Road Crashes Analysis in Greek Islands

Maratos Filippidis[1], Dimitrios Nikolaou[1*], Katerina Folla[1], George Yannis[1]

1. Department of Transportation Planning and Engineering, School of Civil Engineering – National Technical University of Athens, 5 Iroon Polytechniou Str., GR-15773 Athens, Greece, * dnikolaou@mail.ntua.gr

Abstract

This paper analyzes road crashes on 38 Greek islands to assess the impact of tourism on road safety. Monthly data on tourist arrivals, crashes, and fatalities were collected and grouped by geographical region. Initial insights were drawn by creating charts for four representative islands, highlighting correlations between arrivals and both crashes and fatalities. To further explore these relationships, Generalized Linear Models and Random Forest models were developed using R. The analysis revealed a positive correlation between tourist arrivals and road crashes, indicating that higher visitor volumes often coincide with increased crash and fatality rates. However, the strength of this relationship varied across island groups, potentially reflecting differences in road infrastructure and traffic management. Notably, some high-traffic islands exhibited lower crash rates, suggesting that better road conditions or more effective traffic control measures may mitigate the impact of tourism on road safety.

Keywords: *Road Crashes, Tourism, Greek Islands, Generalized Linear Model, Random Forest.*

1. Introduction

1.1 General Overview

Greece has a plethora of inhabitable islands, a characteristic that places it among the top positions worldwide in this specific list. The beauty of many of these islands is the reason why a large volume of tourists, both domestic and international, choose to visit them during the summer months. It is even observed that the population of these islands increases rapidly during this period, and their municipalities have to adapt to this reality every year at this time. According to research by the National Technical University of Athens, crashes occurring in the Greek islands, whether they involve minor injuries or fatalities, are fewer than those on the mainland. The narrower roads, lower traffic, and even the smaller population contribute to this fact seeming logical. However, the traffic demands of winter differ from those in summer, depending, of course, on the tourist traffic of each place. It is rare for a habitable Greek island not to experience a sharp increase in its population, and many of them show a tripling of their residents in July and August. Studies show that traffic crashes increase correspondingly (Nikolaou et al., 2019).

1.2 Objective

This paper aims to analyze road crashes on the Greek islands. Specifically, it investigates the contribution of increased traffic due to tourism to both crashes with injuries and fatal crashes. Furthermore, models correlating crashes and arrivals were created based on the collected data. To achieve the set goals, the databases of the Hellenic Statistical Authority were used for recording arrivals

at ports and airports of 38 Greek islands, for which there is no other access, ensuring an accurate picture of visitors. To draw reliable conclusions related to seasonality, the arrivals are referenced by the monthly number of tickets used for access to each island. Additionally, the detailed monthly road crash data was used. Based on the collected data, a descriptive analysis was conducted for each indicator and each island individually. Moreover, emphasis was placed on the development and comparison of models through computational software, correlating road crashes and related casualties with the arrivals of tourists.

2. Literature Review

2.1 Introduction

In this chapter, the literature review conducted as part of this paper is presented. Specifically, studies concerning the impact of tourism on road crashes are provided, along with the methodologies used and the conclusions that were drawn.

2.2 Relevant Studies

In 1996, a study in New Zealand was published examining whether tourist crashes are random events or if they can be predicted and prevented (Page & Meyer, 1996). According to the researchers, traveler injuries are not random events and can be prevented with proper education and awareness. However, crash prevention is not widely disseminated in the tourism industry due to a lack of cooperation between health professionals, travel agents, and other stakeholders. Another significant finding was that road crashes are the leading cause of death for Americans traveling abroad. Health professionals can play a role in educating travelers with informational packages on crash prevention, including active and passive strategies. Various measures, such as displaying informational videos or distributing leaflets, could help reduce road crashes. Nevertheless, the tourism industry in New Zealand does not consider tourist crashes a major issue due to insufficient data.

In 2004, a study was published highlighting the impact of tourists on road crashes in the Central Scotland area, noting differences between accidents involving local and foreign drivers (Walker & Page, 2004). The study includes data analysis of road crashes in Scotland, which were obtained from the "STATS19" database of the Central Scotland Traffic Accident Investigation Unit. The data covers 2,841 crashes involving a total of 4,842 vehicles and 7,384 victims within the jurisdiction of Central Scotland Police. The database includes a field with postal codes, used to determine whether drivers are "locals" or "visitors." Residents within the area are classified as "locals," while those living outside the area are classified as "visitors". The study shows that foreign drivers are involved in 28% of crashes, with a higher proportion of serious or fatal crashes compared to locals. These crashes increase during holiday periods, peaking in the summer months. Additionally, visitors tend to use main roads more frequently due to unfamiliarity with local roads, which increases the likelihood of crashes on main roads or highways. Notably, the study emphasizes the need for targeted road safety measures, particularly for visitors, such as "Keep Left" campaigns in multiple languages, highlighting the difficulty foreign drivers have in adjusting to the country's traffic conditions.

In 2019, a study was conducted regarding the impact of tourism on road crashes in Greece (Bellos et al., 2020). This study collects and analyzes data on all road crashes recorded in Greece between 2011 and 2015, based on the ELSTAT database. It includes only crashes involving injuries or fatalities, excluding those with material damage. A total of 39,720 crashes were examined, with information on their date and location. Negative binomial regression, a statistical method suitable for overdispersed data, was used to analyze the frequency of crashes. Independent variables include the purpose of the trip (touristic or not) and the area (touristic or not), while data from major cities and transition months (May and October) were excluded. The analysis showed that crashes peak during the tourist season (June-September), with

tourist areas, especially islands, experiencing more crashes. Tourists are at higher risk due to unfamiliar traffic conditions. The study concludes with practical interventions to improve road safety, such as stricter alcohol controls and better signage in tourist areas, especially during the summer months.

In 2019, another study focused on the connection between tourism and road crashes (Nikolaou et al., 2019), using data from ELSTAT regarding injuries in traffic crashes, excluding those with only material damage. The analysis was divided into two stages: the first by type of area (touristic vs. non-touristic) and the second by the nationality of the injured or killed (Greek, foreign tourists, and immigrants). The method of induced exposure to risk was used. The study concludes with interesting findings. The number of road crash victims increases in the summer, with a significant peak in August in tourist areas. The victims in these areas are nearly twice as high compared to other months. In tourist areas, younger individuals (aged 15-44) are more frequently involved in crashes, and two-wheelers constitute the majority of the injured (52%). In non-tourist areas, older individuals (45+) are more involved in crashes, and passenger car drivers make up the majority of injured individuals (54%). The study also reveals that foreign tourists are more likely to be involved in crashes, especially in non-tourist areas. In conclusion, the study suggests measures such as stricter road controls, better signage, and targeted road safety campaigns. It also highlights differences in road safety between tourist and non-tourist areas, offering useful information for decision-makers to improve road safety policies in Greece.

In 2023, a study was conducted with the aim of identifying patterns or relationships between various factors related to traffic accident injuries, both in island and mainland areas of Greece (Ziakopoulos et al., 2023). In this study, data from the ELSTAT database were analyzed, including 41,541 road crash injuries, separated into minor injuries and fatal or serious injuries. The apriori algorithm was also used to discover association rules. The results revealed the frequency with which factors such as clear weather, urban road environments, male drivers, and vulnerable road users are associated with road crashes. These associations appear at high frequencies, usually above 70% or 80% of the total injuries, and provide insights into how certain patterns in injuries are expected due to the high exposure to these factors. Interestingly, the island environment did not show significant differences in the association rules compared to mainland Greece, suggesting that any differences require further detailed analysis.

3. Data Collection

As previously mentioned in Section 1, the purpose of this paper is to analyze road crashes on the Greek islands. In the initial stage, the aim is to investigate the impact of tourism on the increase in crashes during the summer months, as evidenced by the literature presented in Section 2. This section describes the method of collecting the necessary data to achieve the objectives of this study, as well as the appropriate processing of this data. An initial explanation of the collected data and its significance is provided, followed by the first round of data processing using Microsoft Excel, during which the data are formatted appropriately for input into the R programming language.

In order to obtain a satisfactory result that clearly demonstrates the relationship between tourism and road crashes, it was necessary to take a large sample of Greek islands. A set of 38 Greek islands was collected, which are geographically distributed into the following groups: (1) Dodecanese (2) Cyclades (3) Ionian (4) Central and Northern Aegean.

In this paper, monthly data for the decade 2009-2018 was collected, with the aim of accounting for the seasonality of the phenomenon in the analyses. The data on arrivals at the ports of the aforementioned islands were obtained from the Hellenic Statistical Authority, while the data on arrivals at the airports (for those islands with an airport) were sourced from the official website of the Institute of Greek Tourism Enterprises. Since road access to these islands is not possible, the data from the ports and airports were combined to produce the total arrivals per month. It is worth noting that yacht arrivals were not considered in the total arrivals per island for this study. Monthly data on road crashes with at least one

injury for the decade 2009-2018 were also obtained from the Hellenic Statistical Authority. This data was divided into (1) total road crashes and (2) the number of fatalities.

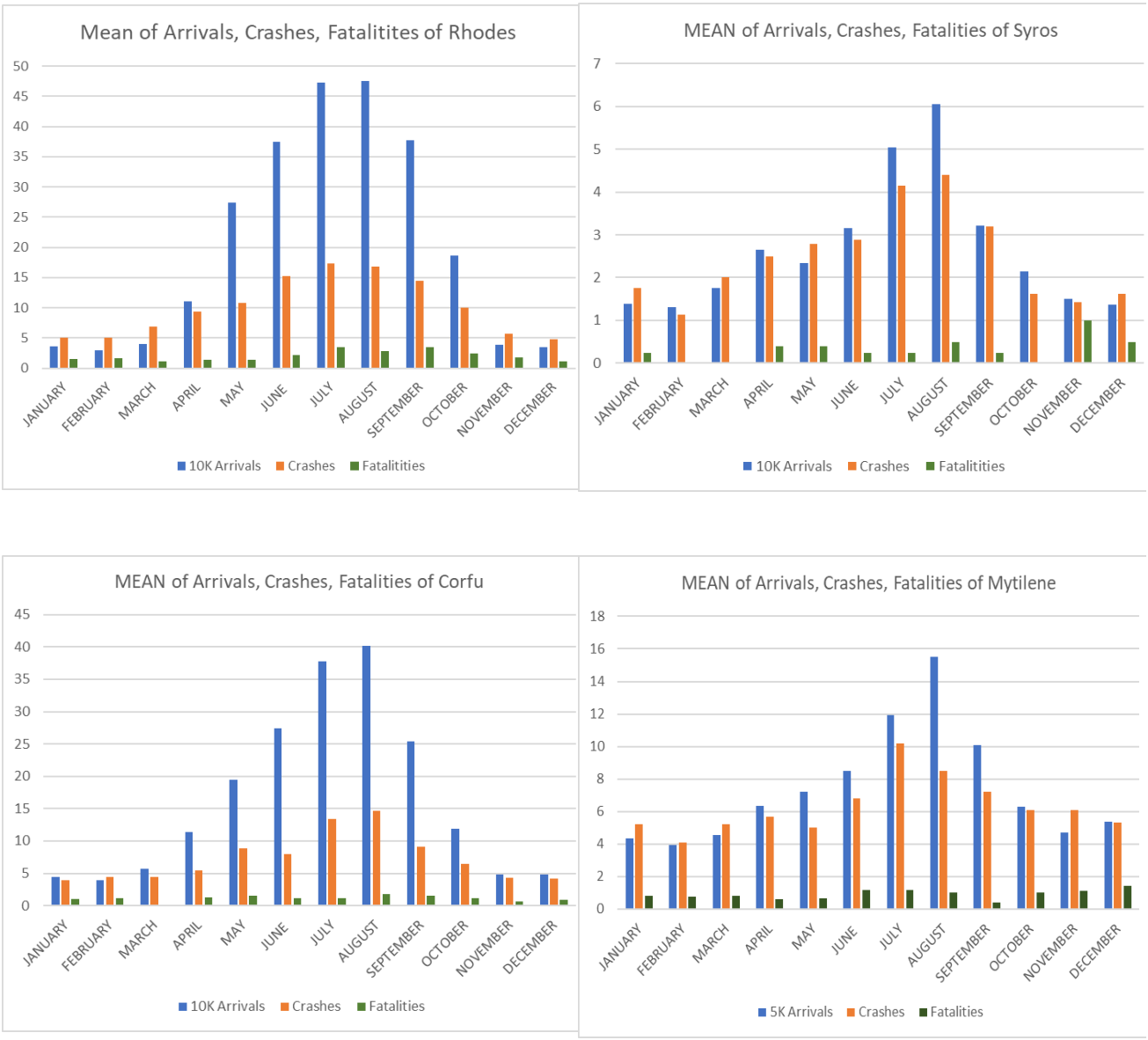

***Figure 1**: Mean of monthly arrivals, crashes, and road fatalities in selected islands (2009-2018)*

The Rhodes graph shows that arrivals on the island increase during the period from May to October, with the highest number of arrivals observed in July and August. In April, there are also more arrivals compared to the winter months, but significantly fewer than during the May–October period. Over the years, tourist arrivals have been increasing. The number of crashes and fatalities seems to follow the trend of tourist arrivals and rises significantly in July and August. The Syros graph shows an increase in arrivals during the period from March to August, followed by a gradual decrease in September and October. The traffic in July and August is clearly much higher than that observed in the other months. The number of crashes, like arrivals, increases during the summer months.

The Corfu graph shows a steady increase in arrivals from April to August and a decrease from September to November, while the winter months remain at stable levels. The trends in arrivals are mirrored by the numbers of monthly crashes and fatalities, showing an increase during the same months. The Mytilene

graphs show an increase in arrivals during the summer months, with the July-September quarter being the period in which the island experiences the highest visitor numbers.

4. Statistical Background

In this section, the statistical background of the methods used in this paper is presented.

Generalized Linear Models (GLMs) extend traditional linear models to accommodate data that do not meet the assumptions of normality and homoscedasticity, making them suitable for a wide range of distributions, including Poisson, binomial, and gamma. They consist of three key components. First, the distribution function allows the dependent variable to follow various distributions from the exponential family, such as normal, binomial, and Poisson, enabling flexibility in modeling different types of data (e.g., binary, count). Second, the linear predictor assumes a linear relationship between the independent variables and a transformed version of the dependent variable, expressed as $\eta=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p$, where η is the linear predictor. Third, the link function connects the expected value of the dependent variable to the linear predictor, allowing for different relationships depending on the data. Common examples include the log link for Poisson models, the logit link for binary outcomes, and the identity link for normally distributed data.

Random Forest is an ensemble machine learning method for classification and regression, known for its accuracy and robustness. It builds multiple decision trees using random subsets of data and features to reduce overfitting and improve generalization. The final prediction is made by aggregating the outputs of these individual trees, using majority voting for classification and averaging for regression. This approach significantly enhances model stability and reliability, especially in high-dimensional data contexts.

Model acceptance criteria are essential for evaluating model performance. McFadden R^2 is a measure of goodness-of-fit for models like logistic regression, calculated as $R^2 = 1 – \ln(Lmodel)/\ln(Lnull)$, where higher values indicate better fit, typically ranging from 0.2 to 0.4 for acceptable models. The AICc (Corrected Akaike Information Criterion) is an adjusted version of AIC that accounts for small sample sizes, aiming to balance model fit and complexity, where a lower AICc indicates a more parsimonious model. Additionally, the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to assess prediction accuracy by measuring the average and squared average differences between observed and predicted values, respectively, providing complementary insights into model performance.

For further information on these methods, the reader is referred to Washington et al. (2020).

5. Results

The following Table summarizes the results of the developed models.

The models for all island groups showed a positive correlation between the number of arrivals and the number of deaths and crashes. Specifically, the highest correlation coefficient for the models of deaths is observed in the Cyclades, with the Dodecanese coming second and the Ionian Islands third, while the low McFadden R^2 of the GLM model for the Central and Northern Aegean group leads to the conclusion that it is not statistically significant, thus unacceptable. Regarding the models with road crashes as the dependent variable, the highest correlation coefficient is found for the model of the Central and Northern Aegean Island group, followed by the Cyclades and the Dodecanese in descending order, with the Ionian Islands coming last. Using GLMs to correlate tourist arrivals with fatalities in road crashes, the smallest prediction errors are found for the Cyclades, followed by the Ionian Islands, with the Dodecanese at the

end. In the corresponding models of crashes, the smallest prediction errors are observed for the Cyclades, with the Central and Northern Aegean, Ionian Islands, and Dodecanese following.

***Table 1**: Results of the developed statistical and machine learning models*

	Generalized Linear Model (GLM)					
	Fatalities			Crashes		
Island Groups	**Συντελεστής**	**McFadden R^2**	**AICc**	**Συντελεστής**	**McFadden R^2**	**AICc**
Dodecanese	0.00909 < p = 0.001	0.16	954.66	0.0123	0.1193	2638.03
Kyklades	0.01122 < p = 0.001	0.11	628.61	0.0182	0.1081	2491.14
Ionio	0.00714 < p = 0.001	0.10	634.42	0.0095	0.1062	1853.99
Central/ North Aegean	0.03354 < p = 0.001	0.06	708.54	0.0574	0.100	2227.65

Island Groups	**Fatalities MAE**	**Fatalities RMSE**	**Crashes MAE**	**Crashes RMSE**
Dodecanese	0.64	2.42	15.40	87.42
Kyklades	0.16	0.34	3.40	15.83
Ionio	0.59	1.07	5.78	18.60
Central/ North Aegean	0.27	0.51	4.35	17.52

	Random Forest			
	Fatalities		Crashes	
Island Groups	**Mean of Squared Residual**	**% of Var explained**	**Mean of Squared Residual**	**% of Var explained**
Dodecanese	0.454	23.16	5.977	63.15
Kyklades	0.099	-26.53	2.185	16.53
Ionio	0.471	2.84	8.015	53.14
Central/ North Aegean	0.245	-21.47	4.658	15.18

Island Groups	**Fatalities MAE**	**Fatalities RMSE**	**Crashes MAE**	**Crashes RMSE**
Dodecanese	0.29	0.65	1.06	2.10
Kyklades	0.16	0.40	0.96	1.91
Ionio	0.40	0.72	1.92	3.42
Central/ North Aegean	0.24	0.50	1.73	3.26

The Random Forest method was also used to correlate tourist arrivals with road crashes and deaths across the same island groups. Regarding the models for the number of deaths in road crashes, only one satisfactory model was obtained for the Dodecanese group, with an acceptable prediction error. For the models correlating arrivals with crashes, acceptable models were obtained for all island groups. It is noted, however, that for the Cyclades and the Central and Northern Aegean, the explained variance is relatively low. The smallest prediction errors were calculated for the Cyclades and Dodecanese, with the Central and Northern Aegean and Ionian Islands following.

In general, comparing the two methods in terms of prediction errors, lower MAE and RMSE indicators are observed for the crash-arrival models using the Random Forest method.

6. *Conclusions*

The analysis has highlighted several important insights into the relationship between tourist arrivals and road crashes on Greek islands. Tourism, and more broadly the volume of arrivals, appears positively correlated with both the number of crashes and fatalities during peak periods. This link is particularly evident in regions like the Central and Northern Aegean, where the influence of tourist arrivals on crash frequency is more pronounced compared to areas like the Ionian Islands, the Dodecanese, and the Cyclades. This disparity likely reflects differences in road infrastructure quality, driver behavior, and the volume of foreign tourists, who may struggle more with adapting to unfamiliar traffic conditions. In

regions with fewer foreign tourists, the absence of targeted measures for their adaptation may also play a role.

However, the relationship between arrivals and fatalities is less consistent. In some cases, no significant correlation was observed, likely due to the low absolute number of fatalities, which limits the statistical power of the models. This suggests that factors beyond just the volume of traffic, such as dangerous driving behaviors, road design, and the availability of emergency medical care, may have a stronger influence on the severity of crashes. Moreover, the fact that crashes tend to increase with arrivals while fatalities do not imply that many incidents involve drivers unfamiliar with local conditions, but not necessarily engaging in extremely risky behavior.

To deepen the understanding of these dynamics, future research could benefit from including additional variables beyond tourist arrivals, potentially exploring the correlation between fatalities and the number of severe injuries on a monthly basis. This would help clarify whether more serious crashes are driving the observed trends. Expanding the analysis to include mainland regions could also provide valuable context by allowing comparisons between island and mainland traffic patterns.

Authorities can play a critical role in reducing crashes and fatalities by implementing targeted road safety measures in tourist areas. This includes ensuring the presence of clear and prominent signage, improving road surface quality, and enhancing street lighting in high-risk zones. Additionally, increased traffic enforcement could help promote safer driving behaviors among both tourists and locals, ultimately contributing to safer travel experiences in these destinations.

7. References-Bibliography

1. Bellos, V., Ziakopoulos, A., & Yannis, G. (2020). Investigation of the effect of tourism on road crashes. Journal of Transportation Safety & Security, 12(6), 782-799.
2. Nikolaou, D., Folla, K., Bellos, E., & Yannis, G. (2019). Tourism and Road Accidents in Greece. In 9th International Congress on Transportation Research (pp. 1-12).
3. Page, S. J., & Meyer, D. (1996). Tourist accidents: An exploratory analysis. Annals of Tourism Research, 23(3), 666-690.
4. Walker, L., & Page, S. J. (2004). The contribution of tourists and visitors to road traffic accidents: a preliminary analysis of trends and issues for Central Scotland.
5. Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.
6. Ziakopoulos, A., Michelaraki, E., Nikolaou, D., Folla, K., & Yannis, G. (2023). Association rule mining for island and mainland road crash injuries in Greece. Transportation research procedia, 72, 163-170.