

# FREE THE DATA API

Para mi proyecto final he desarrollado una API a la que he llamado free the data. Vamos a comenzar viendo la aplicación sobre el front de streamlit y luego veremos un PPT con algunos detalles técnicos sobre su funcionamiento.

Antes de comenzar, aunque todos los que estáis aquí ya sois analistas de datos expertos y conocéis todos los conceptos, voy a refrescar lo que es una API.

Os incluyo aquí la definición de chatGPT que ha sido casi como un compañero más de clase.



Good morning! An API, or Application Programming Interface, is a set of rules and protocols that enables different software programs to communicate with each other. It facilitates interaction and data exchange between applications, which is essential for software development and task automation.

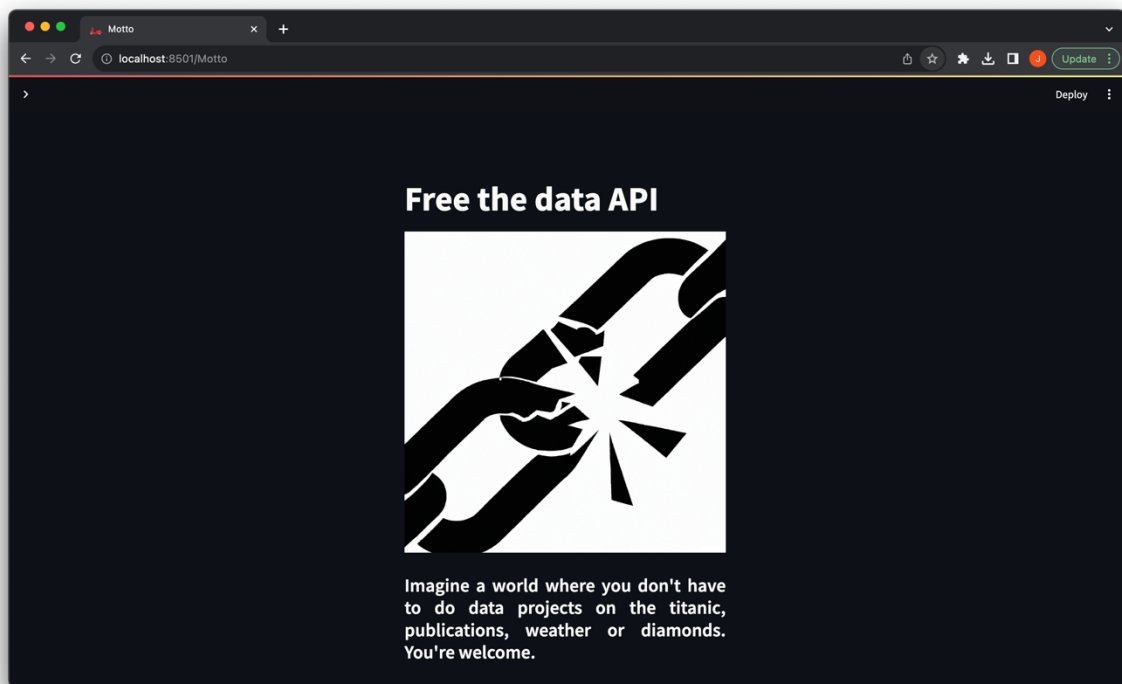
¿Qué es un API?

Una API o Application Programming Interface (interfaz de programación de aplicaciones) es un conjunto de reglas y protocolos que permite que diferentes programas se comuniquen entre sí. Facilita la interacción y el intercambio de datos entre aplicaciones, lo que es esencial para el desarrollo de software y la automatización de tareas.

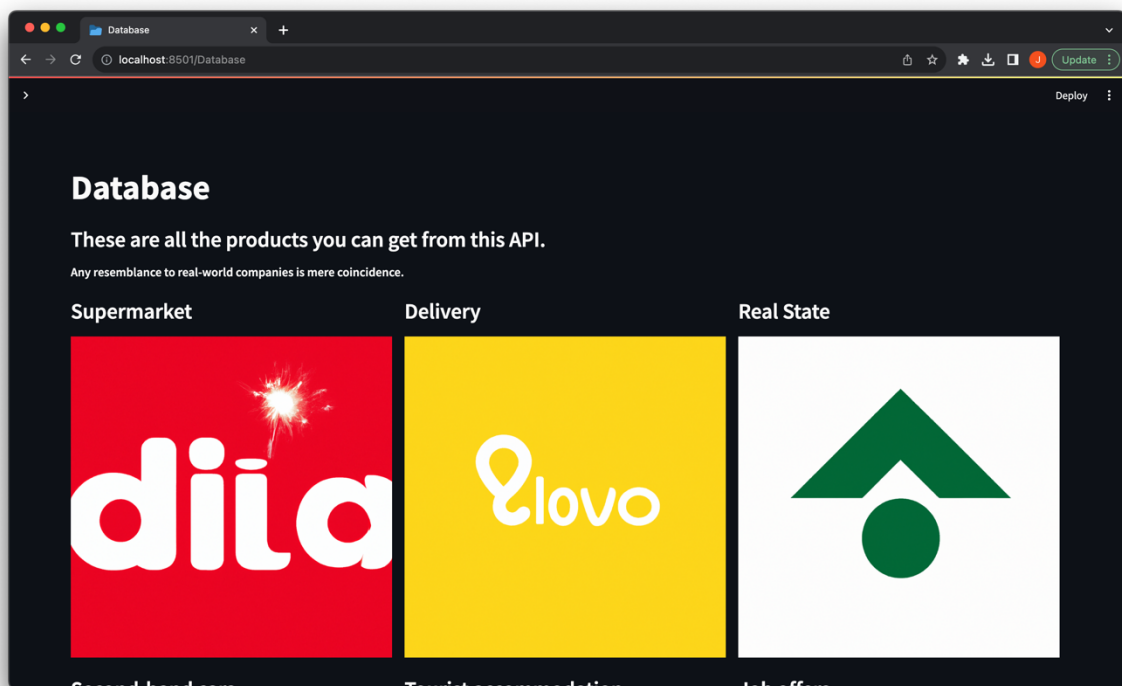
¿Por qué he llamado así a la API/aplicación y cuál es la motivación que persigue mi proyecto?

Durante este bootcamp, irónicamente, uno de los mayores retos a los que me he tenido que enfrentar es encontrar datos que me motivasen para elaborar un proyecto sobre ellos. **Y como bien sabe Jeancha, en cuanto algo no me motiva, no lo hago.** Obviamente no digo que no existan bases de datos gratuitas en Internet, pero sí que creo que escasean y eso hace que los proyectos se repitan y sean siempre los mismos. No hay más que meterse en Kaggle para ver el mismo proyecto repetido mil veces. Si te preguntas porqué todas o la mayoría de las bases de datos gratuitas son siempre las mismas, la respuesta es sencilla: los datos tienen un precio y las empresas no los regalan.

Precisamente de aquí es de donde sale la motivación de mi proyecto y para aportar mi granito de arena a la comunidad, he creado esta API que permite a los usuarios descargar datos que scrapeo de Internet. Esto queda reflejado en el motto (lema) de mi aplicación: ***Imagine a world where you don't have to do data projects on the titanic, publications, weather or diamonds. You're welcome.***

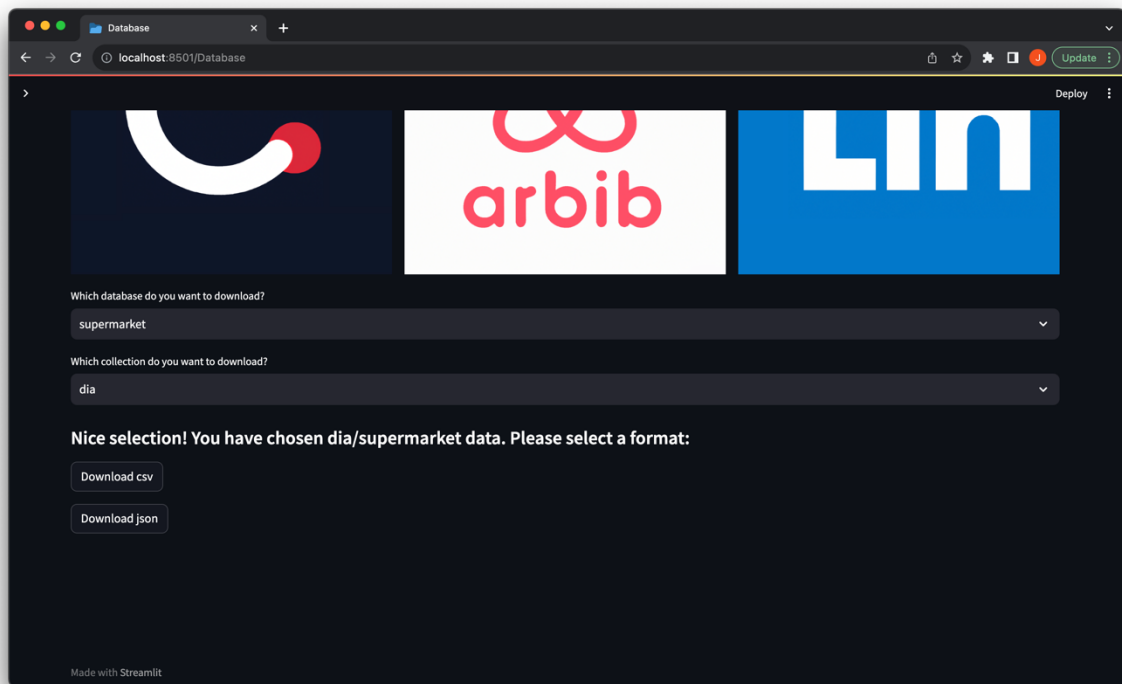


Si vemos la aplicación, en primer lugar, podéis ver en esta pantalla todos los productos que están disponibles para descargar. Cualquier parecido con empresas del mundo real es pura coincidencia.

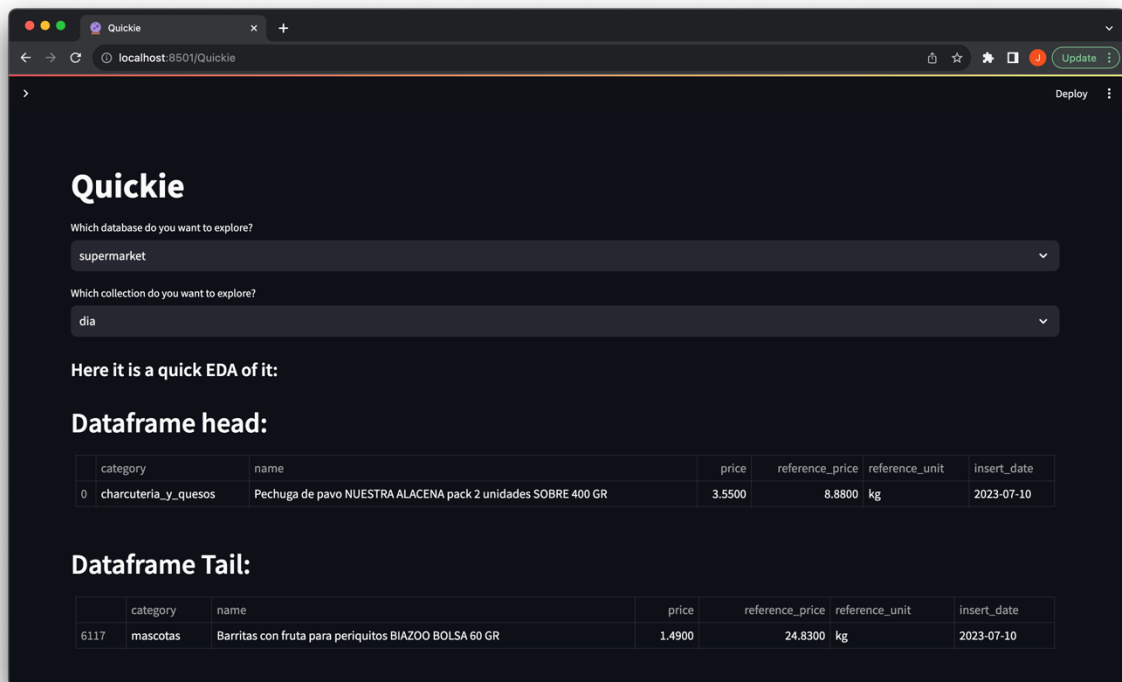


Actualmente solo la primera fila de productos está disponible, pero he añadido una segunda, de productos en los que todavía estoy trabajando como ejemplo de lo que será en un futuro.

El usuario tiene que seleccionar la base de datos que le interesa y dentro de la base de datos la colección que desea descargar. Se pueden descargar los datos tanto en formato csv como en formato json.

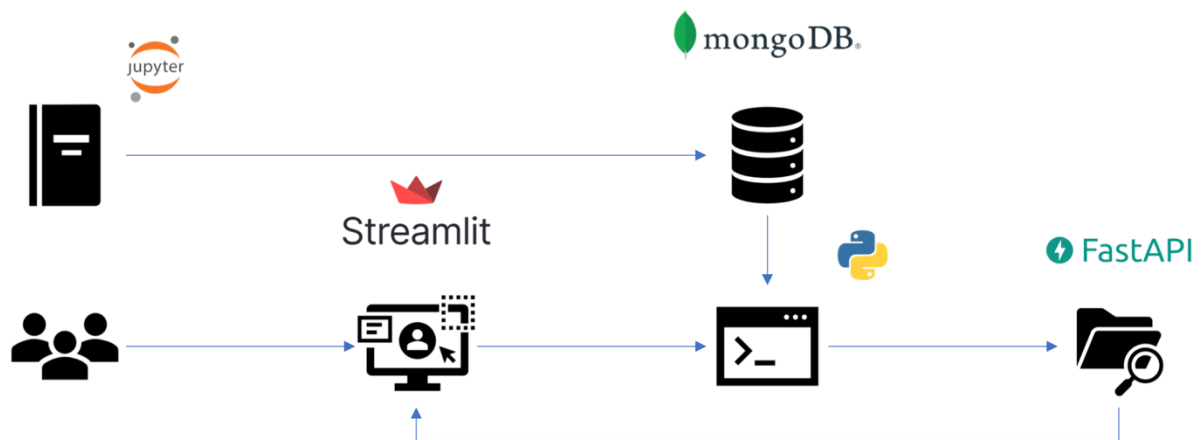


Además, de permitirte descargar datos, he incluido otra funcionalidad en la aplicación llamada quickie, que te hace un EDA rápido sobre los datos que se pueden descargar.



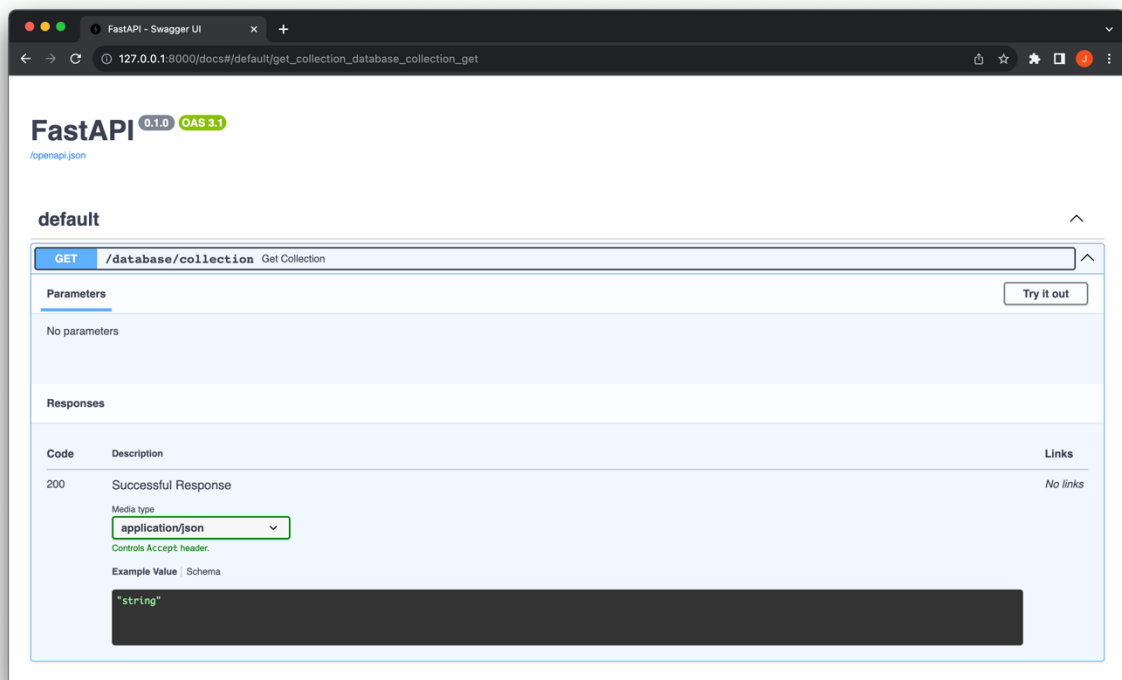
¿Cómo funciona la API?

Vamos a ver ahora el PPT, en el que he generado un diagrama de flujo sencillo que explica el funcionamiento de la aplicación.



**database/collection** → **csv/json**

Toda la estructura de la aplicación sigue el formato de base de datos de MongoDB. Los datos se almacenan en colecciones que a su vez componen bases de datos. Es por ello, que la API crea endpoints dinámicos en función de la solicitud del usuario como se muestra en la siguiente foto.



#### Next steps

- Deployear la aplicación en Internet
- Desarrollar un sistema de actualización automático de los registros
- Aumentar el número de productos