

# Dataset: Covid-19, estadísticas diarias por países en el mundo

*Jhon H. Loaiza G.*

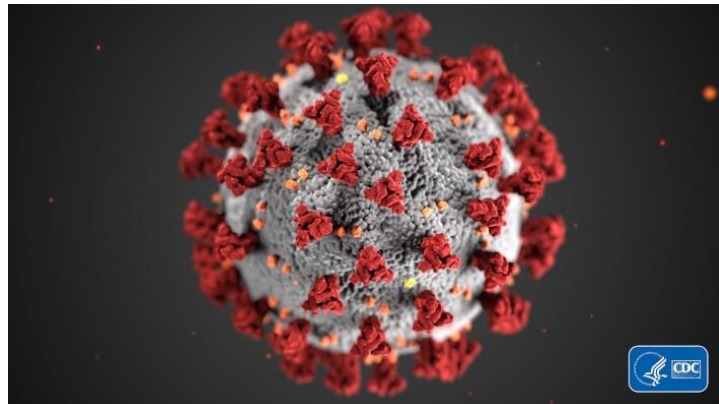


Figura 1. Imagen de referencia del virus Covid-19, CDC

## Descripción

Dada la coyuntura mundial por la emergencia sanitaria, provocada por el Covid-19 y su rápido crecimiento en contagios a nivel global, decidí hacer un aporte al poder descargar los datos de una de las páginas más importantes para analizar el comportamiento del covid-19. Este dataset es creado a partir del sitio web [Worldometers](https://www.worldometers.info/covid-19/) la práctica de Webscraping del curso de Tipología y ciclo de vida del dato, del master en Ciencia de datos de la UOC.

El dataset consta de la información de casos reportados por países a nivel mundial, fallecimientos, pruebas realizadas, casos recuperados, población, entre otras.

## Contexto

Dada la emergencia sanitaria actual, es una necesidad de los gobiernos nacionales y entidades de salud a nivel mundial, reportar las estadísticas del Covid-19 en sus territorios, principalmente de casos nuevos, fallecimientos, recuperados, así como también el número de pruebas o test realizados a la población. El dataset muestra la información actualizada al día (nuevos casos y fallecimientos del día) y acumulado de casos totales por país.

## Contenido

Para cada país, se recopila la siguiente información:

- #: Posición del país en el listado basado en número de casos totales

- **Country/Other:** Nombre del país
- **TotalCases:** Casos totales acumulado
- **NewCases:** Casos nuevos del día
- **TotalDeaths:** Fallecimientos totales acumulado
- **NewDeaths:** Fallecidos nuevos del día
- **TotalRecovered:** Total de recuperados
- **NewRecovered:** Recuperados nuevos del día
- **ActiveCases:** Casos activos acumulado
- **Serious,Critical:** Casos críticos (en UCI u hospitalizados)
- **TotCases/1M pop:** Total de casos por cada millón de habitantes
- **Deaths/1M pop:** Fallecimientos por cada millón de habitantes
- **TotalTests:** Cantidad total de tests realizados
- **Tests/1M pop:** Tests realizados por cada millón de habitantes
- **Population:** Población o cantidad de habitantes
- **Continent:** Continente al cual pertenece el país
- **1 Caseevery X ppl:** 1 caso cada X cantidad de Habitantes
- **1 Deathevery X ppl:** 1 fallecimiento cada X Habitantes
- **1 Testevery X ppl:** 1 test cada X Habitantes

La web de [Worldometers](#) habilito este sitio desde el inicio de la pandemia y se asume que estará activo hasta que termine la misma. [Worldometers](#) recopila y valida la información de distintas fuentes acerca del covid-19, haciéndola una fuente confiable para el seguimiento de la pandemia.

## Agradecimientos

Como se ha mencionado anteriormente, los datos han sido extraídos de la web de Worldometers, dadas su validez y controles de las fuentes de información, lo hace un sitio referencia en materia de estadísticas de Covid-19. La extracción se realizó usando tecinas de Webscraping con Python y las librerías requests y lxml.

## Inspiración

Al ser una información valiosa dada la situación actual a nivel global, y a pesar que puede ser consultada en la web, el poder tener los datos en archivos, con las cifras actualizadas diariamente, permite ver tendencias de nuevos casos en los distintos países, así como el comportamiento de las medidas y controles que están haciendo los países y su efectividad en la disminución de los casos, basado en un histórico de estos datos. Este dataset podría ser interesante para muchos ámbitos, principalmente epidemiólogos, científicos de datos que quieran modelar el comportamiento del covid-19, medios de comunicación y revistas especializadas en salud. El script puede automatizarse para generar los archivos con las cifras de Covid-19.

## Licencia

Para este conjunto de datos, se escoge la licencia **Attribution 4.0 International (CC BY 4.0)**, la cual se considera idónea, dado que los datos que se extraen son de carácter público y este proyecto facilita la obtención y posible almacenamiento de esa información. La licencia cuenta con una única clausula:

- **Atribución** — Debe darse el crédito a la persona o creador de los datos en este caso, indicando que cambios se le han realizado. Es un reconocimiento al trabajo realizado.

Esta licencia permite usar los datos, transformarlos y compartirlos, además se pueden usar para fines comerciales sin ninguna restricción.

## Código fuente y dataset

El código fuente y el dataset resultante pueden ser descargados del siguiente link: <https://github.com/JhonHarry/WebScraping>.

## Recursos

- Subirats, L., Calvo, M. (2018). *Web Scraping*. Editorial UOC.
- Lawson, R. (2015). *Web Scraping with Python*. Packt Publishing Ltd. Chapter 2. Scraping the Data.