



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**La influencia de la publicación de resultados
en la cotización de empresas comparables**

Autor: Álvaro de Artiñano Gutiérrez

5º E-3 Analytics

Tutor: Carlos Bellón Núñez-Mera

Madrid

Junio 2022

Resumen

Este trabajo se centra en el estudio del efecto en la cotización de las empresas de la publicación de resultados de empresas comparables. El foco concreto se halla en analizar cómo la actualización de la información disponible por el mercado se puede reflejar en el corto plazo, y si el mercado valora de manera significativa esta información o si por el contrario pasa desapercibida y queda por lo tanto relegada por eventos que los participantes del mercado consideran de mayor importancia. Para llevar a cabo este análisis, el estudio se apalanca en la utilización de herramientas de Data Analytics, que permitirán la obtención de resultados más fiables y que aporten una información más reveladora.

En este contexto, se erigirá como un área de estudio de gran importancia el análisis de los distintos métodos de valoración existentes, que permitirán entender desde un punto de vista teórico cómo habría de afectar a la valoración de las empresas comparables la publicación de resultados. Igualmente se indagará en la relación entre el valor y el precio de las empresas, al igual que se estudian los distintos engranajes que moverán ambos elementos.

En definitiva, se realiza un análisis teórico y empírico de la materia para lograr entender de una manera holística la cuestión planteada

Palabras Clave: Análisis de datos, métodos de valoración, eficiencia del mercado, anuncio de resultados, algoritmos de clasificación

Abstract

This paper analyzes the effect on a company's stock price of the earnings announcements of comparable companies. The main area of study is to empirically examine how the market reacts to an update of the available information in the short term, and to observe if the market weights this as a relevant event or if it is considered something minor when compared to other pieces of information or broader financial trends. In order to analyze these events, the study will leverage itself on Data Analytics tools that will enable the obtention of more insightful and reliable results.

In this context, it will be of utter importance to explore the different valuation methods, in order to understand from a theoretical position how should earnings announcements affect the valuation of comparable companies. Furthermore, this paper will explore the relationship between price and value and the different levers that affect both.

This paper thus focuses on a theoretical and empirical analysis of the subject in order to holistically understand the matter

Key Words: Data analytics, valuation methods, market efficiency, earnings announcement, classification algorithms

Índice

CAPÍTULO 1. OBJETIVOS DEL ANÁLISIS	7
CAPÍTULO 2. INTRODUCCIÓN A LA VALORACIÓN DE EMPRESAS – MARCO TEÓRICO DEL ESTUDIO	9
2.1 MÉTODOS DE VALORACIÓN INTRÍNSECA	10
2.2 MÉTODOS DE VALORACIÓN RELATIVA	12
2.2.1 <i>Método de los múltiplos comparables</i>	13
2.2.2 <i>Método de las transacciones precedentes comparables</i>	15
2.3 TENDENCIAS MACROECONÓMICAS Y OTRAS CONSIDERACIONES QUE AFECTARÁN AL PRECIO ASIGNADO A UNA EMPRESA	15
2.3.1 <i>Existencia de ciclos macroeconómicos</i>	15
2.3.2 <i>Política monetaria de los Bancos Centrales</i>	16
CAPÍTULO 3. OBTENCIÓN Y TRATAMIENTO DE DATOS	18
3.1 OBTENCIÓN DEL TICKER Y DE LAS FECHAS DE PUBLICACIÓN DE RESULTADOS	20
3.1.1 <i>Central Index Key (CIK)</i>	21
3.1.2 <i>Fechas de publicación de resultados</i>	22
3.1.3 <i>Obtención de las subindustrias y creación de los grupos de empresas comparables.</i>	23
3.1.4 <i>Métrica correspondiente al tamaño de la empresa observada</i>	24
3.1.5 <i>Métricas de ingresos netos predichas y publicadas</i>	25
3.1.6 <i>Métricas analíticas de los resultados publicados en comparación con el trimestre previo</i>	26
3.1.7 <i>Métricas de desempeño de las empresas comparables en el momento de publicación de resultados de la empresa observada – variable objetivo</i>	27
3.2 BASE DE DATOS OBTENIDA TRAS EL PROCESO DE LIMPIEZA Y PROCESAMIENTO DE LAS BASES	29
CAPÍTULO 4. ANÁLISIS DE LOS DATOS Y ESTUDIO DE LAS VARIABLES	30
4.1 ESTUDIO ESTADÍSTICO DE LAS VARIABLES	30
4.2 CREACIÓN, PARAMETRIZACIÓN Y RESULTADOS DEL MODELO DE CLASIFICACIÓN	33
4.2.1 <i>Método del Random Forest</i>	34
4.2.2 <i>Análisis discriminante lineal</i>	35
4.2.3 <i>Potenciación del gradiente</i>	36

CAPÍTULO 5.	FUTURAS LÍNEAS DE INVESTIGACIÓN Y LIMITACIONES DEL ESTUDIO	38
5.1	INTERRELACIÓN DE LAS PRESENTACIONES DE RESULTADOS	38
5.2	EXISTENCIA DE ELEMENTOS DISTORSIONADORES DE LOS RESULTADOS	38
5.3	COMPARABILIDAD DE LAS EMPRESAS	39
5.4	INEXISTENCIA DE INFORMACIÓN DE LAS RAZONES SUBYACENTES AL <i>EARNINGS SURPRISE</i>	40
5.5	TRATAMIENTO DE LAS PREDICCIONES DE LOS ANALISTAS	40
5.6	INCLUSIÓN DE MEDIDAS TENDENCIALES	41
CAPÍTULO 6.	CONCLUSIONES.....	42
CAPÍTULO 7.	ANEXO.....	43
CAPÍTULO 8.	BIBLIOGRAFÍA.....	48

Capítulo 1. Objetivos del análisis

En este trabajo se estudia el efecto en la cotización de las empresas de la publicación de resultados de compañías comparables. Se parte por lo tanto de la premisa de que el mercado valora el rendimiento de las empresas equiparables a una determinada compañía a la hora de establecer su precio de cotización. Esta premisa cobra sentido si se asume que a un inversor no le debe resultar indiferente lo que les ocurre a los competidores de la empresa que está valorando.

La publicación de resultados se presenta por lo tanto como un evento que aporta información nueva y valiosa a los participantes del mercado, en tanto que otorga datos sobre las tendencias que pueden afectar a las empresas comparables, o al mercado en general. Esto será aún más cierto si la empresa en cuestión que publica resultados tiene un efecto transversal a lo largo del mercado, como ocurriría con empresas conglomeradas con participación en diferentes sectores de la economía (como ocurriría por ejemplo con Berkshire Hathaway, que participa directa e indirectamente en numerosas empresas, que representan los cimientos de la economía estadounidense¹). Igualmente, cobraría especial valor la información provista por la publicación de resultados en el caso de empresas con una amplia influencia vertical en una industria o sector, como ocurriría ante la publicación de resultados negativos por parte de Apple, lo cuál extendería su influencia a todo el sector tecnológico, además de a industrias paralelas como la de los semiconductores.

Partiendo de esta base, es importante realizar una delimitación del objeto de estudio. En este caso se ha escogido el análisis del comportamiento de las empresas pertenecientes al SP500 en la actualidad (en este sentido, es importante mencionar que, dado que la composición del SP500 varía año tras año con la incorporación y exclusión de empresas del índice, el número total de empresas estudiadas en este análisis será inferior a los 504 miembros que componen el SP500 en el momento de realización de este trabajo), dado que son aquellas de las que se dispone de una mayor cantidad de información, al tratarse de las empresas más grandes de la economía más fuerte del planeta.

En cuanto a la existencia de análisis previos sobre la materia, se ha estudiado ampliamente el efecto de la presentación de resultados en la cotización de la propia empresa que los publica², mientras que el campo de estudio del efecto en las empresas comparables se mantiene

¹ Berkshire Hathaway es el mayor propietario y operador de infraestructura en Estados Unidos, que tienen en libros un valor de aproximadamente \$158.000.000.000. Es además tenedor de en torno a un 0.5% de la “*publicly-held national debt*”, y aporta casi un 1% de los fondos totales recaudados por la Tesorería Estadounidense en concepto de impuestos sobre los beneficios de las empresas (Buffet, 2021).

² (Fatmasari, Imam, Fuad., Abdul Kharis, & Nurcahyono, 2020), (Dodonov, 2020), (European Central Bank, 2020), (Mohamed, 2010), (Paul & Richard J., 2004) y (Stankevi & Simas, 2014).

relativamente inexplorado y es por ello por lo que se ha decidido focalizar el estudio en esta rama de análisis. En este sentido, se ha estudiado el caso de aquellas empresas con periodicidades escasas en la publicación de sus resultados financieros, habiéndose encontrado que las empresas que publican resultados tan solo dos veces al año son hasta dos veces más sensibles al anuncio de resultados de los líderes de su industria que el resto de las empresas de su sector (Arif, 2016).

Si bien es cierto que la cotización de las empresas se verá intuitivamente más afectada por sucesos internos de la propia empresa como la publicación de sus resultados y la existencia de “sorpresas” en comparación con las predicciones de los analistas, o sucesos externos relevantes, como las tendencias macroeconómicas o la política monetaria (o más recientemente, la aparición de una pandemia global), resulta interesante analizar si los *earnings surprises* extienden sus efectos más allá de la empresa que publica los resultados de manera genérica.

Tendría a priori sentido que, las empresas comparables se vieran afectadas por los anuncios de resultados sorprendentes de las empresas con las que se comparan. Por ejemplo, unos débiles resultados no achacables a la propia empresa y que radican en un cambio en la mentalidad de los consumidores, aportaría valiosa información a los inversores racionales, que actualizarían sus predicciones sobre los resultados de las empresas comparables, y modificarían presumiblemente a la baja las valoraciones de las mismas. Sin embargo, también puede ocurrir lo contrario, un *earnings surprise* achacable únicamente a la empresa observada podría mostrar debilidades en dicha empresa que podrían ser aprovechadas por sus competidores, y por ende se transmitiría una señal positiva a los inversores con respecto a las empresas comparables.

Capítulo 2. Introducción a la valoración de empresas – marco teórico del estudio

Teniendo en cuenta el objetivo de entender los distintos engranajes que mueven la cotización de las empresas, es importante comprender dónde se haya el valor de una empresa (valor que podrá coincidir o no con el precio de la misma), y sobre todo cómo se puede calcular y se calcula dicho valor. Mediante la comprensión de los distintos métodos de valoración y de la forma en que los participantes del mercado asignan un valor a las acciones y generan su consecuente precio, se podrá entender cómo puede afectar la publicación de resultados de una empresa a la cotización en bolsa de empresas comparables.

En un mercado eficiente, la cotización de una compañía reflejaría el valor teórico de una participación teniendo en cuenta toda la información públicamente disponible sobre dicha empresa en el momento de elaboración del juicio de valoración (Fama, *Efficient Capital Markets: a review of Theory and Empirical work*, 1970). Es en aquellas situaciones en las que el valor intrínseco de las empresas difiere de su cotización dónde se producen oportunidades de inversión, y dónde el inversor informado y racional debe actuar. Dicho de otra forma, cuándo el riesgo que deba asumir el inversor difiera del retorno esperado para dicho riesgo en el momento concreto, se encontrarán oportunidades de inversión.

Mediante este proceso, el mercado eficiente, a través de la interacción entre la oferta y la demanda retorna el valor de mercado de las empresas a su valor teórico, considerando además en cada momento la nueva información disponible. De no producirse dicha regresión a la media, se producirían situaciones de arbitraje teórico.

Por ello, se entiende que en el momento concreto de publicación de resultados se genera una divergencia entre el valor de la compañía previo a la publicación de resultados (basado en unas valoraciones que teóricamente deberían partir de unos datos operativos semejantes a aquellos que los analistas predicen de media) y aquel que los participantes del mercado establecerán con la actualización de los datos y el desvelamiento de las incógnitas (Alzahrani & Skerratt, 2009).

Teniendo en cuenta lo anterior, se explica la importancia de conocer el valor teórico de las empresas de cara a analizar y predecir el comportamiento de las empresas en el mercado y su cotización, al igual que será importante entender las palancas que mueven dichas valoraciones y su por qué. Por ello, resulta conveniente estudiar los distintos métodos de valoración que existen, su funcionamiento y aplicación.

2.1 Métodos de valoración intrínseca

Una primera forma de entender la valoración de una empresa es considerando que su valor reside en su capacidad de generar dinero. Entendido de esta forma, se podría considerar que una empresa debe valer tanto como el valor descontado de sus flujos de caja futuros. Este método de valoración se conoce como *Discounted Cashflows*. El mecanismo por el que funcionaría este método sería mediante la aplicación de una tasa de descuento a los flujos de caja disponibles para el accionista durante la vida operativa de la empresa (Morgan Stanley Investment Management, 2021).

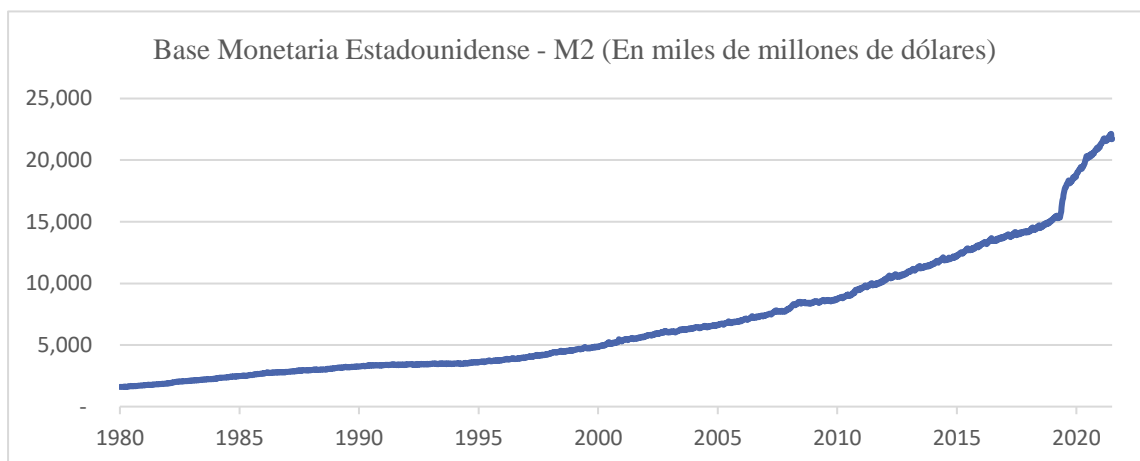
Es conveniente en este sentido comprender qué es exactamente lo que recibe el inversor por su capital, y por qué ha de medir su rentabilidad de la forma en que lo hace. Hay que entender que el dinero es un bien (Alden, 2022) i) **divisible** en tanto que puede ser subdividido sin que las partes pierdan valor con respecto al todo, ii) **portátil** en tanto que se puede trasladar de un lugar, físico o virtual a otro, iii) **fungible** e intercambiable de forma que el dinero no sea especificable y valga lo mismo toda unidad de dinero especificada en una misma moneda y iv) **verificable** en tanto que se pueda comprobar el origen del dinero y su procedencia adecuada (en el caso de las monedas Fiat, esto se materializaría en poder asegurar que el dinero haya sido emitido por el Banco Central correspondiente). A estas características podría añadirse la cualidad del dinero como bien “escaso” en tanto que se mantenga relativamente estable su valor en el tiempo por no alcanzar niveles inestables de *stock-to-flow*³. Por todo ello es el dinero la forma en que medimos la riqueza y la rentabilidad sobre la inversión.

La razón por la que se aplica una tasa de descuento a los flujos de caja futuros es múltiple, y reside en la premisa de que una cantidad de caja en el presente tiene un valor diferente (y generalmente mayor) al que se le debería asignar a un flujo de caja futuro. Esto ocurre por una serie de razones (Damodaran, *The Time Value of Money*) que resulta conveniente enumerar y entender:

- Diferencia entre el consumo presente y el futuro. Generalmente el ser humano prioriza el consumo presente al consumo futuro, y simplemente por esto se encuentra reticente a asignar un mismo valor a una misma cantidad de dinero presente que a una cantidad futura.

³ El stock to Flow se define como el ratio que mide el número de años que se tardaría en reemplazar la totalidad del volumen existente de la unidad de dinero concreto teniendo en cuenta el aumento anual del volumen. En este sentido, el oro se considera una forma de dinero conveniente entre otras cosas por su alto *stock-to-flow* ratio, que se ha situado históricamente en 67x (lo que equivale a 67 años de producción anual equivalen de media al volumen total de oro en un momento dado). En el caso de las monedas Fiat, dicho ratio se encuentra en torno a 20x (Alden, 2022)

- Coste de oportunidad. Como se ha explicado anteriormente, siendo el dinero un bien líquido por definición, hay que tener en cuenta que una unidad de dinero en la actualidad puede ser destinado a multitud de fines y oportunidades de inversión. Por el contrario, durante el periodo de de inversión de dicha unidad monetaria en una empresa, el dinero se transforma en una acción o participación y por ende pierde parte de sus cualidades líquidas. En este sentido, se concluye que un dinero invertido en una empresa no puede estar a la vez invertido en otra, y por ello se considera que existe un coste de oportunidad que se debe tener en cuenta.
- Capacidad de reinversión del capital. Además, es importante tener en cuenta que el dinero invertido genera rentas, que a su vez se pueden emplear para generar más dinero, provocando de esta forma un ciclo virtuoso. Si bien es cierto que esto ocurre igualmente con las inversiones (dado que dicho ciclo compuesto ocurre también con las inversiones de las empresas), si uno confía más en su capacidad de generar rentas con el propio dinero que en la de aquellas empresas en las que se plantea invertir, esta capacidad de reinversión del capital pasaría a ser un factor que considerar en la ponderación del dinero futuro en relación al presente.
- Base monetaria. Es importante considerar que la cantidad de dinero varía a lo largo del tiempo en función de las políticas monetarias que implementan los Bancos Centrales. En este sentido, es interesante el caso de Estados Unidos y su M2 (base monetaria que incluye el efectivo, los depósitos).



- Certeza de los flujos de caja. Finalmente, se debe tener en cuenta un importante factor a la hora de valorar el dinero futuro en relación con el presente. Dicho factor es la certeza en la obtención del dinero, dado que el presente es lo único cierto y verificable, mientras que en el futuro nos es accesible únicamente la incertidumbre (en mayor o menor medida dependiendo de la entropía que rodee a la inversión, pero incertidumbre de todas formas). Si bien esto es cierto y debe tenerse en cuenta, es necesario mencionar la intrínseca

relación entre el riesgo y el beneficio en un mercado eficiente. De esta forma, la mayor incertidumbre debería traer consigo una mayor recompensa en forma de retornos, con lo que la certeza en los flujos de caja debería tener un efecto neutral en la valoración del dinero futuro. Sin embargo, el mercado no es siempre eficiente, ni el ser humano racional. Ciertos sesgos hacen al hombre inclinarse por lo presente en lugar de el incierto futuro pese a darse condiciones en las que los beneficios esperados sean equivalentes.

Entendida de esta forma la valoración de una empresa y la importancia del *time value of money*, juega un papel importante el factor de descuento que se utilice en el cálculo del valor intrínseco de una empresa. En definitiva, dicho factor de descuento es una aproximación a la diferencia en valor de una unidad monetaria futuro con respecto al presente. El valor de descuento se puede asimilar al coste ponderado del coste de capital del inversor (“WACC”), que se calcula en base al coste del capital de los fondos propios y al coste de la deuda.

En función del descuento que se aplique a los flujos de caja, perderán o ganarán valor. Hay por lo tanto una fuerte correlación entre el valor intrínseco de una empresa y los distintos elementos que lo conforman. Siendo el coste de dichos fondos un elemento estrechamente relacionado con la política monetaria, las relaciones internacionales y la macroeconomía, existirán tendencias externas al ámbito de la propia empresa que afectarán a su cotización, tal y como se explica más adelante.

2.2 Métodos de valoración relativa

En segundo lugar, se encuentran los métodos de valoración de naturaleza relativa. Éstos se basan en la premisa de que el mercado de media logra establecer un precio a una empresa que se asemeja a su valor. De esta forma se entiende además que aquellas empresas que tienen una serie de características semejantes, como por ejemplo el volumen de negocio, la ubicación de sus operaciones, los márgenes sobre sus ventas o los productos o servicios que comercializan, deben tener un precio relativo similar. Sumando ambas hipótesis, se llega a la conclusión de que entre una serie de empresas comparables, el mercado de media valorará correctamente un sector.

Teniendo en cuenta lo anterior, surgen dos formas diferenciadas de valorar una empresa, basándose ambas en la comparabilidad entre empresas.

Identificar los comparables es de gran importancia, ya que terminará determinando el múltiplo con el que se valora la empresa en cuestión. En este sentido, una serie de características

(Damodaran, *The Little Book of Valuation: How to Value a Company, Pick a Stock and Profit*, 2011) que deberían compartir dos empresas para ser consideradas comparables son:

- Tasas de crecimiento.
- Capacidad de generación de caja.
- Riesgo.

La conjunción de estas características hará que el riesgo, y por ende el retorno esperado, de dos compañías sea similar. Si atendemos al Five Factor Model, son algunos de estos factores los que se ponderan para explicar el retorno de las acciones, incluyendo el efecto del tamaño o la geografía representada a través de la tasa libre de riesgo por ejemplo (Fama & French, *A five-factor asset pricing model*, 2014).

En cuanto al tamaño, la escalabilidad de los negocios y el efecto red generan fosos económicos que permiten a las empresas de mayor tamaño distinguirse de aquellas que operan en una menor escala, creando de esta forma ventajas competitivas (Dorsey, 2008). En relación con la escala, ésta permite apalancar los costes fijos operativos sobre una mayor base de productos, lo cual resulta en una reducción en el coste marginal de los bienes o servicios provistos. En cuanto al efecto red, se genera cuándo una empresa sostiene una base clientelar suficientemente grande como para generar que la utilidad de los bienes o servicios ofrecidos aumente con el número de usuarios (esto mismo es lo que ocurre con redes sociales, que aumentan el valor ofrecido a los usuarios cuándo aumenta el número de usuarios que las utilizan). Por lo anterior, es razonable que al inversor se le deba “premiar” cuándo participa de una empresa de menor tamaño, a mayor riesgo, mayor ha de ser el retorno esperado para atraer un mismo capital (aquello conocido como el “small cap premium”). A modo de ejemplo, entre 1927 y 1981, las empresas estadounidenses de menor tamaño obtuvieron una rentabilidad superior a aquellas de gran tamaño en 3.1 puntos porcentuales al año (Financier Worldwide Magazine, 2017).

Cabe destacar que técnicamente no es necesario que dos empresas pertenezcan a una misma industria para ser consideradas comparables, ya que teniendo un mismo riesgo, capacidad de generación de caja y tasas de crecimiento, las valoraciones relativas en un mercado eficiente deberán ser similares.

2.2.1 Método de los múltiplos comparables

De acuerdo con esta metodología, se calcula el valor de una empresa en base a los múltiplos con los que cotizan empresas operativamente similares.

En primer lugar, deberemos determinar qué es lo que queremos valorar de la empresa analizada. En función de si queremos valorar el *equity* o el *entreprise value* de la empresa, deberemos utilizar

unos múltiplos u otros. Esto ocurre porque el *enterprise value* es la valoración de las operaciones de la empresa, sin tener en cuenta la estructura de capital existente, mientras que el *equity value* es el valor de la empresa que es directamente atribuible a los accionistas de la misma. El valor del *equity value* de una empresa en un momento determinado se calcula como el precio por acción que se ofrece en bolsa, multiplicado por el número de acciones. En cuanto al valor del *enterprise value*, éste se calcula sumando al *equity value* la deuda contraída por la empresa, restándole el efectivo sin fines operativos⁴, sumando además el valor de las participaciones preferentes en la empresa y los intereses minoritarios.

En este sentido, de entrar a valorar el *equity value*, deberemos utilizar métricas operativas apalancadas, en tanto que deben ser dependientes de la estructura de capital empleada por la empresa.

De esta forma, una métrica apalancada debe considerar los ingresos generados por la empresa, el coste de generar dichos ingresos, los intereses pagados por la deuda contraída en forma de apalancamiento operativo, y los impuestos que se han de pagar sobre los ingresos de explotación de la empresa (considerando igualmente el *tax shield* que genera el empleo de deuda en la estructura de capital de la empresa mediante la deducibilidad de los intereses pagados). Un ejemplo de métrica apalancada sería por lo tanto el resultado del ejercicio de la compañía.

Por otro lado, las métricas desapalancadas empleadas para el cálculo del *enterprise value* de una empresa son aquellas ajenas a las elecciones que haya tomado la administración de la empresa con respecto a la forma de financiar sus inversiones y operaciones. Un ejemplo de dicha clase de métricas sería el importe neto de la cifra de negocios, al igual que lo es el EBITDA o el EBIT.

Teniendo en cuenta lo anterior, se entiende que los múltiplos más utilizados para valorar relativamente una empresa sean

- Para calcular el *equity value*, se suele utilizar el Price to Earnings (P/E, cotización de la empresa en bolsa dividido por los resultados del ejercicio de la empresa o *net income*)
- Para calcular el *enterprise value*, se utiliza normalmente o el EV to EBITDA o el EV to EBIT. En ambos se divide el *enterprise value* de una empresa por una métrica desapalancada.

⁴ Para entender el *enterprise value*, es necesario partir de la base de que éste representa el valor de las operaciones de la empresa observada, que en definitiva equivale a la combinación del valor neto de los activos empleados en las operaciones de la empresa, y las futuras capacidades de crecimiento de dichas operaciones. Por ende, el valor neto de los activos no operativos no se tiene en cuenta para el cálculo del *enterprise value* (al sí estar incluido en el *equity value*, debemos sustraer el valor de dichos activos no operativos).

El segundo paso, por lo tanto, es identificar el universo de empresas comparables, aquellas que tienen unos rasgos similares a la empresa valorada y por ende se estima que se deben valorar de una forma relativamente similar los ingresos o beneficios que se generan en una u otra empresa. Este primer paso se debe realizar siguiendo los criterios anteriormente enunciados.

2.2.2 Método de las transacciones precedentes comparables

Este método consiste en calcular el valor de una empresa como aquel que resulta de multiplicar por una métrica operativa (ventas, EBIT, EBITDA, resultado del ejercicio etc.) el múltiplo que se ha pagado en una transacción en la cuál la empresa objetivo pueda considerarse como comparable a la empresa valorada. Los principios y pasos enunciados anteriormente para el método de los múltiplos comparables anteriormente explicados, aunque le son de aplicación una serie de particularidades.

2.3 Tendencias macroeconómicas y otras consideraciones que afectarán al precio asignado a una empresa

En el contexto de este análisis, es importante estudiar los distintos eventos o tendencias que podrán afectar a la cotización de las empresas, y que por lo tanto distorsionarán los resultados obtenidos, y dificultarán la obtención de conclusiones sobre la relación entre la publicación de resultados de una empresa y el precio de las acciones de sus empresas comparables.

2.3.1 Existencia de ciclos macroeconómicos

A lo largo del tiempo, los mercados financieros y la economía recorren repetitivos ciclos, “*bull markets*” se suceden de “*bear markets*”. Tal y como Theodor Reik (1965) lo describía: “*There are recurring cycles, ups and downs, but the course of events is essentially the same, with small variations. It has been said that history repeats itself. This is perhaps not quite correct; it merely rhymes.*”, lo que viene a decirnos que a lo largo del tiempo el mercado financiero recorre periodos alcistas y periodos de declive, y el inversor se encuentra durante ellos en situaciones que aún no siendo similares, se parecen. La existencia de dichos ciclos implicará que el precio de las acciones se verá afectado por sucesos ajenos a la nueva información provista por la publicación de resultados. Dos periodos serán por lo tanto difícilmente comparables únicamente por la métrica de apreciación o depreciación del valor de una acción. Es importante en este sentido recordar la enorme variabilidad que hay en el mercado entre años; a modo de ejemplo, en el periodo entre 1967 y 2022, el SP500 ha tenido unos retornos medios anualizados del 10%, y sin embargo

durante ese mismo periodo de 65 años, en tan solo seis años el retorno anual se ha encontrado entre el 8% y el 12% (Marks, Bull Market Rhymes, 2022).

2.3.2 Política monetaria de los Bancos Centrales

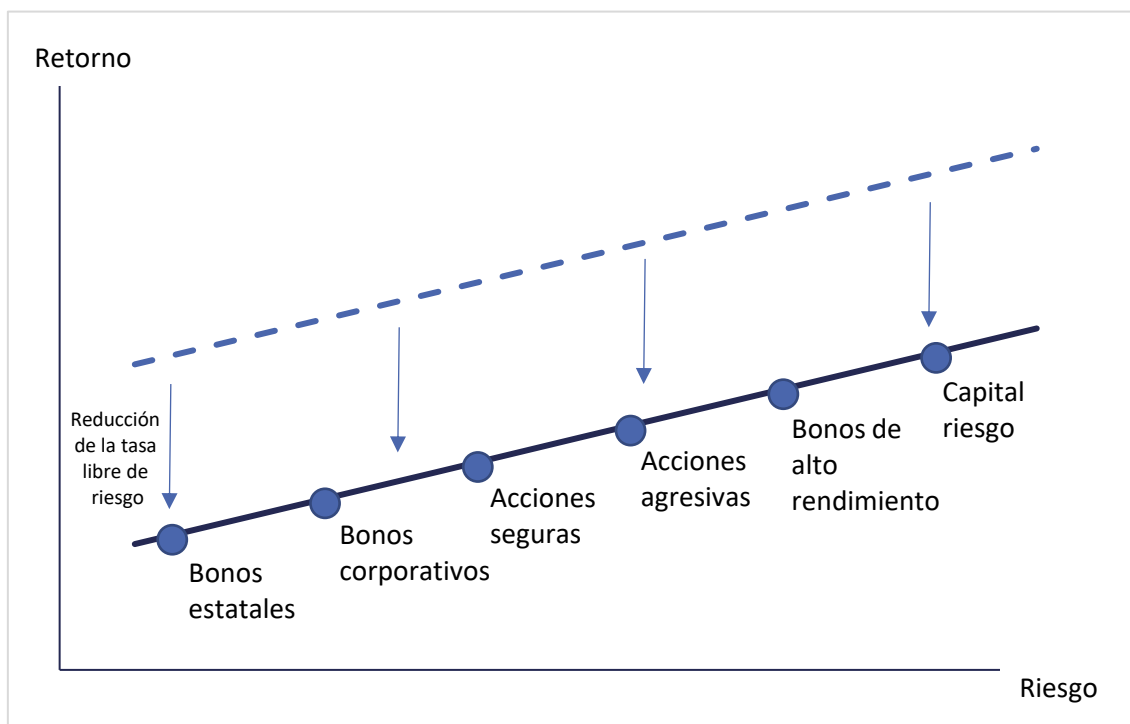
Otro importante factor que afectará los precios de las acciones más allá de sus resultados financieros, o de la información que revele sobre una empresa el hecho de que una empresa comparable publique unos resultados que difieran de las predicciones de los analistas, es la política monetaria que se esté llevando a cabo por parte de los Bancos Centrales.

Entre otras cosas, el tipo de interés establecido por el regulador jugará un importante papel en las valoraciones y precios de las acciones (Marks, Coming into Focus, 2020). La forma en la que los tipos de interés fijados pueden afectar enormemente la valoración de las empresas, y por ende ser factores externos a los resultados operativos de las empresas que afectan a la cotización de estas, se explica a continuación.

En primer lugar, existe el efecto de los tipos de interés como estímulo de la economía, al incentivar el consumo mediante el abaratamiento de la financiación y la reducción de los retornos asociados al ahorro. Este abaratamiento de la financiación permite mayores inversiones de capital, que resultan en la estimulación de la economía.

En segundo lugar, una reducción de los tipos de interés conlleva la disminución de la tasa libre de riesgo, y por ende, un aumento en valor presente de los flujos de caja futuro. Este suceso conlleva a una mayor valoración intrínseca de las empresas, afectando en particular a las empresas de tipo *growth*, empresas en rápido crecimiento y cuyos retornos se encuentran en el futuro (este sería el caso por ejemplo de las tecnológicas que aún no generan beneficios). Lo contrario es también cierto, unos mayores tipos de interés acaban destruyendo valor de las empresas en crecimiento, mientras que aquellas maduras y establecidas prevalecen.

En tercer lugar, una reducción de los tipos de interés reduce los retornos esperados a lo largo de las distintas opciones de inversión. Esto ocurre porque al reducir la rentabilidad de los bonos estatales, el activo más seguro, aumenta la demanda de activos que entrañan un mayor riesgo, y por ende al aumentar la demanda, aumenta el precio y se reducen las rentabilidades esperadas.



En último lugar, la reducción de los retornos demandados afecta directamente a la valoración de las empresas, aumentando los múltiplos a los que cotizan las empresas (Marks, Coming into Focus, 2020). Por ejemplo, cuándo la rentabilidad de los bonos del tesoro es de un 3%, los inversores en acciones demandarán ese 3% más un *market risk premium*. Asumiendo en el ejemplo que dicho premium se situará en el 3.5%, los retornos que demandaría un inversor en *equity* serían de 6.5%. Esta rentabilidad representa indirectamente los flujos recibidos divididos por el precio pagado, que es la fracción inversa del ratio precio-beneficios (*PE ratio*). De esta forma el PE implícito de un retorno esperado del 6.5% es de 100 dividido por 6.5%, o 15.4x. Al reducirse el retorno esperado de los bonos de un 3% a un 1%, el ratio implícito vendría a ser (asumiendo un mismo *market risk premium*) 22.2x.

De esta forma, los tipos de interés afectan enormemente a la cotización de las acciones, y siendo un factor externo a las propias operaciones de la empresa, termina condicionando el desenvolvimiento de las empresas en bolsa.

Capítulo 3. Obtención y tratamiento de datos

Con todo lo anterior, está claro que para realizar una valoración relativa de una empresa, es necesario el minucioso análisis de las empresas del sector en el que opera la empresa valorada. Sin embargo, esta clase de análisis en el contexto de un análisis de *business analytics* resulta en la práctica inviable. Al fin y al cabo, una de las premisas necesarias para realizar dicha clase de estudio es la existencia de un volumen suficiente de datos. Recordemos que las 5Vs del Big Data, que se postulan como características innatas, son *Volume*, *Velocity*, *Variety*, *Veracity* y *Value*. En definitiva, para realizar un análisis riguroso de empresas comparables es necesario analizar pormenorizadamente cada una de las empresas candidatas a ser empresa comparable, pero esto es algo tremendamente costoso en tiempo, con lo que existe un *tradeoff* entre la cantidad de datos en uso y su calidad.

Existiendo, sin embargo, una aproximación relativamente buena de la comparabilidad de dos empresas en el sector al que pertenecen, parece conveniente decantarse en esta cuestión por la utilización de un volumen suficiente de datos, empleando una comparabilidad aceptable. De esta forma, se opta en este análisis por estudiar el comportamiento de una empresa en función del comportamiento relativo de las empresas comparables, y se define una empresa comparable como aquella que pertenece a un mismo sector que la empresa valorada y analizada.

Además, tal y como se explicaba en el capítulo anterior, se comprende la importancia de los ingresos netos generados por las empresas en el análisis de la rentabilidad que podrá obtener el inversor. Igualmente, queda explicado por qué al mercado no le será irrelevante la publicación de resultados y se vislumbra cuáles serán las métricas más relevantes para asignar valor a una empresa.

En cuanto a las bases de datos preexistentes que van a ser utilizadas para realizar el análisis que atañe a este trabajo, son las siguientes:

- SP500 Quarterly. Base de datos de las distintas métricas financieras publicadas por cada empresa (incluyendo aquellas pertenecientes al SP500) cada cuatrimestre. Incluye además de las distintas partidas de balance y cuenta de pérdidas y ganancias que presenta cada empresa en cada momento de publicación de resultados, la industria y sub-industria a la que pertenece cada empresa (que a su vez viene acompañada de su correspondiente ticker, que nos servirá en el análisis como identificador concreto de cada compañía). Tal y como se explicó anteriormente, es esta sub-industria la que nos servirá para agrupar las empresas del SP500 por grupos de empresas comparables.

- Forecasts IBES Analysts. Incluye las estimaciones de los resultados de los ingresos netos por acción que se prevé publicarán las empresas en cada cuatrimestre y año. Se incluye a su vez identificadores únicos para cada analista que realiza las estimaciones, al igual que para cada compañía a la que pertenece el analista que realiza la predicción.

Partiendo de esta base, idealmente la información que se debería tener para llevar a cabo el análisis del efecto en el precio y cotización de una acción de la publicación de resultados de sus empresas comparables, debería tener la siguiente forma y columnas:

- Ticker. Esta primera columna se referiría a un identificador particular e irrepetible de cada empresa, de forma que nos permita saber a qué compañía se refiere cada serie de datos.
- Sub-industria a la que pertenece cada empresa observada. Tal y como se ha explicado anteriormente, esta variable nos servirá como aproximación de la comparabilidad entre empresas, para así generar grupos de empresas relativamente equivalentes.
- Fecha de publicación de resultados. En relación a esta segunda columna, sería esencial para estudiar los efectos del anuncio de resultados en la cotización, la precisión de la fecha de publicación.
- Métrica representativa de la predicción de los analistas de los ingresos netos por acción para el trimestre y año concretos.
- Ingresos netos por acción publicados en el trimestre y año que atañen a cada observación.
- Métrica representativa del desenvolvimiento en bolsa de las empresas comparables tras la publicación de resultados de la empresa observada.
- Métricas analíticas de los resultados publicados en comparación con el trimestre previo.
- Métrica representativa del tamaño de la empresa observada. Esto nos será de utilidad para estudiar el efecto en las empresas comparables de la publicación de resultados considerando el tamaño de la empresa observada, en este caso, mediante los ingresos publicados en el trimestre concreto.

La combinación de un Ticker y una fecha concreta será el identificador único para cada observación de nuestra base de datos.

En cuanto a las herramientas utilizadas para la obtención de la base de datos, se ha empleado mayoritariamente Python, aunque también se ha usado para aplicaciones puntuales Excel por su simplicidad y facilidad de uso para el tratamiento de datos en aspectos concretos del trabajo.

3.1 Obtención del ticker y de las fechas de publicación de resultados

El primer paso para la creación de la base de datos es la instalación de los paquetes de Python necesarios para el tratamiento de datos y la realización del *webscrapping* que se detallará más adelante. Los paquetes utilizados han sido:

- BeautifulSoup4 en su versión 4.10.0
- Pandas en su versión 1.4.0
- Requests en su versión 2.27.1
- Tqdm en su versión 4.62.3

Habiendo explicado las herramientas que han sido utilizadas, es necesario entender de dónde se obtienen los datos de la publicación de resultados. Cada trimestre, las empresas cotizadas en Estados Unidos deben presentar sus resultados trimestrales a la Securities Exchange Comission, que después las publica en su base de datos, EDGAR. Es de EDGAR por lo tanto de dónde se obtienen las fechas de publicación de resultados, además del tipo de resultados que se presentan (10-K en el caso de tratarse de resultados anuales, y 10-Q en el caso de ser resultados cuatrimestrales los que se presentan).

El primer problema al que nos enfrentamos es la accesibilidad de la base de datos de la SEC. Al explorar la página web de la base de datos, se descubrirá que el formato subyacente en el que se presentan los datos no es el adecuado para realizar un *webscrapping*, dado que no se proporcionan los datos en formato de tabla. Sin embargo, tras investigar e inspeccionar la página web y su código, se averigua que existe una dirección alternativa a través de la cuál se puede acceder a los datos por empresa en formato tabla con n filas que dan detalles de cada publicación realizada por la compañía buscada. Dicha dirección web alternativa corresponde a la antigua *interface* de EDGAR, y tiene una estructura propicia para la labor que nos atañe.

La página web en cuestión es la siguiente, en la que el campo “CIK” corresponde al código identificador concreto de la empresa buscada, y el campo “count” corresponde al número de resultados por página que se desean obtener.

https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK=#####&type=10-K&dateb=&owner=exclude&count=####&search_text=

Cabe destacar que esta página web no está operativa para la búsqueda ordinaria de información de empresas, sino que puede ser utilizada únicamente mediante la manipulación de los parámetros

de la dirección web. Al tratar de navegar esta dirección en el navegador, se remite al usuario a la nueva versión de la página web.

3.1.1 Central Index Key (CIK)

Teniendo en cuenta la necesidad de modificar los parámetros de la dirección web para la obtención de datos, es en primer lugar necesaria la tenencia de una lista de códigos identificadores de las empresas del SP500. Para conseguir dicha lista, se realiza un *webscrapping* de la página web “https://en.wikipedia.org/wiki/List_of_S%26P_500_companies”, en la cuál se encuentra el listado de empresas pertenecientes en la actualidad al SP500 junto a una serie de datos relevantes de cada una de las empresas. La forma que tiene la tabla es la siguiente:

Symbol	Security	SEC filings	GICS Sector	GICS Industry	Sub-	Headquarters Location	Date added	first	CIK	Founded
MMM	3M	reports	Industrials	Industrial Conglomerates		Saint Paul, Minnesota	9/8/76		66740	1902
AOS	A. O. Smith	reports	Industrials	Building Products		Milwaukee, Wisconsin	26/7/17		91142	1916
ABT	Abbott	reports	Health Care	Health Equipment	Care	North Chicago, Illinois	31/3/64		1800	1888
ABBV	AbbVie	reports	Health Care	Pharmaceuticals		North Chicago, Illinois	31/12/12		1551152	2013 (1888)
ABMD	Abiomed	reports	Health Care	Health Equipment	Care	Danvers, Massachusetts	31/5/18		815094	1981

Como se puede apreciar, es el parámetro CIK de la tabla que se obtiene en la mencionada dirección web, la que nos será relevante para el análisis que nos atañe, junto con la referencia de cada CIK con cada Ticker-Symbol.

Para realizar el *webscrapping* se opta por la utilización de la función `read_html` que proporciona la librería Pandas (pd), de manera que el código necesario para conseguir la lista de Tickers y Central Index Keys sería:

```
table = pd.read_html('https://en.wikipedia.org/wiki/List_of_S%26P_500_companies')
```

Habiendo hecho esto, se genera un diccionario en el que se guarda para cada key (Ticker), un value (Central Index Key). Esto se consigue mediante el comando:

```
{ticker:cik for ticker, cik in sp500[['Symbol','CIK']].values.tolist() }
```

3.1.2 Fechas de publicación de resultados

Teniendo ya un diccionario en el que almacenamos cada uno de los códigos CIK junto con cada Ticker, la forma de obtener las fechas de publicación de resultados de cada empresa del SP500 es mediante la elaboración de un bucle que realice *scrappings* sistemáticos a EDGAR mientras recorre la lista de empresas del SP500.

El resultado final de este *web scrapping* es la obtención de una base de datos en las que se encuentran las 100 últimas publicaciones (25 años) de resultados para cada una de las empresas del SP500, o en el caso de tratarse de empresas que cotizan en bolsa desde hace menos tiempo, sus publicaciones desde su *listing*.

La forma de la base de datos es la siguiente:

Filings	Format	Description	Filing Date	File/Film Number	ticker	cik
10-Q	Documents Interactive Data	Quarterly report [Sections 13 or 15(d)]Acc-no:...	26/10/21	001-03285211347847	MMM	66740
10-Q	Documents Interactive Data	Quarterly report [Sections 13 or 15(d)]Acc-no:...	27/7/21	001-03285211117640	MMM	66740
10-Q	Documents Interactive Data	Quarterly report [Sections 13 or 15(d)]Acc-no:...	27/4/21	001-0328521858374	MMM	66740
10-K	Documents Interactive Data	Annual report [Section 13 and 15(d), not S-K I...	4/2/21	001-0328521590417	MMM	66740
10-Q	Documents Interactive Data	Quarterly report [Sections 13 or 15(d)]Acc-no:...	27/10/20	001-03285201264704	MMM	66740

En este código se definen dos funciones, en primer lugar aquella que se encarga de obtener datos de archivos HTML y XML mediante la utilización del paquete Beautiful Soup, y en segundo lugar aquella función que se encarga de recorrer mediante un bucle *for* la lista de *tickers* y códigos CIK, que son a continuación introducidos en la función `get_soup` una dirección web específica de cada empresa del SP500 y una serie de credenciales que EDGAR solicita para realizar el *scrapping* (headers). A continuación, para cada iteración y resultado de la función `get_soup`, se ordenan los datos de acuerdo a una serie de columnas previamente nombradas (headings). Finalmente, se realiza un bucle dentro del primer bucle que recorre cada dirección web para que se añadan a una lista inicialmente vacía cada una de las filas que se obtiene del *scrapping* para cada una de las empresas.

El resultado final de la extracción de datos es una base de datos con 42.970 filas. El tiempo de extracción de toda la base de datos es de 4 minutos y 22 segundos, lo que significa que se logra extraer de media 1,92 empresas por segundo. Es este otro ejemplo más del gran potencial del

business analytics, que permite simplificar laboriosas y repetitivas tareas, haciendo por lo tanto más accesible la extracción de conclusiones a través del análisis de datos, tras agilizar la propia extracción de los datos. Sin ir más lejos, de haberse realizado este ejercicio hace 40 años, habría sido necesario consultar mecánicamente las bases de datos, lo cuál implicaría un trabajo que se habría de medir en magnitudes de días y semanas, no de minutos y segundos. Más detalles sobre el código de Python que obtiene las fechas en el **Anexo I**.

3.1.3 Obtención de las subindustrias y creación de los grupos de empresas comparables.

Es la comparabilidad entre empresas un punto crítico para este análisis, y por ello el correcto tratamiento de los datos se erige como un proceso esencial de este trabajo. Para realizar este análisis se ha utilizado una base de datos, “sp500_quarterly”, que incluye datos trimestrales para cada compañía (identificada mediante su correspondiente ticker) de sus ingresos y de la industria y subindustria a la que pertenecen.

Partiendo de esta base de datos, se ha decidido que para el objetivo concreto de este análisis, lo más conveniente sería obtener una lista de las empresas constituyentes del SP500, a la que se le añadiesen tantas columnas como empresas de su misma sub-industria haya, al igual que una columna que incluya la propia sub-industria a la que pertenece la empresa observada. Realizar el análisis de esta manera nos permitirá más adelante realizar un bucle que recorra cada observación de la base de datos y para ella recorra sus columnas correspondientes a cada una de sus empresas comparables, para después obtener los datos de cotización deseados para cada empresa comparable en el momento de publicación de resultados de la empresa objetivo.

Considerando la forma de la base de datos deseada, se ha optado por realizar el tratamiento de los datos para obtener la lista de empresas comparables en Excel. La forma en que se ha realizado dicho análisis ha sido la siguiente:

1. En primer lugar, se han filtrado los datos para conservar únicamente los datos de empresas que en la actualidad cotizan en el SP500, y la sub-industria a la que pertenecen.
2. Teniendo la lista de empresas y sub-industrias lista, se ha realizado una matriz en la que las filas corresponden a cada una de las empresas del SP500, siendo las columnas una versión traspuesta de las filas, de manera que hay una relación $n \times n$ de filas a columnas.
3. Una vez se ha obtenido la matriz, se ha formulado una función que permite comprobar para cada una de las celdas de la matriz si la sub-industria a la que pertenece la empresa que representa la fila es igual o no a la sub-industria a la que pertenece la empresa de su correspondiente columna. En caso de compartir sub-industria, se ha impreso en cada

celda el ticker de la empresa de la columna, y en caso contrario se ha dejado la celda en blanco. Para realizar este análisis se han utilizado las funciones IF, INDEX y MATCH integradas en Excel, que han permitido un ágil y cómodo tratamiento de los datos.

- Tras completar la matriz con los tickers comparables a cada fila, el problema se encuentra en que se trata de una matriz de aproximadamente 500 filas por 500 columnas. Para simplificar la matriz, se ha exportado el documento de Excel como csv y se ha abierto con Text Edit. Una vez en Text Edit se ha proseguido con la sustitución de parejas de comas por comas individuales. De esta forma, se consigue mediante la sustitución iterativa la eliminación de los espacios vacíos en cada fila, de manera que prevalecen únicamente el ticker observado y los tickers que pertenecen a su misma subindustria, cada uno de ellos separado por exactamente una coma (lo cuál permitirá la extracción a Python más adelante).

El resultado final es una base de datos en la que hay tantas columnas como el máximo de empresas que pertenecen a una misma sub-industria (en este caso 26, que es el número de empresas categorizadas como bancos regionales). Cabe destacar que el número total de subindustrias es de 126, lo cuál implica que existen de media aproximadamente 4 empresas por subindustria, y por ende cada empresa tiene de media tres empresas comparables

Subindustry	Ticker	Peer1	Peer2	Peer3	...	PeerN
Airlines	AAL	ALK	DAL	LUV	...	Null
Electric Utilities	PNW	AEP	DUK	NEE	...	Null
Health Care Equipment	ABT	BAX	BDX	DHR	...	Null
Semiconductors	AMD	SWKS	ADI	INTC	...	Null
Industrial Gases	APD	LIN	Null	Null	...	Null
...
Interactive Media & Services	TRIP	TWTR	GOOGL	FB	...	Null

Si estudiamos por ejemplo la primera observación de la base de datos, AAL, ésta corresponde a American Airlines. En cuanto a las empresas que son consideradas comparables de acuerdo al presente análisis, serían Alaska Air Group, Delta Airlines, Southwest Airlines y United Airlines, lo cual parece un resultado razonable.

3.1.4 Métrica correspondiente al tamaño de la empresa observada

En relación al tamaño de la empresa observada, es ésta una métrica significativa para el estudio que nos atañe. Es esperable que la publicación de resultados de una empresa de gran tamaño resulta más significativa para las expectativas de los inversores que la publicación de resultados

de una empresa menor. En este sentido, cabe destacar que se ha optado por la utilización de los ingresos como forma de medición del tamaño, aunque habría sido igualmente razonable utilizar otras métricas como el EBITDA o la capitalización bursátil. De todas formas, al tratarse este análisis de un estudio general del SP500, no parece que exista ninguna métrica que logre aportar una visión razonable, equilibrada y justa del tamaño de cada empresa. De utilizarse, por ejemplo, el EBITDA, aquellas empresas cuyos EBITDAs están distorsionados por técnicas contables, o miden sus niveles operativos con métricas alternativas (como sería el caso de las empresas de Real Estate al utilizar métricas EPRA), no verían en el EBITDA representatividad alguna del tamaño. Igualmente, la capitalización bursátil, al tener en cuenta la estructura de capital de la empresa, no mide realmente el tamaño operativo de la compañía observada como si lo haría su Enterprise Value. Con todo esto, se ha optado por utilizar los ingresos, aunque de nuevo, tiene sus limitaciones al variar enormemente entre industrias la capacidad de generación de flujos de caja en base a unos mismos ingresos. Sin embargo, se puede por lo menos esperar que aquellas empresas que pertenecen a una misma sub-industria puedan ver en los ingresos una medida objetiva y no distorsionada de su tamaño.

En cuanto a la obtención de esta variable, se ha optado por utilizar la misma base de datos que se ha empleado para la obtención de las subindustrias. La diferencia ha radicado en que para esta variable concreta se ha conservado la dimensión temporal del año y trimestre al que se pertenece. En cuanto a la obtención del trimestre concreto de cada observación, éste estaba incluido en la propia base de datos.

3.1.5 Métricas de ingresos netos predichas y publicadas

La comparación entre la predicción de los analistas sobre los ingresos netos de las empresas observadas y el dato real de ingresos netos presentado en el momento de publicación de resultados de las empresas nos otorga una valiosa información sobre la “sorpresa” que recibe el mercado. Debemos tener en cuenta que el precio de la acción se forma mediante la conjunción de la oferta y la demanda en un momento concreto y ponderando la información disponible para los distintos participantes del mercado. Es por ello que en un momento determinado el precio de una acción se creará teniendo en cuenta las expectativas de ingresos y niveles generales de las operaciones de dichas empresas (al menos en un mercado eficiente y racional), y dichas expectativas se asemejarán a las predicciones que realizan los analistas sobre los ingresos netos por acción. Por lo tanto, la publicación de resultados supone una actualización de la información, que afectará a la cotización de la propia empresa, y a las empresas que participan de un mismo mercado y sector que la observada.

Es por lo tanto esta métrica una pieza clave en el análisis que atañe a este trabajo. Para la obtención de esta variable, se ha optado por la base de datos del Institutional Broker's Estimate System (IBES). Creada por primera vez en 1976, y comprada por Thomson Reuters en el año 2000⁵, esta base de datos nos aporta observaciones individuales de cada estimación realizada por cada analista para cada empresa seleccionada y para cada trimestre y año. Esto significa que para cada empresa y trimestre existen a su vez numerosas observaciones, que corresponden a cada analista (identificado con un código único), especificándose a su vez la empresa para la que trabaja el analista.

Es por lo tanto una vasta base de datos, con más de 4 millones de observaciones que abarca desde 1976 a la actualidad, y que por lo tanto aporta valiosa información sobre el sentimiento del mercado en relación a los ingresos netos de las empresas y por ende nos permitirá descubrir la existencia o inexistencia de sorpresas tras la publicación de las cifras de negocio reales. Cabe destacar además que el número de empresas cubiertas en la base de datos de IBES cubre 23.000 compañías de más de 90 países y proveniente de más de 23,000 analistas. Limitándose este análisis a las empresas del SP500, se reducirá la base de datos para excluir empresas que no son relevantes para este trabajo.

En cuanto a la forma de medir la sorpresa que recibe el mercado propiamente dicha, se ha calculado mediante la división del resultado real de los datos publicados de ingresos netos por acción entre la mediana de las predicciones de los analistas

3.1.6 Métricas analíticas de los resultados publicados en comparación con el trimestre previo

Teóricamente resultará interesante el análisis del propio desempeño de las operaciones de la empresa que publica resultados. Es cierto que debemos considerar que el mercado tiende a la eficiencia según la información se democratiza, y que en un mercado eficiente los participantes establecerán el precio de las empresas de acuerdo con la información de la que dispongan en cada momento. Sin embargo, la inexistencia de predicciones ad hoc para métricas como el total de activos o la variación de los ingresos netos en términos absolutos significa que la información será recibida en el momento de publicación de resultados, y que dicha información supondrá una actualización de una información sobre los datos del anterior trimestre directamente (sin la existencia de predicciones interinas). Esto hará que el mercado asigne a esa información una relevancia que le hará actualizar sus valoraciones y por ello se ha optado por la inclusión de dichas métricas en el análisis que nos atañe.

⁵ (Refinitiv, 2022)

En cuanto a la fuente de los datos, éstos provienen de la base de datos de Quarterly Results. Se han obtenido tras tratar la base de datos inicial para mostrar el trimestre y el año al que pertenecen, de manera que condicionalmente se han dividido los resultados del año m y trimestre n entre los resultados del año m y trimestre $n-1$ (excepto para el caso en que se tratara del primer trimestre, en cuyo caso se han dividido entre los resultados del año $m-1$ y trimestre $n+3$).

Igualmente, nos aportará información sobre la tendencia que lleva la economía y el sector. En este sentido, cabe mencionar que se ha estudiado la debilidad en la correlación entre el crecimiento del PIB (que en este caso vendría ligado indirectamente a la publicación de unos mejores resultados) y los retornos obtenidos por la inversión en acciones (Dimson, Marsh, & Staunton, 2003). Sin embargo, hay que destacar que el objeto de este análisis no se centra en los retornos a largo plazo, sino en el efecto a corto plazo de la aparición de nueva información.

Hay también que considerar que la existencia de variaciones en esta clase de métricas se puede deber igualmente a la estacionalidad de la demanda. Este sería por ejemplo el caso de las aerolíneas, cuyos vuelos son más demandados durante las épocas estivales del año (ACI Insights, 2018).

3.1.7 Métricas de desempeño de las empresas comparables en el momento de publicación de resultados de la empresa observada – variable objetivo

La forma de medir el desempeño de las empresas tras la publicación de resultados de la empresa observada ha sido mediante el estudio de la cotización en bolsa de las acciones correspondiente a las empresas comparables. Para realizar este análisis, se ha utilizado la lista de empresas comparables anteriormente obtenida a través de la sub-industria, que convenientemente consistía en una serie de columnas que correspondían a las empresas comparables 1 a n que tenía cada empresa observada.

Partiendo de esta base, se ha decidido estudiar la cotización de las empresas de una forma particular. Se ha procedido a dividir el precio de cotización mediano de los 10 días posteriores a la publicación de resultados entre la cotización mediana de los 10 días anteriores a la fecha de publicación. La razón por la que se ha optado por este método de aproximación de los efectos de la nueva información en las empresas comparables ha sido evitar la volatilidad diaria que sufre la cotización de las empresas. De esta forma, se puede por lo tanto analizar no ya cambios puntuales, sino cambios de tendencia. Cabe destacar sin embargo que no es esta una métrica perfecta ya que tiene limitaciones en tanto que la cotización de las empresas atiende a numerosos factores que

actualizan la información disponible a los participantes del mercado continuamente. Tal y como se ha explicado anteriormente, existen varios niveles de tendencias que afectan al precio de las acciones. Para mayor detalle sobre la forma en que opera el programa de Python que obtiene estos datos, ver **Anexo II**.

Cabe destacar que en el proceso de extracción de estos datos se ha tenido que considerar igualmente la existencia o inexistencia de días festivos en los cuáles los mercados permanecen cerrados. Al poder haber n días festivos entre la fecha de publicación de resultados y el décimo día posterior, se ha empleado un paquete de Python que permite determinar si una fecha concreta fue festiva o no en Estados Unidos. El proceso empleado para tener en cuenta dichas fechas festivas ha consistido en la creación iterativa de una lista vacía para cada fecha de publicación, a la cuál se adicionaban tantos días festivos como hubiera habido en el intervalo, y se suma la cifra resultante de días festivos al número de días estudiados para cada observación. De esta forma se consigue siempre considerar 10 días de cotización para cada empresa.

La selección de 10 días en lugar de una cantidad superior o inferior de días se ha debido que 10 días permiten apreciar cambios de tendencia en el corto plazo en la cotización de las empresas observadas. Además, de haber escogido una cantidad superior de días, los resultados podrían quedar aún más distorsionados por elementos externos a la propia observación, como lo sería la publicación de resultados de otras empresas comparables o de la propia empresa cuya cotización en los diez días posteriores se está estudiando. Si bien esto es cierto, el programa elaborado se ha creado de manera que se puede modificar fácilmente el número de días abarcados en el análisis.

Esta variable será considerada la categoría objetivo en el análisis que nos atañe. De existir un modelo que pudiera predecir el desenvolvimiento de las empresas comparables tras la publicación de resultados de una empresa concreta, éste permitiría a su tenedor la obtención de beneficios mediante el *trading* de las posiciones.

El problema concreto se ha planteado como uno de clasificación, en el que el modelo que se ha planteado trata de predecir si la cotización de las empresas tras la publicación de los resultados de las empresas comparables aumenta o disminuye en el corto plazo. Por ello, se ha creado una nueva columna en la base de datos que muestra un 0 en el caso de que los resultados de cotización fueran negativos y un 1 en caso de que los resultados de cotización fueran positivos.

3.2 Base de datos obtenida tras el proceso de limpieza y procesamiento de las bases

Tras la realización de todos los procesos anteriormente mencionados la base de datos que obtenemos tiene como variable objetivo la cotización positiva (1) o negativa (0) de las empresas comparables a aquella que publica resultados en los días posteriores al evento que nos aporta información. En cuanto a las variables que nos ayudarán a clasificar cada situación, tenemos:

- La “sorpresa” que recibe el mercado en el momento de publicación de resultados, entendida como la división entre el resultado obtenido y las predicciones de los analistas menos uno
- El volumen de ingresos de la empresa que publica resultados, lo cuál nos aportará una estimación del tamaño de la empresa, tal y como se ha explicado anteriormente.
- El aumento en el tamaño de los activos de la empresa que publica resultados con respecto al periodo anterior.
- El aumento de los ingresos netos totales de la empresa con respecto al anterior trimestre.
- Finalmente, se añade la industria como variable categórica, dado que no será indiferente que los resultados se obtengan con respecto a una empresa de una industria u otra. Es posible que haya sectores más interrelacionados que otros.

El resultado final es una base de datos con la siguiente forma:

Industry	Revenue	Surprises	PeersYield	%AssetsTotal	%NetIncome
Airlines	9789	0.002255	1	0.003248	-0.60787
Airlines	9630	-0.029844	1	-0.006158	0.93798
Airlines	10160	-0.268192	1	-0.009101	-0.366242
Airlines	7337	0.032403	1	0.578715	-7.923875
Specialty Retail	2082.891	-0.088301	0	-0.011838	-0.452189

En cuanto a la cantidad de datos, en las sucesivas iteraciones de combinación de bases de datos, se ha perdido una gran cantidad de observaciones, siendo finalmente en torno a tres mil.

Se ha esterilizado la base de datos para que sea agnóstica del periodo o año en el que se publican los resultados dado que, pese a que afectarán a los resultados por la existencia de ciclos económicos tal y como se explicaba anteriormente, a posteriori no se podrán utilizar como variables en tanto que si se utiliza este modelo fuera del entorno de los datos presentes, perderá valor el año al representar el incierto presente y no el pasado.

Capítulo 4. Análisis de los datos y estudio de las variables

4.1 Estudio estadístico de las variables

Una vez se han recopilado los datos, es importante hacer un estudio estadístico de las variables con el fin de observar las posibles dispersiones de los datos, sus distribuciones y sus características esenciales. De esta forma se podrá estudiar en cierto modo la calidad de los datos y la significancia de los mismos. Si bien es cierto que estos datos aún no se ponen en relación con la variable objetivo, podremos observar si hay variables que sería conveniente modificar; por ejemplo, mediante la transformación logarítmica. A modo de ejemplo, se muestra a continuación el resultado del empleo de la función `.describe()` sobre algunas de las variables de nuestro modelo.

	Revenue	Surprises	PeersYield	%AssetsTotal	%NetIncome
count	3018	3018	3018	3018	3018
mean	3840.836525	0.444144	0.599404	0.032174	0.16173
std	6787.549585	25.405475	0.490101	0.19802	8.845983
min	-25623	-93.565111	0	-0.48328	-104.333333
25%	590.339	-0.024134	0	-0.008343	-0.387167
50%	1566.0725	0.011641	1	0.012783	-0.04892
75%	3559.375	0.054906	1	0.039521	0.175431
max	49733	1392	1	7.325257	303.611111

Más adelante, se ha llevado a cabo un análisis de la Coeficiente de Variación de Pearson, que mide la desviación de las observaciones con respecto de la media de los mismos de una forma agnóstica en relación con las magnitudes en las que se representan dichos datos. Se considera que en el momento en que el Coeficiente de Variación de Pearson supera la línea del 30% o 0,3, la media ha dejado de ser representativa, como es el caso para nuestros datos en el análisis que nos concierne. El resultado es el siguiente para nuestros datos:

Variable	Coeficiente
Revenue	1.77
Surprises	57.2
PeersYield	0.82
%AssetsTotal	6.15
%Net Income	54.7

En cuanto a la asimetría o *skewness* de los datos, Python ofrece la función `skew()` que nos permite analizar si los datos tienen asimetría positiva o negativa. La asimetría positiva es aquella que se

da en el caso en que la mediana sea inferior a la media de los datos, y lo contrario ocurre en la asimetría negativa. La mayoría de nuestras variables presentan una asimetría positiva.

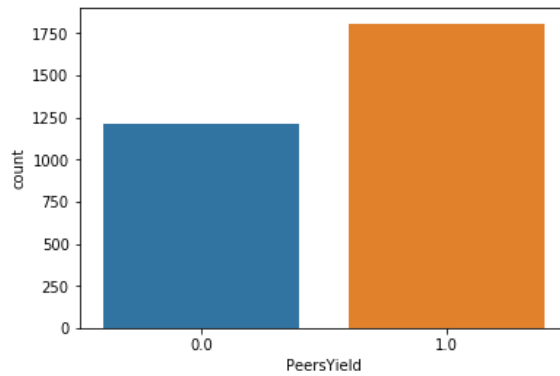
Variable	Skewness
Revenue	3.48
Surprises	54.49
PeersYield	-0.41
%AssetsTotal	22.06
%NetIncome	20.68

En cuanto a las correlaciones entre nuestras variables, apreciamos una gran escasez de correlaciones⁶:

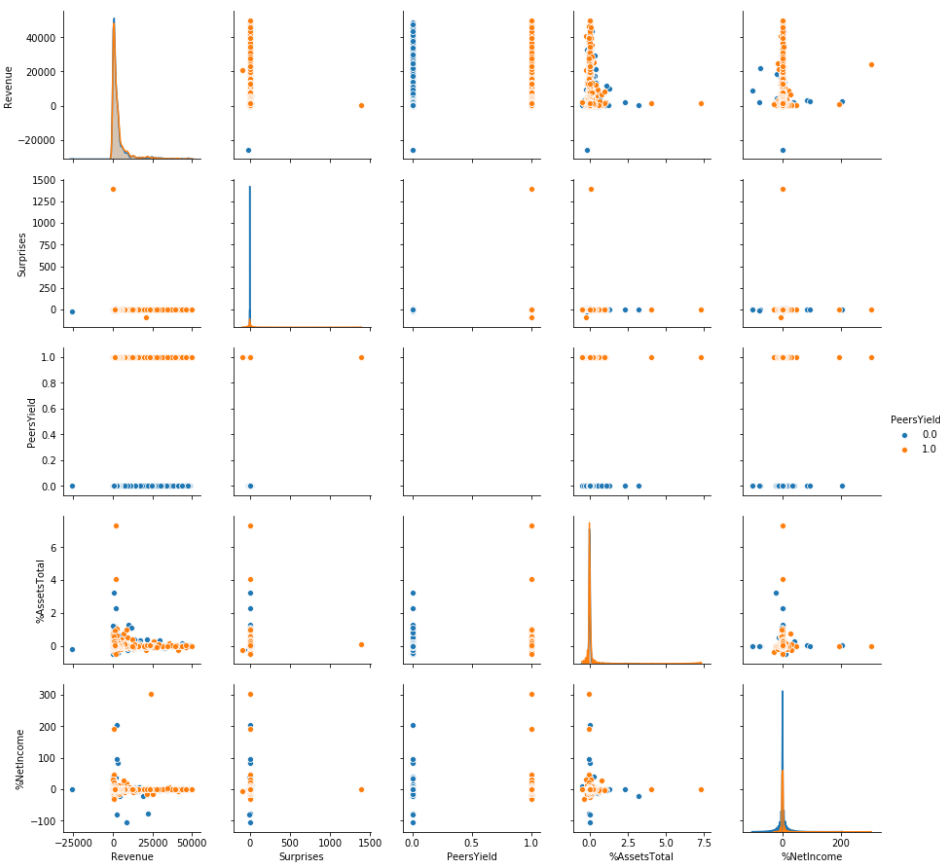


⁶ Correlación medida entre -1 y 1, siendo 1 una correlación positiva absoluta y -1 una correlación negativa absoluta, mientras que un nivel cercano a 0 indicará una escasa correlación.

Además, será interesante el estudio de la variable objetiva. Tal y como se puede observar en el gráfico a continuación, los datos están balanceados en tanto que no hay una abrumadora mayoría de empresas que obtengan resultados positivos ni al contrario. En concreto, un 60% de los datos representan desempeños positivos de la acción, y un 40% negativos (esto es razonable si consideramos que la tendencia del mercado a largo plazo es alcista (Marks, Bull Market Rhymes, 2022)).



La escasa correlación entre la variable objetivo y el resto de variables se puede apreciar además si estudiamos las distintas distribuciones de cada variable y coloreamos los 0s y los 1s. Tal y



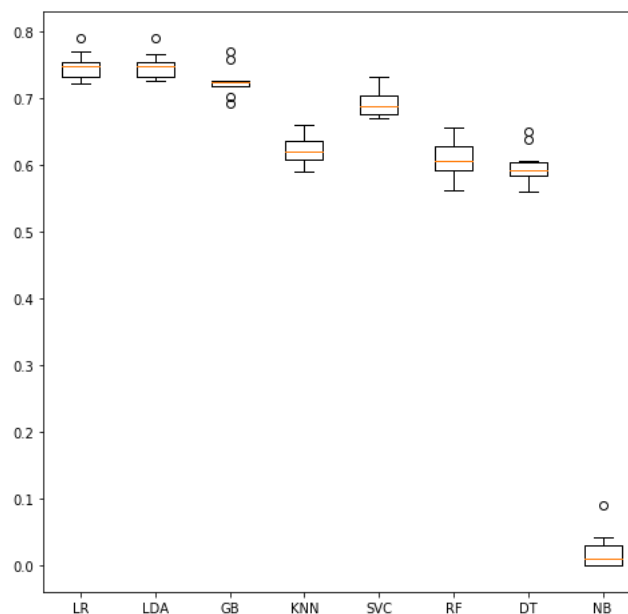
como se aprecia, no existen tendencias claras que nos expliquen el comportamiento de la variable objetivo en función de las variables de las que disponemos.

4.2 Creación, parametrización y resultados del modelo de clasificación

Habiendo estudiado las distintas variables, se ha hecho un estudio de las distintas opciones de modelaje que están disponibles. En particular, se ha estudiado la viabilidad de:

- Regresión Logística – LR
- Linear Discriminant Analysis – LDA
- Gradient Boosting Classifier – GB
- K-Neighbors Classifier – KNN
- Supporting Vectors – SVC
- Random Forest Classifier – RF
- Decision Tree Classifier – DT
- Gaussian NB – NB

El resultado se muestra en el gráfico a continuación, dónde se puede apreciar que claramente el peor modelo de clasificación en el NB. Hay además que tener en cuenta que las categorías en las que clasificar las observaciones son 2, y están distribuidas en un 60% y en un 40% respectivamente.



LR: 0.748805 (0.018787)
LDA: 0.748469 (0.019112)
GB: 0.726995 (0.023057)
KNN: 0.622840 (0.022355)

SVC: 0.748435 (0.020190)
 RF: 0.663582 (0.015946)
 DT: 0.594665 (0.030048)
 NB: 0.021080 (0.027236)

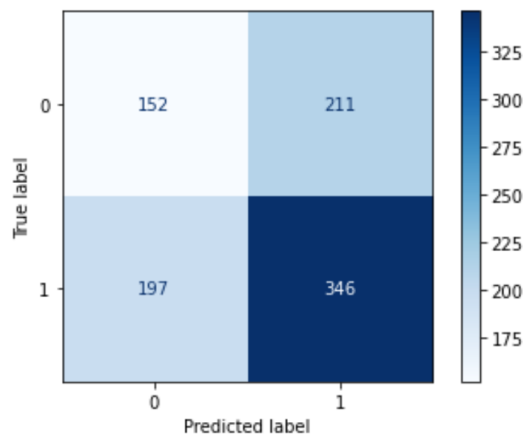
Tras haber realizado este primer análisis, llevamos a cabo la separación de los datos entre los datos de entrenamiento y aquellos que servirán para testear la efectividad del modelo. En este caso se ha optado por dejar un 30% de los datos para el test, y un 70% para construir el modelo de clasificación. Habiendo evaluado la eficacia a priori de los distintos modelos, se ha optado por parametrizar y completar un modelo utilizando el análisis discriminante lineal (LDA) y otro utilizando un modelo de potenciación del gradiente.

4.2.1 Método del Random Forest

Este método de clasificación funciona mediante la aplicación de numerosos decision trees que operan en conjunto. Cada decision tree individual predice una clasificación para cada observación, tras lo cual el algoritmo realiza una media de las estimaciones, siendo aquella clasificación que ha obtenido el mayor número de votos aquella que se acaba escogiendo como observación predicha para cada observación (Yiu, 2019).

Tras haber realizado la parametrización del modelo de random forest con respecto al número de hojas, la profundidad, el número de estimadores y el mínimo de observaciones por hoja y split, los resultados son los siguientes:

	precision	recall	f1-score	support
0.0	0.44	0.42	0.43	363
1.0	0.62	0.64	0.63	543
accuracy			0.55	906
macro avg	0.53	0.53	0.53	906
weighted avg	0.55	0.55	0.55	906



Los resultados de la utilización de un modelo de Random Forest no son los idóneos, pudiendo significar la inexistencia de efectos reales de las variables utilizadas para la predicción del comportamiento positivo o negativo de la acción de una empresa comparable tras la publicación de resultados de una empresa concreta. En cuanto a la efectividad del modelo, de las 543 observaciones que obtienen retornos positivos, ha logrado identificar correctamente 346, a la vez que ha clasificado correctamente 152 de las 363 observaciones con retornos positivos. Estos resultados están por lo tanto en línea con la hipótesis base de que aproximadamente el 60% de las observaciones obtienen retornos positivos y el 40% retornos negativos.

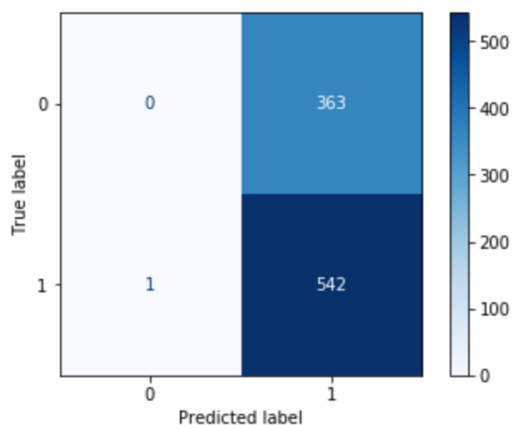
En cuanto a los detalles del modelo, se encuentran en el **Anexo III**.

4.2.2 Análisis discriminante lineal

Este método funciona mediante la aplicación del teorema de Bayes para estimar con qué probabilidad debería una nueva observación pertenecer a una u otra clasificación de entre aquellas que han sido reconocidas a priori (Amat, 2016).

Tras haber realizado una parametrización del análisis discriminante lineal, los resultados del modelo son los siguientes:

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	363
1.0	0.60	1.00	0.75	543
accuracy			0.60	906
macro avg	0.30	0.50	0.37	906
weighted avg	0.36	0.60	0.45	906



Tal y como se puede observar, el modelo ha optado por clasificar como observaciones con retornos positivos en el corto plazo de las empresas comparables 905 de las 906 observaciones. Este es por lo tanto un resultado que no aporta información en relación a la hipótesis base de que aproximadamente el 60% de las observaciones contienen un retorno positivo en las empresas comparables en los diez días próximos (esencialmente, esto se puede entender tal y como se ha explicado anteriormente por la tendencia históricamente alcista del mercado).

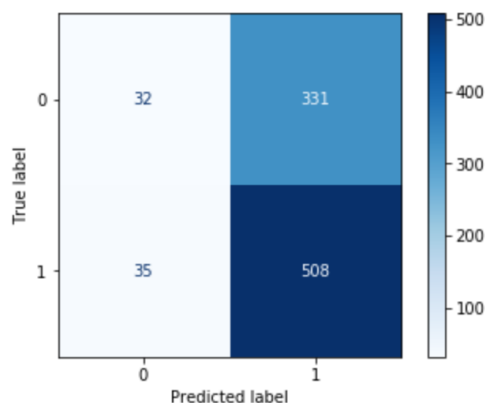
En cuanto a los detalles del modelo, se encuentran en el **Anexo IV**.

4.2.3 Potenciación del gradiente

Este método se basa en los métodos de ensamblaje de modelos, en concreto en los métodos de *boosting* (otra clase de ensamblaje de modelos es la de *bagging*, utilizada por ejemplo en el Random Forest, y basándose en la unión de modelos entrenados con diferentes conjuntos de datos y de variables, lo cuál logra obtener modelos escasamente correlacionados). El método de *boosting* en el contexto del ensamblaje de modelos consiste en el entrenamiento de un modelo, el estudio de los elementos o muestras en los que falla, y volver a entrenar el modelo centrándose en reforzar aquellas muestras en las que erraba (Ke, y otros, 2017).

Los resultados del modelo de potenciación del gradiente son los siguientes:

	precision	recall	f1-score	support
0.0	0.48	0.09	0.15	363
1.0	0.61	0.94	0.74	543
accuracy			0.60	906
macro avg	0.54	0.51	0.44	906
weighted avg	0.55	0.60	0.50	906



Tal y como se puede apreciar en el gráfico de arriba, los resultados, aunque mejores que aquellos obtenidos por el método del análisis discriminante lineal, no son óptimos. El modelo ha clasificado correctamente como retornos positivos 508 observaciones, mientras que ha clasificado correctamente como retornos a corto plazo negativos 32 observaciones. Sin embargo, también ha errado en 331 observaciones a las que ha clasificado como retornos positivos cuándo los retornos eran en realidad negativos, y 35 como retornos negativos cuándo eran en realidad positivos. En definitiva, ha clasificado correctamente aproximadamente un 60% de las observaciones, de nuevo en línea con la hipótesis base de que en torno al 60% de las observaciones tienen retornos positivos. Si bien esto es cierto, también es necesario destacar que de aquellas con retornos positivos, el modelo ha sabido identificar correctamente un 94% de las observaciones.

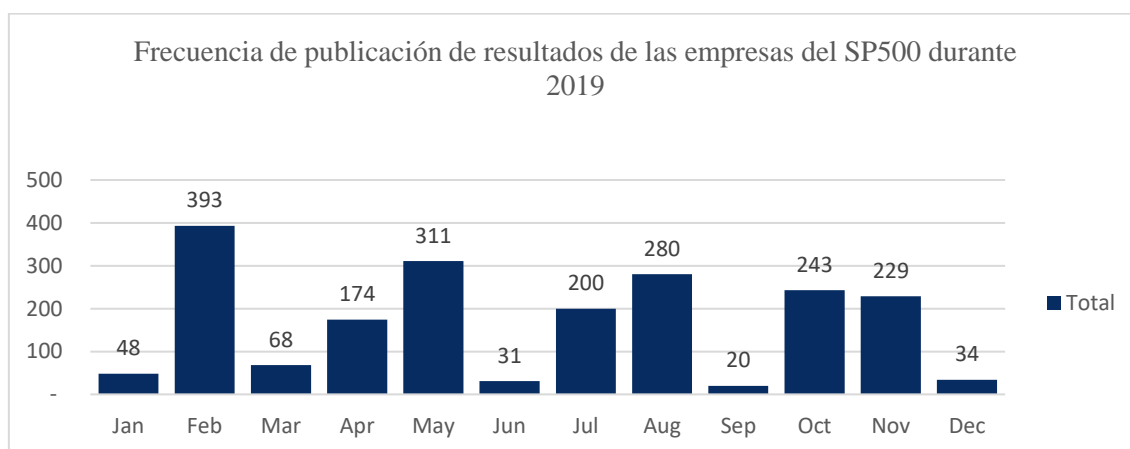
En cuanto a los detalles del modelo, se encuentran en el **Anexo V**.

Capítulo 5. Futuras líneas de investigación y limitaciones del estudio

A lo largo de este análisis, se han detectado una serie de elementos que han imposibilitado la obtención de un resultado más conclusivo en relación al efecto en la cotización de empresas comparables de la publicación del estado financiero trimestral de una empresa observada.

5.1 Interrelación de las presentaciones de resultados

Este análisis no está teniendo en cuenta el hecho de que varias empresas comparables entre sí publiquen resultados en unas fechas parecidas. Así se ve por ejemplo en el siguiente gráfico que la mayoría de las empresas publican resultados durante febrero, mayo, agosto y octubre. Esto implicará que en numerosas ocasiones las publicaciones de resultados de empresas comparables se encontrarán próximas en el tiempo y por lo tanto distorsionará los resultados, en tanto que el mercado ponderará como información más valiosa aquella que atañe directamente a la empresa observada, por encima de aquella que afecte marginalmente a la empresa por provenir de una empresa comparable.



5.2 Existencia de elementos distorsionadores de los resultados

Tal y como se ha explicado anteriormente, en las series temporales el ruido blanco enturbia los resultados, resultando en divergencias entre los valores esperados y los valores observados. En este caso se podría entender como ruido blanco todas aquellas ocurrencias que no son explicadas por la publicación de resultados de las empresas comparables y la existencia o inexistencia de sorpresas con respecto a las predicciones de los analistas. Como es razonable, la mayor parte de

los movimientos de la cotización de las acciones viene explicado por factores ajenos a la publicación de resultados de la empresa comparable.

5.3 Comparabilidad de las empresas

Otro de los factores que habremos de tener en cuenta es la comparabilidad entre las empresas. Tal y como se ha explicado anteriormente, se ha simplificado el problema de la clasificación de las empresas como comparables entre sí mediante la agrupación de empresas que pertenecen a una misma sub-industria. Si bien es cierto que esta es una aproximación a la comparabilidad y que habiendo más de 120 sub-industrias en el SP500 se logra una gran atomización de los grupos, de cara a futuros análisis el perfeccionamiento de los grupos creados habría sido sin duda beneficioso para el trabajo que nos atañe.

Una forma de haber refinado la selección de empresas comparables habría sido el empleo de listas de empresas comparables utilizando herramientas como CapitalIQ, sin embargo dichas herramientas son costosas y su utilización está limitada a empresas dispuestas a pagar el coste que suponen mayoritariamente, como es el caso de los bancos de inversión. Cabe además destacar que la utilización de CapitalIQ podría haber igualmente facilitado el análisis al proveer todas las variables que hemos estudiado de una manera cómoda mediante un add-in sobre Excel; inclusive la cotización en cada fecha concreta de las empresas estudiadas.

La otra forma, sin duda más complicada, habría sido la elaboración de agrupaciones de empresas comparables mediante la selección y agrupación individual de las empresas, lo cuál habría sido no ya muy costoso en dinero sino en tiempo.

Finalmente, se podrían haber elaborado métodos mixtos de agrupación de empresas, como lo habría sido la ponderación de varios factores para asumir la comparabilidad de las empresas, como por ejemplo la existencia de un comportamiento similar en el mercado (estudiado a través del análisis de la cotización histórico y la existencia de correlaciones), o la pertenencia a mismas industrias y la tenencia de unas similares capacidades de generación de flujos de caja, de crecimiento y de características de riesgo intrínseco.

En definitiva, todas estas alternativas habrían sin duda enriquecido el análisis que nos atañe, pero habrían sido más costosas en esfuerzo computacional, económico y/o humano. De todas formas, resulta conveniente su enumeración de cara a futuros análisis de la materia y a la apertura de nuevas líneas de investigación.

5.4 Inexistencia de información de las razones subyacentes al *earnings surprise*

El estudio agnóstico de los ingresos netos por acción y su comparación con las predicciones de los analistas no considera la razón por la cuál se dan esos resultados. Esta información es de gran importancia para entender el comportamiento que podrá tener el mercado en respuesta a los nuevos datos. Es relevante, tal y como se explicaba en la introducción de este trabajo, porque de ser intrínseca a las operaciones de la empresa que publica resultado, unos buenos resultados pueden no significar una futura bonanza en el sector y por ende en las empresas comparables. En cambio, si los resultados responden a un cambio generalizado en el mercado o en el consumidor, éstos sí serán reveladores. El análisis pormenorizado del sector y de las cuentas publicadas nos aportará la información necesaria para entender los resultados y su conocer su significado.

5.5 Tratamiento de las predicciones de los analistas

Además de haber estudiado las predicciones de los analistas como conjunto, utilizando métricas de media y mediana para analizar la información del mercado sobre los resultados de una empresa en comparación con los valores reales obtenidos, habría sido interesante segmentar y ponderar dichas predicciones.

Una forma de haber ponderado por importancia dichas predicciones habría sido por ejemplo segmentar las predicciones por fechas, haciendo n grupos en función de la cercanía de la predicción a la fecha real de publicación de resultados. De esta forma se habría asignado un mayor valor a aquellas predicciones que fueran más cercanas en el tiempo a la publicación de resultados dado que se puede asumir que en primer lugar la información de la que se dispondría sería más amplia según se acercase la fecha de publicación, y que en segundo lugar el mercado establecería eficientemente sus precios en función de la información más reciente.

Podría haber sido igualmente interesante analizar las propias predicciones de los analistas. Dado que la base de datos IBES nos proporciona detalles sobre de dónde proviene cada predicción, se podría haber estudiado qué bancos suelen acertar más en sus estimaciones. Igualmente habría sido curioso estudiar si algunos bancos tienden a sistemáticamente ser más optimistas o pesimistas que la realidad con respecto a sus valoraciones.

Con todo esto, se habría logrado añadir valor a las bases de datos y a la información. Igualmente, mediante la ponderación de las predicciones se podría haber logrado hacer que la variable de las predicciones fuese más valiosa en tanto que se puede esperar teóricamente que el mercado

realizase sus valoraciones con las predicciones más recientes y no con aquellas que tuvieran una mayor antigüedad.

5.6 Inclusión de medidas tendenciales

Tal y como se ha explicado anteriormente, la existencia de retornos positivos o negativos en los diez días posteriores a la publicación de resultados dependerá enormemente de factores externos a la publicación de resultados de la compañía comparable. Es por ello por lo que podría parecer sensato que en futuras investigaciones se incluyesen medidas que explicasen la tendencia existente en el mercado, sector, y en particular en las propias empresas comparables. Por ejemplo, se podría incluir una medida que comunicase cuáles han sido los resultados en la cotización de las empresas comparables en los últimos meses, días e incluso años.

Si bien es cierto que con seguridad estas medidas ayudarían a refinar el modelo y por ende obtener unos mejores resultados en la clasificación, también pasarían a enturbiar el objeto de este trabajo. Al fin y al cabo, el objetivo principal de este trabajo ha sido el estudio del efecto en la cotización de las empresas comparables de la publicación de resultados de una empresa observada. Es por ello que el estudio habría de centrarse de manera agnóstica del resto de circunstancias del mercado en el estudio de la propia publicación de resultados.

En definitiva, dado que la razón principal de realizar este trabajo ha sido el estudio del efecto de la actualización de la información en posesión del mercado con relación a los resultados operativos de una empresa, se ha creído conveniente no incluir variables adicionales que pasasen a explicar los retornos obtenidos con información adicional.

Capítulo 6. Conclusiones

En definitiva, los problemas anteriormente mencionados lastran y enturbian los resultados obtenidos mediante este análisis. La tenencia de unas mayores bases de datos, que además hubieran sido más fácilmente combinables entre sí habría facilitado la labor realizada, sin embargo cabe destacar que estas bases de datos son tan valiosas como caras, lo cuál limita la capacidad de usarlas.

A pesar de todo esto, ha resultado un ejercicio interesante el estudio de este problema, y la respuesta a la pregunta de si influye a la cotización de las empresas la publicación de resultados de empresas comparables se ha logrado responder parcialmente. Al fin y al cabo, se ha logrado mostrar que la existencia de factores distintos a la publicación de resultados acabará ponderando con mayor fuerza en el mercado y en los participantes de él a la hora de establecer sus valoraciones y precios.

Es además destacable el hecho de que la mayoría de empresas publican resultados en épocas cercanas, con lo cuál el análisis de la cotización en los 10 días posteriores podría no haber sido la más adecuada. Tal vez lo propio y correcto habría sido estudiar el efecto de la nueva información en la cotización y volúmenes intra diarios, haciendo una análisis al minuto de la cotización de las empresas comparables. Sin embargo, la inaccesibilidad de estos datos para un particular hace que este análisis sea inviable.

Finalmente, cabe destacar la utilidad que se obtiene de la utilización del código presente en el **Anexo I** para la extracción de las fechas de publicación de las empresas americanas mediante la utilización de los Central Index Keys y las capacidades del *webscrapping*. Es esta una herramienta que en el futuro podrá utilizarse para ampliar y profundizar en el análisis del efecto de la publicación de resultados y los *earnings surprises* en la cotización de las empresas, todo ello sin tener que depender de la utilización de bases de datos de difícil o costoso acceso.

Capítulo 7. Anexo

Anexo I. Código de Python mediante el cual se obtienen las fechas en las cuáles han publicado resultados las empresas del SP500:

```
headers = {Credenciales correspondientes al ordenador empleado}
def get_soup(url, headers):
    """
    returns ...
    """
    page = requests.get(url, headers=headers)
    result = requests.get(url, headers=headers)
    soup = BeautifulSoup(requests.get(url, headers=headers).text, "lxml")
    return soup
headings = ['Filings', 'Format', 'Description', 'Filing Date', 'File/Film Number', 'ticker', 'cik']
def get_web_df(maps, headers, headings):
    all_rows=[]
    for ticker, cik in tqdm(maps.items()):
        url=f'https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK={cik}&type=10-&dateb=&owner=exclude&count=100'
        soup = get_soup(url, headers)
        table = soup.find_all("table", attrs={"class": "tableFile2"})[0]
        # the head will form our column names
        body = table.find_all("tr")
        # Head values (Column names) are the first items of the body list
        body_rows = body[1:] # All other items becomes the rest of the rows
        for table_row in body_rows:
            row = [row_item.text.strip() for row_item in table_row.find_all("td")]
            row += [ticker, cik]
            all_rows.append(row)
    df = pd.DataFrame(data=all_rows, columns=headings)
    df['Filing Date'] = pd.to_datetime(df['Filing Date'].values, format='%Y-%m-%d')
    return df
df = get_web_df(maps_sp500, headers, headings)
df.head()
```

Anexo II. Código de Python mediante el cuál se obtienen los datos de cotización de las empresas comparables en relación a la fecha de publicación de la empresa observada:

```
#Crear la columna PeersYield
CompleteData['PeersYield']=""/>
#Calcular el rango de filas de los datos y el rango de columnas de Peers
RowsRange=range(0, len(CompleteData))
PeerColRange=range(9, 34) #Peers columns
PriorBusinessDays=pd.tseries.offsets.BusinessDay(n = -10)
NextBusinessDays=pd.tseries.offsets.BusinessDay(n = 10)
#Establecer el calendario sobre el que comprobaremos si hay fiestas nacionales
Calendar = USFederalHolidayCalendar()
#Inicializar las listas de variables
PeersYield=[]
ErrorHandler=[]
for x in RowsRange:
    #Seleccionar la fecha de publicación de resultados para la fila x
    FillingDate=CompleteData.iloc[x][3]
    #Establecer el número de días previos y posteriores a la publicación de resultados que se van a solicitar en Yahoo
    PriorRange=-10
    NextRange=10
    def HolidayChecker(m,n,x):
        #Establecer los rangos cumulativos basados en días festivos
        PriorBusinessDays=pd.tseries.offsets.BusinessDay(n = PriorRange+m)
        NextBusinessDays=pd.tseries.offsets.BusinessDay(n = NextRange+n)
        #Fechas
        PriorDate=FillingDate+PriorBusinessDays
        NextDate=FillingDate+NextBusinessDays
        #Contar el número de días festivos
        PriorHolidays=0
        NextHolidays=0
        #Comprobar si hay días festivos en el rango de días previo a la publicación y de haberlos contarlos
        for dt in Calendar.holidays(start=PriorDate, end=FillingDate):
            PriorHolidays-=1
        #Comprobar si hay días festivos en el rango de días posterior a la publicación y de haberlos contarlos
        for dt in Calendar.holidays(start=FillingDate, end=NextDate):
            NextHolidays+=1
        #Llamar de forma recursiva a la función en caso de haber días festivos no tenidos en cuenta
        if (PriorHolidays!=0 or NextHolidays!=0) and (PriorHolidays!=m or NextHolidays!=n):
            HolidayChecker(PriorHolidays,NextHolidays,x)
    #Utilizar el rango de días previos y posteriores correcto para la extracción de las cotizaciones de los Peers
    else:
        print(PriorDate, NextDate)
        STRING=""
        Yields=[]
        for y in PeerColRange:
            PeerTicker=CompleteData.iloc[x][y]
            #Añadir el valor de las columnas Peers a la variable STRING en caso de ser un Ticker
            if PeerTicker!= 'Null':
                STRING=STRING+" "+CompleteData.iloc[x][y]
            #Extraer the Yahoo los precios previos para las fechas útiles para todos los Peers
            PriorDateRangeStock = yf.download(STRING, start=PriorDate, end=FillingDate, progress=False)
            #Extraer the Yahoo los precios posteriores para las fechas útiles para todos los Peers
            NextDateRangeStock = yf.download(STRING, start=FillingDate, end=NextDate, progress=False)
            #Crear una lista de tickers a partir del string de tickers para su iteración
            PeersList=STRING.split(' ')
            #Calcular el rendimiento de cada Ticker para el rango de días útiles posterior y previo
            #Se usa la mediana para evitar las desviaciones
            for z in PeersList:
                if z!="":
                    StockYield=NextDateRangeStock['Close'][z].median()/PriorDateRangeStock['Close'][z].median() -1
                    Yields.append(StockYield)
            #Extraer la mediana de los rendimientos de los Tickers y adjuntar el valor a la lista
            MedianYield=statistics.median(Yields)
            print('x: ', x, 'MedianYield: ', MedianYield)
            PeersYield.append(MedianYield)
            #CompleteData['PeersYield'][x]=MedianYield
try:
    HolidayChecker(0,0,x)
except:
    #Registrar errores en caso de que Yahoo finance no tenga datos para el rango de fechas propuesto
    ErrorHandler.append(x)
    print('x: ', x, 'Error handled')
    PeersYield.append('Error')
    continue
```

Anexo III. Modelo de Random Forest:

```
#Definimos el modelo.
RF_model_param=RandomForestClassifier()

#Definimos combinaciones de parámetros.
parameters = {
    'class_weight':['balanced', 'balanced_subsample'],
    'max_depth': list(range(2, 7)),
    'min_samples_leaf': [1, 2, 4, 6],
    'min_samples_split': [1, 2, 5, 7],
    'n_estimators': [25, 50, 100, 150]}

#Ajustar el modelo en este caso con RandomizedSearchCV.
#En contraste con GridSearchCV, no se prueban todos los valores de los parámetros, el número de ajustes de parámetros es n_iter.

RF_model_param = GridSearchCV(estimator=RF_model_param,
                               param_grid=parameters,
                               cv=8,
                               verbose=5,
                               n_jobs=-1,
                               scoring='precision'
                               )

# Fit the random search model.
RF_model_param.fit(X_train, y_train)

#predicciones sobre datos de prueba.
y_pred=RF_model_param.predict(X_test)

best_random = RF_model_param.best_estimator_
best_random.fit(X_train, y_train)
y_pred = best_random.predict(X_test)

#Dibujamos el los resultados. Interpretarse en base al apartado 3.1) Teoría para la interpretación del modelo.
print(classification_report(y_test, y_pred))
disp = plot_confusion_matrix(best_random,
                             X_test,
                             y_test,
                             display_labels=[0, 1],
                             cmap=plt.cm.Blues,
                             )
```

Anexo IV. Análisis discriminante lineal:

```
#Definimos el modelo.
LDA_model_param = LinearDiscriminantAnalysis()

parameters = {
    #svd es "singular value decomposition". Se utiliza para ocasiones con muchos datos
    #lsqr es "least square solution"
    #eigen proviene de "eigenvalue decomposition"
    'solver':['svd', 'lsqr', 'eigen'],
    #Define si hay o no contracción
    'shrinkage':['auto', None],
    #Número de componentes para la reducción de la dimensionalidad.
    'n_components':[1,2,3,4,5,6,7,8,9,10,11],
    #Para computar o no la matriz de covarianza.
    'store_covariance':[True, False]
}

#Ajustamos el modelo con los parámetros.
LDA_model_param = GridSearchCV(estimator=LDA_model_param, param_grid=parameters, cv=10, scoring='precision')
LDA_model_param.fit(X_train, y_train)

#Predicciones sobre datos de prueba.
y_pred=LDA_model_param.predict(X_test)

#Dibujamos el los resultados. Interpretarse en base al apartado 3.1) Teoría para la interpretación del modelo.
print(classification_report(y_test, y_pred))
disp = plot_confusion_matrix(LDA_model_param,
                             X_test, y_test,
                             display_labels=[0, 1],
                             cmap=plt.cm.Blues
                             )

#Mejores parámetros.
LDA_model_param.best_params_

#Creación del LDA con los mejores parámetros.
LDA_param= LinearDiscriminantAnalysis(n_components=1,
shrinkage=None,
solver='svd',
store_covariance=True)

#Validación cruzada
LDA_results = cross_validate (LDA_param, X_train, y_train, cv=10, scoring=scores)
```

Anexo V. Modelo de Potenciación del Gradiente:

```
#Definimos el modelo.
GB_model_param = GradientBoostingClassifier()

parameters = {
    #La tasa de aprendizaje "hunde" la contribución de cada árbol.
    #Por lo que por un lado tenemos el número de estimadores y por otra la tasa de aprendizaje.
    'learning_rate':[0.1, 0.15],
    #Número mínimo de muestras de cada nodo de hoja
    'min_samples_leaf':np.linspace(0.1, 0.5,4),
    #Optimización de la función de pérdida. el deviance es la logística y el AdaBoost la exponencial.
    'loss':['deviance'],
    #Mide el número mínimo de muestras requeridas en cada nodo internamente- Debe ser ~1% del total de observaciones.
    'min_samples_split':[1,2,5],
    #Profundidad máxima de la regresión individual de los estimadores.
    'max_depth':[1,2,3,-1],
    #Número de gradientes a utilizar. Es robusto en general con el sobreajuste así que un número alto suele dar buen resultado.
    'n_estimators':[5,10,20]
}

#Ajustamos el modelo con los parámetros.
GB_model_param = GridSearchCV(estimator=GB_model_param, param_grid=parameters, cv=10, scoring='precision', n_jobs=-1)
GB_model_param.fit(X_train, y_train)

#Predicciones sobre datos de prueba.
y_pred=GB_model_param.predict(X_test)

#Dibujamos el los resultados. Interpretarse en base al apartado 3.1) Teoría para la interpretación del modelo.
print(classification_report(y_test, y_pred))
disp = plot_confusion_matrix(GB_model_param,
                             X_test, y_test,
                             display_labels=[0, 1],
                             cmap=plt.cm.Blues
                             )

# Mejores parámetros.
GB_model_param.best_params_

#Creación del GB con los mejores parámetros.
GB_model_param= GradientBoostingClassifier(learning_rate=0.15,
loss='deviance',
max_depth=3,
min_samples_leaf=0.1,
min_samples_split=2,
n_estimators=20
)

#Validación cruzada
GB_results = cross_validate (GB_model_param, X_train, y_train, cv=10, scoring=scores)
```

Capítulo 8. Bibliografía

- ACI Insights. (2018). *Airport markets and seasonal variations*.
- Alden, L. (2022). *What is Money, Anyway?*
- Alzahrani, A., & Skerratt, L. (2009). How Markets React to Earnings Announcements in the Absence of Analysts and Institutions Evidence from the Saudi Market.
- Amat, J. (2016). *Análisis discriminante lineal (LDA) y análisis discriminante cuadrático (QDA)*.
Obtenido de Ciencia de Datos.
- Arif, S. (2016). Financial Reporting Frequency and Earnings Information Spillovers around the World.
- Buffet, W. E. (2021). *Warren Buffett's 2021 Letter to Berkshire Shareholders*.
- Damodaran, A. (2011). *The Little Book of Valuation: How to Value a Company, Pick a Stock and Profit*.
- Damodaran, A. (s.f.). *The Time Value of Money*.
- Dimson, E., Marsh, P., & Staunton, M. (2003).
- Dodonov, V. (2020). *Effect of Financial Statement Release on Stock Prices*.
- Dorsey, P. (2008). *The little book that builds wealth*.
- European Central Bank. (2020). *How news affects sectoral stock prices through earnings expectations and risk premia*.
- Fama, E. (1970). *Efficient Capital Markets: a review of Theory and Empirical work*.
- Fama, E., & French, K. (2014). *A five-factor asset pricing model*.
- Fatmasari, S., Imam, G., F. F., Abdul Kharis, A., & Nurcahyono, N. (2020). *Factors Affecting the Stock Price: The Role of Firm Performance*.
- Financier Worldwide Magazine. (2017). A quick review of the literature regarding the 'small-cap premium'.
- Firth, M. (1976). *The Impact of Earnings Announcements on the Share Price Behaviour of Similar Type Firms*.
- Indian Journal of Finance. (2013). *Impact of quarterly results on share prices*.

- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). *LightGBM: A Highly Efficient Gradient Boosting*.
- Marks, H. (2020). *Coming into Focus*.
- Marks, H. (2022). *Bull Market Rhymes*.
- Mohamed, M. H. (2010). *The effect of the earnings announcements on the stock prices of companies listed at the Nairobi Stock Exchange*.
- Morgan Stanley Investment Management. (2021). *Everything Is a DCF Model, A Mantra for Valuing Cash-Generating Assets*.
- Paul, R., & Richard J., T. (2004). – *Are Economically Significant Stock Returns and Trading Volumes Driven by Firm-specific News Releases? .*
- Refinitiv. (2022). *An overview of I/B/E/S Estimates .*
- Reik, T. (1965). *The Unreachables*.
- Stankevi, J., & Simas, A. (2014). *Impact of public announcements on stock prices: relation between values of stock prices and the price changes in Lithuanian stock market*. Riga.
- Vincent, G., & Charles, M. (2021). *How is earnings news transmitted to stock rpices*.
- Wharton School of the University of Pennsylvania. (s.f.). *Linking IBES and CRSP data*.
- Yiu, T. (2019). *Towards Data Science*.