

# Medición del número de viviendas turísticas en España y su capacidad

**Belén González Olmos**

*Subdirectora General de la Subdirección General de Estadísticas de Turismo y Ciencia y Tecnología. INE*

**Carlos Fullea Carrera**

*Jefe de Área de la Subdirección General de Estadísticas de Turismo y Ciencia y Tecnología. INE*

**Alfonso Fernández Bes**

*Director de Programa de la Subdirección General de Estadísticas de Turismo y Ciencia y Tecnología. INE*

## Introducción

Con este proyecto, el INE continúa el camino ya iniciado con otros proyectos similares, de utilización y aprovechamiento de nuevas fuentes de información conocidas como *big data*, para un mejor conocimiento de distintos campos de la sociedad, pero manteniendo los principios de privacidad y preservación del secreto estadístico intrínsecos en el desarrollo del trabajo del INE.

Una de las principales lagunas del actual sistema de estadísticas de turismo en España por el lado de la oferta han sido las viviendas de uso turístico. Según las encuestas de frontera, cerca de un 15% de los turistas no residentes que llegan a España y que utilizan alojamientos de mercado se alojan en viviendas de este tipo; para el turismo residente este porcentaje llega a más de un 20%, según la Encuesta de Turismo de Residentes. El número de turistas que utilizan este tipo de viviendas ha ido creciendo año a año, debido entre otras razones, al desarrollo de las compañías aéreas de bajo coste y a las posibilidades que las nuevas tecnologías brindan para contratar servicios turísticos.

Con el auge que han tenido en los últimos años las plataformas de alojamiento turístico,

la oferta y la demanda de este tipo de alojamientos han crecido de manera considerable en los últimos años. Muchas ciudades y barrios están cambiando su idiosincrasia debido al aumento de viviendas de este tipo, y es por esto que el análisis y estimación del número de este tipo de viviendas turísticas se ha vuelto fundamental, siendo muchos los usuarios que demandan esta información desde hace tiempo.

Las Encuestas de Ocupación en Alojamientos Turísticos que incluyen hoteles, apartamentos, campings y alojamientos rurales dan respuesta al reglamento 692/2011 de la Unión Europea. Este reglamento exige remitir información mensual a Eurostat de las siguientes categorías de la CNAE: 55.1 (Hoteles y alojamientos similares), 55.2 (Alojamientos turísticos y otros alojamientos de corta estancia) y 55.3 (Campings y aparcamientos para caravanas). Para proporcionar la información referente a cada una de estas categorías, se utilizan las encuestas de ocupación mencionadas. Sin embargo, dentro de la categoría 55.2 se encuentran incluidas también las denominadas viviendas de uso turístico, de las que hasta el momento no se dispone de información suficiente que nos permita analizar el impacto que tienen en el sector turístico y mejorar la calidad de los datos sujetos a reglamento. Esta problemática es común en la mayoría de los países europeos y es por eso que Eurostat ha comenzado un piloto con cuatro de las principales plataformas de alojamiento turístico para recuperar esta información.

Desde hace meses, la S. G. de Estadísticas de Turismo, Ciencia y Tecnología del INE empezó a moverse para tratar de recoger información para poder elaborar una estimación de este tipo de viviendas. La primera medida fue contactar con los responsables en materia turística de cada comunidad autónoma, para recabar la información concerniente a este tipo de viviendas de cada una de ellas. La respuesta a este

ejercicio no fue del todo satisfactoria, ya que algunas de las comunidades no disponían de un directorio de vivienda turística y otras lo tenían desactualizado. Por otro lado, desde finales de 2019 se empezaron a desarrollar programas de *web scraping* que extrajeran la información de alojamientos turísticos de las principales plataformas de alojamiento turístico, con el objetivo de ejecutar una estadística experimental que proporcionara el número de viviendas turísticas presentes en España y su capacidad.

### Extracción de los alojamientos mediante *web scraping*

El uso constante de internet por parte de la población provoca que haya una gran cantidad de información en la red disponible para ser utilizada. La utilización de esta nueva fuente para generar estadísticas puede tener un beneficio potencialmente alto, tanto en reducciones de costes y de carga a los informantes, como de acceso a información que sería de difícil acceso de otra forma. Una de las principales técnicas para obtener información procedente de la red, es el *web scraping*, que se trata de una técnica que utiliza programas de *software* para extraer información de sitios *web*. Se basa fundamentalmente en recorrer y captar la información de una página *web*, analizando la estructura particular propia del diseño de cada página y creando un conjunto de datos estructurados que puedan ser almacenados y analizados en una base de datos.

Se ha utilizado esta técnica para obtener la oferta *online* de viviendas de uso turístico en el territorio nacional. Para ello se han diseñado programas de *web scraping* específicos para tres plataformas *web* (Airbnb, Booking y Vrbo), que recorren de forma inteligente el entorno *web*, y extraen únicamente la información necesaria para la estadística.

El proceso de captura de la información se compone principalmente de dos fases. Una inicial que lista todos los alojamientos presentes en el territorio nacional y extrae variables básicas como el identificador del alojamiento, el nombre del anuncio o la url. Y una segunda fase que recorre cada anuncio a partir de la url

extraída en la primera fase y en la que se extraen las variables necesarias para la caracterización de cada alojamiento, como el número de licencia, el nombre, la capacidad o el número de dormitorios, entre otras.

La búsqueda de los alojamientos en las plataformas se ha implementado de dos maneras. Mientras que en una de las plataformas ha sido necesario realizar la búsqueda mediante literales de provincias, en las otras dos se ha utilizado un proceso que divide el territorio nacional en celdas de 10x10 km sobre las que se realizaban las peticiones. Por otro lado, como las plataformas disponen de un límite de alojamientos que mostrar en cada búsqueda, en casos en los que este se supere, en la primera plataforma se aplican filtros en la *web* que permiten bajar el nivel de alojamientos por debajo de este límite. En las otras dos plataformas, se divide la celda en cuatro y se vuelve a repetir la búsqueda sobre las nuevas celdas, así hasta conseguir un conjunto de celdas que estén todas por debajo de ese límite.

*Desde finales de 2019 se empezaron a desarrollar programas de web scraping que extrajeran la información de alojamientos turísticos de las principales plataformas de alojamiento turístico, con el objetivo de ejecutar una estadística experimental*

### Selección de la vivienda turística

En España, no hay una legislación común sobre los tipos de alojamientos turísticos. Este cometido está trasladado a las comunidades autónomas y cada una ella clasifica los alojamientos siguiendo sus propios criterios. Respecto a la vivienda turística, son pocas las comunidades que tenían una legislación propia para este tipo de viviendas hace unos años. Ha sido con el auge de las plataformas, cuando se han visto en la necesidad de legislar sobre este tipo de alojamiento. Así, por ejemplo, en

Canarias tenemos las viviendas vacacionales, en Andalucía las viviendas con fines turísticos y en La Rioja las viviendas de uso turístico.

Las plataformas de las que se extrae la información no solo anuncian alojamientos de vivienda turística, sino que tienen alojamientos de todo tipo, como hoteles, casas rurales, campings o albergues, entre otros. Por otro lado, en los últimos años, las plataformas han empezado a incorporar entre los datos que debe proporcionar cada propietario la licencia del establecimiento. No todos los alojamientos de las plataformas disponen de esta información, pero sí un conjunto amplio sobre el que se puede clasificar cada alojamiento en su tipología de acuerdo a este dato.

De este modo, se ha desarrollado un algoritmo que selecciona la vivienda turística en función de la licencia que declaran los propietarios previamente normalizada. Sobre aquellos alojamientos de los que no se dispone la licencia, la selección de vivienda turística se lleva a cabo utilizando el subtipo del alojamiento reportado en cada comunidad, tratando de cometer el mínimo error posible y

seleccionando únicamente aquellos subtipos que proporcionen una mayor proporción de alojamientos clasificados en la tipología de vivienda turística propia de cada comunidad autónoma.

### Desduplicado de alojamientos

Muchos de los propietarios que quieren disponer de su vivienda para el alojamiento vacacional, deciden publicitar la misma en varias plataformas con el objetivo de darle más visibilidad. Si se suman directamente las viviendas turísticas extraídas de las plataformas, se estaría sobreestimando de manera muy notable el número de estas. Por este motivo, ha sido imprescindible desarrollar un algoritmo que compare las viviendas descargadas de cada plataforma y determine si el alojamiento está presente en más de una.

El algoritmo desarrollado utiliza como variable principal la licencia de los alojamientos, si dos alojamientos de una misma plataforma tienen la misma licencia, se considera que son



el mismo y por lo tanto se excluye uno de ellos del conjunto final analizado. Como no todos los alojamientos tienen la licencia definida, es preciso utilizar otra metodología que permita comparar alojamientos sin licencia entre sí, o con otros en los que sí viene declarada.

Para esta segunda parte del algoritmo, se utilizan variables como el nombre del alojamiento, el nombre del anfitrión, la capacidad, el subtipo o la distancia entre alojamientos, entre otras. Cada una de estas variables tiene una aptitud diferente para desduplicar. Por poner dos ejemplos, es mucho menos probable que dos anfitriones se llamen igual, a que dos viviendas coincidan en el número de plazas que ofertan; también es menos probable que dos anfitriones hablen ruso, a que el subtipo de dos alojamientos sea casa. Después de un análisis individual de cada una de las variables extraídas y de un análisis conjunto mediante ficheros de prueba definidos para diseñar el algoritmo, se definieron una serie de pesos para cada variable que minimizaban tanto el número de alojamientos duplicados que se añaden al conjunto final, como los no duplicados que se dejan de añadir. De este modo, en las comparativas entre dos alojamientos, se definió que si, por ejemplo, los dos tienen el mismo subtipo se les daba un peso de 150, pero si los alojamientos estaban a una distancia de menos de 100 metros se le daba un peso de 1200 debido a la menor probabilidad de este caso. Así, el algoritmo compara todas las variables de dos alojamientos, y se suman los pesos de todas aquellas que coincidan, si esta suma sobrepasa un valor límite, se consideran que son duplicados y por tanto uno de los alojamientos no se incorpora al conjunto final.

Las comparaciones entre alojamientos se realizan a nivel municipio y plataforma a plataforma. Es decir, primero se seleccionan los alojamientos seleccionados como vivienda turística de la primera plataforma y se incluyen todos en el conjunto final. Después, para el mismo municipio se seleccionan las viviendas de la segunda de las plataformas y se compara una a una con los de la primera. Si en algún caso alguna de las viviendas ya estuviera en la primera plataforma, se elimina y no se incluye en el conjunto final. Este proceso se repite, para todos los municipios. Una vez se hayan

comparado las dos primeras plataformas, se incorpora la tercera, y se realiza el mismo procedimiento comparando con las dos primeras.

## Resultados

Una vez se ejecutan estas tres partes del proceso, se obtiene un conjunto final con la unión de las viviendas turísticas de las tres plataformas y habiendo eliminado las viviendas duplicadas. Este conjunto final, dispone para cada vivienda del número de plazas y de una localización geográfica a todos los niveles que permite la generación de las tablas de resultados finales.

*La granularidad geográfica alcanzada es uno de los puntos fuertes de esta estadística, ya que ninguna de las encuestas tradicionales de ocupación llega a estos niveles*

La información que se publica en estadística son el número de viviendas turísticas, las plazas, las plazas por vivienda turística y el porcentaje de vivienda turística sobre el total de las viviendas. Para esta última variable se utiliza el dato total de viviendas del Censo de Población y Viviendas de 2011, que es el último dato oficial publicado al nivel de desagregación geográfico necesario.

Estas variables se desagregan a nivel comunidad autónoma, provincia, municipio, distrito y sección censal. La granularidad geográfica alcanzada es uno de los puntos fuertes de esta estadística, ya que ninguna de las encuestas tradicionales de ocupación llega a estos niveles. Esta información se difunde en forma de tablas que permite a los usuarios realizar sus propios análisis y mediante mapas que reflejan la información de forma más visual.

La primera publicación de esta estadística referente a agosto de 2020, reflejó que el número

de viviendas turísticas en España superaba las 321.000, lo que suponía el 1,3% del total de viviendas censadas y más de un 1,6 millones de plazas. Las comunidades autónomas con mayor número de viviendas turísticas fueron Andalucía, Cataluña y Comunitat Valenciana, representando todas ellas cerca del 60% del total. En cuanto al porcentaje de vivienda turística sobre el total de viviendas, las comunidades que tenían ratios más altos fueron Illes Balears y Canarias con un 5% y un 3,6% respectivamente.

La estadística evidenció que las zonas costeras concentraban el 77% de la vivienda turística, frente al 23% de los municipios de interior. Esto se reflejó en los datos provinciales, ya que de las diez provincias con mayor número de viviendas turísticas solo Comunidad de Madrid era de interior.

A nivel municipal, los municipios de Barcelona y Madrid fueron los que mayor número de viviendas turísticas tenían en España. En torno a

17.000 viviendas turísticas cada uno, y representando un 10,6% del total nacional entre los dos. Los distritos de Eixample y Ciutat Vella en Barcelona, con cerca de 10.000 viviendas turísticas, y el Distrito Centro de Madrid, con más de 7.500 son las zonas que tenían mayor concentración de viviendas de estas ciudades. Otros municipios como La Oliva, Pollença o Begur, destacaron por tener más de mil viviendas turísticas y a su vez representar más del 20% del total de viviendas disponibles en estos municipios.

### Conclusiones

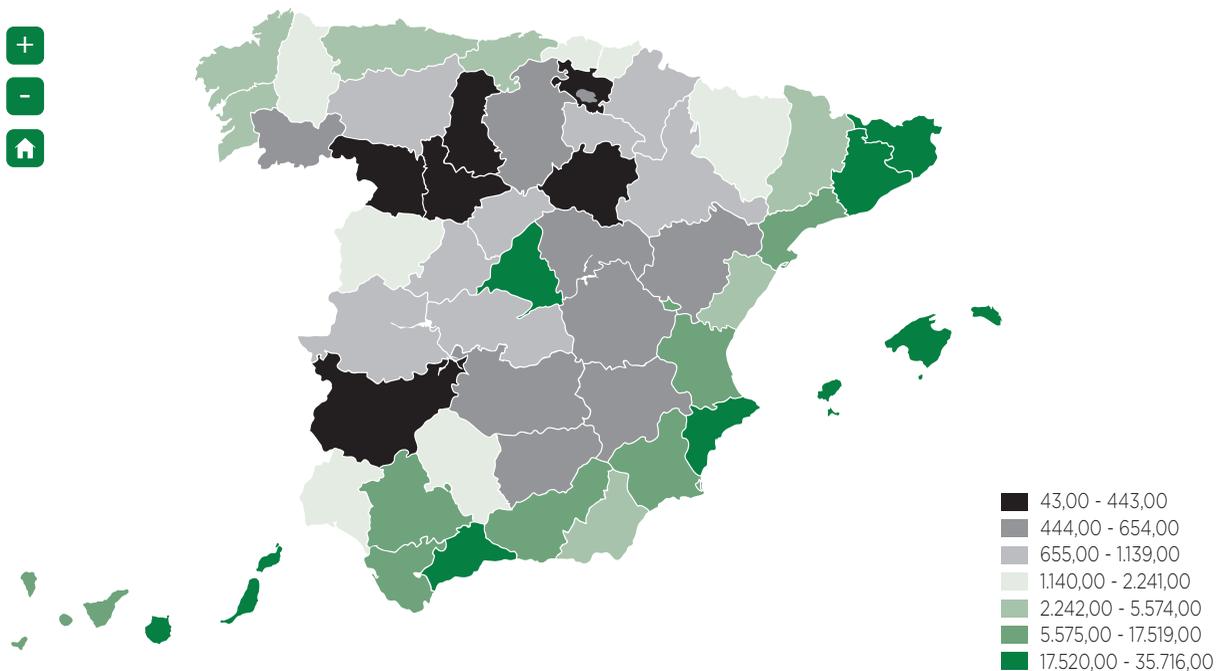
El desarrollo de esta estadística experimental ha permitido obtener por primera vez un dato aproximado del número viviendas turísticas que hay en España, algo que venía siendo altamente demandado por un gran número de usuarios que necesitaban conocer la magni-

**GRÁFICO 1. VIVIENDAS TURÍSTICAS EN ESPAÑA**

Concepto turístico  ▼

Selecciona el ámbito geográfico:  Comunidades y Ciudades Autónomas  Provincias

Área: Albacete (597 Vivienda)



**Fuente:** Medición del número de viviendas turísticas en España y su capacidad. INE.

tud de este fenómeno en el territorio nacional. Esta información no solo se proporciona a los niveles habituales de comunidad autónoma y provincia, sino que la granularidad geográfica que aporta esta operación permite evaluar este fenómeno en municipios, distritos e incluso secciones censales, lo cual puede ser de gran utilidad de cara a aplicar políticas que permitan un turismo más sostenible.

Otra de las aplicaciones posibles de la información obtenida mediante esta operación, es la posibilidad de utilizarla para definir y actualizar los directorios de otros tipos de alojamientos turísticos, o también para la elaboración de índices de precios de alojamientos. Por otro lado, no solo la utilidad de la operación en sí es relevante, sino que el conocimiento adquirido por el organismo en nuevas fuentes de información

como el *web scraping*, o del procesamiento de información proveniente de diversas fuentes mediante el emparejamiento y deduplicado de registros, es de gran importancia de cara a seguir modernizando la estadística oficial y mejorar la calidad y oportunidad de las estadísticas.

Por último, cabe añadir que el futuro de esta estadística pasa por seguir realizando nuevas publicaciones que permitan ir incorporando mejoras al proceso y así oficializarla en un futuro. Uno de los puntos clave sería poder incorporar un dato de ocupación (viajeros y pernoctaciones), para lo cual se necesitaría de una nueva fuente de datos. El proyecto que mantiene Eurostat con las principales plataformas de alojamiento y en el cual el INE colabora, proporcionará en los próximos meses esta información que podría ser incorporada previsiblemente en la estadística. ●

**CUADRO 1.** VIVIENDAS TURÍSTICAS, PLAZAS Y PLAZAS POR VIVIENDA TURÍSTICA

	Agosto 2020					
	Viviendas turísticas		Plazas		Plazas por vivienda	Porcentaje Vivienda turística
	Total	%	Total	%	Total	%
<b>TOTAL</b>	<b>321.496</b>	<b>100,0</b>	<b>1.627.377</b>	<b>100,0</b>	<b>5,1</b>	<b>1,3</b>
Andalucía	67.392	21,0	340.256	20,9	5,0	1,5
Cataluña	63.199	19,7	328.147	20,2	5,2	1,6
Comunitat Valenciana	54.638	17,0	283.900	17,4	5,2	1,7
Canarias	37.791	11,8	162.121	10,0	4,3	3,6
Balears, Illes	29.237	9,1	182.748	11,2	6,3	5,0
Madrid, Comunidad de	19.597	6,1	70.522	4,3	3,6	0,7
Galicia	12.548	3,9	67.050	4,1	5,3	0,8
Murcia, Región de	6.146	1,9	32.079	2,0	5,2	0,8
Castilla y León	5.794	1,8	33.058	2,0	5,7	0,3
Asturias, Principado de	4.891	1,5	24.698	1,5	5,0	0,8
Cantabria	4.886	1,5	24.578	1,5	5,0	1,4
País Vasco	4.598	1,4	19.424	1,2	4,2	0,5
Aragón	3.767	1,2	20.167	1,2	5,4	0,5
Castilla-La Mancha	3.327	1,0	20.317	1,2	6,1	0,3
Extremadura	1.366	0,4	6.594	0,4	4,8	0,2
Rioja, La	1.139	0,4	6.401	0,4	5,6	0,6
Navarra, Comunidad Foral de	1.071	0,3	4.906	0,3	4,6	0,3
Ceuta	66	0,0	253	0,0	3,8	0,2
Melilla	43	0,0	158	0,0	3,7	0,2

**Fuente:** Medición del número de viviendas turísticas en España y su capacidad. INE.