# Proyecto de Minería de Datos: Aplicación del Modelo KDD y Proceso ETL con Pandas y SQL

## Objetivos:

- **Aplicar el modelo KDD (Knowledge Discovery in Databases) para la extracción de conocimiento a partir de datos.**
- **Realizar el proceso ETL (Extract, Transform, Load) utilizando la librería pandas de Python.**
- **Estructurar los datos en niveles Bronze, Silver y Gold en una base de datos SQL para su análisis en Power BI.**

## Ejercicios Prácticos:

### 1. Selección de Datos:

- **Descargar un conjunto de datos público (por ejemplo, datos de ventas de una tienda en línea).**
- **Describir las características del conjunto de datos seleccionado (columnas, tipos de datos, tamaño, etc.).**

### 2. Inserción en SQL (Bronze Layer):

- **Crear una tabla en SQL llamada Bronze donde se insertarán los datos en su forma original.**
- **Insertar los datos extraídos en la tabla Bronze sin realizar ninguna transformación.**

### 3. Preprocesamiento y Limpieza de Datos (Silver Layer):

- **Extraer los datos desde la tabla Bronze en SQL a un DataFrame de pandas.**
- **Identificar y manejar valores faltantes.**

- **Detectar y corregir datos duplicados.**
- **Realizar un análisis exploratorio de datos (EDA) para entender la distribución y relación entre variables.**
- **Guardar los datos limpios y transformados en una nueva tabla en SQL llamada Silver.**

### 4. Transformación de Datos (Gold Layer):

- **Crear nuevas variables derivadas a partir de las existentes.**
- **Estandarizar y normalizar las variables numéricas.**
- **Codificar variables categóricas (one-hot encoding).**
- **Agrupar y resumir los datos según criterios específicos (ejemplo: ventas por mes, categoría de producto, etc.).**
- **Guardar los datos finales en una tabla en SQL llamada Gold, optimizada para análisis en Power BI.**

### 5. Minería de Datos:

- **Aplicar técnicas de minería de datos, como clustering, clasificación o regresión, para descubrir patrones y obtener insights.**
- **Utilizar bibliotecas como scikit-learn para implementar los modelos de minería de datos.**

### 6. Evaluación e Interpretación:

- **Evaluar el rendimiento de los modelos utilizando métricas adecuadas (precisión, recall, F1-score, etc.).**
- **Interpretar los resultados y extraer conclusiones útiles para la toma de decisiones.**

# Instrucciones Detalladas:

### Configuración Inicial:

- **Asegurarse de tener instalado Python, pandas, numpy, scikit-learn y una base de datos SQL.**
- **Descargar el conjunto de datos desde una fuente confiable (ejemplo: Kaggle, UCI Machine Learning Repository).**

### Evaluación:

**✅ Informe Final: Presentar un informe detallado que incluya la descripción del conjunto de datos, los pasos del proceso ETL, el análisis de minería de datos y las conclusiones obtenidas.**
**✅ Presentación Oral: Explicar cada etapa del proceso y justificar las decisiones tomadas.**

## II) Visualización con Power BI

**Link de uno de los documentos:**
https://docs.google.com/spreadsheets/d/1PaHUYlpdg1RiEf09nbtdFLS1B9aT8cFK/edit?usp=sharing&ouid=112367020996334027708&rtpof=true&sd=true

Estos ejercicios te ayudarán a familiarizarte con las funciones básicas y avanzadas de Power BI:

1. Importación de Datos:
   - Importa un conjunto de datos desde un archivo Excel o CSV a Power BI.
   - Explora las opciones de carga, como la carga completa o la carga de datos condicional.

2. Transformación de Datos:
   - Realiza transformaciones básicas, como cambiar tipos de datos, renombrar columnas y eliminar valores nulos.
   - Explora las opciones de transformación más avanzadas, como la agrupación y la fusión de consultas.

3. Creación de Relaciones:
   - Importa dos conjuntos de datos y crea una relación entre ellos.
   - Utiliza las relaciones para crear visualizaciones que involucren datos de ambas tablas.

4. Visualizaciones Básicas:
   - Crea un gráfico de barras para mostrar la distribución de una variable.
   - Utiliza un gráfico de líneas para representar tendencias a lo largo del tiempo.

5. Visualizaciones Avanzadas:
   - Crea una matriz o tabla dinámica que resuma datos de manera efectiva.
   - Experimenta con visualizaciones geoespaciales, como mapas de calor o mapas de burbujas.

6. Creación de Medidas (DAX):
   - Crea medidas simples, como suma o promedio.
   - Desarrolla medidas más complejas utilizando funciones DAX para calcular tasas de crecimiento, acumulados, etc.
- Crea medidas o columnas donde filtres por una categoría o varias categorías.

7. Paneles y Dashboards:
   - Agrupa visualizaciones en un panel para crear un dashboard.
   - Agrega interactividad a tu dashboard mediante la creación de filtros y destaca las visualizaciones clave.

Este ejercicio será evaluado en base a la estética con la que presentas la información y las informaciones relevantes que muestren a partir de los gráficos aplicados.