M2.851

-TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS-

PRA 1

Web scraping

Fecha límite de entrega: 12 abril 2021

Autores:

Olga Garcés Ciemerozum
Carlos Acosta Quintas

Máster Universitario en Ciencia de Datos
Universitat Oberta de Catalunya

INDICE DE CONTENIDOS

Introducción

El presente informe forma parte de la primera práctica de la asignatura M2.851 - Tipología y ciclo de vida de los datos del Máster Universitario en Ciencia de Datos impartido por la Universitat Oberta de Catalunya.

En esta práctica se realizarán técnicas de Web scraping aplicadas a una Web en concreto y se analizarán dichos datos para extraer información relevante y útil.

A su vez, se entregará, junto con la presente memoria, una serie de archivos con el código necesario para la realización de dicho web scraping y varios juegos de datos reales y actualizados con el que el usuario podrá realizar diferentes estudios analíticos a posteriori.

1. Contexto

El presente informe forma parte de la primera práctica de la asignatura M2.851 - Tipología y ciclo de vida de los datos del Máster Universitario en Ciencia de Datos impartido por la Universitat Oberta de Catalunya.

1.1 Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

Contexto en la recolección de la información:

La página web seleccionada para realizar las técnicas de Web scraping es: https://www.bonarea.com/

BonÀrea es empresa con una amplia experiencia en el sector agroalimentario, donde su principal negocio se desarrolla en actividades ganaderas, industriales y comerciales con el fin de poder llegar al consumidor sin intermediarios.

Una de las divisiones de la empresa, bonÀrea Energía, dispone de más de 55 gasolineras que venden más de 430 millones de litros al año con un ratio calidad/precio que hace ahorrar más de 40 millones de euros al año a sus clientes.

Las gasolineras bonÀrea son conocidas por su precio económico debido a sus reducidos márgenes de beneficio y al gran volumen de carburantes vendido, y además debido al uso de economías de escala entre sus diferentes líneas de negocio.

Para ésta práctica, se ha decidido hacer un estudio geográfico y temporal de los precios de las gasolineras de bonÀrea Energía, para determinar, no solamente las variaciones diarias de los precios de los diferentes productos, sino también realizar un registro de las economías de escala asociadas a cada gasolinera (servicios complementarios aportados por cada gasolinera).

De igual forma, y debido a que la página web muestra información descriptiva y relevante de todos los establecimientos que BonÀrea tiene repartido a lo largo del territorio español (supermercados, tiendas, bufets, gasolineras, "Box online" de recogida, Centros de Agricultura y centros Cash para venta al mayor), se ha decidido recolectar y agrupar en un dataset dicha información.

Información extraída del sitio web:

El dominio www.bonarea.com aporta una web dinámica que referencia a todas las divisiones de negocio del grupo, implementando diferentes esquemas de datos para que la experiencia del usuario final a nivel de visualización de la información sea clara y concisa.

Figura 1: Muestra de los diferentes establecimientos de BonÀrea (Fuente: https://www.bonarea.com/)

En el apartado de los establecimientos, para todos ellos y en particular en el de las gasolineras, la página web aporta diferentes datos sobre ellas, tanto a nivel geográfico mediante un mapa interactivo para que el usuario sepa en todo momento dónde está la gasolinera más cercana, como a nivel económico (se muestran los precios de los diferentes carburantes disponibles en cada una de ellas).

La página permite al usuario seleccionar uno o varios tipos de establecimientos y muestra la ubicación sobre el mapa de todos establecimientos de los tipos seleccionados.

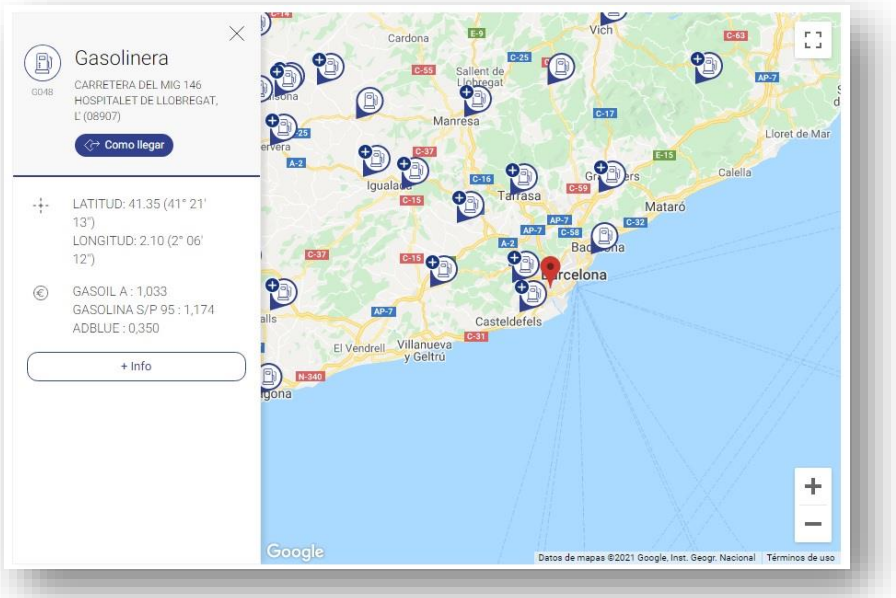


Figura 2: Muestra del mapa interactivo de las gasolineras de BonÀrea (Fuente: https://www.bonarea.com/)

Además, haciendo clic en el icono "**+ info**", cada establecimiento dispone de una página web anidada para que el usuario pueda tener información adicional como el horario de apertura, dirección, teléfono, etc. de cada uno de ellos.

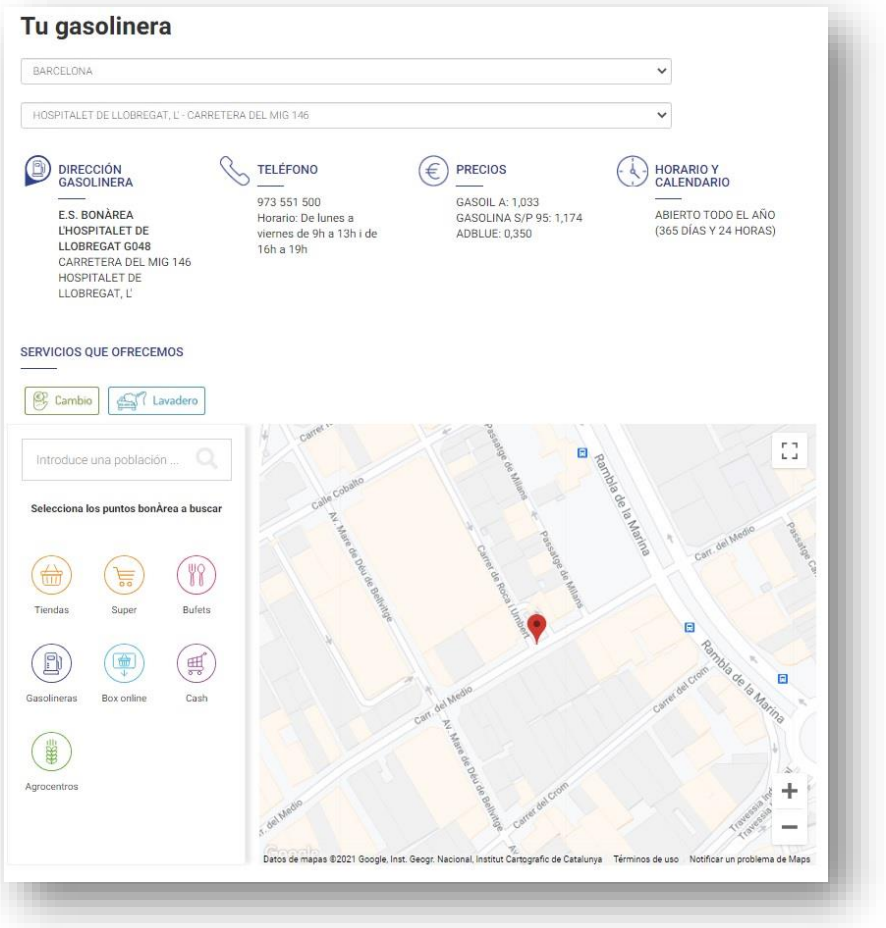


Figura 3: Muestra del panel descriptivo en página anidada de BonÀrea (Fuente: https://www.bonarea.com/)

En definitiva, el sitio web proporciona dicha información con solamente un motivo: conseguir el compromiso y la fidelización del cliente a una página web práctica y bien diseñada, que ayuda a obtener una información "a la carte" útil que puede ser consumida en cualquier aparato móvil con conexión a internet.

La información debe estar disponible en la misma página web para conseguir una máxima transparencia a sus usuarios.

Hay que tener en cuenta **dos puntos claves** relacionados con la interacción del usuario:

-Durante la interacción del usuario con la web, **los datos se generan de forma dinámica**: la web envía una solicitud jquery de información y devuelve un conjunto de puntos en el mapa que cumplen con el criterio impuesto por el usuario.

- Al hacer clic en los distintos puntos que aparecen en el mapa, la página realiza otra solicitud jquery para una instancia de establecimiento específica y visualiza un recuadro con su información básica.

2. Título del dataset

2.1 Definir un título para el dataset. Elegir un título que sea descriptivo.

El proyecto puede generar 3 archivos con formato csv que formarán parte del dataset:

1. **Archivo csv "bonarea_gasolineras_prices.csv":** csv actualizable con histórico de precios diarios de los diferentes carburantes disponibles en las gasolineras.
2. **Archivo csv "bonarea_gasolineras_data_and_prices.csv":** csv actualizable con los datos informativos de las gasolineras e histórico de precios diarios de los diferentes carburantes disponibles.
3. **Archivo csv "bonarea_establecimientos.csv":** csv con la información descriptiva de todos los establecimientos de BonÀrea.

Por tanto, el título propuesto para el conjunto de archivos que conforman el dataset sería:

"Establecimientos bonÀrea: Información descriptiva de los supermercados, tiendas, bufets, gasolineras, Box online, Centros de Agricultura y centros Cash y precios históricos de los carburantes con servicios asociados en sus distintas gasolineras."

3. Descripción del dataset

3.1 Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

"Información descriptiva de las gasolineras bonÀrea, servicios asociados y valores diarios del precio de los carburantes"

El conjunto de datos muestra el identificador de cada gasolinera con la fecha y día de la semana de la extracción de la información.
Además, se incluye información básica de cada gasolinera (url, calle, ciudad, código postal, localización geográfica, servicios que provee en el propio establecimiento y los precios diarios de cada producto (Gasoil A, Gasolina sin plomo 95 y 98 y ADBLUE).

Los atributos del dataset no tienen valores nulos excepto los servicios asociados y los productos de la gasolinera, donde pueden tomar valores nulos si tal servicio o producto no existe en esa determinada gasolinera.

"Establecimientos bonÀrea: Información descriptiva de los supermercados, tiendas, bufets, gasolineras, Box online, Centros de Agricultura y centros Cash."

El conjunto de datos muestra la misma información relevante que el anterior juego de datos para todos los tipos de establecimientos a excepción de los precios diarios por tipo de carburante.

Este juego de datos se ha creado a nivel informativo, para tener en un único dataset toda la información relevante y descriptiva de los establecimientos de BonÀrea.

4. Representación gráfica

4.1 Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.

Figura 4: Diagrama para visualizar el dataset (Fuente: elaboración propia)

5. Contenido

5.1 Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

El dataset contiene datos descriptivos de los distintos tipos de establecimientos BonÀrea, así como los precios diarios de los carburantes en distintas gasolineras de la red BonÀrea en el período de tiempo desde:

23 de marzo de 2021 hasta la fecha de compilación para la entrega de la práctica

El dataset está compuesto por los siguientes ficheros:

'bonarea_establecimientos.csv' - datos descriptivos de los distintos tipos de establecimientos BonÀrea (supermercados, tiendas, bufets, gasolineras, Box online, Centros de Agricultura y centros Cash)

'bonarea_gasolineras_prices.csv' - datos históricos del precio de los carburantes por gasolinera.

'bonarea_gasolineras_data_and_prices.csv' - datos históricos del precio de los carburantes por gasolinera con información detallada de cada gasolinera.

A continuación, se explican los campos que incluye los diferentes archivos csv que forman el dataset:

'bonarea_establecimientos.csv'

- **id** - número identificador de la gasolinera
- **type** - tipo de establecimiento de BonÀrea
- **url** - url de la gasolinera
- **street** - calle donde se ubica la gasolinera
- **city** - ciudad donde se ubica la gasolinera
- **postalCode** - código postal de la dirección de la gasolinera
- **raoSocial** - razón social del establecimiento
- **latitude** - latitud
- **longitude** - longitud
- **minutsLatitude** - latitud en minutos
- **minutsLongitude** - longitud en minutos
- **RENTADOR** - tiene túnel de lavado de coches (1 - sí)
- **CANVI** - tiene cambio para dinero, booleano (1 - sí)
- **SUPER** - tiene supermercado (1 - sí)
- **LAVABO** - tiene lavabo (1 - sí)
- **PARKING** - tiene parking (1 - sí)
- **VENDING** - tiene vending (1 - sí)

'bonarea_gasolineras_prices.csv'

- **id** - número identificador de la gasolinera.
- **Fecha** - fecha
- **Dia_Semana** - día de la semana
- **latitude** - latitud
- **longitude** - longitud
- **minutsLatitude** - latitud en minutos
- **minutsLongitude** - longitud en minutos
- **GASOIL A** - precio diario de este tipo de combustible
- **GASOLINA S/P 95** - precio diario de este tipo de combustible
- **GASOLINA S/P 98** - precio diario de este tipo de combustible
- **ADBLUE** - precio diario de este tipo de combustible

'bonarea_gasolineras_data_and_prices.csv'

- **id** - número identificador de la gasolinera
- **Fecha** - fecha
- **Dia_Semana** - día de la semana
- **type** - tipo de establecimiento de BonÀrea
- **url** - url de la gasolinera
- **street** - calle donde se ubica la gasolinera
- **city** - ciudad donde se ubica la gasolinera
- **postalCode** - código postal de la dirección de la gasolinera
- **raoSocial** - razón social del establecimiento
- **latitude** - latitud
- **longitude** - longitud
- **minutsLatitude** - latitud en minutos
- **minutsLongitude** - longitud en minutos
- **RENTADOR** - tiene túnel de lavado de coches (1 - sí)
- **CANVI** - tiene cambio para dinero, booleano (1 - sí)
- **SUPER** - tiene supermercado (1 - sí)
- **LAVABO** - tiene lavabo (1 - sí)
- **PARKING** - tiene parking (1 - sí)
- **VENDING** - tiene vending (1 - sí)
- **GASOIL A** - precio diario de este tipo de combustible
- **GASOLINA S/P 95** - precio diario de este tipo de combustible
- **GASOLINA S/P 98** - precio diario de este tipo de combustible
- **ADBLUE** - precio diario de este tipo de combustible

Procedimiento de colección de datos

Durante la interacción del usuario con la web, los datos se generan de forma dinámica:

La página permite al usuario seleccionar uno o varios tipos de establecimientos. Esta interacción del usuario con la web genera una solicitud jquery que devuelve todos los establecimientos del tipo seleccionado y los muestra sobre el mapa.

Al hacer clic en los distintos puntos que aparecen en el mapa, la página realiza otra solicitud jquery y devuelve los datos detallados para una instancia de establecimiento específica. La respuesta a la solicitud se hace visible en un recuadro con la información básica del establecimiento.

Pasos para realizar el webscraping:

En la primera interacción recopilamos los datos básicos de los establecimientos BonÀrea y guardamos los valores de Id del establecimiento.

En detalle (Firefox):

Inspeccionamos un elemento de la web seleccionamos la solapa “Network”. Aquí podemos encontrar la solicitud jquery jquery-3.3.1.json.

Una vez seleccionada la línea que contiene jquery jquery-3.3.1.json, podemos ver el request que enviamos y la respuesta que recibimos:

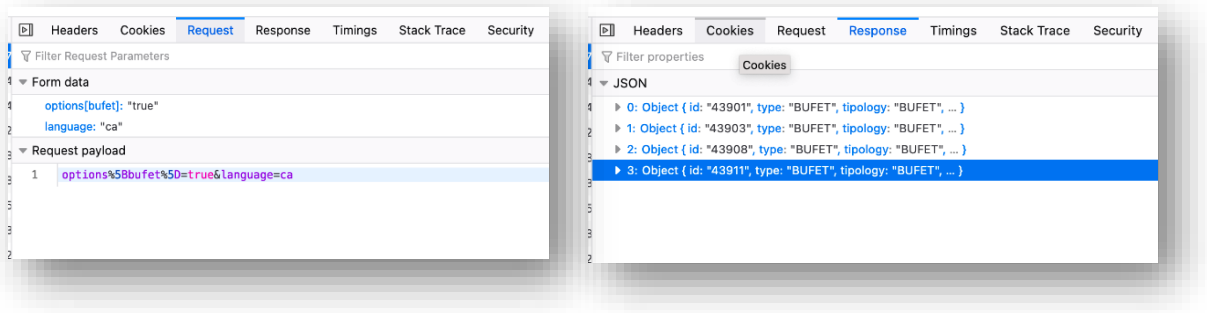


Figura 5 y 6: Request y respuesta mostrada en Firefox (Fuente:Firefox)

Usamos la aplicación **Postman** para generar los datos necesarios para el webscraping de los datos básicos de los establecimientos:

Hacemos clic derecho en jquery-3.3.1.json y seleccionamos “Copy as cURL”. En Postman seleccionamos File --> Import y importamos lo que hemos copiado como Raw text.

Para construir un request necesitamos la url, el string del request y el header para el request. Todos estos datos se encuentran en la solapa Header y la solapa Body de Postman, una vez hayamos importado la url.

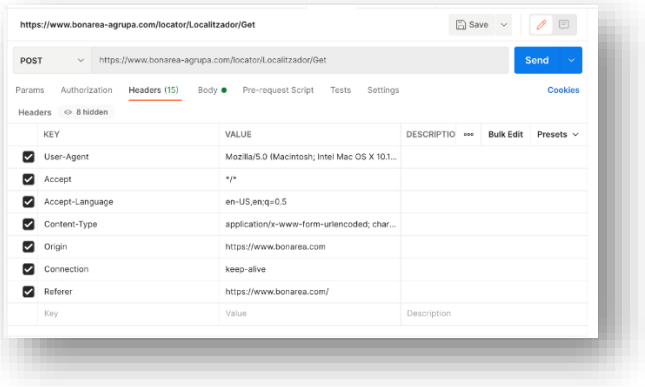
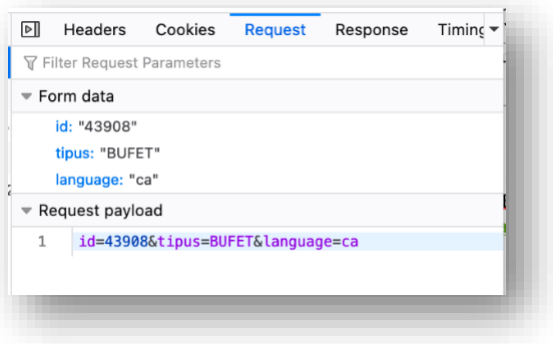


Figura 7: Captura de pantalla de la aplicación Postman (Fuente: https://www.postman.com/)

El request resultante se construye como se muestra a continuación, donde data es un string que resulta de descodificar el componente URI de options%5Bbufet%5D=true&language=ca (options[bufet]=true&language=ca)

```
headers = {'content-type': 'application/x-www-form-urlencoded; charset=UTF-8',
'origin': 'https://www.bonarea.com',
'Referer': 'https://www.bonarea.com/ca/default/locate'}
url = 'https://www.bonarea-agrupa.com/locator/Localitzador/Get'
response = requests.post(url, data=data, headers=headers)
```

En la segunda interacción con la web seleccionamos un establecimiento específico y volvemos a ver la request y la respuesta que ha generado esta acción:



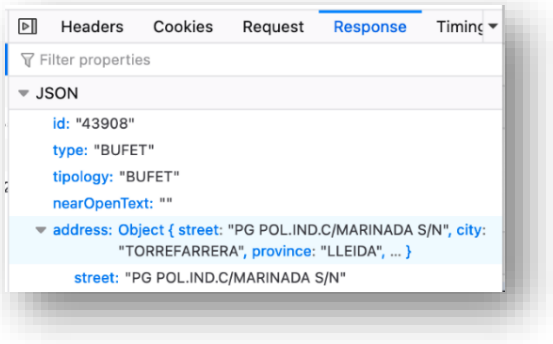


Figura 8 y 9: Request y respuesta mostrada en Firefox (Fuente:Firefox)

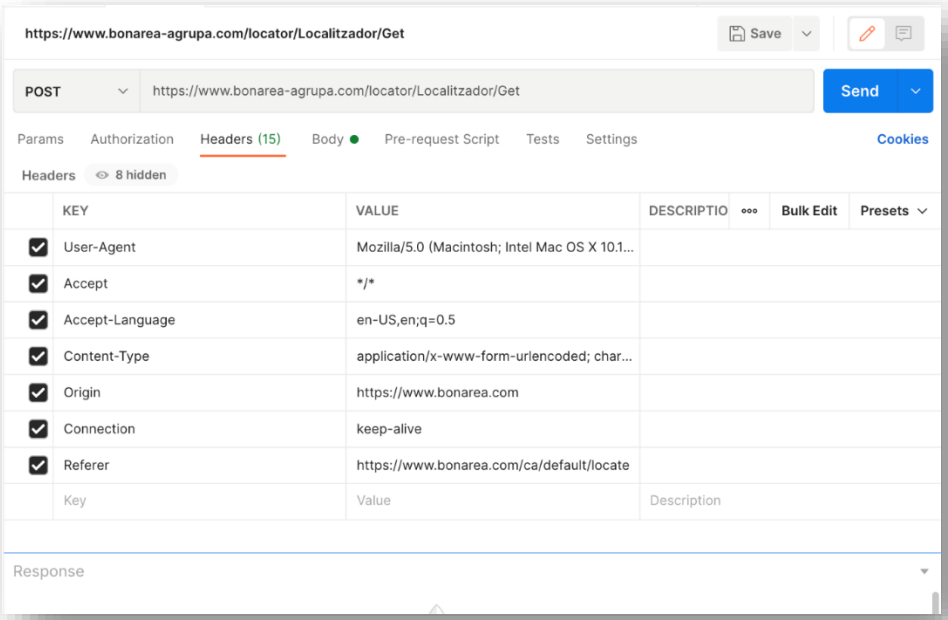
Repetimos el ciclo anterior con los datos nuevos: copiamos la solicitud json como cURL, importamos esta cURL como Raw data en postman, y conseguimos los headers y el string con los parámetros de la solicitud.

Estos pasos nos habrán proporcionado una lista de establecimientos BonÀrea con todos sus datos básicos.

Nota: tenemos la posibilidad de generar los IDs para un tipo de establecimiento específico o para una lista de establecimientos, por lo tanto, **proporcionamos máxima flexibilidad con respecto a la personalización del dataset resultante.**

6. Agradecimientos

Los datos han sido recolectados de la página web de BonÁrea. Para ello se han utilizado `requests` de Python y la aplicación `Postman` (https://www.postman.com/).



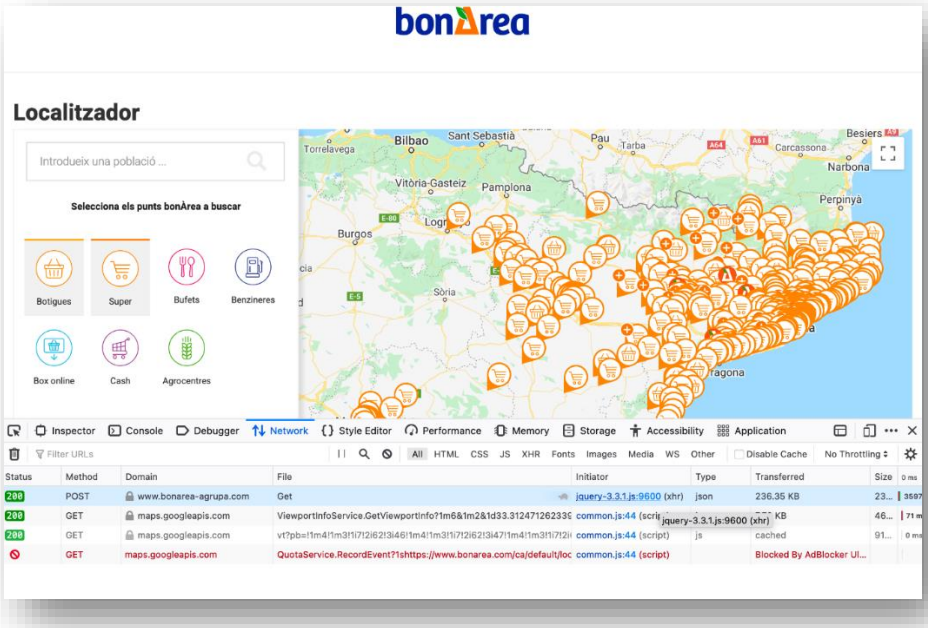


Figura 10 y 11: Captura de pantalla de la aplicación Postman (Fuente: https://www.postman.com/)

Cabe mencionar que www.bonarea.com no dispone de ningún tipo de archivo "robots.txt", hecho que ha facilitado el scraping de la información con total libertad por parte de los estudiantes.

Figura 12: Muestra de la no existencia del archivo robots.txt (Fuente: https://www.bonarea.com/)

6.1 Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.

El fundador y presidente de BonÀrea es **Don Jaume Alsina Calvet**, como tal, la página web y por tanto, el conjunto de datos mostrado en ella es de su propiedad.

Si desglosamos los datos según la estructura empresarial de BonÀrea, podemos argumentar que:

-Los datos descriptivos extraídos para los supermercados, tiendas, bufets, Box online, Centros de Agricultura y centros Cash pertenecen a las marcas de enseña (parte de la División de Alimentación); BonÀrea Agrocentro (Centros de Agricultura), BonÀrea Restaurante (Bufet), BonÀrea Cash & Carry (Box Online y Cash) y BonÀrea Alimenta(supermercados y tiendas).

-Los datos evolutivos extraídos para las gasolineras pertenencen a la División de NO alimentación, en especial BonÀrea Energía.

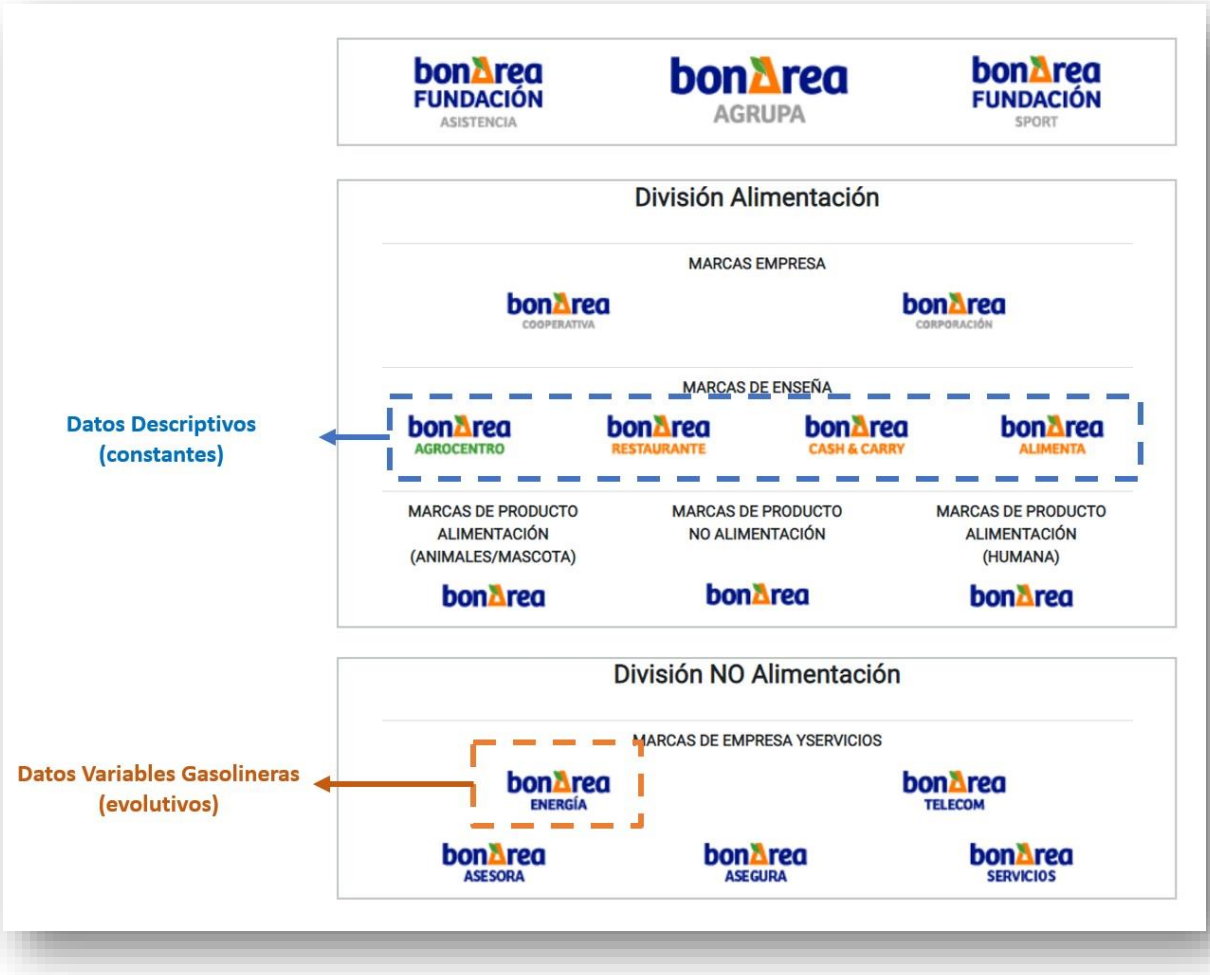


Figura 13: Estructura empresaria BonÀrea (Fuente: https://www.bonarea.com/)

Después de una búsqueda/rastreo de la página web de BonÀrea y diversas websites relacionadas con la empresa, no existe, al menos públicamente un recopilatorio único elaborado por BonÀrea donde se muestre o se pueda descargar toda la información descriptiva de cada establecimiento.

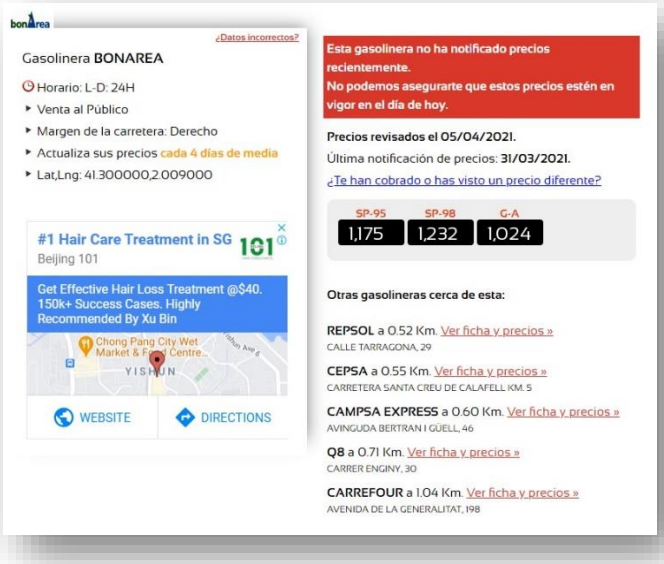
De igual manera, no existe, al menos públicamente, ningún análisis elaborado por BonÀrea de la evolución de los precios de sus carburantes por cada una de sus gasolineras, ni por fechas, ni por días de la semana.

Es cierto que algunas websites tratan de recopilar todos los datos posibles relacionados con ciertas gasolineras, pero a veces la información no es completa. Tómese como ejemplo la siguiente página web: (https://www.dieselogasolina.com/), que muestra los precios de los carburantes de una gasolinera en concreto del grupo BonÀrea, pero a diferencia de la website de BonÀrea, no muestra todos los carburantes disponibles (el ADBLUE no está analizado).

Por tanto, la extracción de la información desde página raíz es necesaria para poder extraer un juego de datos completo y 100% fiable y actualizado.

Ejemplo tomado para la gasolinera de GAVA de BonÀrea donde se observa que otras páginas web no disponen de toda la información:

Figura 14: Muestra la información de la gasolinera G042 (GAVA) de BonÀrea ***con ADBLUE*** *(Fuente: https://www.bonarea.com/)*



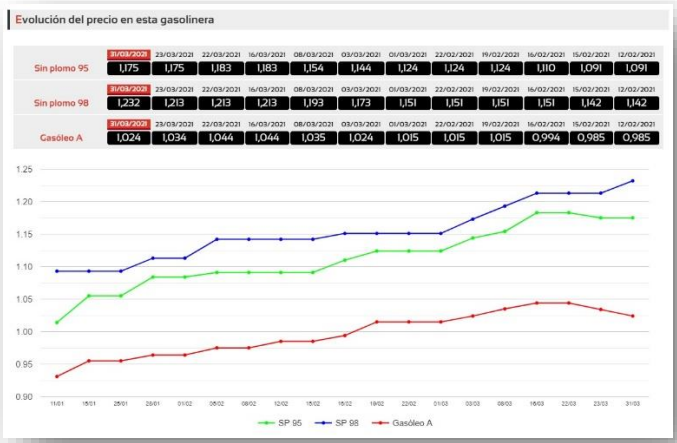


Figura 15 y 16: Muestra la información de la gasolinera G042 (GAVA) de BonÀrea ***sin ADBLUE*** *(Fuente: https://www.dieselogasolina.com /)*

7. Inspiración

7.1 Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

El juego de datos generado también nos aporta la información sobre los distintos establecimientos de la marca BonÀrea, así como sobre los precios diarios de los carburantes (por productos y por gasolineras) en sus gasolineras.

Los datos son accesibles vía web, pero no es posible obtener con facilidad un listado de establecimientos de los que dispone la empresa ya que la información de cada establecimiento se muestra **bajo demanda del usuario.** Los precios de los carburantes también son accesibles vía web como parte de la información básica de las gasolineras y solamente podemos disponer del **último dato**, **pero no del histórico de los precios**. El web scraping proporciona una ventaja añadida: la posibilidad obtener un histórico de los precios de los carburantes a través de la automatización del proceso mediante la implementación de un cron si se desea.

Esta información es realmente interesante para el usuario (y también para la competencia) puesto que se puede generar diferentes aplicaciones o visualizaciones que pueden ser base en toma de decisiones tanto particulares (el usuario puede elegir qué gasolinera repostar por ser más barata) como a nivel estratégico (la empresa competidora puede utilizarlos en su beneficio).

Ejemplos de posibles gráficas que se podría generar, serían las siguientes:

Heatmap geográfico de precios (siendo el color rojo indicador de productos más caros):

El mapa nos mostraría los puntos calientes (carburantes más caros) a nivel geográfico. Esta visualización sería de gran utilidad en toma de decisiones empresariales tanto de BonÀrea como de la competencia.

Un ejemplo sería que todo el repostaje de una flota de una empresa de transporte se hiciese en los lugares más baratos fuera de las grandes ciudades.
Otro ejemplo útil para la competencia sería bajar los precios en sus gasolineras cercanas a los puntos calientes, para poder así competir más eficientemente con BonÀrea.

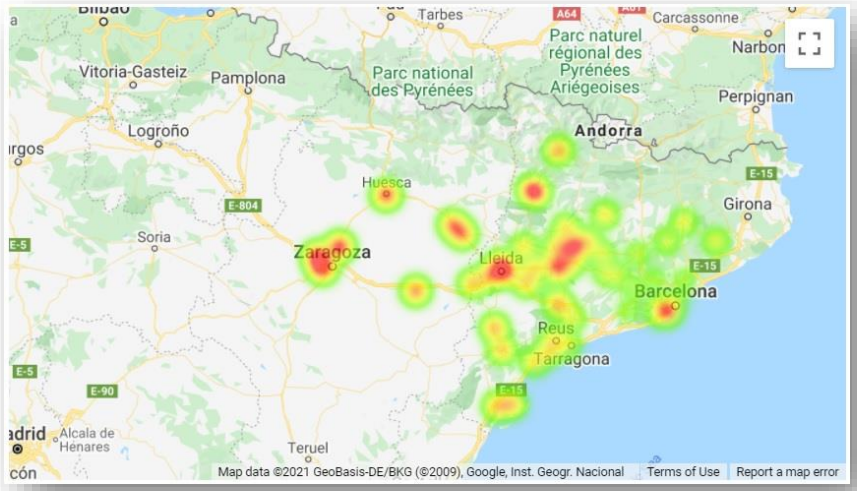


Figura 17: Heatmap creado con el juego de datos (Fuente: Elaboración propia mediante Google Maps)

Evolución temporal del precio del carburante por producto y por gasolinera:

Una de las preguntas que todo conductor nos hacemos alguna vez es cuánto y cuándo suben los precios de los carburantes.

El juego de datos puede solventar esta pregunta mostrando la evolución temporal del precio de **cada tipo** de carburante **por cada gasolinera** existente.

La función creada recupera el input del usuario a nivel de gasolinera y tipo de carburante, resultando una gráfica lineal temporal donde se podría observar ciertos patrones de comportamientos en los cambios de los precios.

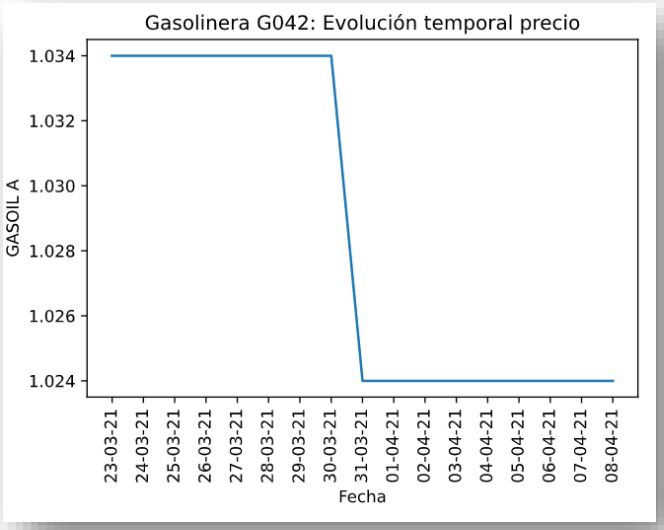


Figura 18: Evolución temporal precio del carburante (Fuente: Elaboración propia a fecha 08 abril 2020)

Estadísticos de los precios de los carburantes por producto y por día de la semana:

El juego de datos resultante también resulta útil si deseamos analizar en profundidad las estadísticas de los precios de los carburantes, y en especial por fecha o por día de la semana.

La gráfica que se muestra a continuación solamente es un ejemplo de los posibles análisis / gráficas estadísticas que se podría realizar con los datos recolectados de la página web.

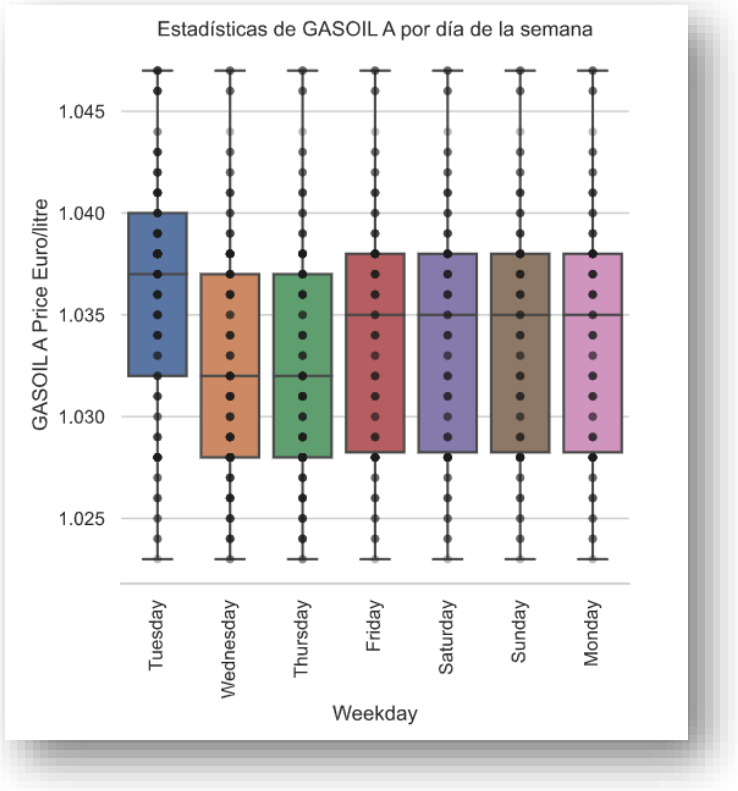


Figura 19: Box Plot de los precios del carburante por día de la semana (Fuente: Elaboración propia a fecha 08 abril 2020)

Por otro lado, el dataset también aporta información de los servicios asociados a cada gasolinera, donde éstos pueden ser:

RENTADOR	- Tiene túnel de lavado de coches
CANVI	- Tiene cambio para dinero, booleano
SUPER	- Tiene supermercado
LAVABO	- Tiene lavabo
PARKING	- Tiene parking
VENDING	- Tiene vending

Y, en consecuencia, además de la información adicional que el usuario pueda obtener de los diferentes servicios que aporta cada gasolinera, estadísticas sobre los precios de los carburantes mediante la agrupación de los diferentes servicios provistos podría aportar información oculta, que a primera vista puede ser no trivial.

Aplicaciones usando Machine Learning:

Hay otra aplicación que los estudiantes ven de hecho como un potencial análisis a realizar. Se ha observado que los precios oficiales en la página web no varían todos los días, aunque es sabido que los precios diarios sí que cambian. Los estudiantes se han realizado la pregunta del porqué esta frecuencia de actualizaciones de precios y si pudiera haber algún patrón escondido detrás.

Debido a que el juego de datos captura no solamente la fecha, sino también el día de la semana, y una vez la recolección de los datos pudiese ser grande (mínimo unos meses de actualizaciones o incluso del orden anual) se podría realizar un algoritmo de clustering para agrupar las diferentes fechas, días de las semanas y gasolineras para poder descubrir un posible patrón de comportamiento.

Comparación con los análisis mostrados en la web https://www.dieselogasolina.com/:

En referencia a los análisis mostrados con el apartado 6 de la presente memoria, cabe destacar que se ha mejorado sustancialmente la elaboración y extracción de información a través de las diferentes gráficas y visualizaciones, además de dar completitud a los datos (considerando todos los tipos de carburantes existentes), puesto que en los análisis anteriores no se mostraba el tipo de carburante ADBLUE.

Nota:
Este proyecto, además de proporcionar los datos y el código para conseguirlos, contiene una serie de herramientas para generar visualizaciones sobre los datos: función que crea un heatmap geográfico de precios y funciones que crean gráficas temporales con la evolución del precio de los carburantes por gasolineras y visualizaciones tipo boxplot con estadísticas para los precios.

8. Licencia

8.1 Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

Released Under CC0: Public Domain License

Released Under CC BY-NC-SA 4.0 License

Released Under CC BY-SA 4.0 License

Database released under Open Database License, individual contents under Database Contents License

Other (specified above)

Unknown License

Razones por la no elección de una licencia Creative Commons:

Primero hay que comentar que normalmente una licencia **Creative Commons** se usa cuando un autor quiere dar a otras personas el derecho de compartir, usar y construir sobre un trabajo que el mismo autor ha creado. Bajo estas circunstancias, los estudiantes **no han creado los datos subyacentes** del juego de datos, sino que **los han recolectado** de una página web, donde estaban expuestos públicamente.

Por este motivo, los estudiantes no creen que una licencia CC pudiese ser la más adecuada para este proyecto.

La licencia escogida es:

Open Data Commons Open Database License (ODbL) v1.0

asociada con:

Database Contents License (DbCL) v1.0

Razones por la elección de una licencia ODbL asociada a DCL:

Las razones de dicha elección son principalmente porque es una licencia creada y basada para dar cobertura a juegos de datos y su contenido.

Con esta licencia, en términos generales se pueden **compartir** (copiar, distribuir y utilizar la base de datos), **crear** (producir obras a partir de la base de datos) y **adaptar** (modificar, transformar y construir sobre la base de datos) siempre que:

- Se atribuya cualquier uso público (o trabajos asociados) de forma expuesta en la licencia
- Se designe la misma licencia a los trabajos resultantes.

A su vez, la licencia DCL cubre el contenido del juego de datos, por tanto, se complementa bien con la OBdL.

9. Código

9.1 Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

La generación del dataset se ha generado utilizando el lenguaje de programación Python.

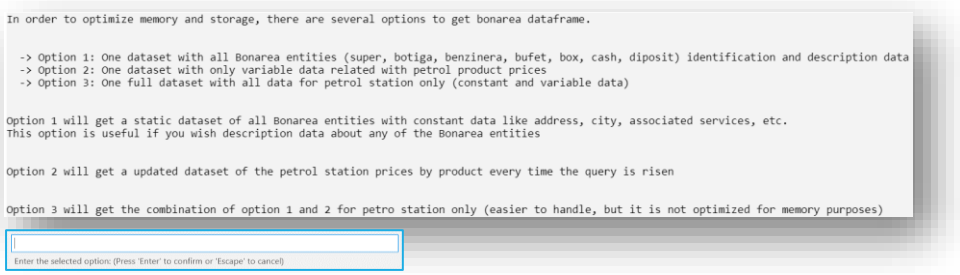
El código utilizado y archivos auxiliares se pueden encontrar en https://github.com/ bajo la siguiente dirección web:

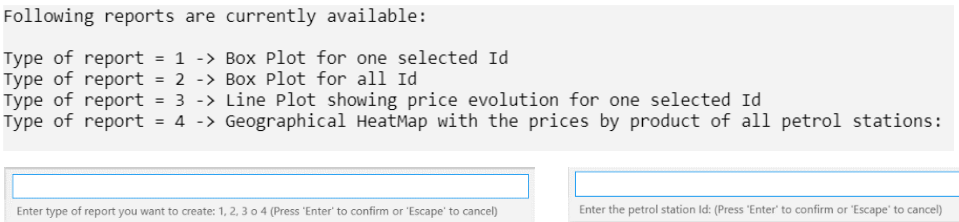
https://github.com/Carlos-Acosta/webscraping.git

Todas las funciones del fichero ***main_functions*** pueden usarse de manera independiente para proporcionar al usuario la mayor flexibilidad de selección del contenido que desea obtener. Si lo desea, el usuario puede realizar el web scraping con las funciones proporcionadas y desarrollar sus propios algoritmos para guardar los datos en formatos diferentes del csv. Además, para usuarios con menor conocimiento de programación, proporcionamos unos ficheros csv con los datos obtenidos hasta el momento y un generador de funciones "a la carte", que permite obtener los datasets predefinidos sin necesidad de programar.

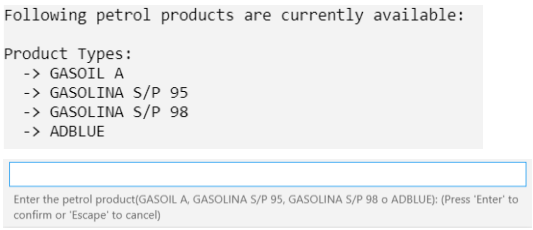
El código utilizado no genera simplemente los datos una vez se ejecutan los archivos, sino que se ha implementado pensando en las preferencias del usuario final, por tanto, se generan diversos inputs que dicho usuario debe de completar para realizar la acción adecuada a la consulta (con sus respectivos controles de consistencia).

En particular, para la obtención de los archivos de datos, el usuario recibirá el siguiente mensaje:

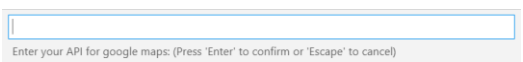




Y para la obtención de las visualizaciones se mostrará una lista con todas los ID de las gasolineras



En el caso del heatmap, se solicitará el API de Google Maps (el usuario deberá crear una desde : https://developers.google.com/maps).



10. Dataset

10.1 Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El dataset se ha publicado a día 08 de abril de 2021 en formato csv en Zenodo.

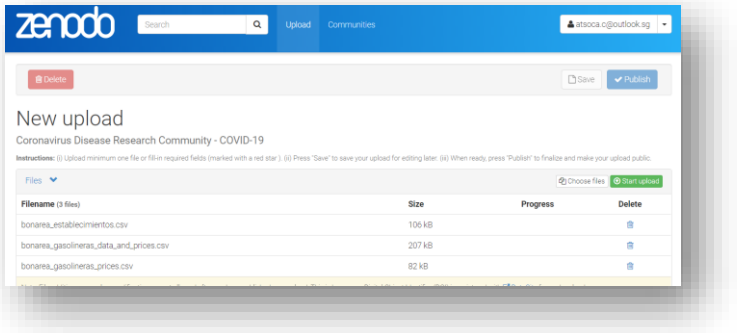
El DOI obtenido se muestra a continuación:

10.5281/zenodo.4671856

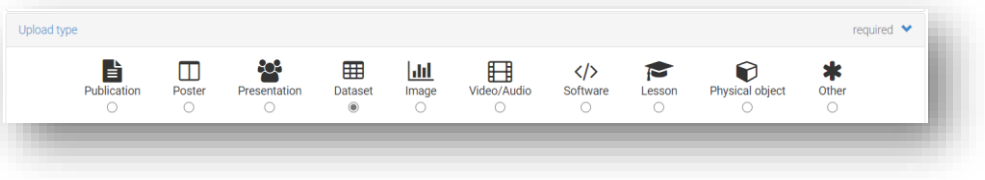
El link de la página de Zenodo donse se dispone el dataset es el siguiente:

https://doi.org/10.5281/zenodo.4671856

Para la obtención del DOI, se han subido los 3 archivos csv correspondientes:

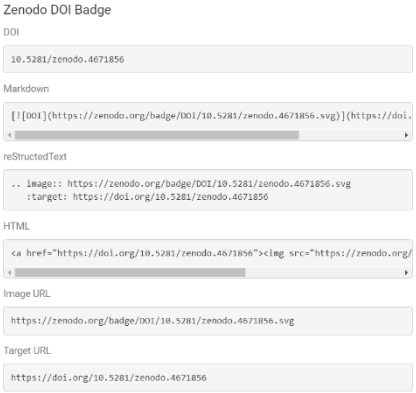


Se ha seleccionado el tipo de datos a subir:



Se han indicado las diferentes informaciones que eran requeridas, incluyendo una breve descripción del juego de datos a subir:

Y finalmente se ha obtenido
la confirmación del DOI suministrado:



11. Tabla de contribuciones al trabajo

Mediante la siguiente tabla, los estudiantes Olga Garcés Ciemerozum y Carlos Acosta Quintas certifican que ambos han colaborado y elaborado conjuntamente tanto en la Investigación previa del proyecto, como en la redacción de las respuestas y el desarrollo del código.

Contribuciones	Firma
Investigación previa	*O. G. / C. A.*
Redacción de las respuestas.	*O. G. / C. A.*
Desarrollo código	*O. G. / C. A.*

Referencias / Fuentes de Información

- -Lawson, R. (2015). _Web Scraping with Python_. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Mitchel, R. (2015). _Web Scraping with Python: Collecting Data from the Modern Web_. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
- Lawson, R. (2015). _Web Scraping with Python_. Packt Publishing Ltd. Chapter 5. Dynamic Data.