

**PRA1**

# Tipología Ciclo de Datos

Máster Universitario de Ciencia de Datos

Francisco José Ramírez Vicente

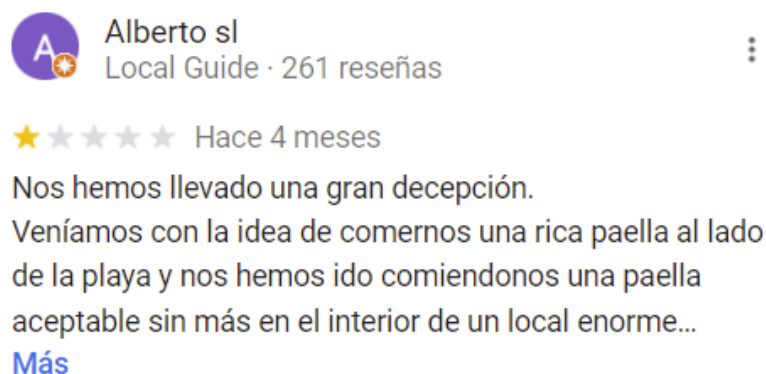
## Pregunta 1 (30% puntuación)

### 1. Contexto

Uno de los mayores problemas que nos encontramos en el ámbito de la IA y más en concreto, dentro del contexto del idioma español, es la falta de datasets específicos. En el procesamiento del lenguaje natural o NLP, tener datasets de textos etiquetados en castellano es una tarea complicada de obtener hoy a la hora de entrenar nuestros modelos. Siempre que busquemos colecciones de datos para algún tipo de proyecto NLP la gran mayoría están en el idioma inglés. Uno de los modelos más utilizados hoy es el análisis de sentimientos y es sobre este tipo de datasets sobre el cual está basado este proyecto.

Existen en Internet un tipo de páginas web que pueden ayudarnos a aportar algo información en idioma español a estos datasets: las páginas con puntuaciones y críticas. El usuario de este tipo de páginas (en el proyecto he seleccionado **Google Reviews** y **TripAdvisor**), en función del tipo de servicio, el usuario o cliente puede aportar una crítica o comentario sobre cómo ha comido (en caso de restaurantes), cómo estaba la habitación (hoteles), etc.

Esto es interesante porque el usuario del servicio ya nos está ayudando a etiquetar la información que está aportando además de puntuarla. Finalmente tendremos por un lado el texto con el comentario (texto descriptivo) y por otro la puntuación (valoración entre 0 y 5 en caso de Google Maps y de 0 a 50 en el caso de TripAdvisor).



*Ejemplo de un comentario negativo en [Google Maps](#)*

Es aquí donde el Webscraping es una gran herramienta que nos permitirá crear nuestro propio dataset para el análisis de sentimientos utilizando esta información. Al tener el contexto (texto de la review) y además la puntuación (las estrellas o valores del 0 a 5) es el mismo usuario el que nos está dando la etiqueta de ese texto, el cual puede ser positivo, negativo o incluso neutral. Aplicando técnicas de Webscraping podemos analizar diferentes páginas web y extraer dicha información y añadirla a un fichero CSV, con la etiqueta de positivo, negativo o neutral.

Las dos páginas webs seleccionadas para este proyecto han sido:

- [Google Maps](#)
- [TripAdvisor.es](#)

Para las reviews de Google Maps, se han probado con Hoteles y Restaurantes sin problemas, aunque es probable que funcione con el resto de servicios. En el caso de TripAdvisor sólo por Restaurantes, por motivos de tiempos de entrega.

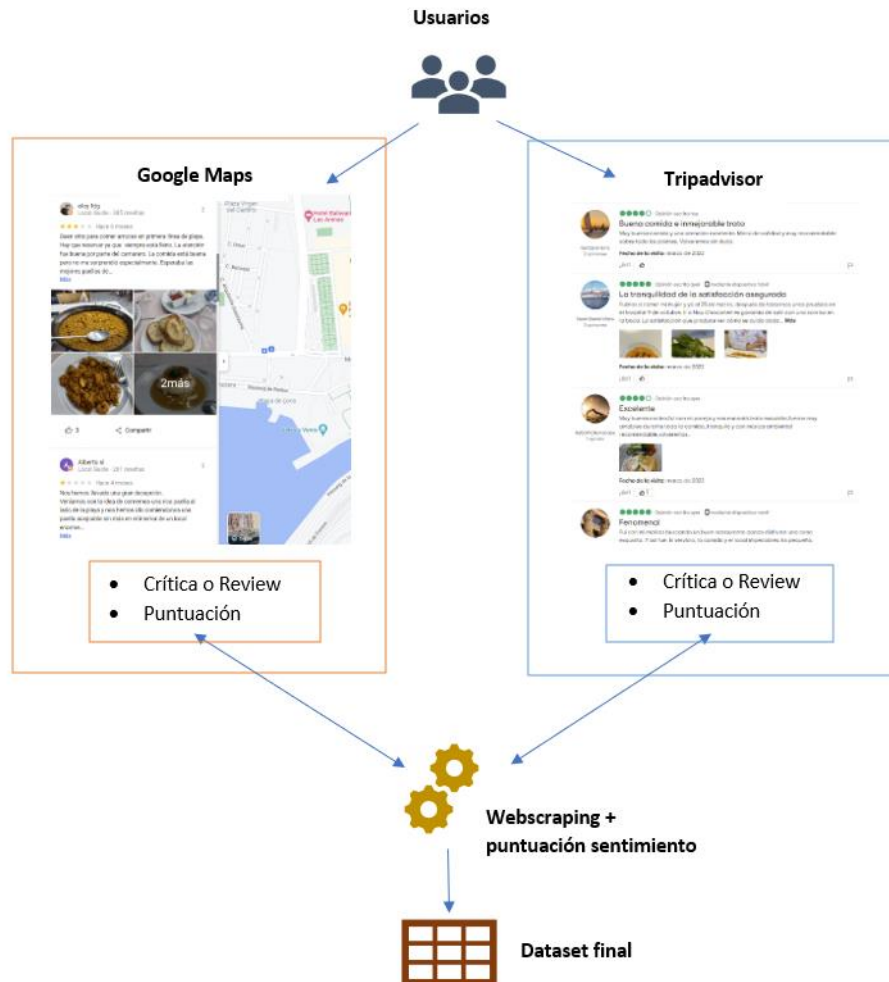
## 2. Título

*Generación de Datasets para el análisis de sentimientos basado en críticas de servicios*

## 3. Descripción del dataset

El dataset obtenido se centra en recopilar las opiniones y calificación de un servicio concreto en alguna de las plataformas que existen para este fin. Este almacenará la información basada en la crítica realizada y la puntuación. Con la puntuación también se calculará el tipo de sentimiento. Se podría almacenar más información (como fecha, nombre del usuario que ha realizado la crítica) pero no se ha considerado relevante y también por motivos de privacidad. Por otro lado, se ha intentado mantener correctamente la codificación para evitar caracteres extraños y el dataset está listo para ser procesado. El fichero resultante es un CSV con delimitador “;”

#### 4. Representación gráfica



#### 5. Contenido

A pesar de haber explorado las webs y entender cómo realizar su “scraping” a varios niveles, finalmente sólo necesitaremos 3 datos, dos extraídos y uno generado directamente por el programa. Para facilitar su análisis se ha creado un dataset (CSV) para cada aplicación, Google Maps y Tripadvisor (se podría modificar de manera sencilla el programa para que los unificara en uno solo). Los campos que se almacenan son:

1. “*puntuación*”. Valor numérico con el cual el usuario ha puntuado el servicio.  
Puede estar entre 0 y 5 (Google Maps) o entre 0 y 50 (Tripadvisor)

2. “review”. Texto con la crítica. Sólo se obtiene el texto resumido, ya que es suficiente para dar contexto del sentimiento.
3. “sentimiento”. Basándose en el valor obtenido en “puntuación”, creamos una nueva columna llamada “sentimiento” la cual calcula de manera sencilla el tipo de valoración, por ejemplo si está entre 0 y 2 será negativa, entre 2 y 3 neutral y entre 3 y 5 positiva. Estos baremos se pueden ajustar en función de nuestras necesidades.

puntuación	review	sentimiento
40	Dispone de espacios acogedores con unas mamparas que hacen la estancia sea	positivo
40	Tanto la comida como los pintxos son buenos, los gintonics los preparan de mar	positivo
50	Hemos comido estupendamente en la terraza con unos amigos y luego hemos tr	positivo
50	Excelente comida en una de las mejores terrazas de Donosti. Atención de 10. Ur	positivo
40	Para comer o cenar algo esta bien, lo malo es que las mesas estas un poco cerc	positivo
40	un sitio muy agradable para tomar un café unas cañas picar algo o incluso come	positivo
50	Conoci este sitio a través de internet y me pareció una opción interesante para	positivo
50	Tras pedirme un café y le pedí consejo al camarero para comer algo para almor	positivo
50	Sitio agradable y familiar . El trato por el personal excelente y cercano, muy am	positivo
50	El bully nunca falla como escenario para nuestras cenas de amigas del equipo de	positivo

*Captura de parte del dataset extraído para el caso de Tripadvisor (las “reviews” están recortadas)*

## 6. Agradecimientos

El programa webscraping creado se ha intentado realizar de la forma más simple y con menos impacto posible a las páginas destino. Por otro lado, el agente o crawler desplegado se conecta en modo incógnito y además se identifica como un “*AcademicCrawler*”. Para de esta forma, ofrecer información al sitio web nuestras intenciones con el webscraping.

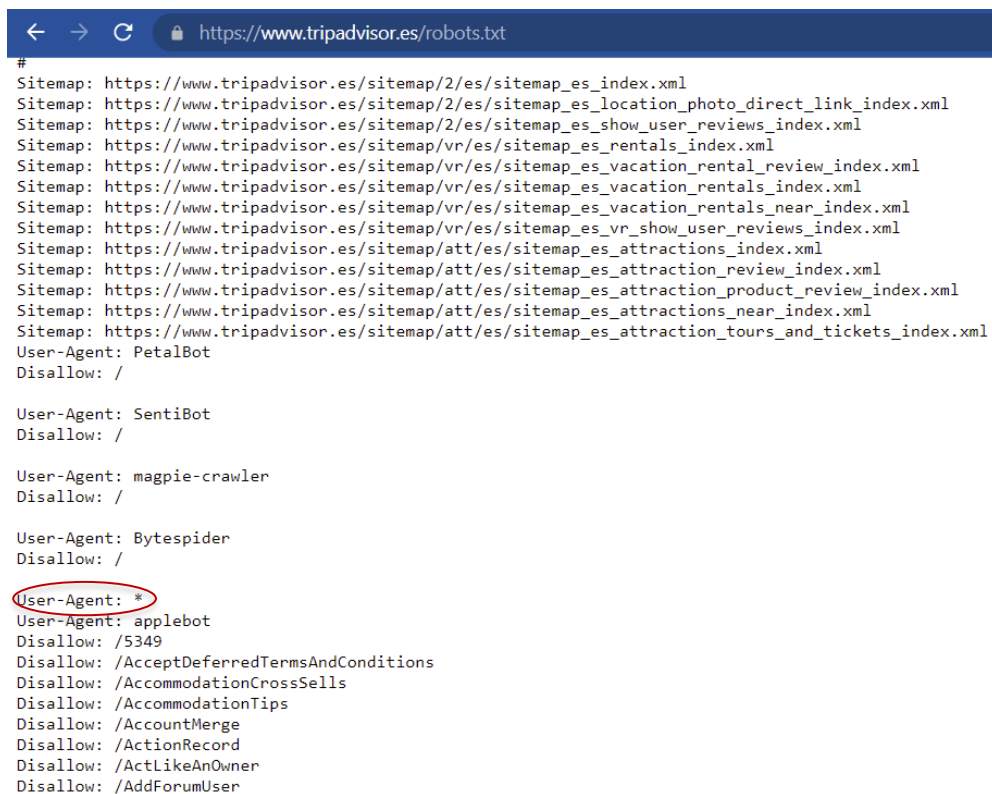
```
option = webdriver.ChromeOptions()
option.add_argument("--incognito")
option.add_argument("user-agent=AcademicCrawler")
driver = webdriver.Chrome(ChromeDriverManager().install(),options=option)
```

*Configuración modo Incógnito y AcademicCrawler del programa*

El programa se ha diseñado para realizar el mínimo posible de llamadas a la página web para evitar afectar de ningún modo a su correcto funcionamiento. Por otro lado, el contenido extraído, tanto de **TripAdvisor** como de **Google Maps** no es Copyright y no incumple los términos de uso de ambas compañías. La información que estamos extrayendo, además de

ser anónimas, son reviews o comentarios escritos por otros usuarios, lo que convierte dicha información en pública. El objetivo del trabajo al ser educativo es otro factor que nos permite realizar el scraping. No se está extrayendo propiedad intelectual y no estamos accediendo a datos personales.

El fichero **robots.txt** de Tripadvisor nos muestra que en principio el acceso desde un crawler está permitido:



```
#
Sitemap: https://www.tripadvisor.es/sitemap/2/es/sitemap_es_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/2/es/sitemap_es_location_photo_direct_link_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/2/es/sitemap_es_show_user_reviews_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/vr/es/sitemap_es_rentals_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/vr/es/sitemap_es_vacation_rental_review_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/vr/es/sitemap_es_vacation_rentals_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/vr/es/sitemap_es_vacation_rentals_near_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/vr/es/sitemap_es_vr_show_user_reviews_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/att/es/sitemap_es_attractions_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/att/es/sitemap_es_attraction_review_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/att/es/sitemap_es_attraction_product_review_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/att/es/sitemap_es_attractions_near_index.xml
Sitemap: https://www.tripadvisor.es/sitemap/att/es/sitemap_es_attraction_tours_and_tickets_index.xml
User-Agent: PetalBot
Disallow: /

User-Agent: SentiBot
Disallow: /

User-Agent: magpie-crawler
Disallow: /

User-Agent: Bytespider
Disallow: /

User-Agent: *
User-Agent: applebot
Disallow: /5349
Disallow: /AcceptDeferredTermsAndConditions
Disallow: /AccommodationCrossSells
Disallow: /AccommodationTips
Disallow: /AccountMerge
Disallow: /ActionRecord
Disallow: /ActLikeAnOwner
Disallow: /AddForumUser
```

*Captura del robots.txt de Tripadvisor donde se muestra que se permite realizar webscraping*

Respecto a **Google Maps**, en sus términos de servicio indica que hacer scraping de sus datos es legal siempre y cuando los datos recolectados no se usen para fines ilícitos. De todas formas, para evitar un posible bloqueo por parte de Google, el proyecto toma pocas muestras de información.

<https://www.google.com/intl/es/permissions/geoguidelines/>

## 7. Inspiración

He trabajado en algunos proyectos basados en el análisis de sentimientos y siempre he echado en falta algún dataset en castellano que nos pueda permitir entrenar algún tipo de modelo basando en este lenguaje. El análisis de sentimiento es realmente útil para crear aplicaciones, tanto para interactuar con el usuario (bots) como parte de una arquitectura que se dedique a detectar fakenews.

Por lo tanto, con este trabajo se pretende encontrar una solución a la pregunta ¿es posible encontrar o generar un dataset de análisis de sentimientos en castellano? Añadiendo nuevas páginas web donde se realicen comentarios de cualquier tipo de servicio, sería posible crear un proyecto de dataset en español para posteriormente publicar con licencia libre y así tener una base sobre la cual crear modelos de IA basados en NLP y el análisis de sentimientos.

## 8. Licencia

La licencia sobre la cual podemos publicar este dataset podría ser:

***CC0: Public Domain License.***

Como he comentado antes, el principal objetivo de la creación de este dataset es ofrecer a todo el mundo, sin ningún tipo de interés privado o comercial, la creación de un dataset de análisis de sentimientos. Por lo tanto, pienso que esta licencia, la cual no tiene ningún tipo de derechos de autoría se dedique íntegramente al dominio público. De esta forma podemos aportar nuestros datasets como base de conocimiento para hacer mayor la comunidad y los progresos de la IA.

Por lo tanto, esta licencia permite modificar, distribuir e interpretar la obra, incluso para propósitos comerciales sin necesidad de pedir permiso.

<https://creativecommons.org/publicdomain/zero/1.0/deed.es>

## 9. Código

Se ha publicado en el github del autor el código fuente. El programa está desarrollado en Python 3.9.12 y requiere Windows 10/11 para funcionar.

## 10. Dataset

Enlace DOI Zenodo:

<https://doi.org/10.5281/zenodo.6425686>

## 11. Vídeo

<https://drive.google.com/file/d/1fPPv4DSun3uyvNGqahz0jONfz1oysR9v/view?usp=sharing>



## APÉNDICE 1. Código fuente

```
# Webscraping
# Análisis comentarios Google Maps y Tripadvisor
# Dataset para NLP análisis de sentimientos
# PRA1 Tipología Ciclo de Datos
# Francisco José Ramírez

from importlib.resources import path
import os
import csv
from selenium import webdriver
import time
from webdriver_manager.chrome import ChromeDriverManager
import urllib
from bs4 import BeautifulSoup
import urllib.request

# Definimos los XPATH principales de los elementos a recuperar la información
# Se eligen dos páginas web donde aparecen comentarios positivos o negativos

# XPATHs TripAdvisor Restaurantes
main_xpath = "//span[@class='taLnk ulBlueLinks']"
rest_find_xpath = ".//div[@class='review-container']"
rest_titulo_xpath = ".//span[@class='noQuotes']"
rest_fecha_xpath = ".//span[contains(@class, 'ratingDate')]"
rest_puntuacion_xpath = ".//span[contains(@class, 'ui_bubble_rating bubble')]"
rest_comentario_xpath = ".//p[@class='partial_entry']"

# XPATHs Google Reviews Restaurantes (puede servir para otros)
google_find_xpath = "//div/div[@data-review-id]"
google_puntuacion_class = 'ODSEW-ShBeI-H1e3jb'
google_comentario_class = 'ODSEW-ShBeI-text'

class Dataset_Comentarios(object):

    def __init__(self, web_objetivo):
        self.objetivo=web_objetivo

    def get_robots(self, miurl):
        response = urllib.request.urlopen(urllib.request.Request(miurl,headers={'User-Agent':
'Mozilla'}))
        soup = BeautifulSoup(response,'html.parser',
from_encoding=response.info().get_param('charset'))
        return soup
```

```
def boton_privacidad_trip(self):
    try:
        time.sleep(3)
        boton_aceptar = driver.find_element_by_id("onetrust-accept-btn-handler")
        driver.execute_script("arguments[0].click();", boton_aceptar)
    except:
        return

def boton_privacidad_google(self):
    try:
        time.sleep(3)
        boton_aceptar =
driver.find_element_by_xpath("/html/body/div[2]/div[3]/form/input[11]")
        driver.execute_script("arguments[0].click();", boton_aceptar)
    except:
        return

def puntuacion_sentimiento_google(self, puntos):
    # En función de la puntuación determinaremos los siguientes varemos de sentimientos:
    # 0 a 2 Negativo
    # 2 a 3 Neutral
    # 3 a 5 Bueno
    sentimiento = ""
    if puntos in range(0,2):
        sentimiento = "negativo"
    elif puntos in range(2,3):
        sentimiento = "neutral"
    elif puntos in range(3,6):
        sentimiento = "positivo"
    return(sentimiento)

def puntuacion_sentimiento_trip(self, puntos):
    # En función de la puntuación determinaremos los siguientes varemos de sentimientos:
    # 0 a 20 Negativo
    # 20 a 30 Neutral
    # 30 a 50 Bueno
    sentimiento = ""
    if puntos in range(0,20):
        sentimiento = "negativo"
    elif puntos in range(20,30):
        sentimiento = "neutral"
    elif puntos in range(30,60):
        sentimiento = "positivo"
    return(sentimiento)
```

```

def fichero_comentarios(self, nombre):
    path_actual = os.getcwd()
    os.chdir(path_actual)
    # fichero csv para almacenar los comentarios

    fichero_csv = open(nombre, 'w', encoding="utf-8-sig")
    fichero_output = csv.writer(fichero_csv, delimiter=';')
    # Añado cabecera al csv
    fichero_output.writerow(["puntuación", "review", "sentimiento"])
    return(fichero_output)

def comentarios_tripadvisor_restaurantes(self, npaginas, fichero_salida):
    # Recorremos las páginas definidas
    for npag in range(0, npaginas):
        # abrimos los comentarios
        # Damos un tiempo de espera para que se abra
        time.sleep(4)
        #driver.find_element_by_xpath(main_xpath).click()
        container_review = driver.find_elements_by_xpath(rest_find_xpath)
        # Ahora recorremos todos los comentarios
        for comentarios in range(len(container_review)):
            puntuacion =
container_review[comentarios].find_element_by_xpath(rest_puntuacion_xpath).get_attribute("class").
split("_")[3]
            puntos_sentimiento = self.puntuacion_sentimiento_trip(int(puntuacion))
            comentario =
container_review[comentarios].find_element_by_xpath(rest_comentario_xpath).text.replace("\n", " ")
            # Almacenamos la salida en el fichero asignado
            fichero_salida.writerow([puntuacion, comentario, puntos_sentimiento])
        # Click Siguiente página hasta fin bucle npaginas:
        button = driver.find_element_by_xpath('//*[a[@class="nav next ui_button primary"]')
        driver.execute_script("arguments[0].click();", button)

def comentarios_google_reviews(self, fichero_salida):
    # Expandimos los resultados de las reviews
    time.sleep(3)
    # Hacemos click en el botón más reseñas para mostrar más comentarios y almacenarlos
    driver.find_element_by_xpath('//button[starts-with(@aria-label,"Más reseñas"]').click()
    # Recorremos las páginas definidas
    time.sleep(4)
    container_review = driver.find_elements_by_xpath(google_find_xpath)
    for comentarios in range(len(container_review)):

```

```

        puntuacion =
container_review[comentarios].find_element_by_class_name(google_puntuacion_class).get_attribute("aria-label").split(" ")[1]
        comentario =
container_review[comentarios].find_element_by_class_name(google_comentario_class).text
        puntos_sentimiento= self.puntuacion_sentimiento_google(int(puntuacion))
        fichero_salida.writerow([puntuacion, comentario, puntos_sentimiento])
        print("Comentarios extraidos:", len(container_review))

if __name__ == "__main__":

    # Parámetros generales
    # Ajustamos las opciones de acceso para evitar problemas legales y de bloqueo
    option = webdriver.ChromeOptions()
    option.add_argument("--incognito")
    option.add_argument("user-agent=AcademicCrawler")
    driver = webdriver.Chrome(ChromeDriverManager().install(),options=option)
    # Fichero para almacenar todos los comentarios
    nombrefile_tripadvisor="dataset_tripadvisor.csv"

    # Visualizamos el fichero robots.txt para su análisis
    #robots = tripadvisor.get_robots("https://www.tripadvisor.es/robots.txt")
    #print(robots)

    print("Google Maps Reviews")

    # Google Maps Reviews
    #####
    nombrefile_googlemaps="dataset_googlemaps.csv"
    website="https://www.google.com/maps/"
    url_googlemaps = "https://www.google.com/maps/place/Restaurante+La+Pepica/@39.4634132,-
0.3252193,17z/data=!3m1!4b1!4m5!3m4!1s0xd6048422f04f9ef:0xf8470a431cc0fe26!8m2!3d39.4634132!4d-
0.3230306"
    googlemaps = Dataset_Comentarios(website)
    driver.get(url_googlemaps)
    googlemaps.boton_privacidad_google()
    fichero_resultados=googlemaps.fichero_comentarios(nombrefile_googlemaps)
    googlemaps.comentarios_google_reviews(fichero_resultados)

    print("Tripadvisor")

    # TripAdvisor Reviews
    #####
    # Número de páginas a comprobar los comentarios TripAdvisor
    paginas = 2

```

```

nombrefile_tripadvisor="dataset_tripadvisor.csv"
# URL TripAdvisor a comprobar
website="https://tripadvisor.es/"
url_tripadvisor = "https://www.tripadvisor.es/Restaurant_Review-g187457-d1026118-Reviews-
Bully-San_Sebastian_Donostia_Province_of_Guipuzcoa_Basque_Country.html"
tripadvisor = Dataset_Comentarios(website)
driver.get(url_tripadvisor)
# Aceptamos las condiciones del mensaje de cookies y privacidad
tripadvisor.boton_privacidad_trip()
fichero_resultados=tripadvisor.fichero_comentarios(nombrefile_tripadvisor)
tripadvisor.comentarios_tripadvisor_restaurantes(paginas,fichero_resultados)

driver.quit()

print("Fin webscraping")

```

## APÉNDICE 2. Algunos ejemplos de la extracción de los valores XPATH.

### Tripadvisor

Fotos COVID-19 Descripción general Opiniones Preguntas y respuestas

ESCAPADAS DE PRIMAVERA  
¡Disfruta de hasta un  
**20%**  
de descuento y cancelación gratuita!

[RESERVA YA](#)

### Opiniones (2.393)

[Escribe una opinión](#)

**Puntuación de viajeros**

☐ Excelente 117  
☐ Muy bueno 12  
☐ Normal 5  
☐ Malo 0  
☐ Pésimo 1

**Tipo de viajero**

☐ Familias  
☐ En pareja  
☐ En solitario  
☐ Negocios  
☐ Amigos

**Época del año**

☐ Mar-may  
☐ Jun-ago  
☐ Sep-nov  
☐ Dic-feb

**Idioma**

☐ Todos los idiomas  
☒ Español (135)  
☐ Inglés (1.982)  
☐ Italiano (96)  
[Más idiomas](#)

Lee lo que dicen los viajeros:

Opinión escrita el 5 de noviembre de 2021 [mediante dispositivo móvil](#)

**Excellent**

Excelente la comida y la atención de Fabio un 10... he is the best!!! El ambiente y el servicio son muy buenos, es para disfrutar de buena música mientras comes [Más](#)

Fecha de la visita: noviembre de 2021

[¿Útil?](#) [1](#)

Manager, Other en Hard Rock Cafe Piccadilly Circus, respondió a esta opinión

DevTools is now available in Spanish!

Always match Chrome's language Switch DevTools to Spanish Don't show again

Elements Console Sources

21 32 35

```

<div class="listContainer hide-more-mobile" data-
contextchoice="DETAIL">
  <div>
    <div class="info hidden" data-expansionsoftgate="fals
e"></div>
    <div class="review-container" data-reviewid="817555371"
data-collapsed="true" data-deferred="false"> == $0
    <!--trKR:1-->
    <!--trKR:817555371-->
    <div class="prw_rup prw_reviews_review_resp" data-
prwridget-name="reviews_review_resp" data-prwridget-
init="handlers">
      <div id="review_817555371" class="reviewSelector"
data-reviewid="817555371">
        <div class="rev_wrap ui_columns is-multiline">
          <div class="ui_column is-2 memberInfoColumn">...
        </div>
        <div class="ui_column is-9">
          <span class="ui_bubble_rating bubble_50">...
        </span>
          <span class="ratingDate" title="5 de noviembr
de 2021">Opinión escrita el 5 de noviembre de
2021 </span>
          <span class="viaMobile" onclick="widgetEvCall
("handlers.viaMobileClick", event, this)">...
        </div>
      </div>
    </div>
  </div>
  ... _location_reviews_list_resp div.listContainer.hide-more-mobile div div.review-container
  
```

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessib

Filter

element.style {

media (min-width: 768px)

DIV.ppr\_rup.ppr\_priv\_location\_reviews\_list\_resp Restaurant\_nd.html:1

.review-container {

border-top: 1px solid #e0e0e0;

blockquote, dd, div, dl, dt, fieldset, form, h1, h2, Restaurant\_land.html

h3, h4, h5, h6, input, legend, li, ol, p, pre, td, textarea, th, ul {

margin: 0;

padding: 0;

ANe\_04.139e23ee.css

margin: 0;

padding: 0;

Fotos COVID-19 Descripción general Opiniones Preguntas y respuestas

Puntuación de viajeros

☐ Excelente 37  
☐ Muy bueno 1  
☐ Normal 1  
☐ Malo 0  
☐ Pésimo 0

Tipo de viajero

☐ Familias  
☐ En pareja  
☐ En solitario  
☐ Negocios  
☐ Amigos

Época del año

☐ Mar-may  
☐ Jun-ago  
☐ Sep-nov  
☐ Dic-feb

Idioma

☐ Todos los idiomas  
☒ Español (39)  
☐ Inglés (332)  
☐ Francés (10)  
[Más idiomas](#)

Lee lo que dicen los viajeros:

Opinión escrita el 24 de julio de 2020 [mediante dispositivo móvil](#)

**Excelente**

Comida riquísima. Me encantó la comida y es un sitio muy acogedor. Los camareros muy simpáticos y amables. El menu muy variado y completo. Recomendando cien por cien el restaurante.

Fecha de la visita: julio de 2020

[¿Útil?](#) [1](#)

DevTools is now available in Spanish!

Always match Chrome's language Switch DevTools to Spanish Don't show again

Elements Console Sources

14 8 25

```

contextchoice="DETAIL">
  <div>
    <div class="info hidden" data-expansionsoftgate="fals
e"></div>
    <div class="review-container" data-reviewid="76134198
data-collapsed="false" data-deferred="false"> == $0
    <!--trKR:1-->
    <!--trKR:761341982-->
    <div class="prw_rup prw_reviews_review_resp" data-
prwridget-name="reviews_review_resp" data-prwridget-
init="handlers">
      <div id="review_761341982" class="reviewSelector"
data-reviewid="761341982">
        <div class="rev_wrap ui_columns is-multiline">
          <div class="ui_column is-2 memberInfoColumn">...
        </div>
        <div class="ui_column is-9">
          <span class="ui_bubble_rating bubble_50">...
        </span>
          <span class="ratingDate" title="5 de noviembr
de 2021">Opinión escrita el 5 de noviembre de
2021 </span>
          <span class="viaMobile" onclick="widgetEvCall
("handlers.viaMobileClick", event, this)">...
        </div>
      </div>
    </div>
  </div>
  ... _location_reviews_list_resp div.listContainer.hide-more-mobile div div.review-container
  
```

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessib

Filter

element.style {

media (min-width: 768px)

DIV.ppr\_rup.ppr\_priv\_location\_reviews\_list\_resp Restaurant\_nd.html:1

.review-container {

border-top: 1px solid #e0e0e0;

blockquote, dd, div, dl, dt, fieldset, form, h1, h2, Restaurant\_land.html

h3, h4, h5, h6, input, legend, li, ol, p, pre, td, textarea, th, ul {

margin: 0;

padding: 0;

ANe\_04.139e23ee.css

margin: 0;

padding: 0;

Fotos COVID-19 Descripción general Opiniones Preguntas y respuestas

ESCAPADAS DE PRIMAVERA  
¡Disfruta de hasta un  
**20%**  
de dto. y cancelación gratuita!

RESERVA YA



## Opiniones (2.393)

Escribe una opinión

### Puntuación de viajeros



### Tipo de viajero

- ☐ Familias
- ☐ En pareja
- ☐ En solitario
- ☐ Negocios
- ☐ Amigos

### Época del año

- ☐ Mar-may
- ☐ Jun-ago
- ☐ Sep-nov
- ☐ Dic-feb

### Idioma

- ☐ Todos los idiomas
- ☒ Español (135)
- ☐ Inglés (1.982)
- ☐ Italiano (96)

### Lee lo que dicen los viajeros:

Buscar opiniones



MoniaBelda  
5 opiniones

★★★★★ Opinión escrita el 5 de noviembre de 2021

mediante dispositivo móvil

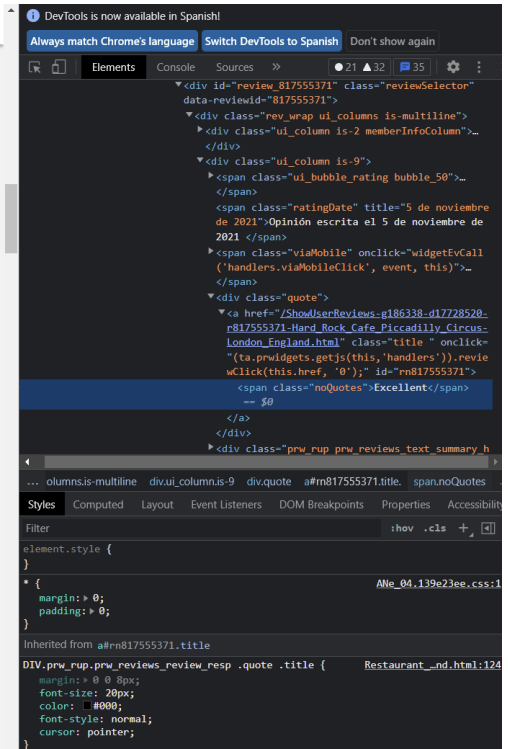
Excellent

Excelente la comida y la atención de Fabio un 10... he is the best!!! El ambiente y el servicio son muy buenos, es para disfrutar de buena música mientras comes **Más**



Fecha de la visita: noviembre de 2021

¿Útil?



Fotos COVID-19 Descripción general Opiniones Preguntas y respuestas

ESCAPADAS DE PRIMAVERA  
¡Regístrate una noche de relax o  
una escapada al sol!

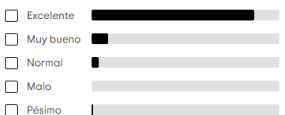
RESERVA YA



## Opiniones (2.393)

Escribe una opinión

### Puntuación de viajeros



### Tipo de viajero

- ☐ Familias
- ☐ En pareja
- ☐ En solitario
- ☐ Negocios
- ☐ Amigos

### Época del año

- ☐ Mar-may
- ☐ Jun-ago
- ☐ Sep-nov
- ☐ Dic-feb

### Idioma

- ☐ Todos los idiomas
- ☒ Español (135)
- ☐ Inglés (1.982)
- ☐ Italiano (96)

### Lee lo que dicen los viajeros:

Buscar opiniones



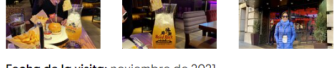
MoniaBelda  
5 opiniones

★★★★★ Opinión escrita el 5 de noviembre de 2021

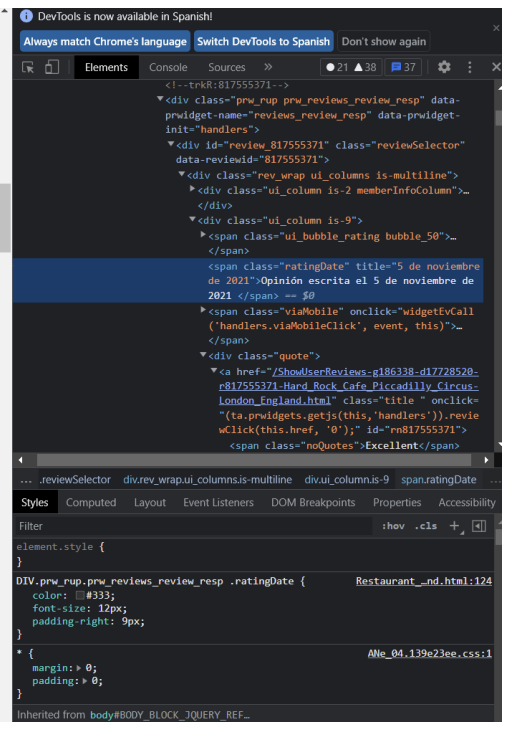
mediante dispositivo móvil

Excellent

Excelente la comida y la atención de Fabio un 10... he is the best!!! El ambiente y el servicio son muy buenos, es para disfrutar de buena música mientras comes **Más**



Fecha de la visita: noviembre de 2021





Fotos
COVID-19
Descripción general
Opiniones
Preguntas y respuestas

**ESCAPADAS DE PRIMAVERA**  
¡Disfruta de hasta un  
**20%**  
de dto. y cancelación gratuita!

> RESERVA YA

**NH**  
HOTELS

## Opiniones (2.393)

Escribe una opinión

Puntuación de viajeros	Tipo de viajero	Época del año	Idioma
<input type="checkbox"/> Excelente 117	<input type="checkbox"/> Familias	<input type="checkbox"/> Mar-may	<input checked="" type="radio"/> Todos los idiomas
<input type="checkbox"/> Muy buena 12	<input type="checkbox"/> En pareja	<input type="checkbox"/> Jun-jago	<input type="radio"/> Español (135)
<input type="checkbox"/> Normal 5	<input type="checkbox"/> En solitario	<input type="checkbox"/> Sep-nov	<input type="radio"/> Inglés (1.982)
<input type="checkbox"/> Malo 0	<input type="checkbox"/> Negocios	<input type="checkbox"/> Dic-feb	<input type="radio"/> Italiano (96)
<input type="checkbox"/> Pésimo 1	<input type="checkbox"/> Amigos		Más idiomas ▼

Lee lo que dicen los viajeros:

Monialbelda  
5 opiniones

● ● ● ● ● Opinión escrita el 5 de noviembre de 2021 ☐ mediante dispositivo móvil

### Excellent

Excelente la comida y la atención de Fabio un 10... he is the best!!! El ambiente y el servicio son muy buenos, es para disfrutar de buena música mientras comes **Más**

**Fecha de la visita:** noviembre de 2021

¿Utilizaste?

The screenshot shows the Chrome DevTools interface. At the top, the address bar displays 'devTools is now available in Spanish!'. Below it, the 'Always match Chrome's language' and 'Switch DevTools to Spanish' buttons are visible, along with a 'Don't show again' link. The 'Elements' panel is active, showing the DOM tree on the left and the rendered HTML on the right. The DOM tree shows a hierarchy starting from 'div.prw\_rup.prw\_reviews\_review\_resp' (data-review-id='817555371') down to 'div.ui\_bubble\_rating' (class='ui\_bubble\_rating\_bubble\_50'). The 'Elements' panel shows the rendered HTML for the selected element, which is a rating bubble. It includes a rating date '5 de noviembre de 2021', a user profile link 'Opinión escrita el 5 de noviembre de 2021', and a review text 'viaMobile' with a 'ShowUserReviews' link. The review text is truncated and includes a 'prwwidgets.getjs' call. The bottom of the screenshot shows the 'Styles' panel with the 'font-size: 16px;' and 'display: inline-block;' styles applied to the selected element.

Fotos

COVID-19

Descripción general

Opiniones

Preguntas y respuestas

ESCAPADAS DE PRIMAVERA

iDisfruta de hasta un

20%

de dño. y cancelación gratuita!

> RESERVA YA

Opiniones (2.393)

Escribe una opinión

Puntuación de viajeros

☐ Excelente
 

117

☐ Muy bueno
 

12

☐ Normal
 

5

☐ Malo
 

0

☐ Pésimo
 

1

Tipo de viajero

☐ Familias

☐ En pareja

☐ En solitario

☐ Negocios

☐ Amigos

Época del año

☐ Mar-may

☐ Jun-ago

☐ Sep-nov

☐ Dic-feb

Idioma

☐ Todos los idiomas

☒ Español (135)

☐ Inglés (1.982)

☐ Italiano (96)

Más idiomas

Lee lo que dicen los viajeros:

Q

Buscar opiniones

Monialbelda

5 opiniones

●●●●●

Opinión escrita el 5 de noviembre de 2021

mediante dispositivo móvil

Excellent

Excelente la comida y la atención de Fabio un 10... he is the best!!!! El ambiente y el servicio son muy buenos, es para disfrutar de buena música mientras comes

Más

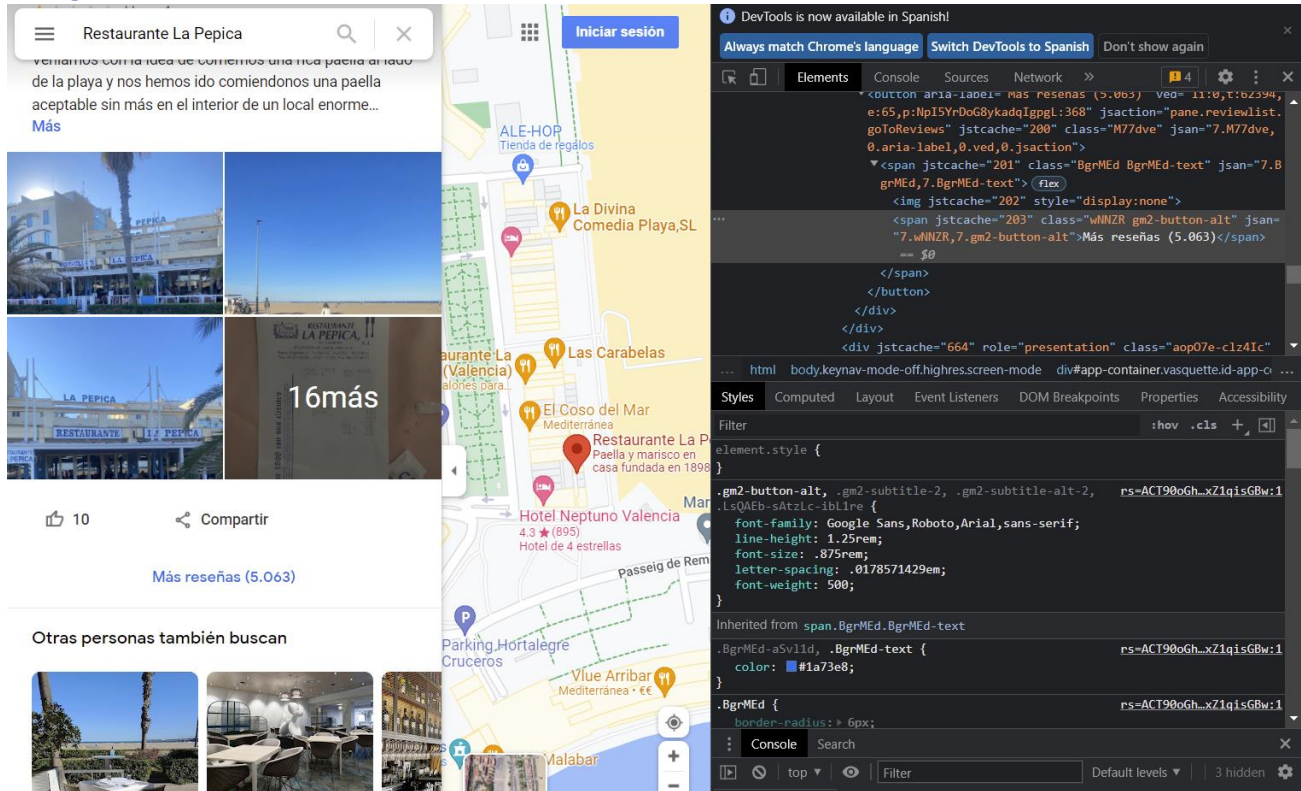
Fecha de la visita:

noviembre de 2021

[illegible]



## Google



The screenshot shows a Google search for "Restaurante La Pepica". The search results include a snippet about paella, a gallery of images, and a map of the restaurant's location in Valencia. A "Botón Más Reseñas" (More Reviews Button) is highlighted. The DevTools overlay shows the HTML structure of the button, which is a `<button>` element with attributes like `aria-label="Más reseñas (5.063)"` and `ved="110,t:02394,e:65,p:NpISYrDoG8YkadqIggL:368"`. The Styles panel shows the default styling for the button, including font-family, line-height, font-size, letter-spacing, and font-weight.

## Botón Más Reseñas

### APÉNDICE 3. Enlaces consultados

- <https://medium.com/codex/web-scraping-with-selenium-in-python-832cf4b827a4>
- <https://www.analyticsvidhya.com/blog/2021/11/web-scraping-a-news-article-and-performing-sentiment-analysis-using-nlp/>
- <https://medium.com/@xyng17/scraping-twitter-and-sentiment-analysis-using-python-c5a44b9288ab>
- <https://towardsdatascience.com/customer-reviews-identify-your-strengths-and-weaknesses-with-the-help-of-web-scraping-data-b87a3636ef55>
- <https://towardsdatascience.com/scraping-google-maps-reviews-in-python-2b153c655fc2>
- <https://curiously.com/posts/create-dataset-for-sentiment-analysis-by-scraping-google-play-app-reviews-using-python/>

### APÉNDICE 4. Autoría

Este trabajo está íntegramente creado y desarrollado por el alumno Francisco José Ramírez Vicente (FJRV).

Contribuciones	Firma
Investigación previa	FJRV
Redacción de las respuestas	FJRV
Desarrollo del código	FJRV



Francisco José Ramírez Vicente