

**1. Título del dataset. Poned un título que sea descriptivo.**

Páginas amarillas de panamá.com.

**2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.**

Listado de empresas agrupadas por 158 categorías de panamá.

**3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente**



**4. Contexto. ¿Cuál es la materia del conjunto de datos?**

Este conjunto de datos recoge las empresas que se han dado de alta en páginas amarillas de panamá .net.

**5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?**

Los campos incluyen el nombre de la empresa, la página web de la misma, la categoría , la localización y el teléfono de la empresa. Al tratarse de una web comercial el periodo de tiempo que abarca no se puede garantizar , por lo que solo aseguramos el momento de recogida, en este caso el 11 de noviembre de 2018, estos datos se han recogido scraping su página web

<http://www.paginasamarillasdepanama.com>

**6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Incluid citas de investigación o análisis anteriores.**

El propietario del conjunto de datos es <http://www.paginasamarillasdepanama.com> no tienen ningún tipo de disclaimer en su página web, anteriormente se intentó analizar otras webs del país como Yoytec, ferrocarriles de Panamá, o horarios de paso de barcos por el canal de Panamá

Yoytec, es una tienda de material informático, se descartó por que no respondieron a la solicitud de permiso de scraping, y no se podía realizar sin el consentimiento , ya que especifican la propiedad intelectual de los contenidos

Ferrocarriles de Panamá se descartó por que la página web tenía numerosos errores y los horarios eran fijos, sin apenas interés como práctica.

El horario de paso de barcos por el canal de Panamá, se descartó ya que el canal de panamá no da ese servicio, y una de las web donde se encuentra, marinetraffic.com, ofrece el servicio de exportación de datos mediante csv por lo que desaparecía el objetivo de la práctica.

**7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?**

Estos datos son interesantes porque aparte de conocer las empresas, todas ellas son propensas a pagar para conseguir publicidad o posicionamiento, por lo que se puede utilizar este listado para desarrollar una campaña de captación de clientes en una empresa de publicidad o similar

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

- Released Under CC0: Public Domain License
- Released Under CC BY-NC-SA 4.0 License
- Released Under CC BY-SA 4.0 License
- Database released under Open Database License, individual contents under Database Contents License
- Other (specified above)
- Unknown License

En este caso , la licencia que creo que sería la más adecuada para este dataset sería la CC BY-SA 4.0 License, ya que esta permite que se utilice comercialmente, que es uno de los usos más directos de este dataset ,Otro punto interesante de esta licencia es que obliga a que se mencione el creador , por lo que evita que se pierda el nombre del creador, y por ultimo obliga a que todos los usos de este dataset así como sus modificaciones tengan que permanecer dentro del mismo tipo de licencia, garantizando la mención del autor.

**9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset**

Se adjunta el código en Python.

**10. Dataset: Dataset en formato CSV Se adjuntan los datos en formato CSV.**

Los campos se separan con ','

