Práctica 1

1. Contexto

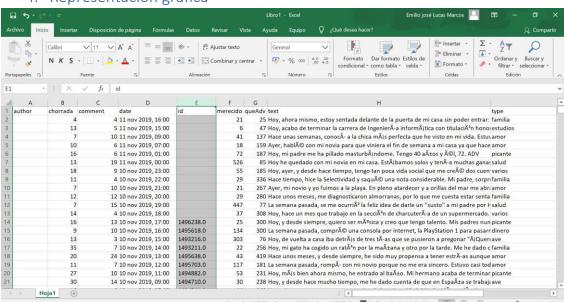
El sitio web elegido es "https://www.ascodevida.com/", se ha escogido este sitio porque sus datos pueden ser utilizados para el estudio de los usuarios de esta página, ya que es una página que funciona como blog para recoger vivencias o comentarios de los usuarios, además no existe archivo robots.txt lo que nos permite acceder a su información.

2. Título del dataset

El título del dataset será: datasetComentariosAdv.csv

3. Descripción del dataset

Se han extraído todos los comentarios publicados en la página, así como toda la información relativa a este que proporciona la página, usuarios, fecha, tipo, etc.



4. Representación gráfica

5. Contenido

- id: identificador único del comentario. (cadena de caracteres)
- autor: autor del comentario. (cadena de caracteres)

- type: tipo de comentario. (cadena de caracteres)
- date: fecha de la publicación del comentario. (fecha)
- text: texto del comentario. (cadena de caracteres)
- comment: cantidad de comentarios recibidos en el comentario. (número)
- queAdv: cantidad de usuarios que están de acuerdo con el comentario. (número)
- merecido: cantidad de personas que están en contra del comentario. (número)
- chorrada: cantidad de personas a las que le da igual el comentario. (número)

Los datos recogidos comprenden el tiempo entre 2014 y la fecha actual, estos se han recogido mediante la librería scrapy de Python, esta crea un proyecto en el cual por medio de spiders haces llamadas a los sitios web y recoges la información.

6. Agradecimientos

La página web de donde se han extraído los datos es una página de entretenimiento a nivel mundial en la que cualquier persona sin estar necesariamente registrada puede dejar un comentario referente a una experiencia de su vida.

Dado el tipo de página web no tiene ningún tipo de estudio anterior.

7. Inspiración

Los datos extraídos en este conjunto de datos son interesantes para el estudio de los usuarios de la página y poder trazar un perfil de estos, saber quienes son los que más lo utilizan y que comentarios son mas propensos a la retro alimentación o que tipo suele ser mejor comprendido.

8. Licencia.

El tipo de licencia sería "Released Under CC BY-NC-SA 4.0 License" ya que el dataset es público y podrá utilizarlo cualquier persona siempre y cuando no sea para lucrarse con él, mencione al creador del dataset y todo los dataset o estudios derivados de este compartan su misma licencia.

Contribuciones	Firma
Investigación previa	EJLM
Redacción de las respuestas	EJLM
Desarrollo código	EJLM