

# Tableau de bord épidémiologie avec PowerBI



Parcours Data Analyste – Projet 10 'Etude sur l'eau' option C

MARCHAND Emmanuel – 09/2024

Fournir des **outils d'aide à la décision** pour les PRO de l'espace vert , **comment ?**



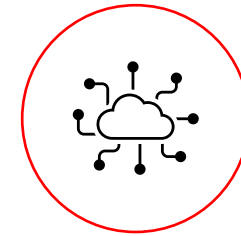
## EPIDEMIO- SURVEILLANCE

**Observations** des bio-agressions sur le territoire grâce aux déclarations de la communauté de professionnels et amateurs.



## MODELES PREDICTIFS

**Calculs algorithmiques** de risque de développement ou d'apparition d'un bio-agresseur à partir de la collecte et du traitement des données Météo.



## HUB CAPTEURS CONNECTES (IOT)

**Intégration et interprétation** des données de capteurs sol et foliaires en temps réel. Plusieurs marques, plusieurs réseaux.

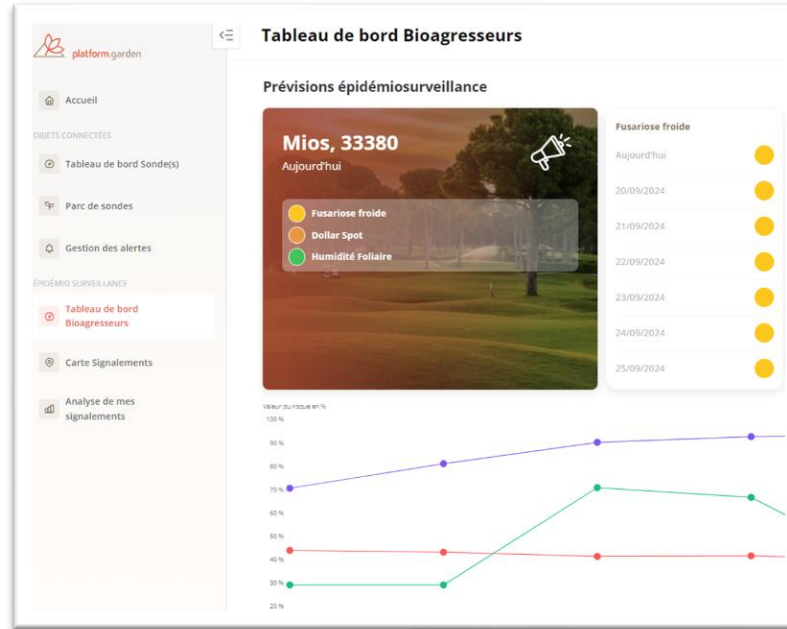
# Le sujet et la problématique du projet : l'épidémio-surveillance



**Appli mobile** pour signaler, consulter, prévisions à 5 jours.

Chaque signalement est géolocalisé et daté.

Visible sur une carte par la communauté



**Portail web – Portail PRO** : Analyse avec données agronomiques sur une semaine.

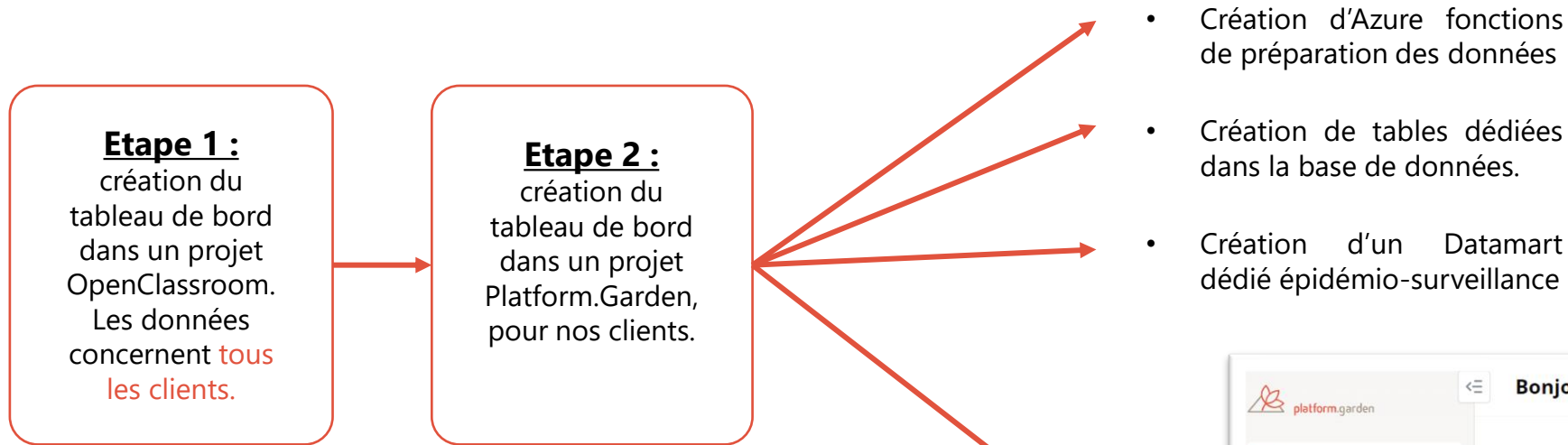
Visualisation des courbes de risques. Les risques sont calculés à partir de données météo et spatialisées grâce à un maillage de 674 stations météo sur le territoire.



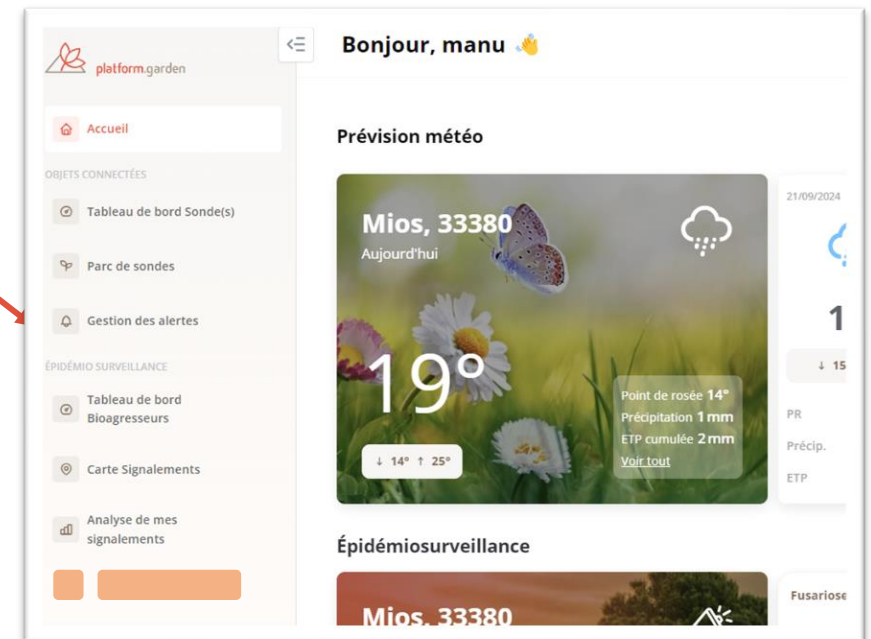
**Problématique** : encourager à signaler en capitalisant sur les données de signalements pour **préparer** la saison, **anticiper** les risques, analyser les évolutions de pressions des bioagresseurs



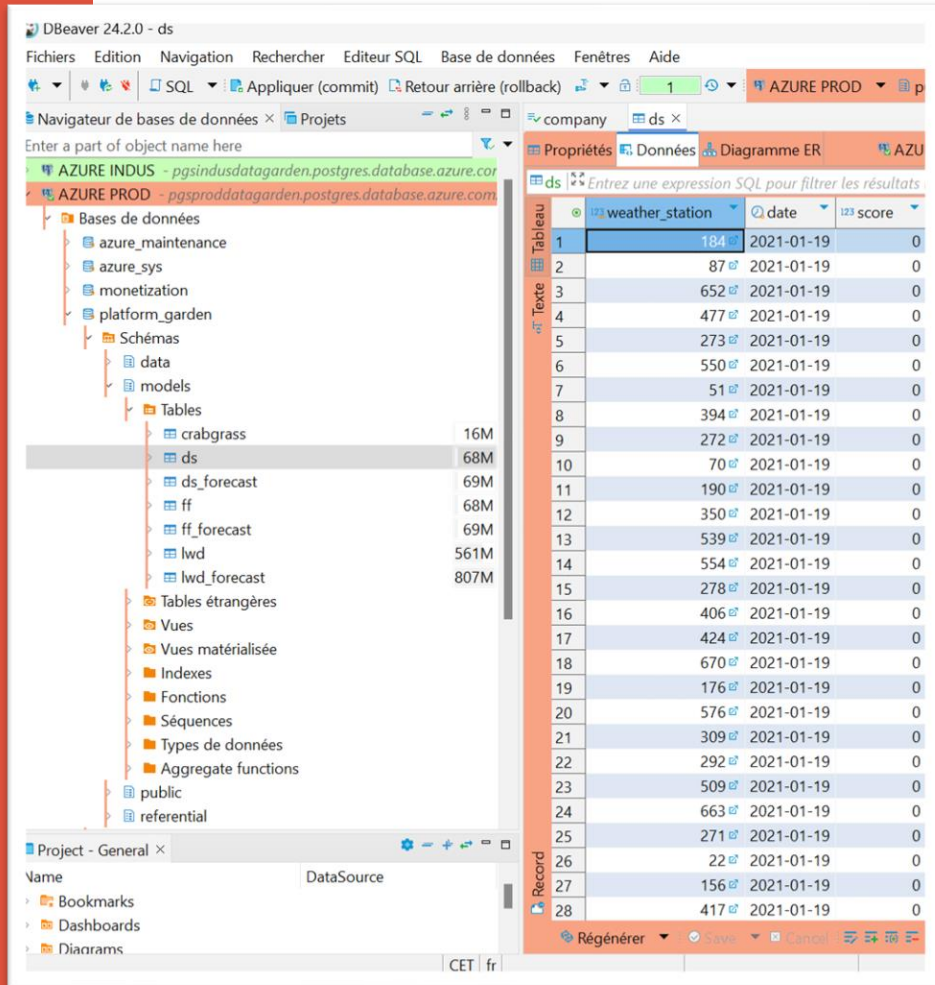
# Une livraison pour nos clients en deux étapes



Création d'un rapport en **Power BI Embedded intégré dans le Portail PRO**. Filtrage sur les données du client uniquement.



# Présentation des données utiles au projet



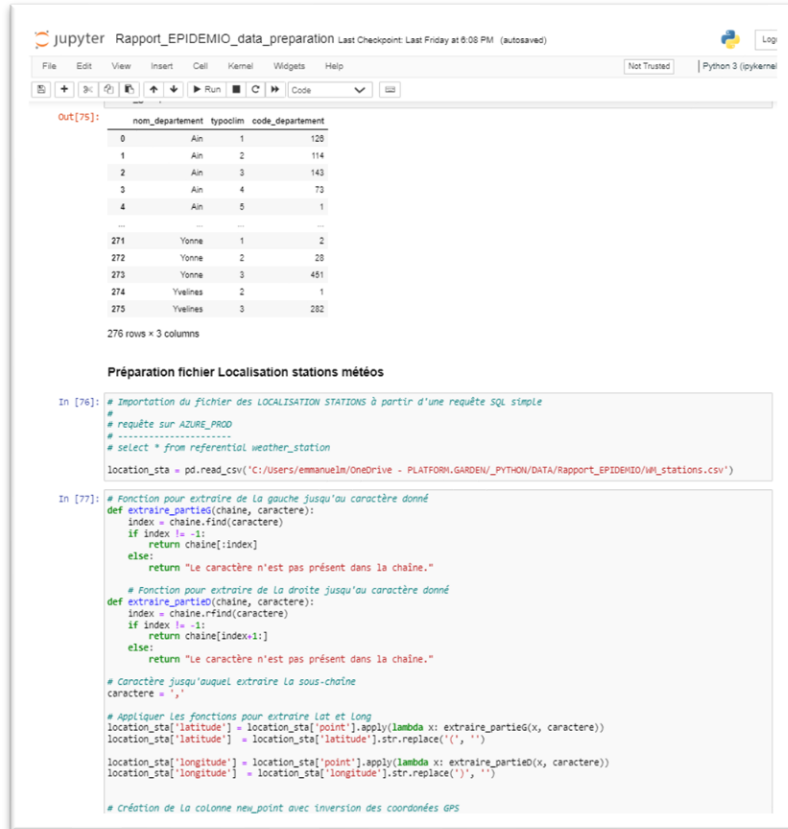
Les données seront extraites de la base de données de PROD de l'application Platform.Garden, par le logiciel Dbeaver sous forme de fichiers csv.

Cela inclue les tables suivantes :

- Tables des signalements (2 772 lignes)
- Table des stations météo (674 lignes)
- Table des risques Dollar Spot (833 034 lignes)
- Table des risques Fusariose froide (835 721 lignes)
- Table des données météo des stations (980 221 lignes)
- Table des types de climats par communes (38 655 lignes)
- Table des libellés de type de climats



# Préparation et nettoyage des données – utilisation de PYTHON



```
Out[75]:
```

	nom_departement	typoclim	code_departement
0	Ain	1	128
1	Ain	2	114
2	Ain	3	143
3	Ain	4	73
4	Ain	5	1
...	...	...	...
271	Yonne	1	2
272	Yonne	2	28
273	Yonne	3	451
274	Yvelines	2	1
275	Yvelines	3	282

276 rows x 3 columns

**Préparation fichier Localisation stations météo**

```
In [76]: # Importation du fichier des LOCALISATION STATIONS à partir d'une requête SQL simple
# requête sur AZURE_PROD
# .....
# select * from referential weather_station
location_sta = pd.read_csv('C:/Users/emmanuel/OneDrive - PLATFORM.GARDEN/PYTHON/Rapport_EPIDEMIO/wl_stations.csv')

In [77]: # Fonction pour extraire de la gauche jusqu'au caractère donné
def extraire_partie(chaine, caractere):
    index = chaine.find(caractere)
    if index != -1:
        return chaine[:index]
    else:
        return "Le caractère n'est pas présent dans la chaîne."

# Fonction pour extraire de la droite jusqu'au caractère donné
def extraire_partie2(chaine, caractere):
    index = chaine.rfind(caractere)
    if index != -1:
        return chaine[index+1:]
    else:
        return "Le caractère n'est pas présent dans la chaîne."

# Caractère jusqu'auquel extraire la sous-chaîne
caractere = ','

# Appliquer les fonctions pour extraire lat et long
location_sta['latitude'] = location_sta['point'].apply(lambda x: extraire_partie(x, caractere))
location_sta['latitude'] = location_sta['latitude'].str.replace(' ', '')

location_sta['longitude'] = location_sta['point'].apply(lambda x: extraire_partie2(x, caractere))
location_sta['longitude'] = location_sta['longitude'].str.replace(' ', '')

# Création de la colonne new_point avec inversion des coordonnées GPS
```

Sous Jupyter Notebook, création d'un script pour préparer les données qui seront utilisées dans le tableau de bord Power BI.

- **Fichier des types de climats**  
par communes : **suppression**  
des colonnes inutiles
- **Fichier des stations météo :**  
**ajout** des deux colonnes pour la longitude et la latitude . **Ajout**  
du type de climat en  
recherchant la commune la plus  
proche de la station.
- **Fichier des signalements :**  
**suppression** de colonnes inutiles,  
et suppression de 5 000  
signalements d'Ambroisie  
chargés manuellement par un  
collègue dans le cadre d'une  
étude de partenariat. **Ajout** des  
colonnes longitude et latitude.  
**Ajout** de la station météo la plus  
proche du signalement pour le  
lier au modèle de prévision de  
maladie.

Pour garantir la sécurité des données, on applique deux types **d'anonymisation** :

**Conformité RGPD** : suppression des informations permettant d'identifier les clients : **Nom, Prénom, Nom de société**, dans le fichier des signalements.

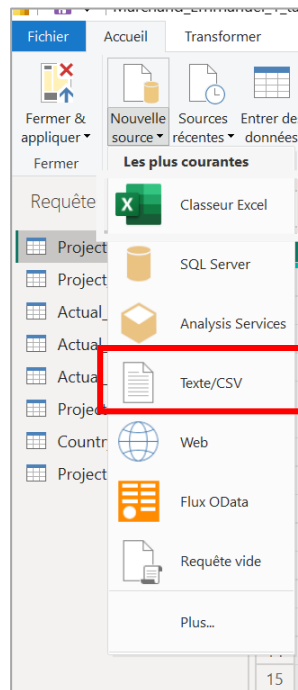


**Données sensibles** : Les résultats de modèles de risques Dollar Spot et Fusariose Froide sont au cœur de l'activité de conseil de Platform.Garden. Application d'un **coefficient arbitraire de distorsion** du résultat.

# Intégration et nettoyage des données – utilisation de Power Query

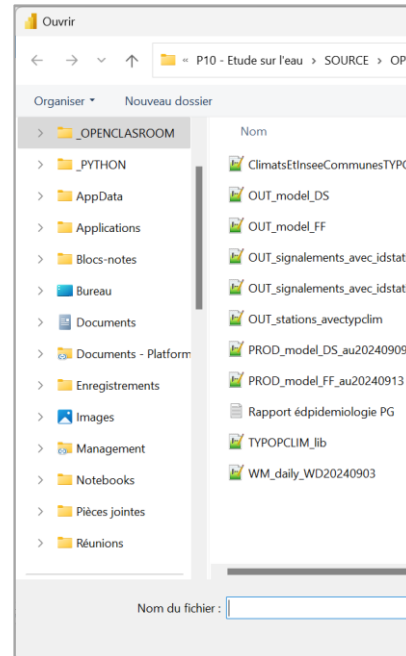
1

Sélection de la source de type fichier CSV



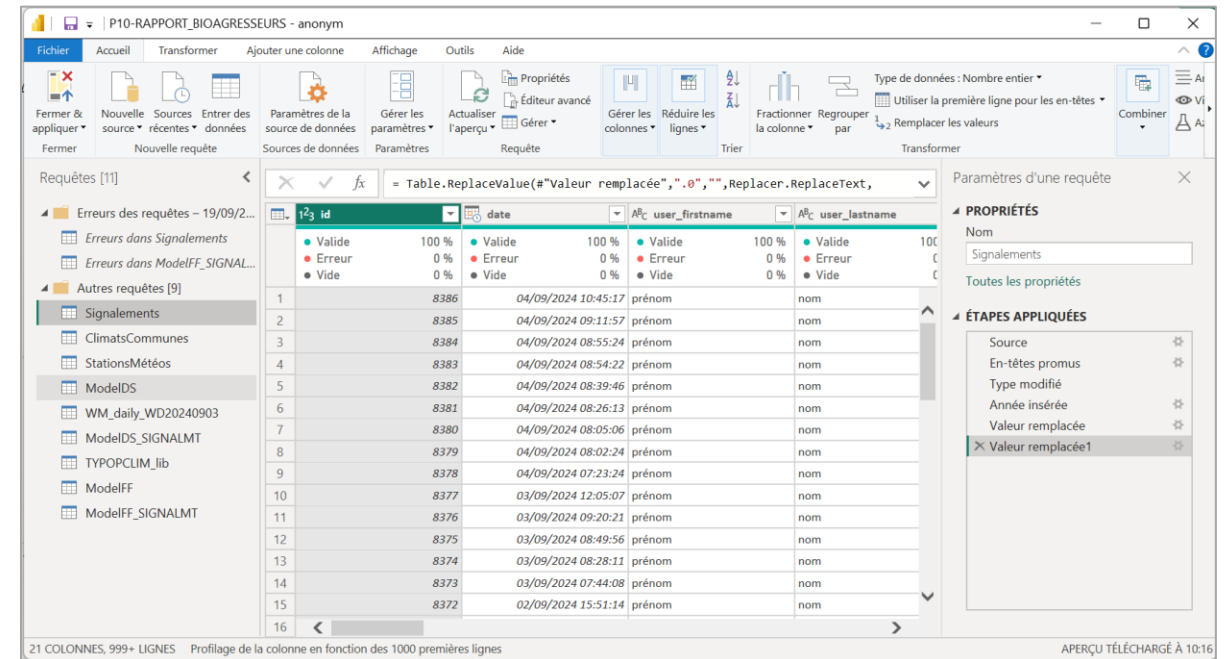
2

Importation de chaque fichier CSV, en sortie du script Python ou directement en extraction de la base de données



3

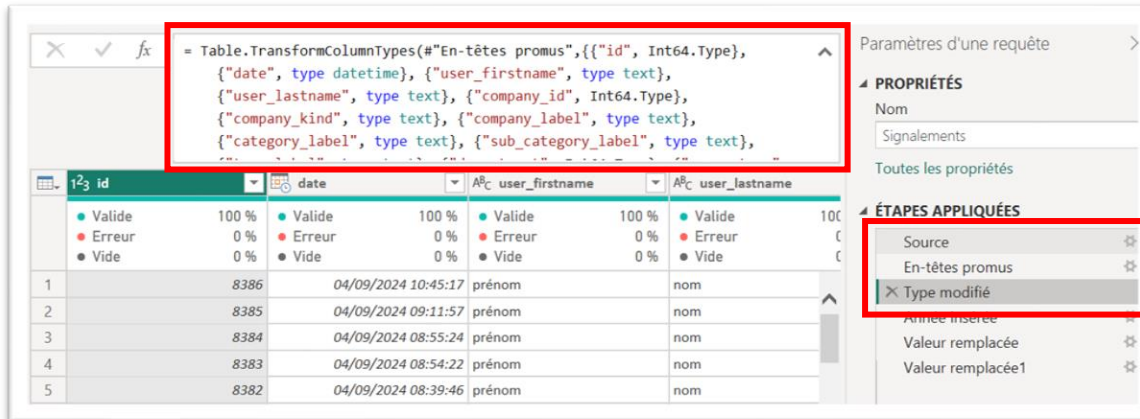
Modification de la structure de table ou des données table par table .





# Préparation Power Query – quelles actions ?

- 1 Etapes commune à tous les fichiers : Promotion des entêtes et type automatiquement affectés.



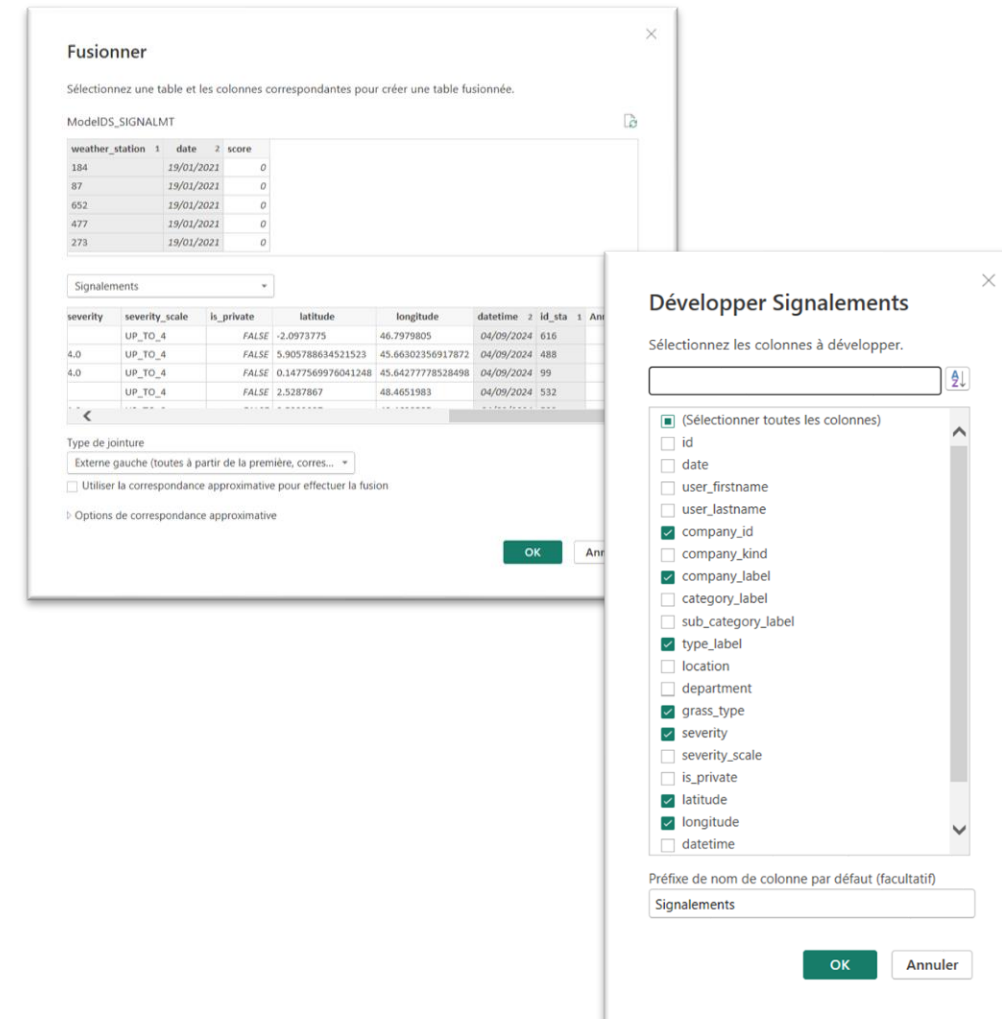
The screenshot shows the Power Query Editor interface. The formula bar contains the following code:

```
= Table.TransformColumnTypes(#"En-têtes promus",{{"id", Int64.Type}, {"date", type datetime}, {"user_firstname", type text}, {"user_lastname", type text}, {"company_id", Int64.Type}, {"company_kind", type text}, {"company_label", type text}, {"category_label", type text}, {"sub_category_label", type text}, {"category_label", type text}, {"sub_category_label", type text}}})
```

The 'ÉTAPES APPLIQUÉES' (Applied Steps) pane on the right shows the following steps:

- Source
- En-têtes promus
- Type modifié (highlighted with a red box)
- Année insérée
- Valeur remplacée
- Valeur remplacée1

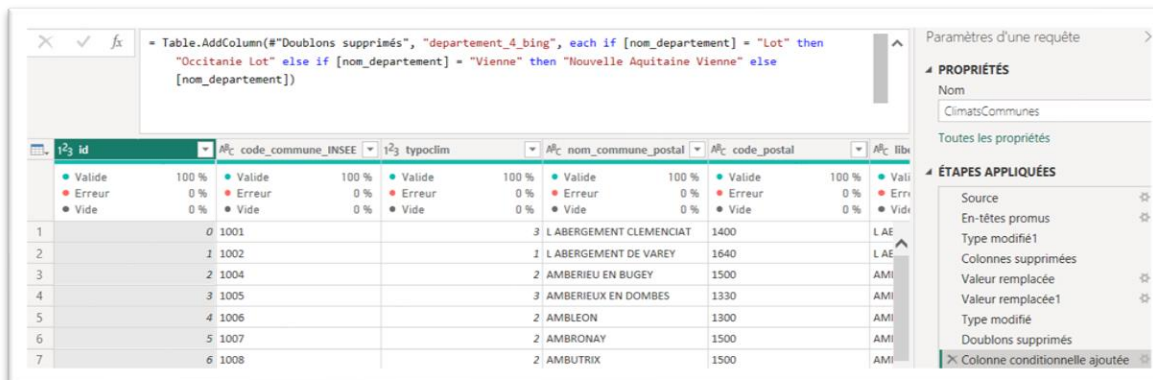
- 3 Création de deux tables **ModelIDS\_SIGNALMT** et **ModelIFF\_SIGNALMT**, pour fusionner les prévisions de risque avec les déclarations de signalement.



The screenshot shows the 'Fusionner' (Merge) dialog box. The 'ModelIDS\_SIGNALMT' table is selected. The 'Signalments' table is also selected. The 'Type de jointure' (Join type) is set to 'Externe gauche' (Left outer). The 'Options de correspondance approximative' (Approximate matching options) are checked.

The 'Développer Signalements' (Expand Signalements) dialog box is also shown. The 'Sélectionner toutes les colonnes' (Select all columns) option is checked. The 'Préfixe de nom de colonne par défaut (facultatif)' (Optional default column name prefix) is set to 'Signalments'.

- 2 Ajout de colonne conditionnelle pour régler un problème d'affichage de département



The screenshot shows the Power Query Editor interface. The formula bar contains the following code:

```
= Table.AddColumn(#"Doublons supprimés", "departement_4_bing", each if [nom_departement] = "Lot" then "Occitanie Lot" else if [nom_departement] = "Vienne" then "Nouvelle Aquitaine Vienne" else [nom_departement])
```

The 'ÉTAPES APPLIQUÉES' (Applied Steps) pane on the right shows the following steps:

- Source
- En-têtes promus
- Type modifié1
- Colonnes supprimées
- Valeur remplacée
- Valeur remplacée1
- Type modifié
- Doublons supprimés
- Colonne conditionnelle ajoutée (highlighted with a red box)

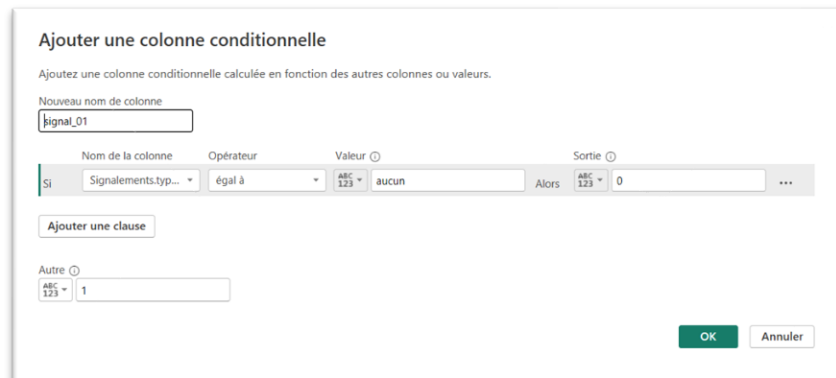
# Préparation Power Query – quelles actions ?

## 4 Modification des deux tables **ModelIDS\_SIGNALMT** et **ModelIFF\_SIGNALMT**.

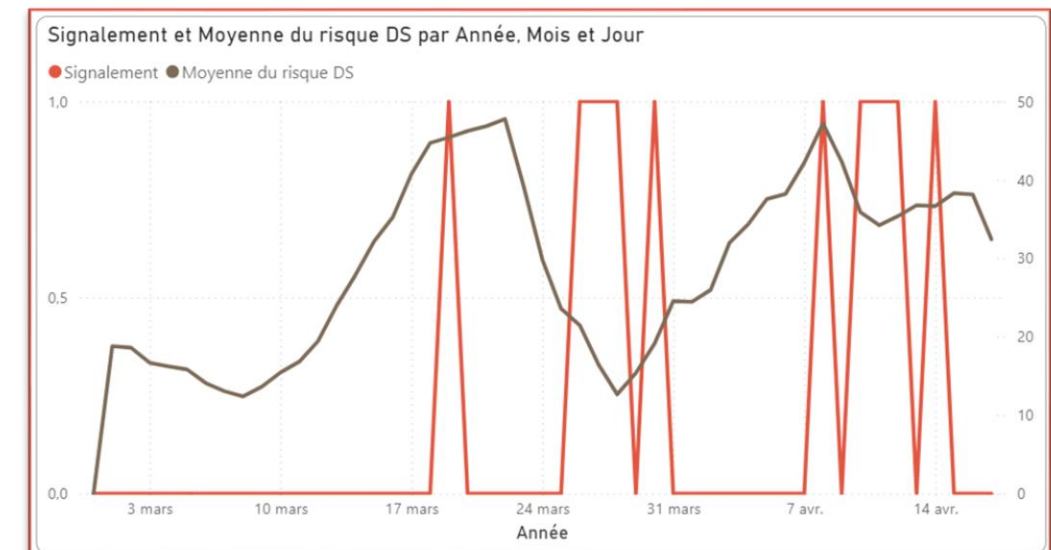
Dans la colonne '**Signalement.type\_label**' on remplace 'null' par aucun.



Création de la colonne '**Signal\_01**' pour savoir si pour un jour donné, une coordonnées GPS données il y a eu un signalement : zéro indique non, un indique oui



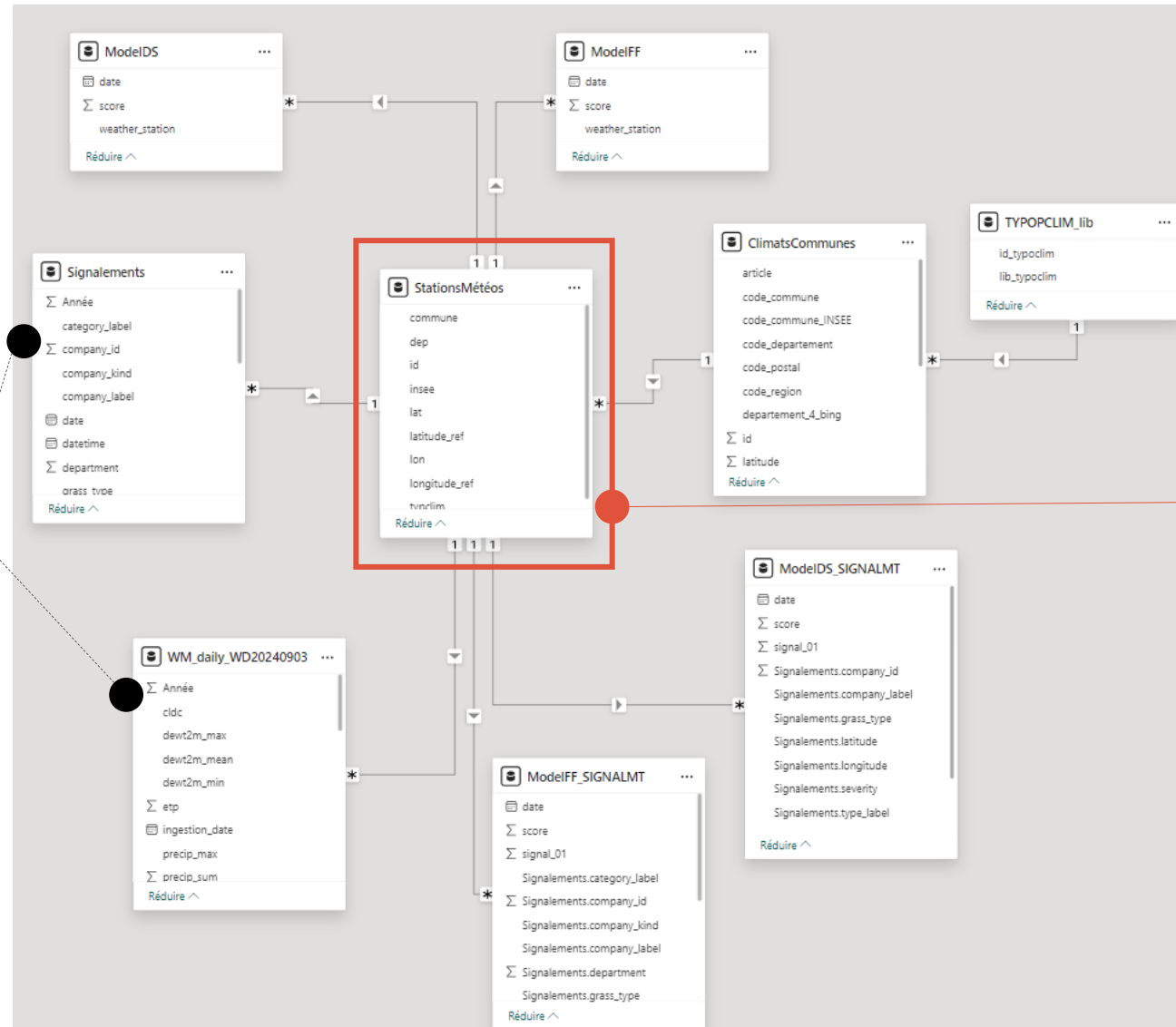
On peut afficher sur un même graphique l'évolution du risque et les jours où il y a eu des signalements



# Le modèle de données en étoile.

Les **tables de FAIT** qui rassemblent les événements connexes de surveillance des bioagresseurs :

- Risques calculés dans les modèles
- Signalements déclarés par la communauté
- Variations des données météo.
- Axe d'analyse selon le type de climat.



La **table de DIMENSION** qui rassemble toutes les infos de référence des stations météo.

# Le blue print du tableau de bord

Besoin utilisateurs	Mesures spécifiques à utiliser	Visualisation	Page/Onglet/Vue*
Zone climatique couverte par les risques	Points GPS des stations météo	Carte de France	Station
L'évolution de la météo.	Moyenne des données météos	Courbe	Station
Localisation des signalements, et nombre de signalements par mois	Point GPS de chaque signalement, Somme des signalements	Carte de France, histogramme	Signalements – Localisation
Identifier les bioagresseurs les plus recensés ?  Comprendre les périodes et les pics de déclaration.	Taux de déclaration par catégories, top des déclarations, zones les plus attaquées, somme de signalements	Secteur, courbe, Histogramme empilés en pourcentage, histogramme	Signalements répartition
Visualiser les risques d'attaque par département et pouvoir naviguer dans l'historique des risques. Voir la corrélation entre le type de climat et la prévision de risque.	Moyenne de risque de chaque département (ensemble des stations du département), la moyenne des risques de chaque station pour une période donnée	Carte choroplèthe de France avec dégradé de couleur selon le taux de risque. Nuage de point	Géographie Dollar Spot et Géographie Fusariose Froide
Faire le lien entre la courbe de risque et les signalements de DS ou FF. Pouvoir anticiper l'attaque en analysant le risque.	Moyenne des risques pour une station et une période donnée.	Carte des signalements uniquement DS ou FF, courbe des risques et des signalements.	Analyse Dollar Spot et analyse Fusariose Froide.

## Tableau de bord épidémiosurveillance (\*) doit permettre de visualiser :

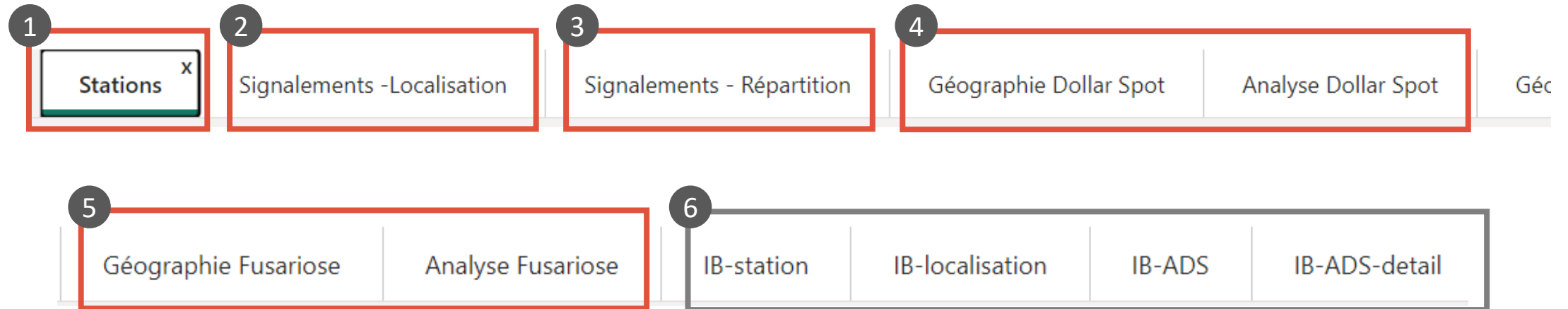
- La localisation des stations météos et le type de climat qu'elles observent.
- Une courbe des données météos clés par station.
- La localisation des signalements
- La répartition des signalements
- Une vision détaillée du modèle Dollar Spot (\*)
- Une vision détaillée du modèle Fusariose Froide. (\*)

(\*) Tableau de bord orienté pour les terrains de sport et le gazon. Cœur de cible de PG

# La structuration du tableau de bord

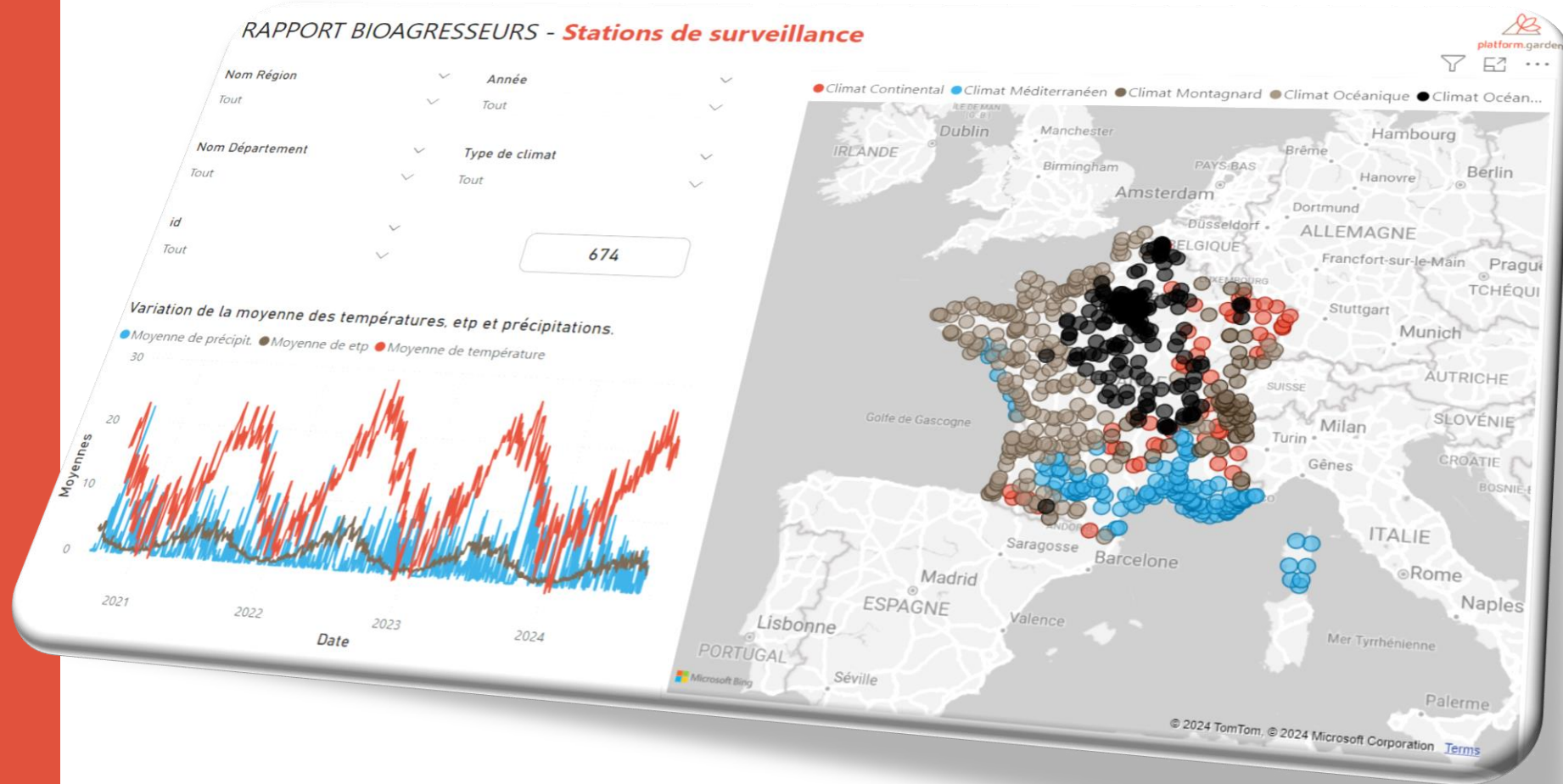
Le tableau de bord est structuré en **6 catégories d'onglets** :

- 1** Description des stations de surveillance
- 2** Géographie des signalements
- 3** Répartition des cibles et des bioagresseurs
- 4** Analyse détaillée du Dollar Spot
- 5** Analyse Détaillée de la Fusariose Froide
- 6** Infos bulles.



Ensemble d'onglets techniques utilisés pour l'info bulles des différents graphiques.

# Présentation du tableau de bord







**MERCI**