Outlier Detection Report: Geospatial Analysis and Outlier Detection in Election Data

Introduction

The objective of this analysis is to identify potential voting irregularities in the recently concluded election by detecting outliers in the voting data. The analysis focuses on polling units where the voting results deviate significantly from neighbouring units, indicating potential influences or rigging.

Methodology

* Data Collection

The dataset provided contains votes for multiple parties across various polling units. The dataset was augmented with geospatial data (latitude and longitude) for each polling unit.

* Geocoding

Geocoding was performed to obtain the latitude and longitude values for each polling unit using the `Awesome Table` plugin in Google Sheets. This plugin allows for easy conversion of addresses into geographical coordinates.

1. Loading the Dataset: The dataset was loaded and inspected for missing geospatial data.

2. Geocoding with Awesome Table: The `Awesome Table` plugin in Google Sheets was used to geocode the addresses of polling units to obtain their latitude and longitude.

3. Validation: The geocoded data was validated to ensure accuracy.

* Neighbour Identification

Neighbouring polling units were identified based on geographical proximity. A radius of 1 km was defined to determine which units are considered neighbours.

* Outlier Detection

Z-scores was used for outlier detection.  
Calculated Z-scores for each polling unit's votes by comparing them to the votes of neighbouring units. The Z-scores help identify polling units with significantly different voting patterns from their neighbours, which can be indicative of potential irregularities.

Results

The output file `/mnt/data/ANAMBRA\_crosschecked\_with\_iqr\_outliers.xlsx` contains the original data with additional columns indicating the outlier values for each party. These columns show the actual vote counts for outliers and `NaN` for non-outliers.

Conclusion

The methodology employed geospatial analysis and robust statistical techniques to identify outliers in election data. By comparing the votes each party received with those of its neighbouring units, potential voting irregularities were detected, providing valuable insights for further investigation.

Recommendations

1. Further Investigation: Polling units identified as outliers should be investigated further to determine the cause of the discrepancies.

2. Improved Data Collection: Ensuring accurate geospatial data for all polling units can improve the robustness of the analysis.

3. Continuous Monitoring: Implementing continuous monitoring and analysis during elections can help in early detection of irregularities.