



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

## Trabajo de Fin de Grado

---

**Ciencia de Datos con Python aplicada al  
Turismo de Canarias**

*Data Science with Python applied to Canary Island's Tourism*

Óscar Hernández Díaz

---

La Laguna, 14 de julio de 2023

D. **José Andrés Moreno Pérez**, Catedrático de Universidad adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

## **C E R T I F I C A**

Que la presente memoria titulada:

*"Ciencia de Datos con Python aplicada al Turismo de Canarias"*

ha sido realizada bajo su dirección por D. **Óscar Hernández Díaz**, con N.I.F. 54.112.806-Q.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 14 de julio de 2023

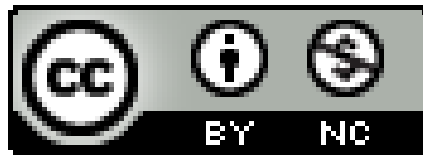
# Agradecimientos

Sobre todo a mi familia y compañeros de carrera por estar siempre presentes durante toda el desarrollo del TFG.

A mi tutor de TFG Jose Andrés Moreno Pérez por su paciencia y apoyo haciendome mejorar en cada paso que daba.

Y por supuesto al profesor Agustín Santana Talavera, de forma voluntaria ha estado presente desde el inicio del desarrollo, aportando ideas como opiniones sobre este trabajo.

# Licencia



© Esta obra está bajo una licencia de Creative Commons Reconocimiento-  
NoComercial 4.0 Internacional.

## **Resumen**

*En este Trabajo de Fin de Grado se estudia el uso de Python para explorar el comportamiento del turismo en las Islas Canarias. Se tienen en cuenta el país de origen de estos, el gasto en conceptos abstractos como la alimentación, el transporte, y más. Se dispone de varios conjuntos de datos desde 2018-2022 desglosados en trimestres derivados de la Encuesta de Gasto Turístico (ISTAC).*

*Para la realización del estudio se ha usado un cuaderno de Jupyter donde emplearemos librerías de Python como Pandas o Numpy. Se incluye una exploración y análisis previo de los datos para identificar que áreas tienen mayor prioridad a ser estudiadas. Con el fin de mejorar la visualización y comprensión de los conceptos estudiados, se han aplicado diversas técnicas y bibliotecas de Python que se adaptaban mejor a los datos y ofrecían un rendimiento óptimo durante el desarrollo del código. Se seleccionaron las herramientas más adecuadas, como NumPy, Pandas, Matplotlib, Scikit-Learn, y se emplearon en combinación con las técnicas apropiadas para lograr resultados efectivos en el análisis y visualización de los datos.*

**Palabras clave:** Datos, Turismo, Ciencia, Análisis, Python, Pandas, Numpy, Jupyter, Scikit-Learn, Matplotlib

## **Abstract**

*This Final Degree Project studies the use of Python to explore the behaviour of tourism in the Canary Islands. We take into account the country of origin of tourists, expenditure on abstract concepts such as food, transport, and more. Several datasets are available from 2018-2022 broken down into quarters derived from the "Encuesta de Gasto Turístico".*

*A Jupyter notebook has been used to carry out the study, where we will use Python libraries such as Pandas or Numpy. A preliminary exploration and analysis of the data is included to identify which areas have the highest priority to be studied. For better visualisation and understanding of the concepts being studied, different techniques and Python libraries have been applied that best fit the data and provided the best performance when developing the code.*

**Keywords:** Data, Tourism, Science, Analytics, Python, Pandas, Numpy, Jupyter

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Ciencia de Datos . . . . .	1
1.2. Antecedentes . . . . .	1
1.2.1. Estado del arte . . . . .	2
1.3. Objetivos . . . . .	3
1.3.1. Objetivo General . . . . .	3
1.3.2. Objetivos Específicos . . . . .	3
<b>2. Método</b>	<b>5</b>
2.1. Obtención de los datos . . . . .	5
2.2. Tecnologías y Herramientas . . . . .	6
<b>3. Desarrollo</b>	<b>8</b>
3.1. Tratamiento y limpieza de los datos . . . . .	8
3.2. Número de turistas . . . . .	9
3.2.1. Número de turistas que han realizado la encuesta en cada aeropuerto	9
3.2.2. Número de turistas según país de residencia en total . . . . .	9
3.2.3. Número de turistas según su nacionalidad y el aeropuerto de entrada	12
3.3. Gasto turístico . . . . .	14
3.3.1. Perfil por país de residencia por gasto separado por edades . . . . .	14
3.3.2. Sankey Diagram aplicado a los turistas . . . . .	17
3.3.3. Gráfico Sunburst por año, edad, país de residencia y gasto . . . . .	18
3.4. Motivación de los turistas en las islas . . . . .	21
3.4.1. Motivación principal de cada isla . . . . .	24
<b>4. Conclusiones y futuras líneas de trabajo</b>	<b>25</b>
4.1. Futuras Líneas de trabajo . . . . .	25
<b>5. Summary and Conclusions</b>	<b>27</b>
<b>6. Presupuesto</b>	<b>28</b>
<b>A. Algoritmos del Gasto Turístico</b>	<b>29</b>
A.1. Algoritmo de perfil por país de residencia por gasto separado por edades . .	29
A.2. Algoritmo para el gráfico Sunburst . . . . .	31

# Índice de Figuras

1.1. IMPACTUR Canarias. Fuente: (2) . . . . .	3
2.1. Datos formato tabla desde VSCode. . . . .	6
3.1. Turistas totales encuestados por años y el aeropuerto origen . . . . .	9
3.2. Turistas encuestados en 2019. . . . .	10
3.3. Turistas encuestados en 2022. . . . .	10
3.4. Numero de turistas encuestados segun su pais de residencia en 2018 . . . . .	11
3.5. Numero de turistas encuestados segun su pais de residencia en 2019 . . . . .	11
3.6. Numero de turistas encuestados segun su pais de residencia en 2021 . . . . .	11
3.7. Numero de turistas encuestados segun su pais de residencia en 2022 . . . . .	11
3.8. Datos formato tabla desde VSCode. . . . .	12
3.9. Datos formato tabla desde VSCode. . . . .	12
3.10 Datos formato tabla desde VSCode. . . . .	13
3.11 Gasto medio en el año 2022 entre 16 y 25 años . . . . .	15
3.12 Gasto medio en el año 2022 entre 26 y 45 años . . . . .	15
3.13 Gasto medio en el año 2022 entre 45 y 65 años . . . . .	16
3.14 Gasto medio en el año 2022 de mas de 66 años . . . . .	16
3.15 Diagrama Sankey. Grupo Edades, país y tipo de gasto . . . . .	17
3.16 Diagrama Sankey. Grupo edades y desglose de gastos . . . . .	17
3.17 Diagrama Sankey. Pais residencia y grupos de edades . . . . .	18
3.18 Gráfico Sunburst Total . . . . .	19
3.19 2022 . . . . .	20
3.20 2022, 26-45 años . . . . .	20
3.21 2022, 26-45 años, España . . . . .	21
3.22 Mapa de calor sobre la media en Aficiones . . . . .	22
3.23 Mapa de calor sobre la media en Descansar . . . . .	22
3.24 Mapa de calor sobre la media en Divertirse . . . . .	22
3.25 Mapa de calor sobre la media en Explorar . . . . .	23
3.26 Mapa de calor sobre la media en Familia . . . . .	23
3.27 Motivación principal de cada isla . . . . .	24



# Índice de Tablas

6.1. Coste del desarrollo . . . . . 28

# Capítulo 1

## Introducción

En este primer capítulo se hace una breve introducción a la Ciencia de Datos y al turismo en Canarias como antecedentes del trabajo, y se exponen los objetivos de la investigación.

### 1.1. Ciencia de Datos

La Ciencia de Datos se trata de un subconjunto de la Inteligencia Artificial (IA). Esta se basa principalmente en el estudio de los datos con el fin de extraer información significativa para después ser mostrada mayoritariamente a entidades o empresas. Es una rama multidisciplinaria que combina principios del campo de las matemáticas, la estadística e ingeniería computacional para analizar grandes bancos de datos.

En esta era tecnológica la ciencia de datos es de suma importancia porque combina herramientas, métodos y tecnologías para generar significados a partir de los datos. Actualmente las organizaciones modernas recopilan y guardan datos constantemente; hay una proliferación de dispositivos electrónicos que recopilan y almacenan información de manera prácticamente automática. Al disponer de grandes cantidades de datos de texto, audio, video e imágenes, las entidades aprovechan este recurso aplicando ciencia de datos para estudiar por ejemplo, el impacto que podría tener un producto en su tienda.

Con el avance de estas tecnologías no cabe duda que se ha aplicado en cada ámbito que ha sido posible. En este documento se va a centrar en el area del turismo y el impacto que ha tenido en las islas Canarias aplicando estrategias analíticas y estadísticas. El banco de datos a los que se han aplicado estas estrategias ha sido obtenido del Instituto Canario de Estadística (ISTAC) [1]. Los datos fueron recopilados a través de una encuesta escrita proporcionada a los turistas en el area de salidas, previo al embarque, en los diferentes aeropuertos de canarias; mas adelante en el capitulo 2 se hablará un poco mas en detalle sobre la encuesta.

Para este proyecto, junto con herramientas como Python para el desarrollo de ciencia de datos, se han realizado una serie de análisis sobre el numero de turistas en las ilas, así como de que paises son más comunes, cuales son sus gastos medios, el tipo de gasto y más.

### 1.2. Antecedentes

En esta sección se expondrán algunos aspectos de la situación actual del turismo en Canarias, el impacto que ha tenido el COVID-19 sobre este y algunas cuestiones de interés

actuales.

### **1.2.1. Estado del arte**

Para poder ponernos en contexto sobre la importancia e impacto que tiene el turismo en las islas Canarias podemos ver que los estudios de impacto económico del turismo IMPACTUR en Canarias cuantifica en 10.802 millones de euros los efectos directos del turismo en 2019 que representaron el 22,9% del total del PIB canario, y generaron un total de 231.575 puestos de trabajo (el 27,2% del total del empleo de las Islas Canarias) antes de la crisis sanitaria. Podemos ver en la imagen 1.1 un infograma aportada por IMPACTUR sobre este impacto.

Después de la pandemia mundial el empleo turístico aumentó un 29,5% en el primer trimestre de 2022 [3], generando así una alta demanda de puestos de trabajos y a medida que se ha ido mejorando la situación del COVID los niveles de turistas entrantes incluso sobrepasan el record pre-pandemia [5].

Javier María Tello Martínez comenta en su trabajo "Potencial aplicación del Big Data al sector turismo con especial referencia a España"[6] los beneficios que se pueden obtener al aplicar ciencia de datos y Big Data en el sector terciario. Principalmente destaca que aplicando las técnicas adecuadas podemos obtener una descripción del perfil del turista altamente precisa, a efectos inmediatos, una empresa que conoce mejor al turista será capaz de ofrecer un producto que se le ajuste, adecuando sus necesidades, acompañada a la capacidad de mejorar la experiencia del turista.

Por último, también podemos ver como organizaciones gubernamentales [7] han hecho uso de estas tecnologías para poder sacarle un uso positivo para la sociedad. Como bien dicen, estas herramientas les proporcionan analizar grandes volúmenes de información sobre las valoraciones y comentarios que realizan en la web los viajeros que visitaron los distintos atractivos turísticos, algo semejante a lo que se hace en este proyecto.

## Impacto económico del Turismo en las Islas Canarias

Año 2019. Niveles prepandemia

Año 2020

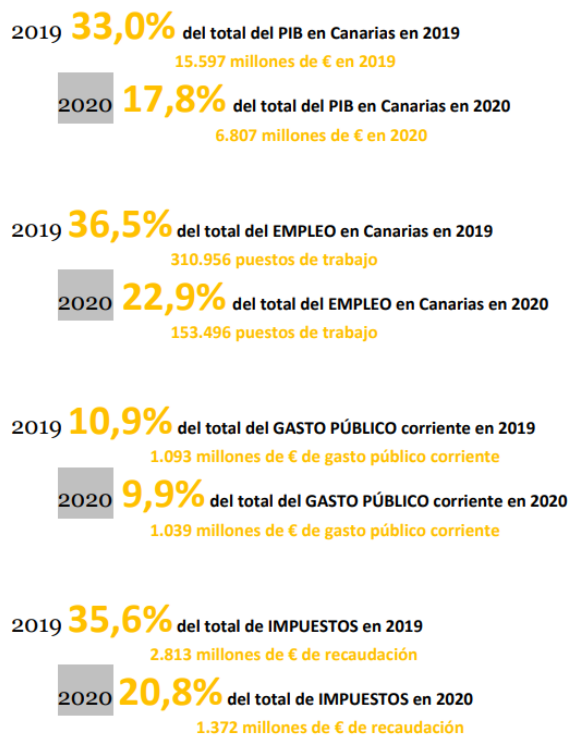


Figura 1.1: IMPACTUR Canarias. Fuente: [2]

## 1.3. Objetivos

### 1.3.1. Objetivo General

El propósito de este Trabajo de Fin de Grado es mostrar cómo, mediante la manipulación y procesamiento de un conjunto de datos, podemos realizar un análisis exhaustivo en diversos aspectos del turismo en Canarias. El objetivo final es obtener un conocimiento detallado que pueda respaldar la toma de decisiones de entidades relacionadas con la industria hotelera, el turismo e incluso en el ámbito gubernamental.

### 1.3.2. Objetivos Específicos

#### Estudio del gasto de los turistas

El propósito de este estudio es realizar un análisis de los turistas procedentes de distintos países y examinar el gasto de viaje de los turistas que visitan Canarias. El objetivo es identificar en qué áreas los turistas muestran una mayor disposición para gastar su dinero. Al obtener esta información, se podrá comprender mejor los patrones de gasto de los turistas y orientar estrategias para impulsar el crecimiento económico en la región.

## **Estudio de la motivación de los turistas**

Por último, se llevó a cabo un análisis para determinar la razón principal por la que los turistas eligen visitar las Islas Canarias. Este estudio tiene como objetivo identificar áreas relevantes que puedan ser utilizadas como base para la toma de decisiones posteriores. Al comprender las motivaciones principales de los turistas, se podrán desarrollar estrategias y acciones específicas para promover y mejorar la experiencia turística en las Islas Canarias, impulsando así el sector turístico.

El siguiente capítulo se dedica al procedimiento desarrollado para alcanzar estos objetivos y los resultados obtenidos. El Trabajo Fin de Grado finaliza con un presupuesto orientativo y la exposición de unas conclusiones derivadas del trabajo y de unas líneas futuras en las que se daría continuidad a la investigación desarrollada.

# Capítulo 2

## Método

En este capítulo se hablará de cómo se han obtenido los datos y qué herramientas/librerías de python se han usado a lo largo del proyecto. En el apartado de Datos Se comentará la distribución de los datos y las variables aportadas. En el apartado Tecnologías y Herramientas se comentará el entorno donde se ha realizado el trabajo y que herramientas se han hecho uso para este.

### 2.1. Obtención de los datos

Se enfocará en el análisis de datos de los años comprendidos entre 2018 y 2022, excluyendo el año 2020 debido a que no es representativo. El año 2020 se descarta debido a las restricciones de movimiento y las consecuencias de la pandemia de COVID-19, que tuvieron un impacto significativo en los datos turísticos. Como se nombró en los capítulos anteriores los datos han sido recogidos por el Instituto Canario de Estadístico (ISTAC) [8]. Esta agencia recoge estos datos a través de una encuesta que se le realiza a los turistas en la vuelta de su viaje. En dicha encuesta se recogen numerosas preguntas de diferentes índoles. Cada pregunta está recogida en un apartado facilitando la tarea de seleccionar aquellas preguntas que se corresponden con los intereses de la investigación. Tenemos apartados por ejemplo del motivo del viaje, organización del viaje, estancia y alojamiento en canarias, gastos del grupo de viaje,... [9]. Una vez recogido los datos ISTAC se encarga de digitalizarlos y volcarlos en un archivo .csv. Para mayor información sobre la metodología de esta encuesta se puede recurrir a su página oficial [13].

En la figura 2.1 podemos ver una parte con los datos que vamos a trabajar en formato de tabla. En esta tabla se ha agrupado todos los años con los que hemos trabajado, consta de 150.515 filas/turistas encuestados y 230 columnas/variables. Es una cantidad bastante grande de variables pero no trabajaremos con todas ellas, sino como se ha dicho, solamente con aquellas que nos interese investigar.

	ID	NUMERO_CUESTIONARIO	OLA	AEROPUERTO_ORIGEN	PAIS_DESTINO	SEXO	EDAD	NACIONALIDAD	PAIS_RESIDENCIA	PROPOSITO	...
0	1		1	4	ACE	DEU276	1	70	DEU276	DEU276	6 ...
1	2		10	4	ACE	DEU276	1	34	ESP724	DEU276	7 ...
2	3		100	4	ACE	GBR826	6	56	GBR826	GBR826	6 ...
3	4		101	4	ACE	GBR826	6	67	GBR826	GBR826	6 ...
4	5		102	4	ACE	GBR826	6	53	GBR826	GBR826	6 ...
...	...		...	...	...	...	...	...	...	...	...
39582	1946		997	10	LPA	PRT620	6	35	ZZZ900	ZZZ900	6 ...
39583	7401		997	12	ACE	ESP724	1	65	ESP724	ESP724	10 ...
39584	7402		998	12	ACE	ESP724	6	49	ESP724	ESP724	7 ...
39585	3570		998	10	TFS	FIN246	6	42	FIN246	FIN246	6 ...
39586	7403		999	12	ACE	ESP724	6	37	ZZZ900	ZZZ900	6 ...

150515 rows × 230 columns

Figura 2.1: Datos formato tabla desde VSCode.

## 2.2. Tecnologías y Herramientas

La ciencia de la información es un ámbito que emplea una variedad de métodos estadísticos, matemáticos y computacionales para examinar y manipular diversas magnitudes de información con el objetivo de obtener conocimientos relevantes que impulsen la toma de decisiones. Para llevar a cabo esta labor, los expertos utilizan técnicas de aprendizaje automático, exploración de datos y presentación visual para realizar un análisis adecuado y descubrir información de gran valor.

Con el fin de llevar a cabo todas estas tareas, se empleó Python, uno de los lenguajes de programación más populares en la actualidad. Además, tanto Python como R son considerados como los lenguajes más apropiados para realizar este tipo de análisis estadísticos y representaciones visuales de datos. Se ha elegido Python por su facilidad de aprendizaje y de programación, lo que hace que sea el lenguaje de programación más apto tanto para principiantes, como para expertos.

Python cuenta con una amplia gama de bibliotecas especializadas en ciencia de datos, que incluyen NumPy, Pandas, Matplotlib, Scikit-Learn y TensorFlow. Estas bibliotecas brindan herramientas poderosas para el análisis y procesamiento de datos. Además, Python es un lenguaje de programación interactivo que permite explorar y analizar datos en tiempo real, facilitando así la experimentación y la toma de decisiones basada en resultados inmediatos. Asimismo, Python es un lenguaje de programación versátil que se utiliza en una amplia variedad de aplicaciones, lo que demuestra su utilidad y adaptabilidad en diferentes contextos.

Además, Python también cuenta con otras bibliotecas especializadas en ciencia de datos, como Plotly, Folium y Geopandas. Estas bibliotecas amplían las capacidades de visualización y análisis de datos al proporcionar herramientas para crear gráficos interactivos, mapas geoespaciales y análisis geográficos. Con estas bibliotecas, Python se convierte en una herramienta aún más poderosa y versátil para el análisis y visualización de datos en diferentes dominios y aplicaciones.

Con el propósito de llevar a cabo esta investigación, se emplearon diversas bibliotecas o paquetes de Python que permitieron realizar el análisis de datos y la visualización correspondiente:

- **Numpy.** Es una biblioteca en Python que proporciona soporte para arreglos multidimensionales y funciones matemáticas de alto rendimiento. Es una de las bibliotecas fundamentales en el ecosistema de ciencia de datos de Python y se utiliza amplia-

mente en diversas áreas, como la manipulación y análisis de datos, la computación científica, el aprendizaje automático y la visualización de datos.

- **Pandas.** Es una biblioteca en Python que proporciona estructuras de datos y herramientas de análisis de datos de alto rendimiento. Se utiliza ampliamente en el campo de la ciencia de datos y análisis de datos para realizar tareas como la manipulación, limpieza, transformación y análisis de datos de manera eficiente.
- **Matplotlib.** Es una biblioteca de visualización de datos en Python ampliamente utilizada. Permite crear gráficos de alta calidad, personalizar su apariencia y estilo, visualizar datos multidimensionales y se integra bien con otras bibliotecas de análisis de datos en Python. Matplotlib es una herramienta esencial para la comunicación efectiva de datos y resultados en diversas aplicaciones de análisis de datos, investigación científica y presentación de informes.
- **Plotly.** Es una biblioteca de visualización interactiva que permite crear gráficos y visualizaciones interactivas en Python. Proporciona una amplia gama de tipos de gráficos, como gráficos de dispersión, gráficos de barras, gráficos de líneas y gráficos de áreas, entre otros. Plotly también ofrece capacidades interactivas, como zoom, desplazamiento y herramientas de selección de datos. Es ampliamente utilizado en el análisis de datos, la generación de informes interactivos y la creación de paneles de control interactivos en aplicaciones web. Plotly también es compatible con múltiples lenguajes de programación, lo que lo convierte en una opción flexible y poderosa para la visualización de datos interactivos.
- **Folium.** Es una biblioteca de Python que permite crear mapas interactivos y visualizaciones geoespaciales. Utiliza la potencia de la biblioteca Leaflet.js para generar mapas interactivos en HTML. Folium facilita la visualización de datos geográficos y su análisis mediante la superposición de capas de datos sobre un mapa interactivo. Permite agregar marcadores, polígonos, líneas y otras formas geográficas, así como personalizar su apariencia y estilos. Folium es ampliamente utilizado en campos como la geolocalización, el análisis de datos espaciales, la visualización de rutas y la presentación de resultados geográficos de una manera accesible y atractiva.
- **Geopandas.** Es una biblioteca de Python que permite trabajar con datos geoespaciales en formato vectorial. Combina las capacidades de manipulación y análisis de datos de pandas con las funcionalidades geoespaciales de shapely y fiona. Con Geopandas, se pueden realizar tareas como leer y escribir archivos geográficos, realizar análisis espacial y manipular geometrías. Es una herramienta útil para realizar operaciones espaciales como unión, fusión y agregación de datos geográficos. Además, Geopandas permite la visualización de datos geoespaciales en mapas, facilitando la comprensión y comunicación de la información geográfica.



# Capítulo 3

## Desarrollo

En este capítulo se explicara en detalle el proceso que hemos seguido en los casos de estudio así como los resultados obtenidos.

### 3.1. Tratamiento y limpieza de los datos

Una vez que se ha recopilado la información necesaria, es fundamental realizar una preparación de los datos para nuestra investigación. El primer paso consiste en analizar la organización y estructura de los datos, con el fin de determinar qué información es relevante para el proyecto y qué datos pueden ser descartados. Una de las principales consideraciones iniciales fue identificar qué variables permitían obtener un perfil del encuestado, lo cual facilitaría la agrupación de un gran número de encuestados en categorías con características específicas. Esta segmentación basada en rasgos concretos proporcionará una base sólida para el análisis posterior de los datos y permitirá obtener información más precisa y significativa para nuestro estudio.

Dentro de los datos recopilados, se identificaron columnas relevantes como el sexo, la edad, el país de residencia, el aeropuerto de origen y los gastos. Estas variables proporcionan una base sólida para realizar una variedad de estudios que se reflejarán en el presente documento. A partir de estas columnas, se podrán realizar análisis detallados que permitirán comprender aspectos como las preferencias de viaje según el género, la distribución de edades de los turistas, las tendencias de origen de los visitantes y los patrones de gasto. Estos estudios proporcionarán una visión más completa y detallada de los perfiles de los turistas en las Islas Canarias.

El esquema de la encuesta [9] fue aprovechado para evaluar y valorar cada bloque de preguntas. En particular, se seleccionaron las columnas relacionadas con los bloques de motivo del viaje y los gastos del grupo de viaje.

Para garantizar la lectura correcta de los datos, la primera modificación se realiza utilizando el formato latin-1 en lugar del decodificador utf-8. De hecho, el decodificador utf-8 no puede manejar correctamente los caracteres del alfabeto latino y los signos diacríticos, como los acentos, la letra ñ o la letra ç. Además, aprovechamos la estructura de datos para concatenar todos los archivos de modo que tengamos un solo conjunto de datos ordenados por año. Esta consolidación nos permitirá trabajar más eficientemente con la información y facilitará el análisis posterior. Es importante tener en cuenta que estos son los cambios iniciales necesarios para comenzar a trabajar con representaciones de datos. Sin embargo, en los siguientes pasos, haremos los cambios necesarios para obtener una imagen precisa de la información.

Para concluir esta sección, la figura 3.1 muestra el número total de turistas encuestados por año y el aeropuerto de origen. Estos datos serán la base principal de nuestra investigación, ya que nos proporcionan información crucial sobre la cantidad de turistas que visitan las Islas Canarias. A partir de estos datos, podremos realizar análisis más detallados y extraer conclusiones relevantes para nuestro estudio.

TRIMESTRE	ACE	FUE	LPA	SPC	TFN	TFS
2018	6661	4814	10944	1526	2584	11375
2019	6753	5293	11023	1520	2948	11224
2021	4813	4200	9477	1551	5291	8931
2022	6134	5360	11343	1818	3724	11208

Figura 3.1: Turistas totales encuestados por años y el aeropuerto origen

## 3.2. Número de turistas

### 3.2.1. Número de turistas que han realizado la encuesta en cada aeropuerto

En un primer análisis de los datos, se ha realizado un pequeño estudio para identificar a qué islas suelen visitar más los turistas y obtener conclusiones sobre cuál podría considerarse como la "favorita". Al observar las figuras 3.2 y 3.3, que representan diferentes intervalos de tiempo separados por 3 años, se puede apreciar que prácticamente no hay cambios significativos en las preferencias de los turistas.

En particular, en el caso de Gran Canaria (LPA), se observa que tanto en 2019 como en 2022, el porcentaje de turistas encuestados en esta isla es casi el mismo. Lo mismo ocurre con los turistas en el Aeropuerto de Tenerife Sur. Estos resultados nos permiten identificar que las islas más propensas a recibir turistas son Tenerife y Gran Canaria, seguidas de Lanzarote y Fuerteventura.

Estos datos son consistentes con lo esperado, ya que Tenerife y Gran Canaria son las islas más desarrolladas turísticamente y tienen una mayor densidad de población por kilómetro cuadrado [11]. Por lo tanto, es comprensible que sean las más populares entre los turistas.

### 3.2.2. Número de turistas según país de residencia en total

Es un punto muy válido considerar el país de residencia de los turistas para obtener un panorama más preciso sobre el origen de los turistas y su comportamiento de gasto. Al elegir la variable "país de residencia" en lugar de "nacionalidad", se tiene en cuenta que los turistas pueden vivir en diferentes países, lo que implica que sus salarios y rangos de gasto también pueden variar según su lugar de residencia.

Al considerar el país de residencia, se pueden obtener valores más precisos y representativos al analizar el comportamiento de los turistas. Esto permitirá tener en cuenta las

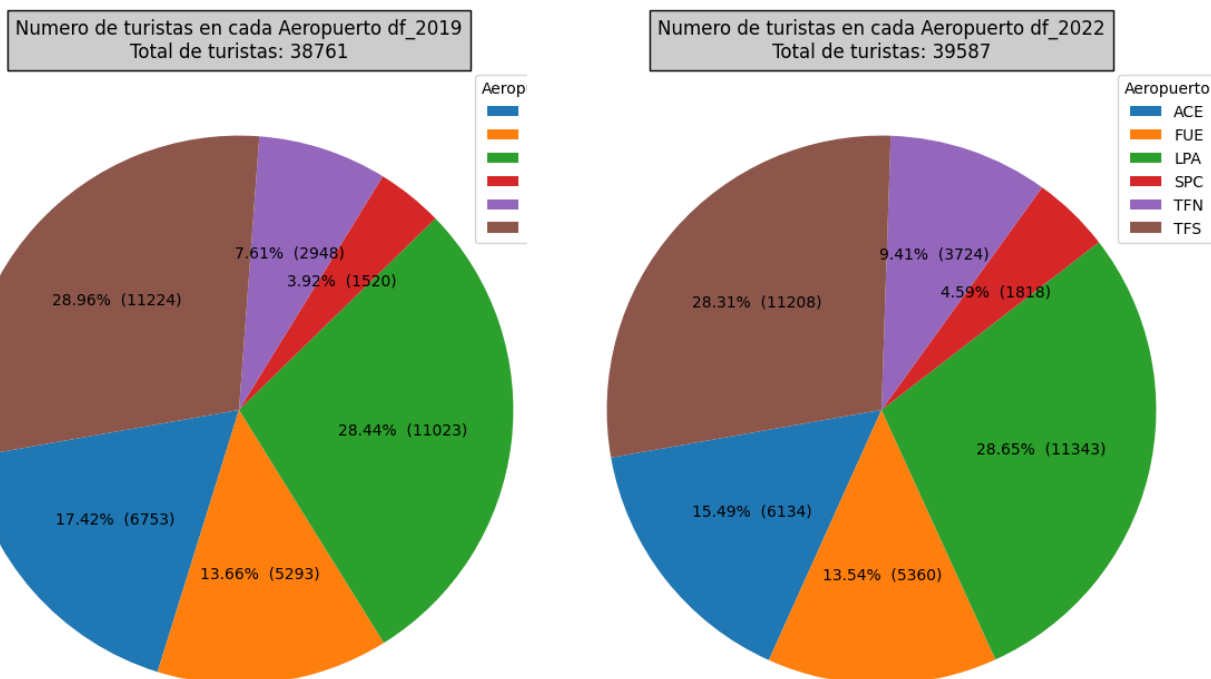


Figura 3.2: Turistas encuestados en 2019. Figura 3.3: Turistas encuestados en 2022.

diferencias económicas y las características específicas de cada país, lo cual puede influir en los patrones de gasto y disponibilidad de los turistas.

En etapas posteriores, al calcular el gasto medio de los turistas, también se utilizará el país de residencia para tener en cuenta estos mismos factores. Esto nos ayudará a obtener resultados más acertados y relevantes para comprender el impacto económico de los turistas en las Islas Canarias.

Al observar las figuras desde la 3.4 hasta la 3.7, se puede visualizar el número de turistas encuestados según su país de residencia a lo largo de cuatro años (2018-2022). Es evidente que los tres países más comunes son Reino Unido (GBR), Alemania (DEU) y España (ESP). En el caso de España, se refiere a aquellos turistas que residen en la península. Es importante mencionar que el país con el código "ZZZ" se refiere a otros países que ISTAC en su metodología [13] especifica que se tratan de países agrupados que no son relevantes individualmente en número, por lo que no es una información válida para nuestro análisis.

A lo largo de estos años, estos tres países (Reino Unido, Alemania y España) dominan en términos de número de turistas, y en general, se mantiene una tendencia similar con cantidades casi iguales de turistas provenientes de cada uno de ellos. Sin embargo, en 2021 se observa una excepción, donde España se convierte en el país predominante con solo 4,573 turistas provenientes de Reino Unido. Esta anomalía se debe a la pandemia mundial de Covid-19, que tuvo un impacto significativo en los viajes internacionales y en el turismo en general durante el año 2020.

Estos datos resaltan la importancia de considerar factores externos, como la situación mundial y los eventos disruptivos, al analizar el número de turistas según su país de residencia.

df_2018	
	TURISTAS
PAIS_RESIDENCIA	
GBR826	10903
DEU276	6654
ESP724	6228
ZZZ900	2955
NLD528	1671
FRA250	1540
SWE752	1370
ITA380	1356
IRL372	1246
BEL056	1090
NOR578	939
DNK208	711
CHE756	621
FIN246	620

Figura 3.4: Numero de turistas encuestados segun su pais de residencia en 2018

df_2019	
	TURISTAS
PAIS_RESIDENCIA	
GBR826	10582
ESP724	7065
DEU276	6907
ZZZ900	2934
FRA250	1714
NLD528	1545
IRL372	1426
SWE752	1364
ITA380	1354
BEL056	1053
NOR578	903
DNK208	735
CHE756	592
FIN246	587

Figura 3.5: Numero de turistas encuestados segun su pais de residencia en 2019

df_2021	
	TURISTAS
PAIS_RESIDENCIA	
ESP724	10749
DEU276	5934
GBR826	4573
ZZZ900	3703
FRA250	2594
NLD528	1255
ITA380	1166
BEL056	966
SWE752	882
CHE756	756
IRL372	585
DNK208	484
NOR578	395
FIN246	221

Figura 3.6: Numero de turistas encuestados segun su pais de residencia en 2021

df_2022	
	TURISTAS
PAIS_RESIDENCIA	
GBR826	9580
ESP724	9090
DEU276	6132
ZZZ900	2965
FRA250	2095
NLD528	1896
IRL372	1553
ITA380	1263
BEL056	1164
SWE752	1105
NOR578	900
DNK208	792
CHE756	551
FIN246	501

Figura 3.7: Numero de turistas encuestados segun su pais de residencia en 2022

### 3.2.3. Número de turistas según su nacionalidad y el aeropuerto de entrada

En base a los datos obtenidos en los apartados anteriores, podemos realizar un análisis unificado. En este estudio, se representa el número de turistas según su país de residencia para cada aeropuerto y año. Dado que los aeropuertos de Tenerife y Gran Canaria son los que reciben la mayor cantidad de turistas, nos centraremos en estos dos aeropuertos. Además, nos enfocaremos en el año 2022, ya que en el intervalo de 2018 a 2021 los resultados son prácticamente iguales sin ninguna anomalía destacable.

En las figuras 3.8 y 3.9, se puede apreciar una clara diferencia entre estos dos aeropuertos predominantes. En el caso de Gran Canaria, la mayoría de los turistas nacionales, mientras que en el aeropuerto de Tenerife Sur hay una clara superioridad de turistas internacionales. Sin embargo, en ambos aeropuertos se observa que los tres países principales son Alemania, Reino Unido y España, como se había previsto.

Estos resultados confirman ciertos perfiles turísticos, donde Gran Canaria tiene una mayor afluencia de turistas nacionales, mientras que en el aeropuerto de Tenerife Sur hay una mayor presencia de turistas británicos. La presencia de Alemania en ambos aeropuertos también destaca como uno de los principales países emisores de turistas.

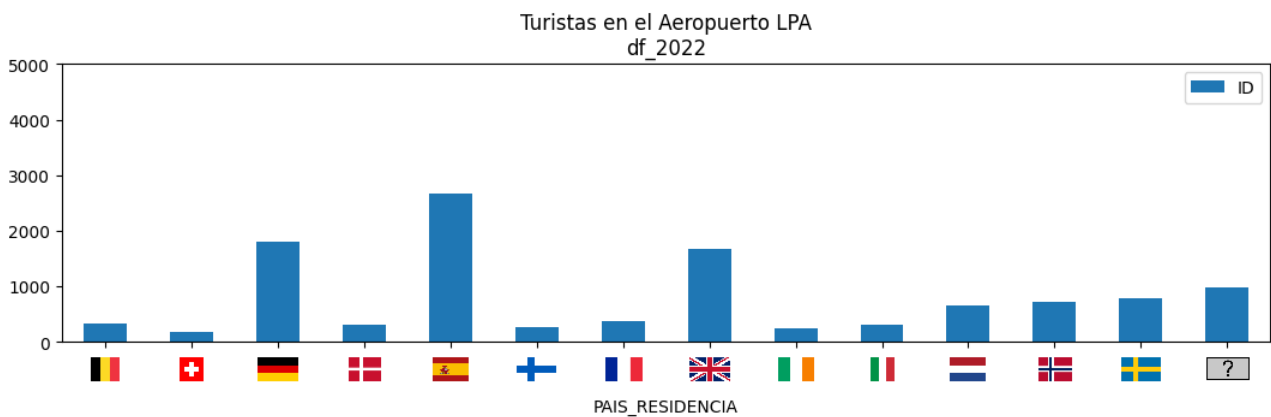


Figura 3.8: Datos formato tabla desde VSCode.

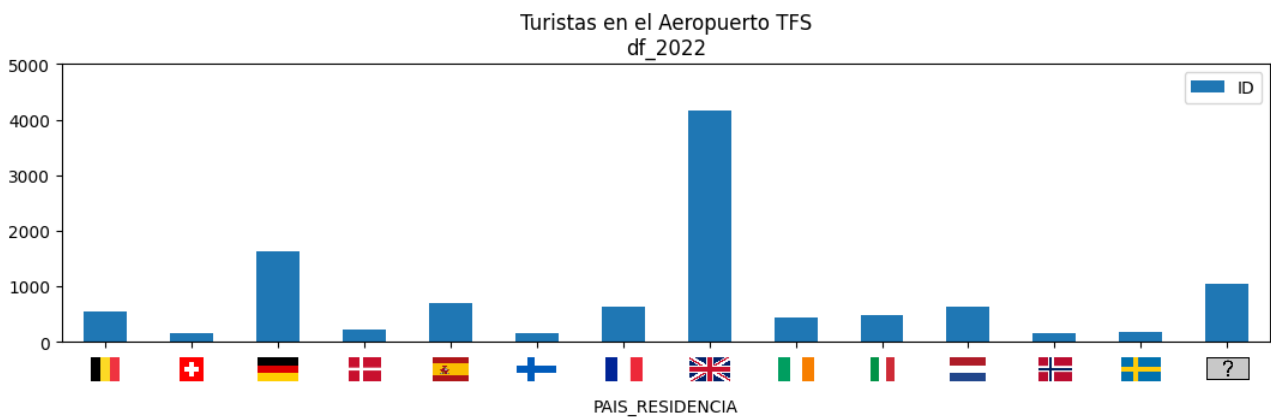


Figura 3.9: Datos formato tabla desde VSCode.

Como dato a destacar sobre los aeropuertos en Tenerife, se observa en la figura 3.10 que prácticamente el único país de origen de los turistas es España. Una de las razones de

esta situación se atribuye a que la zona turística de Tenerife se encuentra principalmente en el sur de la isla. Por lo tanto, es normal que a través del aeropuerto del sur ingresen principalmente turistas extranjeros, mientras que a través del aeropuerto del norte lleguen principalmente turistas nacionales de España. Pero también existe la posibilidad de turistas que vienen por el aeropuerto de Tenerife Sur a alojarse en el norte y viceversa.

Este hallazgo respalda la veracidad de los datos con los que estamos trabajando, ya que refleja la distribución geográfica y las características turísticas de la isla de Tenerife. La concentración de turistas extranjeros en el aeropuerto del sur y la predominancia de turistas nacionales en el aeropuerto del norte están en línea con las expectativas y la comprensión de la estructura turística de la isla.

Estos datos corroboran la validez y confiabilidad de la información con la que estamos trabajando, proporcionando un respaldo adicional a nuestros análisis y conclusiones sobre los patrones de turismo en Tenerife y su distribución según el país de residencia de los turistas.

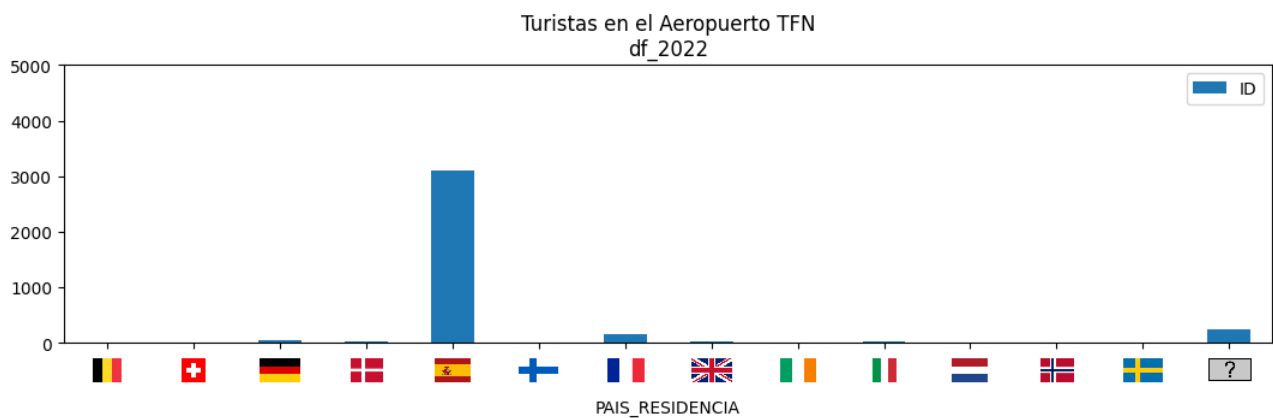


Figura 3.10: Datos formato tabla desde VSCode.

## **3.3. Gasto turístico**

### **3.3.1. Perfil por país de residencia por gasto separado por edades**

Con los perfiles turísticos que hemos creado anteriormente, tenemos una mejor comprensión del tipo de turistas que visitan Canarias. Uno de los aspectos más importantes a considerar es el gasto de estos turistas, se conoce como EGATUR. En la encuesta mencionada en la introducción, hay un apartado que recopila información sobre los gastos del grupo de viaje. Esta sección registra en qué categorías han gastado los turistas y en qué cantidad. Sin embargo, los turistas no proporcionan el gasto total, sino que especifican el gasto en subcategorías, como restaurantes, supermercados, entre otros. Por lo tanto, es necesario agrupar estos subapartados para obtener una visión más general.

Una vez que hayamos agrupado los tipos de gasto, calcularemos su media. La razón por la cual calculamos la media es para comprender mejor el perfil de gasto de cada país. Por ejemplo, aunque un gran porcentaje de turistas provenga de Reino Unido, es posible que cada turista gaste solo 500€ en promedio. En contraste, puede haber un país con pocos turistas entrantes, pero con un gasto medio de 2000€. Es importante considerar este perfil de turista al trabajar con el gasto medio de cada país. Además, para establecer un perfil más completo, hemos seleccionado las variables de país de residencia y edad. Con esto, podremos obtener un perfil más específico y detallado.

Si observamos las figuras del 3.11 al 3.14, todas coinciden en un punto: el mayor gasto de los turistas proviene principalmente de la categoría de ALIMENTACIÓN. Sin embargo, según el país y la edad, el segundo gasto más común varía entre OCIO, TRANSPORTE o COMPRA DE BIENES. Estos tres gastos suelen mantenerse constantes independientemente de la edad o el país. Además, es evidente que a medida que los turistas envejecen, tienden a gastar más, lo cual es comprensible debido a que los jóvenes generalmente no tienen un salario fijo ni ahorros. Por último, en cuanto a los países, se observa que la mayoría de los cinco países con un mayor gasto medio provienen de países con salarios más altos que España [12].

Estos análisis nos permiten identificar los patrones de gasto más comunes entre los turistas, destacando la importancia de la alimentación como el principal gasto y la influencia de la edad y el país de residencia en los patrones de gasto. Además, podemos apreciar la relación entre el nivel de ingresos de los países de origen y el gasto medio de los turistas. Estos insights nos ayudarán a comprender mejor el comportamiento de gasto de los turistas y a orientar estrategias para promover el turismo y maximizar los beneficios económicos en Canarias.

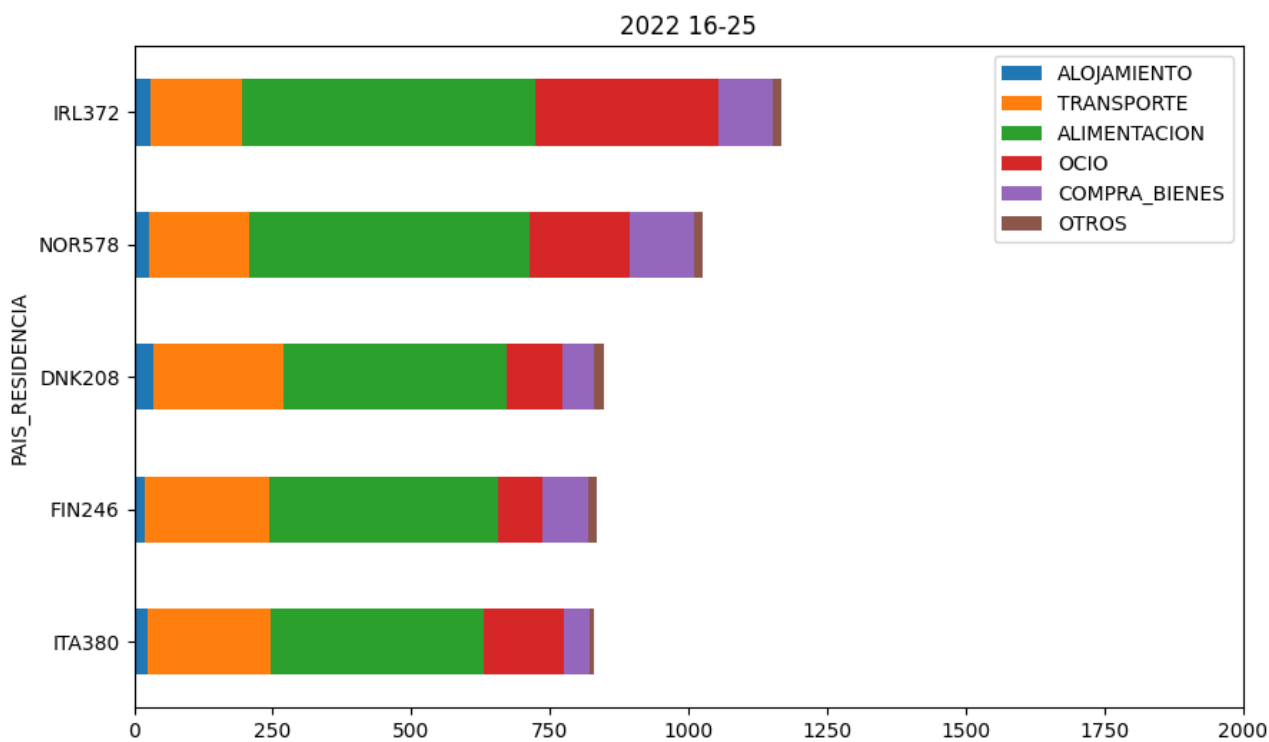


Figura 3.11: Gasto medio en el año 2022 entre 16 y 25 años

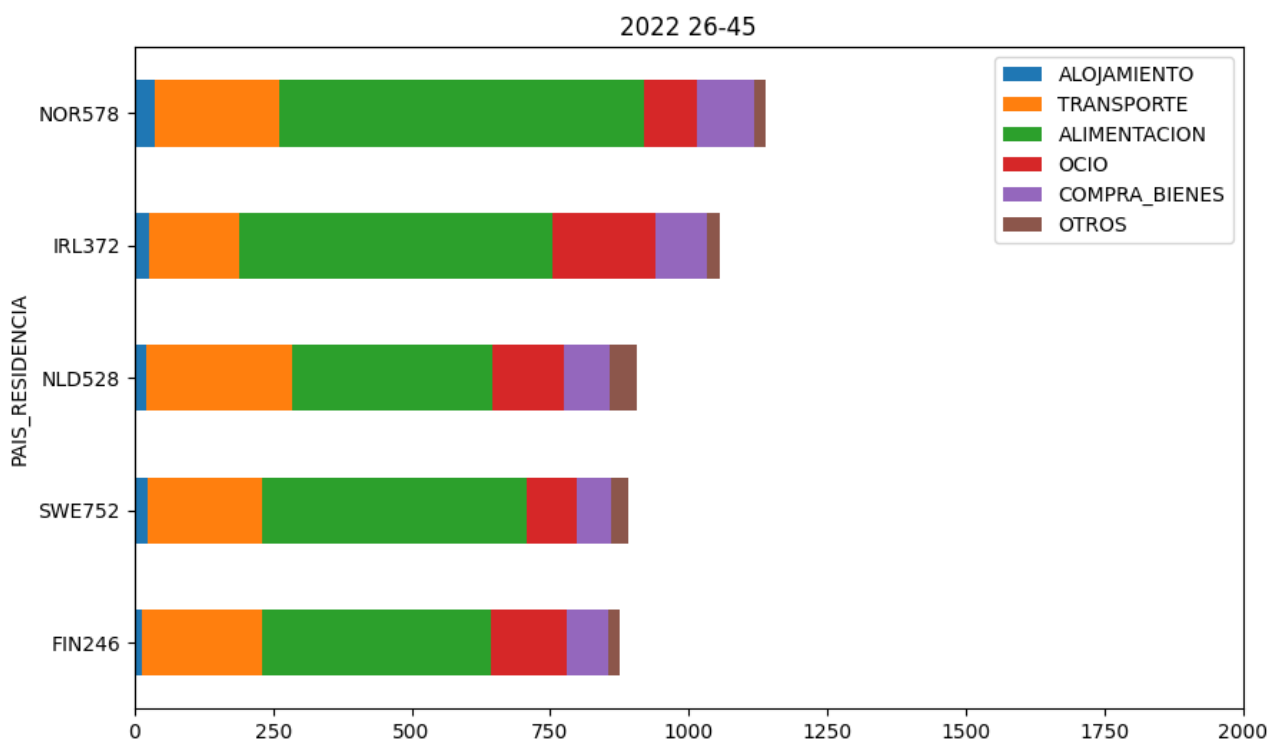


Figura 3.12: Gasto medio en el año 2022 entre 26 y 45 años



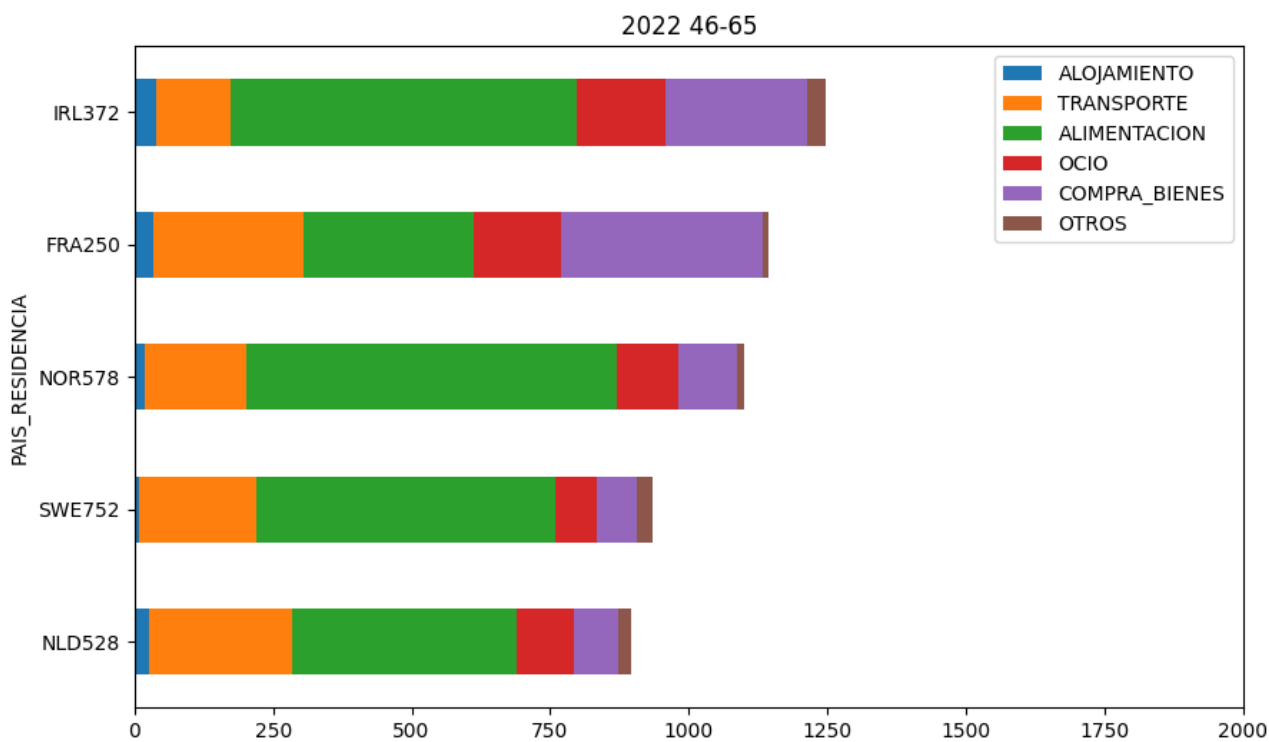


Figura 3.13: Gasto medio en el año 2022 entre 45 y 65 años

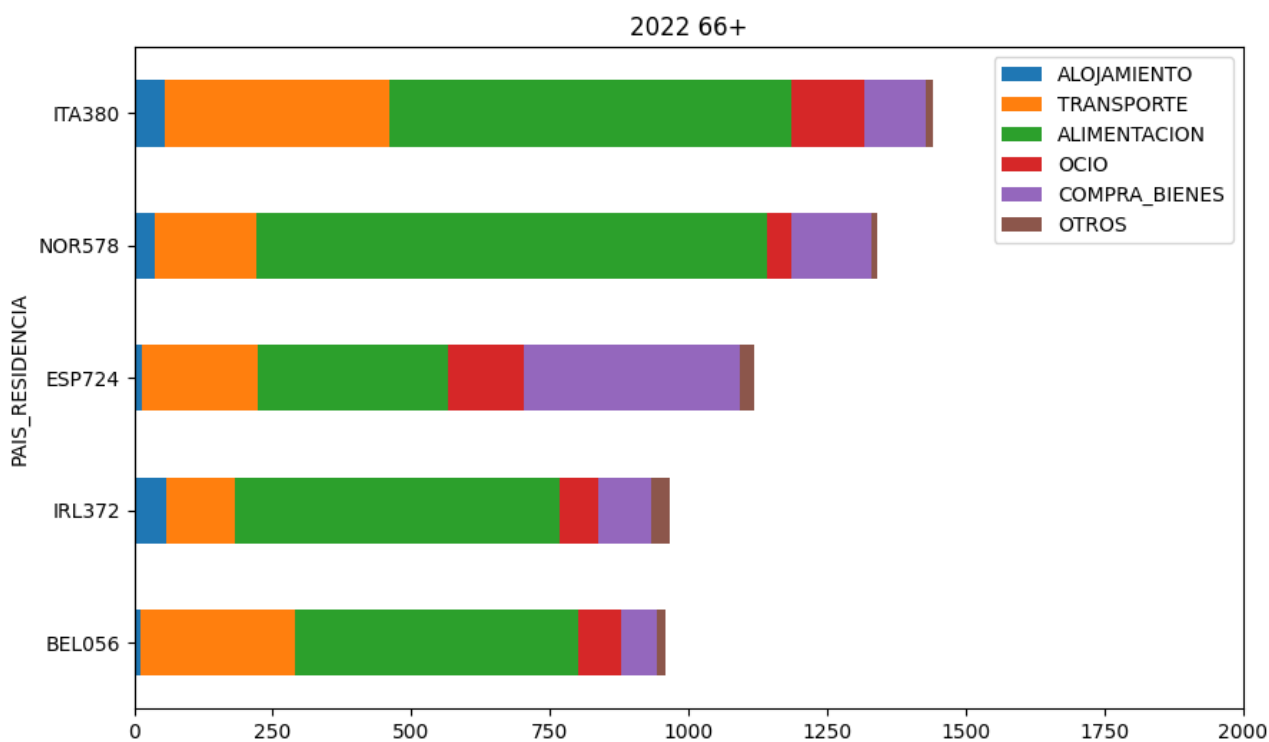


Figura 3.14: Gasto medio en el año 2022 de mas de 66 años

### 3.3.2. Sankey Diagram aplicado a los turistas

Sankey Diagram es una poderosa herramienta de visualización que permite mostrar las conexiones entre diferentes grupos de datos. Sin embargo, al analizar las figuras del 3.15 al 3.17, podemos observar que la visualización en papel no es adecuada para apreciar completamente las complejas redes de conexiones presentes en los datos. Es importante comprender que las Sankey Diagrams son especialmente útiles cuando las conexiones son manejables y se pueden seguir con claridad. Sin embargo, cuando hay una gran cantidad de conexiones, se vuelve difícil identificar y comprender los patrones deseados.

A pesar de los esfuerzos realizados para representar los datos en las figuras proporcionadas, la complejidad y cantidad de conexiones presentes en los datos dificultan su representación de manera clara y concisa. Es crucial reconocer las limitaciones de las herramientas de visualización y adaptarlas según las características y complejidad de los datos analizados. En este caso, puede ser necesario explorar otros enfoques visuales o técnicas analíticas que permitan una mejor identificación y comprensión de las conexiones y patrones presentes en los datos.

Es fundamental recordar que el objetivo principal de cualquier herramienta de visualización es facilitar la comprensión de los datos y obtener información significativa. Por lo tanto, es recomendable evaluar diferentes opciones y seleccionar la técnica visual más apropiada que permita revelar las relaciones y tendencias relevantes en los datos de manera efectiva. Algo que vamos a atribuir a la siguiente sección de este capítulo.

Diagrama de Sankey

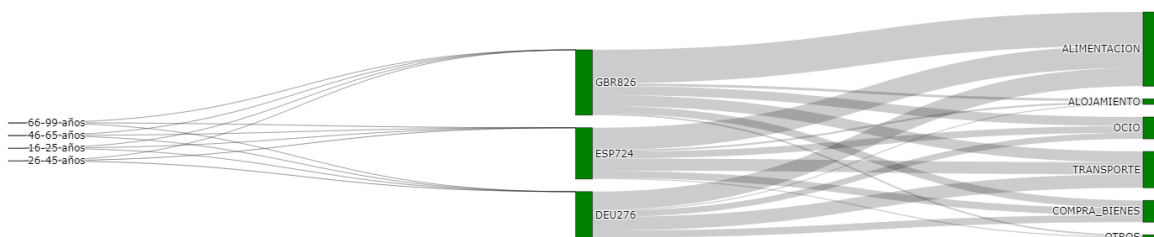


Figura 3.15: Diagrama Sankey. Grupo Edades, país y tipo de gasto

Diagrama de Sankey

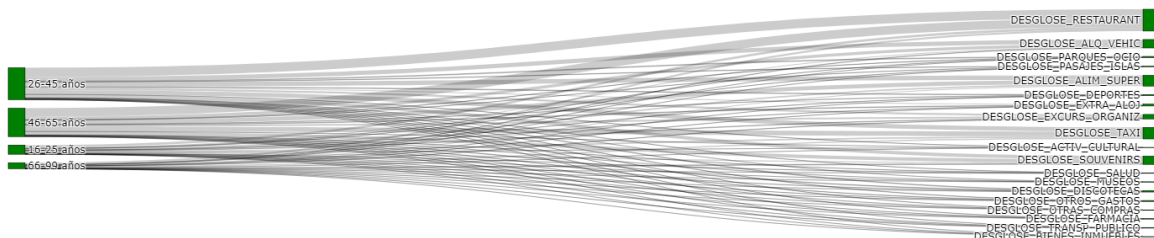


Figura 3.16: Diagrama Sankey. Grupo edades y desglose de gastos

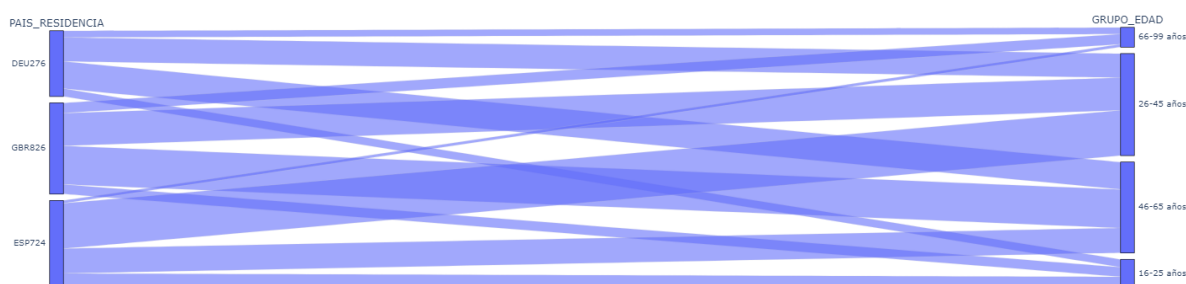


Figura 3.17: Diagrama Sankey. Pais residencia y grupos de edades

### 3.3.3. Gráfico Sunburst por año, edad, país de residencia y gasto

Es una elección acertada utilizar un gráfico Sunburst para visualizar el tipo de gasto, su desglose, el país del turista, la edad y el año. El Sunburst es un gráfico circular que permite agrupar y representar estas variables de manera efectiva. Es una de las posibles alternativas al no poder visualizar los datos con el gráfico Sankey Diagram.

El gráfico Sunburst ofrece múltiples ventajas para nuestra investigación, aunque también presenta algunos inconvenientes. En la figura 3.18, donde se muestra este tipo de gráfico, podemos observar todas las variables mencionadas anteriormente organizadas por años. Con este gráfico, podemos enfocarnos en el aspecto que estamos investigando y obtener información detallada. Además, debido a que el gráfico es interactivo, podemos interactuar con él y seleccionar un año específico, un país en particular y una edad concreta para explorar los datos en mayor detalle.

Sin embargo, como se muestra en la figura, si se muestra el gráfico completo, puede resultar casi irreconocible debido a la gran cantidad de variables que se están teniendo en cuenta. Esto puede dificultar la interpretación y comprensión de los datos, especialmente cuando hay muchas categorías y subcategorías involucradas.

En la figura 3.21, se ha seleccionado en el gráfico Sunburst el año 2022, turistas con edades entre 26 y 45 años y que tienen como país de residencia España. Esta selección nos permite enfocarnos en un grupo específico de turistas y explorar los patrones y desgloses de gasto dentro de ese subconjunto de datos.

Al realizar esta selección, podemos observar cómo se distribuyen los diferentes tipos de gasto dentro de este grupo en particular. El gráfico Sunburst nos permite ver las categorías principales de gasto, así como las subcategorías, lo que nos brinda una visión detallada del comportamiento de gasto de los turistas seleccionados.

Esta capacidad de interacción y filtrado en el gráfico Sunburst nos permite explorar diferentes combinaciones de variables y analizar cómo se relacionan entre sí. En este caso, al seleccionar el año, rango de edad y país de residencia específicos, podemos obtener información valiosa sobre los patrones de gasto de los turistas españoles de ciertas edades en el año 2022.

Desde la figura 3.18 hasta la 3.21 podemos ver una progresión escalada interactuando con el gráfico sunburst hasta llegar al nivel deseado.

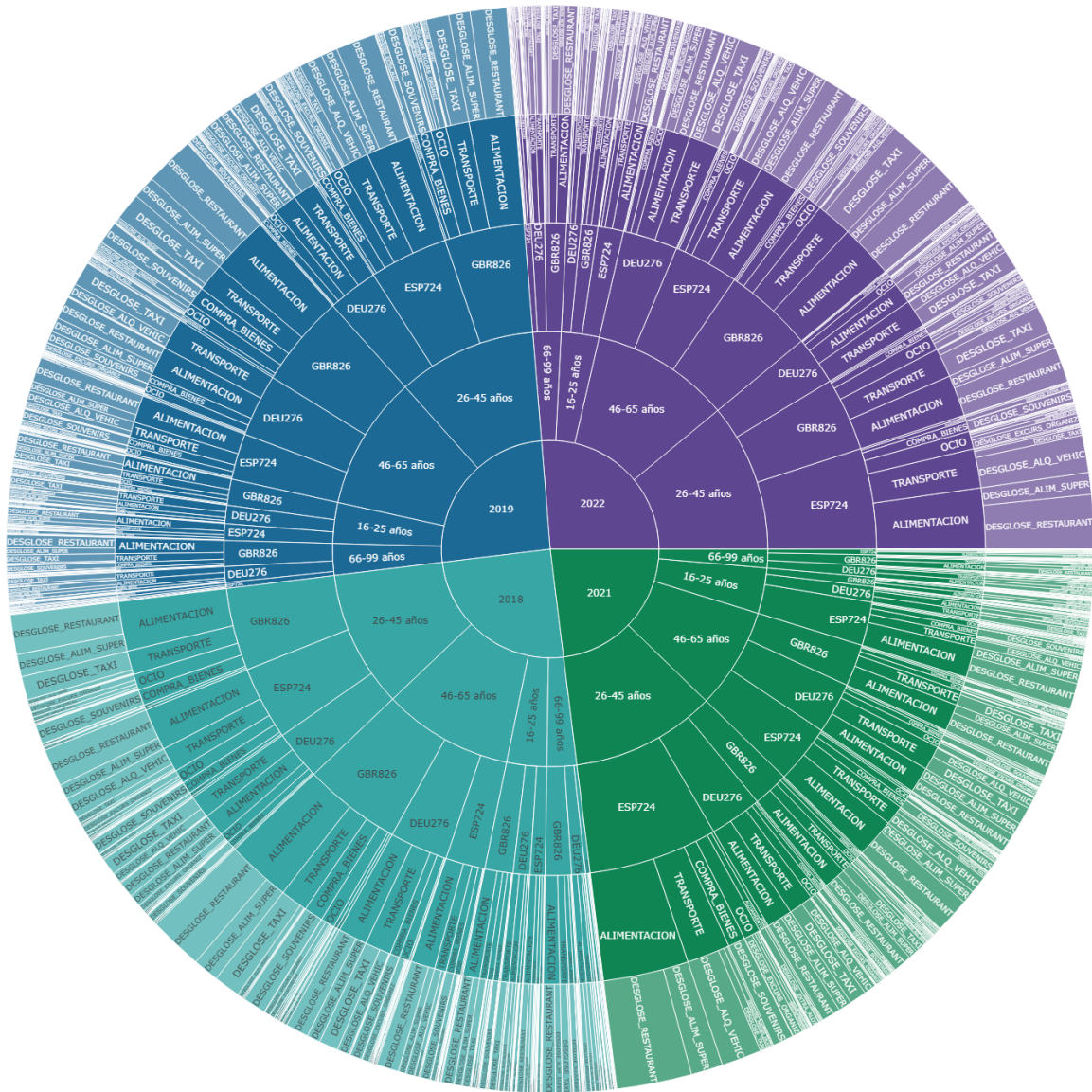


Figura 3.18: Gráfico Sunburst Total





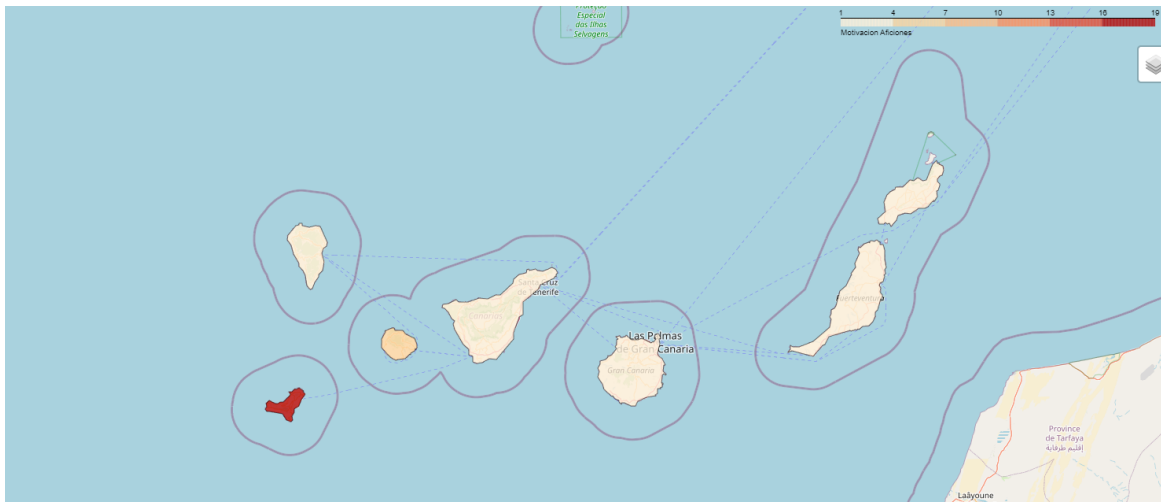


Figura 3.22: Mapa de calor sobre la media en Aficiones

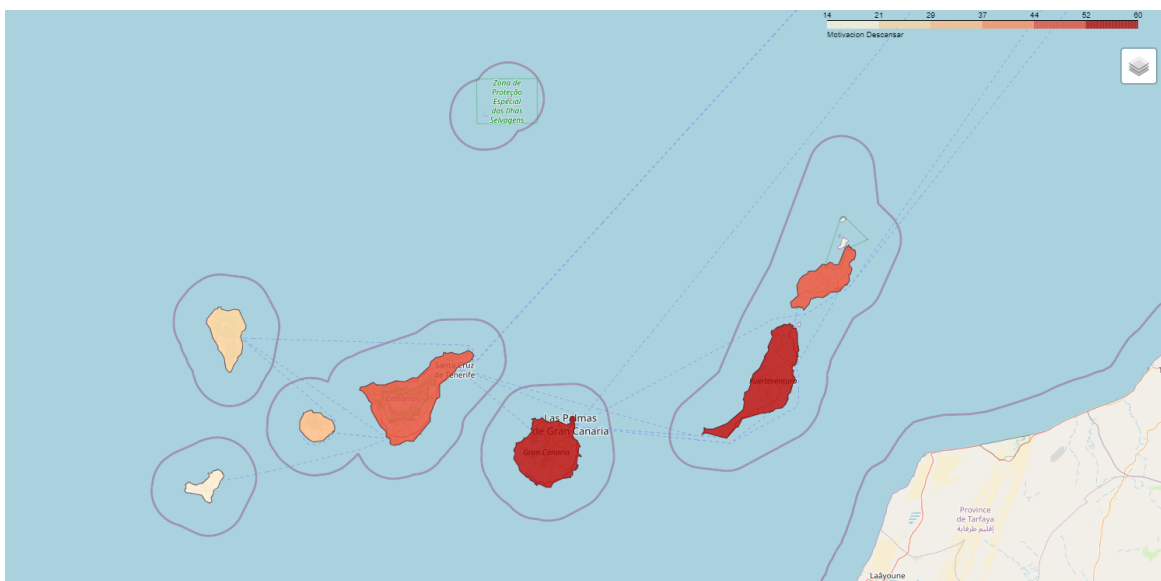


Figura 3.23: Mapa de calor sobre la media en Descansar

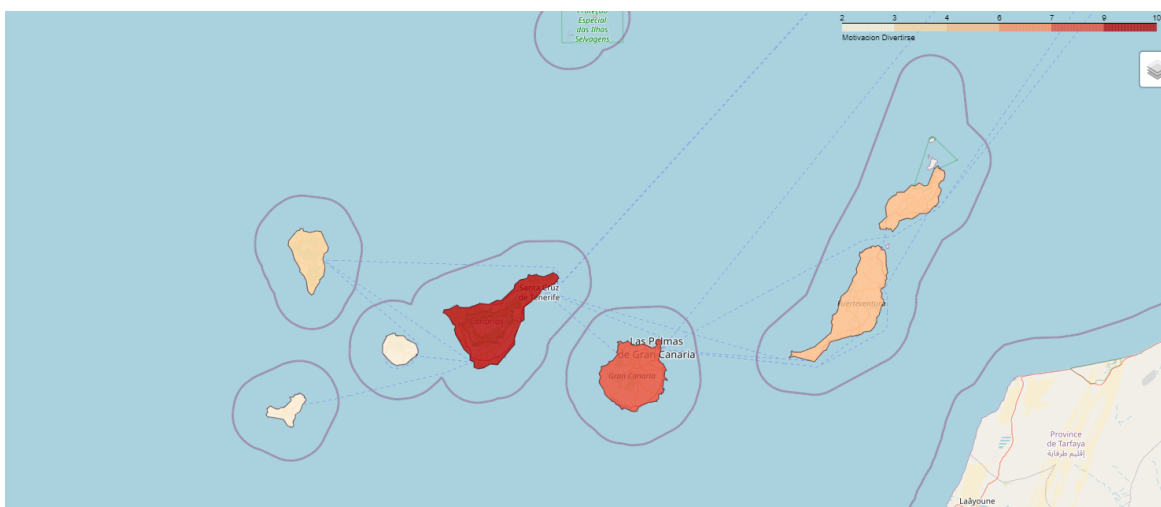


Figura 3.24: Mapa de calor sobre la media en Divertirse

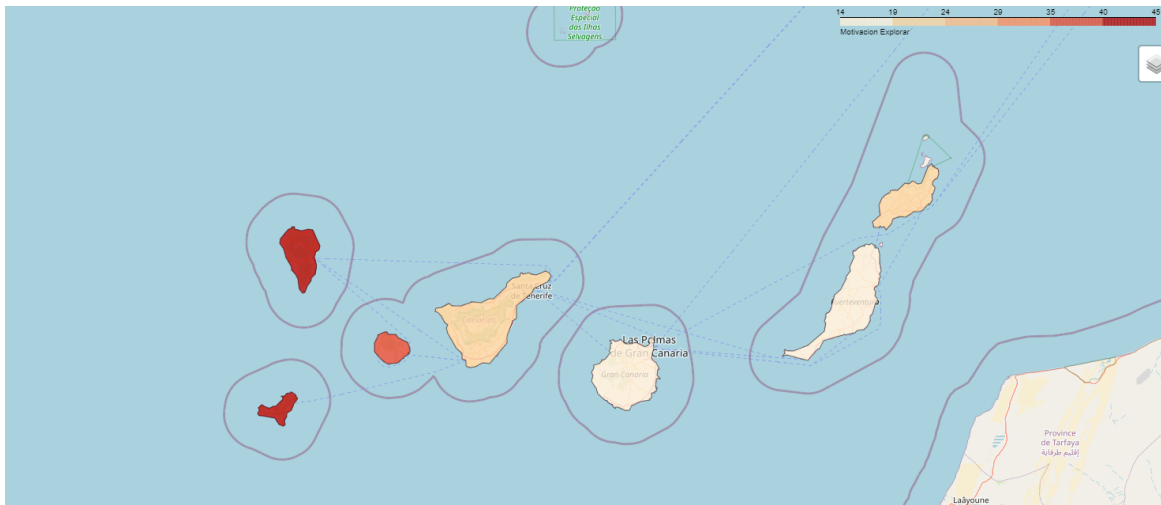


Figura 3.25: Mapa de calor sobre la media en Explorar

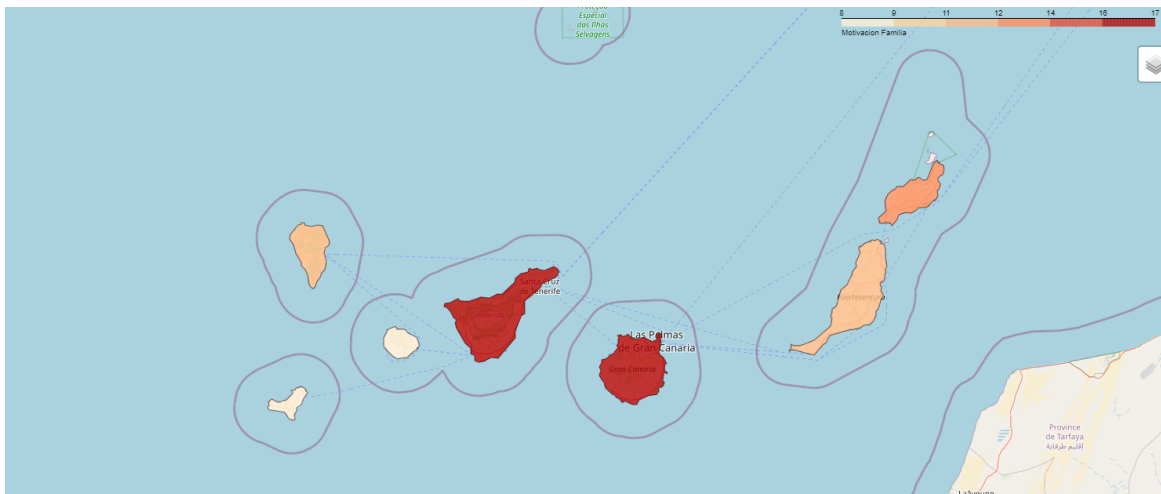


Figura 3.26: Mapa de calor sobre la media en Familia



### 3.4.1. Motivación principal de cada isla

Gracias a los datos obtenidos en el apartado anterior, también podemos destacar cuál es la motivación principal de cada isla. Al tener la media de cada motivación en el análisis anterior, simplemente debemos seleccionar la motivación principal para cada isla y representarla en un mapa.

Si observamos la figura 3.27, que muestra las islas menos desarrolladas turísticamente como La Palma, El Hierro y La Gomera, podemos ver que la motivación principal de los turistas que visitan estas islas es la exploración de las mismas. Por otro lado, en el resto de las islas se observa que los turistas prefieren visitarlas para descansar.

Estos hallazgos nos brindan una comprensión clara de las motivaciones principales de los turistas en cada isla. Nos permite apreciar que las islas menos desarrolladas turísticamente, con un enfoque más natural y de exploración, atraen a turistas que buscan experiencias más auténticas y contacto con la naturaleza. Mientras tanto, las islas más desarrolladas turísticamente, con una infraestructura y servicios más desarrollados, atraen a turistas que buscan un descanso y relajación en sus vacaciones.

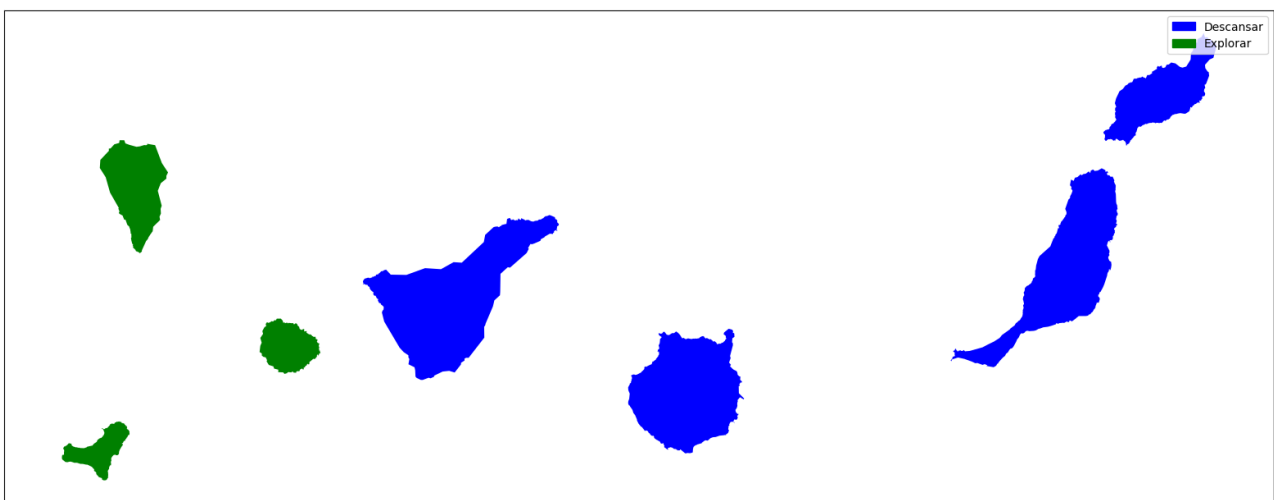


Figura 3.27: Motivación principal de cada isla

# Capítulo 4

## Conclusiones y futuras líneas de trabajo

Este proyecto de finalización de carrera se ha centrado en el estudio y análisis de datos en el contexto del turismo en Canarias. A través de diversas técnicas estadísticas, visualizaciones gráficas y herramientas como el gráfico Sunburst, se ha obtenido una comprensión profunda de los patrones de gasto, motivaciones de los turistas y características demográficas.

Se ha destacado la importancia de variables como el país de residencia, la edad y el año en el comportamiento de los turistas, identificando los países principales de origen, los gastos más comunes y las motivaciones principales para visitar las islas. Asimismo, se ha observado cómo las islas menos desarrolladas turísticamente difieren en cuanto a las motivaciones de los turistas en comparación con las islas más populares.

Se ha reconocido la limitación de ciertas herramientas de visualización, como el Sankey Diagram, cuando se manejan conjuntos de datos complejos con múltiples conexiones. En cambio, se ha resaltado la utilidad del gráfico Sunburst para representar de manera efectiva el desglose de gastos, motivaciones, países de residencia y edades.

Finalmente, se ha planteado la siguiente etapa del proyecto, que implica la implementación de técnicas de Machine Learning e Inteligencia Artificial para un análisis más exhaustivo de los datos, como la predicción de comportamientos de gasto y la extracción de información a partir de comentarios y opiniones de los turistas.

En conjunto, este proyecto ha proporcionado una visión completa y detallada del turismo en Canarias, permitiendo tomar decisiones informadas y diseñar estrategias más efectivas para promover el turismo y optimizar los beneficios económicos en la región.

### 4.1. Futuras Líneas de trabajo

Es una excelente propuesta ampliar el análisis de los datos utilizando herramientas de Machine Learning e Inteligencia Artificial. Estas tecnologías pueden complementar y enriquecer el estudio de los datos, brindando un análisis más exhaustivo y permitiendo descubrir patrones, relaciones y predicciones adicionales.

Con el uso de Machine Learning, se pueden aplicar algoritmos y modelos para realizar tareas como clasificación, regresión, clustering o detección de anomalías. Esto permitiría identificar segmentos de turistas, predecir comportamientos de gasto, analizar tendencias y realizar recomendaciones personalizadas basadas en los datos disponibles.

Además, la utilización de técnicas de Inteligencia Artificial, como el procesamiento del lenguaje natural, puede ayudar a extraer información relevante de los textos abiertos proporcionados en la encuesta, como comentarios y opiniones de los turistas. Esto

permitiría obtener una comprensión más profunda de sus preferencias y experiencias.

Es importante destacar que la implementación de estas herramientas requiere un enfoque metodológico adecuado, incluyendo la preparación y limpieza de los datos, la selección de modelos apropiados, el entrenamiento y la evaluación. También es esencial considerar aspectos éticos y de privacidad en el manejo de los datos.

Con la integración de Machine Learning e Inteligencia Artificial en el proyecto, se puede obtener una visión más completa y detallada de los datos, lo que permitirá tomar decisiones más informadas y diseñar estrategias más efectivas en el ámbito del turismo en Canarias.

# Capítulo 5

## Summary and Conclusions

This final-year project has focused on the study and analysis of data in the context of tourism in the Canary Islands. Through various statistical techniques, graphical visualizations, and tools like the Sunburst chart, a deep understanding of spending patterns, tourist motivations, and demographic characteristics has been achieved.

The importance of variables such as country of residence, age, and year in tourist behavior has been highlighted, identifying the main countries of origin, common expenditures, and primary motivations for visiting the islands. Additionally, it has been observed how less-developed tourist destinations differ in terms of tourist motivations compared to the more popular islands.

The limitations of certain visualization tools, such as the Sankey Diagram, when dealing with complex datasets with multiple connections have been acknowledged. Instead, the usefulness of the Sunburst chart for effectively representing expenditure breakdowns, motivations, country of residence, and age groups has been emphasized.

Lastly, the next stage of the project has been proposed, involving the implementation of Machine Learning and Artificial Intelligence techniques for a more comprehensive analysis of the data, such as predicting spending behaviors and extracting information from tourist comments and reviews.

Overall, this project has provided a comprehensive and detailed insight into tourism in the Canary Islands, enabling informed decision-making and the design of more effective strategies to promote tourism and optimize economic benefits in the region.

# Capítulo 6

## Presupuesto

En el capítulo de presupuesto de nuestro proyecto, recopilaremos las diferentes tareas realizadas junto con sus respectivos costos. Estos costos han sido calculados teniendo en cuenta el número de horas y el esfuerzo invertido en cada tarea. De esta manera, podremos tener una visión clara de los recursos necesarios y asignar de manera adecuada los fondos disponibles para llevar a cabo el proyecto de manera efectiva.

<b>Tareas</b>	<b>Horas</b>	<b>Coste</b>
Obtención de datos	0	0€
Limpieza y exploración de datos	30	225€
Análisis en profundidad de los datos	100	250€
Visualización de los datos	150	700€
Interpretación de los datos	20	50€
<b>Total</b>	300	1200€

Tabla 6.1: Coste del desarrollo

# Apéndice A

## Algoritmos del Gasto Turístico

En este apéndice se incluirá el código completo de alguna de las gráficas realizadas en el capítulo de Gasto Turístico.

### A.1. Algoritmo de perfil por país de residencia por gasto separado por edades

```
# Juntar cada concepto con sus respectivas columnas
alojamientoDesglose = ["DESGLOSE_EXTRA_ALOJ"]
transporteDesglose = ["DESGLOSE_PASAJES_ISLAS", "DESGLOSE_TAXI", "
    DESGLOSE_ALQ_VEHIC", "DESGLOSE_TRANSP_PUBLICO"]
alimentacionDesglose = ["DESGLOSE ALIM_SUPER", "DESGLOSE_RESTAURANT"]
ocioDesglose = ["DESGLOSE_EXCURS_ORGANIZ", "DESGLOSE_DEPORTES", "
    DESGLOSE_ACTIV_CULTURAL", "DESGLOSE_MUSEOS", "DESGLOSE_PARQUES_OCIO", "
    DESGLOSE_DISCOTECAS", "DESGLOSE_SALUD"]
compradebienesDesglose = ["DESGLOSE_SOUVENIRS", "DESGLOSE_BIENES_INMUEBLES", "
    DESGLOSE_OTRAS_COMPRAS"]
otrosDesglose = ["DESGLOSE_FARMACIA", "DESGLOSE_OTROS_GASTOS"]

allDesglose = ['PAIS_RESIDENCIA', 'GASTO_EUROS', 'DESGLOSE_EXTRA_ALOJ', '
    DESGLOSE_PASAJES_ISLAS', 'DESGLOSE_TAXI', 'DESGLOSE_ALQ_VEHIC', '
    DESGLOSE_TRANSP_PUBLICO', 'DESGLOSE ALIM_SUPER', 'DESGLOSE_RESTAURANT', '
    DESGLOSE_EXCURS_ORGANIZ', 'DESGLOSE_DEPORTES', 'DESGLOSE_ACTIV_CULTURAL', '
    DESGLOSE_MUSEOS', 'DESGLOSE_PARQUES_OCIO', 'DESGLOSE_DISCOTECAS', '
    DESGLOSE_SALUD', 'DESGLOSE_SOUVENIRS', 'DESGLOSE_BIENES_INMUEBLES', '
    DESGLOSE_OTRAS_COMPRAS',
    'DESGLOSE_FARMACIA', 'DESGLOSE_OTROS_GASTOS']
# separar por rangos de edades
edades_16_25 = []
edades_26_45 = []
edades_46_65 = []
edades_66_99 = []

for archivo_csv in df_list:
    edades_16_25.append(pd.DataFrame(archivo_csv[(archivo_csv["EDAD"] >= 16) & (
        archivo_csv["EDAD"] <= 25)]))
```

```

edades_26_45.append(pd.DataFrame(archivo_csv[(archivo_csv["EDAD"] >= 26) & (
    archivo_csv["EDAD"] <= 45)]["allDesglose"]))
edades_46_65.append(pd.DataFrame(archivo_csv[(archivo_csv["EDAD"] >= 46) & (
    archivo_csv["EDAD"] <= 65)]["allDesglose"]))
edades_66_99.append(pd.DataFrame(archivo_csv[(archivo_csv["EDAD"] >= 66) & (
    archivo_csv["EDAD"] <= 99)]["allDesglose"]))

main_edades = [edades_16_25, edades_26_45, edades_46_65, edades_66_99]

for idx, edad in enumerate(main_edades):
    for idx1, dataframe in enumerate(edad):
        dataframe = dataframe["allDesglose"].replace(',', '.', regex=True)
        edad[idx1] = pd.DataFrame(dataframe)
    main_edades[idx] = edad

for idx, edades_grupo in enumerate(main_edades):
    for idx1, df_anio in enumerate(edades_grupo):
        df_anio['ALOJAMIENTO'] = df_anio.DESGLOSE_EXTRA_ALOJ.astype(float) *
            df_anio.GASTO_EUROS.astype(float) / 100
        df_anio['TRANSPORTE'] = (df_anio.DESGLOSE_PASAJES_ISLAS.astype(float) +
            df_anio.DESGLOSE_TAXI.astype(float) +
            df_anio.DESGLOSE_TRANSP_PUBLICO.astype(float)
            + df_anio.DESGLOSE_ALQ_VEHIC.astype(float)
            ) * df_anio.GASTO_EUROS.astype(float) /
            100
        df_anio['ALIMENTACION'] = (df_anio.DESGLOSE_ALIM_SUPER.astype(float) +
            df_anio.DESGLOSE_RESTAURANT.astype(float)) * df_anio.GASTO_EUROS.
            astype(float) / 100
        df_anio['OCIO'] = (df_anio.DESGLOSE_EXCURS_ORGANIZ.astype(float) +
            df_anio.DESGLOSE_DEPORTES.astype(float) + df_anio.
            DESGLOSE_ACTIV_CULTURAL.astype(float) +
            df_anio.DESGLOSE_MUSEOS.astype(float) + df_anio.
            DESGLOSE_PARQUES_OCIO.astype(float) + df_anio.
            DESGLOSE_DISCOTECAS.astype(float) +
            df_anio.DESGLOSE_SALUD.astype(float)) * df_anio.
            GASTO_EUROS.astype(float) / 100
        df_anio['COMPRA_BIENES'] = (df_anio.DESGLOSE_SOUVENIRS.astype(float) +
            df_anio.DESGLOSE_BIENES_INMUEBLES.astype(float) +
            df_anio.DESGLOSE_OTRAS_COMPRAS.astype(
                float)) * df_anio.GASTO_EUROS.astype(
                float) / 100
        df_anio['OTROS'] = (df_anio.DESGLOSE_FARMACIA.astype(float) + df_anio.
            DESGLOSE_OTROS_GASTOS.astype(float)) * df_anio.GASTO_EUROS.astype(
                float) / 100
        df_anio['TOTAL'] = df_anio['ALOJAMIENTO'] + df_anio['TRANSPORTE'] +
            df_anio['ALIMENTACION'] + df_anio['OCIO'] + df_anio['COMPRA_BIENES'] +
            df_anio['OTROS']
        edades_grupo[idx1] = df_anio
    main_edades[idx] = edades_grupo

```

```

new_columns = ['PAIS_RESIDENCIA', 'ALOJAMIENTO', 'TRANSPORTE', 'ALIMENTACION', '
OCIO', 'COMPRA_BIENES', 'OTROS', 'TOTAL']

for idx, edades_grupo in enumerate(main_edades):
    for idx1, df_anio in enumerate(edades_grupo):
        edades_grupo[idx1] = edades_grupo[idx1][new_columns]
    main_edades[idx] = edades_grupo

for idx, edades in enumerate(main_edades):
    for idx1, anio in enumerate(edades):

        anio = anio.groupby(['PAIS_RESIDENCIA']).mean()
        edades[idx1] = anio
    main_edades[idx] = edades

edades = ["16-25", "26-45", "46-65", "66+"]
anio = ["2018", "2019", "2021", "2022"]

for x in range(len(edades)):
    for y in range(len(anio)):
        main_edades[x][y] = main_edades[x][y].drop('ZZZ900')
        main_edades[x][y] = main_edades[x][y].sort_values(by=['TOTAL']).tail(5).
            drop(columns=['TOTAL'])
        ax = main_edades[x][y].plot.barh(stacked=True, title=anio[y] + " " +
            edades[x], figsize=(10, 6))
        ax.set_xlim([0,2000])

```

## A.2. Algoritmo para el gráfico Sunburst

```

allDesglose = ['PAIS_RESIDENCIA', 'EDAD', 'TRIMESTRE', 'GASTO_EUROS', '
DESGLOSE_EXTRA_ALOJ', 'DESGLOSE_PASAJES_ISLAS', 'DESGLOSE_TAXI',
    'DESGLOSE_ALQ_VEHIC', 'DESGLOSE_TRANSP_PUBLICO', '
    DESGLOSE ALIM_SUPER',
    'DESGLOSE_RESTAURANT', 'DESGLOSE_EXCURS_ORGANIZ', '
    DESGLOSE_DEPORTES',
    'DESGLOSE_ACTIV_CULTURAL', 'DESGLOSE_MUSEOS', '
    DESGLOSE_PARQUES_OCIO',
    'DESGLOSE_DISCOTECAS', 'DESGLOSE_SALUD', '
    DESGLOSE_SOUVENIRS',
    'DESGLOSE_BIENES_INMUEBLES', 'DESGLOSE_OTRAS_COMPRAS',
    'DESGLOSE_FARMACIA', 'DESGLOSE_OTROS_GASTOS']

copy_df_total = df_total.copy()

copy_df_total = copy_df_total[allDesglose].replace(',', '.', regex=True)

copy_df_total['ALOJAMIENTO'] = copy_df_total.DESGLOSE_EXTRA_ALOJ.astype(float) *
    copy_df_total.GASTO_EUROS.astype(float) / 100
copy_df_total['TRANSPORTE'] = (copy_df_total.DESGLOSE_PASAJES_ISLAS.astype(float)

```



```

+ copy_df_total.DESGLOSE_TAXI.astype(float) +
                                copy_df_total.DESGLOSE_TRANSP_PUBLICO.astype(
                                    float) + copy_df_total.DESGLOSE_ALQ_VEHIC.
                                    astype(float)) * copy_df_total.GASTO_EUROS
                                    .astype(float) / 100
copy_df_total['ALIMENTACION'] = (copy_df_total.DESGLOSE_ALIM_SUPER.astype(float)
+ copy_df_total.DESGLOSE_RESTAURANT.astype(float)) * copy_df_total.GASTO_EUROS
.astype(float) / 100
copy_df_total['OCIO'] = (copy_df_total.DESGLOSE_EXCURS_ORGANIZ.astype(float) +
copy_df_total.DESGLOSE_DEPORTES.astype(float) + copy_df_total.
DESGLOSE_ACTIV_CULTURAL.astype(float) +
                                copy_df_total.DESGLOSE_MUSEOS.astype(float) +
                                copy_df_total.DESGLOSE_PARQUES_OCIO.astype(
                                    float) + copy_df_total.DESGLOSE_DISCOTECAS.
                                    astype(float) +
                                copy_df_total.DESGLOSE_SALUD.astype(float)) *
                                copy_df_total.GASTO_EUROS.astype(float) / 100
copy_df_total['COMPRA_BIENES'] = (copy_df_total.DESGLOSE_SOUVENIRS.astype(float)
+ copy_df_total.DESGLOSE_BIENES_INMUEBLES.astype(float) +
                                copy_df_total.DESGLOSE_OTRAS_COMPRAS.
                                astype(float)) * copy_df_total.
                                GASTO_EUROS.astype(float) / 100
copy_df_total['OTROS'] = (copy_df_total.DESGLOSE_FARMACIA.astype(float) +
copy_df_total.DESGLOSE_OTROS_GASTOS.astype(float)) * copy_df_total.GASTO_EUROS
.astype(float) / 100
copy_df_total['TOTAL'] = copy_df_total['ALOJAMIENTO'] + copy_df_total['
TRANSPORTE'] + copy_df_total['ALIMENTACION'] + copy_df_total['OCIO'] +
copy_df_total['COMPRA_BIENES'] + copy_df_total['OTROS']

copy_df_total['TRIMESTRE'] = copy_df_total['TRIMESTRE'].apply(lambda x: x[:4])
copy_df_total = copy_df_total.drop(copy_df_total[copy_df_total.PAIS_RESIDENCIA ==
'ZZZ900'].index)

def group_age(age):
    if age <= 25:
        return '16-25 anios'
    elif age <= 45:
        return '26-45 anios'
    elif age <= 65:
        return '46-65 anios'
    else:
        return '66-99 anios'

copy_df_total['GRUPO_EDAD'] = copy_df_total['EDAD'].apply(group_age)
copy_df_total.dropna(subset=['ALOJAMIENTO'], inplace=True)

top_10_paises = copy_df_total.groupby('PAIS_RESIDENCIA')['TOTAL'].sum().
    sort_values(ascending=False).head(3).index.tolist()
mask = copy_df_total['PAIS_RESIDENCIA'].isin(top_10_paises)
nuevo_df = copy_df_total[mask]

```

```

df_melted = pd.melt(nuevo_df, id_vars=['TRIMESTRE', 'GRUPO_EDAD', '
    PAIS_RESIDENCIA'], value_vars=['ALOJAMIENTO', 'TRANSPORTE', 'OCIO', '
    COMPRA_BIENES', 'OTROS', 'ALIMENTACION'], var_name='TIPO', value_name='GASTO')

fig = px.sunburst(df_melted,
    path=['TRIMESTRE', 'GRUPO_EDAD', 'PAIS_RESIDENCIA', 'TIPO'],
    values='GASTO',
    title='Gasto en alojamiento por año, grupo de edad, país de
    residencia y tipo de alojamiento',
    color_discrete_sequence=px.colors.qualitative.Prism)

fig.update_layout(
    width=1500, # Ancho de la figura en pixeles
    height=1800 # Altura de la figura en pixeles
)
fig.show()

```

# Bibliografía

- [1] Instituto Canario de Estadística. Available at: <http://www.gobiernodecanarias.org/istac/>. (Accessed 14 de julio de 2023).
- [2] IMPACTUR Canarias. Available at: <https://www.exceltur.org/wp-content/uploads/2022/02/IMPACTUR-Canarias-2020.pdf>. (Accessed 14 de julio de 2023)
- [3] RTVC. Radio Televisión Canaria. Available at: <https://rtvc.es/el-empleo-turistico-aumenta-en-canarias-un-295-en-el-primer-trimestre/#:~:text=El%20n%20mero%20de%20ocupados%20en,de%20Poblaci%20n%20Activa%20>.
- [4] Aplicaciones del Big Data al sector Turístico Español. Available at: <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/33057/TFG-Tello%20MartAnez%2C%20Javier%20MarAa.pdf?sequence=1> (Accessed 14 de julio de 2023)
- [5] Incremento de turistas en Canarias. Available at: <http://www.gobiernodecanarias.org/istac/estadisticas/sectorservicios/hosteleriayturismo/demanda/E16028B.html>. (Accessed 14 de julio de 2023)
- [6] Aplicaciones del Bid Data al sector turístico español. Available at: <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/33057/TFG-Tello%20MartAnez%2C%20Javier%20MarAa.pdf?sequence=1>
- [7] Ciencia de datos aplicada al Turismo en la provincia de Santa Cruz. Available at: <https://t.ly/4QQCU>
- [8] Encuesta sobre Gasto Turístico. Available at: <http://www.gobiernodecanarias.org/istac/estadisticas/sectorservicios/hosteleriayturismo/demanda/C00028A.html>
- [9] En cuestas y metodología. Available at: <https://github.com/alu0101127163/TFG-Ciencia-de-Datos-con-Python-aplicada-al-Turismo-de-Canarias/tree/main/Cuestionarios%20y%20metodolog%20a>
- [10] Numero de turistas en 2022. Available at: <https://www.ine.es/jaxiT3/Tabla.htm?t=10823&L=0>
- [11] Densidad de la población en las islas. Available at: [https://alarcos.esi.uclm.es/per/fruiz/pobesp/ter\\_isl.htm#:~:text=Las%20mayores%20densidades%20de%20poblaci%20n,40%2792%20hbt%2Fkm2](https://alarcos.esi.uclm.es/per/fruiz/pobesp/ter_isl.htm#:~:text=Las%20mayores%20densidades%20de%20poblaci%20n,40%2792%20hbt%2Fkm2).
- [12] Salario medio en Europa. Available at: <https://datosmacro.expansion.com/mercado-laboral/salario-medio>

[13] Encuesta sobre el gasto turístico. <http://www.gobiernodecanarias.org/istac/descargas/C00028A/2018/EGT-2018-metodologia.pdf>