

Prédire les *certifications* de singles en France

Création d'un modèle capable de prédire si un single sera ou non certifié en France

Sommaire

01	<i>Introduction</i> Avant-propos	03
02	<i>Certifications</i> Certifications de singles en France	06
03	<i>Base de données</i> Présentation des variables	11
04	<i>Analyse</i> Etude analytique	15
05	<i>Algorithmes</i> Modèles de prédiction	26
06	<i>Conclusion</i> Bilan et ouverture	29

01

Introduction **Première partie**

—

Avant-propos

le single en France

Présentation du SNEP et des différentes certifications

Le Syndicat National de l'Édition phonographique

*Association interprofessionnelle qui défend les intérêts
de l'industrie française du disque phonographique*

→ Certifie les récompenses et fixe les seuils depuis 1973

→ 2434 certifications remises depuis 1994

Liste des différentes certifications en France

Certifications

**Disque
d'or**

*15 millions d'équivalent
streaming*

DÉSANCHANTÉE
MYLÈNE FARMER

**Disque
de platine**

*30 millions d'équivalent
streaming*

BLANKA
PNL

**Disque
de diamant**

*50 millions d'équivalent
streaming*

MARIE
JOHNNY HALLYDAY

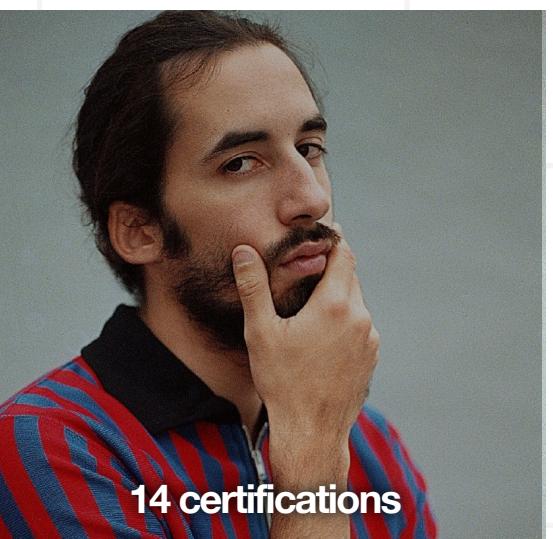
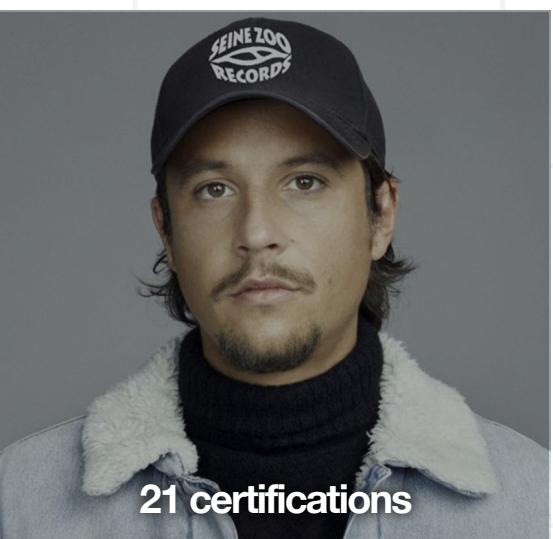
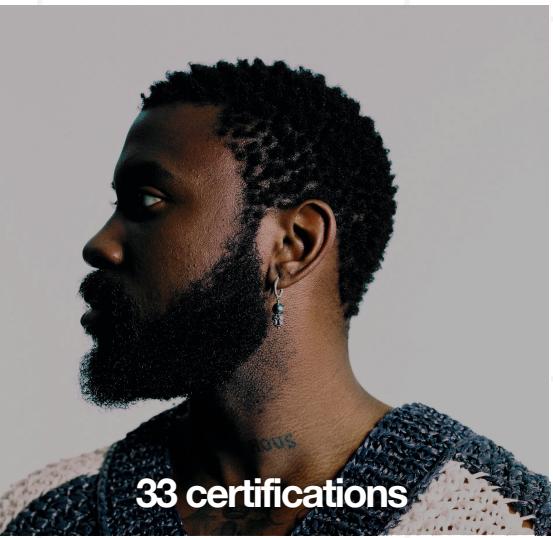
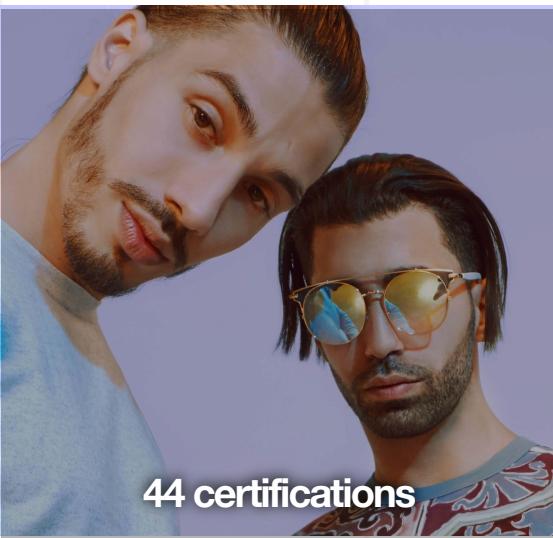
02

Certifications **Deuxième partie**

—

Certifications *de singles en France*

Les 10 artistes les plus certifiés



Le Hip-Hop en tête

Les dix artistes les plus certifiés, sont issus principalement de deux genres : Hip-Hop français et francoton.

À titre de comparaison, Johnny Hallyday ne détient que cinq singles certifiés.

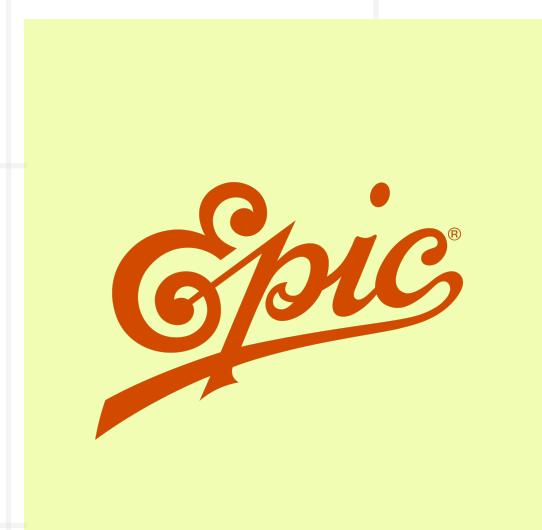
Certifications *de singles en France*

Les 10 labels les plus certifiés

Majors en tête

Sony, Universal et Warner sont les trois labels les plus certifiés en France.

Elles sont également les maisons de disques les plus puissantes au monde



Certifications de singles en France

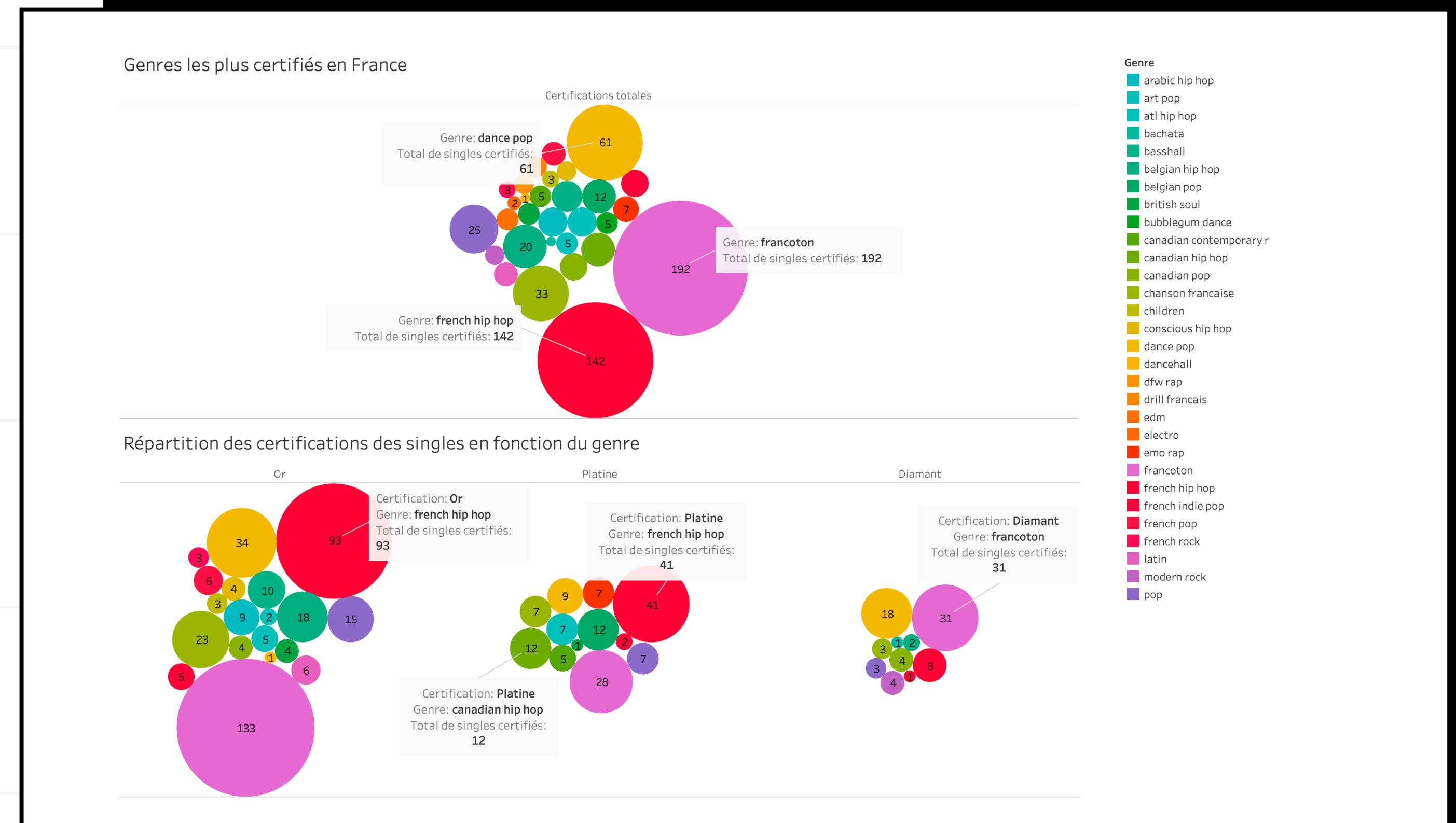
Les 30 genres musicaux les plus certifiés

Trois genres champions

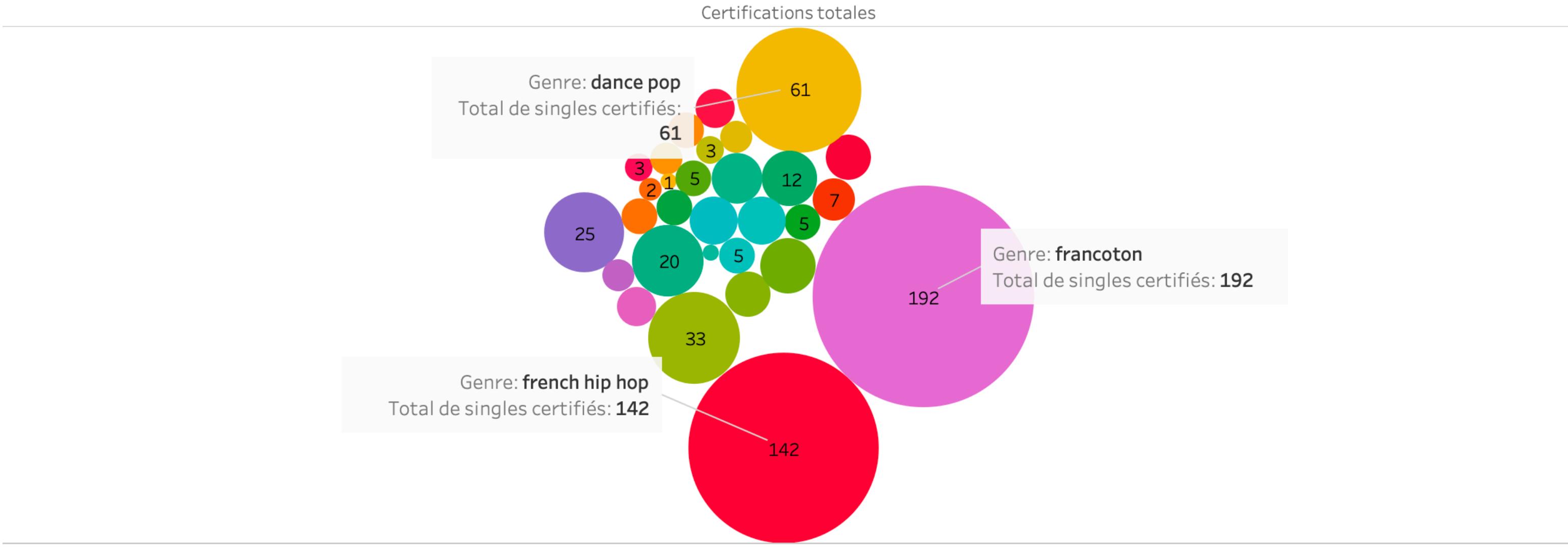
*Les trois genres musicaux les plus certifiés en France sont le **Francoton, le French hip-hop et la dance pop.***

Le French hip-hop obtient plus de disques de platine

Graphique – Les 30 genres musicaux les plus certifiés en France



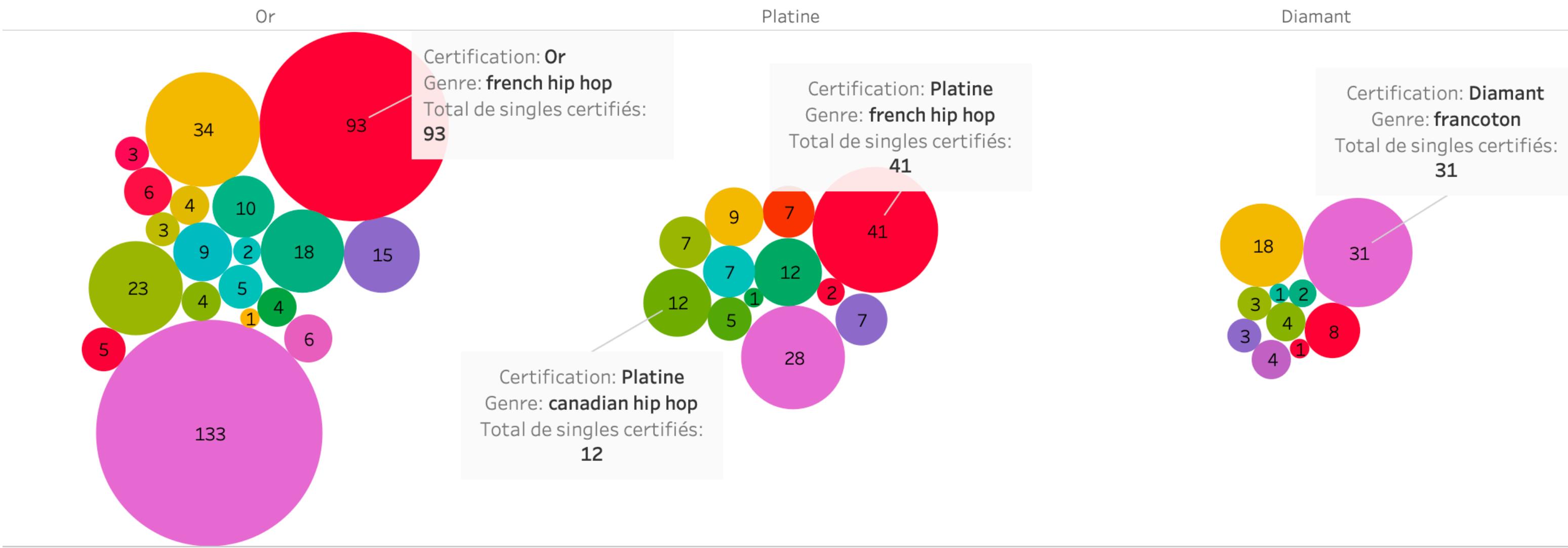
Genres les plus certifiés en France



Genre

- arabic hip hop
- art pop
- atl hip hop
- bachata
- basshall
- belgian hip hop
- belgian pop
- british soul
- bubblegum dance
- canadian contemporary r
- canadian hip hop
- canadian pop
- chanson française
- children
- conscious hip hop
- dance pop
- dancehall
- dfw rap
- drill français
- edm
- electro
- emo rap
- francoton
- french hip hop
- french indie pop
- french pop
- french rock
- latin
- modern rock
- pop

Répartition des certifications des singles en fonction du genre



Le **francoton** est le genre le plus certifié. Il obtient le plus de disques d'or et de diamant en France.

03

Base de données **Troisième partie**

—

Constitution d'une base de données

Spotipy

Librairie Python permettant de récupérer les données fournies par Spotify

Exportify

Application permettant de télécharger les données Spotify des playlists

SNEP

Fichier du SNEP contenant les certifications de singles et d'albums

Constitution d'une base de données

Spotipy

Librairie Python permettant de récupérer les données fournies par Spotify

Exportify

Application permettant de télécharger les playlists Spotify

SNEP

Fichier du SNEP contenant les certifications de singles et d'albums

Base de données contenant 1 383 singles certifiés ou non, ainsi que 36 variables explicatives

Base de données

Présentation des variables

Variables explicatives

Variables
décrivant l'artiste
et ses chansons

artist_name artist_popularity
followers track_name
track_popularity date_release
 genre day_release
 month_release year_release
 weekday_release album_name
 labels explicit_content

Variables
décrivant les
certifications

is_certified certif_single
certif_album nb_jours_certif_album
 nb_jours_certif_singles
 eq_streams_singles eq_ventes_albums

Variables
décrivant les
caractéristiques audios
des chansons

acousticness danceability
energy duration_min
instrumentalness speechiness
 key liveness
 loudness mode
 tempo time_signature
valence end_of_fade_in
 start_of_fade_out

04

Analyse

Quatrième partie

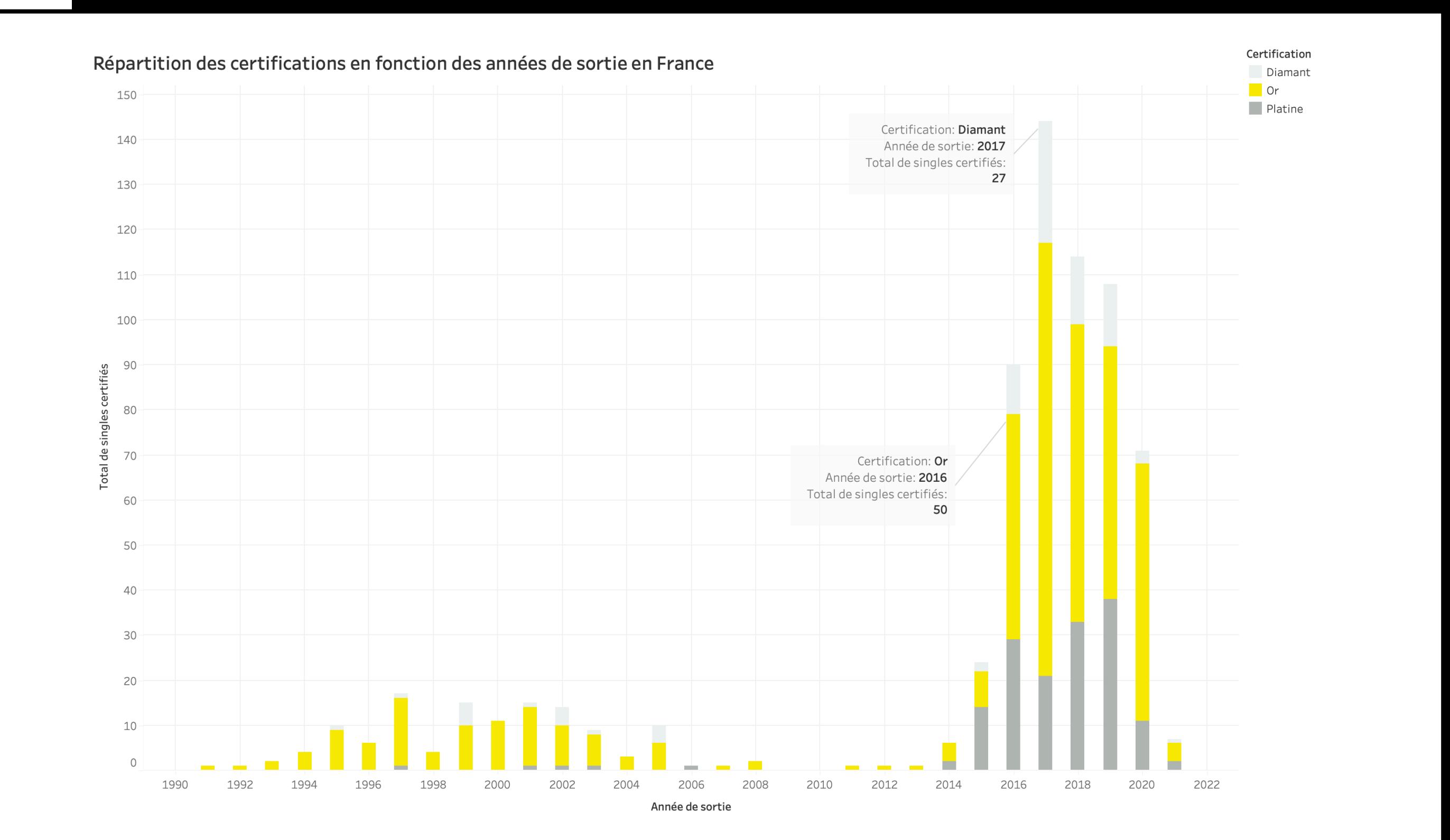
Étude analytique

Analyse des dates de sortie des singles certifiés

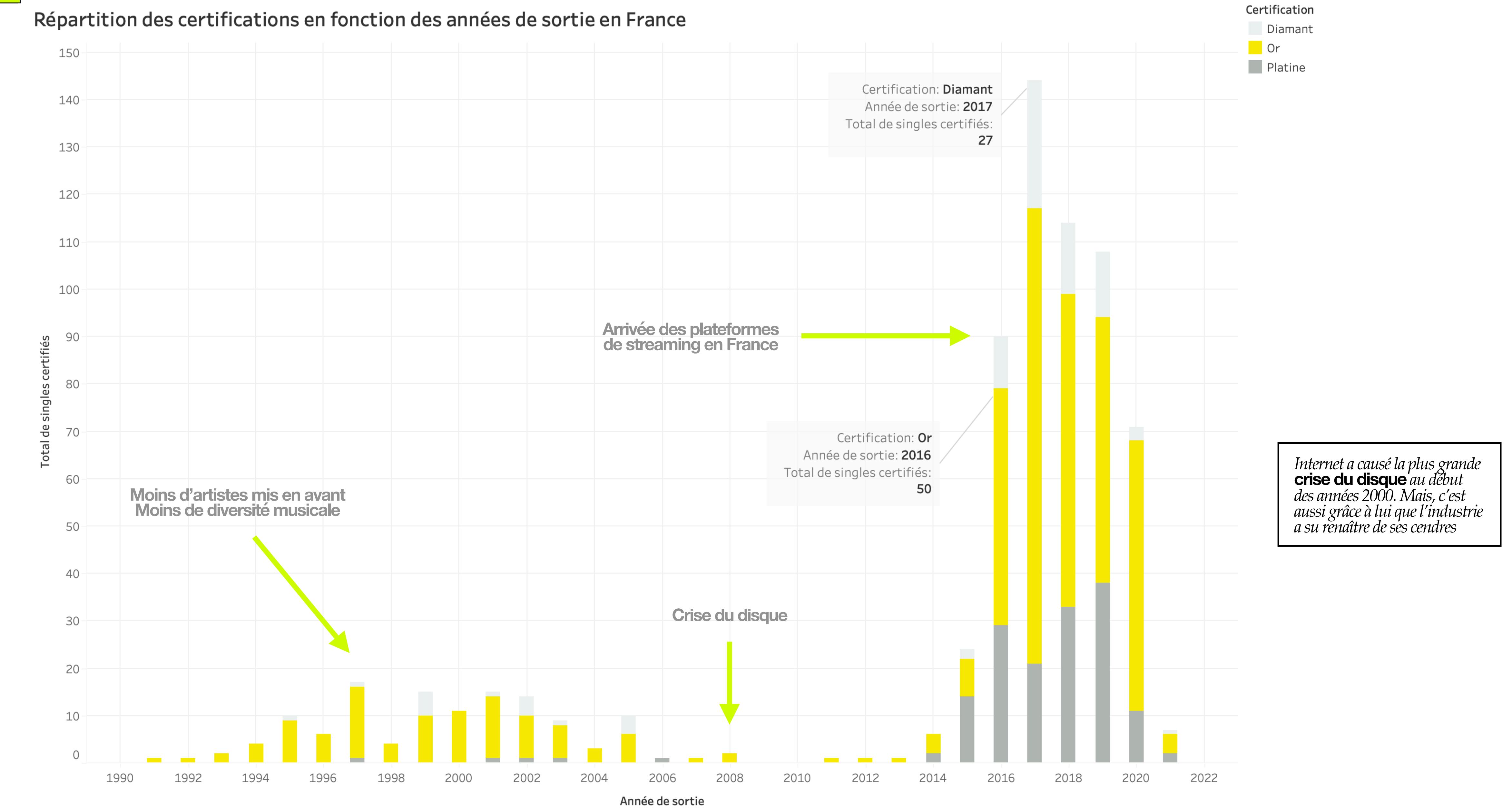
Certifications et streaming
*On observe depuis 2016 une **explosion** des certifications en France. Cela est dû grâce à l'arrivée des plateformes de streaming.*

L'année 2017 enregistre le plus de certifications de singles, avec plus de 140 singles certifiés.

Graphique — Année de sortie des singles certifiés



Répartition des certifications en fonction des années de sortie en France



Étude analytique

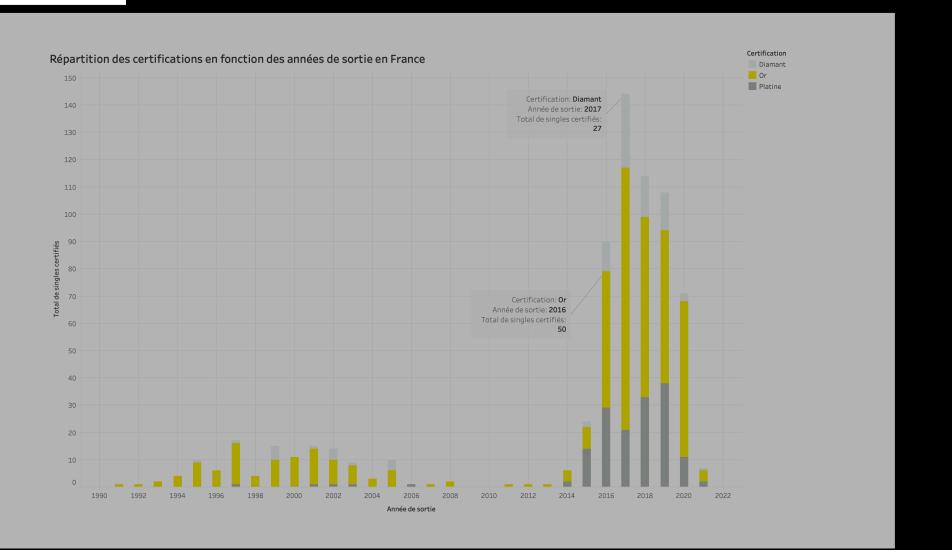
Analyse des dates de sortie des singles certifiés

Saisonnalité

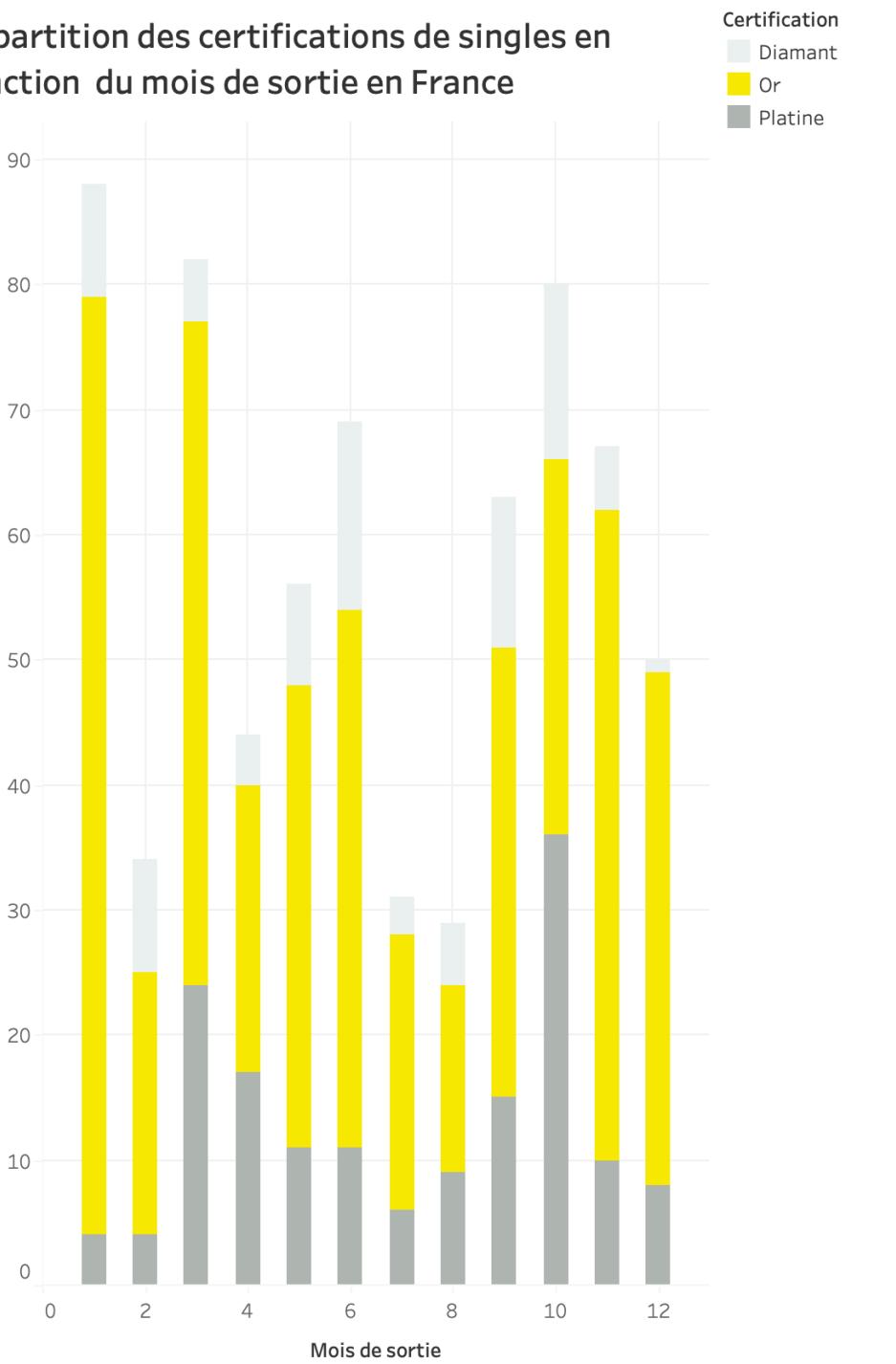
*Les singles les plus certifiés sont sortis en **janvier, mars et octobre**.*

Cela correspond également aux mois comptabilisant le plus de sorties de singles.

Graphique — Année de sortie des singles certifiés



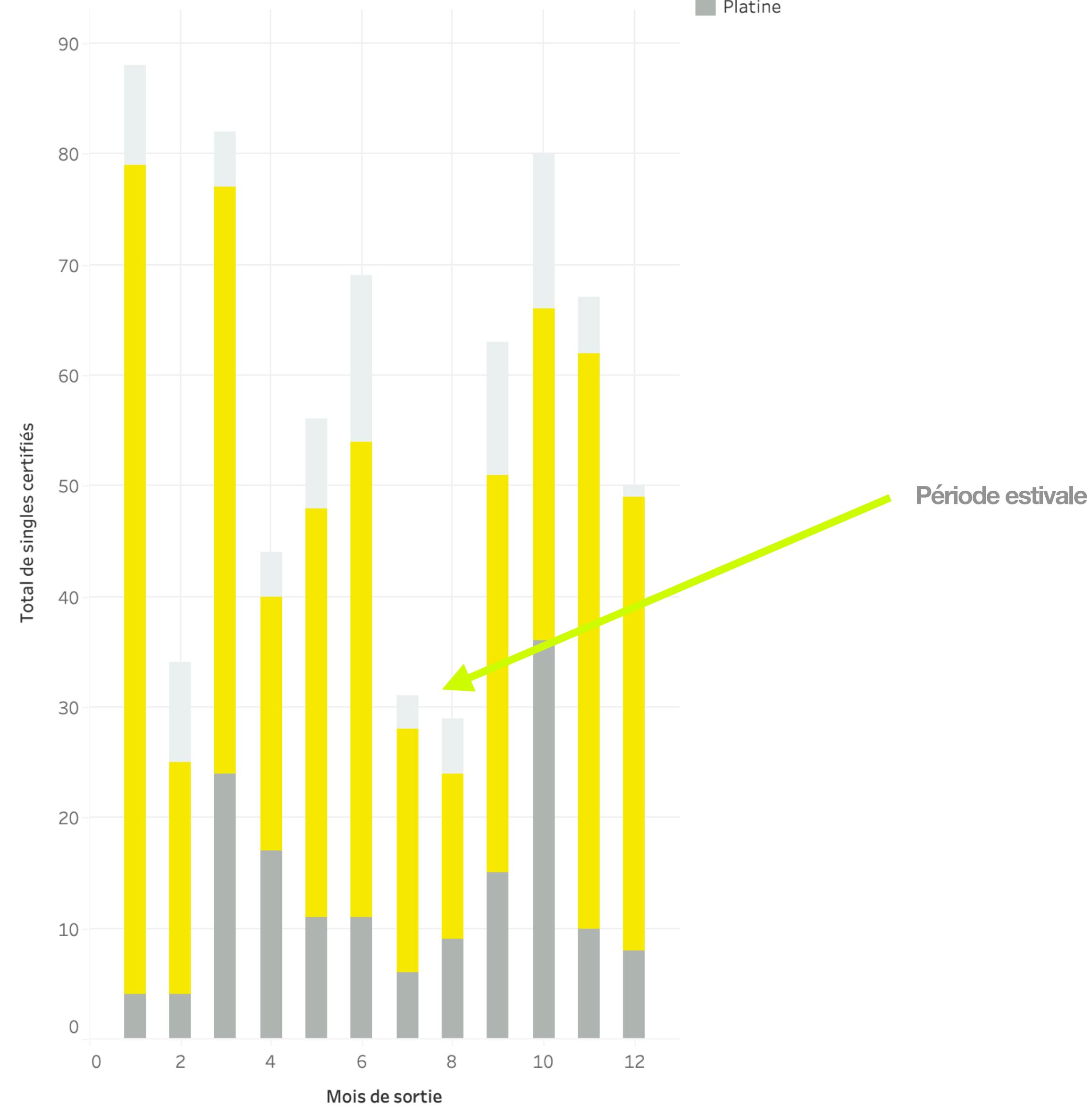
Répartition des certifications de singles en fonction du mois de sortie en France



Graphique — Mois de sortie des singles certifiés

Répartition des certifications de singles en fonction du mois de sortie en France

Certification
Diamant
Or
Platine



On constate moins de certifications de singles sortis en **juillet et août**. Les artistes se concentrent sur juin et septembre, périodes clés.

Étude analytique

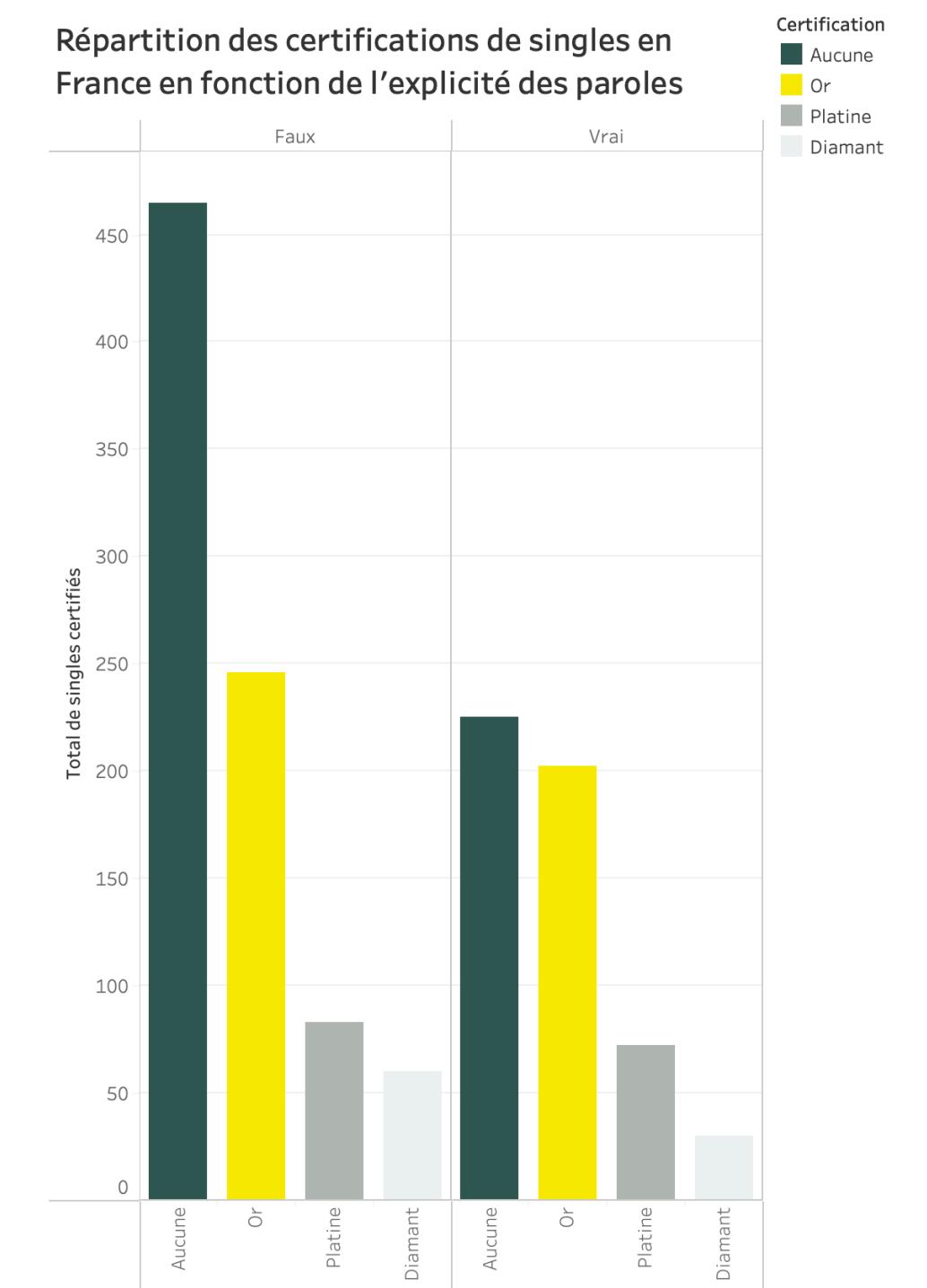
Analyse de l'influence du contenu explicite

Réelle influence ?

La grossièreté des paroles n'influencent pas la certification de singles.

Il a quasiment autant de chances d'être certifié si le single contient du contenu explicite.

Graphique — Explicite des paroles des singles certifiés en France



Étude analytique

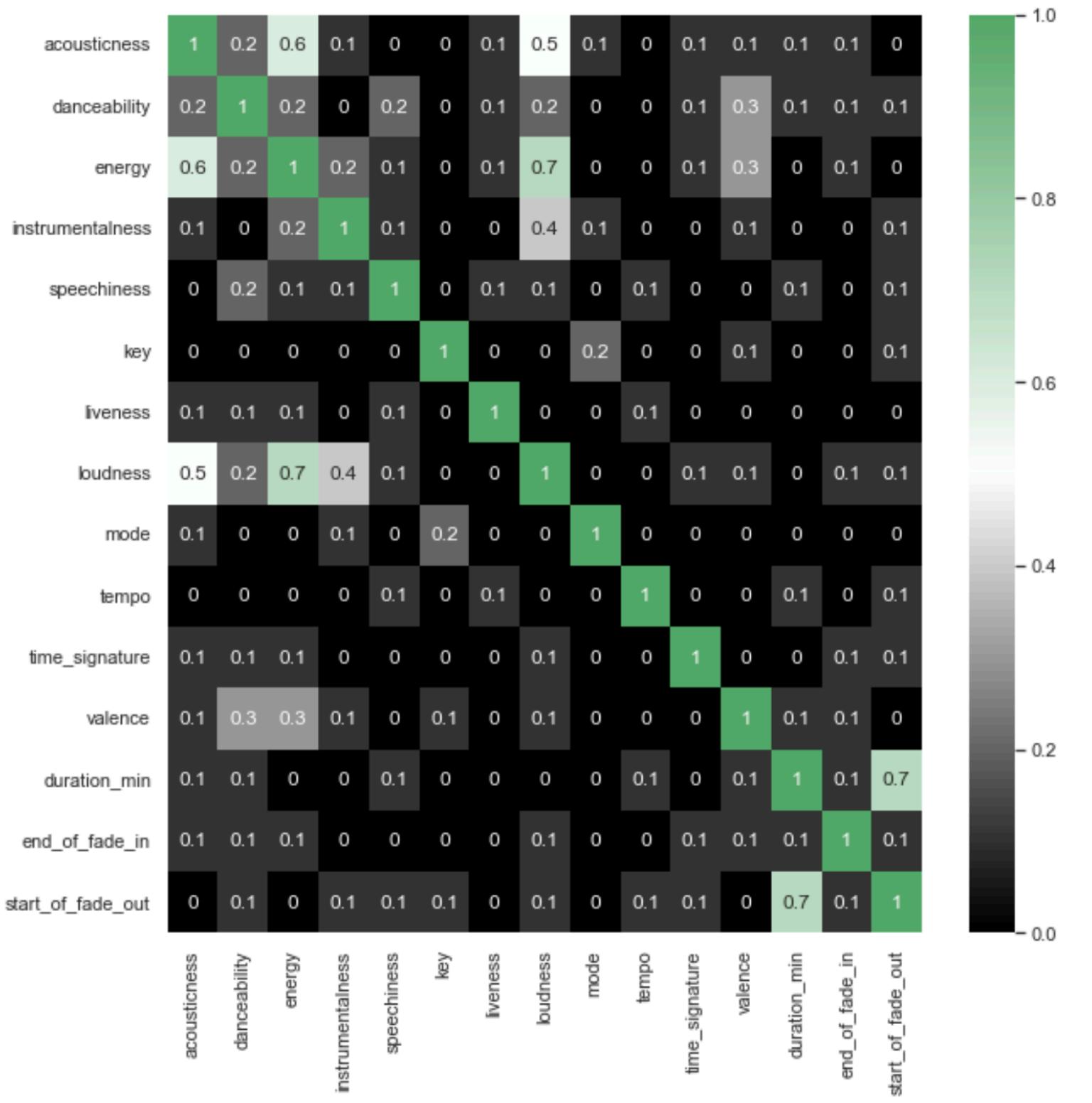
Corrélation entre les caractéristiques audios des singles

Heatmap

On peut observer les corrélations entre les variables continues grâce à une carte de chaleur.

Le coefficient de Pearson permet d'établir les relations entre les variables

Graphique — *Représentation des corrélations entre les caractéristiques audios des singles*

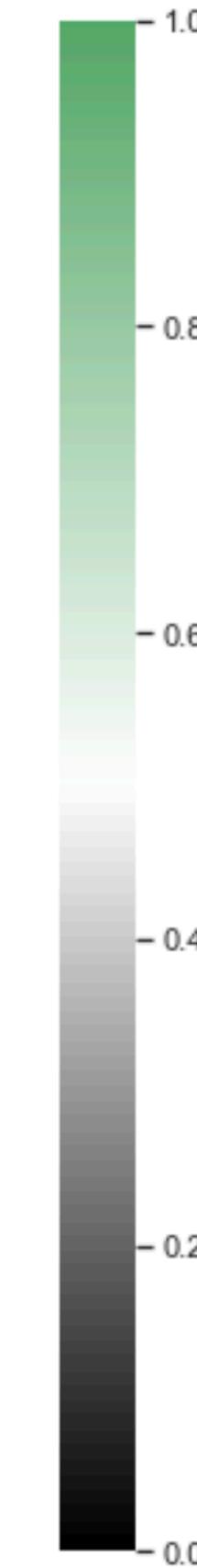
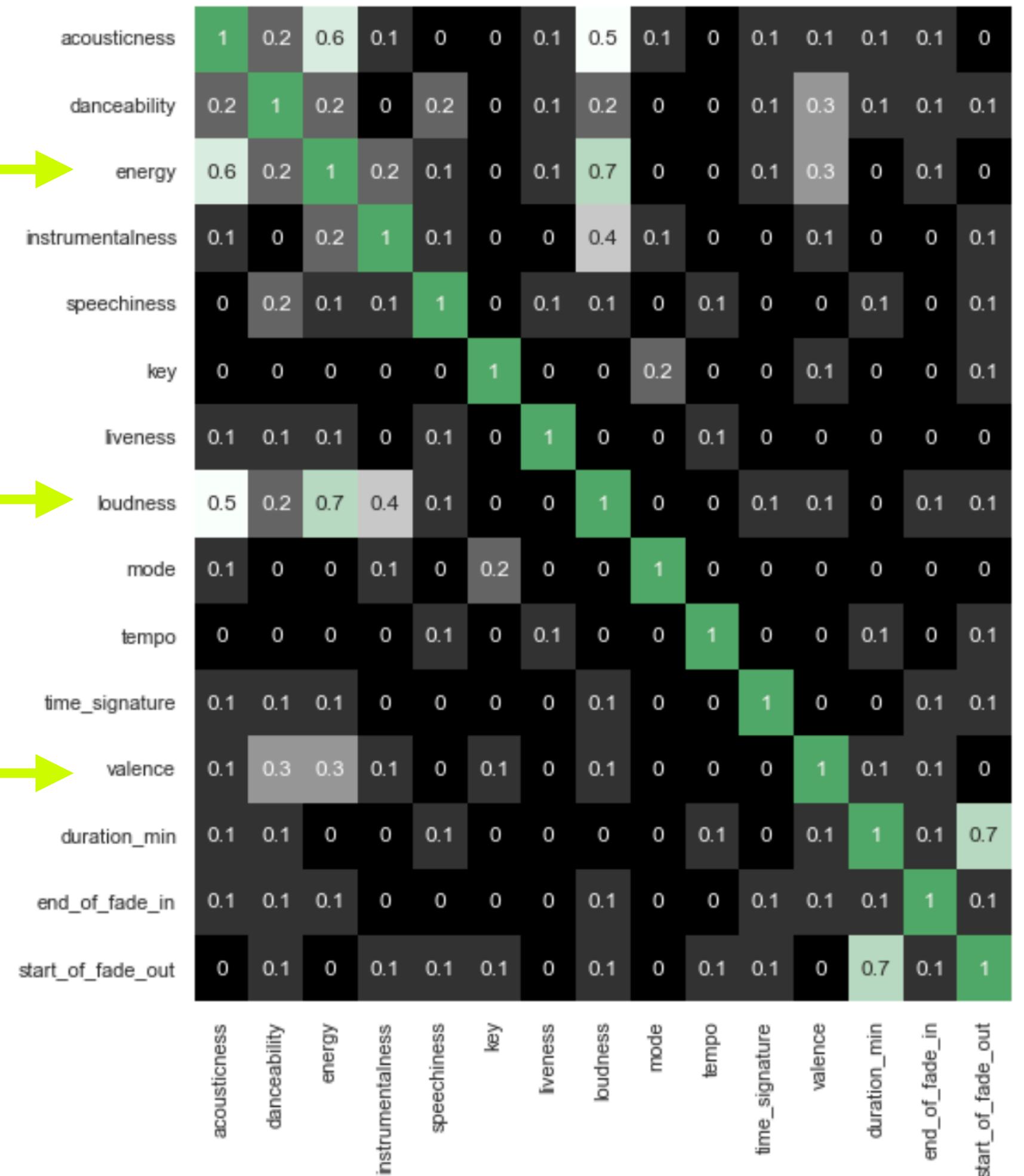


Graphique — Représentation des corrélations entre les caractéristiques audios des singles

une chanson acoustique est moins énergique

plus une chanson a une intensité sonore élevée, plus elle est dansante, énergique et plus elle contient de chants

plus la chanson est positive, plus elle est dansante et plus énergique



On observe plusieurs corrélations positives entre nos variables continues. Elles sont surtout corrélées à l'intensité sonore.

Étude analytique

Analyse des autres types de corrélations

Dépendance entre les différents types de variables

ANOVA/KRUSKAL-WALLIS

Corrélation entre une variable continue et une variable catégorielle.
Variable de référence :
is_certified_single

energy, speechiness, key, liveness et time_signature

KHI-2

Corrélation entre deux variables catégorielles.
Variables de référence :
is_certified_single

genre, is_certified_album, label, artist_popularity et followers

Étude *analytique*

À quoi ressemble un single certifié en France ?

Énergique **Dansant** Moins de 4min **Positivité** Contient des chants

05

Algorithmes Cinquième partie

—

Comparaison entre les différents types d'algorithme

Modèles

Tree-based

XGBoost

Random Forest

Decision Tree

Distance-based

SVM

kNN

Gradient decent-based

Régression logistique

Comparaison entre les différents types d'algorithme

Modèles

Tree-based

XGBoost

Random Forest
85% accuracy
91% precision

Decision Tree

Distance-based

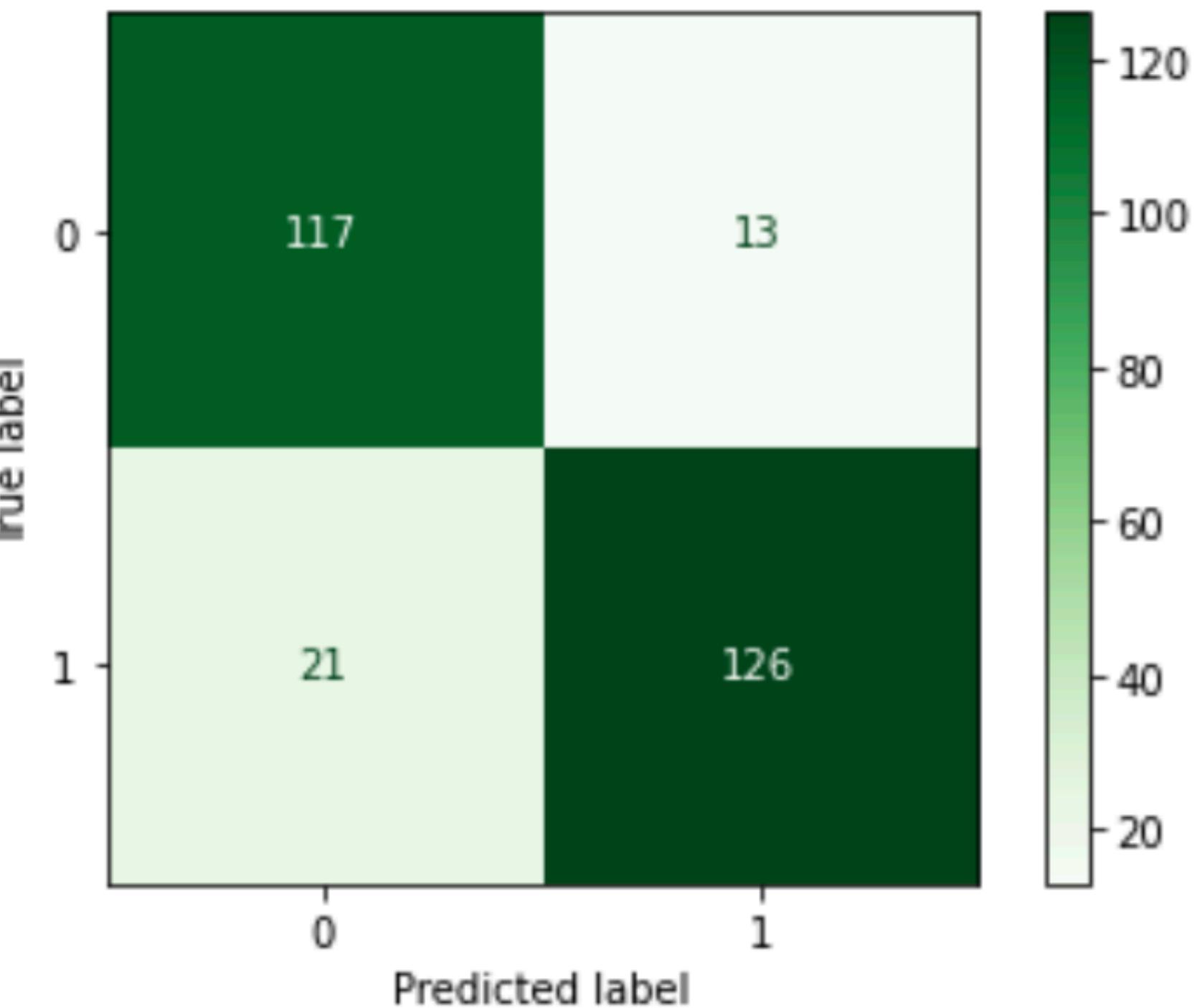
SVM

kNN

Gradient decent-based

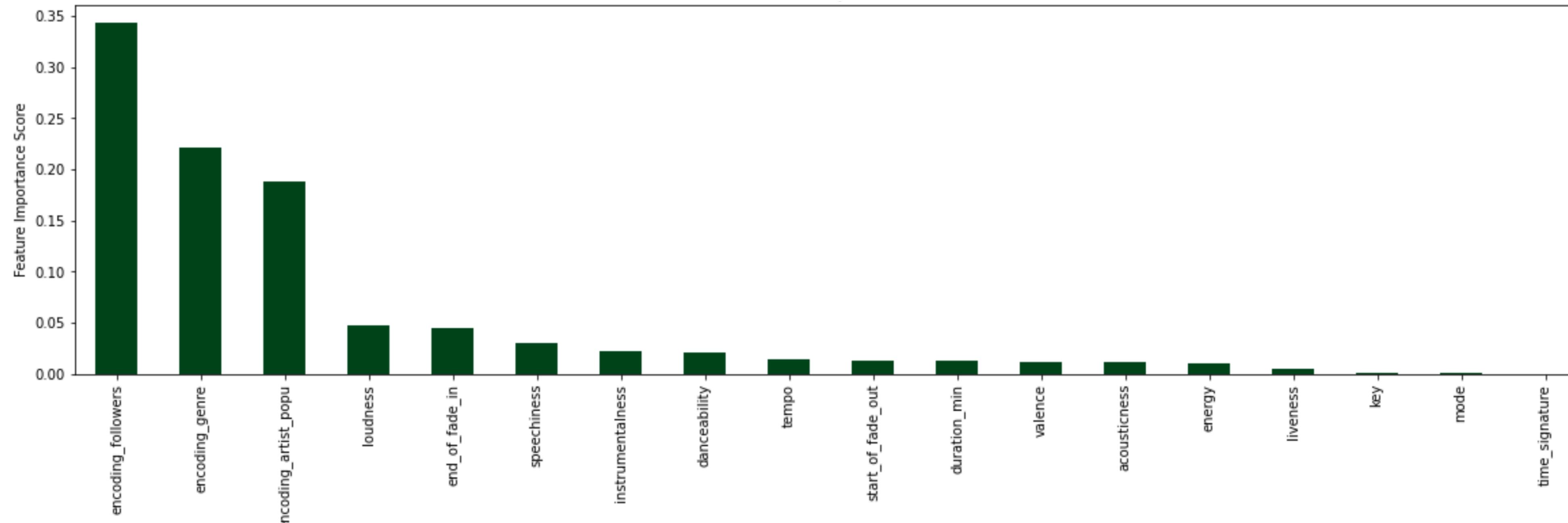
Régression logistique

Graphique – Matrice de confusion pour le modèle de prédiction Random Forest



On obtient de bons scores de prédiction avec seulement 13 faux négatifs et 21 faux positifs

Graphique — Variables les plus déterminantes dans l'obtention d'une certification d'après le modèle Random Forest



Le nombre de **followers**,
le **genre** et la **popularité de l'artiste** sont les éléments les plus déterminants dans l'obtention d'une certification

06

Conclusion

Bilan

—

Prédire une certification de single

En conclusion...

L'obtention d'une certification de single peut être expliquée par différentes variables.

Le genre musical, la popularité de l'artiste, le nombre de followers, l'énergie de la chanson, la tonalité, le label notamment ont un impact sur la certification d'un single en France.

Le francoton, le hip-hop français et la dance pop sont les genres les plus populaires et certifiés.

Les artistes issus de ces genres sont les plus gros vendeurs de singles dans l'Hégaxone.

Avec Random Forest, nous avons créé un modèle capable de prédire si un single sera certifié ou non avec un precision score de 91%.

Merci.