

# Propuesta Inicial

**Integrantes:** Alexis Abreu, Maria Paula Cortés, Ricardo Pretelt, Edgar Balaguera

## **¿Cómo determinar la estructura del mercado de valores en términos de la correlación condicional entre instrumentos financieros para la ayuda en la toma de decisiones de inversión?**

### **Resumen**

Los algoritmos de trading se basan en evaluaciones técnicas segundo a segundo que se realizan sobre los precios y el comportamiento histórico de los mismos, la profundidad del mercado, el volumen de transaccionalidad y la situación económica directamente relacionada con los instrumentos financieros elegidos, entre otros. Esta información, aunada con conocimiento del mercado, permiten desarrollar gráficas de proyección que apoyen las decisiones de inversión en términos de escenarios Optimistas, Esperados y Pesimistas.

Estos algoritmos parten del comportamiento histórico de los precios de las acciones y de su cuantificación segundo a segundo lo que hace que la mayoría de los análisis financieros usando analítica, se trabajan como un problema de Series de Tiempo.

Nuestra propuesta se fundamenta en determinar la correlación condicional entre instrumentos financieros y estar en capacidad de clasificarlos según sus características faciales al cierre del mercado, es decir, en un momento específico del tiempo.

Estas clasificaciones permiten evaluar el riesgo de inversión y tomar decisiones de inversión basados en el Client Suitability o, afinidad entre el componente financiero del portafolio y la expectativa tanto de riesgo como de rentabilidad del agente inversionista.

### **Introducción**

Históricamente las personas con intenciones de inversión han tenido que acercarse a una comisionista o a un inversionista profesional que establezca un vínculo entre las oportunidades de inversión del mercado y el perfil de riesgo del cliente. El deber de asesoría obliga al asesor a mantener las características de apetito de riesgo del cliente alineadas con el portafolio de inversión.

Recientemente se han popularizado las páginas de trading independiente en donde las personas tienen la libertad de invertir en los instrumentos financieros que resultan atractivos. Sin embargo, son pocas las personas con alto Financial Literacy y por ende dependen de las asesorías y consejos de personas que se consideran expertas en el campo.

Un desarrollo de este tipo, habilita la toma de decisiones con base en afinidad entre los instrumentos financieros objeto de estudio o seleccionados por cada persona.

Es por esto, que pretendemos implementar un algoritmo de clustering que permite encontrar grupos de interés basados en los precios de cierre de las acciones en un corte transversal del tiempo, analizar correlaciones condicionadas entre los precios de cierre de cada una de las acciones seleccionadas y así lograr encontrar patrones similares en los movimientos del precio de las acciones y determinar a través de estas similitudes que acciones tienen los mismos comportamientos para diversificar las inversiones.

Se parte de la agrupación de las acciones por el comportamiento en su precio en el mercado, analizar su rendimiento y así poder establecer una medida de comparación con las demás. Una vez establezcamos estas medidas de comparación se realizará un proceso de clustering con K-means.

El clustering, en otras palabras, permite identificar los instrumentos financieros con comportamientos similares e identificar, por ende, las posiciones de mayor rentabilidad al momento de tomar una decisión de inversión. Estos agrupamientos permiten aumentar la posibilidad de rentabilidad positiva a partir del entendimiento de más de 1 cluster, es decir diversificar el portafolio de inversión.

En el proceso de clustering estos grupos con comportamientos similares deben ser mutuamente excluyentes; es decir, cada instrumento debe pertenecer a un sólo grupo y cada grupo debe estar formado por empresas en el que su comportamiento sea lo más parecido posible y a la vez, con comportamientos diferentes entre grupos. En la práctica, se busca encontrar grupos con elementos lo más homogéneos posibles pero heterogéneos entre ellos.

## Revisión preliminar de antecedentes en la literatura

En la literatura asociada con trading algorítmico, se sugiere trabajar con Series de Tiempo para determinar la probabilidad de rentabilidad en términos del comportamiento histórico de los instrumentos. Sin embargo, en las técnicas exploradas se menciona la complejidad de procesamiento, no análisis, dada la dimensionalidad de la información con la que se debe trabajar.

Por otro lado, las técnicas comúnmente usadas en el análisis financiero en el mercado de criptomonedas trabajan a partir de *Affinity Propagation* y *Clustering* para delimitar las criptomonedas en las que se invierte.

A partir de la literatura, concluimos que si bien Series de Tiempo no será la metodología utilizada para el desarrollo de nuestro algoritmo, sí debemos tener en cuenta el comportamiento histórico para entrenar el modelo y tener clusters ajustados y robustos.

## Descripción detallada de los datos

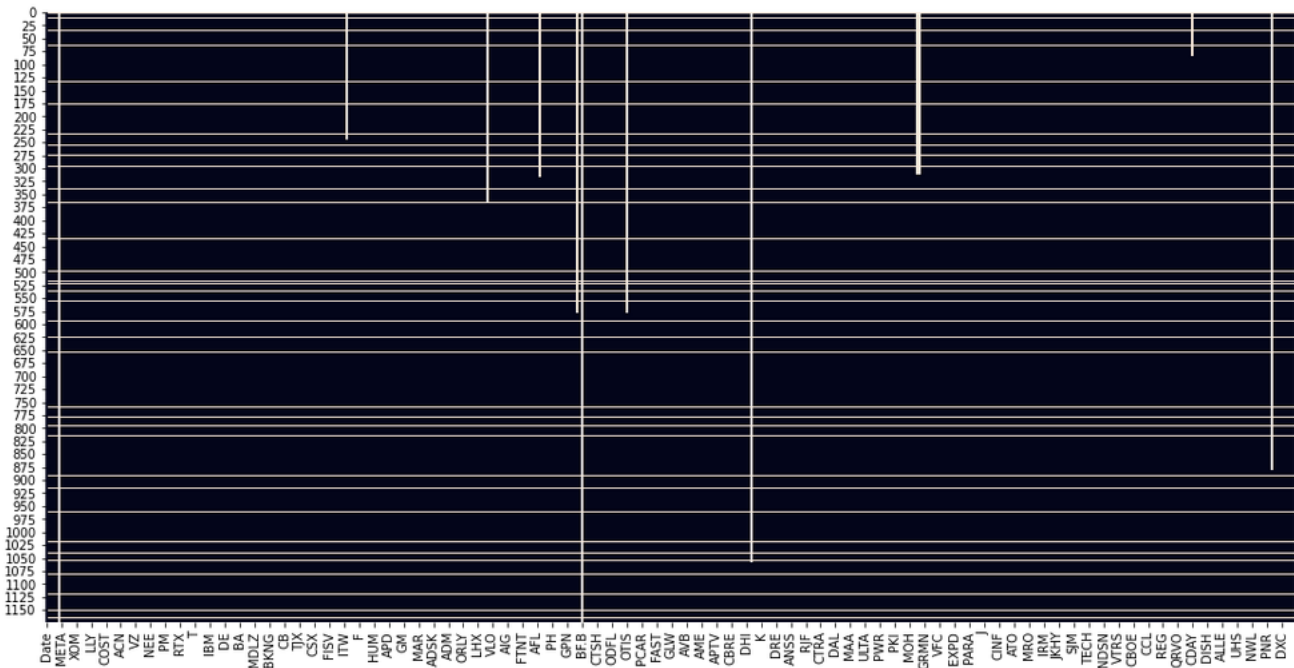
Se usan los datos históricos de valor de acciones en cierre ajustado para empresas de Standard & Poor's 500 (S&P 500) desde 1/01/2018 hasta 30/06/2022, teniendo en cuenta que la bolsa sólo abre en días laborales. Se obtuvieron los datos gracias a la librería de python Yfinance.<sup>1</sup>

Cómo el objetivo del proyecto es encontrar patrones similares en los movimientos del precio de las acciones. Los datos de cierre son expresados como la diferencia de valor entre un día y otro. De esta forma se pueden hacer comparaciones relativas a los cambios de las acciones y no su valor como tal.

---

<sup>1</sup> Documentación Yfinance en python <https://pypi.org/project/yfinance/>

Se pueden observar los datos faltantes con la ayuda de matplotlib:



Se observa que algunas empresas no tienen los datos completos, se procede a quitar los registros de días sin ningún dato. Después de esto se quitan las empresas que aún mantienen datos faltantes ya que son pocas y finalmente se queda con una base completa.

En cuanto a la descripción de los datos se conoce:

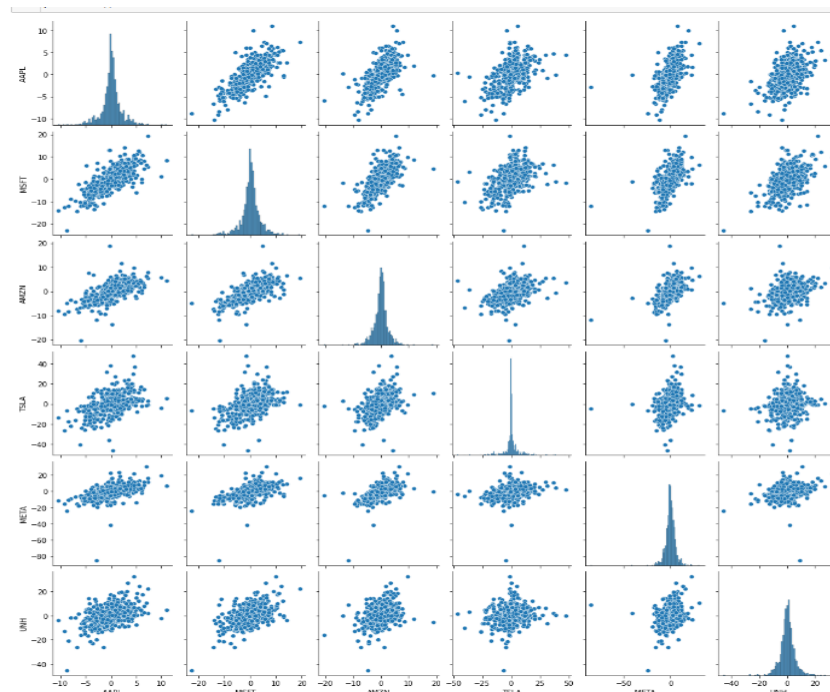
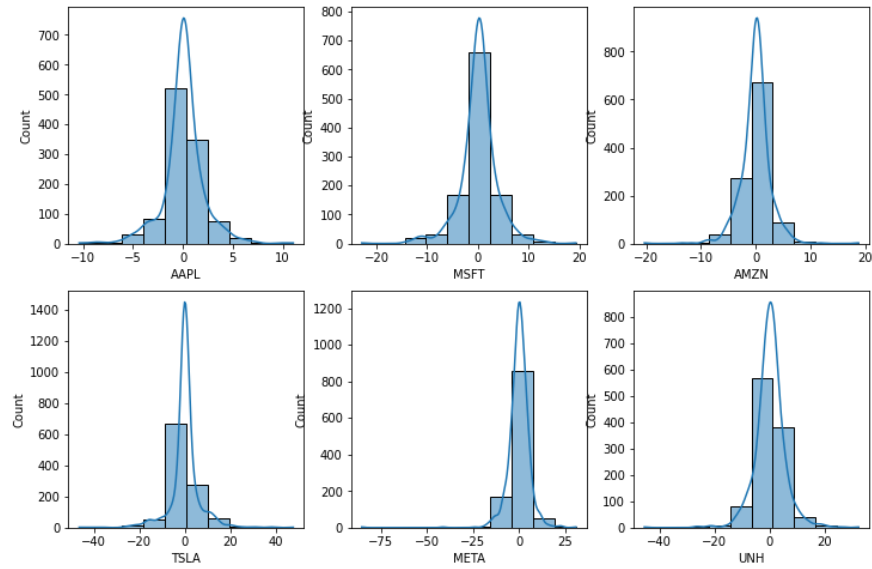
- Matriz de 1089 registros tipo float con 491 columnas que corresponden a las empresas. Estos datos pueden ser positivos o negativos.
- Uso en memoria de 4.1MB, por lo que el espacio ocupado es poco.

También se puede observar un resumen descriptivo de los datos para algunas compañías:

	AAPL	MSFT	AMZN	TSLA	META	UNH	V	NVDA	JNJ	WMT	...	MHK
<b>count</b>	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	...	1089.000000
<b>mean</b>	0.091235	0.166727	0.011612	0.173119	-0.037346	0.268614	0.065767	0.095789	0.064515	0.031585	...	-0.106336
<b>std</b>	2.038065	3.647501	2.637628	6.772296	5.914395	5.578339	3.385888	4.627571	1.799005	1.667199	...	3.988493
<b>min</b>	-10.394508	-22.885284	-20.314995	-46.480011	-85.240005	-45.498077	-23.439178	-20.712128	-13.441071	-16.786896	...	-38.059998
<b>25%</b>	-0.669571	-1.267303	-1.145500	-0.760010	-2.510010	-2.246384	-1.388229	-1.289192	-0.742699	-0.741180	...	-1.929993
<b>50%</b>	0.091190	0.233170	0.121002	0.073999	0.179993	0.326813	0.242432	0.161060	0.080452	0.046661	...	0.020004
<b>75%</b>	0.884510	1.710739	1.179993	1.250000	2.819992	2.742767	1.651642	1.572685	0.934265	0.767731	...	2.050003
<b>max</b>	11.064224	19.326904	18.793991	47.666656	30.779999	32.420227	21.734451	32.013855	9.231430	11.959915	...	15.500000

8 rows x 491 columns

Cómo el número de empresas es muy grande, se grafican las empresas más conocidas y valoradas.



Como se observa las distribuciones son similares y correlaciones son relativamente altas, por lo que otras empresas menos conocidas pero con patrones similares pueden ser prometedoras.

## Propuesta metodológica

GraphLassoCV, Affinity propagation, LocallyLinearEmbedding

Para el desarrollo de este proyecto, inicialmente se tiene planificado el uso de dos a tres técnicas de algoritmos que se describirán a continuación, así como la justificación de su uso.

1. GraphLassoCV: Luego de obtener la variación porcentual para los precios de cierre de los activos durante el periodo analizado, lo que se busca es encontrar que precios de cierre están correlacionados unos con otros de forma condicional, para esto se hará uso de la matriz inversa de covarianza dispersa (sparse inverse covariance estimation), esto nos dará como resultado una lista de símbolos y conexiones.
2. Algoritmo de clusterización: Para este punto no nos limitaremos desde ahora al uso de un algoritmo en específico, aunque dentro de las estimaciones tenemos pensado implementar DBSCAN o affinity propagation, por su facilidad que nos ofrece a la hora de generar automáticamente cluster sin necesidad de definir este parámetro. En este sentido la idea es agrupar precios de cierre que se comportan de forma similar a nivel de su cambio porcentual y su relación condicional, es decir planeamos usar como insumo los resultados del paso anterior.
3. Hasta este punto tenemos contempladas dos técnicas de aprendizaje no supervisado y podemos asignar un cluster a cada valor correspondiente en el data frame pero quisiéramos ir más a fondo, en este caso, dado a que tendremos una matriz con una dimensionalidad bastante alta, quisiéramos realizar una representación visual en dos dimensiones de este output multidimensional, y así reducirlo a dos dimensiones para poder graficar de forma sencilla planeamos el uso de alguno de los algoritmos de sklearn.manifold learning techniques.

Es de esta forma que planeamos de manera generalizada obtener la matriz inversa de covarianza dispersa, clusterizarla y luego graficarla en un espacio de dos dimensiones.

## Bibliografía

- Chen, Yudong, et al. "Clustering financial time series to generate a new method of factor neutralization: An empirical study." *International Journal of Financial Engineering*, vol. 08, no. 2141005, 2021. *World Scientific*, <https://www.worldscientific.com/doi/abs/10.1142/S242478632141005X>.
- Hong, Lei, et al. *Community Identification of Financial Market Based on Affinity Propagation*. Lecture Notes in Electrical Engineering, 2013, [https://www.researchgate.net/publication/287085677\\_Community\\_Identification\\_of\\_Financial\\_Market\\_Based\\_on\\_Affinity\\_Propagation](https://www.researchgate.net/publication/287085677_Community_Identification_of_Financial_Market_Based_on_Affinity_Propagation).
- Marti, Gautier, et al. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets." *Progress in Information Geometry*, 2021, pp. 245-274. *Springer, Cham*, <https://arxiv.org/pdf/1703.00485.pdf>.

- Müller, Florian. *Clustering Financial Market Structures using Affinity Propagation in Python*. 2022. *Relataly*,  
<https://www.relataly.com/crypto-market-cluster-analysis-using-affinity-propagation-python/8114/>.
- Musmeci, Nicolás, et al. *Relation between Financial Market Structure and the Real Economy: Comparison between Clustering Methods*. PLOS ONE, 2015,  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0116201#amendment-0>.
- Nayak, Rudra Kalyan, et al. "Rough Set Based Affinity Propagation Model for Prediction of Future Gold Price in Indian Scenario." *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, 2019, pp. 1-6,  
<https://www.ijitee.org/wp-content/uploads/papers/v9i2/B7584129219.pdf>.
- Tan, Alvin T. "Hierarchical Clustering of the FX Market." *Towards Data Science*, 2020,  
<https://towardsdatascience.com/a-hierarchical-clustering-of-currencies-80b8ba6c9ff3>.
- Wang, Limin, et al. *Affinity propagation clustering algorithm based on large-scale data-set*. International Journal of Computers and Applications, 2018. *Taylor & Francis Online*,  
<https://www.tandfonline.com/doi/abs/10.1080/1206212X.2018.1425184?journalCode=tjca20>.