



Data analytics

3° Trabajo Entregable

Javier Yañez

Comisión 71725

2024

Descripción del tema

La contaminación ambiental es la presencia de componentes nocivos, bien sean de naturaleza biológica, química o de otra clase, en el medioambiente, de modo que supongan un perjuicio para los seres vivos que habitan un espacio, incluyendo, por supuesto, a los seres humanos. Generalmente la contaminación ambiental tiene su origen en alguna actividad humana.

Las partículas en suspensión y gases producidos por el tráfico rodado, la industria y las calefacciones son los principales causantes de la contaminación atmosférica. El incremento de la emisión de CO₂ provoca el calentamiento global que deriva en el cambio climático y por consiguiente provoca los efectos climáticos adversos como olas de calor, sequía, inundaciones, etc. La principal fuente de contaminación atmosférica son los gases ozono troposférico (O₃), óxidos de azufre (SO₂ y SO₃), óxidos de nitrógeno (NO y NO₂), benzopireno (BaP) y las partículas en suspensión (PM). Estos gases se derivan principalmente de las emisiones provocadas por la quema de combustibles fósiles (incluidas las emisiones generadas por el transporte), los procesos industriales, la quema de bosques, el empleo de aerosoles y la radiación.

Objetivo

En el presente proyecto se pretende analizar la evolución de la calidad del aire y la emisión de gases contaminantes, así como el impacto de esta contaminación en la temperatura en los países de Latinoamérica en los últimos años. Se hará un análisis semestral como anual de la temática.

Hipótesis

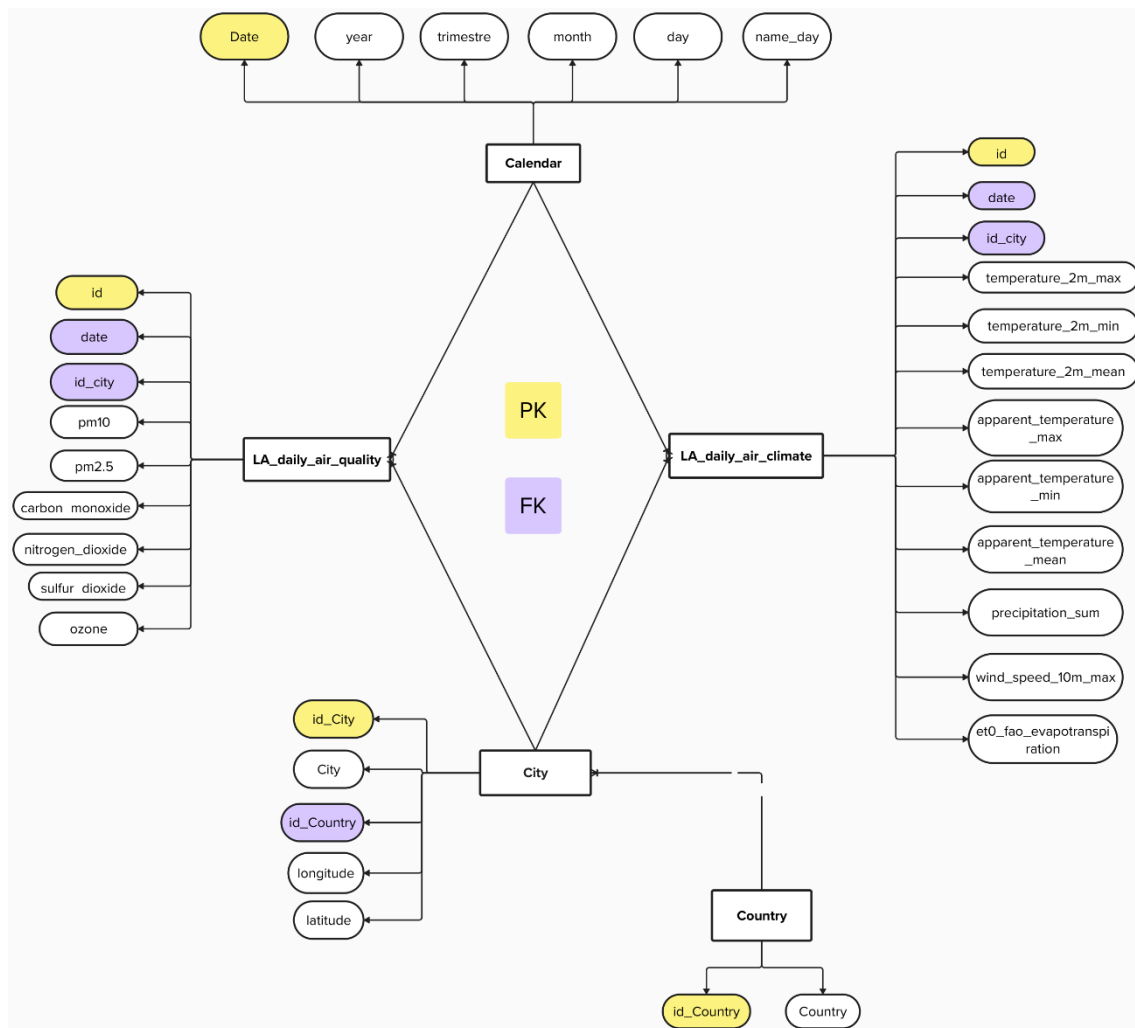
La emisión de gases sigue en aumento, contaminando la atmosfera, lo que contribuye al calentamiento global y, por ende, al aumento de la temperatura.

Nivel de aplicación del análisis

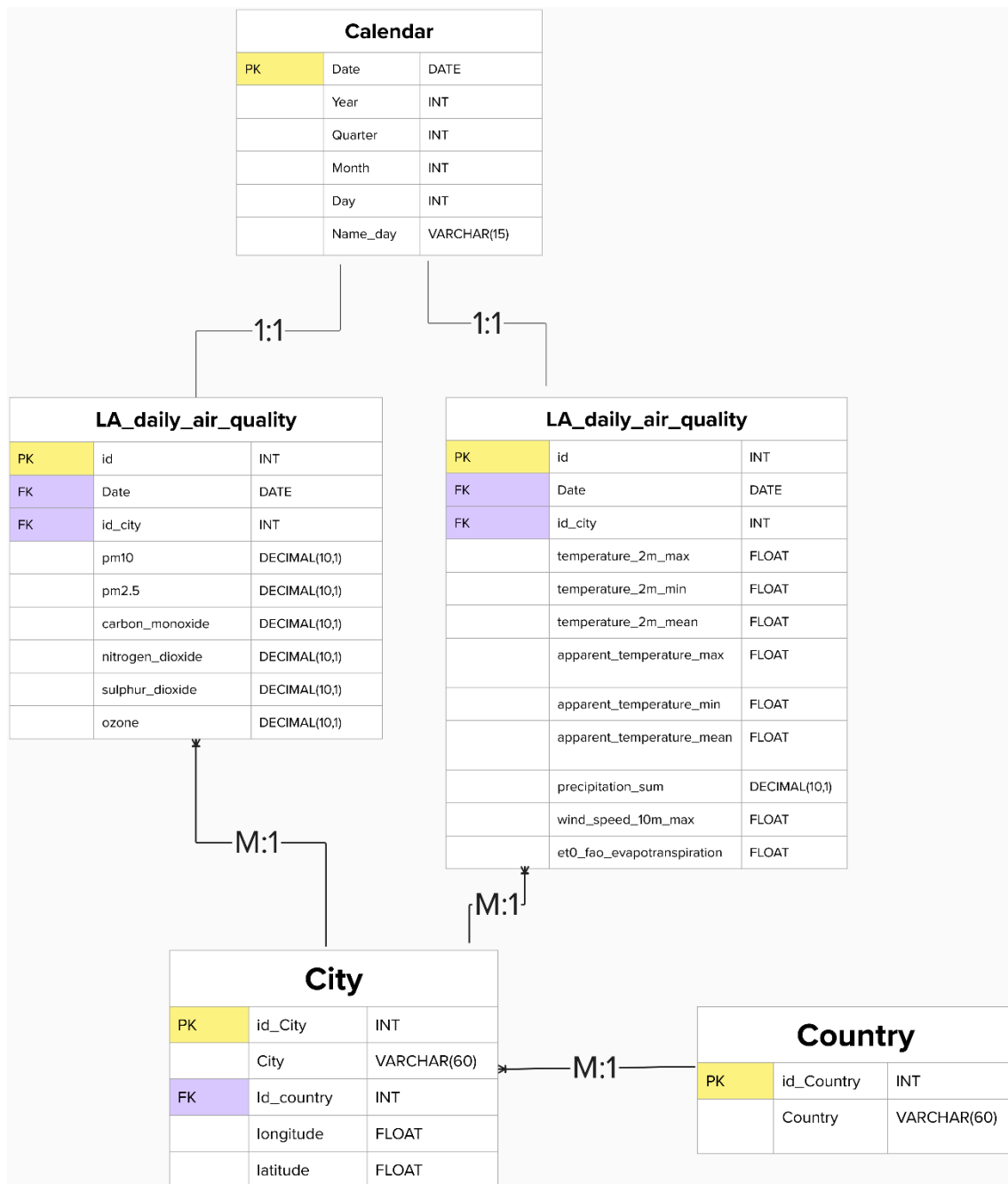
Los tableros serán para el uso a nivel estratégico, siendo los dirigentes de cada país los destinatarios de los mismos para su análisis.

Diagramas

A continuación, se presentará el DER del proyecto a realizar. Se buscará exponer de manera clara las relaciones entre las distintas tablas de hecho y dimensionales de los datos disponibles. Al centro se ubican las dos tablas de hecho, arriba el calendario y abajo la información normalizada, siguiendo la tercera forma normal, sobre la locación del hecho observado dentro de las tablas City y Country



En el siguiente gráfico, se mostrará este diagrama ahondando en el tipo de dato de las tablas y el tipo de relación que hay entre ellas.



Acerca del dataset

Clima y calidad del aire en Latinoamérica

Este dataset contiene 2 archivos CSV generados a partir de la API Open-Meteo y contiene información sobre varios países de Latinoamérica. Los datasets incluyen nombre del país, capital, clima y calidad del aire. Los países incluidos en estos datos son:

Argentina - Buenos Aires

Bolivia - Sucre
Brazil - Brasilia
Chile - Santiago
Colombia - Bogota
Costa Rica - San Jose
Cuba - Havana
Dominican Republic - Santo Domingo
Ecuador - Quito
El Salvador - San Salvador
Guatemala - Guatemala City
Haiti - Port-au-Prince
Honduras - Tegucigalpa
Mexico - Mexico City
Nicaragua - Managua
Panama - Panama City
Paraguay - Asuncion
Peru - Lima
Uruguay - Montevideo
Venezuela - Caracas

Tablas

En este apartado se detallará cada tabla de la base de datos, incluyendo nombre, descripción, tipo de dato y si es key o no cada campo.

Calendario

Contiene información sobre cada fecha de cada año.

Campo	Descripción	Tipo de dato	Key
Date	Fecha	DATE	PK
Year	Año	INT	-
Quarter	Trimestre del año	INT	-
Month	Mes del año	INT	-
Day	Día del año	INT	-
Name_day	Nombre del día de la semana	VARCHAR(15)	-

LA daily air quality

Tabla que contiene las mediciones de los distintos agentes que contaminan el medioambiente en la zona en cada día.

Campo	Descripción	Tipo de dato	Key
Id	Identificador del suceso	INT	PK
Date	Fecha del suceso	DATE	FK
Id_City	Identificador del lugar de medición	INT	FK
Pm10	Cantidad de partículas de materia inferior a 10 µm hasta a 10 mts del suelo	DECIMAL(10,1)	-
Pm2.5	Cantidad de partículas de materia inferior a 2.5 µm hasta a 10 mts del suelo	DECIMAL(10,1)	-
Carbon_monoxide	monóxido de carbono cerca de la superficie (10mts del suelo)	DECIMAL(10,1)	-
Nitrogen_dioxide	dióxido de nitrógeno cerca de la superficie	DECIMAL(10,1)	-
Sulfur_dioxide	dióxido de azufre cerca de la superficie	DECIMAL(10,1)	-
Ozone	Ozono cerca de la superficie	DECIMAL(10,1)	-

LA daily air climate

Tabla que contiene las mediciones de las temperaturas, sensación térmica, precipitaciones, viento y evapotranspiración registradas en la zona en cada día.

Campo	Descripción	Tipo de dato	Key
Id	Identificador del suceso	INT	PK
Date	Fecha del suceso	DATE	FK
Id_City	Identificador del lugar de medición	INT	FK
Temperature_2m_max	temperatura máxima a 2 metros del suelo	FLOAT	-
Temperature_2m_min	temperatura mínima a 2 metros del suelo	FLOAT	-
Temperature_2m_mean	temperatura media a 2 metros del suelo	FLOAT	-
Apparent_temperature_max	Sensación térmica máxima	FLOAT	-
Apparent_temperature_min	Sensación térmica mínima	FLOAT	-
Apparent_temperature_mean	Sensación térmica media	FLOAT	-
Precipitation_sum	Precipitación total del día	DECIMAL(10,1)	-
Wind_speed_10m_max	Velocidad máxima del viento a 10 metros del suelo	FLOAT	-
Et0_fao_evapotranspiration	evapotranspiración potencial calculada	FLOAT	-

City

Contiene el nombre de la ciudad, sus coordenadas y el país al que pertenecen.

Campo	Descripción	Tipo de dato	Key
Id_City	Identificador de la ciudad	INT	PK
City	Nombre de la ciudad	VARCHAR(60)	-
Id_Country	Identificador del país	INT	FK
Longitude	Coordenadas de longitud de la ciudad	FLOAT	-
Latitude	Coordenadas de latitud de la ciudad	FLOAT	-

Country

Contiene el identificador del país y el nombre del país

Campo	Descripción	Tipo de dato	Key
Id Country	Identificador del país	INT	PK
Country	Nombre del país	VARCHAR(60)	-

Unidades

Temperatura -> C°

Precipitación -> mm (1 milímetro de lluvia es igual a 1 litro por metro cuadrado)

Evapotranspiración -> mm (1 mm de evapotranspiración por hora equivale a 1 litro de agua por metro.) El valor presentado es la sumatoria del día

PM -> µg/m³ (microgramos/metro cúbico)

Gases atmosféricos -> µg/m³

Base de datos – SQL Server

Antes de iniciar con el dashboard, procedemos a armar la base de datos desde la cual se va a alimentar el tablero mencionado.

Primero se crearon las tablas con el comando CREATE TABLE, especificando el tipo de dato para cada columna y luego se poblaron con los csv mediante el comando BULK.

Acto seguido, se borraron los registros de las tablas en los cuales habían muchos campos en NULL por lo que no podrían aportar información relevante en estas condiciones.

Luego se creó un campo autoincrementable que funcionarán como primary key en esas tablas.

Una vez hecho esto, se crearon las tablas puente (City y Country) con los campos especificados en el apartado de Tablas. Mediante INSERT INTO y SELECT DISTINCT se poblaron las tablas con los valores únicos que se encuentran en las tablas de origen.

Al tener las tablas puente listas, se añadió el campo id_country en la tabla City, la cual se pobló mediante un INNER JOIN con Country (para ver las correspondencias entre el id de la tabla Country y el nombre de la tabla City) y se estableció el FK en ese campo, referenciando al id de la tabla Country.

Este procedimiento se repitió en las tablas de origen con respecto a la tabla City y luego se eliminaron los campos redundantes en las tablas de origen.

Para más detalle de este procedimiento, se añadió el archivo “SQLQueries” que contiene todo el script aplicado en SQL Server en el que se realizan todos los pasos mencionados.

Dashboard – Power BI

Cuando tenemos la base de datos lista como el esquema planteado, la leemos con Power BI desde la opción “Obtener datos” y seleccionamos SQL Server. Para completar la solicitud es necesario establecer la conexión mediante el usuario, contraseña y nombre de la base de datos.

Calendario

Para conectar las tablas mediante las fechas, se crea una tabla Calendario en Power Query mediante el uso del editor avanzado. Se crea una tabla vacía y en el editor se utilizó el siguiente script:

```
let
    lista_fecha = Table.Column(LA_daily_climate, "Measure_date"),
    min = List.Min(lista_fecha),
    max = List.Max(lista_fecha),
    dias = Number.From(max - min) + 1,
    fechas = List.Dates(min, dias, #duration(1,0,0,0)),
    #"Convertida en tabla" = Table.FromList(fechas, Splitter.SplitByNothing(),
    null, null, ExtraValues.Error),
    #"Columnas con nombre cambiado" = Table.RenameColumns(#"Convertida
    en tabla",{{"Column1", "Fecha"}}),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Columnas con nombre
    cambiado",{{"Fecha", type date}}),
    #"Año insertado" = Table.AddColumn(#"Tipo cambiado", "Año", each
    Date.Year([Fecha]), Int64.Type),
    #"Nombre del mes insertado" = Table.AddColumn(#"Año insertado", "Nombre
    del mes", each Date.MonthName([Fecha]), type text),
    #"Trimestre insertado" = Table.AddColumn(#"Nombre del mes insertado",
    "Trimestre", each Date.QuarterOfYear([Fecha]), Int64.Type),
    #"Nombre del día insertado" = Table.AddColumn(#"Trimestre insertado",
    "Nombre del día", each Date.DayOfWeekName([Fecha]), type text)
in
    #"Nombre del día insertado"
```


Esta tabla se genera utilizando la fecha más antigua y mas reciente de la tabla LA_daily_climate y traza el recorrido día a día de una fecha a otra. Luego se despliegan los campos requeridos de ese total de fechas.

Una vez que tenemos todo relacionado, procedemos al diseño del dashboard

Paleta de colores

Para definir la paleta de colores, se diseñó un logo de una empresa ficticia que solicitaría el reporte que desarrollaremos.



Mediante la herramienta de adobe color ([Rueda de colores, un generador de paletas de colores | Adobe Color](#)) se extrajeron los colores del logo y se diseñó una paleta de colores a utilizar en el dashboard.

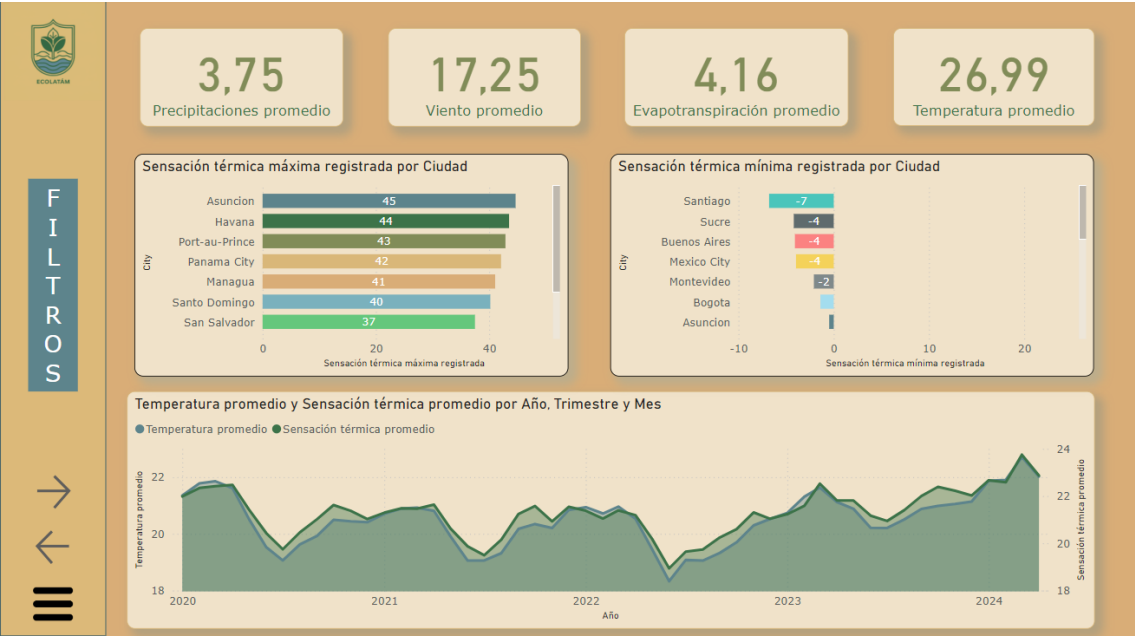
Con nuestra paleta de colores, presentamos las solapas del dashboard realizado.

En todas las solapas se añadió un botón de filtros y uno de menú de navegación desplegables para interactuar con la información mostrada.

Portada

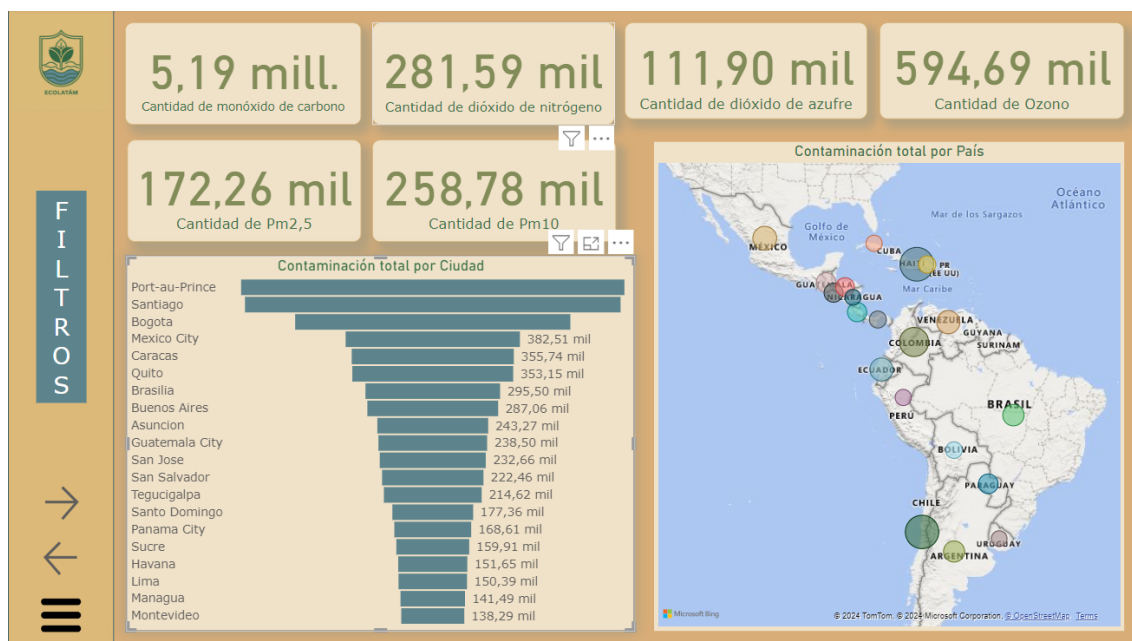


Primera solapa – Condiciones climáticas



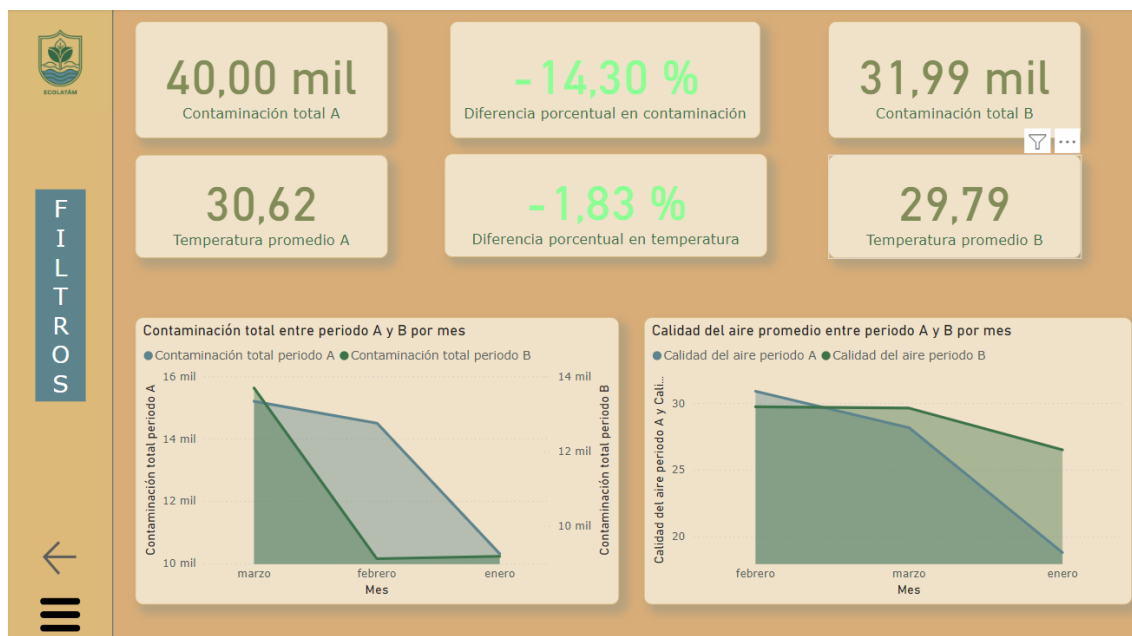
En esta primera solapa se muestra toda la información relativa al clima. Se puede filtrar por tiempo y por país y ciudad (Si bien hay una sola ciudad por país, se deja listo en caso de que se agregue más información a la base de datos en el futuro)

Segunda solapa – Contaminación



En esta solapa, se muestra todo lo relativo a los agentes de contaminación detectados y las partículas que afectan a la calidad del aire detectados en el día a día y un mapa que muestra a simple vista que país emitió más por el tamaño de la burbuja. Se muestra una sumatoria de las mismas en función del tiempo y se puede filtrar por año, trimestre, mes, país y ciudad como la solapa anterior.

Tercera solapa – Impacto y cambios



En esta solapa vemos el cambio que hubo entre dos periodos especificados, si este cambio fue positivo o no, y la comparación de ese periodo con el mismo de otro año especificados. Esta solapa posee dos filtros de tiempo para configurar cada periodo que se compara, siendo el A el más antiguo y el B el más reciente, para que el usuario compare los periodos que requiera. También posee un filtro general por país para hacer foco en una zona en particular.

Medidas

Contaminación total: Suma de los principales agentes contaminantes registrados

$$\text{SUMX}(\text{LA_daily_air_quality}, \text{LA_daily_air_quality}[\text{Carbon_monoxide}] + \text{LA_daily_air_quality}[\text{Nitrogen_dioxide}] + \text{LA_daily_air_quality}[\text{Sulphur_dioxide}] + \text{LA_daily_air_quality}[\text{Ozone}])$$

Precipitaciones promedio: Saca el promedio de las lluvias

$$\text{AVERAGE}(\text{LA_daily_climate}[\text{Precipitation_sum}])$$

Viento promedio: Saca el promedio del viento registrado

$$\text{AVERAGE}(\text{LA_daily_climate}[\text{wind_speed_10m_max}])$$

Temperatura promedio: Promedio de las temperaturas registradas

$$\text{AVERAGE}(\text{LA_daily_climate}[\text{Apparent_temperature_max}])$$

Evapotranspiración promedio: Promedio de la evapotranspiración diaria registrada

$$\text{AVERAGE}(\text{LA_daily_climate}[\text{et0_fao_evapotranspiration}])$$

Diferencia porcentual en contaminación: Calcula la diferencia en porcentaje de la contaminación total entre un periodo de tiempo y otro

$$\text{CALCULATE}(([\text{Contaminación total B}] - [\text{Contaminación total A}]) / ([\text{Contaminación total A}] + [\text{Contaminación total B}] / 2))$$

Diferencia porcentual en temperatura: Calcula la diferencia en porcentaje de la temperatura promedio entre un periodo y otro

$$\text{CALCULATE}(([\text{Temperatura promedio B}] - [\text{Temperatura promedio A}]) / ([\text{Temperatura promedio A}] + [\text{Temperatura promedio B}] / 2))$$

Tablas auxiliares

Para la última solapa se generaron dos tablas auxiliares con fechas para que funcionen independientes y se pueda lograr esa comparación entre dos periodos. Esas tablas contienen sus medidas, se decidió apartarlas de la tabla general de medidas porque funcionan independientemente. Las medidas son las siguientes:

Temperatura promedio A: Calcula la sensación térmica promedio en el periodo especificado por filtros

```
CALCULATE(AVERAGE(LA_daily_climate[Apparent_temperature_max]),USER  
ELATIONSHIP('Periodo A'[Date], 'LA_daily_climate'[Measure_date]))
```

Contaminación total A: Suma los principales agentes de contaminación en el periodo de tiempo especificado por filtros

```
CALCULATE(SUMX(LA_daily_air_quality,  
LA_daily_air_quality[Carbon_monoxide] +  
LA_daily_air_quality[Nitrogen_dioxide] + LA_daily_air_quality[Sulphur_dioxide]  
+ LA_daily_air_quality[Ozone]), USERELATIONSHIP('Periodo A'[Date],  
'LA_daily_air_quality'[Measure_date]))
```

Calidad del aire A: Calcula el promedio de la calidad del aire en el periodo de tiempo especificado por filtros

```
CALCULATE(AVERAGEX(LA_daily_air_quality, LA_daily_air_quality[Pm2_5] +  
LA_daily_air_quality[Pm10]),  
USERELATIONSHIP(LA_daily_air_quality[Measure_date], 'Periodo A'[Date]))
```

Fuente y links relacionados

https://www.kaggle.com/datasets/anycaroliny/latin-america-weather-and-air-quality-data?resource=download&select=LA_daily_climate.csv

https://github.com/anycarolinys/kaggle_weather_dataset