



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**ANÁLISIS Y VISUALIZACIÓN DE LA
OCUPACIÓN HOTELERA EN ESPAÑA
MEDIANTE POWER BI**

Autor/a: **Alejandro Torrecilla Martín**

Tutor/es/as: **D. Eusebio Arenal Gutiérrez y Alfonso
Herrero Tartilán**

Año: **2024/2025**

Agradecimientos:

Quiero expresar mi más sincero agradecimiento a todas las personas que, de una forma u otra, han contribuido a la realización de este trabajo fin de grado. Quiero agradecer al profesorado del grado de estadística y a mi tutor profesional por los conocimientos transmitidos durante estos años, base de este trabajo.

Me gustaría agradecer personalmente a mis compañeros y compañeras de clase, porque son lo mejor que me llevo de esta etapa, por su compañerismo, por sus ánimos y por las tristezas y alegrías compartidas. También me quería acordar especialmente de nuestro compañero Marcos Garcés, D.E.P.

Y sobre todo quería agradecer a mi familia y amigos, y muy especialmente a mis abuelos, que no solo han estado a mi lado en los momentos más difíciles, sino que han sido un motor constante de ánimo, cariño y fe en mí. Ellos han vivido esta etapa con tanta intensidad como yo —o incluso más—, y sé que ver cómo termino esta carrera les llena de un orgullo inmenso, quizá incluso mayor que el mío propio. Gracias por vuestro apoyo incondicional, por vuestra paciencia, y por creer en mí incluso cuando yo dudaba

A todos, gracias.

RESUMEN

El turismo, pilar clave de la economía española, demanda herramientas avanzadas para analizar su dinámica en un entorno digitalizado. Este Trabajo Fin de Grado aborda el análisis de la ocupación hotelera en España mediante una aplicación interactiva en Power BI, basada en los datos de la Encuesta de Ocupación Hotelera del Instituto Nacional de Estadística (INE).

La herramienta permite visualizar indicadores esenciales como demanda, oferta, ocupación y empleo, desglosados por niveles territoriales y temporales. A través de técnicas estadísticas avanzadas, como clustering K-means y análisis de componentes principales, se identifican patrones y tendencias turísticas.

El objetivo es ofrecer una solución práctica que facilite la toma de decisiones estratégicas para profesionales y gestores, promoviendo una gestión eficiente y competitiva del sector hotelero español.

Palabras clave: Ocupación hotelera, Turismo, España, Power BI, INE

ABSTRACT

Tourism, a key pillar of the Spanish economy, requires advanced tools to analyze its dynamics in a digitalized environment. This Bachelor's Thesis addresses the analysis of hotel occupancy in Spain through an interactive Power BI application, based on data from the Hotel Occupancy Survey by the National Statistics Institute (INE).

The tool allows the visualization of key indicators such as demand, supply, occupancy, and employment, broken down by territorial and temporal levels. Using advanced statistical techniques, such as K-means clustering and principal component analysis, it identifies tourism patterns and trends.

The goal is to provide a practical solution that supports strategic decision-making for professionals and managers, promoting efficient and competitive management of the Spanish hotel sector.

Keywords: Hotel occupancy, Tourism, Spain, Power BI, INE

ÍNDICE

RESUMEN	3
ABSTRACT	4
1. INTRODUCCIÓN, MARCO TEÓRICO Y OBJETIVOS	7
2. DESCRIPCIÓN DE LOS DATOS Y METODOLOGÍA DE LA ENCUESTA DE OCUPACIÓN HOTELERA (EOH):	9
1.2 Recogida de datos.....	9
1.2 Unidad estadística	11
1.3 Variables y ratios de estudio.....	11
1.4 Marco de la encuesta y diseño muestral.....	13
1.5 Significatividad estadística	14
2. PROCESO DE CREACIÓN DEL INFORME GRÁFICO	15
2.1 Selección y obtención de los datos.....	15
2.2 Preparación de los datos para su posterior análisis	16
2.2.1 PC AXIS	16
2.2.2 Excel	17
2.2.3 Power BI.....	19
2.3 Modelado, Relaciones y Diseño	20
2.3.1 Relaciones.....	20
3. APLICACIÓN Y MANUAL DE USO	23
3.1 Portada	24
3.2 Overview	25
3.3 Rankings Territoriales.....	29
3.4 Oferta y Estructura del Mercado	30
3.5 Ocupación	32
3.6 Demanda Total	33
3.7 Demanda Residencia.....	34
3.8 Demanda Residencia.....	35
4. ANÁLISIS ESTADÍSTICO AVANZADO	¡Error! Marcador no definido.
4.1 Cluster K-Mean	37
4.2 Análisis de Componentes Principales	39
5. ANÁLISIS PROPUESTOS.....	43
5.1 Valladolid - Provincia.....	43
5.2 Islas Canarias – CCAA.....	45
6. LIMITACIONES Y MEJORAS.....	48
ÍNDICE DE FIGURAS.....	49
Bibliografía	49

APLICACIÓN OCUPACIÓN HOTELERA EN ESPAÑA

Hipervínculo: [Ocupación Hotelera en España](#)

Manual de Uso: [APLICACIÓN Y MANUAL DE USO](#)



1. INTRODUCCIÓN, MARCO TEÓRICO Y OBJETIVOS

Introducción:

El sector turístico es un motor fundamental de la economía española, contribuyendo significativamente al Producto Interior Bruto (PIB) y al empleo. En 2023, según el Instituto Nacional de Estadística (INE), el turismo representó cerca del 12,8% del PIB nacional, consolidando a España como uno de los principales destinos turísticos del mundo. Dentro de este sector, la ocupación hotelera es un indicador clave que muestra las relaciones existentes entre la oferta (establecimientos, habitaciones y plazas) y la demanda (viajeros y pernoctaciones), proporcionando información muy importante para la gestión hotelera, la planificación estratégica y el diseño de políticas públicas.

La digitalización ha transformado la forma en que se analizan los datos turísticos, haciendo que crezca el uso de herramientas de visualización y análisis estadístico que convierten grandes volúmenes de información en conocimiento.

En este contexto, la Encuesta de Ocupación Hotelera (EOH) del INE destaca como una fuente de datos robusta, dando información detallada sobre el comportamiento y movimientos del sector hotelero a nivel nacional, autonómico, provincial y local. La integración de estos datos con plataformas como Power BI permite no solo explorar indicadores clave, sino también aplicar técnicas avanzadas para descubrir patrones y tendencias que apoyen la toma de decisiones. (Exceltur., 2024)

Marco Temporal:

En este trabajo hemos optado por centrarnos en un marco temporal que va desde mayo de 2022 hasta febrero de 2025.

La razón principal para fijar el inicio del análisis en mayo de 2022 es que, a partir de esa fecha, se considera que los efectos de la COVID-19 sobre el turismo en España ya habían finalizado y que el sector comenzaba a estabilizarse, recuperando niveles de actividad similares a los previos a la pandemia.

He optado por no incluir los años más críticos de la pandemia (2020 y 2021), ya que su incorporación podría distorsionar los resultados debido a las restricciones de movilidad, cierres temporales y otras medidas extraordinarias. Del mismo modo, he decidido no considerar los meses anteriores a la pandemia, ya que, al no incluir el periodo directamente afectado por la COVID-19, carecería de sentido comparar datos de contextos tan distintos. Además, esta elección permite trabajar con un volumen de información más manejable tanto en su descarga como en su proceso de depuración.

Por tanto, centrar el análisis en este periodo reciente permite utilizar datos más consistentes y comparables, reflejando una situación turística más estable o “normalizada”. Febrero de 2025 constituye el último mes con información publicada disponible en el momento de realizar el informe en Power BI, lo que garantiza el uso de los datos más actualizados posibles en la fecha de cierre del trabajo (aunque a día de hoy ya existen datos más recientes).

Este Trabajo Fin de Grado (TFG) se enmarca en la tendencia hacia una gestión turística basada en datos, y propone una aplicación interactiva en Power BI que facilita el análisis de la ocupación hotelera en España. A través de visualizaciones dinámicas y técnicas estadísticas

como el clustering K-means y el análisis de componentes principales, el proyecto ofrece una visión general del sector, resaltando diferencias regionales, estacionalidad y evolución reciente del mercado hotelero.

Objetivos:

El objetivo principal de este TFG es desarrollar una herramienta práctica e intuitiva que permita analizar la ocupación hotelera en España, apoyando la toma de decisiones estratégicas en el sector turístico. Los principales objetivos son:

- **Procesar y estructurar los datos** de la Encuesta de Ocupación Hotelera del INE, optimizando su formato para un análisis eficiente en Power BI.
- **Diseñar una aplicación interactiva en Power BI** que visualice indicadores clave (demanda, oferta, ocupación y empleo) con desglose territorial y temporal, incorporando mapas, gráficos y rankings dinámicos.
- **Aplicar técnicas estadísticas avanzadas**, como clustering K-means y análisis de componentes principales, para identificar patrones turísticos y segmentar destinos según su comportamiento.
- **Facilitar la interpretación de datos complejos**, proporcionando una solución accesible para profesionales del turismo, investigadores y responsables de políticas públicas y incluso para personas ajenas a este ámbito.
- **Contribuir al conocimiento del sector hotelero español**, analizando tendencias, estacionalidad y diferencias regionales para promover una gestión sostenible y competitiva.

Este trabajo combina rigor metodológico con innovación tecnológica, ofreciendo una propuesta que no solo responde a las necesidades actuales del sector turístico, sino que también establece un modelo replicable para futuros análisis y expansiones de futuros meses de este.

2. DESCRIPCIÓN DE LOS DATOS Y METODOLOGÍA DE LA ENCUESTA DE OCUPACIÓN HOTELERA (EOH):

Vamos a ver los datos relativos a la Encuesta de Ocupación Hotelera (EOH): encuesta de ocupación, índice de precios e indicadores de rentabilidad.

La información de la encuesta se ofrece mensualmente, de acuerdo con las normas de difusión del INE, 24 días después del cierre de cada mes.

Los datos disponibles permiten abordar el estudio del turismo desde una doble perspectiva: Por un lado, desde el punto de vista de la demanda, se recoge información relativa a los viajeros, las pernoctaciones y la estancia media de los no residentes en España.

Por otro lado, desde la perspectiva de la oferta, se incluyen estimaciones sobre el número de establecimientos abiertos, el número de plazas disponibles, los grados de ocupación y diversos indicadores relacionados con el empleo en el sector.

La información se obtiene a partir de distintos niveles territoriales: ámbito nacional, comunidades y ciudades autónomas, provincias, zonas y puntos turísticos concretos. (Instituto Nacional de Estadística (INE), s.f.), (Instituto Nacional de Estadística, s.f.)

1.2 Recogida de datos

Hoy en día, una gran parte de los establecimientos hoteleros, especialmente aquellos clasificados entre cinco y tres estrellas oro, utiliza sistemas informatizados de gestión llamados PMS (Property Management System). Estos programas permiten controlar de forma centralizada diversos aspectos de la gestión del hotel, como reservas, ocupación, facturación y disponibilidad. Muchos de estos hoteles disponen de integración directa con el Instituto Nacional de Estadística (INE), lo que facilita el envío automatizado de la información necesaria para la Encuesta de Ocupación Hotelera (EOH). Este intercambio de datos se realiza, en su mayoría, a través de ficheros XML, los cuales permiten una estructura uniforme de la información aportada.

El INE no especifica el número exacto de establecimientos que realizan esta integración automatizada, pero se reconoce que los datos recibidos por esta vía son consistentes y fiables.

En el caso de los alojamientos no integrados con esta aplicación, en su mayoría establecimientos de categorías plata y los hoteles de dos y una estrella oro, la participación se realiza de forma diferente. Estos establecimientos no proporcionan datos completos y continuos, sino que rellenan una muestra de variables básicas en un cuestionario físico, que se recogen en una semana aleatoria de cada mes. Esta semana corresponde a siete días consecutivos, seleccionados de forma que, sumando todas las muestras de los distintos alojamientos, se cubra la totalidad del mes.

Además, para los establecimientos clasificados en las categorías tres, cuatro y cinco estrellas oro, se realiza una segunda consulta complementaria, que recoge información agregada sobre

En definitiva, la información puede ser admitida al INE mediante cuatro opciones: a través de un cuestionario en papel o digital, mediante el envío telemático de ficheros XML, o utilizando el sistema en línea ARCE, donde los responsables del establecimiento pueden cumplimentar directamente el formulario a través de la web del INE.

INE
Instituto
Nacional de
Estadística

**Encuesta de Ocupación en Alojamientos
Turísticos. Establecimientos hoteleros**

DOCUMENTO PROTEGIDO
INE
FOLIO 000001/000001

Identificación del establecimiento

Plazo de remisión: Este cuestionario se enviará, una vez cumplimentado, en los cinco días naturales siguientes a los que se refieren los datos, a la Delegación Provincial del Instituto Nacional de Estadística, a la que podrá dirigirse para cualquier aclaración.

1. Modificaciones en la identificación (Cumplimente sólo los apartados sujetos a variación)

Nombre o razón social del establecimiento _____ NIF _____

Dirección (calle, plaza, avenida...) _____

Código Postal _____ Municipio _____ Provincia _____

Cad战略 hotelera a la que pertenece _____

Temporadas de apertura: 1ª del _____ al _____, 2ª del _____ al _____, 3ª del _____ al _____

Persona de contacto a quien dirigirse, en caso necesario, para consultas, aclaraciones o modificaciones sobre este cuestionario

FIRMA O SELLO DEL ESTABLECIMIENTO

Sr. D./Sra. Dña. _____

Cargo o puesto que ocupa: _____

Teléfono: _____ Fax: _____

E-mail (de contacto): _____

Naturalaleza, características y finalidad

Esta estadística se enmarca dentro de las Encuestas de Ocupación en Alojamientos Turísticos. Los objetivos son ampliar y actualizar los servicios estadísticos de los establecimientos de los informes a través de fuentes administrativas (art. 13.1 de la Ley de la Función Estadística Pública de 9 de mayo de 1989, «LFEPI»). Todo el personal estadístico tendrá la obligación de preservar el secreto estadístico (art. 17.1 de la LFEPI).

Legislación

Secreto Estadístico

Serán objeto de protección y quedarán amparados por el secreto estadístico, los datos personales que obtengan los servicios estadísticos tanto directamente de los informantes como a través de fuentes administrativas (art. 13.1 de la Ley de la Función Estadística Pública de 9 de mayo de 1989, «LFEPI»). Todo el personal estadístico tendrá la obligación de preservar el secreto estadístico (art. 17.1 de la LFEPI).

Obligación de facilitar los datos

Las Leyes 4/1990 y 13/1996 establecen la obligación de facilitar los datos que se soliciten para la elaboración de esta Estadística. Los servicios estadísticos podrán solicitar datos de todas las personas físicas y jurídicas, nacionales y extranjeras, residentes en España (artículo 10.1 de la LFEPI).

Todas las personas físicas y jurídicas que suministren datos, tanto si su colaboración es obligatoria como voluntaria, deben contestar de forma veraz, exacta, completa y dentro del plazo a las preguntas ordenadas en la debida forma por parte de los servicios estadísticos (art. 10.2 de la LFEPI).

El incumplimiento de las obligaciones establecidas en esta Ley, en relación con las estadísticas para fines estatales, será sancionado de acuerdo con lo dispuesto en las normas contenidas en el presente Título (art. 48.1 de la LFEPI).

Las infracciones muy graves serán sancionadas con multas de 3.005,07 a 30.050,61 euros. Las infracciones graves serán sancionadas con multas de 300,52 a 3.005,06 euros. Las infracciones leves se sancionarán con multas de 60,10 a 300,51 euros (art. 51.1, 51.2 y 51.3 de la LFEPI).

Nota: Este cuestionario está disponible en las distintas lenguas cooficiales de las comunidades autónomas.

MD-FOUR 99

	Día 1*	Día 2*	Día 3*	Día 4*	Día 5*	Día 6*	Día 7*
4.1 Número de habitaciones ocupadas _____							
4.2 Plazas supletorias y otras plazas adicionales utilizadas (no incluidas en la capacidad del hotel) _____							
5. Personal ocupado							
	Personal no remunerado			Personal remunerado			
	Fijo			Eventual			
Personal ocupado durante la semana de referencia _____							
6. Tarifa Promedio Diaria (ADR) por tipo de cliente (en euros, no incluye IVA)							
Indique el ADR (Average Daily Rate o Tarifa Promedio Diaria) para cada tipo de cliente por día incluyendo sólo alojamiento - sin desayuno, media pensión o pensión completa - de una habitación doble con baño (independientemente de su uso). Se solicita el ADR bruto, es decir, incluyendo rappel y comisiones y excluyendo el IVA.							
Señale también el porcentaje aproximado de habitaciones ocupadas por cada tipo de cliente sobre el total de habitaciones ocupadas - el mismo del apartado 4.1 - en la semana de referencia.							
	ADR en euros						% ²
Tour operador tradicional _____							
Agencia de viajes tradicional ¹ (incluye bonos y talones) _____							
Empresas _____							
Particulares (incluye tarifa normal y tarifa fin de semana) _____							
Grupos _____							
Contratación directa en la web del hotel o en la cadena hotelera _____							
Tour operador on-line _____							
Agencia de viajes on-line _____							
Otro: _____							
							100 %²
¹ Deben considerarse en este tipo de cliente los huéspedes que han entregado bonos y/o talones de hoteles como contraprestación por los servicios prestados. ² Este porcentaje será cero si no ha habido habitaciones ocupadas durante la semana de referencia.							
Observaciones:							

Información básica sobre Protección de Datos	
Responsable	Instituto Nacional de Estadística
Finalidad	Realización de esta Encuesta del Plan Estadístico Nacional
Legitimación	Artículo 10 de la Ley de la Función Estadística Pública y Disposición Adicional Segunda de la Ley 12/1996
Destinatarios	No se ceden datos a terceros salvo lo previsto en la legislación estadística
Derechos	Según el artículo 89.2 del Reglamento 2016/679 relativo a la protección de datos de las personas físicas y el artículo 2.3 de la Ley 15/1999 de Protección de Datos de Carácter Personal no están contemplados los derechos de acceso y rectificación
Información adicional	Puede consultar la información adicional y detallada sobre Protección de Datos en nuestra página web: http://www.ine.es/proteccion_datos/oe/30235

Cabe destacar que, aunque no todos los alojamientos participan con el mismo grado de detalle ni con la misma frecuencia, el impacto de los establecimientos de categorías inferiores (especialmente los de tipo plata) en los resultados globales es relativamente limitado, debido a su menor volumen y capacidad y su mayor probabilidad de error muestral, ya que los cuestionarios manuales que se realizan en este tipo de establecimientos pueden llegar a dar datos falsos.

Además, el INE aplica índices de corrección específicos que permiten minimizar el posible efecto de datos atípicos, valores perdidos o errores, lo que contribuye a mantener la fiabilidad de las estimaciones agregadas (Podemos ver como quizás haya ciertos datos que sean erróneos, pero si vemos las estimaciones agregadas en un cómputo global estas estimaciones se verán minimizadas por el resto de la muestra).

1.2 Unidad estadística

La población objeto de estudio está constituida por todos los establecimientos hoteleros existentes en el territorio nacional. Se entiende por establecimiento hotelero toda unidad productora de servicios de alojamiento hotelero (hotel, hotel apartamento o apartahotel, motel, hostel, pensión, fonda, casa de huéspedes), situada en un mismo emplazamiento geográfico y en la que trabajan una o más personas por cuenta de la misma empresa.

El INE recoge datos de cuestionarios mensuales por parte de todos los Hoteles, desde una a cinco estrellas oro y de hostales de una y dos estrellas plata de España. La categoría del establecimiento viene asignada por las Consejerías de Turismo de las Comunidades Autónomas y varía de unas comunidades a otras.

1.3 Variables y ratios de estudio

1.3.1 ESTABLECIMIENTOS HOTELEROS ESTIMADOS ABIERTOS (INE)

Número de establecimientos hoteleros abiertos de temporada estimados por la encuesta en los que el mes de referencia está comprendido dentro de su periodo de apertura.

1.3.2 HABITACIONES ESTIMADAS (INE)

La habitación se define como el cuarto o conjunto de cuartos que constituyen una unidad indivisible destinada a ser alquilada para alojamiento en el mes de referencia. Estas pueden ser individuales, dobles o múltiples según estén equipadas permanentemente para alojar a una, dos o más personas.

1.3.3 TAMAÑO MEDIO DE LOS ESTABLECIMIENTOS

Esta variable representa una estimación del número medio de habitaciones por establecimiento hotelero abierto. Se obtiene como cociente entre el número medio de habitaciones estimadas y el número medio de establecimientos abiertos estimados.

Al no ser una variable proporcionada directamente por el INE, ha sido necesario calcularla a partir de los datos disponibles para cada estrato.

1.3.4 PLAZAS ESTIMADAS (INE)

El número de plazas equivale al número de camas fijas del establecimiento, sin incluir supletorias y sabiendo que una cama de matrimonio equivale a dos plazas.

1.3.5 VIAJEROS (INE)

Todas aquellas personas que realizan una o más pernoctaciones seguidas en el mismo alojamiento clasificados por su lugar de residencia. En el caso de los residentes en España se solicita información sobre la Comunidad o Ciudad Autónoma de procedencia.

1.3.6 PERNOCTACIONES O PLAZAS OCUPADAS (INE)

Se entiende por pernoctación o plaza ocupada cada noche que un viajero se aloja en el establecimiento.

Al igual que la entrada de viajeros, las plazas ocupadas se desglosan según el lugar de residencia.

1.3.7 ESTANCIA MEDIA

Esta variable estima el número medio de días que los viajeros se alojan en los establecimientos hoteleros. Se calcula como el cociente entre el total de pernoctaciones y el número de viajeros registrados.

Dado que el INE no ofrece esta información de forma directa, ha sido necesario realizar los cálculos pertinentes para cada estrato a partir de los datos disponibles.

1.3.8 GRADO DE OCUPACIÓN POR HABITACIONES (INE)

Relación, en porcentaje, entre la media diaria de habitaciones ocupadas en el mes y el total de habitaciones disponibles en el mismo periodo.

1.3.9 GRADO DE OCUPACIÓN POR PLAZAS (INE)

Relación, en porcentaje, entre el total de las pernoctaciones y el producto de las plazas, incluyendo las camas supletorias, por los días a que se refieren las pernoctaciones.

De esta definición podemos observar que un establecimiento puede tener un grado de ocupación de plazas inferior al cien por cien y sin embargo no disponer de plazas libres, ya que un establecimiento o una habitación o cama doble puede estar ocupado solamente por una persona dando lugar a una sola pernoctación, y sin embargo el establecimiento o la habitación tener una mayor capacidad de plazas.

1.3.10 GRADO DE OCUPACIÓN POR PLAZAS EN FIN DE SEMANA (INE)

Relación, en porcentaje, entre las pernoctaciones del viernes y sábado que caen dentro de la semana de referencia y el producto de las plazas, incluyendo las camas supletorias de esos dos días, por los días a que se refieren las pernoctaciones, en este caso dos.

1.3.11 PERSONAL OCUPADO (INE)

Se define como el conjunto de personas, remuneradas y no remuneradas, que contribuyen mediante la aportación de su trabajo, a la producción de bienes y servicios en el establecimiento durante el mes que incluye el período de referencia de la encuesta, aunque trabajen fuera de los locales de este.

1.3.12 HABITACIONES OCUPADAS POR TRABAJADOR

Esta variable mide el número medio de habitaciones ocupadas por cada persona empleada en el sector hotelero. Se calcula como el cociente entre el número total de habitaciones ocupadas y el promedio de personal ocupado en los establecimientos.

Dado que esta variable no es proporcionada directamente por el INE, ha sido necesario realizar los cálculos correspondientes a partir de los datos disponibles. Su finalidad es ofrecer una aproximación a la productividad o carga de trabajo del personal en relación con la ocupación hotelera.

1.3.13 PERSONAL POR ESTABLECIMIENTO

Esta variable proporciona una estimación del número medio de personas empleadas por cada establecimiento hotelero abierto. Se obtiene dividiendo el promedio de personal ocupado entre el promedio de establecimientos abiertos estimados.

Al no estar disponible directamente en las publicaciones del INE, ha sido necesario calcularla a partir de los datos base. Su utilidad radica en ofrecer una visión aproximada del tamaño medio de la plantilla en los establecimientos, permitiendo valorar la intensidad del empleo en el sector.

1.3.14 VIAJEROS POR ESTABLECIMIENTO

Esta variable estima el número medio de viajeros que se alojan en cada establecimiento hotelero abierto durante un mes. Se calcula dividiendo el total de viajeros registrados entre el número estimado de establecimientos abiertos.

Al no estar disponible directamente en las publicaciones del INE, ha sido necesario calcularla a partir de los datos base. Su utilidad es ofrecer una visión de la afluencia media por establecimiento, permitiendo comparar la intensidad de la demanda entre distintas zonas o periodos.

1.3.15 PERNOCTACIONES POR ESTABLECIMIENTO

Esta variable representa el número medio de pernoctaciones que se registran en cada establecimiento hotelero abierto. Se obtiene como cociente entre el total de pernoctaciones y el número de establecimientos abiertos estimados en el periodo.

Al no estar disponible directamente en las publicaciones del INE, ha sido necesario calcularla a partir de los datos base. Resulta útil para analizar la ocupación media y el rendimiento de la oferta hotelera en diferentes contextos territoriales o temporales.

1.3.16 VIAJEROS POR TRABAJADOR

Indica cuántos viajeros, de media, atiende cada trabajador del sector hotelero durante el mes de referencia.

Se calcula dividiendo el número total de viajeros entre el número medio de personas empleadas. Dado que esta variable no es proporcionada directamente por el INE, ha sido necesario realizar los cálculos correspondientes a partir de los datos disponibles. Permite valorar la carga de trabajo por empleado y comparar la productividad laboral entre regiones o tipos de destino.

1.3.17 PERNOCTACIONES POR TRABAJADOR

Esta variable mide el número medio de pernoctaciones atendidas por cada persona empleada en los establecimientos hoteleros.

Se obtiene dividiendo el total de pernoctaciones entre el promedio de personal ocupado. Dado que esta variable no es proporcionada directamente por el INE, ha sido necesario realizar los cálculos correspondientes a partir de los datos disponibles. Es un buen indicador de la eficiencia operativa del personal y permite estudiar la relación entre empleo y volumen de actividad.

1.3.16 PUNTO TURÍSTICO

Municipio donde la concentración de la oferta turística es significativa. Variable categórica.

1.3.17 ZONA TURÍSTICA

Conjunto de municipios en los que se localiza de forma específica la afluencia turística. Se ofrece información de las principales zonas de interés turístico. Variable categórica.

1.3.18 TIPO DE RESIDENTE

Variable categórica que distingue entre residentes en España y residentes en el extranjero. Permite segmentar la demanda turística según su origen y analizar diferencias entre turismo nacional e internacional. La información es proporcionada directamente por el INE.

1.4 Marco de la encuesta y diseño muestral

El diseño muestral de la encuesta es un muestra aleatoria estratificado, donde los estratos están definidos por el cruce de provincia, categoría y grupo de tamaño. La variable de estratificación “grupo de tamaño” se incorpora al diseño por primera vez en la encuesta del 2025. Su objetivo es buscar una mayor homogeneidad de los establecimientos según número de plazas o capacidad.

Para la formación de estos grupos de tamaño se ha aplicado la regla de la acumulativa de la raíz cuadrada de la distribución de frecuencias (Cochran, 1977)). Esta regla divide a los establecimientos de cada conjunto formado por el cruce provincia-categoría en dos subconjuntos, de manera que en el primero están los establecimientos de menos capacidad, y en el segundo los de mayor.

Cuando el número de establecimientos del directorio por provincia-categoría es inferior a 10, no se aplica la regla anterior.

Los establecimientos que se estudian de manera exhaustiva, y por tanto pertenecen a la muestra con probabilidad 1, son los de categoría 5 estrellas de oro, y los relevantes o con gran capacidad dentro de cada cruce provincia-categoría.

Para el cálculo del tamaño muestral se ha aplicado una afijación óptima, exigiendo unos errores muestrales prefijados al estimador estratificado del total de plazas de 2,5% a nivel provincial y de 3,5%, 4%, 4,5% y 5% a nivel de provincia y 4 estrellas de oro, 3 estrellas de oro, 2 estrellas de oro y resto de categorías, respectivamente.

El tamaño muestral mínimo exigido en cada estrato provincia-categoría-grupo de tamaño, es de 5 establecimientos.

1.5 Significatividad estadística

Para que los resultados de la Encuesta de Ocupación Hotelera del INE sean estadísticamente significativos, se requiere que haya al menos 8 establecimientos en la muestra dentro del dominio analizado (por ejemplo, una provincia y categoría hotelera), con establecimientos abiertos con datos o abiertos sin movimiento, de los cuales al menos 4 deben tener establecimientos abiertos con datos en el módulo mensual. Este criterio se aplica en el módulo mensual para hoteles de 5, 4 y 3 estrellas de oro, y en el módulo semanal para el resto de las categorías.

El error de muestreo debe ser menor que el 20% o en caso de que sea mayor que el 20% el porcentaje de recogida deberá ser mayor que el 80%.

2. PROCESO DE CREACIÓN DEL INFORME GRÁFICO

A lo largo de este epígrafe se detalla el procedimiento seguido para la elaboración del informe gráfico interactivo mediante la aplicación Power BI. Este proceso ha requerido una serie de pasos consecutivos, desde la obtención y transformación de los datos brutos hasta su modelado y visualización final.

2.1 Selección y obtención de los datos

El primer paso de nuestro trabajo consistió en la obtención de los datos necesarios para el análisis hotelero. Para ello, se recurrió a la fuente oficial del Instituto Nacional de Estadística (INE), concretamente a la Encuesta de Ocupación Hotelera (EOH). Comenzamos dirigiéndonos a la página web del INE y vamos a dirigirnos a la sección de “Resultados”, abriendo seguidamente el desplegable de “Mensuales” y seguidamente clicaremos en el enlace que nos lleva a la [Encuesta de ocupación hotelera. Nacional, ccaa, provincias, zonas y puntos turísticos](#) (Establecimientos Hoteleros).



Figura 2 - Encuesta de Ocupación Hotelera - INE BASE

Una vez en este espacio, se encuentran todos los microdatos necesarios para nuestro estudio. Más en concreto, las tablas de datos en las que nos centraremos las podemos dividir en dos grupos: Viajeros y Pernoctaciones (Viajeros y pernoctaciones por comunidades autónomas y provincias, Viajeros y pernoctaciones por zonas turísticas, Viajeros y pernoctaciones por puntos turísticos) y Establecimientos Hoteleros (Establecimientos por comunidades autónomas y provincias, Establecimientos por zonas turísticas, Establecimientos por puntos turísticos).

Por comodidad a la terna de microdatos de Viajeros y Pernoctaciones y Establecimientos Hoteleros a partir de ahora les consideraremos como un bloque y nos referiremos a ellos como Pernoctaciones y Establecimientos.

Como podemos observar, el INE nos ofrece estos datos desagregados territorialmente por CCAA, Provincia, Zona Turística y Punto Turístico. No seleccionaremos la hoja de datos de Viajeros y pernoctaciones por categorías, es decir, estratificado por: Categoría del establecimiento (por ejemplo, 4 estrellas, pensión, etc.) o por Grupo de tamaño, debido a que esto no nos aporta una información relevante a la hora de realizar nuestro análisis.

El Índice de precio por comunidades autónomas es una base de datos que obtendremos en este enlace, también del INE: <https://www.ine.es/jaxiT3/Tabla.htm?t=12156&L=0> y con el que concluiremos nuestro proceso de selección de datos.

Sabemos que el INE dispone de una estructura de datos completa, es decir, si quisiéramos saber, por ejemplo, los datos de las pernoctaciones y viajeros por comunidad autónoma y provincia y además por categorías, el INE dispondría de estos datos, con la contra de que habría que pagar una gran suma de dinero para poder utilizarlos.

2.2 Preparación de los datos para su posterior análisis

2.2.1 PC AXIS

Cuando accedemos a los datos a través de la página web del Instituto Nacional de Estadística (INE), una de las opciones disponibles para su descarga es el formato PC-Axis, un estándar muy utilizado por los institutos de estadística, no solo en España y no solo en el INE, sino también en el ámbito internacional.

Cabe señalar que PC-Axis no es solo un formato de datos, sino que también cuenta con un software asociado que permite abrir, ver y manipular las tablas estadísticas descargadas. Esta herramienta facilita al usuario la selección de variables, la transformación de datos, su exportación a otros formatos (como Excel, Jason o CSV), así como la reorganización de las tablas antes de ser importadas a plataformas de análisis como Power BI.

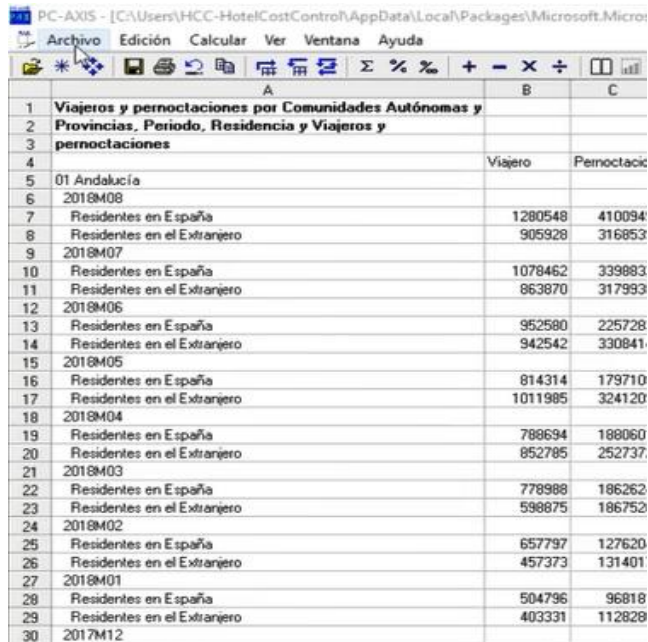
A partir de aquí y en la web del INE los pasos que seguimos consisten en seleccionar el icono de descarga de datos en formato PC-Axis (.px) y descargar cada una de las hojas de datos, con el objetivo de trabajar con ellos y estructurarlos de la manera que vamos a necesitar para nuestro análisis.

Tras abrir el archivo .px, se pueden seleccionar nuevamente las variables y el periodo de interés se nos abre un desplegable que nos permite seleccionar los datos de los que queremos disponer en cada estudio, por ejemplo: en los datos de la hoja de datos Viajeros y pernoctaciones por comunidades autónomas y provincias, se nos abren 4 desplegables:

1. Comunidades Autónomas y Provincias: En el que seleccionaremos aquellas que queremos que estén en nuestro estudio.
2. Viajeros y pernoctaciones: En la que seleccionaremos las variables *Pernoctaciones* y *Viajeros* (Nuestras variables).
3. Residencia: En la que seleccionaremos las variables *Residentes en España* y *Residentes en el Extranjero* (Variables de estratificación).
4. Periodo: El periodo que elijamos para analizar los datos es una decisión clave, ya que condiciona directamente los resultados y las conclusiones que podamos obtener. En este trabajo hemos optado por centrarnos en un marco temporal que va desde mayo de 2022 hasta febrero de 2025.

Ahora, vamos a adaptar nuestra base de datos a una estructura mucho más asumible por Power BI, para ello, accedemos al menú “Edición” → “Cambiar orden de las variables”, donde es posible arrastrar las variables que queremos (“Periodo”, “Comunidad Autónoma”, “Provincia”

o “Residencia”) al espacio destinado a las filas, y poner en las columnas las variables “Viajeros” y “Pernoctaciones”.



	A	B	C
1	Viajeros y pernoctaciones por Comunidades Autónomas y		
2	Provincias, Periodo, Residencia y Viajeros y		
3	pernoctaciones		
4		Viajero	Pernoctacion
5	01 Andalucía		
6	2018M08		
7	Residentes en España	1280548	410094
8	Residentes en el Extranjero	905928	316853
9	2018M07		
10	Residentes en España	1078462	339883
11	Residentes en el Extranjero	863870	317993
12	2018M06		
13	Residentes en España	952580	225728
14	Residentes en el Extranjero	942542	330841
15	2018M05		
16	Residentes en España	814314	179710
17	Residentes en el Extranjero	1011985	324120
18	2018M04		
19	Residentes en España	788694	188060
20	Residentes en el Extranjero	852785	252737
21	2018M03		
22	Residentes en España	778988	186262
23	Residentes en el Extranjero	598875	186752
24	2018M02		
25	Residentes en España	657797	127620
26	Residentes en el Extranjero	457373	131401
27	2018M01		
28	Residentes en España	504796	96818
29	Residentes en el Extranjero	403331	112828
30	2017M12		

Figura 3 - PC-Axis - Base de Datos Transfromada

Una vez estructurada la tabla conforme a nuestras necesidades, guardamos el archivo en formato .xlsx (Excel).

Este proceso se aplica tanto a las hojas de datos relativas a Pernoctaciones como a las correspondientes a Establecimientos. Como resultado, obtendremos un total de seis archivos Excel: tres asociados a Pernoctaciones y tres a Establecimientos, desagregadas en Comunidades Autónomas y Provincias, Puntos Turísticos y Zonas Turísticas. (Por comodidad decidimos separar CCAA y Provincias)

Estas serán finalmente las hojas de datos que vamos a moldear y finalmente exportar a Power BI para su posterior análisis.
(Instituto Nacional de Estadística, s.f.).

2.2.2 Excel

Microsoft Excel es una aplicación de hoja de cálculo de Microsoft, muy utilizada en el ámbito académico, contable y sobre todo estadístico para la gestión, análisis y visualización de datos.

Una vez contamos con todas las hojas de datos necesarias para nuestro análisis, podemos dar comienzo al proceso completo de depuración de nuestra base de datos. El objetivo de esta fase es dejar los datos listos y estructurados adecuadamente para su posterior integración en Power BI.

En primer lugar, nos centramos en revisar cómo están organizados los datos referentes a los diferentes estratos territoriales, es decir, la forma en la que el INE nos aporta los nombres de las categorías. Esta variable la denominaremos *Zona-INE*, ya que recoge la división geográfica utilizada por el Instituto Nacional de Estadística. A continuación, se detallan los distintos formatos en los que se presenta esta variable según el nivel territorial:

- **Comunidades Autónomas:** Se muestran en el formato "Número + Nombre de la Comunidad Autónoma" (1-19 incluyendo las ciudades autónomas de Ceuta y Melilla).
- **Provincias:** Aparecen simplemente con el nombre de la provincia.
- **Zonas Turísticas:** Se estructuran como "Nombre de la Comunidad Autónoma + Zona Turística + (Provincia opcionalmente)".
- **Puntos Turísticos:** Incluyen un código postal seguido del nombre del punto turístico.

Este análisis previo es fundamental para unificar y homogeneizar los nombres y formatos antes de importar los datos al entorno gráfico de Power BI, donde la consistencia de las variables categóricas es clave para un correcto funcionamiento de las relaciones y visualizaciones.

En este punto, el objetivo principal de nuestra depuración es crear una serie de variables que sirvan como filtro para una representación gráfica y visual de los datos lo más potente posible de nuestro informe en Power BI además de comprobar que nuestros datos estén correctamente organizados, sean fiables y controlar los datos faltantes y sobrantes.

Como más adelante explicaremos, uno de los objetivos de nuestro Power BI es la representación gráfica y visual de los datos, y, para ello vamos a crear un mapa mediante el cual el usuario podrá pinchar en cualquier punto o zona geográfica española según unos filtros establecidos con anterioridad y podrá ver de una forma simple y ordenada toda la información relevante.

A la hora de crear las variables “filtro” en nuestra base de datos, nos hemos encontrado con varios problemas derivados tanto de la estructura de los datos como de la presencia de valores perdidos (*missing values*).

Uno de los principales inconvenientes ha sido la gran cantidad de datos faltantes en algunas localidades. En muchos casos, estas zonas no presentaban ni una sola observación con datos en todo el periodo analizado, por lo que nos hemos visto obligados a eliminarlas del análisis. Un ejemplo es el caso de *Centre de Barcelona*, este, no contaba con ningún dato disponible o bien solo contenía registros de años anteriores al periodo de estudio.

Otro problema importante ha sido la duplicidad en la denominación de ciertos territorios. Algunas provincias, como Barcelona, aparecían tanto como “Provincia” como “Punto turístico”, con el mismo nombre en ambos casos.

Este es un proceso, que, si bien no es complicado, resulta tedioso, ya que es necesario ser muy preciso a la hora de crear estas variables.

Tras esta depuración, vamos a crear 3 nuevas variables, las variables “Filtro” (*Filtro Zona*, *CCAA-Asociada* y *Mapa*), que nos van a servir como parámetros de segmentación, como conexión entre tablas de datos y como mapeo para la posterior visualización en Power BI.

En primer lugar, creamos una variable denominada *Filtro Zona*, cuya función es identificar a qué estrato territorial pertenece cada observación. Esta variable toma uno de los siguientes valores dependiendo de la zona: “CCAA”, “Provincia”, “Zona Turística” o “Punto Turístico”. Por último, generamos la variable *Mapa*, construida a partir de la siguiente concatenación: Zona-INE, CCAA-Asociada, y España, separados por comas. Esta variable será especialmente útil a la hora de trabajar con los mapas dentro de Power BI, ya que nos facilita la jerarquización y localización geográfica.

La base de datos de *Pernoctaciones* finalmente se vería así:

1	Mes-Periodo	Periodo	Residencia	Viaje	Pernotacion	Código-INE	CCAA-Asociada	Filtro Zona	Mapa	CONCATENACION
2	febrero	01/02/2025	Residentes en España	683922	1314090	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
3	enero	01/01/2025	Residentes en España	528371	979974	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
4	diciembre	01/12/2024	Residentes en España	646330	1254332	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
5	noviembre	01/11/2024	Residentes en España	675397	1310963	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
6	octubre	01/10/2024	Residentes en España	796871	1661890	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
7	septiembre	01/09/2024	Residentes en España	991041	2493093	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
8	agosto	01/08/2024	Residentes en España	1264244	3999441	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
9	julio	01/07/2024	Residentes en España	1132902	3458067	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
10	junio	01/06/2024	Residentes en España	1107683	2621925	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
11	mayo	01/05/2024	Residentes en España	1013290	2240647	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
12	abril	01/04/2024	Residentes en España	874296	1910803	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
13	marzo	01/03/2024	Residentes en España	865895	1902651	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
14	febrero	01/02/2024	Residentes en España	705236	1373288	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
15	enero	01/01/2024	Residentes en España	569160	1017475	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
16	diciembre	01/12/2023	Residentes en España	710935	1388925	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
17	noviembre	01/11/2023	Residentes en España	673931	1279040	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
18	octubre	01/10/2023	Residentes en España	881182	1817115	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
19	septiembre	01/09/2023	Residentes en España	1040799	2586139	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
20	agosto	01/08/2023	Residentes en España	1352621	4158487	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
21	julio	01/07/2023	Residentes en España	1280780	3610922	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
22	junio	01/06/2023	Residentes en España	1118880	2649406	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
23	mayo	01/05/2023	Residentes en España	920523	1978466	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
24	abril	01/04/2023	Residentes en España	980532	2140172	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
25	marzo	01/03/2023	Residentes en España	804822	1583785	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
26	febrero	01/02/2023	Residentes en España	700088	1313951	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
27	enero	01/01/2023	Residentes en España	571285	1052987	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
28	diciembre	01/12/2022	Residentes en España	648223	1270267	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
29	noviembre	01/11/2022	Residentes en España	666655	1306991	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
30	octubre	01/10/2022	Residentes en España	867863	1821955	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
31	septiembre	01/09/2022	Residentes en España	1092918	2718541	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
32	agosto	01/08/2022	Residentes en España	1370076	4337352	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
33	julio	01/07/2022	Residentes en España	1324060	3799539	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
34	junio	01/06/2022	Residentes en España	1168278	2678849	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA
35	mayo	01/05/2022	Residentes en España	1003842	2214034	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	Andalucía, EspañaCCAA

Figura 4 - Excel Pernotaciones Base de Datos Depurada

La base de datos de *Establecimientos* finalmente se vería así:

CONCATENACIÓN														
1	Mes-Periodo	Periodo	establecimientos abie	ro de plazas e	ro de plazas e	ro de ocupación por	ro de ocupación por	Código-INE	CCAA-Asociada	Filtro Zona	Mapa	Índice	CONCATENACIÓN	
2	febrero	01/02/2025	2124	98535	121805	48,65	58,55	59,69	29061	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	132,68 Andalucía, EspañaCCAA
3	enero	01/01/2025	1983	85987	181357	39,91	47,47	48,62	25421	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	131,47 Andalucía, EspañaCCAA
4	diciembre	01/12/2024	2143	91181	194514	41,2	50,99	48,41	27138	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	144,6 Andalucía, EspañaCCAA
5	noviembre	01/11/2024	2277	99629	212885	48,2	55,37	60,46	30882	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	135,74 Andalucía, EspañaCCAA
6	octubre	01/10/2024	2599	128083	281065	58,02	66,38	70,54	42453	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	143,24 Andalucía, EspañaCCAA
7	septiembre	01/09/2024	2185	141519	316647	62,85	68,68	74,47	51124	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	135,74 Andalucía, EspañaCCAA
8	agosto	01/08/2024	2735	141178	317244	72,22	72,62	76,88	53652	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	143,04 Andalucía, EspañaCCAA
9	julio	01/07/2024	2722	140752	316434	64,8	69,85	70,6	52170	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	146,18 Andalucía, EspañaCCAA
10	junio	01/06/2024	2695	139919	315221	60,57	68,21	69,88	48916	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	144,58 Andalucía, EspañaCCAA
11	mayo	01/05/2024	2717	137443	305411	54,95	67,15	70,28	49387	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	183,27 Andalucía, EspañaCCAA
12	abril	01/04/2024	2643	130357	287820	55,28	63,61	66,19	40837	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	134,32 Andalucía, EspañaCCAA
13	marzo	01/03/2024	2510	118479	259158	51,8	57,17	62,4	35355	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	134,48 Andalucía, EspañaCCAA
14	febrero	01/02/2024	2281	98928	120964	48,55	55,75	58,75	27862	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	130,19 Andalucía, EspañaCCAA
15	enero	01/01/2024	2108	86639	189521	60,56	46,68	49,1	24539	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	126,36 Andalucía, EspañaCCAA
16	diciembre	01/12/2023	2180	91194	194905	43,18	50,78	49,97	26705	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	139,78 Andalucía, EspañaCCAA
17	noviembre	01/11/2023	2302	97950	208937	46,56	56,66	58,71	28282	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	128,29 Andalucía, EspañaCCAA
18	octubre	01/10/2023	2577	127507	279340	57,96	66	69,28	41180	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	136,38 Andalucía, EspañaCCAA
19	septiembre	01/09/2023	2738	140454	315553	62,61	67,42	73,74	49367	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	127,68 Andalucía, EspañaCCAA
20	agosto	01/08/2023	2748	141521	316971	71,74	74,4	76,5	51617	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	134,69 Andalucía, EspañaCCAA
21	julio	01/07/2023	2746	140878	314981	64,64	68,85	69,9	49989	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	136,14 Andalucía, EspañaCCAA
22	junio	01/06/2023	2757	140195	314066	58,21	66,24	67,3	47983	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	133,97 Andalucía, EspañaCCAA
23	mayo	01/05/2023	2707	134521	298069	54,29	62,54	65,57	43362	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	169,24 Andalucía, EspañaCCAA
24	abril	01/04/2023	2585	124752	274610	57,48	64,93	66,38	38475	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	133,05 Andalucía, EspañaCCAA
25	marzo	01/03/2023	2413	106109	229863	46,8	54,93	58,66	29821	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	119,11 Andalucía, EspañaCCAA
26	febrero	01/02/2023	2164	92856	197739	45,65	55,54	55,53	24712	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	122,3 Andalucía, EspañaCCAA
27	enero	01/01/2023	2066	83599	176415	38,04	45,28	46,02	22403	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	117,66 Andalucía, EspañaCCAA
28	diciembre	01/12/2022	2100	85937	181837	38,96	44,67	45,7	24064	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	123,28 Andalucía, EspañaCCAA
29	noviembre	01/11/2022	2132	91274	193103	44,04	54,87	55,32	26639	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	116,31 Andalucía, EspañaCCAA
30	octubre	01/10/2022	2417	123737	268224	54,51	62,5	65,08	38940	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	123,54 Andalucía, EspañaCCAA
31	septiembre	01/09/2022	2630	137946	307027	61,03	68,49	71,78	46759	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	119,89 Andalucía, EspañaCCAA
32	agosto	01/08/2022	2656	138623	309557	71,94	75,46	76,45	49468	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	130,74 Andalucía, EspañaCCAA
33	julio	01/07/2022	2637	137351	306992	65,55	69,17	69,71	48096	01 Andalucía	Andalucía, España	CCAA	Andalucía, España	131,49 Andalucía, EspañaCCAA

Figura 5 - Excel Establecimientos Base de Datos Depurada

Después de resolver los problemas mencionados anteriormente y con todas las hojas de datos ya depuradas procedimos a unificar las hojas en dos únicas tablas: una denominada “Establecimientos” y otra “Pernotaciones” que aúnan sus respectivas tablas de datos. Cabe recalcar que todo este proceso fue seguido de igual manera por la tabla de datos de *Índice de precios hoteleros*, que, aunque menos relevante también forma parte de nuestro estudio. Esta consolidación nos permite trabajar de forma más eficiente en Power BI, centralizando toda la información necesaria para el análisis.

2.2.3 Power BI

Finalmente, como últimos pasos en nuestro proceso de preparación de los datos importamos 3 tablas a nuestro proyecto de PowerBI que serán el núcleo central en nuestro análisis: la tabla de Establecimientos, con variables como *Grado de Ocupación* o *Establecimientos Abiertos*, la

tabla de *Pernoctaciones*, con variables como *Pernoctaciones* o *Viajeros* y la tabla de *Índice del Precio al consumidor*.

Para todas las hojas de datos que vamos a utilizar en el análisis, aplicamos una serie de pasos comunes con el fin de dejarlas completamente preparadas para su integración en Power BI. En primer lugar, promovemos la primera fila de cada tabla al encabezado, ya que contiene los nombres de las variables. A continuación, se procede a cambiar el tipo de dato de cada variable al formato correspondiente en función de su naturaleza. Por ejemplo, la variable *Periodo* se transforma al tipo *Fecha*, *Grado de Ocupación* se convierte en *Número decimal*, *Pernoctaciones* en *Número entero* y *Mapa* a tipo *Texto*.

Después, configuramos el modo en que Power BI debe computar cada variable en las visualizaciones. Así, por ejemplo, se establece que *Grado de Ocupación* se agregue como promedio y en la tabla de *Pernoctaciones* las variables *Viajeros* o *Pernoctaciones* también se agregan mediante la media, salvo que el análisis requiera lo contrario. Este paso es esencial para garantizar que, al representar los datos gráficamente, el valor que se muestre sea el deseado y tenga sentido para la interpretación.

Por último, cabe destacar un detalle importante en la limpieza de los datos: el Instituto Nacional de Estadística (INE) codifica las celdas sin valor con un punto (“.”). En nuestro caso, sustituimos estos puntos por espacios en blanco (“ ”), con el objetivo de que Power BI los ignore automáticamente durante los cálculos y no los interprete como valores erróneos o nulos.

2.3 Modelado, Relaciones y Diseño

Continuamos con las últimas fases en el proceso de nuestro trabajo. En este punto, ya disponemos de todas las tablas limpias, estructuradas y correctamente configuradas, por lo que el siguiente paso es establecer las relaciones entre ellas, definir el modelado de datos y diseñar el entorno visual que nos permita extraer conclusiones relevantes.

2.3.1 Relaciones

Uno de los pasos más importantes en la fase de modelado dentro de Power BI y una de las principales razones por las que esta herramienta destaca es su capacidad para relacionar distintas tablas de datos mediante variables clave. Esta funcionalidad permite que las tablas interactúen dinámicamente entre sí, ayuda a combinar información sin duplicar datos (Por ejemplo, puedes tener una tabla con información de hoteles (establecimientos) y otra con pernoctaciones. Si ambas están relacionadas por una variable común, como “Mapa” o “Zona”, puedes mostrar gráficos o tablas que combinen ambas fuentes sin necesidad de juntar todos los datos en una sola tabla), crear segmentaciones dinámicas (slicers), evitar errores en cálculos y agregaciones, optimizar el modelo de datos.

En nuestro caso, Power BI importa directamente las tablas correspondientes a *Establecimientos*, *Pernoctaciones* e *Índice de Precios Hoteleros (IPH)* a un panel de relaciones que ya tiene la propia aplicación. Las relaciones actuarán como segmentadoras, haciendo que podamos visualizar datos específicos desagregados por las zonas territoriales de España.

Una de las características clave de Power BI es el modelo de datos tabular, que permite optimizar el rendimiento mediante la compresión de datos en memoria, y también el uso del lenguaje DAX (Data Analysis Expressions). DAX es una herramienta muy potente que nos permite crear columnas calculadas, medidas y tablas personalizadas.

Power BI permite configurar relaciones de manera manual o automática, y visualizarlas mediante un diagrama relacional. Para nuestro objetivo, se han creado tres tablas auxiliares para reforzar el modelo y crear en todos los casos relaciones muchos-a-uno y uno-a-muchos:

- 3. La primera contiene los cuatro niveles de desagregación territorial: *Comunidad Autónoma (CCAA)*, *Provincia*, *Zona Turística* y *Punto Turístico*, de forma única en una sola columna. Esta nueva tabla ha sido denominada Filtro Zona, y su función es servir como punto de unión entre las variables geográficas presentes en las tres tablas de datos (Establecimientos, Pernoctaciones e IPH), mediante la variable común “Filtro Zona”. Gracias a esto, es posible aplicar un único filtro que afecte a todos los gráficos relacionados con coherencia de los territorios.
- 4. Del mismo modo, creamos una tabla llamada Mapa, con la columna del mismo nombre. En ella reunimos todos los valores únicos de la columna *Mapa* que aparecen en nuestras bases de datos de Establecimientos, Pernoctaciones e IPH. Esta columna recoge los nombres oficiales de Comunidades Autónomas, Provincias, Zonas Turísticas y Puntos Turísticos, seguidos del país (*España*), con el formato del INE, para que Power BI los reconozca correctamente a la hora de mostrar mapas.

Usamos como punto de unión entre nuestras distintas tablas de datos. Gracias a esta conexión a través de la variable *Mapa*, conseguimos que la información geográfica esté unificada en todo el informe. Así, por ejemplo, si seleccionamos una provincia en un mapa, automáticamente se actualizarán los gráficos relacionados con pernoctaciones, establecimientos y el índice de precios hoteleros que correspondan a esa misma provincia.

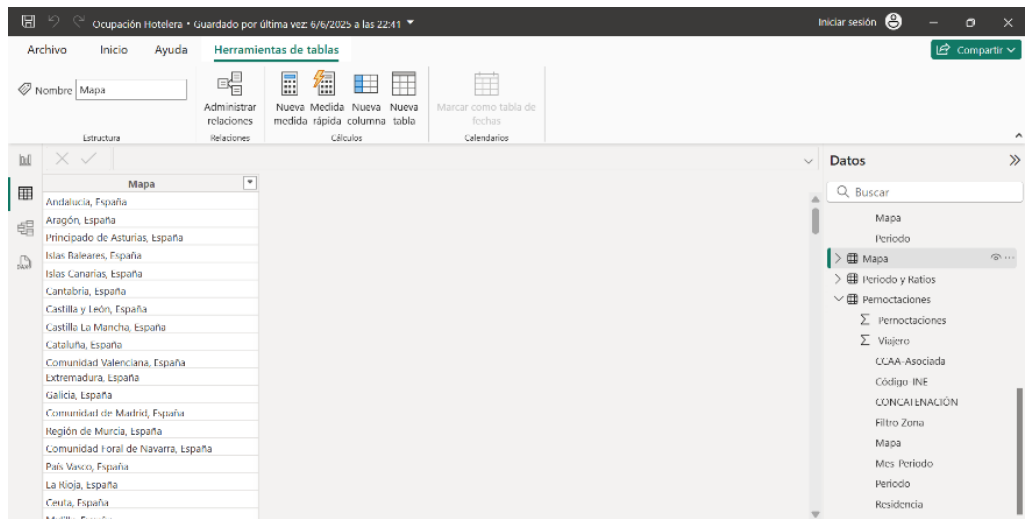


Figura 6 - Tabla de Apoyo Mapa

- Por último, vamos a crear una tabla más que llamaremos Fechas y Ratios. Esta tabla va a ser muy útil porque nos va a permitir trabajar bien con la variable Periodo, que aparece en todas nuestras tablas principales: Pernoctaciones, Establecimientos e IPH (Índice de Precios Hoteleros). Lo que vamos a hacer es juntar los periodos desde mayo de 2022 hasta febrero de 2025 en una única columna, tal y como hicimos antes con *Filtro Zona* o *Mapa*. Así, las tablas quedarán bien conectadas entre sí a través de una misma variable temporal.

Pero no solo nos va a servir como calendario. Esta tabla también la vamos a utilizar como un cajón donde guardaremos todos los ratios, medidas nuevas o variables calculadas que vayamos necesitando a lo largo del análisis.

Con esto nuestra tabla de relaciones quedaría tal que así: y estaría lista para pasar a la última fase de nuestro proyecto, el diseño gráfico.

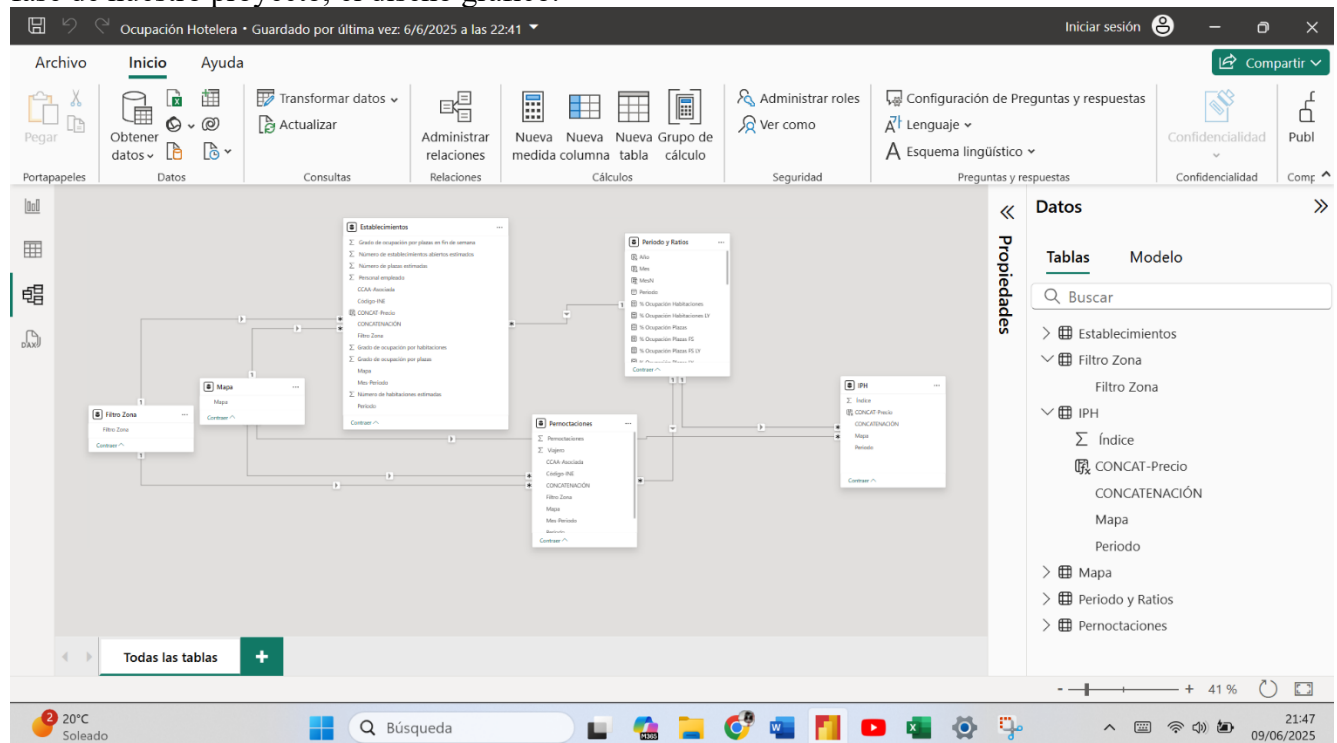


Figura 7 - Tablero de relaciones entre Tablas de Datos

5. APLICACIÓN Y MANUAL DE USO

Una vez completadas las fases de selección, obtención, depuración y modelado de los datos, y establecidas las relaciones necesarias entre las diferentes tablas en Power BI, se da paso a la etapa de diseño de la aplicación interactiva. Esta fase es crucial, ya que permite transformar la complejidad de los datos en una aplicación visual y accesible para cualquier persona —ya sea un profesional del sector (un investigador o un gestor turístico).

RECOMENDACIÓN: La forma más sencilla y recomendable para realizar una exploración visual de la aplicación desarrollada en Power BI es acceder mediante el enlace o el código QR que se presenta a continuación. Una vez abierto, únicamente es necesario iniciar sesión con una cuenta de Microsoft, ya sea utilizando el correo institucional de estudiante (@estudiantes.uva.es) o cualquier otra cuenta personal.

Hipervínculo: [Ocupación Hotelera en España](#)



Instalación:

Primeramente, se deberá instalar la propia aplicación de Power BI desktop: Abre tu navegador y accede al portal de Microsoft 365 de la UVa. Inicia sesión con tu correo académico (@estudiantes.uva.es u otro) y contraseña, y verifica que tengas disponible la licencia estudiantil o de profesor para Power BI, que normalmente está incluida en el paquete educativo.

A continuación, descarga Power BI Desktop desde el portal de Microsoft 365 haciendo clic en "Instalar aplicaciones" y buscando "Power BI Desktop", o directamente desde la página oficial de Microsoft Power BI. Asegúrate de seleccionar la versión PBIDesktopSetup_x64.exe compatible con sistemas de 64 bits.

Al abrir Power BI Desktop, haz clic en "Iniciar sesión" en la esquina superior derecha e ingresa nuevamente tu correo (@estudiantes.uva.es u otro) para vincular tu cuenta universitaria.

Para el correcto funcionamiento de Power BI Desktop, su equipo deberá cumplir con las siguientes especificaciones técnicas mínimas: Sistema Operativo Windows 10/11 (64-bit), memoria RAM de 2 GB, 1.5 GB y conexión a Internet.

Por último, para que los gráficos generados en R referentes al análisis estadístico avanzado se visualicen correctamente en Power BI, es imprescindible que el dispositivo desde el que se accede tenga R instalado. Además, deben instalarse los siguientes paquetes en dicho entorno: **dplyr**, **ggplot2**, **ggrepel**, **janitor** y **factoextra**. Sin estos elementos, las visualizaciones no podrán mostrarse correctamente dentro del informe.

Nota Importante: Power BI Online

Power BI Desktop solo está disponible para sistemas Windows. P macOS o Linux. Para estas plataformas, se recomienda utilizar el servicio Power BI en la nube mediante navegador web.

El objetivo principal de este manual es enseñar a cualquier individuo paso a paso como utilizar esta aplicación. Se facilitan los procesos para la consulta y comprensión de los indicadores clave de la actividad hotelera a distintos niveles geográficos (comunidades autónomas, provincias, zonas y puntos turísticos), permitiendo la segmentación temporal por año, mes o temporada de los datos mensuales obtenidos del INE de mayo-2022 hasta febrero de 2025.

3.1 Portada



Figura 8 - Portada - Índice

La primera diapositiva del informe interactivo desarrollado en Power BI cumple una doble función: por un lado, es la portada del proyecto y, por otro, es el índice de navegación para el usuario.

En la parte derecha, se estructura un menú de secciones o índice que permite al usuario acceder fácilmente a los distintos bloques temáticos del informe. Estas secciones han sido diseñadas para organizar la información según los ejes fundamentales del análisis.

Ahora desarrollaremos cada una de estas secciones una a una. Para poder acceder a cada una de estas diapositivas basta con seleccionarlo en los desplegables que se abren debajo de la aplicación o colocar el ratón encima de cada sección del índice, hacer CTRL+CLIC y la aplicación te llevará directamente a esta diapositiva.

3.2 Overview

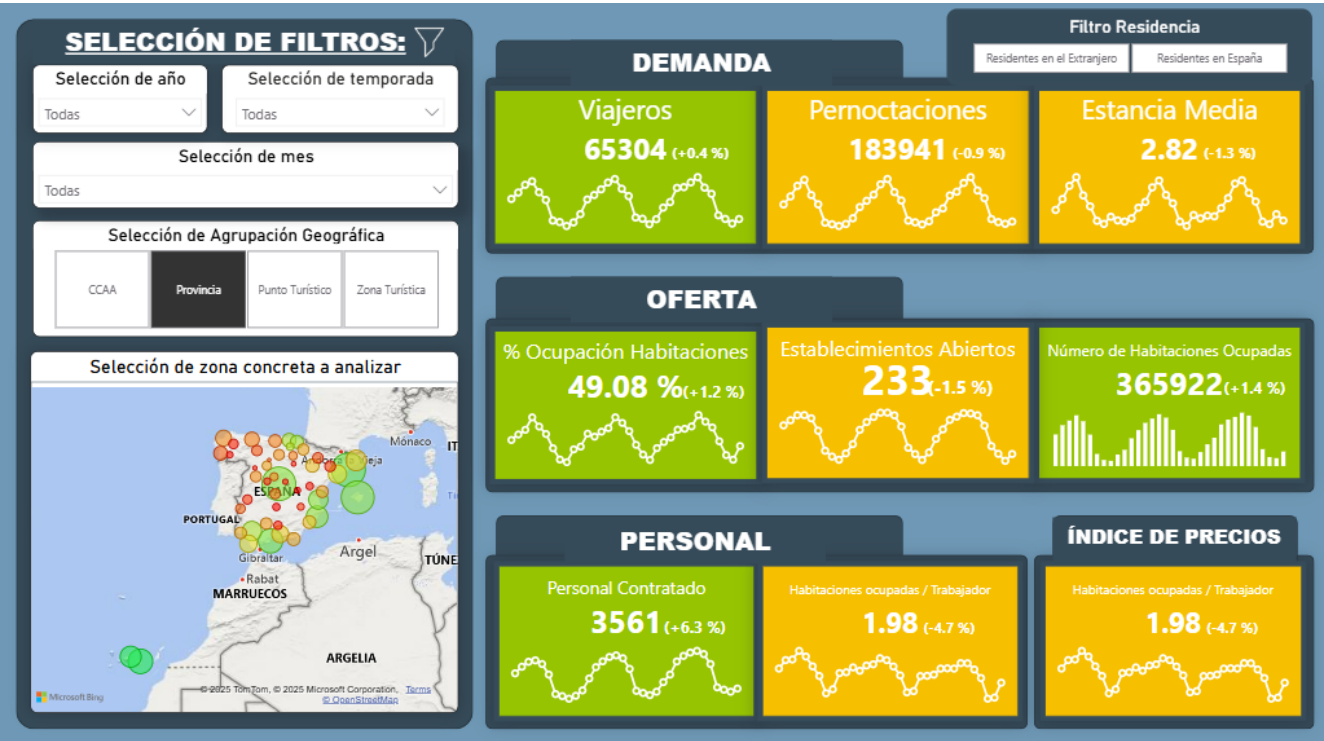


Figura 9 - Overview - Resumen

La segunda diapositiva del informe interactivo en Power BI es un resumen e introducción, desde el cual el usuario puede realizar una visión global del sector hotelero español a través de los indicadores clave, distribuidos en tres bloques: demanda, oferta y empleo. Este panel incorpora filtros personalizables, indicadores dinámicos (KPIs) y gráficos de las variables más relevantes del estudio.

Filtros de segmentación:

En la parte izquierda de la diapositiva de resumen se encuentra el panel de selección de filtros. Estos filtros, implementados mediante segmentadores (slicers) en Power BI, permiten ajustar dinámicamente los indicadores y gráficos del informe según qué datos se seleccionen.

Todos los filtros se pueden desplegar y aparecerá en todos ellos la opción de “Seleccionar todo”, esta opción sirve para seleccionar todas las opciones o segmentaciones temporales de ese filtro, y, si se quieren seleccionar un subconjunto de estas segmentaciones, podremos seleccionar un punto y mediante CNTRL+CLICK en otro punto podremos seleccionarlos simultáneamente para obtener datos agregados de dos o más segmentaciones.

Los filtros disponibles en esta sección permiten segmentar la información según las siguientes variables:

- **Año:** permite seleccionar uno o varios años dentro del periodo disponible (2022–2025), facilitando comparativas temporales o análisis por ejercicio.
- **Mes:** ofrece un filtro más granular, útil para detectar estacionalidad o identificar cambios mensuales.
- **Temporada:** clasifica los datos en función de la estacionalidad turística: verano (alta), diciembre (alta), primavera y otoño (media) y resto de invierno (baja), lo que permite observar patrones estacionales relevantes.
- **Tipo de agrupación geográfica:** mediante botones de selección, el usuario puede elegir el nivel territorial de análisis: comunidad autónoma (CCAA), provincia, zona o punto turísticos. Esta opción activa visualizaciones adaptadas al nivel geográfico seleccionado. Mediante CNTRL+CLIC en las demás secciones se pueden agregar los datos de cada tipo de filtro, aunque esto no se recomienda por la incompatibilidad de estos.
- **Zona concreta para analizar - Mapa:** una vez seleccionada la agregación geográfica, el mapa inferior permite filtrar visualmente la zona específica a analizar mediante un clic directo en el punto geográfico representado.

Además, se incorpora un filtro adicional de residencia, situado en la parte superior derecha del panel. Este selector permite diferenciar entre residentes en España y residentes en el extranjero, útil para segmentar la demanda turística en función del origen de los viajeros (Sólo cambian los datos de las variables relacionadas con la demanda ya que son con las únicas variables con las que está relacionada la variable Residencia).

Cabe destacar que todos estos filtros no solo existen en la diapositiva de resumen, sino que se encuentran en todas y cada una de las diapositivas posteriores. Además, el diseño independiente hace que los filtros que seleccionemos en una diapositiva no se arrastran a las siguientes, adaptando todos los indicadores del panel a las elecciones del usuario en tiempo real y permitiendo que el análisis sea independiente en cada diapositiva de la aplicación.

Por último, se ha añadido, junto al título “SELECCIÓN DE FILTROS”, un pequeño icono de embudo, que actúa como botón para restablecer los sarmentadores. Esta funcionalidad se activa mediante CTRL + CLIC en este icono y permite borrar rápidamente todas las selecciones realizadas.

Mapa interactivo:

Uno de los elementos más relevantes y diferenciales del informe desarrollado en Power BI es la incorporación de un mapa interactivo que permite analizar y comparar territorialmente los indicadores clave del sector hotelero. Esta funcionalidad, presente en todas las páginas del informe, constituye una herramienta especialmente útil para detectar patrones espaciales de comportamiento turístico de forma visual e inmediata.

El mapa ha sido construido mediante el componente de mapas de Power BI basado en Microsoft Bing Maps, que permite superponer puntos georreferenciados sobre el territorio nacional.

Cada punto está codificado visualmente según dos criterios fundamentales:

1. **Tamaño de la burbuja:** se encuentra asociado al número total de viajeros registrados en la zona seleccionada. De esta forma, las burbujas de mayor tamaño representan un mayor volumen de afluencia turística. Esto permite identificar de un solo vistazo los destinos más visitados en el periodo y ámbito seleccionados.
2. **Color de la burbuja:** está vinculado al grado de ocupación por habitaciones (%), siguiendo una escala de color que va del rojo (menor ocupación) al verde (mayor ocupación) pasando por el amarillo (ocupación intermedia).

Esta doble codificación (tamaño y color) permite que el mapa no solo represente dónde se concentra la demanda turística, sino también en qué medida esa demanda se corresponde con un alto grado de ocupación hotelera.

Al igual que con los filtros temporales se permite agregar diferentes zonas territoriales de una misma agrupación geográfica mediante CNTRL+CLIC en las zonas a analizar. Los gráficos se evaluarán de manera que los resultados darán los datos agregados de esas zonas.

Realmente, la principal funcionalidad del mapa es que el usuario pueda seleccionar cualquier punto del mapa para filtrar automáticamente el resto de las visualizaciones del informe en función de esa zona concreta. Esta acción activa un filtrado cruzado, lo que convierte al mapa en una nueva segmentación.

Cabe destacar que, en ausencia de una selección territorial específica, los datos se agregan automáticamente a nivel nacional, sumando o promediando los valores según la agrupación geográfica elegida (comunidad autónoma, provincia, etc.). Esto significa que, si no se selecciona ninguna unidad dentro de una categoría como CCAA o provincia, los resultados mostrados corresponderán al conjunto de España, pero ponderados según la estructura de cada nivel de desagregación. Por tanto, los valores agregados pueden diferir dependiendo de si se ha elegido analizar por comunidades autónomas, provincias u otro nivel territorial, ya que cada uno aporta un peso distinto en la composición global (Para zonas o Puntos turísticos esto no aplica al no cubrirse todo el marco territorial español).

Indicadores clave (KPIs):

En la parte central y derecha de la diapositiva de resumen (*Overview*), se presentan una serie de indicadores clave (Key Performance Indicators, KPIs) que permiten realizar una lectura rápida y estructurada del comportamiento del sector hotelero en la zona y periodos seleccionados. Estos indicadores han sido agrupados temáticamente en tres bloques principales: demanda turística, oferta hotelera y empleo/productividad, junto con un gráfico adicional centrado en la evolución de los precios hoteleros, ya que he considerado que eran los aspectos más importantes para poder tener una visión general del informe. Todos estos KPIS forman parte de una visualización que ha sido obtenida a través del Marketplace de Power BI. Esta visualización se llama “KPI Indicator” y es el pilar fundamental de nuestro análisis por su capacidad explicativa.

Cada KPI se muestra dependiendo de la variable que quiere representar con estos datos:

1. **Valor absoluto / Porcentaje** en la zona y periodo actual seleccionado.
2. **Porcentaje de variación / Valor absoluto de variación** respecto al periodo anterior (en pequeño).

3. **Gráfico de evolución temporal**, en forma de mini sparkline, que permite identificar tendencias, estacionalidades y puntos de inflexión.

Una de las principales utilidades de este KPI es permitir la visualización de la tendencia de distintas variables a lo largo de los años, especialmente al aplicar un filtro por mes específico. Esto se debe a que el indicador se basa en la variación porcentual o absoluta respecto al mismo periodo del año anterior, lo cual asegura una comparación homogénea al mantener constantes los efectos estacionales. En cambio, al analizar periodos agregados o sin filtrado mensual, esta métrica pierde parte de su significancia, ya que se podrían estar comparando momentos del año con comportamientos turísticos muy dispares. Por tanto, su uso es especialmente recomendable en contextos donde se analicen patrones temporales estacionales de forma detallada.

Posibilidades de análisis:

1. Demanda turística

Este bloque presenta una visión general de la actividad turística mediante tres indicadores clave (KPI): viajeros, pernoctaciones y estancia media. Estos indicadores permiten analizar tanto el volumen como el comportamiento de los visitantes, facilitando la detección de variaciones en la demanda y cambios en los hábitos turísticos.

2. Oferta hotelera

Este bloque analiza la capacidad y el uso del sistema hotelero mediante tres indicadores: % de ocupación de habitaciones, número de establecimientos abiertos y habitaciones ocupadas. Estos datos permiten evaluar el grado de aprovechamiento de la infraestructura hotelera y su adecuación a la demanda real, resultando clave para la planificación y el análisis del rendimiento del sector.

3. Empleo y productividad

Este bloque aborda aspectos relacionados con el empleo y la eficiencia del sector hotelero, a través de indicadores como el personal contratado y el número de habitaciones ocupadas por trabajador. Estos datos permiten analizar el impacto laboral del turismo y valorar la productividad del personal en función de la demanda, estableciendo una relación directa entre recursos humanos y rendimiento operativo.

4. Índice de Precios Hoteleros (IPH)

En la parte inferior derecha se encuentra una visualización específica sobre la evolución del Índice de Precios Hoteleros (IPH), indicador proporcionado por el Instituto Nacional de Estadística (INE). Este índice muestra el comportamiento de los precios en el sector hotelero solo a nivel de comunidades autónomas (CCAA).

3.3 Rankings Territoriales

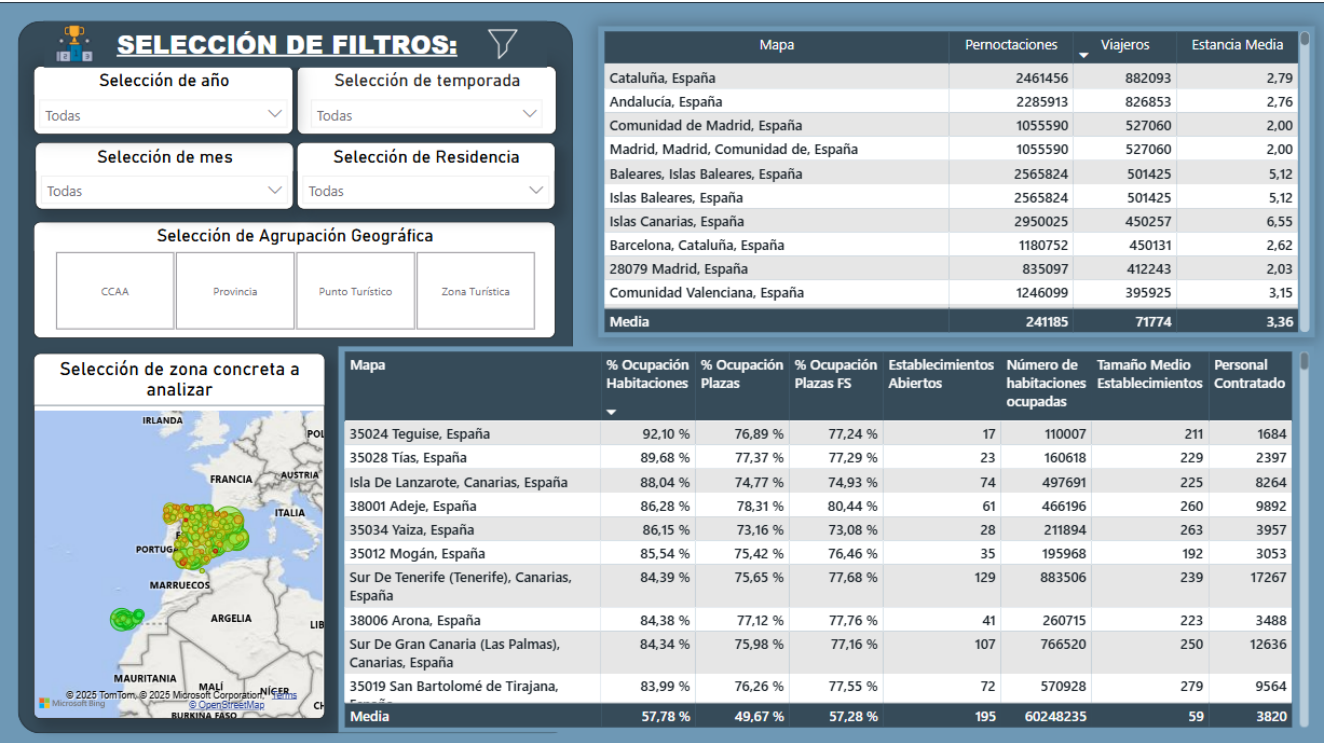


Figura 10 - Rankings Territoriales

La diapositiva de Rankings Territoriales ofrece una perspectiva comparativa y jerárquica del comportamiento turístico entre las distintas comunidades autónomas, provincias, puntos o zonas turísticas. Su objetivo es permitir al usuario identificar, de forma rápida y visual, qué territorios presentan mejores resultados en las diferentes dimensiones del sector hotelero.

Ranking superior – Demanda turística

En la parte superior derecha se dispone un panel de rankings que clasifica las zonas territoriales según tres indicadores clave de demanda turística: pernoctaciones, número de viajeros, estancia media

Este panel permite contrastar volumen (viajeros y pernoctaciones) con intensidad (duración media de las estancias), permitiendo detectar tipologías de turismo diferenciadas. Por ejemplo, regiones como Canarias y Baleares destacan por la elevada estancia media, lo que apunta a un turismo más prolongado y estacional, mientras que otras como Madrid o Castilla y León reflejan estancias más breves, típicas de viajes de trabajo o escapadas cortas.

Tabla inferior – Capacidad, uso y empleo

- % Ocupación de plazas y % Ocupación en fines de semana:** permiten evaluar la intensidad de uso del alojamiento (recordar que en una habitación puede haber 1, 2, 3 o más plazas) en diferentes momentos de la semana, ayudando a detectar patrones de concentración o dispersión de la demanda.
- Tamaño medio de los establecimientos:** refleja el tipo de infraestructura predominante. Regiones como Canarias o Baleares presentan mayores tamaños medios, lo que sugiere

- una concentración en grandes complejos hoteleros, frente a comunidades con estructuras más atomizadas.
3. **Número de establecimientos abiertos y habitaciones disponibles:** proporcionan una visión del volumen de oferta activa.
 4. **Personal contratado:** permite medir el impacto laboral directo del turismo en cada región.
 5. **Ratio de habitaciones ocupadas por trabajador:** actúa como indicador de productividad o carga laboral, útil para analizar la eficiencia operativa del sector.

La última fila de la tabla incluye los valores medios de la desagregación que estemos evaluando, facilitando así una referencia comparativa para interpretar si cada comunidad autónoma se sitúa por encima o por debajo de la media en cada uno de los indicadores.

En este tipo de tablas solo se permite evaluar el ranking para un tipo de variable, que tendremos que seleccionar y toda la tabla se transformará para ofrecer los datos ordenados de mayor a menor, o de menor a mayor si volvemos a hacer clic.

3.4 Oferta y Estructura del Mercado

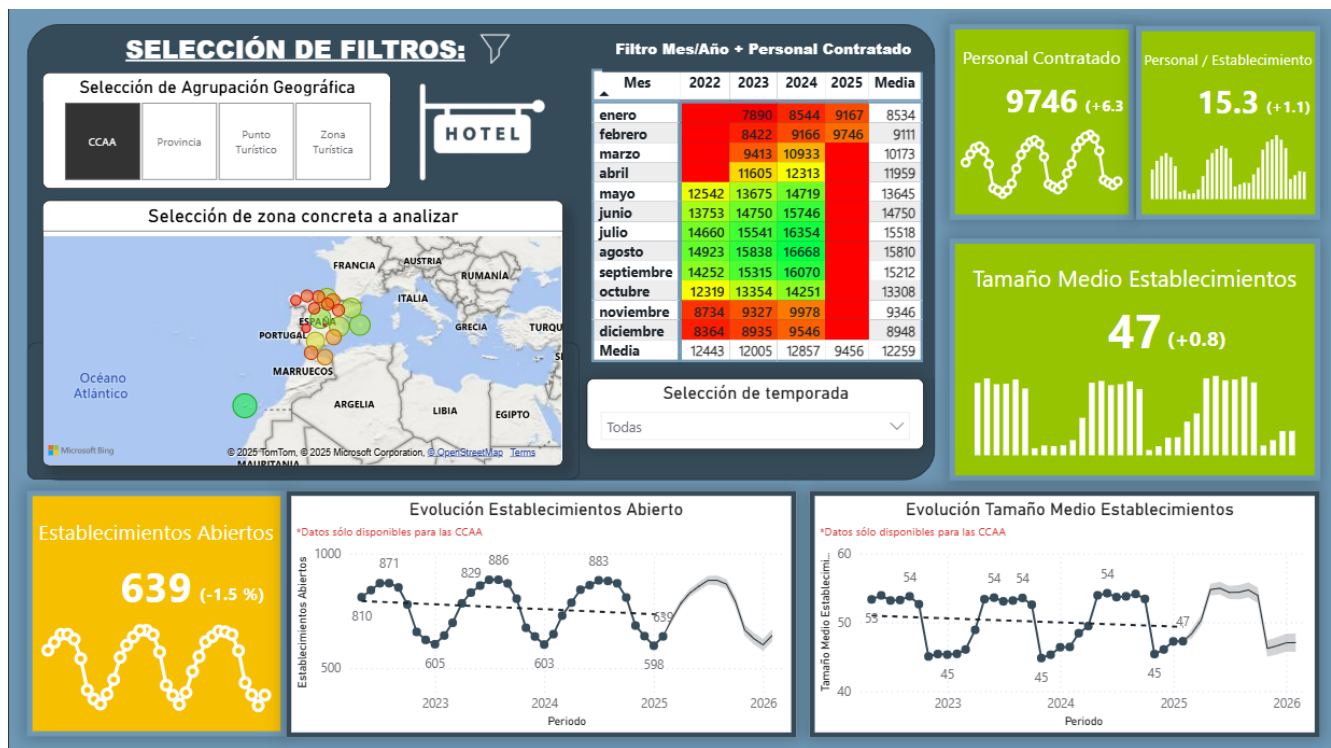


Figura 11 - Oferta y Estructura de Mercado

Esta diapositiva profundiza en la estructura de la oferta hotelera y su relación con el empleo en el sector turístico. El propósito principal es analizar la oferta, y cómo estas se traducen en estructura organizativa y empleo.

En esta diapositiva se mantiene la misma estructura de filtros que en el resto del informe, pero se incorpora un elemento visual nuevo con doble funcionalidad: un *heatmap* del personal contratado. Este gráfico actúa tanto como filtro interactivo como herramienta comparativa. Permite seleccionar directamente el año y el mes deseado haciendo clic sobre el mapa de calor,

lo que actualiza automáticamente el resto de las visualizaciones de la página. Además, ofrece una representación clara de la evolución y estacionalidad del empleo hotelero, facilitando la comparación entre distintos periodos temporales.

En la parte derecha de la visualización se presentan tres indicadores clave (KPI) que permiten medir la dimensión y estructura de los establecimientos. El único nuevo indicador es:

1. **Personal por establecimiento:** se trata de un ratio que estima el tamaño medio de la plantilla en los establecimientos hoteleros abiertos. Se calcula dividiendo el total de trabajadores contratados entre el número de establecimientos abiertos.

En la parte inferior de la diapositiva se incluyen dos gráficos temporales que permiten analizar la evolución a medio plazo de la oferta hotelera:

1. El primer gráfico muestra la evolución del número de establecimientos abiertos entre 2022 y principios de 2025: Se aprecia una tendencia cíclica marcada por los picos de verano, pero también una ligera tendencia descendente desde el año 2024. Esta evolución sugiere que algunos establecimientos, especialmente los más pequeños o menos rentables, han ido abandonando la actividad. Como ejemplo, se observa que en el año 2023 el número medio de establecimientos abiertos fue de 860, mientras que en 2025 la cifra baja a 639, lo que supone una caída del 25,7%.
2. El segundo gráfico presenta la evolución del tamaño medio de los establecimientos: Al igual que el número de alojamientos, esta variable se mantiene estable o incluso muestra un leve decrecimiento en su cómputo general, aunque si vemos en profundidad estos números en la temporada alta de verano veremos como la tendencia es positiva, lo que sugiere que los hoteles más grandes resisten mejor en el tiempo o absorben cuota de mercado. Este fenómeno puede explicarse por economías de escala: los grandes hoteles pueden afrontar mejor los costes fijos y adaptarse a fluctuaciones de demanda.

En estos gráficos se puede observar unas bandas de predicción que aparecen en los gráficos de líneas de Power BI que se generan automáticamente. Estas bandas representan el intervalo de confianza del modelo de predicción ajustado a los datos.

Power BI utiliza un modelo de regresión lineal simple o suavizado exponencial (dependiendo del tipo de datos y configuración del gráfico). Si consideramos el modelo más habitual:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

donde: y_t es la variable de interés (p. ej., número de establecimientos), t es el tiempo, ε_t es el error aleatorio, asumido como normal con varianza constante.

Cuando Power BI proyecta valores futuros (por ejemplo, para 2026), calcula la predicción puntual \hat{y} y un intervalo de confianza de tipo:

$$\{\hat{y}\}_{\{t+h\}} \pm z_{\{1-\frac{\alpha}{2}\}} \cdot \sqrt{\{Var(\{\hat{y}\}_{\{t+h\}} + \varepsilon)\}}$$

Donde: $z_{\{1-\frac{\alpha}{2}\}} = 1.96$ para el 95% de confianza (cola estándar de la normal),

$Var(\{\hat{y}\}_{\{t+h\}} + \varepsilon)$ es la varianza combinada de la estimación y del error futuro (dependiente de h , el número de pasos adelante), esta varianza aumenta con el horizonte temporal, por lo que las bandas se ensanchan conforme avanza el eje temporal.

Nota importante: Estas gráficas se calculan mediante estacionalidad de 12 meses, por lo que cualquier visualización que se haga con menos de estos meses no asegura la validez de los resultados (sobre todo a largo plazo) pero pueden seguir siendo usados para el análisis.

3.5 Ocupación



Figura 12 - Ocupación

Esta diapositiva del informe está dedicada al análisis de la ocupación hotelera, uno de los indicadores más relevantes para evaluar la eficiencia del sector turístico. A través de seis indicadores clave (KPIs), tanto en términos absolutos como relativos, se da una visión detallada de cómo se utiliza la infraestructura hotelera disponible.

En primer lugar, se representa el porcentaje de ocupación de habitaciones, el indicador más importante, que indica la proporción de habitaciones utilizadas respecto al total disponible durante un periodo y zona concreta. A su lado se ha incorporado un gráfico de líneas con un modelo de predicción estacional con una banda de confianza del 95%. La línea de tendencia muestra una clara estacionalidad, con picos de ocupación en los meses de verano (julio y agosto, superando el 80 %) y mínimos en los meses de invierno (enero y febrero, con valores en torno al 52–55 %).

Además de este indicador, se presentan el porcentaje de ocupación de plazas (que mide el uso de camas, no de habitaciones) y el porcentaje de ocupación de plazas en fin de semana, que permite aislar el comportamiento del turismo de corta duración o espontáneo.

Las cifras muestran que la ocupación de plazas es sistemáticamente inferior a la de habitaciones, lógico considerando que no todas las camas disponibles en una habitación se ocupan en cada reserva.

Junto a los indicadores porcentuales, se presentan también valores absolutos: el número total de habitaciones ocupadas, el número de plazas ocupadas y las plazas ocupadas exclusivamente en fines de semana. Estos KPIs ofrecen una dimensión bruta del volumen de turistas alojados como complemento. En el periodo actual seleccionado, se observa un crecimiento respecto al

año anterior: +0,5 % en habitaciones ocupadas, +1,5 % en plazas y +2,2 % en fines de semana, lo que podría interpretarse como una señal de recuperación de la demanda turística o incluso de crecimiento estructural del sector, desde el COVID.

Un gráfico adicional situado en la parte inferior derecha muestra la evolución del Índice de Precios Hoteleros (IPH) mensual, desagregado por año (2022 a 2025). Este indicador, proporcionado por el INE, permite analizar cómo varía el precio medio de los alojamientos en función de la estacionalidad. Se observa una clara correlación entre los picos de ocupación y los meses con mayor IPH, especialmente en los veranos de 2023 y 2024, donde se superan los 160 euros de media. En cambio, los meses de temporada baja, como enero o noviembre, presentan índices por debajo de los 130 euros. Esto sugiere la existencia de estrategias activas de gestión de tarifa estacional, mediante las cuales los precios se ajustan en función de la demanda esperada. Estos datos solo están disponibles para las CCAA.

3.6 Demanda Total

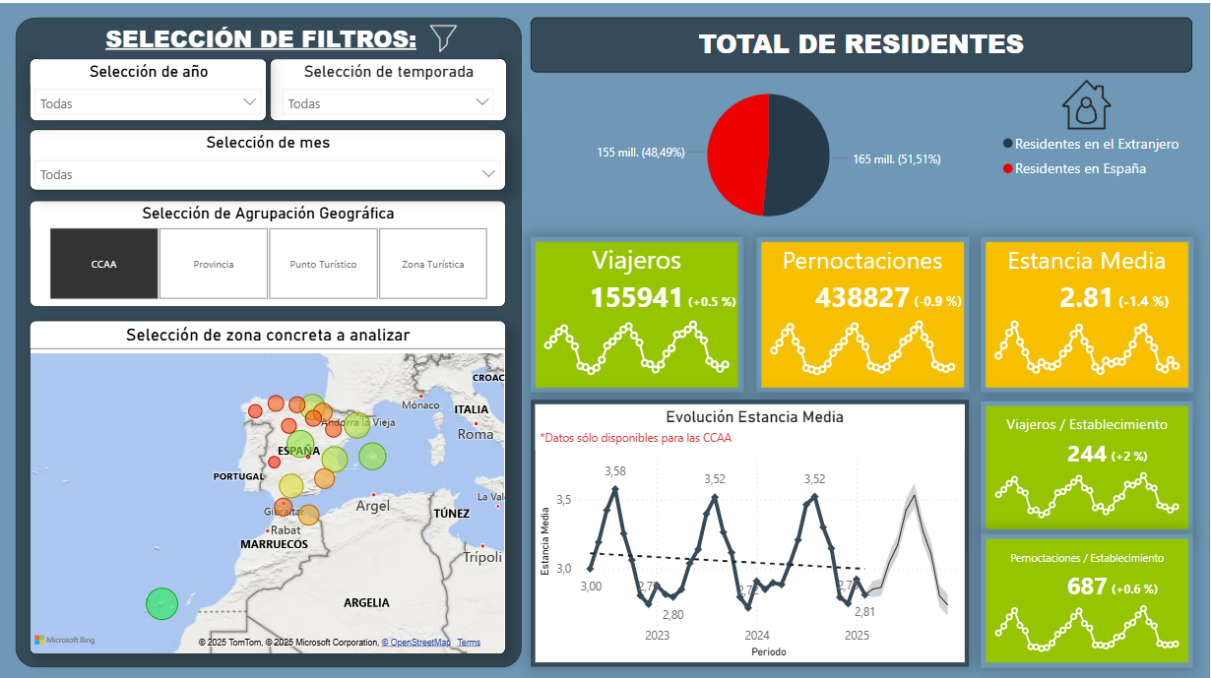


Figura 13 - Demanda Total

Esta diapositiva ofrece una visión global de la demanda hotelera en España, incluyendo tanto a los residentes nacionales (españoles viajando por España) como a los extranjeros.

En el centro superior de la diapositiva se presenta un **gráfico de sectores**, cuya función va más allá de la representación visual del porcentaje y valor absoluto de viajeros por tipo de residencia en España: también actúa como un filtro interactivo que permite alternar rápidamente entre el análisis conjunto de los residentes o focalizar el estudio en uno de los dos grupos —nacionales o extranjeros— mediante el CLIC en cualquiera de los dos sectores del gráfico.

Los indicadores más relevantes desde el punto de vista de la demanda son los siguientes:

1. **Viajeros:** mide el número total de personas que han realizado al menos una pernoctación en un establecimiento hotelero durante el periodo seleccionado.

2.

Pernoctaciones: Representa el número total de noches registradas por los viajeros en los establecimientos. Si dos regiones reciben el mismo número de viajeros pero una acumula el doble de pernoctaciones, podría deducirse que en la segunda los turistas hacen estancias más largas, lo cual tiene mayor impacto económico en el territorio.
3.

Estancia media: Es una medida que se obtiene dividiendo las pernoctaciones entre el número de viajeros. Indica cuántas noches, en promedio, permanece un viajero en el establecimiento. Si se detecta una bajada progresiva de la estancia media en un destino urbano, podría ser reflejo de un cambio hacia el turismo de escapadas breves.
1.

Viajeros por establecimiento: expresa el número medio de personas alojadas por cada alojamiento hotelero abierto en el periodo analizado. Este indicador permite evaluar la concentración de la demanda y detectar posibles desequilibrios entre afluencia y capacidad disponible.
2.

Pernoctaciones por establecimiento: indica la carga media de noches registradas por cada hotel abierto. Este indicador, además de mostrar la intensidad del uso de los alojamientos, puede ser útil para valorar su rentabilidad operativa.

Estos dos últimos indicadores, al estandarizar los datos por establecimiento, permiten realizar comparaciones más homogéneas entre zonas con distinta infraestructura hotelera o tamaño poblacional.

En la parte inferior de la diapositiva se incluye una visualización temporal en forma de gráfico de líneas, que representa la evolución mensual de la estancia media a lo largo de varios años. Este gráfico permite identificar fácilmente los patrones estacionales, como picos en verano o caídas en invierno. Una bajada sostenida de la estancia media podría señalar cambios en los hábitos turísticos, mayor predominancia del turismo nacional de corta duración o mayor rotación de visitantes, con implicaciones relevantes para la gestión de la demanda y la oferta.

3.7 Demanda Residencia



Figura 14 - Demanda Residencia

En esta diapositiva se presenta la comparación entre turistas nacionales e internacionales dentro del contexto de residencias turísticas. Esta segmentación permite identificar diferencias clave en el comportamiento de ambos perfiles mediante la visualización conjunta de estos con el objetivo de crear estrategias personalizadas.

Uno de los usos más relevantes de esta visualización es la identificación de perfiles de demanda: mientras que el turista nacional suele concentrar sus viajes en periodos vacacionales cortos, con estancias más breves y con motivaciones ligadas al ocio o las visitas familiares, con el turista internacional se puede ver una mayor dispersión estacional y una duración media de estancia más elevada. Esta diferencia resulta clave a la hora de diseñar estrategias de precios, estancias mínimas, o campañas promocionales segmentadas.

Además, al permitir aplicar filtros geográficos, esta herramienta facilita el análisis por zonas regionales. Por ejemplo, en zonas donde el turismo internacional es predominante, se puede analizar su evolución con el tiempo y su sensibilidad ante crisis externas, o, por ejemplo, puede ser también útil para detectar oportunidades de desestacionalización si se observa que el turismo internacional mantiene una estancia estable fuera de los meses de verano.

3.8 Demanda Residencia

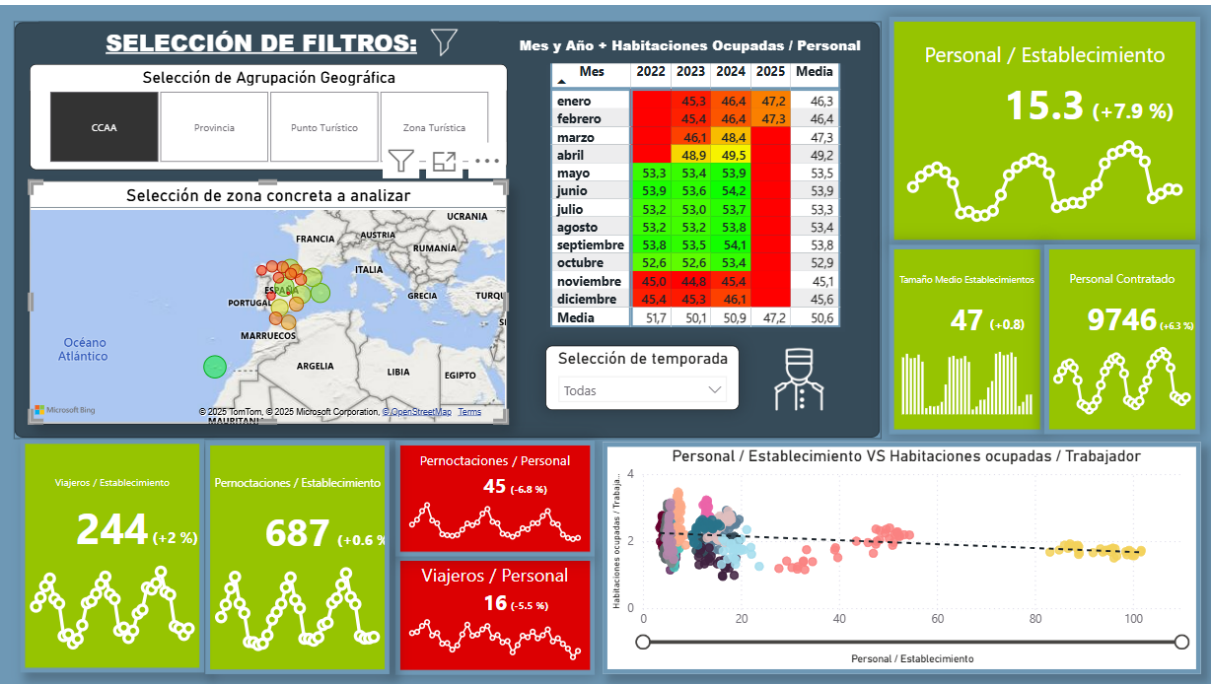


Figura 15 - Empleo y Productividad

Esta diapositiva se centra en el análisis del empleo turístico vinculado al alojamiento y su relación con indicadores clave de productividad. A través de distintos KPIs, gráficos y segmentaciones, permite estudiar el equilibrio entre la demanda turística y los recursos humanos disponibles.

Variables y KPIs representados

1. **Personal / Establecimiento:** Número medio de trabajadores por establecimiento turístico. Refleja la intensidad de contratación y la dimensión de los equipos operativos.
2. **Personal Contratado:** Número total de trabajadores en los establecimientos analizados.
3. **Tamaño medio de los establecimientos:** Media de habitaciones por establecimiento.

4. **Viajeros / Establecimiento y Pernoctaciones / Establecimiento:** Miden la carga de trabajo media que asume cada alojamiento.
5. **Viajeros / Personal y Pernoctaciones / Personal:** Indicadores de productividad individual, útiles para evaluar la eficiencia de los recursos humanos.
6. **Habitaciones ocupadas / Trabajador** (tabla mensual): Permite estudiar la estacionalidad de la carga laboral media. El heatmap sirve para detectar sobrecargas en temporada alta o subutilización en baja o para comparar la evolución entre años para evaluar la eficiencia organizativa.

Gráfico de dispersión: Relaciona el ratio Personal/Establecimiento con Habitaciones Ocupadas/Trabajador, para visualizar dinámicas entre tamaño de plantilla y carga de trabajo. Podemos ver de una forma promediada por los establecimientos para observar mejor las comparaciones entre zonas territoriales; como en la mayoría de los casos a mayor personal contratado por establecimiento menor carga de trabajo para los trabajadores y detectar claramente grupos de zonas territoriales agrupadas por características comunes. Este es especialmente útil desde un punto de vista territorial. Por ejemplo, permite identificar si los establecimientos más grandes logran mantener ratios equilibrados entre personal y carga de trabajo, o si existe una tendencia a la sobrecarga en pequeñas estructuras.

4. ANÁLISIS ESTADÍSTICO AVANZADO

Power BI ofrece la posibilidad de integrar código R directamente dentro de sus visualizaciones. Esta funcionalidad permite a los usuarios ejecutar análisis estadísticos más complejos, aprovechar las capacidades gráficas de R y trabajar con modelos que van más allá de las opciones estándar de Power BI.

Una de las ventajas principales de esta integración es que podemos programar scripts en R utilizando las propias variables cargadas en el modelo de Power BI. Es decir, una vez que seleccionamos los campos relevantes desde la interfaz de Power BI, se genera automáticamente un archivo que R puede leer como un dataset, lo que nos permite realizar transformaciones, cálculos estadísticos y representar gráficamente los resultados con la potencia de paquetes como ggplot2.

El desarrollo de los estadísticos y gráficos presentados se estructuró en tres fases. Primero, se realizó una investigación preliminar para definir las variables clave y el enfoque analítico. Luego, estas ideas se implementaron en RStudio mediante scripts que aplican técnicas como el clustering con K-means. Finalmente, los resultados fueron adaptados a Power BI con el objetivo de visualizarlos de forma clara e interactiva. Este informe se centra en analizar e interpretar en profundidad los patrones identificados a partir de dichas visualizaciones.

4.1 Cluster K-Mean

Con el propósito de identificar y clasificar zonas geográficas según sus características turísticas comunes, hemos aplicado un análisis de *clustering* utilizando el algoritmo K-means en el entorno de R. Esta técnica permite segmentar regiones en grupos homogéneos basados en distintas variables relacionadas con el turismo, ofreciendo una visión más clara y estructurada de los distintos perfiles territoriales.

El algoritmo K-means es uno de los métodos más utilizados en el análisis no supervisado. Su objetivo es organizar un conjunto de observaciones en un número predeterminado de grupos (k), de forma que los elementos dentro de cada grupo sean lo más similares posible entre sí, y a la vez lo más diferentes posible de los elementos de los demás grupos. El proceso consiste según Hastie et al. (Hastie) en estos pasos:

Seleccionar aleatoriamente k centroides iniciales.

Asignar cada observación al centroide más cercano, formando los grupos.

Recalcular los centroides como la media de las observaciones asignadas a cada grupo.

Repetir los pasos 2 y 3 hasta que los grupos no cambien o se alcance la convergencia.

Aplicación del Clustering a la Demanda Turística:

En el análisis de la demanda turística, se seleccionaron tres variables clave: Media de Viajeros, Media de Pernoctaciones y Estancia Media, todas ellas extraídas de la hoja de datos de Pernoctaciones. Estas variables reflejan el comportamiento turístico promedio de cada zona.

Aplicando el algoritmo K-means (función `kmeans()`) con $k = 3$ y tras aplicar una estandarización de las variables con la función `scale()` para que la media sea 1 y la desviación típica 0, logramos identificar tres grupos diferenciados de zonas turísticas, que pueden interpretarse de la siguiente manera:

Zonas con alta afluencia turística y estancias prolongadas.

Zonas con un nivel menor pero también alto de visitantes y estancias largas.

Zonas con menor volumen turístico o visitas más breves.

Esta segmentación ha sido representada en Power BI a través de dos gráficos interactivos:

Estancia Media vs. Número de Viajeros

Estancia Media vs. Pernotaciones

En ambos gráficos, los puntos están coloreados según el clúster al que pertenecen, facilitando una lectura clara de los patrones de comportamiento de los visitantes en cada grupo. La elección de tres grupos se basó en pruebas sucesivas que demostraron que esta división ofrece la mayor claridad interpretativa.

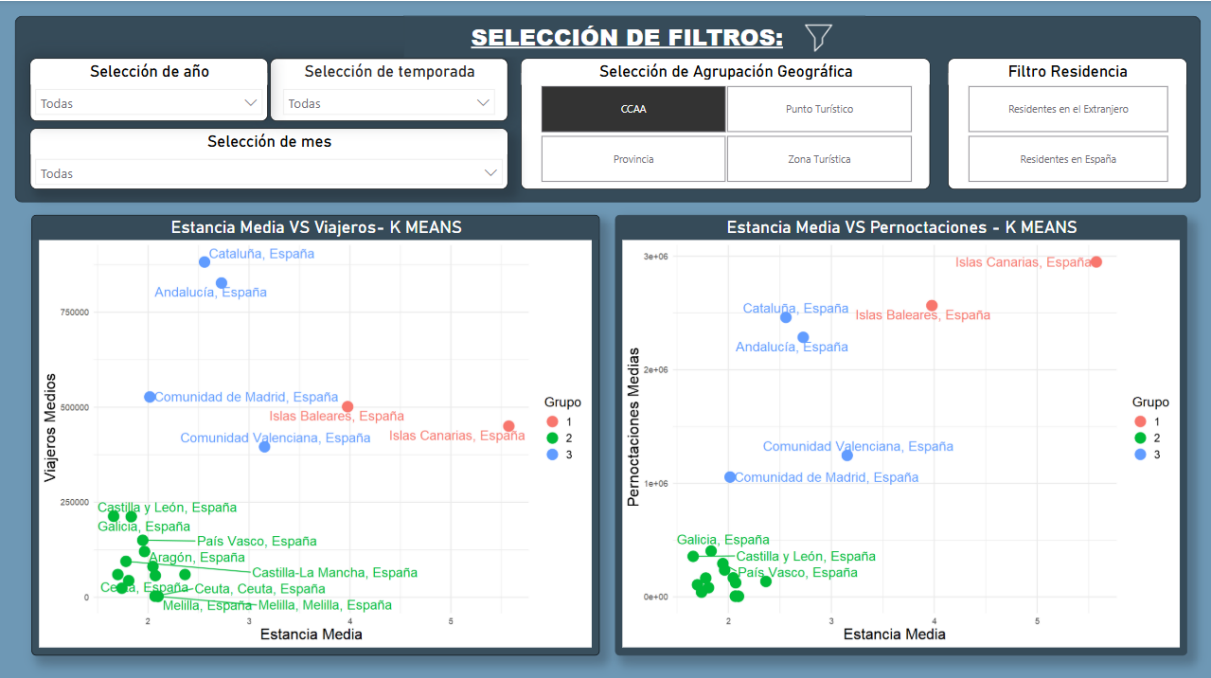


Figura 16 - Cluster Demanda

Aplicación del Clustering a la Oferta Turística:

También se aplicó el algoritmo K-means (función `kmeans()`) para analizar la oferta turística, utilizando como variables: Grado de Ocupación, Número Medio de Plazas y Personal Empleado. Para garantizar una comparación equitativa entre variables, se realizó una estandarización previa de los datos mediante la función `scale()` para que la media sea 1 y la desviación típica 0.

Al igual que en el caso anterior, se estableció $k = 3$, permitiendo identificar tres tipos de zonas en función de su infraestructura y actividad turística:

Zonas con alta ocupación y recursos turísticos consolidados.

Zonas con nivel intermedio de infraestructura y ocupación.

Zonas con infraestructura limitada o baja intensidad turística.

En Power BI, este análisis se visualiza mediante un gráfico en el que:

El eje X representa el Grado de Ocupación por Plazas (%).

El eje Y muestra el Número Medio de Plazas.

El tamaño de los puntos refleja la cantidad de personal empleado, lo que añade una dimensión visual sobre la escala operativa de cada zona.



Figura 17 - Cluster Oferta

4.2 Análisis de Componentes Principales

En este caso hemos decidido hacer un análisis de Componentes principales de la Base de datos de Establecimientos con las variables (Grado de ocupación por habitaciones, Grado de ocupación por plazas, Grado de ocupación por plazas en fin de semana, Número de establecimientos abiertos estimados, Número de habitaciones estimadas, Número de plazas estimadas, Personal empleado).

El Análisis de Componentes Principales es una técnica estadística multivariante que permite reducir la dimensión de un conjunto de datos con muchas variables cuantitativas, manteniendo la mayor parte de la variabilidad presente en los datos originales.

Dado que estas variables se expresan en distintas unidades de medida (porcentajes, recuentos, etc.), se optó por emplear la función `princomp()` de R con el argumento `cor = TRUE`. Esta opción asegura que el análisis se realice a partir de la matriz de correlaciones y no de la matriz de varianzas y covarianzas, lo cual es fundamental cuando las variables tienen escalas y magnitudes heterogéneas. De este modo, cada variable contribuye de igual forma al análisis, independientemente de su rango de valores original.

Interpretación PCA

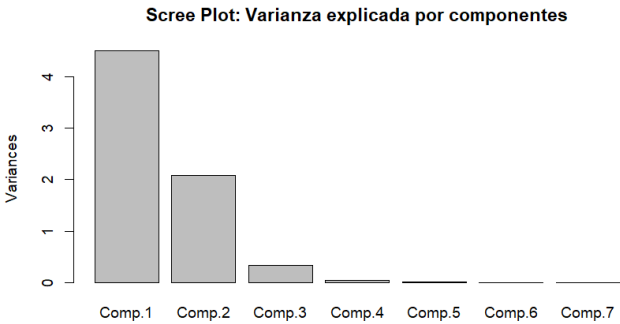


Figura 18 - Scree Plot: Varianza Explicada

Varianza explicada:

Los resultados del PCA (Scree Plot sirve para verlo de forma visual) muestran que la primera componente principal (Componente 1) explica por sí sola un 64,44% de la varianza total de los datos. La segunda componente (Componente 2) aporta un 29,75% más. En conjunto, ambas componentes explican el 94,2% de la variabilidad de los datos originales.

Este valor es notablemente alto y permite afirmar que el biplot en dos dimensiones generado con estas dos componentes recoge prácticamente toda la información relevante del conjunto de datos. Por tanto, no es necesario un gráfico tridimensional ni la interpretación de componentes adicionales para extraer buenas conclusiones.

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	2.120645	1.4429721	0.58038377	0.231908614	0.139181972	0.099204945	
Proportion of Variance	0.642448	0.2974526	0.04812076	0.007683086	0.002767374	0.001405946	
Cumulative Proportion	0.642448	0.9399007	0.98802146	0.995704543	0.998471918	0.999877864	
	Comp.7						
Standard deviation	0.0292396116						
Proportion of Variance	0.0001221364						
Cumulative Proportion	1.0000000000						

Figura 19 - Importancia de las Componentes

Cargas (loadings):

Las cargas indican cuánto aporta cada variable a cada componente.

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
grado_de_ocupacion_por_habitaciones	0.350	0.452		0.545	0.604		
grado_de_ocupacion_por_plazas	0.359	0.441		0.272	-0.771		
grado_de_ocupacion_por_plazas_en_fin_de_semana	0.333	0.467	0.188	-0.784	0.142		
numero_de_establecimientos_abiertos_estimados	0.315	-0.395	0.821			0.238	
numero_de_habitaciones_estimadas	0.424	-0.296	-0.132			-0.544	0.643

Figura 20 - Cargas PCA

Componente 1:

La Componente 1 se caracteriza por tener cargas positivas significativas en prácticamente todas las variables, destacando especialmente:

- Número de plazas estimadas (0,425)
- Número de habitaciones estimadas (0,424)
- Personal empleado (0,422)
- Grado de ocupación por plazas (0,359)
- Grado de ocupación por habitaciones (0,350)

Esto sugiere que esta primera componente representa una dimensión general de tamaño y actividad hotelera. Las zonas con valores altos en esta componente se corresponden con aquellas que cuentan con una mayor infraestructura hotelera (más plazas y habitaciones), mayor ocupación y una plantilla de personal más amplia. Es decir, podríamos interpretarla como un índice global de volumen turístico.

Componente 2:

Por su parte, la Componente 2 presenta una relación inversa entre algunas variables. Tiene cargas positivas altas en:

- Grado de ocupación por habitaciones (0,452)
- Grado de ocupación por plazas (0,441)
- Pero a su vez presenta cargas negativas en:

Número de habitaciones estimadas (-0,296)
 Número de plazas estimadas (-0,288)
 Personal empleado (-0,236)

Esto sugiere que esta segunda componente diferencia entre zonas que, a pesar de tener una infraestructura más reducida, presentan una alta ocupación, frente a otras zonas con gran capacidad, pero menor rendimiento relativo. En este sentido, esta componente puede interpretarse como una medida de eficiencia de ocupación o de intensidad en el uso de la oferta hotelera.

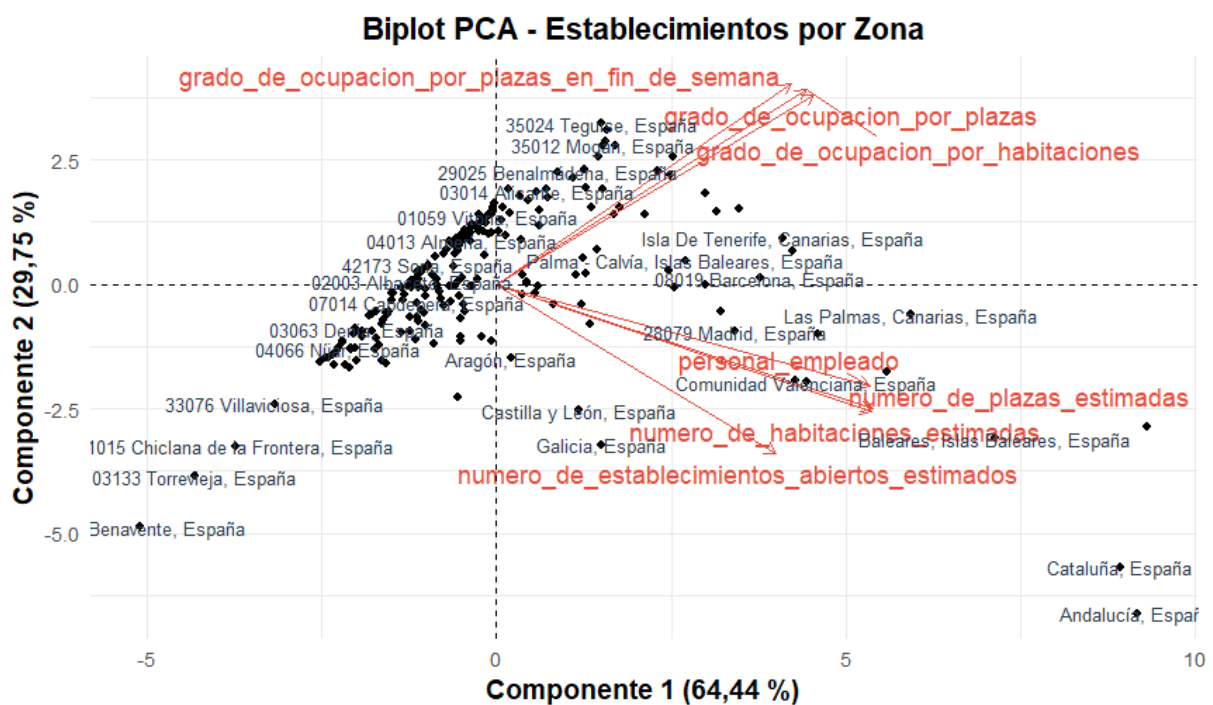


Figura 21 - Biplot PCA

En el eje horizontal (Componente 1), que explica el 64,44% de la varianza total, se observa una clara diferenciación entre zonas en función del tamaño y desarrollo de su infraestructura hotelera. Las zonas situadas en el extremo derecho del gráfico —como Cataluña, Andalucía, Canarias, Baleares o Madrid— presentan valores elevados en variables como el número de establecimientos abiertos, número de plazas y habitaciones estimadas, y el personal empleado. Esta disposición indica que estas comunidades representan los principales focos turísticos nacionales, tanto por su capacidad alojativa como por su volumen de empleo asociado al sector.

Además, zonas como Canarias y Baleares también aparecen alineadas con las variables relativas al grado de ocupación (por plazas, por habitaciones y durante el fin de semana), lo que sugiere que, además de contar con una infraestructura amplia, también muestran altos niveles de utilización y eficiencia del recurso hotelero disponible. Es decir, no solo tienen muchos establecimientos, sino que estos están, en general, bien aprovechados.

En el lado opuesto del gráfico, hacia la izquierda, se encuentran zonas con una menor intensidad turística, como Benavente o Villaviciosa (Torrevieja y Chiclana seguramente estén ahí por falta de datos). Su posición indica bajos valores en las variables de ocupación y volumen, lo que

puede interpretarse como un menor peso del sector hotelero en su economía o una menor capacidad de atracción turística.

En cuanto a la Componente 2, que explica un 29,75% adicional de la varianza, introduce un matiz interesante: diferencia zonas no solo por su tamaño, sino por el rendimiento relativo de su infraestructura hotelera. Así, algunas zonas situadas en la parte superior del gráfico —como Calvià, Isla de Tenerife o Las Palmas— destacan por presentar altos niveles de ocupación con una estructura relativamente menor, lo que sugiere un uso intensivo y eficiente del recurso. En cambio, zonas algo más abajo, como Galicia o Castilla y León, tienden a tener menores grados de ocupación, a pesar de disponer de cierta infraestructura.

La orientación de los vectores también proporciona información clave sobre las correlaciones entre variables. Por ejemplo, las tres variables relacionadas con el grado de ocupación apuntan en direcciones muy similares, lo que indica una alta correlación positiva entre ellas. De igual forma, las variables de infraestructura (plazas, habitaciones, establecimientos y personal) también muestran una fuerte alineación, sugiriendo que las zonas que disponen de una mayor capacidad hotelera tienden a concentrar también más empleo.

5. ANÁLISIS PROPUESTOS

A partir de los datos disponibles y mediante el desarrollo del panel interactivo en Power BI, ha sido posible extraer una serie de conclusiones clave sobre el comportamiento del sector hotelero en España. He seleccionado 2 zonas de la geografía española a analizar, con el objetivo de ver sus similitudes, diferencias y todo aquel dato relevante a analizar. Este es el objetivo de la aplicación, la asimilación de los datos de forma práctica

Importante: Inciso muy importante, todas estas conclusiones están sacadas con datos de los meses de marzo 2022 a febrero de 2025, por lo que es muy probable que ciertos datos estén un poco sesgados, al no disponer de un ciclo de años completo (nos faltan los meses de marzo abril y mayo)

5.1 Valladolid - Provincia

Demanda:

Valladolid ocupa una posición modesta en el ranking nacional, situándose en torno al puesto nº22 en número de viajeros, con una media mensual de 29.000. En cuanto a pernoctaciones, registra unas 110.000 al mes, frente a la media nacional de 50.500, lo que refleja un flujo turístico limitado. Su estancia media, de 1,74 noches, se sitúa bastante por debajo de la media nacional (3,12), lo que responde a un perfil de turismo urbano, de corta duración y con predominio de visitantes nacionales, motivados por actividades culturales, congresos o escapadas de fin de semana.

A lo largo del año se observa una tendencia decreciente en viajeros y estable en pernoctaciones, sin apenas estacionalidad: la actividad se mantiene alta todo el año, con picos en verano, y ligeros descensos en los meses de invierno. Parece que la actividad en estos últimos años se refuerza en los meses de invierno y cae ligeramente en los de verano.

Los viajeros y pernoctaciones por establecimiento aumentan ligeramente, lo que refleja una alta ocupación de la infraestructura hotelera.

Con todo esto obtenemos un perfil como destino estructuralmente estable, ideal para estrategias de desestacionalización.

Por tipo de residencia, crecen tanto los viajeros nacionales como internacionales, pero en los nacionales caen los viajeros. Se puede observar claramente en los turistas extranjeros un pico notable en agosto mientras que los nacionales suele ser una curva más estable durante los demás meses del año. Tanto en nacionales e internacionales, la estancia media crece notablemente, debido sobre todo al aumento del número de pernoctaciones, que crece al igual que el de los viajeros, pero más notablemente. En muchos meses, la demanda nacional es casi 50% superior a la extranjera, lo que confirma la fuerte dependencia nacional de este destino, con sus correspondientes oportunidades y riesgos.



Figura 22 - Valladolid Demanda por Residente

Oferta y Ocupación:

Si analizamos la oferta hotelera en Valladolid, observamos una evolución estable en el número de establecimientos abiertos, que ronda los 103 alojamientos activos durante gran parte del año. Aunque se sitúa en los puestos más bajos del ranking nacional (en torno al nº35), es una cifra coherente con su tamaño y volumen turístico. A diferencia de destinos más estacionales, la oferta no fluctúa excesivamente entre meses, mostrando una cierta continuidad a lo largo del año, pero teniendo picos a la baja en el mes de enero y una tendencia generalizada a la baja. En cuanto al empleo turístico, Valladolid registra una media mensual de 1.100 trabajadores, algo por debajo de la media nacional, aunque con una estructura más constante, sin grandes picos de contratación y manteniéndose estable.

El personal por establecimiento crece, con una ligera tendencia al alza en los últimos años, impulsado por la mejora en la calidad del servicio más que por un aumento significativo de la oferta. El tamaño medio de los establecimientos ronda las 30 habitaciones, sin grandes variaciones, lo que refleja un modelo de alojamiento más funcional y adaptado a una demanda de corta duración, propia del turismo urbano, cultural y de negocios y aumentando en los últimos años.

En cuanto a la ocupación por habitaciones, se mantiene en torno al 60-65 % de media anual, con ligeros ascensos en verano sobre todo en junio y septiembre y peores datos en los meses de primavera y otoño. Las ocupaciones por plazas y plazas en fin de semana no muestran diferencias marcadas a lo largo del año, lo que puede atribuirse a una distribución relativamente equilibrada de la demanda entre días laborables y fines de semana.

Empleo y Productividad:

En cuanto al empleo, Valladolid presenta un modelo más contenido en términos de personal contratado por establecimiento, situándose por debajo de la media nacional. Con un ratio medio de 10-15 empleados por hotel, refleja una estructura operativa ajustada al tamaño y tipo de

establecimiento predominante, generalmente de pequeñas dimensiones y con una oferta de servicios más funcional.

El indicador de habitaciones ocupadas por trabajador se encuentra ligeramente por debajo de la media nacional, lo que indica una productividad relativamente baja, probablemente asociada a una menor complejidad operativa y a un perfil de cliente que no requiere una atención especialmente intensiva.

El personal por establecimiento se mantiene estable, sin grandes variaciones estacionales, lo que facilita la continuidad del empleo a lo largo del año, aunque sin generar crecimientos sostenidos. Por otro lado, el tamaño medio de los alojamientos aumenta, lo que indica una mayor aparición de hoteles más grandes. La ausencia de aperturas de nuevos establecimientos en los últimos años también contribuye a esta estabilidad en los indicadores de empleo.

En consecuencia, los ratios de viajeros y pernoctaciones por trabajador son estables, adaptados a un volumen turístico moderado, sin grandes oscilaciones. El modelo de Valladolid se caracteriza por una gestión laboral eficiente, con plantillas ajustadas y un bajo nivel de estacionalidad, lo que favorece la sostenibilidad del empleo, aunque limita las oportunidades de crecimiento en términos de personal.

5.2 Islas Canarias – CCAA

Demanda:

Canarias ocupa el puesto nº4 en viajeros, con una media mensual de 450.257, pero lidera con diferencia en pernoctaciones, con 2.950.025, frente a la media nacional de 700.029. Su estancia media (6,55 noches) es la más alta de España, por delante de Baleares (5,12) y Comunidad Valenciana (3,15), lo que consolida a Canarias como un destino preferente para estancias largas, especialmente de turistas internacionales.

A lo largo del año se observa una tendencia creciente en viajeros y pernoctaciones, sin apenas estacionalidad: la actividad se mantiene alta todo el año, con picos en enero y verano, y ligeros descensos en marzo y abril. Esto refuerza su perfil como destino estructuralmente estable, ideal para estrategias de desestacionalización.

Los viajeros y pernoctaciones por establecimiento también aumentan, lo que refleja una alta ocupación de la infraestructura hotelera, aunque esta presión podría requerir ajustes en la oferta si la tendencia se mantiene.

Por tipo de residencia, crecen tanto los viajeros nacionales como internacionales, pero en los nacionales caen las pernoctaciones y la estancia media. En los internacionales, aunque también baja ligeramente la estancia media, esto se debe a que el número de viajeros crece más rápido que sus noches de estancia, no por un descenso generalizado. Las diferencias son notables: en muchos meses, la demanda extranjera supera en más del 50 % a la nacional, lo que confirma la fuerte dependencia internacional del destino, con sus correspondientes oportunidades y riesgos.

Oferta y Ocupación:

Si analizamos la oferta hotelera en Canarias, observamos un crecimiento sostenido en el número de establecimientos abiertos, situándose en el puesto nº10 del ranking nacional, lo cual es destacable para una comunidad no tan extensa territorialmente. Por ejemplo, en junio se pasa de 564 a 570 establecimientos, según las previsiones, lo que confirma una tendencia expansiva.

Sin embargo, donde realmente destaca Canarias es en el empleo turístico: ocupa el puesto nº1 en personal contratado, con una media mensual de 54.723 empleados, muy por encima de la media nacional (12.255). Este dato, junto al heatmap de personal contratado, refuerza la idea de que en Canarias no hay estacionalidad, ya que la contratación se mantiene estable y creciente durante todos los meses del año. Además, el indicador de personal por establecimiento también muestra una evolución claramente al alza.

En cuanto al tamaño medio de los establecimientos (habitaciones por alojamiento), se observa una tendencia a la baja, lo que podría deberse a la entrada de nuevos establecimientos más pequeños o a un cambio en el modelo de alojamiento hacia formatos más segmentados o especializados.

Respecto a la ocupación por habitaciones, la evolución es claramente creciente, con pequeñas caídas en mayo, pero el modelo predice un valor cercano al 91 % para noviembre de 2025, lo que reflejaría una utilización casi plena de la capacidad disponible. En las ocupaciones por plazas y por plazas en fin de semana, no se aprecian grandes diferencias a lo largo del año, lo que podría deberse a una ocupación muy equilibrada entre semanas y fines de semana, sin concentraciones bruscas de demanda. A nivel absoluto, Canarias lidera en habitaciones ocupadas, plazas ocupadas y plazas en fin de semana, lo que, sumado a su promedio de tamaño de establecimiento más alto de España, refleja una oferta hotelera muy activa, robusta y eficiente.

Empleo y Productividad:

En cuanto al empleo, Canarias destaca por tener el mayor ratio de personal por establecimiento de toda España, lo que refleja una alta intensidad de contratación por alojamiento. Sin embargo, el indicador de habitaciones ocupadas por trabajador se sitúa en torno a la media nacional, lo que sugiere que el incremento del personal no se traduce directamente en una mayor productividad individual, posiblemente por una apuesta del sector por ofrecer un servicio más completo o personalizado.

Si analizamos el heatmap de habitaciones ocupadas por trabajador, se observa una tendencia descendente en los últimos años, lo que puede deberse tanto al aumento progresivo del empleo como a la estabilización en el ritmo de crecimiento de la ocupación, lo que reduce la carga de trabajo media por persona.

Además, el personal por establecimiento no solo es el más alto, sino que aumenta de forma constante, mientras que el tamaño medio de los establecimientos (en habitaciones) muestra una tendencia a la baja. Esto puede estar relacionado con la apertura de alojamientos más pequeños y especializados, que requieren más personal por unidad que los grandes complejos tradicionales.

Este aumento en el número de empleados provoca, de forma natural, una reducción en los ratios de viajeros por personal y pernoctaciones por personal, ya que la plantilla crece más rápido que la demanda total. Sin embargo, esta creciente demanda de empleo responde a una evolución paralela de la demanda por establecimiento, tanto en viajeros como en pernoctaciones, que sigue creciendo de forma clara.

En conclusión, Canarias presenta un modelo turístico caracterizado por una alta contratación, baja estacionalidad y una estructura cada vez más intensiva en mano de obra, lo que refleja su consolidación como destino líder, pero también plantea retos en términos de eficiencia y sostenibilidad laboral a medio plazo.

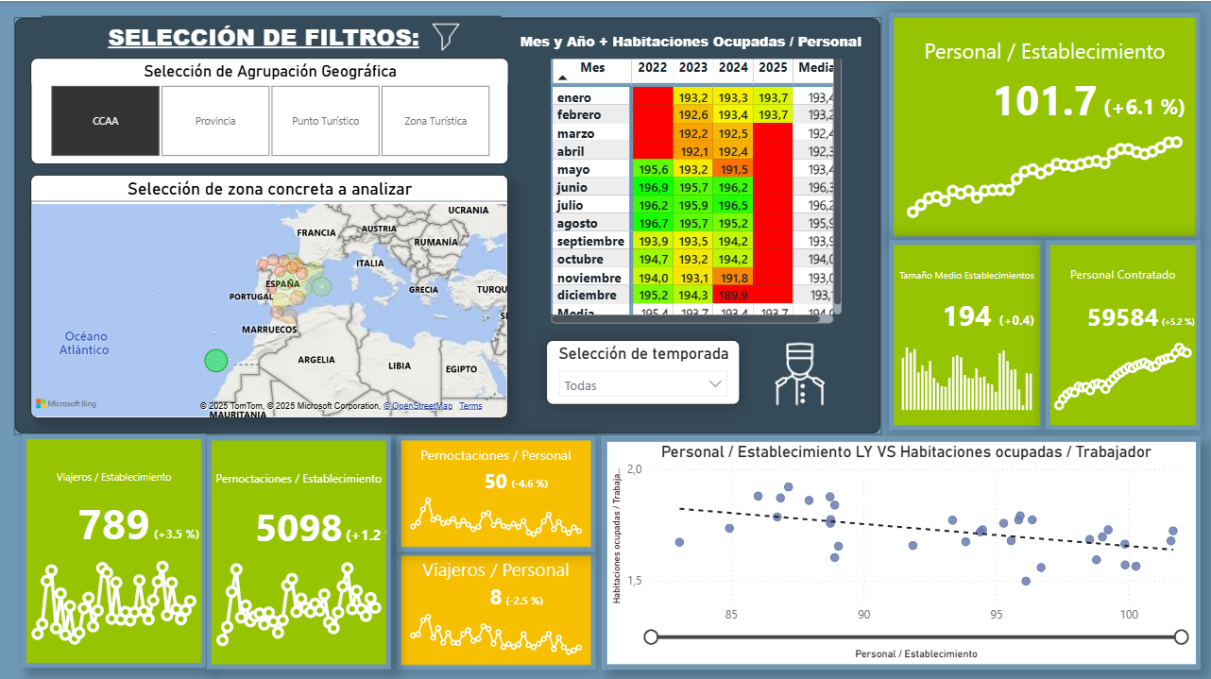


Figura 23 - Islas Canarias Empleo y Producción

6. LIMITACIONES Y MEJORAS

Una de las principales limitaciones y mejoras del proyecto ha sido la falta de datos desagregados en las fuentes oficiales, en particular en los ficheros publicados por el Instituto Nacional de Estadística (INE). Aunque la Encuesta de Ocupación Hotelera ofrece una gran cantidad de información agregada, no fue posible acceder a ciertos niveles de detalle que habrían enriquecido notablemente el análisis ya que no están disponibles de forma gratuita. Por ejemplo, hubiera sido muy valioso disponer de:

Datos desagregados por categoría del hotel (sus estrellas).

Datos por tipo de residente vinculados directamente a los establecimientos.

Datos sobre la procedencia geográfica de los viajeros (país o comunidad autónoma de origen), que permitirían una mejor segmentación del mercado turístico.

En cuanto al desarrollo técnico, la aplicación creada en Power BI, si bien ha resultado eficaz para representar y analizar visualmente los datos, también presenta limitaciones inherentes a la herramienta, como:

Restricciones en el manejo de grandes volúmenes de datos o cálculos complejos en tiempo real.

Dificultad para realizar análisis estadísticos avanzados directamente dentro de la plataforma.

El análisis cubre de mayo de 2022 a febrero de 2025, lo que excluye los meses de marzo y abril de 2025 y datos previos a la pandemia (2020-2021). Esta limitación puede introducir sesgos, especialmente en la evaluación de la estacionalidad, ya que no se dispone de un ciclo anual completo para 2025. Además, la exclusión de 2020-2021 impide analizar el impacto y recuperación post-COVID en profundidad.

ÍNDICE DE FIGURAS Y RESULTADOS

FIGURA 1 - CUESTIONARIO ENCUESTA OCUPACIÓN HOTELERA INE PÁG. 1 Y 4.....	10
FIGURA 2 - ENCUESTA DE OCUPACIÓN HOTELERA - INE BASE	15
FIGURA 3 - PC-AXIS - BASE DE DATOS TRANSFORMADA	17
FIGURA 4 - EXCEL PERNOCTACIONES BASE DE DATOS DEPURADA.....	19
FIGURA 5 - EXCEL ESTABLECIMIENTOS BASE DE DATOS DEPURADA	19
FIGURA 6 - TABLA DE APOYO MAPA.....	21
FIGURA 7 - TABLERO DE RELACIONES ENTRE TABLAS DE DATOS	22
FIGURA 8 - PORTADA - ÍNDICE	24
FIGURA 9 - OVERVIEW - RESUMEN.....	25
FIGURA 10 - RANKINGS TERRITORIALES	29
FIGURA 11 - OFERTA Y ESTRUCTURA DE MERCADO.....	30
FIGURA 12 - OCUPACIÓN	32
FIGURA 13 - DEMANDA TOTAL	33
FIGURA 14 - DEMANDA RESIDENCIA	34
FIGURA 15 - EMPLEO Y PRODUCTIVIDAD.....	35
FIGURA 16 - CLUSTER DEMANDA	38
FIGURA 17 - CLUSTER OFERTA	39
FIGURA 18 - SCREE PLOT: VARIANZA EXPLICADA	39
FIGURA 19 - IMPORTANCIA DE LAS COMPONENTES	40
FIGURA 20 - CARGAS PCA.....	40
FIGURA 21 - BIPLLOT PCA	41
FIGURA 22 - VALLADOLID DEMANDA POR RESIDENTE.....	44
FIGURA 23 - ISLAS CANARIAS EMPLEO Y PRODUCCIÓN.....	47

BIBLIOGRAFÍA

Cochran. (1977)). W.G. Sampling Techniques (3rd ed.). Wiley.

Europe, U. N. (n.d.). *Common metadata framework part C: Metadata concepts, standards and models*. Retrieved from <https://unece.org/statistics/publications/common-metadata-framework>

Exceltur. (2024). Perspectivas del sector turístico español.

Hastie, T. T. (n.d.). 2009. In *The Elements of Statistical Learning*. Springer.

Instituto Nacional de Estadística (INE). (s.f.). *Encuesta de Ocupación Hotelera: Metodología y Resultados*. Obtenido de https://www.ine.eshttps://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177015&menu=metodologia&idp=1254735576863

Instituto Nacional de Estadística. (n.d.). *Herramientas para el tratamiento de tablas estadísticas (PC-Axis)*. Retrieved from <https://www.ine.es/ss/Satellite?c=Page&pagename=MetodologiaYEstandares%2FINELayout&cid=1259941200926>

Jain, A. K. (2010). *ata clustering: 50 years beyond K-means*. *Pattern Recognition Letters*, 31(8), 651–666. Retrieved from <https://doi.org/10.1016/j.patrec.2009.09.011>

Jain, A. M. (1999). Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 264–323.

Kabacoff, R. (2015). *R in Action: Data Analysis and Graphics with R*. Manning.

Microsoft. (n.d.). *Microsoft Excel: Software de hojas de cálculo*. Retrieved from <https://www.microsoft.com/es-es/microsoft-365/excel>

Sweden, S. (n.d.). *PC-Axis family*. Retrieved from <https://www.scb.se/en/services/statistical-programs/pc-axis-family/>