Demographic and Economic Information Applied to Business Decision Making[1]

Garry S. Meyer
Donnelley Marketing Information Services
Stamford, Connecticut

ABSTRACT

With the release of data from the 1980 Census the demand for timely application of this information for management decision systems has rapidly increased. The work that we report on here involves the integration of virtually millions of individual statistics from both the 1970 and 1980 Census into a system which allows retrieval of information for user-defined areas of interest. These areas are constructed through a user-friendly interface which allows either existing geo-political areas to be selected and combined or geometric shapes defined (for example, a circle describes all the area within a fixed distance from a specified point). There are a wide range of applications both public and private sector, for such information. This paper will focus, however, on the utilization of this system for multi-site consumer outlets where the areas defined approximate market areas. SAS programs and IBM TSO procedures are described which allow the demographic and economic characteristics of a user's series of current facilities to be merged with actual client data. Types of statistical analysis and modeling that follow from this capability are then discussed.

BACKGROUND

Business decision making has changed radically over the past several decades with the rate of change increasing to unprecedented highs. Two keys have been responsible for unlocking the door to advanced scientific business decision making, the development of high speed computers with their associated operating and application software (and particular statistical systems such as SAS), and the massive data collection efforts in both public and private sectors. Computer and software advances have meant that previously unattainable information systems could be built at a reasonable cost. The availability of data such as the 1980 Census statistics and proprietary data bases such as Donnelley Marketing's Residential Data Base have provided the other critical component. With the availability of these systems it has become a fact of business life that the ability to stay on the leading edge of a dynamic and rapidly changing market place will spell success or failure for the business executive in general, and most especially for today's marketing managers.

In this paper we discuss how demographic and economic information is used to assist marketing executives in analyzing and then reacting to changing market conditions for products and services in the retail trade sector. The tools and data bases discussed are by no means limited to this particular application, and indeed underlie information requirements of all levels of corporate decision making including top management.[2] Executives of retail establishments have several different problems that they face. They must look at existing locations to see what can be done to increase profits. This will require an understanding of current demographic characteristics within an outlet's trade area in order to be able to make decisions involving the type of merchandise lines and their display. It will also include strategic planning, sales promotion, target marketing, etc. Existing locations often provide the best available source of information to utilize for choosing other locations which will yield good profit potential. Analysis of existing locations provides evidence for relocating or closing these existing sites. In a franchising environment, such analysis may be critical in order to legally establish the firm's right to site new establishments in a marketplace. For each of these applications, accurate and timely small area demographics are an essential component of any analysis.

Donnelley Marketing Information Services, a company of the Dun & Bradstreet Corporation, first became involved in business information support as a result of their development of small area demographic systems. For more than fifty years, Donnelley has demonstrated an ability to produce accurate information products. Utilizing extensive resources, both personnel and equipment, Donnelley's Geographic Division has produced and maintained the country's most complete Address Coding Guide (ACG) that identifies streets, street segments, address ranges, and other pertinent small area data. The ACG enables Donnelley to maintain the largest commercially available residential data base in the United States, with over 72,000,000 individual households.

Building on Donnelley's and Dun & Bradstreet Corporation's expertise and resources, products and services have been developed to meet the demographic information requirements of business, government, and non-profit organizations. These include on-line systems such as AmericanProfile℠, MarketPotential℠, CensusPlus℠, and X/PROFILE℠, as well as other services such as ClusterPlus℠ marketing, Market Profile Analysis (MPA℠), and GraphicProfile℠ (thematic mapping for Census tracts, counties, states, Standard Metropolitan Statistical Areas (SMSA's) etc.). AmericanProfile℠, for example, provides access to basic Census demographics, as well as to Donnelley proprietary demographic updates and projections by client specified areas of interest. These areas may be described via an easy to use interactive computer system (Meyer, 1981); they allow for definition by geo-political and Census areas: States, Counties, Minor Civil Divisions, Census Places, Census Tracts, etc., by marketing areas: ADI's (Areas of Dominent Influence), Metro Markets, SAMI's, DMA's, etc, and by geometrically defined areas (circles, complete polygon shapes, travel contours, etc.) as required by application.

A typical analysis starts out with data from existing retail outlets along with key characteristics as established by the firm. These will certainly include revenue and profit, and may also include both quantitative and qualitative data such as proximity to major highway access roads, size of store, size of inventory, distance to competitors, type of location (standalone, shopping center, mall, etc.), types of merchandise sold, average dollar amount per sale, number of employees, sales employee's man hours, parking availability, and even characteristics of store personnel such as their experience and educational background. The actual market areas served by each store are either established by utilizing the distribution of the residences of the customers that make use of the establishment under analysis or are approximated by using industry-specific guidelines. In many cases, either primary and secondary or primary, secondary and tertiary trade areas are utilized; i.e, by selecting all residents within 1, 3 and 5 miles from the location of an establishment (the exact distance from the site may vary from place to place; polygons may be employed to further refine the definition by excluding parts of the surrounding territory which may be inaccessible due to physical barriers). Although the circular trade areas may be, of necessity, arbitrary, they are based on the client-specific knowledge of his own industry and are therefore believed to be quite reasonable. The best approach would be to use actual customer records or sample studies thereof. While this is much more expensive, many clients have taken this more refined approach. Entry and geo-coding of this information to small areal units such as Census Block Groups or enumeration districts (in rural areas) is then utilized to empirically determine the actual distribution of customers.

The market areas are entered into the AmericanProfile℠ system which utilizes a geometric access method. The system uses a geographic data base containing the latitude and longitude of population centroids. These locations were created by Donnelley for each of 253,000+ block groups and enumeration districts from the most recent 1980 Census maps and location files (DIME files). Table 1 contains a sample session in which several trade areas for multiple establishments are defined. The AmericanProfile℠ system will generate a variety of reports which include data from the 1980 Census (both 100% items as released on STF1 and sample items from STF3), the 1970 Census with information converted to 1980 geography, and Donnelley's proprietary current year estimates of key characteristics. In addition, the system will produce a computer-readable extract file. It is this file which we utilize for analysis since it can be readily interfaced to statistical packages such as SAS.

Although the client has available a wide range of demographic and economic characteristics (a detailed data base dictionary takes over 50 pages and includes hundreds of individual counts, percentages, means, and medians), most analyses rely on a relatively small number of variables. This is necessary in many cases due to the relatively small number of existing sites for the average firm, and because of statistical considerations such as multicollinearity In many cases the variables which enter an analysis include: 1) current total population, 2) average home value, 3) average rent, 4) average family size, 5) percent of family households, 6) percent households with children 18 or younger, 7) percent black, 8) percent Hispanic, 9) several age group percent categories (1-5 years, 6-17, 18-34, etc.), and 10) various income categories (percent less than $10,000, percent $10,000 to $24,999, percent $25,000 to $74,999, percent greater than $75,000, etc.).

In an application of these techniques, the AmericanProfile extract is processed to make use of those data items relevant to the client-specific problem. In Table 2 we provide an example of a SAS program which reads and processes a computer readable extract for a typical client application. This allows the user to create additional statistics from the data base prepared by American-Profile℠. In Table 3 we provide a partial listing of a sample SAS run stream (if more than one) for processing store characteristic data and demographic information to produce regression models of store performance.

After performing and analyzing various alternative regression equations, the client can then utilize the results for several phases of the decision process. Based upon analysis of the residuals, stores that perform below what would be expected are identified. Such establishments can then be targeted for more extensive management review. Those which perform above the average are also looked at with the goal of identifying possible causal relationships. This may involve transfer of either ideas or personnel. The range of areas for which one has reasonable expectation of proper application of models is ascertained and then used in conjunction with the characteristics of potential new site locations to provide evaluation when appropriate.

RESULTS

The approach described in this paper represents a state of the art application of information systems with the most timely and accurate small area statistical data available. It has resulted in an analytic service with significant successes to its credit and a vast range of new and promising future opportunities. It has been applied to a wide spectrum of retail establishments and is utilized throughout the entire United States without any geographic restrictions (i.e., unlike other approaches and systems it is not limited to certain highly urbanized regions or constrained within SMSA's).

The specific application of retail site analysis developed in this paper is but a small part of the other successful uses which have been employed in business decision making. Other applications include, but by no means are limited to: 1) media strategy and advertising, 2) product positioning, 3) sales territory alignment, 4) merchandising, and 5) direct marketing targeting. These and other areas of business decision making are dealt with by an arsenal of tools and techniques which include: 1) retrieval of timely demographic and economic information of user defined marketing areas whereby both small and large areas of interest are not constrained by fixed geopolitical boundaries (i.e., the areas of interest are tailored to specific areas with a great deal of flexibility), 2) market potential estimates which gauge the expected purchases which consumer units are likely to incur (these are available for a wide range of store types, and within each type breakdowns are displayed for fairly detailed major product offerings), 3) life style clusters for very small geographic units (block groups and enumeration districts), 4) cluster information combined with detailed product and services usage information keyed to massive direct mail data bases to provide much more refined and accurate target marketing (Simmons Market Research Bureau and Donnelley Marketing Information Services provide this unique ClusterPlus service), 5) thematic mapping of key demographics and other uses of computer graphics which plays an important role in the overall process of decision making, 6) data base retrieval to permit various levels of geography to be scanned by region and by multiple criteria in order to find areas that fit a particular client profile, and 7) geocoding of customer and competitive locations to provide knowledge of their geographic distribution.

In this paper we have had the opportunity to touch upon only a thin surface of the vast expanse of tools and techniques being used in business decision making. The successful application that we have described has been and will continue to be replicated; these and additional information services will be a necessary and standard fixture for the winning business executive.

[1]I would like to express my appreciation for the work of Dr. S. Haynie, Prof. R. W. Hodge, and Mr. M. Boss in developing and implementing the specific application discussed herein. Mr. G. Hill, Mr. P. Hirtes, Ms. D. Marcus and Mr. S. Speier provided assistance in the preparation of this paper. The entire staff of Advanced Demographic Systems, Donnelley Marketing Information Services is responsible for the development of many of the products and services described, specifically, AmericanProfile, MarketPotential, GraphicProfile, and CensusPlus.

2 For further information see "Mapping Demographics in Market Research Applications" by Carry S. Meyer and "Proper Use of Graphic Information Systems Use" by Daniel S. Raker; both papers presented at the Annual meeting of the National Computer Graphics Association, San Diego, CA., 1982.

```
 1:  COMPANY NAME  (FIRM XYZ)
 2:  USER NAME (GARRY MEYER)
 3:  GEOG(1),CIRC(2),POLYG(3),RTE(4),CORR(5),CONT(6),NONE(7)?2
 4:  TITLE (ENCLOSE WITHIN PARENTHESIS)?(FIRM XYZ 1)
 5:  CENTER OF CIRCLE?22,6645,228,9451
 6:  RADIUS?0/1,1/3,3/5
 7:  SECTORS-CHORDS(1),DIVISION(2),NEITHER(3)?
 8:  GEOG(1),CIRC(2),POLYG(3),RTE(4),CORR(5),CONT(6),NONE(7)?2
 9:  TITLE (ENCLOSE WITHIN PARENTHESIS)?(FIRM XYZ 2)
10:  CENTER OF CIRCLE?22,4866,228,9050
11:  RADIUS?0/1,1/3,3/5
12:  SECTORS-CHORDS(1),DIVISION(2),NEITHER(3)?
13:  GEOG(1),CIRC(2),POLYG(3),RTE(4),CORR(5),CONT(6),NONE(7)?2
14:  TITLE (ENCLOSE WITHIN PARENTHESIS)?(FIRM XYZ 3)
15:  CENTER OF CIRCLE?22,7072,229,3614
16:  RADIUS?0/1,1/3,3/5
17:  SECTORS-CHORDS(1),DIVISION(2),NEITHER(3)?
18:  GEOG(1),CIRC(2),POLYG(3),RTE(4),CORR(5),CONT(6),NONE(7)?2
```

.
.
.

Table 1 - Sample of AmericanProfile Definition

```
//EXEC SAS
//EXT DD DSN=XYZ1234,FIRM.EXT,UNIT=ONLINE,DISP=OLD
//OUT DD DSN=XYZ1234,EXTSAS.FIRM,DISP=(NEW,CATLG,DELETE),
//    UNIT=ONLINE,SPACE=(TRK,(15,2),RLSE)
DATA OUT.EXT;INFILE EXT;
   IF _N_ = 1 THEN DO; INPUT; END;
INPUT
 @18 V1 $15. //
 @1  V2 2.
 @18 (V3-V9)(9.)
 @81 V10 9.2
 @90 (V11-V18)(9.) //
 @117 (V19-V22)(9.) /
 @27 (V23-V42)(9.) /
 @117 V43 9.
 @180 V44 9. /
 @36 (V45-V52)(9.)
 @117 (V53-V59)(9.) ///
 @18 VV1 $15. //
 @1  VV2 2.
 @18 (VV3-VV9)(9.)
 @81 VV10 9.2
 @90 (VV11-VV18)(9.) //
 @117 (VV19-VV22)(9.) /
 @27 (VV23-VV42)(9.) /
 @117 VV43 9.
 @180 VV44 9. /
 @36 (VV45-VV52)(9.)
 @117 (VV53-VV59)(9.) ///
 @18 VVV1 $15. //
 @1  VVV2 2.
 @18 (VVV3-VVV9)(9.)
 @81 VVV10 9.2
 @90 (VVV11-VVV18)(9.) //
 @117 (VVV19-VVV22)(9.) /
 @27 (VVV23-VVV42)(9.) /
 @117 VVV43 9.
 @180 VVV44 9. /
 @36 (VVV45-VVV52)(9.)
 @117 (VVV53-VVV59)(9.) //
;
PBLK1 = V5/V3;
PBLK2 = VV5/VV3;
PBLK3 = VVV5/VVV3;
PSPN1 = V7/V3;
PSPN2 = VV7/VV3;
PSPN3 = VVV7/VVV3;
PNFAM1 = 1.0 - (V43/V8);
PNFAM2 = 1.0 - (VV43/VV8);
PNFAM3 = 1.0 - (VVV43/VVV8);
INC251 = V16 + V17;
INC252 = VV16 + VV17;
INC253 = VVV16 + VVV17;
CWR1 = (V35+V36+V37) / (V23+V32);
CWR2 = (VV35+VV36+VV37) / (VV23+VV32);
CWR3 = (VVV35+VVV36+VVV37) / (VVV23+VVV32);
PROC PRINT DATA=OUT.EXT;
```

Table 2 - Sample SAS Program to "Process" Computer-Readable Demographic Extract

```
DATA AGES;
 INPUT @1 STORNUM 2. (A1-A3 AA1-AA3 AAA1-AAA3) (4.1);
CARDS;
01 176 254 232 196 318 265 245 332 217
02 285 244 209 149 269 223 194 267 220

                    .
                    .
                    .

;
PROC PRINT;
DATA MONTHS;
INPUT @1 STORNUM 2.
 @4 M780 4.1 @9 M781 4.1 @14 M880 4.1 @19 M881 4.1 @24 M980 4.1
 @29 M981 4.1 @34 M1080 4.1 @39 M1081 4.1 @44 M1180 4.1
 @49 M1181 4.1 @54 M1280 4.1 @59 M1281 4.1;
CARDS;
01 3928 2759 4330 3668 4302 3145 2395 1633 2078 2187 4016 3665
02 1684 1445 2358 2907 2996 2538 2764  956 1654 3165 2209 2738

                              .
                              .
                              .

;
DATA STORES;
INPUT @1 STORNUM 2. @3 STORCODE $1. @5 SALES 3.2
 @8 SOFT 3.2 @11 MONTHOP 2. @13 YEAROP 2.;
CARDS;
 1A 2837956404
 2B 1785392634

       .
       .
       .

;
DATA TOT; MERGE MONTHS STORES AGES; BY STORNUM;
DATA FIRM.REGR (KEEP=V3 YY3 YYY3 V52 YY52 YYY52 Y22 YY22 YYY22
 PNFAM1-PNFAM3 PINC1-PINC3 SUM804 STORCODE
   Y44 YY44 YYY44 XY44 XYY44 XYYY44
  PINC11 PINC22 PINC33 PINC111 PINC222 PINC333 PBLK1-PBLK3
   PSPN1-PSPN3 V59 YY59 YYY59 A1-A3 AA1 AA2 AA3 AAA1 AAA2 AAA3
  VV3 VVV3 VV52 VVV52
  INVPOP3 POP13 POP35
 SUM816 SALESCH SALEPOP1 SALEPOP2 SALEPOP3 TOTMON SQFT
  STORNUM);  MERGE FIRM.EXT TOT;
* TOTAL POP;
 YY3=V3+VV3;
 YYY3=YY3+VVV3;
*AVG HOME VALUES;
 C1=(V45+V46+V47+V48+V49+V50+V51);
 C2=(VV45+VV46+VV47+VV48+VV49+VV50+VV51);
 C3=(VVV45+VVV46+VVV47+VVV48+VVV49+VVV50+VVV51);
 YY52=C1/(C1+C2) * V52 + C2/(C1+C2)*VV52;
 YYY52=C1/(C1+C2+C3)*V52 + C2/(C1+C2+C3)*VV52
        + C3/(C1+C2+C3)*VVV52;
* AVERAGE RENT;
 CR1=SUM(OF V53-V58);
 CR2=SUM(OF VV53-VV58);
 CR3=SUM(OF VVV53-VVV58);
 YY59=(CR1*V59 + CR2*VV59)/(CR1+CR2);
 YYY59=(CR1*V59+CR2*VV59+CR3*VVV59)/(CR1+CR2+CR3);
* AVG FAM SIZE;
 Y22=V19/V43;
 YY22=(V19+VV19)/(V43+VV43);
 YYY22=(V19+VV19+VVV19)/(V43+VV43+VVV43);
* PERCENT NONFAMILY HOUSEHOLDS;
```

```
 PNFAM1=(V43/V8)*100;
 PNFAM2=(V43+VV43)/(V8+VV8)*100;
 PNFAM3=(V43+VV43+VVV43)/(V8+VV8+VVV*)*100;
* PERCENT BLACK AND SPANISH;
 PBLK1=V5/V3*100;
 PBLK2=(V5+VV5)YY3 * 100;
 PBLK3=(V5+VV5+VVV5)/YYY3 * 100;
 PSPN1=V7/V3*100;
 PSPN2=(V7+VV7)/YY3 * 100;
 PSPN3=(V7+VV7+VVV7)/YYY3 * 100;
*PERCENT INCOME LESS THAN 10000;
 C5=V12+V13+V14+V15+V16+V17;
 CC5=VV12+VV13+VV14+VV15+VV16+VV17;
 CCC5=VVV12+VVV13+VVV14+VVV15+VVV16+VVV17;
 PINC1=(V12+V13)/C5*100;
 PINC11=(V12+V13+VV12+VV13)/(C5+CC5)*100;
 PINC111=(V12+V13+VV12+VV13+VVV12+VVV13)/(C5+CC5+CCC5)*100;
 PINC2=(V14+V15)/C5 * 100;
 PINC22=(V14+15+VV16+VV17)/(C5+CC5)*100;
 PINC222=(V14+V15+VV14+VV15+VVV14+VVV15)/(C5+CC5+CCC5)*100;
 PINC2=(V14+V15)/C5 * 100;
 PINC22=(V14+V15+VV16+VV17)/(C5+CC5)*100;
 PINC222=(V14+V15+VV14+VV15+VVV14+VVV15)/(C5+cc5+CCC5)*100;
 PINC3=(V16+V17)/C5*100;
 PINC33=(V16+V17+VV16+VV17)/(C5+CC5) * 100;
 PINC333=(V16+V17+VV16+VV17+VVV16+VVV17)/(C5+CC5+CCC5) * 100;
* HHS WITH CHILDREN;
  Y44 = V44/V8 * 100;
  YY44=(V44+VV44)/(V8+VV8)*100;
 YYY44=(V44+VV44+VVV44)/(V8+VV8+VVV8)*100;
 XY44=V44/V8*100;
 XY44=VV44/VV8*100;
 XYYY=VVV44/VVV8*100;
* SALES ;
 SUM814=M981+M1081+M1181+M1281;
 SUM816=SUM814+M781+M811;
SUM804=M980+M1080+M1180+M1280;
 SALESCH=SUM814/SUM814/SUM804*100;
SALEPOP1=SUM816/V3 * 1000; SALEPOP2=SUM816/YY3 * 1000;
  SALEPOP3=SUM816/YYY3 * 100;
 TOTMON=(81-YEAROP)*12 + (13-MONTHOP);
POP13= V3/YYY3 * 100;
POP35= YY3/YYY3*100;
INVPOP3=1/YYY3;
PROC SYSREG DATA=FIRM.REGR S ;
  MODEL SALEPOP3=XY44 XYY44 XYYY44 INVPOP3 POP13 POP35;
PROC PRINT DATA=FIRM.REGR;
```

Table 3 - SAS Sample Program for Merging Client Data, Processed AmericanProfile℠ Demographic and Economic Data, and Generation of Regression Models