HNG 11

**Outlier Detection in Election Data Using Geospatial Analysis**

OTUENE HEPHZIBAH REGINALD

horizontal line

# 

# 

# INTRODUCTION

In the aftermath of a contentious election, concerns over the accuracy and integrity of voting results have surfaced, prompting a rigorous investigation by the Independent National Electoral Commission (INEC). Allegations of vote manipulation and irregularities have underscored the need for a comprehensive analysis to uncover potential anomalies and ensure transparency in electoral outcomes. This report focuses on employing geospatial analysis techniques to identify outlier polling units where voting patterns deviate significantly from their neighboring units.

# METHODOLOGY

I used geospatial analysis to detect outliers in polling unit data. The process involved the following steps:

1. Data Collection and Cleaning: - Polling unit data, including latitude and longitude coordinates, were sourced from a provided dataset. Geocoding was conducted using Google Sheets via Awesome Table to augment the dataset with accurate geospatial information.

- The dataset was meticulously cleaned to ensure completeness and accuracy, eliminating any inconsistencies that could affect subsequent analyses.

2. Outlier Detection Using Python on Jupyter Notebook: - The Isolation Forest algorithm was applied to detect anomalies in voting patterns exhibited by each political party across polling units.

- Outlier scores were computed for each polling unit, indicating the degree of deviation from the expected voting behavior based on its features.

3. Geospatial Analysis: - Distances between polling units were calculated to assess spatial relationships and identify neighboring units.

- The top three outliers were determined for each political party based on their outlier scores, providing insights into areas with potentially irregular voting behaviors.

- Closest neighboring polling units to each outlier were identified to contextualize the anomalous voting patterns geospatially.

## SUMMARY

The analysis uncovered substantial outliers within polling units associated with each political party. The sorted list of polling units by outlier scores for each party can be found in the accompanying Excel sheet, facilitating a deeper understanding of the electoral landscape.

- [cleaned\_polling\_units](https://docs.google.com/spreadsheets/d/1EmhdXzVcyaX3KRu5zFQbjxTKeA4MSQ2cdHEgVRt6NG0/edit?usp=sharing): Contains the cleaned dataset with latitude, longitude, and outlier scores for each polling unit.

- [sorted\_polling\_units\_by\_outlier\_scores](https://docs.google.com/spreadsheets/d/1xOwv9jdTxms5OkuoKyOkwLOxyUlXqearP1P_aGE2DjE/edit?usp=sharing): Excel sheet presenting polling units sorted by outlier scores for APC, LP, PDP, and NNPP.

## Conclusion

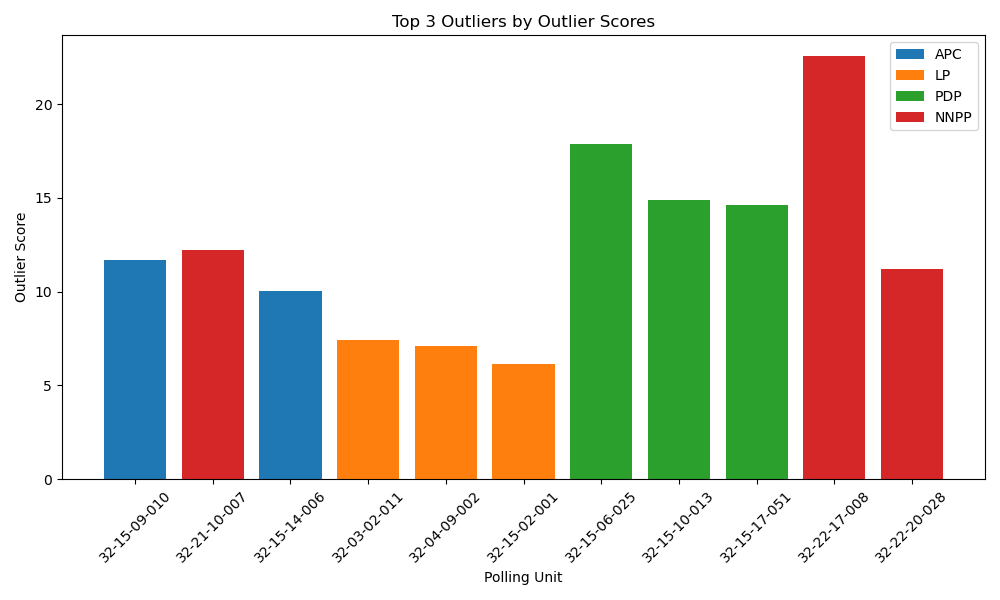
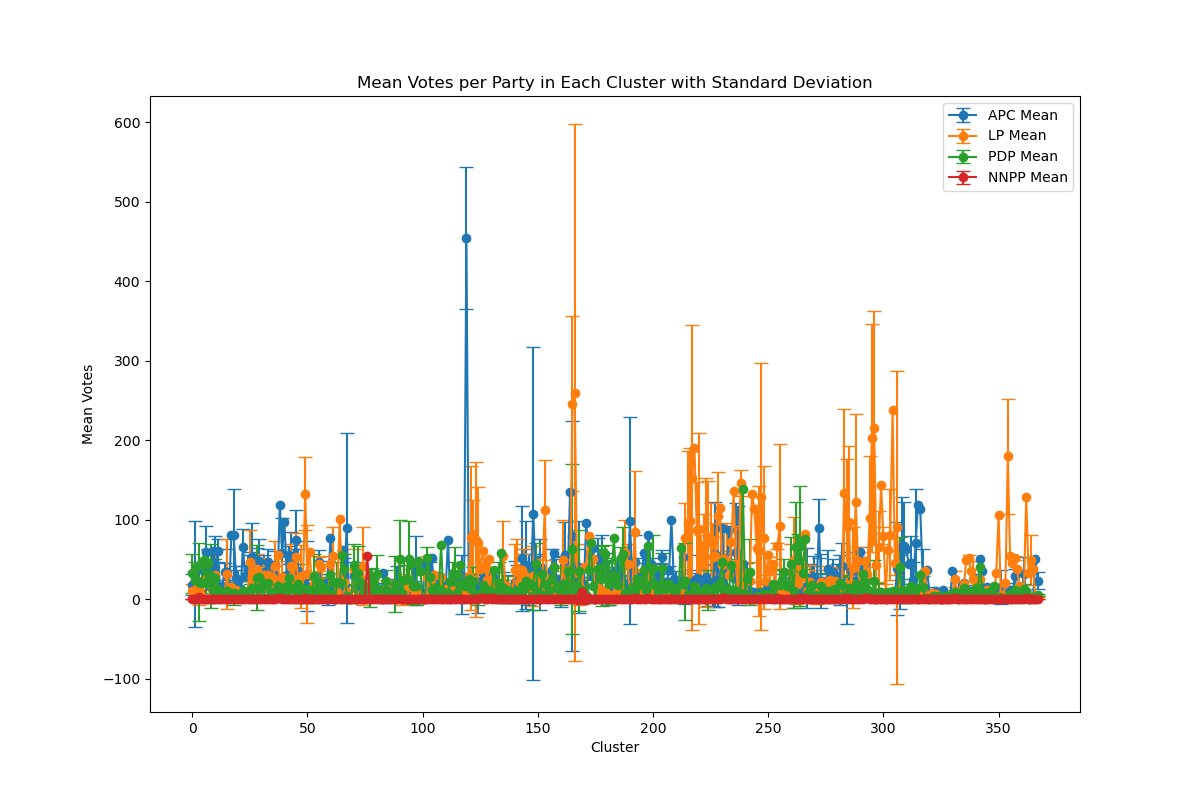
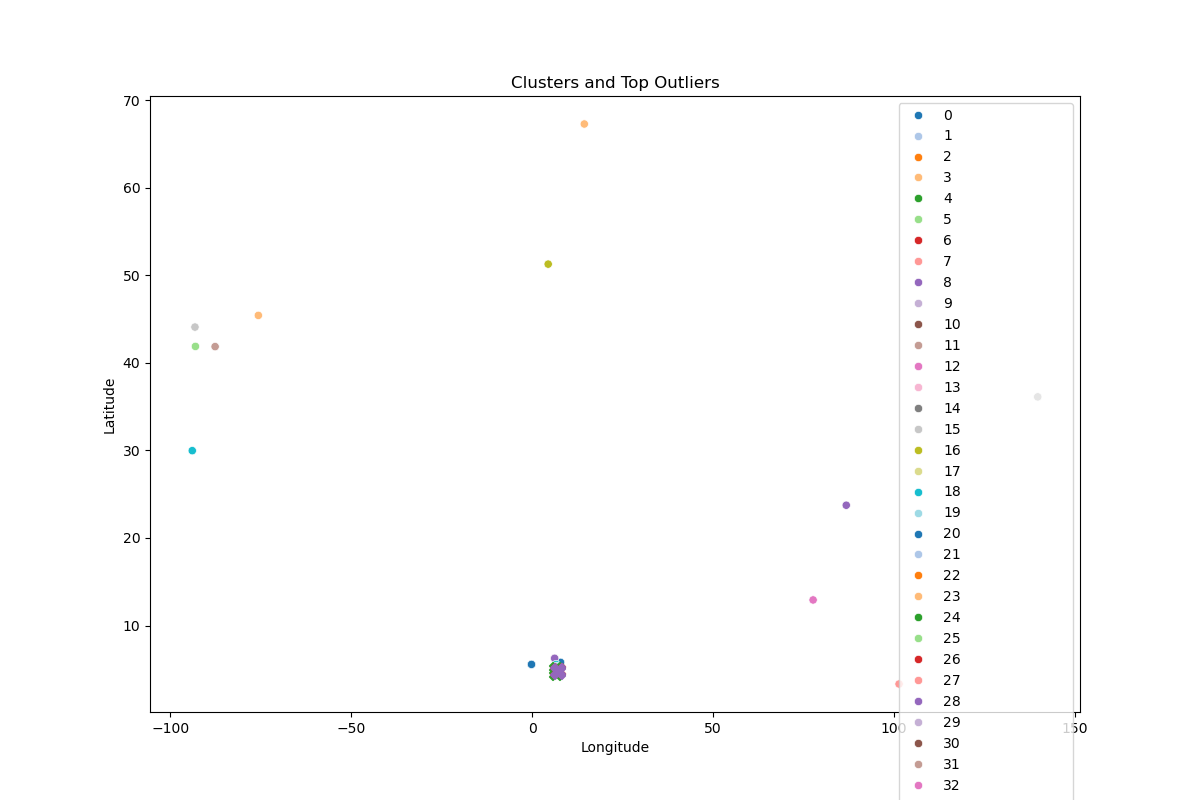
The geospatial analysis and outlier detection provided valuable insights into potential irregularities within the election data:

1. Identification of Significant Outliers: Distinct polling units were identified where voting behaviors significantly deviated from their neighboring units, indicating potential irregularities.

2. Analysis of Top Outliers: The top outliers exhibited voting patterns that were markedly different from their surroundings, highlighting areas warranting further scrutiny and investigation.

3. Spatial Context: Geospatial visualization and proximity analysis elucidated the spatial distribution of outliers, offering a clear depiction of areas with anomalous voting behaviors.

The visualizations included in this report help illustrate these findings and provide a clear understanding of the spatial distribution of outliers to contribute to the transparency and integrity of electoral processes.



This comprehensive approach underscores the importance of geospatial analysis in uncovering potential electoral irregularities and ensuring fair and transparent election outcomes.

For additional insights and learning opportunities, check out [HNG Internship](https://hng.tech/internship) and [HNG Premium](https://hng.tech/premium)

You can access the dataset and further details from [my github](https://heptuene123098.github.io/Rivers-State-Election/)