**République Algérienne Démocratique et Populaire**

**Ministère de l’Enseignement Supérieur et de la Recherche Scientifique**

**Université Ben Youcef BENKHEDDA-Alger1**

**Master 1 Ingénierie Des Systèmes Informatiques Intelligent ISII**



**Faculté des Sciences**

**Département informatique**

**Module : entrepôt de données**

Rapport projet entrepôt de données

**Présenté par :**

* TOUIL FADIA
* CHIKHI MOHAMED AMINE
* BELKESSA SAMY
* BOUNOUA SARAH

**Année Universitaire : 2023/2024**

# Sommaire :

1. Introduction
2. Le schéma en étoile du Data Warehouse.
3. Le processus traité par le Data Warehouse.
4. La granularité de la table des faits.
5. Les mesures et les dimensions.
6. Le tableau de bord.
7. Comment faire marcher le projet (environnement).
8. Conclusion

# Introduction :

L'objectif de ce projet est de créer un entrepôt de données pour analyser les données climatiques et fournir des visualisations interactives à l'aide de Dash. Ce rapport détaille le schéma en étoile de l'entrepôt de données, le processus de traitement des données, la granularité de la table des faits, les mesures, les dimensions et le tableau de bord.

# Le schéma en étoile du Data Warehouse :

Le schéma en étoile est une technique de modélisation des données largement utilisée dans les entrepôts de données en raison de sa simplicité et de son efficacité pour les requêtes analytiques. Il se compose d'une table des faits centrale reliée à plusieurs tables de dimensions. Dans notre projet, le schéma en étoile est conçu pour stocker et analyser les données climatiques.

## Choix de la table des faits et des dimensions :

La table des faits Weather\_Fact est au cœur de notre schéma en étoile. Elle contient les mesures climatiques quantitatives, telles que les précipitations, la température moyenne, la température maximale et minimale, la neige, etc. Ces mesures sont essentielles pour analyser les tendances climatiques et les conditions météorologiques extrêmes.

Les dimensions sélectionnées, Station\_Dim et Date\_Dim, fournissent le contexte nécessaire pour analyser ces mesures :

Station\_Dim : contient des informations sur les stations météorologiques, telles que le code de la station, le nom, la latitude, la longitude, l'élévation et le pays. Cette dimension permet d'analyser les données climatiques par localisation géographique.

Date\_Dim : contient des informations sur les dates, y compris l'année, le mois et le jour. Cette dimension permet d'analyser les données climatiques par période temporelle.

# Le processus traité par le Data Warehouse :

Le processus de traitement des données climatiques dans l'entrepôt de données comprend plusieurs étapes :

1. **Extraction des données :** Les données climatiques sont extraites des fichiers brutes fournis pour la réalisation de ce projet.
2. **Transformation des données :** Les données sont nettoyées et transformées pour correspondre à la structure des tables de l'entrepôt de données. Cette étape comprend la gestion des valeurs manquantes, la conversion des types de données et l'ajout des identifiants nécessaires.

Le script python utilisé pour la transformation des données brutes est décrit dans le fichier etl.ipynb.

P.S : il ne faut pas exécuter ce fichier. Il a déjà été exécuter une fois pour créer le fichier Weather\_data.csv qui contient maintenant les données nettoyées et avec lequel on a rempli notre entrepôt de données.

1. **Chargement des données :** Les données transformées sont chargées dans les tables de faits et de dimensions de l'entrepôt de données.

Le script suivant montre comment les données sont chargées dans les tables :



# Granularité de la table des faits :

La granularité de la table des faits Weather\_Fact est définie par la combinaison des dimensions StationID et Date\_ID. Chaque enregistrement dans la table des faits représente les mesures climatiques d'une station particulière pour un jour précis. Cette granularité permet une analyse détaillée des données climatiques à un niveau quotidien et par station. Elle est choisie pour offrir une vue précise des conditions météorologiques sur des périodes spécifiques et des emplacements particuliers.

# Les mesures et les dimensions :

## Les Mesures :

Les mesures dans la table des faits Weather\_Fact incluent les variables climatiques quantitatives suivantes :

* **PRCP :** Précipitations (en millimètres) - Mesure des précipitations pour analyser les tendances des précipitations et des conditions de sécheresse ou de pluie.
* **TAVG :** Température moyenne (en degrés Celsius) - Mesure de la température moyenne pour analyser les variations de température au fil du temps.
* **TMAX :** Température maximale (en degrés Celsius) - Mesure de la température maximale pour identifier les jours de chaleur extrême.
* **TMIN :** Température minimale (en degrés Celsius) - Mesure de la température minimale pour identifier les jours de froid extrême.
* **SNWD :** Profondeur de la neige (en millimètres) - Mesure de la profondeur de la neige pour analyser les conditions de neige.
* **PGTM :** Pression atmosphérique (en hectopascals) - Mesure de la pression atmosphérique pour comprendre les variations de la pression.
* **SNOW :** Chute de neige (en millimètres) - Mesure de la quantité de neige tombée.
* **WDFG :** Direction du vent (en degrés) - Mesure de la direction du vent pour étudier les patterns de vent.
* **WSFG :** Vitesse maximale du vent (en km/h) - Mesure de la vitesse maximale du vent pour identifier les jours de vents forts.

Ces mesures sont essentielles pour fournir une vue complète des conditions climatiques et permettre des analyses approfondies.

## Les dimensions :

Les dimensions fournissent le contexte nécessaire pour interpréter les mesures climatiques. Les dimensions sélectionnées pour ce projet sont :

* **Station\_Dim :** Cette dimension contient des informations sur les stations météorologiques telles que leur code, nom, latitude, longitude, élévation et pays. Cela permet d'analyser les données climatiques en fonction de l'emplacement géographique.
* **Date\_Dim :** Cette dimension contient des informations temporelles telles que la date, l'année, le mois et le jour. Cela permet d'analyser les tendances climatiques sur différentes périodes.

Les dimensions sont choisies pour leur capacité à fournir un contexte géographique et temporel, ce qui est crucial pour des analyses climatiques complètes et détaillées.

# Tableau de bord :

Le tableau de bord interactif est conçu pour permettre une exploration intuitive des données climatiques. Il utilise Dash pour fournir des visualisations interactives. Les éléments clés du tableau de bord incluent :

* **Dropdowns pour la sélection des stations et des années :** Ces contrôles permettent à l'utilisateur de filtrer les données en fonction de la station météorologique et de l'année.
* **Dropdown pour la sélection des mesures :** Ce contrôle permet à l'utilisateur de choisir la mesure climatique à visualiser.
* **Graphiques interactifs :**
  + **Line Chart :** Montre l'évolution des mesures climatiques au fil du temps pour une station et une année sélectionnée.
  + **Map Graph :** Visualise les mesures climatiques sur une carte géographique pour toutes les stations.
  + **Histogram :** Montre la distribution des mesures climatiques pour une station et une année sélectionnée.

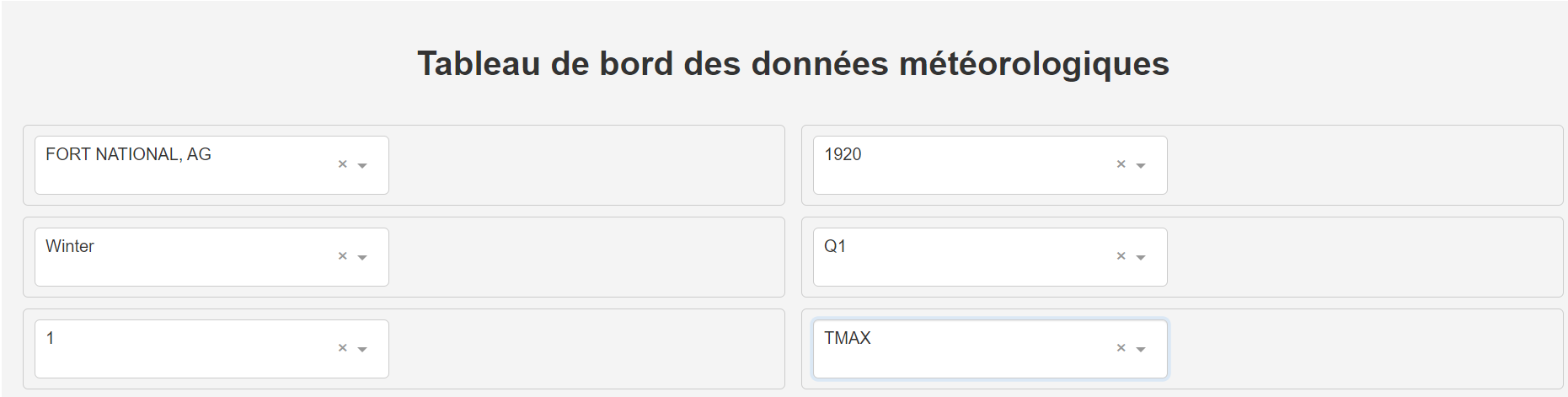
Les graphiques sont mis à jour dynamiquement en fonction des sélections de l'utilisateur, permettant une exploration approfondie et personnalisée des données climatiques.

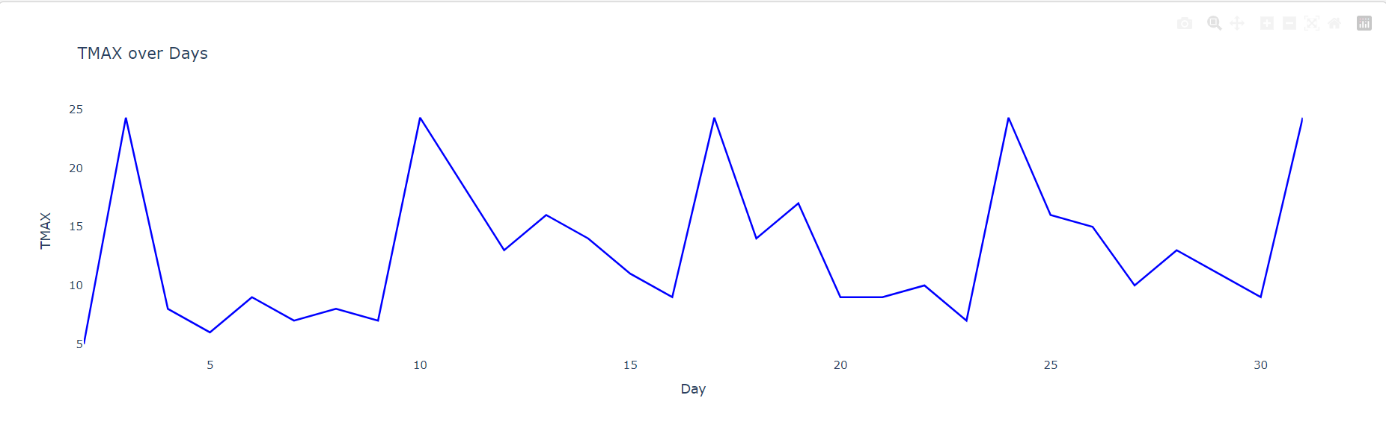
Quand vous sélectionner un mois dans l’interface, il faut aussi sélectionner le trimestre et la saison correspondante a ce mois pour un affichage correcte des graphes.

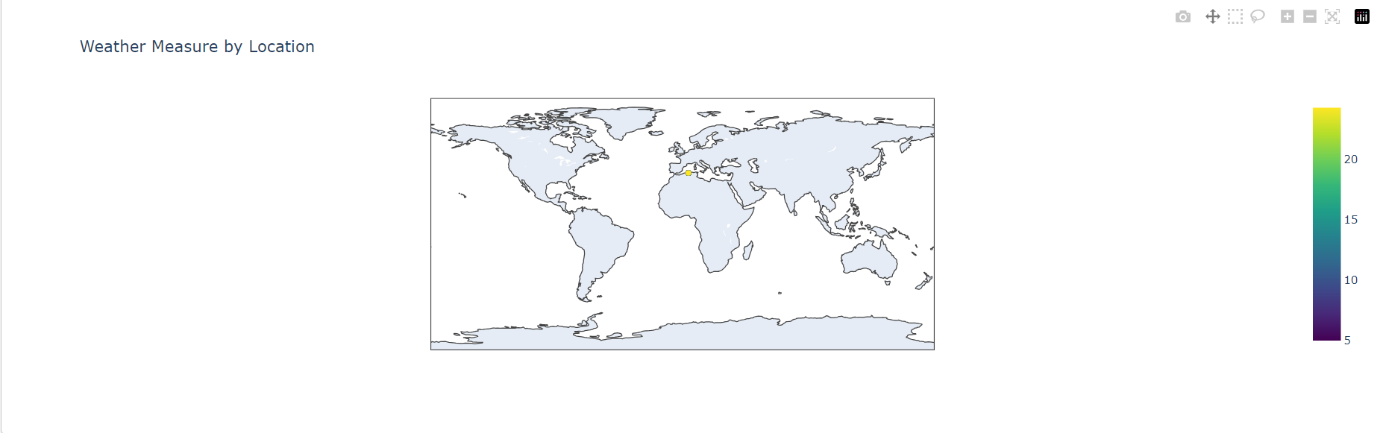
Si aucun graphe ne s’affiche pour des sélections données, cela veut dire qu’aucune donné n’existe pour ces sélections précisément.

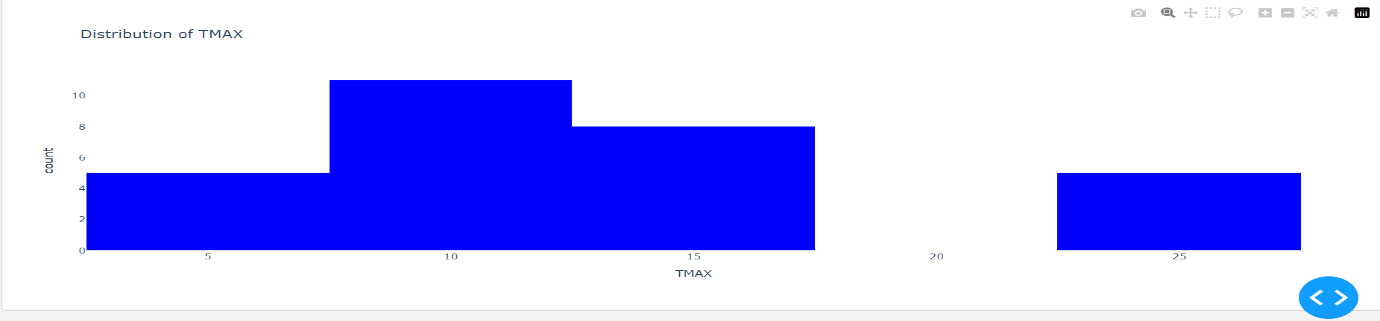
Le code de l’interface est situé dans le fichier interface.py.

Les figures suivantes montrent le tableau de bord.









# Comment faire marcher le projet (environnement) :

## Télécharger et installer Python 3.x :

Pour télécharger Python, il faut se rendre sur ce lien et cliquer sur download : <https://www.python.org/downloads/>

Pour installer correctement python, vous pouvez suivre ce tutoriel : <https://www.youtube.com/watch?v=2faQMy72k3A&t=165s&ab_channel=MikailAltundas>

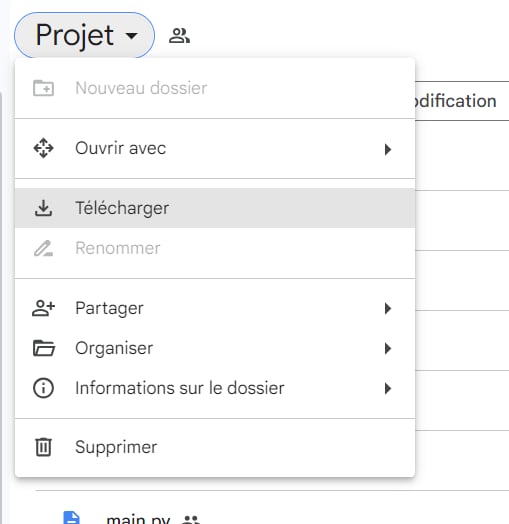
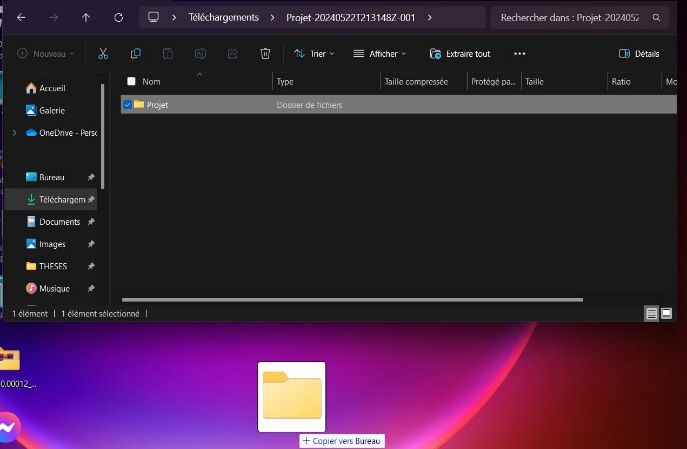
## Télécharger et installer WampServer :

Il faut installer l’application WampServer pour avoir accès à PhpMyAdmin ou sera créer notre base de données (notre datawarehouse).

Vous pouvez suivre ce tutoriel pour installer WampServer : <https://www.youtube.com/watch?v=M2at7D-lciw&t=757s&ab_channel=GeekyScript>

## Télécharger et exécuter le projet :

* **Etape 1 :** commencez par télécharger le dossier du projet à partir du lien google drive qu’on vous partagé par mail, ensuite mettez le dossier projet qui se trouve dans le dossier que vous avez télécharger sur le bureau.

* **Etape 2 :** Lancer l’application WampServer en cliquant sur l’icône correspondante.



* **Etape 3 :** Faite clique droit sur le dossier projet et cliquez sur ouvrir dans le Terminal et exécuter les commandes suivantes une par une pour installer tous les paquets nécessaires.

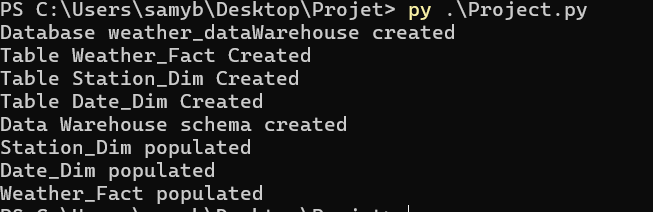
pip install dash

pip install plotly

pip install pandas

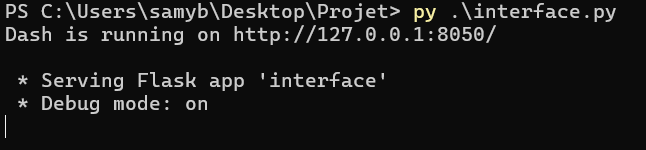
pip install pymysql

* **Etape 4 :** écrire la commande suivante dans l’invite de commande pour exécuter le fichier Project.py pour créer le schéma du Data Warehouse et peupler les tables. L’exécution peut prendre quelque longes minutes en raison de la masse de données.



py .\Project.py

* **Etape 5 :** écrire la commande suivante dans l’invite de de commande pour exécuter le fichier interface.py pour afficher le tableau de bord. L’exécution peut prendre quelques minutes. Cliquer sur l’url afficher comme le montre la figure en dessous.



py .\interface.py

# Conclusion :

Ce projet met en œuvre un entrepôt de données climatiques avec un schéma en étoile pour faciliter l'analyse des données. Le processus de traitement des données comprend l'extraction, la transformation et le chargement des données dans les tables de faits et de dimensions. La granularité de la table des faits permet une analyse détaillée des mesures climatiques. Les dimensions fournissent un contexte géographique et temporel essentiel pour les analyses. Le tableau de bord interactif offre une visualisation intuitive des données, permettant une exploration approfondie des tendances climatiques.