

## CONTEXTO

Nos situamos a principios del 2025. El hotel ficticio VelaMaris\*\*\*\* está a punto de abrir sus puertas por primera vez en la incipiente temporada de la Costa del Sol, donde la competencia es feroz. La directiva quiere que la inauguración sea un éxito y, ante tal joven proyecto, está destinando buena parte de sus recursos en una campaña de marketing a nivel europeo. Pero antes de lanzarse al vacío, ha decidido encomendar a su departamento de Data Analysis un estudio que será la clave del éxito de la campaña y, por ende, del recién estrenado hotel.

## OBJETIVO

¿En qué país o zonas geográficas deberíamos poner el foco de dicha campaña a fin de impulsar la ocupación? Nos interesa toda la temporada, que abarca los meses desde abril hasta octubre, ambos incluidos; pero sobre todo los meses previos y posteriores al ajetreado verano, es decir, abril, mayo, septiembre y octubre.

## HIPÓTESIS INICIAL

Por mi experiencia en el sector, los turistas del Reino Unido son los que más visitan la Costa a lo largo de toda la temporada. Otros países como Alemania también destacan, aunque solo en los meses de pre y post temporada debido en parte a que el clima es más moderado. Comprobaremos si esto es cierto.

## FUENTE DE ESTUDIO

Analizaremos el registro oficial de turistas que visitaron la provincia de Málaga en la temporada 2024 (de abril a octubre), mismo período de apertura que en la actualidad.

No apoyaremos principalmente en el endpoint “*Procedencia de los turistas extranjeros que visitan España a nivel de provincia*” de la API de DATAESTUR (perteneciente al Ministerio de Industria y Turismo del Gobierno de España).

Para la presentación de los datos, emplearemos cifras de registros turísticos obtenidos de la API del INE (Instituto Nacional de Estadística); en concreto, del endpoint “*Viajeros y pernoctaciones por comunidades autónomas y provincias*”.

El acceso a ambas API permite la descarga de los ficheros de una manera amigable, tras la selección y ejecución de los códigos en la propia web. Es por esto que no ha sido necesario recurrir al módulo request en Pyhton.

Referencias y pantallazos a continuación:

API DATAESTUR

www.dataestur.es/apidata

VIAJES Y OCIOECONOMÍATRANSPORTEALOJAMIENTOSOSTENIBILIDADCONOCIMIENTO

GET /TURISMO\_RECEPTOR\_PROV\_PAIS\_DL Procedencia de los turistas extranjeros que visitan España a nivel de Provincia

Para el filtrado de fechas:

- Si pone sólo año de inicio, entonces se filtra ese año completo hasta la última fecha en la base de datos.
- Si pone sólo año de inicio y mes de inicio, se filtra desde ese año y mes de inicio hasta la última fecha en la base de datos.
- Si pone año de fin, se tiene que poner mes de fin y viceversa. Además se debe poner año de inicio y mes de inicio.

Parameters

Cancel

Name	Description
desde (año) integer (query)	Dato disponible desde 2019-07
desde (mes) integer (query)	Dato disponible desde 2019-07
hasta (año) integer (query)	Límite superior para el año.
hasta (mes) integer (query)	Límite superior para el mes.
Provincia string (query)	La provincia.

Execute

Clear

Responses					
Curl	<pre>curl -X 'GET' \ 'https://datastur.azure-api.net/API-SEGITTUR-v1/TURISMO_RECEPTOR_PROV_PAIS_DL?desde%20%28aXC3%81oX29~2024&amp;desde%20%28mesX29~01&amp;hasta%20%28aXC3%81oX29~2024&amp;hasta%20%28mesX29~12&amp;Provincia=M%C3%A1laga' \ -H 'accept: application/octet-stream'</pre>				
Request URL	<pre>https://datastur.azure-api.net/API-SEGITTUR-v1/TURISMO_RECEPTOR_PROV_PAIS_DL?desde%20%28aXC3%81oX29~2024&amp;desde%20%28mesX29~01&amp;hasta%20%28aXC3%81oX29~2024&amp;hasta%20%28mesX29~12&amp;Provincia=M%C3%A1laga</pre>				
Server response	<table><thead><tr><th>Code</th><th>Details</th></tr></thead><tbody><tr><td>200</td><td><div>Response body</div><div>Download file</div><div>Response headers</div></td></tr></tbody></table>	Code	Details	200	<div>Response body</div> <div>Download file</div> <div>Response headers</div>
Code	Details				
200	<div>Response body</div> <div>Download file</div> <div>Response headers</div>				

**API INE:**

Registros de Málaga, Barcelona, Madrid y Canarias.

## Encuesta de ocupación hotelera

### Establecimientos hoteleros

### Viajeros y pernoctaciones por comunidades autónomas y provincias

Unidades: Viajeros, Pernoctaciones

**Selección valores para consultar**

**Comunidades Autónomas y Provincias**


- Total Nacional
- 01 Andalucía
- 04 Almería
- 11 Cádiz
- 14 Córdoba
- 18 Granada

Seleccionados: 70 Total: 70

**Viajeros y pernoctaciones**


- Viajero
- Pernoctaciones

Seleccionados: 1 Total: 2

**Residencia**


- Total
- Residentes en España
- Residentes en el Extranjero

Seleccionados: 1 Total: 3

**Periodo**


- 2025M03
- 2025M02
- 2025M01
- 2024M12
- 2024M11
- 2024M10

Seleccionados: 1 Total: 315

**Elija forma de presentación de la tabla**

Viajeros y pernoctaciones
Residencia
Periodo
Comunidades Autónomas y Provincias

Viajeros y pernoctaciones	Residencia	Periodo	Comunidades Autónomas y Provincias
-	-	-	-
-	-	-	-
-	-	-	-

Decimales a mostrar: Por defecto

Notas ( 3 )

Total: 70 series y 70 datos

Consultar selección Consultar todo

# ANÁLISIS DE DATOS

El proyecto contiene Jupyter notebooks con el código desarrollado para el análisis y presentaciones. Además, incluye scripts de Python con las funciones empleadas para las visualizaciones, así como el main con el muestrario de datos básicos del estudio.

Veamos una breve descripción del dataset utilizado antes de adentrarnos en materia de análisis:

**DATASET:** Análisis del registro de turistas europeos en la provincia de Málaga durante los meses de temporada media-alta de 2024 desglosados en función de los países de procedencia de los viajeros.

**DESCRIPCIÓN:** Tras desglosar el dataset en tres temporadas (previa: abril-mayo, alta: junio-julio-agosto; post: septiembre-octubre), nos hemos quedado con los 15 registros de turistas más altos de cada una de ellas.

**PERIODO:** Los datos comprenden los meses abril a octubre de 2024 (ambos inclusive).

**ÍNDICE:** No tenemos columna índice, ninguna variable tiene una cardinalidad que permita establecerla como tal.

**OBJETIVO DEL ANÁLISIS:** Trabajamos para el departamento de análisis de datos del hotel VelaMaris\*\*\*\* en la Costa del Sol, que está a punto de abrir sus puertas esta temporada 2025. La directiva nos ha pedido que estudiemos los registros de turistas en 2024 para decidir en qué países o zonas geográficas deberíamos poner el foco de la campaña de marketing, con el objetivo de incrementar los porcentajes de ocupación y reservas esperados.

**IMPORTANICA DE LAS VARIABLES:** Nos fijaremos sobre todo en el número de turistas y las pernотaciones, seguidos por el país de origen y la temporada. Éste es el criterio de clasificación:

**\*\*target/directora (0), agrupación importante (1), agrupación interesante (2), agrupación secundaria (3)\*\***

Columna/Variable	Descripción	Tipo_de_Variable	Importancia inicial	Nota
MES	Mes del registro de las pernотaciones	Catégorica	2	Similar a TEMPORADA, por lo que MES apenas la analizaremos
PAIS_ORIGEN	País del que proceden los viajeros	Catégorica	1	Inicialmente contiene muchos registros distintos; tras el tratamiento de datos, nos centraremos en los más frecuentes para tratarla como una catégorica
TURISTAS	Registro total de turistas / viajeros	Númerica continua	0	Posible directora, nos interesa captar el mayor núm posible de turistas
PERNOCTACIONES	Registro total de pernотaciones (noches de estancia en hoteles)	Númerica continua	1	Similar a turistas, a priori van de la mano. Nos centraremos en el análisis de TURISTAS
ESTANCIA_MEDIA	Media de pernотaciones por turista	Númerica continua	2	Es de tendencia central, por lo que no refleja la totalidad de los datos: num pernотaciones correspondiente a cada turista (distinción que por cierto no tenemos)
TEMPORADA	Alta, pre o post temporada en función del mes del año	Catégorica	2	Dividiremos el análisis por temporadas para ver si cambian los resultados

# ANÁLISIS UNIVARIANTE

Registros de la provincia de Málaga extraídos de nuestro dataset tras la limpieza y tratamiento de datos. Solo incluye a turistas europeos durante los meses de temporada en 2024 (de abril a octubre)

Total de TURISTAS: 3.208.322

Total de PERNOCTACIONES: 23.511.240

Duración ESTANCIA MEDIA en 2024: 7.33 noches

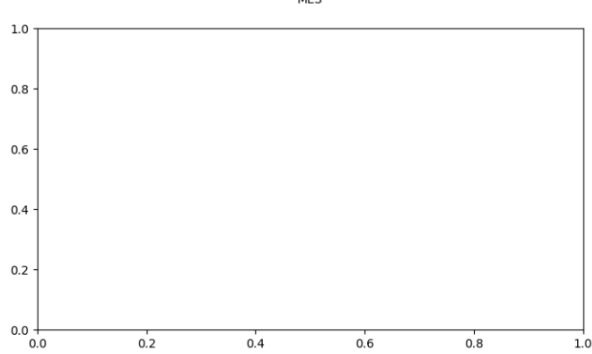
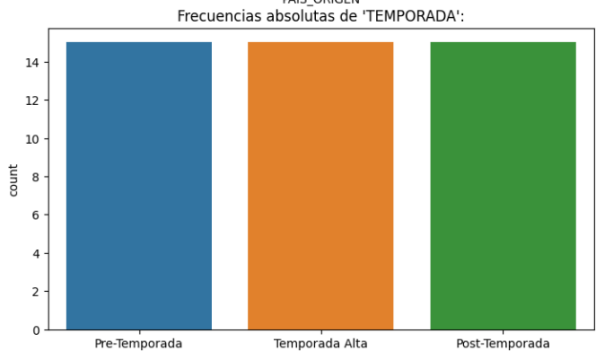
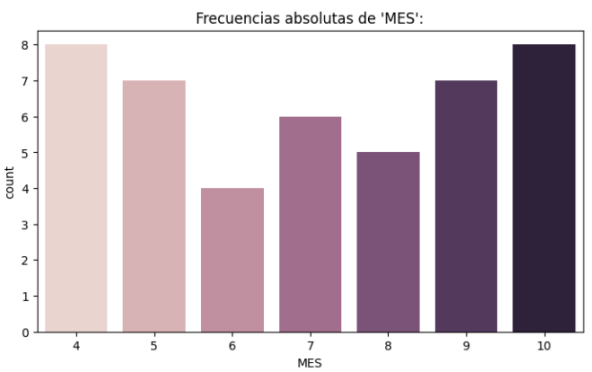
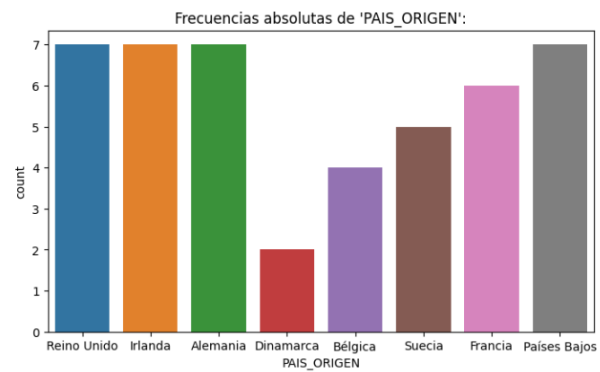
## Tendencia central

MODA PAÍSES DE ORIGEN: Irlanda, Países Bajos, Alemania y Reino Unido

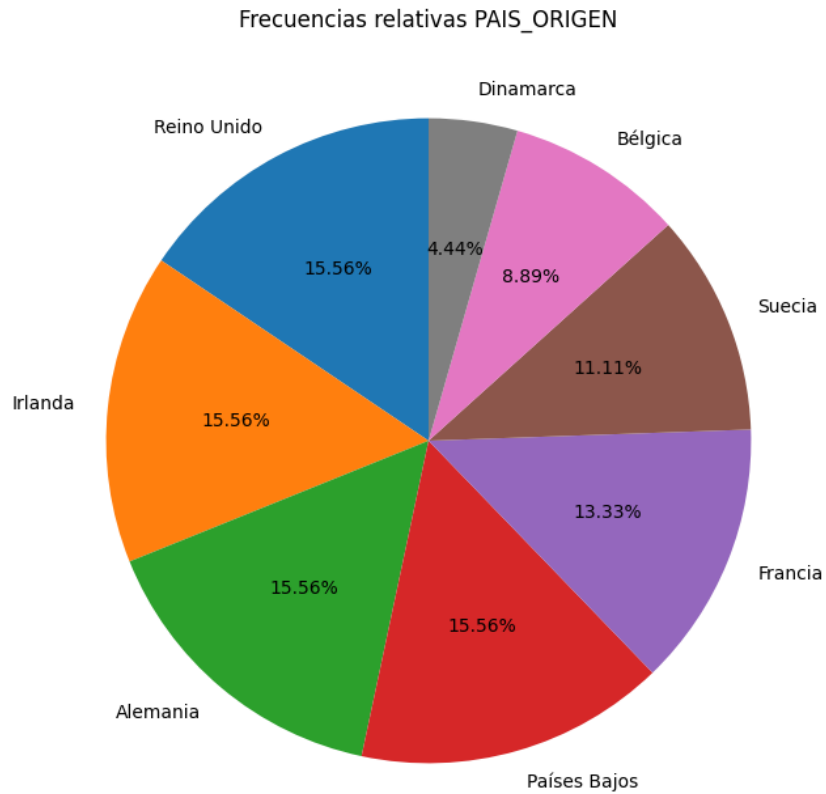
MEDIA, MEDIANA, VALOR MAX: Observamos una gran diferencia entre la media y la mediana, lo que indica que esperamos una gran dispersión en los datos.

MEDIA	MEDIANA	MAX
TURISTAS 66.567,19	TURISTAS 48.578	TURISTAS 228.875
PERNOCTACIONES: 524.915,81	PERNOCTACIONES 349.762	PERNOCTACIONES 1.938.169.
ESTANCIA_MEDIA 11.16	ESTANCIA_MEDIA 7.2	ESTANCIA_MEDIA 45.0

## Frecuencias absolutas categóricas



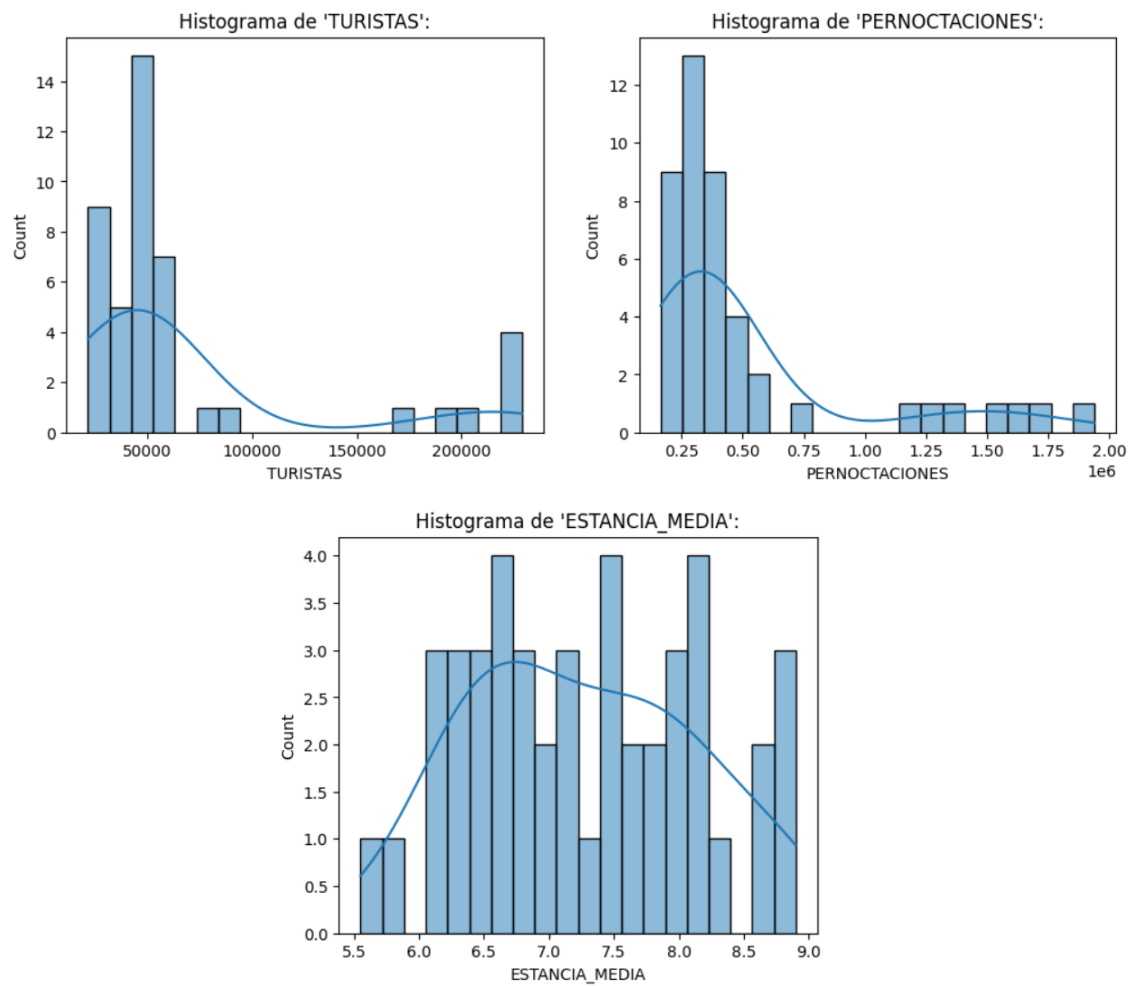
## Frecuencias relativas



Tras el análisis de las frecuencias, obtenemos los mismos datos que observábamos con la moda, los turistas que más nos visitan provienen de Alemania, Irlanda, Países Bajos y Reino Unido.

Los datos están distribuidos equitativamente por temporadas, pero no por meses. Recordemos que de cada temporada hemos escogido los 15 mayores registros numéricos de turistas.

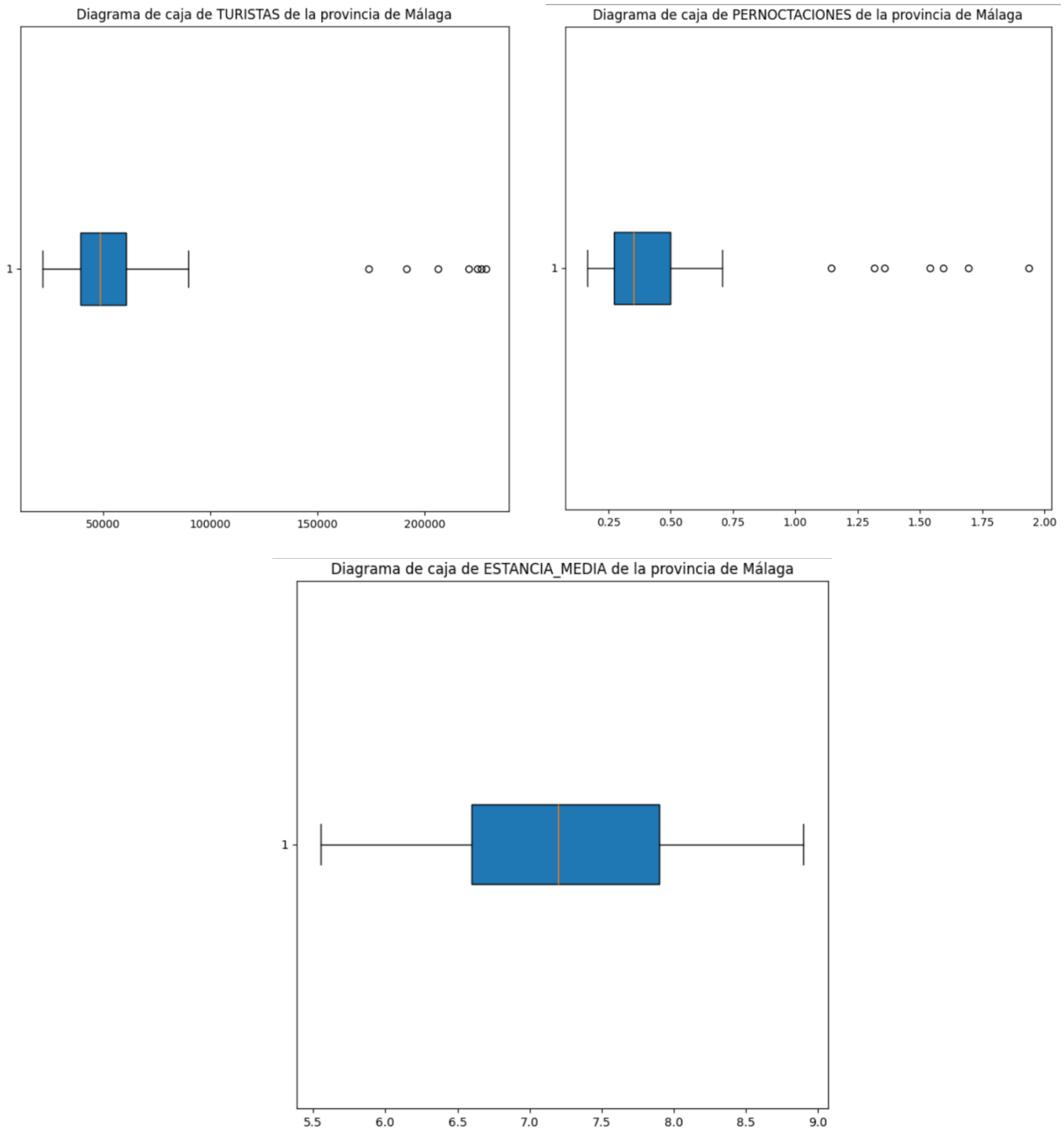
## Histogramas



Las curvas de las variables TURISTAS y PERNOCTACIONES son prácticamente iguales. La mayor parte de los registros se concentran en los valores correspondientes al tercio más bajo de la gráfica.

La curva de ESTANCIA\_MEDIA es similar, aunque, sin ser perfecta, recuerda más a una forma Gaussiana. Los datos son más consistentes y se concentran en un mismo bloque.

## Boxplot o Diagrama de cajas

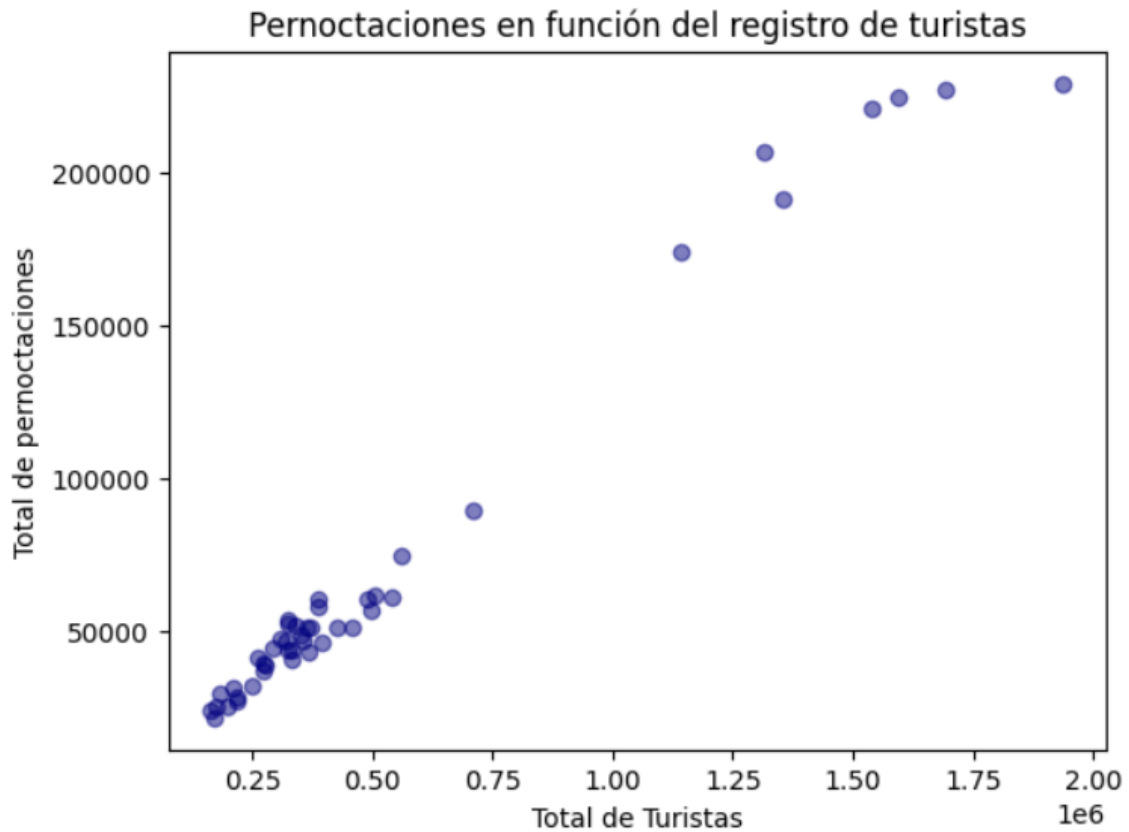


Vemos que los registros se concentran alrededor de los 50.000 (por encima y por debajo). Encontramos outliers muy alejados del umbral max, que rozan los 200.000. La distribución de datos es prácticamente pareja a la de turistas, salvo que los outliers no están tan concentrados. Todos los datos de estancia media se encuentran en el mismo bloque de 6.5 a 8 días.



## ANÁLISIS BIVARIANTE

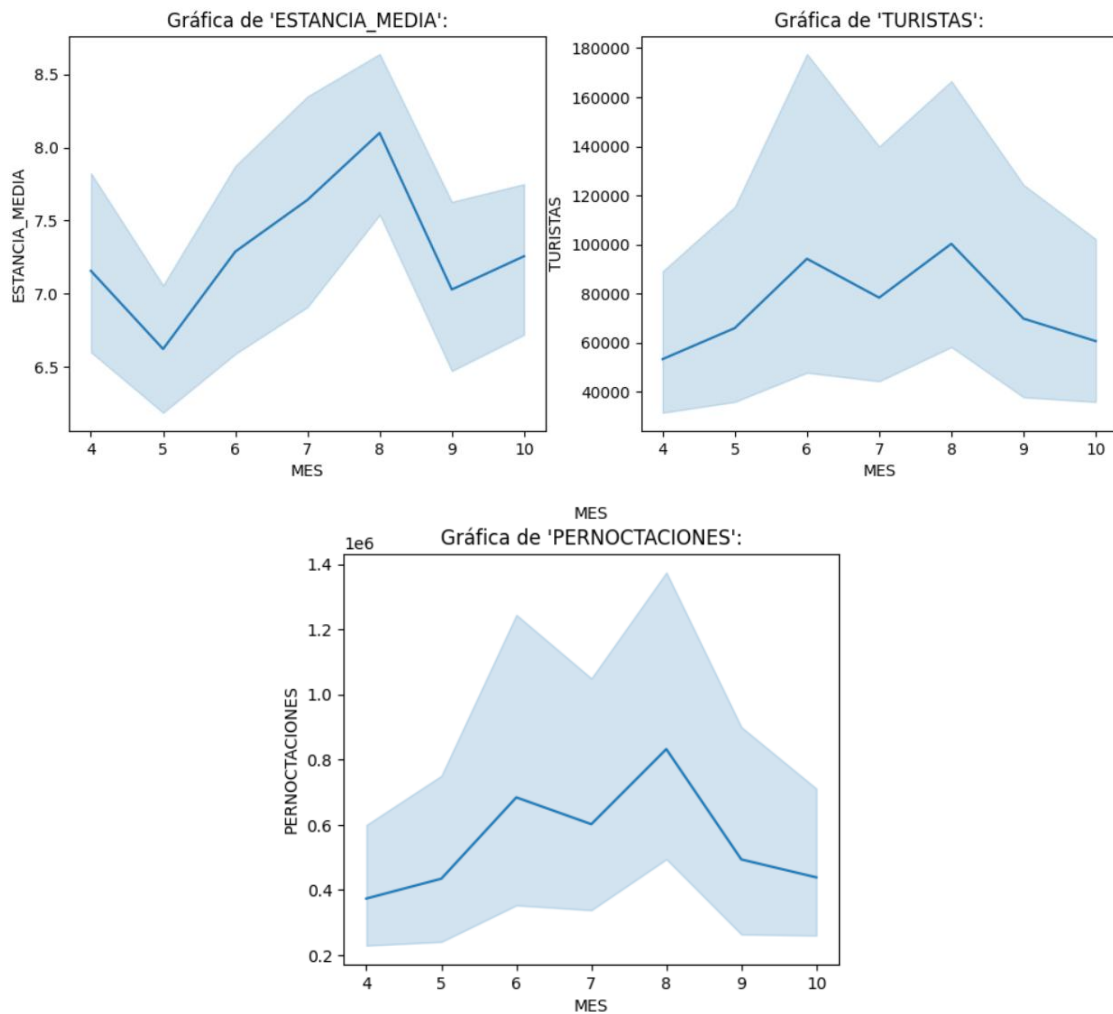
### Scatter plot TURISTAS vs PERNOCTACIONES



Visualmente ya se puede apreciar una fuerte correlación positiva entre ambas, lo cual es lógico, a mayor número de turistas mayor número de total de pernoctaciones.

Tras aplicar el test de Pearson: Estableciendo el umbral de correlación en el 0.05, observamos que el p-value obtenido está muy por debajo, por lo que podemos afirmar rotundamente que ambas variables están fuertemente relacionadas; o dicho de otro modo, podemos refutar la hipótesis nula 'las dos variables son independientes'

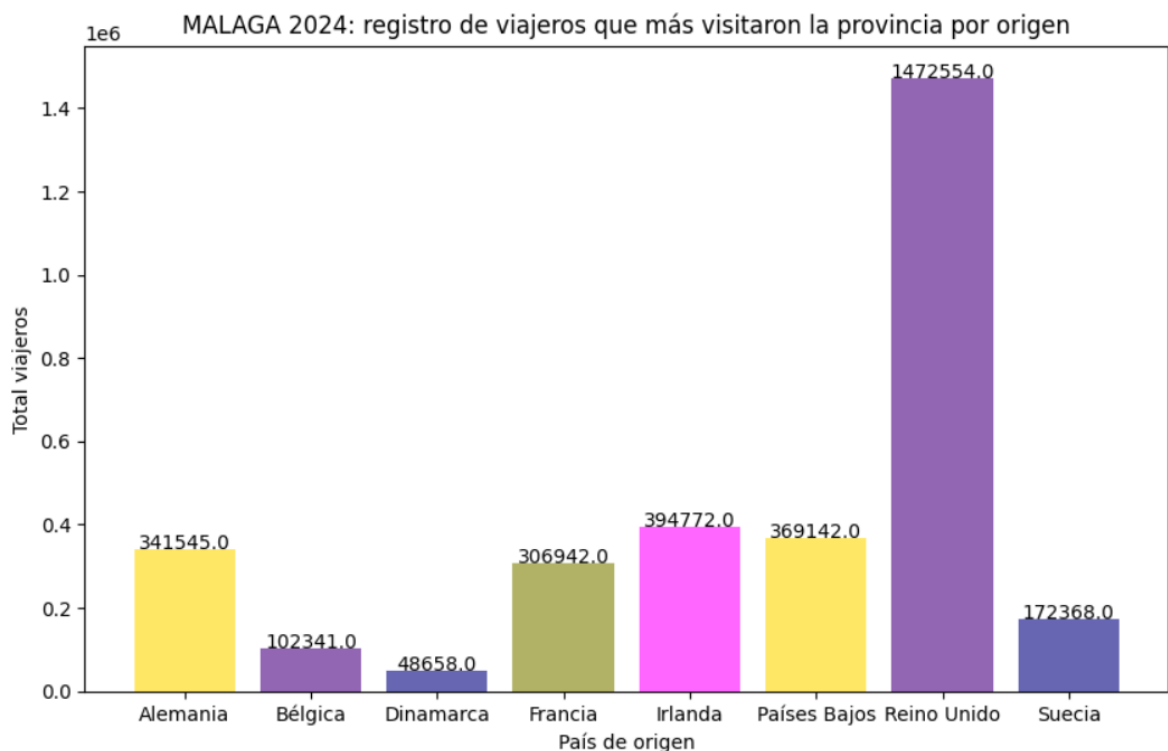
## Timeplot MES vs CATEGORICAS



Sorprendentemente, observamos en TURISTAS y en PERNOCTACIONES un descenso en el mes de Julio, estando el pico máximo en Junio y Agosto. Esto se puede explicar con la gráfica de ESTANCIA\_MEDIA, pues ésta llega a su pico máximo entre Julio y Agosto. La estancia de los viajeros es más larga en el mes de Julio, lo que puede significar que hay menos registros de entrada ya que los clientes ya se encuentran en la provincia.

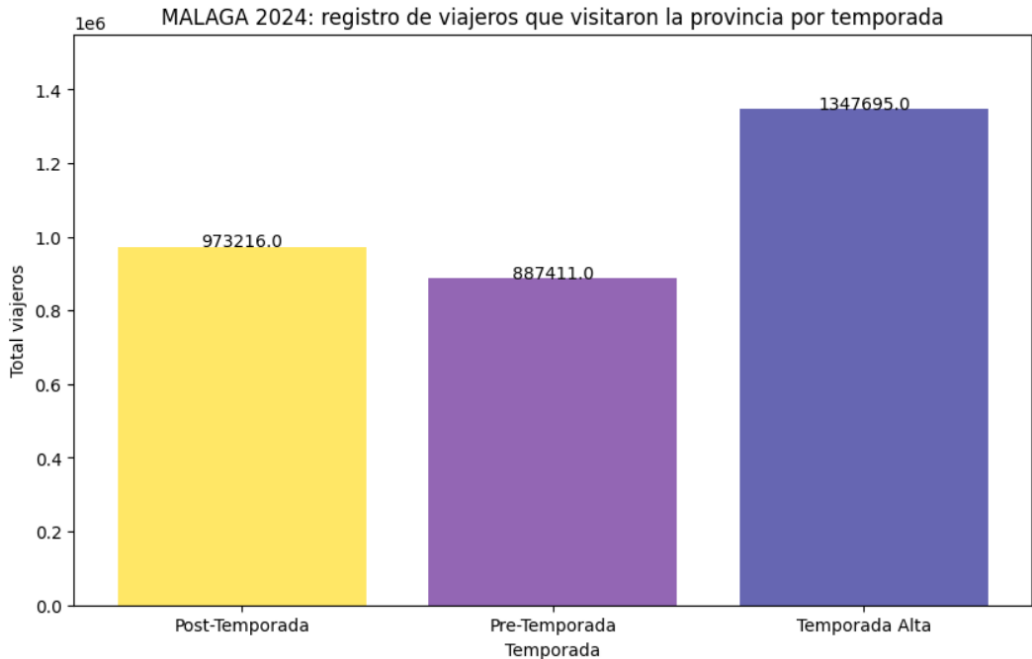
Por otra parte, observamos como tanto en TURISTAS como en PERNOCTACIONES los meses pre-temporada (Abril y Mayo) y post-temporada (Septiembre y Octubre), son muy similares.

Diagrama de barras PAIS\_ORIGEN vs TURISTAS



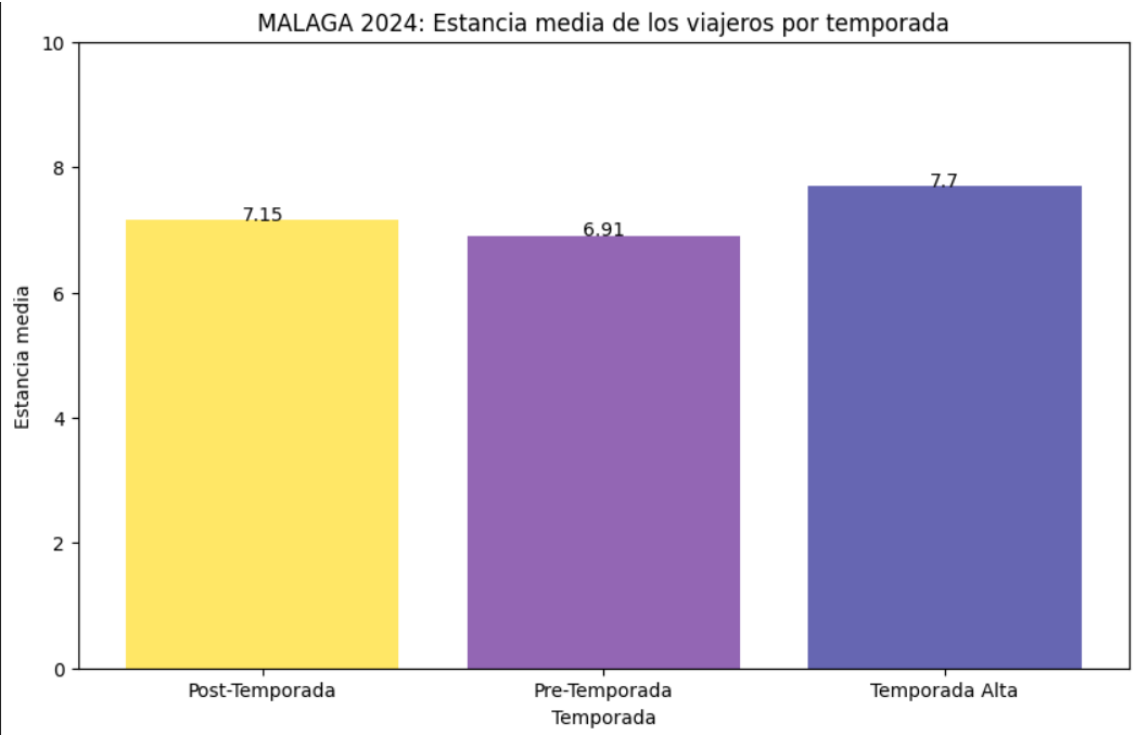
Reino Unido es de lejos el país que más viajó a Málaga en 2024. La diferencia con el segundo puesto, ocupado por Irlanda, Países Bajos y Alemania, es abismal. Será interesante ver en el análisis multivariante si estos datos se mantienen en función de las temporadas.

Diagrama de barras TURISTAS vs TEMPORADA



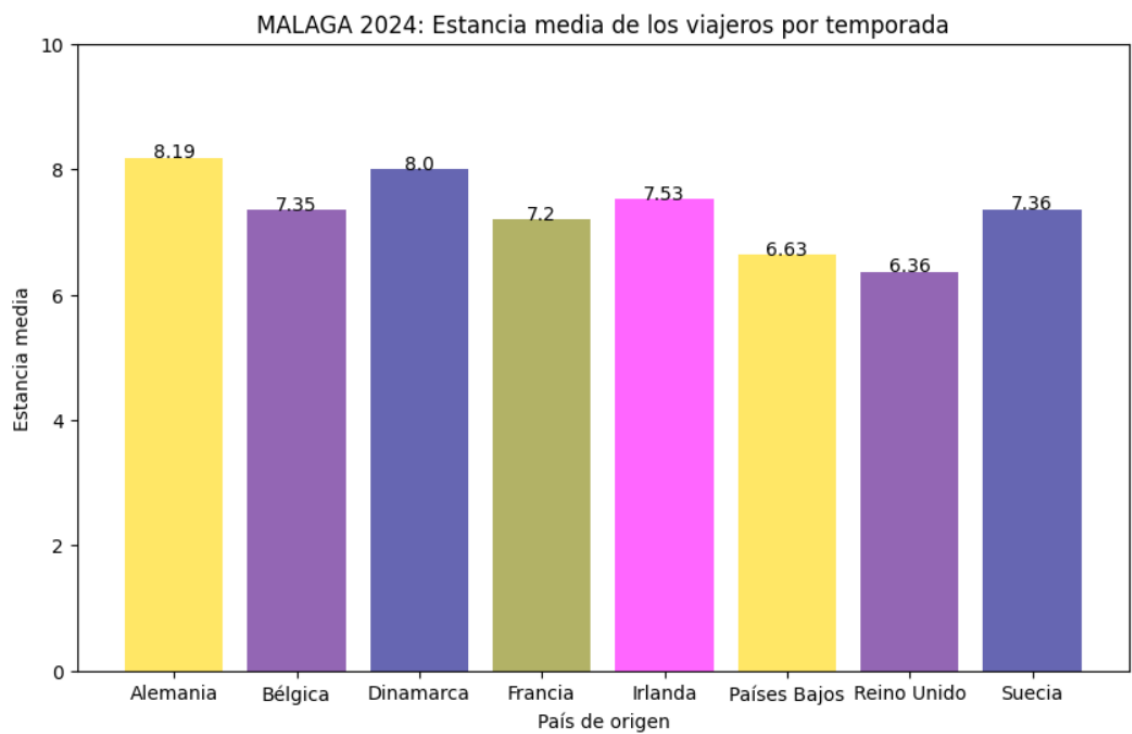
Como era de esperar, en temporada alta encontramos el mayor número de registro de turistas. La pre y post temporada no está nada mal, nuestro objetivo es aumentar estas cifras.

Diagrama de barras ESTANCIA\_MEDIA vs TEMPORADA



No hay gran diferencia en la media de estancia según la temporada, aunque es cierto que en temporada alta los clientes tienden a hospedarse más tiempo. Como siempre, cuidado con las tendencias centrales. No tenemos los datos de cuánto tiempo se ha quedado cada turista así que no podemos indagar más.

**Diagrama de barras PAIS\_ORIGEN vs ESTANCIA\_MEDIA**

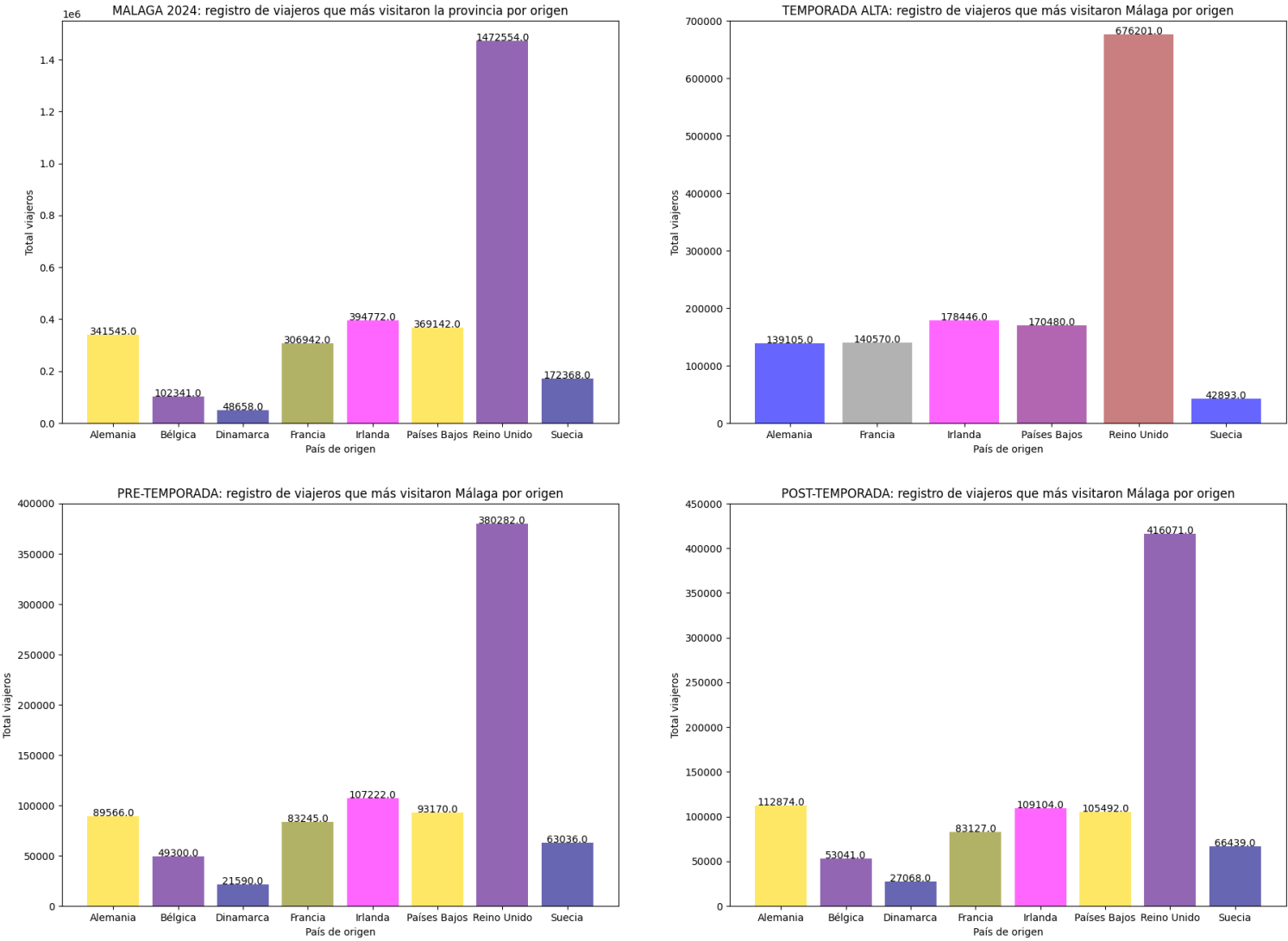


Llama la atención cómo Reino Unido, siendo el país con más registro de turistas, es el que menos días de media se hospeda en Málaga. Esto se puede explicar precisamente por el gran número de viajeros que visitan la provincia, ya que muchos de ellos puede que se queden menos de 3 noches. Al no tener estos datos disponibles, no podemos saber mucho más.

De lo contrario, llama la atención que países con menor presencia de turistas sean los que mayor tiempo de media se queden en Málaga (por ejemplo: Dinamarca, Suecia, Bélgica).

# ANÁLISIS MULTIVARIANTE

## Diagrama de barras entre 'PAIS\_ORIGEN' y 'TURISTAS' en función de 'TEMPORADA'



Seguimos viendo los mismos resultados:

Reino Unido encabeza indiscutiblemente el ranking de total de turistas que visitaron Málaga, tanto en la temporada completa 2024, como en las subdivisiones establecidas.

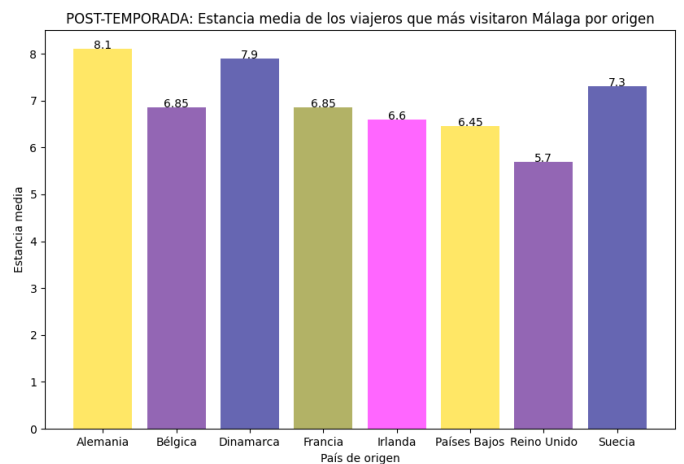
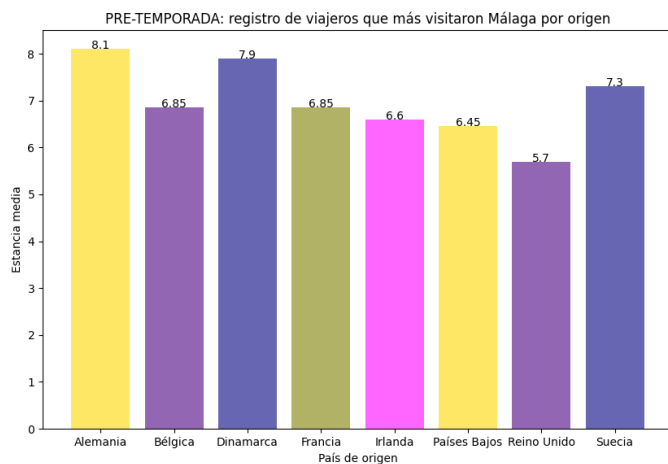
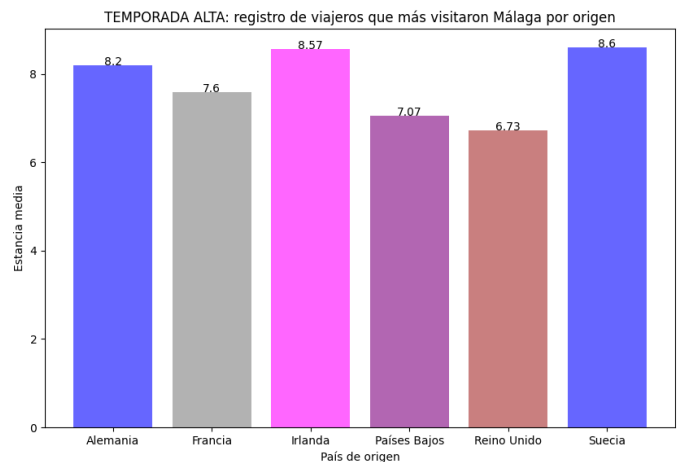
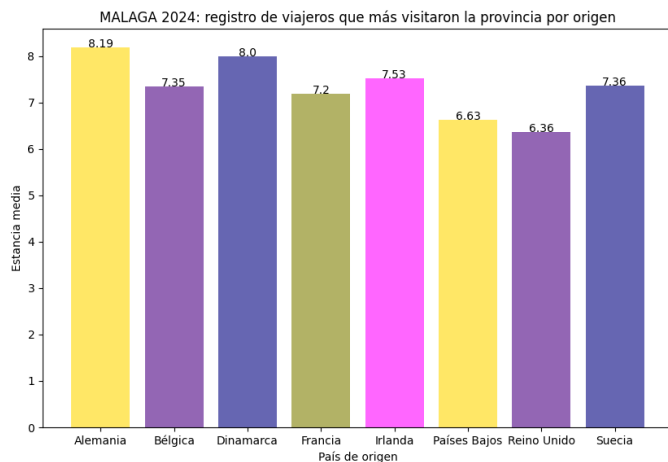
Irlanda, Francia, Alemania y Países Bajos le siguen de lejos, pero con registros bastante elevados. Ocuparían el segundo lugar en el ranking.

Es curioso observar que países como Suecia, Dinamarca y Bélgica prefieren visitar Málaga fuera de temporada. ¿Será una cuestión climática? Las temperaturas en verano pueden ser demasiado extremas.

Vemos que el país de origen afecta claramente en el número de turistas que nos visitan o, al menos, eso podemos pensar a priori. Hemos querido aplicar el test de ANNOVA para corroborar nuestra hipótesis: ambas variables están relacionadas. Como la distribución es muy semejante en las divisiones por temporada (exceptuando la temporada alta pese a que se mantiene la similitud), aplicamos el test al dataset entero.

Sin embargo, la muestra de datos es tan pequeña que no es válida para establecer una correlación. Además, los datos que estamos estudiando no son una muestra, sino toda la población (es decir TODOS los registros de turistas en la provincia de Málaga en 2024). Pese a haber limpiado y acotado los datos, éstos siguen representando la TOTALIDAD, por lo tanto, son lo que son y no tenemos que demostrar necesariamente que hay correlación entre ellos.

## Diagrama de barras entre 'PAIS\_ORIGEN' y 'ESTANCIA\_MEDIA' en función de 'TEMPORADA'



De nuevo las mismas conclusiones:

Es interesante ver que Reino Unido encabeza por debajo el ranking de media por estancia. Este dato hay que cogerlo con pinzas pues es una tendencia central y, al haber tanta dispersión en cuanto al número de viajeros, no podemos tomarlo como una verdad absoluta.

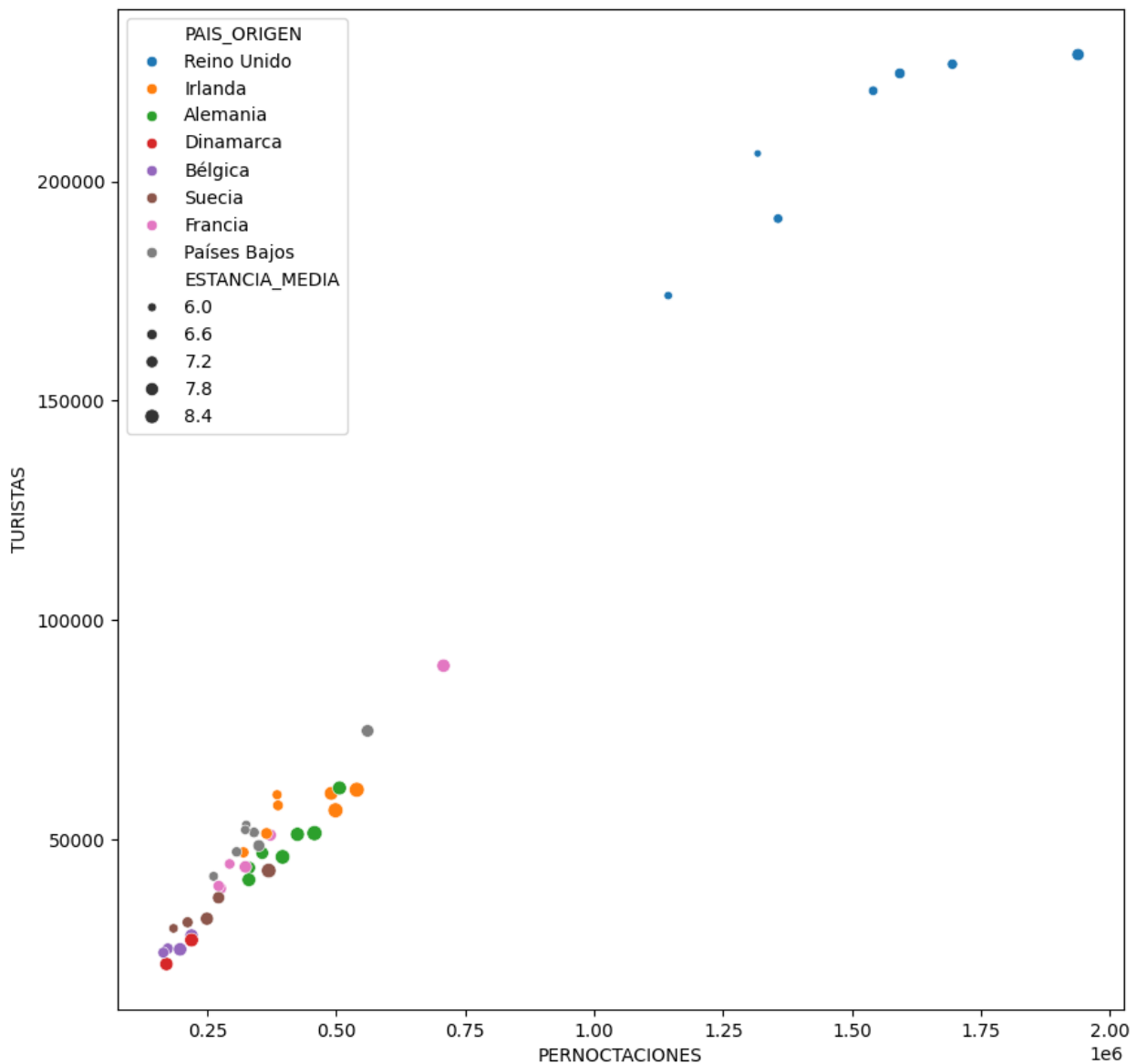
Países como Suecia y Dinamarca, que en ocupaban los últimos puestos en la gráfica PAIS\_ORIGEN vs TURISTAS, destacan en este análisis como dos de los países junto a Alemania que más días de media se hospedan.

De cara a la efectividad del negocio nos interesa captar clientes que prolonguen sus estancias, ya que generan mayor beneficio en todos los sentidos: se dejan más dinero en el rate por noche, gastan más en los outlets del hotel y, por último, housekeeping pierde menos recursos que con las estancias más cortas.

De nuevo, la muestra de datos es demasiado pequeña para aplicar un test de ANNOVA favorable.



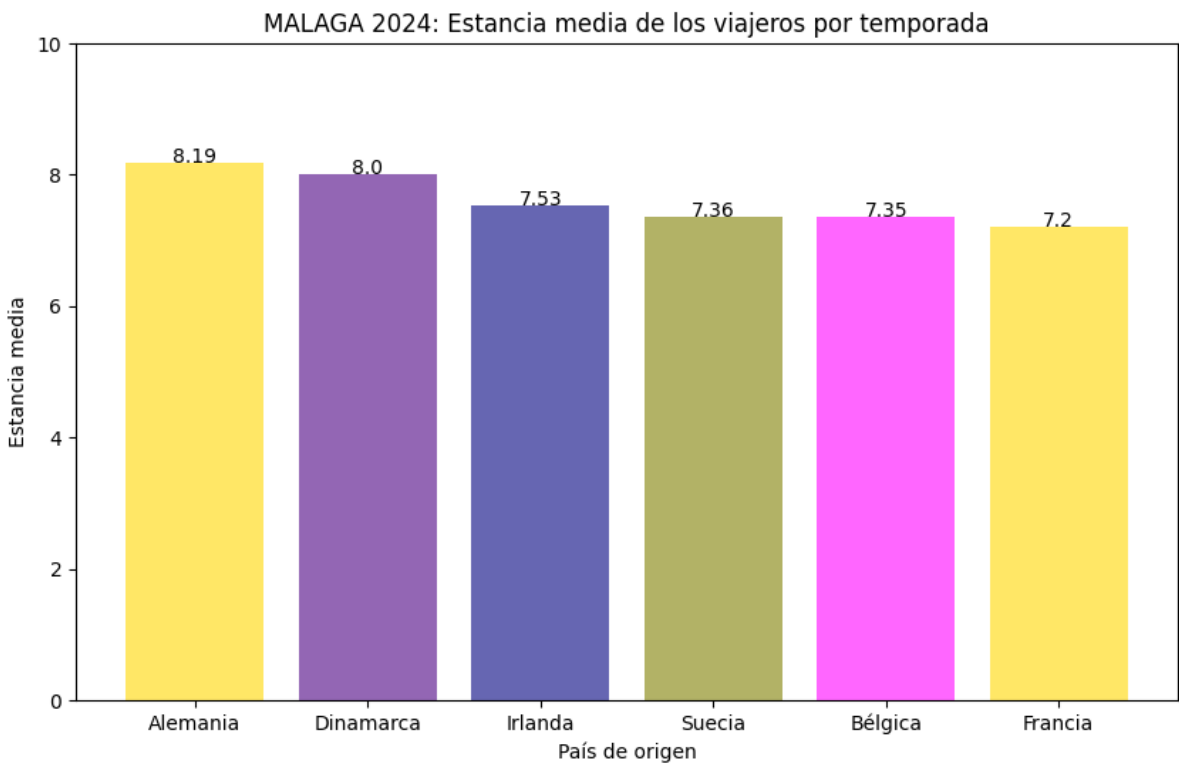
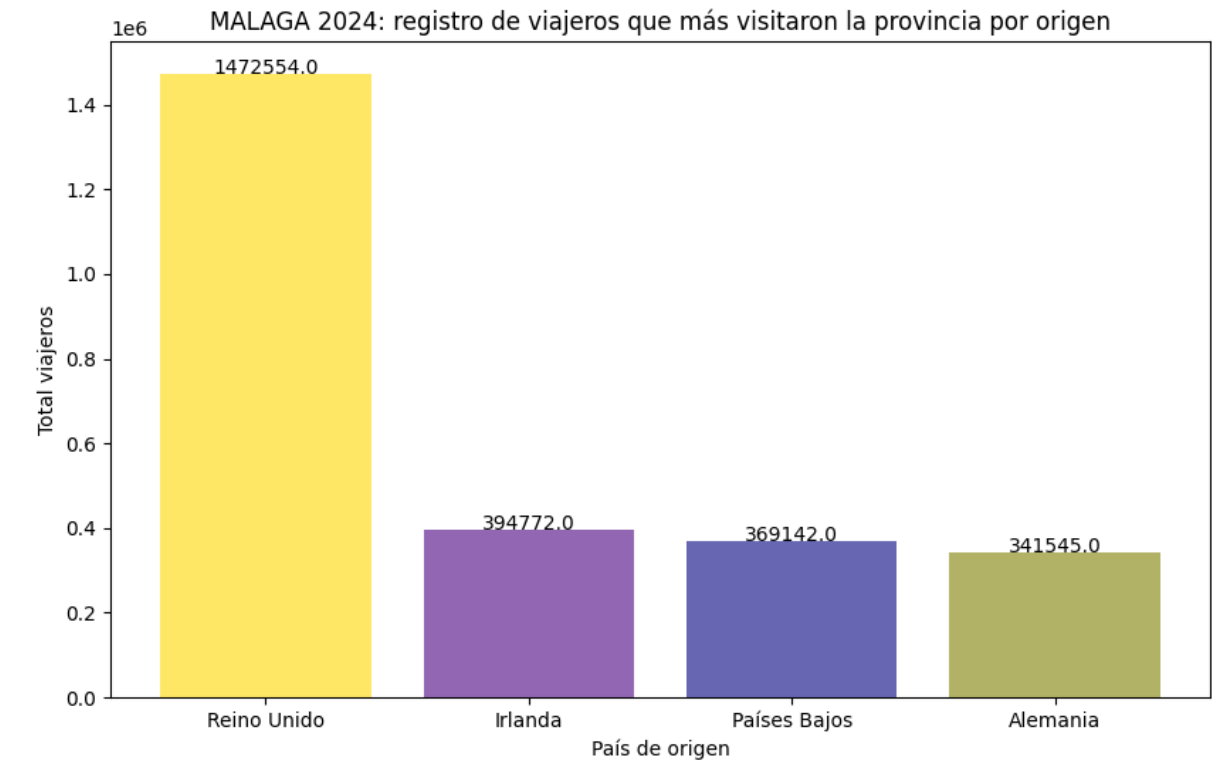
## Análisis de 4 variables TURISTAS, PERNOCTACIONES, ESTANCIA\_MEDIA y PAIS\_ORIGEN



La gráfica corrobora lo que hemos ido viendo hasta ahora:

El menor num de turistas y pernoctaciones se concentran en las pre y post temporadas. En temporada alta destacan el mayor num de turistas, pernoctaciones y también estancia media. En cuanto a los outliers que rozan los 200000, el pico se sitúa obviamente en la temporada alta.

# VISUALIZACIONES RELEVANTES PARA LA PRESENTACIÓN



## CONCLUSIONES FINALES Y RECOMENDACIONES

Se confirma nuestra hipótesis inicial, a la que hemos sacado mucho más jugo tras el análisis:

Reino Unido encabeza de manera arrolladora el total de registros turísticos de viajeros que visitaron Málaga en la temporada 2024. Una opción es optar por una campaña de marketing exclusiva y más agresiva en el mercado anglosajón. A su vez, ésta podría extenderse a Irlanda, el segundo país con más registros.

Alemania, Suecia, Dinamarca y Bélgica encabezan los países que más tiempo de media se alojan en Málaga. Hablamos de una media, no de términos absolutos, pero sigue siendo un dato muy interesante. Si lo que queremos es diversificar nuestro mercado, otra opción es enfocar nuestra campaña de marketing por zonas geográficas: países Escandinavos, francófonos y germánicos (lo que nos permitiría incluir países como Francia y Países Bajos, que también se encuentran en el top de registros turísticos).

De cara a la efectividad del negocio nos interesa captar clientes que prolonguen sus estancias, ya que generan mayor beneficio y nos permiten optimizar los recursos humanos del hotel.

Lo ideal, siempre que el presupuesto lo permita, sería diversificar la campaña en dos focos: 1. El principal para el Reino Unido; 2) Uno secundario pero no menos importante para los mercados escandinavos, francófonos y germánicos.