Exploring Safe and Eco Driving Behavior through Large-scale Data using Unsupervised Learning

Virginia Petraki[1*], Dimitris Nikolaou[1], George Yannis[1]

[1]*National Technical University of Athens, Department of Transportation Planning and Engineering,*
5 Heroon Polytechniou str., 15773, Athens, Greece

* vpetraki@mail.ntua.gr

Abstract

This study aims to detect and analyze sustainable driving styles with respect to road safety and fuel economy using real-world trip data collected via smartphone sensors. A two-level clustering approach was applied using K-means: first, trips were segmented by average speed and road type share; second, behavioral indicators such as harsh braking events, mobile phone use, acceleration variability, and fuel consumption were used. Principal Component Analysis was applied for dimensionality reduction, and the Silhouette method for optimal cluster selection. The analysis revealed distinct driving profiles across urban, rural-dominant and highway-oriented driving. In rural settings, safe and eco-driving behaviors were strongly aligned. On highways, fuel efficiency sometimes coincided with riskier behaviors such as distraction. In urban contexts, however, some less fuel-efficient drivers exhibited relatively safer behavior, suggesting a trade-off. These findings underscore the context-dependent nature of sustainable driving and highlight the need for strategies that address both safety and environmental goals.

Keywords: *driving behavior, fuel efficiency, safe driving, K-means, PCA*

1. Introduction

Road crashes remain a major and growing global challenge, contributing to approximately 1.19 million fatalities annually (WHO, 2023). Simultaneously, the transport sector accounts for around 25% of the European Union's (EU) total carbon dioxide emissions and 31% of its total energy consumption, with road transport being the primary contributor (European Commission, 2025; Eurostat, 2024). Road safety, environmental sustainability, and economic efficiency form the three foundational pillars of sustainable transportation (European Court of Auditors, 2020).

Among the various factors influencing transportation outcomes, driving behavior is the most critical, directly affecting road safety (Singh, 2018), fuel efficiency, and vehicle emissions (Zhang et al., 2022; Zhou et al., 2016). Vehicle speed, acceleration, braking, location, mileage, and temporal variation has a significant correlation with the crash risk. Whereas engine RPM, idling duration, and jerks have a significant correlation with fuel economy and vehicle emission (Singh and Kathuria, 2019). The rapid technological development in naturalistic driver recording has also brought about an increasing availability of data from sensors in vehicles and smartphones that can be used to extract various SSMs such as Time To Collision (TTC), harsh braking events, and harsh acceleration events (Nikolaou et al., 2023).

Numerous studies in the driving behavior literature have aimed to classify drivers into behavioral profiles based on operational patterns, frequently identifying categories such as aggressive, distracted, risk-taking, eco-conscious, and safe drivers (Mantouka et al., 2021). However, the interplay between eco-driving and safe-driving behaviors remains underexplored, particularly in terms of whether these styles consistently align or diverge across different driving contexts.

This study seeks to explore the intersection of safe and eco-driving behavior using real-world trip data and unsupervised learning techniques. The following research questions guide the analysis:

Q1. How can trips be meaningfully clustered according to driving context?

Q2. What are the key behavioral parameters that characterize safe and eco-efficient driving styles?

Q3. Do safe and eco-driving behaviors consistently align?

The paper is structured as follows: Section 2 presents the methodology, detailing the two-level K-means clustering approach, the use of PCA for dimensionality reduction and data collection. Section 3 discusses the results, interpreting driving behavior clusters across different road contexts. Section 4 concludes with key findings.

2. Methodology

2.1 Overall methodology

This study employs a two-level K-means clustering approach to holistically identify driving behavior patterns in relation to both road safety and fuel efficiency. Each clustering level followed a structured process: (1) determining the optimal number of clusters using the Silhouette method, (2) applying dimensionality reduction techniques, primarily Principal Component Analysis (PCA), (3) performing clustering using K-means, and (4) evaluating the classification quality. PCA was applied where appropriate, particularly in the second-level clustering, to reduce dimensionality and emphasize the most significant behavioral variance.

The first-level clustering was motivated by the premise that both safe and eco-driving behaviors are significantly influenced by road type. Urban areas, for instance, are associated with higher crash frequency (Ryder et al., 2019), increased fuel consumption, and greater emissions (Zhu et al., 2022) compared to rural and expressway environments. Therefore, for each trip, the proportion of time spent in different road types, along with the average driving speed, was used to form the initial clusters.

Following this segmentation, a second-level clustering was applied using driving behavioral parameters identified in the literature as indicators of safe and eco-driving. According to Li et al. (2019), eco-safe driving involves reducing fuel consumption and emissions while maintaining safety through smooth acceleration and deceleration, speed compliance, and safe following distance. Jain and Mittal (2023) similarly define eco-safe driving as minimizing fuel use without compromising safety.

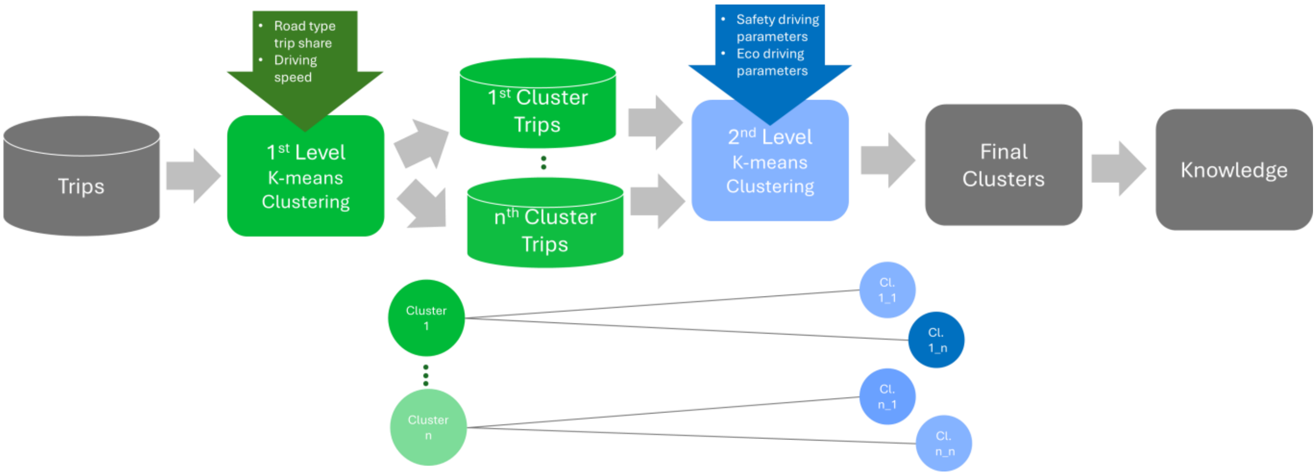


Figure 1: *Methodology Framework*

2.2 K-Means

Cluster analysis is a powerful statistical technique used to group similar data points based on certain features or characteristics (Ziakopoulos et al., 2020). In driving behavior analysis, cluster analysis helps identify patterns and understand road safety implications. A widely used algorithm for clustering is K-means clustering. It partitions data into 'k' clusters and minimizes the within-cluster sum of squares (WCSS) (Hartigan & Wong, 1979). 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 (Yannis et al., 2007; Mantouka et al., 2019). Various customized methods have also been developed (Kanungo et al., 2002; Likas et al., 2003).

Determining the optimal number of clusters ('k') is crucial, and methods like the Silhouette Coefficient can help (Et-Taleby et al., 2020; Rousseeuw, 1987). For a given cluster, Xj (j = 1,....,c), this method assigns to each sample of Xj a quality measure, s(i) (i = 1,....,m), known as the Silhouette width, which is defined as:

$$s(i)= \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad \textbf{(1)}$$

where a(i) is the average distance between the i^{th} sample and all of the samples included in Xj; 'max' is the maximum operator, and b(i) is the minimum average distance between the i^{th} sample and all of the samples clustered in X_k (k = 1,....,c; k ≠ j). From this formula, it follows that s(i) takes values among -1 and 1. When a s(i) is close to 1, the i^{th} sample has been assigned to an appropriate cluster. A silhouette value s(i) close to 0 indicates that the i-th observation lies near the boundary between two clusters and could potentially belong to either. A value near −1 suggests that the observation may have been misclassified. Based on this, a cluster-level Silhouette index can be computed to reflect the cohesion within a cluster and its separation from others.

2.3 Principal component analysis (PCA)

PCA is a linear dimensionality reduction technique during which the data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified. In the framework of this paper, PCA was applied solely for dimensionality reduction, transforming the original variables into a new set of orthogonal components that retain the most significant variance. When performing PCA, the first principal component of a set of "p" variables is the derived variable formed as a linear combination of the original variables that explains the most variance. The second principal component explains the most variance in what is left once the effect of the first component is removed, and we may proceed through "p" iterations until all the variance is explained.

2.4 Data Collection

The data was collected through an innovative data collection scheme, developed by the OSeven Telematics Company (www.oseven.io), which records personalized driving behavior analytics in real time, using smartphone sensors. An integrated system is used for recording, collection, storage, evaluation and visualization of driving behavior data, using smartphone applications and advanced Machine Learning algorithms. The system includes specially developed smartphone applications (apps) for data collection and transmission, as well as for providing feedback to the participants on their driving behavior. The data are stored in the OSeven backend system using advanced encryption and data security techniques, in compliance with the national laws and EU Directives for the protection of personal data. The APIs (Application Programming Interfaces) used support user authentication and encryption to prevent unauthorized data access. For more information about the collection, storage, management, and processing of data, the reader can refer to previous research papers (Papadimitriou et al., 2019).

The dataset captures various driving parameters, including GPS speed, harsh acceleration and deceleration events, mobile phone use and fuel consumption, providing valuable insights into driving styles. The data was collected in an anonymized format from 16,118 trips over a three-month period (March to May) during the years 2023, and 2024. Key descriptive statistics for all years are summarized in the following table.

Table 1 presents the key variables selected for the two-level clustering analysis, along with their descriptions and summary statistics.

***Table 1:** Dataset description*

Variable Name	Description	Summary Statistics			1st Clustering	2nd Clustering
		Min	Median	Q3		
Urban_prc	Percentage of trip duration spent on urban roads (Speed Limit<50)	0.00	59.56	81.98	●	
Rural_prc	Percentage of trip duration spent on rural roads (50<Speed Limit<80)	0.00	35.14	55.18	●	
Highway_prc	Percentage of trip duration spent on highways (Speed Limit>80)	0.00	4.16	0.00	●	
Speed_Avg	Average speed of the trip (km/h)	4.73	26.24	36.98	●	
Harsh_Brk_per_km	Number of harsh braking events per kilometer	0.000	0.080	0.270		●
MobilePhone_min_per100km	Minutes of mobile phone usage per 100 kilometers	0.000	0.000	6.410		●
Speed_Q90	90^{th} percentile of speed during the trip	16.820	53.600	73.200		●
Acc_QCV	Coefficient of variation of acceleration (QCV= $100 \times \frac{Q_3-Q_1}{Q_3+Q_1}$)	0.000	0.600	0.640		●
Fuel_lit_per100km	Fuel consumption measured in liters per 100 kilometers	2.393	8.280	10.791		●

The first clustering, which segments trips based on driving context, relies on variables such as the percentage of trip duration spent on urban, rural, and highway roads, as well as average trip speed. The second clustering focuses on driving style and includes variables related to safety and fuel efficiency, such as harsh braking events per kilometer, mobile phone usage, high-speed tendencies (90th percentile of speed), variability in acceleration, and fuel consumption. The summary statistics reveal considerable variability, particularly in safety-related behaviors like harsh braking and phone use, which are sparse but critical in characterizing riskier driving profiles. This combination of contextual and behavioral features ensures a comprehensive evaluation of driving patterns with respect to both safety and fuel economy.

3. Results and Discussion

3.1 First Level Cluster Analysis – Road type

3.1.1 Number of Clusters

Based on the Silhouette method (Fig. 1), the optimal number of clusters was determined to be three. The model demonstrated good overall clustering performance, with an average silhouette width above 0.4, indicating well-separated and cohesive clusters.

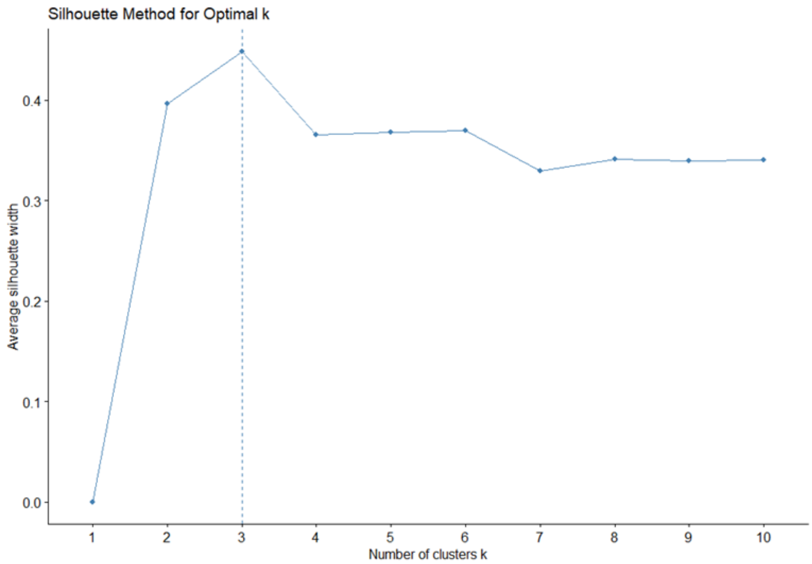


Figure 2: *Optimal number of clusters– 1st level clustering*

3.1.2 Variable Selection

The first-level clustering was performed using four key variables: Urban_prc, Rural_prc, Highway_prc, and Speed_Avg. These variables represent the share of trip time spent on each road type and the average driving speed, which are known to influence both safety and fuel efficiency. The selection of these variables allowed for the segmentation of trips according to dominant driving environments.

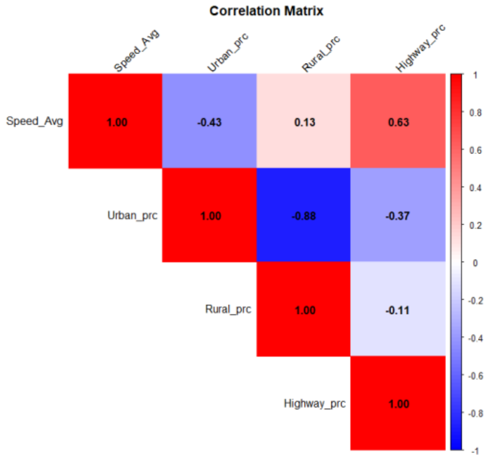


Figure 2: *Correlation Matrix for the selected variables – 1st level clustering*

3.1.3 Clustering Results

Table 3 and Figure 2 summarize the results of the first-level clustering analysis based on average speed and road-type distribution. The K-means algorithm identified three distinct clusters that reflect different driving contexts: urban, rural, and highway-dominant trips.

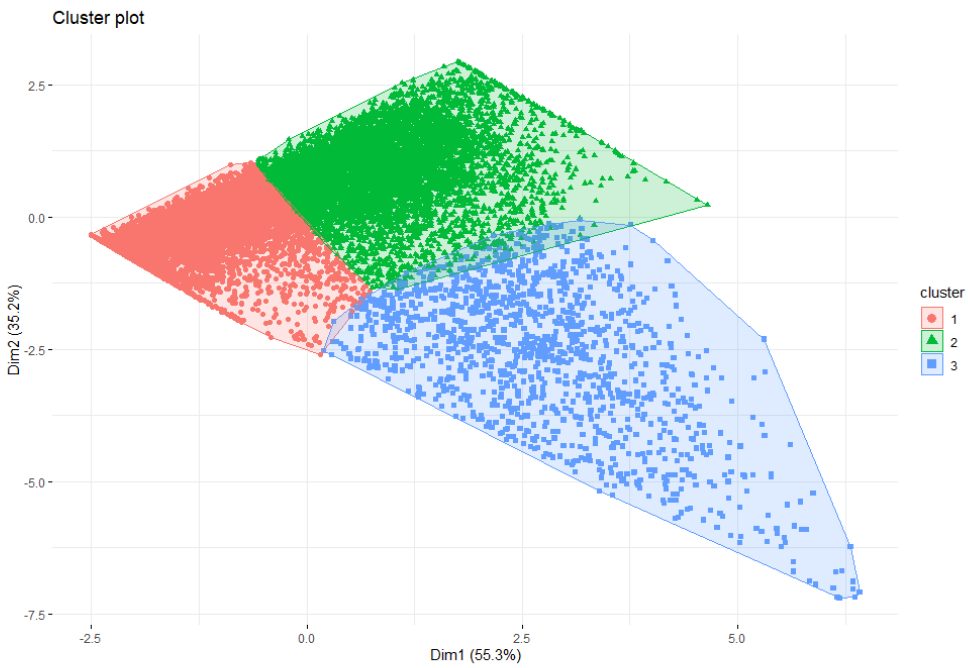


Figure 3: *Cluster plots – 1st level clustering*

Table 2: *Cluster Centers – 1st level clustering*

Cluster		Speed_Avg	Urban_prc	Rural_prc	Highway_prc	Trips
1	Urban Driving	23.48	83.49	16.09	0.52	7675
2	Rural-Dominant Driving	31.23	39.32	59.25	1.33	7145
3	Highway-Oriented Driving	62.30	28.95	28.70	41.27	1298

Cluster 1, labeled Urban Driving, is the largest group with 7,675 trips. These trips are characterized by a dominant share of urban road usage (83.49%) and the lowest average speed (23.48 km/h), reflecting traffic-dense environments with frequent stops and low speed limits. Cluster 2, referred to as Rural-Dominant Driving, includes 7,145 trips. These trips show a higher proportion of rural road exposure (59.25%) and a moderate average speed of 31.23 km/h. This suggests mixed traffic conditions, typically associated with less congestion than urban roads but lower speeds than highways. Cluster 3, identified as Highway-Oriented Driving, contains 1,298 trips. It features the highest average speed (62.30 km/h) and the highest highway share (41.27%), indicating longer-distance trips conducted mostly on high-speed roads with fewer interruptions.

Figure 3 illustrates the separation of these three clusters in two-dimensional principal component space. The clusters are visibly distinct, confirming the suitability of the selected features in differentiating trip contexts. Cluster 3 (highway-oriented) is clearly separated along Dim1, likely driven by speed-related metrics, while Clusters 1 and 2 show more overlap along Dim2, reflecting their closer similarity in road-type proportions.

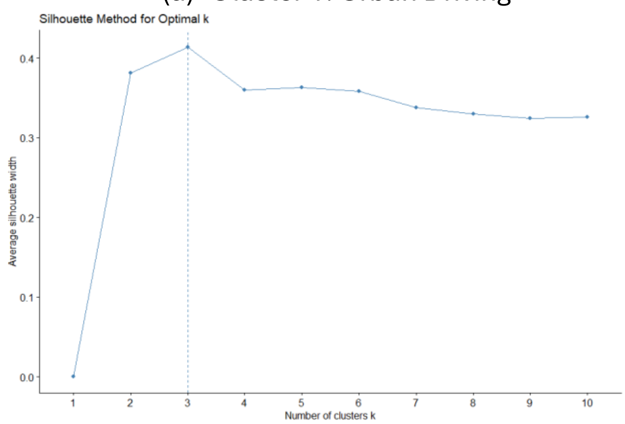
3.2 Second Level Cluster Analysis – Driving Behavior

In order to detect driving behaviors based on road safety and efficiency, a further categorization of the trips was made. A second level clustering was performed for urban driving, rural-dominant driving and highway-oriented driving trips.

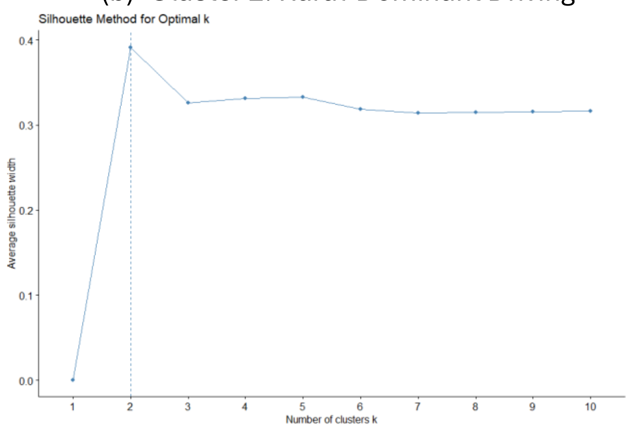
3.2.1 <u>Number of Clusters</u>

The optimal number of clusters from the first level clustering was determined using the Silhouette method. As shown in the bottom row of Figure 4, the silhouette scores supported the selection of three clusters for urban and highway trips, and two clusters for rural trips. The silhouette widths were consistently acceptable (above 0.23), with urban and rural clusters demonstrating particularly strong separation (above 0.45), indicating good clustering cohesion and distinctiveness.

(a) Cluster 1: Urban Driving



(b) Cluster 2: Rural-Dominant Driving



(c) Cluster 3: Highway-Oriented Driving

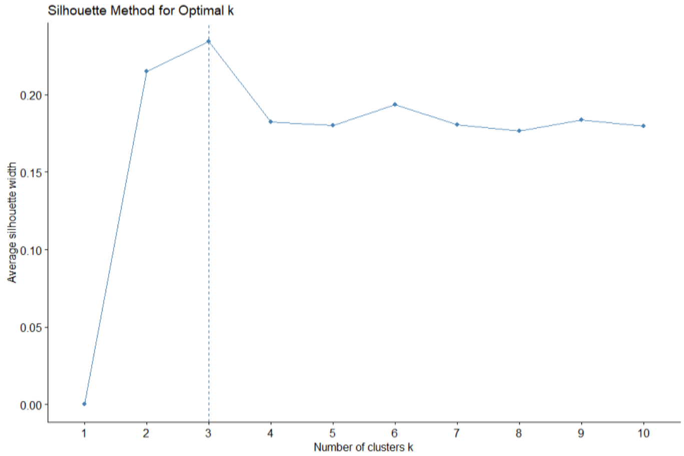


Figure 4: *Optimal number for clusters– 2nd level clustering*

3.2.2 Variable Selection

For the second-level clustering, behavioral parameters related to safety and fuel efficiency were selected. These included: Harsh_Brk_per_km (braking aggressiveness), MobilePhone_min_per100km (driver distraction), Speed_Q90 (upper-speed behavior), Acc_QCV (acceleration consistency), and Fuel_lit_per100km (fuel consumption).

To enhance interpretability and reduce dimensionality, PCA was applied to the urban and rural datasets, retaining the first two principal components which explained approximately 60% of the variance in both cases. For highway trips, PCA was not applied, as clustering directly on the original variables yielded clearer separation, suggesting strong, well-structured input variables.

As presented in Table 3, for urban driving dataset, the first principal component (PC1) explains 33.4 % of the variance, the second principal component (PC2) explains 20.5 %, and subsequent components explain lesser amounts: 19.9 %, 18.3 %, and 8 %. For the rural-dominant driving, the percentages are: PC1 explains 33.8 %, PC2 explains 21 %, followed by 18.7 %, 17.1 %, and 9.4 %.

Table 3: *Principal Component Analysis results– 2nd level clustering*

Loadings	**Cluster 1: Urban Driving**		**Cluster 2: Rural-Dominant Driving**	
	PC1 (33.4 %)	PC2 (20.5 %)	PC1 (33.8 %)	PC2 (21.0 %)
Harsh_Brk_per_km	0.09	0.65	0.28	0.60
MobilePhone_min_per100km	0.21	0.58	0.33	-0.04
Speed_Q90	-0.67	0.09	-0.62	0.11
Acc_QCV	-0.23	0.48	-0.12	0.79
Fuel_lit_per100km	0.67	-0.02	0.64	0.01

Based on Table 3, in the urban driving dataset, Fuel_lit_per100km (0.67) and Speed_Q90 (-0.67) are the most influential variables in PC1, indicating that this component primarily reflects differences in fuel consumption and high-speed behavior. Drivers with higher PC1 scores tend to exhibit higher fuel usage and lower top speeds, whereas lower PC1 scores are associated with higher speeds and lower fuel consumption. In PC2, Harsh_Brk_per_km (0.65) and MobilePhone_min_per100km (0.58) are the dominant contributors, highlighting a behavioral dimension related to driving aggressiveness and distraction.

For the rural-dominant dataset, Fuel_lit_per100km (0.64) and Speed_Q90 (-0.62) again dominate PC1, suggesting a similar focus on fuel consumption and speed characteristics. However, PC2 is most strongly influenced by Acc_QCV (0.79), which captures variability in acceleration—an indicator of less consistent or more reactive driving styles commonly observed in varying rural traffic or road conditions.

3.2.3 *Clustering Results*

Cluster1: Urban Driving

The urban segment, comprising 7,675 trips, was clustered into three behavioral groups using PCA and K-means.

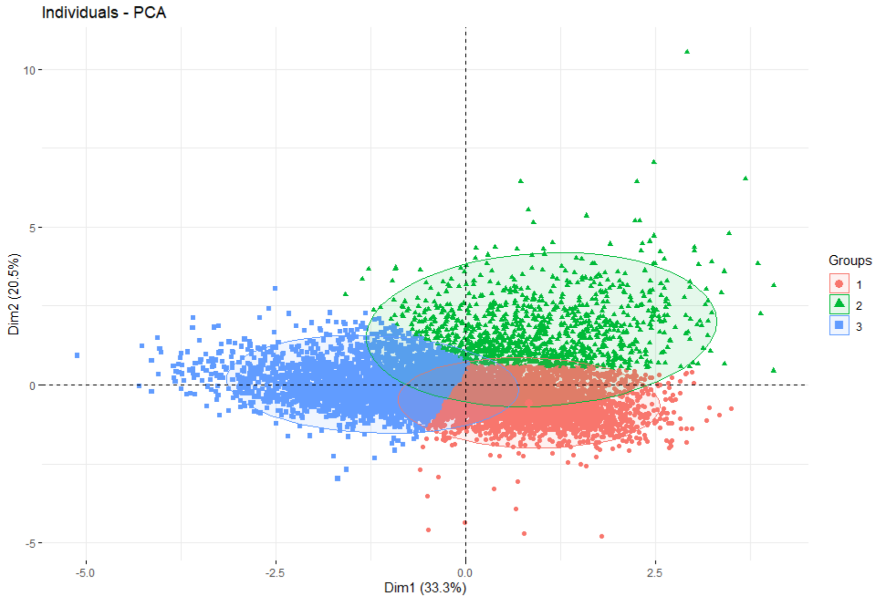


***Figure 5:** Cluster plots – Urban Driving – 2nd level clustering*

***Table 4:** Cluster Centers – Urban Driving – 2nd level clustering*

Cluster		Harsh_Brk_ per_km	MobilePhone_ min_per100km	Speed_ Q90	Acc_ QCV	Fuel_lit _per100km	Trips
1	1_1 Conservative Fuel-Intensive	0.116	7.023	37.482	0.570	12.025	**3418**
2	1_2 Aggressive Distracted	0.630	73.235	40.746	0.612	11.712	**1065**
3	1_3 Eco-Safe	0.178	6.497	60.786	0.616	7.285	**3192**

Cluster 1_1 – Conservative Fuel-Intensive Trips: This is the largest group, with 3,418 trips, characterized by low harsh braking and moderate distraction (7.02 min/100km), suggesting cautious driving. However, their low average high-end speed (Speed_Q90 = 37.5 km/h) and highest fuel consumption (12.03 L/100km) indicate inefficient driving, likely due to prolonged low-speed operation or poor energy management in dense traffic.

Cluster 1_2 – Aggressive Distracted Trips: Representing 1,065 trips, this group shows the highest rate of harsh braking (0.63 per km) and extremely high mobile phone use (73.2 min/100km). Despite moderate speed and acceleration variability, these drivers exhibit the riskiest and most distracted urban behavior profile. Fuel consumption is also elevated (11.71 L/100km), further underlining inefficiency.

Cluster 1_3 – Eco-Safe Trips: With 3,192 trips, this cluster stands out for its highest cruising speed (Speed_Q90 = 60.79 km/h), low harsh events, minimal distraction, and lowest fuel consumption (7.29 L/100km). The combination of efficiency and safety makes this group the most favorable from both an environmental and road safety perspective.

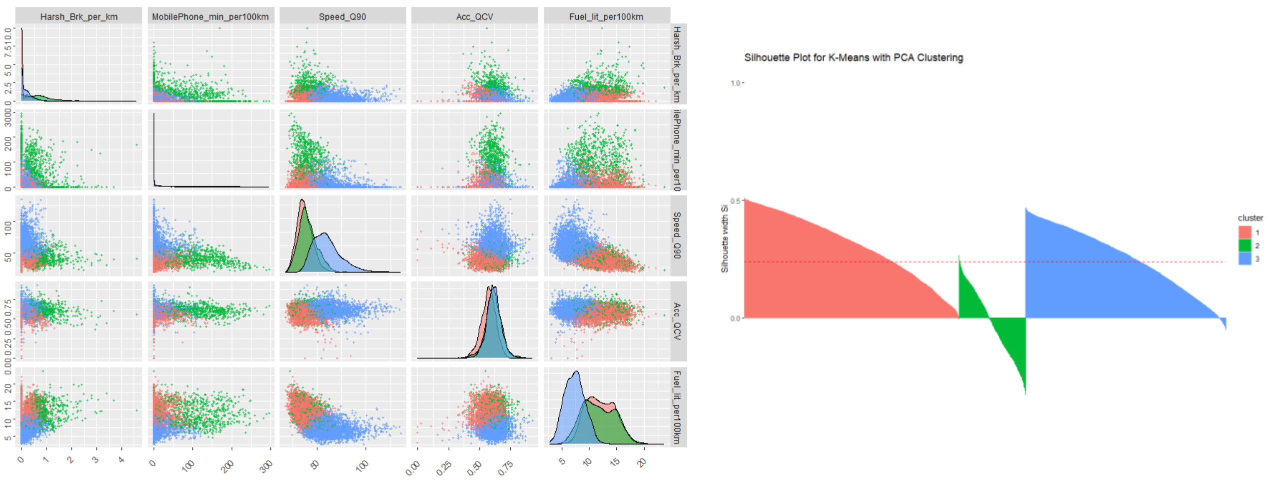


Figure 6: *Clusters based on pairs of the available variables (left), plot of silhouette index per cluster (right)– Urban Driving – 2nd level clustering*

The silhouette plot confirms these patterns: Clusters 1_1 and 1_3 have high silhouette scores, suggesting well-defined and cohesive groupings. Cluster 1_2, however, shows low and negative scores, implying behavioral overlap and inconsistency. These findings highlight that while eco-safe and conservative driving are distinct, aggressive behavior is more variable and less clearly separated.

Cluster 2: Rural-Dominant Driving

The rural driving segment, composed of 7,145 trips, was clustered into two behavioral groups using PCA and K-means.

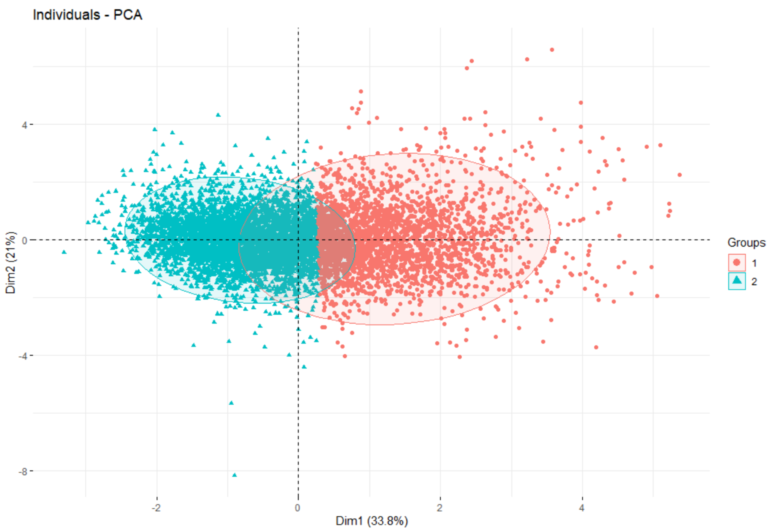


Figure 7: *Cluster plots – Rural-Dominant Driving – 2nd level clustering*

Table 5: *Cluster Centers – Rural-Dominant Driving – 2nd level clustering*

Cluster		Harsh_Brk_per_km	MobilePhone_min_per100km	Speed_Q90	Acc_QCV	Fuel_lit_per100km	Trips
1	2_1 Risky-Inefficient	0.268	21.504	45.504	0.589	10.764	**2709**
2	2_2 Eco-Safe	0.127	4.511	73.393	0.600	6.523	**4436**

Cluster 2_1 – Risky-Inefficient Trips: Comprising 2,709 trips, this cluster exhibits higher harsh braking and notable mobile phone use, indicating a more aggressive and distracted driving style. The average Speed_Q90 is lower, and fuel consumption is significantly higher. These trips tend to operate under less consistent and more reactive conditions.

Cluster 2_2 – Eco-Safe Trips: This larger group of 4,436 trips demonstrates smoother and more efficient behavior, with low harsh braking, minimal distraction, and the highest cruising speeds. Despite similar acceleration variability, their fuel consumption is the lowest. This indicates a profile of consistent, attentive, and fuel-efficient driving typically associated with more experienced or disciplined rural drivers.

Overall, the clustering reveals a clear behavioral distinction: one group follows safe and eco-efficient driving practices, while the other displays more aggressive and less efficient tendencies. These patterns suggest targeted interventions could improve both safety and fuel performance for rural drivers in Cluster 2_1.

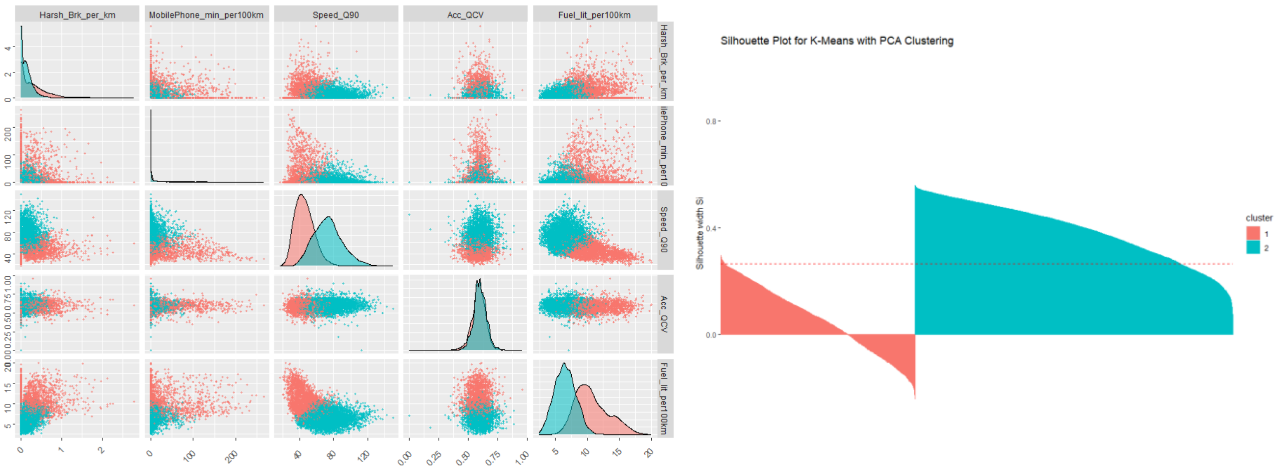


Figure 6: *Clusters based on pairs of the available variables (left), plot of silhouette index per cluster (right)– Rural-Dominant Driving – 2nd level clustering*

The scatterplot matrix reveals clear differences between the two rural driving behavior clusters across key variables. The accompanying silhouette plot supports this distinction. Most observations in Cluster 2_2 have positive silhouette scores, reflecting good internal cohesion and a clear separation from Cluster 2_1. This confirms the behavioral consistency of smooth and eco-safe rural drivers. On the other hand, while Cluster 2_1 contains some negative or borderline silhouette values, these represent a relatively small portion of the total sample. This suggests that although Cluster 2_1 is slightly more variable, the two-cluster solution is still overall valid and meaningful, with the majority of trips well classified.

Cluster 3: Highway-Oriented Driving

The highway segment, consisting of 1,298 trips, was clustered into three groups using K-means. PCA not required due to well-separated variables.

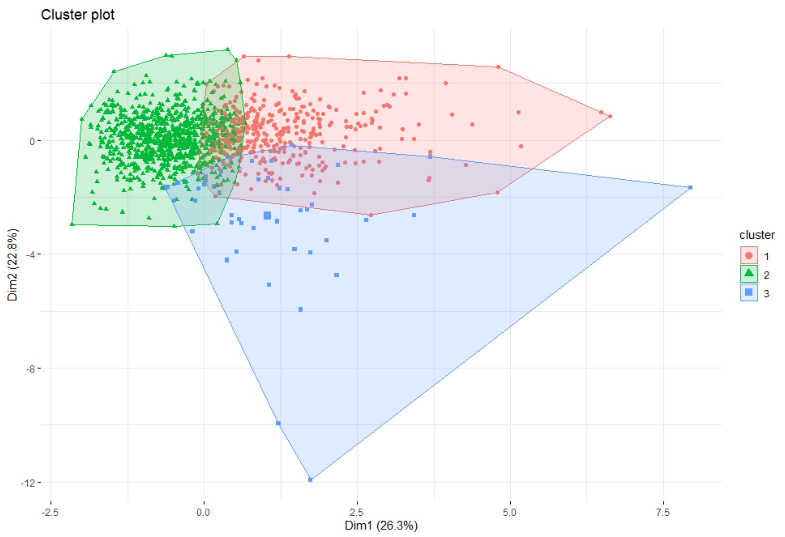


Figure 9: *Cluster plots –Highway-Oriented Driving - 2nd level clustering*

Table 6: *Cluster Centers – Highway-Oriented Driving – 2nd level clustering*

Cluster		Harsh_Brk_per_km	MobilePhone_min_per100km	Speed_Q90	Acc_QCV	Fuel_lit_per100km	Trips
1	**3_1 Moderate**	0.170	1.959	111.866	0.645	7.229	**375**
2	**3_2 Eco-Safe**	0.043	1.649	100.989	0.605	5.442	**876**
3	**3_3 Distracted**	0.089	44.881	105.287	0.602	6.433	**47**

Cluster 3_1 – Moderate Trips: These trips are characterized by the highest cruising speeds , low mobile phone usage, and moderate levels of harsh braking. However, they also exhibit the highest fuel consumption (7.23 L/100 km), suggesting that although these trips are steady and not aggressive, they may involve inefficient fuel use at high speeds.

Cluster 3_2 – Eco-Safe Trips : This is the largest and most favorable cluster, showing minimal harsh braking, low distraction, and the lowest fuel consumption. These trips reflect a consistently safe and more efficient driving style, ideal for highway conditions where smooth and attentive behavior can optimize both safety and economy.

Cluster 3_3 – Distracted Trips =: A small group of trips stands out due to very high mobile phone use (44.88 min/100 km), indicating significant distraction. Although their speed and fuel consumption are moderate, the combination of elevated distraction and moderate harsh braking marks these trips as risk-prone, even in relatively stable highway environments.

Overall, the clustering suggests that while most highway trips exhibit eco-safe characteristics, a minority pose safety concerns due to distraction, highlighting a targeted opportunity for behavioral interventions.

The silhouette analysis confirms the validity of the three-cluster solution in highway settings. The Eco-Safe cluster is both large and well-defined, while the Distracted cluster, though small, is clearly distinct. The Moderate group represents a broader behavioral spectrum with some transitional overlap.

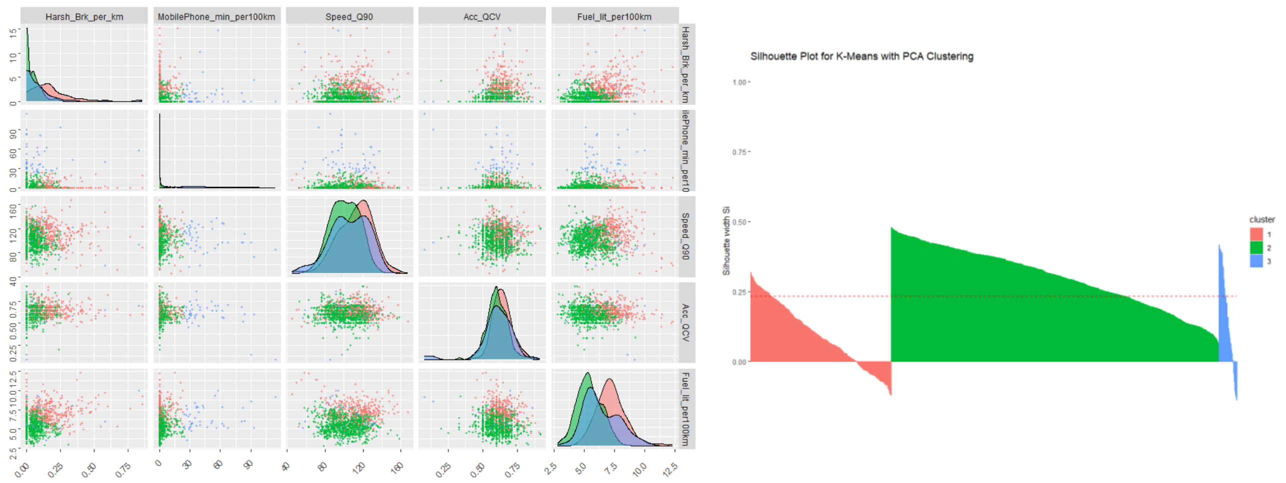
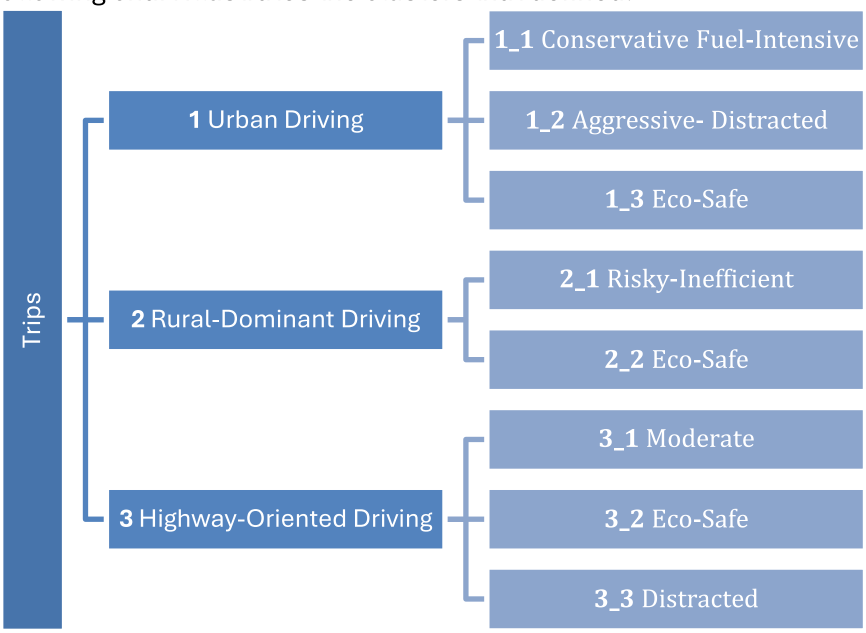


Figure 10: *Clusters based on pairs of the available variables (left), plot of silhouette index per cluster (right)– Highway-Oriented Driving – 2nd level clustering*

In summary, the following chart illustrates the clusters that defined.



4. Conclusion

This study leveraged a two-level K-means clustering framework to analyze over 16,000 real-world trips and identify sustainable driving behaviors across different road contexts. The findings reveal that the alignment between safe and eco-driving styles is not uniform but varies significantly depending on the driving environment. In rural areas, these behaviors are generally aligned, forming a clear eco-safe driving profile. On highways, while a majority of trips demonstrate efficient and safe characteristics, a distinct group displays high distraction levels, indicating that fuel-efficient driving does not always coincide with safety. In urban contexts, a more complex pattern emerged—certain drivers exhibited conservative yet fuel-intensive behavior, while others achieved both safety and efficiency, albeit less frequently.

These results underscore the importance of context-specific strategies for promoting sustainable mobility. Rather than applying uniform behavioral recommendations, interventions should consider the diverse dynamics of driving environments. Future research should build upon this work by incorporating additional behavioral indicators and real-time feedback mechanisms to further improve safety and environmental outcomes in transport systems..

5. Acknowledgements

This project is carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI).

6. References-Bibliography

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.

European Commission (2025). Sustainable urban mobility. Available: https://transport.ec.europa.eu/transport-themes/urban-transport/sustainable-urban-mobility_en (accessed 11/02/2025).

European Court of Auditors, 2020: 'Sustainable Urban Mobility in the EU: No substantial improvement is possible without Member States' commitment'.

Eurostat. (2024, October 8). Energy consumption in transport at pre-pandemic levels. European Commission. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20241008-1

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algo-rithm. Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1), 100-108.

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.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. Safety Science.

Mantouka, E., Barmpounakis, E., Vlahogianni, E., & Golias, J. (2021). Smartphone sensing for understanding driving behavior: Current practice and challenges. International journal of transportation science and technology, 10(3), 266-282.

Nikolaou, D., Ziakopoulos, A., & Yannis, G. (2023). A review of surrogate safety measures uses in historical crash investigations. Sustainability, 15(9), 7580.

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.

Singh, H., & Kathuria, A. (2021). Profiling drivers to assess safe and eco-driving behavior–A systematic review of naturalistic driving studies. Accident Analysis & Prevention, 161, 106349.

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.

World Health Organization (WHO) (2023, 13 December). Road traffic injuries. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modelling for the re-gional effect of enforcement on road accidents. Accident Analysis & Prevention, 39(4), 818-825.

Zhang, Y., Fu, R., Guo, Y., & Yuan, W. (2022). Environmental screening model of driving behavior for an electric bus entering and leaving stops. Transportation Research Part D: Transport and Environment, 112, 103464.

Zhou, M., Jin, H., & Wang, W. (2016). A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing. Transportation Research Part D: Transport and Environment, 49, 203-218.

Ziakopoulos A., Kontaxi, A.,Yannis, G., Fortsakis, P., Kontonasios, K. N., & Kostou-las, G. (2020). Advanced driver monitoring using smartphone applications: The BeSmart project. Proceedings of the 8th Transport Research Arena TRA, 27-30.