XII Conference on Transport Engineering, CIT 2016, 7-9 June 2016, Valencia, Spain
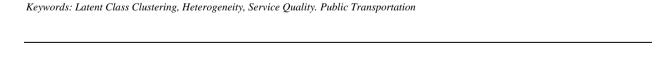
The effect of service attributes' hierarchy on passengers' segmentation. A light rail transit service case study

Francisco Diez-Mesa*, Rocío de Oña, Juan de Oña

*TRYSE Research Group, Department of Civil Engineering, University of Granada,
ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)*

Abstract

Market segmentation can help transit operators to identify groups of passengers that share particular characteristics and specific needs and requirements about the service. Traditionally, socioeconomic variables have been used to perform a simple segmentation, although satisfaction rates about service attributes were not similar among individuals belonging to a group. Cluster analysis emerges as a novel analytical technique for extracting passengers' profiles. This paper investigates passengers' profiles at the metropolitan Light Rail Transit service of Seville (Spain). Latent Class Clustering algorithm is applied and satisfaction rates about different service quality attributes are considered for the segmentation. Particularly, two different cluster analyses are accomplished: first level, with only socioeconomic attributes; and second level, with eight service quality factors and socioeconomic attributes. The service quality factors are obtained through a principal component analysis, at which, the large number of attributes describing the service is reduced into constructs underlying them. Equivalent satisfaction rates are calculated for these service factors. Then, homogeneous groups of passengers are obtained. Additionally, the main differences among cluster are identified.

Keywords: Latent Class Clustering, Heterogeneity, Service Quality. Public Transportation

1. Introduction

One of the central aims of studying Service Quality (SQ) in Public Transport (PT) is increasing the attractiveness of PT and improving its use like a sustainable alternative to private vehicle. The SQ in PT could be measured from two points of view: service operators and passengers. Service operators measure SQ in base on efficiency and effectiveness of transit services concept. Passengers evaluate SQ based on their perceptions and expectations, or by a range of simple disaggregate performance measures that can be used for measuring the ability of the service operator to offer the services that meet passenger expectations. The main methodology to collect passengers' opinions is Customer Satisfaction Surveys (CSS) (de Oña and de Oña, 2015). However, these CSS enjoy major heterogeneity, which leads passengers that present different personal characteristics and context. One solution to reduce this heterogeneity can be to stratify the sample of users on segments of passengers more homogeneous. A wide range of proposals can be found in the literature: traditionally, segmentation is carried out in terms of the population´s socioeconomic and demographic characteristics (i.e., income, gender, etc.) or travel habits profiles (i.e., type of day, frequency of use, etc.) (de Oña et al. 2015). A novel way to reduce the heterogeneity is to use Data Mining techniques such as Cluster Analysis (CA) (de Oña et al. 2014, *in press*; Wen et al. 2008).

In this research, CA is used to reduce this heterogeneity by segmenting the sample of passengers on groups that share common characteristics and have more homogeneous perceptions about the service. It is habitual to use socioeconomic attributes to perform CA (Erdman et al., 2015). Nevertheless, a CA with different typologies of attributes, such as socioeconomic attributes and SQ attributes (i.e., Accessibility, Customer Service, Information, etc.), has not often been used. The main purpose of this study will be to perform two CA for stratifying the sample of users of Metro of Seville (Spain), using, on the one hand, only one typology of variables (socioeconomic) and, on the other hand, two types of attributes (socioeconomic and SQ attributes), in order to compare the different profiles identified by each CA. Data from a CSS collected in 2014 in the Metro of Seville is used. The differences identified among the characteristics of each profiles of each CA are displayed and discussed.

The paper is structure as follows: Section 2 shows the methodology used for stratifying the sample. Section 3 describes the data used for the analysis. The results obtained with the CA are explained and compared in Section 4. Finally, the conclusions are reported in Section 5.

2. Techniques and procedures

2.1. Cluster Analysis

CA is a statistical technique which segments a sample (cases, data or objects such as events, people or things) in different groups (clusters) whose items are heterogeneous among items of others clusters, but homogeneous among items of the same cluster. Latent Class Clustering (LCC) methodology is a technique which has significant advantages over the others (Hair 2010; de Oña et al. 2013), such as: it is possible to consider different attributes without the need for a priori standardization; it allows classifying probabilities by the use of the membership probabilities of each item; it uses measurements which are not based on the distance between data; it does not demand a large space in the memory of computer and the models can incorporate independent attributes (covariates), which could be used to describe the latent class rather than defining them. For a detailed explanation of LCC analyses see Magidson and Vermunt (2002), and Vermunt and Magidson (2005).

The objective in this methodology is to find the optimal number of clusters to align the database with the model. The criteria of selection are based on three information criteria, such as the Bayesian information criterion (BIC) (Raftery 1986), the Akaike information criterion (AIC) (Akaike 1987), and the consistent Akaike information criterion (CAIC) (Fraley and Raftery 1998). The information criteria and criteria of representativeness and of characterization are used to evaluate the optimal number of clusters because it is important that the groups identified present remarkable characteristics and are easily characterized. The optimal number of clusters is the one that

minimizes the score of these criteria, making them more parsimonious and better adapted to the study data (de Oña et al. 2014).

2.2. Principal Component Analysis

PCA is a statistical approach that can be used to analyze interrelationships among a large number of attributes and to explain these attributes in terms of their common underlying dimensions (Coffel 1995).

PCA was used as an exploratory method in order to empirically reduce the number of underlying dimensions by grouping different attributes of the service according to the respondents' SQ perceptions. PCA has been previously used in the development of customer satisfaction scales in the transportation field (de Oña et al. 2015) since it allows the researcher to better understand customer's ratings of SQ and empirically analyze the dimensions that are conceptualized. Therefore, in this paper PCA is used to reveal which are the latent constructs underlying the SQ attributes.

3. Data of study

This study involved 3,198 interviews collected by a CSS which was addressed to users of the Line 1 of Metro of Seville (Spain) in June 2014. The CSS is divided into four main parts:

- Part A: Users' attitude towards the metro. In this section, the user rates the different quality aspects they experience in the metro service.
- Part B: Users' perception towards the service characteristics. In this section, the users directly rates different service aspects of the Metro Sevilla, as well as provide a global score for the service. It contains 37 questions related to various aspects of the metro service, such as availability, accessibility, information, timeliness, etc. The perceived level of quality of each of the 37 attributes was asked on an 11-numeric scale from 0 to 10 (0 being of poor quality and 10 being of the highest quality). Respondents also rated their overall perceived level of quality of the metro service according to the same scale.
- Part C: General information about the trip. In this section, the users score their travel.
- Part D: Socioeconomic and demographic characteristics.

The main characteristics of the sample: more females (53.30%) than males (46.70%). Most respondents were 18-25 years (41.70%), and the next largest groups were 26-40 (28.90%) and 41-65 years (25.60%). There is an underrepresentation of the groups younger than 18 and older than 65 (2.80% and 1.00%, respectively). The main reasons for travelling are studies (38.80%) and work (35.50%), with leisure and other reasons together showing a similar percentage (15.30% and 10.30%, respectively). Most of the users travel daily (52.10%). Users generally have a high school diploma (41.90%) or are university graduates (48.50%), while there is also a small group who has only secondary obligatory education (8.40%). Most of the sample has a low household monthly family income (lower than 1,800 Euros). The sample of users is equally distributed between those who have a private vehicle available to make the trip and those who do not have it (54.78% and 45.22%, respectively). The users in the sample are sufficiently satisfied with the overall service (average rate of 7.6).

4. Results

The first step was to create clusters with only socioeconomic attributes and with socioeconomic and SQ attributes. 11 socioeconomic and 5 general information variables about the trip attributes were selected and, in the case of the perceptions about SQ, which were recorded with 37 attributes, were reduced to 8 dimensions by using a PCA and used like SQ attributes (Table 1).

The information criteria (BIC, AIC, CAIC), representatively and simplicity of the structure criteria were used to select the optimal number of clusters (de Oña et al. 2014). Then, it was obtained that the variation of BIC, AIC and CAIC in % was less than 1% with five clusters. However, following the representativeness and simplicity of the structure criteria six clusters were selected. The main reason for this conclusion was that the representativeness of the six clusters was better than that of the five clusters, as more heterogeneous profiles of users were obtained among clusters and, in turn, greater homogeneity within clusters. Moreover, in both cases the selection of six clusters reduces the value of information criteria, and the entropy for model 6 in each case was: 0.89 for CA with socioeconomic attributes and 0.9 for CA with socioeconomic and SQ attributes, which indicates a good separation between clusters (McLachlan 2004).

Table 1 only shows the attributes that present appreciable differences in order to characterize the clusters/profiles.

Table 1. Distribution of Clusters.

Typology of variables used	Socioeconomic						Socioeconomic + SQ attributes					
Cluster and Size	C1S (842)	C2S (584)	C3S (534)	C4S (479)	C5S (394)	C6S (365)	C1SQ (683)	C2SQ (630)	C3SQ (528)	C4SQ (600)	C5SQ (409)	C6SQ (349)
V1. Average overall SQ	7.5	8.3	7.6	7.1	7.4	7.7	8,44	7,18	7,46	7,63	9,13	5,21
(Standard deviation)	(1.6)	(1.1)	(1.5)	(1.5)	(1.5)	(1.4)	(0.77)	(1.04)	(1.00)	(1.08)	(0.81)	(1.84)
V2. Tangible Service equipment	-	-	-	-	-	-	8,73	7,42	8,09	8,17	9,65	6,2
(Standard deviation)	-	-	-	-	-	-	(0.69)	(0.94)	(1.00)	(1.05)	(0.39)	(1.64)
V3. Accessibility	-	-	-	-	-	-	8,54	7,12	7,99	7,86	9,6	5,79
(Standard deviation)	-	-	-	-	-	-	(0.81)	(1.06)	(1.12)	(1.18)	(0.44)	(1.59)
V4. Availability of the Service	-	-	-	-	-	-	8,17	6,76	7,2	7,21	8,91	5,09
(Standard deviation)	-	-	-	-	-	-	(0.83)	(1.06)	(1.16)	(1.20)	(0.91)	(1.56)
V5. Customer Service	-	-	-	-	-	-	8,41	6,87	7,49	7,55	9,58	5,08
(Standard deviation)	-	-	-	-	-	-	(0.98)	(1.38)	(1.41)	(1.45)	(0.53)	(2.06)
V6. Security	-	-	-	-	-	-	8,15	6,72	7	7,25	9,27	4,72
(Standard deviation)	-	-	-	-	-	-	(1.00)	(1.25)	(1.51)	(1.60)	(0.78)	(1.85)
V7. Information	-	-	-	-	-	-	8,17	6,79	7,61	7,54	9,42	5,23
(Standard deviation)	-	-	-	-	-	-	(0.97)	(1.26)	(1.26)	(1.38)	(0.63)	(1.69)
V8. Environmental pollution	-	-	-	-	-	-	7,14	5,95	6,21	6,27	8,03	4,36
(Standard deviation)	-	-	-	-	-	-	(1.77)	(1.63)	(1.89)	(1.87)	(2.29)	(2.22)
V9. Individual Space	-	-	-	-	-	-	7,2	5,29	6,14	6,35	8,54	3,64
(Standard deviation)	-	-	-	-	-	-	(1.60)	(1.79)	(1.94)	(1.94)	(1.50)	(1.96)
V10. I use the metro service as:												
Fare	7,5	14,7	10,7	7,1	13,2	9	9,76	7,25	9,97	9,11	21,69	4,6
Comfort	54,7	62,7	46	37,1	49,2	42,5	65,03	49,64	45,4	45,89	65,5	17,18
Quickness	59,2	84,2	65,8	58,3	66	68,5	78,39	60,83	64,65	68,43	79,55	39,69
Frequency	22,8	28,6	32,1	27	36,3	33,2	29,99	19,72	31,46	35,35	43,76	11,7
Ecologic reasons	21,2	19	14,2	8,2	18,5	14,2	21,47	16,48	11,99	12,11	27,3	9,22
I have not driving license	0,1	0	63,2	1,3	0	32,9	1,17	0,21	0,01	63,81	14,81	16,68
I have not own private car	1,8	3,4	46,5	53,5	2	52,6	5,72	7,87	31,32	51,11	22,87	28,45
It is my unique alternative	3,2	1	32,9	30,2	7,1	14,5	1,99	4,3	18,46	27,5	10,84	26,7
Lack of parking	57,5	38,9	6,2	6,3	59,6	5,2	47,87	48,57	29,57	5	21,13	30,38
Traffic jam	40,3	27,6	8	7,3	48	6,8	34,23	33,34	23,59	5,84	23,28	23,44
I can use my own private car by any reasons	8,1	3,4	0,7	10,1	10,9	2,7	4,92	6,24	9,98	0,54	3,63	12,22
V11. Trip Purpose												
Work	61.9	56.8	2.8	4.8	4.1	62.5	59,66	62,52	2,98	15,95	36,14	24,06
Studies	5.3	4.3	77.7	83.9	88.1	2.7	3,69	5,99	87,97	58,58	37,1	55,23
Leisure	18.8	21.1	15.4	8.6	5.1	17.8	21,66	16,49	7,08	15,47	16,43	13,15
Other	14.0	17.8	4.1	2.7	2.8	17.0	14,99	15	1,96	10,01	10,32	7,56
V12. Total length												
< 25 min	6,9	87,3	33,3	28,5	34	27,1	46,99	29,23	31,79	31,78	46,74	17,37
25 - 40 min	57,7	12,7	36,6	37,5	41,9	40,5	36,5	45,05	39,71	37,58	34,02	39,95
> 40 min	35,5	0	30,1	34	24,1	32,3	16,51	25,72	28,5	30,65	19,24	42,68

V13. Type of ticket											
Day ticket	8,9	10,3	12	9,4	7,1	10,4	8,35	8,91	8,11	10,65	11,29	13,14
Bonometro	39,1	41,1	24,9	22,9	25,1	26,8	40,29	40,81	24,52	25,68	27,57	23,36
Consortium card	52	48,6	63,2	67,7	67,8	62,7	51,36	50,28	67,36	63,67	61,14	63,5
V14. Frequency of use												
>4days/week	43.5	42.3	59.4	59.9	66.0	51.2	41,23	45,76	64,17	56,72	54,11	54,11
3-4days/week	17.2	15.2	17.2	21.1	18.3	20.5	16,83	16,51	21,44	17,27	19,65	15,78
1-2days/week	17.6	17.5	11.6	10.9	6.3	13.2	18,04	15,99	7,04	12,67	13,35	14,16
Occasionally	21.7	25.0	11.8	8.1	9.4	15.1	23,91	21,74	7,35	13,33	12,89	15,95
V15. If you could not use your own car, what alternative do you use?												
On foot	1,8	6,7	5	2,1	3,8	3	4	3,17	2,37	4,52	4,22	4,13
Bicycle	5,2	8,6	10,3	11,9	3	7,9	5,18	8,43	7,67	8,46	12,54	4,75
Urban Bus (Tussam)	12,9	27,2	40,2	43,6	18,3	40,8	22,96	16,11	34,04	42,38	36,58	21,88
Interurban Bus	9	5,8	17,9	18	14,7	32,3	8,32	9,08	17,24	22,44	15,65	19,55
Private Car	58,5	38,9	12,7	13	49,7	4,1	48,18	51,39	28,58	9,09	18,34	32,67
Motorcycle	1,9	3,6	1,1	2,1	1	0,5	2,64	2,03	1,52	1,3	2,13	1,01
Tram	0,4	2,4	0,6	1	0,5	2,5	1,5	0,69	0,69	1,18	1,39	1,56
Combination of vehicles	9,1	5,5	10,5	6,9	8,1	7,1	6,38	7,94	6,91	9,01	8,61	10,96
Other	1,2	1,4	1,7	1,3	0,8	1,6	0,85	1,16	0,99	1,62	0,54	3,49
V16. Availability of:												
Driver License	98.8	95.5	0.2	99.6	97.5	39.7	5,41	2,59	2,41	95,71	31,99	27,9
Access to private car	93.8	83.4	2.6	20.0	87.8	4.4	94,59	97,41	97,59	4,29	68,01	72,1
Access to motorcycle	7.2	11.0	3.7	4.6	8.4	4.1	0	0	0	0	0	0
Access to bicycle	38.2	39.6	53.6	45.5	52.0	32.6	14,59	16,97	50,9	98,27	54,07	56,58
None	0.0	0.0	43.8	0.8	0.0	39.7	85,41	83,03	49,1	1,73	45,93	43,42
V17. Age												
<18	0.0	0.0	16.5	0.2	0.0	0.3	0	0	0	11,92	5,67	1,11
18-25	6.3	3.1	78.1	89.1	92.6	13.7	5,56	6,11	92,64	62,81	38,89	59,48
26-40	44.9	47.1	5.2	10.6	6.6	45.5	39,77	51,12	7,35	14,33	29,28	26,07
41-65	47.5	47.3	0.0	0.0	0.8	38.1	52,38	41,95	0,01	10,17	23,98	12,71
>66	1.3	2.2	0.0	0.0	0.0	2.5	1,99	0,81	0	0,78	2,19	0,35
V18. Level of studies completed												
None or Secondary School	6.7	10.3	14.6	1.3	1.3	22.7	8,47	5,66	0,83	19,18	15,05	7,68
High School	26.1	23.3	66.1	63.3	61.2	24.4	22,71	23,91	66,57	57,02	42,22	46,88
Bachelors or higher	66.6	65.6	18.4	35.5	37.3	52.3	68,09	70,43	32,6	23,8	41,02	43,72
V19. Employment situation												
Employee	80,4	76	0,2	3,6	2,3	67,9	77,07	83,2	1,37	16,54	37,23	28,35
Student	0,5	0,9	97	90,6	92,6	0,8	1,94	1,48	93,8	71,32	42,47	55,28
Retired	3,6	6,3	0,2	0	0	3,8	5,49	2,48	0	1,2	4,27	1,47
Other	15,5	16,8	2,6	5,9	5,1	27,4	15,5	12,83	4,84	10,94	16,03	14,91

4.1. Cluster Analysis using socioeconomic variables

Table 1 summarizes the main characteristics for each group, and following are the results:

Cluster 1 (C1S) can be named "High income users with predisposition to use the private car" and it is composed of 26.46% of the whole sample. It is predominantly made of users from 26-65 years old, whose employment status is employed or retired, holding a university degree. Users belonging to C1S predominantly show the availability of own vehicle and driver's license; they make sporadic trips for work or other reasons. Moreover, a notable proportion of these users jointly use car and metro service to make their trip and they mainly consider car as an alternative to metro. Finally, users show a greater level of agreement in considering the lack of parking, traffic congestion and less frequently the unavailability of their vehicle as the main reasons to use the metro. This cluster is characterized by a predominant household monthly income over 2,401€.

Cluster 2 (C2S) represents 18.26% of the complete sample and it is made of "High income users with predisposition to use the metro". This cluster shows a higher proportion of users over 26 years old, whose status is employed or retired. University-level studies predominate. These users notably show the availability of private car and bicycle; they mainly perform sporadic and short trips for work, other reasons, and leisure. Also, they more

frequently consider walking, tram or motorcycle as an alternative to the metro than the general trend of the sample. A notably higher proportion of users consider speed, comfort and fare as the main reasons for using the metro. So, we can consider this group of users as users who present predisposition to use the metro, being attracted by service characteristics. Moreover, this group shows the highest average perception of the overall SQ. Also in this case, a monthly household income level over 2,401€ is more predominant than in the complete sample.

Cluster 3 (C3S) is composed of 16.73% of the sample and is named "Captive young students". In fact, users younger than 26 years old, students, with a high school or professional education degree predominantly compose the cluster. Moreover, this cluster shows a higher proportion of captive users (who have not another transport alternative different from the metro for that trip), and users without availability of driver license and an own vehicle; more than 20% of these users live in large families. They use the metro to reach their place of study.

Cluster 4 (C4S) (14.92% of the whole sample) is predominantly made up of "Captive university students", mainly composed by users between 18-25 years old, students, and with a high school or professional education degree. This cluster shows a distinctively higher proportion of users who used the metro because they are captive, not able to drive (no availability of a private vehicle although they own a driver license), and who use to make long trips for studies. Moreover, users in this cluster show the lowest assessment of the overall SQ.

Cluster 5 (C5S) represents 12.32% of the sample and it can be conveniently named "Non-captive university students" mainly consisting of users aged between 18-25 years old, students, with a high school or professional education degree, who daily travel for studies. This cluster did not show a proportion of captive users higher than the general trend, having almost 90% of users access to a private car. Furthermore, in comparison with the complete sample these users more frequently stated that they were using the metro service due to traffic congestion, lack of parking or unavailability of private vehicle; a certain part of these users reach the metro station by car.

Cluster 6 (C6S) is composed of 11.41% of the whole sample. These are "Users with low income and high predisposition to use the PT". These users more frequently declared that urban and metropolitan bus and tram are transport alternatives to the metro service and that they use the metro service mainly due to unavailability of driver license and owning a private vehicle. About 40% of the users have no other availability of transport modes. They show a relatively higher average perception of the overall SQ. CL6S contains a high proportion of users over 26 years old, without studies. There is a higher proportion of employed users, or in another situation and, less prominently, retired. They mainly travel due to work or other reasons. This cluster shows a predominant household monthly income under 1,200€.

4.2. Cluster Analysis using socioeconomic variables and service quality variables

Table 1 summarizes the main characteristics for each group, and following are the main results:

In general, the clusters which have been achieved using socioeconomic and SQ attributes (CSQ) are similar to the clusters which have been achieved using only socioeconomic attributes (CS). Only clusters 5 and 6 of both analyses differ completely and do not have direct mapping among them. However, it is noticeable that certain characteristics, which are common for both clusters, are more over-represented in CSQ than in CS. Thus, CSQ present more specific and typical/indicative profiles. Moreover, some other characteristics are only represented in CSQ:

Cluster 1 (C1SQ) is similar to C2S ("High income users with predisposition to use the metro"). The proportion of users is over 3% higher in C1SQ than C2S. The main differences between both are: a proportion of users in C1SQ consider comfort, the lack of parking and traffic jams as reasons to use the metro; and the more frequent alternatives to the metro are the tram, the motorcycle and the car. Thus, C1SQ can be described as users with high-income level and inclination to use the private car due to mobility problems. C1SQ has a positive perception of the metro, unlike CS2 in which users are attracted by service characteristics and have predilection to use the metro. Moreover, this is the second group on the scale in reference to satisfaction with SQ attributes.

Cluster 2 (C2SQ) agrees with C1S ("High income users with predilection to use the private car"). The proportion of users is 5.5% lower in C2SQ than in C1S. The main differences between both are: the employment status is only over-represented; the lack of parking and traffic jams (mobility problems) are only highlighted as reasons to use the

metro; and a high proportion of users use the metro sporadically. Consequently, C2SQ can be described as employed users with high level of income who have to use the metro due to mobility problems. The displeasure of C2SQ towards the metro is supported by its low level of satisfaction with SQ. It is the fifth group on the scale.

Cluster 3 (C3SQ) is similar to C4S ("Captive university student"). The proportion of users is over 3% higher in C3SQ than in C4S. In this cluster, the only difference is that in C3SQ the lack of parking is not over-represented as one reason to use the metro.

Cluster 4 (C4SQ) is similar to C3S ("Captive young students"). The proportion of users in C4SQ is the same as in C3S. The main differences between both are: the users choose the urban and interurban bus as an alternative to the metro; the level of education is high school, and the option No response has disappeared.

C3SQ and C4SQ present percentages of categories for each attribute more distributed in a heterogeneous way than C4S and C3S respectively. Both show a medium satisfaction grade with SQ aspects in comparison with the rest of clusters (fourth and third on the scale).

Cluster 5 (C5SQ) represents 12.78% of the total sample and it can be conveniently named "Low income users with high satisfaction and predisposition to use the metro" and does not correspond with any CS. This cluster shows a higher proportion of users under 18 years old and over 65 years old, whose status is retired or other. None or secondary studies predominate and they perform short trips. These users notably consider quickness, comfort, fare and ecology-awareness as the main reasons for their metro choice. Moreover, they have the highest level of satisfaction with each SQ aspect. Therefore, they can be described as the users who opt for the metro, being attracted by service's characteristics. A monthly household income of less than 1,200€ is over-represented.

Cluster 6 (C6SQ) represents 10.91% of the total sample and it can be appropriately named "Captive and dissatisfied users with the metro service" and does not correspond with any CS. This cluster shows a higher proportion of users between 18 and 25 years of age, students, and people with a high school or professional education degree. This cluster shows an over-representation of captive users who make long trips (> 40 min.). They have to use the metro, as it is their only alternative to reach their place for studying. Moreover, they are considerably dissatisfied with the quality of service, giving the lowest value satisfaction in all aspects in comparison with other cluster.

5. Conclusions

In this study we compared the profiles from two different cluster analysis (CA). The first was performed with only socioeconomic attributes and, the second was performed with socioeconomic and SQ attributes. We used data from a survey that was carried out in the Metro of Seville (Spain) in 2014. CA was used for identified different profiles of users that have more homogeneous opinions about the service. This advanced segmentation technique is able to consider at the same time various users' socioeconomic characteristics and SQ aspects, or only socioeconomics characteristics for finding the groups.

The outcomes find out interesting insights. CA with socioeconomics attributes determined six groups of passengers, representing diverse profiles. Cluster 1 grouped high-income users with inclination to use the private car. Cluster 2 grouped high-income users with predisposition to use the metro. Cluster 3 grouped captive young students. Cluster 4 grouped captive university students. Cluster 5 grouped non-captive university students and Cluster 6 grouped users with low income and high predisposition to use the PT. Although, in general, they share the same profiles as the clusters that are performed with the profiles of CA with socioeconomics and SQ attributes, they present some differences in their characteristics. The profiles that are identified with a CA with socioeconomics and SQ attributes present more heterogeneous characteristics than only using socioeconomic characteristics. C1SQ to C4SQ are very similar to profiles C1S to C4S. This is not the case for C5SQ and C6SQ; they form two new interesting profiles: C5SQ groups low-income users with high satisfaction and predisposition to use the metro; and C6SQ groups captive and dissatisfied users with the metro service.

Thus, this type of segmentation allows service operator to focus their investments and market strategies in specific aspects of SQ in order to attract potential passengers with a specific profile or improve their satisfaction.

Moreover, these issues are of interest for transport planners, who, in order to stablish marketing strategies for PT promotion, they could use at the same time socioeconomic characteristics and perceptions of SQ to identify more different heterogeneous and characterized profiles of users than only using socioeconomic characteristics.

Acknowledgements

References

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52(3), 317–332.

Coffel, K. (1995). Customer satisfaction index for the mass transit industry.

de Oña, J., de Oña, R. (2015). Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. Transportation Science, 49(3), 605-622.

de Oña, J., de Oña, R., Eboli, L., Mazzulla, G. (2015). Heterogeneity in perceptions of service quality among groups of railway passengers. International Journal of Sustainable Transportation, 9(8), 612–626.

de Oña, J., de Oña, R., López, G. (in press). Transit service quality analysis using cluster analysis and decision trees: A step forward to personalized marketing in public transportation. Transportation, http://dx.doi.org/10.1007/s11116-015-9615-0.

de Oña, J., López, G., Mujalli, R., Calvo, F.J., (2013). Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accident Analysis & Prevention, 51, 1–10.

de Oña, R., López, G., Diez-Mesa, F., de Oña, J. (2014). Cluster analysis for diminishing heterogeneous opinions of service quality public transport passengers. Procedia-Social and Behavioral Sciences, 162, 459–466.

de Oña, R., Machado, J. L., de Oña, J. (2015). Perceived service quality, customer satisfaction and behavioral intentions: A structural equation model for the metro of Seville, Spain. Transportation Research Record, 2538, 76-85.

Erdman, C., Adams, T., O'Hare, B. C. (2015). Development of interviewer response rate standards for national surveys. Field Methods, , 1-17.

Fraley, C., Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. The Computer Journal, 41(8), 578–588.

Hair, J. F. (2010). Multivariate data analysis (P. Education Ed. 7th Edition ed. ed.). New Jersey: Prentice Hall.

Magidson, J., Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. Canadian Journal of Marketing Research, 20(1), 36–43.

McLachlan, G., Peel, D. (2004). Finite mixture models John Wiley & Sons.

Raftery, A. E. (1986). A note on bayes factors for log-linear contingency table models with vague prior information. Journal of the Royal Statistical Society, Series B, 48, 249–250.

Vermunt, J. K., Magidson, J. (2005). Latent GOLD 4.0 user's guide Statistical Innovations Inc.

Wen, C., Lai, S., Yeh, W. (2008). Segmentation and positioning analysis for international air travel market: Taipei-to-tokyo route. Transportation Research Record: Journal of the Transportation Research Board, (2052), 46–53.