Insights into India's Political Landscape: Analyzing Violence Hotspots

Authors: Sindhuja Baikadi, Veda Sahaja Bandi

The Issues:

The main issue addressed in this analysis is the identification of patterns or clusters in incidents of political violence within India, excluding islands, Kashmir, and Ladakh. The objective is to explore how and why these incidents cluster geographically. Analyzing political violence incidents in India using clustering algorithm poses several challenges, prompting us to ask critical questions such as:

- What is the most appropriate number of clusters in this specific context, and what criteria or methods we employ to determine this?
- Can the clustering results unveil significant patterns or trends in the distribution of political violence incidents in India? If so, what might these patterns indicate regarding the underlying causes or factors contributing to political violence?
- How does the geographical distribution of clusters correlate with known socio-political or economic factors within the regions of interest? Are there external influences or historical contexts that could explain the formation of these clusters?

Addressing these questions is crucial for understanding the spatial distribution of political violence incidents in India comprehensively. Through this analysis, we aim to uncover patterns that could guide policy decisions, enhance security measures, and inspire further research into the dynamics of political violence.

Findings

The findings of the analysis on the political violence in India, specifically excluding islands, Kashmir, and Ladakh, have revealed insightful patterns in the spatial distribution of incidents. The analysis successfully identified four main clusters of political violence incidents across the studied regions of India, indicating significant regional differences in the frequency and nature of these incidents. These findings likely highlight the geographical patterns of political violence within the specified regions of India, with clustering potentially revealing significant concentrations of incidents in specific areas.

Visualization of the clusters on a map highlighted the geographical patterns of political violence, particularly in the North and East regions of India. States like Punjab, Haryana (including New Delhi), West Bengal, Chattisgarh, and Meghalaya emerged as hotspots for political violence, attributed to factors such as political unrest, socio-economic disparities, or historical conflicts. The analysis also hinted at differing patterns of political violence between urban and rural settings, with urban areas being focal points for such incidents due to higher population density, greater political activity, or socio-economic tensions.

The variance within clusters and their geographical spread suggests that political violence in India is influenced by a complex interplay of factors, including socio-economic conditions, political rivalries, historical grievances, and other regional specificities. These findings emphasize the importance of targeted interventions and policies addressing the root causes of political violence. By understanding the spatial dynamics and underlying factors contributing to these clusters, policymakers and researchers can better design strategies to prevent and mitigate political violence, contributing to more peaceful and stable socio-political environments.

Overall, the clustering analysis provides a foundational step toward understanding the regional characteristics and dynamics of political violence in India. It highlights the necessity for further in-depth studies to explore the causes and implications of these patterns, aiming to contribute to more peaceful and stable socio-political environments.

Discussion

The clustering analysis of political violence incidents in India from January 1, 2016, through January 1, 2024, reveals distinct spatial patterns and potential regional disparities. The clustering highlights that incidents are concentrated in specific areas rather than randomly distributed, prompting inquiries into the underlying factors shaping their distribution across regions.

The identified clusters suggest potential correlations with external variables such as urbanization, crime rates, and socio-economic status. Moreover, the high variance within clusters indicates disparities in incident spread within each region, hinting at the presence of distinct subregions with unique characteristics. However, the unsupervised nature of the algorithm limits its ability to establish causation or offer deeper insights into the underlying reasons for these spatial patterns.

Further research is needed to understand the complex interplay of variables contributing to these patterns. Areas for exploration include examining law enforcement protocols, assessing community policing effectiveness, evaluating regional gun laws, and understanding historical socio-political contexts. Despite limitations, the findings underscore the need for nuanced approaches to address spatial patterns of political violence in India.

Appendix A: Method

The dataset was obtained in comma-separated values (.csv) file format and was subsequently imported into a Jupyter Notebook for analysis with no null values. The focal point of this analysis was the dataset containing incidents of political violence within India. The dataset primarily included geographical coordinates (latitude and longitude) for each incident, which are crucial for spatial analysis.

The K-Means clustering algorithm, as implemented in the sci-kit-learn Python library, was selected for its effectiveness in identifying geographical patterns and distributions. To determine the optimal number of clusters for this dataset, the Elbow Method was employed. This method involves plotting the Within-Cluster Sum of Squares (WCSS) against a range of possible cluster counts (from 1 to 10) and identifying the point where the reduction in WCSS becomes less pronounced, indicating the most suitable number of clusters.

Four clusters were selected as the optimal number to strike a balance between simplicity and explanatory power, offering a meaningful interpretation of the data without overfitting or oversimplification. This step involved assigning each incident in the dataset to one of the four clusters, based on the similarity of their geographical locations, thereby grouping incidents that are geographically close to each other.

To visualize the clustering results, the Folium library, which is specifically designed for handling and visualizing geospatial data, was utilized. An interactive map was created to display the clustered incidents, with each cluster represented by a distinct color. This visualization provides a clear and intuitive understanding of the geographical distribution and concentration of political violence incidents across the studied regions of India.

The implementation of K-Means clustering, combined with effective data visualization, provided valuable insights into the geographical distribution of political violence in India, thereby contributing to a deeper understanding of the dynamics of such incidents.

Appendix B: Results

The CSV file contains 36383 data points detailing incidents of political violence in India, focusing on the geographical coordinates of each event. Following the initial data cleaning procedures, records with missing values in the 'latitude' and 'longitude' columns, essential for spatial analysis, were removed from the dataset, ensuring the accuracy of the clustering analysis. However, no null data was found.

The Elbow Method plot (Figure 1) suggests that four clusters provide a balanced representation between the within-cluster sum of squares (WCSS) and the number of clusters, indicating the appropriateness of the chosen clustering approach for the given dataset.

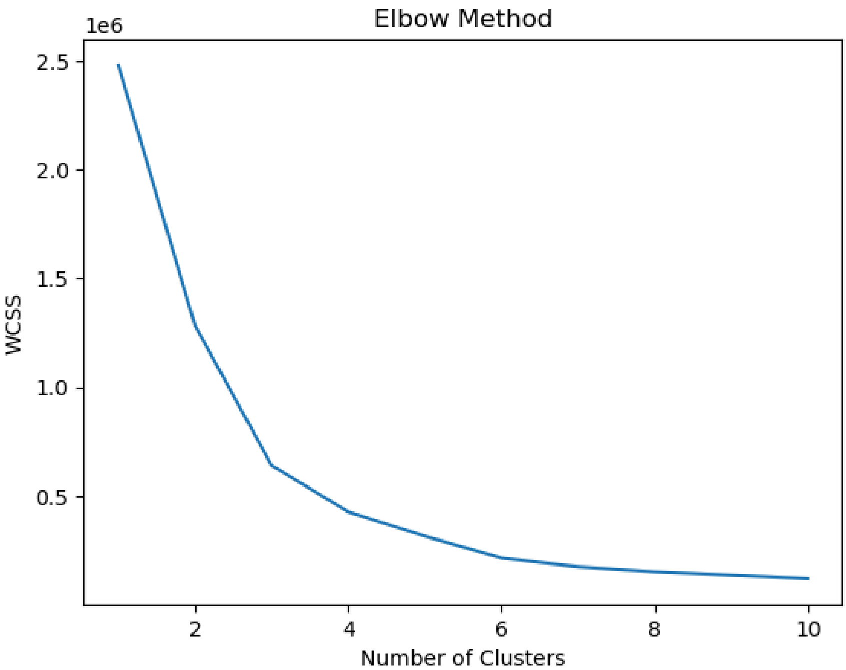

Figure 1: Elbow Method

Through the K-Means algorithm, four distinct clusters were identified within the dataset, each representing different geographic regions within the specified areas of India. These clusters were visualized on an interactive map (Figure 2), with each cluster depicted in different colors: blue (Cluster 1), red (Cluster 2), green (Cluster 3), and yellow (Cluster 4). This differentiation allows for an intuitive understanding of the geographical patterns of political violence, indicating specific areas with heightened activity.

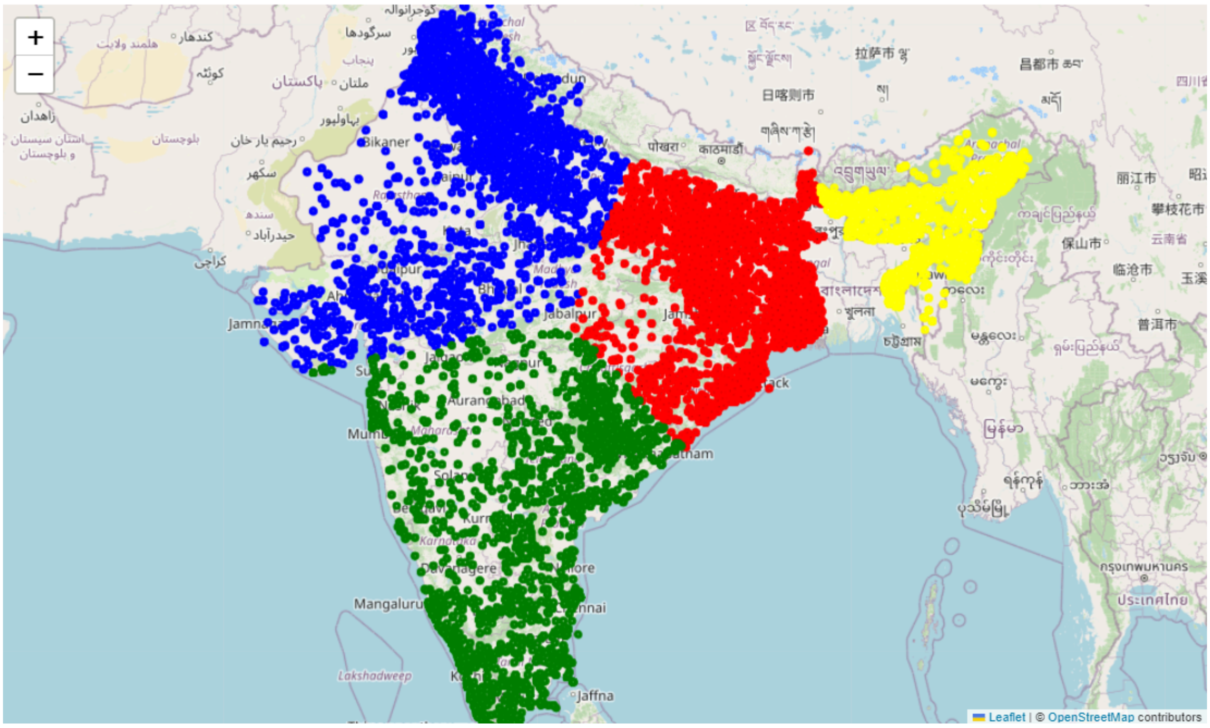

Figure 2: Interactive map with the clusters

The map visualization, enhanced by the color-coded clusters, offers a comprehensive overview of the geographical distribution of incidents across India. Particularly notable are the pronounced concentrations of incidents observed in the North and East regions. This spatial clustering suggests potential regional disparities are due to various factors including political unrest, socio-economic disparities, or historical conflicts, influencing the frequency and distribution of such incidents.

The inertia value of 426895 resulting from the K-Means clustering indicates a degree of cohesion within the clusters. However, the spread within each cluster suggests that incidents are geographically dispersed rather than tightly localized. Despite this dispersion, the clusters highlight distinct patterns of political violence across different regions of India.

The visualization of data points on the interactive map (Figure 3) provides valuable insight into the spatial distribution of political violence incidents, showcasing predominant clusters in states such as Punjab, Haryana (including New Delhi), West Bengal, Chattisgarh, and Meghalaya. These spatial patterns further suggest potential regional disparities in the frequency of political violence, potentially attributed to factors such as higher population density, greater political activity, or socio-economic tensions.

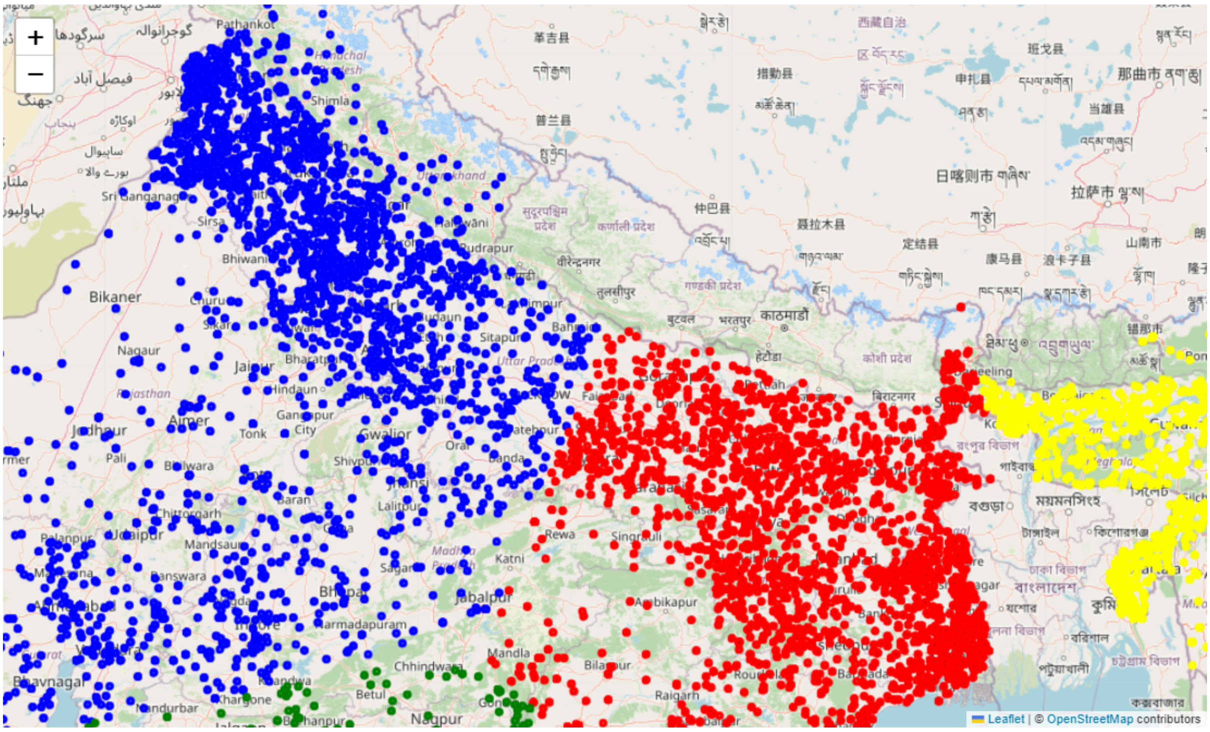

Figure 3: Interactive map with the clusters located at focal points

While the results of the clustering process delineate spatial groupings of political violence, they do not directly illuminate the causative factors behind these incidents. The identified patterns necessitate further investigation into socio-economic, demographic, and political factors to unravel the underlying causes of these regional trends.

In summary, the clustering analysis of political violence incidents in India lays the groundwork for understanding the spatial characteristics of such events within the country. These findings serve as a basis for more in-depth studies into the dynamics influencing these patterns, with implications for policy development, conflict resolution strategies, and peacebuilding efforts.

Appendix C: Code

In this appendix, we document the Python code for performing K-Means Clustering on Police Violations in India.

```
# Importing the libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

import folium

from geopy.distance import geodesic
```

```
from sklearn.cluster import KMeans
```

```
# K-Means Clustering
model = KMeans(n_clusters=4)
y_kmeans = model.fit_predict(df)
df = pd.concat([df, pd.DataFrame(y_kmeans, columns=["y"])], axis=1)
model.inertia_
```

Output

426895.7235738165

```
# Elbow Method and Plot
wcss = []
for i in range(1,11):
    model = KMeans(n_clusters=i, n_init=10)
    y_kmeans = model.fit_predict(df)
    wcss.append(model.inertia_)
plt.plot(range(1,11), wcss)
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.title('Elbow Method')
plt.show()
```

```
# Interactive Map with K-Means Clustering
cluster1 = df[['latitude','longitude']][df['y']==0].values.tolist()
cluster2 = df[['latitude','longitude']][df['y']==1].values.tolist()
cluster3 = df[['latitude','longitude']][df['y']==2].values.tolist()
cluster4 = df[['latitude','longitude']][df['y']==3].values.tolist()
map = folium.Map(location=[data['latitude'].iloc[0],data['longitude'].iloc[0]], zoom_start = 10, tiles="openstreetmap")
for i in cluster1:
    folium.CircleMarker(i, radius=2, color='blue', fill_color='lightblue').add_to(map)
for i in cluster2:
    folium.CircleMarker(i, radius=2, color='red', fill_color='lightred').add_to(map)
for i in cluster3:
```

```
    folium.CircleMarker(i, radius=2, color='green', fill_color='lightgreen').add_to(map)

for i in cluster4:
    folium.CircleMarker(i, radius=2, color='yellow', fill_color='lightyellow').add_to(map)

map
```

Contribution:

Sindhuja Baikadi - 02128756: Worked on the Issues, Findings, Discussion, Method, and Results sections. Also self-plotted the graphs to analyze the data using the various methods discussed in the report.

Veda Sahaja Bandi - 02105111: Outlining key observations and insights, worked on the coding portion of the project to implement necessary functionalities and features