Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2022) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Transport Research Arena (TRA) Conference

Association Rule Mining for Island and Mainland Road Crash Injuries in Greece

Apostolos Ziakopoulos[a]*, Eva Michelaraki[a], Dimitrios Nikolaou[a], Katerina Folla[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

The objective of this paper is to implement association rule mining in order to identify the underlying patterns in road crash injuries recorded in Greece, after separating them into subsets concerning the mainland and the island areas. To that end, an analysis of disaggregated data from the Hellenic Statistical Authority data file for killed, seriously and slightly injured road users was implemented for 2017-2019. The apriori rule was developed and a variety of association rules were obtained. Results indicated a strong association or concurrent presence of the contributory factors at various percentages in both mainland and island road crash injuries. It was also revealed that clear weather, urban area and male road users were dominant interconnected conditions that simultaneously exist in injury crashes.

Peer-review under responsibility of the scientific committee of the Transport Research Arena (TRA) Conference

Keywords: Association rule mining; Apriori algorithm; Road crash injuries; Road safety; Island crashes.

1. Introduction

Despite considerable efforts and rapid technological advancements to date, road safety remains a major issue worldwide. The pursuit for safer road environments has repercussions throughout all aspects of transport. It is estimated that every year the lives of 1.3 million people are cut short as a result of road and traffic crashes (World Health Organization, 2018). Approximately 20 up to 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury. According to the European Commission, around 19,800 road traffic fatalities were reported by the 27 EU Member States in 2021 (European Commission, 2022). This corresponds to an

* Corresponding author. Tel.: +30-210-772-1575;
E-mail address: apziak@central.ntua.gr

Peer-review under responsibility of the scientific committee of the Transport Research Arena (TRA) Conference

increase of 5% in road fatalities compared to 2020, but still represents a decrease of 13% compared to the pre-pandemic period in 2019. Over the period 2010-2020, road fatalities have significantly dropped by 36%, but the target to halve this number by 50% has not been met. In Greece, the most significant road safety improvement in the European Union was recorded, with a decrease of 54% in road fatalities since 2010 (NTUA Road Safety Observatory, 2021). During the same period, the number of serious injuries resulting from road traffic collisions were reduced by 72% and the rate fatalities per vehicles decreased by 56%. Despite these successes, the road safety performance of Greece still remains at a lower level compared to the European average. Thus, more efforts should be made to enhance road safety and reduce fatalities from all modes of transport closer to the goal of Vision Zero.

Based on the aforementioned, Greece was selected as a country-wide case study as a popular destination with an increasing number of annual visitors, especially during the summer period. With a total area of 131,957 km^2, the terrain of the mainland of Greece is mostly (80%) mountainous while its seas are sprinkled with more than 3,000 islands, its climate and weather conditions are generally favorable (Yannis & Papadimitriou, 2012). The Greek islands are one of the main characteristics of the country's morphology, consisting an integral part of its tradition, culture and history. Due to that history, Greek islands have unique, isolated road and often space-starved road environments. These specific attributes of Greece offer unique topics for investigations of the impacts of different geographical regions on road safety.

The aim of the current study is to employ the association rule approach to identify underlying patterns of road crash injuries occurring in Greece, with different examinations for the mainland and island environments. Data on road crashes in Greece for the period 2017-2019 were collected and subsequently underwent processing and analysis. Greek regions were classified as mainland or island areas, while two groups of injured road users were considered according to their severity (i) slight injuries and (ii) fatalities & serious injuries. The data-mining technique of association rules was used to obtain insights into the underlying relationships found in crash injury data. The analysis quantified several interesting findings concerning the relationship among geographical division, severity and crash injuries in Greece.

The paper is structured as follows. Initially, the crucial issue of road crashes in Greece and worldwide as well as the aim of this research is presented. Subsequently, scientific literature with regards to road safety in the Greek island environment is reviewed and an apriori/association rule overview in road safety is also provided. A description of the utilized data then follows. Association rules and the respective discussion between island and mainland areas road crash injuries in Greece is provided. Lastly, the main conclusions and practical considerations are highlighted.

2. Literature review

Few studies have been previously conducted to examine road safety in the Greek island environment or the relationship between foreign and native drivers in terms of road safety in Greece. Yannis et al. (2007) focused on road safety issues of local and non-local drivers in various road environments in Greece. Results confirmed that foreign drivers were found to be at increased risk, while immigrant permanent residents had a lower risk compared to tourists, regardless of the road layout (i.e. urban, rural, highway). Another approach, aiming to investigate the impact of tourism on road crashes in Greece showed that the tourists were more often involved in road crashes (Bellos et al., 2020). In particular, it was revealed that the touristic season as well as tourism as the purpose of traveling led to a spike in road traffic crashes, with the highest increase being identified in touristic regions. A relevant study examined the risk factors for motor-vehicle crashes in the Heraklion District of the island of Crete (Petridou et al., 1997). It was estimated that the corresponding ratio of road crashes for tourists was close to 1:3, underlying the importance of road crashes as the significant health hazard during pleasure traveling.

Moreover, a similar study (Petridou et al., 1999) was also conducted and data from a different Greek island, Corfu, were used in an attempt to indirectly assess the pattern of leisure trips-related injuries among foreign tourists. Interestingly, only 15% of all road crashes concerned residents and Greek tourists, whereas they accounted for 40% of foreign nationalities. An increased likelihood of road traffic crash injuries to young foreign tourists (i.e. especially younger than 24 years old) was also reported. Finally, another study examined the effect of tourism on road crashes in Greece (Nikolaou et al., 2019). It was observed that related casualties and road crashes increased during touristic periods and the highest percentages of road crash casualties in touristic regions consisted of road traffic crashes of young people and powered-two-wheelers (PTWs).

The goal of association rule mining is to examine a group of items that typically occur together in a given event and provide insights for the frequency of associations between items. To date, the association rule has been widely

used in the analysis of road crashes and particular focus was mainly concentrated on the relationship between characteristics and contributory factors and crash risk (Samerei et al., 2021). Xu et al. (2018) explored the contributing factors to serious casualty traffic crashes and their interdependency by conducting an association rule analysis. The results revealed that serious casualty crashes were a result of complex interactions between driver behavior, vehicle factors, road geometric characteristics and weather conditions. Association rule techniques are appropriate to deal with data containing a large number of variables and it can interpret their correlation if a proper confidence and support are given, providing specific and easy venues to describe relationships between crash attributes (Montella et al., 2012). Overall, previous studies have mostly focused on identifying the crash contributory factors, but few studies aimed to explore the potential interdependencies and interactions among crash characteristics in Greece. The road safety problem in mainland and island destinations, affecting both visitors and residents has not been properly addressed. This fact forms the motivation for the current paper, aiming at using the association rules technique in order to investigate those circumstances.

3. Data collection

The required road crash data were collected from the Greek road crash database SANTRA which is provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens (NTUA). The Hellenic Statistical Authority maintains and updates the official road crash database in Greece. This database includes road crashes in which at least one involved road user was injured or killed. In Greece, Traffic Police officers attend the crash site and complete the crash data in high detail in standardized templates (the Crash Data Collection Forms) immediately after the occurrence of a crash, providing information on crash conditions and on characteristics related to the road, the involved persons or vehicles. The Crash Data Collection Forms are then forwarded to the Hellenic Statistical Authority, which is responsible for the final checking and codification into the official National Road Crash Database. Copies of the National Road Crash Database are provided to NTUA (without personal identification), which developed the SANTRA system featuring efficient queries to extract any combination of data.

This database consists of disaggregated data for all road injury crashes in Greece for the period 1985-2019, updated annually. Finally, it should be mentioned that fatalities are defined as all persons killed within 30 days after the crash, seriously injured as all persons hospitalized for at least 24 hours and slightly injured as all road users reported by the police as injured without hospitalization. The variables that were utilized for the objectives of this study are presented in Table 1 and correspond to the 3-year period with the latest available data (2017-2019).

Table 1. List of variables obtained from the road crash database

Variable	Explanation	Values
MONTH	Month of the year	1-12
WEEK	Week of the year	1-53
DAY	Day of the month	1-31
HOUR	Hour of the day	00-23
VEH_NO	Number of vehicles invol	0-99
SEVERITY	Injury severity	1: Killed 2: Seriously Injured, 3: Slightly Injured
POSITION	Position in vehicle	2: Front passenger, 3: 2nd Rear passenger, 4: 3rd Rear passenger, 5: Other, 9: Unknown
WEATHER	Weather conditions	1: Clear, 2: Severe wind, 3: Frost, 4: Fog, 5: Light rain, 6: Rain, 7: Storm, 8: Thunderstorm, 9: Hail, 10: Snow, 11: Smoke, 12: Dust, 13: Other
LIGHTING	Night lighting	1: Adequate artificial lighting, 2: Insufficient artificial lighting, 3: Artificial lighting off, 4: No lighting installation
TRANS_ GROUPED	Transport mode	1: Passenger Car, 2: Powered-Two-Wheeler (PTW), 3: Bicycle, 4: Bus, 5: Heavy Goods Vehicle (HGV), 6: Lorry <3.5t, 7: Other, 8: Pedestrians
AREATYPE	Area type	1: Urban, 2: Non-urban
EQUIPMENT	Use of protective equipm	1: Seat-belt use, 2: Helmet use, 3: Child restraint system use, 4: No seat-belt use, 5: No helmet use, 6: No child restraint system use, 9: Unknown
GENDER	Gender	1: Male, 2: Female, 9: Unknown
AGE_GROUP	Age (5 categories)	1: <18, 2: 18-24, 3: 25-49, 4: 50-64, 5: >65, 6: Unknown

4. Descriptive Statistics

Before proceeding to more detailed analysis, it would be fruitful to present some main descriptive statistics of the variables that were used in the present study. The total number of injured road users for the 3-year period by severity, transport mode and area type is presented in Table 2.

Table 2: Injured road users by severity, transport mode and area type (sum 2017-2019)

Area type	Transport mode	Severity		
		Fatalities	Serious injuries	Slight injuries
Mainland	Passenger Cars	602 (36%)	457 (30%)	10,682 (32%)
	PTW	517 (31%)	719 (47%)	15,612 (47%)
	Bicycles	33 (2%)	32 (2%)	542 (2%)
	Buses	0 (0%)	2 (0%)	386 (1%)
	HGV	16 (1%)	9 (1%)	134 (0%)
	Lorry <3,5 tonnes	101 (6%)	58 (4%)	866 (3%)
	Other	41 (2%)	13 (1%)	197 (1%)
	Pedestrians	355 (21%)	246 (16%)	5074 (15%)
	Total	**1,665 (100%)**	**1,536 (100%)**	**33,493 (100%)**
Island areas	Passenger Cars	152 (33%)	117 (21%)	1,108 (29%)
	PTW	195 (43%)	348 (63%)	2,056 (53%)
	Bicycles	12 (3%)	8 (1%)	74 (2%)
	Buses	0 (0%)	1 (0%)	15 (0%)
	HGV	5 (1%)	1 (0%)	17 (0%)
	Lorry <3,5 tonnes	22 (5%)	8 (1%)	118 (3%)
	Other	14 (3%)	5 (1%)	47 (1%)
	Pedestrians	54 (12%)	61 (11%)	409 (11%)
	Total	**454 (100%)**	**549 (100%)**	**3,844 (100%)**

Based on the figures of Table 2, it can be observed that most killed road users in all areas of Greece correspond to passenger cars and PTWs. However, in the islands, the percentages of killed PTW riders is significantly higher compared to the mainland. In general, it becomes readily apparent that there is a disproportionally high ratio of killed or seriously injured (KSI) road users to slightly injured road users in the islands compared to the mainland, at least at the descriptive statistics level. This hints that the unique, isolated road environment found in Greek islands (Zwier et al., 1995), i.e. narrow roads, sharp turns, reduced visibility, high and sharp cliffs, often populated with foreign/tourist drivers has observable consequences in higher KSI ratios, denoting a less forgiving road transport system.

5. Methodology

The apriori algorithm is the most popular algorithm for mining association rules. Association rules typically comprise of an antecedent part (if X occurs…) and a consequent part (…then Y occurs as well). Apriori reveals the most frequent combinations in a database and identifies association rules between the items. Advantages of apriori include the interpretability of its outcomes, and the ability to easily process larger datasets (Lantz, 2019). Further details regarding the apriori functions are provided in earlier research (Agrawal & Srikant, 1994). In summary, the rules obtained by apriori are evaluated based on three parameters:

1. Support describes the probability that X and Y occur together. Mathematically, X and Y can be considered subsets of the examined set, noted as {X} and {Y}. Support is then calculated as per **Equation 1:**

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Occurrences\ containing\ both\ X\ and\ Y}{Total\ number\ of\ casualties\ (injuries/fatalities)} \quad (1)$$

2. Confidence describes the conditional probability of {Y} occurs by knowing {X} occurs. In other words, how often does {Y} happen when {X} has happened first. Confidence is calculated as per **Equation 2**:

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Support(X,Y)}{Support(X)} \quad (2)$$

3. Lift describes the ratio between support and confidence. For instance, a lift of 2 means that the likelihood of encountering {X} and {Y} together is 2 times the likelihood of just encountering {Y}. Lift is calculated as per **Equation 3**:

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{Occurrences\ containing\ both\ X\ and\ Y\ /\ Occurrences\ containing\ X}{Occurrences\ containing\ Y} = \frac{Confidence(\{X\} \rightarrow \{Y\})}{Support(Y)} \quad (3)$$

It should be noted here that, depending on the dataset, apriori can produce potentially thousands of rules. Typically, researchers set cutoff values for support and confidence so that they can examine the most prevalent rules.

6. Results and Evaluation

Following the application of apriori in the SANTRA database, several informative sets of rules emerged regarding road users injured in road crashes in Greece. Specifically, four subsets of injured road users were examined, corresponding to the geographical division of Greece between island and mainland areas. The analysis was conducted using R-studio and a number of packages, most notably the arules package for association rule mining (Hahsler et al., 2021).

6.1. Slight injuries in the mainland

In total, records for 33,493 mainland road users with slight injuries were analyzed. It is apparent that the most frequent characteristics all have relative frequencies above 0.3. In order to discover more dominant trends in injured road user characteristics, at least a minimum support and confidence value was necessary, to filter out rules that either concern less frequent characteristics or are less informative. After several trials during the modelling process, the selected parameter values were support = 0.4 and confidence = 0.5. The resulting association rules with the top 15 scoring lift values appear on Table 3. In addition to the three apriori metrics, count denotes the absolute instances where this rule is encountered in the dataset.

Table 3. Top 15 association rules for slightly injured road user characteristics in mainland crashes

Rules	Antecedent	Consequent	Support	Confidence	Lift	Count
[1]	{WEATHER=1, AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.439	0.809	1.187	14719
[2]	{AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.468	0.803	1.178	15669
[3]	{WEATHER=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.488	0.759	1.114	16361
[4]	{AREATYPE=1, GENDER=1}	{VEH_NO=2}	0.411	0.706	1.104	13774
[5]	{WEATHER=1, AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.578	0.752	1.103	19350
[6]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.578	0.909	1.096	19350
[7]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{GENDER=1}	0.488	0.768	1.096	16361
[8]	{TRANS_GROUPED=[2,8]}	{GENDER=1}	0.522	0.767	1.094	17489
[9]	{GENDER=1}	{TRANS_GROUPED=[2,8]}	0.522	0.745	1.094	17489
[10]	{TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.616	0.905	1.093	20646
[11]	{AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.616	0.744	1.093	20646
[12]	{WEATHER=1, TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.439	0.900	1.086	14719
[13]	{WEATHER=1, TRANS_GROUPED=[2,8], AREATYPE=1}	{GENDER=1}	0.439	0.761	1.086	14719
[14]	{TRANS_GROUPED=[2,8], AREATYPE=1}	{GENDER=1}	0.468	0.759	1.083	15669
[15]	{TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.468	0.896	1.082	15669

From the examination of the resulting association rules, several informative outcomes can be acquired for crashes with slight injuries in the mainland of Greece. To avoid overloading the reader, the findings of the first five rules as scored by lift are briefly explained indicatively. The first rule reveals that, if a slightly injured male road user is involved in a crash in urban areas under clear weather conditions, then a vulnerable road user (VRU – i.e. pedestrian or PTW occupant) is also involved in more than 80% of the crash injuries. The second, third and fifth rules further support these findings, providing quantifications of the effects of the aforementioned factors in pairs. The second rule, for instance, indicates that if a slightly injured male road user is involved in a crash in urban areas then a VRU is also involved in more than 75% of the crash injuries. The fourth rule reveals that, if a slightly injured male road user is involved in an urban crash, it is a two-vehicle crash 75% of the time. It is evident that the apriori algorithm so far provides noteworthy information quantifying parameters of crash injuries in Greece that are closely linked with exposure, such as the high percentage of male PTW drivers and their frequent involvement in urban crashes (Yannis & Laiou, 2014).

6.2. Fatalities and Serious Injuries in the mainland

In a process mirroring the one outlined in the previous section, 3,201 mainland road users that were killed or seriously injured (KSI) were analyzed in total. After the respective trials, the selected parameter values were support = 0.3 and confidence = 0.5. The slight decrease of support is reasonable, as KSI road users are only a fraction of slightly injured road users. Therefore, lower strictness is required for the apriori algorithm to detect meaningful trends in road user data. The resulting association rules with the top 15 scoring lift values appear on Table 4.

Table 4. Top 15 association rules for KSI road user characteristics in mainland crashes

Rules	Antecedent	Consequent	Support	Confidence	Lift	Count
[1]	{WEATHER=1, AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.352	0.830	1.240	1126
[2]	{AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.378	0.820	1.226	1209
[3]	{WEATHER=1, AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.427	0.816	1.219	1368
[4]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.427	0.694	1.215	1368
[5]	{AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.461	0.806	1.204	1475
[6]	{TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.461	0.689	1.204	1475
[7]	{WEATHER=1, TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.352	0.677	1.184	1126
[8]	{TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.378	0.669	1.170	1209
[9]	{WEATHER=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.520	0.723	1.079	1664
[10]	{VEH_NO=1, WEATHER=1}	{TRANS_GROUPED=[2,8]}	0.354	0.721	1.077	1133
[11]	{VEH_NO=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.311	0.704	1.051	996
[12]	{TRANS_GROUPED=[2,8], AREATYPE=1, GENDER=1}	{WEATHER=1}	0.352	0.931	1.048	1126
[13]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{GENDER=1}	0.520	0.845	1.046	1664
[14]	{TRANS_GROUPED=[2,8]}	{GENDER=1}	0.565	0.844	1.045	1807
[15]	{GENDER=1}	{TRANS_GROUPED=[2,8]}	0.565	0.699	1.045	1807

The examination of the resulting association rules provides information for KSI road users in the mainland of Greece. Specifically, the first rule indicates that KSI male road users in urban environments and clear weather conditions are involved in crashes where VRUs are involved as well, for 80% persons of the crashes of male in urban areas. This finding hints at the increased hazard under which VRUs operate in road networks. In the second, third and fifth rules, fewer of the aforementioned antecedent components are completed but still yield confidence higher than 0.80. As per the fourth rule, a reversal between antecedent and consequent is observed: Mainland crash injuries with KSI male road users under clear weather conditions occur in urban environments approximately 70% of the time. Therefore, this reversal reveals that typically, the factors of clear weather, urban area and male road users are dominant interconnected conditions for injury crashes to occur. Overall, in the top 5 rules, an increase in confidence is observed for KSI road users compared to slightly injured ones.

6.3. Slight injuries in island areas

With respect to the island areas, 3,844 island road users with slight injuries were analyzed in total. After several trials during the modelling process, the selected parameter values were equal to the mainland analysis, i.e. support = 0.4 and confidence = 0.5. The resulting association rules with the top 15 scoring lift values appear on Table 5.

Table 5. Top 15 association rules for slightly injured road user characteristics in island area crashes

Rules	Antecedent	Consequent	Support	Confidence	Lift	Count
[1]	{WEATHER=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.509	0.771	1.083	1958
[2]	{WEATHER=1, AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.489	0.770	1.081	1878
[3]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.489	0.732	1.080	1878
[4]	{AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.517	0.764	1.073	1989
[5]	{TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.517	0.727	1.073	1989
[6]	{GENDER=1}	{TRANS_GROUPED=[2,8]}	0.544	0.763	1.071	2092
[7]	{TRANS_GROUPED=[2,8]}	{GENDER=1}	0.544	0.765	1.071	2092
[8]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{GENDER=1}	0.509	0.763	1.068	1958
[9]	{AREATYPE=1}	{VEH_NO=2}	0.410	0.606	1.062	1577
[10]	{VEH_NO=2}	{AREATYPE=1}	0.410	0.719	1.062	1577
[11]	{VEH_NO=2}	{GENDER=1}	0.426	0.746	1.045	1636
[12]	{GENDER=1}	{VEH_NO=2}	0.426	0.596	1.045	1636
[13]	{WEATHER=1, GENDER=1}	{AREATYPE=1}	0.457	0.692	1.022	1758
[14]	{TRANS_GROUPED=[2,8], AREATYPE=1}	{WEATHER=1}	0.489	0.944	1.017	1878
[15]	{TRANS_GROUPED=[2,8]}	{WEATHER=1}	0.668	0.938	1.011	2567

For the slightly injured road users in the Greek islands, the association rules remained similar with the previous patterns. The first rule indicates that slightly injured male road users under clear weather conditions are involved in crashes where VRUs are involved as well, in 77% of island crash injuries. The second and fourth rules are similar manifestations of the exposure pairings observed in the mainland slight injury crashes. The third and fifth rules are reversals confirming the dominance of clear weather, urban environment and male road users involved in crashes with slight injuries. The confidence and lift of the top five association rules are lower than those discovered from the database of mainland crashes. However, confidence still exceeds 70% for the five top rules by lift.

6.4. Fatalities and Serious injuries in island areas

The final analysis comprised 1,003 island road users that were killed or seriously injured (KSI). The resulting association rules with the top 15 scoring lift values appear on Table 6.

Table 6. Top 15 association rules for KSI road user characteristics in island area crashes

Rules	Antecedent	Consequent	Support	Confidence	Lift	Count
[1]	{AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.373	0.818	1.118	374
[2]	{WEATHER=1, AREATYPE=1, GENDER=1}	{TRANS_GROUPED=[2,8]}	0.351	0.817	1.116	352
[3]	{WEATHER=1, TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.441	0.644	1.112	442
[4]	{WEATHER=1, AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.441	0.805	1.100	442
[5]	{AREATYPE=1}	{TRANS_GROUPED=[2,8]}	0.466	0.804	1.098	467
[6]	{TRANS_GROUPED=[2,8]}	{AREATYPE=1}	0.466	0.636	1.098	467
[7]	{AREATYPE=1, AGE_GROUP=[3,5]}	{TRANS_GROUPED=[2,8]}	0.328	0.799	1.091	329
[8]	{WEATHER=1, AREATYPE=2}	{AGE_GROUP=[3,5]}	0.304	0.816	1.086	305
[9]	{WEATHER=1, AREATYPE=1, AGE_GROUP=[3,5]}	{TRANS_GROUPED=[2,8]}	0.305	0.795	1.086	306
[10]	{WEATHER=1, TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.351	0.629	1.085	352
[11]	{TRANS_GROUPED=[2,8], GENDER=1}	{AREATYPE=1}	0.373	0.625	1.079	374
[12]	{AREATYPE=2}	{AGE_GROUP=[3,5]}	0.340	0.808	1.076	341
[13]	{VEH_NO=2, WEATHER=1}	{TRANS_GROUPED=[2,8]}	0.304	0.780	1.065	305
[14]	{VEH_NO=1, TRANS_GROUPED=[2,8]}	{AGE_GROUP=[3,5]}	0.306	0.795	1.059	307
[15]	{VEH_NO=2}	{TRANS_GROUPED=[2,8]}	0.320	0.772	1.054	321

Lastly, the analysis of KSI road users in the Greek islands yields similar patterns in the association rules remained similar with the previous patterns. KSI crash injuries with male road users in urban areas involve VRUs more than 80% on the time, and an equal finding occurs when clear weather is an additional antecedent. This trend is reversible with respect to urban areas and VRUs with slight loss of confidence, as described by the third rule. The fourth and fifth rules are variations of previously described patterns, with confidence exceeding 80%. As a note, certain age groups (25-49 and >65) were represented in some of the middle-ranking rules (2 times as antecedent and 3 times as precedent). This indicates an overrepresentation of these road user ages in KSI island crash injuries. Furthermore, in none of the 15 rules was the representation of this age group concurrent with a representation of the male gender.

7. Discussion

Apart from individual examination of the association rules, a comparative investigation of the four subsets (slightly injured and KSI road users for mainland and island areas, respectively) reveals interesting findings. On several cases of analyses, the factors of clear weather, urban area and male road users are dominant interconnected conditions that simultaneously exist in injury crashes.

These results can be interpreted in terms of both the presence of specific elements in the obtained rules, and, equally, in terms of absence of other elements. Specifically, the first implication of this result is that a vast majority of road injuries from crashes occur under the most usual conditions of transport activities. Another implication concerns the vulnerability of pedestrians and motorcycles. This study presented sixty rules in total for injured road users, however none of them included a road user type that is not vulnerable. In other words, this finding translates into a quantification of the disproportionately high susceptibility of VRUs to injury from road crashes. The predominance of male road users represented in road crash injuries, and the absence of females in any of the produced rules confirms the more aggressive driving behavior of men, although the respective exposure of male drivers/riders is also higher in Greece compared to their female counterparts (Tzortzi et al., 2021). The Greek island environment did not produce drastically different association rules compared to the mainland. This means that any differences between the two environments require further microscopic or macroscopic granularity to be captured in a fuller extent than this algorithm provides.

These findings are admittedly not novel per se in terms of road safety outcomes. However, the contribution of the present work lies towards (i) the examination of the unique road environment of the Greek islands, (ii) its comparison with the mainland data and (iii) the quantification of the strength and frequency of the island and mainland association rules as described by their support, confidence and lift. Indicatively, for the top five of apriori rules when sorted by lift, the aforementioned factors manifest with a frequency of about or exceeding 70% of total crash injuries as expressed by the respective confidence.

A type of limitation of the present work is the exploited database. The parameters used in this study were derived from the SANTRA Database, which lacks data for some factors (drink-and-drive, mobile phone use, fatigue), which could lead to additional association rules. New technological developments can readily be used for automated and

more standardized data collection. Similar research could be conducted using in-depth analysis data from vehicle details, such as mechanical faults. Such data could be combined with variables describing geometric and road network elements, which can become readily available once the need to record them is translated to action. For instance, incorporating other vehicle related information, such as movement due to collision avoidance, vehicle speed at the collision moment or vehicle component failure, could also be helpful and may allow more detailed analysis of crash characteristics. Lastly, human factor data at the moment of crash are considerably harder to obtain for the entirety of a crash database. Nonetheless, apriori rules with specific human factor behaviors, such as glance direction and duration, could be developed as a more comprehensive index of concurrent circumstances for injury crashes.

8. Conclusions

The present study endeavored to utilize association rule mining to discover underlying patterns in road users injured in crashes in Greece, separating them into subsets concerning the (i) mainland and the (ii) island regions. Data from the SANTRA database was analyzed, comprising 41,541 injury crashes, which were separated into slight crash injuries and killed and serious crash injuries.

The apriori rule was employed and various association rules were obtained. Results provide a quantification of the frequency that clear weather, urban road environment, male road user and VRU presence in crash injuries. The apriori algorithm indicates association or concurrent presence of these factors at various percentages in both mainland and island crashes. Such associations manifest in high frequencies, typically more than 70% or 80% of the total injuries, and provide insights on how certain patterns are expected in road crash injuries due to high exposure. Further research with granular data can create more comprehensive rules for concurrent circumstances for injury crashes in Greece.

Acknowledgements

The present research was carried out within the research project “PHOEBE – Predictive ApproacHes fOr SafEr UrBan Environments”, which has received funding from the European Union’s Horizon Europe research and innovation program under grant agreement with ID 101076963.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, (Vol. 1215, pp. 487-499).

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.

European Commission. (2022, May 11). 2021 road safety statistics: what is behind the figures? European Commission. https://transport.ec.europa.eu/2021-road-safety-statistics-what-behind-figures_en

Hahsler, M., Buchta C., Gruen B., Hornik K., Johnson I. & Borgelt C. (2021). Package ‘arules’: Mining Association Rules and Frequent Itemsets.

Lantz, B. (2019). Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd.

Montella, A., Aria, M., D’Ambrosio, A., & Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Analysis & Prevention, 49, 58-72.

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).

NTUA Road Safety Observatory. (2021, July 28). Road fatalities decrease in the pandemic, Greece 2020. https://www.nrso.ntua.gr/road-fatalities-decrease-due-to-the-pandemic-greece-2020/.

Petridou, E., Askitopoulou, H., Vourvahakis, D., Skalkidis, Y., & Trichopoulos, D. (1997). Epidemiology of road traffic accidents during pleasure travelling: the evidence from the island of Crete. Accident Analysis & Prevention, 29(5), 687-693.

Petridou, E., Dessypris, N., Skalkidou, A., & Trichopoulos, D. (1999). Are traffic injuries disproportionally more common among tourists in Greece? Struggling with incomplete data. Accident Analysis & Prevention, 31(6), 611-615.

Samerei, S. A., Aghabayk, K., Mohammadi, A., & Shiwakoti, N. (2021). Data mining approach to model bus crash severity in Australia. Journal of safety research, 76, 73-82.

Tzortzi, A., Kapetanstrataki, M., Evangelopoulou, V., & Behrakis, P. (2021). Driving behavior that limits concentration: a nationwide survey in Greece. International journal of environmental research and public health, 18(8), 4104.

World Health Organization. (2018). Global status report on road safety 2018. World Health Organization.

Xu, C., Bao, J., Wang, C., & Liu, P. (2018). Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. Journal of safety research, 67, 65-75.

Yannis, G., & Laiou, A. (2014). Current road safety trends in Greece. Periodica Polytechnica Transportation Engineering, 42(2), 145-151.

Yannis, G., & Papadimitriou, E. (2012). Road safety in Greece. Procedia-social and behavioral sciences, 48, 2839-2848.

Yannis, G., Golias, J., & Papadimitriou, E. (2007). Accident risk of foreign drivers in various road environments. Journal of safety research, 38(4), 471-480.

Zwier, R., Nijkamp, P., Hiemstra, F., & Montfort, K. V. (1995). Connectivity and isolation in transport networks: a policy scenario experiment for the Greek island economy. In Overcoming isolation (pp. 205-234). Springer, Berlin, Heidelberg.