

K A Y A K®

What's the best destination
for your next **holidays?**



Kayak's challenge and expectations

Kayak's challenge

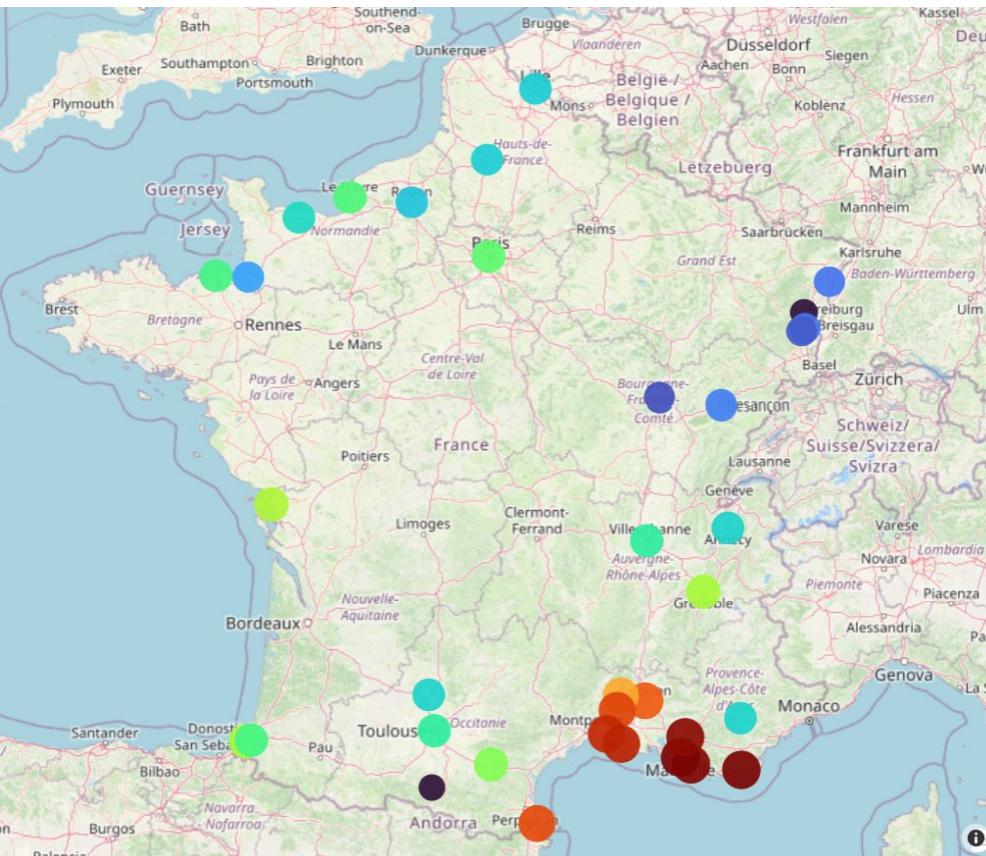
Travel information — weather, location and hotel quality — is **scattered across multiple sources**.

Kayak needs a unified dataset to **rank destinations** and **support travel recommendations**.

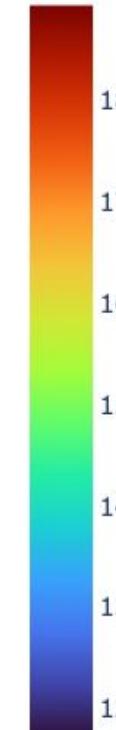
Project objective & steps

- ❖ **Collect data:**
 - Weather & coordinates via APIs
 - Hotel information via web scraping
- ❖ **Clean & merge:**
 - Create a unified table for all the destinations
- ❖ **Store in the cloud:**
 - S3 → data lake for raw files
 - Postgres RDS → database for structured tables analysis
- ❖ **Deliver:**
 - Datasets ready for destination ranking and recommendations

Average temperatures in France (for the next five days)



Temp (°C)



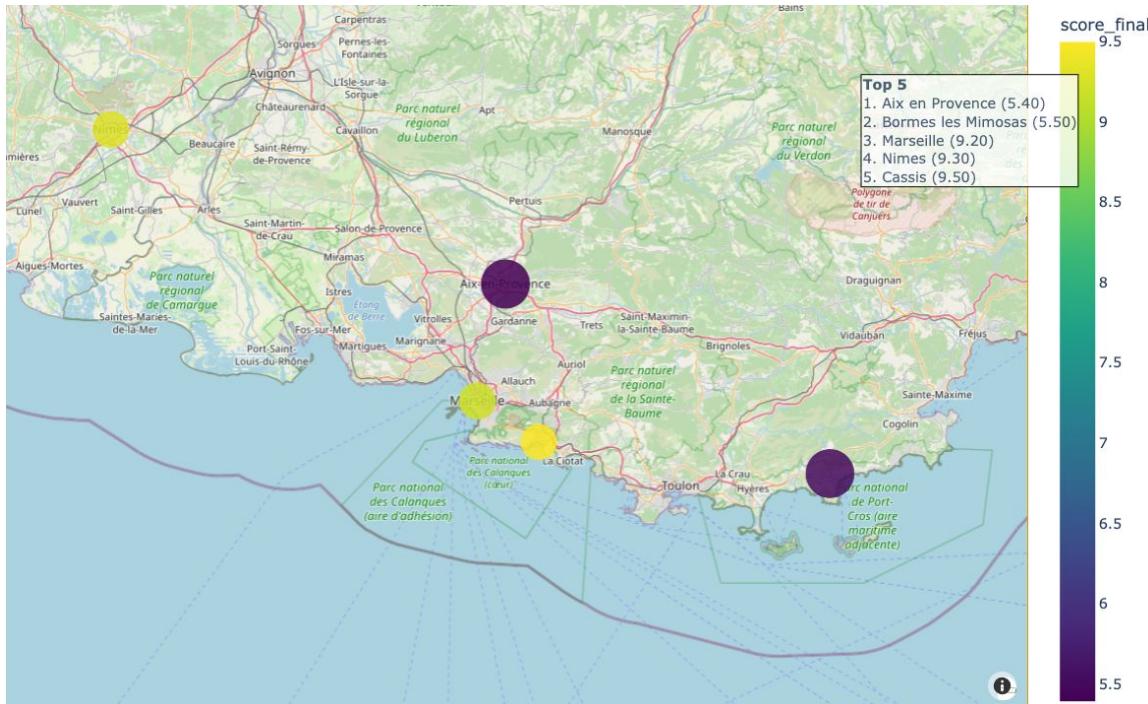
*Five-day temperature averages
computed from the OpenWeather
forecast API.*

Source: openweathermap.org

Top 5 Destinations in France Based on Weather Score

(for the next five days)

- Based on a weighted score combining temperature, wind, cloud cover and humidity.



The top five destinations are:

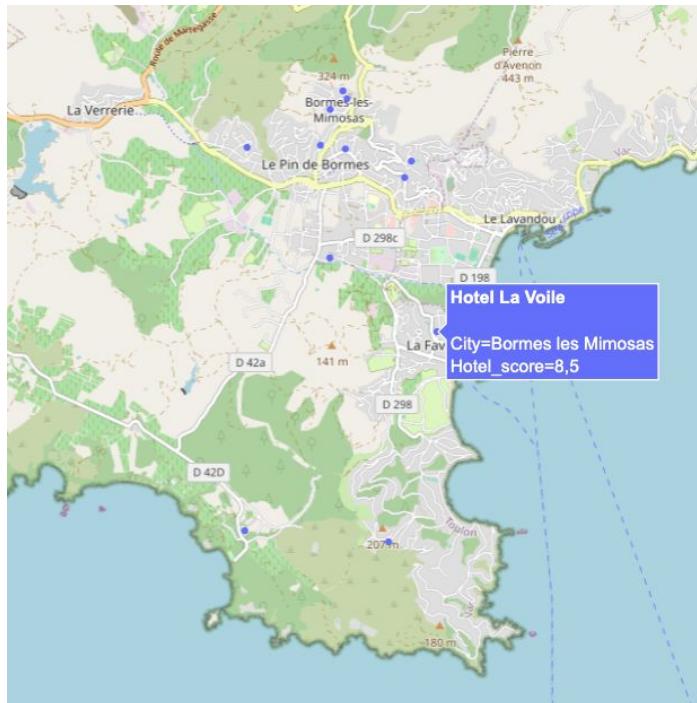
1. Aix en Provence
2. Bormes les Mimosas
3. Marseille
4. Nîmes
5. Cassis

The smaller score is the best

Hotels in the Top 5 Weather Destinations

Hotel data retrieved through web scraping on Booking

Two examples: *Bormes les Mimosas* and *Cassis*





Raw files stored in the S3 data lake (AWS)

Two CSV files are saved in their raw format:

- **Hotel locations and scores** (from the Booking scraper)
- **5-day weather forecast** (from the OpenWeather API)

Raw files are stored untouched so they can be reused or reprocessed later.

kayakbucket699475953240 [Infos](#)

[Objets](#) [Propriétés](#) [Autorisations](#) [Métriques](#) [Gestion](#) [Points d'accès](#)

Objets (2) [Copier l'URI S3](#) [Copier l'URL](#) [Télécharger](#) [Ouvrir](#) [Supprimer](#) [Actions ▾](#) [Créer un dossier](#) [Charger](#)

Les objets sont les entités fondamentales stockées dans Amazon S3. Vous pouvez utiliser l'[inventaire Amazon S3](#) pour obtenir une liste de tous les objets de votre compartiment. Pour que d'autres personnes puissent accéder à vos objets, vous devez leur accorder explicitement des autorisations. [En savoir plus](#)

Rechercher des objets en fonction du préfixe

<input type="checkbox"/>	Nom	Type	Dernière modification	Taille	Classe de stockage
<input type="checkbox"/>	Destinations_infos.csv	csv	19 Sep 2025 06:05:38 PM CEST	417.1 Ko	Standard
<input type="checkbox"/>	moyenne_meteo_5jours.csv	csv	19 Sep 2025 06:05:39 PM CEST	3.2 Ko	Standard



Creating the PostgreSQL database in AWS RDS

Aurora and RDS > Databases > kayakdatabase699475953240

Aurora and RDS

- Dashboard
- Databases**
- Query editor
- Performance insights
- Snapshots
- Exports in Amazon S3
- Automated backups
- Reserved instances
- Proxies

Subnet groups

Parameter groups

Option groups

Custom engine versions

Zero-ETL integrations [New](#)

Events

Event subscriptions

kayakdatabase699475953240

Summary

DB identifier kayakdatabase699475953 240	Status Available	Role Instance	Engine PostgreSQL
CPU  3.19%	Class db.t4g.micro	Current activity  0.00 sessions	Region & AZ eu-west-3b

Connectivity & security Monitoring Logs & events Configuration Zero-ETL integrati

Connectivity & security

Endpoint & port	Networking	Security
Endpoint  kayakdatabase699475953240.c 9csqOyye15e.eu-west-3.rds.amazonaws.com	Availability Zone eu-west-3b	VPC security groups default (sg-0f36d2a2b31029f5f) Active
Port 5432	VPC vpc-andreea-first-vpc (vpc-0f9c97b13e5a1b675)	Publicly accessible Yes



Cleaned tables available in PostgreSQL for analysts

Hotel information

pgAdmin 4

Object Explorer Databases Processes public.destinations/postgres/kayakpostgres@kayakpostgres

Query Query History Scratch Pad

```
1 SELECT * FROM public.destinations
```

Data Output Messages Notifications

Showing rows: 1 to 291 Page No: 1 of 1 14 < << > >> >>

City	HotelName
Nîmes	Logis NIMOTEL Hôtel Cosy
Aix en Provence	Hôtel des Augustins
Grenoble	Hôtel d'Angleterre Grenoble Hyper-Centre
Colmar	Au Grenier à Sel Colmar
Lyon	High-end apartment overlooking the quays AIL
Bormes les Mimosas	Maison Gau Bénat Piscine Vue mer
Strasbourg	Hôtel de l'Europe by HappyCulture
Montauban	"Le cocon de Sapiac" à deux pas du stade, garage
Marseille	BLUE BAY - superbe vue sur le Vieux Port
Cassis	Giovanna N3
Bormes les Mimosas	Appartement climatisé Bord de Mer La Faviere
Chateau du Haut Koenigsbourg	Gites & Camping on the Route des Vins
Chateau du Haut Koenigsbourg	L'élos des vignes - Appartements haut de gamme avec Spa
Chateau du Haut Koenigsbourg	Gite Binnert Bader
Chateau du Haut Koenigsbourg	Hôtel Vall'Vignes Colmar Haut Koenigsbourg Thon Crustacé Balaï

Total rows: 291 Query complete 00:00:226 LF Ln 1, Col 1

Weather information

pgAdmin 4

Object Explorer Databases Processes public.meteo/postgres/kayakpostgres@kayakpostgres

Query Query History Scratch Pad

```
1 SELECT * FROM public.meteo
```

Data Output Messages Notifications

Showing rows: 1 to 35 Page No: 1 of 1 14 < << > >> >>

Ville	Latitude	Longitude	Temperature Moy (°C)
Mont Saint Michel	48.6359541	-1.51146	16.9
St Malo	48.649518	-2.0260409	16.9
Bayeux	49.2764624	-0.7024738	17.3
Le Havre	49.4938975	0.1079732	17.5
Rouen	49.4404591	1.0939658	17.3
Paris	48.8534951	2.3483915	18.8
Amiens	49.8941708	2.2956951	17.1
Lille	50.6365654	3.0635282	17.6
Strasbourg	48.584614	7.507127	18.3
Chateau du Haut Koenigsbourg	48.2494107	7.3443202	16
Colmar	48.0777517	7.3579641	18.6
Eguisheim	48.0447968	7.3079618	18.4
Besançon	47.2380222	6.0249622	17.4
Dijon	47.3215806	5.0414701	17
Angoulême	46.9001240	1.17000247	10.4

Total rows: 35 Query complete 00:00:118 LF Ln 1, Col 1

Final Merged Table

	column_name
name	
1	ville
2	temperature_moyenne
3	humidite_moyenne
4	vent_moyen
5	couverture_nuageuse
6	conditions_meteo
7	latitude_ville
8	longitude_ville
9	hotel
10	score
11	url
12	description
13	latitude_hotel
14	longitude_hotel

Object Explorer public.meteo/post... public.destination... public.meteo/post...

Servers (2)
 kayakpostgres
 Databases (2)
 postgres
 Casts
 Catalogs
 Event Triggers
 Extensions
 Foreign Data Wrappers
 Languages
 Publications
 Schemas (1)
 public
 Aggregates
 Collations
 Domains
 FTS Configurations
 FTS Dictionaries
 FTS Parsers
 FTS Templates
 Foreign Tables
 Functions
 Materialized Views
 Operators
 Procedures
 Sequences
 Tables (3)
 destinations
 kayak_final
 meteo
 Trigger Functions
 Types
 Views
 Subscriptions
 rdsadmin
 Login/Group Roles
 Tablespaces
server_andreea_demo

public.kayak_final/postgres/kayakpostgres@kayakpostgres

No limit

Query Query History

```
1  SELECT * FROM public.kayak_final
2
```

Data Output Messages Notifications

	ville text	temperature_moyenne double precision	humidite_moyenne double precision	vent_moyenne double precision
1	Aigues Mortes	18.2	67.4	
2	Aigues Mortes	18.2	67.4	
3	Aigues Mortes	18.2	67.4	
4	Aigues Mortes	18.2	67.4	
5	Aigues Mortes	18.2	67.4	
6	Aigues Mortes	18.2	67.4	
7	Aigues Mortes	18.2	67.4	
8	Aigues Mortes	18.2	67.4	
9	Aigues Mortes	18.2	67.4	
10	Aigues Mortes	18.2	67.4	
11	Aigues Mortes	18.2	67.4	
12	Aigues Mortes	18.2	67.4	
13	Aigues Mortes	18.2	67.4	
14	Ariege	9.3	86.1	
15	Ariege	9.3	86.1	
16	Ariege	9.3	86.1	
17	Ariege	9.3	86.1	
18	Ariege	9.3	86.1	
19	Ariege	9.3	86.1	
20	Ariege	9.3	86.1	
21	Ariege	9.3	86.1	
22	Ariege	9.3	86.1	
23	Ariege	9.3	86.1	
24	Ariege	9.3	86.1	
25	Ariege	9.3	86.1	
26	Avignon	17.5	66.6	



Thank you for your attention
– any questions?

