10/04/2018

**TIPOLOGIA DE DATOS – PRACTICA 1**

*Nombre de los participantes del grupo: María del Carmen Terrats Reviriego*

1.- *Titulo del dataset*: padrón\_sanse.csv

2. *Subtitulo del dataset*. Datos padrón municipio San Sebastián de los Reyes (Madrid)

3. *Imagen:*

4. *Contexto:* Contiene resumen a fecha de obtención de datos de personas empadronadas en el municipio separadas por rango de edad y sexo.

5. *Contenido*:

El fichero padron\_sanse.csv se ha obtenido a partir de la página web del ayuntamiento del municipio.

En la página web tienen colgada de forma pública datos sin personalizar del número de empadronados en el municipio.

Aplicando fórmulas de scrapping y usando el lenguaje de programación python(a través de jupyter notebook) he desarrollado un fichero software que lo que hace es bajar la página principal (<http://www.ssreyes.org/es/portal.do?TR=C&IDR=358>) y recorrer la página buscando las urls donde están colgados los datos. Una vez descargadas las urls se crea el fichero donde escribir y se escribe la cabecera que explica los campos que contendrán el fichero: Fecha, rango edad, número de hombres y número de mujeres.

Por cada url obtenida se descarga y se obtienen los datos necesarios y se van escribiendo en el fichero añadiendo los nuevos registros.

6. Agradecimientos:

Los datos son propiedad el ayuntamiento de San Sebastián de los Reyes pero los publica para uso público e interés general según pone en su página web

[www.ssreyes.org](http://www.ssreyes.org):

“A través del portal del Ayuntamiento de San Sebastián de los Reyes, situado en la dirección URL www.ssreyes.org, se pone a disposición de los visitantes y usuarios la página Web oficial del municipio de San Sebastián de los reyes (en adelante ssreyes.org) con contenidos y servicios de interés general dirigido a las personas que viven en la ciudad o que desean conocerla, a entidades, a profesionales y a las empresas.”

Para realizar la práctica he usado como base el libro:

Richard, Lawson. Web Scraping with Python, Packt Publishing Ltd, 2015: (Chapter\_2\_Scraping\_the\_Data)

*7. Inspiracion:*  Me pareció interesante los datos porque de ellos se puede obtener una evolución demográfica del municipio. Tienen publicados datos desde el 2004 hasta la actualidad, lo cual fue un dato a favor por la profundidad de datos de la que podía disponer.

*8. Licencia. Seleccionad una de estas licencias y decid porque la habéis seleccionado:*

Released Under CC0: Public Domain License

Licencia pública para que se pueda usar de forma libre puesto que creo que puesto que los datos son públicos en sí mismo la extracción de los datos también las puede usar a quien le interese y pueda aprender de ellos.