

1. Contexto.

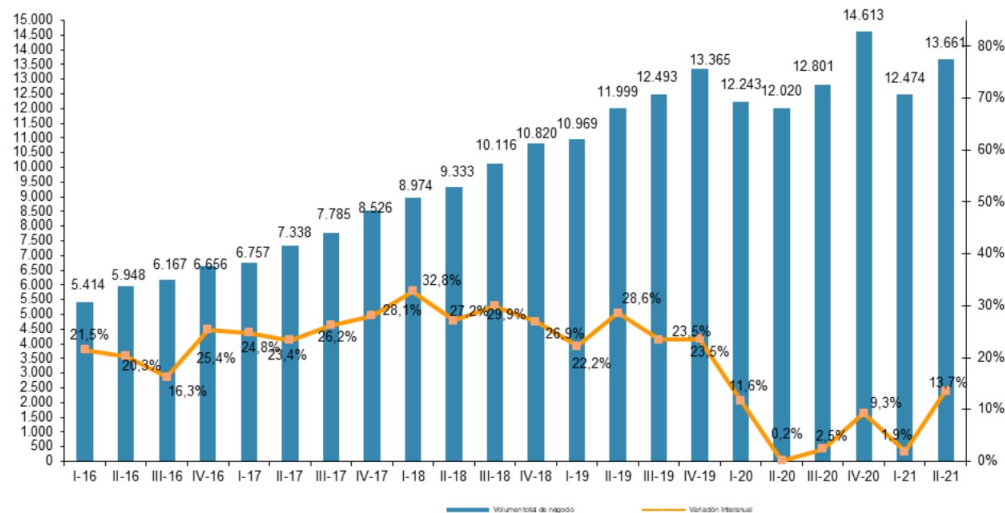
Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web elegido proporciona dicha información.

La empresa **ATiempo** es una empresa de transporte familiar donde el beneficio está más que ajustado. Abaratar costes se ha convertido en un reto muy difícil dada la competencia y el contexto socioeconómico en el que nos encontramos.

Un cambio de los hábitos de consumo ha supuesto que, durante la última década, el comercio electrónico haya experimentado un fuerte crecimiento con una tendencia al alza.

Según los datos publicados por la ¹[Comisión Nacional de los Mercados y la Competencia \(CNMC\)](#), las ventas del negocio en la red en España han superado los 13.600 millones de euros en el segundo trimestre de 2021, un 13,7% más que el mismo período de 2020.

EVOLUCIÓN TRIMESTRAL DEL VOLUMEN DE NEGOCIO DEL COMERCIO ELECTRÓNICO Y VARIACIÓN INTERANUAL (millones de euros y porcentaje)



Fuente: CNMC

Análogamente, se estima que el volumen de negocio del mercado de paquetería empresarial registre una aceleración en su ritmo de crecimiento. Sin embargo, los acontecimientos que se están dando a nivel mundial, como la pandemia por SARS-CoV-2 o la guerra en Ucrania, están provocando la caída del consumo y la actividad industrial, y consecuentemente, una disminución en la paquetería industrial.

Este marco geopolítico en el que nos encontramos, ha provocado una subida extrema y repentina de los precios del combustible; costes que no puede asumir el sector de transportista, dando a lugar a paralizaciones y protestas.

Protestas de transportistas en Italia, Portugal y Alemania por los altos precios que los asfixian

Transportistas en pie de guerra en Europa por los elevados precios del combustible. El sector se ve asfixiado y su protesta espontánea ha traspasado fronteras. Con la guerra de Ucrania el precio del combustible se ha disparado en todas partes, por encima de los 2 euros en Italia.

²Fuente Euronews.com

Precio Medio Anual España -VS- Cotización del Barril Brent



Por lo que, nos encontramos en un sector en auge, como es el de la mensajería, con una sociedad con unos hábitos de consumo diferentes pero con una situación social que hace que el mercado sea inestable.

Por este motivo, **ATiempo** ha iniciado un proceso de transformación digital donde su principal objetivo se centra en la mejora del servicio y trato para con el cliente. Para ello, desarrollarán una app con la que el cliente tenga mayor facilidad para tramitar envíos y, sobre todo una mejor trazabilidad de sus envíos o peticiones. El agilizar las rutas de entrega, seguida por sus transportistas, es primordial. Las unidades de movilidad estarán dotadas de sistemas de geolocalización, el conductor deberá ceñirse a las hojas de rutas preconfiguradas el día anterior y adaptadas a las circunstancias del tráfico y las entregas. Como parte del reparto se incluirá la operación del repostaje. En el momento que se detecte tal necesidad, el sistema incluirá la ruta a la gasolinera más cercana con un precio más bajo, ese es el cometido del presente dataset.

2. Título. Definir un título que sea descriptivo para el dataset.

Información de las estaciones de servicio y del precio de los carburantes.

3. Descripción del dataset.

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

El dataset resultante proporciona la información de las estaciones de servicio que se encuentran cerca de la posición de la unidad móvil, y muestra la información básica de la estación de servicio así como los precios de los carburantes...

Información de la gasolinera y localización:

- Provincia
- Localidad
- Dirección
- Horario
- Empresa
- Precio
- Rótulo

Datos de productos:

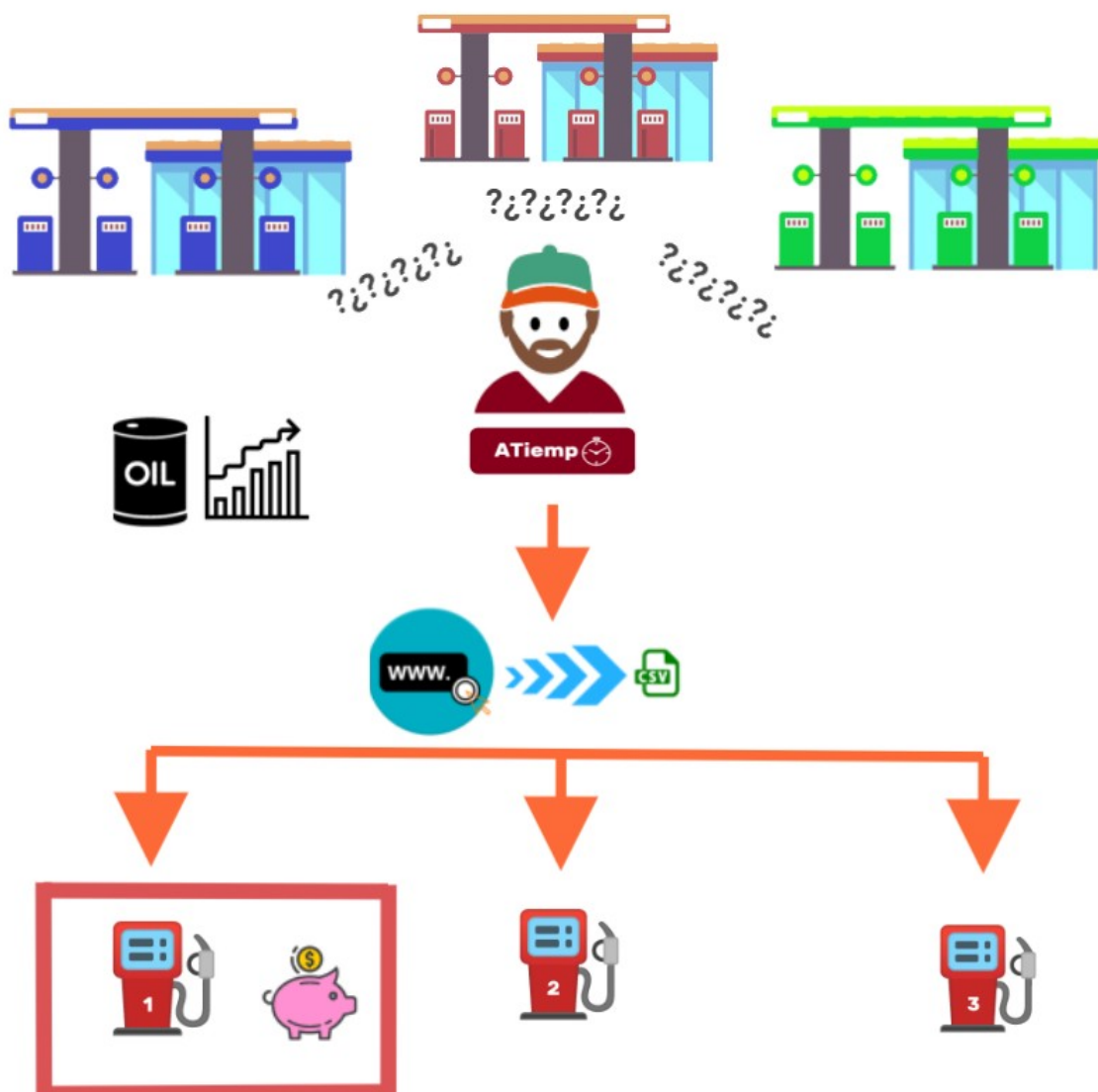
- Gasolina 95 E5
- Gasolina 95 E10
- Gasolina 95 E5 Premium
- Gasolina 98 E5
- Gasolina 98 E10
- Gasóleo A habitual
- Gasóleo Premium
- Gasóleo B
- Gasóleo C
- Bioetanol
- Biodiésel
- Gases licuados del petróleo
- Gas natural comprimido
- Gas natural licuado
- Hidrógeno

Ejemplo Dataset:

```
1 Cod|Provincia|Localidad|Dirección|Fecha|Precio|Venta|Remisión|Empresa|Horario
2 1|TERUEL|ALCAÑIZ|CARRETERA ZARAGOZA KM. 117|11/04/2022|1.779|0|0|ALCAMPO|L-D: 24H|L-D: 00:00-08:30 (D), 08:30-21:30 (A), 21:30-23:59 (D)
3 2|TERUEL|ALCAÑIZ|AVENIDA ZARAGOZA, 100|09/04/2022|1.789|0|0|REPSOL|L-D: 06:00-22:45|L-D: 06:00-22:45 (A)
4 3|TERUEL|ALCAÑIZ|AVENIDA MAESTRAZGO, 4|11/04/2022|1.794|0|0|CEPSA|L-D: 07:00-22:45|L-D: 07:00-22:45 (A)
5 4|TERUEL|ALCAÑIZ|POLIGONO LAS HORCAS FINCA 4 1/A-9, 4|10/04/2022|1.799|0|0|GBA|L-D: 24H|L-D: 24H (A)
```

4. Representación gráfica.

Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.



5. Contenido.

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

Campo	Tipo	Descripción	Ejemplo
Provincia	string	Provincia donde se encuentra la estación de servicio	Teruel
Localidad	string	Localidad donde se encuentra la estación de servicio	Teruel
Dirección	string	Dirección donde se encuentra la estación de servicio	AVENIDA SAGUNTO, 5
Fecha	string	Horario de apertura del establecimiento	22/03/2022
Precio	string	Precio actual del carburante seleccionado	1.769 €/l
Empresa	float	Empresa propietaria de la estacion de servicio	REPSOL
Horario	string	Fecha de actualización de los datos	09/04/2022

6. Agradecimientos.

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. Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.

El precio actualizado de los carburantes se obtiene desde el portal que ofrece el [4Ministerio para la Transición Ecológica y el Reto Demográfico](#).

La información se actualiza cada 5 minutos. Los operadores mayoristas deben remitir al Ministerio el precio que comunique al gestor y éste a su vez también debe remitirlo al Ministerio el precio que fije. [5Geoportal de hidrocarburos](#) ha sido galardonado con diferentes premios, como puede ser el premio CNIS 2015 al mejor servicio en movilidad.

Nos decantamos por esta fuente de información por ser la información oficial del Ministerio. Barajamos otras fuentes, como puede ser el portal [6dieselogasolina.com](#) pero estudiando de donde toman los datos vemos que acuden a la misma fuente oficial.

Examinamos el mapa del portal web y detectamos que proporciona mecanismos para poder consultar y descargar la información sobre el precio de los carburantes en las distintas estaciones de servicio. La información está disponible y de acceso libre y no necesitamos un acceso que nos puedan bloquear.

De todos modos, examinamos el archivo “robots.txt” para detectar posibles restricciones.

Utilizamos la función definida en el fichero robots.py de la siguiente forma:

```
print(robot_txt('https://geoportalgasolineras.es/#'))

<!DOCTYPE html>
<html data-ng-app="geoportalApp" lang="es">
<!-- Faro_Geoportal -->
<head>
<meta http-equiv="Content-Type" content="text/html" charset="UTF-8">
<meta http-equiv="X-UA-Compatible" content="IE=edge" />
<meta name="keywords" content="Estaciones de Servicio, precio, gasolina, di&eacute;sel,
gas&oacute;leo" />
<meta name="rating" content="GENERAL" />
<meta name="subject" content="Computers:Internet:WWW" />
<meta name="description" content="Geoportal de precios de carburantes en las Estaciones de
Servicio" />
<meta name="robots" content="ALL" />
<meta name="distribution" content="GLOBAL" />
<meta name="classification" content="Online Resource" />
<meta name="author" content="MINETUR" />
<meta http-equiv="Cache-Control" content="no-cache, no-store, must-revalidate" />
<meta http-equiv="Pragma" content="no-cache" />
<meta http-equiv="Expires" content="0" />
<meta name="google-signin-scope" content="profile email">
<meta name="google-signin-client_id" content="736064239674-
c2eg0fjfailoaglimjo13n38sioc1rfd.apps.googleusercontent.com">
<link rel="image_src" href="resources/imagenes/chosen/chosen-sprite.png"></link>
<link rel="shortcut icon" type="image/png" href="resources/imagenes/favicon.ico"></link>
<title>Geoportal</title>
```

El portal está alojado en un servidor Apache Tomcat, está desarrollado en Angular y utiliza la api de Google Maps.

```
import builtwith
print(builtwith.builtwith('https://geoportalgasolineras.es'))

{'web-servers': ['Apache', 'Apache Tomcat'], 'programming-languages': ['Java'], 'operating-
systems': ['Windows Server'], 'web-server-extensions': ['mod_jk'], 'widgets': ['AddThis'],
'javascript-frameworks': ['AngularJS', 'Moment.js', 'Underscore.js', 'jQuery', 'jQuery UI'],
'maps': ['Google Maps', 'OpenLayers']}
```

7. Inspiración.

Explicar 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 interés de analizar este conjunto de datos es que a causa de los acontecimientos que se están dando a nivel mundial, como son la pandemia por SARS-CoV-2 y la guerra en Ucrania, están aumentando sustancialmente el precio de los carburantes. Esto está afectando de especial manera a pequeñas empresas de transporte que están viendo incrementados sus costes directos y disminuyendo su margen de beneficio.

Concretamente, las preguntas que se pretenden responder son: ¿Cuál es la gasolinera más barata? ¿Cuál es el precio actual de los diferentes combustibles? ¿A qué distancia se encuentra la gasolinera de la ruta del transportista?

8. Licencia.

Seleccionar una de estas licencias para el dataset resultante y justificar 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.

Una posible licencia para este conjunto de datos puede ser CC BY-SA 4.02 License. La elección se basa en los siguientes puntos:

- Se provee el nombre del creador del conjunto de datos generado y se indican los cambios realizados sobre este.
- Las nuevas contribuciones deben ser publicadas bajo la misma licencia, lo que permite que se le reconozca al autor original en todo momento y bajo los mismos términos que fueron planteados por él.

9. Código.

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

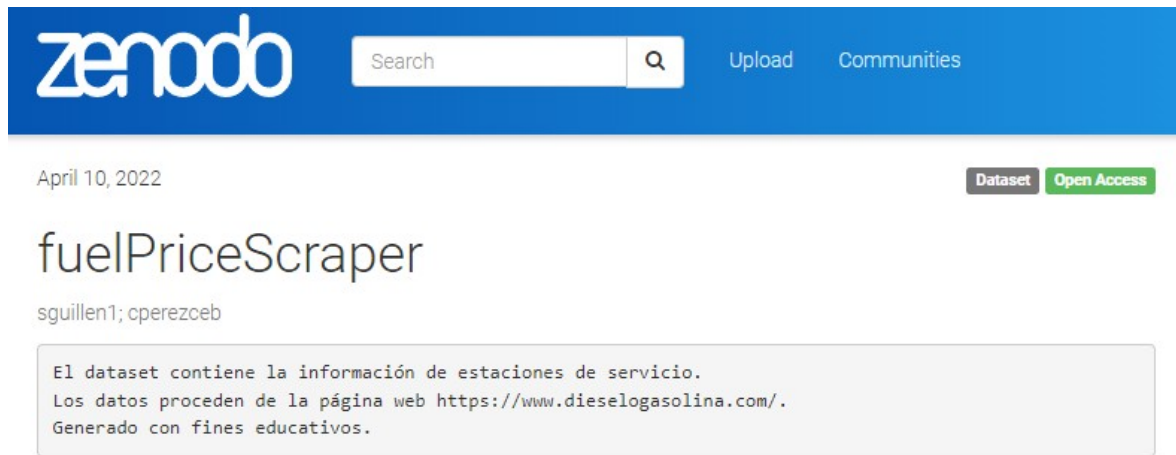
El código Python y el resto de documentación se puede consultar en el repositorio:

<https://github.com/sguillenuoc/fuelPriceScraper>

10. Dataset.

Publicar el dataset obtenido(*) en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.

El dataset generado está publicado en la web de Zenodo y se puede consultar a través del siguiente link:

<https://doi.org/10.5281/zenodo.6436284>

The screenshot shows the Zenodo interface. At the top is the Zenodo logo and a search bar. Below the logo, the date 'April 10, 2022' is displayed. To the right of the date are buttons for 'Dataset' and 'Open Access'. The title of the dataset is 'fuelPriceScraper', and the authors are 'sguillen1; cperezceb'. A description box contains the following text: 'El dataset contiene la información de estaciones de servicio. Los datos proceden de la página web <https://www.dieselogasolina.com/>. Generado con fines educativos.'

Zenodo proporciona un API para subir directamente documentos y publicarlos. En el código de python hay una función que carga el dataset generado en la web de Zenodo, a falta de entrar en la web y sí procede publicarlo.

A modo de ejemplo, adjuntamos el json resultado de una de las pruebas realizadas:

```
{'mimetype': 'text/csv',
  'updated': '2022-04-10T17:30:28.333090+00:00',
  'links':
  {
    'self':
      'https://zenodo.org/api/files/ccff8a90-2edb-4742-b382-8277343b294f/StationData.csv',
    'version':
      'https://zenodo.org/api/files/ccff8a90-2edb-4742-b382-8277343b294f/StationData.csv?versionId=c66d7f44-bec1-4a0c-9ae3-1e6ad79a5968',
    'uploads': 'https://zenodo.org/api/files/ccff8a90-2edb-4742-b382-8277343b294f/StationData.csv?uploads'
  },
  'is_head': True,
  'created': '2022-04-10T17:30:28.327374+00:00',
  'checksum': 'md5:d4d6f8ffeac68b90b56cad177a86b56b',
  'version_id': 'c66d7f44-bec1-4a0c-9ae3-1e6ad79a5968',
  'delete_marker': False,
  'key': 'StationData.csv',
  'size': 2426}
```

11. Vídeo.

Se debe hacer entrega de un vídeo explicativo de la práctica en donde cada uno de los integrantes del grupo explique con sus propias palabras tanto las respuestas del proyecto como el código utilizado para llevar a cabo la extracción. El vídeo debe ser enviado a través de un enlace a Google Drive que deben proporcionar, junto con el enlace al repositorio Git, al momento de entregar la práctica.

12. Referencias

1. Comisión Nacional de los Mercados y la Competencia (CNMC)
<https://www.cnmc.es/>
2. Euronews.com
<https://es.euronews.com/2022/03/14/protestas-de-transportistas-en-italia-portugal-y-alemania-por-los-altos-precios-que-los-as>
3. US Energy Information Administration
<https://www.eia.gov/>
4. Ministerio para la Transición Ecológica y el Reto Demográfico
<https://energia.gob.es/es-es/Paginas/index.aspx>
5. Geoportal de hidrocarburos
<https://geoportalgasolineras.es/#/Inicio>
6. dieselogasolina.com
<https://www.dieselogasolina.com/>