

---

## Projet d'Analyse de Données

---

À vous de mener en **autonomie** un projet d'analyse de données !

### Option n°1 : Style de musique et popularité

Le but de ce projet est d'étudier la base de données disponibles sur **kaggle** : 30000 Spotify Songs

`kaggle.com/datasets/joebeachcapital/30000-spotify-songs`

Vous êtes libres de vos questionnements, une attention particulière sera portée lors de l'évaluation sur la motivation de ceux-ci.

Pour vous aider, quelques questions que vous pouvez *par exemple*<sup>1</sup> vous poser :

- Qualité du jeu de données proposé ? Un pre-processing des données est-il nécessaire ?
- Comment se comportent les différentes variables en fonction des années, du style de musique ?
- Plus généralement : Comment se comportent les différentes variables entre elles ?
- La popularité des morceaux dépend-elle de leur style ?
- Comment les différentes mesures influent-elles sur l'appréciation du morceau considéré ?
- Les différentes mesures permettent-elles de retrouver les genre musicaux ? Autre chose ?

On pourra, au besoin, créer une variable catégorielle encodant la popularité à partir du score de popularité disponible. Le nombre de classes ainsi que le choix des seuils devront être motivés.

---

### Option n°2 : Le jeu de données de votre choix

Vous pouvez également choisir d'étudier un autre jeu de données pour ce projet.

- Votre jeu de données devra cependant être suffisamment complexe pour vous permettre de réinvestir la plupart des méthodes vues en cours ce semestre.
- Des exemples de sites pour trouver des données :
  - **kaggle** : `kaggle.com`
  - **Hugging Face** : `huggingface.co`
  - **Our World in Data** : `ourworldindata.org`
  - ...

Merci dans ce cas de valider votre choix auprès de l'équipe enseignante **a priori**.

---

**Remarque importante** : Il n'est pas interdit de réinvestir des méthodes vues dans d'autres cours (EMS et/ou Machine Learning typiquement). Ces méthodes ne doivent cependant pas se substituer à celles vues en cours d'Analyse de Données.

---

1. En particulier, il n'est pas requis de répondre à toutes ces questions