

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 Ureña (2018-1669)

Junior Hernández (2018-0999)

Luis Angel Ramírez (2018-0570)

Presentado a:

Maximo Emiliano Perez Medrano

Asignatura:

Base de datos II

**SANTIAGO DE LOS CABALLEROS
REPÚBLICA DOMINICANA**

1. Descripción del Objetivo General

Las emisiones de contaminantes en EE.UU. han crecido exponencialmente a lo largo del tiempo. De este modo, el segundo país más contaminante del mundo se aleja drásticamente de los objetivos del acuerdo de París según los cuales debería reducir drásticamente la emisión de gases contaminantes antes del 2025. 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 su objetivo pautado para 2025. Mediante la estructuración y creación de un modelo de dimensiones planteando el análisis de mayores casos de contaminación y cálculos concerniente a dicho análisis.

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

3. Descripción del Proceso a Modelar

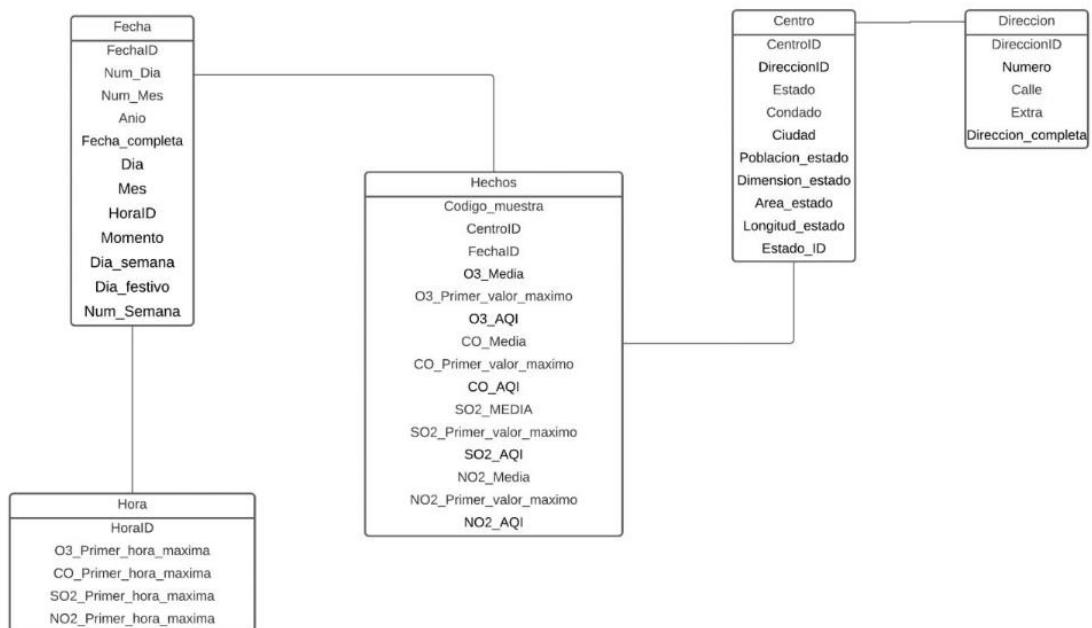
En la actualidad la sociedad es en una considerable medida ignorante del aumento peligroso de los contaminantes procedentes de EE.UU y las personas consciente que saben de susodicha problemática se preocupan que el país no pueda cumplir lo pautado para 2025. Esta ignorancia y desconocimiento del procedimiento 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.

4. Descripción del Grano del modelo

Análisis de cálculo de contaminantes por ciudad

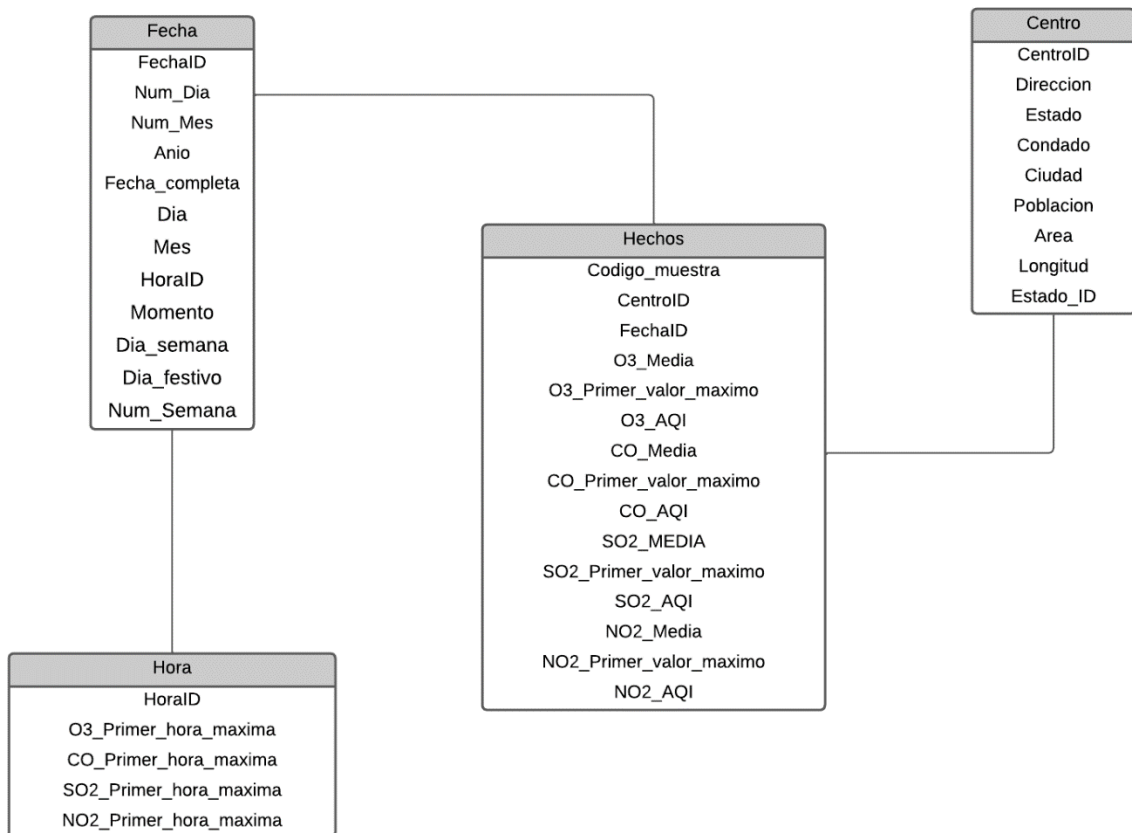
El grano de este modelo usado para el análisis del aumento o decremento de los niveles de contaminación por ciudades tiene un alto detalle por el motivo de que hay diversos factores que contribuyen al aumento o decremento de este (tales el tamaño del lugar y su número de población). Se busca implementar los conceptos teóricos sobre la granularidad de datos para que los mismos no se viesen 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

Air pollution Data Set
Diagram



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

```

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 rellenamos 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.

En cuanto a complicaciones, entre la más que nos dio trabajo 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.

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?

```
/*
create function dbo.InlineMax(@val1 int, @val2 int)

returns int

as

begin

    if @val1 > @val2

        return @val1

    return isnull(@val2,@val1)

end

SELECT Fecha.Año as Año, dbo.InlineMax( Hechos.O3_Media, dbo.InlineMax( Hechos.Co_Media,
dbo.InlineMax (Hechos.SO2_Media ,Hechos.NO2_Media)) ) as Valor

FROM Fecha, Hechos

Where Fecha.FechaID = Hechos.FechaID and fecha.año = 2006
```

	Año	Valor
1	2006	5
2	2006	10
3	2006	10
4	2006	8
5	2006	6
6	2006	5
7	2006	6
8	2006	5
9	2006	9
10	2006	15
11	2006	19
12	2006	14
13	2006	12
14	2006	16
15	2006	14
16	2006	3
17	2006	14
18	2006	20
19	2006	21

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

```
SELECT TOP 5 Centro.Estado, Round(SUM(Hechos.O3_Media + Hechos.CO_Media + Hechos.SO2_Media + NO2_Media),2) as total FROM Centro,Hechos WHERE Hechos.DataID = Centro.DataID GROUP BY Centro.Estado ORDER BY total desc
```

	Estado	total
1	California	2184975.77
2	Pennsylvania	552637.14
3	District of Columbia	542440.63
4	Arizona	366460.25
5	New York	355490.57

--Qué momento del día (mañana, tarde, noche) y fecha presenta mayores niveles de contaminación O3 por ejemplo?

```
SELECT distinct Fecha.FechaID, Fecha.Momento, Fecha.Fecha_completa ,MAX(Hechos.NO2_Media + Hechos.O3_Media + Hechos.SO2_Media+CO_Media) FROM (SELECT Hechos.FechaID, SUM(Hechos.NO2_Media + Hechos.O3_Media + Hechos.SO2_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.Momento, Fecha.Fecha_completa
```

	FechaID	Momento	Fecha_completa	Total_Contaminacion
1	93764	MAÑANA	2004-01-02	29.6535490080714
2	93767	MAÑANA	2004-01-05	51.6225480786525
3	93772	MAÑANA	2004-01-10	27.4647650178522
4	93775	MAÑANA	2004-01-13	41.7962689837441
5	93780	MAÑANA	2004-01-18	37.2311970097944
6	93783	MAÑANA	2004-01-21	23.3598518427461
7	93788	MAÑANA	2004-01-26	42.0126449819654
8	93791	MAÑANA	2004-01-29	41.7005470534787
9	93799	MAÑANA	2004-02-06	43.1313059469685
10	93802	MAÑANA	2004-02-09	42.685952055268
11	93807	MAÑANA	2004-02-14	39.3026170702651
12	93810	MAÑANA	2004-02-17	49.4411089261994
13	93815	MAÑANA	2004-02-22	23.6359710115939
14	93818	MAÑANA	2004-02-25	32.7699959725142
15	93823	MAÑANA	2004-03-01	35.8981721159071
16	93826	MAÑANA	2004-03-04	25.5627359189093
17	93829	MAÑANA	2004-03-07	31.5436570160091
18	93833	TARDE	2004-03-11	37.9924691226333
19	93834	MAÑANA	2004-03-12	19.1848329864442

--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	0.200000002980232	2	0.0833330005407333	1	1	5	12	11	2004-09-04
2	0.0358819998800735	0.0489999987185001	45	0.0125000001862645	0.100000001490116	1	0.333332985639572	1	1	4.54166698455811	16	15	2004-09-06
3	0.01770599314785	0.0219999998807907	20	0	0	0	0.260870003660721	1	1	5.1304349899292	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	0.0377059988677502	0.0439999997615814	41	0	0	0	1.16666698455811	6	9	0	0	0	2004-09-04
6	0.0360589995980263	0.0469999993906746	44	0	0	0	0.375	3	4	0	0	0	2004-09-05
7	0.0336470007896423	0.0430000014603138	40	0	0	0	0	0	0	0	0	0	2004-09-06
8	0.00888200011104345	0.0160000007599592	15	2.02916693687439	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	0.0295880008488894	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	0.0407649986445904	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	0.333332985639572	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

--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.SO2_Media+CO_Media) FROM (SELECT Hechos.FechaID, SUM(Hechos.NO2_Media + Hechos.O3_Media +
Hechos.SO2_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
```

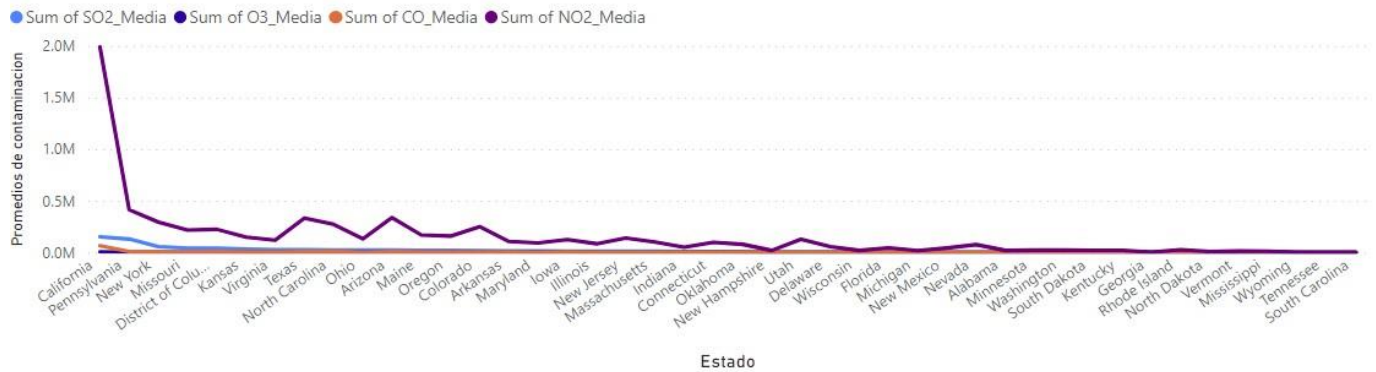
/*

	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

10. Integración inicial a una herramienta analítica

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

Promedios por Estado



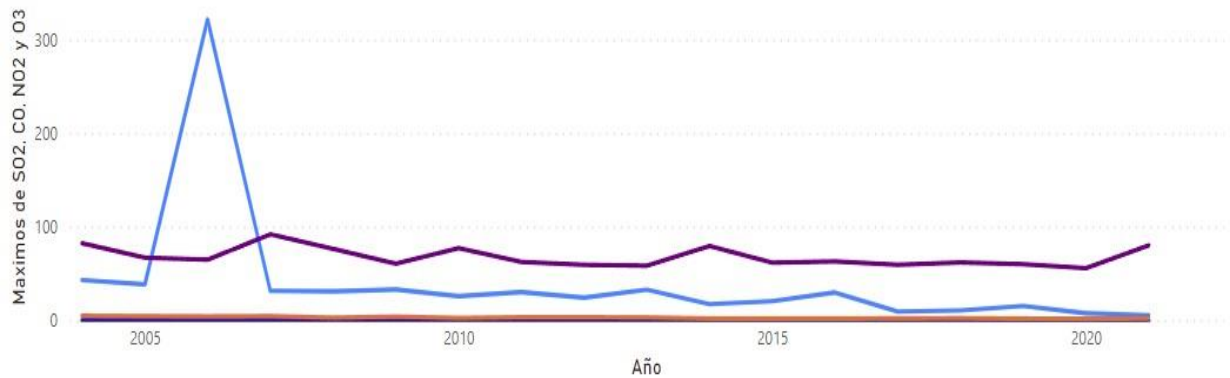
Estado	Sum of O3_Media	Sum of SO2_Media	Sum of CO_Media	Sum of NO2_Media
Alabama	56.46	2,128.58	522.71	18,740.78
Arizona	607.05	21,171.35	8,470.69	336,326.32
Arkansas	314.23	14,304.93	4,617.73	105,722.57
California	5,041.33	151,221.69	63,641.09	1,990,704.50
Colorado	359.83	15,240.49	5,070.12	249,041.44
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	2,859.55	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:

Maximos por Fecha

Max of SO2_Media Max of O3_Media Max of CO_Media Max of NO2_Media



Anio	Max of O3_Media	Max of SO2_Media	Max of CO_Media	Max of NO2_Media
2004	0.09	43.21	4.97	82.43
2005	0.08	38.50	4.58	67.04
2006	0.08	321.63	4.29	64.88
2007	0.09	31.50	4.57	92.00
2008	0.10	31.00	2.90	76.45
2009	0.08	33.00	4.05	60.61
2010	0.08	25.91	2.58	77.25
2011	0.09	30.24	3.38	62.48
2012	0.09	24.30	3.41	59.39
2013	0.08	32.80	3.16	58.43
2014	0.08	17.39	2.03	79.48
2015	0.09	20.65	2.03	61.67
2016	0.07	29.82	1.99	63.02
Total	0.10	321.63	4.97	92.00

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.