

# Rapport - Analyse des données météorologiques de quatre villes à l'aide d'un Système d'Information Décisionnel

SAE Intégration de données dans un DataWarehouse

Adèle DALLE

Abinaya SUBRAMANIAM

John-Etan UZAN

Célia ZUK

BUT SD 2A EMS FI

Projet encadré par :

**M. Jean TSHIBANGU**

## Table des matières

<b>CONTEXTUALISATION .....</b>	<b>3</b>
PROBLEMATIQUE.....	3
OBJECTIFS.....	3
<b>MODELISATION DU SID .....</b>	<b>4</b>
2.1. CHOIX DU PROCESSUS METIER A MODELISER .....	4
2.2. DEFINITION DE LA GRANULARITE .....	4
2.3. CHOIX DES DIMENSIONS.....	4
2.4. IDENTIFICATION DES FAITS (MESURES) .....	5
<b>INTEGRATION DES DONNEES (PROCESSUS ETL) SOUS POWER BI .....</b>	<b>6</b>
3.1. EXTRACTION DES DONNEES.....	6
3.2. TRANSFORMATION DES DONNEES (POWER QUERY).....	6
3.3. MODELISATION DU SID SOUS POWER BI.....	7
3.4. CREATION DES MESURES AVEC DAX .....	7
3.5. CREATION DE VISUALISATIONS SOUS POWER BI.....	8
<b>RESTITUTION.....</b>	<b>9</b>

## Contextualisation

Dans un contexte où les conditions météorologiques influencent de nombreux secteurs d'activité, tels que l'agriculture, le transport, l'énergie et la santé, l'analyse des données climatiques est devenue un enjeu majeur. Les variations de température, les précipitations ont des impacts significatifs sur la gestion des ressources naturelles, l'organisation des infrastructures et la planification des activités économiques. Une compréhension fine de ces phénomènes permet d'anticiper les risques, d'optimiser la prise de décision et d'améliorer la résilience face aux événements climatiques extrêmes, tels que les vagues de chaleur, les tempêtes ou les inondations.

Ce projet s'inscrit dans cette dynamique en proposant la conception d'un **Système d'Information Décisionnel (SID)** dédié à l'analyse des données météorologiques de **quatre villes**. En utilisant **Power BI**, nous mettons en place une solution permettant d'explorer les tendances climatiques à travers des tableaux de bord interactifs.

L'objectif principal est de **faciliter l'interprétation des données climatiques** en offrant une vision claire et synthétique des conditions météorologiques dans ces différentes localités. Ce système permettra de **comparer les tendances météorologiques** entre les villes, d'identifier les variations saisonnières et d'apporter des informations précieuses aux décideurs, qu'il s'agisse de **météorologues, d'agriculteurs, d'urbanistes ou de gestionnaires d'infrastructures**.

En exploitant des techniques de **modélisation de données**, de **transformation (ETL)** et de **visualisation dynamique**, ce projet permet de **transformer des données brutes en informations exploitables**.

**Problématique** : Comment exploiter efficacement les données météorologiques de plusieurs villes afin de visualiser les tendances climatiques et d'aider les décideurs à interpréter ces informations de manière optimale ?

## Objectifs

Ce projet a pour principaux objectifs :

- **Collecter et intégrer** les données météorologiques quotidiennes de quatre villes.
- **Modéliser un Système d'Information Décisionnel (SID)** adapté à l'analyse des données climatiques.
- **Développer des tableaux de bord interactifs sous Power BI**, permettant d'explorer et de comparer les tendances météorologiques entre les différentes villes.
- **Faciliter la prise de décision** en mettant à disposition des visualisations claires et pertinentes.

# Modélisation du SID

## 2.1. Choix du processus métier à modéliser

Le processus métier retenu pour ce projet est **l'analyse des conditions climatiques et la comparaison des tendances météorologiques entre plusieurs villes**. L'objectif principal est d'identifier les variations climatiques, de détecter des anomalies et de fournir des indicateurs permettant d'anticiper d'éventuels phénomènes extrêmes (tempêtes, vagues de chaleur, fortes précipitations, etc.).

Grâce à ce système, les utilisateurs pourront suivre l'évolution des conditions météorologiques, comparer les tendances entre différentes villes et prendre des décisions éclairées en fonction des données analysées.

## 2.2. Définition de la granularité

La granularité des données est définie au **niveau quotidien** pour chaque ville. Chaque enregistrement correspond à un ensemble de mesures météorologiques relevées à une date précise et pour une ville donnée.

Afin de faciliter l'analyse des tendances sur le long terme, des agrégations mensuelles et annuelles pourront être calculées pour fournir des indicateurs synthétiques (ex. température moyenne mensuelle, cumul des précipitations par année, etc.).

## 2.3. Choix des dimensions

Dans la modélisation du SID, plusieurs **dimensions** sont nécessaires pour structurer l'analyse des données :

- **Dimension "Date"** : Cette dimension permet d'analyser les évolutions des données météorologiques dans le temps. Elle comprend les attributs suivants :
  - Jour
  - Mois
  - Année
  - Jour de la semaine
- **Dimension "Ville"** : Cette dimension représente les différentes villes étudiées. Chaque enregistrement est associé à :
  - Nom de la ville
  - Latitude et Longitude
  - Pays
- **Dimension "Saison"** : Cette dimension indique la saison selon la date :
  - SaisonID,
  - Nom de Saison
  - Mois de début,
  - Mois de Fin

## 2.4. Identification des faits (Mesures)

Les faits correspondent aux **mesures météorologiques enregistrées quotidiennement** pour chaque ville. Ces données sont quantitatives et permettent d'effectuer des analyses comparatives. Voici les principales mesures utilisées :

- **Températures (°C)** : Température minimale, maximale et moyenne enregistrée chaque jour.
- **Humidité (%)** : Niveau d'humidité de l'air relevé quotidiennement.
- **Pression atmosphérique (hPa)** : Indicateur permettant d'analyser les variations de pression selon les conditions météorologiques.
- **Vitesse du vent (km/h)** : Intensité du vent mesurée chaque jour.
- **Direction du vent (°)** : Orientation du vent (ex. Nord, Sud-Ouest, etc.).
- **Précipitations (mm)** : Quantité de pluie tombée sur une journée.
- **Ensoleillement (heures)** : Nombre d'heures d'ensoleillement par jour.

# Intégration des données (Processus ETL) sous Power BI

## 3.1. Extraction des données

- **Sources de données :**
  - Fichiers **CSV** : Nos différentes bases de données proviennent des sites [www.kaggle.com](http://www.kaggle.com), [meteo.data.gouv.fr](http://meteo.data.gouv.fr) et [historique-meteo.net](http://historique-meteo.net). Ces bases de données sont libres de droit et chaque base de données correspond à une ville.
- **Importation dans Power BI :**
  - Nous avons ensuite importé les quatre bases de données (Rome, Londres, Paris et Berne).

## 3.2. Transformation des données (Power Query)

1. **Nettoyage des données**
  - Nous avons supprimé les doublons : **Transformer > Supprimer les doublons**.
  - Vérifier les valeurs aberrantes (ex : température extrême de +100°C ou -50°C) : pour la ville de Paris, nous avons des températures maximales aberrantes (plus de 1000°C). Étant donné que les températures minimales et moyennes étaient correctes, nous avons recalculé la température maximale =  $2 \times \text{température\_moyenne} - \text{température\_minimale}$
  - Nous avons ensuite changé le format des dates pour les mettre dans le bon format.
2. **Traitement des valeurs manquantes**
  - Nous n'avons pas de problèmes de valeurs manquantes.
3. **Conversion des unités**
  - Toutes les données de températures sont bien en degrés Celsius (°C).
  - Toutes les données de précipitations sont bien en millimètres (mm).
4. **Création de nouvelles colonnes**
  - Nous avons ajouté une colonne VilleID qui permet de connaître la ville pour chacune des bases de données.
5. **Création de la table calendrier (Table Date)**
  - En allant dans **Modélisation > Nouvelle table** et en entrant la formule DAX, nous avons générer des dates :

```
DAX
Table_Calendar = ADDCOLUMNS (
    CALENDAR (DATE(2010,1,1), DATE(2020,12,31)),
    "Année", YEAR([Date]),
    "Mois", FORMAT([Date], "MMMM"),
    "Jour", DAY([Date]),
    "Semaine", WEEKNUM([Date])
)
```

Ici l'année sera 2020 car nous avons que des bases de données pour cette année-là.

- Nous avons relié cette table aux données météo pour analyser les tendances temporelles.

## 6. Fusion des Tables Météo

Nous avons combiné les quatre bases de données (Paris, Berne, Rome, Londres) en une **table de faits unique (Faits\_Météo)** via l'option "**Combiner, Ajouter des requêtes**" dans Power BI.

- **Ajout d'une colonne "Ville"** dans chaque table pour identifier l'origine des données.
- **Fusion des tables** pour obtenir un jeu de données unifié.
- **Nettoyage et harmonisation** : vérification des types de données, gestion des valeurs manquantes.

## 3.3. Modélisation du SID sous Power BI

### 1. Structure du modèle relationnel (Modèle en étoile)

- **Table de faits** :
  - Faits\_Météo (Température, Précipitations, Humidité, Vent, etc.).
- **Tables de dimensions** :
  - Date (Jour, Mois, Année, Semaine).
  - Ville (Ville, Pays, Latitude, Longitude).
  - Saison (SaisonID, Mois de début, Mois de Fin,).

### 2. Création des relations dans Power BI

- Nous avons relié la table Faits\_Météo aux **tables de dimensions** (Date, Ville, etc.) via les **clés primaires**.

## 3.4. Création des mesures avec DAX

### • Température moyenne mensuelle

DAX

```
Temp_Moyenne_Mois = AVERAGEX( VALUES('Date'[Mois]),
  AVERAGE('Faits_Météo'[Température_Moyenne]) )
```

### • Nombre de jours de pluie par mois

DAX

```
Jours_Pluie = CALCULATE(COUNT('Données_Météo'[Date]),
  'Faits_Météo'[Précipitations] > 0)
```

### • Comparaison des températures maximales entre deux périodes

DAX

```
Diff_Temp_Max =
  VAR Temp_Actuelle = MAX('Données_Météo'[max_temp])
```

```
VAR Temp_Passee = CALCULATE(MAX('Faits_Météo'[max_temp]),  
PREVIOUSMONTH('Date'[Année]))  
RETURN Temp_Actuelle - Temp_Passee
```

- **Code DAX pour catégorisation par mois**

Journee\_Type\_Mois =

```
VAR TempMaxMois = AVERAGE('Faits_Météo'[max_temp])
```

RETURN

```
SWITCH(  
    TRUE(),  
    TempMaxMois > 25, "Mois chaud",  
    TempMaxMois < 10, "Mois froid",  
    "Mois tempéré"  
)
```

### 3.5. Création de visualisations sous Power BI

### 1. Carte des températures par ville

- **Visuel : Carte (Map ou Filled Map).**
- Ajouter Ville dans **Localisation** et Température Moyenne en **Valeur**.

## 2. Graphique d'évolution des températures

- **Visuel : Courbe de tendance.**
- Axe X : Date[Mois], Axe Y : Température Moyenne Mensuelle.

### 3. Histogramme des précipitations

- **Visuel : Histogramme.**
- **Axe X : Mois, Axe Y : Précipitations Totales.**

#### 4. Comparaison inter-villes

- **Visuel : Tableau récapitulatif.**
- Colonnes : Ville, Température Moyenne, Précipitations

### Schéma du modèle en étoile





## 4. Restitution

### 4.1 Tableau de bord

