Template feedback Hackathon#1

Notions évaluées : Extraction, traitement et consolidation de données



Feedback - Hackathon

Nom des élèves : Léa, Karim

Nom du projet : Sentiment Social

Nom du professeur évaluateur : Clara Martinez

Date: 30/06/2025



Feedback global (2 phrases maximum)

Points positifs:

- → Excellente **structuration et lisibilité**: Le projet suit une démarche logique et très claire, allant du nettoyage des données à l'analyse approfondie en plusieurs étapes numérotées. Le code est propre, bien segmenté et facile à suivre.
- → Profondeur de l'analyse exploratoire (EDA) : L'analyse ne se contente pas de produire des graphiques, elle les interprète systématiquement avec une "lecture métier". La capacité à transformer des observations statistiques (ex: corrélation, bimodalité du karma) en insights concrets sur le comportement des utilisateurs Reddit est le point fort de ce projet.
- → **Rigueur statistique :** L'utilisation de tests statistiques (test t, ANOVA, Kruskal-Wallis) pour valider les hypothèses (ex: différence entre "super_users" et "regular_users") démontre une grande maturité analytique et dépasse le cadre d'une simple exploration.
- → Qualité des visualisations et du storytelling : Les graphiques sont non seulement pertinents (boxplot, KDE, heatmap, bar chart horizontal), mais aussi très soignés et systématiquement accompagnés d'une analyse détaillée qui raconte une histoire cohérente sur les dynamiques communautaires.

Points à améliorer :

- → Passer à la modélisation prédictive : L'analyse est exhaustive, mais elle reste descriptive. Une excellente piste de progression serait d'utiliser les features créées (log_karma, sentiment, user_type, etc.) pour construire un modèle de Machine Learning (ex: prédire le **score** d'un commentaire ou le classifier comme "viral" / "non viral").
- → **Modularité du code :** Bien que le code soit clair, certaines sections d'analyse pourraient être encapsulées dans des fonctions pour éviter la répétition et rendre le notebook encore plus concis et réutilisable (par exemple, une fonction qui génère un boxplot et son interprétation pour différentes variables).

Évaluation notée (notes sur 5)

(Ces notes servent à départager les groupes pour la sélection du meilleur projet)

Critère	Note /5	Commentaire
Clarté du code et commentaires pertinents	[5]	Code bien segmenté, clair, lisible et ordonné. Les sections en Markdown servent de commentaires de très haute qualité.
Utilisation efficace des notions vues pendant le bootcamp (Pandas, visualisation, EDA, etc.)	[5]	Utilisation experte de Pandas (groupby, agg), de la visualisation (Seaborn/Matplotlib) et du feature engineering (log transform, sentiment). Le niveau dépasse les attentes.
Résultats corrects et visualisations cohérentes	[5]	Les visualisations sont non seulement correctes mais particulièrement bien choisies pour illustrer les points de l'analyse. L'interprétation est toujours cohérente avec le visuel.
Structuration du notebook / workflow clair et reproductible	[5]	La structure en 5 étapes, de la préparation à l'analyse approfondie, est exemplaire. La fonction build_final_df est une excellente pratique pour la reproductibilité.
Pertinence de l'analyse exploratoire (EDA) et capacité à tirer des insights	[5]	Point fort du projet. Chaque analyse est transformée en un

Critère	Note /5	Commentaire
		"insight métier" actionnable sur la dynamique des communautés Reddit. L'identification des "super_users" est brillante.
Capacité à répondre aux questions posées via des analyses/statistiques simples (NumPy/SciPy)	[5]	Excellente utilisation des tests statistiques (t-test, ANOVA) pour valider les hypothèses, montrant une compréhension qui va audelà de la simple description des données.

XX Commentaire final (optionnel)

Projet exceptionnel qui démontre une maîtrise complète de la chaîne de valeur de la data analyse. La démarche est rigoureuse, la profondeur de l'analyse est impressionnante et, surtout, la capacité à traduire chaque résultat en une interprétation métier pertinente est remarquable. Le groupe a non seulement répondu à la problématique, mais a su raconter une histoire convaincante et riche en insights à partir des données. C'est un travail qui établit une nouvelle référence pour les projets du hackathon. Félicitations