

## SAÉ S2-04

### Exploitation d'une base de données

#### Par groupe de 3 ou 4

**Vous devez envoyer la composition de votre groupe par mail à votre enseignant de bases de données lors de la première séance.**

► **Sujet.** Dans cette SAÉ, vous allez devoir manipuler des données ouvertes et libres. Vous devez choisir un jeu de données sur internet (<https://www.data.gouv.fr/fr/> ou <https://www.kaggle.com/datasets>). Vous êtes libres de choisir le jeu de données que vous voulez, mais vous devez faire valider votre choix par votre enseignant de bases de données.

L'objectif est d'analyser et de structurer ces données afin d'en tirer des graphiques et des statistiques.

► **Travail demandé.** Dans un premier temps, vous devez lister les questions que vous vous posez sur ces données, ainsi que les statistiques que vous souhaitez extraire. Cela peut-être par exemple dans le cadre d'une étude de marché pour la mise en place d'un nouveau produit ou d'une nouvelle entreprise.

Vous devrez ensuite proposer un modèle permettant de structurer vos données. Nous vous encourageons à choisir un jeu de données qui requièrent plusieurs tables ou à fusionner plusieurs jeux de données. Écrivez un script permettant de créer la base de données et d'y importer les données collectées. Selon les données choisies, il sera peut-être nécessaire de les traiter au préalable (pour supprimer les doublons par exemple).

Enfin, vous devrez analyser vos données et présenter sous forme visuelle les statistiques obtenues permettant de répondre aux questions identifiées dans la première partie.

Cette SAÉ s'appuie sur des notions vues en R2.06 (Exploitation d'une base de données), R2.08 (Outils numériques pour les statistiques descriptives), R2.10 (Gestion de projet et des organisations) et R2.12 (Anglais S2). Nous vous encourageons néanmoins à aller plus loin en explorant la documentation des packages Python (`pandas`, `matplotlib`, ...).

► **Livrables.** À l'issue de cette SAÉ, les différents livrables que vous devez fournir sont :

— Un rapport au format PDF contenant :

1. Une présentation du/des jeu(x) de données : Que contient votre jeu de données ? Quels types de données sont à l'intérieur ? Quel pré-traitement avez-vous réalisé (suppression des doublons, transformation de valeurs, ...). Fournissez notamment un lien vers le jeu de données en ligne.
2. Un modèle de données (MCD ou MLD).

3. Une analyse des données : Quelles informations avez-vous extraits de ces données ? Cette partie doit s'appuyer sur des graphiques permettant de visualiser les données.

Ne vous contentez pas de décrire chaque graphique que vous avez réalisé. Prenez du recul et indiquez les informations globales que vous avez obtenu grâce à votre analyse.

- Le script Python permettant de créer votre base de données à partir du/des fichier(s) du jeu de données.
- Le script Python permettant de générer les statistiques et les graphiques.

► **Oral.** En plus des livrables, vous serez évalués lors d'un oral. Lors de cette présentation en groupe, vous devrez expliquer votre analyse du jeu de données en décrivant les statistiques et les graphiques produits pendant la SAÉ. Cet oral sera en anglais et durera 10 min. Il sera suivi par des questions. Bien entendu, les membres du groupe doivent se partager le temps de parole de façon équitable.

► **Barème de notation.** Vous serez évalué sur les différents éléments produits selon le barème suivant :

- Rapport, partie 1 (présentation) : 1 points
- Rapport, partie 2 (modèle de données) : 4 points
- Rapport, partie 3 (analyse) : 6 points
- Script de création : 3 points
- Script de visualisation : 4 points
- Exposé en anglais : 2 points

► **Modalités de rendu.** Les livrables seront à rendre avant le mercredi 17 mai à minuit sur Moodle (un seul dépôt par groupe) sous forme d'archive compressée (**.zip** ou **.tar.gz**). L'oral aura lieu lors du dernier TP de bases de données.

► **Compétence et apprentissages critiques concernés :**

Compétence 4 : Concevoir et mettre en place une base de données à partir d'un cahier des charges client

Apprentissages critiques :

- AC1 : Mettre à jour et interroger une base de données relationnelle
- AC2 : Visualiser des données
- AC3 : Concevoir une base de données relationnelle à partir d'un cahier des charges