

Analyse de dataset d'animés - Projet Data

Contexte du projet

Ce projet s'inscrit dans un travail de montée en compétences en **Python et data analysis**, réalisé dans un cadre pédagogique à **HETIC**. L'objectif était d'appliquer concrètement les notions de **nettoyage de données, analyse statistique et visualisation** à partir d'un jeu de données culturel.

Le dataset étudié regroupe **73 animés**, allant des œuvres les plus populaires aux plus atypiques, afin d'en analyser la qualité globale et d'identifier celles présentant une **forte valeur éditoriale**.

Objectifs

- Nettoyer et structurer un dataset brut
 - Créer des indicateurs pertinents d'analyse
 - Produire des visualisations claires et impactantes
 - Identifier les animés de **très haute qualité** et **réguliers**
 - Rendre le dataset exploitable sur **Python, Excel et SQL**
-

○ Données utilisées

Le dataset contient notamment les informations suivantes : - Nom de l'animé - Genre(s) - Studio de production - Nombre d'épisodes - Statut (fini, en cours, etc.) - Notes globales et par épisode

Après nettoyage, seules les colonnes utiles à l'analyse ont été conservées.

Nettoyage des données

Les étapes de nettoyage ont inclus : - Vérification des valeurs manquantes et incohérentes - Normalisation des formats (dates, catégories) - Harmonisation des statuts et genres - Suppression des colonnes non pertinentes

L'objectif était d'obtenir un **dataset fiable, lisible et standardisé**.

Enrichissement du dataset

Des variables calculées ont été ajoutées afin d'améliorer l'analyse : - **Stabilité de qualité** : écart entre la meilleure et la pire note d'épisode - **Score éditorial** : indicateur combinant note globale et régularité

Ces indicateurs permettent une lecture plus fine que la simple note moyenne.

Analyse & visualisation

Plusieurs visualisations ont été réalisées avec **Matplotlib** : - Répartition des animés par statut - Classement des meilleurs animés (barres horizontales) - Donut chart pour les proportions - Lollipop chart pour une lecture comparative élégante

Ces graphiques facilitent la compréhension des résultats et la communication des insights.

Dataset final

Le projet aboutit à un **dataset clean** : - Structuré pour Excel (tableaux, tris, graphiques) - Exploitable en Python et SQL - Prêt pour des analyses futures ou un rendu académique

Les dates sont normalisées au format **ISO 8601 (YYYY-MM-DD)** lorsque nécessaire.

Résultats

L'analyse permet d'identifier : - Les animés les plus réguliers - Ceux à **forte valeur éditoriale** - Les œuvres combinant excellence globale et cohérence narrative

○ Technologies utilisées

- Python
 - Pandas
 - Matplotlib
 - Jupyter Notebook
 - Excel (analyse complémentaire)
 - SQL (structuration et requêtes)
-

Conclusion

Ce projet démontre la capacité à transformer des données culturelles en **analyses exploitables**, tout en mettant en pratique des compétences clés en **data cleaning, data analysis et data visualization**.

Il illustre l'importance d'un dataset bien structuré pour produire des **insights pertinents et argumentés**.

- Projet réalisé dans un cadre pédagogique – HETIC