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*

1. Introduction

The pandemic of COVID-19 has been affecting human activities, since December 2019. The majority of governments around the world implemented harsh measures, such as lockdowns, from the start in order to decrease human activity and thus prevent the pandemic spread. Existing research has found that there was a significant shift in the mode of transportation, particularly during the first pandemic wave, and that this resulted in a shift in traffic volumes [1]. Driving behavior was also affected as a result of the pandemic, which implies a significant impact on road safety. Many studies found that driving behavior indicators were significantly affected [2]–[4]. According to [3], greater driving speed (6-11%) was recorded in Greece and Saudi Arabia during the first lockdown period, with more frequent harsh accelerations and brakings per distance. Nonetheless, few studies have looked into driving behavior in greater depth by analyzing and modeling driving data [2]. The literature findings revealed that the observed values of three driving behavior indicators (average speed, speeding, and harsh braking per 100 km) were higher than forecasted values compared to the observations prior to the Greek first lockdown [2].

In this direction, the current study intends to identify and examine the most important elements including COVID-19 pandemic parameters (i.e., COVID-19 cases, fatalities, and reproduction rate) that influenced driving behavior in the year 2020. Naturalistic driving data for a 12-month period was used and analyzed for this purpose. The indicators studied were harsh acceleration and harsh braking events before, during, and after the imposition of lockdown measures in Greece. The goal is to fill a gap in the research by providing insights into these two driving behavior indicators and how they changed over the course of 2020. A cross-lockdown analysis was also conducted in order to reveal how indicators differed across the various restrictive conditions (i.e., no restrictions, 1st lockdown, 2nd lockdown).

2. Methodology

2.1 Data Overview

OSeven offered a random dataset with naturalistic driving trips from its database in order to associate driving behavior with COVD-19 parameters and restrictions. The database covered thousands of trips around Greece from January 1, 2020 to December 31, 2020. OSeven Telematics (oseven.io) uses its specially developed smartphone application to exploit data from smartphone sensors (such as GPS, accelerometer, and gyroscope data). The OSeven dataset was used to extract five specific variables (harsh accelerations (HA)/100km, harsh brakings (HB)/100km, mobile use/driving duration, driving during risky hours, and distance), which are explained in Table 1.

Three other datasets with daily observations were used in addition to the OSeven dataset, which supplied a total of 305,000 trips (randomly chosen) in order to correlate them with COVD-19 metrics and restrictions. One dataset was obtained from “Our World in Data” (OWD) (Our World in Data, 2020), which was used to capture the daily evolution of COVID-19 metrics in 2020, such as new cases, new fatalities, and the COVID-19 reproduction rate. The stringency index metric compiled and calculated by Oxford University was used to quantify the Greek government's response actions. This index ranges from 0 to 100 (i.e., 100 for the strictest measures) and can be found in the COVID-19 government response tracker [6], [7]. The driving requests from Apple [8] were used as a

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

surrogate measurement of the traffic mobility deriving from the mobility data reports. The data is compiled by combining the number of daily driving requests from Apple Maps users who requested navigation. These demands are expressed as a percentage change from a baseline of 100% on January 13th, 2020, before the appearance of COVID-19. Table 1 summarizes the driving indicators studied.

Table 1: Variables Description

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
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

2.2 COVID-19 Restriction Measures

The two 2020 lockdowns are depicted in gray shades in Figure 1. The plot also shows how driving mobility volumes (i.e., driving requests) has changed over time in relation to COVID-19 new cases, stringency index of measures, and lockdown periods.

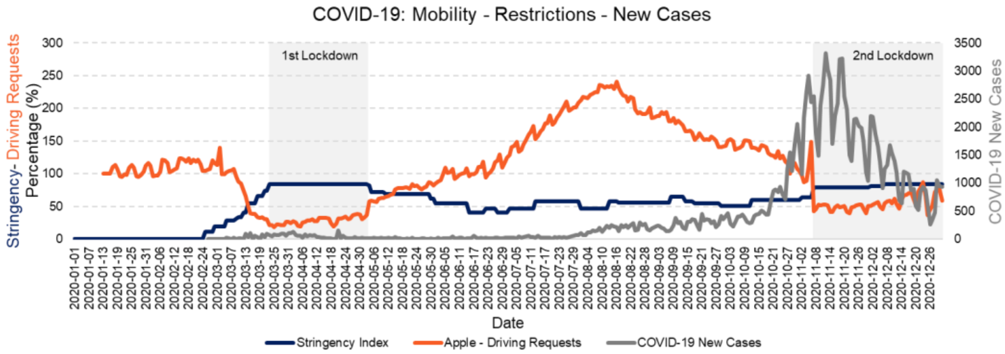

Figure 1: COVID-19 Overview of mobility, restrictions, and COVID-19 cases

2.3 XGBoost Analysis

In order to assess the feature importance of the aforementioned variables, specifically the correlation between mobility, COVID-19 metrics, and restrictions in relation to naturalistic driving behavior indicators, Extreme Gradient Boosting (XGBoost) algorithms were used. The frequency of harsh events, such as harsh braking and acceleration per distance (100km), were used as naturalistic driving behavior indicators. The following three variable important measurements (gain, cover, and frequency) were retrieved using the XGBoost algorithm [9]. The XGBoost algorithms, in the analysis, used these metrics to demonstrate which variables were informative in characterizing the driving behavior indicators (HA and HB /100km).

3. Analysis and Results

Firstly, the data was split randomly; 75% of the data was used for training, while the remaining 25% was used for testing. Furthermore, all outliers were found and deleted from the dataset, resulting in a clean, undistorted study. In addition, multiple learning rate (eta) values were tested (0.01-0.3) for each XGBoost in order to find the best model for harsh events. The number of best iterations inside the XGBoost algorithm was also determined using K-fold cross validation.

3.1 Harsh Acceleration Events

The results of the XGBoost algorithm for Harsh Accelerations (HA) /100km are reported, and the achieved error, ME=0.081, RMSE = 17.314 and MAE = 12.012 can be used to extract the predictive power and accuracy by the application of the XGBoost algorithms on the test subset.

Figure 2 shows the obtained feature importance. Distance, smartphone use/driving time, and driving requests were the top three variables that influenced the HA/100km model the most.

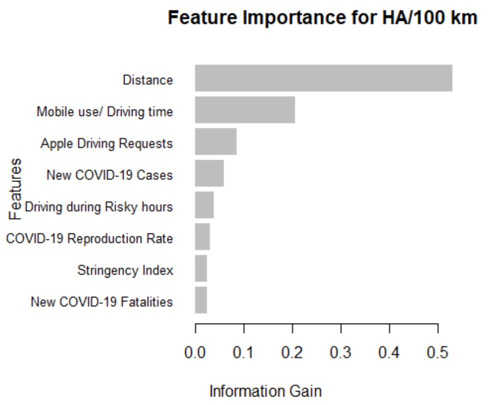

Figure 2: HA/100km feature importance

Boxplots were built as an addition to XGBoost, as shown in Figure 3, to depict the trend of harsh accelerations under these three different 2020 restriction measures, which XGBoost could not reveal directly from feature importance. For each metric, the boxplot displays the median, interquartile range, lowest, and maximum values of HA events. The boxplot for HAs, which includes the entire dataset, is shown in Figure 3 (left). By removing the zero values from the dataset, an additional boxplot in Figure 3 (right) was generated, which only shows the trips where harsh events occurred, and the findings are explained below. Figure 3 (right) shows the highest median at the first lockdown, then the second, and finally no restrictions.

Figure 3: (left) HA/100km for different measures
(right) HA/100km for different measures (excluding zero values)

3.2 Harsh Braking Events

The results for Harsh Brakings (HB)/100km are shown in this subsection. The errors of the developed XGBoost model were ME=-0.025, RMSE = 19.529, and MAE = 14.561.

Figure 4 shows the feature importance. The top three variables that impacted the frequency of HB the most, similar to the hard accelerations model, were distance, mobile use/ driving time, and driving requests.

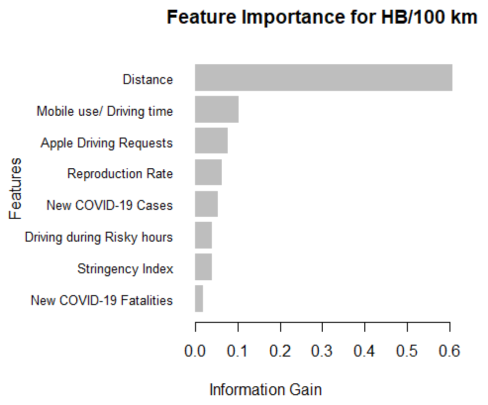

Figure 4: HB/100km feature importance

The highest median value was obtained during the first lockdown, as shown in Figure 5 (left). Then when there were no restrictive conditions, and it is worth noting that the median for the second lockdown is zero. Figure 5 (right) shows that the highest median was seen at the first lockdown, then at the second, and finally without restrictions.

Figure 5: (left) HB/100km for different measures (right) HB/100km for different measures (excluding zero values)

4. Discussion

Distance, mobile phone use/driving time, and driving requests (requested in Apple Maps) were the top three variables that influenced HA and HB events the most. More specifically, out of the eight variables evaluated, trip distance and mobile use duration were the two most important indicators that influenced HA and HB, and this finding is consistent with [2] which focused on the influence on HB events during the first lockdown in Greece. It is worth noting that trip distance had a significant impact on HA and HB events, due to the lengthier trips that are driven on highways and rural roads. As a result, changes in road type are likely to affect braking and accelerating habits of drivers, which may induce 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, more research is needed to support these outcomes. Furthermore, the need for drivers to remain undistracted in order to avoid HA and HB events is demonstrated by mobile phone use, which indicates their distraction. After the important indicators of trip time and mobile phone use, driving requests follow which are a driving exposure measurement and an indication of prevailing traffic levels [10], indicating the relationship between this exposure measurement and HA and HB events. Driving during risky nighttime hours (00:00 - 05:00) made a small contribution to HA and HB, indicating that there was a change in events during nighttime driving, which is consistent with existing literature [2].

The top three variables were unrelated to COVID-19 characteristics, which is understandable given that the COVID-19 pandemic had no direct effect or causality on driving behavior. Despite this, four COVID-19-related variables were discovered to have an effect on HA and HB events. In terms of HA events, new COVID-19 cases in Greece appeared to be ahead of other COVID-19-related variables. In contrast to HA, COVID-19 Reproduction Rate was revealed to have the greatest influence on HB occurrences. COVID-19 Reproduction Rate, Stringency

Index, and New COVID-19 Fatalities and Cases were the COVID-19-related variables that influenced the HA and HB events in Greece, proving that COVID-19 metrics and restriction measures influenced driving behavior.

Figure 1 shows that driving demands were much lower during both lockdowns as compared to the no-restrictions baseline. The first lockdown had the highest decline compared to the second. This means that the traffic volume during the first lockdown was smaller than in the other conditions, allowing drivers to accelerate more freely, as seen in Figure 3 (left), where the highest percentile was higher than in the other conditions. Furthermore, the median in Figure 3 (right) for trips with harsh accelerations was higher than the other conditions, indicating that HA events were more common even on trips with maximum values. This result is also supported by the literature [3]. The median was higher for trips with harsh accelerations during the second lockdown compared to trips without restrictions due to lower traffic volume, but not by the same magnitude as the first lockdown where traffic volume was lower.

Figure 5 (left) shows that the upper percentile is higher than other conditions for HB events, and Figure 5 (right) shows that the median is higher than other conditions for trips with harsh brakings, indicating that HB events were more common even on trips with maximum values. This can be explained by the fact that the traffic flow during the first lockdown was lower than in the other conditions, and with fewer vehicles ahead, the drivers were able to maintain faster speeds, as mentioned in [3], and were thus more likely to be involved in a harsh braking event. In terms of the second lockdown, which followed the same reasoning as HA, the median was higher for trips with harsh brakings compared to no restrictions due to reduced traffic volume, but not by the same magnitude as the first lockdown, which had lower traffic volume.

5. Conclusions

Focusing on the COVID-19-related variables, this research identified the most influential factors in the year 2020 that influenced the relationship between COVID-19 pandemic metrics (i.e., COVID-19 new cases, new fatalities, and reproduction rate), restrictions (i.e., stringency index) with driving behavior. The findings of the XGBoost exploratory study show a correlation between COVID-19 metrics and restriction measures with harsh brakings and accelerations. Furthermore, different HA and HB event patterns were discovered for all three analyzed conditions, namely no restrictions, 1st lockdown, and 2nd lockdown. Due to their correlation with driving exposure measurements (i.e., Apple driving requests), it can be stated that HB and HA (for trips with the occurrence of harsh events) were increased and more frequent during lockdown restrictions.

Acknowledgments

The authors thank OSeven Telematics, London, UK (oseven.io) for supplying all of the necessary naturalistic driving data for this study.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.*, p. Rhodes, Greece, 2021.

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

[6] 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.

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

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

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

[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.