

Rapport sur l'Application Qlik pour l'Analyse du Streaming de Musique dans le Monde via Spotify

CISSOKO, Houmana
CHIPAN, Lucas

1. Objectif de l'Application :

L'objectif principal de notre application Qlik est d'analyser les données de musique dans le monde, plus particulièrement sur Spotify. Nous cherchons à répondre à plusieurs questions clés, notamment :

- Quelles sont les chansons les plus populaires dans différents pays au fil du temps ?
- Comment la popularité des artistes varie-t-elle dans différentes régions?
- Existe-t-il des tendances de streaming spécifiques à certaines villes ou pays?

Nous comptons répondre à ces questions en utilisant des graphiques spécifiques, tels que des graphiques de popularité des chansons, des cartes géographiques des tendances de streaming, et des analyses temporelles de la popularité des artistes.

2. Jeu de Données :

Nous avons utilisé plusieurs sources de données pour notre analyse :

- [Fichier plat venant de kaggle pour les informations à propos des pays et des villes](#)
- [Fichier plat venant de kaggle pour la musique](#) : Un fichier plat contenant des informations sur différentes pistes provenant de Spotify. Ces données comprennent des détails tels que l'ID de la piste, le nom, la durée, l'ID de l'album, la date de sortie, et la popularité. Mais qui contient aussi quelques données provenant de différentes plateformes telles que Deezer et Apple Music.
- [Spotify Web Api](#) : Initialement, nous avons utilisé l'API Spotify pour obtenir des données sur les données des musiques et des albums.
Cependant, en raison de problèmes tels que la restriction (rate limit) de Spotify, nous avons rencontré des limitations dans la quantité de données que nous pouvions récupérer. En conséquence, nous avons complété nos données en utilisant le fichier plat de Spotify.
- [Spotify Charts](#) : Le site de spotify qui nous permet d'avoir le classement des titres en fonction du pays, de la date et même de la ville. Ce site à été utilisé pour récupérer les informations principales à propos des titres et Spotify Web Api à été utilisé pour compléter.

Les données provenant du fichier plat "*world cities*" et de l'API ont été stockées séparément dans nos bases de données.

Nous aurions pu refaire les recherches en utilisant l'API Spotify pour les stocker sous le même format que le fichier plat mais nous avons décidé de ne pas le faire.

Cette décision a été prise pour éviter de refaire toutes les requêtes à l'API Spotify après

avoir rencontré des problèmes de rate limit, assurant ainsi que nous avons une base de données complète malgré ces limitations.

Nous avons aussi trouvé une méthode pour obtenir des classements Spotify basés sur les villes mais étant déjà face à des problèmes liés au rate limit nous n'avons pas utilisé ces données. La méthode utilisée pour extraire ces données est documentée dans notre fichier Python, accessible pour consultation.

Les différents fichiers ont été fournis, pour lancer celui de l'API spotify il vous faudra mettre votre CLIENT_ID et CLIENT_SECRET récupérables sur le [dashboard de l'API](#)

Dashboard > projetQlik > Settings > Basic Information

P Basic Information

Basic Information User Management Extension Requests

Client ID

d8697567e29b4147afc35e4a558f96f

App Status

Development mode

[View client secret](#)

Puis il faudra les fournir dans le fichier .env

Sans cela vous ne pourrez pas générer le token nécessaire pour accéder à l'API Spotify.

Il faudra également fournir le Bearer Token récupérable depuis le navigateur

▼ General

Request URL:

https://charts-spotify-com-service.spotify.com/auth/v1/overview/GLOBAL

Request Method:

GET

Status Code:

● 200 OK

Remote Address:

[2600:1901:1:7c5::]:443

Referrer Policy:

strict-origin-when-cross-origin

► Response Headers (14)

▼ Request Headers

:authority:

charts-spotify-com-service.spotify.com

:method:

GET

:path:

/auth/v1/overview/GLOBAL

:scheme:

https

Accept:

application/json

Accept-Encoding:

gzip, deflate, br, zstd

Accept-Language:

fr-FR,fr;q=0.9,en-US;q=0.8,en;q=0.7

App-Platform:

Browser

Authorization:

Bearer BQDQ-

3. Détails du Jeu de Données :

Pour la version utilisant l'API Spotify nous avons sauvegardé les informations dans différentes tables SQL.

Les tables principales comprennent des informations détaillées sur les pistes, les classements quotidiens, les artistes, les albums, et les relations entre pistes et artistes. Ces tables sont liées par des clés primaires et étrangères, permettant une exploration complète des relations entre les données.

Les données sur les pistes (tracks) contiennent des informations sur les chansons telles que leur durée, leur popularité, et la date de sortie.

Les données de classement quotidien (dailyrankings) nous donnent des informations sur la quantité de streaming pour chaque piste dans différents pays à différentes dates.

Les données sur les artistes (artists) contiennent simplement l'identifiant et le nom de l'artiste. Nous avons aussi une table (tracks_artists) qui fait la liaison entre le(s) artiste(s) et la chanson.

Les données sur les albums (albums) comprennent des informations sur les albums, y compris le type, le nombre total de pistes, et la durée.

Les données sur la relation entre les pistes et les artistes (tracks_artists) nous permettent de comprendre les collaborations artistiques.

Pour la version utilisant le fichier plat nous avons laissé le fichier CSV tel qu'il était.

4. Modélisation de Données :

Nous avons modélisé les données en utilisant un modèle en flocon de neige pour permettre une normalisation efficace tout en maintenant des relations claires entre les entités. Ce choix de modèle facilite l'analyse multidimensionnelle et l'exploration des données.

Mais également un modèle en étoile, ce choix de modèle dans le contexte du streaming musical est idéal pour l'agrégation de mesures telles que le nombre de streams. Ce modèle simplifie l'analyse en permettant aux utilisateurs de se concentrer facilement sur des dimensions clés comme le pays, les artistes, etc. Il offre également une meilleure maintenabilité en facilitant les modifications et les mises à jour. En résumé, le modèle en étoile est une solution efficace pour répondre aux besoins d'analyse dans le domaine du streaming musical ce qui a motivé notre choix d'utiliser ce modèle..

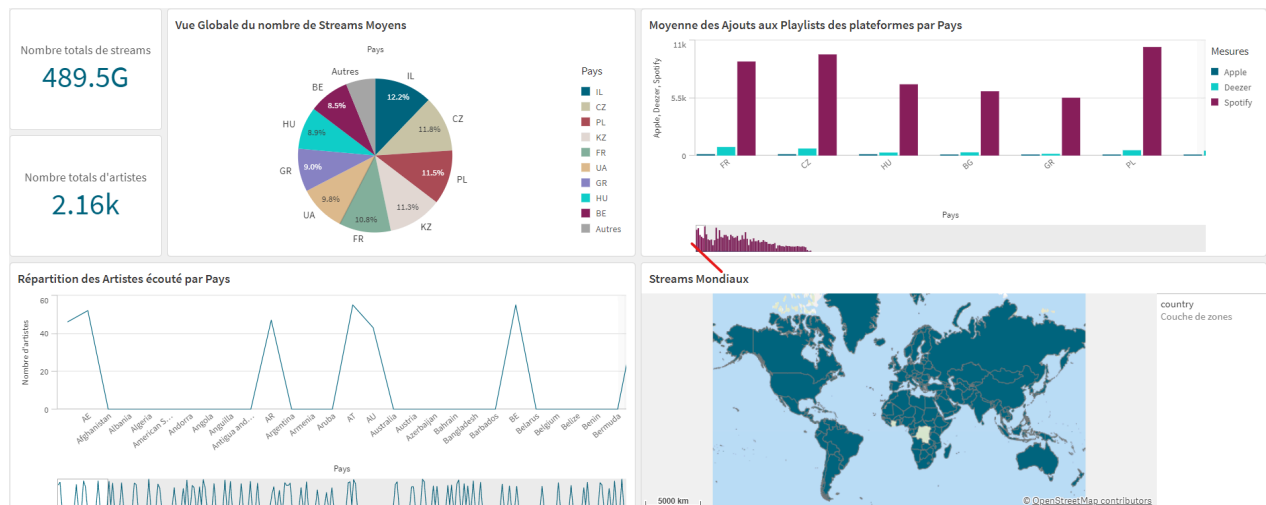
5. Considération du Modèle Multidimensionnel :

Dans Qlik Sense, nous avons mis en œuvre le modèle multidimensionnel en utilisant les fonctionnalités de modélisation de Qlik Sense. Les dimensions incluent les pistes, les artistes, les pays, et les dates, tandis que les mesures comprennent la popularité, la quantité de streaming...

Ne voulant pas gâcher les données récupérées via API nous avons décidé de faire une feuille avec chacun des jeux de données.

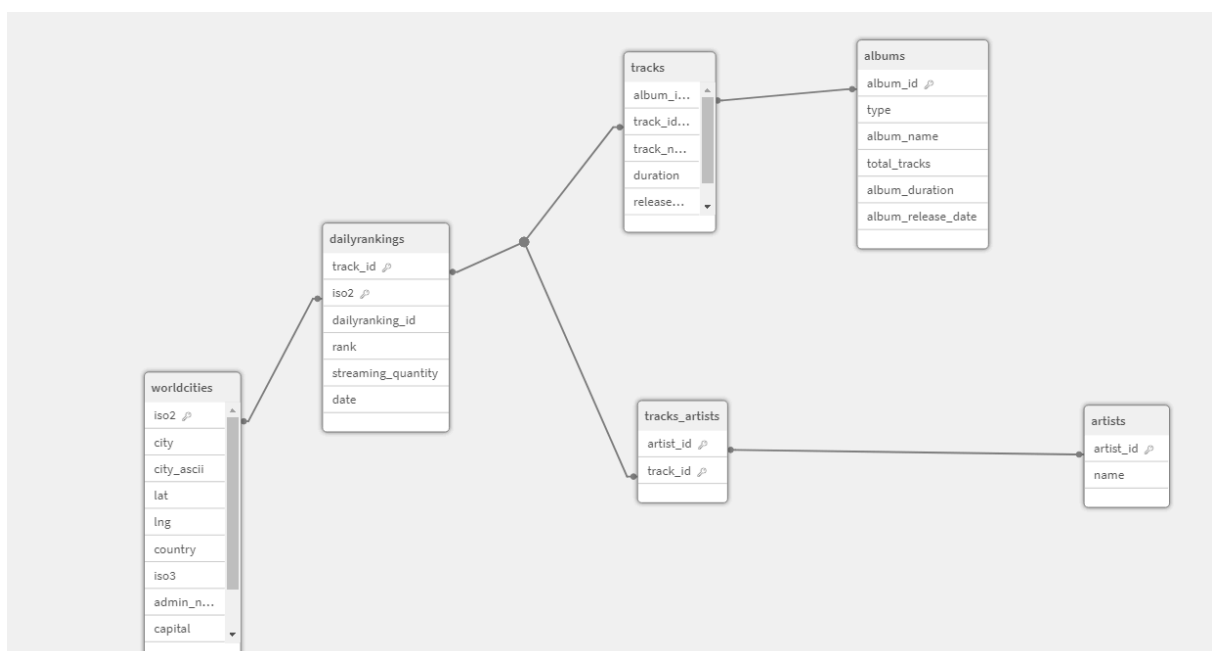
Le 1er jeu de données a été fait avec un modèle en étoile tandis que le second a été fait avec un modèle en flocon

Capture d'écran du modèle de données (avec fichier plat):

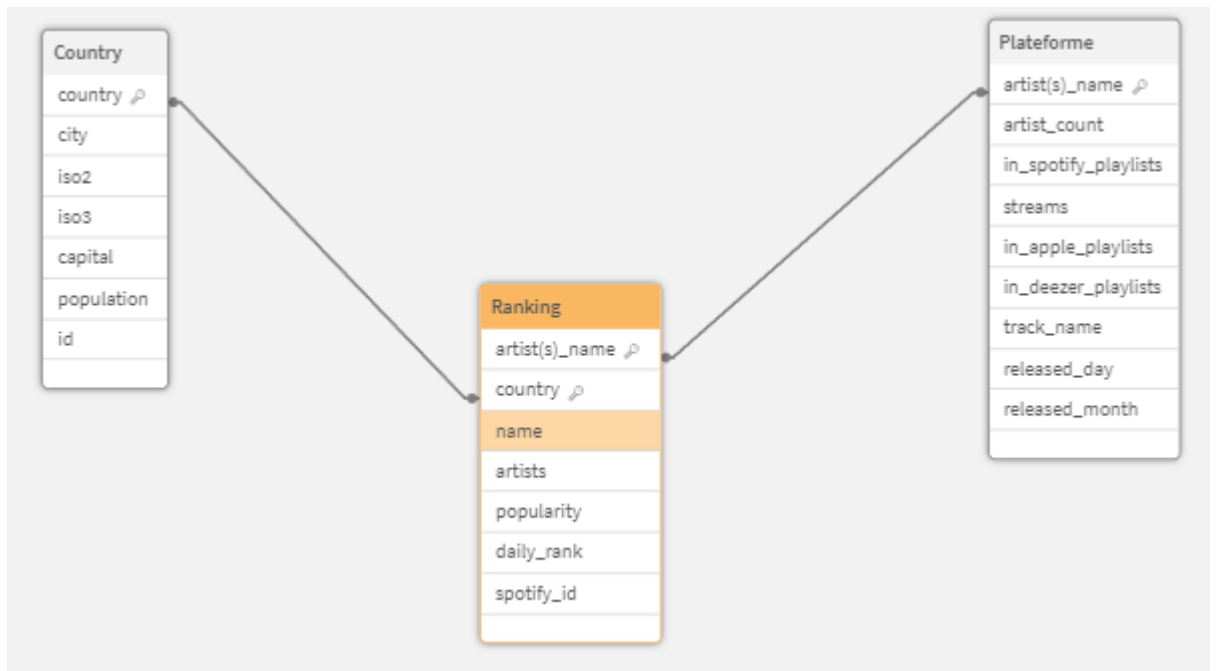


Voici nos différents modèles

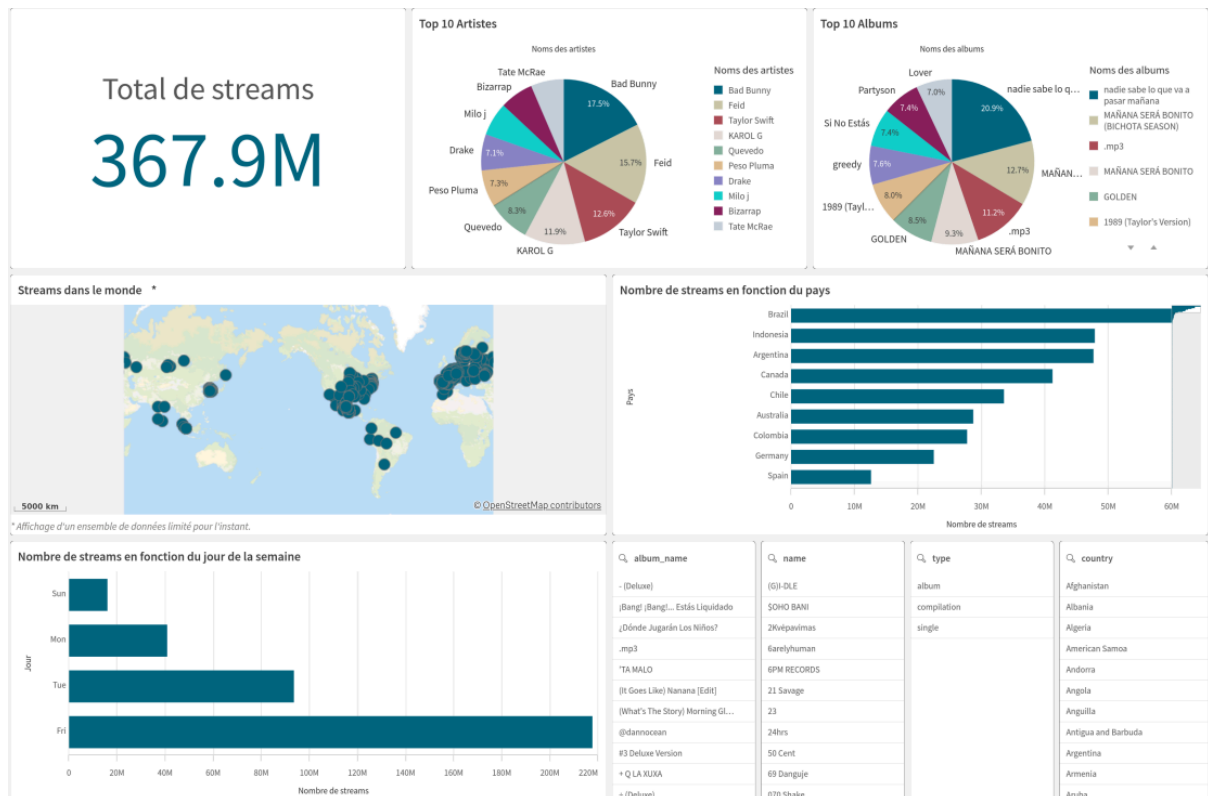
- Celui en flocon



- Celui en étoile



Capture d'écran du modèle de données (avec API):



Ce rapport résume notre approche dans la création de l'application Qlik sur une analyse du streaming musical dans le monde, mettant en avant la modélisation des données, l'utilisation du modèle multidimensionnel, et la qualité des graphiques générés pour répondre efficacement à notre problématique.