



Projet

Collecte et traitement de la donnée

Contexte

Le but de ce projet est de mettre en place un ETL pour :

- Collecter données via API Openfoodsfact
- Les transformées
- Les stockées dans Bigquery
- Faire des analyses

Etape 1

J'ai créé mon compte de service

J'ai données les autorisations pour interagir avec bigquery

Après j'ai téléchargé la clé json
je l'ai utilisé dans mon code

Comptes de service

+ Créer un compte de service

Supprimer

+ Gérer l'accès

Actualiser

Appre

Comptes de service pour le projet "Project final laka 93110"

Un compte de service représente une identité de service Google Cloud, par exemple du code exécuté sur des VM Compute Engine, des applications App Engine ou des systèmes exécutés en dehors de Google. [En savoir plus sur les comptes de service](#)

Les règles d'administration peuvent être utilisées pour sécuriser les comptes de service, bloquer les fonctionnalités de compte de service à risque, telles que l'attribution de rôles IAM automatique, la création/l'importation de clés, ou empêcher complètement la création de comptes de service. [En savoir plus sur les règles d'administration des comptes de service](#)

Filtrer

Saisissez le nom ou la valeur de la propriété

<input type="checkbox"/>	E-mail	État	Nom ↑	Description	ID de clé	Date de création de la clé	ID client OAuth 2	Actions
<input type="checkbox"/>	aissatou11@project-final-laka-93110.iam.gserviceaccount.com	Activé	aissatou11		e8574653495e141fb74103eed5fe7e4a2b8a04a	2 juin 2025	108062147745262814949	
<input type="checkbox"/>	laka10@project-final-laka-93110.iam.gserviceaccount.com	Activé	laka10		e98c1369c8cc3a508437445c4c453fa7520e5696	2 juin 2025	111813461460104023445	

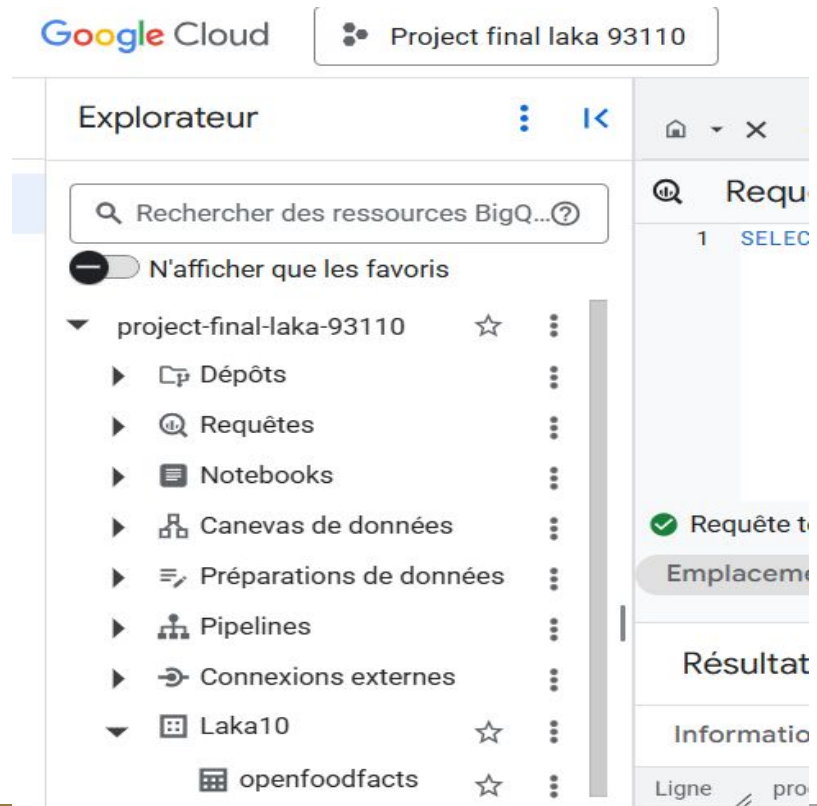
Etape 2

Pour mener à bien ce projet j'ai fait le choix d'utiliser gcp et son entrepot de données bigquery

J'ai créé mon project

Ensuite mon dataset

Après ma table



Extraction

Dans cette partie j'ai récupéré les données via l'api openfoodfacts avec un code python

The screenshot displays a Google Colab environment. On the left, a menu bar includes 'Insérer', 'Exécution', 'Outils', and 'Aide'. Below it, a file explorer shows a folder named 'openfood_referentiel.csv'. The main area contains a Python script that imports necessary libraries, sets up authentication, and defines a function to fetch product data from the OpenFoodFacts API. The script uses the 'requests' library to make a GET request to the API endpoint, checks the status code, and prints a success message if the connection is successful. It also includes a function to fetch all products in a paginated manner.

```
1 import requests
2 import pandas as pd
3 import time
4 import os
5 from google.cloud import bigquery
6
7 # Authentification GCP
8 os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "/content/project-final-laka-93110-e98c1369c8cc.json"
9
10 # Extraction globale
11 PAGE_SIZE = 1000
12 NUM_PAGES = 20
13 CSV_PATH = "/content/openfood_referentiel.csv"
14
15 # BigQuery
16 PROJECT_ID = "project-final-laka-93110"
17 DATASET_ID = "Laka10"
18 TABLE_ID = "openfoodfacts"
19 BQ_TABLE = f"{PROJECT_ID}.{DATASET_ID}.{TABLE_ID}"
20
21 # Vérification de connexion à l'API
22 def check_api_connection():
23     try:
24         response = requests.get("https://world.openfoodfacts.org/", timeout=5)
25         if response.status_code == 200:
26             print("Connexion à l'API OpenFoodFacts établie.")
27             return True
28         except Exception as e:
29             print(f"Échec de connexion à l'API : {e}")
30             return False
31
32 # Fonction : récupération sécurisée des produits
33 def fetch_all_products(page, page_size=1000):
```

On the right, a preview of the 'openfood_referentiel.csv' file is shown. The table has four columns: 'product_name', 'brands', 'stores', and 'nutrisc'. The data is displayed in a grid format, showing various products and their associated information.

product_name	brands	stores	nutrisc
Sidi Ali	Sidi Ali		a
Perly	Jaouda		unknown
Sidi Ali	sidi ali		a
Eau minérale naturelle	Les Eaux Minérales d'oulmès, Sidi Ali		a
الكوافيا	pepsi, PepsiCo	Marjane, Carrefour	a
Lait	Jaouda		b
CRISTALINE Eau De Source 1.5L	Cristaline	Carrefour, Leclerc, Auchan, Intermarché, Super U, E. Leclerc	a
Fromage blanc nature	MILKY FOOD PROFESSIONAL		a

Transformation

D



Insérer Exécution Outils Aide



Partager

Gemini



RAM
Disque



Importer

d....

:13...

responsable(s)

```
95 | return df
96 |
97 | def transform_data(df):
98 |     if df.empty:
99 |         print(" DataFrame vide : aucune transformation possible.")
100 |         return df
101 |
102 |     print(" Transformation des données...")
103 |
104 |     df['energy_kcal'] = pd.to_numeric(df['energy_kcal'], errors='coerce')
105 |     df['fat_100g'] = pd.to_numeric(df['fat_100g'], errors='coerce')
106 |     df['sugars_100g'] = pd.to_numeric(df['sugars_100g'], errors='coerce')
107 |     df['nutriscore_grade'] = df['nutriscore_grade'].str.upper()
108 |     df['product_name'] = df['product_name'].fillna("Inconnu")
109 |     df['has_label_bio'] = df['labels'].str.contains("bio", case=False, na=False)
110 |
111 |     df["nutrient_density"] = df["energy_kcal"] / (
112 |         df["proteins_100g"].astype(float).fillna(0) +
113 |         df["fat_100g"].astype(float).fillna(0) +
114 |         df["sugars_100g"].astype(float).fillna(0)
115 |     ).replace(0, 1)
116 |
117 |     print(" Données transformées.")
118 |     return df
119 |
120 | # Main pipeline
121 | def main():
122 |     if not check_api_connection():
123 |         print(" Arrêt du pipeline car l'API OpenFoodFacts est injoignable.")
124 |         return
125 |
126 |     all_products = []
127 |     for page in range(1, NUM_PAGES + 1):
128 |         print(f" Téléchargement page {page}/{NUM_PAGES}")
129 |         products = fetch_all_products(page, PAGE_SIZE)
130 |         if not products:
131 |             print(f" ⚠ Page {page} vide ou invalide. Passage à la suivante.")
```

openfood_transformed.csv x

21 to 30 of 1000 entries			
product_name	brands	stores	nutriscore_
Perrier	Nestlé, Nestlé Waters, Perrier	Magasins U,Auchan,E.leclerc,Intermarché	A
Ghayt			A
Eau de source naturelle La Source	Aïcha		A
Sidi Ali	Sidi Ali		A
المكرونة	pepsi,PepsiCo	Marjane, Carrefour	A
CRISTALINE Eau De Source 1.5L	Cristaline	Carrefour,Leclerc,Auchan,Intermarché,Super U,E.Leclerc	A
Ain Saiss	Danone	marjane	A
Eau minéral naturelle	Ain atlas		A
6111035000027	Sidi Ali		A
Ain Saiss Eau Minérale Naturelle	Danone		A

Transformation

Conversion des colonnes numériques :

energy_kcal

fat_100g

sugars_100g

Traitement des valeurs manquantes :

product_name : supprimer les valeurs "Inconnu"

Colonne calculée : Scoring nutritionnel personnalisé



$\text{energy_kcal} + \text{sugars_100g} + \text{saturated_fat_100g} - (\text{proteins_100g} + \text{fiber_100g})$

Transformation

Google Cloud

Project final laka 93110

Tapez / pour rechercher des ressources, des documents, des produits, etc

Recherche

Explorateur

Rechercher des ressources BigQuery

N'afficher que les favoris

Notebooks

Canevas de données

Préparations de données

Pipelines

Connexions externes

Laka10

openfoodfacts

vue_looker_export

vue_nutriscore_resume

vue_nutrition_detaillee

vue_nutrition_par_categorie

vue_produits_bio

vue_produits_par_enseigne

vue_produits_risques

vue_nu... lee

*Requêt... tre

Requête sans titre

Exécuter

Enregistrer

Télécharger

Partager

Programmer

Ouvrir dans

1 SELECT * FROM 'project-final-laka-93110.Laka10.vue_nutrition_detaillee' LIMIT 1000

Requête terminée

Résultats de la requête

Enregistrer les résultats

Ouvrir dans

Informations sur le job

Résultats

Graphique

JSON

Détails de l'exécution

Graphique d'exécution

nutriscore_grade		energy_kcal	sugars_100g	saturated_fat_100g	proteins_100g	fiber_100g	has_label_bio	scoring_nutrition...
Mor...	A	712.0	3.0	6.8	15.5	6.8	false	699.5
	A	358.0	4.21	0.526	11.8	10.0	false	340.936

Résultats par page: 50 1 - 50 sur 618

Chargement

Exécution Outils Aide

```
125 # === Étape 4 : Chargement dans BigQuery ===
126
127 os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "/content/project-final-laka-93110-e98c1369c8cc.json"
128
129 PROJECT_ID = "project-final-laka-93110"
130 DATASET_ID = "Laka10"
131 TABLE_ID = "openfoodfacts"
132 BQ_TABLE = f"{PROJECT_ID}.{DATASET_ID}.{TABLE_ID}"
133
134 client = bigquery.Client(project=PROJECT_ID)
135
136 job_config = bigquery.LoadJobConfig(
137     source_format=bigquery.SourceFormat.CSV,
138     skip_leading_rows=1,
139     autodetect=True,
140     write_disposition="WRITE_TRUNCATE"
141 )
142
143 with open(cleaned_csv_path, "rb") as source_file:
144     load_job = client.load_table_from_file(source_file, BQ_TABLE, job_config=job_config)
145
146 load_job.result()
147
148 table = client.get_table(BQ_TABLE)
149 print(f"{table.num_rows} lignes importées dans {BQ_TABLE}.")
150
```

Chargement

google Cloud Project final laka 93110

Tapez / pour rechercher des ressources, des documents, des produits, etc Recherche

Explorateur

Ouvrir le sélecteur de projets (Ctrl O)

Rechercher des ressources BigQ...?

N'afficher que les favoris

- project-final-laka-93110
 - Dépôts
 - Requêtes
 - Notebooks
 - Canevas de données
 - Préparations de données
 - Pipelines
 - Connexions externes
 - Laka10
 - openfoodfacts

Dépôt **Béta**

Aucun dépôt sélectionné
Sélectionnez un dépôt et un espace de travail pour afficher son contenu.

Requête sans titre

Exécuter Enregistrer Télécharger Partager Programmer Ouvrir dans Plus

1 SELECT * FROM `project-final-laka-93110.Laka10.openfoodfacts` :
Requête terminée

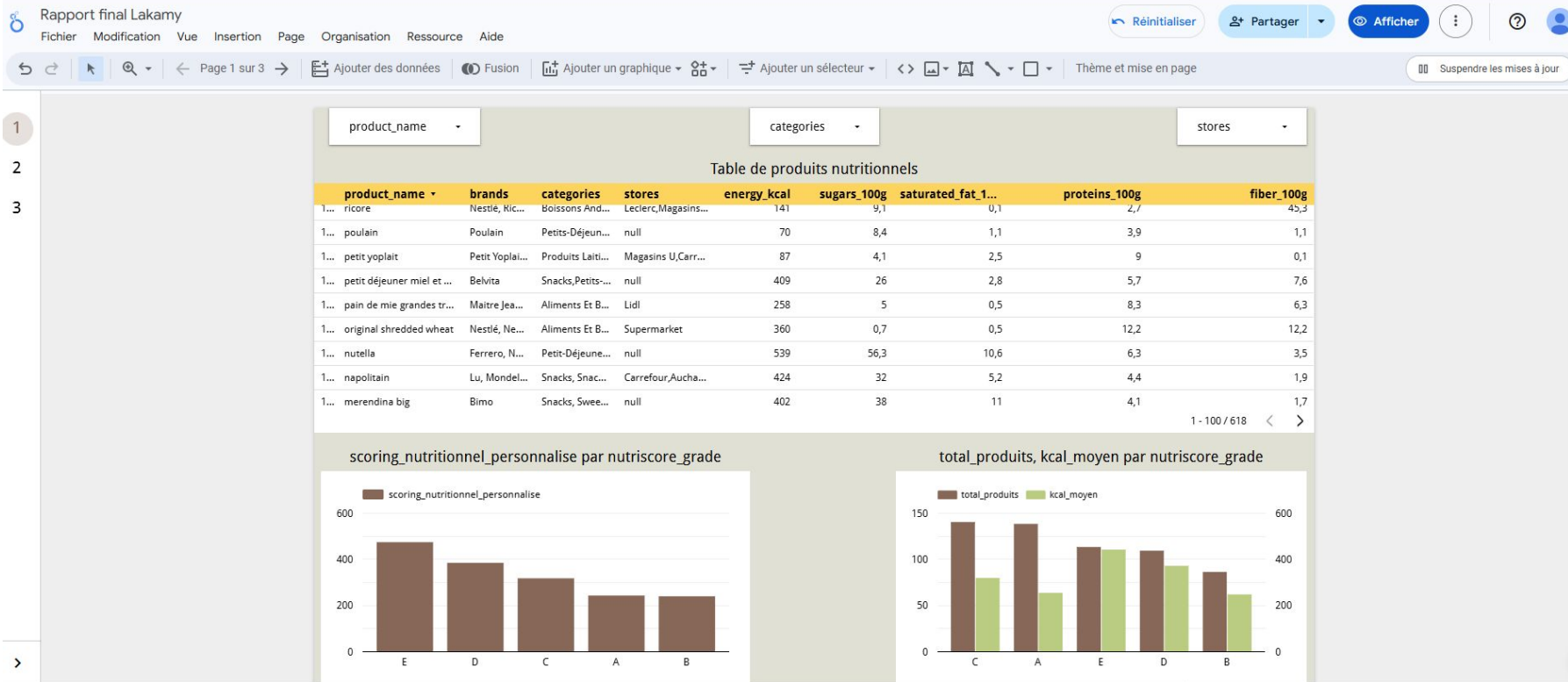
Résultats de la requête

Enregistrer les résultats Ouvrir dans

Informations sur le job Résultats Graphique JSON Détails de l'exécution Graphique d'exécution

Ligne	product_name	brands	stores	nutriscore_grade	nutrition_score_fr	energy_kcal
301	Grilletine au froment sans sucre...	Pasquierbroche Pasquier	Auchan,Cora,Lidl,Métro,Carrefo...	B	null	425.0
302	zero sin azúcares Fibra	Gullón	null	B	null	431.0
303	Figue & son	Gerblé	Franprix,Magasins U,Carrefour,...	B	null	450.0

Analyse



Analyse



Rapport final Lakamy

Fichier Modification Vue Insertion Page Organisation Ressource Aide

Réinitialiser

Partager

Afficher



Page 2 sur 3 Ajouter des données Fusion Ajouter un graphique Ajouter un sélecteur Thème et mise en page

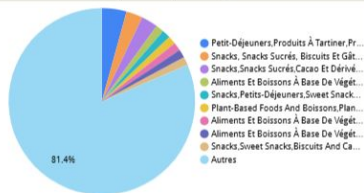
Suspendre les mises à jour

1

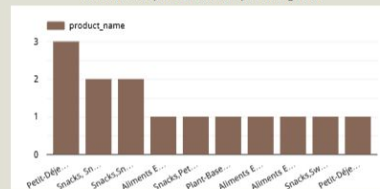
2

3

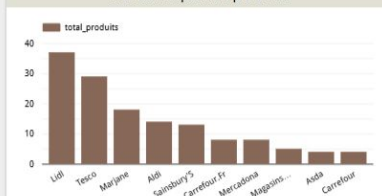
répartition par catégorie



Nombre de product_name par categories



Nombre de produits par Store



scoring_nutritionnel_personnalise

product_name	brands	categories	scoring_nutrition...
1. Huile de table équilibrée	Lesieur, Cristal	Matières Gr...	915
2. Huile de table lio	null	En-Plant-Ba...	914
3. Houllor	Huillor	Plant-Based...	912.6
4. Huile d'olive	Oued Souss	Aliments Et ...	912
5. Huile de tournesol	Lesieur Cristal	Aliments Et ...	911.3
6. Huile Lesieur 3 Graines	Lesieur	Aliments Et ...	910
7. Huile d'olive vierge extra	Lidl, Primadona	Plant-Based...	842

1 - 100 / 127

Analyse

Rapport final Lakamy

Fichier Modification Vue Insertion Page Organisation Ressource Aide

Réinitialiser

Partager

Page 3 sur 3 Ajouter des données Fusion Ajouter un graphique Ajouter un sélecteur Thème et mise en page

Produits très faibles en sucre

	product_name	categories	sugars_100g
1.	Granatapfel Sauerkirsche Fr...	Imbiss, Süßer Snack, Süßwar...	0,4
2.	Raw kimchi	Plant-Based Foods And Boiss...	0,4
3.	Chickpeas In Water	Plant-Based Foods And Boiss...	0,4
4.	Bitter Extra Kraftig	Snacks,Snacks Sucrés,Cacao ...	0,4
5.	Inconnu	En:Boissons And Boissons Pr...	0,4
6.	Sabra Houmous Extra	Plant-Based Foods And Boiss...	0,4
7.	Instant Nudeln - Chicken Fla...	Pflanzliche Lebensmittel Un...	0,4
8.	Organic Vegetable Stock Cubes	Aliments Et Boissons À Base ...	0,3
9.	Pork Sausages	Viandes And Their Products, ...	0,27
1 - 100 / 169			

Vue globale

	categories	stores	nutriscor...	scoring_n...
77.	Snacks,Sweet Snacks,Cocoa And Its Pro...	Tesco	E	627
78.	Snacks,Snacks Sucrés,Biscuits Et Gâtea...	null	E	459,53
79.	Petits-Déjeuners,Spreads,Sweet Spread...	Tesco	E	413,49
80.	Condiments, Sauces, Moutardes, Mout...	Amazon	E	null
81.	Condiments,Sauces,Sauces Barbecue,S...	Marks & Spe...	E	null
82.	Snacks,Sweet Snacks,Biscuits And Cake...	null	E	560,4
83.	Produits Laitiers, Matières Grasses, Pro...	Cora,Franpri...	E	781,1
84.	Petit-Déjeuners, Produits-À-Tartinier, Pr...	null	E	596,1
95.	Petit-Déjeuners, Produits-À-Tartinier, Pr...	Ladbro's M...	E	605,1
1 - 100 / 844				