

PONTIFICIA UNIVERSIDAD CATÓLICA MADRE Y MAESTRA

Departamento de Ciencias e Ingeniería. Escuela de Ingeniería
Telemática y Sistemas



Base de datos II

ISC 349 - 101

Proyecto Final

Presentado por:

Nicol Cristal Ureña (2018-1669, ID 10135720)

*Junior Hernández (2018-0999,
ID 10135069)*

*Luis Angel Ramírez (2018-0570
ID 10134655)*

Presentado a:

Maximo Emiliano Perez Medrano

Asignatura:

Base de datos II

**SANTIAGO DE LOS CABALLEROS
REPÚBLICA DOMINICANA**

Índice

Primera Entrega

1. Descripción de objetivo general.....	3
2. Explicación breve de datasets.....	3
3. Sentencias SQL	4
4. Proceso a modelar.....	4
5. Grano del modelo.....	6
6. Diagrama modelo de datos del Ambiente Analítico.....	8
7. Diagrama modelo de datos inicial.....	9
8. ETL.....	13
9. Análisis con sentencias SQL.....	16
10. Inicio de integración de la herramienta de análisis.....	19

Segunda Entrega

11. Sentencias SQL creación estructuras parte 2.....	21
12. Preguntas de análisis de diferentes naturalezas.....	21
13. Decisiones significativas	24
14. Herramientas utilizadas.....	24
15. Descripción de datasets y fuentes de datos.....	25
16. Integración de Herramientas parte 2.....	28
17. Reflexión general de la experiencia.....	31

1. Descripción del Objetivo General

El objetivo principal de la elaboración de este proyecto es determinar y monitorear los datos acerca de los distintos contaminantes emitidos por Estados Unidos para así concluir si es posible que puedan lograr el objetivo en el acuerdo de París según los cuales se debería reducir drásticamente la emisión de gases contaminantes antes del 2025.

2. Explicar brevemente a manera general archivos fuentes (datasets). Mostrar la composición (estructura) de dichos datasets haciendo una breve explicación de los atributos que los componen.

La mostrada a continuación es la composición de las tablas que pertenecen a la Base de Datos de este proyecto, y cómo es observable están desplegadas las columnas en las tablas principales. Por otro lado, también se agregaron otros datasets para enriquecer los datos y permitir mayor espectro de análisis, los cuales son: DatasetCensus, para un estudio de la población, DatasetCiudad, para información de las ciudades utilizadas, DatasetFestivos, que es el conjunto de días no laborables o festivos en los Estados Unidos, DatasetGeo, que es un conjunto de informaciones de la composición geográfica de los estados(cantidad de espacio terrestre, acuático, y costero en millas y kilómetros), y el dataset principal.

[-]	[Table Icon]	dbo.Centro
	[-]	Columns
		DataID (int, null)
		Direccion (varchar(100), null)
		Estado (varchar(50), null)
		Condado (varchar(50), null)
		Ciudad (varchar(50), null)
		Poblacion_estado (float, null)
		Area_estado (float, null)
		Longitud_estado (float, null)
		EstadoID (varchar(50), null)
	[+]	Keys
	[+]	Constraints
	[+]	Triggers
	[+]	Indexes
	[+]	Statistics
[+]	[Table Icon]	dbo.DatasetCensus
[+]	[Table Icon]	dbo.DatasetCiudad
[+]	[Table Icon]	dbo.DatasetFestivos
[+]	[Table Icon]	dbo.DatasetGeo
[+]	[Table Icon]	dbo.DatasetPrincipal
[-]	[Table Icon]	dbo.Fecha
	[-]	Columns
		FechaID (int, null)
		Num_Dia (int, null)
		Num_Mes (int, null)
		Anio (int, null)
		Fecha_completa (date, null)
		Dia (varchar(50), null)
		Mes (varchar(50), null)
		HorasID (int, null)
		Momento (varchar(50), null)
		Dia_semana (int, null)
		Dia_Festivo (varchar(3), null)
		Num_Semana (int, null)

[-]	[Table Icon]	dbo.Hechos
	[-]	Columns
		Codigo_muestra (int, not null)
		DataID (int, null)
		FechaID (int, null)
		O3_Media (float, null)
		O3_Primer_valor_maximo (float, null)
		O3_AQI (int, null)
		CO_Media (float, null)
		CO_Primer_valor_maximo (float, null)
		CO_AQI (int, null)
		SO2_Media (float, null)
		SO2_Primer_valor_maximo (float, null)
		SO2_AQI (int, null)
		NO2_Media (float, null)
		NO2_Primer_valor_maximo (float, null)
		NO2_AQI (int, null)
[+]	[Folder Icon]	Keys
[+]	[Folder Icon]	Constraints
[+]	[Folder Icon]	Triggers
[+]	[Folder Icon]	Indexes
[+]	[Folder Icon]	Statistics
[-]	[Table Icon]	dbo.Hora
	[-]	Columns
		HoraID (int, null)
		O3_Primer_hora_maxima (int, null)
		CO_Primer_hora_maxima (int, null)
		SO2_Primer_hora_maxima (int, null)
		NO2_Primer_hora_maxima (int, null)

3. Descripción del Proceso a Modelar

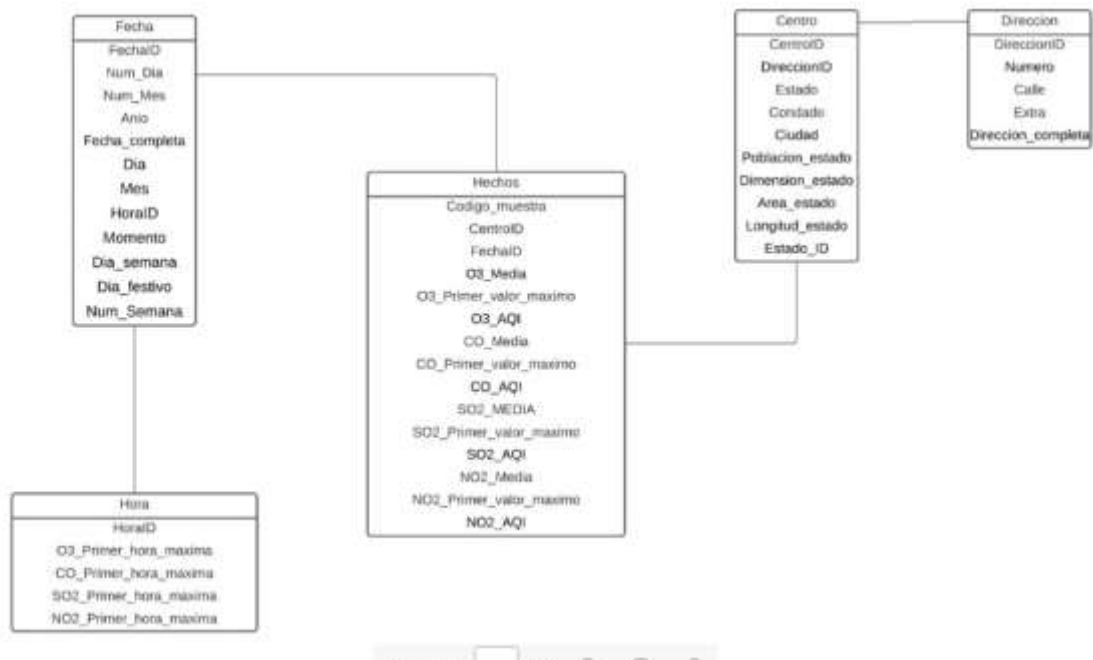
En la actualidad la sociedad es en una considerable medida ignorante del aumento peligroso de los contaminantes procedentes de EE.UU y las personas consciente que saben de susodicha problemática se preocupan que el país no pueda cumplir lo pautado para 2025. Esta ignorancia y desconocimiento del procedimiento para la visualización de información se debe a la difícil comprensión de los datos brindados por los agentes contra la contaminación del país. Por ello es necesario de una aplicación en la cual se pueda monitorear, ingresar, editar y actualizar de una manera más fácil y comprensible los aumentos o decrementos de los contaminantes por ciudad y la fecha en que fueron tomados. EL modelo registra diariamente los niveles de contaminantes en las ciudades de Estados Unidos como el dióxido de azufre (SO₂), el monóxido de carbono (CO), el óxido de nitrógeno (NO₂), y el ozono (O₃) para determinar si es posible reducir la emisión de gases contaminantes según el acuerdo París, determinando, para ello, los aumentos y decrementos en los niveles de contaminación.

4. Descripción del Grano del modelo

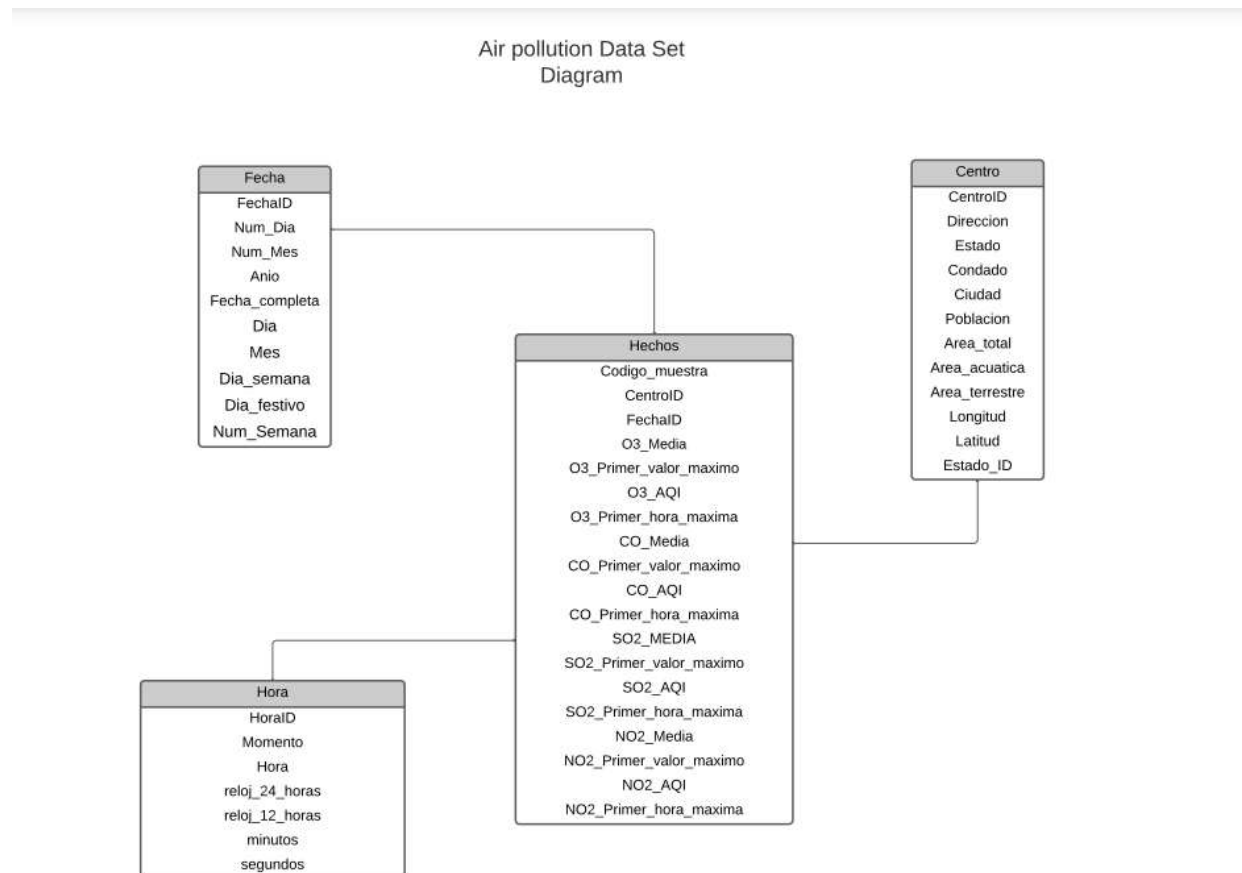
Análisis de cálculo de contaminantes por ciudad

El grano de este modelo usado para el análisis del aumento o decremento de los niveles de contaminación por ciudades tiene un nivel de detalle medio, por el motivo de que se estudia en particular las ciudades y su contaminación diaria. Se busca implementar los conceptos teóricos sobre la granularidad de datos para que los mismos no se viesan afectados a la hora de realizar las sentencias deseadas y estructurar el modelo de dimensiones originado del dataSet principal. Tomando así las referencias de los distintos atributos y dividiéndolos como se explica con anterioridad y se ven en la imagen de las tablas de dimensiones y la de hechos.

5. Diagrama de Modelo Inicial



6. Diagrama del Ambiente analítico



7. Creación de las tablas e insertado

Por otro lado, los queries que se utilizaron para la creación de las tablas fueron:

```
/*CREATE TABLE Fecha(
    FechaID int,
    Num_Dia int,
    Num_Mes int,
    Anio int,
    Fecha_completa date,
    Dia varchar(50),
    Mes varchar(50),
    HoraID int,
    Momento varchar(50),
    Dia_semana int,
    Dia_Festivo varchar(3) CHECK (Dia_Festivo IN ('SI', 'NO')),
    Num_Semana int
)

CREATE TABLE Hechos(
    Codigo_muestra int Identity(1, 1),
    CentroID int,
    FechaID int,
    O3_Media float,
    O3_Primer_valor_maximo float,
    O3_AQI int,
    CO_Media float,
    CO_Primer_valor_maximo float,
    CO_AQI int,
    SO2_Media float,
    SO2_Primer_valor_maximo float,
    SO2_AQI int,
    NO2_Media float,
    NO2_Primer_valor_maximo float,
    NO2_AQI int
)

CREATE TABLE Hora(
    HoraID int,
    O3_Primer_hora_maxima int,
    CO_Primer_hora_maxima int,
    SO2_Primer_hora_maxima int,
    NO2_Primer_hora_maxima int
)

CREATE TABLE Centro(
    CentroID int Identity(1, 1),
    DireccionID varchar(100),
    Estado varchar(50),
    Condado varchar(50),
    Ciudad varchar(50),
    Poblacion_estado float,
    Area_estado float,
    Longitud_estado float,
    EstadoID varchar(50)
)

)*/
```

Otras consultas empleadas fueron:

```
INSERT INTO
Hora(HoraID,O3_Primer_hora_maxima,CO_Primer_hora_maxima,SO2_Primer_hora_maxima,NO2_Primer_hora_maxima)
SELECT * from tabla_hora
```

```
INSERT INTO
Centro(DireccionID,Estado,Condado,Ciudad,Poblacion_estado,Area_estado,Longitud_estado,EstadoID)
SELECT * from tabla_centro
```

```
INSERT INTO
Fecha(FechaID,Num_Dia,Num_Mes,Anio,Fecha_completa,Dia,Mes,HoraID,Momento,Dia_semana,Dia_Festivo,Num_Semana)
SELECT * from tabla_fecha
```

```
INSERT INTO
Hechos(CentroID,FechaID,O3_Media,O3_Primer_valor_maximo,O3_AQI,CO_Media,CO_Primer_valor_maximo,CO_AQI,
SO2_Media,SO2_Primer_valor_maximo,SO2_AQI,NO2_Media,NO2_Primer_valor_maximo,NO2_AQI)
SELECT * from tabla_hechos
```

Y tomando en consideración las vistas, se crearon de la siguiente manera:

```
CREATE VIEW tabla_fecha AS
SELECT DISTINCT DatasetPrincipal.DataId, DatasetPrincipal.Day, DatasetPrincipal.Month,
DatasetPrincipal.Year, DatasetPrincipal.Date, DATENAME(dw,DatasetPrincipal.Date) as dia_sema,
DATENAME(month, DatasetPrincipal.Date) as Mes,Hora.HoraID,(CASE WHEN Hora.O3_Primer_hora_maxima
BETWEEN 6 AND 12 THEN 'MAÑANA' WHEN Hora.O3_Primer_hora_maxima BETWEEN 12 AND 19 THEN 'TARDE'
WHEN Hora.O3_Primer_hora_maxima BETWEEN 19 AND 24 THEN 'NOCHE'
WHEN Hora.O3_Primer_hora_maxima BETWEEN 1 AND 6 THEN 'MADRUGADA'
WHEN Hora.CO_Primer_hora_maxima BETWEEN 6 AND 12 THEN 'MAÑANA'
WHEN Hora.CO_Primer_hora_maxima BETWEEN 12 AND 19 THEN 'TARDE'
WHEN Hora.CO_Primer_hora_maxima BETWEEN 19 AND 24 THEN 'NOCHE'
WHEN Hora.CO_Primer_hora_maxima BETWEEN 1 AND 6 THEN 'MADRUGADA'
WHEN Hora.SO2_Primer_hora_maxima BETWEEN 6 AND 12 THEN 'MAÑANA'
WHEN Hora.SO2_Primer_hora_maxima BETWEEN 12 AND 19 THEN 'TARDE'
WHEN Hora.SO2_Primer_hora_maxima BETWEEN 19 AND 24 THEN 'NOCHE'
WHEN Hora.SO2_Primer_hora_maxima BETWEEN 1 AND 6 THEN 'MADRUGADA'
WHEN Hora.NO2_Primer_hora_maxima BETWEEN 6 AND 12 THEN 'MAÑANA'
WHEN Hora.NO2_Primer_hora_maxima BETWEEN 12 AND 19 THEN 'TARDE'
WHEN Hora.NO2_Primer_hora_maxima BETWEEN 19 AND 24 THEN 'NOCHE'
WHEN Hora.NO2_Primer_hora_maxima BETWEEN 1 AND 6 THEN 'MADRUGADA' END) as momento, DATEPART(WEEKDAY,
DatasetPrincipal.Date) as Dia_semana,(CASE WHEN DatasetFestivos.Date = DatasetPrincipal.Date THEN 'SI'
ELSE 'NO' END) as es_festivo,DATEPART(ISO_WEEK, DatasetPrincipal.Date) as Num_semana
FROM DatasetPrincipal, Hora, DatasetFestivos
WHERE DatasetPrincipal.DataId = Hora.HoraID AND DatasetPrincipal.Year = DatasetFestivos.Year and
DatasetPrincipal.Day = DatasetFestivos.Day and DatasetPrincipal.Month = DatasetFestivos.Month
```

```
CREATE VIEW tabla_hora AS
SELECT DatasetPrincipal.DataID ,DatasetPrincipal.O3_1st_Max_Hour, DatasetPrincipal.CO_1st_Max_Hour,
DatasetPrincipal.SO2_1st_Max_Hour, DatasetPrincipal.NO2_1st_Max_Hour From DatasetPrincipal
```

```
CREATE VIEW tabla_hechos AS
SELECT Centro.CentroID,Fecha.FechaID, DatasetPrincipal.O3_Mean,
DatasetPrincipal.O3_1st_Max_Value, DatasetPrincipal.O3_AQI, DatasetPrincipal.CO_Mean,
DatasetPrincipal.CO_1st_Max_Value, DatasetPrincipal.CO_AQI, DatasetPrincipal.SO2_Mean,
DatasetPrincipal.SO2_1st_Max_Value, DatasetPrincipal.SO2_AQI,
DatasetPrincipal.NO2_Mean, DatasetPrincipal.NO2_1st_Max_Value,
DatasetPrincipal.NO2_AQI
FROM Centro,DatasetPrincipal,Fecha
WHERE Centro.CentroID = DatasetPrincipal.DataID AND DatasetPrincipal.DataID =
Fecha.FechaID
```

```

CREATE VIEW tabla_centro AS
SELECT DatasetPrincipal.DataID, DatasetPrincipal.Address as Direccion, DatasetCiudad.state as Estado,
DatasetPrincipal.City as Ciudad, DatasetPrincipal.County as Condado, DatasetCiudad._2010_census as
Poblacion, DatasetCiudad.land_area_km as Area , DatasetGeo.Longitud, (CASE WHEN DatasetCiudad.state =
'Alabama' THEN 'AL' WHEN DatasetCiudad.state = 'Alaska' THEN 'AK' WHEN DatasetCiudad.state = 'Arizona'
THEN 'AZ' WHEN DatasetCiudad.state = 'Arkansas' THEN 'AR' WHEN DatasetCiudad.state = 'California' THEN
'CA' WHEN DatasetCiudad.state = 'Colorado' THEN 'CO' WHEN DatasetCiudad.state = 'Connecticut' THEN
'CT' WHEN DatasetCiudad.state = 'Delaware' THEN 'DE' WHEN DatasetCiudad.state = 'District of Columbia'
THEN 'DC' WHEN DatasetCiudad.state = 'Florida' THEN 'FL' WHEN DatasetCiudad.state = 'Georgia' THEN
'GA' WHEN DatasetCiudad.state = 'Hawaii' THEN 'HI' WHEN DatasetCiudad.state = 'Idaho' THEN 'ID' WHEN
DatasetCiudad.state = 'Illinois' THEN 'IL' WHEN DatasetCiudad.state = 'Indiana' THEN 'WV' WHEN
DatasetCiudad.state = 'Iowa' THEN 'IA' WHEN DatasetCiudad.state = 'Kansas' THEN 'KS' WHEN
DatasetCiudad.state = 'Kentucky' THEN 'KY' WHEN DatasetCiudad.state = 'Louisiana' THEN 'LA' WHEN
DatasetCiudad.state = 'Maine' THEN 'ME' WHEN DatasetCiudad.state = 'Maryland' THEN 'MD' WHEN
DatasetCiudad.state = 'Massachusetts' THEN 'MA' WHEN DatasetCiudad.state = 'Michigan' THEN 'MI' WHEN
DatasetCiudad.state = 'Minnesota' THEN 'MN' WHEN DatasetCiudad.state = 'Mississippi' THEN 'MS' WHEN
DatasetCiudad.state = 'Missouri' THEN 'MO' WHEN DatasetCiudad.state = 'Montana' THEN 'MT' WHEN
DatasetCiudad.state = 'Nebraska' THEN 'NE' WHEN DatasetCiudad.state = 'Nevada' THEN 'NV' WHEN
DatasetCiudad.state = 'New Hampshire' THEN 'NH' WHEN DatasetCiudad.state = 'New Jersey' THEN 'NJ' WHEN
DatasetCiudad.state = 'New Mexico' THEN 'NM' WHEN DatasetCiudad.state = 'New York' THEN 'NY' WHEN
DatasetCiudad.state = 'North Carolina' THEN 'NC' WHEN DatasetCiudad.state = 'North Dakota' THEN 'ND'
WHEN DatasetCiudad.state = 'Ohio' THEN 'OH' WHEN DatasetCiudad.state = 'Oklahoma' THEN 'OK' WHEN
DatasetCiudad.state = 'Oregon' THEN 'ORE' WHEN DatasetCiudad.state = 'Pennsylvania' THEN 'PA' WHEN
DatasetCiudad.state = 'Rhode Island' THEN 'RI' WHEN DatasetCiudad.state = 'South Carolina' THEN 'SC'
WHEN DatasetCiudad.state = 'South Dakota' THEN 'SD' WHEN DatasetCiudad.state = 'Tennessee' THEN 'TN'
WHEN DatasetCiudad.state = 'Texas' THEN 'TX' WHEN DatasetCiudad.state = 'Utah' THEN 'UT' WHEN
DatasetCiudad.state = 'Vermont' THEN 'VT' WHEN DatasetCiudad.state = 'Virginia' THEN 'VA' WHEN
DatasetCiudad.state = 'Washington' THEN 'Wash.' WHEN DatasetCiudad.state = 'West Virginia' THEN
'W.Va.' WHEN DatasetCiudad.state = 'Wisconsin' THEN 'Wis.' WHEN DatasetCiudad.state = 'Wyoming' THEN
'Wyo.' END) as EstadoID
FROM DatasetCensus, DatasetGeo, DatasetPrincipal, DatasetCiudad
WHERE DatasetPrincipal.Stat = DatasetGeo.Estado AND DatasetPrincipal.Stat = DatasetCensus.STNAME AND
DatasetPrincipal.City = DatasetCiudad.city and DatasetPrincipal.County=DatasetCensus.CTYNAME

```

Las restricciones se aplicaron en el DatasetPrincipal de la siguiente manera:

```

alter table DatasetPrincipal
ADD CONSTRAINT Chk_Anio CHECK(Year >= 2000), CONSTRAINT Chk_Mes CHECK(Month >= 1 and Month <=12),
CONSTRAINT Chk_Day CHECK(Day >= 1 and Day <=31),CONSTRAINT Chk_Addres CHECK(Address <>''),CONSTRAINT
Chk_State CHECK(Stat='Alabama' or Stat='Alaska' or Stat='Arizona' or Stat='Arkansas' or
Stat='California' or Stat='Colorado' or Stat='Connecticut' or Stat='Delaware' or Stat='District of
Columbia' or Stat='Florida' or Stat='Georgia' or Stat='Hawaii' or Stat='Idaho' or Stat='Illinois' or
Stat='Indiana' or Stat='Iowa' or Stat='Kansas' or Stat='Kentucky' or Stat='Louisiana' or Stat='Maine' or
Stat='Maryland' or Stat='Massachusetts' or Stat='Michigan' or Stat='Minnesota' or Stat='Mississippi' or
Stat='Missouri' or Stat='Montana' or Stat='Nebraska' or Stat='Nevada' or Stat='New Hampshire' or
Stat='New Jersey' or Stat='New Mexico' or Stat='New York' or Stat='North Carolina' or Stat='North
Dakota' or Stat='Ohio' or Stat='Oklahoma' or Stat='Oregon' or Stat='Pennsylvania' or Stat='Rhode Island'
or Stat='South Carolina' or Stat='South Dakota' or Stat='Tennessee' or Stat='Texas' or Stat='Utah' or
Stat='Vermont' or Stat='Virginia' or Stat='Washington' or Stat='West Virginia' or Stat='Wisconsin' or
Stat='Wyoming'), CONSTRAINT Chk_County CHECK(County <>''), CONSTRAINT Chk_City CHECK(City <>''),
CONSTRAINT Chk_O3Mean CHECK(O3_Mean >= 0),CONSTRAINT Chk_O31stValue CHECK(O3_1st_Max_Value >=
0),CONSTRAINT Chk_O3Maxhour CHECK(O3_1st_Max_Hour >= 0 and O3_1st_Max_Hour <= 24),CONSTRAINT Chk_O3AQI
CHECK(O3_AQI >= 0),CONSTRAINT Chk_COMean CHECK(CO_Mean >= 0),CONSTRAINT Chk_CO1stValue
CHECK(CO_1st_Max_Value >= 0),CONSTRAINT Chk_COMaxhour CHECK(CO_1st_Max_Hour >= 0 and CO_1st_Max_Hour <=
24),CONSTRAINT Chk_COAQI CHECK(CO_AQI >= 0), CONSTRAINT Chk_SO2Mean CHECK(SO2_Mean >= 0),CONSTRAINT
Chk_SO21stValue CHECK(SO2_1st_Max_Value >= 0),CONSTRAINT Chk_SO2Maxhour CHECK(SO2_1st_Max_Hour >= 0 and
SO2_1st_Max_Hour <= 24),CONSTRAINT Chk_SO2AQI CHECK(SO2_AQI >= 0),CONSTRAINT Chk_NO2Mean CHECK(NO2_Mean
>= 0),CONSTRAINT Chk_NO21stValue CHECK(NO2_1st_Max_Value >= 0),CONSTRAINT Chk_NO2Maxhour
CHECK(NO2_1st_Max_Hour >= 0 and NO2_1st_Max_Hour <= 24),CONSTRAINT Chk_NO2AQI CHECK(NO2_AQI >= 0),
CONSTRAINT UN_DataID unique(DataId)

```

8. ETL

Para el ETL, el proceso y las consultas para la actualización y agregación de datos fueron las siguientes:

En primer lugar, se tomó el dataset a utilizar, DatasetPrincipal en nuestro caso, lo dividimos en 3 tablas de dimensiones, Hora, Fecha y Centro, y en 1 tabla de hechos, Hechos, para así tener mejor distribución de datos y para analizar los mismos en un futuro. Luego rellenos las tablas de dimensiones con dataset adicionales, uno para los estados y sus atributos, para fechas, fuesen festivas o no y también para mediciones sobre los estados. Estas tablas de dimensiones, así como la de hechos, fue llenada con las vistas mencionadas anteriormente.

Ya para la actualización de las tablas de hechos y dimensiones se utilizaron las siguientes consultas:

```
/*  
CREATE PROCEDURE insertar_especifico  
(@ID int)  
AS  
BEGIN  
SET NOCOUNT ON  
  
INSERT INTO  
Hora(HoraID,O3_Primer_hora_maxima,CO_Primer_hora_maxima,SO2_Primer_hora_maxima,NO2_Primer_hora_maxima)  
SELECT * from tabla_hora WHERE DataID = (@ID)
```

```

INSERT INTO
Fecha(DataID,Num_Dia,Num_Mes,Anio,Fecha_completa,Dia,Mes,HoraID,Momento,Dia_semana,Dia_Festivo,Num_Semana)
SELECT * from tabla_fecha WHERE DataID = (@ID)

INSERT INTO
Centro(DataID,Direccion,Estado,Condado,Ciudad,Poblacion_estado,Area_estado,Longitud_estado,EstadoID)
SELECT * FROM tabla_centro WHERE DataID= (@ID)

INSERT INTO
Hechos(DataID,FechaID,03_Media,03_Primer_valor_maximo,03_AQI,CO_Media,CO_Primer_valor_maximo,CO_AQI,SO2_Media,SO2_Primer_valor_maximo,SO2_AQI,NO2_Media,NO2_Primer_valor_maximo,NO2_AQI)
SELECT * from tabla_hechos WHERE DataID = (@ID)

END

GO

CREATE PROCEDURE actualizar_especifico
(@ID int)
AS
BEGIN
SET NOCOUNT ON

UPDATE Hora SET Hora.03_Primer_hora_maxima = tabla_hora.03_1st_Max_Hour, Hora.CO_Primer_hora_maxima =
tabla_hora.CO_1st_Max_Hour, Hora.SO2_Primer_hora_maxima = tabla_hora.SO2_1st_Max_Hour,
Hora.NO2_Primer_hora_maxima = tabla_hora.NO2_1st_Max_Hour
FROM Hora, tabla_hora
WHERE Hora.HoraID = tabla_hora.DataID AND Hora.HoraID = (@ID)

UPDATE Fecha SET Fecha.Num_Dia = tabla_fecha.Day, Fecha.Num_Mes = tabla_fecha.Month, Fecha.Anio =
tabla_fecha.Year, Fecha.Fecha_completa = tabla_fecha.Date, Fecha.Dia = tabla_fecha.dia_sema, Fecha.Mes
= tabla_fecha.Mes, Fecha.HoraID = tabla_fecha.HoraID, Fecha.Momento = tabla_fecha.momento,
Fecha.Dia_semana = tabla_fecha.Dia_semana, Fecha.Dia_festivo = tabla_fecha.es_festivo,
Fecha.Num_Semana = tabla_fecha.Num_Semana
FROM Fecha, tabla_fecha
WHERE Fecha.FechaID = tabla_fecha.DataID AND Fecha.FechaID = (@ID)

UPDATE Centro SET Centro.Direccion = tabla_centro.Direccion, Centro.Estado = tabla_centro.Estado,
Centro.Condado = tabla_centro.Condado, Centro.Ciudad = tabla_centro.Ciudad, Centro.Poblacion_estado =
tabla_centro.Poblacion, Centro.Area_estado = tabla_centro.Area, Centro.Longitud_estado =
tabla_centro.Longitud, Centro.EstadoID = tabla_centro.EstadoID
FROM Centro, tabla_centro
WHERE Centro.DataID = tabla_centro.DataID AND Centro.DataID = (@ID)

UPDATE Hechos SET Hechos.DataID = tabla_hechos.DataID, Hechos.FechaID = tabla_hechos.FechaID,
Hechos.03_Media = tabla_hechos.03_Mean, Hechos.03_Primer_valor_maximo = tabla_hechos.03_1st_Max_Value,
Hechos.03_AQI = tabla_hechos.03_AQI, Hechos.CO_Media =
tabla_hechos.CO_Mean,Hechos.CO_Primer_valor_maximo = tabla_hechos.CO_1st_Max_Value,Hechos.CO_AQI =
tabla_hechos.CO_AQI, Hechos.SO2_Media = tabla_hechos.SO2_Mean,Hechos.SO2_Primer_valor_maximo =
tabla_hechos.SO2_1st_Max_Value,Hechos.SO2_AQI = tabla_hechos.SO2_AQI,Hechos.NO2_Media =
tabla_hechos.NO2_Mean,Hechos.NO2_Primer_valor_maximo = tabla_hechos.NO2_1st_Max_Value,Hechos.NO2_AQI =
tabla_hechos.NO2_AQI
FROM Hechos, tabla_hechos
WHERE Hechos.DataID = tabla_Hechos.DataID AND Hechos.DataID = (@ID)
END

GO

CREATE PROCEDURE actualizar
AS
BEGIN
SET NOCOUNT ON

UPDATE Hora SET Hora.03_Primer_hora_maxima = tabla_hora.03_1st_Max_Hour, Hora.CO_Primer_hora_maxima =
tabla_hora.CO_1st_Max_Hour, Hora.SO2_Primer_hora_maxima = tabla_hora.SO2_1st_Max_Hour,
Hora.NO2_Primer_hora_maxima = tabla_hora.NO2_1st_Max_Hour
FROM Hora, tabla_hora
WHERE Hora.HoraID = tabla_hora.DataID

UPDATE Fecha SET Fecha.Num_Dia = tabla_fecha.Day, Fecha.Num_Mes = tabla_fecha.Month, Fecha.Anio =
tabla_fecha.Year, Fecha.Fecha_completa = tabla_fecha.Date, Fecha.Dia = tabla_fecha.dia_sema, Fecha.Mes
= tabla_fecha.Mes, Fecha.HoraID = tabla_fecha.HoraID, Fecha.Momento = tabla_fecha.momento,
Fecha.Dia_semana = tabla_fecha.Dia_semana, Fecha.Dia_festivo = tabla_fecha.es_festivo,
Fecha.Num_Semana = tabla_fecha.Num_Semana

```

```
FROM Fecha, tabla_fecha
WHERE Fecha.FechaID = tabla_fecha.DataID
```

```
UPDATE Centro SET Centro.Direccion = tabla_centro.Direccion, Centro.Estado = tabla_centro.Estado,
Centro.Condado = tabla_centro.Condado, Centro.Ciudad = tabla_centro.Ciudad, Centro.Poblacion_estado =
tabla_centro.Poblacion, Centro.Area_estado = tabla_centro.Area, Centro.Longitud_estado =
tabla_centro.Longitud, Centro.EstadoID = tabla_centro.EstadoID
FROM Centro, tabla_centro
WHERE Centro.DataID = tabla_centro.DataID
```

```
UPDATE Hechos SET Hechos.DataID = tabla_hechos.DataID, Hechos.FechaID = tabla_hechos.FechaID,
Hechos.O3_Media = tabla_hechos.O3_Mean, Hechos.O3_Primer_valor_maximo = tabla_hechos.O3_1st_Max_Value,
Hechos.O3_AQI = tabla_hechos.O3_AQI, Hechos.CO_Media =
tabla_hechos.CO_Mean, Hechos.CO_Primer_valor_maximo = tabla_hechos.CO_1st_Max_Value, Hechos.CO_AQI =
tabla_hechos.CO_AQI, Hechos.SO2_Media = tabla_hechos.SO2_Mean, Hechos.SO2_Primer_valor_maximo =
tabla_hechos.SO2_1st_Max_Value, Hechos.SO2_AQI = tabla_hechos.SO2_AQI, Hechos.NO2_Media =
tabla_hechos.NO2_Mean, Hechos.NO2_Primer_valor_maximo = tabla_hechos.NO2_1st_Max_Value, Hechos.NO2_AQI =
tabla_hechos.NO2_AQI
FROM Hechos, tabla_hechos
WHERE Hechos.DataID = tabla_Hechos.DataID
```

```
END
```

```
GO
```

```
CREATE PROCEDURE insertar
AS
BEGIN
SET NOCOUNT ON
```

```
INSERT INTO
Hora(HoraID,O3_Primer_hora_maxima,CO_Primer_hora_maxima,SO2_Primer_hora_maxima,NO2_Primer_hora_maxima)
SELECT DISTINCT t1.DataID,
t1.O3_1st_Max_Hour,t1.CO_1st_Max_Hour,t1.SO2_1st_Max_Hour,t1.NO2_1st_Max_Hour from tabla_hora t1 LEFT
JOIN Hora t2 ON t2.HoraID = t1.DataID WHERE t2.HoraID is NULL
```

```
INSERT INTO
Fecha(FechaID,Num_Dia,Num_Mes,Anio,Fecha_completa,Dia,Mes,HoraID,Momento,Dia_semana,Dia_Festivo,Num_Semana)
SELECT DISTINCT t1.DataId,
t1.Day,t1.Month,t1.Year,t1.Date,t1.dia_sema,t1.Mes,t1.HoraID,t1.Momento,t1.Dia_semana,t1.es_festivo,t1
.Num_Semana from tabla_fecha t1 LEFT JOIN Fecha t2 ON t2.FechaID = t1.DataId WHERE t2.FechaID is NULL
```

```
INSERT INTO
Centro(DataID,Direccion,Estado,Condado,Ciudad,Poblacion_estado,Area_estado,Longitud_estado,EstadoID)
SELECT DISTINCT
t1.DataID,t1.Direccion,t1.Estado,t1.Condado,t1.Ciudad,t1.Poblacion,t1.Area,t1.Longitud,t1.EstadoID
FROM tabla_centro t1 LEFT JOIN Centro t2 ON t2.DataID = t1.DataID WHERE t2.DataID = NULL
```

```
INSERT INTO
Hechos(DataID,FechaID,O3_Media,O3_Primer_valor_maximo,O3_AQI,CO_Media,CO_Primer_valor_maximo,CO_AQI,SO
2_Media,SO2_Primer_valor_maximo,SO2_AQI,NO2_Media,NO2_Primer_valor_maximo,NO2_AQI)
SELECT DISTINCT
t1.DataID,t1.FechaID,t1.O3_Mean,t1.O3_1st_Max_Value,t1.O3_AQI,t1.CO_Mean,t1.CO_1st_Max_Value,t1.CO_AQI
,t1.SO2_Mean,t1.SO2_1st_Max_Value,t1.SO2_AQI,t1.NO2_Mean,t1.NO2_1st_Max_Value,t1.NO2_AQI from
tabla_hechos t1 LEFT JOIN Hechos t2 ON t2.DataID = t1.DataID WHERE t2.DataID is NULL
```

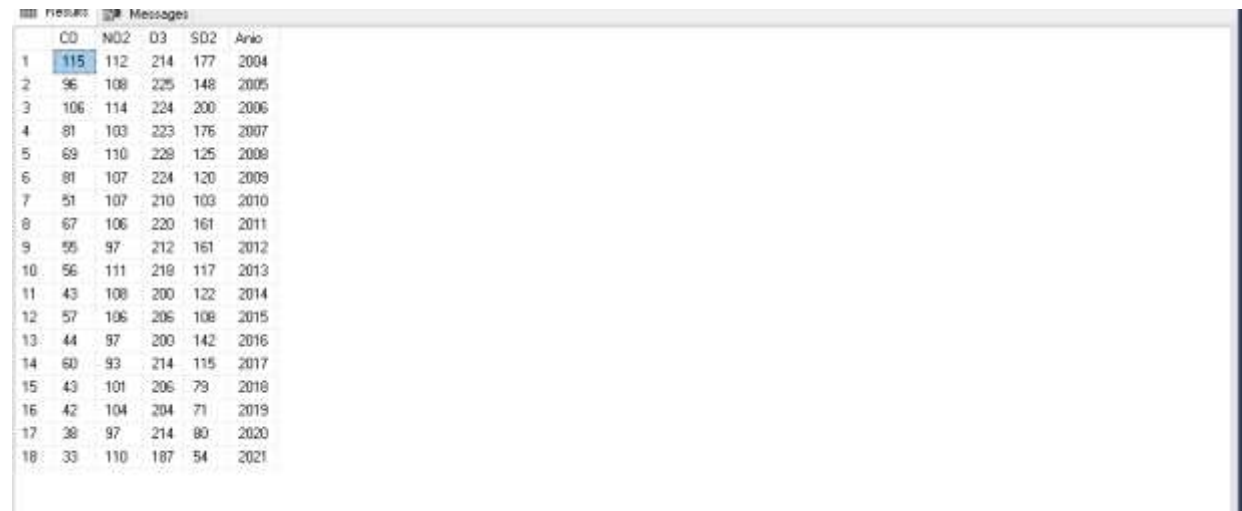
```
END
```

9. Preguntas Analíticas

En cuanto a las preguntas analíticas realizamos las siguientes:

--¿Cuáles son los valores entre los contaminantes más altos por año?

```
/*  
SELECT MAX(Hechos.CO_AQI) as CO, MAX(Hechos.NO2_AQI) as NO2, MAX(Hechos.O3_AQI) as O3,  
MAX(Hechos.SO2_AQI) as SO2, Fecha.Anio  
FROM Hechos, Fecha  
WHERE Fecha.FechaID = Hechos.FechaID GROUP BY Fecha.Anio ORDER BY Fecha.Anio
```



	CO	NO2	O3	SO2	Anio
1	115	112	214	177	2004
2	96	108	225	148	2005
3	106	114	224	200	2006
4	81	103	223	176	2007
5	69	110	228	125	2008
6	81	107	224	120	2009
7	51	107	210	103	2010
8	67	106	220	161	2011
9	55	97	212	161	2012
10	56	111	219	117	2013
11	43	108	200	122	2014
12	57	106	206	108	2015
13	44	97	200	142	2016
14	60	93	214	115	2017
15	43	101	206	79	2018
16	42	104	204	71	2019
17	38	97	214	80	2020
18	33	110	187	54	2021

--¿Cuáles estados presentan más contaminación? Rank de los primeros 5

```
SELECT top (5) Centro.Estado, MAX(Hechos.CO_AQI) as CO, MAX(Hechos.NO2_AQI) as NO2, MAX(Hechos.O3_AQI) as  
O3, MAX(Hechos.SO2_AQI) as SO2  
FROM Hechos, Centro  
WHERE Centro.DataID = Hechos.DataID GROUP BY Centro.Estado ORDER BY sum(Hechos.CO_AQI +  
Hechos.NO2_AQI+Hechos.O3_AQI+Hechos.SO2_AQI) desc
```

	Estado	CO	NO2	O3	SO2
1	California	115	112	228	153
2	Pennsylvania	38	106	217	177
3	District of Columbia	45	107	177	107
4	Texas	65	103	214	126
5	Arizona	47	88	187	92

--- ¿Qué momento del día (mañana, tarde, noche) presenta mayores media (que ya viene en la tabla de hechos) de contaminación?

```
SELECT TOP 1 Fecha.FechaID, Fecha.Momento, Fecha.Fecha_completa, SUM(Hechos.NO2_Media + Hechos.O3_Media + Hechos.SO2_Media+CO_Media) as Maximo
FROM Fecha, Hechos
WHERE Fecha.FechaID = Hechos.FechaID
GROUP BY Fecha.FechaID, Fecha.Momento, Fecha.Fecha_completa
ORDER BY Maximo DESC
```

	FechaID	Momento	Fecha_completa	Maximo
1	103498	MAÑANA	2004-02-14	505 637735001743

--¿Qué días de la semana presentan los niveles de contaminación más altos?

```
SELECT Fecha.FechaID, Fecha.Dia, Fecha.Fecha_completa ,MAX(Hechos.NO2_Media + Hechos.O3_Media +
Hechos.SO_Media+CO_Media) FROM (SELECT Hechos.FechaID, SUM(Hechos.NO2_Media + Hechos.O3_Media +
Hechos.SO_Media+CO_Media) AS totalContaminacion FROM Hechos GROUP BY Hechos.FechaID) AS suma, Fecha,
Hechos
```

```
WHERE Fecha.FechaID = Hechos.FechaID
```

```
GROUP BY Fecha.FechaID, Fecha.Dia, Fecha.Fecha_completa
```

```
ORDER BY FechaID
```

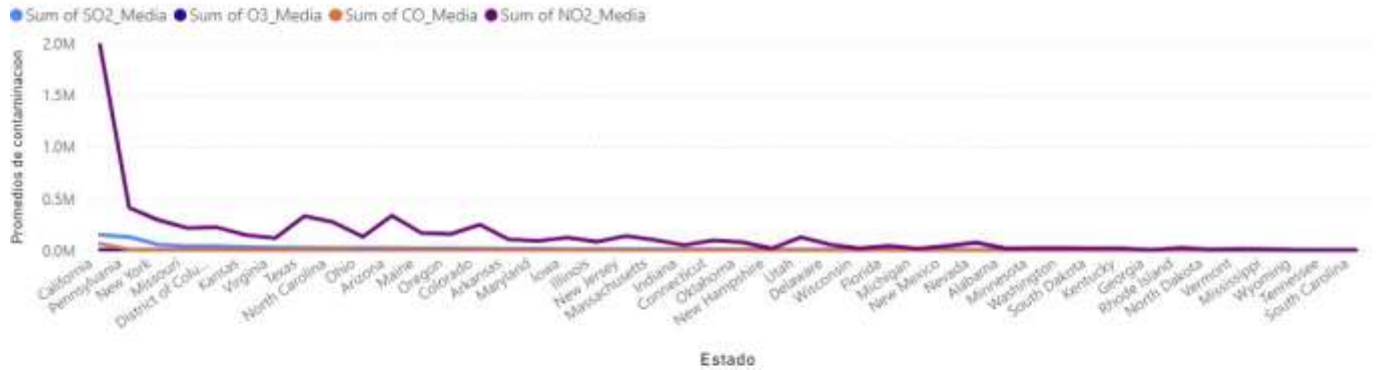
	FechaID	Dia	Fecha_completa	Total_Contaminacion
1	93770	Thursday	2004-01-08	57.9784120237455
2	93773	Sunday	2004-01-11	37.1303720185533
3	93774	Monday	2004-01-12	34.0568457385525
4	93782	Tuesday	2004-01-20	36.1679400820285
5	93784	Thursday	2004-01-22	23.7964500645176
6	93788	Monday	2004-01-26	42.0126449819654
7	93790	Wednesday	2004-01-28	37.8760139187798
8	93792	Friday	2004-01-30	38.665284043178
9	93793	Saturday	2004-01-31	27.8276969566941
10	93805	Thursday	2004-02-12	31.8104609958828
11	93808	Sunday	2004-02-15	35.1404219921678
12	93809	Monday	2004-02-16	42.5561940772459
13	93817	Tuesday	2004-02-24	32.8797959201038
14	93819	Thursday	2004-02-26	37.0468120593578
15	93823	Monday	2004-03-01	35.8981721159071
16	93825	Wednesday	2004-03-03	26.4016829449683
17	93827	Friday	2004-03-05	30.8577590640634
18	93828	Saturday	2004-03-06	33.0481170229614
19	93839	Friday	2004-03-19	43.2327445065603

En el dataset utilizado para la consulta anterior los valores ya han sido calculados mediante las formulas propicias según la agencia de medioambiente de Estados Unidos.

10. Integración inicial a una herramienta analítica

Para la integración del ETL en el análisis, realizamos los siguientes:

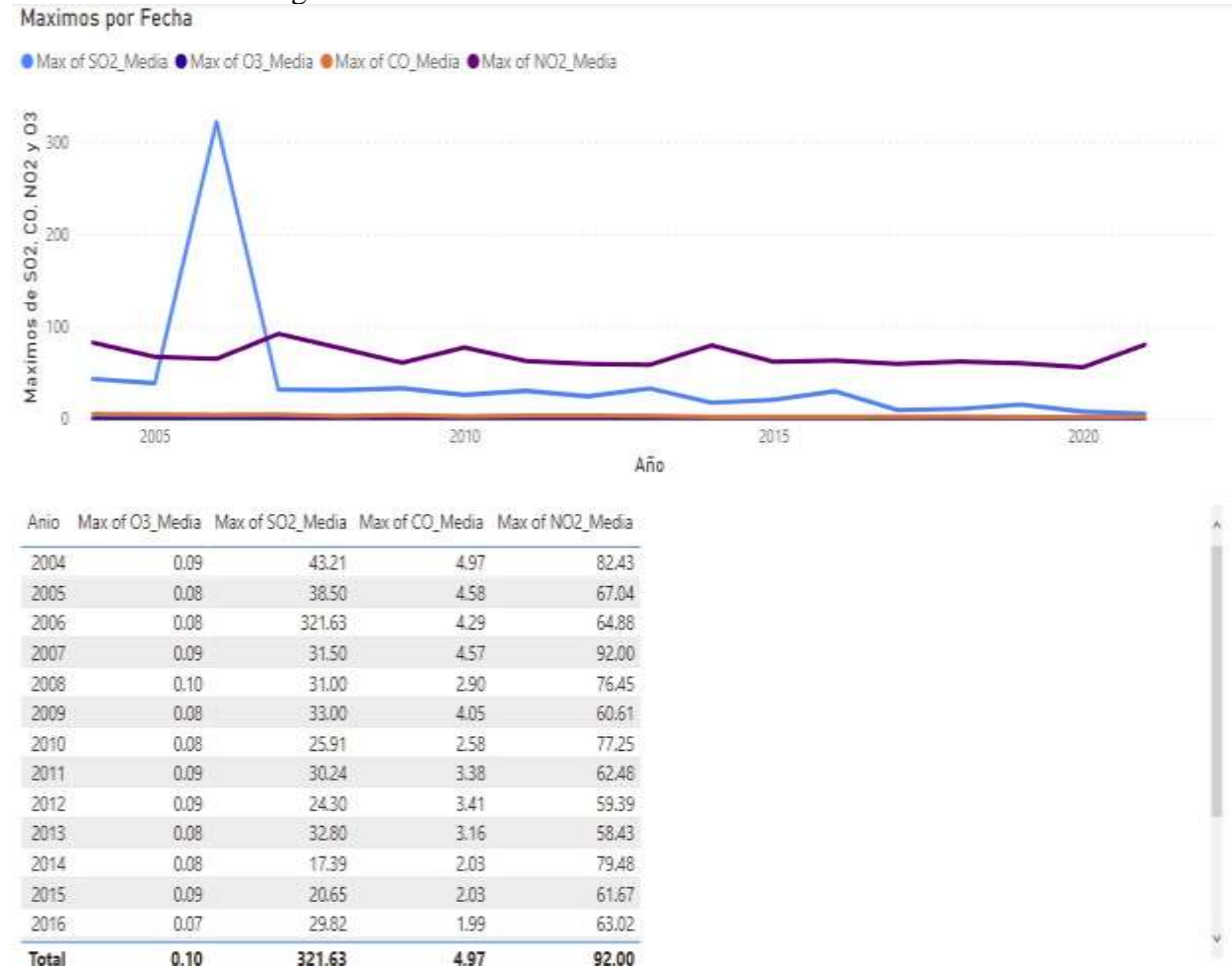
Promedios por Estado



Estado	Sum of O3_Media	Sum of SO2_Media	Sum of CO_Media	Sum of NO2_Media
Alabama	56.46	2,128.58	522.71	18,740.78
Arizona	607.05	21,171.35	8,470.69	336,326.32
Arkansas	314.23	14,304.93	4,617.73	105,722.57
California	5,041.33	151,221.69	63,641.09	1,990,704.50
Colorado	359.83	15,240.49	5,070.12	249,041.44
Connecticut	307.88	7,907.68	2,536.46	97,696.12
Delaware	145.62	3,673.51	1,302.57	55,496.83
District of Columbia	462.84	39,371.11	9,085.23	223,403.13
Florida	241.49	2,734.85	2,659.55	44,816.66
Georgia	16.31	1,661.59	143.89	1,794.50
Illinois	153.79	11,761.32	1,718.14	83,743.71
Indiana	145.45	8,879.53	1,450.08	49,687.36
Iowa	591.84	11,875.76	4,968.15	123,356.47
Total	14,916.77	666,007.67	164,893.32	5,878,936.47

Aquí se puede ver la suma de las medias por estado, viendo así mismo cuales son los más contaminados.

También hicimos el siguiente:



En este podemos ver los valores máximos de contaminación en cuanto a los años.

Como se puede observar, estos fueron algunas tablas y gráficos que se obtuvieron como resultado de herramientas analíticas. Son pruebas básicas con las que se pretendía obtener valores de contaminación por año y por estados, dando en inicio unos gráficos que permitieran visualizar estos factores, y a la vez encaminar las habilidades de los integrantes del grupo para dar paso a análisis más profundos.

11. Sentencias SQL creación estructuras de datos, funciones, procedimientos y cualquier otro objeto creado en la base de datos para el Ambiente Analítico.

Adjuntadas en punto 7. Ver [aquí](#).

12. Elaboración de 9 preguntas de análisis de diferentes naturalezas. Hay que tratar en lo posible que estas preguntas NO sean preguntas triviales sino al contrario preguntas profundas y significativas. La elaboración de estas preguntas debe ser claramente descriptiva de manera que pueda entenderse de forma efectiva lo que quiere decir. Incluir sentencias SQL que responden las preguntas y captura de imagen de los resultados. Las sentencias deben ser lo más eficientes posibles y reducidas a lo necesario, esto último significa que si una pregunta puede responderse con una sola sentencia/instrucción SQL no debería incluirse código adicional con otras instrucciones.

--¿Cuales ciudades presentan un índice de calidad del aire dañino para la salud?

```
SELECT DISTINCT MAX(Fecha.Fecha_completa) as [Fecha mas actual], Centro.Ciudad, (CASE WHEN  
MAX(Hechos.CO_AQI + Hechos.NO2_AQI + Hechos.O3_AQI + Hechos.SO2_AQI) BETWEEN 101 AND 150 THEN 'Dañina a  
la salud para grupos sensibles' WHEN MAX(Hechos.CO_AQI + Hechos.NO2_AQI + Hechos.O3_AQI +  
Hechos.SO2_AQI) BETWEEN 151 AND 200 THEN 'Dañina a la salud' WHEN MAX(Hechos.CO_AQI + Hechos.NO2_AQI +  
Hechos.O3_AQI + Hechos.SO2_AQI) BETWEEN 201 AND 300 THEN 'Muy dañina a la salud' WHEN MAX(Hechos.CO_AQI  
+ Hechos.NO2_AQI + Hechos.O3_AQI + Hechos.SO2_AQI) > 300 THEN 'Peligrosa' ELSE 'No dañina' END) as  
Calidad  
FROM Centro, Hechos, Fecha  
WHERE Centro.DataID = Hechos.DataID AND Fecha.FechaID = Hechos.FechaID GROUP BY Centro.Ciudad ompleta
```

	Fecha mas actual	Ciudad	Calidad
25	2021-06-30	Aroostook	Dañina a la salud
26	2021-06-30	New Haven	Muy dañina a la salud
27	2017-03-31	Hartford	Muy dañina a la salud
28	2021-06-30	Minnehaha	Dañina a la salud para grupos sensibles
29	2021-09-30	Garrett	Dañina a la salud para grupos sensibles
30	2013-01-02	Adair	Dañina a la salud
31	2009-10-31	Montgomery	Muy dañina a la salud
32	2021-09-30	Prince Geor...	Muy dañina a la salud
33	2015-06-30	Fremont	No dañina
34	2021-06-30	Essex	Muy dañina a la salud
35	2021-06-23	Camden	Muy dañina a la salud

--Según la densidad de población, ¿que ciudad presenta mayor AQI?

```
SELECT Centro.Densidad_Ciudad AS Densidad, Centro.Ciudad, MAX(Hechos.CO_AQI + Hechos.NO2_AQI  
+ Hechos.O3_AQI + Hechos.SO2_AQI) as [AQI Acumulado]  
FROM Centro, Hechos  
WHERE Centro.DataID = Hechos.DataID  
GROUP BY Centro.Ciudad, Centro.Densidad_Ciudad
```

order by [AQI Acumulado] desc

	Densidad	Ciudad	AQI Acumulado
1	451.35	Kansas City	425
2	2130.341	Pittsburgh	332
3	1274.74	Essex	331
4	915.997	Charlotte	330
5	1370.268	Rubidoux	319

--Tomando en consideración los últimos 10 años, qué estado ha tenido el mayor incremento en sus niveles de contaminación

```
SELECT TOP 1 Centro.Estado, SUM(Hechos.CO_AQI + Hechos.NO2_AQI + Hechos.O3_AQI +
Hechos.SO2_AQI) as Niveles
FROM Centro, Hechos, Fecha
WHERE Centro.DataID = Hechos.DataID AND Hechos.FechaID = Fecha.FechaID AND
Fecha.Fecha_completa > DATEADD(year,-10,GETDATE())
GROUP BY Centro.Estado ORDER BY Niveles DESC
```

	Estado	Niveles
1	California	6431456

--Qué contaminante específico es el que mayor nivel de contaminación presenta?

```
SELECT DISTINCT (CASE WHEN SUM(Hechos.CO_AQI) > SUM(Hechos.NO2_AQI) AND
SUM(Hechos.CO_AQI) > SUM(Hechos.O3_AQI) AND SUM(Hechos.CO_AQI) >
SUM(Hechos.SO2_AQI) THEN 'CO'
WHEN SUM(Hechos.NO2_AQI) > SUM(Hechos.CO_AQI) AND SUM(Hechos.NO2_AQI) >
SUM(Hechos.O3_AQI) AND SUM(Hechos.NO2_AQI) > SUM(Hechos.SO2_AQI) THEN 'NO2'
WHEN SUM(Hechos.O3_AQI) > SUM(Hechos.CO_AQI) AND SUM(Hechos.O3_AQI) >
SUM(Hechos.NO2_AQI) AND SUM(Hechos.O3_AQI) > SUM(Hechos.SO2_AQI) THEN 'O3'
WHEN SUM(Hechos.SO2_AQI) > SUM(Hechos.CO_AQI) AND SUM(Hechos.SO2_AQI) >
SUM(Hechos.NO2_AQI) AND SUM(Hechos.SO2_AQI) > SUM(Hechos.O3_AQI) THEN 'SO2'
END) as [Contaminante con mayor niveles durante los ultimos años]
FROM Hechos
```

Contaminante con mayor niveles durante los ultimos años	
1	O3

--Comparando con otros días, qué niveles de contaminación se capturan en los días feriados

```
SELECT Hechos.O3_Media, Hechos.O3_Primer_valor_maximo, Hechos.O3_AQI, Hechos.CO_Media,
Hechos.CO_Primer_valor_maximo, Hechos.CO_AQI, Hechos.SO2_Media, Hechos.SO2_Primer_valor_maximo,
Hechos. SO2_AQI, Hechos.NO2_Media, Hechos.NO2_Primer_valor_maximo, Hechos. NO2_AQI,
Fecha.Fecha_completa FROM Fecha, Hechos WHERE Hechos.FechaID = Fecha.FechaID AND Fecha.Dia_Festivo =
'SI'
```

	O3_Media	O3_Primer_valor_maximo	O3_AQI	CO_Media	CO_Primer_valor_maximo	CO_AQI	SO2_Media	SO2_Primer_valor_maximo	SO2_AQI	NO2_Media	NO2_Primer_valor_maximo	NO2_AQI	Fecha_completa
1	0.0408819988369942	0.0529999993741512	49	0.0916669964790344	0.200000002900232	2	0.0833330005407333	1	1	5	12	11	2004-09-04
2	0.0358819998800755	0.0489999997185001	45	0.0125000001862645	0.100000001490116	1	0.33332985639572	1	1	4.54166698455811	16	15	2004-09-06
3	0.017705999314785	0.0219999996807907	20	0	0	0	0.260070009660721	1	1	5.1304348895292	9	8	2004-10-11
4	0.0268820002675056	0.0320000015199184	30	0	0	0	3.23809504508972	9	13	3.68421101570129	5	5	2004-06-19
5	0.0377059998677502	0.0439999997615814	41	0	0	0	1.16666698455811	6	9	0	0	0	2004-09-04
6	0.0360589995800263	0.0469999993906746	44	0	0	0	0.375	3	4	0	0	0	2004-09-05
7	0.036470007896423	0.0430000014603138	40	0	0	0	0	0	0	0	0	0	2004-09-06
8	0.00888200011104345	0.0160000007596682	15	2.02816693687438	3.5	40	16.375	26	37	44.7916679382324	77	76	2004-02-14
9	0.0383529998362064	0.0439999997615814	41	0	0	0	0.25	2	3	4.125	9	8	2004-05-31
10	0.0295880000488894	0.0370000004768372	34	0	0	0	0.375	3	4	2.5	5	5	2004-06-19
11	0.0178239997476339	0.0209999997168779	19	0.5625	0.600000023841858	7	8.125	20	29	24.7083320617676	34	32	2004-04-11
12	0.0164710003882647	0.0309999994933605	29	0.49583300948143	0.600000023841858	7	5.58333301544189	12	17	13.7916669845581	23	22	2004-04-11
13	0.0388819985091686	0.046000000089407	43	0.349999994039536	0.600000023841858	7	8.33333301544189	12	17	13.7083330154419	34	32	2004-06-19
14	0.0407649996445904	0.0469999993906746	44	0.44166699051857	0.5	6	3.29166698455811	5	7	6.66666698455811	13	12	2004-05-31
15	0.0176470000296831	0.0209999997168779	19	0	0	0	0.33332985639572	3	4	5.08333301544189	6	6	2004-11-11
16	0.0231180004775524	0.046000000089407	43	0.712499976158142	0.899999976158142	10	2.75	9	13	25	42	40	2004-09-04
17	0.0301179997622967	0.0439999997615814	41	0.970833003520966	1.70000004768372	19	1.33333301544189	2	3	17.5	40	38	2004-09-05
18	0.0151760000735521	0.0189999993890524	18	0.408333003520966	0.5	6	1.58333301544189	4	6	11.0416669845581	21	20	2004-09-06

13. Elaboración de 3 posibles decisiones significativas basadas en el modelo de datos analítico.

-Determinar que leyes, medidas o incluso sanciones se deben de tomar en consideración para ser aplicadas en los lugares que tienen un nivel de contaminación clasificado como dañino para la salud, y así garantizar el derecho humano al aire limpio.

-Crear organizaciones o entidades –o asegurar la efectividad de las existentes- que velen porque se mantenga la calidad del aire, y que identifiquen y regulen factores que impacten negativamente en la misma, como el costo ambiental de ciertos combustibles, las emisiones de gases de ciertas industrias, etc.

- Aumentar la concientización de la contaminación y sus efectos tanto en la salud de presentes y futuras generaciones, como en el medio ambiente y la capa de ozono, mediante campañas, cursos educativos, y medios de comunicación en las distintas comunidades y la sociedad como tal.

14. Indicar herramientas utilizadas (Lenguajes, DBMS, utilitarios, librerías, APIs, etc.).

Para la elaboración de todo el proyecto, como gestor de base de datos, utilizamos Microsoft SQL Server Management Studio 18, como lenguaje se utilizó SQL, y para la visualización y análisis de esta, utilizamos Power BI versión 2.104.

15. Describir de manera general y particular (atributos, columnas o campos) las fuentes de datos (ejs. Datasets) de forma que el lector comprenda dichas fuentes de datos. Incluir la dirección o lugar de donde se obtuvieron dichas fuentes de datos.

Las tablas enriquecidas gracias al dataset principal y los otros, las cuales pertenecen al diagrama presentado anteriormente, son las siguientes:

-Centro (*DataID, Direcccion, Estado, Condado, Ciudad, Poblacion_estado, Area_estado, Longitud_estado, Latitud EstadoID*): En la misma se recopilan datos concernientes a los centros donde se mide la calidad del aire.

-Fecha (*FechaID, Num_Dia, Num_Mes, Anio, Fecha_completa, Dia, Mes, Dia_semana, Dia_Festivo, Num_Semana*): En la misma se recopilan datos concernientes a la fecha.

-Hora(*HoraID,Momento, Hora, reloj_24_horas, reloj_24_horas, minutos, segundos*):En la misma se recopilan datos concernientes al tiempo.

-Hechos (*Codigo_muestra, DataID, FechaID, O3_Media, O3_Primer_valor_maximo, O3_AQI, CO_Media, CO_Primer_valor_maximo, CO_AQI, SO2_Media, SO2_Primer_valor_maximo, SO2_AQI, NO2_Media, NO2_Primer_valor_maximo, NO2_AQI*): En la misma se recopilan datos concernientes a los contaminantes.

El dataset principal que sirvió de inspiración para este proyecto fue tomado de Kaggle, y presenta “Datos de contaminación por CO, NO2, O3 y SO2 en los EE. UU. entre 2000 y 2021 de la EPA” Es decir, de la Agencia de Protección Ambiental de los Estados Unidos (US EPA). Este conjunto de datos tiene alrededor de 69 mil entradas y 24 columnas, y fue editado por última vez en febrero de 2022, por Angela Kim. Se puede acceder mediante el siguiente enlace, o dando click [aquí](https://www.kaggle.com/datasets/alpacanonymous/us-pollution-20002021).

<https://www.kaggle.com/datasets/alpacanonymous/us-pollution-20002021>

Otros datasets empleados para proporcionar más datos y un mejor estudio y análisis de la

situación tomando en cuenta el objetivo, fueron:

-DatasetCensus (*sumlev, región, división, stateid, county, stname, ctynome, estimatebase2000, popestimate2000, popestimate2001, popestimate2002, popestimate2003, popestimate2004, popestimate2005, popestimate2006, popestimate2007, popestimate2008, popestimate2009*):
Especifica por estado la población total, a través de los años.

Recopilado de la Oficina de Censo de Estados Unidos (United States Census Bureau). Edición del 8 de Octubre del 2021. Se puede acceder mediante el siguiente enlace, o dando click [aquí](#).

<https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-state.html>

-DatasetCiudad (*rank, city, state, _2020_census, _2010_census, percent_change, land_area_mi, land_area_km, pop_density_mi, pop_density_km, degrees_north, degrees_west*):
En la misma se recopilan todas las ciudades y otros datos de los estados de Estados Unidos.

Contiene datos sobre las 330 ciudades más pobladas de los Estados Unidos, incluida la información del censo de 2020, la población del censo de 2010, la superficie terrestre, la densidad de población y la ubicación. Fue editado por última vez en diciembre 2021, por Axel Torbenson. Se puede acceder mediante el siguiente enlace, o dando click [aquí](#).

<https://www.kaggle.com/datasets/axeltorbenson/us-cities-by-population-top-330>

-DatasetFestivos(*date,holiday, weekday, month, day, year*):En la misma se recopilan datos concernientes a los días no laborables o festivos de los Estados Unidos.

Este conjunto de datos incluye 18 años de feriados en EE. UU. con fecha entre el 1 de enero de 2004 y el 31 de diciembre de 2021. Cada registro tiene fecha, feriado, día de la semana, mes, día y año. Fue editado por última vez en 2021, por Donnetew Oluwari. Se puede acceder mediante el siguiente enlace, o dando click [aquí](#).

<https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021>

-DatasetGeo(*Area Total Mi, Area Total Km, Area terrestre Mi, Area terrestre Km, Area*

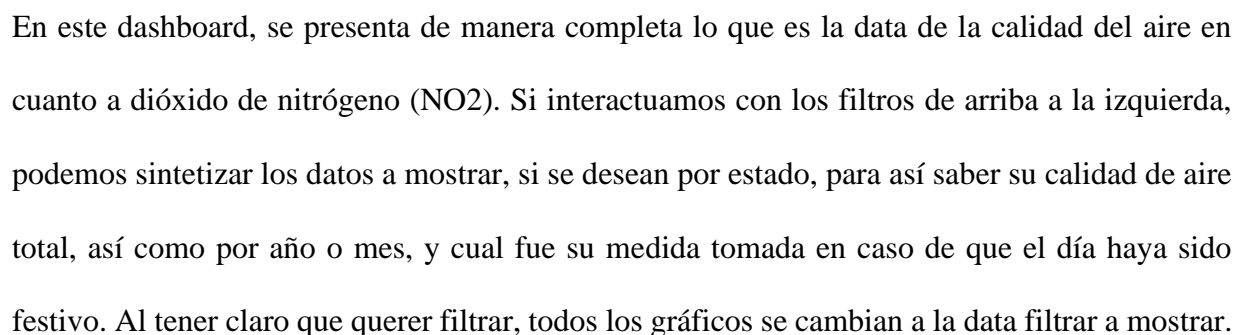
acuatica Mi, Area acuatica Km, Interior Mi, Interior Km, Costero Mi, Costero Km, Lagos Mi, Lagos Km, Territorial Mi, Territorial Km, Latitud, Longitud): En la misma se recopilan datos concernientes a la geografía —el área en cada una de sus formas— de los estados de Estado Unidos, tanto en millas como en kilómetros.

Proporciona las medidas de tierra, agua y área total para los 50 estados, el Distrito de Columbia, Puerto Rico y las Áreas Insulares. Las medidas del área se derivaron de la base de datos del Archivo maestro de direcciones/Codificación y referencia geográfica integrada topológicamente (MAF/TIGER®) de la Oficina del Censo. Los límites de los estados y áreas equivalentes son a partir del 1 de enero de 2010. Las áreas terrestres y acuáticas, incluidas sus clasificaciones, reflejan las actualizaciones de características básicas realizadas en la base de datos MAF/TIGER® hasta agosto de 2010.

Recopilado de la Oficina de Censo de Estados Unidos (United States Census Bureau). Última revisión del 16 de diciembre del 2021. Se puede acceder mediante el siguiente enlace, o dando click [aquí](#).

<https://www.census.gov/geographies/reference-files/2010/geo/state-area.html>

Dashboard de NO2

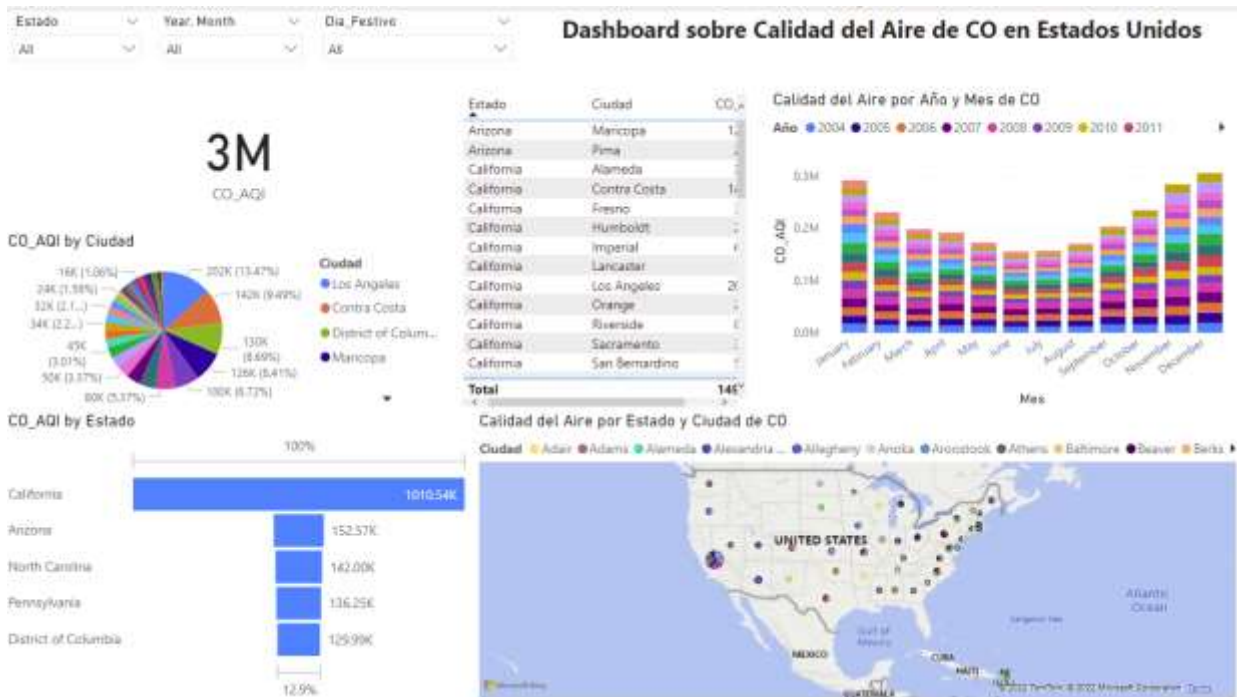


Segunda Ventana



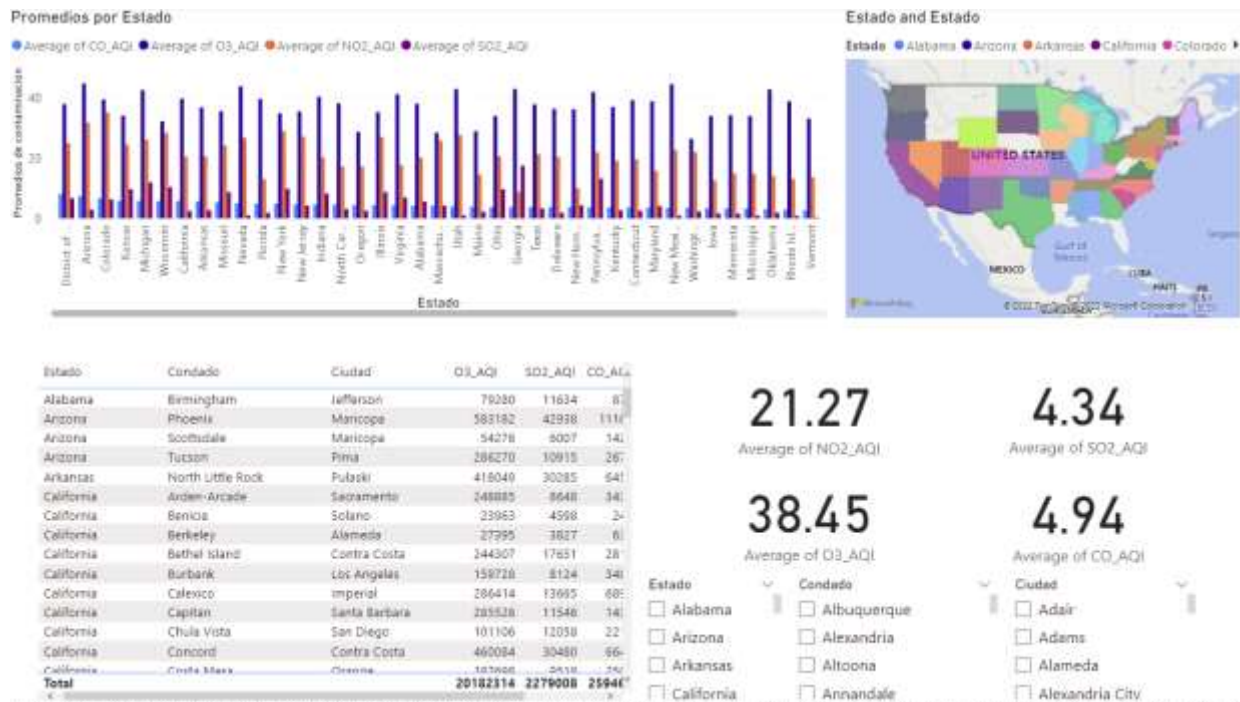
Continuando con el contaminante NO₂, aquí decidimos sacar el promedio de cada estado. A diferencia del anterior, aquí no tomamos en cuenta filtrar por fecha, sino más bien desde estado hasta ciudad, de tal forma que se pueda ver que ciudad en ese estado tiene un promedio mayor. En este caso, el estado de colorado posee un 23.15% de un top 5 de los estados.

Dashboard de CO



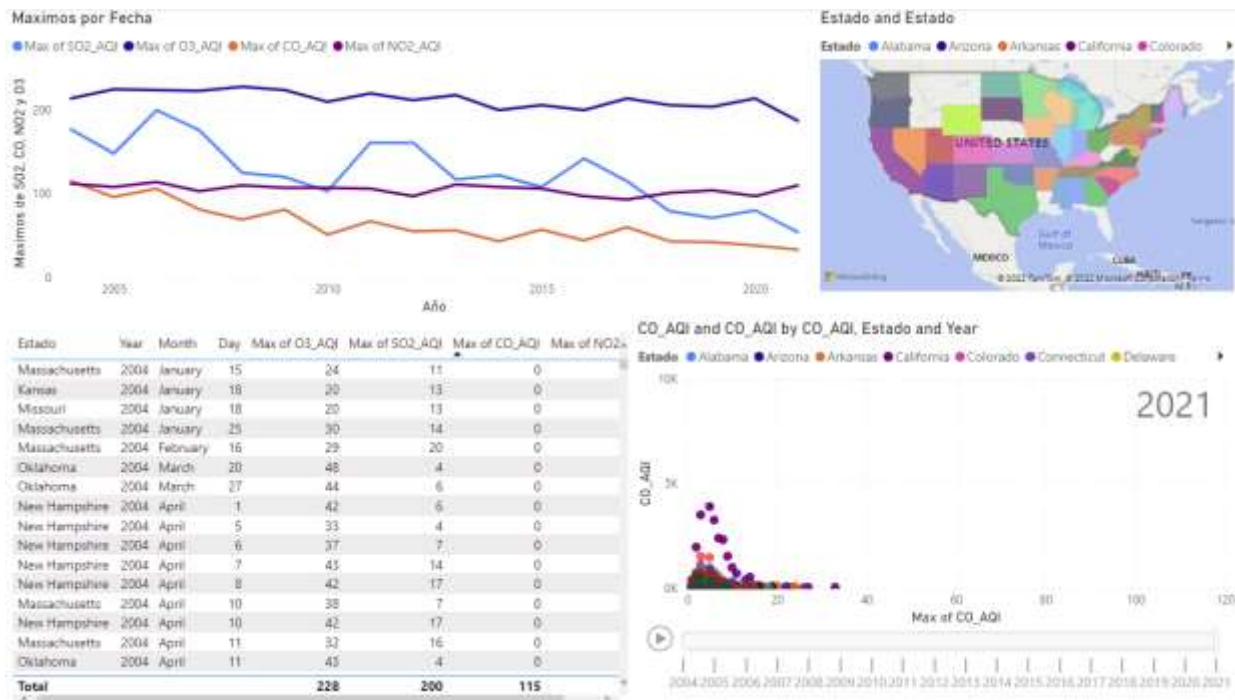
Esta grafica parecida al del contaminante NO₂, pero en este caso se habla del contaminante monóxido de carbono (CO), en la cual interactuamos con los filtros y de ahí se puede pasar al grafico en barras para ampliar la data del mes, en caso de que se haya indicado en el filtro.

Cuarta Ventana



Ya en esta pantalla se ve de manera general los promedios de todas las calidades del aire de cada contaminante, en el cual se puede filtrar desde estado a ciudad. Como se puede ver en la tarjeta, el contaminante Ozono (O3), posee un promedio más alto que los demás contaminantes, que, aunque este dentro del margen aceptable, sigue siendo el mayor entre todos.

Quinta Ventana



En esta última ventana, podemos ver los máximos de los contaminantes, desglosados por año, mes y día. En el mapa se puede especificar qué estado se quiere ver, ya que también más abajo se puede visualizar una línea de tiempo basada en los años en cuanto a cómo fue incrementando o decrementando el máximo de cada contaminante.

17. Reflexión general de la experiencia. La reflexión debe expresar y comunicar un proceso crítico-analítico con ciertos niveles de profundidad, o sea que no debe ser algo trivial o superficial, y que exponga los aspectos de mayor relevancia del trabajo como por ejemplo las principales actividades llevadas a cabo, las principales decisiones de diseño tomadas, los mayores retos o inconvenientes encontrados, entre otros.

Finalmente, luego de reflexionar sobre los procesos realizados a fin de obtener este proyecto, se evidencia que a través del uso de los datos se puede llegar a la respuesta de interrogantes importantes para la sociedad, conociendo entonces si es posible para Estados Unidos cumplir el objetivo planteado en el acuerdo París para la fecha indicada, el 2025. La evaluación de la respuesta a la interrogante planteada se implementó aplicando de forma práctica los elementos teóricos tratados a lo largo de la asignatura, iniciando con la investigación para desglosar la problemática elegida por su relevancia ambiental, e interés social. Se utilizaron distintos conjuntos de datos para enriquecer los análisis y obtener mayor asertividad en las conclusiones resultantes, aplicando reglas de integridad (como la regla de integridad referencial, donde la base de datos no debe contener valores de clave ajena sin concordancia, entre otras) pues la cantidad sin calidad no es un camino viable si se quieren resultados certeros. Continuando con el ETL, nos garantizó la utilidad de los datos para el fácil acceso a la información, y permitiendo que la base de datos sea eficiente y efectiva para la realización de análisis. Se pudo comprobar el impacto de los datos en un estudio, y cómo gracias al correcto manejo de estos se pueden tomar decisiones significativas que ayuden a aliviar la problemática, e identificar factores prudentes para llevar esto a cabo. Gracias a las herramientas analíticas, estos datos pueden ser desmenuzados, y comprendidos bajo distintos puntos de comparación, lo que asegura que la información llegue a ser entendida para el público general.

En cuanto a complicaciones, entre las más difíciles de resolver, fue el producto cartesiano que se hacía entre las tablas al popular la tabla de hechos. Este lo resolvimos revisando el DatasetPrincipal y comparando este con el dataset de los estados, y notamos que había ciudades que no estaban en el de ciudades que, si estaban en el principal, por lo que agregamos estas ciudades y paulatinamente se fue reduciendo el número de filas hasta el número que esperábamos.