

Proyecto:

Efectos de la desigualdad socioeconómica en la percepción de la democracia en Latinoamérica: periodo 1995-2023

Producto 1: Consolidación de la base de indicadores y rutinas automáticas

El presente informe contempla el periodo del 1 al 31 de Octubre del 2025, en donde se contemplan los siguientes entregables:

1. Informe de revisión inicial de la base de datos

- Documento que describa la situación de la versión 0 de la base, lista de indicadores presentes, inconsistencias y vacíos detectados.
- Incluye un esquema actualizado del modelo entidad–relación de la base de datos.

2. Propuesta documentada de indicadores complementarios

- Documento con la lista de indicadores adicionales (socioeconómicos y de democracia).
- Fuentes consideradas (ENEMDU, Latinobarómetro, VDEM y otras).
- Factibilidad de cálculo, variables requeridas, periodicidad y sustento metodológico.

3. Rutinas automatizadas de descarga, procesamiento e ingestión de datos

- Scripts en repositorio con control de versiones (GitHub/GitLab/otro).
- Código comentado y estructurado.
- Documento técnico de dependencias, instrucciones de ejecución y mantenimiento de los scripts.

4. Informe de validación metodológica y cálculo de indicadores

- Documento que muestre la validación de las metodologías aplicadas.
- Ejemplos de cálculos realizados y resultados verificados en la base.

5. Inventario final de indicadores

- Tabla de metadatos consolidada (nombre, descripción, fórmula, fuente, frecuencia de actualización, fecha de última actualización, responsable).
- Documento explicativo sobre criterios de selección y exclusión.

San Miguel de Urcuquí, Ecuador

31 de Octubre, 2025

Contents

1	Antecedentes	3
1.1	INEC – Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU)	3
1.2	Latinobarómetro	4
1.3	V-Democracia	4
2	Desarrollo	5
2.1	Informe de revisión inicial de la base de datos	5
2.1.1	Documento que describa la situación de la versión 0 de la base, lista de indicadores presentes, inconsistencias y vacíos detectados	5
2.1.2	Incluye un esquema actualizado del modelo entidad–relación de la base de datos	6
2.2	Propuesta documentada de indicadores complementarios	7
2.2.1	Documento con la lista de indicadores adicionales (socioeconómicos y de democracia)	7
2.2.2	Fuentes consideradas (ENEMDU, Latinobarómetro, VDEM y otras)	8
2.2.3	Factibilidad de cálculo, variables requeridas, periodicidad y sustento metodológico	8
2.3	Rutinas automatizadas de descarga, procesamiento e ingestión de datos	10
2.3.1	Scripts en repositorio con control de versiones (GitHub/GitLab/otro)	11
2.3.2	Código comentado y estructurado	12
2.3.3	Documento técnico de dependencias, instrucciones de ejecución y mantenimiento de los scripts	12
2.4	Informe de validación metodológica y cálculo de indicadores	14
2.4.1	Documento que muestre la validación de las metodologías aplicadas	14
2.4.2	Ejemplos de cálculos realizados y resultados verificados en la base	14
2.5	Inventario final de indicadores	16
2.5.1	Tabla de metadatos consolidada (nombre, descripción, fórmula, fuente, frecuencia de actualización, fecha de última actualización, responsable)	17
2.5.2	Documento explicativo sobre criterios de selección y exclusión	19

1 Antecedentes

Este informe se enmarca en el proyecto de investigación titulado **Efectos de la desigualdad socioeconómica en la percepción de la democracia en Latinoamérica: periodo 1995-2023**, que tiene como objetivo identificar los factores estructurales que explican esta percepción negativa, considerando el impacto de eventos políticos y económicos clave, como crisis económicas y cambios de gobierno.

En virtud del contrato civil de prestación de servicios profesionales **NRO. UITEY-DGTH-LOES-2025-0097-SP**, la Universidad de Investigación de Tecnología Experimental Yachay contrató al Ing. **Patricio Joshue Mendoza Núñez** para apoyar en la ejecución del proyecto.

1.1 INEC – Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU)

La Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) es una operación estadística periódica realizada por el Instituto Nacional de Estadística y Censos (INEC) del Ecuador. Su objetivo es proporcionar información continua y actualizada sobre el comportamiento del mercado laboral ecuatoriano, incluyendo indicadores clave como tasa de empleo, subempleo, informalidad y condiciones socioeconómicas.

Justificación de uso en el proyecto La ENEMDU constituye la principal fuente oficial de datos socioeconómicos comparables en el Ecuador. Permite construir indicadores robustos sobre condiciones de vida, pobreza, empleo e informalidad, los cuales serán utilizados como variables explicativas en los análisis multivariados que buscan relacionar desigualdad socioeconómica con percepción democrática.

Su periodicidad y nivel de detalle permiten un análisis continuo y se puede complementar con la información sobre el Ministerio de Finanzas para el cálculo de indicadores socioeconómicos.

Origen y organización de los datos Los datos utilizados en este proyecto provienen del Banco de Datos Abiertos del Instituto Nacional de Estadística y Censos (INEC)¹. Esta plataforma pública permite consultar y descargar información estadística oficial del Ecuador en diferentes dominios, entre ellos trabajo, educación, salud y vivienda.

En particular, la sección correspondiente a trabajo incluye varios conjuntos de datos. Para este estudio se descargaron los siguientes conjuntos de datos principales:

- **ENEMDU Personas:** base de datos individual con información socioeconómica y laboral de los encuestados.
- **ENEMDU Vivienda:** contiene características estructurales y de servicios del hogar de los entrevistados.
- **Manual del encuestador:** documento que explica las variables, estructura de la encuesta y proceso de recolección.
- **Archivos de codificación:** tablas de referencia para interpretar las variables categóricas de la encuesta.
- **Fichas metodológicas:** instructivos para el cálculo de los indicadores socioeconómicos.

Estos datos se constituyen como la base empírica para el análisis posterior, ya que contienen información detallada tanto del entorno del hogar como de las condiciones laborales de los individuos encuestados.

¹<https://aplicaciones3.ecuadorencifras.gob.ec/BIINEC-war/index.xhtml>

1.2 Latinobarómetro

El Latinobarómetro es una encuesta de opinión pública aplicada periódicamente en países de América Latina. Para este proyecto se utilizará a nivel país-año, empleando los identificadores X_23 (país) y X_24 (año). Dado que los ítems pueden cambiar de denominación entre olas (p.ej., p20 en un año y p31 en otro), se construirá un *diccionario maestro* que mapee temáticas comparables (apoyo y satisfacción con la democracia, confianza institucional, percepción de corrupción y participación cívica) a las columnas efectivas de cada ola. Las tasas se calcularán con ponderadores muestrales cuando estén disponibles, documentando universos válidos y tratamiento de no respuesta.

Justificación del proyecto. La incorporación de Latinobarómetro permite capturar percepciones y actitudes ciudadanas que no están disponibles en fuentes institucionales (p.ej., apoyo a la democracia, confianza, percepción de corrupción, participación). Estos indicadores son complementarios a los índices estructurales de democracia (p.ej., V-Dem), y son críticos para entender mecanismos de legitimidad, consentimiento y desempeño subjetivo del régimen, así como para generar alertas tempranas sobre erosiones de apoyo o tolerancia al autoritarismo a nivel país-año.

Origen y organización de los datos. Para cada ola anual, se descarga el archivo de encuesta en formato .dta (Stata) y su correspondiente *libro de códigos*. Se realiza una normalización por ola mediante un *diccionario maestro* (mapa ola → variable) que alinea las preguntas comparables a través del tiempo. El procesamiento incluye: (i) armonización de categorías de respuesta, (ii) definición de universos válidos por ítem, (iii) tratamiento de no respuesta y (iv) cálculo con ponderadores muestrales. Los resultados se consolidan a nivel país-año y se documentan metadatos de cobertura (margen de error, tamaño muestral, notas de campo) cuando estén disponibles.

1.3 V-Democracia

V-Dem provee series anuales país-año de índices compuestos de calidad democrática en distintas dimensiones. Se integrarán al tablero los índices de alto nivel v2x_polyarchy (electoral), v2x_libdem (liberal), v2x_partipdem (participativa), v2x_delibdem (deliberativa) y v2x_egaldem (igualitaria), así como los índices de nivel medio especificados en la hoja de códigos. Cuando existan intervalos de codificación (valores *low/high*), se reportarán como bandas de incertidumbre. Todas las series se trabajarán en su escala estandar en [0, 1] y se documentará la variación interanual Δ y promedios móviles (3-5 años) para alertas tempranas.

Justificación del proyecto. V-Dem proporciona una medición sistemática y desagregada de la democracia en múltiples dimensiones, con metodología transparente y replicable. Su inclusión permite: (i) disponer de una base comparativa de *alto nivel* (p.ej., democracia liberal, participativa) y (ii) abrir la caja de mecanismos mediante indicadores de *nivel medio* (p.ej., libertad de expresión, independencia judicial). Esto habilita el análisis de trayectorias, rupturas y convergencias entre dimensiones, y su contraste con actitudes ciudadanas provenientes de Latinobarómetro.

Origen y organización de los datos. Los datos consolidados de V-Dem se descargan en un archivo .csv con cobertura país-año. La organización del repositorio de trabajo incluye: (i) una lista blanca de indicadores a integrar (alto y medio nivel), (ii) un esquema de validación de

rangos y de valores perdidos, (iii) cálculo de transformaciones auxiliares (variación interanual Δ , promedios móviles), y (iv) metadatos de versión/documentación. El CSV se integra como fuente canónica para el tablero y la matriz de indicadores, preservando los nombres de columnas originales para trazabilidad.

2 Desarrollo

Durante octubre de 2025 se desarrollaron las siguientes actividades conforme a los entregables descritos en el contrato de prestación de servicios. A continuación, se presenta un desarrollo explicativo y detallado de los avances en la revisión de la base de datos y en la propuesta de indicadores complementarios de democracia.

2.1 Informe de revisión inicial de la base de datos

2.1.1 Documento que describa la situación de la versión 0 de la base, lista de indicadores presentes, inconsistencias y vacíos detectados

Panorama general del dataset v0. La arquitectura actual comprende tres fuentes: (i) ENEMDU (microdatos de personas y hogares, con sus correspondientes agregaciones por unidad geográfica y periodo), (ii) Latinobarómetro (módulo de preguntas codificadas por año) y (iii) V-Dem (serie país-año con índices compuestos de calidad democrática). En términos operativos, la consolidación se apoya en tablas base, *materialized views* para agregados y vistas de salida para tableros. La estrategia de tipos de datos privilegia el uso de `Nullable(...)` para preservar valores faltantes y evitar sesgos en cálculos posteriores.

Claves de integración temporal y geográfica. Para ENEMDU, la granularidad temporal está dada por `periodo` (formato YYYYMM), derivándose `año` y `mes` cuando se requiere. En Latinobarómetro, los identificadores de `país` y `año` existen y se encuentran codificados en las columnas `X_23` (país) y `X_24` (año). En V-Dem la unidad analítica es país-año. La integración inter-fuente se plantea, por tanto, al menos a nivel nacional y anual; las desagregaciones subnacionales se restringen a ENEMDU.

Diccionarios de códigos y estandarización semántica. Las tres fuentes requieren un diccionario de interpretación para traducir códigos a categorías analíticas: en ENEMDU, los diccionarios ya están incorporados en la metodología de indicadores (p.,ej., empleo/informalidad/pobreza); en V-Dem y Latinobarómetro se debe completar la definición de indicadores de democracia a partir de sus variables y escalas originales. En el caso de Latinobarómetro, además, *la codificación de preguntas varía por ola*: la misma temática puede cambiar de nombre (por ejemplo, p20 en un año y p31 el siguiente). Por ello, la normalización de ítems es **manual y curada**, sustentada en un mapeo de equivalencias por año y versión del cuestionario.

Indicadores presentes y vacíos.

- **ENEMDU:** se cuenta con tasas y proporciones estándar (participación, desempleo, empleo total/formal/informal, empleo adecuado, pobreza por ingresos y extrema, brechas por sexo/edad), además de indice de pobreza multidimensional (IPM) y necesidades básicas insatisfechas (NBI) a nivel de hogar. Estos indicadores ya están calculados conforme a metodología establecida.

- **Latinobarómetro:** existen las respuestas codificadas de múltiples preguntas (A_1, A_1-1, ..., X_27), y los identificadores X_23 (país) y X_24 (año). Falta homogenizar ítems de democracia entre olas para construir indicadores comparables.
- **V-Dem:** la tabla contiene índices compuestos (p.,ej. v2x_polyarchy, v2x_libdem, v2x_partipdem, entre otros). Falta seleccionar y derivar los indicadores que se incorporarán al tablero, y definir reglas de lectura (p.,ej., bandas de incertidumbre).

Inconsistencias y riesgos de calidad (selección).

- **Latinobarómetro:** heterogeneidad de nombres de preguntas por año; se requiere un mapeo de equivalencias por ola y un *codebook* maestro para normalizar series de democracia (apoyo, satisfacción, confianza institucional).
- **V-Dem:** verificación de correspondencias país–año con el universo de interés para Ecuador y la consistencia de series en años límite.
- **Todas las fuentes:** presencia de valores nulos o categorías no aplicables; se sugiere tratamiento explícito con funciones nullIf/ifNull en cada cálculo, y *chequeos de cobertura* por periodo y dominio geográfico.

2.1.2 Incluye un esquema actualizado del modelo entidad–relación de la base de datos

Descripción narrativa del modelo. El modelo conceptual se organiza en tres entidades principales:

1. **ENEMDU** (persona y vivienda/hogar), con relaciones internas por periodo y claves operativas de hogar/persona, y con agregaciones por nivel geográfico (*nacional, área, provincia, cantón, parroquia*) y año.
2. **Latinobarómetro** (unidad: individuo–encuesta), con identificación de X_23 (país) y X_24 (año), y un conjunto de ítems codificados que cambian entre olas; su integración se prevé por país–año y comparabilidad temática normalizada.
3. **V-Dem** (unidad: país–año), con índices de democracia electoral, liberal, participativa, deliberativa e igualitaria; su integración se hará por país–año, con posibilidad de comparativos regionales.

La interoperabilidad entre fuentes se logra en el eje **país–año** (Latinobarómetro–V-Dem) y, cuando se requiera contraste con condiciones socioeconómicas, se acota al **nivel nacional/anual** de ENEMDU para asegurar comparabilidad.

Table 1: Esquema lógico (campos principales) por fuente y notas de interpretación

Tabla/Fuente	Campos principales	Notas y relación con preguntas/indicadores
ENEMDU (Persona)	periodo, anio, mes, area, geo, id_hogar, id_persona, fexp, variables laborales (ocupación, condición de actividad), ingresos (ingpc, ...)	Indicadores estándar ya calculados (participación, desempleo, informalidad, empleo adecuado). Diccionarios de códigos ya incorporados en la metodología.
ENEMDU (Vivienda/Hogar)	periodo, anio, area, geo, id_hogar, carencias (NBI), componentes IPM, características de vivienda	Agregación a nivel de hogar para IPM/NBI y cruces geográficos. Control de pesos y consistencia hogar-persona.
Latinobarómetro	X_23 (país), X_24 (año), ítems de encuesta: A_1, A_1-1, A_2, ..., X_27	Las <i>mismas temáticas</i> cambian de nombre entre olas (p.,ej., <i>apoyo a la democracia</i> puede ser p20 o p31). Requiere normalización manual por ola con diccionario maestro.
V-Dem	country_id/country_text_id, year, índices: v2x_polyarchy, v2x_libdem, v2x_partipdem, v2x_delibdem, v2x_egaldem, (...)	Serie país-año. Los indicadores de democracia se derivarán a partir de estos índices, incorporando bandas de incertidumbre cuando aplique.

Observaciones de diseño.

- El **eje temporal** se estandariza en **anio** para cruce Latinobarómetro–V-Dem; en ENEMDU se preserva **periodo** (YYYYMM) para explotar estacionalidad y luego agregar a año.
- La **interpretación de códigos** se realiza mediante diccionarios por fuente. En ENEMDU están incorporados; en Latinobarómetro y V-Dem se completará como parte del cálculo de indicadores de democracia.
- **Diagrama ER:** diagrama ER formal (PDF) con cardinalidades exactas y tablas auxiliares (dimensiones de tiempo y geografía). Ver diagrama completo en [Modelo ER](#).

2.2 Propuesta documentada de indicadores complementarios

2.2.1 Documento con la lista de indicadores adicionales (socioeconómicos y de democracia)

Alcance y criterio de selección. De acuerdo con la metodología actual del proyecto, **no se proponen nuevos indicadores basados en ENEMDU** (a excepción del indicador "Tasa de Escolaridad"), dado que éstos ya se producen conforme a una metodología oficial establecida. En su lugar, se plantean **indicadores de democracia** derivados de V-Dem y Latinobarómetro, con el objeto de: (i) capturar la evolución institucional (V-Dem) y (ii) aproximar percepciones y actitudes ciudadanas (Latinobarómetro), habilitando análisis comparados y cruces narrativos con el contexto socioeconómico cuando corresponda (a nivel nacional/anual).

Propuesta de indicadores de democracia.

1. **Democracia electoral (Polyarchy)**: nivel anual (`v2x_polyarchy`) y variación interanual (Δ). Permite monitorear la base electoral de la democracia.
2. **Democracia liberal**: nivel anual (`v2x_libdem`), con *banda de incertidumbre* cuando existan estimaciones baja/alta. Útil para alertas de retrocesos.
3. **Democracia participativa, deliberativa e igualitaria** (`v2x_partipdem`, `v2x_delibdem`, `v2x_egaldem`): tendencias y comparación con benchmarks regionales.
4. **Apoyo/satisfacción con la democracia** (Latinobarómetro): porcentaje por categoría y cambio interanual, a partir de *normalización manual* de ítems por ola (p.,ej., *apoyo a la democracia* con `p20/p31` según año).
5. **Confianza institucional** (Latinobarómetro): niveles de confianza en instituciones clave (Congreso, partidos, sistema judicial), armonizados por ola para construir series homogéneas.
6. **Alertas de deterioro democrático**: bandera si $\Delta \leq -X$ p.p. en cualquiera de los subíndices V-Dem respecto del promedio móvil de 3–5 años.

2.2.2 Fuentes consideradas (ENEMDU, Latinobarómetro, VDEM y otras)

ENEMDU (INEC). Se utiliza como contexto socioeconómico a nivel nacional/anual cuando se requiera contrastar narrativas (sin generar nuevos indicadores fuera de la metodología oficial).

Latinobarómetro. Encuesta de opinión pública con identificadores `X_23` (país) y `X_24` (año). Para construir series comparables, se elaborará un **diccionario maestro por ola** que mapee las preguntas temáticas (apoyo, satisfacción, confianza) a las columnas efectivas de cada año (p.,ej., `p20`, `p31`, ...). La normalización es **manual** y documentada.

V-Dem. Base país-año con índices de calidad democrática (electoral, liberal, participativa, deliberativa, igualitaria). Se emplean los valores principales y, cuando existan, sus intervalos de codificación para reflejar incertidumbre. La lectura se realiza en la escala estándar de V-Dem (0–1).

2.2.3 Factibilidad de cálculo, variables requeridas, periodicidad y sustento metodológico

Visión general de factibilidad. La factibilidad es **alta** para V-Dem (valores listos por país-año) y **media-alta** para Latinobarómetro (por la necesidad de normalización manual de preguntas y verificación de ponderadores por ola). La periodicidad natural es **anual** para V-Dem y Latinobarómetro, mientras que para ENEMDU la periodicidad es mensual (trimestral en años anteriores).

Variables requeridas (síntesis).

- **ENEMDU**: `periodo`, `zona`, `sector`, `vivienda`, `ciudad`, `fexp` y otras preguntas para el cálculo de indicadores
- **V-Dem**: `country_text_id` o `country_id`, `year`, `v2x_polyarchy`, `v2x_libdem`, `v2x_partipdem`, `v2x_delibdem`, `v2x_egaldem` (y variantes de incertidumbre cuando existan).

- **Latinobarómetro**: X_23 (país), X_24 (año), ítems de *apoyo, satisfacción y confianza institucional* según la ola (p.ej., p20/p31 para apoyo).

Periodicidad.

- **ENEMDU**: Trimestral en años anteriores, mensual desde 2021.
- **V-Dem**: anual (serie continua).
- **Latinobarómetro**: según ola; los resultados se reportarán en serie anual con nota metodológica cuando falten olas.

Sustento metodológico.

- **ENEMDU**: metodología heredada desde la INEC.
- **V-Dem**: series en escala [0,1]; se informará la variación interanual (Δ), promedios móviles (3–5 años) y bandas de incertidumbre cuando apliquen.
- **Latinobarómetro**: construcción de tasas/porcentajes a partir de ítems normalizados por ola; uso de ponderadores muestrales si están disponibles. Se documentará el **diccionario maestro** (ola → variable), las categorías de respuesta y los tratamientos de valores faltantes/no respuesta.

Table 2: Plan de cálculo para indicadores de democracia

Indicador	Variables requeridas	Periodicidad	Estado/Notas
Democracia electoral (Polyarchy)	V-Dem: country_id, year, v2x_polyarchy	Anual	Listo para cálculo
Democracia liberal (+ banda)	V-Dem: v2x_libdem, (low/high si existe)	Anual	Definir umbrales de alerta
Participativa/ Delib./ Igualitaria	V-Dem: v2x_partipdem, v2x_delibdem, v2x_egaldem	Anual	Listo para cálculo
Apoyo/Satisfacción democracia	Latinobarómetro: X_23, X_24	Anual	Requiere normalización manual
Confianza institucional	Latinobarómetro: ítems por institución + ponderadores	Anual	Requiere codebook final
Alertas de deterioro	V-Dem: de subíndices + regla de umbral	Anual	Requiere revisión de variables

Tabla de planificación y requerimientos técnicos (resumen).

Notas finales de implementación.

- Se mantendrá la **compatibilidad semántica** entre olas de Latinobarómetro mediante un *diccionario maestro versionado*; cada versión documentará qué columnas concretas alimentan cada indicador.

- Para V-Dem, los indicadores se derivarán directamente de sus índices estandarizados y, cuando corresponda, se publicarán junto con **bandas de variabilidad** para transparencia.
- Para análisis narrativos con contexto socioeconómico, se mantendrá el **nivel nacional/anual** de ENEMDU, evitando extrapolaciones subnacionales que no correspondan a la unidad país-año de V-Dem y Latinobarómetro.
- Para los indicadores calculados para ENEMDU, se incluye (i) la metodología correspondiente para cada indicador (definiciones, universos y ponderadores) y (ii) matriz comparativa de los indicadores calculados vs los indicadores reales.

2.3 Rutinas automatizadas de descarga, procesamiento e ingestión de datos

Esta sección documenta la arquitectura de automatización del proyecto, el flujo de trabajo desde la captura hasta la exposición analítica, y los lineamientos de operación y mantenimiento. La solución está **contenedora** y orquestada con `docker compose`; el despliegue estándar se realiza con:

```
docker compose --env-file .env up --build
```

donde `.env` define credenciales, rutas, y parámetros de ejecución. El repositorio se organiza en módulos claramente separados para descarga, limpieza, ingestión, docs y *web app*.

Estructura del repositorio y responsabilidades. La automatización se encuentra en `./automatization` y cubre:

- `download_scripts`: descarga y limpieza inicial (ENEMDU, VDEM, Latinobarómetro).
- `ingest`: cálculo de indicadores, serialización y carga a ClickHouse; registro de `logs` y `errors`.
- `init-scripts`: inicialización de ClickHouse (`create_table.sql`) y Superset (`init_superset_db.py`, `superset_config.py`).

Los datos se versionan por estado y fuente en `./data`:

- `raw/ENEMDU`, `raw/latinobarometro`, `raw/VDEM`: *snapshots* originales por año.
- `enemdu_persona/enemdu_vivienda`: `unprocessed` → `processed`.
- `latinobarometro`: `unprocessed_dta` / `unprocessed_csv` → `processed_csv` / `processed`.

Los volúmenes persistentes para servicios se ubican en `./volumes/clickhouse` y `./volumes/superset`.

Servicios en docker compose y flujos de datos. La orquestación define, al menos, un servicio de descarga/limpieza y servicios de base de datos y *BI*. A modo de referencia:

- `automatic_download`: construye desde `./automatization/download_scripts`, monta volúmenes de datos crudos y *staging*, y ejecuta de forma secuencial los scripts de descarga y limpieza (`enemdu_descarga.py`, `vdem_descarga.py`, `latinobarometro_descarga.py`, `limpieza_.py`).

- **ClickHouse** y **Superset**: inicializables con los scripts en `init-scripts`; exponen almacenamiento columnar y visualización, respectivamente.
- `ingest`: construye desde `./automation/ingest`, ejecuta `ingest_.py`, registra logs y errors.

Table 3: Componentes de automatización y responsabilidades

Módulo/Servicio	Propósito	Entrada / Salida	Observaciones
<code>automatic_download</code>	Descarga y limpieza inicial (ENEMDU, VDEM, Latino)	Entrada: data/raw/; Salida: data/processed	Secuencia de <code>_descarga.py</code> y <code>limpieza_.py</code>
<code>ingest</code>	Cálculo e ingestión en ClickHouse	Entrada: processed; Salida: tablas ClickHouse	<code>ingest_.py</code> , <code>logs/errors</code>
<code>clickhouse_server</code>	Almacenamiento analítico	DDL en init-scripts, clickhouse	<code>create_table.sql</code> ; particiones/orden
<code>superset</code>	Exposición de tableros	Conexión a ClickHouse	<code>init_superset_db.py</code> , <code>superset_config.py</code>

2.3.1 Scripts en repositorio con control de versiones (GitHub/GitLab/otro)

El proyecto emplea control de versiones para garantizar trazabilidad y revisiones por *pull requests*. La carpeta `automation` contiene:

- `download_scripts`:
 - `enemdu_descarga.py`, `vdem_descarga.py`, `latinobarometro_descarga.py`: descarga por API/URL, gestión de años y reintentos.
 - `limpieza_persona.py`, `limpieza_vivienda.py`, `limpieza_vdem.py`, `limpieza_latinobarometro.py`: homogeneización de columnas, tipos, y normalización de valores.
 - `Dockerfile`, `requirements.txt`: *runtime* reproducible.
- `ingest`:
 - `ingest_persona.py`, `ingest_vivienda.py`, `ingest_vdem.py`, `ingest_latinobarometro.py`: serialización de `processed` a ClickHouse.
 - `ingest_codigos.py`: ingestas temáticas.
 - `Dockerfile`, subcarpetas `logs/errors`.
- `init-scripts`:
 - `clickhouse/create_table.sql`: DDL base.
 - `superset/init_superset_db.py`, `superset_config.py`: inicialización y configuración.

El repositorio se encuentra en el siguiente enlace a [Github](#).

2.3.2 Código comentado y estructurado

El código sigue una estructura por **etapas de *data engineering***: *extract* (*_descarga.py*), *clean* (*limpieza_.py*) y *load* (*ingest_.py*). Las funciones se documentan con *docstrings* descriptivos, manejo explícito de errores (registro en *automatization/ingest/errors*) y logging con marcas temporales (en *automatization/ingest/logs*).

- **Nomenclatura:** prefijos semánticos (*limpieza_*, *ingest_*, *calcular_**) y separación por dominio (persona, vivienda, vdem, latinobarometro).
- **Tipificación:** conversión explícita a tipos ClickHouse (*Nullable(...)*) y validaciones de integridad (longitud de códigos, dominios válidos).
- **Idempotencia:** scripts reentrables; no repiten carga si los artefactos de *staging* ya existen (o reescriben según *flag*).

2.3.3 Documento técnico de dependencias, instrucciones de ejecución y mantenimiento de los scripts

Dependencias.

- **Docker Compose:** versión 28.3.0, build 38b7060.
- **Python:** versión definida dentro de cada contenedor. Paquetes base definidos en *automatization/download_scripts/requirements.txt* y *automatization/ingest/Dockerfile*.
- **ClickHouse:** versión del servidor (latest: 25.9-alpine) e imagen (*clickhouse/clickhouse-server:latest*).
- **Superset:** versión de la imagen (*apache/superset:latest*).

Instrucciones de ejecución.

1. Configurar el archivo *.env* en la raíz del proyecto (endpoints, credenciales, rutas locales y de contenedor).
2. Construir e iniciar los contenedores:

```
docker compose --env-file .env up --build
```

3. Verificar logs:

- Contenedor de descarga/limpieza: `docker logs -f automatic_download`
- Contenedor de ingesta: `docker logs -f ingest`.

4. Confirmar estructuras en ClickHouse (*init-scripts/clickhouse/create_table.sql*) y conectar Superset (*init-scripts/superset*).

Mantenimiento y operación.

- **Rotación de logs/errors:** política de retención y compresión.
- **Data refresh:** cron de docker compose run o servicio automatic_download con parámetros de ventana temporal; recomendación: batch mensual para ENEMDU y anual/por ola para Latinobarómetro; V-Dem anual.
- **Versionado de schemas:** cambios en DDL deben versionarse en init-scripts/clickhouse con migrations numeradas (p.,ej., V001_...sql); documentar cambios en docs/CHANGELOG.md.
- **Monitoreo de calidad:** checks automáticos post-ingesta (conteos esperados por año/área, cobertura de nulos, dominios válidos). Resultados en una tabla de control o reporte en docs.

Table 4: Flujo E2E: extracción, normalización, ingestión y publicación

Etapa	Scripts clave	Entradas → Salidas	Puntos de control
Extracción	enemdu_descarga.py, vdem_descarga.py, latinobarometro_descarga.py	raw/ unprocessed/	→ Conteos por año, hash de archivos
Limpieza y Normalización	limpieza_persona.py, limpieza_vivienda.py, limpieza_vdem.py, limpieza_latinobarometro.py	unprocessed/ processed/	→ Tipos, diccionarios, mapeos por ola (Latino)
Ingesta	ingest_persona.py, ingest_vivienda.py, ingest_latinobarometro.py, ingest_vdem.py, ingest_codigos.py	processed/* ClickHouse	→ DDL aplicado, ORDER BY, deduplicación
Derivación Indicadores	MV de ClickHouse	Tablas base → vistas de salida	Validación con cifras oficiales
Publicación	Superset (init_superset_db.py)	ClickHouse → dashboards	Permisos, caché, ALLOWED_DASHBOARDS

Flujo de extremo a extremo (E2E) resumido.

Notas específicas para Latinobarómetro y V-Dem.

- **Latinobarómetro:** los identificadores X_23 (país) y X_24 (año) están presentes; las **preguntas cambian entre olas** y se normalizan manualmente (mapeo p20/p31/...a la temática “apoyo a la democracia”, etc.). Esta normalización alimenta el cálculo de *indicadores de democracia* (no presentes aún) junto con un diccionario maestro de respuestas.
- **V-Dem:** los índices compuestos (v2x_polyarchy, v2x_libdem, v2x_partipdem, v2x_delibdem, v2x_egaldem) permiten derivar indicadores de democracia anuales; se recomienda documentar bandas de incertidumbre cuando existan variables *low/high*.

Documentación técnica adjunta.

- **Guía de ejecución:** este documento + README.md con parámetros, ejemplos y .env de referencia.
- **Dependencias:** requirements.txt y versiones de imágenes en docker-compose.yml.

2.4 Informe de validación metodológica y cálculo de indicadores

Esta sección documenta la **validación de las metodologías** empleadas para el cálculo de indicadores y presenta **ejemplos trazables de cómputo** que pueden reproducirse directamente sobre las tablas agregadas por nivel geográfico y período. La validación se sustenta en dos insumos: (i) una *matriz de indicadores* con sus fichas metodológicas, periodicidad y fuente; y (ii) un *reporte de contraste* que compara indicadores calculados vs. cifras de referencia (“reales” u oficiales), cuantificando el porcentaje de error.

2.4.1 Documento que muestre la validación de las metodologías aplicadas

Enfoque de validación. El proceso se organiza en cuatro etapas, con resultados verificables y trazables a nivel de consulta:

1. **Definición y homologación:** se elabora una *matriz de indicadores* que fija nombre, unidad de medida, fórmula, periodicidad, universo, fuente y supuestos (tratamiento de nulos, denominadores válidos, ponderadores, filtros etarios/ocupacionales).
2. **Implementación reproducible:** las fórmulas se codifican en vistas y/o consultas ClickHouse, aplicando explícitamente nullIf/ifNull y denominadores válidos (*valid universe*) para cada indicador.
3. **Contraste con referencia:** para cada indicador y período, se compara el valor calculado con la cifra de referencia y se reporta el error relativo.
4. **Criterios de aceptación:** se establecen umbrales por tipo de indicador (p.,ej., para tasas nacionales; para desagregaciones cantonales), documentando excepciones y causas (cambios muestrales, redondeos oficiales, ventanas temporales).

Matriz de indicadores. La tabla ilustra el formato consolidado de fichas metodológicas (para ver el documento completo véase [Matriz de Indicadores](#)).

Reporte de contraste (validación). El reporte presenta, para cada indicador–nivel–período, el valor calculado, el valor de referencia y el error relativo. Para ver el documento completo véase [Validación de Indicaadores INEC](#).

2.4.2 Ejemplos de cálculos realizados y resultados verificados en la base

Indicadores de persona (tasas y shares). Para cada combinación de *nivel geográfico* (nacional, provincia, cantón, parroquia) y *período*, se emplean los acumulados ponderados precomputados. Las tasas se calculan con los siguientes cocientes (reproducidos aquí para referencia):

Vistas de indicadores PERSONA A partir de `persona_sums` se exponen vistas por nivel geográfico con etiquetas de mes:

- **Nacional/Provincia/Cantón/Parroquia:** tasas y shares:

- $TPG = 100 \cdot \frac{sw_pea}{sw_pet}$,
- $TPB = 100 \cdot \frac{sw_pea}{sw_pop}$,
- Desempleo = $100 \cdot \frac{sw_pea - sw_occ}{sw_pea}$,
- Empleo = $100 \cdot \frac{sw_occ}{sw_pea}$,
- Formal = $100 \cdot \frac{s_formal_w}{sw_occ}$,
- Informal = $100 \cdot \frac{s_informal_w}{sw_occ}$,
- Adecuado = $100 \cdot \frac{s_emp_adecuado}{sw_pea}$,
- Subempleo = $100 \cdot \frac{s_subempleo}{sw_pea}$,
- NoRemu = $100 \cdot \frac{s_no_remu}{sw_pea}$,
- OtroNoPleno = $100 \cdot \frac{s_otro_no_pleno}{sw_pea}$,
- NiNi = $100 \cdot \frac{s_youth_nini}{s_youth}$,
- Desempleo Juvenil = $100 \cdot \frac{s_juv_des}{s_juv_pea}$,
- Trabajo Infantil = $100 \cdot \frac{s_ti}{s_kids}$,
- Manufactura = $100 \cdot \frac{s_occ_manu}{sw_occ}$,
- Pobreza Ingresos = $100 \cdot \frac{s_under_lp}{s_valid_lp}$,
- Pobreza Extrema = $100 \cdot \frac{s_under_lp_ext}{s_valid_lp}$.

- **Brechas por sexo:**

- Brecha Adecuado_{H-M} = $100 \cdot \frac{\frac{ade_h - ade_m}{ade_h}}{\frac{pea_h - pea_m}{pea_h}}$,
- Brecha Salarial_{H-M} = $100 \cdot \frac{\frac{occ_inc_w_h - occ_inc_w_m}{occ_inc_w_h}}{\frac{occ_w_h - occ_w_m}{occ_w_h}}$.

Se publican: `indicadores_persona_nacional`, `_provincia`, `_canton`, `_parroquia`, `_canton_anual`, `_nacional_por_periodo`, `_parroquia_por_periodo`.

Vistas de indicadores VIVIENDA Se publican vistas por nivel (`_nacional`, `_provincia`, `_canton`, `_parroquia`, `_nacional_por_periodo`, `_parroquia_por_periodo`, `_canton_anual`) con:

$$TPM = 100 \cdot \frac{\sum s_TPM}{\sum w_TPM}, \quad TPEM = 100 \cdot \frac{\sum s_TPEM}{\sum w_TPEM}, \quad NBI = 100 \cdot \frac{\sum s_NBI}{\sum w_NBI}.$$

Las vistas incluyen etiqueta de mes y respetan `area/geo` según `level`.

Buenas prácticas aplicadas en el cómputo.

- **Universos válidos y denominadores consistentes:** se calculan denominadores específicos por indicador (`sw_pea`, `sw_occ`, `s_valid_lp`, etc.), evitando sesgos por valores faltantes mediante `ifNull/nullIf`.
- **Ponderación:** todos los numeradores y denominadores emplean **pesos muestrales** correspondientes al nivel de agregación.
- **Trazabilidad:** cada resultado publicado incluye referencia al período, nivel y consulta que lo generó (bitácora de ejecución y `logs`).

Table 5: Ejemplos de resultados verificados por nivel y período

Indicador	Período	Nivel	Valor	Valor Ref.	% Error
Desempleo	2025-7	Nacional	3,2788	3,3	-0,64
Informalidad	2025-7	Nacional	51,3	51,2725	+0,05
TPM	2024-12	Nacional	40,8	40,8475	-0,12

Ejemplo de reporte de resultados verificados (formato).

Checklist de validación (para auditoría interna).

- Fórmulas revisadas y consistentes con la matriz de indicadores (*one source of truth*).
- Denominadores y universos válidos documentados por indicador, nivel y período.
- Tratamiento de nulos/no respuesta explícito y uniforme.
- Ponderadores aplicados y verificados.
- Bitácora de ejecución y `logs` archivados por corrida.
- Reporte de contraste generado con % de error y criterios de aceptación documentados.

Entregables asociados.

- **Matriz de indicadores (fichas)** — Véase [Matriz de indicadores](#).
- **Reporte de validación (calc vs ref)** — Véase [Reporte de validación](#).

2.5 Inventario final de indicadores

Esta sección presenta el **inventario consolidado de indicadores** empleados en el proyecto, con su tabla de metadatos (nombre, descripción, fórmula, fuente, frecuencia de actualización, fecha de última actualización y responsable) y un **documento explicativo de criterios** de selección y exclusión. La “Matriz de indicadores” constituye la *fuente maestra* de dicha tabla de metadatos; aquí se incluye su *plantilla LaTeX* y un extracto representativo, así como los criterios que guían el inventario en coherencia con el objetivo del proyecto.

2.5.1 Tabla de metadatos consolidada (nombre, descripción, fórmula, fuente, frecuencia de actualización, fecha de última actualización, responsable)

Table 6: Matriz de indicadores

Indicador	Fuente	Periodicidad	Descripción/Pregunta
Tasa Participación Global	ENEMDU Persona	Mensual	Porcentaje de la PEA sobre la PET (12+ según norma).
Tasa Participación Bruta	ENEMDU Persona	Mensual	Porcentaje de la PEA sobre la población total.
Tasa Desempleo	ENEMDU Persona	Mensual	Personas de 15 años y más en condición de desocupación.
Empleo Total	ENEMDU Persona	Mensual	Personas de 15 años y más ocupadas.
Formal	ENEMDU Persona	Mensual	Personas que trabajan en empleo formal (según definición oficial).
Informal	ENEMDU Persona	Mensual	Personas que trabajan en empleo informal (según definición oficial).
Adecuado	ENEMDU Persona	Mensual	Empleo adecuado (criterios de ingreso y horas trabajadas).
Subempleo	ENEMDU Persona	Mensual	Personas ocupadas con insuficiencia de horas/ingreso.
No Remun.	ENEMDU Persona	Mensual	Personas con empleo no remunerado.
Otro No Pleno	ENEMDU Persona	Mensual	Condición “otro no pleno” dentro de la PEA.
Empleo Adecuado Mujeres	ENEMDU Persona	Mensual	Porcentaje de empleo adecuado en mujeres (sobre PEA mujeres).
Empleo Adecuado Hombres	ENEMDU Persona	Mensual	Porcentaje de empleo adecuado en hombres (sobre PEA hombres).
Brecha Adecuado H–M	ENEMDU Persona	Mensual	Diferencia relativa de empleo adecuado entre hombres y mujeres.
Ingreso Medio Hombres	ENEMDU Persona	Mensual	Ingreso promedio de ocupados hombres.
Ingreso Medio Mujeres	ENEMDU Persona	Mensual	Ingreso promedio de ocupados mujeres.
Brecha Salarial H–M NiNi	ENEMDU Persona	Mensual	Diferencia relativa del ingreso promedio H vs M.
Desempleo Juvenil	ENEMDU Persona	Mensual	Población de 15–24 años que no estudia ni trabaja.
Trabajo Infantil	ENEMDU Persona	Mensual	Desocupación entre la población joven (universo joven).
Empleo Sector Manufactura	ENEMDU Persona	Mensual	Población de 5–14 años que realiza trabajo infantil.
Pobreza Por Ingresos	ENEMDU Persona	Mensual	Población ocupada en la rama de actividad manufactura.
			Población bajo la línea de pobreza oficial.

(continúa en la siguiente página)

Indicador	Fuente	Periodicidad	Descripción/Pregunta
Pobreza Extrema Por Ingresos	ENEMDU Persona	Mensual	Población bajo la línea de pobreza extrema oficial.
Tasa de Acceso a la Educación	ENEMDU Persona	Mensual	Cobertura educativa (definición operativa aplicada).
Pobreza Multidimensional	ENEMDU Persona + Vivienda	Mensual	Porcentaje de hogares pobres multidimensionales (IPM).
Pobreza Extrema Multidimensional	ENEMDU Persona + Vivienda	Mensual	Porcentaje de hogares en pobreza extrema multidimensional.
Necesidades Básicas Insatisfechas	ENEMDU Persona + Vivienda	Mensual	Porcentaje de hogares con al menos una NBI.
Apoyo a la democracia	Latinobarómetro	Anual	Porcentaje que declara preferir la democracia sobre cualquier otra forma de gobierno.
Satisfacción con la democracia	Latinobarómetro	Anual	Promedio (0–1) de satisfacción con el funcionamiento de la democracia en el país.
Confianza institucional (promedio)	Latinobarómetro	Anual	Promedio de proporciones con “mucha / algo” de confianza en instituciones seleccionadas.
Tolerancia al autoritarismo	Latinobarómetro	Anual	Porcentaje que considera aceptable un gobierno autoritario bajo ciertas circunstancias.
Participación cívica (índice)	Latinobarómetro	Anual	Media de participación auto-reportada en actividades (votar, marchas, firmas, reuniones).
Percepción de corrupción alta	Latinobarómetro	Anual	Porcentaje que percibe altos niveles de corrupción en instituciones del país.
Evaluación económica positiva	Latinobarómetro	Anual	Porcentaje que evalúa como buena/muy buena la situación económica del país u hogar.
Electoral democracy index (D)	V-DEM	Anual	To what extent is the ideal of electoral democracy in its fullest sense achieved?
Liberal democracy index (D)	V-DEM	Anual	To what extent is the ideal of liberal democracy achieved?
Participatory democracy index (D)	V-DEM	Anual	To what extent is the ideal of participatory democracy achieved?
Deliberative democracy index (D)	V-DEM	Anual	To what extent is the ideal of deliberative democracy achieved?
Egalitarian democracy index (D)	V-DEM	Anual	To what extent is the ideal of egalitarian democracy achieved?

(continúa en la siguiente página)

Indicador	Fuente	Periodicidad	Descripción/Pregunta
Clean elections index (M)	V-DEM	Anual	Extent to which elections are free and fair, free from irregularities and coercion.
Freedom of expression index (M)	V-DEM	Anual	Guarantees for alternative sources of information and freedom of expression.
Freedom of association index (M)	V-DEM	Anual	Extent of associational and party freedom and civil society organization.
Judicial independence (M)	V-DEM	Anual	Perceived independence of the judiciary from government influence.
Equality before the law & individual liberty (M)	V-DEM	Anual	Protection of civil liberties and equal treatment under the law.
Executive constraints (M)	V-DEM	Anual	Institutional checks on the executive by legislature, judiciary, and media.
Civil liberties index (M)	V-DEM	Anual	Respect for core civil liberties (expression, movement, religion, etc.).

2.5.2 Documento explicativo sobre criterios de selección y exclusión

El inventario final se rige por criterios **alineados con el objetivo del proyecto *Efectos de la desigualdad socioeconómica en la percepción de la democracia en Latinoamérica: periodo 1995–2023***. Los criterios garantizan *relevancia sustantiva, comparabilidad temporal y espacial, calidad metodológica y factibilidad operativa*.

Criterios de selección (inclusión).

1. **Pertinencia sustantiva** (*theory-driven*): el indicador captura dimensiones clave del nexo desigualdad–percepción democrática (p.,ej., empleo, pobreza por ingresos y multidimensional, así como actitudes/valores democráticos e índices institucionales).
2. **Cobertura temporal (1995–2023)**: el indicador presenta series suficientemente extensas o con continuidad razonable para análisis intertemporal.²
3. **Cobertura espacial (Latinoamérica)**: disponibilidad para el universo regional o, al menos, para Ecuador con posibilidad de comparación regional (V-Dem/Latinobarómetro a nivel país–año).
4. **Comparabilidad inter-ola/inter-anual**: definiciones consistentes; en Latinobarómetro, **normalización manual** por ola para preservar el constructo (apoyo, satisfacción, confianza).

²Para Latinobarómetro se respetan vacíos de ola; para V-Dem, continuidad anual; para ENEMDU, consolidación anual para comparabilidad.

5. **Calidad y transparencia metodológica:** indicadores basados en fuentes con documentación pública de muestreo, instrumentos y construcción (p.,ej., *codebooks*, fichas V-Dem).
6. **Robustez y replicabilidad:** existencia de fórmula explícita, universos válidos, ponderadores y tratamiento de nulos; reproducible a partir de la *Matriz* y los scripts de ingesta/cálculo.
7. **Relevancia analítica y parsimonia:** priorización de indicadores con alto poder explicativo sobre el *nexo* de interés, evitando redundancia excesiva (p.,ej., escoger un conjunto canónico de índices V-Dem).
8. **Actualización y mantenibilidad:** frecuencia de actualización conocida (mensual/anual/ola) y factible dentro de las rutinas automatizadas (`docker compose`).
9. **Integridad y derechos de uso:** disponibilidad licenciataria para su uso analítico y publicación en tableros (cuando aplique).

Nota: El documento con la información completa se encuentra en [Matriz de indicadores](#).

Criterios de exclusión.

1. **Baja pertinencia** con el *nexo* desigualdad–percepción democrática (indicadores colaterales sin relación teórica o empírica clara).
2. **Cobertura insuficiente** en el periodo 1995–2023 o con *gaps* que impidan análisis de tendencias mínimas.
3. **Incomparabilidad** por cambios definicionales no corregibles (p.,ej., ítems Latinobarómetro sin mapeo consistente entre olas).
4. **Calidad metodológica cuestionable:** falta de documentación de muestreo, sesgo de no respuesta sin mitigación o ausencia de ponderación adecuada.
5. **Redundancia alta:** indicadores altamente colineales o duplicados en su significado, cuando ya existe un representante canónico (parquedad del tablero).
6. **Mantenibilidad limitada:** imposibilidad práctica de actualización con las rutinas existentes o licencias restrictivas para su explotación.

Organización del inventario por dominios. Para facilitar la lectura y la toma de decisiones, el inventario se agrupa en tres dominios que reflejan el objetivo del proyecto:

- **Condiciones socioeconómicas (ENEMDU):** participación, empleo/desempleo, informalidad, pobreza por ingresos, IPM/NBI (publicados a nivel nacional/anual para comparabilidad con democracia).
- **Calidad democrática (V-Dem):** *polyarchy*, democracia liberal/participativa/deliberativa/igualitaria, con variación interanual y, cuando esté disponible, bandas de incertidumbre.
- **Percepción democrática (Latinobarómetro):** apoyo, satisfacción y confianza institucional, normalizados manualmente por ola (país–año), con documentación explícita del diccionario maestro.

Fecha de entrega:	31 de octubre de 2025
--------------------------	-----------------------

RUBRO	CARGO	FIRMA
ELABORADO POR:	Patricio Mendoza, Ing. (Técnico Especialista)	
APROBADO POR:	Erick Cuenca, PhD. (Director del Proyecto)	