

Template feedback Hackathon#1

Notions évaluées : Extraction, traitement et consolidation de données

Feedback – Hackathon

Nom des élèves : Jaled Tazit, Thierry Danois

Nom du projet : Battle of the Cities : Quality of Life Analysis

Nom du professeur évaluateur : Clara Martinez

Date : 02/07/25

Feedback global (2 phrases maximum)

Points positifs :

- **Présentation originale et engageante** avec l'utilisation de personas pour contextualiser le projet.
- **Excellente initiative de collecte de données très diversifiées** via de multiples sources (fichier plat, APIs complexes, scraping), démontrant une grande ambition technique.
- **Mise en œuvre de techniques avancées**, notamment l'utilisation de **BallTree** pour un matching géospatial efficace et la gestion robuste des appels API (erreurs, pagination).
- **Très bonne prise de recul sur les résultats**, en identifiant et commentant les biais potentiels dans les données collectées (Ticketmaster, Transit.land) et en proposant de bonnes perspectives en conclusion.

Points à améliorer :

- **Renforcer la rigueur méthodologique et la validation des insights.** Bien que les biais des données aient été identifiés, leur impact sur les conclusions n'est pas entièrement maîtrisé. Le processus de nettoyage pourrait être plus approfondi (ex: comment gérer les données plafonnées de l'API de transport ?) et les conclusions statistiques mieux validées pour répondre aux questions du type "comment validez-vous ce résultat ?".
 - **Améliorer la structure de la présentation et la lisibilité des supports.** Le flux de la présentation pourrait être plus linéaire (l'enchaînement "impact business → défis techniques → retour au business" était un peu déroutant). Il est crucial de s'assurer que toutes les visualisations sont parfaitement lisibles (ex: l'histogramme coupé) et de respecter le temps imparti (gros dépassement de 7 minutes) pour maximiser l'impact du message.
-

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 très bien structuré en sections logiques (collecte, nettoyage, analyse). L'utilisation de cellules de texte pour expliquer la méthodologie entre les blocs de code est une excellente pratique et rend le notebook très lisible et ordonné.
Utilisation efficace des notions vues pendant le bootcamp (Pandas, visualisation, EDA, etc.)	[4]	Excellente utilisation de Pandas et des bibliothèques de visualisation (Plotly, Seaborn). L'intégration de plusieurs APIs et l'utilisation de techniques avancées comme BallTree dépassent les attentes. La note est légèrement modérée car la rigueur dans le nettoyage et l'interprétation statistique pourrait être renforcée, ainsi que une plus grande diversité d'outils de visualisation.
Résultats corrects et visualisations cohérentes	[4]	Les visualisations géographiques avec Plotly sont particulièrement réussies et percutantes. Cependant, un histogramme était illisible lors de la présentation, et certains résultats sont basés sur des données dont les biais ont été identifiés mais pas entièrement corrigés (ex: num_routes), ce qui affecte légèrement la justesse globale.

Critère	Note /5	Commentaire
Structuration du notebook / workflow clair et reproductible	[5]	Le workflow est exemplaire : collecte de données par source, consolidation, nettoyage, puis analyse exploratoire et statistique. Le notebook est facile à suivre et les étapes sont reproductibles (à l'exception des clés API qui sont, à juste titre, masquées).
Pertinence de l'analyse exploratoire (EDA) et capacité à tirer des insights	[3]	L'analyse exploratoire est très riche, notamment avec les cartes thématiques détaillées. Le groupe a su générer de nombreuses observations. Cependant, les conclusions manquent parfois de validation statistique robuste (comme soulevé lors des questions) et la corrélation de 0.44 a été qualifiée de "forte", ce qui est un peu exagéré (elle est modérée).
Capacité à répondre aux questions posées via des analyses/statistiques simples (NumPy/SciPy)	[3]	Le groupe a utilisé des tests statistiques (t-tests) pour comparer des groupes, ce qui est une bonne initiative. Toutefois, l'interprétation des résultats (notamment la p-value limite à 0.0501) et la capacité à défendre la validité des conclusions face aux questions pourraient être améliorées.

Excellent travail sur un projet très ambitieux. Le groupe a démontré une maîtrise technique impressionnante dans la collecte et l'agrégation de données complexes provenant de sources hétérogènes. La clarté du code et la structure du notebook sont exemplaires. Les principaux axes de progression concernent la validation rigoureuse des insights tirés de données parfois biaisées et la structuration lors de la présentation pour rendre les conclusions encore plus percutantes. Félicitations pour cet excellent travail.