



Rapport projet Data Mining

HAOUD Anas, NGETH Laurent, PAZITHNOV Artemii

02 mars 2025

Table des matières

Introduction	2
Description des Données	2
Prétraitement et exploration des données	4
Chargement et aperçu des données	4
Analyse de la distribution des genres.....	4
Corrélation entre les variables.....	5
Analyse par genre.....	6
Analyse en composantes principales.....	7
Machine learning.....	8
Séparation des données et prétraitement des variables	8
Analyse des Résultats de Modélisation et Importance des Variables.....	9
Importance des Variables	10
Bonus : Modèle CNN pour la classification de genres musicaux.....	11
Annexes	13
1. Descriptions.....	13
2. Distributions des colonnes	15
3. Analyse par genre.....	17
4. Matrices de Confusion	19
5. Tableau des corrélations des features les plus importantes.	21
6. Mel Spectrogrammes	22

Introduction

Dans un contexte où la musique occupe une place prépondérante dans notre quotidien – que ce soit dans les services de streaming, les systèmes de recommandation ou l’analyse musicale – la capacité à classifier automatiquement les genres musicaux représente un enjeu majeur. L’objectif de ce projet est de développer un modèle prédictif capable d’identifier le genre musical d’un extrait audio en se basant sur des caractéristiques extraites automatiquement.

Pour ce faire, nous avons utilisé le dataset GTZAN. Les caractéristiques extraites – telles que les coefficients MFCC, les mesures de la brillance spectrale, le tempo, et d’autres indicateurs issus de l’analyse du signal – fournissent une représentation quantitative du contenu audio qui sera exploitée par divers algorithmes d’apprentissage automatique.

En outre, nous avons construit un modèle CNN simple pour obtenir une classification plus précise des genres.

Ce rapport détaille l’ensemble des étapes du projet, depuis l’exploration initiale des données (EDA) et le prétraitement, jusqu’à la modélisation et l’évaluation des performances des différents modèles.

Description des Données

Le jeu de données utilisé pour ce projet concerne la **classification de genres musicaux**, composé de **caractéristiques audio extraites de plusieurs pistes appartenant à différents genres**. Ce jeu de données couvre **10 genres musicaux distincts : blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae et rock**. Chaque piste est décrite par un ensemble de **caractéristiques numériques** issues de l’analyse audio, capturant différents aspects du profil sonore.

Aperçu des caractéristiques

Les caractéristiques se répartissent en plusieurs catégories :

- **Caractéristiques spectrales**

Ces caractéristiques décrivent la répartition fréquentielle du signal audio, apportant des informations sur la **tonalité** et la **texture** de chaque piste. On retrouve notamment :

- **Chroma** : Distribution de l’énergie sonore par hauteur de note.
- **Centroïde spectral** : Position moyenne des fréquences, donnant une indication de la **brillance** de la piste.
- **Largeur de bande spectrale** : Mesure de la dispersion des fréquences autour du centroïde.
- **Rolloff spectral** : Fréquence en dessous de laquelle se trouve la majorité de l’énergie du signal.

- **Caractéristiques temporelles**

Issues directement de l'onde sonore, elles décrivent la dynamique et le niveau d'énergie de la piste. Exemples :

- **RMS (Root Mean Square)** : Mesure de la puissance moyenne du signal.
- **Taux de passage par zéro** : Fréquence à laquelle le signal change de signe, indicateur du caractère percussif ou bruité du son.

- **Coefficients cepstraux de Mel (MFCC)**

Les MFCC sont une représentation compressée du spectre audio, adaptée à la perception auditive humaine. Le jeu de données contient les **moyennes** et **variances** des 20 premiers MFCC, offrant ainsi une description complète de l'évolution spectrale dans le temps.

- **Caractéristiques rythmiques et harmoniques**

Ces variables capturent des aspects plus larges liés à la structure musicale, incluant :

- **Tempo** : Estimation de la vitesse (battements par minute) de la piste.
- **Harmonie** : Quantification de la richesse harmonique.
- **Caractéristiques perceptuelles** : Variables psychoacoustiques combinant informations spectrales et harmoniques.

Enfin, le dataset inclut également une colonne filename (nom du fichier audio) et la variable cible label indiquant le genre musical. Notez que la variable length a été supprimée dès le chargement des données. Un tableau avec l'explication intuitive des variables est joint en annexe [\[1\]](#).

Prétraitement et exploration des données

Chargement et aperçu des données

Les données utilisées dans ce projet proviennent d'un ensemble de fichiers audio couvrant plusieurs genres musicaux. Chaque morceau a été transformé en un vecteur de caractéristiques via l'extraction de descripteurs audio (MFCC, chroma, spectral centroid, rolloff, etc.), créant une matrice où chaque ligne correspond à un morceau et chaque colonne à une caractéristique acoustique.

Un premier aperçu statistique des données a permis d'identifier la diversité des échelles et distributions entre les différentes variables. Certaines variables, comme les coefficients MFCC, sont centrées autour de valeurs proches de zéro, tandis que d'autres, comme le spectral rolloff ou le spectral centroid, présentent des échelles beaucoup plus larges.

Analyse de la distribution des genres

L'analyse globale des distributions des variables montre une diversité importante entre les indicateurs, reflétant la variété des caractéristiques sonores selon les genres musicaux [\[2\]](#).

La majorité des variables présentent des distributions légèrement asymétriques avec une concentration des valeurs sur une plage restreinte. C'est particulièrement vrai pour les variables liées à l'énergie (rms), la tonalité (chroma_stft) et certaines des mesures spectrales (centroïde, rolloff), ce qui traduit la tendance naturelle des morceaux à se regrouper autour de profils sonores types, avec quelques extraits plus extrêmes.

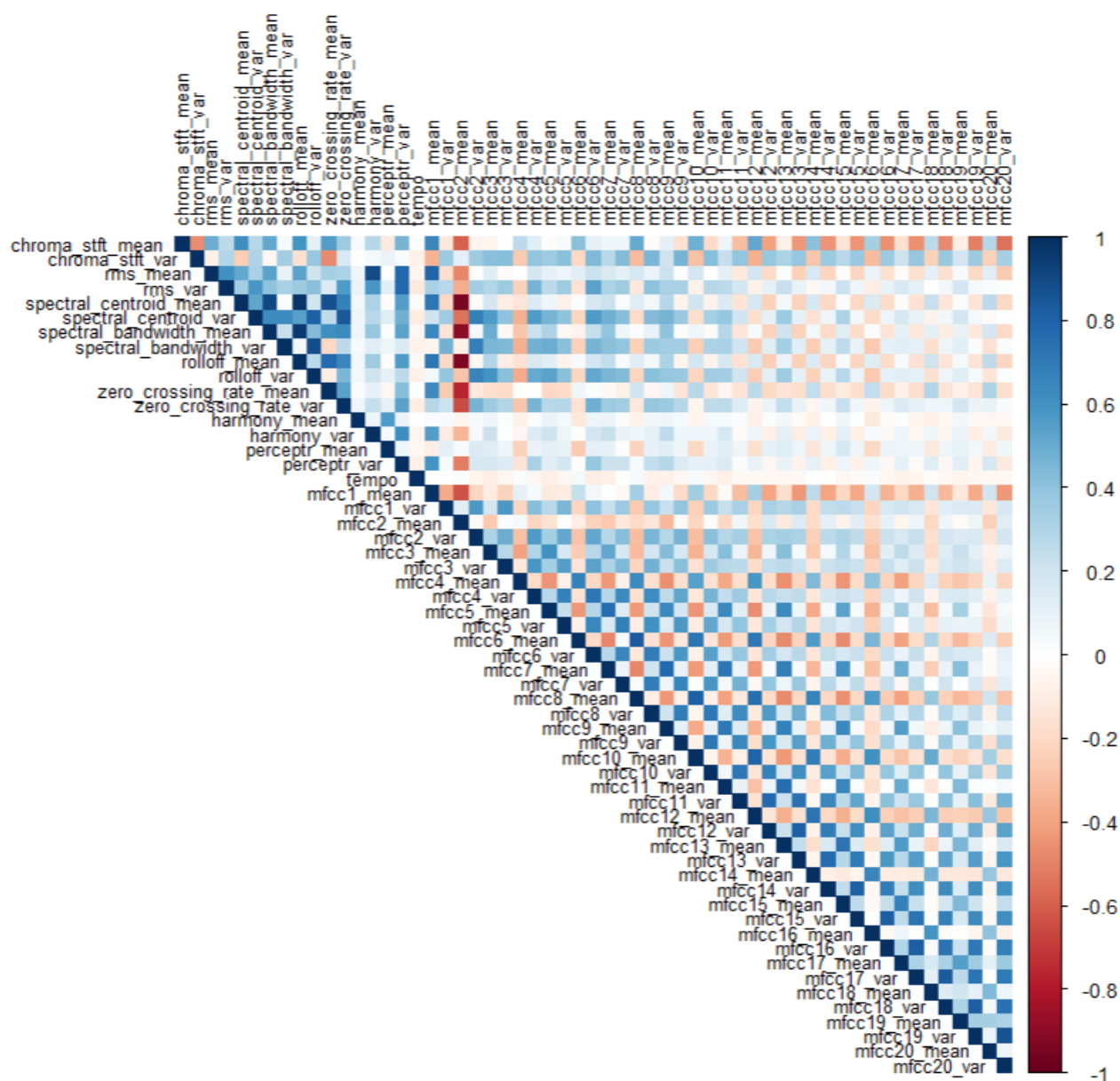
Les coefficients MFCC (moyennes et variances) se distinguent avec des distributions relativement équilibrées, souvent proches d'une forme normale, ce qui est cohérent avec leur rôle de synthèse globale du spectre sonore.

En revanche, certaines variables, comme les taux de passage par zéro (zero_crossing_rate) ou les indicateurs harmoniques, montrent des concentrations très marquées vers zéro. Cela indique que beaucoup de morceaux ont peu ou pas de contenu dans ces dimensions, ce qui est attendu pour certains genres plus mélodiques ou acoustiques.

Globalement, cette exploration met en évidence une forte hétérogénéité entre les descripteurs, traduisant la richesse et la complexité de la base de données, mais aussi la nécessité d'un prétraitement adapté pour éviter que certaines variables trop concentrées ou extrêmes ne biaisent l'apprentissage.

Corrélation entre les variables

Avant d'entraîner nos modèles, nous avons analysé les relations entre les différentes variables grâce à une **matrice de corrélation** :



L'analyse de la corrélation entre les différentes variables a permis de mettre en évidence une forte redondance entre plusieurs descripteurs audio. Cette redondance est particulièrement marquée parmi certaines familles de variables, notamment les **MFCCs**, ainsi que les indicateurs spectraux (centroïde spectral, rolloff, etc.), qui mesurent souvent des caractéristiques similaires sous des angles légèrement différents.

Une visualisation via une matrice de corrélation a montré plusieurs blocs de variables présentant des corrélations supérieures à 0.8, traduisant une forte similarité dans l'information capturée. Ce phénomène est courant dans les analyses audios, où les coefficients de type MFCC et les descripteurs spectraux capturent des aspects proches de la texture sonore ou de la distribution fréquentielle.

Par exemple, **spectral_centroid** et **rolloff** sont deux mesures fortement corrélées, toutes deux reflétant la distribution de l'énergie dans le spectre, mais à travers des méthodes de calcul différentes.

De même, plusieurs paires de coefficients MFCC adjacents présentent des corrélations élevées, reflétant la continuité spectrale et la dépendance entre bandes fréquentielles successives.

Pour limiter la redondance et simplifier le modèle, une étape de filtrage a été appliquée. Les variables présentant une corrélation absolue supérieure à 0.8 ont été identifiées. Parmi chaque paire fortement corrélée, une seule variable a été conservée. Ce processus a conduit à la suppression de **15 variables**, réduisant ainsi le nombre total de caractéristiques de **57 à 42**, ce qui améliore la stabilité du modèle et réduit le risque de surapprentissage sans perte significative d'information pertinente.

Analyse par genre

Des **boxplots par genre** ont été réalisés pour plusieurs indicateurs clés [\[3\]](#). Cette analyse montre :

Les genres comme la **musique classique** se distinguent par des valeurs globalement plus faibles sur des indicateurs liés à la densité spectrale et aux variations fréquentielles, traduisant une **structure sonore plus épurée** et une dynamique plus contrôlée. En revanche, des genres comme le **hip-hop** ou le **disco** présentent des valeurs nettement plus élevées sur la **largeur de bande spectrale** et le **taux de croisement zéro**, ce qui reflète une **densité sonore accrue** et une plus forte présence de composantes aiguës.

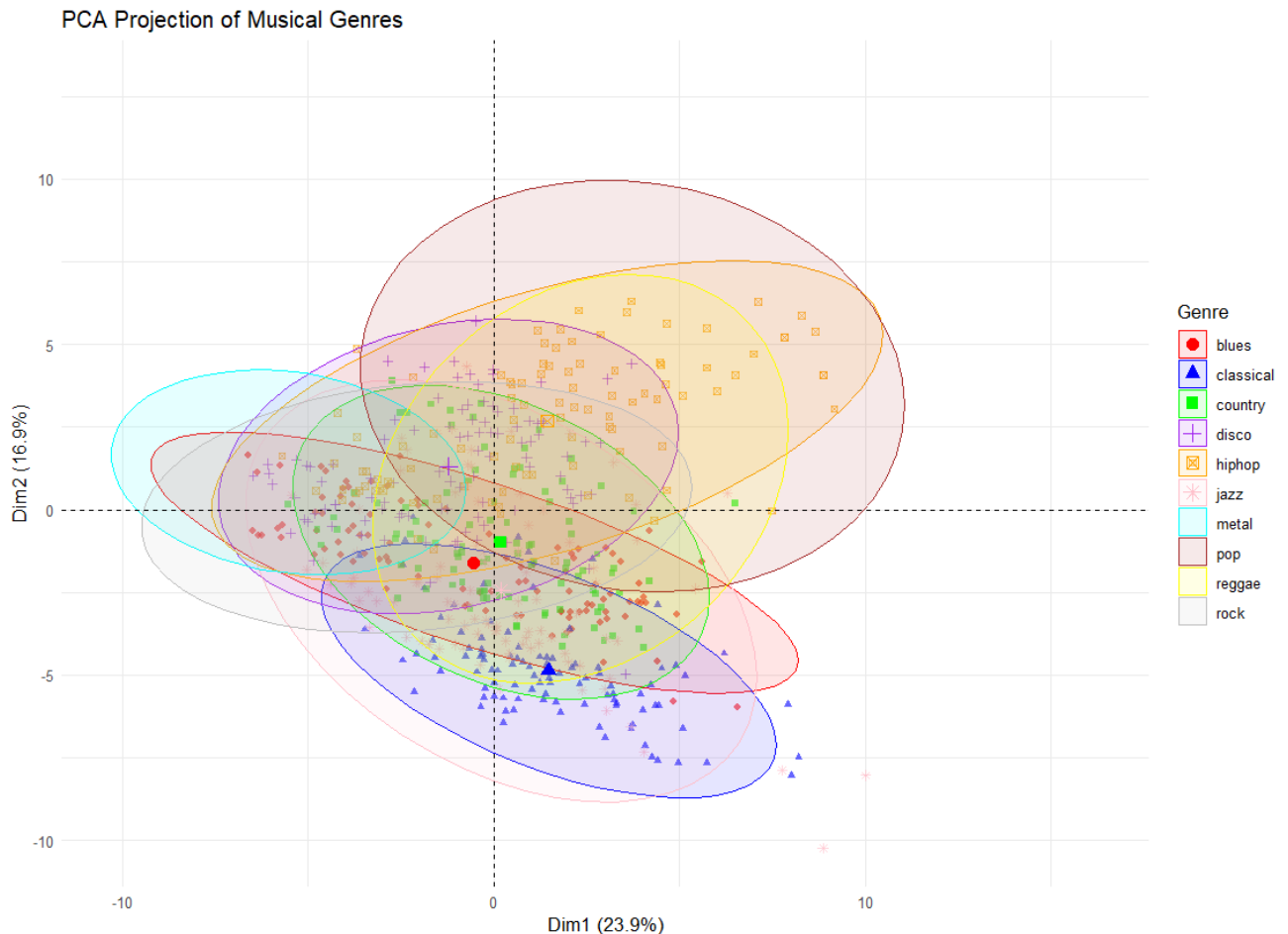
Les **MFCC**, qui capturent la texture globale du signal audio, révèlent également des profils distincts selon les genres. Cependant, certains styles comme **le rock, le disco et le country** présentent des profils relativement proches, rendant plus complexe une séparation nette entre eux.

- Les genres comme classical, jazz et metal présentent des distributions plus concentrées avec une variabilité interne plus faible.
- Les genres modernes et électroniques (hip-hop, pop, disco) affichent une plus forte dispersion, signe d'une plus grande diversité de production.

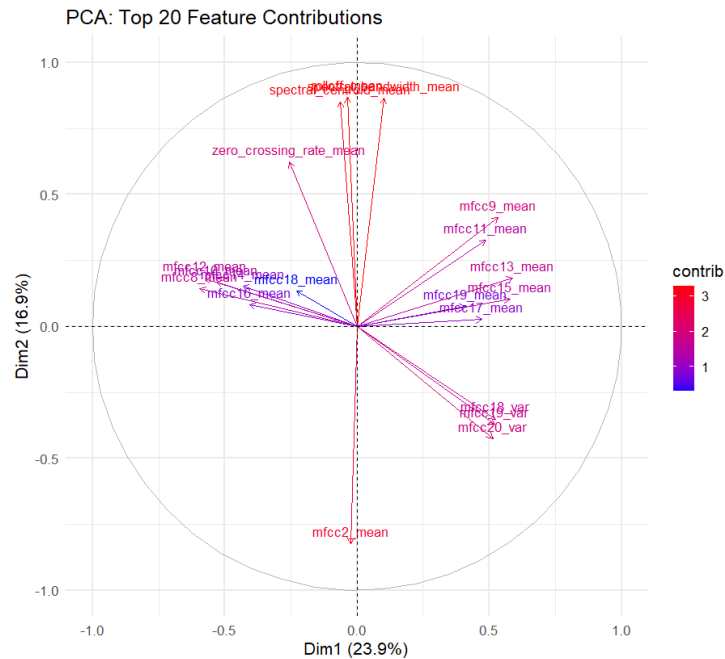
Analyse en composantes principales

L'analyse de la variance expliquée montre que les deux premières composantes principales capturent environ **40,8%** de la variance totale, indiquant qu'une part importante mais non suffisante de l'information est concentrée dans ces axes. Il est donc probable que plusieurs dimensions soient nécessaires pour différencier pleinement les genres.

Le graphique de projection montre que certaines classes, comme **classical**, se distinguent relativement bien des autres, se regroupant dans une zone distincte. Une autre remarque est que les genres comme le reggae et la pop ont une forme plus symétrique, couvrant un espace plus large sur le graphique. Les points de données de ces deux genres sont donc très mélangés avec d'autres, ce qui pourrait compliquer la classification de ces genres au final.



Les indicateurs liés au spectre, comme `spectral_bandwidth_mean`, `spectral_centroid_mean` et `rolloff_mean`, jouent un rôle majeur dans la première composante. Les coefficients MFCC jouent un rôle clé dans les deux dimensions, les valeurs moyennes ayant plus d'influence sur la deuxième dimension et la variance contribuant de manière égale aux deux axes.



Machine learning

Séparation des données et prétraitement des variables

La séparation des données a été réalisée en utilisant un échantillonnage stratifié afin de garantir une répartition équilibrée des genres musicaux entre l'ensemble d'apprentissage (80%) et l'ensemble de test (20%). Cette approche est essentielle pour éviter un déséquilibre de classes qui pourrait biaiser les performances du modèle, notamment en présence de genres sous-représentés.

La stratification préserve la proportion de chaque genre dans les deux ensembles.

En ce qui concerne le prétraitement des variables, un processus de centrage et de mise à l'échelle (scaling) a été appliqué aux caractéristiques numériques.

- Le centrage soustrait la moyenne de chaque variable
- La mise à l'échelle divise chaque variable par son écart-type, permettant de ramener toutes les variables sur une échelle comparable.
- Ce type de normalisation est particulièrement crucial pour les algorithmes sensibles à la distance, comme les SVM ou k-NN.

Analyse des Résultats de Modélisation et Importance des Variables

Les performances globales montrent que le SVM est le modèle le plus performant avec une précision de 74,5%, suivi de près par XGBoost à 73,5% et Random Forest à 71%. Le KNN, avec une précision de 66,5%, est nettement en retrait. Ces résultats confirment que des modèles plus complexes, comme SVM et XGBoost, parviennent mieux à capturer les relations non linéaires entre les caractéristiques audio et les genres musicaux. Toutes les matrices de confusion sont trouvées dans l'annexe [\[4\]](#).

Random Forest

- Points forts : bonne performance globale, notamment sur les genres bien représentés comme classical, jazz et metal.
- Faiblesses : des confusions marquées entre des genres proches comme rock et country ou reggae et hiphop.
- L'importance des variables montre que les caractéristiques spectrales (chroma_stft_mean, spectral_bandwidth_mean) et les MFCC (mfcc4_mean, mfcc5_var) jouent un rôle clé. Les variables liées au timbre (zero_crossing_rate_mean, spectral_centroid_var) sont également significatives.

SVM

- Points forts : très bonnes performances sur les genres classical, jazz et pop, avec une très bonne spécificité pour presque tous les genres.
- Faiblesses : difficultés à différencier certains genres aux caractéristiques proches, comme disco et rock ou reggae et hiphop.
- Ce modèle semble bien tirer parti de la normalisation appliquée sur les caractéristiques, ce qui est typique des modèles SVM utilisant un noyau radial.

XGBoost

- Points forts : bonne performance générale, avec des scores équilibrés sur la plupart des genres.
- Faiblesses : encore quelques confusions sur des genres voisins (disco/rock, reggae/hiphop).
- L'analyse d'importance des variables montre une forte contribution de **perceptr_var**, **chroma_stft_mean**, **spectral_bandwidth_mean**, et plusieurs MFCC (mfcc4_mean, mfcc9_mean, mfcc17_mean). Cela confirme l'importance des descripteurs de timbre et de la texture spectrale pour la classification.

Importance des Variables

L'importance des variables, à la fois dans Random Forest et XGBoost, converge vers un même constat : les caractéristiques spectrales globales (chroma, spectral_bandwidth, spectral_centroid) et les coefficients MFCC sont les plus déterminantes. Ces variables décrivent la texture harmonique, la richesse spectrale et les aspects timbraux, qui sont des signatures fortes permettant de distinguer les genres.

Top 5 variables selon XGBoost :

- perceptr_var
- chroma_stft_var
- chroma_stft_mean
- spectral_bandwidth_mean
- mfcc4_mean

Top 5 variables selon Random Forest :

- chroma_stft_mean
- perceptr_var
- chroma_stft_var
- rms_mean
- spectral_bandwidth_mean

Ces résultats sont cohérents avec les connaissances en analyse audio, où les variations spectrales et les caractéristiques de texture (MFCC, zero crossing rate) sont reconnues comme des indicateurs clés pour l'identification de genres musicaux.

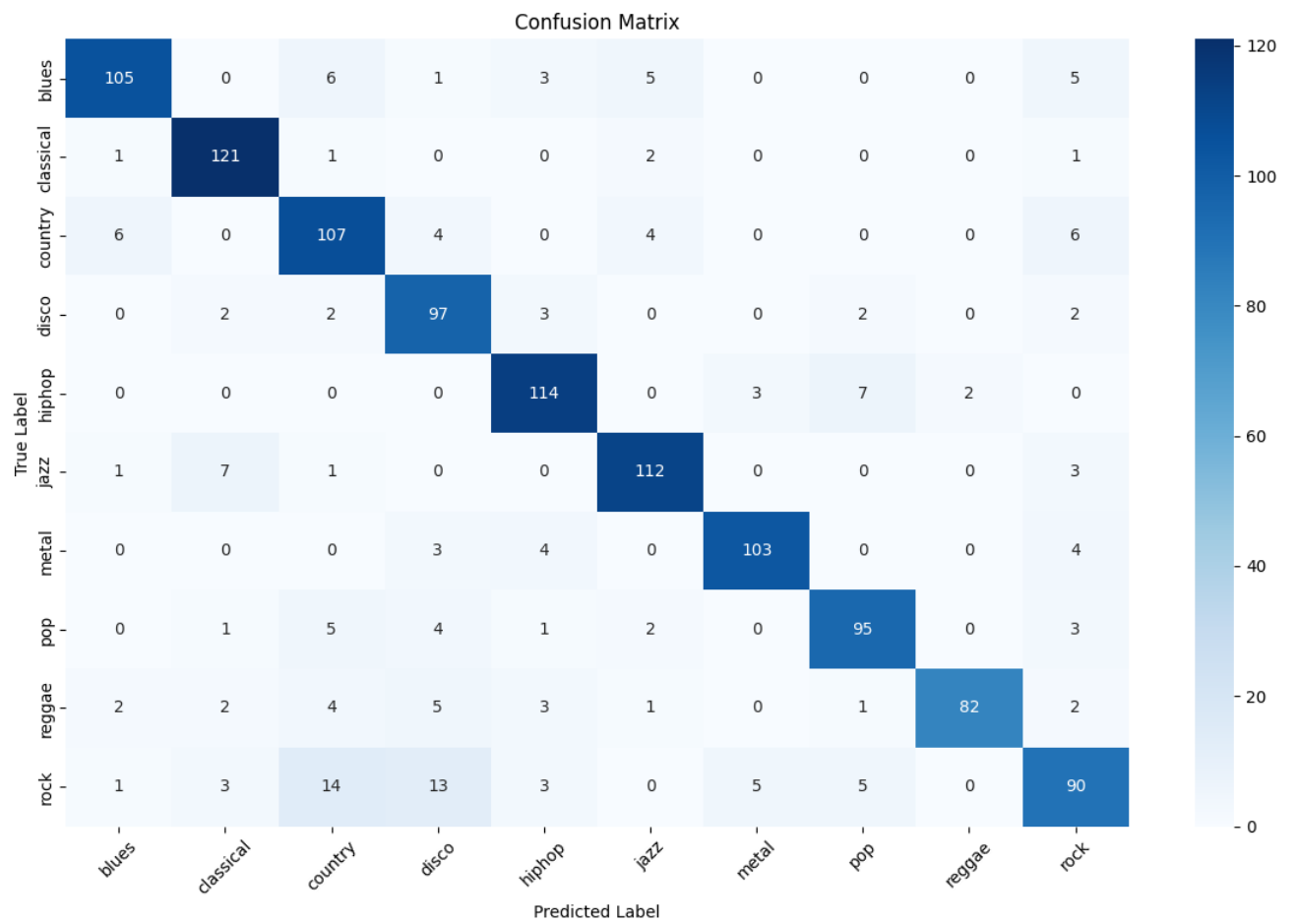
Note : Une analyse exploratoire ciblée a également été réalisée sur les six caractéristiques les plus importantes identifiées par XGBoost (perceptr_var, chroma_stft_var, chroma_stft_mean, spectral_bandwidth_mean, mfcc4_mean et rms_var). L'objectif était de mieux comprendre comment ces variables varient selon les genres et si elles contribuent effectivement à leur différenciation. Cette analyse a confirmé que certaines variables, en particulier celles liées à la texture spectrale et au timbre, permettent de mieux discriminer certains genres comme classical et jazz, tandis que d'autres genres plus proches (pop, disco, reggae) restent plus difficiles à séparer [\[5\]](#).

Bonus : Modèle CNN pour la classification de genres musicaux

En complément des modèles classiques, nous avons construit et entraîné un modèle CNN (Convolutional Neural Network) pour la classification de genres musicaux à partir de spectrogrammes Mel [6]. Ce modèle s'inscrit comme une approche complémentaire aux algorithmes classiques testés précédemment (Random Forest, SVM, KNN, XGBoost).

Prétraitement des données

- **Extraction des spectrogrammes Mel :**
Chaque fichier audio est découpé en segments de 5 secondes.
Pour chaque segment, un spectrogramme Mel est calculé, transformé en échelle logarithmique (dB) puis normalisé.
- **Formatage :**
Les spectrogrammes sont stockés sous forme de tenseurs 4D : (n_samples, 1, 128, 216) (canal unique, 128 bandes de fréquences, 216 fenêtres temporelles).
- **Architecture :**
 - 4 blocs convolutifs (Conv2D, BatchNorm, ReLU, MaxPool, Dropout)
 - 3 couches denses (512 → 256 → 10 genres)
- **Entraînement :**
 - **50 époques**, optimiseur Adam, batch size 32.
 - Meilleure précision test : **86,39%** (epoch 49).
- **Insights :**
 - Excellente précision pour **classical, metal et blues**.
 - Plus de confusion entre genres similaires comme **rock et reggae**.
- **Modèle sauvegardé :** complete_CNN_model.pth



Annexes

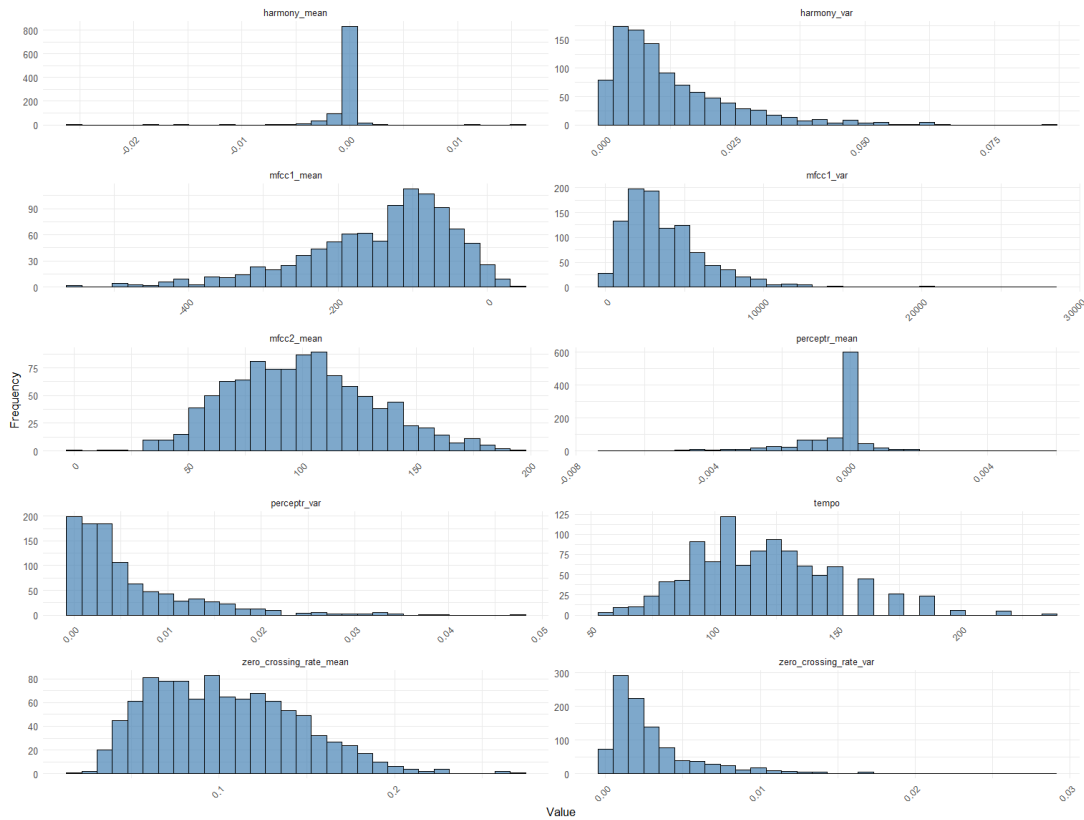
1. Descriptions

Feature	Description
Nom de Fichier	Nom de Fichier
Longueur	Longueur
chroma_stft_mean	Intensité Moyenne des Notes - Mesure la fréquence moyenne d'apparition de chaque note musicale (comme C, D, E) dans la chanson. Donne une idée de l'harmonie ou de la tonalité générale de la chanson.
chroma_stft_var	Variabilité de l'Intensité des Notes - Montre à quel point la présence des notes musicales change tout au long de la chanson, indiquant s'il y a des variations importantes dans l'harmonie.
rms_mean	Moyenne de l'Intensité Sonore - Représente le niveau moyen de volume de la chanson.
rms_var	Variabilité de l'Intensité Sonore - Montre la variation du volume tout au long de la chanson, capturant des sections qui sont particulièrement calmes ou fortes.
spectral_centroid_mean	Moyenne de la Brillance - Indique la "brillance" moyenne de la chanson. Les valeurs plus élevées suggèrent des sons aigus, tandis que des valeurs plus faibles indiquent des sons plus graves et riches en basses.
spectral_centroid_var	Variabilité de la Brillance - Montre comment la brillance change tout au long de la chanson. Les chansons avec une forte variabilité peuvent avoir des transitions dynamiques entre des tons clairs et profonds.
spectral_bandwidth_mean	Moyenne de la Richesse Sonore - Mesure la gamme de fréquences dans la chanson, en lien avec le timbre et la "richesse" perçue du son.
spectral_bandwidth_var	Variabilité de la Richesse Sonore - Indique la manière dont la richesse sonore change, suggérant si la chanson alterne entre des textures simples et complexes.
rolloff_mean	Moyenne de l'Équilibre Graves-Aigus - Indique où se concentre la majeure partie de l'énergie sonore en termes de hauteur. Des valeurs faibles suggèrent une prédominance de basses, tandis que des valeurs élevées indiquent un accent sur les aigus.
rolloff_var	Variabilité de l'Équilibre Graves-Aigus - Montre comment cet équilibre change tout au long de la chanson, indiquant des passages alternant entre basses et aigus.
zero_crossing_rate_mean	Moyenne de la Percussivité - Mesure à quel point la chanson est "hachée" ou "bruyante", souvent associée à de la musique plus rapide et percussive.
zero_crossing_rate_var	Variabilité de la Percussivité - Montre les variations de la percussivité, capturant les transitions entre des sections plus fluides et plus rythmées ou percussives.
harmony_mean	Moyenne du Contenu Harmonique - Reflète le niveau global d'harmonie dans la chanson, ou la présence d'éléments stables et tonaux comme des accords ou des notes tenues.
harmony_var	Variabilité du Contenu Harmonique - Indique la mesure dans laquelle le niveau d'harmonie change. Une forte variabilité suggère des changements d'accords ou de tonalité, comme dans les chansons qui varient harmonieusement.
perceptr_mean	Moyenne de la Netteté - Mesure la clarté ou le "tranchant" du son. Des valeurs plus élevées suggèrent un son plus net et plus défini.
perceptr_var	Variabilité de la Netteté - Montre la façon dont cette netteté change au fil du temps, en reflétant les sections qui sont plus fluides par rapport à celles qui sont plus définies.
tempo	Tempo - La vitesse de la chanson, mesurée en battements par minute (BPM). Des valeurs plus élevées signifient une chanson plus rapide, tandis que des valeurs plus faibles indiquent une chanson plus lente.

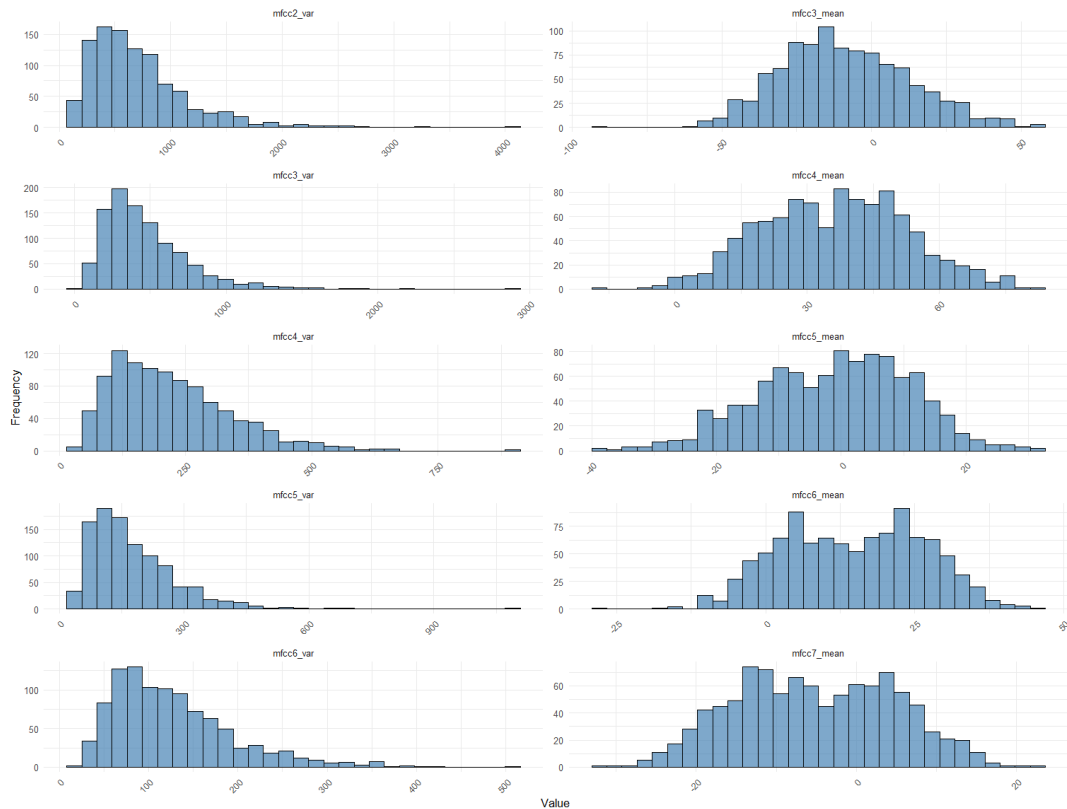
MFCCs	Coefficients Cepstraux en Fréquence Mel (MFCCs) sont des caractéristiques utilisées en analyse audio pour capturer les caractéristiques sonores uniques d'un enregistrement. Ils représentent la manière dont l'énergie sonore est distribuée sur les fréquences en fonction de l'audition humaine. Les premiers MFCCs (mfcc1 à mfcc5) capturent les caractéristiques générales, tandis que les MFCCs suivants (mfcc5+) saisissent des détails plus fins et la texture. Ensemble, les MFCCs fournissent un résumé compact ou "empreinte" des qualités tonales et de timbre d'une chanson, les rendant utiles pour des tâches comme la classification de genre.
mfcc1_mean à mfcc5_mean	Moyenne de la Forme Sonore Générale - Ces premiers MFCCs capturent les caractéristiques générales du son, telles que sa qualité tonale globale, sa hauteur et son énergie. Ils sont liés à la forme et à la profondeur tonale principale de la chanson.
mfcc1_var à mfcc5_var	Variabilité de la Forme Sonore Générale - Montre dans quelle mesure ces caractéristiques générales changent tout au long de la chanson, en reflétant des variations importantes de ton ou d'énergie.
mfcc6_mean à mfcc20_mean	Moyenne des Caractéristiques Tonales Détaillées - Ces coefficients saisissent des détails plus fins, comme des textures subtiles et des qualités de timbre. Ils aident à différencier les éléments uniques du son, en capturant des nuances qui peuvent varier selon le genre.
mfcc6_var à mfcc20_var	Variabilité des Caractéristiques Tonales Détaillées - Indique à quel point ces détails fins changent tout au long de la chanson, en montrant des variations de texture tonale et de complexité.

2. Distributions des colonnes

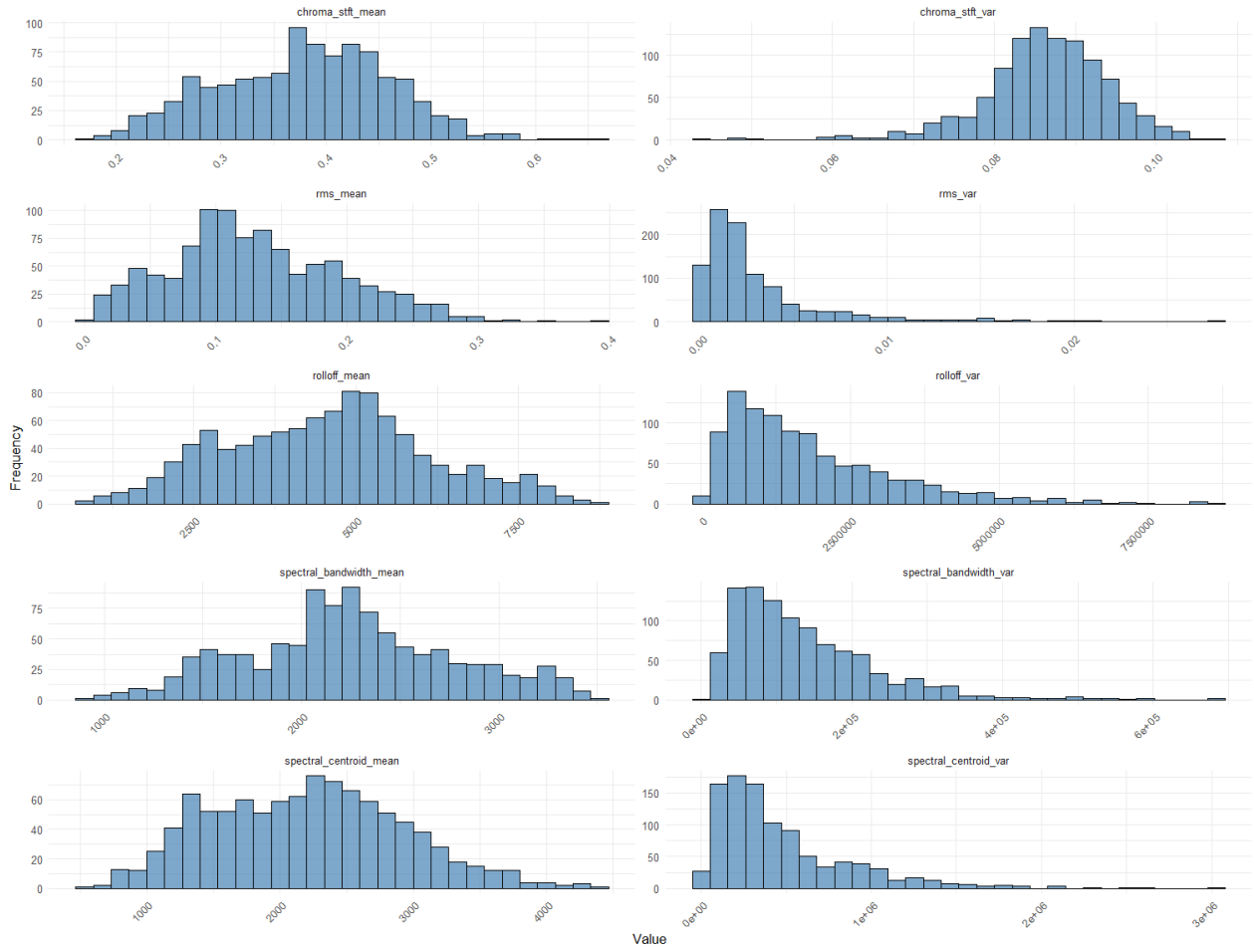
Feature Distributions (zero_crossing_rate_mean, zero_crossing_rate_var, harmony_mean, harmony_var, percept_r_mean, percept_r_var, tempo, mfcc1_mean, mfcc1_v



Feature Distributions (mfcc2_var, mfcc3_mean, mfcc3_var, mfcc4_mean, mfcc4_var, mfcc5_mean, mfcc5_var, mfcc6_mean, mfcc6_var, mfcc7_mean)

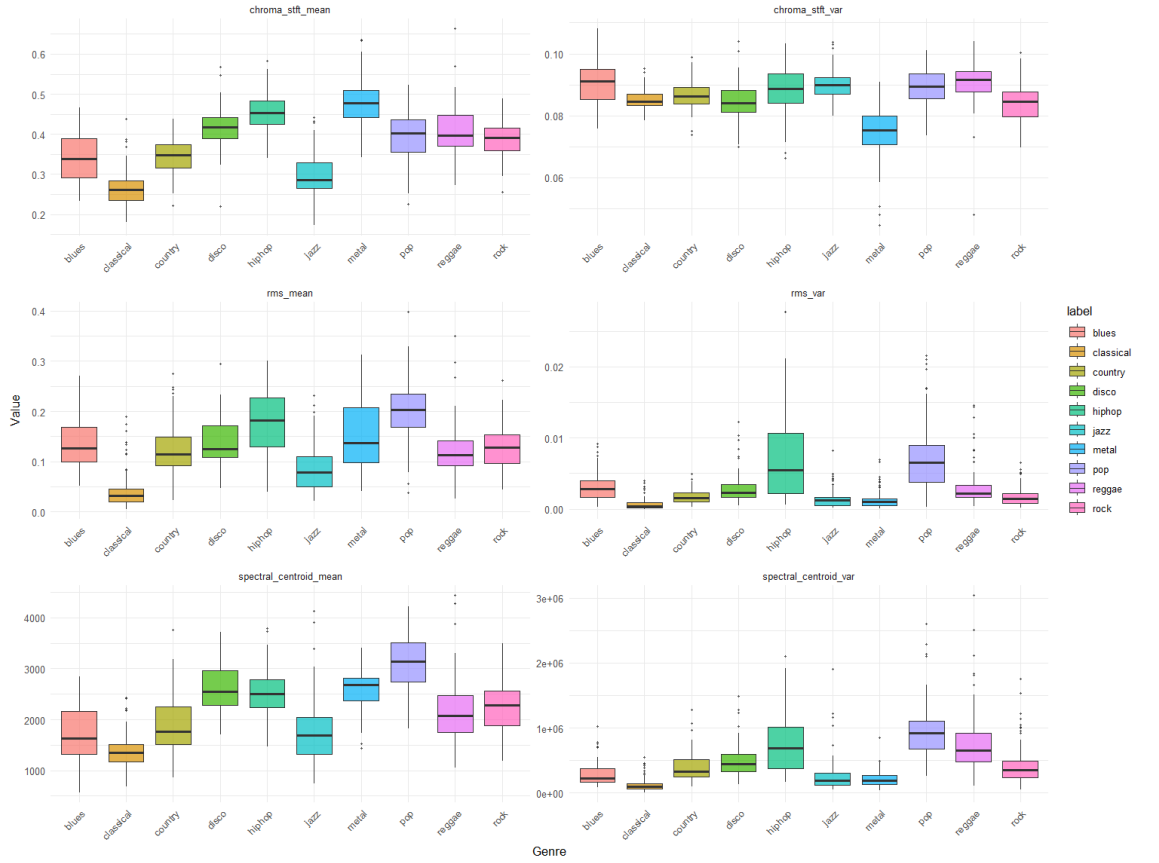


Feature Distributions (chroma_stft_mean, chroma_stft_var, rms_mean, rms_var, spectral_centroid_mean, spectral_centroid_var, spectral_bandwidth_mean, spectral_

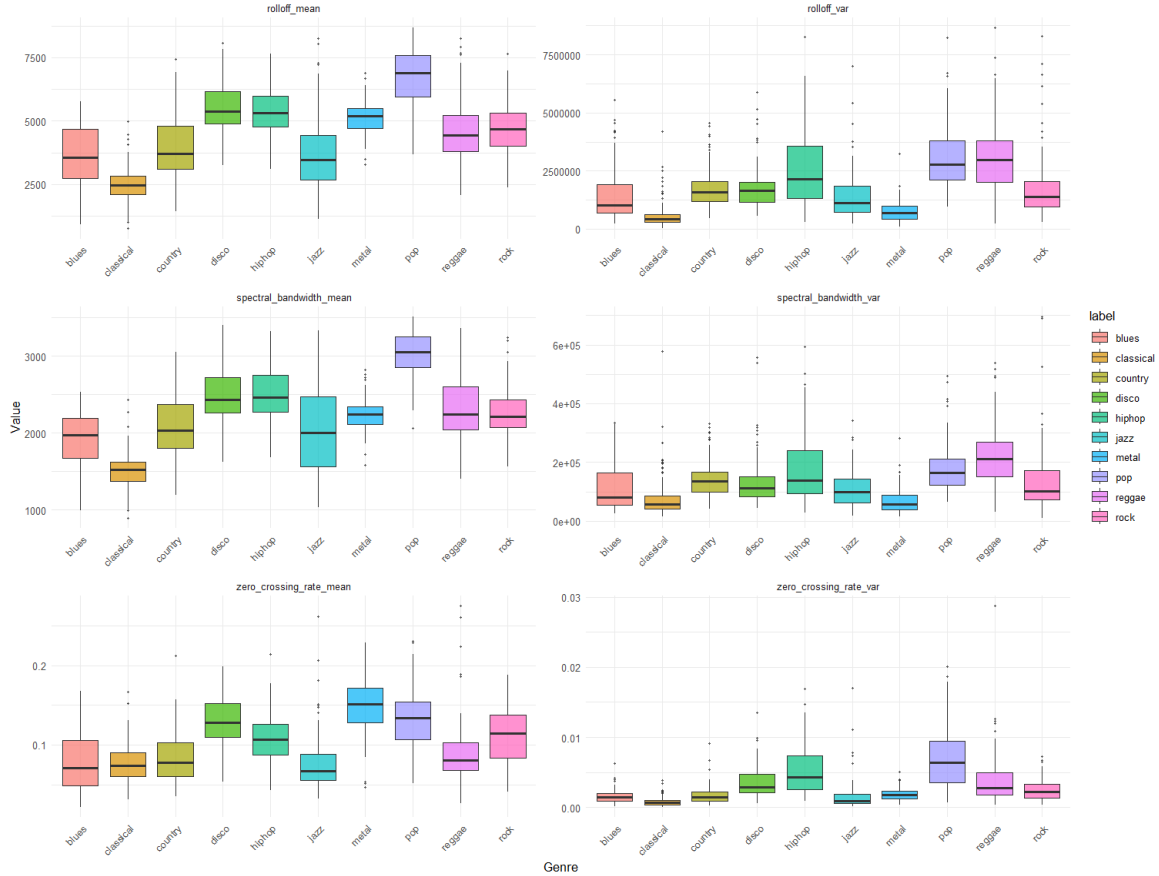


3. Analyse par genre

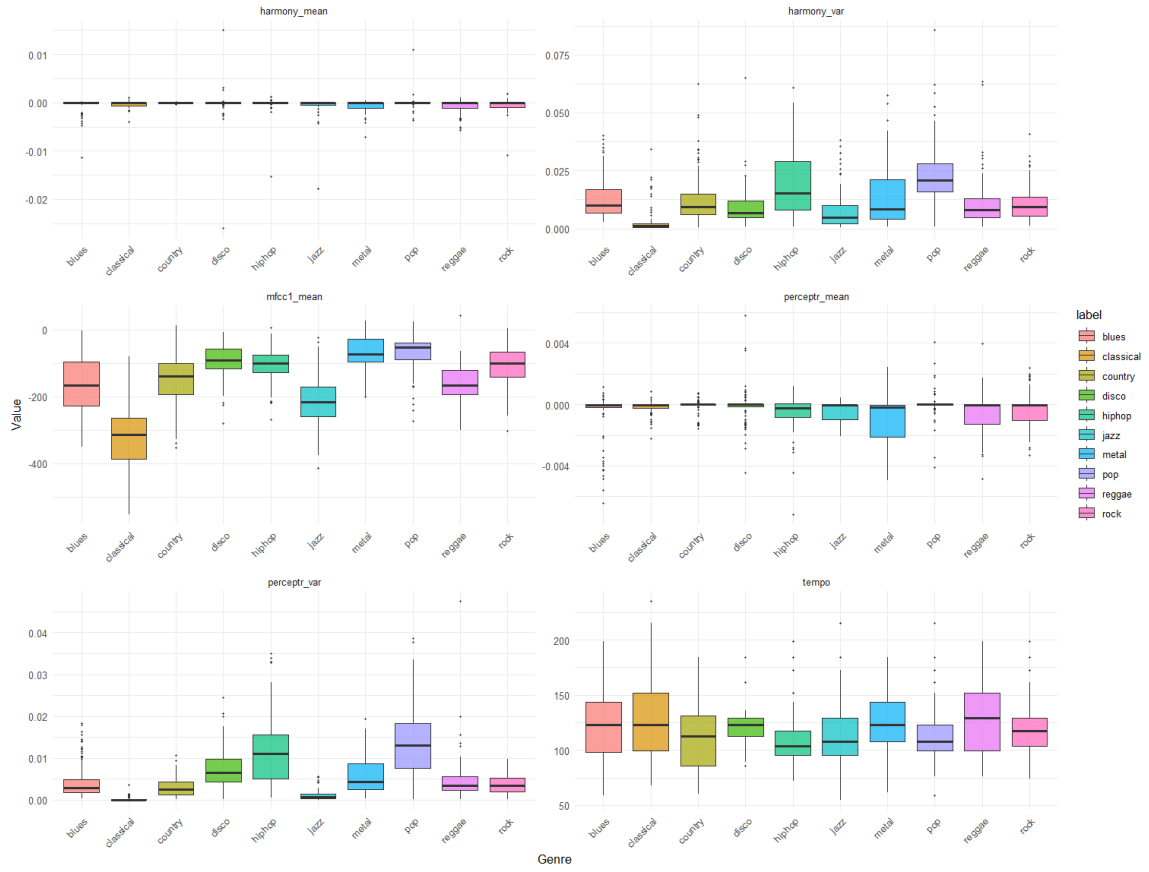
Boxplots of Features by Genre (chroma_stft_mean, chroma_stft_var, rms_mean, rms_var, spectral_centroid_mean, spectral_centroid_var)



Boxplots of Features by Genre (spectral_bandwidth_mean, spectral_bandwidth_var, rolloff_mean, rolloff_var, zero_crossing_rate_mean, zero_crossing_rate_var)



Boxplots of Features by Genre (harmony_mean, harmony_var, perceptr_mean, perceptr_var, tempo, mfcc1_mean)



4. Matrices de Confusion

Random Forest Confusion Matrix:

```
> print(confusionMatrix(pred_rf, test_labels))
```

Confusion Matrix and Statistics

	Reference									
Prediction	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	14	0	2	0	0	0	0	0	0	2
classical	0	17	0	0	0	1	0	0	0	0
country	3	0	10	0	0	0	0	1	0	1
disco	0	0	3	12	1	0	0	0	2	5
hiphop	0	0	0	3	14	0	0	1	4	0
jazz	0	3	2	0	0	17	0	0	0	1
metal	1	0	0	2	1	0	18	0	0	1
pop	0	0	1	1	3	2	0	18	0	0
reggae	1	0	1	1	1	0	0	0	13	1
rock	1	0	1	1	0	0	2	0	1	9

Overall Statistics

Accuracy : 0.71
95% CI : (0.6418, 0.7718)
No Information Rate : 0.1
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6778

SVM Confusion Matrix:

```
> print(confusionMatrix(pred_svm, test_labels))
```

Confusion Matrix and Statistics

	Reference									
Prediction	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	15	0	0	0	0	0	0	0	0	0
classical	0	19	0	0	0	1	0	0	0	0
country	4	0	14	0	0	0	0	1	1	1
disco	0	1	3	12	3	0	0	0	2	5
hiphop	0	0	0	0	13	0	0	0	3	0
jazz	0	0	0	0	0	18	0	0	0	2
metal	0	0	0	2	0	0	18	0	0	1
pop	0	0	1	0	2	1	0	18	0	0
reggae	0	0	1	1	2	0	0	1	13	2
rock	1	0	1	5	0	0	2	0	1	9

Overall Statistics

Accuracy : 0.745
95% CI : (0.6787, 0.8039)
No Information Rate : 0.1
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.7167

KNN Confusion Matrix:

```
> print(confusionMatrix(pred_knn, test_labels))
```

Confusion Matrix and Statistics

	Reference									
Prediction	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	13	0	0	0	1	0	0	0	0	0
classical	0	18	1	0	0	2	0	0	0	0
country	4	0	11	1	1	1	0	0	0	5
disco	1	0	4	16	1	2	1	0	3	4
hiphop	1	0	0	0	11	0	0	0	4	0
jazz	0	1	0	0	0	13	0	0	0	1
metal	0	0	0	0	0	0	12	0	0	0
pop	0	0	0	0	5	2	0	19	2	0
reggae	1	0	3	1	1	0	1	1	11	1
rock	0	1	1	2	0	0	6	0	0	9

Overall Statistics

Accuracy : 0.665
95% CI : (0.595, 0.73)
No Information Rate : 0.1
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6278

XGBoost Confusion Matrix:

```
> print(confusionMatrix(pred_xgb_factor, test_1
```

Confusion Matrix and Statistics

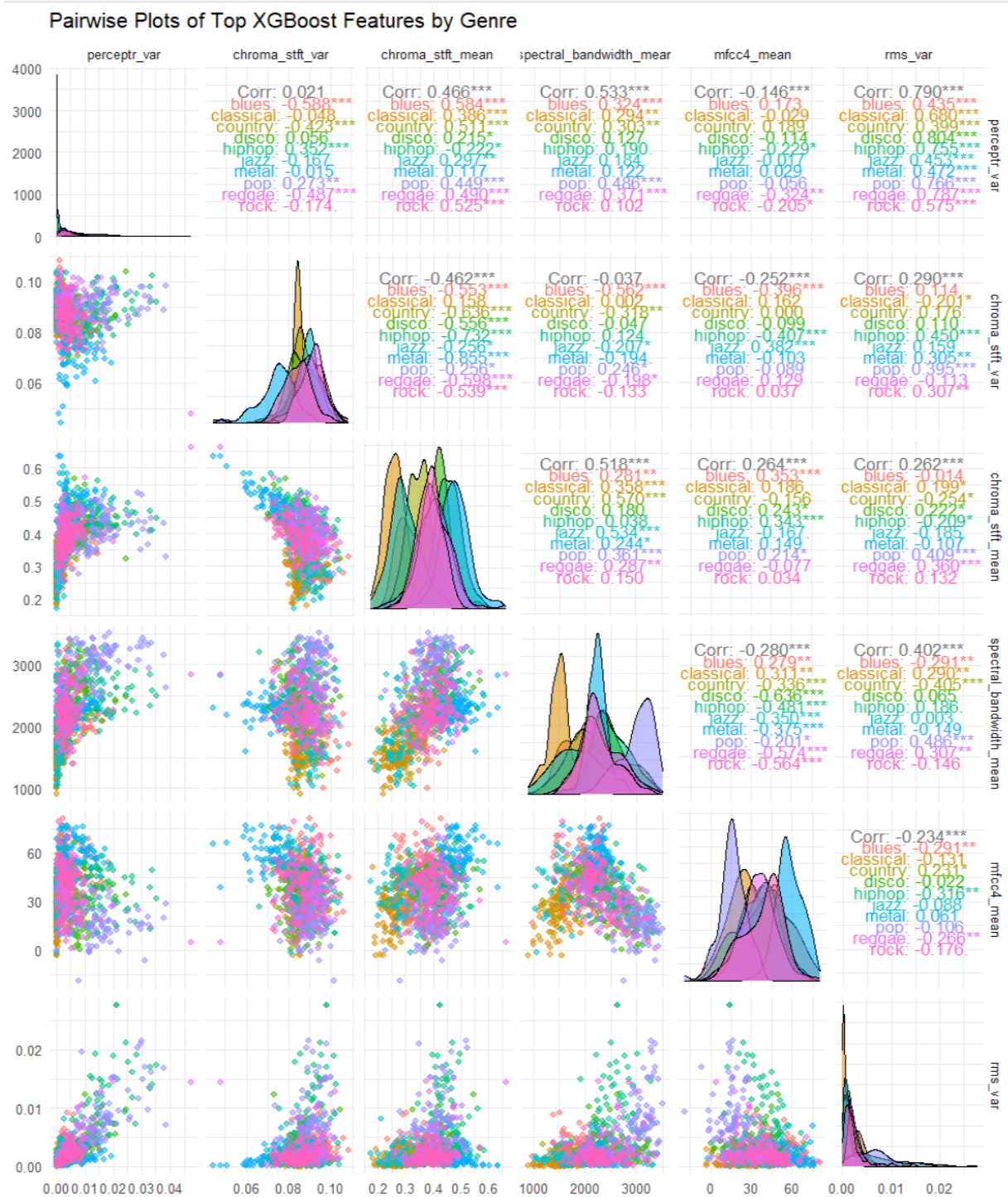
	Reference										
Prediction	0	1	2	3	4	5	6	7	8	9	
0	13	0	1	0	0	0	0	0	0	1	
1	0	17	0	0	0	0	0	0	0	0	
2	4	0	12	0	0	1	0	1	0	1	
3	0	0	0	14	2	0	0	0	2	5	
4	0	0	0	2	16	0	0	0	4	0	
5	1	1	1	0	0	16	0	0	0	2	
6	1	0	0	1	0	0	17	0	0	0	
7	0	1	2	1	2	3	0	19	0	0	
8	1	0	2	1	0	0	0	0	14	2	
9	0	1	2	1	0	0	3	0	0	9	

Overall Statistics

Accuracy : 0.735
95% CI : (0.6681, 0.7948)
No Information Rate : 0.1
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.7056

5. Tableau des corrélations des features les plus importantes.



6. Mel Spectrogrammes

