Election Data Outlier Detection Report (Jigawa State)

Introduction

The purpose of this report is to identify the outliers in the voting patterns for different parties in a dataset of polling units. The parties considered are APC, LP, PDP, and NNPP. In Jigawa state Outliers are determined based on the votes each party received at a polling unit compared to the average votes received by that party within a 1 km radius of the polling unit.

Methodology

1. Data Preparation: Load and examine the dataset containing the polling units' data.
2. Data cleaning: Polling units at which points (latitude and longitude) were not gotten. Records were dropped as there was no justifiable method to fill in missing values.

3. Distance Calculation: Calculate the distance between polling units to identify neighbors within a 1 km radius range.

4. Outlier Calculation: For each polling unit, calculate the deviation of its votes from the average votes of its neighbors.

5. Sorting and Analysis: Sort the polling units based on their outlier scores and identify the top 3 outliers for each party.

Data Preparation

The dataset is read into a DataFrame and inspected to understand its structure. Key columns include `PU-Name`, `latitude`, `longitude`, and the votes for each party (`APC`, `LP`, `PDP`, `NNPP`).

Data Cleaning

Polling units at which points (latitude and longitude) were not gotten. Records were dropped as there was no justifiable method to fill in missing values.

Distance Calculation

A distance matrix is created to calculate the distances between all pairs of polling units. This helps in identifying the neighbors within a 1 km radius of each polling unit.

Outlier Calculation

For each polling unit, the outlier scores for each party are calculated as follows:

- Find neighboring polling units within a 1 km radius.

- Calculate the mean votes for each party among the neighbors.

- Compute the absolute deviation of the polling unit's votes from these means.

Sorting and Analysis

The polling units are sorted based on their outlier scores in descending order. The top 3 outliers for each party are identified and presented.

Results

1. Top 3 APC Outliers:

```python
sorted_apc = outlier_scores.sort_values(by='APC_outlier', ascending=False).head(3)
```

```
print("Top 3 APC Outliers:")

print(sorted_apc)
```

```
Top 3 APC Outliers:
                                                    PU-Name   latitude  longitude
1548      SABON GARI PRI. SCH. SABON GARI  I  11.111372   7.731787
12                 KUDIGIN PRI. SCH. KUDIGIN  12.359314   9.989061
824   R/MAGUZAWA PRI. SCH. R/MAGUZAWA KARSHI  12.278974   9.430907

      APC_outlier  LP_outlier  PDP_outlier  NNPP_outlier  \
1548   290.227273    0.272727    67.636364      4.590909
12      262.095669    0.756303     7.532644     26.577246
824    257.663196    0.934930    94.684539     14.668922

                                                  Neighbours
1548  [GIDAN BAKWARE K. FADA GIDAN BAKWARE, KAFUR PR...
12    [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
824   [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
```

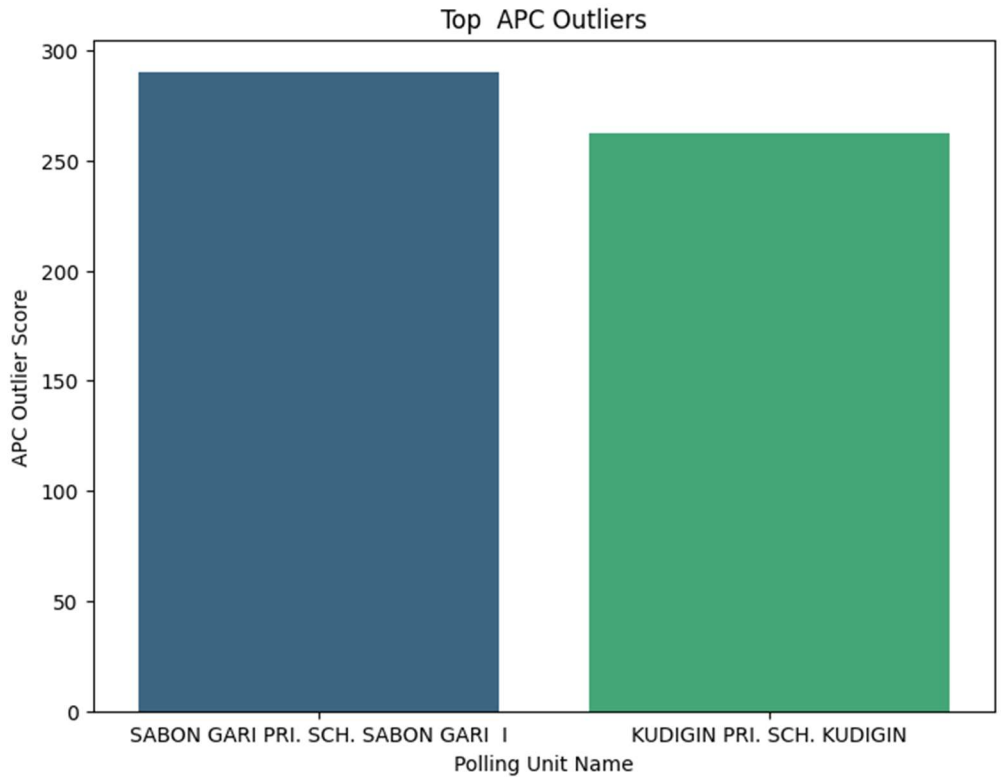

Figure 1: Bar chart showing top outliers for APC

2. Top 3 LP Outliers:

```python
sorted_lp = outlier_scores.sort_values(by='LP_outlier', ascending=False).head(3)
print("Top 3 LP Outliers:")
print(sorted_lp)
```

```
Top 3 LP Outliers:
|     |     |     |     |     |   PU-Name   latitude  longitude  APC_outlier  \
114  UKU DA SIS I PRI. SCH. /UKU DA SISI  12.779998   9.016330    89.956679
338  DIKUKAWA PRI. SCH./D. KUKAWA/FULANI  11.969001   8.870860    88.499348
516                 ADUWA PRI.SCH/ADUWA  12.420000   9.168658     8.706250
|
|     LP_outlier  PDP_outlier  NNPP_outlier  \
114  168.003610    16.417870     20.303249
338  144.872229     9.284876      9.778357
516  140.055208    79.892708     19.412500
|
|     |     |     |     |     |     |     | Neighbours
114  [AYAMA PRI. SCH. AYAMA I, AYAMA PRI. SCH. AYAM...
338  [TSURUTAWA, TSURUTAWA, GIDAN BAKWARE K. FADA G...
516  [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
```

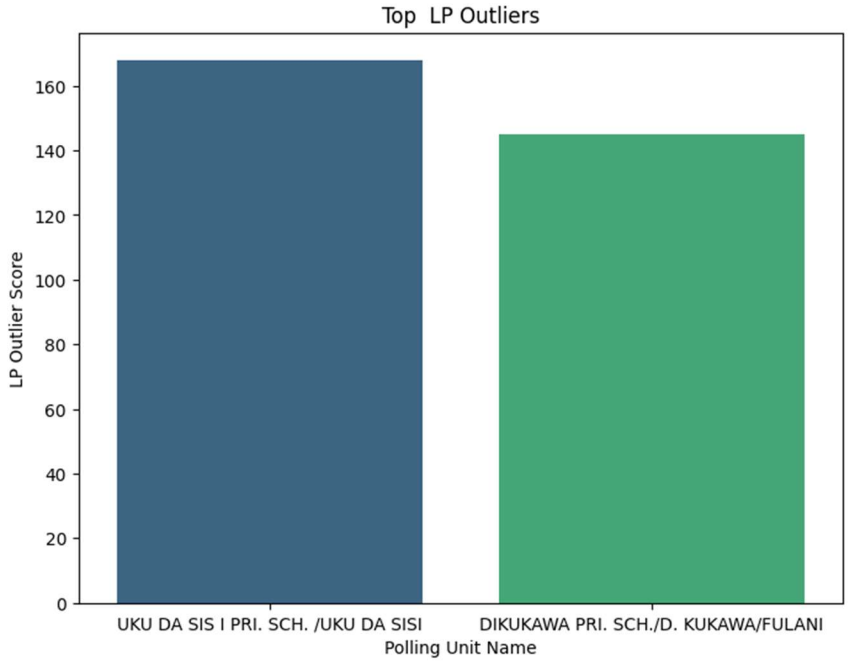

Figure 2: Bar chart showing top outliers for LP

3. Top 3 PDP Outliers:

```python

sorted_pdp = outlier_scores.sort_values(by='PDP_outlier', ascending=False).head(3)
```

```
print("Top 3 PDP Outliers:")

print(sorted_pdp)
```

```
Top 3 PDP Outliers:
|     |     |     |     |  PU-Name   latitude  longitude  APC_outlier  \
412            FAGOJI PRI. SCH. I  11.720236   9.350186    19.831853
55   GATAFA PRI. SCH. GATAF MADAKI  12.397197   9.974453    66.974009
431           GARU AREWA PRI. SCH  11.737472   9.348710    12.506982
|
|     LP_outlier  PDP_outlier  NNPP_outlier  \
412    0.888692   338.867969     30.287744
55     0.760234   292.994802      4.570500
431    0.856606   285.298604     18.422664
|
|     |     |     |     |     |     | Neighbours
412  [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
55   [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
431  [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
```

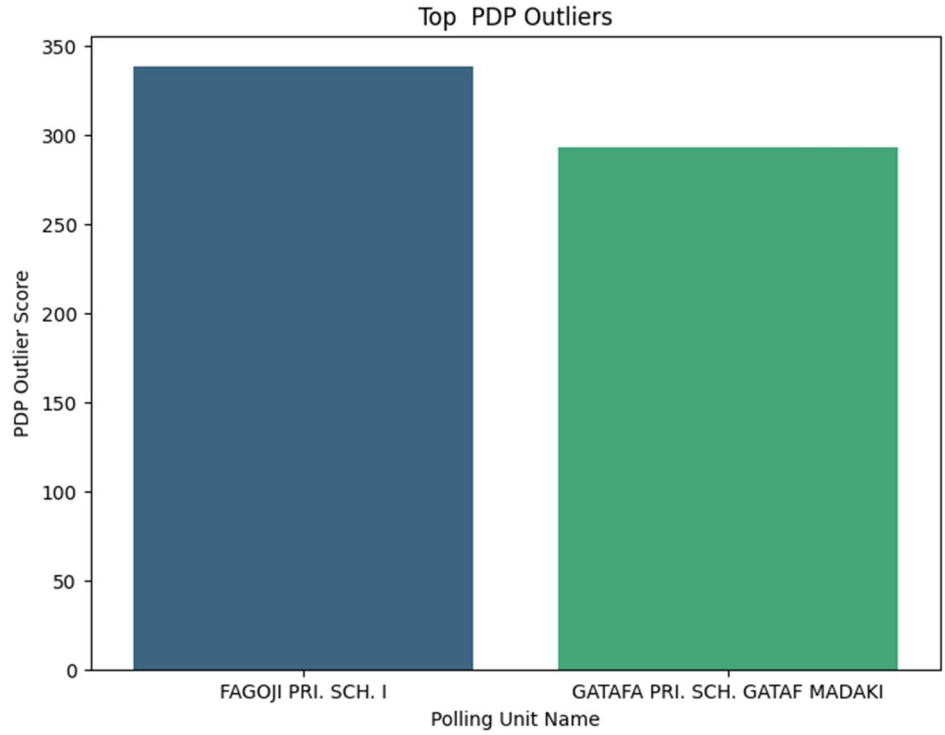

Figure 3: Bar chart showing top outliers for PDP

4. Top 3 NNPP Outliers:

```python
sorted_nnpp = outlier_scores.sort_values(by='NNPP_outlier', ascending=False).head(3)
print("Top 3 NNPP Outliers:")
print(sorted_nnpp)
```

```
Top 3 NNPP Outliers:
      PU-Name   latitude  longitude  APC_outlier  \
1571  NASARAWA PRI. SCH.  NASARAWA II  12.149826   9.154147     1.858680
702    GALAMBI PRI. SCH /GALAMBI FADA  11.270772   9.880277    11.443133
630        LAFIYA PRI. SCH. LAFIYA  I  12.731637  10.420828   151.416048

      LP_outlier  PDP_outlier  NNPP_outlier  \
1571    0.980929    47.570171    455.338386
702     1.039700    39.172747    334.183476
630     0.350669    56.661218    217.224368

                                                   Neighbours
1571  [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
702   [TSURUTAWA, TSURUTAWA, GASTIFI K. FADA , GASTI...
630   [AUYO SPECIAL PRI. SCH. AUYO KUKA, AUYO BAYI, ...
```

E

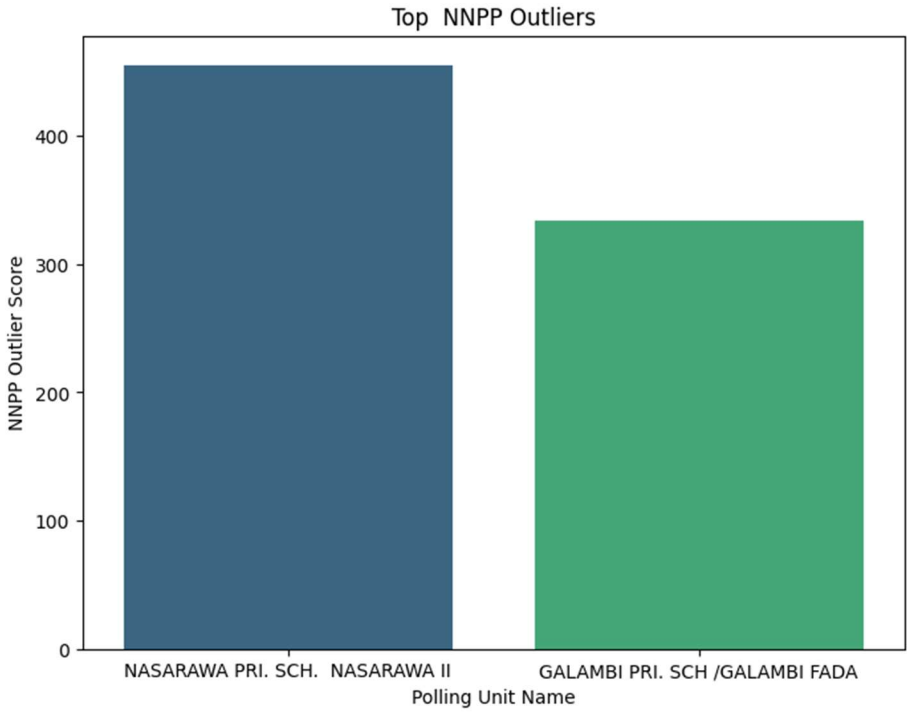

Figure 4: Bar chart showing top outliers for NNPP

Summary

This analysis provides insights into the polling units with the most significant deviations in voting patterns for each party. These outliers could indicate potential anomalies or areas with distinct voting behaviors. Further investigation may be needed to understand the reasons behind these outliers.