

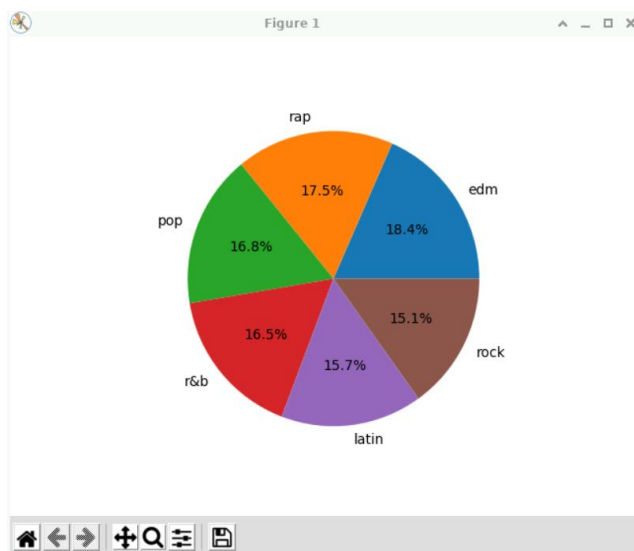
SAE S2-04 Exploitation d'une base de données

Présentation : Notre jeu de données contient des informations sur un peu plus de 30000 sons de l'application Spotify. Nous exploitons ces données pour un label de musique. Bien que l'année ne soit pas une donnée dans notre jeu de données, les sons sont d'années différentes. Nous avons comme différent type de données, beaucoup de « character » avec des données comme les noms de la musique, l'artiste qui l'a fait, le nom de l'album et sa date de réalisation, le nom de la playlist et son genre. Ensuite les autres types de données que nous avons sont des « doubles » avec la popularité du son, sa danceability, energy, loudness, speechiness, acousticness, liveness, valence, tempo et duration. Le pré-traitement que nous avons réalisé sont la suppression de certaines données, instrumentalness, mode et key que nous trouvons inutiles pour notre étude de données. Ensuite nous avons également transformé la colonne duration en duration_m et donc convertit la durée de la musique qui était en millisecondes en minutes. Pour les autres données, nous n'avons pas modifié le nom qui était déjà donné car il nous convient.

Lien vers le jeu de données en ligne :

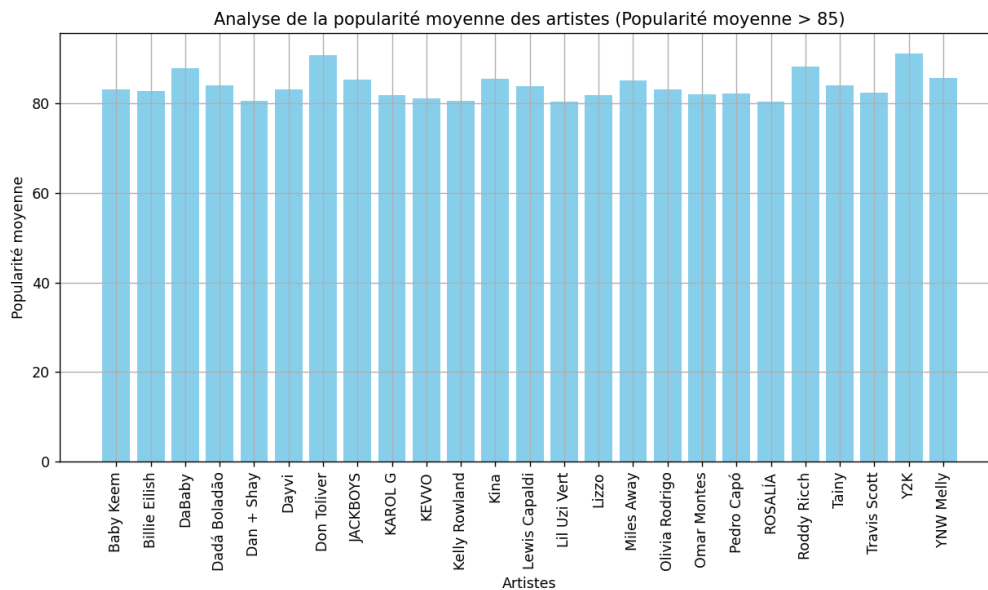
<https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs>

Analyse :



Pour ce premier graphique qui représente la répartition des différents styles de musique dans notre base de données, nous pouvons voir que cela est très bien réparti, cela peut s'expliquer par la grande diversité d'années des musiques qui figure dans ces données. Pour notre label de musique qui peut alors vouloir miser sur un style de musique en

particulier, ce graphique peut permettre de montrer que tous les styles sont bon à prendre, mais en ce moment ce qui fonctionne le mieux en terme de style c'est le rap et l'edm

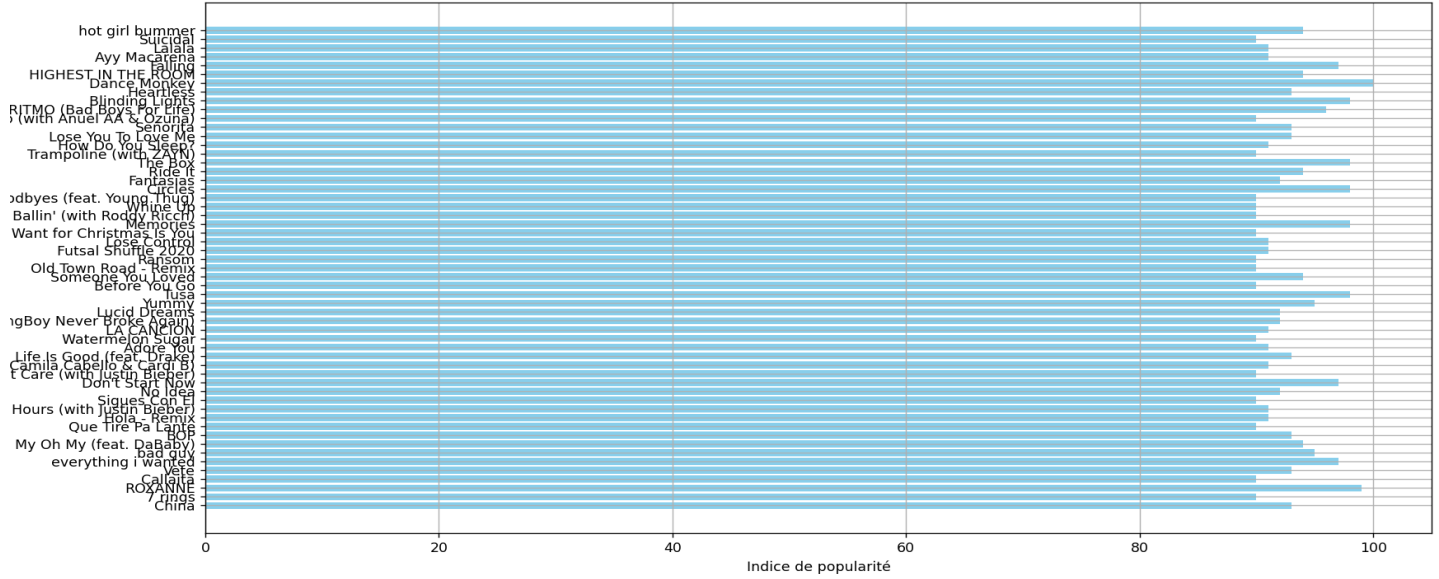


Le graphique ici représente les artistes les plus populaires de la plateforme Spotify en 2019. La popularité est calculée grâce à la moyenne d'écoute que les artistes ont par mois.

Dans notre cas ce graphique va nous servir pour savoir qui nous allons signer car les indices de popularité nous servent à savoir qui sont les meilleurs artistes. Une fois que les artistes nous auront donné leur tarif, en les comparant avec ces données là nous aurons un avant-goût de qui serait les meilleurs artistes à signer pour pouvoir réaliser un maximum de bénéfice.

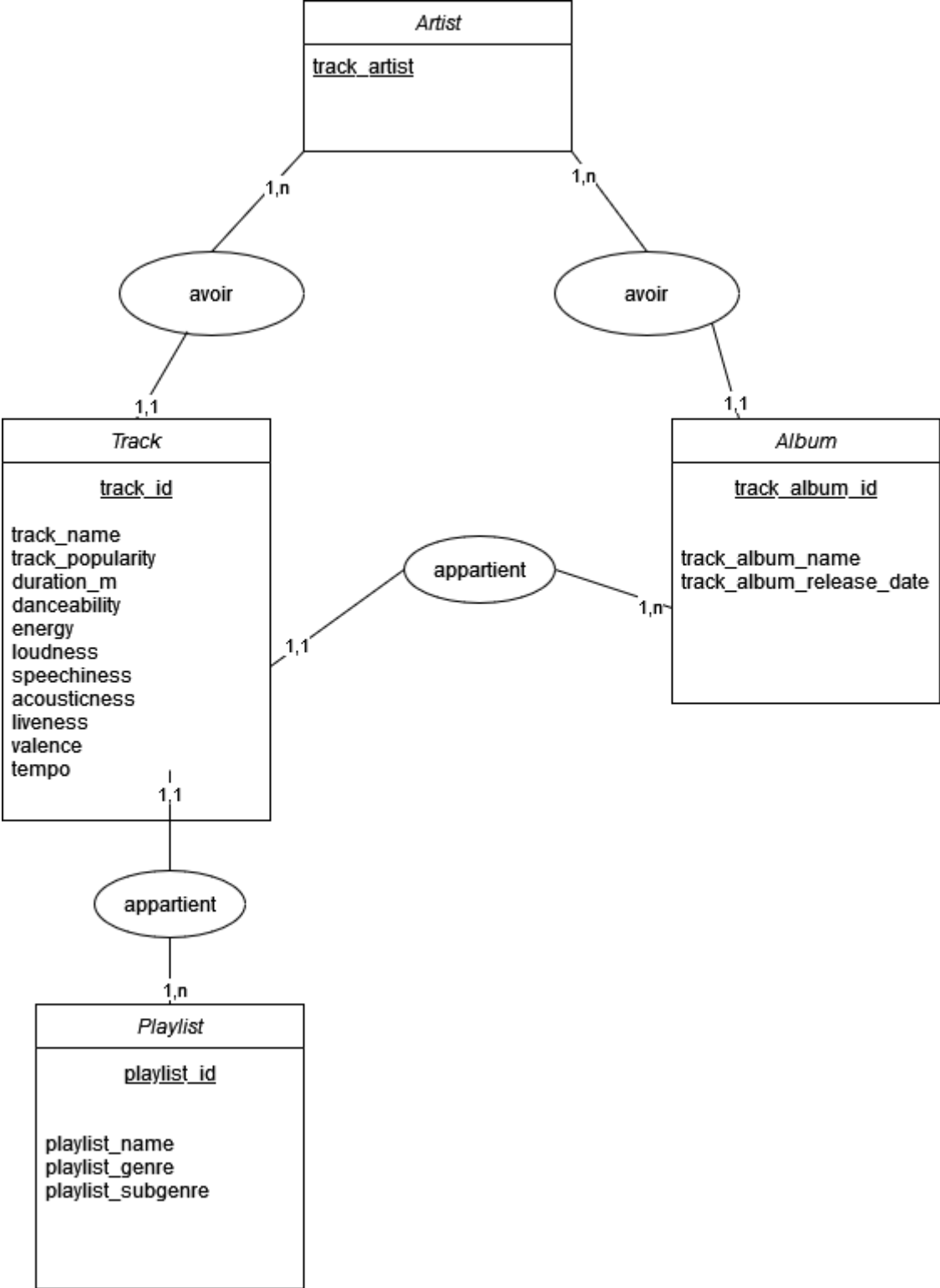
Nous avons décidé de garder uniquement les artistes ayant au minimum 85 d'indices de popularité afin de garder uniquement les artistes les plus intéressants.

Analyse des indices de popularité des chansons



Ce graphique nous montre les musiques les plus streamées, encore une fois avec un indice de popularité. En le combinant avec les premiers graphiques on va pouvoir se donner une idée de qui sont les artistes les plus intéressants à engager dans notre label. Dans un second temps le graphique va aussi nous servir à savoir qui sont les artistes les plus intéressants à signer pour notre label car c'est ceux qui ont sorti un HIT qui a bien marché et s'ils ont réussi à le faire une fois alors pourquoi pas 2. Enfin ce graphique va nous donner une idée de quelle sont les genres musicaux les plus écoutés et ceux qui ont le plus de chance de faire apparaître un hit pour l'année.

Modèle Conceptuel de Données :



c.f. graphique ligne 143

Le graphique représente le taux de possibilité de danser par rapport au tempo de la musique. On peut distinguer que le tempo sur lequel la possibilité de danser est la plus élevée est quand la musique à un tempo aux alentours de 130 BPM. Sinon c'est plutôt homogène comme le montre le nuage de point au centre.