



TEC. AZUAY
INSTITUTO UNIVERSITARIO

**TECNOLOGÍA SUPERIOR EN DESARROLLO DE
SOFTWARE**

ASIGNATURA:

... POO 2 ...

TEMA:

... PROYECTO FINAL ...

ESTUDIANTE:

MARLON MAURICIO MARCA MEJICANGO

DOCENTE:

.....ING. DANIEL ORTIZ

CURSO:

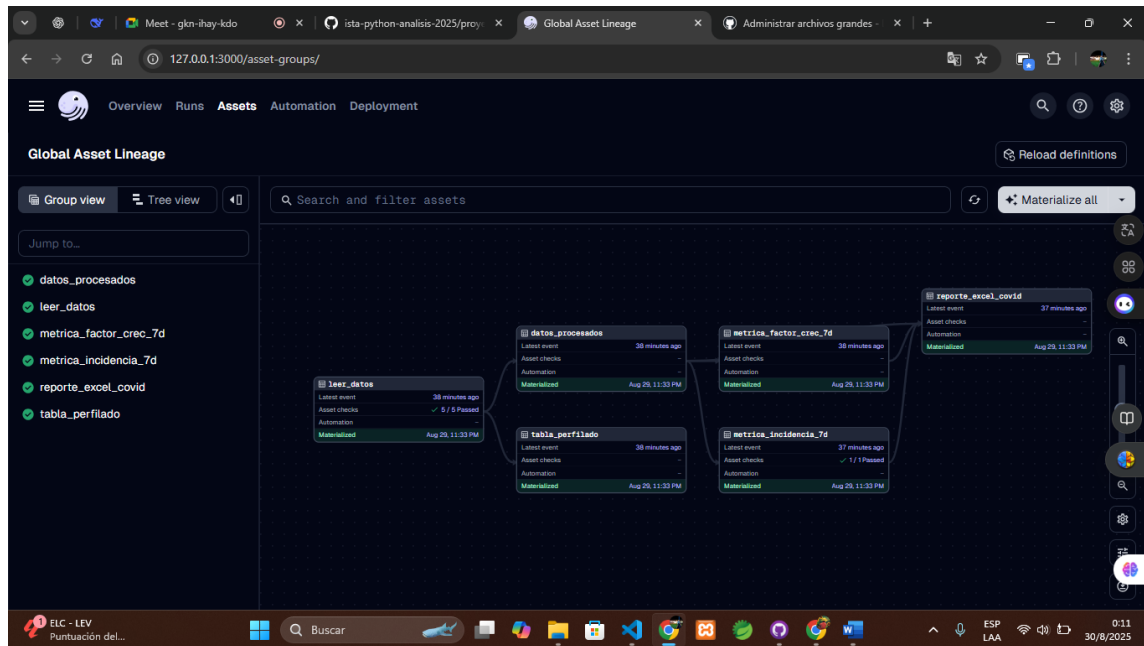
N6A

AÑO LECTIVO:

ABRIL – AGOSTO 2025

1) Arquitectura del pipeline

Diagrama:



Assets y propósito (resumen):

- **leer_datos:** ingesta de compact.csv (OWID o archivo local vía OWID_LOCAL_PATH). Sin transformar.
- **tabla_perfilado:** perfilado mínimo (filas/columnas, min-max new_cases, % nulos, rango de fechas). Exporta tabla_perfilado.csv.
- **datos_procesados:** limpieza (coerción numérica; fechas tz-naive), drop de nulos críticos y duplicados, filtro a **Ecuador y país comparativo** (p. ej., Perú), selección de columnas esenciales.
- **metrica_incidencia_7d:** $(\text{new_cases} / \text{population}) * 100000$ con **media móvil 7d**.
- **metrica_factor_crec_7d:** $\text{rolling_sum_7d}(\text{new_cases})$ vs. semana previa → **factor de crecimiento**.
- **reporte_excel_covid:** empaqueta resultados finales en reporte_covid.xlsx (hojas: datos procesados, incidencia_7d, factor_crec_7d).

2) Decisiones de validación

Entrada y motivación:

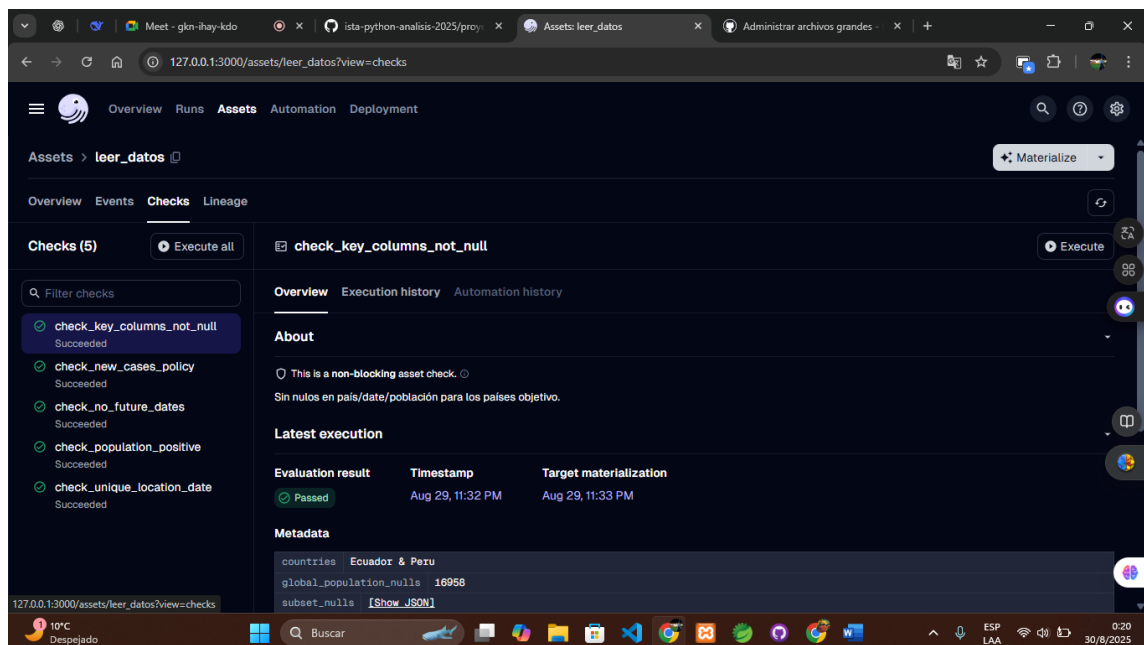
- **Fechas no futuras (en el subconjunto con datos):** evita analizar placeholders o fechas anticipadas del CSV global.
- **Claves no nulas (país/date/población) en el subconjunto:** asegura completitud mínima para métricas per cápita.
- **Unicidad (país, fecha):** previene doble conteo diario.
- **Población > 0:** evita divisiones inválidas.
- **Política de negativos en new_cases:** OWID puede registrar revisiones negativas; la política es configurable (permitir o alertar) y se documenta.

Salida (checks sobre metrica_incidencia_7d) y motivación:

- **Rango razonable** $0 \leq \text{incidencia_7d} \leq 2000$: control básico de outliers por errores de datos o escalas.

Descubrimientos relevantes del dataset:

- Cobertura y frecuencia de reporte heterogéneas entre países → conviene suavizar con ventanas 7d.
- Presencia ocasional de valores negativos en new_cases por revisiones.
- El CSV global puede incluir fechas futuras; al acotar al subconjunto con datos reales, los checks pasan.



3) Consideraciones de arquitectura

- **pandas (elegido en este MVP):** sencillo para **ventanas móviles** y volumen reducido (2 países); excelente ergonomía para limpieza rápida.
- **DuckDB:** ideal si escalamos a muchos países, necesitamos joins o queremos SQL reproducible y rendimiento en analítica tabular local.

```

1 from __future__ import annotations
2 import os
3 import io as pyio
4 import requests
5 import pandas as pd
6 from dagster import asset, asset_check, AssetCheckResult
7
8 # =====
9 # Config
10
11 OMID_URL = os.getenv(
12     "OMID_URL",
13     "https://catalog.ourworldindata.org/garden/covid/latest/compact/compact.csv",
14 )
15
16 OMID_LOCAL_PATH = os.getenv("OMID_LOCAL_PATH", "")
17 COMPARISON_COUNTRY = os.getenv("COMPARISON_COUNTRY", "Peru")
18 ALLOW_NEG = os.getenv("ALLOW_NEGATIVE_NEW_CASES", "false").lower() == "true"
19 OUTPUT_DIR = os.getenv("OUTPUT_DIR", "final_project/output")
20
21 # =====
22 # Helpers
23
24 def _loc_col(df: pd.DataFrame) -> str:
25     """Devuelve el nombre de la columna de país según el dataset:
26     'location' (oid-covid-data) o 'country' (compact.csv)."""
27     if "location" in df.columns:
28         return "location"
29     if "country" in df.columns:
30         return "country"
31     raise KeyError("No se encontró columna de país: 'location' o 'country'.")
32
33 def _scope_rows(df: pd.DataFrame) -> pd.DataFrame:
34     """Filtra solo Ecuador + país comparativo (para checks)."""
35     loc = _loc_col(df)
36     return df[df[loc].isin(["Ecuador", COMPARISON_COUNTRY]).copy()]
37
38 def _process_hace(df: pd.DataFrame, n_ais: str, n_ais_h: str) -> pd.DataFrame:

```

4) Resultados

Los resultados esperados al trabajar con los CSV son reportes claros y comprensibles que muestren cómo ha evolucionado la pandemia en Ecuador y en el país comparativo. En los archivos generados se espera ver tablas limpias, sin duplicados ni datos inconsistentes, donde se reflejen tendencias como el aumento o disminución de casos, así como la cobertura de vacunación. El objetivo es que estos resultados permitan tener una visión resumida y confiable de la situación, para compararla fácilmente entre países y entender de forma sencilla si la pandemia está en crecimiento, estabilización o descenso.

reporte_covid - Excel

ArchivoInicioInsertarDisposición de páginaFórmulasDatosRevisarVistaAyudaCompartir

Portapapeles

Fuente

Alineación

Número

Formato condicional

Dar formato como tabla

Estilos de celda

Estilos

Celdas

Edición

Co

fx

	A	B	C	D	E	F	G	H	I	J
1	location	date	new_cases	ple_vaccin	population					
2	Ecuador	20/1/2021	924	17	17823857					
3	Ecuador	21/1/2021	1747	3008	17823857					
4	Ecuador	22/1/2021	1874	3811	17823857					
5	Ecuador	23/1/2021	969	3848	17823857					
6	Ecuador	25/1/2021	3060	5116	17823857					
7	Ecuador	26/1/2021	275	5876	17823857					
8	Ecuador	27/1/2021	579	6580	17823857					
9	Ecuador	28/1/2021	1389	6802	17823857					
10	Ecuador	29/1/2021	2465	7091	17823857					
11	Ecuador	31/1/2021	3092	7133	17823857					
12	Ecuador	1/2/2021	1049	8521	17823857					
13	Ecuador	2/2/2021	158	9722	17823857					
14	Ecuador	3/2/2021	293	12357	17823857					
15	Ecuador	4/2/2021	1111	12425	17823857					
16	Ecuador	5/2/2021	949	12428	17823857					
17	Ecuador	6/2/2021	2073	12429	17823857					
18	Ecuador	8/2/2021	1167	12431	17823857					
19	Ecuador	10/2/2021	1176	12431	17823857					
20	Ecuador	11/2/2021	293	12466	17823857					
21	Ecuador	12/2/2021	2237	12483	17823857					

datos_procesados

tabla_perfilado - Excel

ArchivoInicioInsertarDisposición de páginaFórmulasDatosRevisarVistaAyudaCompartir

Portapapeles

Fuente

Alineación

Número

Formato condicional

Dar formato como tabla

Estilos de celda

Estilos

Celdas

Edición

Co

fx

	A	B	C	D	E	F	G	H
1	metric,valor							
2	filas,523599							
3	columnas,61							
4	new_cases_min,0.0							
5	new_cases_max,8401906.0							
6	pct_null_new_cases,3.3069199902979185							
7	pct_null_people_vaccinated,84.91001701683922							
8	fecha_min,2020-01-01 00:00:00							
9	fecha_max,2025-12-31 00:00:00							

tabla_perfilado

10°C
Despejado

Buscar

ESP
LAA

0:13
30/8/2025