CollegeBoard

Segment Analysis Service™:

An Educationally Relevant Geodemographic Tagging Service

Table of Contents

Segment Analysis Service™ is a record-tagging service that provides educationally relevant geodemographic information to enrollment managers and other educational practitioners interested in knowing more about college selection, choice, and persistence.

An Introduction to Geodemography

The basic tenet of geodemography is that people with similar cultural backgrounds, means, and perspectives naturally gravitate toward one another or form relatively homogeneous communities; in other words, birds of a feather flock together. When they are living in a community, people emulate their neighbors, adopt similar social values, tastes, and expectations, and—most importantly for consumer marketers—share similar patterns of consumer behavior toward products, services, media, and promotions.

The primary appeal of geodemography from the marketer's perspective is that, with just an address, s/he can begin to craft an image about a particular set of individuals based on the values, tastes, expectations, and behaviors associated with their geographic community. This is done by mapping small bounded geographical regions, typically at a nine-digit zip-code level, against data from credit card agencies, U.S. Census data, and other consumer databases that track consumer characteristics, attitudes, and behaviors. The result is a series of geodemographic "clusters" that represent types of individuals based on a unique set of characteristics, attitudes, and behaviors.

Here is an example of this type of cluster:

> **Rustbelt, USA**: Picture a small town, where life once centered on the now-defunct local mill. Today, empty-nesters spend their evenings on the front porch overlooking quiet, tree-lined streets. Old family-owned businesses struggle to compete with new discount superstores and fast-food restaurants; big American cars are the preferred means of transportation. Healthcare and pharmaceuticals are leading consumer expenditures and housing start-ups rank in the bottom quartile. Income ranks 37th among the 50 clusters.

As with most traditional geodemographic clusters, our example includes a name (Rustbelt, USA) that captures the essence of the consumer group being described, as well as additional text intended to quickly paint a picture of this group. The example also includes some specific references to consumer behavior and spending capability.

At a broader, more action-oriented level, geodemographic clustering allows marketers to target sets of individuals that can best be served through the products and services they offer and, more importantly, to communicate more effectively with these individuals—particularly as they are building new relationships with them. For this reason, organizations targeting an older demographic, such as insurance companies, drug manufacturers, and discount retailers, probably would want to connect with individuals who are part of the Rustbelt, USA, cluster described above.

Although proven and accepted in the commercial marketplace, this type of consumer-focused geodemography has some obvious flaws when applied to the college-choice process. For example:

- Traditional geodemographic systems base their modeling on general data collections tools such as credit card companies and U.S. Census and/or consumer surveys, which have little to do with the phenomenon of college choice.
- Traditional geodemographic systems base their clustering on data for the entire adult population of the country—over 270 million individuals—rather than the subset of only two million traditional-age college-bound students.

- Traditional geodemographic systems produce clusters related specifically to home address but may miss other geodemographic constructs that are important for understanding college-bound students and the factors that impact their choice of colleges (such as the prospective student's high school!).

Educationally Relevant Geodemography

To fully capture the relevant characteristics and behaviors of college-bound students (and their families), while simultaneously addressing the uniqueness of the college-choice process, the College Board offers enrollment managers an educationally relevant geodemographic tagging service called Segment Analysis Service. This service avoids the use of standard consumer-focused neighborhoods that are thinly populated with college-bound students. Instead, it creates and builds on a new set of geodemographic communities composed entirely of college-bound students—referred to throughout this paper as *educational neighborhoods*.

2011 Revision

The original version of Segment Analysis Service, Descriptor PLUS™, defined these neighborhoods based the zip +4 address. Neighboring nine-digit zip codes were combined based on size and similarity to achieve a sufficient sampling of college-bound students. However, zip codes were never intended to represent physically bounded areas--they exist for the convenience of the post office. Hence, they are a less-than-ideal basis for a geodemographic clustering system for a number of reasons: They do not strictly respect political boundaries such as counties or states; they are subject to frequent and arbitrary changes; and they are either too small (zip +4) or too large (5-digit zips) to stand alone as a unit of analysis for college-bound students. The 2011 revision introduced a new set of neighborhoods derived from Census tracts that are persistent (do not change) physically bounded regions. On average, these regions have a total population of about 4,000, of which about 150 individuals turn 18 years old annually. Tracts have the additional advantage of being locally defined to correspond to true neighborhoods while still strictly respecting city, county, and state boundaries. They are also associated with actual, physically bounded property areas, making them suitable for use by GIS-mapping applications. Roughly two-thirds of the new neighborhoods correspond to a single tract; the rest are formed by combining adjacent tracts based on size and similarity to achieve an optimal number of college-bound students.

These new educational neighborhoods are associated with aggregate socioeconomic data, like traditional geodemographic neighborhoods, but we also have included our proprietary, educationally related information such as academic performance, curricular interests, and college-choice behaviors—all of which significantly enhance value in an enrollment-management context. The essential information represented by the full range of these data elements is then distilled into a smaller group of orthogonal descriptive factors, which in turn, allow the College Board to develop the *educationally relevant* geodemographic neighborhood clusters that comprise the Segment Analysis Service system. These clusters represent unique and relevant characteristics of prototypical college-bound students and their behaviors related to college choice. Yet because college choices are guided not only by the characteristics of the student in the context of his/her neighborhood, but also by the characteristics of the student in the context of his/her high school, Segment Analysis Service offers a second view of college-bound students—from the perspective of their high schools—when constructing its final clustering model.

The high-school perspective replaces the role of the educational neighborhood with each individual student's high-school community, and then uses the same statistical techniques and methods used to construct high-school base clusters, which offer a complementary view to the neighborhood clusters. As with neighborhood clusters, high-school clusters are defined by interacting descriptive factors that begin with academic/curricular indicators, historical patterns

of college choice, and student interests, and are then complemented by additional socioeconomic and mobility variables.

The original 30 neighborhood and high-school clusters have been replaced with updated clusters as part of the 2011 revision. The 33 new educational neighborhood clusters are numbered 51-83, while the 29 new high-school clusters are numbered 51-79. The new numbers differ from the original clusters to avoid confusion for those colleges transitioning to the revised clusters. Detailed migration data showing how students classified in the original system would be reclassified in the new system are also available. The geodemographic tagging that is available as part of our dual-cluster approach provides great value to enrollment managers because it is:

1. Specific to student recruitment and retention within higher education.
2. More powerful, sophisticated, and focused in its clustering than existing commercial systems.
3. Easy to understand and apply to strategic and tactical enrollment-management decisions.

The geodemographic clustering done by Segment Analysis Service allows enrollment managers to identify different types of students that are drawn to each institution and to develop an appropriate set of differentiated strategies, messages, and activities for these students that build on what is known about them through their cluster affiliations. These messages can be used to target and manage appropriate subgroups within the prospective student pool as part of the relationship-building process. More importantly, Segment Analysis Service permits this targeted marketing to be conducted at the earliest stages of each recruitment cycle by knowing only the student's address and high school.

Attributes, Factors, and Clusters

The development of Segment Analysis Service begins with the most detailed pieces of information available to the College Board about college-bound students: their actual responses to a series of detailed, academically relevant questions that are asked when they register for various assessments (PSAT/NMSQT, SAT®, SAT II®, AP®).

This information is unique to each individual and addresses a range of characteristics and tendencies that are directly related to the college-choice process. It forms the core set of data attributes affiliated with each student record in the Segment Analysis Service database. At this point, other relevant data elements, such as individual student test-performance results and individual student test-score sending patterns, are attached to each record, and the records are grouped in two ways:

- **High Schools**: The first grouping is done at the high-school level. The values for the student-level attributes are aggregated and averaged across each of the 33,000+ identifiable high schools in the United States. The result is a database with unique behavioral profiles for each of these 33,000+ schools.

- **Educational Neighborhoods**: The same process is used to group students at the neighborhood level using a combination of physical contiguity *and* educational similarity. The result of the neighborhood grouping is a database with unique behavioral profiles for approximately 44,000 "educational neighborhoods" that cover the United States. Each of these profiles represents a set of college-bound students with similar educational, demographic, and socioeconomic attributes. An individual is located or "geocoded" within a single neighborhood based on their home address.

The high-school and neighborhood groupings are then enhanced with additional attributes, including:

- Detailed characteristics of the colleges to which each prospective student expects to apply.

- Neighborhood-level population characteristics, including educational and occupational attainment, income, housing values and socioeconomic indicators, and population demographics such as age distribution and ethnicity.

In total, over 150 attributes, or individual pieces of information, exist for each high school and educational neighborhood in the Segment Analysis Service database (see Appendix A).

From this point, the enormous amount of data available within the Segment Analysis Service database is refined into something more manageable and more "actionable". A smaller set of descriptive "factors" that—when viewed holistically—represent most of the unique information used to group students at the high-school and neighborhood levels is identified through a process called factor analysis. The result is a set of 40 unique high-school factors and 38 unique educational neighborhood factors that can each be weighted to capture the essential information contained in the much larger set of individual attributes associated with each of the two groupings (See Appendix B for the factor sets).

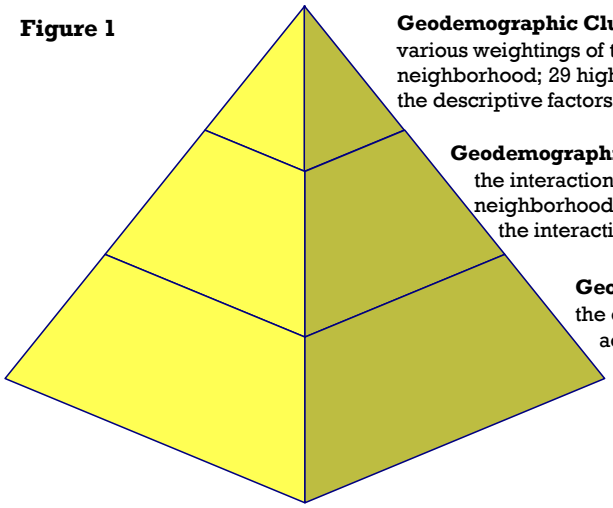
However, successful market segmentation and communication strategies rely on the use of a manageable number of prototypical high schools and a manageable number of prototypical neighborhoods that describe college-choice behaviors. To arrive at this manageable number, the factors are used to group the 33,000+ high schools and 44,000 neighborhoods into 29 unique high-school types and 33 unique neighborhood types referred to as clusters.

The statistical method for this grouping is called cluster analysis. Cluster analysis takes a broad set of characteristics and creates groups (clusters) whose members are more like each other than they are like members of other clusters. Rather than attempting to manage over 33,000 high-school possibilities and over 44,000 neighborhood possibilities, cluster analysis allows enrollment managers to limit each set of possibilities to a reasonable number of prototypical clusters.

Ultimately, each student with a home address can be mapped to one of the neighborhood clusters, and the student's high school can be mapped to one of the high-school clusters, providing detailed insights into his/her likely characteristics and college-bound behaviors.

The development of Segment Analysis Service clusters is illustrated in Figure 1 below:

Figure 1



Geodemographic Clusters: 33 neighborhood clusters emerge from various weightings of the descriptive factors associated with each neighborhood; 29 high-school clusters emerge from various weightings of the descriptive factors associated with each high school.

Geodemographic Factors: 38 descriptive factors are created from the interaction between the attributes within each educational neighborhood and 40 descriptive factors are created from the interaction between the attributes within each high school.

Geodemographic Attributes: 156 unique attributes describe the characteristics and behaviors of college-bound students across 44,000 educational neighborhoods and 33,000+ high schools across the country.

Cluster Characteristics

Calculating the factor scores is a key step in the process of developing the Segment Analysis Service system. Factor scores concisely represent unique dimensions, which when taken together, enable us to sort prospective students into a set of behavioral "types" called clusters. The factor scores bridge the gap between the several hundred individual data elements that comprise the Segment Analysis Service database and the individual cluster into which a specific student is classified. However, like most statistical attributes, they are more useful in mathematical calculations, such as clustering and predictive modeling, than for describing actual people. As an alternative, we have created a set of 51 key cluster characteristics, derived from the individual database elements. These key characteristics describe differences between the clusters in terms that are more understandable to most admissions professionals and span a wide range of areas, including academic performance indicators, socioeconomic and demographic attributes, educational attainment, and college choice. The 33 neighborhood clusters and 29 high-school clusters are listed below along with sample characteristics. The full set of key characteristics is provided in Appendix C.

Neighborhood Cluster Sample Characteristics													
2011 D+ Cluster	SAT Math	SAT CR	Going Out of State	Percent NonWhite	Need Financial Aid	Med Income		2011 D+ Cluster	SAT Math	SAT CR	Going Out of State	Percent NonWhite	Need Financial Aid	Med Income
51	546	533	32%	30%	57%	$ 95,432		68	552	558	52%	35%	65%	$ 57,902
52	480	470	30%	58%	71%	$ 63,578		69	534	521	37%	19%	65%	$ 88,100
53	561	544	32%	50%	55%	$ 92,581		70	613	598	65%	29%	61%	$ 86,381
54	458	443	25%	83%	76%	$ 38,977		71	405	408	39%	97%	68%	$ 42,661
55	566	565	52%	24%	63%	$ 71,576		72	399	397	31%	87%	47%	$ 32,708
56	420	411	29%	93%	66%	$ 35,308		73	528	514	29%	42%	62%	$ 90,849
57	541	519	52%	47%	43%	$ 67,394		74	433	435	29%	84%	79%	$ 44,065
58	533	489	28%	87%	69%	$ 68,213		75	459	457	28%	85%	72%	$ 50,421
59	561	562	52%	24%	74%	$ 54,750		76	514	509	27%	38%	64%	$ 61,332
60	589	590	63%	37%	36%	$ 104,174		77	502	492	26%	18%	75%	$ 62,372
61	585	567	51%	30%	40%	$ 123,858		78	594	578	56%	26%	39%	$ 134,400
62	596	595	67%	24%	72%	$ 59,824		79	550	551	57%	32%	74%	$ 40,909
63	548	541	39%	23%	65%	$ 69,347		80	534	527	39%	39%	65%	$ 49,877
64	466	466	48%	34%	29%	$ 49,829		81	491	483	27%	57%	72%	$ 63,030
65	440	433	23%	93%	78%	$ 45,081		82	496	491	29%	21%	75%	$ 53,465
66	499	492	20%	12%	76%	$ 50,453		83	500	490	19%	26%	71%	$ 49,335
67	519	501	27%	53%	59%	$ 60,960		Total	512	502	32%	43%	65%	$ 70,231

High School Cluster Sample Characteristics														
2011 D+ Cluster	SAT Math	SAT CR	Going Out of State	Percent NonWhite	Need Financial Aid	Med Income		2011 D+ Cluster	SAT Math	SAT CR	Going Out of State	Percent NonWhite	Need Financial Aid	Med Income
51	462	457	14%	33%	68%	$ 40,918		66	498	515	37%	37%	73%	$ 60,272
52	489	496	81%	99%	77%	$ 64,730		67	526	546	48%	41%	69%	$ 71,279
53	471	484	28%	38%	62%	$ 60,833		68	541	540	41%	26%	62%	$ 79,260
54	376	371	33%	96%	38%	$ 38,146		69	390	395	36%	92%	74%	$ 43,391
55	489	481	39%	46%	44%	$ 71,845		70	595	581	56%	33%	48%	$ 105,721
56	536	508	73%	43%	49%	$ 63,967		71	400	412	57%	98%	80%	$ 43,137
57	434	435	29%	82%	79%	$ 48,301		72	528	544	35%	25%	64%	$ 70,018
58	592	577	51%	27%	32%	$ 104,509		73	451	438	24%	89%	76%	$ 48,406
59	499	489	19%	18%	74%	$ 47,685		74	654	579	76%	80%	46%	$ 59,089
60	523	549	23%	30%	33%	$ 70,175		75	514	502	31%	20%	71%	$ 72,850
61	485	370	33%	89%	9%	$ 61,385		76	600	584	72%	50%	28%	$ 90,265
62	474	473	34%	92%	67%	$ 55,515		77	595	508	64%	75%	39%	$ 39,490
63	440	427	28%	86%	72%	$ 49,238		78	473	468	48%	43%	22%	$ 56,703
64	606	542	37%	89%	57%	$ 81,911		79	594	585	61%	26%	71%	$ 65,180
65	515	503	28%	43%	65%	$ 72,692		Total	514	502	32%	44%	65%	$ 70,223

Changes Over Time

As obvious as it may seem, it is important to mention that characteristics of neighborhoods and high schools can and do change over time. At a macro level, this is best illustrated by the population shift across the United States from Northeast to Southwest that began in the last third of the 20th century and continues today.

At the micro, or neighborhood, level, these changes can result in shifts in the values of various academic and socio-economic attributes. As these values shift, from a geodemographic modeling perspective, it is possible that the definitive characteristics and behaviors that emerge from these attributes—reflected through the factor weightings—may also shift. In some cases, it is also likely that, due to the shifting weights, certain educational neighborhoods will actually change enough in their composition to warrant reassignment to a different neighborhood cluster. For example, a specific neighborhood previously might have been affiliated with neighborhood cluster #53 and is now affiliated with a different neighborhood cluster, #72. Therefore, the college-bound students living in that particular educational neighborhood will also change affiliations from neighborhood cluster #53 to neighborhood cluster #72.

This geodemographic shift should make intuitive sense. Over periods of time, neighborhoods become more or less diverse or move up or down in terms of socioeconomic indicators. These shifts typically do not occur quickly, but we are all familiar with neighborhoods that have changed over time. These changes will be reflected in the annual update to the Segment Analysis Service system.

What is most important to note regarding how changes impact the strategic use of neighborhood clusters within Segment Analysis Service is that, in spite of any changes, the *actual definitions of each cluster remain constant*. Therefore, prospective students who are affiliated with each neighborhood cluster today represent the same *types* of students that would have been affiliated with each neighborhood cluster five years ago (and will be affiliated with it five years from now). As a result, the relationship-building actions that make the most sense with regard to students in a particular cluster will always make sense for those students—even though the specific set of educational neighborhoods that are mapped to the particular neighborhood cluster may reflect some minor differences when compared to educational neighborhoods mapped to the same neighborhood cluster five years ago.

It is also important to note that these changes do not impact students in historical datasets that have previously been assigned to clusters. If you are conducting trend analysis based on clusters, the students that were identified five years ago as part of a particular cluster are directly comparable to students in your database today identified as part of that

same cluster. This is true regardless of changes to the specific educational neighborhoods affiliated with the cluster over time.

By regularly updating the values for individual attributes that define neighborhoods and high schools, the College Board is able to monitor shifts that occur within and across the Segment Analysis Service database. These relatively minor changes are reflected through updated text descriptions that are associated with each of the neighborhood clusters. The algorithms used to map student records to both high-school and neighborhood clusters will also be updated to accommodate any shifts in the percentage of college-bound students represented by some of the neighborhood clusters. Obviously, the number of prospective students affiliated with each neighborhood cluster will change from year to year. This is due primarily to the fact that the size and the geographic distribution of the overall college-bound cohort will change from year to year. Therefore, the actual number of college-bound students that can be mapped to the set of neighborhood clusters within Segment Analysis Service will vary each year. In addition, the changes referenced above will also result in some shifts in the mapping of some educational neighborhoods to neighborhood clusters. Therefore, from an institutional perspective, it should not be surprising to see differences in the relative number of prospects, applicants, admits and matriculants who are affiliated with each of the Segment Analysis Service neighborhood clusters from year to year.

At a tactical level, there should be no changes needed on the part of colleges to account for the impact of time on the geodemographic modeling in Segment Analysis Service. Institutions will want to continue targeting the same people, for the same reasons, using the same messages—assuming they work—from year to year. Obviously, it is important to evaluate the effectiveness of all tactical marketing efforts on a continuous basis and make changes when warranted. This is also true of efforts that use Segment Analysis Service information.

Using Segment Analysis Service Information

Segment Analysis Service is unique in its ability to deliver information that describes student characteristics on the basis of academic, socioeconomic, and student-interest attributes collected by the College Board and supplemented with data collected through the U.S. Census. This type of descriptive information can be very valuable at the beginning of the recruitment process when nothing is known about a prospective student other than his or her name, address, and high school.

Neighborhood cluster affiliations and high-school cluster affiliations help enrollment managers understand the types of students that are currently attracted to an institution, and identify those students the institution would like to recruit in the future. By attaching clusters to student records as they move from prospect to enrolled student status, enrollment managers can more effectively assess which types of students persist through the various stages of the recruitment process—and even understand more about students who persist at an institution once they enroll.

The descriptive information about each neighborhood and high-school cluster allows enrollment managers to segment and target their recruitment efforts at prospective students on the basis of educationally relevant characteristics and likely college-choice behaviors. This saves an admission office staff time and money by allowing them to focus their efforts on those students who will not only help make the class, but also help the institution realize specific enrollment goals within the context of its mission.

For example, most institutions are interested in targeting more high-achievers, more economically secure students, and/or more qualified minority students. Through the use of Segment Analysis Service neighborhood and high-school clusters, an enrollment manager can easily identify and target these students at the beginning of the recruitment

process by using Segment Analysis Service as a way of qualifying the license of names from Student Search Service® and building a subsequent contact sequence and communication flow that is driven, in part, by cluster affiliations. Special messages, brochures, and contacts from key faculty or staff can be directed at those students who are viewed as "desirable" on the basis of the goals the institution has established.

To allow enrollment managers to interpret and apply Segment Analysis Service geodemographic tagging in an efficient, easy-to-use manner, the College Board has created two manageable sets of geodemographic clusters represented by codes, rather than arbitrary descriptive titles, which can be applied to prospect records in an institutional database and/or used as additional parameters to license student names through the College Board's Student Search Service®. There are 33 unique educational neighborhoods that reflect unique living environments and 29 high-school clusters that reflect unique learning environments (e.g., high school). Each of these clusters is numbered beginning with 51 and defined through a cluster profile.

There are four sections to each cluster profile: The text description of students represented by the cluster, a set of values and rankings for key attributes used to define the cluster, four of the most dominant factors that influenced the clustering, and a map illustrating the distribution of the cluster throughout the country. Figure 2 below provides an example of one of the educational neighborhood profiles and one of the high-school cluster profiles available in Segment Analysis Service.

Figure 2

Neighborhood Cluster 51

Residents of this neighborhood have relatively high incomes and almost always own their homes which are valued well above average. They are moderately diverse, hold professional and managerial jobs, and most have at least some college with many having graduate degrees. Students attend primarily public high schools, avail themselves of AP/honors course work, and have above average scores on standardized tests. They prefer colleges in state and, although interested in at least one public, will generally apply to a number of modestly selective privates where financial aid will be sought.

Values & Rankings of Key Attributes

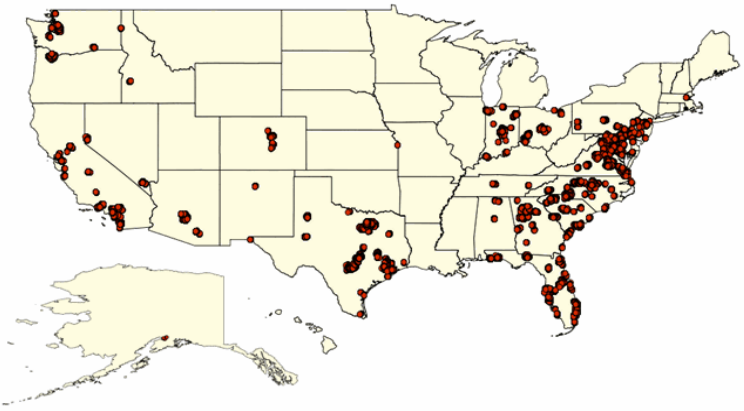
	value	rank		value	rank		value	rank
Median family income (x $1000)	$95.4	4	% of Population college-aged	8%	8	Mean SAT Critical Reading Score	533	12
% Adults w/ professional jobs	50%	7	% of Students 1st generation	30%	27	Mean SAT Math Score	546	12
% Speaking English only	85%	16	% likely to apply out of state	32%	17	Mean SAT Writing Score	522	12
% of Population non-White	17%	16	% interested in Financial Aid	57%	26	Ave cost targeted colleges (x $1000)	$12.8	25

Dominant Cluster Factors

College Prep Culture | Professional and Affluent | Highly Educated | Coed

Distribution of Neighborhood

Number of Neighborhoods	1427
% of All Neighborhoods	3.20%
% 18-21 year olds	3.90%



High School Cluster 51

These high schools are predominantly public and serve traditional, blue-collar communities with very low home values. Families are mature and own their homes but have relatively low incomes. Students often will be the first in their family to graduate from college and have modest curricular preparation, below average test scores, and low degree aspirations. They submit relatively few applications and set their sights on low cost, less selective institutions and local community colleges within their home state. Many will be applying for financial aid, particularly if they are going away to school.

Values & Rankings of Key Attributes

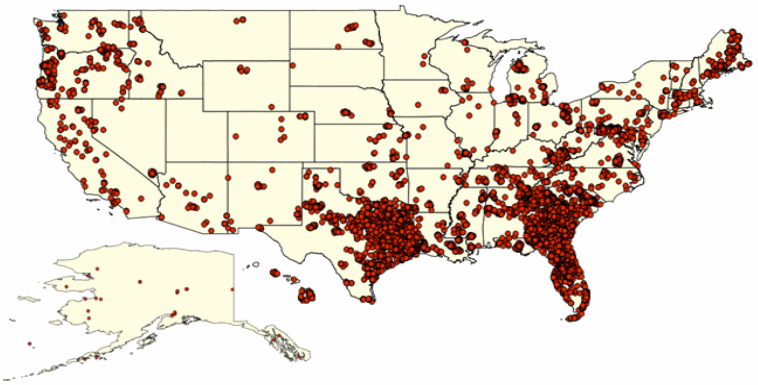
	value	rank		value	rank		value	rank
Mean SAT Critical Reading Score	457	22	% of Students 1st generation	71%	7	Ave cost targeted colleges (x $1000)	$8.3	29
Mean SAT Math Score	462	23	Ave Admit Rate at Targeted Colleges	61%	3	% of students non-White	33%	22
Mean SAT Writing Score	445	22	Ave Number of AP Exams per Student	0.91	22	% of families below poverty	14%	8
Ave Number of Advanced Courses	0.55	18	% likely to apply out of state	14%	29	% interested in Financial Aid	68%	12

Dominant Cluster Factors

Focused/Early Decision | Few AP/Honors | College Interest: Local Technical | Cost Not an Object

Distribution of High Schools

Number of High Schools	859
% of All High Schools	2.56%



Segment Analysis Service cluster tags can easily be attached to historical data files for prospects, applicants, admits, and matriculants to examine trends across the enrollment funnel by cluster affiliation(s) and identify patterns of behavior that can be encouraged or discouraged at various points in the recruitment process. In particular, colleges find it very helpful to use Segment Analysis Service tags as a way of segmenting their early communication with students as they attempt to build an appropriate prospect pool and a desirable applicant pool.

When doing this, it is important to remember that the cluster profile represents the *average student* in the cluster. Therefore, some students in the cluster will have attribute values that are above the values that define the cluster, while other students in the cluster will have attributes below the values that define the cluster. Because of this, it is possible that talented students can emerge from average clusters—and vice versa.

In some cases, an institution—by virtue of its location and historic mission—may draw heavily from clusters that they would not be likely to target when recruiting high-achievers and/or economically secure prospects and still successfully address enrollment objectives related to these metrics. This requires that the institution be able to identify the top students from within average clusters—potentially through the use of additional Student Search Service® parameters—and then be able to communicate in an appropriate manner that respects their unique academic strengths while simultaneously recognizing other aspects of their cluster affiliations that are consistent with their peers. This is particularly true for institutions that may be located in older cities or rural communities where clusters with lower academic and socioeconomic attributes—and in some cases very little diversity—may dominate the landscape within a 150-mile radius of the institution.

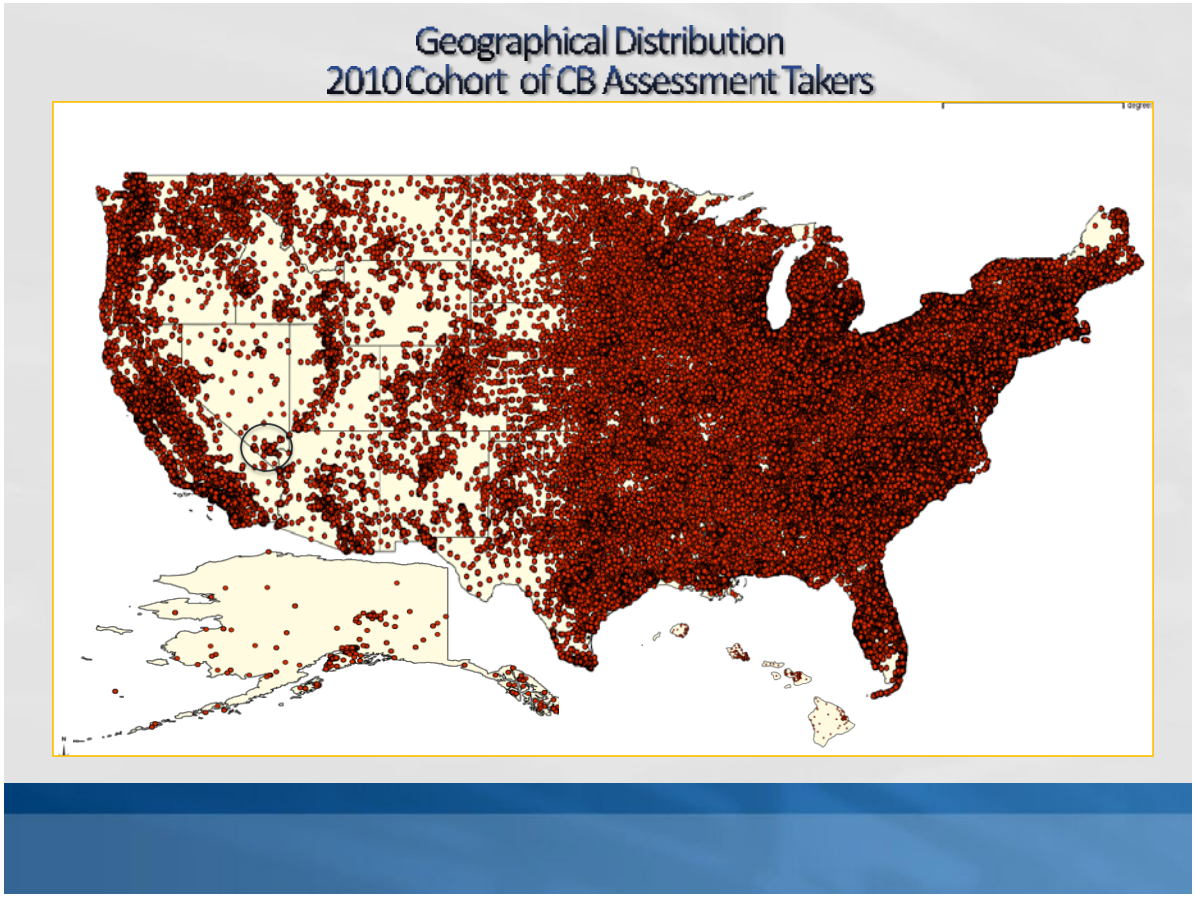
Finally, keep in mind that, when predicting the likelihood of a student attending and succeeding at an institution, actual student behaviors almost always tend to be the best indicators of success. While the descriptive information in the Segment Analysis Service clusters is very valuable, especially early in the process when very little is known about a student, the actual behavior of the student as s/he moves through the process becomes progressively more relevant. It

is important to understand which specific behaviors are important for your institution (i.e., visiting the campus, responding to certain requests, coming to a college night, etc.) and how they impact conversion and yield rates across various sub-cohorts.

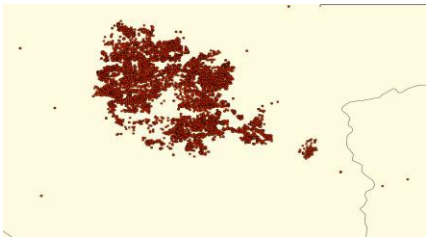
Mapping and the Geographical Distribution of Clusters

The 2011 revision was more extensive than prior annual revisions. In addition to updating the underlying data with the most current information, we replaced the Census-related data with information from the American Community Survey (ACS). The ACS is similar to Census data, but will updated annually as opposed to the every ten years of the decennial Census, thus allowing us to better track and adjust for demographic shifts as they occur. We also defined new neighborhoods and revised and renamed the clusters. Finally, the system was redesigned from the ground up to support sophisticated mapping and visualization applications. These new visualization and mapping capabilities developed for Segment Analysis Service can be employed at all levels of analysis--from the whole nation to down individual states, geomarkets, counties, cities, and the new Segment Analysis Service neighborhoods. The following examples illustrate the type of information that might be delivered through Segment Analysis Service.

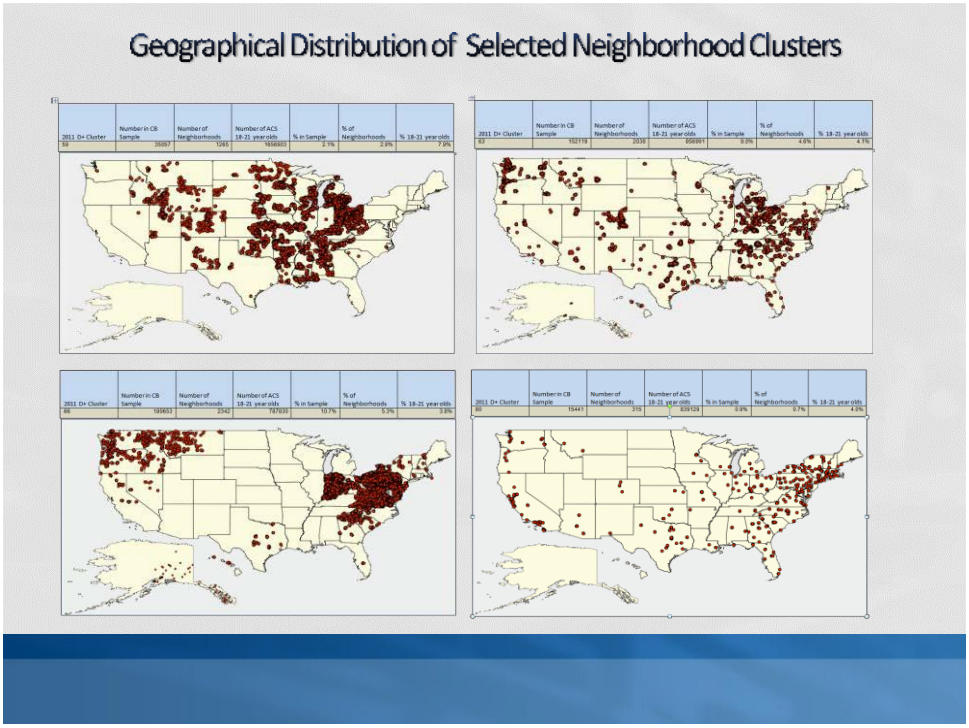
At the most global level, it is useful to consider a map of individual students that have been drawn from 2010 Cohort of College Board assessment takers and tagged with both cluster and location information. Not surprisingly, as with the general population, the density is highest in the East and along the West Coast. But, there are also large numbers of students through the Midwest and even substantial numbers in the relatively unpopulated Western mountain and desert regions.

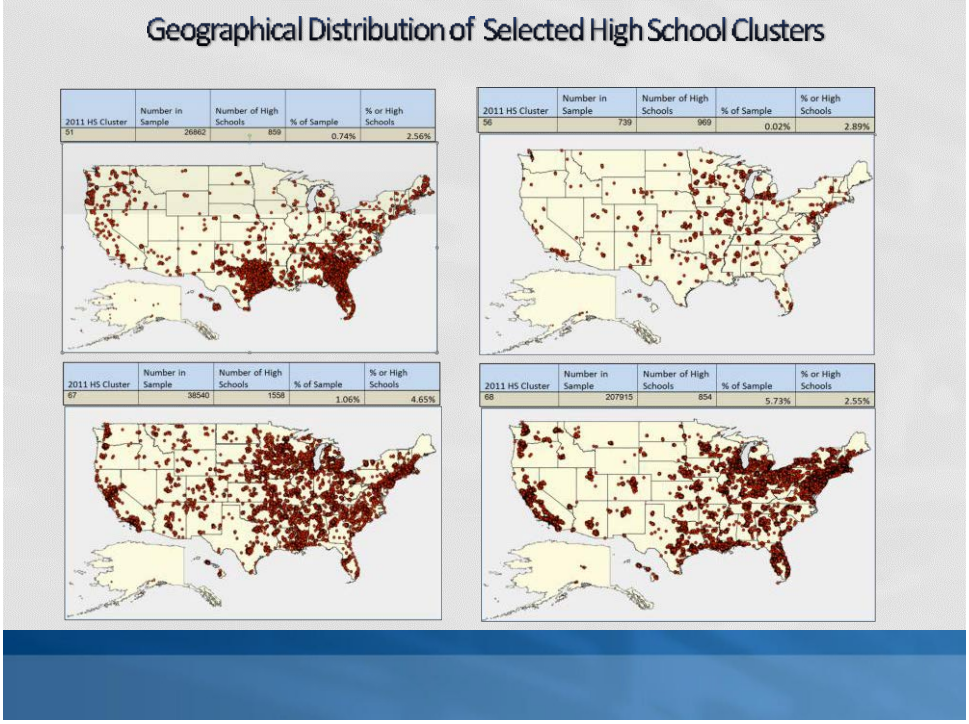


This overall map provides a sense of the distribution of students across the country. However, an important concept when working with mapping data is that of scale. The information that can be depicted on a map is constrained by the scale. If one looks at a MapQuest map of the whole United States, it is impossible to pick out individual houses or, for that matter, street details within cities. In order to observe the finer detail, one must adjust the scale by zooming in. Likewise, since there are several million students being represented in a comprehensive map of all College Board assessment participants, each dot necessarily corresponds to multiple students. If we drill down on an area that appears to be relatively sparse, we will see that there are actually more students that it may seem at first. For example, consider Las Vegas (the area surrounded by the blue circle above). On the national map, it looks like there are only 15 or 20 dots, but if we zoom in we see that in fact there are actually several hundred students being represented by fewer dots at this larger scale.

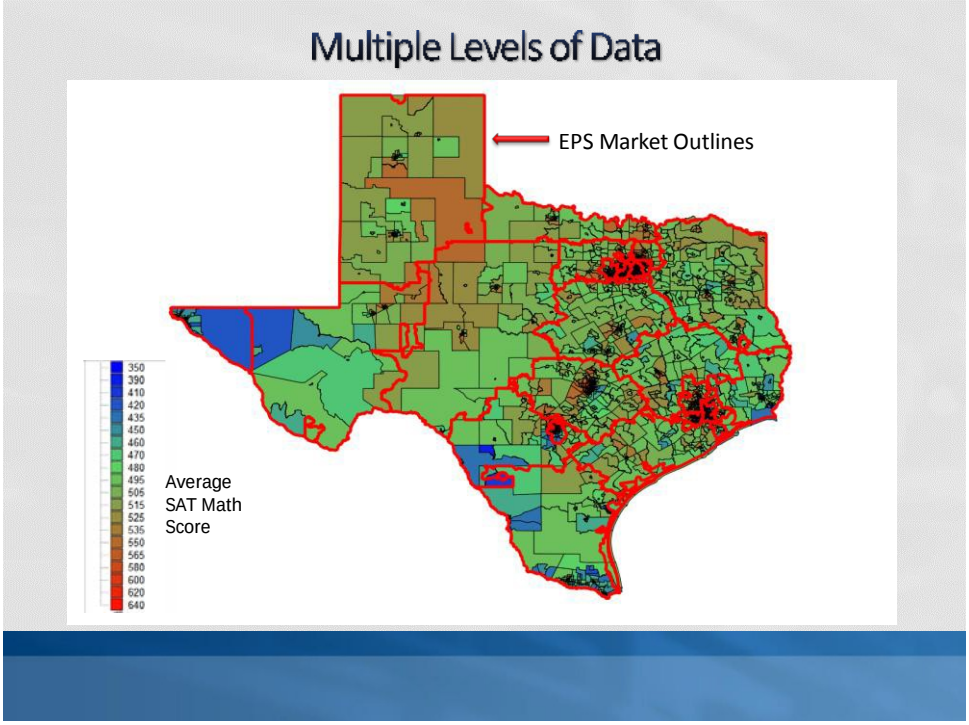


The ability to display different types of information at different scales is useful component of GIS-compatible systems such as Segment Analysis Service. For example, at a national level, we might be most interested in how different clusters distribute themselves differently across the country. This might be one strategy to help identify new geographical areas where we might look to find students that are most similar to groups that we know are well-suited to our campus environment.

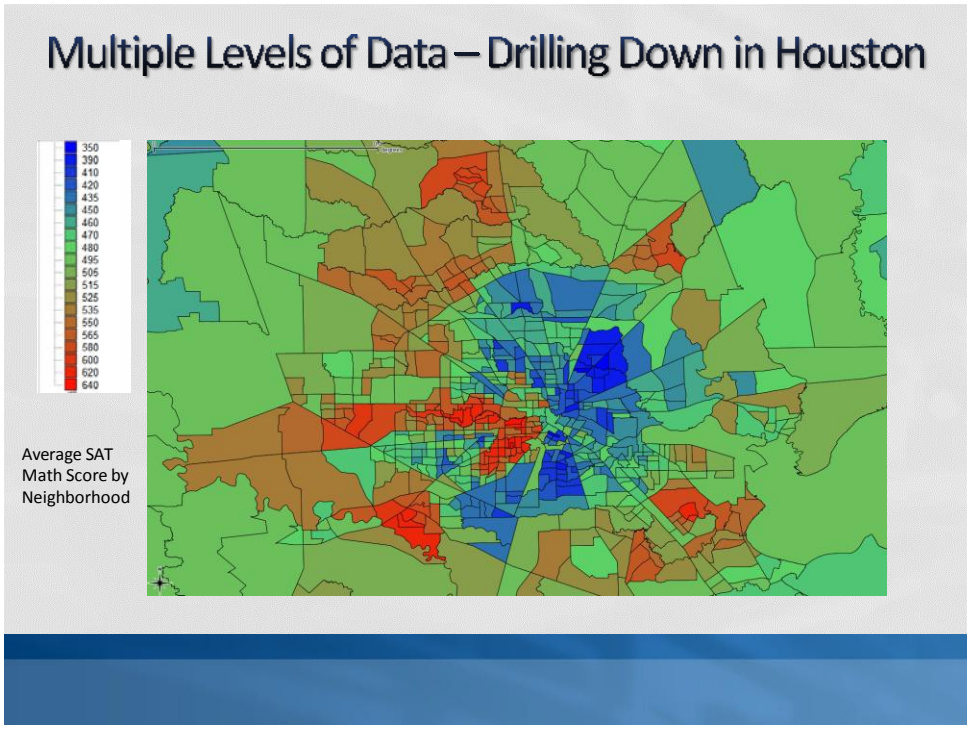


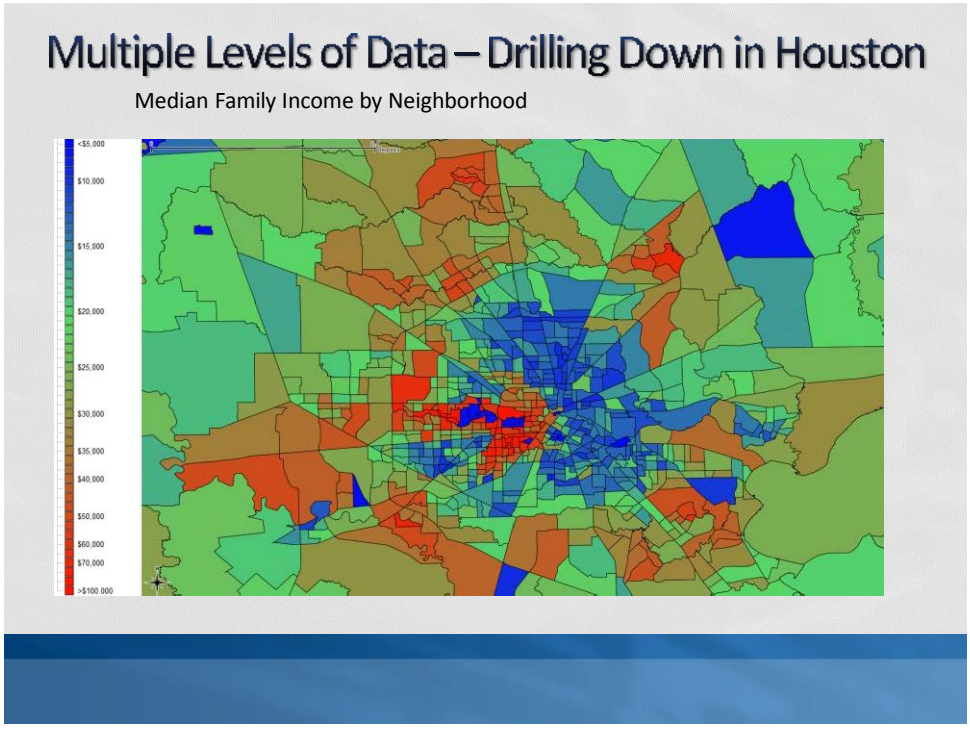


Once we identify a particular state--Texas for example--we can begin to drill down into particular regions. At the state level, EPS markets provide some guidance. However, we quickly notice that there is significant variability on key attributes within each large market. This variability is of interest and we might want tailor our messaging and recruitment activities to submarkets based on local attributes in order to achieve a better match with the interests and capabilities of prospects.



In less populated suburban and rural areas, EPS submarkets might be as far we need to go. But in urban areas, we might want to fine-tune our targeting even more and consider differences between specific neighborhoods (or high schools) when developing communications or planning recruitment travel as a way to optimize our resources. In principle, mapping and GIS analysis tools could utilize any of the factors in Appendix B, or the key characteristics in Appendix C, providing a wealth of actionable information that builds on, but goes beyond, the basic cluster assignment.

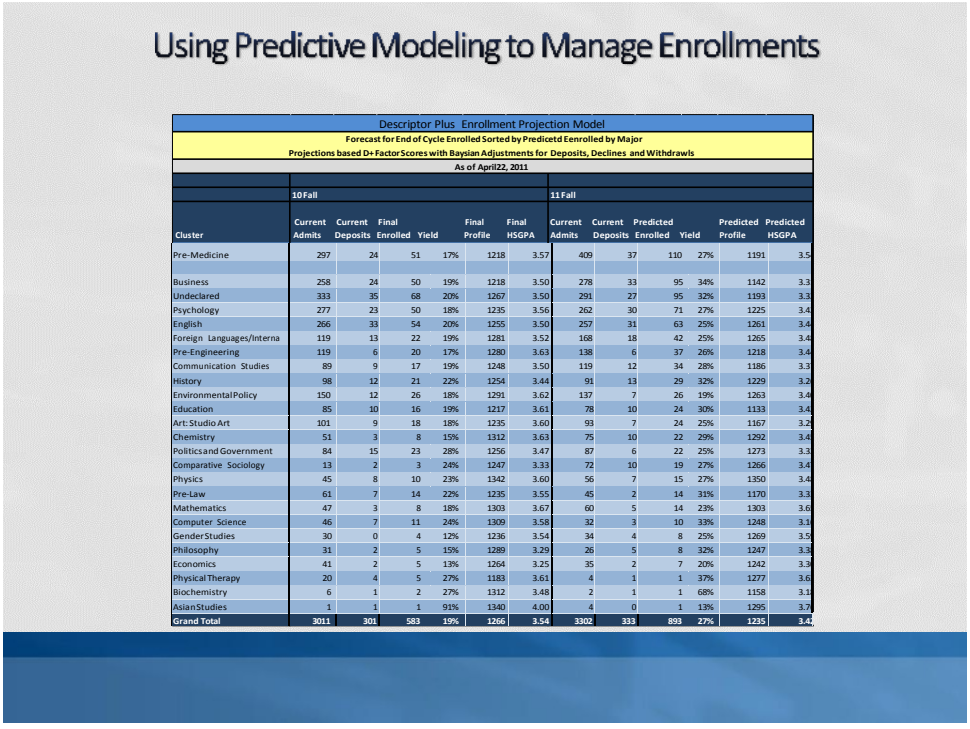




Beyond Description

Segment Analysis Service, as the name implies, is primarily a qualitative tool for describing neighborhoods, high schools, and prospective students. However, within the Segment Analysis Service system lays a rich set of quantitative attributes including the factor scores and key characteristics. These attributes can be associated with an individual prospect at high degree of specificity: Each of the 44,000 neighborhoods and 27,000 highs schools has its own unique set of characteristics. A group of prospects all within the same cluster then can be tagged with a set of characteristics that more uniquely describes each individual. These student-level characteristics then serve as continuous predictors in a statistical modeling process that calculates the probability of given student making a given enrollment-related decision. For example, the probability that a prospect inquiry will become an active inquiry, an inquiry will convert to an application, an admitted student will choose to enroll, or enrolled student will be retained.

A full discussion of predictive modeling is beyond the scope of the current paper, and the College Board does not currently offer modeling services. However, many Segment Analysis Service customers have worked with independent consulting groups to leverage the Segment Analysis Service system in their modeling services. Their experience demonstrates that the Segment Analysis Service is value tool in predicting the likelihood that prospects will actively engage in the recruitment relationship. When the models are normalized, the resulting probabilities have successfully predicted yield outcomes, and waitlist requirements that ultimately translate into a predicted class profile. Such a predicted profile can include academic qualifications and financial aid demand, as well as demographic distributions such as gender, ethnicity, and major, among others.



Descriptor Plus Enrollment Projection Model

Forecast for End of Cycle Enrolled Sorted by Predicetd Eenrolled by Major

Projections based D+ Factor Scores with Baysian Adjustments for Deposits, Declines and Withdrawls

As of April22, 2011

	10 Fall						11 Fall					
Cluster	Current Admits	Current Deposits	Final Enrolled	Yield	Final Profile	Final HSGPA	Current Admits	Current Deposits	Predicted Enrolled	Yield	Predicted Profile	Predicted HSGPA
Pre-Medicine	297	24	51	17%	1218	3.57	409	37	110	27%	1191	3.5
Business	258	24	50	19%	1218	3.50	278	33	95	34%	1142	3.3
Undeclared	333	35	68	20%	1267	3.50	291	27	95	32%	1193	3.3
Psychology	277	23	50	18%	1235	3.56	262	30	71	27%	1225	3.4
English	266	33	54	20%	1255	3.50	257	31	63	25%	1261	3.4
Foreign Languages/Interna	119	13	22	19%	1281	3.52	168	18	42	25%	1265	3.4
Pre-Engineering	119	6	20	17%	1280	3.63	138	6	37	26%	1218	3.4
Communication Studies	89	9	17	19%	1248	3.50	119	12	34	28%	1186	3.3
History	98	12	21	22%	1254	3.44	91	13	29	32%	1229	3.2
Environmental Policy	150	12	26	18%	1291	3.62	137	7	26	19%	1263	3.4
Education	85	10	16	19%	1217	3.61	78	10	24	30%	1133	3.4
Art: Studio Art	101	9	18	18%	1235	3.60	93	7	24	25%	1167	3.2
Chemistry	51	3	8	15%	1312	3.63	75	10	22	29%	1292	3.4
Politics and Government	84	15	23	28%	1256	3.47	87	6	22	25%	1273	3.3
Comparative Sociology	13	2	3	24%	1247	3.33	72	10	19	27%	1266	3.4
Physics	45	8	10	23%	1342	3.60	56	7	15	27%	1350	3.4
Pre-Law	61	7	14	22%	1235	3.55	45	2	14	31%	1170	3.3
Mathematics	47	3	8	18%	1303	3.67	60	5	14	23%	1303	3.6
Computer Science	46	7	11	24%	1309	3.58	32	3	10	33%	1248	3.1
Gender Studies	30	0	4	12%	1236	3.54	34	4	8	25%	1269	3.5
Philosophy	31	2	5	15%	1289	3.29	26	5	8	32%	1247	3.3
Economics	41	2	5	13%	1264	3.25	35	2	7	20%	1242	3.3
Physical Therapy	20	4	5	27%	1183	3.61	4	1	1	37%	1277	3.6
Biochemistry	6	1	2	27%	1312	3.48	2	1	1	68%	1158	3.1
Asian Studies	1	1	1	91%	1340	4.00	4	0	1	13%	1295	3.7
Grand Total	3011	301	583	19%	1266	3.54	3302	333	893	27%	1235	3.4

Conclusion

Segment Analysis Service provides an institution with valuable market data from the College Board that can be used to target students, shape marketing communications, and improve the return-on-investment of overall recruitment effort. This information, presented through geodemographic clusters, is extremely valuable during the early part of each recruitment cycle when very little information has been collected from or about prospects.

The Segment Analysis Service high-school clusters and educational neighborhoods are comprised of complex data factors built on academic, socioeconomic, and student-interest attributes collected annually by the College Board and supplemented by socioeconomic and demographic data collected through the U.S. Census. These clusters describe the population of college-bound students from two perspectives: where they live (educational neighborhood) and where they learn (high-school clusters).

Educational neighborhoods and high-school clusters help institutions understand and visualize the students they are currently attracting, and identify the types of students they would like to attract in the future. The cluster tags can be attached to current prospect data files to allow for segmentation and targeting on the basis of relevant college-choice characteristics and likely behaviors—from the very beginning of the recruitment cycle. Admissions officers can save time and money by focusing their efforts on those students who will not only help make their class from a quantitative perspective, but also help realize specific qualitative enrollment goals developed in support of the institution's mission.

Segment Analysis Service clients can also use educational neighborhoods and high-school cluster information as additional criteria when defining the names they license. This enables institutions to more effectively manage their student list licenses and to enhance response rates by developing more effective, targeted messages that build on the likely characteristics and behaviors of the students within sets of related clusters.

Appendix A

Segment Analysis Service™ Data Attributes

Attributes Used to Construct Educational Neighborhood Clusters:

Census
- Median age
- Median Household Income
- Median Household Size
- Per-capita Income
- Year-moved-in
- % Age 12-18
- % Age 18-21
- % Blue Collar
- % Speak 2nd Language
- % High School Ed or Less
- % College Ed or More
- % English Only
- % Families
- % Moved Recently
- % Own Home
- % Professional
- % Private School
- % Public School
- % White

College Aspirations
- % Accepted
- Average SAT Math
- Average SAT Verbal
- Degree Goal
- Median College Cost
- Number of Colleges Targeted

Student Characteristics
- Honors Classes
- Honor Society
- Leadership Activities
- Science Activities
- Work Activities
- Fraternity/Sorority
- Religious Activities
- No Activities

Academic
- Cumulative GPA
- SAT Math
- SAT Verbal
- Avg. SAT II
- AP test(s) taken
- Avg. grade: English
- Avg. Grade: Foreign
- Avg. Grade: Math
- Avg. Grade: Natural Science
- Avg. Grade: Social Science
- Avg. Grade: Art & Music
- Self-reported Math Ability
- Self-reported Science Ability
- Self-reported Writing Ability

High-School Experience
- % PSAT-takers: Junior
- % PSAT-takers: All
- % SAT takers
- Total HS Enrollment
- Art Experience
- Art Activities
- Athletic Activities
- Computer Experience
- Public School
- Years of Art
- Years of Social Science
- Years of Math & Science

Ethnic Mix
- % Asian
- % African American
- %Hispanic
- % Caucasian
- Ethnic Activities
- English as a Second Language
- Plan to take ESL Courses

College Choice
- Medium city
- Small City
- Suburban
- Rural
- Bordering State
- Large City/Metro Area
- Close-to-home
- Coed
- % Out-of-State
- % Private
- % Religious
- Avg. Distance from Home
- % Two-year
- Enrollment Size (undergrad)

Parental Income/Education
- Father's Education
- Mother's Education
- Parental Income
- Parent Ed Level I
- Parent Ed Level II
- Need for Financial Aid

ACS Demographics
- Population Density
- College Attainment
- Educational Attainment
- Home Ownership
- Mobility
- Ethnicity
- Citizenship
- Age Distribution
- Per Capita Income
- Median Family Income
- Percent Below Poverty Line
- Percent College Age

Attributes Used to Construct High-School Clusters:

Academic
- Cumulative GPA
- SAT Math
- SAT Verbal
- Avg. SAT II
- % College-bound HS
- AP test(s) taken
- Avg. Grade: English
- Avg. Grade: Social Science
- Avg. Grade: Natural Science
- Avg. Grade: Math
- Avg. Grade: Foreign
- Avg. Grade: Art & Music
- Self-reported Science
- Self-reported Math Ability
- Self-reported Writing

Parental Income/Education
- Father's Education
- Mother's Education
- Parental Income
- Parent Ed Level I
- Parent Ed Level II
- Need for Financial Aid

College Aspiration
- % Accepted
- Average SAT Math
- Average SAT Verbal
- Degree goal
- Median College Cost
- Number of Colleges

High-School Opportunity
- Art Experience
- Art Activities
- Athletic Activities
- Computer experience
- % PSAT-takers: junior
- %PSAT-takers: All
- Public School
- % SAT-takers
- Total HS Enrollment
- Years of Art
- Years of Social Science
- Years of Math & Science

Ethnic mix:
- % Asian
- % African American
- % Hispanic
- % Caucasian
- Ethnic Activities
- English as a Second
- Plan to take ESL Courses

Student Characteristics
- Honors Classes
- Honor Society
- Leadership Activities
- Science Activities
- Work Activities
- Fraternity/Sorority
- Religious Activities
- No Activities

College Choice
- Medium City
- Small City
- Suburban
- Rural
- Bordering State
- Large City/Metro Area
- Close-to-home
- Coed
- % Out-of-State
- % Private
- % Religious
- Avg. Distance from Home
- % Two-year
- Enrollment Size

ACS Demographics
- Population Density
- College Attainment
- Educational Attainment
- Home Ownership
- Mobility
- Ethnicity
- Citizenship
- Age Distribution
- Per Capita Income
- Median Family Income
- Percent Below Poverty Line
- Percent College Age

Appendix B

Factor Scores

Neighborhood-Related Factors			
Academic 1	High School Grades a Strength	HS Opportunity 4	Math Science Curriculum
Academic 2	Good Standardized Testers	SES 1	High Educational
Academic 3	Consistently Excels Academically	SES 2	High Need
Academic 4	Content Area Strengths	Student Achievement 1	Academic Orientation
Demographic 1	Christian Orientation	Student Achievement 2	Activist/Community Oriented
Demographic 2	Catholic Orientation	Student Achievement 3	Work/vocational Orientation
Demographic 3	Jewish Orientation	Student Achievement 4	Leadership/Organizational Activities
Demographic 4	Academic Orientation	Student Achievement 5	Athletics
Demographic 5	Single Gender	Student Achievement 6	Arts Activities
College Application 1	Private Selective	College Interest 1	National Selective
College Application 2	Sectarian	College Interest 2	Flagship Public
College Application 3	Public/Nonselective	College Interest 3	Large Urban
Diversity 1	Hispanic/Mexican ESL	College Interest 4	Single Gender
Diversity 2	Primarily African-American	ACS 1*	Professional and Affluent
Diversity 3	Large Asian ESL population	ACS 2*	Traditional Blue Collar
Diversity 4	Puerto Rican/Caribbean ESL	ACS 3*	White Suburban
HS Opportunity 1	College Prep Curriculum	ACS 4*	Urban Families
HS Opportunity 2	Liberal Arts Curriculum	ACS 5*	Rural Stability
HS Opportunity 3	HS AP/Honors Curriculum	ACS 6*	Older/Few College Age Children
* Tract Level Factors			

High School-Related Factors			
Academic 1	Good Standardized Testers	HS Opportunity 3	Religious Curriculum
Academic 2	High School Grades a Strength	HS Opportunity 4	Liberal Arts Curriculum
Academic 3	High Academic Aspirations	HS Opportunity 5	Math Science Curriculum
Academic 4	Content Area Strengths	SES 1	High Educational
Demographic 1	Christian Orientation	SES 2	High Need
Demographic 2	Catholic Orientation	Student Achievement 1	Activist/Community Oriented
Demographic 3	Jewish Orientation	Student Achievement 2	Leadership/Organizational Activities
Demographic 4	Single Gender	Student Achievement 3	Arts Activities
Demographic 5	Academic Orientation	Student Achievement 4	Athletics
College Application 1	Private Highly Selective	Student Achievement 5	Religious Activities
College Application 2	Selective Cost Conscious	Student Achievement 6	Ethnic Activities
College Application 3	Sectarian	College Interest 1	Private Selective
College Application 4	Focused/Early Decision	College Interest 2	Flagship Public
Diversity 1	Large Asian ESL population	College Interest 3	Sectarian
Diversity 2	Primarily African-American	College Interest 4	Single Gender
Diversity 3	Hispanic ESL	ACS 1*	Professional and Affluent
Diversity 4	Mexican ESL	ACS 2*	Traditional Blue Collar
Diversity 5	Puerto Rican/Caribbean ESL	ACS 3*	White Suburban
HS Opportunity 1	College Prep Curriculum	ACS 4*	Urban Families
HS Opportunity 2	HS AP/Honors Curriculum	ACS 5*	Rural Stability

Appendix C

Key Characteristics

Selected Program Derived Values Available at Neighborhood/High-School Level		Selected ACS Values Available at Tract Level
SAT Math Score	% Expecting to Attend Public College	% of Population College Age
SAT Critical Reading	% Expecting to Attend Private College	% Households with Children
SAT Writing	% Parents High School Only	% White
% Attending Out-of-State College	% Parents College Graduates	% Black
% Asian	% Parents with Grad Degree	% Hispanic
% White	Student Estimated Parent Income	% Bluecollar
% Hispanic	Student Goal BA Only	% Professional
% Black	Student Goal MA	% Moving in past 5 years
Number of AP Courses	Student Goal PhD	% Speaking English Only
% Attending Private High School	% of Applications to Sectarian College	% Management
% Attending Church-Affiliated High School	Average Admit Rate for College Applications	Median Family Income
% Applying for Financial Aid	Average Cost for College Applications	% Below Poverty Line
Number of Honor Courses	% of Applications to 4-Year College	Median House Value (% of State Avg)
% Christian	Average Test Scores at Colleges Applied	% Home Owners
% Catholic	Average Retention at Colleges Applied	
% Female	Average Number of Apps at Colleges Applied	
Average Number of College Applications	% Receiving Pell at Colleges Applied	
% English as a Second Language	Average Loan at Colleges Applied	
% Expecting to Attend Sectarian College		