

L'IA au cœur du quotidien calédonien

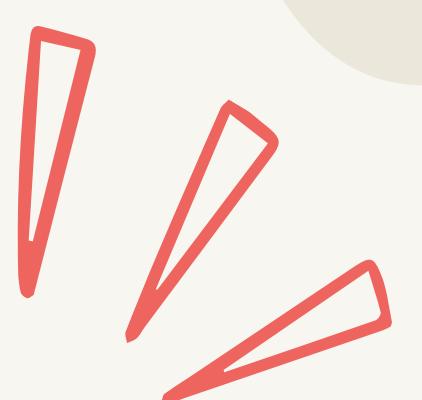
Optimiser les prix et les paniers alimentaires avec PRIX.NC

Présenté par Dévi BERGES



Le contexte : Un enjeu calédonien majeur

Face à la hausse du coût de la vie en Nouvelle-Calédonie, la question de l'accès à des produits alimentaires de qualité, à prix raisonnables, et disponibles localement devient un enjeu central pour les consommateurs.





Objectifs



Deux axes de travail majeurs :

- Détection automatique des **anomalies de prix** (avec Machine Learning)
- Construction de “**paniers idéals**” (économique, équilibré, local/import)

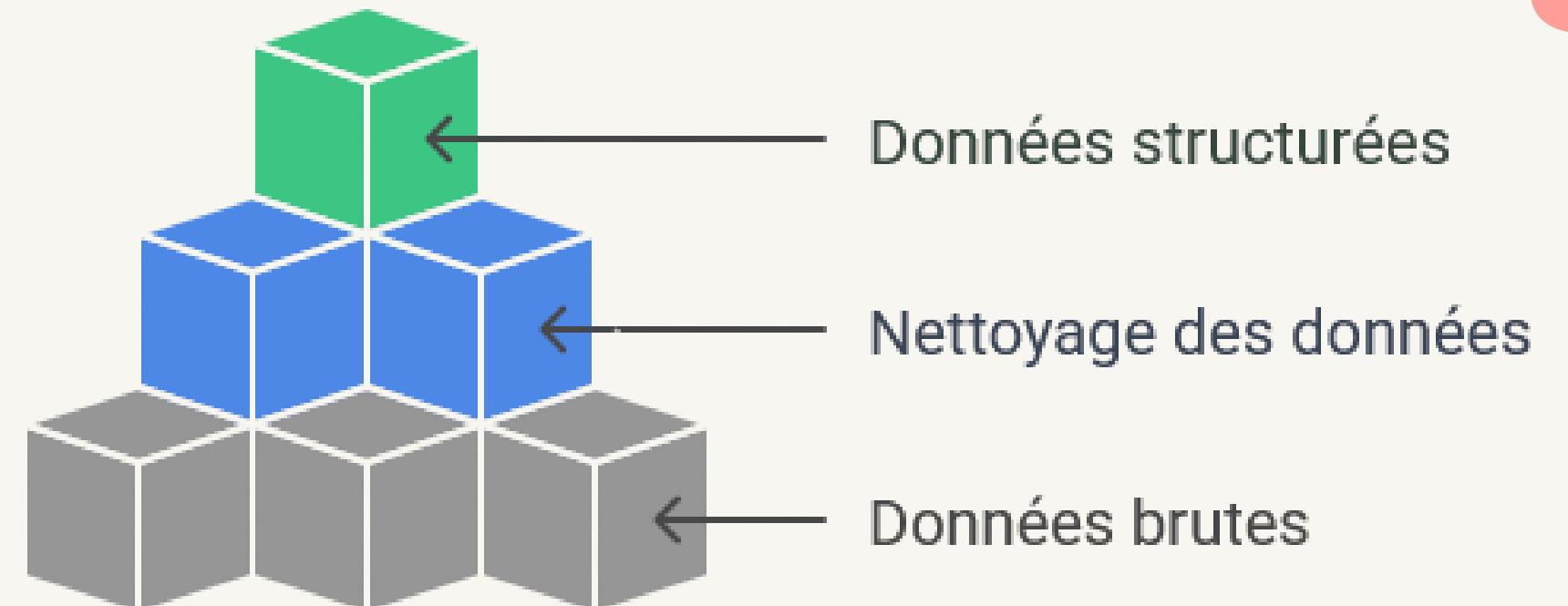


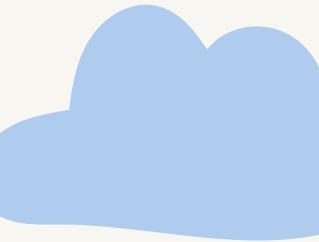
La préparation des données : Le pilier fondamental

Le travail en amont : **nettoyage, structuration, enrichissement**

Souvent **80% du temps** en Data Science

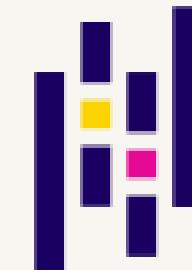
Collaboration essentielle avec l'équipe
Data Prep





Notre arsenal technique

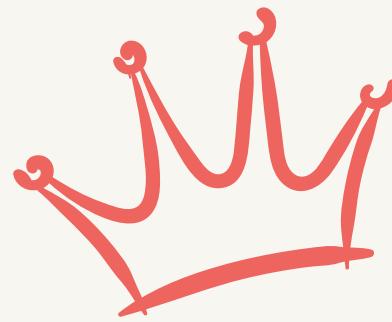
- **DuckDB** : base de données analytique rapide pour gros volumes
- **Pandas** : manipulation et analyse des tables de données en Python
- **Scikit-learn & XGBoost** : Machine Learning (prédiction, classification)
- **Folium** : visualisation cartographique interactive
- Autres : Matplotlib, Seaborn, Numpy...



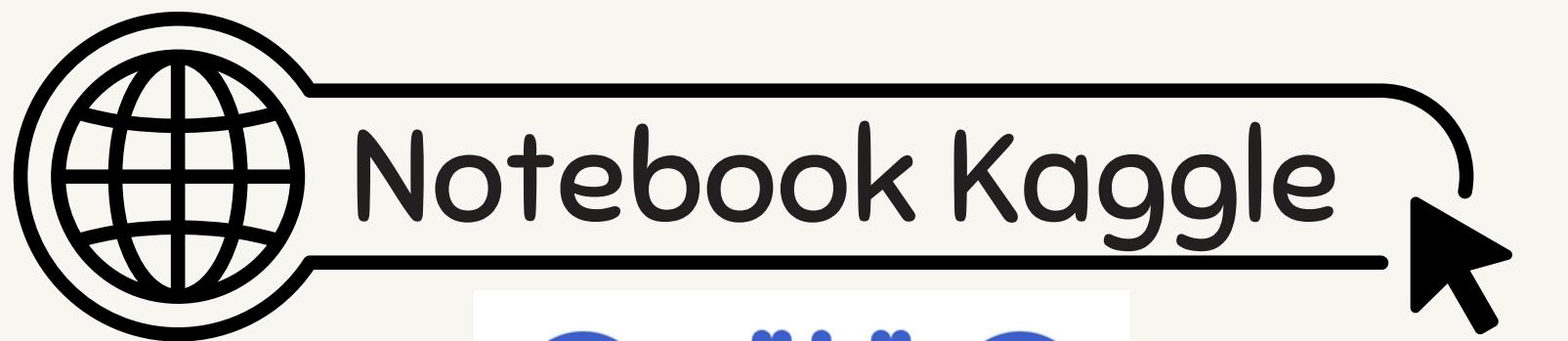
pandas



Folium



Cas d'usage 1: Détection automatique d'anomalies de prix



Le Challenge : Repérer les prix "bizarres"



Comment détecter automatiquement les erreurs
ou promotions exceptionnelles sur Prix.nc ?



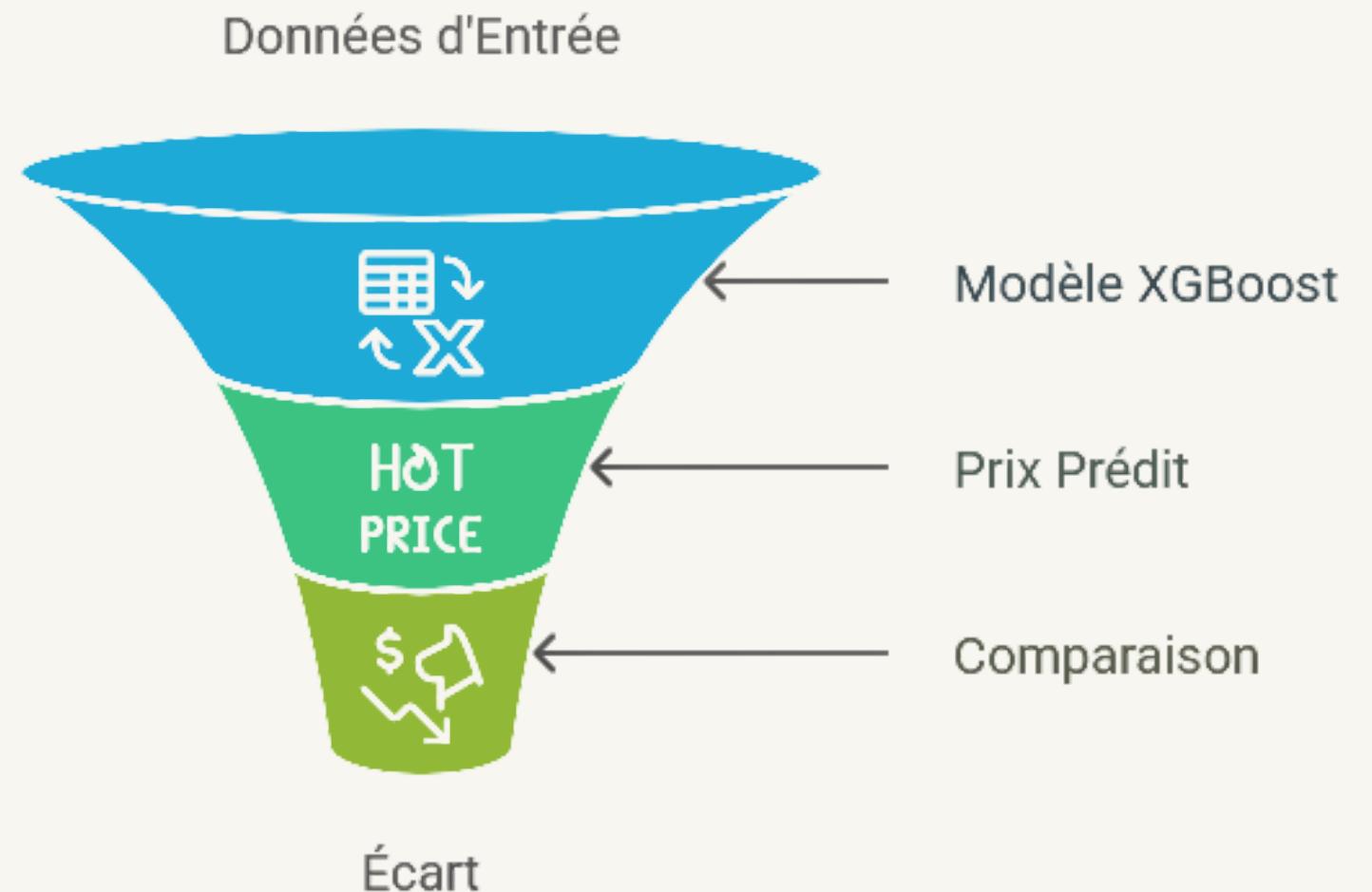
XGBoost Regressor : Notre modèle prédictif



Modèle supervisé : apprentissage sur les données historiques

Force de XGBoost : précision, robustesse (données hétérogènes), rapidité

Principe : prédire un prix “normal” puis identifier les écarts anormaux



Exemples frappants d'anomalies détectées



Super Promo : conserve de cassoulet (4,2kg)
à 500 FCFP (au lieu de ~4000 FCFP) !

Erreur de saisie : pot de yaourt à **84 millions**
FCFP/kg !





Lorsque vous saisissez des mots séparés par des espaces, vous obtenez les produits qui renferment tous les mots demandés. Plus vous saisissez de mots et plus votre résultat sera précis.

163 RÉSULTAT(S) POUR VOTRE RECHERCHE

cassoulet Mutular au porc conserve 4.2 kg

Meilleur prix constaté

500 F CFP

Prix par Kilogramme

119 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER

Produit Local

Points de vente

Prix relevé

Prix par unité

Date du relevé

DISCOUNT PLUS NERA

500 F CFP

119 F CFP / kg

27/05/2025

SUPER U KOUMAC

4,120 F CFP

980 F CFP / kg

02/06/2025

CARREFOUR ECO - KIM CHI

4,320 F CFP

1,028 F CFP / kg

19/05/2025

Les résultats ci-dessous affichent les produits les moins proches de votre recherche.

cassoulet Mutular au porc conserve 1,25 kg

Meilleur prix constaté

1,070 F CFP

Prix par Kilogramme

856 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER

cassoulet Mutular au porc conserve 2 kg

Meilleur prix constaté

1,755 F CFP

Prix par Kilogramme

877 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER

cassoulet Mutular au porc conserve 420 g

Meilleur prix constaté

350 F CFP

Prix par Kilogramme

833 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER

cassoulet Mutular au porc conserve 840 g

Meilleur prix constaté

725 F CFP

Prix par Kilogramme

863 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER

cassoulet Top Budget au porc conserve 840 g

Meilleur prix constaté

495 F CFP

Prix par Kilogramme

589 F CFP / kg

VOIR LES PRIX

AJOUTER AU PANIER



Exemples frappants d'anomalies détectées



Super Promo : conserve de cassoulet (4,2kg)
à 500 FCFP (au lieu de ~4000 FCFP) !

Erreur de saisie : pot de yaourt à **84 millions**
FCFP/kg !



Rechercher et comparer | +

prix.nc/recherche-produits;produit=yaourt%20nature%20tenn;commune=&origine=&type=

All Bookmarks

Rechercher et comparer Localiser un magasin Mon panier Login

PRIX • nc

À PROPOS PANIERS ACTUALITÉS BOUCLIERS QUALITÉ-PRIX PRIX MINI RÉGLEMENTATIONS FAQ RECHERCHE DE PRODUITS

Accueil → Rechercher et comparer

TÉLÉCHARGER L'APP

Sur Play Store Télécharger

Sur Apple Store Télécharger

RECHERCHER UN PRODUIT (Par mot-clés *)

yaourt nature Tennessee Farm pot 4 x 110 g

PAR LOCALISATION PAR ORIGINE PAR TYPE

Communes ▾ Toutes Tous les types de produits ▾ Trouver !

Les prix affichés sont à titre indicatif, sur la base d'une transmission à la DAE par les professionnels / enseignes / commerces. Ils correspondent à un relevé pour une date donnée, qui n'exclut pas leur évolution ultérieure.

Lorsque vous saisissez des mots séparés par des espaces, vous obtenez les produits qui renferment tous les mots demandés. Plus vous saisissez de mots et plus votre résultat sera précis.

172 RÉSULTAT(S) POUR VOTRE RECHERCHE

yaourt nature Tennessee Farm pot 4 x 110 g Meilleur prix constaté 270 F CFP Prix par Kilogramme 613 F CFP / kg VOIR LES PRIX AJOUTER AU PANIER

Produit Local	Points de vente	Prix relevé	Prix par unité	Date du relevé
AUCHAN SAVANNAH SUPERMARCHÉ	270 F CFP	613 F CFP / kg	02/06/2025	
GEANT DUMBEA MALL	270 F CFP	613 F CFP / kg	02/06/2025	
GEANT SAINTE-MARIE	270 F CFP	613 F CFP / kg	02/06/2025	
LEADER PRICE BOURAIL	270 F CFP	613 F CFP / kg	02/06/2025	
LEADER PRICE CONCEPTION	270 F CFP	613 F CFP / kg	02/06/2025	
LEADER PRICE KONE	270 F CFP	613 F CFP / kg	02/06/2025	

LOCKED

Exemples frappants d'anomalies détectées



Super Promo : conserve de cassoulet (4,2kg)
à 500 FCFP (au lieu de ~4000 FCFP) !

Erreur de saisie : pot de yaourt à **84 millions**
FCFP/kg !





DÉVI +2 · 10 AGO (SCHEDULED) · 89 VIEWS



2



:

PRIX.NC_DATAUSE_Détection_anomalies_XGBoost



Copied from james Langeron (+714,-204)

Notebook Input Output Logs Comments (1) Settings

Version 31 of 31



Runtime

53s

Input

DATASETS

dataset-prixnc-nettoye

Tags

Scheduled



Language

Python

Table of Contents

Détection automatique d'anomalies de prix avec XGBoost Regressor

Contexte

🔍 Détection automatique d'anomalies de prix avec XGBoost Regressor

📌 Contexte

La transparence des prix est un enjeu clé pour les consommateurs, les distributeurs et les décideurs publics. En Nouvelle-Calédonie, la plateforme [prix.nc](#) met à disposition une base de données publique sur les prix des produits de consommation courante.

Ces données, bien que précieuses, peuvent parfois contenir des anomalies de prix : erreurs de saisie, promotions très agressives, ruptures de stock, ou hausses brutales non expliquées.

👉 Notre objectif dans ce notebook est de détecter automatiquement ces anomalies à l'aide d'un modèle de machine learning supervisé.

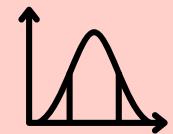
Cela permettra notamment de :

- Mieux comprendre le comportement des prix sur le territoire,
- Repérer des erreurs potentielles dans les relevés,
- Mettre en évidence des situations exceptionnelles (ex : promotions extrêmes),

Accès à cette information, à vos analyses et à votre tableau de bord

Critères pour une vraie anomalie

Un double critère pour qualifier une vraie anomalie :



- Erreur de prédiction **statistiquement rare** (95e percentile des erreurs)
- Prix réel **décalé de >50%** par rapport à la moyenne locale du produit



Concentration des anomalies : où agir ?



Secteurs les plus impactés : épicerie sucrée,
épicerie salée, boissons...

Produits récurrents : chocolats, cafés solubles,
huiles...

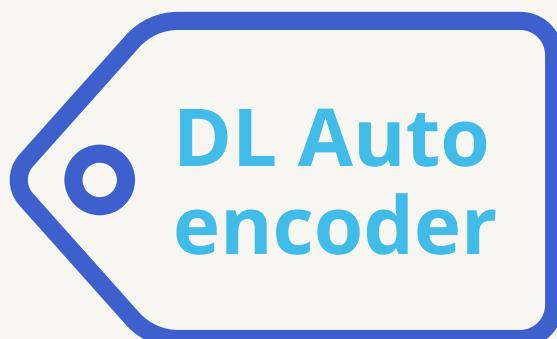
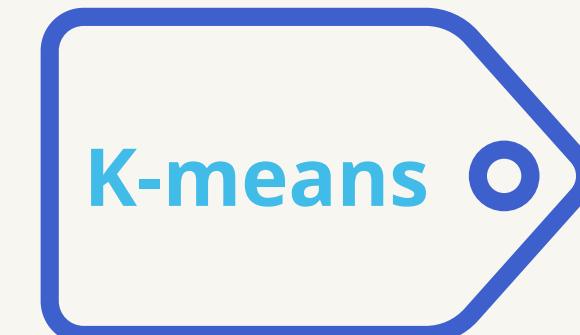
Localisation : certaines communes et enseignes
concentrent plus d'anomalies



La validation croisée : Renforcer la confiance

Autres approches : utilisation de Random Forest, de K-means et de DL Autoencoder pour une analyse similaire

Résultat : Convergence des détections, renforçant la fiabilité des anomalies





L'impact concret de la détection d'anomalies

Amélioration de la qualité de la base Prix.nc

Extraction de signaux économiques (erreurs vs. stratégies de prix)

Démontre l'intérêt de l'analyse de données, même à l'échelle d'un projet étudiant

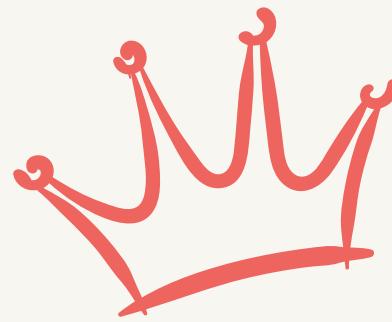
Au-delà du ponctuel : Vers l'analyse des tendances

Et si... ? Analyser les tendances des anomalies
(fréquence, type, localisation)

Révélérer des dynamiques économiques
profondes en Nouvelle-Calédonie

Aider à anticiper les évolutions de prix





Cas d'usage 2: Construction de Paniers idéals



Le quotidien calédonien : Choisir son panier



Un dilemme : coût, santé, consommation locale... comment concilier ?

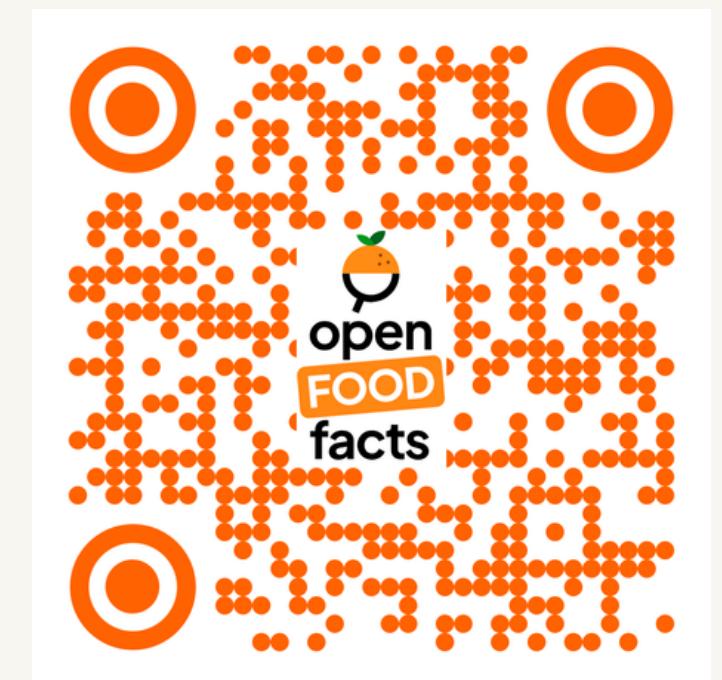




Notre réponse : Les paniers idéals

Construire des paniers optimisés grâce aux données ouvertes de Prix.nc et OpenFoodFacts :

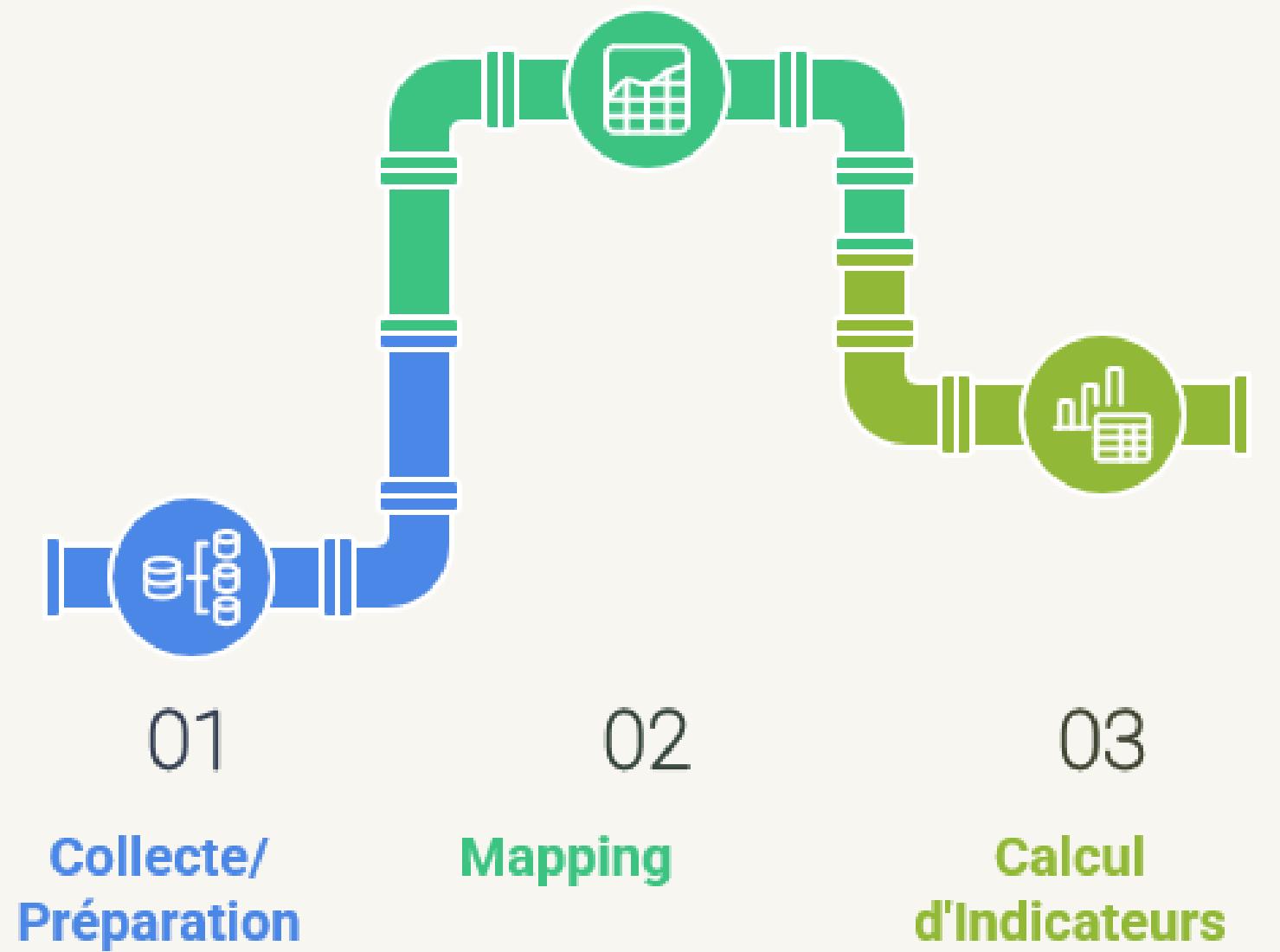
- **Panier Économique** : le moins cher
- **Panier Équilibré** : avec le meilleur Nutri-Score
- **Paniers par origine** : local vs. importé



La méthode : Des datasets aux insights

Un processus structuré :

1. **Préparation & fusion des données** : Prix.nc
+ OpenFoodFacts (DuckDB, Pandas)
2. **Mapping des produits essentiels** : liste ciblée pour chaque catégorie
3. **Génération & évaluation des paniers** : magasin par magasin, selon nos critères



- Manipulation de données : `pandas`, `numpy`, `duckdb`, `json`, `os`
- Visualisation & cartographie : `matplotlib`, `seaborn`, `folium`, `ipywidgets`, `branca`
- Machine Learning & prétraitement : `scikit-learn` (modèles, encodage, pipelines, imputations, évaluations)
- Widgets et export interactif : `IPython.display`, `widgets` (pour les cartes, les listes déroulantes, l'export CSV)
- Gestion des alertes et nettoyage : `warnings`

Ces bibliothèques seront mobilisées tout au long du notebook, notamment pour :

- entraîner un modèle de prédiction du Nutri-Score
- construire des visualisations interactives sur les paniers idéaux
- générer des exports exploitables pour une diffusion ou réutilisation des données.

>Show hidden code

Etape 1 : Chargement des données

Nous commençons par établir une connexion en mémoire avec DuckDB pour attacher les trois bases de données suivantes :

- `prixnc` : base principale contenant les prix, les produits, les magasins, etc.
- `openfood` : enrichissements nutritionnels issus d'OpenFoodFacts
- `entreprises` : rattachement des magasins aux établissements RIDET

Nous chargeons également un export JSON plus riche provenant d'OpenFoodFacts, afin d'exploiter des variables comme les ingrédients, labels, pays d'origine ou Nutri-Score.

Les données sont converties en DataFrame `pandas` pour être fusionnées ensuite avec les relevés de prix.

Table of Contents

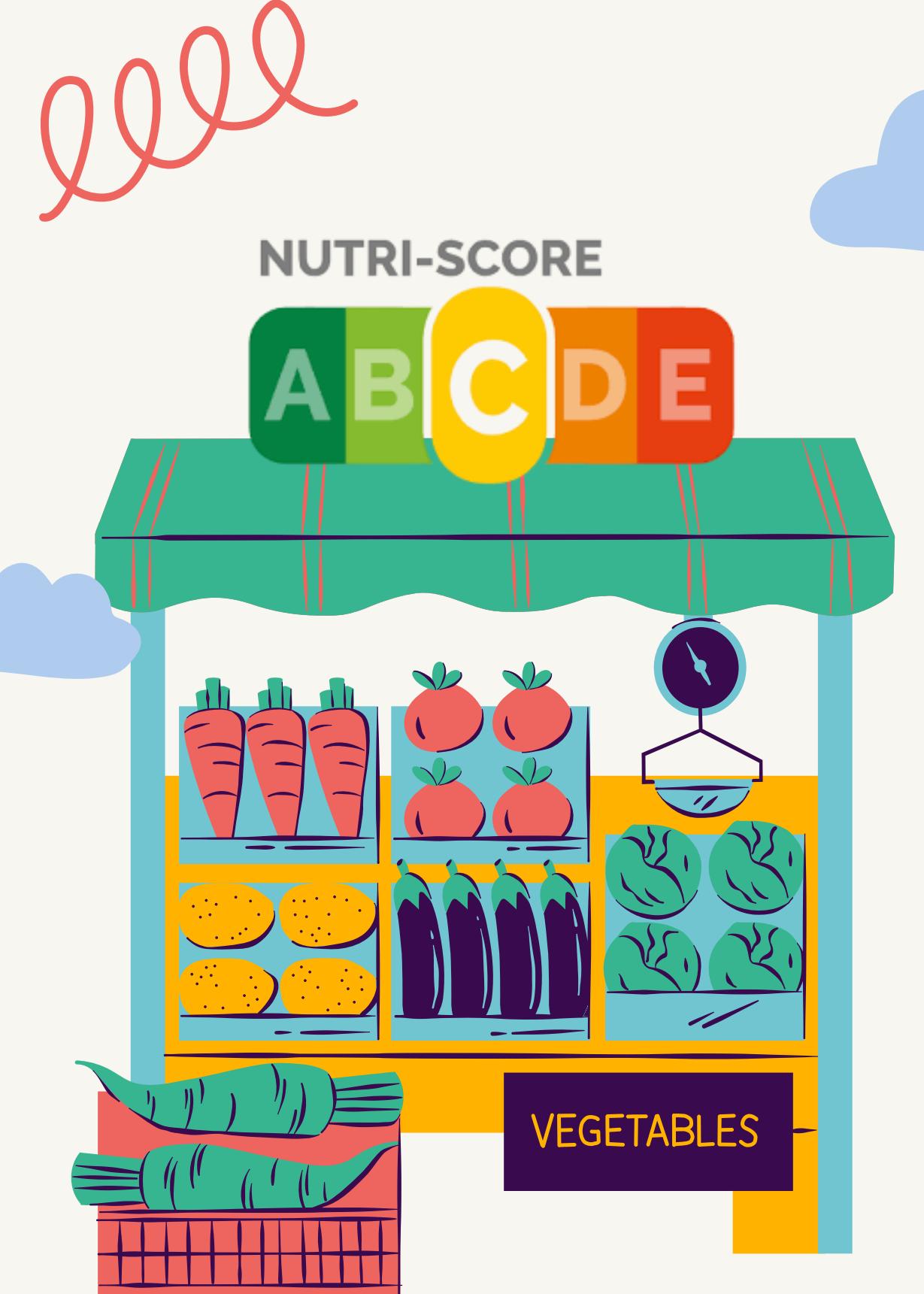
- Paniers idéals en Nouvelle-Calé...
- Contexte
- Objectif
- Méthodes utilisées
- Données utilisées
- Liens vers les datasets utilisés
- Chargement et préparation des ...
- Etapes du notebook
- Etape 0 : Import des bibliothèques
- Etape 1 : Chargement des donné...
- Etape 2 : Nettoyage et prétraite...
- Etape 3 : Prédiction des Nutri-S...
- Analyse des performances du m...
- Complétion des Nutri-Scores m...
- Etape 4 : Définition des critères
- Mapping produit → variétés
- Composition du panier

Prédiction du Nutri-Score manquant

Problématique : nombreux produits sans Nutri-Score

Solution : modèle **Random Forest Classifier** pour prédire les valeurs manquantes

Performance : précision d'environ **99%**, permettant des comparaisons fiables





PRIX.NC_DATAUSE_Panier_ideal

Copied from Dévi (+1397,-669)

Notebook Input Output Logs Comments (1) Settings

⌚ Version 59 of 59



⌚ Paniers idéals en Nouvelle-Calédonie avec Prix.nc

✖ Contexte

Face à la hausse du coût de la vie en Nouvelle-Calédonie, la question de l'accès à des produits alimentaires de qualité, à prix raisonnables, et disponibles localement devient un enjeu central pour les consommateurs.

La plateforme [prix.nc](#) fournit une base de données publique très riche sur les relevés de prix des produits dans les différents magasins du territoire. Couplée à des données issues d'[OpenFoodFacts](#) (valeurs nutritionnelles) et des fichiers de référence des établissements, elle ouvre la voie à des analyses poussées de la consommation et de l'accessibilité alimentaire.

Runtime

▶ 49s

Input

DATASETS

dataset-duckdb-fusion-prixnc-openf

dataset-prixnc-nettoye

prix-nc-dataset-entreprises

Tags

Scheduled



🎯 Objectif

Dans ce notebook, nous explorons la construction de différents types de paniers idéals, en nous appuyant sur plusieurs critères :

Language

Python



PRIX.NC_DATAUSE_Panier_ideal



Copied from Dévi (+1397,-669)

Notebook Input Output Logs **Comments (1)** Settings

⌚ Version 59 of 59



🕒 Paniers idéals en Nouvelle-Calédonie avec Prix.nc

📌 Contexte

Face à la hausse du coût de la vie en Nouvelle-Calédonie, la question de l'accès à des produits alimentaires de qualité, à prix raisonnables, et disponibles localement devient un enjeu central pour les consommateurs.

La plateforme [prix.nc](#) fournit une base de données publique très riche sur les relevés de prix des produits dans les différents magasins du territoire. Couplée à des données issues d'[OpenFoodFacts](#) (valeurs nutritionnelles) et des fichiers de référence des établissements, elle ouvre la voie à des analyses poussées de la consommation et de l'accessibilité alimentaire.

🎯 Objectif

Dans ce notebook, nous explorons la construction de différents types de paniers idéaux, en nous appuyant sur plusieurs critères :

Runtime

▶ 49s

Input

DATASETS

dataset-duckdb-fusion-prixnc-openf

dataset-prixnc-nettoye

prix-nc-dataset-entreprises

Tags

Scheduled



Language

Python

Local vs Importé : Un équilibre délicat

Une dynamique nuancée :

- **Panier 100% local** : souvent plus difficile à composer (complétude plus faible)
- **Prix** : pas de règle unique. Par exemple les protéines locales (viande, poisson) sont moins chères que l'importé, sauf lorsqu'elles sont surgelées

Permet de voir où le local est compétitif et où il l'est moins



PRIX.NC_DATAUSE_Panier_ideal

Copied from Dévi (+1397, -669)

Notebook Input Output Logs Comments (1) Settings

⌚ Version 59 of 59



⌚ Paniers idéals en Nouvelle-Calédonie avec Prix.nc

📌 Contexte

Face à la hausse du coût de la vie en Nouvelle-Calédonie, la question de l'accès à des produits alimentaires de qualité, à prix raisonnables, et disponibles localement devient un enjeu central pour les consommateurs.

La plateforme [prix.nc](#) fournit une base de données publique très riche sur les relevés de prix des produits dans les différents magasins du territoire. Couplée à des données issues d'[OpenFoodFacts](#) (valeurs nutritionnelles) et des fichiers de référence des établissements, elle ouvre la voie à des analyses poussées de la consommation et de l'accessibilité alimentaire.

🎯 Objectif

Dans ce notebook, nous explorons la construction de différents types de paniers idéals, en nous appuyant sur plusieurs critères :

Runtime

▶ 49s

Input

DATASETS

dataset-duckdb-fusion-prixnc-openf

dataset-prixnc-nettoye

prix-nc-dataset-entreprises

Tags

Scheduled



Language

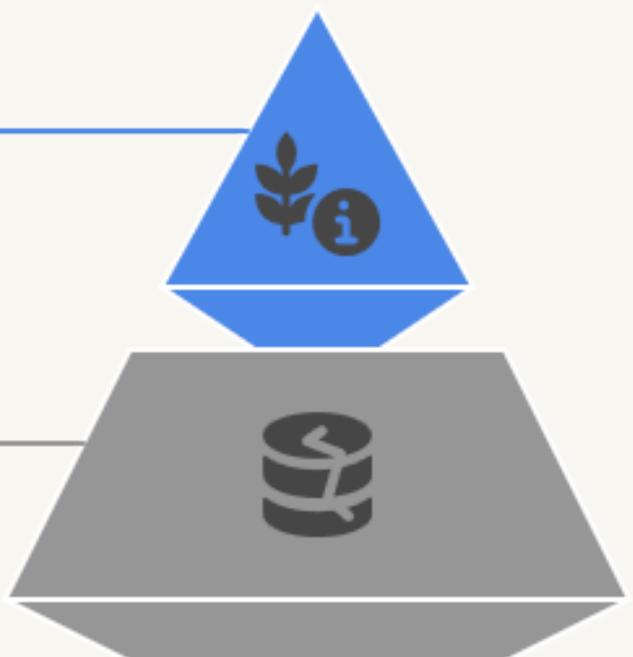
Python



L'impact concret des paniers idéals

Informations concrètes

- coût, qualité, origine



Données brutes

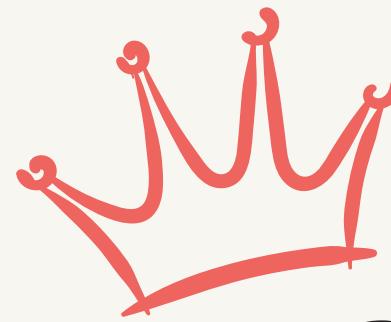
-

Des chiffres bruts à l'éclairage concret

Comprendre le coût réel, la qualité nutritionnelle et l'origine des aliments

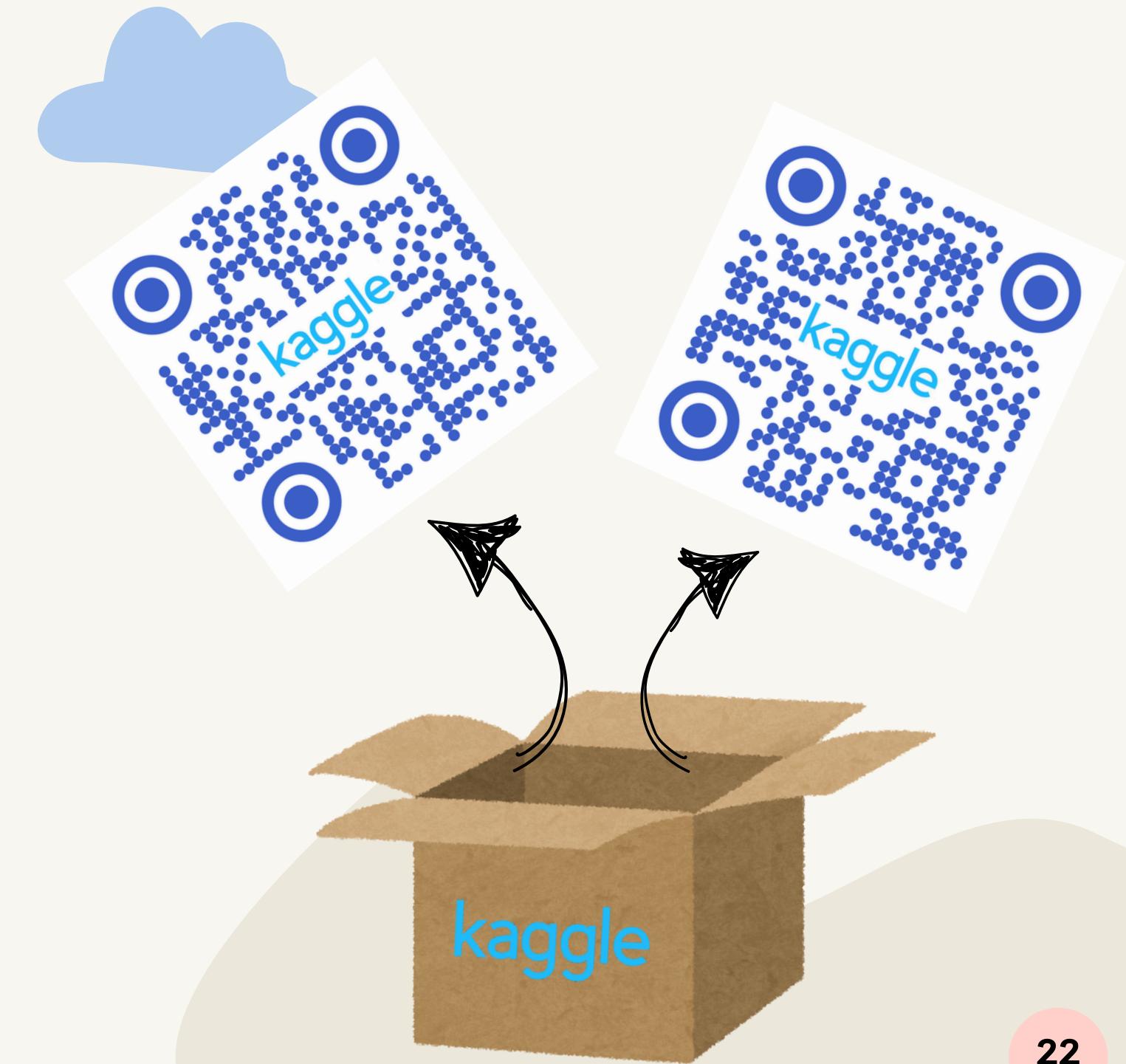
Outil pour des stratégies d'achat objectives

Base solide pour la discussion et les politiques publiques

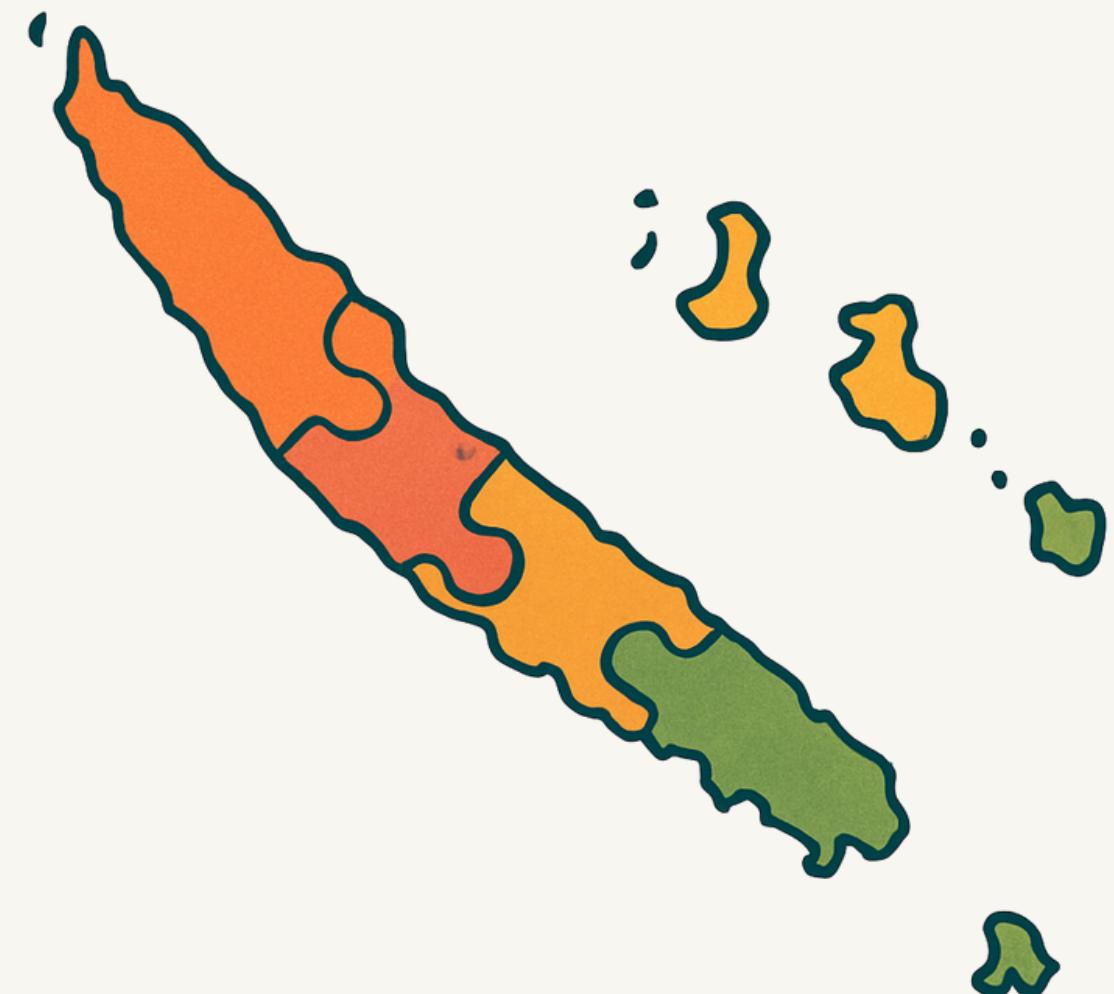
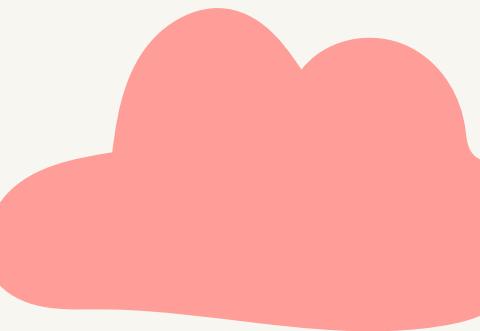


Partage et perspectives futures

- **Résultats disponibles** : exports CSV/Parquet pour réutilisation
- **Pistes d'exploration** : intégrer l'accessibilité géographique, la saisonnalité des produits locaux



Un projet étudiant à l'impact local



- Exploitation des données ouvertes de **Prix.nc** pour comprendre le coût de la vie et les habitudes de consommation
- Un projet qui part **du local, pour le local**

Les deux piliers : Compétences tech & collaboration



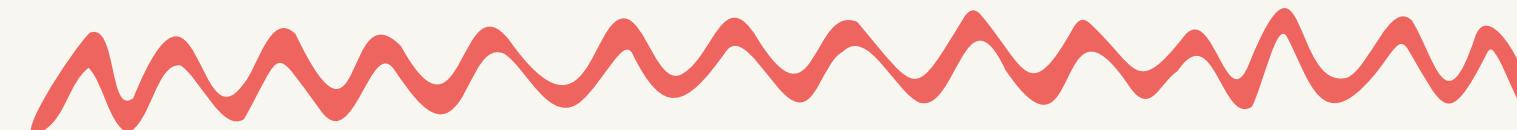
- **Application concrète des compétences techniques** (ML, Data Viz) sur des problèmes réels calédoniens
- **La donnée “parle”** : révélations sur des anomalies (yaourt à 84M, cassoulet à 500F)
- **Collaboration inter-équipes** : clé pour naviguer la complexité du projet et se préparer au monde professionnel



Leçons apprises & perspectives pour la vraie vie



- **Projet très formateur** : confrontation à la complexité du réel (données imparfaites, délais, travail d'équipe)
- **Motivation renforcée** pour des projets à impact
- **Limites** : pas de recommandations personnalisées ni critères sociaux fins pour l'instant



La synthèse : Deux points essentiels

La puissance de l'analyse de données à l'échelle locale : un éclairage concret sur la consommation en Nouvelle-Calédonie

L'importance cruciale des compétences transversales (soft skills) : communication, collaboration, adaptation



Perspectives d'avenir : Un potentiel illimité

Comment l'analyse continue des données ouvertes peut outiller une communauté entière ?

Mieux comprendre et influencer les dynamiques économiques du quotidien



Merci pour votre attention !



Dévi BERGES

Étudiant M2 MIAGE & porteur du projet
de la team DATA USE



Adrien SALES

Enseignant vacataire référent
d'interopérabilité des SI & encadrant du
projet

