

Top 50 des chansons sur Spotify en 2019

ANALYSE ET FOUILLE DES DONNÉES

Rédigez un rapport documenté : vous expliquerez votre démarche et interprétez vos résultats pour chaque question. Vous remettrez votre travail sur le devoir Moodle prévu à cet effet.

1. Créez-vous un compte Kaggle : <https://www.kaggle.com/>
Prenez le fichier <https://www.kaggle.com/datasets/leonardopena/top50spotify2019>
Lisez les informations sur le fichier top 50 de Spotify ;
2. Ecoutez la vidéo qui explique comment transformer un fichier csv en fichier Excel ;
3. Détaillez chaque variable en donnant leur nom et leur signification ;
4. Repérez, s'il y en a, les données manquantes et les données aberrantes et traitez-les.
5. Quels sont les genres musicaux les plus représentés dans le top 50 en 2019 ? Illustrez ;
6. Quelle est la durée moyenne (en minutes) des chansons dans le top 50 en 2019?
7. Quelle est la durée médiane (en minutes) des chansons du top 50 en 2019?
8. Sur base d'un graphique en nuage de points, semble-t-il exister une corrélation entre l'énergie et la danceability en 2019?
9. Comment se distribue la popularité des chansons dans le top 50 en 2019?
10. Quelles sont les chansons les plus longues et les plus courtes du top 50 en 2019?
11. Quels sont les artistes les plus fréquents dans le top 50 en 2019?
12. Comment se distribue la danceability des chansons en 2019?
13. A vous de trouver une question pertinente à creuser.

Bon travail.