



10/04/2023



Search



Home



Library

Introduction aux statistiques: ETUDE STATISTIQUE SUR SPOTIFY

Mathilde GARDE, Sarah FARGES, Paul ESCALIER, Victor GELINEAU





Sommaire



- 1) Introduction
- 2) Présentation des jeux de données + application des tests de cohérence
- 3) Problématique et hypothèses
- 4) Etude statistique
- 5) Conclusion

1) Quelques mots sur Spotify



Plateforme
de **streaming**

-> Créée en
2006 en
Suède

Accès à **des M**
de chansons,
podcasts...

- Grande
variété de
plateformes

+ de **365 M**
d'utilisateurs
(actifs
mensuels)

-> **170** pays

2 versions:
gratuite
(pubs) et
payante (ex:
écoute hors
ligne)

**Algorithme de
recommandation**

-> suggérer de la
musique aux
utilisateurs en
fonction de leurs
préférences





2) Présentation des jeux de données



Jeu 1

113 999 titres
20 colonnes



- Artiste
- Titre
- Indice de popularité (0 à 100)
- Durée
- Explicite
- Degré dansant
- Indice d'énergie
- Clé
- Puissance sonore
- Mode
- Degré de paroles
- Degré acoustique
- Taux d'instrumentation
- Valence
- Signature rythmique
- Genre musicale

Jeu 2

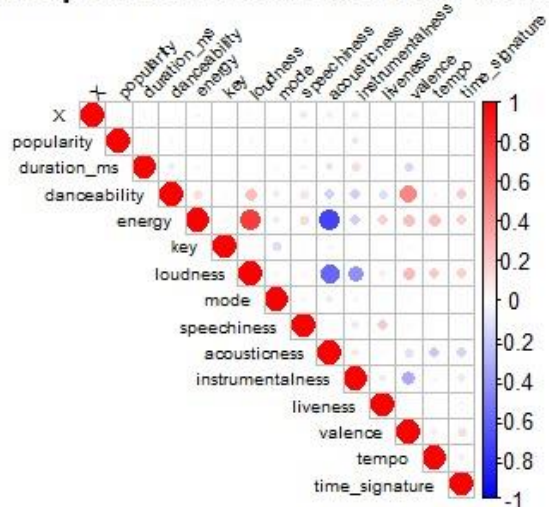
(année de
sortie en +)
2000 titres
18 colonnes



Tests de cohérence



Heatmap de Corrélation entre les variables

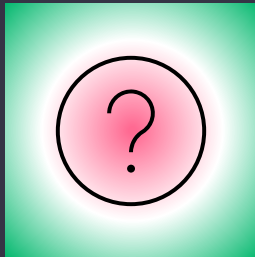


Heatmap du jeu 1

- Idée: avoir deux jeux de données complémentaires
- Mettre en avant les liens de corrélation entre les variables
- Identifier les variables inutiles (nom artiste, chansons de 1998...)



3) Problématique et Hypothèses



Problématique:

Les chansons à succès partagent-elles des caractéristiques communes? Pouvons-nous utiliser ces similarités pour prédire quelles chansons deviendront des hits?

Hypothèses:

- Musiques avec beaucoup de rythme et forte intensité sonore
- Le genre qui prédomine est celui de la pop
- + de son artificiels et - d'acoustique
- Les chansons qui fonctionnent le mieux sont celles sur lesquelles on peut danser + facilement
- Pas besoin de beaucoup de paroles
- Tonalité majeure

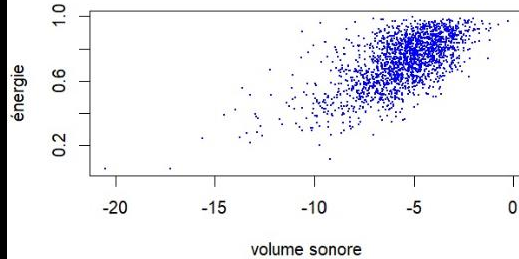


Hit: chanson avec indice de popularité ≥ 70

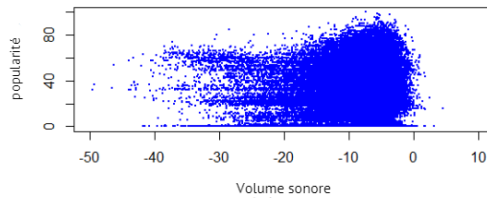


Hypothèse n°1: Les musiques qui plaisent le plus ont tendance à être plus dansantes, énergiques, rythmées et de forte intensité sonore:

Comparaison de l'énergie en fonction du volume sonore



Comparaison de la popularité en fonction du volume sonore



- Les chansons populaires ont tendance à avoir un volume sonore élevé
- Plus le volume sonore est important, plus l'énergie émanée l'est aussi
- Moins d'acoustique dans les hits

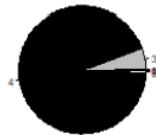




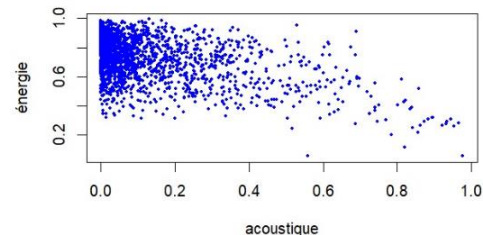
- Plus l'énergie est importante, plus la part d'acoustique est faible
- Même si grande majorité des hits avec beaucoup d'énergie -> développement de chansons + intimistes et + calmes
- Musique les + populaires très rythmées et dynamiques



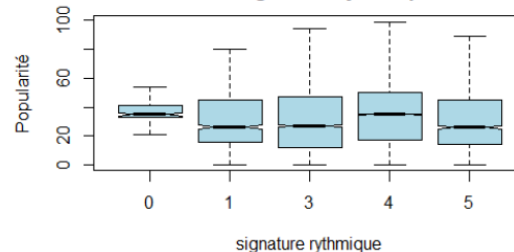
Signatures rythmiques dans les musique à forte popularité (≥ 70)



Comparaison de la énergie en fonction de l'acoustique

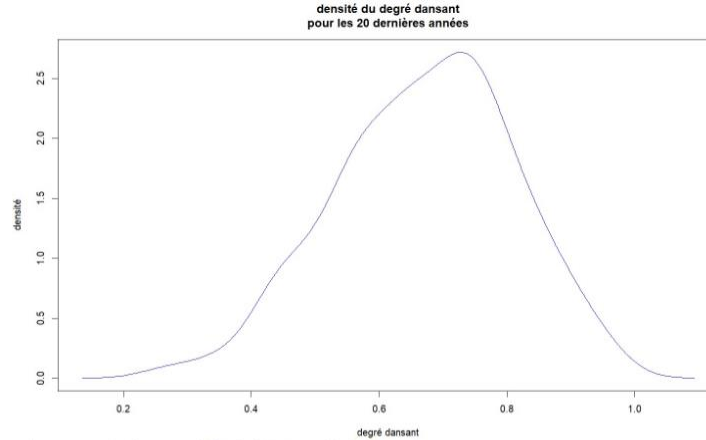


Comparaison de la popularité en fonction de la signature rythmique





Point de passage application du cours:



```
> mean(Spotify$danceability)
[1] 0.6708686
> var(Spotify$danceability)
[1] 0.01953399
```

Application de la loi normale à la moyenne de la densité du degré dansant, calcul intervalle de confiance à 95% :

$$IC_{95\%} = \left[\mu - x \times \frac{\sigma}{\sqrt{n}}, \mu + x \times \frac{\sigma}{\sqrt{n}} \right]$$
$$IC_{95\%} = \left[0,67 - 1,96 \times \frac{0,14}{\sqrt{236}}, 0,67 + 1,96 \times \frac{0,14}{\sqrt{236}} \right]$$
$$IC_{95\%} = [0,6521 ; 0,6878]$$

Hypothèse n°2: Nous sommes portés à écouter des musiques joyeuses le plus souvent de tonalité majeure:

#1

Réduction anxiété et douleurs



#2

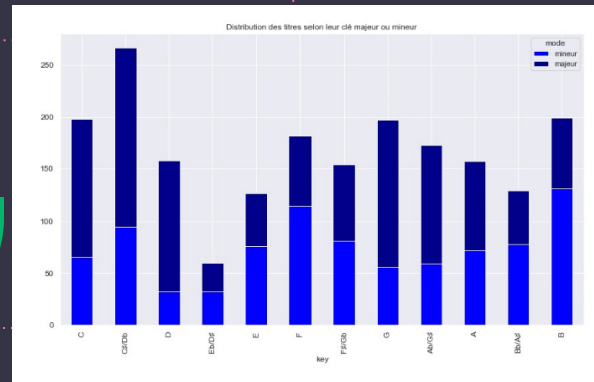
Baisse tension artérielle



Forte présence de musiques majeures

#4

Réparation cérébrale chez les grands prématurés



Prédominance de la clé C#/D \flat

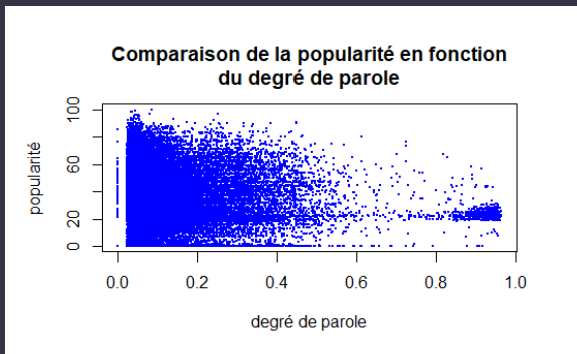
#3

Diminution certains effets secondaires

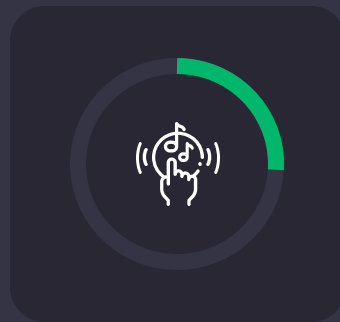


Bienfaits de la musique

Hypothèse n°3: Il n'y a pas forcément besoin de beaucoup de paroles pour qu'un titre fonctionne bien et généralement ces dernières sont peu explicites:

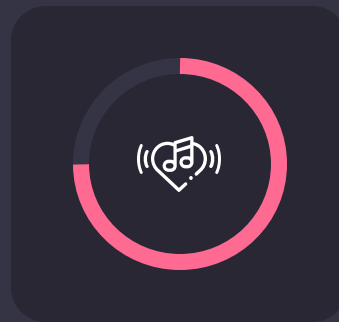


peu de chansons avec un degré de parole important



25%

Explicite



75%

Non-Explicite

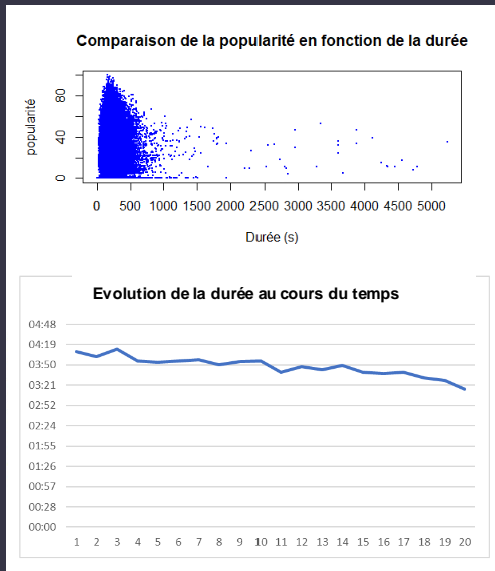
Touche un public plus large

Majorité des chansons populaires non-explicites

Pour les chansons d'indice de popularité ≥ 70



Hypothèse n°4: Une grande partie des musiques font la même durée et la durée des musiques tend à diminuer:



Concentration autour
d'une même durée-> pour
presque l'intégralité des
titres
-> au-delà de la
popularité => tendance

Légère diminution de la
durée au cours des
années

Hypothèse n°5: Au fur et à mesure des années, les tendances changent et les musiques tendent à s'uniformiser (elle sont de plus en plus semblables):

Les genres musicaux les plus écoutés dans le monde

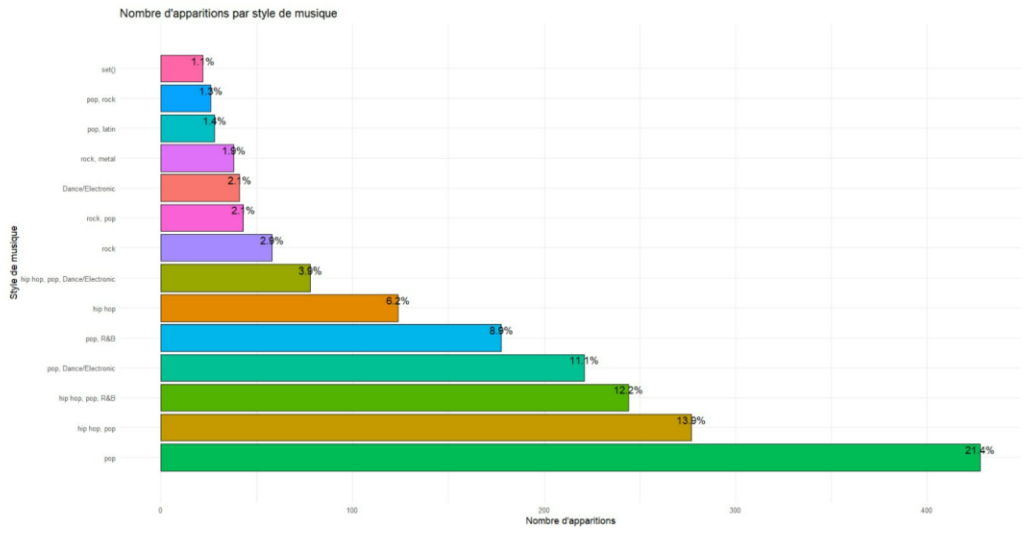
Part des consommateurs de 15 pays qui écoute les genres musicaux suivants*



En moyenne, les consommateurs passent 17,8 heures à écouter de la musique chaque semaine.

* Sondage mené en avril-mai 2018 auprès de 19 000 consommateurs de 18 pays, âgés de 16 à 64 ans.
Source : IFPI

statista



Prédominance de la **pop** (« pop culture »)
Pop combinée à d'autres styles: **73,7%** des écoutes
-> uniformisation des genres





5) Conclusion



- Impossible de prédire avec exactitude si une chanson va devenir un hit
- Mais identification de plusieurs caractéristiques communes aux hits
 - > + de chance de réussir en les regroupant
- Traits communs aux hits:
 - Rythmés
 - Énergiques
 - Dansants
 - Faible variété instrumentale
 - Pas beaucoup de paroles
 - Intensité sonore élevée
- Omniprésence de la pop
- Uniformisation des styles et genres musicaux
 - > s'étend à d'autres domaines: mode, cinéma...

