



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Análisis y Visualización de Datos del IBEX 35:
Identificación de Factores Predictivos para el Éxito
Empresarial

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Miulescu Alecu, Fabian Cristian

Tutor/a: Martínez Plumed, Fernando

CURSO ACADÉMICO: 2023/2024

Agradecimientos

A mi madre, por ser la persona que más ha creído en mi en todo este proceso, por inspirarme y ser el mayor ejemplo de superación que he podido tener.

A Inés, por confiar en mi como nunca pensé que alguien lo haría, por celebrar cada pequeño avance como si fuese tuyo.

A mi tutor Fernando, por trasmitirme la confianza y darme inspiración en momentos donde todo parecía oscuro, por ser un modelo a seguir.

A las maravillosas clases de Micro I, por descubrirme el maravilloso mundo de la economía y las finanzas, un mundo libre.

Os quiero a todos,

One Lova.



Resumen

El interés en este tema surge debido a la importancia del IBEX 35 como el principal indicador de la salud económica y financiera del mercado bursátil español, proporcionando una visión instantánea del estado de la economía española, influyendo en decisiones económicas, políticas y de inversión a nivel nacional e internacional.

Este trabajo tiene como objetivo analizar las variables que influyen en el éxito de las empresas cotizadas en el IBEX 35 en los últimos 10 años. Este estudio se centra en recopilar y examinar contenidos relacionados con cifras de explotación, datos financieros, encuestas y valoraciones externas de las empresas objetivo, así como posibles factores externos igualmente influyentes, utilizando para ello herramientas de visualización y análisis de datos.

El trabajo se estructura en diversas fases, empezando por la creación de un Dataset, seguido por el análisis tanto descriptivo como predictivo de los datos, y culminando con la visualización e interpretación de los datos obtenidos. Este enfoque integrado asegura una comprensión profunda de los factores de éxito en el mercado bursátil español, proporcionando una herramienta útil para inversores, analistas y gestores empresariales.

Palabras clave: análisis de datos, datos financieros, herramientas de visualización, IBEX 35

Resum

L'interès en este tema sorgeix a causa de la importància de l'IBEX 35 com el principal indicador de la salut econòmica i financera del mercat borsari espanyol, proporcionant una visió instantània de l'estat de l'economia espanyola, influint en decisions econòmiques, polítiques i d'inversió a nivell nacional i internacional.

Este treball té com a objectiu analitzar les variables que influeixen en l'èxit de les empreses cotitzades en l'IBEX 35 en els últims 10 anys. Este estudi se centra en recopilar i examinar continguts relacionats amb xifres d'explotació, dades financeres, enquestes i valoracions externes de les empreses objectiu, així com possibles factors externs igualment influents, utilitzant per a això ferramentes de visualització i anàlisi de dades.

El treball s'estructura en diverses fases, començant per la creació d'un Dataset, seguit per l'anàlisi tant descriptiu com predictiu de les dades, i culminant amb la visualització i interpretació de les dades obtingudes. Este enfocament integrat assegura una comprensió profunda dels factors d'èxit en el mercat borsari espanyol, proporcionant una ferramenta útil per a inversors, analistes i gestors empresarials.

Paraules clau: anàlisi de dades, dades financeres, ferramentes de visualització, IBEX 35

Abstract

The interest in this topic comes from the importance of the IBEX 35 as the main indicator of the economic and financial health of the Spanish stock market, providing a snapshot of the situation of the Spanish economy, influencing economic, political and investment decisions at national and international level.

The aim of this work is to analyze the variables that influence the performance of companies listed on the IBEX 35 over the last 10 years. This study focuses on compiling and examining content related to operating figures, financial data, surveys and external valuations of the target companies, as well as possible external factors that are equally influential, using data analysis and visualization tools.

The work is structured in several phases, starting with the creation of a Dataset, followed by the descriptive and predictive analysis of the data, and culminating with the visualization and interpretation of the data obtained. This integrated approach ensures a deep understanding of the success factors in the Spanish stock market, providing a useful tool for investors, analysts and business managers.

Keywords: data analytics, financial data, visualization tools, IBEX 35

Índice de contenido

1.	Introducción	12
1.1.	Motivación	12
1.2.	Objetivos	13
1.3.	Metodología	13
1.4.	Estructura	13
2.	Estado del arte	14
2.1.	Análisis de datos sobre mercados bursátiles	14
2.2.	Análisis predictivos sobre el ROE.....	16
2.3.	Estudios sobre el IBEX 35	17
2.4.	Análisis crítico.....	17
3.	Herramientas utilizadas	19
3.1.	Base de datos	19
3.1.1.	Excel	19
3.2.	Lenguaje de programación	19
3.2.1.	Python.....	20
3.2.2.	Librerías de Python.....	20
3.3.	Plataforma	21
3.3.1.	Jupyter Notebook.....	21
3.4.	Herramientas de análisis.....	21
3.4.1.	Análisis descriptivo	21
3.4.2.	Análisis predictivo	23
4.	Diseño de la solución	26
4.1.	Elaboración del Dataset.....	26
4.1.1.	Recopilación de datos	26
4.1.2.	Selección de las empresas.....	29
4.1.3.	Variables identificativas	30
4.1.4.	Variables económicas	31
4.1.5.	Variables exógenas	39
5.	Desarrollo de la solución.....	43
5.1.	Carga del Dataset	43
5.2.	Visualización inicial.....	44
5.3.	Limpieza de datos.....	45
5.3.1.	Eliminación de filas	45
5.3.2.	Eliminación de columnas.....	46



5.3.3.	Valores nulos	46
5.3.4.	Transformaciones	47
5.4.	Vista minable.....	49
5.5.	Análisis descriptivo	50
5.5.1.	Distribuciones de variables numéricas	50
5.5.2.	Distribuciones de variables categóricas	54
5.5.3.	Análisis de correlación.....	57
5.5.4.	Análisis de series temporales	58
5.5.5.	Análisis de valores anómalos.....	63
5.6.	Preprocesado	67
5.6.1.	Normalización.....	67
5.6.2.	Codificación binaria.....	67
5.6.3.	Codificación One-Hot.....	68
5.6.4.	Manejo de valores anómalos.....	68
5.7.	Modelado.....	70
5.7.1.	Selección de variables de los diferentes Dataset.....	70
5.8.	Evaluación.....	75
5.8.1.	Regresión Lineal	76
5.8.2.	Random Forest Regressor	78
5.8.3.	Modelo elegido	80
6.	Análisis de resultados.....	81
7.	Conclusiones	83
7.1.	Ampliaciones y estudios futuros	84
	Bibliografía	85
	Anexos.....	87
	Anexo A – Análisis	87
	Anexo B – Objetivos de Desarrollo Sostenible.....	88

Índice de tablas y figuras

Tabla 1. Páginas web de todas las empresas utilizadas en el análisis.	28
Tabla 2. Composición del IBEX entre 2014 y 2023 con variaciones.	30
Figura 1. Carga de datos	43
Figura 2. Importación de librerías de Python.....	43
Figura 3. Visualización inicial del Dataset.....	44
Figura 4. Visualización estructura inicial Dataset.....	45
Figura 5. Eliminación de filas	46
Figura 6. Eliminación de columnas.....	46
Figura 7. Valores anómalos Ratio Liquidez.....	47
Figura 8. Transformación sectores	47
Figura 9. Visualización final Dataset.....	49
Figura 10. Distribución Ratio Endeudamiento.....	50
Figura 11. Distribución Margen.....	51
Figura 12. Distribución Efecto Fiscal	52
Figura 13. Distribución Presión Fiscal.....	52
Figura 14. Distribución ROA.....	53
Figura 15. Distribución ROE	53
Figura 16. Distribución Dividendo	54
Figura 17. Distribución Beneficio a Reservas.....	54
Figura 18. Distribución Gasto Personal	55
Figura 19. Distribución Estructura de Costes.....	55
Figura 20. Distribución Ciclo de Vida	56
Figura 21. Distribución Sector según ROE.....	56
Figura 22. Análisis correlación ROE	58
Figura 23. Evolución temporal ROA	59
Figura 24. Evolución temporal ROA por sectores	59
Figura 25. Evolución temporal ROE.....	60
Figura 26. Evolución temporal ROE por sectores.....	61
Figura 27. Evolución temporal Margen y Rotación de activos.....	61
Figura 28. Evolución temporal Margen y Rotación de activos por sectores.....	62
Figura 29. Evolución temporal Ratio endeudamiento.....	62
Figura 30. Evolución temporal Ratio endeudamiento por sectores	63
Figura 31. Valores anómalos ROE.....	63
Figura 32. Valores anómalos ROA	64
Figura 33. Valores anómalos ROA operativa	64
Figura 34. Valores anómalos Rotación activos.....	64
Figura 35. Valores anómalos Margen Operativo	64
Figura 36. Valores anómalos Margen	65
Figura 37. Valores anómalos Ratio endeudamiento.....	65
Figura 38. Valores anómalos Ratio liquidez	65
Figura 39. Valores anómalos Multiplicador capital	65
Figura 40. Valores anómalos Efecto financiero	66
Figura 41. Valores anómalos Apalancamiento	66
Figura 42. Valores anómalos Variación precio de las acciones	66
Figura 43. Librerías Python para el preprocesado.....	67



Figura 44. Normalización de las variables numéricas.....	67
Figura 45. Codificación binaria de las variables categóricas	67
Figura 46. Codificación One-Hot de las variables categóricas	68
Figura 47. Método de identificación de valores anómalos.....	68
Figura 48. Método de eliminación de valores anómalos.....	69
Figura 49. Método de sustitución por la media de valores anómalos	69
Figura 50. Método de imputación con IterativeImputer de valores anómalos	69
Figura 51. Selección de variables para modelado	70
Figura 52. Variables seleccionadas Dataset original.....	71
Figura 53. Variables seleccionadas Dataset sin anómalos	72
Figura 54. Variables seleccionadas Dataset con imputación con la media	73
Figura 55. Variables seleccionadas Dataset con imputación con modelo iterativo	74
Figura 56. Evaluación de los modelos	75
Figura 57. Resultados Regresión Lineal	76
Figura 58. Distribución de resultados Regresión Lineal	77
Figura 59. Mejores modelos Regresión Lineal	77
Figura 60. Resultados Random Forest Regressor	78
Figura 61. Distribución de resultados Random Forest Regressor	79
Figura 62. Mejores modelos Random Forest Regressor	79
Figura 63. Objetivos de Desarrollo Sostenible relacionados	88

1. Introducción

En los últimos años, el IBEX 35 se ha consolidado como un referente bursátil tanto a nivel nacional como internacional, siendo además un indicador clave de la salud económica de España. Este índice agrupa a las 35 empresas más grandes que cotizan en la Bolsa de Madrid y ofrece una visión panorámica del funcionamiento de los sectores más importantes de la economía nacional. A lo largo de los últimos diez años, el IBEX 35 ha sido fundamental para evaluar la evolución del tejido empresarial en España, especialmente en tiempos de crisis e incertidumbre económica.

En un contexto marcado por desafíos recurrentes, donde la volatilidad de los mercados, la imprevisibilidad o la globalización están a la orden del día, comprender los factores que impulsan el éxito de las empresas del IBEX 35 es clave. Estos factores no solo han sido fundamentales para que las empresas se mantengan competitivas, sino que también han influido en la capacidad de estas para superar situaciones adversas, como la reciente crisis ocasionada por la pandemia del COVID-19.

En este trabajo de fin de grado se centrará en el análisis de las variables que han influido en la rentabilidad y el éxito de las empresas durante la última década. Se estudiará cómo diferentes factores económicos, financieros y exógenos han influenciado el desempeño de las empresas. Se usarán herramientas de análisis de datos y modelos predictivos, implementados con Python, para identificar patrones y tendencias relevantes que puedan ser útiles para los diferentes Stakeholders.

1.1. Motivación

Tras la realización de la asignatura de SIE (Sistemas de Información Estratégicos), en la que se trataron dos conceptos que para mí eran nuevos en aquel entonces, fueron el análisis y la minería de datos. Supe que quería orientar tanto mi trabajo de fin de grado como mi futuro laboral hacia este sector concretamente. Por ello, no dudé en contactar con Fernando, quien fue profesor tanto de prácticas como de teoría de la asignatura mencionada para así plantearle la posibilidad de realizar el trabajo de fin de grado.

Tras unas semanas de investigación sobre los posibles temas a los que poder aplicar los conocimientos aprendidos a lo largo de la asignatura, le presenté a Fernando la opción de realizar un análisis histórico del IBEX 35, ya que hace de nexo entre mi nueva pasión por el análisis de datos y mi conocimiento sobre inversión y finanzas.

Desde un punto de vista meramente académico, creí que era el momento oportuno para realizar un estudio de qué era lo que más influía en las empresas de nuestro país a la hora de obtener resultados y rentabilidad para sus accionistas. Tras múltiples crisis y situaciones de inestabilidad a nivel europeo y mundial, vi necesario analizar qué aspectos han sido más afectados por estos desafíos y, al mismo tiempo, determinar cuáles han ayudado a superarlos. Este enfoque busca comprender mejor cómo las empresas cotizadas son capaces de desenvolverse en un entorno económico cambiante y desafiante.

Por todo lo anteriormente mencionado, he decidido enfocar mi futuro profesional hacia esta rama comenzando por la realización de un master en Big Data con la finalidad de dedicarme íntegramente a este sector en un futuro próximo.

1.2. Objetivos

Este trabajo tiene como objetivo analizar las variables influyentes en el éxito de las empresas cotizadas en el IBEX 35 durante los últimos diez años. Los objetivos específicos que derivan del estudio son:

- Identificar las variables económicas, financieras y exógenas que afectan a la rentabilidad de las empresas del IBEX 35.
- Evaluar y medir cómo los diferentes sectores del IBEX 35 han respondido ante situaciones de crisis e incertidumbre, como la pandemia del COVID-19.
- Desarrollar modelos predictivos que permitan estimar la rentabilidad de las empresas.

1.3. Metodología

El primer paso que se llevó a cabo al comenzar el proyecto fue realizar una búsqueda exhaustiva de estudios similares.

Posteriormente, se elaboró la recopilación de los datos históricos del IBEX 35 para la confección del Dataset en el que se incluyen variables económicas, financieras y exógenas. Después de esto, se realizó una limpieza preliminar de los datos con el propósito de prepararlos para el análisis. Una vez los datos estuvieron listos, se pasó a la fase del análisis descriptivo utilizando las herramientas de Python mencionadas en la sección 3.

Para finalizar, una vez conocíamos en profundidad los datos a partir del análisis descriptivo, realizamos un preprocesado con el objetivo de maximizar el desempeño de los modelos predictivos. Finalmente, se implementaron modelos de regresión lineal y árboles de decisión para desarrollar las distintas predicciones sobre la rentabilidad de las empresas.

1.4. Estructura

El trabajo de fin de grado se estructura de la siguiente manera:

- **Estado del arte:** Se realizará una revisión de la literatura existente en relación con el análisis de datos en los mercados bursátiles y los métodos predictivos en finanzas, así como las herramientas empleadas para los mismos. Posteriormente, se llevará a cabo un análisis crítico de esta literatura posicionando las bondades de este estudio.
- **Herramientas utilizadas:** Se describirán las herramientas y métodos utilizados en el estudio, incluyendo justificaciones para su sección.
- **Diseño de la solución:** Este capítulo se basará en la metodología y las decisiones tomadas para la elaboración del *Dataset* compuesto por datos de las empresas del IBEX 35 de los últimos diez años.
- **Desarrollo de la solución:** A lo largo de esta sección, se realizará la limpieza inicial de los datos para su posterior análisis descriptivo, seguido de un preprocesado que preparará el Dataset para obtener un mejor desempeño en el análisis predictivo realizado al final de la sección.
- **Análisis de resultados:** Se presentarán las conclusiones del estudio, destacando las implicaciones de los hallazgos para los inversores.
- **Ampliaciones:** Finalmente se expondrán posibles ampliaciones al trabajo realizado.



2. Estado del arte

En este segundo capítulo, se presenta una revisión de los estudios e investigaciones existentes relacionados con el análisis de los mercados bursátiles y los métodos predictivos empleados en finanzas. Poniendo el foco en aquellos que tengan relación con el IBEX 35 y predicciones sobre la rentabilidad. El objetivo de esto es posicionar este trabajo de fin de grado dentro del contexto académico y profesional actual, destacando sus aportaciones y bondades en comparación con los estudios realizados previamente. También realizaremos un análisis de las tecnologías y metodologías empleadas en estos análisis, finalizando con un diagnóstico crítico que tiene como meta identificar las nuevas perspectivas y mejoras que este trabajo ofrece frente a lo que hay actualmente en el mercado.

2.1. Análisis de datos sobre mercados bursátiles

En primer lugar, evaluaremos aquellos estudios en los que se hayan empleado métodos de análisis de datos similares a los que hemos utilizado en este trabajo aplicado a los mercados bursátiles. Entre los más relevantes podemos encontrar el artículo de Cherif, M. & Gazdar, K. [1], el cual: Investiga los factores que impulsan el desarrollo del mercado de valores en la región MENA (Oriente Medio y el Norte de África) en el periodo 1990-2007.

Para ello, examina la influencia tanto del entorno macroeconómico (nivel de ingreso, tasa de ahorro, tasa de interés...) como de la calidad institucional (a través de un índice de riesgo político) en el desarrollo del mercado de valores de la región. Parte de la premisa que un entorno favorable y unas instituciones fuertes fomentan el crecimiento.

Hace uso de un modelo de regresión en el que el MCAP (Capitalización de Mercado) es la variable dependiente del modelo, los diferentes factores macroeconómicos e institucionales actúan como variables predictivas.

Tras obtener una explicabilidad del 79% en el modelo de regresión, llega a las conclusiones de que: los factores macroeconómicos influyen significativamente en el desarrollo del mercado de valores según lo esperado en la teoría económica y, que los sectores bursátiles y bancarios funcionan como complementarios. Por otro lado, el índice de riesgo político no parece ser un factor relevante.

En segundo lugar, hallamos el artículo de Ou, P. & Wang, H. [2] que aplica y compara diez técnicas de minería de datos con el objetivo pronosticar si el índice de Hang Seng (Hong Kong) subirá o bajará en un día determinado.

Para ello, emplea valores predictivos como: el precio de apertura del índice, el precio más alto y bajo en el periodo anterior al que se quiere predecir, el precio de cierre del mercado americano S&P 500 (compuesto por las 500 mejores empresas del Estados Unidos) y el tipo de cambio entre dólar de Hong Kong y dólar estadounidense.

Las técnicas predictivas empleados son:

1. Análisis discriminante lineal (LDA)
2. Análisis discriminante cuadrático (QDA)
3. Clasificación de k vecinos más cercanos
4. Naïve Bayes
5. Modelo Logit
6. Clasificación basada en árboles
7. Red neuronal
8. Clasificación bayesiana con procesos gaussianos
9. Máquina de vectores de soporte (SVM)
10. Máquina de vectores de soporte por mínimos cuadrados (LS-SVM)

El estudio utiliza datos diarios de precio de cierre del índice junto con los cinco predictores mencionados anteriormente. El periodo estudiado va desde el 3 de enero del año 2000 hasta el 29 de diciembre de 2006 con un total de 1732 días, destinando los primeros 1482 días para el entrenamiento del modelo y los últimos 250 correspondientes al año 2006 para el test.

Los resultados muestran que todas las técnicas empleadas generan una alta tasa de aciertos superior al 80%, donde las máquinas de vectores superan al resto de modelos en términos de capacidad predictiva.

Un último estudio de Maini, S. S. & Govinda, K [3] utiliza técnicas de aprendizaje automático, en concreto Random Forest y SVM, para anticipar si el precio de una acción aumentará o no en el futuro, basándose en datos históricos y análisis de noticias.

El conjunto de datos utilizados abarca el periodo de 2000 a 2016 y se utiliza para predecir el índice americano Dow Jones Industrial Average. La información fue recopilada de:

- Noticias: Titulares históricos del canal de Reddit World News y la API (Application Programming Interface) de noticias del periódico The Guardian
- Datos bursátiles: Datos del Dow Jones Industrial Average de *Yahoo! Finance*.
- Etiquetas: Indican si los titulares de noticias condujeron a un aumento (1) o una disminución/estancamiento (0) en el precio de las acciones al día siguiente.

En cuanto a los resultados, ambos modelos mostraron un buen rendimiento en la predicción del índice, mostrando precisiones superiores al 86% en el caso del Random Forest y al 85% en el caso del SVM.

Podemos ver como la tónica general en cuanto a los estudios sobre índices bursátiles es la predicción del desempeño de este en términos globales, ya sea en base a modelos regresivos o de clasificación.



2.2. Análisis predictivos sobre el ROE¹

Una vez visto cuales son los estudios más relevantes en términos de mercados bursátiles, pasemos a vislumbrar de manera concreta los estudios sobre la predicción del ROE. Para ello, estudiaremos el siguiente artículo de Yang, Y. & Wang, Z. [4] el cual: Presenta un modelo de aprendizaje diseñado para predecir la rentabilidad sobre el capital (ROE) en empresas del sector energético chino. Este estudio asume como premisa que la estructura de capital de una empresa impacta de manera significativa en su desempeño futuro.

Para ello, utiliza el algoritmo XGBoost, que tiene la capacidad de combinar varios predictores simples para construir un predictor más potente y preciso. La información empelada proviene de la base de datos de Guotai'an y contiene hasta 12072 indicadores de hasta 1084 empresas energéticas chinas entre 2003 y 2021. Entre estos indicadores se encuentran, características del capital, rentabilidad, situación financiera, solvencia o la cantidad de activos totales.

La parte de preprocesado es clave dentro del estudio incluyendo manejo de valores faltantes, reducción de la dimensionalidad mediante análisis de componentes principales o el escalado de datos.

Los resultados son prometedores, se obtiene un coeficiente de determinación (R^2) de 0.9460, hecho que implica que este modelo explica una proporción significativa de la variabilidad en el ROE.

Por otro lado, un estudio de Balci, T. & Ogul, H. [5] se basa en el análisis del desempeño de los bancos estatales de depósito en Turquía, centrándose en la predicción del ROE.

Para lograr esto, recurren a redes neuronales artificiales, un fuerte enfoque de inteligencia artificial. En este caso, las redes fueron entrenadas con datos trimestrales del periodo 2005-2015, incluyendo tanto variables exógenas como el tipo de cambio o la tasa de inflación y variables financieras como los depósitos totales o patrimonio.

Los resultados obtenidos revelan que se obtuvo un rendimiento superior al 80% en la predicción del ROE con todos los modelos.

Podemos concluir que, en lo que respecta a la predicción de la rentabilidad, hay una tendencia generalizada a enfocarse en sectores determinados de la economía de un país, así como emplear múltiples herramientas predictivas para obtener los mejores resultados.

¹ ROE (Return On Equity)

2.3. Estudios sobre el IBEX 35

Finalmente, como estudio relacionado con el IBEX 35 podemos encontrar el trabajo de Consoli, S., Negri, M., Tebbifakhr, A., Tosetti, E. & Turchi, M. [6] que explora el uso de las emociones extraídas de las noticias económicas y financieras para mejorar la precisión de los modelos de series temporales. El estudio se centra únicamente en el IBEX 35 y utiliza técnicas de *Deep Learning* para extraer el contenido emocional de las noticias.

El proceso implica la clasificación de millones de artículos de noticias en función de las emociones básicas de Ekman (miedo, ira, alegría, tristeza y sorpresa), mediante la adaptación de un sistema de traducción neuronal entrenado en paralelo tanto inglés como en español. Esto permite tener clasificadas una cantidad masiva de noticias con etiquetas emocionales.

Además, se emplea DeepAr, un método de pronóstico basado en redes neuronales, para predecir las fluctuaciones diarias del IBEX 35. Esta herramienta da estimaciones probabilísticas de los movimientos del índice facilitando así la gestión de la incertidumbre.

Los resultados finales muestran como los indicadores emocionales mejoran el rendimiento de predicción de los cambios diarios del IBEX 35, confirmando, por ejemplo, cómo las emociones negativas tienen un mayor impacto en el modelo.

2.4. Análisis crítico

Por último, vamos a realizar un análisis crítico de los estudios encontrados destacando las aportaciones relevantes.

En cuanto a los estudios sobre los mercados bursátiles, podemos observar como la gran mayoría de los artículos existentes tienen un mismo enfoque: la predicción de la dirección del índice. Es decir, se centran en una visión macro del índice, indicando el comportamiento de las empresas en conjunto. Esta visión es susceptible a varias críticas, en primer lugar, la pérdida de información de las empresas. Al dar un enfoque global, y estar sumando o promediando el total de los desempeños, aceptamos malos resultados por parte de algunas empresas y excelentes por partes de otras. Esto no llega a ser del todo eficiente si el objetivo es que la información obtenida sea útil para los inversores. En segundo lugar, este tipo de enfoques generalistas limita la inversión al índice en su conjunto, ya que no se conocen de manera desagregada los desempeños de los sectores o las empresas. Por nuestra parte, somos capaces de solucionar ambos problemas a través del enfoque micro establecido en este trabajo, donde las empresas son el objeto final de análisis, proporcionando información del desempeño de cada una de ellas de manera histórica y predictiva. Además, centrándonos en los inversores, también se realizan estudios con un nivel de agregación superior (sectores), dando así un amplio abanico de información para que estos y/o interesados puedan tomar decisiones de mayor calidad.

En cuanto a las predicciones del ROE existentes en la literatura actual que tengan un enfoque similar al utilizado en este trabajo podemos sacar una conclusión clara. El principal camino por el que optan estos estudios es por el análisis predictivo enfocado en un único sector de una economía determinada. Con esto, se está viendo el funcionamiento de una única parte de la economía, obviando, por ejemplo, la interacción entre los diferentes sectores y cómo esto afecta al desempeño. En nuestro caso, al realizar el estudio para todo el índice, damos un enfoque más holístico, permitiéndole al inversor o analista realizar juicios propios sobre qué decisiones tomar en base a la información de todos los sectores, así como ver la influencia de algunos de ellos en mayor o menor medida en la predicción del ROE.

Por último, hablando exclusivamente de los trabajos relacionados con el IBEX 35, observamos que, en términos generales, existe una falta de análisis de datos tanto descriptivo como predictivo. La literatura disponible no solo tiene un enfoque diferente, como es el caso de las emociones en el ámbito bursátil, si no que vuelven a tener un enfoque generalista con respecto a si el índice aumenta o disminuye. Respecto a nuestra incumbencia, traemos un análisis descriptivo histórico el cual no se había realizado previamente, pudiendo entender de esta manera cuál ha sido la situación de la económica española a nivel micro como macro en los últimos años. Además, el enfoque predictivo que ofrecemos es innovador, ya que trata de manera particular las rentabilidades de las empresas, proporcionando muchas más herramientas para la toma de decisiones.

En conclusión, este trabajo aporta una perspectiva original al análisis del mercado bursátil español, ofreciendo recursos que contribuyan con gran utilidad al proceso de decisión.

3. Herramientas utilizadas

En este apartado se hablará sobre las herramientas que han sido utilizadas para realizar el análisis. Comenzaremos explicando cuál ha sido la herramienta utilizada para la recolección, gestión de información y posterior generación del *Dataset*. A continuación, nos centraremos en el procesamiento de datos y en la parte estrictamente analítica, mostrando cuál ha sido el lenguaje de programación empleado, así como las librerías utilizadas, junto con el entorno en el que se ha realizado dicho análisis. Por último, explicaremos cuáles han sido las técnicas descriptivas y predictivas empleadas para la obtención de resultados. En la sección 4, se proporcionará una guía detallada y estructurada para llevar a cabo el análisis y la visualización, usando las herramientas aquí descritas.

3.1. Base de datos

En esta primera sección se expondrá cuál ha sido la herramienta utilizada para la creación y gestión del *Dataset*, así como las razones detrás de la elección.

3.1.1. Excel

En este estudio, hemos optado por utilizar Excel como herramienta principal para la creación y gestión del *Dataset* formado por los datos de las empresas del IBEX 35 durante los últimos 10 años. Microsoft Excel², desarrollada por Microsoft como parte del paquete Microsoft Office, es una herramienta muy conocida y utilizada tanto en el ámbito profesional como en el académico, hecho que facilita su adopción sin necesidad de una formación especializada previa.

Otra de las razones, es su integración y compatibilidad con Python, la herramienta de análisis de datos empleada. Esto facilita una transferencia de datos adecuada para la realización de los diferentes análisis.

Debido a la cantidad moderada de datos que manejamos, la fácil accesibilidad y compatibilidad con el entorno de trabajo, y sobre todo, gracias a la replicabilidad, Excel es la herramienta elegida.

3.2. Lenguaje de programación

En esta sección se expondrá cuál ha sido el lenguaje de programación utilizado a lo largo de todo el proceso de análisis, desde la limpieza de datos hasta la elaboración de modelos predictivos, pasando por el análisis descriptivo y el preprocesado. Daremos razones sobre la elección del lenguaje, así como sobre las librerías empleadas.

² <https://www.microsoft.com/es-es/microsoft-365/excel/>



3.2.1. Python

Python³ es un lenguaje de programación de alto nivel y orientado a objetos, creado a finales de los años ochenta por Guido van Rossum. Es conocido por su fácil accesibilidad y legibilidad, lo que facilita la escritura y el mantenimiento del código. Python es ampliamente conocido entre la comunidad de los datos debido a su amplia gama de librerías, lo que lo convierte en una herramienta ideal para emplear en todas y cada una de las fases del análisis de datos.

La elección de Python como lenguaje de programación principal se debe a varias razones. Primeramente, a la accesibilidad de este, pudiendo alcanzar resultados competentes en un corto periodo de tiempo. En segundo lugar, la versatilidad, pudiéndose utilizar para diferentes tareas a lo largo de todo el proceso. Por último, la comunidad de Python es muy activa, proporcionando infinidad de documentación, tutoriales y soporte a las diferentes casuísticas que surgen.

Por todo ello, al no tener experiencia previa en la utilización de lenguajes de programación orientados a análisis de datos, se ha considerado preferible dedicar tiempo al aprendizaje de Python, pudiendo ser útil en futuros trabajos fuera de este ámbito.

3.2.2. Librerías de Python

Como se ha mencionado anteriormente, una de las principales ventajas de utilizar Python como lenguaje de programación es la cantidad de librerías que dispone. Aquellas que se han utilizado a lo largo del estudio son:

- **Pandas**⁴: Una librería usada para la carga, manipulación y tratamiento del Dataset, tiene una función estructural.
- **NumPy**⁵: Esta librería proporciona un soporte para *arrays* y matrices, junto con una serie de funciones matemáticas.
- **Matplotlib**⁶: Una librería usada para la generación de diferentes gráficos empleados en el análisis.
- **Seaborn**⁷: Construida a partir de Matplotlib. En conjunto se utilizan para ofrecer una amplia gama de posibilidades de visualización.
- **Missingno**⁸: Herramienta utilizada fundamentalmente para la identificación de valores faltantes para su posterior tratamiento.
- **Scipy**⁹: Proporciona algoritmos y herramientas estadísticas avanzadas.
- **Statsmodels**¹⁰: Una librería que ofrece clases y funciones para la estimación de diferentes modelos estadísticos.
- **Scikit-learn**¹¹: Proporciona herramientas tanto para el preprocesado de datos así como una gran cantidad de modelos y algoritmos.

³ <https://www.python.org/>

⁴ <https://pandas.pydata.org/>

⁵ <https://numpy.org/>

⁶ <https://matplotlib.org/>

⁷ <https://seaborn.pydata.org/>

⁸ <https://github.com/ResidentMario/missingno>

⁹ <https://scipy.org/>

¹⁰ <https://www.statsmodels.org/>

¹¹ <https://scikit-learn.org/>

3.3. Plataforma

En esta sección explicaremos la plataforma utilizada para desarrollar, documentar y ejecutar el análisis de datos y el modelo predictivo.

3.3.1. Jupyter Notebook

Jupyter Notebook¹² es una aplicación de código abierto extremadamente valorada dentro del ámbito del análisis de datos y desarrollo de modelos de aprendizaje, ya que facilita la integración de diferentes componentes dentro de un solo documento.

La elección de Jupyter Notebook como plataforma para este TFG se basa en varios puntos principales. Primeramente, permite la ejecución de código de manera interactiva, permitiendo a los usuarios ejecutarlo individualmente, observar los resultados de manera inmediata y ajustar el análisis según sea necesario.

Otra de las principales ventajas de Jupyter Notebook, es la capacidad de integrar texto, código y visualizaciones en un único entorno. Esto nos resulta esencial a la hora de documentar el proceso de análisis llevado a cabo a lo largo del proyecto, ya que permite añadir explicaciones junto al código, facilitando de esta manera la replicabilidad y la comprensión del proceso [7].

Por último, la comunidad de Jupyter Notebook y su compatibilidad con diversas plataformas, tanto en local, como en la nube, como Google Colab, facilita el acceso a recursos de computación adicionales, pensando en una posible escalabilidad. Todas estas características la convierten en una herramienta fundamental para el análisis de datos, el desarrollo y documentación de este proyecto.

3.4. Herramientas de análisis

En esta última sección vamos a describir las herramientas que se han utilizado para examinar los datos de las empresas del IBEX 35 y extraer conclusiones relevantes. El análisis se divide en dos partes principales: análisis descriptivo y predictivo mencionado en la sección anterior.

3.4.1. Análisis descriptivo

El análisis descriptivo [8] es una fase fundamental del análisis estadístico de datos, también conocido como análisis descriptivo de datos o estadísticas descriptivas. Su propósito es proporcionar una comprensión clara y concisa sobre la distribución y las principales características de un conjunto de datos. Este tipo de análisis ayuda a los investigadores a identificar excepciones y errores, comprender la relación entre variables y fundamentar investigaciones estadísticas futuras. Transforma los datos brutos en un formato que facilita su interpretación y análisis al reorganizar, clasificar y modificar los datos para extraer información útil. Esta fase crucial en el análisis de datos permite detectar patrones, relaciones y tendencias dentro de los datos, preparándolos para estudios más avanzados.

¹² <https://jupyter.org/>

Existen diversas metodologías empíricas que apoyan la práctica del análisis descriptivo. Las herramientas descriptivas más populares son estadísticas simples que representan tendencias centrales y variaciones (como medias, medianas y modas), útiles para explicar los datos. Estas metodologías pueden incluir medidas de frecuencia, que proporcionan un recuento o porcentaje sobre la periodicidad con que un evento ocurre; medidas de tendencia central, como la media, mediana y moda usados para calcular el valor promedio; medidas de dispersión, como el rango o la desviación estándar, que miden cómo se distribuyen los datos a lo largo de un rango; y medidas de posición, como cuartiles y percentiles, que identifican la posición de un valor en relación con otros datos.

Realizar un análisis descriptivo implica varias fases críticas. La primera es la recolección de datos, que conlleva identificar fuentes de datos, seleccionar métodos apropiados de recolección y verificar que los datos representen con precisión la población o el tema de interés. La segunda fase es la preparación de datos, que asegura un conjunto de datos limpio, consistente y listo para el análisis. Este paso incluye tareas como la limpieza de datos, transformación de datos a un formato adecuado y reducción de datos si el conjunto es demasiado grande.

Los métodos aplicados en la tercera fase abarcan un análisis de distribución de frecuencia, creando tablas de frecuencia o gráficos de barras; medidas de tendencia central, calculando la media, mediana y moda; y medidas de dispersión, calculando el rango, varianza y desviación estándar. También se incluyen medidas de posición, identificando la posición de los valores respecto a otros. El análisis descriptivo emplea varias técnicas y estadísticas para proporcionar una visión general de los datos, lo que facilita la identificación de patrones y relaciones importantes.

Por último, el análisis descriptivo culmina con la presentación de estadísticas resumidas y visualizaciones. Estas visualizaciones pueden incluir histogramas, diagramas de caja, gráficos de dispersión o de líneas para datos numéricos y gráficos de barras o de sectores para datos categóricos. Los gráficos que hemos utilizado para nuestro análisis son los siguientes:

- **Gráficos de barras:** Estos gráficos nos han permitido comparar diferentes categorías o grupos dentro del conjunto de datos así, pudiendo identificar de esta manera distribuciones y frecuencias.
- **Gráficos de dispersión:** Utilizados para visualizar la relación entre dos variables numéricas, ayudando de esta manera a determinar relaciones o posibles patrones.
- **Gráficos lineales:** Esta herramienta nos ha permitido mostrar la evolución de varias series temporales a lo largo del tiempo, señalando tendencias o acontecimientos clave.
- **Gráficos caja y bigote:** Su uso principal ha sido la detección de valores anómalos en las variables numéricas para su posterior tratamiento.

En definitiva, resumir los datos de manera clara y concisa, utilizando visualizaciones efectivas, ayuda a comunicar los hallazgos a audiencias no técnicas, lo que facilita la toma de decisiones basada en datos.

3.4.2. Análisis predictivo

El análisis de datos (o aprendizaje) predictivo es una rama del aprendizaje automático que se centra en la construcción de modelos capaces de hacer predicciones sobre datos futuros basados en patrones identificados en datos históricos. Este proceso involucra el uso de algoritmos y técnicas estadísticas para analizar conjuntos de datos y descubrir relaciones entre variables que pueden ser utilizadas para anticipar resultados desconocidos. El objetivo principal del aprendizaje predictivo es proporcionar una base sólida para la toma de decisiones basada en datos y mejorar la precisión de los pronósticos.

A continuación, explicaremos dos modelos ampliamente utilizados en la literatura y que usaremos en nuestro proyecto: la regresión lineal, como *baseline*, y los bosques aleatorios (Random Forests), como una opción más avanzada y potente, destacando sus características, ventajas y aplicaciones específicas.

REGRESIÓN LINEAL

La regresión lineal [9] es una técnica de análisis de datos utilizada para predecir valores inciertos mediante el uso de otros valores de datos relacionados y conocidos. Esta técnica modela matemáticamente la relación entre una variable dependiente y una independiente mediante una ecuación lineal. Por ejemplo, si se tienen datos sobre gastos e ingresos del año pasado, la regresión lineal puede determinar que estos gastos son la mitad de los ingresos y predecir futuros gastos al reducir a la mitad un ingreso conocido futuro.

Los modelos de regresión lineal son relativamente simples y proporcionan una fórmula matemática fácil de interpretar para generar predicciones. Esta técnica estadística está establecida y se puede aplicar fácilmente al *software* y a la computación. Las empresas la utilizan para convertir datos sin procesar en inteligencia empresarial y conocimiento práctico. Además, los científicos de diversas disciplinas, incluyendo biología y ciencias sociales, utilizan la regresión lineal para realizar análisis de datos preliminares y predecir tendencias futuras. Métodos avanzados, como el *machine learning* y la inteligencia artificial, también emplean esto para resolver problemas complejos.

Una regresión lineal simple trata de trazar un gráfico lineal entre dos variables de datos, x e y . La variable independiente x se representa en el eje horizontal y la variable dependiente y se representa en el eje vertical. La forma más simple de la ecuación es $y = c*x + m$, donde c y m son constantes. Por ejemplo, si los datos de entrada son (1,5), (2,8) y (3,11), se identifica la ecuación de regresión lineal como $y = 3*x + 2$ y se puede extrapolar que y es 14 cuando x es 4.

En el ámbito del *machine learning* [10], los algoritmos analizan grandes conjuntos de datos y calculan la ecuación de regresión lineal a partir de esos datos. Estos algoritmos, entrenados en conjuntos de datos conocidos o etiquetados, se utilizan para predecir valores desconocidos. Los datos de la vida real son más complicados, por lo que la regresión lineal debe transformar matemáticamente los valores para cumplir con cuatro supuestos: (1) relación lineal, donde las variables independientes y dependientes tienen una relación lineal aparente; (2) independencia residual, donde los residuos no muestran patrones identificables; (3) normalidad, donde los residuos deben estar distribuidos normalmente, verificado mediante gráficas Q-Q; y (4) homocedasticidad, donde los residuos deben tener una variación constante a lo largo de los valores de x .



Existen varios tipos de regresión lineal, dependiendo de la complejidad del conjunto de datos y la relación entre las variables. La regresión lineal simple utiliza una función lineal con una única variable independiente: $Y = \beta_0 * X + \beta_1 + \varepsilon$, donde β_0 y β_1 son constantes y ε es el término de error. Esta puede modelar relaciones como la lluvia y el rendimiento de cultivos, o la temperatura y la expansión del mercurio en un termómetro. Por otro lado, la regresión lineal múltiple involucra una variable dependiente y múltiples variables independientes, con la ecuación $Y = \beta_0 * x_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_N * x_N + \varepsilon$. Esta modela múltiples variables y su impacto en un resultado, como la lluvia, temperatura y uso de fertilizantes en el rendimiento de cultivos, o el crecimiento salarial y la inflación en las tasas de préstamos hipotecarios.

RANDOM FOREST

Los Random Forest, también conocidos en castellano como Bosques Aleatorios, son una combinación de árboles predictores en la que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno. Esta técnica es una modificación sustancial de *bagging* [11] que construye una gran colección de árboles no correlacionados y luego los promedia para obtener una predicción más robusta y precisa. El algoritmo para inducir un Random Forest fue desarrollado por Leo Breiman y Adele Cutler [12], y el término proviene de la primera propuesta de Random Decision Forests hecha por Tin Kam Ho de Bell Labs en 1995 [13]. Este método combina la idea de *bagging* y la selección aleatoria de atributos para construir una colección de árboles de decisión con variaciones controladas.

La selección de un subconjunto aleatorio de atributos es un ejemplo del método *random subspace*, formulado inicialmente por Ho como una manera de llevar a cabo la discriminación estocástica propuesta por Eugenio Kleinberg. En muchos problemas, el rendimiento de los Random Forests es muy similar al de boosting, siendo más simples de entrenar y ajustar, razones para convertirse en una técnica bastante popular y ampliamente utilizada.

La idea esencial del *bagging* es promediar muchos modelos “ruidosos” (i.e., modelos muy sensibles a las variaciones en los datos de entrada), pero aproximadamente imparciales (i.e., sin sesgos) para reducir la variación. Los árboles de decisión son candidatos ideales para el *bagging* porque pueden capturar estructuras de interacción compleja en los datos y, si crecen lo suficientemente profundos, tienen una baja parcialidad. Debido a que los árboles son notoriamente ruidosos, se benefician enormemente al promediar.

El proceso de construcción de cada árbol en un Random Forest funciona de la siguiente manera: se elige un conjunto de entrenamiento para el árbol y se usan los casos restantes para estimar el error. Para cada nodo del árbol, se utilizan aleatoriamente m variables a partir de las cuales se basa la decisión, y luego se calcula la mejor partición del conjunto de entrenamiento en base a estas m variables. Para la predicción, un nuevo caso se empuja hacia abajo por el árbol y se asigna la etiqueta del nodo terminal. Este proceso se itera para todos los árboles en el ensamblado, y la etiqueta que obtiene la mayor cantidad de incidencias se reporta como la predicción.

Las ventajas de los Random Forests incluyen:

1. Alta precisión en clasificación para conjuntos de datos suficientemente grandes.
2. Eficiencia en el manejo de grandes bases de datos.
3. Capacidad para manejar cientos de variables de entrada sin exclusión.
4. Estimaciones de la importancia de cada variable en la clasificación.
5. Métodos eficaces para estimar datos perdidos y mantener la exactitud cuando una gran proporción de los datos está perdida.
6. Computación de prototipos que ofrecen información sobre la relación entre variables y la clasificación.
7. Computación de proximidades entre pares de casos para uso en agrupación, localización de valores atípicos y visualizaciones interesantes de los datos.
8. Métodos experimentales para detectar interacciones entre las variables.



4. Diseño de la solución

En esta sección detallaremos el proceso de diseño de la solución, centrado en la elaboración del *Dataset* correspondiente a los diez últimos años de las empresas del IBEX 35. Este proceso está formado por varias etapas que incluyen la recopilación de datos, la selección de las empresas de cada año, así como los criterios empleados y la selección de variables identificativas, económicas y exógenas.

4.1. Elaboración del Dataset

Para la elaboración del *Dataset* hemos decidido dar un enfoque histórico, por ello, utilizamos datos que van desde el 31 diciembre de 2013 al 31 de diciembre de 2023. A continuación, explicaremos la forma en la que se han seleccionado las empresas y los criterios tanto para la selección de estas como para la de las variables que componen el *Dataset*.

4.1.1. Recopilación de datos

Empezaremos explicando cómo ha sido el proceso de recopilación de datos, detallando así el lugar de procedencia y la forma de acceso a los mismos. Los datos se han recopilado de manera manual a partir de las páginas web y los documentos presentados a continuación.

DATOS DE EMPRESAS

Para los datos específicos de las empresas, hemos utilizado diversas fuentes del ámbito financiero:

- **Informes anuales y de gestión:** Estos documentos proporcionan una amplia información sobre el desempeño, tanto operativo como financiero de cada una de las empresas, siendo utilizados así para dar forma a la mayoría de los datos del *Dataset*. Para la obtención dichos documentos se ha recurrido a dos sitios principales:
 - **Página web de la Comisión Nacional del Mercado de Valores¹³ (CNMV):** La CNMV recoge los datos y documentación de todas las empresas objeto de estudio, es una fuente oficial y centralizada para obtener documentación e información relacionada con las empresas del país. Se puede acceder en el apartado de “Búsqueda por entidad” y poniendo el nombre de la empresa.
 - **Páginas web de las corporaciones:** Por cuestiones relacionadas principalmente con la forma de los documentos, se recurre a las páginas web oficiales de las corporaciones. En estas páginas, las empresas muestran este tipo de documentación con el fin de facilitar el acceso a los datos a accionistas e interesados. Esto se realizaba en caso de que los documentos presentados a la CNMV sufrieran un cambio de forma posterior, que agilizase la tarea de recopilación de datos. Podemos ver en la tabla 1, todas las páginas web de todas las empresas del estudio.

¹³ <https://www.cnmv.es/>

Empresa	Web
ABERTIS	https://www.abertis.com
ABENGOA B	https://www.abengoa.com
ACS	https://www.grupoacs.com
ACERINOX	https://www.acerinox.com
AENA	https://www.aena.es
ALMIRALL	https://www.almirall.com
AMADEUS	https://www.amadeus.com
ACCIONA	https://www.acciona.com
ACCIONA ENERGIA	https://www.acciona-energia.com
BBVA	https://www.bbva.com
BANKIA	https://www.bankia.com
BANKINTER	https://www.bankinter.com
BOLSAS Y MERCADOS	https://www.bolsasymercados.es
CAIXABANK	https://www.caixabank.com
CIE-AUTOMOTIVE	https://www.cieautomotive.com
CELLNEX	https://www.cellnextelecom.com
INMOB COLONIAL	https://www.inmocolonial.com
DIA	https://www.diacorporate.com
ENDESA	https://www.endesa.com
ENCE ENERGIA	https://www.ence.es
ENAGAS	https://www.enagas.es
FCC	https://www.fcc.es
FLUIDRA	https://www.fluidra.com
FERROVIAL	https://www.ferrovial.com
GRIFOLS	https://www.grifols.com
IAG	https://www.iairgroup.com
IBERDROLA	https://www.iberdrola.com
INDRA	https://www.indracompany.com
INDITEX	https://www.inditex.com
JAZZTEL	https://www.jazztel.com
LOGISTA	https://www.logista.com
MAPFRE	https://www.mapfre.com
MASMOVIL	https://www.masmovil.es
MELIA HOTELS	https://www.melia.com
MERLIN PROPERTIES	https://www.merlinproperties.com
ARCELORMITTAL	https://www.arcelormittal.com
NATURGY	https://www.naturgy.com
OBRASCON HUARTE LAIN	https://www.ohla-group.com
PHARMA MAR	https://www.pharmamar.com
BANCO POPULAR	https://www.bancopopular.es
REDEIA	https://www.redeia.com
REPSOL	https://www.repsol.com

LABORATORIOS ROVI	https://www.rovi.es
BANCO DE SABADELL	https://www.bancsabadell.com
BANCO SANTANDER	https://www.santander.com
SACYR	https://www.sacyr.com
G. DE ALQUILER DE MAQUINARIA	https://www.gamrentals.com
SOLARIA ENERGIA	https://www.solariaenergia.com
TELEFONICA	https://www.telefonica.com
MEDIASET ESPAÑA	https://www.mediaset.es
TECNICAS REUNIDAS	https://www.tecnicasreunidas.es
UNICAJA	https://www.unicajabanco.es
VISCOFAN	https://www.viscofan.com

Tabla 1. Páginas web de todas las empresas utilizadas en el análisis.

- **Bases de datos financieras:** De este tipo de fuente se ha extraído principalmente la variación anual del precio de las acciones de cada una de las empresas. Dado que este tipo de información no lo incluyen todas las empresas dentro de sus informes anuales o lo hacen de una forma no estandarizada, se ha recurrido a las principales bases de datos financieras, tanto nacionales como internacionales utilizadas ampliamente en el sector bursátil. Estas son:
 - Página web de Bolsas y Mercados Españoles¹⁴
 - Página web de Investing¹⁵
 - Página web de PcBolsa¹⁶

Todas estas fuentes han sido de gran ayuda para la recopilación de todos los datos relacionados directamente con cada una de las empresas del IBEX 35 en los últimos diez años.

DATOS EXTERNOS

Además de los datos que tienen que ver directamente con las empresas, hemos recopilado datos externos que pueden llegar a influir en el desempeño de estas. Estos datos incluyen:

- **Datos macroeconómicos:** Este tipo de datos son esenciales para comprender cuál es el contexto económico y social en el que se desarrollan las empresas cada año, que incluyen indicadores como el Producto Interior Bruto (PIB), tasas de interés y variación del desempleo o del Índice de Precios de Consumo (IPC). Estos datos han sido extraídos del Instituto Nacional de Estadística¹⁷ (INE) y el Banco de España¹⁸.

¹⁴ <https://www.bolsasymercados.es/>

¹⁵ <https://es.investing.com/>

¹⁶ <https://pcbolsa.com/>

¹⁷ <https://www.ine.es/>

¹⁸ <https://www.bde.es/>

- **Regulaciones y políticas:** Datos provenientes principalmente del Boletín Oficial del Estado¹⁹ (BOE), del que se recogen documentos referentes a la incorporación de nueva legislación en materia laboral o la convocatoria de nuevas elecciones dentro del país. Ambos sucesos pueden llegar a generar cambios o variaciones en el desempeño o comportamiento de las empresas, por ello, las hemos considerado fundamentales para el análisis. Aquí podemos ver el listado de los decretos usado:
 - Nueva legislación laboral con Real Decreto-ley 13/2022
 - Elecciones 26 de octubre de 2015 con Real Decreto 977/2015
 - Elecciones 3 de mayo de 2016 con Real Decreto 184/2016
 - Elecciones 4 de marzo de 2019 con Real Decreto 129/2019
 - Elecciones 24 de septiembre de 2019 con Real Decreto 551/2019
 - Elecciones 29 de mayo de 2023 con Real Decreto 400/2023
- **Análisis de empresas externas:** Para aportar un valor añadido, también se han utilizado datos externos procedentes de organizaciones internacionales. En este caso, se ha utilizado el informe anual de la Fundación Heritage²⁰ sobre libertad económica. De dicho informe, no solo se ha utilizado el dato global de libertad económica para cada uno de los 10 años, sino que también se han usado alguna de las medidas que se emplean para la ponderación final de dicho índice, como la libertad laboral, libertad de negocios o derechos de propiedad. Todo ha sido extraído de la página web oficial de la fundación Heritage.

4.1.2. Selección de las empresas

Para la selección de las empresas incluidas en el análisis, nos hemos basado en la información recogida en el apartado Histórico de Índices, de Bolsas y Mercados Españoles²¹ (BME), el cual contiene todas las entradas y salidas del IBEX 35 en cada una de sus revisiones semestrales, desde su creación en 1991.

La forma en la que se ha determinado la composición del índice para cada año ha sido empleando un método de eliminación y sustitución. Partiendo de la composición inicial, se pasan a analizar las salidas y entradas de cada año. En la composición del año siguiente, se incluyen las mismas empresas que el año anterior, eliminando aquellas que han sido excluidas del índice así como agregando las nuevas incorporaciones.

Principalmente se ha tenido en cuenta que a fecha 31 de diciembre del año correspondiente, la empresa se encontrase dentro del índice, sin importar si su incorporación había sido hacía justo un año o en los seis meses previos en la segunda revisión realizada por el IBEX 35. Esto se hace para garantizar una consistencia en cuanto a los criterios a través de los cuales las empresas presentan sus cuentas anuales y otros documentos informativos.

Este enfoque permite que el análisis refleje de manera fiel las dinámicas del IBEX 35 a lo largo del tiempo, así como garantiza los mismos criterios a la hora de extraer los datos.

¹⁹ <https://www.boe.es/>

²⁰ <https://www.heritage.org/index/>

²¹ <https://www.bolsasymercados.es/bme-exchange/es/Indices/Ibex/Composicion-historica>

COMPOSICIÓN HISTORCA DEL IBEX 35 2014-2023				EXCLUSIÓN	EN t+1	INCLUSIÓN	EN t	CAMBIO NOMBRE	DE t A t+1
2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
ABE	ABE	ABE	ABE	ACS	ACS	ACS	ACS	ACS	ACS
ABG.P	ACS	ACS	ACS	ACX	ACX	ACX	ACX	ACX	ACX
ACS	ACX	ACX	ACX	AENA	AENA	AENA	AENA	AENA	AENA
AMS	AENA	AENA	AENA	AMS	AMS	ALM	ALM	AMS	AMS
ANA	AMS	AMS	AMS	ANA	ANA	AMS	AMS	ANA	ANA
BBVA	ANA	ANA	ANA	BBVA	BBVA	ANA	ANA	ANE	ANE
BKIA	BBVA	BBVA	BBVA	BKIA	BKIA	BBVA	BBVA	BBVA	BBVA
BKT	BKIA	BKIA	BKIA	BKT	BKT	BKIA	BKT	BKT	BKT
BME	BKT	BKT	BKT	CABK	CABK	BKT	CABK	CABK	CABK
CABK	CABK	CABK	CABK	CIE	CIE	CABK	CIE	CLNX	CLNX
DIA	DIA	CLNX	CLNX	CLNX	CLNX	CIE	CLNX	COL	COL
ELE	ELE	DIA	COL	COL	COL	CLNX	COL	ELE	ELE
ENG	ENG	ELE	DIA	ELE	ELE	COL	ELE	ENG	ENG
FCC	FCC	ENG	ELE	ENC	ENC	ELE	ENG	FDR	FDR
FER	FER	FER	ENG	ENG	ENG	ENG	FDR	FER	FER
GAM	GAM	GAM	FER	FER	FER	FER	FER	GRF	GRF
GAS	GAS	GAS	GAS	GRF	GRF	GRF	GRF	IAG	IAG
GRF	GRF	GRF	GRF	IAG	IAG	IAG	IAG	IBE	IBE
IAG	IAG	IAG	IAG	IBE	IBE	IBE	IBE	IDR	IDR
IBE	IBE	IBE	IBE	IDR	IDR	IDR	IDR	ITX	ITX
IDR	IDR	IDR	IDR	ITX	ITX	ITX	ITX	LOG	LOG
ITX	ITX	ITX	ITX	MAP	MAP	MAP	MAP	MAP	MAP
JAZ	MAP	MAP	MAP	MEL	MAS/MAS	MEL	MEL	MEL	MEL
MAP	MRL	MEL	MEL	MRL	MEL	MRL	MRL	MRL	MRL
MTS	MRL	MRL	MRL	MTS	MRL	MTS	MTS	MTS	MTS
OHL	OHL	MTS	MTS	NTGY	MTS	NTGY	NTGY	NTGY	NTGY
POP	POP	POP	REE	REE	NTGY	PHM	PHM	RED	RED
REE	REE	REE	REP	REP	REE	REE	REE	REP	REP
REP	REP	REP	SAB	SAB	REP	REP	REP	ROVI	ROVI
SAB	SAB	SAB	SAN	SAN	SAB	SAB	ROVI	SAB	SAB
SAN	SAN	SAN	SGRE	SGRE	SAN	SAN	SAB	SAN	SAN
SCYR	SCYR	TEF	TEF	TEF	SGRE	SGRE	SAN	SCYR	SCYR
TEF	TEF	TL5	TL5	TL5	TEF	SLR	SGRE	SLR	SLR
TL5	TL5	TRE	TRE	TRE	TL5	TEF	SLR	TEF	TEF
TRE	TRE	VIS	VIS	VIS	VIS	VIS	TEF	UNI	UNI

Tabla 2. Composición del IBEX entre 2014 y 2023 con variaciones.

4.1.3. Variables identificativas

A partir de toda la información recopilada de las anteriores fuentes, se han seleccionado una serie de variables identificativas, siendo estas un elemento esencial para poder estructurar y organizar el *Dataset* de la manera adecuada, proporcionando un contexto claro sobre cada una de las empresas incluidas en el análisis. Estas variables son fundamentales para la identificación y clasificación de las empresas, siendo esenciales en análisis comparativos, temporales o segmentaciones. Las variables identificativas que hemos utilizado son las siguientes:

ETIQUETA

Esta variable hace referencia al identificador que usan las empresas dentro de los mercados bursátiles, la cual actúa como identificador de cada empresa dentro del *Dataset*. A su vez, esta puede ser útil para diferentes análisis comparativos en los que poner el nombre completo de la empresa no haga tan sencilla su interpretación.

NOMBRE

El nombre de la empresa hace referencia a la denominación que estas corporaciones utilizan en sus respectivos registros en la Comisión Nacional del Mercado de Valores, así como en los diferentes documentos que han sido utilizados como fuente de datos. Esta variable, no solo facilita

la fácil identificación de la documentación necesaria, sino que también asegura que cada una de las variables esté asignada a su corporación correspondiente.

AÑO

Esta variable indica el año en el que se recogen los datos de cada una de las empresas. Se ha decidido tomar un periodo de diez años, desde el 31 de diciembre de 2013 al 31 de diciembre de 2023. La decisión de limitarse a diez años para el análisis se toma en base a dos criterios. En primer lugar, hemos intentado garantizar la uniformidad de los datos, dado que ampliar el periodo temporal podría llevar a que los documentos utilizados para el análisis tuviesen formatos diferentes o no se mostrase el mismo tipo de información. Por otro lado, diez años es un periodo suficientemente amplio para poder garantizar la calidad del análisis llevado a cabo, pudiendo identificar de esta forma las tendencias temporales, evaluar la evolución o analizar cómo los diferentes sucesos históricos han afectado al desempeño de las empresas del IBEX.

SECTOR

La variable sector hace referencia a la industria o área de actividad a la que pertenece la empresa. Este campo ha sido extraído de los datos de la entidad según figura en la Comisión Nacional del Mercado de Valores. Esta clasificación por sectores no solo facilita la agrupación de los datos para su posterior análisis en un nivel más desagregado, sino que da pie a la realización de numerosos análisis comparativos a nivel sectorial.

4.1.4. Variables económicas

Todas las variables económicas seleccionadas para esta sección tienen su fundamento en literatura utilizada para el análisis económico de las empresas [14][15]. No forman parte de esta jerga DIVIDENDO, BENEFICIO A RESEVAS Y VARIACION PRECIO ACCIONES.

RATIO DE ENDEUDAMIENTO

El ratio de endeudamiento es un indicador importante de la estructura de capital de una organización. Esta ratio mide la proporción de deuda que posee una empresa en relación con su activo total, es decir, muestra qué parte de sus activos están financiados con deuda y qué parte con capital propio de los accionistas. Proporciona una visión clara de la composición financiera de la empresa, así como de su capacidad para afrontar sus obligaciones tanto a corto como a largo plazo.

La situación financiera de las organizaciones, así como su nivel de endeudamiento son factores que pueden afectar a su capacidad de crecimiento y estabilidad económica. Un exceso de endeudamiento puede limitar la capacidad de las empresas de afrontar situaciones imprevistas, así como de invertir en nuevos proyectos, por ello, lo ideal es mantener un equilibrio entre deudas y activo.

El ratio de endeudamiento es igual al total de la deuda que posee la empresa, tanto a corto como a largo plazo, partido por el total de su pasivo, es decir, la suma de las deudas más el patrimonio neto.

$$\text{Ratio de endeudamiento} = \frac{\text{Total deudas}}{\text{Total patrimonio neto más pasivo}}$$

En definitiva, la elección del ratio de endeudamiento se debe principalmente a su relación con la salud financiera de las empresas, permitiendo a todos los interesados evaluar el riesgo de estas.

ROTACIÓN ACTIVOS

La rotación de activos es un indicador que mide la capacidad de una empresa para generar ventas a partir de sus activos totales. Este ratio proporciona una imagen de cómo de eficiente se están empleando los activos de una empresa para generar ingresos.

En otro orden de ideas, una alta rotación de activos no solo muestra una gestión efectiva, sino también una mayor competitividad. Esto se debe a que las organizaciones con estas características son capaces de generar más ingresos con menos cantidad de inversión, mejorando de esta manera su rentabilidad y competitividad dentro del mercado.

El ratio de rotación de activos es igual al Importe Neto de la Cifra de Negocios (INCN), es decir, la facturación obtenida ese año a partir de la actividad antes de gastos, impuestos e intereses, partido por el valor total de los activos de la empresa.

$$\text{Ratio de rotación de activos} = \frac{\text{Importe Neto de la Cifra de Negocios}}{\text{Total activo}}$$

La relación estrecha entre la rotación de activos y la competitividad es algo que nos ha hecho considerar esta variable como fundamental para el éxito empresarial.

MARGEN OPERATIVO

El margen operativo es un indicador clave que mide la rentabilidad de una empresa antes de deducir intereses e impuestos (BAII). Hace referencia al porcentaje de las ventas que se convierte en beneficios como consecuencia de su actividad, dando así una visión precisa de la eficiencia con la que la empresa gestiona sus gastos operativos.

El margen operativo es fundamental porque permite evaluar no solo la rentabilidad, sino también lo eficiente que es la empresa en sus operaciones. Un margen operativo alto, hace referencia a que la empresa controla de manera efectiva sus costes y es capaz de sacar rédito a partir de sus ventas. Además de emplearlo con la finalidad de evaluar la eficiencia operativa, el margen operativo también puede ser una herramienta clave para medir la estabilidad financiera de una empresa en tiempos de fluctuaciones económicas, permitiendo de esta manera a gestores e inversores entender cómo una empresa es capaz de mantener su rentabilidad a pesar de un cambio en los ingresos o en los gastos.

El margen operativo es igual al Beneficio antes de Intereses e Impuestos (BAII) entre el Importe Neto de Cifra de Negocios (INCN), todo esto multiplicado por cien, obteniéndolo de esta manera en forma porcentual:

$$\text{Margen operativo} = \frac{\text{Beneficio Antes de Intereses e Impuestos}}{\text{Importe Neto de la Cifra de Negocios}} \times 100$$

La comprensión holística de la eficiencia y la rentabilidad de una empresa a través de un solo ratio ha sido el principal motivo de la elección del margen operativo como una de las variables clave del análisis.

MARGEN

El margen neto es un ratio financiero que, al igual que el margen operativo, mide la rentabilidad de la empresa, pero este lo hace una vez se han deducido todos los gastos, tanto los operativos como los intereses y los impuestos.

La relevancia del margen radica en su capacidad de mostrar de forma global la rentabilidad de la empresa, es decir, mostrar en qué medida afectan los gastos operativos, intereses e impuestos a la obtención de beneficios.

El margen neto se obtiene al dividir el Beneficio Neto (BN) entre el Importe Neto de la Cifra de Negocios:

$$\text{Margen neto} = \frac{\text{Beneficio Neto}}{\text{Importe Neto de la Cifra de Negocios}} \times 100$$

ROA OPERATIVO

El ROA operativo, también conocido como retorno sobre los activos operativos, es un ratio que mide cuál es la capacidad de una empresa de generar beneficios operativos a partir de sus activos.

La ROA operativa toma un papel clave ya que combina un análisis desde dos perspectivas, mide lo rentable que llegan a ser las operaciones, así como la calidad de la gestión de los activos, factores clave para determinar el crecimiento a largo plazo y la sostenibilidad de la empresa.

La forma de calcular la ROA operativa es a través de la combinación de dos de los ratios mencionadas anteriormente. La multiplicación del margen operativo y la rotación de activos da lugar a una división entre el Beneficio antes de Impuestos e Intereses entre los Activos Totales:

$$\text{ROA Operativa} = \frac{\text{Beneficio Antes de Intereses e Impuestos}}{\text{Total activo}} \times 100$$

En definitiva, permite obtener un análisis global de la situación de la empresa en términos de eficiencia y rentabilidad todo en un solo ratio.

ROA

El Retorno sobre Activos (ROA) es un indicador que, a diferencia de la ROA Operativa, mide la rentabilidad total de la empresa en relación con sus activos, es decir, se tiene en cuenta los beneficios netos finales obtenidos por las empresas y no solo aquellos generados a partir de sus operaciones propias.

Proporciona una visión directa de la forma en la que una empresa es capaz de utilizar sus activos para generar beneficios netos, hecho que lo convierte en uno de los principales indicadores usados para medir la confianza en una empresa, así como la posibilidad de realizar una inversión en la misma.

Pese a todo, es importante entender que tiene ciertas limitaciones, ya que no se está teniendo en cuenta la forma en la que la empresa obtiene financiación ni el impacto del sistema tributario. Por ello, y también debido a su variabilidad dependiendo del sector, es importante que las comparaciones se realicen dentro de un mismo sector.

La forma en la que se calcula es a través de la división entre el Beneficio Neto entre la totalidad de los activos de la empresa:

$$\text{ROA} = \frac{\text{Beneficio Neto}}{\text{Total activo}} \times 100$$



MULT CAPITAL

El multiplicador de apalancamiento, también conocido como multiplicador de capital, es un indicador que mide el grado de apalancamiento financiero de una empresa.

Este dato, es importante porque proporciona una foto de cómo se está llevando a cabo el financiamiento de empresa, mostrando qué proporción de sus activos están financiados por deuda y cuanta por capital propio de los accionistas. Este indicador también, de forma subyacente, sirve como indicador de riesgo, donde aquellas empresas con una mayor proporción de deuda pueden ser empresas en las que se tenga menos confianza.

Su cálculo se realiza a través de la división entre los activos totales de las empresas entre los fondos propios de la misma:

$$\text{Multiplicador de Capital} = \frac{\text{Activos Totales}}{\text{Fondos Propios}}$$

EFECTO APALANCAMIENTO

El efecto del apalancamiento mide de manera efectiva como los gastos e ingresos financieros influyen en la rentabilidad de la empresa.

Esta ratio se puede utilizar para varias funciones. En primer lugar, es una herramienta que sirve para medir cómo los costes de la deuda o los intereses que hay que pagar están afectado a la empresa, de esta manera poder clarificar si el endeudamiento está sirviendo para aumentar la rentabilidad de los accionistas o, por lo contrario, está mermando la capacidad de crecimiento de la empresa debido a sus elevados costes. En segundo lugar, se utiliza como indicador para decidir si es momento de cambiar la estructura de financiamiento o si el camino actual es el adecuado, dando pie así a posibles renegociaciones de las condiciones de la deuda o búsqueda de otros proveedores de esta.

La forma de calcularlo es a través de la relación entre el Beneficio antes de Impuestos (BAI) y el Beneficio antes de Impuestos e Intereses (BAII):

$$\text{Efecto apalancamiento} = \frac{\text{Beneficio Antes de Impuestos}}{\text{Beneficio Antes de Intereses e Impuestos}}$$

Un valor alto de este ratio indica que los costes están siendo cubiertos de manera efectiva y que dicho endeudamiento puede ser positivo para la rentabilidad de los accionistas. Por el contrario, un valor inferior a la unidad da indicios sobre la necesidad de renegociar ciertas condiciones o la necesidad de un cambio en la estructura de financiación.

EFECTO FISCAL

El efecto fiscal es un indicador que mide el impacto que tienen los impuestos sobre la rentabilidad de una empresa.

Es otro indicador que mide de forma específica cómo afecta, en este caso los impuestos, a los resultados obtenidos por las empresas, pudiendo de esta manera hacer un análisis completo de cuál es la parte que más influye en la obtención final de resultados.

La fórmula para calcular el Efecto Fiscal es la relación entre Beneficio Neto (BN) y el Beneficio antes de Impuestos (BAI):

$$\text{Efecto fiscal} = \frac{\text{Beneficio Neto}}{\text{Beneficio Antes de Impuestos}}$$

PRESION FISCAL

De igual manera que el efecto fiscal mide la proporción de las ganancias que se destinan a beneficios netos, la presión fiscal, mide aquella parte que acaba siendo destinada a pagar impuestos.

Tanto este ratio como el efecto fiscal son indicadores complementarios que se utilizan para ver cuál es la carga fiscal real de la empresa, así como para poder buscar nuevas oportunidades de optimización. Pudiendo optar por deducciones o incentivos fiscales que ayuden a reducir la carga tributaria y, por tanto, aumentar la rentabilidad de la empresa.

La fórmula se calcula como el complementario del efecto fiscal:

$$\text{Presión fiscal} = 1 - \text{Efecto Fiscal}$$

ROE

El Return on Equity (ROE) es un indicador financiero que mide la rentabilidad obtenida por parte de la empresa en base al capital invertido por parte de los accionistas

Primeramente, muestra de manera directa el retorno que obtienen los accionistas por cada unidad monetaria invertida en la empresa. Un ROE elevado muestra que la empresa está generando un buen rendimiento en base al capital de los accionistas, lo cual no solo se puede traducir en un aumento del precio de las acciones, sino también en un gran indicador de confianza capaz de atraer nuevas inversiones que hagan posible la escalabilidad de la empresa.

Por otro lado, es una herramienta fundamental a nivel comparativo, ya que realiza un análisis global, permite que los inversores lo usen el ROE para comparar entre las empresas, incluso fuera del sector, con el fin de maximizar el rendimiento de su capital.

La forma en la que se calcula el ROE es a través de la relación que existe entre el Beneficio Neto (BN) y los Fondos Propios de la empresa:

$$\text{ROE} = \frac{\text{Beneficio Neto}}{\text{Fondos Propios}}$$

Una forma alternativa de cálculo es a través de algunos de los ratios mencionadas anteriormente. Este desglose permite ver qué ha sido aquello que más ha influido en la rentabilidad y poner el foco en ello:

$$\text{ROE} = \text{Margen} \times \text{Rotación Activos} \times \text{Multiplicador Capital} \times \text{Apalancamiento} \times \text{Efecto Fiscal}$$

RATIO LIQUIDEZ

El ratio de liquidez mide la capacidad de la empresa de cumplir con sus obligaciones de pago a corto plazo.

El objetivo principal de este ratio es la solvencia de la empresa, donde un ratio superior a la unidad muestra que la empresa tiene suficientes activos para cubrir todas sus deudas a corto plazo.



Por otro lado, un ratio inferior a la unidad puede significar que hay diversos problemas de solvencia para atender proveedores o algunas de sus obligaciones más cercanas en el tiempo.

Los inversores y acreedores también empujan el ratio de liquidez para medir de forma efectiva el riesgo de invertir o prestar capital a una determinada empresa. Por ello, un ratio de liquidez adecuado puede aumentar la confianza de los inversores, así como facilitar el acceso a financiamiento barato y de calidad.

Se calcula dividiendo el total de sus Activos Corrientes, es decir, aquellos que tienen un vencimiento inferior a un año, entre su Pasivo Corriente:

$$\text{Ratio de liquidez} = \frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$$

Su uso para determinar la confianza de los diferentes actores que interactúan con una organización, así como la utilización para determinar la solvencia en el corto plazo de la empresa, es aquello que la convierte en una herramienta fundamental para tener en cuenta dentro de un análisis sobre el éxito de las empresas.

% SERVICIOS EXTERIORES

Los servicios exteriores hacen referencia a aquellos servicios que una empresa subcontrata a terceros en lugar de realizarlos internamente. El porcentaje de ventas que una empresa destina a estos refleja cómo una empresa es capaz de gestionar sus costes operativos, así como de su dependencia de terceros.

Las partidas que pueden estar incluidas dentro de los servicios pueden ser muy amplias. En general, las más comunes son los diferentes servicios en publicidad o marketing que no son realizados de forma interna, gastos en asesoría o consultoría, tanto financiera como operativa o los diferentes mantenimientos o reparaciones que los realicen técnicos ajenos a la empresa.

La fórmula utilizada es la que relaciona la partida Servicios Exteriores con el Importe Neto de Cifra de Negocios (INCEN):

$$\% \text{ Servicios Exteriores} = \frac{\text{Servicios Exteriores}}{\text{Importe Neto de la Cifra de Negocios}} \times 100$$

Conocer como de dependiente es una empresa del exterior, así como los costes que estos suponen en relación con sus ingresos es lo que nos hace considerar esta variable como esencial a la hora de incluirla dentro del análisis.

DIVIDENDO

Dividendo es una variable binaria (SI/NO) que indica si una empresa ha distribuido los beneficios obtenidos en un determinado año entre sus accionistas en forma de dividendo. El empleo de esta variable en el análisis tiene dos principales enfoques, el primero de ellos, es que sirve como un indicador del perfil de riesgo de la empresa a analizar. Las empresas que optan por repartir dividendos, incluso en tiempos de pérdidas, dan indicios de un perfil empresarial más riesgoso que aquellas que deciden guardar sus beneficios en reservas. Por otro lado, el reparto o no de dividendos, puede ser una muestra de la salud tanto financiera como operativa de la empresa, más allá de su perfil más o menos arriesgado.

$$\text{Dividendo} = \text{SI reparte dividendos o NO reparte dividendos}$$

BENEFICIO A RESERVAS

La variable Beneficio a Reservas es de tipo binario (SI/NO) y muestra si una empresa ha decidido destinar sus beneficios anuales a alguna de las reservas durante un año en concreto. Destinar parte de los recursos a reserva puede tener diferentes fines, desde su utilización a modo de reinversión hasta un cumplimiento normativo.

La decisión de utilizar esta variable se basa fundamentalmente en su relación directa con la estrategia financiera de la empresa, así como el perfil de riesgo de esta.

Beneficio a reservas = SI destina a reservas o NO destina a reservas

ESTRUCTURA DE COSTES

La estructura de costes de la empresa hace referencia a cuáles son los tipos de coste que predominan en el proceso operativo de una empresa, las variables, es decir, aquellos ligados directamente a la producción de esta y que aumentan o disminuyen en función de esta, o los costes fijos, aquellos que se dan siempre independientemente de la actividad que haya.

La distinción entre estos dos tipos de coste se ha hecho comparando las partidas de Pérdidas y Ganancias, donde el apartado Aprovisionamientos constituía la parte de costes variables, y los apartados de Sueldos, Deterioros y Amortizaciones formaban la parte de coste fijo. En función de cuál de las dos partes fuese mayor, se le asignaba el valor de (FIJO/VARIABLE).

La principal decisión de añadir este tipo de variable ha sido determinar si había algún tipo de relación entre la forma que las empresas gestionan sus costes y los posibles resultados finales.

Estructura de costes = Estructura VARIABLE o Estructura FIJA

AUMENTA GASTOS PERSONAL

El aumento del gasto del personal se muestra también en forma de variable binaria (SI/NO), y determina si el gasto en personal ha aumentado de un año a otro, ya sea por aumento del número de personas trabajando o por la mejora de condiciones laborales.

Esto se hace visualizando la partida de *Sueldos y Salarios* de Pérdidas y Ganancias, se compara si ha habido un aumento en dicha partida con respecto al año anterior, en caso afirmativo se le asignará el valor SI, en caso de no aumentar o de mantenerse estable, se le asignará el valor NO.

Dado que uno de los factores que influyen de manera directa en la productividad de la empresa son el número de trabajadores que esta tiene o lo eficientes que sean estos trabajadores, la forma de reflejarlo es a través de esta variable, buscando si hay alguna relación entre el aumento del gasto y una mejora tangible en la rentabilidad de la empresa.

Gasto personal = SI aumenta gasto o NO aumenta gasto

CICLO DE VIDA DE LA EMPRESA

El ciclo de vida de la empresa es una variable que pretenden ubicar la situación de una empresa en un año determinado dentro de unas secciones previamente definidas.

El análisis se hace a partir del Estado de Flujos de Efectivo, en el que conforme sea la configuración de los diferentes apartados, se le asigna un valor u otro.

El Estado de Flujos de Efectivo este compuesto por tres partidas, el flujo operativo, de inversión y de financiación, los cuales pueden tener valores positivos o negativos en función de si entra efectivo o sale en cada una de las partidas. A partir de esto, se crean seis posibles combinaciones que dan lugar a los seis posibles valores que toma la variable. Estas combinaciones son las siguientes:

- **INTRODUCCIÓN:** Operativo = Negativo; Inversión = Negativo; Financiación = Positivo
- **CRECIMIENTO:** Operativo = Positivo; Inversión = Negativo; Financiación = Positivo
- **MADUREZ:** Operativo = Positivo; Inversión = Negativo; Financiación = Negativo
- **DECLIVE:** Operativo = Positivo; Inversión = Positivo; Financiación = Negativo
- **REESTRUCTURA:** Operativo = Negativo; Inversión = Positivo; Financiación = Positivo
- **LIQUIDACIÓN:** Operativo = Negativo; Inversión = Positivo; Financiación = Negativo

La decisión de introducir esta variable se basa principalmente en la información que puede llegar a aportar en relación con la rentabilidad de la empresa o con las otras variables mencionadas anteriormente.

BENEFICIO

El beneficio de la empresa se ha mostrado en este análisis en forma de variable binaria (POSITIVO/NEGATIVO) haciendo referencia a como había sido el beneficio de una empresa en un año determinado.

La introducción de esta variable se debe principalmente al alto valor agregado que puede proporcionar a la hora de hacer análisis segmentados o de elegir un criterio para diferenciar dentro de un análisis.

VARIACIÓN PRECIO ACCIONES

La variación del precio de las acciones muestra cuál ha sido el cambio de valor porcentual de una acción en un determinado periodo de tiempo. En este caso, se coge el precio de cierre de las acciones a 31 de diciembre de un año y se compara con el precio de cierre a 31 de diciembre del año siguiente. Se han elegido estas fechas con el objetivo coincidir con la fecha de presentación de las cuentas anuales de cada una de las empresas del IBEX 35.

La razón por la que se ha elegido esta variable es principalmente por la información que aporta al respecto de la percepción del mercado sobre el valor del mercado o de la empresa, pudiendo influir esto significativamente en la rentabilidad de las empresas.

La forma en la que se ha calculado la variación ha sido a través de la siguiente fórmula:

$$\%Var Stock Price = \left(\frac{Precio\ acción_{t+1} - Precio\ acción_t}{Precio\ acción_t} \right) \times 100$$

4.1.5. Variables exógenas

Las variables exógenas son factores que están fuera del control directo de una empresa pero que pueden influir de manera significativa en el desarrollo financiero y operativo de la empresa. Estas variables hacen referencia al entorno macroeconómico, político y social en el que operan las empresas [16].

En este apartado, analizaremos cuáles han sido las variables externas elegidas para este estudio, remarcando su importancia y el impacto que pueden llegar a tener sobre las empresas del IBEX 35.

ELECCIONES

Las elecciones generales de un país son un evento fundamental que puede influir de manera significativa en el sector empresarial del país. En este caso se expresa en forma de variable binaria (SI/NO) marcando con un SI los años en los que en España ocurrieron elecciones generales y un no aquellos en los que no.

Las elecciones generales pueden llevar a cambios en las políticas económicas o en regulaciones que afecten de manera directa a las empresas. El gobierno que resulte como vencedor en las elecciones podría implementar nuevas reformas fiscales, introducir nuevos impuestos o poner ciertos incentivos a las corporaciones aparte de poder regular en materia laboral y medioambiental [17].

Por todo ello y por ser un factor clave que puede influir en la confianza en los mercados, consideramos que es un factor fundamental para tener en cuenta dentro del análisis.

NUEVA LEGISLACIÓN LABORAL

En el caso de la nueva legislación laboral, tiene como finalidad marcar la relevancia de la reforma laboral introducida por el gobierno de España en el año 2022. Esta variable marca con un SI el año 2022 en el que se introdujo la reforma y con un NO los otros años.

Un cambio en la legislación laboral vigente puede influir significativamente en el ámbito operativo de la empresa, ya que, estas reformas pueden incluir desde cambios o variaciones en el salario mínimo lo cual afecta directamente a la estructura de costes de la empresa, hasta modificaciones en la regulación sobre tipo de contratos u horas trabajadas lo cual implica una necesidad de cambio en la estructura operativa y organizacional de la empresa [18].

Dado que este tipo de regulaciones pueden generar cambios significativos en las organizaciones y dan pie a ver como estas se pueden adaptar al cambio, será de gran ayuda a lo largo del análisis.

PANDEMIA

En cuanto a la variable, lo que pretende es reflejar el inicio de la pandemia del COVID-19 en el año 2020, marcando de esta manera con un SI el año 2020 y con un no los demás años.

Tanto el inicio de la pandemia como la posterior cuarentena fue un momento que afectó significativamente a la gran mayoría de las empresas a nivel global. Los parones en las cadenas de producción, la subida de precios o los cambios organizativos fueron algunos de los efectos causado por este fenómeno [19].

Con su inclusión dentro de las variables buscamos medir de manera real el impacto que tuvo en términos de rentabilidad y determinar cuáles fueron los tipos de empresa más afectados y las que mejor sobrepasaron este momento.

GUERRA

La variable guerra hace referencia a otro de los eventos que más han marcado el desempeño de las empresas a nivel europeo en los últimos años, el estallido de guerra entre Rusia y Ucrania. A través de una variable binaria, marca con un SI el inicio del conflicto en el año 2022 dejando los otros años con un NO.

A raíz de este conflicto, se produjeron altas subidas de precio tanto en materia prima como en la energía, siendo ambos dos elementos fundamentales para el funcionamiento de las empresas y la subida de los cuales modificó de forma drástica sus estructuras de costes así como su previsión operativa [20].

A fin de medir cuál ha sido el impacto real de este evento en las empresas del IBEX 35 hemos decidido introducir esta variable dentro del análisis.

ÍNDICE LIBERTAD ECONÓMICA

La libertad económica de un país es un factor fundamental que influye en el desarrollo del entorno empresarial de determinado país. En este caso, esta se mide a través del *Index of Economic Freedom* elaborado por la Fundación Heritage, la cual elabora este índice para diversos estados del mundo.

El índice abarca diversos aspectos de la economía del país, como la protección de la propiedad privada, las regulaciones, el nivel de gasto público o la intervención del gobierno en la economía [21].

Todo esto se mide a través de un sistema de pesos que acaban dando lugar al índice final. Dado que el índice permite realizar un análisis tan amplio, reflejado en una sola variable, creemos oportuno introducirla en el análisis por su gran aporte descriptivo sobre el entorno en el que operan las empresas del país.

DERECHOS DE PROPIEDAD

Los derechos de propiedad es uno de los componentes que forma parte del *Index of Economic Freedom* que mide cuál es la protección legal ofrecida por un país a la propiedad privada. Este ratio mide hasta qué punto los individuos y las empresas pueden adquirir una propiedad y cómo de protegidos están por la legislación del gobierno.

Los derechos de propiedad bien definidos y protegidos garantizan a las empresas que puedan operar sin miedo a que su propiedad se vea vulnerada de alguna forma, lo cual fomenta la estabilidad del entorno y de una proyección sólida a largo plazo. Esto se acaba traduciendo en seguridad de cara a los inversores y accionistas, ofreciendo de esta manera garantías de que el entorno donde se va a destinar el capital es seguro y confiable [21].

Los derechos de propiedad dentro del *Index of Economic Freedom* se basan en los siguientes subfactores:

- Riesgo de Expropiación
- Respeto a los Derechos de Propiedad Intelectual
- Calidad de la Aplicación de Contratos, Derechos de Propiedad y Aplicación de la Ley

GASTO PÚBLICO

El gasto público es otro de los componentes del *Index of Economic Freedom* que evalúa la cantidad de gastos que efectúa el Estado, es decir, trata de medir como el tamaño del Estado y la asignación de los recursos públicos termina impactando en la economía

La forma en la que se asigna el gasto público es de vital importancia para el entorno empresarial, estando directamente relacionado con la capacidad de crecimiento futura de la organización. Un exceso de gasto público mal asignado puede implicar déficits en las cuentas, así como un aumento de la deuda. Esto se puede traducir en una pérdida de confianza por parte de los inversores en el país, disminuyendo de esta manera la inversión privada y cortando de raíz el potencial de crecimiento de las empresas [21].

Este ratio se calcula de manera no lineal, donde los valores más cercanos a cero reciben ciertas penalizaciones en la puntuación, aquellos superiores a un 30% del Producto Interior Bruto obtienen puntuaciones significativamente más bajas y solo las superiores a un valor del 58% del Producto Interior Bruto reciben una puntuación de cero.

LIBERTAD DE NEGOCIOS

La libertad de negocios es el tercer componente del *Index of Economic Freedom* que hemos decidido usar para el análisis. En este caso, trata de medir la forma en la que el entorno regulatorio permite la operación de las empresas, es decir, lo fácil o difícil que supone iniciar, operar o cerrar un negocio.

Este tipo de libertad regulatoria es fundamental para las empresas ya que un entorno flexible y eficiente con pocas restricciones burocráticas permite a las empresas operar de forma más ágil y competitiva, traduciéndose esto en menores costes de cumplimiento y mayor rapidez a la hora de poder operar [21].

Los subfactores en los que se basa el ratio de libertad de negocios son los siguientes:

- Acceso a la Electricidad
- Riesgo del Entorno Empresarial
- Calidad Regulatoria
- Inclusión Económica de la Mujer

LIBERTAD LABORAL

La libertad laboral es también un elemento del *Index of Economic Freedom* y mide aspectos relacionados con el marco regulatorio en el ámbito laboral del país.

Es esencial para las empresas disponer de un mercado laboral flexible y eficiente que les permita adaptarse a cualquier situación cambiante. Un alto valor de este ratio refleja menos restricciones burocráticas, lo cual hace la gestión de la fuerza laboral una tarea más sencilla, traduciéndose esto en un aumento de la competitividad de las empresas pudiendo reducir costes y mejorar la optimización de sus plantillas [21].

Hay nueve subfactores en los que se basa la libertad laboral:

- Salario Mínimo
- Derecho de Asociación
- Vacaciones Pagadas Anuales
- Periodo de Preaviso para Despido
- Indemnización por Despido
- Productividad Laboral
- Tasa de Participación Laboral
- Restricciones sobre Trabajo Extra
- Despidos Permitidos por Ley

TASA DE PARO

La tasa de paro en este caso mide el porcentaje de la población activa que se encuentra desempleada a 31 de diciembre de cada año desde 2013 a 2023.

Una tasa de paro elevada puede ser señal de problemas estructurales en la economía, esta afecta directamente tanto a la mano de obra necesaria para las empresas como a la demanda de productos y servicios de la cual las empresas dependen [22].

Su estrecha relación con el entorno macroeconómico de las empresas es lo que nos ha hecho elegir el paro como uno de los factores para tener en cuenta.

VARIACIÓN DEL PIB

La variación del PIB es un indicador que mide el cambio anual en el Producto Interior Bruto de un país, mostrando de esta manera lo que ha crecido o disminuido la productividad de un país en términos porcentuales a lo largo de un año. Generalmente está relacionado con la estabilidad y salud de la economía, por ello, cuando hay un aumento de este indicador, es un claro reflejo de seguridad que incentiva a las empresas hacia la expansión. También está relacionado con la demanda de bienes y servicios a causa de un aumento de la productividad y por tanto del poder adquisitivo de las personas, lo cual también está fuertemente relacionado con el desempeño de las empresas.

VARIACIÓN DEL IPC

Por último, la variación del IPC hace referencia a cómo ha cambiado en términos porcentuales el Índice de Precios al Consumidor, es decir, el índice que mide el aumento de los precios de una serie de productos útiles en la economía. Dentro de estos se incluyen alimentos, combustibles, energía o vivienda, entre otros.

La relación entre el IPC y las empresas es clara, si se produce un aumento en este índice, puede ir directamente relacionado con un aumento de los costes de la empresa, en términos de materias primas o servicios necesarios para llevar a cabo su proceso productivo.

5. Desarrollo de la solución

En esta sección, se llevará a cabo toda la parte del análisis necesaria para el proyecto. El proceso comienza con una limpieza inicial del *Dataset*, seguida de un análisis descriptivo, que nos permitirá ver las características y las distribuciones de las diferentes variables. Posteriormente, se realizará un preprocesamiento de los datos para prepararlos para el análisis predictivo final. Todo este flujo de trabajo se ha implementado a través de un *pipeline* en Python, garantizando de esta manera la eficiencia y la replicabilidad del proceso.

5.1. Carga del Dataset

Lo primero que haremos para iniciar el *Pipeline*, será cargar los datos y las librerías necesarias para realizar tanto la fase de limpieza como la fase del análisis descriptivo.

La carga de la imagen se hace a través de la librería Pandas y su función *read_excel* en la cual se lee el Excel elaborado anteriormente como podemos (ver Figura 1).

```
df = pd.read_excel('PREPROCESADO.xlsx')
```

Figura 1. Carga de datos

A través de la sentencia *import* se añaden todas las librerías necesarias para realizar el trabajo (ver Figura 2).

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
import plotly.offline as py
import plotly.express as px
import plotly.graph_objs as go
import warnings
from IPython.display import display
from scipy import stats
py.init_notebook_mode(connected=True)
from scipy.stats import skew, kurtosis
from scipy.stats import chi2_contingency
import statsmodels.api as sm
```

Figura 2. Importación de librerías de Python

5.2. Visualización inicial

La visualización inicial de los datos nos va a proporcionar una primera imagen de la composición y estructura del *Dataset* así como de los tipos de datos que la conforman.

En primer lugar, a través de la sentencia *display* y llamando a la función *info()* veremos una primera imagen del *Dataset*, donde se nos indica el número de columnas, así como la cantidad de elementos que la conforma junto al tipo de dato que se trata (ver Figura 3).

```

Información del DataFrame:
RangeIndex: 350 entries, 0 to 349
Data columns (total 37 columns):
#   Column                                     Non-Null Count  Dtype
---  ---                                     -
0   ETIQUETA                                 350 non-null    object
1   AÑO                                       350 non-null    int64
2   NOMBRE                                   350 non-null    object
3   SECTOR                                   339 non-null    object
4   RATIO ENDEUDAMIENTO                     339 non-null    float64
5   ROTACION ACTIVOS                        339 non-null    float64
6   MARGEN OPERATIVO                        339 non-null    float64
7   MARGEN                                   339 non-null    float64
8   ROA OPERATIVA                           339 non-null    float64
9   ROA                                       339 non-null    float64
10  MULT CAPITAL                             339 non-null    float64
11  EFE FINANCIERO                           339 non-null    float64
12  APALANCAMIENTO                           339 non-null    float64
13  EFECTO FISCAL                           339 non-null    float64
14  PRESION FISCAL                           339 non-null    float64
15  ROE                                       339 non-null    float64
16  RATIO LIQUIDEZ                           323 non-null    float64
17  % SERV EXT                               277 non-null    float64
18  DIVIDENDO                                339 non-null    object
19  BNFO A RESERVAS                           339 non-null    object
20  ESTRUCTURA DE COSTES                     339 non-null    object
21  AUMENTA GASTOS PERSONAL                  339 non-null    object
22  CICLO DE VIDA                             339 non-null    object
23  BENEFICIO                                 339 non-null    object
24  AÑO ELECCIONES                           339 non-null    object
25  NUEVA LEG LAB                             339 non-null    object
26  PANDEMIA                                  339 non-null    object
27  GUERRA                                    339 non-null    object
28  LIBERTAD ECONOMICA                       339 non-null    float64
29  DERECHOS DE PROPIEDAD                    339 non-null    float64
30  GASTO PUBLICO                             339 non-null    float64
31  LIBERTAD DE NEGOCIOS                     339 non-null    float64
32  LIBERTAD LABORAL                         339 non-null    float64
33  TASA DE PARO                             339 non-null    float64
34  VAR% PIB                                  339 non-null    float64
35  VAR% IPC                                  339 non-null    float64
36  VAR% PRICE STOCK                         339 non-null    float64
dtypes: float64(23), int64(1), object(13)

```

Figura 3. Visualización inicial del Dataset

Como podemos ver en la Figura 4 , hay un total de 37 columnas y 350 filas, pero hay ciertas columnas que tienen menos valores de los habituales, para una mayor identificación, usaremos la librería *Missingno* que nos permite ver una imagen real de los valores faltantes dentro del *Dataset*.

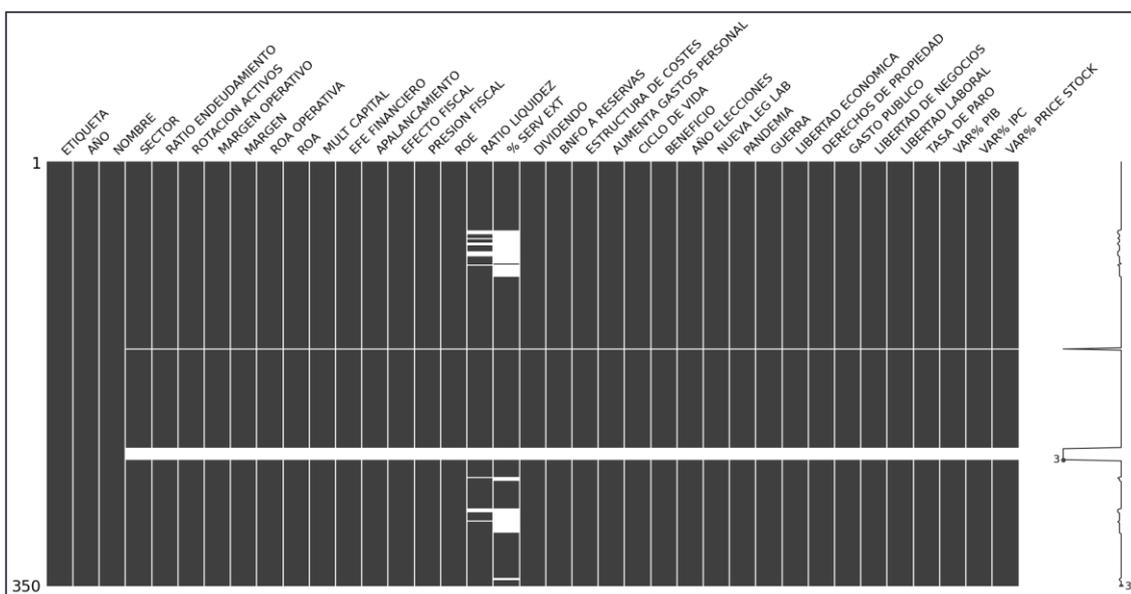


Figura 4. Visualización estructura inicial Dataset

Aquí podemos observar como las columnas “Ratio Liquidez” y “% Serv Ext” son las que presentan valores faltantes. Por otro lado, el hecho de que otras filas también tuviesen 339 valores, es decir, 11 menos, se debe a que las filas de las empresas *Arcelormittal* y *Ferrovial* no tienen datos durante unos periodos.

Una vez averiguado esto, pasaremos a la fase de limpieza donde trataremos los valores faltantes y realizaremos transformaciones en el *Dataset* con el fin de prepararlo para el análisis descriptivo.

5.3. Limpieza de datos

En la fase de limpieza de datos haremos principalmente dos cosas, en primer lugar, trataremos los valores faltantes del *Dataset* a través de la eliminación de filas o columnas, o sustitución de estos por valores similares. Por último, realizaremos transformaciones en alguna de las variables con el fin de facilitar el análisis descriptivo.

5.3.1. Eliminación de filas

El primer paso que vamos a realizar es la eliminación de aquellas filas que no contienen datos para poder realizar el análisis. En este caso se trata de las filas de las empresas de *Arcelormittal* y *Ferrovial*. En el caso de la primera, hay diez filas faltantes dado que es una empresa extranjera que está en el IBEX 35 pero no tributa en España, por tanto, no se rige bajo las mismas reglas contables y uniformidad del análisis se podría ver afectada. Por otro lado, el caso de *Ferrovial* es similar, ya que solo hay una fila faltante y esto es debido a que en el año 2023 España dejó de ser la residencia fiscal de esta empresa, volviendo a caer de nuevo en una disparidad de criterios (ver Figura 5).

```
df = df[df['NOMBRE'] != 'ARCELORMITTAL']
df = df[~((df['NOMBRE'] == 'FERROVIAL') & (df['AÑO'] == 2023))]
df = df.reset_index(drop=True)
```

Figura 5. Eliminación de filas

5.3.2. Eliminación de columnas

En segundo lugar, procederemos a hacer el tratamiento a las columnas. En este caso nos encontramos con dos casuísticas diferentes, en primer lugar, la eliminación de alguna columna debido a que su presencia en el análisis final no es relevante. Y, en segundo lugar, la eliminación debido a la gran cantidad de valores faltantes.

El primero de los casos corresponden a las columnas de *Nombre* y *Etiqueta*, dado que el tipo de análisis que se busca realizar tiene un nivel de agregación mayor al de las propias empresas, junto la dificultad de extraer análisis claros al tratar cada empresa una a una, nos lleva a la decisión de eliminar estos identificadores.

El segundo de los casos hace referencia a la columna *% Serv Ext* la cual tiene una gran cantidad de valores faltantes, esto se debe principalmente a que algunos tipos de empresas, por ejemplo, las entidades bancarias, no tienen dentro de sus Pérdidas y Ganancias expresados los servicios externos. Estos pueden tener distribuidos dentro de otros campos los cuales pueden ser variables en función de la empresa y el tipo de servicio que se contrate a terceros. Por ello, ante la no uniformidad de esta variable a lo largo de las diferentes empresas y, que la solución de rellenar estos valores nulos podría sesgar el análisis al no conocer de manera exacta donde se encuentran todos los servicios exteriores o en qué campos, se ha decidido suprimir la columna (ver Figura 6).

```
df = df.drop('ETIQUETA', axis=1)
df = df.drop('NOMBRE', axis=1)
df = df.drop('% SERV EXT', axis=1)
```

Figura 6. Eliminación de columnas

5.3.3. Valores nulos

Una vez hemos eliminado las filas y las columnas la cantidad de valores nulos se ha visto considerablemente reducida, en este caso trataremos aquellos que corresponden a la columna *Ratio Liquidez* como bien hemos podido observar en la visualización inicial de los datos.

Para la eliminación de estos valores se ha decidido sustituirlos por la mediana de los demás valores de *Ratio Liquidez* que posee una misma empresa. Las empresas que presentan valores faltantes en esta columna son bancos, esto se debe a que hasta el año 2016 no se introdujo una regulación que obligase a las entidades financieras hacer público cuál es su ratio de liquidez, por ello, hay años en los que no se tiene el dato. En cuanto a la elección de la mediana, considerando que las variables económicas pueden no ajustarse a una distribución normal y que medidas como la media pueden verse afectadas por valores atípicos, hemos decidido utilizar la mediana como una opción más robusta para tratar estos valores faltantes (ver Figura 7).

```

def identificar_empresas_con_nulos(df, columna):
    return df[df[columna].isnull()]['NOMBRE'].unique()

empresas_con_nulos = identificar_empresas_con_nulos(df, 'RATIO LIQUIDEZ')
df_filtrada = df[df['NOMBRE'].isin(empresas_con_nulos)]
valores_mediana = df_filtrada.groupby('NOMBRE')['RATIO LIQUIDEZ'].median()

def sustituir_con_mediana(fila):
    if pd.isnull(fila['RATIO LIQUIDEZ']):
        return valores_mediana[fila['NOMBRE']]
    else:
        return fila['RATIO LIQUIDEZ']

df_filtrada.loc[:, 'RATIO LIQUIDEZ'] = df_filtrada.apply(sustituir_con_mediana, axis=1)
df.update(df_filtrada)

```

Figura 7. Valores anómalos Ratio Liquidez

5.3.4. Transformaciones

Por último, vamos a realizar transformaciones para poder facilitar el análisis. En este caso se trata de una agrupación de los sectores de las diferentes empresas.

Dado que la cantidad de sectores en los que se dividían las empresas en base a la Comisión Nacional del Mercado de Valores eran excesivos, hemos decidido agruparlos en un número de grupos menos en base al tipo de producto o servicio que prestan (ver Figura 8).

```

df['SECTOR'] = df['SECTOR'].map({
    'APARCAMIENTO Y AUTOP': 'Construcción e Infraestructura',
    'NUEVAS TECNOLOGÍAS': 'Tecnología y Comunicaciones',
    'CONSTRUCCIÓN': 'Construcción e Infraestructura',
    'METÁLICAS BÁSICAS': 'Industria y Manufactura',
    'TRANSPORTES': 'Comercio y Servicios',
    'INDUSTRIA QUÍMICA': 'Industria y Manufactura',
    'ENERGÍA ELÉCTRICA': 'Energía y Recursos',
    'BANCOS': 'Servicios Financieros',
    'COMERCIO Y OTROS SERVICIOS': 'Comercio y Servicios',
    'COMUNICACIONES': 'Tecnología y Comunicaciones',
    'INMOBILIARIAS': 'Construcción e Infraestructura',
    'PAPEL Y ARTES GRÁFICAS': 'Industria y Manufactura',
    'AGUA Y GAS': 'Energía y Recursos',
    'OTRAS INDUST. MANUFACTURA': 'Industria y Manufactura',
    'SEGUROS': 'Servicios Financieros',
    'PETRÓLEO': 'Energía y Recursos',
    'INDUSTR. TRANSF.DE METALES': 'Industria y Manufactura',
    'MEDIOS DE COMUNICACIÓN': 'Tecnología y Comunicaciones',
    'ALIMENT. BEBIDAS Y TABACO': 'Consumo'
})

```

Figura 8. Transformación sectores

Los grupos nuevos junto con la correspondencia anterior son los siguientes:

- **Construcción e Infraestructura**
 - APARCAMIENTO Y AUTOP
 - CONSTRUCCIÓN
 - INMOBILIARIAS
- **Tecnología y Comunicaciones**
 - NUEVAS TECNOLOGÍAS
 - COMUNICACIONES
 - MEDIOS DE COMUNICACIÓN
- **Industria y Manufactura**
 - METÁLICAS BÁSICAS
 - INDUSTRIA QUÍMICA
 - PAPEL Y ARTES GRÁFICAS
 - OTRAS INDUST. MANUFACTURA
 - INDUSTR. TRANSF.DE METALES
- **Comercio y Servicios**
 - TRANSPORTES
 - COMERCIO Y OTROS SERVICIOS
- **Energía y Recursos**
 - ENERGÍA ELÉCTRICA
 - AGUA Y GAS
 - PETRÓLEO
- **Servicios Financieros**
 - BANCOS
 - SEGUROS
- **Consumo**
 - ALIMENT. BEBIDAS Y TABACO

5.4. Vista minable

Una vez hemos realizado toda la limpieza de valores faltantes así como las pertinentes transformaciones con tal de facilitar el análisis, hacemos un visualizado final, mostrado en la Figura 9 con tal de conocer el estado final del *Dataset*, asegurándonos que está listo para el análisis.

```
Información del DataFrame:
RangeIndex: 339 entries, 0 to 338
Data columns (total 34 columns):
#   Column                                     Non-Null Count  Dtype
---  ---                                     -
0   AÑO                                         339 non-null    int64
1   SECTOR                                     339 non-null    object
2   RATIO ENDEUDAMIENTO                       339 non-null    float64
3   ROTACION ACTIVOS                          339 non-null    float64
4   MARGEN OPERATIVO                          339 non-null    float64
5   MARGEN                                     339 non-null    float64
6   ROA OPERATIVA                             339 non-null    float64
7   ROA                                         339 non-null    float64
8   MULT CAPITAL                              339 non-null    float64
9   EFE FINANCIERO                            339 non-null    float64
10  APALANCAMIENTO                            339 non-null    float64
11  EFECTO FISCAL                             339 non-null    float64
12  PRESION FISCAL                            339 non-null    float64
13  ROE                                         339 non-null    float64
14  RATIO LIQUIDEZ                            339 non-null    float64
15  DIVIDENDO                                  339 non-null    object
16  BNFO A RESERVAS                           339 non-null    object
17  ESTRUCTURA DE COSTES                      339 non-null    object
18  AUMENTA GASTOS PERSONAL                   339 non-null    object
19  CICLO DE VIDA                              339 non-null    object
20  BENEFICIO                                  339 non-null    object
21  AÑO ELECCIONES                            339 non-null    object
22  NUEVA LEG LAB                              339 non-null    object
23  PANDEMIA                                   339 non-null    object
24  GUERRA                                     339 non-null    object
25  LIBERTAD ECONOMICA                         339 non-null    float64
26  DERECHOS DE PROPIEDAD                     339 non-null    float64
27  GASTO PUBLICO                              339 non-null    float64
28  LIBERTAD DE NEGOCIOS                       339 non-null    float64
29  LIBERTAD LABORAL                           339 non-null    float64
30  TASA DE PARO                               339 non-null    float64
31  VAR% PIB                                   339 non-null    float64
32  VAR% IPC                                   339 non-null    float64
33  VAR% PRICE STOCK                           339 non-null    float64
dtypes: float64(22), int64(1), object(11)
```

Figura 9. Visualización final Dataset

Aquí podemos observar la composición final del *Dataset*, con un total de 339 filas y 34 columnas esta vez ya sin valores faltantes.

5.5. Análisis descriptivo

Entramos en la fase de análisis primeramente con la parte descriptiva del mismo. Este apartado tiene como objetivo resumir y visualizar las características principales del conjunto de datos. Se analizará su distribución, tendencias y posibles relaciones entre las variables dando así una serie de *Insights* clave para el estudio.

5.5.1. Distribuciones de variables numéricas

En primer lugar, analizaremos cuáles son las distribuciones de las variables numéricas clave que representan las diferentes áreas de una empresa, optando por aquellas que proporcionan la información más relevante y manejable, junto con un análisis segmentado por sectores, viendo el impacto de estos en las variables. Analizaremos las distribuciones, dispersiones, valores anómalos y patrones relevantes, así como las lógicas económicas que subyacen a dicho análisis.

RATIO ENDEUDAMIENTO

En cuanto al ratio de endeudamiento mostrado en la Figura 10 podemos observar varias cosas. En primer lugar, podemos ver que se trata sustancialmente de una asimetría con sesgo positivo, donde la gran mayoría de los valores se encuentran en la parte derecha del gráfico, los cuáles nos muestran principalmente que dentro del IBEX 35 en los últimos diez años ha habido una clara tendencia a recibir financiación de forma externa a la empresa. Aun así, los datos se encuentran bastante dispersos, por lo que muestra que no existe una uniformidad de criterio sobre si recibir o no financiación a través del endeudamiento.

Es importante destacar dos puntos en particular, el primero situado en torno al rango de 0.4 - 0.6 donde aproximadamente una cuarta parte de los valores se encuentra en ese rango, mostrando de esta manera que una de cada cuatro empresas opta por tener un endeudamiento equilibrado. El segundo punto se encuentra en torno al 0.8 - 0.9 donde se concentran en torno a un tercio de los valores de totales, dando a entender que hay un amplio sector del IBEX 35 que a lo largo de estos últimos diez años ha considerado el hiper endeudamiento como su forma habitual de financiarse.

Por último, en cuanto al análisis sectorial y en referencia al último punto mencionado, es el sector financiero el que más ha decidido endeudarse en los últimos años estando la totalidad de sus valores en el rango anteriormente mencionado. El otro sector claramente marcado es el de la Construcción, donde la mayoría de sus valores se encuentran en los rangos intermedios, mostrando una clara apuesta por el equilibrio entre la financiación interna y externa. En lo que respecta al resto de sectores, no hay una distribución clara en un rango determinado.

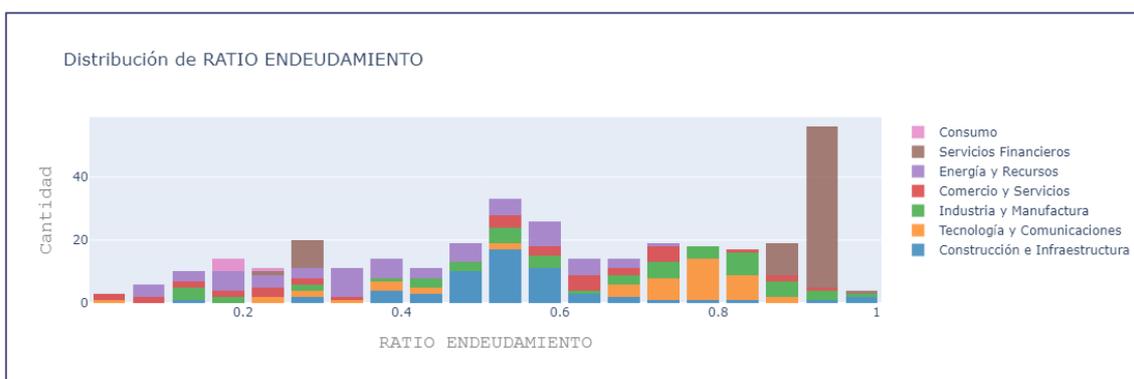


Figura 10. Distribución Ratio Endeudamiento

MARGEN

En relación con la variable margen, podemos ver en la Figura 11, que hay una distribución bastante simétrica, con cierto sesgo positivo y con algunos valores anómalos. Esto nos muestra que, pese a que la gran mayoría de las empresas a lo largo de estos años hayan tenido un margen sobre beneficios positivo, no ha habido una uniformidad dentro de esos márgenes positivos

También es importante destacar la presencia de valores anómalos, tanto positivos como negativos. Los valores negativos pertenecen principalmente al sector de Industria y Manufactura, lo que puede indicar o bien un año negativo para el sector o una racha de varios años negativos para una o varias empresas. Por otro lado, los valores anómalos positivos, no presentan ningún tipo de patrón respecto al sector, por lo que se trata de momentos específicos en los que las empresas tuvieron beneficios por encima de lo habitual.

Por último, respecto a los sectores, cabe destacar que tanto el sector financiero como el energético han sido de los sectores que siempre han mantenido un margen positivo a lo largo de los años. Pese a haber sufrido momentos inciertos a lo largo de los años, han seguido funcionando con aparente normalidad. Esto da a entender que son altamente relevantes debido a que otros sectores generan dependencia de ellos.



Figura 11. Distribución Margen

EFECTO Y PRESIÓN FISCAL

Las variables efecto y presión fiscal son otras variables cuya distribución nos da una información muy relevante, en este caso cómo de afectadas se ven las empresas por las tasas impositivas.

En cuanto al efecto fiscal (ver Figura 12), en primer lugar, recordemos que esta variable muestra qué parte de los beneficios antes de impuestos son transformados realmente en beneficios netos y qué parte no. Podemos observar que la mayoría de los valores se encuentran en torno al valor 1 o superiores, mostrando de esta manera que el 100% de los beneficios acabaron siendo beneficios netos, aumentando incluso en algunos casos a causa de que la hacienda pública le devolvía dinero. Observamos también una serie de valores anómalos, donde ese aumento de beneficios a coste de los impuestos fue sustancial.

En contrapartida, tenemos la variable complementaria presión fiscal (ver Figura 13), donde podemos observar efectivamente cómo la mayoría de las empresas no han sufrido la presión fiscal correspondiente en términos de impuestos sobre sociedades. Dicha distribución debería tener la mayoría de sus datos en torno al 25%, que corresponde con el tipo de gravamen vigente en España.

Estos indicadores se centran en cómo afecta el Impuesto Sobre Sociedades a estas empresas, no hacen un cómputo global de la presión fiscal total soportada por la empresa. Por ello, podemos concluir que el hecho de que casi ninguna empresa, independientemente del sector al que pertenezca, llega a ese 25% estipulado como tipo fijo a las sociedades, se debe principalmente por la internacionalización de las empresas del IBEX 35. Esto se debe a que la gran mayoría de ellas, operan en el extranjero, viéndose obligadas así a tributar la parte proporcional de los beneficios que obtienen en un determinado país. Los beneficios mostrados en las cuentas auditadas son globales por sobre los impuestos que se les retienen que corresponden únicamente a la facturación en España.

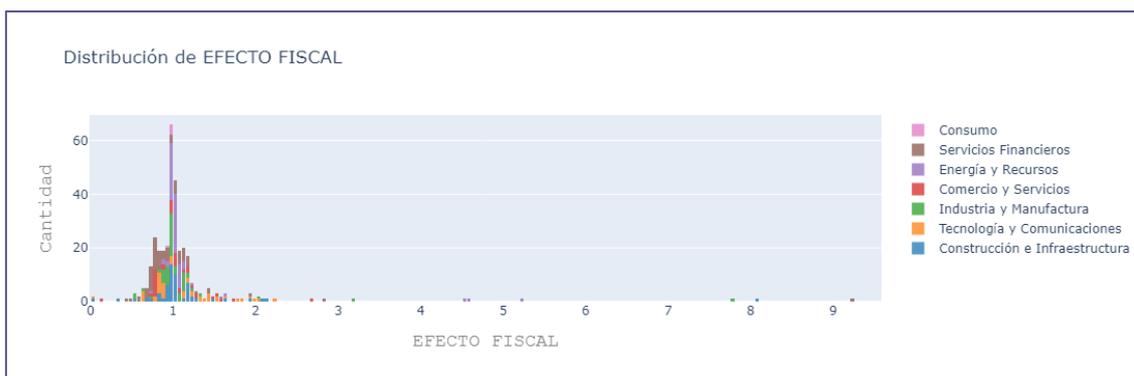


Figura 12. Distribución Efecto Fiscal

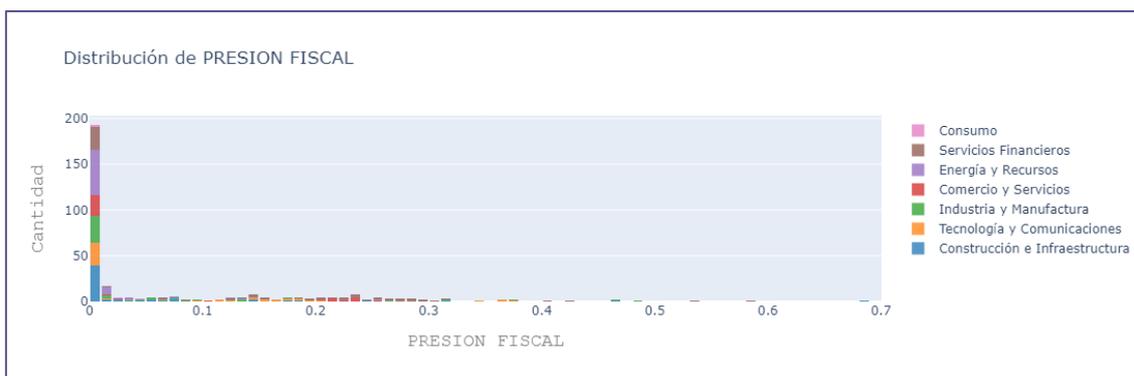


Figura 13. Distribución Presión Fiscal

ROA

Una de las variables que más información da sobre el funcionamiento real de las empresas es la ROA (ver Figura 14), presenta una ligera asimetría positiva, donde la gran mayoría de los valores se encuentran posteriores al cero, mostrando de esta manera que a lo largo de los últimos diez años las empresas de IBEX 35 han sabido transformar en beneficios sus activos, mostrando una gestión operativa óptima.

La existencia de algunos valores anómalos se origina, en gran medida, por periodos inusuales para algunas compañías, ya sea por una reducción en los activos junto a una producción constante o por un aumento en los beneficios derivado de una mayor productividad.

En cuanto al análisis sectorial, si bien no hay patrones claros definidos en según los sectores, podemos ver como el sector financiero es aquel que menos réditos ha obtenido en proporción con

la cantidad de activos empleados, mostrando una aparente improductividad generalizada dentro del sector.

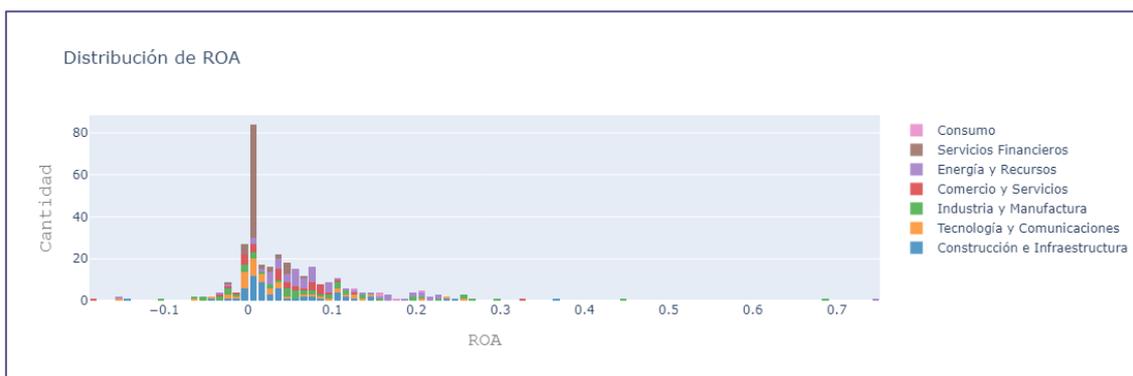


Figura 14. Distribución ROA

ROE

Por último, el ROE es la variable de referencia en términos de desempeño, que sirve para ejercer comparaciones con otros tipos de inversiones. La distribución del ROE expuesta en la Figura 15 presenta, al igual que la ROA, una asimetría positiva, donde la mayoría de los valores son superiores al cero. Una muestra de que el retorno que han sufrido los accionistas e inversores de las diversas empresas del IBEX 35 en los últimos diez años ha sido significativo.

Respecto a esto, y dado que el ROE se utiliza como ratio de comparación al involucrar los mismos elementos que en cualquier otro tipo de inversión, es decir, capital total invertido y rédito obtenido. Podemos observar que la rentabilidad de la gran mayoría de las empresas del IBEX 35 ha sido superior a cualquier otro tipo de inversión. Los fondos monetarios o cualquier inversión de renta fija a lo largo de estos años han estado proporcionando una rentabilidad de entre el 3% al 5% en las épocas con tipos de interés más altos. Por otro lado, la mayor cantidad de las empresas del IBEX 35 se sitúan entre el 5% o superior, convirtiendo la participación directa en la mayoría de las empresas de la bolsa española en los últimos diez años en una inversión altamente rentable y atractiva.

Por último, todos los sectores han mostrado una rentabilidad negativa a lo largo de estos años, no estableciendo un patrón claro.

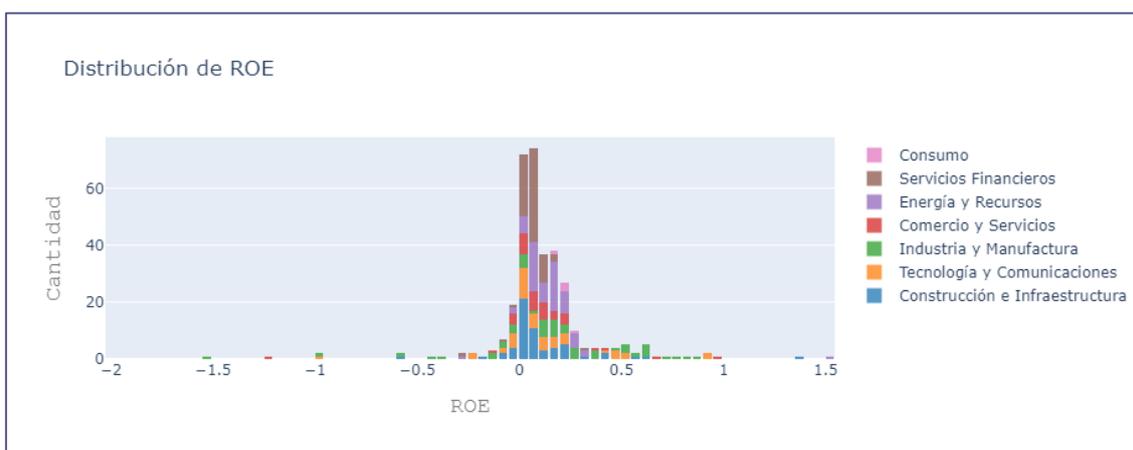


Figura 15. Distribución ROE

5.5.2. Distribuciones de variables categóricas

En segundo lugar, pasaremos a analizar las distribuciones de las variables categóricas introducidas en el *Dataset* que dependen directamente de las empresas. Analizaremos las diferentes variables binarias, la estructura del coste y el ciclo de vida, así como la distribución dentro de los diferentes sectores conforme la rentabilidad obtenida.

DIVIDENDO, BENEFICIO A RESERVAS Y PERSONAL

En relación con el reparto de dividendo (ver Figura 16), podemos ver como la tónica general dentro del índice a lo largo de estos años ha sido realizar el reparto. En torno al 90% de las ocasiones, las diferentes empresas han destinado parte de sus beneficios, o en caso de ser negativos, parte de sus reservas para remunerar a sus accionistas. Segregando, obtenemos que los sectores Financieros y de Construcción han sido los que más dividendos han repartido, independientemente de las duras situaciones vividas a lo largo de los años.

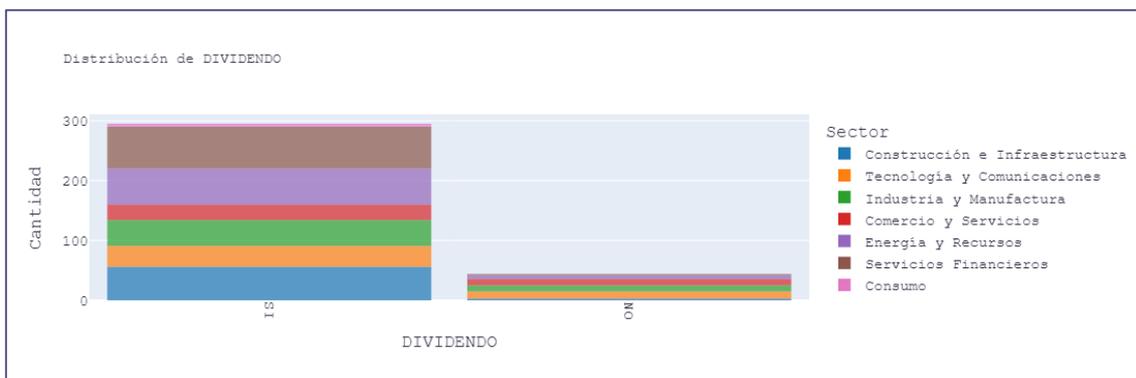


Figura 16. Distribución Dividendo

En relación con los beneficios obtenidos (ver Figura 17), las veces que estos se han destinado a reservas, parcial o totalmente ha sido considerablemente superior a las veces en las que no. Podemos ver cómo las empresas del IBEX 35 optan también por guardar parte de los réditos obtenidos en las reservas, o bien con la finalidad de tener un colchón el cual poder utilizar para comprar activos en un futuro, o bien para poder seguir pagando dividendos a los accionistas pese a no tener beneficios en un periodo determinado. Los diferentes sectores muestran un comportamiento similar a lo largo de los años en cuanto a distribución de los beneficios a las reservas de la empresa.

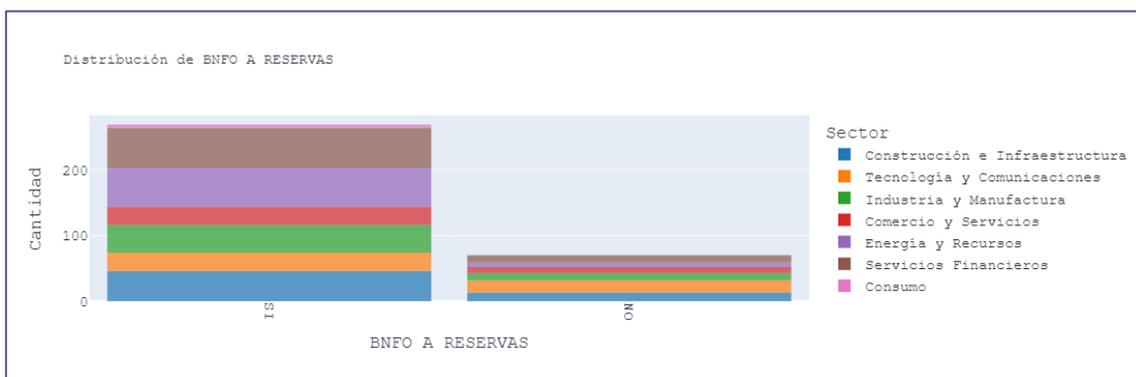


Figura 17. Distribución Beneficio a Reservas

Por último, dentro de las variables binarias relacionadas con la empresa nos encontramos con la decisión de aumentar o no el gasto en persona de un año a otro, mostrado en la Figura 18, esto puede ser a través de nuevas contrataciones o por la mejora de las condiciones laborales de los trabajadores actuales. Podemos destacar cómo, si bien la tendencia ha sido aumentar dicho gasto, está mucho más ajustado con respecto a las otras dos variables analizadas, donde en este caso, en torno al 65% de las ocasiones se ha decidido optar por mejorar la productividad en términos de capital humano. Esto nos lleva a entender que el crecimiento que ha experimentado el índice a lo largo de estos años no siempre se ha debido a que la productividad haya aumentado gracias a un incentivo a los trabajadores o por el aumento de la mano de obra. Factores como la constante digitalización y automatización a lo largo de los años han podido ser clave para experimentar aumentos sustanciales en la productividad sin aumentar el gasto en personal. Esto es una tónica general a lo largo de todos los sectores involucrados.

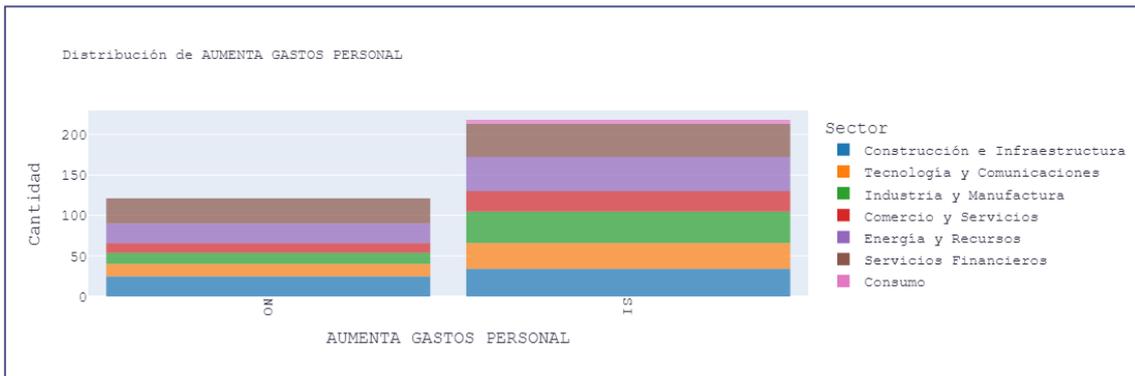


Figura 18. Distribución Gasto Personal

ESTRUCTURA DE COSTES

En cuanto a la estructura de costes (ver Figura 19), podemos ver como la gran mayoría de las empresas tienen una estructura de costes fija, alejándose de esta manera de la gestión de aprovisionamientos relacionada con la producción. Esto sucede en parte debido a que muchas de las empresas analizadas tienen filiales o empresas del grupo a través de las cuales pueden realizar la compra de los materiales necesarios para la producción, por ello, esto no se ve reflejado en las cuentas anuales de la empresa dominante. Los sectores que tiene parcialmente una estructura de costes variables son los sectores industrial y energético.

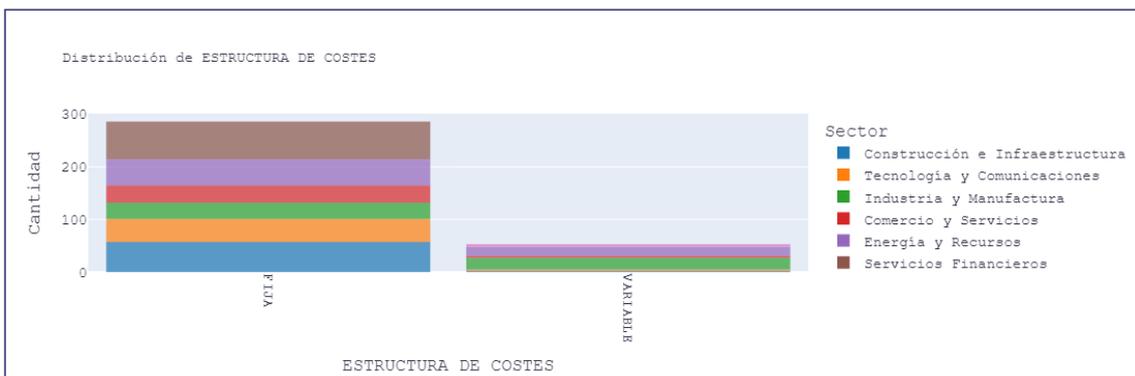


Figura 19. Distribución Estructura de Costes

CICLO DE VIDA

En lo que se refiere al ciclo de vida de la empresa, podemos observar en la Figura 20 como la mayoría de los valores se encuentran en las etapas de Crecimiento, Madurez y Declive. Estas tres etapas tienen en común que están generando flujos de efectivo a partir de sus operaciones, en otras palabras, están produciendo de manera adecuada y su proceso es rentable. Aquello que les diferencia es respecto a la inversión y la financiación, es decir, si están recibiendo financiación en ese momento o simplemente están pagando por la financiación pasado y si están realizando inversiones o deshaciéndolas. Esto muestra que la mayoría de las empresas que conforman el IBEX 35 son empresas asentadas que son capaces de sacar rentabilidad de sus procesos productivos y por tanto, generar rédito a los accionistas.

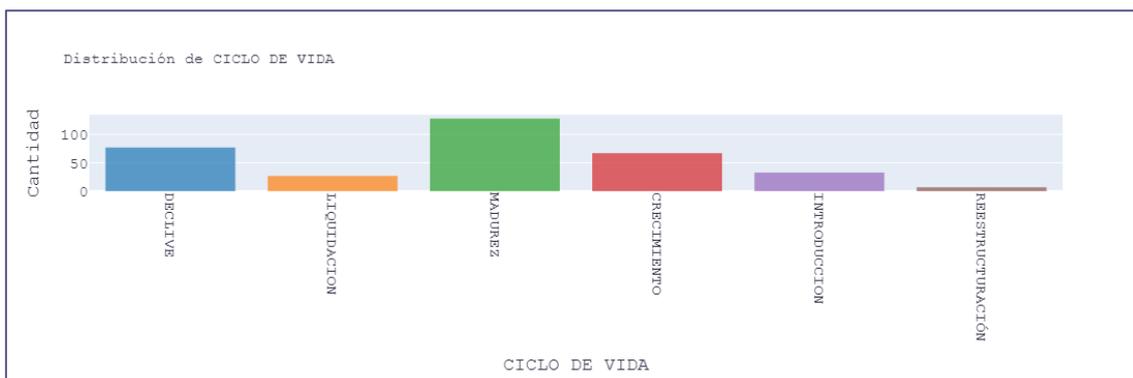


Figura 20. Distribución Ciclo de Vida

SECTOR

A continuación, analizamos la relación entre el sector al que pertenecen las empresas y la rentabilidad que estas ofrecen, utilizando el ROE como métrica principal (ver Figura 21). Para ello, hemos dividido a las empresas en dos grupos, basándonos en la mediana del ROE: aquellas con un ROE superior a la mediana y aquellas con un ROE inferior. Los resultados muestran que los sectores de energía e industrial destacan por tener una mayor concentración de empresas con alta rentabilidad, lo que indica que estos sectores tienden a ser más rentables en comparación con otros. Por otro lado, sectores como el financiero y el de construcción presentan una mayor proporción de empresas con rentabilidades más bajas o incluso negativas. Esta tendencia podría estar relacionada con los desafíos específicos que enfrentan estos sectores, lo que impacta negativamente en su capacidad para generar valor para los accionistas.

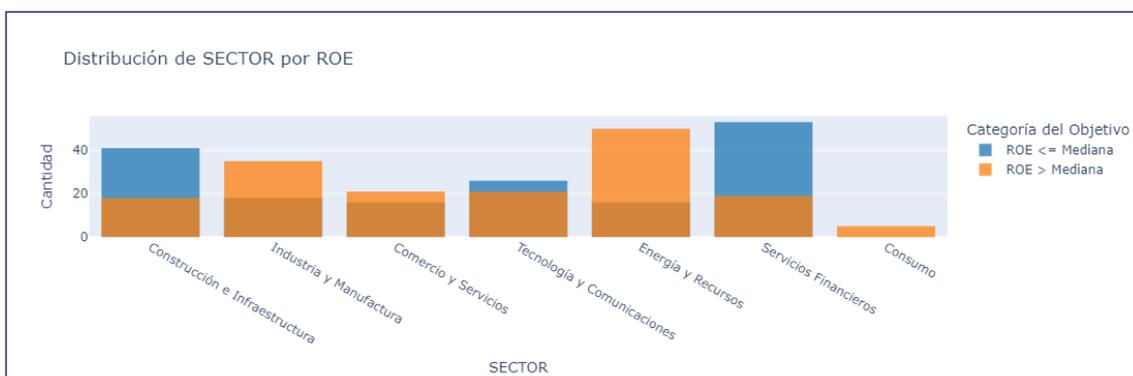


Figura 21. Distribución Sector según ROE

5.5.3. Análisis de correlación

En el análisis de correlación con el ROE (ver Figura 22), hemos identificado aquellas variables que muestran una mayor correlación, tanto positiva como negativa. Este tipo de análisis nos permite entender cuáles son los factores que impulsan o limitan la rentabilidad de los accionistas.

En primer lugar, pasemos a analizar aquellas que tienen una correlación positiva, es decir, aquellas que favorecen que la rentabilidad sea alta. Estas son:

- **Margen:** Esto indica que las empresas que logran mantener o aumentar sus márgenes de beneficios tienden a ofrecer una mayor rentabilidad a sus accionistas, mostrando una gestión de los costes tanto operativos como financieros óptima.
- **Rotación de activos:** La rotación de activos está relacionada positivamente, nos lleva a pensar que una gestión eficiente de los activos para generar ventas hace que el retorno final de los accionistas se vea incrementado.
- **Margen operativo:** Un poco en la misma línea que el margen general, pero esta vez mostrando que la importancia de gestionar adecuadamente los costes operativos es crucial en la obtención de rentabilidad.
- **Variación del precio de las acciones:** Esto muestra que un aumento en la rentabilidad de las empresas se ve reflejado claramente en la confianza que depositan los accionistas en esta, haciendo subir de esta manera el precio de sus acciones.

Las relacionadas negativamente son:

- **Multiplicador de capital:** La relación negativa entre la proporción de deuda sobre el los fondos propios y la rentabilidad muestra que un alto nivel de apalancamiento frena la obtención final de beneficios, esto se puede deber o bien a una mala utilización de esa deuda o a un coste excesivo de la misma.
- **Ratio de endeudamiento:** En relación con la anterior, pero esta vez sobre la totalidad del patrimonio, muestra de nuevo que la deuda empleada no es de la mejor calidad en términos de coste o que la utilización de esta no se está viendo traducida en términos de productividad.
- **Tasa de paro:** Esto se debe a que una alta tasa de desempleo generalmente indica una economía débil y descapitalizada, lo que reduce la demanda general de bienes y servicios reduciendo de esta manera la rentabilidad de las empresas al limitar su capacidad de vender.

Por último, hay que recalcar que variables como la inflación, los derechos de propiedad o la presión fiscal, son factores cuya variación no afecta significativamente en la obtención de rédito por parte de los accionistas.

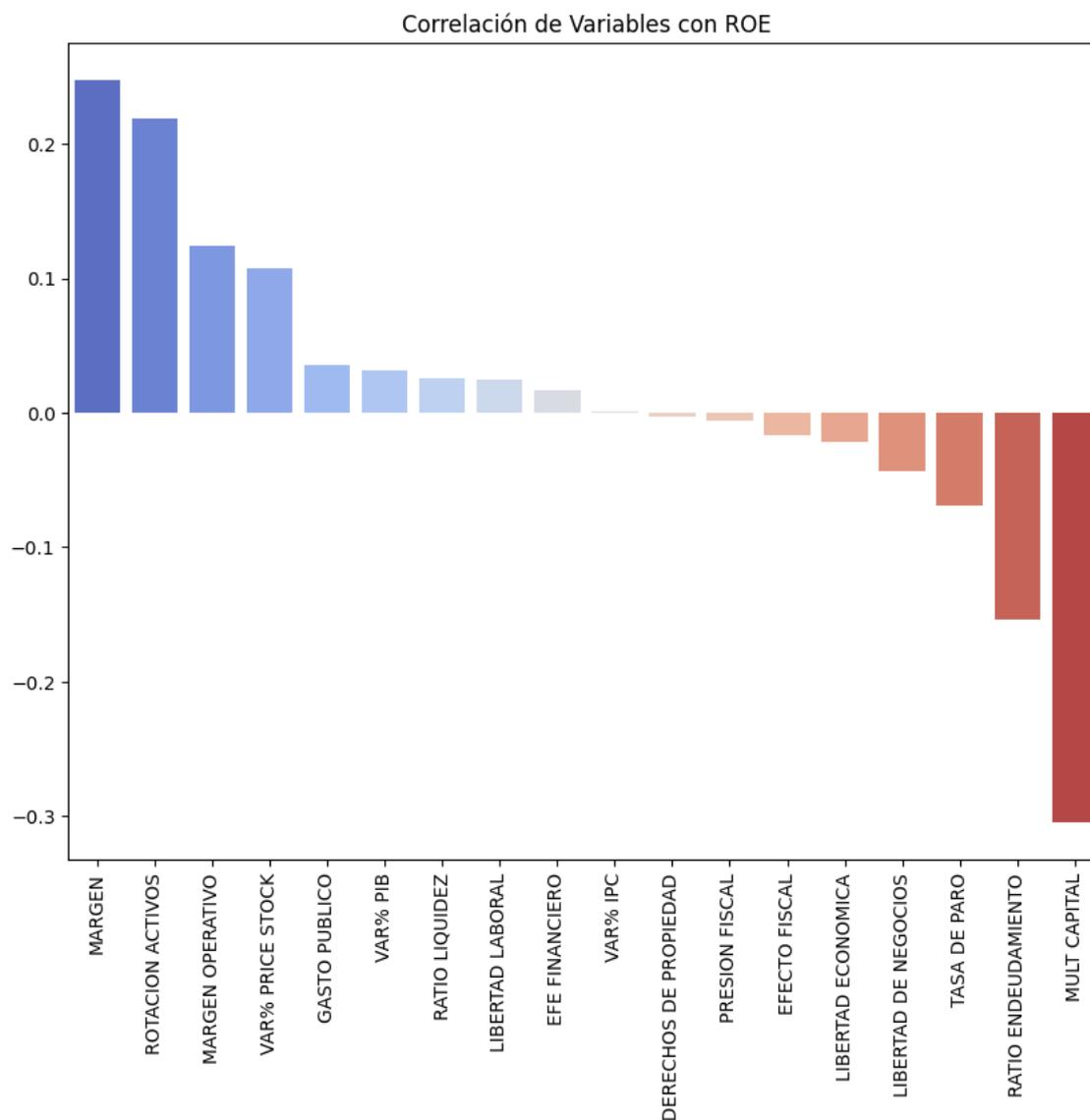


Figura 22. Análisis correlación ROE

5.5.4. Análisis de series temporales

En esta sección, realizaremos un análisis de las series temporales para algunas de las variables que cubren todas las áreas de la empresa y dan una información más clara, observando cómo han evolucionado a lo largo de los años. Este análisis nos permitirá identificar tendencias, ciclos y patrones que pueden influir en el desempeño de las empresas. Adoptaremos un enfoque tanto global como desglosado por sectores, para comprender de esta manera como afecta el paso del tiempo a cada sector de la economía.

ROA

En primer lugar, analizando la evolución de la ROA, podemos observar en la Figura 23 cómo en términos globales en el índice había una tendencia creciente hasta alcanzar su máximo en 2020, donde a causa de la pandemia sufrió una caída que le hizo perder en torno al 50% de su valor. Tras esa caída, vino un periodo de dos años de recuperación para una siguiente caída causada principalmente por las constantes tensiones a nivel geopolítico, el aumento de costes de la materia prima y el aumento del coste de la deuda a causa del aumento de los tipos de interés.

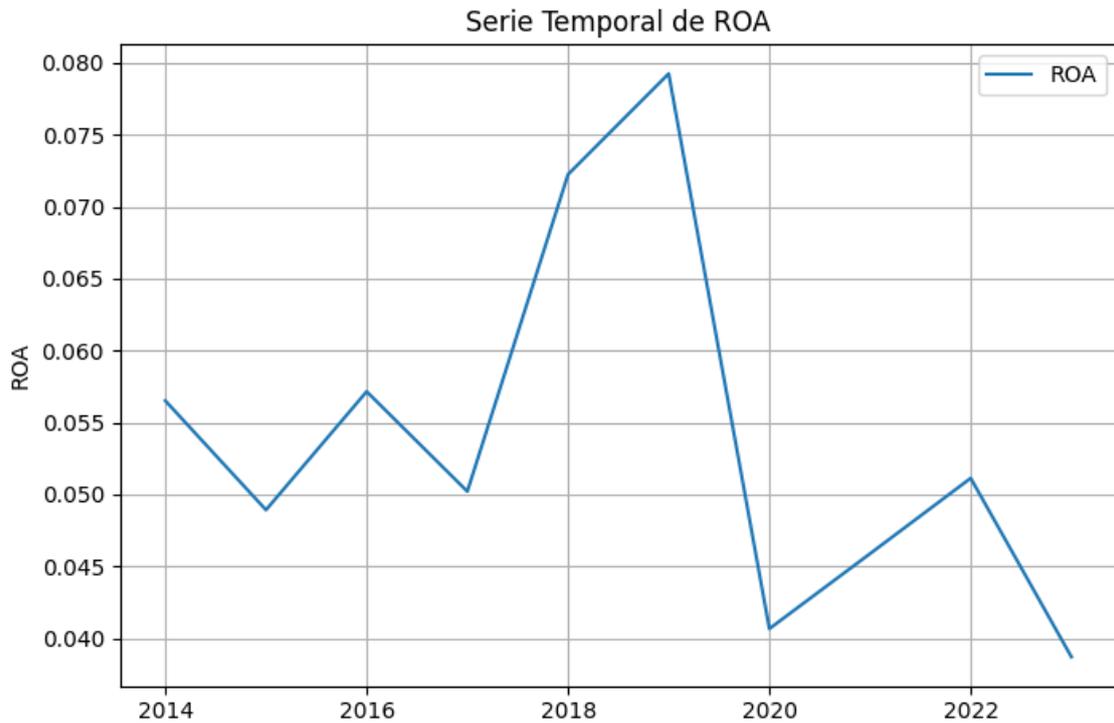


Figura 23. Evolución temporal ROA

En cuanto al análisis por sectores (ver Figura 24), podemos ver cómo los sectores que más sufrieron esta caída tras la pandemia fueron el sector tecnológico y el sector servicios, llegando a tener ROA negativo durante el año 2020. Por otro lado, el sector financiero, el energético y el de la construcción no se vieron afectados prácticamente por el parón global, ya que no son sectores dependientes de la presencia de las personas. Tras la caída, vemos un repunte de los sectores más afectados durante la pandemia volviendo a sus valores habituales.

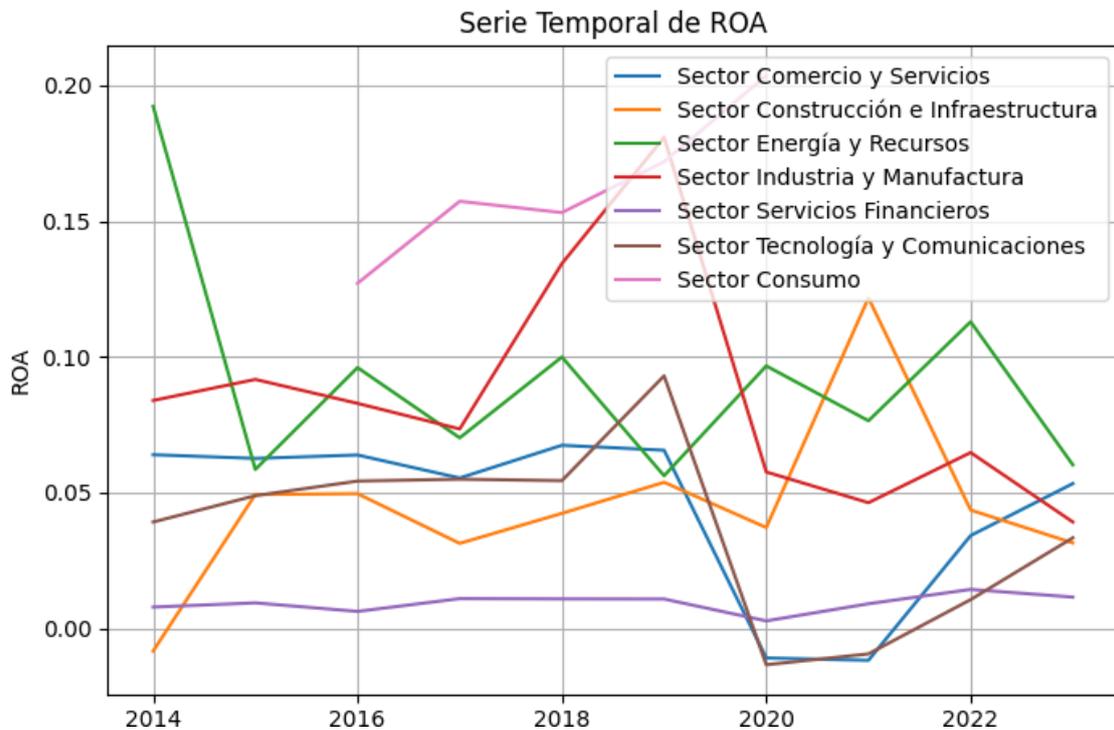


Figura 24. Evolución temporal ROA por sectores



ROE

En lo que se refiere al ROE (ver Figura 25), podemos ver un patrón similar al mencionado anteriormente en la ROA, con varias apreciaciones para tener en cuenta.

En primer lugar, hay que destacar que pese a existir una caída de casi el 50% en el promedio del IBEX 35, las rentabilidades mostradas son positivas, mostrando valores en torno al 7% en medio de una crisis a nivel global. En segundo lugar, lo que podemos ver, es que a diferencia de la ROA el ROE sufrió una fuerte caída en el año 2014 llegando a tener rentabilidades negativas en términos generales.

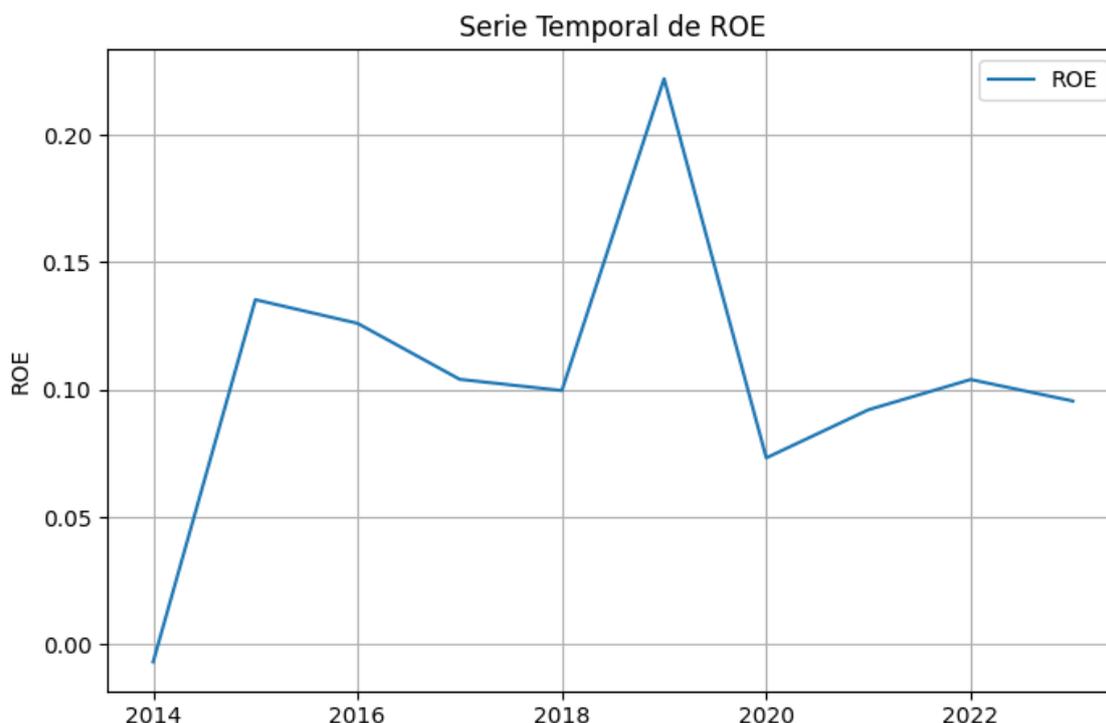


Figura 25. Evolución temporal ROE

Analizando esta vez por sectores, nos es más fácil contestar a las dudas que se nos planteaban cuando vemos la gráfica general.

Lo primero que cabe mencionar, respecto a los años del COVID-19 es que sigue el mismo patrón que la ROA, son los sectores de servicios y tecnológico los más afectados, mientras que el sector industrial y el energético son los que amortiguan la caída en términos generales mostrando rentabilidades cercanas al 10%.

La otra duda que surgía era con respecto al año 2014, y a través de la segmentación por sectores mostrada en la Figura 26, podemos ver como pese a que los sectores energético, tecnológico e industrial muestren unas rentabilidades altísimas, la construcción y el comercio se hundieron totalmente, mostrando pérdidas de entorno al 60% y al 25% respectivamente. Estas fuertes caídas hacen que el cómputo global del IBEX 35 en ese año sea negativo.

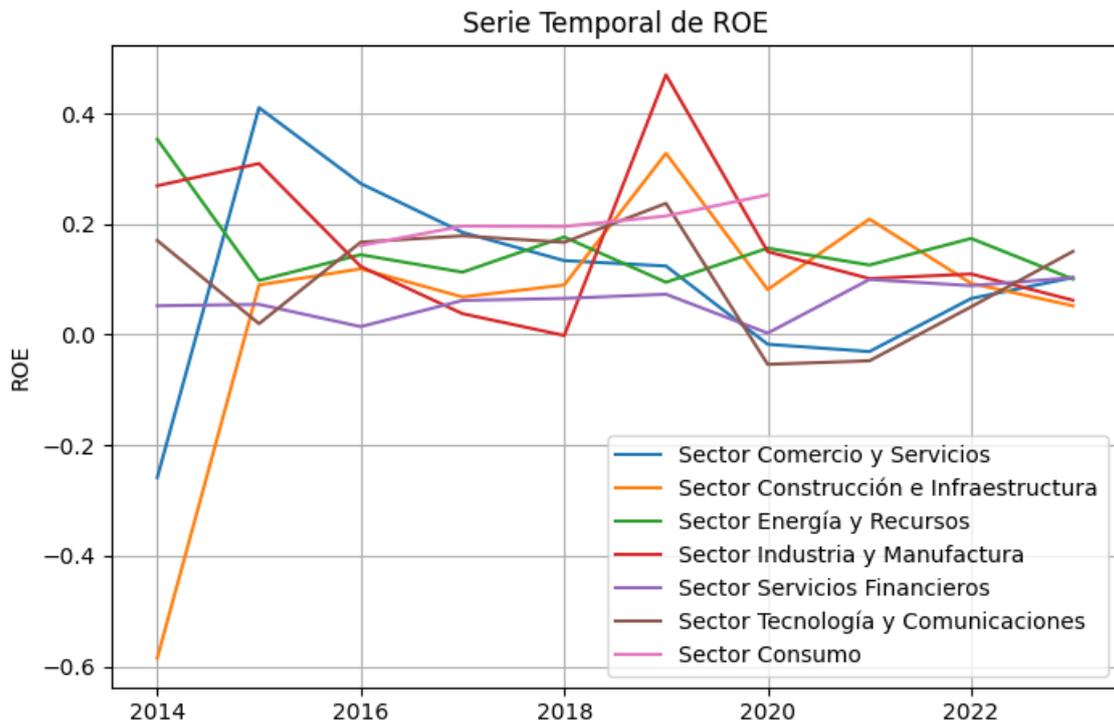


Figura 26. Evolución temporal ROE por sectores

MARGEN Y ROTACIÓN DE ACTIVOS

Analizando ahora el margen y la rotación de activos podemos observar varias cosas a nivel general del índice (ver Figura 27).

En primer lugar, podemos observar cómo de nuevo la crisis afecta de manera significativa a ambas variables, registrando en el año 2020 el punto más bajo en los últimos diez años. Esto muestra cómo la capacidad de gestión operativa y la eficiencia se vieron mermadas por las constantes trabas generadas por una cuarentena a nivel global.

Por otro lado, cabe mencionar que la recuperación del margen es mucho más rápida que la de la rotación de los activos. Existen dos lecturas posibles, o bien la gestión del aumento de costes por la crisis fue mucho más eficiente que la vuelta a la normalidad en términos de producción o el aumento de rentabilidad se debe principalmente en un cambio de estrategia comercial, optando por vender más cara en vez de vender más productos o servicios.

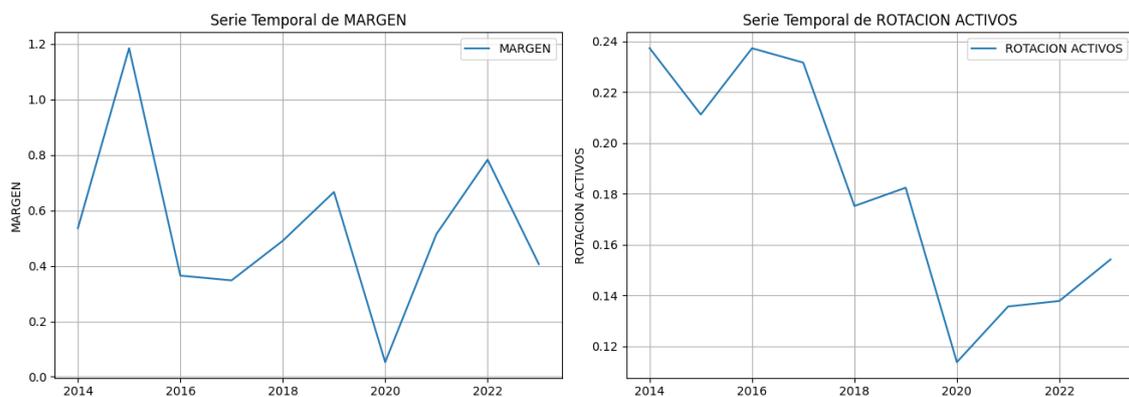


Figura 27. Evolución temporal Margen y Rotación de activos

A raíz del análisis por sectores mostrado en la Figura 28, podemos ver que la caída en términos de márgenes se debe principalmente por una fuerte caída del sector servicios, siendo este otra vez el más afectado por las restricciones de movilidad impuestas durante ese año. Por otro lado, vemos que la recuperación de la rotación de activos únicamente fue efectiva en el sector industrial, por el contrario, los demás sectores no han conseguido alcanzar la misma productividad que en los años posteriores a la pandemia.

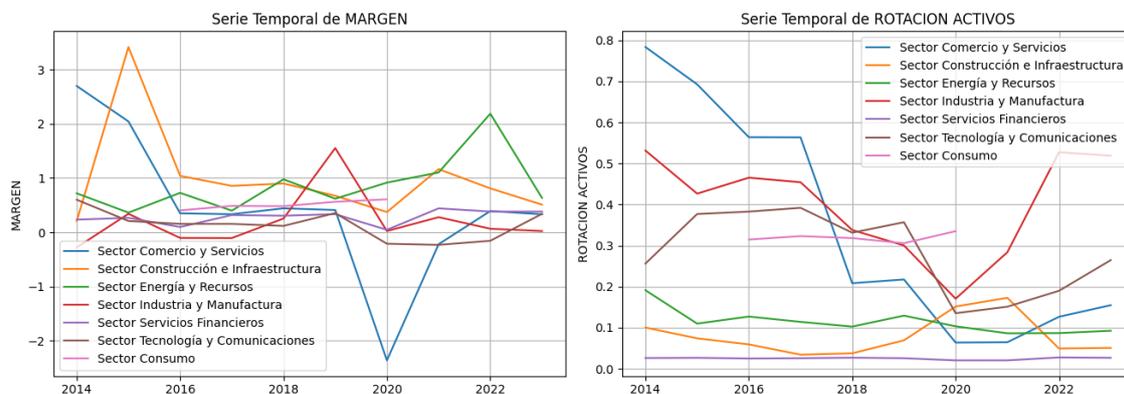


Figura 28. Evolución temporal Margen y Rotación de activos por sectores

RATIO DE ENDEUDAMIENTO

Por último, vemos una clara tendencia decreciente en la deuda desde que alcanzara su punto máximo en 2015 (ver Figura 28), alcanzando su punto mínimo en el año 2021. El punto más bajo no se produce un año después de la pandemia por casualidad, la fuerte incertidumbre alrededor de las empresas y de la situación general hace que la concesión de préstamos por parte de los bancos tuviese un coste superior, así como la subida de tipos de interés anunciada por parte del Banco Europeo, convirtieron la deuda en una opción menos atractiva.

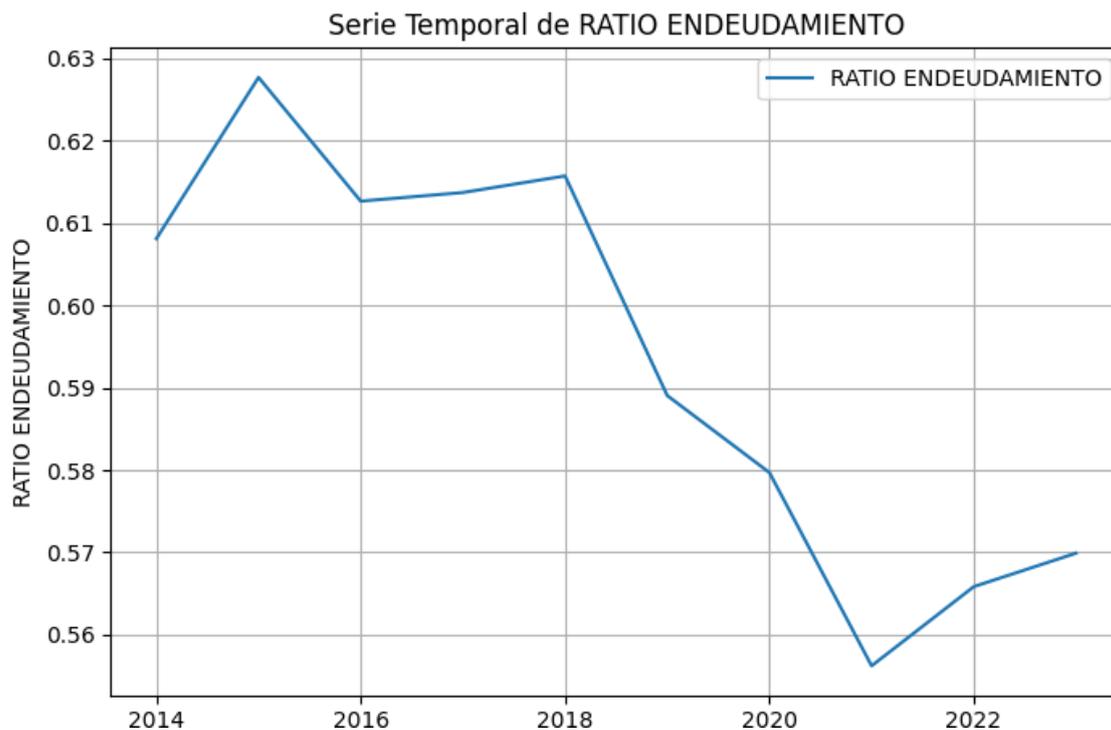


Figura 29. Evolución temporal Ratio endeudamiento

En el análisis por sectores (ver Figura 30), vemos como el sector financiero es el más endeudado a lo largo de todos los años con apenas variaciones dentro de su estructura de deuda. Por otro lado, si bien hay sectores que mantienen más o menos constantes sus niveles de deuda, el sector industrial y el sector servicios reducen de forma considerable su nivel de deuda.

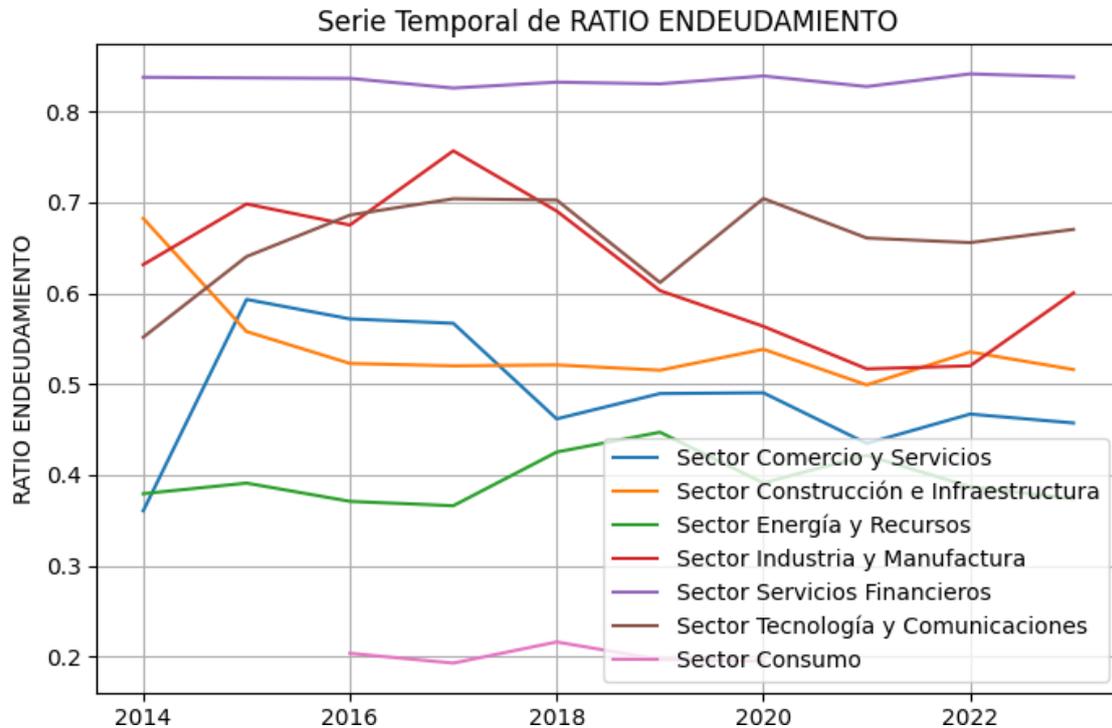


Figura 30. Evolución temporal Ratio endeudamiento por sectores

5.5.5. Análisis de valores anómalos

En esta sección, realizaremos un análisis de los valores anómalos para identificar y evaluar datos que se desvían significativamente de las tendencias generales mostradas por el conjunto de datos. Estos valores anómalos, pueden darse debido a variaciones extremas de los datos o eventos extraordinarios que impactan de manera decisiva en las variables.

Lo que podemos observar es que, en términos generales, todas las variables económicas que dependen directamente de las empresas tienen un número considerable de valores anómalos, a causa de eventos puntuales que generen unos valores fuera de la tendencia o bien por errores a la hora de realizar determinados cálculos (ver Figuras de la 31 a la 42).

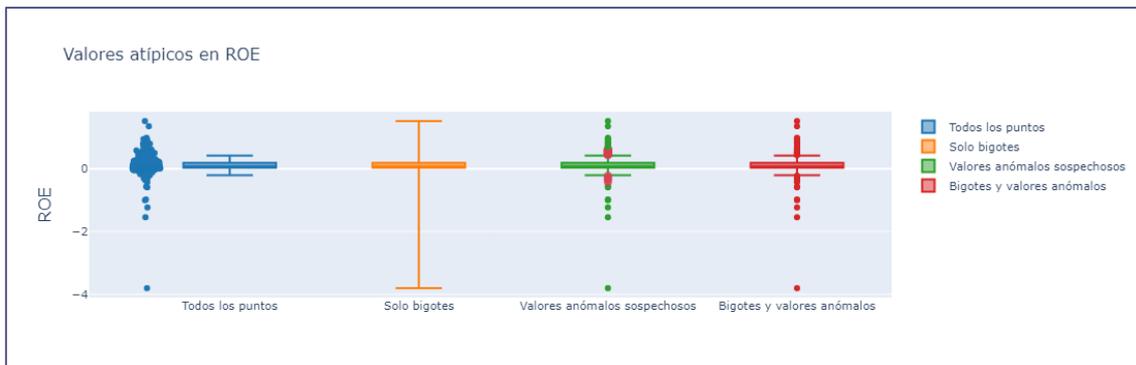


Figura 31. Valores anómalos ROE



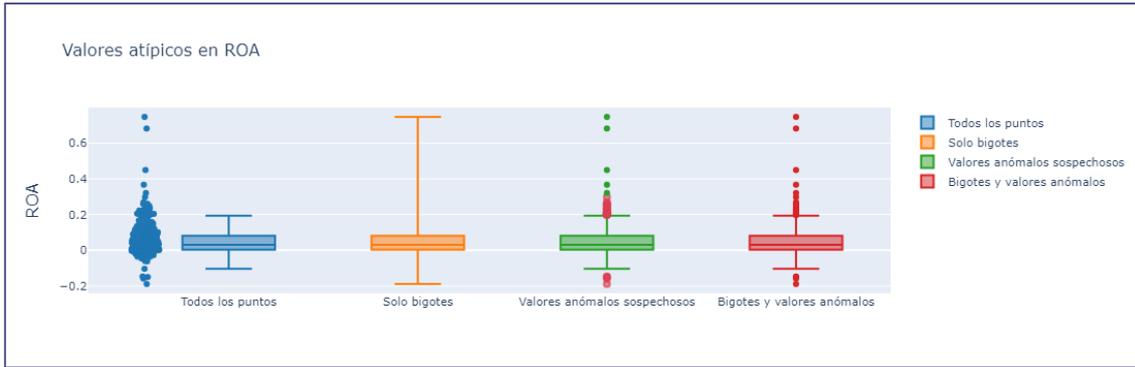


Figura 32. Valores anómalos ROA

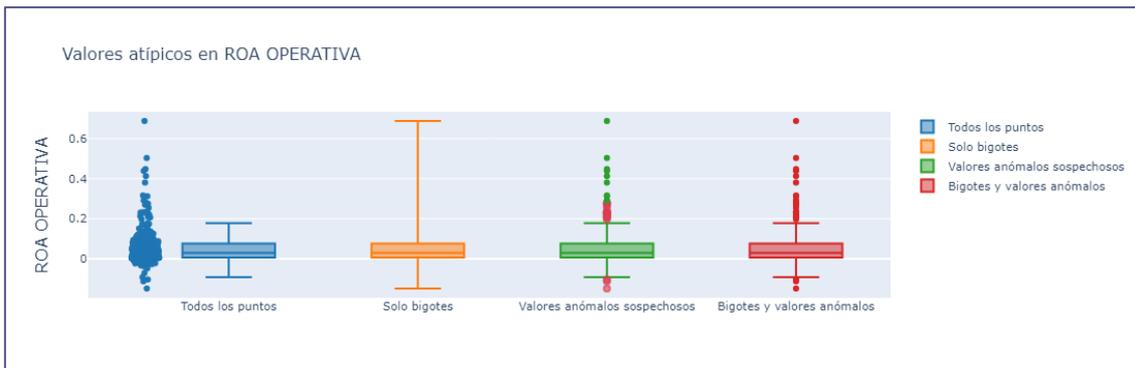


Figura 33. Valores anómalos ROA operativa

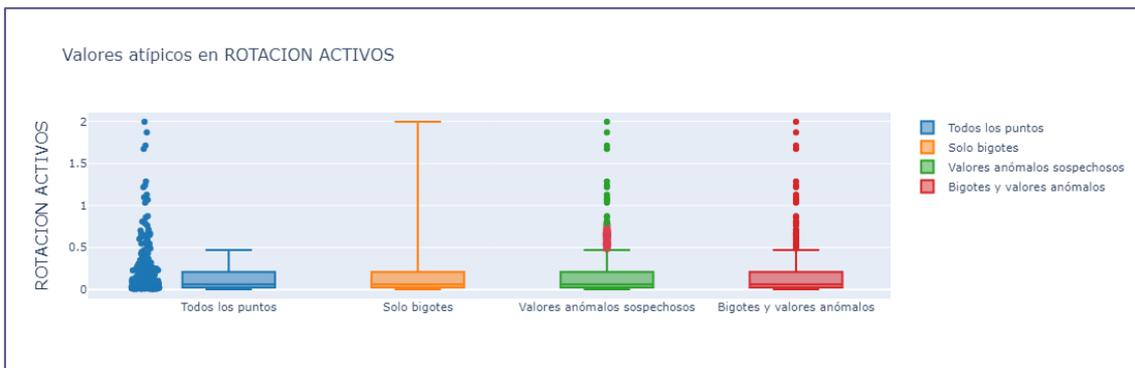


Figura 34. Valores anómalos Rotación activos

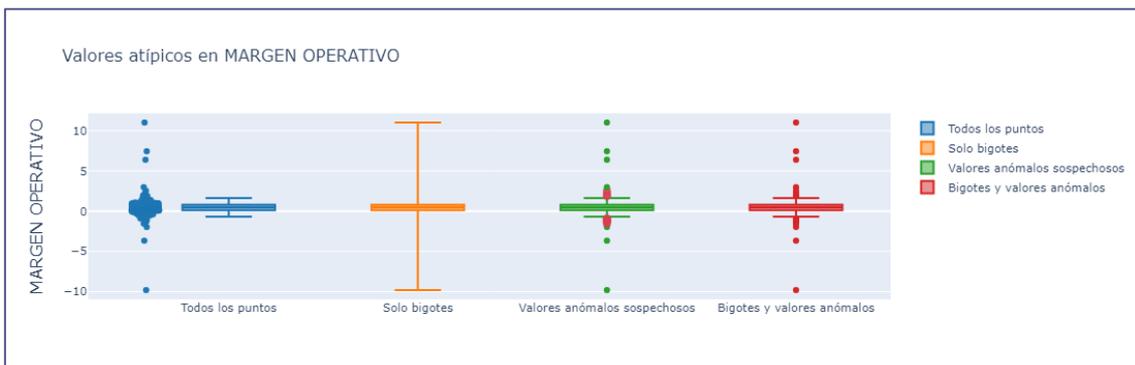


Figura 35. Valores anómalos Margen Operativo

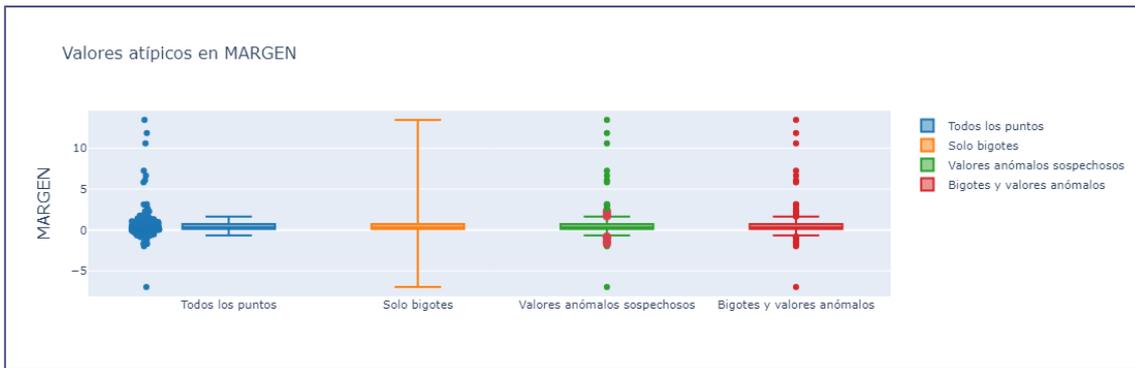


Figura 36. Valores anómalos Margen

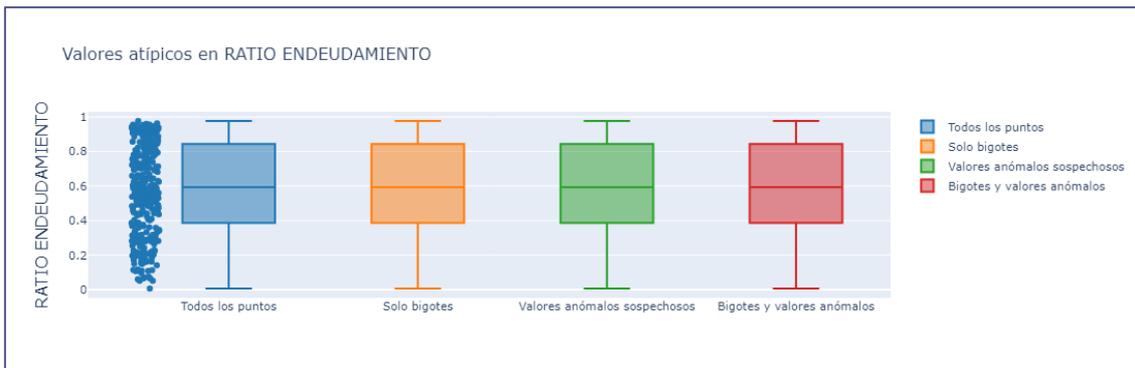


Figura 37. Valores anómalos Ratio endeudamiento

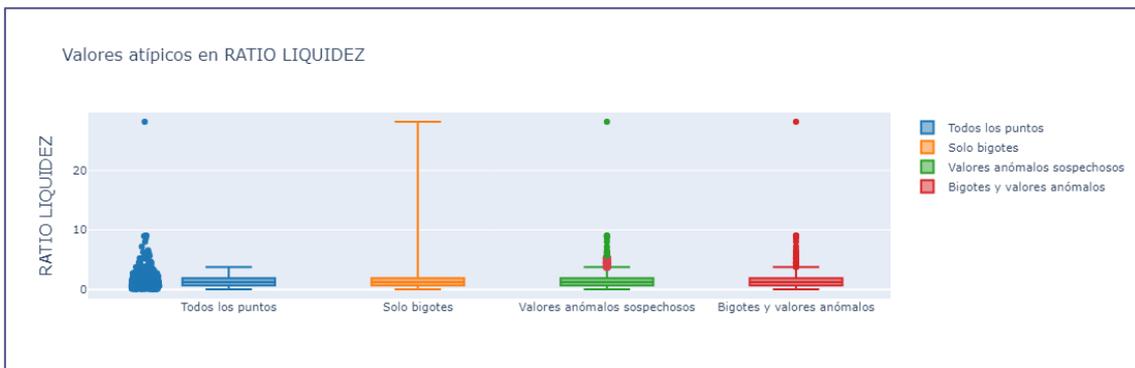


Figura 38. Valores anómalos Ratio liquidez

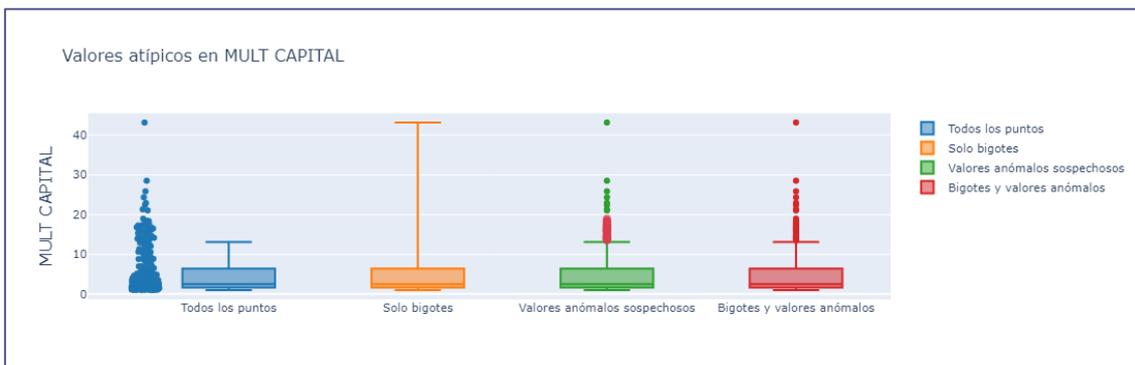


Figura 39. Valores anómalos Multiplicador capital



Figura 40. Valores anómalos Efecto financiero

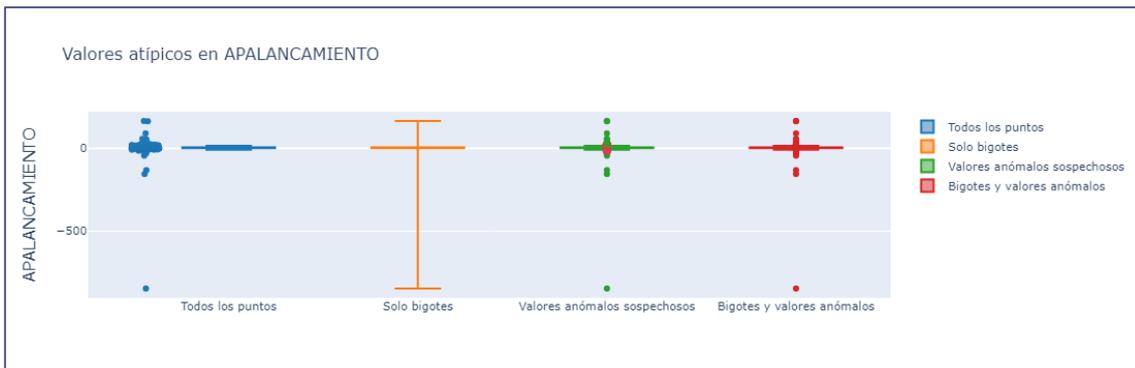


Figura 41. Valores anómalos Apalancamiento

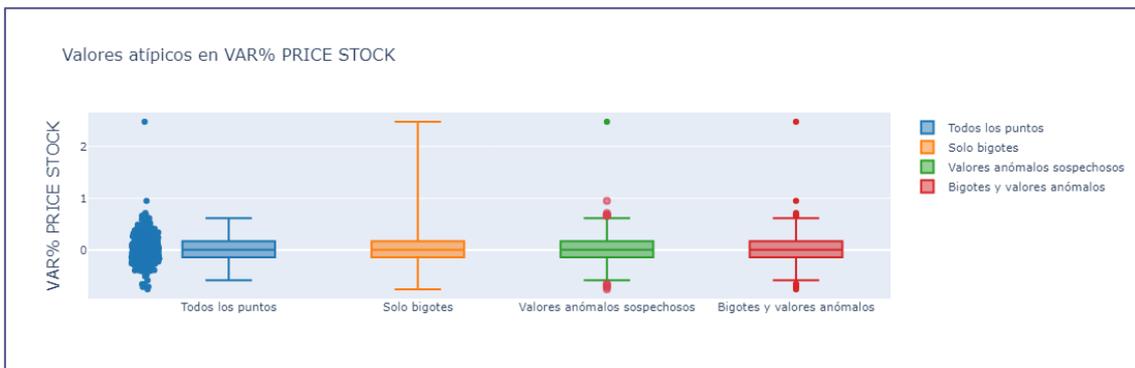


Figura 42. Valores anómalos Variación precio de las acciones

En conclusión, dado la cantidad de valores anómalos, es necesario un tratamiento de estos para facilitar de esta manera el análisis predictivo.

5.6. Preprocesado

En esta parte, abordaremos el preprocesado de los datos, un paso fundamental para preparar el conjunto de datos para la aplicación de cualquier modelo predictivo. Esto no asegura que los datos tengan la forma adecuada para que los algoritmos aplicados tengan la precisión adecuada.

Para ello, importamos las siguientes librerías de Python que utilizaremos a lo largo de todo el proceso (ver Figura 43).

```
from sklearn.preprocessing import OrdinalEncoder
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
from sklearn.preprocessing import Normalizer
```

Figura 43. Librerías Python para el preprocesado

5.6.1. Normalización

El primero de los pasos por realizar es la normalización de las variables numéricas a través de la clase *Normalizer* (ver Figura 44). Esta clase se utiliza para escalar los valores de las variables numéricas de manera que se encuentren en un rango específico, generalmente cero y uno. Esto lo realizamos con la finalidad de mejorar el rendimiento del análisis, asegurando de esta manera una uniformidad en las escalas.

```
def normalizar_variables(df, columnas_numericas, norm='l2'):
    normalizer = Normalizer(norm=norm)

    df[columnas_numericas] = normalizer.fit_transform(df[columnas_numericas])

columnas_numericas = [
    'ROE', 'ROA', 'ROA OPERATIVA', 'ROTACION ACTIVOS', 'MARGEN OPERATIVO', 'MARGEN',
    'RATIO ENDEUDAMIENTO', 'RATIO LIQUIDEZ', 'MULT CAPITAL', 'EFE FINANCIERO', 'APALANCAMIENTO',
    'VAR% PIB', 'VAR% IPC', 'VAR% PRICE STOCK'
]

normalizar_variables(df, columnas_numericas)
```

Figura 44. Normalización de las variables numéricas

5.6.2. Codificación binaria

Lo segundo que vamos a realizar es la codificación binaria de las variables categóricas. Es un método sencillo que consiste en transformar aquellas variables que contengan únicamente dos valores posibles en cero y uno para así garantizar su compatibilidad con los modelos predictivos (ver Figura 45).

```
columnas_binarias = [col for col in df.columns if df[col].isin(['SI', 'NO']).any()]
original_columns = df[columnas_binarias].copy()
def si_no_to_binary(value):
    if value == 'SI': return 1
    elif value == 'NO': return 0
    return value
df[columnas_binarias] = df[columnas_binarias].map(si_no_to_binary)
df['ESTRUCTURA DE COSTES'] = df['ESTRUCTURA DE COSTES'].map({'FIJA': 1, 'VARIABLE': 0})
df['BENEFICIO'] = df['BENEFICIO'].map({'POSITIVO': 1, 'NEGATIVO': 0})
df['ESTRUCTURA DE COSTES'] = df['ESTRUCTURA DE COSTES'].fillna(0).astype('int64')
```

Figura 45. Codificación binaria de las variables categóricas

5.6.3. Codificación One-Hot

Siguiendo con la transformación de las variables categóricas, vamos a transformar aquellas que tengan más de dos valores posibles. Para ello, vamos a aplicar el método *One-Hot* que convierte cada categoría en una columna separada, donde el valor de la columna es uno si dicha fila pertenece a esa categoría y cero en caso contrario (ver Figura 46).

```
df = pd.get_dummies(df, columns=['SECTOR', 'CICLO DE VIDA'], drop_first=True)
```

Figura 46. Codificación One-Hot de las variables categóricas

5.6.4. Manejo de valores anómalos

Finalizando el preprocesado, y debido a la necesidad mostrada en el 5.5.5 Análisis de valores anómalos, vamos a realizar un tratamiento de los valores anómalos a partir de tres aproximaciones. En primer lugar, probaremos a eliminar esos valores anómalos, en segundo lugar, probaremos a sustituirlos por la media de dicha columna y en último lugar, usaremos un modelo predictivo que asigna un valor calculado a través de un modelo a partir de los demás datos del *Dataset*. Tras la implementación de los tres métodos, los utilizaremos por separado en el modelado para decidir cuál de las tres aproximaciones da un mejor resultado predictivo.

Primero de todo, para identificar valores anómalos en un conjunto de datos, hemos decidido utilizar el método de *fences* propuesto por Tukey, el cual define los límites de los valores anómalos utilizando un multiplicador de 1.5 veces el rango intercuartílico. Según Tukey, los valores se encuentran más allá de las *inner fences* se consideran anómalos mientras que los que están más allá de las *outer fences* se consideran extremos [23].

Para ello, hemos implementado una función en Python que identifica los valores anómalos de las diferentes variables del *Dataset* (ver Figura 47).

```
def IdentificarValoresAnomalos(df, nombre_de_variable, multiplicador=1.5):
    valor_de_variable = df[nombre_de_variable]

    Q1 = np.percentile(valor_de_variable, 25.)
    Q3 = np.percentile(valor_de_variable, 75.)
    paso = (Q3 - Q1) * multiplicador

    indices_anomalos = valor_de_variable[~((valor_de_variable >= Q1 - paso) & (valor_de_variable <= Q3 +
    paso))].index.tolist()

    return indices_anomalos
```

Figura 47. Método de identificación de valores anómalos

ELIMINACIÓN

La primera aproximación consiste en la eliminación de los valores anómalos previamente identificados reduciendo de esta manera el tamaño del *Dataset* a 184 filas y 43 columnas (ver Figura 48).

```

def eliminar_outliers(df, columnas_numericas, multiplicador=1.5):
    indices_totales_anomalos = []

    for variable in columnas_numericas:
        indices_anomalos = IdentificarValoresAnomalos(df, variable, multiplicador)
        indices_totales_anomalos.extend(indices_anomalos)

    indices_totales_anomalos = list(set(indices_totales_anomalos))
    df_sin_anomalos = df.drop(df.index[indices_totales_anomalos]).reset_index(drop=True)

    {df_sin_anomalos.shape[1]} características cada una."

    return df_sin_anomalos

```

Figura 48. Método de eliminación de valores anómalos

IMPUTACIÓN CON MEDIA

La segunda aproximación se basa en la sustitución de los valores anómalos identificados por la media de la columna a la que pertenece dicho valor. El tamaño del *Dataset* no varía ya que estos valores se están sustituyendo por otros dejando el conjunto de datos con un tamaño de 339 filas y 43 columnas (ver Figura 49).

```

def sustituir_outliers(df, columnas_numericas, multiplicador=1.5):
    indices_totales_anomalos = []
    df_sustituidos = df.copy()

    for variable in columnas_numericas:
        indices_anomalos = IdentificarValoresAnomalos(df_sustituidos, variable, multiplicador)
        indices_totales_anomalos.extend(indices_anomalos)

        media = df_sustituidos[variable].mean()
        df_sustituidos.loc[indices_anomalos, variable] = media

    indices_totales_anomalos = list(set(indices_totales_anomalos))

    return df_sustituidos

```

Figura 49. Método de sustitución por la media de valores anómalos

IMPUTACIÓN CON MODELO ITERATIVO

La última aproximación consiste en la imputación de los valores anómalos identificados a través de un modelo que sustituye los valores existentes por otros obtenidos a través de una predicción hecha en base a los demás valores del *Dataset*. En este caso tampoco se modifica el tamaño de los datos manteniendo de esta forma las 339 filas y 43 columnas (ver Figura 50).

```

def imputar_outliers(df, columnas_numericas, multiplicador=1.5, max_iter=100, tol=1e-4):
    indices_totales_anomalos = []
    df_imputados = df.copy()

    for variable in columnas_numericas:
        indices_anomalos = IdentificarValoresAnomalos(df_imputados, variable, multiplicador)
        indices_totales_anomalos.extend(indices_anomalos)

        df_imputados.loc[indices_anomalos, variable] = np.nan

    imputer = IterativeImputer(random_state=0, max_iter=max_iter, tol=tol)
    df_imputados[columnas_numericas] = imputer.fit_transform(df_imputados[columnas_numericas])

    indices_totales_anomalos = list(set(indices_totales_anomalos))

    return df_imputados

```

Figura 50. Método de imputación con IterativeImputer de valores anómalos

5.7. Modelado

La fase de modelado la hemos decidido abordar utilizando dos enfoques diferentes mediante el empleo de dos modelos: Regresión Lineal y Random Forest Regressor. Estos modelos han sido seleccionados para aprovechar sus diferentes características en la predicción del ROE.

Por un lado, la regresión lineal es un modelo estadístico que predice el valor de una variable dependiente, como en este caso el ROE, mediante una combinación lineal de una o más variables independientes. Este modelo ha sido escogido por su simplicidad e interpretabilidad, lo que le convierte en una referencia inicial útil para comparar con modelos más complejos. Su funcionamiento es óptimo en los contextos donde las relaciones son aproximadamente lineales, facilitando de esta manera ver cómo las diferentes variables elegidas afectan al ROE.

Por otro lado, el Random Forest Regressor, es un modelo de aprendizaje basado en la construcción de múltiples árboles de decisión de los cuales se realiza el promedio de sus predicciones para obtener el resultado final. Este modelo es efectivo para detectar las interacciones no lineales entre las variables superando de esta manera las limitaciones presentadas por modelos más simples como la Regresión Lineal presentada anteriormente. Hemos decidido emplearlo dentro del análisis predictivo con la finalidad de comparar su desempeño con el de la Regresión Lineal y evaluar de esta manera si la capacidad de detectar patrones no lineales tiene un impacto importante en la predicción del ROE.

Por último, hay que mencionar que la fase de modelado se ha realizado a través de la correlación de las variables con el ROE, empleando umbrales de correlación entre el 0.05 - 0.30 (ver Figura 51) y seleccionado las variables que se encontraban por encima de los diferentes umbrales para entrenar el modelo. Todo esto además se ha hecho con las aproximaciones mencionadas en el apartado 5.6.4. con respecto a los valores anómalos además de realizarse con el *Dataset* original.

```
def seleccionar_caracteristicas(df, umbral):
    matriz_correlacion = df.corr()
    variables = matriz_correlacion[matriz_correlacion['ROE'].abs() > umbral].index.tolist()
    if 'ROE' in variables:
        variables.remove('ROE')
    return variables

umbrales_correlacion = [0.05, 0.10, 0.15, 0.20, 0.25, 0.30]
```

Figura 51. Selección de variables para modelado

5.7.1. Selección de variables de los diferentes Dataset

En esta sección veremos cuáles han sido las variables seleccionadas para las diferentes aproximaciones realizadas en función del tratamiento de los valores anómalos. Este conjunto de variables ha sido utilizado para entrenar ambos modelos, tanto el de Regresión Lineal como el Random Forest Regressor.

ORIGINAL

Vemos en la Figura 52 cuáles fueron las variables seleccionadas para los diferentes umbrales de correlación para el *Dataset* original, sin ningún tipo de tratamiento a los valores anómalos.

CORRELACIÓN	VARIABLES
>0.05	AÑO, RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, APALANCAMIENTO, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, ESTRUCTURA DE COSTES, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, NUEVA LEG LAB, PANDEMIA, GUERRA, LIBERTAD ECONOMICA, DERECHOS DE PROPIEDAD, LIBERTAD LABORAL, TASA DE PARO, VAR% PIB, VAR% PRICE STOCK, SECTOR_Construcción e Infraestructura, SECTOR_Energía y Recursos, SECTOR_Industria y Manufactura, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ
>0.10	AÑO, RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, APALANCAMIENTO, RATIO LIQUIDEZ, DIVIDENDO, ESTRUCTURA DE COSTES, BENEFICIO, PANDEMIA, DERECHOS DE PROPIEDAD, VAR% PIB, SECTOR_Energía y Recursos, SECTOR_Industria y Manufactura, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_MADUREZ
>0.15	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, APALANCAMIENTO, DIVIDENDO, ESTRUCTURA DE COSTES, BENEFICIO, VAR% PIB, SECTOR_Industria y Manufactura, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION
>0.20	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO, SECTOR_Servicios Financieros
>0.25	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO
>0.30	ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO

Figura 52. Variables seleccionadas Dataset original

SIN ANÓMALOS

En segundo lugar (ver Figura 53), vemos las variables seleccionadas para el *Dataset* en el que se han eliminado por completo los valores anómalos.

CORRELACIÓN	VARIABLES
>0.05	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, MULT CAPITAL, EFE FINANCIERO, APALANCAMIENTO, EFECTO FISCAL, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, NUEVA LEG LAB, PANDEMIA, GUERRA, LIBERTAD ECONOMICA, GASTO PUBLICO, TASA DE PARO, VAR% PIB, VAR% IPC, VAR% PRICE STOCK, SECTOR_Construcción e Infraestructura, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, SECTOR_Tecnología y Comunicaciones, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ, CICLO DE VIDA_REESTRUCTURACIÓN
>0.10	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, MULT CAPITAL, EFE FINANCIERO, APALANCAMIENTO, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, LIBERTAD ECONOMICA, GASTO PUBLICO, VAR% PIB, VAR% IPC, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_REESTRUCTURACIÓN
>0.15	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, MULT CAPITAL, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, BENEFICIO, AÑO ELECCIONES, LIBERTAD ECONOMICA, VAR% PIB, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION
>0.20	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, BNFO A RESERVAS, BENEFICIO, VAR% PIB, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION
>0.25	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO, VAR% PIB, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros
>0.30	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO, VAR% PIB, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros

Figura 53. Variables seleccionadas Dataset sin anómalos

IMPUTACIÓN CON MEDIA

En tercer lugar (ver Figura 54), vemos como queda el modelado del *Dataset* en el que se ha realizado una imputación con la media a los valores anómalos.

CORRELACIÓN	VARIABLES
>0.05	AÑO, RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, MULT CAPITAL, EFE FINANCIERO, EFECTO FISCAL, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, ESTRUCTURA DE COSTES, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, NUEVA LEG LAB, PANDEMIA, GUERRA, LIBERTAD ECONOMICA, DERECHOS DE PROPIEDAD, VAR% PIB, VAR% IPC, VAR% PRICE STOCK, SECTOR_Construcción e Infraestructura, SECTOR_Energía y Recursos, SECTOR_Industria y Manufactura, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ, CICLO DE VIDA_REESTRUCTURACIÓN
>0.10	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, PANDEMIA, LIBERTAD ECONOMICA, VAR% PIB, VAR% PRICE STOCK, SECTOR_Construcción e Infraestructura, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ
>0.15	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, PANDEMIA, LIBERTAD ECONOMICA, VAR% PIB, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_MADUREZ
>0.20	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BENEFICIO, VAR% PIB, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION
>0.25	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, DIVIDENDO, BENEFICIO, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros
>0.30	ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO

Figura 54. Variables seleccionadas Dataset con imputación con la media

IMPUTACIÓN CON MODELO ITERATIVO

Por último (ver Figura 55), mostraremos cuáles han sido las variables elegidas para el *Dataset* cuyo tratamiento de valores anómalos se ha hecho a través de una imputación con un modelo iterativo.

CORRELACIÓN	VARIABLES
>0.05	AÑO, RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, MULT CAPITAL, EFE FINANCIERO, APALANCAMIENTO, EFECTO FISCAL, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, ESTRUCTURA DE COSTES, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, NUEVA LEG LAB, PANDEMIA, GUERRA, LIBERTAD ECONOMICA, DERECHOS DE PROPIEDAD, LIBERTAD LABORAL, VAR% PIB, VAR% PRICE STOCK, SECTOR_Construcción e Infraestructura, SECTOR_Energía y Recursos, SECTOR_Industria y Manufactura, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ, CICLO DE VIDA_REESTRUCTURACIÓN
>0.10	AÑO, RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, EFECTO FISCAL, PRESION FISCAL, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, AÑO ELECCIONES, PANDEMIA, GUERRA, LIBERTAD ECONOMICA, DERECHOS DE PROPIEDAD, VAR% PIB, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_DECLIVE, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ
>0.15	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BNFO A RESERVAS, AUMENTA GASTOS PERSONAL, BENEFICIO, PANDEMIA, LIBERTAD ECONOMICA, VAR% PIB, VAR% PRICE STOCK, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_LIQUIDACION, CICLO DE VIDA_MADUREZ
>0.20	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, RATIO LIQUIDEZ, DIVIDENDO, BENEFICIO, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros, CICLO DE VIDA_INTRODUCCION, CICLO DE VIDA_MADUREZ
>0.25	RATIO ENDEUDAMIENTO, ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, DIVIDENDO, BENEFICIO, SECTOR_Energía y Recursos, SECTOR_Servicios Financieros
>0.30	ROTACION ACTIVOS, MARGEN OPERATIVO, MARGEN, ROA OPERATIVA, ROA, EFE FINANCIERO, BENEFICIO, SECTOR_Servicios Financieros

Figura 55. Variables seleccionadas Dataset con imputación con modelo iterativo

5.8. Evaluación

A lo largo de esta sección mostraremos y evaluaremos los resultados obtenidos por los modelos de Regresión Lineal y Random Forest Regressor aplicados a las diferentes variaciones del *Dataset*. Además, explicaremos las decisiones tomadas a lo largo del proceso, incluyendo la elección del método de evaluación Cross-Validation con cinco *folds*, así como la selección de las métricas RMSE (*Root Mean Squared Error*) y R^2 (Coeficiente de Determinación) para medir el rendimiento de los modelos (ver Figura 56).

K-fold Cross-Validation es una técnica de validación que divide el conjunto de los datos en k subconjuntos o *folds* más o menos con la misma cantidad de datos. En este caso, al tener cinco subconjuntos, el conjunto de datos se divide en cinco partes, de las cuales se utilizan cuatro para el entrenamiento del modelo y se evalúa con la parte restante. Este proceso se repite cinco veces, rotando la parte que se usa como validación, de forma que cada subconjunto se ha utilizado como validación una única vez [24].

La elección de esta técnica se fundamenta principalmente en dos razones. Por un lado, proporciona un buen equilibrio entre la precisión de la evaluación y la eficiencia computacional. Aunque con un mayor número de subconjuntos, como diez, podría ofrecer estimaciones más precisas, también aumentaría el coste computacional de manera significativa. Por otro lado, el uso de cinco subconjuntos ayuda a gestionar el coste de oportunidad entre el sesgo y la varianza en las estimaciones, proporcionando de esta manera resultados con menos varianza en comparación con otros modelos, lo que se traduce en una evaluación más estable y confiable del desempeño del modelo.

En cuanto a las métricas utilizadas, el RMSE es muy útil en los contextos donde los errores grandes son muy críticos para la aplicación del modelo como es en el caso del ROE, penalizando de esta manera estos errores debido a la cuadratura entre los valores predichos y los observados. Además, al estar en las mismas unidades que la variable a predecir, proporciona una lectura fácil y directa del error promedio, facilitando de esta manera la comunicación con las personas no técnicas que tengan que interpretar los datos como los accionistas o posibles inversores. Por todo esto, en situaciones donde minimizar los errores de la predicción es algo crucial como es en el caso de las finanzas, el RMSE es preferido sobre otras métricas para medir el error producido en la predicción [25].

En lo que se refiere al R^2 , sirve para medir la proporción de la variabilidad total de la variable dependiente que puede ser explicada por el modelo. Un R^2 con un valor cercano a la unidad indica que el modelo es capaz de explicar la mayor parte de la variabilidad, mientras que un valor cercano a cero muestra que el modelo utilizado tiene poca capacidad explicativa.

```
num_folds = 5
rmse_metricas = make_scorer(lambda y_true, y_pred: np.sqrt(mean_squared_error(y_true, y_pred)),
greater_is_better=True)
r2_metricas = make_scorer(r2_score)

kfold = KFold(n_splits=num_folds, shuffle=True, random_state=0)
rmse_resultados_cv = cross_val_score(modelo, X, Y, cv=kfold, scoring=rmse_metricas)
r2_resultados_cv = cross_val_score(modelo, X, Y, cv=kfold, scoring=r2_metricas)
```

Figura 56. Evaluación de los modelos

5.8.1. Regresión Lineal

En esta sección presentaremos a través de la Figura 57 los resultados obtenidos por el modelo de Regresión Lineal utilizando los cuatro conjuntos creados según el tratamiento de los valores anómalos, cada uno seleccionado según los diferentes umbrales de correlación con el ROE (0.05, 0.1, 0.15, 0.2, 0.25, 0.3). Compararemos los resultados de RMSE y R^2 para cada uno de los *Dataset*, destacando los mejores en términos de precisión y capacidad explicativa.

	Dataset	Umbral Corr	RMSE	R²	Número de Variables Seleccionadas
0	Original	0.05	0.0140	0.7715	34
1	Original	0.10	0.0137	0.7776	22
2	Original	0.15	0.0139	0.7712	15
3	Original	0.20	0.0149	0.7268	9
4	Original	0.25	0.0148	0.7337	8
5	Original	0.30	0.0152	0.7388	7
6	Sin Anómalos	0.05	0.0057	0.8156	35
7	Sin Anómalos	0.10	0.0059	0.8007	27
8	Sin Anómalos	0.15	0.0061	0.7910	21
9	Sin Anómalos	0.20	0.0061	0.7855	15
10	Sin Anómalos	0.25	0.0061	0.7829	11
11	Sin Anómalos	0.30	0.0061	0.7829	11
12	Anómalos Media	0.05	0.0096	0.6313	35
13	Anómalos Media	0.10	0.0094	0.6473	23
14	Anómalos Media	0.15	0.0096	0.6358	21
15	Anómalos Media	0.20	0.0094	0.6459	14
16	Anómalos Media	0.25	0.0095	0.6437	11
17	Anómalos Media	0.30	0.0095	0.6438	7
18	Anómalos Imputados	0.05	0.0059	0.8455	36
19	Anómalos Imputados	0.10	0.0061	0.8367	28
20	Anómalos Imputados	0.15	0.0060	0.8394	21
21	Anómalos Imputados	0.20	0.0062	0.8276	14
22	Anómalos Imputados	0.25	0.0063	0.8224	11
23	Anómalos Imputados	0.30	0.0069	0.7872	8

Figura 57. Resultados Regresión Lineal

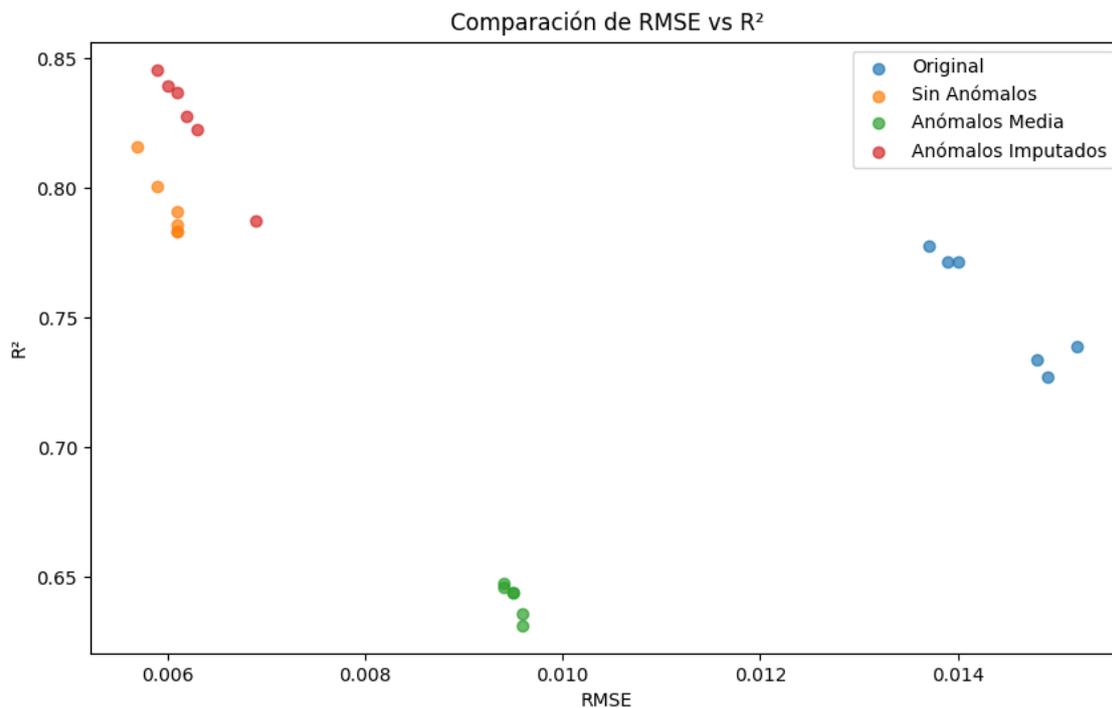


Figura 58. Distribución de resultados Regresión Lineal

Visualizando el gráfico de dispersión de RMSE y R^2 de la Figura 58 podemos ver como el modelo con los datos imputados por el modelo iterativo es en términos generales el que mejor resultados ofrece, tanto en términos de error como en explicabilidad. A este modelo le sigue el que emplea el *Dataset* sin los valores anómalos, el cual había visto reducido su tamaño en más de la mitad de los datos. El modelo original, si bien puede llegar a ofrecer niveles similares de explicabilidad, el error es muy alto, lo cual es inadmisibles en variables tan sensibles como la rentabilidad. Por último, observamos como la imputación con la media es la peor opción en términos de explicabilidad de los cuatro modelos y la tercera peor en términos de error en líneas generales.

Pasemos a seleccionar el mejor modelo en términos de error y de explicabilidad para cada uno de los *Dataset* (ver Figura 59).

	Dataset	Umbral Corr	RMSE	R^2	Número de Variables Seleccionadas
18	Anómalos Imputados	0.05	0.0059	0.8455	36
13	Anómalos Media	0.10	0.0094	0.6473	23
1	Original	0.10	0.0137	0.7776	22
6	Sin Anómalos	0.05	0.0057	0.8156	35

Figura 59. Mejores modelos Regresión Lineal

Vemos como el mejor modelo es aquel que emplea el *Dataset* con los anómalos imputados con un modelo y un umbral de correlación del 0.05 con un total de 36 variables seleccionadas.

Ahora analizaremos si un cambio en la herramienta predictiva muestra una mejora significativa en las métricas del modelo.



5.8.2. Random Forest Regressor

Analicemos ahora los resultados obtenidos por el Random Forest en las mismas condiciones que los modelos anteriores, con los cuatro conjuntos de datos distintos y las distintas correlaciones con el ROE (ver Figura 60).

	Dataset	Umbral Corr	RMSE	R²	Número de Variables Seleccionadas
0	Original	0.05	0.0132	0.8229	34
1	Original	0.10	0.0132	0.8235	22
2	Original	0.15	0.0131	0.8221	15
3	Original	0.20	0.0142	0.7862	9
4	Original	0.25	0.0145	0.7759	8
5	Original	0.30	0.0155	0.7431	7
6	Sin Anómalos	0.05	0.0047	0.8737	35
7	Sin Anómalos	0.10	0.0047	0.8763	27
8	Sin Anómalos	0.15	0.0053	0.8421	21
9	Sin Anómalos	0.20	0.0055	0.8316	15
10	Sin Anómalos	0.25	0.0053	0.8421	11
11	Sin Anómalos	0.30	0.0053	0.8421	11
12	Anómalos Media	0.05	0.0079	0.7471	35
13	Anómalos Media	0.10	0.0079	0.7497	23
14	Anómalos Media	0.15	0.0078	0.7530	21
15	Anómalos Media	0.20	0.0077	0.7607	14
16	Anómalos Media	0.25	0.0078	0.7605	11
17	Anómalos Media	0.30	0.0082	0.7297	7
18	Anómalos Imputados	0.05	0.0052	0.8788	36
19	Anómalos Imputados	0.10	0.0056	0.8603	28
20	Anómalos Imputados	0.15	0.0056	0.8603	21
21	Anómalos Imputados	0.20	0.0055	0.8651	14
22	Anómalos Imputados	0.25	0.0056	0.8610	11
23	Anómalos Imputados	0.30	0.0064	0.8195	8

Figura 60. Resultados Random Forest Regressor

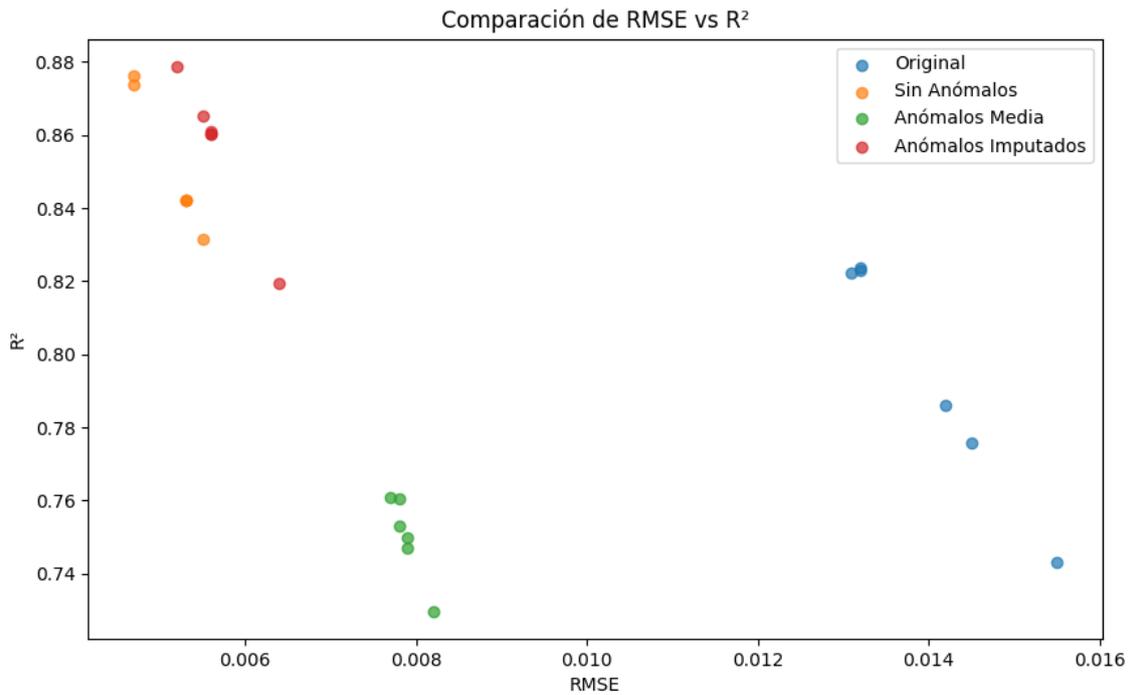


Figura 61. Distribución de resultados Random Forest Regressor

Este gráfico de dispersión (ver Figura 61) sigue una estructura parecida al anterior, sobre todo en lo que respecta al conjunto original y al imputado por la media. Sobre los otros dos, vemos como esta vez la cercanía del *Dataset* en el que se eliminan los valores anómalos con el que se sustituyen de forma predictiva en términos de explicabilidad es mucho mayor, llegando incluso a mostrar mejores valores en lo que al error se refiere.

Veamos ahora si los mejores valores de cada *Dataset* son mejores que aquellos que hemos obtenido con el modelo de Regresión Lineal (ver Figura 62).

	Dataset	Umbral Corr	RMSE	R²	Número de Variables Seleccionadas
18	Anómalos Imputados	0.05	0.0052	0.8788	36
15	Anómalos Media	0.20	0.0077	0.7607	14
1	Original	0.10	0.0132	0.8235	22
7	Sin Anómalos	0.10	0.0047	0.8763	27

Figura 62. Mejores modelos Random Forest Regressor

Podemos observar cómo sí ha habido una mejora en términos de explicabilidad y error con respecto al modelo predictivo anterior en todos los *Dataset*. Esto nos indica que la capacidad para identificar relaciones no lineales entre las variables ha sido clave para observar una mejor en el desempeño del modelo.

5.8.3. Modelo elegido

A la hora de elegir el modelo definitivo hay que tener en cuenta en primer lugar cuál ha sido el *Dataset* que mejor desempeño ha ofrecido para cada uno de los modelos empleados, y luego cuál de los dos modelos ha dado mejores resultados.

En cuanto al *Dataset*, vemos que tanto para el modelo de Regresión Lineal como para el Random Forest Regressor el *Dataset* que ofrece mejores resultados es aquel que tiene los datos anómalos imputados a través del modelo iterativo. Respecto a al modelo sin anómalos, que es el segundo que mejores resultados ofrece, llegando incluso a ser el que menos error presenta, no consideramos apropiada su elección ya que el hecho de reducir a la mitad el *Dataset* hace que el estudio deje de tener sentido ya que gran parte de los datos de todos los años no se ven reflejados en el análisis. Por ello, dado que el *Dataset* con el modelo iterativo mejora los resultados, es el elegido finalmente para ambos modelos como el que mejor desempeño realiza.

Finalmente, en cuanto a los modelos, observamos que el modelo en el que empleamos el Random Forest Regressor obtiene casi un 3% más de explicabilidad que el de Regresión Lineal y consigue disminuir el error en 0.0007, por tanto, el modelo final elegido para predecir el ROE es el siguiente:

- Modelo utilizado: Random Forest Regressor
- Dataset: Imputación de valores anómalos con IterativeImputer
- Umbral de correlación: 0.05
- Número de variables: 36
- RMSE (error): 0.0052
- R^2 (explicabilidad): 0.8788 (87.88% de la variabilidad es explicada por este modelo)

6. Análisis de resultados

En esta sección, presentaremos las conclusiones derivadas de los dos enfoques de análisis aplicados en este estudio. Primeramente, analizaremos la parte descriptiva, focalizándonos en conclusiones relacionadas con los diferentes sectores, las tendencias o el impacto de la pandemia en el índice. A continuación, examinaremos aquellas variables que no han sido utilizadas para la predicción del ROE destacando por qué no son relevantes para el estudio. Finalmente, dedicaremos un apartado específico a los inversores, mostrando los hallazgos realizados y formulando recomendaciones en base a estos.

ANÁLISIS SECTORIAL

Fijándonos en el desempeño de los diferentes sectores del IBEX 35, hemos observado que los sectores energético y financiero han mantenido de manera consistente a lo largo de los últimos diez años un ROE positivo y unos márgenes razonables, incluso en periodos de crisis como la pandemia o la guerra de Ucrania. Esta robustez ha convertido a estos sectores en pilares fundamentales de la economía, de los que dependen en gran parte el resto de las empresas, tanto las del IBEX 35 como las que se encuentran fuera del índice. A diferencia de otros sectores que han mostrado una mayor volatilidad, su desempeño constante les ha permitido consolidar su posición como sectores clave de la economía.

TENDENCIAS

En cuanto a las tendencias observadas en el IBEX 35 a lo largo de los últimos diez años podemos destacar las siguientes:

- **Operación internacional:** En base a la carga fiscal de las empresas a lo largo de estos años, podemos ver como la gran mayoría no solo operan en España, sino que tienen una presencia significativa en mercados internacionales. Esto implica que pagan una parte proporcional a sus beneficios en los países donde operan, lo que muestra una creciente internacionalización de las grandes corporaciones españolas
- **Conservadurismo financiero:** También se observa una tendencia hacia un conservadurismo dentro del IBEX 35. Esto se ve reflejado en una disminución del endeudamiento promedio de las empresas, pasando de ser empresas con un alto nivel de deuda, cercano al 60% de toda su financiación, a empresas con valores más moderados en torno al 50%. Además, en el 75% de las ocasiones en las que las empresas han tenido la oportunidad de hacerlo, las empresas han optado por destinar parte de sus beneficios a reservas, en lugar de repartirlo íntegramente entre los accionistas, fortaleciendo de esta manera la estructura financiera interna.
- **Rentabilidad creciente:** A pesar de las dificultades económicas recientes, como es el caso de la pandemia, dentro del IBEX 35 existe una tendencia creciente en términos de rentabilidad, mostrando una gran capacidad para recuperarse y mejorar su desempeño, lo que es un fiel reflejo de una estructura resiliente y adaptativa en términos generales.

IMPACTO DE LA PANDEMIA

Durante la pandemia, el sector tecnológico y de servicios fueron los más afectados dentro del IBEX 35. El sector tecnológico, compuesto por empresas como Amadeus, Indra o Telefónica, que

disfrutaba de rentabilidades alrededor del 10%, experimentó una caída hasta registrar rentabilidades negativas del -1%, lo que representa una pérdida del 110% de su rentabilidad en tan solo un año, de 2019 a 2020. En la misma línea, el sector servicios, compuesto por empresas como AENA y Meliá Hotels, que había alcanzado rentabilidades de hasta el 20% en el año 2019, vio caer sus márgenes hasta obtener rentabilidades negativas, significando de nuevo una pérdida superior al 100% en términos de rentabilidad.

En contraposición, el sector energético, liderado por empresas como Iberdrola, Endesa o Naturgy, demostró una gran resiliencia a la crisis causada por la pandemia. No solo logró resistir, sino que aumentó su rentabilidad durante este periodo en términos promedios. Esto se debe principalmente a que la energía es uno de los elementos esenciales para la gran mayoría de las actividades de una economía, hecho que provocó que la demanda siguiese constante, permitiendo de esta manera al sector seguir manteniendo los niveles de rentabilidad que venían proporcionando.

VARIABLES PREDICTIVAS

En el proceso de modelado para predecir el ROE, hay varias variables que no forman parte del modelo predictivo final, analicemos a que se puede deber su ausencia:

- **Variables Macroeconómicas (Gasto Público, Tasa de Paro, VAR% IPV):** Estas variables macroeconómicas no mostraron tener una alta correlación con la rentabilidad de las empresas. En el caso del gasto público y la tasa de paro se debe principalmente a que la mayoría de las empresas, al ser empresas grandes, casi todas multinacionales, no dependen directamente de lo que suceda en términos de políticas de gasto o dentro del mercado laboral, ya que su capacidad productiva no se verá generalmente afectada si baja el paro por su alta capacidad para atraer capital humano. En el caso de la inflación, esta tiene un impacto diluido en la rentabilidad de las empresas ya que estas tienen la capacidad de ajustar precios y costes en respuesta a un aumento del coste de los insumos para la producción.
- **Libertad de negocios:** Pese a que la libertad de negocios es crucial en el entorno empresarial general, su impacto en este contexto parece ser reducido. Esto se debe principalmente a que los datos introducidos corresponden únicamente a la libertad económica de España, y al ser muchas de las empresas multinacionales, le afectan diferentes entornos empresariales, por ello la relevancia de esta variable es tan pequeña.
- **Sector Tecnología y Comunicaciones:** El sector tecnológico se encuentra en pleno auge en los últimos años, y esto es precisamente lo que hace que se encuentre fuera del modelo. La alta competitividad del sector junto con la constante innovación en términos de tecnologías hace difícil establecer patrones que sean útiles para poder predecir la rentabilidad, a diferencia de otros sectores más estables.

RECOMENDACIONES PARA INVERSORES

Por último, las rentabilidades observadas en las empresas del IBEX 35 los últimos diez años han demostrado ser constantes y resilientes, lo que hace que tanto la inversión en fondos indexados del IBEX 35 como en acciones individuales de las empresas es una opción atractiva de inversión. Estas inversiones pueden considerarse un activo valioso y comparable con otros mercados a nivel europeo o mundial. En particular, se recomienda el sector energético por su destacada resiliencia y capacidad para mantener una buena rentabilidad, siendo una opción sólida para aquellos inversores que busquen un refugio de valor a largo plazo.

7. Conclusiones

El análisis y visualización de datos del IBEX 35 realizado en este Trabajo de Fin de Grado ha permitido obtener una comprensión profunda de los factores que influyen en el éxito empresarial de las empresas cotizadas en este índice. La combinación de métodos descriptivos y predictivos ha permitido, por su parte, extraer conclusiones relevantes que ayudan a entender las dinámicas pasadas y presentes del IBEX 35, proporcionando, a su vez, herramientas útiles para la toma de decisiones a futuro.

En primer lugar, uno de los hallazgos más reveladores del trabajo es la importancia de la gestión eficiente de activos para la rentabilidad empresarial. La rotación de activos, el margen operativo, y la ROA aparecen como variables cruciales para predecir el retorno sobre el capital (ROE). Esto indica que las empresas que pueden maximizar sus ingresos en relación con sus activos y controlar sus costes operativos son más propensas a ofrecer rendimientos positivos a sus accionistas.

Otro aspecto significativo es la influencia negativa del apalancamiento financiero en la rentabilidad. Los análisis realizados demuestran que un alto nivel de endeudamiento puede limitar la capacidad de una empresa para generar beneficios, subrayando la importancia de mantener un equilibrio saludable entre financiamiento propio y de terceros. En este contexto, hemos observado una tendencia hacia un conservadurismo financiero, con las empresas reduciendo sistemáticamente el nivel de deuda en sus estructuras de financiamiento a lo largo de los años.

El impacto de factores externos y situaciones de crisis también ha sido un área clave del análisis realizado. La capacidad de las empresas para adaptarse y recuperarse de eventos como la pandemia del COVID-19 ha demostrado la resiliencia del IBEX 35 en su conjunto. No obstante, la disparidad en la afectación de distintos sectores pone de manifiesto la necesidad de diversificación en la cartera de un inversor para mitigar riesgos asociados a crisis sectoriales.

El trabajo también resalta la pertinencia de la internacionalización de las empresas del IBEX 35. La presencia significativa en mercados extranjeros no solo diversifica sus fuentes de ingresos, sino que también les permite amortiguar el impacto de cambios regulatorios y fiscales en el ámbito nacional. Sin embargo, esto también sugiere que la cuantificación del éxito de estas empresas se complica debido a la diversidad de entornos fiscales y económicos en los que operan.

Desde una perspectiva tecnológica, el uso de técnicas avanzadas de análisis de datos y *machine learning*, como el Random Forest Regressor, ha demostrado ser altamente efectivo en la predicción del ROE. La capacidad de estos modelos para capturar relaciones no lineales y complejas entre variables proporciona un marco robusto para futuros análisis predictivos en finanzas, sugiriendo una dirección prometedora para la investigación y empleo de la analítica avanzada en el sector empresarial.

Por último, el análisis de correlación ha revelado varias variables que no son directamente influyentes sobre el ROE, subrayando la importancia de una elección cuidadosa de variables en la construcción de modelos predictivos. Esto ofrece una valiosa lección metodológica sobre la necesidad de un proceso riguroso de selección de variables en análisis de datos financieros.

7.1. Ampliaciones y estudios futuros

En este capítulo final, exploraremos posibles vías de ampliación para el estudio y sugeriremos otros enfoques que podrían ser complementarios.

Para mejorar los resultados obtenidos, una posible ampliación sería extender el análisis histórico hasta los orígenes del IBEX 35. Esto permitiría un análisis mucho más profundo y una visión a largo plazo de las tendencias y factores que han influido en el desempeño de las empresas a lo largo de los años.

Otra ampliación importante sería incorporar al *Dataset* datos de países en los que las empresas del IBEX 35 tributan. Esto daría una perspectiva más completa al considerar las diferentes situaciones económicas y fiscales que pueden afectar a una empresa.

Además, una posible variación del trabajo sería enfocarlo en un solo sector del IBEX 35. Esto no solo reduciría el coste de obtención de datos, sino que también permitiría un análisis más detallado, obteniendo *Insights* más específicos.

Por último, sería interesante realizar un análisis comparativo con otros índices bursátiles a nivel europeo, como el DAX alemán o el CAC 40 francés. Esto ayudaría a poner en contexto el desempeño del IBEX 35 en un marco europeo, dando pie a identificar las fortalezas y las debilidades relativas que tienen las empresas españolas en comparación con las de otros países.

Bibliografía

- [1] Cherif, M., & Gazdar, K. (2010). *Institutional and macroeconomic determinants of stock market development in MENA region: New results from a panel data analysis*. *International Journal of Banking and Finance*, 7(1), 139-159.
- [2] Ou, P., & Wang, H. (2009). *Prediction of stock market index movement by ten data mining techniques*. *Modern Applied Science*, 3(12), 28-42.
- [3] Maini, S. S., & Govinda, K. (2017, December). *Stock market prediction using data mining techniques*. In 2017 International Conference on Intelligent Sustainable Systems (ICISS), 654-661. IEEE.
- [4] Yang, Y., & Wang, Z. (2023). *Prediction of return on equity of the energy industry based on equity characteristics*. *Frontiers in Energy Research*, 11, 1136914.
- [5] Balci, T., & Ogul, H. (2021, January). *Predicting Bank Return on Equity (ROE) using Neural Networks*. In 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), 279-286. IEEE.
- [6] Consoli, S., Negri, M., Tebbifakhr, A., Tosetti, E., & Turchi, M. (2021, October). *Forecasting the IBEX-35 stock index using deep learning and news emotions*. In International Conference on Machine Learning, Optimization, and Data Science, 308-323. Springer International Publishing.
- [7] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). *Jupyter Notebooks – a publishing format for reproducible computational workflows*. In Positioning and Power in Academic Publishing: Players, Agents and Agendas, 87-90. IOS Press.
- [8] Nick, T. G. (2007). *Descriptive statistics. Topics in biostatistics*, 33-52.
- [9] Lebanon, G. (2010). *Linear Regression*.
- [10] Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press.
- [11] Breiman, L. (1996). *Bagging predictors*. *Machine learning*, 24, 123-140.
- [12] Breiman, L. (2001). *Random forests*. *Machine learning*, 45, 5-32.
- [13] Ho, T. K. (1995, August). *Random decision forests*. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.
- [14] Marí, S.; Mateos, A.; Polo, F. (2013). *Análisis económico – financiero: supuestos prácticos*. Ed. UPV (Ref. 4013).
- [15] Amat, O. (2008). *Análisis de Estados Financieros (8ª ed.)*. Editorial Gestión 2000, Barcelona.

- [16] Boya, C., & Monino, J. L. (2010). *The impact of exogenous information on stock value through the coloration concept: a test model*. Journal of Innovation Economics, (2), 163-180.
- [17] Santa-Clara, P., & Valkanov, R. (2003). *The presidential puzzle: Political cycles and the stock market*. The Journal of Finance, 58(5), 1841-1872.
- [18] Leung, W. S., Mazouz, K., Chen, J., & Wood, G. (2018). *Organization capital, labor market flexibility, and stock returns around the world*. Journal of Banking & Finance, 89, 150-168
- [19] Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). *Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash*. The Review of Corporate Finance Studies, 9(3), 593-621.
- [20] Derindere Köseoğlu, S., Mercangöz, B. A., Khan, K., & Sarwar, S. (2024). *The impact of the Russian-Ukraine war on the stock market: A causal analysis*. Applied Economics, 56(21), 2509-2519.
- [21] The Heritage Foundation. (2023). *Index of Economic Freedom: Methodology*. The Heritage Foundation. https://www.heritage.org/index/pdf/2024/IndexMethodology_2024.pdf
- [22] Levine, L. (2012). *Economic growth and the unemployment rate*.
- [23] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- [24] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in python*. 201-220
- [25] Hodson, T. O. (2022). *Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not*. Geoscientific Model Development Discussions, 2022, 1-10.

Anexos

Anexo A – Análisis

Cuaderno de Jupyter Notebook:

https://colab.research.google.com/drive/1g5PW3S96Z_oUHNg5ROlnJ9LovjCiaYDX?usp=sharing

Base de datos original:

https://docs.google.com/spreadsheets/d/1qbxJ2bSd648fwKYDKu_RV80_s3dLbiP-/edit?usp=sharing&ouid=117084640944424722745&rtpof=true&sd=true

Base de datos para el análisis:

<https://docs.google.com/spreadsheets/d/1M65PMWYEZEVB81hacgxE6J893yYfzG5X/edit?usp=sharing&ouid=117084640944424722745&rtpof=true&sd=true>

Anexo B – Objetivos de Desarrollo Sostenible

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.		x		
ODS 2. Hambre cero.				x
ODS 3. Salud y bienestar.				x
ODS 4. Educación de calidad.			x	
ODS 5. Igualdad de género.				x
ODS 6. Agua limpia y saneamiento.				x
ODS 7. Energía asequible y no contaminante.				x
ODS 8. Trabajo decente y crecimiento económico.	x			
ODS 9. Industria, innovación e infraestructuras.	x			
ODS 10. Reducción de las desigualdades.		x		
ODS 11. Ciudades y comunidades sostenibles.				x
ODS 12. Producción y consumo responsables.				x
ODS 13. Acción por el clima.				x
ODS 14. Vida submarina.				x
ODS 15. Vida de ecosistemas terrestres.				x
ODS 16. Paz, justicia e instituciones sólidas.				x
ODS 17. Alianzas para lograr objetivos.			x	

El trabajo realizado tiene varias conexiones significativas (ver Figura 63) con los Objetivos de Desarrollo Sostenible, veamos cuales son:



Figura 63. Objetivos de Desarrollo Sostenible relacionados

- **Fin de la pobreza:** Aunque es indirecta, el estudio contribuye a la reducción de la pobreza al promover un crecimiento económico sostenible y continuo, identificando estrategias para proteger el empleo en tiempos de crisis.
- **Trabajo Decente y Crecimiento Económico:** De manera fuerte y directa, el análisis de factores que impulsan la rentabilidad y éxito empresarial contribuye a un crecimiento económico que genere trabajo inclusivo y de calidad. Además, obtener factores de resiliencia en situaciones de crisis ayuda a prevenir la pérdida de calidad en el trabajo y a disminuir las caídas empresariales.
- **Industria, Innovación e Infraestructura:** De manera directa, al estudiar las empresas del IBEX 35 y exponer sus fortalezas y debilidades, se genera un efecto en el que las empresas busquen o mantener sus fortalezas o solventar sus debilidades, fomentando la innovación y el desarrollo.
- **Reducción de las Desigualdades:** Al enfocarse en el crecimiento de las empresas, de manera indirecta se está fomentando la creación de empleo a través de la expansión de las empresas, creando oportunidades nuevas de inserción al mercado laboral.

En cuanto a los que tienen relación baja tenemos:

- **Educación de Calidad:** Las empresas que destinan parte de sus beneficios a reservas o a la mejora de sus empleados podrían estar invirtiendo indirectamente en la formación y educación.
- **Alianzas para Lograr Objetivos:** Debido a la correlación entre los diferentes sectores y la codependencia existente para poder llevar a cabo un mayor desempeño.

Los demás no están relacionados ni directa ni indirectamente o no proceden.

En conclusión, el estudio sobre las empresas del IBEX 35 demuestra un impacto positivo en varios de los Objetivos de Desarrollo Sostenible. Su contribución es especialmente fuerte en el ámbito del trabajo y el crecimiento económico, impulsando la creación de empleo y la fortaleza empresarial. Dado que la generación de empleo y riqueza impacta de una manera general en la sociedad, también se contribuye de manera indirecta a la reducción de desigualdades y de situaciones de vulnerabilidad. Por ello, se muestra que el camino para tener una sociedad más rica y sostenible es a través del fomento de la colaboración tanto entre empresas como en relación con el Estado.