**Identifying Critical Accident Hotspots: Why DBSCAN Outperforms KDE for Traffic Safety Planning**

**Introduction**  
Road traffic accidents are a global public safety challenge, resulting in significant loss of life and economic burden. Identifying accident hotspots is crucial for designing effective interventions to enhance road safety. This study applies the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to traffic accident data from January 1, 2022, to September 30, 2024, aiming to locate critical accident hotspots. By focusing on clustering high-density accident regions, DBSCAN provides actionable insights for resource-constrained safety planning.

Traditional methods such as Kernel Density Estimation (KDE) are often employed for spatial analysis of accident data. While KDE provides a useful visualization of density patterns, it has notable limitations in identifying discrete clusters and managing noise. DBSCAN addresses these gaps by offering a density-based approach that identifies clusters of arbitrary shapes and excludes noise points, making it particularly effective for spatially heterogeneous data, such as traffic accidents.

**Theoretical Overview: DBSCAN vs. Kernel Density Estimation (KDE)**

1. **Kernel Density Estimation (KDE): Overview**  
   KDE is a statistical method used to estimate the probability density function of spatial data. It generates a continuous surface (heatmap) by summing kernel functions centered at each data point, with the bandwidth controlling the extent of smoothing.

**Advantages of KDE:**

* + Produces visually intuitive heatmaps for identifying general density trends.
  + Effective for exploratory spatial analysis across large regions.

**Limitations of KDE:**

* + Highly sensitive to bandwidth selection, which can lead to over-smoothing or under-smoothing of data.
  + Does not differentiate between significant clusters and random noise, often diluting meaningful patterns.
  + Lacks the ability to define distinct cluster boundaries, which can make actionable hotspot identification challenging.

1. **DBSCAN: Overview**  
   DBSCAN is a clustering algorithm designed to group points into high-density clusters based on two key parameters:
   * **eps (ε):** The maximum distance between points for them to be considered part of the same cluster.
   * **minPts:** The minimum number of points required to form a cluster.

**Advantages of DBSCAN:**

* + Identifies clusters of arbitrary shapes, which is crucial for spatial data such as road networks.
  + Effectively handles noise by treating low-density points as outliers.
  + Does not require predefining the number of clusters, unlike many other clustering algorithms.
  + More actionable for resource allocation, as it highlights specific high-risk zones rather than generalized density patterns.

**Why DBSCAN is Better Suited for Accident Hotspot Analysis**  
While KDE is effective for visualizing general density trends, its inability to handle noise and identify discrete clusters makes it less suited for targeted safety planning. In contrast, DBSCAN aligns with real-world needs by explicitly identifying localized clusters and excluding random outliers, ensuring that interventions are focused on areas with significant risk.

**Methodology**

**Dataset Description**  
The dataset consists of 7,235 traffic accidents recorded over a 33-month period. Each accident entry includes geographic coordinates (latitude and longitude), time of occurrence, and severity classification (e.g., "Light" or "Severe/Fatal"). Two subsets were analyzed:

1. **All accidents:** Includes the full dataset.
2. **Severe/Fatal accidents:** Focuses on incidents with higher severity, requiring immediate intervention.

**Data Preprocessing**  
The dataset was cleaned to remove missing or inaccurate data points. Accident locations were converted into a geographic coordinate matrix for clustering. Outliers unrelated to meaningful clusters (e.g., accidents in isolated areas) were flagged for exclusion.

**DBSCAN Parameters**  
The DBSCAN algorithm was implemented using the following parameters:

* **eps (ε):** Set to 0.001 degrees (~100 meters) to capture localized hotspots such as intersections or road segments.
* **minPts:** Set to 10, ensuring clusters represent meaningful concentrations of accidents rather than random occurrences.

**App Description: DBSCAN Clustering for Accident Hotspots**

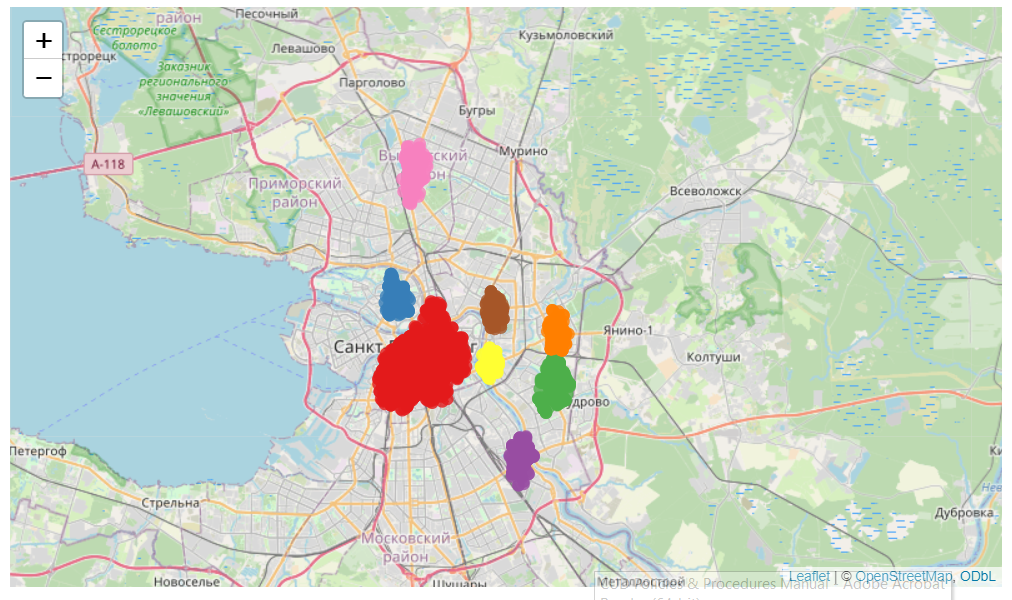
To make the analysis accessible and interactive, the **DBSCAN Clustering for Accident Hotspots** app was developed. This user-friendly tool allows for real-time exploration of accident clusters using adjustable DBSCAN parameters.

**Key Features:**

1. **Dynamic Parameter Adjustment:**
   * **Epsilon (ε):** Users can adjust the radius from 0.0005 to 0.005 degrees (~50 to 500 meters).
   * **Minimum Points (minPts):** Values can be selected between 5 and 15, allowing for fine-tuned cluster detection.
2. **Dual-Tab Clustering Analysis:**
   * **All Accidents:** Displays clusters for all recorded incidents.
   * **Severe/Fatal Accidents:** Focuses on high-severity incidents for targeted analysis.
3. **Interactive Mapping:**
   * Clusters are displayed with distinct colors using Leaflet for easy identification.
   * Users can interact with the map, zooming and panning to explore specific clusters.
   * Pop-ups provide cluster-specific information, such as cluster ID and location details.
4. **Noise Management:**
   * Noise points are excluded from visualizations, ensuring the focus remains on significant hotspots.

**Results and Insights**

DBSCAN successfully identified distinct clusters of high accident density, with noise points effectively excluded. The clusters often corresponded to high-risk areas, such as intersections, curved road segments, or stretches of highways with heavy traffic.



Eight clusters were identified. *Specific examples based on the map to be added*

These clusters align closely with areas requiring urgent interventions, such as improved signage, road redesign, or targeted law enforcement.

**Conclusion**  
This study demonstrates the superiority of DBSCAN for accident hotspot analysis, particularly in scenarios requiring precise cluster identification and the exclusion of noise. Unlike KDE, which smooths data indiscriminately, DBSCAN’s density-based approach ensures that resources are focused on critical, high-density areas. By tuning parameters to reflect local accident patterns, the analysis provides actionable insights for safety planning.

Future work could explore the integration of additional data sources, such as traffic volume, weather conditions, and road quality, to further refine clustering results. DBSCAN’s flexibility and effectiveness make it a valuable tool for data-driven traffic safety interventions.