# Outlier Detection Report — Enugu State

Prepared by: Enoch Chukwuebuka  
Data Analytics Track – HNG Internship (Stage 2B)  
October 2025

## 1. Introduction

This report presents the results of a geospatial outlier-detection analysis performed on Enugu State election polling unit data (crosschecked dataset). The goal is to identify polling units whose vote counts for major parties deviate significantly from their geographical neighbours within a 1 km radius, which may indicate anomalies worth further review.

## 2. Dataset & Preparation

The dataset used for this analysis was sourced from the crosschecked Enugu State election data provided in the HNG Internship repository. Each polling unit record includes details such as LGA, Ward, PU-Code, PU-Name, votes for each major party, and geographic coordinates (Latitude and Longitude). Missing coordinates were retrieved via geocoding techniques. The dataset was then filtered to ensure complete records before further analysis.

A total of 3768 polling units were analyzed, all containing valid coordinates. Neighboring units were identified using a 1 km radius, yielding an average of approximately 149 neighboring polling units per location.

**State analyzed:** Enugu  
**Number of polling units analyzed:** 3768  
**Polling units with valid coordinates (Latitude & Longitude):** 3768  
**Average number of neighbouring polling units (1 km radius):** 148.63

Key dataset fields used: State, LGA, Ward, PU-Code, PU-Name, Latitude, Longitude, APC, LP, PDP, NNPP, Neighbours, and computed outlier score columns (e.g., APC\_Outlier\_Score). Geocoding was performed for missing coordinates prior to neighbour detection.

## 3. Methodology

1. Geocoding: Missing latitude/longitude values for polling units were geocoded to ensure spatial analysis.  
2. Neighbour identification: For each polling unit, neighbouring units were identified using a 1 km radius based on geodesic distances.  
3. Outlier scoring: For each party, an outlier score was computed as a Z-score relative to the neighbours: (PU\_votes - mean(neighbour\_votes)) / std(neighbour\_votes). Units with high positive scores indicate unusually high votes compared to neighbours; high negative scores indicate unusually low votes.  
4. Filtering & sorting: Results were filtered to remove NaN scores and sorted to identify the most significant outliers per party.

## 4. Findings

Below are the top 3 positive outlier polling units for each party (highest positive deviation compared to neighbours). These are primary candidates for further review or verification. The numeric score shown is the computed outlier (Z-score).

### 4.1 Top 3 Outliers — APC

The following table shows the top 3 polling units with the highest APC outlier scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| State | LGA | Ward | PU-Code | PU-Name | APC\_Outlier\_Score |
| ENUGU | ENUGU NORTH | OGUI TOWNSHIP | 14-04-13-001 | SALVATION ARMY PRIMARY SCHOOL I | 13.6817 |
| ENUGU | IGBO ETITI | OHAODO II | 14-07-09-005 | C.P.S. OHODO | 11.0841 |
| ENUGU | AWGU | OWELLI/AMOLI/UGBO/OGUGU | 14-02-09-017 | HEALTH CENTRE AMOLI | 9.6938 |
|  |  |  |  |  |  |

### 4.2 Top 3 Outliers — LP

The following table shows the top 3 polling units with the highest LP outlier scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| State | LGA | Ward | PU-Code | PU-Name | LP\_Outlier\_Score |
| ENUGU | IGBO EZE NORTH | UMUOZZI IV | 14-08-14-004 | U.P.S. AMACHI | 6.7076 |
| ENUGU | ENUGU NORTH | G.R.A | 14-04-07-005 | POLO PARK I | 5.2139 |
| ENUGU | ISI UZO | MBU II | 14-10-06-001 | C.P.S. MBU AKPOTI | 4.7251 |

### 4.3 Top 3 Outliers — PDP

The following table shows the top 3 polling units with the highest PDP outlier scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| State | LGA | Ward | PU-Code | PU-Name | PDP\_Outlier\_Score |
| ENUGU | NSUKKA | OGBOZALLA/IDI | 14-13-01-017 | AMADIM VILLAGE SQUARE | 10.8340 |
| ENUGU | AWGU | OGBAKU | 14-02-11-006 | COMM. SCHOOL EZIAMA | 9.3825 |
| ENUGU | IGBO ETITI | AKU III | 14-07-03-001 | UWANI OTOBO VILLAGE SQUARE | 9.2569 |

### 4.4 Top 3 Outliers — NNPP

The following table shows the top 3 polling units with the highest NNPP outlier scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| State | LGA | Ward | PU-Code | PU-Name | NNPP\_Outlier\_Score |
| ENUGU | UDI | UMUABI | 14-16-04-002 | COMMUNITY PRIMARY SCHOOL ENUGU UMUABI | 13.1061 |
| ENUGU | ENUGU EAST | MBULU-NJODO EAST | 14-03-12-006 | C.P.S. AKPOGA I | 11.8746 |
| ENUGU | NKANU WEST | AKPUGO II | 14-12-09-010 | AMAFOR OBUNO VILLAGE HALL | 10.6963 |

## 5. Statistical Summary of Outlier Scores

APC Outlier Score Summary:

count: 3768.000, std: 0.992, min: -1.681, 25%: -0.565, 50%: -0.380, 75%: 0.222, max: 13.682

LP Outlier Score Summary:

count: 3768.000, std: 0.992, min: -2.376, 25%: -0.791, 50%: -0.132, 75%: 0.594, max: 6.708

PDP Outlier Score Summary:

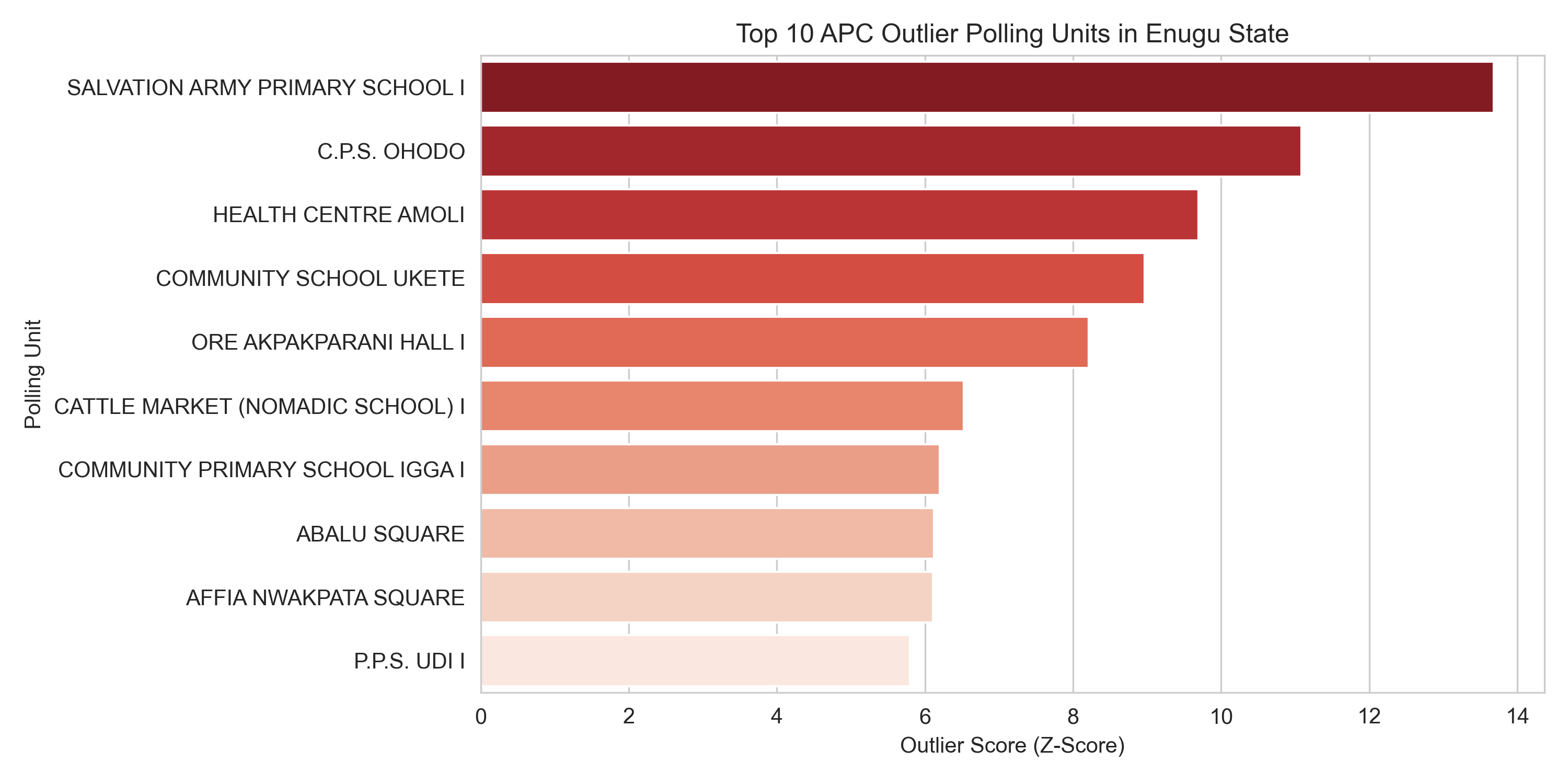
count: 3768.000, mean: 0.000, std: 0.992, min: -1.817, 25%: -0.611, 50%: -0.313, 75%: 0.233, max: 10.834

NNPP Outlier Score Summary:

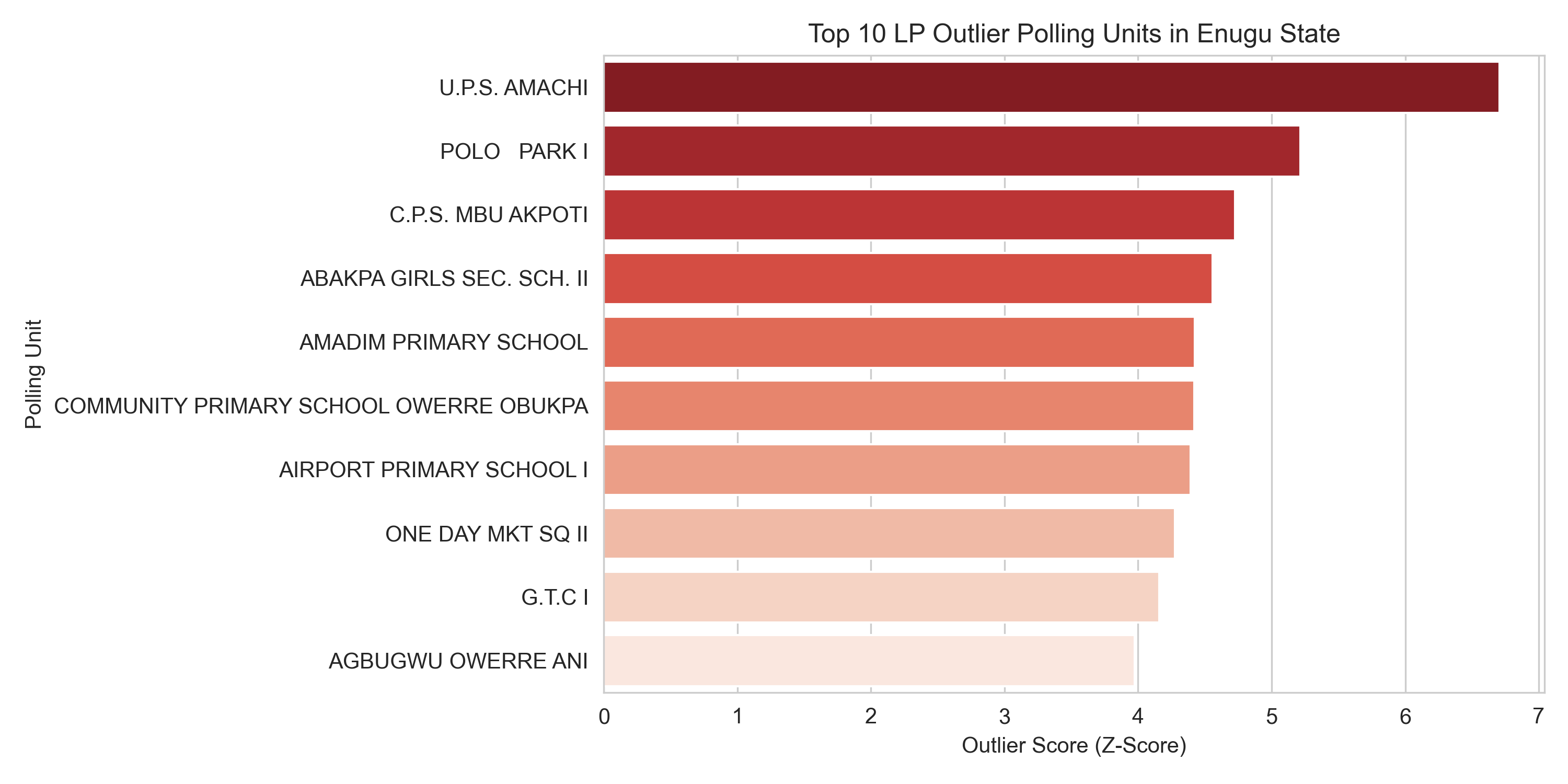
count: 3768.000, mean: 0.000, std: 0.988, min: -1.118, 25%: -0.479, 50%: -0.379, 75%: -0.111, max: 13.106

## 6. Visualizations

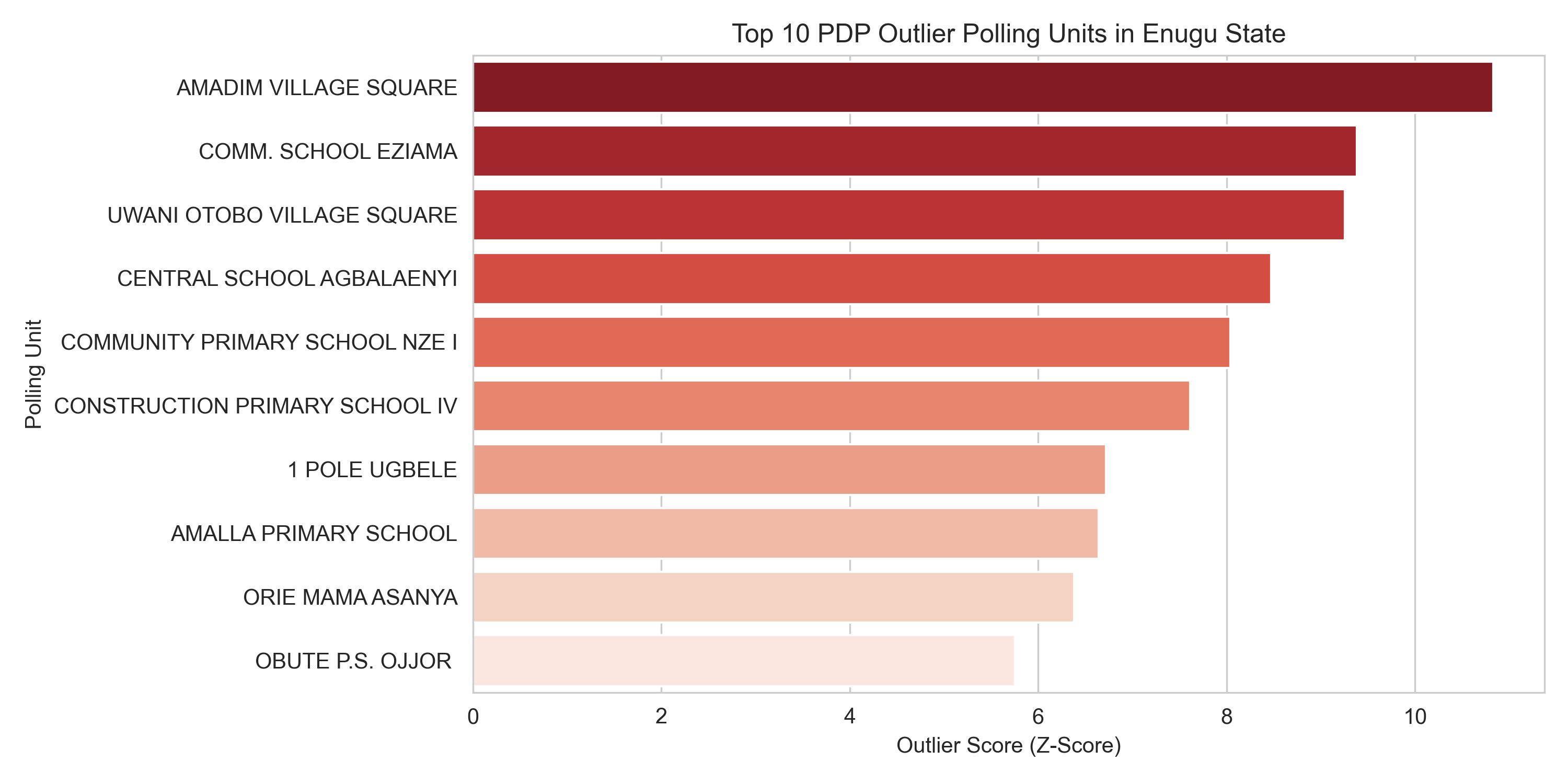
- Top 10 outlier bar charts for each party



As seen in the figure above SALVATION ARMY PRIMARY SCHOOL1 hold the highest APC outlier score followed by C.P.S OHODO school.



As seen in the figure above U.P.S AMACHI has the highest outlier score for LP in Enugu state.



As seen in the figure above AMADIM VILLAGE square has the highest outlier score for PDP.

A graph of red and black color

AI-generated content may be incorrect.

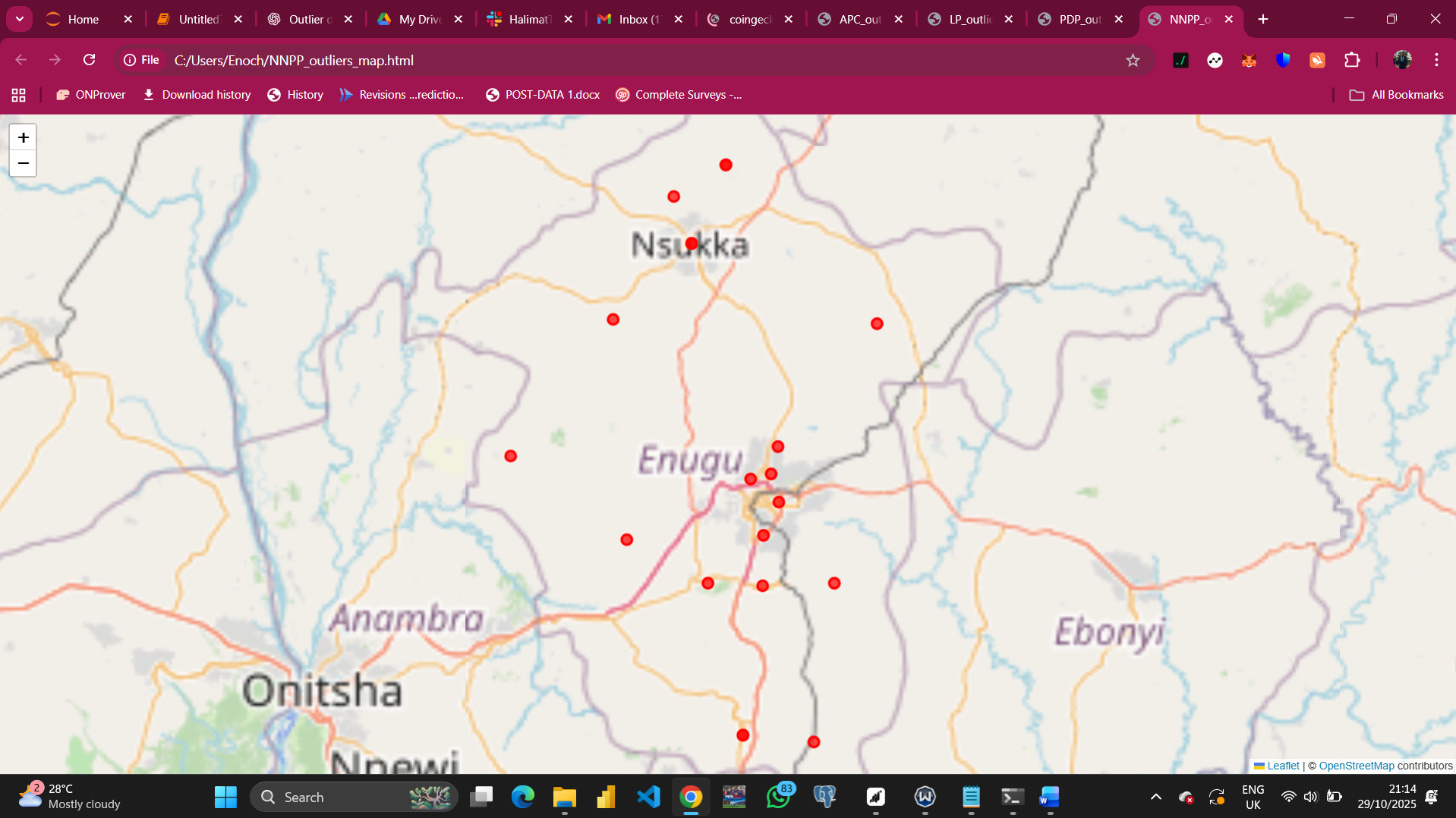
As seen in the figure above COMMUNITY PRIMARY SCHOOL ENUGU UMUABI shows the highest outlier score for NNPP in Enugu state.  
- Interactive folium map screenshots (APC\_outliers\_map.html etc.)  


FIGURE SHOWING NNPP OUTLIER MAP

FIGURE SHOWING PDP OUTLIER MAP

A screenshot of a computer screen

AI-generated content may be incorrect.

FIGURE SHOWING LP OUTLIER MAP

A screenshot of a computer screen

AI-generated content may be incorrect.

FIGURE SHOWING APC OUTLIER MAP  
Place each image under the appropriate section (Top 3 Outliers — [Party]) and include a short caption describing the pattern.

## 7. Interpretation & Recommendations

Interpretation:  
- Polling units with high positive outlier scores are anomalously high compared to their neighbours and should be prioritized for verification.  
- Clusters of outliers within the same LGA/Ward may indicate localized irregularities or genuine demographic concentrations; further local investigation is required.