





### Práctica 7

### Práctica 7:

Lattice de cubo de datos y exploración del datawarehouse del proyecto semestral.

### CONTENIDO

1.	OBJETIVO	2
2.	INTRODUCCIÓN	2
3.	DESARROLLO	3
4.	CONCLUSIÓN	7
5.	REFERENCIAS	8



#### Práctica 7



### 1. OBJETIVO

Comprender el concepto de lattice de cubos de datos, y utilizar la arquitectura de datos para minería.

### 2. INTRODUCCIÓN

Esta práctica se desarrollará con el lenguaje de programación Python, en la cual la principal librería que se utiliza es la librería Pandas.

Las principales características de esta librería son:

- Define nuevas estructuras de datos basadas en los arrays de la librería NumPy, pero con nuevas funcionalidades.
- Permite leer y escribir fácilmente ficheros en formato CSV, Excel y bases de datos SQL.
- Permite acceder a los datos mediante índices o nombres para filas y columnas.
- Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.
- Permite trabajar con series temporales.
- Realiza todas estas operaciones de manera muy eficiente.

Esta librería se utilizará principalmente para la extracción de datos de los archivos xls. MySQL Workbench es un software creado por la empresa Sun Microsystems, esta herramienta permite modelar diagramas de Entidad-Relación para bases de datos MySQL.

Con esta herramienta se puede elaborar una representación visual de las tablas, vistas, procedimientos almacenados y claves foráneas de la base de datos. Además, es capaz de sincronizar el modelo en desarrollo con la base de datos real. Se puede realizar una ingeniería directa e ingeniería inversa para exportare e importar el esquema de una base de datos ya existente el cual haya sido guardado o hecho copia de seguridad con MySQL Administrador.

MySQL Workbench puede generar también el guion necesario para crear la base de datos que se ha dibujado en el esquema; es compatible con los modelos de base de datos de DBDesigner 4 y soporta las novedades incorporadas en MySQL 5.x

La versión utilizada en esta práctica es la 8.0.12.

Tableau.

Tableau es una herramienta de visualización de datos potente utilizada en el área de la Inteligencia de negocios (más conocida como Business Intelligence). Simplifica los datos en bruto en un formato muy fácil de entender.



# ESCOM

#### Práctica 7

La esencia de Tableau es simple y a la vez muy relevante: ayudar a las personas y empresas a ver y comprender todos sus datos. Y esto lo consigue ofreciendo a los usuarios toda una selección de herramientas útiles e intuitivas de inteligencia de negocios.

A través de funciones simples como la de arrastrar y soltar, cualquier persona puede acceder y analizar de forma sencilla datos, e incluso, crear informes y compartir esta información con otros usuarios.

### 3. DESARROLLO

Procedimiento: Desarrollar la automatización de la lattice de cubos de datos para construir el data warehouse del proyecto semestral, y hacer la exploración del mismo.

- 1. Seleccione entre 4 y 6 dimensiones (n) de la tabla de hechos diseñada en la clase anterior. Recuerde que la cantidad máxima de cubo de datos es 2n.
- En esta práctica se utilizarán 4 dimensiones de la tabla de hechos, las cuales serían las siguientes:

Delito, Locación, Categoria Delito y Tiempo\_hechos

- 2. Construya un programa, en el lenguaje de programación de preferencia, donde construya la lattice de cubos de su proyecto. Tienes dos posibilidades sobre la materialización de los cubos.
  - a. Crear vistas en la base de datos, esto ahorraría espacio en disco, pero requerirá tiempo de procesamiento al conectar al cubo a otro software como tableau. También podrá controlar el código fuente con el que la vista o cubo fue realizado.
  - b. Crear tablas en la base de datos. Esto quitará espacio en disco duro, pero ahorraría tiempo de ejecución y procesamiento al exportar el cubo.
  - c. Desarrolle un diagrama de flujo o de procesos, explique el diagrama en el reporte y agregue fragmentos de código en las tareas más importantes.
- Se obtuvieron las tablas por medio de una consulta a la base de datos generada en la práctica anterior, junto como se muestra en la siguiente captura.





#### Práctica 7

- Después de crear las 4 tablas individuales a combinar, procedemos a realizar las distintas combinaciones de la siguiente manera.

```
def Nivel_2(h, f, t, 1):
    ht = pd.merge(h, t, left_on='idHechos', right_on='id')
    print("Creando tabla de Hechos_Tiempo...")
    ht.to_sql(con=dtb, name='hechos_tiempo', if_exists='replace', index=False)
    print("Tabla creada con exito")
```

- En caso de que las tablas que vamos a combinar no tengan algún campo en común, las combinaremos de la siguiente manera.

```
tl = pd.concat([t, l], axis=1)
print("Creando tabla de Tiempo_Locacion...")
tl.to_sql(con=dtb, name='tiempo_locacion', if_exists='replace', index=False)
print("Tabla creada con exito")
```

- Las demás tablas y combinaciones se crearon de forma similar.
  - d. Recuerde explicar las medidas estadísticas usadas: count, avg, max, min, etc.
- 3. Realice los ajustes que considere necesarios, elimine dimensiones, renombre el nombre de las dimensiones. Trate de utilizar valores numéricos en medida de lo posible en su tabla de hechos que usará para crear la lattice.
  - a. Recuerde que después de crear el cubo de datos puede vincularlo con sus respectivas dimensiones o catálogos.
- Se vincularon las dimensiones con sentencias SQL después de haberlas creado por medio del ETL.





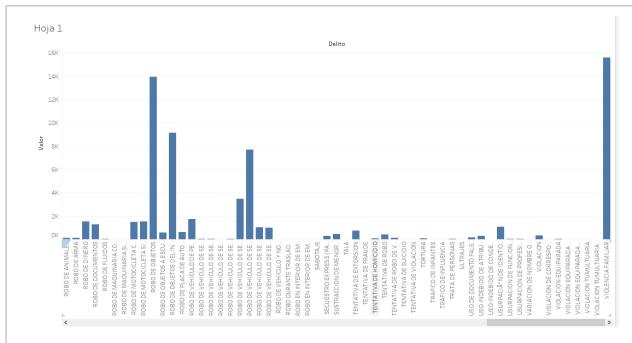
### Práctica 7

Tabl	а	Ω
_	categoria_delito	
	hechos	
⊞	hechos_categoria	
	hechos_locacion	
⊞	hechos_locacion_categoria	
田	hechos_tiempo	
$\blacksquare$	hechos_tiempo_categoria	
⊞	hechos_tiempo_locacion	
⊞	hechos_tiempocion_categoria	
⊞	locacion	
⊞	locacion_categoria	
⊞	tiempo_categoria	
⊞	tiempo_hechos	
	tiempo_locacion	
	tiempo_locacion_categoria	

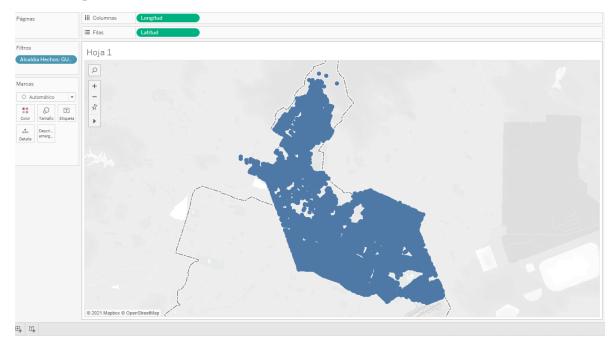
- b. En caso de que el punto anterior no sea posible (e.j un cálculo por delegación debe considerar en el agrupamiento dicho campo de la dimensión)
- 4. Conecte el datawarehouse a tableau y Desarrolle la exploración de los cubos de datos. Desarrolle el análisis sobre los cubos de datos de nivel 3 (dimensiones de espacio y tiempo, más una dimensión diferente a las diferentes (seleccione una dimensión interesante a su criterio)). Posteriormente, agregue otros cubos de distintos niveles que a su criterio considere más importantes.
- Se muestra una dimensión de las tablas de hechos en la dimensión de delito.







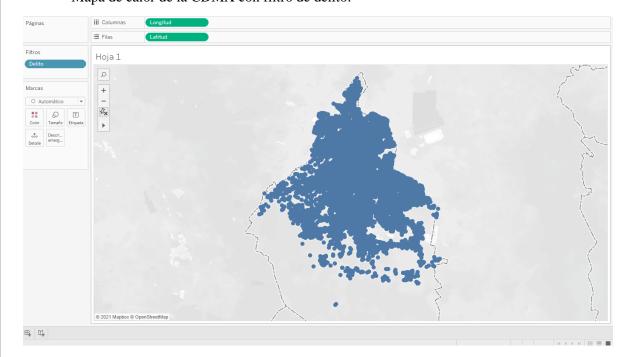
- a. Recuerde que su proyecto debe incluir alguna referencia geográfica, por lo tanto, desarrolle un mapa de calor en Tableau.
- Mapa de calor de la CDMX con filtro de alcaldía de Gustavo A Madero







- Mapa de calor de la CDMX con filtro de delito.



- b. Explique los filtros usados y las razones por las que se usan las dimensiones seleccionadas
- Los filtros que se usaron se utilizaron para mostrar y filtrar las dimensiones y datos de tal manera que se puedan conocer los delitos ocurridos en la alcaldía GAM.

### 4. CONCLUSIÓN

"En esta práctica pudimos trabajar y poner en práctica lo aprendido en clases teóricas con el dataset que elegimos en prácticas anteriores para nuestro proyecto, al cual le aplicamos la técnica ETL para recopilar y manejar directamente los datos para poderlos utilizar, además de la creación de una lattice de datos para un mejor manejo de la información de nuestro proyecto".





### 5. REFERENCIAS

- [1]. Alberca, A. S. (2020, 4 octubre). La librería Pandas. Aprende con Alf. https://aprendeconalf.es/docencia/python/manual/pandas/
- [2]. Pública, A. D. D. I. (s. f.). Portal de Datos Abiertos de la CDMX. Portal de datos Abiertos de la CDMX. Recuperado 23 de mayo de 2021, de <a href="https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx">https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx</a>
- [3]. Nelson Aranibar, Monografias.com. (s. f.). MySQL WorkBench. Monografias.com. Recuperado 3 de marzo de 2021, de <a href="https://www.monografias.com/trabajos88/mysql-workbench/mysql-workbench.shtml">https://www.monografias.com/trabajos88/mysql-workbench.shtml</a>
- [4]. Tableau, una de las principales herramientas de Inteligencia de Negocios. (s. f.). SPnet. Recuperado 28 de marzo de 2021, de <a href="https://softwarepara.net/tableau">https://softwarepara.net/tableau</a>
- [5]. Pública, A. D. D. I. (s. f.). Portal de Datos Abiertos de la CDMX. Portal de datos Abiertos de la CDMX. Recuperado 23 de mayo de 2021, de https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx