Identifying crucial factors of the impact of COVID-19 on driving behaviour using feature analysis on naturalistic driving data

Marios Sekadakis[a1], Christos Katrakazas[a], Eva Michelaraki[a], Apostolos Ziakopoulos[a] George Yannis[a]

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

Abstract

This paper tries to identify and investigate the most significant factors in the entire 2020 that influenced the relationship between the COVID-19 pandemic metrics (i.e., COVID-19 cases, fatalities and reproduction rate) and restrictions (i.e., stringency index and lockdown measures) with driving behaviour. To that aim, naturalistic driving data for a 12-month timeframe were exploited and analyzed. The examined driving behaviour variables were harsh acceleration events and harsh braking events concerning a time period before, during and after the lockdown measures in Greece. The harsh events were extracted using data obtained by a specially developed smartphone application which were transmitted to a back-end telematic platform between 1st of January and 31st of December, 2020. Based on the collected data, XGBoost feature analysis algorithms were deployed in order to obtain the most significant factors. Furthermore, a comparison among the first COVID-19 lockdown (i.e., February to May 2020), the second one (i.e., August to November 2020) and the period without COVID-19 restrictions was drawn. Results revealed the impact of COVID-19 metrics and restrictions on driving behaviour and the indisputable relation with other factors (i.e., distance travelled, mobile use, driving requests, driving during risky hours). Furthermore, the differences and similarities of the harsh events between the two lockdown periods were identified. This paper tries to fill this gap in existing literature concerning a feature analysis for the entire 2020 and including the first and second lockdown restrictions of the COVID-19 pandemic in Greece.

Keywords: COVID-19 pandemic; Driving Behaviour; Harsh Brakings; Harsh Accelerations; Feature Analysis; XGBoost

[1] * Corresponding author. Tel.: +030-210-772-1575;
E-mail address: msekadakis@mail.ntua.gr

1. Introduction

The COVID-19 pandemic has affected human patterns since December 2019 [1] and continues incessantly for more than two years since the beginning. Right from the beginning, many countries around the world imposed strict measures, such as lockdown and suspension of all non-essential movements, in order to reduce human activity which contributes to the spread of the pandemic. In this direction, many studies seek to explore the dynamics of the pandemic in several countries around the world to understand the impact that COVID-19 had on the transport sector [2]. The restriction measures affected typical patterns of travel activities and mobility in urban regions across the world [3]. In the Netherlands, people reduced their outdoor activities due to the pandemic, leading to a decrease in the total number of trips and reduction in distance travelled, with an increase in the proportion of people working from home [4].

Existing studies have shown that there was a major change in the choice of transport mode, especially at the first pandemic wave, and consequently a change in the number of car-driven volumes was observed [5]. Following the restrictive measures taken by governments to restrict the spread of the disease, an unprecedented decline in traffic volumes has been identified [6], [7].

In the context of road safety, during the COVID-19 lockdown measures, the number of road collisions, injuries, and fatalities has significantly decreased, especially during the first lockdown period. In particular, in the Spanish province of Tarragona, a sharp decrease in traffic crashes was revealed [8]. Similarly, Carter [9] showed that during the first COVID -19 period (i.e., from March 15, 2020 to May 16, 2020), the total number of crashes in North Carolina decreased by half, fatalities decreased by 10%, and serious injuries increased by 6%, compared to the pre-closure baseline. This can probably be attributed to the aforementioned general reduction in traffic volumes [10]. A relevant study [11] indicated that traffic crashes, including injury and fatality crashes, decreased on state highways and rural roads. Nevertheless, a more profound study used time-series to predict the road collisions, injuries and fatalities that would have been observed without the existence of the COVID-19 pandemic. Predictions made clear that the reduction of fatalities and injuries was disproportionate taking into account and comparing the reduction in traffic volumes [10].

Driving behaviour has also changed during the pandemic which therefore has a great impact on road safety. Many studies have reported a change in driving behaviour indicators [7], [12], [13]. According to this study [7], by exploiting driving data from the first lockdown period in Greece and Saudi Arabia, increased driving speed (6-11%) was noted, with more frequent harsh accelerations and brakings per distance. Nevertheless, very few studies investigated driver behaviour more profoundly by analyzing and modeling driving data [12]. This innovative study quantified the impact of the pandemic COVID-19 on driving behaviour using SARIMA time series modeling. The results showed that the observed values of three indicators of driving behaviour (i.e., average speed, speeding, and harsh braking per 100 km) were higher than the predicted values based on the corresponding observations before the first lockdown period in Greece. Moreover, in the same study, a machine learning (ML) approach of XGBoost was deployed to identify the most influential COVID-19 indicators.

In this direction, the current study aims to identify and investigate the most significant factors in the entire 2020 that influenced the relationship between the COVID-19 pandemic metrics (i.e., COVID-19 cases, fatalities and reproduction rate) and restrictions (i.e., stringency index and lockdown measures) with driving behaviour. For this purpose, naturalistic driving data for a 12-month timeframe were exploited and analyzed. The examined driving behavior variables were harsh acceleration and harsh braking events concerning a time period before, during and after the lockdown measures in Greece. The motivation is to cover the literature gap by giving insights on these two driving behaviour indicators and how they varied for the entire 2020. A cross-lockdown comparison was also provided insights into how they varied across the examined conditions (i.e., no restrictions, 1st lockdown, 2nd lockdown).

The paper structure is presently briefly: at the beginning, literature findings with regards to mobility, traffic volumes, road safety and driving behaviour change during the COVID-19 pandemic are reviewed. Subsequently, the methodology section follows overviewing the obtained dataset for this study, descriptive statistics of the examined variables, COVID-19 restriction measures and the chosen ML technique background are presented. Then, the analysis results are provided which are divided it into two subsections for each model: i) harsh acceleration events ii) harsh braking events. Finally, the main findings, conclusions for further research and recommendations are also highlighted.

2. Methodology

2.1 Data Overview

In order to correlate driving behaviour with COVD-19 metrics and restrictions, OSeven provided a random dataset with naturalistic driving trips from its database. The time span of the database was from 01/01/20120 to 31/12/2020 and included thousands of trips throughout Greece. The aforementioned one-year dataset contains data before-during-after the appearance of COVID-19 (the first COVID-19 case in Greece was diagnosed on 26/02/2020) and two lockdown restriction measures for non-essential movements. OSeven exploits data from smartphone sensors (e.g. GPS, accelerometer data, and gyroscope data) using the smartphone applications and platform developed by OSeven Telematics (oseven.io). For each trip completed, a large amount of data was recorded, transmitted through Wi-Fi or cellular network and valuable critical information such as features, highlights and driving scores was produced in order to evaluate driving profile and performance. Subsequently, data were sent to the OSeven backend infrastructure where there were evaluated using filtering, signal processing, ML algorithms and safety/eco scoring models. Five dedicated variables (i.e., harsh accelerations (HA) /100km, harsh brakings (HB) /100km, mobile use/ driving time, driving during risky hours, distance) were exploited from the OSeven dataset and their explanation can be found in Table 1. The OSeven platform has clear privacy policy statements and follows strict information security procedures, in compliance with the General Data Protection Regulation (GDPR) and related EU directives. Thus, all data has been provided by OSeven in a completely anonymized format and no geolocation information for the trips has been included in the dataset.

Apart from the OSeven dataset which provided a total of approximately 305,000 trips (randomly chosen) and in order to correlate them with COVD-19 metrics and restrictions, other three datasets with daily observations were exploited. One dataset was the dataset of Our World in Data, 2020 (OWD) which was exploited in order to capture the daily evolution of COVID-19 metrics in 2020 i.e., new cases, new fatalities, the COVID-19 reproduction rate of the pandemic.

The response measures of the Greek government were quantified with an index called the “Stringency Index”. This index was obtained and calculated by Oxford University. This index is open access and can be found in the COVID-19 government response tracker [15], [16]. Specifically, the stringency index ranges between 0 and 100 and represents the government response stringency index and composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (i.e., 100 = strictest response).

In order to be able to provide an overview of the COVID-19 impact on driving patterns, the mobility data reports from Apple [17] were used and specifically the driving requests as a surrogate measurement of the traffic mobility. The aggregated data collected from Apple Maps and show the mobility trends for major cities and several countries or regions. The information is generated by aggregating the number of daily driving requests made by the Apple Maps users who requested navigation. These requests are expressed by the percentage change compared to a baseline of 100% on January 13th, 2020 prior to COVID-19 appearance. The driving indicators examined are summarized in Table 1.

Table 1: Variables Units, Description and Source

Variable	Unit	Description	Source
Harsh accelerations (HA) /100km	events/km	Number of harsh accelerations per distance (100 km)	OSeven
Harsh brakings (HB) /100km	events/km	Number of harsh brakings per distance (100 km)	OSeven
Distance	km	Total trip distance	OSeven
Mobile Use/ Driving Time	0-100 %	Total duration of mobile usage in a trip/ Trip Duration	OSeven
Driving during Risky Hours	km	Distance driven in risky hours (00:00 - 05:00) in a trip	OSeven
New COVID-19 Cases	count	New confirmed cases of COVID-19	OWD
New COVID-19 Fatalities	count	New fatalities attributed to COVID-19	OWD
COVID-19 Reproduction Rate	-	Real-time estimate of the effective reproduction rate (R) of COVID-19	OWD

Variable	Unit	Description	Source
Stringency Index	0-100	Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response)	Oxford
Apple Driving Requests	% change	Requests for driving (%) (100% - baseline on January 13th, 2020)	Apple

Table 2 presents the descriptive statistics of the investigated variables, i.e., mean, standard deviation, maximum value, minimum values for the random subset of trips (305,638 trips). More specifically, 16,927 trips (5.5% of the total) were observed at the 1st lockdown and 42,262 trips (13.8%) at the 2nd. It is worth noting that all the under-investigation variables are continuous and therefore there was no need for special practice during the analysis in contrast to the discrete variables. The sample size was different for COVID-19 metrics, measures and mobility compared to driving data as they had daily observations for the entire 2020. The datasets were merged for analysis purposes, specifically for each trip (provided by OSeven) as well as the daily value of the rest datasets was assigned (i.e., OWD, Oxford and Apple datasets).

Table 2: Descriptive Statistics of Investigated Variables

Variable	*Mean*	*SD*	*Min*	*Max*	*Sample Size*
Harsh Accelerations (HA) /100km	9.36	17.37	0	99.98	305,638
Harsh Brakings (HB) /100km	13.59	19.76	0	99.99	305,638
Distance	13.28	23.65	0.50	648.69	305,638
Mobile Use/ Driving Time	0.05	0.14	0	1.00	305,638
Driving during Risky Hours	0.42	4.32	0	427.70	305,638
New COVID-19 Cases	363.40	662.56	0	3316.00	366
New COVID-19 Fatalities	12.36	26.93	0	121.00	366
COVID-19 Reproduction Rate	0.83	0.52	0	1.48	366
Stringency Index	48.17	28.11	0	84.26	366
Apple Driving Requests	114.35	52.57	18.59	241.14	364

SD: Standard Deviation

2.2 COVID-19 Restriction Measures

Table 3 summarizes the two lockdown periods of non-essential movements due to the COVID-19 pandemic that have been announced by the Greek government.

Table 3: Lockdown Measures and important Dates

Greece – Lockdown Measures	
1st Lockdown restrictions on non-essential movements	23-03-2020→04-05-2020
2nd Lockdown restrictions on non-essential movements	07-11-2020→31-12-2020 (Continued in 2021)

The two lockdowns of 2020 are included in Figure 1 in gray shades. Furthermore, the figure illustrates the evolution through time of driving mobility volumes (i.e., driving requests) in relation to COVID-19 new cases, stringency index of measures, and lockdown periods. An initial observation is that driving requests were significantly reduced during both lockdowns. Nevertheless, the greatest reduction in driving requests was during the first lockdown. Also, there was a spike of nearly 250% (150% more compared to the 100% baseline of January 13th) in the requests in the first half of August. These observations were further analyzed and elaborated by combining the chosen analysis on OSeven naturalistic driving data. In addition, driving data were analyzed descriptively for lockdowns comparison.

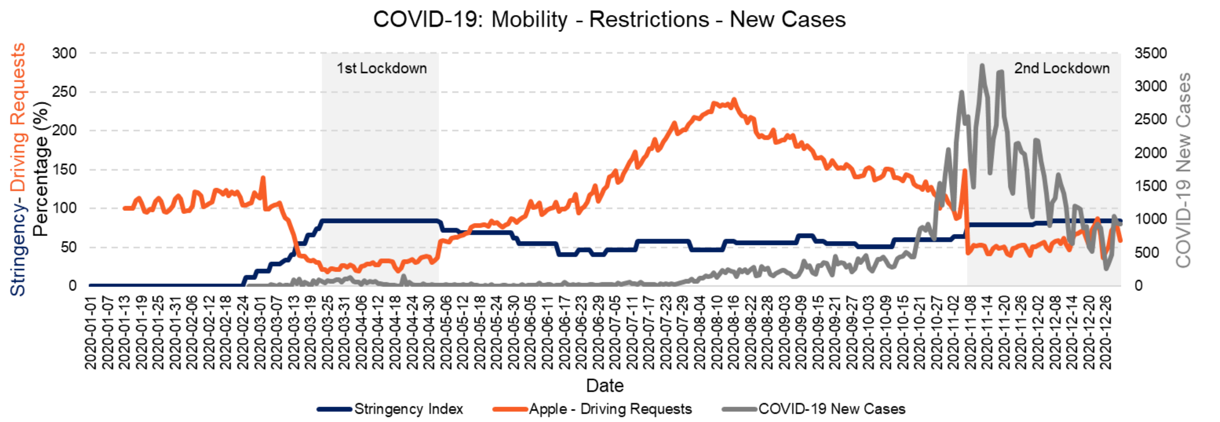

Figure 1: COVID-19 Overview of mobility along with restrictions (lockdown and stringency index) and new COVID-19 cases

2.3 XGBoost Analysis

As an analysis method was chosen the Extreme Gradient Boosting (XGBoost) algorithms. XGBoost algorithms were deployed in order to evaluate the feature importance of the aforementioned variables, i.e., mobility, COVID-19 metrics and restrictions in regards to the naturalistic driving behavior indicators. The naturalistic driving behavior indicators were the frequency of harsh events, such as harsh brakings and harsh acceleration per distance (100km). It should be mentioned that XGBoost is a supervised ML technique and the user defines the independent/dependent variables. The learning process of the algorithm is iterative and consequently involves correcting previous errors in future iterations of the algorithm. A detailed overview of the comprised parameters and technical specifications of the algorithm in the used library can be found in [18]. XGBoost analysis was used because it has been shown to be superior in accuracy compared to logistic regression models or even other ML methods such as Random Forests, Artificial Neural Network, Support Vector Machines, both in the area of traffic safety [19].

In addition, the XGBoost algorithms have the capability to calculate the importance of each predictor variable in the developed model. In the XGBoost algorithm, the following three variable importance metrics were extracted [18]. These variable importance metrics are used by the XGBoost algorithms in the analysis to show which variables are informative in describing the driving behavior indicators (HA and HB /100km):

- Gain describes the enhancement in accuracy that a feature adds to its branches.
- Cover describes the relative amount of observations (or the number of samples) concerned by a feature.
- Frequency describes how often a feature is used in all generated trees.

The gain metric is used for feature importance illustration as shown in Figures 2 and 4.

3. Analysis and Results

An initial description of the parameters as provided in the XGBoost algorithm is given. Firstly, a random split was employed in the data (as described in section 2); 75% was the training set, while the remaining 25% was the test set. Moreover, all the outliers were identified and then removed from the dataset, creating a clean undistorted analysis. Furthermore, multiple values in terms of learning rate (eta) were tested (0.01-0.3) for each XGBoost for extracting the optimal model for harsh events. Additionally, K-fold cross validation was conducted in order to find the number of best iteration within the XGBoost algorithm. Specifically, cross-validation involves splitting the data set into parts; 75% to train the model and the remaining data (25%) is not used for backpropagation but is used to determine a test error. If this error stops improving (or in most cases worsens), it is a sure sign that the model is overfitting - so the training should stop at this point [20].

The defined parameters for the XGBoost model for harsh events are provided as follows:

- Learning rate (eta)= 0.01-0.3
- Gamma= 1
- Maximum depth of a tree= 6
- Subsample ratio of the training instances= 0.8

- Subsample ratio of columns when constructing each tree= 0.5

3.1 Harsh Acceleration Events

In this subsection, the results of the XGBoost for Harsh Accelerations (HA) /100km are presented and these outcomes are further elaborated in the discussion section. Firstly, the predictive power and accuracy provided by the application of the XGBoost algorithms on the test subset can be extracted by the achieved error. Table 4 presents the accomplished error i.e., ME=0.081, RMSE = 17.314 and MAE = 12.012.

Table 4: Errors on Test Predictions

	ME	RMSE	MAE
Test set	0.081	17.314	12.012

The obtained feature importance is provided in Table 5 and Figure 2. The top three variables that impacted the most within HA/100km model were; distance, mobile use/ driving time, driving requests. Also, a small contribution was also provided by driving during risky night-time hours. Then, the COVID-19-related variable of new COVID-19 cases in Greece seems to precede compared to other COVID-19-related variables. Other COVID-19-related variables that influenced the harsh accelerations in Greece were COVID-19 Reproduction Rate, Stringency Index, and New COVID-19 Fatalities. The influence of predicting harsh events is expressed by the gain scores of XGBoost (Table 5 and Figure 2).

Table 5: Feature importance of HA/100km - XGBoost algorithms

Feature	Gain	Cover	Frequency
Distance	0.531	0.364	0.242
Mobile use/ Driving time	0.207	0.212	0.198
Apple Driving Requests	0.086	0.174	0.161
New COVID-19 Cases	0.060	0.083	0.115
Driving during Risky hours	0.039	0.053	0.107
COVID-19 Reproduction Rate	0.031	0.042	0.078
Stringency Index	0.023	0.029	0.045
New COVID-19 Fatalities	0.023	0.043	0.054

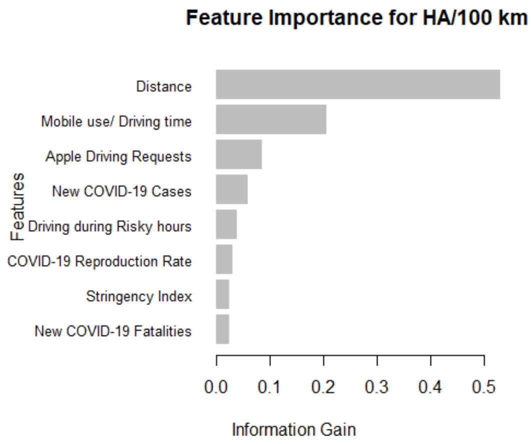

Figure 2: Feature importance of HA/100km - Information Gain

Boxplots were created supplementary to XGBoost, and they can be seen in Figure 3, in order to reveal the trend of harsh accelerations under these three different restriction measures of 2020, something that XGBoost could not reveal directly from feature importance. In general, the boxplot shows the median, interquartile range, minimum, maximum values of completion time for each measure. Figure 3 (left), presents the boxplot for harsh accelerations including the whole dataset. As can be seen in the boxplot, the median values for each condition equal zero. To that end, an additional boxplot in Figure 3 (right) was created by excluding the zero values of the dataset, hence this boxplot presents only the trips with harsh events occurrence and the findings are discussed below. The 2nd lockdown in Greece has a narrower interquartile range than the other conditions (i.e., 1st lockdown and without restrictions). This means that the upper quartile of the 2nd lockdown is lower than the other conditions. Also, the 1st lockdown has a higher upper quartile compared to without restrictions and the 2nd lockdown. Figure 3 (right), as mentioned previously, the zero values were excluded from the dataset and present only the trips with harsh events occurrence, the highest median was observed at the 1st lockdown, then at 2^{nd,} and then without restrictions.

Figure 3: (left) Harsh Accelerations/100km under different restriction measures (right) Harsh Accelerations/100km under different restriction measures by excluding zero values

3.2 Harsh Braking Events

In this subsection, the results for Harsh Brakings (HB)/100km are presented and, as mentioned previously, these outcomes are further elaborated in the subsequent section. Table 6 presents the accomplished error for this model i.e., ME=-0.025, RMSE = 19.529 and MAE = 14.561.

Table 6: Errors on Test Predictions

	ME	RMSE	MAE
Test set	-0.025	19.529	14.561

The obtained feature importance is provided in Table 7 and Figure 4. Similar to the harsh accelerations model, the top three variables that impacted the most the HB were; distance mobile use/ driving time, driving requests. A small contribution was also provided by driving during risky night-time hours. However, the COVID-19-related variable that influenced the most HB in Greece is different than the HA model. COVID-19 Reproduction Rate was found to influence the most HB. Other COVID-19-related variables that influenced the harsh brakings in Greece were New COVID-19 Cases, Stringency Index, and New COVID-19 Fatalities.

Table 7: Feature importance of HB/100km - XGBoost algorithms

Feature	Gain	Cover	Frequency
Distance	0.608	0.368	0.250
Mobile use/ Driving time	0.102	0.132	0.182
Apple Driving Requests	0.078	0.151	0.140
COVID-19 Reproduction Rate	0.063	0.082	0.085
New COVID-19 Cases	0.053	0.101	0.124
Driving during Risky hours	0.039	0.077	0.117
Stringency Index	0.039	0.053	0.051
New COVID-19 Fatalities	0.019	0.035	0.051

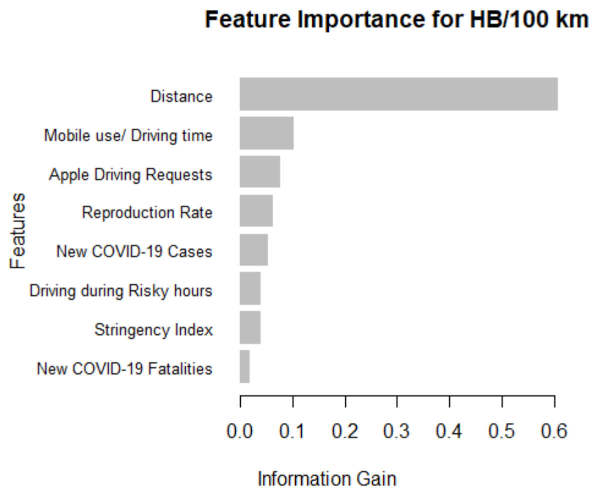

Figure 4: Feature importance of HB/100km - Information Gain

Figure 5 (left), presents the boxplot for harsh accelerations including the entire dataset. As can be seen in this boxplot, the highest median value was observed during the 1st lockdown. Then, the conditions without restrictions follow and it is noteworthy that the median for the 2nd lockdown equals zero. The box of the 2nd lockdown in Greece has a narrower interquartile range than the other conditions (i.e., 1st lockdown and without restrictions). This means that the upper quartile of the 2nd lockdown is lower than the other conditions. Also, the 1st lockdown has a higher upper quartile compared to without restrictions and the 2nd lockdown. Figure 5 (right), similarly to the HA model, the highest median was observed at the 1st lockdown, then at 2nd, and then without restrictions.

Figure 5: (left) Harsh Brakings/100km under different restriction measures (right) Harsh Brakings/100km under different restriction measures by excluding zero values

4. Discussion

The paper aims to identify and investigate the most significant factors in the entire 2020 that influenced the relationship between the COVID-19 pandemic metrics (i.e., COVID -19 cases, fatalities, and reproduction rate) and restrictions (i.e., stringency index and lockdown measures) with driving behaviour. The results of the exploratory analysis by XGBoost suggest a correlation of COVID-19 metrics and restriction measures with driving behaviour. Furthermore, different patterns were revealed for both harsh events among the three examined conditions, i.e., without restrictions, 1st lockdown, and 2nd lockdown.

The top three variables that influenced the most HA and HB events were common for both types of events namely; distance, mobile use/ driving time, and driving requests (requested in Apple Maps). More specifically, trip distance and mobile use duration were the two most important factors out of the eight examined variables that influence HA and HB, and this finding is consistent with [12] in which they investigated the influence on HB events only during the 1st lockdown in Greece. It is worth noting that trip distance had a great impact on HA and HB events probably due to the fact that the longer trips were driven on highways and rural roads. Hence, the change in road type probably influences the drivers' braking and acceleration patterns with more or less frequent harsh events. Another causal factor for the correlation between harsh events and duration was the increasing fatigue by increasing the trip distance. However, these assumptions need further research in order to be validated. Additionally, mobile phone use, which reveals drivers' distraction, shows the importance of the drivers to be undistracted in order to avoid HA and HB events. After trip duration and mobile phone use, driving requests follow which are a driving exposure measurement and is an indication of the prevailing traffic volumes [10] and this finding reveals the relation between this exposure measurement with HA and HB events. A small contribution on HA and HB was also provided by driving during risky nighttime hours indicating that there was a change in events during nighttime driving (00:00 - 05:00) and this finding is consistent with existing literature [12].

The aforementioned top three variables were extraneous with COVID-19 variables, and this is clear since the COVID-19 pandemic had no direct effect and causality on driving behaviour. Nevertheless, four COVID-19-related variables were found to impact HA and HB events. New COVID-19 cases in Greece seemed to precede compared to other COVID-19-related variables in terms of HA events. Interestingly, on the contrary to HA, COVID-19 Reproduction Rate was found to influence the most HB events. This is proof that COVID-19 metrics and restriction measures impacted driving behaviour since the COVID-19-related variables that influenced the HA and HB events in Greece were COVID-19 Reproduction Rate, Stringency Index, and New COVID-19 Fatalities and Cases.

Before analyzing the boxplot findings, it is necessary to investigate how the driving exposure measurement of Apple driving requests in Figure evolved through time in 2020 in Greece. It can be concluded from Figure 1 that driving requests were significantly decreased during both lockdowns compared to the baseline of no restrictions.

The greatest reduction was observed in the first lockdown compared to the second. This means that the traffic volume during the 1st lockdown was lower than the other conditions and hence with fewer vehicles ahead, the drivers could accelerate more easily and this can be revealed based on Figure 3 (left); the upper percentile was higher than other conditions. Additionally, in Figure 3 (right) for trips with harsh accelerations occurrence, the median was higher than the other conditions meaning that the HA events were more frequent even at the trips with maximum values. This finding is also consistent with the literature [7]. With regards to the 2nd lockdown, for trips with harsh accelerations, the median was higher compared to without restriction conditions as a result of the decreased traffic volume but not the same magnitude as the 1st in which the traffic volume was lower.

With regards to HB events, in Figure 5 (left), again, the upper percentile is greater than other conditions as well as for trips with harsh brakings, Figure 5 (right), the median is higher than the other conditions meaning that the HB events were more frequent even at the trips with maximum values. This finding is also consistent with the literature [7]. This can be explained as the traffic volume during the 1st lockdown was lower than the other conditions and hence with fewer vehicles ahead, the drivers could maintain higher speeds as stated in [7] and it was more probable for the drivers to be involved in a harsh braking event with a higher speed. With regards to the 2nd lockdown following the same logic as HA, for trips with harsh brakings, the median was higher compared to no restrictions as a result of the decreased traffic volume but not the same magnitude as the 1st in which the traffic volume was lower.

Nevertheless, this work is not without shortcomings, and therefore, future research could focus on covering the remaining gaps that this work did not cover. Initially, future studies could concentrate on more sophisticated models, such as deep neural networks, e.g., Convolutional neural networks (CNN) or Artificial Neural Networks (ANNs), or other artificial intelligence techniques, which probably can accomplish lower errors and give more insights into driving behaviour variables. In addition, more variables with regards to driving behaviour, i.e., speeding, speeding duration, and speed, could be exploited using the same method in order to give in the same context results. These variables were tested but they presented a great error and therefore did not include in this work. Nevertheless, the aforementioned sophisticated models could successfully predict these variables. Additional data with geolocation information could lead to, an addition to the current method, spatial analysis of the examined variables which would give significant spatial outcomes. Lastly, the current analysis could be combined with road type data and all these could provide insights into driving behaviour for each road type.

5. Conclusions

In order to accomplish this study, naturalistic driving data for a 12-month timeframe were exploited and analyzed. The examined driving behaviour variables were HA and HB events concerning a time period before, during and after the lockdown measures in Greece. The naturalistic driving data were extracted using data obtained by a specially developed smartphone application and were transmitted to a back-end telematic platform (OSeven). The top three variables that influenced the most HA and HB events were mutual for both types of events namely; distance, mobile use/ driving time, and driving requests (as requested in Apple Maps). These top three variables were extraneous with COVID-19 variables, and this is clear since the COVID-19 pandemic had no direct effect and causality on driving behaviour. Furthermore, a small contribution on HA and HB was also provided by driving during risky nighttime hours indicating that there was a change in events during nighttime driving (00:00 - 05:00).

Focusing on the COVID-19-related variables, this study identified the most significant factors in the entire 2020 that influenced the relationship between the COVID-19 pandemic metrics (i.e., COVID-19 new cases, new fatalities, and reproduction rate) and restrictions (i.e., stringency index) with driving behaviour. The results of the exploratory analysis by XGBoost indicate a correlation of COVID-19 metrics and restriction measures with harsh brakings and accelerations. Furthermore, for all the investigated three conditions, i.e., no restrictions, 1st lockdown, and 2nd lockdown, different HA and HB event patterns were revealed. Taking the aforementioned into account, it can be concluded that HB and HA (for trips with harsh events occurrence) were increased and more frequent during lockdown restrictions due to their correlation with driving exposure measurements (i.e., Apple driving requests).

Acknowledgment

The authors would like to thank OSeven Telematics, London, UK (oseven.io) for providing all the necessary naturalistic driving data that assisted in accomplishing this study.

References

[1] N. Zhu *et al.*, “A Novel Coronavirus from Patients with Pneumonia in China, 2019,” *N. Engl. J. Med.*, vol. 382, no. 8, pp. 727–733, 2020, doi: 10.1056/nejmoa2001017.

[2] A. Sharifi and A. Reza Khavarian-Garmsir, “The COVID-19 pandemic : Impacts on cities and major lessons for urban planning , design , and management,” *Sci. Total Environ.*, vol. 749, pp. 1–3, 2020, doi: 10.1016/j.scitotenv.2020.142391.

[3] K. Kim, “Impacts of COVID-19 on transportation: Summary and synthesis of interdisciplinary research,” *Transp. Res. Interdiscip. Perspect.*, vol. Volume 9, no. 100305, 2021.

[4] M. de Haas, R. Faber, and M. Hamersma, “How COVID-19 and the Dutch ‘intelligent lockdown’ change activities, work and travel behaviour: Evidence from longitudinal data in the Netherlands,” *Transp. Res. Interdiscip. Perspect.*, vol. 6, p. 100150, 2020, doi: 10.1016/j.trip.2020.100150.

[5] P. Bucsky, “Modal share changes due to COVID-19 : The case of Budapest,” *Transp. Res. Interdiscip. Perspect.*, vol. 8, p. 100141, 2020, doi: 10.1016/j.trip.2020.100141.

[6] F. Aletta *et al.*, “Analysing urban traffic volumes and mapping noise emissions in Rome (Italy) in the context of containment measures for the COVID-19 disease,” *Noise Mapp.*, vol. 7, no. 1, pp. 114–122, 2020, doi: doi:10.1515/noise-2020-0010.

[7] C. Katrakazas, E. Michelaraki, M. Sekadakis, and G. Yannis, “A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety,” *Transp. Res. Interdiscip. Perspect.*, vol. 7, p. 100186, 2020, doi: 10.1016/j.trip.2020.100186.

[8] Ò. Saladié, E. Bustamante, and A. Gutiérrez, “Transportation Research Interdisciplinary Perspectives COVID-19 lockdown and reduction of traf fi c accidents in Tarragona province , Spain,” *Transp. Res. Interdiscip. Perspect.*, vol. 8, p. 100218, 2020, doi: 10.1016/j.trip.2020.100218.

[9] D. Carter, “Effects of COVID-19 Shutdown on Crashes and Travel in NC,” 2020.

[10] M. Sekadakis, C. Katrakazas, E. Michelaraki, F. Kehagia, and G. Yannis, “Analysis of the impact of COVID-19 on collisions, fatalities and injuries using time series forecasting: The case of Greece,” *Accid. Anal. Prev.*, vol. 162, no. July, p. 106391, 2021, doi: 10.1016/j.aap.2021.106391.

[11] F. Shilling and D. Waetjen, “Impact of COVID19 mitigation on numbers and costs of California Traffic Crashes,” vol. 2020, 2020.

[12] C. Katrakazas, E. Michelaraki, M. Sekadakis, A. Ziakopoulos, A. Kontaxi, and G. Yannis, “Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting,” *J. Safety Res.*, vol. 78, pp. 189–202, 2021, doi: 10.1016/j.jsr.2021.04.007.

[13] E. Michelaraki, M. Sekadakis, C. Katrakazas, A. Ziakopoulos, and G. Yannis, “A four-country comparative overview of the impact of COVID-19 on traffic safety behavior,” *10th Int. Congr. Transp. Res.*, Rhodes, Greece, 2021.

[14] Our World in Data, “Coronavirus Pandemic (COVID-19),” 2020. https://ourworldindata.org/coronavirus (accessed Dec. 08, 2020).

[15] T. Hale *et al.*, “Variation in government responses to COVID-19,” *BSG Work. Pap. Ser. Blavatnik Sch. Gov. Univ. Oxford*, p. Version 8.0, 2020.

[16] T. Hale *et al.*, *Variation in government responses to COVID-19, Version 12.0*. 2021.

[17] Apple, “COVID-19 - Mobility Trends Reports - Apple [WWW Document],” *Accessed 11/06/20*, 2020. https://www.apple.com/covid19/mobility (accessed Jun. 11, 2020).

[18] XGBoost Documentation, “dmlc XGBoost,” 2021. https://xgboost.readthedocs.io/en/latest/index.html#xgboost-documentation (accessed Nov. 26, 2021).

[19] Y. Huang and S. Meng, “Automobile insurance classification ratemaking based on telematics driving data,” *Decis. Support Syst.*, vol. 127, no. January 2019, p. 113156, 2019, doi: 10.1016/j.dss.2019.113156.

[20] Andy Elmsley, “You’re fit and you know it: overfitting and cross-validation,” 2019. https://medium.com/the-sound-of-ai/youre-fit-and-you-know-it-overfitting-and-cross-validation-90a3a9f67c74.