Projet

Octobre, 2021

But : Le but du projet est d'étudier et de décrire un ensemble de données en appliquant des outils rencontrés en TD et également de vous faire chercher un peu au-delà.

Une partie du projet consiste à exécuter des tâches précises (détaillées ci-dessous). Une deuxième partie (**importante !**) du projet est dédiée à la **description et l'analyse des données de la manière que vous trouverez pertinente.**

Consignes générales Le projet doit être rédigé en utilisant R Markdown. La rédaction, les explications, l'interprétation des résultats et la clarté du code seront prises en compte dans la notation. De plus :

- Le rendu final doit être : un rapport sous forme html ou pdf et son fichier source .Rmd. Les titres des fichiers seront de la forme `Nom1_Nom2_Nom3...`.
- La modification externe (sans l'utilisation de R) des fichiers de données est interdite.
- Dans le cas où des `packages` supplémentaires sont utilisés ils doivent être chargés dans un même chunck au début.
- Pour la deuxième partie, les choix d'étude devront être **justifiés**. En particulier, dans le cas où des méthodes statistiques seraient appliquées, inclure également une description sommaire de la méthode.

Données - le fichier `albums.csv` contentient *Rolling Stone's 500 Greatest Albums of All Time* (sorti en 2012). La description du fichier peut être trouvée ici.

- le fichier `tracks.csv` contenient des informations sur différentes chansons. Ce fichier a été proposé par *Machinehack* comme jeu d'entrainement pour un défi de classification. Une description sommaire du fichier se trouve ici.

Greatest Albums

1. Représenter graphiquement le nombre d'albums inclus dans le top par année. Commenter les résultats.
2. Afficher le nom et le nombre d'albums pour les artistes ayant au moins 5 albums dans le top. L'affichage se fera par ordre décroissant de nombre d'albums.
3. Quel est le genre le plus répresenté dans le Top 500 de Rolling Stones?

Chansons

Importer le fichier dans un tableau nommé `tracks`.

1. Faire un top des 1000 chansons les plus dansables (variable `danceability`). En cas d'égalité, départager les chansons en fonction de leur popularité.
    - Quels sont les 9 artistes les plus représentés ?
    - Et la classe (variable `Class`) la plus représentée ? Illustrer la répartition des classes dans ce top.
2. Représenter graphiquement les variables `Popularity` et `danceability` par classe.
3. Trouver le(s) artiste(s) (`Artist Name`) ayant la popularité moyenne la plus elevée parmi ceux qui ont plus de 10 titres inclus dans `tracks`.

4. Enlever les lignes contenant des données manquantes. Répartir les chansons en fonction de leur popularité en quatre catégories de tailles égales (ou presque) nommées : `TresPopulaire`, `Populaire`, `PeuPopulaire`, `PasPopulaire` et ajouter une nouvelle colonne au tableau `tracks` contenant la catégorie de chaque chanson.
    - Quelle est la classe la plus représentée dans la categorie `TresPopulaire` ? Commenter les résultats.
    - Parmi les artistes ayant un nombre de titres supérieur ou égal à la mediane du nombre de titres de chaque artiste de `tracks`, lequel a la plus grande proportion de ses chansons dans la catégorie `TresPopulaire` ?

Chansons et albums

1. Pour le(s) artiste(s) ayant le plus grand nombre d'albums dans le Top 500 de Rolling Stones, représenter par des `boxplots` la popularité de leurs chansons contenues dans `tracks`. Commenter les résultats.
2. Créer un tableau avec les artistes ayant au moins un album dans le Top 500. Pour chaque artiste le tableau contiendra également le nombre d'albums de l'artiste dans le Top 500 ainsi que la position de l'album le mieux classé. Pour ceux qui ont au moins une chanson dans le tableau `tracks` ajouter le nombre de ses chansons apparaissant dans `tracks`, leur popularité moyenne, la dansabilité moyenne, la durée minimale et maximale des chansons, ainsi que la classe dominante (celle attribuée au plus de titres).
3. Pour chaque `Class` apparaissant comme classe dominante dans le tableau précédent, afficher le nombre d'artistes pour lesquels cette classe dominante a été attribuée.