Unveiling driving behavior patterns during a naturalistic driving experiment

Virginia Petraki[a]*, Stella Roussou[a], Christos Katrakazas[a], Muhammad Adnan[b], Kris Brijs[b], Tom Brijs[b], George Yannis[a]

[a] Department of Transportation Planning and Engineering – National Technical University of Athens (NTUA), Athens, Greece.
[b] UHasselt, School for Transportation Sciences, Transportation Research Institute (IMOB), Agoralaan, 3590 - Diepenbeek, Belgium.

Abstract.

This paper aims to provide a detailed overview of driving behaviour indicators during the implementation of the H2020 project i-DREAMS interventions in Greece. To fulfil this aim, a robust methodology utilizing a k-means clustering approach was employed to detect meaningful driving behaviour patterns within a dataset comprising 11,731 trips from 56 Greek car drivers. This exploratory analysis was complemented by an unsupervised pattern recognition algorithm, which aimed at identifying clusters based on safe or dangerous driving behaviour of the users. The assessment of driving behaviour encompassed indicators such as speeding events, harsh braking and accelerating events, and distraction events (phone in hand). This analysis provides valuable insights into the risky driving behaviour among the i-DREAMS naturalistic driving experiment Phases in Greece.

Keywords: i-DREAMS; cluster analysis; risky driving; safety interventions; naturalistic driving experiment.

1 Introduction

Road safety is a major public health issue that requires immediate coordinated efforts and effective prevention. Although several efforts are being made to improve road safety, at a global level the death toll remains very high, estimated at 1.3 million per year [1]. The three main factors contributing to road crashes are the road user, the road environment, and the vehicle, with driver behavior being the main cause of 95% of road crashes [2].

In recent years, advancements in technology have revolutionized the monitoring and analysis of driving behavior. Automotive telematics and driver monitoring systems utilize connected technologies and big data to provide safety interventions and feedback to drivers. These systems leverage sensors in smartphones or On-Board Diagnostics devices to evaluate driver behavior. The primary goal of these interventions is to improve driving behavior and promote road safety and sustainable mobility [3].

The i-DREAMS project, funded by the European Union's Horizon 2020 program, focuses on establishing a platform and system for timely safety interventions, specifically aiming to create a context-aware 'Safety Tolerance Zone' framework for driving.

This paper provides a comprehensive overview of driving behavior indicators observed during the implementation of i-DREAMS safety interventions in Greece, with the objective of identifying significant driving behavior patterns and understanding factors contributing to safe or dangerous driving. The analysis utilized a robust methodology, including a k-means clustering approach, on a dataset comprising 11,731 trips from 56 Greek car drivers. It also employed an unsupervised pattern recognition algorithm to identify clusters based on safe or dangerous driving behavior. Key indicators assessed included speeding events, harsh braking and accelerating events, and distractions like phone usage while driving. These indicators are widely recognized as significant contributors to road accidents and are crucial for evaluating the effectiveness of interventions aimed at improving road safety [4].

2 Methodology

2.1 The i-DREAMS naturalistic driving experiment in Greece

The purpose of the i-DREAMS interventions is to effectively increase driver safety by supporting drivers in their driving task. The experimental design of the i-DREAMS on-road study in Greece consisted of three Phases.

Phase 1 served as the baseline phase, where driving behavior was monitored without interventions for risky driving events. This phase aimed to establish a comparison between driving behavior with and without safety interventions. The baseline measurements were conducted over a 4-week duration. Phase 2 spanned four weeks and involved post-trip interventions. The i-DREAMS post-trip interventions can be qualified as digital-or internet-based interventions via app and are to be understood as combining e-coaching with virtual coaching. Finally, in Phase 3, which lasted six weeks, gamification features were introduced to the drivers. Unlike the previous phase, drivers were rewarded or received benefits for practicing safe driving behavior. A competitive element was introduced through the leader board function.

The system employed flexible thresholds to determine the STZ status, which defines three risk levels: low (crash risk is minimal), medium (risk of crash increases as internal /external events occur), and high (crash risk is further increased if no preventative action taken by driver). In the framework of this paper, it must be noted that events are presented in the following two severity levels ‘medium’, and ‘high’, which correspond to the ‘Danger’ and ‘Avoidable crash’ driving phases of the STZ. The overall objective of the i-DREAMS platform was to keep drivers in the low STZ level for as long as possible and prevent their transition from the medium to the high STZ level.

2.2 Methodological background

Cluster analysis is a powerful statistical technique used to group similar data points based on certain features or characteristics [4]. In driving behavior analysis, cluster analysis helps identify patterns and understand road safety implications. A widely used algorithm for clustering is K-means clustering [5]. It partitions data into 'k' clusters and minimizes the within-cluster sum of squares (WCSS) [5]. Each data point is assigned to the nearest centroid iteratively until the centroids stabilize.

K-means clustering has been applied extensively in transportation and road safety research [6; 7]. Various customized methods have also been developed [8; 9]. Determining the optimal number of clusters ('k') is crucial, and methods like the Silhouette Coefficient can help [10; 11].

3. Results and Discussion

Table 1 provides a comprehensive overview of the descriptive statistics for various driving events per 100 kilometers (km) during each of the three Phases. Those numbers offer valuable insights into how driving behavior evolved in response to safety interventions and gamification features.

In Phase 1, the baseline phase, the trip distance traveled shows considerable variability, with a relatively wide standard deviation (Std) of 15 km. As for total speeding events per 100 km, the high Std of 37 indicates considerable dispersion in these events. Acceleration, deceleration, and distraction events also demonstrate substantial variability, indicating a wide spectrum of driving styles within the baseline phase. Moving to Phase 2, we observe a decrease in the mean trip distance, possibly suggesting a preference for shorter trips. Notably, the reduction in speeding events, indicates the initial effectiveness of post-trip interventions in promoting safer driving. In Phase 3, the mean distance traveled increases to 10 km, implying greater variability in trip lengths. Most significantly, the continued decline in speeding events, with a mean of 27, underscores a sustained shift towards safer driving. The improvements in other examined events further validate the positive impact of combined safety interventions in road safety.

Before starting the analysis, data was cleaned removing trips that were ‘outside phase’, excluding drivers who did not have trip data in all phases, removing the trips that were outliers (defined as the mean +/- three standard deviations), and excluding the trips with less than 1km.

Table 1. Descriptives of events per 100km for each Phase

	Phase 1 - baseline				Phase 2				Phase 3			
	Min	Mean	Max	Std	Min	Mean	Max	Std	Min	Mean	Max	Std
distance_km	1	9	278	15	1	8	230	15	1	10	222	18
duration_min	1	17	193	15	1	17	151	14	1	18	185	15
speeding_total	0	32	625	37	0	30	286	34	0	27	400	33
speeding_H	0	25	625	33	0	24	286	32	0	21	400	29
speeding_M	0	7	200	15	0	6	167	14	0	6	125	13
acceleration_total	0	5	375	16	0	6	250	18	0	4	174	14
acceleration_M	0	3	143	10	0	4	250	13	0	2	119	9
acceleration_H	0	2	375	11	0	2	167	9	0	2	167	9
deceleration_total	0	9	375	20	0	9	222	19	0	8	167	17
deceleration_M	0	6	250	15	0	6	222	14	0	5	167	13

deceleration_H	0	3	167	11	0	3	200	11	0	3	125	9
distraction	0	23	875	60	0	23	667	57	0	17	417	48

*H = STZ high level, M = STZ medium level, total = M+H

Based on the analysis of the Silhouette method, Fig.1 indicates that the optimal number of clusters is two for Phase 1 and 3 while for Phase 2 is three. The overall models' quality is considered good (av. Silhouette width>0.4).

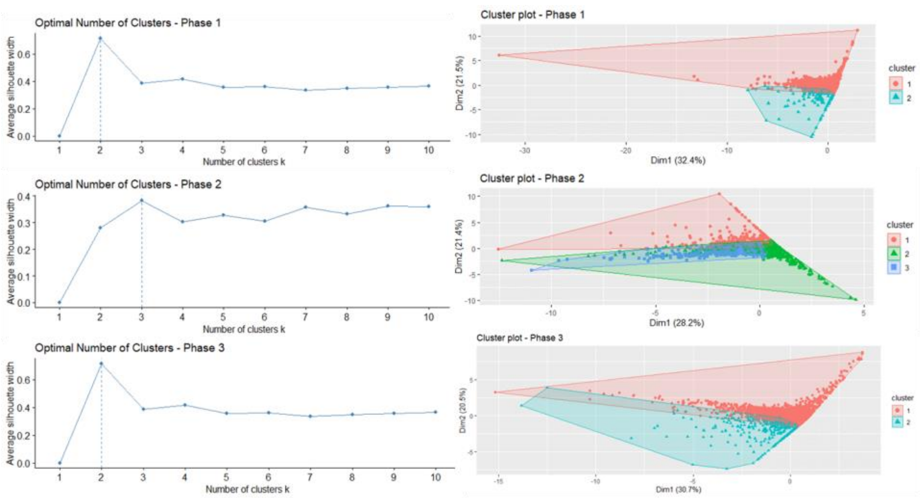

Fig. 1. Optimal number of clusters and the cluster plots for each Phase.

Table 2. Centroid centers and types for driving behavior clusters for each Phase.

Cluster		trip_ distance	speeding total	acceleration _total	deceleration _total	distraction	Cluster_size	
							trips	%
Nr	Behavior	Phase 1						
$1_{,1}$	moderate	9	33	4	9	12	2827	95%
$2_{,1}$	distracted	3	24	11	11	231	143	5%
		Phase2						
$1_{,2}$	distracted	4	27	8	11	179	310	8%
$2_{,2}$	low risk	10	16	3	7	10	2822	69%
$3_{,2}$	speeding	7	79	10	14	17	964	24%
		Phase3						
$1_{,3}$	moderate	11	30	4	8	10	4334	93%
$2_{,3}$	distracted	4	32	9	12	160	331	7%

In the baseline Phase, where no safety interventions or gamification features were introduced to the drivers, the K-means clustering resulted in two distinct clusters. Cluster $1_{,1}$ was characterized by mostly moderate driving behavior within longer trip distances and more frequent speeding events compared to Cluster $2_{,1}$, which displayed distracted driving behavior during shorter trips. Cluster $1_{,1}$ represented most total trips (95%) during Phase 1.

Moving on to Phase 2, where post-trip interventions were introduced via a smartphone app, K-means clustering identified an additional cluster. Comparing Cluster $1_{,1}$ with Cluster $2_{,2}$, we observed an improvement in driving behavior in terms of the number of dangerous driving events per 100 km. This improvement characterized Cluster $2_{,2}$ relatively safer driving behavior, and it still represents most of the trips (69%). Notably, Cluster $1_{,2}$ displayed a significant number of mobile use events, yet it demonstrated a substantial reduction in these events (-23%) compared to Cluster $2_{,1}$. This reduction implies a positive trend towards decreased distracted driving behavior within Cluster $1_{,2}$. Furthermore, a new cluster emerged in Phase 2, comprising drivers with a tendency for speeding behavior, representing 24% of the total trips.

In the final Phase, where drivers received post-trip interventions and were introduced to gamification features, K-means clustering identified two clusters. Cluster $1_{,3}$ demonstrated significant improvements in driving behavior compared to both Phase 1 and Phase 2. Drivers in this cluster exhibited fewer speeding events, fewer aggressive accelerations and decelerations, and reduced distraction events. Cluster $1_{,3}$ can be characterized as relatively moderate driving behavior and represented the vast majority of the total trips. Cluster $2_{,3}$ also showed improvements in driving behavior, although distraction events remained relatively frequent but lower compared to Phases 1 and 2.

4. Conclusion

The analysis of driver behavior plays a crucial role in enhancing road safety and reducing crashes. To that end, a cluster analysis of Greek car driver trips made for each Phase of i-DREAMS experiment using k-means clustering approach to understand the impact of post-trip safety interventions and gamification features on driving behavior. The optimal number of driving behavior clusters was found to be two during both the baseline Phase and the final Phase, which introduced gamification features alongside post-trip interventions. In contrast, Phase 2, which solely featured post-trip interventions, revealed three distinct driving behavior patterns.

The analysis revealed that the clusters differed in terms of distances traveled and various driving behavior indicators such as speeding events, harsh acceleration and deceleration events, and distraction events. Across the three phases of this study, distinct patterns in driving behavior emerged. In the baseline Phase (Phase1) with no safety interventions, two primary profiles emerged. One exhibited moderate driving with longer trips and relatively more frequent speeding events, while the other displayed distracted driving, especially during shorter trips. In Phase 2, post-trip interventions led to safer driving habits, marked by fewer risky driving events like speeding and mobile phone use. A new speeding behavior cluster also emerged. Moving to final Phase 3, where gamification elements were incorporated with post-trip interventions, saw substantial improvements in driving behavior, with reduced speeding, harsh events, and distraction, highlighting the effectiveness of tailored interventions in promoting road safety.

In summary, the three Phases of the i-DREAMS experiment in Greece demonstrated a clear progression in driving behavior improvement as safety interventions and gamification features were introduced. Future research should focus on longitudinal analyses to understand how driving behavior evolves over time. Exploring the relationship between driver characteristics and behavior can inform personalized interventions. Evaluating intervention effectiveness and leveraging advanced technologies can further enhance road safety.

5. Acknowledgements

Part of this work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 814761.

References

1. World Health Organization, 2022. Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. Singh, S., 2018. Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts Crash Stats. Report No. DOT HS 812 506). Washington, DC: National Highway Traffic Safety Administration
3. Zaira, M.M. and Hadikusumo, B.H., 2017. Structural equation model of integrated safety intervention practices affecting the safety behaviour of workers in the construction industry. Safety science, 98, pp.124-135. https://doi.org/10.1016/j.ssci.2017.06.007
4. Ziakopoulos A., Kontaxi, A.,Yannis, G., Fortsakis, P., Kontonasios, K. N., & Kostoulas, G. (2020). Advanced driver monitoring using smartphone applications: The BeSmart project. Proceedings of the 8th Transport Research Arena TRA, 27-30.
5. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1), 100-108.*
6. Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modelling for the regional effect of enforcement on road accidents. Accident Analysis & Prevention, 39(4), 818-825.
7. Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. Safety Science.
8. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), 881-892.
9. Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.
10. A. Et-Taleby, M. Boussetta and M. Benslimane, "Faults detection for photovoltaic field based on K-means elbow and average silhouette techniques through the segmentation of a thermal image", Int. J. Photoenergy, vol. 2020, Dec. 2020.
11. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.