

Trabajo Fin de Grado Ingeniería Aeroespacial

Diseño de un sistema de apoyo a la toma de decisiones sobre la apertura de nuevas rutas aéreas. Caso práctico: Aeropuerto de Jerez.

Autor: Javier Niño Martínez

Tutor: Jorge Juan Fernández de la Cruz

**Dpto. Ingeniería de la Construcción y Proyectos de
Ingeniería
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla**

Sevilla, 2022



Trabajo Fin de Grado
Ingeniería Aeroespacial

Diseño de un sistema de apoyo a la toma de decisiones sobre la apertura de nuevas rutas aéreas. Caso práctico: Aeropuerto de Jerez.

Autor:
Javier Niño Martínez

Tutor:
Jorge Juan Fernández de la Cruz
Profesor Titular

Dpto. Ingeniería de la Construcción y Proyectos de Ingeniería
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2022

Trabajo Fin de Grado: Diseño de un sistema de apoyo a la toma de decisiones sobre la apertura de nuevas rutas aéreas. Caso práctico: Aeropuerto de Jerez.

Autor: Javier Niño Martínez
Tutor: Jorge Juan Fernández de la Cruz

El tribunal nombrado para juzgar el trabajo arriba indicado, compuesto por los siguientes profesores:

Presidente:

Vocal/es:

Secretario:

acuerdan otorgarle la calificación de:

El Secretario del Tribunal

Fecha:

Agradecimientos

A todos los que estaban antes de empezar este duro y largo camino. Mis amigos de siempre y mi familia. Por el esfuerzo que han hecho para que yo haya llegado hasta el final.

A todos los que han comenzado a estar en esta etapa. A Álvaro y Antonio por ser compañeros de éxitos y fracasos. A mis amigos de Estambul por hacerme ver el mundo desde otra perspectiva y a todo el equipo de Zinke por hacer tan sencillo el poder poner fin a esta trayectoria.

Y a Jorge, mi tutor, por guiarme y ayudarme en el último peldaño de una larga escalera.

Han sido años de lágrimas y alegrías y solo puedo decir Gracias.

*Javier Niño Martínez
Sevilla, 2022*

Resumen

Tras la liberalización del mercado del transporte aéreo, tanto aerolíneas como agentes aeroportuarios actúan en un entorno de alta competencia en el que la toma de decisiones en cuanto al posicionamiento en el mercado es clave. Entre estas decisiones se encuentra la de abrir nuevas rutas aéreas que, por su rentabilidad, aportan grandes beneficios a la economía de una compañía.

Los métodos que en la actualidad se llevan a cabo para tomar dicha decisión de apertura de una nueva ruta no están claramente definidos y mucho menos unificados. Surge entonces la necesidad de establecer un sistema que, basado en datos, apoye a los agentes aeronáuticos a tomar decisiones de manera objetiva en cuanto a la apertura de nuevas rutas y sus posibles beneficios.

En este Trabajo de Fin de Grado se diseña un sistema que, dado un conjunto de variables input, ofrezca otro conjunto de variables output que ayuden a estimar la posible rentabilidad o no de abrir una nueva ruta aérea de transporte de pasajeros entre dos aeropuertos. Para ello se establecen una serie de ecuaciones estadísticas y un conjunto de bases de datos que, al ser combinadas, ofrecen una solución al problema de la falta de apoyo para la toma de decisiones objetivas sobre la apertura de nuevas rutas.

En una primera parte se hace un análisis del contexto del mercado aeronáutico, una investigación de las teorías ya propuestas y un repaso de los conceptos estadísticos que se utilizarán. A continuación, se definen las variables y ecuaciones del modelo y por último se plantea un caso práctico sobre la apertura de nuevas rutas desde el aeropuerto de Jerez.

Abstract

Following the liberalization of the air transport market, both airlines and airport agents operate in a highly competitive environment in which decision-making regarding market positioning is a key issue. Among these decisions is the determination to open new air routes that, due to their profitability, bring great benefits to the company revenue.

The methods that are currently carried out to make the decision to open a new route are not clearly defined and much less unified. The need then arises to establish a system that, based on data, supports aeronautical agents to make decisions objectively regarding the opening of new routes and their possible benefits.

In this Final Degree Project, a system is designed that, given a set of input variables, offers another set of output variables that allow deciding whether or not it is profitable to open a new passenger transport air route between two airports. For this, a series of statistical equations and a set of databases are established, offering a solution to the problem of the lack of support for objective decision-making on the opening of new routes.

Firstly, an analysis of the air transport market context is made, followed by an investigation of the theories already proposed and a review of the statistical concepts that will be used. Consecutively, the variables and equations of the model are defined and finally a practical case is presented on the opening of new routes from Jerez airport.

Índice Abreviado

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XIII
1 Introducción y contexto: El problema de la toma de decisiones	1
1.1 Objeto y alcance	2
1.2 Evolución de la apertura de nuevas rutas en aeropuertos españoles	2
1.3 Sistemas de apoyo a la toma de decisiones estratégicas de aerolíneas	4
2 El sistema de apoyo a la toma de decisiones	7
2.1 Introducción al sistema	7
2.2 Definición de conceptos clave	8
2.3 Conceptos estadísticos aplicados	9
3 Definición de las variables que intervienen	15
3.1 Variable demográficas	15
3.2 Variables aeroportuarias de origen y destino	18
3.3 Variables de ruta	20
3.4 Variables de competencia	22
3.5 Variables de costes de operación	24
4 La estructura de datos	27
4.1 Fuentes de adquisición de datos	27
4.2 Arquitectura de datos	38
5 Formulación del modelo	43
5.1 Ecuación de la demanda de pasajeros	43
5.2 Ecuaciones de la elección entre competencia	45
5.3 Ecuación de costes	48
6 Análisis estadístico de las variables	49
7 La estructura de outputs: Índices y accionables	53
7.1 Análisis cualitativo de los índices	53
7.2 Dashboard: Unificación de resultados	54
8 Validez del modelo	63
8.1 Actualización de datos: Validez temporal del modelo	63
8.2 El sistema a nivel europeo: Validez espacial del modelo	64

8.3	Otras hipótesis	64
8.4	Errores estadísticos	65
9	Estudio para un caso concreto: Aeropuerto de Jerez	67
9.1	Contexto: Situación, características y volumen de tráfico	67
9.2	Crecimiento: Evolución del tráfico, crecimiento de la demanda y rutas abiertas en los últimos años	69
9.3	Posibles rutas de apertura	71
9.4	Procedimiento de análisis	71
9.5	Toma de decisiones: La ruta más rentable	76
10	Conclusiones	81
	<i>Índice de Figuras</i>	83
	<i>Índice de Tablas</i>	85
	<i>Índice de Códigos</i>	87
	<i>Bibliografía</i>	89
	<i>Índice alfabético</i>	93
	<i>Glosario</i>	95

Índice

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XIII
1 Introducción y contexto: El problema de la toma de decisiones	1
1.1 Objeto y alcance	2
1.2 Evolución de la apertura de nuevas rutas en aeropuertos españoles	2
1.3 Sistemas de apoyo a la toma de decisiones estratégicas de aerolíneas	4
2 El sistema de apoyo a la toma de decisiones	7
2.1 Introducción al sistema	7
2.2 Definición de conceptos clave	8
2.3 Conceptos estadísticos aplicados	9
2.3.1 Modelo de demanda de elasticidades constantes	9
2.3.2 Modelos de demanda gravitacional	11
2.3.3 Modelos LOGIT y PROBIT	12
3 Definición de las variables que intervienen	15
3.1 Variable demográficas	15
3.1.1 Población	16
3.1.2 PIB: Producto Interior Bruto	16
3.1.3 Ratio de desempleo	16
3.1.4 Número de plazas hoteleras	16
3.1.5 Nivel de ocupación turística	16
3.1.6 Otras variables demográficas	17
3.2 Variables aeroportuarias de origen y destino	18
3.2.1 Capacidad aeroportuaria	18
3.2.2 Clave de referencia de aeródromo	18
3.2.3 Número de rutas activas	18
3.2.4 Número de churn de rutas	19
3.2.5 Número de aerolíneas que operan	19
3.2.6 Índice de accesibilidad al aeropuerto	19
3.2.7 Evolución del volumen de pasajeros	20
3.3 Variables de ruta	20
3.3.1 Distancia Destino-Origen	20
3.3.2 Tiempo de vuelo	21
3.3.3 Restricciones diplomáticas del vuelo	21
3.3.4 Histórico de vuelos Origen-Destino	21
3.3.5 Historial de búsquedas online	22
3.4 Variables de competencia	22

3.4.1	Variables de competencia de otros medios de transporte	22
3.4.2	Variables de competencia de otras aerolíneas	22
3.4.3	Variables de competencia de otros aeropuertos	23
3.5	Variables de costes de operación	24
3.5.1	Costes relativos a las tasas aeroportuarias e impuestos	24
3.5.2	Costes relativos a Handling	25
3.5.3	Costes relativos al combustible	25
3.5.4	Costes relativos al personal	25
4	La estructura de datos	27
4.1	Fuentes de adquisición de datos	27
4.1.1	Eurostat	28
4.1.2	OTLE	30
4.1.3	AENA Estadísticas	32
4.1.4	Google Trends	34
4.1.5	Flight Route Calculator	35
4.1.6	OMIO	37
4.2	Arquitectura de datos	38
4.2.1	Herramientas utilizadas	39
4.2.2	Relaciones entre bases de datos	39
5	Formulación del modelo	43
5.1	Ecuación de la demanda de pasajeros	43
5.2	Ecuaciones de la elección entre competencia	45
5.2.1	Probabilidad frente a otros medios de transporte	45
5.2.2	Probabilidad frente a otras aerolíneas	47
5.2.3	Probabilidad frente a a otros aeropuertos	47
5.3	Ecuación de costes	48
6	Análisis estadístico de las variables	49
7	La estructura de outputs: Índices y accionables	53
7.1	Análisis cualitativo de los índices	53
7.2	Dashboard: Unificación de resultados	54
7.2.1	Herramientas de visualización de datos	55
7.2.2	Metodología de creación de paneles y storytelling	55
8	Validez del modelo	63
8.1	Actualización de datos: Validez temporal del modelo	63
8.2	El sistema a nivel europeo: Validez espacial del modelo	64
8.3	Otras hipótesis	64
8.4	Errores estadísticos	65
9	Estudio para un caso concreto: Aeropuerto de Jerez	67
9.1	Contexto: Situación, características y volumen de tráfico	67
9.2	Crecimiento: Evolución del tráfico, crecimiento de la demanda y rutas abiertas en los últimos años	69
9.3	Posibles rutas de apertura	71
9.4	Procedimiento de análisis	71
9.4.1	Aplicación de los datos a las variables	71
9.4.2	Análisis de las variables demográficas para el caso XRY-BUD	73
9.4.3	Análisis de las variables aeroportuarias para el caso XRY-BUD	74
9.4.4	Análisis de las variables de ruta para el caso XRY-AMS	74
9.4.5	Análisis de las variables de competencia de modo de transporte para el caso XRY-SCQ	75
9.5	Toma de decisiones: La ruta más rentable	76

10 Conclusiones	81
<i>Índice de Figuras</i>	83
<i>Índice de Tablas</i>	85
<i>Índice de Códigos</i>	87
<i>Bibliografía</i>	89
<i>Índice alfabético</i>	93
<i>Glosario</i>	95

Notación

D_i	Demanda de la alternativa i
P_i	Precio de la alternativa
T_i	Tiempo de la alternativa
f_i	Frecuencia de la alternativa
C_i	Coste
X_i	Variable ejemplo
Δ	Incremento
ε	Elasticidad
e	número e
$\frac{\partial D}{\partial P}$	Derivada parcial de D respecto a P
$\log x$	Logaritmo de x
K	Constante gravitacional
M	Masa del objeto mayor
m	Masa del objeto menor
α_i	Coefficientes de elasticidad de la demanda
V	Utilidad de una alternativa
P_{rob}	Probabilidad de una alternativa
dx	Diferencial de x
C_{XY}	covarianza de dos variables aleatorias reales X e Y
R_{XY}	correlación de dos variables aleatorias reales X e Y
ρ_{XY}	Coefficiente de correlación de las variables aleatorias reales X e Y
$\Pr(A)$	Probabilidad del suceso A
σ_X^2	Varianza de la variable aleatoria X
MSE	Minimum square error
$\leq$	Menor o igual
$\geq$	Mayor o igual
Σ	Sumatorio
PIB	PIB de una región
Des	Desempleo en una región
Hot	Plazas Hoteleras
Och	Ocupación Hostelera
Cap	Capacidad
Ref	Clave de referencia
Act	Rutas activas
Chu	Churn de rutas

1 Introducción y contexto: El problema de la toma de decisiones

Desde que el mercado del transporte aéreo de pasajeros comenzó a liberalizarse, los agentes que intervienen en este viven en una constante pugna por hacerse un hueco en este complejo mundo. Dicha libertad, que tanto las aerolíneas como los agentes aeroportuarios fueron adquiriendo alrededor de los años 80, es la que hizo que las compañías pusieran todos sus esfuerzos en hacer sus rutas lo más eficientes y rentables posibles. [1]

Doganis (2001) argumenta que hay una serie de factores claves como la aparición de las compañías de bajo coste, el desarrollo de los sistemas aeroportuarios hub and spoke, las alianzas de compañías aéreas o el sistema de ventas online entre otros, que convierten un mercado que era relativamente estable en un sistema económico completamente volátil y esto hace que la toma de decisiones por parte de las compañías sea un punto crítico. Entre estas decisiones se encuentra la opción de apertura de nuevas rutas aéreas desde un aeropuerto o el propio cierre de las mismas. [2]

La estructura actual del sistema aéreo hace que el beneficio o las pérdidas que una aerolínea pueda obtener al entrar o salir de una ruta aérea no sólo estén directamente relacionada con los costes y beneficios de dicho itinerario, sino que afectará, en mayor o menor medida, al resto de operaciones de la compañía en otros destinos.

Para hacernos una idea de la repercusión que estas decisiones pueden tener, pongamos el ejemplo de las rutas que las tres principales compañías americanas abrieron, tanto nacional como internacionalmente, en el año 2015, y veamos que ocurrió con ellas a finales de 2019. Delta Airlines decidió operar 37 nuevas conexiones, de las cuales 26 siguieron vigentes en 2019. United Airlines entró en 27 rutas nuevas, sobreviviendo al cabo de esos 4 años 23 de ellas. Por último, American Airlines abrió también 50 de las cuales sobrevivieron 38. [3]

En la Figura 1.1, en el que se observan los ratios de éxito en la permanencia de rutas de las tres aerolíneas, podemos observar las diferencias de rentabilidad o viabilidad de dichas acciones entre tres compañías similares. Pequeñas variaciones en una decisión de entrada o salida a una ruta marcan la diferencia entre una aerolínea rentable y un desastre económico de pérdidas millonarias.

Como un ejemplo más concreto podemos comparar dos decisiones de aerolíneas sobre entrada en rutas. En 2019, la combinación Londres Heathrow (LHR)- Nueva York John Fitzgerald Kennedy (JFK) operada por British Airways, fue la ruta con más beneficios del panorama internacional, con más de 1,16 billones de euros vendidos en tickets.[4]

Por otra parte, a finales de 2018, American Airlines decidió la salida de su ruta Chicago O'Hare (ORD)- Pekín (PEK) porque, en palabras de Vasu Raja (vicepresidente de planning de la compañía), "La ruta estaba 80 millones por debajo del beneficio objetivo".[5]

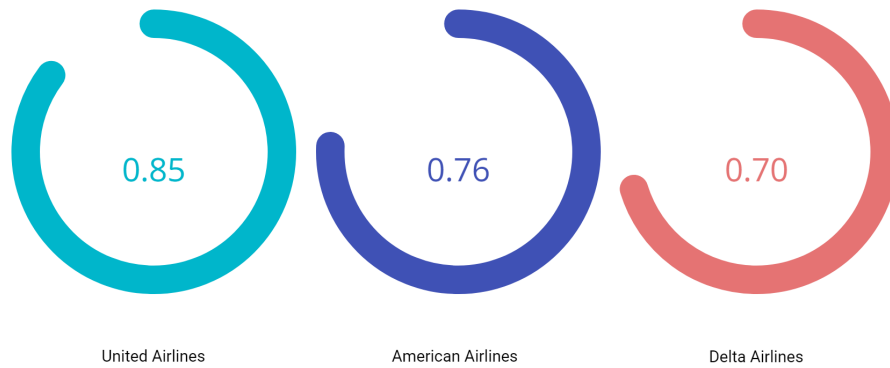


Figura 1.1 Ratios de éxitos de las tres grandes aerolíneas americanas (Elaboración propia).

Como podemos ver, la diferencia en la toma de decisiones de ambas compañías fue determinante y es por todo esto que se ha identificado la necesidad de diseñar un sistema sólido que, basado en los datos de las numerosas variables que intervienen, sea una base para la toma de decisiones de entrada o salida de una compañía en una ruta.

1.1 Objeto y alcance

El objetivo principal de este trabajo de fin de grado es la definición de un sistema que dé apoyo a la toma de decisiones de entrada en una nueva ruta aérea para aerolíneas, enfocándolo primero de forma global y particularizando posteriormente en el caso concreto del uso del sistema para la decisión de entrada de una aerolínea desde el aeropuerto de Jerez. Además de diseñar dicho sistema en sí, se trata de un proyecto de investigación acerca de todos los aspectos que rodean la entrada de aerolíneas, tanto de los puramente económicos de estructura de mercado, como de los aspectos estadísticos que nos permitirán desarrollar nuestro modelo.

Se tendrán en cuenta la tipología del mercado global del transporte aéreo, la situación actual y su evolución con el paso del tiempo, la previsión de mercado esperada y las teorías o métodos que ya usan las aerolíneas para tomar decisiones. Posteriormente, se definirán las variables que intervendrán como input en el sistema, las teorías estadísticas que se utilizarán y la particularización de estas para la obtención de los outputs del sistema. Definiremos simultáneamente las fuentes de datos y la infraestructura que definirá la combinación de dichas bases de datos. Por último, expondremos los outputs, que serán una serie de indicadores resumidos en un cuadro de mando o dashboard. Para particularizar, haremos un análisis concreto de la implementación de dicho sistema para la decisión de entrada de una aerolínea en el aeropuerto de Jerez (España).

Por la complejidad del asunto que se trata, tanto en el sentido de la gran cantidad de factores que intervienen como por la carencia de bases completamente definidas, el proceso previo de investigación es tanto o más definitivo que la propia definición del sistema, por lo que en el caso práctico se ofrece un análisis reducido en cuanto al data set que se utiliza.

En conclusión, este proyecto permitirá conocer, tras un profundo análisis de investigación, cuáles son los factores que determinan la decisión de entrada de una compañía en una ruta aérea, cómo se relacionan dichos factores y cómo se podría mejorar en la toma de decisiones con la unicidad de un sistema.

1.2 Evolución de la apertura de nuevas rutas en aeropuertos españoles

Para comenzar a diseñar dicho sistema, que en nuestro caso se centrará en el panorama aeroportuario español, es necesario hacer un análisis previo de cómo ha evolucionado la apertura de rutas desde los aeropuertos de la red Aena en los últimos años.

Dicha evolución ha venido marcada en los últimos años por las restricciones que la pandemia de Covid-19 dejó en el ámbito aéreo. Y aquí merece recalcar que el asunto de la toma de decisiones se convierte aún en más vital con situaciones como esta.

Haciendo uso de los datos proporcionados por la Dirección de Datos y Mercado Aeronáutico de Aena desde 2018 a 2021 sobre las aperturas, podemos llegar a una serie de conclusiones que nos ayudaran a comprender mejor el contexto del sistema que vamos a diseñar.

2018 se cerró con 336 nuevos trayectos desde aeropuertos españoles. 38 fueron vuelos nacionales, 272 fueron europeos y los 26 restantes extracomunitarios. 2019 fue el año en el que las restricciones aparecieron y debido a esto, el crecimiento se frenó. Se activaron 323 nuevas rutas. Aumentó el tráfico español con 57 nuevas rutas nacionales, pero el europeo (241) y el extracomunitario (25) se mantuvieron en la misma línea de crecimiento.

2020 es el reflejo del desastre que supuso para el mundo de las aerolíneas la crisis del coronavirus. Sólo 71 nuevas rutas nuevas. 47 europeas y 24 nacionales. Tras esta caída, en 2021 comenzó la relajación de medidas y se alcanzaron niveles similares a los años anteriores. 302 nuevas rutas (74 nacionales, 208 europeas y 20 extracomunitarias).[6]

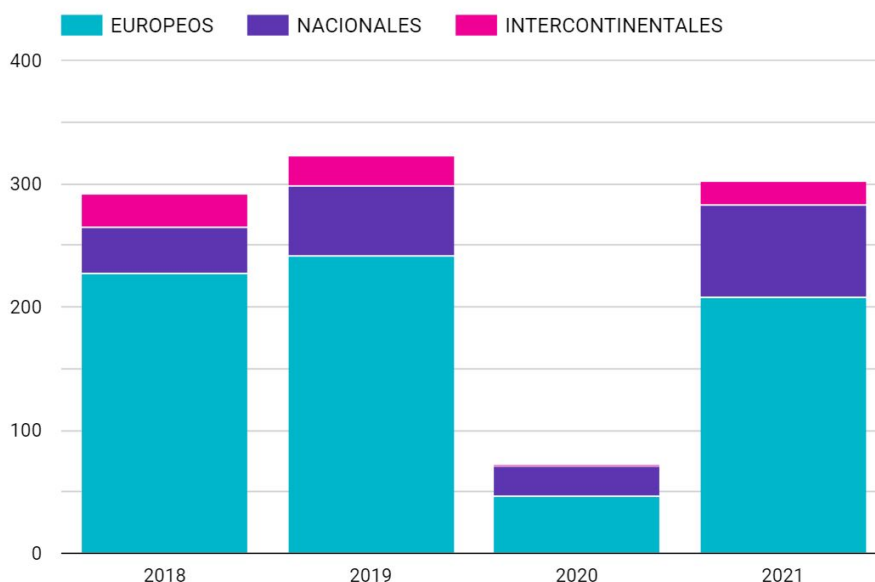


Figura 1.2 Evolución de aperturas de rutas en aeropuertos de la red AENA. (Elaboración propia).

En la Figura 1.2 podemos hacernos una idea más clara de dicha evolución de crecimiento.

En todos los datos comentados anteriormente se considera ruta nueva aquella con más de 5.000 pasajeros en el año actual y menos de 1.000 pasajeros en el año anterior.

El hecho de analizar la evolución anterior nos ayuda a realizar una somera predicción de cómo continuará dicho avance en este periodo y en años venideros. Puesto que nos encontramos en un primer apartado en el que el objetivo no es otro que una contextualización, nos dejaremos guiar por una recta de regresión que, mediante un gráfico de dispersión nos permita hacernos una ligera idea del número de nuevas rutas que se podrían generar en 2022. Como conclusión, podemos decir que nos encontramos en un momento clave para las compañías en el que los índices de crecimiento son elevados y con previsiones de que el número de rutas nuevas va a alcanzar una cifra récord. Es por ello que, en la actualidad, el sistema de apoyo a la toma de decisiones de entrada que vamos a diseñar en el presente proyecto es aún más necesario si cabe.

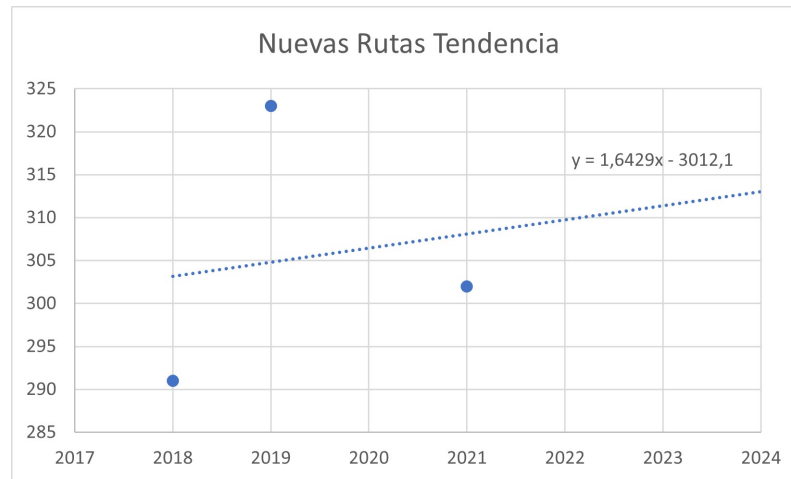


Figura 1.3 Recta de regresión previsión apertura de nuevas rutas en España (Elaboración propia).

1.3 Sistemas de apoyo a la toma de decisiones estratégicas de aerolíneas

Pese a la clara carencia de sistemas sólidos de apoyo a las decisiones estratégicas de entrada en las aerolíneas y teniendo en cuenta que, puesto que es un problema relativamente moderno (en el sentido de que vino de la mano de la liberalización del mercado aéreo), no existe una extensa literatura en el asunto, desde los años 80, se han llevado a cabo diversas teorías, sistemas estadísticos y de previsión y estudios de mercado.

La tónica general de dichos estudios son sistemas de previsión de la demanda de tráfico de pasajeros basados en conceptos estadísticos. En algunos casos sistemas multivariantes complejos y en otros, distribuciones de la estadística clásica aplicados al tema que nos concierne.

Los primeros modelos que se comenzaron a tener en cuenta por las compañías fueron modelos de entrada en mercados concentrados u oligopolios. Estos modelos proponen análisis empíricos de los efectos de entrada a un mercado concentrado y están muy centrados en el estudio de la competencia. Bresnahan y Reiss (1990) estudiaron la relación entre el tamaño del mercado, el número de agentes y la competitividad entre estos y dictaminaron que en mercados de alrededor de 5 competidores, la mayoría de cambios en las funciones de beneficio, vienen dados por entradas o salidas de competidores en el mercado.[7]

Cuando el mercado fue evolucionando y se produjo la “revolución de las low-cost” tal y como lo denomina Doganis (2001), estos estudios se quedan en algo de menos utilidad y aparece el problema del manejo y combinación de un gran número de variables heterogéneas que intervienen en los patrones de decisión.[2]

Es aquí donde aparece el concepto económico-estadístico de los patrones de entrada. La literatura desde entonces se ha basado en establecer modelos empíricos que definan dichos patrones de entrada. Berry (1992) diseña, basándose en los modelos de entrada en oligopolios, una primera metodología empírica que usa una serie de inputs de variables de entrada de la aerolínea en el mercado como indicadores de rentabilidad. Proporciona, en definitiva, tomando como objeto un par de aeropuertos (origen-destino), la estimación de beneficio de dicha ruta.[8]

Boguslaski, Ito y Lee (2004), basándose en el mismo objeto de par de aeropuertos bidireccional identifican empíricamente los factores que hicieron que la aerolínea americana Southwest entrase o saliese de ciertas rutas. Para esto utilizan un modelo Probit basado en datos cruzados.[9]

Desde entonces la bibliografía se ha basado en dichos modelos. Tenemos sobre el papel estudios muy descriptivos como el de Whinston y Collins (1992) en el que desde un estudio de un caso concreto de la aerolínea People Express, llegan a una definición de la estructura competitiva del mercado. [10] Por otro lado, encontramos sistemas completos que nos permiten escoger una base de datos y obtener los outputs de salida, como es el caso del modelo empírico para la entrada de aerolíneas low-cost que propone Oliveira

(2008). [11]

Merece la pena detenerse en este momento, porque ciertos modelos de la literatura tratan el concepto de “Churn” de rutas como algo análogo a la entrada. De Wit y Zuidberg (2016) analizaron las variables que más intervienen en las decisiones de salida de una ruta para una compañía low-cost, llamando “Churn” a dicha salida. La distancia y el número de asientos son las variables más determinantes, aunque otras como la estacionalidad o los años de antigüedad de la ruta también son clave. Con el análisis de dichas variables proponen un modelo de regresión lineal gracias al que son capaces de llegar a una previsión temprana de la decisión de salida, antes de que las pérdidas sean más elevadas.[12]

Aproximaciones más modernas a este concepto del “Churn”, hacen uso de técnicas de machine learning o Deep learning para, en base a datos provenientes de encuestas, predecir la acción de salida de ruta. Park y Kim (2022), utilizan dichas técnicas en base a la opinión de una muestra de usuarios de aerolíneas coreanas. En este caso, en vez de hacer la predicción en base a datos de la ruta en sí (número de asientos, o beneficio), aplican dichos algoritmos para interrelacionar variables como la edad de los viajeros, su propósito de vuelo, su ocupación laboral o la aerolínea con la que vuelan normalmente.[13]

En definitiva, al igual que se ha analizado en profundidad la literatura relativa a los modelos de entrada en el mercado, se podría haber hecho con los modelos de Churn o salida, pero nuestro sistema estará basado en la decisión únicamente de entrada, por lo que no se han analizado con la misma profundidad.

Aunque en este apartado no se haya ahondado en ninguno de los estudios, tanto de entrada, como de salida, el modelo consistirá en combinar y completar dichos estudios de entrada, por lo que en la descripción del modelo se volverán a hacer referencias más detalladas sobre estos.

2 El sistema de apoyo a la toma de decisiones

En este primer apartado se pretende realizar una descripción concisa de cómo será el sistema que se diseña. Quedarán en primer lugar definidos una serie de conceptos claves, hipótesis y especificaciones. A continuación, se hará un análisis desde un marco teórico de los modelos estadísticos en los que se basa.

2.1 Introducción al sistema

El sistema se definirá como una combinación del modelo de demanda gravitacional y el modelo de elección Logit en el que, dadas una serie de variables de entrada, se definirán varias variables de salida, algunas de ellas discretas y otras continuas. El análisis de dichas variables de salida será lo que nos permita tomar la decisión, de ahí que se denomine sistema de apoyo a la toma de la decisión.

Si el output fuese único, combinar un modelo de demanda con otro de elección sería complejo. Cabe destacar que cuando se habla de combinación, estamos hablando de multiplicidad de outputs dados un conjunto de inputs.

En el diagrama de flujo de la figura 2.1 se define la metodología del modelo.

Tal y como se aprecia y como se seguirá en los siguientes apartados de este escrito, se distinguen tres grandes bloques de trabajo. En primer lugar, el sistema de inputs estará compuesto por varias fases: Una primera obtención de los datos, fruto del análisis de diferentes bases de datos, un proceso de limpieza, orden y asociación de todos los datos obtenidos, y una correlación entre estos datos y las correspondientes variables que entrarán en el sistema estadístico. La fase central será la definición de los procesos que dichas variables sufrirán y por último, en una tercera fase se obtendrán unos outputs que serán analizados y que serán los que permitan tomar la decisión.

Se debe comenzar haciendo una serie de apreciaciones que respaldarán los pasos seguidos. En primer lugar, es necesario poner límites a las decisiones a las que nuestro modelo apoyará. Partiendo de la necesidad que se tiene, que no es otra que el tomar una decisión de entrada o no entrada por parte de una aerolínea en una ruta aérea (que puede estar ya siendo operada por otra aerolínea o no) formada por un par de aeropuertos, uno de origen y otro de destino.

Es crucial en este punto recalcar el hecho de que el modelo, al tratarse de un modelo predictivo, no apoyará decisiones de salida de ruta, será únicamente de apoyo a la decisión de entrada. Como se verá en más detalle posteriormente, las opciones de la decisión serán la entrada o la no entrada, nunca se hablará de salida o no salida de una ruta que ya se estuviese operando.

Nuestro modelo tomará como unidad de observación y análisis el par de aeropuertos origen-destino, por lo que todas las variables deberán ir asociadas o bien a un aeropuerto de destino, o bien a un aeropuerto de

origen, o bien a la propia ruta. En cuanto al tiempo de validez de nuestro modelo, será un modelo con validez en un punto de tiempo. Dicho instante puntual será el momento de la toma de la decisión. Hacer un modelo con el tiempo como variable haría mucho más complejo el sistema.

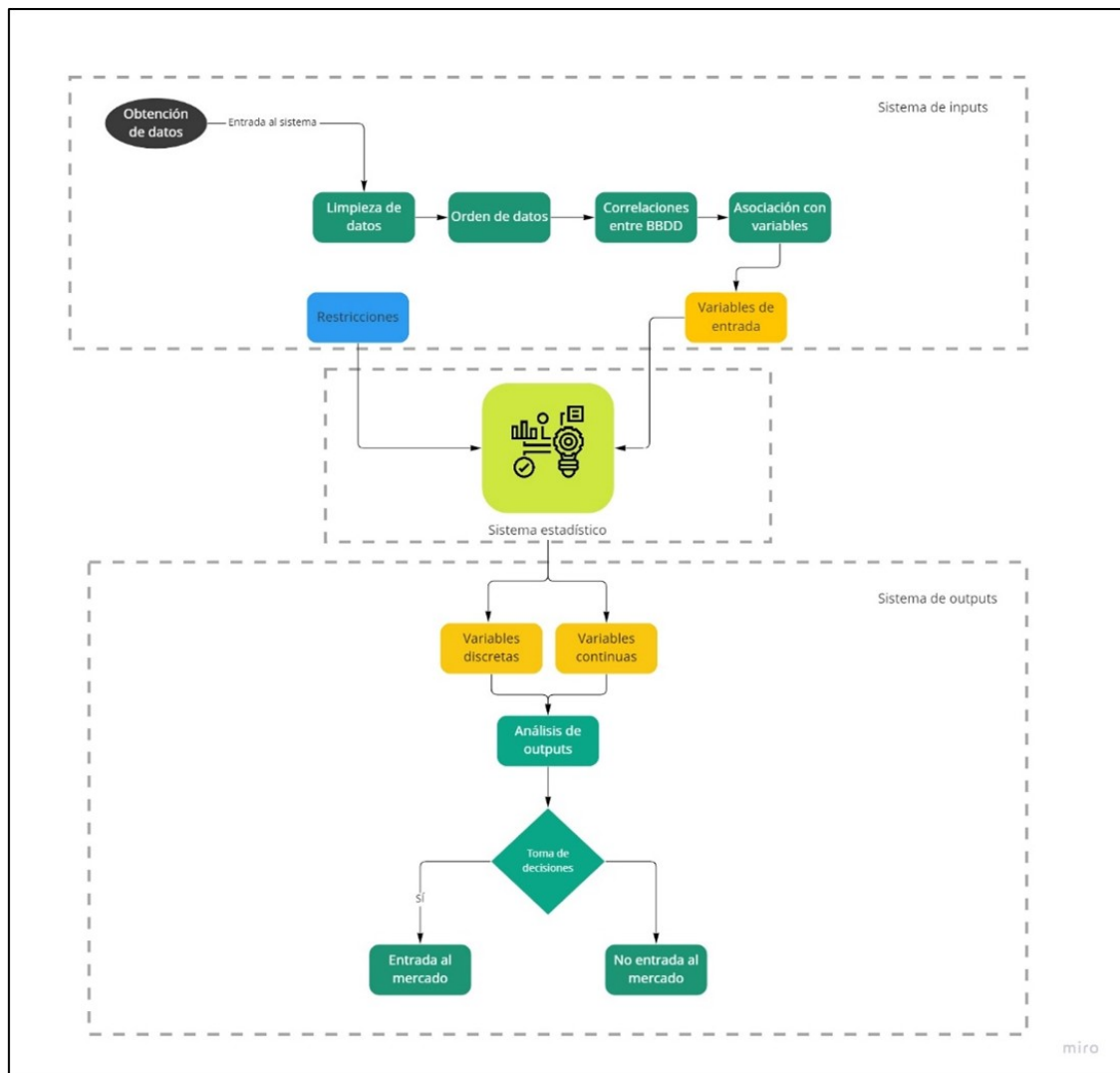


Figura 2.1 Diagrama de flujo de la metodología seguida (Elaboración propia).

2.2 Definición de conceptos clave

Se tratará en este punto de definir una serie de conceptos claves que se repiten durante todo el proyecto y que serán indispensables para la metodología de trabajo que se sigue.

- **Entrada:** Se define como la decisión de un agente aeronáutico de iniciar una nueva ruta aérea con fines comerciales entre un par de aeropuertos. Dicha nueva ruta puede o bien estar operada por otro agente, o haber sido operada anteriormente por el mismo agente. Conceptos equivalentes a entrada son “Apertura de nueva ruta aérea”, “Nueva presencia de la aerolínea” o “Escalado en las operaciones de la aerolínea desde un nuevo par de aeropuertos”.
- **Par aeropuerto Origen-Destino:** Se define como un par de aeropuertos bidireccional a la combinación formada por dos aeropuertos (que serán tanto origen como destino). Será la unidad sobre la que se

tome la decisión.

- *Input*: Se define input a las variables que se obtienen de la observación y toma de datos. Sin sufrir dichas variables ninguna fórmula estadística ni combinación.
- *Output*: Se define output a las variables resultados que se obtienen de la combinación de una serie de inputs mediante la aplicación de fórmulas estadísticas.
- *Demanda*: Se define como "La cantidad de un bien que los compradores quieren y pueden comprar".[14]
- *Población*: Se define como "un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones".[15]
- *Muestra*: Se define como "Cualquier subconjunto de la población".
- *Variable cualitativa*: Se define como "Aquella variable que no es medible, es decir, que toma un valor no numérico".
- *Variable cuantitativa*: Se define como "Aquella variable que es medible, y por tanto toma un valor numérico".
- *Variable discreta*: Se define como "Aquella variable cuantitativa que sólo puede tomar un número finito o numerable de valores; es decir, toman valores enteros."
- *Variable continua*: Se define como "Aquella variable cuantitativa que puede tomar cualquier valor dentro de un Intervalo; esto es, puede tomar cualquier valor real."
- *Variable Dummy*: Se define como "Una variable binaria: Variable que es 0 o 1. La variable binaria se utiliza para indicar un resultado binario. Por ejemplo, X es una variable binaria para el sexo de una persona con valor $X = 0$ si la persona es mujer y con valor $X = 1$ si la persona es hombre." También se le denomina variable ficticia.[16]
- *Variable endógena*: Se define como: "Variable que presenta correlación con el término de error" [16]
- *Variable exógena*: Se define como: "Variable incorrelacionada con el término de error de la regresión." [16]

2.3 Conceptos estadísticos aplicados

En esta sección se hace un repaso de la formulación de los conceptos estadísticos que nos permitirán, en la sección 4 de este documento, desarrollar las ecuaciones que definirán el modelo del sistema.

2.3.1 Modelo de demanda de elasticidades constantes

Se define demanda como "La cantidad de un bien que los compradores quieren y pueden comprar".[17] Es por tanto este concepto, particularizado al ámbito de este estudio, clave para nuestro sistema. En concreto, demanda para nosotros será "Cantidad de transporte que los clientes quieren y pueden comprar". Y cuando se habla de transporte podemos referirnos a número de pasajeros, número de operaciones, etc.

La demanda de un modelo de transporte aéreo puede simplificarse como una función que depende tanto de los precios de los distintos modos de transporte, cómo de otro conjunto de variables relativas al transporte aéreo (población de las ciudades, renta per cápita, etc). Se podría decir que sería una ecuación como esta:

$$D_i = f(P_i, P_j, \dots, P_n, X_1, X_2, \dots, X_m) \quad (2.1)$$

Donde D_i sería la demanda de pasajeros del transporte aéreo, P_i sería el precio de dicho transporte, $P_j \dots n$ sería el precio del resto de alternativas al transporte aéreo (tren, bus, etc) y $X_1 \dots m$ el resto de los factores que influyen (socioeconómicos, operativos, etc).[18]

Y es a la hora de definir dicha función cuando se plantea el concepto de elasticidad de la demanda. En econometría, se define elasticidad como “la respuesta o sensibilidad de una variable con respecto a un cambio en otra variable”.[14] La elasticidad de la demanda define, por tanto, la sensibilidad de la demanda frente a la variación de alguna de las variables de las que depende.

Se expondrá en este apartado el caso particular de la elasticidad precio de la demanda. La elasticidad tendrá la siguiente forma:

$$\varepsilon = \frac{\Delta D}{\Delta P} \quad (2.2)$$

Donde ΔD es la variación de la cantidad demandada, ΔP es la variación del precio y ε la elasticidad.

Cuando hay más de una posibilidad de modo de transporte, se hace referencia la definición de elasticidad cruzada de la demanda: cómo varía la demanda de un producto i con la variación del precio de un producto j :

$$\varepsilon_{ij} = \frac{\partial \log D_i}{\partial \log P_j} \quad (2.3)$$

Para convertir la ecuación 2.1 en una ecuación lineal que se pueda resolver con un ajuste por regresión, surge el modelo de la demanda de elasticidades constantes que establece:

$$D = e^{(\gamma_0)} * (X_1)^{(\gamma_1)} * (X_2)^{(\gamma_2)} * \dots * (X_N)^{(\gamma_N)} \quad (2.4)$$

Dónde $X_1 \dots N$ son las variables que intervienen en la demanda y $\gamma_0 \dots N$ las elasticidades de dichas respectivas variables.

Si en esta ecuación 2.4 se toman logaritmos, llegamos a una ecuación lineal que puede ser resuelta por un ajuste de regresión:

$$\log(D) = \gamma_0 + \gamma_1 \log(X_1) + \gamma_2 \log(X_2) + \dots + \gamma_n \log(X_n) \quad (2.5)$$

Donde se sigue la misma nomenclatura que en 2.4.[19]

Vista esta aproximación, es destacable decir que será de crucial importancia en el modelo estudiar y analizar la correlación entre variables, cómo de significativas son estas y cuáles serán los errores del ajuste en cada caso.

Esta primera definición permite llegar a comprender de mejor manera los modelos de demanda gravitacional que se verán en la siguiente sección.

2.3.2 Modelos de demanda gravitacional

Puesto que, en la actualidad, la predicción de demanda de número de pasajeros es la herramienta más usada tanto por los aeropuertos como por las aerolíneas para hacer planificación de rutas, flotas, horarios y capacidades, es vital tener en cuenta dichas predicciones en el modelo.

Se ha comentado en apartados anteriores el origen de los modelos de demanda simples, pero es cierto que, aunque no existe una certeza de exactitud en ninguna teoría de modelado estadístico de esta índole, los modelos de demanda gravitacional son los más empleados en el mundo del transporte aéreo de pasajeros.

La principal ventaja de los modelos gravitacionales con respecto a los modelos de demanda agregada tradicionales es que son particularmente aplicables a un par de ciudades (o aeropuertos en nuestro caso), que pueden, o no, tener o haber tenido rutas operadas.

Estos modelos, antes de entrar en más detalle, se basan en tres tipologías de variables: Variables explicativas del origen, variables explicativas del destino y variables específicas del par origen-destino. Las dos primeras variables, que son del tipo geográficas-demográficas, relativas al origen o al destino, son características ajenas a la operación y sobre las que el operador no puede tomar decisiones o variaciones. Por ejemplo, la población, nivel económico de la población, etc. En cambio, sobre el tercer tipo de variables, las relativas al par, encontramos por un lado algunas variables sobre las que el operador no puede tomar decisiones (distancia entre ciudades o aerolíneas competidoras) y otras sobre las que sí (aeronave que se utilizará o tiempo del vuelo, por ejemplo).

Estas variables se desglosarán, definirán en profundidad y analizarán en próximos apartados.

El modelo de demanda gravitacional, como su propio nombre indica, procede o está inspirado en la ley de gravitación universal que determina que la fuerza con la que dos objetos se atraen es directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia que los separa:

$$G = \frac{K * (M * m)}{d^2} \quad (2.6)$$

Atendiendo, por otro lado, a la ecuación de la demanda de elasticidades constantes que se trató en capítulos anteriores para n variables:

$$\log(D) = \alpha_0 + \alpha_1 \log(X_1) + \alpha_2 \log(X_2) + \dots + \alpha_n \log(X_n) \quad (2.7)$$

Donde los coeficientes se refieren a la elasticidad de la demanda y los logaritmos a las variables explicativas.[20] Para más detalle sobre este segundo modelo, véase el apartado anterior.

En definitiva, este modelo se ajustará y obtendremos la demanda agregada, dadas la demanda y elasticidad de una serie de variables explicativas que suelen depender de un solo tipo de individuo. Ahora bien, el modelo que se expone depende de más de una entidad en tanto en cuanto depende de un par de aeropuertos origen-destino.

Si se vuelve a la ecuación 2.1 intentando asimilar los dos objetos a ambos aeropuertos, y la distancia a la ruta, si quisiésemos expresar la demanda de pasajeros de esta forma se trataría de algo así:

$$D = k * \frac{(X * Y)^\alpha}{t^\beta} \quad (2.8)$$

Donde D sería la demanda de pasajeros (por ejemplo), k una constante, X serían variables relativas al aeropuerto de origen, Y al aeropuerto de destino, α un parámetro que controla la interrelación entre ambas, t

variables relativas a la propia ruta entre ambos y β otro parámetro que controla la relación entre variables de la ruta y la demanda.

Uniando ambas expresiones en una única expresión en la forma de modelo de elasticidades constantes obtendríamos la expresión siguiente:

$$D = \alpha X_1^\theta X_2^\theta \dots X_n^\theta Y_1^\phi Y_2^\phi \dots Y_n^\phi T_1^\mu T_2^\mu \dots T_n^\mu \quad (2.9)$$

Para comprender dicha expresión, se usará la expresión y variables que propone Grosche (2007) en uno de sus estudios sobre modelos gravitacionales[20]:

$$V_{ij} = e^\varepsilon (P_{ij})^\pi (C_{ij})^\chi (B_{ij})^\beta (G_{ij})^\gamma (D_{ij})^\delta (T_{ij})^\tau \quad (2.10)$$

Donde, aunque la expresión es equivalente a la anterior expuesta, la importancia recae en entender la particularidad de las variables:

Tabla 2.1 Valores de parámetros.

Definición	notación
Población	D_{ij}
Coste	C_{ij}
Índice de poder de compra	B_{ij}
PIB	G_{ij}
Distancia Geográfica	D_{ij}
Tiempo Medio de Viaje	T_{ij}
Volumen de pasajeros entre i y j	V_{ij}

Tras este breve estudio, parece que el modelo de demanda gravitacional puede ajustarse de una manera aproximadamente correcta a nuestro sistema, puesto que la unidad de observación es el par de aeropuertos y las variables que elegiremos serán similares. Puesto que nuestro sistema no será un cálculo de demanda exclusivamente, tendremos que combinar los conceptos vistos en este apartado para llegar a una ayuda más eficiente a la hora de la toma de decisiones.

2.3.3 Modelos LOGIT y PROBIT

Puesto que la toma de decisión es la de entrar o no entrar, estamos ante un output que, si el modelo fuera perfecto, sería una variable dependiente binaria.[21] En el caso estudiado, la variable R (entrada de ruta), será 1 en el caso de que se decida entrar en la ruta, y 0 en caso contrario.

El problema estará entonces en, dadas una serie de variables independientes, establecer la probabilidad de que la variable R sea 1 o la probabilidad de que sea 0.

En el caso de la existencia de varias modalidades de transporte (demanda multiproducto) entre las que se tiene que elegir una alternativa, el modelo más utilizado a la hora de elegir una alternativa es el llamado modelo Logit. Al igual que se aplica para la existencia de multi-modalidad de transporte, también es aplicable a la elección entre varias aerolíneas o entre varios aeropuertos.[22]

Según este modelo, la probabilidad de elegir una alternativa i , tiene la siguiente expresión:

$$P_i = \frac{e_i^V}{\sum_{k=1}^n e_k^V} \quad (2.11)$$

Donde P_i es la probabilidad de que entre las n alternativas se escoja la i , y V_i es la utilidad de la alternativa i .

Se define la utilidad de una alternativa como un índice o parámetro que depende de las variables o atributos del objeto de decisión de la siguiente manera:

$$V_i = \sum_k 1_i^m \theta_{ik} X_{ik} \quad (2.12)$$

Donde θ_{ik} es un coeficiente relativo al atributo k de la alternativa i y X_{ik} es la variable.

Se propone un ejemplo en el que tenemos que calcular el reparto modal (o la probabilidad de elección) de que un pasajero elija entre tres aerolíneas para volar. [23]

- Aerolínea A: $i=1$
- Aerolínea B: $i=2$
- Aerolínea C: $i=3$

Y las variables que se tendrán en cuenta serán:

- Ingresos medios del pasajero: I
- Precio: P
- Aerolínea low cost o tradicional: L (Variable Dummy)
- Antigüedad operando la ruta: A
- Existencia clase Business: B (Variable Dummy)

Y los coeficientes de los atributos serán:

- Ingresos medios del pasajero: θ_{i1}
- Precio: θ_2
- Aerolínea low cost o tradicional: θ_3
- Antigüedad operando la ruta: θ_4
- Existencia clase Business: θ_5

La primera variable I depende del sujeto y el resto de variables solo dependen de la propia alternativa.

Podemos formular las tres utilidades cómo:

$$V_1 = \theta_{11}XI_1 + \theta_2XP_1 + \theta_3XL_1 + \theta_4XA_1 + \theta_5XB_1 \quad (2.13)$$

$$V_2 = \theta_{21}XI_2 + \theta_2XP_2 + \theta_3XL_2 + \theta_4XA_2 + \theta_5XB_2 \quad (2.14)$$

$$V_3 = \theta_{31}XI_3 + \theta_2XP_3 + \theta_3XL_3 + \theta_4XA_3 + \theta_5XB_3 \quad (2.15)$$

De tal manera que, con estos tres valores, podríamos aplicar la fórmula para ver la probabilidad de elección:

$$P_i = \frac{e_i^V}{\sum_{k=1}^n e_k^V} \quad (2.16)$$

En definitiva, dicho modelo nos servirá para descartar aquellas rutas que tengan un porcentaje de probabilidad de elección muy bajo en nuestro modelo. Nos ayudarán a conocer si elegirían nuestra aerolínea o nuestro aeropuerto.

3 Definición de las variables que intervienen

En este apartado se tratará de definir las variables que tomarán parte en el modelo. Se hará una definición descriptiva básica sin entrar en detalles estadísticos o matemáticos (estos detalles se tratarán en el capítulo 5). En este apartado, además, se definirán variables que no se incluirán en el modelo pero que si que tienen influencia en la decisión. Dichas variables serán descartadas por el poco peso en la decisión o por complejidad extrema a la hora de tratar datos.

Dividiremos las variables en una serie de grupos en función de su tipología:

3.1 Variable demográficas

Las variables a las que se llama demográficas en este estudio son aquellas que no están directamente relacionadas con el transporte aéreo o el mercado de las aerolíneas. Son variables relativas a características propias de las áreas de influencia del aeropuerto de origen o destino.

Es difícil definir una regla que nos permita establecer fácilmente el área sobre el que influirá un aeropuerto a priori. Para ello hay sistemas mucho más complejos que este propio proyecto. En el Plan Director de un aeropuerto en España hay siempre un apartado en el que se define dicho área.[24] El apartado 2. Análisis del entorno, contendrá en todo caso un epígrafe (2.2. Entorno Socioeconómico) en el que se define este área en función de una serie de características (similares a las variables que se analizarán en este apartado):

- *Población*
- *Mercado de trabajo*
- *Distribución sectorial de las actividades*
- *Espacios naturales existentes*
- *Sistema de transporte*

En los planes directores se define: "El área de influencia del aeropuerto está formada por las poblaciones alrededor del mismo que pueden considerarse como potenciales usuarios de esta infraestructura, es decir, áreas de población que se supone emplearán los servicios del aeropuerto frente a la posibilidad de utilizar otros medios de transporte".[25]

Esta definición es muy ambigua y para hacer esto más concreto, hay estudios que utilizan diversas técnicas (software en la actualidad) para definir estas regiones. Una teoría muy aplicada es la de los Polígonos de Thiessen que divide las áreas de influencia en polígonos cuyas áreas varían según una serie de variables basándose en la distancia Euclídea.[26]

Puesto que hacer un análisis de polígonos de Thiessen es costoso para un proyecto como este, simplificaremos el modelo suponiendo que las regiones se definirán como un porcentaje de las áreas cercanas al aeropuerto. Esta aproximación y su correspondiente error será tratada en apartados posteriores.

Por tanto, las variables que se definen a continuación, tendrán validez en la región a la que corresponda el aeropuerto de origen y de destino. Establecemos las siguientes variables demográficas:

3.1.1 Población

"Es el número de habitantes que habitan en una región en un instante de tiempo determinado".

La variable Población será una de las más determinantes en nuestro modelo y base de la elección. Es natural y empírico el hecho de que una ruta será más rentable conforme más pasajeros opere y este volumen de pasajeros estará altamente influenciado por la población.

Se podría tratar también el asunto de la evolución de esta población en el tiempo, de hecho, si se fuese a hacer un estudio a muy largo plazo, cómo el de emplazamiento de una nueva terminal o un nuevo aeropuerto, se estaría hablando de un asunto decisivo. La proyección de evolución de las poblaciones y la predicción de población futura se evitan en este modelo para simplificarlo, teniendo en cuenta que la decisión de entrada o salida se tomaría en el momento de reparto del slot del curso siguiente y la ruta entraría a operar en un periodo relativamente corto de tiempo, por lo que la población no va a sufrir una variación destacada en ese periodo.

3.1.2 PIB: Producto Interior Bruto

"El PIB es el valor en número de unidades monetarias de los bienes y servicios producidos por una región en un tiempo determinado"

Al igual que la población, esta variable será definitoria en nuestro modelo. Diversos estudios dan fe de que el crecimiento del volumen de transporte aéreo de una región y el crecimiento económico de dicha área están muy relacionados. No podemos establecer con certeza que una de las dos variables sea la que determine la otra pero la correlación está más que probada, de ahí su elección en este modelo.

3.1.3 Ratio de desempleo

"El Ratio de desempleo es el porcentaje de la población en edad de población activa actualmente en situación de paro o desempleo"

Esta variable es un segundo indicador del bienestar económico de una región o población. Haber escogido solo el PIB de una región como indicador único económico puede inducir a error, de ahí que se complementa con este segundo indicador que definirá que la demanda de transporte aéreo estará relacionada con el número de habitantes en edad de trabajo desempleados. La Unión europea define desempleado como "aquella persona de entre 15 y 64 que no tiene trabajo en el periodo actual o está actualmente disponible para trabajar o que busca activamente trabajo o que ha encontrado un trabajo para comenzar en un período máximo de tres meses".

3.1.4 Número de plazas hoteleras

"Es el número de plazas hoteleras disponibles en los establecimientos preparados para ello en una región en un periodo de tiempo determinado".

La Unión Europea define establecimiento turístico como "cualquier establecimiento que, de forma regular u ocasional, proporciona alojamiento de corta duración a turistas como un servicio retribuido". Tanto esta variable como la que se trata a continuación, son imprescindibles por la correlación turismo-transporte aéreo, que será analizada en el siguiente apartado.

3.1.5 Nivel de ocupación turística

"Es el número de noches que un turista registra en un establecimiento hotelero en un periodo de tiempo determinado".

Se determina que no es necesaria la estancia como tal, sino el registro de que se va a pasar la noche en dicho establecimiento. De nuevo, los datos se expresan por sí solos y es una certeza que, en términos generales, un

gran porcentaje de los pasajeros que vuelan lo hacen por motivos de turismo, lo que nos lleva a pensar que las variables demanda y número de pernoctaciones hoteleras estarán altamente correlacionadas.

Atendiendo a los datos de número de turistas según motivo principal del viaje a España que ofrece el INE (Instituto Nacional de Estadísticas), y, por actualidad, se analizan los datos de 2021 y el primer trimestre de 2022.[27] Los encuestados son preguntados por su motivo de viaje eligiendo una entre las siguientes tres opciones:

- *Ocio, recreo o vacaciones*
- *Negocio, motivos profesionales*
- *Otros motivos*

El resultado se expresa en la siguiente tabla, suma de los 12 meses del año 2021 y el primer trimestre de 2022:

Tabla 3.1 Pasajeros por motivo de viaje.

Motivo	Pasajeros
Total	40.856.225
Ocio, Recreo y Vacaciones	34.340.523
Negocio, motivos profesionales	2.803.363
Otros motivos	3.712.342

Esto implica algo que pone de manifiesto la correlación de la que se hablaba en las dos últimas variables tratadas: Un 84,1 % de los pasajeros que viajaron en España en los últimos 15 meses, lo hicieron por motivos de ocio:

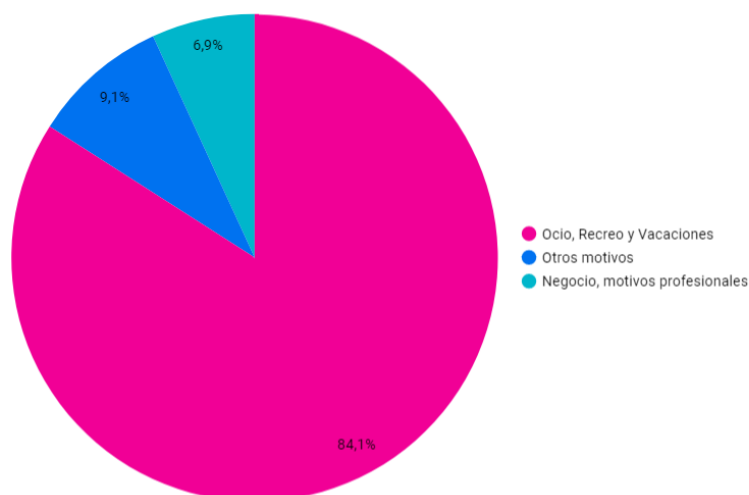


Figura 3.1 Porcentaje de pasajeros en los últimos 15 meses por motivo de viaje (Elaboración propia).

3.1.6 Otras variables demográficas

En este aspecto se comentan una serie de variables que, por simplicidad, no se tratarán en las ecuaciones del modelo pero que si influirían en una situación real.

- *Área de interés cultural:* Partiendo de lo comentado en las dos variables relacionadas con el turismo, si una de las áreas origen-destino es de interés cultural, en el sentido de alojar monumentos patrimonio de la humanidad o atractivos de índole cultural, esto beneficiará a la demanda de transporte. Dicha

variable influirá directamente en el número de plazas hoteleras, PIB y porcentaje activo de la población, por lo que al tener en cuenta las anteriores el error que cometemos al no tener en cuenta esta variable, se verá salvaguardado.

- *Climatología*: Al igual que en el caso del área de interés cultural, el clima puede ser, y es, un atractivo que influye sobre la demanda. Puesto que esta variable climática influirá en el resto de variables tratadas, el error cometido al despreciar este atributo será menor.
- *Nivel educativo*: El porcentaje activo de la población está correlacionado con esta característica por lo que al tener en cuenta dicho porcentaje, de nuevo el error que cometemos al evitar esta variable es menor.

3.2 Variables aeroportuarias de origen y destino

Las variables a las que se llama aeroportuarias en este estudio son aquellas que hacen referencia a una característica intrínseca del aeropuerto. Se tratarán en este apartado variables técnicas que serán limitantes a la operación y por otro lado variables relativas a las operaciones u operadores que realizan su actividad en dichos aeropuertos.

3.2.1 Capacidad aeroportuaria

OACI la define como: " Conjunto de instalaciones aeroportuarias necesarias y como mínimo igual a la demandada que deben ajustarse a la oferta, para dar un servicio adecuado."[28]

Técnicamente lo definiremos como una medida del rendimiento, o sea es el número máximo de operaciones que puede llevarse a cabo en una hora. Será una variable que limitará o no la operación una vez hayamos predicho la demanda y formulando la función de beneficio, actuando como un calibre pasa-no pasa.

3.2.2 Clave de referencia de aeródromo

OACI la define como: "El propósito de la clave de referencia es proporcionar un método simple para relacionar entre sí las numerosas especificaciones concernientes a las características de los aeródromos...La clave está compuesta de dos elementos que se relacionan con las características y dimensiones del avión. El elemento 1 es un número basado en la longitud del campo de referencia del avión y el elemento 2 es una letra basada en la envergadura del avión y en la anchura exterior entre las ruedas del tren de aterrizaje principal".

Para el estudio será una variable limitante en el sentido de que la aeronave que se seleccione para satisfacer la demanda de transporte entre los dos aeropuertos, tendrá que cumplir con la clave de aeródromo correspondiente.

En el Apéndice 1 del Manual de Diseño de Aeródromos de OACI se nos ofrece una clasificación de las aeronaves de cada tipo asociadas a la clave de aeródromo correspondiente.

3.2.3 Número de rutas activas

Se denominará una ruta activa a un trayecto aéreo operado con normalidad y determinada frecuencia entre un aeropuerto de origen y otro de destino, haciendo referencia exclusivamente al transporte de pasajeros. Esta variable será un conteo de rutas distintas desde un aeropuerto.

Tal y como se argumentó en el capítulo 1, "se considera ruta activa aquella con más de 5.000 pasajeros en el año actual". Se podría hacer un análisis más detallado contando rutas de menos pasajeros pero se pone el límite en 5000 pasajeros anuales puesto que se tratan decisiones económicas de volumen. La importancia de esta variable es vital tanto por el volumen de operaciones y la capacidad del aeropuerto, como por la competencia de oferta que puede existir.

3.2.4 Número de churn de rutas

Se denominará una ruta "churneada" a un trayecto aéreo que era operado con normalidad y determinada frecuencia entre un aeropuerto de origen y otro de destino, y en el periodo actual ya no lo hace.

Esta variable será de importancia pues permitirá determinar, en primer lugar, si la ruta que se estudia ya fracasó y, en segundo lugar, permite detectar si es un aeropuerto con poca estabilidad de rutas (muchas rutas churneadas).

3.2.5 Número de aerolíneas que operan

Se define como el número de aerolíneas que tiene alguna ruta activa en el aeropuerto de origen o destino.

Será una variable que tendrá importancia tanto por la componente de competencia, en el sentido de que si hay muchas aerolíneas operando, el posible pasajero tendrá más opciones en las que decidir (aunque por otro lado la posibilidad de combinaciones será mayor), tanto como por el volumen de operaciones que tenga el aeródromo.

3.2.6 Índice de accesibilidad al aeropuerto

Se definirá Índice de accesibilidad como una variable que medirá el nivel de accesibilidad al aeropuerto distinguiendo tres categorías: A: Muy accesible, B: Accesible, C: Poco accesible.

La accesibilidad al aeropuerto es un criterio históricamente relevante. Bao, Hua y Gu (2016) realizan un estudio sobre el impacto de la accesibilidad a la hora de elegir aeropuerto en el mercado Chino, llegando a la conclusión de que el volumen de tráfico podría incrementarse hasta en un 2 % con sólo un 1 % de aumento de los niveles de accesibilidad. Para ello definen una función de accesibilidad, compuesta por el coste monetario de llegar al aeropuerto y el coste en tiempo de llegar al aeropuerto, en la que un aeropuerto será más accesible mientras menor sea dicho coste.[29]

Por simplificación en este estudio no se utilizará dicho tipo de funciones, se designarán tres categorías en función de las opciones de transporte y su coste. Estableceremos los siguientes modos de acceso:

- *Coche propio*
- *Taxi-VTC*
- *Tren*
- *Metro*
- *Autobús*

La clasificación que haremos será de la siguiente modalidad:

Tabla 3.2 Categorías de accesibilidad.

Categoría	Métodos de llegada
A: Alta accesibilidad	Tren o Metro, Autobús y Carreteras desarrolladas
B: Accesibilidad media	Autobús y Carreteras desarrolladas
C: Baja accesibilidad	Carreteras no desarrolladas

Por tanto, esta variable será de importancia en el sentido de que la demanda de una ruta aumentará si la accesibilidad es elevada y disminuirá en el caso contrario.

3.2.7 Evolución del volumen de pasajeros

Se define Evolución del volumen de pasajeros como el porcentaje de crecimiento del número de pasajeros en un aeropuerto con respecto a un periodo anterior.

Tal y como se planteó en el apartado de introducción, podremos hacer estimaciones de cómo evolucionará el tráfico de una región, haciendo una regresión de los índices de evolución de periodos anteriores.

3.3 Variables de ruta

Las variables a las que se llama de ruta en este estudio son aquellas que hacen referencia a una característica exclusiva de la combinación aeropuerto de origen y aeropuerto destino.

3.3.1 Distancia Destino-Origen

Se define cómo la distancia en kilómetros en el mapa entre las coordenadas del aeropuerto de origen y las coordenadas del aeropuerto de destino.

Puesto que la distancia volada en una ruta aérea difiere de dicha distancia absoluta en kilómetros, en esta variable se podría haber tenido en cuenta el cálculo de la ruta en plan de vuelo. Por simplicidad, haremos un pequeño factor de corrección que aumente la distancia absoluta del mapa, atendiendo a que la distancia siguiendo una carta de navegación aérea será menor o igual que a la distancia total.

Se propone para observar dicha diferencia el ejemplo de la ruta SVQ (Sevilla) - PMI (Palma de Mallorca): En primer lugar, medimos la distancia ortodrómica (o distancia en línea recta en el mapa, sin tener en cuenta ninguna ruta aérea). El total es de 787,07 km.



Figura 3.2 Distancia ortodrómica SVQ-PMI. Fuente: Google Maps.

A continuación, haciendo uso del software "Flight Routes Calculator" by Aviapages[30], que hace una simulación de la ruta aérea viable más directa, obtenemos la distancia de 803 km.



Figura 3.3 Ruta de vuelo SVQ-PMI. Fuente: Aviapages.

Esta variable afectará tanto al consumo de combustible, como a la demanda de tráfico aéreo y es una de las variables usadas en casi la totalidad de los modelos.

3.3.2 Tiempo de vuelo

Se define cómo el tiempo en horas, minutos y segundos transcurrido desde el comienzo del rodaje en el aeropuerto de origen hasta el estacionamiento en el aeropuerto de destino.

Podría parecer que esta variable será totalmente dependiente de la distancia entre aeropuertos, pero dependerá en gran parte de la aeronave y las características de operación de manera que no tiene por qué ser proporcional a la distancia en kilómetros entre ambos aeródromos. De ahí que también se tenga en cuenta para el estudio.

3.3.3 Restricciones diplomáticas del vuelo

Se define como la variable que detecta si existen o no restricciones de vuelo (cierres del espacio aéreo) en una determinada región. Esta variable será binaria y será negativa en el caso de que el espacio aéreo sea operable en todas y cada unas de las regiones de la ruta o positiva en el caso de que alguna de las regiones esté restringida.

Como ejemplo, el pasado febrero de 2022, el Ministerio de Asuntos Exteriores español, cerro su espacio aéreo a las aerolíneas rusas debido a su posición en el conflicto ruso-ucraniano. En el caso de que en nuestro sistema una aerolínea rusa quisiese operar en un aeropuerto español, esta variable restringiría dicha posibilidad.[31]

3.3.4 Histórico de vuelos Origen-Destino

Se define como el número de vuelos comerciales de pasajeros que ha habido en un periodo determinado entre los aeropuertos del par origen-destino.

Dicha variable nos permitirá definir si podemos tomar como referencia datos antiguos de otras variables relativas al par origen-destino, o si, por el contrario, si el número de vuelos no es significativo, no las podremos tener en cuenta.

3.3.5 Historial de búsquedas online

Se define como el número de búsquedas de la combinación origen-destino en navegadores de internet.

Con el avance de la tecnología, la venta de tickets online supone un elevado porcentaje de las ventas para las aerolíneas. Un perfecto indicador de la posible demanda, será el número de clientes que se han interesado por esta ruta en un periodo determinado, tecleando en el buscador la combinación seleccionada.

3.4 Variables de competencia

Las variables a las que se llama de competencia en este estudio son aquellas que hacen referencia tanto a la competencia con otros medios de transporte, cómo con otras aerolíneas como con otros aeropuertos. Son variables claves en las ecuaciones de nuestro sistema basadas en los modelos de elección Logit.

3.4.1 Variables de competencia de otros medios de transporte

Tal y como se propuso en el ejemplo que se empleó para la explicación de la teoría del modelo de elección , hay una serie de variables que son claves para definir la probabilidad de que un potencial pasajero escoja el avión en vez de otro tipo de transporte (ferrocarril, por carretera, marítimo...). Estas variables son relativas a las características de la ruta en cada uno de los modos de transporte y por simplicidad compararemos entre: Transporte Aéreo, Ferrocarril, Autobús y Transporte por carretera.[22] Las variables que definiremos en esta tipología serán:

- *Precio de las alternativas*: Se define como el valor en número de unidades monetarias del recorrido del trayecto en un modo de transporte determinado. Esta variable será clave, en tanto en cuanto, el potencial pasajero definirá su alternativa en relación a sus límites económicos. Aparece aquí un concepto económico clave que será WTP: "Willingness to Pay", que es el precio máximo que un cliente está dispuesto a pagar por un servicio de transporte.
- *Tiempo de viaje*: Se define como el tiempo en horas, minutos y segundos transcurrido en el trayecto en cada una de las alternativas. Análogamente al Precio, el tiempo que haya que invertir en el trayecto definirá muchas veces la elección del comprador.
- *Frecuencia de la alternativa*: Se define como el número de servicios prestados por el mismo tipo de transporte en una franja de tiempo determinada. Es de importancia en la selección de alternativas pues una mayor flexibilidad horaria hará que el servicio atraiga a más potenciales clientes.

Existen además variables que por simplicidad no tendremos en cuenta, como podrían ser variables relativas a la comodidad del servicio, a la puntualidad o a la accesibilidad.

3.4.2 Variables de competencia de otras aerolíneas

Al igual que ocurre con el caso de las distintas modalidades de transporte, hay otra serie de variables que definirán la probabilidad de que se elija una aerolínea u otra para realizar el trayecto:

- *Precio medio histórico de ticket*: Se define como el valor en número de unidades monetarias de media de los trayectos similares realizados por dicha aerolínea en un periodo de tiempo determinado.

Esta variable, dada la volatilidad del mercado actual, no será tan definitoria en algunos casos, pero si que es un indicativo a la hora de definir la demanda.

- *Aerolínea low-cost vs tradicional*: Se define como la tipología de modelo económico de cada aerolínea distinguiendo entre low-cost y tradicional. Será una variable binaria y su importancia en el modelo vendrá dada por la tendencia actual del pasajero en una serie de tipos de vuelos por usar las aerolíneas low-cost.[32]

- *Flexibilidad en el ticket*: Se define como la tipología del ticket estándar que ofrece esa aeronave. Será una variable binaria y vendrá muy correlacionada a priori con el hecho de que la aerolínea sea low-cost o tradicional, puesto que las aerolíneas tradicionales suelen ofrecer condiciones de ticketing más flexibles.[33]
- *Antigüedad en la ruta*: Se define como el número de años que una aerolínea lleva operando un trayecto. Será una variable sólo particular de aquellas combinaciones que ya tengan operativo el par de aeropuertos que se estudie.

Otras variables que no se tendrán en cuenta en el modelo por su poco peso serían el componente tradicional que ciertas aerolíneas tienen en ciertas regiones o variables más concretas de cada cliente como la posibilidad de pertenecer a algún programa de afiliación con la compañía.

Además de estas variables, otro tema clave son los slots repartidos en las conferencias semestrales. El hecho de que una compañía posea o no un "slot" o "espacio" tanto en el aeropuerto como en el espacio aéreo, puede ser otro factor determinante que no se tiene en cuenta en este estudio por el hecho de la dificultad y privacidad de dicha información.

3.4.3 Variables de competencia de otros aeropuertos

Por último, en esta sección se plantea una variable clave, que junto con variables ya definidas sobre las características intrínsecas de un aeropuerto, indicarán la probabilidad de que se elija un aeropuerto u otro.

- *Número de aeropuertos cercanos al origen o destino*: Se define como el número de aeropuertos que hay en un radio de kilometraje determinado centrado en el aeropuerto de origen o destino.

Para definir esta variable haremos uso del software "Air Calculator", de modo que seleccionando un código IATA del aeropuerto que estemos estudiando, ponemos como ejemplo el Aeropuerto de Sevilla (SVQ) y el software nos devuelve la siguiente información, cuyo tratamiento se comentará en apartados posteriores:[30]

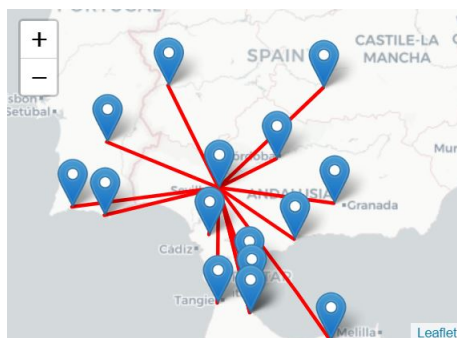


Figura 3.4 Mapa de aeropuertos cercanos al aeropuerto de Sevilla. Fuente: Air Calculator.

Rank	Distance & Direction	Airport Name & IATA Code	Airport Location	Further Actions
1	47.36 miles (76.22 km) S	Jerez Airport (XRY)	Jerez de la Frontera, Spain	
2	64.87 miles (104.40 km) ENE	Córdoba Airport (ODB)	Córdoba, Spain	
3	52.66 miles (84.76 km) SSE	Gibraltar International Airport (GIB)	Gibraltar	
4	52.70 miles (84.82 km) ESE	Málaga Airport (AGP)	Málaga, Spain	
5	110.41 miles (177.69 km) SSE	Ceuta Heliport (JCU)	Ceuta, Spain	
6	113.47 miles (182.60 km) NNW	Badajoz Airport (BJZ)	Badajoz, Spain	
7	116.85 miles (188.06 km) S	Tanger Ibn Battouta Airport (TNG)	Tanger, Morocco	
8	117.10 miles (188.45 km) WSW	Faro International Airport (FAO)	Faro, Portugal	
9	117.66 miles (189.36 km) E	Granada Airport (GRX)	Granada, Spain	
10	120.11 miles (193.30 km) WNW	Béja Airport (BYJ)	Béja, Portugal	
11	130.04 miles (209.27 km) SSE	Sania Ramel Airport (SIL)	Sidi l'ahj, Morocco	
12	130.04 miles (209.27 km) SSE	Sania Ramel Airport (TTU)	Tetouan, Morocco	

Figura 3.5 Listado de aeropuertos cercanos al aeropuerto de Sevilla. Fuente: Air Calculator.

3.5 Variables de costes de operación

Las variables a las que se llama de costes de operación en este estudio son aquellas que hacen referencia al conjunto de costes que conllevan para una compañía la operación de una ruta entre dos aeropuertos. Estas variables intervendrán al final del modelo, cuando se plantee, una vez seleccionado un par origen-destino, si el beneficio esperado comparado con los costes será adecuado para la compañía o no.

Los costes de operación son diversos y variables en función de las leyes aéreas del país de operación, el tipo de compañía, etc. Por simplicidad, establecemos una serie de variables que permitirán estimar de manera sencilla una cuenta del total de coste de la operación. AEA realizó un estudio sobre la estructura característica de costes en una compañía tradicional, sobre el que nos guiaremos.[34] Se resume en la siguiente imagen:

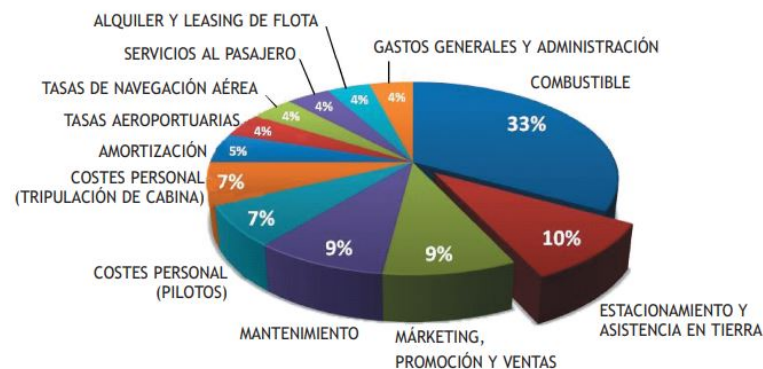


Figura 3.6 Estructura característica de costes en una aerolínea tradicional. Fuente: AEA.

Se han dividido en estas 4 subcategorías:

3.5.1 Costes relativos a las tasas aeroportuarias e impuestos

Estas tasas e impuestos se dividen en las tasas que el operador tiene que pagar por cada pasajero, en las tasas que pagará según el tipo de aeronave que opere, y en una serie de impuestos que el estado establece y que son ajenos a la operación.[35]

- *Tasas de pasajero*: Se define como la cantidad monetaria por pasajero debida al uso de la infraestructura y al servicio de seguridad.
- *Tasas de aeronave*: Se define como la cantidad monetaria por aeronave por los conceptos de: navegación aérea, aterrizaje, estacionamiento, ruido y emisiones, provisión de energía y uso de pasarela telescópica.

- *Tasas de aeronave*: Se define como la cantidad monetaria por aeronave por los conceptos de: navegación aérea, aterrizaje, estacionamiento, ruido y emisiones, provisión de energía y uso de pasarela telescópica.

3.5.2 Costes relativos a Handling

Se define Handling como Asistencia en Tierra a Aeronaves y Pasajeros.

- *Handling en rampa*: Se define como la cantidad monetaria por el siguiente servicio a la aeronave: carga y descarga, control de equipajes, provisión de agua potable, provisión de energía eléctrica, escaneado, arranque de motores, personal cualificado, transferencia de catering, etc
- *Handling al pasajero*: Se define como la cantidad monetaria por los siguientes servicios al pasajero: check-in, ticketing, utilización de salas, etc[36]

3.5.3 Costes relativos al combustible

Se define como la cantidad monetaria que cuesta el combustible requerido para operar un cierto aeronave, con un peso determinado, entre un origen y un destino.

3.5.4 Costes relativos al personal

- *Personal en cabina*: Se define como el coste por operación destinado al equipo de cabina.
- *Pilotos*: Se define como el coste por operación de los pilotos.

Otras variables de costes que no se tienen en cuenta son, por ejemplo, los costes relativos a las negociaciones de compra-venta de slots por parte de la compañía, o los costes relativos a las sanciones que los operadores pueden tener.

4 La estructura de datos

La estructura de datos es un tema tanto o más definitorio que el propio modelo. En este apartado se expondrán los procedimientos del tratamiento de estos datos realizando un análisis más exhaustivo del primer bloque del diagrama de flujo propuesto en la figura 2.1 y que se presenta ampliado en este esquema:

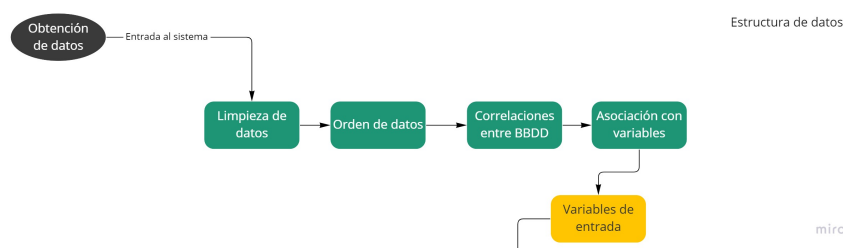


Figura 4.1 Diagrama de bloques de la infraestructura de datos (Creación propia).

Se comenzará, por tanto, por la descripción de las bases de datos escogidas, definiendo el porqué de dicha elección y las características propias de cada una de ellas. Se proseguirá con la definición de una base de datos conjunta en la que se unificarán dichos datos, se hará referencia a las herramientas técnicas empleadas para ello y se terminará con la especificación de los nexos de unión entre las fuentes de datos originales que, mediante relaciones, darán lugar a esta base de datos final.

4.1 Fuentes de adquisición de datos

Number of publications related data science and analytics in aviation.	
Year	Number of Publications
2019-Oct	373
2018	404
2017	290
2016	182
2015	122
2014	65
2013	24
2012	5
2011	2
2010	2
2009	2

Figura 4.2 Evolución sobre las publicaciones sobre datos en aviación. Fuente: Transportation Research.

El tratamiento de fuentes de datos por parte de agentes aeronáuticos mediante técnicas modernas como big data o blockchain ha sido un asunto emergente en los últimos años en el mundo de la aeronáutica. Prueba de ello es este estudio que refleja el aumento de publicaciones sobre este ámbito desde 2009.[37]

Asuntos como predicciones de retrasos, de demanda, optimización de operaciones de handling o gestión de overbooking, son resueltos con el tratamiento de fuentes de datos aportando gran valor a la toma de decisiones.

Es el momento de particularizar sobre las fuentes de datos que serán empleadas. Partiendo sobre el siguiente esquema se dividirá el apartado en las seis subsecciones que se plantean:

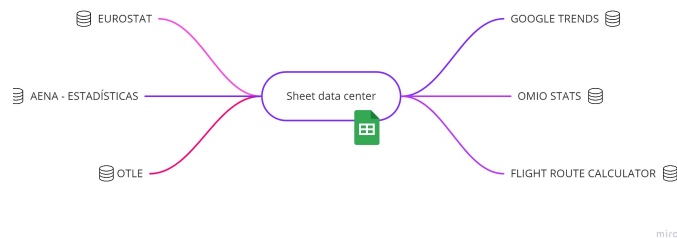


Figura 4.3 Esquema bases de datos. Elaboración propia.

El principal objetivo será la creación de una única base de datos conjunta (lo que en la figura se indica como sheet data center), que será una base de datos en forma de hoja de cálculo con 6 grandes grupos formados por los datos obtenidos.

Antes de continuar es preciso definir la nomenclatura de las etiquetas de unión entre todas las componentes de las bases de datos. Como premisa única y principal se establece que todos y cada uno de los registros de datos, deben contener la etiqueta que definiremos como código OACI del aeropuerto. En el caso de variables exclusivas de un aeropuerto el código será único y en el caso de ser una variable relativa a la combinación el código estará compuesto por el código del primer aeropuerto, un guión y el código del segundo aeropuerto.

IATA define un código determinado y distinto para cada aeródromo, formado por 3 letras, con el fin de establecer un marco común de nomenclatura. OACI propone también otro código de cuatro letras pero en este trabajo se utilizará el indicado por IATA por simplicidad.[38]

4.1.1 Eurostat

Eurostat es la Oficina Estadística de la Unión Europea. Ofrecen una base de datos de diversas cuestiones básicas según numerosos criterios y es el organismo europeo con más fiabilidad en el ámbito estadístico.[39] Puesto que para un primer modelo como el que se plantea sería costoso hacer una base de datos mundial, en este aspecto nos limitaremos al ámbito europeo, aunque sería perfectamente ampliable al ámbito global.

Dentro de estas amplias posibilidades de datos, encontramos una biblioteca de datos con la región como dimensión. Eurostat presenta el concepto de NUTS con el fin de establecer una norma estadística común en Europa. NUTS (Nomenclatura de Unidades Territoriales Estadísticas) es una clasificación del territorio europeo dividida a su vez en 4 niveles[40]:

- **NUTS-0:** Los países son la unidad de separación.
- **NUTS-1:** Las unidades administrativas con un número de habitantes de entre 3 y 7 millones es la unidad de separación.
- **NUTS-2:** Las unidades administrativas con un número de habitantes de entre 800000 y 3 millones es la unidad de separación.
- **NUTS-3:** Las unidades administrativas con un número de habitantes mayor de 150000 es la unidad de separación.

Por ejemplo, en el caso de España, las comunidades autónomas serían la separación en NUTS-3. El porqué de haber escogido dichas regiones para la división recae, además de en la sencillez, en que en varios lenguajes de programación existen bibliotecas con dicha clasificación predefinida por lo que a la hora de combinar o trasladar nuestro sistema a un software en alguno de estos lenguajes, esta implementación estaría ya realizada.

En las siguientes figuras se determinan dichas divisiones en el mapa[41]:

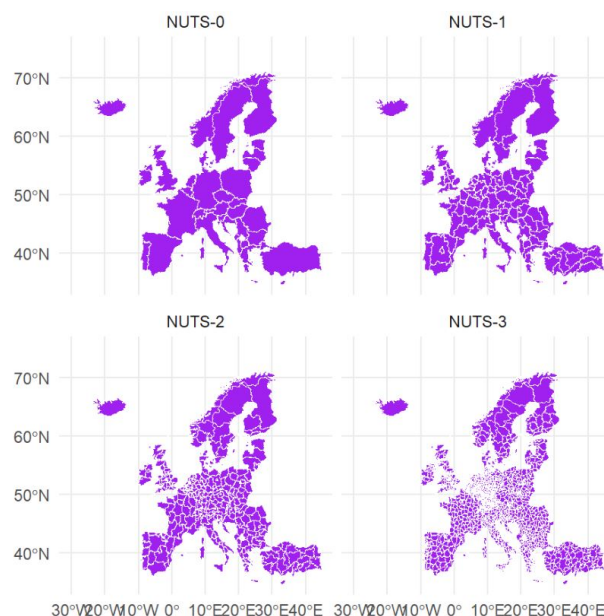


Figura 4.4 Definición regiones NUTS. Fuente:Rubén F. Bustillo-Amazon Web Services.

Se escogerá como región de trabajo la región NUTS-2 que es la más extendida, de manera que los registros de datos relativos a estas bases de datos tendrán como etiquetas el código IATA y el código de la región NUTS-2 que se estime correspondiente al aeropuerto. De esta fuente, obtendremos las siguientes bases de datos que se emplearán principalmente en el apartado de las variables demográficas:

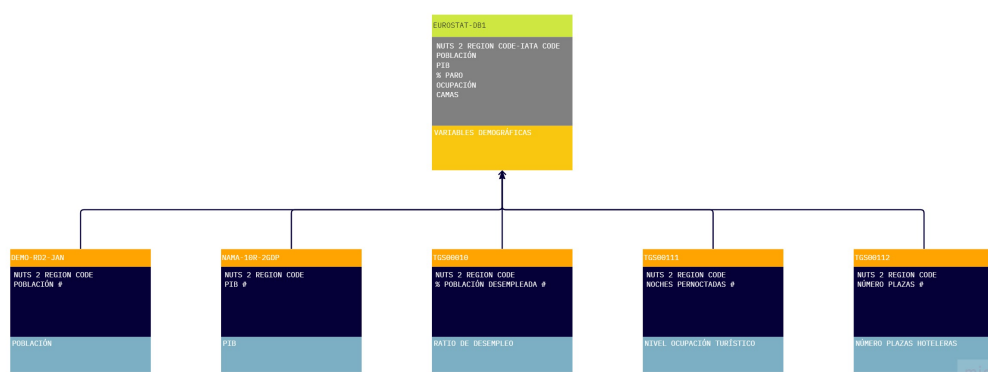
- *"Population on 1 January by age, sex and NUTS 2 region"*: Base de datos que dará la población en cada región NUTS en el 1 de enero de los años que se seleccionen.
- *"Gross domestic product (GDP) at current market prices by NUTS 2 regions"*: Base de datos que dará el producto interior bruto en cada región NUTS en los años que se seleccionen.
- *"Unemployment rate by NUTS 2 regions"*: Base de datos que dará el porcentaje de población desempleada sobre la población activa en cada región NUTS en los años que se seleccionen.
- *"Nights spent at tourist accommodation establishments by NUTS 2 regions"*: Base de datos que dará el número de noches pasadas en un establecimiento turístico en cada región NUTS en los años que se seleccionen. Eurostat define noche pernoctada como: "Se entiende por noche pernoctada cada noche que un huésped/turista (residente o no residente) pasa efectivamente (duerme o se aloja) o está registrado (siendo innecesaria su presencia física) en un establecimiento de alojamiento turístico."
- *"Number of establishments and bed-places by NUTS 2 regions"*: Base de datos que dará el número de camas disponibles en establecimientos turísticos en cada región NUTS en los años que se seleccionen.

Cada base de datos concreta, poseerá un código proporcionado por Eurostat para su búsqueda online y una fecha de última actualización, así como una relación directa con una o varias de nuestras variables elegidas, que en este caso serán todas las variables demográficas. Presentamos dicha información en la siguiente tabla:

Tabla 4.1 Bases de datos de Eurostat.

Base de datos	Código	Fecha actualización	Variable
Population on 1 January by age, sex and NUTS 2 region	DEMO-R-D2JAN	02/06/2022	Población
Gross domestic product (GDP) at current market prices by NUTS 2 regions	NAMA-10R-2GDP	18/04/2022	PIB
Unemployment rate by NUTS 2 regions	TGS00010	07/06/2022	Ratio de desempleo
Nights spent at tourist accommodation establishments by NUTS 2 regions	TGS00111	15/06/2022	Nivel de ocupación turístico
Number of establishments and bed-places by NUTS 2 regions	TGS00112	15/06/2022	Número de plazas hoteleras

En resumen, en esta fuente de datos obtenemos 5 bases de datos que se asociarán entre sí de una manera sencilla puesto que compartirán el código de la región NUTS. El vínculo, al que de aquí en adelante se denominará "join" de esta primera fuente de datos con las siguientes, será como no puede ser de otra manera el código IATA, por lo que aquí se asociará código IATA con código de la región NUTS que se asocie a dicho aeropuerto. Esta "rama" de la estructura de datos queda recogida en el siguiente esquema:

**Figura 4.5** Conexiones Bases de datos Eurostat (DB1). Elaboración propia..

4.1.2 OTLE

El Observatorio de Transporte y Logística de España será la segunda fuente de datos que se expondrá. Es un organismo del Ministerio de Transportes, Movilidad y Agenda Urbana que tiene como objetivo poner a disposición del ciudadano datos sobre transporte en nuestro país.[42] Por simplicidad utilizaremos esta base de datos como primera aproximación a ciertas variables, aunque se conoce que dicha base de datos tendrá que complementarse con otras en el caso de aeropuertos extranjeros.

La operativa de esta fuente de datos es muy similar a los datos establecidos por Eurostat, con la diferencia de que nos ofrecen información directa sobre el aeropuerto y no sobre la región. OTLE nos ofrece infinidad de datos, pero por simplicidad y ajuste a un número de variables definido, tomaremos las siguientes bases de datos:

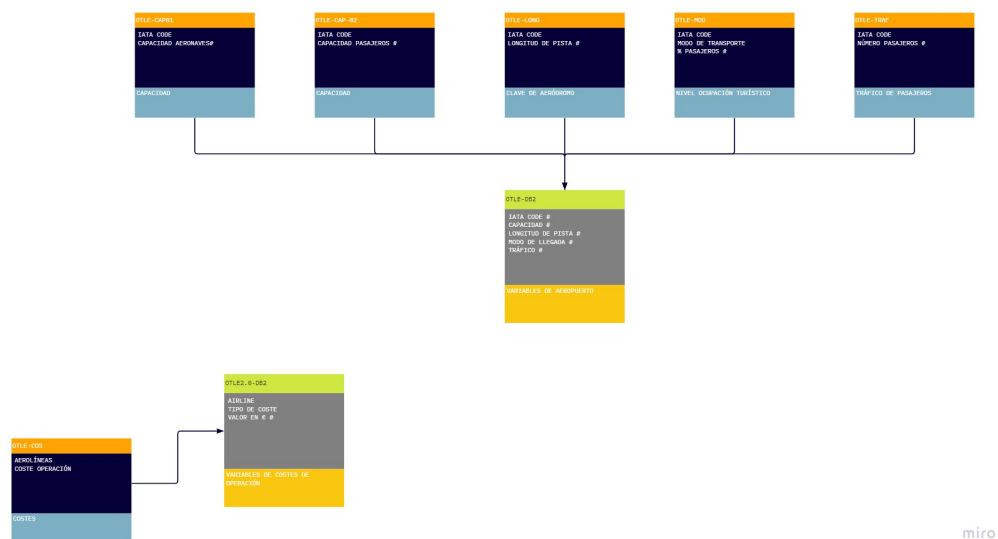
- *"Capacidad del campo de vuelo (aeronaves/hora) de los aeropuertos de la red de Aena por comunidad autónoma"*: Base de datos que dará la medida de capacidad del número de aeronaves que pueden operar (aterrizaje o despegue) en un aeropuerto en una hora.

- *"Capacidad de las terminales (pasajeros/hora) de los aeropuertos de la red de Aena por comunidad autónoma"*: Base de datos que dará la medida del número de pasajeros que una terminal puede acoger en un hora. Capacidad declarada para terminales: capacidad operativa máxima de terminales a aplicar en los procesos de coordinación de horarios. Este valor depende de un nivel de servicio preestablecido y en régimen de demanda continua. OTLE establece las siguientes consideraciones sobre esta base de datos: "-La metodología utilizada por Aena S.M.E., S.A. para la evaluación y determinación de los parámetros de capacidad de terminales que se aplican en los procesos de coordinación de horarios está basada en la utilización de herramientas de simulación y ha sido presentada a la industria (Comité Nacional de Coordinación de Horarios) y aceptada y reconocida por las compañías aéreas como método estándar a utilizar para la evaluación de capacidades en los aeropuertos españoles. -Los aeropuertos Madrid-Barajas, Barcelona-El Prat y Málaga no declaran capacidades operativas máximas de terminales hasta la fecha, dado que, en estos casos, el terminal no es limitante para la venta de slots. En estos aeropuertos la coordinación de slots se realiza con las capacidades declaradas del campo de vuelos."
- *"Longitud (metros) de las pistas de los aeropuertos de la red de Aena"*: Base de datos que dará la longitud de la pista o pistas de dicho aeropuerto.
- *"Modos de acceso (distribución porcentual) a los aeropuertos de la red de Aena"*: Base de datos que nos aporta el porcentaje de pasajeros que acceden desde cada modo de transporte al aeropuerto. OTLE establece las siguientes consideraciones sobre esta base de datos: "-Estos datos se han obtenido mediante encuestas efectuadas en los aeropuertos a pasajeros de vuelos comerciales. -Autobús de cortesía se refiere al servicio gratuito de autobús que hoteles o agencias de viajes ofrecen a los viajeros para el traslado desde o hasta el aeropuerto. -Los datos no están actualizados con la información del año 2020 por no haberse podido realizar encuestas a los pasajeros en los aeropuertos."
- *"Tráfico de pasajeros y mercancías en aeropuertos españoles (número de asientos y operaciones) por tipo de tráfico, aeropuerto y compañías más representativas"*: Base de datos que dará los volúmenes de tráfico en diversos aeropuertos.
- *"Costes de la aviación comercial de las principales compañías aéreas españolas"*: Base de datos que nos aporta información sobre los distintos costes según que aerolínea opere.

Estas bases de datos serán relacionadas principalmente con las variables aeroportuarias de origen. Estableceremos por tanto un código de nuestra base de datos y obtendremos la siguiente clasificación y esquema:

Tabla 4.2 Bases de datos de OTLE.

Base de datos	Código	Fecha actualización	Variable
Capacidad del campo de vuelo (aeronaves/hora) de los aeropuertos de la red de Aena por comunidad autónoma	OTLE-CAP01	Febrero 2022	Capacidad
Capacidad de las terminales (pasajeros/hora) de los aeropuertos de la red de Aena por comunidad autónoma	OTLE-CAP02	Febrero 2022	Capacidad
Longitud (metros) de las pistas de los aeropuertos de la red de Aena	OTLE-LONG	Febrero 2022	Aeródromo
Modos de acceso (distribución porcentual) a los aeropuertos de la red de Aena	OTLE-MOD	Noviembre 2020	Accesibilidad
Tráfico de pasajeros y mercancías en aeropuertos españoles	OTL-TRAF	Febrero 2022	Volúmen de pasajeros
Costes de la aviación comercial de las principales compañías aéreas españolas	OTL-COS	Febrero 2022	Costes de Operación

**Figura 4.6** Conexiones Bases de datos OLTE (DB2). Elaboración propia..

4.1.3 AENA Estadísticas

El portal de Estadísticas de Tráfico Aéreo en los aeropuertos de la red AENA[43], ofrece una plataforma interactiva que permite obtener datos de tráfico en tiempo real sobre los aeropuertos de dicha red. Se ofrece un Dashboard interactivo desde el que se exportarán los datos necesarios. En la figura siguiente se muestra dicha plataforma:

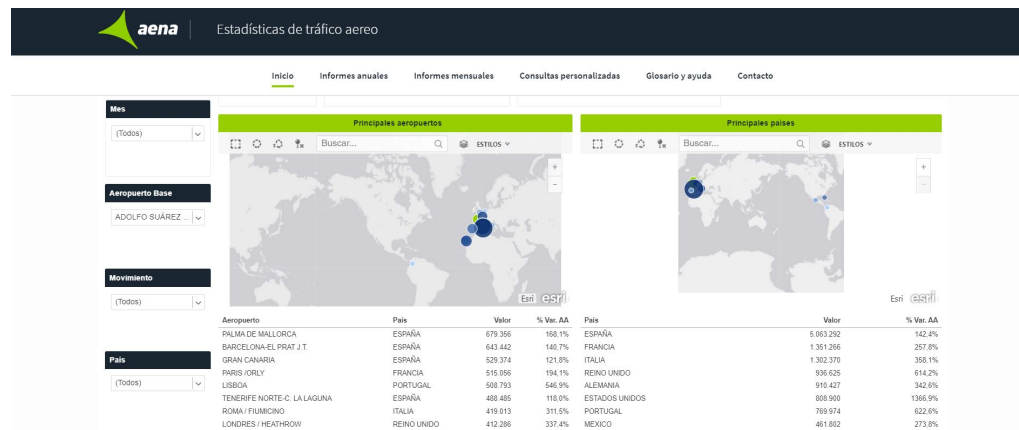


Figura 4.7 Dashboard Interactivo. Fuente: Aena.

Es claro que en este caso escogeremos el código IATA del aeropuerto como dimensión y obtendremos los siguientes listados de datos:

- *"Información de rutas por aeropuertos"*: Base de datos que dará el listado y número de rutas en las que interviene cada aeropuerto, así como el número de operaciones y pasajeros. Cabe decir que se podrían obtener más datos pero por simplicidad se escogen estas dos métricas.
- *"Operaciones O/D"*: Base de datos que dará el número de operaciones por par origen-destino.
- *"Información por compañía aérea"*: Base de datos que dará el número de operaciones que cada aerolínea opera en una ruta.

Estas bases de datos serán relacionadas principalmente con las variables aeroportuarias de origen como en el caso anterior y con las variables de ruta y competencia. Estableceremos por tanto un código de nuestra base de datos y obtendremos la siguiente clasificación y esquema:

Tabla 4.3 Bases de datos de AENA.

Base de datos	Código	Fecha actualización	Variable
Información de rutas por aeropuertos	AENA-NRUT	Abril 2022	Rutas competencia
Operaciones Origen-Destino	AENA-ORDES	Abril 2022	Misma Ruta y Churn de Rutas
Información por compañía aérea	AENA-COMP	Abril 2022	Aerolíneas competencia

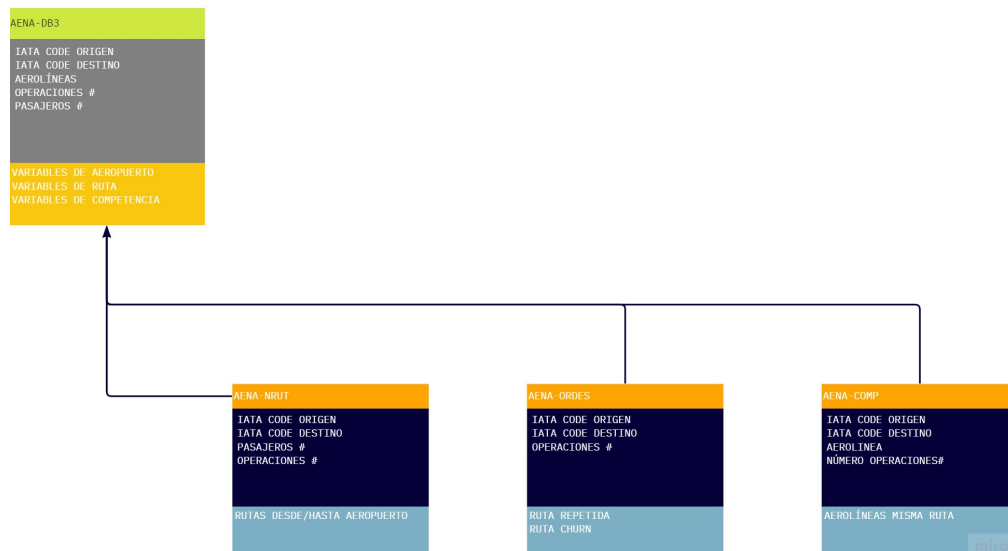


Figura 4.8 Conexiones Bases de datos AENA (DB3). Elaboración propia..

4.1.4 Google Trends

Es bien sabido que ,en la actualidad, Google es el navegador más usado por los usuarios de internet[44]. Sabiendo que ya en 2011 un 80% de los tickets de vuelos se adquirían de forma online[45], y que un 65 % de los usuarios utilizan Google Chrome para buscar sus vuelos, y puesto que adquirir datos de plataformas como Skyscanner es complicado y no existen muchos datos de libre acceso, se decide realizar un estudio de búsquedas en Google del par de aeropuertos origen-destino haciendo uso de Google Trends.

Google Trends es una herramienta de analítica de búsquedas de contenido del propio Google. Utiliza un muestreo significativo de las búsquedas para dar estadísticas de los temas más buscados por palabras. Se establecerá un criterio de búsqueda aeropuerto de origen, aeropuerto de destino y la palabra vuelo para comprobar que nivel de búsquedas tendrá. Cómo ejemplo pondremos "vuelo sevilla madrid" y lo compararemos con "vuelo madrid barcelona". La búsqueda sería tal y como en la siguiente imagen:[46]

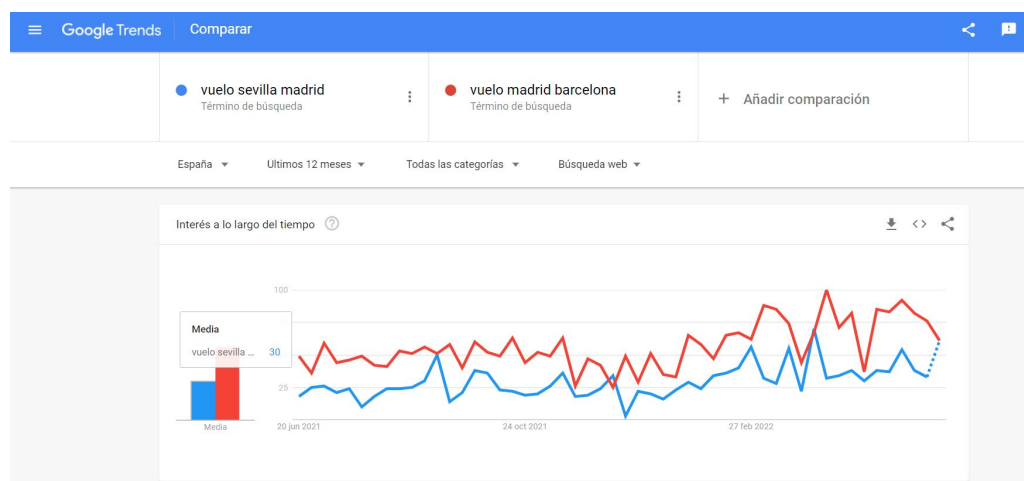
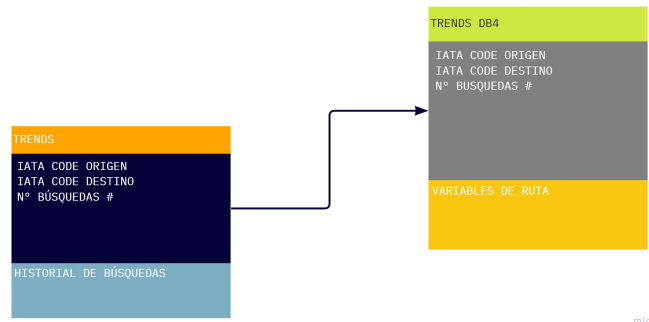


Figura 4.9 Comparativa búsqueda. Fuente: Google Trends.

Esto, por ejemplo, sería un indicativo de que en la comparativa, la ruta ,madrid-barcelona sería más demandada. Es por ello, que lo establecemos como una nueva base de datos en la que exportaremos un conjunto de búsquedas con esa combinación.

Tabla 4.4 Bases de datos de Google Trends.

Base de datos	Código	Fecha actualización	Variable
Búsquedas par origen destino en Google	TRENDS	Tiempo real	Búsquedas Internet

**Figura 4.10** Conexiones Bases de datos Google Trends (DB4). Elaboración propia..

4.1.5 Flight Route Calculator

Puesto que la información técnica sobre la ruta es crucial, en el sentido de que conocer la distancia que separa dos aeropuertos, el tiempo de vuelo o los litros de combustible necesarios será clave en el sistema, tener una base de datos que nos aporte esta información al elegir un origen y un destino, es imprescindible.

Flight Route Calculator es una herramienta de libre acceso de la empresa Aviapages, que nos permite hacer búsquedas de información de rutas. [30] La interfaz es sencilla e intuitiva, de manera que seleccionando las propiedades que en la imagen se exponen, se tendrá disponible la información:

Figura 4.11 Interfaz de búsqueda. Fuente: Flight Routes Calculator..

De manera que al introducir los requisitos de la operación que queramos estudiar obtendremos la siguiente lista de datos útiles:

- *"Airway distance"*: Base de datos que dará la distancia en kilómetros de vuelo entre un origen y un destino.

- "Great Circle distance": Base de datos que dará la distancia real entre origen y destino.
- "Airway time": Base de datos que dará el tiempo en la ruta aérea.
- "Great circle time": Base de datos que dará el tiempo que tardaría en línea recta.

Aunque en el estudio se utilizaran exclusivamente estos parámetros de distancia y tiempo, el software nos aportaría, en un ejemplo cualquiera, la siguiente información:

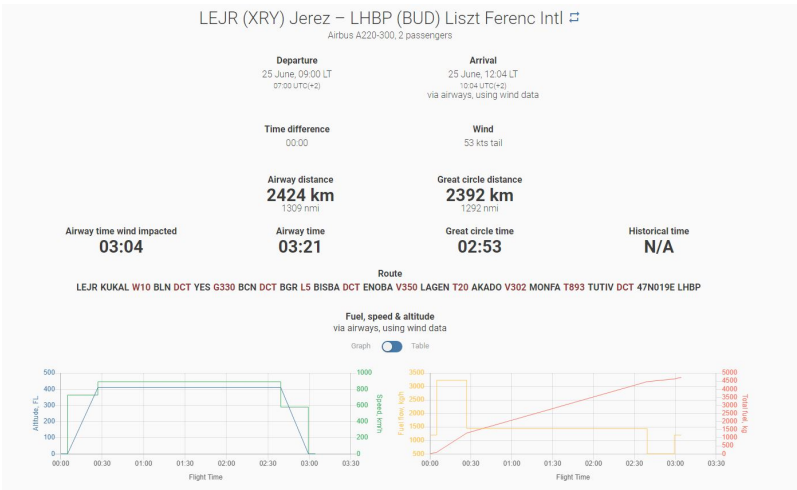


Figura 4.12 Interfaz de resultados. Fuente: Flight Routes Calculator..

Como en los casos anteriores, codificamos estas bases de datos. Por simplicidad, en el modelo tan solo tendremos en cuenta las distancias y tiempos de vuelo:

Tabla 4.5 Bases de datos de Flight Route Calculator.

Base de datos	Código	Fecha actuali- zación	Variable
Airway distance	FRC-DIS	Tiempo real	Distancia
Airway time	FRC-TIM	Tiempo real	Tiempo de vuelo

Y el esquema estructural sería:

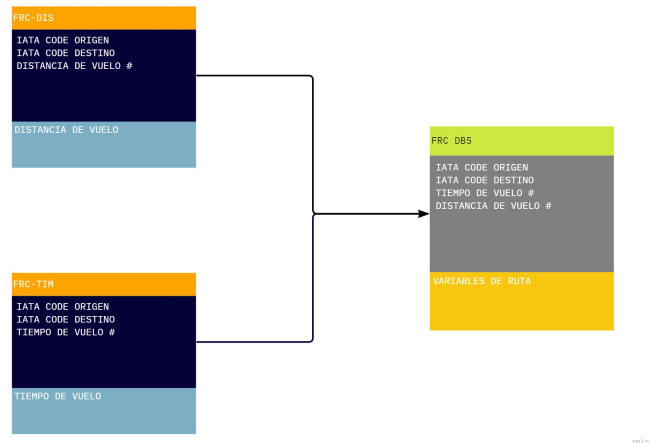


Figura 4.13 Conexiones Bases de datos Flight Route Calculator (DB5). Elaboración propia..

4.1.6 OMIO

Hasta este punto, la única información restante para completar las bases de datos del modelo es la de la competencia de transporte. Es cierto que encontrar una base de datos robusta que recoja la información sobre todos los modos de transporte disponibles en una ruta es complejo, y que en este estudio se busca sencillez en el proceso, por el tiempo y recursos de los que se dispone. Es por ello que se estudiará dicha competencia haciendo uso del almacenaje de datos que la interfaz web de buscador de opciones de transporte OMIO ofrece.[47]

En este caso no tenemos una base de datos como tal, si no un buscador en el que se tendrá que recaudar la información necesaria. El modo de búsqueda consistirá en comparar bus, tren y avión tanto en tiempo como en coste medio del billete:

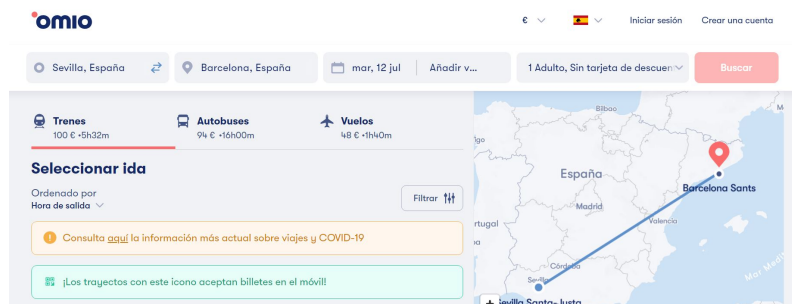


Figura 4.14 Interfaz de búsqueda Omio. Fuente: Omio..

De manera que al introducir el origen y destino:

- "*Precio Alternativa*": Base de datos que dará el precio de cada alternativa entre un aeropuerto de origen-destino.
- "*Tiempo Alternativa*": Base de datos que dará el tiempo de viaje en cada alternativa.
- "*Flexibilidad Alternativa*": Base de datos que determinará si el billete es o no flexible.

Resumiendo y codificando:

Tabla 4.6 Bases de datos de OMIO.

Base de datos	Código	Fecha actualización	Variable
Precio Alternativa	OMIO-P	Tiempo real	Precio
Tiempo Alternativa	OMIO-T	Tiempo real	Tiempo de viaje
Flexibilidad Alternativa	OMIO-F	Tiempo real	Flexibilidad

Cabe destacar en este punto, que se necesitará de un conjunto de fechas lo suficientemente amplio para poder hacer un precio medio fiable en el tiempo. Por último, se presentan las conexiones.

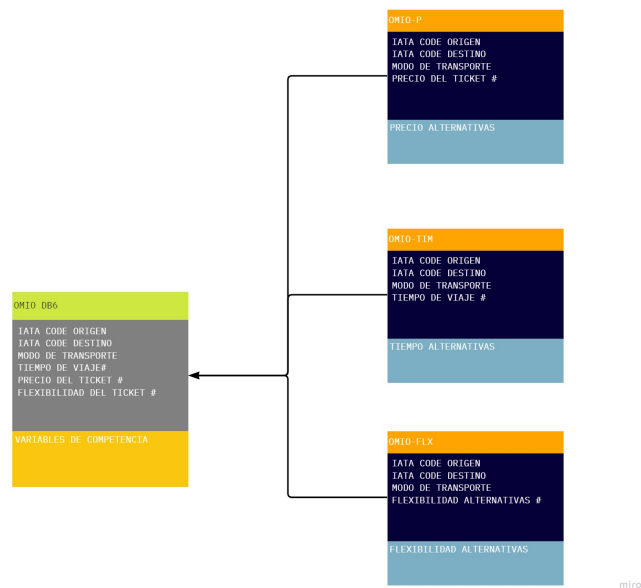


Figura 4.15 Conexiones Bases de datos OMIO (DB6). Elaboración propia..

4.2 Arquitectura de datos

Con el fin último de ofrecer unos outputs lo más reales, completos y actualizados posibles, establecer la estructura de relación entre las diferentes fuentes de datos que se han planteado en la sección anterior es un reto tan necesario como complejo. Una arquitectura de datos sólida se basa en la existencia de un conjunto reducido de dimensiones que hagan que un conjunto más amplio de métricas se relacionen entre sí. Es importante hacer una apreciación sobre los conceptos métrica y dimensión, que se usarán de aquí en adelante:

- **Dimensión:** Una dimensión es un atributo sobre un dato. Suelen ser atributos de tipo texto que aportan información no cuantitativa sobre un dato. Ejemplos: Nombre del alumno, país de origen, aerolínea, aeropuerto, aeronave...
- **Métrica:** Una métrica es una dimensión cuantitativa. Suelen ser atributos numéricos o booleanos, que miden una magnitud. Ejemplo: Nota en el examen, número de habitantes, euros de revenue, capacidad del aeropuerto, consumo de combustible...

Por tanto, el reducir lo máximo posible el número de dimensiones definitorias aportará mucha robustez al modelo. Las variables numéricas se corresponderán con las métricas, mientras que se establecerán las siguientes dimensiones únicas:[48]

Tabla 4.7 Dimensiones del modelo.

Dimensión
Código IATA aeropuerto origen
Código IATA aeropuerto destino
Código NUTS-2
Modo de transporte
Aerolínea

De manera que con 5 dimensiones estaremos codificando más de 30 métricas, algo que es indicativo de la robustez del modelo.

En lo que continúa de capítulo, en primer lugar se establecerán las herramientas que se han utilizado para la organización de datos del modelo y se hará un resumen estructural de relaciones.

4.2.1 Herramientas utilizadas

En cuanto a las herramientas que son necesarias para establecer un modelo sencillo de análisis de datos, no es necesario el manejo de lenguajes de programación complejos. Las herramientas utilizadas han sido:

- *Google Sheets*: Google Sheets es una herramienta de hojas de cálculo en línea de Google, que nos permitirá, en primer lugar, almacenar todos los datos y además, establecer conexiones con el resto de herramientas que usaremos.

Tal y como se definió en el apartado anterior, cada una de las bases de datos codificadas corresponderá en nuestro modelo a una hoja sencilla de cálculo que tendrá como columnas tanto las dimensiones como las métricas correspondientes. Posteriormente, las bases de datos pertenecientes a una misma fuente, se unificarán en una hoja matriz de la fuente. Es por ello que nuestro archivo de cálculo central, en el que se basará todo el cálculo, al que llamaremos "Sheet Data Center", se compondrá en primer lugar de 20 pestañas sencillas individuales (una por cada base de datos comentada), que a su vez se agruparán en 6 pestañas matrices. Posteriormente, habrá otra serie de pestañas que se comentarán en el apartado de outputs para los cálculos necesarios.[49]

- *Google Data Studio*: Google Data Studio es la herramienta de visualización y gestión de datos de Google, que nos permitirá combinar y visualizar los datos almacenados en el archivo maestro "Sheet Data Center".

Esta herramienta no sólo permitirá visualizar, con ella se establecerán los vínculos o joins (como si de SQL se tratase), que harán real la combinación entre fuentes de datos.

- *Miro*: Miro es una herramienta de "pizarra virtual" o esquemática que ha permitido la realización de los diagramas de flujo y bloques que se muestran en este documento.[50]
- *Matlab*: Matlab es un lenguaje de programación que ha servido como apoyo para realizar la regresión lineal de coeficientes en la estimación de demanda.[51]

4.2.2 Relaciones entre bases de datos

Los operadores de unión que se utilizar para establecer dichas relaciones son los siguientes:

- *Combinación Interna*: También llamada "Inner". Dadas dos bases de datos con una dimensión común, tiene que coincidir la misma dimensión en ambas bases de datos para que el registro se tenga en cuenta en la base de datos combinada.
- *Combinación Externa Izquierda*: También llamada "Left Outer". Dadas dos bases de datos con una dimensión común, se traen todos los registros de la base de datos izquierda y aquellos en los que coincida dicha dimensión de la base de datos derecha.

- *Combinación Externa Derecha*: También llamada "Right Outer". Dadas dos bases de datos con una dimensión común, se traen todos los registros de la base de datos derecha y aquellos en los que coincide dicha dimensión de la base de datos izquierda.
- *Combinación Externa Completa*: También llamada "Full Outer". Dadas dos bases de datos, se traen todos los registros de ambas.
- *Cruzada*: También llamada "Cross". Dadas dos bases de datos, se traen todas las posibles combinaciones de los registros de ambas.

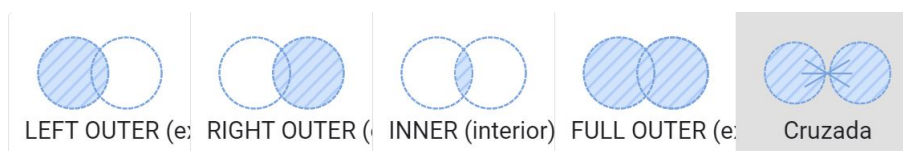


Figura 4.16 Joins o uniones en el modelo. Fuente: Google Data Studio..

De modo que elegiremos los siguientes tipos de nexos en las siguientes uniones[52]

Tabla 4.8 Tipos de uniones.

Bases de datos	Join
EUROSTAT DB1 - SHEET DATA CENTER	Inner
OTLE DB2-SHEET DATA CENTER	Inner
OTLE2.0 DB2-SHEET DATA CENTER	Full Outer
AENA BD3-SHEET DATA CENTER	Full Outer
TRENDS DB4-SHEET DATA CENTER	Inner
FRC DB5-SHEET DATA CENTER	Inner
OMIO DB6-SHEET DATA CENTER	Inner

En la siguiente página queda recogido el esquema total de la estructura de datos:

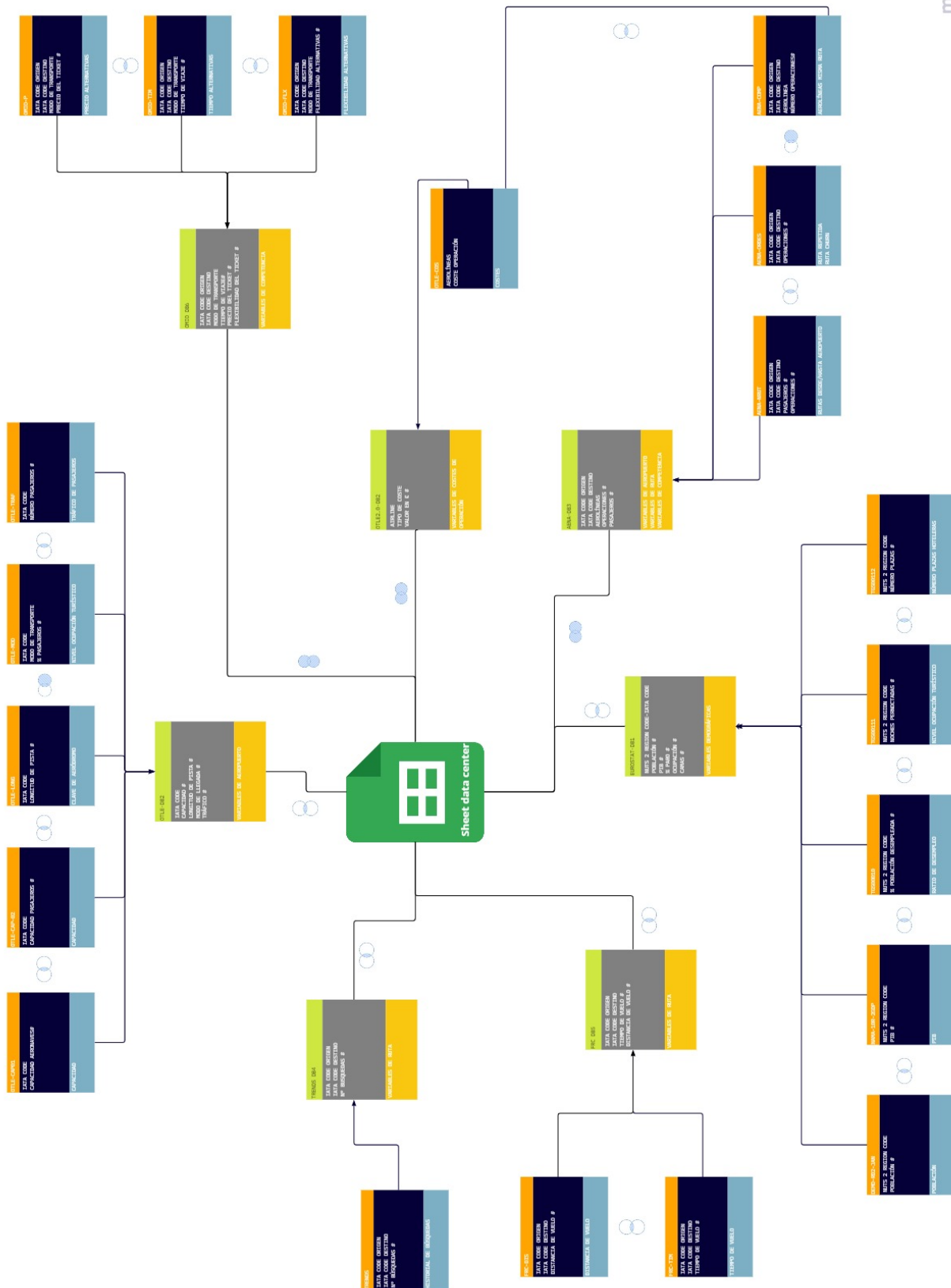


Figura 4.17 Estructura general de datos. Creación propia.

5 Formulación del modelo

Una vez definidas las variables y la estructura de datos, el bloque central del sistema consiste en el tratamiento e interrelación de estas variables en ecuaciones que, basadas en los principios estadísticos expuestos al principio de este trabajo.

El proceso será definir primero una ecuación que nos dé un valor aproximado de la demanda, así como ecuaciones que nos den una estimación de los porcentajes de elección de la competencia, para por último definir una función de beneficio que compare el coste de operación de una ruta, con los ingresos que acarrearía la demanda prevista. Además, se añadirán una serie de ecuaciones límites que deberán cumplirse para que la operación se pueda llevar a cabo.

Es obvio que el modelo tendrá limitaciones, estas serán expuestas en el apartado de validez del modelo, asumiendo en la definición de las ecuaciones ciertos conceptos.

Además de las variables ya definidas haremos uso de un gran número de coeficientes estadísticos (todos ellos expuestos en el glosario). A continuación, se resume este apartado en este diagrama de flujo:

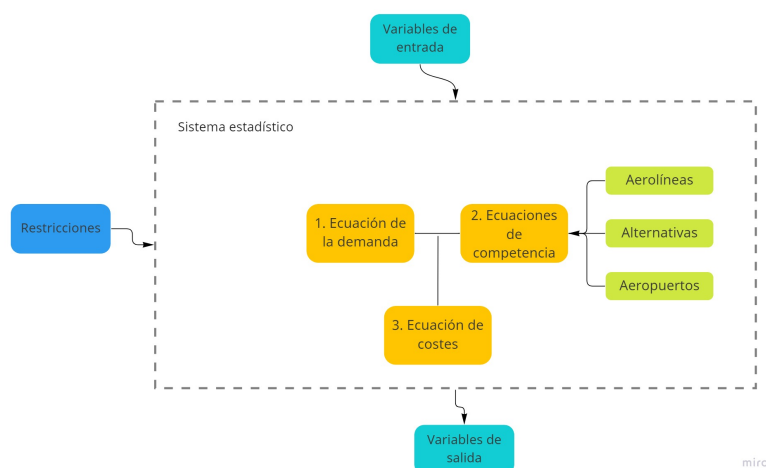


Figura 5.1 Diagrama de flujo del bloque estadístico del sistema (Elaboración propia).

5.1 Ecuación de la demanda de pasajeros

Tal y como quedó definido en el apartado de conceptos estadísticos, cuando se expuso la ecuación de demanda de elasticidades constantes, podríamos decir que nuestra ecuación de demanda tendrá el siguiente aspecto:

$$D = e^{(\beta_0)} * (X_1)^{(\beta_1)} * (X_2)^{(\beta_2)} * \dots * (X_N)^{(\beta_N)} \quad (5.1)$$

Dónde $X(1 \dots N)$ son las variables que intervienen en la demanda y $\beta(0 \dots N)$ las elasticidades de dichas respectivas variables.

El concepto de elasticidad de la demanda quedó definido en ese apartado, y también se vio que aplicando logaritmos en la ecuación se obtenía:

$$\log(D) = \beta_0 + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \dots + \beta_n \log(X_n) \quad (5.2)$$

Ahora bien, esta ecuación tendremos que particularizarla con las variables, datos y coeficientes del modelo. Definamos el conjunto de variables que intervendrán en mayor medida en esta ecuación de demanda y su notación (todas ellas están analizadas en detalle y descritas matemáticamente en el siguiente capítulo):

- P =Población
- PIB =Producto Interior Bruto
- Des =Porcentaje de desempleo
- Hot =Plazas Hostelería
- Och =Ocupación hostelera
- Dis =Distancia entre ciudades
- Bus =Búsquedas en Google

Igualmente, las elasticidades (coeficientes Beta de regresión) de estas

- β_P =Coeficiente de Población
- β_{PIB} =Coeficiente de producto Interior Bruto
- β_{Des} =Porcentaje de desempleo
- β_{Hot} =Plazas Hostelería
- β_{Och} =Ocupación hostelera
- β_{Dis} =Distancia entre ciudades
- β_{Bus} =Busquedas en Google

Nuestra ecuación entonces será:

$$\log(D) = \beta_P \log(P) + \beta_{PIB} \log(PIB) + \beta_{Des} \log(Des) + \beta_{Hot} \log(Hot) + \beta_{Och} \log(Och) + \beta_{Dis} \log(Dis) + \beta_{Bus} \log(Bus) \quad (5.3)$$

Es en este momento preciso ahondar en la definición o el ajuste de los coeficientes β , para lo que emplearemos un ajuste por regresión, usando para ello la variable de histórico de pasajeros.[53]

El procedimiento será extraer los coeficientes de regresión de nuestro conjunto de datos, haciendo uso de la función de Matlab: "regstat". Esta opera de la siguiente manera: En primer lugar, se aporta un vector de demanda de pasajeros conocida en años anteriores al que llamaremos "x". Estos valores del número de pasajeros podemos establecerlos en los últimos años en varias rutas (mientras más exacto y grande sea el conjunto de datos, mejor se ajustarán los coeficientes). En segundo lugar, aportamos una matriz a la que llamaremos "y", en la que añadiremos los datos correspondientes a dichos años de las variables indicadas anteriormente. Por último, seleccionaremos que tipo de regresión queremos que el programa ejecute, y este nos dará unos coeficientes, que serán los que utilicemos en la ecuación.[54]

```
Regresion=regstat(x,y,linear);
```

Por tanto, esta función, necesita como input un vector de demanda real, una matriz de variables y una modalidad de regresión. Veámoslo en detalle en un ejemplo con números ficticios:

Código 5.2 Código 2 Matlab.

```
x=[2000 3000 2500 4000]; %Demanda de pasajeros entre dos aeropuertos del 2000
    al 2003
y=[40000 42000 45000 60000,
100000 120000 115000 150000,
0.15 0.20 0.18 0.09,
1000 1200 1300 1400,
5000 35000 45000 5000,
300 300 300 300,
20 20 20 20] %Cada fila son los datos de las variables comentadas desde 2000 a
    2003
Regresion=regstat(x,y,linear);
disp(regresion.tstat.beta) %vector de las componentes Beta(1 por variable)
```

Con este vector, estaríamos en disposición de programar la función total en nuestro propio archivo excel de la base de datos y tendríamos una primera estimación de la demanda. Pero las posibilidades de la función regstat no acaban aquí. Obtendremos valiosa información sobre la interrelación de las variables y sobre los errores utilizando una serie de comandos, que serán comentados en el siguiente capítulo.

Habiendo definido esta ecuación, ya se estaría en condición de hacer una predicción de la demanda de transporte en una ruta. Ahora se tendrá que estudiar qué competencia tenemos para saber qué porcentaje de la demanda se acogerá a cada modo de elección.

5.2 Ecuaciones de la elección entre competencia

En estas ecuaciones el objetivo será conocer qué porcentaje de la demanda obtenida en la ecuación anterior, elegirá nuestra alternativa para realizar la ruta origen-destino seleccionada. Utilizaremos para ello los conceptos de modelo Logit y utilidades presentados en el principio de este trabajo de fin de grado.

En primer lugar, se comparará el transporte aéreo con otros medios de transporte, de manera que se obtendrá la probabilidad de que un pasajero que desea realizar la ruta, elija el avión como modo de transporte. A continuación, se calculará la probabilidad de que dicho pasajero elija nuestra aerolínea en vez de otra que opere la misma ruta. Por último, también se deberá calcular la probabilidad de que se elija nuestro aeropuerto, en el caso de existir una alternativa cercana similar.

5.2.1 Probabilidad frente a otros medios de transporte

Se definen en primer lugar los 4 modos de transporte que se consideran:

- *Coche*; $i=1$
- *Bus*; $i=2$
- *Tren*; $i=3$
- *Avión*; $i=4$

Posteriormente las variables que intervienen en esta ecuación

- *Precio del ticket*; P_{Alt}

- *Tiempo de viaje*; T_{Alt}
- *Frecuencia*; f_{Alt}

Y finalmente los coeficientes de los atributos a los que llamaremos θ_j

- *Precio del ticket*; θ_P
- *Tiempo de viaje*; θ_T
- *Frecuencia*; θ_F

De tal manera que, las utilidades para cada alternativa quedarán definidas del siguiente modo:

$$V_{coche} = \theta_P P_1 + \theta_T T_1 + \theta_F F_1 \quad (5.4)$$

$$V_{bus} = \theta_P P_2 + \theta_T T_2 + \theta_F F_2 \quad (5.5)$$

$$V_{tren} = \theta_P P_3 + \theta_T T_3 + \theta_F F_3 \quad (5.6)$$

$$V_{avion} = \theta_P P_4 + \theta_T T_4 + \theta_F F_4 \quad (5.7)$$

De tal manera que, con estos tres valores, podríamos aplicar la fórmula para ver la probabilidad de elección de alguno de los modos de transporte:

$$P_i = \frac{e_i^V}{\sum_{k=1}^n e_k^V} \quad (5.8)$$

Es ahora cuando entra en juego el complejo cálculo de los coeficientes θ . Este punto podría dar lugar a un propio trabajo de fin de grado como tal, por lo que trataremos de hacer las estimaciones lo más sencillas posible.

Se comienza haciendo referencia al signo de dicho coeficiente. Si un aumento del valor de la variable conllevaría un aumento de la utilidad, el signo será positivo, en cambio si una disminución del valor de la variable conllevaría un aumento de la utilidad tendría valor negativo. En este caso, la variable precio, tendrá un coeficiente de atributo negativo, porque conforme más caro sea el billete, menos posibilidad hay de que el cliente elija dicha alternativa. Con el tiempo de viaje ocurre lo mismo, y el caso de la frecuencia sería el coeficiente de atributo positivo, de manera que, a más frecuencias, mayor probabilidad habrá de que se escoja dicha alternativa.

La estimación del valor absoluto de estos coeficientes se realiza partiendo de una gran base de datos con una serie de encuestas como base. Cómo la capacidad ha sido limitada, se procederá a realizar una estimación basada en lo que otros modelos han predicho, de esta manera estableceremos unos valores fijos para estos coeficientes.

Lo comentado en estos últimos párrafos sobre los coeficientes de atributos se hace extensible a los dos siguientes apartados y al final de este apartado se establece una tabla con dichos valores y su fuente de estimación.

5.2.2 Probabilidad frente a otras aerolíneas

Se van a definir como ejemplo las probabilidades de elección de 3 aerolíneas, pero es aplicable al número de aerolíneas que se estimase oportuno comparar.

- Aerolínea A; $i=1$
- Aerolínea B; $i=2$
- Aerolínea C; $i=3$

Posteriormente se definen las variables que intervienen en esta ecuación

- Precio medio histórico; P_{ot}
- Aerolínea LowCost; L_{wc}
- Flexibilidad; F_{lx}
- Antiguo; A_{nt}

Y finalmente los coeficientes de los atributos a los que llamaremos θ_j

- Precio medio histórico; θ_{Pot}
- Aerolínea LowCost; θ_{Lwc}
- Flexibilidad; θ_{Flx}
- Antiguo; θ_{Ant}

De tal manera que las utilidades para cada alternativa quedarán definidas del siguiente modo:

$$V_{AerolineaA} = \theta_{Pot}Pot_1 + \theta_{Lwc}Lwc_1 + \theta_{Flx}Flx_1 + \theta_{Ant}Ant_1 \quad (5.9)$$

$$V_{AerolineaB} = \theta_{Pot}Pot_2 + \theta_{Lwc}Lwc_2 + \theta_{Flx}Flx_2 + \theta_{Ant}Ant_2 \quad (5.10)$$

$$V_{AerolineaC} = \theta_{Pot}Pot_3 + \theta_{Lwc}Lwc_3 + \theta_{Flx}Flx_3 + \theta_{Ant}Ant_3 \quad (5.11)$$

En este caso, la variable precio, tendrá un coeficiente de atributo negativo y la variable Low Cost, Flexibilidad y Antiguo, lo tendrá positivo.

5.2.3 Probabilidad frente a a otros aeropuertos

Se van a definir como ejemplo las probabilidades de elección de 3 aeropuertos, pero es aplicable al número de aerolíneas que se estimase oportuno comparar.

- Aeropuerto A; $i=1$
- Aeropuerto B; $i=2$
- Aeropuerto C; $i=3$

Posteriormente se definen las variables que intervienen en esta ecuación

- Ruta ya existente; R_{ex}
- Índice de accesibilidad; A_{cc}

Y finalmente los coeficientes de los atributos a los que llamaremos θ_j

- Ruta ya existente; θ_{Rex}
- Índice de accesibilidad; θ_{Acc}

De tal manera que las utilidades para cada alternativa quedarán definidas del siguiente modo:

$$V_{AeropuertoA} = \theta_{Rex}Rex_1 + \theta_{Acc}Acc_1 \quad (5.12)$$

$$V_{AeropuertoB} = \theta_{Rex}Rex_2 + \theta_{Acc}Acc_2 \quad (5.13)$$

$$V_{AeropuertoC} = \theta_{Rex}Rex_3 + \theta_{Acc}Acc_3 \quad (5.14)$$

En este caso, ambas variables tendrán el coeficiente positivo.

5.3 Ecuación de costes

En este apartado no se pretende hacer un profundo análisis económico de cada operación, es una función que aportará al encargado de tomar la decisión de una idea aproximada del coste que conllevará dicha operación.

También será el mecanismo para comparar las distintas posibilidades de aeronaves una vez se conozca la demanda y se haya decidido el par origen-destino.

Una función de costes de operación estará compuesta de un conjunto de costes fijos y un conjunto de costes variables. Se establecen, como en los casos anteriores, las variables que intervendrán en esta función:

- Precio actual del combustible; P_{comb}
- Distancia de la ruta; D_{is}
- Tiempo de vuelo; T_{vl}
- Modelo aeronave; A_{er}
- Factor de ocupación; F_{oc}
- Coste de tripulación; C_{rw}
- Coste handling en rampa; C_{hlr}
- Coste handling pasajero; C_{hlp}
- Tasas aeronave; T_{av}
- Tasas pasajero; T_{pa}
- Impuestos Operación; I_{op}

Ahora se trata de formular una ecuación lineal de coste:

$$C_{Op} = P_{comb} * f(D_{is}, T_{vl}, A_{er}) + C_{hlr} + T_{av} + I_{op} + g(A_{er}, F_{oc}) * (T_{pa} + C_{hlp}) + C_{rw} \quad (5.15)$$

Donde:

$$f(D_{is}, T_{vl}, A_{er}) = \text{Litros} \quad (5.16)$$

y:

$$g(A_{er}, F_{oc}) = \text{Pasajeros} \quad (5.17)$$

Esta ecuación tendrá unidad de euros y el coste será relativo a un trayecto individual origen-destino.

6 Análisis estadístico de las variables

Se definen estadísticamente en este capítulo las variables del modelo y posteriormente se estudiará su importancia en este:

Tabla 6.1 Variables Demográficas Input.

Variable	Símbolo	Definición	Carácter Estadístico	Unidad	Rango	Periodo	Tipo
Población	P	Número de habitantes que habitan en una región en un instante de tiempo determinado	Variable continua	Individuos	>0	Instante Observación	Demográfica
PIB	PIB	El PIB es el valor en número de unidades monetarias de los bienes y servicios producidos por una región en un tiempo determinado	Variable continua	€	>0	Instante Observación	Demográfica
Ratio de desempleo	Des	El Ratio de desempleo es el porcentaje de la población en edad de población activa actualmente en situación de paro o desempleo	Variable continua	%	0-100	Instante Observación	Demográfica
Plazas Hosteleras	Hot	Es el número de plazas hoteleras disponibles en los establecimientos preparados para ello en una región en un periodo de tiempo determinado	Variable continua	Individuos	>0	Instante Observación	Demográfica
Ocupación Hostelera	Och	"Es el número de noches que un turista registra en un establecimiento hotelero en un periodo de tiempo	Variable continua	Pernoctaciones	>0	Instante Observación	Demográfica

Tabla 6.2 Variables Aeroportuarias Input.

Variable	Símbolo	Definición	Carácter Estadístico	Unidad	Rango	Periodo	Tipo
Capacidad	Cap	Número máximo de operaciones por hora	Variable continua	Operaciones/Hora	>0	Instante Observación	Aeroportuaria
Clave de referencia	Ref	La clave está compuesta de dos elementos que se relacionan con las características y dimensiones del avión. El elemento 1 es un número basado en la longitud del campo de referencia del avión y el elemento 2 es una letra basada en la envergadura del avión y en la anchura exterior entre las ruedas del tren de aterrizaje principal[55]	Variable cualitativa, nominal, combinación de 4 números y 6 letras	-	De 1A a 4F	Instante Observación	Aeroportuaria
Rutas activas	Act	Número de rutas activas distintas desde un aeropuerto.	Variable continua	Rutas	>0	Instante Observación	Aeroportuaria
Churn de rutas	Chu	Número de rutas churneadas distintas desde un aeropuerto.	Variable continua	Rutas	>0	Histórico global	Aeroportuaria
Aerolíneas activas	Aer	Número de aerolíneas activas distintas en un aeropuerto	Variable continua	Aerolíneas	>0	Instante Observación	Aeroportuaria
Índice de Accesibilidad	Acc	Índice de accesibilidad al aeropuerto	Variable discreta que puede tomar: 0-1-2	-	0,1,2	Instante Observación	Aeroportuaria
Evolución volumen de pasajeros	Vol	Porcentaje de crecimiento del número de pasajeros en un aeropuerto con respecto a un periodo anterior	Variable continua	%	>0	Año actual	Aeroportuaria

Tabla 6.3 Variables De Ruta Input.

Variable	Símbolo	Definición	Carácter Estadístico	Unidad	Rango	Periodo	Tipo
Distancia Destino-Origen	Dis	Distancia en kilómetros en el mapa entre las coordenadas del aeropuerto de origen y las coordenadas del aeropuerto de destino.	Variable continua	Km	>0	-	De ruta
Tiempo de vuelo	Tvl	Tiempo en horas, minutos y segundos transcurrido desde el comienzo del rodaje en el aeropuerto de origen hasta el estacionamiento en el aeropuerto de destino.	Variable continua	hh:mm:ss	>0	Instante Observación	De ruta
Restricciones diplomáticas	Res	Variable que detecta si existen o no restricciones de vuelo (cierres del espacio aéreo) en una determinada región	Variable discreta, dicotómica (Dummy): que toma el valor 1 si existen restricciones y 0 en el caso contrario	-	1-0	Instante Observación	De ruta
Histórico de vuelos	His	Número de vuelos comerciales de pasajeros que ha habido en un periodo determinado entre los aeropuertos del par origen-destino.	Variable continua	Operaciones	>0	Histórico 5 años	De ruta
Historial de búsquedas	Bus	Número de búsquedas que ha habido en un periodo determinado entre los aeropuertos del par origen-destino.	Variable continua	Búsquedas	>0	Año actual	De ruta

Tabla 6.4 Variables De Alternativa Input.

Variable	Símbolo	Definición	Carácter Estadístico	Unidad	Rango	Periodo	Tipo
Precio de las alternativas	PAlt	Valor en número de unidades monetarias del recorrido del trayecto en un modo de transporte determinado	Variable continua	€	>0	Instante Observación	De alternativa
Tiempo de viaje	TAlt	Tiempo en horas, minutos y segundos transcurrido en el trayecto en cada una de las alternativas	Variable continua	hh:mm:ss	>0	Instante Observación	De alternativa
Frecuencia de la alternativa	FAlt	Frecuencia mensual de viajes en cada una de las alternativas	Variable continua	Operaciones	>0	Mes Actual	De alternativa
Precio medio histórico	POt	Precio medio de rutas similares en los últimos 5 años	Variable continua	€	>0	Histórico 5 años	De alternativa
Aerolinea low-cost vs tradicional	Lwc	Precio medio de rutas similares en los últimos 5 años	Variable discreta, dicotómica (Dummy): que toma el valor 1 si la aerolinea es Low-Cost y 0 en el caso contrario	-	1-0	Instante Observación	De alternativa
Flexibilidad	Flx	Tipología del ticket estándar que ofrece esa aeronave	Variable discreta, dicotómica (Dummy): que toma el valor 1 si hay flexibilidad en el billete y 0 en el caso contrario	-	1-0	Instante Observación	De alternativa
Antigüedad	Ant	Número de años que una aerolínea lleva operando un trayecto.	Variable continua	años	>0	Histórico 10 años	De alternativa
Aeropuertos cercanos	Aer	Número de aeropuertos que hay en un radio de 200 kilómetros centrado en el aeropuerto de origen o destino.	Variable continua	Aeropuertos	>0	Instante Observación	De alternativa

7 La estructura de outputs: Índices y accionables

El objeto final de la creación del sistema de ayuda a la toma de decisiones no era otro que el de llegar al momento de la toma de decisión y actuar de la manera correcta. Se podría obtener un modelo matemático excelente, un pool de datos extremadamente actualizado y una tecnología punta, pero sin algo que nos permita visualizar al instante los parámetros claves que determinarán la decisión, no se estaría llegando al resultado buscado.

Es por ello, que se plantea un Dashboard de mando como entregable final que apoye la toma de decisiones. Hoy en día, cualquier persona encargada de tomar una decisión en una gran compañía, dispone de un cuadro en el que, con visualizar un pequeño número de indicadores clave, pueda hacerse una idea del estado de su proyecto. En este caso se ha planteado de la misma forma. Se expondrá en este capítulo el Dashboard al que se ha llamado "Airport Decision-Making Dashboard".

Se comenzará definiendo cuáles de los indicadores son claves en la decisión, se argumentarán las herramientas de visualización empleadas y se expondrá en detalle el proceso de creación del Dashboard.

7.1 Análisis cualitativo de los índices

Como en cualquier gran compañía, la tendencia a la digitalización y exactitud es cada vez mayor en Aeropuertos y Aerolíneas. En unos ámbitos en los que hay operaciones tan diversas, es crucial mantener un control y una medida que permita al gestor aeronáutico hacerse una idea de la "temperatura" de su negocio en tiempo real y a simple vista. Puesto que una pequeña decisión tomada por alguno de estos agentes, puede verse reflejada en beneficios o pérdidas importantes, es crucial medir ciertos indicadores accionables sobre los que aplicar pequeños cambios que lleven a la mejora.

Con este objetivo aparece el concepto de "KPI: Key Performance Indicator", que se refiere a un indicador de la actual situación de algún aspecto crítico en una compañía. Es importante tener en cuenta que un KPI no nos permitirá hacer un detallado análisis sobre una materia muy concreta, se trata por el contrario de hacerse una idea real y general de alguna medida de relativa importancia.

En la ciencia de datos (concretamente en la visualización), una de las tareas más complejas es la de elegir un reducido número de indicadores y conseguir que estos plasmen la situación del negocio. En el ámbito puramente aeronáutico, estos KPI's tenían una tendencia a ser exclusivamente económicos en un principio. Con la liberalización del mercado, se fueron centrando en la eficiencia operativa y actualmente, un cuadro de mando de una aerolínea o un aeropuerto, es una combinación de indicadores tanto económicos como operativos.

Humphreys y Graham (2002) realizaron varios análisis sobre encuestas a directivos del mundo de la aeronáutica acerca de los indicadores que más utilidad tenían para ellos y por ejemplo, establecieron KPI's

como Ingresos por Pasajero, Gastos por Pasajero en el Aeropuerto o Tiempo de espera de maletas.[56]

Andersson Granberg (2013) estableció 5 categorías de área de actividad y estableció entre dos y tres KPI's sobre cada una de las áreas, que eran: Operaciones, Economía, Medio Ambiente, Seguridad y Servicio al cliente. Determinando al final de su estudio que es crucial que estos cuadros de mando sean flexibles y adaptados a la evolución.[57]

En definitiva, se realizará a continuación una selección de cuáles serán los KPIs básicos para nuestro modelo, pero se debe tener en cuenta que variarán en función de la evolución del tiempo y el sistema está pensado para ello:

- *Indicador de entrada*: Indicador del tipo: SI/NO que nos recomendará (a priori) si es rentable o no entrar en la ruta.
- *Demanda de pasajeros por año*: Previsión del número de pasajeros que demandan la ruta ese año.
- *Costes de operación*: Coste en euros de realizar la operación entre el par de aeropuertos.
- *Ruta ya operativa*: Número de compañías que operan la misma ruta.
- *Modo más probable*: Modo de transporte más probable a ser elegido para completar la ruta, y dicha probabilidad.
- *Mapa de la ruta*: Representación visual de la ruta.
- *Distancia de la ruta*: Kilómetros de vuelo de la ruta.
- *Tiempo de la ruta*: Horas de vuelo de la ruta.
- *Combustible necesario de la ruta*: Litros de combustible necesarios para completar la ruta.
- *Destinos activos desde el origen*: Listado de destinos actualmente activos desde el origen.
- *Competencia modal*: Otros modos de transporte que son posible competencia y la probabilidad de que sean escogidos y la alternativa avión sea sustituida. Coste de los mismos y tiempo de viaje.
- *Competencia de aerolíneas*: Otras aerolíneas que operan la misma ruta y son posible competencia y la probabilidad de que sean escogidas.
- *Competencia de aeropuerto*: Otros aeropuertos cercanos que pueden ser competencia.
- *Evolución del tráfico*: Evolución del tráfico tanto en el aeropuerto origen como en el destino.

En esta primera aproximación hemos definido los indicadores principales de nuestro Dashboard y que conformarían una primera pestaña en la que, en una visual, se pueda medir el pulso de la decisión. No obstante, plantearemos dos pestañas más, en las que tendremos información más detallada sobre variables relativas al aeropuerto de origen y al aeropuerto de destino.

7.2 Dashboard: Unificación de resultados

Una vez se han seleccionado los indicadores que harán viable la toma de decisiones, es el momento de poner sobre la mesa dichos valores de la forma más visual posible. Para ello se decidirá primero, en base a un estudio comparativo, que herramienta de visualización de datos se ha elegido, y se explicará detalladamente y paso por paso, el proceso de construcción de dicho Dashboard.

7.2.1 Herramientas de visualización de datos

Desde el 2010, con el auge de la digitalización de las compañías y la aparición del big data, el mundo del software de visualización de datos ha crecido rápidamente. Son tres los gigantes de este nicho de mercado:

- *Tableau*
- *Power BI*
- *Google Data Studio*

Tableau es un software francés, con prestaciones interesantes pero de un elevado precio. Es una herramienta que se viene utilizando desde hace bastantes años, instaurada por muchas empresas y con un soporte y una comunidad bastante fuerte. Por otro lado funciona bastante bien en los gráficos tipo mapa y permite el almacenaje de hasta 10 GB de datos.[58]

Power BI es el software de microsoft para la visualización de datos. Es un software que, al igual que Tableau, requiere de unos conocimientos mínimos previos al uso, que tiene una versión reducida gratuita y una más avanzada de pago, y que cómo punto fuerte tendría la gran capacidad de integraciones con otros softwares.[59]

DataStudio es una aplicación de Google para visualización de datos. Es el más moderno de los tres que se comparan y es totalmente gratuito. Es intuitivo y se conecta a la perfección con todas las extensiones de google. El soporte tiene más carencias pero la interfaz permite una gran cantidad de gráficos y el uso es muy sencillo.[60]

En la siguiente imagen se muestra una comparativa de las tres opciones que se plantean.

Power BI	Tableau	Data Studio
		
Muy buenas integraciones y soporte.	Visualizaciones complejas y buenos mapas. Comunidad grande.	Completamente gratuito y ,muy intuitivo
Precio medio y dificultad adquisición conocimientos	Precio muy alto	Soporte nulo y software demasiado joven

miro

Figura 7.1 Comparación software visualización de datos. Creación propia.

Por simplicidad, precio y prestaciones de conexión con la base de datos, además de por la experiencia con esta herramienta, Google Data Studio es la herramienta elegida para la elaboración del dashboard de outputs. A continuación, se definirá el procedimiento de creación del mismo.

7.2.2 Metodología de creación de paneles y storytelling

El informe de DataStudio al que llamaremos Airport Decision-Making Dashboard, tendrá como fuentes de datos las 6 pestañas de nuestro "Sheet Data Center", que cabe recordar estaba en formato de hoja de cálculo de Google Sheets. El informe se dividirá en tres paneles: un primer panel al que llamaremos "Home", que será un resumen de la ruta a estudiar, un segundo panel al que llamaremos "Informe Origen", en el que se especificarán más en detalle variables del origen y un tercero, llamado "Informe Destino", en el que se

especificarán en detalle variables del destino.

Comenzaremos exponiendo como se realiza la conexión con la fuente de datos, que es el primer paso en todo informe de DataStudio que uno desee generar. En nuestro caso, la forma de conexión elegida será con la hoja de cálculo de google, por lo que seleccionaremos dicha opción tal y cómo se presenta en pantalla:

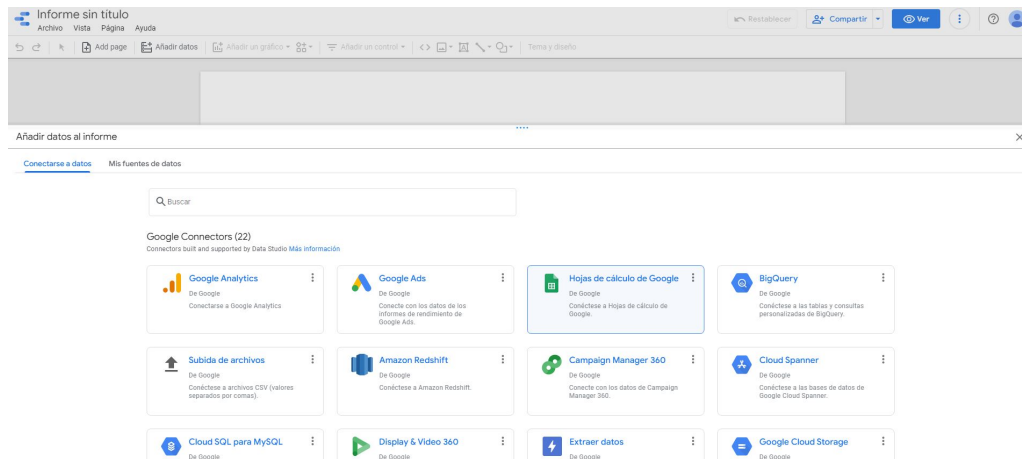


Figura 7.2 Conexión fuente de datos. Fuente: DataStudio.

A continuación, se definen ciertos aspectos puramente estéticos en los que no se hará detenimiento y se comienzan a establecer los seleccionables que deberá fijar el operador aeronáutico que quiera hacer el estudio:

- *Aeropuerto de origen*
- *Aeropuerto de destino*
- *Aeronave de operación*

Y una vez que se hayan seleccionado estos tres valores, toda la información de las tres pestañas del dashboard se modificará de acuerdo a la selección obtenida. Esto es de real importancia porque en cuestión de segundos pasaremos de tener sobre la mesa la información de una ruta a la información de otra, permitiendo esto una gran capacidad de análisis comparativo.

La configuración de estos selectores consistirá sencillamente en elegir la fuente y la dimensión desde la que se desea elegir un registro de datos. En este caso de la imagen, el aeropuerto de destino:

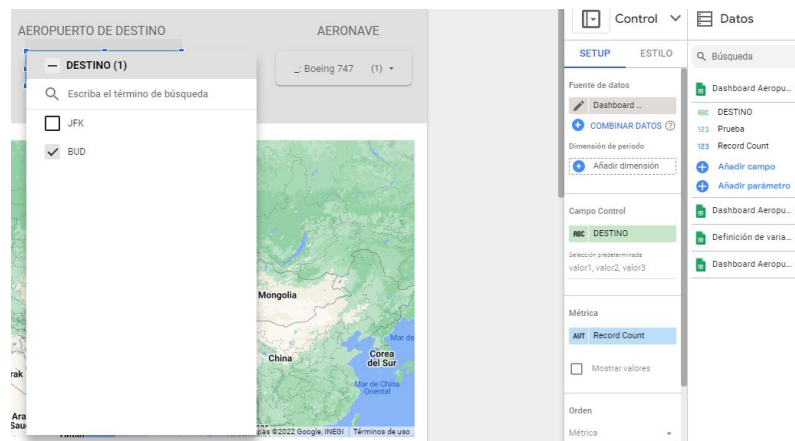


Figura 7.3 Configuración de selectores. Fuente: DataStudio.

Por tanto, el único paso que el usuario del sistema deberá realizar de aquí en adelante será la selección de estos tres índices que se presentan. Comentar en este punto, que los datos que se ven en las capturas de pantalla de este capítulo son ficticios y podrían llevar a confusión. En el último capítulo del caso práctico se verán datos realistas. Por tanto, la primera parte de nuestro Dashboard tendrá la siguiente estructura:

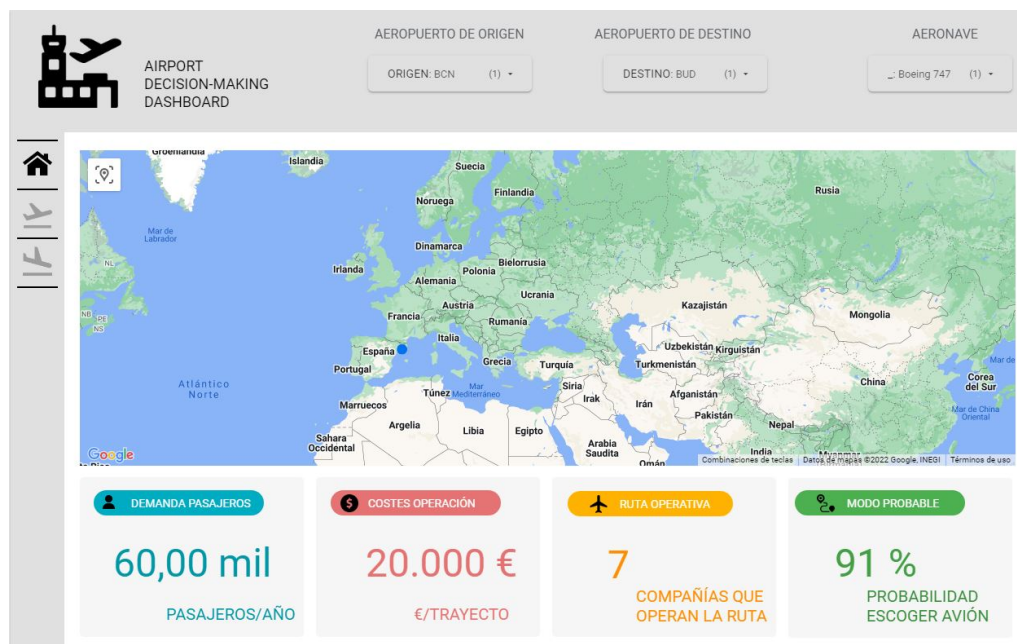


Figura 7.4 Previsualización del encabezado. Fuente: DataStudio.

De aquí en adelante en el capítulo se irá profundizando sobre cada una de las secciones de los distintos paneles, en algunos casos haciendo referencia a su configuración.

Para la creación de un gráfico tipo mapa que aporte una visión espacial del problema, DataStudio se integra con Google Maps, de manera que, conociendo la situación de ambos aeropuertos, genera automáticamente la ruta. Para ello, siempre en nuestra fuente de datos, seleccionamos el campo: coordenadas, y la visualización tanto del campo como de nuestro mapa sería de la siguiente manera:

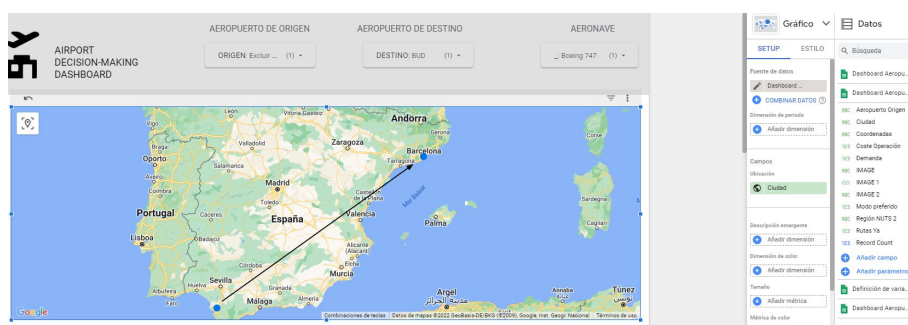


Figura 7.5 Gráfico tipo mapa de ruta. Fuente: DataStudio.

Se prosigue por la configuración de los cuatro indicadores principales de la figura 7.6. Ahora, se tiene que seleccionar en la configuración del gráfico, una métrica además de la dimensión (esta ya fue escogida en los tres selectores superiores). En el caso de la demanda, tal y como se aprecia a la derecha, será una métrica numérica proveniente de la ecuación de demanda. Se podría haber planteado hacer este cálculo en el propio Data Studio con la opción de programar una fórmula, pero se ha hecho en el archivo de google sheets. Para los 3 KPI's siguientes la configuración es exactamente la misma.



Figura 7.6 KPI's numéricos simples. Fuente: DataStudio.

Se continúa por los KPIs relativos a ciertos valores de la ruta en sí. Estos serán indicadores simples que se configuran como los anteriores (Figura 7.7)

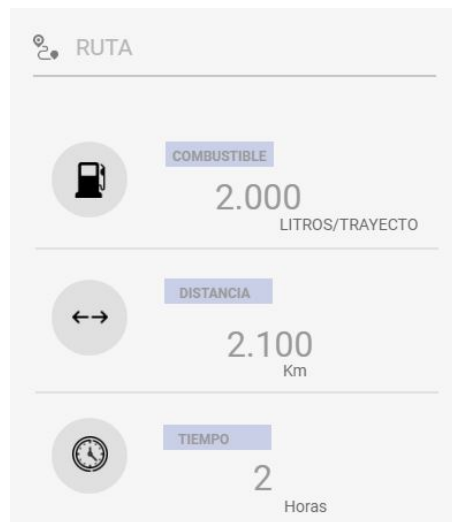


Figura 7.7 KPI's relativos a ruta. Fuente: DataStudio.

También son de importancia los indicadores sobre la competencia tanto en el modo de transporte como en el caso de que existan otras aerolíneas operando la ruta. Estos se muestran en la figura 7.8.



Figura 7.8 KPI's relativos a competencia. Fuente: DataStudio.

En la figura 7.9, se aprecia un listado de las rutas ya operadas en ese destino. El listado es otra de las opciones de gráficos que Google Data Studio nos ofrece.



Figura 7.9 Listado de rutas ya operadas. Fuente: DataStudio.

La última posibilidad de gráfico que se muestra es la de la figura 7.10, en la que el software nos permite obtener un histórico evolutivo del tráfico en un aeropuerto.



Figura 7.10 Histórico de tráfico. Fuente: DataStudio.

Como resultado final de esta pestaña, se obtiene una visión general que será la que permitirá al decisor tener una visión práctica y analítica sobre la ruta en cuestión.

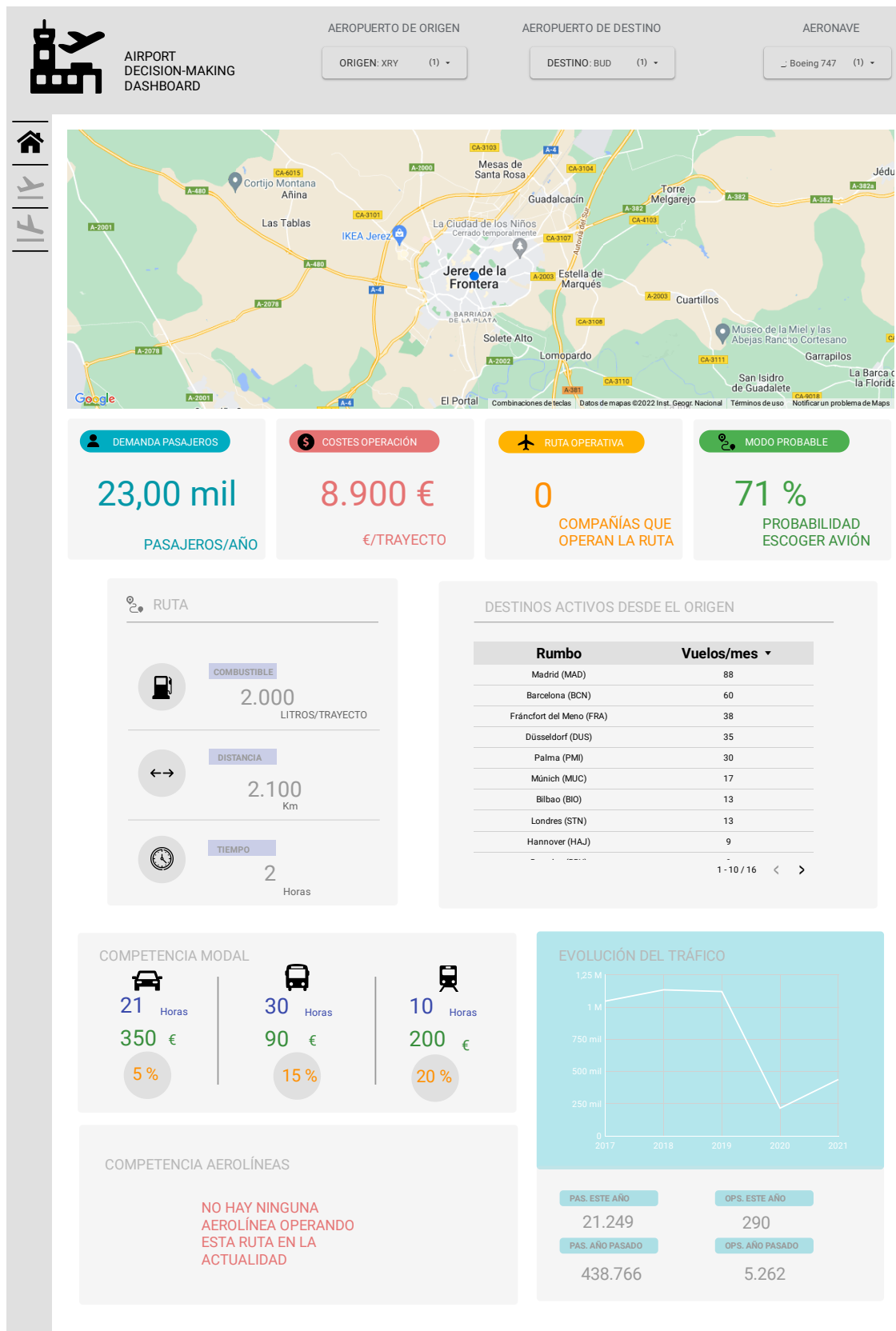


Figura 7.11 Dashboard Home. Fuente: DataStudio.

8 Validez del modelo

En este capítulo se hace referencia a los posibles puntos en los que el modelo puede no ser exacto o perfecto. En los capítulos anteriores se han tratado temas de diversa índole y es preciso repasar que suposiciones e hipótesis hemos hecho.

Son varias las interpretaciones y definiciones que existen sobre la validez de un modelo. Aquí debemos distinguir verificación de validación: La verificación se define como el proceso de comprobar que un sistema o modelo funciona tal y como estaba previsto, mientras que la validación consiste en evaluar el resultado y ver si este es correcto o no.[61]

Se centrará este capítulo en la validación del modelo. El procedimiento seguido para documentar dicha validación será la de recorrer el documento escrito hasta ahora documentando aquellos puntos en los que se encuentra alguna debilidad al sistema.

Con este apartado no se pretende demostrar que el modelo no es válido, si no ser conscientes de que tanto por falta de tiempo o medios, puesto que es una primera aproximación, el sistema no es perfecto.

8.1 Actualización de datos: Validez temporal del modelo

En cualquier modelo estadístico de bases de datos, es clave tener claro con qué frecuencia son actualizados los datos y qué efectividad tienen estos con el paso del tiempo, más aún si son modelos predictivos como el nuestro.

La primera suposición que se hace es el determinar que el instante de la entrada en la ruta es el mismo instante en el que se consulta el Dashboard. Esto puede llevar a conflicto, en el sentido de que debe contemplarse como algo estimatorio. Puede que desde el momento de la discusión hasta el momento en el que la entrada se lleve a cabo, pase demasiado tiempo y esto haga que el modelo ya no tenga la misma validez.

En segundo lugar, debemos tener en cuenta la concordancia en la actualización de datos. Habrá bases de datos como la de búsquedas online o la de precios que variarán con una frecuencia muy alta y deberán ser actualizadas casi diariamente si se quiere mantener el modelo con total efectividad. Otras bases de datos en cambio se actualizarán con menos frecuencia. La población de una región variará poco diariamente o la distancia de vuelo no variará nada. Una frecuencia de actualización mensual sería correcta en principio.

En cuanto a los datos que se visualizan en el Dashboard, estos se actualizan automáticamente desde las bases de datos cada 15 minutos, por lo que esto no supondría ningún problema de validez temporal para el modelo.

En definitiva, un modelo de datos tiene que tener siempre un encargado de su actualización, y al haber bases de datos diversas, esta necesidad se hace aún mayor.

8.2 El sistema a nivel europeo: Validez espacial del modelo

Aunque las bases del sistema están sentadas en términos generales y no habría ningún problema en ampliar a horizontes mayores, está pensado para el ámbito europeo, e incluso en algunos aspectos, su validez a nivel nacional es mayor que a nivel europeo.

Por un lado, el estudio del mercado de aerolíneas europeo difiere del mercado americano o asiático en ciertos sentidos que pueden hacer que el modelo pierda precisión. Por otro lado, hay variables que perderían o ganarían importancia si los vuelos fueran transatlánticos, como por ejemplo en el consumo de combustible o la flexibilidad del billete aumentaría la importancia al ser un vuelo más largo, en cambio, la accesibilidad al aeropuerto no sería tan definitiva por ser superior la utilidad del motivo del vuelo que la comodidad.

Atendiendo a variables concretas, en las variables demográficas, tenemos que tener en cuenta que las regiones NUTs son asociadas a códigos de aeropuerto IATA de una manera aproximada. Aquí puede haber errores de asociación y una futura mejora del modelo podría ser automatizar dichas asociaciones.

El otro aspecto de situación en el que el modelo puede mejorarse es en el estudio de competitividad de aeródromos cercanos. En el sentido de que el área que se considera de influencia ha sido definido como menor o mayor según el agente que lo defina. Por ejemplo, puede que el aeropuerto de Barcelona influya al aeropuerto de Sevilla pese a su lejanía. Aquí el modelo está falto de exactitud y podría ser mejorado.

8.3 Otras hipótesis

Comenzando por cuestiones técnicas, el modelo es limitado en cuanto al cálculo del consumo de combustible de un aeronave. Se han adquirido un listado de las cuatro aeronaves más utilizadas pero no se hacen muchas más especificaciones en cuanto al número de pasajeros, se asume un factor de ocupación medio. Aquí nuestro modelo mejoraría considerablemente haciendo un cálculo más exacto puesto que los litros de combustible dependen directamente del número de pasajeros y este varía con frecuencia.

En cuanto al nivel de accesibilidad de los aeropuertos, se ha estimado un índice básico entre 1 y 3 que quedaba explicado en el capítulo de definición de las variables. Este índice depende del número de modos de acceso al aeropuerto pero no tiene en cuenta ni frecuencias, ni precios ni calidad, por lo que este índice no es del todo fiable.

Una de las hipótesis principales del modelo y que ya se expuso en la introducción, era el hecho de escoger un par de aeropuertos origen-destino como unidad de medida, lo que implica que el tráfico se supone igual en la ida que en la vuelta (se hablaba de un par bidireccional). Esto no es del todo exacto puesto que hay numerosas rutas en las que el número de pasajeros a la ida es totalmente distinto al de la vuelta frecuentemente.

Otro de los aspectos que no se tiene en cuenta es el de los horarios. Un posible pasajero estará dispuesto a pagar más o menos por su billete dependiendo del horario del vuelo y del día de la semana. Además, las frecuencias de vuelos también tendrían una influencia que no se ha tenido en cuenta por simplificar el modelo.

Sobre las hipótesis tomadas a la hora de analizar la competencia también hay varias cuestiones de importancia. En primer lugar, se establecen precios medios temporales para los billetes, cuya validez es muy débil en tanto en cuanto que la estacionalidad influirá totalmente. Otra de las hipótesis es que el viajero realiza el viaje en coche sólo (su coste se convierte en mucho mayor que si compartiese el coche), o por ejemplo se presupone que los viajeros no tienen ningún tipo de descuento en ningún modo de transporte, cosa que no es del todo cierta.

En lo que a la ecuación de costes se refiere, es sabido que las aerolíneas negocian con los agentes aeroportuarios precios distintos a las tarifas estándar por lo que el precio de las tasas que tenemos será una aproximación y dependerá de más factores que no se han tenido en cuenta.

8.4 Errores estadísticos

Pero sin duda, un modelo estadístico siempre tiene como punto crítico de mejora los errores en los coeficientes. En el caso que se estudia, dichos errores provienen por un lado de los coeficientes obtenidos de la recta de regresión para el cálculo de la demanda, y de los coeficientes que nos permitían conocer la utilidad de una alternativa.

En este modelo se ha realizado un primer cálculo de regresión sencillo, pero es sabido que si este cálculo se afina (con el coste que conlleva) el modelo ganaría bastante en utilidad.

En cualquier caso, el objetivo de esta sección no es otro que el de demostrar que, invirtiendo los recursos suficientes, puede hacerse un sistema exacto del modelo básico que se plantea en este trabajo.

9 Estudio para un caso concreto: Aeropuerto de Jerez

En este último capítulo se propone un caso práctico de uso del Dashboard para una supuesta aerolínea ficticia que desea operar una nueva ruta aérea desde el aeropuerto de Jerez. El hecho de no elegir una aerolínea simplifica el modelo de datos en el sentido de que se aparta dicha variable.

El momento de la decisión tendrá fecha del primer semestre de 2022, periodo en el que se ha realizado este Trabajo de Fin de Grado. La elección del Aeropuerto de Jerez se hace por el hecho de ser un aeropuerto mediano, con un tráfico creciente y con relativamente pocas rutas internacionales atendiendo al lugar de interés turístico en el que se emplaza.

El procedimiento seguido en este capítulo será el propio procedimiento que el "decision-maker" de la aerolínea ficticia realizaría para su decisión. En primer lugar, se expondrá el contexto de dicho aeropuerto, sus características, tráfico actual y previsto. Posteriormente, se comentarán brevemente y con un análisis no numérico algunas posibilidades de rutas y se terminará haciendo un análisis que recogerá tanto el tratamiento de dichos datos, como su observación en el Dashboard y su correspondiente análisis.

9.1 Contexto: Situación, características y volumen de tráfico

El Aeropuerto de Jerez (XRY), se encuentra en la provincia de Cádiz, Andalucía, España. Es el único aeropuerto civil de la provincia, que tiene más de un millón de habitantes. Es un aeropuerto de carácter civil-internacional, cuyo código OACI es "LEJR" y el código IATA "XRY".[62] El código de aeródromo según OACI es 4-E, teniendo el aeródromo una sola pista, la 02-20, con una longitud de 2200 m y 45 m de anchura. De Asfalto PCN 51/F/A/W/T, con una elevación de 28 m y una temperatura de referencia de 34 grados.

El edificio de la terminal se construyó en 1992. La planta de salidas dispone de un área de vestíbulo de 1400 metros cuadrados, con 20 mostradores de facturación y 4 controles de seguridad, de un área de embarque de 1390 metros cuadrados y de dos puestos de control de pasaportes.

La zona de recogida de equipajes dispone de cuatro hipódromos, 4 puestos de control de pasaporte y un vestíbulo de 650 metros cuadrados. En la planta alta se tiene un edificio de oficinas técnicas y restaurante.

En cuanto al acceso por carretera, la autovía N-IV se une a que el aeropuerto posee 870 plazas de parking de uso público, por lo que el vehículo propio es el método más usado. Además, está disponible el acceso por ferrocarril, con un tren de cercanías desde Cádiz.



Figura 9.1 Vista aérea del Aeropuerto de Jerez. Fuente: Google Maps.

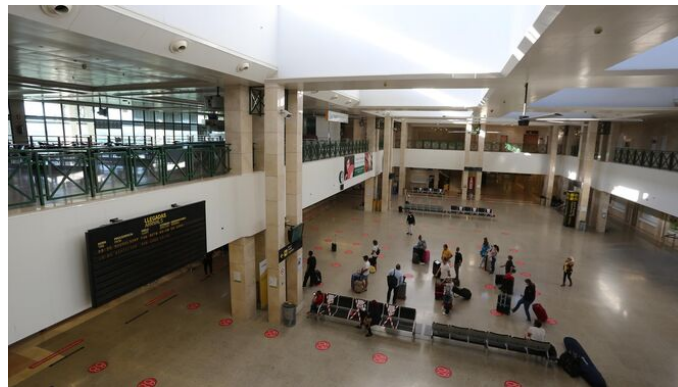


Figura 9.2 Vista interior del Aeropuerto de Jerez. Fuente: Google Maps.

En cuanto a la capacidad y parámetros de diseño del aeropuerto, se encuentran los siguientes valores:

Tabla 9.1 Valores de capacidad y parámetros.

Parámetro	Valor
Capacidad	20 (operaciones/hora)
PHP (Pasajeros Hora Punta)	1421
PHP en llegadas	906
PHP en salidas	984
PHD (pasajeros hora diseño)	1006
AHP (Aeronaves hora punta)	30
AHD (Aeronaves hora diseño)	6

Por último se establecen los niveles de tráfico actuales [62]:

Tabla 9.2 Niveles de tráfico actuales-Mayo 2022.

Parámetro	Valor
Pasajeros totales	95799
Pasajeros comerciales	95241
Operaciones comerciales	895

En definitiva, la situación actual del aeropuerto es la de un aeródromo en un entorno de alto interés turístico pero de pequeño-mediano tamaño, con previsible opciones de crecimiento, de ahí su elección como caso de estudio.[63]

9.2 Crecimiento: Evolución del tráfico, crecimiento de la demanda y rutas abiertas en los últimos años

Tal y como se hizo en el primer capítulo de este trabajo a nivel nacional, se tratará en este apartado de hacer un análisis a nivel aeropuerto de la evolución del tráfico, demanda, rutas y aerolíneas.

El efecto de la pandemia de COVID-19 vivida entre 2020 y 2021 hizo que los datos de tráfico en estos años fueran inferiores al crecimiento que el aeropuerto venía teniendo, aun así, las previsiones son esperanzadoras desde 2022. Prueba de ello es que el aeropuerto este verano de 2022 recuperará las rutas con Londres-Gatwick, Bilbao, Las Palmas y Zúrich.[64]

En lo reportado por AENA como previsión para el verano de 2022, la mitad de las operaciones programadas serán nacionales y la otra mitad internacionales, con Alemania a la cabeza con un tercio de las plazas internacionales.[65]

Se comenzará haciendo un estudio de qué rutas existen en la actualidad y con qué niveles de tráfico:

Rumbo	Vuelos/mes ▾
Madrid (MAD)	88
Barcelona (BCN)	60
Fráncfort del Meno (FRA)	38
Düsseldorf (DUS)	35
Palma (PMI)	30
Múnich (MUC)	17
Bilbao (BIO)	13
Londres (STN)	13
Hannover (HAJ)	9
Bruselas (BRU)	9
París (ORY)	8
Tenerife (TFN)	8
Hamburgo (HAM)	8
Luxembourg (LUX)	4
Zúrich (ZRH)0	4
Stuttgart (STR)	4

Figura 9.3 Rutas aéreas activas desde el aeropuerto de Jerez. Elaboración propia.

Conjunto a este análisis, merece la pena detenerse en ver cuáles han sido las últimas rutas aéreas en añadirse en cada año. Para ello usaremos los datos ya tratados en el apartado primero desde 2018 a 2021.[6]:

- 2018: 4 rutas nuevas: Berlín TXL (Easy Jet), Bilbao (Vueling), Londres LGW (TUI Airlines), Zúrich (Edelweiss)
- 2019: 1 ruta nueva: Manchester (Easy Jet)
- 2020: No hubo nuevas rutas
- 2021: 3 rutas nuevas: Frankfurt (Lufthansa), Palma de Mallorca (Ryanair) y Munich (Lufthansa)

Además del número de rutas, debemos tener en cuenta qué aerolíneas operan ya en el aeropuerto:

- Binter: 1 ruta
- Ryanair: 3 rutas
- Tui: 7 rutas
- Vueling: 5 rutas
- Condor: 3 rutas
- Lufthansa: 1 ruta
- Iberia: 1 ruta
- Luxair: 1 ruta
- Air Nostrum: 1 ruta
- Octavia Airlines: 2 rutas
- Eurowings: 1 ruta

Merece la pena detenerse en varios hechos. En el panorama español-europeo es un caso particular de aeropuerto con diversidad de compañías low-cost. No es el repetido caso de un aeropuerto dominado por una sola de estas compañías, es el caso de un aeropuerto en el que la competencia pervive.

Por último, el indicador que analizaremos será el del número global de operaciones y pasajeros:



Figura 9.4 Evolución del tráfico en el aeropuerto de Jerez. Elaboración propia.

En conclusión, se elige este aeropuerto por los distintos signos de previsión de crecimiento, tanto por el enclave turístico en el que está, cómo por el equilibrio entre tráfico nacional e internacional y por el tamaño del aeropuerto.[66]

9.3 Posibles rutas de apertura

Para realizar a modo de ejemplo el caso de uso de la metodología planteada en este trabajo, se plantean 3 posibles rutas que a priori podrían tener la posibilidad de ser rentables para la compañía ficticia que se comentaba. Se han seguido una serie de criterios totalmente cualitativos para hacer una primera aproximación. Se han descartado, por ejemplo, destinos alemanes, puesto que existen numerosas rutas ya con aeropuertos de este país. También se han excluido destinos costeros por el tipo de turismo, y se han restringido las rutas a Europa.

Las tres rutas que se plantean, con su primera razón cualitativa, son las siguientes:

- *Jerez(XRY)-Ámsterdam Schiphol(AMS)*: Ámsterdam y Frankfurt son los destinos preferidos por los españoles en 2021. Como el caso de Frankfurt tiene ya más de 30 vuelos semanales desde Jerez, se propone Ámsterdam como posible oportunidad.
- *Jerez(XRY)-Budapest(BUD)*: El reciente churn de ruta de Sevilla (SVQ)-Budapest (BUD), a la vez que el crecimiento del turismo en esta capital europea, hace que se presente como nuevo destino probable este de Budapest. También el hecho de turismo económicamente similar hace que se plantee la oportunidad.
- *Jerez(XRY)-Santiago de Compostela(SCQ)*: Madrid y Barcelona son las principales rutas desde el aeropuerto de Jerez. Debido al éxito de las conexiones con el aeropuerto gallego con el resto de grandes aeropuertos del sur, podrían indicarnos que podría ser una ruta rentable.

9.4 Procedimiento de análisis

Esta sección consistirá en ir analizando una a una las tres alternativas hasta llegar a la conclusión de cuál es la más rentable. Se comenzará comparando las variables una a una dependiendo de la ruta, se hará algún análisis más detallado de alguna de ellas y de algunos aspectos cualitativos y, por último, observando la comparativa, se tomará la decisión.

9.4.1 Aplicación de los datos a las variables

A continuación, se presentan los datos divididos según tipo de variable de los distintos elementos extraídos de cada una de las 6 fuentes de datos:

En primer lugar, se obtienen los datos de las variables demográficas de los cuatro aeropuertos que intervienen en el caso de estudio:

Tabla 9.3 Datos de variables demográficas.

Aeropuerto	Población	PIB	% desempleo	plazas hoteleras	pernoctaciones
XRY	1258153	23313,04	21,7	533501	16060434
BUD	1396239	127819,3	4,60	239244	8574123
AMS	3033638	50430,8	2,90	58711	1296526
SCQ	1121758	27466,69	11,70	135698	4272317

En la tabla 9.3 se observa la diferencia entre aeropuertos situados en regiones más prósperas como Ámsterdam, con una tasa de desempleo de un 2,9%, comparado con regiones como la nuestra con un 21,7%, que posteriormente veremos cómo está reflejado en otras variables como precios o número de destinos. En esta propia tabla ya lo podemos comparar con el valor del PIB de ambas regiones.

Los datos anteriores están directamente obtenidos de la base de datos de Eurostat.

A continuación, se exponen los datos de variables de aeropuerto:

Tabla 9.4 Datos de variables aeroportuarias.

Aeropuerto	Capacidad (op/h)	Rutas Activas	Rutas Churn	Aerolíneas	Accesibilidad
XRY	20	17	2	12	2
BUD	74	285	28	74	3
AMS	30	128	15	36	3
SCQ	20	37	7	8	2

En la tabla 9.4 se observa la diferencia entre aeropuertos ahora mayores en número de operaciones, agentes y actividad. Comparándola con la tabla 9.3 podemos ver cómo el tamaño de una región y su poder adquisitivo es relativo en este ejemplo al tamaño de operaciones y agentes de su aeropuerto.

A continuación, pasamos al código de ruta en lugar del de aeropuerto y se exponen los datos de variables de ruta:

Tabla 9.5 Datos de variables de ruta.

Ruta	Distancia (km)	Tiempo(h)	Restricciones	Hist. Vuelos	Hist. Búsquedas
XRY-BUD	2424	3:21	FALSE	0	1209
XRY-AMS	1943	2:44	FALSE	0	1350
XRY-SCQ	713	1:15	FALSE	0	2056

En la tabla 9.5 se observa una curiosa relación cómo es la proporción inversa entre lejanía a partir de una cierta distancia y las búsquedas (este patrón se repite en numerosos casos y este ejemplo es uno de ellos).

Por último, se exponen los datos de la competencia en el transporte:

Tabla 9.6 Datos de variables de competencia.

Ruta	Precio Coche ()	Tiempo Coche(h)	Precio Bus ()	Tiempo Bus (h)	Precio Tren ()	Tiempo Tren (h)
XRY-BUD	502	24	200	45	200	25
XRY-AMS	367	30	130	35	150	17
XRY-SCQ	156	10	74	15	75	10

En esta tabla 9.6 se ha expuesto la diferencia de hacer la ruta en cada uno de los tres transportes que se contemplarán. Estos datos aplicados a la ecuación de competencia de modo de transporte nos darán la probabilidad de que los pasajeros se desplacen en cada modo de transporte.

Una vez se tienen los datos completos, se analizarán en concreto y a modo de ejemplo algunos conjuntos de variables en cada ruta.

Para ello se hará uso del Dashboard y sus diferentes pestañas en modo comparativo. Dividimos dicho dashboard en tres apartados:

- *Home: Información sobre la ruta y la competencia*
- *Aeropuerto de origen: Variables de origen*
- *Aeropuerto de destino: Variables de destino*

9.4.2 Análisis de las variables demográficas para el caso XRY-BUD

Por hacer un análisis más profundo de las variables individuales antes de analizar las variables output genéricas, se comienza por un análisis de las variables demográficas en el caso de la ruta Jerez-Budapest.

En primer lugar, atendiendo a la Población de ambas regiones, tendremos en cuenta la población en la región administrativa NUT-3. En el caso de la suma de población entre destino y origen, en este caso tendremos que las poblaciones de influencia son bastante mayores en Budapest. Aunque el área metropolitana de Budapest es mucho mayor que la de Jerez, hay numerosos núcleos turísticos con influencia en dicho aeropuerto, por lo que el porcentaje será diferente pero no tan definitorio.

En cuanto al PIB, en el caso de Jerez será muchísimo menor al de Budapest. Las actividades principales de la ciudad andaluza son la agricultura y el turismo, mientras que la capital húngara tiene más solidez económica. Aun así, precisamente por el turismo, este PIB no es tan definitorio.

El porcentaje de paro sí que es un punto en contra de la ciudad del aeropuerto de origen, que es bastante mayor que en la de destino. Esto supondrá un pequeño descenso de la demanda en esta ruta. Como en el caso de uso Jerez siempre es el aeropuerto de origen, esto no afectará a la decisión.

En el caso de las plazas hoteleras, los porcentajes son similares a los de población, puesto que se escogen el total de camas de la región de Andalucía (NUTS-2), pero el porcentaje de dichas camas del área de influencia del aeropuerto de origen es similar a la relación de poblaciones.

Se puede hacer la comparativa entre las dos siguientes imágenes:

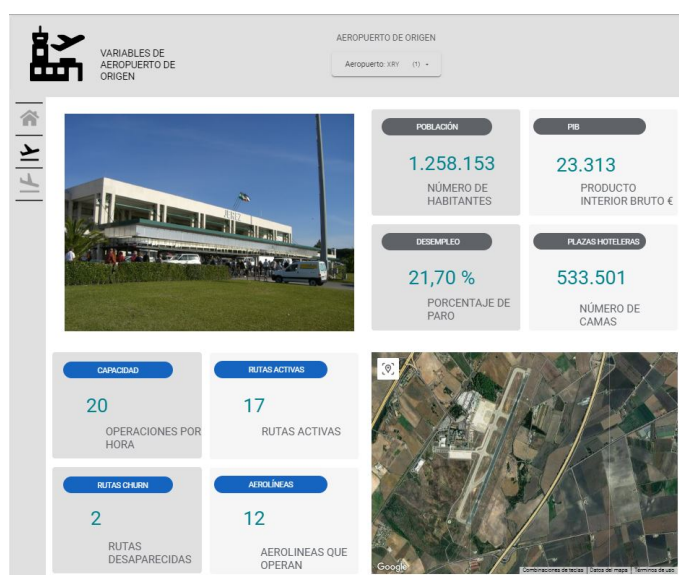


Figura 9.5 Dashboard de aeropuerto de origen. Elaboración propia.

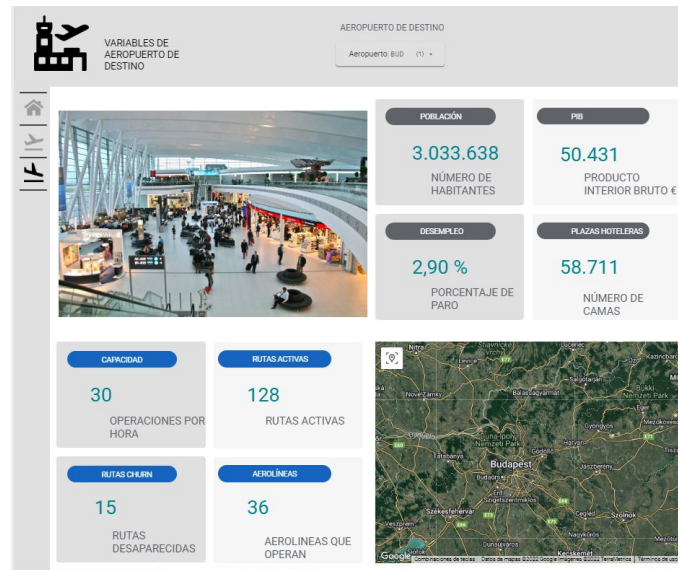


Figura 9.6 Dashboard de aeropuerto de destino. Elaboración propia.

9.4.3 Análisis de las variables aeroportuarias para el caso XRY-BUD

Haciendo uso de los KPIs inferiores de las figuras 9.5 y 9.6 se hace un análisis similar al anterior sobre las variables puramente aeroportuarias.

Con respecto a la capacidad se encuentran unas capacidades de entre 20 y 30 operaciones por hora en ambos aeropuertos. Combinándolas con el número de operaciones y aerolíneas, podemos llegar a la conclusión de que el caso del aeropuerto de Budapest será el más crítico de ambos en el caso de capacidades y oportunidades de conseguir slots.

En el caso de los niveles de accesibilidad, es aceptable en ambos aeródromos, nunca bajando del nivel 2 que se definió.

9.4.4 Análisis de las variables de ruta para el caso XRY-AMS

Analizando ahora el Dashboard de los datos básicos de ruta y suponiendo un gasto de combustible medio aproximado para la aeronave establecida (dependiendo de dicha aeronave el gasto cambiará), obtenemos una serie de valores que se resumen en que la ruta sería de menos de 2000km y se realizaría en unas 3 horas de vuelo. Estas variables tendrán su utilidad en el cálculo de los costes y en el cálculo de la probabilidad que hay de que un pasajero escoja el avión en vez de otro modo de transporte.

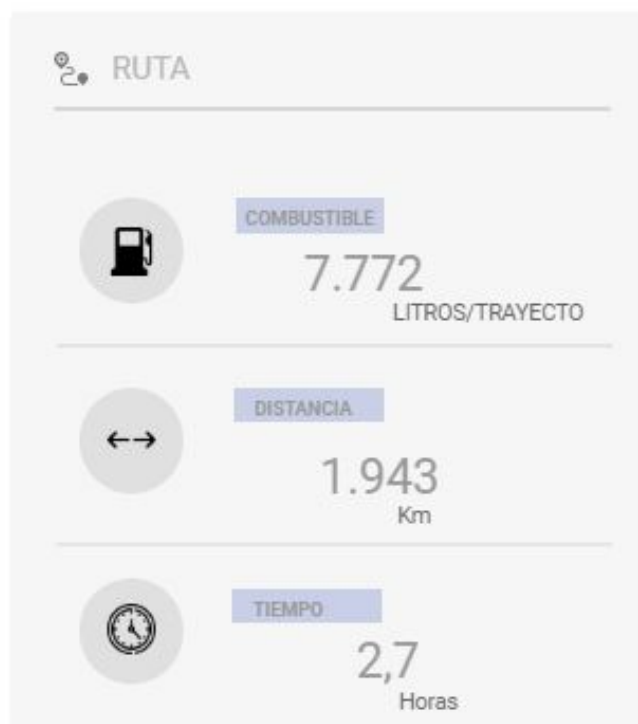


Figura 9.7 Variables de ruta XRY-AMS. Elaboración propia.

9.4.5 Análisis de las variables de competencia de modo de transporte para el caso XRY-SCQ

Por último, se toma como ejemplo el caso de la ruta Jerez-Santiago de Compostela para hacer una comparativa de costes y tiempos de viaje entre estas dos ciudades en tres medios de transporte diferentes:

- *Coche*: Estaríamos hablando de unas 10 horas de viaje y un precio total de ida de 156 €. Aquí merece la pena detenerse de nuevo en el sentido de que nuestro modelo está suponiendo siempre que un pasajero va sólo en su vehículo. Sin embargo, si este fuese acompañado, el precio disminuiría considerablemente y la utilidad de esta variable variaría. En cualquier caso, este trayecto es frecuentemente realizado en coche propio, por lo que la probabilidad de escoger esta alternativa se intuye elevada.
- *Tren*: Estaríamos hablando de unas 10 horas de viaje al existir la necesidad de hacer transbordo. El precio total sería de 75 €, un precio que es competitivo con respecto al coche hasta que en el vehículo van más de tres personas, cuando se vuelve mucho menos útil.
- *Bus*: El autobús es la alternativa que menos probabilidad tendrá pues es la mayor en tiempo, y similar en precio a las anteriores.

En el dashboard tendría el aspecto de la figura 9.8.



Figura 9.8 Variables de competencia XRY-SCQ. Elaboración propia.

9.5 Toma de decisiones: La ruta más rentable

Sólo con hacer uso del Dashboard para los tres casos indicados, observando los KPIs principales, podemos ver cómo tanto en demanda como en precio, la ruta más rentable sería, a priori, la ruta Jerez-Budapest. Este apartado sería un punto tedioso y argumentado en bases poco sólidas sin la ayuda del sistema que se ha elaborado durante este proyecto. A continuación, se muestra la comparativa del Dashboard con las tres opciones que es el resultado de la toma de decisiones a la que se ha hecho referencia durante todo el documento.

Con esto, no se está asegurando ni afirmando que la decisión tomada sea la más acertada, pero si tenemos una base argumentativa sólida y basada en datos reales sobre la que establecer nuestra decisión, que era el principal objetivo de la creación de este sistema.

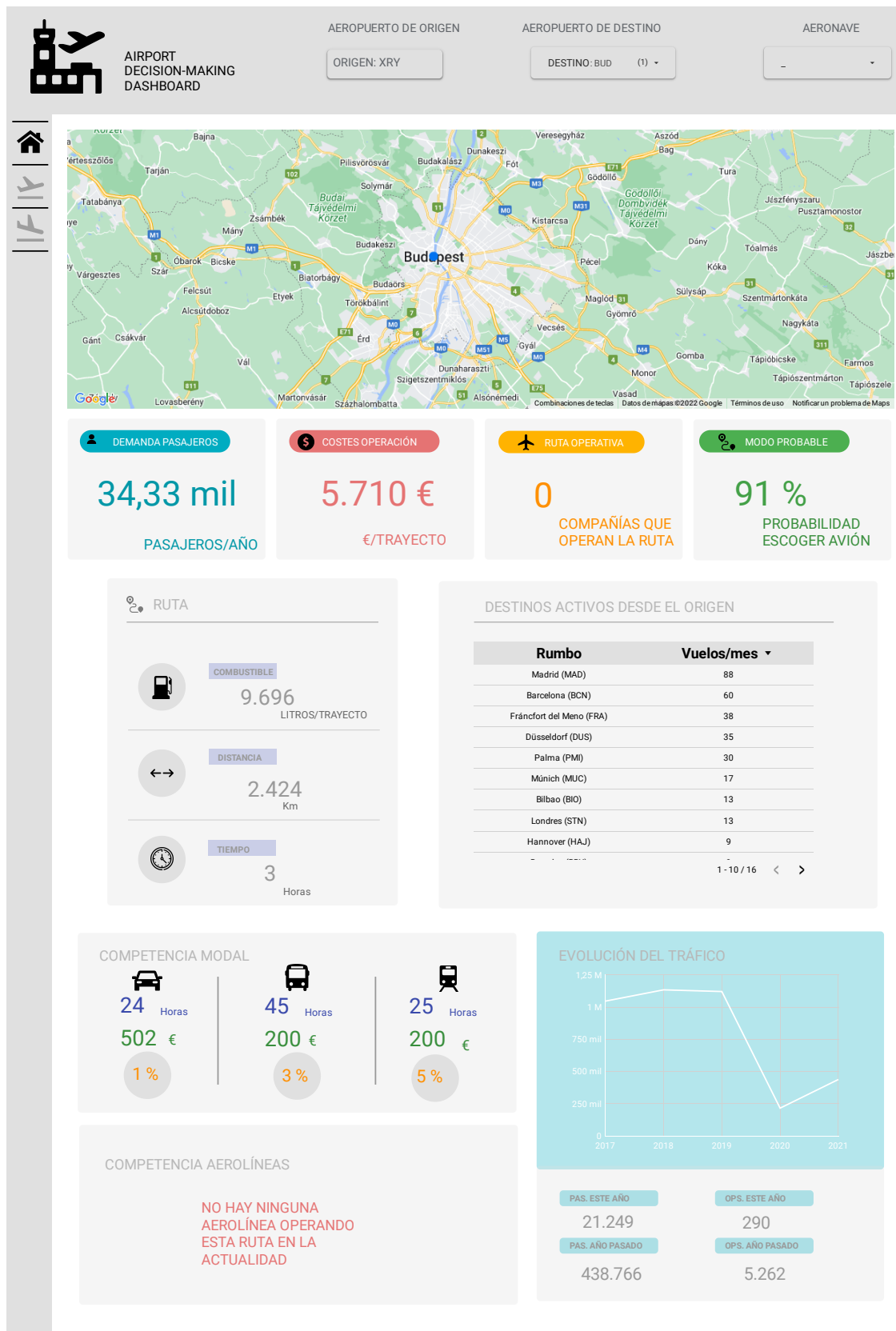


Figura 9.9 XRY-BUD. Fuente: DataStudio.

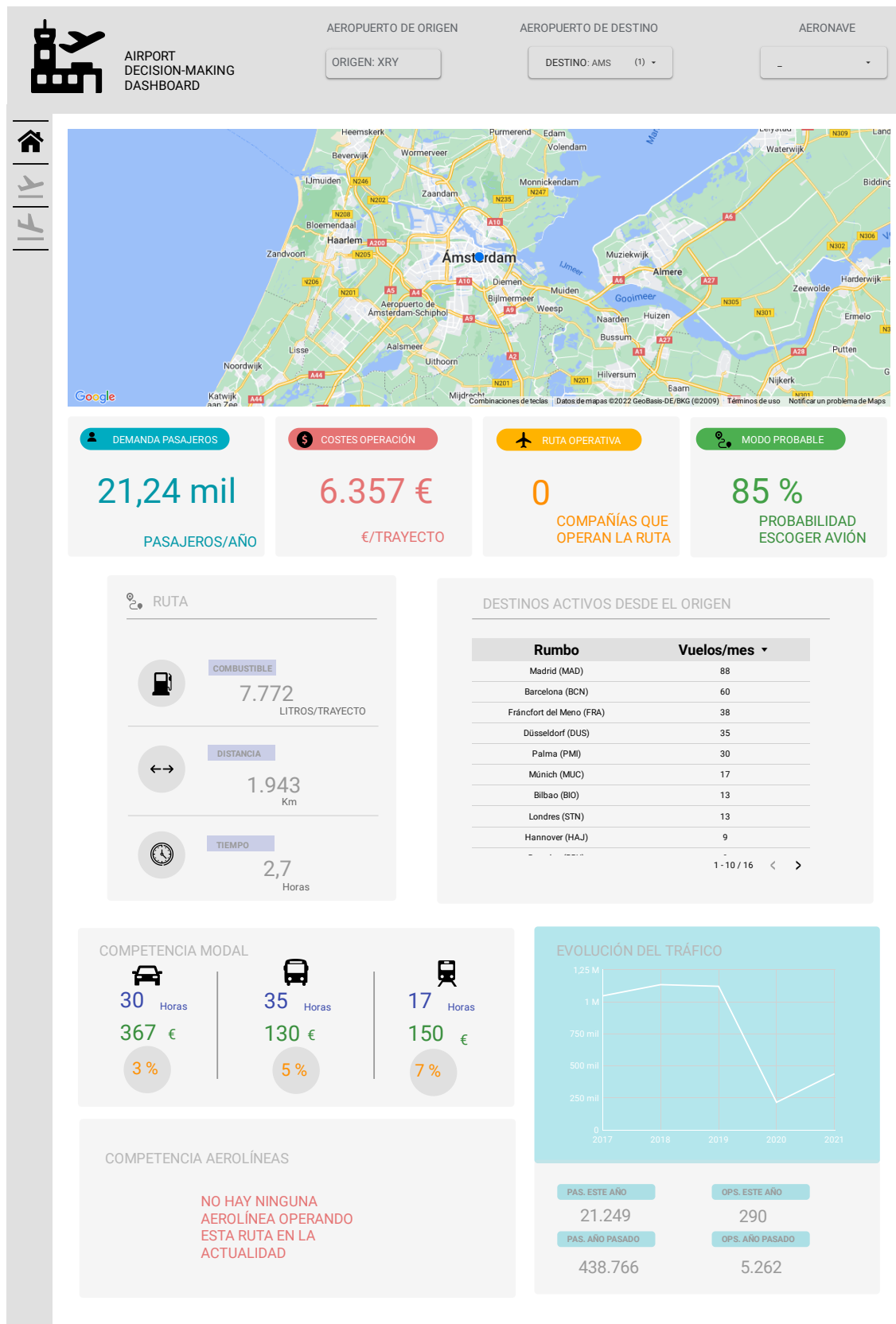


Figura 9.10 XRY-AMS. Fuente: DataStudio.

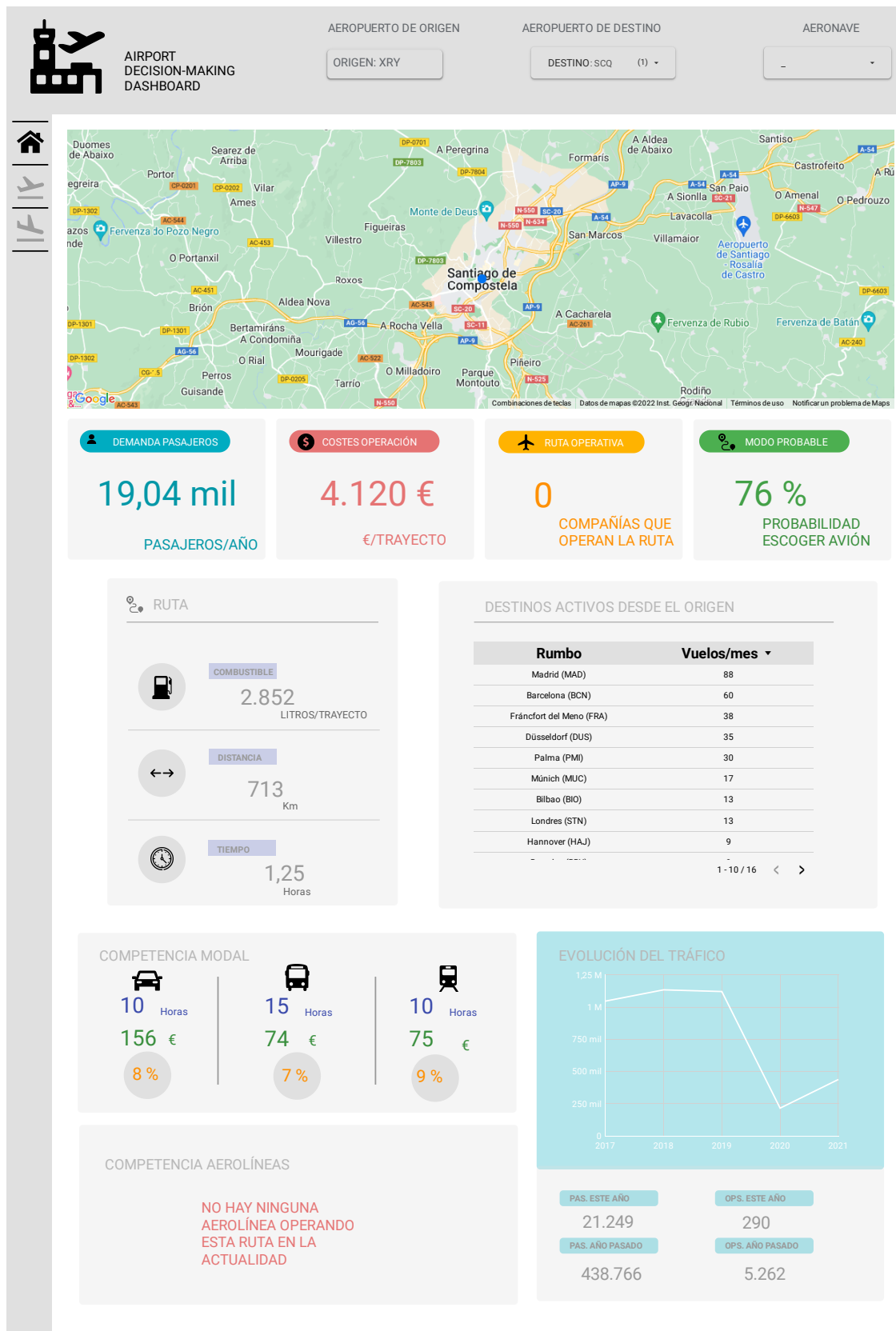


Figura 9.11 XRY-SCQ. Fuente: DataStudio.

10 Conclusiones

Este trabajo de Fin de Grado ha sido una combinación de lo aprendido tanto en mi experiencia laboral como académica. El modelo se basa en el tratamiento de datos, que es una ciencia en auge en cualquier ámbito profesional, dentro y fuera del mundo aeronáutico.

La metodología seguida, es fruto de los conocimientos que se han aportado en la escuela, pero también responde a la metodología que en mi día a día sigo como analista de datos en mi compañía.

Tal y como se ha comentado en diversas ocasiones, se ha presentado una simplificación de lo que sería un modelo mucho más completo cuanto más variables se tratasen y en más profundidad se estudiasen, pero no por ello carece de validez en el ámbito profesional.

Los resultados presentados dan fe de ello, y ponen de manifiesto que el modelo sería de utilidad a los diversos agentes aeronáuticos presentes en una decisión de este tipo, en concreto para un directivo de una aerolínea o un gestor del aeropuerto.

La complejidad que el sistema puede ir adquiriendo se presenta como una oportunidad para seguir profundizando en este tema en una siguiente etapa académica o incluso dentro de una aerolínea.

Con una inversión tanto técnica como humana de un equipo de analistas de datos lo suficientemente experimentado, podríamos hacer escalable este sistema incluso como un negocio que se ofreciese de manera externa a las compañías.

No obstante, el tiempo y los recursos empleados en este trabajo no sólo me han aportado en la creación directa del sistema, me han permitido obtener y mejorar una serie de habilidades técnicas como el manejo de software de visualización de datos, de esquematización de conexiones de bases de datos y, sobre todo, en el ámbito del llamado “Data Storytelling”, una disciplina crucial en cualquier ámbito científico.

Además, en el ámbito de las variables que intervienen en este tipo de decisiones, este trabajo ha servido para hacerse una idea de la complejidad de la interrelación entre estas y de la dificultad de que se mantengan actualizadas las fuentes de datos.

Como conclusión, aunque este trabajo no sea autosuficiente a este nivel de profundidad para permitir tomar una decisión de este calibre (tanto económica como estratégicamente hablando), se han sentado las bases de lo que, en un futuro, podría llegar a ser un sistema definitorio en la toma de decisiones.

Índice de Figuras

1.1	Ratios de éxitos de las tres grandes aerolíneas americanas (Elaboración propia)	2
1.2	Evolución de aperturas de rutas en aeropuertos de la red AENA. (Elaboración propia)	3
1.3	Recta de regresión previsión apertura de nuevas rutas en España (Elaboración propia)	4
2.1	Diagrama de flujo de la metodología seguida (Elaboración propia)	8
3.1	Porcentaje de pasajeros en los últimos 15 meses por motivo de viaje (Elaboración propia)	17
3.2	Distancia ortodrómica SVQ-PMI. Fuente: Google Maps	20
3.3	Ruta de vuelo SVQ-PMI. Fuente: Aviapages	21
3.4	Mapa de aeropuertos cercanos al aeropuerto de sevilla. Fuente: Air Calculator	23
3.5	Listado de aeropuertos cercanos al aeropuerto de sevilla. Fuente: Air Calculator	24
3.6	Estructura característica de costes en una aerolínea tradicional. Fuente: AEA	24
4.1	Diagrama de bloques de la infraestructura de datos (Creación propia)	27
4.2	Evolución sobre las publicaciones sobre datos en aviación. Fuente: Transportation Research	27
4.3	Esquema bases de datos. Elaboración propia	28
4.4	Definición regiones NUTS. Fuente: Rubén F. Bustillo-Amazon Web Services	29
4.5	Conexiones Bases de datos Eurostat (DB1). Elaboración propia.	30
4.6	Conexiones Bases de datos OLTE (DB2). Elaboración propia.	32
4.7	Dashboard Interactivo. Fuente: Aena	33
4.8	Conexiones Bases de datos AENA (DB3). Elaboración propia.	34
4.9	Comparativa búsqueda. Fuente: Google Trends	34
4.10	Conexiones Bases de datos Google Trends (DB4). Elaboración propia.	35
4.11	Interfaz de búsqueda. Fuente: Flight Routes Calculator.	35
4.12	Interfaz de resultados. Fuente: Flight Routes Calculator.	36
4.13	Conexiones Bases de datos Flight Route Calculator (DB5). Elaboración propia.	37
4.14	Interfaz de búsqueda Omio. Fuente: Omio.	37
4.15	Conexiones Bases de datos OMIO (DB6). Elaboración propia.	38
4.16	Joins o uniones en el modelo. Fuente: Google Data Studio.	40
4.17	Estructura general de datos. Creación propia	41
5.1	Diagrama de flujo del bloque estadístico del sistema (Elaboración propia)	43
7.1	Comparación software visualización de datos. Creación propia	55
7.2	Conexión fuente de datos. Fuente: DataStudio	56
7.3	Configuración de selectores. Fuente: DataStudio	56
7.4	Previsualización del encabezado. Fuente: DataStudio	57
7.5	Gráfico tipo mapa de ruta. Fuente: DataStudio	57
7.6	KPI's numéricos simples. Fuente: DataStudio	58
7.7	KPI's relativos a ruta. Fuente: DataStudio	58
7.8	KPI's relativos a competencia. Fuente: DataStudio	59
7.9	Listado de rutas ya operadas. Fuente: DataStudio	59

7.10	Histórico de tráfico. Fuente: DataStudio	60
7.11	Dashboard Home. Fuente: DataStudio	61
9.1	Vista aérea del Aeropuerto de Jerez. Fuente: Google Maps	68
9.2	Vista interior del Aeropuerto de Jerez. Fuente: Google Maps	68
9.3	Rutas aéreas activas desde el aeropuerto de Jerez. Elaboración propia	69
9.4	Evolución del tráfico en el aeropuerto de Jerez. Elaboración propia	70
9.5	Dashboard de aeropuerto de origen. Elaboración propia	73
9.6	Dashboard de aeropuerto de destino. Elaboración propia	74
9.7	Variables de ruta XRY-AMS. Elaboración propia	75
9.8	Variables de competencia XRY-SCQ. Elaboración propia	75
9.9	XRY-BUD. Fuente: DataStudio	77
9.10	XRY-AMS. Fuente: DataStudio	78
9.11	XRY-SCQ. Fuente: DataStudio	79

Índice de Tablas

2.1	Valores de parámetros	12
3.1	Pasajeros por motivo de viaje	17
3.2	Categorías de accesibilidad	19
4.1	Bases de datos de Eurostat	30
4.2	Bases de datos de OTLE	32
4.3	Bases de datos de AENA	33
4.4	Bases de datos de Google Trends	35
4.5	Bases de datos de Flight Route Calculator	36
4.6	Bases de datos de OMIO	38
4.7	Dimensiones del modelo	39
4.8	Tipos de uniones	40
6.1	Variables Demográficas Input	49
6.2	Variables Aeroportuarias Input	50
6.3	Variables De Ruta Input	51
6.4	Variables De Alternativa Input	52
9.1	Valores de capacidad y parámetros	68
9.2	Niveles de tráfico actuales-Mayo 2022	69
9.3	Datos de variables demográficas	71
9.4	Datos de variables aeroportuarias	72
9.5	Datos de variables de ruta	72
9.6	Datos de variables de competencia	72

Índice de Códigos

5.1	Código 1 Matlab	44
5.2	Código 2 Matlab	45

Bibliografía

- [1] Santiago González Varas Ibañez. La liberalización comunitaria del transporte aéreo. *Revista de derecho comunitario Europeo*, 2001.
- [2] Richard Doganis. *The airline industry in the 21st century*. Routledge, 2001.
- [3] U.S. Department of Transportation. Bts releases 2015 airline financial data. *BTS Airline Financial Data*, 2015.
- [4] Dominic Dudley. The 1 billion prize: The world's 10 most lucrative airline routes. *Forbes*, 2019.
- [5] Vasu Raja. After dropping chicao – beijing, american airlines considers ending all chicao – asia flying. *View from the wing*, 2018.
- [6] Aena. Datos de nuevas rutas en aeropuertos de la red aena. *Dirección de Datos y Mercado Aeronáutico*, 2018-2021.
- [7] Bresnahan y Reiss. Entry in monopoly markets. *Review of Economic Studies*, 57, 531-544, 1990.
- [8] S Berry. Estimation of a model of entry in the airline industry. *Econometrica* 60, 889-917, 1992.
- [9] Ito y Lee Boguslaski. Entry patterns in the southwest airlines route system. *Review of Industrial Organization Vol. 25, No. 3 pp. 317-350*, 2004.
- [10] S Whinston, M Collins. Entry and competitive structure in deregulated airline markets: an event study analysis of people express. *Rand Journal of Economics* 23 445-462, 1992.
- [11] Oliveira. An empirical model of low-cost carrier entry. *Transportation Research Part A* 42, 2008.
- [12] de Wit Zuidberg. Route churn: an analysis of low-cost carrier route continuity in europe. *Journal of Transport Geography*, 2015.
- [13] Kim Park. A deeplearning approach to analyze airline customer propensities: The case of south korea. *Appl. Sci.*, 2022.
- [14] Gregory Mankiw. *Principios de Economía*. Cengage Learning Editores,, 2012.
- [15] Levin y Rubin. *Estadística para administración y economía*. Pearson, 1996.
- [16] Stock y Watson. *Introducción a la econometría*. Pearson, 2012.
- [17] Álvarez Becerra Cáceres Osorno Rodríguez. La oferta, la demanda y el mercado. *Introducción a la teoría económica*, 2013.
- [18] Martijn Brons Eric Pels Peter Nijkamp Piet Rietveld. Price elasticities of demand for passenger air travel: a meta-analysis. *Journal of Air Transport Management* 8, 2002.
- [19] Jose María del Castillo Granados. Apuntes de la asignatura explotación del transporte aéreo. *ETSI*, 2021.

- [20] Heinzl Grosche, Rothlauf. Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, 2007.
- [21] Nguyen. Understanding determinants of making airlinerouteentry and exit decisions: An application of logit models. *Collegiate Aviation Review International*, 36, 2018.
- [22] Michal Šimeček. Discrete choice analysis of travel behaviour. *Transactions on Transport Sciences*, 2019.
- [23] Yai Fukuda. Semiparametric specification of the utility function in a travel choice model. *Journal of Air Transport Management*, 2010.
- [24] Movilidad y Agenda Urbana Ministerio de Transportes. Planes directores. <https://www.mitma.gob.es/areas-de-actividad/aviacion-civil/politicas-aeroportuarias/integracion-territorial-aeroportuaria/planes-directores/planes-directores-por-comunidad-autonoma>, 2022.
- [25] Movilidad y Agenda Urbana Ministerio de Transportes. «boe» núm. 27, de 01/02/2021. [https://www.boe.es/eli/es/res/2021/01/21/\(2\)/con](https://www.boe.es/eli/es/res/2021/01/21/(2)/con), 2021.
- [26] Antonio Moreno Jiménez. Manual de autoaprendizaje con arc gis. <https://dialnet.unirioja.es/servlet/articulo?codigo=4057516>, 2008.
- [27] INE. Número de turistas según motivo principal del viaje. <https://www.ine.es/jaxiT3/Tabla.htm?t=13864>, 2022.
- [28] Norberto Cabrera Alonso. Demanda/capacidad en los aeropuertos. ICAO, 2006.
- [29] Danwen Bao. Relevance of airport accessibility and airport competition. *Journal of Air Transport Management*, 2016.
- [30] Flight Route Calculator. Aviapages. <https://aviapages.com/flightroutecalculator/>, 2022.
- [31] Movilidad y Agenda Urbana Ministerio de Transportes. España cierra el espacio aéreo a las aerolíneas rusas como parte del paquete de sanciones impuestas a rusia por el ataque a ucrania. <https://www.mitma.gob.es/el-ministerio/sala-de-prensa/noticias/lun-28022022-0827>, 2022.
- [32] Zhe Yuan. Essays on network competition in the airline industry. *University of Toronto*, 2016.
- [33] Clavé y Saladié. How different are tourists who decide to travel to a maturedestination because of the existence of a low-cost carrier route? *Journal of Air Transport Management*, 2014.
- [34] Centro de publicaciones Secretaría General Técnica Ministerio de Fomento. Ánalysis comparativo de costes de escala en los principales aeropuertos europeo. *Serie Monografías*, 2009.
- [35] AENA. Guía de tarifas. *Edición Enero*, 2022.
- [36] Mariano Domingo. Descubrir el handling aeroportuario - 2ª edición. AENA, 2012.
- [37] Editorial: Chung y Hamsen. Data science and analytics in aviation. *Transportation Research Part E*, 2020.
- [38] IATA. Airline and location code search. <https://www.iata.org/en/publications/directories/code-search/>, 2022.
- [39] Comisión Europea. Eurostat: estadísticas europeas. <https://ec.europa.eu/info/departments/eurostat-european-statistics>, 2022.
- [40] Eurostat. Nuts classification. <https://ec.europa.eu/eurostat/web/nuts/statistics-illustrated>, 2016.
- [41] Rubén F. Bustillo. Working with the nomenclature of territorial units for statistics of eurostat. <https://rpubs.com/rubenfb>, 2020.
- [42] Observatorio de Transporte y Logística de España. Bases de datos otle. <https://apps.fomento.gob.es>, 2022.

- [43] AENA. Estadísticas de tráfico aéreo. <https://www.aena.es/es/estadisticas/inicio.html>, 2022.
- [44] Global Stats. Browser market share worldwide. <https://gs.statcounter.com/browser-market-share>, 2022.
- [45] Blanca Zayas. El 80% de la compra de vuelos en España se realiza por internet, tripadvisor. <https://www.larazon.es/historico/8630-el-80-de-la-compra-de-vuelos-en-espana-se-realiza-por-internet->, 2011.
- [46] Ayuda de Tendencias de búsqueda. Explicación de la página principal. <https://support.google.com/trends/answer/6248105?hl=es&topic=6248052>, 2022.
- [47] OMIO. Viajes que emocionan. <https://www.omio.es/>, 2022.
- [48] Ayuda de Google Analytics. Dimensiones y métricas. <https://support.google.com/analytics/answer/1033861?hl=es-419>, 2022.
- [49] Ayuda de editores de documentos de Google. <https://support.google.com/docs/>, 2022.
- [50] Miro. La pizarra virtual para facilitar la colaboración. <https://miro.com/es/pizarra-virtual/>, 2022.
- [51] MathWorks. <https://www.mathworks.com/discovery/what-is-matlab.html>, 2022.
- [52] Google Support. Como funcionan las combinaciones. *Ayuda de DataStudio*, 2022.
- [53] Belsley y Welsch. Regression diagnostics. *John Wiley y Sons, Inc.*, 1980.
- [54] MathWorks. Help center-regstat. <https://www.mathworks.com/help/stats/regstats.html>, 2022.
- [55] OACI. Volumen i, diseño y operaciones de aeródromos. *Anexo 14, Aeródromos*, 1990.
- [56] Ian Humphreys and Graham Francis. Performance measurement: a review of airports. *International Journal of Transport Management*, 2002.
- [57] Andersson Granberg. Developing key performance indicators for airports. *Int. Workshop on ATM/CNS*, 2013.
- [58] Chris Stolte. ¿qué es tableau? www.tableau.com, 2022.
- [59] Microsoft Editorial. ¿qué es powerbi? www.powerbi.microsoft.com, 2022.
- [60] Google Support. Te damos la bienvenida a data studio. www.support.google.com/datastudio/answer/6283323?hl=es, 2022.
- [61] Yolanda Álvarez Castaño y M^a Isabel Alonso Magdaleno. El proceso de validación de los modelos socio-económicos dinámicos. <https://econo.uniovi.es/c/documentlibrary/getfile?uuid=ee30957a-f2ac-4ab5-9e09-6b86f6d00d67&groupId=746637>, 2000.
- [62] AENA. Propuesta de revisión del plan director del aeropuerto de Jerez. <https://www.mitma.gob.es/recursosmfom/audienciainfopublica/recursos/cap2.xry.situacionactual.pdf>, 2019.
- [63] Javier Rodríguez. Mejorar las conexiones del aeropuerto de Jerez es clave para el turismo de la provincia de Cádiz. *La voz de Cádiz*, 2022.
- [64] E. M. CAÑAS. El aeropuerto de Jerez tendrá una veintena de destinos este verano. *Diario de Jerez*, 2022.
- [65] Redacción. El aeropuerto de Jerez recibió en mayo casi 96.000 pasajeros, la mitad visitantes extranjeros. *Radio Jerez*, 2022.
- [66] Redacción. El patronato de turismo ultima el mayor contrato de su historia: hasta 1,3 millones para atraer vuelos al aeropuerto de Jerez en invierno. *Diario Bahía de Cádiz*, 2022.

Índice alfabético

- β_P , 44
 β_{Bus} , 44
 β_{Des} , 44
 β_{Dis} , 44
 β_{Hot} , 44
 β_{Och} , 44
 β_{PIB} , 44

 2018, 70
 2019, 70
 2020, 70
 2021, 70

 Aerolínea A, 47
 Aerolínea A:, 13
 Aerolínea B, 47
 Aerolínea B:, 13
 Aerolínea C, 47
 Aerolínea C:, 13
 Aerolínea low cost o tradicional:, 13
 Aerolínea low-cost vs tradicional, 22
 Aerolínea LowCost, 47
 Aeronave de operación, 56
 Aeropuerto A, 47
 Aeropuerto B, 47
 Aeropuerto C, 47
 Aeropuerto de destino, 56
 Aeropuerto de destino: Variables de destino, 72
 Aeropuerto de origen, 56
 Aeropuerto de origen: Variables de origen, 72
 Air Nostrum, 70
 Airway distance|hyperpage, 35
 Airway time|hyperpage, 36
 Antiguo, 47
 Antigüedad en la ruta, 23
 Antigüedad operando la ruta:, 13
 Autobús, 19
 Avión, 45

 Binter, 70
 Bus, 44, 45, 75

 Capacidad de las terminales (pasajeros/hora) de los aeropuertos de la red de Aena por comunidad autónoma|hyperpage, 31
 Capacidad del campo de vuelo (aeronaves/hora) de los aeropuertos de la red de Aena por comunidad autónoma|hyperpage, 30

 Climatología, 18
 Coche, 45, 75
 Coche propio, 19
 Combinación Externa Completa, 40
 Combinación Externa Derecha, 40
 Combinación Externa Izquierda, 39
 Combinación Interna, 39
 Combustible necesario de la ruta, 54
 Competencia de aerolíneas, 54
 Competencia de aeropuerto, 54
 Competencia modal, 54
 Condor, 70
 Coste de tripulación, 48
 Coste handling en rampa, 48
 Coste handling pasajero, 48
 Costes de la aviación comercial de las principales compañías aéreas españolas|hyperpage, 31
 Costes de operación, 54
 Cruzada, 40

 Demanda de pasajeros por año, 54
 Demanda:, 9
 Des, 44
 Destinos activos desde el origen, 54
 Dimensión, 38
 Dis, 44
 Distancia de la ruta, 48, 54
 Distribución sectorial de las actividades, 15

 Entrada:, 8
 Espacios naturales existentes, 15
 Eurowings, 70
 Evolución del tráfico, 54
 Existencia clase Business:, 13

 Factor de ocupación, 48
 Flexibilidad, 47
 Flexibilidad Alternativa|hyperpage, 37
 Flexibilidad en el ticket, 23
 Frecuencia, 46
 Frecuencia de la alternativa, 22

 Google Data Studio, 39, 55
 Google Sheets, 39
 Great Circle distance|hyperpage, 36
 Great circle time|hyperpage, 36
 Gross domestic product (GDP) at current market prices by NUTS 2 regions|hyperpage, 29
 Handling al pasajero, 25

 Handling en rampa, 25
 Home: Información sobre la ruta y la competencia, 72
 Hot, 44

 Iberia, 70
 Impuestos Operación, 48
 Indicador de entrada, 54
 Información de rutas por aeropuertos|hyperpage, 33
 Información por compañía aérea|hyperpage, 33
 Ingresos medios del pasajero:, 13
 Input:, 9

 Jerez(XRY)-Budapest(BUD), 71
 Jerez(XRY)-Santiago de Compostela(SCQ), 71
 Jerez(XRY)-Ámsterdam Schiphol(AMS), 71

 Longitud (metros) de las pistas de los aeropuertos de la red de Aena|hyperpage, 31
 Lufthansa, 70
 Luxair, 70

 Mapa de la ruta, 54
 Matlab, 39
 Mercado de trabajo, 15
 Metro, 19
 Miro, 39
 Modelo aeronave, 48
 Modo más probable, 54
 Modos de acceso (distribución porcentual) a los aeropuertos de la red de Aena|hyperpage, 31
 Muestra:, 9
 Métrica, 38

 Negocio, motivos profesionales, 17
 Nights spent at tourist accommodation establishments by NUTS 2 regions|hyperpage, 29
 Nivel educativo, 18
 Number of establishments and bed-places by NUTS 2 regions|hyperpage, 29
 NUTS-0, 28
 NUTS-1, 28
 NUTS-2, 28
 NUTS-3, 28
 Número de aeropuertos cercanos al origen o destino, 23

- Och, 44
 Ocio, recreo o vacaciones, 17
 Octavia Airlines, 70
 Operaciones O/D|hyperpage, 33
 Otros motivos, 17
 Output:, 9

 P, 44
 Par aeropuerto Origen-Destino:, 8
 Personal en cabina, 25
 PIB, 44
 Pilotos, 25
 Población, 15
 Población:, 9
 Population on 1 January by age, sex and NUTS 2 region|hyperpage, 29
 Power BI, 55
 Precio actual del combustible, 48
 Precio Alternativa|hyperpage, 37
 Precio de las alternativas, 22
 Precio del ticket, 45, 46

 Precio medio histórico, 47
 Precio medio histórico de ticket, 22
 Precio:, 13

 Ruta ya existente, 47
 Ruta ya operativa, 54
 Ryanair, 70

 Sistema de transporte, 15

 Tableau, 55
 Tasas aeronave, 48
 Tasas de aeronave, 24, 25
 Tasas de pasajero, 24
 Tasas pasajero, 48
 Taxi-VTC, 19
 Tiempo Alternativa|hyperpage, 37
 Tiempo de la ruta, 54
 Tiempo de viaje, 22, 46
 Tiempo de vuelo, 48
 Tren, 19, 45, 75

 Tráfico de pasajeros y mercancías en aeropuertos españoles (número de asientos y operaciones) por tipo de tráfico, aeropuerto y compañías más representativas|hyperpage, 31
 Tui, 70

 Unemployment rate by NUTS 2 regions|hyperpage, 29

 Variable continua:, 9
 Variable cualitativa:, 9
 Variable cuantitativa:, 9
 Variable discreta:, 9
 Variable Dummy:, 9
 Variable endógena:, 9
 Variable exógena:, 9
 Vueling, 70

 Área de interés cultural, 17
 Índice de accesibilidad, 47

