



Instituto Politécnico Nacional

Escuela Superior de Cómputo

Grupo 3CV19 Data Mining

Profesor Zagal Flores Roberto Eswart

2do Parcial Practica #5: Exploración de Cubos de datos de precipitación pluvial en CDMX

Alumno García González Aarón Antonio

Lunes 03 de mayo de 2021



Índice

Objetivo3

Introducción3

Desarrollo.....4

Conclusiones.....15

Objetivo

Desarrollar cubos de datos principales para construir el data warehouse para la fuente de datos de Precipitación pluvial (PP) con la técnica de recolección para depósito húmedo (H)", durante el periodo "2010 al 2019".

Introducción

A manera de la resolución de esta practica:

- ⇒ Obtener los catálogos y tabla de hechos realizados en la practica anterior
- ⇒ Crear una vista con la información necesaria, omitiendo datos que no serán requeridos para la obtención de los cubos de datos.
- ⇒ Crear cada cubo de datos
- ⇒ Exportar todos los cubos de datos a formato .CSV
- ⇒ Analizar cada cubo de datos de manera gráfica en Tableau

Desarrollo

Procedimiento: Construya cubos de datos para responder a preguntas de minería, basada en la tabla de hechos obtenida en el desarrollo de la herramienta ETL para precipitación pluvial a fin de comenzar con el desarrollo del data warehouse para dicha fuente.

1. Usando los catálogos y la tabla
 - a) Catálogo de estaciones de monitoreo
 - b) Catálogo de elementos de medición (en este caso que solo contenga al elemento de medición en estudio junto con toda su información (e.g. unidad de medición)
 - c) Prepare una table de hechos que contenga la siguiente estructura mínima:
{Elemento}{E},{Delegación}{D},{noSemana}{S},{Anio}{A},{mes}{M},{medición}{P}
- NOTA: LOS VALORES EN PARÉNTESIS SON ABREVIACIONES

```
-- CREACION DE ESTRUCTURA BASICA DE TABLA DE HECHOS
```

```
-- se descarto el ID y el año completo
```

```
SELECT
```

```
    [elemento]  
    , [anio]  
    , [month]  
    , [week]  
    , [localizacion]  
    , [medicion]
```

```
INTO [precipitacion_pluvial].[dbo].[pph_min]
```

```
FROM [precipitacion_pluvial].[dbo].[pph]
```

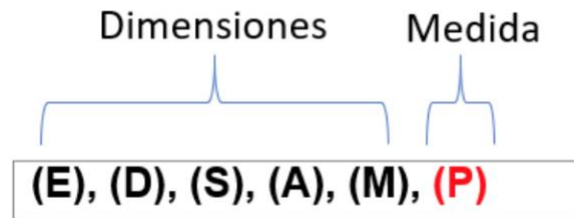
```
-- CONSULTA A LA TABLA ANTERIOR CREADA
```

```
SELECT TOP (1000) [elemento]
```

```
    , [anio]  
    , [month]  
    , [week]  
    , [localizacion]  
    , [medicion]
```

```
FROM [precipitacion_pluvial].[dbo].[pph_min]
```

2. A partir de la tabla de hechos anterior desarrolle cubos que se solicitan. Use solo las dimensiones y la medida estadística de promedio sobre la medición de la precipitación pluvial
3. Grafique los datos y escriba una interpretación para cada uno de los cubos
 - a) Desde los cubos A a la E, utilice la medida de valor máximo y de una interpretación

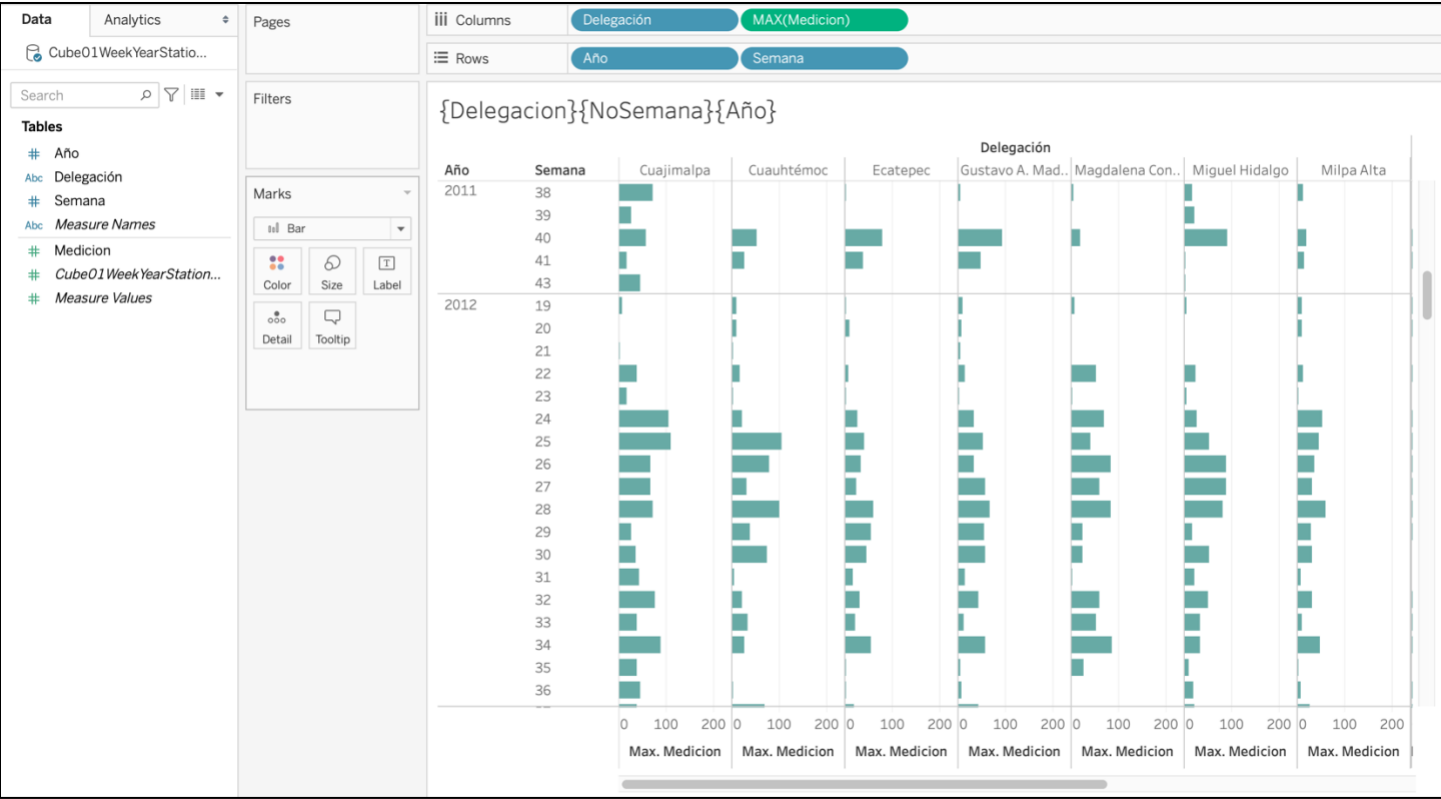


Crearemos una vista:

```
-- Vista Week Year Station Measure
SELECT dbo.estaciones.Delegación_o_Municipio AS Delegación,
       dbo.pph_min.week AS semana,
       dbo.pph_min.anio AS año,
       dbo.pph_min.month AS mes,
       dbo.pph_min.medicion
FROM dbo.pph_min INNER JOIN dbo.estaciones
ON dbo.pph_min.localizacion = dbo.estaciones.id
WHERE (dbo.pph_min.medicion <> - 99)
```

a) {Delegación},{noSemana},{Año}

```
-- {DELEGACION}{NOSEMANA}{ANIO}
SELECT
    [precipitacion_pluvial].[dbo].[estaciones].[Delegación_o_Municipio] AS
    "Delegación",
    [precipitacion_pluvial].[dbo].[pph_min].[week] AS "semana",
    [precipitacion_pluvial].[dbo].[pph_min].[anio] AS "año",
    [precipitacion_pluvial].[dbo].[pph_min].[medicion]
INTO [precipitacion_pluvial].[dbo].[Cube01WeekYearStationMeasure]
FROM [precipitacion_pluvial].[dbo].[pph_min] INNER JOIN
[precipitacion_pluvial].[dbo].[estaciones]
ON [precipitacion_pluvial].[dbo].[pph_min].[localizacion] =
[precipitacion_pluvial].[dbo].[estaciones].[id]
WHERE [precipitacion_pluvial].[dbo].[pph_min].[medicion] != -99
```

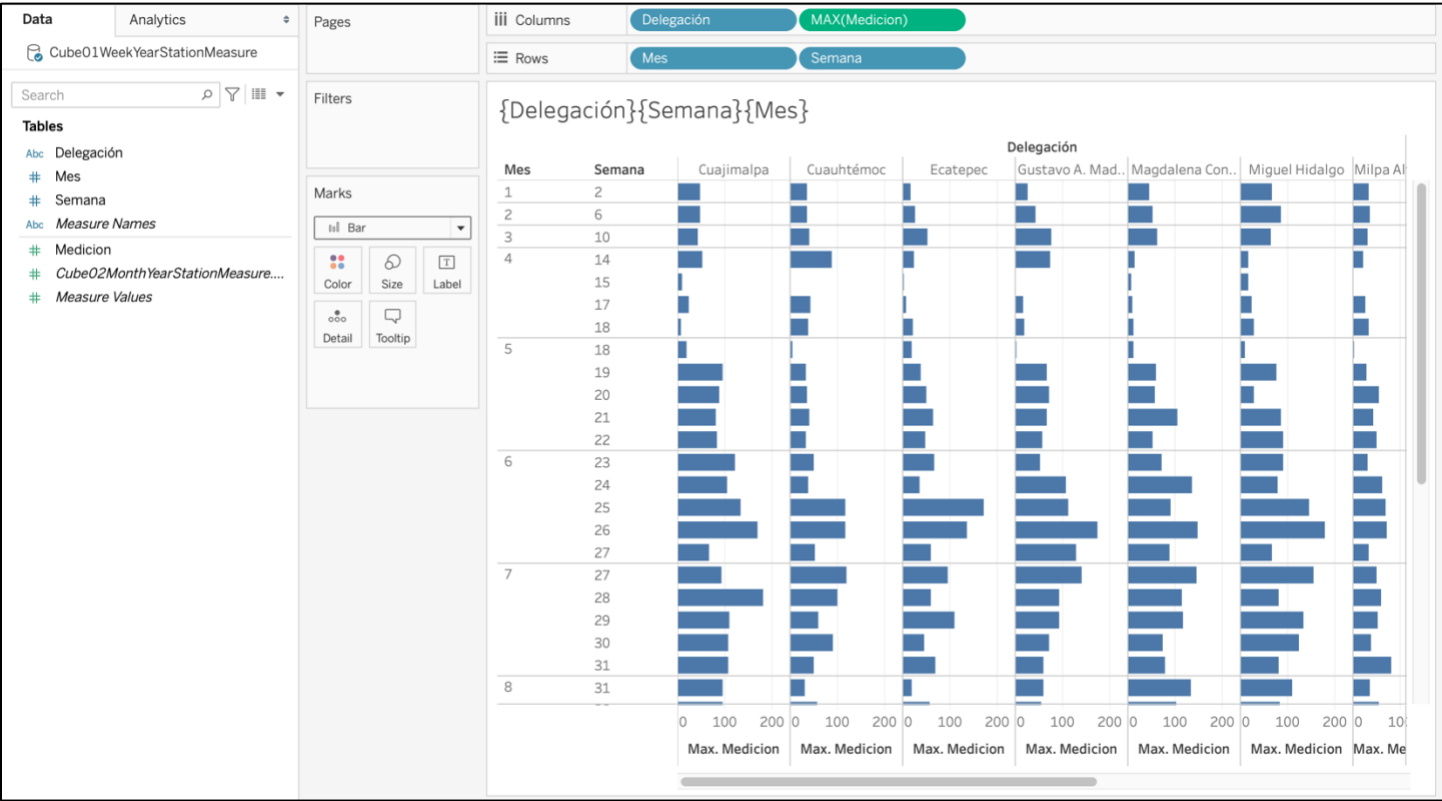


Para el 2010, entre la semana 26 a 32, es en donde más precipitación pluvial se registra, esto es muy acorde a la temporada de lluvias de la zona.

Para el 2011, entre la semana 25 a 35, es en donde más precipitación pluvial se registra, para el 2012 entre la semana 24 a 34, para el caso del 2013, hubo cambios en la precipitación pluvial, ya que se concentran los registros entre la semana 23 a 38, dando un rango de 16 semanas de captación, mientras que en las anteriores fue entre 8 y 10 semanas, esta tendencia de aumentar el los máximos en varias semanas y no concentrarlo en unas pocas como en los primeros años de muestra, continua hasta el 2019 y además pareciera que los máximos se van recorriendo poco a poco en número de semanas, es decir, pareciera que los máximos o lluvias más fuertes tardan más tiempo en llegar ahora.

b) {Delegación},{noSemana},{Mes}

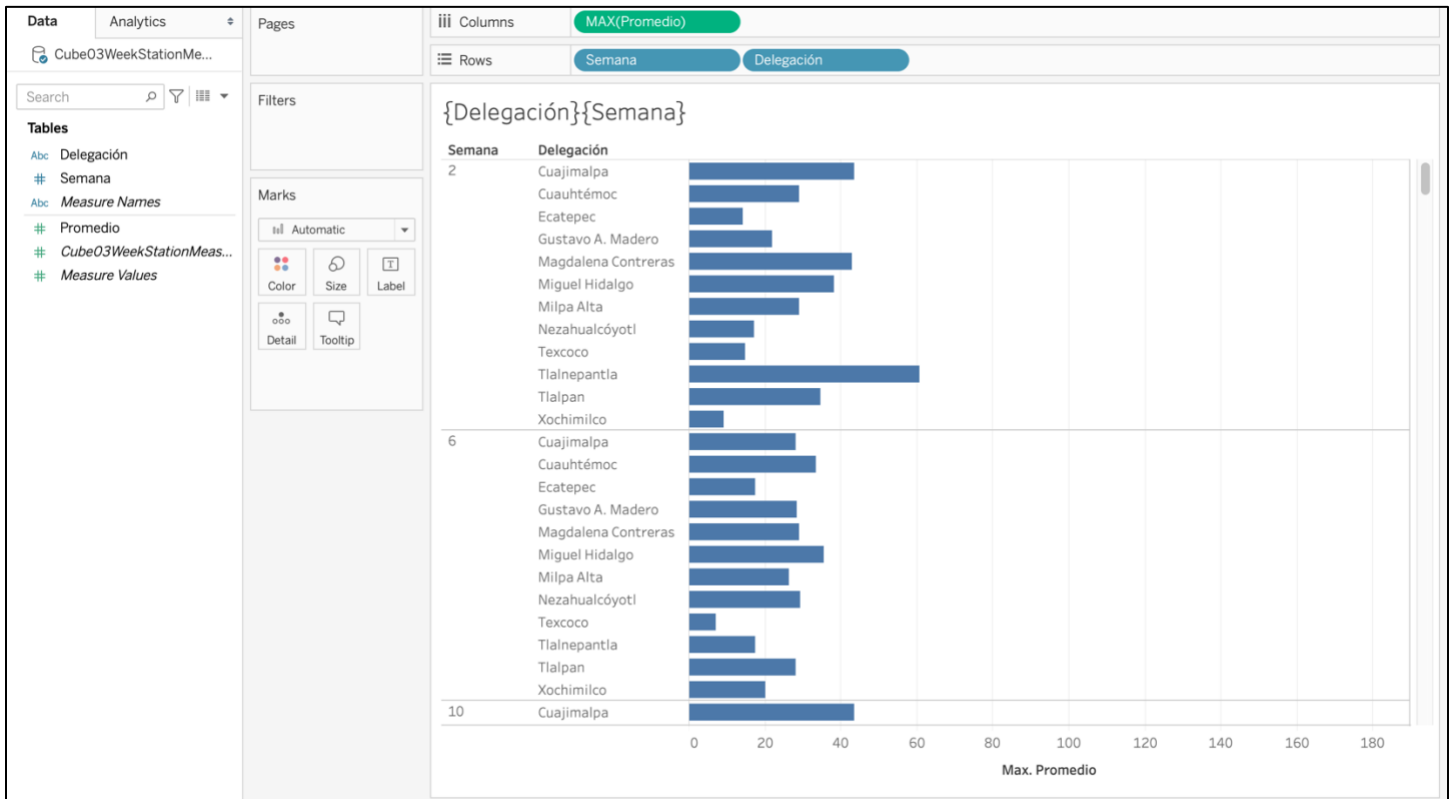
```
-- {DELEGACION}{NOSEMANA}{MONTH}
SELECT
    [precipitacion_pluvial].[dbo].[estaciones].[Delegación_o_Municipio] AS
    "Delegación",
    [precipitacion_pluvial].[dbo].[pph_min].[week] AS "semana",
    [precipitacion_pluvial].[dbo].[pph_min].[anio] AS "año",
    [precipitacion_pluvial].[dbo].[pph_min].[medicion]
INTO [precipitacion_pluvial].[dbo].[Cube02MonthYearStationMeasure]
FROM [precipitacion_pluvial].[dbo].[pph_min] INNER JOIN
[precipitacion_pluvial].[dbo].[estaciones]
ON [precipitacion_pluvial].[dbo].[pph_min].[localizacion] =
[precipitacion_pluvial].[dbo].[estaciones].[id]
WHERE [precipitacion_pluvial].[dbo].[pph_min].[medicion] != -99
```



Este gráfico al estar orientado por meses y semanas, es decir más detallado que año con semanas, se puede apreciar que suelen ocurrir algunas lluvias al inicio de año, luego deja de llover y comienza el periodo de lluvias en la segunda semana del mes de mayo o la semana 19 del año y de ahí se mantiene constante hasta el mes de octubre, los últimos meses del año con algunas lluvias superficiales en los máximos que se muestran, a este nivel ya es posible detectar los lugares en donde hay mas y menos puntos máximos de precipitación, por ejemplo: Cuajimalpa, Gustavo A. Madero, Magdalena Contreras y Tlapan son los que más puntos máximos registran de manera constante. No obstante, lo anterior parece lógico y creíble de acuerdo al clima de la región que hemos observado y notado como ciudadanos.

c) {Delegación},{noSemana}

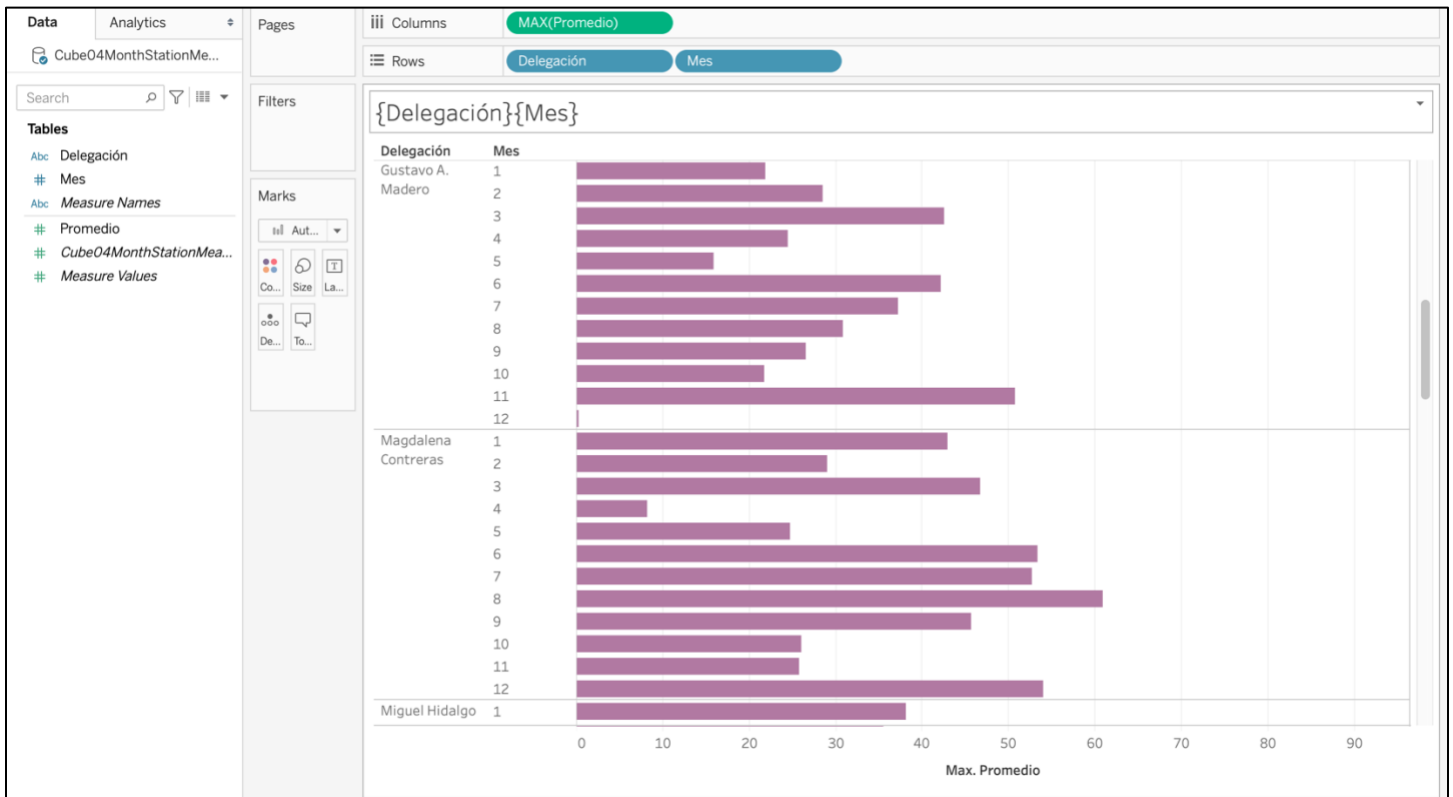
```
-- {Delegacion}{NoSemana}
SELECT [Delegación]
      , [semana]
      ,AVG([medicion]) AS "PROMEDIO"
INTO [precipitacion_pluvial].[dbo].[Cube03WeekStationMeasure]
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]
GROUP BY [Delegación], [semana]
ORDER BY "PROMEDIO" DESC
```



En este gráfico podemos fácilmente destacar por semana cual o cuales delegaciones fueron las que presentaron mayores máximos en precipitación pluvial, por ejemplo, en la semana #2 Tlalnepantla fue el pico más alto, en la #14 fue la alcaldía Cuauhtémoc, en la #27 fue la alcaldía Magdalena Contreras, y en varias semanas, Cuajimalpa mantiene picos máximos mayores respecto a las demás estaciones.

d) {Delegación},{Mes}

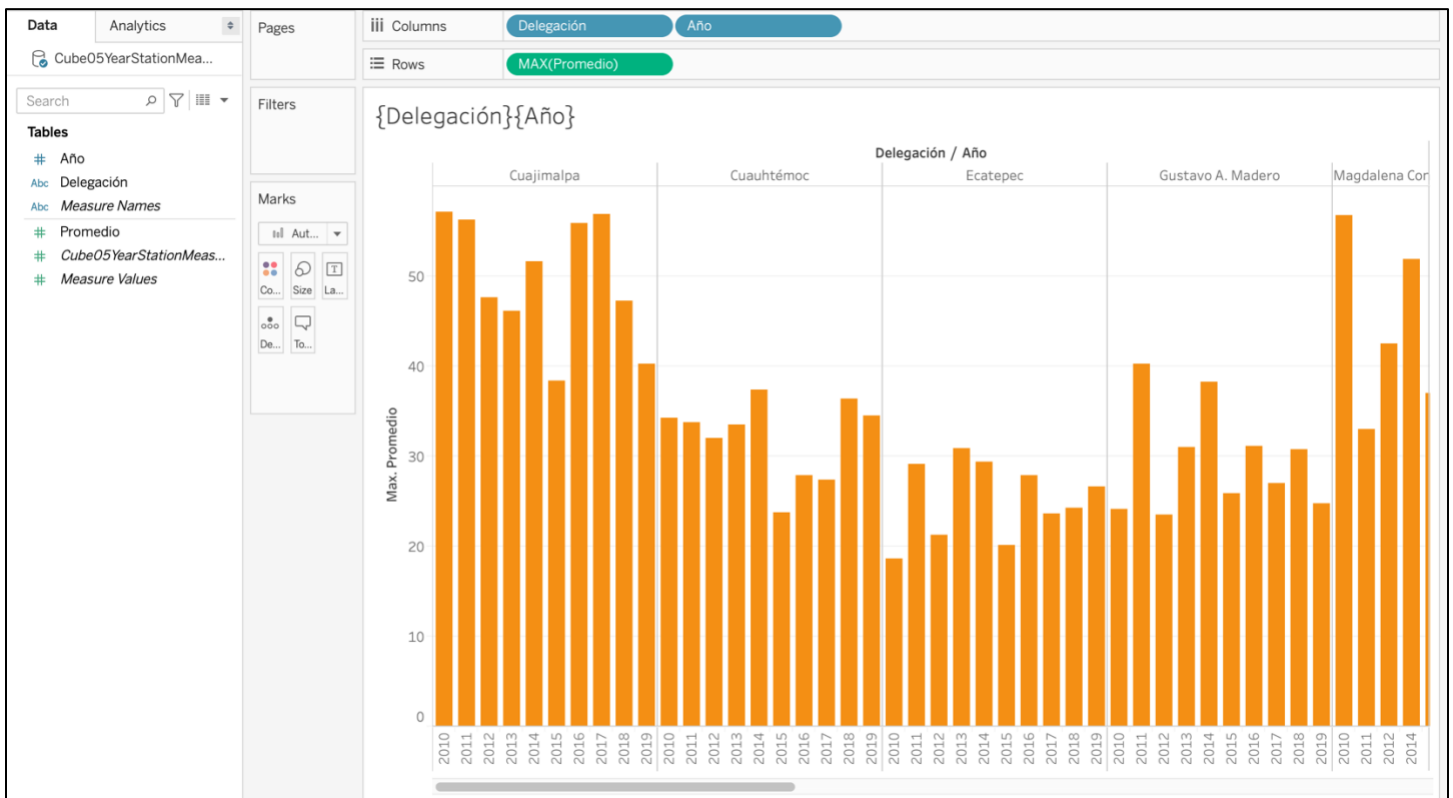
```
-- {Delegacion}{Mes}
SELECT [Delegación]
      , [month]
      , AVG([medicion]) AS "PROMEDIO"
INTO [precipitacion_pluvial].[dbo].[Cube04MonthStationMeasure]
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]
GROUP BY [Delegación], [month]
ORDER BY [Delegación], [month], "PROMEDIO" DESC
```



Para el caso de este gráfico, es muy similar al anterior, pero ahora contrastando estaciones y meses por medición en el máximo registrado de cada estación, se puede apreciar el pico más alto por mes de cada delegación en el periodo del 2010 al 2019, por ejemplo, para Cuajimalpa, Cuauhtémoc, Miguel Hidalgo y Tlalpan, el punto más alto se encuentra en el mes de noviembre, mientras que para Magdalena Contreras y Tlalpan en Agosto, etc.

e) {Delegación},{Año}

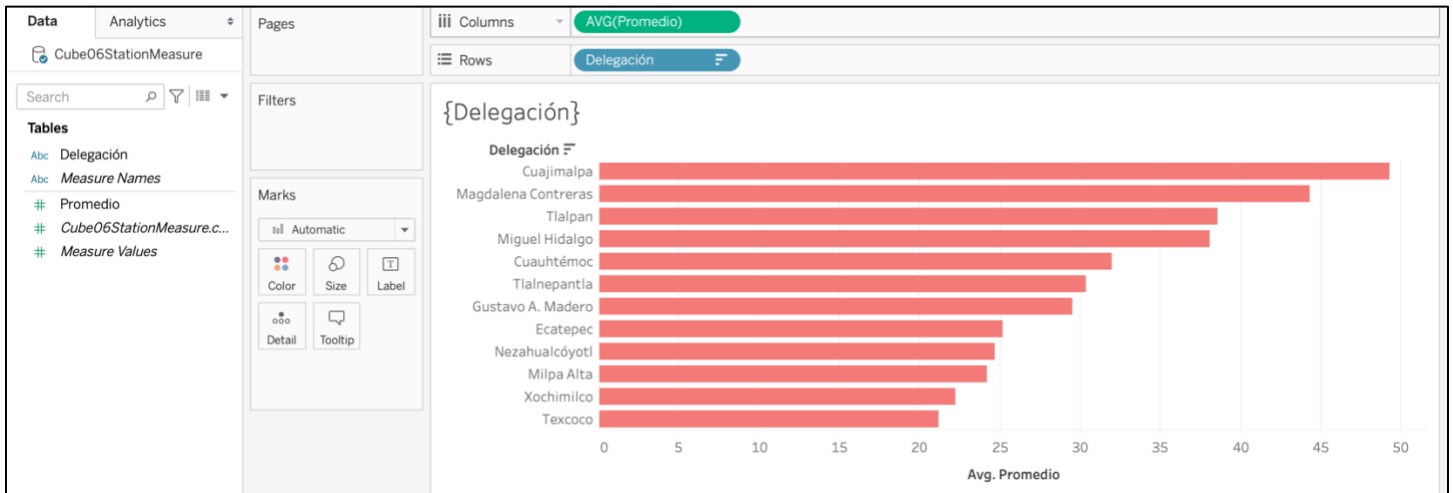
```
-- {Delegacion}{Año}
SELECT [Delegación]
      , [año]
      ,AVG([medicion]) AS "PROMEDIO"
INTO [precipitacion_pluvial].[dbo].[Cube05YearStationMeasure]
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]
GROUP BY [Delegación], [año]
ORDER BY [Delegación], [año], "PROMEDIO" DESC
```



Podemos observar el histórico por año de cada delegación o municipio respecto a los puntos promedio de captación de precipitación pluvial por año, aquí no solo podemos observar que estaciones han captado o registrado más o menos, sino que podemos ver, por ejemplo, Ecatepec tiene de las menores captaciones y además de manera continua a lo largo de 10 años, en Xochimilco pareciera que están disminuyendo la cantidad de precipitación pluvial registrada, que en 2019 ha sido de los máximos de captación con menor cantidad respecto a otros años anteriores en comparación con la misma estación en cuestión.

f) {Delegación}

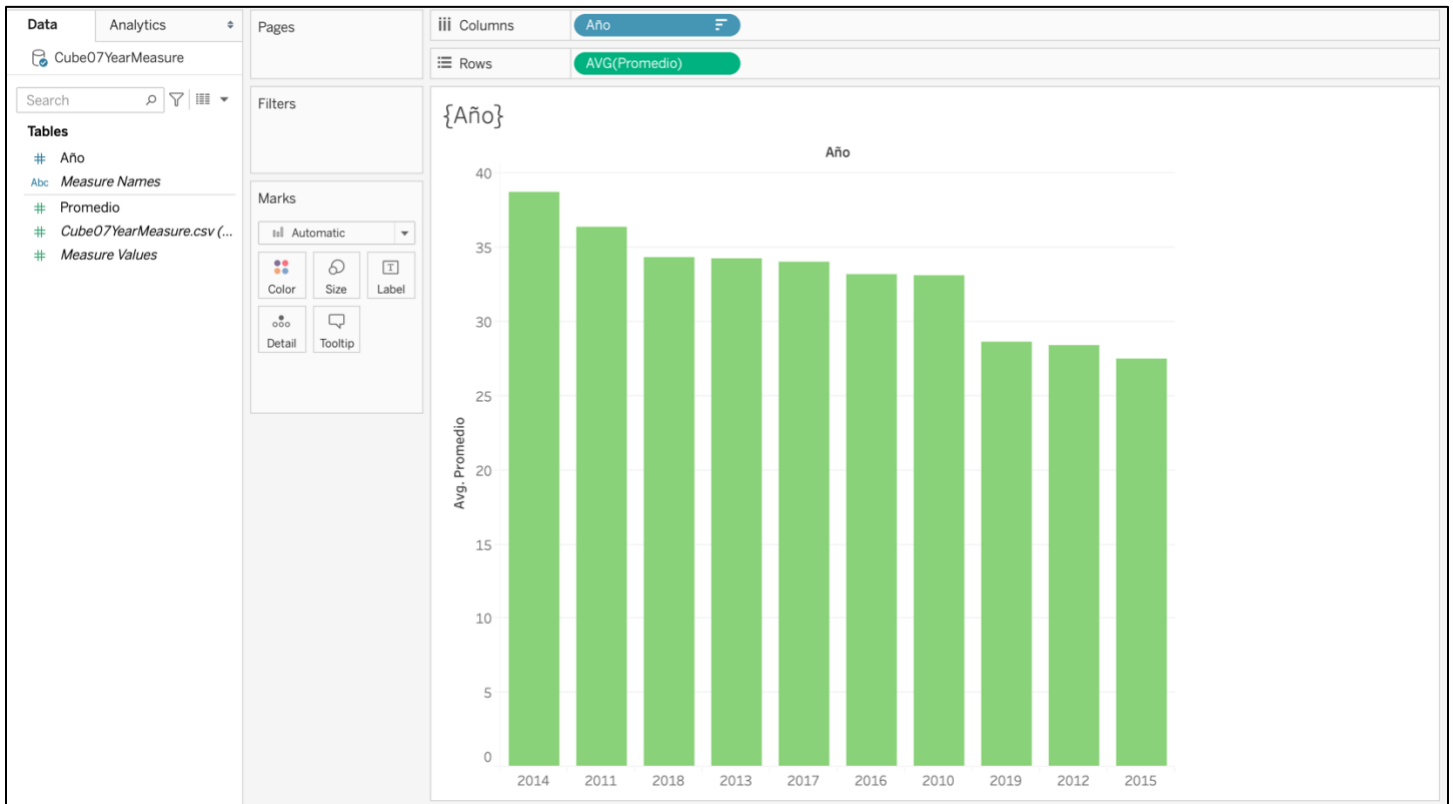
```
-- {Delegacion}
SELECT [Delegación]
      ,AVG([medicion]) AS "PROMEDIO"
INTO [precipitacion_pluvial].[dbo].[Cube06StationMeasure]
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]
GROUP BY [Delegación]
ORDER BY "PROMEDIO" DESC
```



Debo admitir que me sorprende ver a Milpa Alta, Xochimilco y Texcoco como las que tienen en promedio menor cantidad de registro de precipitación pluvial, siendo que Texcoco tiene vegetación, el antiguo lago de Texcoco, Xochimilco ni se diga y Milpa Alta lo mismo, pudiera haber algunos otros factores que ayuden a comprender este fenómeno, como contaminación, ubicación de las estaciones, etc.

g) {Año}

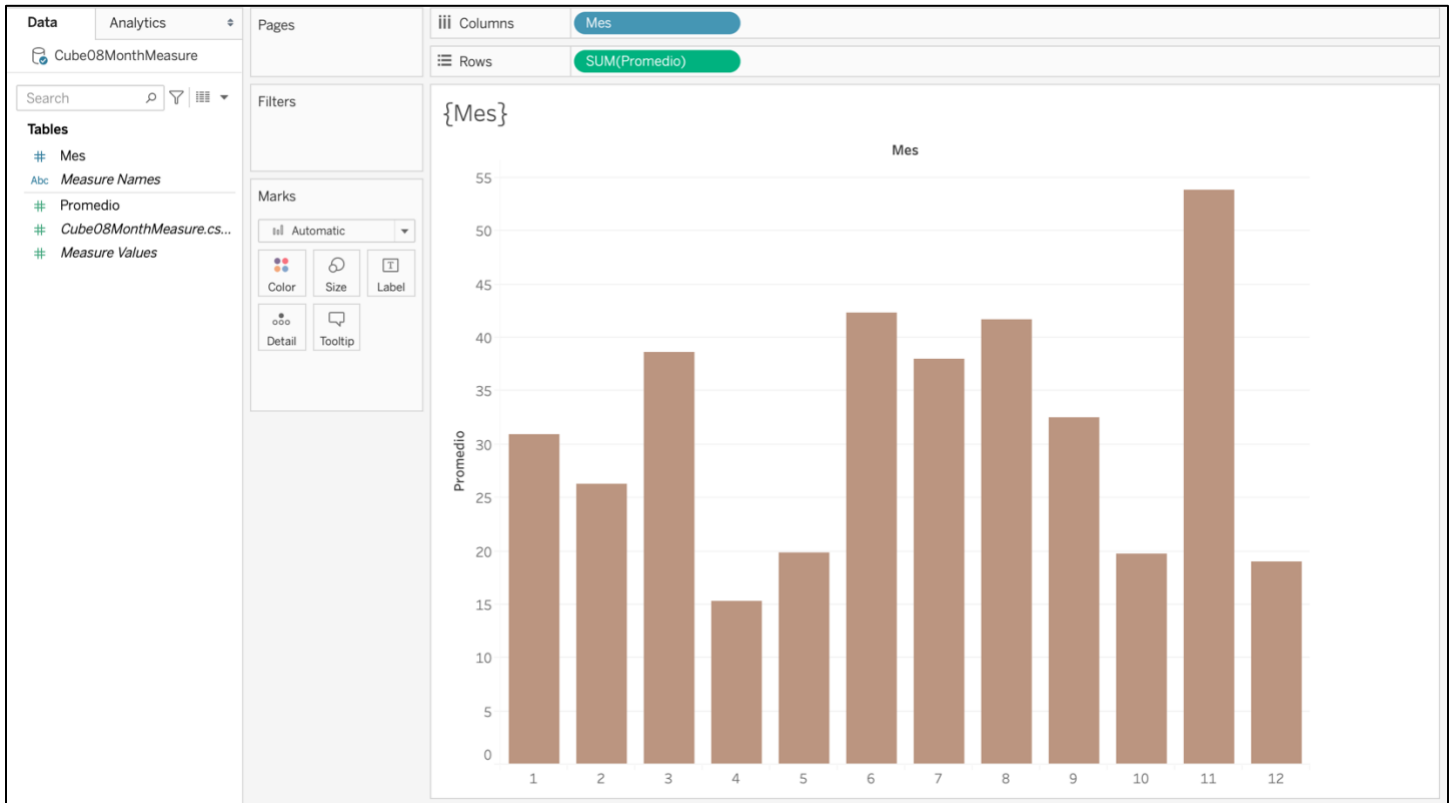
```
-- {Año}  
SELECT [año]  
        ,AVG([medicion]) AS "PROMEDIO"  
INTO [precipitacion_pluvial].[dbo].[Cube07YearMeasure]  
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]  
GROUP BY [año]  
ORDER BY [año] ASC
```



Podemos observar que entre el 2010 y 2019, el 2014 fue el año donde más se registro precipitación pluvial sin importar la estación, este es un análisis a nivel zona metropolitana, y que los años 2019, 2012 y 2015 han sido en donde menor registro de precipitación pluvial se ha hecho en todo este tiempo.

h) {Mes}

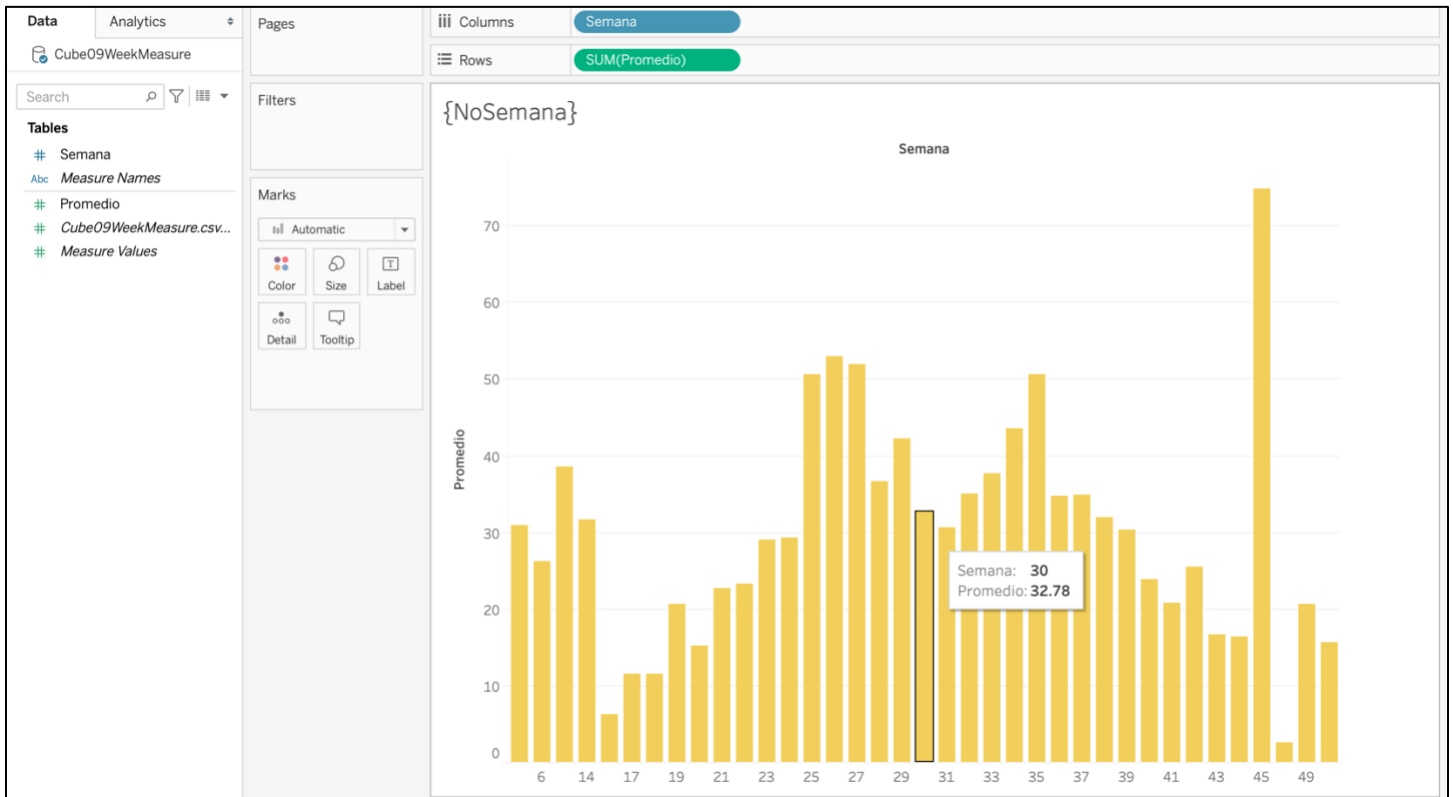
```
-- {Mes}  
SELECT [mes]  
        ,AVG([medicion]) AS "PROMEDIO"  
INTO [precipitacion_pluvial].[dbo].[Cube08MonthMeasure]  
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]  
GROUP BY [mes]  
ORDER BY [mes] ASC
```



Tal y como se mencionó anteriormente para una estación en particular, se confirma que noviembre es el mes donde en promedio se registra la mayor precipitación pluvial a lo largo del año y al menos en los últimos 10 años así ha sido, luego entre Junio y Agosto también se registra de manera considerable.

i) {noSemana}

```
-- {NoSemana}  
SELECT [semana]  
        ,AVG([medicion]) AS "PROMEDIO"  
INTO [precipitacion_pluvial].[dbo].[Cube09WeekMeasure]  
FROM [precipitacion_pluvial].[dbo].[View01WeekYearStationMeasure]  
GROUP BY [semana]  
ORDER BY [semana] ASC
```



Finalmente, este grafico que si lo apoyamos del anterior, la semana con mayor registro es la #45, que se encuentra en noviembre, y justo como lo mencionamos al principio, entre la semana 25 y la 25 es donde se concentra la mayoría de los puntos mas altos de registro de precipitación pluvial para la zona metropolitana más grande del país.

Conclusiones

Como lo hemos comentado en algunas clases anteriores, el analizar datos de esta manera al principio se deberá de saber el para que o al que se quiere encontrar, explícitamente no conocemos eso, yo de manera personal busco en donde y cuando es cuando más ha habido registro de precipitación pluvial, y ahora que he leído en redes sociales sobre la sequia que el rio Cutzamala que alimenta de agua a ala ciudad, pues es muy informativo revisar todo esto.

Es muy buena técnica primero procesar los datos de manera más especifica o a bajo nivel desde el manejador de base de datos, para luego finalmente solo llegar y graficarlo en otro software, quiero pensar que este tipo de información o archivos que se recaudan al realizar los cubos de datos, serán la entrada para los algoritmos de machine learning en predicción por ejemplo.