

# BUILDING MOSAIC

## Introduction

Neighborhood classification or segmentation is one of the cornerstones of geodemographic analysis and is used in a wide range of marketing and site location applications, including: neighborhood description, customer analysis, facility planning, advertising, and direct mail. The attractiveness of neighborhood segmentation stems from the analytical performance and inherent simplicity and understandability of the technique. U.S. Mosaic is the latest in a series of neighborhood classification systems built by Experian, whose international lifestyle segmentation experience spans over twenty years and nearly twenty countries. Mosaic combines the best of Experian's international expertise with the knowledge and experience of some of the most experienced segmentation experts in North America. During the product development and refinement process, Mosaic was compared to other leading U.S. segmentation systems in a wide variety of tests. The results indicated that overall Mosaic performance is either better or equal to other established segmentation systems.

## Segmentation Objectives

Neighborhood segmentation systems are classifications of geographic areas according to their demographic, lifestyle, and other attributes. The goal of classification is to define a set of segments that are as different as possible while ensuring that the neighborhoods assigned to each segment are as similar as possible. In many discussions of segmentation, the words "cluster" and "segment" are used interchangeably. The objectives during the construction of Mosaic were:

- To create a series of segments that provides the most powerful discrimination of consumer behavior, lifestyles, and attitudes.
- To build segments that are as recognizable and meaningful as possible to marketers.
- To ensure that each of the segments contains sufficient numbers of households to be statistically reliable for most analyses.
- To ensure that each segment is homogeneous in terms of demographics and consumer behavior.
- To avoid an excessive concentration of individual segments within particular geographic regions, except where appropriate.



## Methodology

The original formulation of Mosaic was built using 1990 Census demographics at the block group level of geography, which are typically neighborhoods of approximately 500 households. Over two hundred and twenty five thousand such block groups were defined in the 1990 Census, the vast majority of which were included in the analysis to develop the MOSAIC classification. Block groups that did not meet a minimum threshold household count were excluded from the analysis.

The variables chosen for consideration in the model were selected based on several factors and objectives:

- Variables should represent different demographic categories, avoiding overrepresentation of any single category. For example, a system based excessively on income characteristics would fail to discriminate neighborhoods on other characteristics such as household type, size, and age.
- Sets of variables that are highly correlated, or closely related, were avoided by selecting only the most predictive variables from these sets.
- Variables should correlate well with consumer behavior.
- Variables should have sufficient sample size to be statistically valid.
- Variables should not be heavily concentrated in a small number of geographic areas.

The variables used in the development of Mosaic (listed on the page following) reflect a balance of the range of factors that affect consumer behavior, expenditures, and attitudes.

Unlike other systems that rely on data reduction techniques such as factor analysis, the methodology employed allowed each individual variable in its raw form to influence the cluster code given to a particular Block Group. A unique variable weighting facility was used that allowed different levels of influence to be assigned to different variables. This facility is used to weight more heavily the influence of highly predictive variables and variables from categories whose availability is poor. Likewise, over represented categories can be weighted less heavily.

The cluster algorithm used to build Mosaic is known as “iterative relocation” and is based on statistical similarity measures based on least squared differences. Prior to clustering, the variables used are standardized on the basis of means and standard deviations in order to eliminate the effects of measurement scale on the results (e.g. dollars versus percentages). Based on random starting points, proportional to population, the algorithm assigns each neighborhood to the most similar cluster. The average scores of each cluster are recalculated on each input variable, after which neighborhoods are assigned to new clusters



where a better fit is achieved. The process is repeated over many cycles in order to optimize the resulting classification. When complete, the final set of neighborhood segments are as different as possible across the input variables and within each segment, the neighborhoods are as similar as possible. Each of the resulting 62 segments was then classified into groups of segments in order to provide a simple hierarchical structure.

## **Segmentation Variables**

### **Race and Ethnic Origin**

White  
 Black  
 Asian or Pacific Islander  
 Native American  
 Hispanic, Mexican Origin  
 Hispanic, Puerto Rican Origin  
 Hispanic, Cuban Origin  
 Hispanic, Other Origin

### **Age**

Age 5 and Under  
 Age 6-13  
 Age 14-17  
 Age 18-24  
 Age 25-34  
 Age 35-44  
 Age 45-54  
 Age 55-64  
 Age 65 and Older

### **Family Status**

Married Couple Family  
 Single Parent Family

### **NonFamily Household Type**

NonFamily Households  
 College Dormitories  
 Military Quarters

### **Household Size**

5+ persons in household  
 Average household size

### **Travel To Work**

Drive Car to Work  
 Take Public Transportation  
 Work at Home  
 Average Commute Time

### **Employment by Industry**

Unemployed civilian  
 Agriculture/Forestry/Fishing  
 Mining  
 Construction  
 Manufacturing/Transportation  
 Communications/Utilities  
 Trade and Services

### **Occupation**

Manager/Professional  
 Technical  
 Sales

### **Clerical and Administrative**

Service Occupations  
 Farming/Forestry/Fishing  
 Craftsmen/Repairmen  
 Nonfarm Self Employed  
 Farm Self Employed

### **Income**

Median household income  
 Median family income

### **Forms of Income**

Interest/Dividend/Rental  
 Income

Social Security income  
 Public Assistance income

### **Housing Tenure**

Owner occupied  
 Average contract rent

### **Housing Type**

1 unit, detached  
 1 unit, attached  
 2 units  
 3 or more units  
 Mobile home  
 Median home value  
 Seasonal homes

### **Car Ownership**

3 or more vehicles

### **Tenure/Age of Building**

Owned, pre 1950  
 Owned, 1950-1969  
 Owned, 1970-1979  
 Owned, 1980-1990  
 Rented, pre 1950  
 Rented, 1950-1969  
 Rented, 1970-1979  
 Rented, 1980-1990

### **Education**

High school diploma  
 Some college, no degree  
 Associate/Bachelor degree  
 Graduate/Professional degree



## Validation and Description of the Classification

To evaluate the quality of the solution, Experian contracted with several experienced demographers. As part of the validation process, Mosaic types were compared across more than 200 variables, including a large number that were not directly used in the analysis. Geographic distributions and characteristics such as population density were also analyzed to ensure that the Mosaic types were both optimal and valid. The segment descriptions are based on these demographic and geographic analyses. In addition, Mosaic was extensively tested against other segmentation systems in order to evaluate the response lift, cluster differentiation, system parity, coverage, geographic comparability, and minority/ethnicity representation. Experian contracted with a database marketing firm to conduct the tests, which demonstrated that Mosaic performs comparably with other systems.

### Global Mosaic

Global Mosaic is the only segmentation system that classifies over 800 million of the world's consumers and is composed of thirteen lifestyle types that can be found in every modernized country. It is based on the simple proposition that the world's cities share common patterns of residential segregation. Each has their low-income inner city areas, upscale suburbs, and rural communities. Using highly localized statistics across eighteen countries, Experian has identified thirteen types of residential neighborhoods, each with a distinctive set of values, motivations, and consumer preferences. Global Mosaic serves as the common analytical currency between the country specific Mosaic typologies. Each of the Mosaic types in every country is cross-referenced to one of the Global types on the basis of four primary attributes:

- Age structure
- Family structure
- Extent of urbanization
- Income structure

Global Mosaic is used to apply consistent micro-marketing strategies on a global basis. The relationship between the U.S. Mosaic segments and Global Mosaic is shown in the table at right.

Global Group	Global Group Name	U.S. Mosaic Segments
A	Agrarian Heartlands	I43, I44, I45
B	Blue Collar Self Sufficiency	D20, D21, D22, D23, E30, G38, I48
C	Career Focused Materialists	A03, A06, A07, C15
D	De-Industrial Legacy	G37, I47
E	Educated Cosmopolitans	A05, B09, B10, B11, B12, B13, C17, F31
F	Farming Town Communities	D24, G36, I46, J53
G	Greys, Blue Sea & Mountain	C18, E25, E28
H	Hardened Dependency	K58, K59, K60
I	Inner City Melting Pot	F35, J49, J50, J51, J52, J54, K55, K56, K57
L	Lower Income Elderly	H39, H40, H41, H42
M	Midscale Metro Office Workers	B14, C16, C19, E29, F32, F33, F34
N	Non-Private Residences	L61, L62
O	Old Wealth	A01, A02, A04, A08, E26, E27



## **Annual Updates**

Mosaic is updated on an annual basis to reflect changes in the demographic characteristics of existing neighborhoods and to reflect newly developed neighborhoods. In most years, this update consists of the reassignment of neighborhoods to the existing classification system based on changes to the underlying neighborhood demographics. Annual demographic estimates are provided by Applied Geographic Solutions (AGS), a premier provider of demographic and market related data. A key element of the AGS projection methodology is the use of Experian's household level INSOURCE<sup>SM</sup> database, the most comprehensive national household source available. Every few years, Mosaic is subjected to a more rigorous update in which the classification system itself is reanalyzed and restructured in order to accurately reflect changes in household structure, migrational patterns, and income.

## **Readings on Segmentation Analysis**

Much has been written over the last several decades on the methodology and uses of segmentation techniques in target marketing. Among these are: Art Weinstein, "Market Segmentation: Using Demographics, Psychographics and Other Niche Marketing Techniques to Predict and Model Customer Behavior" Michel Wedel and Wagner Kamakura, "Market Segmentation: Conceptual and Methodological Foundations" Chester Swenson, "Selling To A Segmented Market: The Lifestyle Approach" In addition, the manuals that ship with both the SAS and SPSS statistical programs contain excellent introductions to statistical classification methods, but require some understanding of parametric statistical analysis.

