# Portada.

# **EVIDENCIA DE APRENDIZAJE #1**

**CREACIÓN DE UNA BASE DE DATOS ANALÍTICA**

**INTEGRANTES GRUPO #13**

**JOSE LUIS BENITEZ SALCEDO**

**GERMAN EDUARDO PAIBA ROCHA**

**INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA**

**INGENIERÍA DE SOFTWARE Y DATOS**

**ASIGNATURA:**

**BIG DATA**

**NOMBRE DEL TUTOR:**

**ANDRES FELIPE CALLEJAS**

**MAYO 11 DE 2025**

Contenido

[Portada. 1](#_Toc197869757)

[Evidencia de aprendizaje #1 1](#_Toc197869758)

[Introducción. 3](#_Toc197869759)

[Descripción del problema. 3](#_Toc197869760)

[Objetivos. 5](#_Toc197869761)

[**Objetivo General:** 5](#_Toc197869762)

[**Objetivos Específicos:** 5](#_Toc197869763)

[Descripción de los datos disponibles. 6](#_Toc197869764)

[Solución propuesta (Elección del SGBD y esquema diseñado). 7](#_Toc197869765)

[**Elección del SGBD:** 7](#_Toc197869766)

[**Esquema Diseñado:** 8](#_Toc197869767)

[Metodología empleada. 9](#_Toc197869768)

[Resultados y Conclusiones. 10](#_Toc197869769)

[**Resultados Esperados:** 11](#_Toc197869770)

[**Conclusiones Iniciales:** 11](#_Toc197869771)

[Bibliografía 13](#_Toc197869772)

# Introducción.

En el dinámico y a menudo volátil mercado de metales preciosos, el oro se erige como un activo de refugio y una inversión estratégica. Para las empresas que buscan incursionar o expandir sus operaciones en este sector, contar con información precisa, oportuna y centralizada es fundamental para la toma de decisiones informadas. Este documento presenta el diseño y la documentación de un sistema de gestión de base de datos (SGBD) concebido para abordar las necesidades de una empresa en Medellín, Antioquia, en el año 2025, interesada en el análisis y la gestión de datos históricos de precios del oro. El sistema propuesto permitirá consolidar información dispersa, facilitar el análisis de tendencias y, en última instancia, optimizar las estrategias de inversión.

# Descripción del problema.

En el pujante entorno empresarial de Medellín, una compañía visionaria está considerando una inversión estratégica en el mercado del oro. Sin embargo, se enfrenta a un desafío significativo: la falta de un sistema centralizado para la recopilación, el almacenamiento y el análisis de los datos históricos de precios del oro. Actualmente, la información relevante se encuentra dispersa en diversas fuentes, posiblemente en hojas de cálculo, informes web y otras plataformas no integradas.

Esta dispersión de la información conlleva varios problemas críticos:

* **Ineficiencia en la recopilación de datos:** El tiempo y los recursos dedicados a buscar y consolidar datos de diferentes fuentes son considerables, lo que resta eficiencia a los procesos de análisis.
* **Riesgo de inconsistencia e inexactitud:** La manipulación manual de datos y la falta de una fuente única de verdad aumentan la probabilidad de errores e inconsistencias, lo que puede llevar a análisis defectuosos.
* **Dificultad para el análisis de tendencias:** Sin una base de datos estructurada, resulta complejo realizar análisis comparativos a lo largo del tiempo, identificar patrones o predecir tendencias futuras del precio del oro.
* **Limitaciones en la toma de decisiones:** La falta de información consolidada y analizada dificulta la formulación de estrategias de inversión sólidas y oportunas, exponiendo a la empresa a mayores riesgos y oportunidades perdidas.
* **Escalabilidad limitada:** A medida que la empresa crece y la cantidad de datos aumenta, la gestión manual se vuelve cada vez más insostenible y limita la capacidad de adaptación a nuevas necesidades de análisis.

En el contexto específico de Medellín en 2025, una ciudad con un creciente interés en la inversión y el comercio, contar con un sistema robusto de gestión de datos se vuelve crucial para que esta empresa pueda competir eficazmente y tomar decisiones estratégicas basadas en información confiable.

# Objetivos.

## **Objetivo General:**

Desarrollar e implementar un sistema de gestión de base de datos eficiente y robusto para la centralización, el almacenamiento y el análisis de datos históricos de precios del oro, permitiendo a la empresa en Medellín tomar decisiones de inversión informadas y estratégicas.

## **Objetivos Específicos:**

1. **Definir el modelo de datos:** Analizar la estructura de la tabla HTML proporcionada para identificar las entidades relevantes y sus atributos, estableciendo las relaciones necesarias para representar la información de manera coherente en un modelo de datos lógico.
2. **Seleccionar un Sistema de Gestión de Bases de Datos (SGBD):** Evaluar diferentes SGBD disponibles, considerando factores como la escalabilidad, la facilidad de uso, la compatibilidad con posibles herramientas de análisis y la adecuación a la naturaleza de los datos.
3. **Diseñar el esquema de la base de datos:** Traducir el modelo de datos lógico en un esquema físico específico para el SGBD seleccionado, definiendo las tablas, los campos, los tipos de datos, las claves primarias y foráneas, y las restricciones necesarias para garantizar la integridad de los datos.
4. **Implementar la base de datos:** Crear la base de datos y las tablas definidas en el esquema, e importar los datos desde la tabla HTML al sistema de gestión de base de datos.
5. **Documentar el proceso:** Registrar detalladamente cada etapa del desarrollo, desde el análisis de los datos hasta la implementación del sistema, incluyendo el modelo de datos, el esquema de la base de datos y las consideraciones técnicas relevantes.

# Descripción de los datos disponibles.

Los datos disponibles se presentan en una tabla HTML, que representa información histórica de precios del oro. Al analizar la estructura de la tabla, se identifican las siguientes columnas, que constituirán los atributos principales de nuestro modelo de datos:

* **Fecha:** Representa la fecha específica en la que se registraron los precios del oro. Este campo es crucial para el análisis de series de tiempo y la identificación de tendencias. El formato esperado es día mes año (ej: 9 may 2025).
* **Abrir:** Indica el precio del oro al inicio de la jornada de negociación en la fecha especificada. Se espera un valor numérico decimal.
* **Máx.:** Representa el precio máximo alcanzado por el oro durante la jornada de negociación en la fecha especificada. Se espera un valor numérico decimal.
* **Mín.:** Indica el precio mínimo alcanzado por el oro durante la jornada de negociación en la fecha especificada. Se espera un valor numérico decimal.
* **Cerrar:** Representa el precio del oro al final de la jornada de negociación en la fecha especificada. Se espera un valor numérico decimal.
* **Cierre ajustado:** Precio de cierre ajustado para reflejar divisiones de acciones (splits) y distribuciones de dividendos o plusvalías. Este valor es importante para un análisis histórico preciso que tenga en cuenta eventos corporativos. Se espera un valor numérico decimal.
* **Volumen:** Indica la cantidad de contratos o la actividad de negociación del oro en la fecha especificada. Se espera un valor numérico entero.

Estos datos históricos proporcionan una visión del comportamiento del precio del oro a lo largo del tiempo y son fundamentales para realizar análisis técnicos, identificar volatilidades y undamentar decisiones de inversión. La expectativa es que estos datos se proporcionen en un archivo .csv para su importación al sistema de gestión de base de datos.

# Solución propuesta (Elección del SGBD y esquema diseñado).

## **Elección del SGBD:**

Considerando la naturaleza de los datos (series de tiempo), la necesidad de realizar consultas analíticas y la posible escalabilidad futura, se propone la utilización de **PostgreSQL** como Sistema de Gestión de Base de Datos (SGBD). PostgreSQL es un SGBD relacional de código abierto, robusto, escalable y con potentes capacidades para el manejo de datos complejos y el análisis avanzado a través de extensiones. Su confiabilidad, su gran comunidad de soporte y sus características avanzadas lo convierten en una excelente opción para este proyecto.

## **Esquema Diseñado:**

Se diseñará una única tabla para almacenar los datos históricos de precios del oro. Esta tabla contendrá las siguientes columnas, mapeadas a los datos disponibles:

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de la Columna** | **Tipo de Dato** | **Restricciones** | **Descripción** |
| fecha | DATE | PRIMARY KEY | Fecha de la cotización del oro |
| abrir | NUMERIC(10, 2) | NOT NULL | Precio de apertura en la fecha |
| maximo | NUMERIC(10, 2) | NOT NULL | Precio máximo alcanzado en la fecha |
| minimo | NUMERIC(10, 2) | NOT NULL | Precio mínimo alcanzado en la fecha |
| cerrar | NUMERIC(10, 2) | NOT NULL | Precio de cierre en la fecha |
| cierre\_ajustado | NUMERIC(10, 2) | NOT NULL | Precio de cierre ajustado por splits y dividendos |
| volumen | INTEGER | NOT NULL | Volumen de negociación en la fecha |
| fecha\_carga | TIMESTAMP WITH TIME ZONE | NOT NULL DEFAULT CURRENT\_TIMESTAMP | Fecha y hora en que se cargó el registro en la base de datos (para auditoría) |

**Relaciones:**

En este modelo inicial, no se identifican relaciones con otras tablas, ya que el enfoque principal es el almacenamiento y análisis de los datos históricos de precios del oro. Sin embargo, en futuras expansiones del sistema, se podrían considerar tablas relacionadas para almacenar información sobre diferentes mercados, tipos de oro, o incluso análisis y predicciones generadas.

# Metodología empleada.

La metodología empleada se basa en los objetivos específicos definidos previamente, siguiendo un enfoque secuencial y lógico:

1. **Definición del modelo de datos:** Se realizó un análisis exhaustivo de la estructura de la tabla HTML proporcionada para identificar las entidades (en este caso, la cotización diaria del oro) y sus atributos relevantes (fecha, precios de apertura, máximo, mínimo, cierre, cierre ajustado y volumen). Se definieron los tipos de datos apropiados para cada atributo.
2. **Selección del SGBD:** Se llevó a cabo una evaluación de diferentes SGBD, considerando las características de los datos, los requerimientos de análisis y la escalabilidad. La elección de PostgreSQL se fundamentó en su robustez, capacidades analíticas y su naturaleza de código abierto.
3. **Diseño del esquema de la base de datos:** Se diseñó un esquema relacional simple con una única tabla (precios\_oro) que contiene todas las columnas identificadas en el modelo de datos. Se definieron los tipos de datos, las restricciones de integridad (como NOT NULL y PRIMARY KEY), y se añadió una columna para el seguimiento de la fecha de carga.
4. **Implementación de la base de datos:** Se procederá a la creación de la base de datos y la tabla precios\_oro en PostgreSQL utilizando el lenguaje SQL (Structured Query Language). Se definirán los tipos de datos y las restricciones especificadas en el esquema.
5. **Desarrollo del proceso de carga de datos:** Se implementará un script (probablemente en Python utilizando bibliotecas como pandas y psycopg2) para leer los datos del archivo CSV (que simula la información de la tabla HTML), transformar los datos si es necesario (por ejemplo, convertir el formato de fecha), y cargarlos de manera eficiente en la tabla precios\_oro de la base de datos PostgreSQL.

# Resultados y Conclusiones.

## **Resultados Esperados:**

* Un modelo de datos claro y bien definido para la representación de los precios históricos del oro.
* Una base de datos PostgreSQL implementada con una tabla (precios\_oro) que almacena los datos de manera estructurada y eficiente.
* Un proceso automatizado (a través de un script) para la carga de datos desde un archivo CSV a la base de datos.
* Una plataforma centralizada que permite a la empresa en Medellín acceder, consultar y analizar los datos históricos de precios del oro de manera sencilla y confiable.

## **Conclusiones Iniciales:**

La implementación de este sistema de gestión de base de datos proporcionará a la empresa en Medellín una herramienta fundamental para mejorar su proceso de toma de decisiones en relación con las inversiones en oro. Al centralizar la información, se reducirá la ineficiencia en la recopilación de datos y se minimizará el riesgo de errores. La estructura de la base de datos facilitará el análisis de tendencias históricas, la identificación de patrones y la generación de informes relevantes para la estrategia de inversión.

La elección de PostgreSQL como SGBD ofrece una base sólida para el crecimiento futuro del sistema, permitiendo la incorporación de nuevas funcionalidades y la integración con otras herramientas de análisis y visualización de datos. Este proyecto representa un paso importante hacia una gestión de la información más estratégica y basada en datos para la empresa en el competitivo mercado del oro en Medellín.

# Bibliografía

fundation, P. s. (2025). *Python.org*. Obtenido de Python: https://www.python.org/

*Githut*. (11 de 05 de 2025). Obtenido de https://github.com/josebenitezs/bigdata\_2025\_1\_2

Yahoo.com. (11 de 05 de 2025). *yahoo.finance*. Obtenido de https://es.finance.yahoo.com/quote/GC%3DF/history/

URL GITHUB:

<https://github.com/josebenitezs/bigdata_2025_1_2.git>