



**Conoce tu ciudad con Habits<sup>®</sup>  
Big Data**

# ¿Qué es Habits<sup>®</sup> Big Data?



- ❑ Es una colección de indicadores (más de 1.800) que nos ayuda a conocer cómo es una zona: sección censal, distrito, código postal, municipio, provincia y comunidad autónoma.
- ❑ Sus usos son amplios:
  - ❑ Diseño de políticas socio-económicas
  - ❑ Poner información a disposición de la ciudadanía
  - ❑ Mercado potencial
  - ❑ Ayuda a la expansión de una empresa o cadena
  - ❑ Tarificación de primas de seguro
  - ❑ Captación de empresas
  - ❑ Enriquecimiento de la información de clientes
  - ❑ ...

# Módulos



Geodemografía



Property Value



Ingresos



Gastos



Tipologías



Estudios y  
empleo



Indicadores  
económicos



Padrón



Censo Viviendas y  
hogares



Catastro



Actividad empresarial



Meteorología



Redes Sociales



Cartografía



# Metodología general

- Censo imputado a sección censal
- Enriquecimiento del Censo con información externa (encuestas): cada familia estará asociada al registro más adecuado
- Calibración de los datos a partir de los agregados oficiales
- Suavizado de casos extremos y grandes saltos geográficos
- Bajada de otros datos de nivel geográfico alto a Sección censal a partir de variables *proxy*



# Fuentes de información

La Base de datos Habits <sup>(R)</sup> está compuesta por **distintos módulos de información**

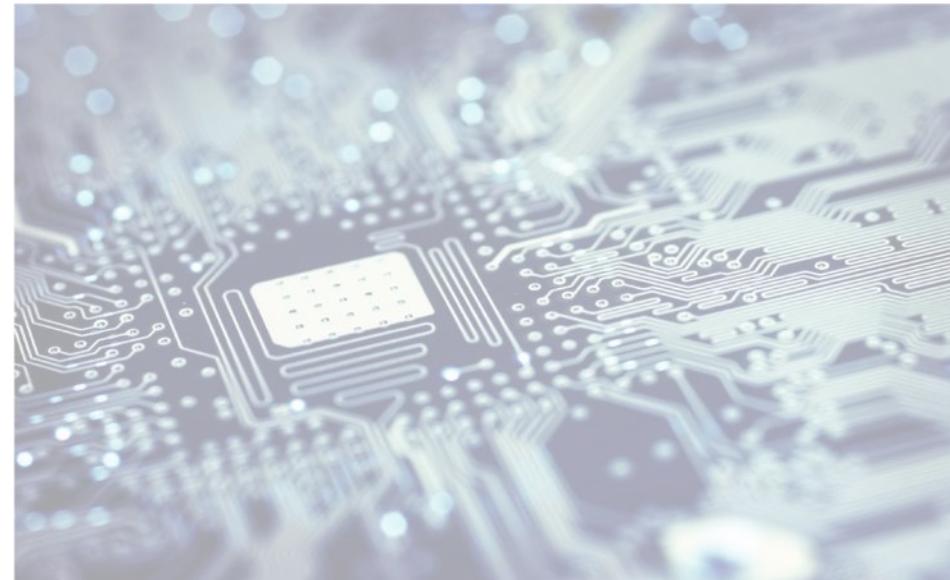
Toda la información utilizada en la construcción de Habits <sup>(R)</sup> tiene su origen en fuentes públicas y se respeta en todo momento la Ley 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal así como la RGPD. Las fuentes de información estándar utilizadas por Habits <sup>(R)</sup> son:

- Censo de Población y Viviendas
- Padrón municipal
- Encuesta de Presupuestos Familiares (EPF)
- Cartografía a nivel de sección censal-
- Encuesta de Condiciones de Vida (ECV)
- Encuesta de población activa (EPA) y Servicio Público de Empleo (SEPE)
- Sede Electrónica del Catastro
- Portales inmobiliarios
- Series de precios inmobiliarios
- Base de Datos de actividad empresarial de Iberinform, Crédito y Caución.
- Redes Sociales a partir de la información suministrada por Brandchats
- Agencia Estatal de Meteorología
- Alojamientos turísticos (extraída de Datahippo)



# Fusión de datos

- Objetivo: fusionar Censo con diferentes encuestas (EPF, ECV) para enriquecer sus datos
- Variables comunes Censo-encuesta
- *Statistical Matching*
  - Métodos paramétricos: presuponen distribución de probabilidad, generamos un modelo con la encuesta y lo aplicamos al Censo para obtener la variable deseada
  - Métodos no paramétricos: cada familia del Censo es asociada con la familia de la encuesta más adecuada, asumiendo sus variables
- Calibración con el dato público



# Validación de la información

- La metodología asegura coincidir con el dato oficial al nivel geográfico en que éste haya sido publicado (usualmente, CCAA): fase de calibración
- Procesos de validación
  - Validación automática de la información, según su naturaleza
  - Validación experta automatizada (valores válidos)
  - Validación experta manual, a nivel visual y analítico (*AIS Data Maps, AIS Modelling Outcome*)



# Bloques de información



## Geodemografía



En el módulo de geodemografía encontramos variables sobre la composición geodemográfica. Es decir, los indicadores sobre cómo se estructura la sociedad española tanto a nivel familiar como a nivel de comunidad.

Es por eso que se incluyen variables relacionadas con el hogar como pueden ser el número de personas por familia y variables relacionadas con las poblaciones donde vivimos tales como el tamaño del municipio, la superficie de la entidad geográfica o la densidad de población

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



## Property Value



Incluimos variables relativas al mercado de la vivienda. Entre ellas podemos hacer una diferenciación entre variables relativas al precio de la vivienda y variables relativas a la superficie. Además, podemos diferenciar entre viviendas de alquiler y viviendas de compra-venta.

Ofrecemos tanto variables históricas de precio y superficie, como variables de estimación de precio tanto de alquiler como de compra-venta, desagregando hasta el precio unitario por metro cuadrado.

Algunos ejemplos de variables podrían ser el precio mensual de alquiler de mercado o el precio de venta estimado del inmueble.

Por lo que se refiere a la granularidad, para las variables históricas tenemos una máxima granularidad a nivel de municipio, **código postal distrito y sección censal** y para las variables de estimación su máxima granularidad llegará al **inmueble y portal**.

# Bloques de información



## Ingresos



Aquí encontramos variables relativas a los ingresos del área administrativa y su distribución.

Por lo que se refiere a la distribución de los ingresos tenemos dos formas de agregación. La primera es a través de un sistema de rangos:

- Ingresos mensuales familiares menores a 1389€
- Ingresos mensuales familiares entre 1389 y 1641€
- Ingresos mensuales familiares entre 1641 y 2122€
- Ingresos mensuales familiares entre 2122 y 2589€
- Ingresos mensuales familiares superiores a 2589€

El segundo sistema de distribución de ingresos son los percentiles de los ingresos mensuales netos.

Por último, el nivel máximo de granularidad será de **municipio, código postal, distrito y sección censal** en todas las variables, excepto en los ingresos netos (mensuales y anuales) que podrá ser de **portal e inmueble**.

# Bloques de información



## Gastos



En este módulo encontramos la distribución de gastos de los residentes del área administrativa agrupada en 12 grandes capítulos.

1. Alimentos y bebidas no alcohólicas
2. Bebidas alcohólicas, tabaco y narcóticos
3. Artículos de vestir y calzado
4. Vivienda, agua, electricidad, gas y otros combustibles
5. Mobiliario, equipamiento del hogar y gastos corrientes de conservación de la vivienda
6. Salud
7. Transportes
8. Comunicaciones
9. Ocio, espectáculos y cultura
10. Enseñanza
11. Hoteles, cafés y restaurantes.

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



## Tipologías



Informamos de la distribución de tipologías y microtipologías Habits que se generan en función de las características socio-económicas de la población española. La segmentación se basa en los siguientes ejes:

- Estructura del hogar
- Zona geográfica
- Tamaño de municipio
- Situación vivienda
- Situación laboral del sustentador
- Ingresos
- Edad
- Educación del sustentador
- Nacionalidad

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



## Estudios y empleo



Conocer la distribución del nivel de estudios, la composición del empleo y a la situación laboral de los habitantes, es lo tenemos en este bloque de información.

Dentro del grupo de variables referentes a la distribución del nivel de estudios podremos por ejemplo ver qué porcentaje de habitantes tienen un nivel de estudios concreto y dónde se localizan.

Entre las variables que nos explican la composición del empleo podremos ver, entre otros indicadores, el total de de personas activas en el área administrativa o el número de estudiantes en cada hogar.

Para la situación actual del empleo tenemos la tasa de paro desagregada según edades.

Por último, en este módulo también se incluye la tasa de desempleo.

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



Indicadores  
económicos



En este módulo disponemos de variables relativas a indicadores económicos que nos permiten tener una visión más profunda de la situación del área administrativa.

Entre los distintos indicadores encontramos en primer lugar las variables que nos muestran la capacidad de pago del área administrativa respecto al municipio, a la provincia, a la comunidad autónoma y al total del país.

Posteriormente tenemos tres indicadores relacionados con los ciudadanos con más dificultades económicas como son la tasa de riesgo de pobreza y el número de personas mayores y niños que están en esta situación.

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



Padrón



Información demográfica sobre la procedencia y las edades de los habitantes del área administrativa es lo que informamos en él.

Para el total de habitantes se dispone de una desagregación por género y otra por edad con rangos de cinco años. Por lo que respecta a la procedencia de los habitantes se distingue por género, lugar de nacimiento y nacionalidad. Además, se hacen agrupaciones por continentes o zonas determinadas (por ejemplo la UE) como variables adicionales.

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



## Censo Viviendas y hogares



Este módulo contiene, principalmente, información obtenida del Censo de Población, Viviendas y Hogares del año 2011. En él encontramos indicadores sobre la procedencia de las personas, estado civil o características de las viviendas, entre otros.

La granularidad disponible de este módulo es de **municipio, código postal, distrito y sección censal**.

# Bloques de información



Catastro



En este módulo encontramos variables relativas al Catastro.

En él podremos ver, entre otros, el número de inmuebles, establecimientos o recintos concretos que hay en el área administrativa. Un ejemplo podría ser determinar cuantas oficinas, hospitales, terrazas o estanques hay en una zona concreta.

Además, distinguiremos los establecimientos hosteleros según sea su tipo.

La granularidad disponible de este módulo es de **portal, inmueble, sección censal, código postal y municipio** ←

# Bloques de información



Actividad empresarial



Número de autónomos o empresas por CNAE, Ventas, Empleados ...

La información no es estadística sino que es agregada, procedente del tratamiento de la BBDD de empresas de Iberinform.

La información puede ser descargada también a Excel.

Se actualiza trimestralmente con las nuevas altas

Información agregada por **Municipio, Código Postal, Distrito y Sección Censal**

# Bloques de información



Meteorología



Información meteorológica histórica y previsiones, procedente de la Agencia Estatal de Meteorología (AEMet).

Este módulo contiene indicadores de temperaturas, precipitaciones, días de sol, nubes, lluvia y nieve, humedad relativa, viento y niveles de radiación ultravioleta. Los datos están desglosados por la estación del año, temperaturas máximas y mínimas.

Las variables de este módulo provienen de las estaciones meteorológicas registradas en la AEMET, generando una granularidad máxima a nivel de **municipio**.

# Bloques de información



Redes Sociales



Promedios de **perfiles** de residentes usuarios de redes sociales

Indicadores de **dispositivo utilizado, Nivel educacional, Estilos de vida**: *gamers, noticias y actualidad, tecnología, deporte, política, amante viajes, belleza, arte & cultura, adrenalina, paternidad, música, ciencia, business & finance, moda, gastronomía, medio ambiente, tv & cine, salud & fitness, motor, religión y espiritualidad, hazlo-tu-mismo, fotografía, mascotas, vida nocturna, consumista, cocina, fan ong, aire libre, decoración, comida rápida*

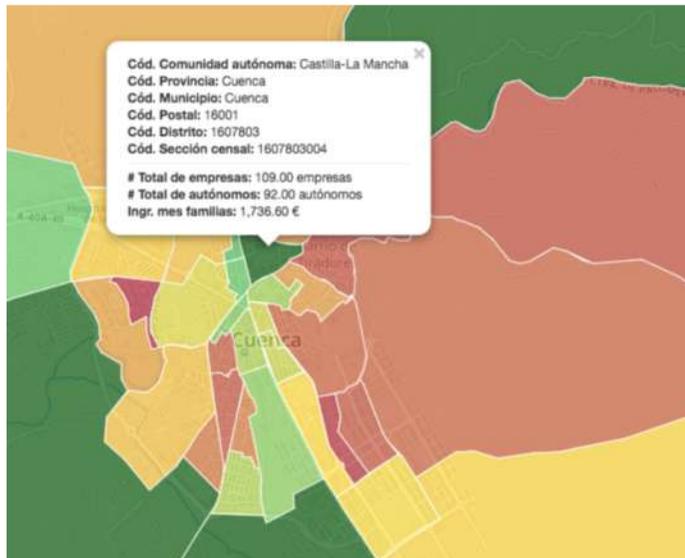
Nivel de popularidad de Aseguradoras (en razón a la cantidad de seguidores), así como de cerca de 60.000 empresas.

Nivel de agregación de **Código Postal**

# Bloques de información



Cartografía



Fichero shapfile con las coordenadas de cada capa de agregación.

Suscripción a Datamaps, webservice en el cloud con la información contratada y la zona geográfica deseada, disponible para todas las agregaciones





# AiS

GROUP

[www.ais-int.com](http://www.ais-int.com)

