

Documentation du Projet Météo-Tourisme

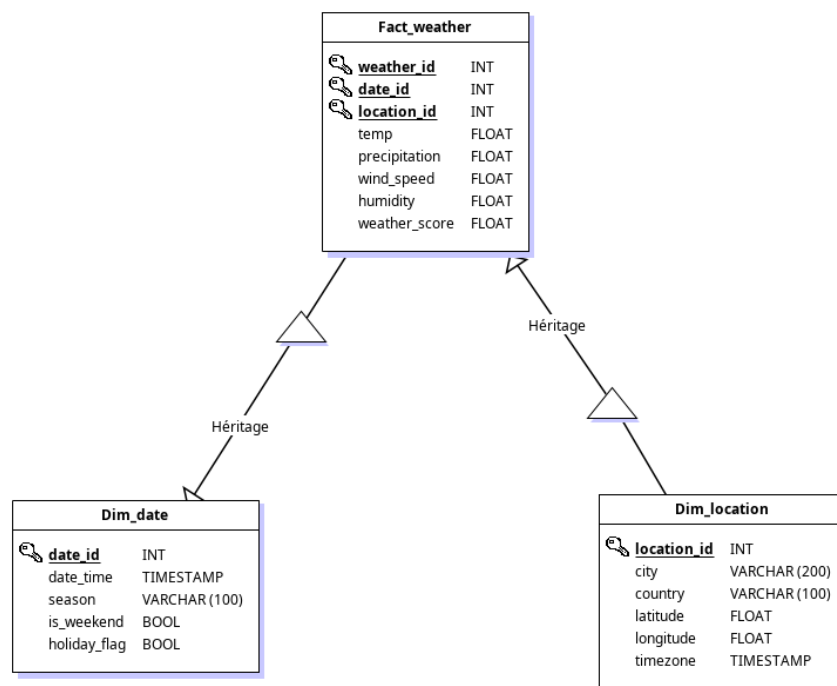
Contexte du Projet

Ce projet vise à recommander les meilleures périodes pour visiter différentes villes en fonction des conditions météorologiques. Nous avons développé un pipeline ETL automatisé qui collecte, traite et analyse les données météorologiques historiques et en temps réel pour calculer des scores de confort météo.

Architecture du Projet

1. Modèle Conceptuel de Données (MCD)

Le MCD du projet:



Explications du modèle en étoile:

1. FACT_WEATHER: Table des faits centrale contenant toutes les mesures météorologiques

- Contient les indicateurs clés (température, précipitations, etc.)
- Score météo calculé (0-1) pour évaluer les conditions
- Clés étrangères vers les dimensions

2. DIM_DATE: Dimension temporelle

- Permet l'analyse par jour/mois/année/saison
- Flags pour weekends/jours fériés
- Optimisé pour le filtrage temporel

3. DIM_LOCATION: Dimension géographique

- Information sur les villes suivies
- Coordonnées géographiques pour cartographie
- Fuseau horaire pour conversions temporelles

2. Pipeline de Données

Composants principaux:

a. Extraction:

- ``historical_data.py``: Génère 7 ans de données historiques simulées
- ``openweather_api.py``: Récupère les données actuelles via l'API OpenWeatherMap

b. Transformation:

- ``clean_data.py``: Nettoie et fusionne les données historiques et actuelles
- ``calculate_scores.py``: Calcule les scores météo basés sur température, précipitation et vent

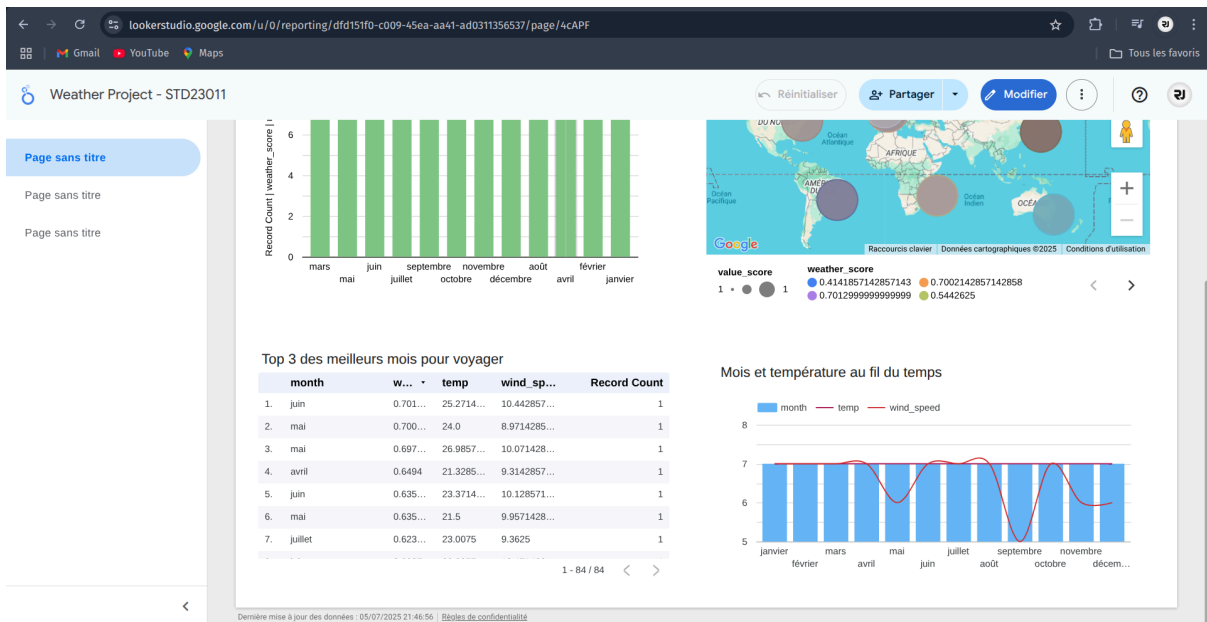
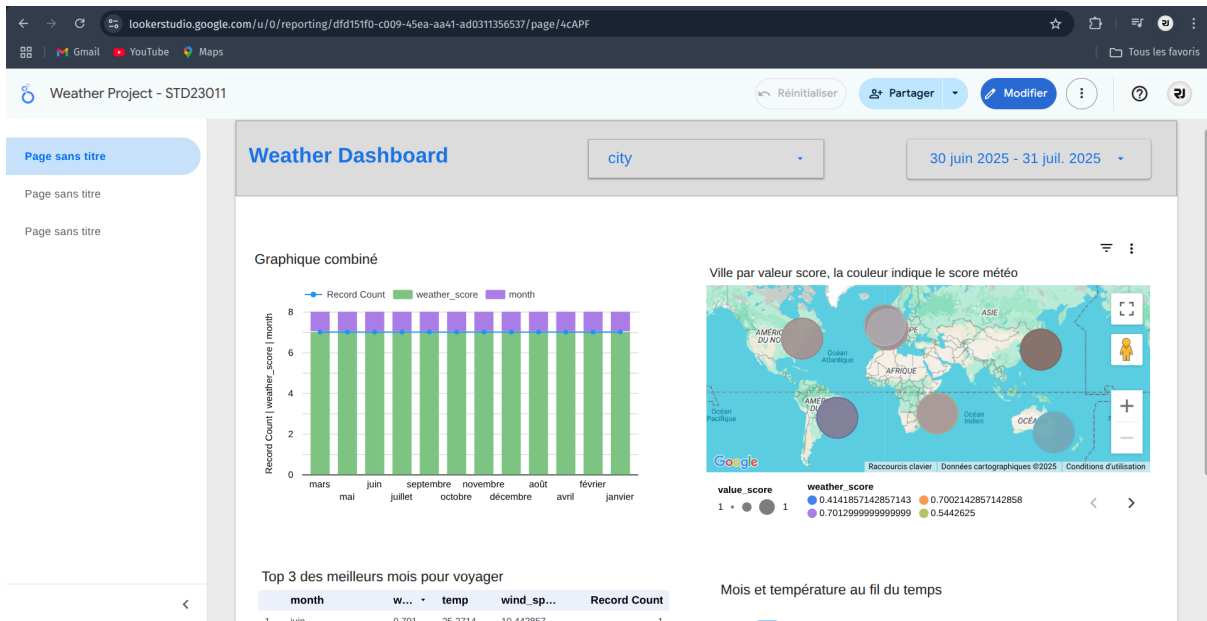
c. Chargement:

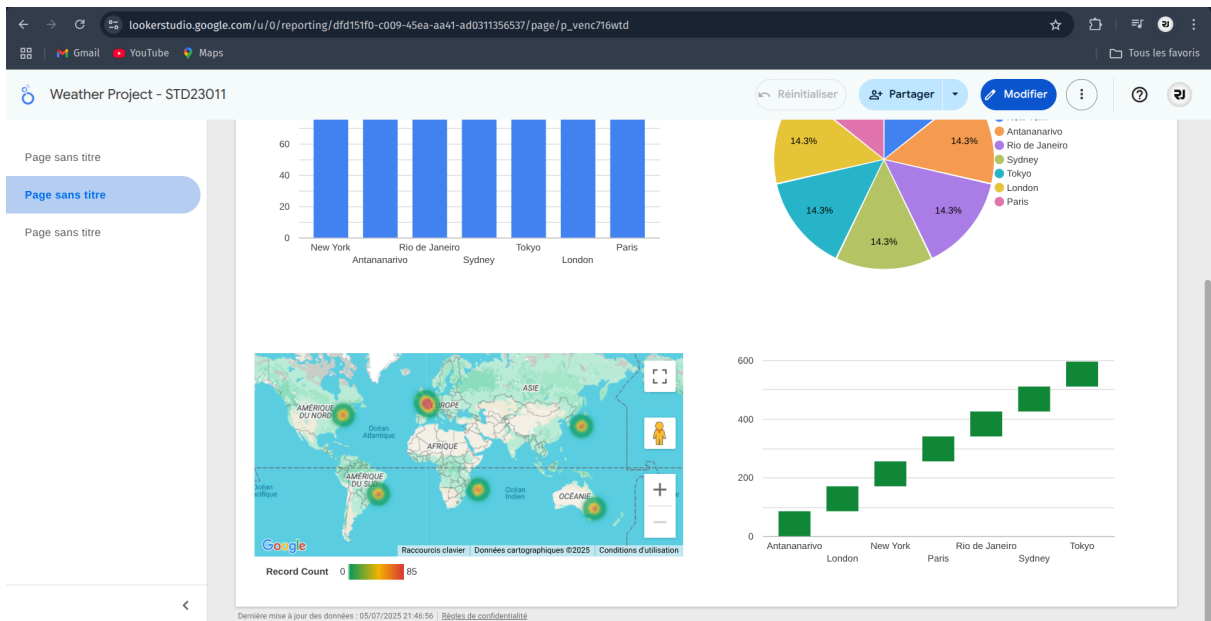
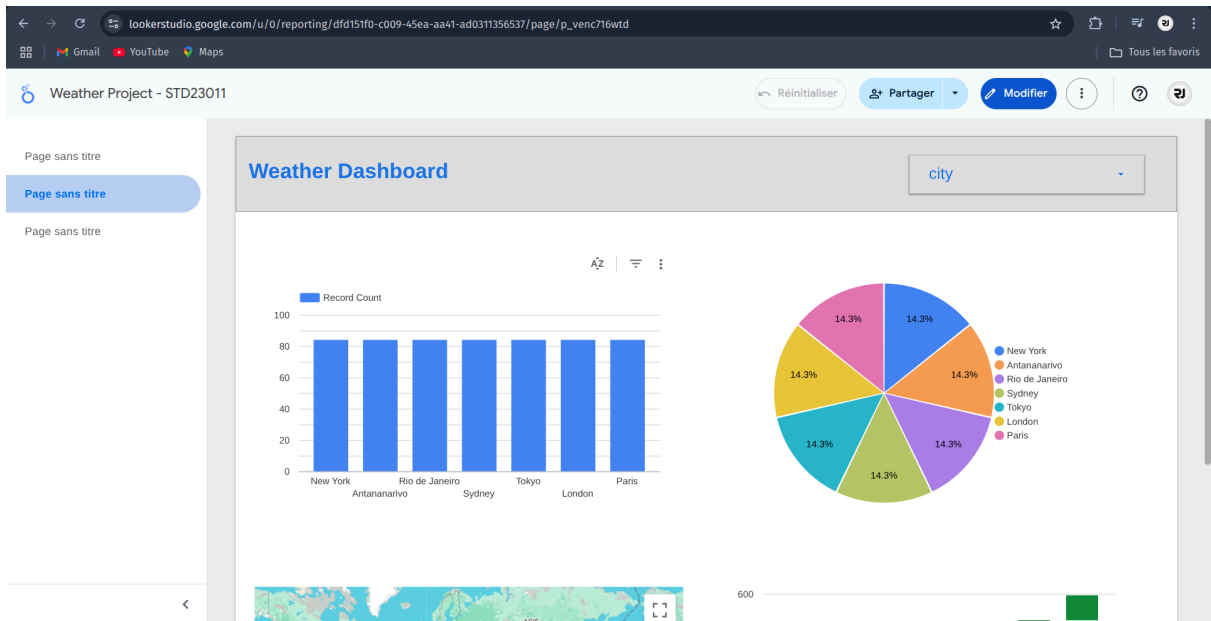
- ``load_to_db.py``: Charge les données dans Google Sheets
- ``update_sheets.py``: Met à jour plusieurs feuilles Google Sheets

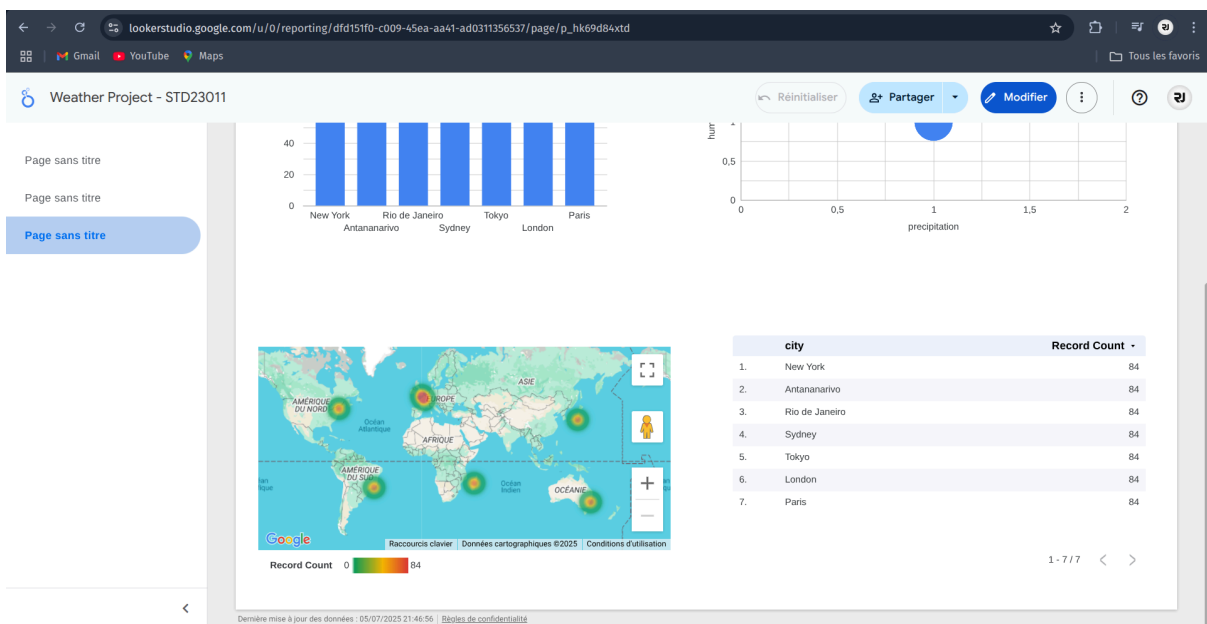
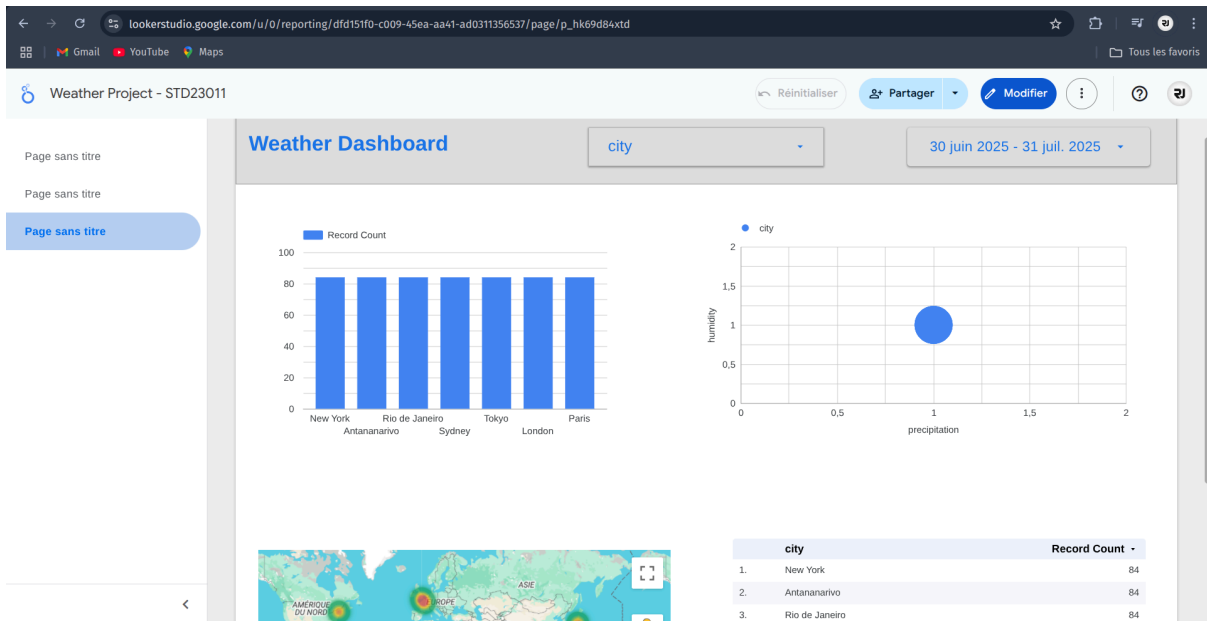
d. Orchestration:

- `meteo_tourism_dags.py`: DAG Airflow qui orchestre l'ensemble du processus

3. Dashboard (Looker Studio)







Le dashboard permet de:

- Visualiser les scores météo par ville et par mois
- Comparer les conditions actuelles avec les moyennes historiques
- Identifier les meilleurs mois pour visiter chaque ville

4. Analyse Exploratoire des Données (EDA)

Principales observations:

a. Variations saisonnières:

- Paris et Londres ont des hivers froids (moyenne 5°C) et des étés modérés (~18°C)
- Rio et Antananarivo ont des températures plus stables toute l'année (~20-25°C)

b. Meilleurs mois pour voyage:

- Mai à Septembre pour Paris (scores 0.3-0.4)
- Novembre à Avril pour Sydney (scores 0.6-0.7)

c. Corrélations:

- Forte corrélation négative entre précipitation et score météo (-0.82)
- Température idéale autour de 25°C (corrélation en cloche)

Exécution du Pipeline

a. Déclenchement quotidien : Via l'outils Apache Airflow

b. Séquence des tâches:

```
...  
[extract_historical, extract_current] >> clean_data >> calculate_scores >>  
[load_data, update_sheets]  
...
```

c. Sorties:

- Fichiers CSV intermédiaires
- Mises à jour des Google Sheets
- Dashboard actualisé

Conclusion

Ce pipeline fournit une base solide pour des recommandations de voyage basées sur des données météorologiques. Le modèle en étoile permet des analyses multidimensionnelles efficaces, tandis que le dashboard offre une interface intuitive pour les utilisateurs finaux.