Práctica1 (35% nota final)

**Alumnos:**

**Sergio Roque Duarte Pérez (XXXXX)**

**Alejandro Guijarro Gallardo (70081435K)**

# 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.

Se han seleccionados estos dos entornos web con objeto desarrollar técnicas de extracción de datos de Web Scraping y mediante entornos API.

# 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\_ESPAÑA**

Registro de datos de coronavirus en España. Resultados obtenidos por Comunidad Autónoma, con evolución por fecha en número de casos, fallecidos, recuperados y datos de hospitalizaciones.

Datos almacenados en un dataset temporal “df” exportados en fichero .csv *Casos\_COVID\_ESPAÑA\_acumulado.csv* tras la ejecución del script *“Casos\_COVID\_Espana.py”*

* **Casos\_COVID\_mundo**

Registro de datos de coronavirus por países, con evolución por fecha en número de casos, fallecidos, recuperados y datos de hospitalizaciones.

Cabe destacar en este caso el mayor desglose en los datos de hospitalizaciones, con registro de los casos críticos. El dataset recoge a su vez datos de infectados, muertes por coronavirus por millos en habitantes.

Datos almacenados en un dataset temporal “df” exportados en fichero .csv *Casos\_COVID\_mundo.csv* tras la ejecución del script *“Casos\_COVID\_mundo.py”*

# 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\_ESPAÑA**

Mostrada a continuación la estructura para el dataset con el registro de los datos de coronavirus en España.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CCAA** | **Fecha** | **Casos** | **Hospitalizados** | **UCI** | **Fallecidos** | **Recuperados** |
|  |  |  |  |  |  |  |

Tabla 3: Estructura Dataset "Casos\_COVID\_ESPAÑA"

**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** | **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 4: Estructura Dataset "Casos\_COVID\_mundo”

## 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.

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

Para la extracción del data set 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.

### DATASETs Casos\_COVID\_ESPAÑA

### DATASETs Casos\_COVID\_ESPAÑA

# Visualización de los resultados

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

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

* Comparativa para meses de enero-febrero-marzo

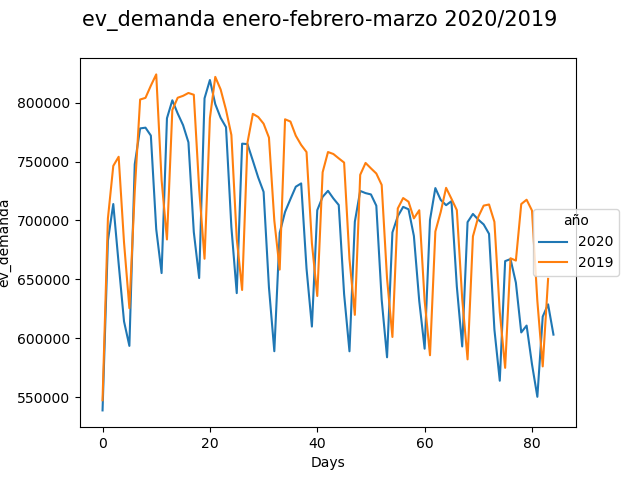


Figura 1: Comparativa Ev. demanda entre 2019-2020 enero-febrero-marzo

* Comparativa para mes de enero

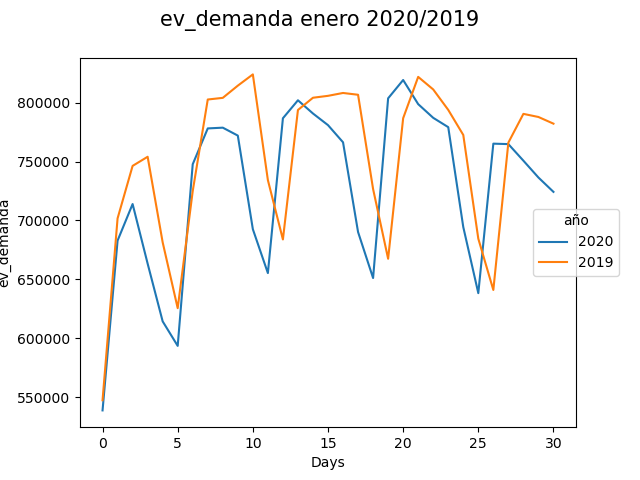


Figura 2: Comparativa Ev. demanda entre 2019-2020 enero

* Comparativa para meses de febrero

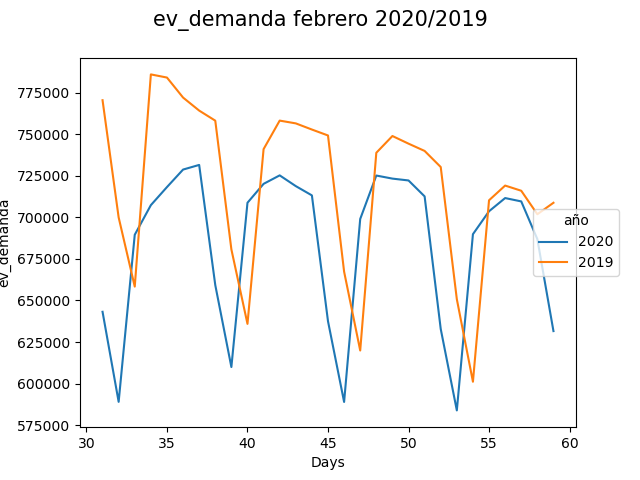


Figura 3: Comparativa Ev. demanda entre 2019-2020 febrero

* Comparativa para meses de marzo

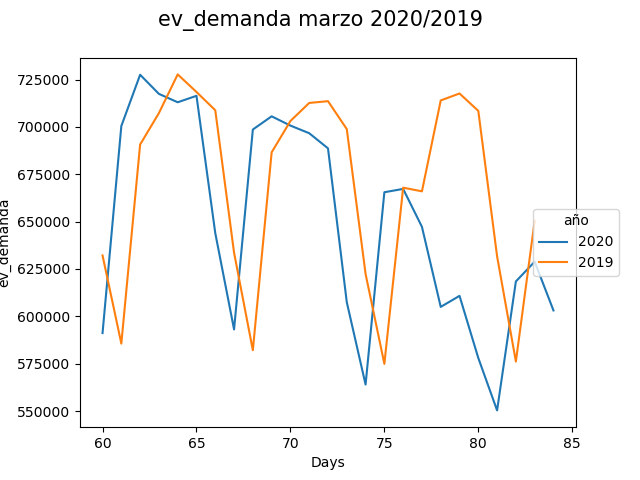


Figura 4: Comparativa Ev. demanda entre 2019-2020 marzo

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

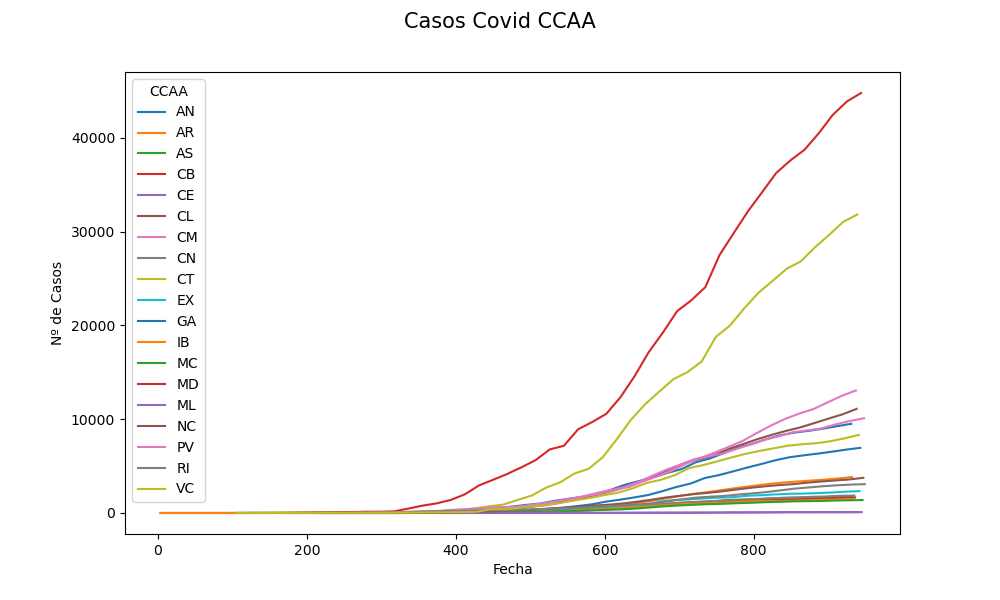


Figura 5: Evolución casos coronavirus por Comunidad Autónoma

# 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.

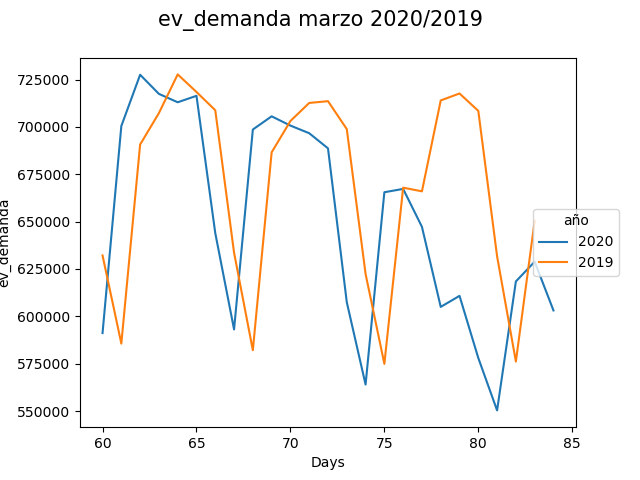


Figura 6: Comparativa Ev. demanda entre 2019-2020 marzo

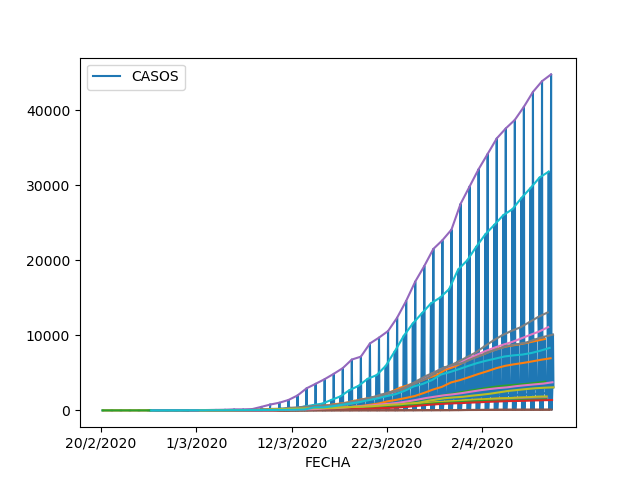


Figura 7: Evolución casos coronavirus en España

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.

# 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.

# Datos: propiedad intelectual y Licencia

Licencia. 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

# 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

# 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\_mundo.csv**
8. **lectura\_datos\_Espana\_Acum.csv**

|  |  |
| --- | --- |
| **Contribuciones** | **Firma** |
| Investigación previa | **SERGIO** |
| Redacción de las respuestas | **SERGIO** |
| Desarrollo código | **SERGIO** |