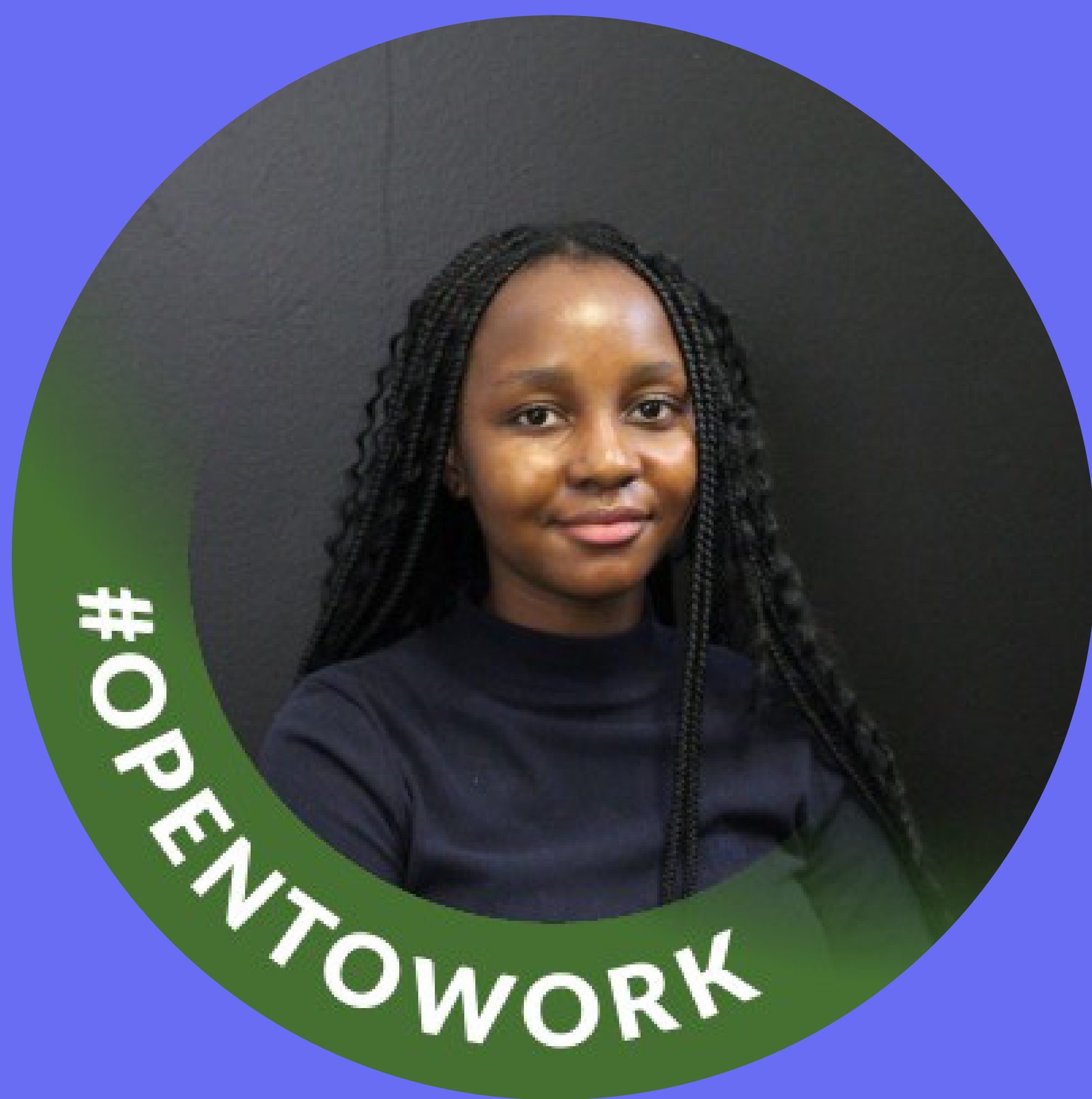


# PROJET FINAL – DATA SCIENCE

## Score Éditorial Anime



**NETTOYER ET ORGANISER DES DONNÉES BRUTE**

**Par:**

**CHELSEA TCHOUMBE  
BACHELOR WEB1  
HETIC**



# OBJECTIF: DE LA DONNEE BRUTE A L'OUTIL STRATÉGIQUE

## MINDSET

Adopter une posture de Data Scientist pour analyser et nettoyer les animes

## ENJEU

Transformer une base de données desordonnée en un outil comprehensive et propre

## METHODOLOGIE

Nettoyage -Analyse-Statistiques  
puis Visualisation

# Bibliothèques Clés : Les piliers de l'analyse

## Pandas

L'outil indispensable pour le Data Wrangling. Utilisé pour le chargement, le tri et le nettoyage rigoureux des données du catalogue.

## Matplotlib & Seaborn

Le duo gagnant pour la Data Visualization. Permet de générer des histogrammes et des graphiques statistiques clairs et parlants.

## WordCloud

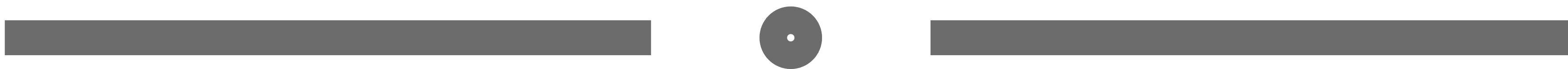
Utilisé pour extraire les tendances thématiques. Implémentation réussie pour visualiser les mots-clés dominants des animés.

# Nettoyage des Données : Garantir la fiabilité

01

## Standardisation

Conversion des notes et épisodes en formats numériques exploitables via la bibliothèque Pandas.



02

## Qualité des Données

Suppression systématique des doublons et gestion rigoureuse des valeurs manquantes pour éviter les biais.

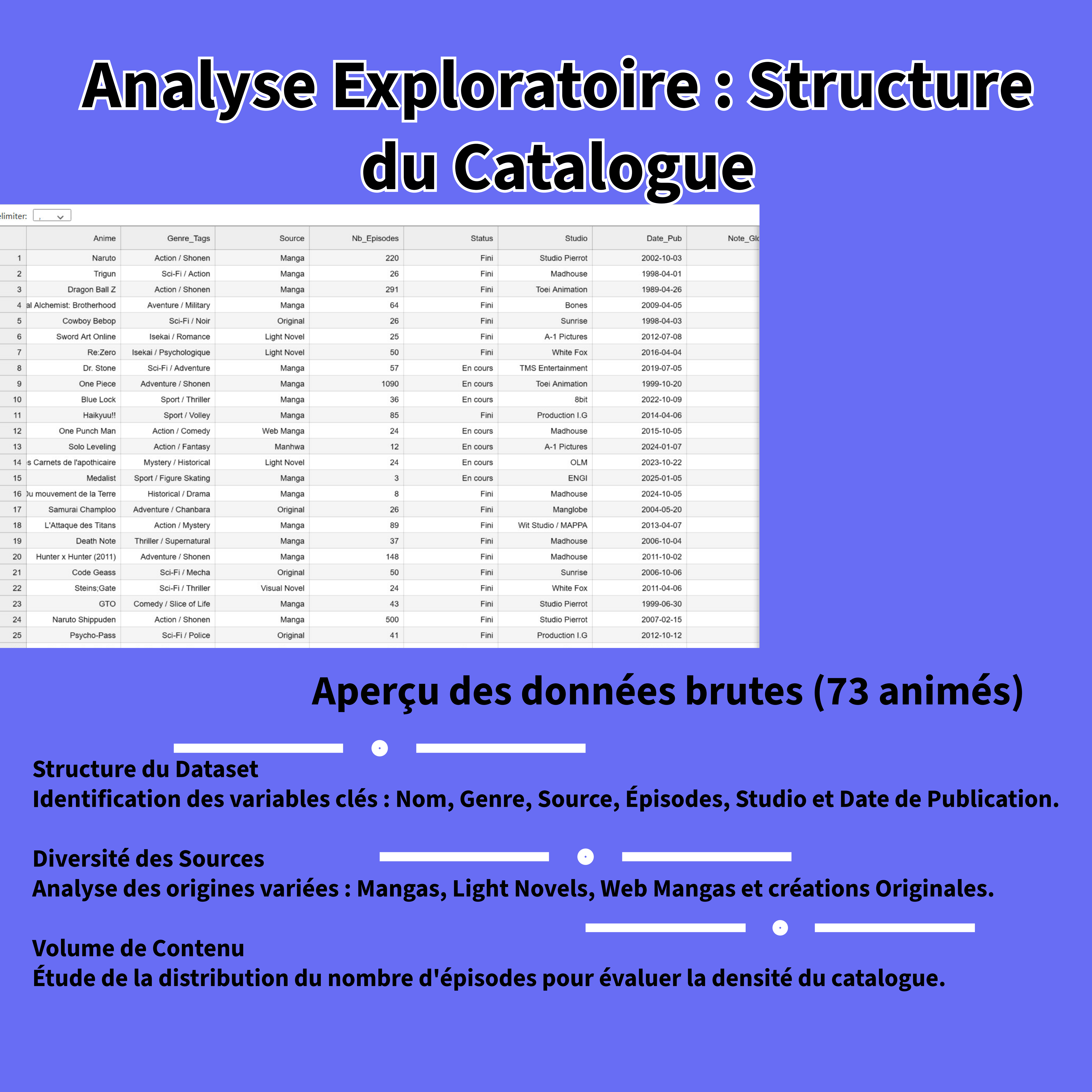


03

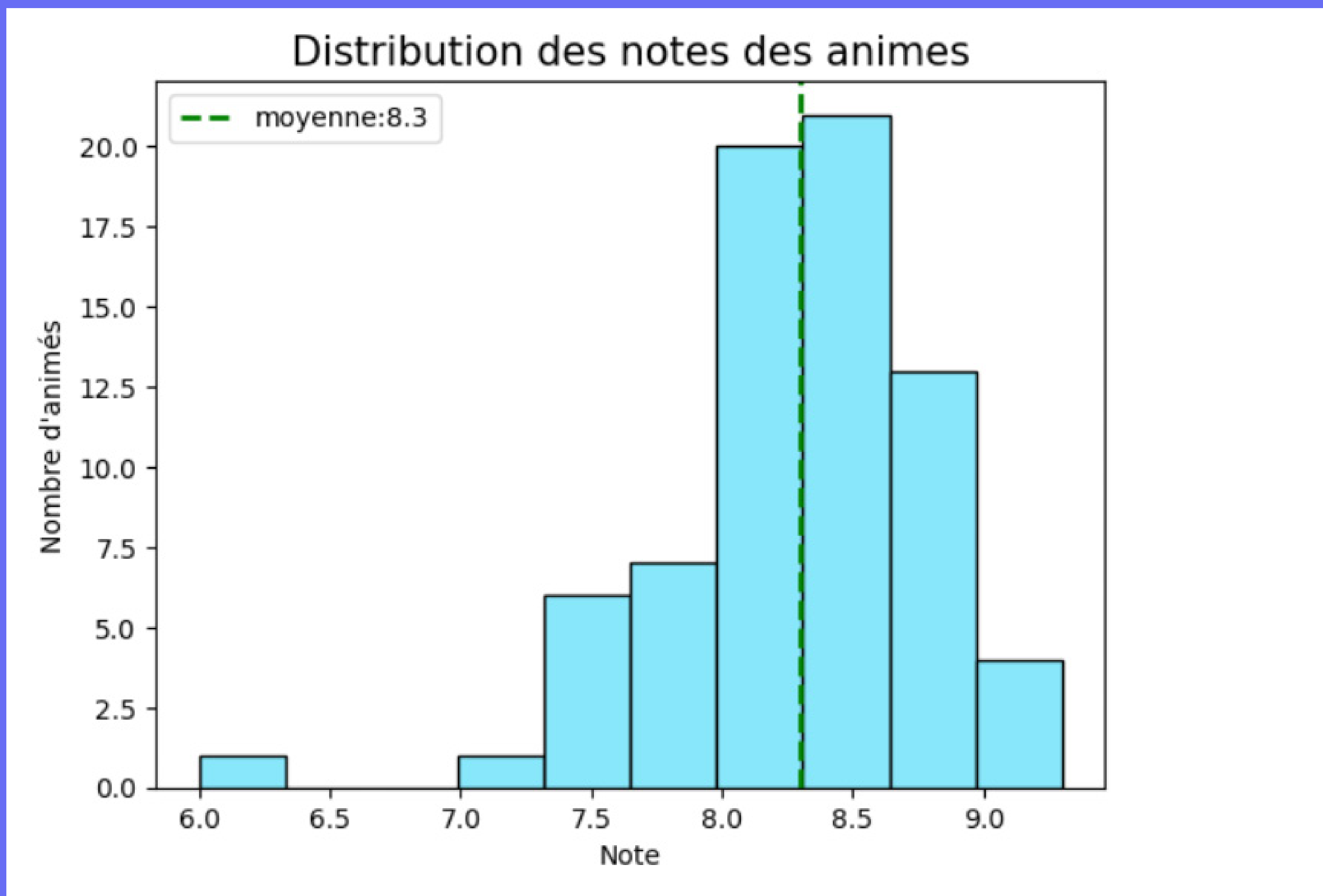
## Professionalisme

Formatage des noms d'animés en MAJUSCULES pour une présentation uniforme et professionnelle.





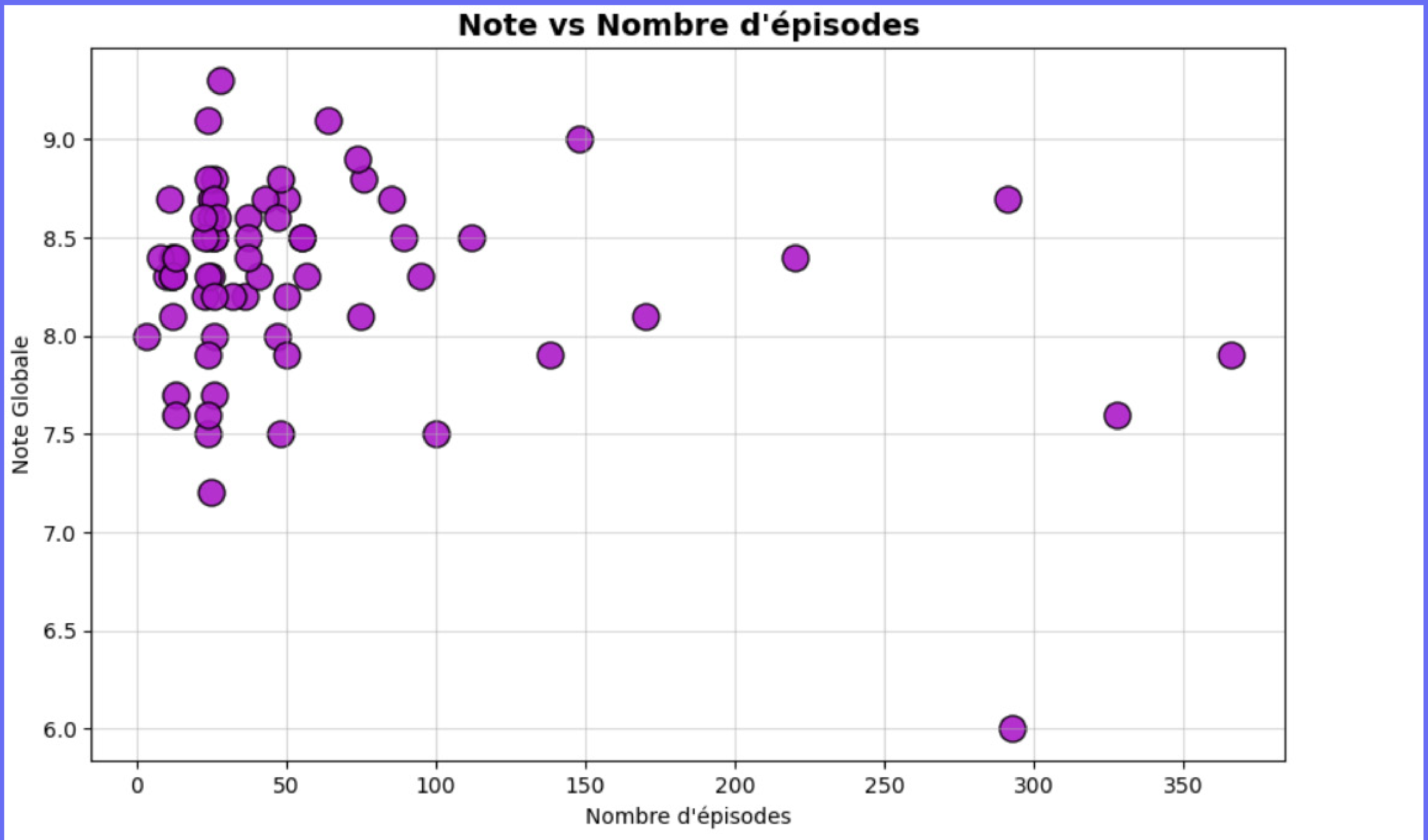
# HISTOGRAMME



Analyse de la  
répartition des notes

:pour identifier les tendances de  
qualité globale du catalogue de 73  
animés.

# NOTE GLOBALE VS NOMBRE D'EPISODE



**NOMBRE D'EPISODE  $\leq$  400**



# TOP 10 DE MES ANIMES

J'ai fais un trie parmi mes  
animes grace qui m'a sortir les  
TOP 10 de mes animes selon la  
note globale



LA 1ere anime selon les votes du public

## TOP 10 ANIMES:

Anime	Note_Globale
FRIEREN: BEYOND JOURNEY'S END	9.3
STEINS;GATE	9.1
FULLMETAL ALCHEMIST: BROTHERHOOD	9.1
HUNTER X HUNTER (2011)	9.0
MONSTER	8.9
COWBOY BEBOP	8.8
LES CARNETS DE L'APOTHICAIRE	8.8
VINLAND SAGA	8.8
HAJIME NO IPPO	8.8
OSHI NO KO	8.7



# Le Défi Technique:

La partie Classification a été l'étape la plus technique et la plus difficile du projet. L'enjeu était de transformer des données qualitatives en prédictions binaires fiables.

```
[148]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
df["Ecart"] = df["Note_Meilleur_Ep"] - df["Note_Pire_Ep"]
df["Regularite"] = 10 - df["Ecart"]
df["Excellent"] = (df["Note_Globale"] >= 8.5).astype(int)
X = df[["Note_Globale", "Regularite", "Nb_Episodes"]]
y = df["Excellent"]
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y)
model = LogisticRegression(class_weight='balanced')
model.fit(X_train, y_train)
print(" Précision du modèle (Accuracy) :", accuracy_score(y_test, y_pred))
print("\n Matrice de confusion :\n", confusion_matrix(y_test, y_pred))
print("\n Rapport détaillé :\n", classification_report(y_test, y_pred, zero_division=0))
```

Précision du modèle (Accuracy) : 0.9090909090909091

Matrice de confusion :

```
[[11  1]
 [ 1  9]]
```

Rapport détaillé :

	precision	recall	f1-score	support
0	0.92	0.92	0.92	12
1	0.90	0.90	0.90	10
accuracy			0.91	22
macro avg	0.91	0.91	0.91	22
weighted avg	0.91	0.91	0.91	22

Cible (Target)  
Prédire si un animé appartient à la catégorie :

"EXCELLENT"  
(Note ≥ 8.5).

# Conclusion : Vers une Stratégie Data-Driven

## Synthèse du Projet

Réussite de la transformation d'un catalogue brut de 71 animés en un véritable **outil d'analyse prédictive**. Ce projet démontre comment la Data Science peut apporter une valeur ajoutée concrète aux décisions d'acquisition de licences.

**Compétences Acquises** Maîtrise du pipeline Data Science complet.

- Nettoyage et standardisation de données réelles.
- Modélisation prédictive avec Scikit-Learn.

# Merci de votre attention !