

## M2.851 – Tipología y ciclo de vida de los datos Aula 2

### Práctica 1

#### Presentación

Elaborar un caso práctico orientado a aprender a identificar los datos relevantes por un proyecto analítico y usar las herramientas de extracción de datos.

#### Competencias

En esta PEC se desarrollan las siguientes competencias del Máster de Data Science:

- Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para resolverlo.
- Capacidad para aplicar las técnicas específicas de web scraping.

#### Objetivos

Los objetivos de esta práctica son:

- Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinarios.
- Saber identificar los datos relevantes que su tratamiento aportan valor a una empresa y la identificación de nuevos proyectos analíticos.
- Saber identificar los datos relevantes para llevar a cabo un proyecto analítico.
- Capturar datos de diferentes fuentes de datos (tales como redes sociales, web de datos o repositorios) y mediante diferentes mecanismos (tales como queries, API y scraping).
- Actuar con los principios éticos y legales relacionados con la manipulación de datos en función del ámbito de aplicación.
- Desarrollar la capacidad de búsqueda, gestión y uso de información y recursos en el ámbito de la ciencia de datos.

#### Desarrollo de la Práctica

El objetivo de esta actividad es la creación de un dataset a partir de los datos contenidos en la web.

##### 1. Título del dataset

Suicidios anuales por cada 100.000 personas a nivel mundial

## 2. Subtítulo del dataset

Información mundial clasificada en rankings por género

## 3. Imagen

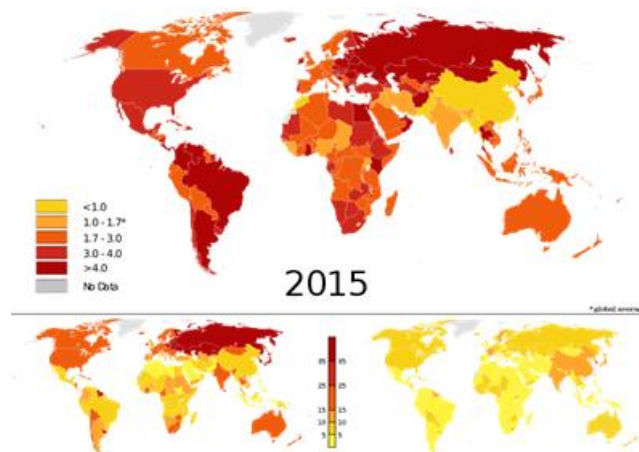


Ilustración 1. Tomado de: [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_suicide\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate)

Suicides per 100,000 people per year (age standardized) <sup>[16]</sup>						
Both sexes rank	Country	Both sexes	Male rank	Male	Female rank	Female
1	Guyana (more info)	30.2	3	46.6	5	14.2
2	Lesotho	28.9	25	22.7	1	32.6
3	Russia (more info)	26.5	1	48.3	31	7.5
4	Lithuania (more info)	25.7	2	47.5	37	6.7
5	Suriname	23.2	6	36.1	13	10.9
6	Cote d'Ivoire	23.0	8	32.0	8	13.0
7	Kazakhstan (more info)	22.8	4	40.1	28	7.7
8	Equatorial Guinea	22.0	9	31.3	15	10.8
9	Belarus	21.4	5	39.3	45	6.2
10	South Korea (more info)	20.2	11	29.6	11	11.6
11	Uganda <sup>b</sup>	20.0	37	21.2	2	18.7
12	Cameroon (more info)	19.5	13	26.9	10	12.5
13	Zimbabwe	19.1	12	29.1	12	11.1
14	Ukraine (more info)	18.5	7	34.5	81	4.7
15	Nigeria	17.3	56	17.5	3	17.1
16	Latvia	17.2	10	31.0	71	5.1
17	Swaziland	16.7	17	25.4	17	9.6
18	Togo	16.6	25	22.7	13	10.9
19	India (more info)	16.5	46	18.5	4	14.5
19	Uruguay	16.5	14	26.8	34	7.1

Ilustración 2. Tomado de: [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_suicide\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate)

## 4. Contexto

La Organización Mundial de la Salud (OMS) consolida anualmente las tasas de suicidio con base en los informes publicados o proporcionados por los países.

El ranking por país está determinado con base en la tasa total ajustada por edad y género por cada 100,000 habitantes de muertes registradas oficialmente como suicidios.

## 5. Contenido

### 5.1 Campos

- Ranking de ambos géneros
- País
- Promedio de edad ambos géneros
- Ranking de suicidio por género masculino
- Promedio de edad por género masculino
- Ranking de suicidio por género femenino
- Promedio de edad por género femenino

### 5.2 Periodo de los datos

La información recopilada corresponde al consolidado realizado por la Organización Mundial de la Salud (OMS) para el año 2016.

### 5.3 ¿Cómo se ha recogido?

La técnica utilizada para la recolección de los datos de la tabla “Suicides per 100,000 people per year (age standardized)[16]” fue “WebScraping”, a través de la construcción de programa desarrollado con Python para almacenarlos en un dataset.

## 6. Agradecimientos

### 6.1 Propietario del conjunto de datos

La información publicada en “[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_suicide\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate)” es proporcionada por el “World Health Organization (WHO)” y otras fuentes.

### 6.2 Citas de investigación o análisis anteriores relacionados en “[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_suicide\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate)”

- Suicide rates by country, 1950-2011 published by the World Health Organization (WHO)
- Top Ten Most Suicidal Countries published by International Business Times
- SA has the eighth-highest suicide rate in the world, psychiatric illnesses are on the rise and stress has become normalised published by Mail & Guardian

## 7. Inspiración

### 7.1 ¿Por qué es interesante este conjunto de datos?

Con esta información los estados pueden compararse con otras naciones, identificar las causas que estén llevando a su población a tomar esta clase de medidas contra su vida y adelantar programas para mejorar la calidad de vida de sus conciudadanos.

### 7.2 ¿Qué preguntas le gustaría responder a la comunidad?

Esta clase de informes le deben resolver a la comunidad inquietudes especialmente sobre el género y rango de edades más vulnerable en su población.

## 8. Licencia

El informe publicado en la dirección [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_suicide\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate) ha sido creado mediante la siguiente modalidad de licenciamiento:

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply.

La cual corresponde a:

Released Under CC BY-SA 4.0 License



Los atributos de este tipo de licencia especifican que:

Usted es libre de:

Atributo	Alcance
Compartir	copie y redistribuya el material en cualquier medio o formato
Adaptar	remezclar, transformar y construir sobre el material. Para cualquier propósito, incluso comercialmente



Bajo los siguientes términos:

Atributo	Alcance
Atribución	debe otorgar el crédito correspondiente, proporcionar un enlace a la licencia e indicar si se realizaron cambios. Puede hacerlo de cualquier manera razonable, pero no de ninguna manera que sugiera que el licenciante lo respalda a usted o a su uso.
ShareAlike	si remezcla, transforma o construye sobre el material, debe distribuir sus contribuciones bajo la misma licencia que el original
Sin restricciones adicionales	no puede aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita la licencia

## 9. Código

El código (main.py) con el que se ha generado el dataset fue elaborado en lenguaje Python el cual se adjunta en la ruta ["https://github.com/cboterouoc/SuicidiosScraper"](https://github.com/cboterouoc/SuicidiosScraper) como se ilustra a continuación:

Manage topics

3 commits

1 branch

0 releases

1 contributor

Branch: master






New pull request

Create new file

Upload files

Find file

Clone or download

 cboterouoc	Scripts y archivo CSV	Latest commit 8c6aa48 a day ago
 DatosSuicidioPais.py	Scripts y archivo CSV	a day ago
 README.md	Nueva actualización	a day ago
 main.py	Scripts y archivo CSV	a day ago
 suicidos_a_nivel_mundial.csv	Scripts y archivo CSV	a day ago

## Librerías

### Instalación requests

```
python -m pip install requests
beautifulsoup
```

```
python -m pip install beautifulsoup
```

```
pandas
```

```
python -m pip install pandas
```

## 10. Dataset

Igualmente, el Dataset (suicidios\_a\_nivel\_mundial.csv) producto de la ejecución del código anteriormente indicado se adjunta en la ruta “<https://github.com/cboterouoc/SuicidiosScraper>” como se ilustra a continuación:

Manage topics

3 commits

1 branch

0 releases

1 contributor

Branch: master






New pull request

Create new file

Upload files

Find file

Clone or download

 cboterouoc	Scripts y archivo CSV	Latest commit 8c6aa48 a day ago
 DatosSuicidioPais.py	Scripts y archivo CSV	a day ago
 README.md	Nueva actualización	a day ago
 main.py	Scripts y archivo CSV	a day ago
 suicidos_a_nivel_mundial.csv	Scripts y archivo CSV	a day ago

## 11. Recursos

Los siguientes recursos fueron consultados para la ejecución de la presente práctica:

- Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.
- Masip, D. El lenguaje Python. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015) Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.
- Tutorial de Github <https://guides.github.com/activities/hello-world> .

<https://github.com/cboteroUOC/SuicidiosScraper>