

Description du projet :

1- Description des utilisateurs :

Les utilisateurs sont des analystes de données, des passionnés de musique, des chercheurs, ou toute personne intéressée par l'étude des tendances dans les paroles de chansons. Ils cherchent à identifier les mots les plus fréquemment utilisés dans des genres musicaux spécifiques, à comparer l'évolution des thèmes et des styles au fil des années, et à explorer les caractéristiques linguistiques des paroles selon le genre musical et l'année de sortie.

2- Liste des tâches visuelles supportées par les utilisateurs et objectifs de visualisation :

Tâches visuelles :

- Identifier les mots les plus courants dans les paroles d'un genre musical spécifique pour une année donnée.
- Comparer les mots-clés dominants entre différentes années.
- Observer l'évolution des thèmes récurrents dans les paroles des chansons d'un même genre au fil du temps.
- Effectuer des comparaisons de popularité de certains mots dans différents genres musicaux.

Objectifs de visualisation :

- Fournir une vue claire et interactive des mots les plus fréquents dans les paroles d'un genre spécifique pour une année donnée.
- Permettre un filtrage par genre et par année pour explorer les différences dans l'utilisation des mots.
- Mettre en évidence les mots-clés dominants grâce à des tailles de police proportionnelles à leur fréquence d'apparition dans le corpus analysé.

3- Liste des attributs (bruts) nécessaires du dataset WASABI :

Pour créer le word cloud, les attributs suivants du dataset WASABI seront nécessaires :

- `_id` : (pour suivre les chansons)
- `publicationDateAlbum` : (pour filtrer les chansons par année)
- `genre` : (pour classer les chansons dans des genres spécifiques)
- `lyrics` : (les paroles des musiques)

4- Description informelle du traitement des données brutes :

Le traitement des données du dataset WASABI pour qu'elles conviennent à la visualisation comprend les étapes suivantes :

- Filtrage par année et genre : L'utilisateur sélectionne une année et un genre, et seules les chansons correspondant à ces critères sont retenues pour l'analyse.
- Extraction des mots : Les paroles des chansons sélectionnées sont analysées pour extraire les mots.
- Nettoyage des données : Les mots inutiles tels que les articles, prépositions et mots courants non pertinents (stop words) sont exclus de l'analyse.
- Comptage de la fréquence : Chaque mot est compté pour déterminer sa fréquence d'apparition.
- Création du nuage de mots : Les mots sont représentés visuellement avec une taille proportionnelle à leur fréquence d'apparition.

5- Technique de visualisation et membre du groupe :

Technique de visualisation : WordCloud (heatmap)

Objectif visuel : Le nuage de mots permet de visualiser rapidement quels termes apparaissent le plus fréquemment dans les paroles d'un genre musical et d'une année donnée. Plus un mot est fréquent, plus il est affiché en grand dans le nuage.

Membre du groupe :

- Loris Drid
- Sayf Eddine Halmi
- Timothee Juillet
- Guillaume Arrigoni
- Skander Meziou

6- Cartographie visuelle des variables :

Attributs des données après traitement :

- Année : L'année sélectionnée par l'utilisateur.
- Genre : Le genre musical choisi pour l'analyse.
- Mots : Les mots extraits des paroles.

- Fréquence des mots : Nombre d'occurrences de chaque mot.

Variables visuelles dans le diagramme en secteurs :

Variables visuelles dans le nuage de mots :

- Texte des mots : Chaque mot du nuage représente un terme des paroles analysées.
- Taille du texte : Proportionnelle à la fréquence du mot dans le corpus analysé.
- Couleur : Utile pour différencier les mots différents entre eux

Ce nuage de mots offre une visualisation engageante et informative, permettant aux utilisateurs de discerner rapidement les thèmes et expressions dominantes dans les paroles des chansons selon le genre et l'année sélectionnés.