

BUT SD2 - SAE 4.02

REPORTING D'UNE ANALYSE MULTIVARIÉE

A. ROLLAND & A. GUILLE, IUT LUMIERE LYON II

BUT SD

OBJECTIFS PÉDAGOGIQUES DE LA SAÉ

- Approfondir pour l'étudiant les enjeux du reporting en l'intégrant dans un processus décisionnel
- Favoriser la compréhension et l'explication de données métiers par leur description et leur classification au travers d'un choix argumenté d'outils d'analyse et de restitution.
- Accompagner l'étudiant dans le choix d'une méthode statistique adaptée et présenter les étapes de préparation des données en vue d'une analyse exploratoire multivariée

STRUCTURE DE LA SAÉ

- un projet en groupes aléatoires de 3 étudiant.e.s mixé VCOD et EMS
- une première phase “analyse multivariée” exploratoire
- une deuxième phase “automatisation” de ou des analyses précédentes
- avec un point formel chaque semaine **en présence à l'IUT**. Premier point ce 15 avril à 15h30.

EVALUATION DE LA SAÉ

- accompagnement du projet par Antoine Rolland
- évaluation par Adrien Guille
 - un 4 pages par groupe, avec problématique, analyses multifactorielles et conclusion sur les phénomènes observés, mais rien de technique, à rendre le 19 juin.
 - une démonstration lors d'une soutenance le 20 juin, avec présentation d'un cas d'usage et commentaire en direct des résultats, avec le choix entre R+Shiny et Python+Streamlit (mais pas PowerBI) pour le développement.

SUJET DE LA SAÉ

typologie des communes de France en fonction des équipements disponibles

- On appelle "équipement" des endroits accessibles au public sur un territoire donné : des bâtiments culturels (musées, cinémas, théâtres, . . .), sportifs (stades, piscines, . . .), des commerces, des écoles, des lieux de santé (médecins, hôpitaux, . . .). L'idée sous-jacente est d'arriver à décrire des types de communes suivant la facilité d'accès à ces équipements pour les habitants.
- Les données proviennent du site de l'INSEE, et sont mises à jour chaque année. On peut donc imaginer une automatisation des traitements pour procéder à des analyses annuelles ; Note : on ne tiendra pas compte des problèmes particuliers de fusion de communes.

SUJET RESSERRÉ DE LA SAÉ

typologie des communes de France en fonction des équipements disponibles

On s'intéresse en particulier aux 550 (environ) communes de 10.000 à 20.000 habitants. La question que l'on cherche à traiter est d'illustrer les différences (ou l'absence de différences) d'équipements entre différents types de communes (par exemple rurales, urbaines et péri-urbaines, ou par région, ou autre) de même taille.

DONNÉES

Vous avez à disposition les fichiers de données concernant :

- la liste des communes avec leur département et leur population totale :
`grille_densite_7_niveaux_2024.xlsx`. Chaque commune est identifiée par un code unique. En outre, chaque commune est caractérisée par un qualificatif
Commencer par sélectionner les communes de 10.000 à 20.000 habitants.
- la liste des équipements par commune, dans le dossier :
`Ensemble-com-2021`
- la liste des modalités possibles des variables
équipements : `BPE_gammes_2021_internet_v2.xlsx`
et `BPE21_Liste_hierarchisee_equipements.pdf`

AIDE AU PROJET

1. choisir une problématique / un angle de vue :
2. construire votre propre base de données. En effectuant des sélections et jointures entre les tables disponibles, vous allez construire une base de données 1) répondant à votre problématique et 2) de nombre de variables raisonnable (nota : le processus de réduction du nombre de variables peut se faire de manière externe - par connaissance du domaine - ou interne - à travers les résultats itérés d'une ACP).
3. procéder à des analyses factorielles et les interpréter en fonction de la problématique initiale. Vous pourrez éventuellement prolonger par un clustering si cela est cohérent avec votre analyse.
4. automatiser ce qui doit l'être, sur *R Shiny* ou *Python*, pour faciliter l'utilisation de ce reporting par l'utilisateur final.