



Facultad de Ciencias Económicas y Empresariales, ICADE

**EFFECTO DE LA PUBLICACIÓN DE ESTADOS
FINANCIEROS DE EMPRESAS COMPARABLES
EN LA COTIZACIÓN DE UNA EMPRESA**

Autor: Ignacio Ortiz Sánchez

Clave: 201808311

Director: Carlos Bellón Núñez-Mera

RESUMEN

Este Trabajo de Fin de Grado tiene por objetivo el análisis del efecto que tiene la publicación de estados financieros de empresas comparables, sobre la cotización y los retornos de una empresa. Se tratará de sacar conclusiones sobre cómo funciona el índice y si de entre todos los factores que afectan a la cotización de una empresa, el desempeño de los *peers* afecta o no.

En este trabajo, se realiza un estudio sobre los factores internos y externos de las empresas que utilizan los inversores para su comparación. Por otro lado, se realizan varios análisis de eventos en los que se comparan los retornos anteriores y posteriores de las empresas cuando ocurre un acontecimiento, que, en este caso, será la publicación de resultados de empresas del mismo subsector. Además, se focalizará en el estudio de las diferencias del efecto que tiene la publicación de resultados en las distintas industrias que forman el Standard & Poor's 500.

En definitiva, haciendo uso de herramientas de análisis y visualización de datos, se realiza un análisis empírico y teórico para comprender uno de los factores que afecta a la cotización de las empresas del índice.

Palabras clave: empresas comparables, valor relativo, análisis de eventos, retorno, capitalización bursátil, subindustria, *dataframe*.

ABSTRACT

This Final Degree Project aims to analyse the effect that the publication of financial statements of comparable companies has on a company's share price and returns. It will draw conclusions on how the index works and if among all the factors that affect a company's share price, the performance of the peers is one of those factors or not.

In this project, a study is carried out on the internal and external factors of the companies used by investors for benchmarking. On the other hand, several event analyses are carried out in which the previous and subsequent returns of companies are compared when an event occurs, which, in this case, will be the publication of financial statements of companies in the same subindustry. In addition, it will focus on the study of the differences in the effect of the publication of results in the different industries that make up the Standard & Poor's 500.

In short, using data analysis and visualization tools, an empirical and theoretical analysis will be carried out to understand one of the factors that affects the share price of the companies in the index.

Keywords: peers, relative value, event analysis, return, market cap, subindustry, dataframe.

Índice

CAPÍTULO I. INTRODUCCIÓN Y OBJETIVOS	7
CAPÍTULO II. VALORACIÓN DE EMPRESAS – MARCO TEÓRICO	9
2.1. Valoración intrínseca	10
2.2. Valoración relativa	13
2.3. Entorno Macroeconómico	16
CAPÍTULO III. OBTENCIÓN, LIMPIEZA, TRATAMIENTO Y ANÁLISIS DE LOS DATOS DEL S&P 500 - METODOLOGÍA	19
3.1. Identificación de empresas	20
3.2. Agrupación por Subindustrias	21
3.3. Publicación de resultados de las empresas	23
3.4. Métricas e información financiera a utilizar	25
3.5. Bases de datos	26
CAPÍTULO IV. ANÁLISIS DE EVENTOS	29
4.1. Análisis de medias con el retorno actual	30
4.2. Análisis de medias y desviación típica con retorno actual	33
4.3. Análisis con retorno anormal	35
4.3.1. Comparación entre fechas de publicación trimestrales y anuales	36
4.3.2. Comparación entre industrias	37
4.3.3. Comparación de subsectores según la cantidad de peers	39
4.3.4. Comparación entre la significatividad con 5 días vs con 10 días	40
CAPÍTULO V. OTRAS LÍNEAS DE INVESTIGACIÓN, LIMITACIONES Y PROBLEMAS	43
CAPÍTULO VI. CONCLUSIONES	45
ANEXOS	47
Anexo I. Bibliotecas Utilizadas	47
Anexo II. Pasos relativos al WebScrapping de Wikipedia	47

Anexo III. Pasos relativos a la obtención del dataframe que contiene los grupos de peers (df_peers).....	47
Anexo IV. WebScrapping EDGAR	48
Anexo V. Creación del dataframe de retorno esperado y retorno anormal	49
Anexo VI. Código tipo para la ejecución del ttest	50
BIBLIOGRAFÍA	51
REFERENCIAS CÓDIGO	53

Índice Ilustraciones

Ilustración 1: Fed Funds Rate vs Inflation (25 years)	17
Ilustración 2: Fed Funds Rate vs Consumption (25years).....	18
Ilustración 3: Webscrapping Compañías S&P 500.....	21
Ilustración 4: Primeras once subindustrias con sus respectivas empresas	22
Ilustración 5: Fechas de publicación compañías (3M)	24
Ilustración 6: Parte del dataframe de peers transformado a Excel.....	27
Ilustración 7: Parte del dataframe de fechas de publicación	27
Ilustración 8: Empresas contenidas para cada umbral de pvalor en la publicación de cuentas trimestrales.....	32
Ilustración 9: empresas contenidas en cada umbral de pvalor en la publicación de cuentas anuales	32
Ilustración 10: Distribución de los pvalores de cada evento (publicación de cuentas trimestrales)	33
Ilustración 11: Distribución de los pvalores de cada evento (publicación de cuentas anuales).....	34
Ilustración 12: Comparación del nivel de significación de la publicación de cuentas trimestrales vs anuales	36
Ilustración 13: Comparación de pvalores para cada industria.....	38
Ilustración 14: Comparación porcentual de grupo de pvalores por industria.....	38
Ilustración 15: Comparación entre subsectores con muchos(izq.) y pocos(dcha.) peers	39
Ilustración 16: Comparación entre significatividad según la duración del event window	41
Ilustración 17: Comparación entre significatividad según la duración del event window (Personal Care)	41

Índice Fórmulas

- (1) Discounted Cash Flows
- (2) Price to Earnings Ratio
- (3) Price to Sales Ratio
- (4) Price to Book Value Ratio
- (5) Enterprise Value/ Ebitda Ratio
- (6) Simple Return
- (7) Abnormal Return
- (8) Beta

CAPÍTULO I. INTRODUCCIÓN Y OBJETIVOS

En este trabajo se pretenderá estudiar si afecta la publicación de los estados financieros de compañías comparables, a la cotización de una empresa, y en tal caso cómo afecta. Ya sea de una forma positiva o negativa, noticias sobre las empresas comparables deberían afectar a la cotización de una empresa, ya que esto es un indicador de cómo está funcionando el mercado y en concreto el sector, y el inversor no se debería sentir indiferente.

Según (Damodaran, 2011), existen diferentes modelos de valoración de un activo, pero únicamente existen dos enfoques principales para realizar dicha valoración. El primero es el intrínseco: el valor intrínseco de un activo está basado en el análisis de los flujos de caja en este caso de la compañía que se esté analizando. El segundo es el relativo, este trata de valorar los activos observando el valor que tienen estos activos o similares, en el mercado. En el caso de las empresas se habla de *peers*, o empresas comparables. Hay diferentes características, que se verán más adelante, en las que habrá que fijarse para determinar cuáles son los *peers* de una empresa. Ambos modelos son complementarios, a la hora de analizar el valor de una empresa.

El hecho de que el enfoque relativo de la valoración sea tan utilizado en el mercado por los inversores, probablemente haga que la publicación de las cuentas de una empresa comparable, refleje un impacto en la cotización de las acciones de otras empresas de diferentes formas. Además, el análisis de empresas comparables, lo utilizan los inversores de diferentes modos:

- En primer lugar, se puede realizar *benchmarking*, así los inversores y analistas usan el rendimiento financiero de las empresas comparables para evaluar si el rendimiento de la empresa está por encima o por debajo de la media de los *peers*. Y según esto analizar cómo se encuentra la empresa con respecto al mercado (Tuovila, 2020).
- Además, a través del análisis de los estados financieros de varias empresas con impacto en el mercado de forma transversal, se puede analizar en el momento del ciclo en el que se encuentra el mercado, el hecho de que los resultados de diferentes empresas de un mismo mercado sean positivos o negativos, puede apuntar a los inversores, que, en general, todo el mercado está en un momento alcista o bajista

del ciclo económico, y por tanto, todas las empresas de este sufran un aumento o disminución en la cotización de sus acciones.

- Por último, el análisis de empresas comparables, suele utilizarse en diferentes modelos de valoración para determinar si el valor de las acciones de una compañía en el mercado realmente es el valor razonable, o en caso contrario si están sobre o infra valoradas.

Por ello, puede afectar directamente de forma que, si el rendimiento financiero de empresas comparables es positivo, se puede pensar que esto signifique que el sector de la empresa analizada se encuentra en un momento alcista del ciclo, y por tanto significará un aumento en la cotización de las acciones. Por otro lado, también tendría sentido pensar que este rendimiento positivo de los *peers* afecte inversamente a la cotización de la empresa, ya que, al realizar *benchmarking*, es probable que la empresa analizada se encuentre peor “posicionada” en el mercado que sus empresas comparables.

Como se puede intuir la publicación de cuentas de empresas es un evento al que el mercado y los inversores le dan gran importancia para analizar sus carteras, así como nuevas oportunidades de inversión. Durante este trabajo se pretenderá resolver el escenario planteado en el párrafo anterior. En resumen, los objetivos de este trabajo son:

- Entender los múltiplos de valoración y en concreto el Price to Earnings ratio y el EV/EBITDA. Saber la importancia que tienen en el mercado y como estos múltiplos se utilizan para la valoración entre las empresas y la comparación con sus *peers* por los inversores.
- Estudiar si realmente la publicación de cuentas trimestrales de empresas comparables afecta a la cotización de una empresa.
- Conocer la significatividad del cambio de los retornos de una empresa en las fechas de publicación de resultados de sus *peers* y comparar según industrias.
- Comparar si afecta más o menos a aquellas subindustrias más o menos grandes dentro del S&P 500 u otras métricas similares.
- Analizar si existe una diferencia entre el uso de una ventana de evento más o menos grande y, por tanto, explicar cuánto tarda el mercado en ajustarse al evento.

CAPÍTULO II. VALORACIÓN DE EMPRESAS – MARCO TEÓRICO

Antes de adentrarse en el análisis de cómo afecta la publicación de cuentas de empresas comparables a la cotización de una empresa en el mercado, hay que entender los distintos factores de los que depende la cotización de las acciones de las empresas. Por ello, es importante conocer qué factores hay que analizar para calcular el valor de una empresa, y, sobre todo, saber cómo calcular el valor de dicha empresa, además del contexto en el que se realizan estos cálculos. Como ya se ha comentado en la introducción, existen dos enfoques para calcular el valor de una empresa, el intrínseco y el relativo (Damodaran, 2011), y cada uno tiene distintos modelos, a través de los cuales analizar el valor.

Todo este análisis del valor de una empresa y de las acciones de la misma se va a hacer bajo la premisa de que el mercado es eficiente. Según Eugene Fama (1970), un mercado es eficiente fuerte, cuando toda la información de las empresas es pública y está disponible para todo el mercado y es por ello por lo que las acciones cotizan siempre al precio objetivo. Por esto, no será posible la compra de acciones infravaloradas ni sobrevaloradas, y será imposible que los inversores obtengan rendimientos más altos que los del mercado a largo plazo, la única forma en la que un inversor podrá obtener márgenes más altos será invertir en activos con mayor riesgo.

Por otro lado, se presentan dos marcos más que representan un mercado eficiente, aunque no sea fuerte. En primer lugar, se presenta una alternativa, que puede ser que refleje más la realidad del mercado, que es el mercado eficiente semifuerte. En este tipo de mercado, se considera que toda la información pública actual, está reflejada en el precio de mercado, pero se acepta que la información privada puede no estar reflejada, lo que haría que pudieran existir oportunidades para obtener beneficios mediante el análisis de información privada. Esta información privada, no son los estados financieros, se da por hecho que las cuentas, en las empresas que cotizan son públicas y accesibles para todo el mundo, por tanto, esta “información privada”, se puede referir, a la gestión de la empresa, decisiones de inversión futuras, planes de expansión o adquisición, en definitiva todo aquello que conocen los ejecutivos de la empresa, y que el mercado no tiene por qué conocer, y será solo cuando esta información se haga pública, cuando el mercado se ajuste.

Por otro lado, el otro marco en el que se puede realizar este análisis, es bajo la premisa de que el mercado es eficiente débil, en este tipo de mercado según Fama (1970), se considera que el precio de mercado refleja toda la información pública y privada del pasado, pero no necesariamente la información actual, esto quiere decir que las tendencias pasadas se pueden utilizar para predecir los movimientos futuros, pero no para obtener beneficios a largo plazo.

En conclusión, existen diferentes opciones de marco teórico, en el que se puede realizar el estudio y en este caso se elige realizarlo sobre el marco del mercado eficiente semifuerte, ya que se ajusta a la realidad del S&P 500, y en este pueden existir oportunidades de inversión. Además, como las cuentas anuales y trimestrales son información pública no existe una incongruencia con el tipo de análisis que se llevará a cabo.

Conociendo esto, será importante también conocer el valor intrínseco de las empresas. A continuación, se explicarán las diferentes formas que existen en el conjunto del mercado para calcular el valor intrínseco de las empresas, cómo estos métodos de valoración funcionan y su aplicación en el mercado.

2.1. Valoración intrínseca

El valor intrínseco se refiere al valor real de un activo, y se calcula de manera objetiva, con modelos financieros, en base a factores fundamentales sin tener en cuenta lo que el mercado pueda considerar sobre este activo. Es más, el precio de mercado de un activo no suele coincidir con el valor intrínseco de ese mismo activo, sin embargo, puede dar una idea a los inversores de si un activo está infravalorado o sobrevalorado. No existe una única forma de calcular el valor intrínseco de una compañía o de un *stock*, las más utilizadas son:

- **El modelo de descuento de dividendos (DDM):** John Burr Williams (1938), en su libro *The Theory of Investment Value*, se centra en el cálculo del valor intrínseco de las acciones, el cual, dice que es el valor presente de los flujos de caja futuros que se esperan recibir de la empresa (los dividendos), y no simplemente en el precio actual de sus acciones. Existen variantes de este modelo como el *Gordon Growth*

Model (GGM), que asume un crecimiento constante en el futuro y una tasa de dividendo constante.

- **Modelo de ingresos residuales (RIM):** este modelo se basa en el concepto de que el valor de una empresa, es igual al valor presente de sus flujos de efectivo futuros menos el valor presente del capital invertido, para generar esos flujos de efectivo. Es decir, se enfoca en el excedente de ingresos o ingresos residuales, que una empresa genera después de cubrir los costos y gastos indispensables para mantener su actividad. Propone una alternativa al modelo de descuento de dividendos, ya que Miller & Modigliani, (1961), piensan en aquellas empresas que no distribuyen dividendos de manera consistente, y en aquellas que no tienen flujos de caja libre positivos.
- **Modelo de descuento de flujos de caja:** por último, el método de valoración más utilizado para la valoración de activos y acciones es el descuento de flujos de efectivo (DCF por sus siglas en inglés). El DCF, es un modelo que utiliza no solo los flujos de caja actuales, sino que proyecta los flujos de caja futuros, para determinar el valor presente de la compañía. El modelo busca los “flujos de caja libres”, estos son todos los flujos de efectivo sin incluir aquellos ingresos y gastos de la cuenta de resultados, que no suponen una entrada o salida real de dinero, como la depreciación o la amortización, y teniendo en cuenta las inversiones necesarias para mantener y expandir el negocio (*Working Capital* y *Capex*) (Damodaran, 2011). Las proyecciones a futuro de estos cash flows, normalmente suele ser de 5 años a 10 años vista, ya que intuir el futuro más allá de los 5 años es complicado, y de los 10 prácticamente imposible, la firma Grant Thornton, aconseja que no debería de superar los 5 años, si realmente se quiere reflejar una imagen fiel (Carroll, 2021).

$$DCF = \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \frac{CF_3}{(1+r)^3} + \dots + \frac{CF_n}{(1+r)^n}$$

(1)

CF= *Cashflow* en el periodo n

r= tasa de descuento (WACC)

Para calcular el valor intrínseco de una compañía o de un activo, no solo se tiene en cuenta factores puramente cuantitativos, que sería el mero cálculo de los flujos de caja. Sino que como dice Damodaran (2011), también entra en juego, una parte cualitativa que es cómo de seguro se está de que se seguirán recibiendo estos flujos de caja en el futuro, es decir, si estos flujos son estables o si por el contrario son volátiles. Por tanto, esta teoría, se podría resumir en que el valor de una empresa reside en la capacidad de esta para generar dinero de forma estable y cierta. Las empresas con *cash flows* estables deberían de ser mejor valoradas que aquellas cuyos *cash flows* son volátiles.

A estos flujos de caja futuros se les deberá aplicar una tasa de descuento y esto se debe a que el futuro es incierto y más arriesgado que el presente, por ello un euro en el presente debe valer más que un euro en el futuro. Además, el fenómeno de la inflación afecta a la depreciación paulatina del dinero. Por último, por el coste de oportunidad, que viene a decir que por el hecho de no poseer dinero en el momento se está dejando de invertir en otros activos, se debe de considerar la oportunidad de inversión perdida al calcular el valor presente de los flujos de efectivo futuros (Damodaran, 2011).

Estas tres razones son por las cuales se aplica una a tasa de descuento a los flujos de caja futuros. En el modelo de descuento de flujos de efectivo, la tasa de descuento para realizar el valor presente de los cash flows futuros de las empresas, se saca del WACC (*Weighted Average Cost of Capital*). Su traducción es el coste medio ponderado del capital, y lo que muestra es la cantidad de dinero que una empresa espera pagar para la financiar sus activos. La financiación de activos se puede hacer a través de deuda o de *equity*, la primera es “más barata” que el segundo, por diferentes factores. El primero es que los accionistas asumen un mayor riesgo que los prestamistas, ya que, en caso de quiebra o liquidación de la empresa, los acreedores tienen derecho a cobrar antes. El segundo factor es que los intereses de la deuda son deducibles de impuestos. Por esto se realiza una media ponderada según la proporción que exista en la estructura de capital de la empresa, y esto es el WACC.

Sin embargo, no solo es importante la valoración intrínseca para valorar una activo, ya que como se ha comentado anteriormente, este valor no suele coincidir con el valor de mercado del activo, pese a que sea en un mercado eficiente, ya que hay muchas otras variables e información que afectan a la cotización de un activo.

2.2. Valoración relativa

La valoración relativa es el segundo enfoque del que hablaba Damodaran (2011), en el que comenta que la mayor parte de los activos y acciones se valoran con un enfoque más relativo. En los distintos modelos valoración que siguen este enfoque, los activos se valoran observando precios de mercado de activos similares. El objetivo es determinar si la empresa en cuestión está infravalorada o sobrevalorada con respecto al mercado.

En el caso de las acciones como ya se ha comentado anteriormente, se comparan con otras acciones de empresas similares, a estas empresas se les llama *peers* o empresas comparables. La elección de estas empresas comparables es crucial para la correcta valoración relativa de una acción, y por tanto para este trabajo. Algunas de las características comunes que deben tener dos empresas para considerarlas comparables son:

- **Sector:** para que dos empresas sean comparables deben pertenecer al mismo sector, y subsector, y tener una actividad lo más similar posible. Es importante saber que no todas las empresas del mismo sector son comparables, ya que pueden tener diferentes características que las hagan realmente diferentes, por ejemplo, hay dos empresas del S&P500 que pertenecen al sector *consumer discretionary*, una es Best Buy y la otra Chipotle, una se dedica al retail de objetos electrónicos y la otra es una cadena de restaurantes, que son dos mercados que no tienen nada que ver el uno con el otro. Por esto habrá que fijarse no solo en el sector, sino también en la subindustria de las compañías.
- **Tamaño:** en este caso no será de tanta importancia ya que las empresas que cotizan en el S&P500 son todas empresas de gran tamaño, pero es una característica a tener en cuenta cuando se comparan dos empresas, ya sea a nivel de balance, de ingresos o de mercado.
- **Modelo de negocio:** las empresas deben tener un modelo de negocio similar, lo que implica que generan ingresos y ganancias de forma similar. Esto es más complicado de comparar, ya que es una característica cualitativa de las empresas, por tanto, no se tendrá tan en cuenta a la hora de escoger los *peers* de una empresa.
- **Ciclo de vida:** un poco en consonancia con el tamaño, las empresas deben estar en un ciclo de vida similar, ya que estas se enfrentan a retos y oportunidades similares. No se puede comparar una *startup* con una empresa que tiene cincuenta años de

experiencia. Este apartado también está relacionado con el crecimiento que experimenta la empresa en la actualidad. Tampoco será de gran importancia ya que las empresas del S&P500 se encuentran en una etapa madura.

- **Geografía:** esta característica tampoco afecta a este estudio, ya que en el S&P500 a todas las empresas les afectan condiciones regulatorias, económicas y sociales similares.

Una vez se conoce cómo agrupar a las diferentes empresas en comparables, hay que saber qué métodos de valoración relativa usan los inversores para conocer mejor el funcionamiento del mercado que se va a analizar, y cómo se relacionan los grupos de *peers*. Existen varios métodos de valoración relativa, según el sector, el tamaño de la empresa y otras características específicas del negocio, puede haber métodos que se ajusten mejor a un subsector que otros, pero los más importantes son los siguientes:

- **Múltiplos de precio beneficio (P/E ratio):** este ratio evalúa a las compañías midiendo el precio actual de la acción en relación al beneficio por acción (EPS por su denominación en inglés) de la compañía, comparándolo con empresas del sector (Ghaeli, 2016). Por definición, muestra cuánto dinero está dispuesto a pagar un inversor para recibir 1\$ del beneficio de la compañía. Este ratio, se suele utilizar para empresas y sectores maduros, ya que al utilizar el Beneficio neto en su fórmula, las empresas y sectores en crecimiento en muchas ocasiones tienen ingresos muy altos pero beneficios pequeños o incluso negativos, debido a la alta inversión que realizan para crecer, por ello un ratio como el P/E ratio no tendría mucho sentido en este tipo de empresas, y sí lo tiene en las empresas más maduras.

$$\frac{P}{E} \text{ ratio} = \frac{\text{Market Value per share}}{EPS} \quad (2)$$

- **Múltiplos de precio sobre ventas (P/S ratio):** en este caso es un ratio financiero que mide el precio actual de una acción en relación con las ventas por acción de una compañía. Este ratio es útil para valorar empresas que están en un momento de rápido crecimiento, cuyos ingresos aumentan, pero que todavía sus resultados o beneficios no han sido significativos o directamente han sido negativos por la inversión que se esté realizando para crecer. La industria donde más se utilizan es la tecnológica, donde la mayor parte de sus empresas están en una etapa temprana de su ciclo de vida y son *startups*. La contraparte de este ratio es que no valora la

salud financiera de una empresa ya que no tiene en cuenta los costes y gastos, ni la rentabilidad de la misma. En el caso de este trabajo, el hecho de que las empresas analizadas estén establecidas, hace que haya mejores opciones para valorar las empresas del S&P 500 (Fernández, 2007).

$$\frac{P}{S} \text{ ratio} = \frac{\text{Market value per share}}{\text{Sales per share}} \quad (3)$$

- **Múltiplos de valor en libros (P/BV ratio):** este ratio compara el precio actual de una acción con su valor contable por acción. Este valor contable se calcula restando el total de los pasivos de una empresa a sus activos. Este ratio lo suelen usar los inversores para evaluar si una compañía está infravalorada (P/BV bajo) o sobrevalorada (P/BV alto). Existe una variante más utilizada que usa el valor de los activos tangibles en vez del de todos los activos (Bragg, 2023). Este ratio se considera efectivo para aquellos sectores o negocios que requieren una gran densidad de capital como pueden ser energéticas, empresas de transporte o empresas de la construcción, ya que este ratio ignora los activos intangibles, los cuáles son abundantes en las empresas del sector servicios y no tanto en los que se han descrito.

$$\frac{P}{BV} \text{ ratio} = \frac{\text{Market price per share}}{\text{Book Value per share}} \quad (4)$$

- **Múltiplos de valor empresarial (EV/ Ebitda):** este es un ratio financiero utilizado para valorar a una compañía, que compara el valor total de una empresa, con el Ebitda (*Earnings Before Interests Taxes Depreciation and Amortization*), que es la capacidad de esta para generar ganancias operativas, es decir, a través de su actividad principal. De forma muy simplificada, el *Enterprise value*, se calcula a través de la suma de la capitalización bursátil de la compañía, más la deuda neta de la misma. Por otro lado, el Ebitda, se calcula restando los gastos operativos, a los ingresos provenientes de las actividades principales de la empresa. El ratio EV/Ebitda, puede variar según el sector, por ello adquiere sentido cuando se compara o bien con empresas del mismo subsector, o con el histórico de una misma empresa. Al comparar estos, si hay un ratio bajo, puede significar que la compañía esté infravalorada, y un ratio alto sobrevalorada. Es de esperar que en los sectores

de alto crecimiento existan altos ratios EV/Ebitda y en sectores con un crecimiento más lento que este sea bajo (Hayes, 2022).

$$\frac{EV}{EBITDA} \text{ ratio} = \frac{\text{Market cap} + \text{Net Debt}}{EBITDA}$$

(5)

$$EV = \text{Market Cap} + \text{Preferred Stock} + \text{Outstanding Debt} + \\ + \text{Minority interest} - \text{Cash and Equivalents}$$

Como bien se ha explicado, todos los ratios descritos, tienen su respectivo uso y cada uno funciona mejor o peor según el entorno. El *Price to earnings* ratio es el más usado entre los inversores por lo que es más fácil y rápido comparar con otras empresas del S&P500. El EV/Ebitda es una alternativa que tiene en cuenta el balance y la deuda de las empresas.

Una vez se conoce cómo funciona el mercado, y más concretamente el S&P500, en cuanto a cómo se valoran las empresas, como se comparan estas con sus *peers*, ya qué se debe la cotización de las empresas en el mercado, sólo faltaría conocer en qué entorno macroeconómico se desarrolla este trabajo, para conocer todos los factores que pueden afectar a los retornos de las acciones del S&P500.

2.3. Entorno Macroeconómico

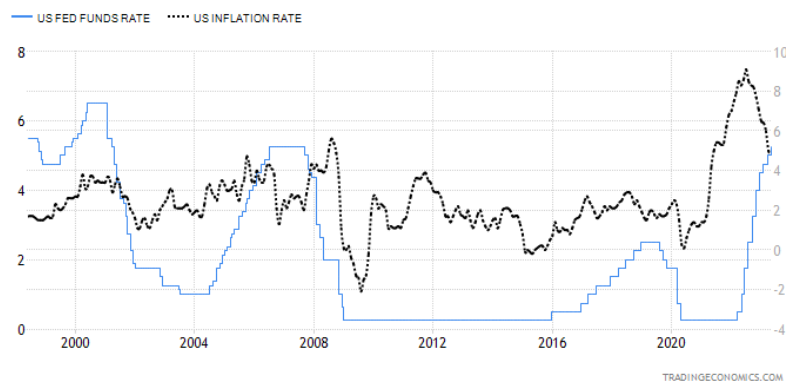
Por último, se quiere describir de forma breve el entorno macroeconómico en el que se desarrollará el trabajo. Ya se está teniendo en cuenta a través de la valoración intrínseca aquello que impulsa el valor de una empresa o acción, además, la valoración relativa ofrece una estimación más realista de este valor en comparación con el sector. Por último, los factores macroeconómicos son los que afectan a todas las empresas de un mercado y que también hay que tener en cuenta.

Primero de todo, describir en qué entorno macroeconómico se va a desarrollar el análisis. Este trabajo, como se ha dejado entrever durante la introducción, se va a llevar a cabo en el mercado estadounidense, más concretamente en el Standard & Poor's 500 (S&P 500), que es un índice bursátil el cual está ponderado por capitalización de mercado

y sigue el desempeño económico de las principales empresas que cotizan en las bolsas estadounidenses. Incluye empresas de diversos sectores y se usa como principal referencia del desempeño del mercado de valores del país.

Los factores macroeconómicos que pueden afectar a las empresas del S&P500 que se estudiarán en este trabajo, al analizar datos pasados de estas compañías, pueden ser: primero, una amenaza de recesión económica debido a la pandemia, la desconfianza generalizada en el mercado desde la crisis de 2008, los factores medioambientales afectarán considerablemente a las empresas que deberán buscar unas emisiones netas cero para 2050, principalmente a los mercados occidentales de los que el S&P500 forma parte. Además, el periodo de 2008 a 2012 que también se va a analizar, está muy marcado por la crisis de 2008. Por último, durante estos últimos años, la inflación que está afectando a EEUU, de forma distinta a Europa, ha supuesto la subida de tipos de interés, los cuales son clave como se ha visto en la valoración intrínseca de empresas, ya que se utiliza para el cálculo del coste de capital en el método del DCF, y esta subida afecta negativamente, al valor presente de los flujos de caja futuros. Además, la subida de tipos de interés afecta al consumo de las personas, ya que lo que se pretende es la congelación parcial de la economía, afectando indirectamente a las empresas.

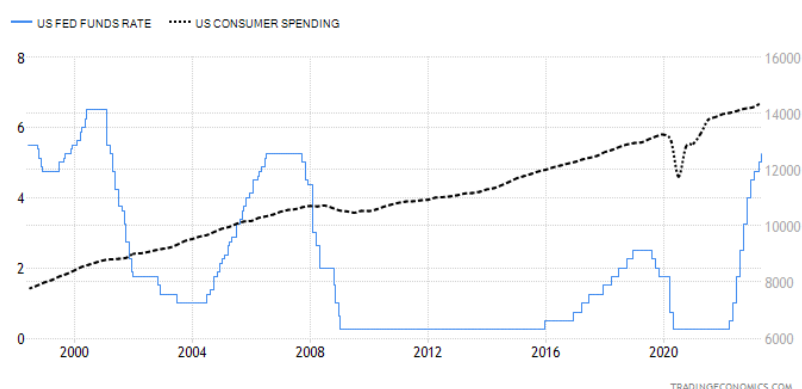
Ilustración 1: Fed Funds Rate vs Inflation (25 years)



Fuente: *Tranding Economics*

Las dos principales finalidades de la Reserva Federal de los Estados Unidos, son la disminución del desempleo, y mantener la inflación entre el 0-2%. Como se puede observar en la imagen, se ha mantenido la inflación entre el 1% y el 4% a lo largo del siglo XXI, con algún que otro pico, pero llama la atención que la inflación en los últimos años, tras la Covid-19, se ha disparado, por esto, la Reserva Federal ha aumentado los tipos de interés.

Ilustración 2: Fed Funds Rate vs Consumption (25years)



Fuente: Trading Economics

Sin embargo, como se puede observar, si bien se han aumentado los tipos para doblegar la curva de la inflación, esto no ha afectado al consumo de las personas, como la teoría dice que debería haberlo hecho, lo cual es una buena noticia, ya que se ha conseguido el propósito, sin congelar la economía, que suele ser el efecto secundario, cuando se aumentan los tipos de interés. Este es el actual entorno macroeconómico en el que se va a desarrollar este trabajo.

Con todo esto, se conoce cómo los inversores valoran las empresas en el mercado, y como las comparan con otras empresas creando grupos de *peers* para un análisis correcto. Y además se presenta el entorno macroeconómico en el que se desarrollará el estudio, se conocen los principales factores que afectan a la cotización de las empresas. A continuación, se puede seguir con la obtención y tratamiento de los datos pertinentes para el estudio de eventos que se realizará, para conocer si la publicación de cuentas trimestrales de empresas comparables afecta a la cotización de las acciones de una empresa.

CAPÍTULO III. OBTENCIÓN, LIMPIEZA, TRATAMIENTO Y ANÁLISIS DE LOS DATOS DEL S&P 500 - METODOLOGÍA

Durante este capítulo se quiere informar y explicar de qué forma se extraerán los datos de las empresas, qué bases de datos se utilizarán, qué lenguaje de programación se usará, y todo lo relativo a los pasos previos al análisis.

Para poder sacar conclusiones de cómo afecta la publicación de cuentas trimestrales de empresas comparables a la cotización de una empresa, se necesitan series de datos de las empresas del S&P500, tanto cuantitativos para analizarlos (Capitalización bursátil, fechas de publicación etc.), como cualitativos para identificar si realmente son o no comparables (sector, subindustria etc.). Por ello, se necesitará hacer uso de herramientas de *Data Analytics*, para analizar los valores diarios de la cotización de las empresas, las fechas clave en las que estas publican las cuentas trimestrales los distintos sectores que se van a analizar. Para realizar este análisis se utilizará Python con bibliotecas como numpy, pandas, scipy, *BeautifulSoup* (para realizar el webscraping), pyplot (para realizar gráficos), entre otras, que se pueden observar en el **Anexo I**. Además, se hará uso de la base de datos Factset, para la obtención de la cotización de las empresas.

Primero de todo se comenzará por agrupar a las 500 compañías, en sectores, se dividirán en los 11 principales que forman el S&P500. Estos sectores son: El sector **Tecnológico**, que es un 20,8% del total del S&P500 en términos de capitalización de mercado; el segundo será el sector de la **Salud**, cuya cuota es del 14,9%; en tercer lugar, el sector **Financiero**, con 13,7%; el último con más de un 10% de cuota es el sector de **Productos de consumo Discrecional** (en este grupo se incluyen aquellos bienes de consumo más premium). Entre el 5% y el 10% de cuota se encuentran sectores como el de las **comunicaciones** (9%), el sector **Industrial** (9,7%), el sector de **Bienes de consumo Básico** (6,7%), y el **Energético** (6%). Por último, los otros 3 sectores con menos de un 5% de cuota, pero que siguen siendo relevantes y existen diferentes empresas entre las que comparar, son el sector **Utilities**, **Materiales** y **Real Estate**.

Más tarde, dentro de cada sector, se dividirán de nuevo en subsectores, existen 126 subsectores en total, por lo que cada industria se compone de 11 subsectores diferentes de media. El subsector, será la característica principal que dividirá en distintos grupos de *peers* al conjunto de las compañías del índice. Otras de las características enumeradas anteriormente, se podrían tener en cuenta para la selección de *peers*, pero no varían mucho

por el hecho de cotizar en el S&P500, esto se debe a que todas las compañías tienen un tamaño muy grande. Además, en el caso de que no tengan su sede en Estados Unidos, sí se ven afectados por una situación socio-económica parecida, y se encuentran en un momento del ciclo de vida muy parecido, ya que son compañías muy establecidas en el mercado. Por tanto, la principal característica que diferencia unas de las otras, es el subsector al que pertenecen, a través del cual se pueden crear grupos de empresas comparables.

Además, para diferenciar cada compañía, en el S&P500 utilizan una secuencia de letras, llamada *ticker*, así como una secuencia de números llamado código CIK, que es diferente para cada compañía y son necesarios para saber a qué compañía se refiere una serie de datos. El *ticker* y el código CIK también serán necesarios.

En conclusión, durante este capítulo se procederá a la recopilación de los datos del S&P500, tanto los datos históricos de la cotización diaria, como los informes financieros o cuentas que las empresas publican trimestralmente, necesarios para realizar el análisis. En segundo lugar, de forma paralela a la recopilación de los datos, se irán tratando los datos, esto puede incluir el eliminar datos faltantes, normalizar valores, y las transformaciones que sean necesarias.

3.1. Identificación de empresas

En primer lugar, para realizar la recopilación de los datos, lo primero será la identificación de las empresas que se van a analizar, y la fecha de comienzo desde la cual se empezará el análisis, que será un intervalo de 15 años, entre el 8 de mayo de 2008, hasta el 8 de mayo de 2023, sin tener en cuenta fines de semana ni festivos, ya que cierra la bolsa americana. La fecha de análisis variará a lo largo del trabajo ya que es dependiente de otros datos, es decir, cuando por ejemplo se calcule el retorno existirán valores faltantes en la primera fecha, y como este caso, otros más. Los datos del *Market Cap*, sí se obtendrán en este intervalo de fechas.

Para la identificación de las empresas, se ha realizado un *webscrapping* de Wikipedia haciendo uso de la función `read_html` de la librería Pandas, como se puede observar en el **Anexo II**. Este es el vínculo de donde se han cogido los datos (link)

Posteriormente se pasa la información a diferentes *dataframes*, este tipo de variables comenzarán por las letras df, y le seguirán algún tipo de palabra que explique su contenido. Además, se pasa el *dataframe* obtenido del *webscrapping*, a diferentes archivos csv, a través de la función `.to_csv` de la biblioteca de Python Pandas. Esto se hace para poder visualizar mejor la información que está contenida en cada *dataframe* que se va creando, ya que Python solo muestra un número determinado de observaciones.

En este caso, en el *dataframe* creado, está contenida la información para cada empresa y en orden de: su *ticker* (**Symbol**), que es único para cada empresa del S&P500; el nombre completo por el que se conoce a cada compañía (**Security**); el sector al que pertenece dicha compañía (**GICS Sector**); la subindustria dentro del sector, a la que pertenece cada compañía (**GICS Sub-Industry**); la localización de la sede principal de cada empresa (**Headquarters Location**); la fecha en la que se incorporó al índice (**Date added**); el Central Key Index (**CIK**), que es un número identificador asignado por la Comisión de Valores y Bolsa de los Estados Unidos, parecido al código CIF en España; y por último, cuándo fue fundada cada compañía (**Founded**). En la siguiente tabla se observa algunos ejemplos de la información descargada:

Ilustración 3: Webscrapping Compañías S&P 500

Symbol	Security	GICS Sector	GICS Sub-Industry	Headquarters Location	Date added	CIK	Founded
MMM	3M	Industrials	Industrial Conglomerates	Saint Paul, Minnesota	04/03/1957	66740	1902
AOS	A. O. Smith	Industrials	Building Products	Milwaukee, Wisconsin	26/07/2017	91142	1916
ABT	Abbott	Health Care	Health Care Equipment	North Chicago, Illinois	04/03/1957	1800	1888
ABBV	AbbVie	Health Care	Pharmaceuticals	North Chicago, Illinois	31/12/2012	1551152	1888
ACN	Accenture	Information Technology	IT Consulting & Other Services	Dublin, Ireland	06/07/2011	1467373	1989
ATVI	Activision Blizzard	Communication Services	Interactive Home Entertainment	Santa Monica, California	31/08/2015	718877	2008
ADM	ADM	Consumer Staples	Agricultural Products & Services	Chicago, Illinois	04/03/1957	7084	1902
ADBE	Adobe Inc.	Information Technology	Application Software	San Jose, California	05/05/1997	796343	1982
ADP	ADP	Industrials	Human Resource & Employment Services	Roseland, New Jersey	31/03/1981	8670	1949
AAP	Advance Auto Parts	Consumer Discretionary	Automotive Retail	Raleigh, North Carolina	09/07/2015	1158449	1932
AES	AES Corporation	Utilities	Independent Power Producers & Energy Traders	Arlington, Virginia	02/10/1998	874761	1981

Fuente: Creación propia (Wikipedia)

Además, con estos datos se crea un diccionario en Python que solo contenga cada número CIK, con el respectivo *ticker* para cada empresa. Esto será útil para futuros apartados, ya que en la mayor parte de bases de datos a las empresas se las reconoce por alguna de estas dos.

3.2. Agrupación por Subindustrias

Como se ha estado adelantando durante la introducción, una parte importante para poder comparar empresas, es agrupar a estas por características parecidas y crear grupos

de *peers*. Para ello se utilizará la subindustria de cada empresa, y se crearán grupos según estas. Como se ha mencionado anteriormente, hay 11 sectores, 126 subindustrias (11 por sector de media) y 500 empresas, por lo que aproximadamente existen 45 empresas por sector, y 4 empresas por subindustria, por lo que los grupos de empresas comparables tendrán de media este tamaño.

Para esto, se va a usar el *webscrapping* realizado para la obtención de las empresas, ya que, como se ha observado, en el anterior link había datos para cada empresa de su sector y subindustria. Una vez realizado el *webscrapping*, de la misma forma, se elegirán las columnas deseadas para un nuevo *dataframe* que se va a crear. En este caso, interesa escoger una columna que contenga un valor único para cada empresa, ya sea el código CIK, o el *ticker*, en mi caso he escogido el *ticker*, ya que es el factor común con los demás *dataframes*. La otra columna necesaria, es la de subsector, una vez escogidas estas dos columnas, se creará un *dataframe*, en el que para cada valor de *ticker* solo existe una subindustria.

Posteriormente, se realizará la agrupación por subindustria a través de la función *.groupby* de la biblioteca Pandas. De esta forma se obtiene un *dataframe* en el que una columna es la industria y la otra es una lista de los *ticker* de las empresas que pertenecen a cada industria. Se puede descargar a un csv para mostrarlo. Esto es un ejemplo de las 11 primeras subindustrias, en este caso ordenadas por orden alfabético:

Ilustración 4: Primeras once subindustrias con sus respectivas empresas

GICS Sub-Industry	Symbol
Advertising	['IPG', 'OMC']
Aerospace & Defense	['AXON', 'BA', 'GD', 'HWM', 'HII', 'LHX', 'LMT', 'NOC', 'RTX', 'TXT', 'TDG']
Agricultural & Farm Machinery	['DE']
Agricultural Products & Services	['ADM', 'BG']
Air Freight & Logistics	['CHRW', 'EXPD', 'FDX', 'UPS']
Apparel Retail	['ROST', 'TJX']
Apparel, Accessories & Luxury Goods	['NKE', 'RL', 'TPR', 'VFC']
Application Software	['ADBE', 'ANSS', 'ADSK', 'CDNS', 'CDAY', 'FICO', 'GEN', 'INTU', 'ORCL', 'PAYC', 'PTC', 'CRM', 'SNPS', 'TYL']
Asset Management & Custody Banks	['AMP', 'BLK', 'BK', 'BEN', 'IVZ', 'NTRS', 'STT', 'TROW']
Automobile Manufacturers	['F', 'GM', 'TSLA']
Automotive Parts & Equipment	['APT', 'BWA']

Fuente: Creación propia

Como se puede observar, hay industrias con mayor cantidad de empresas y otras con menos, ya se puede sacar alguna conclusión, como que las subindustrias que están relacionadas con la tecnología parece que poseen mayor cantidad de empresas, en el caso de la subindustria *Application Software* catorce, que otras como por ejemplo la de maquinaria de agricultura, de la que solo forma parte una empresa del índice.

Como el caso de la subindustria de maquinaria de agricultura hay varios casos más, es decir, existen varias subindustrias, en las cuales sólo está incluida una empresa del índice, y por tanto no existe un grupo de *peers*. Para este estudio, eso no tiene sentido ya que se va a estudiar la relación entre una empresa y sus *peers*, si hay empresas que no tienen *peers* el estudio carece de sentido. Por ello, se decide eliminar, a través de la función `.apply(len)>1`, aquellas industrias de las que solo forme parte una empresa, algunos de estos subsectores, son: el subsector del cobre, oro, distribuidores de comida etc. En total, el número de subindustrias eliminadas, y, por tanto, también el número de empresas eliminadas, es igual a 30.

Tras la limpieza de datos y ajuste del *dataframe*, quedan aquellas subindustrias que integran dos o más empresas del índice, (en total 96 subindustrias), y en vez de ordenar las compañías en listas, se ha reconvertido la columna *Industry* al índice del *dataframe* y se han creado 18 columnas, que es el número máximo de *peers* que tiene una industria, cada columna contiene un *ticker* de un *peer*. Todo el proceso de limpieza y obtención del *dataframe* `df_peers`, que será el utilizado para acceder a los grupos de *peers* se puede observar en el **Anexo III**.

3.3. Publicación de resultados de las empresas

Una vez ya se conoce qué empresas se van a analizar y cuál es su grupo de *peers*, la siguiente parte importante para realizar el estudio de eventos, y saber cómo afecta la publicación de resultados de las empresas comparables a la cotización de una acción, es conocer cuándo se producen esas publicaciones, es decir, las fechas en las que las compañías del S&P500, publican las cuentas trimestrales (10-Q) y anuales (10-K).

Existe una página oficial del SEC (*U.S. Securities and Exchange Commission*) llamada EDGAR (*Electronic Data, Gathering, Analysis and Retrieval*), la cual contiene diferente información de empresas del S&P500, como por ejemplo, cuándo publican las cuentas anuales y trimestrales estas empresas, este es el vínculo que lleva a estas páginas, con el ejemplo de la empresa 3M ([link](#)). Lo único necesario para que cambie la empresa es escribir el código CIK en una parte concreta del vínculo. Por ello, se puede realizar un *webscrapping* de cada empresa, gracias a la biblioteca Beautiful Soup. Habrá que usar el *dataframe* que contiene los códigos CIK obtenidos del primer *webscrapping*, y con un

bucle for, ir iterando sobre cada fila que contiene estos códigos para ir realizando el *webscrapping* para cada empresa. De esta forma se obtendrá una base de datos en las que se encuentran las últimas 80 publicaciones de cuentas trimestrales. Tras importarla a csv, se muestra de la siguiente manera:

Ilustración 5: Fechas de publicación compañías (3M)

Filings	Filing Date	File/Film Number	Ticker	CIK
10-Q	25/04/2023	001-0328523843854	MMM	66740
10-K	08/02/2023	001-0328523598184	MMM	66740
10-Q	25/10/2022	001-03285221328846	MMM	66740
10-Q	27/07/2022	001-03285221109744	MMM	66740
10-Q	26/04/2022	001-0328522853129	MMM	66740
10-K	09/02/2022	001-0328522606093	MMM	66740
10-Q	26/10/2021	001-03285211347847	MMM	66740
10-Q	27/07/2021	001-03285211117640	MMM	66740

Fuente: Creación propia (EDGAR)

El resultado final de este *webscrapping*, crea una base de datos con las fechas de publicación de todas las empresas del S&P500, de más de 46.500 filas. Aunque, aún queda por realizar varios cambios a estos datos. En primer lugar, dado que los datos de Factset son trimestrales, las cuentas anuales (10-K) dejarán de ser necesarias, y no se tendrán en cuenta por el momento, aunque si durante el análisis se quiere acceder a estos datos habrá que realizar el *webscrapping* de EDGAR de nuevo, lo cual no supone un gran problema. Además, se eliminarán aquellas empresas que se han eliminado en el apartado anterior porque no tenían ninguna otra empresa con la que comparar en su subindustria. Para ello se crea un nuevo *dataframe* denominado `cik_tik_filtered` que a través de la función `.isin`, se borran aquellas compañías que se han borrado con anterioridad, porque pertenecían a una industria sin más *peers*, pasando a haber 473 compañías. En el nuevo dataset de las fechas de publicación las filas se han reducido a poco más de 41.000, en el **Anexo IV**, se puede observar directamente cómo se ha realizado el *webscrapping* de EDGAR, de las compañías que pasaban el filtro impuesto, pero lo único que se ha cambiado en este código es allí donde ponía `cik_tik`, ahora se puede observar `cik_tik_filtered`.

3.4. Métricas e información financiera a utilizar

Como se ha dejado entrever durante la introducción de este trabajo hay métricas e información financiera de las empresas, necesarias para poder hacer uso de los métodos de valoración relativa, como pueden ser el *Enterprise Value*, el Ebitda, el Resultado Neto o la Capitalización Bursátil (*Market Cap*). Sin embargo, para el estudio de eventos, sólo se requiere el *Market Cap*. Se utilizarán las dos siguientes métricas para este estudio:

El **Market Cap**: que es la capitalización bursátil de la empresa, es decir el valor que le da el mercado al total de las acciones de la misma. El *Market Cap* sirve para conocer cómo cambia el valor de mercado de una empresa cuando ocurren las publicaciones de los *peers*, será lo equivalente a observarlo en el precio de la acción, pero es más exacto, ya que tiene en cuenta el número de acciones de la empresa. Esto se utilizará para calcular los retornos que tienen las acciones de una empresa a lo largo de los quince años que se están estudiando.

El **Retorno**: el retorno se refiere a la ganancia o pérdida que un inversor recibe cuando posee las acciones de una empresa. El retorno se puede medir durante diferentes periodos de tiempo, en el caso de este trabajo se medirá el retorno diario de las empresas. Para el cálculo del retorno diario, se necesita comparar el precio de una acción en un día concreto y el precio de dicha acción el día anterior. Existen diferentes formas de calcular el retorno, ya que se puede realizar el retorno continuo, que se calcula a través del logaritmo neperiano del valor final entre el inicial. O también, el retorno simple, que se calcula como:

$$\text{Retorno Simple} = \left(\frac{\text{Valor Final} - \text{Valor Inicial}}{\text{Valor Inicial}} \right) - 1 \quad (6)$$

Para este trabajo, se ha utilizado el retorno simple, ya que no ofrece valor añadido el conocer los retornos continuos, utilizando logaritmos. Es importante destacar que los retornos pueden ser positivos o negativos, indicando ganancias o pérdidas en cada caso.

Una parte importante de este proceso de limpieza y orden de los datos, ha sido transponer los *dataframes*, ya que se encontraban en las filas los *tickers* de las empresas y en las columnas las fechas. Para realizar un análisis más cómodo y correcto se requería transponer los *dataframes*. Posteriormente a esto se ordena de fecha más lejana (2008-

05-08) a más cercana (2023-05-08). Aunque parezca algo indiferente, el ordenarlo de esta forma, cobrará sentido a la hora de realizar el análisis.

Además, observando detenidamente el *dataframe* del *Market Cap*, se observan datos faltantes (NaN) de distintos tipos. En primer lugar, se encuentran datos faltantes para todas las fechas de algunas compañías, estas compañías se decide eliminarlas. Por otro lado, existen compañías que tienen datos faltantes en las fechas más lejanas esto se puede deber o bien a que entraron más tarde de 2008 en el índice o porque simplemente no se tenían datos para algunas fechas, en el caso de estas compañías se decide mantenerlas, y dejar los datos faltantes como NaN ya que no afectará al análisis, únicamente se comenzará a analizar desde el momento que se tienen datos. Por último, los otros datos faltantes encontrados han sido algunos huecos vacíos entre medias de dos datos no faltantes, este caso se ha tratado de una forma distinta, ya que se ha rellenado con el mismo valor de cotización que el del día anterior, utilizando la función `.fillna(method='ffill')`.

Una vez se ha ordenado el *dataframe*, y limpiado y rellenado los valores faltantes, se puede proceder al cálculo de los retornos.

3.5. Bases de datos

Tras haber recopilado todos los datos necesarios para realizar el análisis, el panorama y formato final de los distintos es el siguiente. Existen actualmente tres *dataframes* que se utilizarán para el análisis, y estos son:

1. **df_peers**: *dataframe* en el que se recopilan la información de los *peers*. En cada fila hay una industria, y mínimo dos *peer* para cada industria, hay industrias que agrupan hasta 18 empresas, por tanto, existen 19 columnas de las cuales 18 son *peers*. Existen celdas vacías para aquellas industrias que tengan menos *peers*, que son la mayoría.

Ilustración 6: Parte del dataframe de peers transformado a Excel

Industry	Peer1	Peer2	Peer3	Peer4	Peer5	Peer6	Peer7	Peer8	Peer9	Peer10
Advertising	IPG	OMC								
Aerospace & Defense	AXON	BA	GD	HWM	HII	LHX	LMT	NOC	RTX	TXT
Agricultural Products & Services	ADM	BG								
Air Freight & Logistics	CHRW	EXPD	FDX	UPS						
Apparel Retail	ROST	TJX								
Apparel Accessories & Luxury Goods	NKE	RL	TPR	VFC						
Application Software	ADBE	ANSS	ADSK	CDNS	CDAY	FICO	GEN	INTU	ORCL	PAYC
Asset Management & Custody Banks	AMP	BLK	BEN	IVZ	TROW					
Automobile Manufacturers	F	GM	TSLA							

Fuente: Creación propia

- df_date:** todas las fechas de la publicación de cuentas trimestrales y anuales por cada empresa. Solo existen 2 columnas, la primera contiene el *ticker* de la empresa, la segunda contiene una lista con todas las fechas de publicaciones trimestrales recogidas. En la ilustración solo aparecen los primeros valores de las listas de fechas, pero estas, están hasta el 8 de mayo de 2008. Este dataframe necesitará ser transformado para el análisis ya que contiene las fechas de publicación de cada empresa, y lo que interesa son las fechas de publicación de las demás empresas de su subsector, esto realizará en cada análisis.

Ilustración 7: Parte del dataframe de fechas de publicación

MMM	AOS	ABT	ABBV	ACN	ATVI	ADM	ADBE
08/02/2023	14/02/2023	17/02/2023	17/02/2023	12/10/2022	23/02/2023	14/02/2023	17/01/2023
09/02/2022	11/02/2022	18/02/2022	18/02/2022	15/10/2021	25/02/2022	17/02/2022	21/01/2022
04/02/2021	12/02/2021	19/02/2021	19/02/2021	22/10/2020	23/02/2021	18/02/2021	15/01/2021
06/02/2020	24/02/2020	21/02/2020	21/02/2020	29/10/2019	27/02/2020	18/02/2020	21/01/2020
07/02/2019	15/02/2019	22/02/2019	27/02/2019	24/10/2018	28/02/2019	19/02/2019	25/01/2019
08/02/2018	16/02/2018	16/02/2018	16/02/2018	26/10/2017	27/02/2018	16/02/2018	22/01/2018
09/02/2017	17/02/2017	17/02/2017	17/02/2017	28/10/2016	28/02/2017	17/02/2017	20/01/2017
11/02/2016	17/02/2016	19/02/2016	19/02/2016	30/10/2015	29/02/2016	19/02/2016	19/01/2016
12/02/2015	17/02/2015	27/02/2015	20/02/2015	24/10/2014	26/02/2015	20/02/2015	20/01/2015

Fuente: Creación propia

- df_mc_tp:** datos del *Market cap* diario de las empresas que se van a analizar. Los datos del *Market Cap* son diarios, comenzando el 8 de mayo de 2008 y terminando el 8 de mayo de 2023. No se ha podido ilustrar en formato Excel ya que se importaba incorrectamente, sin embargo, así se ve cuando se realiza un *print()* del *dataframe* en Python.

Ilustración 8: Print de los datos del Market Cap en Python

Dates	CTAS	RTX	WEC	MAA	\
2008-05-08	4565.951670	72075.484345	5621.898522	1420.587720	
2008-05-09	4589.004505	71170.404347	5613.713151	1419.537936	
2008-05-12	4639.720215	72571.820454	5621.898522	1460.741007	
2008-05-13	4633.572714	72912.440362	5630.083446	1464.152755	
2008-05-14	4670.456841	73146.014663	5633.591143	1465.989827	
...	
2023-05-02	46668.924828	144463.110367	30045.139915	17522.286357	
2023-05-03	46580.444754	141964.558940	29739.168027	17449.957913	
2023-05-04	46572.306871	140152.734873	30076.682888	17434.791583	
2023-05-05	47430.670355	141117.093923	30284.870840	17777.771381	
2023-05-08	47049.796875	140225.796875	29988.361328	17658.777344	

Fuente: Creación propia (datos de Fectset)

Antes de comenzar la importación y limpieza de datos, se preveía usar las 500 empresas del índice para el análisis, no obstante, se han borrado diferentes empresas por dos razones principales como se ha ido explicando a lo largo del trabajo. La primera razón es que existen empresas que al analizar sus subsectores no había otras empresas que formaran parte de ese subsector, por tanto, no existe un grupo de comparables, esas empresas e industrias se eliminan. Además, para valores de *Market Cap*, en la base de datos FactSet, no había datos para algunas empresas, por ello también se decide eliminar dichas empresas. En total, tras la limpieza de los datos, se concluye en un conjunto de 426 empresas de las que se tienen los datos anteriormente descritos, desde el 8 de mayo de 2008 al 8 de mayo de 2023. Aclarar, que para algunas empresas no se tienen datos tan antiguos como el 8 de mayo de 2008 ya que entraron posteriormente al índice, pero aun así se podrán utilizar para el análisis, aunque sea a partir de la fecha en la que se tienen datos.

CAPÍTULO IV. ANÁLISIS DE EVENTOS

El análisis de eventos nace directamente de la premisa que el mercado es eficiente, ya que la usabilidad de este tipo de estudio existe cuando el mercado reacciona rápidamente al evento, y, por tanto, el efecto en la cotización de las acciones es relativamente inmediato. Este trabajo, tal y como se mencionó al comienzo, se desarrolla en el marco teórico de un mercado eficiente, por tanto, se podrá llevar a cabo un estudio de eventos (MacKinlay, 1997).

Para el desarrollo del estudio de eventos se seguirá el artículo realizado por Craig MacKinlay (1997), el cual es considerado una referencia clásica en el estudio de eventos. Se puede realizar un estudio de muy diversos eventos, pero en este trabajo, los eventos principales a analizar serán las diferentes publicaciones de resultados trimestrales de las empresas comparables, desde el segundo trimestre de 2008 hasta el último trimestre del año 2022, ya que se poseen datos del *market cap* del primer trimestre del 2023 pero no datos de las fechas de publicación de todas las empresas en este trimestre. La finalidad principal será analizar si esta publicación de resultados afecta a los retornos en la cotización de las empresas o no.

En el estudio de eventos, la hipótesis nula (H_0) siempre será que el evento, en este caso la publicación de resultados, no tiene un impacto en la cotización de las empresas, y, por tanto, la hipótesis alternativa (H_1) será que sí tiene un impacto.

Además, un factor importante es definir los periodos que se van a examinar, es decir, cuantos días antes del evento se van a utilizar para analizar los retornos y cuantos días después se van a analizar para observar si estos retornos han variado o no. Para este trabajo se utilizarán, en primer lugar, los cinco días anteriores (incluyendo el del evento) y los cinco días posteriores (sin incluir el del evento), y más tarde con los diez días anteriores y los diez días posteriores, que es lo aconsejado por el profesor de Wharton Xuan Liu (2020).

A este periodo de tiempo se le denomina *event window*. Quiero destacar que cuanto más grande sea esta ventana de evento, existirán más posibilidades de que los resultados se vean tergiversados por otros eventos que afectan a la compañía. Por ello, se elige realizar el análisis con cinco días, porque se puede dar el caso que las empresas que tengan muchos *peers*, en una misma semana, más de una empresa comparable publique cuentas. Esto es porque estas publicaciones tienden a aglutinarse en las mismas fechas para todas

las empresas. Por otro lado, si la ventana es demasiado pequeña, es arriesgado ya que podrían no haberse visto afectados los retornos aún, aunque si el mercado es fuertemente eficiente, cinco días deberían ser más que suficientes para que se pudiera ver reflejado en los retornos, aun así, se decide hacer un análisis con diez días como *event window*, para realizar esta comparación.

Se entiende entonces que para llevar a cabo este análisis se necesitarán varios nuevos *dataframes*, que se sacarán de los datos ya tratados. Estos nuevos *dataframes* son:

- **Retornos:** se llamará de diferente manera a lo largo del código según qué retornos contenga, pero se calculará como un retorno simple ($Market\ Cap\ final - Market\ Cap\ inicial$) / $Market\ Cap\ inicial$ todo menos uno. Para este cálculo, se usará en Python la función `.shift(1)`, que escoge el valor anterior al analizado.
- **Retornos en el *event window*:** en este caso, gracias a la función `.rolling`, se realizará la media de los retornos de los cinco días anteriores a una fecha, incluyendo esta fecha (en total cuatro días), ya que así lo dice MacKinlay en su artículo. Y la media de los retornos de los cinco días posteriores a la fecha, en este caso sin incluir la fecha del evento, y con la misma función en Python.
- **Desviación estándar:** la desviación estándar indica que tan dispersos están unos valores respecto a su media. En el análisis de eventos, se puede hacer uso de la desviación típica para estudiar la volatilidad de la acción en el periodo anterior y posterior al evento, un aumento de la desviación típica significa un aumento en la volatilidad de las acciones.

Estos *dataframes* se crearán para todas las fechas de las que se tiene datos del *market cap* y más tarde se filtrarán para aquellas fechas en las que han ocurrido eventos y se compara.

4.1. Análisis de medias con el retorno actual

Durante este apartado se realizará un análisis distinto y menos sofisticado, lo que se llama en inglés un *naive analysis*, que el que se hará más adelante, es un análisis de eventos predecesor al propuesto por MacKinlay (1997). Es menos sofisticado ya que MacKinlay propone que para una correcta valuación del impacto del evento, se requiere el uso del retorno esperado, para poder calcular el retorno anormal de una compañía. El

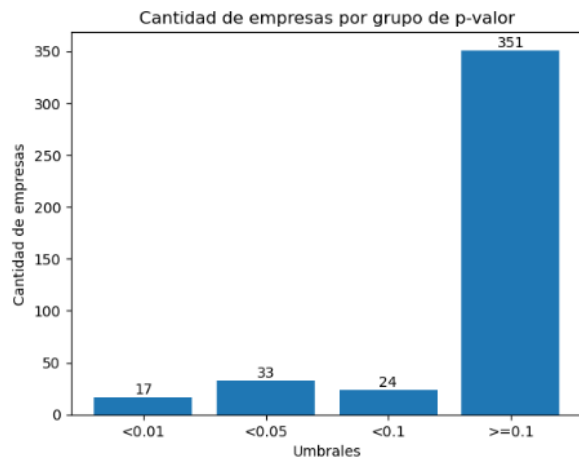
retorno anormal es la diferencia entre los retornos actuales de una empresa y los retornos esperados de la misma. Más tarde se hará uso de este retorno anormal antes del evento y después del evento, para el análisis.

Sin embargo, en este primer análisis se tendrá en cuenta únicamente el retorno actual y no el retorno esperado. Por tanto, se calculará el retorno anterior y posterior con el retorno actual de las empresas, y se realizará un *ttest* para observar si existe una diferencia significativa entre el conjunto de valores de la media de los retornos de cinco días antes del evento, y el conjunto de los valores de los cinco días posteriores a ese mismo evento, para cada compañía, sin tener en cuenta las desviaciones típicas de cada grupo.

Los resultados de este test, se irán guardando en un diccionario en Python en el que la clave será la empresa y el contenido será una lista con la *tstat* y el pvalor. Si el pvalor es inferior a 0,1, se podrá rechazar la hipótesis nula de que la media de los cinco días anteriores al evento es igual que la media de los cinco días posteriores al evento. Hay que tener en cuenta que este análisis se realiza sobre todos los valores a la vez, y no fecha a fecha, por tanto, el *ttest* medirá si para el conjunto de todas las fechas es estadísticamente relevante el cambio en la media.

Es interesante comparar entre dos tipos de eventos, las cuentas anuales y las trimestrales, y así observar si de media para cada compañía, la publicación de cuentas anuales afecta más o menos a la cotización que la publicación de cuentas trimestrales. Para el caso de la publicación de cuentas trimestrales, se puede observar que solo es estadísticamente significativo el cambio en el retorno para el 17% de las compañías, como se puede observar en la siguiente imagen:

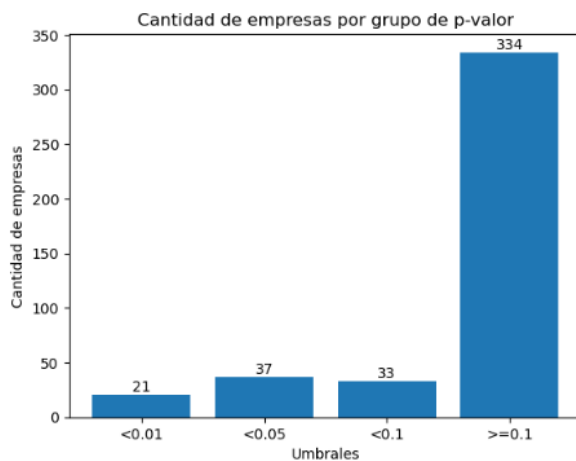
Ilustración 8: Empresas contenidas para cada umbral de pvalor en la publicación de cuentas trimestrales



Fuente: Elaboración propia (Python)

Por otro lado, cuando se realiza el análisis para los eventos de cuentas anuales los resultados son ligeramente diferentes como se puede observar en la siguiente imagen.

Ilustración 9: empresas contenidas en cada umbral de pvalor en la publicación de cuentas anuales



Fuente: Elaboración propia (Python)

En este caso, para el 21% de las compañías, es estadísticamente relevante de media el cambio en todos los retornos posteriores a la fecha de publicación anual. Esto puede sugerir que el mercado reacciona más fuertemente respecto a la publicación de cuentas anuales que a la publicación de cuentas trimestrales.

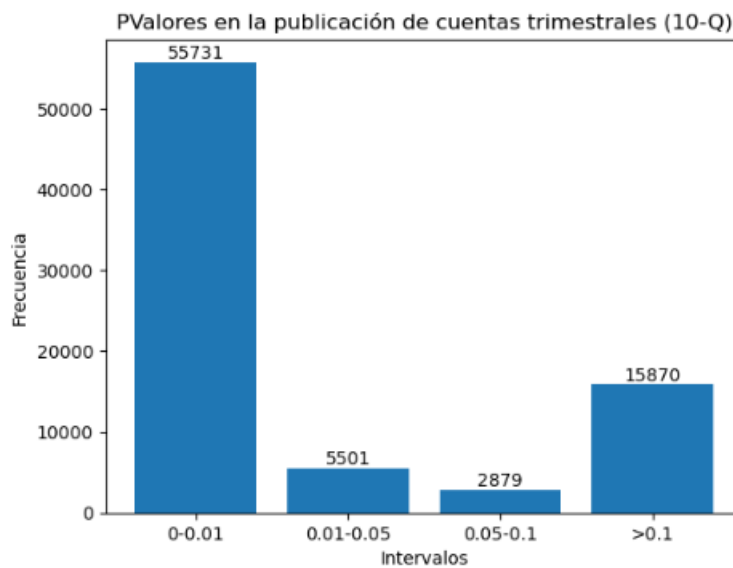
4.2. Análisis de medias y desviación típica con retorno actual

En este apartado no solo se tendrá en cuenta la media de los retornos de los cinco días anteriores y posteriores al evento, sino que también se tendrá en cuenta la desviación típica en estos periodos para observar si cambia o no tras el evento. En este caso, se observan todas las fechas para cada empresa, ya que se estará teniendo en cuenta la variación de la desviación típica durante los varios días antes y después del evento, así como la media de los retornos de los cinco días anteriores al evento y de los cinco días posteriores al evento. La hipótesis nula, es parecida al anterior análisis:

- H0: la publicación de resultados no afecta a los retornos de las empresas, ni a la volatilidad de estos retornos
- H1: la publicación de resultados sí afecta a los retornos y a la volatilidad de estos.

Para realizar este análisis se ha utilizado una función de Python que desarrolla el *ttest*, ya que permite incluir las desviaciones típicas de cada grupo que se compara, esta función es `.ttest_ind_from_stats`. En este caso la función no devuelve solo un pvalor, ya que no está observando si el conjunto de retornos anteriores y posteriores al evento para una empresa son estadísticamente diferentes. Sino que compara cada media de los retornos de los cinco días anteriores y posteriores para cada fecha de publicación de resultados de empresas comparables, teniendo en cuenta la desviación típica de cada empresa en los días anteriores al evento.

Ilustración 10: Distribución de los pvalores de cada evento (publicación de cuentas trimestrales)

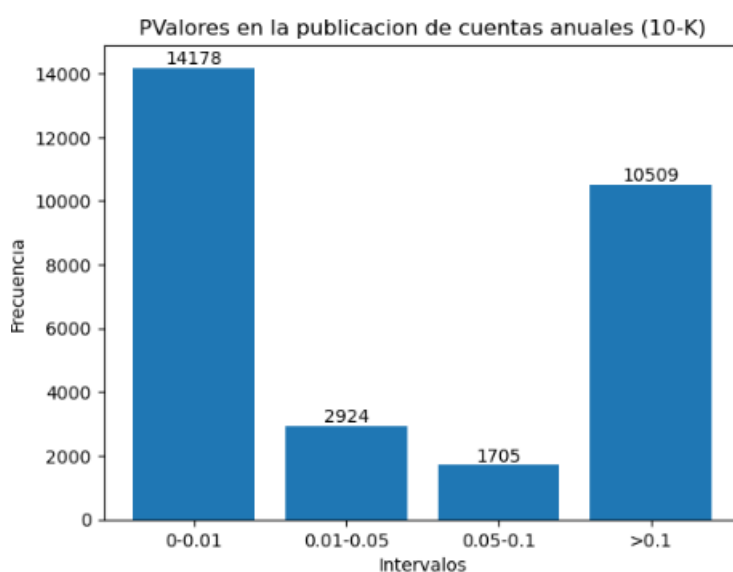


Fuente: Elaboración propia (Python)

Se han analizado todas las fechas de publicación de cuentas trimestrales de *peers* para cada empresa, en total, casi 80.000 eventos. En el 80% de los eventos, es estadísticamente significativo que se produce un cambio en la media y las desviaciones típicas de los retornos de las empresas tras la publicación de cuentas trimestrales.

Se puede llevar a cabo una comparación igual a la realizada en el anterior análisis, para las cuentas anuales. En este caso, como el foco se está poniendo en el evento concreto y no en las compañías, salen menos observaciones que en el análisis de cuentas trimestrales, lo cual tiene sentido, ya que se publican menos cuentas anuales que trimestrales, y por tanto ocurren menos eventos. En este análisis ha ocurrido lo contrario, porcentualmente, las cuentas anuales son menos significativas estadísticamente hablando. Ya que solo para el 64% de los eventos ocurre un cambio entre la desviación típica y la media de los retornos anteriores y posteriores a la publicación de cuentas anuales de los *peers*.

Ilustración 11: Distribución de los pvalores de cada evento (publicación de cuentas anuales)



Fuente: Elaboración propia (Python)

En este análisis se ha podido observar que cuando se tiene en cuenta cada evento, por sí solo, sí es más significativa la variación en la media de los retornos y la desviación típica antes y después del evento. Sin embargo, no se corrobora la conclusión extraída en el primer análisis de que la publicación de cuentas anuales genera un mayor cambio en la media de los retornos y en las desviaciones típicas que la publicación de cuentas trimestrales.

4.3. Análisis con retorno anormal

Tras realizar estos dos análisis *naive* teniendo en cuenta los retornos “brutos” de la cotización de las acciones de las empresas, se va proceder a realizar el análisis como lo propone MacKinlay, utilizando el *abnormal return*, en español retorno anormal de las empresas. Este retorno anormal se refiere al retorno de la empresa sin tener en cuenta el retorno esperado de la misma. Por tanto, como ya se posee el retorno actual de la empresa, que ha sido el utilizado en el análisis anterior, habrá que calcular el retorno esperado para cada compañía en cada fecha.

Este retorno esperado se puede calcular de diferentes formas. MacKinlay (1997) propone diferentes procesos por los que calcularlo. Los dos principales son:

- **Constant Mean Return Model:** es un modelo teórico en el que se asume que los retornos de un activo financiero son constantes a lo largo del tiempo. Teniendo esto en cuenta el rendimiento medio del activo se mantiene estable y no se ve influido por factores externos. (Brown & Warner, 1980).
- **Capital Asset Pricing Model (CAPM):** este es un modelo estadístico que relaciona el retorno de una compañía con el retorno del portfolio del mercado. Se utiliza para determinar el retorno esperado de una compañía y permite el cálculo de este considerando la sensibilidad (beta) de un activo a los movimientos de mercado (retorno de mercado) y el *risk free rate* (que es el retorno que tiene un activo libre de riesgo). Este ha sido el modelo utilizado para el cálculo del retorno esperado.

$$R_i = R_f + \beta(R_m - R_f) + \varepsilon \quad (7)$$

Para obtener todos los valores que aparecen en la fórmula anterior, se han seguido varios pasos. En primer lugar, a través de la biblioteca `yfinance` se obtuvo la cotización de índice S&P500 (`^GSPC`), y se calcularon los retornos diarios de este índice, esto hace referencia al *market return* de la fórmula. En segundo lugar, se obtuvieron los datos del *risk free rate*, utilizando la cotización de los *treasury bills* de Estados Unidos. Por último, una vez se han conseguido estos valores, solo faltaría calcular las betas para cada compañía en cada fecha, para ello, se crea una función en Python (`calc_betas`), en la cual, introduciendo el *ticker* de la compañía y la fecha para la que se quiere calcular la beta, y un *dataframe* que incluye el retorno del mercado y el de todas las compañías, calcula la beta dividiendo: la covarianza del retorno del mercado y de la compañía, entre la varianza

del retorno del mercado. Tras realizar esto para todas las compañías y todas las fechas, se obtiene un nuevo *dataframe* que contiene el retorno esperado. Este proceso se puede observar en el **Anexo V**.

$$Beta = \frac{Covariance (market, stock)}{Variance (market)} \quad (8)$$

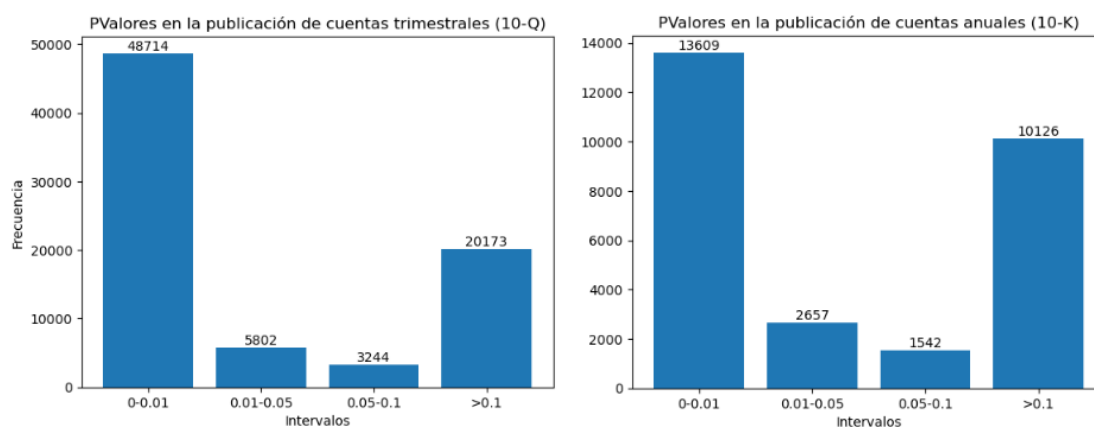
Una vez se han conseguido los retornos esperados, ya se puede calcular el retorno anormal para cada compañía en cada fecha. Y con el retorno anormal ya se puede analizar correctamente el efecto de la publicación de cuentas trimestrales de empresas comparables sobre los retornos de una empresa. Se van a llevar a cabo diferentes análisis que para no solo observar si afecta o no sino también cómo afecta la publicación de resultados.

4.3.1. Comparación entre fechas de publicación trimestrales y anuales

Al igual que para el análisis con el retorno bruto, se realizará una comparación entre el efecto de las fechas de publicación trimestrales y las fechas de publicación anuales para el retorno anormal.

Si se tienen en cuenta sólo las fechas en las que se publican las cuentas trimestrales, en el 74% de los eventos, ocurre un cambio estadísticamente relevante, además, para la mayor parte de este grupo se puede asegurar con mayor certeza ya que el nivel de significancia es muy alto (pvalor < 0,01). Por otro lado, en el caso de la publicación de cuentas anuales, este porcentaje se reduce al 64% de los eventos.

Ilustración 12: Comparación del nivel de significación de la publicación de cuentas trimestrales vs anuales



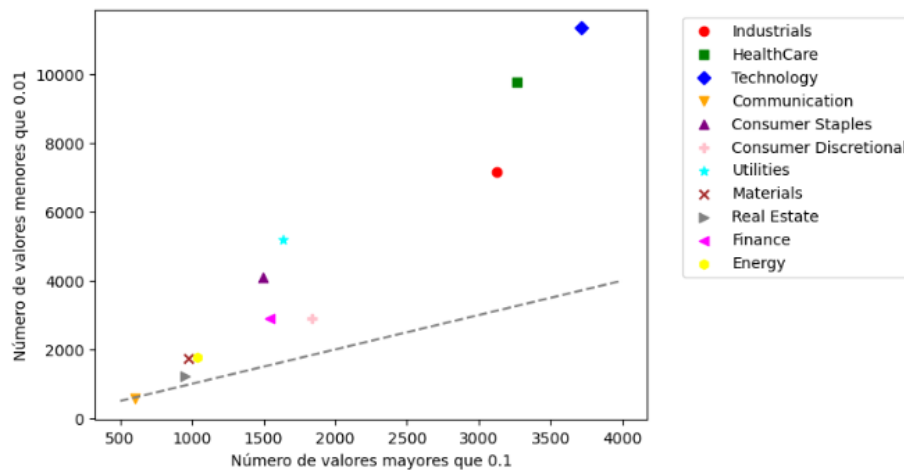
Pese a que se podría pensar que el impacto que tiene la publicación de resultados anuales de los *peers* en la cotización de las acciones debería ser estadísticamente más significativo que la publicación de cuentas trimestrales, ya que las partidas son más grandes y además se publican menos veces, queda demostrado que esto no es así. Esto podría deberse al hecho de que el mercado es eficiente por lo que la “sorpresa” que se lleva el mercado con la publicación de resultados anuales de los *peers* es menor que la que se lleva con la publicación de resultados trimestrales, ya que las cuentas anuales son el agregado de todas las cuentas trimestrales, que ya conoce el mercado. Por tanto, es más fácil proyectar unas cuentas anuales que unas trimestrales, y el mercado se ajusta a estas proyecciones antes de que ocurra el propio evento.

Esta es una hipótesis no corroborada, que intenta explicar los resultados obtenidos de por qué no siempre cambia lo suficiente la cotización de una empresa tras la emisión de cuentas anuales de sus *peers*, y si lo hace tras la publicación de cuentas trimestrales para que estos eventos sean estadísticamente más relevantes.

4.3.2. Comparación entre industrias

Para realizar esta comparación en primer lugar se crean diferentes *dataframes* por sector, en total once, que tiene el mismo formato que *df_peers*, pero solo contiene las subindustrias que pertenecen a cada sector, por lo que el *ttest* sólo se realizará para las empresas pertenecientes a estas subindustrias. En vez de realizar once histogramas como los anteriores uno para cada industria, se ha decidido realizar un *scatter plot*, dónde se puede comparar todas las industrias a la vez. El eje x, mide la cantidad de eventos que tienen un pvalor mayor que 0,1, es decir, aquellos que indican poca significatividad estadística. El eje y, mide la cantidad de eventos muy significativos, es decir, por debajo de 0,01. Se ha querido observar esta relación para todas las industrias. La línea gris marca la función $x = y$, es decir, la relación 50/50 entre los dos grupos que se miden.

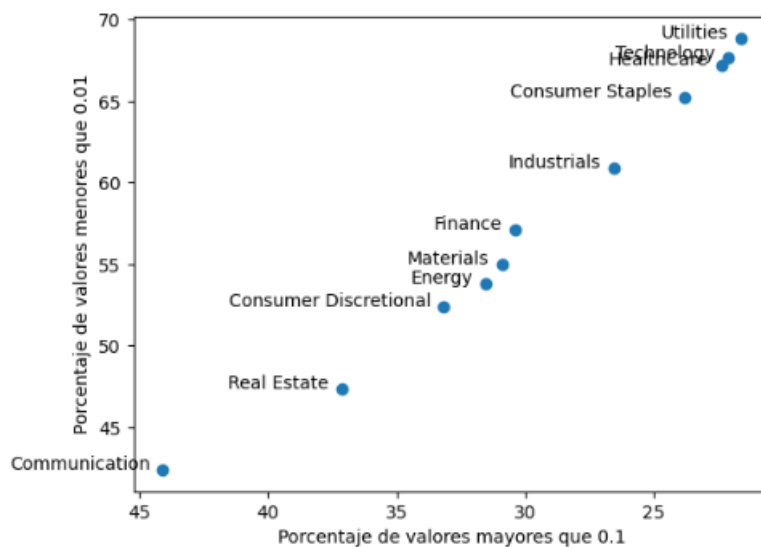
Ilustración 13: Comparación de pvalores para cada industria



Fuente: Elaboración propia (Python)

Cada industria tiene un número diferente de observaciones, por ejemplo, el sector de la comunicación tiene menos de 1.400 observaciones, sin embargo, el tecnológico supera las 15.000 observaciones. Por ello, también se ha realizado un scatter plot porcentual, donde las empresas que se encuentran arriba son las que más porcentaje de observaciones significativas tienen ($<0,01$) y las que se encuentran más a la derecha, las que menos porcentaje de observaciones poco significativas tienen ($>0,1$).

Ilustración 14: Comparación porcentual de grupo de pvalores por industria



Fuente: Elaboración propia (Python)

Si se observan ambos análisis, se puede decir, que el sector donde la publicación de cuentas trimestrales de *peers* afecta más a la cotización de las empresas es el de *Utilities*, y en el escenario contrario se encuentra *Communication Services*. Esto se puede deber a

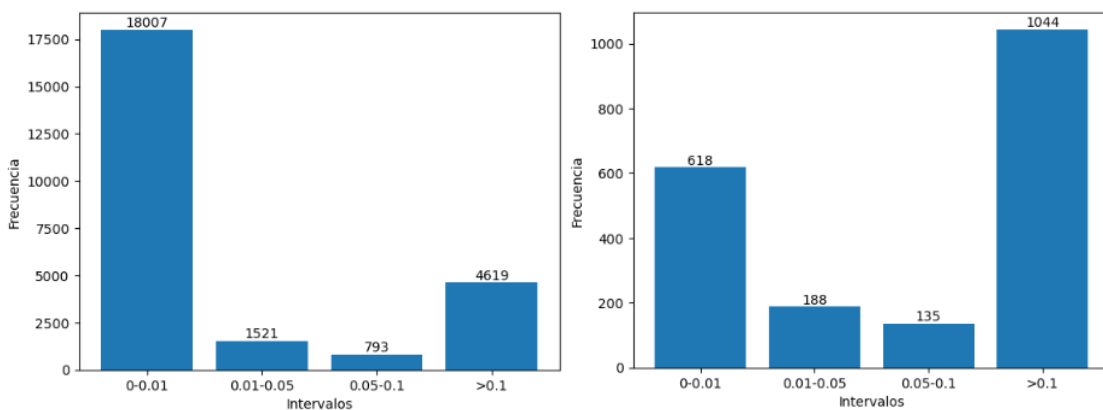
que, el sector de la comunicación, está compuesto por empresas que ofrecen servicios muy heterogéneos, y que pueden tener dinámicas muy diferentes, y, por tanto, los inversores de este sector, le den más importancia a otros factores más sociales o geopolíticos que a los puramente económicos.

Por otro lado, el hecho de que en el sector *Utilities*, sólo existan tres subsectores (es el sector con menos subindustrias), puede significar que todo esté muy concentrado y que las compañías se vean afectadas por los mismos factores, y, por tanto, la significatividad de la publicación de cuentas de *peers* sea más alta, que en los demás sectores.

4.3.3. Comparación de subsectores según la cantidad de peers

Para realizar esta comparación, se ha decidido escoger aquellos subsectores de *df_peers* que únicamente tengan catorce o más sociedades, o tengan dos sociedades, y comparar ambos grupos. Esto se realiza porque se piensa que el hecho de tener un solo *peer* con el que comparar, hará que la publicación de cuentas de ese *peer* provoque cambios mayores en la cotización de la otra empresa. En total, el grupo de subsectores con dos sociedades, está formado por 50 empresas, y el de catorce empresas o más, está formado por 61 empresas.

Ilustración 15: Comparación entre subsectores con muchos(izq.) y pocos(dcha.) peers



Fuente: Elaboración propia (Python)

Como se puede observar, la realidad es completamente distinta a lo que se creía que iba a ser en un inicio. No existe una relación entre el número de empresas que tenga un subsector y la significatividad de la publicación de cuentas anuales, es más, en el caso de que existiera la relación sería al revés, cuantas más empresas existan, más significatividad tendría la publicación de resultados de los *peers*. Cabe destacar que para los datos de las

empresas grandes pese a que se ha tenido en cuenta más empresas que para el otro grupo, solo son cuatro subsectores los que se están teniendo en cuenta. Por tanto, no existe mucha variedad de subsectores y puede ser que en dichos subsectores la significatividad sea muy alta por otros factores que no sea únicamente el número de empresas que los forman.

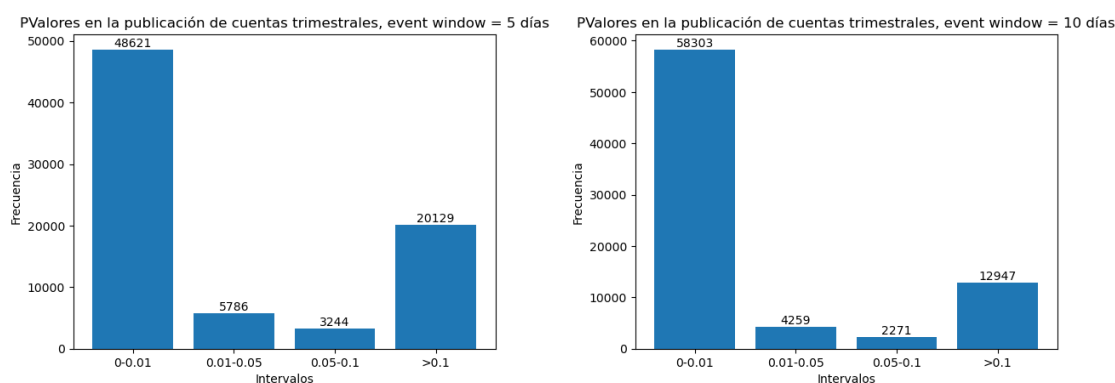
4.3.4. Comparación entre la significatividad con 5 días vs con 10 días

Como se ha comentado anteriormente, existen ventajas e inconvenientes para las diferentes longitudes del *event window*. La elección de un periodo largo para este parámetro puede hacer que se tergiversen los resultados ya que habrá más posibilidades de que existan eventos que afecten a los retornos de una compañía. Sin embargo, un periodo corto puede significar que al mercado no le dé tiempo a ajustarse y por tanto no se observen cambios tras el evento.

Todos los análisis anteriores se han llevado a cabo usando diez días como parámetro para el *event window*, cinco días antes y cinco días después. Pero este parámetro se puede cambiar según cómo sea el mercado y lo que se pretenda analizar. Se ha hecho uso del *event window* de diez días para todos los análisis dado que algunas subindustrias están formadas por más de diez empresas, y estas empresas, publican cuentas trimestrales en fechas parecidas, por tanto, había riesgo de que los eventos se superpusieran unos a otros. Sin embargo, el profesor de Wharton Xuan Liu, recomienda que este periodo sea de diez días, por ello, también se quiere llevar a cabo un análisis del que se pueda extraer alguna conclusión de cuál de las dos opciones se ajusta más a este índice.

Para esto únicamente hay que cambiar en la función `.rolling(x, min_periods=x)`, la variable `x` por el número de días que se quieran analizar antes y después del evento, en este caso cinco y diez días. Se realizan dos histogramas parecidos a los que se han venido haciendo, en los que se pueden observar la cantidad de eventos que tienen una significatividad en el cambio.

Ilustración 16: Comparación entre significatividad según la duración del event window

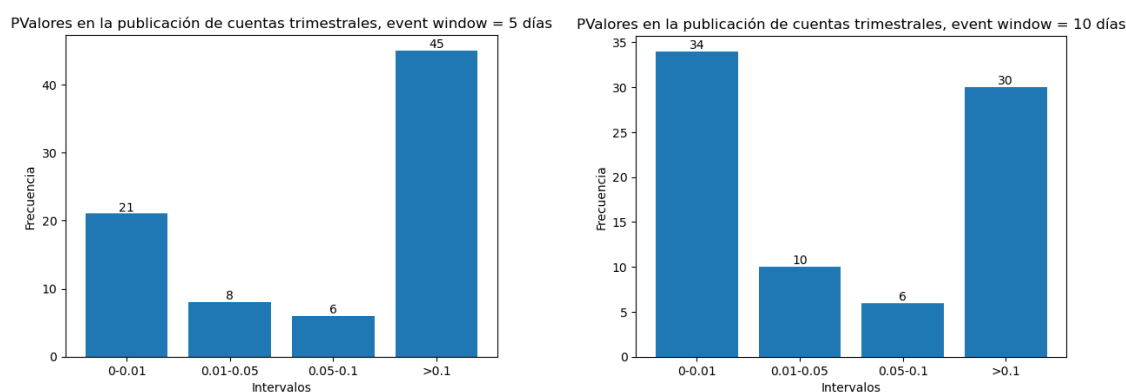


Fuente: Elaboración propia (Python)

Como se puede observar, existe una mayor cantidad de eventos en los que la diferencia entre la media de los diez días antes y de los diez después es significativa. Esto se puede deber a dos razones, la primera es que suponiendo que no ocurre ningún otro evento en esos veinte días de ventana, el mercado tarda en ajustarse. La otra razón puede ser, que ocurran otras publicaciones y por tanto la variación de las medias sea mayor.

Para evitar esto se va a realizar un último análisis en el que se va a comparar para una sola subindustria este cambio, y esta subindustria solo tendrá dos *peers*, para que no pueda ocurrir una variación de los retornos por la publicación de resultados de otros *peers*. Se ha elegido una al azar, la subindustria de cuidados personales, esta subindustria la forman Procter & Gamble y Estee Lauder, y se ha realizado el mismo análisis anterior pero solo para esta subindustria.

Ilustración 17: Comparación entre significatividad según la duración del event window (Personal Care)



Fuente: Elaboración propia (Python)

Como se puede observar, en la industria de cuidado personal, se corrobora el análisis anterior llevado a cabo para todas las industrias. Una venta de evento mayor

permite que el mercado se ajuste durante un periodo más largo de tiempo, y, por tanto, la significatividad de las diferencias entre los días anteriores y posteriores aumenta.

CAPÍTULO V. OTRAS LÍNEAS DE INVESTIGACIÓN, LIMITACIONES Y PROBLEMAS

Se han realizado diferentes análisis de interés para los retornos anormales de las compañías. Además, se abre un abanico de posibilidades de análisis que se pueden realizar, más y menos complejos, y poder mostrar diferentes conclusiones según sea el objeto de estudio.

Sin embargo, hay diferentes análisis que pueden ser de interés y que mostrarían aún más a fondo cómo funciona el mercado y más concretamente el S&P500, en relación a la publicación de resultados de *peers* y del estudio de estos eventos sobre los retornos de una empresa. Algunos estudios pueden ser:

- **Comparación entre cuartiles según diferentes ratios:** un análisis que sería interesante realizar, sería conocer cómo afecta la publicación de cuentas trimestrales o anuales de *peers* a las empresas que tengan un *Price to earnings ratio* dentro del primer cuartil, o para aquellas del último cuartil, y de igual forma con los distintos ratios presentados para la valoración relativa de las empresas. Esto puede mostrar si a aquellas empresas de una subindustria que tienen mejores ratios que las demás les afecta menos la publicación de cuentas o igual que al resto de empresas de su subsector, y por tanto observar si afecta más a los líderes del mercado o a las empresas que no lo son.
- **Comparación por dimensión:** otro análisis que sería de interés y en concordancia al propuesto anteriormente, sería observar si existe una diferencia notable en cómo afecta la publicación de cuentas de empresas comparables a aquellas empresas más o menos grandes, sabiendo que dentro del índice S&P500 todas las empresas tienen un tamaño considerable o si por el contrario no es una característica que influya.
- **Directa o inversamente relacionadas:** el último análisis que se propone, sería poder conocer si, para aquellas empresas en las que sea significativa la diferencia entre la media de los retornos de los días anteriores y la de los días posteriores al evento, si están directamente o inversamente relacionados. Es decir, en el caso de que los estados financieros de los *peers* hayan ido a mejor con respecto al año o al trimestre anterior, si esto se traduce en un cambio a positivo o a negativo de los retornos de la empresa.

Aunque las posibilidades son infinitas, y depende de aquello que se quiera estudiar, estas son algunas de las futuras líneas de investigación que pueden ser de interés y que no se han abarcado en este trabajo. Estas líneas de investigación, se basan en el estudio individual, empresa a empresa, de cómo afecta la publicación de cuentas de *peers* a estas. Ya que en este trabajo se ha realizado un análisis con un punto de vista más generalista, queriendo sacar conclusiones de cómo afecta la publicación de resultados en el índice S&P500 y en las industrias que lo forman, y no individualista, sin ir caso por caso y analizando cómo afecta a cada empresa la publicación de resultados de empresas comparables.

Este tipo de análisis individual no se puede realizar con la base de datos que se ha usado, sino que se tendría que utilizar una muestra del índice y extrapolarlo al índice entero. Esto se comenta ya que cuando se realizan operaciones empresa a empresa, la necesidad computacional necesaria es alta cuando se operan las 500 empresas del índice en los 15 años estudiados. Por ejemplo, en el cálculo de las betas de cada empresa para cada fecha, las operaciones realizadas superaron el millón y medio y las horas invertidas en este cálculo fueron aproximadamente 12h. Por tanto, si el estudio completo se basara en operaciones individuales, se debería utilizar una muestra más pequeña, tanto temporal como de número de empresas, por ejemplo, durante cinco años y para una subindustria por sector.

Por otro lado, el no haber realizado el estudio empresa a empresa, no permite que las conclusiones sacadas se puedan corroborar para todos los casos en el mercado. Cada sector y subsector funciona de una forma muy diferente, y yendo más allá, cada empresa dentro de estos subsectores, también funciona de una manera diferente, por tanto, las conclusiones sacadas no son para todos los subsectores ni para todas las compañías. Sirve para conocer mejor cómo funciona el índice en términos generales y las tendencias generales del mercado, y si se quisiera sacar conclusiones más concretas, habría que fijarse en los subsectores o empresas de interés. Pero este no era el objeto de estudio del trabajo.

CAPÍTULO VI. CONCLUSIONES

Durante este trabajo se ha realizado el estudio de la valoración de empresas y cómo los inversores se fijan en diferentes aspectos intrínsecos de la compañía y utilizan ratios para valorar y comparar una compañía con sus *peers*. Se ha podido conocer que para los diferentes integrantes del mercado las empresas comparables son casi tan importantes como la propia empresa, para valorar ésta en el mercado.

Una vez se ha conocido cómo se relacionan unas empresas con otras de su misma subindustria para los inversores, se ha podido comenzar con la limpieza y tratamiento de los datos que se utilizarían para el análisis. Tras haber terminado con esta parte, que ha sido la que más tiempo ha llevado del trabajo, se han realizado diferentes análisis.

En primer lugar, se ha estudiado si la publicación de cuentas anuales de empresas comparables afecta a la cotización de las empresas. Para esto se ha realizado el estudio de eventos, que primero se realizó de una forma menos elaborada (*naive*) haciendo uso del retorno bruto de las compañías en vez del retorno anormal. Esto ha servido para sacar unas primeras conclusiones del S&P500, para la mayor parte del conjunto de todas las medias de los retornos anteriores y posteriores al evento, no había una diferencia significativa. Sin embargo, cuando se tiene en cuenta la desviación típica y se realiza un estudio fecha a fecha de si existe significatividad, se llega a la conclusión que en la mayor parte de los eventos sí existe una diferencia significativa entre el retorno anterior y posterior al evento. Además, se compara entre los eventos que marcan las publicaciones anuales y las trimestrales, y se llega a una conclusión contraria a la hipótesis del comienzo, las cuentas anuales no suponen una diferencia en los retornos más significativa que las cuentas trimestrales.

En segundo lugar, teniendo en cuenta el retorno anormal para el que anteriormente se ha calculado el retorno esperado de cada empresa en cada fecha, se ha realizado la anterior comparación entre cuentas trimestrales y anuales dando un resultado similar que para los retornos brutos.

En tercer lugar, se han comparado las diferentes industrias que forman el S&P500, observando cómo se relacionan las empresas dentro de estas y si existen divergencias entre industrias. En este caso se ha llegado a la conclusión en que cada industria las empresas tienen dinámicas diferentes, y que, por tanto, se relacionan de diferente forma con sus comparables, afectando en algunas industrias poco la publicación de resultados

trimestrales de los *peers*, como es el caso de la industria de la comunicación. En el otro extremo se encuentran los sectores como la tecnológica o *utilities*, en los que la publicación de resultado afecta significativamente a los retornos de las compañías. En consonancia con las industrias, no solo se ha querido analizar las industrias en términos cualitativos (sector), sino también en términos cuantitativos y observar si en aquellas subindustrias donde solo hay dos empresas comparables, si la publicación de resultados en este tipo de industrias afecta más significativamente por el hecho de solo tener una empresa con la que comparar. La hipótesis inicial era que sí, que debería de afectar más a estas empresas, sin embargo, se ha podido observar lo contrario, que en las empresas donde existen muchos *peers* es en aquellas donde la significatividad de la emisión de resultados es mayor.

Por último, se ha querido analizar si la ventana temporal estudiada significaba una diferencia en los retornos anteriores y posteriores. Se puede comprobar que efectivamente el hecho de elegir una ventana concreta hace que los resultados varíen, sin embargo, esta diferencia no es excesivamente notable como para que los resultados muestren conclusiones completamente opuestas, sólo muestra que el mercado tarda más de lo esperado en ajustarse a los eventos.

En definitiva, se ha realizado un estudio superficial del índice, estudiando las industrias, y sacando conclusiones que han permitido conocer el S&P500, sus sectores y subindustrias.

ANEXOS

Anexo I. Bibliotecas Utilizadas

```
###BIBLIOTECAS UTILIZADAS
import yfinance as yf
import pandas as pd
import requests
from bs4 import BeautifulSoup
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import pandas_datareader as pdr
import datetime as dt
from sklearn.linear_model import LinearRegression
from datetime import timedelta
```

Anexo II. Pasos relativos al WebScrapping de Wikipedia

```
# ANEXO II, WEB SCRAPPING WIKIPEDIA OBTENCIÓN DE NÚMERO CIK, SUBINDUSTRIAS Y OTROS DATOS

table=pd.read_html('https://en.wikipedia.org/wiki/List_of_S%26P_500_companies')
df = table[0]
# Obtener los valores únicos de la primera columna de df_ebitda
valores_ebitda = df_ebitda[df_ebitda.columns[0]].unique()
df = df[df[df.columns[0]].isin(valores_ebitda)] # Filtrar el DataFrame df por valores de la primera columna en df_ebitda

df.to_csv('S&P500-Info.csv', index=False)
#df.to_csv('S&P-NeedeInfo.csv', columns=['Symbol', 'GICS Sector', 'GICS Sub-Industry', 'CIK'], index=False)
#df.to_csv("S&P500-Symbols.csv", columns=['Symbol'], index=False)

col_list = ["Symbol"]
cik_tik= pd.read_csv('S&P500-Info.csv', usecols=['Symbol', 'CIK'])
```

Anexo III. Pasos relativos a la obtención del dataframe que contiene los grupos de peers (df_peers)

```
#AGRUPACIÓN DE SUBINDUSTRIAS A PARTIR DEL WEB SCRAPPING ANTERIOR Y OTROS RETOQUES
columnas_deseadas = ['Symbol', 'GICS Sub-Industry'] # Lista de nombres de las columnas deseadas
df1 = df[columnas_deseadas]
df_agrupado = df1.groupby('GICS Sub-Industry')['Symbol'].apply(list).reset_index()

df_filtrado = df_agrupado[df_agrupado['Symbol'].apply(len) > 1].reset_index(drop=True)
#df_filtrado.to_csv('S&P500-SIGroups-Filtrado.csv', index=False)
cik_tik_filtered = cik_tik[cik_tik['symbol'].isin(df_filtrado['Symbol']).explode()]

num_industrias_eliminadas = sum(df_agrupado['Symbol'].apply(len) == 1)#Cuantas empresas se eliminan

#REORGANIZACIÓN DE LOS TICKERS Y SELECCIÓN DE LA EMPRESA DE REFERENCIA POR INDUSTRIA
df_new = df_filtrado.copy() # Copiar el DataFrame original
df_new['First Symbol'] = df_filtrado['Symbol'].apply(lambda x: x[0])
df_new['Symbol'] = df_filtrado['Symbol'].apply(lambda x: x[1:])
df_new = df_new.reindex(columns=['GICS Sub-Industry', 'First Symbol', 'Symbol'])
#df_new.to_csv('S&P500-SIGroups-Filtrado-Actualizado.csv', index=False)
```



```

#EN LA PRIMERA COLUMNA LA SUBINDUSTRIA, EN LA SEGUNDA LA EMPRESA DE REFERENCIA Y LAS SUCEASIVAS COLUMNAS LOS PEERS
df_peers = df_new[['GICS Sub-Industry', 'First Symbol']].copy() # Crear el DataFrame df_peers con las columnas deseadas
max_length = df_new['Symbol'].apply(len).max() # Obtener la longitud máxima de las listas en la tercera columna
column_names = [f'Peer{i}' for i in range(2, max_length + 2)] # Generar nombres de columna para los peers (Peer1, Peer2, Peer3, ...)
df_peers_split = pd.DataFrame(df_new['Symbol'].tolist(), columns=column_names) # Dividir las listas de la tercera columna en columnas
df_peers = pd.concat([df_peers, df_peers_split], axis=1) # Combinar los DataFrames df_peers y df_peers_split
df_peers.rename(columns={'First Symbol': 'Peer1'}, inplace=True)
df_peers.rename(columns={'GICS Sub-Industry': 'Industry'}, inplace=True)
df_peers=df_peers.set_index('Industry', drop=True)

```

Anexo IV. WebScraping EDGAR

```

#(WEB SCRAPING, EDGAR, OBTENER FECHAS DE PUBLICACIÓN)
headers = {
    'User-Agent': 'nacho.ortizsan@gmail.com',
    # Otros encabezados si son necesarios
}

def get_soup(url, headers):
    page = requests.get(url, headers=headers)
    soup = BeautifulSoup(page.text, "lxml")
    return soup

headings = ['Filings', 'Format', 'Description', 'Filing Date', 'File/Film Number', 'ticker', 'cik']

def get_soup(url, headers):
    page = requests.get(url, headers=headers)
    soup = BeautifulSoup(page.text, "lxml")
    return soup

def get_web_df(cik_tik_filtered, headers, headings):
    all_rows = []

    for _, row in cik_tik_filtered.iterrows():
        ticker = row['Symbol']
        cik = row['CIK']
        url = f'https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK={cik}&type=10-Q&owner=exclude&count=80&search_text='
        soup = get_soup(url, headers)
        table = soup.find("table", class_="tableFile2", summary='Results')

        if table is None:
            continue

        rows = table.find_all("tr")
        for row in rows[1:]:
            cells = row.find_all("td")
            filings=cells[0].text.strip()
            filing_date = cells[3].text.strip()
            file_number = cells[4].text.strip()
            all_rows.append([filings, filing_date, file_number, ticker, cik])

    df_dates = pd.DataFrame(data=all_rows, columns=['Filings', 'Filing Date', 'File/Film Number', 'Ticker', 'CIK'])
    df_dates['Filing Date'] = pd.to_datetime(df_dates['Filing Date'], format='%Y-%m-%d')
    return df_dates

```

Anexo V. Creación del dataframe de retorno esperado y retorno anormal

```
# ABNORMAL RETURN: Obtener La cotización del índice S&P 500
sp500 = yf.Ticker("^GSPC")

# Obtener el historial de precios del índice S&P 500 desde el 8 de mayo de 2008 hasta el 8 de mayo de 2023
historical_data = sp500.history(start="2007-12-31", end="2023-01-01")
df_sp = historical_data['Close']
df_sp.index = df_sp.index.strftime('%Y-%m-%d %H:%M:%S')
df_sp = df_sp.to_frame()
df_sp.index=pd.to_datetime(df_sp.index)
df_sp.rename(columns={'Close': 'spindex'}, inplace=True)

df_sp = df_sp.merge(df_mc_tp, left_index=True, right_index=True)

rsp=(df_sp/(df_sp.shift(periods=1)))-1

def calc_betas(a,rsp,t_date,n=15,window=120):
    end=t_date-dt.timedelta(days=n)
    start=end-dt.timedelta(days=window)
    rspd=rsp.loc[start:end]
    var = rspd.var()
    cov = rspd.cov()
    beta=cov.loc[a, 'spindex']/var.loc['spindex']

    return beta
```

```
###CREACIÓN DEL DATAFRAME ue contiene Los risk free rates

df_rf1=pd.read_csv('long-term-rates-2000-2021.csv')
df_rf1[df_rf1.columns[0]] = pd.to_datetime(df_rf1[df_rf1.columns[0]])
df_rf1 = df_rf1.set_index(df_rf1.columns[0])
df_rf1 = df_rf1.iloc[:, :1]
df_rf1 = df_rf1.rename(columns={df_rf1.columns[0]: 'Rate'})
df_rf1 = df_rf1[df_rf1.index >= '2009-01-01']
df_rf1 = df_rf1[:-1]

df_rf2=pd.read_csv('daily-treasury-rates.csv')
df_rf2[df_rf2.columns[0]] = pd.to_datetime(df_rf2[df_rf2.columns[0]])
df_rf2 = df_rf2.set_index(df_rf2.columns[0])
df_rf2 = df_rf2.iloc[:, :1]
df_rf2 = df_rf2.rename(columns={df_rf2.columns[0]: 'Rate'})
df_rf2 = df_rf2[:-1]

df_rf = pd.concat([df_rf1, df_rf2])

df_rfsp = pd.DataFrame(rsp['spindex']).merge(df_rf, left_index=True, right_index=True)
df_capm = pd.DataFrame(columns=r.columns)
```

```
###CALCULAR EL EXPECTED RETURN
counter = 0
# Incrementar el contador en cada iteración del bucle

for a in df_capm.columns:
    ticker = a
    counter += 1 # Incrementar el contador en cada iteración del bucle
    print(counter,a)
    for index, row in df_rfsp.iterrows():

        # Obtener Los valores necesarios del DataFrame df_rf_sp
        market_return = df_rfsp.iloc[:, 0] # Primera columna: Market Return
        risk_free_rate = df_rfsp.iloc[:, 1] # Segunda columna: Risk-Free Rate

        # Definir La fecha y el ticker para Los cálculos
        fecha = index

        # Calcular La beta utilizando La función calc_betas
        beta = calc_betas(ticker, rsp, fecha)

        # Obtener el Risk-Free Rate y el Market Return para La fecha específica
        risk_free_rate = risk_free_rate.loc[fecha]
        market_return = market_return.loc[fecha]

        # Calcular el Expected Return según el CAPM
        expected_return = risk_free_rate + beta * (market_return - risk_free_rate)
        df_capm.loc[fecha,ticker]=expected_return
```

```

## ANEXO XV Calculo de Abnormal Return a través de Expected return creado anteriormente
df_capm=pd.read_csv('Expected Return.csv')
df_capm=df_capm.set_index(df_capm.iloc[:,0], )
df_capm.index.name='Date'
df_capm= df_capm.iloc[:, 1:]
df_capm.index=pd.to_datetime(df_capm.index)

#Calcular ABNORMAL Return para Las fechas que coincidan
indices_comunes = df_capm.index.intersection(r.index)
df_capm = df_capm.loc[indices_comunes]
r = r.loc[indices_comunes]

##ABNORMAL Return (df_ar)
df_ar=df_capm - r

```

Anexo VI. Código tipo para la ejecución del ttest

```

t_p_values = {}

for index, row in df_peers.iterrows():
    # Seleccionar Los valores de la fila en el DataFrame r
    fila_h = df_peers.loc[index]
    rh = r.T
    rh = rh[rh.index.isin(fila_h)]
    rh = rh.T

    # Seleccionar Las fechas para healthcare
    hdate = df_date.T
    hdate = hdate[hdate.index.isin(fila_h)]
    hdate = hdate.T

    rhpre=rh.rolling(5,min_periods=5).mean()
    rhpre_std = rh.rolling(window=45, min_periods=45).std()
    rh=rh.iloc[:-1]
    rhpos=rh.shift(1).rolling(5,min_periods=5).mean()
    rhpos_std = rh.shift(1).rolling(window=45, min_periods=45).std()
    rh=rh.iloc[:-1]
    rhpos=rhpos.iloc[:-1]
    rhpos_std=rhpos_std.iloc[:-1]

    for i in df_peers.loc[index]:####HACERLO PARA TODOS LOS TICKERS
        if pd.notna(i):
            for col in hdate.columns:
                hdatea = pd.DataFrame(columns=hdate.columns)
                # Obtener Las columnas sin incluir La columna actual
                other_cols = [c for c in hdate.columns if c != col]

                # Obtener Los valores de Las columnas restantes en el orden deseado
                values = hdate[other_cols].values.flatten()

                # Asignar Los valores a La columna correspondiente en hdatep
                hdatea[i] = values
                first_row = hdatea.iloc[0]

                columns_to_drop = first_row[first_row.isna()].index # Obtener Las columnas con valores NaN en la primera fila
                hdatea = hdatea.drop(columns_to_drop, axis=1) # Eliminar Las columnas de hdatep con valores NaN en la primera fila
                hdatea[i] = hdatea[i].fillna(method='ffill') # Rellenar Los NaN en La columna 'Columna1' con el valor anterior
                hdatea = hdatea.drop_duplicates(subset=[i]).dropna(subset=[i]).reset_index(drop=True) # Eliminar Los valores duplica

                ra=rh[i]
                rapre=rhpre[i]
                rapos=rhpos[i]
                rapre_std=rhpre_std[i]
                rapos_std=rhpos_std[i]
                raf = ra.loc[ra.index.isin(hdatea[i])]
                rapref=rapre.loc[rapre.index.isin(hdatea[i])]
                raposf=rapos.loc[rapos.index.isin(hdatea[i])]
                rapref_std=rapre_std.loc[rapre_std.index.isin(hdatea[i])]
                rapref_std=rapref_std.dropna()
                raposf_std=rapos_std.loc[rapos_std.index.isin(hdatea[i])]
                raposf_std=raposf_std.dropna()
                indices_comunes = rapref_std.index
                # Filtra ambas series utilizando Los indices comunes
                raposf_std = raposf_std[indices_comunes]
                rapref=rapref[indices_comunes]
                raposf=raposf[indices_comunes]

                pref= rapref
                pref=pref.dropna()
                prelen=len(pref)
                posf=raposf
                posf=posf.dropna()
                poslen=len(posf)

                t_statistic, p_value=stats.ttest_ind_from_stats(pref,rapref_std,prelen,posf,raposf_std,poslen)
                t_p_values[i]=p_value
print(t_p_values)

```

BIBLIOGRAFÍA

- Bragg, S. (2023). Price to book ratio definition *AccountingTools*.
- Brown, S. J., & Warner, J. B. (1980). Measuring security price performance. *Journal of Financial Economics*, 8(3), 205-258.
- Carroll, S. (2021, diciembre). IAS 36 - Value in use estimating future cash inflows and outflows. *Grant Thornton International Ltd. Home*.
- Cornell Law School (2020) Net Earnings.
- Corporate Finance Institute. (2022). How to Choose Comparable Companies. *Corporate Finance Institute*.
- Damodaran, A. (2011). *The little book of valuation: How to value a company, pick a stock and profit*. John Wiley & Sons.
- Fama, E. (1970). Efficient Capital Markets: a review of Theory and Empirical work.
- Fernández, P. (2007). Company valuation methods. the most common errors in valuations. *IESE Business School*, 449, 1-27.
- Ghaeli, M. R. (2016). Price-to-earnings ratio: A state-of-art review. *Accounting*, 3(2), 131-136.
- García, D. & Norli, O. (2012, marzo) Crawling EDGAR.
- Gordon, M. J. (1959). Dividends, earnings, and stock prices.
- Hayes, A. (2022). Enterprise Multiple (EV/EBITDA): Definition, Formula, Examples. *Investopedia*.

- Kothari, S.P. and Warner, J.B., (2006). The Econometrics of Event Studies. *Handbook of Corporate Finance: Empirical Corporate Finance*
- Liu, X. (2020, marzo) Introduction to Event Study by WRDS. *The Wharton School, University of Pennsylvania.*
- MacKinlay, C. (1997, marzo). Event studies in Economic and Finance. *Journal of Economic Literature.*
- Miller, M. H., & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *Journal of Business*, 34(4), 411-433.
- Cornell Law School (2020) Net Earnings.
- Sarath. (2023, 28 marzo). *Price to Earnings (P/E) Ratios by Industry 2023 | Eqvista.*
- Stern School of Business, New York University. “*Dividend Discount Models*”
- Tuovila, A. (2020). Relative Valuation Model: Definition, Steps, and Types of Models. *Investopedia.*
- Trading Economics. (s. f.). *United States Fed Funds Rate - 2023 Data - 1971-2022 Historical - 2024 Forecast.*
- William, J.B. (1938). The Theory of Investment Value.

REFERENCIAS CÓDIGO

Yfinance. (s/f). PyPI.

Smigel, L. (2022, enero 11). Yfinance python tutorial (2023). Analyzing Alpha.
<https://analyzingalpha.com/yfinance-python>

Importing historical S&P 500 data in python. (s/f). Stack Overflow

Python - yahoo finance download all s&p 500 stocks. (s/f). Stack Overflow.