

Dans cet exercice, vous allez effectuer vos **premières requêtes** de base de données.

Cet exercice est entièrement guidé.

Vous allez réaliser un document technique complet regroupant les résultats de chacune des étapes ci-dessous.

Suivez les étapes ci-dessous.

Étape 1 :

### Instructions

- Consultez les différents fichiers afin de comprendre vos données :
  - Données de [contrats clients](#)
  - Données du [référentiel géographique](#) des régions françaises (extrait du [site de data.gouv.fr](#))
- Remplissez le [dictionnaire des données](#) afin de faire correspondre :
  - le nom de chaque colonne entre le dictionnaire des données et les fichiers CSV ;
  - les **types de variables** (colonne type de données) et les **contraintes** (taille) que vous souhaitez ajouter dans votre modèle de données (exemple : la variable est un nombre alors c'est un integer, la variable ne contient qu'une lettre alors c'est un varchar de contrainte 1, etc.) ;
  - la **description des données** pour chaque ligne.

### Recommandations

- Il faut prendre le temps de **comprendre** les différentes données.
- Vous pouvez faire des **filtres** dans les données pour trouver les différentes valeurs ou des **tableaux croisés dynamiques** dans Excel.

### Résultat attendu

- Le dictionnaire des données explorées **complet**.

### Ressource :

Cours [Requêtez une base de données avec SQL](#)

Étape 2 :

## Instructions

- Prenez connaissance de l'ébauche du [schéma relationnel normalisé fait avec SQL Power Architects](#) (le schéma est en format image plus bas si besoin).
- Faites **correspondre** le schéma relationnel avec le dictionnaire des données directement dans le logiciel :
  - Ajoutez les **colonnes manquantes**.
  - Modifiez les **types de données**.
  - Ajoutez les **contraintes**.
- Générez le code SQL qui permet de créer la base de données directement dans SQL Power Architects.

## Recommandations

- Les variables présentes dans le dictionnaire des données doivent toutes être présentes dans le schéma relationnel normalisé.

## Ressources

- Lien de téléchargement vers [SQL Power Architects](#)
  - Démonstration de [SQL Power Architects](#)
- Si vous souhaitez utiliser un autre logiciel, voici l'ébauche du schéma en fichier [JPG](#).
- Vous pouvez suivre les cours suivants sur la création d'un schéma relationnel fonctionnel :
  - [Décomposez la modélisation de votre BDD en trois étapes clés – Modélez vos bases de données](#)
  - [BDD1 : modèle relationnel, clé, schéma relationnel](#)
  - Suivez cette partie de cours : [Améliorez votre modélisation grâce aux formes normales - Modélez vos bases de données](#)

## Résultats attendus

- **Schéma relationnel** (format .jpg)
- **Code SQL** générant les tables dans une base de données (format fichier texte)

## Étape 3 :

### Instructions

- Sélectionnez un système de gestion de base de données parmi les 3 proposés ci-dessous :
  - **SQLite** — le système de gestion de base de données le plus accessible
  - **MySQL** — le système le plus utilisé dans le monde
  - **PostgreSQL** — la version open source, mais avec toutes les contraintes d'utilisation d'une base de données (client et serveur)
- Installez le logiciel et faites les paramétrages nécessaires.
- Créez vos tables dans la base de données.
- Chargez toutes les données des deux fichiers dans le système de gestion de base de données (SGBD). Le nettoyage et le formatage des données sont effectués.
- Vérifiez qu'il y a **le bon nombre de lignes** entre le fichier CSV et le chargement dans la base de données.

### Ressources

- Consultez les ressources suivantes pour choisir et installer votre SGBD :
  - SQLite : [How To Download & Install SQLite Tools](#)
  - MySQL : [Installez le SGBD MySQL - Implémentez vos bases de données relationnelles avec SQL](#)
  - PostgreSQL : [PostgreSQL](#)
- Consultez votre mentor si besoin.

### Résultats attendus

- La **base de données** créée dans un SGBD avec les données chargées.
- La **capture d'écran de la BDD** montrant que les deux tables sont créées dans la base et que l'ensemble des données est présent (copie d'écran avec **le nombre de lignes** dans la table).

## Étape 4 :

### Instructions

- **Réalisez, sur un document de travail** de votre choix, les **trois analyses** demandées ci-dessous en suivant l'exemple.
  - Voici une [trame en format Word](#) que vous pouvez remplir.

### Recommandations

- Le but de ces analyses est de vous **initier** au langage SQL. Elles doivent vous permettre d'en **construire la logique**. Pour cela :
  - suivez le cours fourni en ressources ci-dessous ;
  - assurez-vous de lire attentivement l'exemple fourni dans le tableau ;
  - prenez le temps de construire votre réflexion sur la rédaction de vos premières requêtes **en identifiant d'abord de quelles variables vous avez besoin** ;
  - verbalisez, **expliquez votre cheminement logique** en détail à votre mentor de façon à ce qu'il puisse détecter une éventuelle compréhension erronée ou une lacune.

### Ressource

- [Requêtez une base de données avec SQL](#)

### Résultats attendus

- Le document rempli pour les trois premières analyses incluant :
  - le code SQL associé ;
  - la copie d'écran des résultats.

## Étape 5 :

### Instructions

- Faites le point avec votre mentor pour **vérifier** si vos 3 premières requêtes sont justes puis lancez-vous dans la suite des analyses.
- Réalisez les analyses suivantes à l'aide de requêtes SQL :
  - Requête 4 : Quels sont les 5 contrats qui ont les surfaces les plus élevées ?
  - Requête 5 : Quel est le prix moyen de la cotisation mensuelle ?
  - Requête 6 : Quel est le nombre de contrats pour chaque catégorie de prix de la valeur déclarée des biens ?

- Requête 7 : Quel est le nombre de formules “integral” sur la région Pays de la Loire ?
- Requête 8 : Lister les numéros de contrats avec le type de contrat et leur formule pour les maisons du département 71.
- Requête 9 : Quelle est la surface moyenne des contrats à Paris ?
- Requête 10 : Classement des 10 départements où le prix moyen de la cotisation est le plus élevé.
- Requête 11 : Liste des communes ayant eu au moins 150 contrats.
- Requête 12 : Quel est le nombre de contrats pour chaque région ?

## Recommandations

- Faites l'effort de **redire à haute voix** ce que vous avez compris.
- Cherchez ensuite le résultat par vous-même.

## Ressources

- Cette partie de cours : [Affichez les données pertinentes avec SELECT - Requêtez une base de données avec SQL](#)

## Résultat attendu

- L'ensemble des requêtes complétées dans le document, avec pour chaque requête :
  - le code SQL associé ;
  - la copie d'écran des résultats.

Étape 6 :

## Instructions

- Rédigez un support de présentation de votre méthodologie afin de prendre du recul sur ce que vous avez appris dans ce projet.
- Détaillez les requêtes et les résultats.

## Recommandations

- La présentation doit être claire, convaincante et professionnelle.
  - Elle doit être synthétique (autour de 10 slides maximum).
  - Elle doit avoir moins de 7 éléments par slide.
- Elle doit présenter la méthodologie en quelques slides avec :
  - le dictionnaire des données ;
  - le schéma relationnel ;
  - la base de données ;

- les requêtes SQL avec les résultats.

## Ressources

- Si vous êtes sur le parcours complet de Data Analyst, inspirez-vous du modèle réalisé sur le projet précédent.

## Résultats attendus

- Le support de présentation.

## Livrables

Envoyez à votre mentor tous vos livrables intermédiaires dans un seul dossier :

1. Le **document technique** de 12 pages maximum au format **pdf** avec :
  - l'exploration des types de données avec la **capture d'écran** du **dictionnaire des données** ;
  - le **schéma** relationnel modifié de la base de données ;
  - le **code SQL** générant les tables ;
  - la **capture d'écran** de votre **base de données** chargée.
2. La **liste** complète des analyses avec les **résultats associés**.
3. La **présentation** de la **méthodologie** suivie présentant les étapes et le rappel des résultats au format **pdf**.

Vous aurez ensuite un temps d'échange avec votre mentor pour comparer les évaluations respectives de ces livrables.