Hi Joseph,

For hierarchical clustering, your idea of grouping crashes by severity, time, and weather conditions could be beneficial for transportation planning, like identifying that 45% of nighttime accidents happen during rain in certain zones (Kumar & Toshniwal, 2016).

The way you used K-means to classify road segments by crash behavior was also well put. If we have data on 500 road segments, we could group them into rural high-speed (200), urban low-speed (250), and mixed-type roads (50). This could help the transport authority focus its safety efforts more effectively (Puspitasari et al., 2020).

Your DBSCAN example was also impressive. Using it to detect hotspots near schools or dark intersections could support smart city planning (Nieto, 2018). Do you think DBSCAN might miss smaller clusters if there are not enough points close together, especially in less populated areas?

Thank you for providing clarity in your post, in a real dataset. It made the application of these algorithms much clearer!

All The Best!

Avinash

References

Kumar, S., & Toshniwal, D. (2016). Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC). *Journal of Big Data, 3*(13). https://doi.org/10.1186/s40537-016-0046-3

Nieto, A. (2018, February 15). *Density-based clustering: Exploring fatal car accident data to find systemic problems*. ArcGIS Blog. Retrieved April 18, 2025, from https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/density-based-clustering-exploring-fatal-car-accident-data-to-find-systemic-problems

Puspitasari, D., Wahyudi, M., Rizaldi, M., Nurhadi, A., Ramanda, K., & Sumanto. (2020). K-means algorithm for clustering the location of accident-prone on the highway. *Journal of Physics: Conference Series, 1641*(1), 012086. https://doi.org/10.1088/1742-6596/1641/1/012086