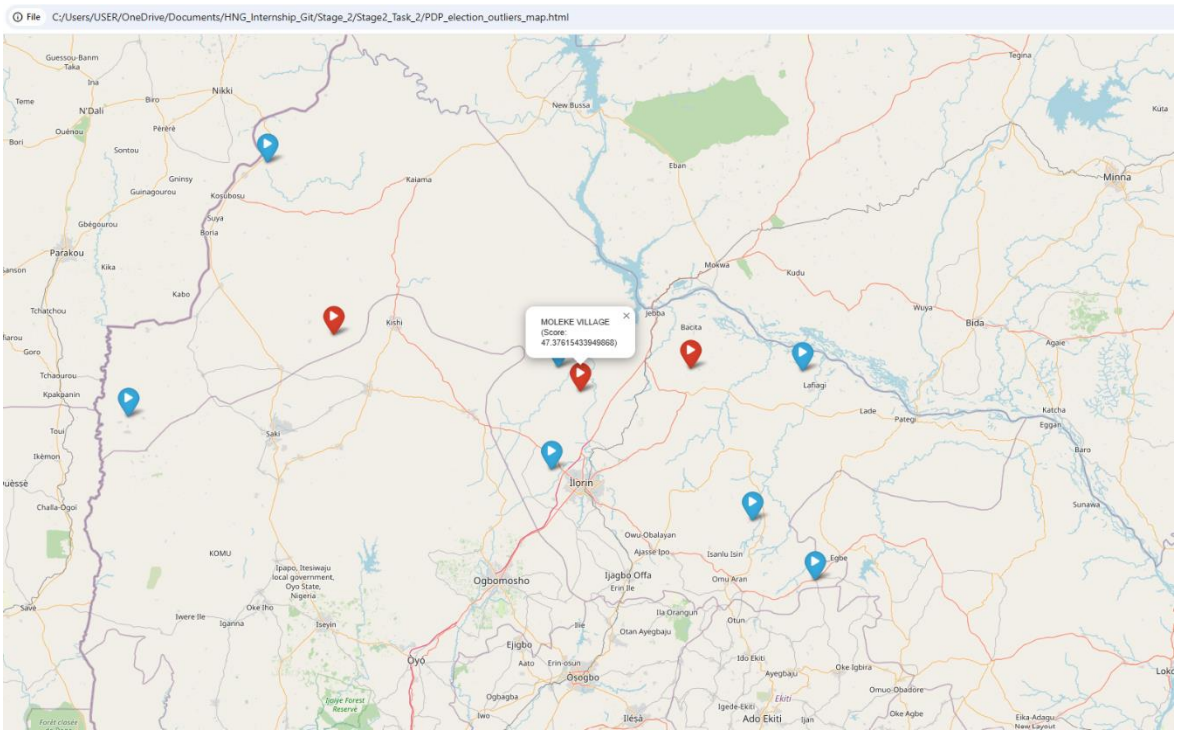

Fig 3: PDP Outlier Map

Observation:

- The polling unit in Moleke Village recorded the highest PDP deviation score (47.37).
- PDP outliers are mainly in central Kwara, near Ilorin and Kishi, with a few closes to the state borders.
- These points are surrounded by polling units with much lower vote counts, highlighting strong local deviations.

Insight:

The clustering of PDP outliers in central locations could point to voter concentration effects or local administrative inconsistencies rather than broad manipulation.

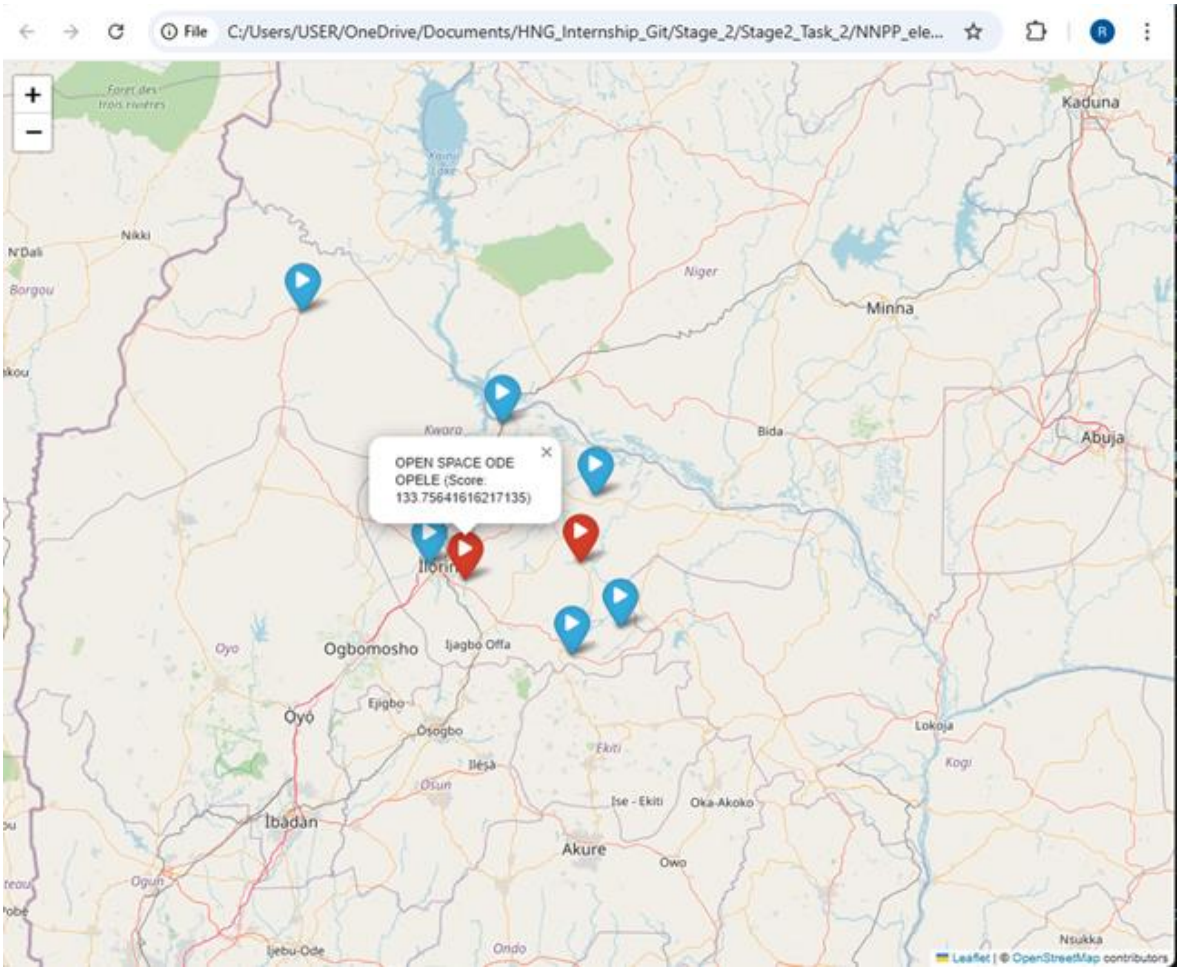

Fig 4: NNPP Outlier Map

Observation:

- The polling unit in the *Open Space Ode Opele* in Kwara State recorded a notably high score of 133.76, standing out compared to nearby units.
- Most of the NNPP outliers are clustered around Ilorin and its surrounding areas, with fewer spread toward the southern and eastern edges of the state.
- This spatial pattern suggests concentrated areas of strong NNPP activity or data variation near urban centers.

Insight:

The NNPP map highlights pockets of elevated scores around Ilorin, indicating localized areas of heightened party presence or reporting anomalies. These clusters may reflect higher engagement levels, data concentration, or potential inconsistencies requiring further verification.

4. Conclusion and Recommendations

This analysis used geospatial techniques to detect outliers in election data.

By comparing vote patterns within small geographical clusters, polling units with unusual voting behaviours were identified. The key Insights from the analysis are as follows:

- A few polling units displayed results that differ sharply from nearby locations.
- These deviations could indicate potential data anomalies or election irregularities.
- Geospatial analysis is a valuable tool for transparency and election integrity monitoring.

For more transparency, INEC and other election regulators should review the identified outlier polling units to confirm data accuracy and investigate possible causes behind the irregularities.