

TP 33

Bases de données

Dans ce TP, on se propose d'interagir avec une base de données sur le thème des communes de France. On va chercher à répondre à des questions comme :

- Dans quelle commune y a-t-il le plus de boulangeries par habitant ?
- Quelles communes ont plus de lettres que d'habitants ?
- Quel est le nom de commune le plus fréquent ?
- Quelles communes ont plus d'habitants que le chef-lieu de leur département ?

A. Environnement de travail

On pourra au choix travailler avec :

- le bac à sable SQL que j'ai utilisé en cours : <https://tchou.github.io/sqlsb/> ;
- le logiciel DB Browser for sqlite qui est normalement installé sur les machines du lycée.

B. Les données

Les données proviennent du site de l'INSEE :

- Les tables communes, departements et regions ont été créées à partir de données téléchargées à l'adresse <https://www.insee.fr/fr/information/2560452>
- Les tables demographie et equipements ont été créées puis téléchargées à partir de l'adresse suivante : <https://statistiques-locales.insee.fr>.

La base de données communes.sql regroupe les informations nécessaires.

Avant de commencer à écrire les requêtes, je vous conseille de bien vous familiariser avec les schémas des tables.

On garantit que les données sont cohérentes (les informations récupérées sur le site de l'INSEE ne concernent pas toutes la même année, mais elles ont été légèrement modifiées pour que les tables soient cohérentes entre elles). Dans toute la suite, on fera comme si toutes les données portaient sur une seule et même année.

Dans toute la suite on demande d'écrire une requête en SQL qui permet de répondre à la question posée. Toutes les requêtes doivent s'exécuter en moins de 2 s. Lorsque rien n'est précisé, le contexte détermine si la réponse peut ou non contenir des doublons. Si l'on demande par exemple *quelles sont les communes qui ...* ou *pour chaque commune donner ...*, la réponse ne doit pas contenir de doublons, on attend une ligne par commune. En revanche si la question est *quels sont les noms des communes qui ...*, sans précision supplémentaire, on attend a priori toujours une ligne par commune. Comme plusieurs communes peuvent avoir le même nom, le résultat de cette requête peut contenir des doublons.

C. Questions administratives

Dans cette section, nous utilisons uniquement les tables communes, départements et régions.

1. Donner la ligne correspondant à la commune de nom *Dijon*.
2. Combien y a-t-il de communes en France ?
3. Combien de communes s'appellent *Saint-Loup* ?
4. Combien de communes contiennent la suite de lettres *Saint* dans leur nom, sans que le nom ne commence par cette même suite de lettres ? On peut tester si a est un facteur de b avec `a LIKE '%b%'` où % désigne une chaîne quelconque.
5. Quels sont, sans doublons, les noms de communes, de départements ou de régions, dans l'ordre lexicographique croissant ?
6. Quels sont les noms des communes qui portent le même nom que leur département ?
7. Quels sont, sans doublons, les noms des communes qui portent le nom d'un département qui n'est pas le leur ?
8. Dans combien de régions différentes existe-t-il une commune s'appelant *Sainte-Marie* ?
9. Quel est le taux (nombre flottant entre 0 et 1) de communes dont le nom est constitué d'au moins trois mots séparés par des tirets ? On pourra multiplier par 1.0 pour avoir une division flottante.
10. Quelles sont les communes (tous les attributs) qui ne sont pas des chefs-lieux (de département ou de région) ?
11. Quels sont les codes et noms des communes dont le chef-lieu départemental et le chef-lieu régional ne sont pas les mêmes ?
12. Donner la table des communes avec comme colonnes : le nom de la commune, le nom du département de cette commune, le nom du chef-lieu du département, le nom de la région de cette commune, le nom du chef-lieu de la région.
13. Donner, sans doublons, le nom des communes qui ne sont pas les seules à avoir ce nom au sein d'une même région, dans l'ordre lexicographique croissant.

D. Questions démographiques

Dans cette section, nous utilisons en plus la table *demographie*.

14. Combien y a-t-il d'habitants en France ?
15. Quelle est la proportion des moins de 25 ans en France, en pourcentage entier.
16. Donner, pour chaque commune, le code de la commune et le nombre d'habitants qu'il y avait dans cette commune l'année précédente, en supposant qu'il n'y a eu ni émigration, ni immigration.
17. Quels sont, dans l'ordre lexicographique croissant, les codes des communes avec strictement plus de naissances que la moyenne des naissances par commune ?
18. Quelles sont les communes sans aucun habitant ? La table aura les mêmes colonnes que la table communes. On pourra utiliser l'étoile préfixée par le nom ou l'alias de la table.
19. Quelles sont les 19 communes les moins peuplées, parmi celles qui comportent au moins un habitant renseigné ? On donnera le nom de la commune, le nom de son département et sa population.
20. Donnez les communes (tous les attributs) ayant autant d'habitants que de lettres (tirets inclus) dans leur nom, avec ce nombre d'habitants. Vous pouvez utiliser la fonction prédéfinie `LENGTH` sur les chaînes de caractères.

21. Quels sont les codes et noms des départements dans lesquels toutes les communes ont vu au moins une naissance et un décès, si cette valeur est renseignée. On considère que le département de *Mayotte* vérifie la propriété et doit donc être présent.
22. Donnez la table des communes ayant plus d'habitants que le chef-lieu de leur département. La table aura pour colonnes : le nom de la commune, sa population, le nom du département, le nom du chef-lieu et la population de son chef-lieu.

E. Questions générales

Enfin, dans cette section, nous utilisons également la table équipements.

23. Combien y a-t-il de boulangeries en France ?
24. Combien y a-t-il, en moyenne, de pharmacies par commune ?
25. Dans combien de communes y a-t-il strictement plus de poissonneries que de crèches ?
26. Quel est le nom de la commune avec le plus faible nombre d'habitants (connu) par boulangerie ? On donnera le nom de la commune avec le nombre d'habitants par boulangerie (partie entière).

F. Autour des agrégats

27. Donnez les noms des communes qui sont utilisés par au moins deux communes, ainsi que le nombre de communes utilisant chacun de ces noms. La table sera triée par ordre décroissant suivant le nombre de communes, puis par ordre croissant des noms des communes.
28. Donner la table des régions, de leur population, du nombre de naissances et du nombre de décès. On souhaite avoir les mêmes colonnes que la table *regions*, avec trois colonnes en plus : une pour la population, une pour les naissances et une pour les décès.
29. Écrire une requête renvoyant le nom de la région contenant le département le plus peuplé ainsi que le nom de ce département.
30. Donnez la table des régions en y ajoutant le pourcentage entier de la population de la région habitant dans le chef-lieu de région.
31. Donner le nombre de lycées par département. On donnera le nom et le nombre de lycées de chaque département, ordonnés par nombre de lycées décroissant puis par nom de département croissant pour l'ordre lexicographique.
32. Quels sont les départements dans lesquels toutes les communes sont dotées d'au moins une pharmacie ?
33. Quel est le nom de la région ayant le plus de poissonneries par habitant de plus de 65 ans ?
34. Quelles sont les régions dont tous les départements ont au moins un cinquième de leurs communes qui ont au moins une crèche pour 30 naissances ?
35. Donner la table des régions avec en plus, pour chaque région, le nom du département contenant le plus de boulangeries, et la proportion, en pourcentage entier, de boulangeries de la région qui sont dans ce département.