

DISTRIBUCIÓN DE AGUA BEBIBLE EN LATINOAMERICA



Galetto Cecilia

CODERHOUSE

Data Analytics

Docente: Rubén Alfredo Parente Aun

Tutora: María Belén Renna

17/12/2024

Contenido

1 Resumen.....	2
2 Introducción	2
3 Hipótesis	3
4 Objetivos	3
5 Usuario final y nivel de aplicación del análisis.....	3
6 Herramientas tecnológicas utilizadas.....	3
7 Modelo relacional	4
7.1 Diagrama de entidad- relación	4
7.2 Listado y descripción de tablas	6
7.3 Listado de columnas de cada tabla	8
8 Modelo relacional en PowerBi.....	9
9 Modificaciones de la base de datos.....	10
9.1 Nuevas Tablas.....	10
9.2 Columnas añadidas	10
9.3 Diagrama final	11
10 Edición de la base de datos en PowerBi.....	11
11 Medidas calculadas generadas.....	11
12 Segmentadores y filtros.....	13
13 Visualización de datos	13
13.1 Físico-Químico	14
13.2 Bacteriológico	14
13.3 Toxicológico	15
13.4 Análisis.....	16
13.5 Análisis por región.....	16
14 Conclusiones finales.....	17
15 Futuras líneas.....	17
16 Bibliografía.....	18

1 Resumen

El conjunto de datos del presente trabajo contiene muestras de agua recolectadas tanto de fuentes naturales como industriales, así como diversas variables analizadas para clasificar el agua como potable o no potable. Nuestro cliente nos ha proporcionado los datos de 16 países en un periodo de 5 años, desde el 2020 al 2024 inclusive, con el objetivo de evaluar el nivel de contaminación del agua en Latinoamérica y tomar las medidas pertinentes al respecto.

2 Introducción

En noviembre de 2022, tras la 12ª Asamblea Latinoamericana de Ecología, se modificaron las regulaciones relativas a los parámetros y valores tolerables para el agua potable en la industria, basándose en los consensos ecológicos y ambientales alcanzados.

Por otro lado, se anticipa una modificación en los valores de las variables estudiadas en las aguas naturales, como resultado de la implementación de leyes ecológicas en cada país y la creciente conciencia global sobre la importancia del cuidado y la remediación del medioambiente.

Los valores tolerables para las variables analizadas, que permiten caracterizar el agua como potable, son los siguientes:

Fisicoquímicas:

- Cloro residual libre: Hasta 1 mg/L
- Turbidez: Hasta 1 NTU
- Fluoruro: Hasta 1 mg/L
- pH: Entre 6 – 9.5

Biológicas:

- Coliformes: <1 UFC/mL
- *Escherichia coli*: <1 UFC/mL

Toxicológicas:

- Agroquímicos: ND (no detectable)

3 Hipótesis

El presente estudio organizará los datos de manera que permita una visualización clara de la información, con el objetivo de evidenciar el grado de contaminación del agua, evaluar la efectividad de los métodos empleados en la industria y, de esta forma, tomar decisiones informadas y dirigir recursos hacia las áreas con mayor nivel de contaminación. Asimismo, se podrá evaluar el impacto de las medidas adoptadas tras la 12ª Asamblea Latinoamericana de Ecología, para determinar si han tenido efectos significativos en la mejora de la calidad del agua.

4 Objetivos

- Organizar los datos y lograr una visualización clara y concisa que permita su presentación en la 13ª Asamblea Latinoamericana de Ecología.
- Proporcionar información útil para los diversos representantes, que les permita implementar medidas ecológicas y de biorremediación, basadas en los resultados obtenidos.

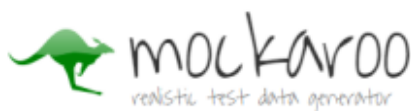
5 Usuario final y nivel de aplicación del análisis

El presente análisis está dirigido a las autoridades de los institutos o ministerios de ecología y medioambiente de cada país participante en las Asambleas Latinoamericanas de Ecología, proporcionando información valiosa para la toma de decisiones a nivel legislativo en materia ambiental y ecológica.

6 Herramientas tecnológicas utilizadas

Las herramientas que se implementaron para realizar el presente trabajo se detallan a continuación:

Herramientas utilizadas para completar el set de datos:



Herramientas utilizadas para el modelado y normalización de la base de datos:



Herramientas utilizadas para la visualización de los datos y generar la documentación correspondiente:

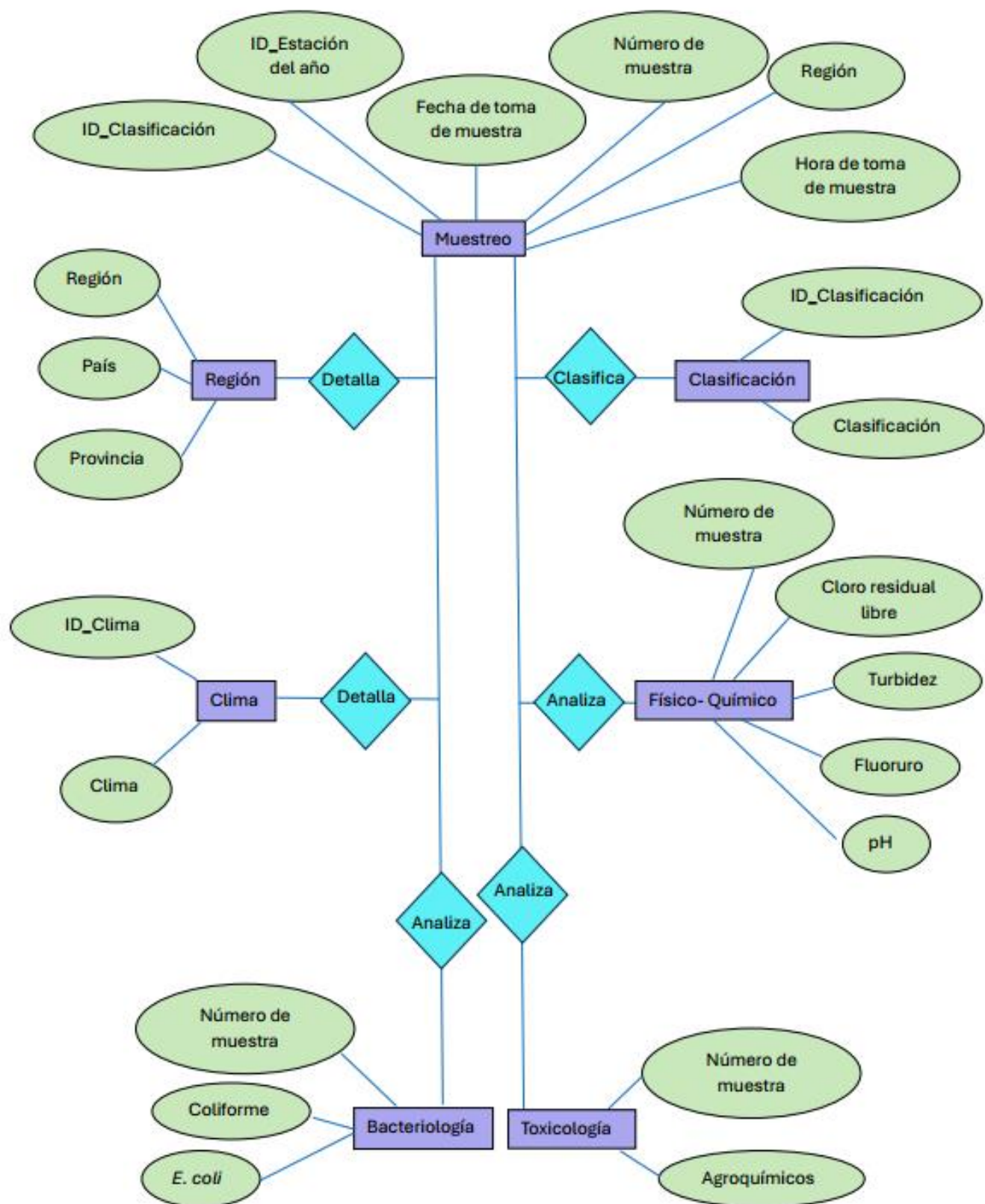
ChatGPT



7 Modelo relacional

7.1 Diagrama de entidad- relación

A continuación, se detalla el diagrama de entidad- relación generado para el presente proyecto.



7.2 Listado y descripción de tablas

En esta sección se describirán cada una de las tablas

- Tabla Reportes

Contiene un número de identificación llamado número de muestra, único para cada ejemplar, con la fecha y hora en que se tomaron las muestras, la localización, el clima general de la región y su clasificación.

💧	PK	Número de muestra
💧	FK	Región
💧	FK	ID_Clasificación
💧	FK	ID_Estación del año
💧		Fecha de toma de muestra
💧		Hora de toma de muestra

- Tabla Región

Contiene el código numérico identificador de cada región en donde fueron recolectadas las distintas muestras con el país y la provincia correspondiente.

💧	PK	Región
💧		País
💧		Provincia

- Tabla Clasificación

Contiene la clasificación de la muestra como natural, si la misma se obtuvo de una fuente Natural como ríos, lagos, entre otras o industrial si la misma se obtuvo de industrias o plantas potabilizadoras.

💧	PK	ID_Clasificación
💧		Clasificación

- Tabla Clima

Contiene una clasificación según el clima general de la región y su identificador numérico.

🔹	PK	ID_Clima
🔹		Clima

- Tabla Físico- Químico

Contiene las variables y concentraciones físico-químicas a analizar para calificar un agua como potable. La medición de cloro libre residual se mide en miligramos por litro (mg/L), la medición de turbidez en Unidades de Turbidez Nephelométrica (UTN), la medición de la concentración de fluoruro en miligramos por litro (mg/L).

🔹	PK	Número de muestra
🔹		Cloro residual libre
🔹		Turbidez
🔹		Fluoruro
🔹		pH

- Tabla Bacteriológico

Contiene el análisis bacteriológicas de las bacterias comúnmente presentes en agua y se miden en unidades formadoras de colonias por 100ml (UFC/100ml).

🔹	PK	Número de muestra
🔹		Coliformes
🔹		<i>E. coli</i>

- Tabla Toxicológico

Contiene el análisis toxicológico que incluye la presencia de agroquímicos medida en miligramos por litro (mg/L).

🔹	PK	Número de muestra
🔹		Agroquímicos

7.3 Listado de columnas de cada tabla

A continuación, se mencionará las columnas que posee cada tabla junto con su tipo de campo y clave:

Reportes		
Campo	Tipo de campo	Tipo de clave
Número de muestra	int	PK
Región	int	FK
ID_Clasificación	int	FK
ID_Estación del año	int	FK
Fecha de toma de muestra	datetime	
Hora de toma de muestra	timestamp	

Región		
Campo	Tipo de Campo	Tipo de clave
Región	int	PK
País	text	
Provincia	varchar (100)	

Clasificación		
Campo	Tipo de Campo	Tipo de clave
ID_Clasificación	int	PK
Clasificación	varchar (100)	

Clima		
Campo	Tipo de Campo	Tipo de clave
ID_Clima	int	PK
Clima	text	

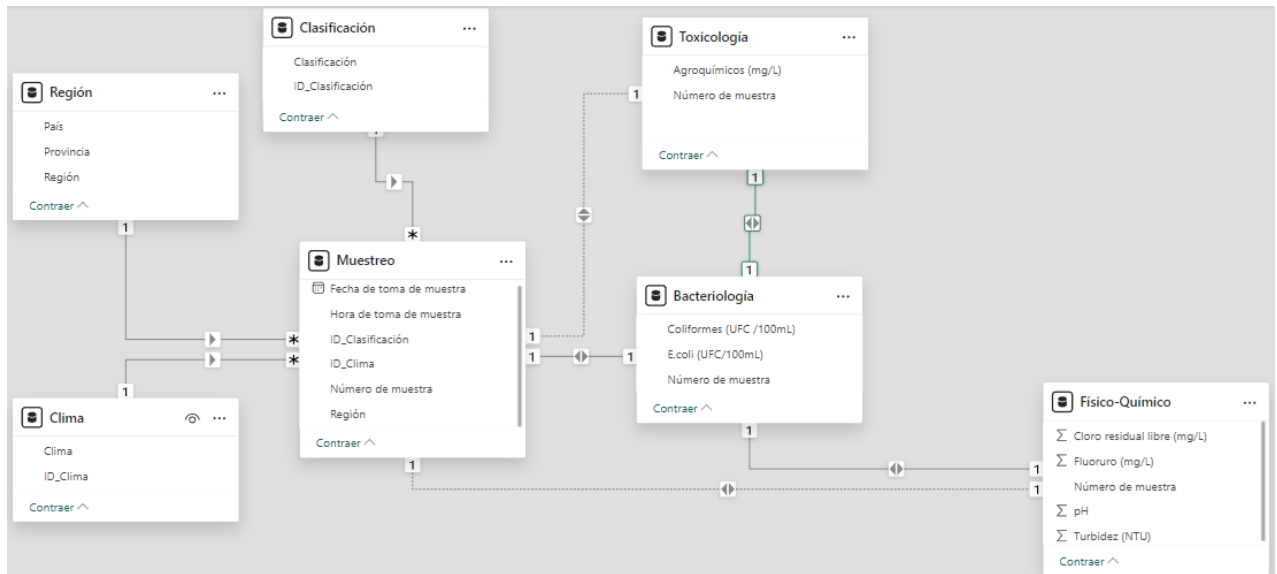
Físico-Químico		
Campo	Tipo de Campo	Tipo de clave
Número de muestra	int	PK
Cloro residual libre	decimal (10,2)	
Turbidez	decimal (10,2)	
Fluoruro	decimal (10,2)	
pH	decimal (10,2)	

Bacteriológico		
Campo	Tipo de Campo	Tipo de clave
Número de muestra	int	PK
Coliformes	varchar (100)	
<i>E. coli</i>	varchar (100)	

Toxicológico		
Campo	Tipo de Campo	Tipo de clave
Número de muestra	int	PK
Agroquímicos	varchar (100)	

8 Modelo relacional en PowerBi

A continuación, se puede visualizar el modelo relacional en PowerBi



9 Modificaciones de la base de datos

Se realizaron las siguientes modificaciones

9.1 Nuevas Tablas

Se crearon 3 nuevas tablas:

1. Tabla Análisis: como combinación de las tablas Físico-Químico, Bacteriología y Toxicología
2. Tabla calendario
3. Tabla medidas

9.2 Columnas añadidas

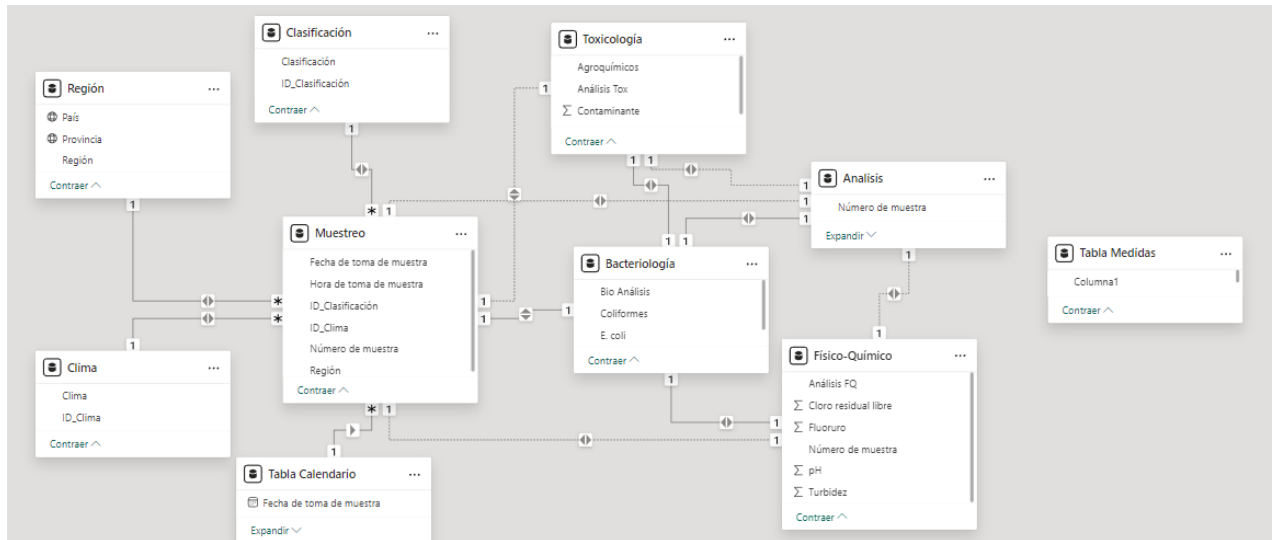
Se agrego una columna en la tabla 'toxiológico' y se la nombro 'contaminante' para generar una medida de la forma: A cada ND (no detectable) se lo relacionó con el número 0

Se agregó una columna en la tabla 'Físico-Químico' y se la denominó 'Análisis FQ' para clasificar las muestras de agua como bebibles o no bebibles según los valores mencionados en la sección 2. De la misma forma, en la tabla 'Bacteriología' y se la denominó 'Bio Análisis' y en la tabla 'Toxicología' y se la denomino 'Análisis Tox'.

Se agrego la columna 'Calidad' en la nueva tabla 'Análisis' para clasificar el agua como bebida o no según los criterios y valores mencionados en la sección 2.

9.3 Diagrama final

A continuación, se presenta el diagrama relacional luego de las modificaciones realizadas.



10 Edición de la base de datos en PowerBi

Antes de comenzar con el análisis, se realizaron modificaciones con numerosas herramientas tales como:

- Columna a partir de ejemplos
- Movimiento y renombre de la nueva columna
- modificar el tipo de columna
- reemplazar valores
- Columna condicional
- Columna personalizada

11 Medidas calculadas generadas

A continuación, se detallan las medidas calculadas generadas y sus fórmulas

En columnas personalizadas:

- En la tabla 'Bacteriológico' para la columna 'Bio Análisis': `if [Coliformes]= "<1" and [E. coli]= "<1" then "bebible" else "No bebible"`

- En la tabla 'Físico-Químico' para la columna 'Análisis FQ': if [Cloro residual libre] <1 and [Turbidez] <1 and [Fluoruro] <1 and [pH] >6 and [pH] <9.5 then "Bebible" else "No bebible"
- En la tabla 'Análisis' para la columna 'Calidad': if [Coliformes]= "<1" and [E. coli]= "<1" and [#"Físico-Químico.Cloro residual libre"]<1 and [#"Físico-Químico.Turbidez"] <1 and [#"Físico-Químico.Fluoruro"] <1 and [#"Físico-Químico.pH"] >6 and [#"Físico-Químico.pH"] <9.5 and [Toxicología.Contaminante]=0 then "Bebible" else "No bebible"

En la tabla calendario:

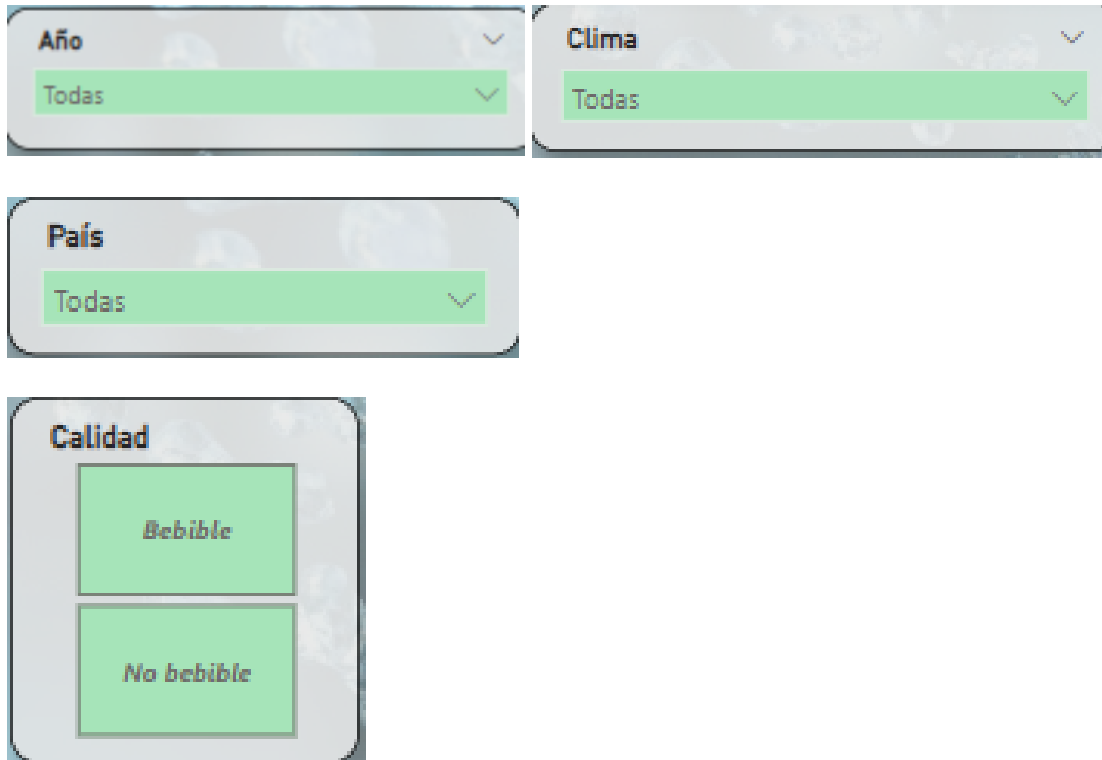
- Fecha = format('Tabla Calendario'[Fecha de toma de muestra],"DD/MM/YYYY")
- Numero de mes = MONTH('Tabla Calendario'[Fecha de toma de muestra])
- Nombre del mes = 'Tabla Calendario'[Fecha de toma de muestra].[Mes]
- Mes corto = FORMAT('Tabla Calendario'[Fecha de toma de muestra],"mmm")
- Año = YEAR('Tabla Calendario'[Fecha de toma de muestra])
- Periodo = FORMAT('Tabla Calendario'[Fecha de toma de muestra],"YYYYMM")
- Día semana N° = WEEKDAY('Tabla Calendario'[Fecha de toma de muestra],2)
- Día semana = FORMAT('Tabla Calendario'[Fecha de toma de muestra],"DDDD")
- Día semana corto = FORMAT('Tabla Calendario'[Fecha de toma de muestra],"DDD")
- Trimestre = "Q" & FORMAT('Tabla Calendario'[Fecha de toma de muestra],"Q")

En la tabla Medidas:

- Cantidad de muestra evaluadas = count (Muestreo[Número de muestra])
- Cantidad bebible = calculate(COUNT(Análisis[Calidad]), FILTER(Análisis,Análisis[Calidad]="Bebible"))
- Porcentaje = DIVIDE([Cantidad bebible],[Cantidad de muestra evaluadas])

12 Segmentadores y filtros

Con el objetivo de alcanzar una visualización eficiente, se implementó el uso los segmentadores que se detallan a continuación:

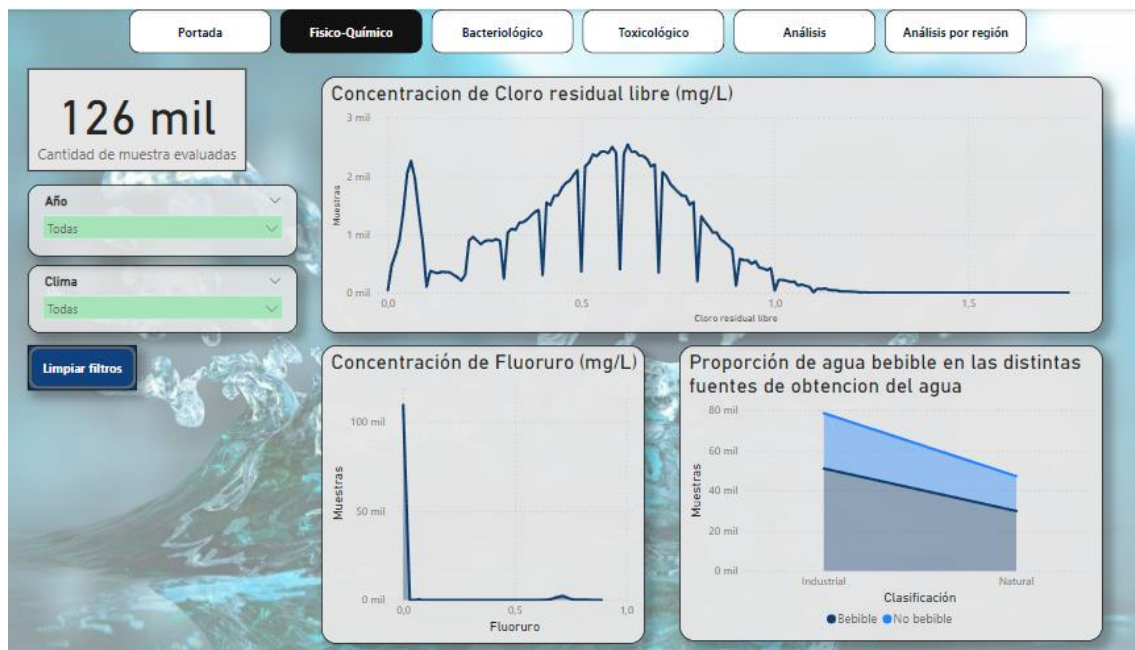


Estos segmentadores no se emplearon en todas las solapas, sino en función de las necesidades y requerimientos específicos de los distintos gráficos presentados en cada una de ellas.

13 Visualización de datos

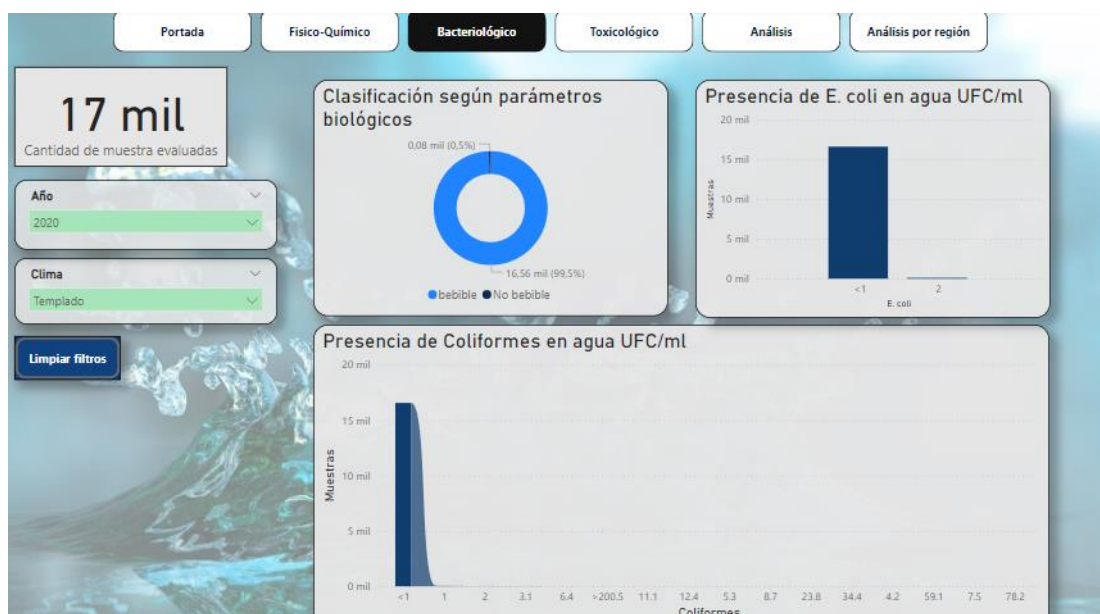
A continuación, se describen las diferentes visualizaciones

13.1 Físico-Químico



En esta solapa se presentan los valores aceptables de los parámetros físico-químicos que clasifican el agua como apta para el consumo o no. La mayoría de las muestras de agua resultaron aptas para el consumo. Al aplicar los filtros y segmentadores, se observa que, tanto en distintos climas como a lo largo de los años, los resultados se mantienen constantes. Asimismo, se identificaron similitudes a lo largo de los años en las concentraciones de cloro y fluoruro.

13.2 Bacteriológico



En esta solapa se presentan los parámetros biológicos utilizados para clasificar el agua como apta para su consumo. En cada uno de los países evaluados, no se detectan grandes cantidades de aguas contaminadas con microorganismos. Sin embargo, se observa que, en las aguas contaminadas, la proporción de bacterias coliformes es mayor que la de *Escherichia coli*. A lo largo de los años, se puede apreciar que *Escherichia coli* ha dejado de ser un contaminante frecuente.

13.3 Toxicológico



En esta solapa se analiza la presencia de agroquímicos en las muestras. Sorprendentemente, en los años 2022 y 2023 se detectó una baja contaminación con agentes tóxicos, mientras que en 2024 se observa nuevamente una elevada concentración de tóxicos en el agua, lo que la hace no apta para el consumo. Además, se evalúa la concentración de estos agentes en aguas industriales, en comparación con las aguas naturales, y se encontró que ambas presentaron niveles de contaminación similares. Aunque se esperaba que la concentración fuera más alta en las fuentes naturales debido al riego y las corrientes de agua que transportan sustancias de los campos a los ríos, la industria también evidencia trazas de agroquímicos.

13.4 Análisis



En esta solapa se presenta un análisis exhaustivo de todos los parámetros que determinan si el agua es apta para el consumo o no. En la mayoría de los casos, el agua es considerada apta para el consumo, y se observa una tendencia similar entre los diferentes países evaluados. Se destaca, sin embargo, una considerable proporción de agua no apta para el consumo en las muestras obtenidas de la industria. No se aprecian variaciones significativas en cuanto al agua apta para el consumo a lo largo de los años.

13.5 Análisis por región



En esta solapa se presenta, en términos porcentuales, la proporción de agua analizada considerada apta para el consumo, así como una comparación entre las diferentes provincias o departamentos de cada país.

14 Conclusiones finales

Se evidencia que a lo largo de los años no se han observado cambios significativos en la calidad del agua, lo que sugiere que las iniciativas planteadas en la 12ª Asamblea Latinoamericana de Ecología no han generado los avances esperados. Sin embargo, se destaca una disminución en la presencia de *Escherichia coli*, lo cual podría atribuirse a una mayor concientización de la población respecto a la higiene y el manejo del agua.

En cuanto a la presencia de agroquímicos, se ha evidenciado una reducción con el paso de los años, aunque el año 2024 muestra una concentración elevada en comparación con 2022 y 2023, lo que sugiere que aún persisten desafíos en el control de estos contaminantes.

Las principales variables que afectan la potabilidad del agua continúan siendo los parámetros físicos y químicos, lo que resalta la importancia de regular y controlar estrictamente estos aspectos, tanto a nivel industrial como en las plantas potabilizadoras. Es crucial implementar normativas más estrictas y asegurar su cumplimiento para garantizar la calidad del agua.

Por último, es fundamental proteger las fuentes de agua natural que se consideran aptas para el consumo, mediante la implementación de leyes medioambientales adecuadas y un esfuerzo continuo de concientización en la población para preservar estos recursos esenciales.

15 Futuras líneas

Para continuar con este estudio, es crucial avanzar hacia una mayor estandarización y sistematización en la recolección de muestras de agua, con el objetivo de lograr estudios comparativos más precisos y efectivos. En este sentido, se planteará en la nueva asamblea la necesidad de normalizar la cantidad de muestras a extraer cada año, tanto en los diferentes territorios como

en las distintas fuentes de agua, lo que permitirá obtener datos consistentes y comparables a lo largo del tiempo.

Con este enfoque, se podrán identificar de manera más eficiente las áreas y fuentes de agua que enfrentan los mayores problemas de contaminación, lo que facilitará la implementación de políticas y medidas específicas para mitigar los riesgos asociados.

16 Bibliografía

1. Kaggle: Your Machine Learning and Data Science Community
2. Parámetros Agua Potable: Conócelos y asegura la calidad | Gedysa Laboratorio
3. <https://learn.microsoft.com/>