**Documento de trabajo**

*Aplicando modelos de ML a BTC*

1. **Presentación de la empresa, organización o problema específico**

Somos un grupo de profesionales de diversas áreas de expertise interesados en el mundo de las criptomonedas en general, y del Bitcoin (BTC) en particular, que decidimos llevar adelante un análisis de las variables que multidimensionalmente influencian la suba y baja del precio de este criptoactivo, a fin generar diversos modelos de machine learning que predigan (con cierto grado de previsibilidad) el comportamiento futuro de su cotización a corto y mediano plazo.

1. **Problema de investigación**

El precio de todo activo financiero es un fenómeno de tipo multicausal que escapa a la mera concepción de la economía racionalista (mayor beneficio al menor costo sobre el fundamento de la eficiencia como principio regulador del mercado).

Esto significa que cualquier intento de aproximación hermenéutica explicativa al dinamismo de un bien económico (cualquiera sea su naturaleza) conlleva la necesidad de recurrir al diseño y ejecución de diversos tipos de análisis del activo en cuestión:

* ***Fundamental***

Este es un tipo de análisis introducido en el mundo económico por Benjamin Graham y David Dodd en 1934, con el objetivo de poder estimar cuál es el valor real de un bien, título o acciones mediante la comparación del resultado obtenido luego del análisis de balances y el valor real del mercado. En nuestro caso concreto, dado que no se trata de un título, acción o bono de un Estado, podríamos realizar una analogía con la estimación del activo en las redes sociales (e.g. Facebook, Twitter, Instagram, Reddit, etc.), portales de noticias (Financieros y no especializados), y buscadores de internet (e.g. Google Trends) a los que toda persona puede tener acceso con un celular o laptop.

* ***Técnico***

Este complementa, y se ve complementado por (circularidad hermenéutica), el análisis fundamental. En síntesis, el análisis técnico, introducido por Charles Henry Dow a finales del siglo XIX, busca detectar la tendencia del precio de un activo según las reglas del mercado (el análisis fundamental se centraba en el estudio de la entidad emisora o portadora del activo). Para lograr un buen conocimiento del dinamismo del precio del bien estimado se suele recurrir a la identificación y análisis de tendencias, momentums, volumen, patrones, volatilidad, y del resto de los elementos de una serie de tiempo (e.g. ciclicidad, estacionariedad, residualidad, auto-correlación, etc.).

Entendiendo el nivel introductorio de esta primera aproximación al análisis del dinamismo de bajas y subas de precio del BTC, solo nos remitiremos al análisis técnico. En lo posible, y si las herramientas adquiridas y el tiempo lo permiten, también sentimos el interés de abordar ciertos elementos propios del análisis fundamental de este criptoactivo, tal como por ejemplo, la tendencia histórica de palabras específicas buscadas en Google Trends. Para el resto de los elementos de este tipo de análisis necesitaríamos tener acceso a diferentes fuentes de big data.

Se entiende que además, que este problema de estudio en particular, posee a su vez otras dos complejidades que deben ser explicitadas en este documento. La primera, el estudio de un activo financiero implica el abordaje de series temporales y la identificación de sus elementos principales. Entendiendo que este tipo de series no son abordadas en el curso, y que en la vida profesional de un científico de datos, esta situación puede repetirse regularmente, aceptamos el desafío de complementar por nuestra cuenta el abordaje de este tipo de problemas.

En segunda instancia somos conscientes de que mediante un ordenador doméstico (pensamos que igual sucede con supercomputadoras) no es posible predecir con un grado de previsibilidad alto el precio exacto de un activo mediante el uso de machine learning. Por dicho motivo, nos propusimos rediseñar nuestro proyecto final adecuado a un problema de clasificación ligado a la predicción de la suba y baja del precio, y predicciones para saber si es un buen momento para comprar o vender el activo según diversas estrategias de trading establecidas de antemano.

1. **Preguntas y objetivos de investigación**

Dada la multi-dimensionalidad del problema de investigación que nos proponemos abordar nos realizamos las siguientes preguntas orientadoras, las cuales esperamos poder contestar mediante el cumplimiento de los objetivos generales y específicos que también se explicitarán debajo.

* ***Preguntas de investigación***

1. ¿Cuáles son las diversas variables que influyen en el dinamismo del precio de BTC?
2. ¿Es predecible el precio de BTC en caso de que se conozcan dichas variables?
3. ¿Son capaces los modelos supervisados de clasificación de machine learning que se pueden ejecutar en computadoras doméstica de predecir con cierto grado suficiente de previsibilidad (>50/55%) el movimiento del precio de este criptoactivo?
4. En caso de que no se puedan correr en ordenadores domésticos, ¿es posible reducir la cantidad de variables y registros procesados mediante algún modelo de machine learning no supervisado que pueda ser ejecutado en computadoras domésticas, y a su vez resulte en predicciones de subidas-bajadas de precio con una previsibilidad superior al 50/55% de verdaderos positivos?
5. ¿Es posible mediante modelos supervisados y no supervisados de machine learning (ejecutados en computadoras domésticas) determinar el momento más adecuado para comprar o vender BTC, sobre la base de diferentes estrategias de trading que permitan superar los falsos positivos que a veces generan los indicadores técnicos?

* ***Objetivos general***

1. Diseñar diversos modelos de machine learning capaces de predecir con cierto grado de previsibilidad positiva (>50/55%) el dinamismo del precio de Bitcoin en el corto plazo, es decir, si el precio de este criptoactivo subirá o bajará en la temporalidad estudiada (1 día).

* ***Objetivos específicos***

1. Detectar diversos factores (multi-dimensionales) que afectan el precio de BTC.
2. Analizar las variables de los factores detectados a fin conocer el modo mediante el cual influencia la variación del precio de este criptoactivo.
3. Diseñar diversos modelos de machine learning capaces de predecir con cierto grado de previsibilidad positiva (>50/55%) a corto plazo, el movimiento alcista o bajista del precio de BTC.
4. Diseñar diversos modelos de machine learning capaces de predecir con cierto grado de previsibilidad positiva (>50/55%) a corto plazo, momentos propicios para la compra-venta este criptoactivo, según una serie de estrategias de trading establecidas de antemano, diseñadas sobre la base del análisis técnico de sus principales variables.
5. **Estructura de Desglose de Trabajo**

En esta sección presentamos nuestro Esquema de Desglose de Trabajo (EDT). Según los estándares internacionales del project management (i.e. PMBOK 6) cada etapa de una EDT debe dividirse en unidades organizacionales menores codificadas bajo el formato ID: i) Cuentas control (1.1., 1.2. y 1.3, etc….); ii) Paquetes de trabajo (1.1.1., 1.1.2., 1.1.3., 1.2.1., 1.2.2., 1.2.3., etc,…), y iii) Tareas de trabajo concretas (1.1.1.1., 1.1.1.2., 1.1.1.3., etc…).

En este caso concreto solo se desarrollarán las dos primeras unidades organizacionales dado el carácter macro de esta sección (en un proceso de consultoría o trabajo real se recomienda llevar a cabo el despliegue de la tercera unidad en un plan de trabajo -sección aparte-). Cada una de las unidades que se describirán debajo han sido proyectadas según los siguientes elementos: a) duración de las unidades de trabajo; b) interrelación entre las tareas diseñadas; c) la metodología de análisis y ejecución de cada unidad y sub-unidad de trabajo; y d) los hitos tentativos de informes y/o entrega de productos.

***Etapas de la EDT***

Etapa 1. Investigación preliminar

Etapa 2. Data acquisition

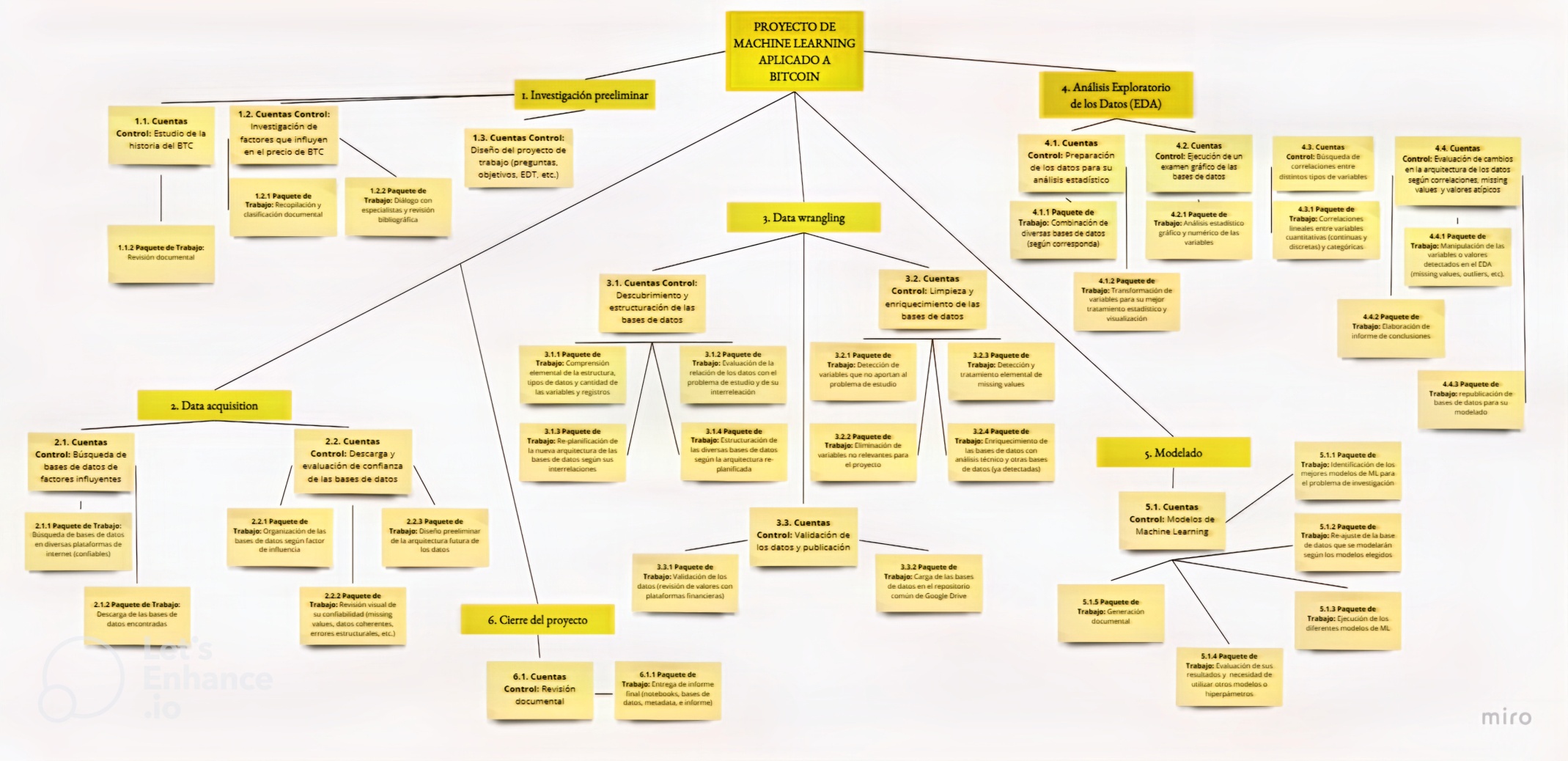
Etapa 3. Data wrangling

Etapa 4. Análisis Exploratorio de los Datos

Etapa 5. Modelos de Machine Learning

Etapa 6. Cierre del proyecto

**Estructura de Desglose de Trabajo (Gráfico)**



**Link original (permiso de visualización):** <https://miro.com/app/board/uXjVPeYRDtw=/?share_link_id=33358300391>

*La imagen original ha sido procesada para mejorar su calidad por:* [*https://letsenhance.io/boost*](https://letsenhance.io/boost)

1. **Data acquisition y data wrangling**

El proyecto cuenta con numerosos data set relacionados con diversos factores económicos que influyen en el dinamismo del precio de BTC. Si bien el detalle específico de cada base de datos se encuentra explicitado en el documento titulado “Metadata.docx” (fecha de creación, tipo de datos, diccionario, ubicación del archivo, notebooks, ubicación del notebook, bases de datos sin tratamiento, y base de datos finales), a continuación se ofrece una descripción general de los mismos en lo relativo con sus aspectos generales.

* ***Base de datos: BTC\_final***

*Fuente de los datos en bruto*: <https://www.investing.com/crypto/bitcoin/historical-data>

*Descripción del data set*: se extrajeron los datos más antiguos que permite bajar la página web (julio de 2010) hasta el día en que se descargaron los datos (20 de julio de 2022). En el notebook respectivo (BTC\_final.ipynb) se puede observar los comentarios de las variables originales y como estas fueron tratadas. La temporalidad del data set es de un 1 día y contaba con las variables: Date, Price, High, Low, Open, Vol., y Change %.

*Criterio de selección*: es la base principal que se concatenó con el resto de los data sets.

*Tratamiento de la base de datos*:

1. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
2. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
3. Se re-diseñaron las columnas “Vol.”, y “Percentage\_diff” por nomenclatura inapropiada para su posterior tratamiento (signo de %, y letras finales B, M, K).
4. Se modificó el tipo de datos (“Date”, “Vol.”, y “Percentage\_diff”).
5. Se completaron missing values en la columna Vol. con un promedio de días anteriores (3 días anteriores a los valores y 3 días posteriores).
6. Se exportó como csv y se guardó en un repositorio digital (Google Drive).
7. Se construyó la variable target cuando y se dispuso la siguientes variables categóricas: 1 (precio subió) y 0 (precio bajó).

* ***Base de datos: BTC\_metals***

*Fuente de los datos en bruto*: <https://www.investing.com/commodities/metals>

*Descripción del data set*: se extrajeron los datos más antiguos que permite bajar la página web (junio 2014) hasta el día en que se descargaron los datos (20 de julio de 2022). En el notebook respectivo (BTC\_metals.ipynb) se puede observar los comentarios de las variables originales y como estas fueron tratadas. La temporalidad del data set es de un 1 día. Se descargaron las bases de datos de los siguientes metales: oro, plata, cobre, paladio, aluminio y platino.

*Criterio de selección*: se eligieron estos metales porque se consideran los más importantes al momento de producir bienes de alto valor y porque en algunos casos se los considera reserva de valor (e.g. oro). Además en el proceso de investigación documental se adquirió el conocimiento de que había una fuerte ligazón entre el cobre y el BTC (relación que se nota en el gráfico respectivo de correlación en el precio de ambos activos).

*Tratamiento de la base de datos*:

1. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
2. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
3. Se eliminó la columna “Vol.” de cada data set porque no se analizará por ponerse el acento en el precio de los metales y sus diferencias diarias.
4. Se re-diseñó la variable “Percentage\_diff” por nomenclatura inapropiada para su posterior tratamiento (signo de %, y letras finales B, M, K).
5. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
6. Se concatenaron todas las bases individuales de cada metal.
7. Se realizó un merge con la base de datos de metales (VI) con BTC\_final.
8. Se exportó como csv y guardó en un repositorio digital (Google Drive).

* ***Bases de datos: BTC\_ind\_trend, BTC\_ind\_trend\_ET, BTC\_ind\_mom, BTC\_ind\_mom\_ET, BTC\_ind\_vol, BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_patterns\_indicator, BTC\_statistic\_functions, y BTC\_cycle\_indicators***

*Fuente de los datos en bruto*: BTC\_final.csv y librería Ta-lib (link en “Metadata.docx”).

*Descripción de los data sets*: poseen las variables y registros correspondientes a la aplicación del análisis técnico a los valores de BTC\_final.csv. Entre los indicadores técnicos los hay de tendencia (trend), momento (momentum), volatilidad, volumen, ciclicidad, patrones y funciones estadísticas básicas (e.g. desviación estándar, varianza, etc.) y complejas (e.g. (regresiones de diverso tipos, entre otros). Las bases de datos de igual nombre pero diferenciadas por un “\_ET” final implica el agregado de variables categóricas que refieren el momento de compra o venta del activo según estrategias básicas de trading (explicadas en el notebook BTC\_ind\_trading.ipynb).

*Criterio de selección*: se construyeron estas bases de datos para lograr un conocimiento acabado de la base de datos BTC\_final.csv en lo respectivo con sus diversos aspectos técnicos (tendencia, momentos, etc.), y para detectar las señales de compra y venta del activo según estrategias de trading básicas.

*Tratamiento de la base de datos*:

1. Se procesaron los diversos indicadores y se los agrupó por tipo (tendencia, momento, volatilidad, volumen, etc.).
2. Se aplicaron estrategias de trading básicas y se construyeron las variables categóricas correspondientes (comprar, vender, mantener posición).
3. Se realizó un merge con la base de datos con BTC\_final.
4. Se exportó cada base de dato como csv y se guardaron en un repositorio digital común (Google Drive).

* ***Base de datos: BTC\_Forex***

*Fuente de los datos en bruto*: <https://www.investing.com/currencies/>

*Descripción del data set*: se extrajeron los datos más antiguos que permite bajar la página web (junio 2010) hasta el día en que se descargaron los datos (20 de julio de 2022) de los pares de monedas EUR/USD (Euro y Dólar Norteamericano), USD/JPY (Yuan japonés) y USD/CHY (Yen chino). Las variables originales eran: Date, Price, High, Low, Open, Vol., y Change %. Finalmente se realizó un merge entre todas las bases de datos de Forex y con BTC\_final a fin de tratar los valores en conjunto.

*Criterio de selección*: se eligieron estos pares de monedas debido a la relevancia económica de Estados Unidos, Europa, Japón (moneda fuerte en estabilidad con el Dólar) y China.

*Tratamiento de la base de datos*:

1. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
2. Se modificó el nombre de la columna “Change %” por “Percentage\_diff”.
3. Se eliminó la columna “Vol.” de cada data set porque no se analizará por ponerse el acento en el precio de los pares de monedas y sus diferencias diarias.
4. Se re-diseñó la variable “Percentage\_diff” por nomenclatura inapropiada.
5. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
6. Se concatenaron todas las bases individuales de cada para de monedas.
7. Se realizó un merge con la base de datos BTC\_final.
8. Se exportó como csv y guardó en un repositorio digital (Google Drive).

* ***Base de datos: BTC\_US\_Finance***

*Fuente de los datos en bruto*: <https://fred.stlouisfed.org/series/FEDFUNDS> (Tasa de interés de la Reserva Federal de los Estados Unidos), <https://fred.stlouisfed.org/series/CPIAUCSL> (Tasa del índice de precios al consumidor de los EEUU), <https://www.investing.com/rates-bonds/u.s.-10-year-bond-yield-historical-data> (Bono a 10 años de los Estados Unidos).

*Descripción del data set*: se extrajeron los datos más antiguos que permite bajar la página web (junio 2010) hasta el día en que se descargaron los datos (20 de julio de 2022) de los datos de inflación, tasas de interés y precio de los bonos a 10 de los Estados Unidos. Finalmente se realizó una concatenación con la base de datos de BTC\_final.

*Criterio de selección*: se eligieron estos datos influyentes en la macroeconomía mundial a fin de conocer si tiene algún tipo de correlación y/o causalidad con el dinamismo del precio del BTC.

*Tratamiento de la base de datos*:

1. Se realizó un análisis estructural básico: tipos de datos, missing values, etc.
2. Se modificó el nombre de la columna relativa con los cambios porcentuales de los valores a “Percentage\_diff”.
3. Se modificó el tipo de datos (“Date” y “Percentage\_diff”).
4. Se concatenaron todas las bases individuales de cada tabla de datos.
5. Se realizó un merge con la base de datos BTC\_final.
6. Se exportó como csv y guardó en un repositorio digital (Google Drive).
7. **Análisis univariado, bivariado y multivariado de los diversos data sets**

En esta sección se describe a nivel general la etapa del Análisis Exploratorio de los Datos (EDA). Para el detalle específico del análisis se remite a los respectivos notebooks del repositorio de notebooks de este proyecto (Google Drive). Los elementos propios de la meta data de este proceso se encuentran desarrolladas en el archivo “Metadata.docx”.

Para cada base de datos anteriormente mencionadas se realizó un EDA univariado, bivariado y multivariado. El momento de la entrega de este documento se han analizado las bases de datos: BTC\_metals, BTC\_ind\_trend\_ET, BTC\_ind\_mom\_ET, y BTC\_patterns\_indicator. Restaría un análisis de las bases de datos BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_ Statistic\_Functions (en un solo análisis conjunto), y BTC\_US\_Finance y BTC\_Forex (otro análisis conjunto).

***Análisis univariado***

Se debe tener en cuenta que al ser una serie de tiempo algunos análisis que se deberían ejecutar en esta etapa se realizaron en la fase de análisis bivariado por sumarse la variable “Date”. Los resultados de este análisis se pueden observar en los notebooks.

En general, se llevaron adelante las siguientes acciones:

* Conocimiento de la arquitectura de los datos (cantidad de variables y registros).
* Identificación de missing values (tratamiento si se requirió).
* Reconocimiento de tipos de datos.
* Descripción estadística básica (cuartiles, mediana, media, desviación estándar, etc.)
* Gráficos de barras e histogramas para conocer distribución de datos.
* Se ejecutó un profiling a cada uno de ellos (para ejecutarlos en los notebooks se debe borrar el # delante -> generó problemas de memoria a medida que se avanzaba en el análisis bivariado y multivariado con gráficos y procesamientos más complejos).

Librearías utilizadas: pandas, matplotlib, seaborn.

***Análisis bivariado***

Este tipo de análisis se focalizó en las variables “Date”, “Price”, “Percentage\_diff” y “Target”, las cuales se las estudió de a pares con el resto de las variables. Cabe destacar que se ejecutó un desarrollo de elementos propios de una serie de tiempo: ciclicidad, estacionariedad, tendencia, y residualidad. Los resultados específicos de este análisis se pueden observar en los notebooks.

En general, se llevaron adelante las siguientes acciones:

* Gráficos de líneas, barras, violín y dispersión para la distribución temporal de los datos.
* Se ejecutó una descomposición de la serie de tiempo en sus diversos elementos. Estos se anexaron a la base de datos en análisis (ciclos, residualidad, tendencia y estacionariedad).
* Se demostró por medio de la variación de la varianza, media y estructura de auto-correlación que las series temporales en análisis no eran estacionarias.
* Se realizaron mapas de calor para conocer la correlación entre los pares de variables.
* Se ejecutaron diversos análisis de la estructura de auto-correlación (pueden servir para determinar los hiperparámetros en los modelos de ML).

Librearías utilizadas: numpy, pandas, matplotlib, seaborn, plotly, pandas\_profiling, statsmodels, pandas.plotting, statsmodels.graphics y statsmodels.graphics.tsaplots.

***Análisis multivariado***

Este tipo de análisis se focalizó en todas variables de los data sets. Los resultados específicos de este análisis se pueden observar en los notebooks correspondientes.

En general, se llevaron adelante las siguientes acciones:

* Gráficos de dispersión entre dos variables con la discriminación por colores de la variable Target. No solo permitieron corroborar los valores de correlación de los heatmaps de la fase de análisis bivariado sino también conocer la distribución de las subidas y bajadas de precio de BTC en las mismas.
* Se graficaron diversos heatmaps entre variables categóricas y numéricas (continuas y discretas). Las variables categóricas (correspondientes a Target y las diversas variables con terminación “\_ET”), en general, fueron manipuladas como variables dummies.

Librearías utilizadas: numpy, pandas, matplotlib, seaborn, plotly, pandas\_profiling, statsmodels, pandas.plotting, statsmodels.graphics y statsmodels.graphics.tsaplots.

***Conclusiones del análisis univariado, bivariado y multivariado***

Para mayor información se remite a la lectura y análisis de los diferentes notebooks. Aquí solo se condensan las conclusiones generalísimas correspondientes a cada uno de los data sets.

* *Análisis de BTC\_metals*

1. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En relación con el precio desde los 157.3 USD (2014) BTC alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 260 operaciones (mínimo valor registro del vol. del período 2014-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
2. Al momento de modelar los datos se necesitará normalizar (según el modelo).
3. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
4. Los precios de los metales presentan correlaciones positivas altas y muy altas.
5. Los precios de BTC encuentran correlaciones positivas bajas, medias y altas con distintos metales. Se destaca la correlación con el cobre (tal como anticipó la fase de revisión bibliográfica de la etapa de investigación preliminar).
6. Como era de esperar la correlación entre las variables [Price] y [Trend] es muy alta. Lo mismo sucede con las variables de [Price], [Low], [High] y [Open].
7. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.
8. Se dieron subidas y bajadas en todos los valores de estacionalidad aunque mayor densidad en valores cercanos a -30, 0, 15 y 25.
9. En el gráfico de residualidad se puede observar que cuando esta alcanzó picos mínimos se dieron bajadas de precio, y por el contrario, cuando alcanzó picos máximos hubo subidas de precio. En el gráfico de correlación (heatmap) solo se percibe una relación de |0.23| (muy débil) (correlación negativa para los valores 0.0 de la variable Target y correlación positiva para sus valores 1.0).
10. Los mayores outliers en [Percentage\_diff] se dieron con estacionariedad muy baja, y viceversa.
11. La diferencia de precio de BTC tiene un comportamiento inverso a la diferencia de precios de los metales en cuanto respecta sus respectivos máximos y mínimos. Esto podría significar que cuando los metales incrementan o decrecen en gran proporción, el BTC se toma como posible inversión o resguardo de valor (habría que estudiarlo), y viceversa. Al menos esa hipótesis tiene sentido a la inversa, es decir, cuando BTC varía poco de valor es porque el capital de los inversores va a la seguridad de los metales (oro, por ejemplo).
12. A medida que la diferencia del precio del cobre es positiva, la de BTC tiende a serlo también.

* *Análisis de BTC\_ind\_trend\_ET*

1. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Información más precisa al respecto se puede encontrar en el notebook BTC\_ind\_trading.ipynb (link disponible en: Metadata.docx).
2. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el caso del precio de los 0.00 USD (2014) BTC alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
3. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
4. Al momento de modelar los datos se necesitará normalizar los valores del data set (según el modelo que se termine aplicando).
5. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
6. Como era de esperar, cuanto más días de rolling fueron tenidos en cuenta al momento de calcular los indicadores de tendencias, más suaivazada son las líneas de sus gráficos a lo largo del tiempo.
7. Existe una correlación positiva muy fuerte entre los indicadores de tendencia y la variable [Price] (>0.90).
8. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días), la auto-correlación disminuye, como era de suponer.
9. La variable [Target] tiene una correlación débil con los valores [Comprar] y [Vender] de las diversas variables categóricas de terminación “\_ET” (<=0.30 y >-0.3, respectivamente).
10. Con respecto a la residualidad se podría pensar en una estrategia de trading en valores distintos a 0. En caso de filtrarse los valores cercanos a 0, parece haber una correlación fuerte entre [Residuals] y [Target] (positiva y negativa según los valores de Target).
11. El par ["Price"] ["Residuals"] nos indica que en general, que cuando los residuos se encuentran en el rango -2500/-5000 el precio bajó, y cuando se encuentra en el otro extremo, el precio subió.
12. El par ["Residuals"] ["Seasonality"] nos indica que en general, cuando la residualidad fue alta, sin importar si la estacionalidad subió o bajó, el precio de BTC subió. Por el contrario, cuando la residualidad fue muy baja, el precio bajó.

* *Análisis de BTC\_ind\_mom\_ET*

1. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver notebook: BTC\_ind\_trading.ipynb (link disponible en el documento: Metadata.docx).
2. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
3. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
4. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
5. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
6. Los pares [Date] y las variables de indicadores de momento denotan la posibilidad de estacionariedad. En general, no tienen tendencia. Habría que evaluar la posibilidad de ciclicidad (quizás pueda ser útil para el proyecto).
7. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.
8. El método estocástico de tipo k [STOCHRSI\_fastk] parece ser más sensible a la variación de precio que el estocástico de tipo d [STOCHRSI\_fastd]. Por lo tanto, al momento del modelado se podría utilizar solo [STOCHRSI\_fastk].
9. Se nota una muy fuerte correlación positiva entre [Price] y [ADX\_PDM] (0.91).
10. La variable [Percentage\_diff] tiene una correlación positiva débil (0.34) con [STOCHRSI\_fastk].
11. A valores más altos de [Price], la variable [RSI] subió más que lo bajó, y con [RSI] más bajos hay más bajas de precio de la variable [Price] (esto se da especialmente en valores mayores a 50). Esto podría generar una estrategia de trading que tenga el valor 50 como base de compra.
12. Con la variable [STOCHRSI\_fastk] acontece lo mismo que en XI, y es mucho más marcado que con respecto al [STOCHRSI\_fastd].
13. Con la variable [ULTOSC] acontece lo mismo que en XI y XII (por encima del valor 50). Debajo del valor=25 sería un buen indicador de que hay que vender el activo porque el precio tiene tendencia a bajar en esa franja.
14. Con valores muy bajos del indicador [WILLR] (valores negativos pero cercanos a 0), el precio de BTC subió más que lo que bajó. Con valores muy bajos (valores cercanos a -100) hubo más bajas que subidas de precio del activo en cuestión.
15. No se observaron correlaciones significativas de la variable [Percentage\_diff] con el resto de las variables numéricas excepto con las variables con [ROC] (positiva, débil, 0.23), [RSI] (positiva, débil, 0.24), [STOCHRSI\_fastk] (positiva, débil, 0.34), WILLR (positiva, débil, 0.25).
16. Solo se observan las siguientes correlaciones significativas de la variable [Target] con las variables numéricas del data set: [RSI] (débil, positiva, 0.33), [WILLR] (débil, positiva, 0.29), [STOCHRSI\_fastk] (moderada, positiva, 0.54).
17. La variable [Target] no tiene correlaciones significativas en relación con las variables categóricas del data set excepto con los registros “Comprar” de la variable [STOCHRSI\_fastk\_ET] (débil, positiva, 0.33) y con los registros “Vender” de la misma variable (débil, negativa, -0.36).
18. Con respecto a la residualidad se podría pensar en una estrategia de trading en valores distintos a 0. En caso de filtrarse los valores cercanos a 0, parece haber una correlación fuerte entre [Residuals] y [Target].
19. El par ["Price"] ["Residuals"] nos indica que en general, que cuando los residuos se encuentran en el rango -2500/-5000 el precio bajó, y cuando se encuentra en el otro extremo, el precio subió.
20. El par ["Residuals"] ["Seasonality"] nos indica que cuando la residualidad fue alta, sin importar la estacionalidad el precio tendió a subir. Por el contrario, cuando la residualidad fue muy baja, el precio bajó.
21. Se pueden deducir numerosas estrategias de trading que podrían ser modeladas:
    1. En cualquier valor de [ADX] si el valor de [ADX\_MDI] y [ADX\_MDM] se mantiene cercano a 0, el precio sube.
    2. En cualquier valor de [ADX] si el valor de [ADX\_MDI] y [ADX\_MDM] se mantiene cercano a su máximo (100), el precio baja.
    3. Si [ADX] es mayor a 10 y el [ADX\_DX] es mayor a 30/40 y menor a 75, el precio sube (caso contrario el precio tiende a bajar).
    4. Si [ADX\_MDM] y [ADX\_MDI] son menores a 20 y [ADX\_PDI] es mayor a 40, el precio sube (salvo en los extremos de ADX\_PDI]).
    5. Si [ADX\_DX] es mayor a 65 y [ADX] es mayor a 35 el precio tiende a subir.
    6. Si [ADX\_DX] es mayor a 40 y [ADX\_PDI] es mayor a 35, el precio tiende a subir.
    7. Si [AroonDown] es menor que 20 y [ADX\_PDI] es mayor que 40, el precio sube.
    8. Entre [MACD] y [MOM] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors. Se podría trazar una línea en la que queden muchas más bajas y compras de un lado y otro de la línea respectivamente. Lo mismo se observa con claridad entre [MACD] y [RSI].
    9. También se percibe que en los extremos del estocástico rápido (cualquiera de sus dos versiones), el precio sube o baja respectivamente a medida que lo hace el indicador. Lo mismo sucede en los extremos de [WILLR].
    10. Si el [MOM] es mayor a 0 y el [ULSTOC] es mayor a 40, el precio tiende a subir.
    11. Si el [MOM] es mayor a 0 y [RSI] mayor a 50, el precio parece subir.
    12. Entre los estocásticos se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors para ciertos rangos a fin de obtener zonas relativamente seguras de suba y baja de precio de BTC.
    13. Si [ADX] es mayor a 0 y [RSI] es mayor a 70, el precio tiende a subir. En caso de que [ADX] es menor a 50 y [RSI] es menor a 40, el precio tiende a bajar.
    14. Entre [RSI] y [ADX\_MDI], al igual que entre [RSI] y [ADX\_MDM] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors dado que existe correlación negativa fuerte. Se podría trazar una línea en la que queden muchas más bajas/compras de un lado y otro de líneas imaginarias respectivamente.
    15. Entre [RSI] y [ADX\_PDI] se podría pensar en una regresión lineal o un algoritmo de K-Nearest Neighbors dado que existe correlación positiva fuerte. Se podría trazar una línea en la que queden muchas más bajas/compras de un lado y otro de una línea imaginaria, respectivamente.
    16. En el par RSI y los Aroon\_Oscillator podrían generarse algunos modelos de ML para predecir aumento o descenso del precio.

* *Análisis de BTC\_patterns\_indicator*

1. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Información más precisa al respecto se puede encontrar en el notebook BTC\_ind\_trading.ipynb (link disponible en: Metadata.docx).
2. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD del precio de BTC, este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo período).
3. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
4. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos que se terminen eligiendo para aplicar).
5. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
6. Luego de realizar un gráfico de los patrones en el tiempo ([Date]) se realizó un filtrado de variables. El análisis se continuó con las variables que se detallan a continuación. El resto de las columnas se descartó porque no mostraban valores != de 0. Variables que se mantuvieron:
   1. [CDL3OUTSIDE]
   2. [CDLADVANCEBLOCK]
   3. [CDLBELTHOLD]
   4. [CDLCLOSINGMARUBOZU]
   5. [CDLDOJI]
   6. [CDLDOJISTAR]
   7. [CDLDRAGONFLYDOJI]
   8. [CDLENGULFING]
   9. [CDLGRAVESTONEDOJI]
   10. [CDLHAMMER]
   11. [CDLHANGINGMAN]
   12. [CDLHARAMI]
   13. [CDLHARAMICROSS]
   14. [CDLHIGHWAVE]
   15. [CDLHIKKAKE]
   16. [CDLLONGLEGGEDDOJI]
   17. [CDLLONGLINE]
   18. [CDLMARUBOZU]
   19. [CDLMATCHINGLOW]
   20. [CDLRICKSHAWMAN]
   21. [CDLSEPARATINGLINES]
   22. [CDLSHOOTINGSTAR]
   23. [CDLSHORTLINE]
   24. [CDLSPINNINGTOP]
   25. [CDLTAKURI]
   26. [CDLXSIDEGAP3METHODS]
7. En un gran porcentaje de los patrones filtrados se observa su aparición a la largo de toda la serie de tiempo (gráfico -> [Pattern x] y [Price]).
8. La correlación entre [Percentage\_diff] y [CDLBELTHOLD] es positiva y de 0.35 (débil). La correlación entre [Percentage\_diff] y [CDLLONGLINE] es positiva y de 0.21 (débil).
9. La variable [Target] tiene una correlación débil con [CDLBELTHOLD] de 0.38 y una correlación moderada con [CDLLONGLINE] de 0.40. En caso de los valores 0 y 1 de [Target] los signos de las correlaciones son negativo y positivo respectivamente si se los analiza por separado.
10. La variable [Target] tiene una correlación débil con [CDLSHORTLINE] de 0.37 y una correlación moderada con [CDLSPINNINGTOP] de 0.39. En caso de los valores 0 y 1 de [Target] los signos de las correlaciones son negativo y positivo respectivamente si se los analiza por separado.
11. Luego de realizar el análisis de [Seasonality] y de los patrones filtrados arriba, se volvió a hacer una selección de aquellos patrones que muestran con claridad que al cumplirse la forma invocada no había falsos positivos o negativos. Así el nuevo listado de patrones que se pueden utilizar para modelar con ML son los siguientes:
    1. [CDL3OUTSIDE]
    2. [CDLADVANCEBLOCK]
    3. [CDLBELTHOLD]
    4. [CDLCLOSINGMARUBOZU]
    5. [CDLDOJISTAR]
    6. [CDLENGULFING]
    7. [CDLHARAMI]
    8. [CDLHARAMICROSS]
    9. [CDLHIGHWAVE]
    10. [CDLHIKKAKE]
    11. [CDLLONGLINE]
    12. [CDLMARUBOZU]
    13. [CDLSEPARATINGLINES]
    14. [CDLSHOOTINGSTAR]
    15. [CDLSHORTLINE]
    16. [CDLSPINNINGTOP]
    17. [CDLXSIDEGAP3METHODS]

* *Análisis de BTC\_multiple\_indicators*

*Se analizó en un solo notebook las siguientes bases de datos:*

*BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_statistic\_functions, y BTC\_ind\_cycle\_indicators, y variables de una serie de tiempo obtenidas al momento del estudio concreto.*

1. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver notebook: BTC\_ind\_trading.ipynb (link disponible en el documento: Metadata.docx).
2. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo).
3. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
4. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
5. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
6. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.
7. Existe una correlación positiva muy fuerte entre [TSF], [Price], [Low], [High] y [Open], [Lineraregrregresion], y [Linearincetercept].
8. Se nota una muy fuerte correlación positiva entre [Price] y [ATR] (volatilidad).
9. Se nota una fuerte correlación positiva entre [Price], [VAR], y [STDDEV].
10. Hay una correlación negativa débil entre [Vol.] y [ADOSC] (volumen).
11. La residualidad y ciclicidad de BTC se encuentran fuertemente correlacionadas (no así la estacionariedad).
12. Se pueden deducir numerosas oportunidades para aplicar modelos de ML y de generar estrategias de trading que también podrían ser modeladas a partir de la relación entre las variables:
    1. El par [ADOSC] - [HT\_DCPERIOD] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0,4 y 1, y [HT\_DCPERIOD] entre 0 y 20, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
    2. El par [ADOSC] - [HT\_DCPHASE] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0 y 1, y [HT\_DCPHASE] entre 100 y 200, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
    3. El par [ADOSC] - [HT\_PHASOR\_quadrature] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo entre 0 y 1,5, y [HT\_PHASOR\_quadrature] entre -500 y 500, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
    4. El par [ADOSC] - [HT\_SINE\_sine] refleja que en determinadas franjas el precio subió y en otras bajó. Ej. mientras [ADOSC] se mantuvo por encima de 0, el precio de BTC subió. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC).
    5. El par [HT\_PHASOR\_quadrature] - [HT\_SINE\_sine] refleja que en determinadas franjas el precio subió y en otras bajó. Se podría pensar en una estrategia de trading fundada en KNN.
    6. El par [HT\_PHASOR\_quadrature] - [HT\_PHASOR\_leadsine] refleja que en determinadas franjas el precio subió y en otras bajó. Se podría pensar en una estrategia de trading fundada en KNN.
    7. El par [Linearslope] - [Linearangle] refleja que en determinadas franjas el precio subió y en otras bajó. Esto podría ayudar a pensar una estrategia de trading (compra y venta de BTC) mediante algún modelo de KNN (por su modo de predecir).
    8. Se nota que la variable [Residuals] podría utilizarse con diversos modelos para predecir la subida o bajada de precio con [Price], [Percentage\_diff], [VAR], [STDDEV], [TSF], [Trend], [Seasonality] y [BTC\_cycle]. Lo mismo acontece con BTC\_cycle. Se hipotetiza que modelos de clasificación podrían ofrecer resultados alineados con los objetivos de este estudio (KNN, SVM, etc).
    9. Se nota que la variable [Residuals] podría utilizarse con diversos modelos para predecir la subida o bajada de precio con [BTC\_cycle], [HT\_PHASOR], [HP\_SINE\_sine], [HP\_SINE\_leadsine]. Se hipotetiza que modelos de clasificación podrían ofrecer resultados alineados con los objetivos generales y específicos de este estudio (KNN, SVM, etc).

* *Análisis de BTC\_US*

*Se analizó en un solo notebook las siguientes bases de datos:*

*BTC\_Forex, BTC\_US\_Finance y variables de una serie de tiempo obtenidas al momento del estudio concreto.*

1. Los valores NaN en cada una de las columnas corresponde con la cantidad de valores que se requirieron para ejecutar el algoritmo de análisis técnico de la variable [Price]. Para mayor información ver los notebooks: BTC\_Forex.ipynb y BTC\_US\_Finance.ipynb (link disponible en el documento: Metadata.docx).
2. Se nota un incremento notable en el precio y el volumen total de transacciones diarias. En el primer caso, desde los 0.00 USD (2010) del precio de BTC este alcanzó un valor de 68990.6 USD (2021). Al día de la fecha (2022), su valor se redujo a 25.000 USD. De 80 operaciones (mínimo valor registro del vol. del período 2010-2022) se alcanzó el record de 4470000000 (última fase del mismo).
3. La mayor subida diaria de precio fue de 336.84% y su caída mínima de -57.21%.
4. Al momento de modelar los datos se necesitará normalizar el data set (según los modelos de ML que se apliquen para predecir).
5. El par de las variables [Date] y [Price] es una serie de tiempo no estacionaria: su mediana, varianza y estructura de auto-correlación no son constantes.
6. La estructura de auto-correlación de la serie tiempo en lo que respecta a la variable [Price] se ajusta muy bien a un día de diferencia. A medida que se va agrandando el número de lags (dif. días) la auto-correlación disminuye, como era de suponer.
7. Se percibe un incremento notable en las bolsas de valores de US hasta finales del 2021.
8. En lo que respecta a las diferencias de precio intradiaria en las bolsas de valores se percibe un máximo y mínimo histórico a principios del año 2020 con el inicio de la pandemia del COVID-19.
9. La inflación acumulada (ACPI) creció exponencialmente a partir del 2020 (pandemia).
10. Durante la pandemia del COVID-19 se mantuvieron en valores cercanos a cero las tasas de interés de los US.
11. Se puede observar como el USD se empodera a partir del 2022 con las tres monedas de comparación (EUR, JPY, CNY).
12. Se nota una muy fuerte correlación positiva entre [Price], [Open], [High], [Low], [Price\_SP500], [Price\_Nasdaq], [Price\_DJ].
13. Se nota una correlación negativa débil entre [Price] y [Price\_US\_Bond].
14. Se nota una correlación positiva fuerte entre [Price] y [I\_Percentage\_diff\_ACPI].
15. Se nota una correlación positiva moderada entre [Price] y [M\_Percentage\_diff \_MCPI].
16. Se nota una correlación negativa débil entre [Price] y [EUR\_USD].
17. Se nota una correlación positiva débil entre [Price] y [USD\_JPY].
18. Se nota una correlación positiva débil entre [Vol.] y [I\_Percentage\_diff\_ACPI].
19. Se nota una correlación positiva moderada entre [Trend] y [I\_Percentage\_ diff\_ACPI].
20. Se nota una correlación muy fuerte entre los precios de las tres bolsas de EEUU.
21. Entre la variable [Price] y [Percentage\_diff\_NASDAQ] se puede percibir una relación entre subidas y bajadas de precio. A mayor diferencia de NASDAQ el precio tendió a subir, mientras que a diferencia más negativa del valor de NASDAQ, el precio tendió a bajar.
22. Dado que no se observan patrones definidos con claridad más que algunas zonas de densidad de bajas y ventas del precio del BTC, se podría pensar en modelos supervisados de ML de clasificación (SVM, KNN, etc.).
23. **Data sets para modelar con ML**

De la ejecución de diversos análisis univariado, bivariado y multivariado de los diversos data sets en cuestión se obtuvieron algunas conclusiones que permiten reducir sus variables a un número menor, lo cual terminará favoreciendo el desarrollo de modelo de ML más eficientes (al menos en teoría dada la menos cantidad de datos que debe procesar).

Los criterios de selección de las variables numéricas (continuas y discretas) y categóricas que se conservan para el modelado de ML son aquellas que:

* Guardan una correlación débil, moderada, fuerte o muy fuerte con las variables [Price], [Percentage\_diff] y/o [Target].
  + En el caso de correlaciones negativas la escala aplicada es la siguiente: -0.4<x<=-0.2 (débil), -0.6<x<=-0.4 (moderada), -0.8<x<=-0.6 (fuerte), y -1<= x<=-0.8 (muy fuerte).
  + En el caso de correlaciones positivas la escala aplicada es contraria en signo: 0.2<=x<0.4 (débil), 0.4<=x<0.6 (moderada), 0.6<=x<0.8 (fuerte), 0.8<=x<=1 (muy fuerte).
* Ofrecen la posibilidad de generar modelos de ML relacionados que favorezcan la discriminación de subidas y bajadas de precio más allá de la correlación que exista con las variables [Price], [Percentage\_diff] y/o [Target]. Serían casos de correlaciones indirectas (para llamarlas de algún modo).
* Permiten eficientizar las estrategias de trading respectivas eliminando falsos negativos y falsos positivos.

Al momento de diseñar los modelos de ML se modificarán las bases de datos en cuestión. Debajo se especifican las variables que se conservarán cuando se proceda con el modelado.

* ***BTC\_metals.csv***

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Date]
  + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Vol.]
  + [Target]
  + [Percentage\_diff']
  + [Price\_gold]-> por estimarse mundialmente como reserva de valor
  + [Perc\_diff\_gold]-> por estimarse mundialmente como reserva de valor
  + [Price\_copper]
  + [Perc\_diff\_copper]
  + [Price\_aluminium]-> por guardar correlación positiva moderada (0.34) con [Vol.].
  + [Perc\_diff\_aluminium]-> por guardar correlación positiva moderada (0.34) con [Vol.].
  + [Trend]
  + [Residuals]

El resto de los metales, a pesar de guardar correlación fuerte con la variable [Price] se descartaron dado que la mayor correlación con esta variable la lleva el cobre [Price\_ copper] con correlación de tipo positiva y muy fuerte (0.88).

* ***BTC\_ind\_trend.csv***

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Vol.]
  + [Percentage\_diff']
  + [Target]
  + [EMA\_7d]
  + [HT\_TRENDLINE]
  + [BBANDS\_upper\_band]
  + [BBANDS\_middle\_band]
  + [BBANDS\_lower\_band]
  + [PSAR]
  + [EMA\_7d\_ET]
  + [HTTL\_ET]
  + [PSAR\_ET]
  + [Year]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Month]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Day]-> se conserva por si es necesario ajustar la temporalidad de los modelos de ML.
  + [Trend]

El resto de las medias móviles (distinta temporalidad) se eliminaron por mejor adaptación al precio de las medias de menor tiempo. Además se tuvo en cuenta que la mayor auto-correlación de la serie de tiempo [Date] y [Price] se da con menor cantidad de lags. Se prefiere la media móvil exponencial (EMA) a la media móvil simple (SMA) por ajustarse mejor al precio dado que pondera en su fórmula el peso de los últimos valores de la serie.

* ***BTC\_ind\_mom.csv***

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Vol.]
  + [Percentage\_diff']
  + [Target]
  + [ADX]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ADX\_MDI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ADX\_MDM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ADX\_DX]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ADX\_PDI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ADX\_PDM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [Aroon\_Oscillator]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [MACD] -> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [MOM]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [ROC]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [RSI]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [STOCHRSI\_fastk]
  + [ULTOSC]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [WILLR]-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [STOCHRSI\_fastk\_ET]
  + [Year]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Month]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Day]-> se conserva por si es necesario ajustar la temporalidad de los modelos de ML.
  + [Trend]
  + [Residual]

Se prefiere el método estocástico rápido de tipo k al d porque se ajusta mejor al precio por fórmula propia.

* ***BTC\_ind\_mom.csv***

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Vol.]
  + [Percentage\_diff']
  + [Target]
  + [CDL3OUTSIDE]
  + [CDLADVANCEBLOCK]
  + [CDLBELTHOLD]
  + [CDLCLOSINGMARUBOZU]
  + [CDLDOJISTAR]
  + [CDLENGULFING]
  + [CDLHARAMI]
  + [CDLHARAMICROSS]
  + [CDLHIGHWAVE]
  + [CDLHIKKAKE]
  + [CDLLONGLINE]
  + [CDLMARUBOZU]
  + [CDLSEPARATINGLINES]
  + [CDLSHOOTINGSTAR]
  + [CDLSHORTLINE]
  + [CDLSPINNINGTOP]
  + [CDLXSIDEGAP3METHODS]
  + [Year]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Month]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Day]-> se conserva por si es necesario ajustar la temporalidad de los modelos de ML.
  + [Trend]
  + [Residual]

La explicación de porqué se conservaban estos patrones fue realizada arriba.

* ***BTC\_multiple\_indicators (conjunto de bases de datos):***

*BTC\_ind\_vol\_ET, BTC\_ind\_volat, BTC\_statistic\_functions, y BTC\_ind\_cycle\_indicators, y variables de análisis de serie de tiempo (Seasonality, Trend, etc) obtenidas al momento del estudio en particular.*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Vol.]
  + [Percentage\_diff']
  + [Target]
  + [ADOSC] (BTC\_ind\_volat\_ET)
  + [ATR] (BTC\_ind\_volat)
  + [Linearregression] (BTC\_statistic\_functions)
  + [Linearintercept] (BTC\_statistic\_functions)
  + [Linearslope] (BTC\_statistic\_functions)
  + [Linearangle] (BTC\_statistic\_functions)
  + [VAR] (BTC\_statistic\_functions)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [STDDEV] (BTC\_statistic\_functions) )-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [HT\_SINE\_sine] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [HT\_PHASOR\_leadsine] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [HT\_DCPHASE] (BTC\_ind\_cycle\_indicators)-> se conserva por posible estrategia de trading (descrita arriba en 6.)
  + [Year]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Month]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Day]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Seasonality] (análisis de serie de tiempo)
  + [Trend] (análisis de serie de tiempo)
  + [Residual] (análisis de serie de tiempo)
  + [BTC\_cycle] (análisis de serie de tiempo)
* ***BTC\_multiple\_indicators (conjunto de bases de datos):***

*BTC\_Forex, BTC\_US\_Finance, y variables de análisis de serie de tiempo (Seasonality, Trend, etc) obtenidas al momento del estudio en particular.*

Cumpliendo los criterios generales de selección arriba mencionados y debido a algunos criterios extras que se describirán oportunamente se conservan las siguientes variables:

* + [Price]
  + [Open]
  + [High]
  + [Low]
  + [Percentage\_diff']
  + [Target]
  + [Price\_SP500] (BTC\_US\_Finance)
  + [Price\_NASDAQ] (BTC\_US\_Finance)
  + [Price\_DJ] (BTC\_US\_Finance)
  + [Price\_US\_Bond] (BTC\_US\_Finance)
  + [I\_Percentage\_diff\_ACPI] (BTC\_US\_Finance)
  + [M\_Percentage\_diff\_MCPI] (BTC\_US\_Finance)
  + [FEDfunds] (BTC\_US\_Finance)
  + [Percentage\_diff\_USD\_JPY] (BTC\_Forex)
  + [Year]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Month]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Day]-> se conserva por si es necesario ajustar la temporalidad de los modelos.
  + [Seasonality] (análisis de serie de tiempo)
  + [Trend] (análisis de serie de tiempo)
  + [Residual] (análisis de serie de tiempo
  + [BTC\_cycle] (análisis de serie de tiempo)

1. **Modelado de ML**

El próximo paso es dar inicio a la etapa 5 (Modelado) de este proyecto. Se tendrán en cuenta modelos de ML supervisados (en primera instancia) y no supervisados (en segundo término) a fin de poder cumplimentar con el objetivo general descrito arriba, y con el objetivo específico de hacer más eficientes las estrategias de trading evitando falsos positivos y negativos.

Así se espera poder predecir la subida y bajadas futuras de precio de BTC en el corto plazo y si conviene o no comprar o vender BTC (estrategias de trading del objetivo específico respectivo).

1. **Metadata**

La metadata de este proyecto se puede encontrar en el archivo: “Metadata.docx”. Link al repositorio de Google Drive: <https://docs.google.com/document/d/12yfdnVTwmsX9Y6gXGe3_Ew-xuha35iBk/edit?usp=sharing&ouid=101791130377378188501&rtpof=true&sd=true>