

Contrôle qualité du fichier qui a été utilisé comme base pour mon jeu de données.

Le lien : [Charts des top chansons au Brésil à la semaine du 25 janvier 2024](#)

Point de contrôle qualité	Problème observé et commentaire	Correction : comment ? importance ?
1. le jeu de données est difficilement accessible (format "image", PDF, HTML...), le fichier est mal formé	Le fichier est très accessible, il est en format csv donc lisible pour la majorité des logiciels. Tu le télécharges sur la page d'un classement et le bouton est très visible puisqu'avant l'affichage du classement.	Je n'ai rien à dire de plus. Le format csv est adapté à la manipulation de données.
2. la licence est absente ou inhabituelle, le jeu de données n'est pas "open data"	Pour avoir accès au fichier, il faut posséder un compte Spotify (premium ou pas). Ensuite aller sur la page Spotify Chart et choisir un chart et on peut le télécharger directement.	Malheureusement les données sont accessibles que si l'on possède un compte Spotify, ce qui ne rend pas les jeux de données très accessibles pour ceux n'ayant pas de compte.
3. le fichier fait peu appel aux standards répandu et aux données pivots	Rien à signaler, le fichier est en format csv et les données sont bien répertoriées. Par exemple on peut directement le traiter sur Excel pour en faire des tableaux dynamiques.	Ca aurait été bien d'avoir plus de format mais le csv convient déjà bien.
4. le fichier est mal documenté	Pour moi toutes les données à ce qu'on attend d'un classement est là et sont claires.	
5. il existe des problèmes de syntaxe	Rien à signaler.	

Point de contrôle qualité	Problème observé et commentaire	Correction : comment ? importance ?
6. Valeurs aberrantes, suspectes, inexplicables, pas crédibles	Rien à signaler, les données sont vérifiables directement sur la page du site au pire s'il on veut vérifier les cohérences.	
7. Il manque des données et cela n'est pas documenté (trous, données tronquées, valeurs vides, granularité / fréquence / maillage / fraîcheur)	Rien à signaler.	
8. Trop de données : doublons, inutilement vieilles, précision / fréquence / maillage / fraîcheur	Rien à signaler.	
9. Données posant problème avec la réglementation (données perso, relatives à la santé, la religion..., propriété littéraire et artistique, etc.)	Non rien à signaler.	
10. Les contenus posent problèmes : synonymies, non traduits (USA), cryptique (DAECP), utilisation du 0 au lieu du "null"...	Rien à dire.	