



Facultad, Escuela o Instituto

EL NEGOCIO DE LOS DATOS EL IMPACTO DEL BUSINESS ANALYTICS EN LA EMPRESA

Autor/a: Ignacio Lopez Rey
Director/a: Mahmoud Aymo

MADRID | Junio 2020

ÍNDICE

I.	RESUMEN EJECUTIVO.....	3
II.	PARTE I: LA ERA DE LOS DATOS	
	1. Introducción al dato.....	4
	2. Beneficios de los datos.....	4
	3. La ética del negocio de los datos.....	5
	4. Marco jurídico: Regulación.....	6
III.	PARTE II: BUSINESS ANALYTICS EN LA EMPRESA	
	5. Definición de business analytics.....	8
	6. Evolución del business analytics y la empresa.....	9
	7. Big Data.....	11
	a. Netflix.....	12
	b. Verisk Analytics.....	13
	c. AWS, el futuro de Amazon.....	14
	8. Análisis de business analytics en la empresa	
	a. Investigación.....	15
	b. Metodología.....	17
	c. Fuentes, datos y variables empleadas.....	18
	d. Análisis descriptivo.....	23
	e. Análisis exploratorio.....	26
IV.	CONCLUSIONES Y DESARROLLO FUTURO.....	30
V.	BIBLIOGRAFÍA.....	32
VI.	ANEXOS.....	35

Resumen Ejecutivo

Las últimas décadas han sido testigo del surgimiento de una nueva forma de tomar decisiones en el ámbito empresarial. La irrupción de los datos en el panorama ejecutivo ha alterado profundamente la manera en la que las empresas llevan a cabo sus operaciones. El *business analytics* es el conjunto de procesos y herramientas de colección y análisis de datos destinados a facilitar la toma de decisiones en el entorno empresarial. A pesar de no ser un concepto nuevo, el análisis de datos para la toma de decisiones es un tema del que todavía queda mucho por aprender. No obstante, el *business analytics* ha experimentado una profunda evolución en los últimos años. Muchas empresas a lo largo del mundo emplean el análisis de datos a la hora de la toma de decisiones tanto a nivel operativo como estratégico. Este profundo interés por los datos ha propiciado que este mercado sea uno de los puntos focales para órganos reguladores y las propias empresas. El negocio de los datos está en auge. Hoy más que nunca, los CEO de las grandes empresas se basan en los datos proporcionados por el *business analytics* para la toma de decisiones a corto y largo plazo. Esto ha generado situaciones de desequilibrio debido a las grandes posibilidades de recopilación de las grandes multinacionales frente a aquellas de menor tamaño.

En este trabajo se investigará cómo el uso de datos en la empresa ha evolucionado a lo largo de los años, se analizarán los abusos de poder (Cambridge Analytica) y su regulación. A continuación, se estudiará en detalle específicamente el campo del *business analytics* y se analizará el impacto que tiene en la empresa, poniendo numerosos casos como ejemplo. Entre ellos, se explorará cómo empresas como Netflix, Verisk o Amazon han implementado el *business analytics* en sus operaciones y cómo han sido beneficiadas por ello. Finalmente, se llevará a cabo un análisis de varias empresas que hayan incorporado el *business analytics* a sus operaciones (bien como departamento o como parte de cualquier otro departamento). El objetivo de este análisis es el de comprobar que el *business analytics* contribuye al incremento de ingresos en la empresa. Para esto, se empleará una regresión lineal cuya variable dependiente son los ingresos. Este modelo indicará, en base a una serie de variables independientes (grado de implementación, por ejemplo) la cantidad en la que varían los ingresos por ventas de la empresa por cada unidad invertida en *analytics*.

Palabras clave: toma de decisiones, datos, business analytics, rendimiento

PARTE I: LA ERA DE LOS DATOS

1. Introducción al dato

Cómo dirigir una empresa de manera efectiva nunca ha sido una ciencia exacta. La toma de decisiones en el entorno de la empresa siempre se ha llevado a cabo apoyándose en información objetiva, pero siempre confiando en la experiencia e intuición. De acuerdo con un estudio llevado a cabo por Accenture, el 40% de las grandes decisiones dentro de una empresa se toman en base a la intuición (Vujanic, 2008). No obstante, la década de los 60 trajo consigo la introducción de una revolución en la dirección y gestión de empresas que todavía hoy contribuye enormemente al crecimiento económico y la transformación de la sociedad (COTEC, 2017). ARPANET, un proyecto constituido por una red de ordenadores distribuidos geográficamente para mantener el flujo constante de información entre bases militares se convertiría, más tarde, en la base bajo la cual se planteó Internet, lo que inició la revolución de los datos (Craig & Ludloff, 2011). Esta revolución está siendo caracterizada mayoritariamente por la disponibilidad y acceso ilimitado de datos, propulsada gracias a la llegada de Internet (Fernandez-Manzano, 2016).

El dato se ha convertido para muchos en el “petróleo del siglo XXI.” De acuerdo con Deloitte, “un minuto de rastro virtual equivale a millones de datos” (COTEC, 2017). Un dato corresponde a cada uno de los valores y hechos derivados de la actividad de individuos o empresas. Este concepto no es nuevo, ya que siempre se ha acudido a los datos para conocer mejor las propias operaciones de la empresa, cómo desarrollar sus productos o servicios de manera eficiente y para estudiar a los consumidores. Sin embargo, la llegada de la computación ha permitido no solo automatizar el proceso de generación de datos, sino también la recopilación de estos. El 90% de los datos en el mundo se han generado en los últimos diez años. Es más, se estima que empresas como Walmart gestionan más de un millón de transacciones cada hora, el equivalente a 2,5 petabytes, más de 1×10^{12} la cantidad de datos que se necesitó para viajar a la luna (Asllani, 2014). Lo cierto es que esta tendencia creciente es exponencial. De acuerdo con la Ley de Moore, la cantidad de datos existente se dobla cada dos años, lo que indicaría que la cantidad de datos disponibles será incluso mayor en los próximos años (Asllani, 2014). Este hecho pone de relieve la importancia del manejo y análisis de datos y garantiza su estabilidad como fuente de información vital.

2. Beneficios de los datos

Todas aquellas empresas que han apostado por la recopilación de datos desde sus comienzos han podido apalancar esta actividad como ventaja competitiva respecto al resto. Estas son

empresas pioneras, tales como Google, Microsoft o Netflix, que comenzaron a invertir en estos recursos como actividad propulsora y que ahora pueden ver cómo esa inversión da sus frutos. De acuerdo con un informe publicado por la agencia de comunicación IDC, los ingresos directamente derivados del manejo de datos habrán crecido en más de sesenta mil millones de dólares en menos de cuatro años, el equivalente a un incremento del 50% (COTEC, 2017). Numerosos estudios ponen de relieve tan solo algunas de las ventajas que una política de datos abiertos podría suponer tanto a nivel social como económico. Un estudio llevado a cabo por Omidyar Network señalaba que una política de datos abiertos supondría un aumento de los ingresos registrados por los países del G20 de entre 700 y 950 mil millones de dólares, así como una mayor transparencia, mejores condiciones de trabajo, mayor eficiencia energética y menores barreras de entrada para emprendedores y pequeñas empresas. De acuerdo con este mismo estudio, la ciudadanía se beneficiaría enormemente del uso de datos mediante la optimización del tráfico, la mejora de la sanidad o incluso consumo personalizado.

En el ámbito empresarial, existe una amplia gama de posibilidades de mejora tanto en el campo operacional como en el financiero. La colección de datos da acceso a herramientas para un mejor entendimiento de las propias mecánicas de la empresa, tales como entender por qué una empresa desempeña de una manera u otra en función de la situación económica y cambios en el mercado. Los datos abren un abanico de posibilidades a la empresa, las cuales pueden emplear métodos estadísticos para llevar a cabo inferencias en base a una pequeña muestra. El análisis de datos puede permitir conocer qué medidas están teniendo un impacto real en la empresa y qué medidas empujan en dirección contraria a la visión propuesta. El análisis de datos permite recortar gastos y potenciar la eficiencia mediante la optimización de técnicas que minimicen el consumo de activos o mediante el uso de modelos que predigan cambios en el mercado. Asimismo, el análisis de datos puede permitir manejar el riesgo a través del uso de métricas precisas (Davenport et al, 2010).

3. La ética del negocio de los datos

Como ya se ha mencionado con anterioridad, el concepto de dato no es nuevo, a pesar de que la economía ligada a este concepto es una relativamente emergente. Las organizaciones dentro de esta economía serán más o menos exitosas dependiendo de la capacidad de dicha organización para aprovechar las herramientas de gestión y análisis de datos. De acuerdo con Deloitte, “el valor estimado de la economía del dato en Europa alcanzó los 272 mil millones de euros en 2015, lo que representa el 1,87% del PIB de los países miembros y podía ascender al 4,7% en 2020” (COTEC, 2017). Este hecho, junto con la sensibilidad que supone llevar a

cabo esta actividad, pone de relieve diversos problemas en cuanto a la finalidad y ética de las empresas participantes. Efectivamente, el negocio de los datos es una gran oportunidad emergente, pero también corre el riesgo de convertirse en un factor más que contribuya a la desigualdad. En el transcurso de los últimos cinco años, ha sido posible observar este tipo de factor en diversas ocasiones. Este es el caso de Cambridge Analytica, por ejemplo. La joven empresa, establecida en 2013 en Londres, epitomiza la problemática detrás del negocio de los datos. La idea bajo la cual dicha empresa fue fundada era la de proporcionar un servicio único mediante el uso conjunto de herramientas avanzadas de análisis de datos junto al uso de las redes sociales para la recogida de datos. A pesar del enorme potencial detrás de esta empresa, no es demasiado descabellado afirmar que desde el principio se había preparado para el fracaso. Fundada por una de las figuras públicas involucradas en la política estadounidense, Steve Bannon, así como Robert Mercer, multimillonario conocido por aportar fondos para el partido republicano de los Estados Unidos, el objetivo era apalancar sus activos para acceder y manipular los procesos electorales tanto en Reino Unido (con la campaña del Brexit) como en Estados Unidos (en las elecciones generales de noviembre de 2016). Cuando el caso salió a la luz en 2017, se inició una investigación en la empresa, que acabó por clausurarse en 2018. Sin embargo, a pesar de que la organización ya no está activa, se ha comprobado que su influencia sobre los actos electorales mencionados anteriormente fue vital en el resultado final. Por otro lado, la otra gran implicada por permitir el acceso no autorizado a datos de hasta 87 millones de usuarios fue Facebook. No obstante, la conclusión para Facebook fue mucho más positiva, siendo la reprimenda social y la pérdida de usuarios su peor consecuencia (Cadwalladr & Graham-Harrison, 2018).

El negocio de los datos es un fenómeno revolucionario en todos los aspectos. Sin embargo, si estas herramientas no son utilizadas teniendo en cuenta la protección de los derechos de los individuos y estando estas sujetas a un marco regulatorio, es posible que se acentúen los problemas ligados a la desigualdad en lugar de mitigarlos (COTEC, 2017). La Unión Europea tiene en esta nueva economía el papel de asegurar un marco jurídico y unas políticas que garanticen la interoperabilidad, la protección de los datos, la seguridad y los derechos de propiedad intelectual. Estas medidas son las que darán seguridad reglamentaria para las empresas y confianza al consumidor de las tecnologías de datos.

4. Marco jurídico: Regulación

A pesar de la dimensión global de internet, el concepto de “privacidad” varía entre naciones, por lo que las leyes de privacidad y organismos regulatorios varían de país a país (Craig &

Ludloff, 2011). Estados Unidos parece empujar el acceso a información personal en materia pública. Al contrario que en la Unión Europea, actualmente no existe una legislación específica que regule el tratamiento y recogida de datos a nivel nacional. Asimismo, su Constitución no recoge específicamente el término de “privacidad.” En su lugar se han promulgado una serie de leyes que protegen los datos personales a nivel federal y estatal. El Acta Patriota es una ley promulgada en 2001 por George W. Bush, que permite la entrega de toda la biblioteca de registros de un individuo sin el conocimiento de este ya que las solicitudes son secretas (Craig & Ludloff, 2011). La nueva economía alrededor de los datos no dispone en España de iniciativas específicas, articuladas por el Gobierno y apoyadas sobre los principales actores económicos (COTEC, 2017). Sin embargo, en mayo de 2016, la Unión Europea pasó el primer paquete de protección de datos para adaptar a “Europa a la nueva edad digital” (European Union, 2016). La directiva de la Ley General de Protección de Datos, que nace como respuesta a la creciente presión de los europeos para recibir protección sobre el proceso de datos personales, recoge todas aquellas acciones que involucren la recogida y proceso de datos de personas naturales ligadas a procesos criminales. Dicha ley procura que la información personal de víctimas, testigos y sospechosos de crímenes esté protegida. Esta ley también facilitará la cooperación transnacional en la lucha contra el crimen y el terrorismo (European Union, 2016). La Ley General de Protección de Datos ha sido adoptada por todos los países pertenecientes a la Unión Europea, lo que ha afectado a todas aquellas empresas que operan en dichos países. Una cosa es segura, la presión para la creación de leyes que protejan la privacidad de los datos personales está creciendo, y en un mundo dictado por el control y manejo de datos, las organizaciones deben adaptar sus procesos tecnológicos para asegurar que sus actividades cumplen con las nuevas regulaciones (Tabesh et al, 2019).

PARTE II: BUSINESS ANALYTICS EN LA EMPRESA

5. Definición de *business analytics*

El *business analytics*, también conocido como *business intelligence*, se entiende como el conjunto de procesos, dentro del ámbito empresarial, destinados al análisis de datos para la toma de decisiones cuyo objetivo es el de mejorar el desempeño de la empresa en ciertas áreas y entornos de negocio. Este proceso requiere del cotejo, proceso y clasificación de datos a través de distintos modelos estadísticos y metodologías iterativas para la transformación de datos en *insights*. El *business analytics* puede ser incorporado en cualquier área funcional de la empresa tal como marketing, operaciones, finanzas o atención al cliente. Este análisis requiere de dos componentes básicos: analistas de datos y herramientas de *analytics*. Todos aquellos individuos que han completado un grado en estadística, ingeniería informática, matemáticas o temáticas relacionadas, tales como el grado en métodos cuantitativos, pueden llevar a cabo análisis de datos a distintos niveles. Estos profesionales son los encargados de construir sets de datos, diseñar y correr los modelos, interpretar los resultados y visualizarlos. Los conocidos como *data scientists* son “aquellos profesionales que han completado entrenamiento y poseen la curiosidad para llevar a cabo descubrimientos en el mundo del *big data*” (Davenport & Patil, 2012). En cuanto a las herramientas de *data analytics*, pueden variar enormemente entre ellas y ofrecer distintas funcionalidades. Es necesario hacer una distinción entre el software y el lenguaje de programación. Dentro de esta última categoría es común encontrar lenguajes como R, Python o SQL, pero el número de lenguajes es muy vasto y crece por minutos. Por lo general, la diferencia entre lenguajes es mínima, reduciéndose a diferentes nomenclaturas o capacidad de visualización y gestión de los sets de datos. Por otro lado está el software. El software es el programa que lee e interpreta el lenguaje de datos y responde de una manera u otra. Los programas disponibles son numerosos, habiendo empresas que desarrollan su propio software de análisis que se ajusta más a su forma de operar. No obstante, algunos de los programas más utilizados incluyen Microsoft Power BI, SAP Business, SAS, Tableau, Oracle Analytics Cloud o IBM Watson. Empresas como Amazon, Oracle, Microsoft o IBM dedican gran cantidad de sus recursos al desarrollo de herramientas comerciales de *analytics*.

Las posibilidades que brinda el marco de *analytics* son muy amplias, aunque también lo son los posibles desafíos que puede traer consigo. En términos de gestión de empresa, se habla mucho de la facilidad en la toma de decisiones derivada de la disponibilidad y control de datos. Sin embargo, existen una serie de dimensiones y ámbitos dentro del mundo de la gestión de

empresas que serán impactados y será importante tener en cuenta: liderazgo, gestión de talento y adaptación a las nuevas tecnologías. La mejora en la toma de decisiones basadas en *insights* resultantes del análisis de datos genera diversas incógnitas, tales como quién debería de tomar las decisiones o incluso podría llegar a poner en duda la importancia de los altos cargos. Es, por tanto, una necesidad tener un equipo de liderazgo eficiente, con una visión clara que sepa incorporar estos resultados. Las decisiones más importantes con relación a la política de datos de la empresa ha de ser llevada a cabo por la cúpula de liderazgo. Dado que no todos los datos van a cumplir con los criterios de calidad, las decisiones de liderazgo son vitales para el desarrollo óptimo dentro de la empresa (Davenport et al, 2010). Este desafío se pasa por alto habitualmente. En una investigación llevada a cabo por Kaggle en 2017 a más de siete mil *data scientists*, se reveló que cuatro de cada siete barreras tenían que ver con incidencias en el último tramo del proceso y no incidencias técnicas. Algunas de las incidencias incluyen “falta de liderazgo,” “falta de cuestiones a responder” o “resultados no llegados a ser implementados en la solución final” (Berinato, 2019). La incorporación del análisis de datos no elimina el factor humano, sino que enfatiza la necesidad de un componente humano organizado y claro (McAfee et al, 2012). Este hecho es un factor importante a tener en cuenta, especialmente si consideramos a los escépticos que todavía consideran que la implementación de *analytics* es un proceso complejo “que eleva más dudas que resultados” (Germann et al, 2013). Asimismo, cuanto más se incrementa la disponibilidad de datos, más se encarecen los métodos de análisis. El análisis de datos requiere de visualización y entendimiento que permita explicar y acortar la brecha entre correlación y causalidad.

6. Evolución del *business analytics* y la empresa

Como se ha venido mencionando, la evolución del análisis de datos ha sido muy rápida, en una franja temporal relativamente corta. El académico Thomas Davenport defiende la clasificación de *analytics* en tres segmentos, dependiendo en las características del trabajo que se quiere llevar a cabo a través de *analytics*. Cada uno de estos segmentos se basa en el anterior. De esta manera, se pueden distinguir el *analytics* descriptivo, prescriptivo y predictivo (Asllani, 2014). El *analytics* descriptivo se basa estrictamente en un análisis exploratorio de los datos para determinar la situación actual o pasada de la empresa, tanto en general como en un ámbito específico. Este tipo de análisis está generalmente apoyado en la generación de gráficos y otros agentes visuales. En general, las empresas emplean este tipo de análisis para identificar factores de fortaleza y factores de debilidad en la empresa, así como para la identificación de patrones de conducta en los clientes. El análisis predictivo está enfocado en la previsión de eventos

futuros a través de modelos estadísticos y técnicas de *machine learning*. Este tipo de *analytics* está basado en los modelos descriptivos. Su funcionalidad se basa en extrapolar la probabilidad de que un evento actual o pasado se repita en el futuro. El objetivo del *analytics* prescriptivo es el de proveer recomendaciones operacionales que permitan revertir una situación actual o futura. Este tipo de análisis tiene un fuerte componente algorítmico y visual. Las compañías competidoras de *analytics* que ofrecen software a sus clientes se basan generalmente en este tipo de análisis a la hora de proponer recomendaciones en tiempo real.

De esta manera, a pesar de su corta historia, el *business analytics* ha pasado por tres etapas distintas desde su concepción en los años 60. Thomas Davenport describe la etapa de *Analytics 1.0* como la era del *business intelligence*. Esta etapa está caracterizada por una introducción a las primeras herramientas de análisis, centradas principalmente en las ventas y los procesos de producción. Estas herramientas tienen únicamente capacidad descriptiva, enfocándose en el análisis de hechos pasados. Debido a las restricciones tecnológicas, estos análisis podían llegar a durar semanas, por lo que era vital formular las preguntas adecuadas. El paradigma tecnológico sufrió una rápida evolución que permitió desarrollar el concepto de *big data*, y con ello, la nueva era de *analytics*. El *Analytics 2.0* se apoya en la idea de la colección de datos ajenos al entorno de la empresa, es decir, provenientes de fuentes de información secundarias. Esta etapa permitió el acceso a un mercado de datos mucho más amplio que el de la primera etapa, así como la necesidad de desarrollo de herramientas de análisis más avanzadas. La necesidad de técnicas más rápidas y precisas incitaron al desarrollo del *machine learning*, el desarrollo y testeo de modelos semi-automatizados, lo que permitió la introducción del análisis de datos predictivo. El salto generacional a la siguiente etapa no fue tan marcado como en la etapa anterior, pero este supuso un gran cambio en el ámbito competitivo de la empresa. De esta manera se dio paso a la etapa de *Analytics 3.0*, caracterizada por la fuerte adopción de *analytics* por las empresas como parte de su abanico competitivo. Uno de los puntos más importantes que destaca Davenport es que no solo las empresas tecnológicas son las únicas que pueden aprovechar las bondades del análisis de datos, sino que cada empresa en cada industria puede apalancar estos métodos para desarrollar nuevos productos o mejorar sus servicios (Davenport, 2013). El análisis llevado a cabo en 125 empresas pertenecientes a una variedad de sectores dentro del índice Standard & Poor's 500 revela que la gran mayoría de empresas

incorporan herramientas de *analytics* en sus operaciones e incluso cuentan con departamentos destinados enteramente a labores de *analytics*.¹

7. Big Data

El *big data* es un concepto que hace referencia a una colección de datos complejos de gran tamaño y que requieren de una gestión de costes efectiva, así como del análisis de estos para la extracción de *insights* (Tabesh et al, 2019). En comparación con los sets de datos tradicionales, el *big data* típicamente incluye grandes cantidades de datos sin estructurar que necesitan de análisis en tiempo real (Chen, 2014). La distinción entre un set de datos regular y uno de *big data* se basa en las cuatro uves:

- Velocidad: hace referencia a la tasa a la que se generan o actualizan los datos.
- Volumen: gran escala de datos, que requieren de herramientas actualizadas de colección, almacenamiento y análisis.
- Veracidad: hace referencia a las complejas estructuras que pueden hacer de los activos ambiguos, imprecisos o inconsistentes. Por ejemplo, los datos correspondientes a las opiniones de consumidores pueden estar sesgados o ser ambiguos.
- Variedad: variación en los tipos de datos. Los datos pueden provenir de fuentes de información tan variadas como textos, hojas de cálculo, videos o sensores.

Irónicamente, el *big data* era un serio problema para las empresas tan solo hace unos años. En el comienzo del siglo XXI, la capacidad de almacenamiento y las unidades de procesamiento de los ordenadores no podían hacer frente al creciente volumen de datos (Russom, 2011). La confluencia de la proliferación de la economía del dato, los avances algorítmicos y una potencia computacional más avanzada ha propulsado la adopción del *big data*. Las compañías se están alineando con esta tendencia, apresurándose para adaptarse a esta tecnología y apalancar la información de grandes cantidades de datos (Tabesh et al, 2019).

Un artículo de Harvard Business Review define el proceso como algo cíclico, y pone de relieve cuatro fases distintas en el análisis de *big data*. En una primera fase se recogen grandes colecciones de datos, generalmente sin estructurar. Estos datos son “limpiados” y analizados a través de las herramientas de *business analytics*, e interpretados por un equipo de *data scientists*. En una segunda fase, los *insights* observados de los datos son transformados en decisiones. La tercera fase consiste en traducir estas decisiones en acciones operativas. En teoría, estas acciones se deberían de ver transformadas en un incremento extraordinario de

¹ Anexo 1. Análisis de Empresas

ingresos que, a su vez, deberán de ser reinvertidos en el proceso de recogida de datos en la cuarta y última fase (Tabesh et al, 2019). El artículo recalca que ciclo perpetuo de *big data* puede beneficiar a la organización de manera significativa. No obstante, un informe de McKinsey Global Institute estima que la economía derivada del *big data* podría generar un valor adicional anual superior a tres mil millones de dólares en siete industrias analizadas a nivel mundial. El informe estima que alrededor de la mitad repercutirá de forma directa en los ciudadanos por la mejora de servicios públicos y privados (COTEC, 2017). Por si esto fuera poco evidencia, a continuación se procede a señalar tres ejemplos de casos de “éxito *big data*.”

a. Netflix

Netflix, Inc. es una empresa de suscripción de entretenimiento digital. La compañía ofrece una vasta gama de series de televisión, documentales y películas que varían en género e idiomas. A pesar de que la compañía se constituyó originalmente como un servicio de alquiler de películas y series en formato físico, Netflix aplica el uso de la tecnología de *business analytics* a todas sus áreas funcionales en su actividad regular. Desde su inicio, Netflix ha sido una compañía pionera en la toma de decisiones basadas en datos. De acuerdo con la propia compañía, el modus operandi de Netflix consiste en intentar solucionar cuestiones empresariales ambiguas sumergiéndose en enormes colecciones de datos complejos. Su nivel de integración es tal que opera de manera transdepartamental, compartiendo los descubrimientos y valoraciones de nuevas oportunidades a lo largo de las distintas áreas funcionales de la empresa, creando nuevos segmentos de crecimiento y siempre dando prioridad a los datos. Esto es posible gracias a un equipo entrenado en *data science*, así como a través del desarrollo de herramientas analíticas (Netflix, 2020). Esto les ha permitido alcanzar un número de suscripciones a su servicio de entorno a 182 millones en 2020. Por si fuera poco, su agresiva inversión en *business analytics* les ha permitido ser un referente en el sector del entretenimiento, llegando a revolucionar por completo un sector que estaba ligado a antiguas convenciones.

¿Qué diferencia a Netflix del resto de su competencia? Principalmente, su apuesta por los análisis predictivos desde sus comienzos. Tomemos el caso del desarrollo de la serie “House of Cards” como ejemplo. *House of Cards* es una serie producida por Netflix que debutó en 2013. La serie levantó titulares desde el principio gracias al enorme presupuesto que Netflix devotó a su desarrollo: 100 millones de dólares. A pesar de que se aseguró el mejor talento posible para la producción, se empleó el uso de *big data analytics*, hasta el punto de contratar al actor principal y a su director sin llevar a cabo un episodio piloto (Wu, 2015). Estas

decisiones fueron tomadas gracias a que su algoritmo predictivo mostraba que el producto final sería un éxito basado en el tema tratado en la serie y la popularidad de su elenco. El resultado final no solo fue un éxito ante las críticas, si no que ha probado ser una de las series más longevas de la compañía de *streaming*, llegando a alcanzar 2,8 millones de espectadores durante la primera semana. Por si esto fuera poco, Netflix ha demostrado que su formula es reproducible, repitiendo el éxito de la serie a través del estreno de la película “*Bird Box*” (Markman, 2019).

b. Verisk Analytics

Cuando uno piensa en grandes casos de éxito de integración de *analytics*, quizás Verisk no sea el primer ejemplo que venga a la mente. Sin embargo, en muchos ámbitos, Verisk es una empresa pionera que fue capaz de observar el potencial del análisis de datos desde sus comienzos, y que ha sido capaz de potenciar su motor de crecimiento basado en el uso de este. Verisk Analytics, Inc. es una empresa encargada de proveer soluciones de *data analytics* a empresas a nivel internacional. Entre la gama de servicios ofrecidos, Verisk favorece los análisis predictivos y da soporte de toma de decisiones a empresas en un amplio rango de ámbitos. Verisk comenzó su andadura como una empresa de seguros establecida en Nueva York, bajo el nombre de Insurance Services Offices (ISO). Distintos cambios en el marco regulatorio de la industria aseguradora en 1945 y 1971 permitieron que las agencias de rating se unificasen. De esta manera, la unión de cinco corredurías y un centro de análisis de datos dio paso a la creación de ISO. El conglomerado nació con una propuesta de valor único extraordinariamente anómala para la época. Estos cambios regulatorios habían hecho posible que ISO tuviera la base de datos de asegurados más grande del mundo. Tras el atentado del 11 de septiembre, ISO se dio cuenta de la importancia que sus clientes le daban al modelado de catástrofes en el marco de los seguros. A mediados de los 2000, la compañía comenzó a estudiar planes de crecimiento basados en nuevas industrias verticales. La robusta experiencia de ISO en el análisis de datos podría ser extrapolada a otros segmentos como el financiero o el energético. Verisk nació como respuesta a la posibilidad de hacer la empresa pública, como subsidiaria de ISO. Verisk se hizo pública en 2009, lo que constituyó la OPA más grande llevada a cabo este mismo año. Desde entonces, Verisk ha llevado a cabo una política de gestión en la que se premia la innovación y la iniciativa. Como resultado directo de esta filosofía y de la implementación de *analytics*, Verisk ha experimentado un crecimiento en sus ingresos por ventas del 30% en los últimos tres años, y su desempeño en la bolsa de valores ha superado consistentemente el desempeño del S&P 500 desde 2011 (Wasynczuk et al, 2018).

c. AWS, el futuro de Amazon

Amazon.com es una empresa de comercio electrónico y suscripciones que opera a nivel internacional a través de tres segmentos: América del Norte, Amazon Internacional y Amazon Web Services. Amazon es una empresa pionera en el uso de *big data analytics*. Amazon.com almacena información extensa sobre cómo interactúan los usuarios con los distintos elementos disponibles a lo largo de la página web. Esta información es analizada y expuesta a distintos modelos de *machine learning*, tales como el de reglas de asociación, el cual estudia los distintos elementos ofrecidos en la web de Amazon, analiza cómo han interactuado los distintos usuarios con estos elementos, y como resultado despliega las opciones que cuentan con mayor probabilidad de compra en base a aquellos usuarios que interactuaron con ciertos elementos. Gracias a este modelo, Amazon puede ofrecer recomendaciones personalizadas basadas en la actividad de cada usuario. No obstante, a pesar de que este aspecto forma parte de la gran identidad de Amazon, es otro de sus segmentos que realmente le convierten en una empresa competidora de *analytics*, Amazon Web Services. AWS es la vertiente de Amazon que ofrece servicios en la nube a empresas e individuos que varían desde almacenamiento de datos hasta servicios de *machine learning* e inteligencia artificial integrables sin conocimiento previo. Este es, sin duda, el segmento de Amazon que muestra mayor potencial. Un estudio de mercado llevado a cabo por Flexera Software (2019) a más de 786 empresas revela que, en 2019, el 67% de empresas encuestadas empleaba Amazon Web Services.

Gráfico 1. Evolución del crecimiento de la plataforma online de Amazon frente al crecimiento de Amazon Web Services



Fuente: Amazon.com

Tal y como se puede observar en el gráfico propuesto, a pesar de su corta vida, AWS muestra un crecimiento superior al de la plataforma online, aunque sus ingresos difieran enormemente

con respecto a este segmento. Empresas que confían en el uso de datos como piedra angular de su actividad, tales como Dow Jones, Nasa o Intuit, hacen uso de Amazon Web Services en sus operaciones diarias, factor que demuestra el potencial de crecimiento de dicha plataforma.

8. *Análisis del business analytics en la empresa*

a. *Investigación*

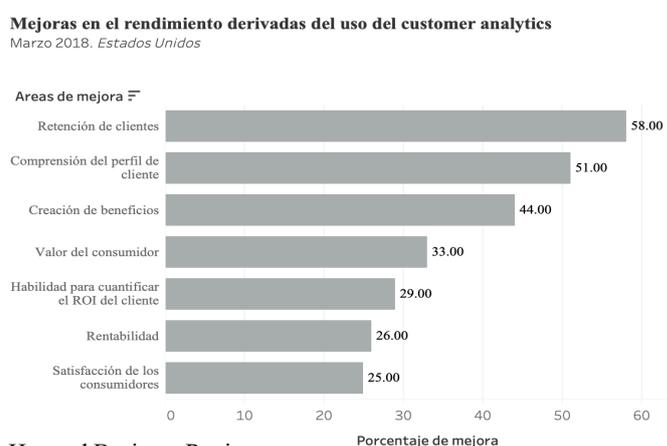
El *business analytics* pone a disposición del empresario una serie de herramientas que, de ser utilizadas por un individuo entrenado, pueden producir grandes resultados dentro de la empresa. Los autores del ensayo *Emerging Trends in Business Analytics* (2002) pusieron de relieve una serie de variables que pueden afectar negativamente a la implementación del *business analytics* si no están bien definidas en la empresa. En una época en la que el acceso a datos es tan sencillo, la colección de datos se vuelve un problema. La facilidad para la recogida de datos provoca que se coleccionen grandes sets de datos que, por lo general, quedan sin utilizar ya que nunca se definieron los parámetros ni las preguntas a responder. Este hecho se añade a expectativas irrealistas sobre los resultados que se pueden inferir a través del análisis de datos. La gran variedad de fuentes de datos disponibles provoca que los sets de datos varíen en formato y tamaño. Esto hace que el proceso de limpieza, integración y análisis de datos se complique (Kohavi et al, 2002). Es necesario que la implementación del *business analytics* identifique estos desafíos y se definan soluciones a los mismos.

Cómo definir el rendimiento de una empresa no es tarea fácil. Existen numerosas formas de medir distintos tipos de rendimiento. El doctor Andy Neely, vicedecano en la Universidad de Cambridge hace distinción de hasta cuatro tipos de rendimiento: desde el punto de vista financiero, desde el punto de vista de marketing, desde un punto de vista operacional y desde el punto de vista de gestión empresarial (2007). Una medida para el rendimiento ha de ser un valor cuantitativo estándar, que permita compararlo en el horizonte temporal, así como compararlo competitivamente (Kellen & Wolf, 2003). Bajo este criterio, teniendo en cuenta la subjetividad de algunas medidas, así como la facilidad para el acceso de ciertos datos, y con el objetivo de simplificar este problema de multidimensionalidad del concepto, a lo largo del análisis se hará referencia al concepto de rendimiento empresarial desde el punto de vista operacional, entendiendo que este atributo puede recoger el impacto más directo del *business analytics*. Tal y como se ha podido observar en casos como los de Netflix, Microsoft o Amazon, las empresas que incorporan las herramientas del *business analytics* en la gestión y operaciones diarias, ven beneficios en el largo plazo. Las organizaciones se han vuelto mucho más competitivas gracias al empleo del *business analytics*. Un estudio llevado a cabo en 2013 por

el MIT Sloane Management Review pone de relieve que un 67% de las compañías utilizan el análisis de datos como ventaja competitiva frente al 37% en 2010. Esto viene a confirmar que, efectivamente, el negocio de los datos está en auge y que cada vez más empresas se benefician de las herramientas que el *business analytics* puede aportar (Asllani, 2014).

¿Cómo puede afectar exactamente la incorporación de dichas herramientas al resultado de la empresa? Como se ha venido anunciando hasta el momento, el *business analytics* es un concepto relativamente moderno, por lo que sus aplicaciones no han sido adoptadas por la gran mayoría (a pesar de ser un movimiento con gran potencial de expansión), por lo que todas aquellas empresas que lo adopten pueden aprovechar una ventaja competitiva por avance tecnológico. Funcionaría de la siguiente manera: la empresa recopila datos que analiza para sacar *insights* sobre consumidores. El empresario toma decisiones basándose en estos *insights*. Los *insights* también se traducen en un mayor entendimiento de los patrones de comportamiento de los clientes, lo que favorece un trato personalizado. Los consumidores observan esto y atribuyen a la empresa un valor añadido, que servirá como elemento diferenciador frente a los competidores. Con esta ventaja competitiva, la empresa puede elevar sus precios, lo que implica un ingreso por ventas mayor (Hill et al, 2014). Un estudio llevado a cabo por Harvard Business Review en colaboración con Intel y Accenture a más de 560 líderes del mundo de la empresa en 2018 revela que el 70% de las organizaciones participantes ha incrementado su gasto en soluciones de *business analytics* en los últimos años, y de ellas, el 44% ha experimentado un incremento en los ingresos por ventas directamente relacionados. Este mismo estudio revela que los sectores de la banca, textil y el sector de telecomunicaciones poseen el mayor potencial de crecimiento como resultado de la adopción de las herramientas de *business analytics*. A continuación, se pueden observar los segmentos desglosados:

Gráfico 2. Estudio del impacto en el rendimiento empresarial basado en la implementación del *business analytics* en 560 empresarios



Fuente: Harvard Business Review

Adicionalmente, Erik Brynjolfsson, director de Initiative on the Digital Economy del prestigioso Instituto Tecnológico de Massachusetts (MIT), sostiene que las empresas que adoptan decisiones basadas en datos logran entre 5% y 6% más productividad que aquellas que no lo hacen. (COTEC, 2017). Para más inri, un estudio llevado a cabo por McKinsey en 2009 revela que, a pesar de que el número de líderes que emplean métodos de análisis predictivo es sustancialmente pequeño, aquellos que lo utilizan experimentan un rendimiento sustancialmente superior frente al resto (Germann, 2013).

b. Metodología

La metodología que se ha llevado a cabo para la preparación del análisis y su posterior conclusión ha sido, en primer lugar, un estudio teórico de la literatura ligada al concepto de *business analytics*, *Big Data* y el empleo de este en las empresas. Este estudio ha elevado cuestiones relacionadas con la inclusión de este tipo de herramientas en la empresa. A continuación, se ha procedido al desarrollo del modelo mediante la elección de variables a tener en cuenta. Bajo este concepto, se ha creado un set de datos en Excel constituido por 100 empresas seleccionadas de forma aleatoria dentro del índice económico S&P 500. Con este set de datos, se condujo un análisis descriptivo, para observar tendencias iniciales. Para ello se han empleado gráficos, elaborados a través del software RStudio, Microsoft Excel y Tableau. A continuación, se ha formulado el siguiente modelo matemático:

$$\begin{aligned} \text{Ln (Ingresos por ventas)} = & \beta_0 + \beta_1 (\text{ROA}) + \beta_2 (\text{Puntuación gobierno}) + \beta_3 \\ & (\text{Número de trabajadores}) + \beta_4 (\text{Expectativa de crecimiento}) + \beta_5 (\text{EV/R}) + \beta_6 (\text{Grado} \\ & \text{de integración 2}) + \beta_7 (\text{Grado de integración 3}) + \beta_8 (\text{Grado de integración 4}) + \beta_9 \\ & (\text{Grado de integración 5}) \end{aligned}$$

El modelo econométrico que se ha utilizado para la consecución de la investigación es estático, uniecuacional y lineal. Uniecuacional porque hay una sola ecuación, estático porque no se está considerando la variable del tiempo y de regresión lineal porque el modelo puede verse como una función que para cada valor de las X ofrece la media condicionada de las Y. El principal objetivo de este modelo es identificar qué variables no son significativamente relevantes, y así eliminarlas. El modelo de regresión lineal permite identificar en qué medida afectan cada una de las variables independientes a la variable dependiente. El modelo de regresión lineal obtiene esta solución mediante la minimización de la distancia entre los datos estimados y los datos reales, a través de un método conocido como Mínimos Cuadrados Ordinarios. De esta forma,

se busca comprobar que dichas variables tienen un impacto (es decir, la beta es distinta de cero) sobre el resultado de la empresa. Este resultado será contrastado con los resultados gráficos.

El objetivo es que ambos análisis muestren un resultado similar. Por tanto, este análisis será dividido en dos partes: un análisis descriptivo que permita identificar tendencias, así como ajustar y limpiar el set de datos, y un análisis exploratorio. A través de este análisis exploratorio se busca verificar que el *business analytics* tiene un impacto (bien positivo o negativo) en el resultado final de la empresa. A primera instancia, atendiendo a la literatura propuesta y sin intención de caer en falsas inferencias no basadas en evidencia experimental, parece ser que la respuesta a esta hipótesis es que, en efecto, el *business analytics* afecta al resultado de la empresa de manera significativa.

H-1. La incorporación de analytics afecta a la empresa positiva o negativamente

$$H_0: \beta_{\ln(\text{rendimiento empresa})} = 0$$

$$H_1: \beta_{\ln(\text{rendimiento empresa})} \neq 0$$

c. Fuentes, datos y variables empleadas

De esta forma se procede a enunciar algunos de los detalles detrás de la investigación. El proceso de investigación se ha llevado a cabo a través de la obtención de datos de corte transversal. Los datos de corte transversal consisten en observar a distintos individuos (empresas pertenecientes al S&P 500) en el sentido estadístico del término. Se ha tenido en cuenta tanto las observaciones de las variables explicativas como de la variable endógena. Los datos se han obtenido principalmente de fuentes de información secundarias. Dado que se busca probar el impacto de la implementación de herramientas de *business analytics* en la empresa, no se le ha dado importancia a la situación geográfica de las mismas, entendiendo que el mismo resultado debería darse bajo las mismas circunstancias independientemente de la situación geográfica. Por esta misma razón, no se han tenido en cuenta variables de índole macroeconómica que puedan afectar los resultados del modelo y dificulten su reproducibilidad. La razón por la que se acabó seleccionando el índice económico S&P 500 para la selección de empresas se debe a que el tamaño de las organizaciones que lo componen son lo suficientemente grandes, en términos de capitalización, como para poseer o no un departamento destinado a *business analytics* dentro de la empresa, y que, por tanto, el impacto en el resultado final sea significativo. Adicionalmente, se han eliminado todas aquellas empresas que forman parte del sector financiero (a excepción de Capital One, debido a su estructura altamente analítica), ya que sus complejas cuentas de resultados embarran el enlace directo que suponen los ingresos por ventas. La muestra cuenta con 125 organizaciones

distintas seleccionadas aleatoriamente. Del total de la muestra, 100 empresas serán utilizadas para entrenar el modelo, mientras que las 25 restantes serán empleadas para verificar los resultados. Algunos de las fuentes consultadas para la obtención de datos incluyen Yahoo! Finance y el portal estadístico Statista. Por tanto, se puede concluir que la muestra es lo suficientemente variada aleatoriamente para la estimación y consecución de ambos modelos.

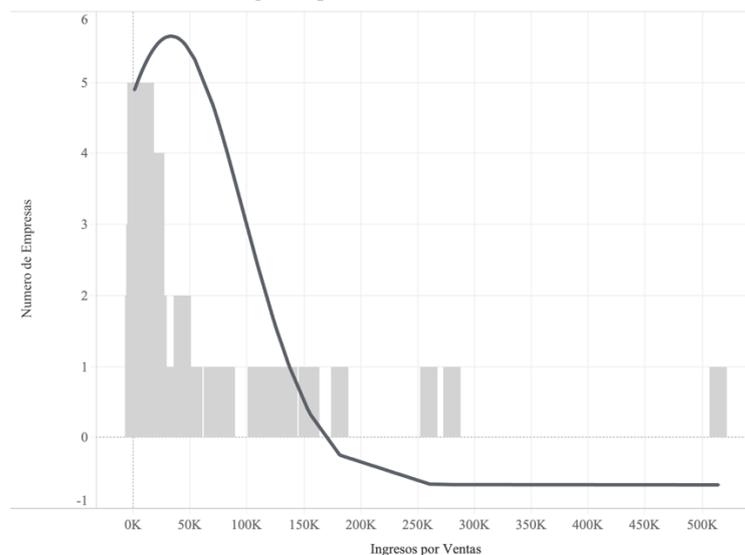
Existen múltiples fuentes que pueden afectar al resultado final de la empresa. Es importante tener en cuenta la mayor parte de variables que pueden afectar al análisis con la intención de individualizar el efecto de la variable a medir y que esta no recoja el efecto de otras variables no tenidas en cuenta. Adicionalmente, hay que cuidar que, por intentar reducir lo mencionado anteriormente, se consiga el efecto contrario, entrando en un problema de multicolinealidad. Se entiende la multicolinealidad como el problema derivado de la dependencia lineal entre las variables que conforman el modelo econométrico. Este hecho puede dificultar precisar el efecto que ejerce cada una de las variables independientes sobre la variable dependiente (Vega Vilca & Guzmán, 2011).

Es difícil evaluar en qué cantidad exacta afecta el hecho de que una empresa utilice herramientas de *business analytics* en sus operaciones en cuanto a beneficios. Sin embargo, este análisis está basado en el trabajo de distintos académicos en la materia, que han venido definiendo las variables a tener en cuenta a la hora de analizar el impacto del uso de *analytics* en desempeño de dicha empresa. El siguiente apartado está destinado a explicar la naturaleza de las variables empleadas en el modelo, así como el por qué de las mismas.

- **Ln (Ingresos por ventas):** Como se ha mencionado con anterioridad, hay muchas maneras de entender el rendimiento de una empresa, y hay el mismo número de maneras de medir el desempeño de una empresa. Su diversidad es tal, que se podrían llevar a cabo numerosos proyectos al respecto. El objetivo es demostrar que dicho desempeño se ve mejorado en base al grado de incorporación de herramientas de *business analytics* (en mayor o menor medida) en la empresa. En este sentido, otros estudios de la misma índole (Germann et al, 2013) han empleado variables tales como el Return on Investment (RoI) que viene a indicar la proporción en la que una inversión produce beneficios, o en caso contrario, pérdidas. El cálculo de dicho término sería enormemente complicado en este caso dado el limitado acceso a los datos de las empresas, que se limita a los datos públicos. Otros estudios emplean el EBITDA como forma de medida de impacto. Esta medida, no obstante, se ha mostrado como poco reflectante del posible impacto marginal de la incorporación del *business analytics*. Con el objetivo de buscar la mayor relación directa, se ha decidido

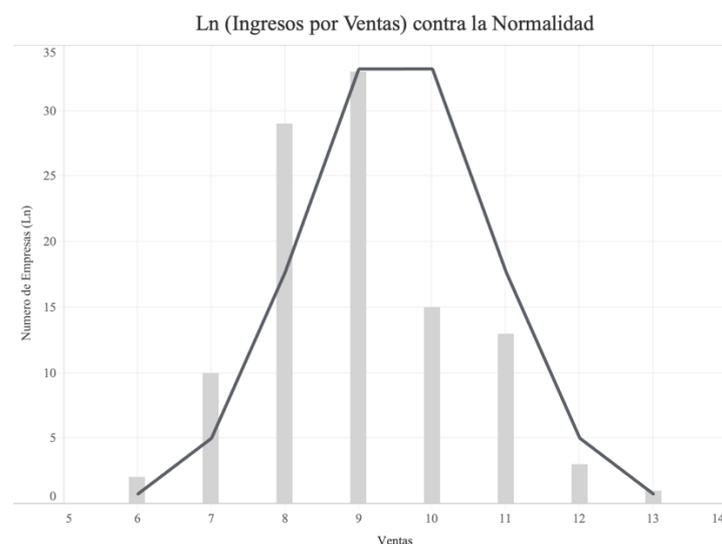
emplear el ingreso por ventas como fuente de medición. Dicho dato se ha extraído de las distintas cuentas de pérdidas y ganancias correspondientes al año 2019 incorporados a la página de la Comisión de Valores Estadounidense (SEC por sus siglas en inglés). El ingreso por ventas es una variable cuantitativa, endógena y discreta. No obstante, debido a la variedad de ingresos de las empresas seleccionadas, el rango de datos es amplio como se puede observar en el siguiente gráfico:

Gráfico 3. Dispersión de la variable “Ingresos por ventas” contra la normalidad
Ingresos por Ventas contra la Normalidad



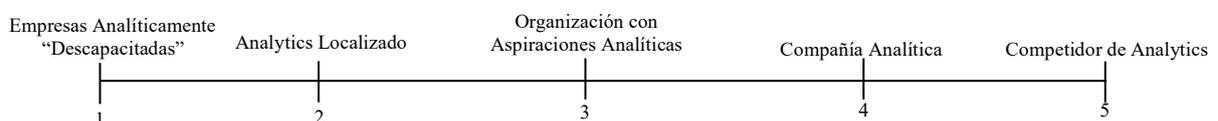
Por esta razón, se atenderá a la presente variable como si fuera una variable continua y se procederá a expresarse en forma de escala logarítmica. Esto permitirá que los valores se contraigan, disminuyendo así el rango de datos, quedando representado de la siguiente manera:

Gráfico 4. Dispersión de la variable “Ln (Ingresos por ventas)” contra la normalidad



- **Rentabilidad de los Activos:** Variable cuantitativa y discreta. El RoA por sus siglas en inglés (Return on Assets) es un indicador financiero que mide la rentabilidad de la empresa en relación con los activos que posee la misma y se calcula mediante la división de los ingresos netos percibidos por la empresa entre sus activos totales. Es un indicador vital en este caso porque señala cómo de eficiente es el liderazgo de la empresa a la hora de gestionar los activos para la generación de ingresos. Es el máximo exponente en cuanto a medida de eficiencia. A mayor RoA, mayor eficiencia. En teoría, de acuerdo con la literatura propuesta, el uso de *analytics* debe de incrementar la eficiencia de los activos.
- **Grado de incorporación de Business Analytics en la empresa:** En su libro “*Analytics at work: Smarter decisions, better results,*” el académico Thomas Davenport hace una distinción de hasta cinco diferentes etapas en el espectro de empresas que adoptan el *Business Analytics* en su actividad diaria (2010):
 - Empresas analíticamente “discapacitadas”: aquellas organizaciones que carecen de datos, habilidades analíticas o interés dentro del equipo de liderazgo.
 - Analytics localizado: aquellas empresas que poseen centros focalizados de actividad analítica pero que no están alineados con los objetivos estratégicos.
 - Organizaciones con aspiraciones analíticas: todas aquellas empresas que visualizan un futuro más analítico dentro de la organización. Dichas empresas han establecido “hubs” de análisis y poseen ciertas iniciativas en desarrollo, pero el progreso es lento.
 - Compañías analíticas: la organización posee tanto el capital humano como los recursos tecnológicos para aplicar *analytics* de forma regular y a lo largo de su actividad empresarial. Sin embargo, su estrategia no está enfocada en el *analytics*, y no emplea el mismo como ventaja competitiva frente a la competencia.
 - Competidor de analytics: la empresa posee las mismas características que las compañías analíticas, pero aprovechan su capacidad de análisis como ventaja competitiva frente al resto de la competencia.

En el Anexo se puede encontrar el proceso de investigación que se ha llevado a cabo para clasificar cada una de las 125 empresas que forman parte del modelo. Siguiendo las pautas de clasificación del Dr. Davenport, se ha procedido a dividir las empresas en base a criterios de estructura de datos (qué clase de datos puede tener la empresa), unicidad y gobernabilidad (Davenport et al, 2010).



En términos prácticos, a la hora de incluir la variable en el modelo, se ha procedido a crear cuatro variables distintas, a las que se le ha dado valores de 1 o 0 en base a si la observación pertenece a esa categoría o no.

- **Puntuación del Gobierno Corporativo:** Variable cuantitativa y discreta. Dicha variable se presenta en una escala de valores que varía desde el decil 1 al decil 10. Una puntuación más baja (cercana al decil 1) representa prácticas de gobierno de calidad relativa mayores que aquellas empresas con una puntuación más alta. La puntuación de la gestión corporativa viene recogida por el Institutional Shareholder Services como medida de análisis de riesgo de gobierno corporativo dentro del portafolio de compañías. La presente variable recoge el sentimiento de los trabajadores en relación con el desempeño y decisiones tomadas por la cúpula corporativa de la empresa. Teóricamente, de acuerdo con la literatura presentada, es de esperar que una cúpula efectiva (y por tanto con mejor *scoring*) integre las herramientas y gestione a los analistas de datos de manera más eficiente y, por tanto, esta mejora se vea reflejada en un ingreso por ventas mayor.
- **Número de empleados:** Variable cuantitativa, explicativa y discreta. Esta variable recoge el número de empleados registrados en la empresa a tiempo completo. Teóricamente el número de empleados no se debería de ver afectado por el grado de análisis de la empresa, pero sí que afecta a los ingresos, por lo que es necesario tenerlo en cuenta dentro del modelo con la intención de aislar el efecto del resto de variables sobre los ingresos. Los datos para la variable han sido recogidos de Yahoo! Finance.
- **Expectativa de crecimiento:** Variable cuantitativa, explicativa y continua. De acuerdo con la literatura propuesta, el *business analytics* debería de ser un elemento diferenciador que potencie el crecimiento de la empresa. De esta forma, se espera que aquellas empresas analíticamente avanzadas tengan gran potencial de desarrollo futuro, por lo que tendría sentido que el grado de integración afecte positivamente a la expectativa de crecimiento de la empresa. Los datos correspondientes a la variable se han recogido de Yahoo! Finance. Dicha variable viene expresada en términos porcentuales, representando el crecimiento (o decrecimiento) anual de la empresa para los próximos cinco años. La razón por la que se ha optado por tomar la expectativa de crecimiento de los próximos cinco años y no la del próximo año es debido a la situación económica provocada por el virus COVID-19 que

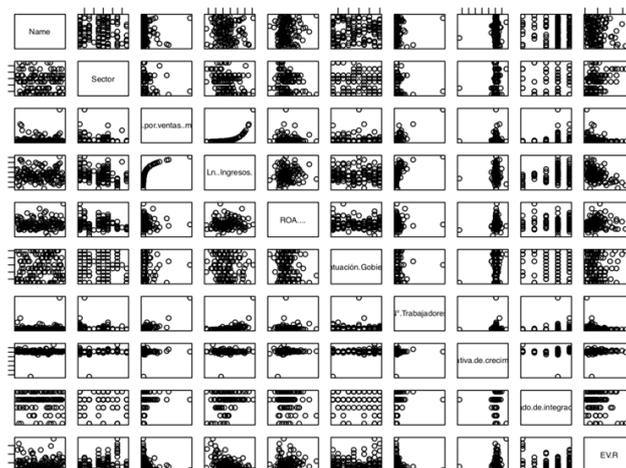
puede afectar a las expectativas de crecimiento en el corto plazo. Al escoger la expectativa para los próximos cinco años se diversifica el riesgo generado por dicho virus.

- **Enterprise Value / Revenue:** Variable cuantitativa, explicativa y continua. Este múltiplo cuantifica el valor de mercado de una empresa en relación con sus ingresos por ventas. Esta ratio pone en términos relativos la capacidad de la empresa para generar ventas frente a su tamaño. Se ha decidido optar por añadir esta variable ante la posibilidad de que el tamaño de las empresas sesgue la influencia del grado de implementación de *analytics* (es posible que una empresa de gran tamaño no tenga un gran grado de implementación, pero que perciba grandes ingresos procedentes de ventas). Asimismo, dado que esta variable es la combinación lineal de los ingresos por ventas con el valor de mercado de la empresa, cabe la posibilidad de que se entre en un problema de multicolinealidad con la variable “ingresos por ventas.”

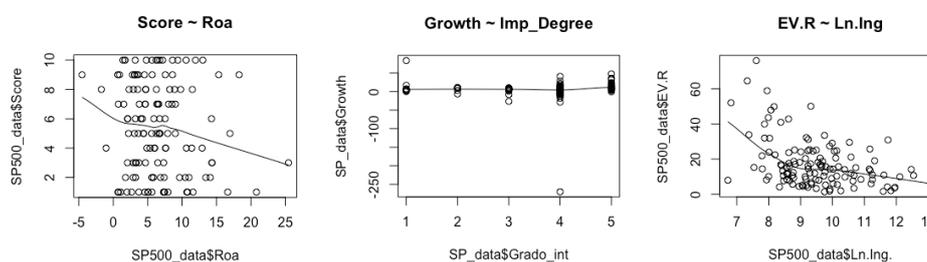
d. Análisis descriptivo

La primera dimensión del análisis consta de un estudio preliminar de ciertos aspectos del set de datos, tanto en términos de gráficas como de estadísticos principales. El objetivo de este análisis es el de eliminar potenciales valores que incurran en problemas de multicolinealidad, identificar posibles *outliers* en el set de datos y observar tendencias basadas los modelos gráficos. Estos se utilizarán en contraste con los resultados expuestos por los modelos de regresión llevados a cabo en el análisis exploratorio. Para llevar a cabo este análisis se ha procedido a cargar el set de datos en la aplicación RStudio. La función *pairs* permite observar una matriz de gráficos compuesta por la combinación lineal de cada una de las variables que componen el set de datos. A través de este atajo es posible observar con rapidez *outliers* o patrones en cualquiera de las combinaciones lineales de las variables:

Gráfico 5. Matriz de Combinación Lineal de Variables

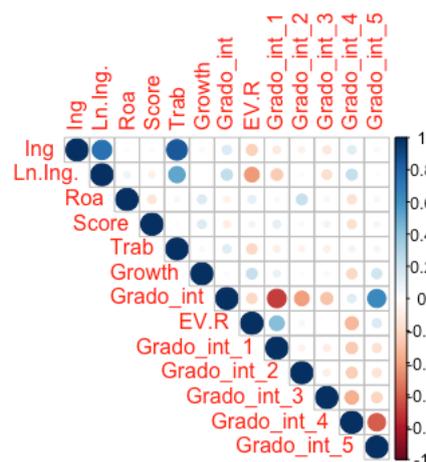


De esta matriz se pueden extraer distintas observaciones. Las variables “Ingresos por ventas” y “Ln (Ingresos por ventas)” parecen tener una correlación negativa con la variable que recoge la ratio del valor de mercado de la empresa con relación a sus ingresos por ventas. Esto tiene sentido dado que esta ratio esta formada por la combinación lineal de la variable “Ingresos por ventas.” También existe una correlación, esta vez positiva, entre la puntuación del gobierno corporativo y el RoA, una mejor puntuación de la gestión se traduce en una mayor rentabilidad de los activos. Asimismo, el gráfico compuesto por el grado de integración de *analytics* y la expectativa de crecimiento apunta a que un mayor grado de integración se traduce en una mayor expectativa de crecimiento.



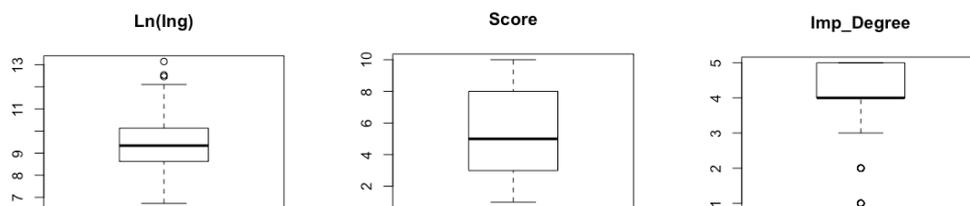
Para poder constatar esta relación en mayor medida, se ha procedido a formular una matriz de correlación con todas las variables. Esta tabla pone de relieve la correlación de la combinación de cada una de las variables del set. Como práctica común, se ha decidido emplear un límite de 0,8 para determinar si existe una relación lineal significativa entre dos variables, y que por tanto haya un problema de multicolinealidad.

Gráfico 6. Matriz de Correlación de las Variables para las Observaciones 1-125



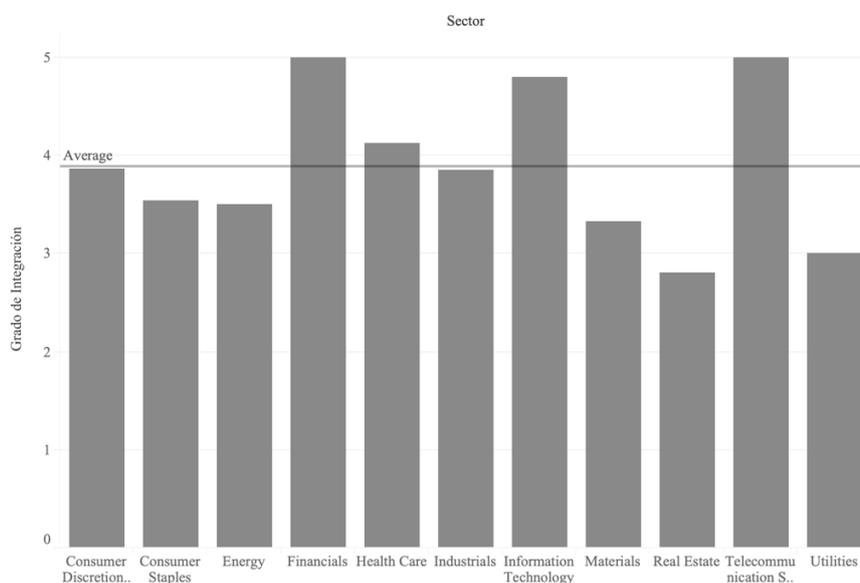
La matriz de correlación pone de relieve la necesidad de eliminar la variable “Ingresos por ventas,” ya que, como se venía intuyendo hasta el momento, está estrictamente relacionada con la variable “Ln (Ingresos por ventas)” y parece guardar relación con el número de trabajadores también. Naturalmente, la matriz señala la relación entre la variable “Grado de integración” y

sus sucesivas variables binarias. Por esta razón, se ha eliminado las variables “Ingresos por ventas” y “Grado de integración.”



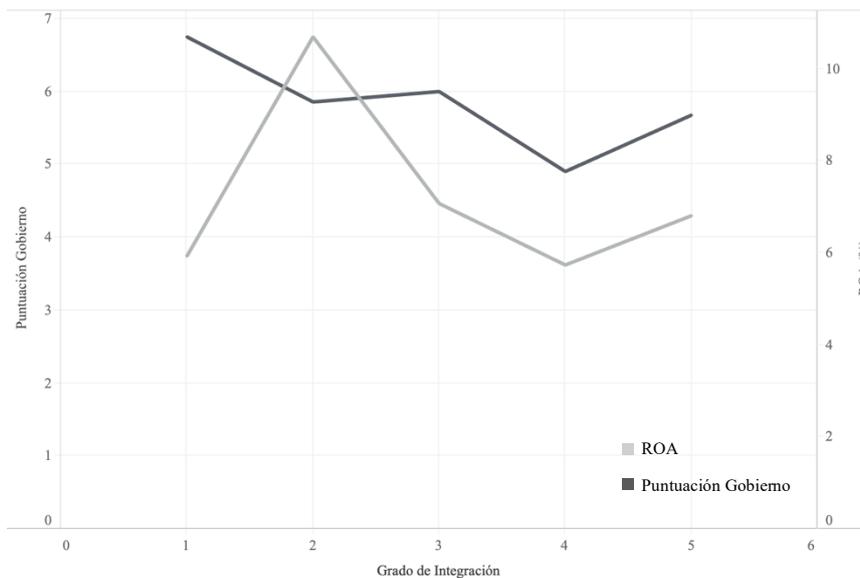
Una vez descartados la existencia de *outliers* significativos y habiendo suprimido las variables que podían alterar el resultado de las regresiones, se puede atender al análisis gráfico de tendencias. Anteriormente se hacía mención a un estudio llevado a cabo por Harvard Business Review con Intel que apuntaba a que las empresas del sector financiero, textil y telecomunicaciones son aquellas que aprovechan mejor las capacidades de *analytics*. Esta tendencia parece cumplirse de acuerdo con el análisis de las empresas del S&P 500. El gráfico desplegado a continuación pone de relieve otros sectores crecientes, como el sector de seguros médicos.

Gráfico 7. Distribución de Grado de Integración por Sector



La literatura que se ha citado hasta el momento señala la importancia de una buena cúpula que gobierne y que gestione los recursos hacia un futuro fuertemente apoyado en la integración de *analytics* en las operaciones de la empresa. Anexo a este concepto se encuentra la idea de que, aquellas empresas que cuentan con un mejor cuerpo de gobierno, y por tanto incorporan herramientas de *analytics* en mayor medida, deberían de registrar mejores volúmenes de ventas que aquellas que no.

Gráfico 8. Distribución de Empresas de según el Grado de Integración



El gráfico propuesto pone de relieve información muy interesante. Parece que, tal y como apuntaba el gráfico de correlación, existe una relación directa entre eficiencia y liderazgo. De media, en su mayor parte, aquellas empresas cuyas cúpulas de gobierno toman decisiones de mayor calidad perciben un mayor rendimiento de sus activos. Existe bastante disparidad entre las empresas que integran *analytics* en términos de gestión, es decir, no parece haber una correlación directa entre una mejor gestión y un mayor grado de integración de *analytics*. Por otro lado, el gráfico si que apunta a este tipo de relación entre rendimiento de los activos y grado de integración, donde por lo general, un mayor grado de integración se traduce en mayor rentabilidad de los activos. No obstante, cabe destacar un pico en la rentabilidad de aquellas empresas que cuentan con focos no desarrollados de *analytics*. Esto se puede deber al hecho de que la rentabilidad de algunos de los activos gestionados por dichas compañías es intrínsecamente alta, como puede ser el caso de las empresas del sector inmobiliario. El análisis exploratorio llevado a cabo a continuación arrojará más luz sobre este efecto.

e. Análisis exploratorio

A través del análisis exploratorio se busca encontrar una relación más profunda en el efecto que tienen las distintas variables sobre la variable dependiente, Ln (Ingresos por Ventas). La regresión lineal calcula una recta optimizada para minimizar la distancia entre los puntos predichos y los datos actuales del set. Además, se va a poder estudiar qué variables afectan significativamente a esta variable endógena. Para estudiar esta significatividad más a fondo, se van a llevar a cabo regresiones con seis modelos distintos. En cada iteración del modelo se incluirá una nueva variable de manera gradual. La regresión lineal se ha construido mediante la partición del set de datos en “Training,” con el 75% de los datos, y en “Test,” con el 25%

restante. Las distintas iteraciones de los modelos de regresión se calculan dentro del set de entrenamiento para no “contaminar” el set de testeo y así poder medir su capacidad predictiva.

Gráfico 9. Regresión Lineal. Modelos Estimados para las Observaciones 1-100

	Modelo I		Modelo II		Modelo III		Modelo IV		Modelo V		Modelo VI	
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Constante	8,4333	16,319***	8,21008	15,199***	8,2727	13,601***	8,314	16,072***	8,328	15,952***	9,300	16,140***
Grado de Integración 2	0,8983	1,229	0,71216	0,962	0,70608	0,948	7,254e-01	1,145	7,275e-01	1,142	1,845e-01	0,295
Grado de Integración 3	0,6822	1,023	0,61600	0,925	0,61236	0,914	5,470e-01	0,960	5,481e-01	0,957	-2,030e-02	-0,036
Grado de Integración 4	1,3240	2,407*	1,33766	2,442*	1,33105	0,0179*	1,069	2,269*	1,074	2,267*	5,663e-01	1,193
Grado de Integración 5	1,0478	1,834	0,98180	1,720	0,97815	1,704	7,306e-01	1,491	7,22e-01	1,464	4,751e-01	1,004
RoA			0,03903	1,353	0,03855	1,326	7,306e-01	1,120	2,643e-02	1,043	3,160e-02	1,315
Puntuación Gobierno					-0,00999	-0,229	2,778e-02	-0,335	-1,413e-02	-0,374	-1,961e-02	-0,548
Número Trabajadores							2,638e-06	5,822***	2,632e-06	5,774***	2,263e-06	5,074***
Expectativa Crecimiento									1,173e-03	0,297	2,543e-03	0,676
EV/R											-3,182e-02	-3,280**
R ²	0,07429		0,09337		0,09393		0,3523		0,3529		0,4272	
R ² corregido	0,03221		0,04127		0,03071		0,2989		0,2913		0,3651	
Estadístico F	1,765		1,792		1,486		6,604		5,727		6,878	

Para los estadísticos t, se rechaza la hipótesis nula, y se concluye que la variable es significativa *** al 1% de significación ($p < 0.01$) ** al 5% de significación ($p < 0.05$) * al 10% de significación ($p < 0.1$)
 Para el estadístico F, se rechaza la hipótesis nula, y se concluye que el modelo es conjuntamente significativo *** al 1% de significación ($p < 0.01$) ** al 5% de significación ($p < 0.05$) * al 10% de significación ($p < 0.1$)

Cada uno de los modelos ha sido evaluado en función de su validez a través de los contrastes de significación individuales (estadístico t) y globales (estadístico F), así como la bondad de ajuste (medida a través de R^2). Lo primero que llama la atención de la tabla propuesta es que, cuando se tiene en cuenta el modelo en su conjunto, ninguna de las variables que registran el grado de integración de la empresa en *analytics* afecta significativamente a los ingresos, por lo que no se puede rechazar la hipótesis nula a ningún nivel de significatividad. Si, por otro lado, consideramos el primer modelo, donde solo se tienen en cuenta las variables que recogen el nivel de integración, se puede observar que no solo es una de las variables significativa a un nivel de significación del 10%, sino que afecta positivamente a la variable endógena. En otras palabras, de acuerdo con este modelo, cuando las empresas alcanzan cierto nivel de inversión en *analytics*, estas pueden esperar un impacto del 275% de media y caeteris paribus en los ingresos por ventas por cada unidad invertida en integración de *analytics*. Esta evolución tan drástica en los parámetros estimados de las variables responde a la omisión de importantes variables de control (ver evolución de coeficientes en Modelo I contra el Modelo VI). Llama la atención que ninguno de los modelos es significativo en su conjunto, y el valor R^2 es relativamente bajo, por lo que las variables independientes explican un porcentaje bajo de la variable dependiente. Asimismo, la volatilidad en la evolución de las estimaciones revela un problema de estabilidad e interpretabilidad. Si se atiende al estudio de precisión, el modelo estima correctamente el 92,7% de los datos.

La regresión Ridge toma este modelo de regresión lineal de partida y aplica un procedimiento de restricción de ciertos parámetros a través de la penalización de los coeficientes de regresión (λ). Cuanto mayor es el grado de penalización, menores serán los coeficientes y, por

tanto, más robustos a problemas de multicolinealidad. Esto hace que el modelo estime los datos actuales con menor exactitud que el método de Mínimos Cuadrados Ordinarios, pero que este desempeñe mejor generalizando, ya que es menos sensible a la variación extrema de datos como los *outliers*. Se espera, por tanto, que a través del cálculo de la regresión Ridge se consigan resultados más estables, con coeficientes más significativos.

Gráfico 10. Regresión Ridge. Modelos Estimados para las Observaciones 1-100

	Modelo	
	Coef.	t
Constante	9,554	NA
Grado de Integración 2	-5,97e-02	0,279
Grado de Integración 3	-1,728e-01	0,987
Grado de Integración 4	1,3240	1,979*
Grado de Integración 5	1,0478	0,607
RoA	1,569e-02	1,376
Puntuación Gobierno	-9,816e-03	0,540
Número Trabajadores	1,231e-03	0,677
Expectativa Crecimiento	1,234e-06	5,660***
EV/R	-1,799e-02	4,248***
Ridge Parameter	1,031635	

Como se puede observar del gráfico expuesto, los parámetros que penalizan a los coeficientes han provocado que algunos de estos sean negativos. No obstante, dicho modelo presenta valores más significativos que el generado a través de los Mínimos Cuadrados Ordinarios, especialmente para los valores asociados a la variable que recoge las empresas con un grado de integración en *analytics* 4. En este caso, el valor t de la variable es significativo a un nivel del 10%, por lo que se puede asegurar con un nivel de seguridad del 90% que aquellas empresas que alcanzan cierto nivel de inversión en *analytics* perciben más ingresos por cada unidad invertida. Adicionalmente, se verifica que este modelo presenta ligeramente peor precisión que el de regresión lineal, con una capacidad de estimación del 92.19%. No obstante, es conveniente comparar la capacidad predictiva de ambos modelos para determinar cuál de los dos está mejor preparado.

Gráfico 11. Estudio de Predictibilidad. Modelo Observado contra Modelo Predicho

K-Pliegues	Model RMSE	RMSE New Data
Mínimos Cuadrados Ordinarios	1,177583	0,9246839
Regresión Ridge	1,224964	0,9239218
Train/Test Split	Model RMSE	RMSE New Data
Mínimos Cuadrados Ordinarios	0,9686316	0,9246839
Regresión Ridge	0,9686316	0,978912

Una de las razones que motivaron la decisión de optar por la regresión lineal entre otros modelos, es porque esta ofrece la posibilidad de estudiar la capacidad predictiva. Para poder comprobar esta predictibilidad, se ha de llevar a cabo una prueba que consiste en comparar la

desviación estándar de la diferencia entre los resultados estimados y los datos actuales (residuos) con la desviación estándar de los residuos del modelo predicho. Para ello se han empleado dos metodologías: Train/Test Split y K-Pliegues. Ambas metodologías requieren de la partición de los datos en dos sets distintos. La validación cruzada permite establecer la exactitud del modelo a la hora de predecir. El primer modelo consiste en una única iteración en la que se calcula el error cuadrático medio de los sets de entrenamiento y examinación. El segundo método opera “K” iteraciones y computa un grado medio para el error cuadrático medio, lo que ofrece una propuesta más plausible. De la tabla del estudio de predictibilidad se puede extraer que el modelo derivado de la regresión Ridge opera peor que la regresión lineal a la hora de estimar coeficientes de datos reales, mientras que tiene mejor resultado en la predictibilidad de datos no conocidos.

CONCLUSIONES Y DESARROLLO FUTURO

Li & Fung, una compañía que gestiona el proceso de cadena de valor para numerosas empresas, observaba en 2009 un flujo de datos en su red que apenas alcanzaba 100 gigabytes. Un año después, esta cantidad se había multiplicado por diez (The Economist, 2010). La cantidad de información disponible se ha disparado exponencialmente en los últimos diez años, y esta tendencia solo se está acentuando. La tecnología que permite la colección y análisis de datos ha experimentado una evolución similar, abaratando los costes asociados a estas tareas. El análisis llevado a cabo sobre las empresas pertenecientes al índice S&P 500 señala la importancia capital del *analytics*. Numerosas empresas han basado su plan de negocio en ofrecer servicios de análisis de datos para facilitar la toma y gestión de decisiones de terceros. En su artículo *Analytics 3.0*, Thomas Davenport (2013) describe cómo empresas centenarias, tales como UPS o Schneider Electric, han sufrido cambios estructurales para incorporar *analytics*, no solo en el sentido tradicional, para la mejora en la toma de decisiones, sino también para la creación de valor en sus productos o servicios. El análisis de empresas llevado a cabo da fe de esta tendencia. Es más, el análisis no solo revela una mayor adopción, sino que, además, aquellas empresas que implementan *analytics* ven una mejora más pronunciada en sus ingresos que aquellos que no. A pesar de que los resultados reflejan que el *business analytics* tiene un impacto en el resultado final de la empresa, este análisis eleva otras preguntas. ¿Debe tener la empresa cierto tamaño para poder aprovechar eficientemente los resultados que proporciona el *business analytics*? ¿Servirán los modelos actuales para solidificar los cimientos de las nuevas técnicas que aún están por venir?

Una cosa está clara, la tendencia de adopción parece acelerar con el tiempo, y con cada vez más empresas aprovechando las bondades que ofrece el *analytics*, las compañías han de innovar en nuevas formas de explotar su capacidad analítica. Actualmente se desconoce cómo o si la actual etapa de *analytics* será la última. No obstante, las empresas han de permanecer flexibles a los cambios y adaptarse a los mismos con rapidez para así mantener su competitividad en un entorno cambiante. Hoy más que nunca es vital para las empresas entender cómo aprovechar las posibilidades del *analytics*. A pesar de que el análisis propuesto ha arrojado luz sobre el impacto de esta nueva tecnología en el entorno empresarial, es necesario expandir el estudio, dando acceso a información de calidad. A través de este análisis se puede observar una pequeña introducción al concepto, pero conviene ampliar las fronteras del entendimiento. Futuros estudios con mayor acceso a datos más sensibles, tales como la inversión en *analytics* de las distintas organizaciones, y a través de una muestra de empresas

mucho mayor, deberán atender a las cuestiones planteadas anteriormente. De esta manera será potencialmente viable establecer el crecimiento de cualquier empresa en base a su inversión en *analytics*.

REFERENCIAS BIBLIOGRÁFICAS

- [1] (2016). *Data protection in the EU*. European Union. Retrieved from: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en
- [2] (2020). *Analytics. Driving insights from data*. Netflix Research. Retrieved from: <https://research.netflix.com/research-area/analytics>
- [3] (2019). RightScale 2019 State of the Cloud Report. Flexera Software. Retrieved from: <https://www.statista.com/statistics/511508/worldwide-survey-public-coud-services-running-applications-enterprises/>
- [4] Asllani, A. (2014). *Business Analytics with Management Science Models and Methods*. FT Press.
- [5] Berinato, S. (2019). Data science & the art of persuasion. *Harvard Business Review*, 97(1), 126-137.
- [6] Cadwalladr, C., & Graham-Harrison, E. (2018). *The Cambridge analytica files*. The Guardian, 21, 6-7.
- [7] Chae, B. K., Yang, C., Olson, D., and Sheu, C. (2014). *The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective*. *Decision Support Systems* (59), pp 119-126.
- [8] Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). *Business Intelligence and Analytics: From Big Data to Big Impact*. *MIS quarterly* (36:4), pp 1165-1188.
- [9] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209.
- [10] Chen, W. & Quan-Haase, A. (2020). *Big Data Ethics and Politics: Toward New Understandings*. *Social Science Computer Review*, 38(1), 3–9. <https://doi-org.sandiego.idm.oclc.org/10.1177/0894439318810734>
- [11] Craig, T., & Ludloff, M. E. (2011). *Privacy and big data: the players, regulators, and stakeholders*. " O'Reilly Media, Inc."
- [12] Davenport, T. H. (2013). Analytics 3.0. *Harvard business review*, 91(12), 64-72.
- [13] Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.

- [14] Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard business review*, 90(5), 70-76.
- [15] Davis, K. (2012). *Ethics of Big Data: Balancing risk and innovation*. " O'Reilly Media, Inc."
- [16] Departamento de Estudios y Gestión del Conocimiento Fundación Cotec (2017). *Generación de Talento Big Data en España*. Deloitte Digital.
- [17] Dove, E. S. (2018). The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. *Journal of Law, Medicine & Ethics*, 46(4), 1013–1030. <https://doi-org.sandiego.idm.oclc.org/10.1177/1073110518822003>
- [18] Economist, T. (2010). *Data, data everywhere: A special report on managing information*. The Economist.
- [19] Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). *Measuring the effects of business intelligence systems: The relationship between business process and organizational performance*. *International Journal of Accounting Information Systems*, 9(3), 135-153.
- [20] Fernández-Manzano, Eva-Patricia (2016). *Big data. Eje estratégico en la industria audiovisual*. Editorial UOC.
- [21] Germann, F., Lilien, G. L., and Rangaswamy, A. (2013) *Performance implications of deploying marketing analytics*. *International Journal of Research in Marketing* (30:2), pp 114-128.
- [22] Hill, C. W., Jones, G. R., & Schilling, M. A. (2014). *Strategic management: Theory & cases: An integrated approach*. Cengage Learning.
- [23] Hürtgen, H. & Mohr, N. (2018). *Achieving business impact with data*. Digital McKinsey.
- [24] Kellen, V., & Wolf, B. (2003). Business performance measurement. *Information Visualization*, 1(312), 1-36.
- [25] Kohavi, R., Rothleder, N. J., and Simoudis, E. (2002) *Emerging trends in business analytics*. *Communications of the ACM* (45:8), pp 45-48.
- [26] Lohr, S. (2012). *The age of big data*. *New York Times*, 11(2012).
- [27] Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

- [28] Markman, J. (2019). *Netflix Harnesses Big Data To Profit From Your Tastes*. Forbes Magazine. Retrieved from: <https://www.forbes.com/sites/jonmarkman/2019/02/25/netflix-harnesses-big-data-to-profit-from-your-tastes/#1551d26466fd>
- [29] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). *Big data: the management revolution*. Harvard business review, 90(10), 60-68.
- [30] Neely, A. (Ed.). (2007). *Business performance measurement: Unifying theory and integrating practice*. Cambridge University Press.
- [31] Provost, F., & Fawcett, T. (2013). *Data science and its relationship to big data and data-driven decision making*. Big data, 1(1), 51-59.
- [32] Russom, P. (2011). *Big data analytics*. TDWI best practices report, fourth quarter, 19(4), 1-34.
- [33] Song, B. (2018). *Big Data As the Next Public Good*. NPQ: New Perspectives Quarterly, 35(3), 12–14. <https://doi-org.sandiego.idm.oclc.org/10.1111/npqu.12153>
- [34] Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., & Philippart, L. (2014). Using big data to improve customer experience and business performance. *Bell labs technical journal*, 18(4), 3-17.
- [35] Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*, 62(3), 347-358.
- [36] Vega Vilca, J. C., & Guzmán, J. (2011). Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple. *Revista de Matemática: Teoría y Aplicaciones Vol. 18 Núm. 1 2011*.
- [37] Vujanic, A. (2008) *Most U.S. Companies Say Business Analytics Still Future Goal, Not Present Reality*. Accenture
- [38] Wasynczuk, A., Gino, F., & Sameh, K. (2018). *Verisk: Trailblazing in the Big Data Jungle*. Harvard Business School Case 919-014
- [39] Wu, T. (2015) *Netflix's Secret Special Algorithm Is A Human*. The New Yorker. Retrieved from: <https://www.newyorker.com/business/currency/hollywoods-big-data-big-deal>

ANEXO I. Análisis de empresas

*Información obtenida de Yahoo! Finance, así como de web corporativas y casos académicos. *

ResMed Inc. (RMD): ResMed Inc. desarrolla, fabrica y distribuye tanto instrumentos médicos como software basado en la nube para el diagnóstico, tratamiento y gestión de desórdenes respiratorios que comprende enfermedades respiratorias pulmonares crónicas, enfermedades neuromusculares y otras enfermedades crónicas. A través de su segmento de venta de software, ResMed aplica herramientas de business intelligence. El equipo de business intelligence ofrece soporte mediante la producción y visualización de datos para la plataforma de distribución de información. ResMed tiene una posición que les permite obtener información única de sus clientes. Dado que el segmento de *analytics* ofrece soporte para la rama principal de negocio, cabe esperar que se trate de una compañía analítica.

National Oilwell Varco, Inc. (NOV): National Oilwell Varco es una empresa encargada de diseñar, construir, fabricar y vender sistemas, componentes y productos derivados de petróleo y gas. National Oilwell ofrece un servicio que recibe el nombre de GoConnect. Dicho servicio provee datos y análisis a tiempo real a los clientes, que pueden determinar la condición y rendimiento de sus componentes. NOV entiende que el *analytics* predictivo cambia rápidamente, y por esta razón están cometidos a evolucionar sus capacidades analíticas, así como adaptar su gama de productos a este segmento. Por tanto, se trata de una compañía analítica.

Home Depot (HD): The Home Depot es una empresa de bricolaje y ferretería. Opera a través de tiendas físicas en las que venden material de hogar, material de jardinería, productos de decoración y servicios de instalación y mantenimiento del hogar. De acuerdo con el portal digital de Adobe, Home Depot cuenta con un departamento destinado únicamente a business intelligence y analytics en línea que se dedica al análisis de métricas en línea y en tiendas físicas (tales como número de llamadas, actividad web, volumen de devoluciones o número de cancelaciones) con el objetivo de tomar las mejores decisiones que más se ajusten a las necesidades del consumidor.

FirstEnergy Corp. (FE): FirstEnergy es una empresa que, a través de sus subsidiarios, genera y distribuye electricidad a lo largo de los Estados Unidos. Opera a través de la generación de energía hidráulica, gas natural, energía eólica y energía solar. Asimismo, la empresa ofrece productos relacionados con la energía y servicios de distribución eléctrica. A pesar de que la organización ofrece puestos de trabajo para el análisis de datos, no cabe constancia de que se utilice esta información como ventaja competitiva, lo que le clasifica como una empresa de *analytics* localizado.

A.O. Smith Corporation (AOS): A. O. Smith es una organización que fabrica y comercializa calderas de gas y agua, tanques y productos de tratamiento de agua en Norte América, Europa, China e India. A través de perfiles de trabajadores actuales en LinkedIn, se puede observar que la organización posee focos de actividad analítica enfocados principalmente a la recolección y visualización de datos. Bajo estas características se puede clasificar como una organización con aspiraciones analíticas.

Regeneron Pharmaceuticals (REGN): Regeneron es una empresa biofarmacéutica que estudia, desarrolla y comercializa medicamentos para el tratamiento de diversas condiciones a nivel global. Asimismo, está desarrollando productos candidatos para el tratamiento de pacientes con enfermedades oculares, enfermedades alérgicas e inflamatorias, cáncer y enfermedades cardiovasculares. La naturaleza de la organización hace que el análisis de datos sea parte de su ADN. Regeneron cuenta con múltiples líneas que se dedican al análisis de datos

de clientes con diversas condiciones. Dado que su actividad principal no es el análisis y manejo de datos, se clasificará como compañía analítica.

TJX Companies (TJX): TJX es un conglomerado de empresas que opera en Norte América, Europa y Australia bajo diversas tiendas físicas en las que ofrece líneas de ropa familiares, líneas de hogar, mobiliario o instrumentos de cocina, entre otros. La compañía cuenta con un equipo de business intelligence que apoya al equipo comercial mediante el análisis de clientes, pero la falta de información pública lleva a pensar que el desarrollo de estos equipos de análisis es lento.

Walgreens Boots Alliance, Inc. (WBA): Walgreens es una organización basada en la venta de productos farmacéuticos y de bienestar. Walgreens es operada nacionalmente mediante la venta de productos a través de sus tiendas físicas, repartidas por todo Estados Unidos, mientras que a nivel internacional colabora con diversas empresas de seguros de salud. Walgreens lleva incorporando las herramientas de Big Data Analytics desde el año 2014 a través de la colaboración con empresas tecnológicas como Inovalon. Esta integración les permite analizar alrededor de 8.3 mil millones de incidencias médicas que pueden llevar a su línea de seguros médicos. El programa, denominado “Healthcare Clinic,” les permite realizar un servicio totalmente personalizado y ajustado al perfil del cliente.

PG&E Corp. (PCG): Pacific Gas and Electric es una empresa que participa en la comercialización y distribución de gas natural y electricidad en áreas residenciales, comerciales, fábricas y entornos agrícolas en el centro y norte de California. PG&E no cuenta con un departamento especializado, sin embargo, sí que hay constancia de que existan focos localizados de *analytics*. Estos focos trabajan en conjunción con el área operacional a la hora de preparar reportes y visualizaciones.

Estée Lauder Companies, Inc. (EL): Estée Lauder es una compañía cosmética que fabrica y comercializa productos de cuidados cutáneos, maquillaje, fragancias y productos de cuidado del cabello. Estée Lauder ofrece sus productos a través de tiendas físicas, departamentos, aeropuertos, farmacias y outlets, entre otros. Un artículo de The Wall Street Journal en 2019 revela que la compañía había llevado a cabo una remodelación completa de su departamento de IT, pero que aún estaba considerando incorporar *business analytics* a sus líneas de negocio.

Gilead Sciences, Inc. (GILD): Gilead es una empresa biofarmacéutica que estudia, desarrolla y comercializa medicamentos en áreas médicas que no han sido respondidas. De esta forma, Gilead desarrolla tratamientos para el VIH, hipertensión, y actualmente para el COVID-19. Gilead entiende la transformación digital como un proceso evolutivo continuo, donde las plataformas proveen un entendimiento de datos profundo y visualizaciones de datos más intuitivas. Gilead vislumbra un futuro en el que la inteligencia artificial y la inclusión de la automatización faciliten el proceso de desarrollo y la creación de soluciones. Asimismo, en Gilead están comprometidos a desarrollar nuevas herramientas de *analytics*, así como incorporar habilidades de *machine learning* a su portafolio de herramientas actual. Gilead posee una gran capacidad analítica que prometen desarrollar más en el futuro. No obstante, debido al hecho de que no se dedican exclusivamente al análisis de datos y a su comercialización, se clasificará como una compañía analítica.

Illumina, Inc. (ILMN): Illumina es una empresa biotecnológica encargada de secuencias y soluciones analíticas para el análisis genético y genómico. Illumina cuenta con un equipo de desarrollo de herramientas de *business intelligence*, así como de un equipo de *data scientists* en constante expansión. De manera similar a Gilead, Illumina posee herramientas de análisis personalizadas tales como BaseSpace Informatics Suite, cuyo objetivo es acelerar el análisis de datos genómicos. De acuerdo con la web corporativa de Illumina, este software está

disponible para todos aquellos laboratorios que deseen adquirir dicho software. Este hecho hace de Illumina un competidor de *analytics*.

Yum! Brands, Inc. (YUM): Yum! Es un conglomerado empresarial compuesto por algunas de las franquicias de comida rápida más conocidas a nivel internacional, como KFC, Pizza Hut o Taco Bell. Un reporte de 2009 revelaba que Yum! empezaría a colaborar con XFormity Technologies, un referente en *analytics* de restaurantes, con la intención de mejorar las medidas métricas operacionales. No obstante, la falta de información y la naturaleza de la empresa en franquicias apuntan a que estos focos están muy localizados y su desarrollo es lento. Por tanto, se puede clasificar como empresa con *analytics* localizado.

Cerner Corporation (CERN): Cerner ofrece soluciones tecnológicas para seguros médicos a nivel de Estados Unidos e internacionalmente. Entre los servicios ofrecidos se encuentran Cerner Millennium, una arquitectura informática que incluye sistemas de gestión de información clínica y financiera, así como HealtheIntent, una plataforma basada en la nube que incorpora las distintas herramientas y permite una integración acorde al modelo operativo de la empresa. De acuerdo con su web oficial, Cerner opera en el segmento de *analytics* a través de dos soluciones: HealtheAnalytics y HealtheEDW. La primera solución se trata de un paquete de *analytics* que ofrece la capacidad de examinar contenido de la población y asimilarlos a casos conocidos del espectro médico. La iniciativa HealtheEDW identifica patrones y variables ocultas en un set de datos. Dado que Cerner opera y gestiona soluciones analíticas, y las emplea como ventaja competitiva, se le puede clasificar como competidor de *analytics*.

Alliant Energy Corporation (LNT): Alliant Energy es una organización que proporciona electricidad y gas natural a parte de los Estados Unidos, principalmente en Iowa. Alliant ofrece una herramienta de *analytics* en línea que proporciona insights básicas a los clientes que lo utilizan. Entre las soluciones que se ofrecen se incluye coste y energía utilizada, desglose de gastos por categoría e impacto del clima. Este sistema obliga a Alliant a mantener un equipo de analistas e ingenieros de software que mantengan al día el sistema, pero no se llevan a cabo estudios de análisis para la empresa. Por esta razón, la empresa se puede clasificar como empresa con aspiraciones analíticas.

Capital One Financial Corporation (COF): Capital One es un banco que ofrece diversos productos financieros en Estados Unidos, Reino Unido y Canadá. Capital One opera a través de tres segmentos: tarjetas de crédito y banca comercial. En un artículo escrito por el académico Thomas Davenport para la revista Forbes, Davenport pone de manifiesto la capacidad analítica de Capital One desde el año 2002. Davenport recalca específicamente que Capital One se trata de una empresa que compite en *analytics* debido a su incorporación en sus servicios financieros.

Republic Services, Inc. (RSG): Republic Services se encarga de la recogida, transporte y reciclaje de desechos en Estados Unidos. La organización ofrece contenedores de distintos tamaños y funcionalidades a municipios y áreas residenciales. La compañía cuenta con un equipo de *business analytics* encargado de la gestión, implementación y optimización de herramientas de *business analytics* y su integración con las operaciones de la empresa.

General Mills, Inc. (GS): General Mills desarrolla y comercializa productos de consumo alimenticio a nivel internacional. La compañía vende sus productos a través de venta directa en tiendas físicas y a través de intermediarios. Un artículo escrito en 2019 pone de manifiesto que General Mills lleva a cabo labores de análisis a través de Nielsen. Mediante esta colaboración, General Mills busca acelerar el desarrollo de productos y poner soluciones a las deficiencias de la cadena de valor.

Apartment Investment & Management (AIV): Apartment Investment & Management (Aimco) es un trust de inversión inmobiliaria enfocado en la adquisición y gestión de comunidades residenciales de calidad localizadas en áreas selectas en Estados Unidos. No hay información pública o anuncios por parte de la empresa que indiquen de la existencia de un equipo de *business analytics* o de focos localizados dentro de la compañía.

Netflix, Inc. (NFLX): Netflix, Inc. es una empresa de suscripción de entretenimiento digital. La compañía ofrece una vasta gama de series de televisión, documentales y películas que varían en género e idiomas. A pesar de que la compañía se constituyó originalmente como un servicio de alquiler de películas y series en formato físico, Netflix aplica el uso de la tecnología de *business analytics* a todas sus áreas funcionales en su actividad regular. Desde su inicio, Netflix ha sido una compañía pionera en la toma de decisiones basadas en datos. De acuerdo con la propia compañía, el modus operandi de Netflix consiste en intentar solucionar cuestiones empresariales ambiguas sumergiéndose en enormes colecciones de datos complejos. Su nivel de integración es tal que opera de manera transdepartamental, compartiendo los descubrimientos y valoraciones de nuevas oportunidades a lo largo de las distintas áreas funcionales de la empresa, creando nuevos segmentos de crecimiento y siempre dando prioridad a los datos. Esto es posible gracias a un equipo entrenado en *data science*, así como a través del desarrollo de herramientas analíticas.

Digital Realty Trust, Inc. (DRT): Digital Realty ofrece soporte de centros de datos e interconexión de estrategias a sus clientes a lo largo del continente americano. Sus productos varían desde servicios de soporte en la nube a soporte de redes de servicios financieros, energía, seguros de salud y productos de consumo. Dado que la compañía ofrece soporte y soluciones de datos y *business analytics* como principal actividad, se le puede reconocer como *competidor de analytics*.

Church & Dwight Co, Inc. (CHD): Church & Dwight es una empresa que comercializa productos domésticos y de cuidado íntimo en Estados Unidos e internacionalmente. En 2012, se reveló que Church & Dwight no lleva a cabo sus operaciones de *business analytics*, si no que las subcontrata a SAP. No obstante, son conscientes de la importancia de los datos como fuente de venta y su necesidad de análisis. Desde entonces, Church & Dwight se ha comprometido a desarrollar una reestructuración completa de su política de IT. Diversas ofertas de trabajo han puesto de relieve la existencia de posiciones de *business intelligence* dentro de la compañía. La falta de información adicional al respecto lleva a pensar que hay focos localizados pero su desarrollo es lento.

Caterpillar, Inc. (CAT): Caterpillar fabrica y vende equipo de construcción y minería, motores diésel y de gas natural y turbinas industriales. De acuerdo con su página corporativa, Caterpillar dispone de “laboratorios de innovación de datos” a través de los que integran las herramientas de *business analytics* con el desarrollo de productos para ofrecer la mejor experiencia personalizada para cada cliente. A través del uso del *analytics* predictivo pretenden adoptar una posición proactiva y adelantarse a las necesidades de mantenimiento de sus productos.

Halliburton Company, Inc. (HAL): Halliburton es una empresa que provee servicios y productos a una red internacional de compañías petrolíferas y de gas natural. Halliburton implementa el análisis de datos en cada una de las fases de sus operaciones. De acuerdo con el chief data scientist de Halliburton, a pesar de que cada fase emplea *analytics*, algunas fases se ven amplificadas por estas herramientas, mientras que otras dependen en su totalidad del *data analytics*. Estas herramientas se integran a través de la plataforma conocida como Landmark.

Fiserv, Inc. (FISV): Fiserv provee servicios de tecnología financieros a nivel internacional. Fiserv incorpora el data analytics en sus servicios ofrecidos. A través de su línea “Information Management Solutions,” Fiserv colecciona, organiza y gestiona datos procedentes del cliente. Con estos datos, Fiserv lleva a cabo análisis que les permite proporcionar al cliente soluciones basadas en su información, insights para mejorar el rendimiento económico, mejorar la eficiencia y reducir el riesgo.

ConocoPhillips (COP): ConocoPhillips es una empresa productora, transportadora y comercializadora de petróleo y gas natural. En sus propias palabras, ConocoPhillips se está convirtiendo en una superpotencia en materia de *data analytics* con más de cuatro mil profesionales destinados únicamente a esta rama y cientos de aplicaciones. Algunos casos de uso incluyen mejorar la eficiencia en operaciones, minimizar el tiempo de perforación, optimizar sus diseños y aumentar el conocimiento sobre las características subterráneas. Asimismo, ConocoPhillips entiende el business analytics como un proceso evolutivo que están dispuestos a seguir y expandir.

Deere & Company (DE): Deere & Co. fabrica y distribuye equipamiento agrícola y de construcción. Entre el equipamiento que distribuye se encuentran tractores, colectores de algodón y cosechadores de azúcar. La compañía también cuenta con una rama de servicios financieros a través de la cual financian alquileres y ventas de su equipamiento. En términos de análisis de datos, John Deere integra herramientas de *big data* junto a su equipamiento que permite a los granjeros obtener información en tiempo real sobre el rendimiento de su equipamiento. Todos los datos se encuentran conjuntos de manera personalizada en el portal en línea myjohndeere.com. Adicionalmente, Deere & Co. cuenta con el Centro de Operaciones John Deere, que ayuda a los granjeros a formar decisiones relacionadas con el tipo de cosecha y la idoneidad del terreno.

Sealed Air Corporation (SEE): Sealed Air proporciona productos de seguridad alimenticia, así como soluciones protectivas. Dentro de su línea de seguridad alimenticia ofrecen empaquetado integrado, y soluciones automáticas que alargan la vida del producto. De ofertas de puestos de trabajo se puede extraer que la compañía posee un equipo de *analytics*, pero destinados únicamente a mejorar los procesos internos.

Verisk Analytics, Inc. (VRSK): Verisk Analytics, Inc. es una empresa encargada de proveer soluciones de *data analytics* a empresas a nivel internacional. Entre la gama de servicios ofrecidos, Verisk favorece los análisis predictivos y da soporte de toma de decisiones a empresas en un amplio rango de ámbitos.

BorgWarner, Inc. (BWA): BorgWarner es una empresa que ofrece soluciones para vehículos híbridos, eléctricos y de combustión a nivel internacional. Actualmente, BorgWarner cuenta con un departamento de Business Analytics destinado a la mejora en la optimización de proyectos internos, así como la mejora la productividad.

Host Hotels & Resorts Inc. (HST): Host Hotels & Resorts es el mayor conglomerado inmobiliario del S&P 500, propietario de hoteles internacionales de alta gama. Es una compañía pionera en su segmento por su fuerte apuesta en *analytics* predictivo a través de la colaboración con IBM. El objetivo de esta colaboración es unir la capacidad analítica de IBM con la plataforma de datos de Host para el desarrollo de modelos predictivos que mejoren las decisiones de inversión y extraer *insights* de datos sin estructurar.

Oracle Corporation (ORCL): Oracle es una compañía que ofrece productos y servicios para el entorno informático empresarial a nivel global. Parte de su gama de productos incluye software de *Business Analytics* que facilita a las empresas implementar el análisis de datos en sus operaciones de manera sencilla y orgánica con su actividad regular.

ONEOK, Inc. (OKE): Oneok es una compañía que lleva a cabo tareas de recogida, procesamiento, almacenamiento y transporte de gas natural en Estados Unidos. No cabe constancia de que Oneok cuente con un departamento o focos aislados de *analytics*.

Xylem, Inc. (XYL): Xylem se encarga del diseño y servicio de soluciones de agua y residuos acuáticos internacionalmente. Xylem divide su actividad en tres ramas principales: infraestructura, soluciones de medida y de control. Como parte de su familia de productos, Xylem ofrece soluciones inteligentes ante los desafíos de gestión de agua, Xylem Analytics. Esta gama está caracterizada por el uso de *analytics* para el estudio de soluciones personalizados a todo tipo de desafíos de gestión.

The Williams Companies, Inc. (WMB): Williams Companies es un conglomerado de empresas que opera como una compañía de infraestructura energética que opera principalmente en Estados Unidos. No cabe constancia de que lleven a cabo actividades de *analytics* o que existan focos localizados dentro de la empresa.

Boston Scientific Corporation (BSX): Boston Scientific desarrolla, fabrica y comercializa instrumentos médicos para intervenciones quirúrgicas a nivel internacional. En un artículo de prensa publicado por la propia compañía en 2018 se pone de manifiesto la importancia del *big data* dentro de la empresa. El presidente de la rama de Investigación y Desarrollo define su uso de *big data* como el uso de tecnología para analizar grandes cantidades de datos provenientes de distintos proveedores médicos, así como de los propios instrumentos de Boston Scientific y transformarlos en “inteligencia accionable.” Por tanto, se puede caracterizar como una compañía analítica.

Newell Brands, Inc. (NWL): Newell Brands diseña, fabrica y distribuye productos comerciales a nivel mundial. Entre su gama de productos se encuentran productos de cocina, productos domésticos y productos infantiles. A través de diversas ofertas de trabajo, se puede deducir que Newell Brands cuenta con un pequeño equipo de *business intelligence*, que colabora con el equipo de *supply chain* para optimizar el proceso de cadena de valor, lo que le convierte en una compañía con aspiraciones analíticas.

CMS Energy Corporation (CMS): CMS Energy es una empresa energética que opera principalmente en el estado de Michigan, Estados Unidos a través de cuatro segmentos: electricidad, gas, soluciones para empresas y EnerBank. Se puede deducir, a través de distintas ofertas de trabajo que, CMS cuenta con un pequeño equipo de *business intelligence* integrado en el análisis de clientes que ayuda a la toma de decisiones inteligentes en relación a clientes.

Stryker Corporation (SYK): Stryker es una compañía tecnología médica que comercializa productos ortopédicos, equipo quirúrgico y productos neurotecnológicos. De forma similar a Host Hotels & Resorts, Stryker colabora con IBM, quien les ofreció un equipo de gestión de rendimiento para controlar y mejorar la captura de clientes. Gracias a su colaboración con IBM, Stryker tiene acceso a análisis de datos en tiempo real. No obstante, no cabe constancia de que Stryker tenga un equipo propio de *analytics*.

United Parcel Service, Inc. (UPS): United Parcel Services, también conocida como UPS, es una compañía que ofrece servicios de transporte, logística y envíos de cartas y paquetes a nivel internacional. Un comunicado de prensa emitido por UPS en 2019 señala la importancia de la implementación de herramientas de *business analytics* en una empresa como UPS. Explican que la naturaleza de la empresa les obliga a analizar datos para mejorar áreas críticas en la entrega de pedidos. Es por esta razón que UPS incorpora *analytics* en todas las etapas de su actividad.

FMC Corporation (FMC): FMC es una compañía que ofrece soluciones agrarias basadas en la protección de cosechas, salud del terreno y control de plagas. FMC cuenta con focos dentro

de la empresa destinados a la recopilación de datos de sentimiento de mercado, análisis de comportamiento de clientes y análisis de tendencias agroquímicas. No obstante, parece que los focos de *analytics* están muy localizados y operan de manera lenta.

Harley Davidson, Inc (HOG): Harley Davidson es una fabricante de motocicletas establecido en Estados Unidos y opera a nivel internacional. Un artículo de Harvard Business Review pone de manifiesto cómo Harley Davidson integró un software de machine learning para hacer crecer sus ventas en un 2.930% en el año 2016. El artículo explica cómo Harley Davidson incorpora software que analiza datos de clientes existentes, identificando patrones de comportamiento. Una vez ha identificado estos patrones, el software identifica perfiles de clientes similares, así como elementos que tienen éxito y elementos que no, creando la fórmula perfecta. A través de este software, Harley Davidson es capaz de conseguir una ratio de ventas constante.

Universal Health Services, Inc. (UHS): Universal Health Services es una empresa que ofrece apoyo a hospitales de cuidado, así como a instalaciones de salud cognitiva. El grado de incorporación de *analytics* es bajo, destinado principalmente al análisis de nuevos potenciales clientes.

Equinix, Inc. (EQIX): Equinix es una empresa de servicios que ofrece centros de datos a sus clientes para conectar a los líderes empresariales con sus consumidores.

United Rentals, Inc. (URI): United Rentals es una de las mayores empresas de alquiler de equipamiento del mundo. Sus dos mayores segmentos están concentrados en el alquiler de materiales y equipamiento de construcción. Como sucede con otras empresas del mismo sector, United Rentals delega sus actividades de *analytics* a empresas especializadas, en este caso Teradata.

General Electric Company (GE): General Electric es una compañía que opera a nivel internacional a través de distintos segmentos dentro del ámbito de la tecnología punta. Entre sus productos se encuentra GE Digital, donde ofrecen soluciones digitales a empresas que afrontan los desafíos del presente. Uno de sus principales productos dentro de esta rama es Predix Platform, una plataforma de software que ponen a servicio de las empresas capaz de llevar a cabo tareas de *machine learning* y *analytics* avanzado.

FLIR Systems, Inc. (FLIR): FLIR Systems es una organización que diseña, desarrolla y comercializa sistemas de reconocimiento térmico de imágenes, sistemas de medición y diagnóstico y detección de amenazas. Dentro de su segmento comercial, FLIR ofrece una gama de productos basados en el *analytics* de consumidores. Para ello, FLIR pone a la venta un set de cámaras de reconocimiento que, junto con su software de *analytics*, es capaz de definir perfiles de clientes en tiempo real y así poder crear experiencias más personalizadas para el consumidor.

Molson Coors Beverage Company (TAP): Molson Coors es un conglomerado de empresas encargado de producir y comercializar cerveza y otras bebidas de malta a nivel internacional. Molson Coors cuenta con un departamento destinado a *business intelligence*. Un artículo publicado en SAP en el que se entrevista a Alexandre Rivet, explica el viaje que ha experimentado la empresa en el ámbito tecnológico. En dicho artículo, Rivet explica cómo se ha favorecido la empresa de la colaboración con SAP. A través de las herramientas ofrecidas por SAP, Coors ha acelerado el proceso de recogida y limpieza de datos. Rivet también señala cómo se han visto mejoradas sus ventas gracias a la información recopilada de cada una de las ventas.

eBay, Inc. (EBAY): eBay es una empresa digital que conecta a compradores y vendedores a nivel mundial. eBay cuenta, a día de hoy, con un equipo completo de big data analytics. Su

importancia dentro de la empresa es tal que Seshu Adunuthula, director de infraestructura de analytics en eBay, explica que los datos son su activo más importante. Adunuthula explica cómo los datos de más de diez millones de compradores y vendedores semanales potencian, por ejemplo, un estudio de cómo eBay puede apoyar a sus vendedores durante el proceso de venta.

Salesforce.com, Inc. (CRM): Salesforce es una compañía que desarrolla soluciones de computación en la nube con un foco en gestión de relaciones con consumidores. Salesforce ofrece soluciones completas a problemas de datos. Su gama de productos varía desde el almacenamiento de datos a software de análisis en tiempo real. Sus soluciones de *analytics* están completamente personalizadas a las necesidades de datos de las empresas.

The Walt Disney Company (DIS): The Walt Disney Company es una organización que opera en el segmento del entretenimiento bajo distintas cadenas de televisión, así como parques, experiencias y productos. La compañía ha llevado a cabo una fuerte inversión en *business analytics* en los últimos años con el objetivo de mejorar la experiencia del consumidor a lo largo de sus áreas de negocio. Teddy Benson, director de integración de producto explica que el uso de *analytics* es integral para mejorar la experiencia del consumidor. El equipo de *analytics* está comprendido por más de mil empleados, divididos en distintos focos. Benson explica que el equipo lleva a cabo tareas de *analytics* de distinta índole, desde incorporación de datos a reconocimiento facial.

Valero Energy Corporation (VLO): Valero es una empresa energética que ofrece producción, venta y transporte de productos petroquímicos a nivel internacional. En este sentido, Valero es una empresa pionera dado que decidió por apostar por el potencial de las herramientas de *business intelligence* desde principios de los 2000. Distintas ofertas de trabajo revelan que Valero emplea el *business analytics* para mejorar deficiencias en el proceso de cadena de valor y para la mejora de rendimiento comercial.

Waters Corporation (WAT): Waters Corp es una empresa de medición especializada de flujos líquidos que provee soluciones analíticas a las empresas. De manera similar a Xylem, Waters ofrece una gama de hardware y software analíticos.

DaVita, Inc. (DVA): DaVita es una empresa de servicios médicos que ofrece diálisis para pacientes que sufran fallo renal o que se encuentren en fase terminal de fallo renal. DaVita lleva a cabo sus operaciones a través de centros de diálisis y servicios de laboratorio relacionados. DaVita implementa herramientas de *analytics* predictivo para estudiar el riesgo de sus clientes y así reducir el número de hospitalizaciones. Esto, además, es complementado con sistemas de inteligencia artificial que ayudan a realizar recomendaciones a los pacientes.

PACCAR, Inc. (PCAR): PACCAR es una empresa que produce, distribuye y comercializa camiones ligeros, medios y pesados a nivel internacional. PACCAR cuenta con focos dispersos de analytics a lo largo de la empresa que operan bajo el paraguas del departamento de IT. PACCAR define el Centro de la Excelencia de Analytics Avanzado como un centro en el que se llevan a cabo ideas impulsadas por datos y que son llevada a IT para su estudio y posterior implementación.

Ventas, Inc. (VTR): Ventas es un trust inmobiliario que cuenta con aproximadamente 1200 activos en Estados Unidos, Reino Unido y Canadá. Ventas delega sus tareas de Business Analytics a una empresa consultora llamada Tiger Analytics, que cuenta con 250 empleados. Ventas dedica una gran parte de su actividad a la recogida masiva de datos. El análisis big data le ha permitido a Ventas llevar a cabo predicciones sobre dónde desarrollar o comprar nuevas comunidades, qué tipo de arrendamiento emplear o incluso qué operador escoger.

Boston Properties, Inc. (BXP): Boston Properties es uno de los trust inmobiliarios más grandes de Estados Unidos. Boston Properties desarrolla, gestiona, y posee un portafolio diverso de espacios de oficinas. De manera similar a otros trust inmobiliarios, Boston Properties colabora con una empresa que les ofrece una plataforma de analytics, Enel X, a través de la cual BXP estudia cómo reducir su consumo de energía además de ofrecer otros insights relacionados con eficiencia energética.

Conagra Brands, Inc. (CAG): Conagra Brands es una empresa que produce y comercializa productos alimenticios. La página corporativa de Conagra pone de relieve la existencia de un equipo de Insights & Analytics dentro de la empresa. El objetivo de este equipo es el de apoyar al cuerpo de gobierno corporativo en la toma de decisiones y la optimización del proceso de cadena de valor.

Helmerich & Payne, Inc. (HP): Helmerich & Payne opera principalmente en la perforación de gas y petróleo. Helmerich & Payne posee un equipo de data scientist que operan bajo el paraguas del departamento de Marketing.

Altria Group, Inc. (MO): El Grupo Altria es un conglomerado de empresas que produce y comercializa productos derivados del tabaco y vino internacionalmente. Altria cuenta con data scientists que operan dentro de otros grupos más amplios, por lo que el impacto es menor, y su labor se reduce a la de dar apoyo a estos departamentos.

Kohl's Corporation (KSS): Kohl's es una empresa que ofrece una gran variedad de productos a través de sus grandes almacenes a lo largo de los Estados Unidos. Se puede observar, a través de su web corporativa, que Kohl's está llevando a cabo una fuerte inversión en tecnología de inteligencia artificial. A través de las herramientas de machine learning, Kohl's logra extraer insights de cada una de las 1100 tiendas con las que cuenta, contribuyendo a crear una experiencia única para el consumidor.

Xilinx, Inc. (XLNX): Xilinx diseña y desarrolla instrumentos programables y tecnología asociada a nivel mundial. Como parte del software que Xilinx pone a disposición de sus clientes se encuentra Xelera Analytics. Se trata de un paquete de con soporte en la nube a través del cual el consumidor puede acceder al potencial de las herramientas de machine learning sin necesidad de poseer un gran hardware que lo acompañe, gracias al uso de la nube.

Consolidated Edison, Inc. (ED): Consolidated Edison es un conglomerado empresarial que distribuye electricidad, gas y vapor. La empresa es pionera en su segmento gracias a la fuerte apuesta por el analytics predictivo basado en el IoT (Internet de las cosas). El objetivo es operacionalizar los datos presentados en los más de cinco millones de sensores de los que dispone Consolidated Edison. Esto les permite reducir sus costs operativos y mejorar la experiencia de usuario.

Fortune Brands Home & Security, Inc. (FBHS): Fortune Brands Home & Security produce y comercializa productos de seguridad del hogar, productos de reparación del hogar, remodelado y aplicaciones de seguridad. No cabe constancia de que Fortune Brands haga un uso activo y continuado de las herramientas de analytics.

Alexandria Real Estate Equities, Inc. (ARE): Alexandria Real Estate Equities es un trust inmobiliario que cuenta con numerosas propiedades a lo largo de Estados Unidos. No cabe constancia de que Alexandria Real Estate Equities haga un uso activo y continuado de las herramientas de analytics.

Laboratory Corporation of America Holdings (LH): Laboratory Corporation of America es un holding que opera a través de laboratorios clínicos independientes a nivel internacional. Parte de su compromiso de excelencia implica que cada ensayo clínico cumple con los más

exactos requisitos. Sus soluciones analíticas y de business intelligence ofrecen a las empresas la seguridad de decisiones exactas basadas en los datos.

Union Pacific Corporation (UNP): Union Pacific Corporation es una empresa que, a través de Union Pacific Railroad, lleva a cabo servicios relacionados con el ferrocarril en Estados Unidos. A través de la propia web corporativa se puede observar que Union Pacific recopila miles de datos cada día y los incorpora en un modelo que, en colaboración con los ingenieros, predice las vías que potencialmente corren más riesgo. En este sentido, Union Pacific cuenta con un amplio equipo de ingenieros y data scientists que actualizan el modelo.

Accenture plc (ACN): Accenture es una empresa que ofrece servicios de consultoría, tecnología y outsourcing de servicios. Como parte de sus servicios de tecnología, Accenture ofrece a las instituciones la posibilidad de llevar a cabo análisis y extraer datos basados en distintos modelos e inteligencia artificial.

Colgate-Palmolive Company (CL): Colgate-Palmolive es una compañía que produce y comercializa productos domésticos a nivel internacional. A través de múltiples ofertas de empleo se puede observar que Colgate cuenta con un equipo de data scientists destinado a optimizar el proceso de cadena de valor, así como dar soporte al diseño y comercialización de los distintos productos.

Edwards Lifesciences Corporation (EW): Edward Lifesciences es una empresa tecnológica que comercializa productos relacionados con enfermedades cardiovasculares, así como instrumentos quirúrgicos a nivel internacional. Diferentes ofertas de empleo ponen de relieve que Edwards Lifesciences cuenta con distintos focos de analytics que llevan a cabo funciones tal como ofrecer soporte a productos existentes o monitorizar el inventario.

Aptiv PLC, Inc. (APTV): Aptiv es una empresa que diseña, fabrica y comercializa componentes de vehículos a nivel mundial. De esta manera, Aptiv ofrece soluciones eléctricas, electrónicas y de seguridad en los mercados automovilísticos. Aptiv ofrece servicios de datos como parte de su rama de soluciones tecnológicas. A través del software de Aptiv, los consumidores tienen acceso a análisis de distintos aspectos del vehículo.

L Brands, Inc. (LB): L Brands comercializa productos textiles íntimos femeninos, productos cosméticos y accesorios. L Brands cuenta con un departamento de *analytics* denominado DS&A. El equipo de DS&A ofrece soporte a los distintos departamentos a través de distintos proyectos relacionados con data science.

Fidelity National Information Services, Inc. (FIS): Fidelity National opera dentro del Mercado de servicios financieros a nivel internacional a través de tres segmentos distintos: Soluciones Comerciales, Soluciones de Banca y Soluciones de Mercado Capital. Fidelity cuenta con un equipo dedicado. Tara Bonell, vicepresidenta de soluciones de datos, explica cómo Fidelity national analiza la ingente cantidad de datos generados para detectar fraude, implementar la automatización de ciertos procesos o para el análisis de préstamos. Asimismo, Bonell pone de manifiesto la importancia de mantener un objetivo común.

Monster Beverages Corporation (MNST): Monster Beverage es un conglomerado que desarrolla y comercializa bebidas y concentrados a nivel internacional. No cabe constancia de que Monster Beverage cuente con un equipo dedicado de analistas, aunque si cuentan con distintas posiciones aisladas que ofrecen apoyo a las operaciones de la empresa.

PVH Corporation (PVH): PVH es una empresa que comercializa productos textiles a nivel internacional a través de tres segmentos: Tommy Hilfiger, Calvin Klein y Heritage Brands. Los data scientists en PVH operan bajo el paraguas del equipo de Business Intelligence & Database.

Dicho equipo lleva a cabo múltiples proyectos para la plataforma de Big Data y Analytics de PVH.

Devon Energy Corporation (DVN): Devon Energy es una empresa energética que se encarga de la exploración, desarrollo y producción de petróleo en Estados Unidos. De acuerdo con Harvard Business, Devon Energy es una empresa pionera en su sector por su fuerte inversión en iniciativas impulsadas por los análisis de datos en el marco de descubrimiento, perforación y extracción de petróleo. A través de múltiples sensores, Devon es capaz de recoger datos sobre numerosas métricas, que son analizadas a través de distintos modelos.

Nielsen Holdings plc (NLSN): Nielsen es un holding que opera como una compañía de medición y data analytics.

Norwegian Cruise Line Holdings (NCLH): Norwegian Cruise Line es un holding que cuenta con una diversa línea de cruceros que operan a nivel global. Su uso de data analytics se limita al análisis y pronóstico de inventario, así como para el análisis de clientes bajo el paraguas del departamento de marketing.

Occidental Petroleum Corporation (OXY): Occidental es una empresa que lleva a cabo actividades relacionadas con la adquisición, exploración y desarrollo de propiedades de petróleo y gas natural. La propia web corporativa de Occidental revela que, ya en 2017 anunciaron la adopción de métodos de colección de datos y de visualización, así como someter el sometimiento de datos bajo marcos de métodos computacionales. Los insights extraídos por el conjunto de métodos les permite llevar a cabo decisiones basadas en estos datos.

International Business Machines Corporation (IBM): IBM es una compañía que comercializa soluciones integradas y servicios tecnológicos a nivel global. Uno de los segmentos bajo los que compete IBM es el de analytics. IBM ofrece distintos paquetes de analytics que incorporan Inteligencia Artificial para desarrollar patrones que normalmente no serían visibles.

Public Storage (PSA): Public Storage es una empresa inmobiliaria dedicada a la adquisición, Desarrollo y operación de instalaciones de almacenamiento en Estados Unidos. No cabe constancia de que Public Storage cuente con un equipo de business analytics.

EOG Resources, Inc. (EOG): EOG Resources es una empresa petroquímica que desarrolla actividades relacionadas con la exploración, extracción y comercialización de petróleo y gas natural. EOG cuenta con acceso a análisis de datos en tiempo real desde 2015. Sandeep Bhakhri, director de información y tecnología, explica que los datos son el mayor recurso con el que la compañía cuenta. EOG recolecta datos de los más de 5000 pozos en Estados Unidos. Esto les permite un acceso a yacimientos más exactos y extracciones más rápidas.

Apple, Inc. (AAPL): Apple es una empresa tecnológica dedicada al diseño, producción y comercialización de smartphones, ordenadores, tablets, accesorios y servicios relacionados. Apple analiza los datos recogidos de los miles de millones de dispositivos en manos de los usuarios. El equipo de analytics y big data se encarga de llevar a cabo análisis sobre estos datos con el objetivo de encontrar patrones de uso y nuevas tendencias entre los usuarios. Asimismo, Apple colabora con IBM para los datos de salud de sus dispositivos y así proponer nuevas soluciones y mejorar el entendimiento del usuario sobre su propia salud.

The Kraft Heinz Company (KHC): Kraft Heinz es una empresa que produce y comercializa productos alimenticios y bebidas a nivel internacional. Hasta 2012, Kraft contaba con ingentes cantidades de datos que no usaba eficientemente. Kraft recoge datos a través de diversas plataformas. Estos datos son procesados y convertidos en insights, lo que les permite

desarrollar perfiles únicos de usuarios y así segmentar sus productos de manera individual en lugar de en masa.

Automatic Data Processing, Inc. (ADP): Automatic Data Processing (ADP) es una empresa que ofrece gestión de capital a empresas a través de servicios en la nube. A través de su ADP Research Institute llevan a cabo análisis predictivos de los empleados de la empresa y ofrecen insights tales como el movimiento de personal estimado.

Amgen, Inc. (AMGN): Amgen es una empresa biotecnológica que desarrolla, produce y comercializa productos terapéuticos a nivel global. De acuerdo con su web corporativa, Amgen incorpora data analytics a través de su equipo de Digital Health & Innovation. De esta manera, Amgen es capaz de mejorar la efectividad de departamentos tales como marketing, producción o I+D.

United Continental Holdings (UAL): United Airlines es una compañía que ofrece servicios de transporte aéreo a nivel internacional. United Airlines cuenta con un equipo encargado de llevar a cabo las tareas de analytics, Enterprise Analytics Team. Para ello, United cuenta con los servicios de empresas como Teradata y Hortonworks, los cuales proveen el almacenamiento de ingentes cantidades de datos.

D. R. Horton, Inc. (DHI): D.R. Horton es una empresa constructora de viviendas que opera en gran parte de Estados Unidos. D.R. Horton cuenta con algunos focos de analytics aislados que operan bajo otros departamentos, pero no cabe constancia de que formen parte de sus operaciones.

Alaska Air Group, Inc. (ALK): Alaska Air es una compañía que ofrece servicios de transporte aéreo de pasajeros y mercancía en Estados Unidos, México, Canadá y Costa Rica. Alaska Air es un referente en eficiencia gracias, en parte, a su uso de datos. De esta manera, Alaska basa sus decisiones estratégicas en el análisis derivado de los datos de rendimiento. Esto les ha permitido alcanzar una excelente eficiencia operacional.

The Sherwin-Williams Company (SHW): Sherwin-Williams es una compañía que desarrolla, produce y distribuye pinturas, revestimientos y productos relacionados de uso profesional. Diversas ofertas de empleo revelan que Sherwin Williams cuenta con un equipo de Business Intelligence/Data Warehouse. La función de este equipo es la de dar soporte a la aplicación y optimización de soluciones relacionadas con las distintas áreas funcionales de la empresa.

Charter Communications, Inc. (CHTR): Charter Communications es una empresa que ofrece servicios de televisión e Internet a clientes comerciales y residenciales en Estados Unidos. A través de su subsidiario Spectrum, Charter Communications ofrece un paquete de conexión a sus clientes especialmente para aquellos usuarios que requieran de business analytics en sus negocios. Charter no ofrece analytics a sus clientes, sino que ofrece una conexión que permite a sus clientes atender a las necesidades técnicas del big data analytics. Por esta razón, dado que la empresa opera para este sector, será considerada como competidor de analytics.

Amazon.com, Inc (AMZN): Amazon.com es una empresa de comercio electrónico y suscripciones que opera a nivel internacional a través de tres segmentos: América del Norte, Amazon Internacional y Amazon Web Services. Amazon es una empresa pionera en el uso de *big data analytics*. Amazon.com almacena información extensa sobre cómo interactúan los usuarios con los distintos elementos disponibles a lo largo de la página web. Esta información es analizada y expuesta a distintos modelos de *machine learning*, tales como el de reglas de asociación, el cual estudia los distintos elementos ofrecidos en la web de Amazon, analiza cómo han interactuado los distintos usuarios con estos elementos, y como resultado despliega

las opciones que cuentan con mayor probabilidad de compra en base a aquellos usuarios que interactuaron con ciertos elementos. Gracias a este modelo, Amazon puede ofrecer recomendaciones personalizadas basadas en la actividad de cada usuario. No obstante, a pesar de que este aspecto forma parte de la gran identidad de Amazon, es otro de sus segmentos que realmente le convierten en una empresa competidora de *analytics*, Amazon Web Services. AWS es la vertiente de Amazon que ofrece servicios en la nube a empresas e individuos que varían desde almacenamiento de datos hasta servicios de *machine learning* e inteligencia artificial integrables sin conocimiento previo.

SBA Communications Corporation (SBAC): SBA Communications es una operadora telefónica de que opera en América y Sudáfrica. No cabe constancia de que SB emplee activamente o incorpore herramientas de *analytics* en sus operaciones.

Ingersoll-Rand, Inc (IR): Ingersoll Rand es una empresa que comercializa equipamiento de compresión y productos y servicios derivados a nivel internacional. Ingersoll cuenta con un equipo de profesionales que trabajan bajo el paraguas del equipo de operaciones de Data y Analytics. Dicho equipo ofrece soporte a los distintos departamentos dentro de Ingersoll a través del desarrollo de aplicaciones de business analytics y gestión de almacenamiento de datos.

3M Company (MMM): 3M desarrolla, fabrica y comercializa productos de muy diversa índole a nivel mundial. 3M opera a través de tres segmentos: Seguridad e Industrial, Transporte y Electrónica, Health Care y Consumo. 3M ha experimentado un cambio drástico hacia la integración de *analytics* predictivo. De esta manera, 3M ha sido capaz de incorporar big data analytics en cada uno de sus segmentos operacionales, con importante hincapié en el segmento de Health Care.

Allergan plc (AGN): Allerga es una empresa farmacéutica que desarrolla y comercializa farmacéuticos, instrumentos quirúrgicos y medicina regenerativa a nivel mundial. Sus operaciones están divididas en tres segmentos: Área terapéutica especializada, Medicina General e Internacional. El empleo de *analytics* está dedicado principalmente a la provisión de una dirección estratégica de marca y en el apoyo a la gestión de marcas en el área terapéutica.

Kinder Morgan, Inc. (KMI): Kinder Morgan es una compañía de infraestructura energética que opera principalmente en Estados Unidos. Las tareas de *analytics* se desarrollan principalmente bajo el ámbito del área de riesgos. El equipo de *analytics* se encarga del desarrollo, implementación y mejora de soluciones aplicadas al modelado de análisis cuantitativo de riesgos.

Intel Corporation (INTC): Intel es una empresa que ofrece soluciones computacionales, de red, de almacenamiento de datos y de comunicación a nivel mundial. Intel ofrece distintos paquetes de software de big data analytics que se adaptan a las necesidades de los clientes, permitiendo un análisis rápido, sencillo y profundo.

NiSource, Inc. (NI): NiSource es un holding energético que provee gas natural y electricidad en Estados Unidos. NiSource ha llevado a cabo una integración paulatina de *analytics* desde 2014. Estas actividades se basan principalmente en la prevención de daño producido por excavación.

Starbucks Corporation (SBUX): Starbucks es una compañía que comercializa café y derivados de café a través de sus establecimientos a nivel global. Starbucks recopila datos derivados de las más de noventa millones de transacciones semanales que se producen en sus más de 25.000 tiendas a nivel internacional. El equipo de Data Analytics en Starbucks es el encargado de gestionar esta ingente cantidad de información y transformarla en directrices de marketing, ventas y soporte a la toma de decisiones. Gran parte de la estrategia de expansión

que está siguiendo Starbucks está estrictamente basada en los insights extraídos de los distintos modelos de analytics.

The Western Union Company (WU): Western Union es una compañía que ofrece servicios de pago y movimiento de dinero a nivel mundial. Western Union hace hincapié en la importancia del manejo de datos en una compañía con unas características como las suyas. De acuerdo con su web corporativa, Western Union procesa de media 32 transacciones por segundo, lo que hace necesario un control en tiempo real de los datos gestionados. Western Union emplea data analytics para el análisis en tiempo real de los datos proporcionados por las transacciones y así determinar actividad fraudulenta. El equipo de analytics avanzado es capaz de extraer patrones irregulares dentro de transacciones salientes y entrantes entre Francia y Turquía.

Noble Energy, Inc. (NBL): Noble Energy es una empresa que lleva a cabo actividades relacionadas con la extracción y comercialización de petróleo y gas natural. Noble cuenta con un equipo destinado a tareas de big data analytics. Noble emplea las herramientas de analytics principalmente para la predicción y prevención de horas muertas en sus sistemas y arquitectura.

Adobe Systems, Inc. (ADBE): Adobe es una empresa tecnológica que ofrece una vasta gama de productos de software a nivel global. Adobe ofrece empresas y consumidores paquetes de software que organizan los datos e incorporan herramientas de analytics.

Polo Ralph Lauren Corporation (RL): Ralph Lauren es una empresa que diseña, comercializa y distribuye productos textiles internacionalmente. En 2014, Ralph Lauren decidió adentrarse en el mundo del “Internet de las Cosas” a través de la comercialización de un polo inteligente. Este polo lleva incorporadas directamente a la tela unas fibras que permite a Ralph Lauren recopilar unidades de medición únicas que son analizadas y ofrecen al usuario insights sobre su capacidad y estado físico a través de una app especializada.

Eaton Corporation plc (ETN): Eaton es una compañía de suministro de energía que opera a nivel global. Eaton ofrece, como parte de su gama de productos, paquetes de software de reporte y analytics conocido como Power Xpert Reporting. Este software consolida datos complejos extraídos de una multitud de dispositivos repartidos a lo largo del edificio o infraestructura y los convierte en reportes gráficos de fácil comprensión.

Qualcomm, Inc. (QCOM): Qualcomm es una empresa que diseña, desarrolla, produce y comercializa productos de comunicación a nivel global a través de circuitos integrados y software. Qualcomm es una empresa competidora de analytics dado su compromiso con la producción de microchips que integren tecnología de inteligencia artificial y actúen como centro de datos. La compañía espera que, para 2025, los chips con capacidad de análisis potenciados con inteligencia artificial crecerán en un 100%.

Weyerhaeuser Company (WY): Weyerhaeuser se dedica a la compra y gestión de áreas forestales, así como a la producción de productos derivados de madera. Múltiples ofertas de trabajo revelan que Weyerhaeuser posee diversas posiciones destinadas a data scientists, operando bajo el paraguas del departamento de IT. Estos puestos son dispares y se limitan al uso de modelos analíticos para estudiar la factibilidad de ciertas decisiones, así como para el estudio de la información financiera para la toma de decisiones.

Illinois Tool Work, Inc. (ITW): Illinois Tool Works es una empresa que fabrica y comercializa productos industriales y equipamiento a nivel mundial. No hay evidencia que indique que Illinois Tool hace uso activo de las herramientas de analytics en sus operaciones.

Mettler-Toledo International, Inc. (MTD): Mettler-Toledo es una empresa que fabrica instrumentos de precisión y servicios relacionados a nivel mundial. Parte de su oferta es un paquete de software que ofrece gestión de datos y la extracción de insights.

Waste Management, Inc. (WM): Waste Management ofrece servicios de gestión de residuos a clientes comerciales, residenciales, industriales y municipales en todo Estados Unidos. Waste Management cuenta con un equipo de enterprise analytics y gestión de datos. El objetivo de este equipo es el de mejorar la optimización de logística y transporte. El equipo de analytics es capaz de emplear software Spitfire para mejorar la velocidad de sus operaciones y llegar a más clientes.

Microsoft Corporation (MSFT): Microsoft es una empresa que desarrolla productos tecnológicos y ofrece soluciones y soporte relacionados. Microsoft opera a través de distintos segmentos, entre los que se encuentra su servicio de Intelligent Cloud. A través de dicho segmento, Microsoft ofrece paquetes de analytics a través de Azure, su plataforma en la nube. Estos paquetes permiten aprovechar la capacidad analítica de la propia empresa o aprovechar el potencial de otros servicios enteramente gestionados por Microsoft.

CSX Corporation (CSX): CSX es una empresa que ofrece servicios ferroviarios. CSX recopila los datos generados a través de los numerosos dispositivos que se encuentran a lo largo de sus ferrocarriles y en las vías ferroviarias. El equipo de ingenieros aplica modelos de machine learning sobre estos datos para optimizar tiempos de llegada, mejores rutas o incluso para mejorar la calidad de las inspecciones.

The Boeing Company (BA): Boeing es una compañía que fabrica y comercializa aeronaves militares, aviones comerciales y satélites, entre otros servicios. Boeing ofrece distintos servicios de analytics. A través de Boeing Analytics, Boeing analiza los datos de los clientes y lo complementa con su experiencia en aeronáutica para extraer insights que ayudan a sus clientes en la toma de decisiones. Adicionalmente, Boeing ofrece Analytx, su software de analytics.

AT&T, Inc. (T): AT&T es una empresa de telecomunicaciones que opera a nivel global. AT&T combina un equipo especializado que da servicio de análisis e insights a sus clientes en conjunción con software especializado cargado con distintos modelos de machine learning.

Costco Wholesale Corporation (COST): Costco es una cadena de venta mayorista que opera a nivel internacional a través de grandes almacenes. Costco cuenta con un equipo especializado de analytics que se encarga de analizar los datos de sus productos para mejorar la productividad y sostenibilidad de su sección de productos frescos. Para llevar a cabo esta tarea, Costco hace uso de la tecnología SAP a través de un algoritmo que calcula la demanda que ayuda a la gestión de productos.

The Coca-Cola Company (KO): Coca Cola es una empresa que ofrece una amplia gama de productos de bebidas no alcohólicas a nivel global. De acuerdo con un artículo de Harvard Business Review, Coca Cola genera una ingente cantidad de datos derivados de sus actividades. Coca Cola somete estos datos a distintos modelos de machine learning, a través de los cuales obtienen resultados sobre patrones, que utilizan para sacar nuevos productos al mercado. Coca Cola recoge datos de sus consumidores más jóvenes a través de la página MyCokeRewards.com. A través de esta página, la compañía cree que ha logrado alcanzar un alto grado de conocimiento que les ha permitido dirigirse a consumidores como individuos. La web atrae alrededor de 3.000 visitantes al día. Este tipo de datos, unido a los datos proporcionados por los estudios de mercado que lleva a cabo Coca Cola sobre su extensiva base de consumidores, hace que sus datos sean únicos.

Wal-Mart (WMT): Wal-Mart es una cadena de venta mayorista que opera a través de distintos formatos a nivel mundial. La incorporación de analytics en Wal-Mart es tal que han creado lo que denominan como “Data Café,” un centro de datos de analytics que se encuentra en Arkansas. Wal-Mart Labs es la rama analítica de Wal-Mart. Este centro de trabajo aprovecha las ventajas del *analytics* a través del análisis de grandes volúmenes de datos en tiempo real procedentes tanto de su plataforma en línea como de ventas físicas. El análisis de estos datos generalmente revela patrones de conducta en los consumidores que permite a Wal-Mart implementar cambios en sus tiendas de acuerdo con estos *insights*.

Johnson & Johnson (JNJ): Johnson & Johnson es una empresa que investiga, desarrolla y comercializa distintos productos en el ámbito sanitario a nivel global. De acuerdo con Miao Song, CIO de Johnson & Johnson hasta 2018, el factor clave que permite a Johnson & Johnson ser una compañía digital es su capacidad de analytics. Johnson & Johnson integra el análisis de datos derivado de plataformas internas con fuentes externas para ofrecer contenido personalizado a consumidores y a proveedores de seguros médicos.

General Motors Company (GM): General Motors diseña, construye y comercializa automóviles a nivel global. General Motors cuenta con un equipo dedicado exclusivamente a la gestión de datos, inteligencia artificial y servicios de analytics. General Motors ha construido una plataforma de analytics predictivo que genera insights sobre distintas estrategias internas, tales como anticipar la demanda de mercado de vehículos autónomos.

Ford Motor Company (F): Ford Motor es una empresa que diseña, desarrolla, fabrica y comercializa automóviles a nivel global. Ford ha contado con el análisis de datos para la toma de decisiones desde hace años. No obstante, no fue hasta 2015 cuando decidió crear la unidad global de insights de datos y analytics, un equipo centralizado de data scientists cuyo objetivo es el de compartir buenas prácticas de analytics y propagar la optimización de toma de decisiones basadas en datos a lo largo de la organización.

Facebook, Inc. (FB): Facebook es una red social que permite a sus usuarios conectar a través de dispositivos móviles y ordenadores a nivel global. Desde su concepción, Facebook ha sido un generador de big data, llegando a convertirse en uno de los repositorios de datos personales más grandes del mundo. A través de distintos algoritmos y modelos de machine learning, Facebook es capaz de identificar y recomendar otros usuarios que, con alto grado de probabilidad, corresponden con características similares a las del usuario en cuestión. Asimismo, Facebook pone a disposición de empresas y usuario ciertas herramientas de analytics que permiten destinar los anuncios a distintos segmentos en particular.

Chipotle Mexican Grill (CMG): Chipotle Mexican Grill opera una cadena de restaurantes en Estados Unidos. Chipotle ha colaborado con Accenture para digitalizar sus servicios y ofrecer a sus consumidores una experiencia digital sobresaliente. No obstante, Chipotle cuenta con un equipo encargado de llevar a cabo tareas de analytics. El objetivo de este equipo es el de idear un mapa que guíe las decisiones a tomar.

Advanced Micro Devices, Inc. (AMD): Advance Micro Devices es una compañía que desarrolla y comercializa semiconductores a nivel mundial. A través de Cloudera SDX, AMD ofrece a sus usuarios la capacidad tecnológica para llevar a cabo distintos procesos de analytics incluso de manera simultánea. SDX hace más fácil el desarrollo de aplicaciones de datos multidisciplinarias y abarata su incorporación.

Whirlpool Corporation (WHR): Whirlpool es una compañía que fabrica y comercializa dispositivos de hogar y productos relacionados a nivel internacional. A través de su web corporativa se puede observar que, en 2016, Whirlpool llegó a un acuerdo con IBM para usar su software de machine learning y así conectar los dispositivos de Whirlpool a través del

Internet de las Cosas. De esta manera, Whirlpool será capaz de optimizar el consumo de agua y electricidad, y en el futuro esperan poder ofrecer a sus consumidores nuevas experiencias innovadoras.

Hasbro, Inc. (HAS): Hasbro es una compañía que fabrica y comercializa productos de entretenimiento y juguetes. En 2017, Hasbro empezó a implementar herramientas de analytics a través de su equipo para así sacar provecho de los datos proporcionados por las redes sociales y apelar a los gustos de los más pequeños.

ANEXO II. Código RStudio

```
#Carga de documento CSV con los datos del S&P500
SP500_data =read.delim("S&P500 Dataset.csv",sep=",",header=TRUE)
#Comprobamos que los datos se han trasladado correctamente
head(SP500_data)
str(SP500_data)
#La variable "Sector" no nos aporta nada, así que será eliminada
SP500_data$Sector <- NULL
SP500_data$Name <- NULL
#Ahora procederé a comprobar una serie de gráficas que contienen la
combinación lineal de las variables
pairs(SP500_data)
#Llaman la atención ciertas combinaciones:
plot(SP500_data[, 'Ing'], SP500_data[, 'EV.R'])
scatter.smooth(x=SP500_data$Ln.Ing., y=SP500_data$EV.R, main="EV.R ~
Ln.Ing")
plot(SP500_data[, 'Roa'], SP500_data[, 'Score'])
scatter.smooth(x=SP500_data$Roa, y=SP500_data$Score, main="Score ~
Roa")
plot(SP500_data[, 'Grado_int'], SP500_data[, 'Growth'])
scatter.smooth(x=SP500_data$Grado_int, y=SP500_data$Growth,
main="Growth ~ Imp_Degree")
#Parece que hay probabilidad de que algunas variables estén
significativamente relacionadas
#Para comprobar este hecho voy a construir una tabla de correlaciones
que solucionará cualquier duda
#install.packages("corrplot")
library("corrplot")
df = data.frame(SP500_data)
M <- cor(df)
corrplot(M, type="upper")
#La matriz apunta a una correlación fuerte entre las variables
"Ingresos" y las variables "Ln(Ing)" y "Num. Trabajadores"
#Dado que el set cuenta con la variable "Ln(Ing)", procederé a
eliminar la variable "Ingresos".
SP500_data$Ing <- NULL
#Como he codificado la variable que recoge el grado de integración en
distintas variables binarias, procederé a eliminarla.
SP500_data$Grado_int <- NULL
#Comprobar que no existen outliers
boxplot(SP500_data$Score, main="Score")
boxplot(SP500_data$Ln.Ing., main="Ln(Ing)")
boxplot(SP500_data$Growth, main="Growth")
boxplot(SP_data$Grado_int, main="Imp_Degree")
#ANÁLISIS EXPLORATORIO#
#Aqui llevaré a cabo las dos regresiones. Para ello, comenzaré
dividiendo el set de datos en dos: Training y Test.
#75% será destinado al set de entrenamiento y el 25% restante para el
set de testeo.
smp_size <- floor(0.75 * nrow(SP500_data))
## set the seed to make your partition reproducible
set.seed(123)
train_ind <- sample(seq_len(nrow(SP500_data)), size = smp_size)
train <- SP500_data[train_ind, ]
test <- SP500_data[-train_ind, ]
###REGRESIÓN LINEAL###
lm1 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5, data = train)
summary(lm1)
lm2 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa, data = train)
summary(lm2)
```

```

lm3 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score, data = train)
summary(lm3)
lm4 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Trab, data = train)
summary(lm4)
lm5 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Trab + Growth, data = train)
summary(lm5)
lm6 = lm(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Trab + Growth + EV.R, data = train)
summary(lm6)
#Predictibilidad del modelo
predicho_lm <- predict (lm6, test)
comparacion_lm <- cbind (actual=test$Ln.Ing., predicho_lm)
mean (apply(comparacion_lm, 1, min)/apply(comparacion_lm, 1, max))
###REGRESIÓN RIDGE###
install.packages("ridge")
library(ridge)
lrm <- linearRidge(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Growth + Trab + EV.R, data = train)
summary(lrm)
#Predictibilidad del modelo
predicho_lrm <- predict(lrm, test)
comparacion_lrm <- cbind (actual=test$Ln.Ing., predicho_lrm)
mean (apply(comparacion_lrm, 1, min)/apply(comparacion_lrm, 1, max))
###Simulación MCO vs Ridge. Estudio de predictibilidad###
#Estudio de predictibilidad
library(caret)
library(kableExtra)
#Train/Test Split
#MCO
p <- predict(lm6, train)
error <- (p- train$Ln.Ing.)
RMSE_Model <- sqrt(mean(error^2))
ptest <- predict(lm6, test)
error1 <- (ptest- test$Ln.Ing.)
RMSE_NewData <- sqrt(mean(error1^2))
Method <- c("Train/Test Split")
ModelRMSE <- c(RMSE_Model)
RMSENewData <- c(RMSE_NewData)
table_1 <- data.frame(Method, ModelRMSE, RMSENewData)
kable(table_1) %>% kable_styling(c("striped", "bordered"))
%>%column_spec(2:3, border_left = T)
#Ridge
p_lrm <- predict(lrm, train)
error_lrm <- (p_lrm - train$Ln.Ing.)
RMSE_lrm <- sqrt(mean(error^2))
error2 <- (predicho_lrm- test$Ln.Ing.)
RMSE_NewData_lrm <- sqrt(mean(error2^2))
Method <- c("Train/Test Split")
ModelRMSE_lrm <- c(RMSE_lrm)
RMSENewData_lrm <- c(RMSE_NewData_lrm)
table_2 <- data.frame(Method, ModelRMSE_lrm, RMSENewData_lrm)
kable(table_2) %>% kable_styling(c("striped", "bordered"))
%>%column_spec(2:3, border_left = T)
#K-fold Cross Validation
#MCO
lm_cv <- train(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Trab + Growth + EV.R, data=train,method =
"lm",trControl = trainControl(method = "cv", number = 10))
RMSE_Modelcv <- lm_cv$results$RMSE
summary(lm_cv)

```

```

pcv <- predict(lm_cv, test)
errorcv <- (pcv- test$Ln.Ing.)
RMSE_NewDatacv <- sqrt(mean(errorcv^2))
Method2 <- c("K-fold")
ModelRMSEcv <- c(RMSE_Modelcv)
RMSENewDatacv <- c(RMSE_NewDatacv)
table_3 <- data.frame(Method2, ModelRMSEcv, RMSENewDatacv)
kable(table_3) %>% kable_styling(c("striped", "bordered"))
%>%column_spec(2:3, border_left = T)
#Ridge
lrm_cv <- train(Ln.Ing. ~ Grado_int_2 + Grado_int_3 + Grado_int_4 +
Grado_int_5 + Roa + Score + Trab + Growth + EV.R, data=train,method =
"ridge",trControl = trainControl(method = "cv", number = 10))
RMSE_lrm_cv <- lrm_cv$results$RMSE
pcv_lrm <- predict(lrm_cv, test)
errorcv_lrm <- (pcv_lrm - test$Ln.Ing.)
RMSE_NewDatacv_lrm <- sqrt(mean(errorcv_lrm^2))
ModelRMSEcv_lrm <- c(RMSE_lrm_cv)
RMSENewDatacv_lrm <- c(RMSE_NewDatacv_lrm)
table_4 <- data.frame(Method2, ModelRMSEcv_lrm, RMSENewDatacv_lrm)
kable(table_4) %>% kable_styling(c("striped", "bordered"))
%>%column_spec(2:3, border_left = T)

```

ANEXO III. Output Regresiones Lineales

o Modelo I

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5})$$

Output:

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.7611 -0.8611 -0.1473  0.6227  3.3927

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.4333     0.5168  16.319  <2e-16 ***
Grado_int_2  0.8983     0.7308   1.229  0.2223
Grado_int_3  0.6822     0.6672   1.023  0.3093
Grado_int_4  1.3240     0.5502   2.407  0.0182 *
Grado_int_5  1.0478     0.5713   1.834  0.0700 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.266 on 88 degrees of freedom
Multiple R-squared:  0.07429, Adjusted R-squared:  0.03221
F-statistic: 1.765 on 4 and 88 DF, p-value: 0.143

```

o Modelo II

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5}) + \beta_5 (\text{RoA})$$

Output:

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.6807 -0.8339 -0.1225  0.7814  3.3724

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.21008     0.54017  15.199  <2e-16 ***
Grado_int_2  0.71216     0.74031   0.962  0.3387
Grado_int_3  0.61600     0.66584   0.925  0.3574
Grado_int_4  1.33766     0.54768   2.442  0.0166 *
Grado_int_5  0.98180     0.57074   1.720  0.0889 .
Roa          0.03903     0.02884   1.353  0.1794
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.26 on 87 degrees of freedom
Multiple R-squared:  0.09337, Adjusted R-squared:  0.04127
F-statistic: 1.792 on 5 and 87 DF, p-value: 0.1228

```

○ *Modelo III*

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5}) + \beta_5 (\text{RoA}) + \beta_6 (\text{Puntuación de Gobierno})$$

Output:

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.6972 -0.8797 -0.1565  0.7403  3.3791

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.272770    0.608236  13.601  <2e-16 ***
Grado_int_2  0.706083    0.744852   0.948  0.3458
Grado_int_3  0.612360    0.669687   0.914  0.3631
Grado_int_4  1.331052    0.551439   2.414  0.0179 *
Grado_int_5  0.978158    0.574091   1.704  0.0920 .
Roa          0.038557    0.029071   1.326  0.1882
Score       -0.009998    0.043664  -0.229  0.8194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.267 on 86 degrees of freedom
Multiple R-squared:  0.09393, Adjusted R-squared:  0.03071
F-statistic: 1.486 on 6 and 86 DF, p-value: 0.1926

```

○ *Modelo IV*

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5}) + \beta_5 (\text{RoA}) + \beta_6 (\text{Puntuación de Gobierno}) + \beta_7 (\text{Número de Trabajadores})$$

Output:

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.59536 -0.65483 -0.03823  0.63252  2.40582

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.314e+00    5.173e-01  16.072  <2e-16 ***
Grado_int_2  7.254e-01    6.335e-01   1.145  0.2554
Grado_int_3  5.470e-01    5.697e-01   0.960  0.3397
Grado_int_4  1.069e+00    4.711e-01   2.269  0.0258 *
Grado_int_5  7.306e-01    4.901e-01   1.491  0.1397
Roa          2.778e-02    2.479e-02   1.120  0.2657
Score       -1.243e-02    3.714e-02  -0.335  0.7386
Trab        2.638e-06    4.530e-07   5.822  1e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.077 on 85 degrees of freedom
Multiple R-squared:  0.3523, Adjusted R-squared:  0.2989
F-statistic: 6.604 on 7 and 85 DF, p-value: 3.096e-06

```

○ *Modelo V*

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5}) + \beta_5 (\text{RoA}) + \beta_6 (\text{Puntuación de Gobierno}) + \beta_7 (\text{Número de Trabajadores}) + \beta_8 (\text{Expectativa de$$

Output:

```
Residuals:
      Min       1Q   Median       3Q      Max
-2.59317 -0.66190 -0.05259  0.61651  2.41273

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.328e+00  5.221e-01  15.952 < 2e-16 ***
Grado_int_2  7.275e-01  6.369e-01   1.142  0.257
Grado_int_3  5.481e-01  5.727e-01   0.957  0.341
Grado_int_4  1.074e+00  4.741e-01   2.267  0.026 *
Grado_int_5  7.224e-01  4.935e-01   1.464  0.147
Roa          2.643e-02  2.533e-02   1.043  0.300
Score       -1.413e-02  3.777e-02  -0.374  0.709
Trab        2.632e-06  4.558e-07   5.774 1.27e-07 ***
Growth      1.173e-03  3.949e-03   0.297  0.767
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.083 on 84 degrees of freedom
Multiple R-squared:  0.3529,    Adjusted R-squared:  0.2913
F-statistic: 5.727 on 8 and 84 DF,  p-value: 8.149e-06
```

○ *Modelo VI*

Input:

$$\text{Ln (Ingresos por ventas)} = \beta_0 + \beta_1 (\text{Grado de integración 2}) + \beta_2 (\text{Grado de integración 3}) + \beta_3 (\text{Grado de integración 4}) + \beta_4 (\text{Grado de integración 5}) + \beta_5 (\text{RoA}) + \beta_6 (\text{Puntuación de Gobierno}) + \beta_7 (\text{Número de Trabajadores}) + \beta_8 (\text{Expectativa de Crecimiento}) + \beta_9 (\text{EV/R})$$

Output:

```
Residuals:
      Min       1Q   Median       3Q      Max
-3.05427 -0.75459 -0.00618  0.53969  2.30509

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.300e+00  5.762e-01  16.140 < 2e-16 ***
Grado_int_2  1.845e-01  6.252e-01   0.295  0.76860
Grado_int_3 -2.030e-02  5.692e-01  -0.036  0.97164
Grado_int_4  5.663e-01  4.747e-01   1.193  0.23626
Grado_int_5  4.751e-01  4.732e-01   1.004  0.31830
Roa          3.160e-02  2.403e-02   1.315  0.19223
Score       -1.961e-02  3.579e-02  -0.548  0.58513
Trab        2.263e-06  4.459e-07   5.074 2.34e-06 ***
Growth      2.543e-03  3.761e-03   0.676  0.50079
EV.R        -3.182e-02  9.703e-03  -3.280  0.00152 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.025 on 83 degrees of freedom
Multiple R-squared:  0.4272,    Adjusted R-squared:  0.3651
F-statistic: 6.878 on 9 and 83 DF,  p-value: 2.461e-07
```