CLUSTERISATION AND ACCIDENT ANALYSIS OF THE TRANSMILENIO BRT SYSTEM, IN BOGOTÁ 2005-2010

Juan Luna Monroy[1], Juan Miguel Velásquez[2]

[1,2]Universidad de los Andes, Bogotá (Colombia), Carrera 1a Este # 19A-40

Email for correspondence: jc.luna135@uniandes.edu.co

ABSTRACT

1. Objective

Within the context of strong public transport demand growth, and the current need for expansion of the TransMilenio system in Bogotá, Colombia, it is important to analyse how accidents within the network have evolved.
Accidents are complex phenomena with origins in simultaneity of causes; therefore cluster analysis is useful to indentify these heterogeneous patterns of occurrence in order to make recommendations by recognising common characteristics of different types of accidents.

2. Data/Methodology

Accident information for years 2005 to 2010 was provided by TransMilenio S.A., the city-owned company that is in charge of supervising the system's operation.
The K-means algorithm implemented in the clustering process groups the accidents into clusters according to the similarity in their characteristics (two groups were created: Time-spatial and General circumstantial), therefore the method maximises the homogeneity of accidents within the clusters while simultaneously maximising the heterogeneity between the aggregates. The algorithm was developed using a Visual Basic model provided by Universidad de los Andes.

3. Results/Findings

The study was divided into three biannual periods, indicating that collision with pedestrians is the leading cause of deaths and injuries from traffic accidents for the two types of buses of the system. That means that the main victims of the events are non-users with accidents taking place mainly in two of the nine trunk roads of the network.

The evolution of accidents does not follow a strict trend of increase or decrease, especially for fatal accidents, which have remained almost constant around the years at a level of 8 fatalities per year.

This study demonstrates the usefulness of clustering techniques even when the interaction between characteristics shows there is no unique combination associated to the occurrence of the accidents.

4. Implications for Research/Policy

The nature of the results provides further thought into the interaction of different factors within the clusters in order to understand the relationships that remain hidden in the accident database. The study also aims for the necessity to strike a proper balance between the amount of detail on field records in order to optimise and make information as reliable as possible. This is especially relevant now since in the short term the BRT will be used by and increasing number of citizens because of the new public transportation unification plan that will start operations in Bogotá in 2012.

Keywords: Accident evolution, BRT system, Cluster, K-means algorithm

INTRODUCTION

Bogota's administration implemented TransMilenio in 2001, a bus rapid transit system (or BRT) that runs on a segregated bus-corridor. It was built looking for a solution to the transport problems that had been present in the city along its evolution, with a conception based on the experience from the Brazilian city of Curitiba with its "Rede Integrada de Transporte", but adding modifications which significantly increased its operational efficiency and capacity.

TransMilenio represents the largest investment in urban mobility in the country in the last decade and nowadays, ex-post evaluation of the system reveals how in the last decade the city has made significant progress in road safety, achieving a significant reduction in accident rates, along clear improvements in travel times, emissions, passengers per hour per direction (PPHPD), among others.

By 2010 the system carried 29% of all journeys made by public transport in the city, with a network conformed by 84 km of trunk roads spread over 9 different corridors: Caracas, Autonorte, Suba, Calle 80, NQS Central, Americas, NQS Sur, Caracas Sur and Eje Ambiental, additionally has 663 km of feeder routes spread over 83 lines.

Figure 1 – Trunk lines and feeders distribution within the system

Source: TransMilenio S.A. (2011)

It is important to analyse the accident evolution within the TransMilenio network, taking into account that the system presents annual growths of 9.5% in the number of passengers (Cámara de Comercio de Bogotá & Uniandes, 2010).

Generally TransMilenio accidents are defined by the own company as events which cause deaths, injuries or damages above five times the statutory minimum wage (1.200 USD in 2010), which in 2010 accounted for 1.3% over the total proportion of events in Bogotá. A fact that indicates how the security requirements on the buses, feeders, and the control over the segregated pathways results in favourable conditions for road safety.

This paper aims to use the clustering method by implementing the K-means algorithm, working databases obtained from TransMilenio entity, seeking to understand the accidents for the years 2005-2010, by assessing three biannual periods, 2005-2006, 2007-2008 and 2009-2010.

General Objective

The aim of the study is to identify factors which contribute to the accident cases, looking forward to point out more recurring patterns, and thus, facilitate the generation of prevention policies.

Specific objectives

- Demonstrate the utility of clustering methods, specifically of the k-means algorithm, for purposes of organization, and further analysis of the variables simultaneous interaction within the cause of accidents.

- Recognition of the characteristics of the types of accidents on which greater efforts should be concentrated.

- Generation of useful recommendations over fatality reduction within the system, and considerations for future TransMilenio expansions and new BRT systems.

INFORMATION MANAGMENT

Driving conditions vary depending on the place or mode of transport (Fotoughi & M. Montazeri, 2011), and so, for our particular case, the conditions contributing to the accidentality of the TransMilenio system can be ordered by the Haddon matrix criteria, one of the most used and useful in the classification of information within the field of wounds prevention (Haddon, 1981):

Variables specific to the severity of the accident: Number of deaths and injuries, it is important to clarify how the severity of the accident is determined according to the level of the occupant in the worse condition (Milton, 2006).
Geographical location: Features the corridor where the accident occurred (trunk line, city area of occurrence)
Circumstances Information: Event details (type of accident, time of the accident, month, vehicles involved)
Parties Involved: User or non-user

With the data from accident database, features and characteristics of the events were classified as follows:

Table 1 – Considered time-spatial characteristics

	Variable	Option	Convention
TIME-SPATIAL CHARACTERISTICS	***Time***	Morning rush hour (05-09)	1
		Morning (10-12)	2
		Afternoon (13-16)	3
		Evening rush hour (17-19)	4
		Evening - Night (20-21)	5
		Night (22-04)	6
	Weekend	Yes	1
		No	2
	Trunk road	Caracas (Calle 76 -Tercer Milenio)	1
		Autonorte (Portal norte -Héroes)	2
		Suba (Portal Suba - San Martín)	3
		Calle 80 (Portal 80 -Polo)	4
		NQS Central (NQS calle 75 -Ricaurte)	5
		Américas (De la Sabana - Portal Américas)	6
		NQS Sur (Comuneros - Portal Sur)	7
		Caracas Sur (Hospital - Portal Tunal)	8
		Eje Ambiental (Museo del oro - Las aguas)	9
		NA	10

Source: Authors elaboration

Table 2 – General Circumstantial characteristics

	Variable	Option	Convention
GENERAL CIRCUMSTANTIAL CHARACTERISTICS	***Type of Vehicle***	Bus	1
		Feeder	2
	Type of accident	Type 1 (Death)	1
		Type 2 (Injuries)	2
		Simple (Economic damages only)	3
	Condition	Injured	1
		Fatality	2
		Injured + Fatality	3
		No consequences	4
	Affected characteristics	User	1
		Non user	2
		User + Non user	3

		No consequences	4
	Type of collision	TransMilenio vehicle (bus or feeder)	1
		Private vehicle	2
		Infrastructure	3
		NA	4
		TransMilenio vehicle + Private vehicle	5
	Hit	Pedestrian	1
		Cyclist	2
		Motorcyclist	3
		NA	4
	Other events	Touch	1
		Braking hits[1]	2
		Approximations[2]	3
		NA	4

Source: Authors elaboration

CLUSTRING PROCESS

Clusterisation or cluster analysis, is the name of a group of multivariate techniques which groups individuals or objects into clusters such that objects in the same cluster are more similar to each other than to other objects on different clusters. These methods intention is to maximise the homogeneity of objects within the clusters while simultaneously maximising the heterogeneity between aggregates, so that if the classification is correct, the objects within groups are very next when rendered graphically, and different groups are far apart (Hair, 2008).

The K-MEANS algorithm

K-means is a hierarchical clustering algorithm and it is one of the most widely used algorithms for clustering. The hierarchical clustering iteratively builds clusters by agglomeration or by division from the previous iteration, so it can be understood as the disintegration of the elements incorporated in a single group division generating groups as a result.

This algorithm requires three user-specified parameters: number of clusters K, a matrix mxn and the number of rows and columns the matrix has. It consists of the following steps:

[1] Accidents produced by the braking operation of the buses
[2] Accidents produced on arrival at the bus stations

Step 1. The algorithm starts by reading the data. Then each variable is compared with the minimum possible value, divided by the maximum value minus this minimum value, all this in order to normalize the data:

Step 2. Given n objects, the method first select k objects as initial k clusters. It then iteratively assigns each object to the most similar cluster based on the mean value of the objects in each cluster. Based on this, the algorithm calculates a first objective function as a reference for the next iteration.

Step 3. Compute the clusters centres and generates a new partition by assigning the closest accident to each one of the cluster centres.

Step 4. The algorithm calculates a second objective function. This function is the sum of the absolute distance of each of the accidents compared to the centroid in each cluster. The value of this function is a measure of distance that is calculated for each variable and for all accidents. What is sought is that the distance is the minimum possible to ensure that accidents that make up each cluster have a minimum separation between them.

Step 5. Assigns each accident to the most similar cluster.

ALGORITH APPLICATION & BIANNUAL RESULTS

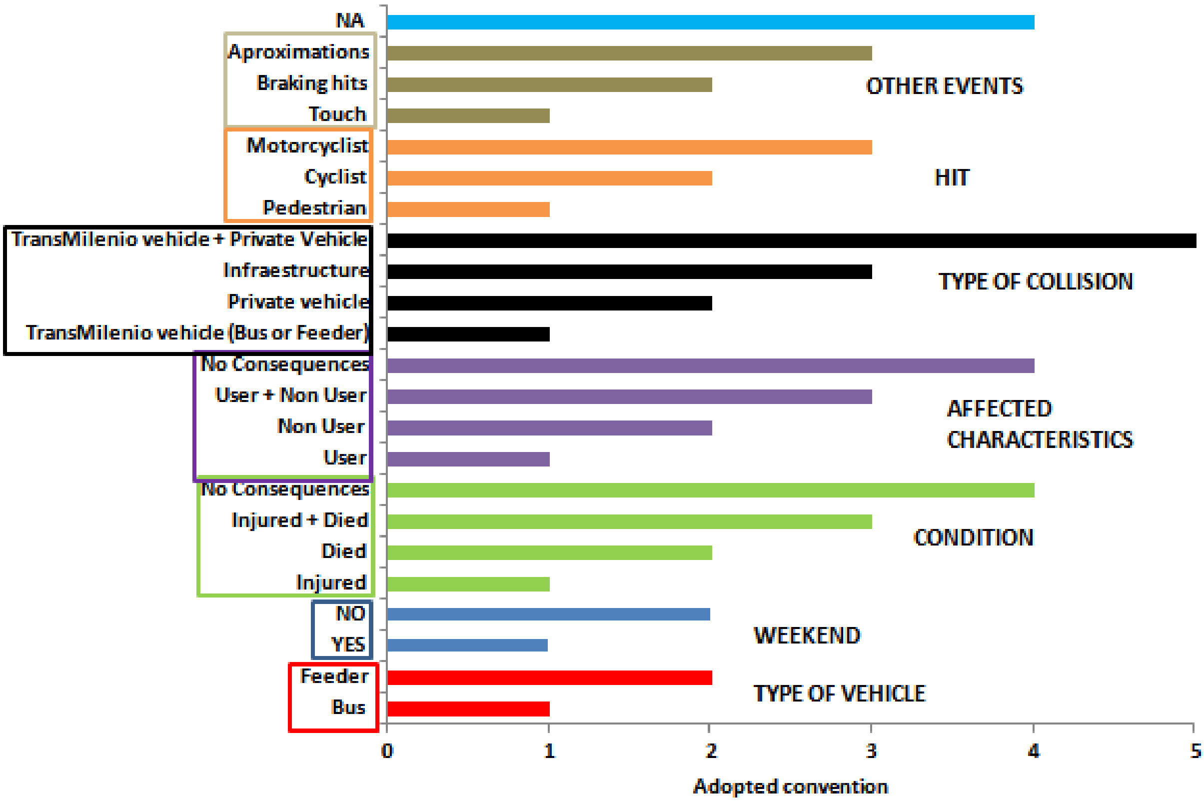
For purposes of a detailed reading of the databases, the study was divided into pairs of years. Thus the clustering algorithm was performed biannually for the years covered between 2005-2010.

For the generation of the clusters, the input data were:

- Database accident with conventions for parameter
- Size of the matrix
- Number of clusters to generate (20 for all periods)

Up next the conventions of the parameters included in the clustering process:

Figure 2 – Adopted conventions for the cluster analysis



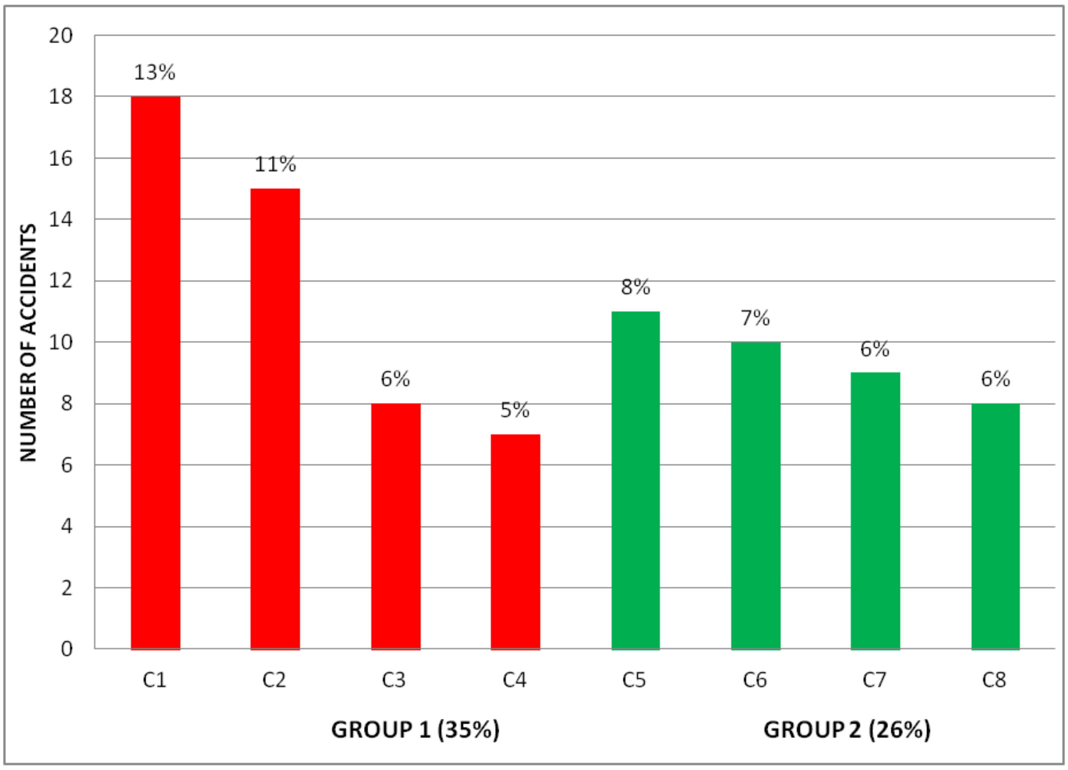
Source: Authors elaboration

First period: 2005-2006

To study the effects of recurrent parameters, from the 20 clusters, 8 were selected for this period due to being representative. The selected groups represent 62% of the 141 Type 2 accidents studied.

From the observed characteristics, 2 study groups were formed: one for accidents involving TransMilenio buses and another for the feeder vehicles:

Figure 3 – Formed groups from selected clusters, Period 2005-2006



Source: Authors elaboration

Both groups share a common detail: the victim, which for the selected clusters, always were non users.

Below are the final parameters for the selected clusters:

Group 1 – Buses

Table 3 – Clusters characterization for the buses group, Period 2005-2006

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
1	13	1	2	2	4	1	4
2	11	1	2	2	2	4	4
3	6	1	1	2	2	4	4
4	5	1	1	2	4	1	4

Source: Authors elaboration

The first group presents interesting interactions between the causality of the event and the day of the event. Thus, cluster 1 indicates that 13% of accidents occurred on weekdays with a common cause: pedestrian hit. It is interesting to see the change in the accident rate for events of the same cause, but this time occurred on weekends: cluster 5, group that unites these characteristics, holds a much smaller percentage with only 5% of the total.

Group 2 - Feeders

Table 4 – Clusters characterization for the feeders group, Period 2005-2006

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
5	8	2	2	2	2	4	4
6	7	2	1	2	4	1	4
7	6	2	1	2	2	4	4
8	6	2	2	2	4	1	4

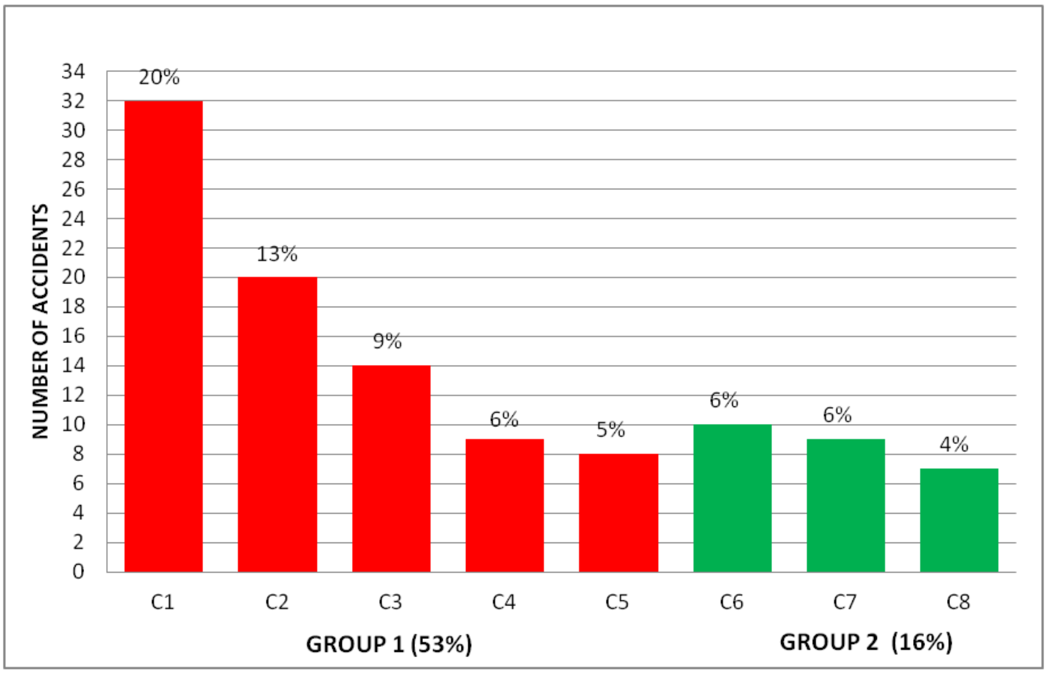
Source: Authors elaboration

Interactions within Group 2 show a similar trend. However, this time we see a majority of accidents resulting from collisions with a private vehicle and secondly accidents caused by hitting pedestrians. Analysing the impact of the day on which the event occurs, it is easy to see how the accident rate for this period was concentrated further within the weekdays.

Second period: 2007-2008

To study the effects of recurrent parameters from the 20 clusters, 8 were selected for this period due to being representative. The selected groups represent 69% of the 160 Type 2 accidents studied. Once again from the observed characteristics two groups were formed: one for accidents involving TransMilenio buses participation and another for the feeder vehicles:

Figure 4 – Formed groups from selected clusters, Period 2007-2008



Source: Authors elaboration

Figure 4 shows the percentage distribution of the total of accidents studied and the way they are distributed within each cluster. In comparison with the first period we appreciate groups with a higher rate of the total representation, which is explained by the greater sample size for this second period of analysis.

Both groups share a common characteristic: It always involved non-users.

Below, the final parameters for the selected clusters:

Group 1 - Buses

Table 5 – Clusters characterization for the buses group, Period 2007-2008

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
1	20	1	2	2	4	1	4
2	13	1	1	2	4	1	4
3	9	1	1	2	2	4	4
4	6	1	2	2	2	4	4
5	5	1	2	2	4	2	4

Source: Authors elaboration

Clusters belonging to this majority group, indicates the strong participation of the accidents caused by pedestrians hitting. Events framed under this fact represent the formation of the most representative cluster with 20% of the accidents occurring on weekdays, and 13% for those occurring on weekends.
In a significantly lesser extent, the characteristics of the accident indicate that those events generated by the collision with private vehicles, represented by clusters 3 and 4, contributes with 15% of the total, mostly happening during the weekends. This apparently contradicts what happened in the first period, where for this type of cause, the largest share came from the events that occurred on weekdays.

Finally, although it is not a group of mass participation for the period, it is important to see that accidents caused by cyclists hitting formed a cluster containing 5% of the reported events, which is a non-negligible percentage if we take into account the growth of users who choose to use this mode of transport. This implies the consequent need to develop strategies for prevention and education for users who prefer this mode.

Group 2 – Feeders

Table 6 – Clusters characterization for the feeders group, Period 2007-2008

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
6	6	2	2	2	4	1	4
7	6	2	1	2	4	1	4
8	4	2	2	2	4	2	4

Source: Authors elaboration

The second group shows that clusters of greater participation were originated by accidents caused by pedestrians hitting. Thus, once again the dynamic of events arising from this cause is repeated, and this time accounts for 12% of the accidents in the period, which shows that the percentage of occurrence is partly evenly between weekdays and weekend, each with 6% of the total.

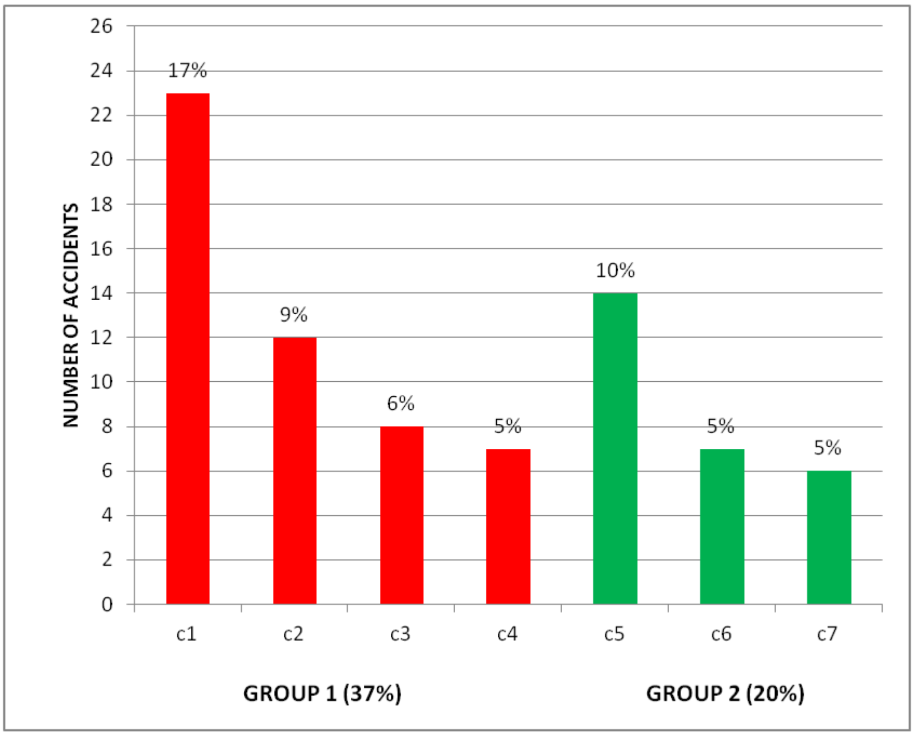
A final cluster in a lesser extent, with 4% of the total, evidence that the collision with private vehicles on weekdays is the second most recurring accidents between feeders of the TransMilenio system.

Third period: 2009-2010

Once again, to study the effects of recurrent parameters from the 20 clusters, 7 were selected for this period due to being representative. The selected groups represent 57% of the 135 Type 2 accidents studied.

From the observed characteristics was continued the formation of two study groups: one for accidents involving TransMilenio buses participation and another for the feeder vehicles:

Figure 5 – Formed groups from selected clusters, Period 2009-2010



Source: Authors elaboration

The presented trend evidences a decline in the share of each type of vehicle compared with the second period. The percentages obtained indicate a stabilisation of the results after a considerable increase in the second period. Now the results for this third period are similar to those obtained in the first.

The common feature between clusters of greater participation (like it was observed for the first two periods) implies that non-users are involved in most accidents

Below are the final parameters for the selected clusters:

Group 1 - Buses

Table 7 –Clusters characterization for the buses group, Period 2009-2010

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
1	17	1	2	2	4	1	4
2	9	1	1	2	4	1	4
3	6	1	2	2	4	3	4
4	5	1	1	2	2	4	4

Source: Authors elaboration

The evidenced dynamics from the clusters generated for accidents caused by TransMilenio buses demonstrate that events most frequently were originated by pedestrians hitting (first two clusters), where the vast majority of these events (17%) occurred on weekdays.

Now, it is important to see how the accidents involving motorcyclists are relevant in the study of this period, which shows a growing problem within the system with 6% of participation over the total of counted events.

Finally, the group formed from the collision accidents with a private vehicle (cluster 4), in general shows a decrease compared to the percentages evidenced for the previous periods within the buses group (P1: 17%, P2: 15%, P3: 5%)

Group 2 – Feeders

Table 8 –Clusters characterization for the feeders group, Period 2009-2010

Cluster	% from total of accidents	*Type of Vehicle*	Weekend	*Affected characteristics*	*Collision*	Hit	Other
5	10	2	1	2	2	4	4
6	5	2	1	2	4	1	4
7	5	2	2	2	2	4	4

Source: Authors elaboration

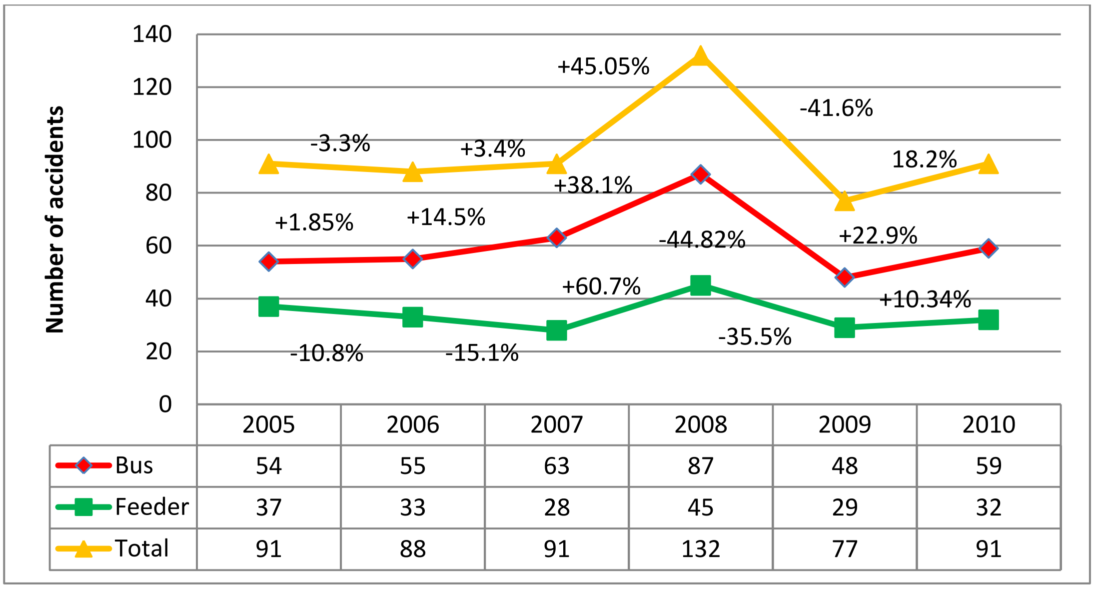
The dynamics evidenced for feeders shows how the accidents caused by collision with private vehicles not only ranks as the dominant cause in this period, but that in turn, represents the largest share of the indicator through the study of the accident feeders for all periods.

It also highlights the greater overall participation rate of accidents on the weekend, where the analysis of this special time factor shows how accidents no longer occur mostly on weekdays.

OVERALL RESULTS

Up next is shown the general trend in the number of accidents within the TransMilenio system. It is important to clarify that accidents only covers those events with serious injuries and human consequences, not those events without direct effects involved.

Figure 6 – Evolution of the total number of accidents within the system



	2005	2006	2007	2008	2009	2010
Bus	54	55	63	87	48	59
Feeder	37	33	28	45	29	32
Total	91	88	91	132	77	91

Source: Authors elaboration with the database provided by TransMilenio S.A.

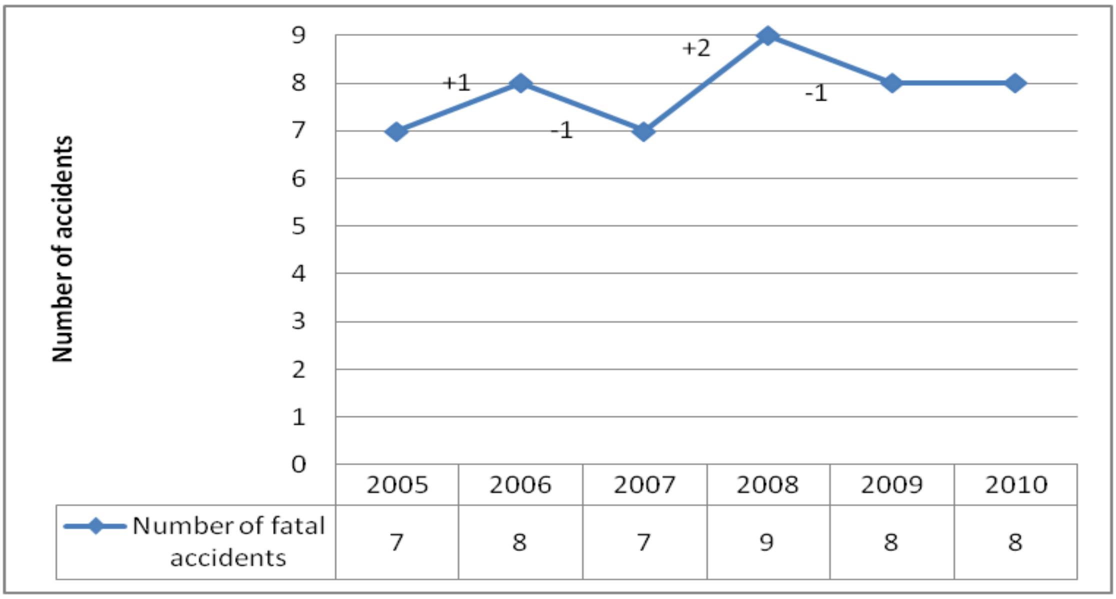
The first three years of study shows a stable trend in the number of accidents, which is explained by the behaviour of feeders and buses. We see how from 2005 to 2007 while the buses accidents increased, feeders accidents decreased at a greater rate. This is an effect that in general compensates these rates of accidents on the early years.

At this point it is critical to note the significant increase in the number of accidents for both modes between 2007 and 2008, with increases of 38.1% for buses and 60.7% for feeder buses, this period was the one with the largest changes, which makes 2008 the year with the most accidents in the system. That year saw a sharp increase in the number of passengers among feeder buses (20%), which could explain their increased contribution in the accidents for that year.

Then there is a significant decrease, which affected the trend that had been presented until 2007, showing significant reduction percentages for both type of bus, this brought the overall decline in accidents by 2009, with a 41.6% less events respect to 2008. Finally, Figure 6 shows a significant increase in the accident rate for 2010, constituting itself in the second highest peak in the system, only exceeded by the 2008 peak.

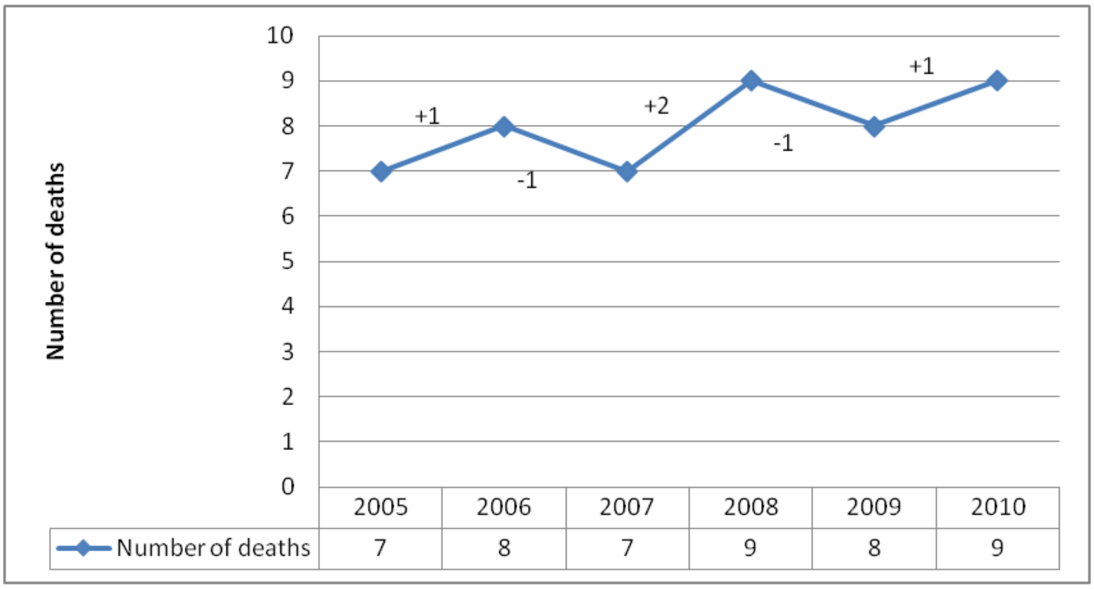
TYPE 1 ACCIDENT ANALYSIS

Figure 7 – Evolution of the total number of fatal accidents within the system



Source: Authors elaboration with the database provided by TransMilenio SA

Figure 8 – Evolution of the total number of deaths within the system



Source: Authors elaboration with the database provided by TransMilenio SA

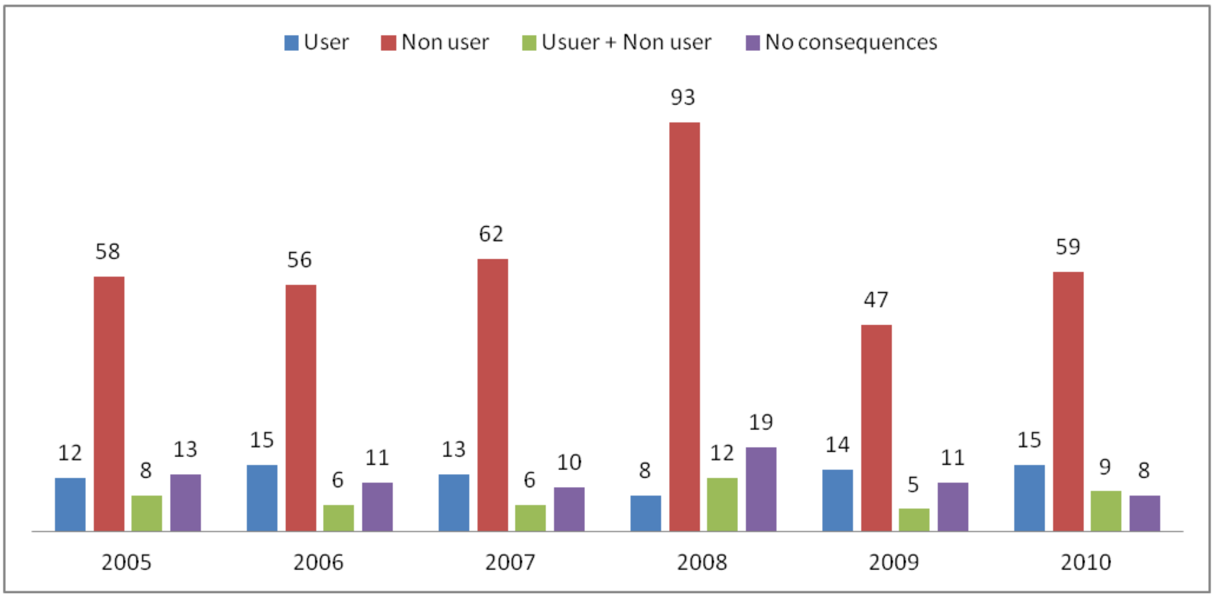
.

From the type 1 analysis figures, it is worrying to see how the trend, if is not increasing markedly over the years, has remained almost constant in number of accidents as death toll. These results arise in the need for greater control and implementation of measures to reduce fatal accidents, mainly during evening rush hour (17-19) of weekdays and generally in the morning hours of the weekend days (05-12), which is when most fatal crashes occurred.

The analysis of fatal accidents in the system showed that in all periods most accidents were produced by pedestrian hitting, which is a fact that validates it again as one of the leading causes of general system events, followed by cyclist and motorcyclists hitting.

EVOLUTION OF ACCIDENTS BY VICTIMS

Figure 9 – Evolution of the total number of Type 2 accidents by affected victims



Source: Authors elaboration with the database provided by TransMilenio SA

The analysis for the affected characteristics shows how major accidents involve individuals outside the TransMilenio system service. This relates directly with the information we extract from the collisions and hitting events around the study, in which it is clear how private vehicles and pedestrians are the ones primarily involved in the crashes, which means the non-users are the most affected ones. This leads us to believe that despite the high frequency of service and constant interaction, BRT vehicles rarely crash with each other.

Finally we see how accidents with economic consequences only (Type 3), are minor compared to those involving any human consequence. So measures should be taken to make this kind of accident the only one allowable within the system.

DISCUSSION & CONCLUSIONS

Initially this study shows a system in which the improvements which occur in some of the years are isolated events, and do not correspond to strict regulations tracking a marked and constant decrease. Because of this trend, the accident rate of the system, along with its various causes and effects, does not respond to a direct increase or decrease form, instead, their tendency responds to an oscillatory behaviour for the study years. Despite this behaviour, it is important to note how even for a system that reports annual growth of 9.5% passengers annually (Cámara de Comercio de Bogotá & Uniandes, 2010), there is no corresponding increase in accident numbers.

The study of databases allowed to fully implementing the clustering K-means algorithm, which was used to study the type of accident more representative within the sample, corresponding to events with injuries (Type 2). The development of the algorithm was used to classify these accidents into specific groups, which were organized from the dynamics of homogenisation and simultaneous occurrence in the time of the event.
This study demonstrates the usefulness of clustering techniques even when the interaction between characteristics shows there is no unique combination associated to the occurrence of the accidents.

At this point it is important to frame the limitations introduced in the implementation of the method, in relation to the use of the information in the database. Thus the modelling showed that the low number of accidents within the system leads to a problem of creating enough representative clusters. This created the difficulty to evaluate information as the occurrence trunk line, month, and time of occurrence range, and so it was not possible to identify a dominant trend.

The results show that non-users are those mainly affected by accidents, especially pedestrians. There is also a worrisome trend of increasing collisions with private vehicles, especially in feeder buses where services run with the mixed traffic. This implies that investment is needed in the infrastructure adjacent to the system, like pedestrian crossings and sidewalks. It is also important to focus traffic law enforcement on junctions to avoid private vehicles to invade the busways.

It is clear now that the accidents studied, because of its origin, must concentrate more efforts on mitigating the pedestrian accidents. This idea is based not only on the participation of this cause in the accident with injuries, but also because of its impact over the evolution of fatal accidents as well, where at length it was the fundamental cause.

Finally it is important to note that this study focuses on the current cause of accidents within the system. It is not a study on how accident rates changed in the city before and after the implementation of BRT lines. It therefore has a limited scope, and does not intend to underscore the important contribution of BRTs to the road safety policies of Colombian and Latin American cities.

BIBLIOGRAPHY

Alcaldia Mayor de Bogotá. (2010). *Movilidad en cifras 2010.* Bogotá.

Brenac. (2005). The indirect involvement of buses in traffic accident processes,.

Cámara de Comercio de Bogotá & Uniandes. (2010). *Observatorio de movilidad y la región. Reporte anual de movilidad 2010.*

Depaire & Vanhoof. (2008). Traffic accident segmentation by means of latent class clustering. En *Accident Analysis & Prevention* (págs. 1257-1266). Dieoenbeek.

Dirección de Planeación Nacional. (2010). *Plan Nacional de Desarrollo 2010-2014. .*

Elvik. (2009). Why some road safety problems are more difficult to solve than others. Gaustadallén.

Fondo de Prevención Vial & Uniandes. (2010). *Desarrollo de Metodología de Valoración del Costo Económico de la Accidentalidad Vial en Colombia y Cálculo para el periodo 2008-2010.*

Fondo de Prevención Vial. (2010). Anuario estadístico de accidentalidad nacional. Colombia.

Fotoughi & M. Montazeri. (2011). *Traffic condition recognition using the K-means clustering method.*

Fraser. (2005). *Traffic accidents scar Latin America's roads.*

Guojun. (1979). Data clustering : theory, algorithms, and applicationa . Journal compilation.

Haddon. (1981). *Injury control, Preventive and community medicine.*

Hair. (2008). Multivariate Data Analysis. Lousiana: Pearson, Prentice Hall.

Hartigan, J. (1975). The k-means algorithm. En *Clustering Algorithms* (págs. 84-107). Connecticut: Yale university.

Instituto Nacional de Medicina Legal y Ciencias Forenses. (2010). *forensi 2010.*

Kopelias. (2007). *Urban freeway crash analysis.*

Martensen, D. &. (2008). Risk and protection factors in fatal accidents .

Milton. (2006). *Highway accident severities and the mixed logit model: An exploratory empirical analysis.* Washington D.C: Pennsylvania State University.

PIARC. (2004). Road safety manual : recommendations from the World Road Association . Switzerland: Route 2 Market.

Turner & Kooshian. (2012). *Case study: Colombia's Bus Rapid Transit (BRT) Development and Expansion.*