

# Dataset: Recetas de cocina

## Descripción

El conjunto de datos en cuestión reúne un listado de recetas de cocina, clasificadas en función de su categoría y proporciona tanto el link a la receta completa como información que las caracteriza: la valoración media dada por los usuarios, la dificultad y el tiempo de realización, el número de comensales para los que está pensada, entre otros.

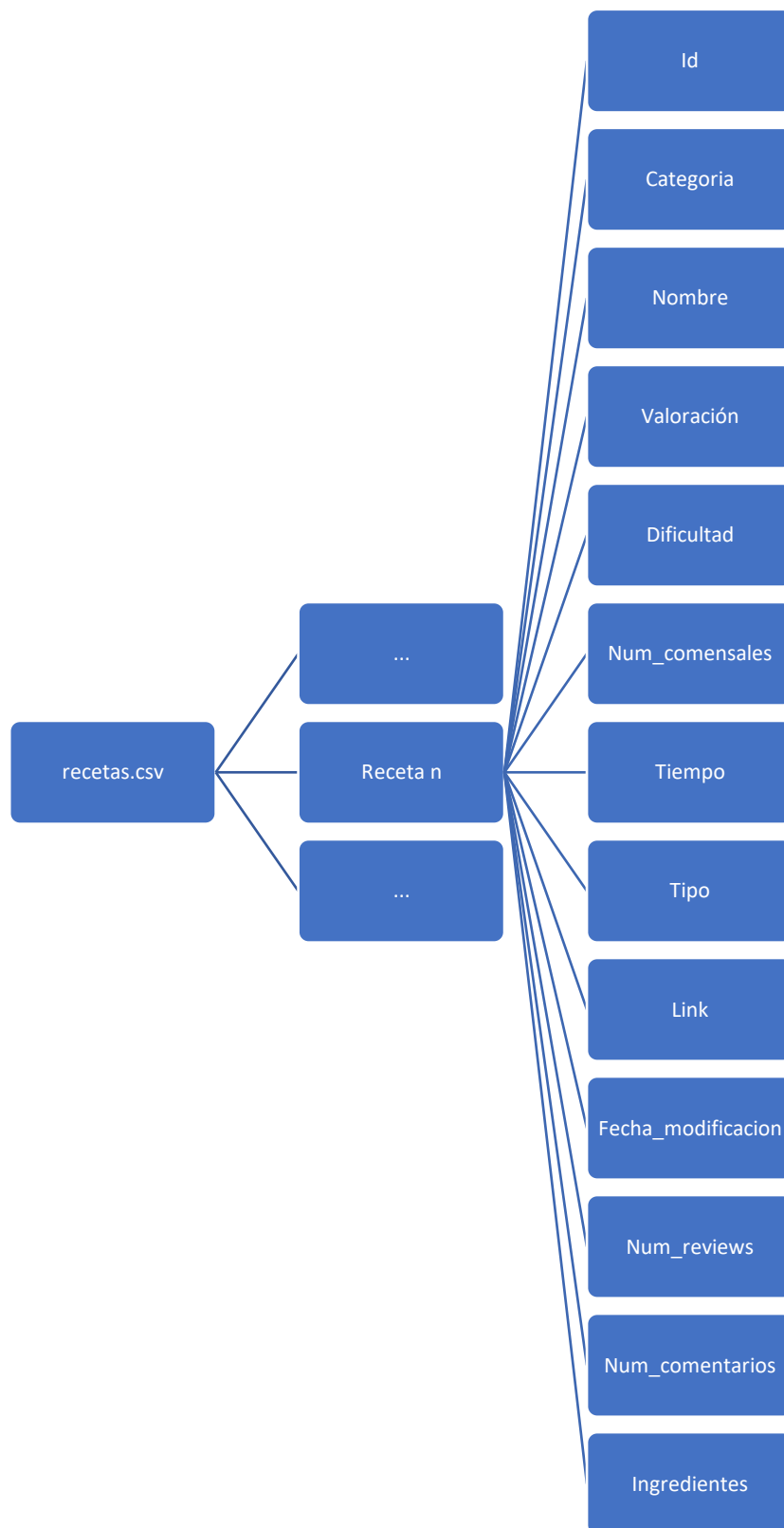
## Contexto

La información sobre recetas de cocina se ha recolectado en tiempos del COVID-19, unos tiempos difíciles para el mundo entero y en los que ha sido necesario el confinamiento de las personas en sus propias casas para evitar el contagio. En dichos tiempos, han surgido multitud de iniciativas para tratar de mantener a la gente entretenida, y una de ellas ha sido la cocina.

## Imagen identificativa y representación gráfica



*Ilustración 1: Mesa llena de platos de comida en visión cenital.*



*Ilustración 2: Diagrama de la estructura del conjunto de datos.*

## Contenido

El conjunto de datos incluye los siguientes campos:

- **Id:** Identificador único de la receta
- **Categoría:** La categoría a la que pertenece la receta.
- **Nombre:** El nombre de la receta.
- **Valoración:** La media de la valoración dada por los usuarios.
- **Dificultad:** La dificultad de realización de la receta.
- **Num\_comensales:** El número de comensales para los que está pensada la receta.
- **Tiempo:** El tiempo de realización de la receta.
- **Tipo:** El tipo de receta.
- **Link:** El link a la receta completa.
- **Fecha\_modificacion:** La fecha en la que la receta ha sido creada.
- **Num\_reviews:** Cantidad de valoraciones recibidas para esta receta.
- **Num\_comentarios:** Interacciones en forma de comentarios para esta receta.
- **Ingredientes:** Listado de ingredientes necesarios para seguir la receta.

Dichos datos provienen de la página web <https://www.recetasgratis.net/>, cuyos autores llevan recopilando recetas desde el año 2008. Para su extracción, se ha hecho uso del lenguaje de programación Python y de técnicas de web scraping para obtener la información alojada en las páginas HTML.

## Agradecimientos

Los datos han sido recolectados desde la página web <https://www.recetasgratis.net/>, cuyo equipo está formado por Eva López, Olga Gallego (equipo de redacción) y varios colaboradores entre los que se encuentran cocineros profesionales, estudiantes de cocina y amateurs: Laura Durán, Alix Hernández, María Angélica Manrique, Vanessa Romero, Ana Victoria Vázquez, Alicia Terán, Neus Jordi, Cris GRX, Miriam Hernández Medina, Eleonor Fischer, por nombrar unos pocos.

## Inspiración

El presente conjunto de datos podría utilizarse en diferentes aplicaciones. Por un lado, mediante técnicas de minería de datos se podrían analizar las tendencias en el mundo de la cocina a partir de la fecha de creación de cada receta y de la interacción de los usuarios (su valoración). Por otro lado, se podrían analizar los ingredientes más utilizados o los ingredientes que están relacionados y los supermercados podrían utilizar dicha información para situar los productos de manera estratégica o crear promociones específicas con la finalidad de incrementar sus ventas. Otro uso interesante sería el de las empresas que producen comidas precocinadas, que podrían buscar inspiración también en dichas tendencias para lanzar nuevos productos al mercado.

## Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **CC BY-SA 4.0** (<https://creativecommons.org/licenses/by-sa/4.0/>). Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de sus cláusulas:

- *Se permite la copia y redistribución del material en cualquier medio o formato. Además, se permite adaptar, transformar y construir sobre el material para cualquier fin, incluso su uso comercial.* De esta forma la posibilidad de que el conjunto de datos se utilice se incrementan enormemente.
- *Se debe proporcionar el nombre del creador del conjunto de datos, un link a la licencia e indicar si se produjeron cambios.* De esta manera se reconoce el trabajo ajeno.
- *Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma.* De esta manera el trabajo del autor original seguirá distribuyéndose bajo los mismos términos que planteó inicialmente.



## Código fuente y dataset

Tanto el código fuente escrito para la extracción de los datos como el dataset generado pueden ser accedidos a través del siguiente enlace al repositorio Github:

<https://github.com/DaniCorretja/web-scraping>

Además, el dataset también se puede encontrar en Zenodo mediante el siguiente enlace:

<https://zenodo.org/record/3749328>

## Recursos

1. Lawson, R. (2015). *Web Scraping with Python*. Packt Publishing Ltd. Chapter 2. Scraping the Data.
2. Mitchel, R. (2015). *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
3. Visual Studio Live Share (<https://visualstudio.microsoft.com/es/services/live-share/>)

## Autores

Contribuciones	Firma
<b>Investigación previa</b>	DCA, XGG
<b>Redacción de las respuestas</b>	DCA, XGG
<b>Desarrollo del código</b>	DCA, XGG