Habits®
METHODOLOGY AND INFORMATION

13-09-2021

Version 002.000

Version control

Version	Date	Revision	Comments
001.000	15-01-2015	Germán Sánchez, Miquel Torrens, Ferran Carrascosa	Initial version
002.000	13-09-2021	Germán Sánchez	Updated to new methodology and version 2021-1

Overview

1. Introducción

The objective of this document, prepared by AIS, is to introduce the methodology used to generate Habits®, as well as to detail the information modules contained in the latest version of Habits®.

2. Habits®

Habits® is a set of economic and sociodemographic indicators that provide an accurate portrait of Spanish society and its way of life. It's a key tool when defining marketing and geomarketing strategies.

The database is updated every six months, following the updates of the different sources of information from which it is fed. More specifically, two main updates are made annually, at the time when the maps and the Municipal Register (*Padrón*) are updated, on the one hand, and the Family Budget Survey (EPF, from Spanish *Encuesta de Presupuestos Familiares*), and the Living Conditions Survey (ECV, from Spanish *Encuesta de Condiciones de Vida*) by the other, main sources of information for the Habits® generation.

2.1. Objectives

- **Habits®** segments Spanish households into typologies based on their sociodemographic and economic characteristics, offering a wide range of family economic indicators such as income, expenses, savings, and assets.
- **Habits®** makes it possible to know the presence and expense profile of each type of household in each geographical microzone throughout Spain, allowing the precise calculation of business opportunities in the territory.
- **Habits®** also makes it possible to assign its information to your own databases for enrichment and segmentation.

2.2. Sources of information

All the information used in the construction of Habits® has its origin in public sources and Law 15/1999, of December 13, on the Protection of Personal Data is always respected. The standard information sources used by Habits® are the following.

1. **Population and Housing Census:**

   - <u>Objective</u>: population count and knowledge of its structure.

   - <u>Variables studied:</u> regarding the population: age, sex, nationality, residence and marital status, place of birth, migration variables, education, relationship with economic activity, socioeconomic status, marriage, fertility, kinship relationships, area, size of the municipality, structure of the homes and family nuclei.

   Regarding real state: class, area, facilities, useful area in square meters, year of construction, number of rooms, tenure regime and owner class; and by type, number of floors, number of dwellings, class of owner, condition and year of construction of the building.

   <u>Latest available results:</u> 2011.

Source: INE, Statistics National Institute.

Update: decennial.

2. Municipal Register (*Padrón*):

- Objective: to provide the official population figures, approved by Royal Decree, of all Spanish municipalities on January 1 of each year.

- Variables studied: population according to different territorial breakdowns by age, sex, origin, and nationality.

Latest available results: 2020.

Source: INE, Statistics National Institute.

Update: annual.

3. Household Budget Survey (EPF):

- Objective: to provide annual information on the nature and destination of consumer spending and on various characteristics related to the living conditions of households

- Variables studied: total household expenditure, average expenditure per household and average expenditure per person on goods and services in monetary terms according to 12 expenditure groups, characteristics of the households and the main breadwinner. More than 500 expenditure indicators.

Latest available results: 2019.

Source: INE, Statistics National Institute.

Update: annual.

4. Cartography at the census section level:

- Objective: to provide information on the polygon into which the Spanish state is divided, perfectly identifying the different administrative partitions: census section, census district, municipality, province, and Autonomous Community.

- Variables studied: cartography of the census sections, including identifiers, names, perimeters, and surfaces.

Latest available results: 2020.

Source: INE, Statistics National Institute.

Update: annual.

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

2

5. Survey of life conditions (ECV):

- Objective: to provide comparative statistics on income distribution and social exclusion at the European level

- Variables studied: family income, poverty rates, population at risk of social exclusion.

Latest available results: 2019.

Source: INE, Statistics National Institute.

Update: annual.

6. Active Population Survey (EPA), Public Employment Service (SEPE)

- Objective: to provide information on the population related to the labour market: employed, active, unemployed, and inactive.

- Variables studied: number of employed, active, unemployed, and inactive by age and sector.

Latest available results: first quarter of 2021.

Source: INE, Statistics National Institute; SEPE, Public Employment Service.

Update: quarterly / annual (respectively).

7. Electronic Office of the Cadastre:

- Objective: to provide information on all properties in the State (excluding the Basque Country).

- Variables studied: counts by type of property, characteristics of the properties (surface area, age, participation coefficient, geolocation...).

Latest available results: first half of 2021.

Source: Electronic Office of the Cadastre.

Update: biannual.

8. Real estate portals:

- Objective: to obtain information on the portfolio of properties currently for sale and rent.

- Variables studied: characteristics of the properties, both residential and non-residential (flats, houses, shops, offices, garages, storage rooms, industrial ships, land), including the offer price and geolocation; also, for residential properties, counts by number of rooms and bathrooms.

Latest available results: May 2021.

Source: different real estate portals.

Update: monthly.

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

3

9. Real estate price series:

- <u>Objective:</u> to provide information on the real estate market in terms of sale prices, assessment prices, number of transactions, and price indices.

- <u>Variables studied:</u> house price indices, number and price of assessments and transactions for residential and non-residential real estate. Within the residential ones, distinction between free and protected housing, new and not new, etc.

<u>Latest available results:</u> 2021.

<u>Source:</u> Housing Price Index (IPV, INE), Ministry of Development (Fomento), Notaries.

<u>Update:</u> quarterly or monthly.

10. Weather information:

- <u>Objective:</u> to provide historical meteorological information.

- <u>Variables studied:</u> historical data at the municipality level (meteorological station) relative to temperatures, atmospheric pressure, wind, rainfall, sunny days, etc.

<u>Latest available results:</u> first trimester of 2020.

<u>Source:</u> State Meteorological Agency (AEMet).

<u>Update:</u> annual.

11. Business activity:

- <u>Objective:</u> to provide information on business activity.

- <u>Variables studied:</u> natural and legal persons, counts by CNAE, last invoice figure and number of employees, age of the company and debt.

<u>Latest available results:</u> 2021.

<u>Source:</u> Camerdata

<u>Update:</u> biannual.

12. Criminality:

- <u>Objective:</u> to provide information on the volume of crimes broken down by type of crime.

- <u>Variables studied:</u> counts at the municipal level of different types of crimes: homicides, injuries, kidnappings, robberies, drugs...

<u>Latest available results:</u> 2020.

<u>Source:</u> Statistical portal of crime, Minister for home affairs

<u>Update:</u> quarterly.

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

4

Group

2.3. Definition of Habits® Typologies

The basic typologies are made up of 13 segments or groupings of families depending on the composition of the household (number of members, number of children and adults) and the ages of its members.

The resulting typologies are the following:

Mates

Households in which more than two adults without children and all of them under 65 years of age share a home.

Singles

Households made up of a single person. According to the age of this person, a distinction is made between a **young single** (under 35 years old) and an **adult single** (between 35 and 64 years old).

DINKs

Households made up of an adult couple without children and both with earned income. Depending on the age of the main breadwinner, we distinguish between **young DINKs** (under 35 years old) and **adult DINKs** (between 35 and 64 years old).

Full nest

Households formed by an adult couple with children. The age of the children makes it possible to differentiate between the following three types: **Full nest with young children**, where all children are under 16 years old, **Full nest with adolescent children**, where all children are between 16 and 34 years old, and **Full nest with children of different ages**, where children under 16 are mixed with others between 16 and 34 years old.

Single parent

Families made up of a single adult with at least one child, all under the age of 25.

Intergenerational

Families made up of at least three members, one under 25 years old, another between 25 and 64 years old, and another 65 years old or older.
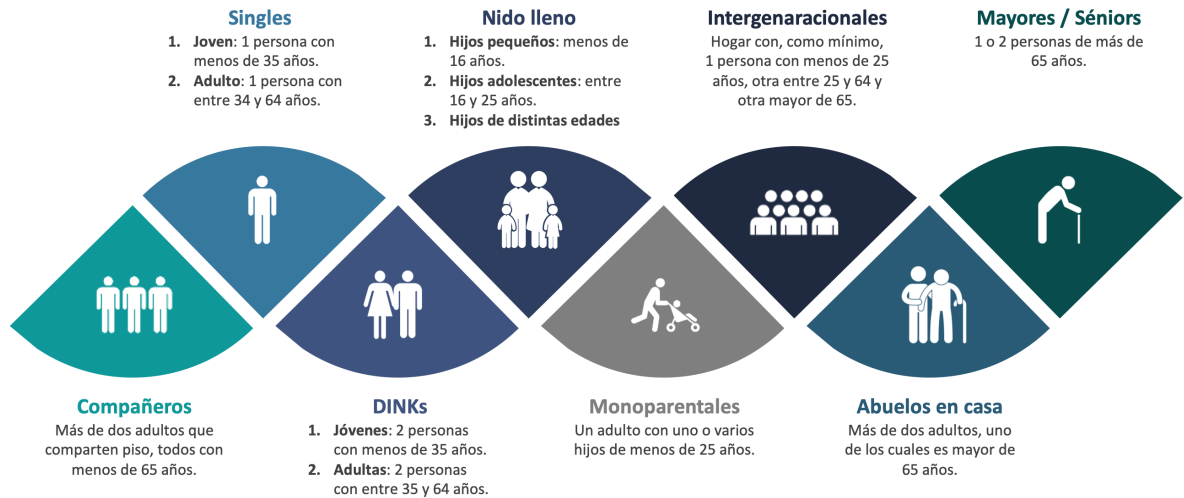
Grandparents at home

Families without children consisting of more than two members, where at least one of them is 65 years or older.

Senior

Households made up of one or two adults aged 65 or over and without children. The number of family members defines the typologies **One senior** or **Two seniors**.

The following figure shows (in Spanish) a graphic representation of the typologies, generally ordered by the age of the main breadwinner or its members.

Singles
1. **Joven**: 1 persona con menos de 35 años.
2. **Adulto**: 1 persona con entre 34 y 64 años.

Nido lleno
1. **Hijos pequeños**: menos de 16 años.
2. **Hijos adolescentes**: entre 16 y 25 años.
3. **Hijos de distintas edades**

Intergenaracionales
Hogar con, como mínimo, 1 persona con menos de 25 años, otra entre 25 y 64 y otra mayor de 65.

Mayores / Séniors
1 o 2 personas de más de 65 años.

Compañeros
Más de dos adultos que comparten piso, todos con menos de 65 años.

DINKs
1. **Jóvenes**: 2 personas con menos de 35 años.
2. **Adultas**: 2 personas con entre 35 y 64 años.

Monoparentales
Un adulto con uno o varios hijos de menos de 25 años.

Abuelos en casa
Más de dos adultos, uno de los cuales es mayor de 65 años.

It is important to note that the typologies are defined sequentially. Thus, if a family is classified in a certain typology, the following typologies will not be considered even if the family also meets its requirements. The order of application of the typologies, which defines a priority among them, is as follows:

1. Single parent families
2. DINKs
3. Two seniors
4. Singles
5. One senior
6. Full nest
7. Intergenerational
8. Mates
9. Grantparents at home

2.4. Data fusion

We have information from the Census and the family budget survey (EPF). However, this information cannot be directly crossed, so it is necessary to carry out a process of fusion of the information.

For this, AIS has its own algorithm based on Statistical Matching or Data Fusion techniques. And that consists of combining information from different sources that do not contain common observation units.

Subsequently, through the public information available by census section and population units, together with the municipal registers, the information at the census section level is completed through ecological inference processes.

The results by census sections and population units provide the marginal distributions of the infra-municipal data (census section) and the municipal registers provide crosses by sex, age, origin, and nationality.

The main steps that are carried out in this phase of data fusion are introduced below.

4.2.1. Initial state and target

Each new update of Habits starts from the same initial state: the families included in the Population and Housing Census are distributed in each of the census sections into which the Spanish territory is divided. The general objective of this phase is to perform a statistical matching between the new source to be incorporated (in particular, the EPF and the ECV) and the Census, so that we indirectly impute each record of the survey to each of the census sections considered.

4.2.2. Definition of family typologies

To carry out this statistical matching between the Census, the EPF and the ECV (or any source of information to be incorporated), the most probable Habits typology is defined for each family in the Census. Section 2.3 details the set of variables considered to make this unambiguous definition. On the other hand, the same process is carried out on the families of the source to be incorporated. This step can be carried out in two ways: in case of having the information (social and demographic variables) that defines the typology in the same way as in the Census (as in the case of the EPF), the same method is applied to identify the most likely typology for each of the families. On the other hand, if the information included in the new source is not the same as that used to define the typology, or it is not possible to adapt it, a Machine Learning process is gathered with the Census data, considering the common variables between the Census and the new source, and this model is applied to the families of the new source, thus obtaining the most probable typology for each family of the new source.

4.2.3. Information fusion

Once the most probable type of family has been defined for all records in Census and in the source of information to be incorporated into Habits, the propensity score matching methodology is applied, by means of which a Mahalanobis distance is defined through the common variables between Census and the new source (including the Habits typology). After obtaining this score, each family of the Census is associated with the family of the most appropriate source to be incorporated, always respecting the geographical restrictions (thus making the most of the geographical information available in both databases) and other types of restrictions linked to the nature of the new data source to be incorporated.

With this methodology we can achieve the initial objective, which was to impute each record of the information source to be incorporated into each of the considered census sections. From here, the aggregations of the social, demographic and / or economic information provided by the new source of information can be made, towards the geographical levels that are considered appropriate (census section, census district, municipality, province, and Autonomous Community).

Finally, it should be noted that this methodology ensures that the weighted average of the different estimates at precise geographical levels coincides with the published data. This is achieved with a calibration phase. In addition, basically three information validation processes are carried out:

1. Automatic validation of information, according to its nature
   For example, percentage variables must add 1 (or 100), certain variables cannot be negative, and so on.

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

7

2. Automated expert validation
A series of expert thresholds are defined for each variable, so that an alert is triggered if there are values outside of that reasonable range.
3. Manual expert validation
Visual validation processes of the geographical distribution of the different estimated variables are carried out (through AIS Data Maps®), in addition to analytical validation processes (through AIS Master®).

2.5. Habits® update

Habits® is updated biannually, although some modules may be updated more frequently.

Both the municipal registers and the EPF and the ECV have annual periodicity. However, the Census has a ten-year periodicity and not taking this aspect into account could distort the information over time.

Reason why, AIS updates the population information contained in its Habits® database through weighting processes based on the information contained in the municipal registers, the EPF and the ECV. Through this update, the information contained in Habits® is adjusted considering possible population migrations, possible changes in the distribution of typologies and its aging / rejuvenation. All this at the census section level.

2.6. Census section update

Since the population evolves over time, the definition of the census tracts does too. Every year census sections are born and die and there are also census sections that see the x | y coordinates that determine their borders modified.

AIS is aware of this census update process and adjusts the sectioning content in Habits® through annual street maps and census polygon maps. Process that is carried out through classification processes generated by AIS.

2.7. Cadastre

The variables included in this block are downloaded every six months and processed from the Cadastre registry (the provincial communities are excluded). In general, all the variables refer to the count of properties with certain characteristics except:

- The type of area where it is reported only if the majority of properties are urban or rural.

- Areas, participation coefficients and antiques that are calculated as an average of the observed properties.

2.8. Habits 2021-1_T1

The latest version of Habits® published is the one corresponding to the first quarter of 2021 (first semester), whose sources are the following:

Source	Date data	Publication date	Update periodicity
Maps INE	2020-01-01	2020-02	Annual
Municipal register	2020-01-01	2020-01	Annual
EPF	2019	2020-07-30	Annual

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

8

ECV	2020	2021-07-21	Annual
EPA	2021T1	2021-04	Quarterly
SEPE	2020-12	2020-12	Monthly
INE - IPV	2020T4	2021-05	Quarterly
Ministry of Development	2020T4	2021-05	Quarterly
Notaries	2021-03	2021-05	Monthly
Census	2011	2014	Unique
Catastre	2021Q1	2021-02	Biannual
Properties	2020-06 a 2021-05	2021-05	Monthly
Camerdata	2021Q1	2021-05	Quarterly
DGT	2021-03	2021-04	Monthly
AEMet	2021-03	2021-04	Monthly

The column "Data date" indicates which period the data obtained from each source refers to, while the column "Publication date" indicates the most recent publication date of the source.

2.9. Information blocks

The different blocks of information contained in Habits® are detailed below.

- Geodemographics: identifiers, names, and density of the different administrative partitions.
- Family typologies: percentage of presence of each of the Habits® typologies.
- Family income.
- 12 Expense groups of the family basket:
  1. Food and non-alcoholic beverages
  2. Alcoholic beverages, tobacco, and narcotics
  3. Clothing and shoes
  4. Housing, water, electricity, natural gas and other fuels
  5. Furniture, housing equipment and running costs of housing maintenance
  6. Healthcare
  7. Transportation
  8. Communications
  9. Leisure, entertainment, and culture
  10. Education
  11. Hotels, cafés, and restaurants
  12. Other goods and services
- *Property Value*: purchase and rental real estate prices, absolute and unitary
- AVM: valuation of specific properties, offering both asking and closing prices
- Housing prices: from INE, Ministry of Development and Notaries
- Employment and educational level
- Municipal register: sexes, ages, nationalities, origin
- Cadastre, with all the properties of the State
- Cadastre valuation: valuation of all properties in the Cadastre with our AVM models
- Census 2011

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

9

- Economic Indicators (IE), with economic capacity indicators (ICE) and poverty indicators
- Business activity
- Weather
- Shops
- Use of ICT

The following figure shows a summary of the modules currently included in Habits®:

Diputació, 246 bajos · 08007 Barcelona · España · Tel: +34 93 414 35 34 · Fax: +34 93 414 10 28 · **w w w . a i s - i n t . c o m**

10