Habits® for ESRI

Methodology

Version control

Version	Date	Revision	Comments
001.000	15/01/2015	Germán Sánchez, Miquel Torrens, Ferran Carrascosa	Initial version

Overview

1 Introduction

Habits® is a database that contains a full description of the Spanish population and its precise geographical location. This information provides a wide range of possibilities in developing marketing and sales strategies, vastly improving decision-making in many areas.

Habits® segments the Spanish families typologies based on their demographic and economic characteristics, offering a large amount of economic indicators of Spanish households as income, expenses and property value.

Habits® can know the presence and expenses profile of each type of household in each geographic micro area of Spain, facilitating precise calculation of the business opportunities in the territory. Habits® also enables information to be assigned to your own databases for enrichment and segmentation.

Habits® variables are built on public databases, mostly provided by the INE[1]. All data used are suitably anonymised and therefore comply scrupulously the LOPD. The use of information from INE guarantees the reliability and consistency of the data, in addition to its completeness and regular updating.

This document prepared by AIS aims to facilitate the understanding of the methodology used by AIS to obtain the variables delivered in this project. The document is estructured as follows: Section 2 enumerates the data sources used to generate the considered variables, while Section 3 details the methodology employed to build each set of variables provided by AIS.

---

[1] INE: National Institute of Statistics (from Spanish, *Instituto Nacional de Estadística,* www.ine.es*)*

2 Sources

Habits® variables are related to one of the administrative entities that divide the territory of Spain: census area, census district, municipality, province and autonomous community. The INE annually publishes a set of digital maps available in all these administrative levels. The main sets of information published by the INE are the Population and Housing Census[2], the Household Budget Survey[3] (HBS), the Municipal Register[4] (*Padrón*), and the Economically Active Population Survey[5] (EAPS), existing other sources of information as the Consumer Price Index[6] (CPI).

Other external sources of information are the Public Service of Employment[7] (PSE), the General Directorate for Cadaster[8] and a set of real estate portals.

Below is detailed the approximate frequency of updating of each set of information that have been used for developing data provided by AIS.

- Digital maps: annual
- Census: 10 years
- HBS: annual
- Municipal Register: annual
- EAPS: annual
- CPI: annual
- PSE: quarterly
- Cadaster: quarterly
- Real estate portals: monthly

---

[2] Population and Housing Census (from Spanish, Censo de Población y Viviendas):
http://www.ine.es/en/censos2011 (INE)
[3] HBS, Household Budget Survey (in Spanish, *Encuesta de Presupuestos Familiares, EPF*):
http://www.ine.es/en/prensa/epf_prensa_en.htm (INE)
[4] Municipal Register (in Spanish, *Padrón*):
http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t20/e245/&file=inebase (INE)
[5] EAPS, Economically Active Population Survey (in Spanish, *Encuesta de Población Activa, EPA*):
http://www.ine.es/inebaseDYN/epa30308/epa_inicio.htm
[6] CPI, Consumer Price Index *(*in spanish, *Indice de Precios de Consumo*, IPC)
http://www.ine.es/ipc
[7] PSE, Public Service of Employment (in Spanish, *Servicio Público de Empleo Estatal, SEPE*):
http://www.sepe.es
[8] General Directorate for Cadaster (in Spanish, *Dirección General del Catastro*):
http://www.catastro.meh.es

3 Methodology

Habits® variables are computed through techniques of missing data, statistical matching, calibration and segmentation. This section introduces the methodology used to build each provided set of information.

3.1 Identifiers

This set contains the different identifier codes and names of the considered administrative entities. The census area ID implicitly contains the IDs of the superordinate geographical authorities (district, municipality and province), except the Autonomous Community. All these IDs make reference to the year 2013.

3.2 Typologies

Habits typologies are computed from the imputation of households described in the Census published by the INE, in each of the census areas that divide the Spanish territory. The data associated with each household are updated annually through the HBS, using statistical matching techniques to link each household of the Census to a household in the HBS. At the end of this process, data of each census area are computed by aggregating the data related to the households assigned to it. This methodology enables the exploitation of any variable published in the HBS, including incomes, spending and some geodemographics variables.

The specific Habits® typologies A to R have been obtained by applying a clustering technique (K-Means) to households from Census and HBS. Microtypologies have been obtained by analysing households' income within each typology.

The data used for the preparation of this version were the most updated at the time of publication: Maps 2013, Census 2001, HBS 2012 and Municipal Register 2013.

3.3 Inhabitants

All variables in this set have been directly obtained by downloading and processing annual information from the Municipal Register. These variables include the population by sex, age, age by sex, nationality, nationality by sex, birthplace and birthplace by sex and refer to 2013.

3.4 Incomes

Income variables are obtained by adding the incomes of the households imputed in each census area with a statistical weighting procedure. The remaining variables of this set of information are obtained by analysing the distribution of household incomes in each section area.

It's important to note that, due to the fact that the specific monetary values from HBS had date of 2012, these values have been updated to the reality of 2013 by applying the CPI 2013 values.

3.5 Spending

Spending variables are obtained in an analogous way to the previous set of Incomes, i.e., by using the spending data reported by each household imputed in each census area. In the same way as before, monetary values have also been updated to 2013 using the CPI.

3.6 Geodemographics

Variables included in this set are computed by employing different processes:

- `OTPRCAP` describes if the census area belongs to a municipality that is capital of the province.
- `OTMUPOP` is related to the size of the municipality of the census area.
- `OTSUPSC` displays the area surface of each census area.
- `OTACTME`, `OTOCCME` and `OTSTUDE` are computed by characterising each household and estimating their number of active, employed and student members.
- `OCTST???` are obtained by taking into account the level of education of the main representative of each household.
- `OTPP???` and `OTFOR??` are a summary of the data from Municipality Register.
- `OTUN???` variables are estimated by crossing the aggregated results at the provincial level of the EAPS, the updated Municipality Register, monthly publications disaggregated at the municipal level of the PSE and Habits® itself.
- `OTDENSIT = PEOPLE / OTSUPSC`

3.7 PropValue

All variables related to property value are obtained by analysing, for each census area, a sufficient number of properties for sale and rent published in several property websites. Data is downloaded and processed monthly and the processes ensure that the characteristics of the properties taken to inform each census area are consistent with the environment in which they are located.

3.8 Cadaster

Variables included in this block are fully obtained by quaterly downloading and processing properties registered in the from Cadaster. In general, all variables refer to count properties with certain features, except the following:

- `CATYPEA` indicates if most of properties of the census area are considered to be urban or rural.
- `CAMESUP`, `CACOEFF` and `CAANTIQ` are averages of surface, sharing ratio and age of the properties, respectively.

3.9 Census

The variables included in this block are directly provided by the Census, after performing some processes to complete missing data.