

Université de Bordeaux
Sciences et Technologies

Master de Bioinformatique - Parcours Biologie Computationnelle

Année 2019-2020

Projet de Bases de Données
TD Insee

Abdelghani NEUHAUS

Sommaire

Introduction	3
1 Base de données	5
2 Réponses aux questions	6
2.1 Première partie	6
2.2 Deuxième partie	7
3 Discussions, remarques et conclusion	8
Annexes	9

Introduction

L'INSEE (Institut National Statistiques Etudes Economiques) publie régulièrement de nombreuses bases de données et est en charge du recensement de la population française. Parmi les données issues du référencement, nous allons travailler sur des données sociales et environnementales sur les départements français de 2008 à 2016. Il s'agira de construire une base de données contenant la géographie française actuelle (régions et départements) et les données sociales et environnementales associées à chaque département.

Partie 1

Base de données

Ci-dessous, voici un schéma de la base de données créée et utilisée :

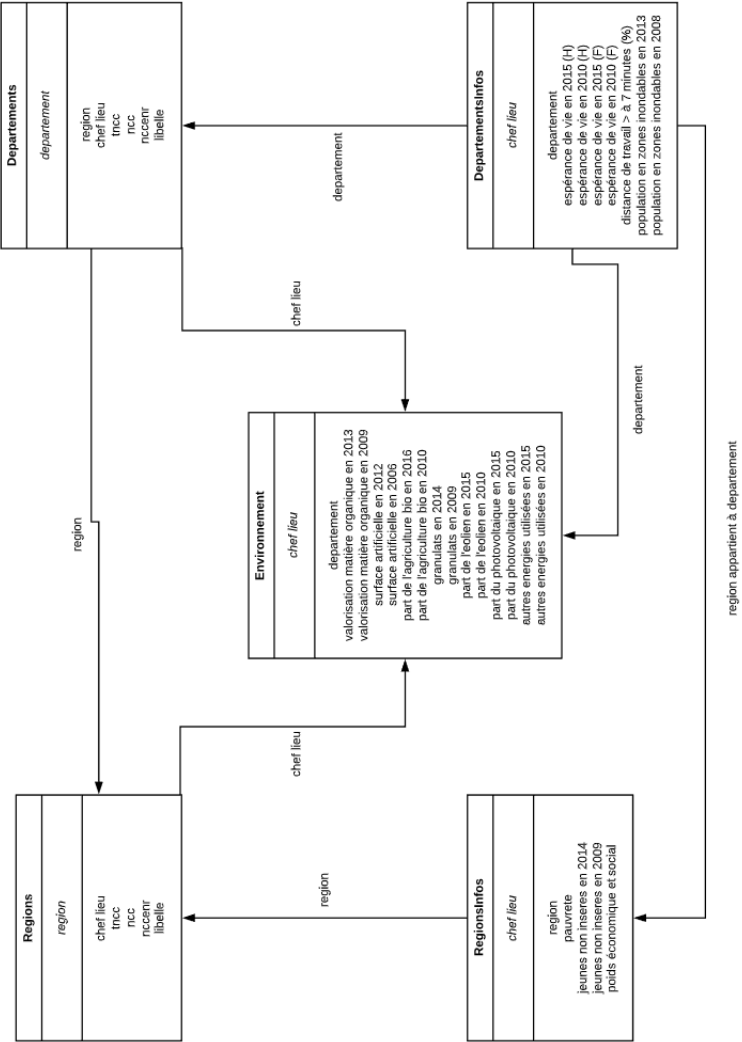


Figure 1 : Schéma de la base de données (noms des relations en gras et clés primaires en italique).

Partie 2

Réponses aux questions

Les réponses écrites ci-dessous sont bien évidemment dans le script de réponses **questionsBBD.py** présent dans l'archive fournie.

2.1 Première partie

1) **Liste des régions**

```
SELECT nccenr FROM Regions ;
```

2) **Liste des départements**

```
SELECT nccenr FROM Departements ;
```

3) **Données de la région saisie par l'utilisateur**

```
SELECT nccenr, cheflieu, tncc FROM Regions ;
```

4) **Données demandées (sociales ou environnementales) pour le département choisi par l'utilisateur**

```
SELECT nccenr FROM Departements ;
```

Puis, selon le choix :

Social

```
SELECT cheflieu, espH2015, espH2010, espF2015, espF2010, distravailsup7, inondable2013, inondable2008  
FROM DepartementsInfos WHERE dep = '%s'; (%s est remplacé par la saisie de l'utilisateur)
```

Environnemental

```
SELECT cheflieu, valorga2013, valorga2009, surfarti2012, surfarti2006, agribio2016, agribio2010, gra-  
nulats2014, granulats2009, eolienne2015, eolienne2010, photovoltaïque2015, photovoltaïque2010, au-  
trenergie2015, autrenergie2010 FROM Environnement WHERE dep = '%s'; (%s est remplacé par la  
saisie de l'utilisateur)
```

5) **Liste des départements où la part de l'énergie choisie a augmenté entre les deux années de référence, classés de la plus forte augmentation à la plus faible**

```
SELECT dep FROM environnement ORDER BY energieAprès/NULLIF(energieAvant,0) DESC ; (ener-
```

gieAprès pouvant être par exemple eolienne2015 et energieAvant pouvant être eolienne2010 (cf. script).)

2.2 Deuxième partie

1) **Départements dont la région a eu une production de granulats supérieure à 25 000 000 tonnes en 2014**

```
SELECT D1.nccenr FROM departements D1 WHERE D1.reg IN (SELECT R1.reg FROM departements D1 JOIN environnement E1 ON D1.nccenr = E1.dep JOIN regions R1 ON R1.reg = D1.reg GROUP BY R1.reg HAVING SUM(granulats2014) > 25000000);
```

2) **Afficher les 5 départements avec le plus grand taux d'énergie éolienne comme source de la puissance électrique en 2015**

```
SELECT dep FROM environnement ORDER BY eolienne2015 DESC LIMIT 5;
```

3) **Département ayant le plus faible taux de valorisation matière et organique en 2013**

```
SELECT dep FROM environnement WHERE valorga2013 != -1 ORDER BY valorga2013 ASC LIMIT 1;
```

4) **Part (en %) de l'agriculture biologique dans la surface agricole totale du département contenant le plus grand pourcentage de population éloignée de plus de 7 minutes des services de santé de proximité**

```
SELECT E1.dep, agribio2016, disSanteSup7 FROM environnement E1 JOIN departementsinfos D1 ON E1.dep = D1.dep WHERE disSanteSup7 != 1 ORDER BY disSanteSup7 DESC LIMIT 1;
```

5) **Taux de pauvreté en 2014 des régions dont la part des jeunes non insérés est supérieure à 30% en 2014**

```
SELECT reg, pauvrete FROM regionsinfos WHERE jeunesnoninseres2014 > 30 AND pauvrete != -1;
```

6) **Poids de l'économie sociale dans les emplois salariés de la région dont la source de la puissance électrique en énergies renouvelables provenait à au moins 10% de l'énergie photovoltaïque et dont la part de l'agriculture biologique dans la surface agricole totale était d'au moins 5% en 2015**

```
SELECT R1.nccenr, R2.poidsecosoc FROM environnement E1 JOIN departements D1 ON D1.nccenr = E1.dep JOIN regions R1 ON R1.reg = D1.reg JOIN regionsinfos R2 ON R1.nccenr = R2.reg GROUP BY R1.nccenr, R2.poidsecosoc HAVING (SUM(E1.photovoltaique2015)/COUNT(R2.reg)) > 10 AND (SUM(E1.agribio2016)/COUNT(R2.reg)) > 5;
```

Partie 3

Discussions, remarques et conclusion

Tout d'abord, concernant l'utilisation de la base de données, je l'ai créée manuellement via mon terminal (comme pour le TD Dragons) et je l'ai nommé `insee_bd`. Ensuite, j'ai entré mon identifiant et mon mot de passe (ligne 6 du fichier `creationInsee_BdB.py` et ligne 5 du fichier `questionsBDD.py`). Il faudra modifier ces champs (et éventuellement le nom de la base de données si besoin) pour pouvoir tester le fonctionnement de celle-ci.

Ensuite, une première remarque est le fait que, n'arrivant pas à prendre en charge les cases "nd", "nc" et les cases vides, j'ai par défaut mis -1 dans ces dernières. J'ai évidemment pris en compte ceci lors de certaines requêtes (*exemple : question 5 de la partie 2*).

En ce qui concerne la saisie des régions et départements par l'utilisateur, j'ai par défaut sélectionné le format "minuscule" (*exemple : Nouvelle-Aquitaine*). Ainsi, si l'utilisateur entre une région correcte en majuscule, même si elle existe, il y'a retour au menu. Notez qu'il faut mettre la première lettre en majuscule. On peut copier-coller un nom depuis la liste des régions pour se faciliter la tâche (j'affiche aussi aux questions 1 et 2 de la partie 1 tous les éléments en "minuscules")

Un autre problème que j'ai rencontré concerne le numéro des départements, qui sont pour la plupart uniquement des chiffres, a été de gérer les numéros pour la Corse. J'ai donc mis comme type de variable `varchar`.

Annexes

