

**UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA**

**MÁSTER EN INGENIERÍA Y CIENCIA DE DATOS**

**Análisis Predictivo del Precio del Oro:**

**Correlaciones con Indicadores Técnicos y Fundamentales en el Contexto de Mercados Financieros**

FERNANDO DEVÍS RODRÍGUEZ

Dirigido por: FERNANDO LÓPEZ OSTENERO

Curso 2023-2024: Convocatoria de Junio

Tabla de contenido

[Resumen 6](#_Toc168293014)

[Abstract 6](#_Toc168293015)

[CAPITULO 1: El valor del oro y contexto histórico 8](#_Toc168293016)

[¿Por qué el oro? 8](#_Toc168293017)

[El oro en los mercados 9](#_Toc168293018)

[CAPITULO 2: Análisis de datos y estados financieros 11](#_Toc168293019)

[El precio de un activo. El valor del oro. 11](#_Toc168293020)

[Indicadores fundamentales: 11](#_Toc168293021)

[ Oferta y demanda 11](#_Toc168293022)

[ Coyuntura económica: 12](#_Toc168293023)

[ Políticas monetarias: 12](#_Toc168293024)

[ Factores geopolíticos: 12](#_Toc168293025)

[ Valor del dólar: 12](#_Toc168293026)

[ Expectativas del mercado y sentimiento inversor: 12](#_Toc168293027)

[Indicadores para el análisis (técnico) 13](#_Toc168293028)

[1. Media Móvil (MA) 13](#_Toc168293029)

[2. Índice de Fuerza Relativa (RSI) 13](#_Toc168293030)

[3. Bandas de Bollinger 13](#_Toc168293031)

[4. Convergencia/Divergencia de Medias Móviles (MACD) 13](#_Toc168293032)

[5. Oscilador Estocástico 13](#_Toc168293033)

[6. Retrocesos de Fibonacci 13](#_Toc168293034)

[7. Índice de Movimiento Direccional (DMI) y el Índice Direccional Promedio (ADX) 13](#_Toc168293035)

[8. Volumen 14](#_Toc168293036)

[9. Indicador de Acumulación/Distribución (A/D) 14](#_Toc168293037)

[10. Índice de Flujo de Dinero (MFI) 14](#_Toc168293038)

[Técnicos vs Fundamentales 14](#_Toc168293039)

[DESCARGA DE LOS DATOS 16](#_Toc168293040)

[DATASETS OBTENIDOS 19](#_Toc168293041)

[CAPITULO 2: El escenario actual del análisis financiero 20](#_Toc168293042)

[Modelos y situación actual del oro. Regresión lineal 20](#_Toc168293043)

[El dataset “GC=F”, de Yahoo Finance. 22](#_Toc168293044)

[Las variables explicativas (MA 3 y 9). Justificación de su uso 24](#_Toc168293045)

[La variable objetivo “next day price” 26](#_Toc168293046)

[División de los conjuntos de entrenamiento y prueba 26](#_Toc168293047)

[Modelo de Regresión Lineal: 26](#_Toc168293048)

[Resultados y evaluación del modelo: 27](#_Toc168293049)

[Resumen: 30](#_Toc168293050)

[Avanzado en la interpretación del análisis: estrategias de inversión con medias móviles 30](#_Toc168293051)

[Conclusiones y limitaciones del modelo 32](#_Toc168293052)

[**Horizonte Temporal Óptimo para Invertir en Oro** 33](#_Toc168293053)

[**Mejoras con Redes Neuronales LSTM** 33](#_Toc168293054)

[**Resumen** 34](#_Toc168293055)

[CAPITULO 3: Análisis predictivo con Redes Neuronales LSTM 34](#_Toc168293056)

[Introducción: 34](#_Toc168293057)

[La importancia de los datos 34](#_Toc168293058)

[Redes Neuronales LSTM 35](#_Toc168293059)

[El proceso ETL y la construcción de nuestro dataset completo 36](#_Toc168293060)

[Extracción (Extract): 36](#_Toc168293061)

[Transformación (Transform): 37](#_Toc168293062)

[Carga: 37](#_Toc168293063)

[GoldStock.csv 38](#_Toc168293064)

[Dataset GLD de Yahoo Finance. 45](#_Toc168293065)

[Stock Market Dataset 2019 – 2024 48](#_Toc168293066)

[XAUSD: Los primeros indicadores. RSI 14 y SMA14 54](#_Toc168293067)

[El MACD (Moving Average Convergence / Divergence) 62](#_Toc168293068)

[Los tipos de interés y la FED. Análisis Fundamental 65](#_Toc168293069)

[Fusionando el dataset completo. 68](#_Toc168293070)

[Implementación de la Red Neuronal LSTM 73](#_Toc168293071)

[Preparación de los datos 73](#_Toc168293072)

[Resultados obtenidos y comparativa 73](#_Toc168293073)

[Conclusiones 78](#_Toc168293074)

[Bibliografía 79](#_Toc168293075)

[Anexo: instalación y despliegue. 79](#_Toc168293076)

[Disponibilidad del código fuente 79](#_Toc168293077)

[Uso y mantenimiento 79](#_Toc168293078)

TABLA DE FIGURAS:

[Figura 1 - Uso del oro como moneda a lo largo de la historia 9](#_Toc168289753)

[Figura 2: Investing.com Precio del Oro al contado 17](#_Toc168289754)

[Figura 3: Kaggle, apartado DATASETS 18](#_Toc168289755)

[Figura 4: Tipos de interés FED -Banco de la Reserva Federal - 19](#_Toc168289756)

[Figura 5: Histogramas dataset Oro Yahoo Finance 24](#_Toc168289757)

[Figura 6: Evolución del precio del oro 24](#_Toc168289758)

[Figura 7 - Resultado previsión modelo RL 28](#_Toc168289759)[Gráfico, Gráfico de líneas

Descripción generada automáticamenteFigura 8- Estudio de medias móviles 3 y 9 días 31](#_Toc168289760)

[Figura 9: A CONVENIR 36](#_Toc168289761)

[Figura 10: Muestra XML DBLP 36](#_Toc168289762)

# Resumen

Este trabajo de fin de máster titulado "**Análisis Predictivo del Precio del Oro: Correlaciones con Indicadores Técnicos y Fundamentales en el Contexto de Mercados Financieros**" abarca un estudio detallado del precio del oro utilizando el conocimiento adquirido de todo el máster. El proyecto se estructura en varios capítulos, comenzando con un análisis histórico del valor del oro y su relevancia en el contexto económico actual.

En la fase de recopilación de datos, se examinan diversos indicadores fundamentales como la oferta y demanda, políticas monetarias y factores geopolíticos, así como indicadores técnicos como la media móvil, el RSI, y el MACD. Estos datos se descargan y organizan para su análisis posterior, mediante procesos ETL.

El siguiente paso es la construcción de un modelo sencillo de regresión lineal utilizando datos históricos del oro obtenidos de Yahoo Finance. Se evalúan las variables explicativas y se prueba la eficacia del modelo para predecir el precio del oro del día siguiente. A pesar de sus limitaciones, este modelo proporciona una base para comparaciones posteriores.

Posteriormente, se implementa un modelo avanzado utilizando Redes Neuronales LSTM, que permite capturar patrones más complejos en los datos. El proceso ETL (Extract, Transform, Load) se describe en detalle, y se realiza una fusión de varios conjuntos de datos para construir un dataset completo. Los resultados de este modelo se comparan con los obtenidos mediante regresión lineal, destacando las mejoras y la precisión alcanzada. Finalmente, se presentan las conclusiones del estudio, resaltando las limitaciones encontradas y las posibles mejoras futuras.

**PALABRAS CLAVE:** Análisis predictivo, modelado de series temporales, LSTM, redes neuronales recurrentes, regresión lineal, minería de datos, indicadores financieros, media móvil simple, índice de fuerza relativa, convergencia/divergencia de medias móviles, bandas de Bollinger, oscilador estocástico, retrocesos de Fibonacci, índice de movimiento direccional, volumen de trading, indicador de acumulación/distribución, índice de flujo de dinero, política monetaria, factores geopolíticos, oferta y demanda, coyuntura económica, valor del dólar, expectativas del mercado, sentimiento inversor, procesos ETL, extracción de datos, transformación de datos, carga de datos, dataset de Yahoo Finance, evaluación de modelos, precisión predictiva, análisis de datos financieros

# Abstract

This master's thesis titled "Predictive Analysis of Gold Prices: Correlations with Technical and Fundamental Indicators in the Context of Financial Markets" encompasses a detailed study of gold prices using the knowledge acquired throughout the master's program. The project is structured into several chapters, starting with a historical analysis of the value of gold and its relevance in the current economic context.

In the data collection phase, various fundamental indicators such as supply and demand, monetary policies, and geopolitical factors are examined, along with technical indicators like moving averages, RSI, and MACD. These data are downloaded and organized for subsequent analysis through ETL processes.

The next step involves constructing a simple linear regression model using historical gold data obtained from Yahoo Finance. Explanatory variables are evaluated, and the model's effectiveness in predicting the next day's gold price is tested. Despite its limitations, this model provides a foundation for subsequent comparisons.

Subsequently, an advanced model using LSTM Neural Networks is implemented, allowing for the capture of more complex patterns in the data. The ETL (Extract, Transform, Load) process is described in detail, and multiple datasets are merged to construct a complete dataset. The results of this model are compared with those obtained through linear regression, highlighting the improvements and accuracy achieved.

Finally, the study's conclusions are presented, highlighting the limitations encountered and potential future improvements.

**KeyWords:**

Predictive analysis, time series modeling, LSTM, recurrent neural networks, linear regression, data mining, financial indicators, simple moving average, relative strength index, moving average convergence/divergence, Bollinger bands, stochastic oscillator, Fibonacci retracements, directional movement index, trading volume, accumulation/distribution indicator, money flow index, monetary policy, geopolitical factors, supply and demand, economic conditions, dollar value, market expectations, investor sentiment, ETL processes, data extraction, data transformation, data loading, Yahoo Finance dataset, model evaluation, predictive accuracy, financial data analysis.

# CAPITULO 1: El valor del oro y contexto histórico

Uno de los deseos más perseguidos de la humanidad, ha sido la riqueza. Independientemente de la época, situación, perfil o trasfondo., cualquier persona, de una forma u otra, necesita poder adquisitivo. Unos tal vez, no consiguen ni el necesario para vivir, otros lo codician de forma enfermiza, otros incluso creen que no es tan importante, incluso en algunos ámbitos religiosos o culturales puede llegar a tener hasta una connotación negativa

El propósito de este trabajo no es, perseguir un objetivo especulativo, ni mucho menos pretende juzgar el uso del dinero, sin embargo, el concepto de “dinero”, en multitud de sus diferentes formas y variantes es perfecto para estudiarlo como dato, y ¿acaso no estamos estudiando Ciencia de Datos?

Vamos a tratar una de las formas de dinero primigenio, el oro, y estudiaremos su precio atendiendo a multitud de factores que afectan a su variabilidad. Estos factores suelen tener una división principal: **FUNDAMENTALES y TÉCNICOS** (Weinstein, 1988) (como veremos más adelante, al definir el análisis técnico y el análisis fundamental)

## ¿Por qué el oro?

El oro ha jugado un papel fundamental en la historia económica de la humanidad, siendo el **primer metal precioso utilizado como dinero**. (Ammous, 2018) Su descubrimiento y uso datan de tiempos prehistóricos, cuando las primeras civilizaciones empezaron a valorar este metal por su belleza, durabilidad y maleabilidad. La capacidad del oro para no oxidarse o corroerse lo convertía en un medio ideal para **preservar la riqueza** a través del tiempo y del espacio.

En diversas culturas, el oro comenzó a utilizarse no solo como adorno o símbolo de poder y divinidad, sino también como **moneda de cambio**. Los egipcios, por ejemplo, acuñaron las primeras piezas de oro que funcionaron como dinero hacia el año 3400 a.C. Su aceptación como medio de intercambio fue casi universal debido a que podía ser fácilmente dividido en unidades más pequeñas sin perder su valor por unidad de peso. (Davies, 2002)

La universalidad del oro como moneda facilitó el comercio entre diferentes regiones y culturas, estableciendo un sistema económico que permitía comparar y valorar bienes y servicios a escala global. Así, el oro no solo fue un precursor del dinero como lo conocemos hoy, sino que también fue un agente fundamental en la creación de un mercado mundial interconectado.

## El oro en los mercados

El oro, desde su inicial utilización como moneda en las antiguas civilizaciones, ha evolucionado considerablemente en su función económica, transformándose en un bien fundamental en los mercados financieros modernos. A medida que los sistemas monetarios evolucionaron, el oro dejó de usarse como **moneda circulante**, pero su valor intrínseco y su rareza continuaron haciéndolo un **refugio seguro para la inversión** y la preservación de la riqueza.

Imagen que contiene foto, tabla, llenado, montón

Descripción generada automáticamente

Figura 1 - Uso del oro como moneda a lo largo de la historia

En el siglo XX, especialmente después de la **disolución del patrón oro[[1]](#footnote-1)**, que vinculaba directamente el valor de la moneda a reservas específicas de oro, este metal precioso encontró una nueva vida en los mercados cotizados. Las bolsas de valores y los mercados de **commodities** comenzaron a incluir el oro como un activo financiero más, sujeto a compra y venta, no diferente de acciones o bonos.

Hoy en día, el oro se comercializa en diversas formas en los mercados financieros. Los inversores y especuladores pueden operar con oro físico, en forma de lingotes o monedas, o mediante instrumentos financieros derivados como futuros, opciones y fondos cotizados en bolsa (ETFs) que replican el precio del oro. Esta versatilidad ha hecho del oro una herramienta esencial para la gestión de riesgos y la diversificación de carteras, permitiendo a los inversores protegerse contra la inflación, la volatilidad del mercado y las incertidumbres económicas.

Así, el oro continúa siendo no solo un símbolo de riqueza y seguridad, sino también un activo dinámico en la economía global, capaz de ser negociado y valorado en tiempo real a través de mercados financieros internacionales.

# CAPITULO 2: Análisis de datos y estados financieros

## El precio de un activo. El valor del oro.

Para entender en qué se basa el precio de un activo, como el oro, es importante considerar varios factores fundamentales que influyen en su valoración en los mercados financieros.

Estos mercados (los financieros) son plataformas donde se compran y venden activos financieros, como acciones, bonos, divisas y **commodities[[2]](#footnote-2)**, como el oro en nuestro caso. La interacción entre compradores y vendedores, crea así el flujo de capital y de recursos a través de la economía, y permiten un mecanismo para que los inversores inviertan su dinero, para que las empresas capten recursos para expandirse y para que los gobiernos financien sus actividades, contribuyendo al crecimiento económico y a la estabilidad financiera.

## Indicadores fundamentales:

El precio del oro, al igual que otros activos, se determina por una combinación de **oferta y demanda**, en estos mercados financieros, así como por factores macroeconómicos y geopolíticos.

* Oferta y demanda: Es el factor más importante para determinar el precio. La oferta del oro proviene principalmente de la minería y el reciclaje de productos de oro, mientras que la demanda proviene de su uso en joyería, tecnología y **como inversión** fundamentalmente. Cabe destacar que el oro, pese a que es un bien que no se agota (no como BTC, que sí que es finito), su extracción puede ser cada vez más costosa, lo que influye también en el equilibrio de la oferta y la demanda. Un bien que es fácil de producir es mucho más económico que uno escaso. No obstante, debemos recalcar que el principal uso del oro, como hemos comentado anteriormente es como moneda de cambio y reserva de valor
* Coyuntura económica: El oro suele considerarse como una inversión segura, y más durante tiempos de incertidumbre económica o crisis financiera. En períodos de inflación alta o cuando las monedas pierden valor, los inversores tienden a *refugiarse* en activos tangibles como el oro, lo que puede incrementar su demanda y, por tanto, su precio.
* Políticas monetarias: Las políticas implementadas por los bancos centrales, especialmente en lo que respecta a los *tipos de interés* y a la *impresión de dinero,* también pueden influir en el precio del oro. Un entorno de bajos tipos de interés puede hacer que el oro sea más atractivo, ya que reduce el *costo de oportunidad* de mantener oro en comparación con otros activos que generan intereses
* Factores geopolíticos: Los conflictos, las tensiones internacionales y la inestabilidad política pueden llevar a un aumento en la demanda de oro como 'activo refugio'. Los inversores buscan seguridad en el oro cuando hay incertidumbre sobre la estabilidad global.
* Valor del dólar: Existe una relación inversa entre el valor del dólar y el precio del oro. Cuando el dólar se debilita, el oro generalmente se vuelve más caro en términos de otras monedas, lo que puede aumentar la demanda y el precio del oro.
* Expectativas del mercado y sentimiento inversor: Las expectativas de los inversores sobre la dirección futura de la economía mundial y los mercados financieros pueden influir en el precio del oro. El sentimiento del mercado puede ser afectado por una variedad de indicadores económicos, reportes de analistas y tendencias del mercado. **Greedy / Fear Index[[3]](#footnote-3)**

## Indicadores para el análisis (técnico)

### 1. Media Móvil (MA)

**Media Móvil Simple (SMA)**: Promedio del precio de cierre durante un período específico. Es útil para identificar la dirección de la tendencia.

**Media Móvil Exponencial (EMA)**: Da más peso a los precios recientes, reaccionando más rápidamente a los cambios de precios que la SMA.

### 2. Índice de Fuerza Relativa (RSI)

El RSI mide la velocidad y el cambio de los movimientos de precios, oscilando entre 0 y 100. Un valor por encima de 70 indica que el activo está sobrecomprado, mientras que un valor por debajo de 30 indica que está sobrevendido.

### 3. Bandas de Bollinger

Consisten en una media móvil y dos desviaciones estándar que se trazan por encima y por debajo de la media móvil. Las Bandas de Bollinger se utilizan para identificar niveles de sobrecompra y sobreventa.

### 4. Convergencia/Divergencia de Medias Móviles (MACD)

Es un indicador de seguimiento de tendencias que muestra la relación entre dos medias móviles de precios. La línea MACD se calcula restando la EMA de 26 días de la EMA de 12 días. Una línea de señal (EMA de 9 días del MACD) se traza sobre el MACD para funcionar como una señal de compra/venta.

### 5. Oscilador Estocástico

Este indicador compara el precio de cierre de un activo con su rango de precios durante un período de tiempo específico. El valor del oscilador oscila entre 0 y 100. Un valor por encima de 80 indica que el activo está sobrecomprado, mientras que un valor por debajo de 20 indica que está sobrevendido.

### 6. Retrocesos de Fibonacci

Basado en los números de Fibonacci, se utilizan para identificar posibles niveles de soporte y resistencia. Los niveles clave son 23.6%, 38.2%, 50%, 61.8%, y 100%.

### 7. Índice de Movimiento Direccional (DMI) y el Índice Direccional Promedio (ADX)

El DMI consta de dos líneas: +DI y -DI, que indican la presión de compra y venta, respectivamente. El ADX mide la fuerza de la tendencia.

### 8. Volumen

El volumen es un indicador crucial que confirma la dirección de la tendencia. Un aumento en el volumen puede indicar una continuación de la tendencia actual, mientras que una disminución puede señalar una posible reversión.

### 9. Indicador de Acumulación/Distribución (A/D)

Este indicador mide el flujo acumulado de volumen de dinero. Ayuda a identificar divergencias entre el precio y el volumen.

### 10. Índice de Flujo de Dinero (MFI)

Similar al RSI, pero incorpora el volumen en su cálculo. Oscila entre 0 y 100, con niveles por encima de 80 indicando sobrecompra y por debajo de 20 indicando sobreventa.

Estos indicadores pueden ser utilizados de manera conjunta para obtener una visión más completa del comportamiento del precio del oro. Obviamente ningún indicador es infalible, no obstante, en nuestro modelo intentaremos elegir aquellos valores que puedan parecer más relevantes, dependiendo del contexto temporal

## Técnicos vs Fundamentales

Acabamos de ver que en el análisis de valores (acciones, o en nuestro caso el precio del oro), se utilizan dos enfoques principales para evaluar y seleccionar inversiones: el análisis técnico y el análisis fundamental. A continuación, se explican las diferencias clave entre estos dos tipos de análisis:

***Análisis Técnico***

El análisis técnico se centra en el estudio de los movimientos de precios y volúmenes de negociación de una acción a través de gráficos y otras herramientas estadísticas. Se basa en la idea de que los movimientos de precios históricos pueden predecir futuros movimientos de precios. Su objetivo es identificar tendencias y patrones de comportamiento del mercado para tomar decisiones de compra o venta a corto plazo.

Supuestos Clave:

* Los precios se mueven en tendencias.
* La historia tiende a repetirse.
* Los precios de mercado reflejan toda la información disponible (aunque este punto es más debatido). (Weinstein, 1988)

***Análisis Fundamental***

El análisis fundamental evalúa el valor intrínseco de una acción mediante el estudio de factores económicos, financieros y otros cualitativos y cuantitativos. Se basa en la idea de que los mercados pueden no valorar correctamente las acciones a corto plazo, pero los precios se ajustarán al valor intrínseco de la acción a largo plazo. Su objetivo es determinar el valor intrínseco de una empresa para tomar decisiones de inversión a largo plazo.

Supuestos Clave:

* Los precios de las acciones reflejarán eventualmente su valor intrínseco.
* Los inversores pueden obtener beneficios comprando acciones infravaloradas y vendiendo acciones sobrevaloradas.
* Los factores económicos fundamentales afectan el rendimiento de las acciones.

Las diferencias entre ambos modelos, se basan principalmente en el **Horizonte Temporal:** El análisis técnico tiende a ser a corto plazo, mientras que el análisis fundamental es a largo plazo. Respecto al foco de estudio, el análisis técnico se centra en los movimientos de precios y volúmenes, mientras que el análisis fundamental se centra en los datos financieros y económicos de la empresa. Las herramientas también suelen ser diferentes: el análisis técnico utiliza gráficos e indicadores técnicos, mientras que el análisis fundamental utiliza estados financieros, ratios y evaluaciones cualitativas.

Ambos enfoques pueden ser complementarios y muchos inversores los utilizan en conjunto para tomar decisiones más informadas.

## DESCARGA DE LOS DATOS

En esta sección veremos los lugares donde comenzar nuestro trabajo de ETL Extracción, Transformación y Carga (load), ya que una de las primeras tareas de un proyecto de Análisis de Datos, es obviamente, la recopilación de estos. Como no lo vamos a hacer de forma manual (no, por lo menos, la mayoría de ellos), debemos recurrir a diferentes sitios web donde podemos realizar la descarga según nuestras necesidades, para posteriormente realizar todas aquellas operaciones necesarias. Estos sitios de extracción de datos pueden ser dedicados exclusivamente a información financiera o a datos en general:

Sitios Exclusivamente Financieros

Son plataformas que se especializan en proporcionar información y datos relacionados con los mercados financieros. Estos sitios ofrecen datos sobre precios de acciones, índices bursátiles, commodities, divisas, bonos, entre otros. Además, suelen incluir análisis financieros, noticias del mercado, herramientas de inversión y gráficos en tiempo real.

Ejemplos:

**Investing.com** (<https://es.investing.com/currencies/xau-usd>):

Proporciona información sobre una amplia gama de instrumentos financieros, incluyendo acciones, divisas, materias primas, índices y criptomonedas. Ofrece noticias, análisis y gráficos en tiempo real.

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Figura 2: Investing.com Precio del Oro al contado

**Yahoo Finance:** Ofrece cotizaciones de acciones, noticias financieras, análisis de mercado, informes financieros y datos históricos.

**Bloomberg**: Proporciona noticias financieras, datos de mercado, análisis y herramientas para la gestión de carteras.

**Morningstar:** Ofrece análisis, datos de mercado, investigación sobre fondos mutuos y ETFs, y herramientas para inversores.

**MarketWatch**: Proporciona noticias financieras, datos de mercado, análisis y opiniones sobre inversiones.

Sitios de Datos Generales

Son plataformas que recopilan y proporcionan conjuntos de datos de diversas áreas, no limitados únicamente al ámbito financiero. Estos sitios abarcan datos sobre temas como ciencia, tecnología, salud, economía, medio ambiente, deportes, entre otros. Los datos pueden ser utilizados para análisis, investigación, proyectos académicos y aplicaciones comerciales.

Ejemplos:

**Kaggle.com**: Una plataforma que ofrece conjuntos de datos sobre una amplia variedad de temas, desde datos financieros hasta imágenes médicas, datos de redes sociales, datos de clima, entre otros. También es conocida por sus competencias de ciencia de datos.

**Data.gov:** El portal de datos abiertos del gobierno de los Estados Unidos, que proporciona acceso a conjuntos de datos sobre temas diversos como agricultura, educación, energía, salud y seguridad pública.

**UCI Machine Learning Repository**: Un repositorio de conjuntos de datos que se utilizan comúnmente para el desarrollo y prueba de algoritmos de aprendizaje automático y minería de datos.

**Google Dataset Search**: Un motor de búsqueda que permite encontrar conjuntos de datos disponibles públicamente en la web sobre diversos temas.

**World Bank Open Data**: Proporciona acceso a datos sobre indicadores de desarrollo económico y social de diferentes países.

Interfaz de usuario gráfica, Sitio web

Descripción generada automáticamente

Figura 3: Kaggle, apartado DATASETS

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura 4: Tipos de interés FED -Banco de la Reserva Federal -

## DATASETS OBTENIDOS

Como podemos observar, la extracción de datos la podemos realizar de diferentes formas a la hora de elaborar un “dataset”, en nuestro caso, lo hemos realizado de tres formas diferentes:

* **Método de extracción de datos desde archivos CSV**: Utilización de archivos CSV para importar datos financieros almacenados localmente.
* **Método de extracción de datos desde una API:** Uso de la API de FRED para obtener datos de tipos de interés y otros indicadores financieros en tiempo real.
* **Método de extracción de datos mediante el paquete yfinance:** Empleo de la biblioteca yfinance para acceder a datos financieros de Yahoo Finance, como cotizaciones históricas de acciones.

Finalmente, para nuestras primeras modelizaciones, utilizaremos estas fuentes de datos, en cuyo momento se detallará su estructura y características:

* **Yahoo Finance: import de Python**
* **GOLDSTOCK.csv**
* **STOCK MARKET DATASET.csv**
* **XAUUSD\_2010\_2023.csv**
* **TIPOS DE INTERES FED: API DESDE FRED.**

# CAPITULO 2: El escenario actual del análisis financiero

## Modelos y situación actual del oro. Regresión lineal

Antes de profundizar sobre una modelización compleja, tal vez sea necesario realizar un trabajo previo más sencillo, con el fin de poder comparar en el futuro el comportamiento y la mejora al respecto del modelo final.

En el campo del análisis predictivo, la **regresión lineal** se destaca por su simplicidad y eficacia, sirviendo como punto de partida para muchos proyectos de modelado estadístico y machine learning.

Este método busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. En nuestro caso, aplicaremos un modelo de regresión lineal para predecir los precios del oro, esta vez utilizando como *dataset* los precios del oro obtenidos mediante la biblioteca **yfinance** de *Yahoo Finance*, y utilizando como variables, las medias móviles, que calcularemos “in situ”, ya que es un indicador técnico, que, pese a no facilitarse en los datasets, lo podemos calcular.

Históricamente, la regresión lineal ha demostrado ser eficaz en diversos contextos financieros, desde la previsión de ingresos y gastos hasta la valoración de activos y la evaluación del riesgo crediticio. Su uso recurrente en estudios académicos y aplicaciones prácticas refuerza su credibilidad y utilidad.

Los ejemplos que ha servido de partida son este trabajo, son estudios contrastados, que han utilizado la regresión lineal para previsiones financieras:

* GOLD PRICE PREDICTIONS USING ML ALGORITHMS (Radhamani V, Manju D y otros). Department of Computing, Coimbatore Institute of Technology, Tamil Nadu, India: En dicho paper utilizan algoritmos de machine learning, entre ellos los **árboles de decisión,** y la **regresión lineal**, obteniendo unos resultados bastante positivos teniendo en cuenta valores como los precios de apertura, cierre y volumen. (http:ymerdigital.com), Julio 2022.
* TIME SERIES FORECASTING: (Umut Toygar Goz), 2023 Kaggle AI Report: En dicho estudio similar al anterior, tiene un cuenta un número ampliado de indicadores técnicos y se adentra en la utilización de las medias móviles. Utiliza algoritmos más enfocados al análisis de series temporales (LSTM), tal y como haremos nosotros en el modelo más avanzado de este trabajo.

A pesar de la complejidad y la inherente imprevisibilidad de los mercados financieros, existen algoritmos contrastados de regresión lineal y otros métodos de aprendizaje automático que han demostrado tener cierto éxito en la predicción de tendencias y comportamientos del mercado. Sin embargo, estos modelos no son infalibles y su capacidad para prever con precisión los movimientos del mercado es limitada debido a la naturaleza caótica y dinámica de los mismos.

En nuestro proyecto, nos proponemos ampliar los modelos predictivos existentes mediante el uso de técnicas más avanzadas como las Redes Neuronales de Memoria a Largo Plazo (LSTM), que son especialmente adecuadas para el análisis de series temporales complejas. Además, generaremos conjuntos de datos más completos y variados, incorporando una mayor cantidad de variables, incluyendo indicadores técnicos y fundamentales. Esto permitirá capturar de manera más efectiva las múltiples dimensiones que afectan los mercados financieros.

Asimismo, ajustaremos nuestros métodos para seleccionar horizontes temporales más apropiados para cada tipo de predicción, lo que mejorará la precisión y la utilidad de nuestros modelos. Al combinar estas innovaciones, esperamos superar las limitaciones de los enfoques actuales y proporcionar previsiones más robustas y precisas, lo que representa una contribución significativa al campo de las predicciones financieras.

No obstante, los primeros resultados que obtengamos de un modelo más sencillo nos servirán como línea base para comparar la eficacia de modelos más complejos como las redes neuronales, que son el modelo que implementaremos más adelante, y con una amplia gama de variables.

## El dataset “GC=F”, de Yahoo Finance.

Este dataset contiene los valores principales de los precios del oro, tiene la ventaja de estar bastante optimizado para la cantidad de datos que tiene, ya que data prácticamente del principio de las cotizaciones del oro. Además, al ser una biblioteca proporcionada por Yahoo!, podemos realizar la importación de una forma flexible, en nuestro caso, y siendo un ejemplo básico, se ha realizado de la siguiente forma:

| **Descripción del Dataset** |  |
| --- | --- |
| **Período de Tiempo** | Desde el 2 de enero de 2008 hasta el 20 de mayo de 2024 |
| **Índice** | DatetimeIndex (Fechas de días hábiles) |
| **Número de Columnas** | 5 |
| **Número Total de Entradas** | 4122 |
| **Uso de Memoria** | Aproximadamente 193.2 KB |

| **Columnas del Dataset** | **Tipo de Dato** | **Descripción** | **Valores Nulos** |
| --- | --- | --- | --- |
| Open | float64 | Precio de apertura del oro para cada día hábil | Ninguno (4122 valores no nulos) |
| High | float64 | Precio más alto del oro alcanzado durante cada día hábil | Ninguno (4122 valores no nulos) |
| Low | float64 | Precio más bajo del oro alcanzado durante cada día hábil | Ninguno (4122 valores no nulos) |
| Close | float64 | Precio de cierre del oro para cada día hábil | Ninguno (4122 valores no nulos) |
| Volume | int64 | Volumen de transacciones del oro para cada día hábil | Ninguno (4122 valores no nulos) |

Una vez extraídos, los datos, el paso natural es realizar un análisis estadístico de los mismos (que siendo un ejemplo de entrada, realizaremos de la forma más sencilla posible). Los datos estadísticos que arrojan la instrucción *describe()* del dataset **df\_gold** son los siguientes:

Descripción estadística del dataset:

Open High Low Close Volume

count 4122.000000 4122.000000 4122.000000 4122.000000 4122.000000

mean 1432.017613 1439.487215 1424.105313 1431.867345 5342.086851

std 334.022331 335.393239 332.936455 334.190696 28429.414848

min 705.000000 715.500000 681.000000 704.900024 0.000000

25% 1211.749969 1217.800049 1205.300049 1211.749969 46.000000

50% 1330.049988 1337.000000 1323.500000 1330.250000 152.000000

75% 1727.000000 1735.424988 1718.800049 1727.074982 488.000000

max 2415.800049 2435.800049 2409.699951 2433.899902 386334.000000

En resumen, los precios del oro presentan una considerable variabilidad, con valores que van desde alrededor de 700 hasta más de 2400. El volumen de transacciones también muestra una alta dispersión, con valores que van desde 0 hasta más de 380,000, lo que indica la presencia de días con muy alta actividad en comparación con otros.

También suele ser habitual, trazar las gráficas estadísticas mediante la instrucción **hist() y plt.show():**

Histograma

Descripción generada automáticamente

Figura 5: Histogramas dataset Oro Yahoo Finance

Con los datos extraídos, podemos trazar la gráfica del precio del oro a lo largo del tiempo, gracias a la librería “mathplot”:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Figura 6: Evolución del precio del oro

## Las variables explicativas (MA 3 y 9). Justificación de su uso

Hemos visto que, en el ámbito del análisis financiero, el uso de modelos de regresión lineal para predecir el precio el oro es una práctica común. Sin embargo, muchos de estos modelos tradicionales tienden a simplificar los datos, basándose únicamente en variables como el precio y el volumen de negociación. Aunque estos enfoques pueden ofrecer una visión general del comportamiento del mercado, a menudo carecen de la profundidad necesaria para capturar las dinámicas subyacentes que influyen en los movimientos de los precios.

El modelo que se propone a continuación intenta realizar un avance al incorporar medias móviles de 3 días (MA3) y 9 días (MA9) como variables predictoras.

Las medias móviles son herramientas técnicas esenciales en el análisis financiero que permiten suavizar las fluctuaciones de precios a corto y medio plazo, proporcionando así una visión más clara de las tendencias subyacentes. A continuación, se detallan las razones por las que este enfoque mejora y avanza sobre los modelos tradicionales:

**Captura de Tendencias Temporales:** Las medias móviles de diferentes periodos permiten capturar las tendencias de precios en diferentes horizontes temporales. Mientras que la MA3 refleja las variaciones a corto plazo, la MA9 ofrece una perspectiva a medio plazo. Este doble enfoque permite que el modelo sea más sensible a las variaciones recientes y, al mismo tiempo, mantenga una visión más estable de la tendencia general.

**Reducción de la Volatilidad del Ruido:** Al suavizar las fluctuaciones diarias, las medias móviles ayudan a reducir el "ruido" del mercado, es decir, las pequeñas variaciones que no necesariamente indican un cambio en la tendencia general. Esto mejora la precisión del modelo al centrarse en los movimientos de precios más significativos y evitar predicciones erróneas causadas por datos atípicos.

**Mejor Identificación de Patrones de Mercado:** La utilización de MA3 y MA9 permite identificar patrones de cruce de medias móviles, que son indicativos de cambios en la dirección de la tendencia. Por ejemplo, un cruce de la MA3 por encima de la MA9 puede señalar un cambio de tendencia alcista, mientras que lo contrario puede indicar una tendencia bajista. Estos patrones son valiosos para realizar predicciones más informadas y precisas.

**Adaptabilidad a Cambios del Mercado**: El uso de medias móviles hace que el modelo sea más adaptable a las condiciones cambiantes del mercado. A medida que los precios fluctúan, las medias móviles se ajustan, permitiendo que el modelo mantenga su relevancia y precisión en diferentes contextos de mercado, desde periodos de alta volatilidad hasta fases de estabilidad relativa.

**Complementación de Datos Simples:** Si bien los modelos tradicionales basados en precio y volumen ofrecen una base útil, la incorporación de medias móviles añade una capa adicional de análisis que puede complementar y enriquecer la información obtenida. Esto permite realizar predicciones más robustas y confiables, proporcionando una ventaja competitiva en la toma de decisiones financieras.

## La variable objetivo “next day price”

La variable objetivo en el modelo es **next\_day\_price**, que representa el precio de cierre del oro al día siguiente. La idea detrás de este enfoque es predecir el precio de cierre del día siguiente utilizando las medias móviles de 3 y 9 días como variables explicativas.

## División de los conjuntos de entrenamiento y prueba

Una vez definidas las variables, podemos realizar la división del dataset en un conjunto de entrenamiento y otro de prueba. Utilizaremos la partición habitual del 80% entrenamiento y 20% prueba. Esta división asegura que el modelo tenga suficientes datos para aprender patrones durante el entrenamiento, mientras que el conjunto de prueba permite evaluar objetivamente su rendimiento en predicciones futuras.

## Modelo de Regresión Lineal:

Ahora implementaremos un modelo de **regresión lineal** con la librería **Sci-Kit Learn.**

Un modelo de regresión lineal es un método estadístico utilizado para predecir el valor de una variable dependiente (u objetivo) basándose en los valores de una o más variables independientes (o explicativas). La regresión lineal asume que existe una relación lineal entre las variables independientes y la variable dependiente, es decir, que la variable dependiente puede expresarse como una combinación lineal de las variables independientes más un término constante.

* *y* es el precio del oro al contado del día siguiente.
* β0 es el **intercepto** (Precio Oro con variables dependientes a cero)
* β1 es el coeficiente de la media móvil de 3 días.
* β2 es el coeficiente de la media móvil de 9 días.
* *S*3 y *S*9​ son los valores de las medias móviles de 3 y 9 días

En nuestro caso, el objetivo es predecir el precio del oro al contado del día siguiente utilizando dos variables explicativas: la media móvil de 3 días y la media móvil de 9 días del precio de cierre del oro.

## Resultados y evaluación del modelo:

Una vez aplicado el algoritmo y teniendo el modelo establecido de Regresión Líneal, llega el momento de hacer las comparativas, y entrar en la fase de obtener los resultados y evaluar el modelo respecto al conjunto de datos de prueba. En nuestro caso, el gráfico parece prometedor:

Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

Figura 7 - Resultado previsión modelo RL

A continuación, vamos a describir brevemente las principales métricas para evaluar este modelo, y analizar los resultados obtenidos en cada uno de ellos, que en un futuro compararemos con los obtenidos con un modelo más complejo y con más variables , utilizando como motor predictivo una red neuronal de tipo LSTM.

**Coeficiente de Determinación (R² Score)**:

* Mide la proporción de la **varianza** en la variable dependiente que es explicada por el modelo.

**Resultado: 98.24%**

* **Interpretación**: El coeficiente de determinación (R²) indica que el 98.24% de la variabilidad en el precio del oro al contado puede ser explicada por el modelo de regresión lineal utilizando las medias móviles de 3 y 9 días. Este es un valor muy alto, lo que sugiere que el modelo se ajusta muy bien a los datos de entrenamiento y tiene un alto poder explicativo.
* **Conclusión**: El modelo tiene una excelente capacidad para explicar la variabilidad en los datos, lo cual es un indicador positivo.

**Error Cuadrático Medio (Mean Squared Error, MSE)**:

* Mide el promedio de los errores al cuadrado entre los valores observados y los predichos.

**Resultado: 404.76**

* **Interpretación**: El MSE mide el promedio de los errores al cuadrado entre los valores predichos y los valores reales. Un valor de 404.76 sugiere que, en promedio, el cuadrado de los errores es relativamente bajo, aunque este valor depende de la escala de la variable dependiente (precio del oro).
* **Conclusión**: Un MSE bajo indica que las predicciones están generalmente cerca de los valores reales, lo cual es positivo.

**Raíz del Error Cuadrático Medio (Root Mean Squared Error, RMSE)**:

* Es la raíz cuadrada del MSE y tiene la misma unidad que la variable dependiente.

**Resultado (RMSE): 20.12**

* **Interpretación**: El RMSE es la raíz cuadrada del MSE y tiene la misma unidad que la variable dependiente (USD en este caso). Un RMSE de 20.12 sugiere que el error típico de las predicciones del modelo es de aproximadamente 20.12 USD.
* **Conclusión**: El RMSE proporciona una medida intuitiva del error de predicción promedio. En el contexto del precio del oro, es importante considerar si un error de 20.12 USD es aceptable o significativo.

**Error Absoluto Medio (Mean Absolute Error, MAE)**:

* Mide el promedio de los errores absolutos entre los valores observados y los predichos.

**Resultado: 15.28**

* **Interpretación**: El MAE mide el promedio de los errores absolutos entre las predicciones y los valores reales. Un MAE de 15.28 indica que, en promedio, las predicciones del modelo difieren de los valores reales por 15.28 USD.
* **Conclusión**: El MAE es fácil de interpretar y sugiere que el modelo tiene un error promedio moderado en sus predicciones.

**Error Absoluto Medio Porcentual (Mean Absolute Percentage Error, MAPE)**:

* Mide el error absoluto medio como un porcentaje de los valores observados.

**Resultado: 0.81%**

* **Interpretación**: El MAPE mide el error absoluto medio como un porcentaje de los valores reales. Un MAPE de 0.81% sugiere que, en promedio, las predicciones del modelo son muy precisas, con un error medio inferior al 1% del valor real.
* **Conclusión**: Un MAPE tan bajo indica que el modelo tiene una precisión muy alta y es capaz de predecir el precio del oro con una desviación muy pequeña relativa a los valores reales.

### Resumen:

En conjunto, las métricas indican que el modelo de regresión lineal tiene un rendimiento excelente:

* **R² Score alto**: Muestra una gran capacidad explicativa del modelo.
* **MSE y RMSE bajos**: Indican que las predicciones del modelo están generalmente cerca de los valores reales.
* **MAE bajo**: Refuerza que el error promedio de las predicciones es pequeño.
* **MAPE extremadamente bajo**: Demuestra una alta precisión relativa del modelo.

## Avanzado en la interpretación del análisis: estrategias de inversión con medias móviles

Para no limitarse a dejar el estudio en el apartado de los resultados, vamos a extendernos en el análisis de los datos, y concretamente, centrarnos en las variables explicativas que nos van a permitir realizar un análisis pormenorizado desde dos puntos de vista ya diferenciados del precio objetivo del día siguiente. Estas variables son las ya definidas por nosotros: las medias móviles a corto y muy corto plazo. (3 y 9)

Para poder entenderlas, previamente debemos realizar un gráfico sobre ellas, valga el ejemplo de un mes para poder ver las oscilaciones y los cruces:

Gráfico, Gráfico de líneas

Descripción generada automáticamenteFigura 8- Estudio de medias móviles 3 y 9 días

La estrategia básica basada en estas medias móviles se llama "cruce de medias móviles", y nos indican ciertos momentos “señales” en los que realizar compras o ventas.

**Señal de Compra**:

Ocurre cuando la media móvil de corto plazo (S\_3) cruza por encima de la media móvil de largo plazo (S\_9).

Esto indica que el precio está comenzando una tendencia alcista.

Los inversores pueden considerar comprar oro cuando se observa este cruce, ya que se espera que el precio continúe subiendo.

**Señal de Venta:**

Ocurre cuando la media móvil de corto plazo (S\_3) cruza por debajo de la media móvil de largo plazo (S\_9).

Esto sugiere que el precio está comenzando una tendencia bajista.

Los inversores pueden considerar vender oro cuando se observa este cruce, ya que se espera que el precio continúe bajando.

Por ejemplo, supongamos que el 15 de enero de 2021, la media móvil de 3 días (S\_3) cruza por encima de la media móvil de 9 días (S\_9). Un inversor podría considerar esto como una señal de compra y adquirir oro. Si más adelante, el 25 de enero de 2021, la media móvil de 3 días (S\_3) cruza por debajo de la media móvil de 9 días (S\_9), el mismo inversor podría interpretar esto como una señal de venta y decidir vender el oro comprado anteriormente.

## Conclusiones y limitaciones del modelo

Aunque los resultados obtenidos con el modelo de regresión lineal parecen satisfactorios a primera vista, con un R² Score de 98.24% y errores relativamente bajos (MSE: 404.76, RMSE: 20.12, MAE: 15.28, MAPE: 0.81%), debemos considerar varias limitaciones importantes:

1. **Simplicidad del Modelo**:
   * **Naturaleza Lineal**: La regresión lineal asume una relación lineal entre las variables explicativas (medias móviles) y la variable objetivo (precio del oro). En la realidad, las relaciones entre los factores que influyen en el precio del oro son mucho más complejas y no lineales.
   * **Variables Limitadas**: El modelo solo considera dos variables explicativas (medias móviles de 3 y 9 días), ignorando otros factores importantes como tasas de interés, inflación, políticas monetarias y eventos geopolíticos que pueden tener un impacto significativo en el precio del oro.
2. **Horizonte Temporal de Inversión**:
   * **Predicción a Muy Corto Plazo**: El modelo está diseñado para predecir el precio del oro al día siguiente, lo cual es útil para análisis a muy corto plazo. Sin embargo, esta información no es necesariamente suficiente para desarrollar estrategias de inversión robustas a medio o largo plazo.
   * **Volatilidad del Mercado**: La alta volatilidad del precio del oro puede hacer que las predicciones a corto plazo sean insuficientes para tomar decisiones de inversión efectivas.

### **Horizonte Temporal Óptimo para Invertir en Oro**

Basándonos en los resultados del modelo de regresión lineal, un horizonte de inversión relativamente optimista podría ser el de corto a mediano plazo (semanas a meses). Sin embargo, dado que las predicciones se basan solo en tendencias de precios recientes, este enfoque puede no captar adecuadamente los cambios repentinos en el mercado del oro.

Para mejorar la efectividad de las inversiones en oro, es recomendable considerar horizontes temporales más amplios y complementar el modelo con análisis fundamentales y técnicos adicionales que tengan en cuenta una gama más amplia de variables.

### **Mejoras con Redes Neuronales LSTM**

Las redes neuronales de tipo Long Short-Term Memory (LSTM) son una variante de las redes neuronales recurrentes (RNN) especialmente adecuadas para modelar datos secuenciales y dependencias a largo plazo. Estas redes pueden mejorar significativamente la precisión de las predicciones de precios del oro debido a su capacidad para:

1. **Capturar Relaciones No Lineales**:
   * **Aprendizaje de Patrones Complejos**: Las LSTM pueden aprender y modelar relaciones no lineales complejas entre múltiples variables explicativas, lo que las hace más adecuadas para capturar la dinámica real del mercado del oro.
2. **Incorporar Más Variables**:
   * **Factores Externos**: A diferencia del modelo de regresión lineal, las LSTM pueden incorporar una amplia gama de variables adicionales, como tasas de interés, índices de inflación, indicadores económicos, y eventos geopolíticos, lo que proporciona una visión más holística y precisa del mercado.
3. **Mejor Manejo de Datos Temporales**:
   * **Dependencias Temporales**: Las LSTM están diseñadas para manejar secuencias de datos y pueden recordar información relevante durante largos periodos, lo que mejora su capacidad para hacer predicciones basadas en datos históricos extensos.
4. **Predicciones a Largo Plazo**:
   * **Horizonte Temporal Ampliado**: Con su capacidad para manejar dependencias a largo plazo, las LSTM pueden proporcionar predicciones más precisas no solo para el corto plazo, sino también para horizontes temporales más amplios, lo que es crucial para estrategias de inversión a mediano y largo plazo.

### **Resumen**

En resumen, aunque el modelo de regresión lineal ha mostrado resultados prometedores, sumando a que hemos realizado una mejora respecto a modelos habituales implementados, su simplicidad y limitación a pocas variables lo hacen insuficiente para estrategias de inversión seguras.

Un enfoque más efectivo incluiría el uso de redes neuronales LSTM, que pueden incorporar múltiples variables y capturar relaciones no lineales complejas, proporcionando así predicciones más precisas y útiles para decisiones de inversión en el mercado del oro a diferentes horizontes temporales.

# CAPITULO 3: Análisis predictivo con Redes Neuronales LSTM

## Introducción:

### La importancia de los datos

Para desarrollar un modelo de predicción más robusto, debemos unificar diversas fuentes de datos en un dataset más complejo y completo.

Este dataset incluirá no solo el precio del oro, sino también otros valores de inversión como la plata y el bitcoin, que pueden ofrecer información valiosa sobre las tendencias del mercado de metales preciosos y criptomonedas. Además, integraremos una variedad de indicadores técnicos: algunos ya existentes y otros que calcularemos nosotros mismos. Estos indicadores técnicos ayudarán a capturar patrones de precios y tendencias del mercado que no son evidentes a simple vista. Al combinar estos diversos tipos de datos, esperamos mejorar significativamente la precisión y la capacidad predictiva de nuestro modelo, proporcionando una herramienta más efectiva para la toma de decisiones de inversión.

### Redes Neuronales LSTM

Un enfoque más efectivo que el anterior modelo sencillo de Regresión Lineal, incluiría el uso de redes neuronales LSTM, que pueden incorporar múltiples variables y capturar **relaciones no lineales complejas**, proporcionando así predicciones más precisas y útiles para decisiones de inversión en el mercado del oro a diferentes horizontes temporales.

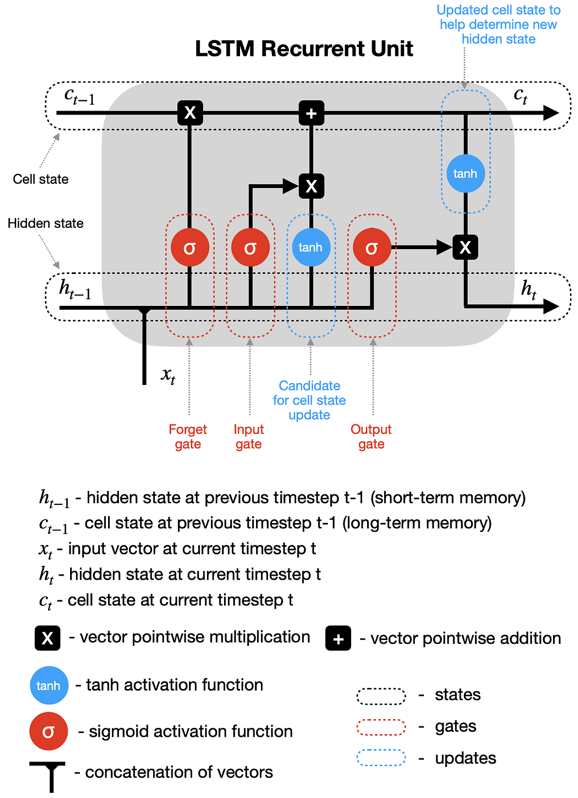


Figura 9: Diagrama de una red LSTM (Medium: <https://pub.aimind.so/long-short-term-memory-a-detailed-note-from-scratch-2660e476cf42>)

Las redes neuronales de tipo Long Short-Term Memory (LSTM) son una variante de las redes neuronales recurrentes (RNN) especialmente adecuadas para modelar datos secuenciales y dependencias a largo plazo. Estas redes pueden mejorar significativamente la precisión de las predicciones de precios del oro debido a su capacidad para:

1. **Capturar Relaciones No Lineales**:
   * **Aprendizaje de Patrones Complejos**: Las LSTM pueden aprender y modelar relaciones no lineales complejas entre múltiples variables explicativas, lo que las hace más adecuadas para capturar la dinámica real del mercado del oro.
2. **Incorporar Más Variables**:
   * **Factores Externos**: A diferencia del modelo de regresión lineal, las LSTM pueden incorporar una amplia gama de variables adicionales, como tasas de interés, índices de inflación, indicadores económicos, y eventos geopolíticos, lo que proporciona una visión más holística y precisa del mercado.
3. **Mejor Manejo de Datos Temporales**:
   * **Dependencias Temporales**: Las LSTM están diseñadas para manejar secuencias de datos y pueden recordar información relevante durante largos periodos, lo que mejora su capacidad para hacer predicciones basadas en datos históricos extensos.
4. **Predicciones a Largo Plazo**:
   * **Horizonte Temporal Ampliado**: Con su capacidad para manejar dependencias a largo plazo, las LSTM pueden proporcionar predicciones más precisas no solo para el corto plazo, sino también para horizontes temporales más amplios, lo que es crucial para estrategias de inversión a mediano y largo plazo.

## El proceso ETL y la construcción de nuestro dataset completo

ETL es un acrónimo de Extract, Transform, Load (Extracción, Transformación, Carga), y consiste en un proceso clave en la gestión de datos que se utiliza para integrar datos de muchas fuentes y convertirlos en un conjunto unificado. A continuación, explicamos las tres etapas del proceso:

### Extracción (Extract):

Aquí recolectaremos datos de diversas fuentes. Estas fuentes pueden ser bases de datos, archivos, servicios web, APIs, y más, tal y como hemos comentado con anterioridad. El objetivo principal es obtener los datos necesarios de las fuentes originales, asegurando que todos los datos relevantes sean capturados para su posterior procesamiento.

### Transformación (Transform):

En esta etapa,, los datos extraídos se limpian, formatean y transforman para adecuarse al modelo de datos de nuestro modelo.

El objetivo de esta fase es mejorar la calidad de los datos y convertirlos en un formato adecuado y consistente para el análisis y el almacenamiento. Las tareas habituales de esta etapa suelen ser las siguientes:

* Limpieza de Datos: Eliminación de valores nulos, duplicados y corrección de errores.
* Normalización: Ajuste de los datos a un estándar común (por ejemplo, convertir todas las fechas al mismo formato).
* Enriquecimiento: Agregar datos adicionales derivados o calculados.
* Integración: Combinación de datos de diferentes fuentes en una estructura unificada.
* Agregación: Resumir o consolidar datos (por ejemplo, calcular promedios o totales).

### Carga:

Esta etapa implica insertar los datos transformados en el sistema de destino, que suele ser un almacén de datos, teniendo en cuenta que existen multitud de tipos de estos últimos, nuestro objetivo es asegurar que los datos transformados estén disponibles para su utilización. En nuestro caso, dicha carga se convertirá en un conjunto de datos del tipo *DataFrame* de *Pandas*

Las actividades comunes de esta etapa consisten en la inserción de los datos en el destino, la actualización o modificación, la indexación y la validación de estos para comprobar que sean correctos y consistentes.

Teniendo en cuenta que nuestra variable objetivo es el precio futuro del oro, el primer conjunto de datos que vamos a tratar para construir nuestro dataset, estará compuesto por características inherentes a este valor.

Dicho conjunto de datos proviene de Kaggle (ya visto anteriormente), y consta de un CSV, cuya descripción vamos a tratar a continuación.

### GoldStock.csv

Lugar de descarga: <https://www.kaggle.com/datasets/sahilwagh/gold-stock-prices>

Este conjunto de datos ofrece un registro exhaustivo de los precios diarios del oro desde el 19 de enero de 2014 hasta el 22 de enero de 2024. Los datos son proporcionados por Nasdaq e incluyen métricas financieras clave para cada día de negociación.

El DATASET consta de las siguientes columnas:

* Fecha: Una fecha única para cada día de negociación registrado.
* Cierre: El precio de cierre del oro en la fecha correspondiente.
* Volumen: Volumen de negociación del oro en la fecha pertinente.
* Apertura: El precio de apertura del oro en la fecha pertinente.
* Máximo: El precio más alto registrado del oro durante el día de negociación.
* Mínimo: El precio más bajo registrado del oro en el día de negociación.

Como vemos, aquí tenemos una de las primeras limitaciones que nos podemos encontrar al construir un dataset, que es la delimitación de las fechas. Debemos considerar que el conjunto de datos que posea un rango de fechas más limitado, va a influir decisivamente en el resto, ya que todos van a estar supeditados al conjunto con menor rango de fechas. No obstante, para nuestro propósito académico, y para no necesitar una excesiva potencia computacional, va a ser suficiente el rango de tiempo con el que vamos a trabajar.

Mediante Python, aplicaremos las primeras instrucciones del proceso ETL. En nuestro, caso, al describir el conjunto de datos obtenido, tenemos estos resultados:

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 2511 entries, 0 to 2510

Data columns (total 7 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Unnamed: 0 2511 non-null int64

1 Date 2511 non-null object

2 Close 2511 non-null float64

3 Volume 2511 non-null float64

4 Open 2511 non-null float64

5 High 2511 non-null float64

6 Low 2511 non-null float64

dtypes: float64(5), int64(1), object(1)

memory usage: 137.4+ KB

None

Descripción estadística del dataset:

Unnamed: 0 Close Volume Open High \

count 2511.000000 2511.000000 2511.000000 2511.000000 2511.000000

mean 1260.792911 1498.726085 185970.770609 1498.725528 1508.451454

std 729.262879 298.824811 97600.769382 299.118187 301.262244

min 0.000000 1049.600000 1.000000 1051.500000 1062.700000

25% 630.500000 1249.850000 126693.500000 1249.500000 1257.300000

50% 1259.000000 1332.800000 175421.000000 1334.000000 1342.400000

75% 1888.500000 1805.850000 234832.000000 1805.600000 1815.450000

max 2532.000000 2093.100000 787217.000000 2094.400000 2098.200000

Low

count 2511.000000

mean 1488.869932

std 296.417703

min 1045.400000

25% 1242.350000

50% 1326.600000

75% 1793.050000

max 2074.600000

Primeras filas del dataset:

Unnamed: 0 Date Close Volume Open High Low

0 0 2024-01-19 2029.3 166078.0 2027.4 2041.9 2022.2

1 1 2024-01-18 2021.6 167013.0 2009.1 2025.6 2007.7

2 2 2024-01-17 2006.5 245194.0 2031.7 2036.1 2004.6

3 3 2024-01-16 2030.2 277995.0 2053.4 2062.8 2027.6

4 4 2024-01-12 2051.6 250946.0 2033.2 2067.3 2033.1

Últimas filas del dataset:

Unnamed: 0 Date Close Volume Open High Low

2506 2528 2014-01-28 1250.5 81426.0 1254.9 1261.9 1248.0

2507 2529 2014-01-27 1263.5 63419.0 1269.9 1280.1 1252.0

2508 2530 2014-01-24 1264.5 34998.0 1264.3 1273.2 1256.9

2509 2531 2014-01-23 1262.5 41697.0 1235.1 1267.1 1230.8

2510 2532 2014-01-22 1238.6 80262.0 1240.5 1243.5 1235.5

Valores nulos por columna:

Unnamed: 0 0

Date 0

Close 0

Volume 0

Open 0

High 0

Low 0

dtype: int64

Con esta primera extracción, podemos inferir las siguientes conclusiones:

1. **Estructura del Dataset:**
   * **Total de entradas:** 2511
   * **Total de columnas:** 7
   * **Tipos de datos:**
     + Enteros: 1 columna (**Unnamed: 0**)
     + Objetos (strings): 1 columna (**Date**), que vamos a tener que cambiar su formato.
     + Flotantes: 5 columnas (**Close**, **Volume**, **Open**, **High**, **Low**): es un formato adecuado para valores monetarios. Dichos valores están representados en USD, es decir dólares americanos.
2. **Descripción Estadística:**
   * **Unnamed: 0:** Es un simple identificador de filas, podríamos plantearnos su eliminación.
   * **Close, Open, High, Low:**
     + Representan los precios del oro respecto al cierre, apertura, valor más alto y valor más bajo.
     + Los valores están en un rango desde aproximadamente 1045 hasta 2098.
     + La media de los precios está alrededor de 1498.
   * La diferencia entre el precio máximo y mínimo es de aproximadamente 1053.8, lo que indica variabilidad significativa en los precios.
   * **Volume:**
     + Representa el volumen de transacciones. Es uno de los indicadores técnicos más importantes.
     + Varía considerablemente con un rango de 1 a 787,217 y una media alrededor de 185,970.
   * El volumen tiene una gran desviación estándar, lo que sugiere que hay días con volúmenes de transacciones extremadamente altos en comparación con otros días. Podríamos realizar un estudio más profundo en relación a los mismos, por si los consideramos como **valores atípicos (outliers)**
3. **Valores Nulos:**
   * No hay valores nulos en el dataset, lo cual es ideal para el análisis, ya que no se requiere limpieza adicional en este aspecto.

**Cambios a realizar:**

A tenor del primer análisis descriptivo, vamos a tener que realizar dos acciones muy importantes en este primer conjunto de datos:

* Eliminar la columna del índice que no aporta información relevante
* Convertir la fecha a un formato adecuado, ya que hemos visto que aparece como si fuera un “String”
* Investigar que ocurre con los volúmenes.

Los dos primeros cambios son fácilmente realizables con unas líneas sencillas de código que aparecen comentadas en el entregable. Sin embargo, para observar el volumen intercambiado, podemos comenzar con un gráfico en el que nos aparezcan los datos de **precio, fecha y volumen.**

Para su visualización se ha utilizado, al igual que hicimos con el modelo de regresión lineal, el módulo de Matplot, generando un gráfico de líneas, que marca los precios, intercalándolo con uno de barras, para mostrar los volúmenes, de esta forma tal vez podamos observar algún patrón al respecto:

Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamenteFigura 10- Precio y Volumen del Oro (14/24)

Este gráfico nos ofrece una visión preliminar, con una suerte de patrón; que indica que los volúmenes anteceden a un siguiente movimiento alcista, tal y como podemos observar en el período que va desde 2016 a 2020, anticipando una subida en dicho período. No obstante, esta es una visión general que abarca demasiado tiempo, y en la que no podemos ver realmente que es lo que ocurre con los valores extremadamente bajos.

Para seguir con análisis más minucioso, podemos establecer un gráfico individual para cada año, con el fin de ver si existe cierta estacionalidad. Una vez ejecutado el código, obtenemos los siguientes gráficos:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamente Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamenteFigura 11 - Detalle anual Precio y Volumen Oro

Los gráficos mensuales nos muestran que no existe una estacionalidad respecto a los volúmenes de negociación, y que es están más sujetos a la variabilidad de los precios, evidenciando una correlación entre grandes variaciones de los precios y los volúmenes, pero no en un sentido claro, ya que una oscilación del precio, tanto ascendente como descendente realiza un aumento considerable del intercambio de valores, algo que no ocurre sólo con el oro, si no con todos los valores financieros intercambiables.

Obviamente tendremos que discernir en qué sentido va la influencia; es decir, si los precios varían porque varía el volumen, o viceversa. Algo, que todavía no estamos en posición de decidir, porque, no lo olvidemos, estamos tratando apenas con dos variables, y recordemos la cantidad de factores tanto técnicos como fundamentales que influyen en el valor de una cotización.

También podemos observar cómo existen irregularidades, o más formalmente “outliers”, o **valores atípicos**; la más marcada es el caso de octubre de 2021, en el que apenas hay volumen. Sin embargo, este volumen se multiplica el mes siguiente.

Si investigamos, podemos comprobar como el dataset apenas contiene datos de dicho período. Vamos además a comprarlo con el siguiente y con la media total del conjunto de datos:

Octubre de 2021: 4 filas

Noviembre de 2021: 38 filas

Promedio mensual del dataset: Aproximadamente 20.82 filas por mes

Es posible que aquí tengamos un error de extracción de datos, por lo que, llegados a este punto, tal vez sea necesario realizar una comparación con los datos extraídos de Yahoo Finance, con el fin de ver sus posibles diferencias.

### Dataset GLD de Yahoo Finance.

Anteriormente, en el modelo introductorio de regresión lineal ya hemos utilizado este dataset. Sin embargo, ahora vamos a diseccionarlo más para poder discernir cuál de los dos conjuntos de datos iniciales puede ser más relevante para nuestro propósito, teniendo en cuenta la falta de datos que hemos visto en el anterior “goldstock.csv”, obtenido de Kaggle.

Mediante las instrucciones habituales, aquí tenemos los resultados estadísticos principales obtenidos:

[\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*100%%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*] 1 of 1 completed

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

DatetimeIndex: 2513 entries, 2014-01-22 to 2024-01-18

Data columns (total 5 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Open 2513 non-null float64

1 High 2513 non-null float64

2 Low 2513 non-null float64

3 Close 2513 non-null float64

4 Volume 2513 non-null int64

dtypes: float64(4), int64(1)

memory usage: 117.8 KB

None

Descripción estadística del dataset:

Open High Low Close Volume

count 2513.000000 2513.000000 2513.000000 2513.000000 2513.000000

mean 1494.751015 1501.760924 1487.833825 1494.778789 5311.463589

std 295.904912 297.662285 294.268456 295.918402 29358.987472

min 1053.699951 1062.000000 1046.199951 1050.800049 0.000000

25% 1248.000000 1253.199951 1241.699951 1247.599976 46.000000

50% 1331.099976 1337.500000 1326.199951 1331.300049 172.000000

75% 1795.500000 1805.400024 1788.199951 1798.099976 532.000000

max 2081.600098 2130.199951 2066.500000 2081.899902 386334.000000

Si realizamos una comparativa de volúmenes, podemos observar como las diferencias son notorias en todo el dataset, pero vamos a analizar el mayor “outlier”, que ocurría en octubre de 2021:

| **Fecha** | **Volumen (Goldstock)** | **Volumen (Yahoo Finance)** |
| --- | --- | --- |
| 2021-10-01 | 148,890 | 8,371,088 |
| 2021-10-04 | 163,944 | 9,000,000 |
| 2021-10-29 | 248,553 | 9,900,000 |
| 2021-10-29 | 248,553 | 9,900,000 |

En esta parte de nuestro análisis, hemos podido comprobar que una de las claves en el análisis de datos va a ser EXTRAER DATOS DE FUENTES SEGURAS Y CONFIABLES.

En caso de examinar otras fuentes, podemos observar que los datos reales son los de Yahoo Finance. Cierto es que los precios son los mismos, pero por alguna razón en el dataset extraído de Kaggle, los volúmenes son extremadamente bajos.

Un inciso muy importante que hemos aprendido en el master, es la importancia de consultar o investigar bien la materia objeto de análisis, y aunque no soy un experto en la materia, mi pequeña experiencia en el ámbito de las inversiones, me dice que probablemente los datos del primer dataset no corresponden al mercado al contado del Oro, si no seguramente, a uno de los ETFs de inversión en Oro.

Un ETF (Exchange-Traded Fund) es un tipo de fondo de inversión que se negocia en las bolsas de valores, similar a las acciones. Los ETFs poseen una suerte de bolsa o canasta de activos, que pueden incluir acciones, bonos, commodities (como el oro) o una combinación de estos. Son herramientas valiosas para inversores que buscan exposición al oro con ventajas de diversificación y facilidad de transacción. Sin embargo, los volúmenes negociados de un ETF de oro serán inevitablemente más bajos en comparación con el mercado total de oro debido a su alcance limitado y la preferencia de algunos inversores por otras formas de inversión en oro.

Por todo lo expuesto anteriormente, vamos a tener que desechar el dataset “goldstock.csv”, por el mas fiable de Yahoo Finance.

No está de más cara al análisis, comprobar el resultado gráfico del nuevo dataset elegido:

Por un lado, vamos a ver el gráfico general:

Gráfico, Gráfico de líneas

Descripción generada automáticamenteFigura 12- Precio del Oro y Vol. (Yahoo! Finance)

Y aunque los números respecto al volumen eran más prometedores, es curioso que si vamos a cualquier año de forma individual siempre tenemos el mismo patrón, con grandes diferencias entre meses:

Gráfico

Descripción generada automáticamente

Los pocos datos del período agosto – octubre son realistas, porque efectivamente, algunos mercados cierran y no hay apenas movimiento. No obstante, y después de ver varios orígenes de datos, los volúmenes, al contrario del precio, muchas veces son una apreciación bimensual. Por lo que vamos a tener que escalar y normalizar los datos próximamente, con el fin de minimizar las diferencias.

Este escalado de datos también va a ser crucial para ver el horizonte temporal de inversión, ya que, en un modelo predictivo basado en series temporales, no es lo mismo intentar realizar inversiones de muy corto plazo, que, a largo plazo, sobre todo teniendo en cuenta que va a ser muy difícil tener el dato real del volumen negociado.

En caso de normalizar por trimestres, las gráficas del volumen ya ofrecen unos datos más coherentes:

Gráfico

Descripción generada automáticamenteFigura 13- Precio y Vol. Oro TRIMESTRAL (Yahoo! Finance)

Es un patrón lógico y repetitivo, ya que generalmente el primer trimestre de cada año suele tener siempre un mayor volumen de transacciones.

Es bastante probable que tengamos que agregar una nueva columna que realice una media o similar de los valores del volumen, con el fin de normalizar y escalar los datos. No obstante, antes de ello, vamos a analizar otros datos financieros relacionados.

### Stock Market Dataset 2019 – 2024

El siguiente conjunto de datos es muy interesante, ya que gracias a él vamos a disponer de un gran número de alternativas de inversión, además de poder seguir comprobando si los datos que disponemos del oro son confiables.

Este conjunto de datos se ha obtenido también desde Kaggle, concretamente en esta dirección:

<https://www.kaggle.com/datasets/saketk511/2019-2024-us-stock-market-data>

Y está es la definición, según su propio autor:

*“Este conjunto de datos recoge un examen detallado de la dinámica del mercado durante un periodo de cinco años, centrándose en la fluctuación de los precios y los volúmenes de negociación de una cartera diversificada. Abarca varios sectores, entre ellos materias primas energéticas como el gas natural y el petróleo crudo, metales como el cobre, el platino, la plata y el oro, criptomonedas como Bitcoin y Ethereum, e índices bursátiles y empresas clave como S&P 500, Nasdaq 100, Apple, Tesla, Microsoft, Google, Nvidia, Berkshire Hathaway, Netflix, Amazon y Meta Platforms. Este conjunto de datos constituye un valioso recurso para analizar tendencias y patrones en los mercados mundiales.”*

La estructura según el enunciado, es la siguiente:

Date: The date of the recorded data, formatted as DD-MM-YYYY.

Natural\_Gas\_Price: Price of natural gas in USD per million British thermal units (MMBtu).

Natural\_Gas\_Vol.: Trading volume of natural gas

Crude\_oil\_Price: Price of crude oil in USD per barrel.

Crude\_oil\_Vol.: Trading volume of crude oil

Copper\_Price: Price of copper in USD per pound.

Copper\_Vol.: Trading volume of copper

Bitcoin\_Price: Price of Bitcoin in USD.

Bitcoin\_Vol.: Trading volume of Bitcoin

Platinum\_Price: Price of platinum in USD per troy ounce.

Platinum\_Vol.: Trading volume of platinum

Ethereum\_Price: Price of Ethereum in USD.

Ethereum\_Vol.: Trading volume of Ethereum

S&P\_500\_Price: Price index of the S&P 500.

Nasdaq\_100\_Price: Price index of the Nasdaq 100.

Nasdaq\_100\_Vol.: Trading volume for the Nasdaq 100 index

Apple\_Price: Stock price of Apple Inc. in USD.

Apple\_Vol.: Trading volume of Apple Inc. stock

Tesla\_Price: Stock price of Tesla Inc. in USD.

Tesla\_Vol.: Trading volume of Tesla Inc. stock

Microsoft\_Price: Stock price of Microsoft Corporation in USD.

Microsoft\_Vol.: Trading volume of Microsoft Corporation stock

Silver\_Price: Price of silver in USD per troy ounce.

Silver\_Vol.: Trading volume of silver

Google\_Price: Stock price of Alphabet Inc. (Google) in USD.

Google\_Vol.: Trading volume of Alphabet Inc. stock

Nvidia\_Price: Stock price of Nvidia Corporation in USD.

Nvidia\_Vol.: Trading volume of Nvidia Corporation stock

Berkshire\_Price: Stock price of Berkshire Hathaway Inc. in USD.

Berkshire\_Vol.: Trading volume of Berkshire Hathaway Inc. stock

Netflix\_Price: Stock price of Netflix Inc. in USD.

Netflix\_Vol.: Trading volume of Netflix Inc. stock

Amazon\_Price: Stock price of Amazon.com Inc. in USD.

Amazon\_Vol.: Trading volume of Amazon.com Inc. stock

Meta\_Price: Stock price of Meta Platforms, Inc. (formerly Facebook) in USD.

Meta\_Vol.: Trading volume of Meta Platforms, Inc. stock

Gold\_Price: Price of gold in USD per troy ounce.

Gold\_Vol.: Trading volume of gold

Como podemos observar existen infinidad de variables, incluso la del oro, por lo que, de una de las primeras tareas será, estudiar su distribución (únicamente del oro) y ver las diferencias respecto al dataset de yfinance, que de momento era el mas viable.

Además, en este dataset vamos a tener que realizar bastantes acciones, ya que al ver su definición, podemos ver como se mezclan diferentes tipos de datos, existe una columna de índices que no necesitamos, y además nos encontramos con el problema de los valores nulos, ya que hay bastantes en algunos de los campos:

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1243 entries, 0 to 1242

Data columns (total 39 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Unnamed: 0 1243 non-null int64

1 Date 1243 non-null object

2 Natural\_Gas\_Price 1243 non-null float64

3 Natural\_Gas\_Vol. 1239 non-null float64

4 Crude\_oil\_Price 1243 non-null float64

5 Crude\_oil\_Vol. 1220 non-null float64

6 Copper\_Price 1243 non-null float64

7 Copper\_Vol. 1206 non-null float64

8 Bitcoin\_Price 1243 non-null object

9 Bitcoin\_Vol. 1243 non-null float64

10 Platinum\_Price 1243 non-null object

11 Platinum\_Vol. 636 non-null float64

12 Ethereum\_Price 1243 non-null object

13 Ethereum\_Vol. 1243 non-null float64

14 S&P\_500\_Price 1243 non-null object

15 Nasdaq\_100\_Price 1243 non-null object

16 Nasdaq\_100\_Vol. 1242 non-null float64

17 Apple\_Price 1243 non-null float64

18 Apple\_Vol. 1243 non-null float64

19 Tesla\_Price 1243 non-null float64

20 Tesla\_Vol. 1243 non-null float64

21 Microsoft\_Price 1243 non-null float64

22 Microsoft\_Vol. 1243 non-null float64

23 Silver\_Price 1243 non-null float64

24 Silver\_Vol. 1196 non-null float64

25 Google\_Price 1243 non-null float64

26 Google\_Vol. 1243 non-null float64

27 Nvidia\_Price 1243 non-null float64

28 Nvidia\_Vol. 1243 non-null float64

29 Berkshire\_Price 1243 non-null object

30 Berkshire\_Vol. 1243 non-null float64

31 Netflix\_Price 1243 non-null float64

32 Netflix\_Vol. 1243 non-null float64

33 Amazon\_Price 1243 non-null float64

34 Amazon\_Vol. 1243 non-null float64

35 Meta\_Price 1243 non-null float64

36 Meta\_Vol. 1243 non-null float64

37 Gold\_Price 1243 non-null object

38 Gold\_Vol. 1241 non-null float64

dtypes: float64(30), int64(1), object(8)

memory usage: 378.9+ KB

None

Las tareas de borrar la columna de índices, y de convertir los valores son triviales, sin embargo, precisamente existen multitud de valores nulos en el dato del volumen de la plata, que precisamente es uno de los activos más correlados con el oro, ya que después del él, es el material precioso más intercambiado en los mercados financieros.

Para ello, tenemos varias opciones, como rellenar los valores con el dato anterior o posterior existente, o realizar una media de la semana, mes, etc. También existiría la opción de borrar las filas, pero como existen los datos de otros valores, no sería una buena decisión, ya que perderíamos bastante información. Por lo tanto, optaremos por utilizar el mismo valor inmediatamente posterior existente, y ya puestos, haremos lo mismo para todos los demás valores nulos del conjunto de datos.

Una vez aplicadas las correcciones, comprobamos que ya no existen valores nulos en todo el dataset:

Valores nulos por columna:

Date 0

Natural\_Gas\_Price 0

Natural\_Gas\_Vol. 0

Crude\_oil\_Price 0

Crude\_oil\_Vol. 0

Copper\_Price 0

Copper\_Vol. 0

Bitcoin\_Price 0

Bitcoin\_Vol. 0

Platinum\_Price 0

Platinum\_Vol. 0

Ethereum\_Price 0

Ethereum\_Vol. 0

S&P\_500\_Price 0

Nasdaq\_100\_Price 0

Nasdaq\_100\_Vol. 0

Apple\_Price 0

Apple\_Vol. 0

Tesla\_Price 0

Tesla\_Vol. 0

Microsoft\_Price 0

Microsoft\_Vol. 0

Silver\_Price 0

Silver\_Vol. 0

Google\_Price 0

Google\_Vol. 0

Nvidia\_Price 0

Nvidia\_Vol. 0

Berkshire\_Price 0

Berkshire\_Vol. 0

Netflix\_Price 0

Netflix\_Vol. 0

Amazon\_Price 0

Amazon\_Vol. 0

Meta\_Price 0

Meta\_Vol. 0

Gold\_Price 0

Gold\_Vol. 0

Llegados a esta punto, se hace necesaria una selección de valores, ya que no vamos a tratar todo el conjunto de columnas, si no aquellas, que por tradición pueden estar relacionadas con el oro, dependiendo del tipo de inversor. Por ejemplo, ya sabemos que la plata es otro valor muy relacionado, siendo complementario al oro, pero además, vamos a elegir otros valores que puedan ser sustitutivos; es decir, aquellos valores que los inversores eligen, en caso de no resultar atractivo el oro.

Un ejemplo de ello, sería buscar valores en tecnológicas, y precisamente tenemos el Nasdaq disponible, que aúna los principales índices bursátiles tecnológicos, e incluso para irnos al otro lado del espectro del perfil inversor, también lo haremos con BitCoin, y comprobaremos si existe algún tipo de correlación entre los diferentes valores.

Finalmente, y después de todas las modificaciones realizadas, esta sería la descripción del dataset “stock\_market\_filtered”:

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1243 entries, 0 to 1242

Data columns (total 4 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 1243 non-null datetime64[ns]

1 Bitcoin\_Price 1243 non-null float64

2 Silver\_Price 1243 non-null float64

3 Nasdaq\_100\_Price 1243 non-null float64

dtypes: datetime64[ns](1), float64(3)

memory usage: 39.0 KB

None

Descripción estadística del dataset:

Bitcoin\_Price Silver\_Price Nasdaq\_100\_Price

count 1243.000000 1243.000000 1243.000000

mean 25241.903057 21.588977 12037.318101

std 16029.009055 3.859288 2887.069742

min 3397.700000 11.772000 6904.980000

25% 10014.600000 17.998500 9298.730000

50% 23055.100000 22.758000 12381.170000

75% 37784.200000 24.512000 14563.250000

max 67527.900000 29.418000 17642.730000

Date Bitcoin\_Price Silver\_Price Nasdaq\_100\_Price

0 2024-02-02 43194.7 22.796 17642.73

1 2024-02-01 43081.4 23.236 17344.71

2 2024-01-31 42580.5 23.169 17137.24

3 2024-01-30 42946.2 23.225 17476.71

4 2024-01-29 43299.8 23.134 17596.27

... ... ... ... ...

1238 2019-02-08 3661.7 15.809 6913.13

1239 2019-02-07 3397.7 15.713 6904.98

1240 2019-02-06 3404.3 15.701 6997.62

1241 2019-02-05 3468.4 15.836 7023.52

1242 2019-02-04 3462.8 15.886 6959.96

[1243 rows x 4 columns]

También podemos observar los gráficos, con el fin de comprobar visualmente si los valores son coincidentes con la realidad. Podemos observar gráficos de diferentes sitios para revisar la equivalencia de los gráficos:

Gráfico

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura 14 - Comparativa valores propios / Investing- ORO

Interfaz de usuario gráfica, Gráfico, Gráfico de líneas

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Figura 15- Comparativa valores propios / Investing - PLATA

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

Figura 16- Comparativa valores propios / Investing – NASDAQ

### XAUSD: Los primeros indicadores. RSI 14 y SMA14

El propósito de este data set no es conseguir precios o volúmenes como hasta ahora, si no comenzar a obtener indicadores técnicos. Este dataset tiene la particularidad de incluir el RSI 14 y el SMA 14, que consisten en lo siguiente:

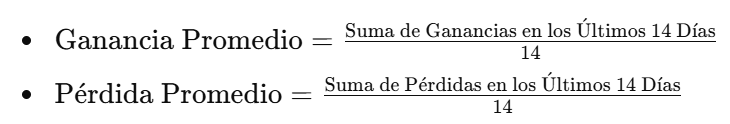
#### El RSI (Relative Strength Index):

Es un indicador de impulso que mide la velocidad y el cambio de los movimientos de precios. Se utiliza para identificar condiciones de sobrecompra o sobreventa en un mercado.

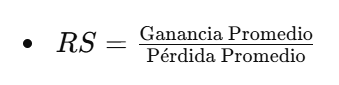
#### Fórmula Matemática:

El RSI se calcula en varias etapas, utilizando la siguiente fórmula para un período de 14 días:

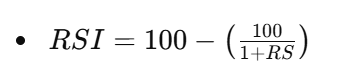
1. **Cálculo del cambio promedio en los precios de cierre:**



1. **Cálculo de la Fuerza Relativa (RS):**



1. **Cálculo del RSI:**



#### Explicación de la Fórmula:

* **Ganancia Promedio y Pérdida Promedio:** Se calculan tomando la media de las ganancias y pérdidas de los precios de cierre durante los últimos 14 días.
* **RS (Relative Strength):** Es la relación entre la ganancia promedio y la pérdida promedio.
* **RSI (Relative Strength Index):** Se escala en un rango de 0 a 100.

**Interpretación del RSI (Relative Strength Index)**

El RSI es un indicador técnico de momentum o impulso que oscila entre 0 y 100. Ayuda a identificar condiciones de sobrecompra o sobreventa en un mercado. Aquí se explica cómo interpretar los valores del RSI:

Escala del RSI (0-100):

RSI > 70 (Zona de Sobrecompra):

**Interpretación**: Cuando el RSI está por encima de 70, indica que el activo puede estar sobrecomprado o sobrevalorado. Esto sugiere que puede haber una corrección o retroceso en el precio.

**Acción** Posible: Los traders pueden considerar vender o tomar ganancias si otros indicadores también señalan una posible reversión.

RSI < 30 (Zona de Sobreventa):

**Interpretación**: Cuando el RSI está por debajo de 30, indica que el activo puede estar sobrevendido o subvalorado. Esto sugiere que puede haber un repunte o recuperación en el precio.

**Acción** Posible: Los traders pueden considerar comprar o abrir nuevas posiciones largas si otros indicadores también señalan una posible reversión.

*RSI entre 30 y 70 (Zona Neutra):*

**Interpretación**: Un RSI entre 30 y 70 generalmente indica que el activo no está ni sobrecomprado ni sobrevendido. El mercado puede estar en una tendencia estable.

**Acción** Posible: Los traders pueden usar otros indicadores o esperar señales más claras antes de tomar decisiones de trading.

**Señales Adicionales del RSI:**

Divergencias:

*Divergencia Alcista*: Ocurre cuando el precio del activo está en una tendencia bajista, pero el RSI empieza a subir. Esto puede indicar una reversión alcista inminente.

Divergencia Bajista: Ocurre cuando el precio del activo está en una tendencia alcista, pero el RSI empieza a bajar. Esto puede indicar una reversión bajista inminente.

**Cruce de Niveles Clave:**

***Cruce hacia Arriba del Nivel 30*:** Indica que el activo puede estar saliendo de una condición de sobreventa, lo cual es una señal potencial de compra.

***Cruce hacia Abajo del Nivel 70***: Indica que el activo puede estar saliendo de una condición de sobrecompra, lo cual es una señal potencial de venta.

Ejemplos de Interpretación del RSI:

Ejemplo de Sobrecompra:

**Situación**: El RSI de una acción sube por encima de 70.

**Interpretación**: La acción podría estar sobrecomprada.

**Acción**: Podrías considerar vender la acción o poner una orden de stop para proteger las ganancias.

Ejemplo de Sobreventa:

**Situación**: El RSI de una acción cae por debajo de 30.

**Interpretación**: La acción podría estar sobrevendida.

**Acción**: Podrías considerar comprar la acción o abrir una posición larga si otros indicadores confirman la señal.

Importancia del Contexto:

El RSI no debe usarse de manera aislada. Es más efectivo cuando se combina con otros indicadores técnicos y el análisis del contexto del mercado. Por ejemplo, una acción puede permanecer en una condición de sobrecompra o sobreventa durante períodos prolongados en mercados de fuerte tendencia. Por lo tanto, siempre es útil considerar el RSI junto con otros factores y herramientas de análisis técnico.

Conclusión:

El RSI es una herramienta valiosa para identificar condiciones de sobrecompra y sobreventa, así como para detectar posibles reversiones de tendencia. Sin embargo, es importante utilizarlo en combinación con otros indicadores y análisis para tomar decisiones de trading bien informadas.

#### La SMA (Simple Moving Average)

Es un indicador técnico que muestra la media aritmética de los precios de cierre de un activo durante un período específico. En este caso, se toma la media de los últimos 14 días.

#### Fórmula Matemática:

La SMA de 14 días se calcula de la siguiente manera:

**Cálculo de la SMA:**

****

#### Explicación de la Fórmula: Es sencillamente la media aritmética de los precios de cierre durante los últimos 14 días.

Interpretación de la SMA:

1. \*\*Identificación de Tendencias:\*\*

- \*\*SMA Ascendente:\*\* Indica una tendencia alcista si los precios de cierre actuales están por encima de la media móvil de 14 días. Esto sugiere que el precio está subiendo de manera sostenida.

- \*\*SMA Descendente:\*\* Indica una tendencia bajista si los precios de cierre actuales están por debajo de la media móvil de 14 días. Esto sugiere que el precio está bajando de manera sostenida.

2. \*\*Cruce de Medias Móviles:\*\*

- \*\*Cruz Dorada (Golden Cross):\*\* Ocurre cuando una SMA de corto plazo, como la SMA de 14 días, cruza por encima de una SMA de largo plazo, como la SMA de 50 días. Esto es una señal de compra.

- \*\*Cruz de la Muerte (Death Cross):\*\* Ocurre cuando una SMA de corto plazo, como la SMA de 14 días, cruza por debajo de una SMA de largo plazo, como la SMA de 50 días. Esto es una señal de venta.

3. \*\*Soporte y Resistencia:\*\*

- \*\*Soporte:\*\* En una tendencia alcista, la SMA de 14 días puede actuar como un nivel de soporte dinámico. El precio puede retroceder hacia la SMA y luego rebotar.

- \*\*Resistencia:\*\* En una tendencia bajista, la SMA de 14 días puede actuar como un nivel de resistencia dinámico. El precio puede subir hacia la SMA y luego retroceder.

### Ejemplos de Uso de la SMA de 14 días:

1. \*\*Ejemplo de Tendencia Alcista:\*\*

- \*\*Situación:\*\* La SMA de 14 días está ascendiendo y el precio del activo se mantiene por encima de esta media.

- \*\*Interpretación:\*\* Indica una tendencia alcista sostenida.

- \*\*Acción:\*\* Podrías considerar mantener posiciones largas o buscar oportunidades de compra en retrocesos hacia la SMA de 14 días.

2. \*\*Ejemplo de Cruz Dorada:\*\*

- \*\*Situación:\*\* La SMA de 14 días cruza por encima de la SMA de 50 días.

- \*\*Interpretación:\*\* Señal de una posible tendencia alcista.

- \*\*Acción:\*\* Podrías considerar abrir nuevas posiciones largas o añadir a las existentes.

3. \*\*Ejemplo de Soporte Dinámico:\*\*

- \*\*Situación:\*\* El precio de un activo retrocede hacia su SMA de 14 días durante una tendencia alcista y luego rebota.

- \*\*Interpretación:\*\* La SMA actúa como un nivel de soporte.

- \*\*Acción:\*\* Podrías considerar comprar el activo en el retroceso hacia la SMA de 14 días.

#### Cálculo a partir de Precios:

Acabamos de ver que sería posible realizar el cálculo de estas métricas, si necesidad de obtenerlas de un dataset, no obstante, podemos hacer una investigación para comprobar que los datos extraídos en este CSV son fiables.

Este dataset es bastante especial, ya que sus datos son obtenidos **cada 5 minutos**, de forma que, en este caso, tendremos que realizar tareas de agrupamiento para poder convertirlo a una escala que nos permita en un futuro fusionarlo con el resto de los datos que disponemos. En una primera observación estadística, que nos arroja los siguientes datos:

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 986004 entries, 0 to 986003

Data columns (total 7 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 time 986004 non-null object

1 open 986004 non-null float64

2 high 986004 non-null float64

3 low 986004 non-null float64

4 close 986004 non-null float64

5 rsi14 986004 non-null float64

6 sma14 986004 non-null float64

dtypes: float64(6), object(1)

memory usage: 52.7+ MB

None

Descripción estadística del dataset:

open high low close \

count 986004.000000 986004.000000 986004.000000 986004.000000

mean 1480.687806 1481.201203 1480.169858 1480.688379

std 269.235947 269.360461 269.108012 269.235929

min 1046.920000 1048.080000 1044.180000 1046.940000

25% 1252.750000 1253.080000 1252.400000 1252.750000

50% 1373.580000 1374.180000 1373.055000 1373.580000

75% 1732.680000 1733.320000 1732.010000 1732.680000

max 2138.000000 2145.140000 2125.980000 2138.410000

rsi14 sma14

count 986004.000000 986004.000000

mean 50.260961 1480.681750

std 11.309953 269.230025

min 3.000000 1048.840000

25% 42.810000 1252.740000

50% 50.250000 1373.580000

75% 57.730000 1732.660000

max 96.800000 2116.810000

realizamos las tareas análogas de anteriores dataset, excepto de la del agrupamiento, que sería la novedad de este nuevo conjunto de datos:

* Convertir la fecha
* Comprobar valores nulos
* Agrupar en días
* Eliminar los precios, ya que serían redundantes y hemos comprobado que son correctos comprobando con anteriores datasets.

Por lo que finalmente, obtenemos el dataframe “xauusd\_daily” deseado:

Información general del dataset:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 4353 entries, 0 to 4352

Data columns (total 3 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 date 4353 non-null datetime64[ns]

1 rsi14 4353 non-null float64

2 sma14 4353 non-null float64

dtypes: datetime64[ns](1), float64(2)

memory usage: 102.1 KB

None

Descripción estadística del dataset:

rsi14 sma14

count 4353.000000 4353.000000

mean 50.407766 1484.841644

std 4.177207 271.763375

min 31.470667 1056.329722

25% 47.776389 1253.075735

50% 50.326812 1380.867206

75% 52.791520 1738.184130

max 79.080667 2091.748333

date rsi14 sma14

0 2010-01-03 56.271806 1097.092500

1 2010-01-04 57.883830 1115.956454

2 2010-01-05 50.596500 1122.012893

3 2010-01-06 51.419929 1130.732456

4 2010-01-07 46.572527 1130.280890

date rsi14 sma14

4348 2023-12-25 65.670000 2058.894583

4349 2023-12-26 50.831920 2062.687101

4350 2023-12-27 56.223732 2074.287174

4351 2023-12-28 45.525290 2074.604529

4352 2023-12-29 48.272500 2066.676961

Ahora podemos ver los gráficos y comprobar como estos indicadores técnicos suelen tener una forma de onda, ya que se basan en los impulsos de mercado. Reflejan las fluctuaciones periódicas en los precios de los activos. El RSI oscila entre 0 y 100 en función de los cambios de precios, detectando sobrecompra y sobreventa, lo que genera un patrón ondulatorio a medida que el precio sube y baja. De manera similar, la SMA suaviza las variaciones de precios al calcular el promedio de los precios de cierre sobre un período determinado. Este promedio móvil se actualiza continuamente, siguiendo las tendencias y correcciones del mercado, lo que da lugar a una curva ondulante que refleja la media de los movimientos de precios.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamenteFigura 17- Curvas RSI y SMA

Sin embargo, aquí tenemos un problema bastante grave, que resolveremos más adelante; si observamos las gráficas, únicamente llegamos hasta el 2021, con lo que al fusionar dataframes, tendremos que recortar bastante los datos para adecuarlos a este rango de fechas

### El MACD (Moving Average Convergence / Divergence)

El MACD es un indicador de tendencia y momentum que muestra la relación entre dos medias móviles de los precios de cierre. Es ampliamente utilizado para identificar posibles puntos de entrada y salida en el mercado. Técnicamente es un poco más complicado que los anteriores, aunque puede ser calculado únicamente teniendo en cuenta los precios de cierre de la cotización

Sus componentes son los siguientes:

* Línea MACD: Diferencia entre la EMA (Exponential Moving Average) de 12 días y la EMA de 26 días.
* Línea de Señal: EMA de 9 días de la línea MACD.
* Histograma MACD: Diferencia entre la línea MACD y la línea de señal.

Para calcular la EMA, se hace de la siguiente forma:

Imagen que contiene Calendario

Descripción generada automáticamente

Donde Pt es el precio de cierre del día actual y k es el factor de suavizado, calculado como k=2/(n+1) y n es el número de períodos (por ejemplo, 12, 26 o 9 días).

**Línea MACD**: MACD = EMA12 - EMA26

**LÍNEA DE SEÑAL**= EMA9(Línea MACD)

**HISTOGRAMA MACD=** Línea MACD – Línea de Señal

Teniendo pues disponible la forma de construir el MACD, podemos utilizar nuestro dataset y Python para construir las columnas correspondientes desde el dataframe pandas de referencia “df\_gold”, con el que obtenemos la siguiente estructura:

<class 'pandas.core.frame.DataFrame'>

DatetimeIndex: 2513 entries, 2014-01-22 to 2024-01-18

Data columns (total 12 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Open 2513 non-null float64

1 High 2513 non-null float64

2 Low 2513 non-null float64

3 Close 2513 non-null float64

4 Volume 2513 non-null int64

5 Year 2513 non-null int64

6 Month 2513 non-null int64

7 EMA\_12 2513 non-null float64

8 EMA\_26 2513 non-null float64

9 MACD 2513 non-null float64

10 Signal\_Line 2513 non-null float64

11 MACD\_Histogram 2513 non-null float64

dtypes: float64(9), int64(3)

memory usage: 255.2 KB

Y, por supuesto obtener los gráficos:

Interfaz de usuario gráfica, Histograma

Descripción generada automáticamente

Para una mejor comprensión, podemos explicar cómo interpretar los resultados de dicho indicador:

1. **Cruces de la Línea MACD y la Línea de Señal:**
   * **Señal de Compra:** Ocurre cuando la Línea MACD cruza por encima de la Línea de Señal. Esto indica un cambio de tendencia hacia arriba y puede ser una buena oportunidad para comprar.
   * **Señal de Venta:** Ocurre cuando la Línea MACD cruza por debajo de la Línea de Señal. Esto indica un cambio de tendencia hacia abajo y puede ser una buena oportunidad para vender.
2. **Movimiento del Histograma:**
   * **Incremento en el Histograma:** Cuando las barras del histograma están aumentando, esto indica que la distancia entre la Línea MACD y la Línea de Señal está creciendo, lo que sugiere un fortalecimiento de la tendencia.
   * **Decremento en el Histograma:** Cuando las barras del histograma están disminuyendo, esto indica que la distancia entre la Línea MACD y la Línea de Señal está reduciéndose, lo que sugiere un debilitamiento de la tendencia.
3. **Divergencias:**
   * **Divergencia Alcista:** Ocurre cuando el precio del activo está haciendo nuevos mínimos, pero el MACD no está confirmando estos mínimos (es decir, el MACD está haciendo mínimos más altos). Esto puede indicar una posible reversión alcista.
   * **Divergencia Bajista:** Ocurre cuando el precio del activo está haciendo nuevos máximos, pero el MACD no está confirmando estos máximos (es decir, el MACD está haciendo máximos más bajos). Esto puede indicar una posible reversión bajista.

Adjuntamos el gráfico correspondiente a agosto de 2020, en el que podemos ver una clara señal de venta:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

### Los tipos de interés y la FED. Análisis Fundamental

Los tipos de interés de la FED son las tasas a las que los bancos pueden pedir prestado dinero a corto plazo de la Reserva Federal. La FED ajusta estos tipos de interés para controlar la economía

La relación entre los tipos de interés y la inflación es inversa. Cuando los tipos de interés aumentan, generalmente se espera que la inflación disminuya, ya que los préstamos se vuelven más costosos y el gasto y la inversión se desaceleran. Por otro lado, cuando los tipos de interés disminuyen, se espera que la inflación aumente, ya que los préstamos son más baratos y estimulan el gasto y la inversión. La Reserva Federal ajusta los tipos de interés para controlar la inflación y mantener la estabilidad económica.

Respecto al oro, tienen un impacto significativo en su precio, ya que existe, al igual que con la inflación , una relación inversa debido a que el oro es un activo que no genera intereses ni dividendos. Por esta razón, cuando la FED aumenta los tipos de interés, los activos que sí generan rendimientos (como los bonos y cuentas de ahorro) se vuelven más atractivos en comparación con el oro. Esto tiende a reducir la demanda de oro, lo que a su vez puede hacer que su precio baje.

Cabe destacar llegados a este punto que este indicador es diferente respecto a los anteriores ,ya que ahora estamos hablando de un indicador del tipo **fundamental,** un concepto que ya comentamos al principio de esta memoria.

El **análisis fundamental** se centra en evaluar el valor intrínseco de un activo mediante el examen de factores económicos, financieros y otros cualitativos y cuantitativos relevantes.

En el caso que nos ocupa respecto al oro, además de los tipos de interés, serían factores fundamentales:

Inflación: Las expectativas inflacionarias pueden aumentar la demanda de oro como un refugio seguro y reserva de valor.

Fortaleza del Dólar: Dado que el oro se cotiza en dólares, la relación inversa entre el valor del dólar y el precio del oro es un factor fundamental clave.

Condiciones Económicas Globales: La estabilidad económica y política mundial también afecta la demanda de oro como refugio seguro.

En nuestro análisis nos centraremos en el tipo de interés de la FED, y para variar el tipo de extracción, esta vez utilizaremos una API desde la web:

https://api.stlouisfed.org/fred/series/observations

Una API (Application Programming Interface) permite la interacción entre diferentes sistemas de software. Para obtener datos de una API, generalmente realizamos una solicitud HTTP (GET) y recibimos una respuesta en un formato estructurado, comúnmente JSON. Consta de tres pasos principalmente:

* Realizar la solicitud: Usamos la biblioteca requests de Python para enviar una solicitud HTTP a la API.
* Recibir la respuesta: La respuesta generalmente estará en formato JSON.
* Convertir la respuesta JSON a un DataFrame de Pandas: Usamos la biblioteca pandas para convertir el JSON a un DataFrame.

Usar una API es ideal para datos en tiempo real, CSV es conveniente para datos estáticos y locales, y los imports son cruciales para proyectos modulares y escalables

Una vez hemos extraído el código mediante la API y convertido al formato habitual del DataFrame de Pandas, obtenemos la siguiente estructura:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 25535 entries, 0 to 25534

Data columns (total 2 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 date 25535 non-null datetime64[ns]

1 value 25535 non-null float64

dtypes: datetime64[ns](1), float64(1)

memory usage: 399.1 KB

None

value

count 25535.000000

mean 4.606077

std 3.599016

min 0.040000

25% 1.790000

50% 4.250000

75% 6.250000

Y su gráfico sería el siguiente:

Gráfico, Histograma

Descripción generada automáticamente

Para realizar un análisis de estos datos, se ha de contar con cierto conocimiento experto del mundo financiero, no obstante y haciendo un resumen sencillo de los últimos años, podríamos decir que antes de la pandemia de COVID-19, los tipos de interés de la FED se mantuvieron relativamente estables y bajos, como parte de una política monetaria expansiva destinada a apoyar el crecimiento económico tras la crisis financiera de 2008. En este período, los tipos de interés rondaron entre el 1.5% y el 2.5%.

Durante la pandemia de COVID-19, la FED redujo rápidamente los tipos de interés a casi cero (0.00%-0.25%) para mitigar el impacto económico de las restricciones y cierres. Esta política extremadamente laxa ayudó a proporcionar liquidez y apoyo financiero en un momento de incertidumbre y contracción económica.

A medida que la economía comenzó a recuperarse y la inflación empezó a aumentar significativamente debido a interrupciones en la cadena de suministro y políticas fiscales expansivas, la FED inició un ciclo de aumentos de los tipos de interés para controlar la inflación. En 2022 y 2023, los tipos de interés aumentaron rápidamente, alcanzando niveles superiores al 4%, con incrementos sucesivos en un esfuerzo por enfriar la economía y reducir las presiones inflacionarias.

### Fusionando el dataset completo.

Hasta el momento disponemos de los siguientes datasets, convertidos en DataFrames de Panda, que es el formato que vamos a utilizar. Un DF de Pandas, consiste en una estructura bidimensional, similar a una tabla en una base de datos o una hoja de cálculo de Excel. Un DataFrame está compuesto por filas y columnas, donde cada columna puede ser de un tipo de datos diferente (numérico, cadena, etc.). Los DataFrames son la estructura más utilizada en Pandas y permiten realizar operaciones complejas de manipulación y análisis de datos, como filtrado, agrupamiento y transformación

Pese a que el primer dataframe fue el llamado “goldstock.csv” de Kaggle, y lo utilizamos en el primer modelo de Regresión Lineal, posteriormente, tras investigar aspectos como la fiabilidad, la decisión acerca de obtener los datos del oro, serían a cargo del uso de la herramienta que ofrece Yahoo Finance (yfinance), con ella hemos obtenido el dataframe “df\_gold”, que extendimos además con los datos obtenidos mediante funciones estadísticas con el MACD.

Posteriormente, obtuvimos los datos de otros referentes en los mercados financieros, en el que seleccionamos valores tanto para un perfil inversor conservador y complementarios al oro, como la plata y otros que podrían ser sustitutivos o preferibles por otro sector inversor más arriesgado o tecnológico, como el BTC o la cesta del Nasdaq.

Si realizamos la tarea de fusionar ambos Datasets, obtendremos el siguiente esquema:

<class 'pandas.core.frame.DataFrame'>

Int64Index: 1232 entries, 0 to 1231

Data columns (total 16 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 1232 non-null datetime64[ns]

1 Open 1232 non-null float64

2 High 1232 non-null float64

3 Low 1232 non-null float64

4 Close 1232 non-null float64

5 Volume 1232 non-null int64

6 Year 1232 non-null int64

7 Month 1232 non-null int64

8 EMA\_12 1232 non-null float64

9 EMA\_26 1232 non-null float64

10 MACD 1232 non-null float64

11 Signal\_Line 1232 non-null float64

12 MACD\_Histogram 1232 non-null float64

13 Bitcoin\_Price 1232 non-null float64

14 Silver\_Price 1232 non-null float64

15 Nasdaq\_100\_Price 1232 non-null float64

dtypes: datetime64[ns](1), float64(12), int64(3)

memory usage: 163.6 KB

El nuevo dataframe unifica los datos obtenidos en los anteriores, obviamente vamos a tener la restricción de fechas y ceñirnos al que menos rango tiene, ya que en estos casos por razones obvias no podemos inventar datos.

Por razones de comprobación, también podemos realizar una comparativa de los valores de forma independiente y escalada. De esta forma podemos observar que los patrones de precios siguen la forma esperada:

Gráfico

Descripción generada automáticamente

Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

Cuando vimos anteriormente los indicadores técnicos del SMA 14, y del RSI 14, percibimos gracias a las gráficas que el rango de datos era bastante limitado, ya que llegaba únicamente hasta 2021, cuando nuestro objetivo deberá ser por lo menos llegar hasta finales del 2023.

Si fusionásemos ambos datasets, nuestro dataframe sería ya demasiado limitado, como muestra la información que se muestra a continuación, fruto del “merge”:

<class 'pandas.core.frame.DataFrame'>

Int64Index: 706 entries, 0 to 705

Data columns (total 18 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 706 non-null datetime64[ns]

1 Open 706 non-null float64

2 High 706 non-null float64

3 Low 706 non-null float64

4 Close 706 non-null float64

5 Volume 706 non-null int64

6 Year 706 non-null int64

7 Month 706 non-null int64

8 EMA\_12 706 non-null float64

9 EMA\_26 706 non-null float64

10 MACD 706 non-null float64

11 Signal\_Line 706 non-null float64

12 MACD\_Histogram 706 non-null float64

13 Bitcoin\_Price 706 non-null float64

14 Silver\_Price 706 non-null float64

15 Nasdaq\_100\_Price 706 non-null float64

16 rsi14 706 non-null float64

17 sma14 706 non-null float64

dtypes: datetime64[ns](1), float64(14), int64(3)

memory usage: 104.8 KB

Como ocurre con el MACD, también vimos que estos indicadores técnicos están basados en formulas estadísticas de variables sobre las que ya disponemos de información, así que para resolver esta limitación, descartaremos el conjunto de datos anterior, para incluir en nuestro dataframe la información de ambos indicadores técnicos, programando dichas fórmulas en Python:

<class 'pandas.core.frame.DataFrame'>

Int64Index: 1232 entries, 0 to 1231

Data columns (total 18 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 1232 non-null datetime64[ns]

1 Open 1232 non-null float64

2 High 1232 non-null float64

3 Low 1232 non-null float64

4 Close 1232 non-null float64

5 Volume 1232 non-null int64

6 Year 1232 non-null int64

7 Month 1232 non-null int64

8 EMA\_12 1232 non-null float64

9 EMA\_26 1232 non-null float64

10 MACD 1232 non-null float64

11 Signal\_Line 1232 non-null float64

12 MACD\_Histogram 1232 non-null float64

13 Bitcoin\_Price 1232 non-null float64

14 Silver\_Price 1232 non-null float64

15 Nasdaq\_100\_Price 1232 non-null float64

16 SMA\_14 1219 non-null float64

17 RSI\_14 1219 non-null float64

dtypes: datetime64[ns](1), float64(14), int64(3)

memory usage: 182.9 KB

Quedaría, por último, agregar el indicador fundamental de los tipos de interés, teniendo finalmente nuestro dataframe “final\_merged\_df” construido:

<class 'pandas.core.frame.DataFrame'>

Int64Index: 1232 entries, 0 to 1231

Data columns (total 19 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Date 1232 non-null datetime64[ns]

1 Open 1232 non-null float64

2 High 1232 non-null float64

3 Low 1232 non-null float64

4 Close 1232 non-null float64

5 Volume 1232 non-null int64

6 Year 1232 non-null int64

7 Month 1232 non-null int64

8 EMA\_12 1232 non-null float64

9 EMA\_26 1232 non-null float64

10 MACD 1232 non-null float64

11 Signal\_Line 1232 non-null float64

12 MACD\_Histogram 1232 non-null float64

13 Bitcoin\_Price 1232 non-null float64

14 Silver\_Price 1232 non-null float64

15 Nasdaq\_100\_Price 1232 non-null float64

16 SMA\_14 1219 non-null float64

17 RSI\_14 1219 non-null float64

18 Interest\_Rate 1232 non-null float64

dtypes: datetime64[ns](1), float64(15), int64(3)

memory usage: 192.5 KB

## Implementación de la Red Neuronal LSTM

Hasta el momento, hemos visto que el mercado del oro es influenciado por una vasta gama de factores económicos, políticos y sociales, haciendo que su predicción sea un desafío complejo. Aunque el modelo de regresión lineal simples que hemos diseñado al principio, pueden ofrecer una visión inicial valiosa, su capacidad para capturar la complejidad de las relaciones entre múltiples variables es limitada.

Ahora ya estamos en disposición de utilizar un modelo más complejo utilizando redes neuronales Long Short-Term Memory (LSTM), que integrará la variedad de variables provenientes de las diferentes fuentes de datos que hemos consultado y de los diferentes indicadores técnicos que hemos implementado.

Intentaremos seguir los cánones que hemos visto durante todo el master para preparar el modelo (preparación de datos, construcción del modelo LSTM, entrenamiento y evaluación y comparación del modelo)

### Preparación de los datos

## Resultados obtenidos y comparativa

**Ampliación del Conjunto de Variables**

El nuevo modelo no se limitará a medias móviles simples del precio del oro. En su lugar, incorporará una serie de **variables explicativas adicionales** que tienen el potencial de mejorar significativamente la precisión de las predicciones. Estas variables incluirán, pero no se limitarán a:

1. **Indicadores Económicos**:
   * **Tasas de Interés**: Las tasas de interés influyen en el costo de oportunidad de mantener oro en lugar de activos generadores de intereses.
   * **Índices de Inflación**: El oro es tradicionalmente visto como una cobertura contra la inflación, por lo que los índices de inflación tienen un impacto directo en su precio.
2. **Datos del Mercado Financiero**:
   * **Índices Bursátiles**: La relación inversa entre el oro y los mercados bursátiles puede proporcionar información valiosa.
   * **Tipo de Cambio**: Las fluctuaciones en el valor del dólar estadounidense afectan el precio del oro, ya que se cotiza en dólares.
3. **Factores Geopolíticos**:
   * **Eventos Políticos y Sociales**: Crisis políticas, conflictos internacionales y otros eventos pueden causar incertidumbre económica y afectar el precio del oro.
4. **Datos Históricos y Técnicos**:
   * **Volatilidad del Mercado**: Medidas de volatilidad pueden ayudar a capturar el riesgo percibido en el mercado del oro.
   * **Indicadores Técnicos**: Como el Índice de Fuerza Relativa (RSI), el Oscilador Estocástico, entre otros.

**Integración de Datos desde Diversas Fuentes**

Para lograr una predicción precisa y robusta, se recopilarán datos desde múltiples fuentes, incluyendo:

* **Bases de Datos Económicos**: Fuentes como el Banco Mundial, la Reserva Federal y otras instituciones financieras internacionales.
* **Plataformas Financieras**: Yahoo Finance, Bloomberg, Reuters, entre otras, para obtener datos del mercado en tiempo real.
* **APIs Financieras**: APIs como Alpha Vantage, Quandl y similares para acceso programático a datos históricos y actuales.
* **Noticias y Redes Sociales**: Herramientas de análisis de sentimientos y scraping de noticias para captar el impacto de eventos recientes y la percepción del mercado.

**Unificación y Preprocesamiento del Dataset**

Una vez recopilados, los datos serán unificados en un dataset estructurado, siguiendo estos pasos:

1. **Limpieza y Normalización**:
   * **Eliminación de Datos Faltantes**: Manejo de valores nulos y faltantes para asegurar la calidad del dataset.
   * **Normalización de Variables**: Escalado y normalización de variables para garantizar que todas las características contribuyan equitativamente al modelo.
2. **Transformación de Datos**:
   * **Cálculo de Indicadores**: Derivación de indicadores técnicos y económicos adicionales a partir de los datos brutos.
   * **Creación de Características**: Generación de nuevas características que capturen relaciones temporales y dependencias entre las variables.
3. **Almacenamiento y Acceso**:
   * **Base de Datos Centralizada**: Almacenamiento del dataset en una base de datos centralizada para facilitar el acceso y el procesamiento.
   * **Herramientas de Análisis**: Uso de herramientas y bibliotecas avanzadas como Pandas, NumPy y SQL para el manejo eficiente de datos.

**Implementación del Modelo LSTM**

El modelo LSTM será implementado utilizando bibliotecas de deep learning como TensorFlow y Keras, aprovechando su capacidad para:

* **Manejar Secuencias de Datos**: Capturar dependencias temporales y secuenciales en los datos.
* **Aprender Representaciones Complejas**: Modelar relaciones no lineales entre múltiples variables.
* **Adaptarse a Cambios en el Mercado**: Mejorar continuamente mediante el entrenamiento con nuevos datos.

**Conclusión**

La implementación de este modelo avanzado de predicción del precio del oro, basado en LSTM y enriquecido con una amplia gama de variables explicativas, tiene el potencial de superar las limitaciones de los modelos tradicionales. Al integrar datos de diversas fuentes y utilizar técnicas avanzadas de machine learning, esperamos desarrollar una herramienta poderosa para la toma de decisiones de inversión en el mercado del oro

ESTA PÁGINA HA SIDO INTENCIONADAMENTE DEJADA EN BLANCO

# Conclusiones

ESTA PÁGINA SE HA DEJADO INTENCIONADAMENTE EN BLANCO

# Bibliografía

Ammous, S. (2018). *El Patrón Bitcoin.* DEUSTO.

Davies, G. (2002). *A History of Money: From Ancient Times to the Present Day.* University of Whales Press.

Weinstein, S. (1988). *Secrets for profiting in Bull and Bear Markets.* McGrawHill .

Wikipedia. (2017). *Wikipedia*. Obtenido de ISO/IEC 12207: https://es.wikipedia.org/wiki/ISO/IEC\_12207

## Anexo: instalación y despliegue.

### Disponibilidad del código fuente

Nótese que el enlace también está disponible en el menú superior de la aplicación, así como esta memoria en formato PDF.

## Uso y mantenimiento

1. La disolución del patrón oro culminó con los acuerdos de Bretton Woods en 1944, estableciendo un nuevo sistema financiero internacional [↑](#footnote-ref-1)
2. Una "commoditie" es un bien o recurso básico y fungible, como el petróleo, el trigo, o el oro, que se intercambia en mercados internacionales y cuyo precio es determinado por la oferta y la demanda. [↑](#footnote-ref-2)
3. El Greed/Fear Index, desarrollado por CNNMoney, mide el sentimiento del mercado, indicando si los inversores están motivados por el miedo o la codicia en un momento dado. [↑](#footnote-ref-3)