

Taller Entornos de Ejecución

Integrantes:

- Guillermo Pinto
- Brayan Yesid Quintero Santander

Material de apoyo:

- Diapositivas - **Do not just use Colab:** [Ver presentación](#)
- Notebooks - **Hugging Face and Kaggle Datasets**
 - Colab version: [Ver notebook](#)
 - Kaggle version: [Ver notebook](#)

Entornos de ejecución

Los entornos de ejecución son las plataformas donde podemos ejecutar nuestros experimentos de manera gratuita. Generalmente, solo se conoce Google Colaboratory para este fin, sin embargo, actualmente existen diferentes alternativas, unas mejores y otras no tanto, que nos permiten sacar adelante alguna idea.

En este taller revisamos cada una de las alternativas disponibles en Colombia en el momento. Vimos cómo usarlas, las comparamos y definimos un *tier list* de cuáles, desde la perspectiva de Guille, son las mejores.

Al final revisamos cómo hacer uso de los *secretos*, los cuales son claves como *tokens* o *api keys* que nos permiten acceder a contenido privado de manera segura. Todo esto lo encuentras en las diapositivas **Do not just use Colab**.

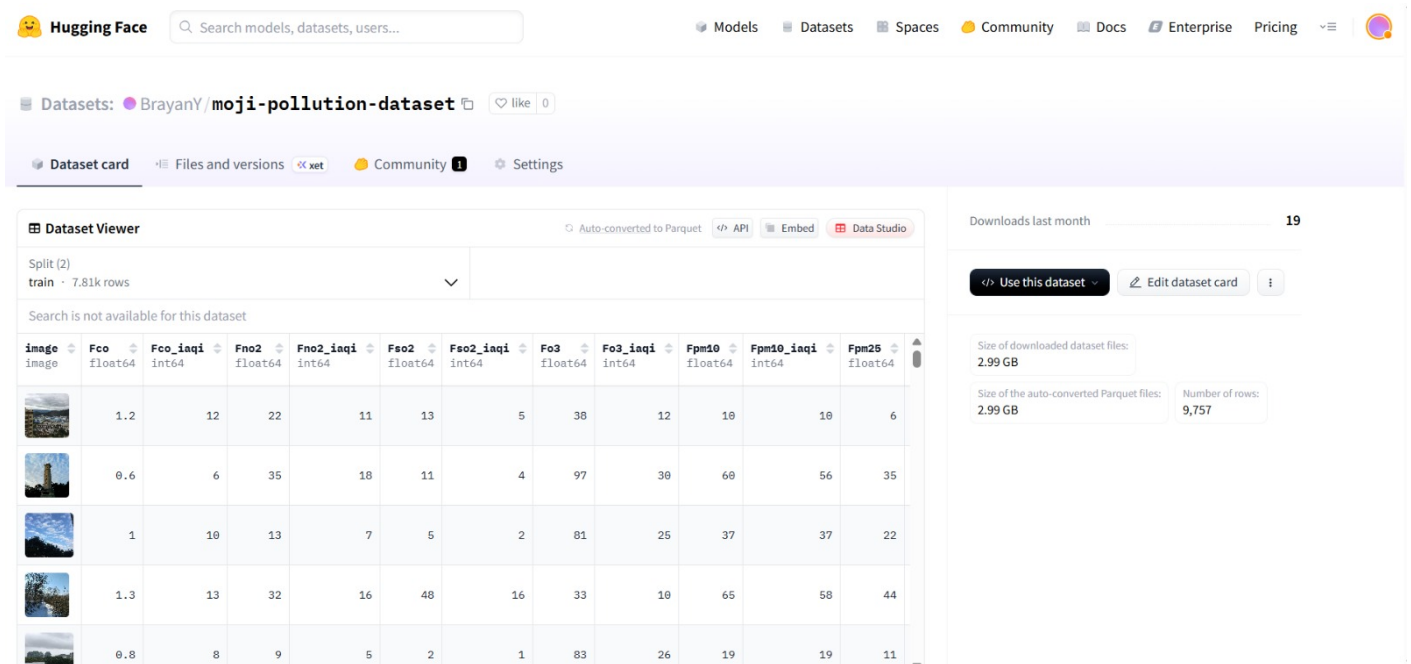
Datasets

Cuando estamos trabajando en un proyecto necesitamos datos. Pero más importante aún, necesitamos un lugar donde guardar estos datos de manera que podamos accederlos sin muchas complicaciones.

Entre las mejores alternativas encontramos *Hugging Face Datasets* y *Kaggle Datasets*. Dos plataformas donde podemos alojar nuestros datos de manera gratuita y luego accederlos, desde cualquier lugar, a través de los secretos (*api keys* o *tokens*).

¿Cómo subir un dataset?

Por último, se subió un **dataset de manera privada** siguiendo la estructura recomendada por *Hugging Face Datasets*, el conjunto de datos quedó organizado con la siguiente estructura:



The screenshot shows the Hugging Face dataset viewer for the 'moji-pollution-dataset'. The dataset is split into 'train' with 7.81k rows. The 'Dataset Viewer' section shows a table with columns for 'image', 'Fco', 'Fco_iaqi', 'Fno2', 'Fno2_iaqi', 'Fso2', 'Fso2_iaqi', 'Fo3', 'Fo3_iaqi', 'Fpm10', 'Fpm10_iaqi', and 'Fpm25'. The table contains 5 rows of data. On the right, there are statistics: 'Downloads last month: 19', 'Size of downloaded dataset files: 2.99 GB', and 'Size of the auto-converted Parquet files: 2.99 GB' with 'Number of rows: 9,757'.

image	Fco	Fco_iaqi	Fno2	Fno2_iaqi	Fso2	Fso2_iaqi	Fo3	Fo3_iaqi	Fpm10	Fpm10_iaqi	Fpm25
	1.2	12	22	11	13	5	38	12	19	19	6
	0.6	6	35	18	11	4	97	38	68	56	35
	1	10	13	7	5	2	81	25	37	37	22
	1.3	13	32	16	48	16	33	10	65	58	44
	0.8	8	9	5	2	1	83	26	19	19	11

Esta estructura asegura que cada subconjunto (train y test) tenga sus propias imágenes y metadatos, donde cada *metadata.csv* contenga únicamente la información correspondiente a cada subconjunto. Puedes descubrir de manera práctica cómo hacerlo en los dos notebooks que encuentras en el material de apoyo.

— Guillermo Pinto y Brayan Quintero

“📖 Recursos adicionales

- [“Lightning AI | Turn Ideas Into AI, Lightning Fast.” Lightning AI, lightning.ai.](https://lightning.ai/)
- [Kaggle: Your Machine Learning and Data Science Community. www.kaggle.com](https://www.kaggle.com/)
- [SageMaker Studio Lab. studiolab.sagemaker.aws.](https://studiolab.sagemaker.aws/)

- ["GitHub Codespaces." GitHub, 2025, github.com/features/codespaces.](https://github.com/features/codespaces)
- ["Modal: High-performance AI Infrastructure." Modal, modal.com.](https://modal.com)
- [AI Code King. "These Are the Best Google Colab Alternatives! \(Free Tiers With GPUs\)." YouTube, 11 May 2024, www.youtube.com/watch?v=yvvNtkfjhGI.](https://www.youtube.com/watch?v=yvvNtkfjhGI)
- [Datasets. huggingface.co/docs/datasets/en/index.](https://huggingface.co/docs/datasets/en/index)
- [Datasets Documentation. www.kaggle.com/docs/datasets.](https://www.kaggle.com/docs/datasets)

Revision #6

Created 30 October 2025 22:21:39 by Brayan Quintero

Updated 2 November 2025 12:50:33 by Guillermo Pinto