

CALIDAD DEL DATO

- Descargar el fichero Santander del servidor Azure y realizar un control de calidad de sus datos.
 - Localiza el total de valores perdidos por columna y fila.
 - ¿Cuántos registros y columnas tienen más de un 5% de valores perdidos?
 - En una primera aproximación de limpieza, quédate con los registros SIN valores nulos.
 - Partiendo del conjunto de datos sin limpiar, realiza una imputación en las 3 columnas con más valores perdidos. Utiliza 3 métodos distintos de imputación, en función del tipo de variable de cada columna (media, moda, mediana, ...)

WEBSCRAPPING

- Al final de los scripts de scrapping (1-Introducción.py/ipybn, 2-BeautifulSoup.py/ipybn) están planteados unos sencillos ejercicios de scrappeo
- Prueba a scrappear la web de rastreator, concretamente la sección relativa a tarifas de luz para particulares. Utiliza los siguientes datos: 3 personas en la vivienda, 80m2 de vivienda y 48001 código postal
 - Si te animas, automatiza el proceso de extracción para las 3 capitales de provincia de Euskadi.
- Escrapea la web de meneame.net. Elige la versión clásica o nueva de la web. De la web extrae el título de la noticia, su texto, la url de la noticia, el número de vistas y comentarios. La solución que he preparado para este ejercicio está desarrollada para la versión nueva de meneame.

USO API

- Prueba a descargar de la página de AEMET, los valores climatológicos extremos de viento registrados en Bizkaia en la estación de Punta Galea. (sección valores climatológicos – extremos registrados)
- Extrae la información, pero esta vez para la estación de Machichaco
- Compara ambos sets de datos extraídos, en qué estación se han registrado las mayores rachas de viento.

ACCESO A BBDD

- Crea una base de datos (dale el nombre que quieras) con los siguientes campos:

Variable	Tipo variable
id	Número entero y clave primaria
Fecha	Dato fecha
Cantidad	Número entero
Artículo	Cadena de texto de tamaño max 50
Descuento	Número decimal
Proveedor	Cadena de texto de tamaño max 100
Lugar_Venta	Cadena de texto de tamaño max 30

-