

## Template feedback Hackathon#1

Notions évaluées : Extraction, traitement et consolidation de données

---

### Feedback – Hackathon

**Nom des élèves** : Agathe, Walid

**Nom du projet** : Comprendre la gastronomie mondiale

**Nom du professeur évaluateur** : Clara Martinez

**Date** : 03/07/25

---

### Feedback global (2 phrases maximum)

#### Points positifs :

→ **Grande curiosité et ambition analytique** : Le groupe ne s'est pas contenté d'une seule approche. Ils ont mené deux analyses distinctes et complexes : une première sur un modèle prédictif et une seconde, très large, sur de l'analyse exploratoire (EDA). Cela démontre une forte motivation et une volonté d'aller au fond des choses.

→ **Large palette de techniques mobilisées** : Le groupe a utilisé un éventail impressionnant d'outils : manipulation avancée (fusion, transformation de types), une grande variété de visualisations, et même des techniques plus spécifiques comme l'analyse de fréquence de mots sur du texte (**collections.Counter**) ou la création de features simples (**estimate\_calories**). C'est la marque d'un travail de recherche approfondi.

#### Points à améliorer :

→ **Rigueur méthodologique et compréhension des données** : C'est le principal axe de progression. Le groupe applique des fonctions techniques sans toujours en valider le sens. Par exemple, convertir **toutes** les colonnes texte (comme **Name** ou **Text Description**) en codes numériques, puis calculer des moyennes ou des corrélations sur ces codes, est une erreur méthodologique. Une moyenne de "noms de plats" n'a pas de sens statistique. Il faut bien distinguer l'encodage pour un modèle de Machine Learning de l'analyse statistique directe.

→ **Analyse critique et interprétation des résultats** : Ce point est confirmé par la deuxième analyse.

\* **Modèle prédictif** : Un score  $R^2$  de 1.0 et des prédictions identiques pour tous les pays auraient dû être des signaux d'alerte de surapprentissage ou d'un problème dans les données d'entrée, nécessitant une investigation.

\* **Analyse de texte** : L'analyse de fréquence des mots est une bonne idée, mais elle inclut des "stop words" (comme 'and', 'the', 'is'), qui devraient être retirés pour ne garder que les mots porteurs de sens (ingrédients, etc.).

\* **Erreur de code** : L'analyse par région a échoué à cause d'une erreur de logique

(`df[region] == region`), ce qui montre un besoin de tester et déboguer son code plus attentivement.

---

### 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
Utilisation efficace des notions vues pendant le bootcamp (Pandas, visualisation, EDA, etc.)	[4]	Le groupe utilise une très grande variété de notions, ce qui est excellent. La note n'est pas à 5 car l'application n'est pas toujours "efficace" ou correcte (ex: analyse statistique sur des codes catégoriels), mais la connaissance technique est là.
Résultats corrects et visualisations cohérentes	[ 3]	Les visualisations sont techniquement bien générées, mais elles représentent souvent des données dont l'analyse est incorrecte (ex: corrélation de noms de plats). Les résultats du modèle prédictif sont également erronés (overfitting).
Structuration du notebook / workflow clair et reproductible	[5]	Le workflow reste très facile à suivre, même à travers les différentes analyses. C'est très bien structuré.
Pertinence de l'analyse exploratoire (EDA) et capacité à tirer des insights	[4]	L'intention de l'EDA est très pertinente (chercher les plats populaires, les

Critère	Note /5	Commentaire
		ingrédients clés). Ils tirent des insights corrects de leurs premières visualisations. La note est légèrement réduite car une partie de l'EDA plus avancée est basée sur une méthodologie fragile.
Capacité à répondre aux questions posées via des analyses/statistiques simples (NumPy/SciPy)	[3]	Cette note est confirmée. Le groupe utilise des statistiques, mais peine à interpréter les résultats de manière critique (t-stat infinie, corrélation sur des données nominales). C'est le lien entre l'outil statistique et sa signification concrète qui doit être renforcé.

### ❏ Commentaire final (optionnel)

→ Travail très impressionnant par son ambition et la diversité des techniques explorées. Le groupe fait preuve d'une grande autonomie et d'une curiosité qui les a poussés à aller bien au-delà des attentes minimales. Leurs compétences en matière de codage et de structuration sont excellentes.

Le principal axe de développement est la consolidation de la **rigueur méthodologique**. Il est crucial de s'assurer que l'analyse statistique appliquée est adaptée à la nature des données (on ne traite pas des identifiants comme des valeurs numériques continues). De même, en modélisation, des résultats "trop parfaits" doivent déclencher une analyse critique.

Avec une attention accrue à la signification profonde des méthodes utilisées, ce groupe a le potentiel pour produire des analyses de très haut niveau. Félicitations pour cet effort remarquable.