

SAE SQL 101 : SAISON BALNEAIRE 2024

HAMZAOUI

JAMEL HAMZAOUI

VEGA

BUT Sciences des données

2025

Nom SAE : Reporting à partir de données stockées dans un SGBD relationnel

Titre SAE : Rapportage de la saison balnéaire 2024

CODE : SAE 101

SOMMAIRE

I. INTRODUCTION	1
II. CONFIGURATION DE L'ENVIRONNEMENT DE TRAVAIL.....	2
III. INTERCONNEXION ET VALIDATION DU SYSTÈME	3
IV. ANALYSES EXPLORATOIRES.....	4

I. INTRODUCTION

Dans le cadre de notre SAE 101, ce rapport porte sur l'analyse de la saison balnéaire 2024. Cette étude s'appuie sur la directive européenne 2006/7/CE, qui impose de surveiller la qualité sanitaire des eaux pour protéger la santé des baigneurs face aux risques de pollution .

Notre tâche majeure consiste à regrouper l'ensemble des relevés des zones de baignade et les résultats des suivis sanitaires au sein d'un serveur PostgreSQL.

Après avoir organisé la base de données, nous exploiterons les données à l'aide de Metabase, afin de transformer les chiffres en tableaux et en graphiques. Et cela nous permettra d'interpréter les résultats.

Pour ce faire, nous détaillerons dans un premier temps l'installation des différents environnements de travail nécessaires (PostgreSQL, Java et Metabase). Ensuite, nous expliquerons la procédure de connexion entre le serveur et l'outil d'analyse. Enfin, nous terminerons par la présentation de nos analyses exploratoires portant sur les deux indicateurs bactériologiques majeurs : les Entérocoques intestinaux et les Escherichia coli.

II. CONFIGURATION DE L'ENVIRONNEMENT DE TRAVAIL

1) Installation de PostgreSQL

Le système de gestion de base de données PostgreSQL stocker vos données.

1. **Téléchargement :** Tout d'abord allez sur le site officiel de [PostgreSQL](#). Puis cliquez sur votre système (Windows/Mac/Linux), cliquez sur "Download the installer" et choisissez la version la plus récente.
2. **Installation :** Lancez le fichier téléchargé. Suivez les étapes en laissant les options par défaut.
3. **Sécurité :** Lors de l'installation, un mot de passe vous sera demandé pour l'utilisateur postgres. Mettez votre mot de passe. **Notez-le bien**, il sera indispensable pour connecter Metabase plus tard
4. **Port :** Vérifiez que le port est bien fixé sur **5432**.
5. **L'installation est donc terminée.**

2) Crédation et Peuplement de la Base de Données

Nous allons maintenant injecter les données de la SAÉ dans le serveur.

1. Lancer pgAdmin 4 : Ouvrez l'application pgAdmin. Dans le menu de gauche, cliquez sur Servers et entrez votre mot de passe qui a été créé à l'étape précédente.
2. Créer la base : Faites un clic droit sur Databases > Create > Database.... Nommez votre base sae101 et cliquez sur Save.
3. **Importer les données :**
 - Faites un clic droit sur votre nouvelle base (sae101) et sélectionnez Query Tool .
 - Cliquez sur l'icône Dossier en haut à gauche pour ouvrir votre fichier de script SQL décompressé : [lien](#)
 - Cliquez sur l'icône Play pour exécuter le script
 - La base de données est peuplée.

3) Mise en place de Metabase

Metabase nécessite Java pour fonctionner sur votre ordinateur.

1. **Installer Java :** Téléchargez la dernière version du [JDK Java](#). Choisissez la version adaptée à votre système et installez là.
2. **Installer Metabase :** Téléchargez le fichier metabase.jar : [Télécharger metabase.jar](#)
3. Sur cette page, cherchez le bouton "**Download metabase.jar**".
4. Après créez un dossier nommé "**Metabase**" sur votre bureau et placez-y le fichier .jar.

5. Lancer le serveur :

- Ouvrez votre dossier "Metabase".
- Dans la barre d'adresse du dossier (en haut), effacez le texte, tapez cmd et appuyez sur Entrée.
- Dans la fenêtre noire qui s'ouvre, tapez : java -jar metabase.jar.
- Dès que vous voyez "Metabase Initialization COMPLETE", le serveur est prêt.
- Ne fermez pas la fenêtre noire (l'invite de commande) pendant que vous utilisez Metabase, sinon le site ne fonctionnera plus.
- *Dans certains cas : Si vous avez une erreur indiquant que le port 3000 est déjà utilisé, lancez la commande java -DMB_JETTY_PORT=3001 -jar metabase.

III. INTERCONNEXION ET VALIDATION DU SYSTÈME

1) Connexion de Metabase à PostgreSQL

1. Ouvrez votre navigateur et allez sur : <http://localhost:3000> (ou <http://localhost:3001>)
2. Suivez les premières étapes pour créer votre compte administrateur.
3. Ajouter la base : Allez dans les Réglages administrateur > Bases de données > Ajouter une base de données.
4. Remplissez les champs avec précision :
 - **Type** : PostgreSQL
 - **Nom de la base** : sael01 (le nom créé à l'étape 2)
 - **Hôte** : localhost
 - **Port** : 5432
 - **Utilisateur** : postgres
 - **Mot de passe** : Celui créé à l'étape 1.
 - Cliquez sur **Sauvegarder**.

2) Génération d'un premier graphique

1. Sur l'accueil, cliquez sur + Nouveau > Requête SQL.
2. Saisissez une requête simple pour tester.
3. Cliquez sur Visualisation en bas de l'écran.
4. Choisissez le type de graphique (Barres, Camembert, etc.) dans le menu de droite.
5. Utilisez les Options pour faire vos modifications puis cliquez sur Enregistrer.

IV. ANALYSES EXPLORATOIRES

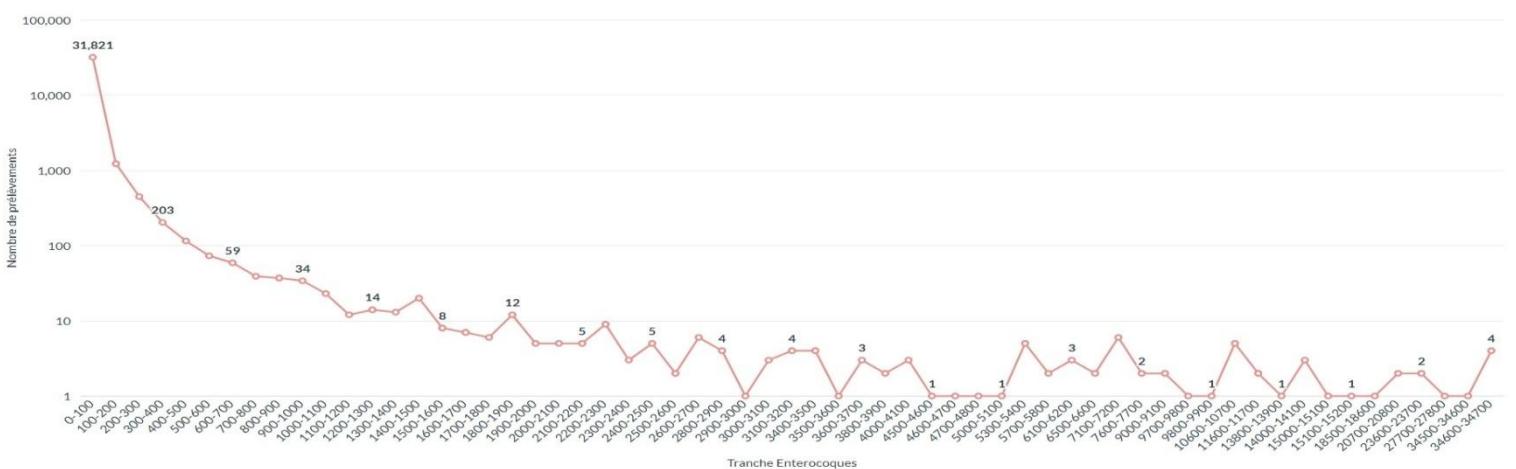
Etude 1 :

L'objectif de notre travail exploratoire est d'évaluer la qualité sanitaire globale des eaux françaises pour la saison 2024 selon la directive 2006/7/CE en comparant les indicateurs Entérocoques et E. coli. Pour traiter efficacement plus de 30 000 lignes de données sans perdre de lisibilité, j'ai mis en œuvre une discréétisation par tranches de 100 via la fonction SQL FLOOR et utilisé une échelle logarithmique en ordonnée. Ce choix technique est indispensable car il permet de visualiser sur un même graphique les volumes massifs de prélèvements conformes et les rares situations de pollution extrême qui seraient restées invisibles sur une échelle classique.

Les Entérocoques intestinaux

```
SELECT  
    FLOOR(enterocoques/100)*100 || ' - ' || (FLOOR(enterocoques/100)*100  
    + 100) AS "Tranche Enterocoques",  
    COUNT(*) AS "Nombre de prélèvements"  
  
FROM  
    Analyses  
  
GROUP BY  
    FLOOR(enterocoques/100)*100  
  
ORDER BY  
    FLOOR(enterocoques/100)*100;
```

graphique 1 :graphique de la pollution fécale par tranches de concentration en Entérocoques.

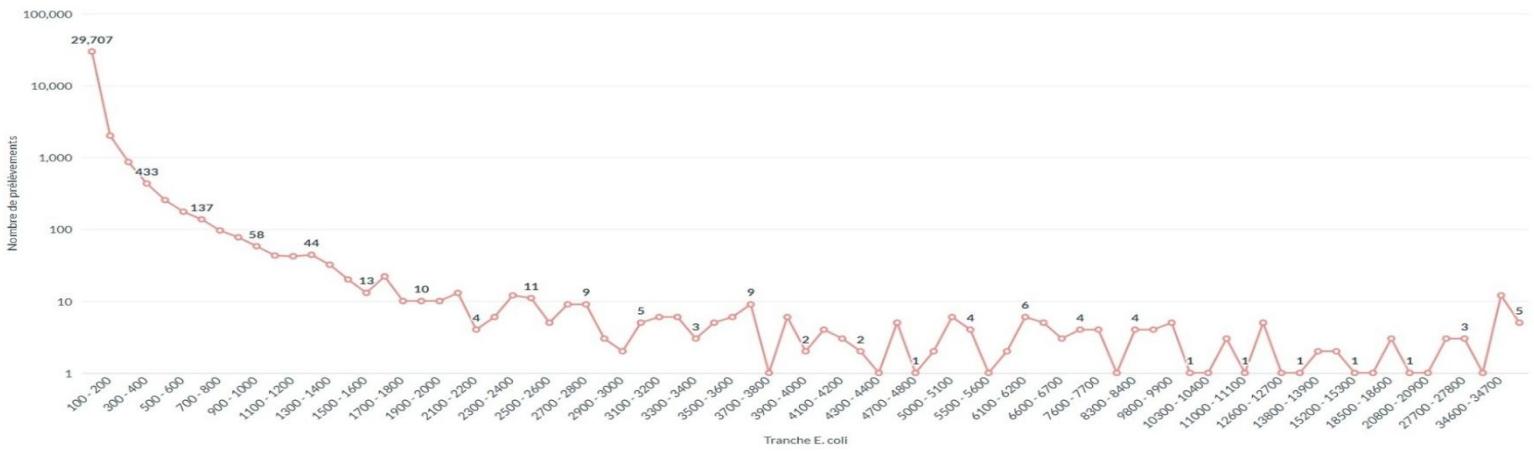


```

COUNT(*) AS "Nombre de prélèvements"
FROM
Analyses
GROUP BY
FLOOR(escherichia/100)*100
ORDER BY
FLOOR(escherichia/100)*100;

```

graphique 2 :graphique de la pollution par tranches de concentration en



Escherichia coli.

Conclusion globale : L'analyse comparative révèle une corrélation très forte entre les deux bactéries : la première tranche (0-100 UFC/100ml) est largement dominante avec 31 821 relevés pour les Entérocoque (graphique 1) contre 29 707 pour les E. coli (graphique 2), confirmant que la qualité "Excellent" est la règle sur la majeure partie du territoire. En cumulant ces données avec la deuxième tranche (100-200), on constate que l'immense majorité des sites respecte les seuils d'excellence européens fixés à 100 et 250 UFC/100ml. Cependant, si le nombre de prélèvements chute drastiquement dans les zones intermédiaires (avec par exemple seulement 58 relevés pour E. coli dans la tranche 900-1000), les deux graphiques mettent en évidence des pics de pollution critique en fin de courbe atteignant 34 700 UFC/100ml. En conclusion, cette étude prouve que si les normes sanitaires sont globalement très bien respectées, l'existence de ces points noirs de pollution massive justifie une surveillance prioritaire et ciblée pour garantir la sécurité des baigneurs.

Étude 3 : Identification et analyse des zones critiques de pollution par commune

Pour cette troisième étude, j'ai voulu aller plus loin que les statistiques générales en cherchant où se situent les zones les plus critiques. Mon objectif était de mettre des noms de communes sur les chiffres pour identifier les cas graves de pollution. J'ai donc fait une requête SQL utilisant des jointures JOIN entre les tables Analyses, Sites et Communes. J'ai regroupé les données par ville GROUP BY et après j'ai aussi calculé des moyennes AVG arrondies pour faciliter la comparaison.

```

SELECT
communes.nom AS "Commune",

```

```

COUNT(analyses.idAnalyse) AS "Nombre d'analyses",
ROUND(AVG(analyses.escherichia), 2) AS "Moyenne E. coli",
ROUND(AVG(analyses.enterocoques), 2) AS "Moyenne Entérocoques"
FROM Analyses
JOIN Sites ON analyses.idsite = sites.idsite
JOIN Communes ON Sites.codeCommune = communes.code
GROUP BY communes.nom
ORDER BY AVG(analyses.escherichia) DESC
LIMIT 10;

```

Détail des prélèvements et moyennes bactériennes par commune (Saison 2024)

Commune	Nombre d'analyses	Moyenne E. coll	Moyenne Entérocoques
QUEND	14	2,610.5	196.57
FORT-MAHON-PLAGE	15	2,455.73	233.53
THARAUX	5	2,409	423.2
MOUTIERS-EN-RETZ (LES)	12	2,367.75	447.25
SAINTE-LAURENT-DU-VAR	56	2,201.09	1,153.48
ROCHEGUDE	5	2,193.2	347
SAINTE-YRIEIX-LA-PERCHE	5	1,762	552.2
PIERRE-PERCEE	5	1,590.2	130.2
MENDE	9	1,578.22	1,319.11
CHEMILLE-SUR-INDROIS	6	1,537.33	141.17

Top 10 des communes présentant les plus fortes concentrations moyennes en bactéries



Conclusion : Cette troisième étude permet de mettre en évidence des zones critiques où la qualité de l'eau est vraiment préoccupante, ce qui tranche avec la tendance globale plutôt propre observée au début. En analysant le classement des dix communes les plus polluées, on remarque que les moyennes d'Escherichia coli et d'Entérocoques dépassent largement les seuils de sécurité. Par exemple, la commune de Quend arrive en tête du classement avec une moyenne de 2 610,5 UFC/100 mL pour l'indicateur E. coli, ce qui représente plus de dix fois le seuil d'excellence fixé à 250 par la directive européenne. Toutefois, il faut rester nuancé avec ces résultats. Comme je l'ai indiqué dans mon tableau détaillé, certaines localités comme Tharaux ou Rochegude ne comptent que 5 prélèvements sur toute la saison. Avec un échantillon aussi petit, une unique pollution accidentelle peut suffire à faire grimper la moyenne de façon spectaculaire. À l'inverse, le cas de Saint-Laurent-du-Var est beaucoup plus parlant : sa moyenne de 2 201,09 repose sur 56 analyses, ce qui rend le résultat bien plus représentatif d'un véritable problème de fond.