# Session du 02/03/2023

PySpark

Spotfire

SQL -> Etre capable de répondre en direct a des questions

Ne pas s’attacher au language, il faut s’attacher à la technique

RIGEUR

PROJET 9

1. Receuille de données : Méthode PESTEL.
   1. Choisir une dizaine de variable pertinente Qui ne sont pas corrélé entre elles dans une année intéressante. dans le choix de l’exportation (pays). -> PAS US, Norvège. Ne pas choisir les pays riches. A moi de trouver des axes discriminants.
   2. Après une première itération, rajouter/modifier/supprimer des variables pour améliorer le modèle.
   3. Se renseigner sur le Scalaire. (robus scalaire, standard scalaire). Essayer de les comprendre.

Itération 1 (Pour une année)

* Choix des variables
* Nettoyage des données
* Analyse Univariée
* Analyse Bivarié (heatmap, test)
* Scaleur
* Kmeans pour avoir des groupes
* ACP pour comprendre les groupes + Boxplot

Itération 2 : On recommence encore et encore

Le but est d’obtenir un modèle qui permet d’avoir un groupe bien défini qui est interessant pour l’entreprise. Le but est d’avoir une fiche d’identité claire par pays.

Par la suite, le faire pour plusieurs années.

Pour la semaine pro, choisir des variables, les nettoyer et les comprendre.

Comprendre comment le Kmine et le dendrogramme font pour créer des groupes. Comprendre comment choisir le nombre de groupe. Comprendre les scaleurs. Comprendre l’ACP et tous ces outils.

Faire un avant après visuel des scaleurs.