

Proyecto 01



**Comparador de precios de vuelos:
encuentra las mejores ofertas para tus
vacaciones de navidad**

Equipo G: Daniel Muñoz, José Dos Reis y Pamela J. Colman Vega

Project 01



**Flight Price Comparator:
find the best deals for your Christmas
holidays.**

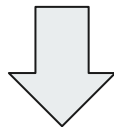
Team: Daniel Muñoz, José Dos Reis y Pamela J. Colman Vega

Motivaciones

Se acerca la navidad y año nuevo, ¿qué mejor que disfrutarlo en otro país?

22/12/23 - 26/12/23

22/12/23 - 26/12/23



Preguntas específicas

Motivaciones



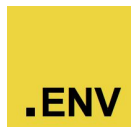
- Investigar la capacidad de extracción y análisis de datos sobre vuelos.
- Automatizar la búsqueda y extracción masiva de opciones de destinos y vuelos.
- Aplicar análisis y estadística sobre los datos y resultados para obtener información relevante, tendencias, conclusiones.
- Hallar las mejores opciones de vuelo basado en análisis de datos.

Alcance del proyecto



- Inicialmente, herramienta personalizada y personal de filtrado y explotación de datos, concretamente destinos y vuelos.
- Potencialmente expansible a un uso más profesional y comercial.
- Podría ser componente o base para servicios como comparadores de portales de vuelos, agencias de turismo, buscador de vuelos baratos, etc.

Herramientas Utilizadas



FOURSQUARE



Tools and Libraries



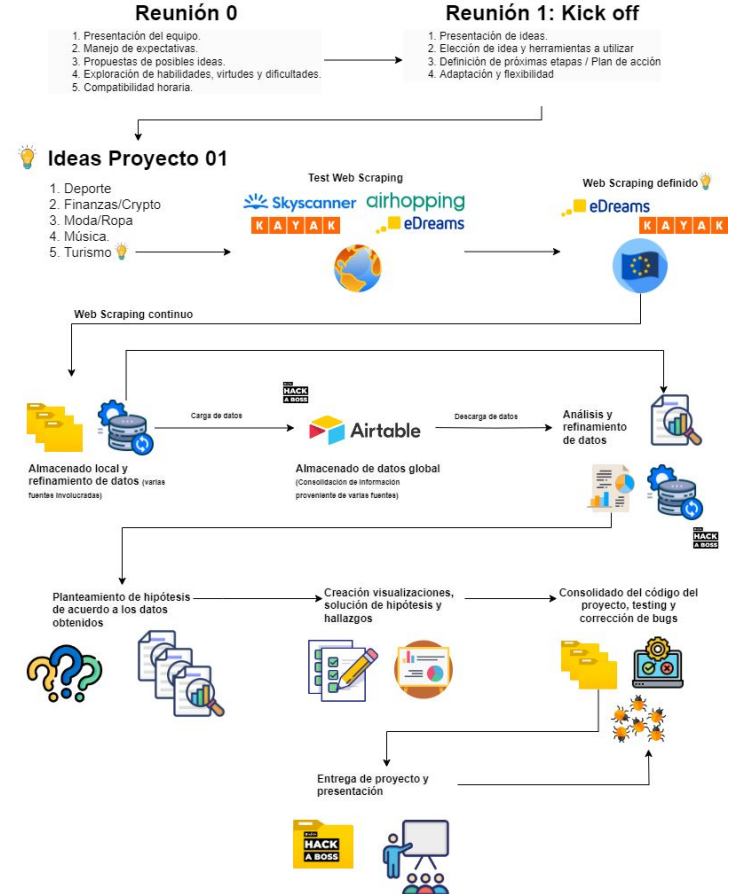
FOURSQUARE



Esquema del proyecto

- Sesiones grupales (daily meet).
- Sesiones individuales.
- Tutorías y asesoramiento.
- Metas cortas, alcanzables y con deadlines.

Proyecto 01



Desafíos durante el proyecto



1. **Alcance claro y tiempo ajustado:**
Establecer un alcance preciso desde el inicio para optimizar recursos y tiempo debido a plazos ajustados.
 2. **Extracción de datos compleja:**
Dificultades en la extracción debido a HTML/CSS dinámico, y medidas de seguridad de los portales.
 3. **Tratamiento de datos diversos:**
Unificación de datos de diferentes orígenes y decisiones sobre datos a utilizar.
 4. **Estudio y análisis de datos:**
Definición de preguntas específicas, agrupación de datos y manejo de variabilidad en la cantidad de datos.
- **Extra:** detección de funciones deprecadas en el programa entregado.

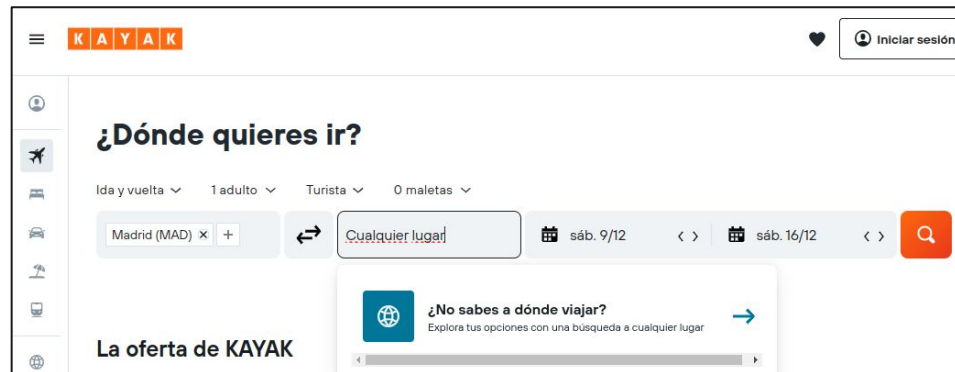
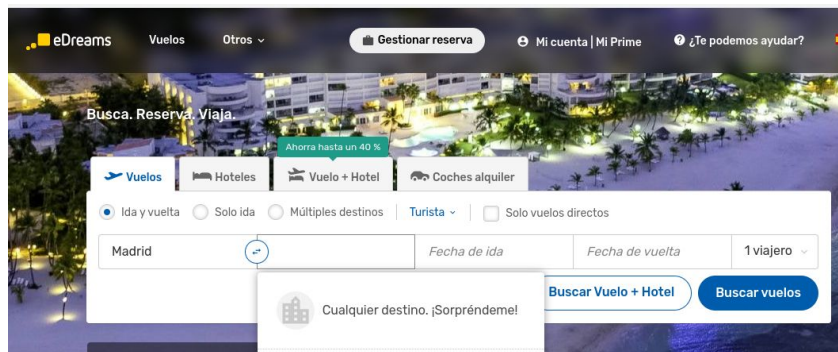
Solución a los desafíos



- 1. Gestión de alcance y entrega:**
 - Definir un alcance realista y conseguible debido a límites de tiempo y recursos.
 - Equitativa distribución de tareas para paralelizar el avance.
- 2. Abordar la extracción de datos compleja:**
 - Estudio del código fuente HTML y metadatos.
 - Uso de Selenium para realizar acciones en las webs y evitar controles y alertas.
 - Programación de una lógica personalizada para las acciones específicas de búsqueda y los patrones estáticos en el HTML.
- 3. Tratamiento de datos diversos:**
 - Establecer un modelo de datos común, basado en el origen más polivalente y adecuado.
 - Adaptación de los datos al modelo común independientemente del origen.
 - Apoyo en el uso de ficheros locales para evitar consumo de APIs y sus tiempos de espera .
- 4. Estudio y análisis de datos complejos:**
 - Establecimiento de criterios sobre los datos y variables relevantes dentro del conjunto de información.
 - Agrupación de variables según criterios.
 - Reconocimiento de limitaciones de datos.

Extracción de los datos: web scraping en eDreams y KAYAK

Ambos portales permiten búsqueda multidesino, por lo que el objetivo es, partiendo de un(os) origen(es) y fecha(s) preestablecidos, lanzar una búsqueda multidesino y, mediante scraping, recopilar todas las posibilidades que nos brinda cada uno de los portales.

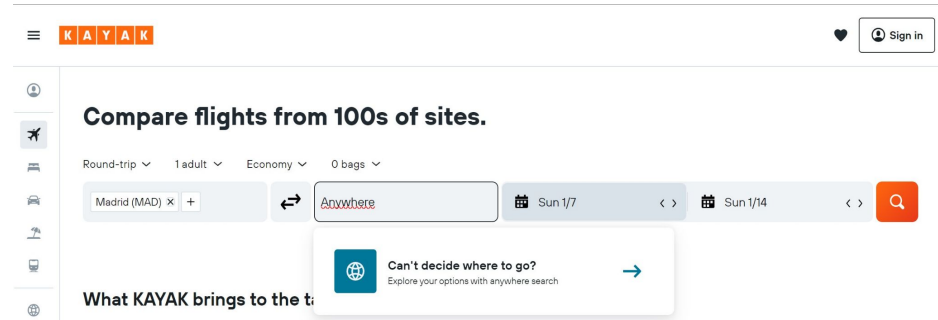
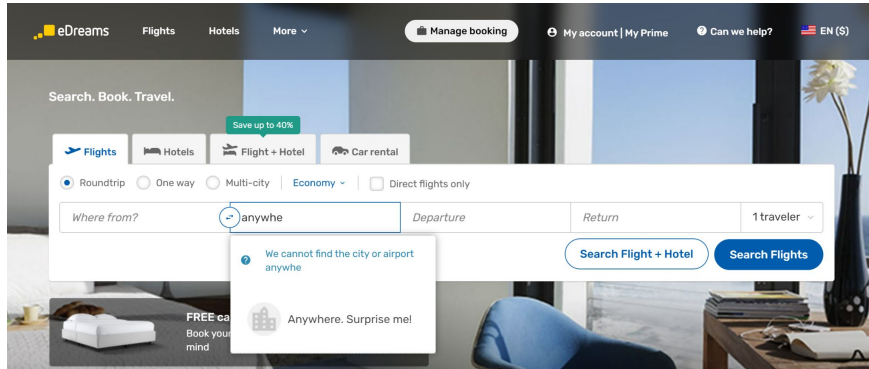


Los datos obtenidos son todas las combinaciones de destinos y vuelos ajustada al origen y fechas, con todos los detalles típicos de los vuelos:

→ Horarios, aeropuertos, aerolíneas, escalas, vuelos, clases, política de equipaje...

Data Extraction: Web Scrapping on eDreams and KAYAK

Both websites allow multi-destination searches, so the goal is to launch a search starting from one or more pre-established origin(s) and date(s), and through scraping, collect all the possibilities that each of the sites provides.



The obtained data includes all combinations of destinations and flights adjusted to the origin and dates, with all typical flight details:

→ *Schedules, airports, airlines, layovers, flights, classes, baggage policies, and more....*

Busca. Reserva. Viaja.

Ahorra hasta un 40 %

✈️ **Vuelos**

🛏️ Hoteles

✈️ Vuelo + Hotel

🚗 Coches alquiler

☒ Ida y vuelta ☐ Solo ida ☐ Múltiples destinos | **Turista ▾** | ☐ Solo vuelos directos

Madrid



Fecha de ida

Fecha de vuelta

1 viajero ▾



Cualquier destino. ¡Sorpréndeme!

Buscar Vuelo + Hotel

Buscar vuelos



KAYAK



Iniciar sesión



¿Dónde quieres ir?

Ida y vuelta ▾

1 adulto ▾

Turista ▾

0 maletas ▾

Madrid (MAD) × +



Cualquier lugar



sáb. 9/12



sáb. 16/12



La oferta de KAYAK



¿No sabes a dónde viajar?

Explora tus opciones con una búsqueda a cualquier lugar




DataFrame resultado

	url	origen	destino	fecha_inicio	fecha_fin	pasajeros	inicio_ida	fin_ida	inicio_vuelta	fin_vuelta
0	https://www.edreams.es/travel/#results/type=R;...	LCG	VLC	2024-01-05	2024-01-09	1	09:20	12:45	17:05	13:00
1	https://www.kayak.es/flights/MAD-MIL/2023-12-2...	MAD	MIL	2023-12-22	2023-12-26	1	12:20	14:35	6:00	14:55
2	https://www.edreams.es/travel/#results/type=R;...	PMI	BCN	2023-12-22	2023-12-26	1	06:40	07:35	17:15	18:05
3	https://www.edreams.es/travel/#results/type=R;...	MAD	TCI	2023-12-22	2023-12-26	1	12:20	17:30	17:20	21:10
4	https://www.edreams.es/travel/#results/type=R;...	LCG	LON	2023-12-29	2024-01-02	1	09:20	13:50	13:10	22:10

DataFrame resultado

inicio_vuelta	fin_vuelta	escala_ida	escala_vuelta	duracion_ida	duracion_vuelta	aerolineas	equipaje_mano	equipaje_bodega	precio	clase
17:05	13:00	1	3	3h 25m	19h 55m	[Iberia, Air Europa]	1	0	245	NaN
6:00	14:55	0	1	2h 15m	8h 55m	[Wizz Air Malta, Iberia, Ryanair]	0	0	181	Value, Básica
17:15	18:05	0	0	0h 55m	0h 50m	[Vueling]	0	0	254	NaN
17:20	21:10	1	0	6h 10m	2h 50m	[Iberia]	0	0	811	NaN
13:10	22:10	1	1	5h 30m	8h 00m	[Iberia]	1	0	1103	NaN

DataFrame filtrado



url	object
origen	object
destino	object
fecha_inicio	datetime64[ns]
fecha_fin	datetime64[ns]
pasajeros	int64
inicio_ida	object
fin_ida	object
inicio_vuelta	object
fin_vuelta	object
escala_ida	int64
escala_vuelta	int64
duracion_ida	timedelta64[ns]
duracion_vuelta	timedelta64[ns]
aerolineas	object
equipaje_mano	int64
equipaje_bodega	int64
precio	int64
clase	object
pagina_web	object
ciudad_origen	object
ciudad_destino	object
num_aerolineas	int64
dtype:	object

DataFrame



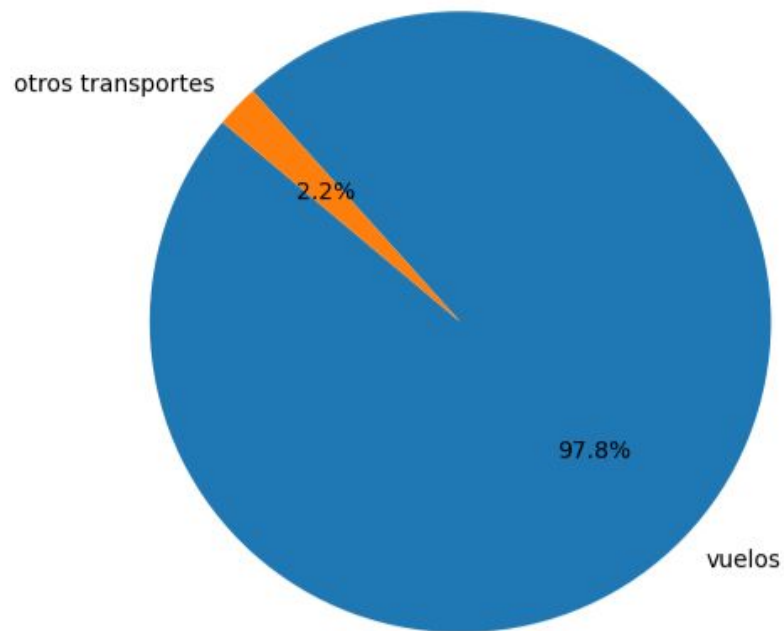
url	object
origen	object
destino	object
fecha_inicio	datetime64[ns]
fecha_fin	datetime64[ns]
pasajeros	int64
inicio_ida	object
fin_ida	object
inicio_vuelta	object
fin_vuelta	object
escala_ida	int64
escala_vuelta	int64
duracion_ida	timedelta64[ns]
duracion_vuelta	timedelta64[ns]
aerolineas	object
equipaje_mano	int64
equipaje_bodega	int64
precio	int64
clase	object
pagina_web	object
ciudad_origen	object
ciudad_destino	object
num_aerolineas	int64
dtype:	object

¿Preguntas específicas sobre los datos?

Distribución de precios de vuelos por ciudad de origen y destino

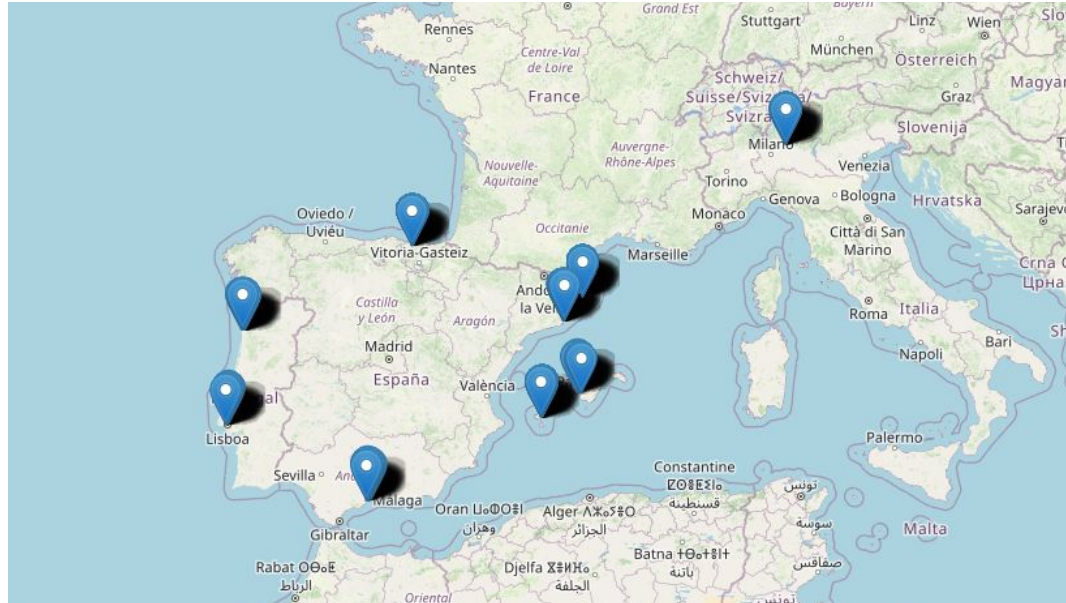


Proporción de vuelos y otros transportes



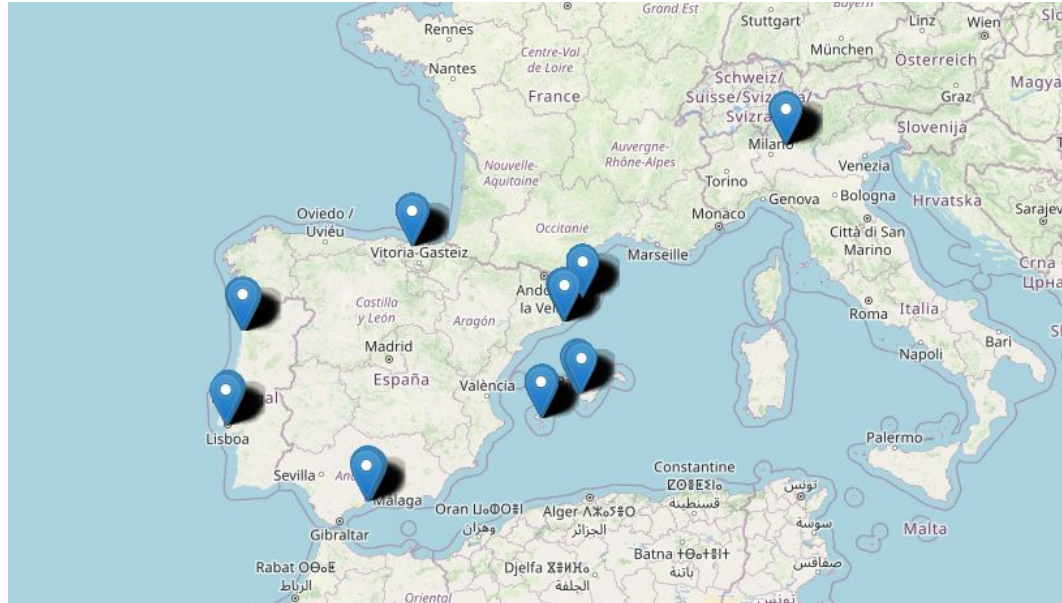
Del total de vuelos que nos ofrecen las dos páginas webs, un 2.2% se corresponden a viajes en trenes, buses y afines.

Mapa de puntos de interés en las ciudades halladas



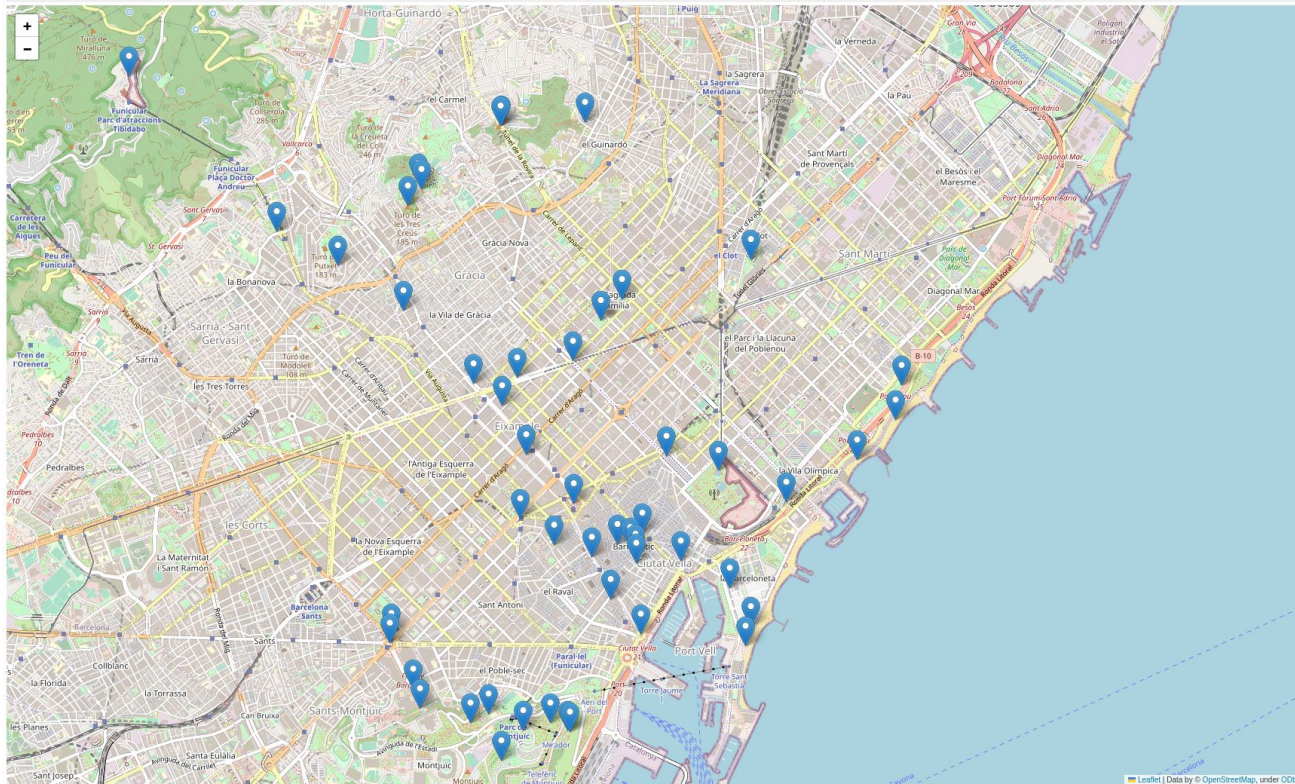
	ciudad_destino	precio
0	Barcelona	102.980551
1	Lisboa	137.808511
2	Milán	166.123457
3	Bilbao	168.653955
4	Málaga	172.450075
5	Ibiza	174.982544
6	Oporto	179.655063
7	Palma	184.856018
8	Gerona	195.000000
9	Bérgamo	212.475676

Point of Interest (POI) Map for the Cheapest Flight Tickets



	ciudad_destino	precio
0	Barcelona	102.980551
1	Lisboa	137.808511
2	Milán	166.123457
3	Bilbao	168.653955
4	Málaga	172.450075
5	Ibiza	174.982544
6	Oporto	179.655063
7	Palma	184.856018
8	Gerona	195.000000
9	Bérgamo	212.475676

Barcelona detailed POI





¿Cuántos destinos se obtienen en cada página web en cada una de las fechas establecidas?
¿Cuántos vuelos por destino se obtienen?

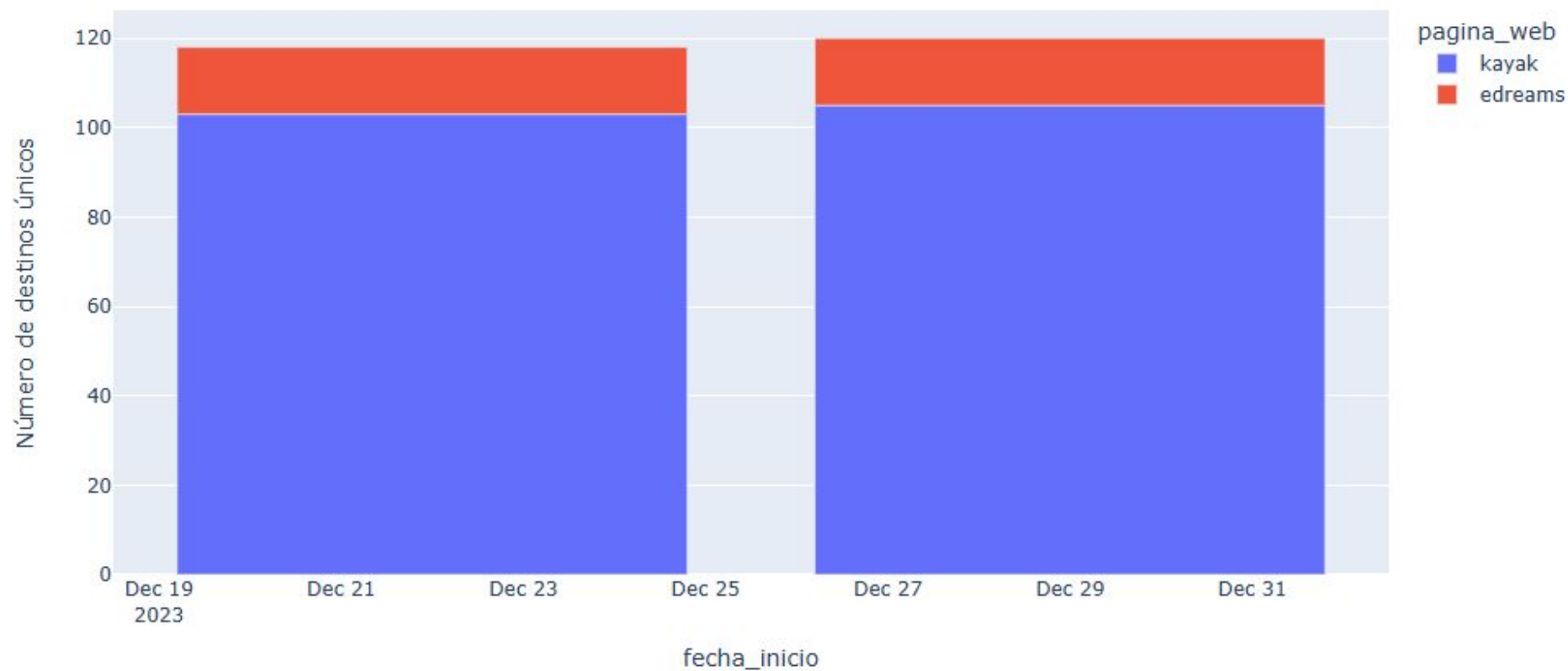


	pagina_web	fecha_inicio	destino
0	edreams	2023-12-22	15
1	edreams	2023-12-29	15
2	kayak	2023-12-22	103
3	kayak	2023-12-29	105

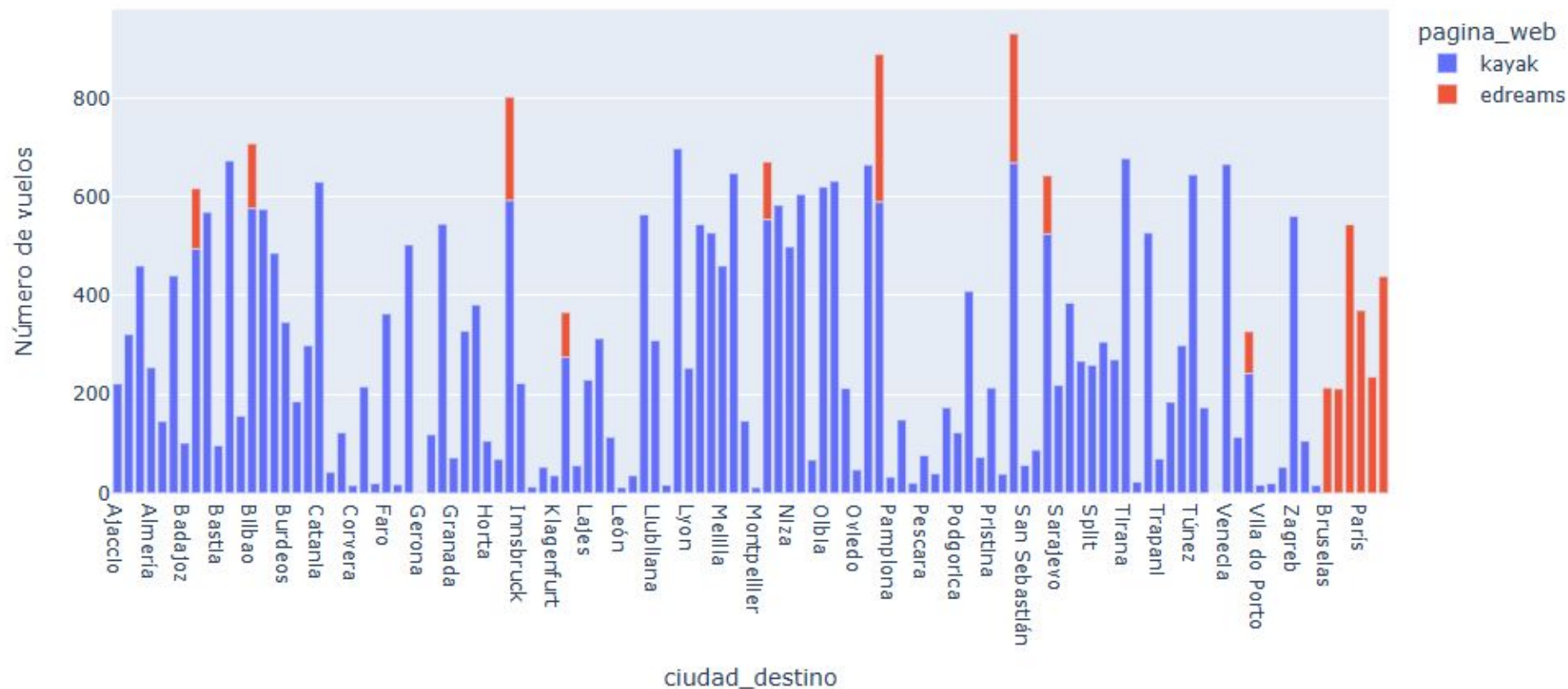
	pagina_web	ciudad_destino	num_vuelos
0	edreams	Barcelona	122
1	edreams	Bilbao	131
2	edreams	Bruselas	213
3	edreams	Gran Canaria	211
4	edreams	Ibiza	209
...
118	kayak	Vitoria	19
119	kayak	Zadar	52
120	kayak	Zagreb	561
121	kayak	Zakynthos	105
122	kayak	la Seu d'Urgell	15

123 rows × 3 columns

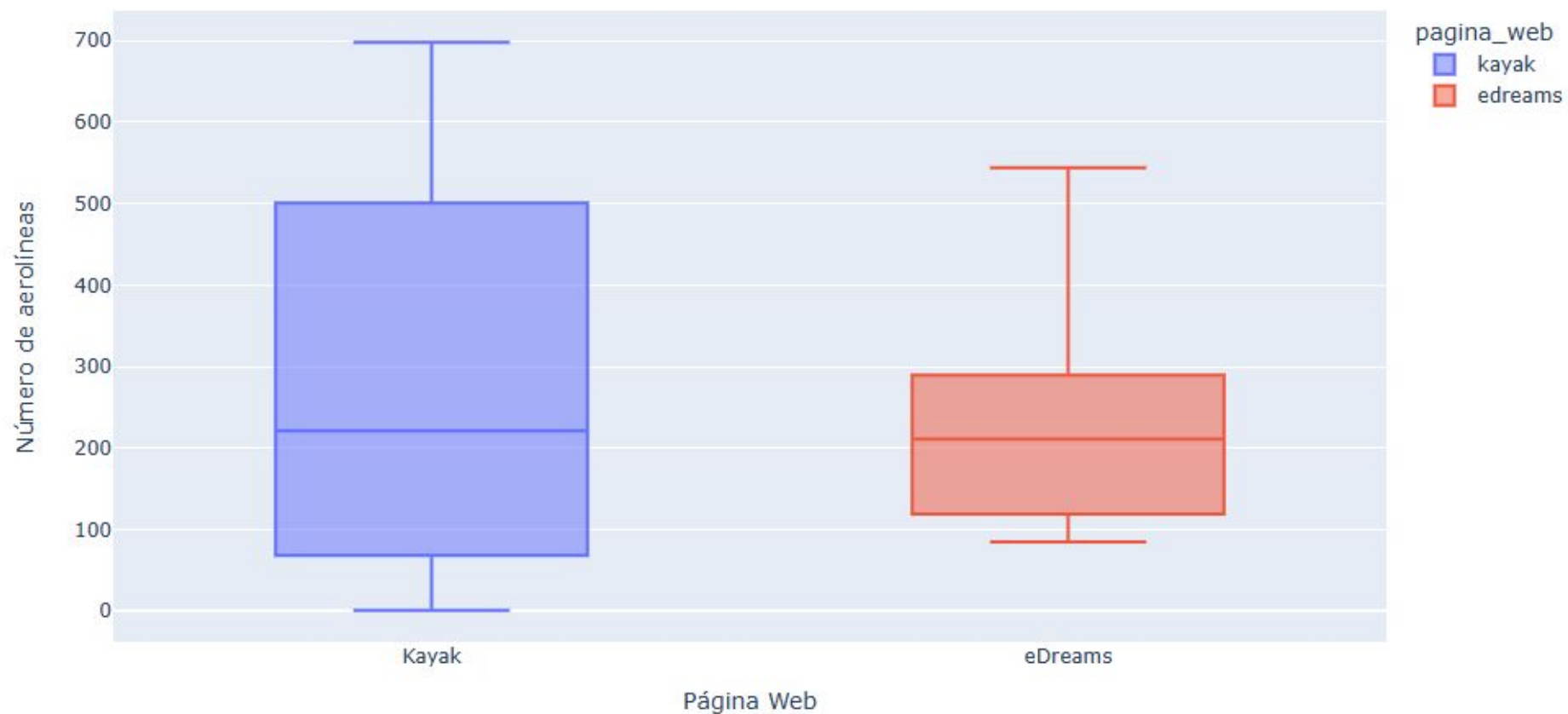
Número de destinos por fecha



Número de vuelos por destino



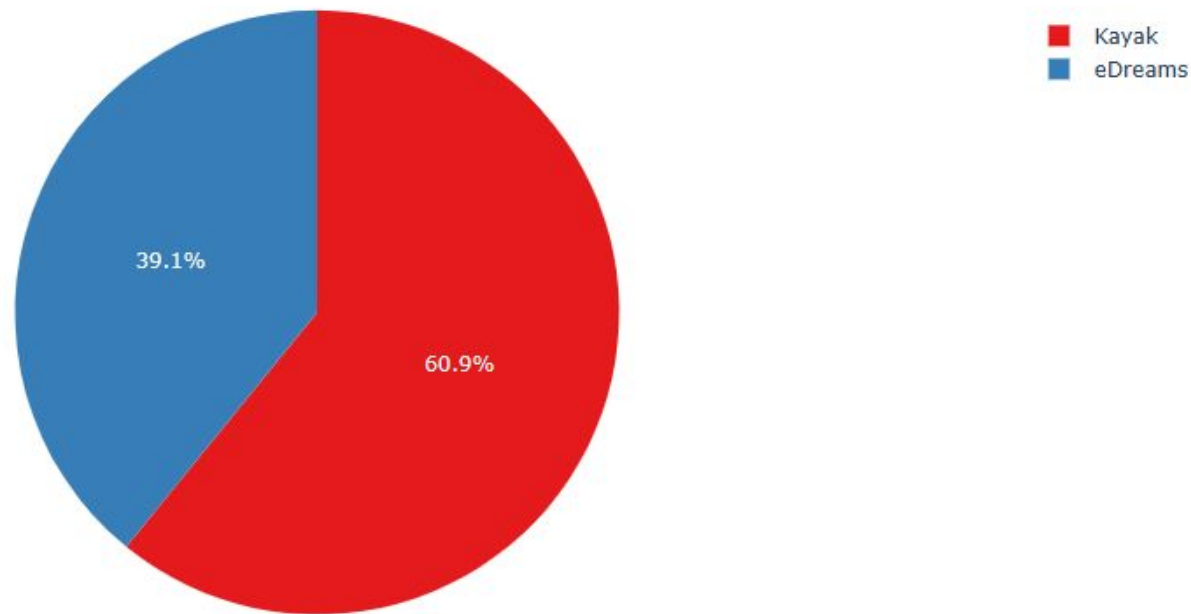
Número de aerolíneas por destino





De todos estos vuelos: ¿cuál es el porcentaje de vuelos que opera con más de una aerolínea?

Porcentaje de vuelos con más de una aerolínea

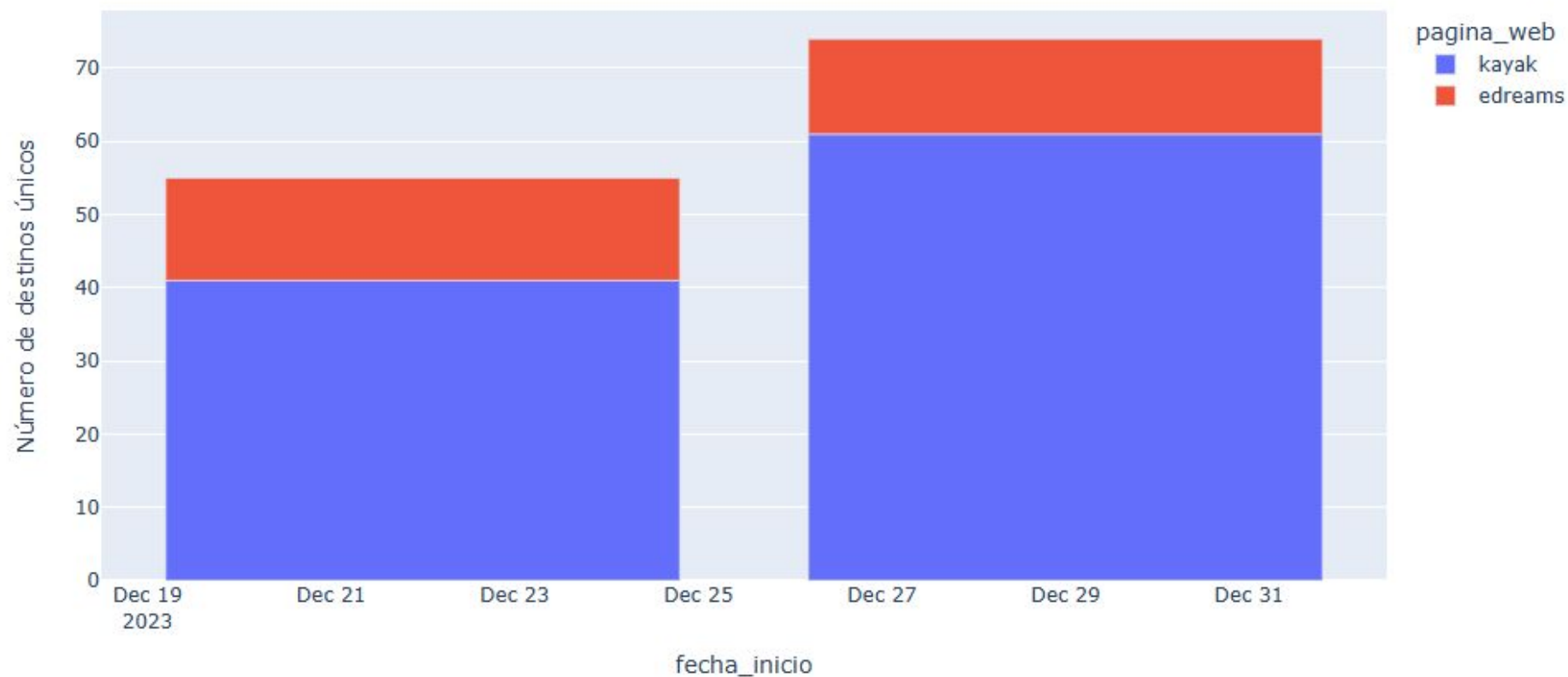


A lo mejor nos interesa que el vuelo sea gestionado por una sola aerolínea en caso de que haya retrasos o se deba pedir indemnizaciones. O bien, nos gusta el servicio de una aerolínea en concreto.

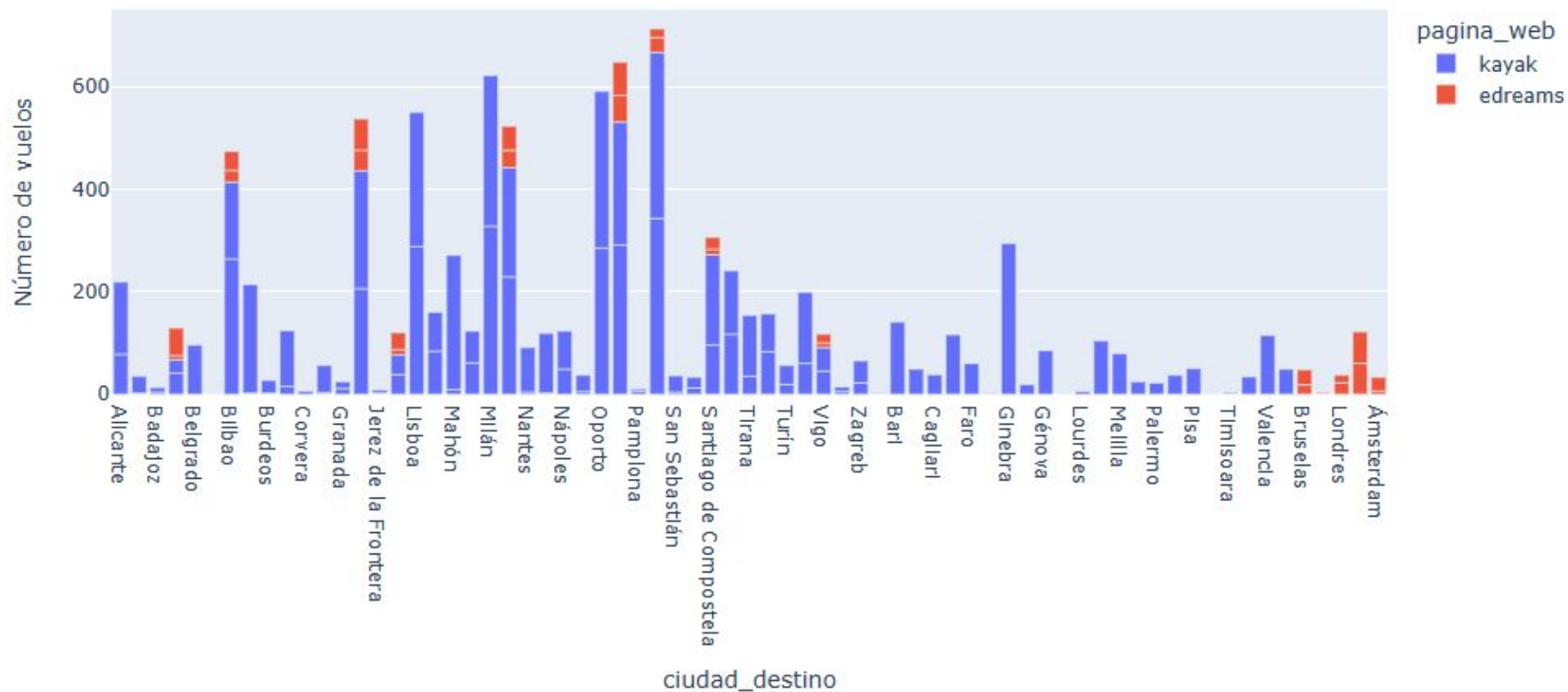


**¿Cuál de ellos me ofrecen destinos más
baratos?**

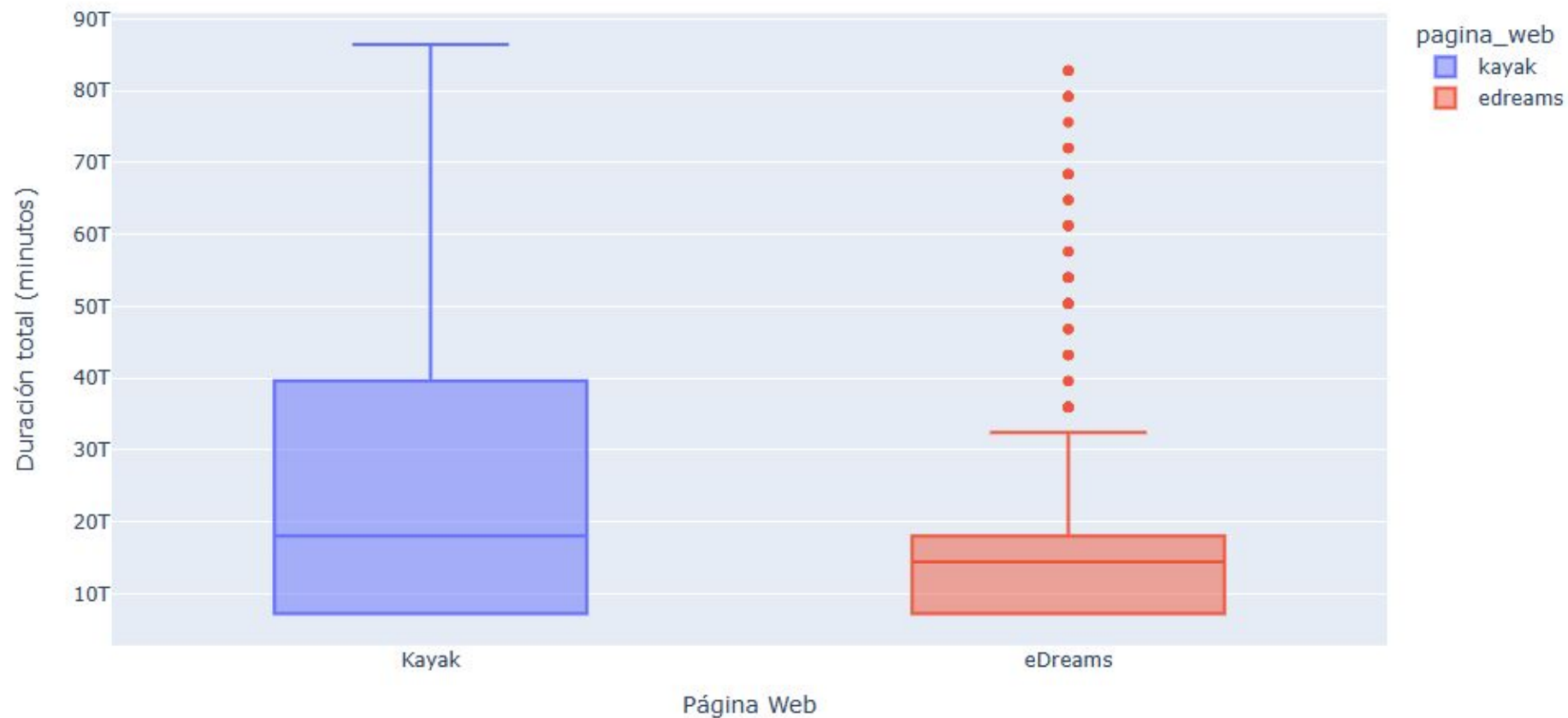
Número de destinos por fecha



Número de vuelos por destino



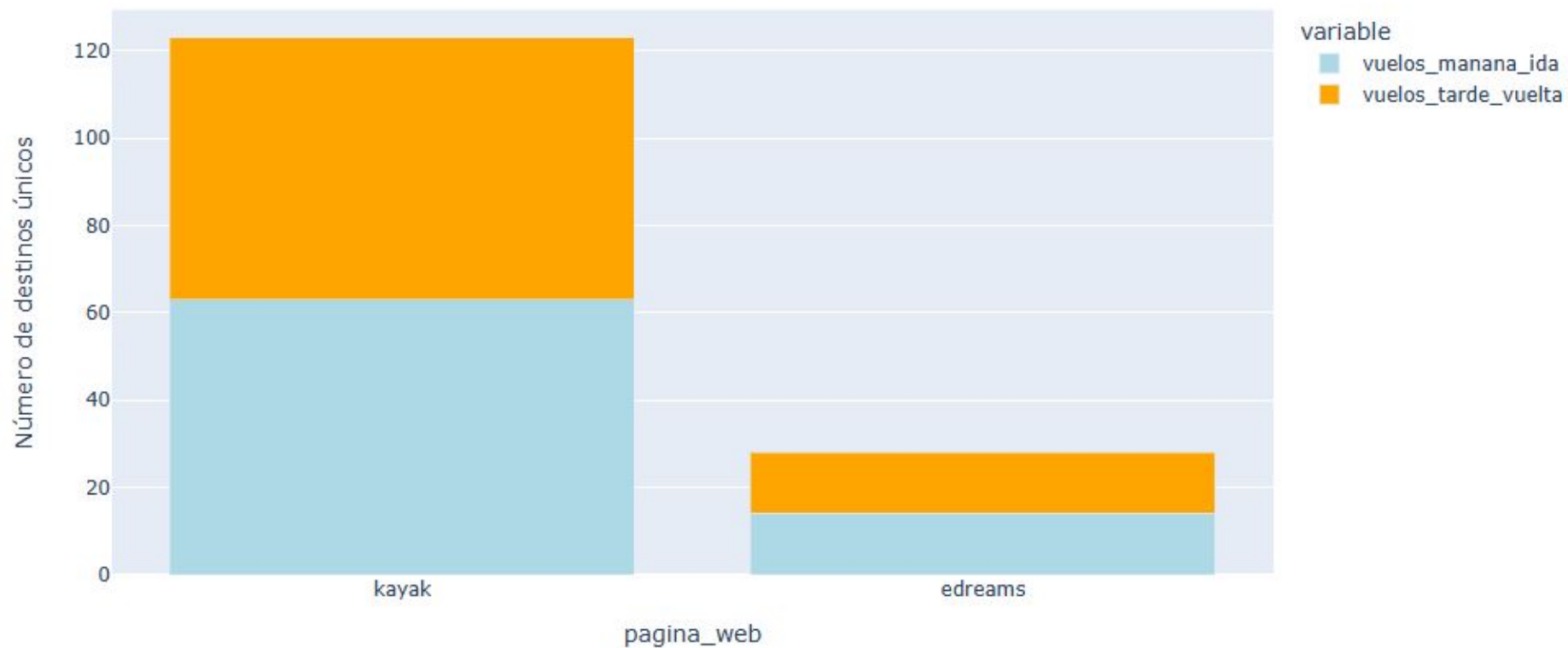
Distribución de la duración total de vuelos (Ida-Vuelta)





**Evaluaremos los horarios de ida y vuelta que
nos ofrece kayak y edreams**

Número de destinos únicos por categoría horaria





**Evaluaremos las condiciones de escalas,
equipaje, las aerolíneas que operan:**

Comparación de condiciones de vuelo: Kayak vs eDreams



Aerolíneas que operan en los vuelos: ['Wizz Air Malta' 'Iberia' 'Ryanair' 'ITA Airways' 'Vueling' 'Lufthansa' 'Iberia Express' 'Air Europa' 'Volotea' 'TAP AIR PORTUGAL' 'easyJet' 'Air Malta' 'Transavia France' 'easyJet Europe' 'Air Serbia' 'Wizz Air' 'KLM' 'Klm Royal Dutch Airlines' 'SWISS' 'Brussels Airlines' 'Aeroitalia' 'Air France' 'Pegasus Airlines' 'Easyjet' 'Turkish Airlines' 'British Airways' 'Austrian Airlines']



¿Qué página web utilizaremos para nuestro próximo viaje?