

# Acceso al Agua Potable en Países de Latinoamérica y África (2012–2022)

Autor: Arevalo, Juan Pablo

**Coderhouse - Data Analytics /  
Comisión 79090**

# Índice

## Hipótesis

Comprender y analizar, durante el período 2012–2022, la brecha existente entre los diez países latinoamericanos y africanos en el acceso básico al agua potable.

Además, identificar los países más vulnerables y la evolución de crecimiento o decrecimiento de su nivel de acceso en dicho período, como también las diferentes brechas

## Objetivo

La escasez de agua potable es una problemática central que afecta a millones de personas en el mundo, particularmente en regiones de América Latina y África. El acceso al agua potable es un factor clave en el desarrollo humano y está directamente vinculado con la salud, la calidad de vida y el crecimiento de una población.

Este proyecto busca realizar un análisis entre las diferencias en el acceso al agua potable de 10 (diez) países latinoamericanos y 10 (diez) africanos durante el período 2012–2022.

La limitación de diez países se debe a que en el dataset base hay muchos actores que no tienen los valores completos, a partir de la simplificación de estos datos se construyó un dataset depurado y enriquecido donde cada fila representa la combinación país, año e indicador, donde incluye tanto el valor original del indicador como una serie de variables derivadas que permiten entender la dinámica del acceso al agua en el tiempo y comparar de manera directa la situación relativa de cada país.

El objetivo es específico porque se centra únicamente en el acceso al agua potable y en un grupo de países, teniendo en cuenta los años e indicadores. Es medible porque se observan variables cuantitativas expresadas en porcentaje de población, a partir de las cuales analizamos variaciones interanuales, las variaciones acumuladas desde 2012, las brechas entre distintos niveles de servicio (mejorado, seguro, limitado) y un ranking anual de países para cada indicador. El análisis es alcanzable debido a lo que proporciona el dataset que contiene países, regiones, años y valores porcentuales para lo requerido. Es relevante debido a que en algunas regiones es un problema crítico para el desarrollo humano, visualizarlos permiten identificar las desigualdades entre dichas regiones. Está acotado en el tiempo porque el estudio se limita deliberadamente al período 2012–2022 y a este conjunto específico de países.

## Alcance

El análisis tiene un alcance significativo, ya que se limita a abordar dos regiones, el Latinoamericano y el Africano, entre los periodos 2012-2022, en el dataset base hay varios indicadores que no son tenidos en cuenta pero los abordados son especialmente cinco, que son el “uso de fuentes mejoradas, disponibles cuando se necesitan”, “uso de servicios básicos”, “uso de fuentes mejoradas”, “uso de servicios gestionados de forma segura” y “uso de servicios limitados”, el resultante de esta transformación de los datos permitió enriquecer con variaciones, analizar las brechas entre sus indicadores, y establecer un ranking entre los países.

Los indicadores que no van a ser tenidos en cuenta denotan por ejemplo datos por provincia o departamentos, zonas rurales y urbanas, de saneamiento e higiene, pero la mayoría está vacía o sin informar por ende no se abordan en este análisis.

## Usuario final y nivel de aplicación del análisis

L

# Descripción de la temática de los datos

Para el análisis se extrajo información sobre el acceso de la población a servicios de agua potable en los hogares en Kaggle (UN Global Water Data 2012-2022).

Se analizarán 20 países (10 de América Latina y 10 de África) y cubre el período entre los años 2012–2022, permitiendo visualizar niveles de acceso, calidad del servicio y evolución en el tiempo.

Los datos incluyen cinco indicadores principales de agua potable a nivel de población:

- **WS\_PPL\_W-AVA:** uso de fuentes de agua potable mejoradas disponibles cuando se necesitan. (red de agua dentro de la casa o del terreno que casi nunca se corta)
- **WS\_PPL\_W-B:** uso de servicios básicos de agua potable. (cañilla pública mejorada o un pozo protegido al que hay que ir caminando)
- **WS\_PPL\_W-I:** uso de fuentes de agua potable mejoradas. (agua de red, pozo entubado protegido, bomba manual o grifo público mejorado que no esté dentro de la vivienda)
- **WS\_PPL\_W-SM:** uso de servicios de agua potable gestionados de forma segura. (agua de red dentro del hogar, disponible cuando se necesita y que además cumple criterios de calidad)
- **WS\_PPL\_W-L:** uso de servicios limitados de agua potable. (fuente mejorada (pozo protegido, grifo público, etc.) pero que está lejos o con restricciones fuertes de horario, por lo que el acceso es "limitado")

Cada indicador está expresado como porcentaje de la población (%), por país y año.

Además, el dataset incorpora variables derivadas, como:

- Cambios interanuales y acumulados desde 2012 en el acceso al agua.
- Brechas entre porcentajes de indicadores.

# Transformación, normalización y enriquecimiento de los datos.

Primero que nada en el dataset base identificamos los indicadores que tiene a través de la consulta:

```
1 |  SELECT DISTINCT [INDICATOR Indicator]
2 |  FROM dbo.water
3 |  ORDER BY [INDICATOR Indicator];
```

INDICATOR Indicator
1 Ascension and Tristan da Cunha"
2 excluding Australia and New Zealand"
3 Hong Kong Special Administrative Region"
4 Macao Special Administrative Region"
5 Sint Eustatius and Saba"
6 WS_HCF_W-B: Proportion of health care facilities with basic water services
7 WS_HCF_W-L: Proportion of health care facilities with limited water services
8 WS_HCF_W-N: Proportion of health care facilities with no water service
9 WS_PPL_H-N: Proportion of population with no handwashing facility at home
10 WS_PPL_W-AVA: Proportion of population using improved drinking water sources available when needed
11 WS_PPL_W-B: Proportion of population using basic drinking water services
12 WS_PPL_W-I: Proportion of population using improved drinking water sources
13 WS_PPL_W-L: Proportion of population using limited drinking water services
14 WS_PPL_W-NP: Proportion of population using non-piped improved drinking water sources
15 WS_PPL_W-P: Proportion of population using piped drinking water sources
16 WS_PPL_W-PRE: Proportion of population using improved drinking water sources located on premises
17 WS_PPL_W-QUA: Proportion of population using of improved drinking water sources free from faecal and priority chemical contamination
18 WS_PPL_W-SM: Proportion of population using safely managed drinking water services
19 WS_PPL_W-SW: Proportion of population using surface water
20 WS_PPL_W-UI: Proportion of population using unimproved drinking water sources
21 WS_SCH_W-L: Proportion of schools with limited drinking water services
22 WS_SCH_W-N: Proportion of schools with no drinking water service

Luego de tener identificado los indicadores vamos a crear una nueva columna:

```
1 |  DROP TABLE dbo.watersimplificadoenriquecido;
2 |
3 |  SELECT
4 |      [REF_AREA Geographic area],
5 |      [INDICATOR Indicator],
6 |      [TIME_PERIOD Time period],
7 |      [OBS_VALUE Observation Value],
8 |      [UNIT_MEASURE Unit of measure]
9 |      INTO dbo.watersimplificadoenriquecido
10 |      FROM dbo.water
11 |      WHERE
12 |          [INDICATOR Indicator] LIKE '%WS_PPL_W-AVA:%'
13 |          OR [INDICATOR Indicator] LIKE '%WS_PPL_W-B:%'
14 |          OR [INDICATOR Indicator] LIKE '%WS_PPL_W-I:%'
15 |          OR [INDICATOR Indicator] LIKE '%WS_PPL_W-SM:%'
16 |          OR [INDICATOR Indicator] LIKE '%WS_PPL_W-L:%';
```

Tomamos valores de “REF\_AREA Geographic area” y lo transformamos en “country\_code” (CHL por ejemplo) y “country\_name” (Chile) y “región” creando nuevas columnas dándole los valores AFR en caso de los africanos y LAT para los latinoamericanos:

```
1  ALTER TABLE dbo.watersimplificadoenriquecido
2    ADD country_code varchar(3) NULL,
3        country_name varchar(200) NULL,
4        region      varchar(3) NULL;
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
```

---

```
1  UPDATE w
2  SET
3    country_code = LEFT(w.[REF_AREA Geographic area], 3),
4    country_name = LTRIM(SUBSTRING(
5      w.[REF_AREA Geographic area],
6      CHARINDEX(':', w.[REF_AREA Geographic area]) + 1,
7      200
8    )),
9    region = CASE
10      WHEN LEFT(w.[REF_AREA Geographic area], 3) IN
11        ('BRA', 'CHL', 'COL', 'CRI', 'DOM', 'ECU', 'GTM', 'HND', 'PER', 'PRY')
12        THEN 'LAT'
13
14      WHEN LEFT(w.[REF_AREA Geographic area], 3) IN
15        ('CIV', 'COD', 'DZA', 'ETH', 'GHA', 'MAR', 'MDG', 'NGA', 'SEN', 'UGA')
16        THEN 'AF'
17      ELSE NULL
18    END
19  FROM dbo.watersimplificadoenriquecido AS w;
```

---

En estas consultas básicamente creamos columnas y las enriquecemos con los valores que vamos a trabajar en este proyecto, teniendo en cuenta que los 10 países Latinoamericanos son:

BRA → Brasil  
CHL → Chile  
COL → Colombia  
CRI → Costa Rica  
DOM → República Dominicana  
ECU → Ecuador  
GTM → Guatemala  
HND → Honduras  
PER → Perú  
PRY → Paraguay

Los 10 países Africanos son:

CIV → Costa de Marfil (Côte d'Ivoire)  
COD → República Democrática del Congo  
DZA → Argelia  
ETH → Etiopía  
GHA → Ghana  
MAR → Marruecos  
MDG → Madagascar  
NGA → Nigeria  
SEN → Senegal  
UGA → Uganda

Eliminamos la columna “REF\_AREA Geographic area” con DROP COLUMN

Si realizamos la siguiente consulta:

```
1  SELECT DISTINCT
2      country_code,
3      country_name,
4      region
5  FROM dbo.watersimplificadoenriquecido
6  ORDER BY region, country_code;
7
```

Vamos a notar los otros países que continúan en la columna, entonces ejecutamos el siguiente comando para que solo queden los que tengan un valor asignado en “region”:

```
1  DELETE FROM dbo.watersimplificadoenriquecido
2  WHERE region IS NULL;
```

Una vez ejecutado esa consulta ya vamos a tener una mejor visualización de los países que vamos a analizar.

Renombraremos la columna “TIME\_PERIOD time period” y le asignamos “year”, también vamos a separar la columna “INDICATOR Indicator” por “indicator\_code” e “indicator\_label” creando las ya mencionadas, donde uno tendrá el código como por ejemplo “WS\_PPL\_W-AVA” y el otro el indicador la descripción del mismo.

Ejecutamos las siguientes consultas:

```
1  ALTER TABLE dbo.watersimplificadoenriquecido
2    ADD indicator_code varchar(50) NULL,
3        indicator_label varchar(500) NULL;
4
5  UPDATE w
6  SET
7      indicator_code = LEFT(w.[INDICATOR Indicator],
8          CHARINDEX(':', w.[INDICATOR Indicator]) - 1),
9      indicator_label = LTRIM(SUBSTRING(
10         w.[INDICATOR Indicator],
11         CHARINDEX(':', w.[INDICATOR Indicator]) + 1,
12         500
13     ))
14  FROM dbo.watersimplificadoenriquecido AS w;
15
16 ALTER TABLE dbo.watersimplificadoenriquecido
17   DROP COLUMN [INDICATOR Indicator];
```

También vamos a borrar la tabla “UNIT\_MEASURE Unit of measure”.

Luego crearemos las columnas “variacion\_x\_anio” y le asignamos valores.

```
1  ALTER TABLE dbo.watersimplificadoenriquecido
2    ADD variacion_x_anio float NULL;
3
4  WITH c AS (
5      SELECT
6          *,
7          CAST([OBS_VALUE Observation Value] AS float) AS val_num,
8          LAG(CAST([OBS_VALUE Observation Value] AS float)) OVER (
9              PARTITION BY country_code, indicator_code
10             ORDER BY CAST(year AS int)
11         ) AS prev_val
12     FROM dbo.watersimplificadoenriquecido
13 )
14 UPDATE c
15 SET variacion_x_anio = val_num - prev_val;
```

También creamos otra columna llamada “variacion\_desde\_2012” y le asignamos valores.

```
1  ALTER TABLE dbo.watersimplificadoenriquecido
2    ADD variacion_desde_2012 float NULL;
3
4  WITH c AS (
5      SELECT
6          *,
7          CAST([OBS_VALUE_Observation Value] AS float) AS val_num,
8          FIRST_VALUE(CAST([OBS_VALUE_Observation Value] AS float)) OVER (
9              PARTITION BY country_code, indicator_code
10             ORDER BY CAST(year AS int)
11         ) AS val_2012
12     FROM dbo.watersimplificadoenriquecido
13 )
14 UPDATE c
15 SET variacion_desde_2012 = val_num - val_2012;
```

También creamos otra columna “brechaentre\_wl\_y\_resto” y le asignamos valores.

```
ALTER TABLE dbo.watersimplificadoenriquecido
ADD brechaentre_wl_y_resto float NULL;

WITH g AS (
    SELECT
        *,
        MAX(
            CASE
                WHEN indicator_code = 'WS_PPL_W-L'
                    THEN CAST([OBS_VALUE Observation Value] AS float)
            END
        ) OVER (PARTITION BY country_code, year) AS val_L,
        CAST([OBS_VALUE Observation Value] AS float) AS val_cur
    FROM dbo.watersimplificadoenriquecido
)
UPDATE g
SET brechaentre_wl_y_resto = val_L - val_cur;
```

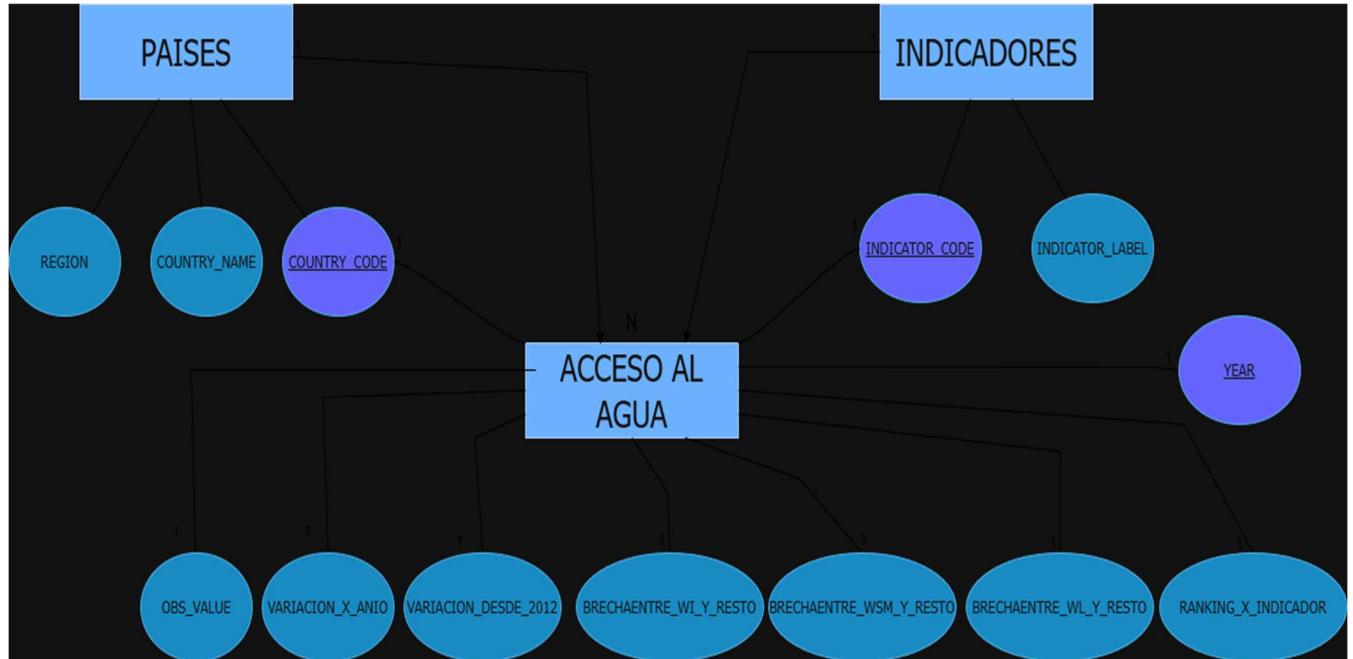
Por último vamos a crear una tabla llamada “ranking\_x\_indicador” y le asignamos valores.

```
ALTER TABLE dbo.watersimplificadoenriquecido
ADD ranking_x_indicador int NULL;

WITH r AS (
    SELECT
        *,
        RANK() OVER (
            PARTITION BY indicator_code, year
            ORDER BY CAST([OBS_VALUE Observation Value] AS float) DESC
        ) AS rk
    FROM dbo.watersimplificadoenriquecido
)
UPDATE r
SET ranking_x_indicador = rk;
```

De esta forma conseguimos el dataset “waterv2.csv”.

# Diagrama de entidad-relación



## Listado de tabla

En la tabla “dbo.watersimplificadoenriquecido” se concentra una sola única tabla, que la misma agrupa toda la información de los países, indicadores y observaciones del acceso al agua.

La misma posee una sola clave primaria que a su vez es una clave concatenada que esta formada por las tablas “country\_code”, “indicator\_code” y “year”.

La combinación de las mismas identifican de manera única un registro de país, indicador y año en el análisis.

## Listado de las columnas de la tabla

TIPO DE CLAVE	CAMPO	TIPO DE CAMPO
PK	country_code	varchar
PK	indicator_code	varchar
PK	year	varchar
	OBS_VALUE Observation Value	varchar
	country_name	varchar
	region	varchar
	indicator_label	varchar
	variacion_x_anio	float
	variacion_desde_2012	float
	brechaentre_wi_y_resto	float
	brechaentre_wsm_y_resto	float
	brechaentre_wl_y_resto	float
	ranking_x_indicador	int

country\_code: Código de tres letras que identifica al país (ej.: BRA, CHL, GHA).

indicator code: Código del indicador de acceso al agua. (ej.: WS\_PPL\_W-AVA)

year: Año al que corresponde la observación (entre 2012 y 2022).

OBS\_VALUE Observation Value: Valor observado del indicador en ese país y año. Es el **porcentaje de la población** que cumple la condición del indicador (0–100).

country\_name: Nombre del país al que pertenece el registro (ej.: Chile, Brasil, Ghana).

region: Región geográfica del país dentro del análisis: LAT para Latinoamérica y AF para África.

indicator\_label: Descripción textual del indicador, explicando qué tipo de acceso al agua mide ese código.

variacion\_x\_anio: Diferencia del valor del indicador respecto al **año anterior**, para el mismo país e indicador. Expresado en puntos porcentuales.

variacion\_desde\_2012: Diferencia del valor del indicador respecto al **año 2012**, para el mismo país e indicador. Mide cuánto mejoró o empeoró el acceso en todo el período.

brechaentre\_wi\_y\_resto: Brecha entre el indicador **WS\_PPL\_W-I** (uso de fuentes mejoradas) y el resto de los indicadores, para el mismo país y año.

brechaentre\_wsm\_y\_resto: Brecha entre el indicador **WS\_PPL\_W-SM** (servicios gestionados de forma segura) y el resto de los indicadores, para el mismo país y año.

brechaentre\_wl\_y\_resto: Brecha entre el indicador **WS\_PPL\_W-L** (servicios limitados de agua potable) y el resto de los indicadores, para el mismo país y año.

ranking\_x\_indicador: Posición del país en un **ranking del 1 al 20** para ese año e indicador, según el valor de OBS\_VALUE.

