



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Data Mining

Práctica 5

Exploración de Cubos de datos de precipitación pluvial en CDMX.

Grupo:3CV18

Integrantes:

Cazares Martínez Maximiliano

Ramos Nieves Adrián

Profesor:

Roberto Eswart Zagal Flores

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 práctica:

- Obtener los catálogos y tablas de hechos realizados en la práctica anterior.
- Crear una vista con la información necesaria.
- Crear cada cubo de datos
- Exportar todos los cubos de datos
- Analizar cada cubo de datos 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 tabla

- a) Catálogo de estaciones de monitoreo
- b) Catálogo de elementos de medición
- c) Prepare una tabla de hechos que contenga la siguiente estructura mínima:
{Elemento}(E), {Delegación}(D), {noSemana}(S), {Anio}(A), {mes}(M), {medición}(P)

```
1  -- Tabla de hechos
2  select M.Elemento,
3  EM.Delegacion_municipio,
4  M.Semana, M.Anio, M.Mes, M.Medicion
5  into TablaHechos
6  from EstacionesMonitoreo EM, Mediciones M
7  where EM.Id_estacion = M.Id_estacion
8  select * from TablaHechos;
```

2. A partir de la tabla de hechos anterior desarrolle cubos que se solicitan. Use sólo 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

a) {Delegación}, {Semana}, {Anio}

```

1  -- Cube 01
2  -- {Delegación},{noSemana},{Anio}
3  select Delegacion_municipio, Semana, Anio, Medicion
4  into Cube01 from TablaHechos where Medicion != -99
5  select * from Cube01;

```

{Delegación},{noSemana},{Anio}



Como se puede observar la precipitación pluvial a lo largo de los años es constante y se nota una tendencia entre las semanas 20 y 40 de cada año. Sin embargo, entre las semanas 25 y 35 es donde más precipitación pluvial existe.

b) {Delegación}, {noSemana}, {Mes}

```

1  -- Cube 02
2  --{Delegación},{noSemana},{Mes}
3  select Delegacion_municipio, Semana, Mes, Medicion
4  into Cube02 from TablaHechos where Medicion != -99
5  select * from Cube02;

```

{Delegación},{noSemana},{Mes}



En la anterior gráfica podemos observar que la temporada de lluvias empieza en el mes de abril y termina a finales de octubre. Además, podemos observar cómo en algunas Delegaciones la precipitación es más abundante que en otras. Tal es el caso de, Cuajimalpa, Gustavo A. Madero, Magdalena Contreras y Tlalpan

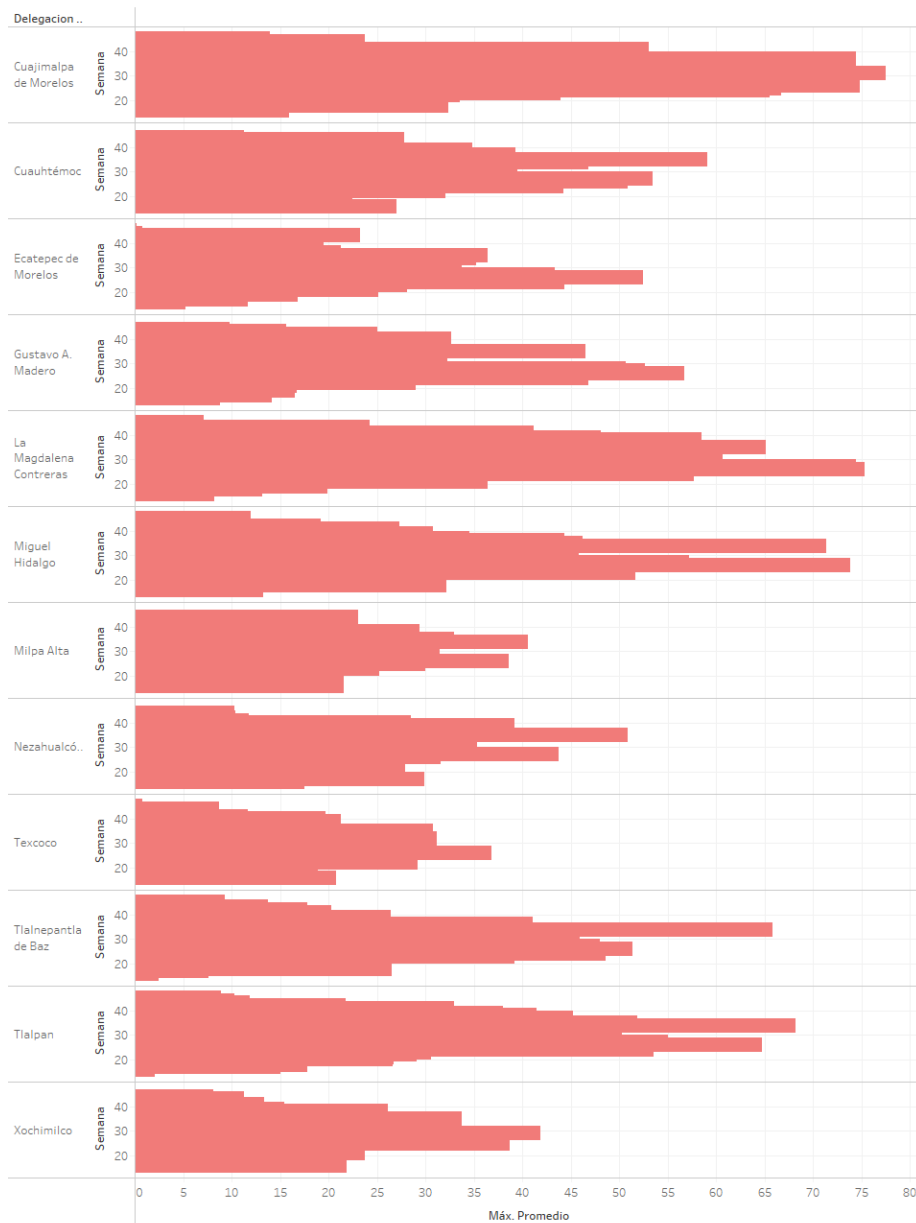
c) {Delegación}, {noSemana}

```

1  -- Cube 03
2  --{Delegación},{noSemana}
3  select Delegacion_municipio, Semana, avg(Medicion) as Promedio
4  into Cube03 from Cube01
5  group by Delegacion_municipio, Semana order by Promedio desc
6  select * from Cube03;
7

```

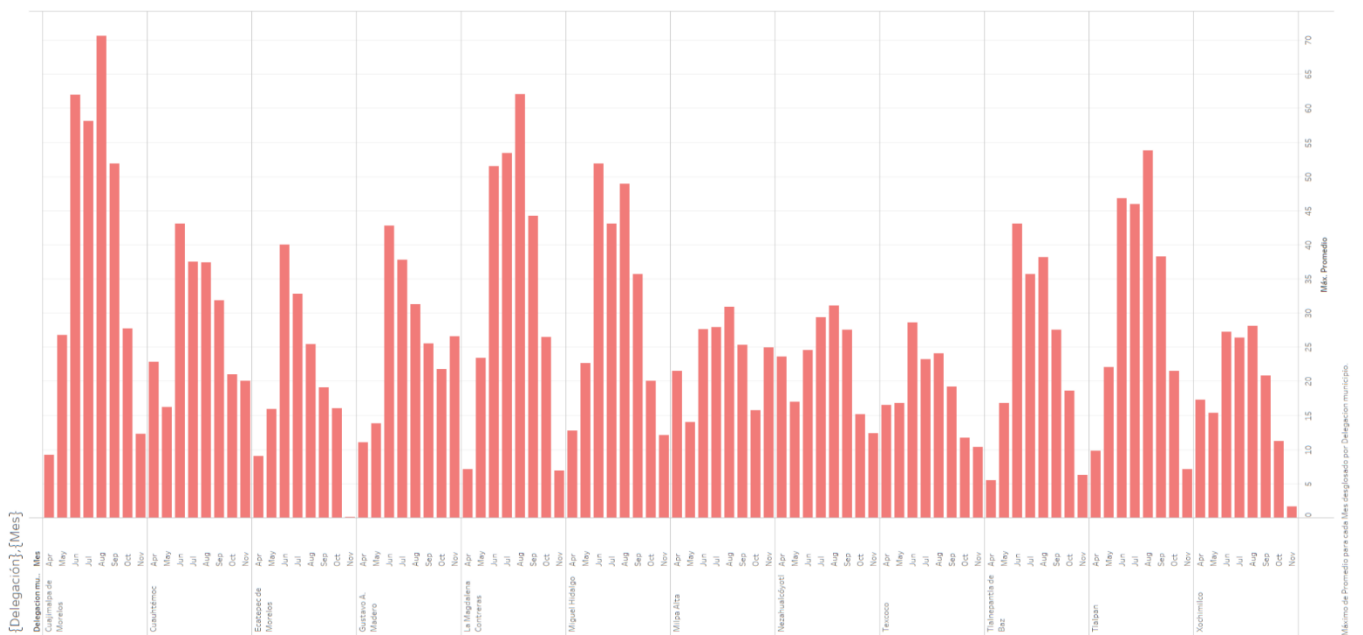
{Delegación},{noSemana}



En este gráfico podemos apreciar las delegaciones donde se presenta mayor precipitación pluvial por semana. De nuevo podemos observar cómo las delegaciones Cuajimalpa, Magdalena Contreras y Tlalpan entre las semanas 25 y 35 es cuando más llueve.

d) {Delegación}, {Mes}

```
1  -- Cube 04
2  --{Delegación},{Mes}
3  select Delegacion_municipio, Mes, avg(Medicion) as Promedio
4  into Cube04 from TablaHechos where Medicion != -99
5  group by Delegacion_municipio, Mes order by Delegacion_municipio, Mes, Promedio asc;
6  select * from Cube04;
```



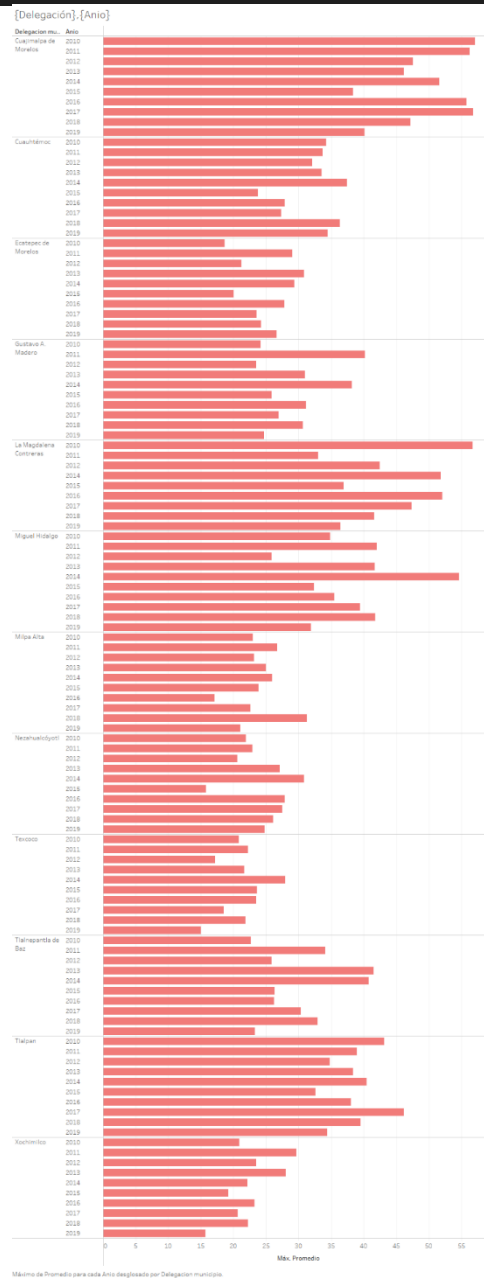
En esta gráfica podemos observar algo similar a la anterior, solo que esta vez tomamos como referencia el mes. Podemos distinguir que en la delegación Tlalpan, Magdalena de las Contreras y Cuajimalpa tienen más precipitaciones en el mes de agosto, a diferencia de las otras delegaciones en el cual varía en qué momentos se pueden observar estos datos.

e) {Delegación}, {Año}

```

1  -- Cube 05
2  --{Delegación},{Año}
3  select Delegacion_municipio, Año, avg(Medicion) as Promedio
4  into Cube05 from Cube01
5  group by Delegacion_municipio, Año order by Delegacion_municipio, Año, Promedio asc
6  select * from Cube05;

```

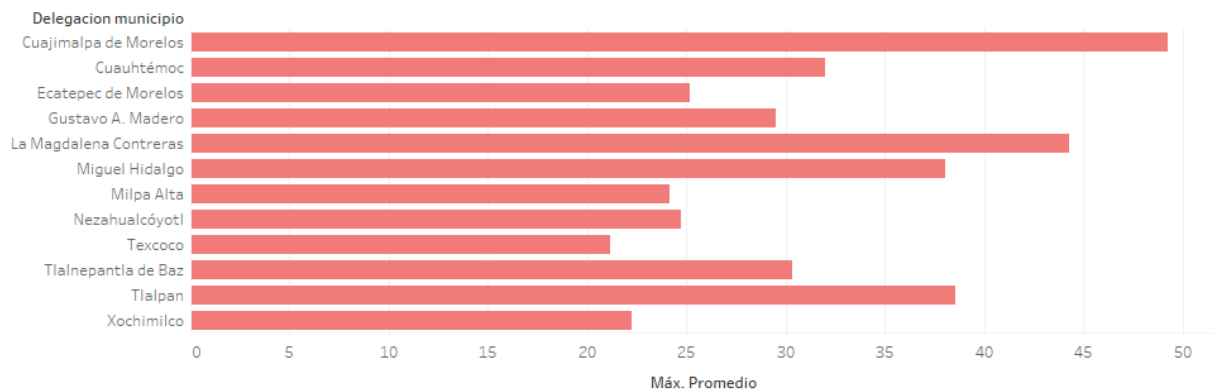


En la siguiente gráfica podemos ver el histórico de cada año dependiendo de la delegación en la que se encuentren los datos, podemos precisar la cantidad de datos que se han obtenido a lo largo de los años acerca de la precipitación, dándonos un análisis de los posibles lugares con menos PP, o donde es más frecuente captar este tipo de datos para su recolección. Podemos observar que en el año 2010 en varias delegaciones se encuentra el pico de mayor PP en 10 años.

f) {Delegación}

```
1  -- Cube 06
2  --{Delegación}
3  select Delegacion_municipio, avg(Medicion) as Promedio
4  into Cube06 from Cube01
5  group by Delegacion_municipio order by Promedio asc
6  select * from Cube06;
```

{Delegación}

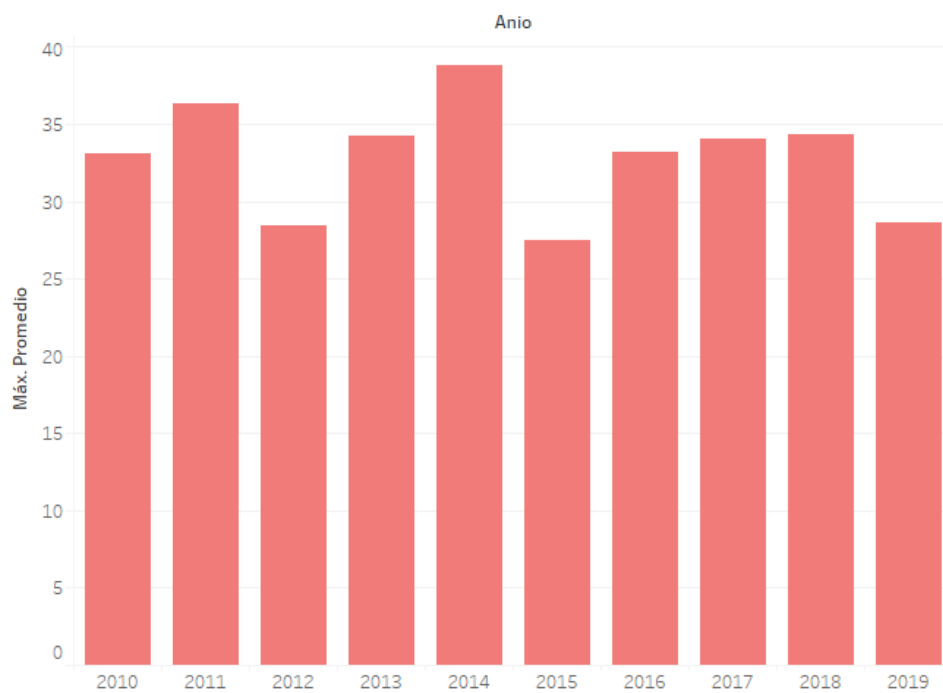


Con esta gráfica captamos que el lugar con más PP es Cuajimalpa de Morelos, con el segundo puesto es Magdalena de Contreras, que pueden ayudar a resolver problemáticas en estos lugares donde posiblemente sufre de más inundaciones en el año, y que posiblemente tengan algunos percances por esto.

g) {Anio}

```
1  -- Cube 07
2  --{Anio}
3  select Anio, avg(Medicion) as Promedio
4  into Cube07 from Cube01
5  group by Anio order by Anio asc
6  select * from Cube07;
```

{Anio}

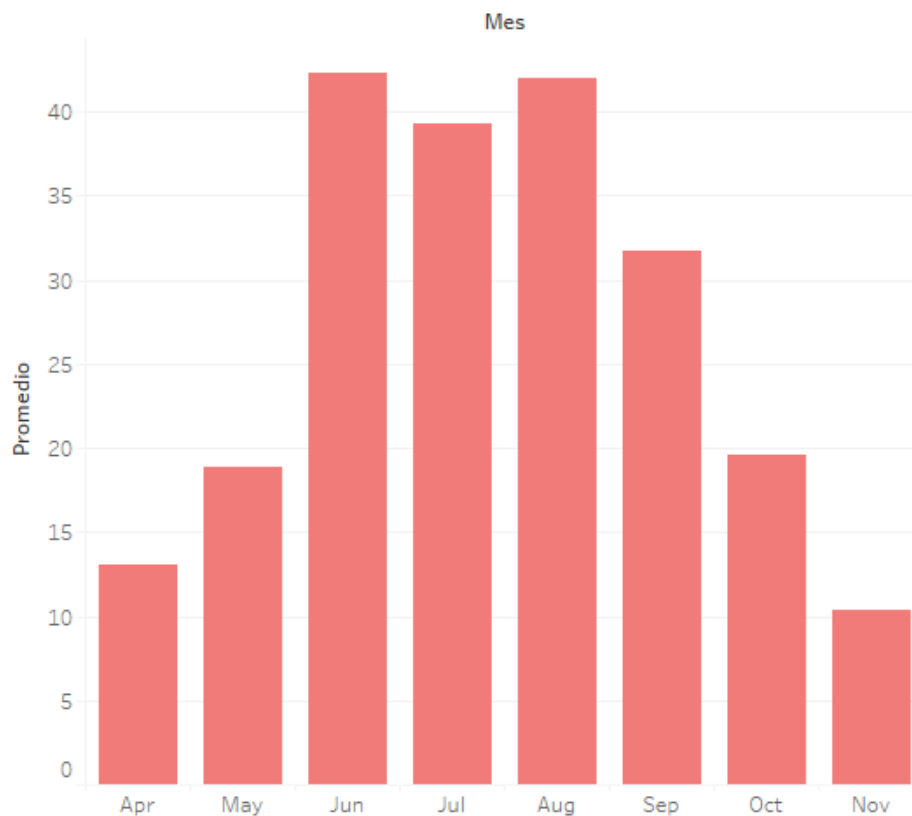


Podemos observar que el año en el cual se obtuvieron más datos de PP fue en 2014, probablemente son los datos máximos recolectados hasta este momento, algo que nos puede dar una pista sobre el cambio climático que hemos estado sufriendo en los últimos años ya que se tienen en los próximos años unas altas y bajas considerables.

h) {Mes}

```
1  -- Cube 08
2  --{Mes}
3  select Mes, avg(Medicion) as Promedio
4  into Cube08 from TablaHechos where Medicion !=-99
5  group by Mes order by Mes asc
6  select * from Cube08;
```

{Mes}

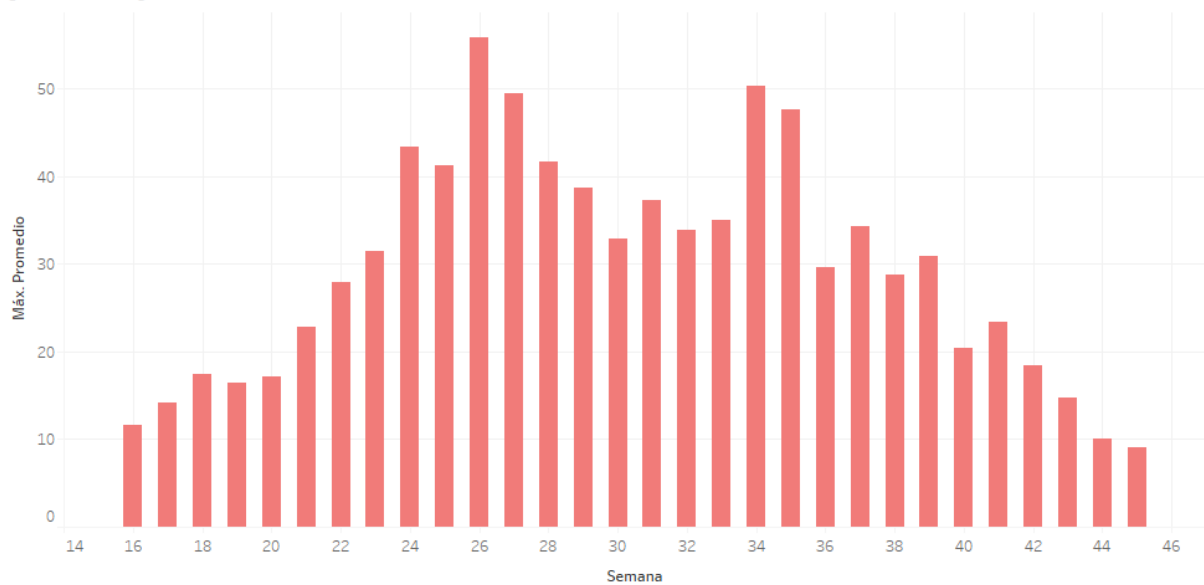


En la gráfica anterior se muestra que la mayor cantidad de PP se encuentra entre los meses de junio y agosto, coincidiendo con la temporada de lluvias de cada año y durante 10 años esta temporada no se ha visto significativamente alterada.

i) {noSemana}

```
1  -- Cube 09
2  --{noSemana}
3  select Semana, avg(Medicion) as Promedio
4  into Cube09 from Cube01
5  group by Semana order by Semana asc
6  select * from Cube09;
```

{noSemana}



Finalmente, con la última gráfica podemos respaldar nuestros datos anteriores del punto h, ya que, si observamos las semanas en el calendario y las comparamos con el mes, obtenemos que son aproximadas al periodo junio-julio y agosto-septiembre en los picos más altos, correspondiendo a la temporada de lluvias.

Conclusiones

A lo largo de la práctica aplicamos los conocimientos vistos en clase acerca de los cubos de datos. Es una gran técnica cuando se pretende analizar los datos de manera específica, ya que al hacer distintas combinaciones con todos los indicadores de la tabla y al hacer uso de un software graficador podemos observar la información específicamente. De esta manera, podemos hacer uso de los datos para proponer soluciones a diversas problemáticas, y en el caso de esta práctica las inundaciones son casualmente el mayor problema de las delegaciones en donde se presenta mayor precipitación pluvial a lo largo del año.