

Realizado por:

Marta Sanjuán de Caso

Práctica 1

<https://github.com/martasanjuan/TourismInfoScraper>

1. **Título:** Actividades turísticas
2. **Subtítulo:** Conjunto de actividades turísticas que se llevan a cabo en la ciudad de Sevilla por una de las más importantes empresas del sector turístico de actividades y tours, civitatis.com
3. **Imagen:**



Derechos de autor: <https://www.voyagetips.com/wp-content/uploads/2018/01/things-to-do-in-Seville.jpg>

4. El contexto de los datos obtenidos se basa en los diferentes servicios turísticos que ofrece el portail civitatis de la ciudad de Sevilla.

5. Los campos que se incluyen son:
- Título del servicio turístico
 - Idioma en que se ofrece el servicio turístico
 - Duración del servicio turístico
 - Puntuación que le dan sus usuarios

No existe un periodo de tiempo concreto, sino que se han recogido en una fecha concreta. No existe la posibilidad de recogida de servicios turísticos según un periodo concreto.

6. El propietario del conjunto de datos es el propietario del portal de servicios turístico www.civitatis.com

No he encontrado ninguna referencia sobre investigaciones o análisis que se hayan realizado sobre esta web.

7. Este conjunto de datos es interesante fundamentalmente para la competencia del este tipo de actividad, ya que, gracias a los datos obtenidos, podemos hacer un seguimiento de qué actividades se ofrecen por ciudad y cuáles son las más demandadas por los usuarios del mismo portal.

8. Le podría asignar una licencia pública, en este caso la primera licencia denominada Release under CC0: Public Domain License, debido a que no se trata de información relativa a comentario ni información de índole personal, siendo información publicada con carácter público y no teniendo en el archivo robots.txt ningún tipo de deshabilitado para los enlaces o páginas de los que se ha obtenido la información, tal y como se puede ver en la siguiente imagen:

```
User-Agent: *
Disallow: /include
Disallow: /contactar
Disallow: /vendor
Disallow: /api
Disallow: /lib
Disallow: /cron
Disallow: /actividades/voucher
Disallow: /actividades/reserva_confirmada
Disallow: /traslados/reserva_confirmada
Disallow: /add-activity-to-cart
Disallow: /add-transfer-to-cart
Disallow: /*?rating*
Disallow: /*?cart*
Disallow: /*?origen*
Disallow: /foro
Disallow: /blog/tag/
Disallow: /blog/author/
Disallow: /blog/category/
Disallow: /blog/2008/
Disallow: /blog/2009/
Disallow: /blog/2010/
Disallow: /blog/2011/
Disallow: /blog/2012/
Disallow: /blog/2013/
Disallow: /blog/2014/
Disallow: /blog/2015/
Disallow: /blog/2016/
Disallow: /blog/2017/
Disallow: /blog/2018/
Disallow: /blog/2019/
Disallow: /blog/2020/
Disallow: /blog/2021/
Disallow: /blog/2022/

Allow: /

User-agent: MJ12bot
Disallow: /

Sitemap: https://www.civitatis.com/sitemap.xml
Sitemap: https://www.civitatis.com/sitemap_images.xml
```