



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



# Kayak's challenge and expectations

## Kayak's challenge

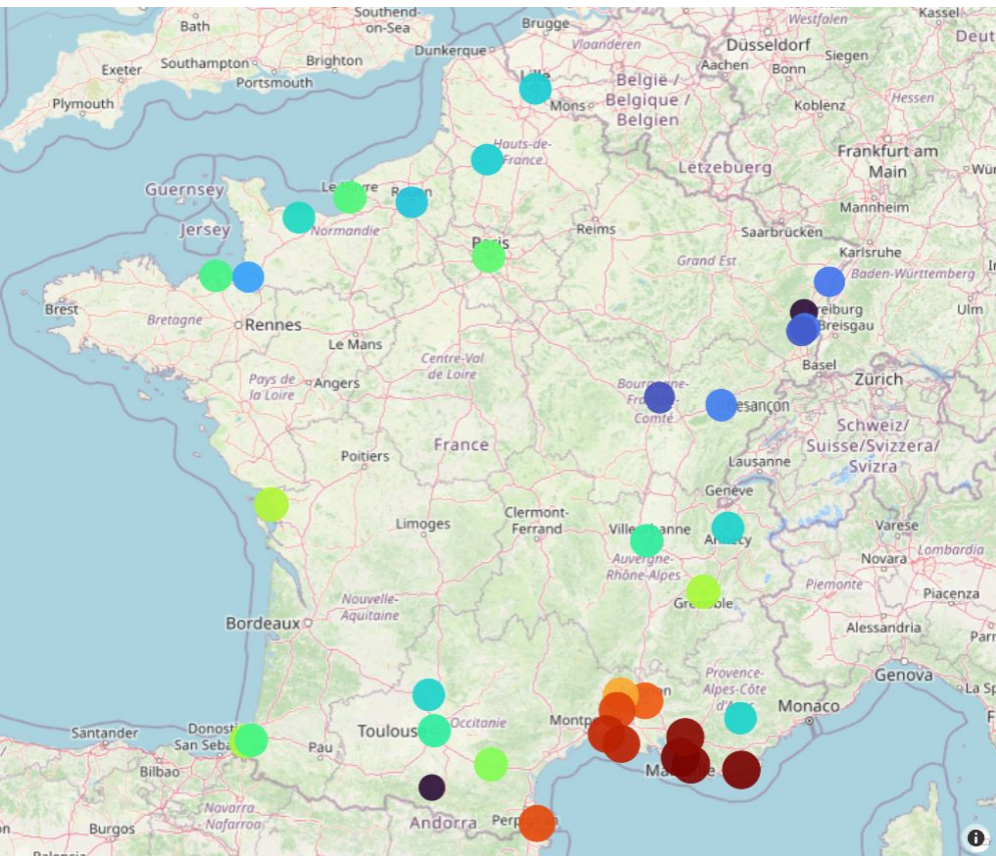
Travel information — weather, location and hotel quality — is **scattered across multiple sources**.

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

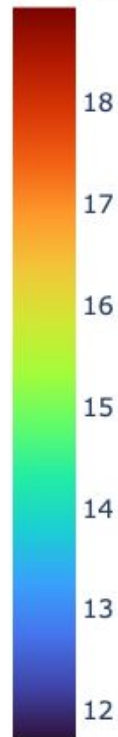
## Project objective & steps

- ❖ **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)



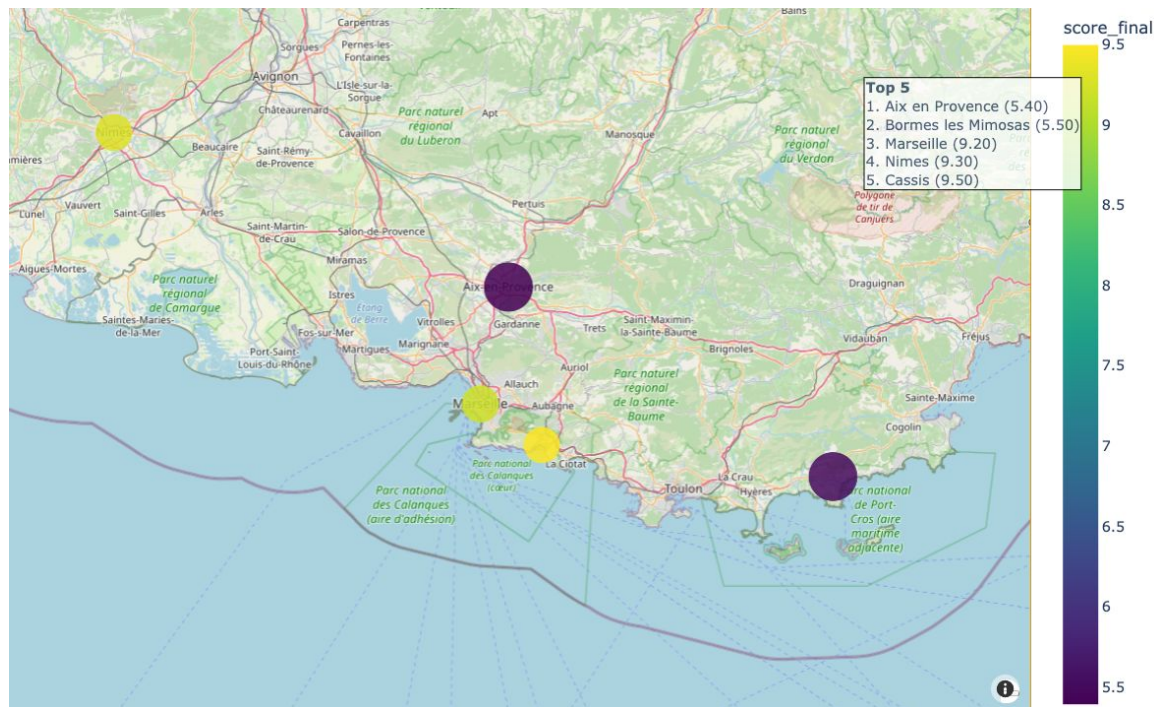
*Five-day temperature averages  
computed from the OpenWeather  
forecast API.*

Source: [openweathermap.org](https://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:

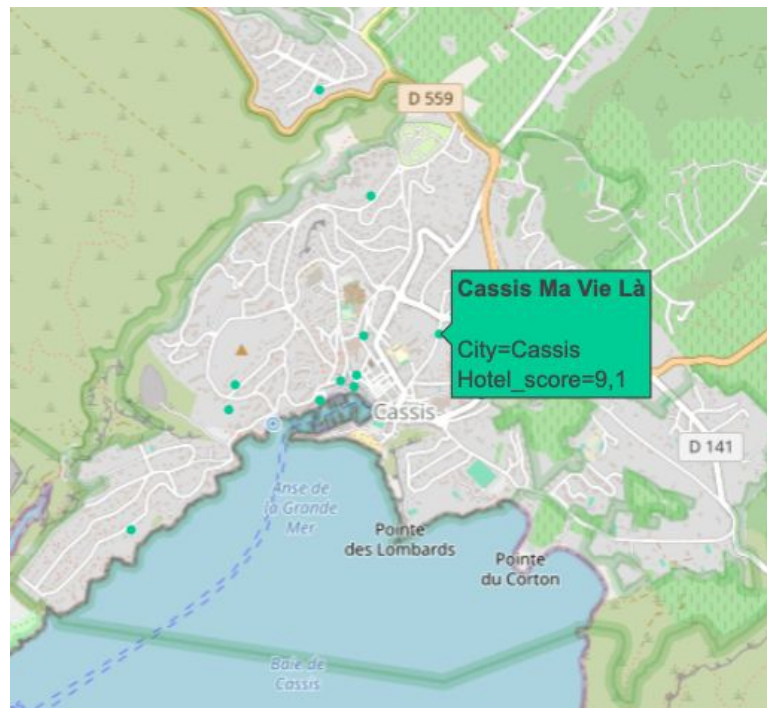
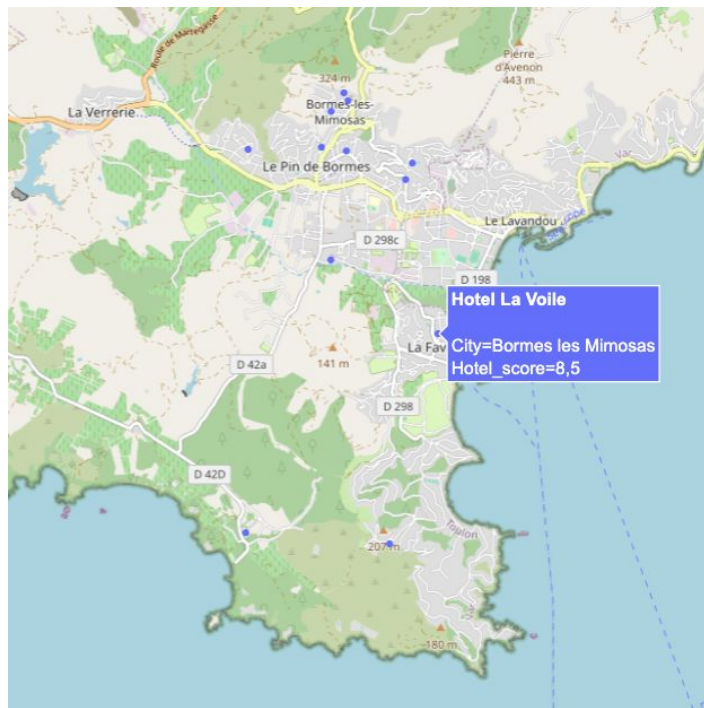
1. Aix en Provence
2. Bormes les Mimosas
3. Marseille
4. Nîmes
5. Cassis

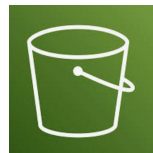
*The smaller score is the best*

# 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

< 1 >

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



# Creating the PostgreSQL database in AWS RDS

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

Events

Event subscriptions

kayakdatabase699475953240

Summary

DB identifier  
kayakdatabase699475953240

CPU  
3.19%

Status  
Available

Class  
db.t4g.micro

Role  
Instance

Current activity  
0.00 sessions

Engine  
PostgreSQL

Region & AZ  
eu-west-3b

Connectivity & security

Monitoring

Logs & events

Configuration

Zero-ETL integrati

Connectivity & security

Endpoint & port

Endpoint  
kayakdatabase699475953240.c9csq0yye15e.eu-west-3.rds.amazonaws.com

Port  
5432

Networking

Availability Zone  
eu-west-3b

VPC  
vpc-andreea-first-vpc (vpc-0f9c97b13e5a1b675)

Security

VPC security groups  
default (sg-0f36d2a2b31029f5f)  
Active

Publicly accessible  
Yes

# Cleaned tables available in PostgreSQL for analysts



## Hotel information

pgAdmin 4

Object Explorer

- postgres
  - Casts
  - Catalogs (2)
  - Event Triggers
  - Extensions (1)
  - Foreign Data Wrappers
  - Languages (1)
  - Publications
  - Schemas (1)
    - public
      - Aggregates
      - Collations
      - Domains
      - FTS Configurations
      - FTS Dictionaries
      - FTS Parsers
      - FTS Templates
      - Foreign Tables
      - Functions
      - Materialized Views
      - Operators
      - Procedures
      - Sequences
      - Tables (2)
        - destinations
        - meteo
      - Trigger Functions
      - Types
      - Views
      - Subscriptions
      - rsdadmin
      - Login/Group Roles
      - Tablespaces
      - server\_andrea\_demo

Query

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

Data Output

	City text	Hotel_Name text
1	Nîmes	Logis NIMOTEL Hôtel Cosy
2	Aix en Provence	Hôtel des Augustins
3	Grenoble	Hôtel d'Angleterre Grenoble Hyper-Centre
4	Colmar	Au Grenier à Sel Colmar
5	Lyon	High-end apartment overlooking the quays AIL
6	Bormes les Mimosas	Maison Gaou Bénat Piscine Vue mer
7	Strasbourg	Hôtel de l'Europe par HappyCulture
8	Montauban	"Le cocon de Sapiac" à deux pas du stade, garage
9	Marseille	BLUE BAY - superbe vue sur le Vieux Port
10	Cassis	Giovannina N3
11	Bormes les Mimosas	Appartement climatisé Bord de Mer La Favière
12	Chateau du Haut Koenigsbourg	Gîtes & Camping on the Route des Vins
13	Chateau du Haut Koenigsbourg	Éclôses des vignes - Appartements haut de gamme avec Spa
14	Chateau du Haut Koenigsbourg	Gîte Binnert Bader

Total rows: 291 Query complete 00:00:00.226 LF Ln 1, Col

## Weather information

pgAdmin 4

Object Explorer

- postgres
  - Casts
  - Catalogs (2)
  - Event Triggers
  - Extensions (1)
  - Foreign Data Wrappers
  - Languages (1)
  - Publications
  - Schemas (1)
    - public
      - Aggregates
      - Collations
      - Domains
      - FTS Configurations
      - FTS Dictionaries
      - FTS Parsers
      - FTS Templates
      - Foreign Tables
      - Functions
      - Materialized Views
      - Operators
      - Procedures
      - Sequences
      - Tables (2)
        - destinations
        - meteo
      - Trigger Functions
      - Types
      - Views
      - Subscriptions
      - rsdadmin
      - Login/Group Roles
      - Tablespaces
      - server\_andrea\_demo

Query

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

Data Output

	Ville text	Latitude double precision	Longitude double precision	Temperature Moy (°C) double precision
1	1 Mont Saint Michel	48.6359541	-1.51146	16.9
2	2 St Malo	48.649518	-2.0260409	16.9
3	3 Bayeux	49.2764624	-0.7024738	17.3
4	4 Le Havre	49.4938975	0.1079732	17.5
5	5 Rouen	49.4404591	1.0939658	17.3
6	6 Paris	48.8534951	2.3483915	18.8
7	7 Amiens	49.8941708	2.256951	17.1
8	8 Lille	50.6365654	3.0635282	17.6
9	9 Strasbourg	48.584614	7.7507127	18.3
10	10 Chateau du Haut Koenigsbourg	48.2494107	7.3443202	16
11	11 Colmar	48.0777517	7.3579641	18.6
12	12 Eguisheim	48.0447968	7.3079618	18.4
13	13 Besancon	47.2380222	6.0243622	17.4
14	14 Dijon	47.3215806	5.0414701	17

Total rows: 35 Query complete 00:00:00.118 LF Ln 1, Col 1

# Final Merged Table

## Tables (3)

- > destinations
- > kayak\_final
- > meteo

	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

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
      - rdadmin
      - Login/Group Roles
      - Tablespaces
      - server\_andreea\_demo

public.kayak\_final/postgres/kayakpostgres@kayakpostgres

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_moyen 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?

