**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:

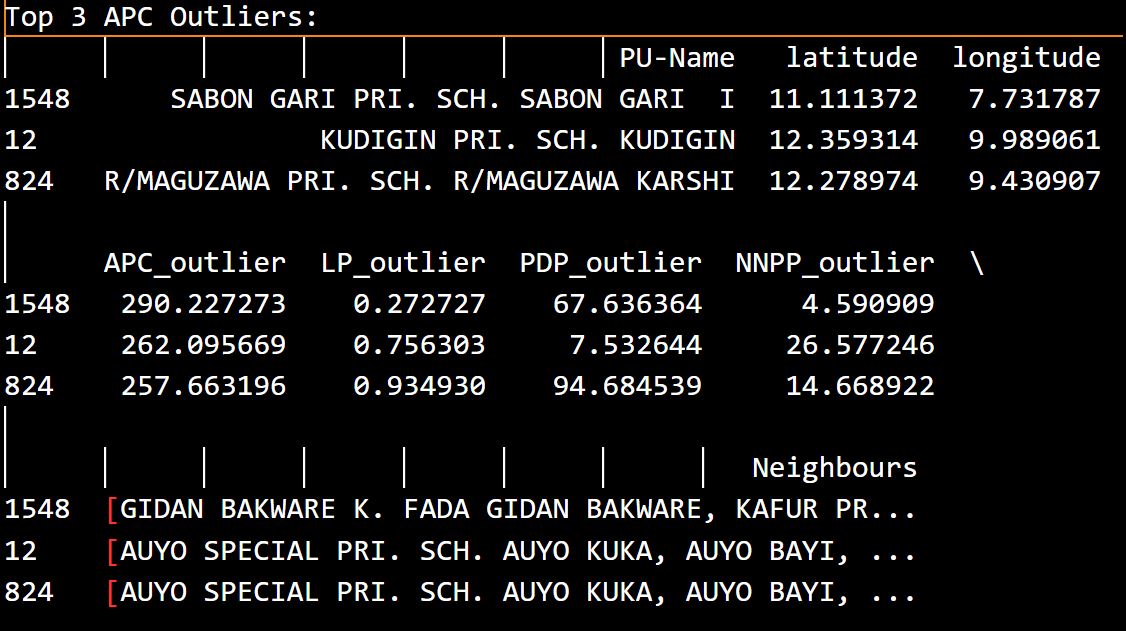
```python

sorted\_apc = outlier\_scores.sort\_values(by='APC\_outlier', ascending=False).head(3)

print("Top 3 APC Outliers:")

print(sorted\_apc)

```



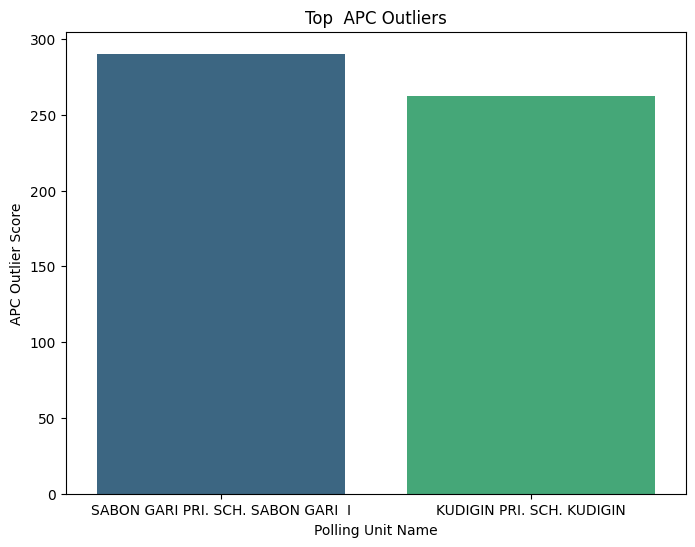


Figure 1: Bar chart showing top outliers for APC

2. Top 3 LP Outliers:

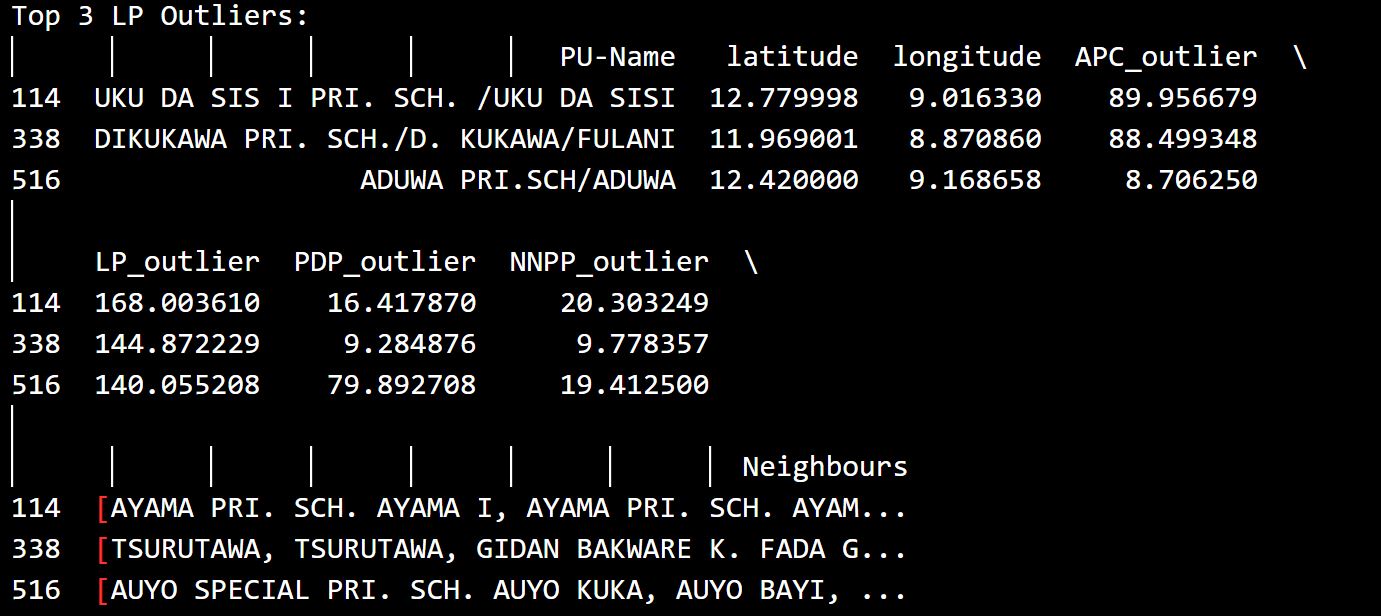
```python

sorted\_lp = outlier\_scores.sort\_values(by='LP\_outlier', ascending=False).head(3)

print("Top 3 LP Outliers:")

print(sorted\_lp)

```



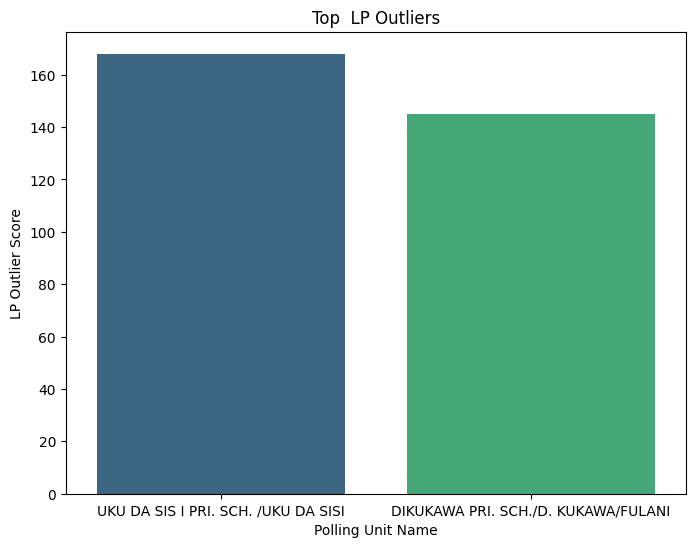


Figure 2: Bar chart showing top outliers for LP

3. Top 3 PDP Outliers:

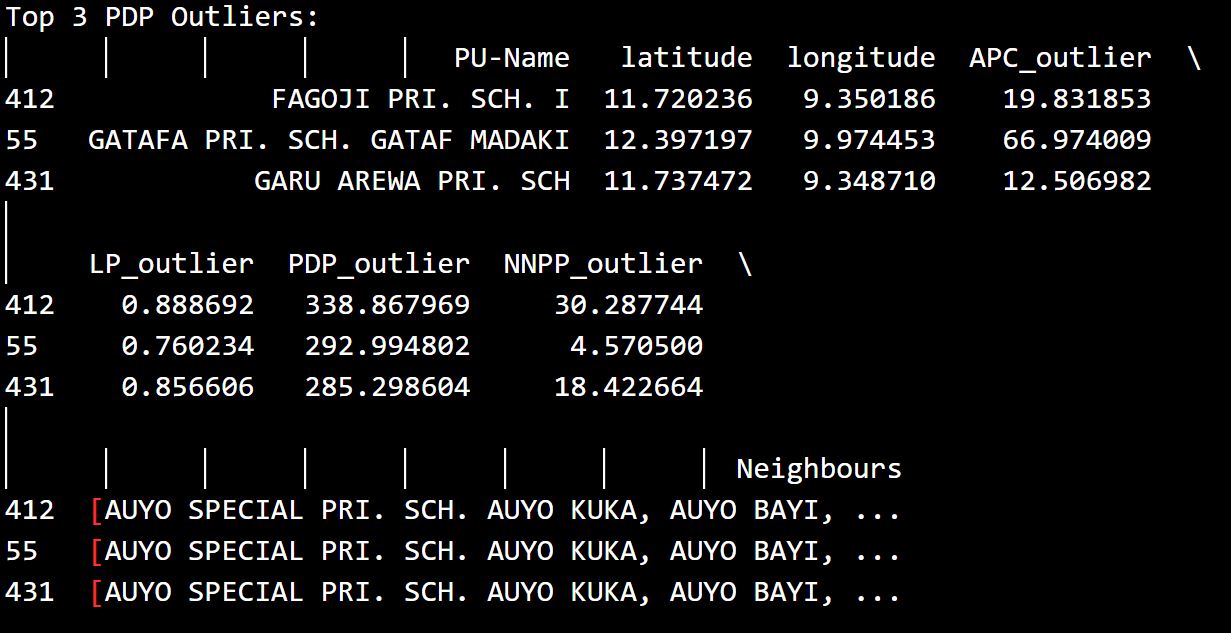
```python

sorted\_pdp = outlier\_scores.sort\_values(by='PDP\_outlier', ascending=False).head(3)

print("Top 3 PDP Outliers:")

print(sorted\_pdp)

```



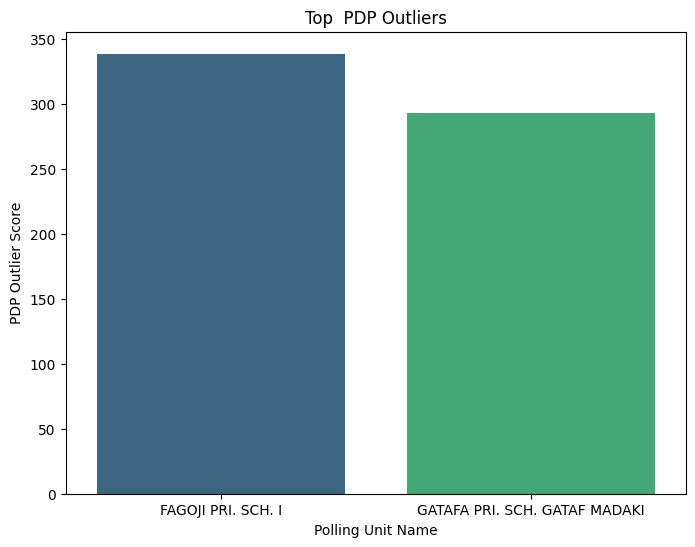


Figure 3: Bar chart showing top outliers for PDP

4. Top 3 NNPP Outliers:

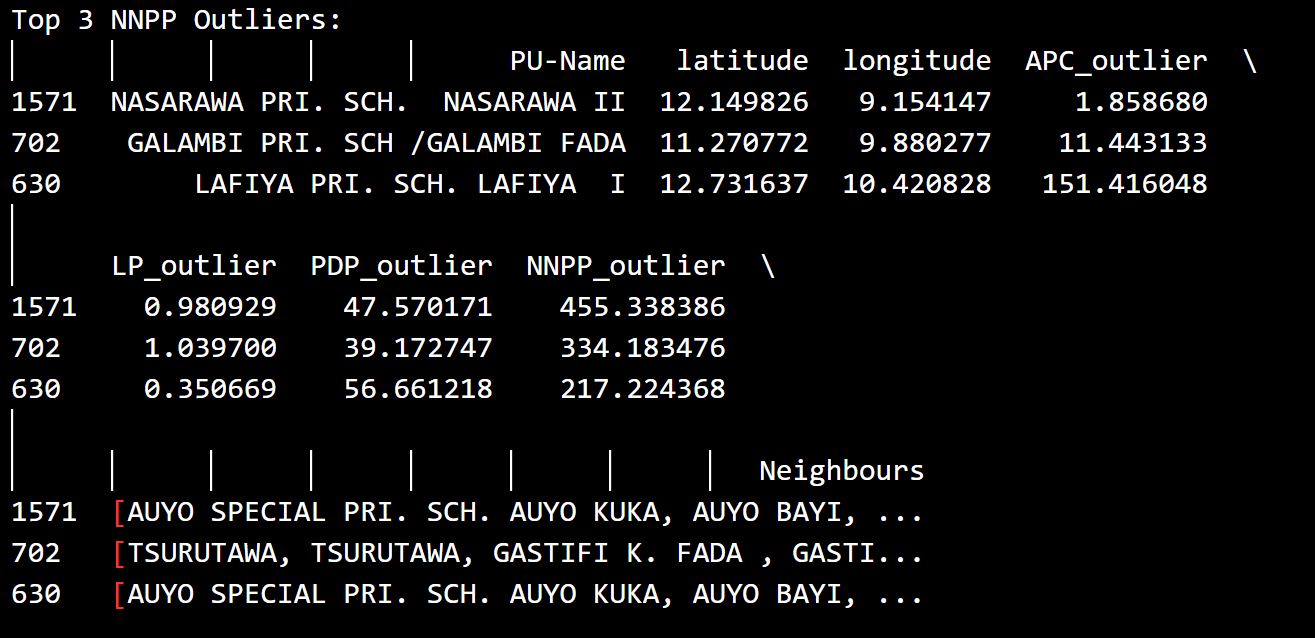
```python

sorted\_nnpp = outlier\_scores.sort\_values(by='NNPP\_outlier', ascending=False).head(3)

print("Top 3 NNPP Outliers:")

print(sorted\_nnpp)

```

E

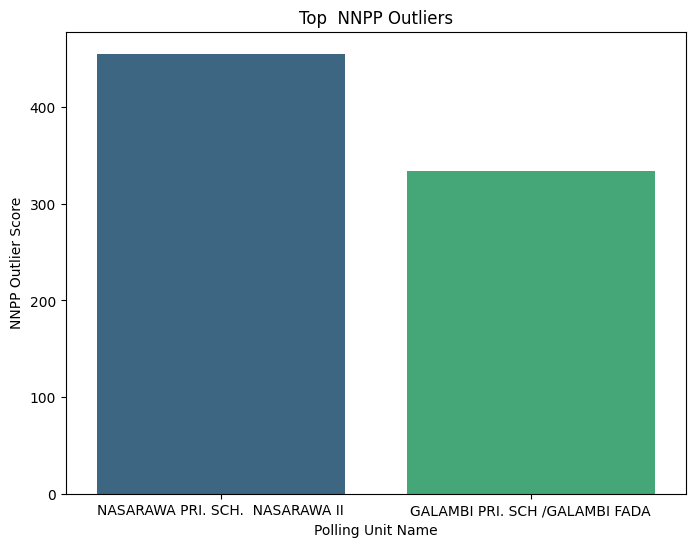


Figure 4: Bar chart showing top outliers for NNPP

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.