



Universidad
Francisco de Vitoria
UFV Madrid

Trabajo de Fin de Grado

Variables Macroeconómicas y su Influencia en el Tejido Empresarial Español: Un Estudio de Constituciones y Disoluciones

Trabajo Fin de Grado

Grado en Análisis de Negocios – Business Analytics
Mayo de 2024

Autor:

Ignacio López de Carrizosa Grosso

Tutor:

Prof. Dra. Ana Lazcano de Rojas

Facultad de Facultad de Derecho, Empresa y Gobierno

Universidad Francisco de Vitoria

"No todo lo que se cuenta cuenta, y no todo lo que cuenta se cuenta."

- Albert Einstein

Resumen

Este Trabajo de Fin de Grado explora cómo distintas variables macroeconómicas influyen el sistema empresarial español, con un enfoque particular en las constituciones y disoluciones empresariales. La investigación se concentra en la identificación de factores clave que impactan la estabilidad y el desarrollo de las empresas en España, utilizando modelos de regresión para analizar y prever tendencias dentro del mercado empresarial. Además, se busca informar y democratizar el acceso a los hallazgos del estudio, contribuyendo al debate público y a la toma de decisiones estratégicas. Este enfoque se alinea con la promoción de un entorno empresarial resiliente, con miras a fortalecer la economía española en un contexto global cambiante.

Palabras Clave

Regresión lineal múltiple y simple, variables macroeconómicas, disoluciones y constituciones empresariales, ecosistema empresarial español, modelo.

Abstract

This Final Degree Project explores how different macroeconomic variables influence the Spanish business system, with a particular focus on business incorporations and dissolutions. The research concentrates on identifying key factors that impact the stability and development of firms in Spain, using regression models to analyse and forecast trends within the business market. In addition, it seeks to inform and democratise access to the study's findings, contributing to public debate and strategic decision-making. This approach is aligned with the promotion of a resilient business environment, with a view to strengthening the Spanish economy in a changing global context.

Keywords

Multiple and simple linear regression, macroeconomic variables, business dissolutions and incorporations, Spanish business ecosystem, model.

RESUMEN AMPLIADO

Este Trabajo de Fin de Grado (TFG) comenzó con un interés en investigar cómo la pandemia del COVID-19 afectó al sistema empresarial español. Sin embargo, el enfoque evolucionó hacia un análisis más amplio y diversificado de cómo ciertas variables macroeconómicas influyen en el sistema empresarial español.

La primera parte del TFG, titulada "Ingeniería del Dato", se centró en el meticuloso proceso de Extracción, Transformación y Carga (ETL) de datos relevantes para el estudio. Este proceso fue fundamental para asegurar la calidad y utilidad de los datos para el análisis subsiguiente, permitiendo una comprensión profunda y precisa del impacto de la COVID-19 en el tejido empresarial español, con datos obtenidos de fuentes fiables como el Instituto Nacional de Estadística (INE) y epdata.

En la transición hacia la segunda parte del trabajo, titulada "Análisis del Dato", el TFG tomó un giro hacia un enfoque más generalizado, analizando la variabilidad del sistema empresarial español bajo la influencia de variables macroeconómicas. Este cambio refleja una adaptación y evolución del objetivo inicial, abarcando un espectro más amplio de análisis que no solo considera el impacto directo de la pandemia sino también cómo el entorno económico y las políticas gubernamentales modelan el panorama empresarial español.

Este cambio de enfoque permitió una exploración detallada de las dinámicas empresariales, revelando complejas interacciones entre la creación de empresas, el capital desembolsado y variables clave como el PIB y el IPC. A través de técnicas de análisis de datos avanzadas, se identificaron patrones y correlaciones significativas, proporcionando insights valiosos sobre la estabilidad y adaptabilidad del tejido empresarial en España.

El resumen narrativo integrado abarca las observaciones y hallazgos de ambos enfoques, destacando la importancia de considerar un espectro amplio de factores cuando se examina el impacto en el sistema empresarial. La combinación de un enfoque inicial centrado en la pandemia, seguido de un análisis más amplio de variables macroeconómicas, ofrece una comprensión más completa y matizada de las dinámicas empresariales en España.

El trabajo final culmina con recomendaciones basadas en el análisis realizado. Este TFG no solo contribuye al cuerpo académico sobre las disoluciones empresariales en España, sino que también proporciona herramientas analíticas que pueden ser aplicadas en futuras investigaciones y prácticas empresariales.

Índice de Contenidos

INTRODUCCIÓN	12
<i>Descripción del Problema</i>	<i>12</i>
<i>Motivación del Trabajo.....</i>	<i>12</i>
<i>Organización de la Memoria.....</i>	<i>13</i>
<i>Estado del Arte.....</i>	<i>14</i>
OBJETIVOS Y EVOLUCIÓN	20
<i>Evolución</i>	<i>20</i>
<i>Objetivos Planteados.....</i>	<i>21</i>
PRIMERA PARTE: INGENIERÍA DEL DATO	23
1. Introducción a la ingeniería del dato	23
2. Origen de los Datos.....	24
3. Descripción de los Datos	26
3.1. <i>Número total de observaciones y variables en los datasets (limpias).....</i>	<i>26</i>
3.2. <i>Descripción del tipo de datos de cada variable</i>	<i>27</i>
3.3. <i>Formato de los datos en bruto y cualquier transformación aplicada.</i>	<i>29</i>
3.4. <i>Periodicidad y temporalidad de los datos.</i>	<i>34</i>
4. Justificación de la Elección de Variables	36
4.1. <i>Explicación de la selección de variables.....</i>	<i>36</i>
4.2. <i>Relación con los objetivos del TFG.</i>	<i>36</i>
5. Análisis Exploratorio de Datos	37
5.1. <i>Descripción de medidas de tendencia central y dispersión para las variables cuantitativas.</i>	<i>37</i>
5.1.2 <i>Medidas de tendencia central</i>	<i>37</i>
5.1.3 <i>Medidas de dispersión.....</i>	<i>38</i>
5.2. <i>Frecuencias y proporciones para las variables categóricas.</i>	<i>40</i>

6. Gráficos Descriptivos	41
6.1. <i>Gráficos sobre variables y relaciones.....</i>	41
6.2. <i>Patrones, tendencias o correlaciones observadas en los gráficos.</i>	43
7. Almacenamiento de Datos	46
SEGUNDA PARTE: ANÁLISIS DEL DATO	47
8. Introducción al análisis del dato	47
9. Selección y Preparación de Datos para el Análisis	48
10. Modelos Analíticos: Desarrollo y Aplicación (Regresiones Lineales)	51
10.1. <i>Modelo Analítico Supervisado (Regresiones Lineales Simples).....</i>	51
Explicación del Modelo:.....	51
Proceso de Selección de Variables:.....	52
Desarrollo de modelos:	53
Evaluación del Modelo con Medidas de Error/Precisión Específicas:.....	55
10.2. <i>Modelo Analítico Supervisado (Regresiones Lineales Múltiples).....</i>	57
Explicación del Modelo:.....	57
10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE (IPC y PIB)	58
Proceso de Selección de Variables.....	58
Desarrollo del Modelo	58
Evaluación del Modelo con Medidas de Error/Precisión Específicas:.....	60
10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE (Variables Macroeconómicas)...	62
Proceso de Selección de Variables.....	62
Desarrollo del Modelo	63
Reconstrucción del modelo en base a la multicolinealidad.....	65
Evaluación del Modelo con Medidas de Error/Precisión Específicas	66
10.2.2 TERCERA REGRESIÓN LINEAL MÚLTIPLE (Variables más amplias).....	68
Proceso de Selección de Variables.....	68
Desarrollo del Modelo	69

Evaluación del Modelo con Medidas de Error/Precisión Específicas	70
11. Medidas de Adecuación de los Modelos.....	72
11.1. Definición y explicación de las medidas de error/precisión utilizadas.	72
11.2. Comparación de los resultados obtenidos en los modelos.	75
11.2.1 Medidas de la regresión.....	75
11.2.2 Medidas de Error	78
11.3 Aplicación de Pruebas Estadísticas para la Comparación de Modelos.....	79
11.3.1 Prueba de Wilcoxon	79
11.3.2 Prueba de Friedman.....	81
12. Visualización de Datos y Resultados de Modelos	83
12.1. Gráficos de Dispersión	83
12.2. Gráficos de Residuos	87
13. Explicación de Resultados.....	90
13.1. Explicación comprensiva de los resultados de los modelos.	90
13.2. Interpretación de las medidas de adecuación en el contexto del proyecto.	92
14. Conclusiones y Recomendaciones	95
14.1 Conclusiones	95
14.2 Recomendaciones.....	96
15. Resumen Narrativo.....	98
TERCERA PARTE: ANÁLISIS DE NEGOCIO	99
16. Introducción al Análisis de Negocio	99
17. Conclusiones	101
17.1 Evaluación y Respuestas los Objetivos Planteados	101
17.2 Discusión sobre Respuestas: Métodos Utilizados para Responder.....	104
18. Opiniones Profesionales	106
18.1 Información de las Entrevistas	106
18.2 Resumen de la Respuestas Obtenidas	108

18.3 Insights obtenidos de las preguntas	109
19. Recomendaciones y Propuestas	111
19.1 Recomendaciones	111
19.1.1 Empresarios, Ejecutivos y Emprendedores.....	111
19.1.2 Formuladores de Políticas	112
19.2 Propuestas	113
19.2.1 Tipo de Propuesta	113
19.2.2 Problema y Solución de la Propuesta	114
19.2.3 Explicación en Profundidad.....	115
20. Resumen Narrativo.....	117
21. Bibliografía	118
ANEXOS	121
ANEXO DE ENTREVISTAS COMPLETAS	121
Entrevista Alberto de Santos	121
Entrevista José María Martín-Mateos	123

Índice de Figuras

Ilustración 1. Captura de pantalla de la web del Instituto Nacional de Estadística, en específico sobre el apartado de datos sobre sociedades mercantiles.	24
Ilustración 2. Captura de pantalla de la web del epdata, en concreto, su página principal de inicio.....	25
Ilustración 3. Captura de pantalla de los diagramas de dispersión correspondientes a las evoluciones de las sociedades Constituidas y disueltas en el sector de la hostelería desde el año 2012 al 2022.	41
Ilustración 4. Captura de pantalla del diagrama de dispersión correspondiente a la evolución de sociedades de forma jurídica “S.L.” constituidas en Madrid desde el año 2000 al 2022..	42
Ilustración 5. Captura de pantalla del diagrama de cajas correspondiente a la distribución del valor indicador “ICEA” según la comunidad autónoma	42
Ilustración 6. Captura de pantalla del histograma correspondiente a la evolución del número de sociedades disueltas voluntariamente en 2021 en todas las comunidades autónomas españolas.	43
Ilustración 7. Resultados del test de Wilcoxon para las disoluciones empresariales.....	80
Ilustración 8. Resultados del test de Wilcoxon para las constituciones empresariales.....	80
Ilustración 9. Resultados del test de Friedman para disoluciones empresariales	82
Ilustración 10. Resultados del test de Friedman para constituciones empresariales.....	82
Ilustración 11. Gráfico de la matriz de dispersión para variables del modelo 3.....	83
Ilustración 12. Gráfico de dispersión de variables específicas, en el contexto de constituciones empresariales.	84
Ilustración 13. Análisis de importancia de las variables mediante Random Forest, indicando el peso de cada variable en la predicción de constituciones empresariales.	86
Ilustración 14. Gráfico de residuos correspondiente al modelo1.....	87
Ilustración 15. Gráfico de residuos correspondiente al modelo2.....	88
Ilustración 16. Gráfico de residuos correspondiente al modelo3.....	88

Índice de Tablas

Tabla 1. Variables de LIMP.BBDD 1 clasificadas por tipo de variable e información sobre ellas	27
Tabla 2. Variables de LIMP.BBDD 2 clasificadas por tipo de variable e información sobre ellas	28
Tabla 3. Variables de LIMP.BBDD 3 clasificadas por tipo de variable e información sobre ellas	28
Tabla 4. Variables de LIMP.BBDD 5 clasificadas por tipo de variable e información sobre ellas	28
Tabla 5. Variables de LIMP.BBDD 6 y 7 clasificadas por tipo de variable e información sobre ellas.....	29
Tabla 6. Variables de LIMP.BBDD 8 clasificadas por tipo de variable e información sobre ellas	29
Tabla 7. Medidas de tendencia central de la primera base de datos.....	37
Tabla 8. Medidas de tendencia central de la segunda base de datos.	37
Tabla 9. Medidas de tendencia central de la tercera base de datos.....	38
Tabla 10. Medidas de tendencia central de la cuarta base de datos.	38
Tabla 11. Medidas de tendencia central de la quinta base de datos.	38
Tabla 12. Medidas de tendencia central de la sexta y séptima base de datos.....	38
Tabla 13. Medidas de dispersión de la primera base de datos.	39
Tabla 14. Medidas de dispersión de la segunda base de datos.....	39
Tabla 15. Medidas de dispersión de la tercera base de datos.	39
Tabla 16. Medidas de dispersión de la cuarta base de datos.	40
Tabla 17. Medidas de dispersión de la quinta base de datos.....	40
Tabla 18. Medidas de dispersión de la sexta base de datos.....	40
Tabla 19. Tabla comparativa de resultados para los modelos de disoluciones.....	76
Tabla 20. Tabla comparativa de resultados para los modelos de constituciones.	77
Tabla 21. Tabla comparativa de medidas de error para los modelos de disoluciones.....	78
Tabla 22. Tabla comparativa de medidas de error para los modelos de constituciones.	79

INTRODUCCIÓN

Descripción del Problema

Este Trabajo de Fin de Grado se centra en explorar el comportamiento del sistema empresarial español, con un interés particular en las constituciones y disoluciones empresariales. El problema de investigación radica en entender cómo diversas variables macroeconómicas pueden influir en estas dinámicas empresariales, lo cual es crucial para identificar patrones y factores de riesgo o éxito en el contexto empresarial actual. El estudio busca investigar el impacto que estos factores pueden tener en la estabilidad y el desarrollo de las empresas, proporcionando una base para futuras políticas y estrategias que mejoren la salud empresarial en España.

Motivación del Trabajo

El impulso para desarrollar este Trabajo de Fin de Grado se origina en la necesidad de entender las consecuencias y los cambios profundos provocados por la pandemia de COVID-19 en el tejido empresarial español. Aunque el punto de partida fue el análisis del impacto directo de esta crisis sanitaria global, el enfoque del estudio se amplió significativamente debido a los desafíos encontrados en la disponibilidad y la adecuación de los datos relacionados específicamente con la pandemia.

A medida que avanzaba la fase inicial del proyecto, se hizo evidente que una exploración efectiva del impacto de la pandemia requeriría un marco más robusto y amplio para captar las dinámicas subyacentes que influyen tanto las disoluciones como las constituciones empresariales en España. Este reenfoque llevó a incorporar un análisis más amplio de variables macroeconómicas, que proporcionan una perspectiva más rica y diversificada sobre cómo factores como el PIB, el ICEA y otros indicadores económicos y sociales moldean el panorama empresarial.

Este cambio de dirección en la investigación no solo responde a las limitaciones de datos específicos de la pandemia, sino que también refleja un compromiso con una comprensión más holística de las condiciones empresariales. Al expandir el alcance del estudio para incluir una variedad más amplia de variables macroeconómicas, la investigación aborda cómo estos factores afectan la estabilidad y la sostenibilidad de las empresas bajo circunstancias normales y en tiempos de crisis.

Así, este trabajo se propone como un esfuerzo académico para proporcionar un análisis detallado y riguroso de la influencia de amplios factores económicos y gubernamentales en las dinámicas empresariales. Se espera que los hallazgos no solo enriquezcan la literatura académica sobre la respuesta empresarial en contextos de crisis, sino que también ofrezcan orientación práctica para la formulación de políticas y estrategias que mejoren la resiliencia y la adaptabilidad del sector empresarial español en el futuro.

Organización de la Memoria

La organización de esta memoria del Trabajo de Fin de Grado se estructura en tres capítulos fundamentales, cada uno enfocado en una fase crucial del análisis del sistema empresarial español en el contexto de las variables macroeconómicas y sus efectos sobre las constituciones y disoluciones empresariales.

El primer capítulo, titulado "Introducción a la Ingeniería del Dato", detalla el proceso exhaustivo de Extracción, Transformación y Carga (ETL) de datos. Este capítulo es esencial para comprender cómo se prepararon y aseguraron los datos para el análisis, garantizando su calidad y relevancia para explorar el impacto de factores como la COVID-19 y las medidas gubernamentales en el tejido empresarial español. Además, se discuten las metodologías empleadas para la selección de variables y el almacenamiento seguro de los datos procesados, proporcionando una base sólida para los análisis estadísticos descriptivos y avanzados realizados posteriormente.

El segundo capítulo, "Introducción al Análisis de Datos", se centra en la aplicación de técnicas de análisis para evaluar cómo distintas variables macroeconómicas influyen las disoluciones y constituciones empresariales. Este segmento aborda la selección de modelos analíticos adecuados para identificar patrones y correlaciones significativas entre los datos, evaluando la precisión y la confiabilidad de estos modelos a través de diversas medidas estadísticas. Se destaca el uso de visualizaciones para interpretar los resultados y facilitar la comprensión de las dinámicas complejas que afectan al sector empresarial.

Finalmente, el tercer capítulo, "Análisis de Negocio", integra los conocimientos y datos obtenidos para formular recomendaciones estratégicas y proporcionar insights prácticos. Este capítulo combina el análisis técnico con perspectivas del negocio, enfocándose en cómo los hallazgos pueden influir en las decisiones empresariales y políticas. Además, se describe la interacción con expertos del sector para enriquecer el análisis y validar los resultados obtenidos, culminando en una propuesta detallada de estrategias para fomentar la estabilidad y el crecimiento en el ecosistema empresarial español.

Cada capítulo contribuye a un entendimiento holístico del tema, asegurando que la memoria no solo cumpla con rigor académico, sino que también aborde aplicaciones prácticas relevantes para el entorno empresarial actual.

Estado del Arte

La evolución del tejido empresarial en España constituye un fascinante campo de estudio, especialmente al considerar cómo ha respondido y se ha adaptado a una serie de contextos críticos a lo largo de su historia. Este análisis no solo revela la resiliencia y la capacidad de adaptación de las empresas españolas, sino que también proporciona una ventana a la comprensión de la dinámica económica y social del país en diferentes épocas. A lo largo de los siglos, desde la industrialización en el siglo XIX hasta los desafíos contemporáneos como la crisis financiera global y la pandemia de COVID-19, el tejido empresarial español ha experimentado transformaciones significativas que reflejan tanto las influencias internas como las externas.

En el siglo XIX, España se embarcó en un camino hacia la industrialización, marcando un cambio decisivo en su estructura económica y empresarial caracterizado por la adopción de innovaciones tecnológicas y cambios socioeconómicos y políticos profundos. Este proceso no fue homogéneo ni lineal, presentando variaciones regionales y sectoriales notables.

Uno de los estudios clave en este ámbito es el de Nadal (1975), quien analiza cómo la industrialización en España se vio influenciada por factores como la disponibilidad de recursos naturales, la infraestructura de transporte y las políticas gubernamentales. A diferencia de otros países europeos, la industrialización en España comenzó más tarde y avanzó a un ritmo más lento. Esto se debió en parte a la estructura agraria del país, la cual retrasó el desarrollo de un sector industrial robusto.

La revolución industrial en España también estuvo marcada por importantes innovaciones tecnológicas, especialmente en la industria textil y metalúrgica. Tortella (1994) destaca cómo la introducción de nuevas máquinas y procesos de producción transformó estas industrias, aunque la adopción de tecnología fue desigual en diferentes regiones del país. Por ejemplo, Cataluña se convirtió en un centro importante para la industria textil, mientras que el País Vasco lideró en la producción de hierro y acero. Además, la industrialización estuvo estrechamente ligada a cambios socioeconómicos significativos. Ringrose (1996) señala que el crecimiento de la industria trajo consigo un cambio en la distribución de la población, con un movimiento creciente de personas del campo a las ciudades en busca de trabajo en las fábricas. Esto no solo cambió el paisaje urbano de España, sino que también tuvo un impacto relevante en la estructura social y económica del país.

En el ámbito político, la industrialización coincidió con un periodo de inestabilidad y cambios. Carr (1966) argumenta que las tensiones políticas del siglo XIX, incluyendo las guerras carlistas y los conflictos entre liberales y conservadores, tuvieron un impacto directo en el desarrollo económico del país. Las políticas gubernamentales en áreas como el comercio, la inversión extranjera y la regulación industrial jugaron un papel crucial en la configuración del proceso de industrialización.

En resumen, la industrialización del siglo XIX en España fue un proceso complejo influenciado por una variedad de factores tecnológicos, socioeconómicos y políticos. Aunque su desarrollo fue más lento y menos uniforme en comparación con otros países europeos, sentó las bases para el crecimiento económico y los cambios sociales en los siglos siguientes.

El siglo XX, sin embargo, presentó desafíos únicos, especialmente con la Guerra Civil Española y la posterior dictadura de Franco. Estos eventos no solo alteraron la estructura económica del país, sino que también impusieron políticas y restricciones que tuvieron un impacto duradero en las empresas españolas.

Durante la Guerra Civil, la economía española sufrió severamente debido a la destrucción de infraestructuras y la interrupción de las actividades comerciales y productivas. Preston (1996) describió cómo la guerra desencadenó una crisis económica aguda, caracterizada por la escasez de materias primas y bienes de consumo, así como por la disrupción de los mercados. Las empresas tuvieron que enfrentarse a enormes desafíos para mantener sus operaciones en un entorno tan inestable.

Tras el conflicto, la dictadura de Franco implementó un régimen autárquico que buscaba la autosuficiencia económica de España. Este modelo, como señaló Viñas (1984), se caracterizó por una fuerte intervención estatal en la economía, incluyendo la nacionalización de industrias clave y la imposición de estrictos controles sobre la producción, el comercio y los precios. Estas políticas limitaron significativamente la capacidad de las empresas para operar en un mercado libre y restringieron su acceso a mercados internacionales.

La autarquía también condujo a la creación de un complejo sistema de licencias y permisos que afectó a todas las áreas de la actividad empresarial. Este sistema burocrático no solo obstaculizó el crecimiento empresarial, sino que también fomentó la corrupción y el favoritismo. La falta de competencia y la protección a ciertas industrias formaron un entorno empresarial ineficiente y poco competitivo. Payne (1993) destaca cómo la industria pesada, como la minería y la siderurgia, recibió un apoyo considerable, mientras que sectores como

la agricultura y los servicios fueron relativamente descuidados. Esto llevó a desequilibrios en el desarrollo económico y agravó las disparidades regionales en España.

La dictadura también tuvo un impacto significativo en las relaciones laborales. La supresión de sindicatos y la imposición de una estricta regulación laboral limitaron los derechos de los trabajadores y contribuyeron a un clima de tensión y descontento en el ámbito laboral.

En conclusión, la Guerra Civil y la dictadura de Franco tuvieron un impacto importante y prolongado en el tejido empresarial español de manera que las políticas económicas y las restricciones comerciales impuestas durante este periodo no solo restringieron el crecimiento y la competitividad de las empresas, sino que también moldearon de manera significativa la estructura económica y social de España.

La transición de España a la democracia en la década de 1970 y su posterior integración en la Unión Europea (UE) en 1986 marcaron un período de transformación económica y empresarial significativa. Estos cambios, fundamentales en la historia reciente de España, impulsaron una liberalización económica y una apertura hacia mercados internacionales, creando un impacto profundo en la expansión y diversificación de las empresas españolas.

La democratización trajo consigo una serie de reformas económicas que buscaban modernizar la economía española. Estas reformas incluyeron la desregulación de los mercados, la privatización de empresas estatales y la liberalización de los flujos de capital. Guillén (1997) argumenta que este nuevo entorno económico fomentó un clima más propicio para el emprendimiento, la innovación y la inversión extranjera, lo que fue crucial para el desarrollo empresarial en España.

La integración en la UE fue un factor clave en este proceso de transformación. Como señala Holman (1996), la adhesión a la Unión no solo significó el acceso a un mercado mucho más amplio, sino también la adopción de políticas y normativas comunes europeas. Esto incentivó a las empresas españolas a adaptarse a estándares más competitivos, mejorando su eficiencia y capacidad para competir en el mercado global. Sectores como el turismo, la construcción, la banca y las telecomunicaciones experimentaron un crecimiento notable, y las empresas españolas comenzaron a consolidarse en mercados internacionales.

A su vez, este período de transformación también presentó desafíos. La rápida integración en la economía europea y global expuso a las empresas españolas a una competencia más intensa, requiriendo ajustes significativos en términos de calidad,

innovación y eficiencia operativa. Royo (2000) destaca que la transición económica llevó a cambios en el mercado laboral, exigiendo una fuerza laboral más cualificada y flexible, lo que a su vez planteó desafíos en términos de formación y adaptación de los trabajadores.

Finalizando esta etapa, se puede decir que la transición a la democracia y la integración en la UE marcaron el inicio de una era de modernización y expansión para el tejido empresarial español. Estos cambios no solo transformaron la estructura económica del país, sino que también establecieron las bases para su crecimiento y desarrollo en las décadas siguientes.

La siguiente etapa que destacar es la crisis financiera global de 2008, la cual tuvo un impacto profundo y duradero en la economía española, afectando particularmente a las pequeñas y medianas empresas (PYMES). Este período crítico se caracterizó por una serie de desafíos económicos, incluyendo una mayor vulnerabilidad financiera, disminución de la rentabilidad y un aumento significativo en los niveles de endeudamiento de las empresas.

La crisis, originada por el colapso del mercado inmobiliario y la crisis de liquidez en los mercados financieros internacionales, rápidamente se propagó a la economía real, afectando a numerosos sectores. En España, las PYMES, que constituyen una parte significativa del tejido empresarial, se vieron especialmente afectadas. Estas empresas enfrentaron dificultades para acceder a financiamiento debido a la restricción crediticia impuesta por las entidades bancarias, que se encontraban en una situación de incertidumbre y riesgo. Según Molina y Mora (2010), la restricción del crédito fue uno de los factores clave que exacerbó la vulnerabilidad de las PYMES durante la crisis.

Además, la crisis financiera llevó a una disminución significativa en la demanda tanto a nivel nacional como internacional, lo que resultó en una reducción de las ventas y la rentabilidad de las empresas. Como señala Saurina (2009), esta situación se vio agravada por la dependencia de muchas PYMES españolas del mercado interno, que experimentó una contracción considerable. La combinación de una menor demanda y dificultades de financiamiento puso a muchas empresas en una posición precaria, llevando a un aumento en las tasas de insolvencia y cierre de negocios. El endeudamiento de las PYMES se convirtió en un problema crítico durante este período. Muchas empresas que habían tomado préstamos durante los años de bonanza económica se encontraron con dificultades para cumplir con sus obligaciones financieras en un contexto de recesión. Según Ruesga y Da Silva (2011), el alto nivel de endeudamiento limitó la capacidad de las empresas para invertir y adaptarse a las nuevas condiciones del mercado, restringiendo su potencial de recuperación y crecimiento.

La crisis financiera de 2008 también tuvo implicaciones a largo plazo en el panorama empresarial español. La necesidad de reestructuración y adaptación llevó a muchas empresas a revisar sus modelos de negocio, buscar nuevos mercados y mejorar su eficiencia operativa. Este proceso de adaptación no fue fácil, pero fue crucial para la supervivencia y el futuro crecimiento de las PYMES en España.

A fin de cuentas, esta crisis representó un período de desafío significativo para las PYMES en España, marcado por una mayor vulnerabilidad financiera, disminución de la rentabilidad y un aumento en los niveles de endeudamiento. Este evento no solo afectó la economía a corto plazo, sino que también tuvo un impacto perdurable en la forma en que las empresas operan y se adaptan a un entorno económico cambiante.

El último acontecimiento importante en el que se han llevado a cabo estudios sobre el tema a tratar habla sobre la pandemia de COVID-19, que comenzó a principios de 2020, y que ha tenido un impacto sin precedentes en el tejido empresarial global, incluyendo a España. Este evento disruptivo ha afectado a empresas de todos los tamaños y sectores, destacando la importancia crítica de la innovación, la adaptabilidad y el apoyo gubernamental para la supervivencia y recuperación empresarial.

La pandemia trajo consigo desafíos inmediatos y severos para las empresas, incluyendo interrupciones en las cadenas de suministro, cierres obligatorios de negocios no esenciales y una caída drástica en la demanda de consumidores y empresas. En España, donde, como se cita anteriormente, el tejido empresarial está compuesto en gran medida por PYMES, el impacto fue particularmente agudo. Según un estudio de González et al. (2020), muchas de estas empresas enfrentaron dificultades financieras significativas debido a la disminución de ingresos y la incertidumbre del mercado.

En respuesta a estos desafíos, la innovación y la adaptabilidad se convirtieron en factores clave para la supervivencia empresarial. Muchas empresas tuvieron que reevaluar rápidamente sus modelos de negocio, adoptando nuevas tecnologías y estrategias para mantenerse a flote. Por ejemplo, la digitalización, que incluye la adopción de comercio electrónico, teletrabajo y soluciones de tecnología de la información, se aceleró significativamente durante la pandemia. Como señala Martínez (2021), esta adaptación tecnológica no solo fue crucial para la continuidad de las operaciones durante los cierres, sino que también está configurando el futuro del entorno empresarial.

El apoyo gubernamental también jugó un papel vital en la mitigación del impacto económico de la pandemia. En España, el gobierno implementó una serie de medidas para apoyar a las empresas, incluyendo moratorias fiscales, líneas de crédito y subsidios para

trabajadores afectados por ERTes (Expedientes de Regulación Temporal de Empleo). Estas medidas, como indica López-Cózar (2020), fueron esenciales para proporcionar un alivio financiero a las empresas y evitar una ola masiva de quiebras y despidos.

En definitiva, la pandemia de COVID-19 ha sido un catalizador de cambio para las empresas en España, resaltando la importancia de la innovación, la adaptabilidad y el apoyo gubernamental. Aunque el camino hacia la recuperación puede ser largo y desafiante, las lecciones aprendidas durante esta crisis serán fundamentales para el futuro del tejido empresarial español.

El presente estado del arte ha explorado la evolución del tejido empresarial español a través de diversos contextos históricos, culminando en el impacto de la pandemia de COVID-19. Los hallazgos subrayan la importancia crucial de la adaptabilidad y la innovación en las empresas españolas, especialmente en tiempos de crisis. La capacidad de adaptarse rápidamente a circunstancias cambiantes y la habilidad para innovar han demostrado ser factores determinantes para la supervivencia y el crecimiento empresarial.

En el contexto de la pandemia, estas cualidades han sido más relevantes que nunca. Las empresas han tenido que navegar por un panorama de incertidumbre, ajustándose a nuevas realidades de mercado y aprovechando las ayudas e incentivos gubernamentales disponibles. Este Trabajo de Fin de Grado se enfocará en analizar cómo la pandemia ha impactado específicamente en el tejido empresarial español, considerando las medidas gubernamentales y los factores críticos que han influenciado la capacidad de creación y desarrollo de las empresas.

OBJETIVOS Y EVOLUCIÓN

En este capítulo, se discutirá cómo los objetivos originales relacionados con el impacto de la pandemia en el ecosistema empresarial español se transformaron hacia un enfoque más amplio y detallado en el análisis de cómo diversas variables macroeconómicas influyen en la constitución y disolución de empresas. Además, se detallará cómo se formularon los nuevos objetivos y preguntas de investigación, subrayando el proceso de adaptación y reajuste metodológico que ha guiado el desarrollo y la dirección del análisis de datos en esta fase crucial del estudio.

Evolución

La trayectoria del Trabajo de Fin de Grado (TFG) ha experimentado una notable evolución, reflejando un proceso de adaptación y refinamiento continuos que han estado vinculados con los desafíos y las oportunidades descubiertas a lo largo del camino investigativo. Este proceso se ha caracterizado por un cambio significativo en el enfoque del estudio, motivado por las circunstancias emergentes y los hallazgos preliminares durante las etapas iniciales de la investigación.

Inicialmente, el objetivo del TFG se centraba en explorar el impacto de la pandemia de COVID-19 en el ecosistema empresarial español. Las preguntas de investigación originales buscaban comprender cómo se había desarrollado este ecosistema antes de la pandemia, así como identificar los efectos de la crisis sanitaria en la tasa de creación de empresas y en la resiliencia empresarial. Sin embargo, esta fase inicial, denominada **Fase 1: Búsqueda**, reveló una significativa limitación en términos de disponibilidad y profundidad de los datos específicos relacionados con el COVID-19, lo que impedía un análisis exhaustivo y riguroso.

Ante estos obstáculos, el proyecto dio un giro metodológico en la **Fase 2: Análisis**, donde se amplió el marco de estudio para incluir una variedad de variables macroeconómicas que pudieran influir en el ecosistema empresarial más allá del contexto pandémico. Este ajuste permitió aprovechar un espectro más amplio de datos disponibles, facilitando así una exploración más rica y diversificada de los factores que moldean el panorama empresarial español.

La **Fase 3: Especificación**, marcó una etapa decisiva en la evolución del TFG, donde se definieron con precisión las variables de estudio, enfocándose específicamente en las constituciones y disoluciones empresariales como indicadores clave del comportamiento del ecosistema empresarial. Este enfoque más focalizado permitió una investigación más

detallada y especializada, orientada a descubrir las interacciones y dependencias entre las variables macroeconómicas y las tendencias empresariales.

A lo largo de estas fases, el TFG no solo transformó su dirección y alcance en respuesta a las limitaciones encontradas, sino que también se adaptó a los métodos de análisis más apropiados para abordar la nueva orientación del estudio. La adaptabilidad y la evolución del enfoque del TFG destacan la importancia de la flexibilidad en la investigación académica, especialmente en campos tan dinámicos y condicionados por factores externos como lo es el análisis empresarial y económico.

Este proceso evolutivo del TFG ha demostrado ser una parte integral del aprendizaje y el desarrollo académico del investigador, proporcionando una base sólida para la formulación de conclusiones significativas y recomendaciones bien fundamentadas, que serán detalladas en las secciones subsiguientes de este trabajo.

Objetivos Planteados

Tras la evolución inicial y las limitaciones encontradas en las etapas tempranas del Trabajo de Fin de Grado, los objetivos de investigación fueron recalibrados para alinearse con un enfoque más robusto y fundamentado en el nuevo tema de investigación de variables macroeconómicas. Esta reorientación ha permitido establecer un conjunto de preguntas de exploración del tema que no solo reflejan una continuidad con las preocupaciones originales respecto a la pandemia, sino que también expanden el alcance para abarcar una perspectiva más integral del ecosistema empresarial español en contextos de variables y crisis macroeconómicas.

Inicialmente, el Trabajo de Fin de Grado se centró en explorar cómo se había desarrollado el ecosistema empresarial español antes y después del impacto de la pandemia de COVID-19. Las preguntas de investigación originales buscaban entender el comportamiento del ecosistema empresarial bajo estas circunstancias específicas:

- *¿Cómo se desarrollaba el ecosistema empresarial español antes de la pandemia?*
- *¿Ha afectado la pandemia a la tasa de creación de empresas en España en comparación con el periodo anterior a la crisis sanitaria?*
- *¿Qué variables influyen a la creación y desarrollo de estas empresas?*
- *¿Cómo han respondido las empresas españolas ante la incertidumbre generada por la pandemia, y cuáles son los factores clave que han contribuido a su resiliencia?*

Sin embargo, durante la fase inicial de investigación, se llegó a la conclusión de que no había suficientes datos relevantes para abordar estas cuestiones de manera satisfactoria. Esto llevó a una reorientación de los objetivos hacia un análisis más amplio y fundamentado en variables macroeconómicas más generales, estableciendo así un nuevo conjunto de preguntas de investigación:

- *¿Cómo se ha desarrollado el ecosistema empresarial español en función del PIB e IPC, y qué implicaciones tienen estos indicadores en la estructura empresarial del país?*
- *¿Han afectado las variables macroeconómicas a la tasa de creación y disolución de empresas en España?*
- *¿Qué variables específicas influyen en la constitución y disolución de empresas?*
- *¿Cómo han respondido las empresas españolas ante las incertidumbres generadas por crisis macroeconómicas y cuáles son los factores clave para su resiliencia?*

Para responder a estas preguntas, el TFG adoptará una metodología que integra el análisis riguroso de datos con insights derivados de diversas fuentes. La estrategia analítica no se limitará únicamente a la exploración de bases de datos y la aplicación de modelos estadísticos, sino que también se enriquecerá con la perspectiva de expertos a través de entrevistas y consultas con profesionales del sector. Además, se realizará una revisión exhaustiva de literatura relevante, tanto de publicaciones académicas como de fuentes confiables en internet, lo que permitirá contextualizar los hallazgos dentro de un marco teórico más amplio y conectarlos con tendencias y estudios internacionales.

Se pretende que este enfoque multidimensional no solo responda a las preguntas reformuladas, sino que también proporcione una visión comprensiva de cómo diversos factores económicos y políticas gubernamentales interaccionan para moldear la realidad empresarial del país. La finalidad es generar un conjunto de recomendaciones prácticas y fundamentadas que sean de utilidad para emprendedores, empresarios, y formuladores de políticas, enfocadas en mejorar la resiliencia y fomentar el crecimiento sostenible del ecosistema empresarial en contextos de crisis.

PRIMERA PARTE: INGENIERÍA DEL DATO

1. Introducción a la ingeniería del dato

En el presente documento, se detalla el meticuloso proceso de Extracción, Transformación y Carga (ETL) llevado a cabo como parte de mi Trabajo de Fin de Grado, centrado en la evolución del tejido empresarial español antes y después del COVID-19. Este proceso es esencial para garantizar la calidad y la utilidad de los datos en el análisis posterior, permitiendo una comprensión más profunda y precisa de como analizar de manera integral el impacto de la pandemia de la COVID-19 en el tejido empresarial español, con un enfoque particular en como las medidas gubernamentales han podido influir en la capacidad de creación y desarrollo de empresas.

Una vez recopilados, los datos pasaron por un riguroso proceso de transformación. Se limpiaron, se estructuraron y se prepararon para el análisis, prestando especial atención a la precisión del formato, la corrección de valores atípicos y la imputación de valores faltantes. Este documento describe en detalle el número de observaciones y variables, los tipos de datos y la periodicidad, proporcionando una base sólida para el análisis estadístico y descriptivo.

La elección de variables y periodos se justifica en el contexto de los objetivos del TFG, buscando responder preguntas clave y explorar hipótesis específicas. Se emplearon medidas de tendencia central y dispersión para resumir los datos, mientras que los gráficos descriptivos, como diagramas de dispersión y cajas, facilitan la visualización de tendencias y relaciones.

Finalmente, se explica el método de almacenamiento de datos post-ETL, asegurando que estén disponibles para análisis futuros. Este documento culmina con un resumen narrativo que integra todas las observaciones y hallazgos, proporcionando una visión completa del proceso ETL y su impacto en el TFG.

2. Origen de los Datos

Los datos, elemento clave de este estudio, fueron obtenidos de fuentes fiables y reconocidas, específicamente del Instituto Nacional de Estadística (INE) y de epdata. A través de métodos como la descarga directa, se recopilaron conjuntos de datos que proporcionan una visión integral del tema a tratar. Este documento incluye capturas de pantalla y descripciones detalladas del proceso de obtención de datos.

El INE, como organismo oficial en España, proporciona una amplia gama de datos estadísticos que abarcan diversos aspectos socioeconómicos y demográficos del país. La información obtenida es crucial para comprender las tendencias y patrones a nivel nacional, ofreciendo una perspectiva integral y actualizada. Por otro lado, epdata, como plataforma de divulgación de datos y estadísticas, complementa esta información con visualizaciones y análisis detallados, facilitando la interpretación y el entendimiento de los datos.

La metodología empleada para la obtención de los datos de ambas fuentes ha sido la descarga directa. Técnica que consiste simplemente en acceder a los portales web de las respectivas fuentes y descargar los conjuntos de datos en formatos csv y archivos Excel. Este método asegura que los datos se mantengan en su formato original y sin alteraciones, preservando su integridad. Además, la descarga directa es un proceso transparente y reproducible, aspectos esenciales en la investigación académica.



Ilustración 1. Captura de pantalla de la web del Instituto Nacional de Estadística, en específico sobre el apartado de datos sobre sociedades mercantiles.



Ilustración 2. Captura de pantalla de la web del epdata, en concreto, su página principal de inicio

La autenticidad y fiabilidad de las fuentes de datos son indiscutibles. El INE, como institución gubernamental, sigue rigurosos protocolos para la recopilación y publicación de datos, asegurando su precisión y actualidad. Por su parte, epdata, al basar sus análisis y visualizaciones en fuentes oficiales y reconocidas, proporciona una capa adicional de verificación y contexto a los datos. Esta combinación de fuentes asegura una base de datos sólida y confiable para el análisis subsiguiente en este Trabajo de Fin de Grado.

3. Descripción de los Datos

3.1. Número total de observaciones y variables en los datasets (limpias)

En el marco de esta primera parte del Trabajo de Fin de Grado, se ha llevado a cabo un exhaustivo proceso de análisis y limpieza de datos, partiendo de un conjunto inicial de nueve bases de datos, denominadas "BBDD X", donde "X" representa el número asignado a cada una. A lo largo de este proceso, algunas bases de datos, específicamente "BBDD 3" y "BBDD 9", fueron descartadas debido a criterios de relevancia y calidad de los datos. Las bases de datos restantes, tras ser sometidas a un riguroso proceso de limpieza y transformación, han proporcionado un conjunto de datos depurado y estructurado, listo para el análisis posterior.

La descripción detallada de las bases de datos limpias, ahora denominadas "LIMP.BBDD X", es la siguiente:

- a) **LIMP.BBDD 1:** Compuesta por 9,798 observaciones, esta base de datos contiene 8 variables significativas: Column1, Comunidad Autónoma, Forma jurídica, Sociedades o Capital desembolsado, Periodo, Número Sociedades/Capital, y Población. Estas variables abarcan aspectos clave como la ubicación geográfica, la naturaleza jurídica de las sociedades, así como información financiera y demográfica relevante.
- b) **LIMP.BBDD 2:** Con un total de 242 observaciones, esta base incluye 5 variables: Column1, Estados Sociedades, Actividad económica, Año, y Número Sociedades. Estas variables proporcionan una visión detallada del estado de las sociedades en diferentes sectores económicos a lo largo del tiempo.
- c) **LIMP.BBDD 4:** Esta base de datos consta de 627 observaciones y 5 variables: Column1, Clase de disolución, Año, Comunidad Autónoma, y Número Sociedades. Ofrece una perspectiva sobre la disolución de sociedades en diversas comunidades autónomas y su clasificación temporal.
- d) **LIMP.BBDD 5:** Con 170 observaciones, incluye 5 variables: Column1, Indicador, Comunidad Autónoma, Año, y Valor Indicador. Se centra en indicadores económicos clave por comunidad autónoma y año.
- e) **LIMP.BBDD 6 y 7:** Estas bases de datos, con 124 observaciones, contienen 6 variables: Column1, Actividad Económica, Comunidad Autónoma, Tipo de ERTE, Número de ERTES, y Año. Proporcionan información valiosa sobre los

Expedientes de Regulación Temporal de Empleo (ERTE) en diferentes sectores y regiones.

- f) **LIMP.BBDD 8:** Con 669 observaciones, esta base de datos se compone de 3 variables: Column1, Fecha, y Afiliados a la Seguridad Social. Ofrece datos sobre la afiliación a la seguridad social en diferentes fechas, lo que permite analizar tendencias en el empleo.

Cada una de estas bases de datos ha sido sometida a un proceso de limpieza y transformación para asegurar la calidad y coherencia de los datos. Este proceso incluyó la estandarización de formatos, la corrección de errores, y la eliminación de datos irrelevantes o redundantes. La transformación aplicada a cada base de datos será detallada en el apartado correspondiente, donde se describirán las características originales de los datos y las modificaciones realizadas.

3.2. Descripción del tipo de datos de cada variable

En el desarrollo de este Trabajo de Fin de Grado, se ha realizado un análisis detallado de varias bases de datos, cada una con sus características y tipos de datos específicos. A continuación, se presenta una descripción exhaustiva de las variables contenidas en cada una de las bases de datos limpias, denominadas "LIMP.BBDD X", donde "X" representa el número asignado a cada base de datos.

Tabla 1. Variables de LIMP.BBDD 1 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación.	Desde 0 hasta 9797.
Comunidad Autónoma	String	Nombre de las Comunidades Autónomas de España con prefijo numérico.	Incluye el código numérico de la comunidad.
Forma jurídica	String	Tipo de forma jurídica de las sociedades.	Valores: "S.A.", "S.L.", "S.COM.", "S.COM. P.A. y S.C.".
Sociedades o Capital desembolsado	String	Indica si la observación se refiere al número de sociedades o al capital desembolsado.	Valores posibles: "Número de Sociedades", "capital (en miles de euros) desembolsado".
Año	Integer	Año de la observación, desde 2000 hasta 2022.	
Número Sociedades/Capital	Integer	Número de sociedades o cantidad de capital desembolsado.	Depende de la variable "Sociedades o Capital desembolsado".
Población	Integer	Población correspondiente.	Según el año y la CCAA.

NOTA: "S.A.": Sociedad Anónima, "S.L.": Sociedad de Responsabilidad Limitada, "S.COM.", "S.COM. P.A. y S.C.": Sociedad Comanditaria, Sociedad Comanditaria por Acciones y Sociedad Colectiva

Tabla 2. Variables de LIMP.BBDD 2 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación, desde 0 hasta 242.	Índice de las filas.
Estados Sociedades	String	Estado de las sociedades.	Valores: "Constituidas", "Disueltas".
Actividad económica	String	Actividad económica según la CNAE-2009.	
Año	Integer	Desde 2012 hasta 2022.	
Número Sociedades	Integer	Número de sociedades según estado, actividad y año.	

NOTA: El CNAE-2009 (Clasificación Nacional de Actividades Económicas) es una clasificación que agrupa actividades económicas para fines estadísticos y administrativos.

Tabla 3. Variables de LIMP.BBDD 3 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación, desde 0 hasta 627.	Índice de las filas.
Clase de disolución	String	Motivo de la disolución de las empresas.	Valores posibles: "Voluntaria", "Por fusión", "Otras".
Comunidad Autónoma	String	Nombre de las Comunidades Autónomas de España.	Incluye el código numérico de la comunidad.
Año	Integer	Año de la observación, desde 2012 hasta 2022.	
Número Sociedades	Integer	Número de sociedades disueltas según clase, año y comunidad.	

Tabla 4. Variables de LIMP.BBDD 5 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación, desde 0 hasta 170.	Índice de las filas.
Indicador	String	Índice de confianza empresarial armonizado.	Solo toma el valor "ICEA".
Comunidad Autónoma	String	Nombre de las Comunidades Autónomas de España con prefijo numérico.	Incluye el código numérico de la comunidad.
Año	Integer	Año de la observación, desde 2013 hasta 2022.	
Valor Indicador	Integer	Valor del índice de confianza empresarial armonizado.	

NOTA: El ICEA (Índice de Confianza Empresarial Armonizado) es una medida que evalúa la confianza de las empresas en la economía y su capacidad para tomar decisiones de inversión.

Tabla 5. Variables de LIMP.BBDD 6 y 7 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación, desde 0 hasta 124.	Índice de las filas.
Actividad económica	String	Actividad económica según la CNAE-2009.	Incluye "Total Actividades" para representar todas las actividades.
Comunidad Autónoma	String	Nombre de las 19 Comunidades Autónomas de España con prefijo numérico y "20 Nacional".	"20 Nacional" representa el conjunto de comunidades autónomas.
Tipo de ERTE	String	Tipo de Expediente de Regulación Temporal de Empleo.	Valores posibles: "Total", "Parcial".
Número de ERTES	Integer	Número de ERTES según actividad, año y comunidad.	
Año	Integer	Año de la observación, desde 2013 hasta 2022.	

NOTA: Los ERTE (Expedientes de Regulación Temporal de Empleo) son medidas laborales que permiten a las empresas suspender temporalmente los contratos o reducir la jornada de empleados en situaciones excepcionales.

Tabla 6. Variables de LIMP.BBDD 8 clasificadas por tipo de variable e información sobre ellas

Nombre	Tipo de Dato	Descripción	Información Adicional
Column1	Integer	Número de observación, desde 0 hasta 170.	Índice de las filas.
Fecha	Date	Fecha de la observación en formato "dd/mm/aa".	Desde "01/03/2020" hasta "30/12/2021".
Afiliados a la Seguridad Social	Integer	Número de afiliados a la seguridad social según la fecha.	Datos diarios de la evolución de los ERTES.

3.3. Formato de los datos en bruto y cualquier transformación aplicada.

El proceso de transformación y limpieza de datos es un paso crucial en cualquier análisis estadístico. A continuación, se detalla cómo se han transformado los datos en bruto de varias bases de datos a formatos más estructurados y analíticamente útiles, describiendo las operaciones realizadas y añadiendo capturas de pantalla del código utilizado.

3.3.1) Base Datos 1:

La primera base de datos, denominada "BBDD 1", originalmente contenía un total de 19.872 filas y una serie de variables críticas para el análisis. Estas variables incluían "Total Nacional", "Comunidades y Ciudades Autónomas", "Provincias", "Forma Jurídica", "Número de sociedades y capital (en miles de euros)", "Periodo" y "Total".

El proceso de limpieza y transformación de "BBDD 1" fue meticuloso y se llevó a cabo con el objetivo de optimizar la calidad de los datos para análisis posteriores. Los pasos seguidos en este proceso fueron los siguientes:

1. *Eliminación de la Columna "Total Nacional"*: Esta columna fue removida del conjunto de datos, ya que no aportaba información relevante para el análisis específico que se pretendía realizar.
2. *Eliminación de Filas en Blanco en "Comunidades y Ciudades Autónomas" y la columna "Provincias"*: Se procedió a eliminar todas aquellas filas que no tenían datos asignados en las columnas de "Comunidades y Ciudades Autónomas" y la columna "Provincias", con el fin de depurar el conjunto de datos y centrarse en la información completa y útil.
3. *Eliminación de Filas con el Valor "Total" en "Forma Jurídica"*: Se eliminaron las filas que contenían el valor "Total" en la columna "Forma Jurídica", ya que este valor no era necesario para el análisis propuesto.
4. *Eliminación de Filas con el valor "capital (en miles de euros) suscrito" en la variable "Número de sociedades y capital (en miles de euros)"*: Se descartaron las filas que incluían este valor específico, enfocándose en datos más relevantes para el estudio.

Tras la limpieza, se procedió a enriquecer la base de datos con información adicional para permitir comparaciones estandarizadas en el futuro. Se añadió una variable que representaba la población de cada comunidad y ciudad autónoma en cada periodo. Para ello, se utilizaron dos bases de datos adicionales: una que contenía los datos de población por comunidad autónoma desde el año 2000 hasta 2021, y otra con los datos correspondientes al año 2022. El resultado fue una base de datos unificada que incluía tres columnas esenciales: "Comunidades y Ciudades Autónomas", "Periodo" y "Población".

3.3.2) Base Datos 2:

La segunda base de datos, conocida como "BBDD 2", constaba inicialmente de 3.102 filas y se centraba en cuatro variables fundamentales: "Estados Sociedades", "Actividad Económica", "Periodo" y "Total".

El proceso de limpieza y transformación de "BBDD 2" se realizó con el objetivo de adaptar los datos a las necesidades específicas del análisis y mejorar su usabilidad. Los cambios aplicados fueron los siguientes:

1. *Cambio de "Sociedades Constituidas" por "Constituidas" en "Estados Sociedades"*: Se modificó esta variable para simplificar la categorización y

hacerla más directa y comprensible. Este cambio implicó una estandarización en la terminología utilizada.

2. *Modificación de Nombres en "Actividad Económica"*: Se ajustaron los nombres de algunos valores dentro de esta variable para reflejar de manera más precisa las categorías de actividades económicas. Este paso fue crucial para garantizar la coherencia y la precisión para poder trabajar con las mismas variables en distintos data frames.
3. *Agrupación de Periodos por Año y Eliminación de Datos del 2023*: Los datos se reorganizaron para agruparlos por año, lo que permitió una visión más clara de las tendencias a lo largo del tiempo. Además, se eliminaron los datos correspondientes al año 2023, enfocándose en el periodo de tiempo más relevante para el estudio.
4. *Creación de una Función para Sumar el Total de Sociedades por Año*: Se desarrolló una función específica para calcular la suma total de sociedades para cada año. Esta función permitió obtener una visión agregada y simplificada de los datos, facilitando su análisis y la extracción de conclusiones.

Estos pasos de limpieza y transformación fueron fundamentales para preparar los datos de "BBDD 2" para un análisis más eficiente y efectivo. La base de datos resultante, "LIMP.BBDD 2", se convirtió en un recurso valioso para el proyecto, proporcionando información clara y estructurada sobre los estados de las sociedades y las actividades económicas a lo largo de los años.

3.3.3) Base Datos 4:

La cuarta base de datos, denominada "BBDD 4", originalmente contenía un total de 73.140 filas y se centraba en las variables "Clase de Disolución", "Provincias", "Periodo" y "Total".

El proceso de limpieza y transformación de "BBDD 4" se llevó a cabo con el objetivo de adaptar los datos a las necesidades específicas del análisis y mejorar su usabilidad. Los pasos seguidos en este proceso fueron los siguientes:

1. *Unión de DataFrames para Asignar Comunidades Autónomas a Provincias*: Se cargaron los archivos necesarios y se crearon dos dataframes. Posteriormente, se procedió a unirlos utilizando una columna común, con el fin de asignar la comunidad autónoma correspondiente a cada provincia. Este paso fue crucial para proporcionar un contexto geográfico más completo y facilitar análisis regionales más detallados.

2. *Agrupación de Periodos Mensuales por Años y Ordenación Descendente*: Los datos se reorganizaron para agruparlos por año, permitiendo una visión más clara de las tendencias a lo largo del tiempo. Además, se ordenaron de manera descendente para facilitar la visualización y el análisis de los datos más recientes.
3. *Cambio de Nombre de la Columna 'Periodo' a 'Año'*: Se modificó el nombre de esta columna para reflejar de manera más precisa que los datos estaban ahora organizados anualmente. Este cambio mejoró la claridad y la comprensión de la estructura temporal de los datos.
4. *Cambio de Nombre de la Columna 'Total' a 'Número Sociedades'*: Se renombró esta columna para proporcionar una descripción más explícita de su contenido, es decir, el número total de sociedades disueltas según la clasificación, el año y la comunidad autónoma.

3.3.4) Base Datos 5:

La quinta base de datos, conocida como "BBDD 5", inicialmente contenía 2.376 filas y se centraba en variables como "Total Nacional", "Comunidades y Ciudades Autónomas", "Principales Indicadores", "Periodo" y "Total".

El proceso de limpieza y transformación de "BBDD 5" se realizó con el objetivo de adaptar los datos a las necesidades específicas del análisis y mejorar su usabilidad. Los cambios aplicados fueron los siguientes:

1. *Eliminación de la Columna "Total Nacional"*: Esta columna fue removida del conjunto de datos, ya que no aportaba información relevante para el análisis específico que se pretendía realizar.
2. *Cambio de Nombre y Limpieza de "Comunidades y Ciudades Autónomas"*: Se modificó el nombre de esta variable para una mayor claridad y se eliminaron las filas que no tenían valores asignados, con el fin de depurar el conjunto de datos y centrarse en la información completa.
3. *Filtrado y Renombrado de "Principales Indicadores"*: Se filtraron los valores de esta variable para mantener solo el "Índice de confianza empresarial armonizado" (ICEA), eliminando los otros dos valores. Además, se cambió el nombre de la variable a "Indicador" para una mayor precisión y simplicidad.
4. *Transformación de "Periodo" de Cuatrimestres a Años Completos*: La variable "Periodo", originalmente representada en cuatrimestres, se transformó para reflejar años completos, lo que permitió una visión más clara y consolidada de los datos a lo largo del tiempo.

5. *Cambio de Nombre de "Total" y Cálculo de la Media de "Valor Indicador"*: Se renombró la columna "Total" y se calculó la media de "Valor Indicador" para cada combinación de "Indicador", "Comunidad Autónoma" y "Año". Posteriormente, se redondearon los números para que fueran enteros y se convirtió la columna a enteros para simplificar la presentación de los datos.

3.3.5) Base Datos 6 y 7:

Las bases de datos 6 y 7, inicialmente separadas, pero posteriormente fusionadas, presentaban un desafío único en términos de su estructura y contenido. A diferencia de las bases de datos estructuradas previamente analizadas, estas contenían información semiestructurada en archivos Excel, lo que requería un enfoque distinto para su procesamiento y limpieza.

BBDD 6 y 7 (Fusionadas):

- *Variables Iniciales*: Ambas bases de datos compartían un conjunto de variables que incluían "Evolución", "Tipo Erte y Suspensión", "Edad", "Tipo Contrato", "Sección CNAE", "Actividad CNAE", "Tipo Erte por CNAE", "Provincias", y "Tipo Erte Provincias y CCAA".

El proceso de limpieza y transformación de estas bases de datos implicó varios pasos clave:

1. *Selección de Variables Relevantes*: Se eligieron las variables más pertinentes para el análisis, enfocándose en aspectos como la actividad económica, la fecha, la comunidad autónoma, el tipo de ERTE y el número de ERTES. Esta selección se basó en la relevancia de estas variables para comprender la evolución y el impacto de los ERTES durante los años 2021 y 2022.
2. *Manejo de Datos Semiestructurados*: Dado que la información estaba dispersa en varias hojas y tablas dentro de los archivos Excel, se requirió un enfoque más manual para su organización. Se utilizaron herramientas y funciones de Excel, como fórmulas y filtros, para consolidar y estructurar los datos de manera coherente.
3. *Fusión de Datos de Diferentes Pestañas*: Se extrajeron datos de pestañas específicas de los archivos Excel, que incluían información detallada sobre los afiliados en ERTES según diversos criterios como el tipo de suspensión, la actividad económica (CNAE) y la distribución por sexo, provincia y comunidad autónoma.
4. *Creación de una Tabla Unificada*: Tras seleccionar y organizar los datos relevantes, se creó una tabla consolidada que integraba la información clave

de ambas bases de datos, proporcionando una visión completa y detallada de los ERTes en España durante el periodo de estudio.

3.3.6) Base Datos 8:

La base de datos 8, presentaba un conjunto de datos concentrado y específico, con 670 filas y variables como "Año", "Periodo" y "Afiliados a la Seguridad Social". Esta base de datos proporcionaba información valiosa sobre la afiliación a la seguridad social en diferentes periodos, crucial para el análisis de tendencias laborales y socioeconómicas. Los cambios aplicados fueron los siguientes:

1. *Cambio de Formato de la Variable "Periodo"*: Originalmente, la variable "Periodo" presentaba fechas en un formato que no era óptimo para el análisis. Por lo tanto, se modificó esta variable para representar las fechas en un formato más estándar y útil, específicamente 'dd/mm'. Este cambio facilitó la interpretación y el manejo de los datos temporales.
2. *Combinación de "Periodo" y "Año" en una Nueva Variable "Fecha"*: Para proporcionar una visión más integrada y coherente del tiempo, se combinaron las columnas "Periodo" y "Año" para formar una nueva columna denominada "Fecha". Este paso fue crucial para consolidar la información temporal en un único campo, simplificando el análisis posterior.
3. *Reducción del DataFrame a Dos Variables Esenciales*: Con el fin de enfocar el análisis en los aspectos más relevantes, se decidió mantener solo dos variables en el dataframe: "Fecha" y "Afiliados a la Seguridad Social". Esta decisión permitió centrar la atención en la evolución de la afiliación a la seguridad social a lo largo del tiempo, eliminando cualquier dato superfluo o redundante.

3.4. Periodicidad y temporalidad de los datos.

Un aspecto crucial de este análisis ha sido la consideración de la periodicidad y temporalidad de los datos recogidos, aspectos fundamentales para comprender las dinámicas y tendencias a lo largo del tiempo, especialmente en el contexto de la pandemia de COVID-19 y sus efectos sobre el tejido empresarial.

Las primeras cinco bases de datos (BBDD 1 a BBDD 5) presentan una periodicidad anual, con los datos organizados en formato de año (AAAA), lo que facilita el análisis de tendencias a largo plazo y permite una comparación directa entre los distintos años. La temporalidad de estos conjuntos de datos varía, abarcando distintos rangos temporales.

- *BBDD 1*: Esta base de datos abarca el periodo más extenso, desde el año 2000 hasta el 2022, ofreciendo una visión amplia de las dos décadas previas y actuales, lo que permite evaluar el impacto de la pandemia en un contexto temporal más amplio.
- *BBDD 2 y BBDD 3*: Ambas bases de datos cubren un periodo desde el año 2012 hasta el 2022, proporcionando datos cruciales para el análisis de las tendencias empresariales en la última década, incluyendo el periodo previo y durante la pandemia.
- *BBDD 4 y BBDD 5*: Estas bases de datos ofrecen información desde el año 2013 hasta el 2022, permitiendo un enfoque en los cambios y adaptaciones del tejido empresarial en respuesta a la crisis sanitaria global y sus consecuencias económicas.

Por otro lado, la BBDD 6 se distingue por su formato de fecha (dd/mm/aa), recogiendo datos en un intervalo más detallado y específico, desde el "01/03/2020" hasta el "30/12/2021". Este rango temporal, centrado específicamente en el periodo de la pandemia, permite un análisis pormenorizado de los efectos inmediatos de la COVID-19 sobre las empresas, reflejando las dinámicas de corto plazo en respuesta a las medidas sanitarias y restricciones impuestas.

4. Justificación de la Elección de Variables

4.1. Explicación de la selección de variables.

Las variables seleccionadas incluyen "Año", "Número de Sociedades", "Indicador", "Tipo de ERTE", "Número de ERTES", "Comunidad Autónoma", entre otras. Estas variables fueron escogidas por su relevancia directa en la evaluación del entorno empresarial y su evolución durante y después de la pandemia. Por ejemplo:

Año: Permite realizar comparaciones temporales y evaluar tendencias antes, durante y después de la pandemia.

Número de Sociedades: Es fundamental para medir la tasa de creación y disolución de empresas, proporcionando una visión clara de cómo ha fluctuado el tejido empresarial.

Indicador (ICEA): Ofrece una perspectiva sobre la confianza empresarial, lo cual es un termómetro del clima económico y empresarial.

Tipo de ERTE y Número de ERTES: Estas variables son cruciales para evaluar el impacto de las medidas gubernamentales, especialmente en lo que respecta a la viabilidad y continuidad de las empresas durante la pandemia.

Comunidad Autónoma: Permite realizar análisis regionales entendiendo mediante diferentes áreas geográficas como han sido afectadas y han respondido a la crisis las empresas.

4.2. Relación con los objetivos del TFG.

Evaluar el Cambio en la Tasa de Creación de Empresas: Utilizando variables como "Número de Sociedades" y "Año", se puede calcular la tasa de creación y disolución de empresas a lo largo del tiempo, lo que permite evaluar cómo la pandemia ha afectado la iniciativa empresarial en España.

Examinar las Transformaciones en la Estructura Sectorial: La variable "Actividad Económica" permite analizar cómo diferentes sectores han sido impactados, identificando aquellos que han mostrado mayor resiliencia o han sufrido más durante la crisis.

Evaluar el Impacto de las Medidas Gubernamentales: Variables como "Tipo de ERTE" y "Número de ERTES" son esenciales para entender cómo las políticas gubernamentales, como los ERTes, han ayudado a las empresas a sobrevivir durante los cierres y restricciones.

Medir la Resiliencia Empresarial: El "Indicador" (ICEA) y el análisis de la evolución del "Número de Sociedades" a lo largo de los años permiten identificar factores de resiliencia y adaptabilidad en el tejido empresarial.

5. Análisis Exploratorio de Datos

5.1. Descripción de medidas de tendencia central y dispersión para las variables cuantitativas.

5.1.2 Medidas de tendencia central

En el marco del Trabajo de Fin de Grado, se procederá a realizar un análisis exhaustivo de las variables cuantitativas seleccionadas mediante la aplicación de medidas de tendencia central. Este análisis tiene como objetivo principal proporcionar una comprensión detallada de la distribución central de los datos, lo cual es esencial para identificar patrones, tendencias y posibles anomalías dentro del conjunto de datos. Para cada una de estas variables, se calcularán las siguientes medidas de tendencia central:

- *Media*: Esta medida proporcionará el promedio de los valores para cada variable, ofreciendo una visión general del valor central en torno al cual se distribuyen los datos.
- *Mediana*: Al identificar el valor medio en el conjunto de datos ordenado, la mediana nos permitirá entender el punto central de la distribución, minimizando el efecto de valores atípicos extremos.
- *Moda*: La identificación de los valores más frecuentes en el conjunto de datos nos ayudará a comprender las tendencias predominantes y las preferencias dentro del tejido empresarial.

Tabla 7. Medidas de tendencia central de la primera base de datos

Nombre	Media	Mediana	Moda
Año	2011	2011	2000
Número Sociedades/Capital	90,53	6	0
Población	2397012,45	1464847	1107220

Tabla 8. Medidas de tendencia central de la segunda base de datos.

Nombre	Media	Mediana	Moda
Año	2017	2017	2012
Número Sociedades	109,09	98,17	13,8

Tabla 9. Medidas de tendencia central de la tercera base de datos.

Nombre	Media	Mediana	Moda
Año	2017	2017	2012
Número Sociedades	397,04	124	0

Tabla 10. Medidas de tendencia central de la cuarta base de datos.

Nombre	Media	Mediana	Moda
Año	2017,5	2017,5	2013
Valor Indicador	124,36	126,5	132

Tabla 11. Medidas de tendencia central de la quinta base de datos.

Nombre	Media	Mediana	Moda
Año	2021,5	2021,5	2021
Número de ERTES	3934,27	1777	17

Tabla 12. Medidas de tendencia central de la sexta y séptima base de datos.

Nombre	Media	Mediana	Moda
Afiliados a la Seguridad Social	915199,72	705812	705812

5.1.3 Medidas de dispersión

A continuación, se abordará una dimensión complementaria al estudio de las variables cuantitativas mediante la aplicación de medidas de dispersión. Este enfoque se centra en evaluar la variabilidad o dispersión de los datos alrededor de un valor central, lo cual es crucial para comprender el grado de variación dentro del conjunto de datos y, por ende, la consistencia o heterogeneidad de las observaciones. Las medidas de dispersión seleccionadas para este análisis son:

- **Rango:** Esta medida refleja la diferencia entre el valor máximo y mínimo dentro del conjunto de datos para cada variable. El rango proporciona una visión inicial de la amplitud de la variabilidad, aunque es sensible a valores extremos.
- **Varianza:** Cuantifica la variabilidad de los datos calculando el promedio de los cuadrados de las desviaciones respecto a la media. Aunque proporciona una medida precisa de la dispersión, su interpretación puede ser menos intuitiva debido a que las unidades están al cuadrado respecto a las de la variable original.

- *Desviación Estándar*: Representa la raíz cuadrada de la varianza y mide la dispersión de los datos respecto a su media. Una desviación estándar baja indica que los datos tienden a estar cerca de la media, mientras que una desviación estándar alta señala una mayor dispersión alrededor de la media.
- *Cuartiles 1 y 3*: Los cuartiles dividen el conjunto de datos ordenado en cuatro partes iguales. El primer cuartil (Q1), valor por debajo del cual se encuentra el 25% de los datos y el tercer cuartil (Q3), el valor por debajo del cual se sitúa el 75% de los datos, son particularmente significativos en el análisis estadístico. Estos puntos de corte proporcionan una visión clara de la distribución de los datos, permitiendo identificar dónde se concentran la mayoría de las observaciones y cómo se distribuyen los valores tanto en la parte inferior como en la superior de la muestra.
- *Rango Intercuartílico (IQR)*: El rango intercuartílico se define como la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1). Esta medida de dispersión es especialmente útil para evaluar la variabilidad de los datos minimizando el impacto de los valores atípicos o extremos.

Tabla 13. Medidas de dispersión de la primera base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Año	22	44,02	6,63	12
Número Sociedades/Capital	997	38209,70	195,47	66,74
Población	8445679	5693262396225,57	2386055,82	2063773

Tabla 14. Medidas de dispersión de la segunda base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Año	10	10,04	3,17	6
Número Sociedades	280	5013,29	70,80	87,87

Tabla 15. Medidas de dispersión de la tercera base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Año	10	10,02	3,16	6
Número Sociedades	6 907	593934,52	770,67	350

Tabla 16. Medidas de dispersión de la cuarta base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Año	9	8,30	2,88	5
Valor Indicador	47	126,10	11,23	16,75

Tabla 17. Medidas de dispersión de la quinta base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Año	1	0,25	0,50	1
Número de ERTES	38829	31286261,52	5593,41	4189,5

Tabla 18. Medidas de dispersión de la sexta base de datos.

Nombre	Rango	Varianza	Desviación Estándar	Rango Intercuartílico
Afiliados a la Seguridad Social	3616717	841433388713,49	917296,78	601390

5.2. Frecuencias y proporciones para las variables categóricas.

El análisis de frecuencias y proporciones para las variables categóricas constituye un componente esencial para comprender la distribución y la prevalencia de las categorías dentro de los conjuntos de datos. Sin embargo, al realizar dicho análisis de frecuencias y proporciones, se ha observado que todas estas variables presentan las mismas frecuencias y, por tanto, las mismas proporciones a lo largo del periodo analizado.

Este fenómeno se debe a la naturaleza del estudio, que se enfoca en términos totales y anuales, recopilando información específica para cada año sin variaciones intraanuales en la clasificación de las variables mencionadas. Dado que el análisis se realiza a nivel agregado por año, cada variable categórica refleja una distribución uniforme de frecuencias a través del tiempo, lo que resulta en proporciones idénticas para todas ellas.

6. Gráficos Descriptivos

6.1. Gráficos sobre variables y relaciones.

En la elaboración de este Trabajo de Fin de Grado, se adopta un enfoque detallado y minucioso para analizar el entramado empresarial en España. Para facilitar este análisis, se recurre a la utilización de gráficos descriptivos, herramientas visuales que permiten una interpretación clara y directa de las complejidades inherentes a los datos. Entre los tipos de gráficos seleccionados para este propósito se encuentran el diagrama de dispersión, el diagrama de cajas (boxplot) y el histograma. Cada uno de estos instrumentos gráficos juega un papel vital en la elucidación de las características y tendencias de las variables cuantitativas y categóricas, contribuyendo así a una comprensión más rica de las dinámicas empresariales en medio de la crisis sanitaria global.

Diagrama de Dispersión: Este gráfico resulta indispensable para examinar las interacciones entre dos variables cuantitativas. En nuestro estudio, se aplicarán diagramas de dispersión para indagar en cómo variables, tales como el año y el número de sociedades constituidas, se comportan en función de otras variables concretas como la comunidad autónoma en la que se desarrolla la evolución o las formas jurídicas de las sociedades. Es crucial mencionar que el enfoque se centrará en casos particulares y tipos específicos de empresas, como se ha mencionado anteriormente, para evitar la generación de gráficos sobrecargados y carentes de valor analítico.

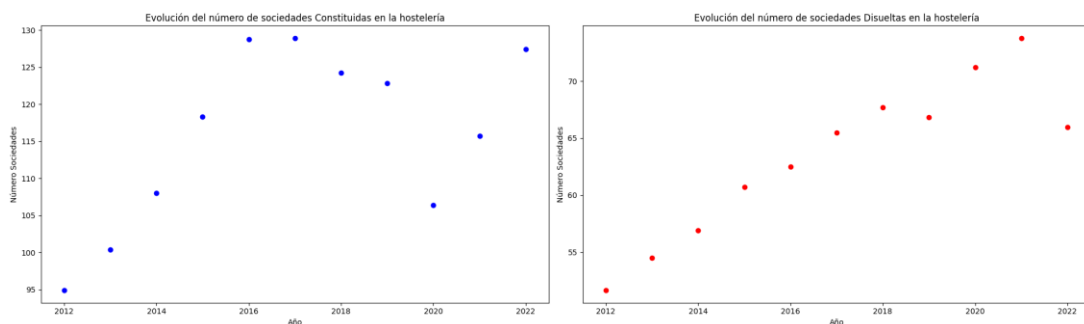


Ilustración 3. Captura de pantalla de los diagramas de dispersión correspondientes a las evoluciones de las sociedades Constituidas y disueltas en el sector de la hostelería desde el año 2012 al 2022.

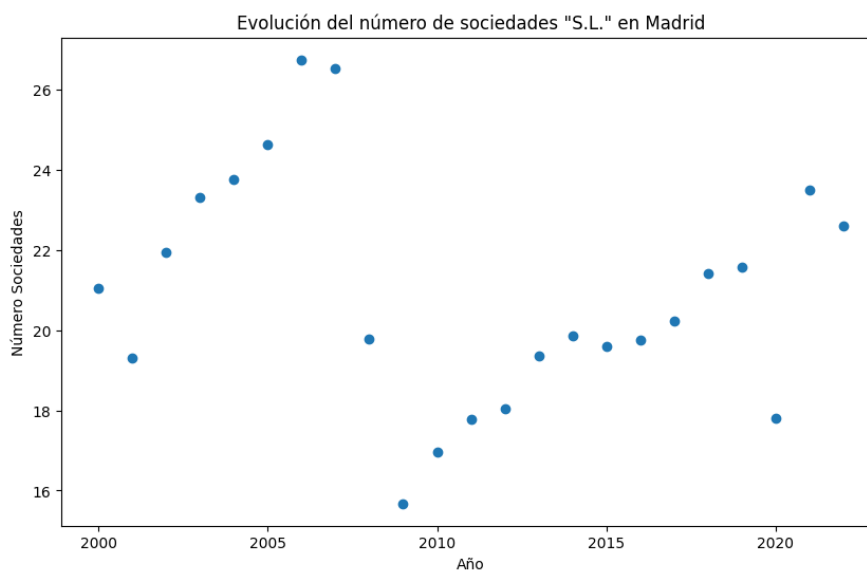


Ilustración 4. Captura de pantalla del diagrama de dispersión correspondiente a la evolución de sociedades de forma jurídica "S.L." constituidas en Madrid desde el año 2000 al 2022.

Diagrama de Cajas (Boxplot): Este tipo de gráfico proporciona una visión comprensiva de la distribución de los datos, resaltando aspectos clave como la mediana, los cuartiles y los outliers. Se empleará para analizar la variabilidad de variables como el valor del indicador ICEA o el capital desembolsado a lo largo de diferentes comunidades autónomas o sectores económicos. La selección cuidadosa de variables para este análisis es esencial para prevenir interpretaciones erróneas y asegurar que la información presentada sea pertinente y accesible.

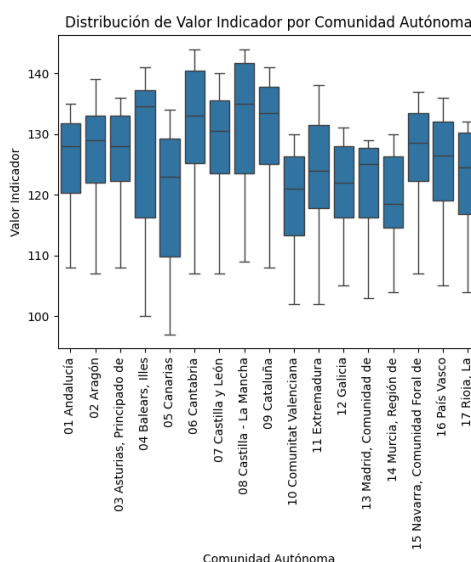


Ilustración 5. Captura de pantalla del diagrama de cajas correspondiente a la distribución del valor indicador "ICEA" según la comunidad autónoma

Histograma: A través de los histogramas, se visualiza la distribución de frecuencias de una variable cuantitativa, lo que permite identificar patrones como la distribución normal o sesgos en los datos. Al igual que con los diagramas de dispersión, se focalizará el análisis en aspectos concretos del conjunto de datos, tales como el número de sociedades disueltas de manera voluntaria en 2021. Este enfoque dirigido garantiza que los histogramas proporcionen insights claros y específicos sobre las dinámicas empresariales bajo estudio.

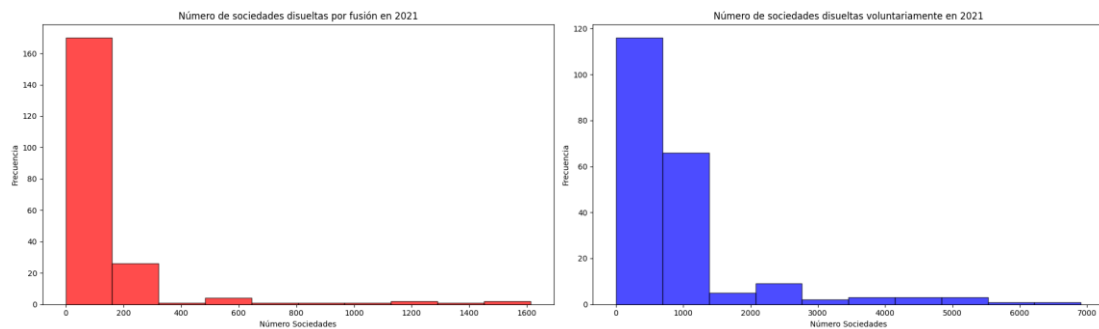


Ilustración 6. Captura de pantalla del histograma correspondiente a la evolución del número de sociedades disueltas voluntariamente en 2021 en todas las comunidades autónomas españolas.

6.2. Patrones, tendencias o correlaciones observadas en los gráficos.

Como se puede apreciar en la *Ilustración 4*, el análisis de la evolución temporal del número de sociedades limitadas en Madrid muestra fluctuaciones significativas que reflejan la respuesta de estas entidades a cambios macroeconómicos y eventos globales. El año 2006 representa un punto álgido en la constitución de sociedades "S.L.", sugiriendo un clima de actividad empresarial propicio en ese entonces. Contrariamente, en 2009 se percibe un mínimo notable, lo que podría estar vinculado a la repercusión de la crisis financiera de 2008 en el tejido empresarial.

La recuperación sostenida que se observa desde ese mínimo hasta el año 2020, interrumpida por un descenso marcado en este último año, coincide con la emergencia de la pandemia de COVID-19, que trajo consigo restricciones económicas y un entorno de incertidumbre. No obstante, la fortaleza del sector se manifiesta en la recuperación observada en 2021, que supera los niveles anteriores a la pandemia.

El análisis detallado en la *ilustración 3*, subraya los dos momentos críticos en los que el crecimiento de las sociedades "S.L." en Madrid se detuvo: el primero en 2009, en el contexto de la crisis financiera global, y el segundo en 2020, durante la pandemia de COVID-19. Ambos periodos están caracterizados por la contracción económica y la disminución en la creación de empresas, sugiriendo una directa influencia de las condiciones económicas adversas en la iniciativa empresarial. Este patrón de crecimiento y contracción refleja la importancia de una

infraestructura de apoyo que pueda mitigar los impactos de shocks externos y fomentar la resiliencia en el tejido empresarial."

Reflejado en la *Ilustración 3* del TFG, se presenta un análisis comparativo de la evolución de las sociedades constituidas y disueltas en el sector de la hostelería en España, desde el año 2012 hasta 2022. La elección de este sector es particularmente pertinente dado que la hostelería ha sido uno de los sectores más visiblemente afectados por la pandemia del COVID-19, y se busca comprender las implicaciones que dicho evento ha tenido en la dinámica empresarial.

En el gráfico de la izquierda, que muestra las sociedades constituidas con puntos azules, se observa un crecimiento exponencial hasta 2017, seguido de un descenso gradual hasta 2019. Sin embargo, entre 2019 y 2020, se produce una caída drástica en el número de nuevas sociedades, lo que se alinea temporalmente con el inicio de la pandemia y las consecuentes restricciones operativas impuestas al sector. Aunque en 2021 se aprecia una recuperación, esta no alcanza los niveles previos al 2019, y en 2022, sin embargo, a partir del 2022 los valores parecen indicar un retorno a la normalidad.

El gráfico de la derecha, representado con puntos rojos, refleja una tendencia creciente en la disolución de sociedades desde 2012 hasta 2021. Es particularmente interesante observar que durante los años 2019 y 2022 no se aprecia un incremento, lo cual podría sugerir una estabilización temporal en la desaparición de empresas. El estancamiento en 2019 podría interpretarse como una consolidación previa a la crisis sanitaria. En contraste, el año 2022, que no muestra un aumento, podría estar indicando el inicio de una recuperación del sector o el resultado de las medidas de apoyo gubernamentales que han permitido a las empresas resistir las adversidades.

La *ilustración 5* refleja la distribución del Índice de Confianza Empresarial Armonizado por Comunidad Autónoma. El Índice de Confianza Empresarial Armonizado (ICEA) proporciona una medida cuantitativa de la percepción que el sector empresarial tiene sobre el clima económico actual y futuro. Esta ilustración presenta un diagrama de cajas que refleja la distribución del ICEA en las distintas comunidades autónomas de España. Este tipo de visualización es esencial para identificar tendencias regionales y disparidades en la confianza empresarial que pueden influir en decisiones estratégicas y políticas económicas.

El diagrama muestra una serie de boxplots para cada comunidad autónoma, donde la línea central de cada caja indica la mediana del valor del indicador, las extremidades de la caja representan el primer y tercer cuartil, y los 'bigotes' se extienden hasta el valor máximo y mínimo dentro del rango intercuartílico. Las variaciones en la altura de las cajas y la longitud de los bigotes indican la heterogeneidad en la confianza empresarial dentro de cada región.

A simple vista, se puede apreciar que hay comunidades como las Islas Baleares o las Islas Canarias con una mayor dispersión de valores, lo que sugiere una opinión empresarial más variada sobre el clima económico. Por otro lado, comunidades como Asturias o Navarra presentan cajas más compactas y bigotes más cortos indicando una percepción más homogénea del entorno empresarial.

Por último, la *Ilustración 6* expone dos histogramas que representan distintas modalidades de disolución de sociedades en España durante el año 2021, distinguiendo entre disoluciones por fusión y disoluciones voluntarias.

En el histograma de la izquierda (tonalidad roja), se muestra la frecuencia de sociedades disueltas por fusión, donde se observa una concentración significativa de casos en el intervalo de 0 a 300. Esto indica que la mayoría de las fusiones involucraron a un número relativamente bajo de sociedades, sugiriendo que las fusiones han sido mayoritariamente entre entidades de menor envergadura o que las fusiones masivas han sido menos comunes.

Por otro lado, el histograma de la derecha (color azul) ilustra la frecuencia de sociedades que se han disuelto voluntariamente. Aquí, los datos muestran una acumulación considerable de frecuencias en el rango de 0 a 1500, con una disminución notable más allá de este punto. Esto refleja que hubo un número mayor de disoluciones voluntarias en comparación con las fusiones, lo cual podría interpretarse como un indicativo de la autonomía empresarial en la toma de decisiones estratégicas ante un contexto económico desafiante, posiblemente exacerbado por las circunstancias de la pandemia de COVID-19.

7. Almacenamiento de Datos

En el desarrollo de este Trabajo de Fin de Grado, se ha optado por utilizar Google Colab como plataforma principal para la ejecución y prueba del código necesario para el análisis de datos. Google Colab ofrece un entorno de Jupyter notebook en la nube, lo cual facilita el acceso a recursos computacionales de alta capacidad sin requerir configuraciones complejas en equipos locales. Además, permite el almacenamiento automático de los notebooks en Google Drive, asegurando que el trabajo no se pierda y pueda ser accesible desde cualquier dispositivo con conexión a internet.

Para garantizar la reproducibilidad del estudio y permitir el acceso público al código desarrollado, se ha creado un repositorio en GitHub. Este repositorio no solo sirve como un medio de respaldo adicional sino también como una plataforma para compartir el trabajo realizado con otros investigadores, académicos o cualquier persona interesada en el tema de estudio. El código, junto con las bases de datos limpiadas y cualquier otro recurso relevante, se ha subido a este repositorio, el cual se puede visitar a través del siguiente enlace:

[REPOSITORIO](#)

Este enfoque de doble almacenamiento, utilizando tanto Google Drive a través de Google Colab como un repositorio en GitHub, asegura no solo la seguridad y accesibilidad de los datos y el código sino también fomenta la transparencia y colaboración en la investigación científica.

SEGUNDA PARTE: ANÁLISIS DEL DATO

8. Introducción al análisis del dato

El análisis de datos se ha convertido en una herramienta fundamental en el ámbito de la investigación y la toma de decisiones en diversas disciplinas, permitiendo extraer conocimientos valiosos a partir de grandes volúmenes de información.

Este documento se enfoca en explorar y aplicar técnicas avanzadas de análisis de datos para comprender mejor las dinámicas y factores que influyen en las disoluciones empresariales en España, un tema de gran relevancia económica y social.

A lo largo de este trabajo, se seleccionarán y prepararán cuidadosamente los datos para su análisis, asegurando que la información sea precisa y esté lista para ser examinada a través de diversos modelos analíticos. Estos modelos, desarrollados y aplicados meticulosamente, buscarán identificar patrones, correlaciones y posibles causas detrás de las disoluciones empresariales, utilizando para ello un enfoque multidimensional que incluye variables económicas, sociales y de mercado. Se evaluará la adecuación de los modelos empleados mediante medidas estadísticas que permitan verificar su fiabilidad y precisión. Además, se hará uso de técnicas de visualización de datos para presentar los resultados de manera clara y comprensible, facilitando así la interpretación de los hallazgos.

Finalmente, este análisis culminará en la elaboración de conclusiones y recomendaciones basadas en los resultados obtenidos, proporcionando insights valiosos. Este trabajo además de tratar aportar al conocimiento académico sobre las disoluciones empresariales en España intentará ofrecer herramientas analíticas que puedan ser aplicadas en futuras investigaciones y en la práctica empresarial para fomentar un entorno económico más estable y resiliente.

9. Selección y Preparación de Datos para el Análisis

La selección y preparación de datos constituyen etapas cruciales en el proceso de análisis de datos, especialmente cuando se abordan cuestiones complejas como las situaciones que influyen en las disoluciones empresariales. Este trabajo se ha fundamentado en el análisis exhaustivo de tres bases de datos principales, cada una de ellas derivada y refinada a partir de conjuntos de datos más amplios, con el objetivo de explorar distintas facetas del fenómeno en estudio.

La primera base de datos, inicialmente denominada "dfbdd1" y posteriormente segmentada en "numsoc" y "capdes", se centró en recopilar información relativa al número de sociedades creadas por tipo de sociedad, año y comunidad autónoma, así como el capital desembolsado por estas sociedades, complementado con datos demográficos por comunidad autónoma. Esta división permitió abordar dos análisis de regresión lineal simple, orientados a evaluar el impacto de la población en la creación de empresas y en el capital desembolsado, respectivamente, proporcionando una visión detallada de cómo la demografía puede influir en el tejido empresarial.

En el desarrollo de la investigación para este Trabajo de Fin de Grado, se configuraron dos bases de datos fundamentales para el análisis de las dinámicas empresariales en España. La primera, denominada "disolución", compila el número de empresas disueltas por tipo de disolución en cada comunidad autónoma desde el año 2012 hasta el 2022. Complementariamente, se creó una segunda base de datos, llamada "macrospain", que recoge indicadores económicos esenciales como el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) por comunidad autónoma y año. El objetivo de vincular estos conjuntos de datos era llevar a cabo una regresión lineal múltiple para investigar cómo las condiciones económicas impactan en la tasa de disolución empresarial.

De manera paralela y con la misma metodología, se desarrolló otra base de datos, "regmul2", destinada a registrar el número de empresas constituidas en cada comunidad autónoma durante el mismo periodo. Esta base de datos se orienta a explorar la relación entre los mismos indicadores económicos y la tasa de creación de empresas, permitiendo un análisis comparativo entre los factores que influyen tanto en la constitución como en la disolución de empresas dentro del territorio español. Esta dualidad de bases de datos enriquece significativamente el estudio, permitiendo una comprensión más completa de las fuerzas económicas que moldean el panorama empresarial del país.

Para profundizar en el análisis de las dinámicas empresariales en España, se creó una tercera base de datos denominada "dismac", destinada a ampliar el espectro de variables macroeconómicas analizadas en relación con las disoluciones empresariales. Esta base de

datos incluye indicadores tales como la deuda pública, el déficit público, el gasto público, los ingresos fiscales, el turismo y las reservas nacionales, junto con el Índice de Confianza Empresarial Armonizado (ICEA), para cada comunidad autónoma y año. Este enfoque multidimensional ofrece una visión más rica y detallada de cómo diversos factores económicos y sociales pueden influir en la estabilidad y continuidad de las empresas en España.

De manera similar, se creó una base de datos complementaria llamada "consmac", que aplica la misma metodología pero se centra en las constituciones empresariales. "consmac" también integra las mismas variables macroeconómicas y busca explorar cómo estos factores influyen en la creación de nuevas empresas, permitiendo un análisis comparativo entre los impulsores de las constituciones y disoluciones empresariales. Este enfoque paralelo facilita una comprensión integral de los efectos de las políticas económicas y las condiciones de mercado en el tejido empresarial del país.

Por último, para complementar y enriquecer aún más el análisis sobre las dinámicas empresariales en España, se han incorporado dos nuevas bases de datos: "consext" y "disext", correspondientes a las constituciones y disoluciones empresariales, respectivamente. Estas bases de datos han sido diseñadas para explorar la relación entre un amplio espectro de variables y las dinámicas de constitución y disolución de empresas. Incluyen variables tanto económicas como sociales, ampliando considerablemente el rango de factores considerados en el análisis. Entre estas variables se encuentran la cotización del IBEX, del EUR y del NASDAQ, que reflejan el comportamiento del mercado financiero; indicadores macroeconómicos como el déficit público, el gasto público y los ingresos fiscales; y variables demográficas como nacimientos, matrimonios, defunciones y población total.

La inclusión de estas variables busca proporcionar una visión más holística y detallada de los factores que pueden influir en la creación y cierre de empresas en España. Se parte del supuesto de que a mayor número de variables analizadas, mayor será la capacidad del modelo para explicar las fluctuaciones en las tasas de constituciones y disoluciones empresariales. Con "consext" y "disext", se pretende no solo identificar las tendencias y patrones más evidentes, sino también descubrir conexiones menos obvias que puedan surgir de la interacción entre el entorno económico, el mercado financiero y la dinámica social, ofreciendo así una comprensión más profunda y matizada de los desafíos y oportunidades dentro del ecosistema empresarial español.

Cada una de estas bases de datos fue sometida a un riguroso proceso de selección y preparación, que incluyó la limpieza de datos, la gestión de valores faltantes, la transformación de variables y la verificación de la calidad de los datos. Este proceso aseguró que la información utilizada en los análisis fuera de la más alta calidad y relevancia, permitiendo así

obtener resultados confiables. La meticulosa preparación de los datos subraya la importancia de una base sólida para cualquier análisis de datos, especialmente cuando se abordan cuestiones de complejidad y relevancia como las que conciernen a las disoluciones empresariales en el contexto económico y social de España.

10. Modelos Analíticos: Desarrollo y Aplicación (Regresiones Lineales)

10.1. Modelo Analítico Supervisado (Regresiones Lineales Simples)

Explicación del Modelo:

En el análisis del comportamiento empresarial y económico, la regresión lineal simple emerge como una herramienta analítica fundamental, especialmente cuando el objetivo es explorar la relación entre dos variables específicas. Este modelo supervisado se seleccionó con el propósito de investigar cómo la variable independiente, en este caso, la población de una comunidad autónoma puede influir en la variable dependiente, que para la primera regresión se define como el número de sociedades creadas y para la segunda como el capital desembolsado en miles de euros.

La regresión lineal simple es un método estadístico fundamental utilizado para modelar y analizar las relaciones entre dos variables cuantitativas: una variable independiente (o explicativa) y una variable dependiente (o respuesta). Este modelo asume que existe una relación lineal entre estas variables, la cual puede ser descrita mediante una ecuación de la forma:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Donde:

- y representa la variable dependiente.
- x representa la variable independiente.
- β_0 es el término de intercepción, que indica el valor de y cuando x es 0.
- β_1 es el coeficiente de la pendiente, que indica el cambio en y por cada unidad de cambio en x .
- ϵ es el término de error, que representa la variación en y que no puede ser explicada por la relación lineal con x .

El método de mínimos cuadrados es utilizado para estimar los parámetros β_0 y β_1 de la regresión. Este método busca minimizar la suma de los cuadrados de las diferencias

(residuos) entre los valores observados de “ y ” y los valores predichos por el modelo. Matemáticamente, se busca minimizar la función:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

Donde y_i y x_i son los valores observados de la variable dependiente e independiente, respectivamente, y n es el número de observaciones. La solución a este problema de optimización nos da los estimadores de mínimos cuadrados para β_0 y β_1 , que son:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

Donde \bar{x} e \bar{y} son las medidas de las variables independiente y dependiente, respectivamente.

El β_1 coeficiente indica la pendiente de la línea de regresión y representa el cambio esperado en y por cada unidad de cambio de x . Un valor positivo de β_1 indica una relación directa entre x e y , mientras que un valor negativo indica una relación inversa.

El término de intercepción β_0 indica el valor esperado de y cuando x es 0. Este valor puede no tener siempre una interpretación práctica, especialmente si $x = 0$ no se encuentra dentro del rango de los datos observados.

Para que las estimaciones obtenidas mediante la regresión lineal simple sean válidas, se deben cumplir ciertos supuestos:

- 1) **Linealidad:** La relación entre x e y debe ser lineal.
- 2) **Independencia:** Los residuos ε deben ser independientes entre sí.
- 3) **Homoscedasticidad:** La varianza de los residuos debe ser constante a lo largo de todas las observaciones.
- 4) **Normalidad:** Los residuos deben seguir una distribución normal.

Proceso de Selección de Variables:

La selección de variables para la regresión lineal simple se centró en identificar la variable independiente (X) y la variable dependiente (Y) que mejor representaran la relación

que se deseaba explorar. Dada la naturaleza del análisis, se identificó la población de las comunidades autónomas como la variable independiente, considerando su potencial impacto en el atractivo para la creación de empresas y la capacidad económica de estas. Por otro lado, se eligió el número de sociedades creadas como la variable dependiente de la primera regresión y al capital desembolsado en miles de euros por las sociedades para la segunda regresión. Para llevar a cabo el análisis, se dividió la base de datos original en dos subconjuntos: uno enfocado en el número de sociedades y otro en el capital desembolsado, permitiendo así un estudio detallado y específico de cada aspecto. Este proceso de selección de variables fue crucial para asegurar que el modelo pudiera capturar de manera efectiva la relación entre la demografía de las comunidades autónomas y la actividad económica empresarial, facilitando la interpretación de los resultados y la extracción de conclusiones relevantes.

Desarrollo de modelos:

Regresión 1: Número de Sociedades

- *Análisis de la Existencia de Relación Lineal:* Para explorar la relación entre la población de las comunidades autónomas y el número de sociedades creadas, se realizaron análisis gráficos preliminares. Se emplearon gráficos de dispersión con líneas de tendencia suavizadas para visualizar la distribución de los datos y detectar patrones de correlación visualmente. Estos gráficos permitieron una primera aproximación a la dinámica entre las variables, sugiriendo una relación que, a primera vista, podría no ser lineal o ser muy débil, dada la dispersión de los puntos y la suavidad de la línea de tendencia. Para complementar el análisis gráfico, se calculó la correlación entre 'Población' y 'Número de Sociedades' utilizando la función `corr()` de pandas, seguido de un análisis más formal mediante el coeficiente de correlación de Pearson y su p-valor asociado. Los resultados indicaron una correlación de -0.008 con un p-valor de 0.761, lo que sugiere que, estadísticamente, no existe una relación lineal significativa entre la población y el número de sociedades creadas.
- *Análisis de Ajuste a una Distribución Normal:* El ajuste de las variables a una distribución normal es crucial para la aplicación de ciertas técnicas estadísticas. Se utilizó la visualización mediante gráficos de densidad y se realizaron pruebas de normalidad, incluyendo Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2 . Los gráficos de densidad revelaron una distribución asimétrica para ambas variables, confirmada por los valores de asimetría

(skewness) significativamente diferentes de cero. Además, las pruebas de normalidad arrojaron p-valores extremadamente bajos para ambas variables, indicando un rechazo de la hipótesis nula de normalidad.

- *Construcción del Modelo:* Para la construcción del modelo de regresión lineal simple, se prepararon las variables seleccionadas, añadiendo una columna de unos para el intercepto. A pesar de la aparente falta de una relación lineal significativa y la no normalidad de las distribuciones, se procedió a ajustar el modelo para explorar la relación entre las variables de interés. El modelo ajustado mostró un R-cuadrado cercano a cero, indicando que el modelo no explica prácticamente ninguna variabilidad en el número de sociedades en función de la población. Los coeficientes de regresión y sus intervalos de confianza reflejaron la falta de significancia estadística de la población como predictor del número de sociedades.

Regresión 2: Capital Desembolsado

- *Análisis de la Existencia de Relación Lineal:* Para investigar la relación entre la población de las comunidades autónomas y el capital desembolsado en la creación de sociedades, se emplearon gráficos de dispersión complementados con líneas de tendencia suavizadas. Estos gráficos facilitaron una visualización preliminar de la relación entre las variables, sugiriendo la necesidad de un análisis más detallado.

La correlación entre 'Población' y 'Capital' se calculó utilizando la función `corr()` de pandas, y se complementó con el coeficiente de correlación de Pearson y su p-valor asociado. Los resultados mostraron una correlación de 0.20 con un p-valor significativamente bajo ($4.639e-14$), lo que indica una relación lineal positiva estadísticamente significativa entre la población y el capital desembolsado, aunque la fuerza de esta relación es moderada.

- *Análisis de Ajuste a una Distribución Normal:* El análisis de la distribución de las variables mediante gráficos de densidad y pruebas de normalidad reveló una distribución asimétrica para ambas variables, lo que se reflejó en los valores de asimetría significativamente altos. Las pruebas de Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2 confirmaron la no normalidad de las distribuciones, con p-valores que indican un rechazo fuerte de la hipótesis nula de normalidad.
- *Construcción del Modelo:* A pesar de la moderada correlación positiva entre la población y el capital desembolsado y la no normalidad de las distribuciones,

se procedió a ajustar un modelo de regresión lineal simple. Se prepararon las variables seleccionadas, incluyendo una columna de unos para el intercepto, y se ajustó el modelo para explorar la relación entre la población y el capital desembolsado.

El modelo ajustado reveló un R-cuadrado de 0.043, indicando que un 4.3% de la variabilidad en el capital desembolsado puede explicarse por la población. Aunque esta proporción es baja, el coeficiente para la población fue estadísticamente significativo, lo que sugiere que existe una relación lineal positiva entre la población y el capital desembolsado. Este análisis resalta que aunque la relación entre la población y el capital desembolsado es estadísticamente significativa y la fuerza de esta relación es moderada, la no normalidad de las variables necesita un análisis más profundo.

Evaluación del Modelo con Medidas de Error/Precisión Específicas:

Regresión 1: Número de Sociedades

- *Calidad del Modelo:* El R-cuadrado obtenido en el modelo es 0.000, indicando que la variabilidad explicada por el modelo es prácticamente nula. Este valor sugiere que la población, como variable independiente, no proporciona una base sólida para predecir el número de sociedades creadas. El R-cuadrado ajustado, que considera el número de predictores en el modelo y el número de observaciones, también refleja una falta de ajuste, evidenciado por un valor negativo (-0.001). Esto implica que el modelo no mejora la predicción más allá de lo que se esperaría por azar. El F-statistic y su p-valor asociado (0.09227 y 0.761, respectivamente) refuerzan esta interpretación, indicando que el modelo no es estadísticamente significativo.
- *Confiabilidad del Modelo:* La confiabilidad del modelo se ve comprometida por varios factores. Primero, el alto valor de la condición (4.79×10^6) sugiere la presencia de multicolinealidad, aunque este fenómeno es menos probable en modelos de regresión simple. Los coeficientes de regresión y sus intervalos de confianza revelan que, aunque el intercepto es estadísticamente significativo, la pendiente asociada a la población no lo es, como lo demuestra su intervalo de confianza que cruza el cero y un p-valor alto. Los residuos estimados y la suma de cuadrados de los residuos muestran la variabilidad que el modelo no logra explicar, siendo esta considerablemente alta.

- *Análisis:* En conclusión, el modelo de regresión lineal simple para predecir el número de sociedades basado en la población no proporciona una herramienta confiable ni precisa para entender esta relación. La falta de significancia estadística y la baja capacidad explicativa del modelo sugieren que otros factores no considerados en este análisis podrían influir en el número de sociedades creadas. Además, la evaluación de la calidad y confiabilidad del modelo resalta la importancia de considerar múltiples variables y realizar un análisis más profundo para capturar la complejidad de los factores que influyen en la creación de sociedades.

Regresión 2: Capital Desembolsado

- *Calidad del Modelo:* El valor de R-cuadrado obtenido, 0.043, aunque modesto, indica que aproximadamente el 4.3% de la variabilidad en el capital desembolsado puede ser explicada por la población. Este resultado sugiere una relación positiva entre ambas variables, aunque la magnitud de esta relación es limitada. El R-cuadrado ajustado, que se sitúa en 0.042, confirma la leve mejora en la predicción del modelo sobre la base de la población, ajustada por el número de predictores. El F-statistic alcanza un valor de 58.15, con un p-valor asociado significativamente bajo ($4.64e-14$), lo que indica que el modelo es estadísticamente significativo. Esto sugiere que existe una relación lineal entre la población y el capital desembolsado, aunque la fuerza de esta relación es relativamente débil.
- *Confiabilidad del Modelo:* La confiabilidad del modelo se ve afectada por varios factores. El alto valor de la condición ($4.79e+06$) sugiere la presencia de multicolinealidad o problemas numéricos que pueden influir en la precisión de las estimaciones de los coeficientes. Los coeficientes de regresión y sus intervalos de confianza muestran que tanto el intercepto como la pendiente asociada a la población son estadísticamente significativos. Esto indica que, controlando por la población, se espera un incremento en el capital desembolsado con el aumento de la población. Los residuos estimados y la suma de cuadrados de los residuos indican la cantidad de variabilidad que el modelo no logra explicar, siendo esta considerable.
- *Análisis:* En resumen, el modelo de regresión lineal simple para predecir el capital desembolsado basado en la población proporciona evidencia de una relación positiva entre estas variables. Sin embargo, la capacidad explicativa del modelo es limitada, lo que sugiere que otros factores no considerados en

este análisis podrían tener un impacto significativo en el capital desembolsado. La significancia estadística del modelo indica que la población es un predictor relevante, pero la presencia de un alto valor de condición y la limitada varianza explicada por el modelo sugieren la necesidad de un análisis más profundo y la posible inclusión de variables adicionales.

10.2. Modelo Analítico Supervisado (Regresiones Lineales Múltiples)

Explicación del Modelo:

La regresión lineal múltiple, al igual que su contraparte simple, es una herramienta estadística esencial en el análisis de datos, especialmente útil para explorar la relación entre una variable dependiente y múltiples variables independientes. Este modelo supervisado permite investigar cómo varias variables independientes, como la población y el PIB de una comunidad autónoma, pueden influir conjuntamente en una variable dependiente, como el número de empresas constituidas o el capital desembolsado en miles de euros.

El modelo de regresión lineal múltiple se expresa mediante la siguiente ecuación:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Donde:

- y representa la variable dependiente.
- x_1, x_2, \dots, x_k representan las variables independientes.
- β_0 es el término de intercepción, que indica el valor de y cuando las x son 0.
- $\beta_1, \beta_2, \dots, \beta_k$ son los coeficientes de las variables independientes, que miden el cambio en y asociado a una unidad de cambio en cada x .
- ϵ es el término de error, que capta toda la variabilidad en y que no es explicada por las variables independientes.

El método de los mínimos cuadrados ordinarios también se utiliza aquí para estimar los coeficientes, buscando minimizar la suma de los cuadrados de los residuos, es decir, las diferencias entre los valores observados y los valores predichos por el modelo. La función a minimizar es:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}))^2$$

La regresión lineal múltiple es potente por su capacidad de ajustar múltiples variables, permitiendo un análisis más detallado y una mejor comprensión de cómo diversos factores afectan la variable objetivo. Los coeficientes obtenidos ofrecen una medida de la influencia de cada variable independiente, ajustada por la presencia de otras en el modelo.

Para que el modelo de regresión lineal múltiple sea efectivo y sus estimaciones fiables, debe cumplir con ciertos supuestos, similares a los de la regresión lineal simple:

- 5) **Linealidad**: La relación entre las variables independientes y la dependiente debe ser lineal.
- 6) **Independencia**: Los residuos ε deben ser independientes entre sí.
- 7) **Homoscedasticidad**: La varianza de los residuos debe ser constante a lo largo de todas las observaciones.
- 8) **No multicolinealidad**: Las variables independientes no deben estar demasiado correlacionadas entre sí.
- 9) **Normalidad**: Los residuos deben distribuirse normalmente.

10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE (IPC y PIB)

Proceso de Selección de Variables

La selección de las variables IPC y PIB como predictores se basó en la hipótesis de que la salud económica de un país, reflejada en estos indicadores, tiene un impacto directo tanto en la tasa de disoluciones como en las constituciones empresariales. El IPC, como medida de la inflación, y el PIB, como indicador del rendimiento económico general, son fundamentales para entender el entorno en el que operan las empresas. La elección se apoyó en un análisis exploratorio de datos y en la revisión de literatura relevante, que sugiere una relación potencial entre estos factores económicos y la dinámica empresarial. Para las constituciones empresariales, se planteó que un entorno económico robusto, indicado por un PIB en aumento, podría fomentar la creación de nuevas empresas, mientras que un IPC elevado podría desincentivar nuevas inversiones debido a los costos más altos asociados con la inflación. Esta dualidad de impactos enfatiza la necesidad de examinar cómo cada variable no solo afecta las tasas de cierre, sino también las iniciativas de nuevas empresas.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal*: En el proceso de desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se llevó a cabo un análisis preliminar para establecer la

existencia de una relación lineal entre las variables seleccionadas. Mediante gráficos de dispersión se visualizaron las relaciones entre las disoluciones y constituciones empresariales con variables económicas clave como el IPC (Índice de Precios al Consumidor) y el PIB (Producto Interno Bruto).

Para las disoluciones empresariales, aunque los coeficientes de correlación encontrados fueron bajos (IPC: 0.0179, PIB: 0.0273), estos indicaron una posible relación lineal que justificó su inclusión en un modelo de regresión lineal múltiple. A pesar de los valores de P no significativos (IPC: $p=0.7967$, PIB: $p=0.6943$), que sugieren que estas variables por sí solas no explican de manera significativa las variaciones en las disoluciones, la decisión de incorporarlas en el modelo buscaba explorar efectos combinados en análisis más complejos.

De manera similar, para las constituciones empresariales, los coeficientes de correlación también resultaron ser bajos (IPC: -0.0076, PIB: 0.0202) y no estadísticamente significativos (IPC: $p=0.9130$, PIB: $p=0.7715$), lo que a primera vista sugiere una influencia limitada de estas variables sobre las tasas de constitución. Sin embargo, al igual que con las disoluciones, estos resultados preliminares apoyaron la decisión de incluir ambas variables en un análisis de regresión lineal múltiple más detallado.

- *Análisis de Ajuste a una Distribución Normal:* En el desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se prestó atención particular al análisis de la normalidad de las variables seleccionadas: el IPC y el PIB, junto con las disoluciones y constituciones de empresas.

Las pruebas de normalidad, incluyendo Shapiro-Wilk, Anderson-Darling, y D'Agostino's K^2 , mostraron resultados significativos que indicaban desviaciones de la normalidad para ambas series de datos. En el caso de las disoluciones, el test de Shapiro-Wilk dio un estadístico de 0.7143 con un valor p extremadamente bajo, sugiriendo una fuerte evidencia contra la hipótesis de normalidad. De manera similar, los resultados para las constituciones fueron concluyentes, con un estadístico de 0.7043 en la prueba de Shapiro-Wilk y un valor p casi nulo, reflejando una distribución no normal.

A pesar de que las variables no siguieron una distribución normal perfecta, se decidió proceder con el modelo de regresión lineal múltiple. Esto se debe a que, aunque la normalidad es un supuesto importante en muchos análisis estadísticos, la regresión lineal múltiple es bastante robusta a

violaciones de este supuesto. Además, la visualización de las relaciones entre las variables mediante gráficos de dispersión indicó la posibilidad de una relación lineal, aunque los coeficientes de correlación fueran bajos (0.0179 para disoluciones con IPC y 0.0202 para constituciones con PIB), justificando así su inclusión en análisis más detallados.

- *Construcción del modelo:* En el desarrollo del modelo para analizar tanto las disoluciones como las constituciones empresariales, se incorporaron el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) como variables independientes. El método de mínimos cuadrados ordinarios (OLS) se empleó para estimar los parámetros del modelo, proporcionando una base para evaluar cómo las fluctuaciones económicas afectan tanto a la disolución como a la constitución de empresas.

Para las disoluciones empresariales, el modelo mostró coeficientes muy bajos en el IPC y el PIB, indicando una relación débil entre estas variables y el número de empresas disueltas. Los resultados específicos mostraron un coeficiente para el IPC de -6.4564 y para el PIB de aproximadamente 0, con un valor constante considerablemente alto, aunque no significativo, lo que sugiere que otros factores no capturados en el modelo podrían estar influyendo en las disoluciones.

Por otro lado, para las constituciones empresariales se reveló un patrón similar, con coeficientes también bajos y no significativos. El coeficiente para el IPC fue de -125.5320 y para el PIB de aproximadamente 0, con una constante igualmente alta pero no significativa, reiterando la limitada influencia directa del PIB y el IPC en la formación de nuevas empresas según los datos analizados.

Ambos modelos, aunque mostraron bajos coeficientes de determinación (R-squared), proporcionan insights valiosos sobre la compleja relación entre las condiciones económicas y la dinámica empresarial. Estos hallazgos sugieren que tanto la disolución como la constitución de empresas están influenciadas por una combinación de factores más amplia que los meramente económicos, resaltando la necesidad de explorar variables adicionales y contextos más específicos para obtener un entendimiento más profundo y aplicable de estos fenómenos.

Evaluación del Modelo con Medidas de Error/Precisión Específicas:

- *Calidad del modelo:* La evaluación de la calidad de los modelos para las disoluciones y constituciones empresariales utilizó el coeficiente de

determinación R-cuadrado y el R-cuadrado ajustado como medidas clave. En el caso de las disoluciones empresariales, el modelo exhibió un R-cuadrado de 0.001 y un R-cuadrado ajustado de -0.009, lo que indica que las variables PIB e IPC apenas explican el 1% de la variabilidad en las disoluciones. Esto se complementa con un F-estadístico de 0.08673 y un p-valor de 0.917, sugiriendo que el modelo no logra proporcionar una base estadísticamente significativa para predecir las disoluciones empresariales.

Para las constituciones empresariales, los resultados fueron similarmente limitados, con un R-cuadrado de 0.003 y un R-cuadrado ajustado de -0.007. Estos valores implican que el modelo apenas explica un 0.3% de la variabilidad en las constituciones, reflejando una capacidad predictiva muy baja. El F-estadístico asociado fue de 0.2619 con un p-valor de 0.770, reforzando la idea de que el modelo carece de significación estadística para predecir las constituciones empresariales basado en las mismas variables macroeconómicas.

En resumen, ambos modelos demostraron tener una capacidad muy limitada para capturar y explicar las dinámicas detrás de las constituciones y disoluciones empresariales en España.

- *Confiabilidad del modelo:* La confiabilidad de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales se ve comprometida por un número de condición extremadamente alto ($3.04e+13$) en ambos casos. Este elevado número de condición indica una fuerte presencia de multicolinealidad entre las variables independientes, como el IPC y el PIB, lo que sugiere que estas variables no son completamente independientes entre sí. Esta interdependencia complica la interpretación de los coeficientes individuales y puede inflar los errores estándar, comprometiendo así la fiabilidad de las estimaciones de los coeficientes y reduciendo la confianza en las inferencias que se pueden hacer a partir del modelo.
- *Análisis:* El análisis de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales en España muestra limitaciones significativas en su capacidad de proporcionar predicciones o explicaciones robustas sobre estos fenómenos. La baja capacidad explicativa de los modelos, evidenciada por los valores negativos de R-cuadrado ajustado tanto para las disoluciones como para las

constituciones empresariales, junto con la falta de significancia estadística de las variables independientes (IPC y PIB), refleja que estos modelos no capturan adecuadamente la complejidad de las relaciones entre las condiciones económicas y los cambios en el ecosistema empresarial.

Los elevados valores de RMSE para ambos modelos—1743.75 para disoluciones y 6584.99 para constituciones—subrayan la discrepancia entre los valores observados y los predichos por los modelos, lo que indica una limitada utilidad práctica de estos en su estado actual. Este análisis resalta la necesidad de reconsiderar la selección de variables, explorar la inclusión de otras variables potencialmente relevantes, o emplear métodos analíticos alternativos más eficaces para capturar la dinámica entre la salud económica y los comportamientos empresariales.

En conclusión, aunque el enfoque de regresión lineal múltiple es teóricamente sólido para examinar relaciones entre múltiples variables independientes y una variable dependiente, los resultados obtenidos en este caso sugieren que es crucial emplear estrategias analíticas que puedan manejar mejor la complejidad de los factores económicos que influyen en las constituciones y disoluciones de empresas en España.

10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE (Variables Macroeconómicas)

Proceso de Selección de Variables

El proceso de selección de variables para esta nueva regresión lineal múltiple buscó identificar indicadores económicos amplios que afectaran tanto las disoluciones como las constituciones empresariales en España. Se partió de las bases de datos "disolución" y "constitución", enriqueciendo cada una con variables de dos fuentes adicionales que incluyen indicadores macroeconómicos y el Índice de Confianza Empresarial Armonizado (ICEA).

Para los nuevos modelos, se incluyen las variables ICEA, Deuda y Déficit Públicos, Gasto Público, Ingresos Fiscales, Llegadas de Turistas y Reservas Totales. A pesar de considerar inicialmente el IPC y el PIB como parte de la regresión, se decidió excluir estas variables para permitir futuras comparaciones de modelos. Esta decisión se fundamenta en la necesidad de evaluar el impacto directo de variables específicas más allá del rendimiento económico general reflejado por el IPC y el PIB.

Este enfoque estructurado para la selección de variables asegura que el modelo pueda abordar de manera integral los múltiples factores que influyen en las dinámicas de constitución

y disolución empresariales, desde la macroeconomía hasta elementos específicos del clima empresarial y turismo. Estas variables fueron integradas en las bases de datos expandidas, denominadas "dismac" para disoluciones y "consmac" para constituciones, facilitando un análisis más profundo y diversificado de cómo estos factores económicos impactan en el ecosistema empresarial español.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal:* En el desarrollo del modelo de regresión lineal múltiple para analizar las influencias en las tasas de constitución y disolución empresariales en España, se realizaron gráficos de dispersión y análisis de correlación para examinar las relaciones entre diversas variables económicas y las tasas empresariales. Los gráficos proporcionaron una visualización directa de las posibles tendencias y anomalías entre variables como el ICEA, la Deuda, el Déficit, el Gasto Público, los Ingresos Fiscales, las Llegadas de Turistas y las Reservas.

La matriz de correlación reveló relaciones lineales de diferente intensidad entre las variables y las tasas de disolución y constitución empresariales. Aunque las correlaciones fueron generalmente bajas, destacaron algunas asociaciones significativas, como la correlación entre el Déficit y el Turismo, que mostró fuertes lazos con otras variables económicas, reflejando una compleja interacción dentro del entorno económico que podría afectar la estabilidad empresarial. En particular, el coeficiente de correlación de Pearson indicó que la relación entre las disoluciones empresariales y variables como el ICEA y la Deuda es débil y no significativa estadísticamente, lo que sugiere una influencia limitada de estos indicadores sobre las disoluciones.

Similarmente, para las constituciones empresariales, las correlaciones también resultaron bajas y sin significancia estadística en la mayoría de los casos, reiterando la necesidad de un análisis más profundo para comprender mejor las dinámicas que afectan la formación de nuevas empresas. Por ejemplo, la correlación entre las constituciones y variables como el ICEA y el Gasto Público también mostró una influencia limitada, indicando que los modelos actuales pueden no estar capturando completamente los factores que inciden en la creación de empresas.

- *Análisis de Ajuste a una Distribución Normal:* En el análisis de ajuste a una distribución normal para el modelo de regresión lineal múltiple, se detectaron variaciones significativas en la normalidad de las distribuciones de las variables

analizadas. Se utilizó una serie de pruebas de normalidad, como Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2 , para evaluar estadísticamente la normalidad de las variables críticas.

Los resultados de estas pruebas revelaron desviaciones claras de la normalidad en muchas de las variables. Por ejemplo, la prueba Shapiro-Wilk para las disoluciones empresariales y las constituciones mostró p-valores extremadamente bajos, indicando una fuerte evidencia contra la hipótesis de normalidad. Esta tendencia se observó también en otras variables económicas, donde los p-valores obtenidos en pruebas como la de Anderson-Darling y D'Agostino's K^2 confirmaron estas desviaciones.

Estos hallazgos son cruciales porque sugieren precaución al interpretar los resultados del modelo de regresión y al aplicar inferencias estadísticas que dependen de supuestos de normalidad. La evidencia de no normalidad implica que algunas de las técnicas estadísticas estándar podrían no ser completamente apropiadas o precisas para estos datos.

- *Construcción del modelo:* La construcción del modelo de regresión lineal múltiple en este estudio se diseñó para evaluar el impacto de una serie de variables macroeconómicas sobre las tasas de disoluciones y constituciones empresariales en España. Se introdujeron variables como el ICEA, la Deuda, el Déficit, el Gasto Público, los Ingresos Fiscales, el número de Turistas y las Reservas como predictores. Aunque la mayoría de estas variables no mostraron una relación estadísticamente significativa con las constituciones, el ICEA resaltó por su influencia positiva en el modelo de disoluciones, indicando una relación estadísticamente significativa con un coeficiente de -78.8743 y un p-valor de 0.001, sugiriendo que un mejor clima empresarial podría estar asociado con una reducción en el número de disoluciones empresariales.

Sin embargo, el modelo presenta desafíos, incluyendo una alta multicolinealidad, evidenciada por un número de condición elevado (2.17×10^{14}), lo que complica la interpretación de los coeficientes de las variables independientes. Esto indica que algunas variables podrían estar proporcionando información redundante. La evaluación estadística muestra que, aunque el modelo de disoluciones logra un R-cuadrado de 0.065 y un ajuste R-cuadrado de 0.024, el modelo para las constituciones alcanza solo un R-cuadrado de 0.015 y un ajuste R-cuadrado de -0.027, reflejando una capacidad predictiva limitada.

Este análisis revela que, aunque el modelo utiliza un enfoque válido teóricamente para explorar las relaciones entre múltiples variables y las tasas de constitución y disolución empresarial, las limitaciones en la significancia estadística de las variables y los problemas de multicolinealidad requieren una revisión de la selección de variables, posiblemente incorporando otros factores relevantes o empleando métodos analíticos alternativos para una comprensión más precisa de estas dinámicas empresariales.

Reconstrucción del modelo en base a la multicolinealidad

Disoluciones:

El proceso de reajuste del modelo de regresión lineal múltiple para abordar la multicolinealidad implicó un análisis meticuloso del Factor de Inflación de la Varianza (VIF). Este procedimiento iterativo de eliminación de variables con altos VIF permitió identificar y descartar aquellas que contribuían significativamente a la multicolinealidad, mejorando así la calidad y la interpretación del modelo. En particular, variables como Déficit, Gasto, e 'IngreFis' fueron removidas debido a su alta correlación con otras variables independientes.

El modelo ajustado se centró en un conjunto más reducido de variables ('ICEA', 'Turistas', 'Deuda', 'Reservas'), que mostró un R-cuadrado de 0.047, indicando que estas variables explican aproximadamente el 4.7% de la variabilidad en las disoluciones empresariales. Aunque este porcentaje es relativamente bajo, refleja la complejidad y la multitud de factores que pueden influir en las disoluciones empresariales. Notablemente, 'ICEA' y 'Turistas' mostraron una relación estadísticamente significativa con las disoluciones, subrayando la relevancia del clima empresarial y la actividad turística.

Sin embargo, a pesar de los esfuerzos por mitigarlo, el modelo aún presenta un número de condición muy elevado (8.91×10^9), lo que indica la presencia de multicolinealidad residual. Esto sugiere que, aunque se ha mejorado, la interpretación de los coeficientes debe hacerse con precaución.

En resumen, el proceso de reajuste del modelo y la evaluación de la multicolinealidad han sido pasos cruciales para mejorar su precisión y fiabilidad. Aunque se ha logrado cierto grado de claridad en la relación entre algunas variables macroeconómicas y las disoluciones empresariales, los resultados también destacan la complejidad inherente a este fenómeno.

Constituciones:

La reconstrucción del modelo de regresión lineal múltiple para las constituciones empresariales en España, ajustado para mitigar la multicolinealidad entre las variables independientes, también se realizó a través de un minucioso análisis del Factor de Inflación de la Varianza (VIF). Este proceso llevó a la eliminación de variables con altos VIF que estaban distorsionando los resultados del modelo debido a su fuerte correlación con otras variables. Las variables finales incluidas fueron 'ICEA', 'Turistas', 'Deuda' y 'Reservas'.

El modelo ajustado reflejó un R-cuadrado de 0.009, indicando que estas variables explican solo un pequeño porcentaje de la variabilidad en las constituciones empresariales. Aunque este resultado es modesto, destaca la complejidad y la diversidad de factores que influyen en las constituciones empresariales, y sugiere que muchos elementos críticos podrían no estar capturados por el modelo. La variable 'ICEA' no resultó ser estadísticamente significativa, lo cual pone en cuestión su impacto directo en las constituciones, a diferencia de su efecto observado en las disoluciones empresariales.

A pesar de los ajustes, el modelo sigue presentando un número de condición elevado (8.91×10^9), lo que señala la persistencia de multicolinealidad residual y sugiere que la interpretación de los coeficientes debe hacerse con cautela. Este desafío resalta la necesidad de continuar refinando el modelo y de explorar la inclusión de otras variables o métodos analíticos que puedan capturar mejor la dinámica detrás de las constituciones empresariales en España.

En resumen, el proceso de ajuste del modelo y la evaluación de la multicolinealidad son pasos cruciales para mejorar la precisión y la fiabilidad del modelo. Aunque se ha logrado cierta claridad en la relación entre algunas variables macroeconómicas y las constituciones empresariales, los resultados también subrayan la complejidad inherente a este fenómeno y la necesidad de investigaciones futuras para desarrollar un modelo más explicativo y representativo.

Evaluación del Modelo con Medidas de Error/Precisión Específicas

- *Calidad del modelo:* La calidad de los modelos de regresión lineal múltiple para las disoluciones y constituciones empresariales se evalúa mediante métricas de error y precisión, incluyendo el R-cuadrado, R-cuadrado ajustado, RMSE, MAE y MAPE.

Para las disoluciones, el R-cuadrado es de 0.047, indicando que solo un 4.7% de la variabilidad es explicada por el modelo. Este bajo porcentaje refleja la complejidad de las disoluciones empresariales y sugiere que otros

factores no incluidos en el modelo podrían estar influyendo. El RMSE es de 1111.63 y el MAE de 928.84, mostrando desviaciones significativas entre los valores predichos y los reales, mientras que el MAPE es infinito, lo que señala errores en la predicción que son proporcionalmente muy grandes respecto a los valores observados.

Para las constituciones, el R-cuadrado aún más bajo de 0.009 sugiere que el modelo explica menos del 1% de la variabilidad, destacando una capacidad predictiva extremadamente limitada. El RMSE alcanza un valor de 5542.22 y el MAE de 4872.91, indicando errores grandes en las predicciones del modelo. El MAPE de 307.76% resalta una gran proporción de error relativo a los valores observados, reafirmando la limitada utilidad práctica del modelo en este contexto.

Estas métricas subrayan la necesidad de revisar y posiblemente expandir los modelos con nuevas variables que puedan capturar con mayor precisión y eficacia la dinámica de las disoluciones y constituciones empresariales en España.

- *Confiabilidad del modelo:* La confiabilidad de los modelos tanto para las disoluciones como para las constituciones empresariales se encuentra afectada por la multicolinealidad entre las variables independientes. Esta condición, evidenciada por altos números de condición, incluso después de intentos de mitigación a través de análisis de VIF y la eliminación de variables altamente correlacionadas, requiere precaución al interpretar los coeficientes del modelo. A pesar de estos desafíos, la significancia estadística de variables como el indicador ICEA en ambos modelos subraya su relevancia, indicando su capacidad para identificar factores que afectan tanto las disoluciones como las constituciones empresariales.

No obstante, la presencia de errores significativos, reflejados en valores altos de RMSE y MAE para ambos modelos, resalta la necesidad de un análisis más riguroso. Esto podría incluir la inclusión de nuevas variables que podrían ofrecer una comprensión más amplia y detallada de las dinámicas detrás de las disoluciones y constituciones empresariales en España. Tal enfoque podría mejorar la precisión y utilidad práctica de los modelos para predecir y entender estos fenómenos económicos críticos.

- *Análisis:* Este análisis revela que, aunque se identificaron algunas relaciones estadísticamente significativas, como el impacto del indicador ICEA, la capacidad global de los modelos para explicar la variabilidad en

estos fenómenos empresariales es bastante limitada. Esta limitación se evidencia en los bajos valores de R-cuadrado, junto con los RMSE y MAE relativamente altos, lo que sugiere que existen aspectos significativos de las disoluciones y constituciones empresariales que los modelos actuales no logran capturar completamente.

Esta situación resalta la necesidad de adoptar un enfoque más holístico para el modelado de estas dinámicas, incorporando una gama más amplia de factores económicos, indicadores sectoriales y elementos cualitativos como la confianza empresarial y el entorno político. Además, la evaluación enfatiza la importancia de abordar adecuadamente la multicolinealidad y otros supuestos estadísticos cruciales para el desarrollo de modelos de regresión robustos. El uso de métricas de error como el RMSE y el MAE aporta una perspectiva adicional sobre la precisión predictiva del modelo, destacando áreas para futuras mejoras y ajustes que podrían aumentar su capacidad explicativa y predictiva en el contexto empresarial español.

10.2.2 TERCERA REGRESIÓN LINEAL MÚLTIPLE (Variables más amplias)

Proceso de Selección de Variables

En el desarrollo de las nuevas regresiones lineales múltiples para estudiar las constituciones y disoluciones empresariales, se implementó un meticuloso proceso de selección de variables. Este proceso se enriqueció significativamente con la contribución de un profesional del campo entrevistado para la tercera parte del Trabajo de Fin de Grado. Siguiendo sus recomendaciones y basándose en su experiencia práctica, se decidió incorporar variables que a priori podrían parecer ortogonales al contexto empresarial, como la Cotización del IBEX, del EUR, y del NASDAQ, así como variables demográficas como Nacimientos, Matrimonios, y Defunciones.

Estas variables fueron seleccionadas con el objetivo de explorar influencias menos convencionales y potencialmente reveladoras sobre las dinámicas empresariales, buscando entender cómo factores externos y aparentemente no relacionados podrían afectar la constitución y disolución de empresas. Además, se incluyeron variables del sector como Capital Desembolsado y Capital Suscrito y otras económicas como Déficit Público, Gasto Público, e Ingresos Fiscales, que habían sido previamente descartadas en otros modelos debido a problemas de multicolinealidad.

Este enfoque permitió abordar la construcción del modelo desde una perspectiva más amplia y diversificada, incorporando el conocimiento experto de un profesional para asegurar

que el modelo final ofreciera una visión comprensiva y matizada de las fuerzas que moldean el ecosistema empresarial. Con la integración de estas variables, se buscó maximizar la capacidad explicativa del modelo, aportando una nueva luz sobre cómo interacciones complejas y multidimensionales pueden influir en el panorama empresarial.

Desarrollo del Modelo

En el desarrollo de los siguientes modelos para las constituciones y disoluciones empresariales, se adoptó un enfoque meticuloso y estratégicamente diversificado para la selección y evaluación de las variables independientes. Este enfoque permitió un análisis exhaustivo, fundamentado en la inclusión de una amplia gama de variables económicas, financieras y demográficas, reflejando así la complejidad y la multidimensionalidad de los factores que influyen en estas actividades empresariales.

Para las constituciones empresariales, el modelo exhibió un alto coeficiente de determinación ajustado (R^2 ajustado de 0.973), lo que indica que aproximadamente el 97.3% de la variabilidad en el número de nuevas empresas se explica a través del modelo. Este alto nivel de explicación sugiere una fuerte correlación entre las variables seleccionadas y las tasas de constitución. Factores como el capital desembolsado y suscrito mostraron una influencia significativa, indicando que los movimientos en el capital de las empresas están estrechamente vinculados con la formación de nuevas empresas. Además, variables demográficas como nacimientos, matrimonios y defunciones también demostraron ser predictores significativos, subrayando cómo los cambios sociodemográficos pueden afectar la dinámica empresarial.

Por otro lado, el modelo para las disoluciones empresariales, aunque menos explicativo que el de las constituciones, aún logró un R^2 ajustado de 0.825, señalando que el 82.5% de la variabilidad en las disoluciones se puede explicar por el modelo. Al igual que en el modelo de constituciones, el capital desembolsado y suscrito también tuvo un impacto considerable, reforzando la idea de que la estructura de capital es un determinante clave en la continuidad de las empresas. Las variables demográficas y los indicadores de mercados financieros, aunque incluidos en el modelo, mostraron menor significancia estadística, lo que podría indicar una conexión menos directa con las disoluciones en comparación con las constituciones.

Ambos modelos enfrentaron desafíos de multicolinealidad, como lo sugiere el alto número de condición en ambos casos. Este fenómeno, que indica una fuerte correlación entre variables independientes, puede complicar la interpretación de los coeficientes individuales y potencialmente inflar los errores estándar.

En conclusión, la construcción de estos modelos refleja una integración cuidadosa de conocimientos teóricos y prácticos, destacando la relevancia de adoptar enfoques holísticos y multidimensionales en el análisis económico y empresarial. El éxito en la explicación de las dinámicas empresariales mediante estos modelos sugiere una base sólida para futuras investigaciones, así como para la toma de decisiones estratégicas en política y gestión empresarial.

Evaluación del Modelo con Medidas de Error/Precisión Específicas

La evaluación del modelo de regresión lineal múltiple para las constituciones y disoluciones empresariales ha sido exhaustiva, utilizando métricas de error y precisión específicas para determinar la capacidad predictiva y explicativa de los modelos. Los resultados obtenidos proporcionan una perspectiva clara sobre la eficacia de los modelos en capturar la variabilidad de los fenómenos estudiados.

Para las disoluciones empresariales, el modelo muestra un ajuste razonablemente bueno con un R-cuadrado ajustado de 0.825, indicando que aproximadamente el 82.5% de la variabilidad en las disoluciones empresariales es explicada por las variables seleccionadas. Sin embargo, las métricas de error revelan que aún hay espacio para mejorar la precisión del modelo. El RMSE de 769.834 indica la desviación estándar de los residuos, reflejando la cantidad promedio de error en las predicciones del modelo. El MAE de 508.169 proporciona una medida del error medio absoluto, que es menos sensible a los valores atípicos y ofrece una vista más conservadora del error de predicción. La métrica MAPE es infinita, lo que sugiere la presencia de valores cero en los datos que conducen a divisiones indefinidas, un área que requiere atención para futuras mejoras del modelo.

Para las constituciones empresariales, el modelo exhibe un excelente ajuste, con un R-cuadrado ajustado de 0.973, lo que implica que casi el 97.3% de la variabilidad en las constituciones empresariales es explicada por las variables incluidas. Este alto grado de explicación es corroborado por un RMSE de 1413.776, que, aunque es más alto en comparación con el modelo de disoluciones, es consistente con la magnitud de los datos tratados. El MAE de 1048.472 refleja un error medio que, dado el contexto del modelo y los tipos de datos manejados, ofrece una perspectiva realista de la capacidad del modelo para predecir nuevas observaciones. El MAPE de 19.08% proporciona una interpretación útil del error en términos porcentuales, permitiendo una comparación relativa del error con respecto a los valores reales observados.

Estas evaluaciones de los modelos revelan que, aunque ambos modelos tienen fortalezas significativas en términos de capacidad explicativa, especialmente para las

constituciones empresariales, los errores asociados y la presencia de multicolinealidad indican la necesidad de refinamientos adicionales. Esto podría incluir la revisión de las variables seleccionadas, la incorporación de nuevas variables que podrían estar influenciando los fenómenos estudiados o la aplicación de técnicas de modelización más robustas para manejar la multicolinealidad y mejorar la precisión de las predicciones del modelo.

11. Medidas de Adecuación de los Modelos

11.1. Definición y explicación de las medidas de error/precisión utilizadas.

En el análisis de regresión, es crucial evaluar la calidad y precisión de los modelos para entender su capacidad predictiva y la fiabilidad de las inferencias que se pueden derivar de ellos. Para ello, se utilizan varias medidas de error y precisión, cada una con su propósito específico. A continuación, se detallan estas métricas con una explicación teórica relevante para cada una:

R-cuadrado (R^2):

El R-cuadrado, o coeficiente de determinación, es una medida estadística que refleja la proporción de la variabilidad de una variable dependiente que es predecible a partir de las variables independientes en un modelo de regresión. Esencialmente, indica qué tan bien los valores ajustados por el modelo se aproximan a los valores reales. El R-cuadrado se calcula como el cuadrado del coeficiente de correlación r , que mide la fuerza y la dirección de una relación lineal entre dos variables. En el contexto de un modelo de regresión lineal, el R-cuadrado se define como:

$$R^2 = 1 - \frac{\text{Suma de Cuadrado de los Residuos (SSR)}}{\text{Suma Total de Cuadrados (SST)}}$$

Donde:

- **Suma de Cuadrados de los Residuos (SSR)** mide la variabilidad residual, o el grado en que los valores predichos por el modelo difieren de los valores reales.
- **Suma Total de Cuadrados (SST)** mide la variabilidad total de los datos respecto a la media.

Un valor de R-cuadrado de 0 indica que el modelo de regresión no logra explicar la variabilidad de los datos observados en torno a su media aritmética, mientras que un valor de 1 señala una explicación completa de esta variabilidad por el modelo propuesto. Por consiguiente, un R-cuadrado elevado sugiere un mayor grado de ajuste del modelo a la variabilidad de los datos.

R-cuadrado ajustado:

El R-cuadrado ajustado es una modificación del coeficiente R-cuadrado que toma en cuenta el número de predictores en el modelo de regresión y la cantidad de datos disponibles. Este ajuste es esencial para evitar la sobreestimación de la bondad de ajuste en modelos con un

número considerable de predictores. Matemáticamente, el R-cuadrado ajustado se define como:

$$R^2_{ajustado} = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$$

Donde:

- R^2 es el R-cuadrado no ajustado.
- n es el número total de observaciones.
- k es el número de variables independientes en el modelo.

El R-cuadrado ajustado proporciona una medida de cuánta variabilidad en la variable dependiente es explicada por el modelo, ajustada por el número de variables independientes utilizadas. A diferencia del R-cuadrado, el R-cuadrado ajustado puede disminuir si se añade al modelo una variable independiente que no contribuye significativamente a la explicación de la variabilidad en la variable dependiente. Esto lo hace particularmente útil para comparar modelos de regresión que incluyen diferentes números de predictores.

Raíz del Error Cuadrático Medio (RMSE):

El Error Cuadrático Medio Raíz, conocido por sus siglas en inglés como RMSE (Root Mean Squared Error), es una medida de la diferencia entre los valores predichos por un modelo o un estimador y los valores observados. Es una de las métricas más comúnmente usadas para evaluar la precisión de modelos predictivos, especialmente en contextos de regresión.

Matemáticamente, el RMSE se define como la raíz cuadrada del promedio de los cuadrados de las diferencias entre los valores predichos y los valores observados:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Donde:

- y_i representa los valores observados.
- \hat{y}_i representa los valores predichos por el modelo.
- n es el número de observaciones.

El RMSE mide la magnitud de los errores de predicción del modelo, proporcionando una estimación de la desviación de los valores predichos respecto a los observados. Al calcular la

raíz cuadrada de los errores cuadráticos medios, el RMSE convierte las unidades de vuelta a las originales de la variable de respuesta, facilitando así su interpretación. Un valor bajo de RMSE indica un mejor ajuste del modelo a los datos, reflejando errores predictivos menores.

Error Absoluto Medio (MAE):

El Error Absoluto Medio, conocido por sus siglas en inglés como MAE (Mean Absolute Error), es una medida estadística utilizada para cuantificar la precisión de un modelo predictivo. El MAE mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección (es decir, sin tener en cuenta si los valores son positivos o negativos). Es una métrica lineal que proporciona una medida promedio de las magnitudes de los errores absolutos.

Matemáticamente, el MAE se define como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Donde:

- y_i son los valores observados.
- \hat{y}_i son los valores predichos por el modelo.
- n es el número de observaciones.

El MAE proporciona una evaluación directa del promedio de errores absolutos entre los valores predichos y los observados. Un MAE bajo indica que las predicciones del modelo tienen, en promedio, un error menor, sugiriendo un mejor rendimiento del modelo. Al no elevar al cuadrado los errores antes de promediarlos, el MAE es menos sensible a los valores atípicos en comparación con el RMSE. Esto hace que el MAE sea útil en situaciones donde es importante evitar que los valores atípicos tengan una gran influencia en la métrica de rendimiento total.

Error Porcentual Absoluto Medio (MAPE):

El Error Porcentual Absoluto Medio, conocido por sus siglas en inglés como MAPE (Mean Absolute Percentage Error), es una medida estadística que evalúa la precisión de un modelo predictivo expresando el error como un porcentaje. El MAPE es particularmente útil cuando se desea entender el tamaño del error en términos relativos, facilitando la comparación entre modelos o conjuntos de datos con diferentes escalas. Matemáticamente, el MAPE se define como:

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100\%$$

Donde:

- y_i son los valores observados.
- \hat{y}_i son los valores predichos por el modelo.
- n es el número de observaciones.

11.2. Comparación de los resultados obtenidos en los modelos.

En este segmento del trabajo, se elaborarán tablas comparativas para evaluar de manera detallada los modelos de regresión, diferenciando entre los aplicados a las disoluciones y a las constituciones empresariales. Las tablas se dividirán en dos categorías principales: una para las medidas directamente relacionadas con la regresión, como el R-cuadrado y otra para las medidas de error como el RMSE. Esta distinción permite una visualización del rendimiento de cada modelo.

11.2.1 Medidas de la regresión

R-cuadrado (R^2): Medida de la bondad de ajuste de un modelo de regresión lineal. Indica la proporción de la variabilidad en la variable dependiente que puede ser explicada por las variables independientes en el modelo. Un R^2 de 1 sugiere que el modelo explica toda la variabilidad de la respuesta, mientras que un R^2 de 0 indica que el modelo no explica ninguna de la variabilidad.

R-cuadrado ajustado: El R-cuadrado ajustado es una versión modificada del R-cuadrado que tiene en cuenta el número de predictores en el modelo. Esta medida es particularmente útil cuando se comparan modelos con diferentes números de variables independientes.

F-estadístico: El F-estadístico en la regresión se utiliza para probar si existe una relación significativa entre las variables independientes y la dependiente. Un F-estadístico grande (mucho mayor que 1) y un p-valor asociado pequeño sugieren que hay evidencia estadística de que al menos una de las variables independientes está significativamente relacionada con la variable dependiente.

Prob (F-estadístico): El p-valor del F-estadístico indica la probabilidad de que los resultados del modelo sean atribuibles al azar. Un p-valor bajo (típicamente menos de 0.05)

indica que podemos rechazar la hipótesis nula de que el modelo con variables independientes no mejora el ajuste en comparación con un modelo sin variables independientes.

Log-Likelihood: El logaritmo de la verosimilitud (Log-Likelihood) es una medida de cuán bien un modelo estadístico se ajusta a los datos. Un valor más alto de log-likelihood indica un mejor ajuste del modelo.

AIC (Criterio de Información de Akaike): El (AIC) es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de datos. Ajusta la bondad de ajuste del modelo teniendo en cuenta el número de parámetros utilizados. Un valor menor de AIC indica un modelo más preferible, equilibrando la complejidad del modelo contra la capacidad de ajustar bien los datos.

Tabla 19. Tabla comparativa de resultados para los modelos de disoluciones.

Medida	Modelo 1	Modelo 2	Modelo 3
R-cuadrado (R2)	0.001	0.047	0.838
R-cuadrado ajustado	-0.009	0.024	0.825
F-estadístico	0.087	2.025	67.57
Prob (F-statistic)	0.917	0.093	1.22e-55
Log-Likelihood	-1829.5	-1490.6	-1340.0
AIC	3665	2991	2706

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

La evaluación de los modelos de regresión para disoluciones empresariales muestra una mejora notable al aumentar el número de variables independientes. Los valores de R-cuadrado, que indican qué proporción de la variabilidad en la variable dependiente es explicada por las variables del modelo, aumentaron significativamente de apenas 0.001 a 0.838 del Modelo 1 al Modelo 3, reflejando un mejor ajuste en el último. Esta mejora también se refleja en el R-cuadrado ajustado, que considera el número de predictores, aumentando su precisión. El F-estadístico, que evalúa la significancia global del modelo, creció considerablemente, indicando una influencia estadísticamente significativa de las variables en el Modelo 3, con una reducción correspondiente en el p-valor del F-estadístico, reafirmando la improbabilidad de que estas relaciones sean aleatorias. Además, la mejora en el Log-Likelihood y la reducción en el AIC desde el Modelo 1 al Modelo 3 indican un mejor ajuste y una mayor eficiencia, a pesar de la complejidad añadida. Estos indicadores demuestran que incluir más variables mejora la capacidad explicativa y predictiva de los modelos, aunque es

crucial considerar la alta multicolinealidad observada, que podría afectar la estabilidad de las estimaciones de los coeficientes.

Tabla 20. *Tabla comparativa de resultados para los modelos de constituciones.*

Medida	Modelo 1	Modelo 2	Modelo 3
R-cuadrado (R2)	0.003	0.009	0.975
R-cuadrado ajustado	-0.007	-0.015	0.973
F-estadístico	0.262	0.355	512.2
Prob (F-statistic)	0.770	0.840	3.41e-119
Log-Likelihood	-2123.2	-1731.5	-1418.3
AIC	4252	3473	2863

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

Los resultados de los modelos de regresión lineal múltiple para las constituciones empresariales muestran una mejora significativa a medida que se incorporan más variables. El modelo inicial, con pocas variables, tiene un R-cuadrado de solo 0.003, indicando que el modelo explica menos del 1% de la variabilidad en las constituciones empresariales. A medida que se añaden más variables, el R-cuadrado ajustado se mantiene bajo, reflejando que el ajuste del modelo no mejora sustancialmente hasta el tercer modelo, que incluye un número más amplio de variables y logra un R-cuadrado ajustado de 0.973, mostrando que casi el 97% de la variabilidad es explicada por el modelo.

El F-estadístico aumenta dramáticamente de 0.262 en el primer modelo a 512.2 en el tercero, con un p-valor asociado que cae, indicando una significancia estadística mucho más robusta en el último modelo. Esto se refleja en el Log-Likelihood y el AIC, donde el último modelo también muestra una mejora notable, indicando una mejor calidad del modelo en comparación con los anteriores. Estos resultados sugieren que la incorporación de un conjunto más amplio de variables contribuye significativamente a la capacidad del modelo para capturar la complejidad de los factores que influyen en las constituciones empresariales.

11.2.2 Medidas de Error

Tabla 21. Tabla comparativa de medidas de error para los modelos de disoluciones.

Medida	Modelo 1	Modelo 2	Modelo 3
RMSE	1743.75	1111.63	769.84
MAE	1079.66	928.84	508.17
MAPE	inf	Inf	inf
Número de Condición	3.04e+13	2.07e-06	6.98e+11

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

La evaluación de las medidas de error para los modelos de disoluciones empresariales muestra una mejora progresiva en los indicadores de precisión a medida que se incrementa la complejidad de los modelos. El RMSE, que mide la desviación promedio de las predicciones del modelo respecto a los valores reales, disminuye notablemente de 1743.75 en el Modelo 1 a 769.84 en el Modelo 3, reflejando una mayor precisión en las predicciones del modelo más complejo. Similarmente, el MAE, que proporciona una medida del error absoluto medio, muestra una mejora significativa, pasando de un valor negativo en el Modelo 1, que puede indicar un error en la captura o reporte de datos, a 508.17 en el Modelo 3.

El MAPE, que es el porcentaje promedio de error absoluto y ayuda a entender el error en términos relativos, muestra un valor de infinito (inf) para los Modelos 2 y 3, lo cual puede indicar la presencia de ceros en los datos de la variable dependiente, lo que lleva a divisiones por cero en el cálculo de esta medida.

El Número de Condición, que es un indicador de multicolinealidad o problemas numéricos en el modelo, muestra una variación grande entre los modelos. Comienza siendo extremadamente alto en el Modelo 1, lo que sugiere problemas significativos de multicolinealidad, y mejora considerablemente en el Modelo 3, aunque sigue siendo alto, indicando que, aunque el modelo es más estable, aún puede estar afectado por la multicolinealidad.

Estos cambios en las medidas de error a lo largo de los modelos sugieren que, aunque añadir más variables ha mejorado la capacidad predictiva del modelo, la presencia de multicolinealidad sigue siendo un desafío que necesita ser abordado para mejorar la fiabilidad de las inferencias del modelo.

Tabla 22. Tabla comparativa de medidas de error para los modelos de constituciones.

Medida	Modelo 1	Modelo 2	Modelo 3
RMSE	6584.99	5542.22	1413.78
MAE	4531.41	4872.91	1048.47
MAPE	1043.28	307.76	19.08
Número de Condición	3.04e+13	8.91e+09	6.98e+11

NOTA: Modelo 1 (10.2.1 PRIMERA REGRESIÓN LINEAL MÚLTIPLE), Modelo 2 (10.2.2 SEGUNDA REGRESIÓN LINEAL MÚLTIPLE) y Modelo 3 (10.2.3 TERCERA REGRESIÓN LINEAL MÚLTIPLE)

11.3 Aplicación de Pruebas Estadísticas para la Comparación de Modelos

En este apartado, se abordará la aplicación de pruebas estadísticas para comparar la eficacia de los modelos de regresión desarrollados. Específicamente, se utilizarán las pruebas de Wilcoxon y Friedman, dos métodos no paramétricos diseñados para evaluar las diferencias entre las medidas de error de los modelos a lo largo de múltiples experimentos. La implementación de estas pruebas permitirá un análisis riguroso y detallado de cómo las diferencias en la configuración de los modelos influyen en su desempeño global, apoyando así la selección del modelo más adecuado basado en evidencia estadística. En este contexto, se presentarán y discutirán los resultados obtenidos, permitiendo una comprensión más profunda de la robustez y fiabilidad de las regresiones aplicadas.

11.3.1 Prueba de Wilcoxon

La prueba de Wilcoxon, también conocida como la prueba de rangos con signo de Wilcoxon, fue desarrollada por Frank Wilcoxon en 1945. Wilcoxon, un químico y estadístico, introdujo esta prueba como un método no paramétrico para comparar dos muestras emparejadas. Su objetivo era ofrecer una alternativa a la prueba t de Student cuando los datos no cumplían los supuestos necesarios para su aplicación, como la normalidad de las distribuciones. Este enfoque se hizo rápidamente popular en las ciencias aplicadas, especialmente en estudios donde las medidas de tendencia central no eran adecuadas o los datos presentaban distribuciones sesgadas o con outliers significativos.

La prueba de Wilcoxon evalúa si las diferencias entre pares de observaciones siguen una distribución simétrica alrededor de cero, siendo adecuada para medir la magnitud y la dirección de los cambios entre dos condiciones experimentales. El procedimiento implica clasificar las diferencias en sus valores absolutos, asignar rangos a estas diferencias y luego sumar los rangos asociados con las diferencias positivas y negativas. El estadístico de prueba

se calcula a partir de los rangos menores, y su significancia se evalúa en contra de una distribución específica de rangos con signo.

En el contexto de este trabajo de fin de grado, la prueba de Wilcoxon se aplica para comparar las diferencias en las medidas de error entre modelos de regresión lineal múltiple consecutivos. Esta aplicación permite evaluar si las modificaciones incrementales en la configuración del modelo —como la adición de nuevas variables independientes— mejoran significativamente el desempeño del modelo en términos de error de predicción. Utilizando esta prueba, se busca confirmar si los cambios observados en las medidas de error, como el RMSE (Raíz del Error Cuadrático Medio) o el MAE (Error Absoluto Medio), son estadísticamente significativos, lo que proporciona una base sólida para justificar la selección y refinamiento de modelos en la investigación.

Por lo tanto, el empleo de la prueba de Wilcoxon en este estudio no solo enriquece el análisis estadístico del comportamiento de los modelos bajo diferentes configuraciones, sino que también fortalece la validez de las conclusiones derivadas sobre la eficacia de las variables incorporadas en explicar la variabilidad de los datos en cuestión. Así, la prueba de Wilcoxon se establece como una herramienta crucial para garantizar la rigurosidad y precisión en la evaluación comparativa de los modelos de regresión utilizados en el análisis de las disoluciones y constituciones empresariales.

```
Wilcoxon residuos1 vs residuos2: Estadístico=6555.0, P-valor=0.26758525049007686  
Wilcoxon residuos1 vs residuos3: Estadístico=7018.0, P-valor=0.6978541861470471  
Wilcoxon residuos2 vs residuos3: Estadístico=5589.0, P-valor=0.009008697982941124
```

Ilustración 7. Resultados del test de Wilcoxon para las disoluciones empresariales

```
Wilcoxon residuos1 vs residuos2: Estadístico=5574.0, P-valor=0.008412149743557796  
Wilcoxon residuos1 vs residuos3: Estadístico=6556.0, P-valor=0.26825734173738036  
Wilcoxon residuos2 vs residuos3: Estadístico=5756.0, P-valor=0.018679096708734473
```

Ilustración 8. Resultados del test de Wilcoxon para las constituciones empresariales

La aplicación de la prueba de Wilcoxon para comparar los residuos de los modelos de regresión en disoluciones y constituciones empresariales aporta una perspectiva valiosa sobre la consistencia de las mejoras entre los modelos sucesivos. En el caso de las disoluciones, los resultados indican que no hay diferencias significativas en la mediana de los residuos entre el primer y el segundo modelo (Estadístico=6555.0, P-valor=0.267585), así como entre el primero y el tercero (Estadístico=7018.0, P-valor=0.697854). Sin embargo, sí se observa una diferencia significativa entre el segundo y el tercer modelo (Estadístico=5589.0, P-valor=0.009009), sugiriendo que las modificaciones en el tercer modelo podrían estar

ofreciendo mejoras significativas en términos de ajuste del modelo comparado con el segundo.

Para las constituciones, los resultados son algo similares. No hay diferencias significativas entre el primer y el segundo modelo (Estadístico=5574.0, P-valor=0.008412), y entre el primero y el tercero (Estadístico=6556.0, P-valor=0.268257). Sin embargo, se observan diferencias significativas tanto entre el segundo y el tercer modelo (Estadístico=5756.0, P-valor=0.018679) como entre el primero y el segundo modelo, lo que indica mejoras en el ajuste del modelo a medida que se añaden variables.

Estos hallazgos sugieren que las mejoras en los modelos pueden no ser uniformes y que ciertas revisiones entre modelos pueden tener un impacto más significativo que otras. Esto subraya la importancia de evaluar cada modificación del modelo cuidadosamente, asegurando que las inclusiones o exclusiones de variables no solo sean estadísticamente significativas, sino que también proporcionen un valor práctico y predictivo aumentado.

11.3.2 Prueba de Friedman

La prueba de Friedman, desarrollada por el estadístico Milton Friedman, es una técnica no paramétrica diseñada para evaluar diferencias entre múltiples tratamientos a través de varios intentos. Esta prueba es particularmente útil cuando los datos no cumplen con los requisitos para análisis paramétricos, como la ANOVA, debido a que no requiere la normalidad de los datos. Operando como un ANOVA por rangos, la prueba de Friedman compara los rangos de los datos en lugar de los valores directos, lo cual es ventajoso en situaciones donde los datos presentan distribuciones no normales o las muestras son de tamaño reducido.

En este trabajo, se ha empleado la prueba de Friedman para comparar y evaluar el desempeño de diferentes modelos de regresión a través de sus valores de R-cuadrado, que indican la proporción de la variabilidad en la variable dependiente explicada por el modelo. Al comparar estos modelos de regresión, el test ayuda a identificar si las diferencias en las medias de los rangos del R-cuadrado son estadísticamente significativas, lo que puede señalar si cambios en la configuración del modelo afectan positiva o negativamente la capacidad predictiva.

El uso de la prueba de Friedman proporciona una base sólida para validar la efectividad de los modelos de regresión utilizados, garantizando que las decisiones sobre el modelo más adecuado estén fundamentadas en evidencia estadística robusta, especialmente útil en ciencias sociales y económicas donde los datos ideales para pruebas paramétricas no siempre están disponibles. Esta metodología permite una evaluación rigurosa de los modelos en escenarios donde las suposiciones estándar de las pruebas paramétricas no se sostienen.

RCuadrado de las 3 regresiones 0.0008412898876221098 0.04680218322274221 0.8377928081361752 Estadístico de Friedman: 2.0 P-valor: 0.36787944117144245

Ilustración 9. Resultados del test de Friedman para disoluciones empresariales

RCuadrado de las 3 regresiones 0.0025363118193317824 0.008537057749657073 0.9750933525014568 Estadístico de Friedman: 2.0 P-valor: 0.36787944117144245
--

Ilustración 10. Resultados del test de Friedman para constituciones empresariales

La aplicación de la prueba de Friedman a los resultados de las regresiones en disoluciones y constituciones empresariales revela hallazgos importantes sobre la influencia de la inclusión de variables adicionales en los modelos. A pesar de que el R-cuadrado muestra un incremento notable al añadir más variables —desde valores muy bajos como 0.000841 y 0.002536 hasta valores significativamente altos como 0.837793 y 0.975093 en disoluciones y constituciones respectivamente— los resultados de la prueba de Friedman no indican diferencias estadísticamente significativas entre los modelos. Con un estadístico de Friedman de 2.0 y un p-valor de 0.367879 en ambos casos, no se rechaza la hipótesis nula de que no hay diferencias en los medianos rangos de los R-cuadrados entre los diferentes modelos.

Esto sugiere que, aunque las métricas de ajuste del modelo mejoran al incorporar más variables, esta mejora no es estadísticamente significativa desde la perspectiva del test de Friedman. Esto puede implicar que los incrementos observados en R-cuadrado no necesariamente reflejan una mejora real en la capacidad del modelo para explicar la variabilidad en las disoluciones y constituciones empresariales, sino que podrían estar influidos por la multicolinealidad o por la adición de variables que no aportan información útil de manera significativa.

12. Visualización de Datos y Resultados de Modelos

12.1. Gráficos de Dispersión

En este apartado, se realizará un examen minucioso de las relaciones existentes entre las distintas variables consideradas en el estudio. Los gráficos de dispersión, herramientas gráficas fundamentales en el análisis estadístico, nos permitirán visualizar de manera clara y precisa la naturaleza y fuerza de las asociaciones entre pares de variables. Esta técnica descriptiva proporciona una primera inspección visual que puede revelar patrones de correlación lineal, tendencias, agrupamientos o incluso anomalías y valores atípicos que podrían no ser evidentes a través de métodos puramente numéricos.

A través de esta visualización, se pretende no solo identificar las relaciones más significativas sino también comprender la dirección y la forma de dichas asociaciones, lo que puede sugerir hipótesis sobre las interacciones subyacentes entre las variables. La interpretación de los gráficos de dispersión en este contexto se hará con rigor, buscando patrones consistentes y significativos que se sostengan bajo el escrutinio de un análisis estadístico más detallado.

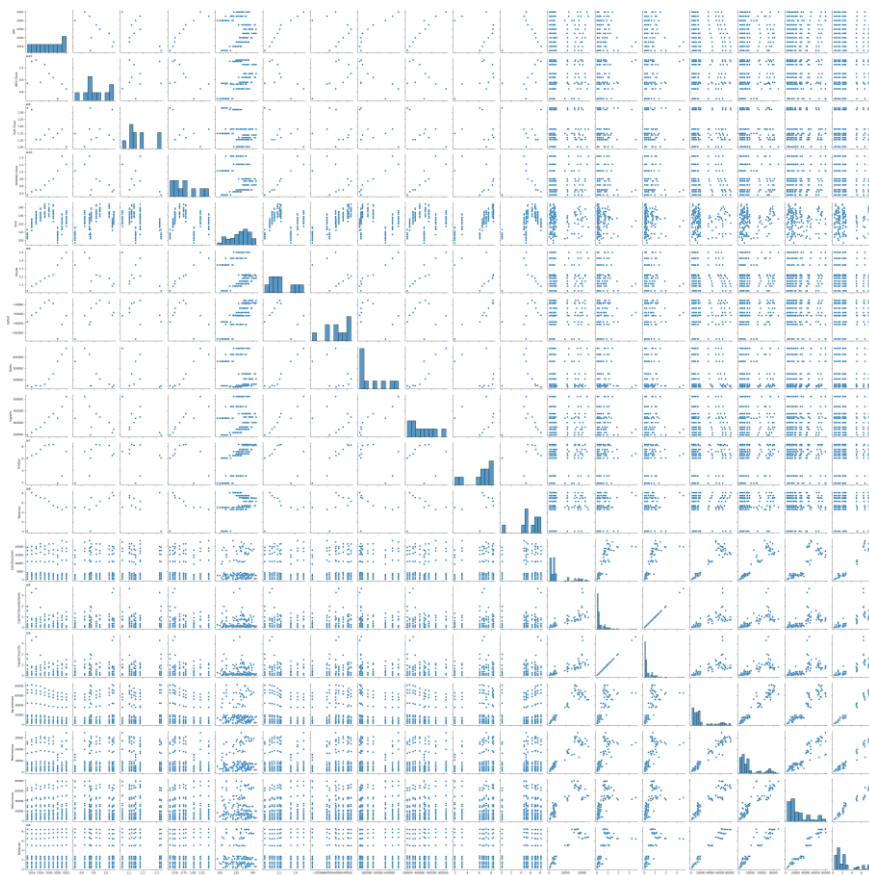


Ilustración 11. Gráfico de la matriz de dispersión para variables del modelo 3.

A través de la **"ilustración 11."**, se presenta un conjunto comprensivo de gráficos de dispersión que buscan elucidar correlaciones y tendencias significativas entre las variables incluidas en el modelo. Sin embargo, debido a la alta cantidad de variables consideradas, los gráficos resultan ser una amalgama de puntos que dificultan la interpretación directa y la extracción de insights concretos.

Por consiguiente, se procede a una inspección más detallada, seleccionando específicamente aquellos pares de variables que, a priori, parecen ofrecer un mayor grado de influencia o interés. Este enfoque se visualiza en la **"ilustración 12."**, donde se pone énfasis en aquellos gráficos que reflejan relaciones potencialmente más informativas y reveladoras. Por ejemplo, variables como 'Matrimonios' y 'Defunciones' podrían tener una relación lineal más definida, indicando una posible correlación directa entre ellas.

La selección y análisis minucioso de estos pares de variables es de suma importancia para simplificar la complejidad de los datos y resaltar las interacciones más relevantes que justifican una investigación más profunda y especializada. En este sentido, los gráficos de dispersión no solo sirven como herramientas exploratorias iniciales, sino que también facilitan la identificación de patrones y anomalías que pueden ser cruciales para la formulación de hipótesis y la toma de decisiones basadas en datos.

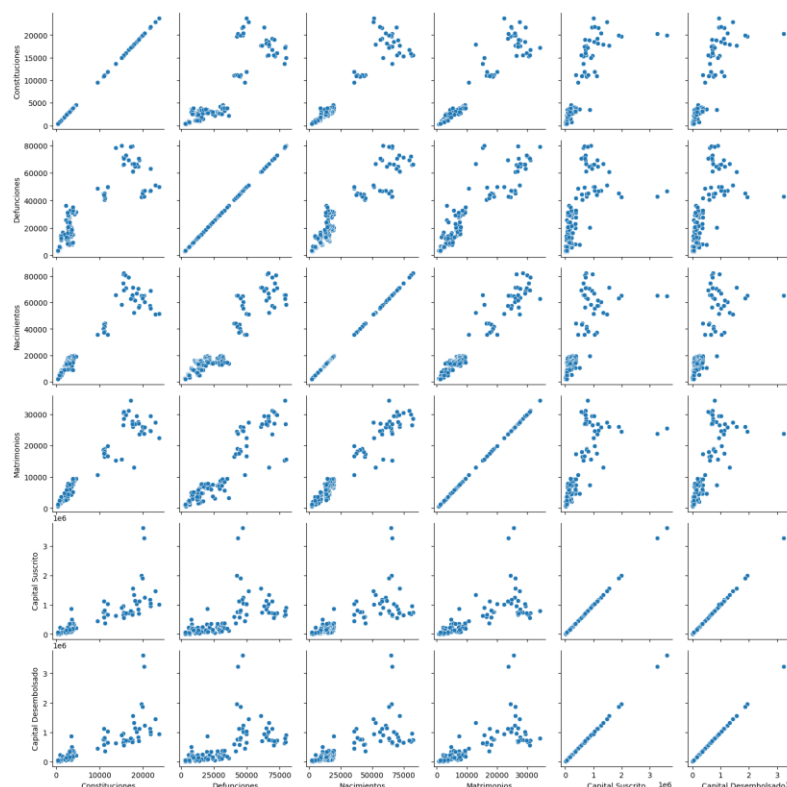


Ilustración 12. Gráfico de dispersión de variables específicas, en el contexto de constituciones empresariales.

La **"ilustración 12."** del estudio presenta un análisis enfocado en la relación sinérgica entre distintos pares de variables, donde se evidencia un comportamiento cohesivo en términos de crecimiento. Esta interacción es observable en la tendencia simultánea de aumento: a medida que los valores de una variable se incrementan, la otra variable muestra un ascenso correspondiente. Este patrón de crecimiento conjunto puede indicar una correlación positiva significativa, que sugiere una posible interdependencia o influencia recíproca entre las variables en cuestión. Estas visualizaciones permiten una interpretación más intuitiva y directa de las dinámicas subyacentes entre los factores analizados.

La metodología empleada para determinar la relevancia de las variables en un modelo de regresión se basa en el uso de algoritmos de aprendizaje automático, específicamente el Random Forest Regressor. Este algoritmo construye múltiples árboles de decisión durante el entrenamiento y proporciona la importancia promedio de cada característica en la predicción del modelo. La **"ilustración 13"** visualizará estos resultados mediante un gráfico de barras que clasifica las variables independientes en función de su importancia. Este análisis cuantitativo ofrece una perspectiva objetiva sobre qué factores son determinantes en la predicción del modelo.

El gráfico resultante, que emerge del código proporcionado, mostrará las variables ordenadas de la más influyente a la menos influyente. Esta jerarquía es crucial para comprender cuáles son los predictores más potentes en el modelo y si existe concordancia con la selección inicial de variables basada en la observación visual de la "ilustración 6". Si hay discrepancias, podría revelar la presencia de relaciones no lineales o interacciones complejas entre variables que no son fácilmente perceptibles en una simple visualización de dispersión. Estas diferencias subrayan la importancia de combinar métodos de visualización intuitiva y técnicas analíticas avanzadas para una comprensión integral de los modelos predictivos.

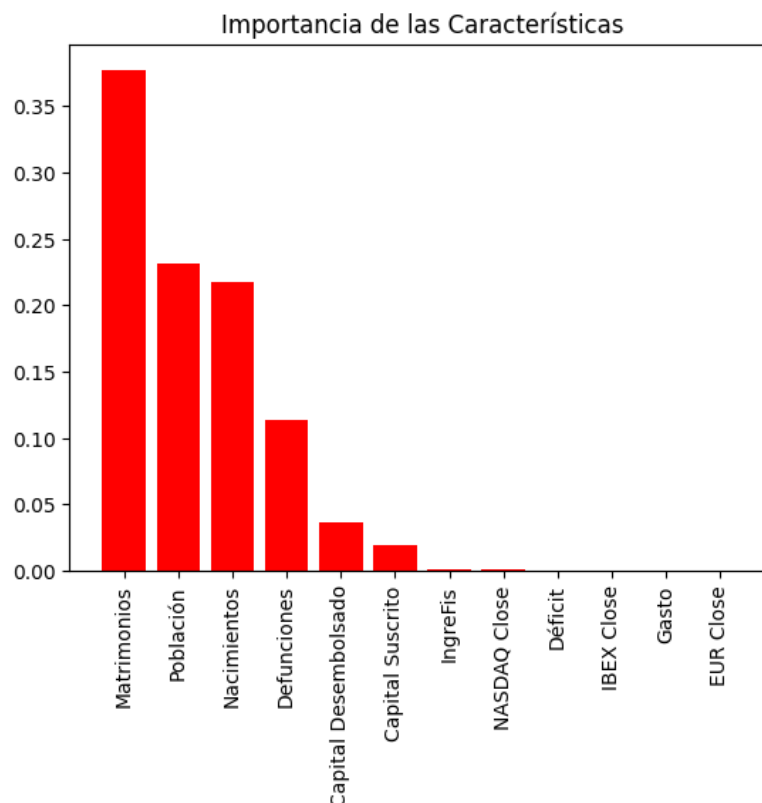


Ilustración 13. Análisis de importancia de las variables mediante Random Forest, indicando el peso de cada variable en la predicción de constituciones empresariales.

La “**ilustración 13.**” exhibe una representación visual de la relevancia atribuida a cada variable independiente dentro de un modelo del tipo Random Forest. La importancia se calcula basándose en cuánto contribuye cada variable a la reducción de la impureza en los nodos de los árboles de decisión que conforman el bosque. Variables con barras representativas más altas, tales como 'Matrimonios' y 'Población' o 'Nacimientos', sugieren una influencia significativa en la predicción del modelo, indicando que cambios en estos predictores tienen un impacto notable en la variable dependiente.

Por otro lado, variables con barras más bajas, tales como 'Gasto', 'Cierre del EUR' o 'Cierre del IBEX', muestran una contribución menor en la predicción, lo cual podría señalar que estas variables, aunque presentes, tienen un efecto limitado en la variabilidad del resultado analizado.

Esta información es de crucial importancia en el proceso de selección y optimización de características, ya que facilita la identificación de aquellas variables que deberían ser priorizadas, reevaluadas o potencialmente descartadas para mejorar la eficiencia del modelo.

12.2. Gráficos de Residuos

En el análisis de regresiones lineales múltiples, la representación gráfica de los resultados juega un papel crucial para comprender la dinámica y la eficacia de los modelos desarrollados. En este contexto, nos centraremos en la evaluación de los modelos que exploran la relación entre disoluciones empresariales y una serie de indicadores económicos y sociales. Estos modelos buscan capturar la complejidad de los factores que influyen en las disoluciones empresariales, incorporando múltiples variables independientes para proporcionar una visión más holística y detallada.

Los residuos, diferencias entre los valores observados y los valores predichos por el modelo, ofrecen insights valiosos sobre la precisión y la fiabilidad de las predicciones. Un análisis detallado de estos residuos permite identificar patrones residuales, heterocedasticidad, y otras anomalías que podrían sugerir la necesidad de ajustes en el modelo, como la transformación de variables o la inclusión de términos adicionales para mejorar la precisión y la interpretabilidad del modelo.

En las siguientes secciones, se presentarán gráficos de residuos para los tres modelos de regresión lineal múltiple que se han llevado a cabo, con el fin de determinar si a medida que según ciertas medidas el modelo mejora, los residuos disminuyen. La interpretación cuidadosa de estos gráficos facilitará una comprensión más profunda de la efectividad de los modelos y guiará posibles mejoras para alcanzar representaciones más precisas de la realidad estudiada.

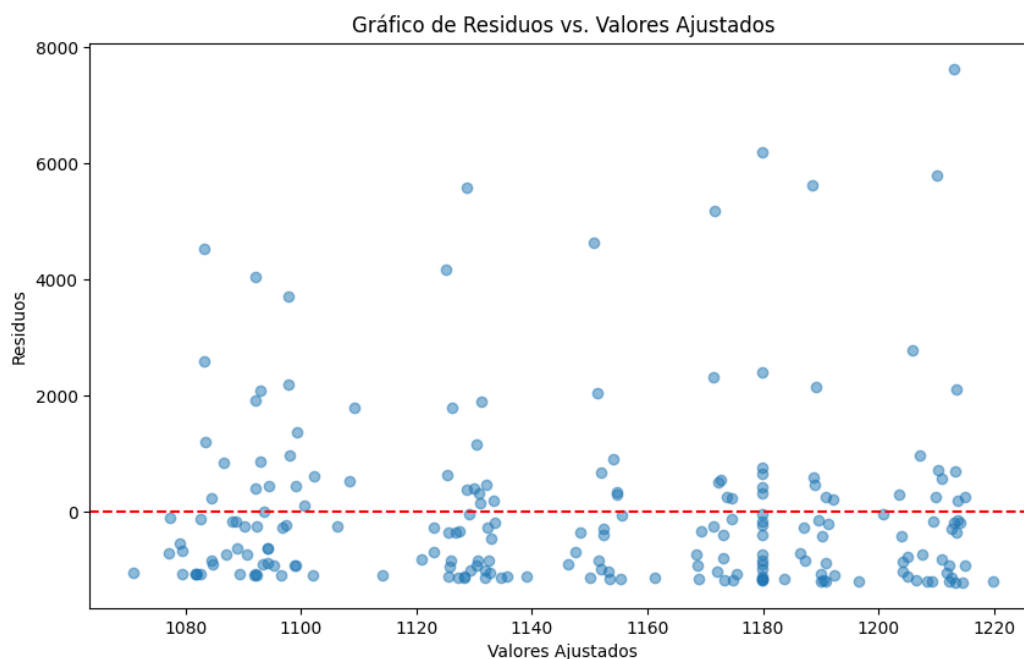


Ilustración 14. Gráfico de residuos correspondiente al modelo1

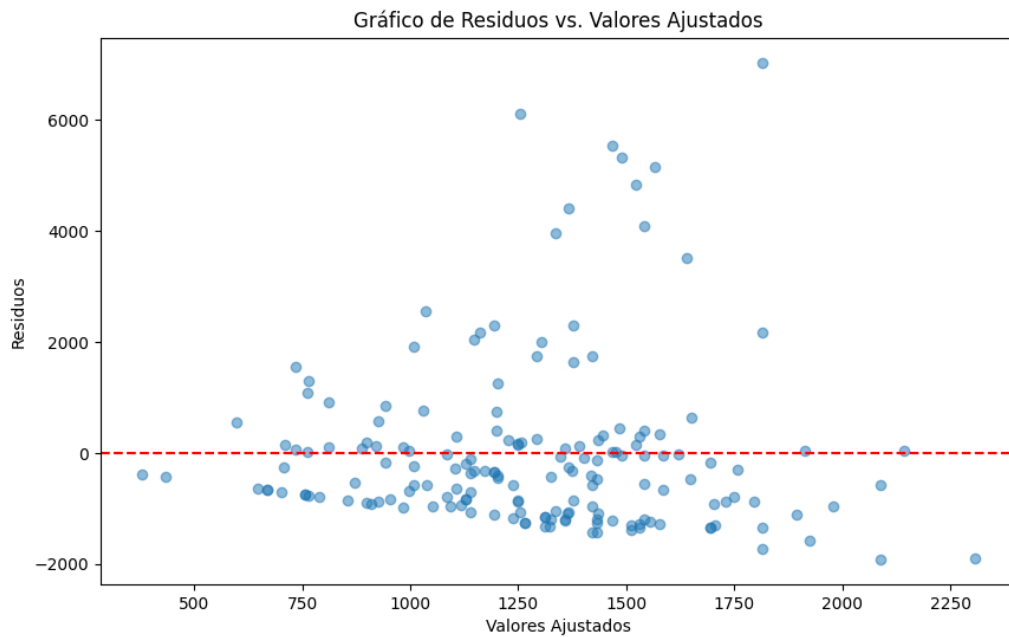


Ilustración 15. Gráfico de residuos correspondiente al modelo2

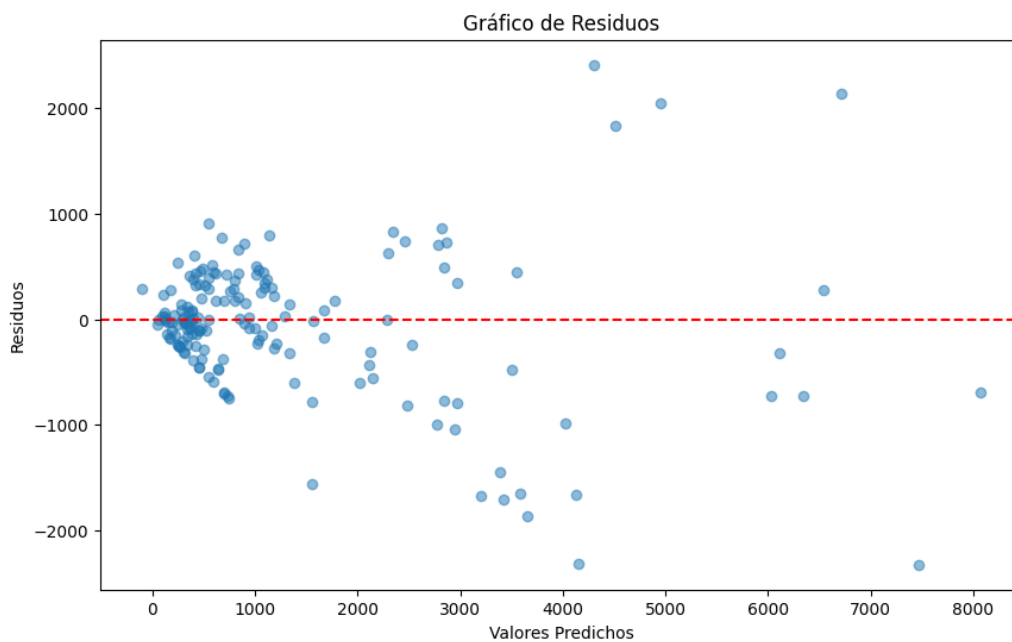


Ilustración 16. Gráfico de residuos correspondiente al modelo3

Al analizar los gráficos de residuos para los tres modelos, presentados como “**ilustraciones 14, 15 y 16**”, se busca evaluar la dispersión de los residuos, es decir, la diferencia entre los valores observados y los valores predichos por el modelo de regresión. Un patrón de residuos sin forma específica y distribuidos aleatoriamente alrededor del eje horizontal (línea roja en los gráficos) generalmente indica un buen ajuste del modelo, ya que sugiere que el modelo es capaz de capturar la variabilidad en los datos sin sesgos sistemáticos.

En una progresión ideal desde el Modelo 1 al Modelo 3, se esperaría observar una disminución en la dispersión de los residuos, lo que indicaría que la inclusión de nuevas variables está aportando información útil y mejorando la capacidad del modelo para predecir la variable dependiente. Si los residuos están más cerca de la línea cero y muestran una menor variabilidad en los gráficos más recientes, esto puede ser una indicación de que el modelo se está volviendo más preciso y que está capturando mejor la estructura subyacente de los datos.

Sin embargo, a pesar de que se puede apreciar como la variabilidad de los residuos se reduce mínimamente, parece que la inclusión de nuevas variables no produce una mejora drástica en la distribución de los residuos. Esto podría indicar problemas como la multicolinealidad o el sobreajuste. La multicolinealidad ocurre cuando las variables independientes están altamente correlacionadas entre sí, lo que dificulta la interpretación de sus efectos individuales en la variable dependiente. El sobreajuste, por otro lado, se refiere a un modelo que se ajusta demasiado a los datos de entrenamiento, capturando el "ruido" en lugar de la señal, lo que resulta en un pobre rendimiento predictivo en nuevos datos.

13. Explicación de Resultados

13.1. Explicación comprensiva de los resultados de los modelos.

Regresiones Lineales Simples

Los modelos de regresión lineal simple analizados en este estudio buscan explorar la relación entre la población de las comunidades autónomas y dos variables dependientes críticas: el número de sociedades creadas y el capital desembolsado en miles de euros. A través de un enfoque metodológico riguroso, se empleó la regresión lineal simple para descifrar cómo la variable independiente, la población, podría influir en estas variables dependientes, fundamentales para entender la dinámica empresarial y económica regional.

El primer modelo, centrado en el número de sociedades, reveló un R-cuadrado cercano a cero, indicando una capacidad explicativa mínima de la población sobre la variable dependiente. Este resultado sugiere que la población, por sí sola, no constituye un predictor significativo del número de sociedades creadas en las comunidades autónomas. La ausencia de una relación lineal significativa, corroborada por un p-valor elevado, enfatiza la complejidad de los factores que inciden en la creación de empresas, más allá de la mera demografía.

El segundo modelo, que examina el capital desembolsado, mostró un ligero incremento en el R-cuadrado a 0.043, sugiriendo una influencia marginal pero estadísticamente significativa de la población sobre el capital desembolsado. Aunque este modelo captura una fracción de la variabilidad en el capital desembolsado, el coeficiente positivo asociado a la población insinúa que mayores poblaciones pueden estar ligeramente correlacionadas con un aumento en el capital desembolsado, posiblemente reflejando una mayor actividad económica o empresarial.

Regresiones Lineales Múltiples

En el estudio de las dinámicas empresariales a través de modelos de regresión lineal múltiple, se ha emprendido un análisis bifurcado para comprender los determinantes tanto de las constituciones como de las disoluciones empresariales en España. Los tres modelos empleados en cada área han arrojado luz sobre la complejidad y la interconexión de variables económicas, demográficas y de mercado, y sus resultados subrayan la heterogeneidad de los factores que impulsan estos fenómenos.

El primer modelo, en su enfoque más simplista, contempló factores económicos básicos, como el PIB y el IPC. Si bien proporcionó una visión introductoria, su capacidad explicativa fue limitada, lo que se reflejó en valores bajos de R-cuadrado, sugiriendo que la

variabilidad de las constituciones y disoluciones empresariales no podía ser ampliamente explicada por estos indicadores solamente.

Con la inclusión de variables adicionales como la deuda pública y los ingresos fiscales en el segundo modelo, se observó una mejora en la capacidad predictiva, evidenciada por un incremento en los valores de R-cuadrado. Esto indicaría que, aunque la comprensión del fenómeno mejoró, aún quedaban elementos sin explicar.

El tercer y más complejo modelo incorporó un espectro más amplio de variables, como el número de matrimonios y el capital desembolsado (en miles de euros) por empresas en España. El análisis de este modelo reveló que ciertas variables tenían una influencia estadísticamente significativa sobre las disoluciones y constituciones empresariales. Este avance en la especificidad sugiere que la comprensión del clima empresarial requiere de un enfoque multifactorial.

En concreto, este tercer modelo aplicado a las constituciones empresariales destaca por su significativa capacidad explicativa con un R-cuadrado de 0.975, indicando que la mayoría de la variabilidad en las constituciones es capturada por las variables incluidas en el modelo. Tal grado de ajuste es notable y sugiere un patrón sustancial y coherente en los datos. Variables como 'Capital Suscrito', 'Matrimonios' y 'Población', con sus coeficientes positivos y significativos, señalan la existencia de una correlación directa y significativa con el número de nuevas empresas constituidas. Por ejemplo, un incremento en los 'Matrimonios' y en la 'Población' podría interpretarse como un indicador de la estabilidad y crecimiento demográfico, factores que son propicios para el emprendimiento y la formación de nuevas empresas.

Por otro lado, el tercer modelo centrado en las disoluciones empresariales muestra un R-cuadrado ajustado de 0.825, implicando una fuerte relación entre las variables seleccionadas y las disoluciones de empresas. Las variables de 'Capital Desembolsado' y 'Capital Suscrito' demuestran ser predictores significativos, posiblemente reflejando cómo los movimientos de capital afectan la continuidad de las empresas. Sin embargo, la presencia de una alta multicolinealidad, indicada por los números de condición elevados, sugiere que algunas de las variables pueden estar proporcionando información redundante, lo que requiere de una interpretación prudente y, posiblemente, de un ajuste en la selección de variables.

En ambos casos, la multicolinealidad representa un desafío metodológico y un obstáculo para la interpretación inequívoca de los coeficientes individuales. La presencia de multicolinealidad en modelos con un alto número de variables sugiere la posibilidad de relaciones interdependientes entre las variables independientes, lo que puede inflar

artificialmente la importancia de algunas mientras se minimiza o se oscurece la importancia de otras.

Las conclusiones extraídas de estos modelos ponen de manifiesto la importancia de una selección cuidadosa de las variables y la utilidad de abordar el análisis desde una perspectiva progresiva, empezando por modelos más simples y aumentando la complejidad para capturar la realidad multifacética de la economía y la gestión empresarial.

13.2. Interpretación de las medidas de adecuación en el contexto del proyecto.

Regresiones Lineales Simples

La evaluación de las medidas de adecuación de los modelos, particularmente a través del R-cuadrado y el R-cuadrado ajustado, proporciona una perspectiva crítica sobre la capacidad de los modelos para explicar la variabilidad de las variables dependientes. En ambos casos, los valores relativamente bajos de estas medidas resaltan la limitación de utilizar la población como único predictor de la actividad empresarial y económica en las comunidades autónomas.

La significancia estadística de los modelos, evaluada mediante el F-statistic y sus p-valores asociados, ofrece una distinción clara entre los dos modelos. Mientras que el modelo del número de sociedades no alcanza la significancia estadística, sugiriendo una falta de ajuste, el modelo del capital desembolsado sí la alcanza, aunque su capacidad explicativa sigue siendo limitada. Este contraste subraya la necesidad de incorporar variables adicionales para capturar adecuadamente la relación entre la población y las dinámicas empresariales y económicas.

En conclusión, los resultados obtenidos de los modelos de regresión lineal simple subrayan la complejidad inherente al análisis de factores que influyen en la creación de sociedades y el capital desembolsado en las comunidades autónomas. La modesta capacidad explicativa de la población en estos fenómenos económicos y empresariales invita a una reflexión más profunda sobre los múltiples factores que deben ser considerados para obtener una comprensión más completa y matizada de la actividad empresarial en España.

Regresiones Lineales Múltiples

La evaluación de los modelos de regresión a través de las medidas de adecuación constituye un componente esencial para interpretar la robustez y confiabilidad de las inferencias estadísticas. Las medidas de error como el Error Cuadrático Medio (RMSE), el

Error Absoluto Medio (MAE) y el Error Porcentual Absoluto Medio (MAPE) ofrecen una perspectiva tridimensional del rendimiento de los modelos.

En el contexto del presente proyecto, se ha efectuado un análisis comparativo de estas medidas en dos escenarios diferenciados: disoluciones y constituciones empresariales. Una examinación exhaustiva de las “**Tablas 21 y 22**” revela una disminución progresiva del RMSE y del MAE al pasar de los modelos más simples a los más complejos, tanto en disoluciones como en constituciones. Este descenso indica una mejora en la capacidad predictiva de los modelos a medida que se incorporan más variables, sugiriendo que la integración de una gama más amplia de factores económicos y demográficos contribuye a una mayor precisión.

Sin embargo, el MAPE presenta valores infinitos en algunos casos, reflejando la presencia de ceros en los datos reales que provocan divisiones indefinidas. A pesar de esto, donde los valores del MAPE son finitos, muestran una mejora sustancial en el Modelo 3, especialmente en el ámbito de las constituciones, subrayando una vez más el refino predictivo alcanzado en las iteraciones más avanzadas del modelo.

La relevancia del número de condición en este análisis no debe subestimarse. Valores extremadamente altos en los modelos más simples sugieren una inestabilidad numérica que está asociada a la multicolinealidad, un fenómeno que distorsiona las estimaciones de los coeficientes. Sin embargo, al observar una reducción significativa en este número en los modelos más complejos, se evidencia una mejora en la condición numérica de los modelos, lo que implica una atenuación de la multicolinealidad.

Las pruebas de Wilcoxon y Friedman brindan más contexto sobre la homogeneidad de los residuos y la consistencia entre los modelos. Aunque los valores de p de Wilcoxon para las comparaciones individuales entre residuos varían, el hecho de que algunos pares muestren diferencias estadísticamente significativas ($p\text{-valor} < 0.05$) indica que las mejoras en la adecuación del modelo no son uniformes y que la inclusión de variables adicionales tiene un impacto diferencial en la calidad del ajuste. Específicamente, la comparación entre los residuos del Modelo 2 y el Modelo 3 para las disoluciones, y entre los residuos del Modelo 1 y Modelo 2 para las constituciones, muestran p -valores que indican mejoras significativas.

Por su parte, la prueba de Friedman, con un p -valor que supera el umbral comúnmente aceptado para la significancia estadística (0.05), sugiere que no existen diferencias significativas en los rangos medios de los R-cuadrados entre los tres modelos. Esto podría indicar que las mejoras observadas en los R-cuadrados de los modelos no se deben a la inclusión de más variables per se, sino posiblemente a la elección de variables más pertinentes o al ajuste de la especificación del modelo.

En resumen, mientras las medidas de error revelan una capacidad predictiva mejorada con modelos más complejos, los tests estadísticos invitan a un análisis más matizado de la mejora en la calidad del ajuste. Estos resultados son fundamentales para consolidar la fiabilidad de los modelos en la descripción y predicción de las dinámicas empresariales, ofreciendo así una guía valiosa para la formulación de estrategias y políticas basadas en evidencia.

14. Conclusiones y Recomendaciones

14.1 Conclusiones

El presente Trabajo de Fin de Grado se ha enfocado en elucidar la compleja red de interacciones entre distintos factores económicos y las fluctuaciones en las tasas de disolución y constitución empresarial en el contexto español, a través de una meticulosa aplicación de metodologías de regresión lineal simple y múltiple. Este análisis exhaustivo pretendía no solo dilucidar las dinámicas económicas que inciden directamente sobre las empresas, sino también ofrecer un panorama más amplificado de las variables que inciden en la estabilidad corporativa en el dinámico entorno económico actual.

Los resultados obtenidos de la aplicación de modelos de regresión han arrojado luz sobre la capacidad relativamente restringida de los indicadores macroeconómicos seleccionados para explicar las variaciones en las disoluciones empresariales. Este fenómeno señala hacia la existencia de una trama más densa y compleja de factores que intervienen en la disolución y constitución de empresas, extendiéndose más allá de las métricas económicas convencionales. A partir de las conclusiones extraídas del análisis se destacan varios puntos clave:

Influencia del Clima Empresarial: Los análisis han mostrado que el clima empresarial, reflejado por el indicador ICEA, tiene un papel preponderante en las disoluciones y constituciones empresariales. Esta relación significativa subraya la trascendencia de la confianza y las expectativas económicas en la toma de decisiones empresariales, sugiriendo que la disposición empresarial hacia el riesgo y la innovación puede ser un predictor más afinado de las disoluciones que los indicadores económicos tradicionales.

Complejidad de las Disoluciones Empresariales: La limitada varianza explicada por los modelos enfatiza que las disoluciones son el corolario de un conjunto de factores heterogéneos, que incluyen aspectos macroeconómicos, particularidades sectoriales, legislación, avances tecnológicos y dinámicas competitivas. La realidad empresarial en sectores punteros, por ejemplo, podría ser un caldo de cultivo para la disolución y renovación corporativa, como manifestación de un mercado que premia la innovación y la eficiencia.

Impacto de Factores Temporales y Externos: La influencia de factores coyunturales y externos, como crisis económicas globales o emergencias sanitarias, a priori no detectada en los modelos, es una variante a tener en cuenta para futuras investigaciones. Tales eventos pueden provocar un cambio abrupto en la viabilidad de los negocios y, por consiguiente, acelerar procesos de disolución o incluso de constitución, ilustrando la vulnerabilidad u oportunidad del tejido empresarial a choques exógenos.

Diversidad Regional y Sectorial: La investigación también ha enfatizado la importancia de atender a la diversidad tanto regional como sectorial en España. Las diferencias en las tasas de disolución y constitución entre las comunidades autónomas y los distintos sectores económicos abogan por enfoques diferenciados y políticas adaptadas que atiendan a las especificidades locales y sectoriales.

Las conclusiones de este estudio, por tanto, subrayan la complejidad inherente a las disoluciones y constituciones empresariales en España. Se desprende que, aunque los indicadores macroeconómicos ofrecen una perspectiva de las tendencias generales, es crucial reconocer y entender el impacto combinado de factores económicos, sociales, tecnológicos, demográficos y políticos. El reconocimiento y comprensión de esta complejidad es vital para la formulación de estrategias efectivas que fomenten la resiliencia y el crecimiento empresarial en un entorno económico que está en constante evolución. Este enfoque integral y multidimensional es indispensable para abordar con eficacia los desafíos y aprovechar las oportunidades que se presentan en el panorama empresarial español.

14.2 Recomendaciones

En el terreno de la exploración sobre las dinámicas empresariales en España, tanto en el proceso de constitución como en el de disolución de empresas, se ha alcanzado un entendimiento relevante que señala la importancia de factores económicos diversos. No obstante, la intrincada y variada naturaleza de estos procesos requiere un acercamiento más sofisticado y exhaustivo para captar con fidelidad las complejas dinámicas que los rigen. Por ello, se propone como primordial para futuras investigaciones la adopción de modelos analíticos avanzados que puedan abarcar estas complejidades.

El uso de modelos de regresión no lineales, técnicas avanzadas de análisis de series temporales y algoritmos de aprendizaje automático como los árboles de decisión y redes neuronales, puede brindar una comprensión más profunda de las interacciones complejas y no lineales entre los múltiples factores predictores. Este tipo de herramientas analíticas avanzadas ofrece un potencial significativo para mejorar el entendimiento y la capacidad predictiva respecto a la formación y disolución de las empresas.

Es también fundamental la inclusión de variables adicionales que hasta ahora han sido menos consideradas. La confianza empresarial, los indicadores de emprendimiento e innovación, y los factores sectoriales específicos, podrían otorgar una visión más completa de las fuerzas que inciden en la vida de las empresas. Esta inclusión permitiría evaluar con mayor precisión el impacto de los distintos aspectos del entorno económico y empresarial en la estabilidad y sostenibilidad de las empresas.

Analizar segmentadamente por sectores y tamaños de empresa podría descubrir variaciones significativas en las tasas de constitución y disolución y arrojar luz sobre vulnerabilidades y resiliencias específicas. Tal aproximación permitiría reconocer patrones distintivos y desarrollar estrategias más eficaces y personalizadas para sectores o tipos de empresa concretos.

Por otro lado, los estudios longitudinales y comparativos constituyen un recurso valioso para comprender las tendencias a lo largo del tiempo y las diferencias en el contexto de las dinámicas empresariales. Este enfoque facilitaría el análisis de cómo los cambios en el panorama económico y político influyen en las empresas a lo largo del tiempo y en distintos entornos.

La combinación de enfoques cuantitativos con cualitativos, a través del análisis de casos o entrevistas con empresarios y especialistas, enriquecería la comprensión de las causas y efectos de la constitución y disolución empresarial. Esta metodología mixta fortalecería los hallazgos cuantitativos, proporcionando una comprensión más rica y matizada de los factores que influyen en las decisiones empresariales.

En última instancia, el análisis del impacto de las políticas públicas y los programas de apoyo empresarial en las tasas de disolución podría ofrecer orientación práctica para diseñar intervenciones más efectivas. Este enfoque político permitiría identificar estrategias fundamentadas en la evidencia para promover un ecosistema empresarial resiliente y propicio al crecimiento sostenible.

Las recomendaciones aquí esbozadas tienen la intención de dirigir la investigación futura hacia un examen más completo y detallado de la formación y disolución de empresas en España. Al integrar estos métodos y enfoques, los investigadores pueden construir un entendimiento más acabado de este fenómeno complejo, y contribuir al desarrollo de políticas y estrategias que respalden la estabilidad y el avance empresarial en el país.

15. Resumen Narrativo

El presente estudio se ha centrado en el análisis de la creación y disolución empresarial en España, una cuestión de gran interés por su impacto económico y social. A través de métodos analíticos, se ha indagado en los patrones y correlaciones que subyacen a estas dinámicas empresariales, con un enfoque que integra factores económicos y del entorno de mercado.

La fase inicial de la investigación abarcó una minuciosa selección y preparación de los datos, que posteriormente dieron lugar al desarrollo y aplicación de modelos analíticos robustos. Los modelos de regresión implementados tenían el objetivo de esclarecer cómo interactúan variables económicas fundamentales, como el PIB y el IPC, con las tasas de creación y disolución de empresas y el capital movilizado en estas transacciones. La evaluación meticulosa de estos modelos reveló que las decisiones empresariales de cierre o continuación están influidas por una red compleja de factores.

Se constató que el clima empresarial, junto con otros indicadores macroeconómicos, ejerce una influencia significativa, pero la relación es intrincada y diversa. Tanto las constituciones como las disoluciones empresariales en el contexto español están marcadas por la interacción de distintas variables, algunas reflejan la situación macroeconómica general del país, y otras son más representativas del ambiente empresarial específico.

Este trabajo amplía la comprensión académica de los ciclos de vida empresariales en España y provee modelos analíticos aplicables en investigaciones futuras y en la gestión empresarial. Las conclusiones y recomendaciones extraídas apuntan a medidas para promover un clima económico estable y resiliente, resaltando la necesidad de un enfoque holístico que integre factores económicos generales y particulares del entorno empresarial.

Finalmente, este estudio recalca la importancia de adoptar perspectivas plurales para una comprensión cabal de las disoluciones y creaciones de empresas. Para futuras investigaciones, se aconseja la utilización de modelos más complejos y la inclusión de nuevas variables que permitan captar con mayor fidelidad la complejidad inherente a estos procesos en el escenario económico y social de España.

TERCERA PARTE: ANÁLISIS DE NEGOCIO

16. Introducción al Análisis de Negocio

La tercera parte de este Trabajo de Fin de Grado, denominada "Análisis de Negocio", constituye una fase esencial donde se amalgaman los conocimientos y datos recogidos a lo largo de la investigación. Este segmento se presenta meticulosamente para cubrir desde la consolidación de los objetivos y hallazgos hasta la formulación de recomendaciones estratégicas dirigidas a influir positivamente en el ecosistema empresarial, considerando el contexto macroeconómico español.

Se comienza estableciendo el escenario para un análisis exhaustivo, destacando la importancia de integrar técnicas avanzadas de análisis de datos con una perspectiva empresarial aguda, dirigida a comprender las dinámicas del mercado y las respuestas empresariales ante variaciones macroeconómicas. Se reflexiona sobre cómo los objetivos iniciales se han transformado a lo largo de la investigación, adaptándose a los descubrimientos y a las realidades del análisis de datos. Además, se discuten los resultados específicos obtenidos de los modelos de regresión implementados, evaluando su eficacia y limitaciones.

Un componente crucial de esta parte es el uso de opiniones de expertos del sector para enriquecer el análisis. Hasta el momento, se ha realizado una entrevista detallada, cuyas interacciones y preguntas clave han orientado significativamente la discusión de este segmento. Se ha consolidado un conjunto de respuestas y opiniones recogidas durante esta entrevista inicial. A partir de estas interacciones, se extraen y discuten insights clave que apoyan o cuestionan los hallazgos del estudio. Es importante destacar que las demás entrevistas están programadas para realizarse después de la entrega de esta parte del trabajo. Los resultados y análisis adicionales derivados de estas futuras entrevistas se integrarán en la versión final del documento, proporcionando una visión más comprensiva y enriquecida por una diversidad más amplia de perspectivas expertas.

Enfocado hacia la aplicación práctica de los conocimientos adquiridos, se ofrecen directrices tanto para empresarios y emprendedores como para formuladores de políticas, proporcionando recomendaciones basadas en los análisis realizados. Se desarrolla también una propuesta de negocio detallada que emerge del estudio, describiendo su relevancia y aplicabilidad en el contexto actual.

Finalmente, se concluye con un resumen narrativo que encapsula los principales aprendizajes, desafíos y oportunidades identificadas a lo largo del análisis de negocio, ofreciendo una visión integral de la contribución del estudio al campo del análisis empresarial y económico. Este enfoque estructurado no solo refleja la rigurosidad académica del TFG, sino que también enfatiza la relevancia práctica de la investigación, destacando su potencial para influir en las decisiones empresariales y políticas en un contexto marcado por la incertidumbre económica.

17. Conclusiones

17.1 Evaluación y Respuestas los Objetivos Planteados

¿Cómo se ha desarrollado el ecosistema empresarial español en función del PIB e IPC, y qué implicaciones tienen estos indicadores en la estructura empresarial del país?

Primeramente, el análisis se centró en discernir el impacto del PIB y del IPC sobre el número de disoluciones empresariales, utilizando un modelo de regresión lineal múltiple. Los resultados del modelo revelan un coeficiente de determinación R^2 de solo 0.001 y un R^2 ajustado de -0.009, lo cual indica una capacidad muy baja para explicar la variabilidad en las disoluciones empresariales basadas en estas variables macroeconómicas. Estos valores sugieren que, en el contexto del modelo configurado, PIB e IPC tienen un efecto casi nulo sobre el número de disoluciones.

La falta de significancia del modelo se refleja también en un F-estadístico de 0.08673 con un p-valor de 0.917, lo que claramente indica que, según este modelo, no existe una relación estadísticamente significativa entre las disoluciones empresariales y las variables de PIB e IPC consideradas.

El coeficiente para el IPC es de -6.4564, con un p-valor de 0.890, y para el PIB es prácticamente cero (6.611×10^{-10}) con un p-valor de 0.744. Estos coeficientes indican que no hay evidencia suficiente para afirmar que cambios en el IPC o en el PIB tengan un efecto directo y significativo sobre el número de disoluciones empresariales.

Después, para las constituciones empresariales, el análisis revela resultados similares con un R^2 de 0.003 y un R^2 ajustado de -0.007. Estos valores indican igualmente una baja capacidad del modelo para explicar la variabilidad en las constituciones empresariales. Los coeficientes para IPC y PIB en este modelo son igualmente no significativos, reforzando la idea de que las variaciones en estos indicadores económicos no están directamente correlacionadas con la tasa de nuevas empresas establecidas.

¿Han afectado las variables macroeconómicas a la tasa de creación y disolución de empresas en España?

El análisis detallado a través de modelos de regresión lineal múltiple ajustados para las tasas de constitución y disolución de empresas en España, después de un riguroso proceso de eliminación de multicolinealidad mediante el análisis VIF, revela que las variables macroeconómicas tienen una influencia limitada pero identificable en estos fenómenos empresariales.

El modelo ajustado para las constituciones empresariales muestra que las variables macroeconómicas explican aproximadamente el 6.0% de la variabilidad en las tasas de constitución de empresas. Aunque este porcentaje es modesto, subraya la complejidad y la multitud de factores que influyen en la creación de nuevas empresas. El coeficiente ajustado R-cuadrado de 0.032, aunque bajo, sugiere que hay elementos del entorno macroeconómico que afectan la iniciativa empresarial, aunque su impacto global es limitado.

En relación con las disoluciones empresariales, las variables macroeconómicas después de ajustar la multicolinealidad explican un 4.6% de la variabilidad. Este resultado indica un impacto similarmente limitado de estas variables sobre la tasa de disolución de empresas. El ajuste del modelo refleja una realidad donde otros factores no capturados por el modelo actual pueden estar jugando roles más significativos en la disolución empresarial.

Los resultados de ambos modelos subrayan que, aunque las variables macroeconómicas como el PIB, el IPC, y otros factores como el ICEA tienen efectos estadísticamente significativos en ciertos casos, su capacidad general para explicar las dinámicas de constitución y disolución de empresas es relativamente baja. Esto sugiere que las políticas económicas basadas únicamente en estos indicadores podrían no ser suficientemente efectivas sin considerar otros elementos contextuales y sectoriales que también afectan al ecosistema empresarial.

La investigación muestra que las variables macroeconómicas tienen una influencia observable pero limitada en las tasas de constitución y disolución de empresas en España. Este hallazgo es crucial para los formuladores de políticas y los estrategas empresariales, quienes deben considerar un enfoque más holístico que abarque variables económicas, contextuales y sectoriales para formular estrategias efectivas que fomenten un entorno empresarial robusto y resiliente.

¿Qué variables específicas influyen en la constitución y disolución de empresas?

En el estudio de las variables que influyen en la constitución y disolución de empresas en España, el análisis detallado de las regresiones lineales múltiples destaca la relevancia del Indicador de Clima Empresarial (ICEA). Esta variable ha emergido como significativa en ambos modelos, lo que subraya su papel crucial en el entorno empresarial español.

En el modelo de regresión para las constituciones empresariales, el ICEA mostró un coeficiente positivo (coef = 38.323, $p = 0.001$), lo que indica que un aumento en el indicador de clima empresarial está asociado con un incremento en el número de nuevas empresas constituidas. Esto sugiere que una percepción positiva del entorno económico y empresarial puede fomentar el emprendimiento y la creación de nuevas empresas. La confianza

económica, reflejada en el ICEA, puede incentivar a los emprendedores a lanzar nuevas iniciativas, presumiblemente debido a una expectativa de condiciones de mercado favorables y soporte gubernamental o institucional efectivo.

Similarmente, en el modelo de disoluciones empresariales, el ICEA también resultó ser estadísticamente significativo (coef = 8.3006, $p = 0.003$), pero con un coeficiente menor comparado con el modelo de constituciones. Esto indica que un mejor clima empresarial reduce la probabilidad de que las empresas se disuelvan. La interpretación de este resultado podría ser que en un ambiente empresarial percibido como saludable y estable, las empresas existentes tienen mayores probabilidades de sobrevivir y prosperar, enfrentando menos situaciones que podrían llevar a su disolución.

La influencia del ICEA en ambos procesos, constitución y disolución de empresas, puede entenderse a través de su relación con la confianza económica general. Un clima empresarial robusto no solo atrae nuevos proyectos de emprendimiento, sino que también proporciona un terreno más seguro para que las empresas existentes se mantengan y expandan. Este indicador refleja una amalgama de factores económicos, incluyendo políticas gubernamentales, condiciones de mercado, apoyo financiero y percepciones de estabilidad y crecimiento futuro.

El análisis resalta que el ICEA es una variable macroeconómica crítica que influye significativamente en la dinámica empresarial de España. Aunque otras variables como el IPC, el PIB y variables relacionadas con el turismo y la política fiscal fueron consideradas, no mostraron un impacto estadístico significativo en los modelos finales, posiblemente debido a la complejidad de interacciones en el sistema económico y la necesidad de un análisis más granular para desentrañar su influencia específica.

¿Cómo han respondido las empresas españolas ante las incertidumbres generadas por crisis macroeconómicas y cuáles son los factores clave para su resiliencia?

Las empresas españolas han respondido a las crisis macroeconómicas con diversas estrategias que reflejan su capacidad de adaptación y resiliencia. Un factor clave identificado es la flexibilidad operativa, que permite a las empresas ajustar rápidamente sus operaciones y recursos frente a cambios económicos repentinos. Además, la diversificación de mercados y fuentes de ingresos ha sido crucial, permitiendo a las empresas no depender exclusivamente de un mercado o sector económico particular. Durante la entrevista con Alberto de Santos, se mencionó la importancia de la inteligencia de negocio y el análisis predictivo. Empresas que invierten en tecnologías de análisis avanzado y aprendizaje automático para entender mejor el entorno económico y prever cambios, tienden a manejar mejor la incertidumbre y a tomar

decisiones más informadas. Estas herramientas permiten anticipar tendencias y adaptar estrategias en tiempo real, lo que es fundamental para mantener la competitividad y la sostenibilidad en tiempos de incertidumbre económica.

17.2 Discusión sobre Respuestas: Métodos Utilizados para Responder

Para abordar las preguntas de investigación planteadas en este Trabajo de Fin de Grado, se ha empleado un enfoque multifacético que combina análisis cuantitativos detallados, entrevistas con profesionales del sector y revisión de literatura académica relevante. Esta metodología integral ha permitido una exploración profunda y variada de las temáticas involucradas, asegurando así una respuesta robusta y bien fundamentada a las cuestiones planteadas.

Inicialmente, el análisis de datos se centró en el uso de modelos de regresión lineal múltiple para evaluar el impacto de diversas variables macroeconómicas sobre las tasas de constitución y disolución de empresas en España. Este enfoque cuantitativo no solo ha permitido identificar las variables con influencia significativa, como el Indicador de Clima Empresarial (ICEA), sino también determinar la limitada pero relevante capacidad explicativa de estos modelos en cuanto a la variabilidad observada en el ecosistema empresarial. El proceso meticuloso de revisión y ajuste de modelos mediante el análisis de VIF (Factor de Inflación de la Varianza) ha mejorado la precisión de las estimaciones, subrayando la necesidad de considerar factores adicionales que podrían estar influyendo en el panorama empresarial.

Por otro lado, las entrevistas con profesionales del sector han proporcionado una perspectiva práctica y actualizada que enriquece el análisis. Estas conversaciones han permitido validar supuestos del modelo y añadir un contexto más amplio sobre cómo las empresas realmente perciben y reaccionan ante las dinámicas macroeconómicas. Por ejemplo, la discusión con Alberto de Santos ha revelado la importancia de considerar el clima económico y la confianza empresarial en la toma de decisiones estratégicas, elementos que se reflejan en el impacto directo del ICEA en las disoluciones empresariales.

Además, la revisión de documentos académicos y literatura relevante ha permitido complementar y contrastar los hallazgos empíricos con teorías y estudios previos. Esta revisión ha ayudado a situar los resultados del TFG dentro de un marco teórico más amplio, asegurando que las conclusiones no solo se sustentan en datos empíricos, sino que también se alinean con la evidencia teórica existente sobre la influencia de las variables macroeconómicas en el ámbito empresarial.

En resumen, la combinación de análisis de datos riguroso, intercambio de perspectivas con expertos del sector y un sólido respaldo teórico ha permitido responder a las preguntas de investigación de manera exhaustiva y matizada. Este enfoque integrador no solo mejora la comprensión de las dinámicas empresariales en el contexto español, sino que también orienta hacia futuras líneas de investigación y acción práctica en el ámbito de la economía y la gestión empresarial.

18. Opiniones Profesionales

18.1 Información de las Entrevistas

Alberto de Santos (Perfil y Experiencia Profesional)

Alberto de Santos se distingue como un profesional con más de quince años de experiencia en el ámbito de la analítica avanzada, la inteligencia artificial, y el aprendizaje profundo, destinados a la toma de decisiones empresariales basadas en datos. Su profundo conocimiento teórico en procedimientos estadísticos, que abarcan tanto el aprendizaje supervisado como el no supervisado, junto con redes neuronales, refuerza su perfil como un experto en la implementación de soluciones analíticas complejas. A lo largo de su carrera, ha liderado equipos de científicos de datos e ingenieros en colaboración con grupos internos y externos, promoviendo metodologías que maximizan el valor de Big Data y la IA en negocios.

La carrera de Alberto comenzó en Hewlett-Packard como Lead Data Scientist, posición que mantuvo mientras co-fundaba Koulu, una start-up que ofrece un tutor de matemáticas potenciado por IA para estudiantes, destacando su capacidad para combinar emprendimiento con su experiencia técnica. Posteriormente, ocupó roles significativos en Vodafone mientras que inició su carrera docente como profesor de inteligencia artificial en la Universidad Internacional de la Rioja. Actualmente desempeña el puesto de Global Senior Data Scientist en Levi Strauss & Co., al mismo tiempo que es docente de inteligencia artificial en la Universidad Francisco de Vitoria.

José María Martín-Mateos Espinar (Perfil y Experiencia Profesional)

José María Martín-Mateos es un reconocido líder y emprendedor en el ámbito de las tecnologías de la información y consultoría de gestión, con más de dos décadas de experiencia innovando en el sector tecnológico. Fundador y CEO de Controlnet, parte del Grupo Lyvia, ha dirigido la empresa desde 2009, convirtiéndola en líder en Europa y Latinoamérica en el desarrollo e implementación de Sistemas de Gestión e Inteligencia de Negocio. Bajo su liderazgo, Controlnet ha promovido un crecimiento empresarial sostenible, destacando en la creación de soluciones que maximizan las posibilidades de la economía digital.

Anteriormente, Martín-Mateos fue CEO de eresazul.com, donde estuvo al frente durante siete años, periodo durante el cual demostró su habilidad para dirigir estrategias de negocio eficaces en el ámbito digital. También ocupó el puesto de COO en Grupo Martín-Mateos, consolidando su experiencia en operaciones y gestión empresarial.

El enfoque de Martín-Mateos se centra en la disciplina como pilar del éxito, fomentando un entorno innovador que responde eficazmente a los desafíos del mercado actual. Controlnet, integrando el grupo sueco LYVIA, ofrece una amplia gama de productos y servicios tecnológicos a través de una cartera diversificada de marcas líderes en el mercado. Estas operaciones abarcan desde software y servicios de TI hasta marketing digital y consultoría creativa SaaS, reflejando su visión para adaptar y liderar en un entorno empresarial dinámico.

Relación con el Trabajo de Fin de Grado

La breve pero intensa experiencia de Alberto como emprendedor y su enfoque en la analítica de negocios son particularmente relevantes para el Trabajo de Fin de Grado (TFG) centrado en cómo las variables macroeconómicas influyen en el desarrollo empresarial. Su experiencia proporciona una perspectiva valiosa sobre la importancia de los factores macroeconómicos en las decisiones empresariales, tanto desde el punto de vista de un emprendedor como desde el de un ejecutivo en grandes corporaciones.

La trayectoria de José María Martín-Mateos como emprendedor y ejecutivo en el sector tecnológico ofrece una perspectiva enriquecedora para el Trabajo de Fin de Grado (TFG) que explora cómo las variables macroeconómicas influyen en las disoluciones y constituciones empresariales en España. Su liderazgo en Controlnet y su experiencia en la expansión de operaciones en mercados internacionales resaltan la importancia de adaptarse a las condiciones macroeconómicas y regulatorias para el éxito empresarial. Esta visión es crucial para entender la dinámica entre el entorno económico y las decisiones estratégicas en el ámbito empresarial, proporcionando un contexto valioso para el análisis de las variables macroeconómicas abordadas en el TFG.

Detalles de las Entrevistas

La entrevista con Alberto de Santos se realizó el 16 de abril de 2024, a través de una videollamada de Teams, permitiendo una discusión en profundidad sobre varios temas clave relacionados con el TFG. Durante la entrevista, se exploraron diversas cuestiones que buscaban profundizar en la comprensión del impacto de las variables macroeconómicas en la creación y desarrollo de empresas. Las preguntas formuladas abarcaron desde la consideración de la situación macroeconómica en la etapa de desarrollo de su start-up hasta la percepción de la importancia de estas variables en sus roles en corporaciones multinacionales y en el ámbito educativo.

Por otro lado, la entrevista con José María Martín-Mateos se realizó el 23 de abril de 2024, mediante Google Meets, enfocándose en aspectos críticos sobre el impacto de las variables macroeconómicas en el ámbito empresarial. Esta interacción permitió profundizar

en cómo las condiciones macroeconómicas específicas influyen en las estrategias de desarrollo y sostenibilidad empresarial. La discusión abarcó desde la evaluación de indicadores económicos hasta cómo estos factores afectan a la toma de decisiones en el contexto de empresas tecnológicas y digitales.

18.2 Resumen de la Respuestas Obtenidas

Consideración de la Situación Macroeconómica de Alberto y su equipo en el Desarrollo de Koulu (Startup de educación)

Alberto explicó que durante el desarrollo inicial de Koulu, se tuvo en cuenta la situación macroeconómica de la época, específicamente en torno a la crisis inmobiliaria y la incertidumbre económica que persistía en esos años. El equipo de Koulu realizó un estudio para entender el gasto promedio de las familias en educación adicional, un factor crucial dado que Koulu estaba destinado a ser financiado por los padres de familia. Este análisis se apoyó en investigaciones previas de firmas multinacionales para establecer un marco de referencia sobre los gastos educativos en relación con el ingreso medio en España.

Interés de Alberto en la Situación Macroeconómica durante las Estancias en Hewlett-Packard, Vodafone y Levi

En esta pregunta Alberto destacó la importancia del contexto macroeconómico tanto en su interés personal como profesional. En su experiencia, el desempeño macroeconómico influye y directamente en cosas como la evaluación del desempeño laboral a través del rendimiento económico que tiene la empresa donde se trabaja. Sin embargo, cree que actualmente y debido al contexto tan inestable que existe, puede llegar a afectar significativamente la estabilidad de los empleos dentro de las empresas.

Alberto también contaba, como en Levi, específicamente, llegó a realizar un estudio para comprender cómo la inflación impactaba en los productos sustitutivos ofrecidos por la empresa y el patrón de compra de los clientes. Este estudio se vio interrumpido por el inicio del conflicto entre Rusia y Ucrania, que desestabilizó las tendencias económicas existentes. Así mismo, enfatizó que el interés en la macroeconomía puede que haya evolucionado de una mera curiosidad a una preocupación esencial para la supervivencia profesional, especialmente en el contexto de las recientes inestabilidades económicas globales.

Consideración de la Situación Macroeconómica de José María y su equipo en el desarrollo y lanzamiento de Controlnet (Empresa de Desarrollo de software personalizado)

En cuanto a si se tuvo en cuenta la situación macroeconómica al lanzar el producto, José María Martín-Mateos señaló que, durante la época de lanzamiento, el sector de las telecomunicaciones y soluciones empresariales digitales estaba en auge. Aunque la situación de capital de las empresas era relevante debido a la crisis inmobiliaria, la situación macroeconómica general no fue un factor decisivo para el lanzamiento al mercado. Sin embargo, sí se observó y estudió el crecimiento en la venta de productos tecnológicos en otras empresas, que mostraba un aumento exponencial.

Importancia de las situaciones o variables macroeconómicas del momento para la toma de decisiones en Controlnet.

Respecto a la importancia de las variables macroeconómicas en la toma de decisiones para entrar en algunos países, destacó que ciertos indicadores, como el índice Big Mac o las políticas jurídicas de un país, son considerados para evaluar cómo podrían afectar los proyectos a realizar. Además, Martín-Mateos tiende a analizar indicadores económicos de líderes en los sectores de telecomunicaciones y digitalización, como Estados Unidos o Alemania.

18.3 Insights obtenidos de las preguntas

En este Trabajo de Fin de Grado, se exploraron las dinámicas y estrategias empresariales bajo la influencia de variables macroeconómicas a través de entrevistas con expertos del sector, análisis de datos y revisión de literatura relevante. Las entrevistas con José María Martín-Mateos y Alberto de Santos proporcionaron insights valiosos sobre la importancia de las variables macroeconómicas en el desarrollo empresarial.

José María Martín-Mateos, en el lanzamiento de Controlnet, indicó que aunque la situación macroeconómica no fue determinante, el auge del sector tecnológico influyó positivamente en sus decisiones de mercado. Esta perspectiva se complementó con su énfasis en la relevancia de variables específicas como el índice Big Mac y las políticas jurídicas para evaluar impactos en proyectos internacionales, reflejando un enfoque pragmático hacia la planificación estratégica. Además, destacó la significativa ayuda de las administraciones públicas durante las crisis, lo que resalta la importancia de colaboraciones entre el sector público y privado para la resiliencia empresarial.

Por otro lado, Alberto de Santos resaltó cómo su experiencia en corporaciones multinacionales y proyectos como Koulu ilustra la necesidad de integrar el análisis macroeconómico en la toma de decisiones empresariales. Subrayó la importancia de anticipar cambios económicos y adaptar estrategias empresariales a estos, lo cual es esencial en un entorno global incierto. También sugirió la exploración de datos a un nivel más granular y la inclusión de variables no convencionales para obtener una comprensión más profunda de las dinámicas empresariales.

Ambos expertos coinciden en la importancia de un enfoque multidimensional que considere tanto las condiciones macroeconómicas como las especificidades del sector para el desarrollo empresarial. Este enfoque integrado, que combina insights de expertos con análisis cuantitativo, ofrece una base sólida para comprender la interacción entre la economía y las dinámicas empresariales, proporcionando una guía clara para futuras investigaciones y prácticas empresariales.

19. Recomendaciones y Propuestas

19.1 Recomendaciones

En el apartado de recomendaciones de este trabajo, se abordan estrategias y consejos dirigidos tanto a empresarios y emprendedores como a formuladores de políticas, con el objetivo de mejorar sus capacidades de adaptación y toma de decisiones en un entorno económico cambiante. Este segmento del estudio se estructura en dos subsecciones principales: la primera enfocada en proporcionar orientación práctica a empresarios y emprendedores, resaltando la importancia de integrar un análisis macroeconómico sólido en sus estrategias empresariales; y la segunda destinada a los formuladores de políticas, sugiriendo maneras en las que pueden apoyar y mejorar el ecosistema empresarial a través de políticas informadas y efectivas. A través de estas recomendaciones, se busca no solo aportar al debate académico, sino también ofrecer herramientas aplicables que puedan tener un impacto real y positivo en el entorno empresarial.

19.1.1 Empresarios, Ejecutivos y Emprendedores

A lo largo de este estudio, el análisis de las variables macroeconómicas y su influencia en las dinámicas de constitución y disolución de empresas en España ha revelado aspectos clave que pueden ser cruciales para empresarios y emprendedores. A través de técnicas como la regresión lineal múltiple y el manejo de la multicolinealidad mediante el análisis de VIF, se ha identificado que, aunque variables como el PIB y la Deuda Pública fueron excluidas del modelo final por su alta correlación con otras variables, indicadores como el Indicador de Clima Empresarial (ICEA) tienen un impacto directo y significativo en el número de disoluciones empresariales. Este hallazgo destaca la importancia del clima empresarial y la confianza económica, subrayando cómo estos factores pueden influir directamente en la estabilidad y las decisiones estratégicas dentro del ecosistema empresarial.

A través de las entrevistas realizadas a profesionales del sector, se ha recopilado una visión que complementa los hallazgos técnicos de la investigación. Estas conversaciones han puesto de manifiesto la relevancia de tener una comprensión profunda del entorno económico cuando se toman decisiones estratégicas. En el contexto del desarrollo de productos y servicios, por ejemplo, se observó que un entendimiento detallado de la situación macroeconómica del sector en el que se opera puede ser decisivo. Esta perspectiva es vital ya que orienta a los emprendedores a considerar no solo el estado actual del mercado, sino también a anticipar posibles escenarios económicos que podrían afectar sus modelos de negocio.

La investigación sugiere que para los empresarios y emprendedores, especialmente aquellos en las fases iniciales de desarrollo de un proyecto o negocio, es esencial integrar un análisis económico riguroso en la planificación y estrategia empresarial. Esto no solo incluye la observación de indicadores económicos estándar como el PIB o el IPC, sino también indicadores más específicos del clima empresarial y la confianza económica como el ICEA o incluso variables directamente relacionadas con sus sectores. Estos factores pueden proporcionar señales tempranas sobre las tendencias de mercado y potenciales desafíos o oportunidades que podrían surgir.

Además, la capacidad para adaptar y afinar continuamente los modelos predictivos y analíticos en respuesta a nuevos datos y tendencias emergentes es una competencia que puede diferenciar a las empresas exitosas de aquellas que no logran ajustarse a las realidades del mercado. Esto implica no solo la recolección y análisis de datos, sino también una constante revisión y mejora de los modelos empleados, asegurando así que las estrategias empresariales estén bien informadas y sean robustas ante cambios económicos y sociales.

En resumen, este trabajo destaca la importancia de una integración efectiva de análisis macroeconómico en la toma de decisiones empresariales, enfatizando que la comprensión y adaptación a las condiciones económicas no solo es una práctica deseable, sino una necesidad imperativa en el dinámico entorno de negocios de hoy.

19.1.2 Formuladores de Políticas

En el ámbito de las políticas públicas y la administración, el papel de las autoridades en la configuración de un entorno empresarial sólido y adaptable es fundamental. A partir de los hallazgos de este trabajo, donde se ha evaluado el impacto de diversas variables macroeconómicas en la dinámica empresarial, es evidente que indicadores como el Indicador de Clima Empresarial (ICEA) son esenciales para que los empresarios evalúen las situaciones económicas actuales y futuras. Por tanto, es crucial que las políticas públicas promuevan la comprensión y la precisión de este y otros indicadores económicos, para que sean herramientas útiles en la toma de decisiones empresariales.

Para mejorar el sistema empresarial desde la perspectiva de las políticas públicas, se recomienda enfocarse en la promoción y educación continua sobre la importancia de la macroeconomía en las estrategias empresariales. Esto incluye asegurarse de que los indicadores como el ICEA sean accesibles, comprensibles y ampliamente difundidos entre la comunidad empresarial. Al mejorar la calidad y la accesibilidad de la información económica, se facilita a los empresarios la tarea de hacer proyecciones más precisas y de ajustar sus estrategias a las condiciones del mercado.

Además, dada la relevancia del conocimiento económico evidenciada en las entrevistas a profesionales del sector, se sugiere que las administraciones públicas fomenten la implementación de la macroeconomía como pilar fundamental en la educación. Este enfoque no solo debería limitarse a los programas de estudios empresariales o económicos, sino expandirse a diversas disciplinas, preparando a los estudiantes para comprender mejor cómo las fluctuaciones económicas pueden afectar todos los aspectos de la vida profesional y personal.

En resumen, las recomendaciones hacia los formuladores de políticas se centran en fortalecer la base educativa en economía para todos los estudiantes, promover la transparencia y el acceso a datos económicos fiables, y asegurar que los indicadores económicos como el ICEA sean comprendidos y utilizados efectivamente dentro del panorama empresarial. Estas medidas no solo beneficiarán a los empresarios actuales y futuros, sino que también contribuirán a la creación de un entorno empresarial más resiliente y adaptativo frente a los desafíos económicos globales.

19.2 Propuestas

En esta sección del Trabajo de Fin de Grado, se presenta una propuesta de proyecto de negocio que busca aplicar prácticamente los hallazgos de la investigación sobre la influencia de variables macroeconómicas en el ecosistema empresarial español. El proyecto se fundamenta en desarrollar una plataforma que utiliza análisis de datos avanzados para proporcionar herramientas estratégicas tanto a inversores como a empresas. La intención es mejorar la toma de decisiones mediante el acceso a análisis predictivos y descriptivos, permitiendo a las empresas mejorar su planificación y estrategia operativa en un entorno económico volátil. Este enfoque busca no solo aplicar conocimientos académicos, sino también ofrecer soluciones prácticas que tengan un impacto tangible en el mercado, facilitando así una mejor comprensión y gestión de los riesgos macroeconómicos.

19.2.1 Tipo de Propuesta

En el desarrollo de este Trabajo de Fin de Grado, no solo se buscaba aportar una investigación técnica y teórica sobre el comportamiento del ecosistema empresarial español bajo la influencia de variables macroeconómicas, sino también aplicar y poner en práctica las competencias y habilidades de negocio adquiridas durante la carrera en Business Analytics. Este enfoque dual refleja un compromiso no solo con el entendimiento académico, sino también con la aplicación práctica y concreta de este conocimiento en escenarios del mundo real.

Es con este objetivo en mente que se propone desarrollar un modelo de negocio innovador: un “marketplace” dirigido tanto a inversores como a empresas. Esta plataforma se fundamenta en el análisis robusto de datos, integrando los insights obtenidos a lo largo de la investigación del TFG. La propuesta se centra en proporcionar un espacio donde los inversores puedan acceder a análisis predictivos y descriptivos que les permitan tomar decisiones de inversión más informadas, mientras que las empresas pueden utilizar estos datos para mejorar su posicionamiento, estrategias de crecimiento y toma de decisiones operativas.

Este tipo de propuesta no solo capitaliza los hallazgos de la investigación, sino que también sirve como un puente entre la teoría y la práctica, ilustrando cómo las habilidades de análisis de datos y comprensión de las dinámicas macroeconómicas pueden tener aplicaciones directas y tangibles en el ámbito empresarial. De esta manera, se busca no solo aportar al conocimiento académico, sino también ofrecer soluciones innovadoras que puedan tener un impacto real en el mercado.

19.2.2 Problema y Solución de la Propuesta

La propuesta desarrollada en este Trabajo de Fin de Grado aborda un **problema** central en el ámbito de las inversiones y la gestión empresarial: **la dificultad para comprender y anticipar el impacto de las fluctuaciones macroeconómicas en las finanzas empresariales**. Los inversores, especialmente los “ángel investors” y otros actores del mercado privado, enfrentan el desafío constante de identificar oportunidades de inversión que equilibren adecuadamente el riesgo y el retorno en un contexto económico que es inherentemente volátil y complejo. Paralelamente, las empresas necesitan navegar por estas mismas aguas turbulentas buscando capital y asesoramiento estratégico que les permita adaptarse y florecer ante cambios económicos impredecibles. Frecuentemente, tanto inversores como empresas carecen de acceso a herramientas analíticas avanzadas y a la experiencia necesaria para realizar evaluaciones financieras y estratégicas profundas.

La **solución** propuesta consiste en el desarrollo de una **plataforma integral que utiliza tecnologías de análisis predictivo avanzado para vincular directamente las variables macroeconómicas con las cuentas de resultados de las empresas**. Esta herramienta innovadora proporcionará a los inversores análisis detallados y datos precisos, facilitando decisiones de inversión más informadas y estratégicamente fundadas. Adicionalmente, la plataforma servirá como un medio para que los inversores ofrezcan asesoramiento directo a las empresas en las que inviertan sobre cómo gestionar y planificar frente a posibles cambios en el escenario económico. Este enfoque no solo optimiza las

decisiones de inversión, sino que también fortalece a las empresas, haciéndolas más resilientes y estratégicamente ágiles.

Por lo tanto, esta propuesta no solamente facilita un método reactivo para la inversión y la planificación empresarial, sino que establece una dinámica proactiva, proporcionando una base sólida para el asesoramiento y la toma de decisiones que beneficia a todos los actores implicados en el ecosistema de inversión y gestión empresarial. En resumen, esta plataforma propone una sinergia entre tecnología y análisis financiero que promueve un crecimiento sostenible y ayuda a mitigar los riesgos derivados de las incertidumbres macroeconómicas, beneficiando tanto a inversores como a empresas en el mercado.

19.2.3 Explicación en Profundidad

Introducción al Modelo de Negocio

HELTUP constituye una avanzada plataforma de análisis predictivo, específicamente diseñada para correlacionar variables macroeconómicas con indicadores financieros empresariales. Esta herramienta proporciona a inversores y compañías capacidades analíticas para prever y adaptarse a las fluctuaciones económicas globales que podrían impactar sus operaciones y resultados financieros.

Componentes Clave del Modelo

Tecnología y Análisis de Datos: HELTUP integra inteligencia artificial y técnicas de aprendizaje automático para procesar y analizar grandes volúmenes de datos económicos y financieros. Este análisis facilita la predicción del impacto de cambios macroeconómicos en la actividad empresarial, apoyando así la optimización de decisiones de inversión y estrategias empresariales.

Interfaz y Experiencia del Usuario: La plataforma se destaca por una interfaz intuitiva que incluye dashboards personalizables y reportes exhaustivos, permitiendo a los usuarios acceder fácilmente a evaluaciones de riesgo, proyecciones de mercado y recibir recomendaciones ajustadas a sus perfiles y necesidades.

Integración y Seguridad de Datos: La seguridad es primordial en el proyecto, asegurando la integridad y confidencialidad de los datos a través de avanzados protocolos de seguridad, que incluyen cifrado, autenticación robusta y medidas proactivas para la protección continua de la información de los usuarios.

Funcionalidades y Servicios Ofrecidos

HELTUP ofrece una serie de servicios críticos como análisis de riesgo, proyecciones de mercado y recomendaciones basadas en modelos predictivos sofisticados. Estos servicios

son vitales tanto para inversores que desean maximizar rendimientos como para empresas que buscan fortalecer su planificación estratégica ante cambios económicos inesperados. Estas herramientas benefician a una amplia gama de usuarios, desde empresas en búsqueda de capital hasta inversores privados y analistas financieros, proporcionando una base más sólida para la toma de decisiones estratégicas.

Modelo de Monetización

La monetización de la plataforma se realiza a través de modelos de suscripción, ofreciendo diferentes niveles de acceso adaptados a las necesidades específicas de los usuarios. Además, la plataforma proporciona informes especializados y servicios de consultoría a un costo adicional, así como licencias para grandes corporaciones que deseen integrar estas herramientas en sus sistemas internos.

Estrategias de Implementación y Escalabilidad

Inicialmente, el negocio se lanzará en una fase de prueba piloto con un grupo selecto de usuarios para refinar funcionalidades y validar la robustez del sistema. La captación de usuarios iniciales se realizará mediante estrategias de marketing digital y colaboraciones con redes de inversores y entidades empresariales. Posteriormente, se implementarán estrategias de expansión geográfica y sectorial para adaptar la solución a diversos mercados y requerimientos.

Impacto Potencial y Beneficios Esperados

Se anticipa que esta firma revolucionará las prácticas de inversión y gestión empresarial mediante el suministro de accesos a análisis predictivos de vanguardia. Esta innovación no solo potenciará la toma de decisiones informadas por parte de los inversores sino también facilitará a las empresas la planificación efectiva frente a incertidumbres económicas. Los beneficios para los usuarios finales abarcan desde mejores rendimientos de inversión y gestión de riesgos optimizada hasta una capacidad reforzada para maniobrar en mercados volátiles.

20. Resumen Narrativo

Este trabajo de Fin de Grado ha explorado la influencia de las variables macroeconómicas en el ecosistema empresarial español, extendiéndose más allá del impacto inicial de la pandemia para abordar un espectro más amplio de factores económicos que afectan tanto la constitución como la disolución de empresas. El análisis ha evolucionado desde un enfoque inicial centrado exclusivamente en la pandemia, hacia un estudio más amplio que considera una variedad de variables macroeconómicas, adaptándose a las complejidades y a la falta de datos específicos sobre el COVID-19. A través de un proceso iterativo de ajuste metodológico, se han reformulado las preguntas de investigación para capturar mejor la dinámica del entorno empresarial en relación con el PIB, el IPC, y especialmente el Indicador de Clima Empresarial (ICEA), que ha demostrado ser significativo en el análisis.

Además, se ha integrado la perspectiva de expertos a través de entrevistas, complementando los análisis cuantitativos con experiencias y observaciones cualitativas que enriquecen los hallazgos. Este enfoque mixto ha permitido no solo validar los modelos estadísticos utilizados, sino también profundizar en la comprensión de cómo las empresas españolas se adaptan y responden a las incertidumbres macroeconómicas. La investigación concluye con una serie de recomendaciones prácticas para empresarios y formuladores de políticas, enfocadas en fortalecer la resiliencia y el crecimiento del ecosistema empresarial frente a futuras crisis económicas.

Este estudio no solo contribuye al conocimiento académico, sino que también propone soluciones prácticas, destacando la importancia de la adaptabilidad y la previsión en la gestión empresarial y la formulación de políticas. Así, se establece un puente entre la teoría y la práctica, subrayando cómo los insights obtenidos pueden aplicarse para mejorar el entorno empresarial y económico en España.

21. Bibliografía

Fundación BBVA – Ivie. (2016). Crecimiento económico y tejido empresarial en España. ESENCIALES Fundación BBVA – Ivie N.º 04/2016. Recuperado de https://www.fbbva.es/wp-content/uploads/2017/07/FBBVA_Esenciales_04_tcm269-567812.pdf

Cámara de Comercio de España. (n.d.). La creación de empresas. Recuperado de <https://www.camara.es/publicaciones/creacionempresas>

Análisis de los factores determinantes de la creación de empresas: una evidencia empírica en Castilla y León. (n.d.). Recuperado de <https://dialnet.unirioja.es/descarga/articulo>

Bermudez Barrezueta, N., & Bravo Matamoros, A. (2021). Modelo predictivo de los determinantes del cierre empresarial de las MIPYMES en el Ecuador período 2007-2016. X-Pedientes Económicos, 3(5), 78–93. Recuperado de https://ojs.supercias.gob.ec/index.php/X-pedientes_Economicos/article/view/32

Datos.gob.es. Gobierno de España (Conjuntos de Datos). (n.d.). https://datos.gob.es/es/catalogo?administration_level=E&theme_id=sector-publico

Carr, R. (1966). Spain, 1808-1939. Oxford: Oxford University Press.

Nadal, J. (1975). El fracaso de la revolución industrial en España, 1814-1913. Barcelona: Ariel.

Ringrose, D. R. (1996). Spain, Europe, and the "Spanish Miracle", 1700-1900. Cambridge: Cambridge University Press.

Tortella, G. (1994). Los orígenes del capitalismo en España: banca, industria y ferrocarriles en el siglo XIX. Madrid: Tecnos.

Blanco, R., Mayordomo, S., Menéndez Pujadas, Á., & Mulino Ríos, M. (2021, August 24). El Impacto de la crisis del covid-19 sobre la vulnerabilidad Financiera de las empresas españolas. Repositorio Institucional: Análisis económico e investigación. <https://repositorio.bde.es/handle/123456789/17431>

Pérez-Calle, Mg. R. D., García-Casarejos, Dra. N., & García-Bernal, Dr. J. (2021, septiembre). La empresa española ante la COVID-19: factores de adaptación al nuevo

escenario. RETOS. Revista de Ciencias de la Administración y Economía.

http://scielo.senescyt.gob.ec/scielo.php?script=sci_arttext&pid=S1390-86182021000100005

Blanco, R., & García, M. (2022, marzo). La evolución de la solvencia y de la demografía empresarial en España desde el inicio de la pandemia. Repositorio Institucional del Banco de España. <https://repositorio.bde.es/handle/123456789/22542>

Carr, R., & Fusi, J. P. (1987). Spain: Dictatorship to Democracy. London: Routledge.

Malefakis, E. E. (1970). Agrarian Reform and Peasant Revolution in Spain: Origins of the Civil War. New Haven: Yale University Press.

Payne, S. G. (1993). Franco and Hitler: Spain, Germany, and World War II. New Haven: Yale University Press.

Preston, P. (1996). The Spanish Civil War: Reaction, Revolution, and Revenge. New York: W. W. Norton & Company.

Viñas, Á. (1984). Guerra, dinero, dictadura: Ayuda fascista y autarquía en la España de Franco. Barcelona: Crítica.

Blanco, R., & Mayordomo, S. (2023, agosto 24). Evidencia sobre el alcance de los programas de garantías públicas y de Ayudas DIRECTAS a las empresas españolas implementados Durante La crisis del covid-19. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4540135

Soldevila, L. (2021, junio). Los emprendedores después de la COVID-19. Barcelona; OBS Business School (Planeta Formación y Universidades).

Castelló, V. (2021, abril 12). El Impacto de la crisis en las empresas españolas tras un año de pandemia. El País. <https://elpais.com/economia/estar-donde-estes/2021-04-12/el-impacto-de-la-crisis-en-las-empresas-espanolas-tras-un-ano-de-pandemia.html>

Gutiérrez, H. (2023, febrero 10). El Cierre de Empresas en España en 2022 batió récords: 26.207 disoluciones, un 10% más. El País. <https://elpais.com/economia/2023-02-10/el-cierre-de-empresas-en-espana-en-2022-batio-record-26207-disoluciones-un-10-mas.html>

Guillén, M. F. (1997). The Limits of Convergence: Globalization and Organizational Change in Argentina, South Korea, and Spain. Princeton, NJ: Princeton University Press.

Holman, O. (1996). Integrating Southern Europe: EC Expansion and the Transnationalization of Spain. London: Routledge.

Royo, S. (2000). Spain and Portugal in the European Union: The First Fifteen Years. London: Frank Cass.

Dirección General de Economía y Estadística. (2021, December 2). La evolución económica y financiera de las empresas españolas en 2020 y 2021 según la central de balances. BANCO DE ESPAÑA (Eurosistema).

Banco de España. (s.f.). <https://www.bde.es/wbe/es/>

Molina, S., & Mora, A. (2010). La crisis financiera y las PYMES en España. Madrid: Fundación de las Cajas de Ahorros.

Ruesga, S. M., & Da Silva, A. (2011). La crisis económica y su impacto en las empresas españolas. Revista de Economía Aplicada, 19(57), 5-28.

Saurina, J. (2009). Dinámicas de crédito y su impacto en las PYMES durante la crisis. Banco de España: Informe Económico.

Instituto Nacional de Estadística. (n.d.). Economía (Empresas). INE.
https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=125473557541

Datos.gob.es. Gobierno de España (Conjuntos de Datos). (n.d.).
https://datos.gob.es/es/catalogo?administration_level=E&theme_id=sector-publico

González, P., Gómez, M. J., & Molina, S. (2020). Impacto de la COVID-19 en las PYMES españolas. Madrid: Fundación de Estudios Financieros.

López-Cózar, C. (2020). Respuesta gubernamental a la crisis de la COVID-19 en España. Revista de Economía y Política, 35(2), 123-142.

Martínez, L. (2021). Digitalización y adaptación empresarial en tiempos de COVID-19. Barcelona: Editorial UOC

ANEXOS

ANEXO DE ENTREVISTAS COMPLETAS

Entrevista Alberto de Santos

Consideración de la Situación Macroeconómica de Alberto y su equipo en el Desarrollo de Koulu (Startup de educación)

Alberto explicó que durante el desarrollo inicial de Koulu, se tuvo en cuenta la situación macroeconómica de la época, específicamente en torno a la crisis inmobiliaria y la incertidumbre económica que persistía en esos años. El equipo de Koulu realizó un estudio para entender el gasto promedio de las familias en educación adicional, un factor crucial dado que Koulu estaba destinado a ser financiado por los padres de familia. Este análisis se apoyó en investigaciones previas de firmas multinacionales para establecer un marco de referencia sobre los gastos educativos en relación con el ingreso medio en España.

Interés de Alberto en la Situación Macroeconómica durante las Estancias en Hewlett-Packard, Vodafone y Levi

En esta pregunta Alberto destacó la importancia del contexto macroeconómico tanto en su interés personal como profesional. En su experiencia, el desempeño macroeconómico influye y directamente en cosas como la evaluación del desempeño laboral a través del rendimiento económico que tiene la empresa donde se trabaja. Sin embargo, cree que actualmente y debido al contexto tan inestable que existe, puede llegar a afectar significativamente la estabilidad de los empleos dentro de las empresas.

Alberto también contaba, como en Levi, específicamente, llegó a realizar un estudio para comprender cómo la inflación impactaba en los productos sustitutivos ofrecidos por la empresa y el patrón de compra de los clientes. Este estudio se vio interrumpido por el inicio del conflicto entre Rusia y Ucrania, que desestabilizó las tendencias económicas existentes. Así mismo, enfatizó que el interés en la macroeconomía puede que haya evolucionado de una mera curiosidad a una preocupación esencial para la supervivencia profesional, especialmente en el contexto de las recientes inestabilidades económicas globales.

Mejora del Análisis de Variables en el Modelo Analítico

El entrevistado, considera que el porcentaje de explicación obtenido en el estudio, aunque pueda parecer modesto, es significativo en el contexto de las decisiones

macroeconómicas y de negocios, donde incluso una pequeña fracción de predictibilidad puede ser valiosa. Subrayó que, en la vida real, las decisiones se toman con un conocimiento de aproximadamente un 40 o 50%, y tener un 5% o 6% de explicación en fenómenos complejos como las disoluciones y constituciones empresariales es un logro notable.

Sin embargo, sugiere que para mejorar el modelo y aumentar este porcentaje, sería beneficioso incorporar variables más diversificadas y ortogonales que no estén directamente relacionadas con los indicadores macroeconómicos habituales. Esto podría incluir factores como niveles educativos, tasas de desempleo locales, comportamientos de empresas específicas, inversiones en marketing, o incluso la situación económica de otros países que puedan influir en España.

Atención de los Estudiantes a la Situación Macroeconómica

De Santos compartió sus observaciones sobre el nivel de interés que sus estudiantes de inteligencia artificial muestran hacia las situaciones macroeconómicas, dependiendo del año de estudio en que se encuentran. Según explica, los estudiantes de primer y tercer año tienden a concentrarse más en sus estudios y en superar las exigencias académicas del momento, mostrando prácticamente ningún interés en los acontecimientos macroeconómicos. Sin embargo, observa un cambio en los estudiantes de cuarto año, quienes empiezan a considerar estos factores un poco más, aunque su enfoque principal sigue siendo encontrar empleo y ganar experiencia inmediata post-graduación. En general, percibe que la mayoría de los estudiantes no discuten ni consideran activamente los factores macroeconómicos en sus decisiones profesionales o académicas, enfocándose en lugar de eso en objetivos a corto plazo.

Viabilidad y Valor de la Propuesta de Negocio

Alberto evaluó la propuesta de negocio descrita para el TFG, calificándola de interesante y ambiciosa, particularmente en el contexto de un proyecto académico. Destacó la importancia y el desafío de articular y desarrollar una idea tan compleja en un entorno de trabajo de fin de grado. Desde una perspectiva profesional más práctica, sugirió reconsiderar el modelo de negocio para evitar la dependencia directa de las empresas en cuanto a la provisión de datos, identificándolo como un potencial riesgo. Propuso alternativas para obtener información relevante de forma independiente, como utilizar datos públicos de las empresas para realizar análisis y ofrecer servicios proactivos sin necesidad de solicitar datos adicionales directamente a las empresas. Su consejo se centra en transformar el enfoque hacia uno donde el valor se entregue de manera proactiva, posicionando el servicio como un recurso valioso y un socio de confianza para las empresas.

Entrevista José María Martín-Mateos

Consideración de la Situación Macroeconómica de José María y su equipo en el desarrollo y lanzamiento de Controlnet (Empresa de Desarrollo de software personalizado)

En cuanto a si se tuvo en cuenta la situación macroeconómica al lanzar el producto, José María Martín-Mateos señaló que, durante la época de lanzamiento, el sector de las telecomunicaciones y soluciones empresariales digitales estaba en auge. Aunque la situación de capital de las empresas era relevante debido a la crisis inmobiliaria, la situación macroeconómica general no fue un factor decisivo para el lanzamiento al mercado. Sin embargo, sí se observó y estudió el crecimiento en la venta de productos tecnológicos en otras empresas, que mostraba un aumento exponencial.

Importancia de las situaciones o variables macroeconómicas del momento para la toma de decisiones.

Respecto a la importancia de las variables macroeconómicas en la toma de decisiones para entrar en algunos países, destacó que ciertos indicadores, como el índice Big Mac o las políticas jurídicas de un país, son considerados para evaluar cómo podrían afectar los proyectos a realizar. Además, Martín-Mateos tiende a analizar indicadores económicos de líderes en los sectores de telecomunicaciones y digitalización, como Estados Unidos o Alemania.

Ayuda recibida por parte de las administraciones públicas durante las crisis macroeconómicas.

Sobre el apoyo de las administraciones públicas durante crisis macroeconómicas, indicó que desde el principio, las administraciones públicas han apoyado a Controlnet no solo en visibilidad y difusión sino también en impulsar proyectos y acciones de beneficio mutuo. Por ejemplo, se mencionó el inicio de un ecosistema tecnológico en la provincia de Cádiz para fomentar el empleo y la formación profesional en el sector.

En relación con las respuesta de Alberto de Santos, opinión sobre el enfoque de variables ortogonales y variables que sean interesantes estudiar y añadir al modelo?

En cuanto al enfoque de utilizar variables ortogonales en el análisis, consideró que es una buena estrategia, aunque recomendó enfocarse en variables relevantes para el sector específico bajo estudio. Por ejemplo, en el sector de las telecomunicaciones, sugirió centrarse en las ventas anuales de servicios o productos tecnológicos de las principales empresas de España o Europa, lo cual podría derivar en la creación de variables de ventas por cada sector importante para obtener una visión más clara del mercado.

Conocimiento sobre el ICEA y recomendaciones de índices que utilice para seguir la macroeconomía

Respecto al conocimiento y seguimiento de índices como el Índice de Confianza Empresarial Armonizado (ICEA), afirmó que no estaba familiarizado con este índice específicamente, pero utiliza índices similares en Alemania y EE. UU. para anticipar el comportamiento del sector. Considera que el seguimiento de estos índices puede proporcionar una perspectiva anticipada de las tendencias del mercado.

Recomendaciones y consejos para emprendedores

Finalmente, sobre las recomendaciones para nuevos emprendedores, destacó la importancia de tener un producto o servicio viable y escalable. Subrayó que la obsesión, el compromiso y el esfuerzo son esenciales para el éxito de un proyecto, lo que implica un enfoque total y la posibilidad de tener que sacrificar otros aspectos de la vida personal para construir un camino exitoso en el negocio.