

# Dataset: Pandemia Covid-19: Datos a nivel mundial

*Raquel Gómez Pérez y Jorge Serra Planelles*

*14 de abril de 2020*

## Descripción

El dataset generado contiene la información obtenida sobre la pandemia mundial Covid-19 a nivel de número de personas afectadas (infectadas, fallecidas...).

## Representación gráfica



## Contexto

Esta información se ha obtenido de tres páginas web:

- <https://www.worldometers.info/coronavirus/>: de esta web se ha obtenido la información total de todos los países y la información por estados de los Estados Unidos. Worldometers es una web que publica estadísticas del mundo en tiempo real. Por ejemplo cifras sobre población mundial, gobiernos y economía, sociedad y medios, medio ambiente, alimentos, agua, energía y salud. Actualmente, Worldometer es el proveedor de datos globales de COVID-19 para gobiernos e instituciones de todo el mundo.
- <https://covid19.isciii.es/>: de esta web se ha obtenido la información por regiones de España. en este caso solo existía información de casos totales. Esta web es propiedad del Instituto de Salud Carlos III (ISCIII) el cual es la referencia nacional e internacional en investigación biomédica y salud pública en España. Por tanto esta página recoge la notificación diaria de casos agregados de Covid-10 al Ministerio de sanidad.
- [https://it.wikipedia.org/wiki/Pandemia\\_di\\_COVID-19\\_del\\_2020\\_in\\_Italia](https://it.wikipedia.org/wiki/Pandemia_di_COVID-19_del_2020_in_Italia): de esta web se ha obtenido la información por regiones de Italia. Wikipedia es una enciclopedia libre, políglota y editada de manera colaborativa, que tiene información de múltiples categorías, incluida la categoría 'Salud', en la cual estaría incluida esta información.

## Contenido

Los campos que se han recogido en el dataset son los siguientes:

- **País:** país para el cual vamos a obtener los datos de la pandemia Covid-19.
- **Región:** región del país al que corresponden los datos. En caso que los datos sean del total del país en este campo aparecerá 'Total'.
- **Casos totales:** número total de personas registradas portadoras del Covid-19.
- **Fallecimientos:** número de personas fallecidas a causa del virus.
- **Recuperados:** número de personas que han sido portadoras del virus y se han recuperado.
- **Casos activos:** número de personas portadoras del virus actualmente (casos totales - fallecimientos - recuperados).
- **Casos cerrados:** número de personas que han portado el virus pero no son casos activos actualmente (recuperados + fallecimientos).
- **Fecha:** fecha en la cual se han actualizado los datos en formato DD/MM/YYYY hh:mm.

Esta información se ha recogido a nivel mundial, con datos totales de países, y a nivel más específico según regiones para los tres países con más casos actualmente: Estados Unidos, España e Italia. Se trata de datos totales desde que comenzó la expansión del virus hasta la última fecha de actualización indicada en el campo fecha.

## Agradecimientos

En primer lugar agradecemos al equipo de Worldometer, un equipo internacional de desarrolladores, investigadores y voluntarios que tienen como objetivo de hacer que las estadísticas mundiales estén disponibles en un formato de reflexión y tiempo relevante para una amplia audiencia de todo el mundo. Para el caso de Covid-19 recopilan datos de informes oficiales, directamente de los canales de comunicación del Gobierno o indirectamente, a través de fuentes de medios locales cuando se consideran confiables.

En segundo lugar a la comunidad de wikipedia, por recopilar tantos temas de interés de forma libre y colaborativa.

En tercer lugar al Instituto de Salud Carlos III, es la referencia nacional e internacional en investigación biomédica y salud pública en España. Es el Organismo Público de Investigación (OPI) del Gobierno encargado de financiar y llevar a cabo la investigación biomédica nacional. Depende del Ministerio de Ciencia, Innovación y Universidades, aunque también está adscrito al Ministerio de Sanidad, Consumo y Bienestar Social.

## Inspiración

El contenido del dataset es un tema de interés mundial actualmente, debido a la crisis sanitaria por la que estamos pasando. Hay muchos aspectos que pueden analizarse con estos datos, tanto desde un punto de vista matemático y estadístico como en relación a su impacto en las decisiones que han tomado diversos gobiernos regionales o nacionales. El procesamiento de estos datos es importante para la adopción de medidas orientadas a la reducción del número de contagios, o a la ralentización de los mismos. También podría ser interesante para realizar modelos de predicción de la evolución del brote

## Licencia

La licencia escogida para este conjunto de datos es Released Under CC0: Public Domain License. Es una licencia libre de derechos, permite la libertad total de reproducción para terceros. Escogería esta porque, en primer lugar, los datos que hemos obtenidos ya son públicos, por lo que no tendría sentido pedir más

derechos sobre ellos que los de las fuentes de origen. Además, requerir una licencia más estricta podría hacer que un posible usuario de los datos se echará para atrás para evitar líos legales. En segundo lugar, son datos que tienen interés público y general, por lo que la libre distribución de los mismos es más beneficiosa.

## **Código fuente y dataset**

Tanto el código fuente escrito para la extracción de datos como el dataset generado pueden ser accedidos a través de este enlace.

## **Recursos**

1. Reitz, K. 2018. Requests-HTML: HTML Parsing for Humans. From: <https://requests.readthedocs.io/projects/requests-html/en/latest/>.
2. Richardson, R. 2020. BeautifulSoup Documentation. From Crummy: <https://www.crummy.com/software/BeautifulSoup/>
3. Schafer, C. 2017. Python Tutorial: Web Scraping with BeautifulSoup and Requests. From YouTube: <https://www.youtube.com/watch?v=ng2o98k983k&t=917s>.
4. Schafer, C. 2019. Python Tutorial: Web Scraping with Requests-HTML. From YouTube: <https://www.youtube.com/watch?v=a6flbtFB46g&t=645s>.
5. Creative Commons. 2020. From: <https://creativecommons.org/>.