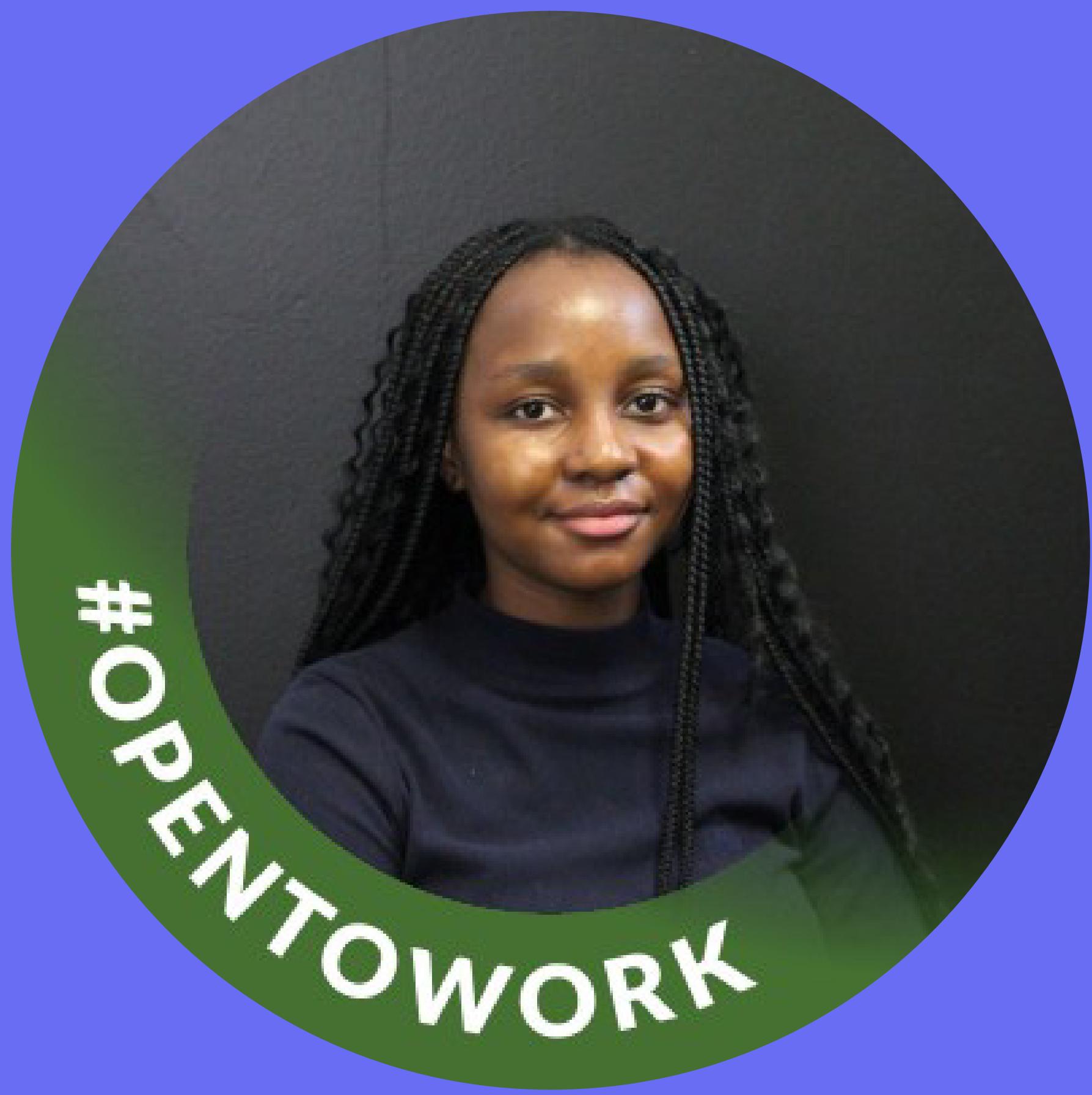


PROJET FINAL - DATA SCIENCE

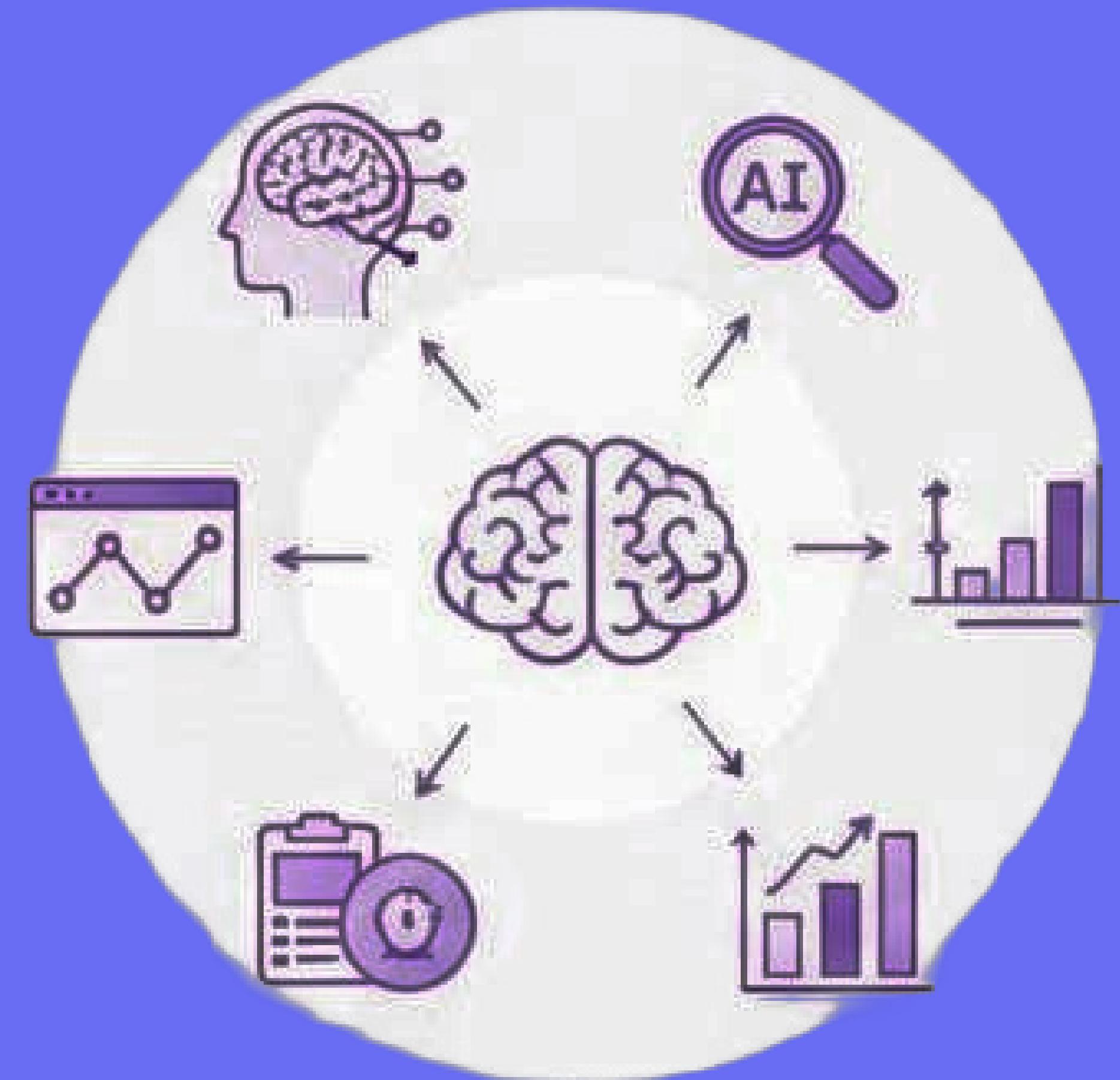
Score Editorial Anime



NETTOYER ET ORGANISER DES DONNÉES BRUTE

Par:

CHELSEA TCHOUMBE
BACHELOR WEB1
HETIC



OBJECTIF: DE LA DONNÉE BRUTE A L'OUTIL STRATÉGIQUE

MINDSET

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

EN JEU

Transformer une base de données désordonnée 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

Professionnalisme

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

Analyse Exploratoire : Structure du Catalogue

	Anime	Genre_Tags	Source	Nb_Episodes	Status	Studio	Date_Pub	Note_Globale
1	Naruto	Action / Shonen	Manga	220	Fini	Studio Pierrot	2002-10-03	
2	Trigun	Sci-Fi / Action	Manga	26	Fini	Madhouse	1998-04-01	
3	Dragon Ball Z	Action / Shonen	Manga	291	Fini	Toei Animation	1989-04-26	
4	Fullmetal Alchemist: Brotherhood	Aventure / Military	Manga	64	Fini	Bones	2009-04-05	
5	Cowboy Bebop	Sci-Fi / Noir	Original	26	Fini	Sunrise	1998-04-03	
6	Sword Art Online	Isekai / Romance	Light Novel	25	Fini	A-1 Pictures	2012-07-08	
7	Re:Zero	Isekai / Psychologique	Light Novel	50	Fini	White Fox	2016-04-04	
8	Dr. Stone	Sci-Fi / Adventure	Manga	57	En cours	TMS Entertainment	2019-07-05	
9	One Piece	Adventure / Shonen	Manga	1090	En cours	Toei Animation	1999-10-20	
10	Blue Lock	Sport / Thriller	Manga	36	En cours	8bit	2022-10-09	
11	Haikyuu!!	Sport / Volley	Manga	85	Fini	Production I.G	2014-04-06	
12	One Punch Man	Action / Comedy	Web Manga	24	En cours	Madhouse	2015-10-05	
13	Solo Leveling	Action / Fantasy	Manhwa	12	En cours	A-1 Pictures	2024-01-07	
14	Les Carnets de l'apothicaire	Mystery / Historical	Light Novel	24	En cours	OLM	2023-10-22	
15	Medalist	Sport / Figure Skating	Manga	3	En cours	ENGI	2025-01-05	
16	Le mouvement de la Terre	Historical / Drama	Manga	8	Fini	Madhouse	2024-10-05	
17	Samurai Champloo	Adventure / Chanbara	Original	26	Fini	Manglobe	2004-05-20	
18	L'Attaque des Titans	Action / Mystery	Manga	89	Fini	Wit Studio / MAPPA	2013-04-07	
19	Death Note	Thriller / Supernatural	Manga	37	Fini	Madhouse	2006-10-04	
20	Hunter x Hunter (2011)	Adventure / Shonen	Manga	148	Fini	Madhouse	2011-10-02	
21	Code Geass	Sci-Fi / Mecha	Original	50	Fini	Sunrise	2006-10-06	
22	Steins;Gate	Sci-Fi / Thriller	Visual Novel	24	Fini	White Fox	2011-04-06	
23	GTO	Comedy / Slice of Life	Manga	43	Fini	Studio Pierrot	1999-06-30	
24	Naruto Shippuden	Action / Shonen	Manga	500	Fini	Studio Pierrot	2007-02-15	
25	Psycho-Pass	Sci-Fi / Police	Original	41	Fini	Production I.G	2012-10-12	

Aperçu des données brutes (73 animés)

Structure du Dataset

Identification des variables clés : Nom, Genre, Source, Épisodes, Studio et Date de Publication.

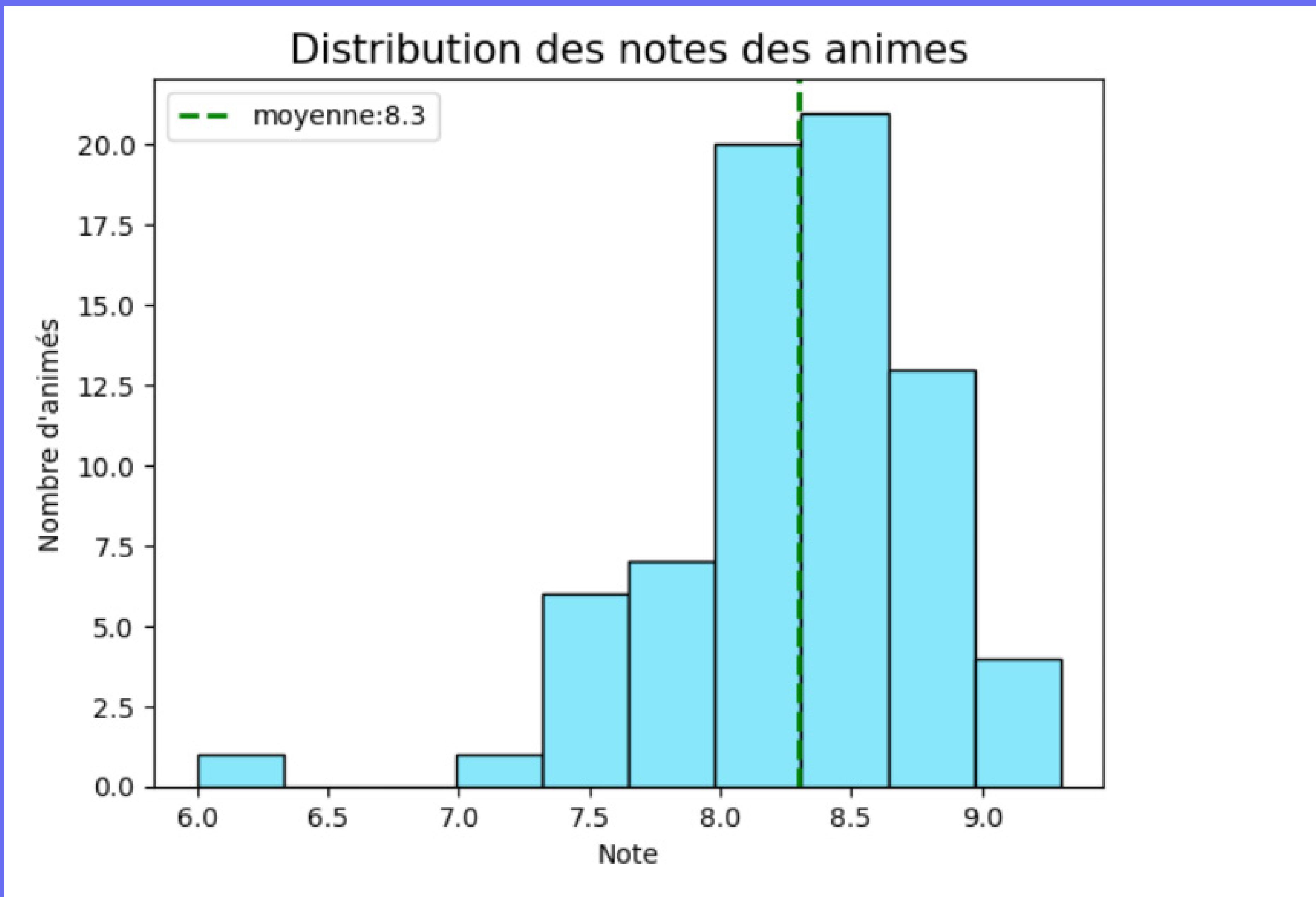
Diversité des Sources

Analyse des origines variées : Mangas, Light Novels, Web Mangas et créations Originales.

Volume de Contenu

Étude de la distribution du nombre d'épisodes pour évaluer la densité du catalogue.

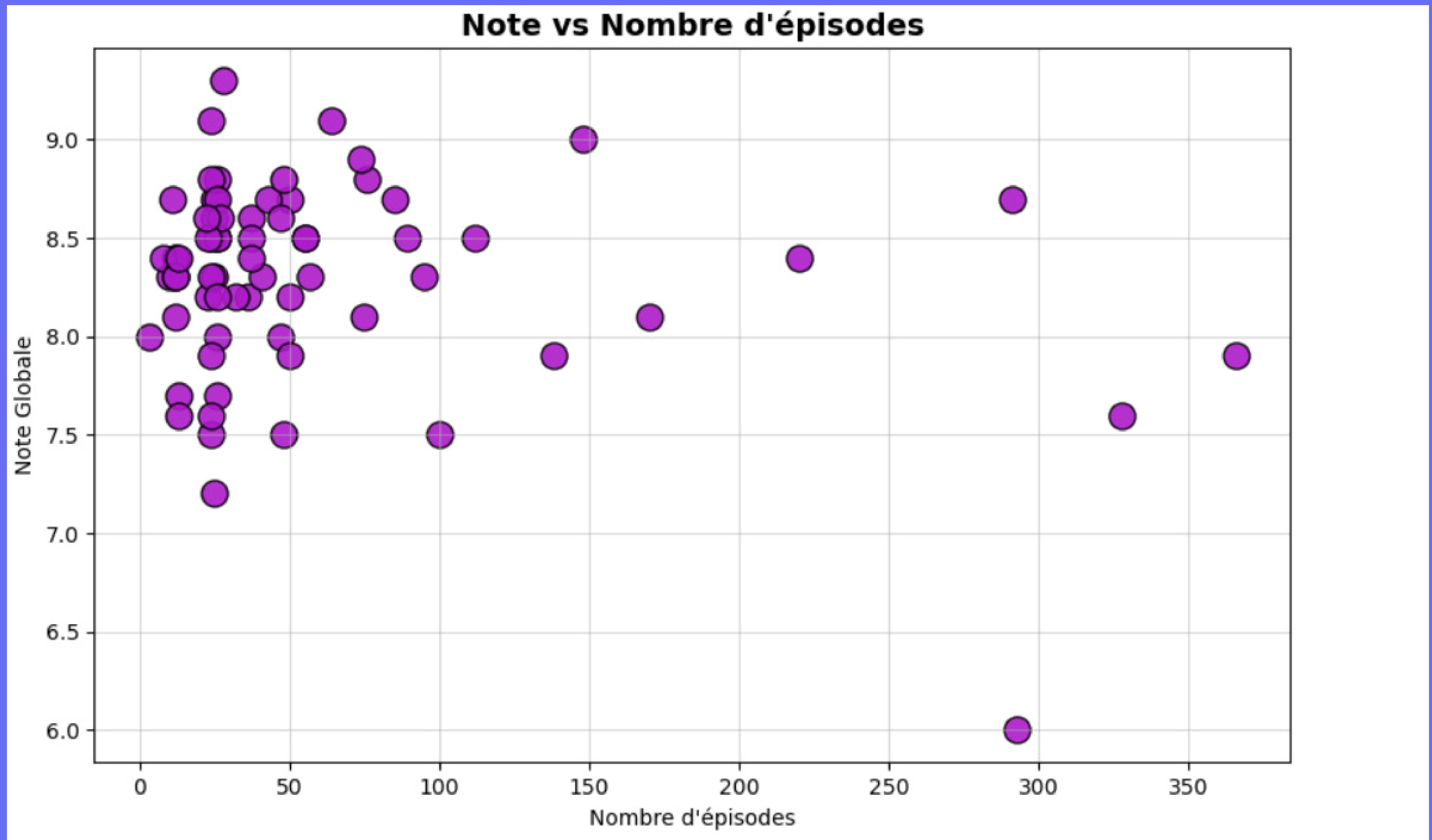
HISTOGRAMME



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 ≤ 400

TOP 10 DE MES ANIMES

J'ai fais un trie parmis 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 IPPÔ		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 !