

# 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

|    | A      | B        | C       | D                     | E         | F        | G      | H   |
|----|--------|----------|---------|-----------------------|-----------|----------|--------|---|
|    | author | chorrada | comment | date                  | id        | merecido | queAdv | text  |
| 1  |        |          |         |                       |           |          |        | type  |
| 2  |        | 4        |         | 4 11 nov 2019, 16:00  |           | 21       | 25     | Hoy, ahora mismo, estoy sentada delante de la puerta de mi casa sin poder entrar: familia   |
| 3  |        | 13       |         | 5 11 nov 2019, 15:00  |           | 6        | 47     | Hoy, acabo de terminar la carrera de Ingeniería informática con titulación honoris causa    |
| 4  |        | 7        |         | 10 11 nov 2019, 09:00 |           | 41       | 137    | Hace unas semanas, conocí a la chica más perfecta que he visto en mi vida. Estuvimos        |
| 5  |        | 10       |         | 6 11 nov 2019, 07:00  |           | 18       | 159    | Ayer, hablé con mi novia para que viniera el fin de semana a mi casa ya que hace amor       |
| 6  |        | 16       |         | 6 11 nov 2019, 01:00  |           | 72       | 187    | Hoy, mi padre me ha pillado masturbándose. Tengo 40 años y sé que ADV es picante            |
| 7  |        | 13       |         | 19 11 nov 2019, 00:00 |           | 526      | 85     | Hoy he quedado con mi novia en mi casa. Estábamos solos y tenía muchas ganas de salud       |
| 8  |        | 18       |         | 9 10 nov 2019, 23:00  |           | 55       | 185    | Hoy, ayer, y desde hace tiempo, tengo tan poca vida social que me crean dos cuen varios     |
| 9  |        | 11       |         | 4 10 nov 2019, 22:00  |           | 29       | 336    | Hace tiempo, hice la Selectividad y saqué una nota considerable. Mi padre, sorprende        |
| 10 |        | 7        |         | 10 10 nov 2019, 21:00 |           | 21       | 267    | Ayer, mi novio y yo fuimos a la playa. En pleno atardecer y a orillas del mar me abrió amor |
| 11 |        | 12       |         | 12 10 nov 2019, 20:00 |           | 29       | 280    | Hace unos meses, me diagnosticaron almorranas, por lo que me cuesta estar sentada familia   |
| 12 |        | 7        |         | 15 10 nov 2019, 19:00 |           | 447      | 77     | La semana pasada, se me ocurrió la feliz idea de darle un "susto" a mi padre por la salud   |
| 13 |        | 14       |         | 4 10 nov 2019, 18:00  |           | 37       | 308    | Hoy, hace un mes que trabajo en la sección de charcutería de un supermercado. Varios        |
| 14 |        | 16       |         | 13 10 nov 2019, 17:00 | 1496238.0 | 25       | 300    | Hoy, y desde siempre, quiero ser médica y creo que tengo talento. Mis padres nunca          |
| 15 |        | 9        |         | 10 10 nov 2019, 16:00 | 1495618.0 | 134      | 300    | La semana pasada, compré una consola por internet, la PlayStation 1 para pasar dinero       |
| 16 |        | 13       |         | 3 10 nov 2019, 15:00  | 1493216.0 | 303      | 76     | Hoy, de vuelta a casa iba detrás de tres tías que se pusieron a pregonar "¡Queríame         |
| 17 |        | 35       |         | 7 10 nov 2019, 14:00  | 1493211.0 | 22       | 256    | Hoy, mi gato ha cogido un ratón por la mañana y otro por la tarde. Me he dado cuenta        |
| 18 |        | 20       |         | 24 10 nov 2019, 13:00 | 1495638.0 | 43       | 419    | Hace unos meses, y desde siempre, he sido muy propensa a tener estrías aunque amor          |
| 19 |        | 11       |         | 7 10 nov 2019, 12:00  | 1495703.0 | 117      | 181    | La semana pasada, rompí con mi novio porque no me era sincero. Estuvo casi todo amor        |
| 20 |        | 27       |         | 10 10 nov 2019, 11:00 | 1494882.0 | 53       | 231    | Hoy, más bien ahora mismo, he entrado al baño. Mi hermano acaba de terminar picante         |
| 21 |        | 30       |         | 14 10 nov 2019, 09:00 | 1494710.0 | 30       | 238    | Hoy, y desde hace mucho tiempo, me he dado cuenta de que en España se trabaja ave           |

## 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  |