

Temporal Analysis of Daily Fatalities Using K-Means Clustering Method

J. S. Ho^{*1}, N. A. Mhd Yunin¹, D. W. Kak¹, Y. Y. Wu² and S. A. Syed Mohamed Rahim¹

¹ Malaysian Institute of Road Safety Research (MIROS), 43000 Kajang, Selangor, Malaysia

² School of Business and Economics, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

*Corresponding author: jsho@miros.gov.my

ORIGINAL ARTICLE

Open Access

Article History:

Received
28 Oct 2025

Accepted
15 Feb 2026

Available online
1 May 2026

ABSTRACT – Malaysians make long-distance travel during festive seasons, when traffic increases by approximately 30%. The high exposure on the roads inevitably increases the risk of traffic crashes. By using daily traffic fatalities between 18th March 2025 to 25th May 2025, this study examines the effect of different types of holidays on traffic fatalities. Holidays were classified into three categories: school holidays, public holidays, and seasonal holidays. The K-Means clustering method was employed for this analysis, and the results revealed that the type of holiday played a significant role in influencing the traffic fatality patterns. Three distinct clusters were deduced in this analysis. The highest risks took place in early June during the seasonal holidays, while the lowest fatalities, with an average of 12.22 fatalities, were observed around September. Moderate risks were reported in the mid of March 2025. Understanding the effect of seasonal variation on traffic fatalities enables relevant government agencies to develop targeted interventions during the respective seasons. As this study only considered limited variables such as day of week and type of holiday, future research can incorporate additional factors such as weather conditions and potential spatial correlations to enhance the predictive insights.

KEYWORDS: Festive seasonal effects, K-Means clustering, traffic fatalities

Copyright © 2026 Society of Automotive Engineers Malaysia - All rights reserved.
Journal homepage: www.jsaem.my

1. INTRODUCTION

Malaysia, a unique multiracial country with a population of 33 million, experiences intense travel activities during festive seasons and long holidays. People from all walks of life often seize these opportunities to travel. According to Low et al. (2017), road traffic volume increases by an average of 30% during festive seasons. This surge in traffic inevitably raises the risk of road crashes, as exposure is a significant contributing factor (Claros et al., 2020). In Malaysia, road traffic crashes are the fourth leading cause of death, with approximately 6,000 to 7,000 fatalities reported annually (Radzuan et al., 2020; PDRM, 2023).

Recognizing the profound societal impacts of motor vehicle fatalities, it is imperative to enhance road safety and work toward the goal of halving traffic fatalities by 2030. Crash prediction models play a critical role in formulating effective safety policies. These models typically rely on crash data aggregated over time (temporal) and space (spatial), both of which are often influenced by unobserved heterogeneity (Kamel & Sayed, 2021).

Despite significant advancements in crash prediction and time series modeling, most studies have focused on monthly or annual data, neglecting the dynamic nature of daily fatal crash occurrences. Failure to account for unobserved heterogeneity can lead to biased parameter estimates and inaccurate predictions (Mannering et al., 2016). Studies have shown that models incorporating spatial and temporal effects perform better than those that do not, as they address this heterogeneity.

Spatial correlations may exist due to similar characteristics in neighboring zones (e.g., infrastructure, topography, socioeconomic factors), resulting in clusters of similar crash frequencies (El-Basyouny & Sayed, 2009; El-Basyouny et al., 2014). Similarly, temporal correlation occurs when crash observations are influenced by seasonal factors such as holidays, school schedules, and weather variations, all of which affect travel behavior (Böcker et al., 2013; Kamel & Sayed, 2021).

While temporal and spatial variations are critical for accurate forecasting, crash analyses are often limited to yearly aggregates due to the unavailability of disaggregated data, resulting in information loss and increased unobserved heterogeneity (Mannering et al., 2016; Claros et al., 2020). Therefore, understanding daily crash trends is crucial for building robust predictive models that reflect the dynamic nature of contributing factors and help improve road safety.

To fill this research gap, the present study applies K-Means Clustering to analyze the seasonal dynamics in daily traffic fatalities, thereby offering new insights into how holidays affect fatal crashes. The structure of this manuscript is arranged in the following.

2. METHODOLOGY

Since 15 March 2024, the Royal Malaysia Traffic Police have begun compiling daily traffic crash and fatality data (NST, 2024). The Federal Traffic Investigation and Enforcement Department releases this data daily, usually before 2:00 PM. This study uses secondary data on daily and weekly traffic fatalities collected between April 2024 and May 2025, comprising 418 daily records. Figure 1 displays the daily time series of traffic fatalities, while the variables used in the dataset are summarized in Table 1.

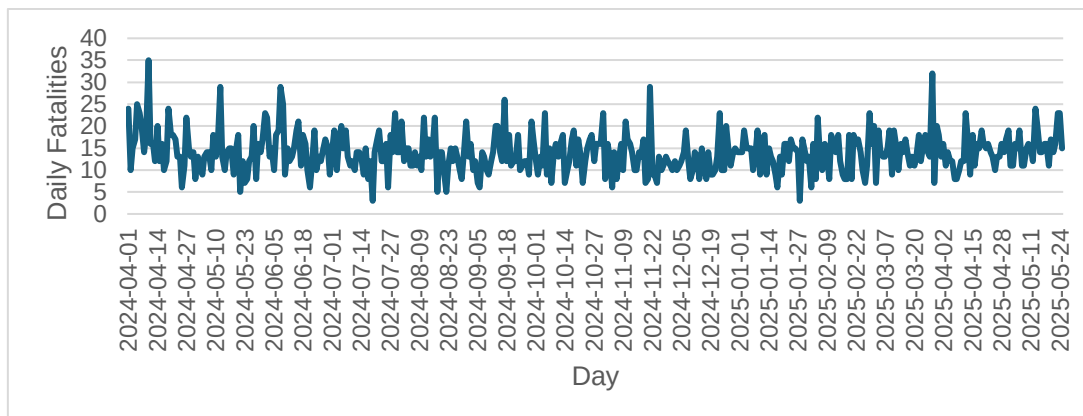


FIGURE 1: Distribution of daily fatalities

TABLE 1: Description of the variables

Variable	Level
Date	-
Fatalities	-
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday
Month	January, February, March, April, May, June, July, August, September, October, November and December
Holiday	Normal days, School holidays, public holidays, and seasonal holiday (School and Public holidays)

2.1 K-Means Clustering Method

This study applies the K-Means clustering method – an unsupervised machine learning algorithm that groups data points with similar patterns into K clusters. Initial centroids are chosen from the dataset, and each cluster's centroid represents the mean of data points within it. The Euclidean Distance is used to determine the distance between each data point and the cluster centroid, as shown in Equation 1:

$$d = |x - y| \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where,

d = calculation of distance to the center of the cluster

x = point coordinates of the data

y = centroid coordinate

$\sum_{i=1}^n$ is the amount of data to be measured the distance, while $i=1$ is the clustering process starting from the first iteration

x_i = coordinate point of the i object

y_i = i centroid coordinate point

In general, there are several steps in the K-Means clustering analysis:

- i. Choose number of clusters (k).
- ii. Randomly initialize k centroids.
- iii. Assign each point to the nearest centroid.
- iv. Recalculate centroids based on the mean of points in each cluster.
- v. Repeat steps 3-4 until centroids do not change (convergence).

3. RESULTS

In this study, daily traffic fatalities from 1st April 2025 to 25th May 2026 are used to examine the temporal seasonal patterns. A total of 6,061 fatalities were reported during the period. The data were first checked for stationarity. Stationary data means that the series has a constant mean. Table 2 presents the stationarity check using the Augmented Dickey-Fuller (ADF) test, and the result indicates that this series is stationary at the 1% significance level. Stationary data is important to make result, specifically clustering in this case more sensible.

TABLE 2: Augmented Dickey-Fuller (ADF) test

Variable	ADF
Fatalities	-19.5876***

Note: *** significant at 1%

The next step is to determine the optimal number of clusters. K is selected when adding more clusters does not decrease the clustering performance significantly. Figure 2 illustrates the distribution of K clusters, and the values are noted in Table 3.

After determining the K clusters, the cluster profiling is conducted, and the results are presented in Table 3. Of the clusters, it is shown that the lowest fatalities, with an average of 12.22 fatalities, occur around September and with minimal holidays. The moderate deaths and low holidays fell in the month of March/April (an average month of 3.76). Cluster 2 has the highest average deaths (15.67), which took place around the month of June (6.05) period, which was seasonal holidays (both school and public holidays).

Figure 3 illustrates the distribution of data points resulting from K-Means clustering. The data points are grouped into three clusters, each represented by a different color: purple, green, and yellow. The purple cluster, located in the upper-left quadrant, corresponds to low-risk, normal days. The green cluster represents moderate-risk periods, while the yellow cluster, concentrated in the lower-right quadrant, indicates days with intense holiday activity and high fatality rates. The clear separation among the

clusters suggests that the K-Means clustering algorithm successfully identified meaningful structure in the dataset.

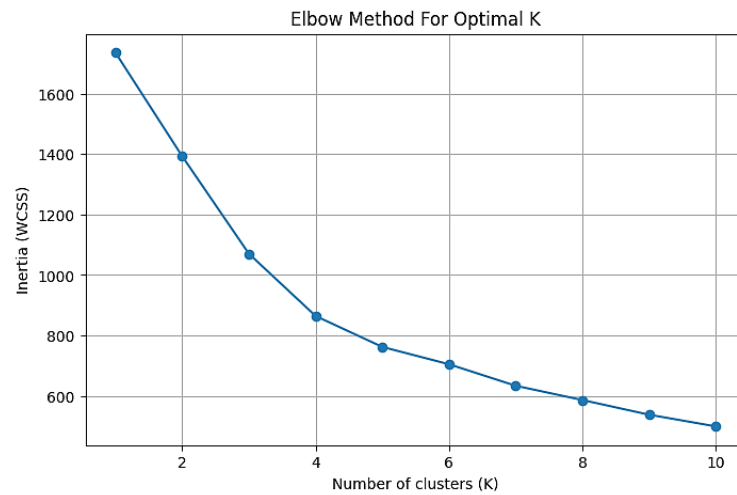


FIGURE 2: Distribution of K clusters

TABLE 3: Silhouette score

K	Silhouette Score
K = 2	0.2660
K = 3	0.2553
K = 4	0.2624
K = 5	0.2634

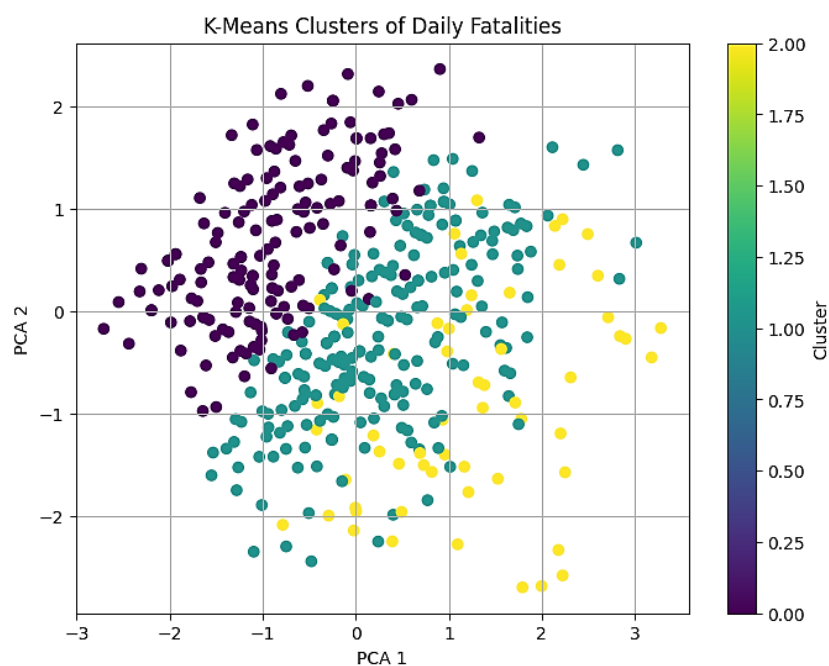


FIGURE 3: Distribution of K clusters

4. DISCUSSION AND CONCLUSION

Malaysia is a unique and vibrant multiracial country, where the cultural richness is reflected in a wide range of religious and traditional celebrations observed throughout the year. These festive periods – such as Chinese New Year, Hari Raya Aidilfitri, Deepavali, and Christmas – are often accompanied by large-scale movement of people as citizens return to their hometowns or travel across states to celebrate with family and friends. Such celebrations, while socially and culturally significant, are typically associated with a surge in traffic volume on the roads, highways, and expressways nationwide. These extensive travel patterns contribute to an increased exposure to road safety risks, particularly over several consecutive days of heightened travel demand.

Statistical records show that on average, between 6,000 and 7,000 people lose their lives on Malaysian roads each year. This persistent and high fatality rate is not only a public health concern but also imposes considerable economic, social, and environmental burdens on the nation. Road traffic deaths result in loss of productivity, emotional trauma for families, costs related to emergency response and medical care, and long-term consequences for community well-being.

Recognizing the seasonal patterns in traffic risks, the Malaysian government has introduced several intervention programs to manage road safety during high-risk periods. One prominent initiative is the Ops Bersepadu program, a coordinated operation involving multiple enforcement and safety agencies – including the Royal Malaysia Police (PDRM) and the Road Transport Department (JPJ). This program is implemented during major festive seasons and is aimed at curbing traffic violations, enhancing visibility of enforcement, and reducing fatal crashes through checkpoints, public awareness campaigns, and compliance inspections.

Despite these efforts, a major shortcoming remains: limited coverage of other high-risk travel periods, particularly school holidays and combinations of public and seasonal breaks, which have been shown to pose equal or even higher risks than traditional festivals. This study identifies a critical gap in the timing and targeting of existing safety interventions.

Building on this observation, the study explores the seasonal and temporal dynamics of traffic fatalities in Malaysia, focusing on how different types of holidays impact daily fatality rates. Utilizing the K-Means Clustering technique, a popular unsupervised machine learning method, this research categorizes holiday periods based on similarities in crash frequency patterns. The findings demonstrate clear differences in crash rates between ordinary days, major festive holidays, and extended seasonal holidays (which often include school breaks and long weekends).

Among these, seasonal holidays emerged as having the highest traffic fatality risk, yet they are typically overlooked in current safety operations. The insights gained from this clustering analysis offer substantial potential to refine and optimize road safety planning, especially in terms of when and how interventions should be deployed. For example, targeted enforcement and awareness campaigns could be expanded to include school holiday periods, not just major festive dates.

However, while the timing of interventions is important, the nature of road crashes and the types of road users involved must also be considered to ensure the effectiveness of safety measures. Travel behaviors vary greatly between regions and contexts. For instance, long-distance travel from urban areas to rural hometowns may be associated with increased highway usage and higher-speed vehicle-to-vehicle collisions. In contrast, short-distance travel within rural or semi-urban towns often involves more vulnerable road users, such as motorcyclists, and takes place on roads with different design and enforcement characteristics. Therefore, intervention strategies must be tailored not only to the timing of high-risk periods but also to local travel patterns and crash profiles. A one-size-fits-all approach may not be sufficient. Data-driven, location-specific, and user-specific measures are essential to improving road safety outcomes.

It is also important to acknowledge the limitations of the current study. The analysis primarily considers temporal variables such as the day of the week and type of holiday, while other potentially significant contributing factors are not included. These may include weather conditions, which can influence visibility and road surface conditions; time of day, as nighttime driving typically carries higher risk; traffic

volume and congestion levels, which fluctuate based on location and event timing; and spatial characteristics, such as road design, traffic control measures, and proximity to high-risk zones.

These additional variables are likely to contribute meaningfully to the variability in crash outcomes. Future research should consider integrating both temporal and spatial dimensions, along with traffic flow and behavioural data, to construct a more comprehensive model of crash risk prediction. Such integrated research would provide policymakers, road safety practitioners, and enforcement agencies with more accurate tools for forecasting and decision-making, enabling them to allocate resources – such as enforcement personnel, equipment, and campaign efforts – more effectively. Most importantly, these insights could be incorporated into strategic frameworks like the National Road Safety Plan (NRSP) to support Malaysia's commitment to achieving the United Nations Sustainable Development Goal (SDG) Target 3.6, which aims to halve the number of global deaths and injuries from road traffic crashes by 2030.

In conclusion, this study highlights the urgent need for seasonally adaptive and data-driven traffic safety interventions in Malaysia. By recognizing and responding to the nuanced patterns of road fatalities across different holiday types, stakeholders can implement more targeted, efficient, and equitable road safety measures – ultimately saving lives and supporting a safer transportation system for all road users.

ACKNOWLEDGEMENT

The authors would like to thank all members of MIROS for their valuable input and strong cooperation throughout this work. This research was supported by MIROS.

REFERENCES

- Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of everyday weather on individual daily travel behaviours in perspective: A literature review. *Transport Reviews*, 33(1), 71–91.
- Claros, B., Chitturi, M., Vorhes, G., Bill, A., & Noyce, D. (2020). Seasonal crash prediction model for urban signalized intersections: Wisconsin southeast region. *Traffic Injury Prevention*, 21(7), 447-452.
- El-Basyouny, K., & Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, 41(4), 820–828.
- El-Basyouny, K., Barua, S., & Islam, M. T. (2014). Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Acc. Analysis & Prevention*, 73, 91–99.
- Kamel, M. B., & Sayed, T. (2021). Accounting for seasonal effects on cyclist–vehicle crashes. *Accident Analysis & Prevention*, 159, 106263.
- Low, S. F., Mohd Soid, N. F., Wong, S. V., & Abu Kassim, K. A. (Eds.). (2017). *Compilation of OPS Bersepadu studies conducted during Hari Raya Aidilfitri 2017 (MRR 319)*. MIROS.
- Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, 11, 1–16.
- NST (2024). *Op Selamat: 35 deaths reported on the first day of Raya*. New Straits Times. Retrieved from <https://www.nst.com.my/news/nation/2024/04/1036802/op-selamat-35-deaths-reported-first-day-raya>
- PDRM (2023). *Laporan perangkaan kemalangan jalan raya Malaysia 2023*. Polis Diraja Malaysia (PDRM).
- Radzuan, N. Q., Hassan, M. H. A., Musa, R. M., Majeed, A. A., Razman, M. M., & Kassim, K. A. (2020). A support vector machine approach in predicting road traffic mortality in Malaysia. *Journal of the Society of Automotive Engineers Malaysia*, 4(2), 135-144.