

Documentación

Resumen del desafío

En un principio, antes de comenzar con la lógica leí el archivo `Electric_Vehicle_Population_Data.csv` para ver que columnas tenía y como era la calidad de los datos (que tan bien estaban definidas las columnas y si tenían datos vacíos).

Una vez tuve el panorama de los datos que tenía vi los requisitos.

- How many electric vehicles are registered per year?
- What are the top 10 electric vehicle models by registration count?
- Where are the CAFV-eligible vehicles concentrated geographically?
- What is the year-over-year change in electric vehicle registrations by county?

Con eso en mente, limpié los datos de `Electric_Vehicle_Population_Data.csv` con transformaciones como pasar todas las columnas a `snake_case`, convertir columnas definidas arbitrariamente a números y a fechas, separar la columna `Vehicle Location` en latitud y longitud y rellenar valores faltantes.

Luego usé un modelo dimensional que, aunque no era necesario, se utiliza mucho en ingeniería de datos para crear un Datawarehouse, con eso en mente creé las tablas

- `dim_vehicle`: Se creó esta tabla dimensión ya que representan características conceptuales de los vehículos (no son necesarios para los cálculos de contabilidad, pero si para poder ampliar la información si se necesita)
- `dim_date`: Necesaria para cualquier modelo dimensional y útil para el posterior análisis de datos
- `dim_location`: Se creó esta tabla dimensión ya que presenta atributos que no cambian mucho con el tiempo
- `dim_electric_type`: Presenta pocos valores posibles (2) que se repiten en muchos registros
- `dim_policy`: Representa una clasificación de cada auto electrico segun politicas CAF por lo que solo si se modifica o agrega una politica esta tabla va a cambiar
- `fact_ev_registration`: Data Warehouse guarda los eventos medibles, por ejemplo, "el vehículo VIN XXXXX fue registrado en tal lugar, en tal fecha, con tales características"

Esto no solo muestra unas tablas útiles para los analistas de datos, sino que también mejoran el rendimiento de la ETL al no tener que estar insertando datos seguidamente en las tablas dimensiones.

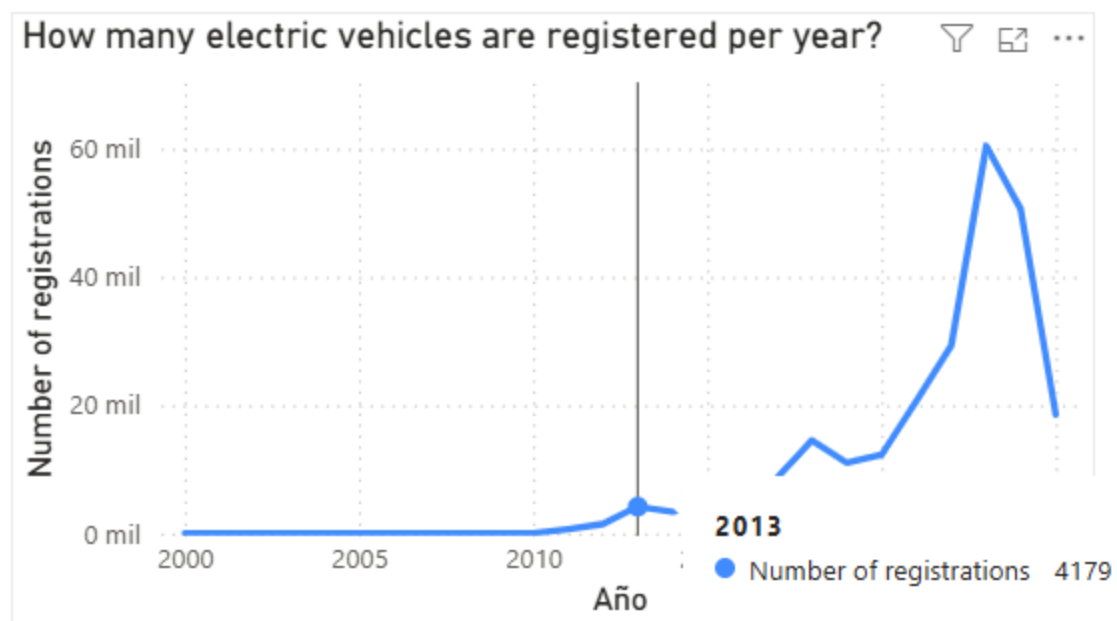
Una vez tenía todo eso, creé la carga de las tablas definidas anteriormente. Teniendo en cuenta criterios como Slowly Changing Dimension (SCD) y una selección arbitraria de inserción de datos en el Data Warehouse (como, por ejemplo, si se intentara insertar datos que no existen en el dim_vehicle se saltea esa inserción);

El desafío no me trajo muchos inconvenientes más allá de que me cuesta un poco el análisis de datos y que no manejo muy bien el inglés

Task:

- How many electric vehicles are registered per year?

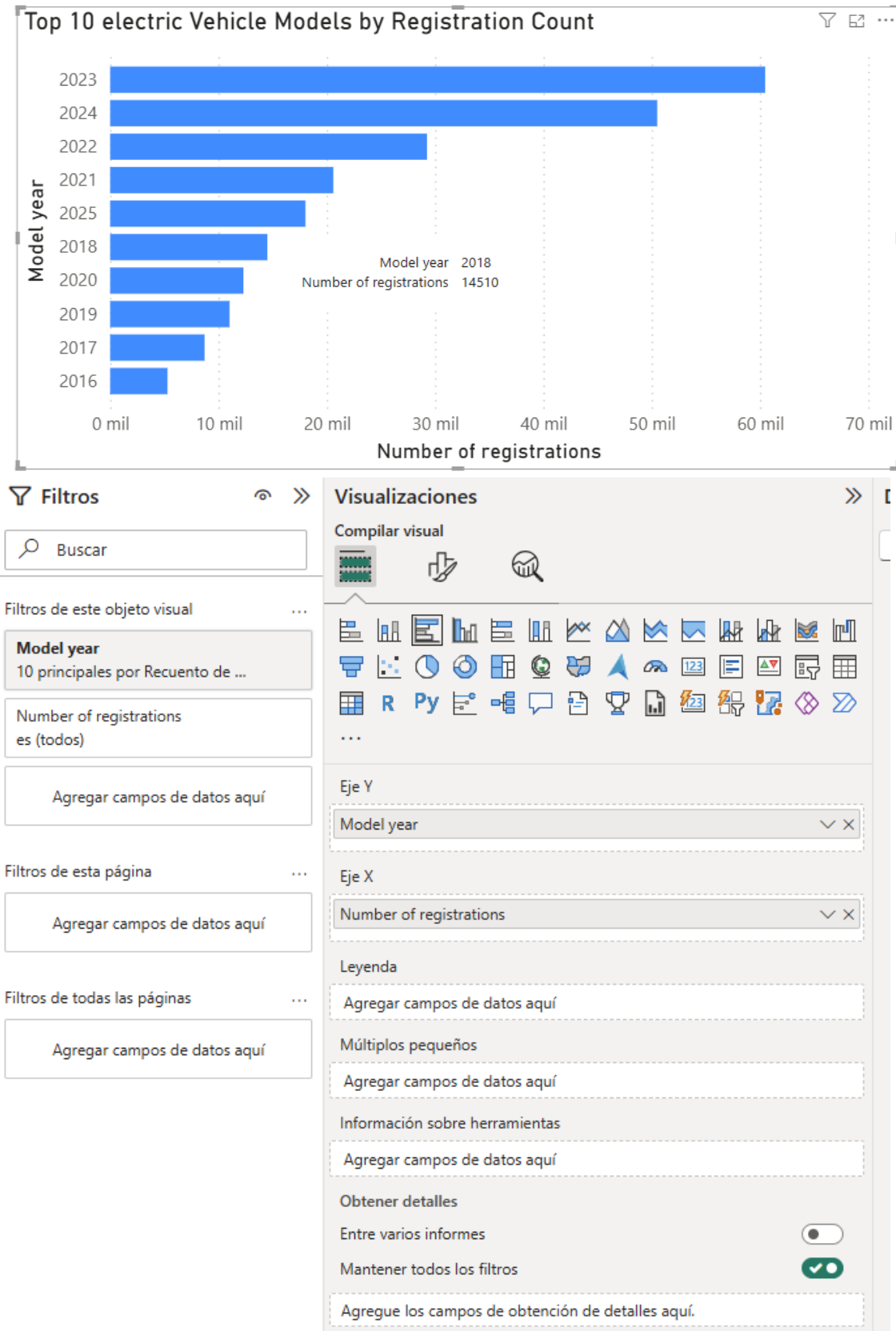
Esto se obtuvo mediante Power BI pasando el año en el eje x y un count, para cada año, de la columna id_reg (PK del Data Warehouse)



- What are the top 10 electric vehicle models by registration count?

Esto se obtuvo mediante Power BI pasando el modelo del vehiculo en el eje x y un count, para cada año, de la columna id_reg (PK del Data Warehouse)

Además, se agregó un filtro para obtener el TOP 10



Where are the CAFV-eligible vehicles concentrated geographically?

Acá usé un mapa y dos medidas, en la primera medida filtro los datos del data Warehouse, haciendo un conteo de cuantos registros de dicha tabla poseen “Clean Alternative Fuel Vehicle Eligible”

Y la segunda medida saca el porcentaje

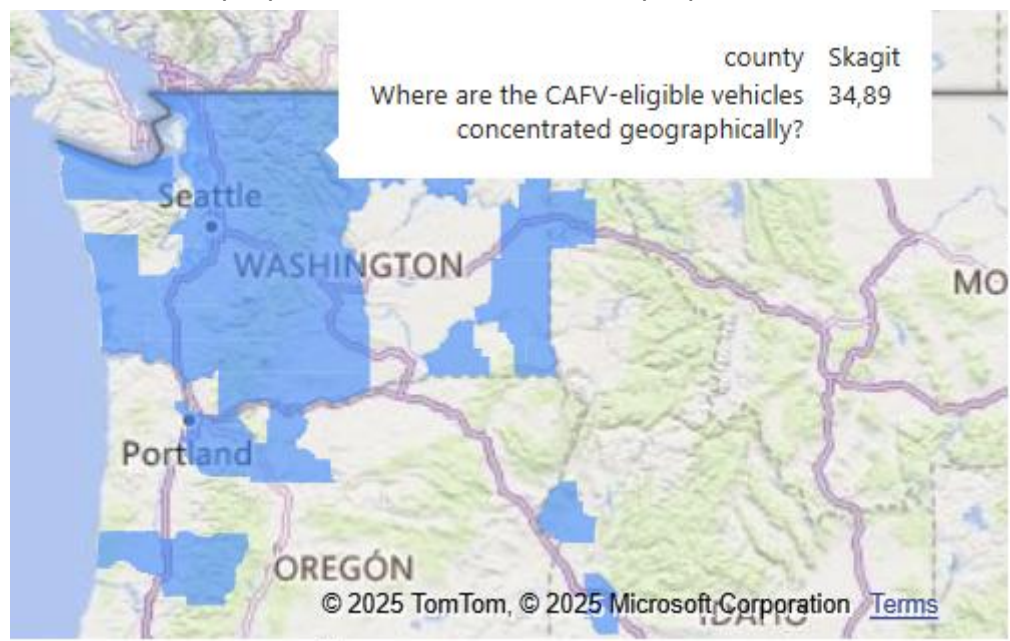
CAFV Eligible Count =

```
CALCULATE(  
    COUNT('public fact_ev_registration'[id_reg]),  
    FILTER(  
        'public dim_policy',  
        'public dim_policy'[cafv_eligibility] = "Clean Alternative Fuel Vehicle Eligible"  
    )  
)
```

CAFV Eligibility % =

```
DIVIDE(  
    [CAFV Eligible Count],  
    COUNT('public fact_ev_registration'[id_reg]),  
    0  
) * 100
```

Cabe mencionar que estos calculos son por condado de Washington, por lo que Seattle tiene su propio 100%, Tacoma tiene su propio 100% etc.



Ubicación	
county	▼ ×
latitude	▼ ×
longitude	▼ ×

Leyenda	
Agregar campos de datos aquí	
Latitud	
Agregar campos de datos aquí	