Memorandum

TO: Valerie Knepper, Metropolitan Transportation Commission

FROM: Cemal Ayvalik, Kimon Prossaligou, Ryan Greene-Roesel, Chris Wornum,

DATE: December 12, 2008

RE: TOD Choices Study - Market Segmentation Results (Deliverable 7b)

This document provides a summary of the final results of the market segmentation exercise undertaken for the TOD Choices study. It describes the analytical steps used in market segmentation, and presents the results of the second round of analysis, which focused on a six-factor solution. Tasks 8 and 9 of the study will enter into greater detail on strategies to attract the targeted segments to TODs.

Introduction

Most service providers understand that not all their customers are the same. The private sector has long used market research to investigate customer attitudes and preferences in order to develop and target products and service at specific market segments.

This study extends private market research techniques to identify the market segments for all those looking for an apartment or house. This research starts with a survey of Bay Area households that have recently moved or are planning to move. By surveying the attitudes of these "movers", it is possible to divide them into market segments, each with distinct attitudes and preferences. The segmentation will ultimately give public agencies critical information regarding public amenities and policies that can make Transit Oriented Developments (TOD) a more attractive choice to specific market segments.

Market segmentation is a multi-step process. This document describes each of the steps, which are:

1. Gathering of information on the attitudes of the target group (new movers).

2. Analysis of the attitudes to identify underlying attitudinal dimensions or "factors", which represent combinations of attitudinal statements. This step includes both *exploratory* and *confirmatory* factor analysis.

3. Structural Equations Modeling (SEM), which links each factor with a set of related socioeconomic characteristics.

4. Cluster analysis, which defines unique market segments based on respondents socioeconomic characteristics and the importance they attach to each factor.

115 South LaSalle, Suite 2200
Chicago, IL 60603
tel 312 346 9907 www.camsys.com **fax** 312 346 9908

It is important to emphasize that the research effort focuses on profiling recent movers and potential movers, not all the households in the region.

New Movers' Attitudes

The first step in market segmentation is to survey the attitudes of the target group; in our case, new movers in five Bay Area counties (San Francisco, San Mateo, Santa Clara, Alameda, and Contra Costa Counties). Specifically, "new movers" were defined as those who had either moved within the last three years or were planning to move within the next year.

Corey, Canapary, and Galanis surveyed the target group and collected 911 complete surveys, each of which contains responses to 35 attitudinal statements. The statements gauge the relative importance of considerations that influenced the respondent's housing location choice. Respondents' rated the attitudinal statements from zero to ten, where zero denotes the strongest level of disagreement with the statement and ten denotes the highest level of agreement. Table 1 below presents the average (mean) score for each attitudinal statement as well as the standard deviation of the score.

Table 1. Means and Standard Deviations for Attitudinal Statements

	Statements (abbreviated)	N	Mean	Std Dev.
1.	Living In a neighborhood where I felt safe enough walking at night…	911	8.5	2.4
2.	Living in a clean neighborhood…	911	8.3	2.3
3.	Living in a neighborhood where it is safe and convenient to walk or bike…	910	8.0	2.5
4.	Living within a short commute to work…	910	7.4	3.0
5.	Living on a quiet street…	911	7.3	2.7
6.	Having plenty of indoor space…	910	7.2	2.6
7.	Living where there are nearby places for outdoor recreation…	911	7.2	2.7
8.	Having pleasant public parks nearby where my family or pets can safely play…	911	7.2	2.9
9.	Living in a neighborhood where there are places to spend time…	911	7.2	2.6
10.	Having only one or fewer dedicated parking spots is sufficient if I could easily travel where I needed to go without my personal vehicle	889	7.0	3.1
11.	Having a dedicated parking spot…	911	7.0	3.4
12.	Having my own garage…	910	6.7	3.5
13.	Even with quick and reliable public transit, I would frequently need access to a car...	890	6.5	3.2
14.	Having a back yard…	911	6.4	3.4
15.	Living within easy access to a freeway…	911	6.4	2.9
16.	Living where I can easily commute using transit…	910	6.1	3.3
17.	Having easy on-street parking…	911	6.1	3.1
18.	I would be willing to spend more money on my home if it met more of my criteria	898	6.1	3.3
19.	Living in a neighborhood where transit is reliable...	906	5.9	3.4
20.	Living in a school district that provides a good education…	909	5.9	4.1
21.	Living within easy access to local bus or MUNI…	908	5.6	3.4
22.	Being able to safely walk to schools…	909	5.5	3.8
23.	Having a front yard or space in front of my home…	910	5.5	3.3
24.	If it would shorten my commute, I would live in a smaller home	907	5.3	3.5
25.	If transit was less expensive, I would take it more often	902	5.2	3.7
26.	Living within an easy transit ride to downtown San Francisco…	910	5.1	3.5
27.	Living within an easy drive to BART…	909	5.0	3.5
28.	Living within easy access from my home to a commuter rail system…	906	4.6	3.5
29.	Living within an easy walk to BART…	907	4.5	3.5
30.	Living in a newer home…	911	4.5	3.3
31.	I would be willing to have a longer commute if I found a home that met more of my criteria	897	4.5	3.4
32.	Living in a neighborhood that offers a mix of housing types …	909	4.4	3.2
33.	Living within an easy drive to downtown San Francisco…	911	4.1	3.2
34.	Living in a new neighborhood as opposed to an older neighborhood…	911	3.9	3.1
35.	Living in San Francisco…	905	3.0	3.6

The mean scores indicate that the majority of the respondents place a high value on safe, quiet, clean, and pedestrian- and cyclist friendly neighborhoods. Furthermore, they prefer short commutes. The respondents gave less importance (less than 5.0) to living in San Francisco and easy access to San Francisco, and living in new homes and neighborhoods.

The standard deviation is another useful statistic in proving information on how scores in each statement were distributed. A smaller standard deviation indicates that responses are grouped tightly around the mean and overall variance is low (in other words, there is less variation in how respondents rated the statement). Statements showing strong agreement present low standard deviations implying tight distributions around the mean. The statements relating to neighborhood cleanliness and safety, for example, all had low standard deviations and high mean scores, indicating that nearly all respondents place a high value on them.

Higher standard deviations indicate that scores are more widely distributed. Statements such as "*Living in a school district that provides a good education*", "*Being able to safely walk to schools…*", and "*If transit was less expensive, I would take it more often*" all had high standard deviations. These statements have more responses concentrated at both ends of the scale mainly due to the variations in importance of schools for households with different life-cycles. Generally, statements with higher standard deviations are better candidates for distinguishing among target markets.

Note that we did not normalize responses to the survey questions. Respondents were allowed to rate agreement with each statement on a scale of zero to ten. They could assign a high or low rating to all statements, and were not forced to rank the importance of the statements. This is precluded in a survey includes more than ten statements. The general rule is that respondents have difficulty ranking more than ten statements and our survey had 32. While a ranking of each statement's relative importance to the others may have been useful, the information gained from the 32 statements was regarded and significantly more valuable.

Approaches to Market Segmentation

The objective of market segmentation is to identify several distinct groups in the population whose members share the same set of values and are distinct from other groups. Each of the "homogeneous" market segments represents a group of people with similar attitudes and socioeconomic characteristics.

There are multiple approaches to market segmentation, ranging from simplistic to complex. The simplest approach is to divide the markets based on one or more socioeconomic characteristics. For example, differences in the beliefs and behavior between men and women may be considered as a basis for market segmentation.

Alternatively, the response to a single attitudinal question can be used for segmentation. For example, responses to the statement, "*Living in a neighborhood where it is safe and convenient to walk or bike…,*" can be used to classify respondents based on their willingness to live in

pedestrian and cyclist friendly neighborhoods and to identify market segments that would be more responsive to improvements in pedestrian and bicycle amenities.

Instead of relying on a single attitudinal statement, respondents can also be segmented based on their reaction to an attitudinal dimension, composed of a set of related attitudinal statements. For example, all those who placed a high value on questions relating to transit accessibility could be grouped together as a market segment.

We use an even more nuanced approach, which is to segment the market using multiple attitudinal dimensions and demographic characteristics. This approach assumes that there are a few key attitudinal dimensions or "factors" behind all observed attitudes that can be used to drive the segmentation process.

Achieving this nuanced approach to segmentation requires a multi-step analytical process consisting of factor analysis, structural equations modeling, and cluster analysis. Each step is described below.

Factor Analysis

The first task is to define key attitudinal dimensions using factor analysis. Factor analysis is an analytical technique that reduces a large number of attitudinal statements into a small number of key factors. Factor analysis is typically a two-step process, in which the first stage is "exploratory" and the second stage is "confirmatory".

Exploratory Factor Analysis (EFA) is used to investigate the factor structure underlying the observed data. This method uses a large set of variables as inputs and does not impose any specific hypotheses about how they are related. Instead, observed correlations in the responses to the statements are used to group them together into factors .

The process results in the assignment of each statement to an underlying factor according to its "loading", which is a measure of how much the statement is related to the factor. Loadings range between -1 and 1, where -1 indicates the statement has a high negative relationship to the factor, zero indicates the statement has no relationship to the factor, and 1 indicates the statement has a very high positive relationship to the factor.

Since the exploratory factor analysis is not guided by expectations or theory, the results are not always easy to interpret. In the next stage, *Confirmatory Factor Analysis,* the results are manually refined to allow for easier interpretation. For example, the number of factors can be altered to improve the statistical fit of the model; an attitudinal statement that does not appear to belong to its assigned factor can be moved to a different factor; and certain statements can be dropped if their removal improves statistical fit. In addition, the analyst develops names for the factors to allow for explanation and interpretation.

We tried several rounds of confirmatory analysis. Initially, we settled on a five-factor solution due to its superior statistical fit.

The five factors identified were given the following labels:

F1 – Transit Accessibility

F2 – Neighborhood Amenity

F3 – Driving Orientation

F4 – School Quality

F5 – San Francisco Access

The adoption of the five-factor solution, however, involved elimination of 12 attitudinal statements that did not fit within the five factors. To allow inclusion of a greater number of statements, we introduced a sixth factor, which we call "travel minimization" because it includes several statements relating to the importance of reducing travel (e.g. having a short commute to work, being within walking distance of shops, etc).

The six-factor solution had a slightly inferior statistical fit overall as compared to the five-factor solution. Table 2 below compares the fit of the two solutions in terms of three measures of statistical goodness-of-fit, as measured by the GFI, AGFI, and RMSEA statistics (defined in table footnote below).

Table 2. Statistical Characteristics of Five- and Six-Factor Solutions

	Five Factor Solution	Six Factor Solution.
Statistical Fit	GFI 0.87 AGFI 0.85 RMSEA 0.05	GFI 0.85 AGFI 0.83 RMSEA 0.06
Explanatory Variables	23 attitudinal statements	29 attitudinal statements

Note: The GFI is the Goodness of Fit Index, A relative measure of the proportion of variance explained by the model. It is analogous to R-Square in regression. Values larger than 0.90 indicate good fit. The AGFI is the Adjusted Goodness of fit Index: GFI is adjusted by the degrees of freedom (i.e. model complexity). Values larger than 0.90 indicate good fit. RMSEA - Root Mean Square Error of Approximation: A measure of discrepancy in the estimation of population covariances. Values smaller than 0.05 indicate good fit.

In spite of the slightly reduced statistical fit of the six-factor solution, and the fact that the sixth factor is highly correlated with the "transit accessibility factor" (see below for more detail), we decided to retain it. This allowed the retention of a number of attitudinal statements that could be used to inform TOD policy.

Six Factor Results

Table 2 below presents the results of the six-factor solution. The loading of each statement is shown in the "factor loading" column at the right of the table. The closer the loading is to -1 or

1, the stronger the relationship between the statement and the underlying factor. All of results are statistically valid based on goodness-of-fit measures.

These factors are not the only forces influencing individuals' housing choices. The factors are a product of the three focus groups, which were intended to screen a universe of possible attitudinal questions and isolate those that were most important. Furthermore, this universe was limited to those attitudes that relate to public policies: e.g., attitudes about safety, schools, access to amenities, parking, neighborhood character, etc. We did not include attitudes towards housing that would be controlled by the developer: e.g., number of bedrooms and bathrooms, quality of the appliances, interior design, etc.

Table 2. Summary of Initial Confirmatory Factor Analysis Results for the Six Factor Solution

Factor Label	Attitudinal Statements Associated with Factor	Strength of Association (factor loading)
Transit Accessibility	Living where I can easily commute using transit…	1.000
	Living within easy access to local bus or MUNI…	0.981
	Living in a neighborhood where transit is reliable...	0.925
	Living within an easy walk to BART…	0.870
	If transit was less expensive, I would take it more often	0.720
	Living within easy access from my home to a commuter rail system…	0.614
Neighborhood Amenity	Having pleasant public parks nearby where my family or pets can safely play…	1.000
	Living on a quiet street…	0.988
	Living in a clean neighborhood…	0.951
	Living where there are nearby places for outdoor recreation…	0.906
	Living In a neighborhood where I felt safe enough walking at night…	0.881
	Having a back yard…	0.781
	Living in a neighborhood where there are places to spend time…	0.631
	Living in a neighborhood where it is safe and convenient to walk or bike to do my errands	0.507
	Living within a short commute to work…	0.354
Driving Orientation	Having my own garage…	1.000
	Having a dedicated parking spot…	0.909
	Living within easy access to a freeway…	0.853
	Having easy on-street parking…	0.829
	Living in a newer home…	0.816
	Living in a new neighborhood as opposed to an older neighborhood…	0.721
	Even if I lived near quick and reliable public transit, I would frequently need access to a car	0.668
	Living in a neighborhood that offers a mix of housing types …	0.566
	Living within easy access from my home to a commuter rail system…	0.458
	Living within an easy drive to downtown San Francisco…	0.378
School Quality	Living in a school district that provides a good education…	1.000
	Being able to safely walk to schools…	0.871
	Having a back yard…	0.243
SF Access	Living within an easy transit ride to downtown San Francisco…	1.000
	Living in San Francisco…	0.742
	Living within an easy drive to downtown San Francisco…	0.607
Travel Minimization	Living in a neighborhood that offers a mix of housing types …	1.000
	... having only one or fewer dedicated parking spots is sufficient.	0.954
	Even if I lived near quick and reliable public transit, I would frequently need access to a car	-0.927
	Living within a short commute to work…	0.793
	Living in a neighborhood where it is safe and convenient to walk or bike to do my errands…	0.662
	Living in a neighborhood where there are places to spend time…	0.571
	Having my own garage…	-0.568
	Being able to safely walk to schools…	0.476
	Living where there are nearby places for outdoor recreation…	0.319

Final Step : Structural Equation Modeling

The factor analysis process tells us that there are a small number of underlying attitudinal dimensions (factors) among the sample population. It does not reveal which types of respondents (in terms of age, gender, etc.) care most about which factors. The next analytical step, Structural Equations Modeling, allows the linking of factors with demographic characteristics, as shown in Figure 1 below.

Another purpose of the SEM step is to generate factor scores for each individual survey respondent. The factor score can be thought of as the importance that individual places on that factor (not to be confused with the factor "loading" on individual attitudinal statements that defines the statement's membership in the factor group, as described above). The development of factor scores for individuals makes the last step (market segmentation) possible.

Table 3 below provides an example of the SEM results for the "school quality" factor. The table shows a set of demographic characteristics that predict the strength of the factor score. As would be expected, the presence of young children in the household has a strong positive impact on the factor score, while older age (above 55) has a negative influence on the factor score.

Figure 1. Schematic Representation of the SEM Specification

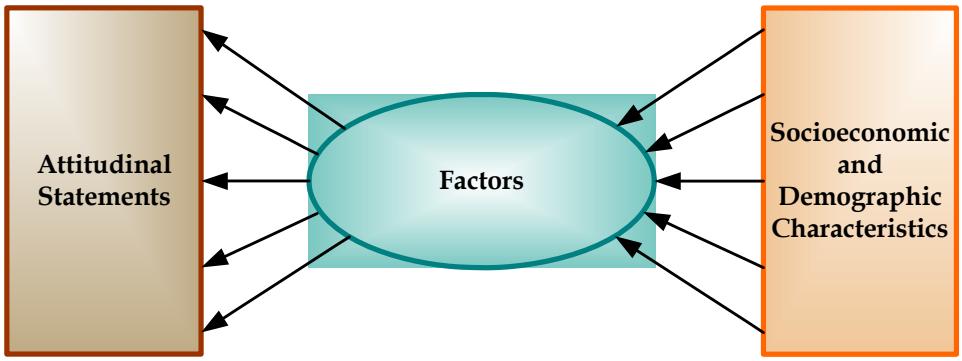


Table 3. Relationship of School Quality Factor with Demographic Characteristics

Demographic Characteristic	Strength of Association with School Quality Factor
Household with 4 or more members	4.232
Three-member household	2.590
Presence of one kid younger than 18 in the household	1.829
Age group for 55 to 64 years olds	-1.357
Resident of San Francisco County	-1.367

Selection of Factors for Segmentation

Although we identified five factors above, it is not necessary to use all of them to divide the market. To help decide which factors should be included in segmentation, we consider the following criteria:

- The selected factors should be those of the greatest interest and most easily interpretable and usable from a policy point of view;
- The selected factors should provide the greatest degree of homogeneity within the resulting market segments;
- There should be a low level of correlation between the selected factors. If two factors are highly correlated to one another, it is preferable to use one or the other in segmentation, not both. Using factors that are distinct helps select independent attitudes that are more likely to yield distinct segments.

Table 4 shows the correlations between the factor scores. The highlighted cells indicate pairs of factors that are less desirable for clustering due to a medium or high correlation level. The following pairs of factors are highly correlated, making it desirable to avoid using both factors for segmentation:

- Neighborhood amenity and driving orientation;
- SF access and transit accessibility;
- SF access and travel minimization; and
- Transit accessibility and travel minimization. This pair had a very high correlation (0.94), indicating that the factors represent essentially the same underlying attitude.

Table 4. Correlations between Factor Scores (Six Factor Solution)

	Pearson Correlation Coefficients					
	Transit Accessibility	Neighborhood Amenity	Driving Orientation	School Quality	SF Access	Travel Minimization
Transit Accessibility						
Neighborhood Amenity	0.242					
Driving Orientation	0.242	0.830				
School Quality	0.240	0.578	0.497			
SF Access	0.812	0.242	0.222	0.174		
Travel Minimization	0.939	0.173	0.174	0.161	0.773	

Based on the patterns of relationships between factors and the criteria discussed above, the following factors were selected to segment the market.

- Travel Minimization
- Neighborhood Amenity
- School Quality

The application of these three factors for segmentation, however, does not mean that they are the most important or that the remaining three are not important. The three selected merely work the best in forming the segments. Once the segments are defined, all six factors information about each segment.

Cluster Analysis

Cluster Analysis is the analytical technique used to achieve segmentation of the sample population into homogenous groups[1] (e.g. the market segments). The success of segmentation is assessed by statistics that provide inferences about the homogeneity of the segments. These

[1] In our analysis, we have adopted a nonhierarchical clustering technique referred to as the K-Means method. This method requires a priori knowledge of the number of clusters to be formed. However, by repeating the analysis with a different number of clusters, the best solution can be identified.

statistics serve as a guide in deciding on the number of segments to be formed and the segmentation scheme. The decision on the optimal number of segments is based not only on these statistical measures but also on the ease of interpretation.

Our analysis indicated an eight cluster solution could be preferable. Figure 2 shows the factor scores (from very low to very high) for each of the eight segments.

Figure 2. Schematic Representation of Mean Factor Scores by Segment

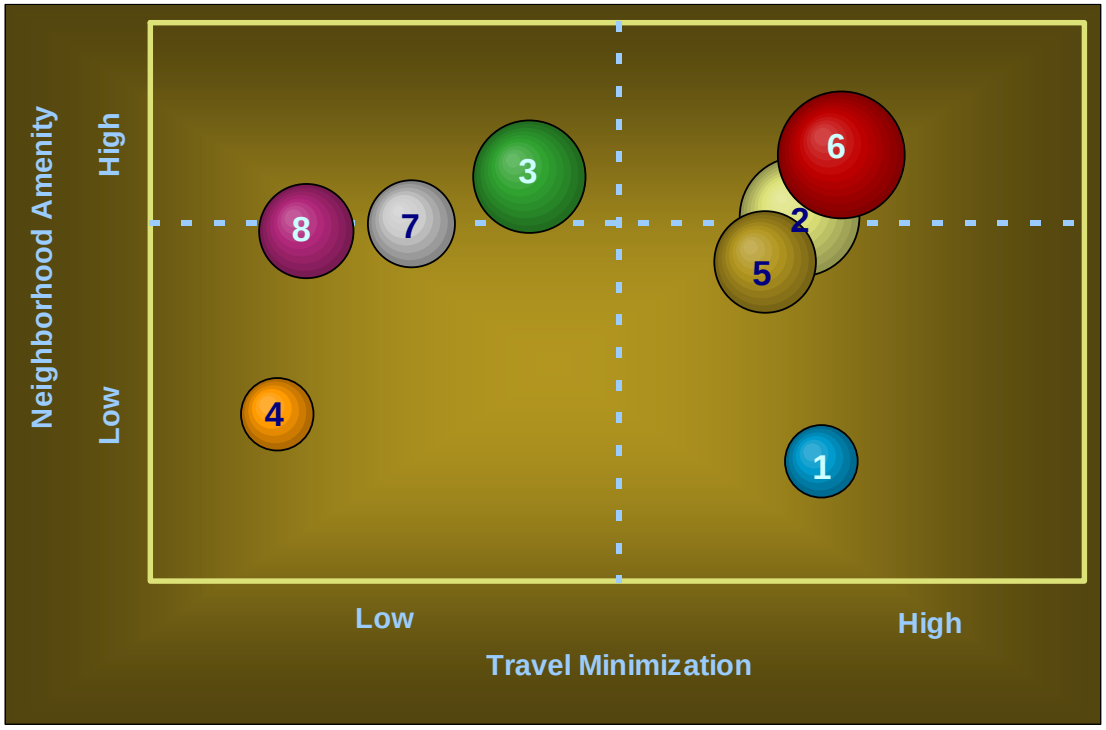
Segment Attributes			
Clusters	Neighborhood Amenity	School Quality	Travel Minimization
1	△△△	▲△	▲▲▲
2	▲△	▲△	▲▲
3	▲▲	▲▲▲	▲△
4	△△△	△△△	△△△
5	△△	△△△	▲▲
6	▲▲	▲▲▲	▲▲▲
7	▲△	△△△	△△
8	▲△	▲△	△△△

△△△ Very low
△△ Low
▲△ Medium
▲▲ High
▲▲▲ Very high

The results show significant differentiation in attitudes among the clusters. Some clusters (particularly 6) appear to place a strong importance on all the factors, while others (particularly 4) do not place a high importance on any of them. This latter response may indicate that factors not included in the survey were of more importance (e.g., location of the house/apartment, ethnic composition of the neighborhood, price, proximity to relatives, etc.). Note that these clusters were not chosen, but result from patterns in the data.

Figures 3 presents the relative positioning of each segment in a two-dimensional space defined by two of the factors that were used in market segmentation. The mean factor scores shown in Figure 2 above were used to position each market segment. The relative size of each market segment is represented by the area of each circle.

Figure 3. Segment Positions Neighborhood Amenity vs. Travel Minimization

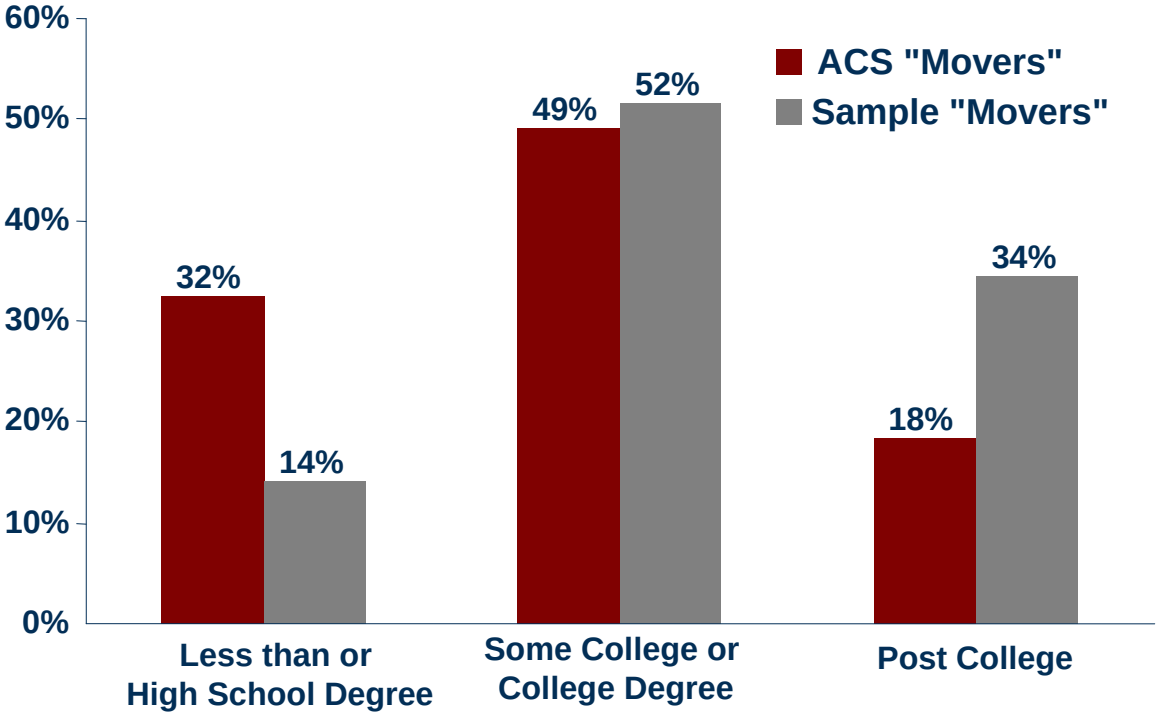


Segment Weighting

An additional step completed during the second round of analysis involved weighting the segments to better reflect the population of Bay Area movers. While the exact characteristics of the population of people looking for new homes ("movers") is not known, a reasonable proxy can be found in American Community Survey (ACS) data.

The ACS provides information on a limited set of demographic characteristics of individuals who have moved in the last year, while our survey collected information on a sample individuals who have moved in the past three years or are planning to move. While these groups are defined slightly differently, they are close enough that the comparison is useful. Figure 3 below compares the educational attainments (one of the few statistics available) of the population of movers in the ACS, as compared with the population of movers in our sample. It appears that lower-educated groups (less than a high school degree) are underrepresented in our sample, while higher-educated groups (more than a college degree) are overrepresented. The size of the market segments was adjusted to account for this discrepancy.

Figure 3. Educational Characteristics of Sample "Movers" versus American Community Survey "Movers"



Notes: Percentages reflect the educational status of "movers" in San Francisco, Alameda, Contra Costa, San Mateo, and Santa Clara. ACS "movers" are those who reported moving in the last year (2007 data). Sample "movers" are those who reported moving in the last year or were planning a move in the next three years (2008 data).

Summary

This document describes the process used to segment the market of new movers in the Bay Area. The results suggest that the selection of housing location is influenced by six main attitudinal dimensions or factors that may be influenced by public sector action:

1. The accessibility and quality of transit around the home (transit accessibility factor);
2. The accessibility and quality of highways and parking around the home (driving orientation factor);
3. The quality of schools around the home (school quality factor);
4. The perceived quality of the neighborhood environment in terms of its quiet, safety, cleanliness, etc.; (neighborhood quality / neighborhood quiet-safety factor);
5. The proximity of the home to San Francisco; and
6. The desire to minimize travel (travel minimization factor, very strongly correlated with the transit accessibility factor).

Respondents place differing values on each of these factors depending on their personal preferences, which may be influenced in part by their demographic characteristics (age, marital status, etc.). The market segmentation process revealed eight distinct groups of individuals (or clusters) where individuals belonging to any one cluster have more in common with their fellow members than they do with individuals in any of the other clusters.

The characteristics of each cluster or segment were presented and described in detail at the TOD Choices Study Advisory Committee on November 14th. The final presentation is appended to this document for reference.