

The background of the cover is a grayscale aerial photograph of a boat moving across a body of water, leaving a long, white wake behind it. The boat is positioned in the upper left quadrant, moving towards the center. The water's surface is textured with ripples and the white foam of the wake. At the top of the page, there is a decorative horizontal band with a wavy, wavy pattern.

Final Working Paper

Market Segmentation for Ridership Forecasting

Prepared by:

Cambridge Systematics

Mark Bradley Research & Consulting

March 27, 2002

Prepared for:

WATER TRANSIT AUTHORITY



Table of Contents

Introduction	1
Identifying Traveler Attitudes	2
Overview of Factor Analysis	2
Exploratory Factor Analysis	2
Confirmatory Factor Analysis	4
Forecasting Traveler Attitudes	5
Overview of Attitude Models	5
Socioeconomic Data	5
Logistic Regression Analysis	7
Combined Factor and Relationship Models	9
Market Segmentation Models	19
Overview of Cluster Analysis	19
Exploratory Cluster Analysis	20
Confirmatory Cluster Analysis	23
Characteristics of Market Segments	28

List of Tables

1.	Attitudinal Factors and Variables in the Structural Equation Modeling (SEM)	11
2.	Structural Equation for the Desire to Help the Environment Factor	13
3.	Structural Equation for the Need for Time Savings Factor	14
4.	Structural Equation for the Need for Flexibility Factor	15
5.	Structural Equation for the Sensitivity to Stress Factor.....	16
6.	Structural Equation for the Sensitivity to Transport Cost Factor	17
7.	Structural Equation for the Sensitivity to Personal Travel Experience Factor	18
8.	Squared Multiple Correlations for Traveler Factors in the Structural Equation Model.....	19
9.	Statistics for All Attitudinal Factors.....	21
10.	Statistics with Various Cluster Numbers	21
11.	Statistics for Variables	22
12.	Cluster Summary	22
13.	Average Factor Scores for Clusters	23
14.	Summary of Market Segment Characteristics	34

List of Figures

1.	SEM Model Structure	10
2.	Final Market Segmentation	24
3.	Size of the Market Segments	25
4.	Average Factor Scores for each Market Segment.....	26
5.	Market Segments by Age Group	29
6.	Market Segments by Household Size	29
7.	Market Segments by Number of Children.....	30
8.	Market Segments by Household Income Category	30
9.	Market Segments by Vehicles Available	31
10.	Market Segments by Workers per Household	31
11.	Market Segments by Gender.....	32
12.	Market Segments by Education Level	33
13.	Market Segments by Marital Status	33

Market Segmentation for Ridership Forecasting

Introduction

The market analysis of potential ferry riders is based on data collected in the fall of 2001 on random sample of people in the Bay Area who travel in corridors that are either currently served by ferries or may be served by ferries in the future. The market analysis is conducted in three steps:

- Identifying Traveler Attitudes (Factor Analysis) – To identify the latent variables that can be used to segment the respondents while retaining the explanatory power of each manifest attitudinal statement.
- Forecasting Traveler Attitudes (Relationship Models) – To relate the socioeconomic and demographic variables to the factors identified in the factor analysis and provide forecasting abilities.
- Segmenting the Traveler Market (Cluster Analysis) – To identify the attitudinal-based market segments of the population based on results from factor analysis.

We adopted a Structural Equation Modeling (SEM) technique in the market research. SEM serves purposes similar to multiple regression analysis, but in a more powerful way which takes into account the modeling of interactions, correlated independents, measurement error, correlated error terms, and multiple latent independents. The power and precision of data analysis for SEM is better than with traditional methods of analysis, because many variables (both manifest and latent variables) may easily be brought into the analysis, and errors of measurement may be controlled for.

This report documents the results of the market segmentation analysis, as conducted in stages, as well as the final results from the structural equation modeling. The primary advantage of the SEM technique over the traditional methods is the fact that one or more of the analyses can be conducted simultaneously rather than sequentially. Due to limitations in the current model estimation software, the structural equation modeling was used to simultaneously estimate the factors and relationships to the socioeconomic data, but the cluster analysis and mode choice models were estimated separately. A software package called AMOS was used to estimate the structural equation models, and SAS was used to estimate the cluster analysis models. The mode choice models are documented in the next section and were estimated using the ALOGIT software.

Identifying Traveler Attitudes

Overview of Factor Analysis

A factor analysis step is performed to analyze the 30 attitudinal variables collected from the sample of potential ferry riders. Factor analysis involves a statistical procedure that transforms a number of possibly correlated variables into a smaller group of uncorrelated variables called principal components or factors. The objectives of performing factor analysis are to reduce the number of variables (data reduction) and to classify variables so as to detect the structural relationships between variables (structure detection).

There are two types of factor analysis we conducted for this market analysis: exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis is a process where the statistics of the data determine the structure and content of the resulting factors. It is used to explore the survey data to determine the nature of factors that account for the co-variation between variables without imposing any a priori hypothesis about the number and structure of factors underlying the data. Confirmatory factor analysis is a process where we apply judgment regarding the structure and content of the factors and then estimate the statistical results of these established factors. It is based on the results from exploratory factor analysis as well as theoretical hypotheses as to which variables are correlated with which factors. In this report, we describe the exploratory factor analysis and the confirmatory factor analysis. We recommend that the confirmatory factor analysis results are the recommended factors for use in the remainder of the market analysis because the confirmatory factor analysis offers a more viable method for evaluating construct validity of the attitudinal survey; it also enables us to explicitly test hypotheses concerning the factor structure of the data, and provides more logical and consistent factors that may provide better information for the cluster analysis.

Exploratory Factor Analysis

There are two classical methods to determine the number of factors to be extracted in the exploratory factor analysis. The most widely used method is Kaiser criterion, according to which, we only extract factors when it explains at least as much as the equivalent of one original variable. Another one is a screen test method proposed by Cattell. It plots the eigenvalues in a simple line plot, from which we find the point where the smooth decrease of eigenvalues appears to level off to the right of the plot. In this study, we would retain 9 factors using the Kaiser criterion, and 7 factors using the screen plot. Since Kaiser criterion tends to retain too many factors, we used the Cattell screen plot method to decide the number of factors to be extracted for this analysis. This resulted in the following seven factors:

- **Factor One** includes eight variables related to *desire to help the environment and preference for walking, ferry, and public transportation* as appose to private automobile driving. This factor accounts for as much as 12.7 percent of the total

variance in the data set. It reflects respondents' willingness to help improve the environment by changing their travel behaviors (Q44, Q45, Q48). It also shows that respondents have a preference for walking (Q42, Q49), ferry (Q51), and public transportation (Q50, Q53). All variables except Q50 have positive loading on the first factor. Negative sign indicates that the statement "I don't feel safe on public transportation" (Q50) has a negative contribution to the overall score for factor one.

- **Factor Two** corresponds to *travelers' sensitivity to the time* spent on traveling. It explained 8.87 percent of the total variance of the 30 attitudinal variables. Q25, Q26, and Q27 indicate that travelers would like to arrive at their destinations as fast as they can, even if the travel cost is higher (Q25), form of travel is less convenient (Q26), or the traveler has to put up with crowds on his trip. These people are usually in a hurry when they make a trip (Q32) or they have a very rigid schedule (Q33).
- **Factor Three** is a dimension shows *travelers' preference for driving or taking the ferry*. This factor only includes three variables (Q43, Q46 and Q52), but it explained an additional 7.27 percent of the total variance. Q43 states the preference to drive than to be driven, and Q46 shows the preference to "make trips alone because I like the time to myself." These respondents most likely have preference for driving over public transportation. Q52 further indicates, "I would ride a ferry, but I wouldn't ride the bus." So taking ferry is another travel mode choice other than driving; but public transportation is not considered in choosing a model.
- **Factor Four** corresponds to the need for flexibility, explaining another 6.06 percent of the total variance. Q34 is a statement that says "generally I make the same types of trips at the same times of the day," so travelers are not flexible in choosing travel time and trip types. But this attitudinal variable has a negative loading on factor four, so the flexibility of travel time and trip type actually contributes positively to factor four. Q36 and Q37 show that travelers "would make trips to a wide variety of locations each week" or "need to have to flexibility to make many trips during the day" respectively, so they have positive loadings.
- **Factor Five** is an attitudinal dimension reflecting respondents' *sensitivity to travel stress*. It explained 5.12 percent of the total variance. All the six variables included in this factor (Q27, Q30, Q35, Q39, Q40 and Q41) are related to the stress on travel.
- **Factor Six** includes three variables representing travelers' *easy-going attitude*, explaining 4.35 percent of the total variance. Q29 indicates that a traveler would make a longer trip if he could be productive while traveling, Q31 suggests that he "wouldn't mind traffic congestion if it is predictable from day to day," and Q38 shows that he "don't mind delays as long as I am comfortable."
- **Factor Seven** is a factor that corresponds to the *sensitivity to travel cost*, including two variables Q47 and Q54, which explain 4.06 percent of the total variance. Q47 says the traveler would "use the most convenient form of transportation regardless of cost," and Q54 suggests that "the worst thing about driving to work is the lack of affordable

parking.” Since the loading of Q47 is negative, and Q54 is positive, both the variables suggest the importance of travel cost in the choice of travel behaviors.

Total variance explained by the exploratory factor analysis is 48.3 percent.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is conducted using the SAS CALIS procedure. CFA aims to verify the specified dimensions of latent variables, which serve as the important underlying decisive factors for travel mode choice.

- **Factor One** includes three variables related to *desire to help the environment*. It reflects respondents’ willingness to help improve the environment by paying more (Q44) or changing their travel mode (Q45). People in the group also believe that use transit can help improve environment (Q48).
- **Factor Two** corresponds to *travelers’ need for time savings*. This factor indicates that travelers would like to arrive at their destinations as fast as they can, even if they have to change the form of travel (Q26), or put up with crowds on their trips (Q28). These people are usually in a hurry when they make a trip (Q32). They would mind walking a few minutes to get to their destination (Q49, negative loading). This factor also indicates that travelers would take the fastest route to their destination even if they have a cheaper alternative (Q25)
- **Factor Three** represents the *need for flexibility*. Both the statements “I would make trips to a wide variety of locations each week” (Q36) and the statement “I need to have the flexibility to make many trips during the day” (Q37) shows the need for travel flexibility in scheduling, so they have positive loadings. But Q34 is a statement that “I make the same types of trips at the same times of the day” and Q33 is a statement that “I need to make trips according to a very rigid schedule,” so they have negative loadings on this factor.
- **Factor Four** is an attitudinal dimension reflecting respondents’ *sensitivity to travel stress*. All the six variables included in this factor (Q35, Q41, Q40, Q39 and Q27) are related to the stress on travel.
- **Factor Five** is a factor that corresponds to the *insensitivity to transport cost*. This is based on traveler’s insensitivity to transport cost and the anchor variable is set to be Q47, which says travelers will use the most convenient form of transportation regardless of cost. Q54 suggests, “The worst thing about driving to work is the lack of affordable parking.” Since the loading of Q47 is set to be 1, the loading of Q54 is negative. Other variables (Q25 and Q44) also suggest the importance of travel cost in the choice of travel behaviors.
- **Factor Six** is a dimension that shows travelers’ *sensitivity to personal travel experiences*. These respondents show a preference for one mode over another. In

some cases, this is a preference for driving alone (Q43 and Q46). In other cases, it is a preference for taking a ferry over a bus (Q52) or a preference for transit in general (Q53). This personal travel experience factor also shows traveler's attitude towards walking (Q42) and riding ferry (Q51).

In each factor, the loading of one variable (called "anchor variable") is fixed to be constant (1.0) based on one's expectations. It is served to scale the loadings of all other variables in the same factor. The t-value shows the degree of statistical significance for the relative contribution of each attitudinal variable to the corresponding factor.

CFA uses analysis of Goodness of Fit Index (GFI) to analyze whether the data confirms the theoretical construct. The Goodness of Fit Index is a measure of the relative amount of variances and co-variances jointly accounted for by the model. It can be thought of as roughly analogous to the R² in multivariate regression. The closer the GFI is to 1.00, the better is the fit of the model to the data. Generally, GFI values should be equal to or greater than .90 to indicate an adequately fitting model. The GFI in this work is 0.8966, indicating that 89.66 percent of the co-variation in the data can be reproduced by the given model. Thus, the confirmatory factor analysis supports the six-dimension attitudinal construct.

Forecasting Traveler Attitudes

Overview of Attitude Models

Attitude models provide a means to forecast traveler attitudes based on existing demographic data that is readily available through the MTC regional travel demand forecasting model. This ensures that we will be able to forecast the traveler factors that were estimated using the factor analysis and in turn forecast the market segments that are important for modal choice analysis. The relationship between the available demographic data and the factors is statistically challenging using traditional methods because we hypothesize that the socioeconomic data available in the current travel demand forecasting model are not necessarily the best variables for forecasting specific traveler attitudes. For example, there is no employment status variable available in the MTC model (full-time, part-time, retired), which would likely be a significant variable in forecasting traveler sensitivity to scheduling. As a result, the linear and logistic regression models developed from these data to forecast traveler attitudes do not provide reliable statistical results. Fortunately, the statistical results of these relationships improved when structural equation modeling was employed.

Socioeconomic Data

The socioeconomic data requirements are that variables exist in both the MTC model and the stated-preference household survey. In some cases, data processing was necessary to provide compatibility. These data are as follows:

- Average household size (HHSIZE) is available as a continuous variable.
- Number of persons under age 18 (HHSU18) per household is available as a continuous variable.
- Number of workers per household (HHWORK) is available as both a continuous variable and a categorical variable (0, 1, and 2+ workers).
- Number of vehicles per household (HHVEH) is available as both a continuous variable and a categorical variable (0, 1 and 2+ vehicles).
- Age is available as a categorical variable (18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+).
- Household income is available as a continuous variable (average) and as a categorical variable (<\$24,999, \$25-49,999, \$50-74,999, \$75,000+). There were more income categories in the stated-preference household survey; these were re-coded to match the income categories in the MTC model.
- Number of college students per household is available as a continuous variable. College students were estimated from the employment status variable in the stated-preference household survey.
- Households with more workers than vehicles are generated from the cross-classification of workers per household and vehicles per household as a dichotomous variable (household with more workers than vehicles, households with same or less workers than vehicles).

There are additional variables in the stated-preference household survey that were used to describe market segments, but were not available in the MTC household survey and were therefore not included in the relationship models.

- Education level (EDUC) is available as a categorical variable (grade 1-11, high school, some college, college grad, post graduate).
- Marriage status (MARSTAT) is available as a categorical variable (single, married/cohabitating, divorced/separated, widowed).
- Job status (EMPSTAT) is available as a categorical variable (full-time, part-time, retired, student).
- Gender (GENDER) is available as a dichotomous variable (male, female).

There were 3,404 cases in the stated-preference household survey, of which 252 cases are excluded from the following logistic regression because of missing data.

Logistic Regression Analysis

About Logistic Regression

We dummy coded the scores of the six factors that were derived from confirmatory factor analysis. Each dummy variable will have the value of 0 when the corresponding factor score is negative and 1 when the factor score is positive. Then we tried to fit logistic regression models to these dummy factors and the socioeconomic as well as demographic variables.

Proc Logistic in SAS does logistic regression analysis. It is often used to investigate the relationship between binary or discrete responses and a set of explanatory variables. It fits linear logistic regression models for binary or ordinal response data by the method of maximum likelihood. In this analysis, the logit link function is defined to be $\log(p/1-p)$ where p is the probability of the factor score to be positive ($D_i = 1$, where $i = 1, 2, 3, 4, 5, 6$).

Logistic Model Results

For the logistic model estimate, it provides detailed analysis on each of the variables that remain in the stepwise logistic regression analysis. It reports the parameter estimate, standard error of the estimate, the chi-square value, its p-value and the odd ratio. If the p-value for a variable is very small (say less than 0.05), we will reject the null hypothesis that the parameter is zero and conclude that the corresponding variable is statistically significant in the analysis.

For factor 1 (desire to help the environmental), there are only two significant socioeconomic variables being included in the model: household vehicle number and college student. It suggests that households with more vehicles tend to care less about environment; and being a college student has negative impacts on environment.

For factor 2 (need for time savings), there are five significant variables remaining in the model. Households with more children, more workers, and higher income have positive coefficients on the need for time savings, while households with more workers than cars have a negative coefficient. The age variable also has a negative coefficient, indicating that there is a less need for time savings as people grow older. Factor 3 (need for flexibility) has three significant explanatory variables: children number and vehicle number in household are positively related to the need for flexibility, while the worker number in household has a negative coefficient.

For factor 4 (sensitivity to stress), household worker number stands out to be the only variable with a positive coefficient. Generally speaking, household with more workers are more sensitive to stress. But if a household has more workers than cars, it tends to reduce the stress sensitivity, as indicated by the negative coefficient of the variable of “more workers than cars.” The negative coefficients for the rest of variables suggest that people with higher income, old age group, and college students have less sensitivity to stress.

For factor 5 (insensitivity to transportation cost), the coefficients for household size and income are estimated to be positive, meaning that larger households and higher income households care less about transportation cost. On the other side, household children number and worker number as well as college student are negative explanatory variables for the cost factor, meaning that transportation cost is an important decision factor for households with more children and more workers, and for college students as well.

For factor 6 (Sensitivity to personal travel experience), household worker number is the only variable with a negative coefficient. Children number and vehicle number in households are positive; so is the age variable.

The probability of positive factor score is defined to be p , while the probability of negative factor score is $(1-p)$. The odd ratio of positive factor score is $p/(1-p)$, while the odd ratio of negative factor score is $(1-p)/p$. In the model estimate, the odd ratio is computed as the exponential of the estimated parameter. It gives the amount of change expected in the odds ratio when there is a one unit of change in the exploratory variable with all other valuables in the model being held constant. For example, in the logistic model of factor 2 (need for time savings), the odd ratio for children number in household is 1.179, that is, for every unit change in children number, the odds for the score of factor 2 to be positive ($D_2 = 1$) is 1.179 times greater than the odds for the score of factor 2 to be negative ($D_2 = 0$). In the same model, the odd ratio for “more workers than cars in household” is 0.757. Take the reciprocal of 0.757, we get $1/0.757=1.321$. So we can predict that the odds of having a negative score in factor 2 for household with more workers than cars are 1.321 times greater than the odds of having a positive score. *The global null hypothesis of BETA being zero for each model is also estimated.* The interpretation of p-value is the same as for the likelihood-ratio test. In our case, the p-value of score statistic for each model is less than 0.0001. So we conclude that all the six estimated models are statistically significant.

The logistic modeling results also present the associations for assessing the predictive ability of a model. They are rank correlation of observed responses and predicted probabilities. A pair of observations with different responses is concordant if the observation with the response 1 has the higher predicted event probability, while a pair of observations with different responses is said to be discordant if the observation with the response 1 has the lower predicted event probability. If a pair of observations with different responses is neither concordant nor discordant, it is a tie. For example, observation with response 1 and the predicted probability 0.4 will be a tie for an observation with response 0 and the predicted probability 0.4. Paired is to measure the total number of distinct pairs. Somer’s D index is used to determine the strength and direction of relation between pairs of variables. Its values range from -1.0 (all pairs disagree) to 1.0 (all pairs agree). The Sonoma’s D indices are positive for all six factor models. The Gamma index does not penalize for ties on either variable. Its values range from -1.0 (no association) to 1.0 (perfect association). Because it does not penalize for ties, its value will generally be greater than the values for Somer’s D. Kendall’s Tau-a is a modification on Somer’s D to take into the account of the difference between the number of possible paired observations and the number of paired observations with different response. Usually Tau-a is much smaller than Somer’s D since there would be many paired observations with the same response. C index is another measure of rank

correlation of ordinal variables. It justifies so that it ranges from 0 (no association) to 1 (perfect association). It is a variant of Somer's D index. The c indices for the six factors range from 0.538 for factor 1 (desire to help the environment) to 0.604 for factor 2 (Need for time saving).

Combined Factor and Relationship Models

Overview of Structural Equation Modeling

Structural equation modeling (SEM) is a modeling technique that enables us to test a set of linear equations simultaneously. It is used in this study to identify the structural attitudes of travel behaviors and to quantify the causal relationships between travelers' socioeconomic status or demographic profile and travel attitudes. The primary objective of structural equation modeling is to improve the statistical reliability of the relationship between the socioeconomic data and the estimation of factors. This process modifies the attitudinal variables in each factor in order to improve the forecasting abilities of the model.

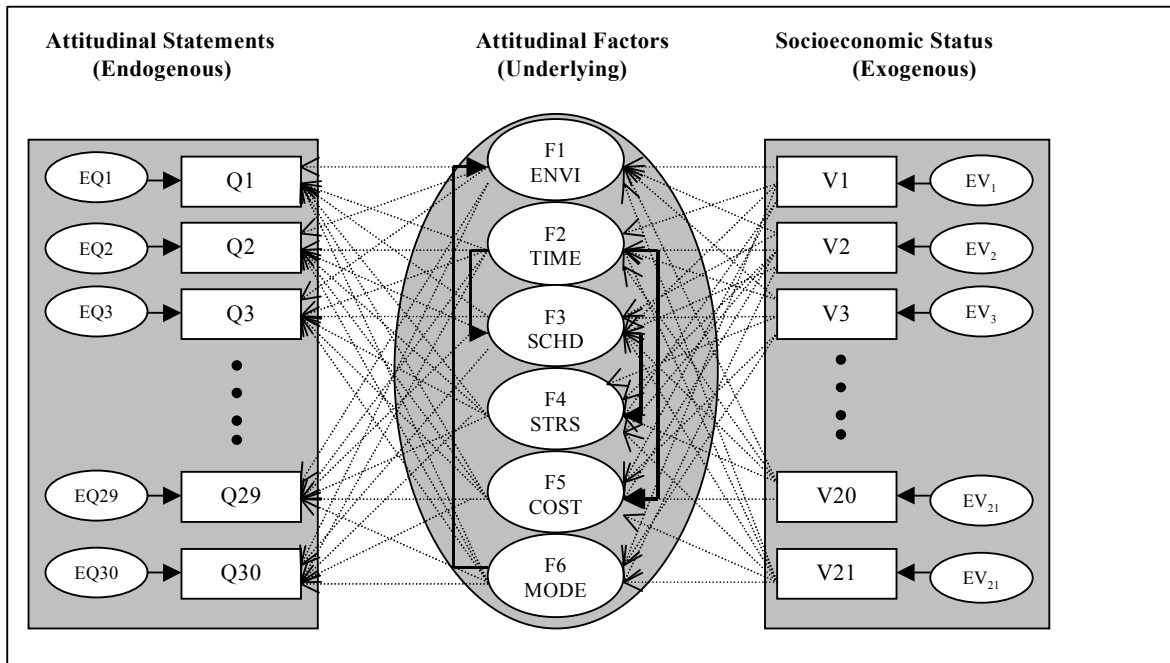
There are two types of variables used in the SEM:

- **Manifest variables** are observed variables that are directly measured from household surveys. In this study there are two main groups of manifest variables: 1) 30 attitudinal variables, which are the ratings that indicate travelers' certain attitude toward travel, and 2) socioeconomic and demographic variables such as household size, household income, vehicle ownership etc.
- **Latent variables** are unobserved variable that are not directly measured but are inferred by the relationships or correlations among manifest variables in the analysis. There are also two groups of latent variables in the SEM: 1) six attitudinal factors representing the most important attitudinal dimensions for traveler behaviors and 2) error terms associated with each variable involved in the SEM model.

Conceptually, every variable has an associated measurement error. So the SEM model includes an error term for each associated variable.

The SEM model is constructed in AMOS 4.0, software developed by Small Water, Inc. AMOS uses path diagrams to represent relationships among manifest and latent variables. Ovals or circles represent latent variables, while rectangles or squares represent manifest variables (see Figure 1). Single-headed arrows in the path diagram represent causal effects.

Figure 1. SEM Model Structure



In the SEM model structure, people's socioeconomic and demographic statuses are regarded as exogenous variables, while the ratings of attitudinal statements are endogenous variables. An SEM structural model is used to capture the causal influences of the exogenous variables on the endogenous variables through sets of underlying attitudinal factors. In order to do so, there are three basic sets of simultaneous equations being estimated concurrently in the SEM:

- Functions between attitudinal factors and socioeconomic, demographic variables.** The attitudinal latent variables are specified as linear equations of observed social economic and demographic variables, which act as the indicators of the underlying attitudinal structure toward travel. The socioeconomic and demographic variables include household size, number of children (under 18) in the household, household vehicle ownership, household worker number, age information, income level, college student, and household worker number compared with vehicle number. All the categorical variables are dummy coded, resulting in 21 socioeconomic or demographic variables being built into the model.
- Functions between ratings on attitudinal statements and underlying attitudinal factors.** Each latent factor is associated with multiple attitudinal statements through a confirmatory factor analysis structure. SEM is a confirmatory rather than exploratory modeling method because the modeler predetermines the model structure. If the modeler assumes no direct relationship between an attitudinal factor and an attitudinal statement, the path coefficient in the diagram is set to be 0. For each attitudinal factor, there is one and only one path coefficient being fixed to be 1. This is

the anchor variable that is used to set the scale of measurement for the latent factor and residuals. The SEM model estimates all other path coefficients.

- **Functions between the latent variables.** There are six attitudinal factors being constructed in the model: desire to help the environment, need for time savings, need for flexibility, sensitivity to travel stress, insensitivity to transport cost, and sensitivity to personal travel experience. The causal influences of latent variables upon one another are also represented as linear equations in the SEM. There are four pairs of causal relationships being modeled: 1) need for flexibility as a function of need for time savings, 2) sensitivity to travel stress as a function of need for flexibility, 3) insensitivity to transport cost as a function of need for time savings, and 4) desire to help the environment as a function of sensitivity to personal travel experience.

All the linear equations in SEM are estimated simultaneously. Attitudinal factors scores are calculated using the estimated coefficients for function set 1 (i.e., functions between attitudinal factors and socioeconomic, demographic variables). The factor scores are then used in market segmentation.

The results of the SEM process are a final series of traveler factors that are estimated simultaneously with the demographic data required to estimate these factors. Table 1 presents the attitudinal factors and variables in the SEM process, with statistics on standard error (Std Error) and significance (t-value) for each variable.

Table 1. Attitudinal Factors and Variables in the Structural Equation Modeling (SEM)

Factor/ Variable	Variable Statements	Coefficient	Std Error	t-value
Factor One	Desire to help the environment			
PAYENVIR	I would be willing to pay more when I travel if it would help the environment.	1.000		
MODENVIR	I would switch to a different form of transportation if it would help the environment.	0.949	0.028	33.447
TRNENVIR	Use of transit can help improve the environment.	0.376	0.018	20.887
Factor Two	Need for time savings			
CHANGMOD	I would change my form of travel if it would save me some time.	1.000		
HURRY	I am usually in a hurry when I make a trip.	0.911	0.023	39.283
FASTEST	I always take the fastest route to my destination even if I have a cheaper alternative.	0.760	0.024	32.044
NOSTRESS	Having a stress-free trip is more important than reaching my destination quickly.	-0.680	0.030	-22.978
CROWDSOK	I'll put up with crowds if it means I'll get to my destination quickly.	0.657	0.020	32.082

Table 1. Attitudinal Factors and Variables in the Structural Equation Modeling (SEM) (continued)

Factor/ Variable	Variable Statements	Coefficient	Std Error	t-value
COMFORT	I don't mind delays as long as I am comfortable.	-0.511	0.021	-23.848
DLDRIVE	I don't like to drive but it is usually the fastest way to get where I need to go.	0.418	0.025	16.791
Factor Three	Need for flexibility			
VARIETY	I need to make trips to a wide variety of locations each week.	1.000		
NEEDFLEX	I need to have the flexibility to make many trips during the day if necessary.	0.841	0.031	27.555
REGULAR	Generally, I make the same types of trips at the same times of the day.	-0.489	0.023	-21.654
Factor Four	Sensitivity to Travel Stress			
ANXIOUS	I am usually anxious and unsettled by the time I reach my destination.	1.000		
NOSTRESS	Having a stress-free trip is more important than reaching my destination quickly.	1.106	0.080	13.883
BRIDGES	Driving on the bridges across the bay is stressful for me.	1.266	0.074	17.099
STRESSFL	I avoid making certain trips at certain times because it is too stressful to make the trip.	1.519	0.080	19.059
Factor Five	Insensitivity to transport cost			
CONVENNT	I use the most convenient form of transportation regardless of cost.	1.000		
PAYENVIR	I would be willing to pay more when I travel if it would help the environment.	0.248	0.032	7.797
FASTEST	I always take the fastest route to my destination even if I have a cheaper alternative.	0.680	0.070	9.716
Factor Six	Sensitivity to personal travel experience			
PRFDRIVE	I would prefer to drive than to be driven.	1.000		
TRANCOMF	The people who ride transit to work are like me.	-1.181	0.081	-14.608
WALKING	I am comfortable walking near my destination during the day.	-0.678	0.055	-12.289
FERNOBUS	I would ride a ferry, but I wouldn't ride the bus.	1.273	0.092	13.887
PRFALONE	I prefer to make trips alone because I like the time to myself.	0.861	0.072	12.025
DLDRIVE	I don't like to drive but it is usually the fastest way to get where I need to go.	-0.352	0.065	-5.413

Table 2. Structural Equation for the *Desire to Help the Environment* Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	18-24	AGE1824	0.838	0.247	3.400	0.001
	25-34	AGE2534	0.583	0.118	4.939	0.000
	35-44	AGE3544	1.145	0.093	12.269	0.000
	45-54	AGE4554	0.935	0.093	10.093	0.000
	55-64	AGE5564	0.664	0.115	5.749	0.000
	65-74	AGE6574	0.782	0.189	4.138	0.000
College Student	None	CSTUDENT	0.457	0.306	1.492	0.136
Household Size	1 Person	HHSIZE_1	-0.358	0.402	-0.892	0.372
	2 Persons	HHSIZE_2	0.319	0.088	3.631	0.000
	3 Persons	HHSIZE_3	0.011	0.104	0.110	0.912
Households with Children (<18 years old)	0 Kid	HHSU18_1	0.020	0.112	0.175	0.861
	1 Kids	HHSU18_2	0.411	0.106	3.868	0.000
	2 Kids	HHSU18_3	0.334	0.229	1.463	0.143
Household Income	\$25-50,000	INC2550K	0.601	0.153	3.937	0.000
	<\$25,000	INC25K	0.838	0.374	2.239	0.025
	\$50-75,000	INC5075K	0.375	0.114	3.281	0.001
Vehicles per Household	0 vehicles	VEHS0	-1.669	0.528	-3.162	0.002
	1 vehicle	VEHS1	0.191	0.115	1.667	0.096
Workers per Household	0 worker	WORK0	-0.568	0.186	-3.055	0.002
	1 worker	WORK1	-0.269	0.095	-2.830	0.005
Households with more Workers than Vehicles	None	WORKCARS	-0.214	0.143	-1.499	0.134
Sensitivity to Personal Travel Experience Factor	None	f6_mode	-0.658	0.059	-11.151	0.000

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories: 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Table 3. Structural Equation for the *Need for Time Savings* Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	18-24	AGE1824	3.432	0.221	15.541	0.000
	25-34	AGE2534	3.340	0.114	29.379	0.000
	35-44	AGE3544	2.896	0.091	31.900	0.000
	45-54	AGE4554	2.661	0.089	29.969	0.000
	55-64	AGE5564	2.790	0.108	25.908	0.000
	65-74	AGE6574	2.068	0.168	12.308	0.000
College Student	None	CSTUDENT	0.086	0.270	0.320	0.749
Household Size	1 Person	HHSIZE_1	0.505	0.352	1.435	0.151
	2 Persons	HHSIZE_2	-0.174	0.074	-2.342	0.019
	3 Persons	HHSIZE_3	0.100	0.090	1.117	0.264
Households with Children (<18 years old)	0 Kid	HHSU18_1	0.233	0.098	2.375	0.018
	1 Kids	HHSU18_2	0.338	0.093	3.639	0.000
	2 Kids	HHSU18_3	1.123	0.201	5.573	0.000
Household Income	\$25-50,000	INC2550K	-0.070	0.134	-0.520	0.603
	<\$25,000	INC25K	-0.997	0.330	-3.027	0.002
	\$50-75,000	INC5075K	-0.058	0.100	-0.579	0.563
Vehicles per Household	0 vehicles	VEHS0	-0.346	0.464	-0.746	0.455
	1 vehicle	VEHS1	0.163	0.101	1.612	0.107
Workers per Household	0 worker	WORK0	-0.023	0.163	-0.143	0.886
	1 worker	WORK1	-0.070	0.082	-0.855	0.392
Households with more Workers than Vehicles	None	WORKCARS	0.496	0.125	3.964	0.000

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories: 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Table 4. Structural Equation for the *Need for Flexibility* Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	18-24	AGE1824	-0.170	0.423	-0.402	0.688
	25-34	AGE2534	-0.974	0.270	-3.603	0.000
	35-44	AGE3544	0.364	0.225	1.618	0.106
	45-54	AGE4554	0.150	0.213	0.706	0.480
	55-64	AGE5564	0.061	0.241	0.252	0.801
	65-74	AGE6574	0.175	0.309	0.565	0.572
College Student	None	CSTUDENT	0.325	0.459	0.708	0.479
Household Size	1 Person	HHSIZE_1	-0.008	0.600	-0.013	0.990
	2 Persons	HHSIZE_2	0.071	0.126	0.566	0.571
	3 Persons	HHSIZE_3	0.149	0.153	0.976	0.329
Households with Children (<18 years old)	0 Kid	HHSU18_1	0.541	0.168	3.223	0.001
	1 Kids	HHSU18_2	0.578	0.159	3.627	0.000
	2 Kids	HHSU18_3	0.783	0.349	2.244	0.025
Household Income	\$25-50,000	INC2550K	0.746	0.226	3.294	0.001
	<\$25,000	INC25K	1.425	0.564	2.526	0.012
	\$50-75,000	INC5075K	0.058	0.170	0.337	0.736
Vehicles per Household	0 vehicles	VEHS0	-1.022	0.791	-1.292	0.196
	1 vehicle	VEHS1	-0.654	0.172	-3.800	0.000
Workers per Household	0 worker	WORK0	0.828	0.278	2.979	0.003
	1 worker	WORK1	0.352	0.140	2.515	0.012
Households with more Workers than Vehicles	None	WORKCARS	-0.914	0.215	-4.256	0.000
Need for Time Savings Factor		f2_time	0.314	0.062	5.103	0.000

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories: 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Table 5. Structural Equation for the *Sensitivity to Stress* Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	AGE1824	0.959	0.172	5.577	0.000	AGE1824
	AGE2534	1.048	0.090	11.676	0.000	AGE2534
	AGE3544	1.145	0.077	14.842	0.000	AGE3544
	AGE4554	0.851	0.071	11.955	0.000	AGE4554
	AGE5564	0.673	0.083	8.116	0.000	AGE5564
	AGE6574	0.626	0.131	4.771	0.000	AGE6574
College Student	CSTUDENT	0.336	0.210	1.597	0.110	CSTUDENT
Household Size	HHSIZE_1	-0.204	0.274	-0.744	0.457	HHSIZE_1
	HHSIZE_2	0.125	0.058	2.159	0.031	HHSIZE_2
	HHSIZE_3	-0.062	0.070	-0.884	0.377	HHSIZE_3
Households with Children (<18 years old)	HHSU18_1	0.120	0.077	1.559	0.119	HHSU18_1
	HHSU18_2	-0.219	0.073	-2.994	0.003	HHSU18_2
	HHSU18_3	0.378	0.157	2.404	0.016	HHSU18_3
Household Income	INC2550K	0.239	0.104	2.292	0.022	INC2550K
	INC25K	0.209	0.257	0.812	0.417	INC25K
	INC5075K	0.399	0.079	5.022	0.000	INC5075K
Vehicles per Household	VEHS0	-1.346	0.365	-3.688	0.000	VEHS0
	VEHS1	-0.002	0.079	-0.031	0.976	VEHS1
Workers per Household	WORK0	-0.063	0.127	-0.495	0.621	WORK0
	WORK1	0.042	0.064	0.656	0.512	WORK1
Households with more Workers than Vehicles	WORKCARS	-0.114	0.098	-1.164	0.245	WORKCARS

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories - 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Table 6. Structural Equation for the Sensitivity to Transport Cost Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	18-24	AGE1824	-0.672	0.338	-1.988	0.047
	25-34	AGE2534	-0.765	0.221	-3.460	0.001
	35-44	AGE3544	-0.843	0.185	-4.552	0.000
	45-54	AGE4554	-0.471	0.174	-2.706	0.007
	55-64	AGE5564	0.024	0.195	0.124	0.901
	65-74	AGE6574	-0.059	0.245	-0.239	0.811
College Student	None	CSTUDENT	-0.053	0.361	-0.147	0.883
Household Size	1 Person	HHSIZE_1	-0.093	0.471	-0.197	0.844
	2 Persons	HHSIZE_2	-0.428	0.100	-4.285	0.000
	3 Persons	HHSIZE_3	0.051	0.120	0.428	0.669
Households with Children (<18 years old)	0 Kid	HHSU18_1	-0.055	0.132	-0.420	0.674
	1 Kids	HHSU18_2	-0.207	0.125	-1.651	0.099
	2 Kids	HHSU18_3	-0.651	0.275	-2.367	0.018
Household Income	\$25-50,000	INC2550K	-1.361	0.181	-7.510	0.000
	<\$25,000	INC25K	0.770	0.444	1.736	0.083
	\$50-75,000	INC5075K	-0.515	0.135	-3.825	0.000
Vehicles per Household	0 vehicles	VEHS0	0.096	0.621	0.155	0.877
	1 vehicle	VEHS1	-0.084	0.135	-0.623	0.533
Workers per Household	0 worker	WORK0	0.704	0.219	3.216	0.001
	1 worker	WORK1	0.384	0.110	3.477	0.001
Households with more Workers than Vehicles	None	WORKCARS	-0.761	0.170	-4.478	0.000
Need for Time Savings Factor		f2_time	0.266	0.052	5.133	0.000

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories: 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Table 7. Structural Equation for the *Sensitivity to Personal Travel Experience* Factor

Variable*	Stratification	Name	Estimate	Std Error	C.R.	P
Age	18-24	AGE1824	-0.589	0.184	-3.210	0.001
	25-34	AGE2534	-0.159	0.087	-1.820	0.069
	35-44	AGE3544	-0.116	0.068	-1.699	0.089
	45-54	AGE4554	-0.152	0.068	-2.233	0.026
	55-64	AGE5564	-0.293	0.085	-3.432	0.001
	65-74	AGE6574	-0.343	0.140	-2.449	0.014
College Student	None	CSTUDENT	0.132	0.226	0.585	0.558
Household Size	1 Person	HHSIZE_1	-0.962	0.300	-3.212	0.001
	2 Persons	HHSIZE_2	-0.532	0.069	-7.749	0.000
	3 Persons	HHSIZE_3	-0.427	0.079	-5.417	0.000
Households with Children (<18 years old)	0 Kid	HHSU18_1	-0.073	0.082	-0.886	0.376
	1 Kids	HHSU18_2	-0.287	0.079	-3.611	0.000
	2 Kids	HHSU18_3	0.310	0.169	1.830	0.067
Household Income	\$25-50,000	INC2550K	-0.155	0.112	-1.385	0.166
	<\$25,000	INC25K	-0.335	0.276	-1.213	0.225
	\$50-75,000	INC5075K	0.195	0.085	2.304	0.021
Vehicles per Household	0 vehicles	VEHS0	-0.489	0.390	-1.253	0.210
	1 vehicle	VEHS1	0.106	0.085	1.249	0.212
Workers per Household	0 worker	WORK0	0.231	0.137	1.681	0.093
	1 worker	WORK1	0.375	0.072	5.202	0.000
Households with more Workers than Vehicles	None	WORKCARS	0.370	0.107	3.473	0.001

* For each categorical variable, there is one category that is not used in the structural equation model. For example, there are actually three workers per household categories: 0 worker households, 1-worker households, and 2+ worker households. Only the first two of these three categories are included in the model.

Even with the use of structural equation modeling, we are limited in our ability to forecast all the factors reliably. Fortunately, three of the six factors have reasonable statistical reliability and are we therefore choose these three factors for use in market segmentation:

- Desire to Help the Environment
- Need for Time Savings
- Sensitivity to Stress

Table 8 presents the squared multiple correlations for the six traveler factors, with the three factors greater than 0.35 considered to be reliable for forecasting purposes. These three factors are used in the cluster analysis that identifies the market segments.

Table 8. Squared Multiple Correlations for Traveler Factors in the Structural Equation Model

Factor	Name	R-squared
Desire to Help the Environment	f1_envir	0.376
Need for Time Savings	f2_time	0.774
Need for Flexibility	f3_sched	0.141
Sensitivity to Stress	f4_stres	0.449
Insensitivity to Transport Cost	f5_cost	0.179
Sensitivity to Personal Travel Experience	f6_mode	0.133

Market Segmentation Models

Overview of Cluster Analysis

The core concept of market segmentation is to view a market as several segments rather than one homogeneous group. Each market segment is unique in its characteristics and attitudes toward travel behaviors. Market segmentation results can provide the basis for strategic marketing plans that involve applying different marketing strategies to different market segments. The objective of cluster analysis is to identify unique travel groups for market profiling. It is useful to the extent that people within the same cluster share similar attitudes toward travel behavior, while people in different clusters hold different views.

There are two types of cluster analysis we conducted for this market analysis: exploratory cluster analysis and confirmatory cluster analysis:

- **Exploratory cluster analysis** is a process where the statistics of the data determine the structure and content of the resulting cluster. It is used to explore the survey data to determine the nature of clusters that account for the co-variation between factors without imposing any a priori hypothesis about the number and structure of clusters underlying the data.
- **Confirmatory cluster analysis** is a process where we apply judgment regarding the structure and content of the clusters and then estimate the statistical results of these established factors. It is based on the results from exploratory cluster analysis as well as theoretical hypotheses as to which factors are correlated with which clusters.

In this report, we describe the exploratory cluster analysis and the confirmatory cluster analysis. We recommend that the confirmatory cluster analysis results are the market-based clusters for use in the remainder of the market analysis because the confirmatory cluster analysis offers a more viable method for evaluating construct validity of the attitudinal factors; it also enables us to explicitly test hypotheses concerning the cluster structure of the data, and provides more logical and consistent clusters that may provide better information for the modal choice models.

Exploratory Cluster Analysis

We use the FASTCLUS procedure in SAS/STAT software for cluster analysis. There are four major steps in the process of developing clusters:

- **Step one** is to identify the basis variables, that is, variables being used in the cluster detection algorithm. From an exploratory point of view, all available attitudinal factors should be included in the cluster analysis as basis variables. From a practical point of view, however, we should only select basis variables that have the greatest potential to be both analytically and strategically useful for market segmentation purpose. All the six factors extracted from factor analysis are used in the initial exploratory steps of cluster analysis, but only the ones with high R2 values are selected as the basis variables. As seen from Table 9, factor 1 (f1, desire to help the environment) and factor 3 (f3, sensitivity to travel stress) have the highest R2, so they are chosen to be the key attitudinal basis for clustering.
- **Step two** is to decide how many clusters to be segmented. With desire to help the environment (f1) and sensitivity to travel stress (f3) as the basis variables, a range of cluster analyses were conducted to decide the best segmentation scheme. Table 10 presents the Cubic Clustering Criterion (CCC) and R2 with different cluster numbers. The CCC can be used for crude hypothesis testing and estimating the number of population clusters. The local peak of the CCC indicates a good clustering number. The CCC has a local peak at 6 to 7 clusters. The pseudo F statistic also indicates 7 clusters, while the pseudo t^2 statistic suggests more clusters.

Table 9. Statistics for All Attitudinal Factors

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
Desire to Help the Environment	2.83706	1.31498	0.785499	3.661975
Need for Time Savings	0.91118	0.83224	0.167049	0.200551
Need for Flexibility	2.61287	1.31039	0.748872	2.982033
Sensitivity to Stress	0.78895	0.74211	0.116575	0.131958
Insensitivity to Transport Cost	0.95266	0.80804	0.281680	0.392137
Sensitivity to Personal Travel Experience	1.44278	1.24104	0.261244	0.353627
Total	1.79434	1.07104	0.644259	1.811038

Table 10. Statistics with Various Cluster Numbers

Cluster No.	CCC	Pseudo R2	Pseudo F
2	-2.675	0.37768	1707.35
3	-18.598	0.66701	2277.97
4	-13.294	0.75068	2572.47
5	-17.460	0.80195	2469.24
6	-19.092	0.83533	2414.39
7	-21.471	0.85895	2312.98
8	-18.513	0.87666	2437.98
9	-17.808	0.89044	2468.44
10	-16.823	0.90146	2511.79

Since we are using a two-dimensional segmentation scheme, six (rather than seven) clusters will be analyzed. The following contents presents results for cluster analysis when cluster number =6 and basis variables = f1, f3. Table 11 displays the table of statistics for the basis variables. The table shows the total standard deviation for each variable, the pooled within-cluster standard deviation, and the R2 for the basis variable.

Table 11. Statistics for Variables

Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
f1	3.01271	1.32631	0.806482	4.167482
f3	2.63632	1.30186	0.756514	3.107011
OVER-ALL	2.83078	1.31414	0.784813	3.647113

Pseudo F Statistic = 2414.39;

Approximate Expected Over-All R-Squared = 0.83533; and

Cubic Clustering Criterion = -19.092.

Table 12 shows the number of records falling in each market segment, and the root mean square standard deviation. The next two columns display the largest Euclidean distance from the cluster seed to any observation within the cluster and the number to the nearest cluster. The last column displays the distance between the centroids of the nearest cluster and the centroid of the current cluster. A centroid is the point having coordinates that are the means of all the observations in the cluster.

Table 12. Cluster Summary

Cluster Number	Frequency	RMS Std Deviation	Max Distance from Seed Cluster	Nearest Cluster	Distance Between Cluster Centroids
1	260	1.5054	4.7472	4	4.2527
2	152	1.6212	3.6834	3	5.0000
3	624	1.3076	3.2347	6	3.9335
4	932	1.2034	2.8176	5	3.6654
5	628	1.3078	4.3107	4	3.6654
6	720	1.3167	4.1044	3	3.9335

Table 13 shows the average factor scores for all six clusters. As we may observe, clusters 1 and 2 have high negative value in factor 1, clusters 3 and 4 have medium factors score in factor 1, and clusters 5 and 6 have high positive scores in factor 1. It is obvious that the desire to help the environment factor (f1) is a good indicator to divide the market into three segments: those favor environment protection, those against it, and those who do not care. If we examine the score of each cluster on f3 (need for flexibility), we observe

another important pattern, that is, each environmental segments can further divided into two groups based on their need for flexibility – those who have high sensitivity (indicated by a high positive score in f3), and those who have low sensitivity to scheduling (negative f3 score).

Table 13. Average Factor Scores for Clusters

Cluster	Desire to Help the Environment Factor	Need for Flexibility Factor
1	-4.74	-2.52
2	-6.53	2.38
3	-1.54	2.67
4	-0.53	-1.87
5	3.12	-2.18
6	2.39	2.42

Confirmatory Cluster Analysis

Confirmatory cluster analysis uses the same tools as exploratory cluster analysis, except that the factors used to define the various clusters are pre-determined. In our study, three factors were selected for inclusion in the cluster analysis based on our ability to reliably forecast these factors based on the structural equation modeling discussed in the previous section:

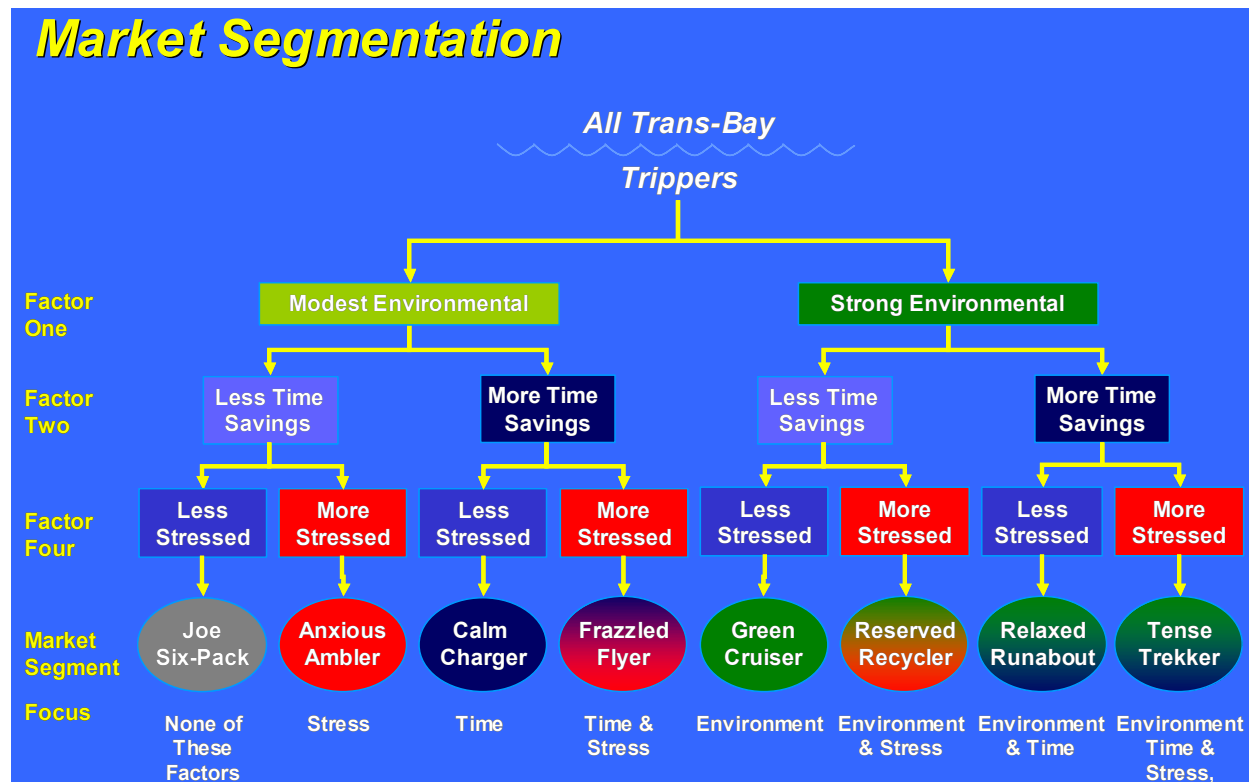
- Desire to Help the Environment,
- Need for Time Savings, and
- Sensitivity to Stress.

The results of this cluster analysis are presented in Figure 2. This demonstrates that each factor is divided into two groups, for a resulting stratification of a total of eight market segments. Each market segment is identified with a descriptive name that invokes the primary drivers behind the traveler attitudes in that segment, as follows:

- **Joe Six-Pack**, less focus in all factors;
- **Anxious Ambler**, higher sensitivity to stress;
- **Calm Charger**, need for more time savings;

- **Frazzled Flyer**, higher sensitivity to stress and need for more time savings;
- **Green Cruiser**, desire to help the environment;
- **Reserved Recycler**, desire to help the environment and sensitivity to stress;
- **Relaxed Runabout**, desire to help the environment and need for more time savings; and
- **Tense Trekker**, desire to help the environment, need for more time savings and sensitivity to stress.

Figure 2. Final Market Segmentation



The market segmentation process produces market segments of various sizes from the original household survey. These are presented in Figure 3. The smallest market segment is the Anxious Ambler, with 6 percent of the total surveyed, and the largest market segment is the Reserved Recycler, with 18 percent of the total surveyed.

One means of evaluating the specific traveler attitudes present in each of the market segments is to calculate average factor scores for the original six factors that are present within each segment. These are compared to the overall mean total factor scores to identify whether each market segment is higher or lower than the overall average.

Figure 3. Size of the Market Segments

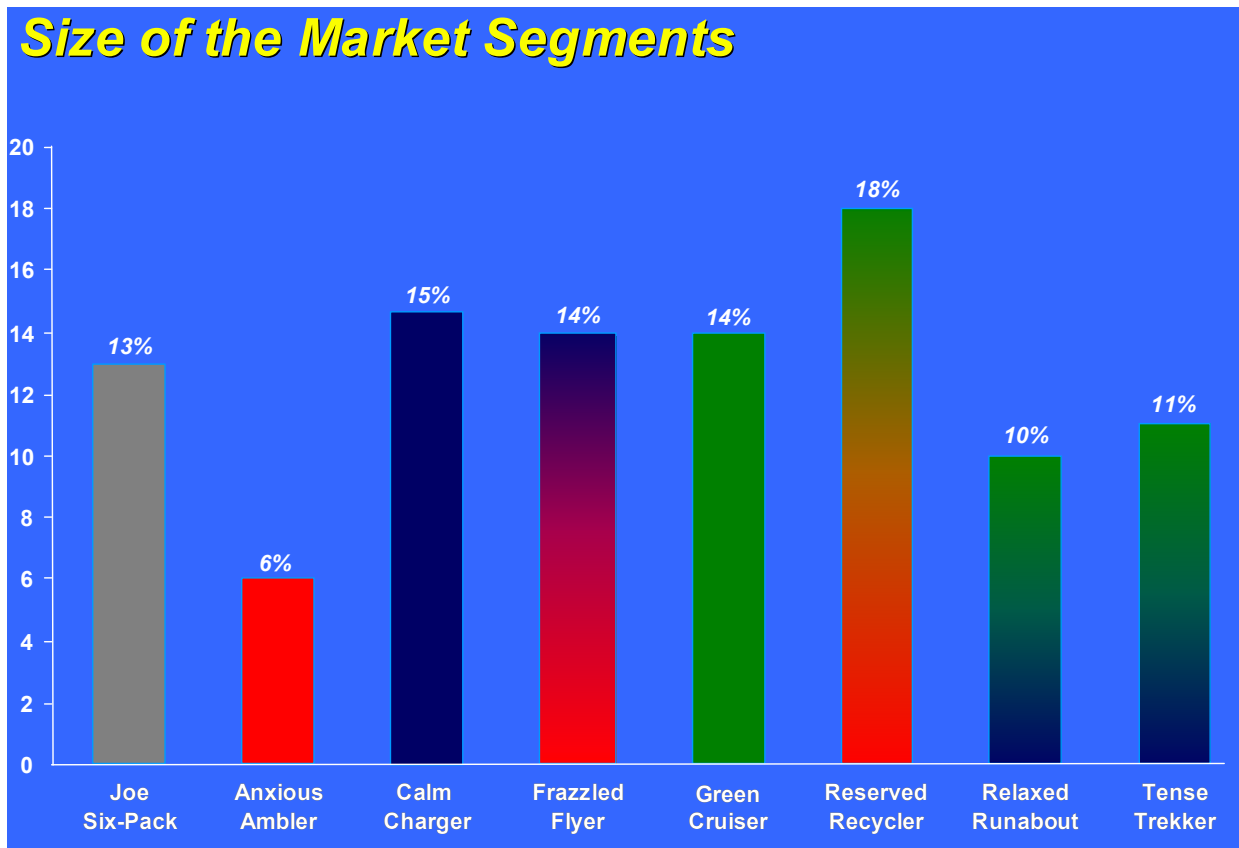


Figure 4 presents the results of this analysis for each of the eight market segments. The following conclusions can be made regarding these results:

- Anxious Amblers, Calm Chargers, and Frazzled Flyers are the most sensitive to personal travel experience, while Green Cruisers, Reserved Recyclers are the least sensitive.
- Tense Trekkers and Reserved Recyclers are the most sensitive to cost (note that this factor identifies *insensitivity* to cost and is therefore reversed in concept from the other factors), while Joe Six Pack and Calm Chargers are the least sensitive to cost.
- Tense Trekkers are more sensitive to stress than all other categories, while Joe Six Packs are the least sensitive to stress.
- Tense Trekkers and Relaxed Runabouts have the highest need for flexibility, while Green Cruisers have the least need for flexibility.
- Frazzled Flyers have the highest need for time savings, while Joe Six Packs have the least need for time savings.
- Reserved Recyclers have the highest desire to help the environment, while Joe Six Packs have the least desire to help the environment.

Figure 4. Average Factor Scores for each Market Segment

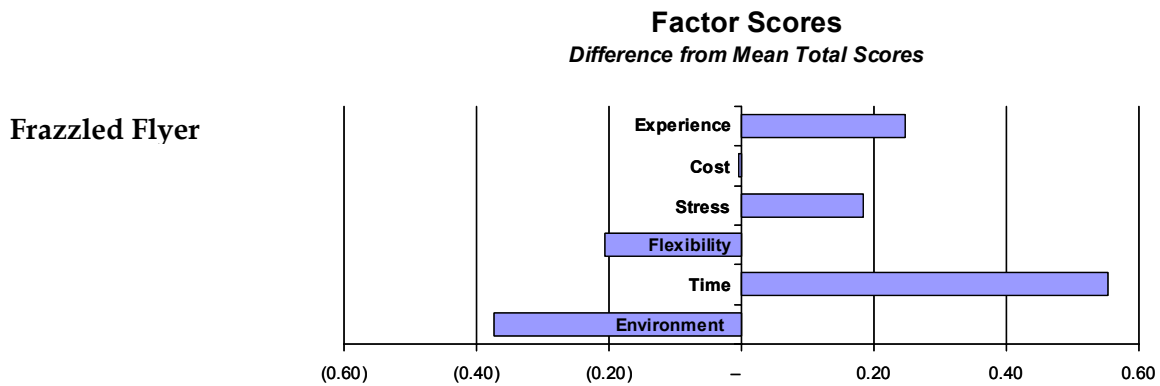
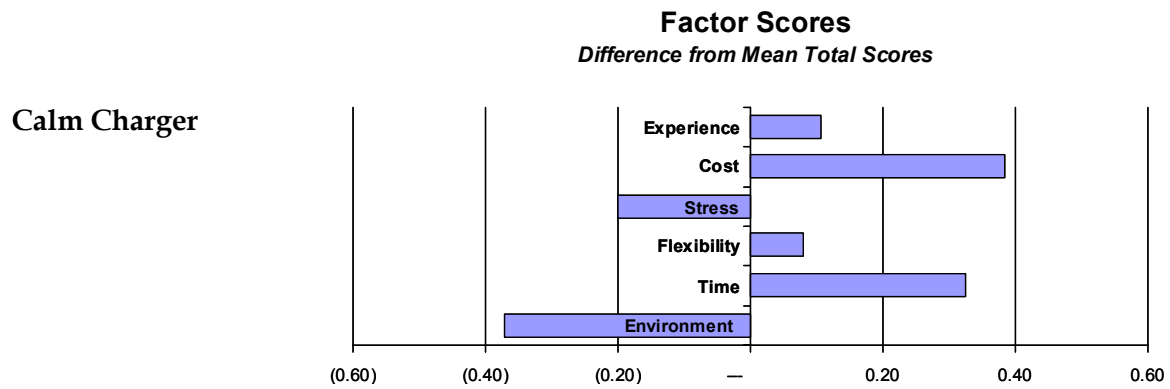
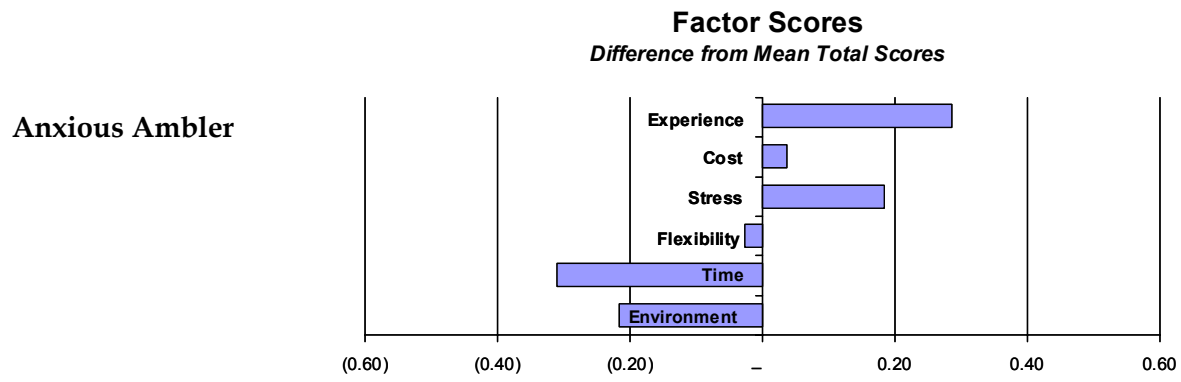
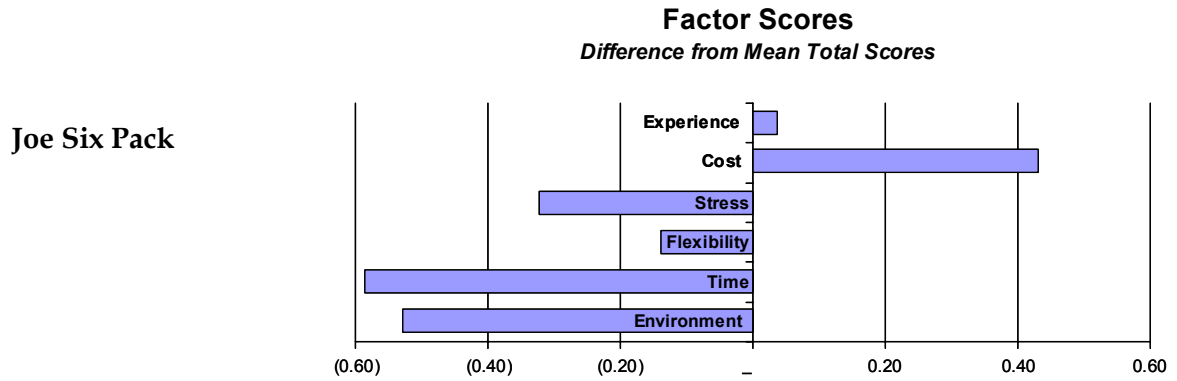
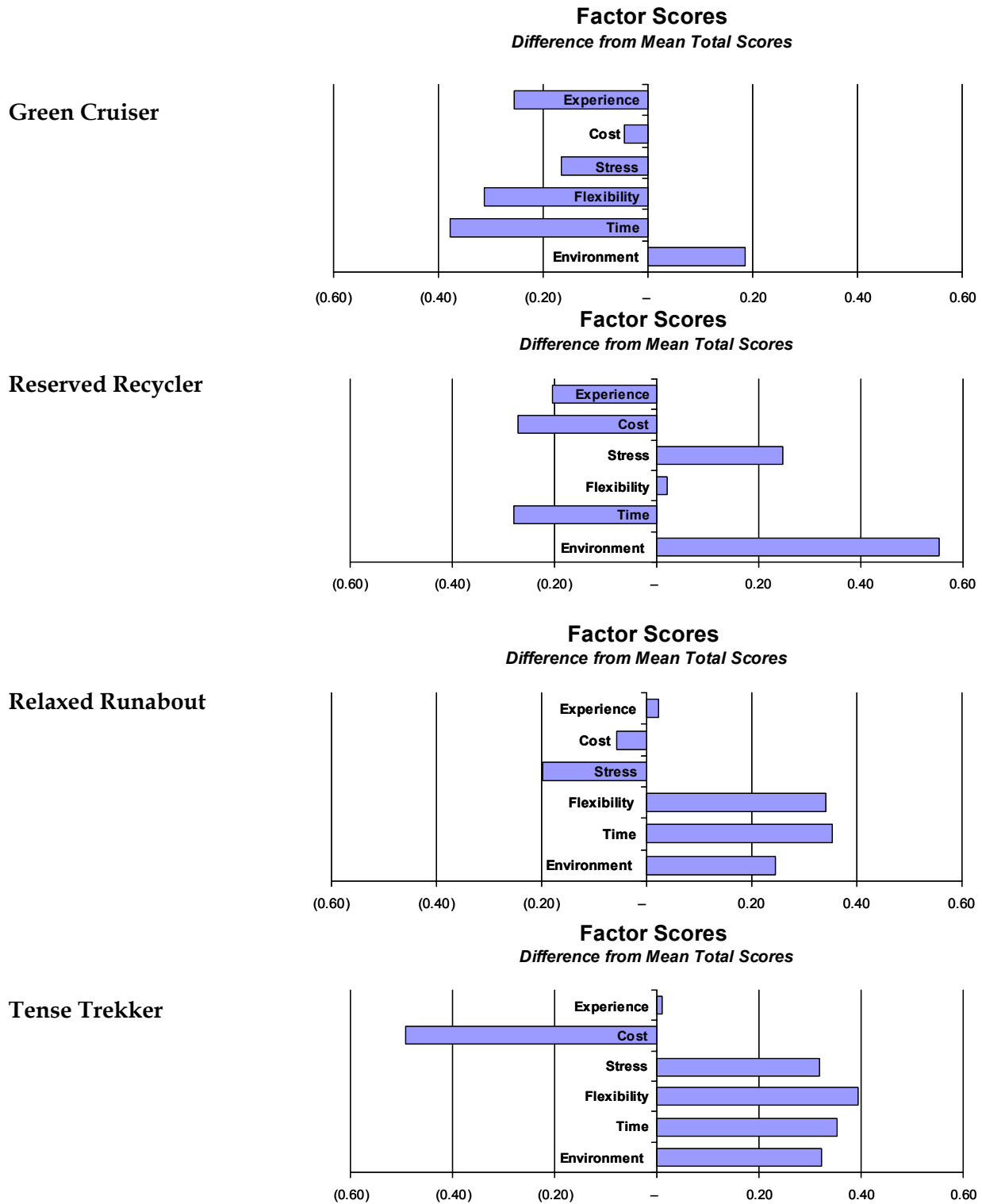


Figure 4. Average Factor Scores for each Market Segment (continued)



This information on market segments can be useful in designing transit services that meet the needs of various market segments, as well as for identifying those market segments that are not likely to take transit, regardless of the transit services provided. For example, market segments with a high need for time savings and a high need for flexibility (such as Relaxed Runabouts and Tense Trekkers) are more difficult to serve with fixed route transit systems. But market segments with a desire to help the environment and sensitivity to stress (such as Reserved Recyclers) are more likely to be well served by proposed ferry services.

Characteristics of Market Segments

The market segments identified through confirmatory cluster analysis have specific demographic characteristics that are used to understand the market segment and to support forecasting market segments for future populations. The demographic data used to forecast market segments is discussed in the previous section on structural equation modeling. This discussion focuses on the socioeconomic characteristics of each market segment that can be used to target each population with services and marketing messages that will appeal to the specific traveler attitudes of each segment.

The following conclusions regarding the relationships between socioeconomic characteristics and market segments can be demonstrated, as follows:

- Younger persons have a higher need for time savings than older persons. (Figure 6)
- Middle age persons have a higher desire to help the environment than either younger or older persons. (Figure 5)
- Households with 3 or more persons have a higher need for time savings than households with 1 or 2 persons. (Figure 6)
- Households with kids have a higher need for time savings than households with no kids. (Figure 7)
- Households with 3 or more kids are more sensitive to stress than households with 2 or 1 kids. (Figure 7)
- Lower income households (less than \$50,000 per year) and middle income households (\$50-\$100,000 per year) are more sensitive to stress than upper income households. (Figure 8)
- Households with only 1 vehicle are more sensitive to stress than households with 2 or more vehicles. (Figure 9)
- Households with 2 or more workers have a higher desire to help the environment than households with 0 or 1 worker. (Figure 10)
- Households with 2 or more workers have a higher need for time savings than households with 0 or 1 worker. (Figure 10)
- There are no significant differences in gender by market segment, although there is a slight tendency for females to be more sensitive to stress than males (Figure 11).

Figure 5. Market Segments by Age Group

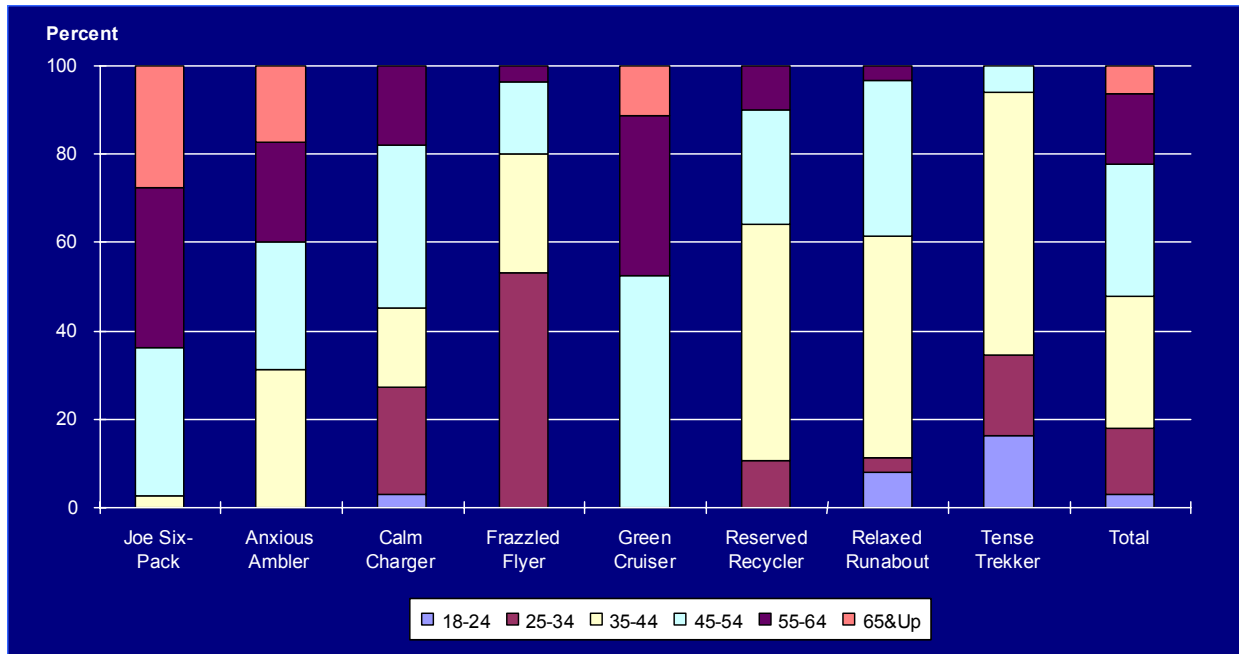


Figure 6. Market Segments by Household Size

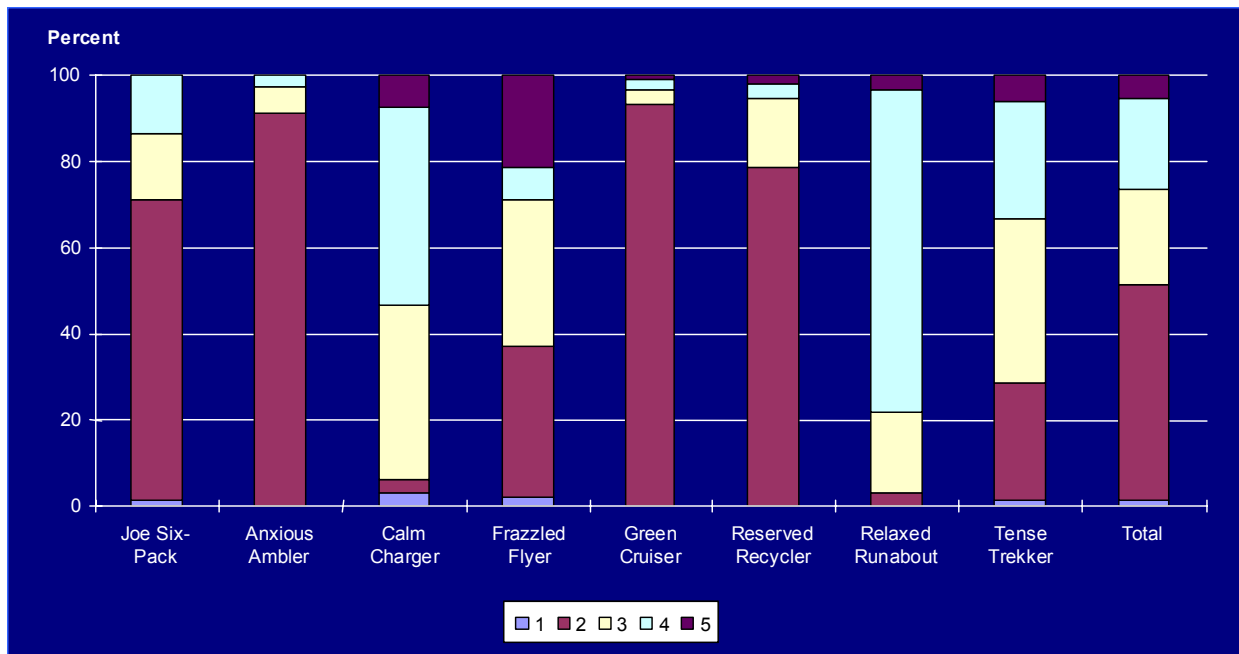


Figure 7. Market Segments by Number of Children

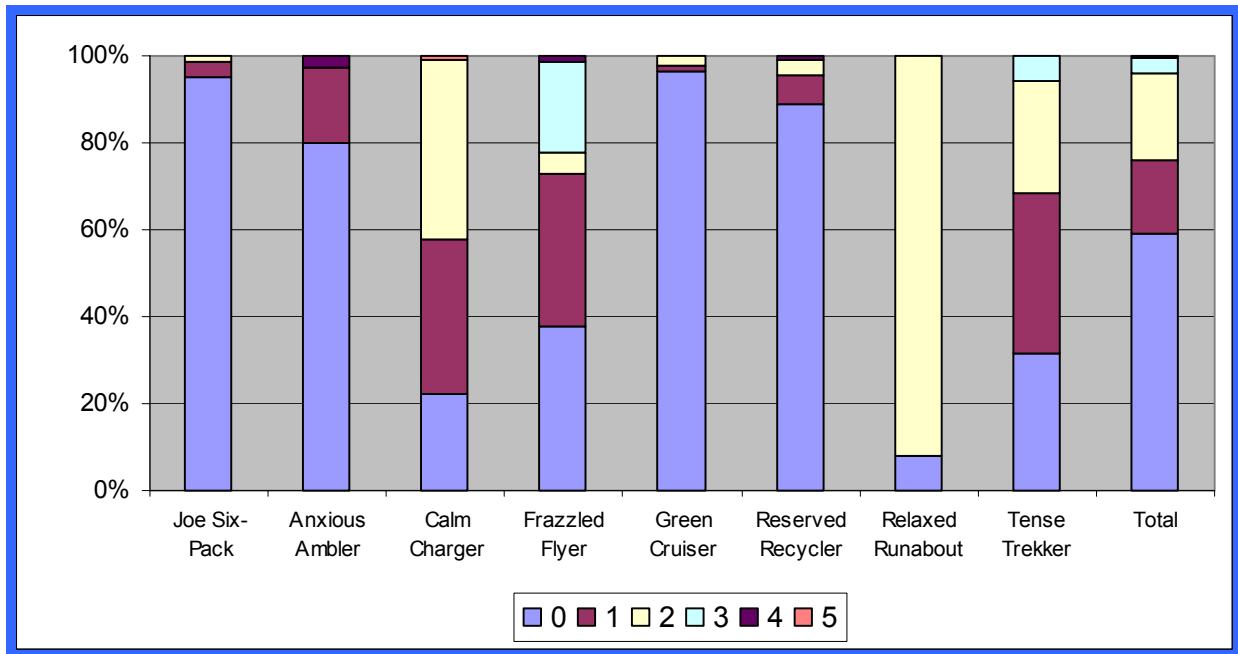


Figure 8. Market Segments by Household Income Category

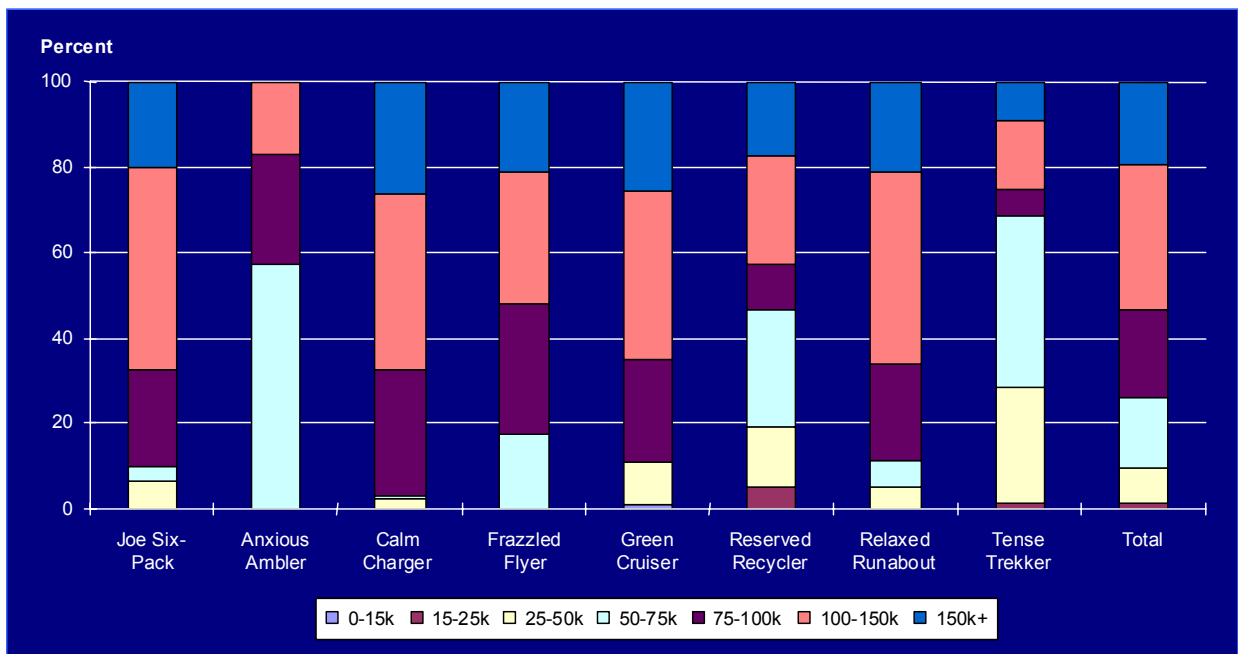


Figure 9. Market Segments by Vehicles Available

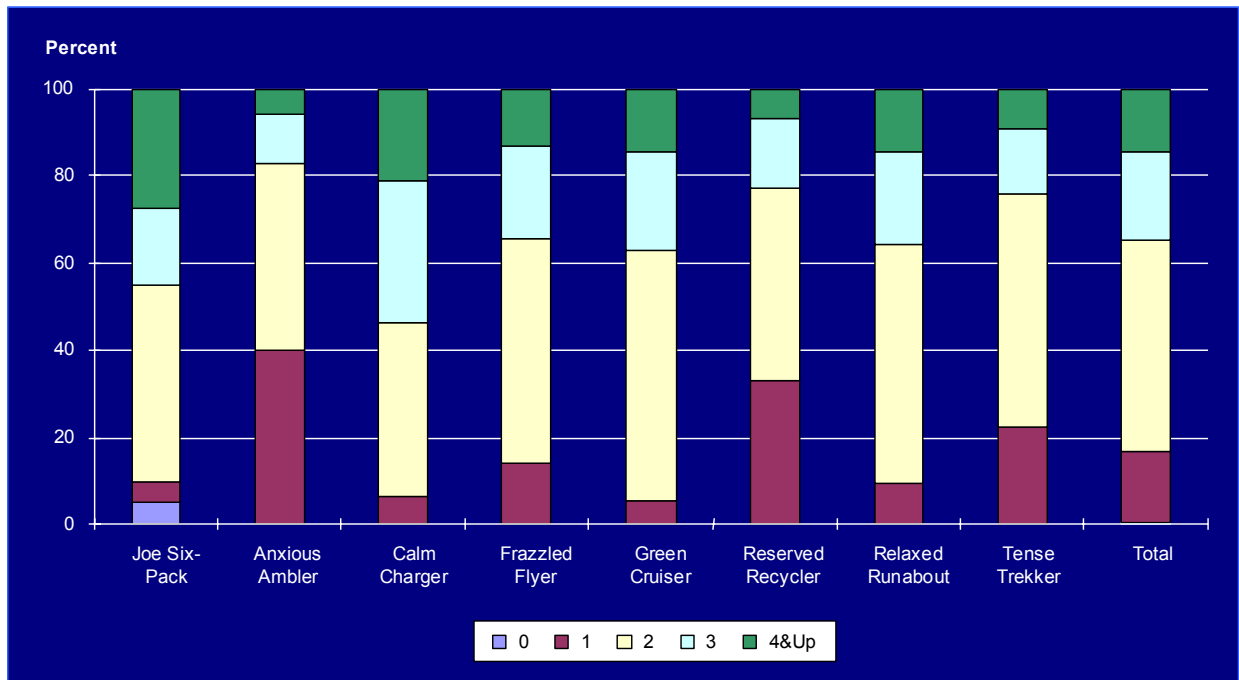


Figure 10. Market Segments by Workers per Household

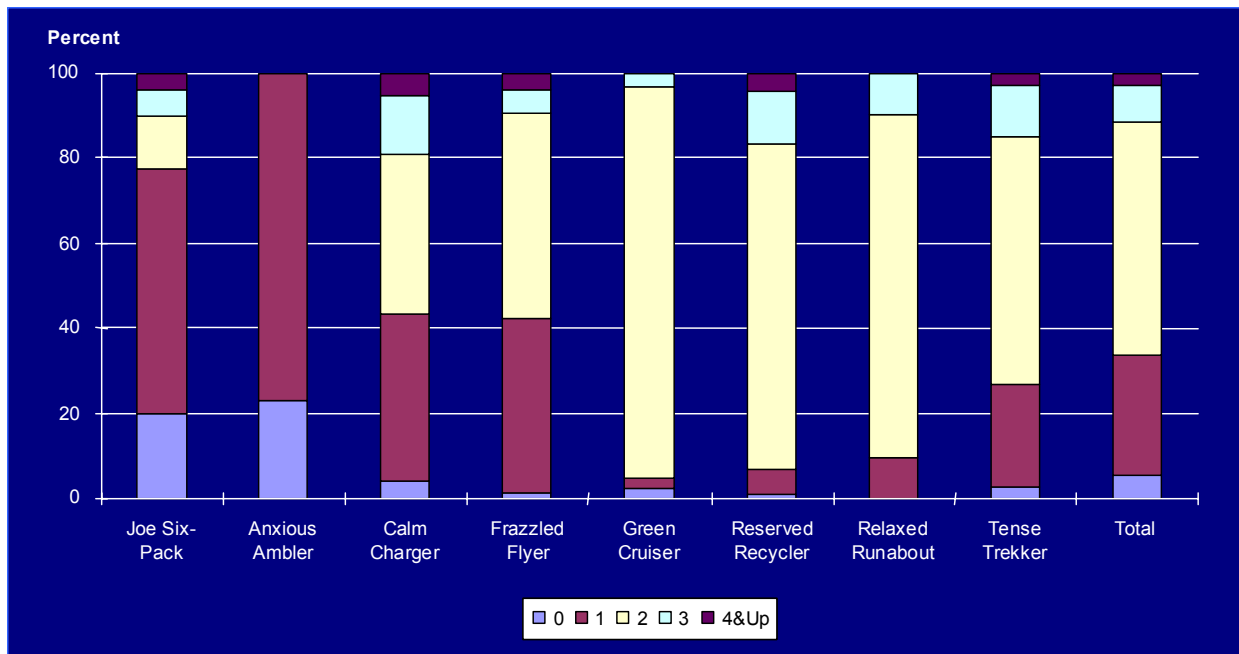
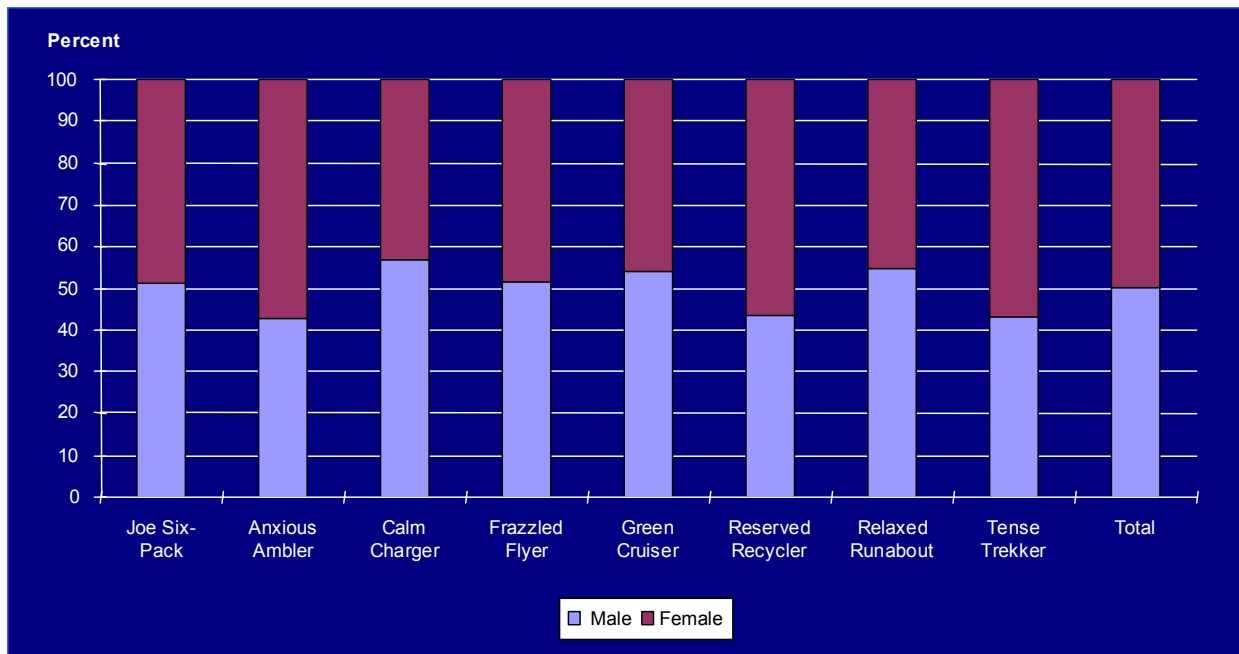


Figure 11. Market Segments by Gender



- College students and college graduates are slightly more likely to be sensitive to stress than post-graduates (Figure 12).
- Single people are slightly more likely to be sensitive to stress than any other marital status group (Figure 13).

Table 14 summarizes these socioeconomic characteristics by market segment and characteristic. This table demonstrates that there are clear demographic differences among each market segment, contributing to the overall differences in traveler attitudes.

Figure 12. Market Segments by Education Level

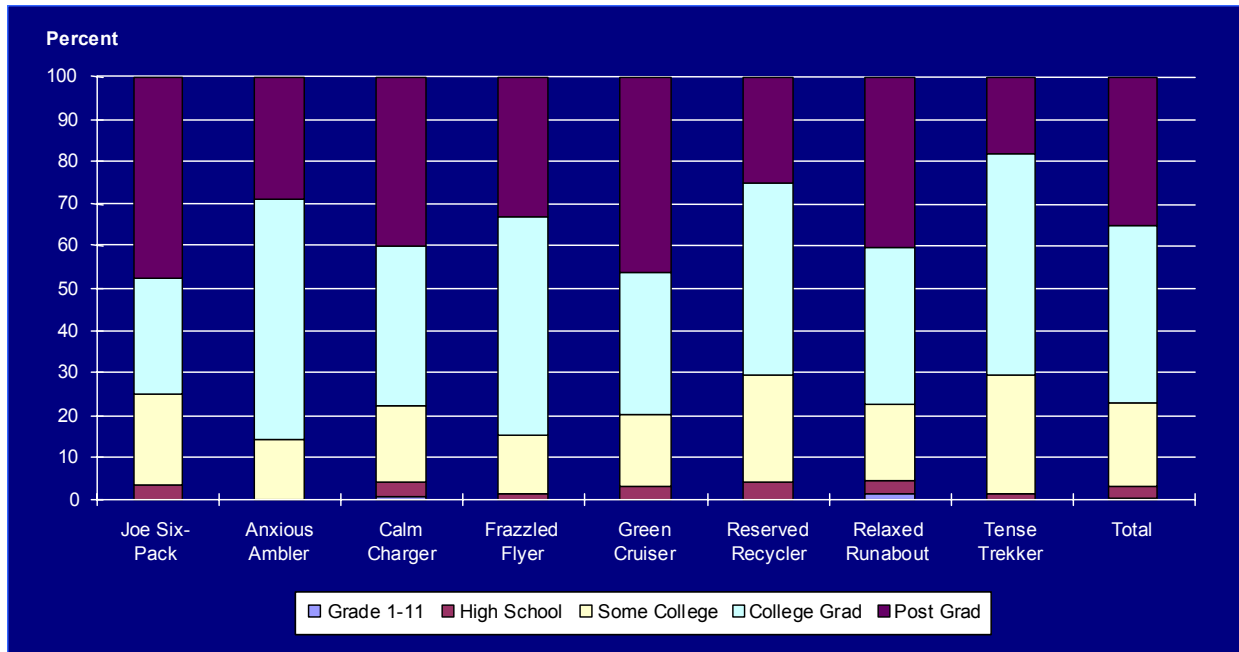


Figure 13. Market Segments by Marital Status

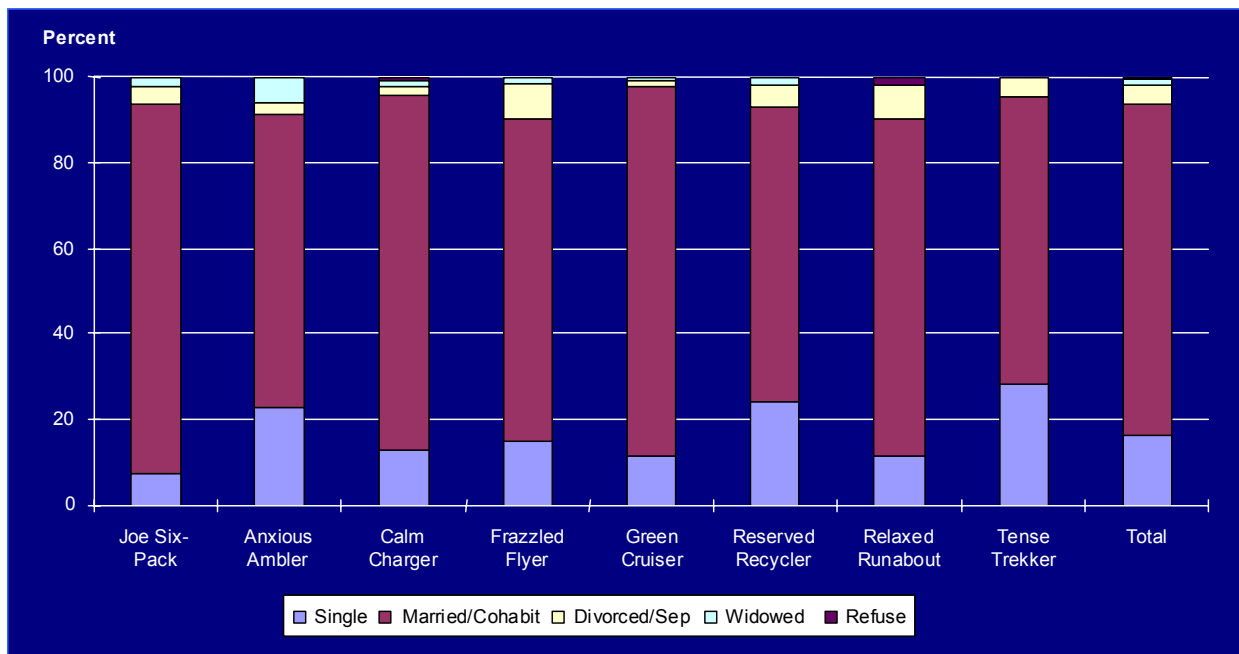


Table 14. Summary of Market Segment Characteristics

Market Segment	Average Age	Average Persons per Household	Average Number of Kids	Average Household Income	Average Vehicles per Households	Average Workers per Household
Joe Six Pack	Oldest	1-2	None	High	0 and 4+	0-1
Anxious Ambler	Older	1-2	None	Middle	1	0-1
Calm Charger	Young	3+	1	High	2+	1
Frazzled Flyer	Young	2-3	1	Middle-High	2+	1
Green Cruiser	Older	1	None	High	2+	2+
Reserved Recycler	Middle	1	None	Low	1	2+
Relaxed Runabout	Young and Middle	3+	2+	High	2+	2+
Tense Trekker	Youngest	2-3	1	Middle	1	2+