

KALEIDOS: a splash of colour for the “silvery haired”

Authors

Linda Pimpinelli, Ignazia Vitali - Giallo Dat@
Guido Cuzzocrea, Salvina Piccarreta - NUNATAC

Abstract

Where did those people heading “for the top” actually get to? Have the “career singles” decided to settle down? Finally, are the “movers, shakers and VIPs” still in their respective positions? Nearly 10 years after the ISTAT census and nearly five after the SEAT Geo-demographic Cluster Segmentation, based on these data, were put on the market, Giallo Dat@ has carried out a project to refresh the colouring of the 320,000 Italian census micro zones.

The available data sources used to reach this ambitious objective are heterogeneous in terms of detail. They also cover diverse information categories, and for comparison purposes, a number of extremely consistent samples have been utilised, with, however, no overlapping.

This paper presents the results of the project, it describes the methodology used which enabled us to successfully use a wide range of sources and to construct a classification process which was at the same time sound and rich with information: SAS users can follow a procedure which leads to an optimal utilisation of the integration potential and the completeness of the available analysis techniques in the system. For marketing users, surprises abound, a splash of colour for the “silvery haired”.

1. Introduction

In 1999 Giallo Dat@ embarked on a project of information gathering called **Ricerca Europea** on Consumption Habits using approximately 400,000 questionnaires. The information gathered is based on categories: *personal data* (age, profession, education...), *behaviour* (sport, reading habits, payment modes...), and *consumption* (products purchased, assets, favourite brands...). The main distribution method was via a questionnaire inserted in the Yellow Pages. Prizes were used as an incentive for people to complete the questionnaire.

In addition to the Ricerca Europea, Giallo Dat@ has also further updated information sources, which are classified as internal and external sources respectively. Namely,

Internal sources

- SEAT private telephone users files, for the year 2000;
- CLUSTER91 socio-demographic classification of the micro-zones;
- SEAT business files, for the year 2000.

External sources

- Automobile owners and related characteristics files (source Targa Services);
- Random telephone survey to validate the representativeness, Italy-wide, of Ricerca Europea.

Nine years after the gathering of the census data and less than five years after the CLUSTER91 Geodemographic Segmentation, based on this data, was put on the market, it was deemed necessary to update this classification using the wealth of information gathered. The objective of the project is to produce a new segmentation at the micro-zone level based both on the socio-demographic variables, thereby guaranteeing continuity of the ‘91 segmentation, and on the consumption habits gathered from Ricerca Europea. By doing this, we are able to colour each of the micro-zones according to the 19 obtained Segments for which we can identify the consumption habits and demographic aspects of the families who live in them. This new product is a useful instrument for the purposes of both Geo-marketing and Target Analysis, namely:

- > to describe territorial shares
- > to locate potential clients
- > to construct potential commercial areas
- > to optimise the composition of selling outlet products
- > to segment prospect files, like telephone files
- > to segment and extend client data bases

2. Presentation of Nunatac and Giallo Dat@

2.1 Giallo Dat@ is the new company of Seat Pagine Gialle (Yellow Pages) Spa, whose activities include business information and direct marketing, which were previously the activities of the Seat Direct division. Due to its commercial network, which is all over Italy, it can provide a range of complete and innovative services in the fields of direct marketing and business information to its clients, of all sizes and sectors, from High Consumption to Financial, from Public Administration to Non-Profit.

Since february 2001, after the acquisition of Consodata S.A. by Seat, Giallo Dat@ is part of Consodata Group.

Its main services are:

- Commercialisation of data banks and business-to-business and business to consumer information
- Management of its clients' data banks and direct marketing campaigns
- Consultation and decision making support for the processing of marketing strategies, using application tools for target analysis, geo-marketing, statistical-territorial and cartographic analysis.

2.2 Nunatac

NUNATAC is a Quality Partner company at the European level, of the SAS Institute, composed of a group of consultants with the following skills:

- applied statistical methods (datamining),
- information development to support decision making systems,
- target marketing.

The defining characteristic of NUNATAC is the combination of specific abilities and a well developed know-how in the field of Database Marketing. Our approach to project responsibility is centred on proposing solutions to needs identified by consultation with management, systems development and training of our clients in the methods we have utilised. In the last 7 years, Nunatac has developed comprehensive solutions for Database Marketing in the following sectors: Banks, Insurance, Telecommunication, Mail Order and the Automobile Industry.

3. Reclassification of the micro-zones

3.1 Preliminary considerations

The data used for the construction of the behavioural clusters was gathered from

GIALLO DAT@ and external sources, with the aim of the optimal utilisation of the wealth of information on the Consumption Habits derived from Ricerca Europea.

The following conclusions were arrived at from ideas validated during the planning stage of the project:

> carry out the Segmentation at a micro-zone territorial level.

> carry out a detailed analysis of the sampling area, so as to identify an analysis process with the aim of the classification of the micro-zones, ensuring robustness and statistical reliability.

> with regards to time-to-market, the availability of prompt data and the existence of heterogeneous sub-markets in the Italian market, to subdivide the project in two successive phases: the first involves the classification of the micro-zones for all of the 15 Regional capitals, which are considered the most important areas. The second phase involves the classification of the micro-zones in the rest of Italy.

> It is important to note that the 15 cities, the other Provincial capitals and the rest of Italy, represent a stratification of the universe, with regard to the socio-demographic and living characteristics described via the CLUSTER91 segmentation. The separate treatment of the sub-universes is therefore desired from the methodological and commercial perspectives.

> the issue of sample covering at the micro-zone level, along with the separate treatment of some sub-universes, requires the identification of a stable and repeatable classification process: it is not a matter of simply dynamically clustering a body of statistical units, but rather the identification of a general classification rule, which can be applied to the different levels of the universe, and always using a common denominator in the interpretation of the segments.

3.2 Information completeness and robustness of the analysis results

Considering what has been said above, some conclusions can be reached on the most appropriate methodological choices to obtain a sound solution to the issue of the classification of the project of the reclassification of the micro-zones.

The wealth of information on life-style habits and consumption behaviour of Italians, obtained via Ricerca Europea, constitutes the main innovative component of the segmentation material: while the CLUSTER91

segmentation was based solely of socio-demographic and living characteristics of families, this time we have at our disposal a wide range of indicators with regard to the following areas:

- cultural habits,
- use of financial products and services,
- level of computerisation,
- willingness to donate to non-profit organisations,
- preferred distributive channels used for the purchase of groceries and other items,
- main free-time activities and how vacations are spent.

The first important methodological choice is the decision to use purchases and life-style as analysis variables, while the socio-demographic characteristics as supplementary variables, so as to better describe to behaviour profiles and to enable them to be individualised.

In terms of statistical reliability, the sample cover considered sufficient for the share of micro-zones for the universe of the 15 cities, is about 20%. In order to strengthen the analysis, to better exploit nearly all of the questionnaires obtained via Ricerca Europea, and finally, to determine a behavioural segmentation rule which can be extended to the whole Italian universe, it was decided to proceed as follows:

- the concept of Macro-zones was defined, which is instrumental to the analysis, in which micro-zones with the same CAP (post/zip codes) and the same CLUSTER91 Segment were aggregated. The resulting Macro-zones are homogeneous with respect to geographical proximity criteria, and to socio-demographic similarity of the families.
- Within the resulting Macro-zones it is possible to identify the relevant features to characterise consumption behaviour and life-style. Having determined these features, the *factor score coefficients* relating to them can be applied to the specific coordinates of the original variables of the micro-zones: what this means, in fact, is percentage presences for single characteristics (e.g. % of Il Sole 24 Ore readers), normalised compared to the number of answers in the geographical unit considered.
- The same concept applies to the cluster rule, constructed at the Macro-zone level and applied to the micro-zones, on the basis of specific coordinates.
- The reasoning is that a more robust Macro-zone rule for the universe is estimated and it is applied to the micro-zones, while complying to the typicalness of each of them.

In order to ensure the continuation of the statistical reliability of the sample cover, the classification of the micro-zones proceeded as follows:

- on the basis of the general classification rule, estimated at a Macro-zone level, the micro-zones with a sufficient level of cover (about 20% of the total micro-zones) were classified considering their individual behaviour coordinates;
- the micro-zones with an insufficient sample cover and which are part of covered Macro-zones (a further 60% of the total micro-zones), were classified, inheriting the cluster of the Macro-zone to which they belong.

3.3 Use of exogenous data banks for the classification of uncovered micro-zones.

For the territorial units whose level of sample covering was considered insufficient in terms of statistical reliability, at both a micro-zone and a macro-zone level, the clusterisation method was different. To this end, it was decided to construct a discriminate model based on:

- Data banks of automobile owners
- Files on SEAT economic operators
- Socio-demographic analysis carried out in '91.

For an even better analytical robustness, it was decided to construct the model from 20% of the sufficiently covered micro-zones.

Through the discriminate model based on external sources, we were able to classify nearly all of the micro-zones, which was previously impossible to colour with only the data from Ricerca Europea.

3.4 The clusterisation in two stages

The wealth of information on consumption and life-styles gathered through Ricerca Europea was studied using factorial analysis techniques in order to identify the relevant themes inherent to this study.

Factorial analyses were carried out on single information categories and on the union of several categories (see paragraph 3.2). The factors obtained were used as input for the successive cluster analysis. In particular, the factorial of the union of several categories was used to:

- reduce the number of input variables to the clusterisation algorithm,
- to assign a larger relative weight to the Culture, Technology and Financial Services categories, which would otherwise have been covered up by an excessive number of “super-structural” factors.

The more analytical factors, constructed on single categories, are, however, used to trace the subsequent profiles of the obtained segments.

On the basis of 27 synthesis factors on consumption and life-style, a classification rule in 15 clusters was attained. These clusters have a very clearly defined socio-demographic descriptive connotation, with the exception of the 4 larger clusters: within each of these, a further split was carried out on the basis of the number of family members and the different age brackets within the family.

The total number of clusters attained is 19.

4. Conclusions and results

To describe the profiles of the 19 clusters, both the variables used in the analyses and the variables from other data banks, were utilised. The availability of such a wide scope of information allowed us to trace a “well rounded” behavioural and socio-demographic profile of the territorial micro units which characterise the Italian market.

What is of significant interest is the possibility of comparing the migratory phenomena of the CLUSTER91 segments to the updated ones: our cities, like the provinces, have changed along with their inhabitants and characteristics.

Within the synthesis results, it is worthwhile analysing the percentage distribution of

Segmenti 2000, ordered according to economic status:

Segment	% Micro-zone	% Families	% Tel. Users
RC 01	3,93%	2,36%	2,63%
RC 02	2,41%	2,82%	3,12%
RC 03	3,14%	3,35%	3,71%
RC 04	1,80%	2,13%	2,34%
RC 05	2,05%	2,78%	3,12%
RC 06	2,90%	3,57%	3,95%
RC 07	2,51%	2,29%	2,56%
RC 08	6,79%	5,01%	5,49%
RC 09	2,75%	2,82%	3,07%
RC 10	5,51%	6,67%	7,29%
RC 11	5,88%	7,22%	7,53%
RC 12	7,77%	9,92%	10,83%
RC 13	3,12%	3,38%	3,47%
RC 14	3,41%	3,76%	3,47%
RC 15	6,37%	8,07%	7,88%
RC 16	4,50%	5,73%	5,34%
RC 17	5,90%	5,75%	5,88%
RC 18	7,37%	7,25%	6,71%
RC 19	9,65%	11,52%	9,89%
RC 20	12,24%	3,60%	1,75%

The graphic below illustrates the relative position of the 19 clusters, using 4 variables and constructed from the linear combination and the standardisation of the used variables. On the horizontal axis there is a variable linked to the richness of the cluster. On the vertical axis there is a variable linked to the tendency of the cluster to adhere to initiatives of a cultural nature, while the size of the symbol represents the size of the families, while the colour of the symbol becomes lighter as there is an increase in the technological knowledge displayed through the questions of the questionnaire.

