

# Práctica1 (35% nota final)

## Alumnos:

**Sergio Roque Duarte Pérez (25598456P)**

**Alejandro Guijarro Gallardo (70081435K)**

## 1 Introducción y Contexto

El presente documento recoge las respuestas a los puntos recogidos en la Práctica 1 de la asignatura Tipología y Ciclo de Vida de los Datos. El informe se ha elaborado siguiendo la estructura propuesta en el enunciado de la práctica.

El desarrollo de la práctica se ha centrado en la extracción de datos relacionados con el virus COVID-19, planteando una evaluación del número de individuos infectados en los últimos meses, junto con el posible impacto en términos de consumo energético.

Los entornos web seleccionados para la extracción de los datos citados son:

- <https://www.worldometers.info/coronavirus/> → estadísticas de individuos infectados.
- <https://www.ree.es/es> --> datos de consumo eléctrico (proporcionados por Red Eléctrica) en España.
- <https://covid19.isciii.es/> → estadísticos sobre la evolución de los casos de COVID-19 por CCAA.

Se han seleccionado estos entornos web con objeto desarrollar técnicas de extracción de datos de Web Scraping y mediante entornos API, entre otros.

Tanto la realización del documento como los aportes de cada uno de los script, los csv y gráficos generados, se han llevado a cabo mediante la plataforma de GitHub en el repositorio: [https://github.com/o92dupes/Web\\_scraping\\_UOC](https://github.com/o92dupes/Web_scraping_UOC)

## 2 Dataset

Los datos obtenidos en ambos entornos web se han almacenado en los siguientes datasets:

- **Consumo elect COVID 2020 & Consumo elect COVID 2019**

Registro con los datos de consumo energético en España para los meses de enero, febrero, marzo, en años 2019 y 2020.

- Ev\_demanda20 → dataset incluyendo datos de demanda eléctrica diaria entre meses de enero-febrero-marzo 2020. Datos exportados en fichero .csv *Consumo\_elect\_COVID\_2020.csv*.
- Ev\_demanda2019 → dataset incluyendo datos de demanda eléctrica diaria entre meses de enero-febrero-marzo 2019. Datos exportados en fichero .csv *Consumo\_elect\_COVID\_2019.csv*.

Generación de ambos dataset mediante la ejecución del script “API - Red Electrica.py”

#### - **Casos COVID mundo**

Registro de datos de coronavirus a nivel mundial. El dataset está agrupado por fecha, número de casos, fallecidos, recuperados y datos de hospitalizaciones, entre otros datos relevantes que nos ayudan a ser conscientes de lo que esta pandemia está suponiendo para el ser humano. Cabe destacar que este conjunto de datos no resulta relevante para el análisis que hacemos sobre el consumo eléctrico, pero pensamos que puede servir como una pequeña introducción o preámbulo para presentar los demás datos.

Datos exportados al fichero “*Casos\_COVID\_mundo.csv*” tras la ejecución del script “*Casos\_COVID\_mundo.py*”

#### - **Casos COVID ESPAÑA**

Registro de datos de coronavirus en España y agrupados por CCAA, número de casos totales, casos en las últimas 24h y su incidencia en los últimos 14 días.

Los datos son exportados al fichero “*Casos\_COVID\_ESPAÑA.csv*” tras la ejecución del script “*Casos\_COVID\_Espana.py*”

#### - **Casos COVID ESPAÑA acumulado**

En este archivo se recogen datos obtenidos mediante descarga directa desde la misma web a la que se le hace el raspado con el script “*Casos\_COVID\_Espana.py*”. En este punto hay que aclarar que se ha optado por el uso de estos datos en el análisis dada la imposibilidad de obtener datos coherentes con los que se aprecian en el código html de la web. Se han incluido ambos scripts en base a la recomendación del profesor encargado de la asignatura.

### 3 Obtención de datos y Contenido Visualización de los resultados,

Se muestran en esta sección de manera esquemática las estructuras para los dataset generados en la práctica.

#### **DATASETS Consumo elect COVID 2019 & Consumo elect 2020**

Se muestra a continuación la estructura para los datasets con los resultados del consumo energético en años 2019 y 2020 (meses enero-febrero-marzo).

Ev Demanda	Date	Year	Month	Day
"dato demanda"	"Year-month-day"	"2020"	1 (enero 2 (febrero 3 (marzo)	1 - 30

*Tabla 1: Estructura Dataset "Consumo\_elect\_COVID\_2020"*

Ev Demanda	Date	Year	Month	Day
"dato demanda"	"Year-month-day"	"2019"	1 (enero 2 (febrero 3 (marzo)	1 - 30

*Tabla 2: Estructura Dataset "Consumo\_elect\_COVID\_2019"*

### **DATASET Casos COVID mundo**

Estructura para el dataset para el dataset con el registro de los datos de coronavirus en el mundo, categorizados por países.

Pais /Otros	Casos totales	Nuevos Casos	Muertes totales	Nuevas muertes	Total recuperados	Casos activos	Casos críticos	Casos 1M pop	Muertes 1M Pop	Total Test	Total Test 1M Pop	Continente

*Tabla 3: Estructura Dataset "Casos\_COVID\_mundo"*

### **DATASET Casos COVID ESPAÑA acumulado**

A continuación se muestra la estructura para el dataset con el registro de los datos de coronavirus en España.

CCAA	Fecha	Casos	Hospitalizados	UCI	Fallecidos	Recuperados

*Tabla 4: Estructura Dataset "Casos\_COVID\_ESPAÑA\_acumulado"*

### **DATASET Casos COVID ESPAÑA**

CCAA	Total	Ult24h	Inc.14d

*Tabla 5: Estructura Dataset "Casos\_COVID\_ESPAÑA"*

## 3.1 Extracción de datos de los entornos web.

Se presenta en este apartado de manera algo más detallada cuales han sido los procedimientos y técnicas utilizados para la extracción de los datos en los entornos web escogidos para el desarrollo de la práctica. Se hace referencia en cada caso a los scripts definidos para generar los datasets.

### 3.1.1 DATASETs Consumo\_elect\_COVID\_2019 & Consumo\_elect\_2020

Para la extracción del dataset con los datos de consumo eléctrico se ha optado por técnicas basadas en aplicaciones API. El script generado para la extracción de este dataset es “*API - Red Electrica.py*”

El entorno web de Red Eléctrica de España <https://www.ree.es/es>, dentro de su sección REData, dispone de un entorno API donde se da acceso a los usuarios al registro de consumo eléctrico en España, por comunidad autónoma, en diferentes franjas temporales, diarias, mensuales y anuales.

Se proporciona dentro del entorno API de la web, una guía con un desglose detallado para el usuario sobre como formular las peticiones de datos en el entorno web. Para el caso utilizado en la práctica, se ha procedido a extraer la evolución de la demanda energética en una franja temporal, entre meses de enero, febrero, marzo, para años 2019-2020. La request formulada es la siguiente:

```
r = requests.get('https://apidatos.ree.es/es/datos/demanda/
#indicamos los datos, en este caso evolución de demanda
    evolucion?
#rango temporal para el que solicitamos los datos
    start_date=2020-01-01T00:00&end_date=2020-0325T22:00&
#fracción temporal en la que queremos visualizar los datos (diario)
    time_trunc=day&
#Zona geográfica sobre la que aplicaremos la extracción de datos.
    geo_limit=peninsular&
    geo_ids=8741')
```

Como resultado, obtenemos un fichero de datos el cual normalizamos a formato Json para su posterior edición.

Este punto es quizás el más tedioso en el desarrollo de este script en concreto, puesto que requiere de un entendimiento del fichero de datos para su posterior tratamiento, en el que se limitará el contenido a los datos de interés.

Concluida la edición del fichero, procedemos a almacenar los datos en un dataframe de acuerdo a la estructura presentada en el apartado 2, para el dataset de consumo eléctrico.

### 3.1.2 DATASET Casos\_COVID\_Mundo

Para efectuar la extracción de los datos de la web <https://www.worldometers.info/coronavirus/> se ha optado por el uso de las librerías BeautifulSoup, Pandas y Requests de Python. El script “Casos\_covid\_mundo.py” adjunto en el repositorio de la práctica 1 de la asignatura, se encuentra totalmente documentado y comentado paso a paso para el perfecto entendimiento de quien lo use.

Pasamos a continuación a describir el funcionamiento del script: en primer lugar, y después de importar las librerías citadas anteriormente, se procedió a definir una función llamada `casos_covid()` que se encargará de obtener el código html de la web mediante la librería Requests, a continuación se define un mensaje de error en caso de que no se pueda acceder a la web, no se encuentre la tabla o no podamos obtener datos de la misma. Lo siguiente será ubicar el tag donde se encuentra el cuerpo de la tabla mediante sentencias de la librería Bs4 para después crear un bucle que iterará sobre todas columnas y filas de la tabla. Por último, se definirá un DataFrame con los nombres de las columnas y lo exportaremos como el archivo “Casos\_COVID\_mundo.csv” deseado.

### 3.1.3 DATASETs Casos\_COVID\_ESPAÑA

En este apartado se hablará del procedimiento seguido para la obtención de los datasets Casos\_COVID\_ESPAÑA y Casos\_COVID\_ESPAÑA\_acumulado.

Para el primero, se ha realizado una adaptación del script para la extracción de los datos “Casos\_COVID\_mundo.csv”. De nuevo hemos de recalcar que, aunque el script funcione y nos genere un archivo .csv con datos de individuos contagiados por comunidades autónomas, los datos no se corresponden con los que se muestran en la página web. Dado que no se encuentra solución para este problema, se ha optado por actuar de otra manera.

Es por lo expuesto anteriormente que se ha decidido descargar de forma directa los datos acumulados desde la misma web para poder realizar el análisis final de este trabajo.

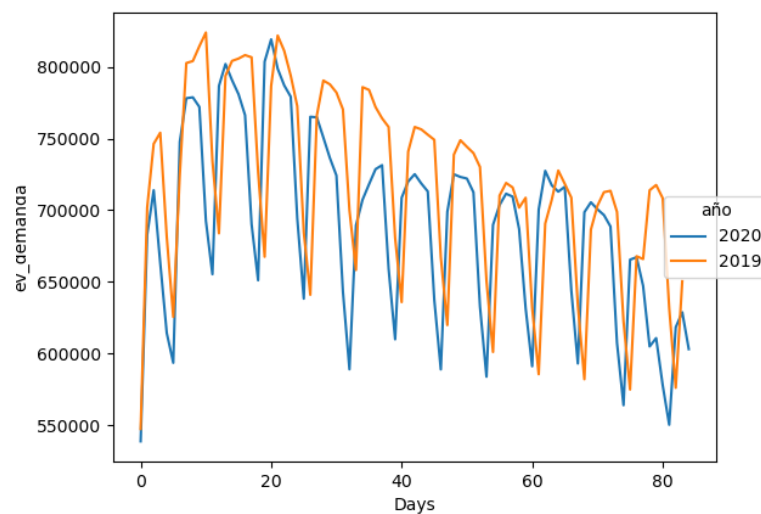
## 4 Visualización de los resultados

Se presenta en este apartado una visualización de los datos a partir de los datasets generados.

En primer lugar, se muestra la evolución en el consumo eléctrico, con una comparativa entre los años 2019 y 2020.

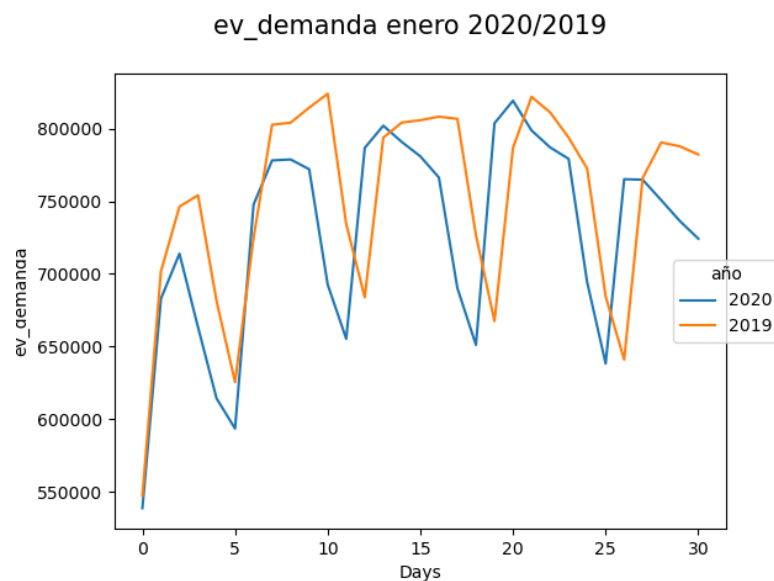
- Comparativa para meses de enero-febrero-marzo

ev\_demanda enero-febrero-marzo 2020/2019



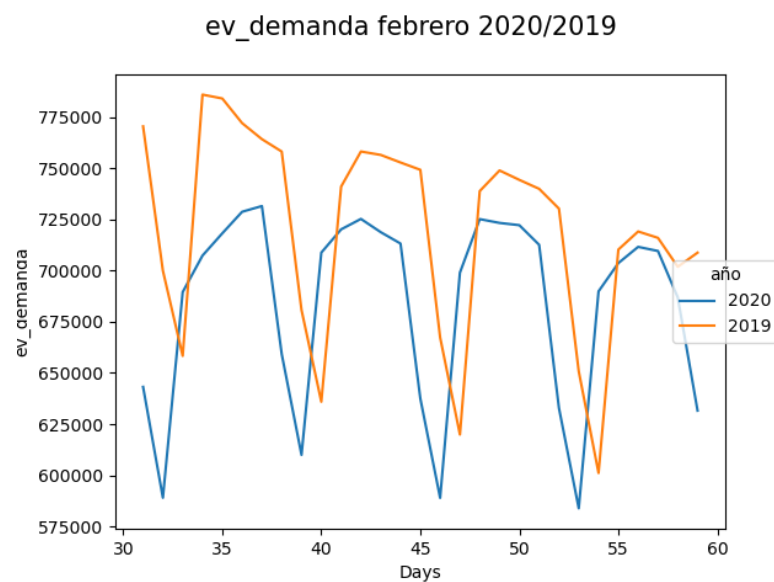
*Figura 1: Comparativa Ev. demanda entre 2019-2020 enero-febrero-marzo*

- Comparativa para mes de enero



*Figura 2: Comparativa Ev. demanda entre 2019-2020 enero*

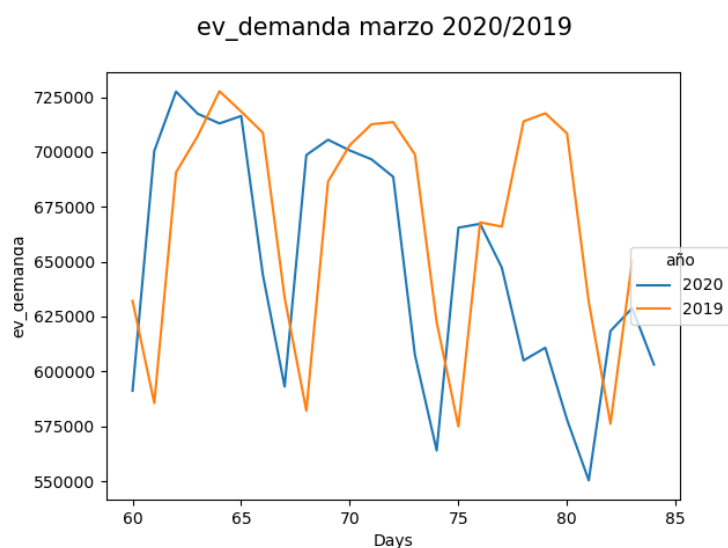
- Comparativa para meses de febrero



*Figura 3: Comparativa Ev. demanda entre 2019-2020 febrero*



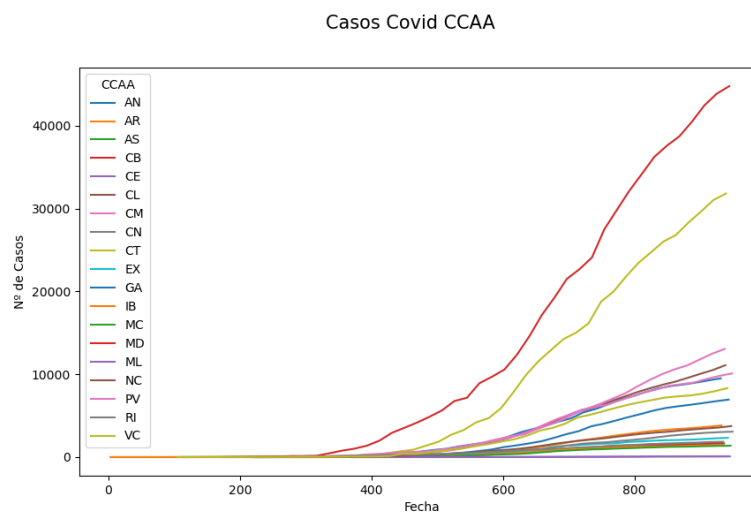
- Comparativa para meses de marzo



*Figura 4: Comparativa Ev. demanda entre 2019-2020 marzo*

En lo que respecta a los datasets que reflejan la evolución de casos de coronavirus, se muestran a continuación los gráficos generados para los datos registrados en España.

- Comparativa para meses de enero-febrero-marzo



*Figura 5: Evolución casos coronavirus por Comunidad Autónoma*

## 5 Conclusiones

Es quizás pronto para identificar una relación y cambio de tendencia claro en una franja temporal tan reducida, pero por una simple comparativa en la evolución de la demanda entre 2019, 2020, si parece identificarse en el mes de marzo un cambio significativo hacia finales de mes que podría estar relacionado con el notable aumento de casos afectados por el virus.

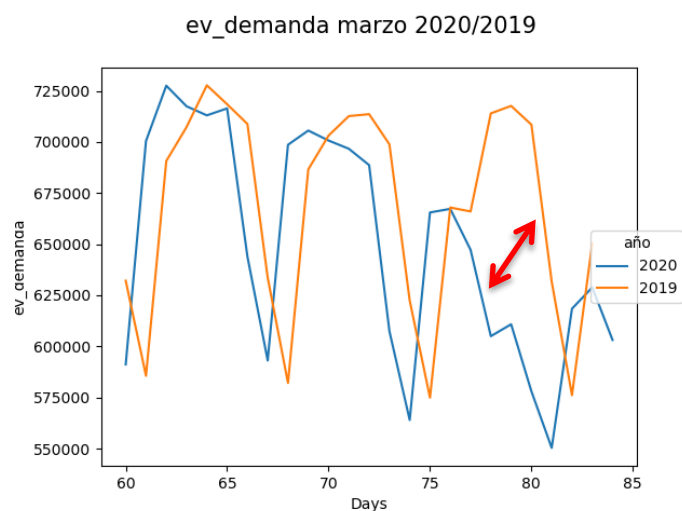


Figura 6: Comparativa Ev. demanda entre 2019-2020 marzo

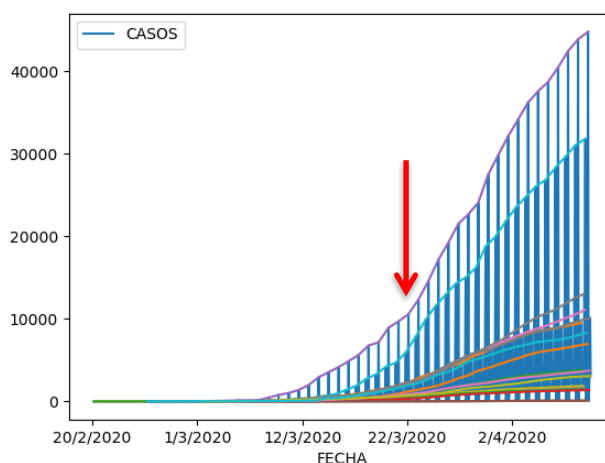


Figura 7: Evolución casos coronavirus en España

Aunque los datos y la perspectiva que como ciudadanos tenemos del problema indican una posible tendencia en este sentido, es pronto para establecer conclusiones en base a los datos. Será necesario llevar a cabo esta comparativa quizás pasados unos

meses, momento en el que dispondremos de datos contrastados sobre la mesa para realizar una valoración del impacto del virus en términos de consumo energético.

## 6 Inspiración y Agradecimientos.

### Inspiración

Una de las razones para desarrollar la presente práctica entorno al COVID-19 es la de intentar responder a algunas de las preguntas que ambos autores de esta práctica, entendemos muchas personas han empezado a hacerse: cual es realmente el impacto de este virus y cuando vamos a ser capaces de cuantificarlo realmente.

El hecho de optar por orientar este análisis del impacto del virus hacia el sector energético responde a los primeros datos comunicados recientemente por algunos medios sobre el notable descenso de la contaminación como consecuencia de la evidente reducción a mínimos de la actividad productiva.

### Agradecimientos

Mencionar en este punto en primer lugar a los entornos web (Red Electrica Worldometers) que han facilitado los datos para llevar a cabo la realización de esta práctica.

Agradecer a su vez al profesorado de la UOC, para la asignatura “Tipología y ciclo de vida de los datos”, por el soporte y feedback proporcionado durante la preparación de la práctica.

## 7 Datos: propiedad intelectual y Licencia

El tipo de licencia para el tratamiento de los datos de la web <https://covid19.isciii.es/> se corresponden con la CC0: Public Domain License. Estos datos se pueden usar ya que son de dominio público.

El tipo de licencia que usa <https://www.worldometers.info/coronavirus/> es de carácter privado. El uso que se quiera hacer de los datos que ofrece el entorno web (y que no sean con fines educativos, como es el caso) deberá ser notificado a los propietarios para que reciban la remuneración correspondiente.

En lo que respecta al entorno web <https://www.ree.es/es>, Lilos datos extraídos del entorno API y utilizados en la presente práctica son propiedad de Red Eléctrica, permitiéndose un uso de los mismos siempre y cuando no se incurra en fines comerciales, sujeto a las condiciones establecidas para licencias-tipo en el artículo 4.2. b de la Ley 37-2007 de 16 de noviembre.

## 8 Citas de investigaciones previas

En lo referente a investigaciones previas, se ha optado por incluir como referencia un estudio desarrollado por Deloitte sobre el impacto del Covid-19 en diferentes sectores, entre ellos el energético.

- Estudio de Deloitte

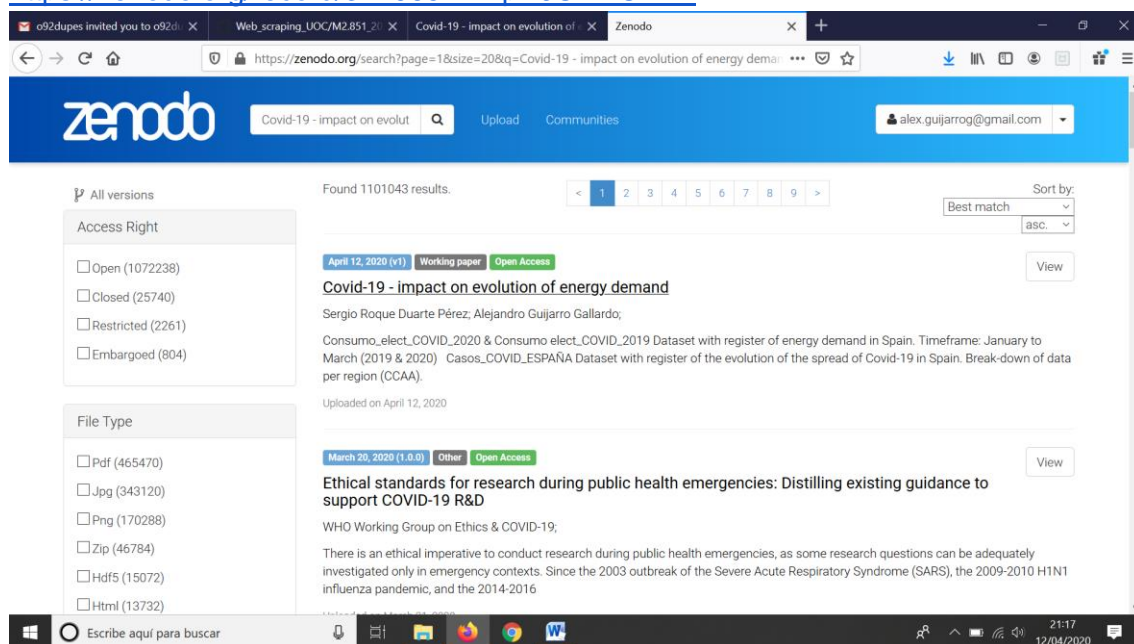
<https://www2.deloitte.com/content/dam/Deloitte/es/Documents/acerca-de-deloitte/Deloitte-ES-Barometro-de-empresas-COVID-19-20200329.pdf>

## 9 Publicación de datos en Zenodo

Los datasets obtenidos en la práctica, exportados en formato .csv se han publicado en Zenodo bajo el título: **Covid-19 - impact on evolution of energy demand**

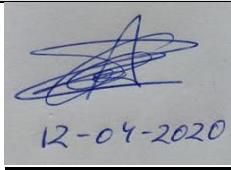

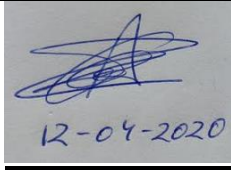

Se incluye a continuación un link a los datos en Zenodo para revisar su contenido:

<https://zenodo.org/record/3749652#.XpRJCZmCE2w>



## 10 Adjuntos

1. **API – Red Electrica.py**
2. **Casos\_COVID\_Espana.py**
3. **Casos\_covid\_mundo.py**
4. **Consumo\_elect\_COVID\_2019.csv**
5. **Consumo\_elect\_COVID\_2020.csv**
6. **Casos\_COVID\_Espana\_acumulado.csv**
7. **Casos\_COVID\_Espana.csv**
8. **Casos\_COVID\_mundo.csv**
9. **Lectura\_datos\_Espana\_Acum.py**

<u>Contribuciones</u>	<u>Firma</u>
Investigación previa	 
Redacción de las respuestas	 
Desarrollo código	