

Article

Análisis de la relación entre la actividad turística y el empleo en el sector servicios en España

Iyán Álvarez, Azahara Martínez, Juan Alcaraz,

* Correspondence:

Abstract: El presente informe analiza la evolución del turismo en España y su relación con la ocupación en el sector servicios, a partir de datos oficiales del INE. Se integraron dos bases de datos —una sobre viajeros y pernoctaciones, y otra sobre empleo por sectores— para realizar un análisis exploratorio, univariante y bivariante a nivel de comunidad autónoma. Los resultados muestran una clara tendencia creciente y estacional en la actividad turística, así como una relación positiva y estadísticamente significativa entre el turismo y el empleo en la mayoría de las comunidades autónomas. Sin embargo, las elasticidades obtenidas son bajas, lo que indica que, aunque ambas variables evolucionan conjuntamente, el empleo responde de forma inelástica a las variaciones del turismo. En conjunto, los resultados sugieren que el turismo constituye un motor relevante del empleo en los servicios, aunque su efecto es moderado y depende de la estructura económica de cada territorio.

1. INTRODUCCIÓN

1.1. CONTEXTO GENERAL

El turismo constituye uno de los principales motores económicos de España, aportando una parte significativa al Producto Interior Bruto y al empleo nacional. La actividad turística no solo influye en la generación de riqueza, sino también en el dinamismo de otros sectores, como el transporte, la hostelería o el comercio.

En los últimos años, la evolución del turismo español ha experimentado variaciones notables, especialmente durante el periodo marcado por la pandemia de la COVID-19, que provocó una caída sin precedentes en los flujos de viajeros y en los niveles de ocupación del sector servicios. A partir de 2021, el proceso de recuperación ha sido desigual entre las comunidades autónomas, lo que hace necesario un análisis detallado que permita comprender las tendencias y relaciones entre las distintas variables implicadas.

En este contexto, el presente trabajo realiza un análisis exploratorio a partir de datos oficiales del **Instituto Nacional de Estadística (INE)**, con el objetivo de examinar la evolución del turismo y su relación con la ocupación en el sector servicios. Este enfoque permite no solo identificar patrones y correlaciones, sino también detectar posibles valores atípicos derivados de eventos excepcionales, como la crisis sanitaria, que pueden distorsionar la interpretación de los resultados.

1.2. DESCRIPCIÓN DE LOS DATOS

Para el desarrollo del presente análisis se han empleado dos bases de datos principales.

La primera recoge información sobre el turismo en España, desglosada por comunidades autónomas y provincias, e incluye variables como el número de viajeros y las pernoctaciones registradas a lo largo de varios años.

La segunda base de datos corresponde al sector servicios, concretamente a la ocupación laboral dentro del mismo, y se analiza con el propósito de explorar su relación con la evolución del turismo.

Citation: . Análisis de la relación entre la actividad turística y el empleo en el sector servicios en España. *Journal Not Specified* **2024**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1.3. CUESTIONES A TRATAR	38
En este estudio se abordan tres cuestiones principales:	39
1. ¿Se observan datos atípicos relacionados con la pandemia del COVID-19? De ser así, ¿Como se tratarían?	40 41
2. ¿Qué evolución ha experimentado el turismo en las distintas comunidades autónomas a lo largo del periodo analizado?	42 43
3. ¿Qué relación existe entre la evolución del turismo y la ocupación en el sector servicios?	44
2. IMPORTACIÓN	45
Para garantizar el uso de datos actualizados, la importación se realiza directamente desde las URLs oficiales del INE, evitando copias locales.	46 47
Antes de la carga definitiva, se comprobó la codificación de los archivos para evitar errores de lectura, optando por el formato UTF-8 .	48 49
Finalmente, los datos se importaron y almacenaron en dos data frames principales: uno para turismo y otro para ocupación.	50 51
3. ANÁLISIS UNIVARIANTE	52
Antes de realizar el análisis univariante, es necesario limpiar y preparar las bases de datos. A continuación, se procede al tratamiento individual de cada una: primero la base de turismo y posteriormente la base de ocupación.	53 54 55
3.1. TURISMO	56
3.1.1. Limpieza	57
Una vez importados los datos, se analizan las características generales del conjunto correspondiente al turismo en España. Este dataset recoge el número de viajeros y pernoctaciones por comunidad y provincia a lo largo del tiempo.	58 59 60
En un inicio, este dataset se forma de 134.820 filas y 8 columnas, cuyo nombre resulta completamente representativo.	61 62
Las variables incluidas son:	63
• Totales.Territoriales : columna con un único valor (“Total Nacional”).	64
• Comunidades.y.Ciudades.Autónomas : variable categórica que representa las comunidades o ciudades autónomas de España.	65 66
• Provincias : variable categórica que muestra la provincia asociada.	67
• Viajeros.y.pernoctaciones : variable categórica. Indica el tipo de registro (viajeros o pernoctaciones).	68 69
• Residencia..Nivel.1 : variable categórica redundante que contiene el valor “Total” en todas sus observaciones.	70 71
• Residencia..Nivel.2 : variable categórica que indica el origen de los viajeros, diferenciando entre residentes en España y residentes en el extranjero.	72 73
• Periodo : variable tipo fecha que representa el año y mes de la observación.	74
• Total : valor numérico de viajeros o pernoctaciones expresado en millones.	75
Las variables Totales.Territoriales, Viajeros.y.pernoctaciones, Residencia..Nivel.1 y Residencia..Nivel.2, no aportan información relevante y se eliminan en fases posteriores.	76 77
A continuación, se utilizó la función <code>glimpse()</code> del paquete <code>dplyr</code> para obtener una vista general del dataset, comprobando que todas las columnas se importaron como tipo carácter y que será necesario ajustar sus tipos en etapas posteriores.	78 79 80

Gracias a esto, se observa que todas las columnas se importaron inicialmente como tipo character. Por ello, en la fase de preparación se realizará la conversión a tipos de datos adecuados (fechas, factores y numéricos) para facilitar el análisis estadístico y gráfico.

Para asegurar la correcta manipulación numérica, se eliminaron los puntos utilizados como separadores de miles en la columna Total.

Posteriormente, se aplicó sapply() junto con unique() para revisar los valores de cada variable y detectar posibles inconsistencias.

##	Totales.Territoriales	Comunidades.y.Ciudades.Autónomas	88
##	1	20	89
##	Provincias	Viajeros.y.pernoctaciones	90
##	51	2	91
##	Residencia..Nivel.1	Residencia..Nivel.2	92
##	1	3	93
##	Periodo	Total	94
##	321	95210	95

Se identificaron varias **incidencias**: la columna Totales.Territoriales presentaba un único valor (“Total nacional”), y Comunidades.y.Ciudades.Autónomas incluía 20 categorías en lugar de 19, debido a registros vacíos o duplicados.

Además, se observó que la variable CCAA presentaba registros vacíos o en blanco, lo que explica las discrepancias detectadas anteriormente. Estos valores se eliminarán en la fase de limpieza del dataset.

El siguiente paso fue eliminar las filas y columnas redundantes (Totales.Territoriales y las filas que incluyen los espacios en blanco), así como los valores en blanco, manteniendo únicamente la información agregada por comunidades autónomas.

Posteriormente, con el fin de facilitar la interpretación y el análisis comparativo, se optó por trabajar con los datos agregados a nivel de comunidades y ciudades autónomas. Este enfoque proporciona una visión más general del comportamiento del turismo en España, adecuada para su posterior comparación con la ocupación en el sector servicios.

En consecuencia, se eliminaron las variables no relevantes para este nivel de análisis (Provincias, Viajeros.y.pernoctaciones, Residencia..Nivel.1 y Residencia..Nivel.2).

Dado que todas las variables se importaron inicialmente como texto, fue necesario convertir Periodo al formato de fecha. Para ello, se sustituyó el carácter “M” por un guion y se aplicó la función ym() para reconocer correctamente el año y el mes.

Además, la variable Total se convirtió a tipo numérico para permitir la realización de cálculos y análisis estadísticos sin errores de formato.

Antes de convertir la variable Comunidades.y.Ciudades.Autónomas a tipo factor, se limpian y estandarizaron sus valores eliminando números, espacios y caracteres innecesarios, haciendo uso de expresiones regulares.

Finalmente, se renombró como CCAA y se definió como variable categórica tipo factor.

Para continuar con el análisis, se creó un nuevo conjunto de datos agregado a nivel de comunidad autónoma.

Tras la limpieza, se comprobó la ausencia de espacios en blanco y valores faltantes (NA) para asegurar la integridad del conjunto de datos. Al no detectarse incidencias, el dataset quedó listo para el análisis y se guardó en formato RData.

3.1.2. Análisis

Como primer paso del análisis univariante, se calcularon los estadísticos descriptivos básicos con el fin de obtener una visión general de las variables del conjunto de datos, utilizando la función summary().

El resumen estadístico muestra que todas las comunidades autónomas tienen el mismo número de observaciones, con registros trimestrales entre 1999 y 2025. La variable Total presenta un mínimo de 0 (probablemente asociado a la pandemia), una media de 6,27

millones y un máximo de 54 millones, reflejando una notable variabilidad entre regiones y periodos.

Dado que la función `summary()` ofrece información limitada para variables categóricas, se amplió el análisis calculando los principales estadísticos descriptivos de Total por comunidad autónoma. Este enfoque permite comparar la magnitud y variabilidad del turismo entre regiones, antes de profundizar en las relaciones en el análisis bivariante.

CCAA	media	mediana	sd	min	max	IQR
Cataluña	20162927	18224322	11081536	0	46181964	17356596
Illes Balears	19915553	14101666	17348312	0	54032304	33005996
Canarias	19889672	20162212	6396172	0	32533664	9837890
Andalucía	19415478	18803052	8278829	0	39493055	12911490
Comunitat Valenciana	10438034	10005161	4122034	0	21030789	5604198
Comunidad de Madrid	9063390	9194200	2976974	0	15096330	4736004

Los resultados muestran que Cataluña presenta la media más alta en el número de turistas, lo que refleja su consolidación como uno de los principales destinos turísticos del país. Le siguen Baleares, Canarias y Andalucía, todas ellas comunidades con una fuerte especialización en el sector turístico.

En el extremo opuesto, Melilla, Ceuta, La Rioja y Navarra registran las medias más bajas, lo que podría deberse a su menor capacidad de atracción turística o a un peso económico menos centrado en este sector.

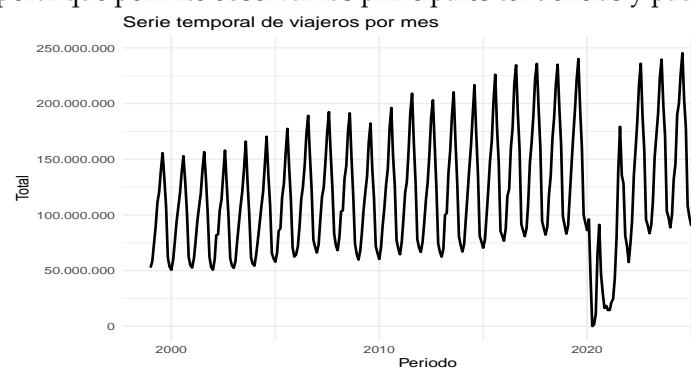
Para obtener una visión general del comportamiento del turismo en el conjunto del país, se calcularon los principales estadísticos descriptivos del Total Nacional, agregando los valores de todas las comunidades por periodo.

Estos, muestran una media de aproximadamente 119 millones de turistas por periodo y una mediana cercana a 112 millones, lo que indica una distribución relativamente equilibrada aunque con cierta asimetría hacia los valores altos.

La desviación típica, de unos 52 millones, junto con una varianza elevada, refleja una alta variabilidad en el volumen de turismo entre periodos.

El valor mínimo (0) corresponde al impacto de la pandemia de la COVID-19, mientras que el máximo (247,9 millones) evidencia los picos alcanzados en los años de mayor actividad turística.

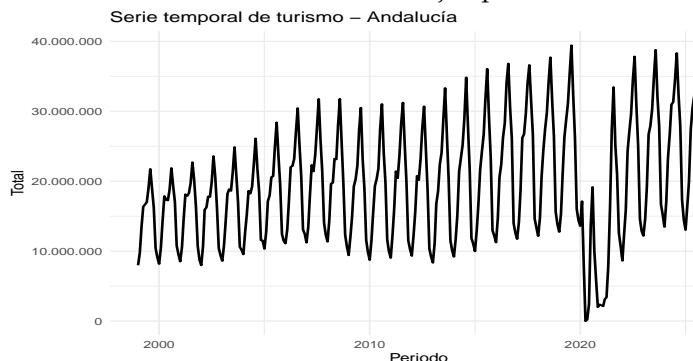
Para ilustrar la evolución del turismo en el conjunto nacional, se elaboró una serie temporal que permite observar las principales tendencias y patrones estacionales.



La gráfica muestra una tendencia general al alza con una clara estacionalidad anual, marcada por picos recurrentes en los meses de mayor actividad turística. También se identifican valores atípicos durante el periodo de la pandemia de la COVID-19, que reflejan la brusca caída del sector en esos años.

A continuación, se representa la serie temporal de cada comunidad y ciudad autónoma para analizar la evolución del turismo y comparar sus comportamientos a lo largo del tiempo.

Para evitar una visualización sobrecargada, los gráficos individuales de cada comunidad autónoma se han guardado automáticamente en la carpeta `graficos_comunidades`. En el informe se muestra únicamente un ejemplo ilustrativo.



En general, todas las comunidades muestran una marcada estacionalidad, con picos de actividad durante los meses de verano. A largo plazo, Madrid, el País Vasco y Canarias presentan una tendencia claramente creciente, mientras que otras, como Ceuta, mantienen una evolución más estable.

Asimismo, se observa heterocedasticidad en regiones como Illes Balears, Asturias, Andalucía o la Comunitat Valenciana, donde la variabilidad aumenta con el número de turistas, a diferencia de Madrid o Canarias, donde la dispersión se mantiene más constante.

Por último, todas las comunidades registran una caída abrupta en 2020, asociada al impacto de la pandemia de la COVID-19. En el siguiente apartado se aplicará el método 3sigma estacional (por mes y comunidad autónoma) para identificar de forma formal los valores atípicos sin confundirlos con la estacionalidad natural de la serie.

3.1.3. Detección de outliers

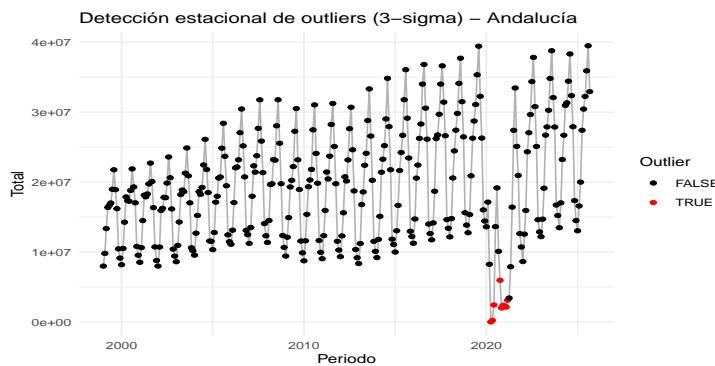
Para la detección de valores atípicos en las series de turismo se aplicaron tres métodos estadísticos complementarios:

- **Regla 3-sigmas:** considera como *outliers* los valores que se alejan más de tres desviaciones típicas respecto a la media.
- **Método de Hampel:** utiliza la mediana y la desviación absoluta mediana (MAD) para detectar observaciones atípicas, siendo más robusto frente a valores extremos.
- **Método del Boxplot:** identifica *outliers* a partir de los límites del rango intercuartílico ($1,5 \times \text{IQR}$).

En todos los casos se aplicó una versión estacional del método, calculando los parámetros (media, desviación, mediana o cuartiles) por mes y comunidad autónoma, con el objetivo de evitar clasificar como atípicos los picos o caídas debidos a la estacionalidad turística.

Los gráficos se generaron automáticamente para cada comunidad autónoma, aunque solo se muestra un ejemplo en el informe. Todos ellos se han guardado en carpetas independientes:

`graficos_outliers_3sigma`, `graficos_outliers_hampel` y `graficos_outliers_boxplot`.



Aplicando los tres métodos de detección de valores atípicos se obtiene el siguiente número de outliers identificados por cada método:

CCAA	n_hampel	n_boxplot	n_sigma3
Ceuta	20	21	8
Illes Balears	18	17	8
Castilla - La Mancha	14	16	11
Castilla y León	13	14	9
Comunitat Valenciana	12	13	10
La Rioja	12	15	10

Durante la aplicación de los distintos métodos de detección de valores atípicos (3-sigma, Hampel y Boxplot) se observó que algunas comunidades autónomas no presentaban outliers, a pesar de mostrar comportamientos anómalos en sus series temporales.

El análisis de los estadísticos descriptivos (media, varianza y desviación típica) evidenció que en regiones como la Comunidad de Madrid la variabilidad dentro del año es reducida, pero la variabilidad entre estaciones es alta debido al crecimiento sostenido del turismo en los últimos años. Esto provoca que métodos como el de 3-sigma no identifiquen correctamente algunos valores que visualmente se perciben como atípicos.

Dado que los métodos tradicionales no captan adecuadamente todas las anomalías y considerando el contexto económico y social, se decidió tratar como outliers comunes el periodo comprendido entre marzo de 2020 y febrero de 2021, correspondiente al impacto más severo de la pandemia de la COVID-19. De este modo, se homogeneiza el tratamiento de valores atípicos en todas las comunidades, evitando sesgos derivados de diferencias estacionales.

3.1.4. Series temporales

El periodo de la pandemia constituye claramente una anomalía dentro de la serie temporal turística. A partir de marzo de 2020 se observa un descenso abrupto en el número de viajeros y pernoctaciones, directamente asociado al confinamiento y las restricciones de movilidad.

Tras ese desplome, las cifras se mantienen bajas durante varios trimestres y comienzan a recuperarse gradualmente hacia 2022, alcanzando niveles similares a los previos a la pandemia. Esto sugiere que el COVID-19 interrumpió temporalmente la tendencia del sector, sin modificar de forma estructural su evolución a largo plazo.

Por tanto, se considera que los datos correspondientes a dicho periodo no aportan información representativa del comportamiento habitual del turismo.

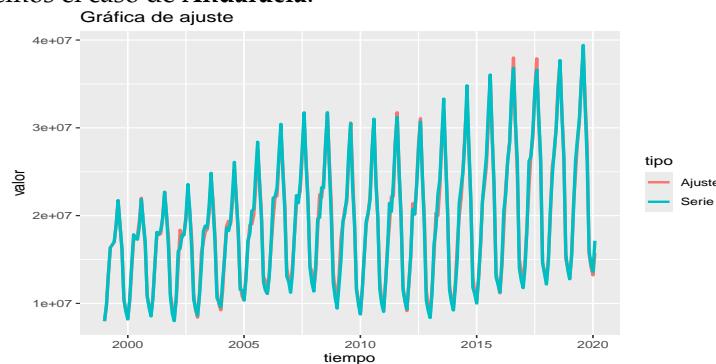
Con esta justificación, se plantea ajustar modelos ARIMA para el análisis de tendencias a largo plazo, utilizando únicamente los datos previos a la pandemia.

El modelo se define como: $ARIMA(p, d, q)(P, D, Q)$, donde los parámetros se ajustan automáticamente mediante la función `auto.arima()`.

Posteriormente, se utiliza la función `forecast()` para proyectar los valores de los 12 meses siguientes a partir del modelo estimado.

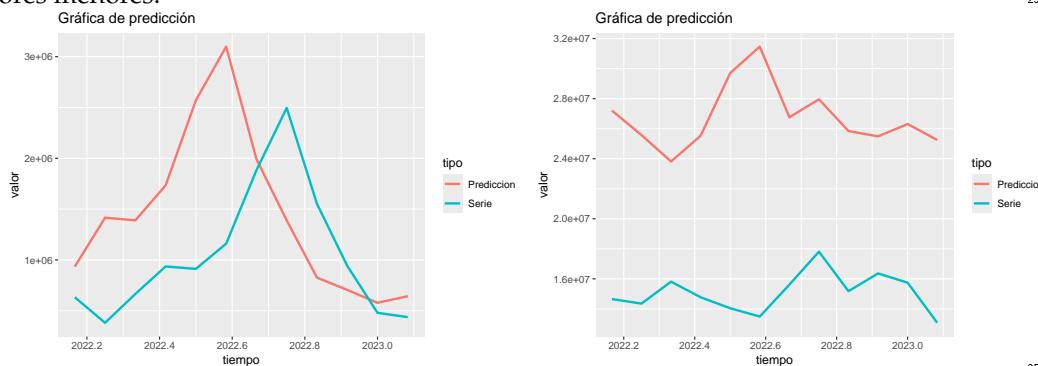
Los resultados para cada comunidad se guardaron en un archivo llamado `AnalisisBox.rsd`. Se trata de una lista, si queremos acceder a ellos se pueden importar con el comando `readRDS("AnálisisBox.rsd")` y asignándolo a una variable.

Se pueden ver las diferentes gráficas de juste para cada comunidad, como son muchos gráficos se muestra solo uno, pero se puede ver que el resto muestran resultados similares. Ponemos el caso de **Andalucía**.



Entonces, tras ver que el modelo aproxima de forma razonable los datos prepandemia, hay razones para confiar en las predicciones, las cuales se pueden interpretar como los datos de pernoctaciones que hubieran sucedido en esos meses siguientes a febrero de 2020 si la pandemia no hubiera azotado.

Lo más interesante ocurre si comparamos esas predicciones con los datos posteriores a la pandemia. Si bien es cierto que hay algunas comunidades para las cuales las predicciones no se ajustan en nada (son los casos de las ciudades autónomas) en el resto de casos se puede ver como la predicción y el valor postpandemia son iguales solo que con cifras algo menores menores.



3.1.5. Imputación de datos

Mediante el modelo ARIMA previamente entrenado, se imputan los valores del período comprendido entre marzo de 2020 y febrero de 2021, sustituyendo los datos originales afectados por la pandemia por las predicciones del modelo. Así se obtiene una serie temporal más coherente y continua, almacenada en el nuevo data frame `df_Turismo_corregido`.

3.2. OCUPACIÓN

El segundo conjunto de datos corresponde a la ocupación por sectores económicos en España, obtenido del INE. Contiene el número de personas ocupadas (en miles) por comunidad o ciudad autónoma, sector y periodo trimestral.

El dataset está compuesto por 21.300 observaciones (filas) y 5 variables (columnas), cuyos nombres son claros y representativos, lo que facilita su comprensión y manipulación posterior.

Las variables incluidas son:

- **Sexo.** Variable categórica que presenta el valor “mujer”, “hombre” y “ambos sexos”.

- **Comunidades.y.Ciudades.Autónomas.** Variable categórica que incluye las comunidades Españolas y el total nacional. 266
267
- **Sector.económico.** Variable categórica que clasifica la actividad económica en agricultura, industria, construcción o servicios. 268
269
- **Periodo.** Variable tipo fecha que muestra el año y trimestre. 270
- **Total.** Variable numérica que muestra las personas ocupadas (en miles). 271

Como primer paso, se utilizó la función `glimpse()` para revisar la estructura del dataset. Se comprobó que todas las columnas se importaron como tipo carácter, por lo que será necesario ajustar los tipos de datos en los siguientes pasos. 272
273
274

3.2.1. Limpieza

Para limpiar y preparar la base de datos de ocupación, ha sido necesario realizar los siguientes pasos: 275
276
277

1. Eliminar la columna “Sexo”. 278
2. Eliminar números y caracteres especiales de los nombres de las comunidades. Además de filtrar y eliminar el valor “Total nacional” de la variable CCAA, por ser redundante respecto al resto de comunidades y ciudades autónomas. Posteriormente, esta variable se convierte a tipo factor. 279
280
281
282
3. Filtrar los registros correspondientes al sector “Servicios” dentro de la columna Sector.económico. Dado que la tabla se centrará exclusivamente en este sector, la variable se elimina por resultar redundante. 283
284
285
4. Convertir la variable “Periodo” del formato YYYYTQ a un formato de fecha trimestral mediante la función `as.yearqtr()`. 286
287
5. Normalizar la variable “Total”, eliminando la palabra “miles” y los puntos de millares, sustituyendo la coma decimal por punto, convirtiendo los valores a numéricos y multiplicándolos por 1.000 para expresarlos en personas. De este modo, se mantiene la coherencia de escala con la base de datos de turismo. 288
289
290
291

3.2.2. Análisis

A continuación, se representa la **evolución trimestral del empleo nacional en el sector servicios**. Para ello, se agrupan los datos por periodo y se calcula el total de personas ocupadas, generando una serie temporal mediante un gráfico de líneas. 292
293
294
295



Se observa una caída entre 2010 y 2014, posiblemente asociada a los efectos de la crisis de 2008, seguida de una recuperación sostenida hasta 2019. 296
297

En 2020 se produce un descenso notable debido al impacto de la pandemia, tras lo cual la ocupación retoma una tendencia creciente, alcanzando niveles superiores a los anteriores. 298
299
300

Además, se aprecia una estacionalidad clara, con aumentos recurrentes en determinados trimestres, probablemente vinculados a la temporada turística. 301
302

A continuación, se analiza la evolución trimestral del empleo en el sector servicios por Comunidad Autónoma, agrupando los datos por región y periodo. Este análisis permite observar las diferencias en la dinámica laboral entre territorios a lo largo del tiempo.

Los gráficos individuales de cada comunidad autónoma se han generado y almacenado automáticamente en una carpeta independiente (*graficos_ocupacion_ccaa*). A modo de ejemplo, a continuación se muestra la evolución correspondiente a Andalucía.



Se observa una tendencia al alza más pronunciada en comunidades como Andalucía, Canarias, Madrid y Cataluña, principales motores del sector servicios. Destaca una fuerte estacionalidad en Illes Balears, asociada a la actividad turística, mientras que en la mayoría de regiones la evolución es más estable y con un crecimiento moderado tras la pandemia.

3.2.3. Detección de outliers

Aplicando los tres métodos de detección de valores atípicos, al igual que en la base de datos del turismo, se obtiene el siguiente número de outliers identificados por cada método:

CCAA	n_hampel	n_boxplot	n_sigma3
Extremadura	6	8	0
Principado de Asturias	5	7	0
Castilla - La Mancha	3	0	0
La Rioja	3	1	0
Comunidad de Madrid	2	1	0
Andalucía	1	3	0

El número de valores atípicos detectados en la serie de ocupación es reducido y no presenta un patrón sistemático entre comunidades autónomas. En la mayoría de los casos, estas observaciones reflejan variaciones reales del empleo, asociadas al crecimiento estructural del mercado laboral o a episodios económicos concretos, como la crisis de 2008.

A diferencia de lo ocurrido en la base de datos turística —donde los outliers correspondientes al periodo de la pandemia de la COVID-19 se eliminaron por representar una disruptión excepcional—, en este caso se considera que las fluctuaciones detectadas forman parte del comportamiento normal de la serie.

Por tanto, no se eliminarán las observaciones atípicas y se mantendrán en el análisis, tomando como referencia el método de detección 3-sigma, que ofrece una estimación más estable y coherente con la naturaleza de estos datos.

4. ANÁLISIS BIVARIANTE

En esta sección se abordará el análisis bivariante, con el objetivo de relacionar las bases de datos de turismo y ocupación para estudiar posibles vínculos entre ambas variables.

Antes de combinar los conjuntos de datos, será necesario estandarizar ciertas variables con el fin de garantizar una fusión correcta (merge) y obtener un marco de datos coherente para el análisis conjunto.

Así pues, se convierte la variable Periodo al formato año-trimestre, luego se agrupan los datos por comunidad y periodo, y se calcula el total de turistas por trimestre y CCAA. Finalmente, se ordenan los registros por comunidad y periodo para facilitar el merge posterior.

Mediante la función `merge()`, se unen ambos conjuntos de datos utilizando como claves comunes las variables CCAA y Periodo.

Antes de la unión, en la base de ocupación se renombra la variable Total como Numero_ocupados para distinguirla del total de turistas y evitar ambigüedades en el nuevo dataset combinado (`df_Turismo_Ocupacion`).

Finalmente, una vez preparadas y combinadas las bases de datos, se guardan los objetos resultantes en formato `.RData`.

Esto permite conservar los datos limpios y listos para su análisis, facilitando su carga posterior sin necesidad de repetir el proceso de limpieza y fusión.

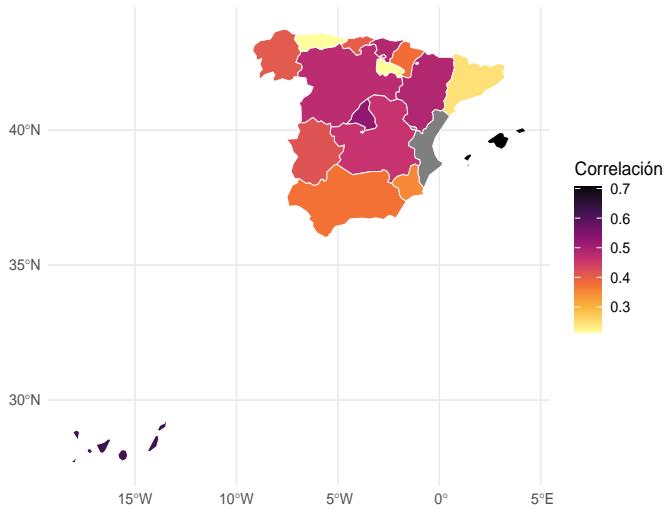
Correlaciones de Pearson

Como primer paso, se calcula la correlación de Pearson entre el número de turistas y el empleo en el sector servicios para cada Comunidad Autónoma.

El objetivo es medir la intensidad y dirección de la relación lineal entre ambas variables, obteniendo tanto el coeficiente de correlación como su p-valor para evaluar la significatividad estadística.

Además, se incorpora un indicador de nivel de significación ($p < 0.05$), que permite identificar en qué comunidades la relación es estadísticamente relevante. Los resultados se ordenan de mayor a menor correlación, destacando aquellas regiones donde el vínculo entre turismo y empleo es más fuerte y significativo.

Coefficiente de correlación por CCAA



Se observa que Illes Balears presenta la mayor correlación positiva ($r = 0.705$) entre turismo y empleo en el sector servicios, reflejando su fuerte dependencia del turismo estival. En el extremo opuesto, Melilla muestra una correlación muy baja ($r = 0.083$), probablemente por una menor relación directa entre la llegada de turistas y el empleo local.

En general, las correlaciones son positivas pero de magnitud moderada, indicando que ambas variables tienden a evolucionar conjuntamente, aunque con distinta intensidad según la región.

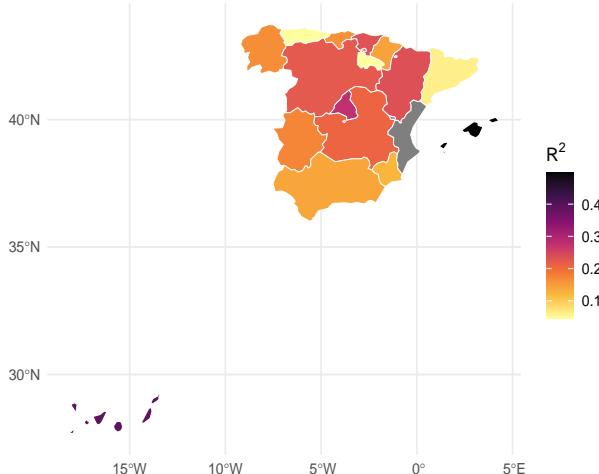
En la mayoría de comunidades, la relación es estadísticamente significativa ($p < 0.05$), mientras que en Ceuta, La Rioja, Cataluña y Asturias no se observa significatividad, posiblemente por el menor peso del turismo en su economía o por una mayor volatilidad en los datos.

Por último, cabe señalar que la correlación no implica causalidad: los resultados reflejan asociación, pero no necesariamente una relación directa de causa y efecto.

Análisis de regresión lineal simple

Además, modelamos para cada comunidad una regresión simple con el objetivo de conocer la **bondad de ajuste (R^2)**, lo cual permitea valorar qué proporción de la variabilidad en el empleo es explicada por el turismo.

Bondad de ajuste por CCAA
Coeficiente de determinación (R^2)



En general, se observa que los valores son moderados o bajos, lo que indica que, aunque existe cierta relación entre ambas variables, el turismo no explica completamente la evolución del empleo, y probablemente influyen otros factores económicos o estructurales.

Las comunidades con mayor ajuste del modelo son Islas Baleares ($R^2 = 0.498$), Canarias ($R^2 = 0.388$), Madrid ($R^2 = 0.276$) y Aragón ($R^2 = 0.234$), donde donde el comportamiento del empleo en el sector servicios parece estar más vinculado a la actividad turística, mostrando una relación más clara entre ambas variables en comparación con el resto de regiones.

Por el contrario, en Ceuta y Melilla ($R^2 \approx 0$) y Asturias ($R \approx 0.045$) el modelo lineal apenas explica la variación en el empleo, lo que sugiere una relación débil o inexistente entre ambas variables en estas regiones.

A partir del modelo de regresión lineal calculado anteriormente, se estima la **elasticidad** del empleo respecto al turismo, utilizando la pendiente de la recta de regresión como medida del impacto del turismo sobre el empleo en cada comunidad autónoma.

CCAA	elasticidad	int
Canarias	0.3962694	Inelástico
Castilla - La Mancha	0.2110888	Inelástico
Comunidad de Madrid	0.2068096	Inelástico
Extremadura	0.1564707	Inelástico
Región de Murcia	0.1302117	Inelástico
Illes Balears	0.1172972	Inelástico

Los valores de elasticidad, comprendidos entre 0.01891369 y 0.3962694 , muestran que en todas las comunidades autónomas la relación entre turismo y empleo es inelástica (ya que <1), es decir, el empleo en el sector servicios apenas varía ante los cambios en la actividad turística. Esto sugiere que un aumento del turismo no se traduce proporcionalmente en un mayor nivel de empleo.

Esta baja respuesta puede deberse a que parte del empleo es estructural o fijo, a la existencia de desfases temporales entre el incremento del turismo y la contratación de personal, o a que las variables se miden en frecuencias distintas (mensual para el turismo y trimestral para el empleo), lo que suaviza la relación observada.

En regiones como Canarias (≈ 0.396), donde la elasticidad es algo mayor, el empleo muestra una mayor sensibilidad a las variaciones del turismo, coherente con su elevada dependencia económica de esta actividad.

Aunque las correlaciones y la bondad de ajuste indicaban una relación fuerte y significativa entre turismo y empleo, las elasticidades calculadas son bajas. Esto no representa una contradicción, sino una diferencia en la interpretación: mientras la correlación mide la fuerza de la relación, la elasticidad cuantifica su intensidad relativa.

En otras palabras, aunque el turismo y el empleo evolucionan de forma coordinada, el impacto porcentual del turismo sobre el empleo es reducido, lo que sugiere que otros factores —como la estacionalidad o la estabilidad estructural del mercado laboral— también influyen en la evolución del empleo.

5. CONCLUSIONES

El análisis realizado pone de manifiesto la fuerte dependencia entre el turismo y la actividad económica en España, especialmente en el ámbito del empleo en el sector servicios. A lo largo del periodo analizado, las series temporales muestran un comportamiento claramente estacional, con picos en los meses de verano y una tendencia general de crecimiento sostenido, interrumpida únicamente por el impacto extraordinario de la pandemia de la COVID-19.

Durante dicho periodo, el turismo experimentó una caída abrupta sin precedentes, que fue tratada como un episodio atípico común en todas las regiones. Esta decisión permitió preservar la coherencia de las series y centrarse en el análisis estructural del fenómeno turístico. Los métodos de detección de outliers aplicados (3-sigma, Hampel y Boxplot) confirmaron que la mayoría de las anomalías se concentran en los años 2020 y 2021, reflejando los efectos temporales de la crisis sanitaria y las restricciones a la movilidad.

En conjunto, los resultados evidencian que el turismo español ha demostrado una notable capacidad de resiliencia, recuperando rápidamente sus niveles de actividad tras la pandemia. Además, se constata una relación positiva y significativa entre la evolución del turismo y la ocupación en los servicios, lo que refuerza la idea de que el sector turístico sigue siendo uno de los motores esenciales del crecimiento económico y del empleo en España.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.