Práctica1 (35% nota final)

**Alumnos:**

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

**Alejandro Guijarro Gallardo (70081435K)**

# Introducción y Contexto

El presente documento abarca 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 por país.
* <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:

* Ev\_demanda20 ‘🡪 dataset incluyendo datos de demanda eléctrica diaria entre meses de enero-febrero 2020.
* Ev\_demanda2019 🡪 dataset incluyendo datos de demanda eléctrica diaria entre meses de enero-febrero 2020.
* Casos\_COVID\_ESPAÑA 🡪 Aún por ver si lo incluimos o no.
* Casos\_COVID\_mundo 🡪 en este dataset se incluyen los datos actualizados del número de casos confirmados diarios y totales, muertes diarias y totales, test realizados y otras variables de interés por país.

# Visualización de los resultados

**Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente**

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

Tabla 1: Ev\_demanda20

# Obtención de datos y contenido

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

# Agradecimientos e Inspiración

**Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).**

**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 personalmente creo 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 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.

Es quizás pronto para identificar una relación y cambio de tendencia claro en una franja temporal tan reducida, pero es muy posible esta temática adopte una gran importancia en futuros estudios de análisis de datos.

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

# Adjuntos

1. API – Red Electrica.py
2. Casos\_COVID\_Espana.py
3. Casos\_covid\_mundo.py
4. Scapy\_spider.py
5. Consumo\_elect\_COVID.csv

**.Dataset. Publicación del dataset en formato CSV en Zenodo con una pequeña descripción.**

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