



Facultad de Ciencias Económicas y Empresariales

# **Análisis de variables financieras para predecir la rentabilidad en el sector bancario: Evidencia del ROE**

Autor: María Iborra Meseguer

Clave: 201808242

Director: Dra. Elisa Aracil

MADRID | Junio 2023

## **ABSTRACT**

Profitability becomes essential for making informed decisions, control risk, and maximize investment performance in today's dynamic and digital world. This is particularly important in the banking industry, where being able to forecast profitability is crucial. Considering this, it is important to investigate which set of variables offers the best model to forecast ROE (Return on Equity). The financial variables that most precisely estimate ROE were identified using descriptive and predictive approaches from a financial database of banks listed on the New York Stock Exchange (NYSE) that collects information on investments. According to the results, the ESG rating has a negative impact on the ROE estimate. In addition, both the PS ratio and the current ratio have a positive and considerable impact on ROE prediction. On the other hand, ROE is not influenced in terms of predictability by ROA and PE ratio. As a result of this outcomes, investors may be able to make better decisions and manage profitability more effectively.

**Key words:** *profitability, ROE, forecasting, banks, financial variables, ESG, investor*

## RESUMEN

En un mundo dinámico y digital, la rentabilidad se vuelve una herramienta esencial para tomar decisiones informadas, gestionar el riesgo y optimizar el rendimiento de las inversiones. Esto es especialmente relevante en el sector bancario, donde la capacidad de predecir la rentabilidad es crucial. Ante este hecho, resulta esencial examinar qué conjunto de variables proporciona el mejor modelo para predecir el ROE (Return on Equity). A partir de una base de datos financieros de bancos de la Bolsa de Nueva York (NYSE) que recopila información sobre inversiones, se utilizaron metodologías descriptivas y predictivas para encontrar los factores financieros que estiman con mayor exactitud el ROE. Según los resultados obtenidos, la calificación ESG impacta negativamente en la proyección del ROE. Además, tanto el ratio PS como el ratio de liquidez tienen un impacto positivo y considerable en la predicción del ROE. Por otra parte, el ROE no se ve influenciado en términos de predicción por el ROA y el ratio PE. Gracias a esta información, los inversores tienen la posibilidad de tomar decisiones más informadas y gestionar la rentabilidad de manera más eficiente.

**Palabras clave:** *rentabilidad, ROE, predicción, bancos, variables financieras, ESG, inversor*

# ÍNDICE

1.	INTRODUCCIÓN .....	6
2.	MARCO TEÓRICO .....	8
2.1	LA INFLUENCIA DE LOS RATIOS FINANCIEROS EN EL SECTOR BANCARIO.....	8
2.2	ROE .....	8
2.3	PREDICCIÓN DEL ROE A PARTIR DE OTRAS VARIABLES.....	9
2.4	FORMULACIÓN DE LAS HIPÓTESIS .....	14
3.	MÉTODOS.....	16
3.1	METODOLOGÍA .....	16
3.2	MUESTRA .....	19
3.3	MEDIDAS .....	19
4.	RESULTADOS .....	22
4.1	ANÁLISIS DESCRIPTIVO DE LOS DATOS .....	22
4.1.1	Distribución de las variables numéricas.....	26
4.1.2	Valores atípicos .....	29
4.1.3	Normalización de las variables .....	35
4.1.4	Análisis de correlación.....	38
4.1.5	Análisis de correlación del ROE .....	41
4.2	ANÁLISIS PREDICTIVO .....	42
4.2.1	Importancia del análisis predictivo .....	42
4.2.2	Modelos predictivos .....	42
4.2.3	Resultados obtenidos en cada modelo predictivo .....	43
5.	CONCLUSIONES .....	51
6.	BIBLIOGRAFÍA.....	54
7.	ANEXO .....	60

## ÍNDICE DE FIGURAS

Figura 1. Resumen de metodologías.....	18
Figura 2. Representación de los bancos recogidos en la base de datos .....	22
Figura 3. Distribución del ROE.....	26
Figura 4. Variables con distribución asimétrica a la derecha.....	27
Figura 5. Variables con distribución asimétrica a la izquierda.....	29
Figura 6. Distribución de variable con doble simetría.....	29
Figura 7. Variables sin valores atípicos en el diagrama de caja.....	30
Figura 8. Variables con valores atípicos.....	31
Figura 9. Gráficas de las variables normalizadas .....	36
Figura 10. Matriz de correlación .....	39
Figura 11. Matriz de correlación con variables eliminadas.....	41

## ÍNDICE DE TABLAS

Tabla 1. Descriptivos de la muestra .....	19
Tabla 2. Resumen estadístico de los datos .....	23
Tabla 3. Contribución de cada variable independiente al modelo predictivo .....	45
Tabla 4. Resultado de los modelo predictivos.....	48

# 1. INTRODUCCIÓN

En un entorno económico altamente competitivo y volátil, los bancos se enfrentan a desafíos constantes para mantener su rentabilidad (Maverick, 2021). La rentabilidad de los bancos es un tema importante en el contexto económico mundial (Adrianzen, 2016), especialmente después de la pandemia del Covid-19 (Li et al., 2021).

En este contexto se ha vuelto crucial comprender y predecir el ROE como una métrica fundamental. (Pennacchi and Santos, 2018). El ROE es un acrónimo que hace referencia a “Return on Equity” o “Rentabilidad sobre el Patrimonio Neto”. (Damodaran, 2007). Para calcular el ROE, se multiplica el beneficio neto por 100 y se divide el resultado entre los fondos propios (European Central Bank Report, 2010).

El ROE es un indicador financiero utilizado para analizar la capacidad de una empresa para generar ganancias a partir de su inversión de capital (Santander Universidad, 2022) y refleja la eficiencia con la que una entidad utiliza su capital para generar beneficios (Jara Quevedo et al., 2021). Es considerado uno de los ratios más importantes para analistas financieros e inversores, al medir la rentabilidad de una empresa (Contreras, 2006).

Los inversores utilizan el ROE como una herramienta para estimar la rentabilidad de sus inversiones en el sector bancario, obteniendo una visión clara de cuánto beneficio se genera a partir del capital invertido (Vera-Gilces et al., 2020). Además, los reguladores y supervisores financieros consideran el ROE como un indicador esencial para evaluar la salud financiera de los bancos (Athanasoglou et al., 2008). Asimismo, analistas financieros, clientes y competidores, prestan especial atención a este ratio ya que proporciona información valiosa sobre la capacidad para generar beneficios y la posición competitiva de una entidad en el mercado (Chidozie and Ayadi, 2017).

Teniendo en cuenta el contexto económico posterior a la pandemia del Covid-19 en el que nos encontramos, el objetivo de la investigación es predecir el ROE a partir de diferentes variables financieras. Además, brinda a las instituciones financieras y a los demás actores del mercado la posibilidad de entender aquellas variables que más información aportan a la hora de estudiar la rentabilidad de los bancos y que les permita enfrentar los desafíos de rentabilidad. Al predecir el ROE, se busca comprender cómo

diferentes variables influyen en la eficiencia con la que una entidad utiliza su capital para generar beneficios.

Con el fin de lograr el objetivo, se trabajará con una base de datos que contiene información financiera y de inversiones del sector bancario, obtenida del NYSE. Se aplicarán diversas técnicas analíticas a este conjunto de datos para contribuir a la consecución del objetivo. En primer lugar, se realizará un análisis descriptivo para proporcionar una visión general acerca de las características y patrones presentes en el conjunto de datos. Durante este análisis se examinaron las distribuciones de las variables, se normalizaron y se identificaron los valores atípicos, entre otras cosas. En segundo lugar, se realizará un análisis predictivo con el propósito de dar la respuesta más precisa al objetivo de la investigación. Para ello, se utilizarán los datos históricos disponibles que servirán como base de los modelos predictivos capaces de hacer estimaciones sobre los resultados futuros de la variable ROE.

Después de completar la investigación, se obtuvieron resultados significativos que rebelan la influencia de distintas variables en la predicción del ROE. Se identificó que ciertas variables, como la calificación ESG, el ratio EPS, el ratio PS, el ratio PB y el ratio de liquidez, tienen una mayor importancia al predecir el ROE ya que muestran coeficientes más altos en los modelos predictivos. Por otro lado, se encontró que variables como el ratio PE, el ROA y la inflación, entre otras analizadas, no presentan influencia significativa en la predicción del ROE.

Estos resultados son de gran utilidad ya que ayudan a la toma de decisiones en el ámbito financiero y de inversiones. Al conocer qué variables son relevantes para predecir el ROE, analistas y profesionales del sector bancario pueden utilizar esta información para mejorar la evaluación de la rentabilidad de los bancos y realizar proyecciones más precisas sobre su desempeño futuro.

## **2. MARCO TEÓRICO**

### **2.1 LA INFLUENCIA DE LOS RATIOS FINANCIEROS EN EL SECTOR BANCARIO**

Los ratios financieros se utilizan para establecer una relación matemática entre dos cantidades que aparecen en los estados financieros con el fin de obtener información sobre el rendimiento global de la empresa, entre otros aspectos (Babalola and Abiola, 2013). Los ratios financieros son una herramienta útil para los participantes en el mercado, ya que su cálculo e interpretación son sencillos (Kharatyan, 2016). Los ratios de rentabilidad miden la capacidad de una compañía para generar ingresos en relación con las ventas, los activos y los fondos propios. Es decir, permiten medir cómo de bien la empresa está utilizando sus recursos para generar ganancias (Tuvadaratragool, 2013). Por ejemplo, un ratio de rentabilidad alto implica que la empresa está generando más ganancias en relación con los recursos que está utilizando y por lo tanto es un buen indicador sobre la eficacia en la gestión de la rentabilidad de la empresa (Chandra et al., 2019).

Según varios autores, incluyendo a Black, Wright y Davies (2001), el objetivo principal y fundamental de cualquier empresa es generar riqueza y crear valor para sus propietarios. Esto se logra cuando los rendimientos del capital de una empresa superan el coste de dicho capital (Black et al., 2001). Considerando lo anterior, el ROE es posiblemente el ratio más importante para una empresa, ya que indica el ritmo al que aumenta la riqueza de los propietarios (Wet and Toit, 2007).

### **2.2 ROE**

El ROE, junto con el ROA, es una de las medidas más populares y utilizadas para medir el rendimiento financiero de las empresas (Malagrino et al., 2018; Rappaport, 1986). El ROE representa el resultado final de un análisis estructurado de ratios financieros conocido como análisis Du Pont, lo que contribuye a su popularidad entre analistas, gestores financieros y accionistas por igual (Stowe et al., 2002; Correia et al., 2003; Firer et al., 2004; Wet and Toit, 2007). Además, el ROE vincula la cuenta de resultados con el balance, lo que aumenta su atractivo (Wet and Toit, 2007).

El ROE se puede calcular de formas muy distintas. Algunas de las fórmulas más comunes para calcular el ROE son:

$$ROE = \frac{\textit{Beneficio neto}}{\textit{Fondos propios}}$$

$$ROE = ROA \times \frac{\textit{Activos totales}}{\textit{Fondos propios}}$$

El ROE indica qué porcentaje de beneficios obtiene la empresa por cada unidad monetaria de fondos propios invertida en ella (Berkalzne and Zelgalve, 2014) y es una medida sólida que refleja la buena gestión de una empresa (Kharatyan 2016). Las empresas con un ROE más alto tienden a tener ventajas competitivas, lo que se traduce en una mayor rentabilidad para los inversores (Kharatyan 2016). Según Berman et al., (2013) el ROE refleja la capacidad de una empresa para generar rendimiento a los accionistas y superar el riesgo potencial al que se incurría al realizar la inversión. No obstante, el ROE no especifica cuánto efectivo se devolverá a los accionistas (Moussu, 2018) pero sigue siendo un indicador importante que se utiliza ampliamente en la banca y en las bonificaciones de los ejecutivos (Moussu and Petit-Romec, 2017).

Es crucial considerar el ROE no como un indicador aislado, sino en relación con otras variables internas y externas de la empresa que también generan un impacto en la rentabilidad (Wet and Toit, 2007). Consecuentemente, el ROE no debe ser el único factor analizado cuando se desea conocer la rentabilidad de una empresa, ya que ignora por completo el riesgo, lo que limita su utilidad como métrica de rendimiento en algunos casos, especialmente en la banca (Klaassen adn Eeghen, 2015). Es pues importante tener en cuenta los ratios e indicadores que pueden tener un impacto significativo en el ROE y la rentabilidad de las empresas (Kharatyan, 2016). Teniendo en cuenta todo esto, parece imperativo estudiar aquellas variables que sirven como impulsoras y predictoras del ROE.

### **2.3 PREDICCIÓN DEL ROE A PARTIR DE OTRAS VARIABLES**

Para lograr una rentabilidad constante, los bancos se ven influenciados por factores internos y externos (Ongore and Kusa, 2013; Kosmidou, 2008). En este sentido, según Jaber and Al-Khawaldeh, (2014), tanto los determinantes endógenos como los exógenos afectan la rentabilidad y los beneficios de las instituciones bancarias.

Los factores internos son cruciales para la ganancia particular del banco y están bajo el control e influencia de la entidad. Engloban los atributos propios del banco que impactan su rendimiento, y son afectados por las decisiones internas tomadas por la gerencia y el consejo directivo (Pradhan and Shrestha, 2016). Por otro lado, los factores externos son independientes de la gestión del banco y reflejan el entorno económico y del sector que influye en el funcionamiento y rendimiento de la industria bancaria (Horobet et al., 2021). Estos factores abarcan aspectos sectoriales y macroeconómicos que están fuera del control de la entidad y tienen un impacto directo en la rentabilidad de los bancos (Pradhan and Shrestha, 2016).

Es una práctica habitual que los inversores busquen comprender y anticipar el rendimiento futuro de sus inversiones en relación con el retorno sobre los recursos propios, empleando variables tanto internas como externas (Kumbure et al., 2022). Por consiguiente, se han llevado a cabo diversos estudios e investigaciones académicas para analizar la capacidad predictiva del ROE y determinar qué variables posibilitan una estimación más precisa de este indicador.

Los ratios de valoración, como el PS (Price-to-Sales), PE (Price-to-Earnings) y PB (Price-to-Book) pueden proporcionar señales útiles tanto a corto plazo como a largo plazo para los inversores (Khan et al., 2012). Si los ratios de valoración indican que el mercado está sobrevalorado, los inversores pueden optar por ser cautelosos y reducir su exposición a acciones. Por el contrario, si los ratios de valoración indican que el mercado está infravalorado, podría ser un buen momento para aumentar la inversión en acciones. Cuando se analiza cómo los ratios de valoración del mercado pueden impactar las decisiones de los inversores en relación con su exposición a acciones, especialmente al evaluar si el mercado está sobrevalorado o infravalorado, se observa que estas decisiones pueden fundamentarse en las expectativas de ganancias futuras (Campbell and Shiller, 2021). Como resultado, estas decisiones pueden tener un impacto significativo en el rendimiento y la trayectoria del ROE de las empresas.

Varios estudios han investigado el poder de diferentes ratios de valoración para predecir el ROE. Circiumaru et al., (2010), por ejemplo, encontraron una correlación positiva entre el ratio PS y el ROE, lo que indica que el ratio PS puede ser un buen predictor del ROE. Wu (2014) descubrió que el ratio PE tiene una relación en forma de U con el ROE, lo que

significa que puede predecir el ROE hasta cierto punto óptimo, pero más allá de ese punto, su capacidad predictiva disminuye

Al analizar la capacidad de predicción del ratio EPS sobre la rentabilidad, Wet and Toit (2007) señala que maximizar el ratio EPS no siempre conduce a una mayor rentabilidad para los accionistas en el mismo año, lo que indica que el poder predictivo del ratio EPS puede ser informativo, pero relativamente débil. Sin embargo, Ige et al. (2020) sugieren que las estrategias de crecimiento interno, como maximizar el ratio EPS, pueden predecir con facilidad los resultados financieros de las empresas en el corto plazo. Taani and Banykhaled (2011) también concluyen que el rendimiento de los fondos propios está fuertemente influenciado por el ratio EPS y que esta variable permite predecir el comportamiento futuro del ROE.

Según diversos autores, se ha analizado la relación y la capacidad de predicción del ratio PB sobre el ROE. Maverick (2022) descubrió que el ratio PB y el ROE tienden a moverse en tándem, lo que permite determinar si una empresa está sobrevalorada (alto ratio PB y bajo ROE) o infravalorada (bajo ratio PB y alto ROE). El autor sostiene que, al conocer esta relación, los inversores pueden estimar el valor del ROE a corto plazo utilizando el ratio PB y considerando los comportamientos y patrones anteriores. Por otro lado, Liem and Basana (2023) afirman que no existe una relación significativa entre el rendimiento de las acciones y el ratio ratio PB, lo que sugiere que el ratio PB no es útil para estimar el rendimiento a corto plazo ni a largo plazo de las acciones. Estos hallazgos plantean distintas perspectivas sobre la capacidad predictiva del ratio PB en relación con el ROE.

Asimismo, algunos de los estudios realizados han demostrado que los ratios financieros pueden ser útiles para predecir el rendimiento de las acciones, siendo el ratio PB el que muestra un mayor poder predictivo (Kheradyar, 2011). De hecho, diversos autores han encontrado que el ratio PB puede predecir de manera sólida los rendimientos de las acciones (Fama et al., 1988, 1992, 1995; Kothari and Shanken, 1997; Pontiff and Schall, 1998). Además, en los resultados de la investigación de Kheradyar, se ha observado que combinar varios ratios financieros, como el ratio PB y el ratio PE, en un modelo de regresión predictiva múltiple puede mejorar aún más la capacidad de predicción del rendimiento de las acciones.

Estos hallazgos respaldan la importancia de los ratios de valoración como herramientas predictivas para los inversores y destacan la relevancia de considerar diferentes ratios en el análisis financiero de una empresa.

Numerosos estudios han rebelado la capacidad de los diferentes ratios de rentabilidad para predecir el desempeño financiero de una empresa. Según Berkalzne adn Zelgalve (2014), la relación de predicción más sólida se encuentra entre el ROE y el ROA. Estos resultados son respaldados por Kabajeh et al. (2012), quienes también encontraron que el ROI y el precio de las acciones son variables predictivas importantes en relación con el ROE.

Por otro lado, Burja y Marginean (2014) concluyen que el margen de beneficio neto no tiene capacidad predictiva para el ROE. Sin embargo, Denle et al. (2013) sostienen que el margen de beneficio neto desempeña un papel significativo en la predicción del ROE.

Estos hallazgos resaltan la complejidad de las relaciones entre los ratios de rentabilidad y subrayan la importancia de considerar múltiples factores y variables al predecir el rendimiento financiero de una empresa.

La liquidez es un factor crucial en la gestión financiera de una empresa, ya que se refiere a su capacidad para cumplir con las obligaciones de pago a corto plazo (Fama and French, 1988). Para evaluar la liquidez, se utiliza el ratio de liquidez, que compara los activos líquidos con los pasivos a corto plazo (Máster Cámara Málaga, 2022). Además, una buena gestión de la liquidez permite a la empresa aprovechar oportunidades de inversión rentables y financiar proyectos que generen ganancias a largo plazo (DeAngelo et al., 2011). Es por ello, que el ratio de liquidez puede ser considerado un buen predictor del ROE.

La rentabilidad y la liquidez son indicadores efectivos para evaluar la salud corporativa y el desempeño de las entidades comerciales, incluyendo los bancos comerciales (Flamini et al., 2009). Estos indicadores adquieren una relevancia especial para los accionistas y depositantes, quienes tienen un gran interés en el rendimiento de un banco (Eljelly, 2004). Teniendo en cuenta la importancia de estos indicadores y su relevancia en la evaluación del desempeño bancario, se han llevado a cabo numerosos estudios para analizar la relación entre la rentabilidad y la liquidez, así como para explorar el poder predictivo de una variable sobre la otra. De acuerdo con Nunes et al. (2012), la rentabilidad de un banco

está influenciada por su liquidez, y se ha encontrado una correlación positiva entre ambas variables. Estos hallazgos respaldan la idea de que la liquidez puede ser un buen predictor del retorno de capital.

En línea con esto, la investigación llevada a cabo por Solano et al. (2022) reveló que las empresas que presentan niveles más altos de liquidez tienden a generar una mayor rentabilidad para sus accionistas. Los autores llegaron a la conclusión de que los inversores pueden utilizar indicadores financieros y macroeconómicos como herramientas para tener una idea del desempeño en términos de rentabilidad de una empresa. Hassan (2019) también respalda esta idea, destacando la utilidad del ratio de liquidez como predictor del ROE. Los resultados de estos estudios son consistentes con las conclusiones de Pradhan y Shresha (2016) y Pasiouras y Kosmidou (2007), quienes han encontrado que el ratio de liquidez posee un alto poder predictivo en relación con el ROE.

Sin embargo, otros autores como Berkalzne and Zelgalve (2014), Hassan (2019) y Saleem and Rehman (2011) no encuentran una significancia estadística determinante para la variable del ratio de liquidez al momento de predecir el ROE. Estas discrepancias sugieren la existencia de diferentes perspectivas y enfoques en la literatura académica sobre la capacidad de la liquidez a la hora de predecir la rentabilidad bancaria.

Los criterios ESG (ambientales, sociales y de gobernanza) son importantes para la inversión sostenible y para el desempeño de las empresas. En el caso de los bancos, el riesgo ESG ejerce una influencia de diversa magnitud sobre los riesgos financieros presentes en un banco. En cuanto a la relación entre los componentes ESG para predecir el desempeño financiero de las empresas, un estudio en Colombia encontró que un mayor puntaje ESG no se ve reflejado ni en un mayor ROE ni en uno menor (Velásquez, 2022). Por el contrario, autores como El Khoury et al. (2023) y Aydogmus et al. (2022) muestran que las inversiones en ESG permiten a una empresa mejorar su rentabilidad y que esta se puede predecir de manera acorde a través de la calificación obtenida. De la misma manera, Clément et al. (2023) utilizó técnicas de aprendizaje automático para estudiar si la puntuación ESG de una empresa permitía estimar su rentabilidad y descubrió que esto sí sucede y que una mayor puntuación ESG está asociada a una mayor rentabilidad.

Algunas investigaciones analizaron el proceso de inversiones de compra y venta ya que este puede afectar indirectamente al ROE de una empresa al influir en la disponibilidad de capital y en el sentimiento de los inversores. Elisabeth encontró que este proceso no afecta directamente a la rentabilidad de la empresa ni a sus ingresos netos, que son la base para calcular el ROE (Casamayou, 2019). Asimismo, al analizar qué sucede con la volatilidad de los precios, no existe un consenso claro sobre su capacidad de predecir el ROE (Usman Arshad, 2021; Singh and Tandon, 2019; Roni Bhowmik and Wang, 2020).

Varios estudios han explorado la capacidad de la inflación como factor externo para predecir la rentabilidad de los bancos. Al respecto, Sufian and Chong (2008) y Ben Khediri et al. (2009) concluyeron que la inflación no es un predictor sólido de la rentabilidad bancaria. Sin embargo, otros autores han encontrado una relación leve entre la inflación y la rentabilidad de los bancos. Por ejemplo, Petria et al. (2015), Athanasoglou et al. (2005), Pasiouras y Kosmidou (2007), Ganesan (2001) y Growe et al. (2014) indican que la inflación puede tener cierta capacidad predictiva, aunque en baja medida. Estos estudios han demostrado que la tasa de crecimiento de la inflación se ha utilizado para predecir la rentabilidad de los bancos, como se evidencia en las investigaciones de Wet and Toit (2021) y Bezawada (2018).

Los estudios existentes han arrojado resultados contradictorios en cuanto a la efectividad de los ratios de valoración, ratios financieros, la liquidez, el proceso de compra y venta y la volatilidad y las variables externas como la inflación para predecir dicha rentabilidad. Ante esta falta de consenso en la literatura, es necesario abordar este tema desde una nueva perspectiva con el fin de aclarar las incertidumbres existentes y avanzar más allá en los estudios previos.

## **2.4 FORMULACIÓN DE LAS HIPÓTESIS**

Diversos autores han examinado las variables con mayor influencia en la predicción del ROE, es decir, esto ha sido ampliamente estudiado en la literatura. Se han llevado a cabo investigaciones que respaldan hipótesis similares a las planteadas a continuación y que ya han sido sometidas a pruebas y contrastes. No obstante, para que este estudio sea provechoso y aporte una perspectiva nueva, es importante resaltar que las hipótesis definidas en esta investigación se han aterrizado específicamente al sector bancario y tratan de completar la literatura existente en este tema a través de diversos modelos

predictivos y la combinación de diferentes variables financieras. Además, el objetivo consiste en analizar qué combinación de variables, al ser incluidas en un mismo modelo, permite predecir el ROE de manera más precisa.

La formulación de estas hipótesis se basará en un análisis riguroso y fundamentado de los datos disponibles, y se buscará que sean coherentes y relevantes para los objetivos de la investigación.

- H1: El ratio PE (Price-to-Earnings) influye positivamente a la predicción del ROE
- H2: El ratio PS (Price-to-Sales) influye positivamente a la predicción del ROE
- H3: El ratio EPS (Earnings per-Share) influye positivamente a la predicción del ROE
- H4: El ratio PB (Price-to-Book) influye positivamente a la predicción del ROE
- H5: El ratio de liquidez influye positivamente a la predicción del ROE
- H6: El ROA influye positivamente a la predicción del ROE

## 3. MÉTODOS

### 3.1 METODOLOGÍA

Con el propósito de realizar un estudio exhaustivo de las hipótesis planteadas y evaluar su aplicabilidad en el conjunto de datos bajo investigación, se llevarán a cabo dos análisis fundamentales.

El primero de ellos es un análisis descriptivo de los datos, una metodología que permite resumir y comprender las características esenciales y los patrones presentes en el conjunto de datos. Para ello, se aplicarán cinco enfoques de estudio distintos, que incluyen la exploración de la distribución de las variables numéricas, la identificación de valores atípicos, la normalización de las variables, el análisis de correlación y el análisis de correlación específicamente relacionado con el ROE. El segundo análisis se basa en un modelo predictivo, cuyo objetivo es pronosticar el valor futuro de la variable ROE mediante el uso de información histórica y otras variables relevantes. En este caso, se proporcionará una explicación teórica sobre la importancia del análisis predictivo, se implementarán los modelos predictivos apropiados y se analizarán en detalle los resultados obtenidos en cada uno de ellos. Todas las tareas realizadas en ambos análisis han sido llevadas a cabo utilizando la plataforma R Studio. Durante el proceso, se ha utilizado código específico y este se muestra en el anexo que aparece al final del documento.

En el estudio realizado, se emplearon modelos predictivos basados en la función  $lm()$  en R, la cual se utiliza para ajustar modelos de regresión lineal. Los modelos predictivos resultantes son modelos de regresión lineal que permiten estimar los valores de la variable de respuesta utilizando las variables predictoras. Estos modelos ajustados incluyen los coeficientes de regresión, los cuales representan las relaciones lineales entre las variables predictoras y la variable de respuesta.

En el contexto de este estudio, se utilizaron modelos de regresión basados en Mínimos Cuadrados Ordinarios (MCO) para analizar la relación entre las variables financieras seleccionadas y el retorno sobre el patrimonio en el sector bancario. Mediante estas regresiones, fue posible modelar y estimar cómo las variables independientes, como la

calificación ESG, el ratio EPS, el ratio PS, el ratio PB y el ratio de liquidez, entre otras variables financieras, se relacionan con la variable dependiente, que es el ROE.

El objetivo de utilizar modelos de regresión por MCO fue identificar las variables que tienen una influencia significativa en la predicción del ROE y determinar el peso relativo de cada variable en el modelo. Esto proporciona información valiosa para comprender qué aspectos financieros son más relevantes para la rentabilidad de las empresas del sector bancario. Seguidamente, se presenta las ecuaciones de los modelos predictivos:

Modelo 1:

$$\text{ROE} = \text{Precio de compra} + \text{Precio de venta} + \text{Volatilidad de compra} + \text{Volatilidad de venta} + \text{Sharpe Ratio} + \text{Retorno esperado} + \text{Calificación ESG} + \text{ratio PE} + \text{ratio EPS} + \text{ratio PS} + \text{ratio PB} + \text{Margen de beneficio neto} + \text{Ratio de liquidez} + \text{ROA}$$

Modelo 2:

$$\text{ROE} = \text{Precio de compra} + \text{Retorno esperado} + \text{Calificación ESG} + \text{ratio PE} + \text{ratio EPS} + \text{ratio PS} + \text{Margen de beneficio neto} + \text{Ratio de liquidez} + \text{ROA}$$

Modelo 3:

$$\text{ROE} = \text{Precio de venta} + \text{Retorno esperado} + \text{Calificación ESG} + \text{ratio PE} + \text{ratio EPS} + \text{ratio PS} + \text{Margen de beneficio neto} + \text{Ratio de liquidez} + \text{ROA}$$

Modelo 4:

$$\text{ROE} = \text{Retorno esperado} + \text{Calificación ESG} + \text{ratio PE} + \text{EPS} + \text{ratio PS} + \text{ratio PB} + \text{Margen de beneficio neto} + \text{Ratio de liquidez} + \text{ROA}$$

Modelo 5:

$$\text{ROE} = \text{Retorno esperado} + \text{Calificación ESG} + \text{ratio EPS} + \text{ratio PS} + \text{ratio PB} + \text{Margen de beneficio neto} + \text{Ratio de liquidez} + \text{ROA}$$

Modelo 6:

ROE = Retorno esperado + Calificación ESG + ratio PE + EPS + ratio PS + ratio PB + Margen de beneficio neto + Ratio de liquidez

Modelo 7:

ROE = Retorno esperado + Calificación ESG + ratio PE + ratio EPS + ratio PS + ratio PB + Margen de beneficio neto +ROA

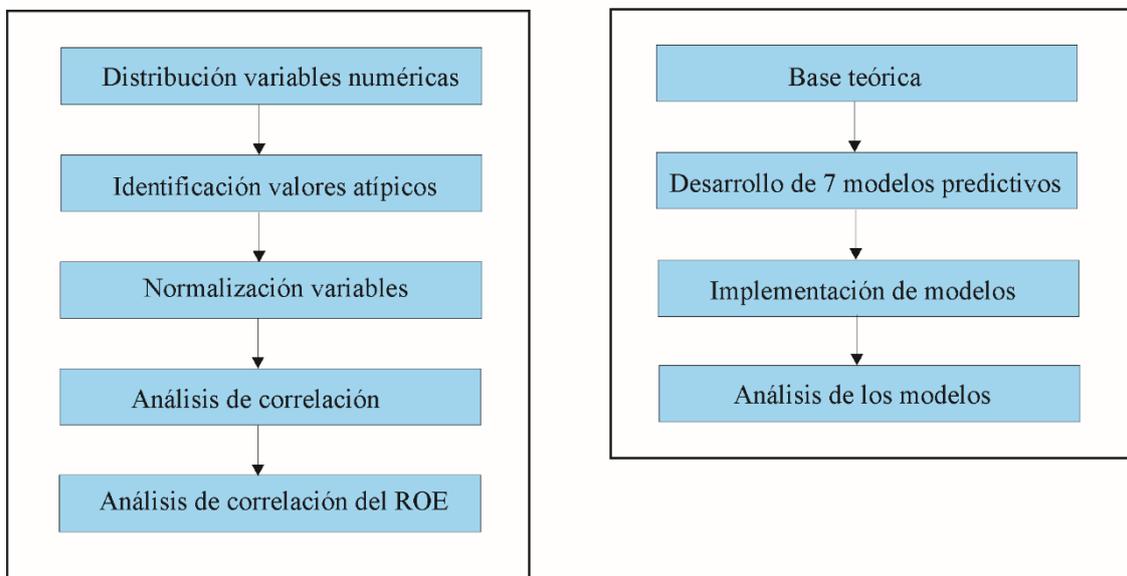
Estos análisis complementarios permitirán obtener una visión completa de los datos, desde una descripción detallada de sus características hasta la capacidad de predecir valores futuros de la variable ROE. Este enfoque garantiza una comprensión profunda del conjunto de datos y proporciona información valiosa para respaldar la toma de decisiones informada y estratégica.

A continuación, se incluye la Figura 1 donde se recogen los pasos a seguir en cada uno de los dos análisis.

Figura 1. Resumen de metodologías

**ANÁLISIS DESCRIPTIVO DE DATOS**

**ANÁLISIS MODELOS PREDICTIVOS**



Fuente: Elaboración propia

### 3.2 MUESTRA

Con el propósito de llevar a cabo el estudio, se empleará la base de datos obtenida de Kaggle. Kaggle es una plataforma en línea establecida en 2010 que facilita la exploración, intercambio y colaboración en proyectos relacionados con el aprendizaje automático, la inteligencia artificial y el análisis de datos en general, mediante una variedad de recursos disponibles.

La base de datos utilizada se denominada " 400k NYSE random investments + financial ratios". La base de datos recoge información sobre inversiones realizadas en empresas listadas en la Bolsa de Valores de Nueva York (NYSE) y además incluye información financiera y ratios de inversión vinculados a los bancos.

De esta muestra, se trabajará exclusivamente con las empresas pertenecientes al sector bancario. Los bancos, incluidos son American Express (AXP), Bank of America (BAC), JP Morgan Chase & Co (JPM), UBS Group AG (UBS), Santander Consumer USA (SC) que es la filial del Banco Santander en Estados Unidos y Deutsche Bank Aktiengesellschaft (DB). A continuación, en la tabla 1 se recoge la información básica de los bancos.

*Tabla 1. Descriptivos de la muestra*

	American Express Company	Bank of America Corporation	JP Morgan Chase & Co	UBS Group AG	Santander Consumer USA	Deutsche Bank Aktiengesellschaft
<b>Acrónimo</b>	AXP	BAC	JPM	UBS	SC	DB
<b>Fundación</b>	1850	1904	2000	1998	2003	1870
<b>Market capitalization (Junio 2023)</b>	\$129,67Bn	\$232,06Bn	\$413,47Bn	\$63,67Bn	\$12,73Bn	\$21,76Bn
<b>Número de empleados</b>	77.300	217.000	293.723	72.597	17.200	84.930
<b>Sede central</b>	Nueva York, EE.UU.	Carolina del Norte, EE.UU.	Nueva York, EE.UU.	Zurich, Suiza	Texas, EE.UU.	Frankfurt, Alemania
<b>Filial</b>	No	No	No	No	Si	No

*Fuente: Elaboración propia*

### 3.3 MEDIDAS

Con el propósito de cumplir con los objetivos de la investigación, se empleará la base de datos donde se incluyen únicamente 6 bancos. La base de datos incluye:

- Duración de inversión en días.

- Cantidad: Número de acciones compradas.
- Precio de las acciones en el momento de compra.
- Precio de las acciones en el momento de venta.
- Volatilidad de la acción al momento de compra.
- Volatilidad de la acción al momento de venta.
- Sharpe Ratio: Métrica financiera que mide el rendimiento de una inversión en relación con el riesgo asumido.
- Retorno esperado: Refleja la estimación de la ganancia o pérdida que se espera obtener de una inversión en el futuro por parte de los inversores.
- Inflación: Nivel de inflación durante el periodo de tiempo en el que se tuvo lugar la inversión.
- Retorno nominal: Refleja el rendimiento real que se obtiene de una inversión después de que se hayan tenido en cuenta factores como la inflación y los costos de transacción.
- Tipo de inversión: Calificación del tipo de inversión realizada según está sea considerada buena o mala.
- Calificación ESG: Calificación ESG de la inversión donde una puntuación más alta indica un mejor desempeño.
- Ratio PE (Price to Earnings): Medida que se utiliza para evaluar la valoración de una acción. Es decir, mide la relación entre el precio de la acción y el beneficio neto por acción de la empresa.
- Ratio EPS (Earnings per-Share): Medida del beneficio neto de una empresa dividido por el número de acciones en circulación. Indica cuánto beneficio genera la empresa por cada acción.
- Ratio PS (Price to Sales): Medida que se utiliza para evaluar la valoración de una empresa en relación con sus ingresos por ventas.
- Ratio PB (Price to Book): Medida que se utiliza para evaluar la relación entre el precio de mercado de las acciones de una empresa y el valor contable de sus activos.

- Margen de beneficio neto: Conocido como margen de beneficio neto. Medida del porcentaje de ingresos que queda después de deducir todos los gastos y costos asociados con la producción y venta de bienes y servicios.
- Ratio de liquidez: Conocido como ratio de liquidez corriente. Medida de la capacidad de una empresa para pagar sus deudas a corto plazo.
- ROA (Return on Assets): Medida de la eficiencia de la empresa en la utilización de sus activos para generar beneficios.
- ROE (Return on Equity): Medida de la eficiencia de la empresa en la utilización de los fondos aportados por los accionistas para generar beneficios. Es decir, que mide la rentabilidad de la empresa en relación con la inversión de los accionistas.

## 4. RESULTADOS

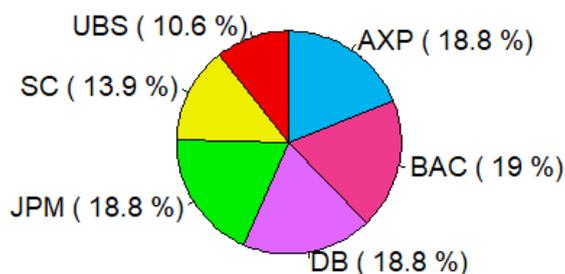
### 4.1 ANÁLISIS DESCRIPTIVO DE LOS DATOS

El análisis descriptivo permite comprender mejor las distribuciones y comportamientos del conjunto de datos. En una primera aproximación, se analiza la composición de la base de datos que revela que hay 84.757 filas y 25 columnas lo que equivale a 2.118.925 observaciones durante un periodo de 7 años (2013-2020). Además, la estructura de la base de datos se compone de 20 variables numéricas, tres variables de tipo carácter y dos variables de tipo fecha.

Asimismo, resulta necesario comprobar si hay algún valor perdido en la muestra. Los datos faltantes afectan, generalmente, al análisis ya que perturban la representatividad de los datos. Con el fin de ver si esto sucedía en nuestra base de datos, utilizamos el comando en R de *is.na()*. Al analizar los resultados, se observa que no hay valores perdidos en nuestro conjunto de datos y que, por lo tanto, podemos trabajar con normalidad sin la necesidad de imputar ningún valor.

En el análisis exploratorio, se ha utilizado un gráfico para visualizar la proporción de bancos presentes en el conjunto de datos. En la figura 2 se observa que la distribución de los bancos es bastante homogénea ya que ningún banco representa una proporción significativamente mayor o menor que los demás. Esto implica que los datos están equilibrados. Sin embargo, se observa que el AXP, BAC, DB y JPM presentan una leve mayor presencia en comparación con el SC y el UBS.

*Figura 2. Representación de los bancos recogidos en la base de datos*



*Fuente: Elaboración propia*

Asimismo, resulta interesante obtener un resumen estadístico de cada variable numérica, incluyendo también las variables de tipo fecha. Para ello, se ha usado el comando *summary()* del conjunto de datos, que nos proporciona datos como la media, la mediana, el valor mínimo, el valor máximo y los cuartiles de cada variable. A continuación, en la tabla 2 se muestra el resultado obtenido en R:

*Tabla 2. Resumen estadístico de los datos*

	Valor mínimo	1er cuartil	Mediana	Media	Desv. Estándar	3er cuartil	Valor máximo
<b>Duración</b>	1,0	15,0	90,0	187,4	210,9	300,0	720,0
<b>Cantidad</b>	50,0	400,0	2000,0	8161,0	12841,0	10000,0	50000,0
<b>Fecha compra</b>	10/10/2013	08/05/2015	25/07/2016	24/06/2016	08/11/2016	31/08/2017	09/10/2018
<b>Fecha venta</b>	11/10/2013	30/10/2015	26/01/2017	28/12/2016	11/13/2017	09/03/2018	31/12/2019
<b>Precio compra</b>	10,00	14,46	22,20	36,87	27,29	57,71	99,95
<b>Precio venta</b>	10,00	14,12	21,90	36,87	28,25	58,71	99,95
<b>Volatilidad compra</b>	0,129	0,196	0,241	0,266	0,090	0,306	0,546
<b>Volatilidad venta</b>	0,1290	0,1993	0,2408	0,2684	0,0920	0,3093	0,5462
<b>Sharpe Ratio</b>	0,2579	0,4025	0,4816	0,5341	0,1710	0,6180	1,0924
<b>Retorno esperado</b>	0,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>Inflación</b>	-0,50	-0,20	-0,15	0,62	1,05	1,68	1,96
<b>Retorno nominal</b>	0,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>ESG ranking</b>	19,8	24,8	26,9	26,0	3,3	27,9	30,0
<b>PE</b>	0,00	7,49	12,14	18,41	57,53	15,42	599,56
<b>EPS</b>	-6,56	0,93	1,97	2,53	2,81	5,22	8,07
<b>PS</b>	0,41	0,79	1,65	1,59	0,74	2,08	3,04
<b>PB</b>	0,00	0,61	0,84	1,31	1,15	1,25	4,44
<b>Margen de beneficio neto</b>	-20,01	6,98	13,69	11,06	8,89	16,61	22,43
<b>Ratio de liquidez</b>	0,61	0,91	0,95	1,13	0,45	1,14	2,23
<b>ROA</b>	-0,46	0,33	0,81	1,22	1,24	1,96	3,82
<b>ROE</b>	-11,69	3,18	8,66	10,04	9,63	16,55	28,81

*Fuente: Elaboración propia*

Los datos muestran que la duración de las inversiones varía significativamente, con una media de 190 días. La inversión más larga duró casi dos años, mientras que la más corta fue de solo un día. En cuanto a la cantidad de acciones compradas y vendidas, los valores oscilaron entre 50 y 50.000, con una media de 8.161.

Al observar las fechas de compra y venta, se observa que la primera operación de compra registrada en la base de datos fue el 10 de octubre de 2013, mientras que la última operación de venta registrada tuvo lugar el 28 de septiembre de 2020. Esto refleja que hay datos recogidos a lo largo de un periodo de tiempo de 7 años.

La base de datos utilizada abarca el periodo comprendido entre 2013 y 2020. Sin embargo, no se ha considerado el impacto del Covid-19 en el análisis. La razón principal de esto radica en que la recopilación de datos se llevó a cabo antes del inicio de la pandemia y no se cuenta con información pública más reciente acerca de estas

inversiones. Por lo tanto, los datos utilizados en este estudio son los más actualizados y disponibles dentro del contexto temporal seleccionado.

Es importante reconocer que durante el Covid-19 se produjeron cambios significativos en las condiciones del mercado y en el comportamiento de las variables financieras. No obstante, los resultados y conclusiones obtenidos reflejan principalmente la realidad y los eventos ocurridos en el periodo, sin considerar directamente el impacto específico de la pandemia. Asimismo, para incorporar el efecto de la pandemia en el análisis se necesitarían datos más recientes y específicos correspondientes al periodo posterior de la pandemia. Dado que esta información no está disponible en la base de datos, es necesario interpretar los resultados de forma rigurosa y reconocer que los hallazgos pueden no reflejar completamente la realidad debido al inicio de la pandemia.

Asimismo, el precio medio de compra y venta de acciones es de aproximadamente 37 dólares. Esto sugiere que el retorno esperado de la inversión no es muy alto. Además, ambas variables presentan valores similares con pequeñas diferencias en los decimales. De hecho, al examinar las variables de retorno esperado y retorno nominal, se observa que los valores son extremadamente bajos y se acercan mucho a cero.

En cuanto a la volatilidad en el momento de compra y la volatilidad en el momento de venta presentan valores similares también, con una volatilidad media de 0.24 aproximadamente en ambos casos. Pese a que el valor mínimo es igual en ambos casos, el valor máximo de la volatilidad en el momento de venta (0.74) es un 34% aproximadamente mayor a la volatilidad en el momento de compra (0.55).

Respecto a la variable de inflación, el valor mínimo registrado es de -0.5, lo que indica un período de deflación en el que los precios de los bienes y servicios estaban disminuyendo. Por otro lado, el valor máximo registrado es de casi 2.0, lo que indica un período de alta inflación en el que los precios estaban aumentando significativamente. El valor medio de la inflación, por su parte, indica que el ritmo general de aumento de precios es del 0.6 en todo el período registrado, lo que sugiere un aumento gradual en los precios.

En lo que refiere a la variable de ESG, se encontró que en promedio todas las inversiones tienen una puntuación alta en este aspecto, con valores alrededor de 26, siendo el valor máximo de 30 y el mínimo de 20. La escala de puntuación va de 0 a 50. Sin embargo, la

base de datos no proporciona información sobre el desglose de esta puntuación en términos de criterios individuales de medio ambiente, social y gobernanza, lo que limita la utilidad de esta variable para fines analíticos.

Se puede observar que el ratio con mayor variación entre el ratio PE, ratio EPS, ratio PS y ratio PB es el ratio PE. La razón principal de esta variabilidad se debe a que el ratio se calcula dividiendo el precio de la acción por las ganancias por acción (ratio EPS), lo que puede provocar que ambos factores varíen significativamente. De hecho, el ratio PE tiene un valor máximo de 600 y un valor medio de 20, lo que nos indica claramente que el valor máximo corresponde a un valor atípico. En contraposición, el ratio EPS, el ratio PS y el ratio PB presentan comportamientos más estables y una distribución más uniforme con valores medios de 2.5, 1.6 y 1.3, respectivamente.

Al mismo tiempo, el análisis de la variable de margen de beneficio neto muestra que el valor mínimo es de -20, lo que sugiere que algunas empresas bancarias de la muestra están operando con pérdidas. Por el contrario, el valor máximo de 22 indica que otras empresas están obteniendo altos beneficios en relación con sus ingresos netos. El valor medio de la variable con un 11 indica que, en promedio, los bancos de la muestra están generando beneficios netos del 11% sobre sus ingresos totales, lo que es considerado un buen margen de beneficio neto.

La muestra de bancos presenta un ratio de liquidez constante, lo que sugiere que estos bancos tienen una capacidad adecuada para cumplir con sus obligaciones a corto plazo, es decir, que pueden pagar sus deudas y compromisos con facilidad.

Por último, se analizan el ROA y el ROE. En primer lugar, el ROA presenta valores medios positivos (1.2), lo que sugiere que las empresas están obteniendo ganancias en relación con sus activos totales. El valor máximo (3.2) indica que algunas empresas tienen un desempeño financiero excepcionalmente bueno, mientras que el valor mínimo (-0.5) indica que algunas empresas están teniendo pérdidas en relación con sus activos, lo que significa que están generando menos ingresos de lo que están gastando para mantener sus operaciones. Por otro lado, el ROE muestra un valor medio de 10 que indica que la empresa está generando un retorno del 10% sobre el capital invertido por los accionistas. Un valor máximo de 29 implica que la empresa está generando un retorno significativamente alto en relación con la inversión de los accionistas, lo cual puede

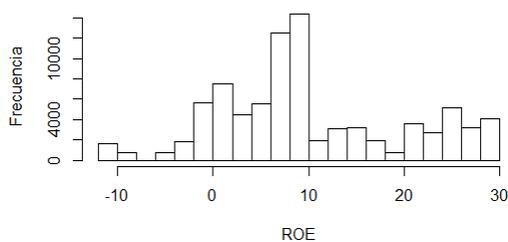
considerarse muy positivo. Sin embargo, el valor mínimo (-11) indica que algunos bancos han incurrido en pérdidas y generado un retorno negativo sobre la inversión de los accionistas, lo cual es preocupante. En general, un ROE positivo y alto es un buen indicador de la rentabilidad y eficiencia financiera de la empresa.

#### 4.1.1 Distribución de las variables numéricas

Para continuar con el análisis descriptivo de los datos, es importante conocer la distribución de las variables ya que nos permite tener una mejor comprensión de los datos. Si una variable tiene una distribución muy sesgada o con una alta concentración de valores en una parte del rango, esto puede indicar que existen factores o características particulares que están influyendo en la variable y que pueden tener un impacto significativo en el análisis. El sesgo, también conocido como asimetría puede ser positivo o negativo, lo que indica la dirección en la que los datos están sesgados. La asimetría positiva, o hacia la derecha, sugiere que hay más valores por debajo de la media que por encima, mientras que la asimetría negativa, o hacia la izquierda, indica lo contrario, es decir, que hay más valores por encima de la media que por debajo. En este conjunto de datos, la mayoría de las variables presentan algún grado de asimetría. Seguidamente, mostramos la distribución de cada variable a través de la función *hist()* en R.

En la figura 3 se puede observar que la única variable que parece seguir una distribución normal y simétrica es el ROE. Es importante tener en cuenta que no es una distribución normal en su estado puro, pero se puede considerar como tal. De esta variable, se puede concluir que la media del ROE es representativa de la mayoría de los casos y puede ser utilizada como un punto de referencia válido para hacer comparaciones entre empresas dentro del mismo sector y conocer el posicionamiento dentro de la industria.

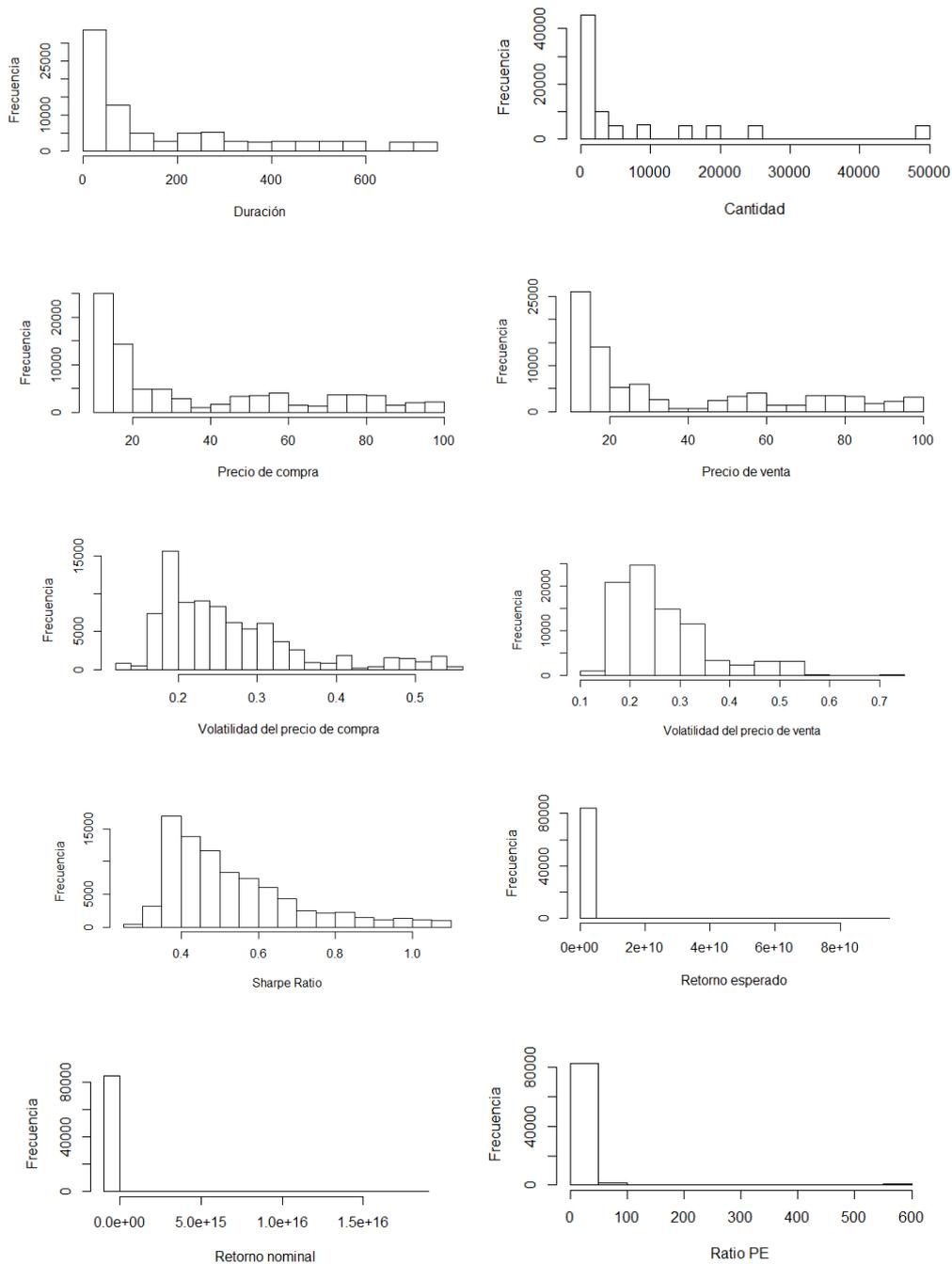
*Figura 3. Distribución del ROE*

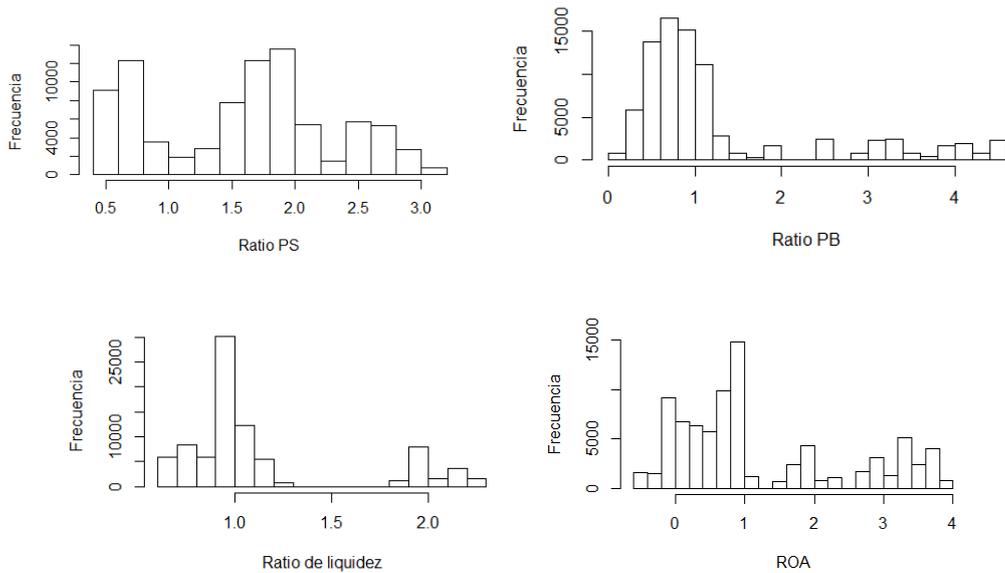


*Fuente: Elaboración propia*

Las variables que muestran una simetría a la derecha son duración, cantidad, precio de compra, precio de venta, volatilidad en el momento de compra, volatilidad en el momento de venta, Sharpe ratio, retorno esperado, retorno nominal, ratio PE, ratio PS, ratio PB, ratio de liquidez y ROA. Esto se observa respectivamente en la figura 4.

*Figura 4. Variables con distribución asimétrica a la derecha*

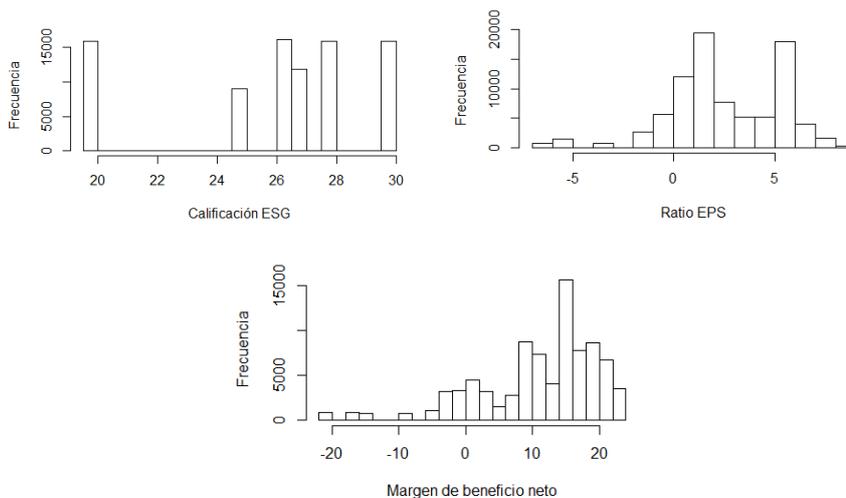




*Fuente: Elaboración propia*

Por el contrario, las observaciones que muestran una distribución asimétrica hacia la izquierda son calificación ESG, ratio EPS y margen de beneficio neto. Esto se observa en la imagen 5. Las conclusiones extraídas de estas variables indican que, en los tres casos, el rendimiento de los bancos es bajo. Se observan puntuaciones relativamente pobres en términos de ESG, lo que sugiere que hay un amplio margen para mejorar las prácticas sostenibles y éticas dentro del sector bancario. Asimismo, el rendimiento de ganancias por acción también es bajo, lo que indica que las empresas se enfrentan a constantes desafíos para generar beneficios significativos y consistentes en relación con sus acciones en circulación. Por último, el margen de beneficio neto que muestra valores bajos sugiere la necesidad de que las empresas revisen y ajusten sus estrategias.

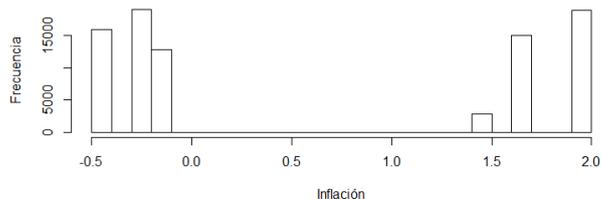
Figura 5. Variables con distribución asimétrica a la izquierda



Fuente: Elaboración propia

Asimismo, la variable de inflación presenta una doble simetría. Esto implica que la distribución se divide en dos partes prácticamente iguales y que existe un punto en el centro que actúa como espejo entre ambas partes. La distribución de esta variable indica la presencia de dos fenómenos o condiciones diferentes que impactan en la inflación durante el periodo estudiado. Esto se observa en la figura 6.

Figura 6. Distribución de variable con doble simetría



Fuente: Elaboración propia

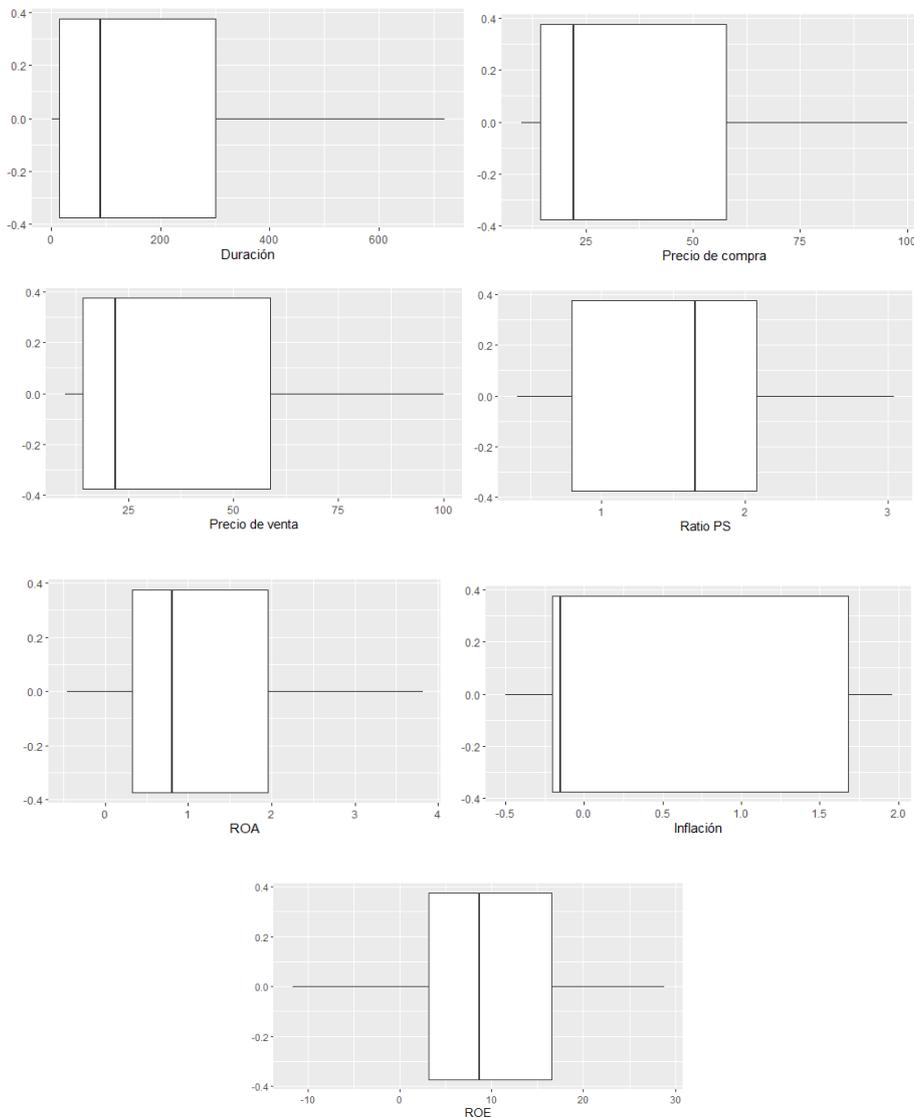
#### 4.1.2 Valores atípicos

Ahora que ya se tiene una idea más clara de cómo se distribuyen los datos de nuestra muestra, es esencial analizar la posible existencia de valores atípicos en los datos ya que nos permite detectar aquellos valores que puedan afectar el análisis y las conclusiones que se obtengan a partir de ellos. La presencia de valores atípicos puede afectar negativamente las estadísticas descriptivas y los modelos predictivos, ya que estos valores extremos pueden distorsionar las conclusiones y disminuir la precisión de los resultados.

Por lo tanto, es fundamental detectar y evaluar los valores atípicos para determinar si deben ser eliminados o si son valores genuinos que deben ser considerados en el análisis. Para ello, se utiliza el diagrama de caja ya que es útil para revelar la presencia de valores atípicos a través de los puntos que están más allá de las líneas verticales, conocidas como bigotes.

Existen variables para las que no se observan valores atípicos en el diagrama de caja. Estas son la duración, el precio de compra, el precio de venta, la inflación, el ratio PS, el ROA y el ROE. Los diagramas correspondientes se observan en la figura 7.

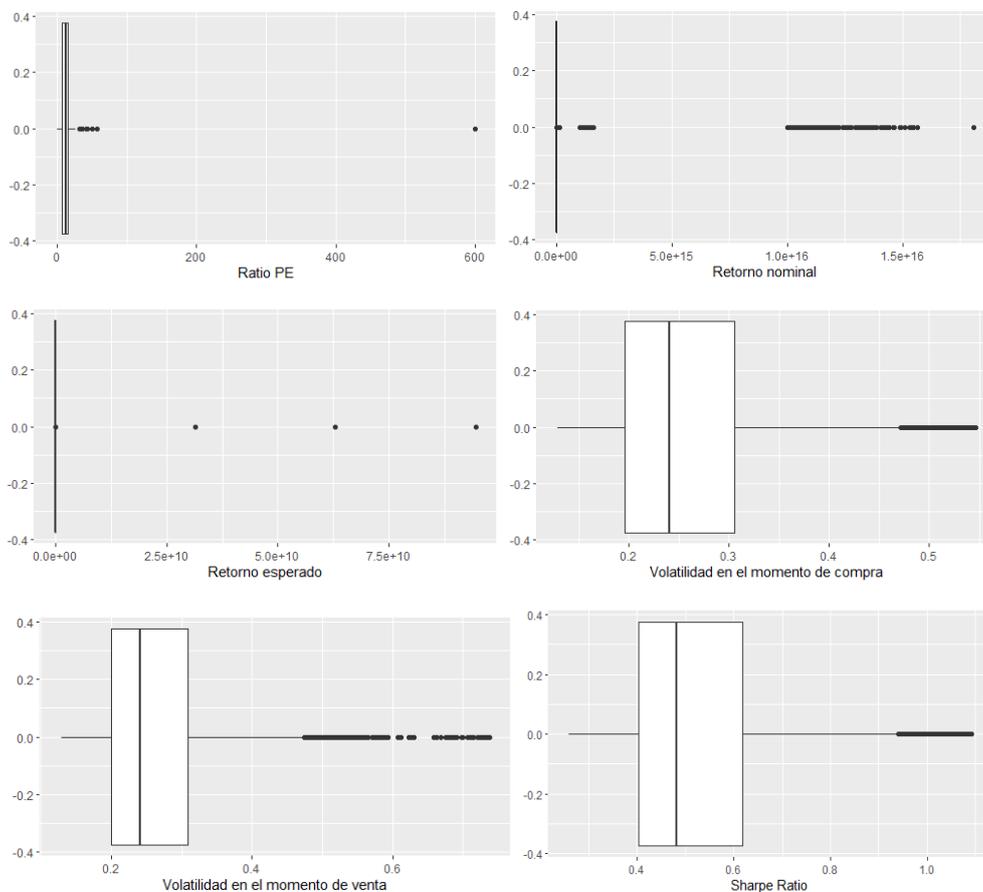
*Figura 7. Variables sin valores atípicos en el diagrama de caja*

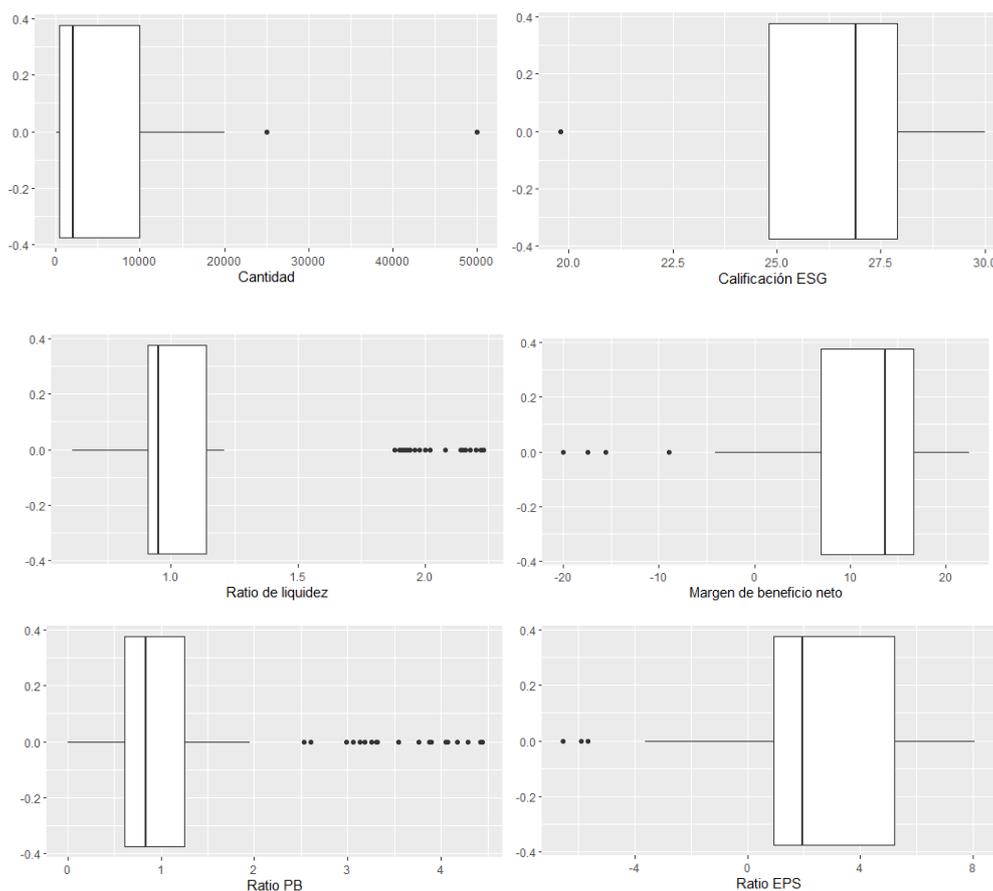


*Fuente: Elaboración propia*

Por el contrario, las demás variables exhiben valores atípicos, por lo que se requerirá de un análisis más detallado. Las variables que contienen valores atípicos son, consecuentemente, la cantidad, la volatilidad en el momento de compra, la volatilidad en el momento de venta, el Sharpe ratio, el retorno esperado, el retorno nominal, la calificación ESG, el ratio PE, el ratio EPS, el ratio PB, el margen de beneficio neto y el ratio de liquidez. Para estos casos en los que hemos identificado valores atípicos, hemos almacenado todas las observaciones que se consideran valores atípicos en vectores específicos para cada una de ellas. Esto nos permite conocer el número de observaciones que se consideran valores atípicos para cada variable y tener una idea aproximada de la proporción de observaciones que se encuentran en cada valor extremo. Estos vectores se han denominado outliers\_X donde la “X” se refiere al nombre de la variable correspondiente. En la figura 8 se observan respectivamente los diagramas de caja para cada una de las variables.

*Figura 8. Variables con valores atípicos*





*Fuente: Elaboración propia*

De igual manera, se proporciona un desglose detallado de cada uno de los vectores mencionados previamente. En el caso del vector `outliers_amount`, contiene un total de 10.042 observaciones consideradas como valores atípicos. Al emplear el comando `table()` en R, se puede determinar la proporción de observaciones en cada valor, lo que rebela que hay 5.007 observaciones en el valor de 25.000 y 5.035 en el valor de 50.000. Por otro lado, el vector `outliers_volbuy` indica que hay 5.454 observaciones consideradas valores atípicos y al analizar la proporción de observaciones se ve que el rango de valores es amplio. Asimismo, el vector `outliers_volsell` muestra que hay un total de 5.527 observaciones que son consideradas valores atípicos. Al evaluar la proporción de observaciones en cada valor, se observa que el rango de valores es extenso, variando desde 0,474 hasta 0,737. Igualmente, el vector `outliers_sr` que hace referencia a la variable de Sharpe ratio, señala que hay 3.623 observaciones que se consideran valores atípicos. Al analizar la proporción de observaciones en cada valor, se observa que el rango de valores es holgado, oscilando entre 0,474 y 0,987.

En quinto lugar, el vector outliers\_ExpR que hace referencia al retorno esperado, refleja que se han almacenado 15.677 observaciones que se consideran valores atípicos. Al analizarlos se observa que el rango de valores es muy grande, lo que tiene sentido ya que las inversiones son muy diversas y dependen de muchos otros factores. En el caso del vector outliers\_NomR que se refiere a la variable retorno nominal, recoge un total de 13.248 observaciones consideradas valores atípicos, y en el análisis se ve que el número de observaciones es muy grande, pero el rango de valores es muy bajo, con números cercanos a 0. Por su parte, en el vector outliers\_ESG rebela 15.955 observaciones consideradas valores atípicos, todas con el mismo valor de 19.8. Esto implica que hay muchas inversiones que reciben la misma puntuación en términos ESG, lo que se considera un valor atípico ya que está bastante por debajo de la media. No obstante, esta calificación sigue siendo positiva y demuestra que los bancos sí están haciendo esfuerzos en este aspecto, pero, aun así, estos no son suficientes en comparación con el resto de los competidores.

Para el caso de los ratios y margen de beneficio neto, se han almacenado todas las observaciones relevantes en diferentes vectores, como outliers\_PE, outliers\_EPS, outliers\_PB, outliers\_NPM y outliers\_CR. Al observar outliers\_PE, se encontraron un total de 7.415 observaciones consideradas como valores atípicos, con un rango de valores de 31.8 a 599 y una media de 785 observaciones por valor. Outliers\_EPS rebeló 2.381 observaciones como valores atípicos, con un rango de valores entre -6.56 y -5.66 y una media de 800 observaciones por valor. Por otro lado, outliers\_PB arrojó un total de 15.955 observaciones como valores atípicos, con un rango de valores de 2.53 a 4.44 y una media de 800 observaciones por valor, aunque algunos valores tienen más repeticiones que otros con 1.400 repeticiones o tan solo 300. También se encontraron 3.128 observaciones como valores atípicos en outliers\_NPM, con un rango de valores de -20 a -9 y una media de 800 observaciones por valor. Finalmente, outliers\_CR mostró un total de 15.955 observaciones como valores atípicos, con un rango de valores entre 1.88 y 2.23.

Tras haber realizado el análisis de valores atípicos de la base de datos, es importante considerar y decidir si estos deben ser eliminados de la base de datos o si en caso contrario, los mantenemos. Tras un análisis cuidadoso, se ha decidido que, en este caso concreto, los valores atípicos son valores legítimos y representan fenómenos reales o verdaderas excepciones en los datos. No obstante, teniendo en cuenta que se va a realizar

un modelo predictivo con los datos, se ha decidido eliminar aquellas observaciones de la base de datos que son consideradas valores atípicos.

Eliminar los valores atípicos de un conjunto de datos es fundamental por varias razones. En primer lugar, los valores atípicos tienen el potencial de distorsionar los resultados y conclusiones de los modelos predictivos. Estos valores extremos pueden introducir ruido en los datos, lo que puede afectar negativamente la precisión del modelo al distorsionar las relaciones entre las variables predictoras.

Al eliminar los valores atípicos, se logra reducir la influencia de estos valores en el análisis. Esto conlleva múltiples beneficios. En primer lugar, se mejora la capacidad del modelo para capturar patrones y relaciones significativas en los datos, ya que se eliminan valores que se alejan significativamente del comportamiento general de los datos. Esto contribuye a una mayor precisión en las predicciones, ya que el modelo se basa en patrones más representativos y menos influenciados por casos excepcionales. Además, se mejora la estabilidad del modelo. Los valores extremos pueden tener un impacto excesivo en la estimación de parámetros y en las predicciones, lo que puede hacer que el modelo sea más sensible a variaciones pequeñas en los datos. Al reducir la influencia excesiva de los valores atípicos, se obtiene un modelo más robusto y estable, que generaliza mejor a nuevos datos. Por último, la eliminación de valores atípicos también facilita la comprensión y la interpretación de las relaciones entre variables. Al eliminar valores extremos, se obtiene un conjunto de datos más coherente y consistente, lo que simplifica el análisis y la interpretación de los resultados. Esto es especialmente importante en situaciones donde se busca entender la importancia relativa de diferentes variables y cómo se relacionan entre sí.

Para ello, se ha utilizado la función *union()* que agrega los índices de las filas que contienen valores atípicos en cada variable. Luego, se ha utilizado el operador de indexación negativo (-) para eliminar las filas correspondientes a los valores atípicos del conjunto de datos original, creando así una versión actualizada del conjunto de datos. El nuevo conjunto de datos contendrá el conjunto de datos original sin las observaciones que contenían valores atípicos en las variables identificadas. Las filas que coinciden en diferentes variables se eliminarán solo una vez.

Después de eliminar los valores atípicos del conjunto de datos, se obtiene un conjunto de datos más limpio y consistente que consta de un total de 34.282 variables. Aunque el conjunto de datos resultante es más pequeño que el original ya que se han eliminado un total de 50.475 variables, este tamaño de muestra sigue siendo lo suficientemente grande como para extraer relaciones y conclusiones interesantes. Al tener un conjunto de datos más depurado, los análisis posteriores pueden centrarse en las variables relevantes y en las relaciones más significativas, lo que facilita la tarea de descubrir patrones y obtener una comprensión más profunda de los datos.

#### **4.1.3 Normalización de las variables**

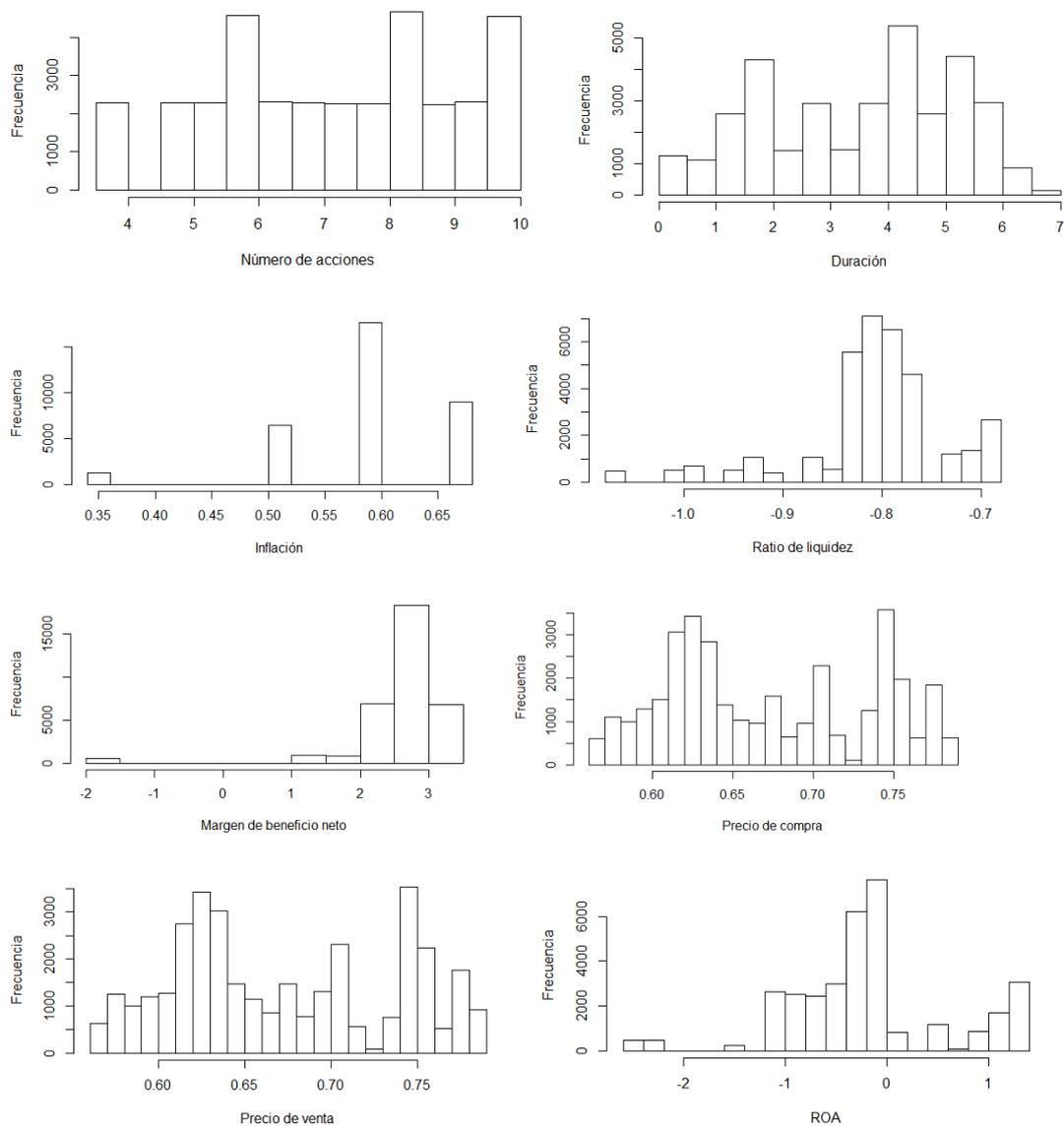
Una vez que se ha tomado una decisión acerca de cómo tratar los valores atípicos, es esencial normalizar los datos para poder comparar y analizar las diferentes variables de la base de datos. El propósito de la normalización es ajustar los valores de los datos a una escala común. Para lograr esto, existen varias técnicas que se pueden emplear. En el caso más simple, cuando los datos siguen una distribución normal, se puede utilizar la función *scale()* en R. En caso contrario, si los datos presentan cualquier tipo de asimetría, se necesita aplicar una transformación logarítmica con la función *log()* en R para ajustar los datos originales. A veces, esto no es suficiente y se debe usar la técnica Box-Cox para obtener una mejor normalización de los datos. Box-Cox utiliza un parámetro conocido como lambda que se estima mediante la maximización de una función de verosimilitud.

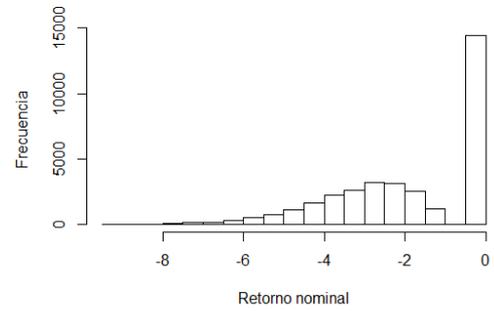
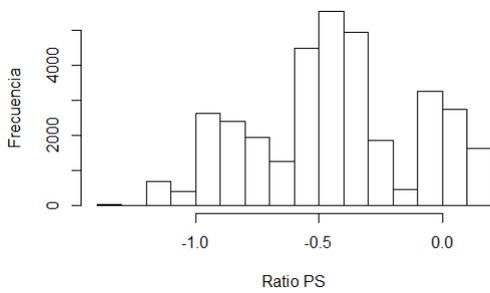
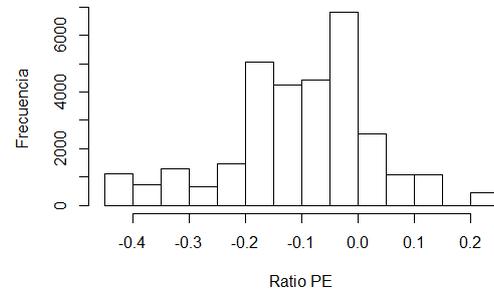
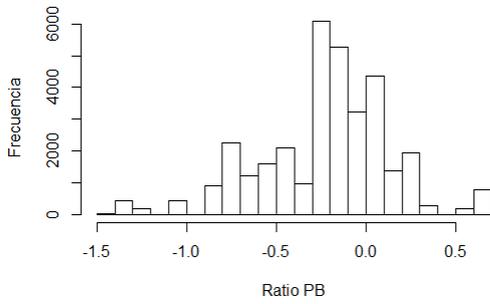
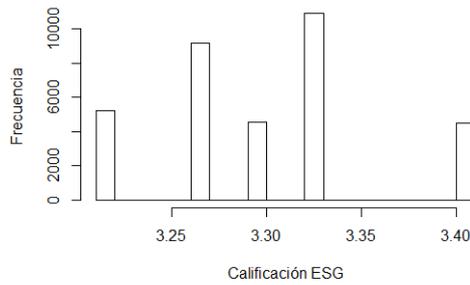
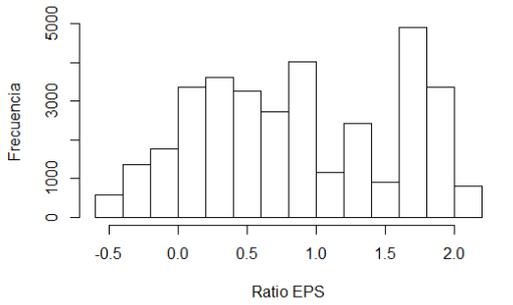
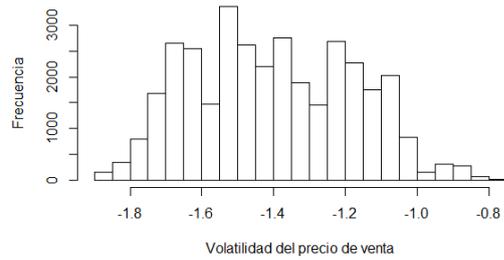
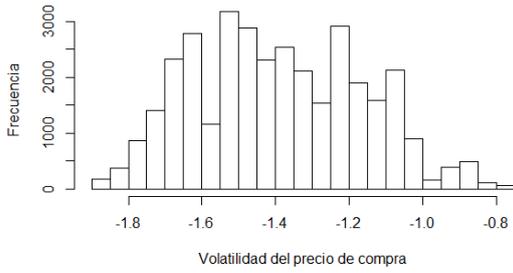
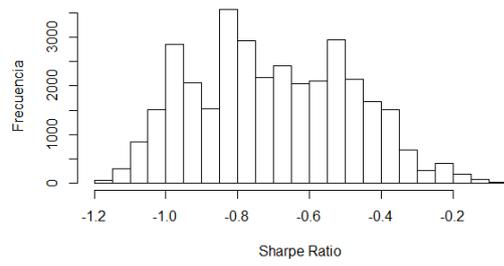
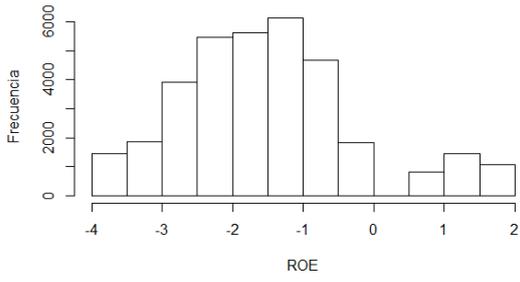
En el conjunto de datos actual, se observa que, en ciertas variables, la transformación logarítmica por sí sola no es suficiente, y se requiere aplicar la técnica Box-Cox para normalizar los datos. Sin embargo, en algunos casos, el resultado obtenido después de aplicar la transformación Box-Cox es menos normal que el obtenido con la transformación logarítmica. Por lo tanto, en esos casos, se mantendrán los datos normalizados mediante la transformación logarítmica. Además, se ha observado que, en dos casos particulares, los resultados no se ajustan a una distribución normal en absoluto. En estas situaciones, se optará por trabajar con la forma logarítmica de los datos, ya que esta preserva de manera más fiel la estructura original.

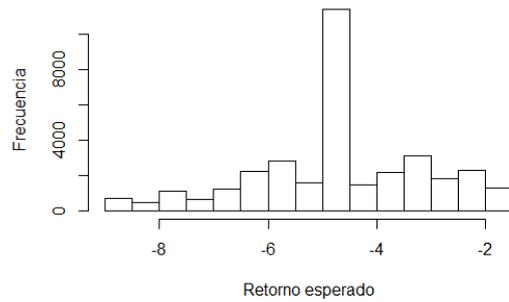
Aquellas variables que han sido normalizadas siguiendo una transformación logarítmica son: duración, cantidad, retorno esperado, inflación, calificación ESG, ratio EPS, ratio PB, margen de beneficio neto, volatilidad en el momento de compra, volatilidad en el

momento de venta, Sharpe ratio, retorno nominal y ROA. Por el contrario, aquellas variables que han sido normalizadas siguiendo una transformación logarítmica y después aplicando la transformación Box-Cox son: precio de compra, precio de venta, ratio PE, ratio PS y ratio de liquidez. En el caso de ROE, como seguía una distribución normal, se utilizará únicamente la función *scale()*. A continuación, se muestran las gráficas de las variables numéricas normalizadas en la figura 9.

Figura 9. Gráficas de las variables normalizadas







*Fuente: Elaboración propia*

De estas distribuciones se extraen diferentes conclusiones. En primer lugar, un sesgo a la derecha en la variable de duración sugiere que hay transacciones que requieren un tiempo prolongado para completarse, lo que puede indicar la presencia de procesos más complejos que demandan una mayor atención y gestión.

En relación con los precios de compra y venta, un sesgo a la derecha implica que hay transacciones que se realizan a precios más altos en comparación con los precios más bajos. Del mismo modo, la presencia de volatilidad en el momento de compra y venta rebela que hay momentos en los que el mercado experimenta cambios significativos y rápidos en los precios. Este aspecto resulta atractivo para los inversores, ya que les brinda oportunidades para aprovechar estas variaciones y obtener ganancias en sus transacciones.

Además, un sesgo a la derecha en medidas financieras como el Sharpe ratio, el retorno esperado, el retorno nominal, el ratio PE, el ratio PS, el ratio PB y el ROA, indica que la mayoría de los valores analizados presentan cifras más altas. Esto puede sugerir un rendimiento financiero superior o una valoración más favorable en comparación con las métricas consideradas.

En resumen, un sesgo a la derecha en estas variables implica una tendencia hacia valores más altos o favorables en las características y métricas analizadas. Esto puede indicar un rendimiento superior, una mayor actividad o una valoración más positiva en relación con estas variables.

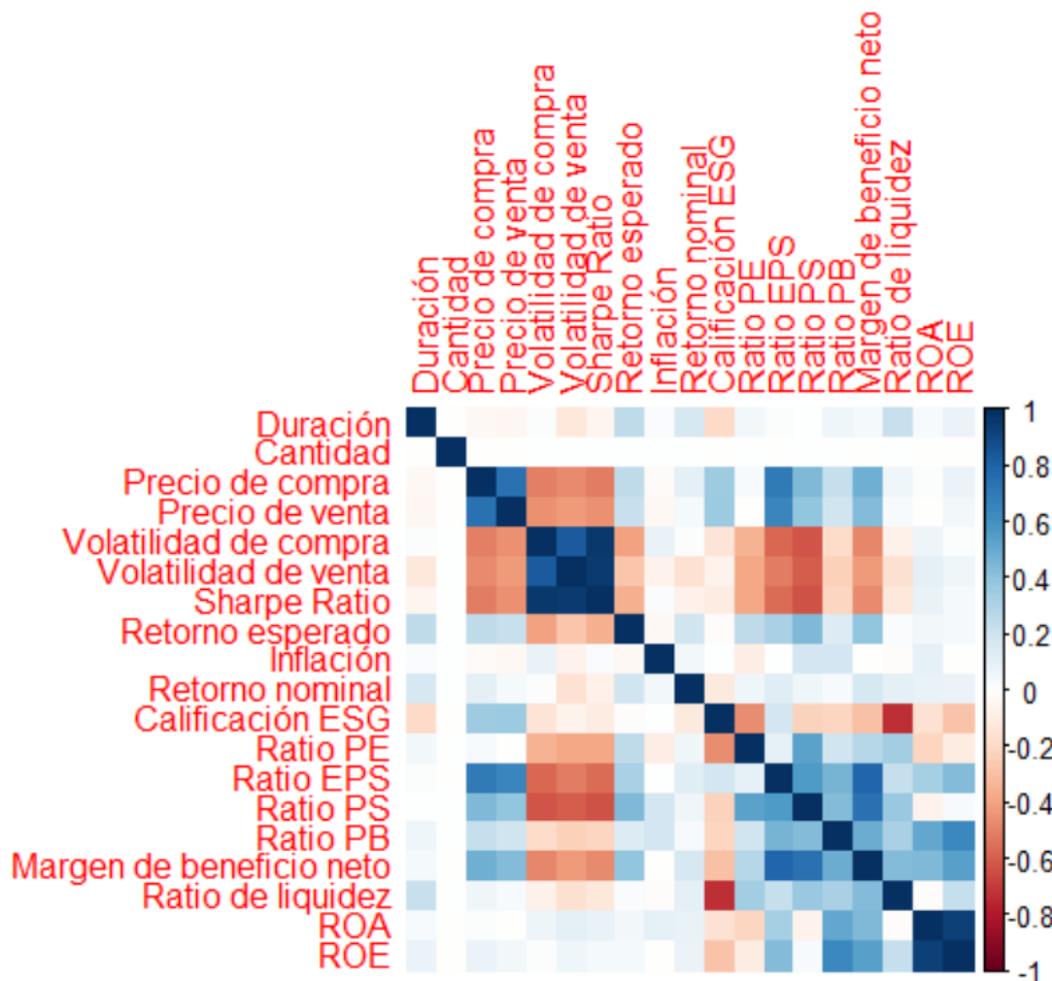
#### **4.1.4 Análisis de correlación**

El análisis de correlación es una técnica estadística que se utiliza para cuantificar la relación entre dos variables. Esta técnica permite identificar patrones y relaciones entre

pares de variables, lo cual puede ser muy útil para comprender cómo funcionan las cosas. En el caso de que exista una correlación significativa (superior a 0,7), se puede inferir cómo una variable influye en la otra. Además, el análisis de correlación ofrece numerosas ventajas como la identificación de variables importantes en un conjunto de datos, la predicción de resultados o la validación de hipótesis. Todo esto resulta interesante para el modelo predictivo que se llevará a cabo posteriormente.

Para llevar a cabo el cálculo de la correlación en el conjunto de datos, se han utilizado en R los comandos de `cor()` para obtener los valores de la correlación entre los pares de variables en forma de matriz y el `corrplot()` para obtener el gráfico que se muestra en la figura 10. En ella se muestra la fuerza y dirección de la correlación entre pares de variables de la matriz de correlación.

Figura 10. Matriz de correlación



Fuente: Elaboración propia

El gráfico rebela diversas conclusiones que se pueden inferir. En primer lugar, se observa que varias variables muestran una correlación mínima con el resto de las variables. Por ejemplo, la duración, la cantidad, la inflación y el retorno nominal exhiben colores muy claros en el gráfico, lo que sugiere que estas variables no tienen un poder de influencia significativo sobre las demás.

Sin embargo, las demás variables presentan distintos niveles de correlación entre sí. Específicamente, la variable ROE muestra una correlación alta con el ratio EPS (0,712), el ratio PB (0,775), el ROA (0,970), la calificación ESG (-0,759) y el margen de beneficio neto (0.548). El precio de compra presenta una correlación alta con el precio de venta (0.749), la volatilidad en el momento de compra (-0.601), la volatilidad en el momento de venta (-0.577) y el ratio EPS (0.808). A su vez, la volatilidad en el momento de compra presenta una correlación fuerte con la volatilidad en el momento de venta (0.839) y el ratio PS (-0.621). De igual manera, el Sharpe ratio está fuertemente correlacionado con el precio de compra (-0.513), el precio de venta (-0.558), la volatilidad en el momento de compra (0.960), la volatilidad en el momento de venta (0.957), el ratio EPS (-0.566) y el ratio PS (-0.647). La variable calificación ESG presenta una correlación significativa con el ratio de liquidez (-0.734) y también el margen de beneficio neto con el ratio EPS (0.792) y el ratio PS (0.743).

Según la matriz de correlaciones, se puede afirmar que no existe multicolinealidad entre las variables independientes, a excepción de aquellas que presentan una correlación excepcionalmente alta, lo que podría conllevar problemas de multicolinealidad. Esto se puede observar en ambas variables de volatilidad con el Sharpe ratio y entre el ROA y el ROE. Sin embargo, estas excepciones no influirán en los resultados de los análisis ni en el número de variables independientes. Esto se debe a que, aunque la correlación sea significativa entre las variables, esta no afecta de manera negativa a la calidad y precisión del análisis.

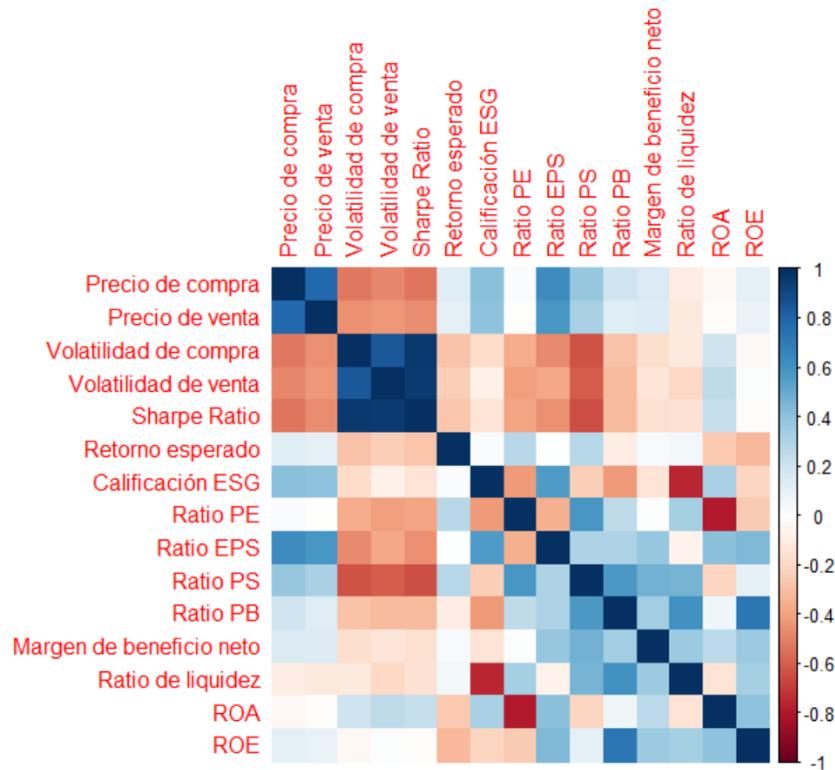
En conclusión, el análisis del gráfico rebela distintos niveles de correlación entre las variables, destacando relaciones significativas entre algunas de ellas, lo que proporciona información valiosa para comprender las interacciones y posibles influencias entre los distintos factores considerados.

#### 4.1.5 Análisis de correlación del ROE

Una vez que se ha obtenido una comprensión más completa de las variables y de sus interacciones, se procederá a realizar un análisis minucioso de la variable objetivo del modelo predictivo, el ROE, y de su relación con las demás variables. De este modo, se podrán descartar aquellas variables que presenten una correlación extremadamente baja con el ROE y que, por tanto, no aporten valor al análisis predictivo. En definitiva, se llevará a cabo una selección de variables que permita obtener resultados precisos y significativos en la predicción del ROE.

Los resultados de la matriz de correlaciones muestran una relación muy cercana a 0, apenas perceptible entre el ROE y la duración, la cantidad, el retorno nominal y la inflación. Por tanto, al eliminar estas variables sin correlación del modelo, se consigue una simplificación y mejora de la capacidad predictiva ya que el modelo se enfoca en aquellas variables que sí tienen una relación significativa con la variable objetivo (Figura 11).

Figura 11. Matriz de correlación con variables eliminadas



Fuente: Elaboración propia

## **4.2 ANÁLISIS PREDICTIVO**

### **4.2.1 Importancia del análisis predictivo**

Un modelo predictivo es una herramienta analítica que utiliza datos históricos para hacer predicciones sobre eventos futuros (Raheja and Jain, 2021). Su objetivo principal es proporcionar información anticipada y estimaciones basadas en patrones y tendencias identificadas en los datos pasados (Lee et al. 2022). El uso de modelos predictivos resulta interesante en varios campos y sectores, incluido el sector bancario, donde desempeñan un papel fundamental y se aplican en diversas variables y ratios financieros.

En el sector bancario, los modelos predictivos tienen una amplia gama de aplicaciones y razones para intentar predecir el ROE. En primer lugar, facilitan la toma de decisiones estratégicas al permitir a los líderes y gerentes evaluar el impacto de diferentes estrategias y asignar recursos de manera informada (Alonso and Carbó, 2020). Además, estos modelos son utilizados para la gestión de riesgos, al identificar y mitigar posibles riesgos financieros y operativos, lo que salvaguarda los activos del banco (Petrooulos et al. 2017).

Asimismo, la predicción del ROE permite a los bancos evaluar su rentabilidad y compararla con la de la competencia (Boyacioglu et al., 2009). Esto les brinda la oportunidad de identificar áreas de mejora y crecimiento, optimizando su rendimiento financiero y aumentando su competitividad en el mercado (Cox et al., 2014). Por último, estos modelos también sirven como herramienta para la planificación financiera a largo plazo, ya que proporcionan estimaciones del ROE que permiten proyectar ganancias y establecer metas realistas, facilitando la elaboración de planes financieros sólidos y la adopción de medidas proactivas para alcanzar los objetivos establecidos (Demyanyk et al., 2010).

### **4.2.2 Modelos predictivos**

Tras la formulación de las hipótesis y el análisis predictivo de los datos, se han desarrollado una serie de ecuaciones para construir un modelo predictivo que permita analizar los factores que influyen en la rentabilidad de los bancos, en concreto en el ROE. En total, se han creado 7 modelos predictivos, variando la inclusión de cada variable y manteniendo algunas constantes, con el propósito de determinar qué variables

contribuyen de manera significativa al rendimiento del ROE. Los modelos predictivos coinciden con las ecuaciones mostradas anteriormente en el apartado de metodología. De esta forma, se identifican los impulsores clave que afectan la rentabilidad de los bancos y comprender mejor su desempeño financiero.

#### **4.2.3 Resultados obtenidos en cada modelo predictivo**

A continuación, se presentan los resultados de los diferentes modelos que han sido definidos previamente con el objetivo de realizar un análisis predictivo en el sector bancario y predecir el retorno sobre el patrimonio (ROE). Para este estudio, se ha utilizado el lenguaje de programación R.

El Modelo 1 incluye el ROE y otras 14 variables. En todo modelo predictivo, los datos se dividen en un conjunto de entrenamiento (70% de las observaciones de la base de datos) y de prueba (30% de las observaciones restante). En este caso, los datos se dividen en el conjunto de entrenamiento con 23.997 observaciones y en el de prueba con 10.285.

Seguidamente, se crea el modelo predictivo con la función *lm()* donde se especifican las variables predictoras (todas las variables incluidas en el modelo que no son el ROE) y la variable respuesta, la que se busca predecir (ROE). Asimismo, para generar las predicciones se utiliza la función *predict()* que realiza predicciones basadas en el modelo ajustado y genera valores predictivos para futuras observaciones.

Para evaluar la eficacia del modelo se utiliza el MCO. Cuanto menor sea el valor del MCO, mejor será la capacidad de predicción del modelo, ya que indica que los valores predichos se acercan más a los valores reales. El MCO del modelo 1 es de 0,11923 lo que indica que, en promedio, el modelo tiene un error cuadrático de aproximadamente 0.11923 unidades al predecir el ROE en el conjunto de datos de prueba. Esto sugiere que la capacidad de predicción es buena.

Finalmente, se calcula el coeficiente de determinación ( $R^2$ ) del modelo utilizando la función *summary()*. El coeficiente de determinación evalúa la capacidad de ajuste del modelo predictivo, es decir, qué tan bien se ajusta a los datos observados. Un valor más cercano a 1 indica que el modelo explica mejor la variabilidad de la variable objetivo por lo que cuanto más alto sea el resultado, mejor. El coeficiente de determinación ( $R^2$ ) del modelo 1 es de 0.92388 lo que significa que aproximadamente el 92.39% de la

variabilidad del ROE en el conjunto de datos puede ser explicada por el modelo construido. Esto indica que el modelo tiene un buen ajuste y es capaz de capturar la mayor parte de la variabilidad presente en la variable dependiente.

En conclusión, para el modelo 1, un MCO bajo y un  $R^2$  alto indican que el modelo tiene una buena capacidad de predicción y ajuste en relación con el ROE.

El Modelo 2 incluye un total de 10 variables. Al evaluar el rendimiento, se obtiene un MCO de 0,18629. Asimismo, el coeficiente de determinación ( $R^2$ ) del modelo 3 es de 0,88089.

Por su parte, el Modelo 3 sigue el mismo esquema que el modelo 2, pero reemplaza la variable de precio de compra por la de precio de venta. El MCO es de 0,19486 y el  $R^2$  es de 0,87596.

El modelo 4 incluye también 10 variables. Se obtiene un MCO de 0,13765 y un  $R^2$  de 0,91292.

Las variables PE ratio y ROA presentan una alta correlación entre sí. Esto sugiere que si se elimina una de ellas la predicción del ROE no debería variar significativamente. En el modelo 5 se ha incluido el ROA y en el modelo 6 el PE ratio. Ambos modelos incluyen un total de 9 variables. El MCO del modelo 5 es de 0,18029, y su  $R^2$  es de 0,88782. Por otro lado, el MCO del modelo 6 es de 0,18830, lo que indica un nivel similar de error promedio en las predicciones en comparación con el modelo 5 y su  $R^2$  es de 0,88096.

Finalmente, para crear el Modelo 7, se ha eliminado la variable de ratio de liquidez ya que se encontró que la correlación con la variable de calificación ESG era considerablemente alta. Se optó por mantener la variable de calificación ESG ya que su correlación con el ROE es mayor. El MCO del modelo es de 0,22629 y su  $R^2$  de 0,85712.

Después de analizar los resultados de los modelos predictivos, se puede determinar qué hipótesis se respaldan en cada uno de ellos. Los modelos 1 y 5 respaldan todas las hipótesis, H1, H2, H3, H4, H5 y H6. Por otro lado, los modelos 2 y 3 respaldan las hipótesis H1, H3, H5 y H6 y rechazan las hipótesis H2 y H4. En el caso del modelo 4, se respaldan las hipótesis H1, H3, H4, H5 y H6 y rechaza la hipótesis H2. El modelo 6 respalda las hipótesis H1, H2, H3 y H4 y rechaza H5 y H6. Por último, el modelo 7 respalda las hipótesis H1, H2, H3, H4 y H6, y rechaza H5.

Al analizar los resultados de R con más profundidad, el modelo nos devuelve el intercepto y los coeficientes de cada variable independiente. Estos valores contienen información sobre la relación lineal entre la variable dependiente (ROE) y variables independientes. Por un lado, el intercepto, representa el valor estimado de la variable dependiente (ROE) cuando todas las variables independientes son iguales a cero, es decir, cuando no hay efecto de las variables independientes en la variable objeto de estudio. Por otro, los coeficientes, indican cómo cambia la variable dependiente en respuesta a un cambio unitario en la variable independiente correspondiente, manteniendo constantes el resto de las variables independientes. Un coeficiente positivo significa que hay una relación directa positiva, lo que implica que un aumento en el valor de la variable independiente está asociado con un aumento en el valor de la variable dependiente y viceversa. Es importante tener en cuenta que los coeficientes se interpretan en relación con las unidades de las variables involucradas. Toda esta información se presenta en la tabla 3 y recoge los valores de cada modelo para cada una de las variables que están en la ecuación.

Toda esta información se recoge en la tabla que se muestra a continuación.

*Tabla 3. Contribución de cada variable independiente al modelo predictivo*

	<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>	<b>Modelo 4</b>	<b>Modelo 5</b>	<b>Modelo 6</b>	<b>Modelo 7</b>
<b>Precio compra</b>	1,640	2,440 *	-	-	-	-	-
<b>Precio venta</b>	0,247	-	1,179	-	-	-	-
<b>Volatilidad de compra</b>	(-0,248)	-	-	-	-	-	-
<b>Volatilidad de venta</b>	1,474 *	-	-	-	-	-	-
<b>Sharpe Ratio</b>	(-0,928)	-	-	-	-	-	-
<b>Retorno esperado</b>	(-0,033)	(-0,054)	(-0,056)	(-0,032)	(-0,0415)	(-0,053)	(-0,061)
<b>Calificación ESG</b>	(-28,06) ***	(-39,34) ***	(-38,84) ***	(-27,45) ***	(-19,79) ***	(-19,30) ***	(-12,40) ***
<b>Ratio PE</b>	0,778 *	1,233 *	1,210	0,709	-	0,025	0,413
<b>Ratio EPS</b>	2,089 ***	2,901 *	2,946 **	2,123 *	1,395 *	1,460 *	1,357 *
<b>Ratio PS</b>	(-3,134) ***	(-3,510) ***	(-3,389) ***	(-2,997) **	(-1,943) *	(-2,224) **	(-2,652) **

<b>Ratio PB</b>	1,409 ***	-	-	1,460 *	2,103 **	2,195 **	1,794 *
<b>Margen de beneficio neto</b>	(-0,002)	(-0,190)	(-0,199)	0,020	0,161	0,242	0,083
<b>Ratio de liquidez</b>	(-8,891) ***	(-10,85) ***	(-10,80) ***	(-9,364) ***	(-7,606) ***	(7,632) ***	-
<b>ROA</b>	0,829	1,298 *	1,254 *	0,753	0,182	-	0,510
<b>ROE</b>	78,981	111,432	110,751	76,960	55,486	53,334	36,078
<b>Número de observaciones</b>	34.282	34.282	34.282	34.282	34.282	34.282	34.282
<b>Variables</b>	15	10	10	10	9	9	9
<b>R<sup>2</sup></b>	0,92388	0,8809	0,8760	0,9130	0,8878	0,8809	0,8571

*Fuente: Elaboración propia*

Según el análisis de la tabla 3, se concluye que todas las variables analizadas en las hipótesis originales tienen alguna influencia, en mayor o menor medida, en la predicción del ROE. Entre estas variables, la calificación ESG y el ratio de liquidez se destacan por tener valores de magnitud más altos, lo que indica una relación más fuerte con la variable dependiente. Además, al presentar un signo negativo, sugiere que un aumento en cualquiera de las dos variables está asociado con una disminución en el ROE. Esto puede ser interesante para los inversores y para estudiar el manejo de las inversiones.

En el caso de la variable del ratio PS, también se observa una relación negativa, aunque la fuerza de esta relación es considerablemente menor en comparación con el ratio de liquidez. Por otro lado, las demás variables estudiadas en las hipótesis (ratio PB, ratio EPS, ratio PE y ROA) presentan una relación positiva con el ROE, aunque su fuerza es más moderada, ya que los valores de los coeficientes son más bajos.

Es importante destacar que en la investigación no se incluyó ninguna hipótesis inicial que considerara el impacto de la calificación ESG de la inversión en la predicción del ROE. Sin embargo, esta variable muestra los valores más altos en la tabla, lo que indica una posible influencia significativa en el ROE. Por lo tanto, sería interesante incorporar esta variable en estudios futuros para evaluar su impacto en la predicción del ROE.

A continuación, se presenta una tabla a modo resumen que recoge toda la información mencionada anteriormente sobre los modelos predictivos y sus resultados. En la tabla 4 se incluye el número de variables por cada modelo, qué variables están incluidas, el resultado del MCO y qué hipótesis respalda cada modelo.

Tabla 4. Resultado de los modelos predictivos

	Variables	MCO	Hipótesis respaldadas
<b>MODELO 1</b>	- ROE - Precio de compra y venta - Volatilidad en la compra y venta - Sharpe ratio - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS, PB - Margen de beneficio neto - Ratio de liquidez - ROA	0,11923	H1, H2, H3, H4, H5 y H6
<b>MODELO 2</b>	- ROE - Precio de compra - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS - Margen de beneficio neto - Ratio de liquidez - ROA	0,18629	H1, H3, H5 y H6
<b>MODELO 3</b>	- ROE - Precio de venta - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS - Margen de beneficio neto - Ratio de liquidez - ROA	0,19486	H1, H3, H5 y H6
<b>MODELO 4</b>	- ROE - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS, PB - Margen de beneficio neto - Ratio de liquidez - ROA	0,13765	H1, H3, H4, H5 y H6
<b>MODELO 5</b>	- ROE - Retorno esperado - Calificación ESG - Ratios: EPS, PS, PB - Margen de beneficio neto - Ratio de liquidez - ROA	0,18029	H1, H2, H3, H4, H5 y H6
<b>MODELO 6</b>	- ROE - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS, PB - Margen de beneficio neto - Ratio de liquidez	0,18833	H1, H2, H3 y H4
<b>MODELO 7</b>	- ROE - Retorno esperado - Calificación ESG - Ratios: PE, EPS, PS, PB - Margen de beneficio neto - ROA	0,22629	H1, H2, H3, H4 y H6

Fuente: Elaboración propia

Teniendo en cuenta toda esta información y analizando los resultados de los diferentes modelos predictivos, se pueden obtener algunas conclusiones generales. En primer lugar, se realiza la comparación de MCO (Error cuadrático medio). En base a los resultados

obtenidos, se puede observar que el Modelo 1 tiene el MCO más bajo de 0,11923, lo que indica un menor error promedio en las predicciones en comparación con los otros modelos. Esto sugiere que el Modelo 1 tiene una mejor capacidad de predicción en relación con el ROE.

En segundo lugar, se compara el valor de  $R^2$  (Coeficiente de determinación). Según los resultados, el Modelo 1 vuelve a tener el valor más alto con un 0,92388, lo que implica que aproximadamente el 92% de la variabilidad del ROE puede ser explicada por las variables incluidas en este modelo. Asimismo, el Modelo 4 tiene un  $R^2$  significativamente alto con un valor de 0,91292, lo que implica que aproximadamente el 91% de la variabilidad del ROE puede ser explicada por las variables incluidas en este modelo. Esto indica que ambos modelos tienen un buen ajuste en relación con el ROE y en comparación con los otros modelos.

En último lugar, se analizan las variables relevantes y el número de variables presente en cada modelo. Al comparar los modelos, se puede observar que ciertas variables son incluidas en varios modelos, lo que sugiere que son consideradas relevantes para predecir el ROE. Estas variables incluyen el retorno esperado, la calificación ESG, el PE ratio, el EPS ratio, el PS ratio, el PB ratio, el margen de beneficio neto, el ratio de liquidez y el ROA. Estas variables muestran una correlación positiva con el ROE y parecen tener un impacto significativo en la predicción del ROE.

Basándonos en estas conclusiones, se puede afirmar que el Modelo 1, el Modelo 4 y el Modelo 5 muestran promesa en términos de capacidad de predicción y ajuste en relación con el ROE. Al elegir entre estos modelos predictivos, es importante considerar diversos factores, ya que no hay una única métrica que determine cuál modelo es mejor en todos los casos. Sin embargo, es relevante tener en cuenta que un modelo con menos variables podría ser preferible si logra un rendimiento similar o mejor que otro modelo con más variables. Un modelo más simple y con menos variables puede ser más fácil de interpretar y menos propenso a problemas de sobreajuste. Por lo tanto, se descartaría el Modelo 1 debido a que incluye un total de 15 variables. No obstante, las diferencias entre el Modelo 4 y el Modelo 5 no son tan significativas en términos de MCO y coeficiente de determinación, y resulta interesante mantener la información que nos proporciona la variable del PE ratio para predecir el ROE.

Después de completar la investigación, se ha determinado que ciertas variables tienen una influencia más significativa en la predicción del ROE en comparación con otras. Estas variables incluyen la calificación ESG, el ratio EPS, el ratio PS, el ratio PB y el ratio de liquidez, ya que muestran coeficientes con magnitudes más altas en los modelos predictivos utilizados. Todas ellas están presentes en los modelos predictivos 4 y 5. Esto implica que las hipótesis 2, 3, 4 y 5 son respaldadas por los resultados obtenidos en la investigación. Por otro lado, las hipótesis 1 y 6, serían rechazadas.

## 5. CONCLUSIONES

Tras completar la investigación y el análisis, se han obtenido conclusiones interesantes que contribuyen al objetivo principal de este estudio. El objetivo era predecir el ROE utilizando diversas variables y comprender cómo afectan a la eficiencia en el uso del capital para generar beneficios.

El enfoque consistió en determinar qué combinación de variables, al incluirse en un mismo modelo predictivo, permite una predicción más precisa del ROE. Para lograr este objetivo, se generaron siete modelos predictivos distintos, cada uno con diferentes variables independientes. Los resultados de estos modelos se analizaron a través del análisis de los coeficientes de las variables independientes, el Error Cuadrático Medio (MCO), el coeficiente de determinación y el número de variables incluidas. A partir de este análisis, se seleccionaron los modelos predictivos con mejores resultados. Es importante reconocer que el uso de regresiones por MCO implica suposiciones y limitaciones inherentes. En esta técnica se asume que la relación entre las variables es lineal, lo cual puede no ser válido en todos los casos por lo que se sugiere probar otro tipo de técnicas estadísticas que aborden el problema desde otra perspectiva.

Los modelos seleccionados fueron el modelo 4 y el modelo 5, respaldando las hipótesis 2 (ratio PS), 3 (ratio EPS), 4 (ratio PB) y 5 (ratio de liquidez), mientras que rechazan las hipótesis 1 (ratio PE) y 6 (ROA). Cada hipótesis afirmaba que la variable que incorporaba contribuía significativamente a la hora de predecir la variable objetivo, el ROE. Estos hallazgos cumplen con los objetivos planteados al inicio de la investigación, lo que demuestra que el proceso metodológico seguido ha sido adecuado.

Los resultados de este estudio tienen importantes implicaciones prácticas y teóricas en el ámbito financiero y en la toma de decisiones. Al predecir el ROE utilizando diferentes variables, se proporciona a las instituciones financieras y a otros actores del mercado una comprensión más profunda de los factores que influyen en la rentabilidad de una entidad.

Cabe destacar que la variable de calificación ESG, que no se incluyó en las hipótesis iniciales, ha mostrado una contribución significativa pero negativa en la predicción del ROE, siendo incluso la variable con mayor influencia. Esto contradice los resultados analizados en la literatura lo que puede sugerir que las diferencias en la metodología y

técnicas estadísticas han conducido a resultados divergentes. Sin embargo, esto plantea la posibilidad de investigar más a fondo la relación entre esta variable y el ROE, así como analizar por separado la importancia de cada pilar (E, S y G) para comprender su influencia en la predicción del ROE. Los resultados obtenidos de ESG pueden revelar que los inversores no consideran estos aspectos como factores determinantes al tomar decisiones de inversión y que no valoran la participación de un banco en iniciativas de responsabilidad social.

Es importante tener en cuenta que la muestra de datos utilizada pertenece al NYSE y abarca observaciones hasta 2019, lo que implica que la base de datos no refleja información reciente. Sería interesante actualizar la base de datos y realizar el estudio nuevamente para verificar si las tendencias y patrones observados varían a lo largo del tiempo o si se mantienen.

Es importante resaltar la importancia de llevar a cabo investigaciones futuras que incluyan información actualizada que abarque el período durante y posterior a la pandemia para ver cómo la pandemia ha afectado las variables de interés y comparar los resultados previos y posteriores para obtener una visión más completa y actualizada. Asimismo, es relevante enfatizar que los hallazgos pueden no reflejar completamente la realidad debido a la ausencia del impacto inicial de la pandemia.

Además, se sugiere explorar la posibilidad de realizar un estudio similar en otro sector distinto al de la banca, o utilizar las variables del conjunto de datos para predecir otras variables relevantes que pueden ayudar a los bancos a gestionar mejor sus inversiones. Esto permitiría ampliar el alcance y la aplicación de los modelos predictivos desarrollados.

A lo largo del proceso de investigación, se identificaron algunas dificultades, como la búsqueda de una base de datos con información relevante y confiable que contase con un número de observaciones y variables considerable, así como el manejo y procesamiento de los datos para garantizar el buen funcionamiento del modelo predictivo. Dependiendo de las decisiones tomadas en relación con los valores atípicos y la normalización de variables, entre otros, los resultados se veían afectados notablemente. Estos desafíos resaltan que es esencial el hecho de comprender las técnicas estadísticas y los métodos de

análisis de datos adecuados para abordar las preguntas de investigación de manera precisa.

En conclusión, este trabajo de investigación ha permitido obtener conclusiones significativas y ofrecer conocimiento valioso sobre la predicción del ROE en el sector bancario. Los resultados tienen implicaciones prácticas y teóricas importantes, abriendo nuevas oportunidades de investigación y aplicaciones en el ámbito financiero y en la toma de decisiones.

## 6. BIBLIOGRAFÍA

- Adrianzen Cabrera, Carlos Manuel (2016). Tesis Doctoral: La rentabilidad de los bancos comerciales y el ambiente macroeconómico: El caso peruano en el periodo 1982-2014. Director: Dr. Jesús Abad Puente Universitat Politècnica de Catalunya Departamento de Organización de Empresas. Programa de Doctorado en Administración y Dirección de Empresas
- Alonso, A. and Carbó, J. M. (2020) Machine learning in credit risk: measuring the dilemma between prediction and supervisory cost. Banco de España. Documento de Trabajo N° 2032.
- Athanasoglou, P., Brissimis, S. N. and Delis, M. D. (2008). Bank-Specific, Industry-Specific and Macroeconomic Determinants of Bank Profitability | Semantic Scholar. Economics Business SSRN Electronic Journal.
- Aydoğmuş, M., Gülay, G., Ergun, K. (2022). Impact of ESG performance on firm value and profitability. *Borsa Istanbul Review* 22(2), 119-127.
- Babalola Y. A. and Abiola F. R (2013). Financial ratio analysis of firms: A tool for decision making. *International Journal of Management Sciences*, 1(4), 132-137.
- Bawack, R. E. and Ahmad, M. O. (2021). Understanding business analytics continuance in agile information system development projects: an expectation-confirmation perspective. *Information Technology & People*, 1.
- Bayrak, T. (2015). A review of business analytics: A business enabler or another passing fad. *Procedia-Social and Behavioral Sciences*, 195, 230-239.
- Ben Khediri, K., Ben Ali, M.S. and Ben-Khedhiri, H. (2010). Bank-specific, Industry-specific and Macroeconomic Determinants of African Islamic Banks' Profitability, *International Journal of Business and Management Science*, 3(1), 39-56.
- Berman, K., Knight, J. and Case, J. (2013). *Financial Intelligence. A Manager's Guide to Knowing What the Numbers Really Mean*. Business Literacy Institute, Inc. USA, 284 pages.
- Berzkalne, I. and Zelgalve, E. (2014). Return on equity and company characteristics: an empirical study of industries in Latvia. The 8th International Days of Statistics and Economics, Prague, September 11-13, 2014.
- Bezawada Brahmaiah, R. (2018). Factors Influencing Profitability of Banks in India. *Theoretical Economics Letters*, 8, 3046-3061.
- Bhowmik R. and Wang S. (2020). Stock Market Volatility and Return Analysis: A Systematic Literature Review. *Entropy (Basel)*, 22(5):522.
- Biecek, P. and Burzykowski, T. (2021). *Explanatory model analysis: explore, explain, and examine predictive models*. CRC Press, Taylor and Francis.
- Black, A., Wright, P. and Davies, J. (2001). *In search of shareholder value*. 2nd Edition. London: Pearson.
- Boyacioglu, Melek Acar, Yakup Kara, and Ömer Kaan Baykan (2009) "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey." *Expert Systems with Applications* 36(2), 3355-3366.

- Brynjolfsson, E., Jin, W. and McElheran, K. (2021). The Power of Prediction: Predictive Analytics, Workplace Complements, and Business Performance. *Workplace Complements, and Business Performance*, 1.
- Burja, V. and Mărginean, R. (2014). The Study of Factors that may Influence the Performance by the Dupont Analysis in the Furniture Industry. *Procedia Economics and Finance*, 16, 213-223.
- Campbell, J. Y. and Shiller, R. J. (2001). Valuation ratios and the long-run stock market outlook: an update. NBER Working Paper No. 8221. JEL No. G12.
- Casamayou Calderón, E. T. (2019). Apalancamiento financiero y su influencia en la rentabilidad de las empresas, periodo 2008-2017 QUIPUKAMAYOC ENERO-ABRIL, 2019 27(53), 33-39.
- Chandra, T., Junaedi, A. T., Wijaya, E., Suharti, S., Mimelientesa, I. and Ng, M. (2019). The effect of capital structure on profitability and stock returns Empirical analysis of firms listed in Kompas 100. *Journal of Chinese Economic and Foreign Trade Studies*, 12(2), 74-89.
- Chidozie, U. E. and Ayadi, F. S. (2017). Macroeconomy and Banks' Profitability in Nigeria. *African Research Review*, 11(46), 121-137.
- Circiumaru, D., Siminica, M. and Marcu, N. (2010). A Study on the Return on Equity for the Romanian Industrial Companies. *Annals of University of Craiova – Economic Sciences Series*, 2(38).
- Clément, A., Robinot, É. and Trespeuch, L. (2023), "The use of ESG scores in academic literature: a systematic literature review", *Journal of Enterprising Communities: People and Places in the Global Economy*, Vol. ahead-of-print No. ahead-of-print.
- Contreras, I. (2006). Análisis de la rentabilidad económica (ROI) y financiera (ROE) en empresas comerciales y en un contexto inflacionario. *Visión Gerencial enero-junio*, 1, 13-28.
- Correia, C., Flynn, D., Uliana, E. and Wormald, M. (2003). *Financial management*. 5th Edition. Cape Town: Juta.
- Cox, Raymond AK, and Grace W-Y. Wang. (2014) "Predicting the US bank failure: A discriminant analysis." *Economic Analysis and Policy*, 44(2), 202-211.
- Damodaran, A. (2007). Return on Capital (ROC), Return on Invested Capital (ROIC) and Return on Equity (ROE): Measurement and Implications. *Stern School of Business*, July 2007.
- DeAngelo, H., DeAngelo, L. and Whited, T. (2011). Capital structure dynamics and transitory debt *Journal of Financial Economics*, 99(2), 235-261.
- Delen, D., Kuzey, C. and Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, 40(10), 3970-3983.
- Demyanyk, Yuliya, and Iftekhhar Hasan (2010). "Financial crises and bank failures: A review of prediction methods." *Omega* 38(5), 315-324.
- El Khoury, R., Nasrallah, N. and Alareeni, B. (2023). ESG and financial performance of banks in the MENAT region: concavity–convexity patterns. *Journal of Sustainable Finance & Investment*, 13(1). Special Issue: Banking Regulation and Post-Covid-19 crisis.

- Eljelly, A. (2004). The cyclical behavior of optimal bank capital. *Journal of Banking and Finance*, 28(3), 1469-1498.
- European Central Bank (2010). Beyond ROE-How to measure bank performance. Appendix to the report on EU banking structures, September, 2010.
- Fama, E. F., and French, K. (1988). "Dividend yields and expected stock return," *Journal of Financial Economics*, 22, 3-25.
- Fama, E. F., and French, K. (1992) "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427-465.
- Fama, E. F., and French, K. (1995) "Size and book-to-market factors in earnings and returns," *Journal of Finance*", 50(1), 131-155.
- Fire, C., Ross, S.A., Westerfield, R.W. and Jordan, B.D. (2004). *Fundamentals of corporate finance*. 3rd South African edition. New York: McGraw-Hill.
- Flamini, V., McDonald, C. and Schumacher, L. (2009). *The Determinants of Commercial Bank Profitability in Sub-Saharan Africa*. International Monetary Fund WP/09/15 IMF Working Paper African Department.
- Ganesan, P. (2001) Determinants of Profits and Profitability of Public Sector Banks in India: A Profit Function Approach. *Journal of Financial Management and Analysis*, 14, 27-37.
- Grove, G., et al. (2014) Profitability and Performance Measurement of U.S. Regional Banks. *Advances in Management Accounting*, 24, 189-237.
- Hassan, S. (2019). Impact of Internal Factors on Banks' ROE. *International Conference on Business and Management*, 534–537. Dhaka. ISBN: 978-984-344-3540.
- Horobet, A., Radulescu, M., Belascu, L. and Dita, S. M. (2021). Determinants of Bank Profitability in CEE Countries: Evidence from GMM Panel Data Estimates. *Journal of Risk and Financial Management* 14:30.
- Ige Abolarinwa, S., Ikechukwu Asogwa, S., Ezenwakwelu, C. A., Court, T. O. and Adedoyin, S. (2020). Corporate growth strategies and financial performance of quoted manufacturing firms in Nigeria: The mediating role of global economic crises, *Cogent Economics & Finance*, 8(1): 1750259.
- Jaber, J. J. and Al-Khawaldeh, A. A. (2014). The Impact of Internal and External Factors on Commercial Bank Profitability in Jordan. *International Journal of Business and Management*; 9(4).
- Jara Quevedo, G., Massuh Arreaga, O., Ibarra Velásquez, A., Castro Carrasco, J., Zurita Vargas, S. and Mendoza Malo, A. (2021). Factores internos y externos relacionados con la ROA de bancos privados ecuatorianos desde la dolarización. *Revista Compendium: Cuadernos de Economía y Administración*, 8(2), 175-190.
- Kabajeh, M. A. M., Nu'aimat, D. S. and Dahmash, D. F. (2012). The Relationship between the ROA, ROE and ROI Ratios with Jordanian Insurance Public Companies Market Share Prices. *International Journal of Humanities and Social Science*, 2(11), 115-120.
- Khan, M. B., Gul, S., Razzaq, N. and Kamran, A. (2012). Financial Ratios and Stock Return Predictability (Evidence from Pakistan). *Research Journal of Finance and Accounting* 3(10).
- Kharatyan. D. (2016). Ratios and indicators that determine return on equity. Final Dissertation presented to Instituto Politécnico de Bragança.

- Kheradyar, S., Ibrahim, I. and Mat Nor, F. (2011). Stock Return Predictability with Financial Ratios. *International Journal of Trade, Economics and Finance*, 2(5).
- Klaassen, P. and Eeghen, I. (2015) Analyzing bank performance –linking RoE, RoA and RAROC: U.S. commercial banks 1992–2014 *The Journal of Financial Perspectives*, July 2015 | 3(2).
- Kosmidou, K. (2008). The determinants of banks’ profits in Greece during the period of EU financial integration. *Journal of economic literature*, 34(3), 146-159.
- Kothari, S. P. and Shanken, J. (1997) “Book-to-market, dividend yield, and expected market returns: a time series analysis,” *Journal of Financial Economics*, 44, 169–203.
- Kumbure, M. M., Lohrmann, C., Luukka, P. and Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review *Expert Systems With Applications*, 197: 116659.
- Lee, C. S., Cheang, P. Y. S. and Moslehpour, M. (2022) Predictive Analytics in Business Analytics: Decision Tree Article. *Advances in Decision Sciences* March.
- Li, X., Feng, H., Zhao, S., Carter, D. A. (2021). The effect of revenue diversification on bank profitability and risk during the COVID-19 pandemic *Finance Research Letters*, 43: 101957
- Liem, P. F. and Basana, S. (2012). Price Earnings Ratio and Stock Return Analysis (Evidence from Liquidity 45 Stocks Listed in Indonesia Stock Exchange) *Jurnal Manajemen dan Kewirausahaan*, 14(1).
- Liu, Z., Zhu, Z., Gao, J., and Xu, C. (2017) Forecast Methods for Time Series Data: A Survey. Digital Object Identifier 10.1109/ACCESS.2017.
- Máster Cámara Málaga (2022). Diferencias entre solvencia, liquidez y rentabilidad.
- Maverick, J. B. (2021). Does a High Price-to-Book Ratio Correlate to ROE? <https://www.investopedia.com/ask/answers/010915/does-high-price-book-ratio-correspond-high-roe.asp>.
- Maverick, J. B. (2022). What Is the Average Return on Equity (ROE) of Banks? <https://www.investopedia.com/ask/answers/040815/what-level-return-equity-common-company-banking-sector.asp>.
- Malagrino, L. S., Roman, N. T., and Monteiro, A. M. (2018). Forecasting stock market in dex daily direction: A Bayesian network approach. *Expert Systems With Applications*, 105, 11–22.
- Moussu, C. and Petit-Romec, A. (2017). ROE in Banks: Performance or Risk Measure? Evidence from Financial Crises. *Finance*, 38(2), 95-133.
- Moussu, C. (2018). Bank capital and RoE: erroneous beliefs and financial instability *Special issue of Réalités Industrielles*.
- Nunes, P.M., Viveiros, A. and Serrasqueiro, Z. (2012). Are the Determinants of Young SME Profitability Different? Empirical Evidence Using Dynamic Estimators. *Journal of Business Economics and Management*, 13(3), 443-470.
- Ongore, V. and Kusa, G. (2013). Determinants of Financial Performance of Commercial Banks in Kenya, *International Journal of Economics and Financial issues*, 3(1), 237-252.

- Pasiouras, F. and Kosmidou, K. (2007) Factors Influencing the Profitability of Domestic and Foreign Commercial Banks in the European Union. *Research in International Business and Finance*, 21(2), 222-237.
- Pennacchi, G. and Santos, J. A. C. (2018). Why Do Banks Target ROE? Federal Reserve Bank of New York *Liberty Street Economics* (blog), October 9, 2018, <http://libertystreeteconomics.newyorkfed.org/2018/10/why-do-banks-target-roe.html>.
- Petriaa, N., Caprarub, B. and Ihnatovc, J. (2015). Determinants of banks' profitability: evidence from EU 27 banking systems. *Procedia Economics and Finance*, 20, 518 – 524. 7th International Conference on Globalization and Higher Education in Economics and Business Administration, GEBA 2013.
- Petropoulos, A., Siakoulis, V., Stavroulakis, E. and Vlachogiannakis, N. E. (2017) Predicting bank insolvencies using machine learning techniques. Bank of Greece, 3 Amerikis, 10250 Athens, Greece.
- Pontiff, J. and L. Schall, L. (1998). "Book-to-market ratios as predictors of market returns," *Journal of Financial Economics*, 49, 141–160.
- Pradhan, R. S. and Shrestha, D. (2016) Impact of Liquidity on Bank Profitability in Nepalese Commercial Banks. *SSRN Electronic Journal*.
- Raheja, S. and Jain, D. (2021) "AN EMPIRICAL STUDY OF SELECTED INDIAN COMPANIES' TALENT MANAGEMENT", *Amity Journal of Professional Practices*. Florida, USA, 1(1).
- Rappaport, A. (1986). *Creating shareholder value*. New York: The Free Press.
- Saleem, Q. and Rehman, R. (2011). Impacts of liquidity ratios on profitability. *Interdisciplinary Journal of Research in Business*, 1(7), 95-98.
- Santander Universidades (2022). ¿Qué es el ROE y cómo se calcula?: <https://www.becas-santander.com/es/blog/que-es-roe.html>.
- Singh, N. P., and Tandon, A. (2019). The Effect of Dividend Policy on Stock Price: Evidence from the Indian Market. *Asia-Pacific Journal of Management Research and Innovation*, 15(1–2), 7–15.
- Solano-Benavides, E., Alandete-Brochero, N., Anguila-Carrillo, A. and Villanueva-Imitola, A. (2022). Índices financieros que determinan el valor de la acción patrimonial en empresas de la Bolsa de Valores de Colombia, 2015-2018 *Información Tecnológica*, 33(3), 81-88.
- Stowe, J.D., Robinson, T.R., Pinto, J.E. and McLeavy, D.W. (2002). *Analysis of equity investments: Valuation*. Baltimore: AIMR.
- Sufian, F. and Chong, R.R. (2008). Determinants of bank profitability in a developing economy: Empirical evidence from the Philippines, *Asian Academy of Management Journal of Accounting and Finance*, 4(2), 91-112.
- Taani, K. and Banykhaled, M. H. H. (2011). The effect of financial ratios, firm size and cash flows from operating activities on earnings per share:(an applied study: on Jordanian industrial sector). *International Journal of Social Sciences and Humanity Studies*, 3(1), 1309-8063.
- Tuvadaratragool, S. (2013). *The role of financial ratios in signalling financial distress: evidence from Thai listed companies* [Southern Cross University]. Thesis

[https://researchportal.scu.edu.au/esploro/outputs/doctoral/The-role-of-financial-ratios-in/99101282152\\_0902368](https://researchportal.scu.edu.au/esploro/outputs/doctoral/The-role-of-financial-ratios-in/99101282152_0902368).

- Usman Arshad, M. (2021). The relevance of earning-to-price and ROE predictability for explaining Shenzhen stock exchange (SZSE), returns in China: A dynamic panel data approach. *The Journal of Corporate accounting & Finance*, 32(3), 94-109.
- Velásquez Acosta, A. (2022) La Relación de los Componentes ESG con el Desempeño Financiero de las Empresas en el Entorno Colombiano. ¿Cuál es la relación entre los componentes ESG y el desempeño financiero de las empresas colombianas? Universidad de los Andes Trabajo Fin de Grado 2022-2.
- Vera-Gilces, P., Camino-Mogro, S., Ordeñana-Rodríguez, X. and Cornejo-Marcos, G. (2020) A look inside banking profitability: Evidence from a dollarized emerging country, *The Quarterly Review of Economics and Finance*, 75, 147-166.
- Wet, J.H.v.H. and Toit, E. (2007). Return on equity: A popular, but flawed measure of corporate financial performance *S.Afr.J.Bus.Manage.*, 38(1), 1-11
- Wu, W8. T. A. (2014). The P/E Ratio and Profitability. *Journal of Business & Economics Research (Online)*, 12(1), 65-76.

## 7. ANEXO

Este anexo muestra parte del código desarrollado en R para el procesamiento de los datos y la creación de los modelos predictivos.

```
###-----TFG BUSINESS ANALYTICS-----  
###-----MARIA IBORRA MESEGUER - 201808242 -----  
###-----CARGA DE LOS DATOS-----  
library(readxl) #para poder leer un archivo excel  
library(tidyr)  
library(ggplot2) #paquete que permite hacer visualizaciones  
library(tidyverse)  
library(MASS) #para poder hacer normalización box-cox  
library(forecast)  
library(Hmisc) #imputar valores perdidos, infinitos etc  
library(FactoMineR) #para hacer PCA  
library(missMDA)  
library(rpart)  
library(imputeTS)  
library(dplyr)  
library(glmnet)  
library(caret)  
library(corrplot)  
  
data <- read_excel("400k_bank.xlsx") #nombre del archivo excel con las observaciones de  
bancos  
  
colSums(is.na(data[, 5:ncol(data)])) #ver si hay algún valor perdido: no hay valores perdidos  
#Hay valores perdidos en el Price_SELL: los reemplazamos con la media de esos precios  
media_precio <- mean(data$Price_SELL, na.rm = TRUE)  
data$Price_SELL <- round(data$Price_SELL %>%  
      replace(is.na(.),media_precio), digits = 2)  
colSums(is.na(data[, 5:ncol(data)])) #ver si hay algún valor perdido: no hay valores perdidos  
View(data) #vemos las observaciones del dataset  
###-----ANÁLISIS DESCRIPTIVO DE TODOS LOS DATOS-----  
#Obtener el tamaño de la base de datos
```

```

dim(data)

#Estructura de la base de datos

str(data)

###-----ANÁLISIS DESCRIPTIVO VARIABLES CATEGÓRICAS-----

#Frecuencia de cada variable categórica - De las variables numéricas la probabilidad de que se
repitan los números es muy baja

bancos <-table(data$Company) #AXP-15.955 BAC-16.101 DB-15.972 JPM-15.908 SC-
11.819 UBS-9.002

colores_bancos <- c("deepskyblue2", "violetred2", "mediumorchid1", "green2", "yellow2",
"red2" )

#Visualización de la variable categórica

pie(bancos,

  col = colores_bancos,

  labels = paste(names(bancos), "(", round(100*bancos/sum(bancos),1),"%)",

  cex = 1.2,

  clockwise = TRUE)

legend(1.7,0.9,

  legend = names(bancos),

  fill = colores_bancos,

  cex = 0.6,

  title = "Bancos")

#Mismo proceso tipo de inversión

inversion <-table(data$`Type of investment`) #BAD-59.592 GOOD-25.165

colores_inversion <- c("red2", "green2")

pie(inversion,

  col = colores_inversion,

  labels = paste(names(inversion), "(", round(100*inversion/sum(inversion),1),"%)",

  cex = 1.5,

  clockwise = TRUE)

legend(1,0.9,

  legend = names(inversion),

  fill = colores_inversion,

  cex = 1.5,

  title = "Tipo de inversión")

```

```

###-----ANÁLISIS DESCRIPTIVO DE VARIABLES NUMÉRICAS-----
summary(data[,5:ncol(data)]) #resumen estadístico de las variables numéricas
###-----ANÁLISIS DE DISTRIBUCIÓN DE VARIABLES Y VALORES ATÍPICOS---
-----
hist(data$Horizon, main="Duración (días) de la transacción ", xlab = "Duración", ylab =
"Frecuencia") #asimetría hacia la derecha - positiva
ggplot(data, aes(x = Horizon)) + geom_boxplot() + labs(title = "Boxplot - Duración de la
transacción") #en el gráfico no parece que haya outliers
outliers_duracion <- boxplot.stats(data$Horizon)$out #el vector nos sale vacío
###-----ELIMINACIÓN DE OUTLIERS-----
indices_a_eliminar <- c() # Crea un vector vacío para almacenar los índices de las filas a
eliminar
indices_a_eliminar <- union(indices_a_eliminar, which(data$Amount %in% outliers_amount))
# Agrega los índices de las filas que contienen outliers en cada variable a 'indices_a_eliminar'
###-----ANÁLISIS DE CORRELACIÓN-----
correlation_data <- cor(data[, c(5,6,9:25)])
corrplot(correlation_data, method = "color")
View(correlation_data)
###-----NORMALIZACIÓN DE VARIABLES NUMÉRICAS-----
#La función scale normaliza los datos en R con distribución normal
#La función log normaliza los datos en R con distribución asimétrica
#Box - Cox: En caso de que estas técnicas no funcionen para crear una distribución normal
#Luego se volverá a hacer el hist() de cada variable para ver si el resultado ha funcionado
#ROE tiene una distribución normal -> scale()
data$ROE <- scale(data$ROE)
#Para las variables que presentan asimetría, es mejor usar una transformación logarítmica
hist(data$ROE, main="Rendimiento de los fondos propios (ROE) del sector", xlab = "ROE",
ylab = "Frecuencia") #normal
#Price BUY
class(data$Price_BUY)
price_buy_norm <- BoxCox(data$Price_BUY, lambda = "auto")
data$Price_BUY <- price_buy_norm
hist(data$Price_BUY, main="Precio de compra de las acciones ", xlab = "Precio de compra",
ylab = "Frecuencia") #asimetría hacia la derecha - positiva
pca_variables <- data[, c(9:14,17:25)]

```

```

View(pca_variables)

# Reemplazar los valores -Inf por el valor mínimo no negativo de la variable
pca_variables$PE[pca_variables$PE == -Inf] <- min(pca_variables$PE[pca_variables$PE > 0],
na.rm = TRUE)

summary(pca_variables$PE)

###-----MODELO PREDICTIVO 1: TODAS LAS VARIABLES-----

modelo1 <- pca_variables
modelo1 <- as.matrix(modelo1)
modelo1[!is.finite(modelo1)] <- mean(modelo1, na.rm = TRUE)
modelo1 <- as.data.frame(modelo1)

#Dividir los datos en un conjunto de entrenamiento y prueba
#70% de los datos al conjunto de entrenamiento y 30% de los datos al de prueba
set.seed(123) #semilla para asegurar que los datos son reproducibles

train_index1 <- sample(1:nrow(modelo1), 0.7*nrow(modelo1)) #seleccionamos el 70% de los
datos

#TRAIN SUB DATASET
train_data1 <- modelo1[train_index1, ] #creamos la variable de entrenamiento
train_data1 <- as.matrix(train_data1)
train_data1[!is.finite(train_data1)] <- mean(train_data1, na.rm = TRUE)
train_data1 <- as.data.frame(train_data1)
any(is.na(train_data1)) #False
train_data1 <- as.matrix(train_data1)
any(!is.finite(train_data1)) #false: no hay valores infinitos
train_data1 <- as.data.frame(train_data1)
#train_data1[!is.finite(train_data1)] <- impute(train_data1[!is.finite(train_data1)], mean)
#any(!is.finite(train_data1)) #siguen habiendo valores infinitos
#inf_cols_train <- colSums(!is.finite(train_data1)) #eliminamos los valores infinitos
# Seleccionar las columnas que no tienen valores infinitos
#train_data1 <- train_data1[, inf_cols_train == 0]
#any(!is.finite(train_data1))

#TEST
test_data1 <- modelo1[-train_index1, ] #el 30% de los datos que no están en la variable de
entrenamiento son del test

test_data1 <- as.matrix(test_data1)

```

```

test_data1[!is.finite(test_data1)] <- mean(test_data1, na.rm = TRUE)
test_data1 <- as.data.frame(test_data1)
any(is.na(test_data1)) #False: no hay valores faltantes
test_data1 <- as.matrix(test_data1)
any(!is.finite(test_data1)) #False: No hay valores infinitos
test_data1 <- as.data.frame(test_data1)
#test_data1[!is.finite(test_data1)] <- impute(test_data1[!is.finite(test_data1)], mean)
#any(!is.finite(test_data1)) #True: siue habiendo valores infinitos
#inf_cols_test <- colSums(!is.finite(test_data1)) #eliminamos los valores infinitos
#test_data1 <- test_data1[, inf_cols_test == 0]
#any(!is.finite(test_data1))
#comprobamos que tienen las mismas dimensiones
dim(train_data1)
dim(test_data1)
#Creación del modelo
modelo1_valoracion <- lm(ROE ~ Price_BUY + Price_SELL + Volatility_Buy +
Volatility_Sell + `Sharpe Ratio` + `Expected Return` + `ESG ranking` + PE + EPS + PS + PB +
`Net Profit Margin` + `Current Ratio` + ROA, data = train_data1) #Ajustamos el modelo de
regresión lineal
predicciones_modelo1 <- predict(modelo1_valoracion, newdata = test_data1) #Nos hace las
predicciones en el test
MCO_modelo1 <- mean((predicciones_modelo1 - test_data1$ROE)^2) # Calcular el error de
predicción
r2_modelo1 <- summary(modelo1_valoracion)$r.squared

```