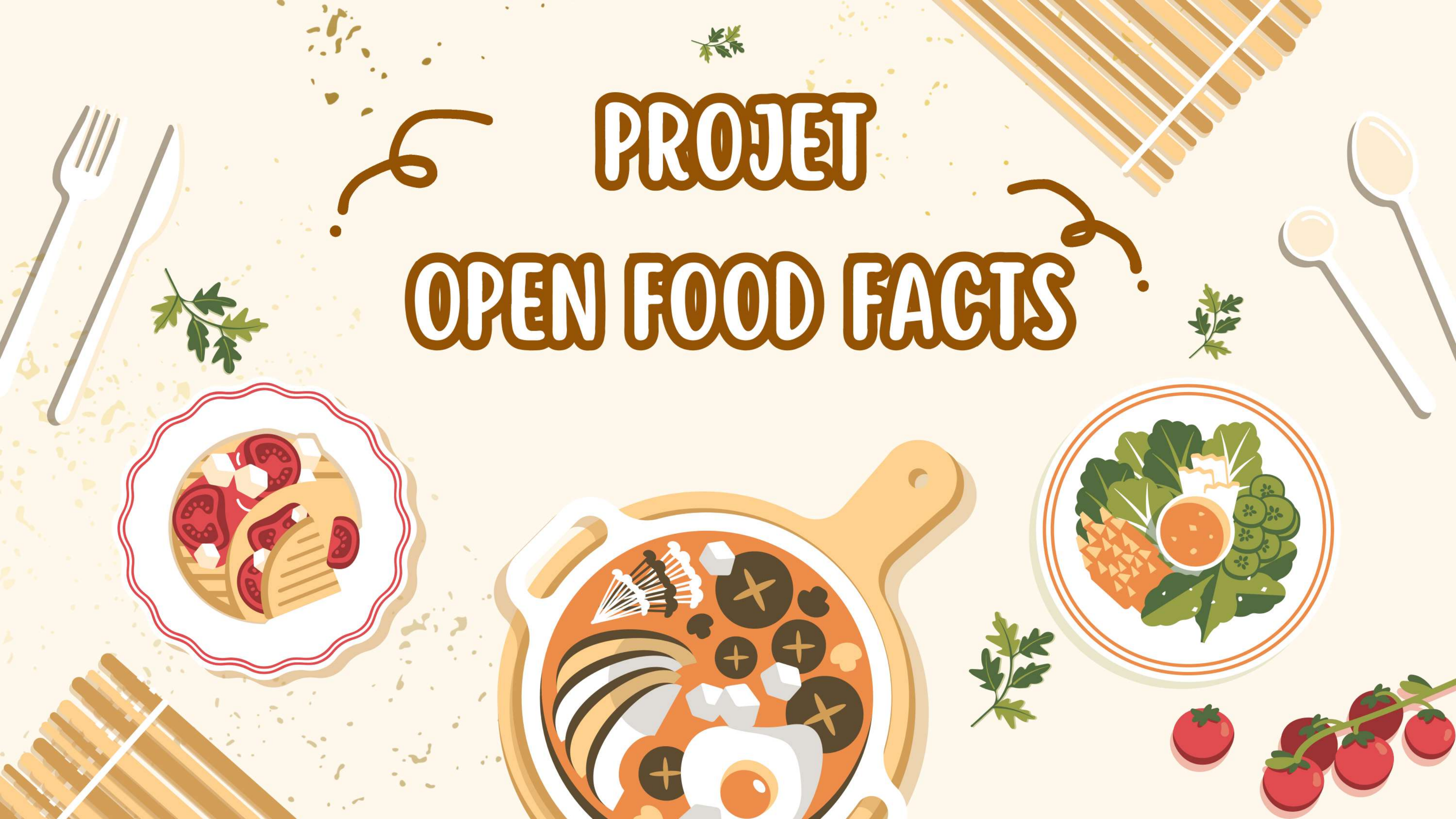


PROJET OPEN FOOD FACTS



SOMMAIRE



- **DATA PREPROCESSING**



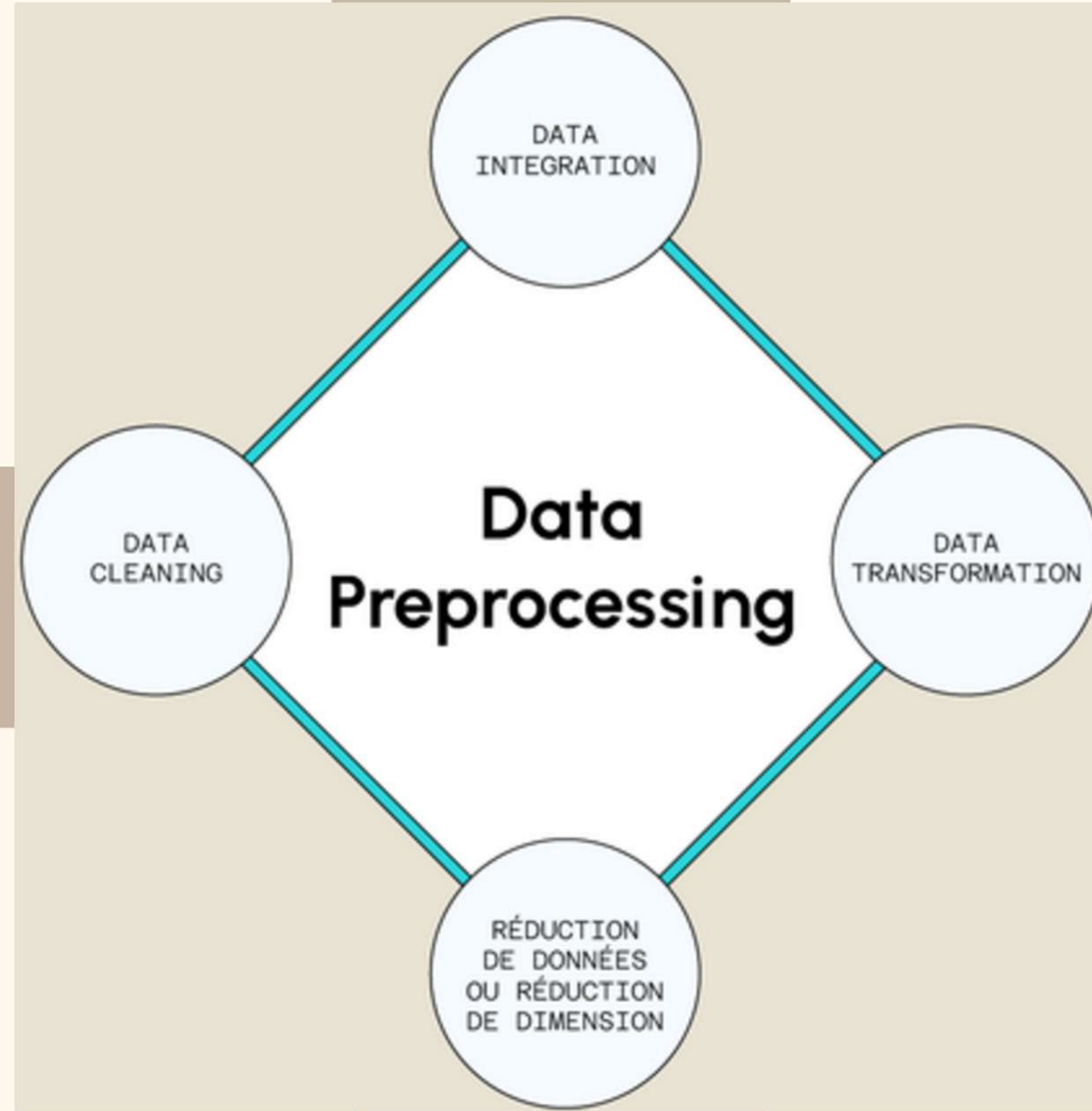
- **LES LIBRAIRIES PYTHON UTILISÉES**

- **AMÉLIORER LA SANTÉ PAR L'IA – ANALYSE DES DONNÉES OPEN FOOD FACTS**

- > **BLOC 3: CONTRIBUTION ET ORIGINE DES DONNÉES**



Combinaison de sources multiples dans un dataset, création et traitement de données différentes de manière à les harmoniser



Changement sur la donnée en elle même en lien avec les algos que nous comptons utiliser (lissage, agrégation, discrétisation, normalisation)

Nettoyage des données incorrectes, incomplètes ou manquantes (remplacement par le mode ou la moyenne, binning...)

Réduction de la quantité ou la dimension des données, de manière à améliorer les capacités de stockage et réduire les coûts de l'analyse



LES LIBRAIRIES PYTHON UTILISÉES

° **PANDAS**

° **MATPLOTLIB**

° **SEABORN**

° **MISSINGNO**

° **NUMPY**





**PANDAS A ÉTÉ UTILISÉE POUR LA MANIPULATION ET L'ANALYSE DE DONNÉES.
ELLE NOUS A PERMIS DE TRAVAILLER AVEC DES TABLEAUX ET DES DONNÉES
STRUCTURÉES, COMME LES FICHIERS CSV.**

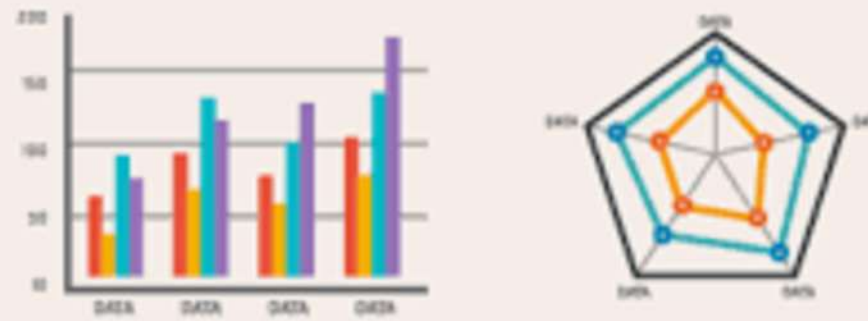




**MISSINGNO EST SPÉCIALISÉE DANS LA VISUALISATION DES DONNÉES MANQUANTES.
CETTE LIBRAIRIE A PERMIS D'IDENTIFIER LES VALEURS NULLES DANS LE JEU DE
DONNÉES.**



matplotlib



MATPLOTLIB A ÉTÉ UTILISÉE POUR CRÉER DES VISUALISATIONS (GRAPHIQUES, DIAGRAMMES, ETC.).C'EST UNE LIBRAIRIE PUISSANTE POUR PRODUIRE DES GRAPHIQUES STATIQUES EN 2D.





BASÉ SUR MATPLOTLIB, NOUS L'AVONS ÉGALEMENT UTILISÉE POUR CRÉER DES VISUALISATIONS STATISTIQUES. ELLE OFFRE DES GRAPHIQUES PLUS ESTHÉTIQUES ET PLUS FACILES À PERSONNALISER.





**NUMPY A ÉTÉ UTILISÉE POUR LA MANIPULATION DE DONNÉES NUMÉRIQUES.
CETTE LIBRAIRIE EST PERFORMANTE POUR TRAVAILLER AVEC DES
TABLEAUX/MATRICES NUMÉRIQUES**

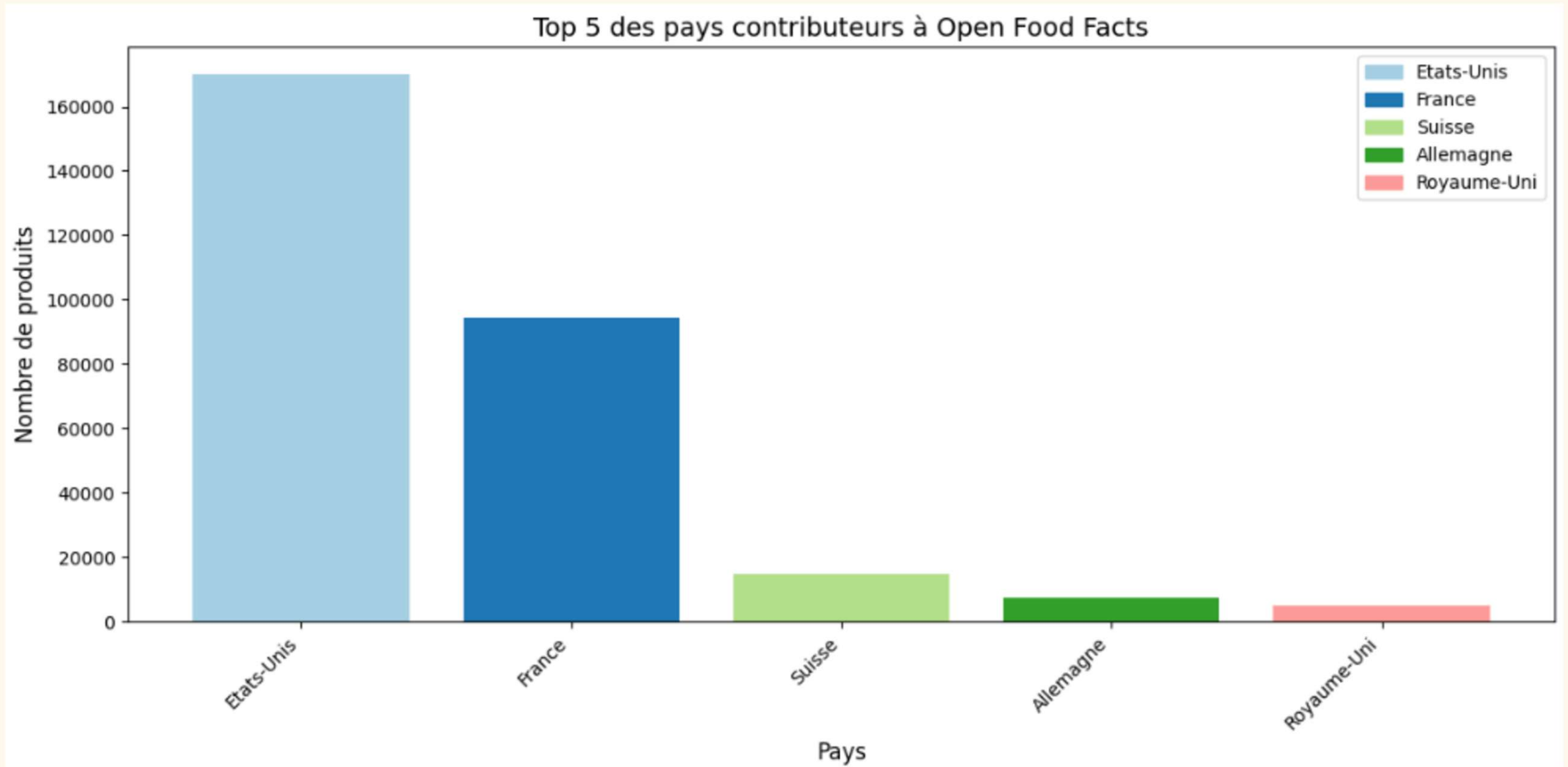


AMÉLIORER LA SANTÉ PAR L'IA – ANALYSE DES DONNÉES OPEN FOOD FACTS

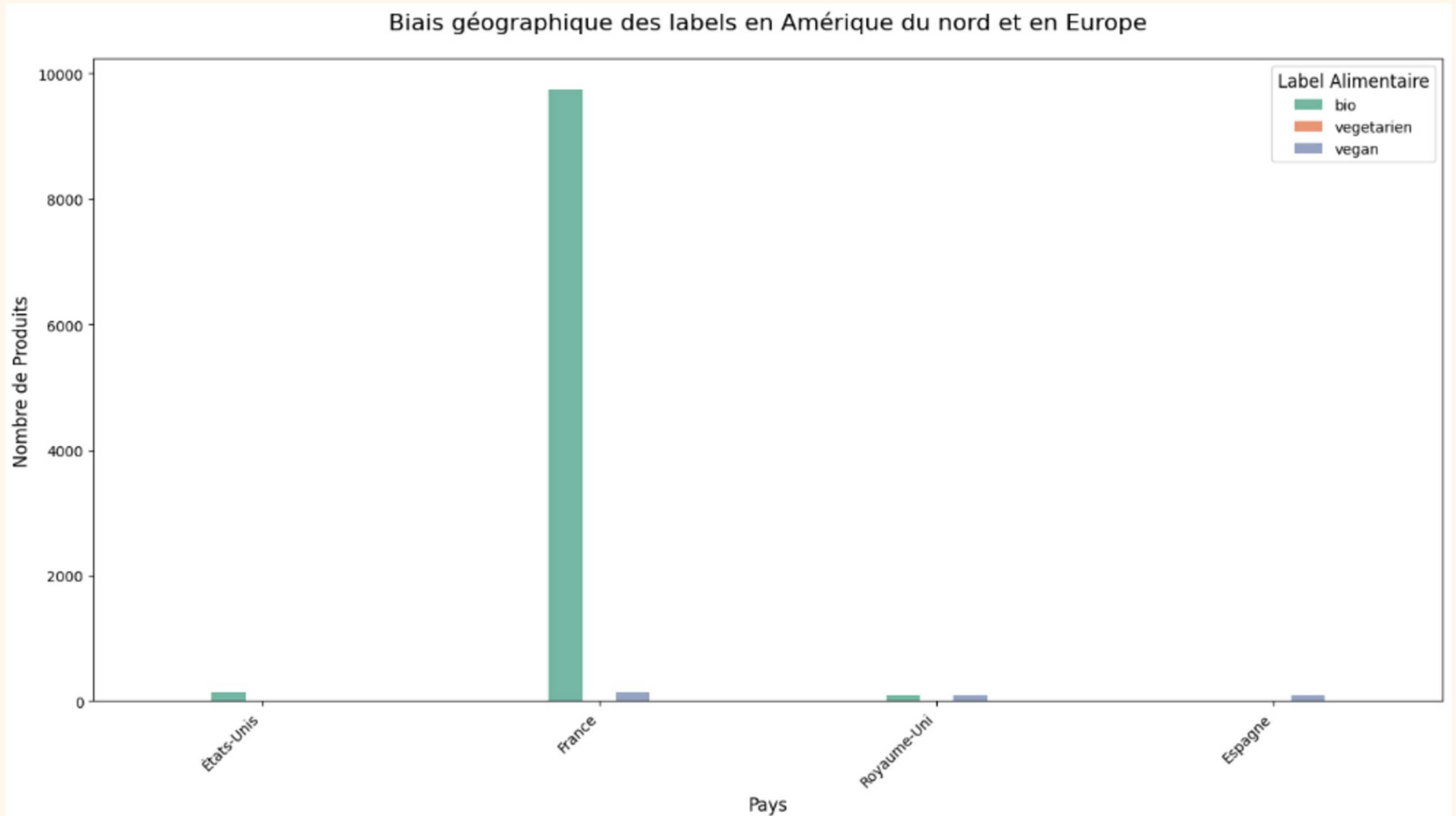
BLOC 3: CONTRIBUTION ET ORIGINE DES DONNÉES



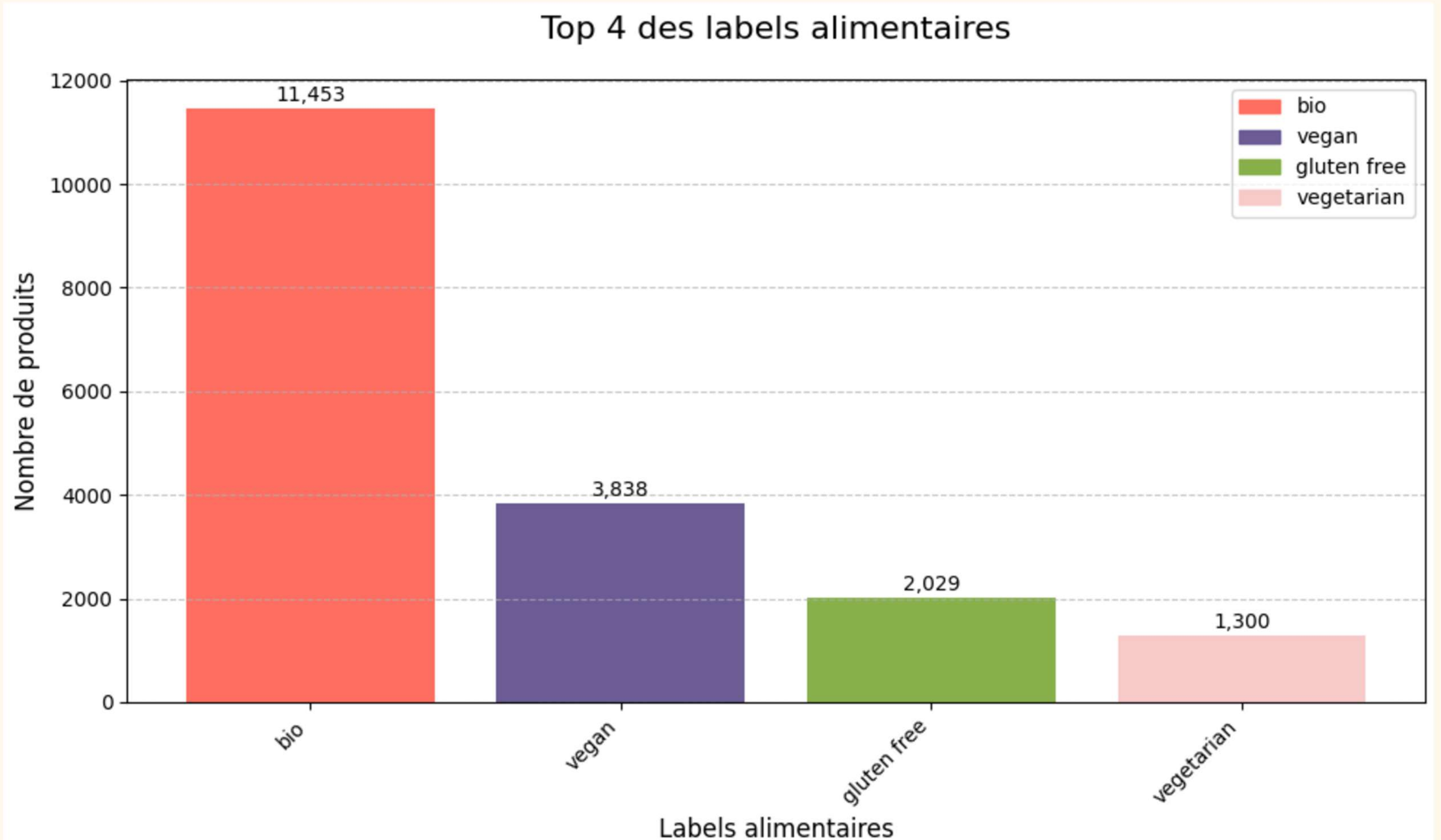
16) QUELS SONT LES PAYS LES PLUS CONTRIBUTEURS À OPEN FOOD FACTS ?



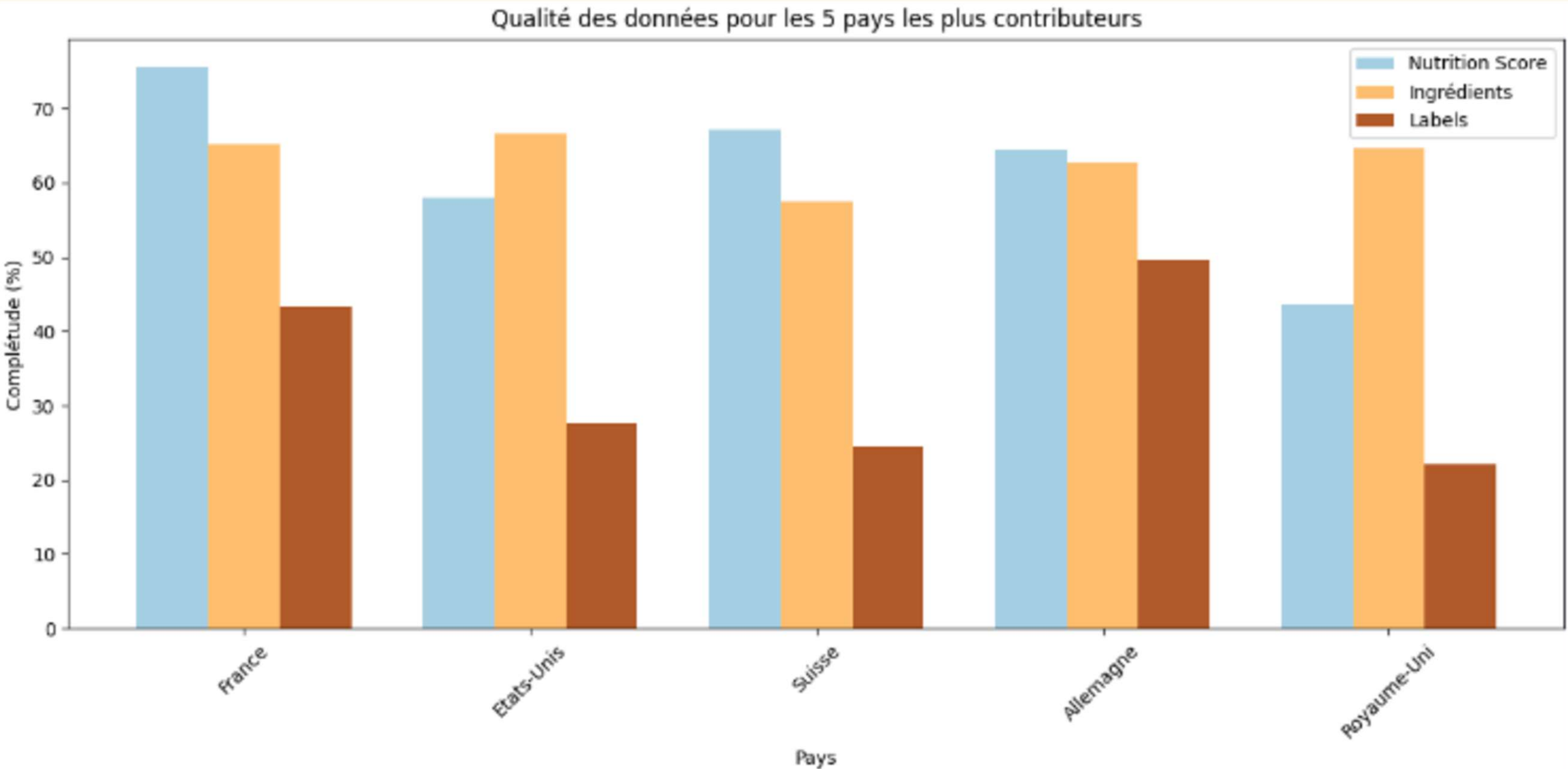
17) EXISTE-T-IL DES BIAIS GÉOGRAPHIQUES DANS LES DONNÉES ?



18) QUELS SONT LES LABELS DE QUALITÉ LES PLUS FRÉQUENTS ? SONT-ILS ASSOCIÉS À DES PRODUITS PLUS SAINS ?

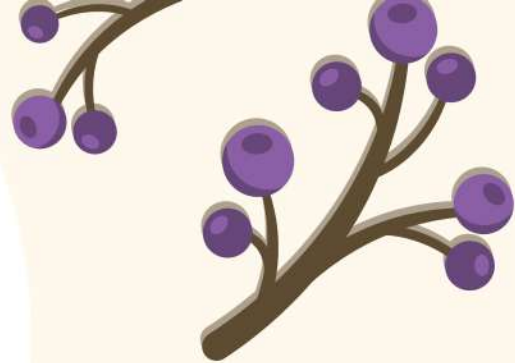


19) COMMENT LA QUALITÉ VARIE T-ELLE EN FONCTION DES CONTRIBUTEURS ?



20) PEUT-ON IDENTIFIER DES TENDANCES DANS L'EVOLUTION DES DONNÉES AU FIL DU TEMPS?





FIN

