**OUTLIER DETECTION IN ELECTION DATA USING GEOSPATIAL ANALYSIS**

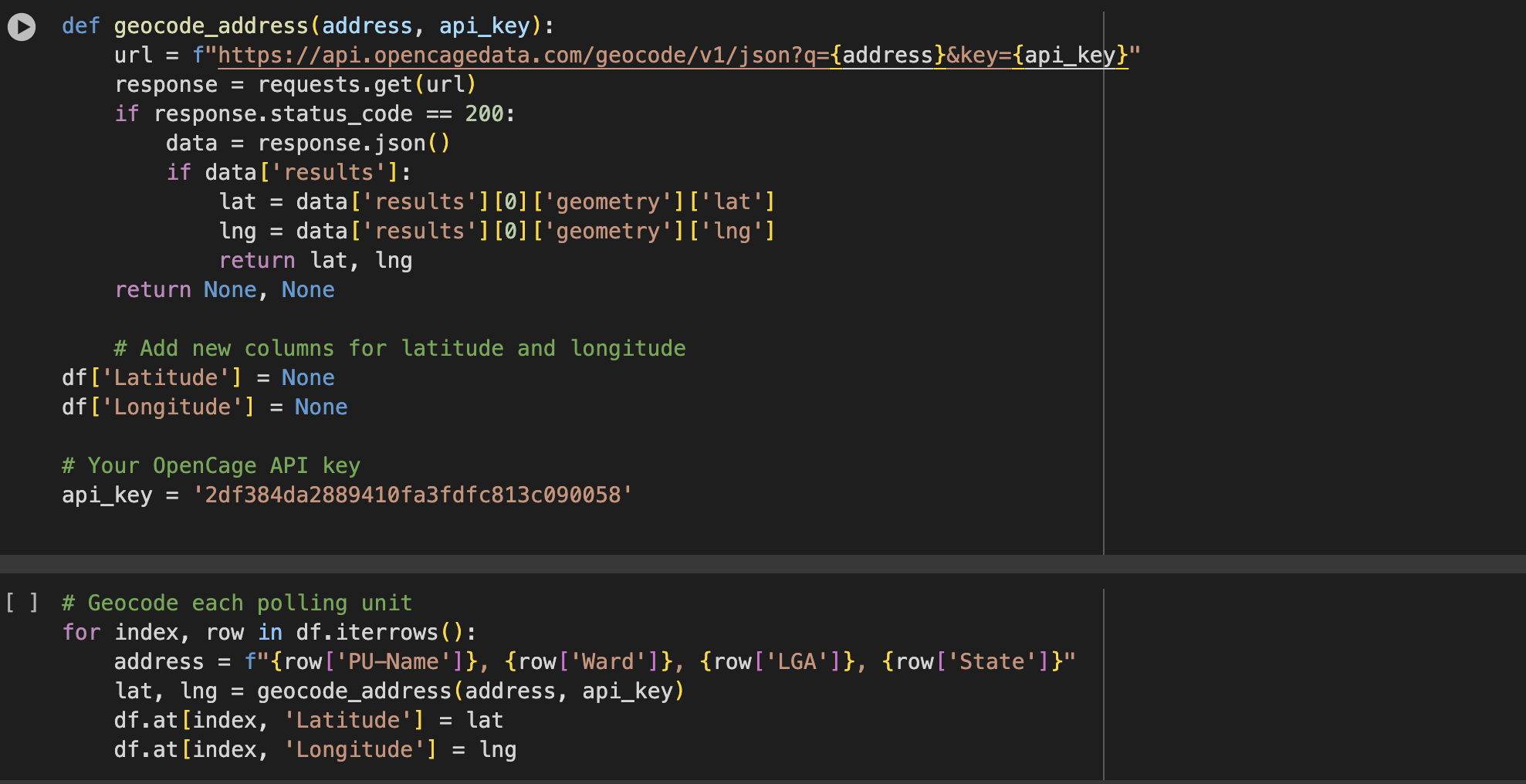
**INTRODUCTION**

The objective of this analysis is to identify potential voting irregularities in the 2024 Osun State concluded election. The Independent National Electoral Commission (INEC) has faced allegations of vote manipulation and irregularities, prompting a thorough investigation. By using geospatial techniques, we aim to uncover outlier polling units where the voting results deviate significantly from neighboring units, indicating potential influences or rigging.

**METHODOLOGY**

**Geocoding techniques**

For polling units without longitude and latitude values, I employed geocoding techniques using the OpenCageData API to obtain the necessary geospatial data.



**Neighbor identification**

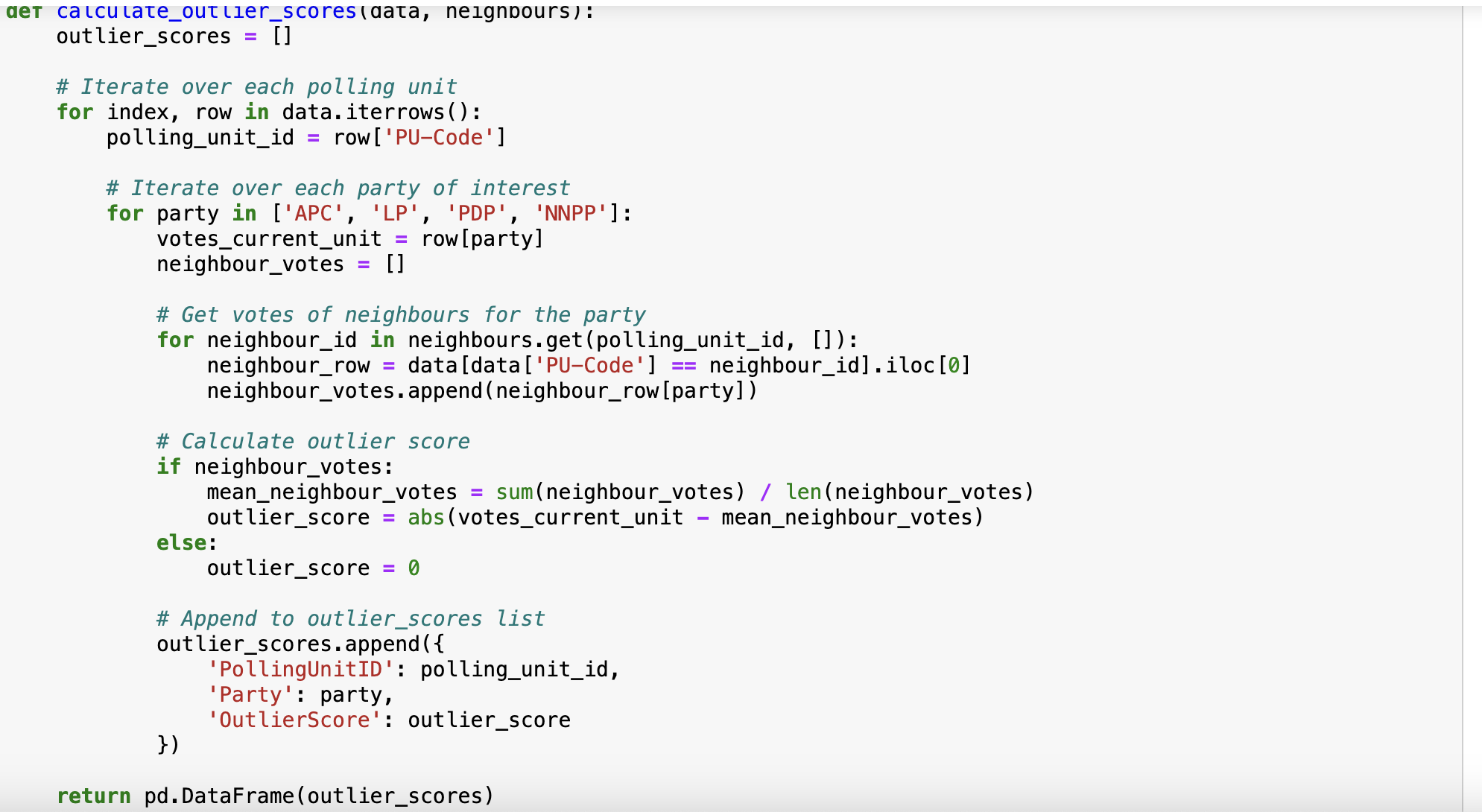
To identify neighboring polling units, I used geographical proximity. The latitude and longitude of each polling unit were utilized to calculate distances. Polling units within a 1km radius were considered neighbors



**Outlier score calculation**

For each polling unit, I compared the votes each party received with those of its neighboring units. The outlier score for each party was calculated based on the deviation of votes from neighboring units. The formula used for the outlier score is

Outlier score = |votes in current unit - mean votes of neighbors|

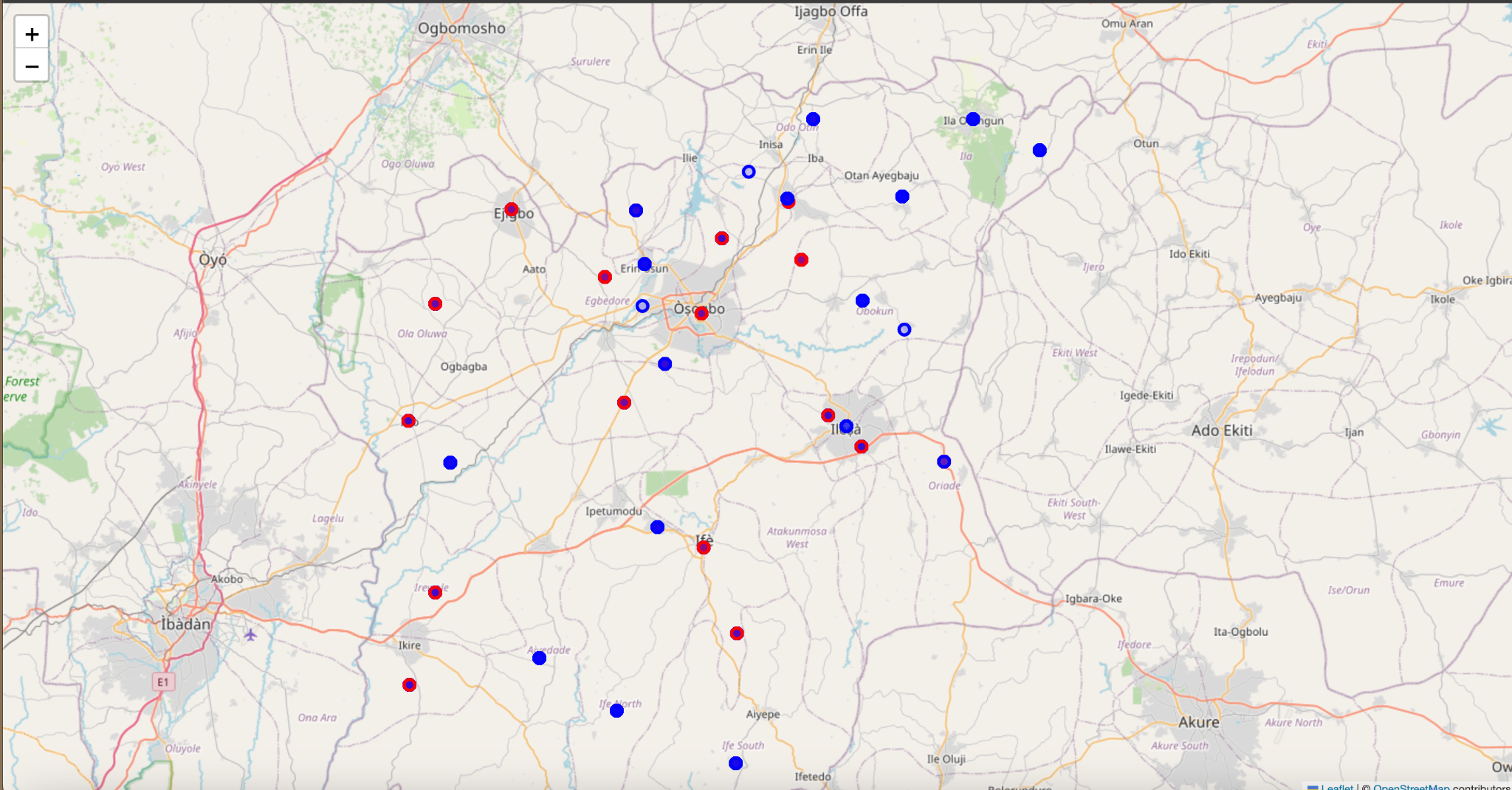


**SUMMARY OF FINDINGS**

The analysis revealed several polling units with significant outlier scores. These units displayed voting results that deviated substantially from the neighbors, suggesting potential irregularities

**Map of polling units and outliers**

Below is a map highlighting the polling units. Significant outliers are marked in red, while others are in blue



### **Detailed Examples of the Top 3 Outliers, with Explanations and Visualizations**

#### **Top 3 Outliers**

Based on the analysis, the top 3 outliers for each party are as follows:

**PDP:**

1. **Polling Unit: 29-07-05-004**
   * **Outlier Score:** 249.02
   * **Explanation:** This polling unit shows a significantly higher number of votes for PDP compared to its neighboring units. This deviation suggests potential irregularities or influences affecting the voting results.
2. **Polling Unit: 29-26-07-001**
   * **Outlier Score:** 228.00
   * **Explanation:** The outlier score for this polling unit is also substantially high, indicating an unusual voting pattern that warrants further investigation.
3. **Polling Unit: 29-26-07-002**
   * **Outlier Score:** 228.00
   * **Explanation:** Similar to the previous unit, this polling unit's results significantly deviate from neighboring units, raising concerns about the integrity of the votes.

**NNPP:**

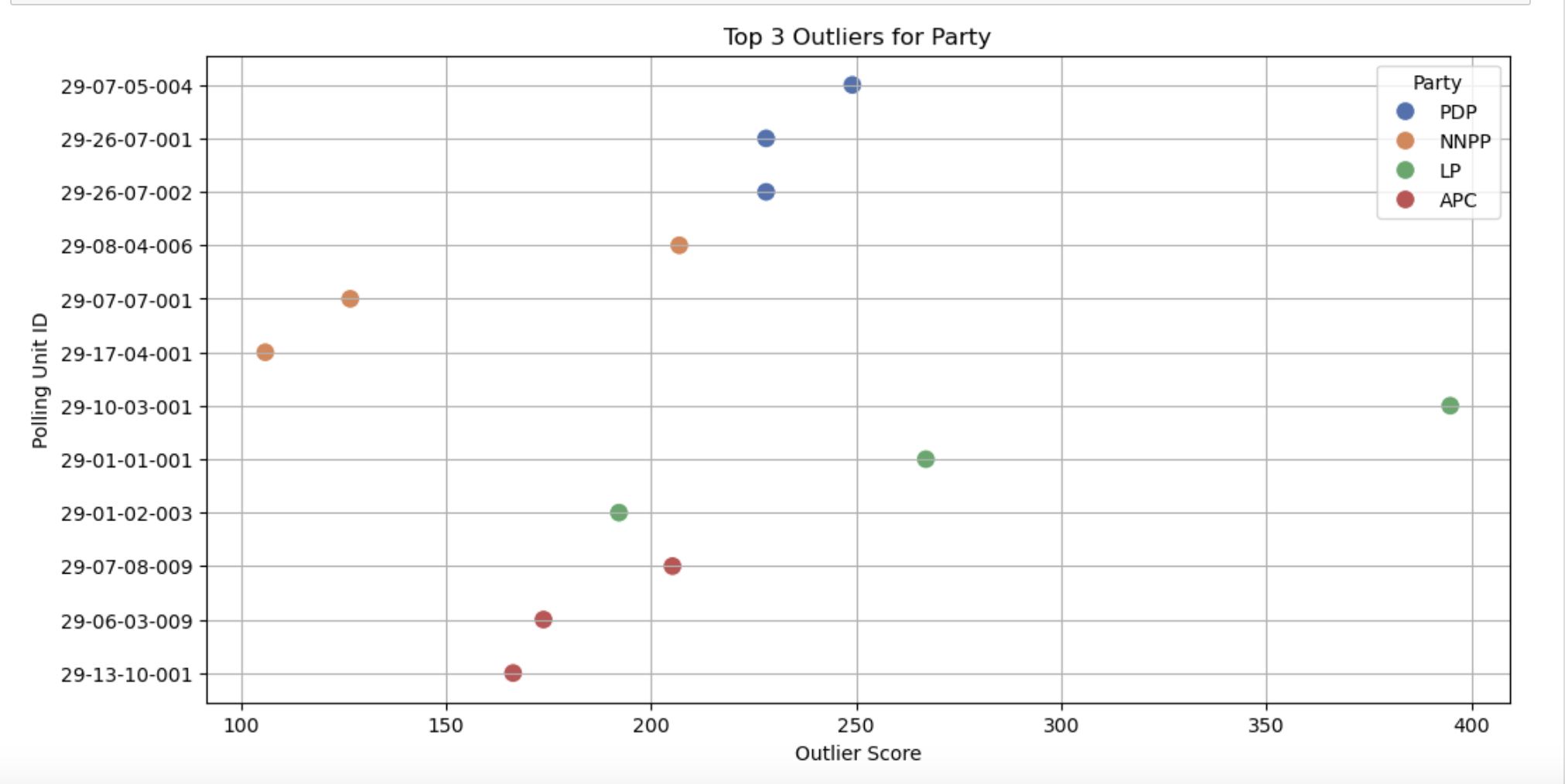
1. **Polling Unit: 29-08-04-006**
   * **Outlier Score:** 206.83
   * **Explanation:** This polling unit has a high outlier score for NNPP, suggesting that the number of votes received by NNPP is unusually high compared to nearby units.
2. **Polling Unit: 29-07-07-001**
   * **Outlier Score:** 126.63
   * **Explanation:** The significant deviation in the number of votes for NNPP at this polling unit indicates potential voting irregularities.
3. **Polling Unit: 29-17-04-001**
   * **Outlier Score:** 105.92
   * **Explanation:** This unit also exhibits a noticeable outlier score, suggesting that the voting results for NNPP may have been influenced or manipulated.

**LP:**

1. **Polling Unit: 29-10-03-001**
   * **Outlier Score:** 394.76
   * **Explanation:** The highest outlier score across all parties, indicating a major deviation in voting results for LP. This polling unit requires urgent attention to determine the cause of this irregularity.
2. **Polling Unit: 29-01-01-001**
   * **Outlier Score:** 266.93
   * **Explanation:** This polling unit also shows a significant deviation, suggesting potential issues with the voting process or result accuracy.
3. **Polling Unit: 29-01-02-003**
   * **Outlier Score:** 192.13
   * **Explanation:** The outlier score indicates a considerable deviation in the number of votes for LP, warranting further investigation.

**APC:**

1. **Polling Unit: 29-07-08-009**
   * **Outlier Score:** 205.17
   * **Explanation:** This polling unit has a high outlier score for APC, suggesting an abnormal voting pattern that needs to be examined.
2. **Polling Unit: 29-06-03-009**
   * **Outlier Score:** 173.73
   * **Explanation:** The significant deviation in APC votes indicates potential irregularities in the voting process.
3. **Polling Unit: 29-13-10-001**
   * **Outlier Score:** 166.33
   * **Explanation:** This unit also exhibits a high outlier score, suggesting that the voting results for APC may not be accurate or have been influenced.



**Significant Findings:**

* The analysis identified several polling units with significant outlier scores, suggesting potential voting irregularities.
* Outliers were found for each party, indicating that no single party was exclusively affected by anomalies.
* The outlier scores for PDP and LP at certain polling units are significantly high, indicating potential irregularities or influences affecting the voting results.
* The highest outlier score was observed for LP at polling unit 29-10-03-001 with a score of 394.76, which is notably higher than other outliers across all parties.
* These findings suggest that specific polling units exhibit unusual voting patterns that deviate significantly from their neighboring units.

#### **Conclusion**

The analysis has identified specific polling units with significant deviations in voting patterns for each party. These outliers suggest potential irregularities that need to be addressed to ensure the integrity of the election results. By investigating these outliers and implementing the recommended actions, future elections can be conducted with greater accuracy and transparency, maintaining public trust in the electoral process.

### **Recommendations**

* Further investigation is recommended for the identified outlier polling units to ensure election integrity.
* Conduct a detailed review of the voting process and results at the identified polling units with high outlier scores to determine the cause of the deviations.
* Verify the accuracy of the recorded votes and ensure that all procedures were followed correctly to maintain election integrity.
* Implementing more robust monitoring and verification processes during elections could help reduce the occurrence of such anomalies.