

**AUTOMATIZACIÓN DE CONSULTAS DE DATOS MULTISECTORIALES:
EMPLEANDO LA API USDA CON PYTHON EN INVESTIGACIÓN
AGROALIMENTARIA¹****AUTOMATION OF MULTISECTORAL DATA QUERIES: USING THE USDA API WITH
PYTHON IN AGRI-FOOD RESEARCH**

Muñoz-Mortera M. del S.; Hernández-Nolasco Z.; Hernández-Cázares A.S.; Hidalgo-Contreras J.V.*

*Colegio de Postgraduados, Campus Córdoba, Carretera Federal Córdoba-Veracruz km 6 348,
Congregación Manuel León, Amatlán de los Reyes, Veracruz, México. C. P. 94946*

* E-mail: jvhidalgo@colpos.mx

Fecha de envío: 12, mayo, 2025

Fecha de publicación: 20, julio, 2025

Resumen:

El acceso automatizado a datos agroalimentarios mediante Interfaces de Programación de Aplicaciones (APIs) públicas representa una herramienta fundamental para fortalecer la investigación científica y la toma de decisiones en el sector agrícola. Este estudio tuvo como objetivo obtener información automatizada relevante sobre la producción de canola (*Brassica napus*) (2020–2026) y su perfil nutricional. Para ello, se diseñó una metodología de extracción y análisis de datos utilizando Python y las APIs proporcionadas por el Departamento de Agricultura de los Estados Unidos (USDA): NASS QuickStats y FoodData Central. La API de NASS QuickStats se empleó para recopilar datos estadísticos sobre la producción de canola, en tres etapas, configuración inicial, adquisición de datos y procesamiento. Por otro lado, la API de FoodData Central permitió obtener y organizar la composición nutricional del aceite de canola. En ambos casos, los resultados se organizaron en archivos .xlsx. La automatización del proceso se realizó mediante el uso de bibliotecas de Python como *requests*, *pandas* y *openpyxl*, optimizando la descarga, el procesamiento y la exportación de datos. Este proceso permitió la sistematización eficiente de grandes conjuntos de datos, redujo la incidencia de errores humanos y mejoró el acceso a información estratégica para el sector agroalimentario. En conclusión, la implementación de esta metodología demostró ser una solución eficaz, rápida y selectiva, con aplicaciones potenciales en la investigación básica y el análisis de exportaciones agroalimentarias.

Palabras clave: Automatización, API, investigación, agroalimentaria, datos.

¹Trabajo derivado de Proyectos de Investigación del Doctorado en Innovación Agroalimentaria Sustentable (DCIAS).

Abstract:

Automated access to agri-food data through public Application Programming Interfaces (APIs) represents a fundamental tool for strengthening scientific research and decision-making in the agricultural sector. This study aimed to obtain relevant automated information on the production of canola (*Brassica napus*) (2020—2026) and its nutritional profile. To this end, a data extraction and analysis methodology was designed using Python and the APIs provided by the United States Department of Agriculture (USDA): NASS QuickStats and FoodData Central. The NASS QuickStats API was used to collect statistical data on canola production, in three stages: initial configuration, data acquisition and processing. On the other hand, the FoodData Central API made it possible to obtain and organize the nutritional composition of canola oil. In both cases, the results were organized in .xlsx files. The automation of the process was carried out through the use of Python libraries such as requests, pandas and openpyxl, optimizing the download, processing and export of data. This process allowed for the efficient systematization of large data sets, reduced the incidence of human errors and improved access to strategic information for the agri-food sector. In conclusion, the implementation of this methodology proved to be an effective, fast and selective solution, with potential applications in basic research and the analysis of agri-food exports.

Keywords: Automation, API, research, agri-food, data.

INTRODUCCIÓN

El análisis de grandes volúmenes de datos agrícolas y alimentarios se ha convertido en una herramienta clave para la investigación científica y la toma de decisiones en el sector agroalimentario. En este contexto, el Departamento de Agricultura de los Estados Unidos (USDA) pone a disposición del público diversas plataformas institucionales, como AgData Commons, Agricultural Resource Management Survey (ARMS), QuickStats y FoodData Central, que contienen información actualizada y estructurada sobre producción agrícola, precios de mercado, composición nutricional y factores económicos (U.S. Department of Agriculture, n.d.-a, b, c).

El Servicio Nacional de Estadísticas Agrícolas de la USDA (USDA/NASS) es la principal fuente de datos agrícolas en EE. UU. Entre sus actividades destaca la realización del censo agrícola, llevado a cabo cada cinco años, con el propósito de recopilar información detallada sobre la producción agrícola, la economía y la demografía del sector (USDA, 2024). Además, publica informes sobre cultivos y ganado, datos fundamentales para el análisis de tendencias del mercado. Asimismo, desarrolla encuestas especializadas que aborda aspectos como el empleo agrícola, el uso de insumos químicos y la producción orgánica. Para facilitar el acceso a estos datos, USDA/NASS dispone de la plataforma en línea QuickStats (Figura 1), que permite la consulta de estadísticas agrícolas de forma

rápida. La Junta Nacional de Estadística Agropecuaria (NASB) supervisa la recopilación y procesamiento de la información, garantizando que los datos generados respondan a las necesidades del sector agroalimentario.

The image shows the USDA National Agricultural Statistics Service (NASS) QuickStats website. The header includes the USDA logo, the text "United States Department of Agriculture National Agricultural Statistics Service", and a search bar labeled "Barra de búsqueda principal". Below the header is a navigation menu with links: Data & Statistics, Publications, Newsroom, Surveys, Census, About NASS, Contact Us, and Help. The main content area features a "Today's Reports" section with a date of Jun 03, 2025, and a "CENSUS OF AGRICULTURE NAVAJO NATION PROFILES NOW AVAILABLE" banner. A "Reporte de datos" section includes a map of the United States with a "Select a State" dropdown and a "Find a Regional or State Field Office" link. A "Visualización de datos" section features a "COMPLETE YOUR SURVEY ONLINE" button and a "DATA VISUALIZATION" graphic. A "Novedades" (News) section lists recent news releases. The footer contains a tagline "Providing Timely, Accurate and Useful Statistics in Service to U.S. Agriculture" and a list of links including NASS Home, USDA.gov, Data.gov, USA.gov, WhiteHouse.gov, ESMIS, No Fear Act - NASS Data, Accessibility Statement, FOIA, Privacy Policy and Assessments, Security Statement, Information Quality, Non-Discrimination Statement, Guidance Documents, USDA Policies and Links, and AskUSDA.

Figura 1. Plataforma en línea QuickStats National Agricultural Statistics Service del USDA para la obtención de datos estadísticos.

Figure 1. QuickStats online platform for obtaining statistical data from the USDA National Agriculture Statistics Service.

Por otro lado, FoodData Central (FDC) del USDA es una plataforma integral lanzada en 2019 que recopila datos científicos y estandarizados sobre la composición de alimentos (Figura 2). Esta herramienta de acceso abierto ofrece información detallada sobre nutrientes, componentes bioactivos, etiquetas de alimentos comerciales y productos analizados en condiciones experimentales. Además, esta plataforma está dirigida a investigadores, profesionales de la salud, formuladores de políticas, desarrolladores de productos y educadores. Estas bases de datos han sido utilizadas ampliamente en estudios científicos sobre calidad alimentaria, predicción de rendimientos, planificación agrícola, nutrición y sostenibilidad alimentaria (Sebastian et al., 2025; Bastian y Bienias, 2023; Ma et al., 2021; Prasai et al., 2021). Sus cinco bases de datos principales son: Foundation Foods, SR Legacy, FNDDS, Experimental Foods y Branded Foods (USDA, 2024).

El uso de APIs del USDA junto con Python optimiza el análisis de datos agroalimentarios al facilitar el acceso automatizado a información agrícola del NASS, incluyendo variables animales, vegetales, económicas y ambientales (Potter, 2019). Python, con bibliotecas como *pandas* y *scikit-learn*, permite manipular datos, aplicar estadísticas y desarrollar modelos de aprendizaje automático para predecir rendimientos e identificar factores que afectan la productividad de los cultivos agrícolas, materias primas y bienes naturales (Lal, n.d.; Yadav et al., 2024). Esta integración impulsa decisiones basadas en la inteligencia artificial (IA), esenciales para la agricultura de precisión y prácticas sostenibles (Elbaşı et al., 2024). Además, promueve la interoperabilidad entre conjuntos de datos, fomentando la innovación en el sector agrícola (Siné et al., 2014). Este enfoque se alinea con la agricultura digital, centrada en estrategias basadas en datos para mejorar la eficiencia y anticipar problemas como las enfermedades de cultivos (Roscher et al., 2023; Prasai et al., 2021).

Aunque útiles, estas herramientas presentan desafíos como el manejo de grandes volúmenes de datos. Las API permiten un acceso eficiente a la información, pero su uso requiere habilidades tecnológicas avanzadas y conocimientos en procesamiento de datos. En este contexto, Python se ha consolidado como una opción robusta y versátil para la automatización de tareas, la visualización de datos y realización de análisis estadísticos, en bibliotecas como *Pandas*, *Matplotlib* y *Seaborn* (Sarkar, 2020).

Barra de búsqueda principal

FoodData Central
USDA's comprehensive source of food composition data with multiple distinct data types

Search all foods in all data types

Foundation Foods
Analytical data/metadata on commodity and minimally processed food samples.
Source: USDA
Update: Twice annually - April & October
[Browse](#) [Download](#) [Read More](#)

SR Legacy
Historical data derived from analyses, calculations, and published literature.
Source: USDA
Update: Final update was in 2018
[Browse](#) [Download](#) [Read More](#)

FNDDS
Data applied to analyze foods/beverages reported in What We Eat in America, NHANES.
Source: USDA Foundation Foods & SR Legacy
Update: Every 2 years
[Browse](#) [Download](#) [Read More](#)

Experimental Foods
Data from foods published in peer-reviewed journals in collaboration with USDA.
Source: Researchers
Update: Twice annually (if available)
[Browse](#) [Read More](#)

Branded Foods
Label data collected through a public-private partnership.
Source: Commercial food brands
Update: Monthly
[Browse](#) [Download](#) [Read More](#)

Secciones de bases de datos disponibles

Novedades

New for April 2025

- Now Available in Experimental Foods; [USDA, FDA and ODS-NIH Database for the Iodine Content of Common Foods](#)
- [FNS Child Nutrition Database Updates](#)
- New Foods: [Inventory and Update Log](#)
- [Future Updates](#)

FNS Child Nutrition Database

The FNS Child Nutrition Database (CNDB) is now hosted on FDC! Find the CNDB and its related documents on the [CNDB page](#).

At this time, only a basic view of search results is available for viewing on mobile devices. Advanced filter features, such as searching by data type, are not yet mobile-enabled and are available only in desktop view. Users are encouraged to use a desktop computer to conduct food item searches.

Quick Start Videos

- [Getting Started on FoodData Central](#)
- [Branded Foods](#)
- [Food and Nutrient Database for Dietary Studies \(FNDDS\)](#)
- [Foundation Foods](#)
- [SR Legacy Foods](#)

View the [About Us](#) page for essential information about FoodData Central data types and how to use this system.

FoodData Central is managed by the Agricultural Research Service's Beltsville Human Nutrition Research Center and hosted by the National Agricultural Library.

Licensing

USDA FoodData Central data are in the public domain and they are not copyrighted. They are published under [CC0 1.0 Universal \(CC0 1.0\)](#)

No permission is needed for their use, but we request that users list FoodData Central as the source of the data, and when possible, notify us of the product that uses the data so we may better track how FDC is being used in the public domain. The suggested citation is:
U.S. Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center.
FoodData Central. [Internet]. [cited (enter date)]. Available from <https://fdc.nal.usda.gov/>.

Información adicional

Enlaces al API

[Return to Top](#)

[Home](#) [Food Search](#) [Component Search](#) [Data Type Documentation](#) [Download Data](#) [API Guide](#) [Help](#) [FAQ](#) [About Us](#) [Contact FoodData Central](#)

Figura 2. Plataforma en línea FoodData Central (FDC) del USDA para la obtención de datos nutricionales.

Figure 2. USDA FoodData Central (FDC) online platform for retrieving nutritional data.

En este contexto, se plantean dos casos de estudio que muestran el uso práctico de las APIs del USDA en el análisis agroalimentario. El primero consiste en recuperar y analizar datos estadísticos sobre la producción de canola a nivel nacional y estatal (Texas) durante el periodo 2020–2026 mediante la API NASS Quick Stats del USDA. El segundo se enfoca en obtener el perfil nutricional del aceite de canola utilizando la API FoodData Central, con el objetivo de establecer una metodología eficiente para la adquisición, extracción, clasificación, análisis y almacenamiento de datos agroalimentarios. Ambas propuestas aprovechan tecnologías de acceso abierto para promover la transparencia y accesibilidad en la investigación agroalimentaria.

MATERIALES Y MÉTODOS

Para la gestión y procesamiento de datos se empleó el lenguaje de programación Python 3.13.1., junto con bibliotecas especializadas como *requests*, *pandas* y *openpyxl*. La obtención de datos se llevó a cabo mediante el acceso de interfaces de programación de aplicaciones (APIs) públicas del Departamento de Agricultura de los Estados Unidos (USDA), con autenticación mediante claves de acceso oficiales. El diagrama de flujo que describe el procedimiento aplicado en los dos estudios de caso se presenta en la Figura. 3.

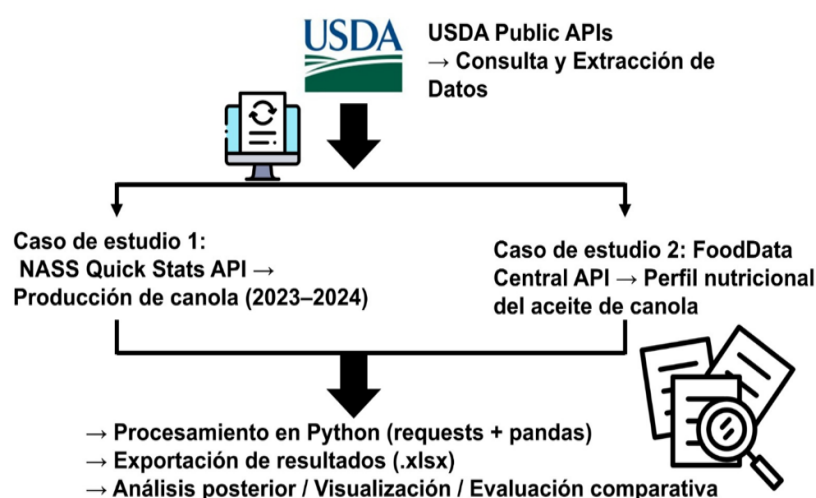


Figura 3. Diagrama de flujo para la automatización de consultas de datos multisectoriales sobre producción de Canola y Aceite mediante la API USDA con Python.

Figure 3. Flowchart for automating multisectoral data queries on Canola and Oil production using the USDA API with Python.

Caso de estudio 1. Canola en el NASS

Se implementó una metodología en tres etapas para adquirir y procesar los datos agrícolas del USDA. En la configuración preliminar se utilizaron las bibliotecas *requests* para realizar solicitudes HTTP a la API y *pandas* para manipular datos, definiendo una clave de API (API_KEY) y la URL base (BASE_URL). Se especificaron las materias primas «canola» y «CANOLA» para un rango de cinco años (2020-2026) y se correlacionaron las variables como «PLANTADO» a «ÁREA PLANTADA» para estandarizar su interpretación. En la adquisición de datos, se desarrolló la función *get_data*, que iteró sobre variables y años, enviando solicitudes GET a la API con parámetros específicos (*commodity_desc*, *año*, *format=JSON*) y manejando errores con un bloque try-except, registrando datos ausentes como «N/A». Los datos se almacenaron en una lista de diccionarios y se convirtieron en un *DataFrame* de *pandas*, consolidándolos en un diccionario de *DataFrames* (*all_data*) por año. En el procesamiento y exportación, se transformaron los datos mediante una tabla dinámica (*pivot_table*), organizando variables en columnas con materias primas y se verificó la integridad de los *DataFrames* antes de exportarlos a un archivo Excel («DATA_CANOLA_USDAV1.api.xlsx») usando PD ExcelWriter con *openpyxl*, asignando cada año a una hoja distinta para mantener consistencia.

Caso de estudio 2 Aceite de Canola en el FoodData Central

En el estudio de Caso 2, se accedió a la API *FoodData Central* para obtener el perfil nutricional del aceite de canola, seleccionado como modelo el aceite vegetal de uso alimentario. El procedimiento incluyó dos etapas: (1) identificación del alimento mediante el endpoint */foods/search*, obteniendo su código único (*fdcId*), y (2) recuperación de su composición nutricional detallada a través del endpoint */food/{fdcId}*. Las variables extraídas fueron , el nombre del nutriente, cantidad y unidad, que se organizaron en un *DataFrame* y se exportaron en un archivo Excel.



RESULTADOS Y DISCUSIÓN

Caso de estudio 1

La ejecución del *script* o serie de comandos genera un archivo Excel estructurado en hojas anuales, cada una conteniendo datos organizados por *commodity* y estado, con columnas que representan las variables seleccionadas. La metodología empleada garantiza la coherencia de los datos, incluso en casos de respuestas incompletas de la API, y permite su uso inmediato en análisis estadísticos o visualizaciones. Si bien el API es una herramienta robusta, presenta ciertas limitaciones en la solicitud de variables. Cuando el número de variables es elevado o las descripciones contienen frases extensas, el acceso a la API puede verse restringido, y mostrará un error de autenticación de la APIKEY como se muestra en la (Figura 4).

La integración o evaluación en el código de solicitud de productos básicos, en relación con las variables utilizadas para componer los lexemas, en sus manifestaciones en minúscula y/o mayúscula, cuando solo se emplea una variante, genera un sesgo informativo, que se limita exclusivamente a las escritas, por ejemplo, en mayúsculas; por lo tanto, es imperativo reconocer también que las consultas deben realizarse en inglés en relación con cualquiera de los productos que se pretenda investigar y examinar. La metodología es escalable y adaptable a otras commodities y/o productos que se disponen en la USDA QuickStats. Sin embargo, depende de la estabilidad de la API y la validez de la clave de API. Para futuras mejoras dentro del manejo y las APIS, es posible incluir la integración de validaciones adicionales para datos anómalos o la automatización de la actualización periódica de los datos.

La secuencia de comandos (*script*) del caso de estudio generó un perfil estadístico que incluye datos sobre el producto, área plantada, área cosechada, rendimiento, producción, existencias, exportaciones, importaciones y precio recibido, como se detalla en el Cuadro 1 para el año 2020, con base en la información extraída de NASS. Para fines prácticos, no se presentan los resultados desglosados por año para el período 2020-2026.

El Cuadro 1 presenta el proceso de recolección de datos sobre la producción de canola mediante una secuencia de comandos (*script*). Este puede ser programado para recuperar información de uno o varios productos (*commodities*) de forma simultánea, aunque su eficacia dependerá de la capacidad del sistema de cómputo utilizado, en particular de la memoria RAM, el procesador y la longitud de la clave API (APIKEY) al

solicitar cada variable. Los datos obtenidos a través del *script* permiten construir una base de datos con múltiples variables y un amplio rango de años. Esta información puede emplearse en distintos contextos de investigación, por ejemplo, para analizar la superficie sembrada, la extensión de los cultivos, la producción agrícola de canola en relación con los precios recibidos, o las existencias disponibles, entre otros. Este proceso es eficiente, ya que la descarga masiva de datos a través de la API permite seleccionar con precisión la información a analizar y prepararla previamente para una visualización eficiente, ahorrando tiempo y eliminando la necesidad de trabajo manual.

Cuadro 1. Perfil Estadístico de producción de canola en USA para el 2020.

Table 1. Statistical profile of canola production in the USA for 2020.

Commodity	State	Area planted (Acres)	Area harvested (Acres)	Yield (Acres)	Production (Bushels)	Stocks (Tons)	Exports (Tons)	Price received (Dollars Per Bushels)
Canola	Idaho	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Canola	Kansas	4,500	2,300	1,790	502,000	N/A	N/A	N/A
Canola	Minnesota	50,000	48,000	1,570	14,168,000	N/A	N/A	N/A
Canola	Montana	9,300	9,300	2,550	38,373,000	N/A	N/A	N/A
Canola	North Dakota	83,500	83,400	2,070	548,497,000	N/A	N/A	N/A
Canola	Oklahoma	10,500	7,000	1,530	1,928,000	N/A	N/A	N/A
Canola	Oregón	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Canola	Washington	91,000	90,000	2,230	40,541,000	N/A	N/A	N/A
Canola	Other states	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Canola	Us total	1,831,000	1,801,300	1,942	644,009,000	481,897,000	N/A	N/A

Nota: Datos obtenidos del Servicio Nacional de Estadísticas Agrícolas del Departamento de Agricultura de Estados Unidos (USDA/NASS), (USDA, 2024). Variables: *Commodity* (producto), *State* (estado), *Area Planted* (área plantada), *Area harvested* (área cosechada), *Yield* (rendimiento), *Production* (producción), *Stocks* (existencias), *Exports* (exportaciones) y *Price received* (precio recibido). N/A = Datos ausentes.



```
# "CÓDIGO PARA EXTRAER DATOS DE LA API DE NASS, UNA DE LAS BASES DE
DATOS DE LA USDA"
# Estudio de caso: Análisis de la producción de Canola en EE.UU.
entre 2020-2025

# Importar Las Librerías necesarias

import requests
import pandas as pd

API_KEY = "TU APYKEY" # Reemplaza con tu propia clave de API (solicitarla
en La página de NASS)
BASE_URL = "https://quickstats.nass.usda.gov/api/api_GET/" # URL base de
La API

commodities = ["CANOLA", "canola"] #lista de commodities
# Lista de años para Los cuales se desea obtener datos
years = list(range(2020, 2026)) # Cambiar el rango de años según sea
necesario (5 años)

# Cambiar nombres al formato deseado
variables_deseadas = {
    "PLANTED": "AREA PLANTED",
    "HARVESTED": "AREA HARVESTED",
    "YIELD": "YIELD",
    "PRODUCTION": "PRODUCTION",
    "STOCKS": "STOCKS",
    "EXPORTS": "EXPORTS",
    "IMPORTS": "IMPORTS",
    "PRICE RECEIVED": "PRICE RECEIVED"
}

def obtener_datos(commodity, year):
    df_list = []
    for short_desc, nombre_columna in variables_deseadas.items():
        params = {
            "key": API_KEY,
            "commodity_desc": commodity,
            "year": year,
            "format": "JSON"
        }
        try:
            response = requests.get(BASE_URL, params=params)
            data = response.json()
            results = data.get("data", [])
            for item in results:
                if short_desc in item.get("short_desc", "").upper():
                    df_list.append({

```

```
                    for item in results:
                        if short_desc in item.get("short_desc", "").upper():
                            df_list.append({
                                "Commodity": item.get("commodity_desc", "N/A"),
                                "State": item.get("state_name", "US TOTAL"),
                                "Year": item.get("year", "N/A"),
                                "Variable": nombre_columna,
                                "Value": item.get("Value", "N/A")
                            })
                    except Exception:
                        df_list.append({
                            "Commodity": commodity,
                            "State": "N/A",
                            "Year": year,
                            "Variable": nombre_columna,
                            "Value": "N/A"
                        })
                    return pd.DataFrame(df_list)

# Recolectar Los datos
all_data = {year: pd.DataFrame() for year in years}

for year in years:
    dfs = []
    for commodity in commodities:
        print(f"Obteniendo datos de {commodity} para {year}")
        df = obtener_datos(commodity, year)
        dfs.append(df)
    all_data[year] = pd.concat(dfs, ignore_index=True)

# Guardar en Excel
nombre_archivo = "Data_Canola_USDAV1.API.xlsx"
with pd.ExcelWriter(nombre_archivo, engine="openpyxl") as writer:
    for year, df in all_data.items():
        if df.empty or not all(col in df.columns for col in ["Commodity",
"State", "Variable", "Value"]):
            print(f" Datos incompletos para el año {year}. Saltando...")
            continue

        pivot_df = df.pivot_table(
            index=["Commodity", "State"],
            columns="Variable",
            values="Value",
            aggfunc="first"
        ).reset_index()

        # Asegurar orden de columnas
        columnas_deseadas = ["Commodity", "State"] +
list(variables_deseadas.values())
        pivot_df = pivot_df.reindex(columns=columnas_deseadas)

        pivot_df.to_excel(writer, sheet_name=str(year), index=False)

print(f" Archivo guardado como '{nombre_archivo}'")
```

Figura 4. Script en Python para la consulta automatizada del perfil estadístico de la Canola en USDA/NASS.

Figure 4. Python script for automated query of the Canola statistical profile from USDA/NASS.

Caso de estudio 2

Se utilizó con éxito la API FoodData Central del USDA para automatizar la recuperación del perfil nutricional del aceite de canola mediante una secuencia de comandos (*script*) en Python. Este código permitió llevar a cabo dos acciones principales, identificar el alimento mediante el parámetro de búsqueda "Canola oil" utilizando el endpoint `/foods/search`; y extraer su composición nutricional detallada a través del endpoint `/food/{fdcId}`. La Fig. 5 muestra el *script* en Python desarrollado para ejecutar esta consulta automatizada. El *script* recuperó y organizó automáticamente los datos en un *DataFrame* de *pandas*, generando una tabla estructurada con los siguientes campos: nombre del nutriente, cantidad y unidad de medida (Cuadro 2). Posteriormente, los datos fueron exportados a un archivo de Excel denominado `nutrientes_aceitecanola.xlsx`, con una hoja individual.

Los resultados obtenidos mostraron una composición nutricional detallada del aceite de canola. Este aceite está compuesto exclusivamente por grasa (100 g/100 g), sin aporte de proteínas ni carbohidratos, lo que le confiere un valor energético de 900 kcal por cada 100 g. Su perfil lipídico se caracterizó por un elevado contenido de ácidos grasos monoinsaturados, principalmente ácido oleico, así como por una proporción alta de ácidos grasos poliinsaturados, destacando el ácido linoleico (omega-6) y el ácido α -linolénico (omega-3). En cuanto a los ácidos grasos saturados, los más representativos fueron el ácido palmítico y el ácido esteárico. En relación con los micronutrientes, el aceite de canola presentó niveles relevantes de vitaminas liposolubles, especialmente vitamina E y vitamina K. No se detectó la presencia de colesterol, minerales ni otras vitaminas en cantidades significativas.



Cuadro 2. Composición nutricional del aceite de canola (100 g), extraída mediante un *script* automatizado en Python a través de la API FDC del USDA.

Table 2. Nutritional composition of canola oil (100 g) extracted using an automated Python *script* through the USDA FDC API.

Nutriente	Cantidad	Unidad
Lípidos totales (grasas)	100	g
Energía	900	kcal
Vitamina E (alfa-tocoferol)	17.3	mg
Colina total	0.2	mg
Vitamina K (filloquinona)	71.3	µg
AGS totales	6.61	g
AGS 14:0	0.047	g
AGS 16:0	3.86	g
AGS 18:0	1.66	g
AGMI 18:1	60.3	g
AGPI 18:2	17.8	g
AGPI 18:3	7.45	g
AGMI 16:1	0.205	g
AGMI 20:1	1.89	g
AGMI totales	62.6	g
AGPI totales	25.3	g

Nota: Datos obtenidos de FoodData Central del Departamento de Agricultura de EE. UU. (USDA, 2025). Abreviaciones: AGS, Ácidos grasos saturados; AGMI, Ácidos grasos monoinsaturados; AGPI, Ácidos grasos poliinsaturados; 4:0, Ácido butírico; 6:0, Ácido caproico; 8:0, Ácido caprílico; 10:0, Ácido cáprico; 12:0, Ácido láurico; 14:0, Ácido mirístico; 16:0, Ácido palmítico; 18:0, Ácido esteárico; 18:1, Ácido oleico; 18:2, Ácido linoleico; 18:3, Ácido α -linolénico; 20:4, Ácido araquidónico; 20:5 n-3, Ácido eicosapentaenoico (EPA); 22:6 n-3, Ácido docosahexaenoico (DHA); 16:1, Ácido palmitoleico; 18:4, Ácido estearidónico; 20:1, Ácido eicosenoico; 22:1, Ácido erúcico; 22:5 n-3, Ácido docosapentaenoico (DPA). Las siguientes variables presentaron un valor de 0 en el registro consultado: Proteína, Carbohidratos, por diferencia, Alcohol etílico, Agua, Cafeína, Teobromina, Azúcares totales, Fibra dietética total, minerales (Ca, Fe, Mg, P, K, Na, Zn, Cu, Se), Retinol, Vitamina A (RAE), Caroteno (beta), Caroteno (alfa), Vitamina D (D2 + D3), Criptoxantina (beta), Licopeno, Luteína + zeaxantina, Vitamina C ácido ascórbico total, Tiamina, Riboflavina, Niacina, Vitamina B-6, Folato total, Vitamina B-12, Ácido fólico, Folato alimentario, Folato (DFE), Vitamina E añadida, Vitamina B-12 añadida, Colesterol, AGS 4:0, AGS 6:0, AGS 8:0, AGS 10:0, AGS 12:0, AGPI 20:4, AGPI 22:6 n-3 (DHA), AGPI 18:4, AGPI 20:5 n-3 (EPA), AGMI 22:1, AGPI 22:5 n-3 (DPA).

Estudios recientes han demostrado la utilidad de FoodData Central del USDA como herramienta para la estimación nutricional (Sebastian et al., 2025; Ma et al., 2021). La base de datos de Alimentos y Nutrientes para Estudios Dietéticos (FNDDS) ha sido empleada para calcular la composición de alimentos simples y complejos, así como el consumo poblacional (Sebastian et al., 2025). Asimismo, Ma et al. (2021) confirmaron que la Base de Datos Global de Productos Alimenticios de Marca (BFPD), combinada con técnicas de aprendizaje automático, permite estimar cuantitativamente los nutrientes en función de los ingredientes. En este contexto, el procedimiento aplicado en el presente

estudio confirmó tanto la funcionalidad de la API como la eficiencia del *script* en Python para la recuperación estructurada de datos nutricionales, evidenciando su aplicabilidad en investigaciones agroalimentarias que requieren la extracción organizada de información a partir de fuentes oficiales.

```
import requests
import pandas as pd

# API Key del USDA
api_key = 'TUAPIKEY'

# Lista de aceites
aceites = [
    'Canola oil'
]

# Crear un escritor de Excel
with pd.ExcelWriter('nutrientes_aceitecanola.xlsx', engine='openpyxl') as writer:
    for aceite in aceites:
        print(f"\n Buscando nutrientes de: {aceite}")
        search_url = f'https://api.nal.usda.gov/fdc/v1/foods/search?query={aceite}&api_key={api_key}'
        respuesta_busqueda = requests.get(search_url).json()

        try:
            fdc_id = respuesta_busqueda['foods'][0]['fdcId']
        except (KeyError, IndexError):
            print(f" No se encontró el alimento: {aceite}")
            continue

        detalles_url = f'https://api.nal.usda.gov/fdc/v1/food/{fdc_id}?api_key={api_key}'
        detalle = requests.get(detalles_url).json()

        nutrientes = []
        for nutriente in detalle.get('foodNutrients', []):
            nombre = nutriente.get('nutrient', {}).get('name')
            valor = nutriente.get('amount')
            unidad = nutriente.get('nutrient', {}).get('unitName')

            if nombre and valor is not None:
                nutrientes.append({
                    'Nutriente': nombre,
                    'Cantidad': valor,
                    'Unidad': unidad
                })

        # Crear DataFrame y guardar en una hoja con el nombre del aceite
        df = pd.DataFrame(nutrientes)
        hoja = aceite.replace(' oil', '').replace(' ', '_') # Nombre corto para la hoja
        df.to_excel(writer, sheet_name=hoja[:31], index=False)
        print(f"Guardado en hoja: {hoja[:31]}")

print("\n Todos los nutrientes han sido exportados a 'nutrientes_aceitecanola.xlsx'")
```

Figura 5. *Script* en Python para la consulta automatizada del perfil nutricional del aceite de canola mediante la API FoodData Central (FDC) del Departamento de Agricultura de los Estados Unidos (USDA).

Figure 5. Python *script* for the automated retrieval of the nutritional profile of canola oil using the FoodData Central (FDC) API from the United States Department of Agriculture (USDA).

Los hallazgos derivados de los casos de estudio sobre el perfil estadístico de la producción de canola y el perfil nutricional de su aceite subrayan la eficacia del uso de API públicas para crear bases de datos organizadas y estructuradas, capaces de proporcionar información pertinente y alineada con los objetivos de investigación. Estos resultados enfatizan la importancia de emplear variables específicas que sean coherentes con dichos objetivos, así como con las terminologías utilizadas en los informes y bases de datos del USDA. La extracción precisa de datos mediante claves API (API keys) depende intrínsecamente de esta coherencia terminológica. La definición adecuada de estas variables asegura la fiabilidad y aplicabilidad de los resultados, lo que refuerza el valor de las API como herramientas estratégicas en la investigación agroalimentaria.

Aunado a estos hallazgos, la literatura señala una tendencia creciente en la adopción de nuevas tecnologías —como el Internet de las Cosas (IoT), la inteligencia artificial y el software especializado— por parte de los agricultores. Encuestas recientes indican que el 77 % de los agricultores en Nebraska, EE. UU., utilizan computadoras, y el 85 % emplean teléfonos inteligentes, superando los promedios nacionales en la adopción tecnológica (Technology Use Growing on Nebraska Farms and Ranches, 2023). Esta integración tecnológica facilita el acceso en línea a los servicios del USDA, con 25% de los agricultores de Nebraska utilizando Internet para los informes del USDA (Technology use growing on Nebraska farms and ranches, 2023). Adicionalmente, el NASS publica más de 300 informes anualmente, proporcionando datos vitales para la toma de decisiones a nivel local, lo cual es crucial para los agricultores y las agroempresas (Iwig, 1995). La combinación de datos accesibles y tecnologías avanzadas respaldan significativamente las prácticas agrícolas en los Estados Unidos (Gibson, 2012).

Dentro del paradigma mexicano, la aplicación de la API del USDA junto con Python surge como un instrumento influyente para el análisis de datos en la investigación agroalimentaria. Potter (2019) subraya que esta integración aprovecha una amplia gama de datos agrícolas, que abarcan las estadísticas agrícolas y económicas, que son cruciales para la toma de decisiones informadas en el sector agroalimentario.

Las bibliotecas de Python, como *pandas* y *scikit-learn*, facilitan la manipulación y el análisis eficaces de conjuntos de datos extensos, lo que las convierte en herramientas ejemplares para los investigadores, tal como explica Lal (s.f.). Esta metodología ha

demostrado su eficacia en las investigaciones sobre la agricultura inteligente y el análisis de la biodiversidad agrícola en México, donde se empleó Python para la extracción y el análisis de datos (Rosales-Soto y Arechavala-Vargas, 2020; Torres et al., 2022). Además, la capacidad de analizar los datos sobre precios de productos básicos agrícolas utilizando Python, como indican Cedillo-Jiménez y Paredes-García (2018), aumenta la comprensión de la dinámica del mercado y fomenta la formulación de estrategias competitivas dentro del sector agroalimentario.

No obstante, Sandoval-Almazán (2019) señala que, si bien México ha avanzado en iniciativas hacia el gobierno abierto y transparencia, la implementación de prácticas de datos abiertos sigue siendo limitada. Las deficiencias en regulaciones y procesos impiden una verdadera apertura gubernamental y restringen la participación ciudadana a través de los portales estatales. En este contexto, el uso de API resulta especialmente útil en diversos escenarios de investigación. Combinado con herramientas como el lenguaje de programación Python y sus librerías especializadas, es posible extraer información masiva, ordenada y relevante, lo que permite su aplicación informada en la resolución de problemas complejos dentro del ámbito agroalimentario.

CONCLUSIÓN

La implementación de las APIs USDA QuickStats y FoodData Central, mediante un *script* en Python, permitió automatizar eficazmente la obtención y organización de datos sobre la producción de canola y el perfil nutricional de su aceite. Este proceso validó la funcionalidad de los endpoints y demostró el potencial de estas herramientas para su aplicación en investigación agroalimentaria. El enfoque desarrollado ofrece una alternativa sistematizada, reproducible y eficiente para diversos actores del sector, desde investigadores, agricultores, exportadores, gobiernos, formuladores de políticas, nutricionistas y tecnólogos de alimentos. Sin embargo, la estructura y nomenclatura de los datos pueden requerir ajustes según los objetivos específicos de cada estudio. Se recomienda ampliar esta estrategia integrando otras fuentes de datos y evaluando su utilidad en distintos contextos agroalimentarios.



Agradecimientos

Al programa de Doctorado en Innovación Agroalimentaria Sustentable (DCIAS) del Colegio de Postgraduados, Campus Córdoba, así como a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI).

LITERATURA CITADA

- Bastian, G. E., & Bienias, B. A. (2023). A novel method for estimating state-grown edible portion fruit and vegetable servings using agricultural census data. *Journal of Nutrition Education and Behavior*, 55(7, Suppl.), 1. <https://doi.org/10.1016/j.jneb.2023.05.005>
- Cedillo-Jiménez, C. A., & Paredes-García, W. J. (2019). Price data for Mexican agricultural products: case study Queretaro state (Mx). <https://doi.org/10.6084/m9.figshare.10009970>
- Elbasi, E., Mostafa, N., Zaki, C., AlArnaout, Z., Topcu, A. E., & Saker, L. (2024). Optimizing agricultural data analysis techniques through AI-powered decision-making processes. *Applied Sciences*, 14(17), 8018. <https://doi.org/10.3390/app14178018>
- Gibson, N. P. (2012). *United States Department of Agriculture (USDA)—A treasure trove of information: National Agricultural Statistics Service (NASS)*. University of Arkansas Libraries. <http://uark.libguides.com/c.php?g=79103&p=505124>
- Iwig, W. C. (1996). The National Agricultural Statistics Service county estimates program. In *Agricultural statistics: Modernization and innovation* (pp. 129–144). Springer. https://doi.org/10.1007/978-1-4612-0721-4_7
- Lal, A. S. (n.d.). *¿Por qué Python para el análisis de datos?* [Why Python for data analysis?].
- Ma, P., Li, A., Yu, N., Li, Y., Bahadur, R., Wang, Q., & Ahuja, J. K. (2021). Application of machine learning for estimating label nutrients using USDA Global Branded Food Products Database (BFPD). *Journal of Food Composition and Analysis*, 100, 103857. <https://doi.org/10.1016/j.jfca.2021.103857>
- Potter, N. (2019). *rmassqs: Acceso a datos agrícolas del USDA mediante API* [rmassqs: Access to USDA agricultural data via API]. <https://doi.org/10.59350/py9hm-2sv48>
- Prasai, R., Schwertner, T. W., Mainali, K., Mathewson, H., Kafley, H., Thapa, S., Adhikari, D., Medley, P., & Drake, J. (2021). Application of Google Earth Engine Python API and NAIP imagery for land use and land cover classification: A case study in Florida, USA. *Ecological Informatics*, 66, 101474. <https://doi.org/10.1016/j.ecoinf.2021.101474>
- Rosales-Soto, A., & Arechavala-Vargas, R. (2020). Agricultura inteligente en México: Analítica de datos como herramienta de competitividad. <https://doi.org/10.29105/vtga6.2-619>

- Roscher, R., Roth, L., Stachniss, C., & Walter, A. (2023). Agricultura digital centrada en los datos: Una perspectiva [Data-centric digital agriculture: A perspective]. *arXiv*. <https://doi.org/10.48550/arXiv.2312.03437>
- Sandoval-Almazan, R. (2019). Gobierno Abierto y Transparencia en México: estudio longitudinal 2015-2018. *Innovar-Revista De Ciencias Administrativas Y Sociales*, 29(74), 115–131. <https://doi.org/10.15446/INNOVAR.V29N74.82097>
- Sarkar, T. (2020). Harvest and analyze agricultural data with the USDA NASS API, Python, and Tableau. Medium. <https://towardsdatascience.com/harvest-and-analyze-agricultural-data-with-the-usda-nass-api-python-and-tableau-a6af374b8138/>
- Sebastian, R. S., Goldman, J. D., & Moshfegh, A. J. (2025). Leveraging USDA databases to estimate population intakes of foods that are commonly consumed individually and in multi-component foods: The case of cheese. *Journal of Food Composition and Analysis*, 142, 107434. <https://doi.org/10.1016/j.jfca.2025.107434>
- Siné, M., Haezebrouck, T.-P., & Emonet, E. (2015). API-AGRO: Una plataforma de datos abiertos y API abierta para promover la interoperabilidad en servicios agrícolas y aplicaciones web [API-AGRO: An open data platform and open API to promote interoperability in agricultural services and web applications]. *Journal of Agricultural Informatics*, 6(4), 56–64. <https://www.cabdirect.org/cabdirect/abstract/20163000246>
- Technology use growing on Nebraska farms and ranches. (2023). University of Nebraska-Lincoln. <https://doi.org/10.32873/unl.dc.cap011>
- U.S. Department of Agriculture. (n.d.-a). FoodData Central API. FoodData Central. <https://fdc.nal.usda.gov/api-guide>
- U.S. Department of Agriculture. (n.d.-b). *Quick Stats API*. National Agricultural Statistics Service. <https://quickstats.nass.usda.gov/api>
- U.S. Department of Agriculture. (n.d.-c). *ARMS Data API*. Economic Research Service. <https://www.ers.usda.gov/developer/data-apis/arms-data-api>
- U.S. Department of Agriculture. (2024). *FoodData Central: About us*. <https://fdc.nal.usda.gov/about-us.html>
- USDA (United States Department of Agriculture). (2025). *Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center. FoodData Central* [Internet]. [Cited 2025 May 4]. <https://fdc.nal.usda.gov/>
- Yadav, R., Seth, A., & Dembla, N. (2024). Optimizing crop yield prediction: Data-driven analysis and machine learning modeling using USDA datasets. *Current Agriculture Research Journal*, 12(1), 272–285. <https://doi.org/10.12944/CARJ.12.1.22>

