

Linkedin

Sur une plateforme de streaming, attendre des mois de données d'engagement pour décider quoi mettre en avant... c'est trop lent.

Je me suis donc lancé sur un projet Data Science : créer un "Score de Confiance" pour identifier rapidement des animés à forte valeur éditoriale, même avec des données limitées.

Point de départ :

- une note globale = réception du public
- mais ça ne suffit pas à mesurer la fiabilité de l'expérience (régularité, maîtrise de prod...) Méthodologie (très résumé)

Nettoyage → EDA (distributions/corrélations/outliers) → features de stabilité → score + tests de sensibilité → segmentation éditoriale → top 10 + dashboard.


Le score : un indice de confiance (pas une popularité)

$\text{Score_Editorial} = \text{Note_Globale} + \text{Bonus_Studio} - (\text{Ecart_Brut} \times 0.1)$

- Note Globale : réception du public
- Bonus Studio : "label de prestige" (savoir-faire / moyens)
- Malus régularité : pénalise une expérience trop inégale

Choix important : l'intégrité de la mesure

Les notes étaient très élevées (moyenne ~8.30). Plutôt que "punir" la note, j'ai gardé la valeur intacte et j'ai mis la sévérité dans la segmentation pour pouvoir isoler l'excellence.

 En bref : distinguer un animé "très bon" d'un chef-d'œuvre techniquement irréprochable.

Au-delà du score, j'ai surtout appris à transformer une intuition "éditoriale" en démarche data : nettoyer, explorer, construire un indicateur, puis le rendre actionnable via une segmentation. Prochaine étape : tester ce score sur un autre dataset / d'autres genres, et voir comment il se comporte dans le temps.

#DataScience #Python #EDA #Jupyter #Scoring #FeatureEngineering #Produit
#Streaming #Anime