

Programa de formación Machine Learning and Data Science MLDS



Módulo 1

Análisis y visualización de datos con Python

Unidad 2

Análisis de datos con *Pandas*

Lugares para encontrar *datasets* de uso libre

Lugares para encontrar datasets de uso libre

En este documento se presentan algunos sitios web donde puede encontrar conjuntos de datos abiertos para su uso en proyectos de *Machine Learning* y *Data Science*.

Kaggle

Sitio web especializado en la ciencia de datos donde puede encontrar *datasets* para diversas tareas de aprendizaje de máquina como clasificación, regresión, *retrieval*, recomendación y demás, relacionados con una amplia variedad de campos de aplicación. Puede encontrar desde *datasets* médicos hasta datos bancarios y recomendaciones de *Netflix*.



<https://www.kaggle.com/datasets>

UCI Machine Learning Repository

Repositorio de aprendizaje computacional administrado y mantenido por la Universidad de California Irvine, que almacena una gran cantidad de *datasets* para todo tipo de tareas y con todo tipo de contenidos como imágenes médicas, textos, entre otros.



<https://archive.ics.uci.edu/ml/datasets.php>

FiveThirtyEight

Este sitio expone de manera libre los datos utilizados en la generación de infografías en distintos medios de comunicación. Muy recomendable para la construcción de visualizaciones de distintas temáticas de interés público.



<https://data.fivethirtyeight.com/>

Lugares para encontrar datasets de uso libre

Datos abiertos Colombia

Si es de su interés acceder a datos regionales, en este sitio web puede encontrar *datasets* proporcionados por entidades estatales y privadas en Colombia, como parte de la iniciativa de Estado Abierto del Ministerio de Tecnologías de la Información y las Comunicaciones - MinTIC.



<https://www.datos.gov.co/>

Datos abiertos Bogotá

En este sitio web puede encontrar la plataforma distrital de Datos Abiertos de la Alcaldía Mayor de Bogotá. Esta iniciativa es similar a la de Datos Abiertos del Gobierno de Colombia y busca mejorar la transparencia de los datos en entidades públicas distritales.



<https://datosabiertos.bogota.gov.co/>

Iniciativa Latinoamericana por los Datos Abiertos

En este sitio puede encontrar reportes y proyectos de investigación realizados para promover y entender el uso de los datos abiertos en América Latina, como parte de la Iniciativa Latinoamericana por los Datos Abiertos en la que participan varios países de la región. En ella, puede encontrar reportes del estado de cada país en el acceso a este tipo de datos.



<https://idatosabiertos.org/>

Lugares para encontrar datasets de uso libre

COVID-19

Varias organizaciones, como el Gobierno de Colombia con la iniciativa de Datos Abiertos, disponen al público los *dataset* con el reporte diario de la evolución de la pandemia de enfermedad por coronavirus COVID-19. La [Organización Mundial de la Salud](#) permite descargar los datos recopilados en su reporte de la situación. Además, instituciones como la [Universidad John Hopkins](#) o iniciativas como [Our World in Data](#) disponen los datos usados en sus proyectos de investigación, en los que se recolectan datos de los países afectados alrededor del mundo.



World Health Organization

<https://github.com/CSSEGISandData/COVID-19>
<https://ourworldindata.org/coronavirus-data>
<https://covid19.who.int/table>

Lista de recursos

- <https://www.kaggle.com/datasets>
- <https://archive.ics.uci.edu/ml/datasets.php>
- <https://data.fivethirtyeight.com/>
- <https://www.datos.gov.co/>
- <https://datosabiertos.bogota.gov.co/>
- <https://idatosabiertos.org/>
- <https://github.com/CSSEGISandData/COVID-19>
- <https://ourworldindata.org/coronavirus-data>
- <https://covid19.who.int/table>
- <https://www.dataquest.io/blog/free-datasets-for-projects/>

Derechos de imágenes

Logo Kaggle <https://www.kaggle.com/arunsankar/kaggle-logo>

Logo UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/assets/logo.gif>

Logo FiveThirtyEight: https://commons.wikimedia.org/wiki/File:FiveThirtyEight_Logo.svg

Logo Datos abiertos Colombia:
https://herramientas.datos.gov.co/sites/default/files/pictures/logo_pda.png

Logo Datos abiertos Bogotá: <https://datosabiertos.bogota.gov.co/img/logo.svg>

Logo Iniciativa Latinoamericana por los Datos Abiertos: https://idatosabiertos.org/wp-content/themes/ilda/dist/img/ilda_logo.svg

Logo Organización Mundial de la Salud: <https://www.who.int/about/who-we-are/publishing-policies/logo>

Créditos

Facultad de

I N G E N I E R Í A

Autor

Felipe Restrepo Calle, PhD

Asistente docente

Alberto Nicolai Romero Martínez

Diseño instruccional

Claudia Patricia Rodríguez Sánchez

Diseño gráfico

Clara Valeria Suárez Caballero

Milton R. Pachón Pinzón

Diagramación PDF

Daniela Duque García

Fecha

2020-II

Programa de formación Machine Learning and Data Science MLDS

Facultad de
I N G E N I E R Í A
Sede Bogotá



UNIVERSIDAD
NACIONAL
DE COLOMBIA