www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 13(4), Dec 2022, 10-25. ISSN: 2007-1558.

Fatal Cyclist-car Accidents Scenarios at Intersections from the Guadalajara Metropolitan Area

Ramon A. Briseño[1], Rocio Maciel Arellano[2], Edgar Cossio[3], Víctor M. Larios[2], Raul J. Beltrán[1], J. Antonio Orizaga T[1]

[1]Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Doctorado en Tecnologías de Información, México.
[2]Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Centro de Innovación en Ciudades Inteligentes, México.
[3]Instituto de Información Estadística y Geográfica de Jalisco, México.

alejandro.bmartinez@alumnos.udg.mx, rmaciel@cucea.udg.mx, edgar.cossio@iieg.gob.mx, vmlarios@cucea.udg.mx, raul.beltran@academicos.udg.mx, jose.orizaga@academicos.udg.mx

Abstract. Faced with the imminent high fatal cyclist-car accident rate in the Guadalajara Metropolitan Area in recent years, it is necessary to implement mechanisms to improve the safety of cyclist mobility. This research analyzes the principal factors and patterns in cyclist-car accidents at intersections in the Guadalajara Metropolitan Area through machine learning algorithms and statistical methods to identify risk scenarios. The data show that the most dangerous intersection consists of one main street and a street. The type of vehicle most involved in accidents with cyclists is public transport. Factors such as the speed limit can increase the risk on some roads. Furthermore, with relevant factors and patterns, some risk scenarios were identified. Also, the scenarios show different interest situations. Women might prefer to travel on main streets, public transport vehicles are hazardous in secondary streets, and cycling infrastructure can decrease the risk at an intersection.

Keywords: Fatal cyclist-car accidents, Intersections, Sustainable mobility, Artificial intelligence, Smart City.

Article Info
Received March 29, 2022
Accepted August 12, 2022

1 Introduction

One of the premises of a smart city is generating clean and sustainable mobility where the citizens can move in an agile and safe way. Cyclist mobility can be part of one solution in clean mobility improvement. However, according to the World Health Organization, bikers are part of the most vulnerable sector of the public road [1]. Therefore, bikers are likely to die or suffer serious harm to their health in a traffic accident.

There are three different types of cyclist accidents. A single-bicycle accident (when a cyclist falls or crashes with an object) [2]. Bicycle-bicycle accidents [3], and cyclist-car accidents [4]. This work focuses on cyclist-car accidents because those are the most reported and dangerous incidents for the cyclist's community [4, 5]. In the Guadalajara Metropolitan Area (GMA), the governmental and academic authorities are working to transform mobility into smart and sustainable mobility. The city of Guadalajara has more than 100km of bike paths [6], the same growing ones. Also, the government, academia, and industry are creating mechanisms to incentive bicycle mobility; one of them is the IoP (Internet of People) Jalisco [7].

Despite the continuous growth of cycling infrastructure in the GMA, the biker community has many fatal accidents yearly. The average of lost lives was 23 per year from 2009 to 2019 [8]. Furthermore, when it seemed that the accident rate tended to decrease, the year 2019 compared with 2018 presented an increase of 53.8% of deaths [9] (see Fig. 1). Years 2020 and 2021 were atypical because the Covid-19 pandemic caused a strong dismissal of mobility in Mexico [10].

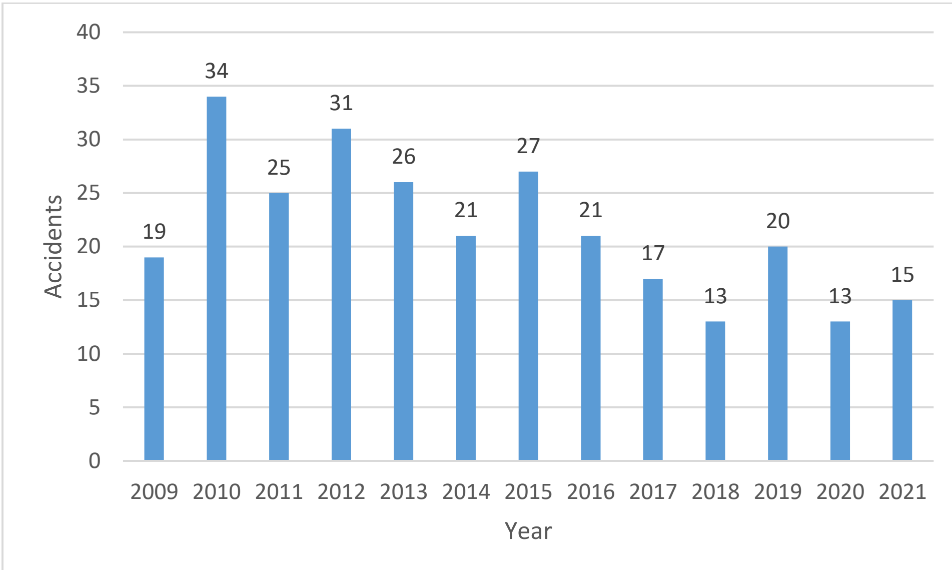

Fig. 1. Death of cyclists on public roads from 2009 to 2021 [8].

This paper reviews the cyclist-car accident literature, emphasizing the intersections of streets and entrances and exits of car parks and shops. Since, from different cities globally, the crossroads and intersections are the most frequent places for cyclist-car accidents [4, 5, 11], the main driver of this work is to find factors and patterns with a high impact on GMA fatal cyclist-car accidents at intersections to identify risk scenarios. For this research, we define scenario as a frequent combination of infrastructure, cyclist, and vehicle factors involved in a cyclist-car accident. Based on historical data, we propose identifying accident risk cyclist-car scenarios with three goals. First, to create tools for the government could identify better how to allocate their resources to improve the overall safety for citizens moving on bicycles. Second, provide simple but valuable information to citizens to increase awareness of where risk scenarios of accidents are part of their trip through a web platform. Third, the possibility of deciding where to place internet of things objects in dangerous crossroads to mitigate as accidents as possible.

This work is organized in the following sections: section two mentions the risk factors for the cyclist-car accidents at intersections found in the literature. Section three relates the GMA problem, explains the primary data, and proposes an analysis with descriptive statistics, contingency tables, and clustering and classification machine learning algorithms to identify the most common factors and patterns in the database [8]. Also, section 3 formulate risk scenarios by integrating relevant found factors and patterns. Section four analyses and discusses the important findings of the proposal. Finally, section five concludes and presents the following steps on this work.

2 Literature Review

Several factors influence the causality of cyclist-car accidents at intersections. These factors correspond to the infrastructure of the streets, type of vehicle involved, drivers maneuver and aptitudes, the time and weather, and the density of cyclists and vehicles.

Some works concluded that the most dangerous drivers' maneuver moving is when a vehicle turns right and impacts a cyclist who comes from the right, left, or the same direction [4, 12, 13, 14, 15]. The situation that creates the most conflicts is when a vehicle starts to drive after the red light, and a cyclist passes through the intersection without stopping [13]. Also, when the involucre vehicle is a truck or a bus, the risk increase [11, 13, 14, 15, 16]. Additionally, most accidents with trucks participation are in traffic lights intersections [13, 15]. The principal attitudes that promote an accident are breaking the traffic rules [4, 16], particularly ignoring the red lights and stop signs. Also, cyclists tend to invade other public road areas such as the pedestrian zone to prioritize crossing and avoid the red light, increasing the risk of an accident [17].

The presence of cyclist infrastructure [4, 11, 2] and the distance between the vehicles and cyclists [14] decrease the accident risk. This situation allows drivers to see the cyclists with more time and a better angle vision [12]. In addition, separating too much the cycle road from the street can cause visual obstacles [13]. Another aspect that reduces the visual angle is the geometry of the intersection. The intersections with angle orthogonal ($85<\times<95$ grades) are safer than no orthogonal intersections ($0<\times\leq 85$ or $95\leq\times<175$ grades). Thus, the visual angle is lower in no orthogonal intersections, and drivers and cyclists have less time to react and evict a collision [11]. Also, the time of the accident affects visibility; accidents at night have a significantly higher risk [11, 16]. Speed limit and traffic flow are two variables in a constant study for traffic accidents. Cyclist-car accidents are not the exception; a speed limit higher than 30 km/h increases the risk [4, 18]. Furthermore, the more significant the traffic flow of both vehicles and cyclists increases the risk [12, 18].

Roundabouts are the intercessions with the highest risk index [18, 19, 20, 21]. That situation is explained by a higher concentration of conflict points [21] and a relationship with an unfavorable geometry, especially in small roundabouts [18, 21]. Also, the streets' width, the importance [20], and the number of exits from a roundabout [21] are highly relevant to the probability that a cyclist suffers an accident with a motorized vehicle.

Besides traffic density, another essential aspect in collisions between vehicles and cyclists is the density of cyclists on the streets. A cyclist is more likely to suffer an accident when traveling alone than in group [22].

To determine the group of relevant variables, the review works utilized techniques like frequencies analysis [4, 12, 13, 14, 19], multivariable logistic regression [11, 15, 16, 18, 21, 23], conditional logistic regression [20], hierarchical regression and exploratory factor analysis with Varimax Rotation [17], and latent class and association rules [23]. With this, we can observe that the most common techniques to determine risk factors in cyclist-car accidents in this literature review are frequency analysis and logistic regression in its multiple modalities.

It is possible to watch that in regression analysis, the target variables most used are the severity of the accident [11, 15, 16, 23] followed by the location [18, 20, 21], and patterns [17, 23]. In some cases, the dependent variable is chosen at the beginning of the study [11, 15, 18, 20, 21]. Moreover, in other cases is dieted by the process of the analysis [17, 23] like in this research.

3 Proposal

The high rate of fatal accidents with cyclists in the GMA represents a latent problem for cyclists. Of the total casualties, 74.1% are at intersections. Therefore, this work proposes to identify frequent risk-fatal cyclist-car accident scenarios from the GMA, analyzing the open database of White Bicycle organization. Section 3.1 describes the structure of the database. Section 3.2 explains the methodology of this work; wherewith the help of machine learning algorithms and statistical methods, it identifies risk scenarios. Section 3.3 points out the intersections that fit the different fatal cyclist-car accident scenarios for a polygon of the GMA.

3.1 Database description

From January 2009 to January 2022, the White Bicycle organization registered 283 cyclists' deaths in the Guadalajara Metropolitan Area in its open database [8]. Two hundred forty-eight records were identified as a cyclist-car accident, six as falls, two collisions with objects, and 27 where the cause is unknown. Of the 248 fatal cyclist-car accidents, 59 occurred on a stretch without intersections, five times the location was not identified, and 184 at an intersection. For the analysis, it's used 184 cyclist-car accidents at intersections.

The Fatal cyclist-car accidents database initially contains the variables of *Sex*, *Type of road*, *Age*, *Type of vehicle*, and *Location* of the accident site. Then, to obtain the most significant number of variables found in the literature review, it's used the street view of Google maps and the map of the Moovit platform for each location. With these tools, the variables: *Number of lanes per direction*, *Speed limit* and *Number of directions* of the most significant road involved, and the *Number of entrances and exits* in each intersection were added. In addition, the binary variables: *Presence of cycling infrastructure*, *Presence of public transport routes*, *Orthogonal intersection* (Fig. 2 explains orthogonal intercessions), *Roundabout*, and *Traffic light intersection* were also added. Finally, the variable *Type of road* that described only one road involved in the intersection was modified to express two roads and was called "*Type of intersection*." Table 1 shows the structure of the database records. For the variable *Type of intersection*, avenues, carriageway, peripherals, and highways were classified as “Main street“. Also, streets were called to all those secondary access roads that do not fall into the avenues. For the variable *Type of vehicle*, the *Private car* class includes sedan-type cars, motorcycles, and pickups. The *Truck* class represents cargo vehicles, pipes, and trailers. Finally, the *Public transport* class contains the public transport buses and trains. On the other hand, the variable *Speed limit* included *30km/h*, *50km/h, 60km/h, and 80km/h* classes. Still, because the classes *30km/h and 60 km/h* only contained one record for each class, the variable *Speed limit* was discretized to 2 classes.

Table 1. Structure of database records [8]

variable	class	class meaning
Sex	M	Male
	F	Female
Type of vehicle	Public transport	Public transport
	Private car	Private car
	Truck	Truck
	Unidentified	Unidentified
Age	0-19	0-19 years old
	20-39	20-39 years old
	40-59	40-59 years old
	60+	60 or more years old
	Unidentified	Unidentified
Number of lanes per direction	1c	One lane
	2c	Two lanes
	3c+	Three or more lanes
Speed limit	50km/h	50 km/h
	80km/h	80 km/h
Number of directions	1s	One direction
	2s	Two directions
Number of entrances and exits	3e	Three entrances and exits
	4e	Four entrances and exits
	5e+	Five or more entrances and exits
Cycling infrastructure	Yes	Yes
	No	No
Public transport routes	Yes	Yes
	No	No
Orthogonal intersection	Yes	Yes
	No	No
Roundabout	Yes	Yes
	No	No
Traffic light intersection	Yes	Yes
	No	No
Type of intersection	Main street-main street	Two main streets
	Main street-street	Main street and a street
	Street-street	Two non-main streets

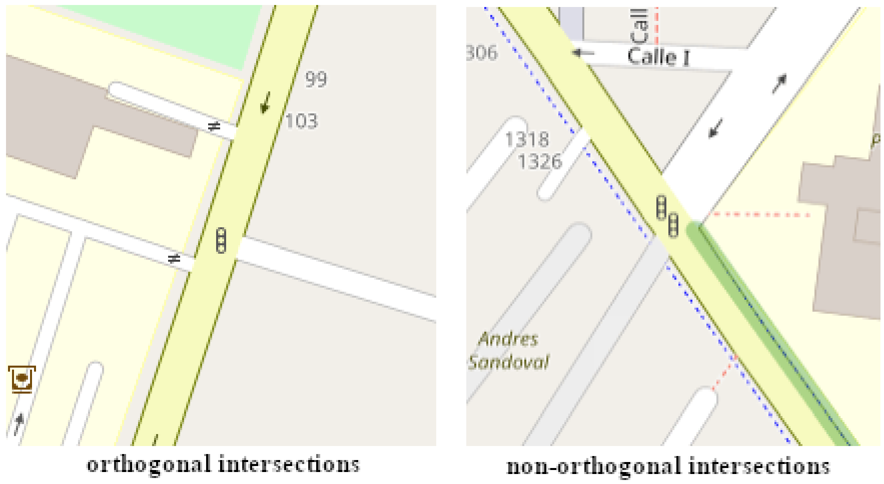

Fig. 2. The intersection is orthogonal when the street angle is between 85 and 95 degrees, and otherwise, the intersection is non-orthogonal. The left side of the figure shows orthogonal intersections, while the right side contains non-orthogonal intersections [11].

3.2 Methodology

In order to identify the most common risk scenarios, the data analysis includes two main steps: find factors and patterns and formulate scenarios. First, to find factors and patterns, four techniques are used. A descriptive statistical analysis (section 3.2.1) highlights the main characteristics of fatal accidents at intersections from the GMA. A contingency tables analysis (section 3.2.2) analyzes the relationship between variables' classes. A clustering algorithm (Section 3.2.3) identifies associations between factors. Also, a classification machine learning algorithm (section 3.2.4) selected the variable that can be better explained, and factors with positive influence for each classification are determined.

Second, it is identified scenarios integrating found factors and patterns from the previous step results. Fig. 2 shows the process of the methodology.

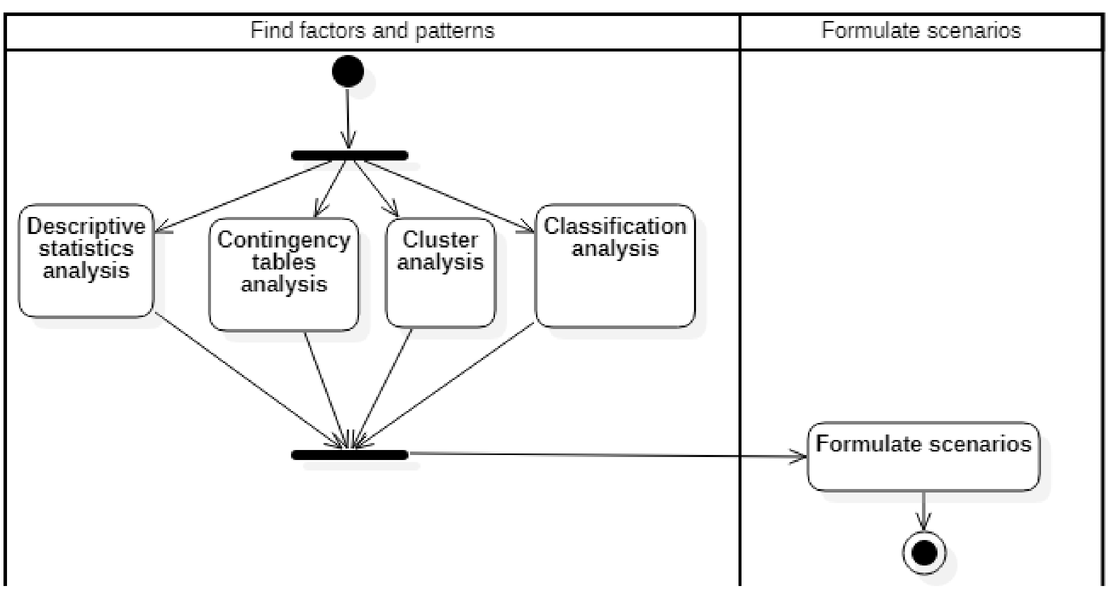

Fig. 2. Diagram of the methodology

3.2.1 Descriptive statistics analysis

The descriptive statistics analysis observed that 168 (91.30%) victims are male, and 16 (8.70%) are female [8]. This makes sense with the information presented by the Institute of Statistical and Geographical Information of Jalisco (IIEG for its acronym in Spanish), which mentions that for the trips of the GMA public bicycle sharing system MiBici (from December 2020 to December 2021), only 21.1% were made for females [24].

In contrast to what is mentioned in [18, 19, 20, 21], roundabouts are not the most dangerous intersections speaking about fatal cyclist-car accidents for the GMA. Because of the 184 incidents, only 7, equivalent to 3.80%, occurred in a roundabout. Furthermore, it was found that, in contrast to [21], increasing the number of entrances and exits from an intersection does not increase the risk of a fatal cyclist-car accident. The predominant number of entrances and exits at intersections was 4 with an occurrence of 55.98%, followed by 3 with 40.22%, intersections with five or more entrances and exits had an event of 3.80%.

On the other hand, it cannot assure that the existence of cycle lanes reduces the number of accidents, as mentioned in [13]. However, the records show that 92.39% of fatal accidents occurred at intersections without bicycle infrastructure.

As mentioned in [11], fatal accidents involving public transport and trucks are more frequent than accidents with small vehicles for the GMA. Accidents with public transport and trucks represent 56.62%. Thus, public transport is the enemy number one for cyclists with 42.39% of incidents, followed by private cars with 40.76%. Similar to that mentioned in [20], the importance of the streets is of great impact since only 25% of the recorded intersections do not contain main roads. The most common intersection consists of one main street and a street, present in 50% of the occasions. In addition, 99.27% of the intersections with at least one main street have public transport routes, and 97.82% are two-way traffic. Also, the number of lanes per street increases according to the importance of the streets in the intersection.

It is detected that cyclists between the ages of 20 and 39 are most vulnerable to fatal accidents. It seems that fatal accidents are caused when the motorized vehicle travels at more than 30 km/h, as mentioned [18]. For this study, 86.41% of the accidents occurred at intersections with a speed limit of 50 km/h and 13.56% with 80 km/h. The classes considered as a factor with at least 85% of incidence in the total of records are:

1) *No*, from *Cycling infrastructure*.
2) *Yes*, from *Public transport routes*.
3) *No*, from *Roundabout*.
4) *M*, from *Sex*.
5) *2s*, from *Number of directions*.
6) *50km/h*, from *Speed limit*.

3.2.2 Contingency tables analysis

A relationship between two or more classes can show a causal factor in increasing or decreasing accidents. It used contingency tables with Standardized Pearson Residuals (SPR) to identify the events that occur greater or lesser than expected. Standardized Pearson Residuals are calculated for each cell of the contingency tables as:

$$SPR_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}(1 - R_i/T)(1 - C_j/T)} \quad (1)$$

where:

SPR_{ij} = The Pearson residual score for the cell in the i column and j row.
O_{ij} = The observed value for the cell in the i column and j row.

E_{ij} = The expected value for the cell in the i column and j row. The expected value is calculated with the quotient between the product of the row and column totals and the grand total.
R_i= The row total divided by the grand total.
C_j = The row total divided by the grand total.
T = The grand total.

SPRs have a mean of 0 and a standard deviation of 1. To select a deviation as relevant must meet two conditions; it has to have an SPR >=2 or SPR<=-2 and show the presence of a phenomenon. For example, suppose more accidents than expected are found at intersections type *Main Street-main street* with a speed limit of *80 km/h*. The SPR of the cell is 3.1. Also, it is suspected that the most common speed limit is *50 km/h* at this type of AMG intersection. In that case, the deviation is relevant because the phenomenon is that speed could be a risk factor, and the SRP is higher than 2.

The following relevant deviations were found in the contingency table for each combination of two and three dependent variables Using Orange software [25]:

1) 6.5 more accidents than expected were found in the type of intersection *Main street-main street*, where the arteries' speed limit is *80 k/m*.
2) 11.9 more accidents than expected were found in intersections where one street has a speed limit of *80 km/h* and *three or more lanes per direction*.
3) 5 more accidents than expected were found in intersections with *5 or more entrances and exits* is a *roundabout*

3.2.3 Cluster analysis

Three groupings were made with the k-means unsupervised algorithm of the Orange software. The algorithm aims to find groups of records where the within-group variance is minimized and is based on the least-squares principle [26]. Given n observation, K number of clusters (K≤n) and C= $\{C_I,\ C_2, C_3 \ldots C_k\}$ groups, the principle of the algorithm is:

$$Arg\ min\ C\ \sum_{i=1}^{K}\sum_{x\epsilon Ci}(x-\mu_i)^2 \qquad (2)$$

where:

$Arg\ min$ = minimum argument
K = number of clusters
C = set of clusters
μ = the mean of points in C_i

It was chosen to work with 2, 3, and 4 clusters because, according to the silhouette method, the association in 2 groups has the best scored, followed by 3 and 4 groups. Also, since the variables involved are nominal categorical of two, three, and four classes, it was possible to observe how they are distributed and associated in the centroids of the clusters. Tables 2, 3, and 4 show the frequency and centroids with variations for each cluster of the different grouping. To recognize an association of factors as a pattern, this association must show the presence of a phenomenon. Once defining an association of different classes as a pattern, it is corroborated that the pattern found has at least a 20% higher incidence than any other combination of classes of the variables involved.

Between the centroids of clusters from the three groupings it was found the following patterns:

1) *Public transport*, *1 line per direction*, *orthogonal intersection*, *traffic light intersection*, *4 Entrances and exits*, and type of intersection *Street-stree*t were found together in one cluster for each aggrupation.
2) *Private car*, *3 or more lanes per direction*, *no orthogonal intersection*, and type of intersection *Main street-street* were found together in one cluster for each aggrupation.

3) *Private car, 3 or more lanes per direction*, *no orthogonal intersection*, *no traffic lights intersection, 3 entrances and exits*, and type of intersection *Main street-street* were found together in one cluster for the three and four clusters aggrupation.
4) *Public transport*, *3 or more lanes per direction*, *traffic lights intersections*, *4 entrances and exits*, and Type of intersection *Main street-main street* were found together in one cluster for the three and four clusters aggrupation.

Table 2. Centroids of the variables that present variation in the K-means of 2 clusters

Attribute	Cluster 0	Cluster 1
Number of records	138	46
Age	20-39	Unidentified
Type of vehicle	Private car	Public transport
Number of lanes per direction	3c+	1c
Orthogonal intersection	No	Yes
Traffic light intersection	Yes	No
Entrances and exits	4e	4e
Type of intersection	Main street- street	Street-street

Table 3. Centroids of the variables that present variation in the K-means of 3 clusters

Attribute	Cluster 0	Cluster 1	Cluster 2
Number of records	70	72	42
Age	40-59	20-39	Unidentified
Type of vehicle	Private car	Public transport	Public transport
Number of lanes per direction	3c+	3c+	1c
Orthogonal intersection	No	Yes	Yes
Traffic light intersection	No	Yes	No
Entrances and exits	3e	4e	4e
Type of intersection	Main street- street	Main street-main street	Street-street

Table 4. Centroids of the variables that present variation in the K-means of 4 clusters

Attribute	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Number of records	54	31	60	39
Age	40-59	20-39	20-39	0-19
Type of vehicle	Public transport	Public transport	Private car	Public transport
Number of lanes per direction	2c	3c+	3c+	1c
Orthogonal intersection	Yes	No	No	Yes
Traffic light intersection	Yes	Yes	No	No
Entrances and exits	4e	4e	3e	4e
Type of intersection	Main street-street	Main street-main street	Main street-street	Street-street

3.2.4 Classification analysis

The classification analysis used a logistic multinomial regression algorithm. The algorithm calculates the probability for each class of the dependent variable. The probabilities for each class are associated, which means that the sum of class probabilities equals 1, expressed as:

$$\sum_{i=1}^{C} p_i = 1 \qquad (3)$$

where:

p_i= probability of class i.
C = classes of the dependent variable.
Also, to calculate the probability for a class, the model implements the function:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \qquad (4)$$

where:

$z_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + \alpha_{i3}x_3 + \ldots \alpha_{in}x_n + \beta_i$
x = independent variable.
n = the number of independent variables.

The variable with the best precision for the classification machine learning algorithm was *Type of intersection* (selected as the dependent variable). The independent variables used to build de model are the *Number of lines per direction*, *Public transport routes*, *Speed limit*, *Orthogonal Intersection*, *Traffic light intersection*, *Number of directions*, *Sex*, and *Number of entrances and exits*. These variables were selected using a backward stepping methodology from all the database variables. The algorithm gets a precision of 76%, an area under the curve of 85.5%, a recall of 76.7%, and an F1 score of 74.7%.

Fig. 3. Classification analysis nomogram.

The followings combination of factors offers a high possibility of good classification for each class of *Type of intersection*:

1) For an intersection type *Main street-main street*, the following factors influence positively descending

order: *2 and 3 or more number of lines per direction*, *80 km/h*, *two directions*, *4 and 5 or more entrances and exits, no orthogonal intersection*, *traffic light intersection*, *Female*, and *public transport routes*. This combination of factors gets at least 93% of the probability of classifying an intersection like *Main street-main street* (see Fig. 3).

2) For an intersection type *Main street-Street*, the following factors influence positively descending order: *two directions, 2 and 3 or more lines per direction, public transport routes, 4 and 3 entrances and exits*, *without traffic lights*, *orthogonal intersection*, *50 km/h*, and *Male*. This combination of factors gets at least 84% of the probability of classifying an intersection like *Main street-street* (see Fig. 3).
3) For an intersection type *Street-street*, the following factors positively influence descending order: *2 and 1 lane per direction*, *one direction*, *50 km/h*, *without public transport routes*, *orthogonal intersection*, *Male*, *without traffic lights*, *and 3 and 4 entrances and exits*. This combination of factors gets at least 93% the probability of classifying an intersection like *Street-street* (see Fig. 3).

3.2.5 Formulate scenarios

To take a series of factors as a scenario, first, collect all factors from each type of intersection's classification and cluster analysis results. Second, it adds factors found by statistic description and corresponding contingency tables. It ensures that collected and added factors do not contradict each other. Suppose a factor found by descriptive statistics or contingency tables controvert a factor found by clustering or classification. In that case, the variable is removed from the scenario. This situation means that the classes of the said variable are distributed uniformly in that scenario and do not show a tendency. Fig. 4 shows the formulate scenarios process.

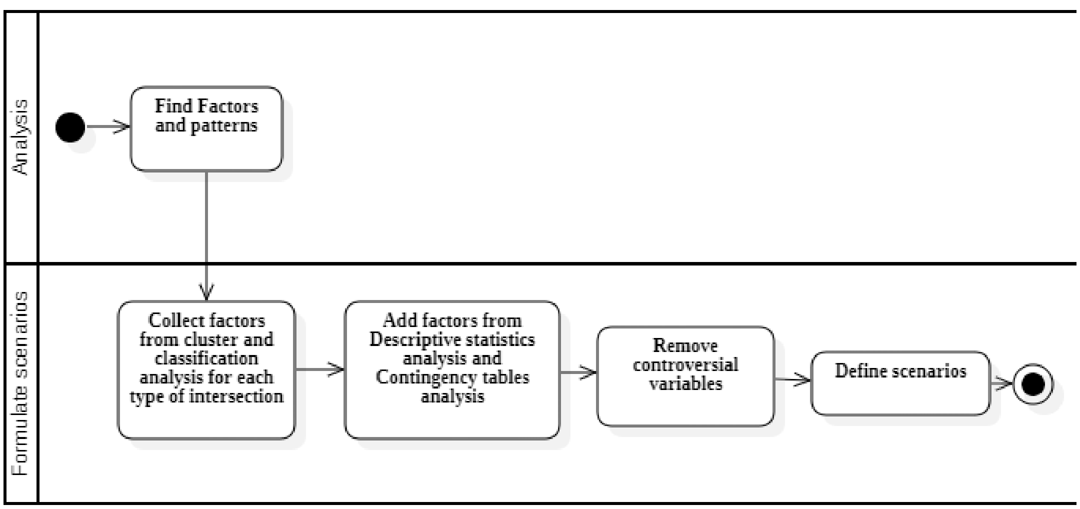

Fig. 4. The process to define scenarios

Following the formulate scenarios process, 3 principal scenarios were identified:

1) For *Main street-main street* type of intersection, it collected the factors *2 and 3 or more number of lines per direction*, *80 km/h, two directions*, *4 and 5 or more entrances and exits*, *no orthogonal intersection*, *traffic light intersection*, *Female*, and *public transport* routes by the classification analysis; *Public transport, 3 or more lanes per direction*, *traffic lights intersections*, and *4 entrances and exits* by the pattern 4 of cluster analysis. Also, it added *no Cycling infrastructure, Public transport routes, no Roundabout, Male, two directions, and 50km/h by* statistic description and *80 km/h, 3 or more lanes per direction, 5 or more entrances and exits,* and *Roundabout* by contingency tables. The variables *Speed Limit, Roundabout, and Sex* were removed from the scenario because their factors found in the different analyses controvert each other. Then, the scenario for the *Main street-main street* intersection contains the following factors: *2 and 3 or more lanes per direction, 2 directions, 4 and 5 or more entrances and exits, non-orthogonal intersection, traffic light intersection, Public transport routes,* and *No cycling infrastructure*. Also, by limiting the variables *Entrances and exits* to 4 and *Number of lanes*

per direction to *3 or more*, it's possible to add *Public transport* type of vehicle to the scenario (see table 5).

2) For *Main street-street* type of intersection, it collected the factors *two directions, 2 and 3 or more lines per direction, public transport routes, 4 and 3 entrances and exits*, *without traffic lights, orthogonal intersection, 50 km/h* and *Male* by the classification analysis; *Private car, 3 or more lanes per direction*, and *no orthogonal intersection* by pattern 2 of cluster analysis*; Private car, 3 or more lanes per direction, no orthogonal intersection, no traffic lights intersection and 3 entrances and exits by pattern 3 of cluster analysis.* Also, it added *no Cycling infrastructure, Public transport routes, no Roundabout, Male, two directions, and 50km/h by* statistic description and *80 km/h, and 3 or more lanes per direction* by contingency tables. The variables *Orthogonal intersection* and *Speed limit* were removed from the scenario because their factors found in the different analyses controvert each other. Then, the scenario for the *Main street-street* intersection contains the following factors: *2 directions, 2 and 3 o more lines per direction, Public transport routes, 4 and 3 entrances and exits, no traffic lights intersection, Male, No roundabout,* and *No cycling infrastructure*. Limiting the variables *Number of lines per direction* to 3 or more, and *Number of entrances and exits* to 3 makes it possible to add a *Private car* type of vehicle to the scenario.

3) For *Street-street* type of intersection, it collected the factors *2 and 1 lane per direction, one direction, 50 km/h, without public transport routes, orthogonal intersection, Male, without traffic lights, and 3 and 4 entrances and exits* by the classification analysis; *Public transport, 1 line per direction, orthogonal intersection, traffic light intersection, 4 Entrances and exits* by the pattern 1 of cluster analysis. Also, it added *no Cycling infrastructure, Public transport routes, no Roundabout, Male, two directions, and 50km/h by* statistic description. The variables *Number of directions, traffic light intersection, and Public transport routes* were removed from the scenario because their factors found in the different analyses controvert each other. Then, the scenario for the S*treet-street* intersection contains the following factors: *2 and 1 line per direction, 50 km/h, Male, 3 and 4 entrances and exits, Orthogonal intersection, No roundabout, and No cycling infrastructure*. Limiting the variables *Number of lines per direction* to 1, and *Number of entrances and exits* to 4, it's possible to add the *Public transport* type of vehicle to the scenario.

Table 5. visual representation of the formulating of scenario 1 for Main street-main street type of intersection.

Processes / variables	Collet factors from classification analysis	Collet factors from cluster analysis	Add factors from descriptive statistic analysis	Add factors from contingency table analysis	Remove controversial variables	Define factors as scenario
Sex	F	-	M	-	Sex	-
Type of vehicle	-	Public transport	-	-	-	Public transport
Age	-	-	-	-	-	-
Number of lanes per direction	2c, 3c+	3c+	-	3c+	-	2c, 3c+
Speed limit	80km/h	-	50km/h	80km/h	Speed limit	-
Number of directions	2s	-	2s	-	-	2s
Number of entrances and exits	4e, 5e+	4e	-	5e+	-	4e, 5e
Cycling infrastructure	-	-	No	-	-	No
Public transport routes	yes	-	yes	-	-	Yes
Orthogonal intersection	No	-	-	-	-	No
Roundabout	-	-	No	Yes	Roundabout	-
Traffic light intersection	Yes	Yes	-	-	-	Yes

Table 6. visual representation of the formulating of scenario 2 for Main street-street type of intersection.

Processes / variables	Collet factors from classification analysis	Collet factors from cluster analysis	Add factors from descriptive statistic analysis	Add factors from contingency table analysis	Remove controversial variables	Define factors as scenario
Sex	M	-	M	-	-	M
Type of vehicle	-	Private car	-	-	-	Private car
Age	-	-	-	-	-	-
Number of lanes per direction	2c, 3c+	3c+	-	3c+	-	2c, 3c+
Speed limit	50km/h	-	50km/h	80km/h	Speed limit	-
Number of directions	2s	-	2s	-	-	2s
Number of entrances and exits	3e, 4e	3e	-	-	-	3e, 4e
Cycling infrastructure	-	-	No	-	-	No
Public transport routes	yes	-	yes	-	-	Yes
Orthogonal intersection	Yes	No	-	-	Orthogonal intersection	-
Roundabout	-	-	No	-	-	No
Traffic light intersection	No	No	-	-	-	No

Table 7. visual representation of the formulating of scenario 3 for Street-street type of intersection.

Processes / variables	Collet factors from classification analysis	Collet factors from cluster analysis	Add factors from descriptive statistic analysis	Add factors from contingency table analysis	Remove controversial variables	Define factors as scenario
Sex	M	-	M	-	-	M
Type of vehicle	-	Public transport	-	-	-	Public transport
Age	-	-	-	-	-	-
Number of lanes per direction	1c, 2c	1c	-	-	-	1c, 2c
Speed limit	50km/h	-	50km/h	-	-	50km/h
Number of directions	1s	-	2s	-	Number of directions	-
Number of entrances and exits	3e, 4e	4e	-	-	-	3e, 4e
Cycling infrastructure	-	-	No	-	-	No
Public transport routes	No	-	yes	-	Public transport routes	-
Orthogonal intersection	Yes	Yes	-	-	-	Yes
Roundabout	-	-	No	-	-	No
Traffic light intersection	No	Yes	-	-	Traffic light intersection	-

3.3 Identification of intersections with risk of fatal cyclist-car accident

The infrastructure and public transport routes variables corresponding to each accident scenario were interpolated at 750 intersections of a polygon from the GMA to identify the points where a fatal accident could occur. The interpolation identified two intersections for scenario 1 (*Main street-main street*), 28 intersections for scenario 2 (*Main street-street*), and 45 for scenario 3 (*Street-street*). Fig. 5 shows the spatial distribution of the points identified using QGIS software.

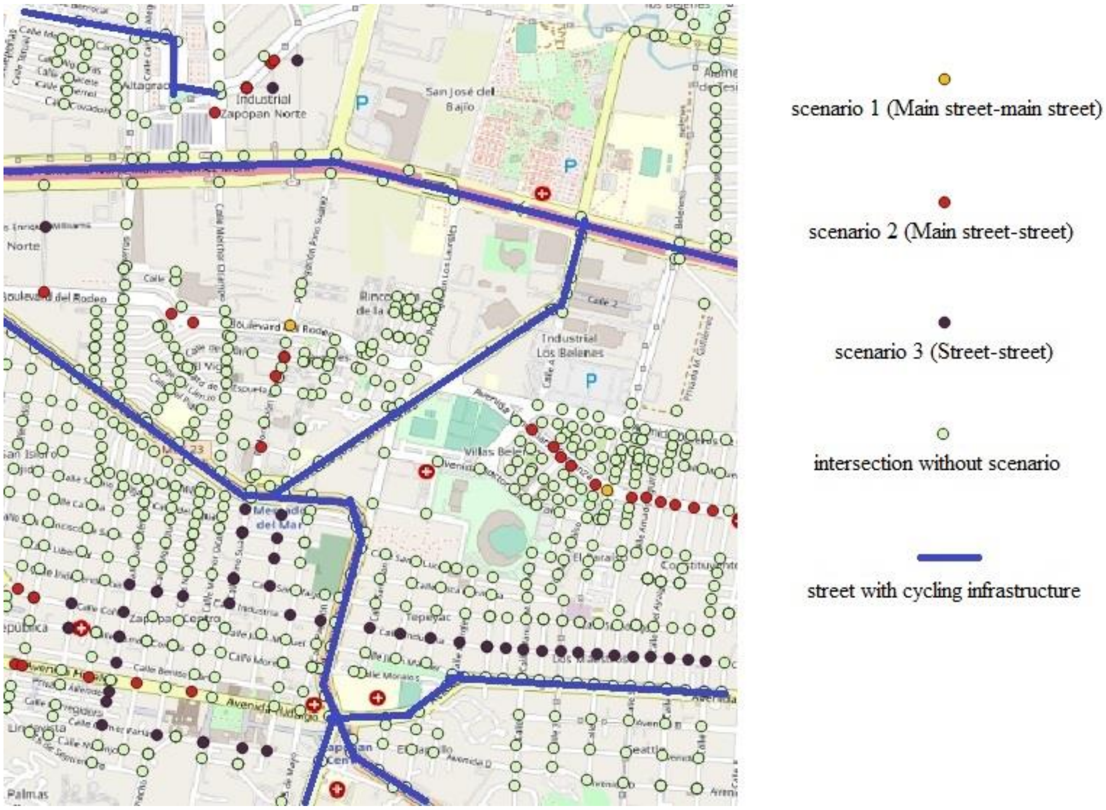

Fig. 5. Identification of possible fatal cyclist-car accident scenarios in a Polygon of the GMA.

4 Results and Discussion

Based on the descriptive statistics analysis, the analysis with contingency tables, the cluster analysis, and the classification analysis, the following relevant results are discussed:
The descriptive statistics analysis shows that the most frequent type of intersection in fatal cyclist car accidents is formed by one main street and a street. Also, enemy number one of the cyclist is the buses and trains. The relevant situation since the number of *public transport* vehicles that travel in the GMA is lower than that of *private vehicles*. Yet they have a greater number of incidents in accidents [27].

In contingency table analysis, It is observed that speed can trigger the risk of suffering a fatal accident at the main streets of *3 or more lines per direction*. In addition, although in the review of the literature some works [18][21] indicated that small roundabouts are the most dangerous, due to the *number of entrances and exits* (5 or more) in intersections with a history of fatal accidents for the GMA indicates that the roundabouts involved are large.

By the clustering analysis, in pattern 1, *Public transport* is the only type of vehicle associated with intersections type *Street-street* and *one lane per direction*. In patterns 1 and 4, *Public transport* is associated with *4 entrances and exits* and *traffic lights intersection*. The latter could indicate that buses are the ones that least respect traffic lights. On the other hand, in patterns 2 and 3, *Private car* is related to *non-orthogonal*, *Main street-street* intersection where at least a road has *3 or more lanes per direction*.

Unlike the clustering analysis, the classification analysis did not consider the *Type of vehicle* as a predictor variable. However, it does consider factors of variables such as *Sex*, *Speed limit*, and the *presence of public transport routes* that are not explained in the cluster analysis, which shows a clearer trend of what kind of intersection we could find these factors.

Finally, the following findings are discussed for each of the 3 scenarios identified with the results of the four analyst techniques used:

1) The scenario for the *Main street-main street* intersection does not define the sex involucrate, considering that males are present in 91.30% of casualties, which can mean that women prefer to travel on main streets. Also, by the relation found between 5 or more entrances and exits and roundabouts, the major concentration of roundabouts is in *Main Street-Main street* intersections. On the other hand, the *speed limit* is not defined. That means this type of intersection has a significant concentration of records with a speed limit of *80 km/h*, even though the most common speed limit is 50 km/h with an incidence of 86.41%.
2) In the *Main street-street* intersection scenario, the variable *Orthogonal intersection* is not defined, which means this variable is distributed uniformly in *Main street-street* intersections. Also, similar to Main street-main street type of intersections the *speed limit* is not defined. That means this type of intersection has a significant concentration of records with a speed limit of *80 km/h.*
3) The scenario for the *Street-street* intersection does not define the variable *Number of directions*, considering that 86.41% of casualties are two directions, which means this scenario presents the mayor's concentration of intersections with one direction. Also, the scenario does not define de variable *Public transport routes*. Considering that 92.39% of records occurred in streets with transport routes, this scenario contemplates the major concentration of causalities in streets without transport routes. In this scenario, where intersections have *4 entrances and exits* and *one line per direction*, the *Public transport* type of vehicle is defined even though the factor of *public transport routes* is not. That situation could indicate that buses are the most dangerous vehicle on small secondary streets with *public transport routes*. Finally, the scenario does not define the variable *traffic light intersection,* which it is distributed uniformly.

The variable *Age* is not present in the scenarios because it has a considerable dispersion for each type of intersection. Class *No* of the variable *cyclist infrastructure* is the only constant in the scenarios. It can mean that bicycle facilities decrease the risk of fatal cyclist-car accidents at an intersection.

In identifying risk intersections, Of the 75 intersections indicated, 45 correspond to *street-street* intersections and only two to *Main street-main* street intersections. Two reasons could explain this phenomenon: one, in the selected polygon, many main streets have cycling infrastructure, which reduces the risk of these streets, and second, there are more *Street-street* intersections than *Main street-main street* and *Main street-street* in the city.

Probably for the intersections found in the polygon studied, the incorporation of cycling infrastructure could improve cycling safety. However, other strategies must be taken in the streets where it is impossible to incorporate bike lines. Some of which may be to define speed limits or priority signaling.

Other works use analysis of frequencies [4, 28] and clustering of unsupervised machine learning techniques [29] for formulating accident scenarios. These techniques efficiently identify frequent events but don't clearly show events that appear little in the dataset but show an essential aspect for a scenario. The methodology of this work analyzes frequencies and clusters but also adds methods such as contingency tables and multinomial logistic regression allowing for identifying events of risk that do not frequently appear in the dataset. For example, thanks to these tools were possible to watch the interaction of the woman and roundabouts in scenario 1 for *Main street-main street* intersections and the speed limit in scenarios 1 and 2 for *Main street-main street* and *Main street-street* intersections.

5 Concluding Remarks and Future Work

This article described the interaction of infrastructure, vehicle, and bicyclist factors involved in fatal accidents occurring at an intersection of the GMA. Likewise, in this work's most important findings, the *Type of intersection* with the highest risk of a fatal accident is formed by one main street and a street. *Public transport* is the vehicle most involved in fatal accidents with cyclists. Roads with a higher speed limit can increase the risk of a cyclist-car fatal accident at intersections type *Main street-main street*, and intersections where at least one road has *three or more lanes per direction*. Also, the following scenarios were identified: in *Main-street-main street* type of intersection, *2 and 3 or more lanes per direction, 2 directions, 4 and 5 or more entrances and exits, non-orthogonal intersection, traffic light intersection, Public transport routes*, and *No cycling infrastructure* form a scenario; For *Main-street-street* type of intersection, *2 directions, 2 and 3 o more lines per direction, Public transport routes, 4 and 3 entrances and exits, no traffic lights intersection, Male, No roundabout*, and *No cycling infrastructure* factors form a scenario; in *Street-street* type of intersection, *2 and 1 line per direction, 50 km/h, Male, 3 and 4 entrances and exits, Orthogonal intersection, No roundabout*, and *No cycling infrastructure* factors form a scenario. Studying the scenarios and their variations in depth shows that women might prefer to travel on main streets, the *Public transport* vehicles are hazardous in secondary streets, and bicycle facilities can reduce the risk at an intersection.

Knowing these scenarios will make it possible to detect the conflict points that can lead to a fatal accident between a cyclist and motor vehicles in the intersections of the GMA. Thus, making public this information, the government can decide better, for example, where to build bicycle infrastructure and where to place bicycle sharing program stations. The citizens could plan their trips to avoid conflict points. In academics, researchers can create internet of things devices that alert drivers in a dangerous situation. In fact, with the interpolation of the scenarios, it was possible to identify 75 of 750 intersections of a polygon from the GMA where a fatal accident scenario could occur.

Also, identifying fatal cyclist-car accidents scenarios at intersections is important because, in future work, the intention is to predict causalities and classify the intersections of a polygon in the GMA based on the risk of suffering a fatal cyclist-car accident. This proposal makes it necessary to take aleatory samples of intersections with no accident history like in [18] that match the scenarios found in this work. Furthermore, for the following steps of this research, it pretends to add variables such as the traffic index and the flow of cyclists that have been predictors for cyclist-car accidents in other studies [4, 18, 30].

References

1. WHO. (s. f.). Global status report on road safety 2018. Retrieved from https://www.who.int/publications-detail-redirect/9789241565684
2. Schepers, P., & Klein Wolt, K. (2012). Single-Bicycle Crash Types and Characteristics. Cycling Research International, 2, 119-135.
3. Wallentin, G., & Loidl, M. (2016). Bicycle-Bicycle Accidents Emerge from Encounters: An Agent-Based Approach. Safety, 2(2), 14. https://doi.org/10.3390/safety2020014
4. Kuehn, M., Hummel, T., & Lang, A. (s. f.). CYCLIST-CAR ACCIDENTS – THEIR CONSEQUENCES FOR CYCLISTS AND TYPICAL ACCIDENT SCENARIOS. 8.
5. Glász, A., & Juhász, J. (2017). Car-pedestrian and car-cyclist accidents in Hungary. Transportation Research Procedia, 24, 474-481. https://doi.org/10.1016/j.trpro.2017.05.085
6. Vega, I. P. (2020, junio 3). Más de 100 km de ciclovías en Guadalajara, motivo para "celebrar" en el Día Mundial de la Bicicleta. UDG TV. Retrieved from https://udgtv.com/noticias/mas-100-km-ciclovias-guadalajara-celebrar-dia-mundial-bicicleta/
7. Hoozie—Users. (s. f.). Hoozie. Retrieved from https://www.hoozie.io/?lang=en
8. Datos Bici Blanca. (s. f.). Datos abiertos. Retrieved from https://docs.google.com/spreadsheets/d/1fXJGPruV8Flzvb_fKoQ39YmqASXXfva8hfgyI80arR0/edit?usp=embed_facebook
9. Carapia, F. (s. f.). Matan a más ciclistas pese a las ciclovías. Retrieved from https://www.reforma.com/matan-a-mas-ciclistas-pese-a-las-ciclovias/ar1861611
10. Prieto, K., Chávez–Hernández, M. V., & Romero–Leiton, J. P. (2022). On mobility trends analysis of COVID–19 dissemination in Mexico City. PLOS ONE, 17(2), e0263367. https://doi.org/10.1371/journal.pone.0263367

11. Asgarzadeh, M., Verma, S., Mekary, R. A., Courtney, T. K., & Christiani, D. C. (2017). The role of intersection and street design on severity of bicycle-motor vehicle crashes. Injury Prevention, 23(3), 179-185. https://doi.org/10.1136/injuryprev-2016-042045
12. Madsen, T. K. O., & Lahrmann, H. (2017). Comparison of five bicycle facility designs in signalized intersections using traffic conflict studies. Transportation Research Part F: Traffic Psychology and Behaviour, 46, 438-450. https://doi.org/10.1016/j.trf.2016.05.008
13. Richter, T., & Sachs, J. (2017). Turning accidents between cars and trucks and cyclists driving straight ahead. Transportation Research Procedia, 25, 1946-1954. https://doi.org/10.1016/j.trpro.2017.05.219
14. Seiniger, P., Gail, J., & Schreck, B. (2015). Development of a Test Procedure for Driver Assist Systems Addressing Accidents Between Right Turning Trucks and Straight Driving Cyclists. https://doi.org/10.13140/RG.2.2.36586.72649
15. Pokorny, P., Drescher, J., Pitera, K., & Jonsson, T. (2017). Accidents between freight vehicles and bicycles, with a focus on urban areas. Transportation Research Procedia, 25, 999-1007. https://doi.org/10.1016/j.trpro.2017.05.474
16. Lin, Z., & Fan, W. (David). (2021). Cyclist injury severity analysis with mixed-logit models at intersections and nonintersection locations. Journal of Transportation Safety & Security, 13(2), 223-245. https://doi.org/10.1080/19439962.2019.1628140
17. Kummeneje, A.-M., & Rundmo, T. (2020). Attitudes, risk perception and risk-taking behaviour among regular cyclists in Norway. Transportation Research Part F: Traffic Psychology and Behaviour, 69, 135-150. https://doi.org/10.1016/j.trf.2020.01.007
18. Harris, M. A., Reynolds, C. C. O., Winters, M., Cripton, P. A., Shen, H., Chipman, M. L., Cusimano, M. D., Babul, S., Brubacher, J. R., Friedman, S. M., Hunte, G., Monro, M., Vernich, L., & Teschke, K. (2013). Comparing the effects of infrastructure on bicycling injury at intersections and non-intersections using a case–crossover design. Injury Prevention, 19(5), 303-310. https://doi.org/10.1136/injuryprev-2012-040561
19. Tan, T., Haque, S., Lee, -Archer Lachlan, Mason, T., Parthiban, J., & Beer, T. (s. f.). Bicycle-friendly roundabouts: A case-study. Journal of the Australasian College of Road Safety, 30(4), 67-70. https://doi.org/10.3316/informit.032179159989363
20. Aldred, R., Kapousizis, G., & Goodman, A. (2021). Association of Infrastructure and Route Environment Factors with Cycling Injury Risk at Intersection and Non-Intersection Locations: A Case-Crossover Study of Britain. International Journal of Environmental Research and Public Health, 18(6), 3060. https://doi.org/10.3390/ijerph18063060
21. Hollenstein, D., Hess, M., Jordan, D., & Bleisch, S. (2019). Investigating Roundabout Properties and Bicycle Accident Occurrence at Swiss Roundabouts: A Logistic Regression Approach. ISPRS International Journal of Geo-Information, 8(2), 95. https://doi.org/10.3390/ijgi8020095
22. Johnsson, C., Laureshyn, A., Dágostino, C., & De Ceunynck, T. (2021). The 'safety in density' effect for cyclists and motor vehicles in Scandinavia: An observational study. IATSS Research, 45(2), 169-175. https://doi.org/10.1016/j.iatssr.2020.08.003
23. Prati, G., Angelis, M. D., Puchades, V. M., Fraboni, F., & Pietrantoni, L. (2017). Characteristics of cyclist crashes in Italy using latent class analysis and association rule mining. PLOS ONE, 12(2), e0171484. https://doi.org/10.1371/journal.pone.0171484
24. IIEG. (s. f.). MIBICI.knit. Uso del medio de transporte MiBici de diciembre del 2020 a diciembre del 2021. Retrieved from https://iieg.gob.mx/ns/wp-content/uploads/2022/02/MIBICI.html
25. Demsˇar, J., Curk, T., Erjavec, A., Demsar, J., Curk, T., Erjave, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M., & Zupan, B. (s. f.). Orange: Data Mining Toolbox in Python. 5.
26. Kriegel, H.-P., Schubert, E., & Zimek, A. (2017). The (black) art of runtime evaluation: Are we comparing algorithms or implementations? Knowledge and Information Systems, 52(2), 341-378. https://doi.org/10.1007/s10115-016-1004-2
27. IIEG. (s. f.). Vehículos de motor registrado en circulación – IIEG. Retrieved from https://iieg.gob.mx/ns/?page_id=2025
28. Distefano, N., & Leonardi, S. (2018). A list of accident scenarios for three legs skewed intersections. IATSS Research, 42(3), 97-104. https://doi.org/10.1016/j.iatssr.2017.07.003
29. Pan, D., Han, Y., Jin, Q., Wu, H., & Huang, H. (2021). Study of typical electric two-wheelers pre-crash scenarios using K-medoids clustering methodology based on video recordings in China. Accident Analysis & Prevention, 160, 106320. https://doi.org/10.1016/j.aap.2021.106320
30. Aldred, R., Goodman, A., Gulliver, J., & Woodcock, J. (2018). Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accident Analysis & Prevention, 117, 75-84. https://doi.org/10.1016/j.aap.2018.03.003