Report

Task

> Allegations of vote manipulation and irregularities have been widespread, prompting a thorough investigation into the matter.Your mission, should you choose to accept it, is to help us uncover potential voting irregularities and ensure the transparency of the election results. You will achieve this by identifying outlier polling units where the voting results deviate significantly from neighbouring units, indicating potential influences or rigging.

Methodology

Step 1

First the data is loaded into mongodb for efficient storage and to use its optimal geoindexing.
To achieve this, the data is pulled and transformed to be suitable for MongoDB. MOngoDB stores location data in this object format {type: "type e.g Points", coordinates:[long, lat]}

Then Created index on the locaion field

Step 2

The data is transformed and key metrics calculated. Mongodb aggregate pipeline is used

- Select the data for the state
- Do a self join, the criteria for this join is those within 3Kmm and not equal to the same pooling unit.
- For each of hte data, calculate summary statistics (mean, std and count) of those within its neighborhood, for each party column
- return the needed data

Step 3

Calculate the Z score which tells you how dispersed a point is from the rest points.
Z values are typically within 3 std values. values above this can be considered outlliers.

A Z score of 4 is achieved by 0.003% of points (i.e 3 in 100k points) There are over 100k pooling unit centers in the whole country, this implies that we have to push the z score boundary.

I have chosn a threshold of 5. (well 3 is still fine)

Step 4

Find the outliers for each party vote counts by subsetting /filtering out the Z score below the threshold, also thos with little numbers of sorrouding neighbours as the count would also influence the result. I have set a count threshold of 3 values minimum.
Those centres with only 1 or 2 neighbours may not be considered outliers (as what is observed in riverrine areas and islands)

Step 5

Plot and observe each party and see locations and influence of their outliers

Step 6

Compare results across the parties

Result

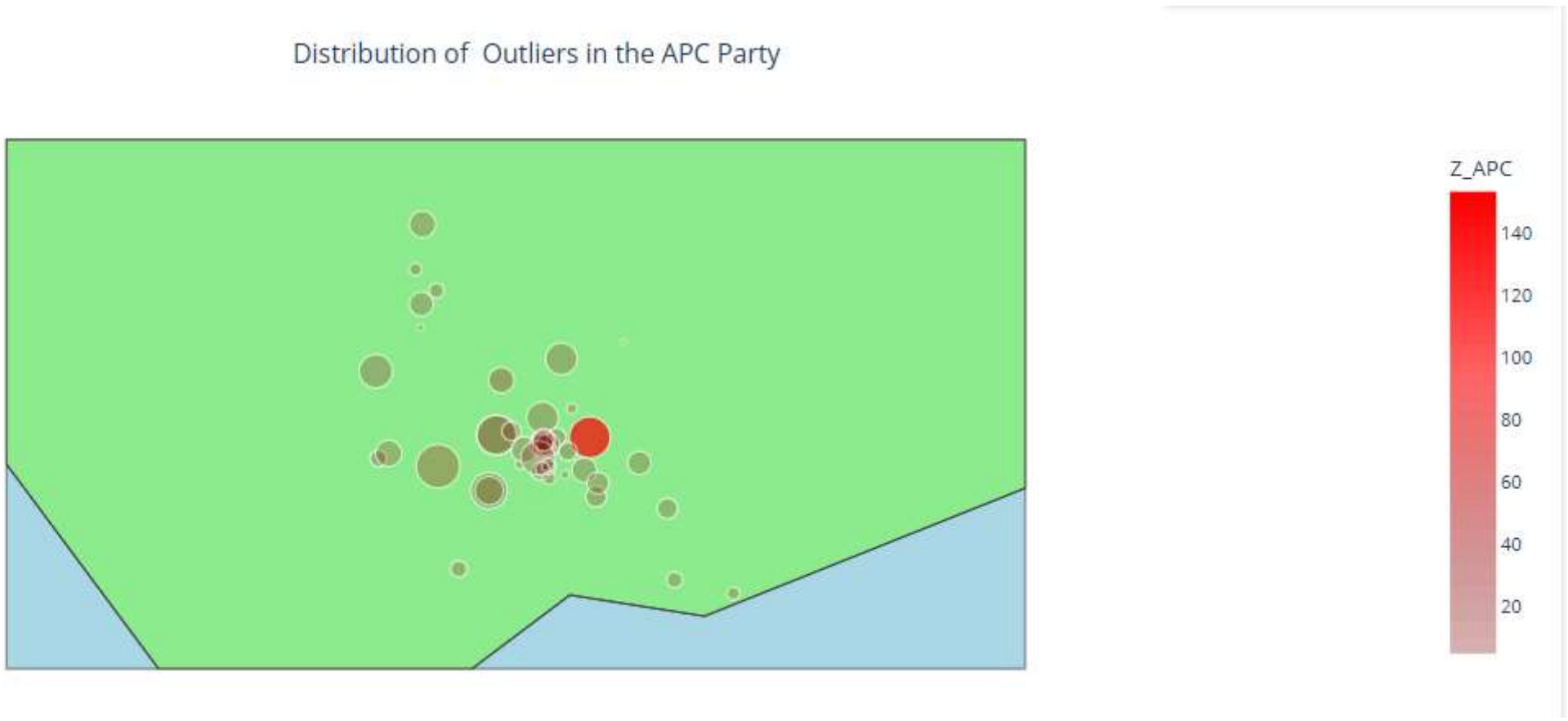
Top 5 Outliers in Each Party

The color sclae shows the values of the Z-score, red meaning higher z scores, the size of the bubble indicate the size of the chart as indicated by the values compared to those around

Top 5 Outliers in APC

Name	Latitude	Longitude	Z_APC	APC	APC_mean	APC_std	count
VILLAGE SQUARE, OYIGBO	4.874907	7.134533	153.202323	500	1.840909	3.251642	132
AGWUO/AWHUA OMAGWA, IKWERRE	5.020634	6.874301	38.061114	191	13.750000	4.656984	4
AGADA I, SCHOOL HALL, ABUA-ODUAL	4.801506	6.687099	33.798165	562	21.800000	15.983116	20
STATE SCHOOL HALL, OKOLOMADE II, ABUA-ODUAL	4.821016	6.511460	13.268363	81	21.333333	4.496913	3
MGBUCHI/ELEDO TOWN RUKPOKU, OBIO/AKPOR	4.924600	6.995200	12.578233	299	16.134615	22.488483	52

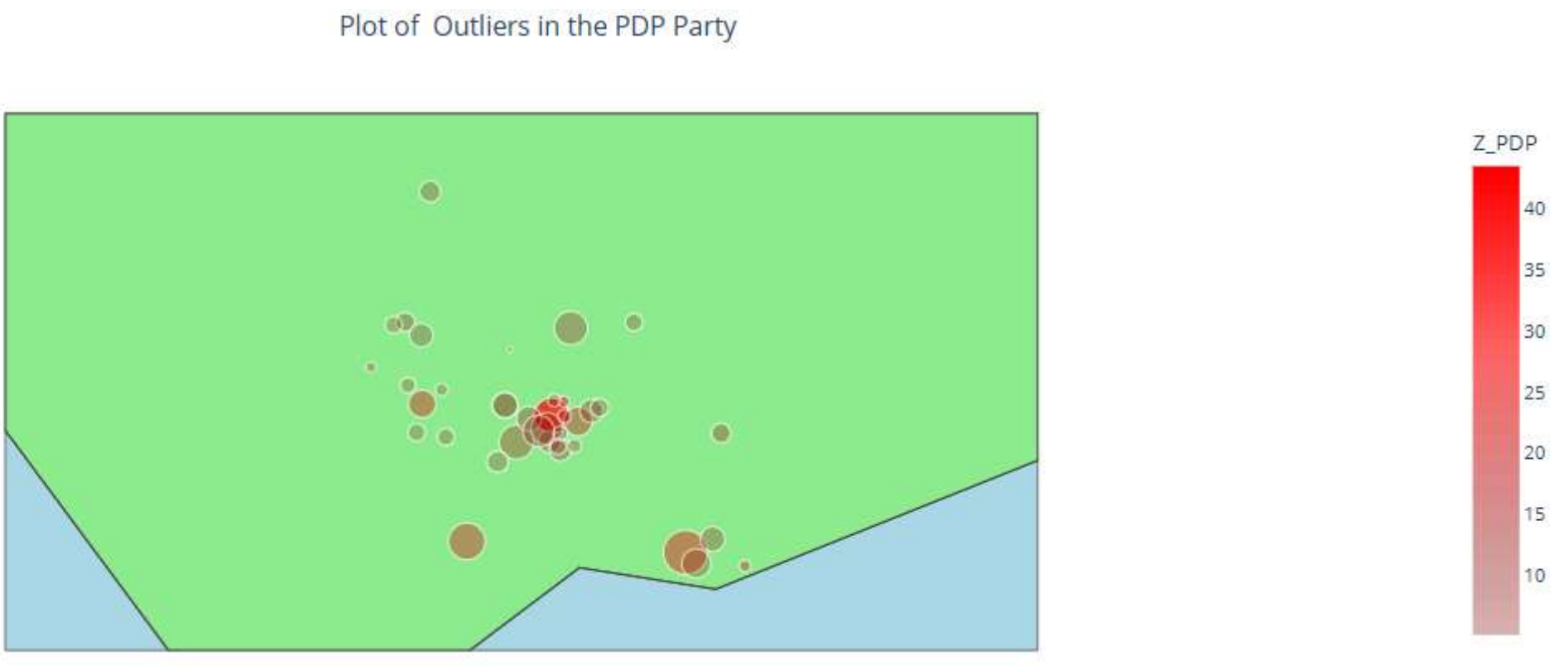
Geographical Distribution of Outliers IN APC



Top 5 Outlier In PDP

Name	Latitude	Longitude	Z_PDP	PDP	PDP_mean	PDP_std	count
0PEN SPACE BY 76 NVUIGWE RD BESIDE POLICE POST...	4.855528	6.994628	43.487272	243	3.411141	5.509402	377
UMUMBRA TOWN HALL UMUOLA, OMUMA	4.890700	7.030380	29.393877	25	1.000000	0.816497	3
OKAMA/ POLOKIRI VILLAGE SQUARE, ANDONI	4.511998	7.382942	26.217515	392	16.896000	14.307382	125
OKOBOH TOWN HALL, ABUA-ODUAL	4.884460	6.618440	22.864577	152	6.142857	6.379175	7
PEDRO'S COMPOUND BENIBO, AKUKU TORU	4.539604	6.747379	22.566498	272	14.480769	11.411573	156

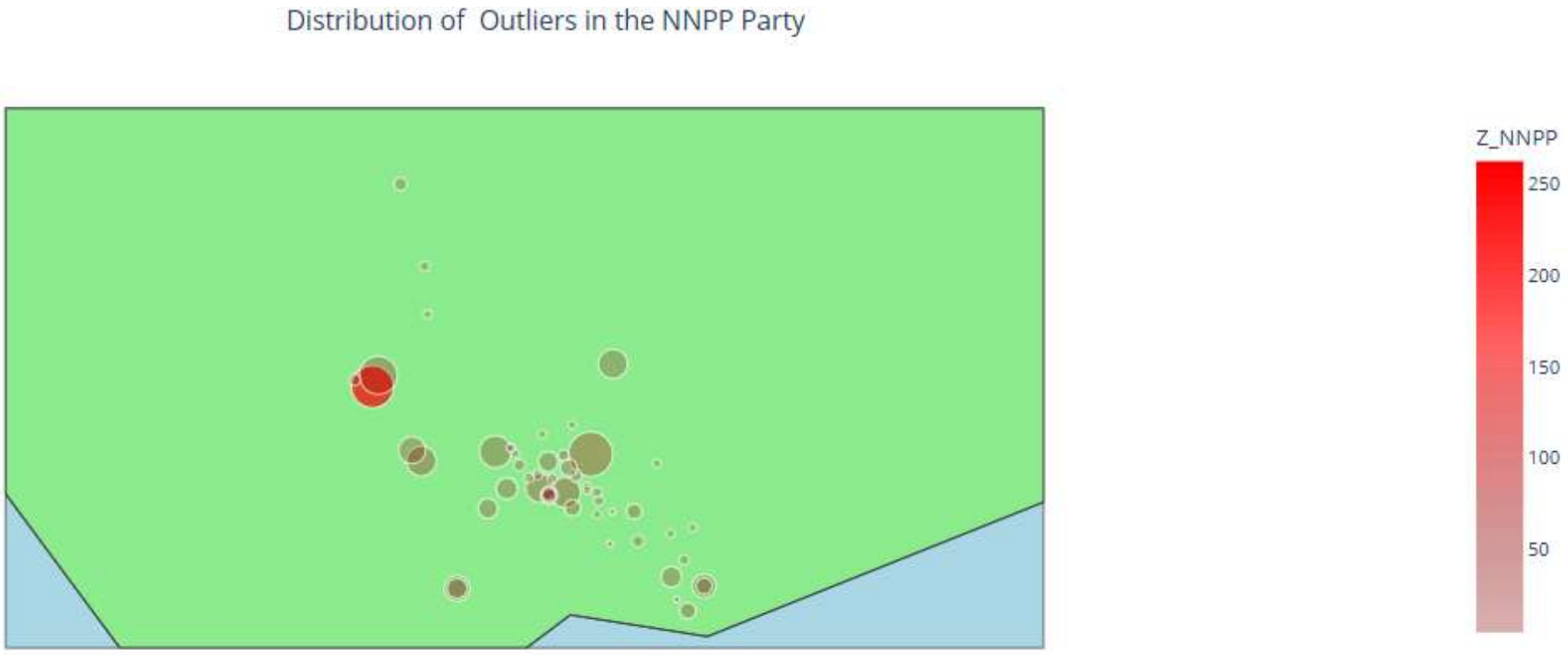
Geographical Distribution of Outliers in PDP



Top 5 Outliers in NNPP

Name	Latitude	Longitude	Z_NNPP	NNPP	NNPP_mean	NNPP_std	count
TOWN HALL, EGBEE, AHOADA WEST	5.043357	6.504660	261.817973	54	0.044444	0.206080	90
VILLAGE SQUARE, OYIGBO	4.874907	7.134533	78.737036	60	0.303030	0.758182	132
ABULOMA UPE III, PORT HARCOURT	4.779366	7.060843	68.967888	28	0.136364	0.404009	44
STATE SCHOOL HALL, OBARANY, ABUA-ODUAL	4.857630	6.645190	67.066220	27	0.140000	0.400500	50
KENNETH COMMERCIAL SCHOOL I, PORT HARCOURT	4.789949	6.988290	50.191187	25	0.204082	0.494029	147

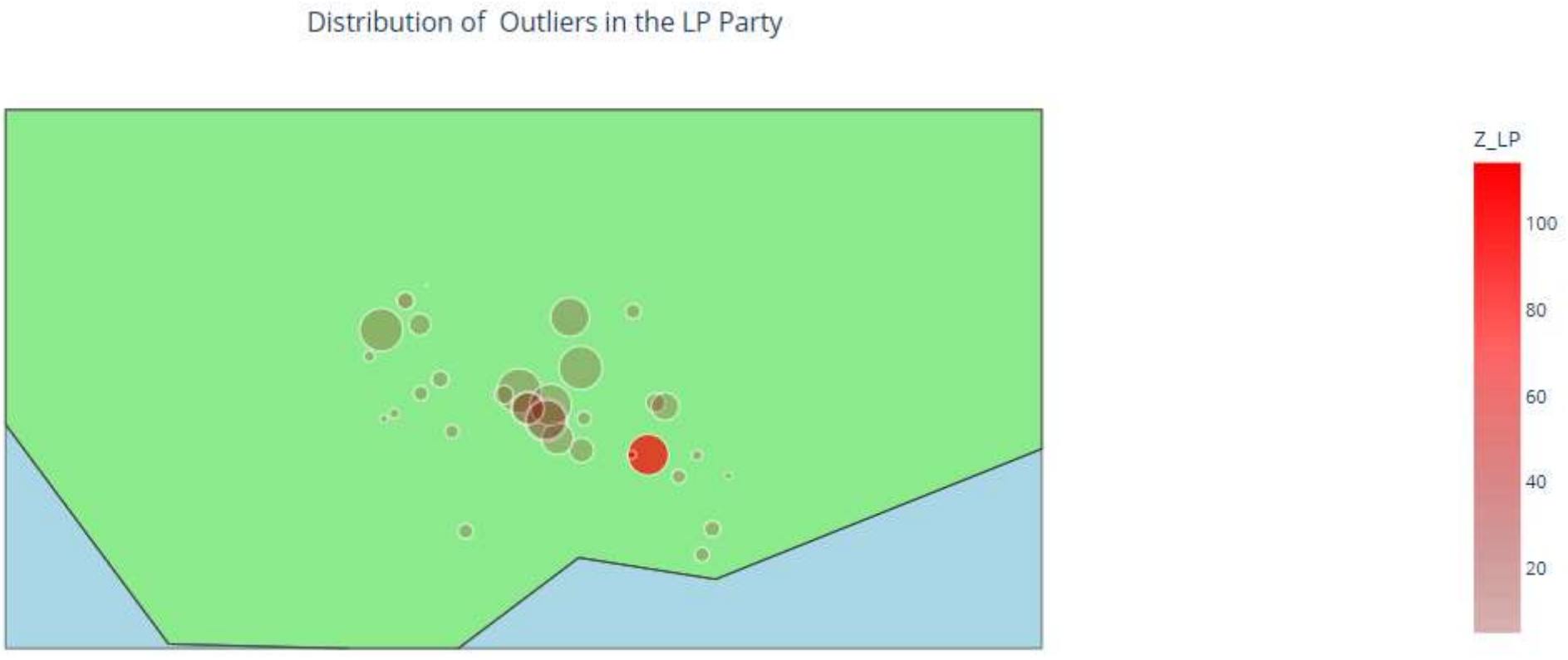
Geographical Distribution of Outliers in NNPP



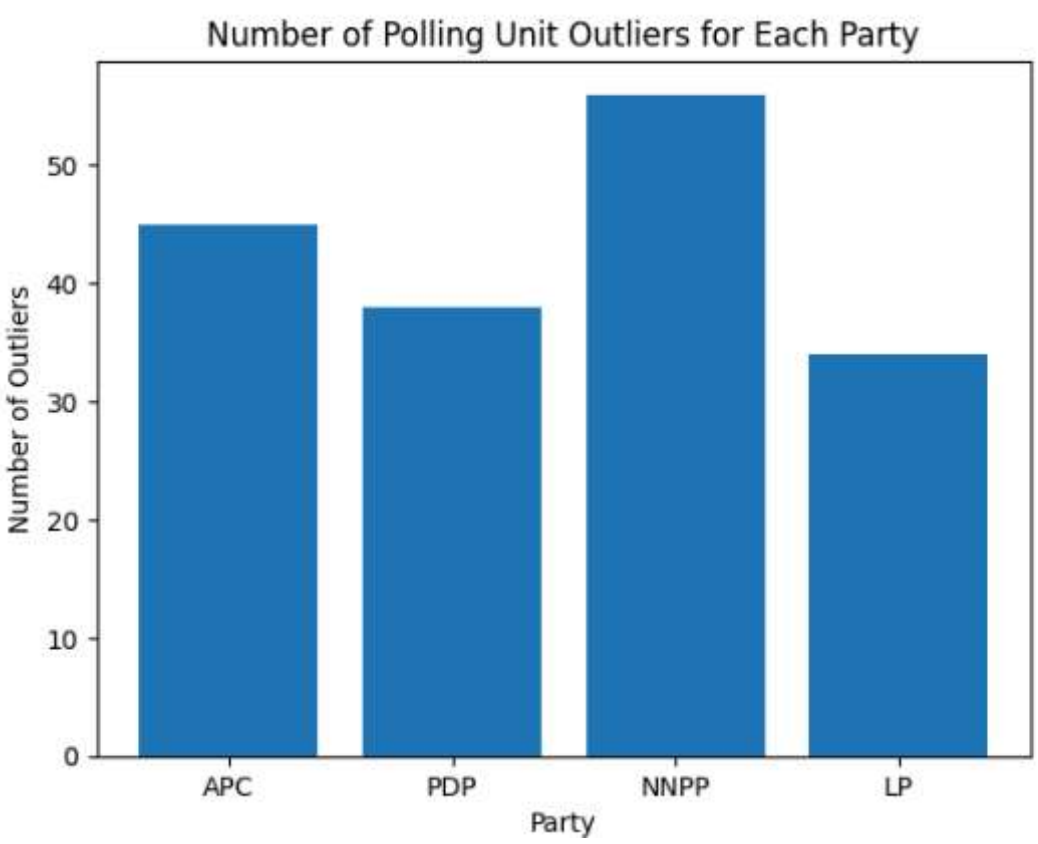
Top 5 Outliers in LP

Name	Latitude	Longitude	Z_LP	LP	LP_mean	LP_std	count
C.P.S I, BOTEM, TAI	4.730900	7.27657	114.054974	443	2.825000	3.859323	80
TOWN HALL, OGODA, AHOADA WEST	5.115790	6.57502	28.730584	76	4.333333	2.494438	3
COMM. PRI. SCH. ISHIAYI, AHOADA WEST	5.043357	6.50466	16.045960	490	24.022222	29.040194	90
STATE SCHOOL HALL, OKOLOMADE II, ABUA-ODUAL	4.821016	6.51146	15.202796	15	0.666667	0.942809	3
OKOBOH TOWN HALL, ABUA-ODUAL	4.884460	6.61844	13.736956	57	4.857143	3.795809	7

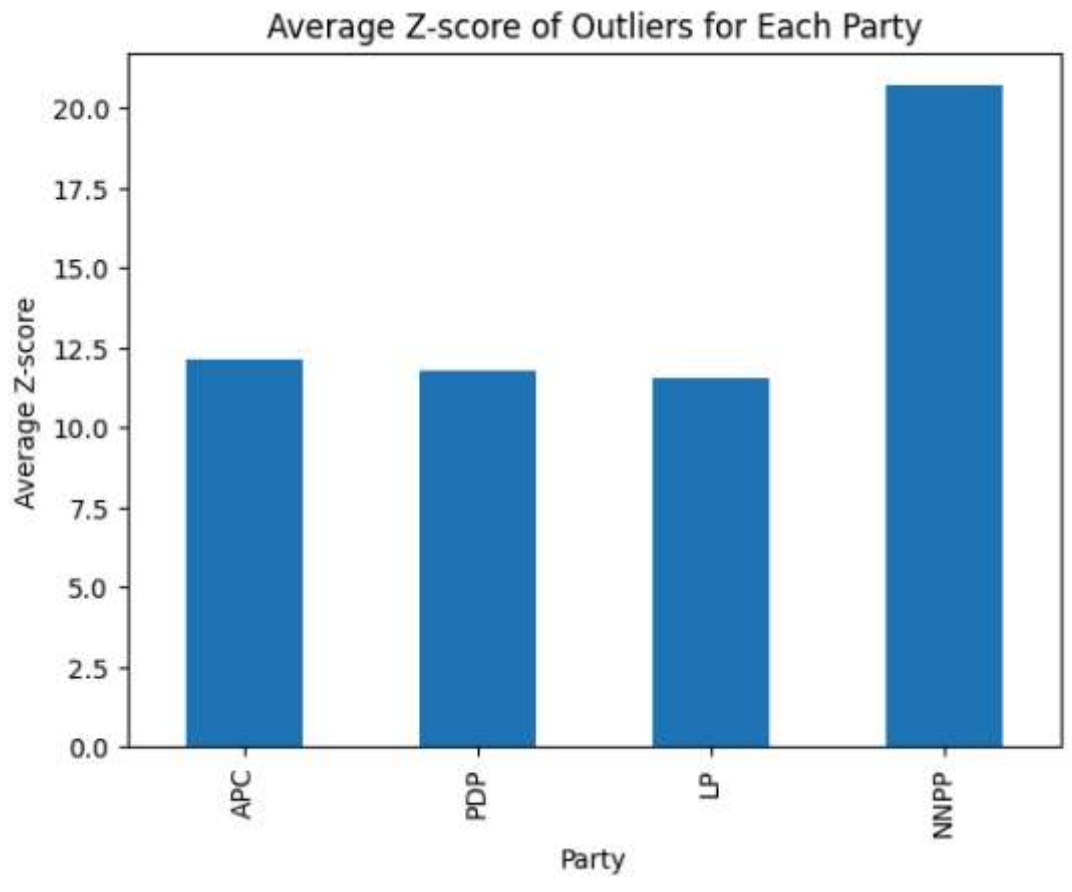
Geographical Distribution oF LP



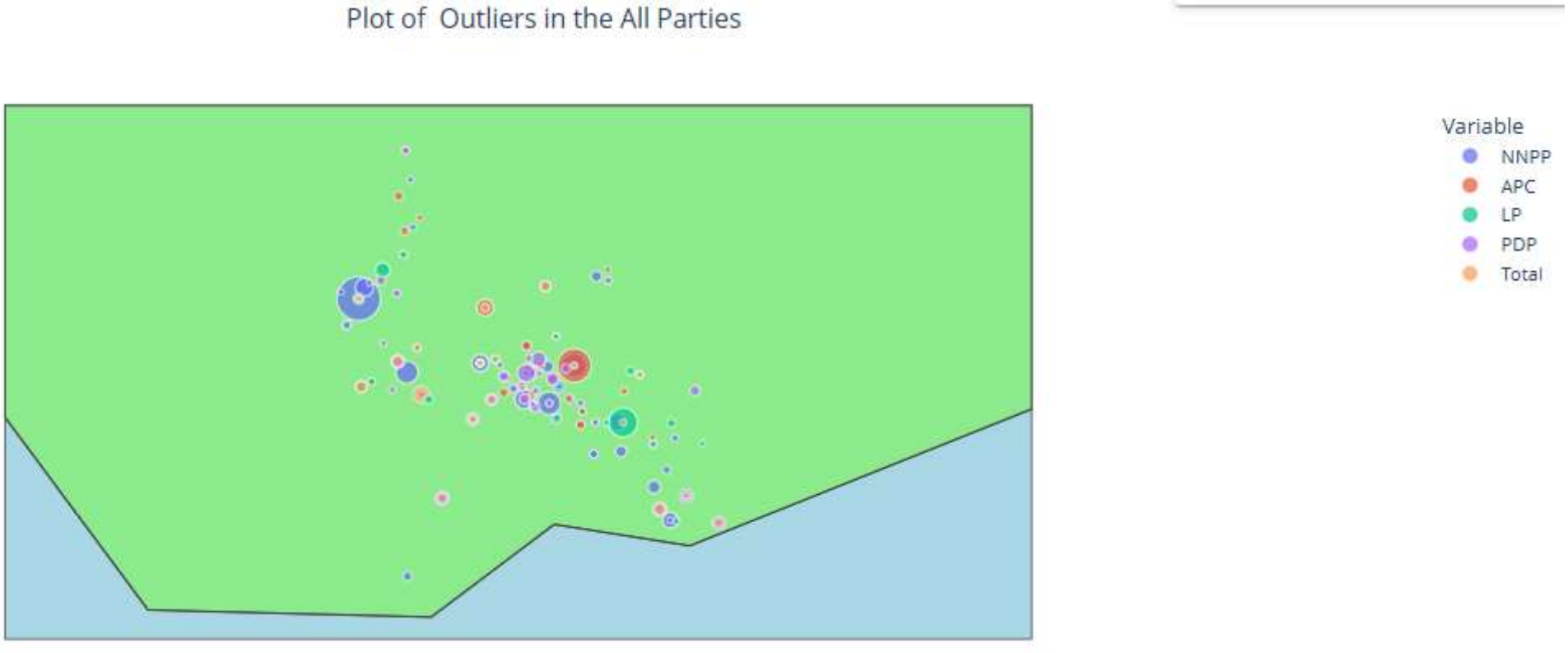
Number OF Pooling Unit Outliers for each Party



Average Outlier Score



Geographical Plot Of All Parties



The color represent different Parties the bubble sizes represent the Outlier Score

Double-click (or enter) to edit

Summary

- Of 4769 polling unit stations, outlier value where detected in 156 different units
- The parties had
    - APC : 45
    - PDP : 38
    - NNPP : 56
    - LP: 34
- NNPP had both highes number in outliers in unit and average

Start coding or generate with AI.