



# Sistemas organizacionales y gerenciales 2[N]

## José Luis Reynoso Tiú

Día, Fecha:	Sábado, 22/02/2025
Hora de inicio:	08:00AM



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

# Clase 5



# Información de contacto



## FOROS DE UEDI

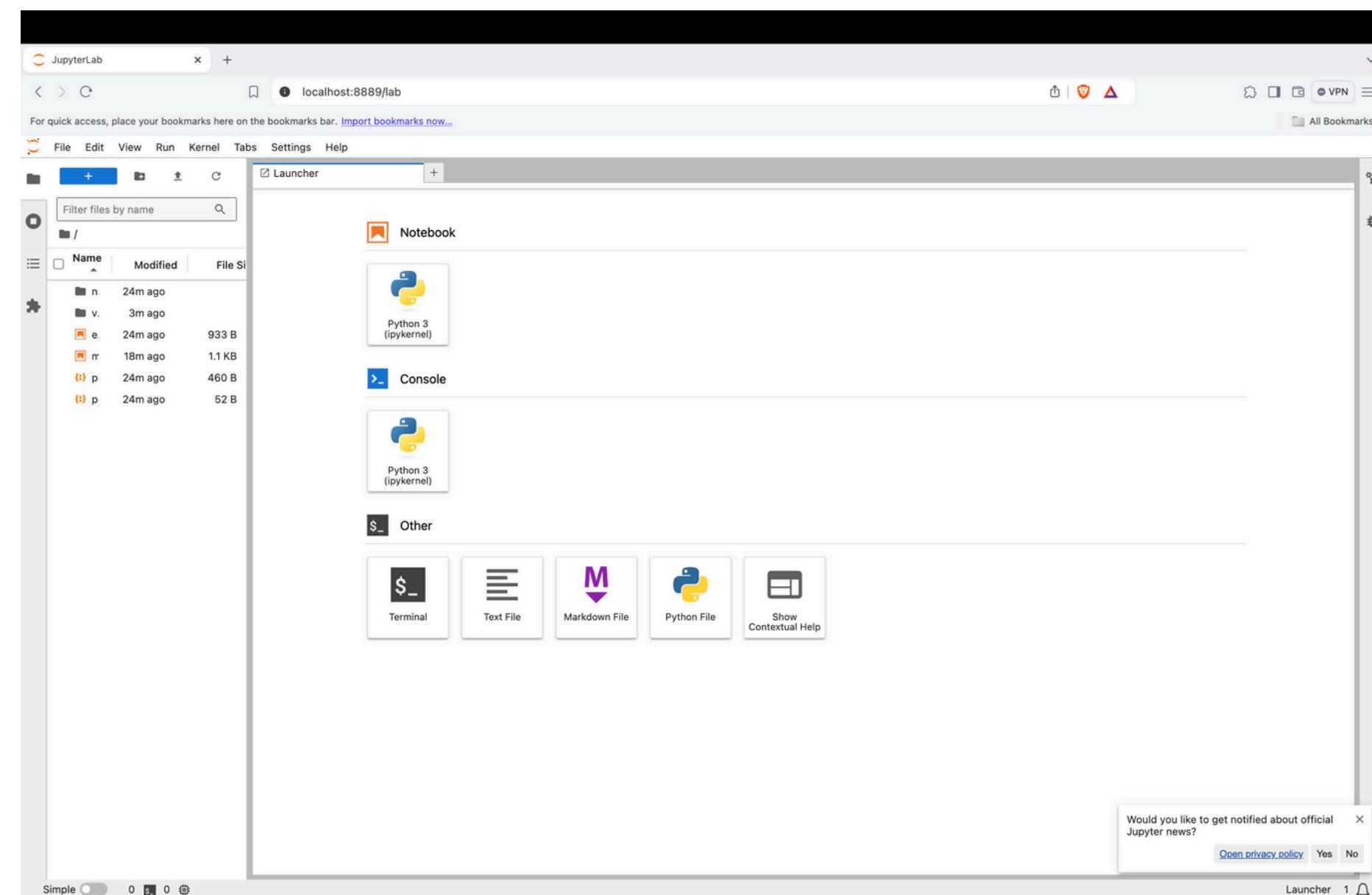
Semanalmente se creará un foro en UEDI para resolver dudas.

**2655896221401@ingenieria.usac.edu.gt**

[SOG2]Asunto

# Jupyter Notebook

Jupyter Notebook es un entorno interactivo para ejecutar código en el navegador. Es una excelente herramienta para el análisis exploratorio de datos y es ampliamente utilizada por científicos de datos.



## JupyterLab

Install JupyterLab with [pip](#):

```
pip install jupyterlab
```

Note: If you install JupyterLab with conda or mamba, we recommend using the [conda-forge channel](#).

Once installed, launch JupyterLab with:

```
jupyter lab
```

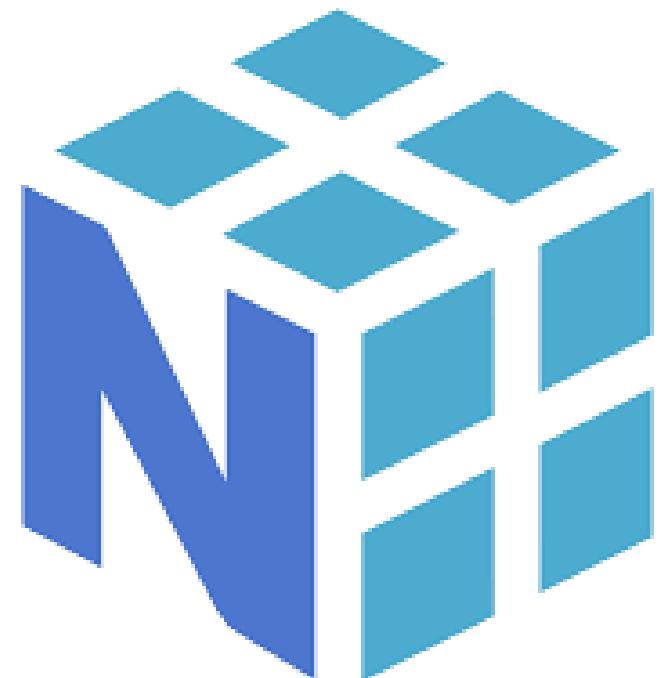


# NumPy

- Es una biblioteca fundamental para la ciencia de datos en Python.
- Se utiliza principalmente para realizar cálculos matemáticos y científicos.
- Contiene funcionalidad para matrices multidimensionales, funciones matemáticas de alto nivel como operaciones de álgebra lineal y la transformadas de Fourier, y generadores de números pseudoaleatorios.



La generación de números pseudoaleatorios consiste en crear una secuencia de números que aparenta ser aleatoria, aunque realmente se produce mediante un algoritmo determinístico



# Pandas

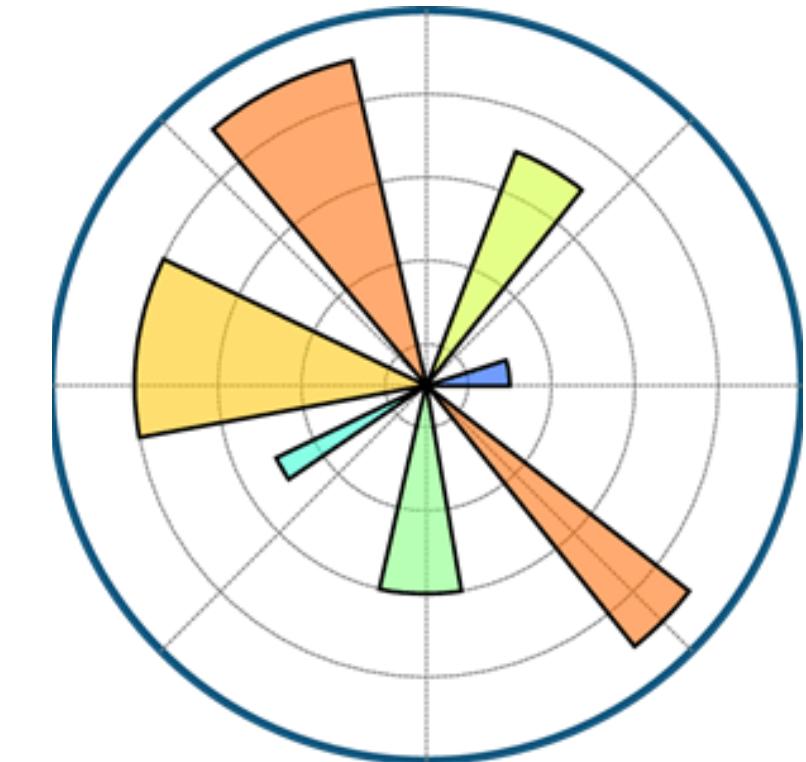
- Pandas es una biblioteca de Python para el análisis y la manipulación de datos.
- Está construida alrededor de una estructura de datos llamada DataFrame que se basa en el DataFrame de R.
- Un DataFrame en pocas palabras es una tabla.
- Permite ingerir datos desde una gran variedad de formatos.



# Matplotlib

- Es la biblioteca principal de gráficos científicos en Python.
- Es una biblioteca completa para crear visualizaciones estáticas, animadas e interactivas en Python.
- **pyplot**: Es un módulo de Matplotlib que ofrece varias funciones simples para agregar elementos como líneas, imágenes o textos a los ejes de un gráfico.

```
In [1]: import matplotlib.pyplot as plt
```



# Seaborn

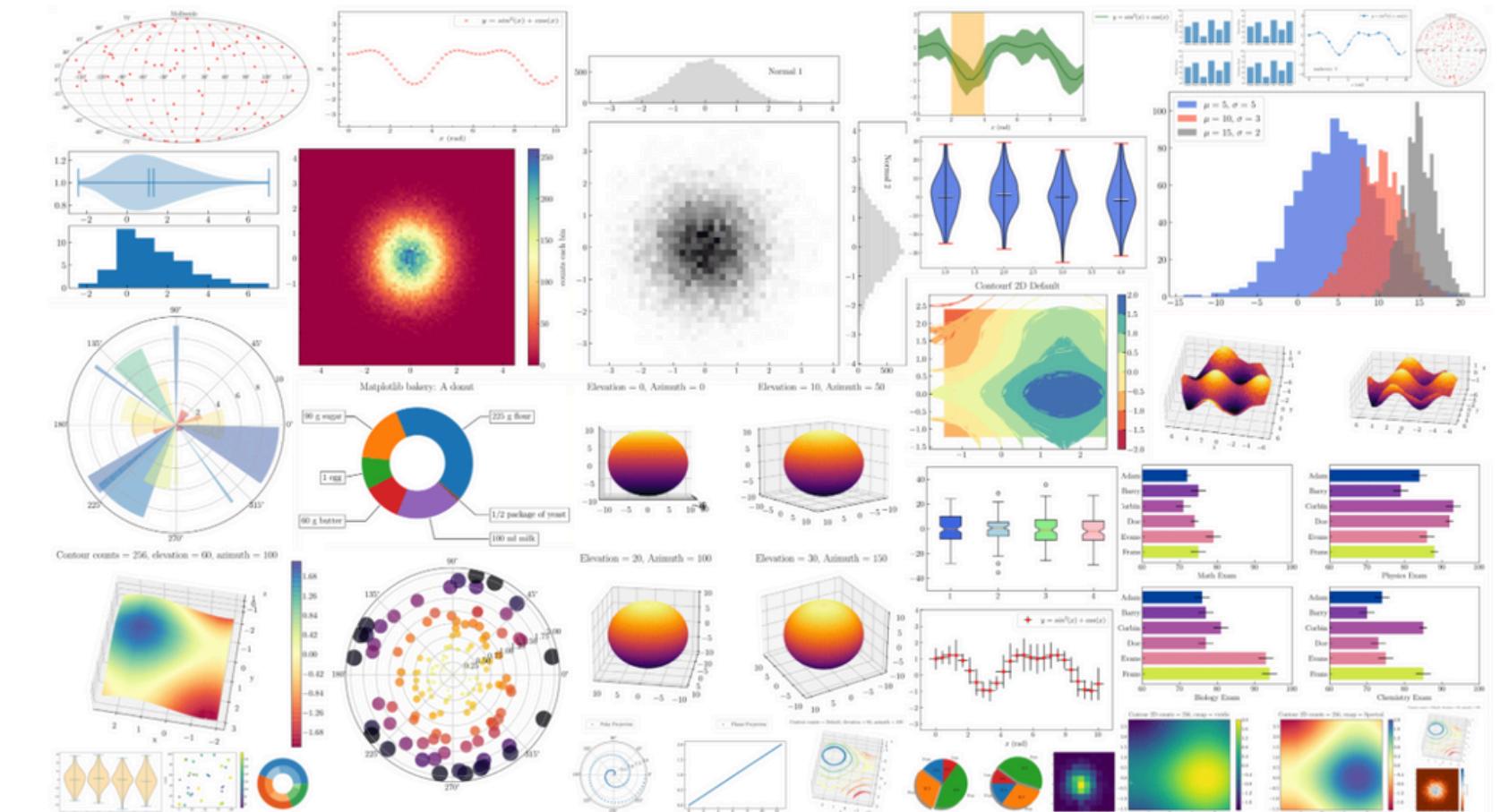
- Es una librería de Python basada en matplotlib para la visualization de datos.
- Está diseñada para trabajar con DataFrames de Pandas.
- Una ventaja de Seaborn es su estrecha integración con Pandas y sus DataFrames, aunque Matplotlib también se integra con Pandas y NumPy.
- En general, Seaborn es la mejor herramienta para visualizaciones estadísticas, mientras que Matplotlib se destaca por su capacidad de personalización.



seaborn

# VIZUALIZACIÓN DE DATOS

- Una parte fundamental del conjunto de herramientas del científico de datos es la visualización de datos. Si bien es muy fácil crear visualizaciones, es mucho más difícil producir buenas visualizaciones.
- Hay dos usos principales para la visualización de datos:
  - Para explorar información
  - Para comunicar información



# NETFLIX

La aplicación de la ciencia de datos en las empresas no solo contribuye a aumentar los ingresos, sino también a mitigar las pérdidas.

Mediante el análisis de datos, la plataforma de streaming desarrollada por la empresa puede sugerir a los usuarios contenido que les resulte atractivo, lo que resulta en que aproximadamente el 80% del contenido consumido en Netflix provenga de estas recomendaciones.



# AIRBNB



Gracias al análisis de datos, la plataforma de reserva de alojamientos vacacionales experimentó un crecimiento del 43,000% en tan solo cinco años.

Una de las particularidades más destacadas de su enfoque para entender a los clientes mediante el data science es que no se limita únicamente a considerar las calificaciones numéricas que estos otorgan a los alojamientos, ya que consideran que estas no son suficientes para comprender su verdadera opinión. Por ello, también emplean técnicas de procesamiento de lenguaje natural para captar realmente el significado de los comentarios y los sentimientos subyacentes en ellos. De esta manera, logran clasificar de manera más precisa cada uno de los alojamientos y desarrollar un algoritmo que realmente sugiera opciones interesantes para el público.

# AMAZON

Al implementar la ciencia de datos, Amazon ha logrado personalizar por completo el proceso de compra de los consumidores, lo que permite que cada usuario vea exactamente lo que desea. Esto implica que los usuarios pueden visualizar productos y contenido específico que les gusta.

Este nivel de personalización se logra mediante un análisis de datos que contiene la información proporcionada por cada usuario. Además, se emplean algoritmos de inteligencia artificial para determinar las preferencias del consumidor.



# DATA WAREHOUSE

Un Data Warehouse es un almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información. Los datos de un data warehouse deben almacenarse de forma segura, fiable, fácil de recuperar y fácil de administrar.

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura conocida ya en muchas empresas modernas.

# ¿Cómo funciona un Data Warehouse?

Un data warehouse funciona como un repertorio central. La información proviene de una o varias fuentes de datos, como sistemas transaccionales u otras bases de datos relacionales.

Los datos pueden ser estructurados, semiestructurados o no estructurados. Una vez que están integrados en el warehouse, son tratados y transformados. Los usuarios pueden luego acceder a ellos con la ayuda de herramientas de Business Intelligence, de clientes SQL o de tablas.

Agregando la información en el mismo lugar, una empresa logra tener una visión de conjunto de su clientela o de otros elementos cruciales

# Tipos de Data Warehouse

- **Enterprise Data Warehouse (EDW)** es un almacén centralizado que proporciona servicios de apoyo a la toma de decisiones en toda la empresa.
- **Operational Data Store (ODS)** es simplemente un almacén de datos necesario cuando ni el Data Warehouse ni los sistemas OLTP soportan las necesidades de reporte de las organizaciones.
- **Data Mart** es un subconjunto del data warehouse, diseñado especialmente para una línea de negocio particular, como ventas, finanzas, etc.



# ¿Quién utiliza un DWH?

Los Data Warehouse son utilizados por todas las empresas que deben tratar grandes volúmenes de datos o que recolectan datos de múltiples fuentes. Son utilizados también por empresas que desean acceder con mayor facilidad a sus datos.

Para cualquier empresa que busque asistencia en la toma de decisiones, los Data Warehouse pueden resultar pertinentes. Este es también el caso para los usuarios que buscan gestionar informes, gráficos y diagramas a partir de datos. Los Data Warehouses tienen su lugar en todos los sectores de actividad. Sin embargo, son utilizados de formas muy diferentes según la industria.

# OLAP Y OLTP

El procesamiento analítico en línea (OLAP) y el procesamiento de transacciones en línea (OLTP) son sistemas de procesamiento de datos que ayudan a almacenar y analizar datos empresariales. Puede recopilar y almacenar datos de múltiples fuentes, como sitios web, aplicaciones, medidores inteligentes y sistemas internos.

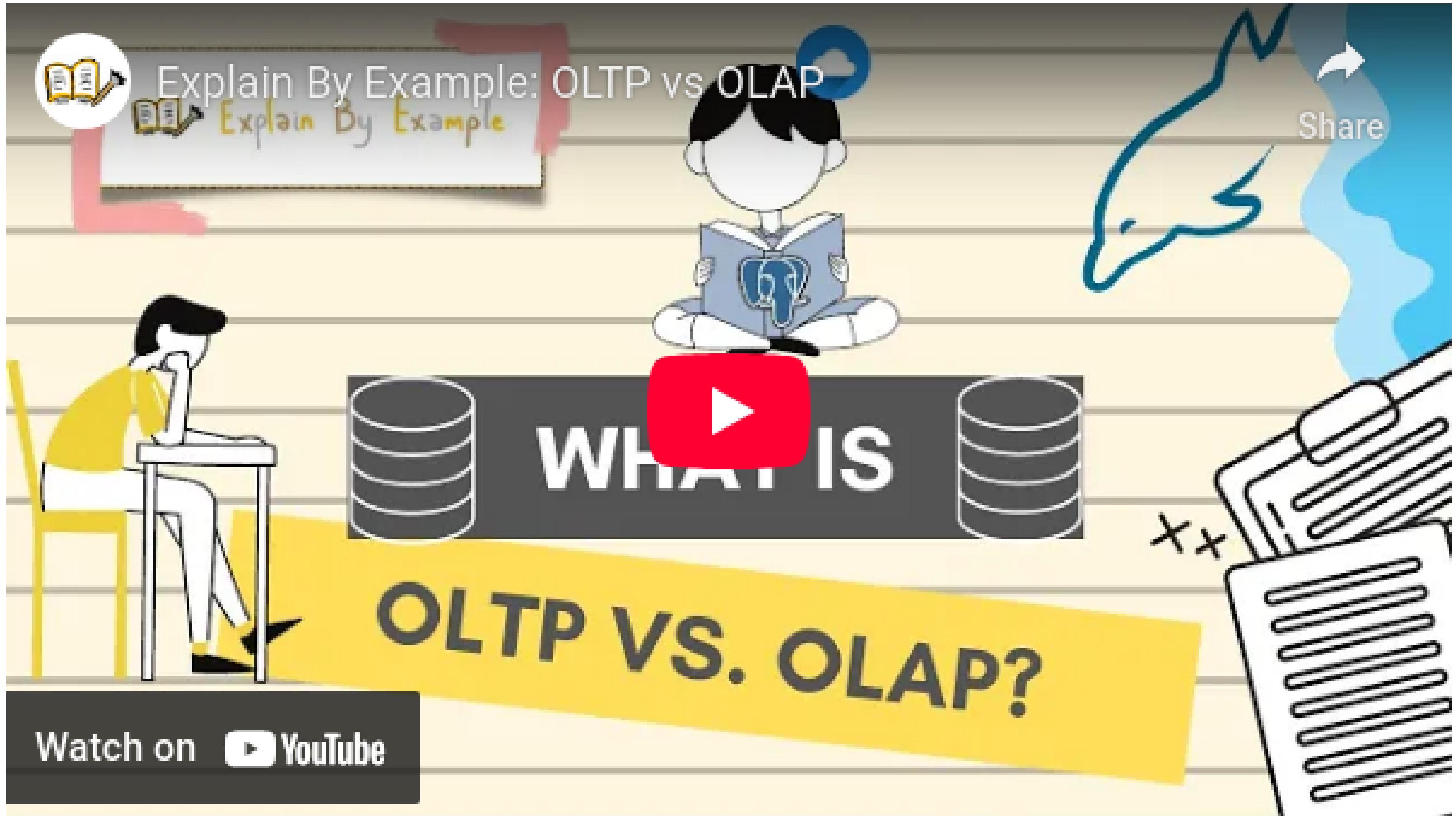
El OLAP combina y agrupa los datos para que pueda analizarlos desde diferentes puntos de vista. Por el contrario, el OLTP almacena y actualiza los datos transaccionales de manera confiable y eficiente en grandes volúmenes. Las bases de datos de OLTP pueden ser uno de los diferentes orígenes de datos de un sistema de OLAP.

El propósito principal del procesamiento analítico en línea (OLAP) es analizar los datos agregados, mientras que el propósito principal del procesamiento de transacciones en línea (OLTP) es procesar las transacciones de bases de datos.

Los sistemas de OLAP se utilizan para generar informes, realizar análisis de datos complejos e identificar tendencias. Por el contrario, los sistemas de OLTP se utilizan para procesar pedidos, actualizar el inventario y administrar las cuentas de los clientes.

Otras diferencias entre OLTP Y OLAP son:

- Formato de datos
- Arquitectura de datos
- Rendimiento
- Requisitos



# Bases de datos, DWH y Data Lake



# Tarea 3

- Importar y cargar los datos del archivo csv.
- Realizar una limpieza y preparación básica de los datos si es necesario.
- Realizar un resumen estadístico de los datos.
- Realizar una matriz de correlación.
- Realizar al menos 3 gráficos.
- Indicar al menos 3 conclusiones de información importante observada de los datos.
- **Observaciones:**
  - Puede utilizar cualquier herramienta para realizar la tarea (Excel, Python, R)
  - Fecha de entrega sábado 08 de marzo a las 23:59 horas.
- **Entregable:**
  - Código fuente (o funciones de Excel utilizadas)
  - Insights o información importante obtenida.
  - **Entregar todo en un único documento PDF**

<b>Nombre de la actividad:</b>	Evaluación tutor
<b>Cantidad de participantes:</b>	28
<b>Doy fe que esta actividad está planificada en dtt (Sí/No):</b>	Sí

<b>Hora de inicio:</b>	08:34
<b>Hora de fin:</b>	08:44
<b>Duración (min):</b>	10

### Participantes: llenar las siguientes cajas de texto (tomar información del chat del meet)

202003892 - Diego Arriaga  
 201807169 - Abdul Chacon  
 202113553 - Kevin García  
 201904016 - Cesar Martínez  
 202105561 - Jeser Rodas  
 202002793 - Hugo Martínez  
 201901756 - Andrés López  
 201902502 - Carlos Soto  
 201314564 - Ronald Ordóñez  
 201800992-Kevin Pozuelos  
 202004770 - Ronaldo Posadas  
 201900956 - Rivaldo Tojín  
 201549059 - Alexander Lux  
 201709282 - Carlos Javier Martinez Polanco

202003894 - Estuardo Son  
 201901604 - Belen Contreras  
 202000424 - Jonatan Garcia  
 201900619 - Yonathan Hernandez  
 201612511 - Luis Garcia  
 201900716 - Kelly Herrera  
 201801460 -Zenaida Chacón  
 201907483-David Eduardo López Morales  
 202012039 - Angel Sique  
 202103718 - Harry Gómez  
 202003959 - Angel Marroquín  
 202006629 - Wilber Zúñiga  
 201902698-Pablo Batz  
 201404202 - Khristian Garcia

--