RISK RATING FROM ROAD TRAFFIC ACCIDENT FATALITIES FOR THE WORLD INSURANCE SECTOR

Mehmet Niyazi ÇANKAYA, PhD in Statistics,
Uşak University
E-mail: mehmet.cankaya@usak.edu.tr
ORCID: https://orcid.org/0000-0002-2933-857X

Ercan ÖZEN, PhD in Economics,
Uşak University, Faculty of Applied Science
E-mail: ercan.ozen@usak.edu.tr
ORCID: https://orcid.org/0000-0002-7774-5153

DOI: https://doi.org/10.36004/nier.cecg.III.2024.18.04

Abstract. *Traffic insurance is a system that compensates the insured for various damages that may arise due to accidents. Insurance companies finance these damages with the premiums received from the insured. Suitable insurance premium tariff. It keeps the customer demand in optimum balance and at the same time affects the financial success of the insurance company. Therefore, there is a need for optimal pricing. In order for insurance companies to establish the optimal risk-pricing balance, they need to rate the risks. Rating the mortality risks due to traffic accidents also allows insurance companies to price appropriate premiums. For this reason, the aim of this study is to rate the global traffic accident related mortality risks by countries. This study examines the global distribution of road traffic mortality by analyzing data from 231 countries using the DBSCAN algorithm. By classifying the number of deaths per 100,000 population, we identify patterns of similarity and show that countries can be grouped into 27, 8 or even as few as 7 distinct classes, depending on the tuning parameters of the DBSCAN algorithm. These results suggest that, despite the large number of countries, road traffic fatalities show similar patterns that can be attributed primarily to human factors. The classification highlights the importance of focusing on vehicle and driver-related issues rather than infrastructure, which appears to be less of a differentiating factor between countries. This findings have important implications for policy makers and insurance companies aiming to reduce the number of road deaths through targeted interventions.*

Key-words: *Classification, Transportation, Insurance, Government Expenditures statistics*
JEL: *G22, H51, R40, R41*
UDC: *364.322*

Introduction. Road traffic accidents are one of the world's leading public health problems. Large cities and heavy traffic can trigger fatal accidents. These accidents have serious economic and social burdens on individuals and society (Özen et. al., 2014). The insurance system offers different types of policies taking into account these risks and premiums are calculated based on these risks. Accurate rating of the mortality risks arising from traffic accidents enables insurance premiums to be determined more appropriately. Insurance premiums are usually calculated by looking at how much risk there is. In traffic accidents, these risks are calculated based on variables such as drivers' age, gender, alcohol use and accident history, but the risk of death is also one of the most important factors. For this reason, the risk of death of drivers is graded in detail. Apart from individual factors, factors that increase the risk of death can be counted as vehicle characteristics, traffic density, road conditions, weather conditions, education level and traffic infrastructure. It is observed that fatal accidents are less in countries with more modern traffic infrastructure (Özen et.al., 2024).

Rating the risks of death due to traffic accidents and including these risks in insurance premium calculations is a multidimensional process. Companies use actuarial and statistical methods to determine mortality risks, which in turn determine insurance premium rates. Accurate determination of mortality risks contributes to the determination of fair premium rates for both drivers and companies.

Optimal determination of risks and adjustment of premiums accordingly may be perceived by drivers as an incentive to be more careful while driving. This, in turn, may lead to a decrease in traffic accidents and thus a decrease in premiums through reinforcing safe driving habits of drivers and harmonisation with technology. Thus, both individual and social benefits can be created.

The rating of mortality-related risks according to countries also supports insurance companies operating both locally and internationally in making decisions on issues such as market entry and pricing. Therefore, the main purpose of this study is to categorise countries worldwide according to their risk of death due to traffic accidents.

Method and Findings. Cluster analysis is a statistical technique used to group data points into clusters, where points within the same cluster are more similar to each other than to those in different clusters. It helps to identify patterns and structures within large data sets, allowing data to be categorised based on inherent similarities (Hennig et al. 2015; Kaufman, and Rousseeuw, 1990; Hartigan, 1975). Clustering is a crucial technique in data analysis and machine learning that lets you group similar data points together. Among the various cluster algorithms, DBSCAN (Density-Based Spatial Cluster of Applications with Noise) is notable for its ability to detect clusters of any shape and to be robust to noise and outliers. DBSCAN was developed in 1996 by Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu. It has become a widely used method in applications ranging from geographic data analysis to image processing.

In this study, we try to determine the optimal number of clusters according to the magnitude of the number of mortalities caused by traffic accidents or injuries throughout the world.

Section 1: Research methodology and data set. DBSCAN is a density-based clustering algorithm that groups data points into clusters based on the density of their local neighborhoods. The algorithm is based on two key parameters:

1. Eps (ε): The maximum distance between two points to be considered as neighbours.
2. MinPts: The minimum number of points required to form a dense region (cluster).

Based on these parameters, DBSCAN classifies data points into three categories: Core Points: Points that have at least \`MinPts\` within a distance of \`eps\`; Boundary Points: Points that are within \`eps\` distance of a core point, but do not have enough neighbours to be a core point themselves; Noise Points: Points that are neither core nor boundary points and do not fit into any cluster.

Section 1.1: The DBSCAN Algorithm Steps.

1. Start with an unvisited point: Randomly select an unvisited point and determine if it is a core point.
2. Expand the cluster: If the point is a core point, a cluster is formed. The algorithm then iteratively adds all reachable points (i.e. points within \`eps\` of any point in the cluster) to the cluster.
3. Mark noise points: If the point is not a core point and cannot be added to an existing cluster, it is marked as noise.
4. Repeat: The process is repeated until all points have been visited.

Section 1.2: The structure of data set. The dataset used in this analysis comes from the World Bank's official database, specifically from the indicator "Mortality caused by road traffic injury (per 100,000 people)". The data span the years 2000 to 2019 and provide a comprehensive view of road traffic deaths worldwide. This indicator is essential for understanding the impact of road traffic injuries on public health and for assessing the effectiveness of road safety measures over time. The data can be accessed directly through the World Bank's data portal and the missing values are omitted while analyzing. (https://data.worldbank.org/indicator/SH.STA.TRAF.P5).

Section 2: Main results. The application of the DBSCAN algorithm to the analysis of road traffic fatalities in 231 countries provides significant insights into the global distribution of these events. The results, which vary according to the tuning parameters used, demonstrate the algorithm's ability to classify countries into distinct clusters based on the number of road deaths per 100,000 population (Ester et al., 1996; Jain et al. 1999).

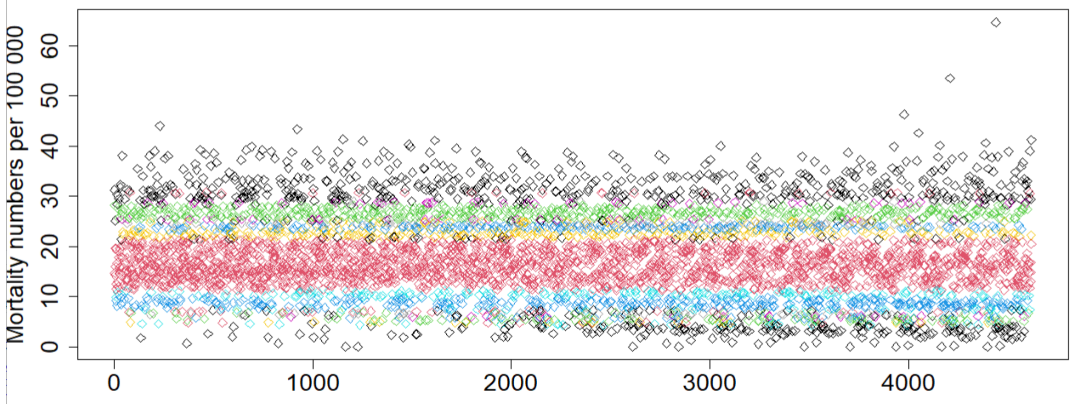

Figure 1. ε, ***eps = 0.1, MinPts = 23, the number of clusters: 27***

Figure 1. shows the results when the DBSCAN parameters are set to ($\varepsilon = 0.1$ and ($\mathbf{MinPts} = 23$). Under these conditions, the data is grouped into 27 distinct clusters. This relatively large number of clusters suggests a more detailed differentiation between countries where fatality rates show subtle differences between regions.

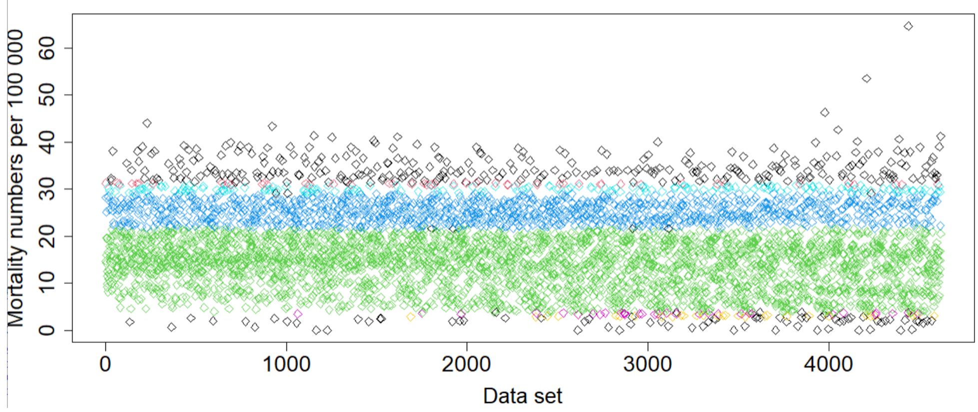

Figure 2. **ε , eps = 0.15, MinPts = 23, the number of clusters: 8**

Figure 2. shows the result of increasing $\varepsilon = 0.15$ while keeping ($\mathbf{MinPts} = 23$). The number of clusters is reduced to 8, indicating that the higher ε value leads to a broader grouping, with countries with moderately similar fatality rates being grouped into fewer clusters. This aggregation reflects a higher tolerance for differences in the number of fatalities, emphasising broader regional similarities.

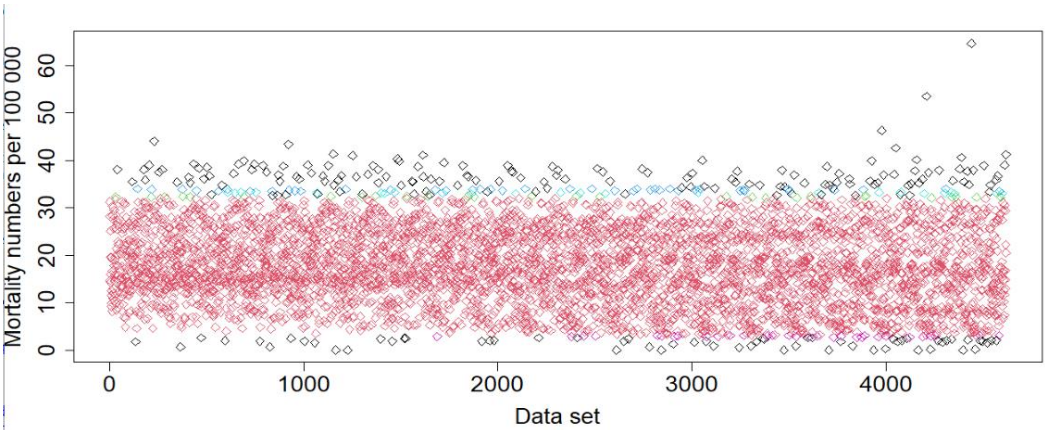

Figure 3. ***ε*, eps = 0.2, MinPts = 23, the number of clusters: 7**

Figure 3. shows the scenario where $\varepsilon = 0.2$ and $(\mathbf{MinPts} = 23)$. Here the data points are clustered into only 7 groups. This further reduction in the number of clusters suggests that as ε increases, the algorithm identifies even broader patterns of similarity between countries, likely driven by dominant human factors rather than regional or infrastructural differences.

These results highlight the ability of DBSCAN to reveal varying degrees of similarity between countries based on road traffic mortality rates. The identification of only 7 clusters across 231 countries underlines the commonality of the underlying factors contributing to road crashes, mainly human factors such as driver behavior and vehicle conditions. The findings suggest that these factors transcend geographical and infrastructural differences, highlighting the need for global strategies focused on improving driver safety and vehicle standards to effectively reduce road deaths worldwide.

Table 1. **The R codes and explanations**

```
# Install and load necessary libraries
install.packages("fpc")
library(fpc)
# Perform DBSCAN clustering
dbscan_result <- dbscan(data_long$Value, eps = 0.1, MinPts = 23) #MinPts = 15:
Specifies the minimum number of points required to form a dense region, which becomes a core
point of a cluster.
dbscan_result$eps
# Plot the clusters
plot(data_long$Value, col = dbscan_result$cluster + 1, pch = 5, cex = 1,
xlab="Data set",ylab="Mortality numbers per 100 000",cex.lab=1.6, cex.axis=1.6)
# Number of unique clusters (including noise as a separate color)
num_colors <- length(unique(dbscan_result$cluster)) + 1 # +1 for noise
```

In Table 1, code snippet performs DBSCAN clustering on a dataset, visualises the results by plotting the clusters, and calculates the number of unique clusters (including noise). DBSCAN is a useful algorithm for identifying clusters of different shapes and sizes in data, especially when noise or outliers are present (Hennig,2015).

In 231 countries, when the number of traffic accident fatalities in 100 000 people in the population is classified within themselves, it is seen that those who show similar numbers or those who give the same numerical pattern can find themselves in 27, 8 and 7 classes according to different tuning parameter values in the DBSCAN algorithm. This result shows that although there are 231 countries in the world in terms of the number of deaths caused by traffic accidents on the basis of numerical magnitudes, it can be observed that the number of deaths in traffic accidents is similar to 27, 8 and even 7. This shows that accidents are generally caused by human phenomenon. We can classify this phenomenon as an infrastructure, two vehicles and a driver. In this sense, it is important for insurance companies to focus especially on the vehicle and driver situation; because the fact that the number of accidents can be reduced to 7 types from 231, countries analyzed in terms of the number of accidents can also be concluded that there are no infrastructure problems or that the infrastructures can be homogeneous between countries.

Discussion and conclusions. The analysis shows that road deaths in 231 countries can be categorized into a surprisingly small number of groups, suggesting that similar factors are at work around the world. The reduction to just 7 classes suggests that human factors, such as driver behavior and vehicle condition, are the main contributors to road crashes, rather than differences in infrastructure. This finding is significant because it shifts the focus of accident prevention strategies towards addressing human factors, which are more manageable and consistent across different regions. For insurance companies, the study highlights the importance of tailoring policies and prevention measures to driver and vehicle-related risks. By focusing on these areas, insurers will be able to develop more effective strategies for the reduction of road traffic fatalities. In addition, the homogeneity of infrastructure across countries suggests that efforts to improve road safety should be global rather than local, with best practice being shared on an international basis.

General proposals. *Focus on driver and vehicle interventions*: Insurers and policymakers should prioritize initiatives aimed at improving driver behavior and vehicle safety. These could include stricter driver training requirements, regular vehicle inspections and incentives for the use of advanced vehicle safety technologies. *Global cooperation on road safety*: Given the similarity in crash patterns between countries, there is an opportunity for international cooperation to standardize safety practices and share successful interventions. Countries with lower mortality rates can serve as models for others. *Further research on infrastructure homogeneity*: While this study suggests that infrastructure may not be a significant differentiator in road fatalities, further research is needed to confirm this finding. If infrastructure is indeed homogeneous across countries, this raises the possibility of global standards in road design and traffic management.

Data-driven policy-making: Governments should continuously monitor and classify road crashes using data analytics and machine learning techniques such as DBSCAN. This will enable dynamic and responsive policy-making that adapts to emerging trends and patterns in road safety.

REFERENCES

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Vol. 96(34), pp. 226-231). https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf

Hartigan, J. A. (1975). *Clustering Algorithms*. Michigan: Wiley.

Hennig, C. (2015). *Fpc: Flexible Procedures for Clustering*. *R package version 2.2-9.* https://data.worldbank.org/indicator/SH.STA.TRAF.P5

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2015). *Handbook of Cluster Analysis*. New York: CRC Press. https://doi.org/10.1201/b19706

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*(3), 264-323. https://doi.org/10.1145/331499.331504

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Antwerpen, Belgia: Wiley. https://doi.org/10.1002/9780470316801

Ozen, E., Genc, E., & Kaya, Z. (2014). Estimation of the costs of traffic accidents in turkey: An evaluation in terms of the insurance and financial system. *Yaşar Üniversitesi E-Dergisi*, *9*(33), 5649-5673. https://doi.org/10.19168/jyu.94397