

Análisis de sentimientos multimodal

Sara Olías Zapico

22 de septiembre de 2022

Índice

1	Recopilación de datos	2
2	Estudio de la correlación de las variables	11
	Bibliografía	13

1. Recopilación de datos

Lo primero que debemos hacer para comenzar el desarrollo del trabajo es recoger el histórico de datos de BTC necesario. Para ello hemos utilizado la librería *ccxt* de python. Esta librería contiene una colección de extracciones de diferentes criptomonedas. Para nuestro trabajo hemos extraído las correspondientes con BTC/USDT.

Hemos realizado dos extracciones para las diferentes pruebas: en la primera se obtiene una fila por día desde el 17 de Agosto de 2017; en la segunda, se obtiene una fila por hora desde el 17 de Agosto de 2017 a las 04:00. En ambos casos las columnas extraídas son: valor de cierre, fecha, valor máximo alcanzado en ese periodo, valor mínimo y volumen medio por transacción.

Como estos indicadores eran insuficientes, hemos añadido nuevos indicadores a nuestro conjunto inicial y, de esta manera, hemos creado tres conjuntos de datos para poder comparar como cambia el rendimiento de nuestros modelos conforme se aumenta el número de indicadores.

CONJUNTO DE DATOS 1

A las columnas iniciales le añadimos las siguientes:

- **Anterior:** precio de cierre del periodo anterior.
- **Diferencia:** diferencia de precio de cierre entre el periodo anterior y el actual.
- **Anterior-1:** precio de cierre de dos periodos atrás.
- **Diferencia-1:** diferencia de precio de cierre entre el periodo anterior y el de antes de éste.
- **Subida:** tiene el valor 1 si del periodo anterior al actual el precio de cierre ha subido, 0 si ha bajado.
- **Subida-1:** tiene el valor 1 si de dos periodos atrás al anterior el precio de cierre ha subido, 0 si ha bajado.

Al añadir estas columnas, algunos datos se quedaban en blanco, ya que no teníamos los datos de los días anteriores de las dos primeras filas; por lo que hemos tenido que eliminar estas filas. Adjuntamos una extracción del primer conjunto de datos creado:

	Open	High	Low	Close	Volume	Anterior	Diferencia	Anterior-1	Diferencia-1	Subida	Subida-1
datetime											
2017-08-17 06:00:00+00:00	4330.29	4345.45	4309.37	4324.35	7.229691	4315.32	9.03	4308.83	6.49	1	1
2017-08-17 07:00:00+00:00	4316.62	4349.99	4287.41	4349.99	4.443249	4324.35	25.64	4315.32	9.03	1	1
2017-08-17 08:00:00+00:00	4333.32	4377.85	4333.32	4360.69	0.972807	4349.99	10.70	4324.35	25.64	1	1
2017-08-17 09:00:00+00:00	4360.00	4445.78	4360.00	4444.00	10.763623	4360.69	83.31	4349.99	10.70	1	1
2017-08-17 10:00:00+00:00	4441.10	4470.00	4399.81	4460.00	24.865399	4444.00	16.00	4360.69	83.31	1	1
...
2022-06-01 16:00:00+00:00	30742.59	30744.19	30291.67	30419.62	14127.395730	30742.59	-322.97	31172.49	-429.90	0	0
2022-06-01 17:00:00+00:00	30419.62	30429.71	29901.29	30179.67	11518.461400	30419.62	-239.95	30742.59	-322.97	0	0
2022-06-01 18:00:00+00:00	30179.68	30445.00	30119.41	30167.69	5554.231140	30179.67	-11.98	30419.62	-239.95	0	0
2022-06-01 19:00:00+00:00	30167.69	30351.74	30002.92	30126.72	6735.208380	30167.69	-40.97	30179.67	-11.98	0	0
2022-06-01 20:00:00+00:00	30126.71	30162.44	29301.00	29636.09	13421.548650	30126.72	-490.63	30167.69	-40.97	0	0

41865 rows × 11 columns

Figura 1: Extracción del conjunto de datos 1.

CONJUNTO DE DATOS 2

Para crear este conjunto de datos, vamos a agregar al conjunto de datos anterior los siguientes indicadores:

- **Precio Medio Ponderado por Volumen(VWAP):** este indicador nos muestra la relacion que hay entre el precio del BTC y el volumen de las operaciones diarias.
- **SMA10:** una media móvil simple (o aritmética) es una media móvil aritmética calculada sumando los elementos de una serie temporal y dividiendo este total por el número de períodos de tiempo. Media SMA con 10 periodos anteriores.
- **SMA20:** Media SMA con 20 periodos anteriores.
- **SMA55:** Media SMA con 55 periodos anteriores.
- **EMA10:** las medias móviles exponenciales dan más peso a los períodos más recientes. Esto los hace más seguras que la SMA y por lo tanto pueden ser usadas para crear una mejor estrategia de trading con medias móviles. Media EMA con 10 periodos anteriores
- **EMA20:** Media EMA con 20 periodos anteriores.
- **EMA55:** Media EMA con 55 periodos anteriores.
- **VMA10:** sirven para frenar el promedio cuando los precios están en el período de consolidación, para evitar señales malas y acelerar el promedio cuando el mercado está en tendencia para aprovechar al máximo los precios en tendencia. Media VMA con 20 periodos anteriores.

- **VMA20:** Media VMA con 20 periodos anteriores.
- **VMA55:** Media VMA con 20 periodos anteriores.

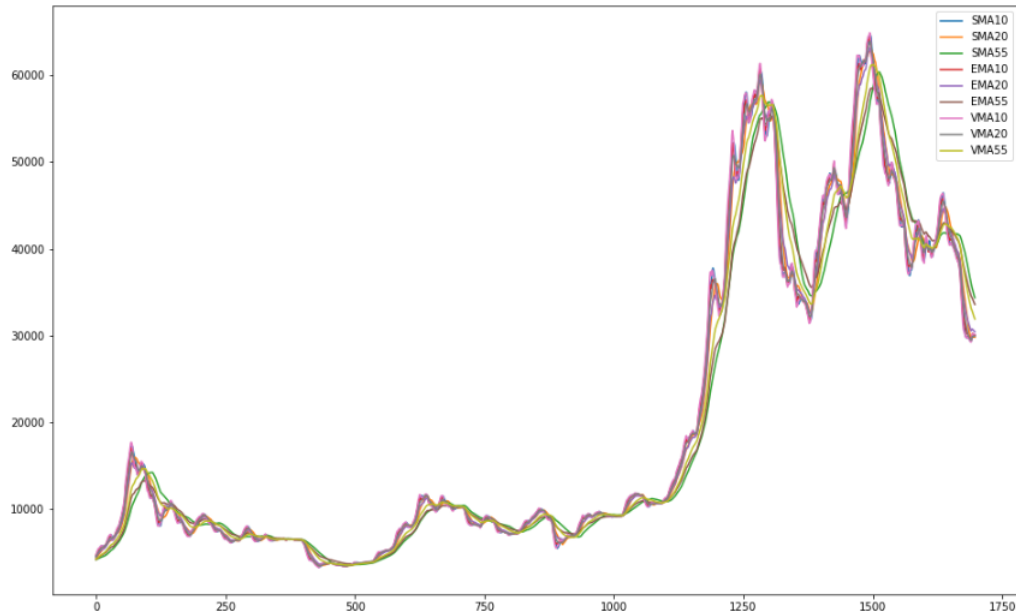


Figura 2: Representación grafica de las medias moviles para el conjunto de datos por días.

Como podemos ver, en el conjunto de datos 2 hemos añadido varias medias móviles. Las medias móviles son uno de los indicadores técnicos más conocidos ya que son fáciles de usar. En estadística, las medias móviles son unas medidas que van tomando los datos referentes a un periodo temporal definido. Conforme se va avanzando en el tiempo, los datos a evaluar también avanzan, ajustándose y actualizándose de forma sucesiva.

Cuando hablamos de series temporales, las medias móviles son muy útiles. Cuanto mayor sea el periodo, la media móvil es más lenta en dar señales, pero las señales serán mas fiables; es por ello que las medias móviles con periodos mas altos, van a ser más útiles para los datos por horas, ya que tenemos mas registros.

Hay varios tipos de medias móviles, en nuestro caso vamos a estudiar la media móvil aritmética (*SMA*), la media móvil exponencial (*EMA*) y la media móvil variable (*VMA*). Veamos cada una de ellas por separado [1]:

– Media Móvil Simple o Aritmética (*SMA*):

Una media móvil simple (o aritmética) es una media móvil aritmética calculada sumando los elementos de una serie temporal y dividiendo este total por el número

de períodos de tiempo. Este es el tipo mas simple de todos los promedios móviles. En esta media, los pesos son los mismos para todos los elementos (si tenemos un periodo n , el peso asociado será $\frac{1}{n}$). Generalmente esta media es útil para identificar la dirección de la tendencia.

$$SMA = \frac{(\text{Suma de los } n \text{ últimos datos})}{(\text{Número total de periodos})}$$

- **Media Móvil Exponencial (EMA)**: Cuando hay picos en el precio a estudiar, las medias móviles simples no funcionan muy bien. Las medias móviles exponenciales dan más peso a los periodos más actuales, su cálculo es:

$$EMA \text{ actual} = [\text{Precio de cierre} - EMA \text{ (día anterior)}] * \text{multiplicador} + EMA \text{ (día anterior)}$$

donde

$$Multiplicador = 2(\text{cantidad de periodos seleccionados} + 1)$$

La media móvil exponencial es más rápida en reaccionar que la SMA, esto se debe a que cada término del periodo de la media móvil tiene un peso exponencialmente mayor que el término anterior.

- **Meda móvil variable (VMA) [2]**:

La *VMA* es una media móvil ponderada exponencialmente, para ello se utiliza un Índice de Volatilidad (*VI*) que mide la rapidez o lentitud con la que los precios cambian a lo largo del tiempo.

El propósito principal del *VMA* es frenar la media cuando los precios se están consolidando y acelerar esta media cuando los precios están en tendencia.

$$VMA = (\alpha * VI * \text{Precio de cierre}) + (1 - (\alpha * VI)) * VMA[1]$$

$$\alpha = \frac{2}{(N + 1)}$$

N = periodo sobre el que se realiza la media

$$VMA[1] = VMA \text{ del periodo anterior}$$

Podemos ver las principales diferencias de ajuste en esta gráfica, creada a partir de las medias de los datos de un día (los últimos 100 registros):

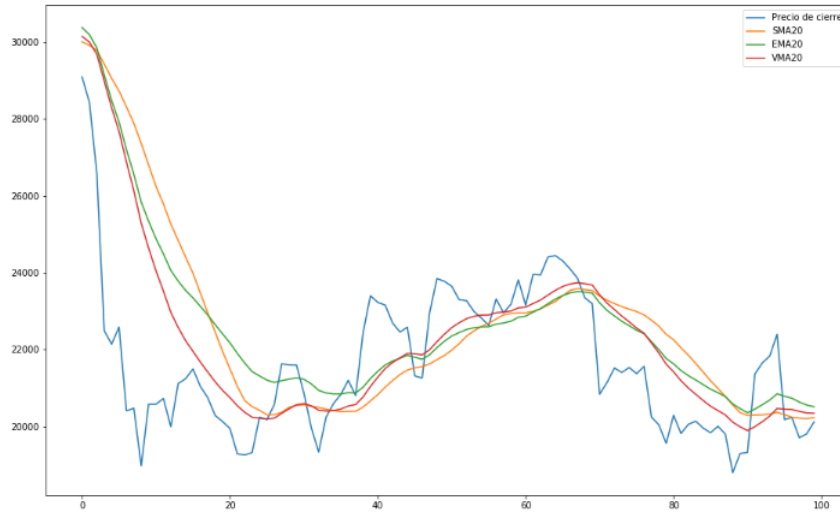


Figura 3: Representacion SMA20, EMA20 y VMA20 frente al precio de cierre

CONJUNTO DE DATOS 3

Para este tercer conjunto, vamos a añadir al conjunto de datos 2, varios indicadores:

- **Indicador MACD (Moving Average Convergene Divergence) [3] :**

Mide la convergencia y divergencia en el tiempo de dos medias móviles del precio de un activo. En otras palabras el MACD señala, en cada momento, la separación entre el valor de dos medias móviles con diferente período de cálculo. Para el cálculo del MACD utilizamos la media móvil exponencial (EMA), primero calculada en un periodo corto y después en un periodo medio, y calculamos la diferencia entre ambas. Normalmente, para la media corta se emplean 12 períodos y, para la otra media, 26 períodos. ($MACD = EMA(12) - EMA(26)$).

Después de obtener el MACD, se calcula, a su vez, su media móvil exponencial. Para realizar este cálculo se suele emplear una media de 9 períodos. Esta media móvil se denomina Señal. El histograma MACD representa la diferencia entre MACD y su EMA de 9 días, la línea de señal.

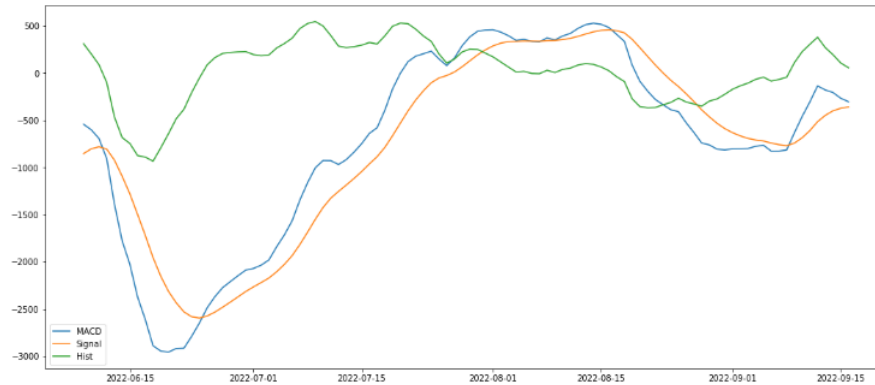


Figura 4: Representacion MACD

■ Indicador KDJ :

Las líneas K y D del oscilador estocástico (señalan una tendencia inminente cuando ocurren divergencias alcistas y bajistas.) y la J muestra la divergencia del valor D de la K . Gráficamente:

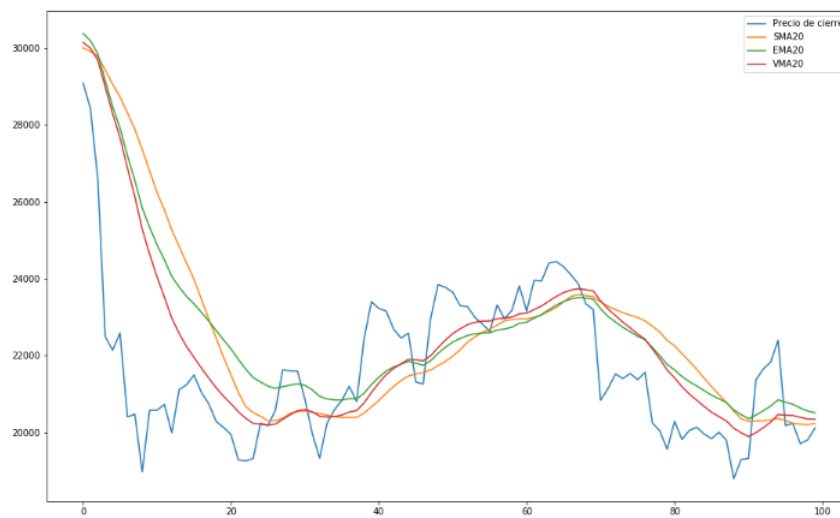


Figura 5: Representacion KDJ

■ Bandas de Bollinger [4]:

El indiador bandas de Bollinger utiliza una medida estadística conocida como la desviación estándar para determinar dónde podría tener lugar un posible nivel de

soporte o resistencia. Estas bandas se conforman por dos líneas, una por encima y otra por debajo de una media central.

La media central es una SMA de periodo N ; la banda superior se dibuja en X desviaciones estándar por encima de la media anterior y la banda inferior, se dibuja en X desviaciones estándar por debajo de la media. Nosotros hemos escogido hacer la media SMA con 20 periodos y con una desviación estándar multiplicada por 2. Para estos parámetros, las estadísticas muestran que el 95 del precio debe permanecer dentro de las Bandas de Bollinger.

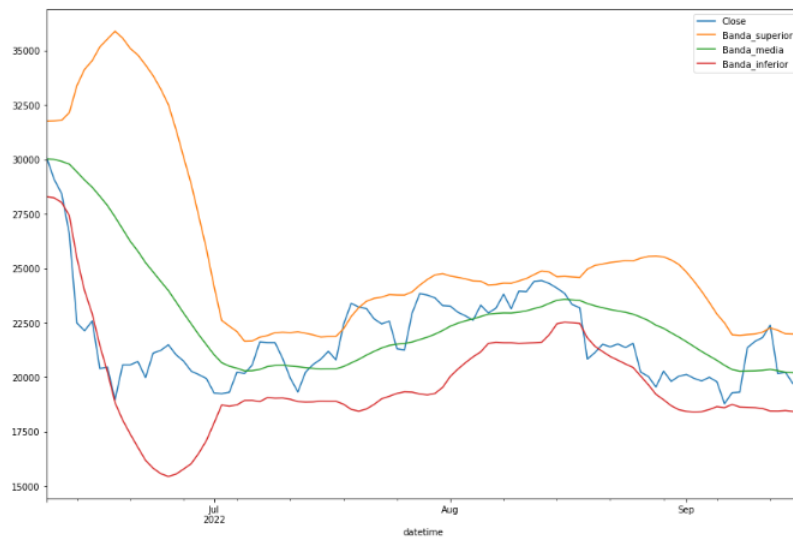


Figura 6: Representacion Bandas de Bollinger

■ RSI(Relative Strength Index) :

es un indicador de tipo oscilador que refleja la fuerza relativa de los movimientos alcistas, en comparación con los movimientos bajistas. Es utilizado por los traders para medir la fuerza de una tendencia e identificar señales de fin de tendencia. El oscilador se mueve entre los valores 0 y 100. Si el indicador esta por debajo de 30 se dice que está en *sobreventa*, y el precio tiende a subir. Cuando se encuentra por encima de 70, decimos que esta en *sobrecompra* y que el precio tenderá a bajar.

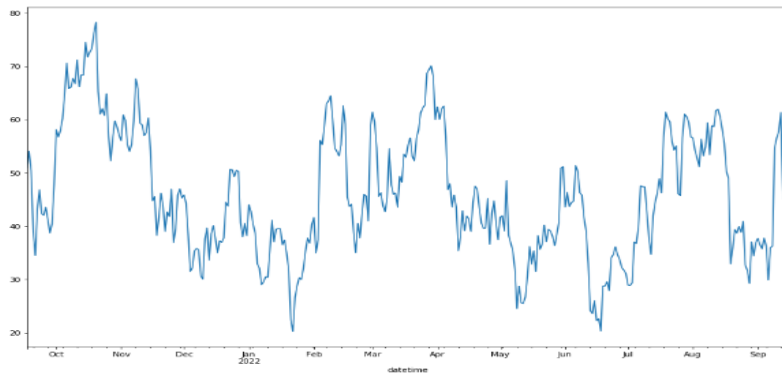


Figura 7: Representacion RSI

■ **Estocastico** [5]:

significa la comparación realizada por el indicador estocástico entre el precio de cierre actual y sus precios de cierre anteriores durante un período elegido. El Oscilador Estocástico se mide usando las líneas K y D, que se definen:

$$K = 100[(C - Ln)/(Hn - Ln)] ; D = \text{es la media móvil de K en N periodos}$$

donde C es el precio de cierre actual; Ln es el precio más bajo durante los últimos "n" periodos; Hn es el precio más alto durante las últimos "n" periodos.

Para interpretarlo sabemos que: si el precio de cierre actual es cercano al precio más alto (Hn) del período en cuestión, el estocástico estará cerca del 100 ; si el precio de cierre actual está cerca del precio más bajo (Bn) del período relevante, el estocástico estará cerca del 0 .

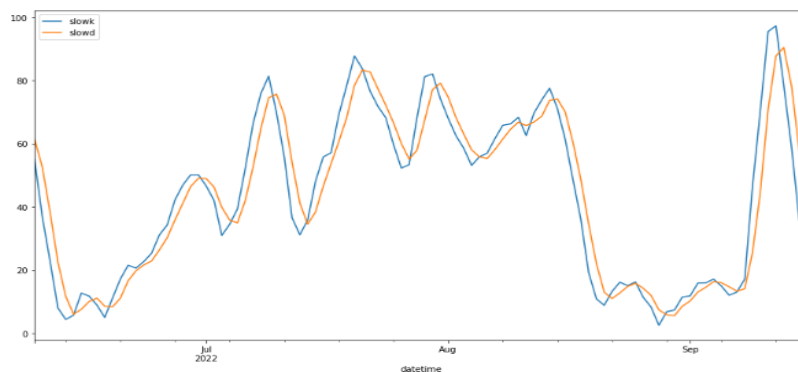


Figura 8: Representacion indicador estocástico

- **Indicador ATR (Average True Range) [6]:**

calcula la volatilidad existente en el periodo actual. El ATR sube o baja cuando el activo tiene mas o menos volatilidad; es decir, cuando el valor de ATR es alto, hay una alta volatilidad y unos valores bajos indican una estabilidad en los precios.

Para calcularlo, llamamos TR a el valor más alto de entre: el maximo menos el mínimo, el maximo menos el cierre del periodo anterior y el minimo menos el cierre del periodo anterior. Una vez tenemos este valor, calculamos lo siguiente:

$$ATR = (ATR \text{ anterior} * (n - 1) + TR(\text{periodo actual}))/n$$

donde n indica el periodo en el que calculamos esta media.

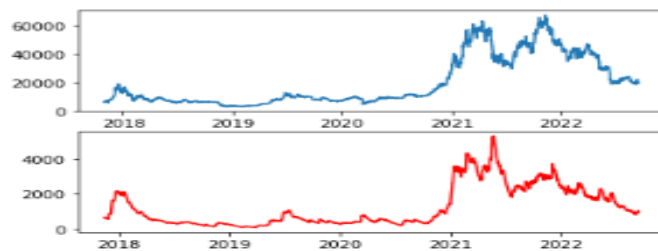


Figura 9: Comparación precio de cierre (arriba) con valores del indicador ATR (abajo)

- **Funciones de Reconocimiento de patrones (BELTHOLD) :**

al ingresar los datos de apertura, máximo, mínimo y cierre de la acción en estudio la función seleccionada retornara uno de tres posibles valores enteros, 0 cuando no reconoce patrón, 100 cuando reconoce un patrón alcista y -100 cuando reconoce un patrón bajista.

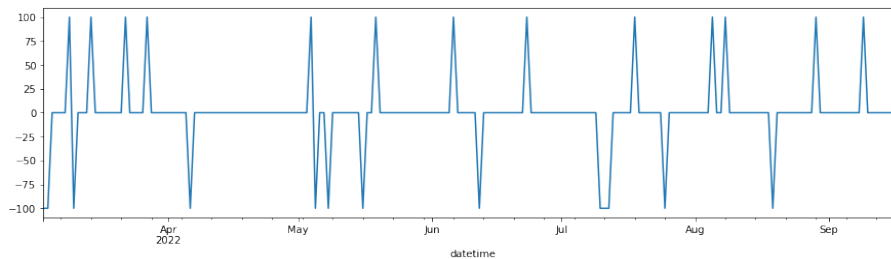


Figura 10: Representacion BELTHOLD

2. Estudio de la correlación de las variables

El estudio de la correlación nos proporciona información sobre la fuerza de la relación que hay entre las variables. Existe la correlación positiva y negativa. La positiva nos indica que las variables se mueven en la misma dirección, cuando una aumenta, la otra también lo hace. La correlación negativa ocurre cuando se mueven en direcciones opuestas, si una aumenta, la otra disminuye.

Para estudiar la correlación que existe entre todas las variables vamos a calcular el coeficiente de correlación de Pearson y representarlo gráficamente en un mapa de correlación(11).

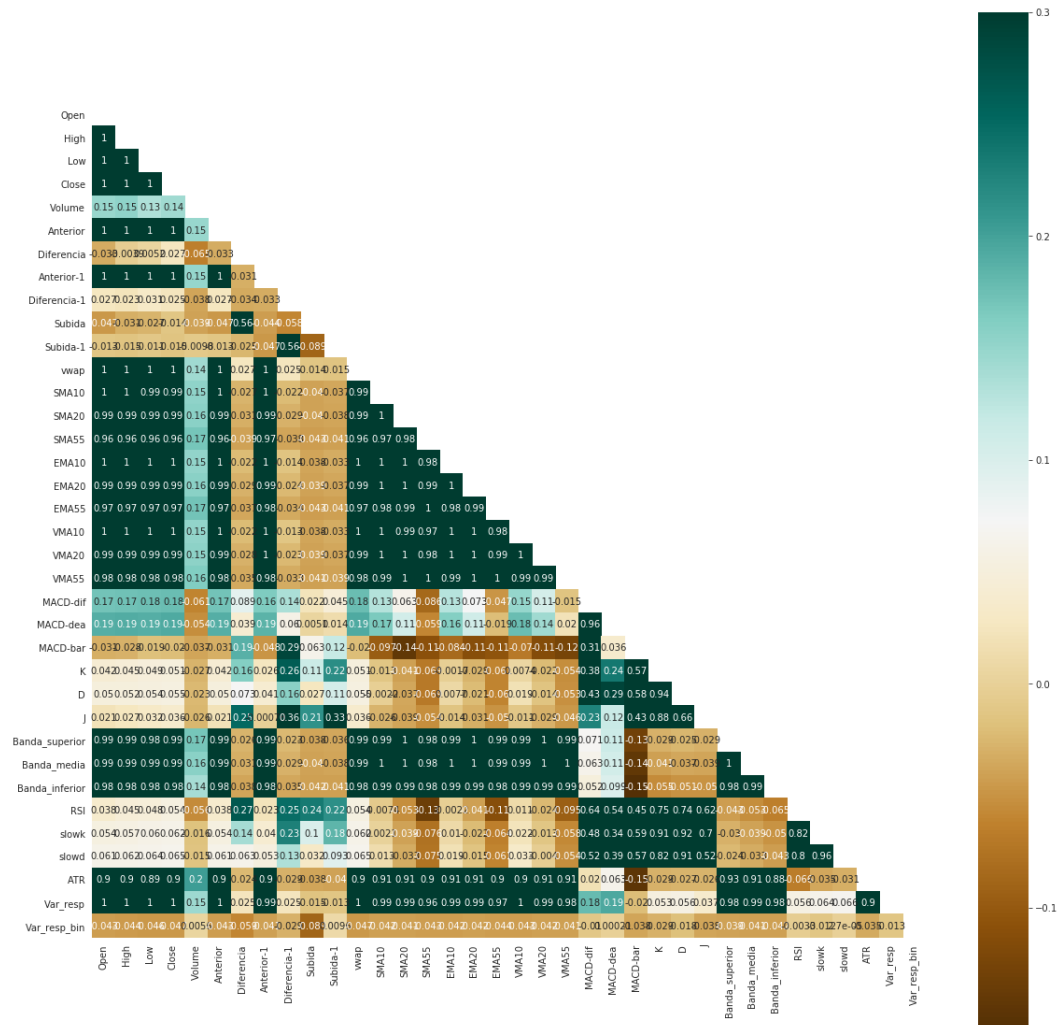


Figura 11: Mapa de calor para el conjunto de datos 1.

El coeficiente de correlación de Pearson entre dos variables, X e Y, se calcula utilizando la siguiente fórmula:

$$Corr(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Donde: \bar{x} es el valor medio de x, \bar{y} es el valor medio de y, x_i es el i-ésimo valor de x e y_i es el i-ésimo valor de y.

Los valores de correlación van entre -1 y 1; si el valor es 1 hay una correlación positiva, si el valor es -1, hay una correlación negativa y, si es 0 indica que no hay correlación entre las variables.

Si nos fijamos en el primer conjunto de datos, podemos observar que tienen muy buena correlación las variables Open, High, Low, Close, Anterior y Anterior-1; tanto entre ellas como con la variable respuesta.

Para el segundo conjunto, vemos que todas las medias móviles que hemos añadido tienen una correlación bastante alta, por lo que podemos suponer que este conjunto va a funcionar mejor que el primero.

Por último, en el tercer conjunto, la correlación no es tan alta, parece que el indicador que más aporta al resultado de la variable respuesta son las Bandas de Bollinger.

Bibliografía

- [1] "*Introduction to Moving Averages For Day Trading*", mayo 2022. Nick,P. Disponible: <https://bullsonwallstreet.com/moving-averages-day-trading/>
- [2] "*Variable Moving Average – Trading Strategy Backtest*", agosto 2022. Oddmund Groette. Disponible: <https://www.quantifiedstrategies.com/variable-moving-average/>
- [3] "*Estrategias de trading MACD*". AVATRADE INTERNATIONAL. Disponible: <https://www.avatrade.es/educacion/professional-trading-strategies/macd-trading-strategies/>
- [4] "*Las bandas de Bollinger*", 2006. John Bollinger . *Visual Chart Group*.
- [5] "*El Indicador Estocástico en profundidad*", 11 julio 2022. Admirals Markets. Disponible: https://admiralmarkets.com/es/education/articles/forex-indicators/indicador-estocastico?utm_source=google&utm_medium=cpc&utm_campaign=ES_ES_performance_max_new&utm_term=&gclid=EAIaIQobChMIm-bP_sKy9wIV1vZRCh0T1gvKEAAYAiAAEgIzZfD_BwE
- [6] "*¿Qué es el Average True Range (ATR) y cómo utilizarlo?*", 28 de Enero 2021. Luis Ángel Hernandez. Disponible: <https://www.rankia.com/blog/bolsa-desde-cero/3529085-que-average-true-range-atr-como-utilizarlo>