KAYAK®

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

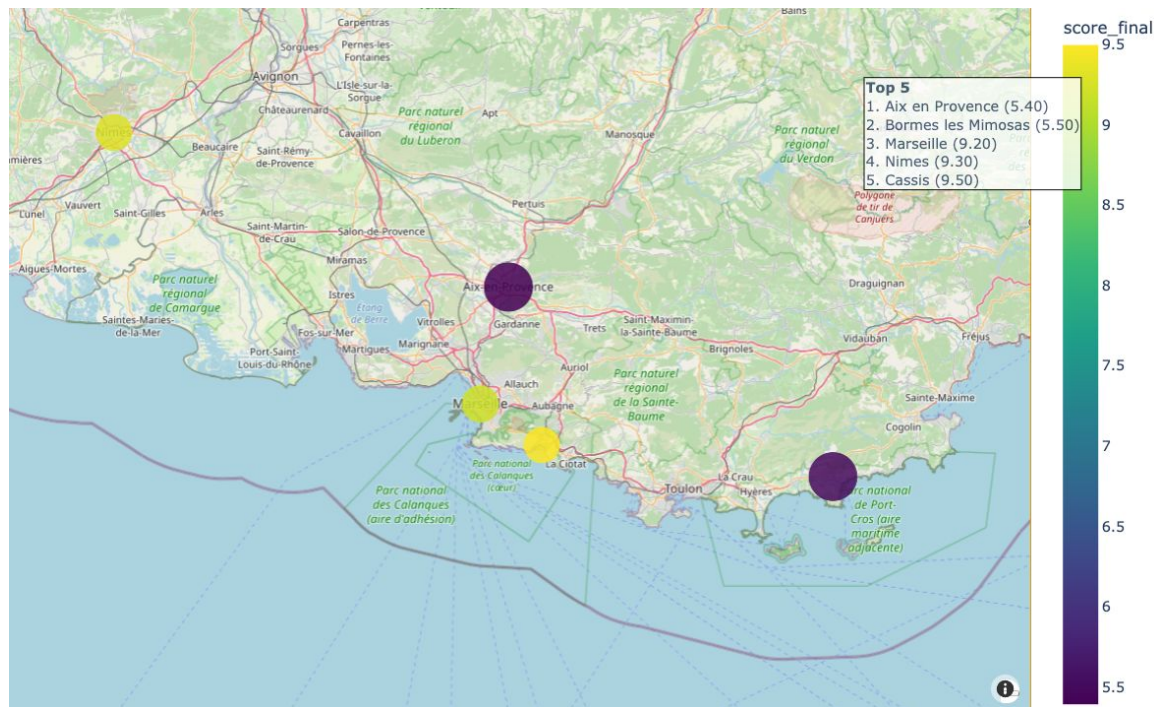# Average temperatures in France (for the next five days)
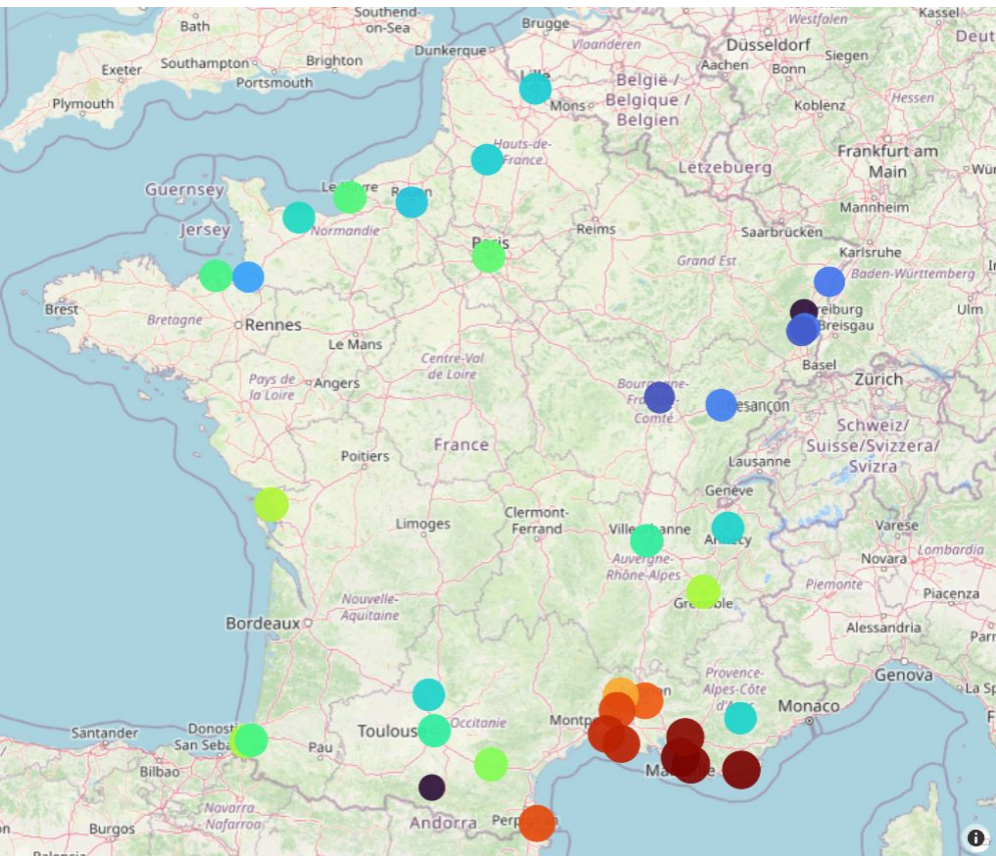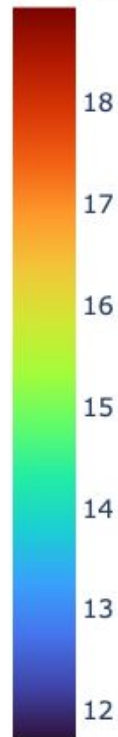


Five-day temperature averages computed from the OpenWeather forecast API.

Source: openweathermap.org

# Top 5 Destinations in France Based on Weather Score

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

**Top 5**
1. Aix en Provence (5.40)
2. Bormes les Mimosas (5.50)
3. Marseille (9.20)
4. Nîmes (9.30)
5. Cassis (9.50)

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*
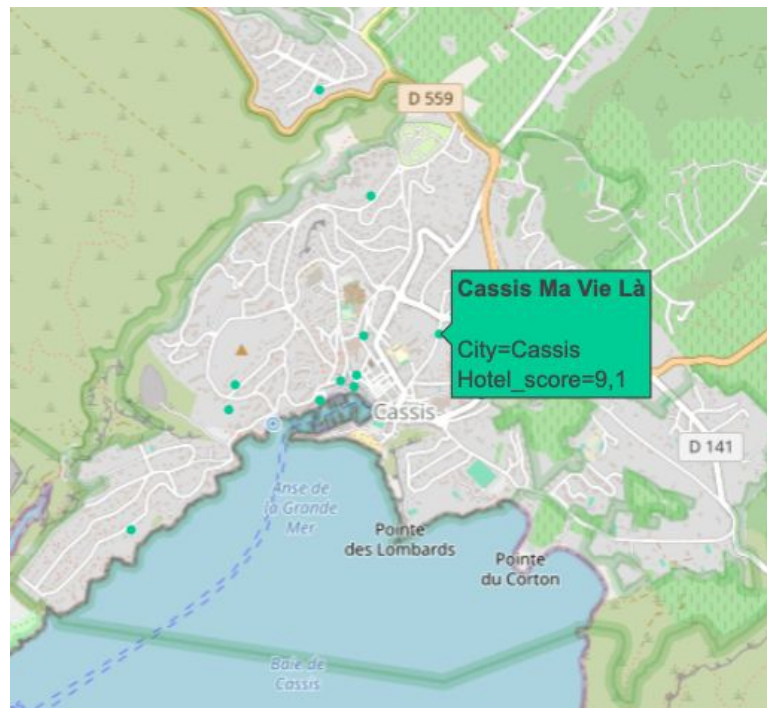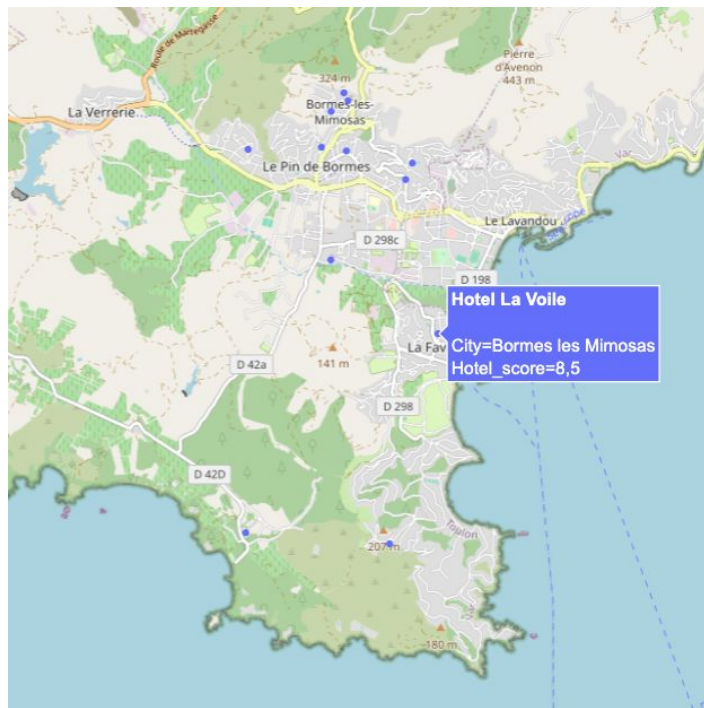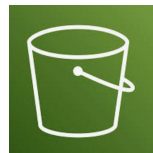
# Raw files stored in the S3 data lake (AWS)

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.

# Creating the PostgreSQL database in AWS RDS

# Cleaned tables available in PostgreSQL for analysts

<u>Hotel information</u>

<u>Weather information</u>

# Final Merged Table

Thank you for your attention
– any questions?