

Template feedback Hackathon#1

Notions évaluées : Extraction, traitement et consolidation de données

Feedback – Hackathon

Nom des élèves : Julianna, Jeremy

Nom du projet : Coût de la santé et facteur de risque

Nom du professeur évaluateur : Clara Martinez

Date : 03/07/25

Feedback global (2 phrases maximum)

Points positifs :

- **Excellente méthodologie et structuration** : Le projet suit une démarche d'analyse de données très claire et professionnelle : nettoyage, analyse exploratoire (EDA), création de nouvelles variables (feature engineering), tests statistiques pour valider les hypothèses, et même une initiation au machine learning.
- **Analyse approfondie et proactive** : Le groupe ne s'est pas contenté d'analyser le dataset fourni. L'initiative d'intégrer des données macro-économiques externes sur les dépenses de santé par État pour contextualiser les résultats est un vrai point fort qui démontre une grande curiosité et une approche analytique mature.
- **Rigueur statistique** : L'utilisation de tests statistiques (t-test, Kruskal-Wallis, χ^2) pour confirmer les observations visuelles est remarquable. Cela ajoute une crédibilité scientifique aux conclusions tirées.
- **Qualité des visualisations** : Les graphiques sont clairs, bien légendés et pertinents. L'utilisation de visualisations interactives (**plotly**, **ipywidgets**) est un excellent bonus qui valorise la présentation.

Points à améliorer :

- **Confiance à l'oral** : La maîtrise technique est indéniable, comme en témoigne la qualité du notebook. Le prochain axe de progression serait de gagner en assurance lors de la prise de parole pour présenter les résultats avec la même conviction que celle démontrée dans l'analyse. La technique est là, il faut maintenant la "vendre" avec confiance.
 - **Justification des choix méthodologiques** : Le notebook utilise des outils statistiques avancés, ce qui est excellent. Pour aller plus loin, il serait bénéfique d'expliquer brièvement *pourquoi* un certain test a été choisi (par exemple, "nous utilisons Kruskal-Wallis car nous comparons plus de deux groupes et ne pouvons pas assumer une distribution normale des charges"). Cela démontrerait une couche de compréhension théorique encore plus profonde.
-

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 conclusions intermédiaires en markdown guident parfaitement le lecteur.
Utilisation efficace des notions vues pendant le bootcamp (Pandas, visualisation, EDA, etc.)	[4]	Le groupe a utilisé un large éventail de bibliothèques (Pandas, Seaborn, Scipy), mais l'explication du choix de certains tests statistiques (ex: Kruskal-Wallis vs ANOVA) pourrait être plus approfondie pour atteindre la maîtrise complète.
Résultats corrects et visualisations cohérentes	[5]	Les visualisations sont claires, bien titrées et toujours pertinentes pour appuyer une conclusion. L'utilisation de graphiques interactifs est un vrai plus.
Structuration du notebook / workflow clair et reproductible	[5]	Le notebook suit une structure logique et facile à suivre, de l'importation des données au modèle prédictif. Le parcours analytique est exemplaire.
Pertinence de l'analyse exploratoire (EDA) et capacité à tirer des insights	[5]	Analyse très complète. L'exploration des interactions (ex: bmi * smoker) et le croisement avec des données externes démontrent une

Critère	Note /5	Commentaire
		excellente capacité à générer des insights qui vont au-delà de la surface.
Capacité à répondre aux questions posées via des analyses/statistiques simples (NumPy/SciPy)	[5]	Utilisation rigoureuse des tests statistiques (t-test, Kruskal-Wallis, Chi ²) pour valider formellement les hypothèses et confirmer les observations visuelles, ce qui est une marque de grande qualité.

🔍 Commentaire final (optionnel)

→ Excellent travail qui dépasse les attentes du hackathon. Le groupe a non seulement démontré une solide maîtrise technique sur l'ensemble du périmètre (de la manipulation de données à la modélisation), mais a aussi fait preuve d'une grande intelligence analytique en enrichissant son étude avec des données externes et en testant rigoureusement ses hypothèses. La conclusion sur l'effet multiplicateur de l'obésité et du tabagisme, surtout dans la région Sud-Est, est particulièrement percutante. Un projet de très haut niveau qui pourrait servir de référence.