Analysis of transit quality of service through segmentation and classification tree techniques

By:, Rocío de Oña & Juan de Oña

Rocío de Oña & Juan de Oña (2015) *Analysis of transit quality of service through segmentation and classification tree techniques*. **Transportmetrica A: Transport Science, 11(5), 365-387** DOI: 10.1080/23249935.2014.1003111

Analysis of transit quality of service through segmentation and classification tree techniques

ABSTRACT

Perceptions about the quality of service are very different among public transport users. Users' perceptions are heterogeneous for many reasons: the qualitative aspects of public transport service, users' socioeconomic characteristics, and the diversity of tastes and attitudes toward public transport. By analyzing different groups of users that share a common characteristic (e.g. socio-economic or travel behavior) it is possible to homogenize user opinions about the quality of service. This paper studies quality as perceived by users of the metropolitan transit system of Granada (Spain) through a classification tree technique (CART) based on five market segmentations (gender, age, frequency of use, reason for traveling, and type of ticket). CART is a non-parametric method that has a number of advantages compared to other methods that require a pre-defined underlying relationship between dependent and independent variables. The study is based on data gathered in several customer satisfaction surveys (non-research oriented) conducted in the Granada metropolitan transit system. The models' outcomes show that some attributes are very important for almost all the market segments (punctuality, information), while others are not very relevant for any of the segments — most notably fare, despite the fact that fare was stated as very important by most of the passengers during the interview.

Keywords: Service quality; quality management; public transport; bus transit; data mining; classification and regression trees (CART); segmentation

1. INTRODUCTION

Service Quality (SQ) in the arena of Public Transport (PT) is a key factor in attracting and retaining new users. An increase in passenger satisfaction translates into retained markets, increased use of the system, newly attracted customers, and a more positive public image (Transportation Research Board, 1999). Therefore, the ability to identify those factors having the greatest impact on SQ has become a guiding principle for PT planners and managers, helping them to decide where to direct their service operations and enhancement efforts.

Many authors hold that SQ must be measured from the customer's perspective> As Berry et al. (1990) point out, “customers are the sole judges of service quality”. PT companies usually want to know how their customers rate them on detailed service attributes, in addition to the relative importance of these attributes for customers. Hence, periodic surveys on user satisfaction related to different aspects of service are conducted to provide quantitative assessment of the key aspects from the users’ point of view (del Castillo and Benitez, 2013). Customer satisfaction surveys (CSS) are the most widely-used technique to obtain this information. Based on the CSS, two separate approaches may be applied to identify the attributes of greatest importance (de Oña et al. 2013). The most common approach is to ask customers to rate each attribute on an importance scale (stated importance). The main alternative, now growing in use, would be to derive the attributes’ importance by statistically testing the strength of the relation between individual attributes and overall SQ or satisfaction. Stated importance is still the most intuitive and simple method, and it is used very often by PT planners and managers. However, this approach has several drawbacks: it increases the length of the survey; yields insufficient differentiation among mean importance ratings, with customers rating nearly all of the measures near the top of the scale; and attributes may be rated as important even though they have little impact on SQ (Weinstein, 2000).

Another important issue in SQ analysis is that customers’ perceptions present heterogeneity among users, which is a problem for many techniques that intend to measure SQ. Such heterogeneity depends on a variety of factors that are difficult to quantify: the qualitative nature of certain aspects that characterize the services (Awasthi et al. 2011; Dad and Pandit 2014), the attitudes passengers have towards the use of PT, the different ways of viewing aspects of the service, and the social and economic characteristics of passengers and their preferences (Eboli and Mazzulla, 2011). Stratified sampling on more uniform segments of passengers may help to reduce heterogeneity. Some authors (Dell’Olio et al., 2010; Chou et al. 2014) propose specific models after conducting stratified sampling based on the social and demographic characteristics of the passengers (i.e. models for women, for the elderly, according to income level, etc.), or independently analysis of some lines of the same public transport service to determine if there are differences in passenger requirements among separate lines (Bordagaray et al., 2014). Other authors (Eboli and Mazzulla, 2008; Cirillo et al., 2011) have proposed mixed logit models to handle heterogeneity. However, most of these models have their own model assumptions and pre-defined underlying relations between dependent and independent variables, such as normal data, linear relationships between dependent and independent variables, low multi-colinearity, and so on. According to Garver (2003), these assumptions are almost always violated in customer satisfaction research. In fact, SQ measured from the passengers’ perspective tends to show multi-colinearity problems, owing to the fact that passengers group items together even when satisfaction surveys address completely different aspects (such as passengers’ evaluation of service punctuality with the evaluation of speed, for instance) (de Oña et al., 2012). If these assumptions are violated, the model could lead to erroneous estimations of the likelihood of SQ.

Some data mining techniques are able to overcome the above-mentioned limitations due to they are non-parametric models with non-predefined underlying relationships between variables. For example, de Oña et al. (2012) applied a decision tree methodology for analyzing the quality of a metropolitan bus public service in Granada (Spain), and subsequently, Garrido et al. (2014) investigated the most influent factors affecting users overall satisfaction about the same bus service operating in Granada by adopting an artificial neural network technique. Decision trees have been widely employed in several fields (business administration, agriculture, industry, engineering, etc.), but also artificial neural networks have been successfully used in other transportation engineering fields such as choice behavior. However, decision trees models could be preferred by public transport managers because its simplicity for interpreting the results due to its graphic representation. Therefore, and given the possibilities afforded by this methodology as opposed to other parametric models this paper uses a decision tree methodology (Classification and Regression Tree algorithm, CART) to identify the key factors affecting SQ in passenger PT. Moreover, in order to reduce the heterogeneity present in passengers' opinions, a stratification of the sample using demographic and travel behavior characteristics was performed, and different CART models were developed for each specific group of passengers.

This paper aims to identify the attributes that have the greatest impact on SQ in metropolitan PT by bus, by considering the heterogeneity present in users' perceptions, so that PT managers may draw up more effective strategies that target specific groups of users (i.e. personalized marketing) depending on the needs and perceptions regarding each one of them.

Models specifically designed for each user segment were developed for this purpose. We used the user segmentation normally applied in other SQ studies (gender, age and frequency of use), as well as others used less frequently (travel reason and type of ticket). Stratifying the sample of users into more homogeneous segments is a method that has been used before to reduce the heterogeneity of passengers' perception about PT (Dell'Olio et al., 2010), but not in the context of metropolitan PT by bus, where performance characteristics and requirements differ widely from those of urban PT and even from metropolitan PT by different modes of transport (e.g., suburban rail). Still, policies for SQ in PT can only succeed if specific measures are applied, focusing on the characteristics of the type of service under study and particular passenger needs; that is, a generic framework of action is not recommended.

The main purpose of this study is therefore to examine whether the evaluation of SQ, as well as the key drivers towards SQ, are different for each one of the market segments studied. The paper uses the data gathered in four CSS conducted in the Granada metropolitan transit system from 2008 to 2011, which were non-research oriented surveys (designed for a rather simple statistical frequency analysis) developed by the Transport Consortium of Granada. The research results contribute to general knowledge on SQ and bring practical value to the public transport industry by identifying the key factors in each market segment, so that planners can better focus their efforts to enhance quality.

Section 2 presents the methodological approach and a description of the available data. Section 3 follows with the model results and discussion. The paper concludes with a summary and the principal conclusions.

2. MATERIALS AND METHODS

2.1. Methodology

Decision trees are one of the most widely-used data mining techniques for classifying and predicting class variables. When the target variable is discrete, a classification tree is developed, whereas a regression tree is developed for continuous variables. CART can be used for both kinds of target variables. In this study the target variable is discrete (bus SQ: Poor, Fair and Good), and therefore a classification tree is developed.

One of the most valuable outcomes provided by CART analysis is the value of the standardized importance of independent variables, which reflects the impact of such predictor variables on the model. Information is obtained for all the independent variables, making it easy to find which ones are the most important.

CARTs are developed in three steps: a) tree growing; b) tree pruning and; c) optimal tree selection. Below the reader will find a short description of each one of these steps. A more detailed description of CART analysis and its applications can be found in Breiman et al. (1998).

2.1.1. Tree growing

The principle behind tree growing is to recursively partition the target variable to maximize "purity" in the child nodes so the data in each child node will be more homogeneous than those in the upper parent node. To achieve this, a set of candidate split rules is created, consisting of all possible splits for all variables included in the analysis. These splits are then evaluated and ranked based on a certain criterion, to choose amongst the available splits at every non-terminal node. The most famous splitting index is the Gini Index. The Gini criterion measures the "worth" of each split in terms of its contribution toward maximizing the homogeneity through the resulting split. If a split results in the splitting of one parent node into B branches, the "worth" of that split may be measured as follows:

$$\boldsymbol{Worth = Impurity\ (Parent\ node) - \sum_{b=1}^{B} P(b) * Impurity\ (b)} \quad (1)$$

where Impurity (Parent node) denotes the Gini measure for the impurity (i.e., non-homogeneity) of the parent node, and P(b) denotes the proportion of observations in the node assigned to branch b. The impurity measure, Impurity (node), may be defined as follows, in which I is the number of classes in the target variable:

$$\boldsymbol{Impurity(node) = 1 - \sum_{i=1}^{I} \left(\frac{number\ of\ class\ i\ cases}{all\ cases\ in\ the\ node}\right)^2} \quad (2)$$

If a node is ‘pure’, all the observations in the node belong to one class and the Impurity (node) will be equal to zero.

This process is applied recursively to achieve child nodes having maximum worth, which in turn become the parents to successive splits, and so on. The splitting process is continued until there is no (or less than a pre-specified minimum) reduction in impurity and/or the limit for a minimum number of observations in a leaf is reached. Following this process a saturated tree is obtained (a tree that overfits the data set from which it is constructed, but that is not useful for classifying another data set). Therefore, to develop a CART model, the data is divided into two subsets, one for training and the other for testing. The training sample is used to split nodes, while the testing sample is used to compare the misclassification. The saturated tree is then constructed from the training data. Overly large trees could result in higher misclassification when applied to classify new data sets.

2.1.2. Tree pruning

To lessen the complexity of the saturated tree that overfits the training data and creates simpler trees, the tree is “pruned” in the second step. This pruning is performed according to the cost-complexity algorithm, which is based on removing the branches that add little to the predictive value of the tree. After pruning a branch, if the increase in the misclassification cost is sufficiently lower than the decrease in the complexity cost, that branch will be pruned, and a new tree is created. As more and more nodes are pruned away, simpler and simpler trees are the result. At the end of the tree pruning process, the relationship between the misclassification costs and tree complexity in terms of the number of terminal nodes is obtained.

2.1.3. Optimal tree selection

The principle behind selecting the optimal tree from the pruned ones is to find a tree on the testing dataset that does not overfit the information in the training dataset.

When the tree grows larger and larger, the misclassification cost for the training data decreases monotonically, indicating that the saturated tree always gives the best fit to the training data. On the other hand, in the misclassification cost for the testing data, first there is a decrease, and then an increased is observed, after reaching a minimum. Thus, the optimal tree is the one that has the least misclassification cost for the test data.

2.2. Data

We investigated the metropolitan PT service in the area of Granada (Spain). This is a medium-sized metropolitan area with a population of around 500,000 inhabitants. A Granada Area Transport Consortium was created in 2003 to coordinate transit bus service management in the metropolitan area. Since 2007, the PT service is provided by a bus system in which 15 bus companies operate in 18 transport corridors. The metropolitan PT system carries around 10 million passengers per year.

The Granada Area Transport Consortium uses a CSS to evaluate the SQ of the metropolitan PT system. The surveys were done by a company specialized in developing this sort of surveys. To ensure the coverage of the area and of customers, the surveys were conducted at the heads of the lines that make up the metropolitan bus network. This network has a radial structure focused on two entrances to Granada —one in the north of the city and the other in the south— owing to the fact that 90% of the trips take place between the nearby municipalities and the city of Granada per se. Respondents were randomly selected establishing a minimum representativeness of certain segments of passengers (minimum stratification representativeness considering gender and age). Around 1,000 face-to-face surveys are conducted annually. We used the data gathered in four CSS conducted from 2008 to 2011 so as to study some market segments with low representation on the total sample (i.e. elderly people or specific travel tickets). Those who did not respond to the overall SQ question were omitted from our analysis. Consequently, the usable sample size was 3,664 respondents.

The CSS are divided into two main sections. The first gives general information about the trip, socioeconomic characteristics of passengers and travel habits. Information is gathered about the time of the interview, the bus stop, the bus route to be taken, the operator, the origin, destination, and so on. The variables used to describe the passengers' demographic profile include gender and age; and the ones that describe the passengers' travel behavior include travel reason, use frequency, type of ticket, private vehicle available, complementary modes from origin to bus stop, and complementary modes from bus stop to destination (see Table 1).

Of the respondents, a majority of 2,493 (68%) were female, and 1,171 were male (32%). Half of the respondents were age 18 to 30 (49%); 1,479 (40%) were age 31 to 60 and only 376 (11%) were older than 60. For 1,027 (28%) of the respondents the reason for travelling was occupation, whereas for 911 (25%) the reason was studies. The rest of the respondents travelled for other reasons, such as doctor visit (11%), shopping (7%), holidays (6%) and other reasons (23%). Most respondents used the service almost daily (57% using it more than four times a week), 22% were frequent passengers (from 1 to 3 times a week). The rest reported using the service occasionally (13% from 1 to 3 times a month) or sporadically (8% using it less than once a month). Altogether, 2,445 respondents used the consortium pass (67%), as opposed to 850 (23%) who used the standard ticket, and 249 (7%) who had the senior citizen pass; 120 (3%) used some other type of ticket. The sample of users is equally distributed among those who had a private vehicle available for making the trip (47% of the respondents) with those who did not have available a private vehicle for making the trip (53% of the sample). Most of the respondents gain access to the bus service on foot (77% of the passengers), while others use the urban bus (18%), the metropolitan bus (2%), a private vehicle (1%), or other complementary modes of transportation (motorbike, bicycle, taxi or others). Moreover, almost all the respondents get to their final destination from the bus stop on foot (95%) rather than by urban bus or private vehicle.

(Insert Table 1 here)

The criteria used for stratifying the sample of passengers were selected among the 8 variables describing passengers' demographic profile and travel behavior (Table 1). A statistical analysis was performed to confirm the existence of statistical differences between the attribute perception rates among the given segments of passengers. Because of the non-normality of the data, non-parametric techniques were used (Kruskal-Wallis and Dunn test). Analysis showed that there were statistically significant differences, with a 95% confidence level, for the segments generated under the following criteria: gender, age, travel reason, use frequency and type of ticket (see Table 2). This indicated a need to calculate different models for 14 different market segments. According to the travel reason criterion, in our analysis only three segments were considered, representing the most important reasons: travelling for occupation, travelling for studies, and all other reasons (including doctor, shopping, holidays and so on). Regarding use frequency, the number of segments was reduced to two in order to ensure adequate representation of each group. Passengers travelling almost daily and frequently were grouped and labeled as frequent passengers, and passengers travelling occasionally and sporadically were also grouped and labeled as sporadic passengers. The rest of the market segments were obtained directly from the categories of criteria defined in the survey: male versus female for the variable gender; [18-30], [31-60] and [>60] were the three age ranges displayed in the survey; and standard ticket, consortium pass, senior citizen pass and other type of ticket were the four options for ticket type.

(Insert Table 2 here)

The second section of the survey focuses on passengers' perception of more specific service characteristics. First, the interviewers asked the passengers about their perception of performance with regards to 12 SQ factors, on an 11-point scale, based on a Likert scale (0 defined as Very low quality and 10 defined as Very high quality). Second, they asked the passengers to identify the three most important SQ factors out of each one of the 12 factors. This question was stated as: "Which three service aspects, among those displayed in the previous question, would you choose as the most important for you?" And finally, they were asked about the overall SQ perception based on a 5-point Likert scale. The variables used to measure the perception of the SQ attributes included provision of information, punctuality, safety on board, driver courtesy, bus interior cleanliness, bus space, bus temperature, accessibility to/from the bus, fare, speed, frequency of service, and the proximity of stops to/from origin/destination. Table 3 shows the average rates and standard deviation calculated from the performance perception rates expressed by the users with regards to the 12 SQ attributes, as well as the number of times that the passengers identified each attribute as being the most important (Option 1), the second most important (Option 2) or the third most important (Option 3).

(Insert Table 3 here)

In general, the average SQ rates suggest that people are fairly satisfied with the service. All the attributes have an average rate higher than sufficient (>6). The service characteristics considered to have the highest SQ are "driver courtesy" (7.91), "safety on-board" (7.59), "bus interior cleanliness" (7.47) and "bus temperature" (7.37). These four characteristics also have the lowest dispersions in their evaluation (all presenting a standard deviation lower than 1.85). On the other hand, the service characteristics with the lowest SQ but with the highest dispersion were "frequency of service" (6.08), "fare" (6.14) and "information" (6.43).

A closer look at the values of importance shows users to judge three attributes as very important (with a frequency of around 50%): "frequency of service", "fare" and "punctuality". Two of them (frequency and fare) were also identified as attributes with the lowest SQ. "Safety" and "speed" were also among the five most important attributes.

3. RESULTS AND DISCUSSION

Fourteen different CART decision trees were built – one for each market segment – and another CART was built for the overall market. In each CART all the service attributes (12) were considered in the models as well as the variables defining socioeconomic characteristics and travel habits of passengers that had not been used for segmentation (eight variables for the overall market model and seven variables for the other 14 models).

To arrive at results easier for PT managers to interpret —and taking into account the low frequency of certain categories of the 5-point Likert scale at target variable (overall SQ) have: 1 (2.5%), 2 (5.9%) and 5 (4.0%) —this dependent variable was recorded in a reduced semantic scale. This three-point scale comprised ratings 1 and 2 as POOR, 3 as FAIR, and 4 and 5 as GOOD. The explanatory constructs included in the analysis were the passengers' demographic profile, the passengers' travel behavior, and the performance perception of several SQ attributes measured on the original eleven-point scale.

In addition, the most important variables in the prediction of the dependent variable were obtained at each developed model using for this purpose a normalized form of the importance index (Kashani and Mohaymany, 2011). The importance index extracts the influence of each independent variable on the model in terms of the improvements that each variable produces when used as a primary or surrogate splitter across all splits in the tree. For the sake of simplicity, here only the five most important variables are shown in each case, highlighting the attributes considered to lead to service quality. Moreover, in order to compare the derived importance of the variables extracted from the models with those stated by the passengers in the survey, Table 4 indicates the stated versus derived importance for the overall market, and for each of the 14 market segments analyzed.

(Insert Table 4 here)

All CART models gave a precision ratio ranging from 63.7 % for the Standard Ticket, up to 79.1% for the Senior Citizen Pass. The precision rates are acceptable for all CART models, and they are higher than the values obtained in other studies in which decision trees were applied for SQ analysis (de Oña et al., 2012; Wong and Chung, 2007).

3.1. CART for the overall market

Figure 1 illustrates the CART for the overall market. The root node (Node 0) is split into two child nodes (Node 1 and Node 2), using the variable that maximizes 'purity' in the two child nodes. In this case, the splitter was INFORMATION. When INFORMATION is rated with a score higher than 6 (Node 2), the overall SQ is likely to be perceived as *GOOD* (75.7%). The fact that 72.1% of the sample is concentrated in this child node (Node 2) demonstrates that it is a great discriminant of the model. The next best splitting criterion for those who scored INFORMATION with a value equal to or lower than 6 is FREQUENCY. This is a key variable for discriminating user perception of overall SQ. It groups those who give a value of *POOR* or *FAIR* on the left side (Nodes 5 and 6), as opposed to those who rate it as *GOOD* or *FAIR*, on the right side (Nodes 8, 9 and 10). The cut-off point for FREQUENCY is a value of 2. When perceived FREQUENCY is very bad (≤2) and PROXIMITY is considered insufficient (≤4), there is a high probability (69.7%) that the passenger will rate SQ as *POOR*. On the other hand, if the FREQUENCY scores higher than 2 and TEMPERATURE has an adequate score (>6), SQ perception will be *GOOD*. When FREQUENCY scores high enough (>6), a rating of *GOOD* is obtained even when the score for TEMPERATURE is 6 or lower. This tree is 68.6% accurate.

(Insert Figure 1 here)

Table 4 shows the normalized importance rates derived from the model (Derived Importance). PUNCTUALITY is the most important attribute for SQ in the overall market, followed by TEMPERATURE, INFORMATION, FREQUENCY and SAFETY.

3.2. CART for the different market segments

We analyzed 14 market segments, stratified according to five variables used as segmentation criteria (Gender, Age, Use Frequency, Travel Reason and Type of Ticket). A CART was built specifically for each one of the segments: Male and Female; Young people (age interval {18-30}), Middle age ({31-60}) and Elderly people ({>60}); Frequent and Sporadic passengers (separated at 1 trip per week); people who take the bus for Working, Studying and Other Reasons; and passengers using different types of tickets (Standard Ticket, Consortium Pass, Senior Citizen Pass and Other Tickets).

3.2.1. CART for Gender segments

The splitter variable that best splits the root node for males and females is different. In the male model (see Figure 2) the variable is INFORMATION. As occurred in the global model (see Figure 1), when the variable obtains a good value (>6) the overall SQ perception is *GOOD* (80.5%). This model uses TEMPERATURE, SAFETY, SPEED and FARE as successive splitter variables, and the accuracy obtained with the model is 72.0%, meaning it is more accurate than the model generated for the entire market.

(Insert Figure 2 here)

If we focus on women, the model developed (65.3% accuracy) uses FREQUENCY to split the tree into two branches. On the right side of the tree, where FREQUENCY has been rated positively (>6), all the leaf nodes obtain *GOOD* values for the variable class, with the exception of Node 25, in which ACCESSIBILITY, COURTESY and TEMPERATURE condition the category selected to a *FAIR* value. On the left side of the tree, TEMPERATURE is the attribute that discriminates SQ the best. For values lower than or equal to 6, the variable class obtains a value of *POOR* or *FAIR*, whereas for TEMPERATURE values higher than 6, the variable class will be mainly *GOOD* or *FAIR*.

Some of the most important variables identified by the importance index coincide for both gender segments (PUNCTUALITY, INFORMATION and FREQUENCY), although women place more importance on TEMPERATURE and CLEANLINESS of the service, and men on SAFETY and COURTESY.

3.2.2. CART for Age segments

In the trees generated for age-related market segments, for Young People, PUNCTUALITY is the variable splitter that best discriminates perceived quality, compared to FREQUENCY for Middle-aged and INFORMATION for Elderly, coinciding with the most important variable at each segment, respectively, according to the importance index (Table 4). This may be because most of the individuals under age 30 (50.4%) are students having non-adaptable schedules, who expect the service to be on time. Individuals aged 30-60 have more flexible schedules, so they attach more importance to FREQUENCY. For example, in Spain, when people use public transport to get to work they are supposed to arrive at a more or less fixed time, although it is acceptable to vary it somewhat. If they miss the bus, but the next one arrives in just a few minutes, they still reach their place of work in due time. For the elderly (>60 years old), most of them retired, PUNCTUALITY and FREQUENCY are less important, and so they focus more on good information regarding service. It is worth pointing out that Node 7, in the model for the elderly (see Figure 3), is a pure Node in which quality is rated as POOR in all cases. This occurs after a series of evaluations on INFORMATION and SPEED that end in the key factor of PROXIMITY —and if the latter it is not rated as good (≤6), the global evaluation of quality will not be good either. The accuracy obtained in this model is 64.7% for Young people, 68.2% for Middle-aged and 70.0% for the Elderly.

(Insert Figure 3 here)

3.2.3. CART for Frequency of Use segments

The results given in Figure 4 reveal that service quality for frequent travelers can be explained by the model created for the overall market, after pruning a few of its branches. This may be because most of the passengers interviewed (78.3%) use the service constantly, and make up a large percentage of the overall sample. It would be erroneous to assume that frequent users are not concerned about the quality of the INFORMATION because they already have ample experience using the system. Quite the contrary, any sudden changes in itineraries and time tables are more upsetting precisely for these everyday users, and therefore the INFORMATION factor is decisive in their assessment of quality. Quality for sporadic passengers, in turn, may be best explained in terms of PUNCTUALITY. When users take a bus occasionally, they are only concerned with the bus being on time, and pay less attention to other features. In fact, a look at the results obtained with the importance index shows PUNCTUALITY to be at the top of the ranking, far from the other variables (CLEANLINESS, COURTESY and TEMPERATURE having normalized importance under 30%). The accuracy of the models for frequent and sporadic passengers respectively, is 68.3% and 69.8%.

(Insert Figure 4 here)

3.2.4. CART for Travel Reason segments

The models determined for the Travel Reason underscore what was interpreted in trees for different age groups. Granted, when we refer to Young people {18-30}, the segment encompasses most of the segment that has Studies as the Travel Reason, and when we target the population aged 30-60, most travel for Work-related reasons. When one travels for Studies (see Figure 5), the most important variable is PUNCTUALITY in light of lessons or exams. Yet when the reason for travel is Work, FREQUENCY becomes the most discriminant variable. In the cases of travel for some other reason (Others), INFORMATION becomes the most important attribute. These models are 65.9%, 67.7% and 69.6% accurate, respectively.

(Insert Figure 5 here)

3.2.5. CART for Type of Ticket used segments

The individuals who used a Senior Citizen Pass discriminate their perception of quality in terms of INFORMATION, ACCESIBILITY and COURTESY (see Figure 6), although SPACE and CLEANLINESS are also important service characteristics for them (see Table 4). The heightened importance of ACCESIBILITY could stem from mobility issues generally faced by the elderly. The discriminant factors for passengers who use Standard tickets, however, are SPEED and COURTESY. In the case of those who use the Consortium Pass, the tree is more complex, owing to a range of attributes, including INFORMATION, FREQUENCY, PROXIMITY, AGE and TEMPERATURE. This tree is very similar to the one built for the overall market, as most passengers (66.7%) use this kind of ticket. The accuracy attained in these models is 79.1% for Senior Citizen Pass, 63.7% for Standard ticket, 67.8% for Consortium pass, and 65.0% for Other tickets.

(Insert Figure 6 here)

3.3 Discussion of the results

Some similarities and differences were found among the CART models built for different market segments, since in some cases the variable used for splitting the root node was repeated across various models (for example the variable INFORMATION), until the size of the tree was very different among these groups of passengers —in terms of both the number of nodes the tree had and the number of levels of these trees. The diverse tree model structure reflects the heterogeneity of the passengers' opinions.

The models reached good rates of prediction, with accuracy values above 63.7%. In some market segments, even higher accuracy is attained (72.0% for Men, 79.0% for Elderly, 69.8% for Sporadic, 69.6% for Other travel reason and 79.1% for Senior Citizen pass). This demonstrates that such segmentations can lead to sample homogeneity, as well as better results from the models.

Concerning the importance frequencies by market segment (Stated Importance in Table 4), the five most important attributes for the overall market are also identified as the most important in all market segments, except for three market segments: Elderly, Senior citizen pass and Other type of ticket. For Elderly and those who use a Senior citizen pass, the three most relevant attributes for the overall market (FREQUENCY, FARE and PUNCTUALITY) remain among the five most important attributes, but SAFETY and ACCESSIBILITY also become highly important. In these market segments SAFETY is considered paramount, and ACCESSIBILITY is among the top five attributes. With regards to "Other type of ticket", the attribute INFORMATION comes third in importance. Therefore, based on the question about importance, no significant differences are observed among the different market segments, contrary to the results from earlier studies (Andreassen, 1995; dell'Olio et al., 2010; Ganesan-Lim et al., 2008). Once again, these results point to the limitations of using stated importance to identify the importance of each attribute (Weinstein, 2000).

With respect to the importance derived from the models, PUNCTUALITY was found among the three most important variables in the overall market and in all the market segments —with the exception of Elderly and those who used a Senior Citizen Pass (two groups that are highly correlated). Most of these people are retired and do not have to comply with a schedule for work or studies, so they concede more importance to INFORMATION, ACCESIBILITY or COURTESY. A number of authors (dell'Olio et al., 2010; Eboli and Mazzulla, 2010; del Castillo and Benitez, 2013) analyzing SQ for bus transport also identified PUNCTUALITY as one of the attributes with the greatest impact on overall SQ. The variables TEMPERATURE, INFORMATION and FREQUENCY are also identified as having a lot of weight on overall market. Two of these (TEMPERATURE and FREQUENCY) coincide with the variables obtained by dell'Olio et al. (2010). In that case, they did not evaluate TEMPERATURE as a separate attribute of the service. Instead, they used overall comfort of the service as a variable that can encompass TEMPERATURE as well. It would be reasonable to suppose that travel comfort is a relevant attribute on long journeys, as in the case of metropolitan transport.

INFORMATION is another attribute that is highly important in most categories (despite the importance stated by the passengers), although it is less important to young people and sporadic passengers. Young people may not attach much importance to INFORMATION because they are skilled at using new technologies and/or interpreting the information panels for travelers. Sporadic passengers only use PT occasionally; what matters most to them is PUNCTUALITY, other attributes being considerably less important.

There are three attributes that appear many times (7) among the five most important ones for various market segments: FREQUENCY, SAFETY and COURTESY. FREQUENCY is identified as one of the five most important attributes for men and women, middle-aged passengers, frequent users, users who travel for work-related reasons and passengers who use standard or other tickets. SAFETY is identified as one of the five most important attributes for males, young people, frequent and sporadic users, passengers who travel for studies or other reasons, and for consortium pass users. The profile of users who attach considerable value to COURTESY would be male, young or old, sporadic users of the service for other travel reasons, and having a standard ticket or senior citizen pass. It seems logical that the elderly or infrequent users, who usually buy their ticket on board and therefore necessarily interact with the driver, would rate COURTESY as a highly important aspect.

A further group of attributes related to travel comfort (SPACE, TEMPERATURE and CLEANLINESS) is repeated in five or six market segments. This is to be expected in the sense that these are metropolitan trips that tend to have longer itineraries. Also noteworthy is the fact that some attributes are considered highly important only for some specific market segments; for instance ACCESSIBILITY attains importance values of 82.8% for Senior citizen pass users, but it is not among the five top attributes in the remaining market segments.

Finally, Table 4 shows that FARE is not among the top five variables for the overall market nor any of the market segments. This contrasts with the results shown in the same table regarding the stated importance, where FARE is among the three most frequent attributes, and in Table 3, where FARE was one of the attributes with the lowest SQ. Interestingly, it also shows that INFORMATION is identified as one of the five most important attributes for the overall market and for eleven market segments, and yet it is among the five most frequent attributes (Stated Importance) in only one market segment (Table 4). INFORMATION is likewise one of the three attributes having the lowest SQ (Table 3). Such differences may be explained by:

- When people are asked to rate the most important variables of the service, FARE is frequently rated as very important for SQ. Notwithstanding, when the weight of the variable

FARE is obtained using models based on the overall SQ, it is found to be less important than other variables having a much greater impact (frequency, punctuality, safety, etc.) (dell'Olio et al., 2010; de Oña et al., 2012), pointing to the real influence that each factor plays in the overall SQ evaluation. This contradiction can be traced to the fact that passengers accept the price of the ticket, but they want to highlight the importance it has for them for choosing that mode of transport for their trips. Moreover, passengers tend to appraise the quality of the FARE as poor even though they are not dissatisfied with it, being afraid of a possible rise in price if their evaluation is positive.

- The contrary would be true with INFORMATION. When users are asked about this variable directly, they do not usually rate it very highly (de Oña et al., 2012; Eboli and Mazzulla, 2010);, yet when it is inferred from the models based on overall SQ, the importance of INFORMATION increases (Andreassen, 1995; de Oña et al., 2012; Eboli and Mazzulla, 2010).

4. SUMMARY AND CONCLUSIONS

The main objective of this study is to demonstrate that a combined use of CART analysis and segmentation can be very useful to learn more about the attributes rated as the most important in an evaluation of SQ in a PT metropolitan service by bus. A secondary aim is to identify differences between the overall market and the various market segments.

Our work was based on the data from several customer satisfaction surveys provided by the administration, which were designed for a rather simple statistical frequency analysis, in order to evaluate transit SQ. However, the application of CART methodology proves that this kind of data may also reveal very interesting information for managers, administrations and PT operators. It also provides valuable insight contributing to effective decision-making to promote the use of PT.

The main advantage of using the CART model is that the outcomes of the analysis are easy to understand, owing to the graph representation afforded by the results. Moreover, the CART analysis allows many explanatory variables to be processed, and the most important ones are easy to identify. In this case, variables of different natures were used, some pertaining to passengers' social and demographic characteristics and travel behavior, and others to service features. This afforded knowledge of the weight of all the variables in the model. A further advantage is that CART analysis effectively handles multi-colinearity problems, which is one of the major drawbacks of regression models.

However, the specific findings of this paper cannot be generalized to other context-related PT services (such us urban PT, or even metropolitan or suburban PT services involving modes of transport other than the one analyzed here) because the performance characteristics and passengers requirements differ widely among transit services. Although these results should not be extrapolated to other types of PT services, the fact that CART methodology represents a powerful and suitable tool for analyzing different public transport systems is a valuable finding with broad implications.

Policies for improving SQ in PT can only succeed if they apply specific measures focusing on the characteristics of the type of service under study and the specific needs of the customers; a generic framework of action is not recommended. In this paper we verify that the main attributes leading to service quality tend to change, depending on the market segment under study. Therefore, when analyzing SQ, it is advisable to take different groups of users into consideration so that transport planners might direct their efforts more accurately to the group of users whose loyalty they seek by attending to their preferences and needs.).

Such segmentation (i.e. using a personalized marketing approach) normally includes the frequent and sporadic users, or else users are grouped by sex, age or minimum income. In this study, in addition to previous segmentations, we also used less common groups (i.e. travel reason and type of ticket) that led to interesting results.

Another aspect addressed in this paper is the drawback of using stated importance methods to identify the significance of each attribute. This type of method increases the length of the survey, yields insufficient differentiation among mean importance ratings, and may rate attributes as important even though they have little influence on SQ (Weinstein, 2000). In the case at hand, however, where the three most important variables in the survey were explored, a stated importance approach presents fewer drawbacks than asking about the importance of all the variables. Moreover, this study adds two more drawbacks to the literature: the possibility of attributes that are important for passengers and yet are not identified as such in the survey (i.e. INFORMATION in this study); and the impossibility of identifying differences between market segments on the basis of users' direct answers with regards to importance.

Our research may be useful to public transport planners in a number of ways. First, it helps clarify the factors that have a noteworthy impact on service quality, either overall or by user segments. Secondly, the factors are not the same for all passengers, and therefore each market segment requires different incentives (i.e. personalized marketing). The ability to differentiate between key service quality factors in the various market segments will enable transport planners to decide which users they want to engage, and plan their loyalty programs accordingly.

ACKNOWLEDGEMENTS

This study is sponsored by the Conserjería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) through the Excellence Research Project denominated “Q-METROBUS-Quality of service indicator for METROpolitan public BUS transport services”. The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

REFERENCES

Andreassen, T.W. (1995). (Dis)satisfaction with public services: The case of public transportation. Journal of Services Marketing, 9, 30-41.

Awasthi, A., Chauhan, S.S., Omrani, H., & Panahi, A. (2011). A hybrid approach based on SERVQUAL and fuzzy TOPSIS for evaluating transportation service quality. Computers and Industrial Engineering, 61(3), 37-46.

Berry, L.L., Zeithaml, V.A., & Parasuraman, A. (1990). Five imperatives for improving service quality. Sloan Management Review Summer 9-38.

Bordagaray, M., dell'Olio, L., Ibeas, A. & Cecín, P. (2014). Modelling user perception of bus transit quality considering user and service heterogeneity, Transportmetrica A: Transport Science, 10:8, 705-721.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1998). Classification and regression trees. Chapman & Hall/CRC.

Chou, P.F., Lu, C.S., & Chang, Y.H. (2014). Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2014.915247

Cirillo, C., Eboli, L., & Mazzulla, G. (2011). On the asymmetric user perception of transit service quality. International Journal of Sustainable Transportation, 5, 216-232.

Das, S. & Pandit, D. (2014). Determination of level-of-service scale values for quantitative bus transit service attributes based on user perception, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2014.910563

De Oña, J., de Oña, R., & Calvo, F.J. (2012). A classification tree approach to identify key factors of transit service quality. Expert Systems with Applications, 39, 11164-11171.

De Oña J., de Oña R., Eboli L., Mazzulla G. (2013). Perceived service quality in bus transit service: A structural equation approach. Transport Policy 29:219-226.

Del Castillo J.M. & Benitez F.G. (2013). Determining a public transport satisfaction index from user surveys. Transportmetrica A: Transport Science, 9:8, 713-741.

Dell'Olio, L., Ibeas, A., & Cecin, P. (2010). Modelling user perception of bus transit quality. Transport Policy, 17(6), 388-397.

Eboli, L., & Mazzulla, G. (2008). Willingness-to-pay of public transport users for improvement in service quality. European Transport, 38, 107-118.

Eboli, L., & Mazzulla, G. (2010). How to capture the passengers' point of view on a transit service through rating and choice options. Transport Review, 30, 435-450.

Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. Transport Policy, 18, 172-181.

Ganesan-Lim, C., Russell-Bennett, R., & Dagger, T. (2008). The impact of service contact type and demographic characteristics on service quality perceptions. Journal of Services Marketing, 22(7), 550-561.

Garrido. C.. de Oña. R. & de Oña. J. (2014). Neural networks for analyzing service quality in public transportation. Expert Systems with Applications. 41. 6830-6838.

Garver, M.S. (2003). Best practices in identifying customer-driven improvement opportunities. Industrial Marketing Management, 32(6), 455-466.

Kashani, A.T., & Mohaymany, A.S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science, 49, 1314-1320.

Transportation Research Board. (1999). A Handbook for Measuring Customer Satisfaction and Service Quality, TCRP Report 47.

Weinstein, A. (2000). Customer satisfaction among transit riders. How customer rank the relative importance of various service attributes. Transportation Research Record, 1735, 123-132.

Wong, J., & Chung, P. (2007). Managing valuable Taiwanese airline passengers using knowledge discovery in database techniques. Journal of Air Transport Management, 13(6), 362-370.

Figures

Tables

CHARACTERISTICS	STATISTICS
1.Gender	Male (32%), female (68%)
2.Age	18-30 (49%), 31-60 (40%), > 61 year-olds (11%)
3.Travel reason	Occupation (28%), studies (25%), doctor (11%), shopping (7%), holidays (6%), others (23%)
4.Use frequency	Almost diary (57%), frequently (22%), occasionally (13%), sporadically (8%)
5.Type of ticket	Consortium pass (67%), standard ticket (23%), senior citizen pass (7%), other ticket (3%)
6. Private vehicle available	Yes (47%), no (53%)
7. Complementary modes from origin to bus stop	On foot (77%), urban bus (18%), metropolitan bus (2%), private vehicle (1%), other mode (2%)
8. Complementary modes from bus stop to destination	On foot (95%), other mode (5%)

Table 1. Sample characteristics

VARIABLE	CATEGORIES
Sex	1.Male ; 2.Female
Age	1.{18-30}; 2.{31-60}; 3.{>60}
Travel reason	1. Occupation; 2. Studies; 3. Other
Use frequency	1. Frequent; 2. Sporadic
Type of ticket	1. Standard ticket; 2. Consortium Pass; 3. Senior Citizen Pass; 4. Other

Table 2. Segments of passengers analyzed

	Performance perceptions		**Importance frequencies**				
	Average Performance Rate	Standard Deviation	Option 1*	Option 2*	Option 3*	Sum	Overall (%)
ACCESIBILITY (accessibility to/from the bus)	7.28	2.03	85	125	90	300	9.2%
CLEANLINESS (bus interior cleanliness)	7.47	1.76	88	184	122	394	12.0%
COURTESY (driver courtesy)	7.91	1.84	115	149	123	387	11.8%
FARE (fare)	6.14	2.40	**603**	**534**	**468**	**1,605**	**49.1%**
FREQUENCY (frequency of service)	6.08	2.51	**618**	**617**	**465**	**1,700**	**52.0%**
INFORMATION (information)	6.43	2.31	159	80	120	359	11.0%
PROXIMITY (stops proximity to/from origin/destination)	7.10	2.21	125	135	159	419	12.8%
PUNCTUALITY (punctuality)	7.27	2.10	**804**	**451**	**322**	**1,577**	**48.2%**
SAFETY (safety on-board)	7.59	1.85	**369**	**329**	**288**	**986**	**30.1%**
SPACE (bus space)	7.09	2.01	93	123	144	360	11.0%
SPEED (speed)	6.97	2.09	**173**	**306**	**314**	**793**	**24.2%**
TEMPERATURE (bus temperature)	7.37	1.81	39	69	57	165	5.0%

Table 3. Service quality performance perceptions and importance frequencies for the overall market

	Stated Importance			Derived Importance		
Market segment	**Category**	**Variable**	**Frequency n. obs. (percentage)**	**Category**	**Variable**	**Normalized Importance**
OVERALL MARKET	OVERALL MARKET (max. obs. available 3,271)	FREQUENCY	1,700 (52.0%)	OVERALL MARKET (n. obs. 3,664; prec. rate 68.56%)	PUNCTUALITY	100.0%
		FARE	1,605 (49.1%)		TEMPERATURE	92.3%
		PUNCTUALITY	1,577 (48.2%)		INFORMATION	91.3%
		SAFETY	986 (30.1%)		FREQUENCY	86.0%
		SPEED	793 (24.2%)		SAFETY	70.3%
GENDER	FEMALE (max. obs. available 2,229)	FREQUENCY	1,201 (53.9%)	FEMALE (n. obs. 2493; prec. rate 65.26%)	PUNCTUALITY	100.0%
		PUNCTUALITY	1,084 (48.6%)		INFORMATION	70.6%
		FARE	1,074 (48.2%)		FREQUENCY	64.7%
		SAFETY	685 (30.7%)		TEMPERATURE	63.2%
		SPEED	521 (23.4%)		CLEANLINESS	57.7%
	MALE (max. obs. available 1,042)	FARE	531 (51.0%)	MALE (n. obs. 1171; prec. rate 71.98%)	INFORMATION	100.0%
		FREQUENCY	499 (47.9%)		PUNCTUALITY	90.4%
		PUNCTUALITY	493 (47.3%)		SAFETY	81.1%
		SAFETY	301 (28.9%)		FREQUENCY	78.6%
		SPEED	272 (26.1%)		COURTESY	72.2%
AGE	YOUNG (max. obs. available 1,599)	FREQUENCY	860 (53.8%)	YOUNG (n. obs. 1,809; prec. rate 64.73%)	PUNCTUALITY	100.0%
		FARE	852 (53.3%)		SAFETY	77.8%
		PUNCTUALITY	844 (52.8%)		SPEED	55.6%
		SPEED	457 (28.6%)		PROXIMITY	37.9%
		SAFETY	382 (23.9%)		COURTESY	34.5%
	MIDDLE (max. obs. available 1,326)	FREQUENCY	702 (52.9%)	MIDDLE (n. obs. 1479; prec. rate 68.15%)	FREQUENCY	100.0%
		FARE	654 (49.3%)		INFORMATION	99.9%
		PUNCTUALITY	598 (45.1%)		PUNCTUALITY	87.8%
		SAFETY	428 (32.3%)		SPACE	85.7%
		SPEED	264 (19.9%)		TEMPERATURE	79.8%
	ELDERLY (max. obs. available 346)	SAFETY	176 (50.9%)	ELDERLY (n. obs. 376; prec. rate 78.98%)	INFORMATION	100.0%
		FREQUENCY	138 (39.9%)		COURTESY	45.8%
		PUNCTUALITY	135 (39.0%)		PROXIMITY	37.5%
		FARE	99 (28.6%)		ACCESIBILITY	37.5%
		ACCESIBILITY	82 (23.7%)		SPEED	32.5%
FREQUENCY OF USE	FREQUENT (max. obs. available 2,541)	FREQUENCY	1,356 (53.4%)	FREQUENT (n. obs. 2870; prec. rate 68.29%)	PUNCTUALITY	100.0%
		FARE	1,303 (51.3%)		INFORMATION	94.2%
		PUNCTUALITY	1,262 (49.7%)		FREQUENCY	92.1%
		SAFETY	726 (28.6%)		SPEED	74.0%
		SPEED	620 (24.4%)		SAFETY	71.4%
	SPORADIC (max. obs. available 730)	FREQUENCY	344 (47.1%)	SPORADIC (n. obs. 794; prec. rate 69.77%)	PUNCTUALITY	100.0%
		PUNCTUALITY	315 (43.2%)		SAFETY	64.5%
		FARE	302 (41.4%)		CLEANLINESS	27.1%
		SAFETY	260 (35.6%)		COURTESY	26.3%
		SPEED	173 (23.7%)		TEMPERATURE	20.0%
TRAVEL REASON	OCCUPATION (max. obs. available 920)	FREQUENCY	496 (53.9%)	OCCUPATION (n. obs. 1027; prec. rate 65.86%)	FREQUENCY	100.0%
		PUNCTUALITY	472 (51.3%)		INFORMATION	93.4%
		FARE	442 (48.0%)		PUNCTUALITY	89.9%
		SAFETY	250 (27.2%)		SPACE	80.8%
		SPEED	198 (21.5%)		CLEANLINESS	80.7%
	STUDIES (max. obs. available 796)	FARE	457 (57.4%)	STUDIES (n. obs. 911; prec. rate 67.67%)	PUNCTUALITY	100.0%
		PUNCTUALITY	436 (54.8%)		TEMPERATURE	90.1%
		FREQUENCY	432 (54.3%)		SPACE	89.7%
		SPEED	241 (30.3%)		SAFETY	71.1%
		SAFETY	175 (22.0%)		INFORMATION	51.3%
	OTHERS (max. obs. available 1,555)	FREQUENCY	772 (49.6%)	OTHERS (n. obs. 1726; prec. rate 69.58%)	INFORMATION	100.0%
		FARE	706 (45.4%)		PUNCTUALITY	71.3%
		PUNCTUALITY	669 (43.0%)		CLEANLINESS	65.2%
		SAFETY	561 (36.1%)		SAFETY	61.2%
		SPEED	354 (22.8%)		COURTESY	51.6%
TYPE OF TICKET	STANDARD (max. obs. available 781)	PUNCTUALITY	408 (52.2%)	STANDARD (n. obs. 850; prec. rate 63.65%)	COURTESY	100.0%
		FREQUENCY	396 (50.7%)		SPEED	83.0%
		FARE	337 (43.1%)		PUNCTUALITY	76.0%
		SAFETY	274 (35.1%)		FREQUENCY	72.7%
		SPEED	207 (26.5%)		INFORMATION	71.8%
	CONSORTIUM PASS (max. obs. available 2,169)	FREQUENCY	1,166 (53.8%)	CONSORTIUM PASS (n. obs. 2445; prec. rate 67.81%)	PUNCTUALITY	100.0%
		FARE	1,160 (53.5%)		INFORMATION	85.9%
		PUNCTUALITY	1,058 (48.8%)		SAFETY	82.6%
		SAFETY	569 (26.2%)		TEMPERATURE	64.3%
		SPEED	524 (24.2%)		SPACE	62.8%
	SENIOR CITIZEN PASS (max. obs. available 232)	SAFETY	127 (54.7%)	SENIOR CITIZEN PASS (n. obs. 249; prec. rate 79.12%)	INFORMATION	100.0%
		FREQUENCY	91 (39.2%)		ACCESIBILITY	82.8%
		PUNCTUALITY	82 (35.3%)		SPACE	61.4%
		ACCESIBILITY	65 (28.0%)		CLEANLINESS	58.1%
		FARE	62 (26.7%)		COURTESY	55.5%
	OTHER (max. obs. available 89)	FREQUENCY	45 (50.6%)	OTHER (n. obs. 120; prec. rate 65.00%)	SPEED	100.0%
		FARE	39 (43.8%)		ACCESIBILITY	43.5%
		INFORMATION	25 (28.1%)		PUNCTUALITY	23.8%
		PUNCTUALITY	23 (25.8%)		SPACE	22.8%
		SAFETY	22 (24.7%)		FREQUENCY	20.8%

Table 4. Stated Importance versus Derived Importance by market segment

Figure 1. CART for the metropolitan public transport in Granada (Spain). Overall market.

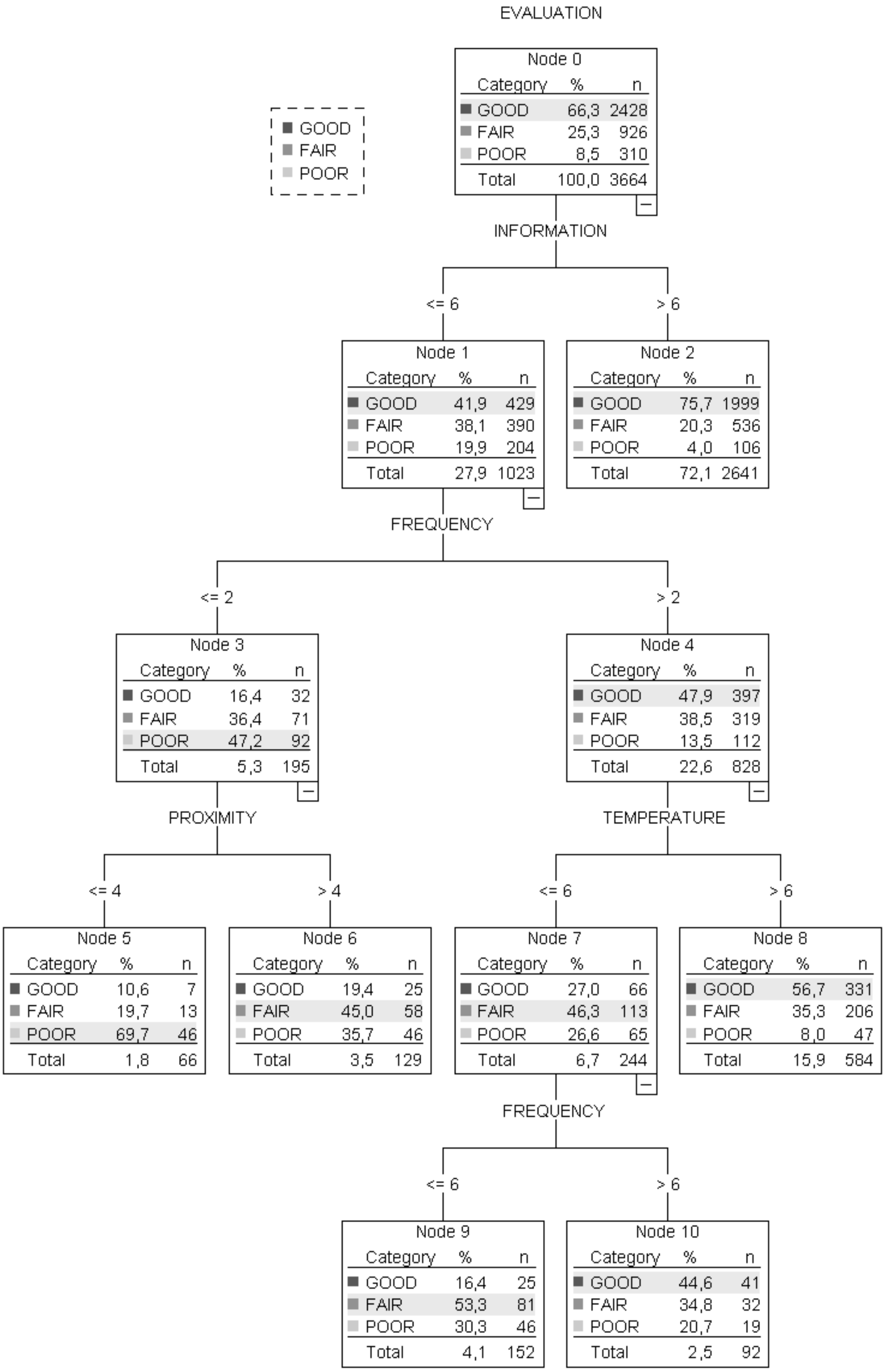

Figure 2. CART for users classified according to the gender (Male)

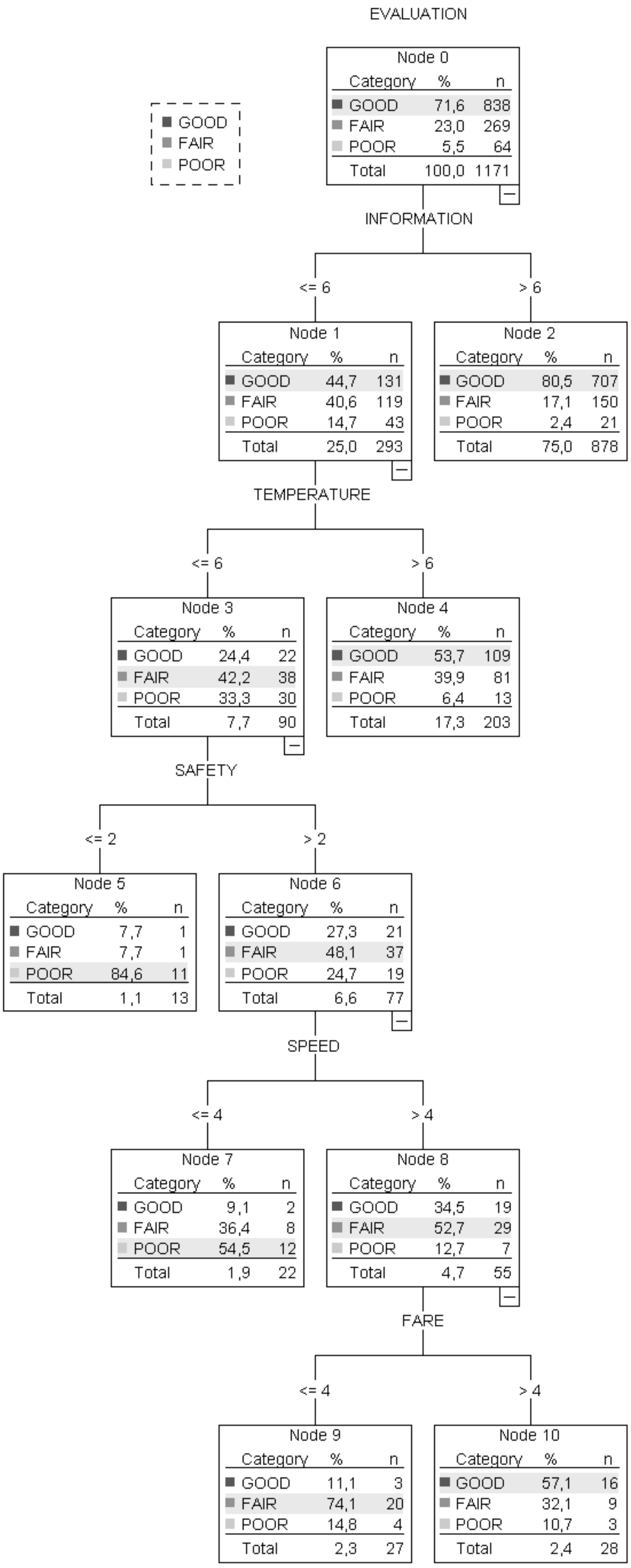

Figure 3. CART for users classified according to the age (Elderly)

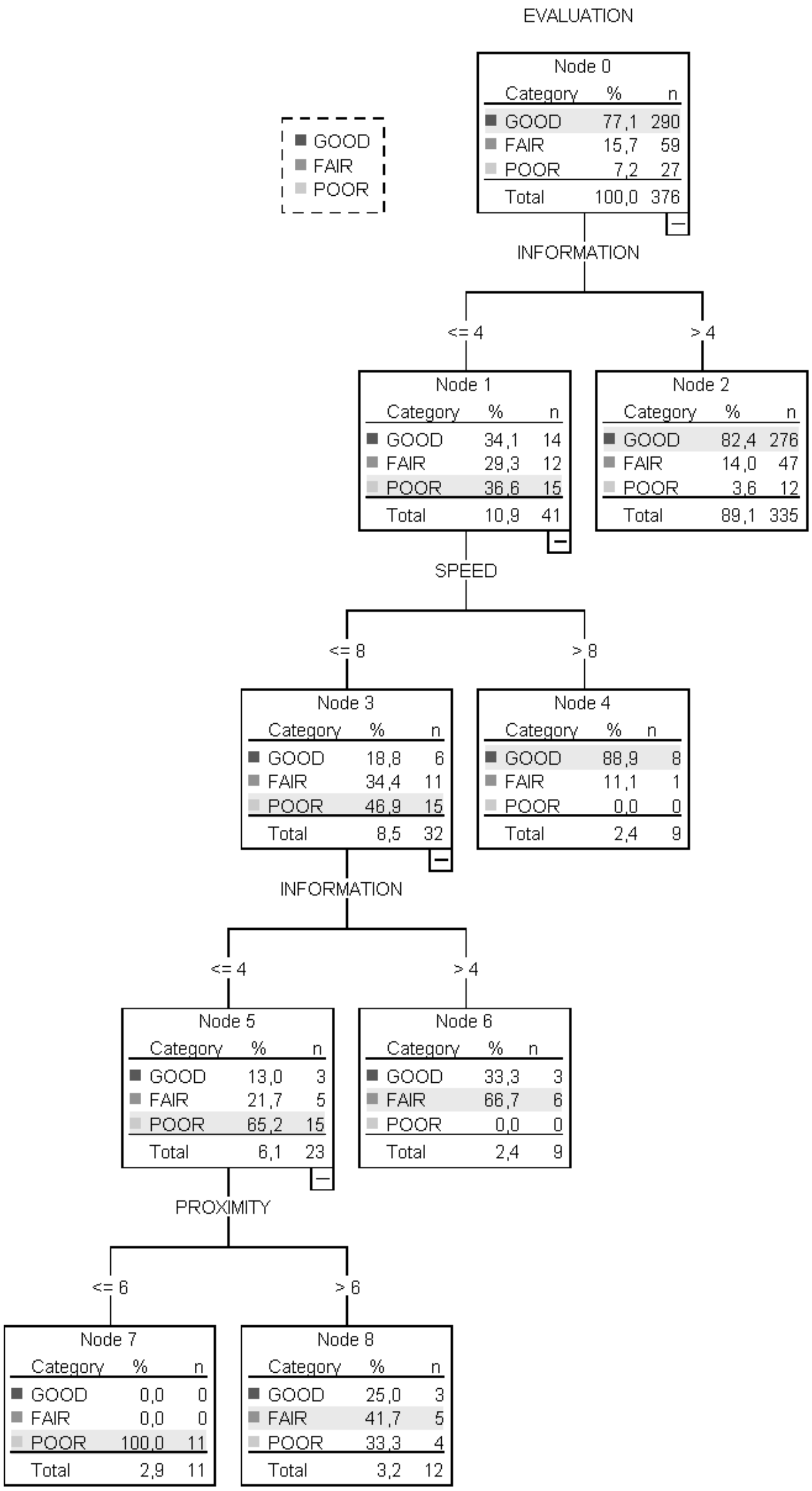

Figure 4. CART for users classified according to the frequency of use (Frequent passengers)

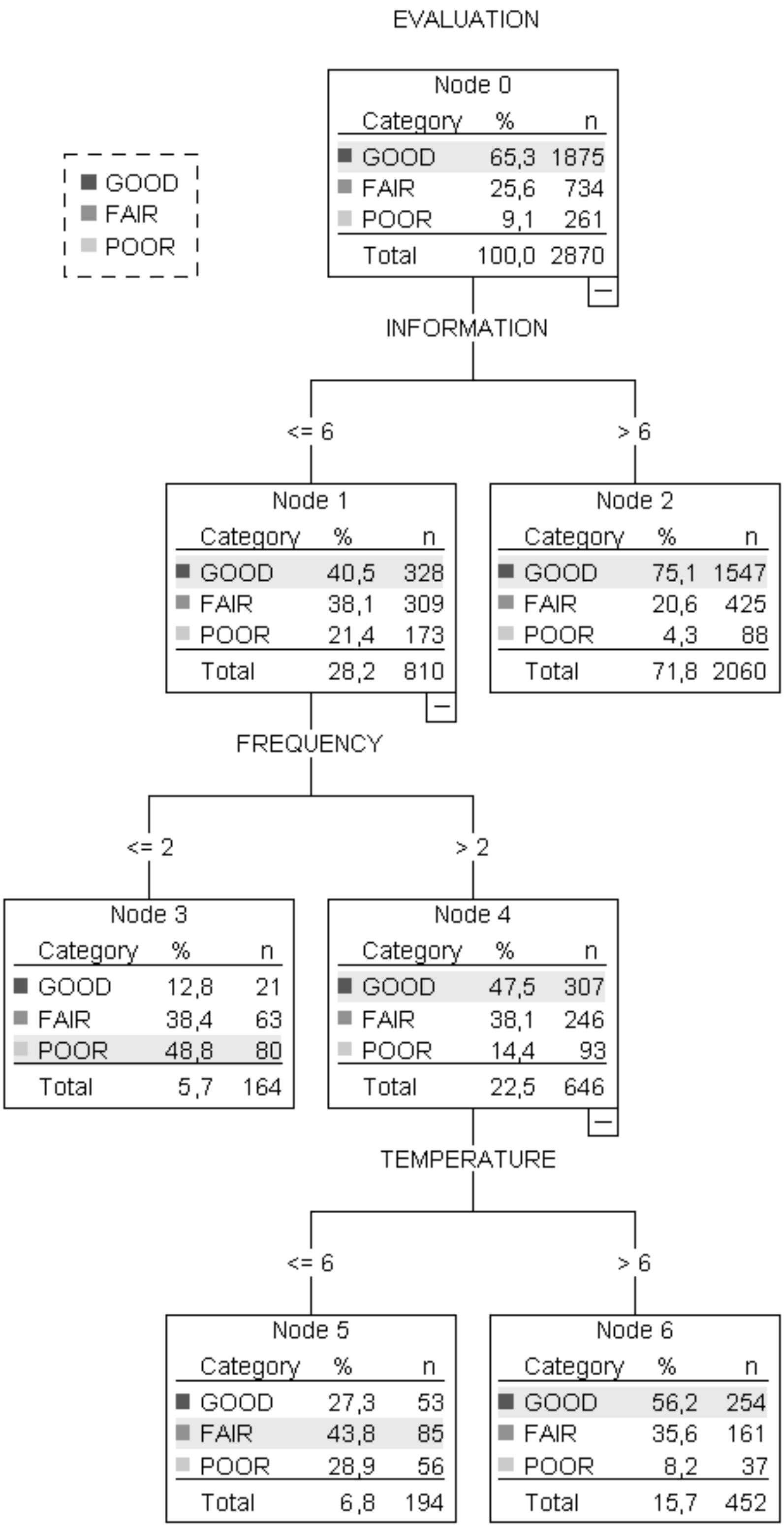

Figure 5. CART for users classified according to the travel reason (Studies)

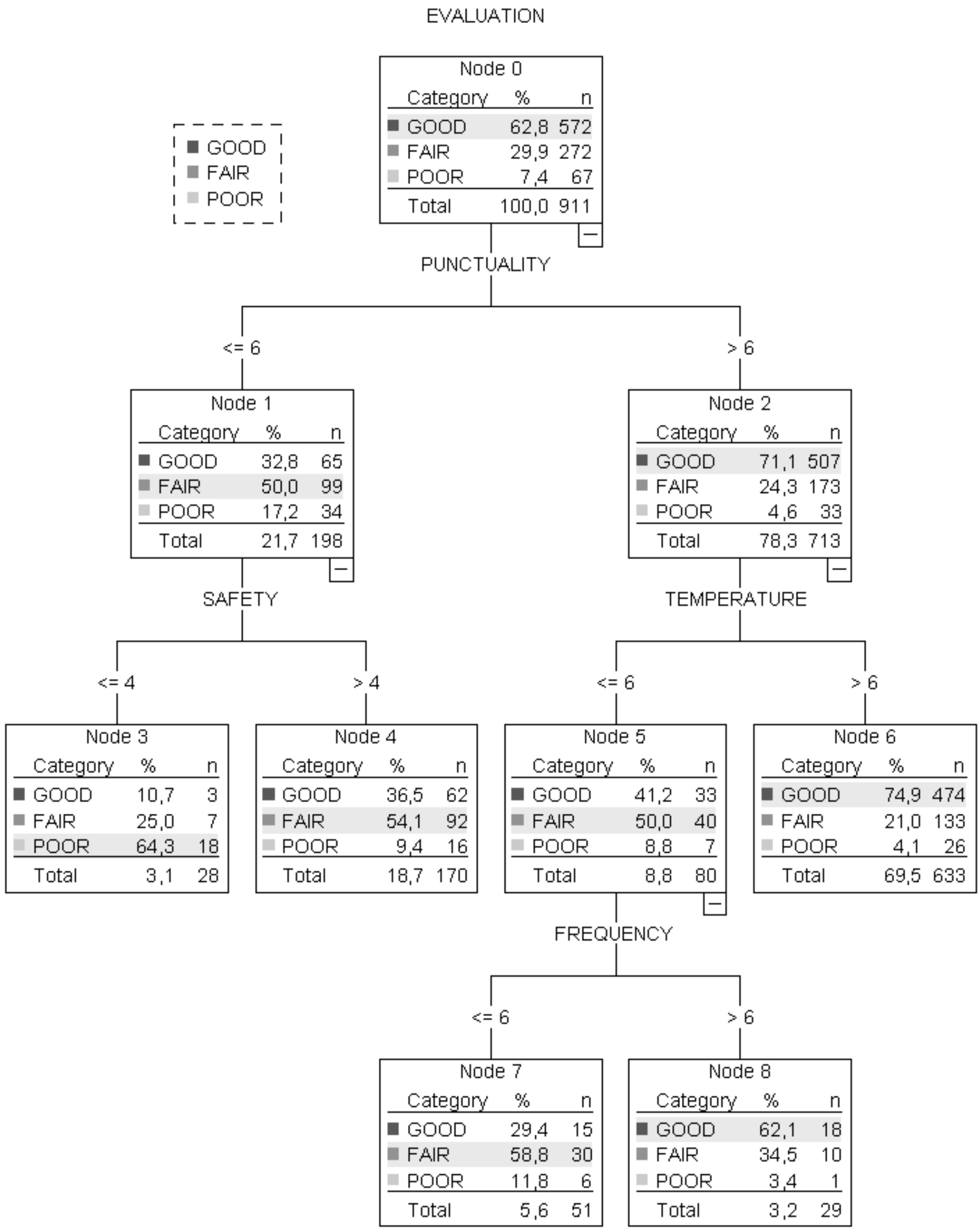

Figure 6. CART for users classified according to the type of ticket (Senior citizen pass)