Universidad
Francisco de Vitoria
UFV Madrid

2ª Parte del Trabajo de Fin de Grado

Análisis de los Datos

Trabajo Fin de Grado
Grado en Análisis de Negocios – Business Analytics
Febrero 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

RESUMEN

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 del Documento

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 e información de 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 e información de 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 4 clasificadas por tipo e información de 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 e información de 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 e información de 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 e información de 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 1. *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 2. *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 3. *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 4. *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 5. *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 6. *Medidas de tendencia central de la quinta 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 1. *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 2. *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 3. *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 4. *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 5. *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 6. *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.

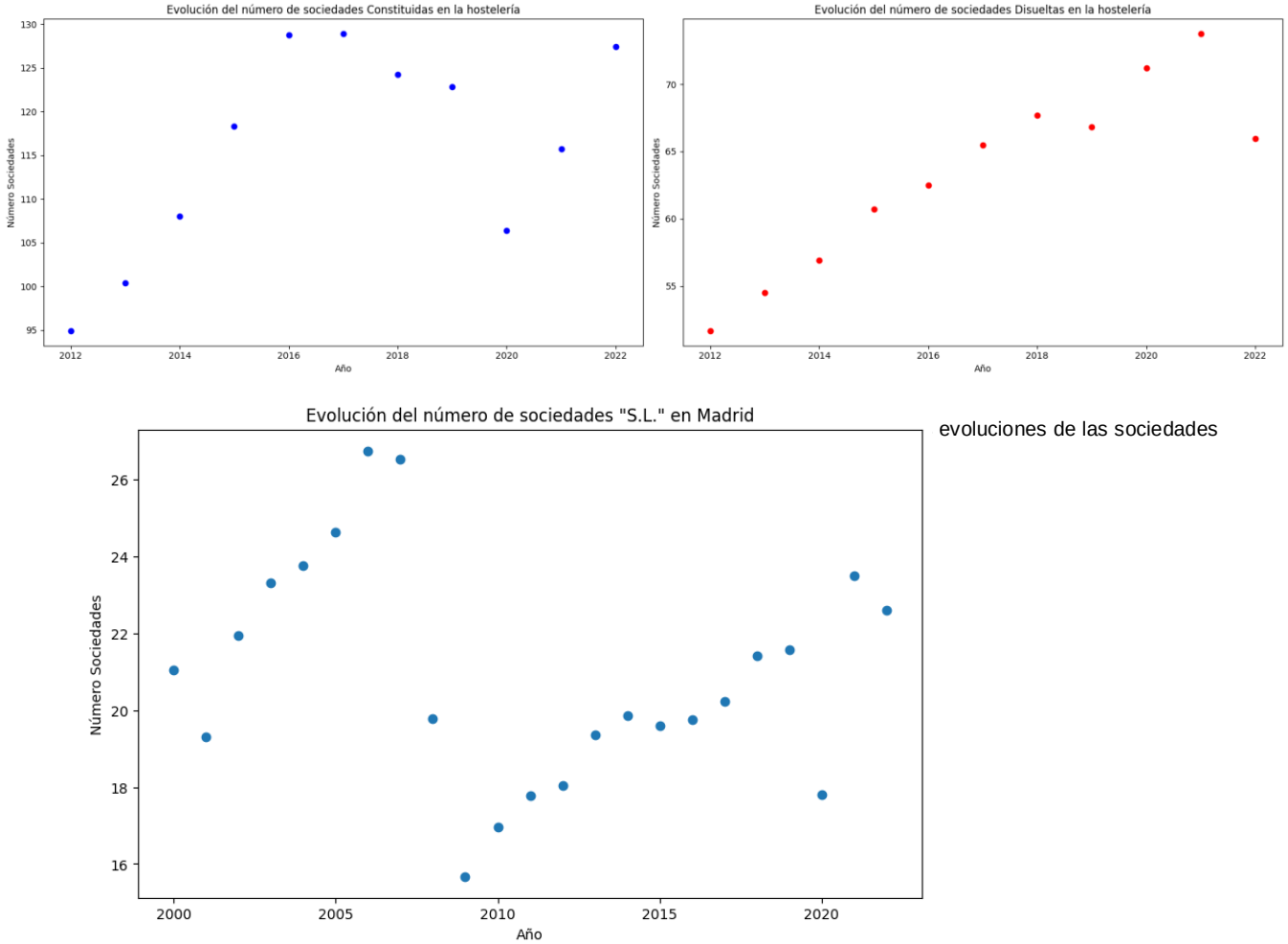
6. Gráficos Descriptivos

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.



evoluciones de las sociedades

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.

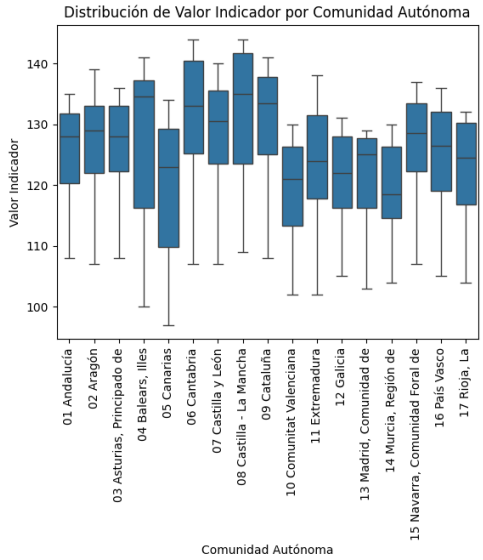


Ilustración 5. Captura de pantalla del diagrama de cajas correspondiente a la distribución del valor indicador "ICEA" según la comunidad autónoma.

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.

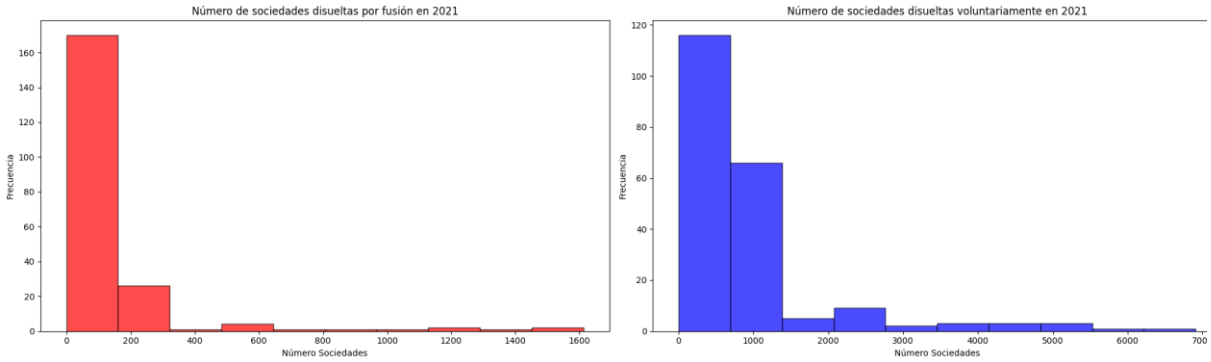


Ilustración 6. Captura de pantalla del histograma correspondiente a la evolución del número de sociedades disueltas voluntariamente en 2021 en todas las comunidades autónomas españolas.

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 Figura 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 "dfbbdd1" 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.

La segunda base de datos, conocida como "disolución", se originó a partir de un conjunto de datos que registraba el número de empresas disueltas por tipo de disolución en cada comunidad autónoma desde 2012 hasta 2022. Para profundizar en el análisis, se generó una base de datos complementaria, "macrospain", que incorporaba indicadores económicos clave como el PIB y el IPC por comunidad autónoma y año. El propósito de esta combinación era realizar una regresión lineal múltiple para investigar cómo el entorno económico, reflejado en el PIB y el IPC, afecta a la tasa de disolución de empresas, buscando patrones y relaciones significativas que pudieran explicar las tendencias observadas.

Finalmente, la tercera base de datos, "dismac", se creó con el objetivo de ampliar el espectro de variables macroeconómicas analizadas en relación con las disoluciones empresariales. Esta base de datos integró indicadores 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 buscaba ofrecer una comprensión más rica y matizada de cómo diversos factores económicos y sociales pueden influir en la estabilidad y la continuidad de las empresas en España.

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 se basa en el principio de mínimos cuadrados, buscando la línea que mejor se ajusta a los datos mediante la minimización de la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. Este enfoque permite obtener una ecuación lineal que describe cómo el cambio en la variable independiente afecta la variable dependiente, proporcionando así una base cuantitativa para la toma de decisiones y el análisis económico.

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 (4.79e+06) 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 (1ª Regresión Lineal Múltiple)

Explicación del Modelo Supervisado Seleccionado

Para investigar la influencia del Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) en el número de disoluciones empresariales, se seleccionó un modelo de regresión lineal múltiple. Este modelo permite examinar cómo múltiples variables independientes (en este caso, PIB e IPC) afectan a una variable dependiente (Disoluciones). La elección de este modelo se fundamenta en su capacidad para proporcionar una comprensión detallada de las relaciones entre variables económicas y su impacto en la estabilidad empresarial, permitiendo así identificar patrones y tendencias significativas en el contexto económico de España.

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 en la tasa de disoluciones 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.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal:* Se realizaron gráficos de dispersión para visualizar las relaciones entre las variables seleccionadas. Aunque los coeficientes de correlación entre Disoluciones e IPC (0.0179) y entre Disoluciones y PIB (0.0273) fueron relativamente bajos, indican una posible relación lineal. Estos resultados preliminares justificaron la inclusión de ambas variables en el modelo de regresión lineal múltiple para un análisis más profundo.
- *Análisis de Ajuste a una Distribución Normal:* Las pruebas de normalidad para las variables Disoluciones, IPC y PIB mostraron desviaciones de la normalidad, como se evidencia en los resultados de las pruebas Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2. Estos hallazgos sugieren que, aunque las variables no siguen perfectamente una distribución normal, el modelo de regresión lineal múltiple aún puede proporcionar insights valiosos.

- *Construcción del modelo:* El modelo de regresión lineal múltiple fue construido incorporando el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) como variables independientes para predecir el número de disoluciones empresariales. La inclusión de estas variables se justificó por la hipótesis de que reflejan aspectos fundamentales de la salud económica que podrían influir en la estabilidad de las empresas. El ajuste del modelo se realizó mediante el método de mínimos cuadrados ordinarios (OLS), resultando en un coeficiente para el IPC de -6.4564 con un error estándar de 46.460 y para el PIB de 6.611e-10 con un error estándar de 2.02e-09. La constante del modelo se estimó en 1017.4258 con un error estándar significativo de 2825.706, indicando la base sobre la cual se evalúa el efecto de las variables independientes.

Evaluación del Modelo con Medidas de Error/Precisión Específicas:

- *Calidad del modelo*: La calidad del modelo se evaluó a través del coeficiente de determinación R-cuadrado, que fue de 0.001, y el R-cuadrado ajustado de -0.009. Estos valores indican que el modelo explica menos del 1% de la variabilidad en el número de disoluciones empresariales, lo que sugiere una capacidad predictiva muy limitada de las variables seleccionadas sobre el fenómeno de interés. El F-estadístico de 0.08673 con un p-valor de 0.917 refuerza la noción de que el modelo, en su estado actual, no proporciona una base estadísticamente significativa para predecir las disoluciones empresariales basándose en el PIB y el IPC.
- *Confiabilidad del modelo**:*** La confiabilidad del modelo se ve cuestionada por un número de condición extremadamente alto (3.04e+13), lo que sugiere la presencia de multicolinealidad entre las variables independientes. Esto implica que las variables seleccionadas pueden no ser independientes entre sí, complicando la interpretación de sus coeficientes individuales y potencialmente inflando los errores estándar. Tal multicolinealidad compromete la fiabilidad de las estimaciones de los coeficientes y, por ende, la confianza en las inferencias realizadas a partir del modelo.
- *Análisis:* El análisis de los resultados del modelo de regresión lineal múltiple revela limitaciones significativas en su capacidad para proporcionar insights predictivos o explicativos robustos sobre las disoluciones empresariales en España. La baja capacidad explicativa,

evidenciada por los valores de R-cuadrado y R-cuadrado ajustado, junto con la falta de significancia estadística de las variables independientes (como lo indican los p-valores elevados) y los problemas de multicolinealidad, sugieren que el modelo actual no captura adecuadamente la complejidad de las relaciones entre las variables económicas y las disoluciones empresariales.

Además, la evaluación del modelo mediante el RMSE (1741.7715) y el R^2 ajustado (-0.0255) en la fase de validación subraya la inadecuación del modelo para predecir con precisión el número de disoluciones empresariales. Estos indicadores apuntan a una discrepancia significativa entre los valores observados y los predichos por el modelo, lo que refleja su limitada utilidad práctica.

En conclusión, aunque el modelo de regresión lineal múltiple representa un enfoque teóricamente válido para explorar las relaciones entre múltiples variables independientes y una variable dependiente, los resultados obtenidos sugieren que es esencial reconsiderar la selección de variables, explorar la inclusión de otras variables potencialmente relevantes o emplear métodos analíticos alternativos que puedan capturar mejor la dinámica entre la salud económica y las disoluciones empresariales.

10.3. Modelo Analítico Supervisado (2ª Regresión Lineal Múltiple)

Explicación del Modelo Supervisado Seleccionado

En el presente análisis, se adopta un enfoque de regresión lineal múltiple avanzado con el objetivo de examinar la influencia de un conjunto ampliado de variables económicas sobre el fenómeno de las disoluciones empresariales, referidas en este contexto como "Disoluciones".

Este modelo incorpora, además del Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC), una serie de variables que se postulan como determinantes potenciales en este proceso, incluyendo el Indicador de Clima Empresarial (ICEA), la Deuda Pública, el Déficit Público, el Gasto Público, los Ingresos Fiscales, las Llegadas de Turistas y las Reservas Totales.

La inclusión de estas variables busca enriquecer el análisis al considerar una gama más amplia de factores que podrían tener un impacto significativo en la tasa de disoluciones empresariales, ofreciendo así una comprensión más profunda y matizada de las dinámicas

que afectan la estabilidad empresarial en España. Este enfoque metodológico no solo amplía el espectro de análisis, sino que también facilita la identificación de relaciones complejas entre las disoluciones empresariales y el entorno económico, permitiendo una evaluación más detallada de cómo diversos factores económicos contribuyen a configurar el panorama empresarial del país.

Proceso de Selección de Variables

El proceso de selección de variables para el modelo de regresión lineal múltiple se llevó a cabo mediante un procedimiento meticuloso y estructurado, con el fin de incorporar un espectro amplio de indicadores económicos que pudieran influir en la tasa de disoluciones empresariales en España. Inicialmente, se partió de la base de datos "disolución", a la cual se le añadieron variables provenientes de dos fuentes adicionales: un conjunto de indicadores macroeconómicos y el Índice de Confianza Empresarial Armonizado (ICEA).

Una vez consolidada la base de datos ampliada, denominada "dismac", se procedió a la selección de variables específicas para el modelo. Esta selección incluyó, además del IPC y el PIB, variables como el ICEA, la Deuda y el Déficit Públicos, el Gasto Público, los Ingresos Fiscales, las Llegadas de Turistas y las Reservas Totales. Esta diversidad de variables buscó capturar la complejidad de los factores económicos que podrían estar incidiendo en las disoluciones empresariales, desde la perspectiva macroeconómica hasta indicadores más específicos del clima empresarial y el turismo.

Desarrollo del Modelo

- *Análisis de la Existencia de Relación Lineal:* En el análisis de la existencia de relación lineal para el modelo de regresión lineal múltiple, se emplearon gráficos de dispersión y la correlación de Pearson para examinar cómo variables económicas como el IPC, PIB, ICEA, entre otras, influyen en el número de disoluciones empresariales. Los gráficos de dispersión ofrecieron una visualización directa de las tendencias entre las variables, facilitando la identificación de patrones o anomalías. La matriz de correlación, obtenida a través de la función corr() de pandas, reveló las asociaciones lineales entre las disoluciones y cada variable independiente, siendo crucial para determinar la fuerza y dirección de estas relaciones.

  Los coeficientes de correlación de Pearson entre las disoluciones y variables como el IPC (0.0179, p-valor=0.7967) y el PIB (0.0273, p-valor=0.6943) indicaron correlaciones débiles y no significativas estadísticamente. Sin

embargo, la variable ICEA mostró una correlación más notable (0.1562, p-valor=0.0239), sugiriendo una influencia potencial sobre las disoluciones empresariales. Estos resultados subrayan la importancia de seleccionar cuidadosamente las variables para el modelo, basándose en evidencia numérica sólida y significancia estadística, para asegurar un análisis robusto y confiable de los factores que afectan la estabilidad empresarial en España.

- *Análisis de Ajuste a una Distribución Normal:* El análisis de ajuste a una distribución normal para el modelo de regresión lineal múltiple reveló variaciones significativas en la normalidad de las distribuciones de las variables analizadas. Mediante gráficos de densidad, se observaron las distribuciones de variables como Disoluciones, IPC, PIB, entre otras, identificando desviaciones de la normalidad esperada. Estas visualizaciones, complementadas con pruebas de normalidad como Shapiro-Wilk, Anderson-Darling y D'Agostino's K^2, proporcionaron una evaluación cuantitativa de la normalidad. Por ejemplo, la prueba Shapiro-Wilk para Disoluciones arrojó un estadístico de 0.714 y un p-valor significativamente bajo (1.15e-18), indicando una desviación clara de la normalidad. Similarmente, otras variables como ICEA y Gasto mostraron p-valores que rechazan la hipótesis de normalidad, con estadísticos de Anderson-Darling y D'Agostino's K^2 que confirman estas observaciones.

  Estos resultados subrayan la evidencia de no normalidad en varias variables críticas sugiere cautela al interpretar los resultados del modelo y al aplicar inferencias estadísticas basadas en suposiciones de normalidad.
- *Construcción del modelo:* La construcción del modelo de regresión lineal múltiple se centró en analizar la influencia de una serie de variables macroeconómicas sobre el número de disoluciones empresariales en España. El modelo ajustado reveló que, aunque la mayoría de las variables no mostraron una relación estadísticamente significativa con las disoluciones, el indicador ICEA presentó un coeficiente positivo de 8.1745 con un p-valor de 0.006, sugiriendo una relación significativa con el número de disoluciones empresariales.

  El análisis de los resultados del modelo presencia multicolinealidad, sugerida por un número de condición elevado (2.17e+14), planteando así desafíos adicionales en la interpretación de los coeficientes individuales e indicando que algunas variables independientes pueden estar proporcionando información redundante sobre las disoluciones.

En conclusión, este modelo de regresión lineal múltiple ofrece una visión preliminar de cómo ciertos factores económicos y turísticos pueden estar relacionados con las disoluciones empresariales en España. Sin embargo, la baja significancia estadística de la mayoría de las variables y la presencia de multicolinealidad sugieren la necesidad de un análisis más profundo y la consideración de otros factores potenciales para comprender completamente las dinámicas detrás de las disoluciones empresariales.

- *Reconstrucción del modelo en base a la multicolinealidad:* La

Reconstrucción del modelo en base a la multicolinealidad

La evaluación del modelo de regresión lineal múltiple, tras el reajuste para abordar la multicolinealidad entre las variables independientes, se llevó a cabo mediante un análisis detallado del Factor de Inflación de la Varianza (VIF). Este proceso iterativo de eliminación de variables con altos VIF permitió identificar y descartar aquellas que contribuían significativamente a la multicolinealidad, como 'PIB', 'Deuda', e 'IngreFis', mejorando así la calidad y la interpretación del modelo. La eliminación de estas variables se basó en su potencial para distorsionar los resultados del modelo debido a su alta correlación con otras variables independientes.

El modelo final, ajustado con un conjunto refinado de variables ('IPC', 'ICEA', 'Déficit', 'Gasto', 'Turistas', 'Reservas'), mostró un R-cuadrado de 0.041, indicando que estas variables explican aproximadamente el 4.1% 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, muchos de los cuales pueden no estar capturados por el modelo. La variable 'ICEA' se destacó como estadísticamente significativa, sugiriendo que tiene un impacto directo en el número de disoluciones empresariales. Este hallazgo subraya la importancia de considerar el clima empresarial y la confianza económica al analizar las disoluciones.

Sin embargo, el alto número de condición (1.6e+10) aún señala la presencia de multicolinealidad residual, lo que sugiere que la interpretación de los coeficientes debe hacerse con cautela. Este desafío subraya la necesidad de continuar refinando el modelo y explorar otras variables o métodos que puedan capturar mejor la dinámica detrás de las disoluciones empresariales.

En conclusión, el proceso de reajuste del modelo y la evaluación de la multicolinealidad han sido pasos cruciales para mejorar su precisión y fiabilidad. Aunque el modelo ha logrado

cierto grado de claridad en la relación entre ciertas variables macroeconómicas y las disoluciones empresariales, los resultados también destacan la complejidad inherente a este fenómeno y la necesidad de investigaciones futuras para desarrollar un modelo más explicativo y representativo de las disoluciones empresariales en España.

Evaluación del Modelo con Medidas de Error/Precisión Específicas

- *Calidad del modelo*: La calidad del modelo de regresión lineal múltiple ajustado se evalúa a través de una combinación de métricas, incluyendo el R-cuadrado, el R-cuadrado ajustado, RMSE y el MAE para una comprensión más profunda del rendimiento del modelo. El R-cuadrado de 0.041 sugiere que solo un 4.1% de la variabilidad en las disoluciones empresariales es explicada por las variables seleccionadas, lo que indica una capacidad limitada del modelo para capturar la complejidad detrás de las disoluciones empresariales. Aunque este porcentaje es bajo, refleja la naturaleza multifacética de las disoluciones empresariales, influenciadas por una amplia gama de factores económicos y contextuales. El RMSE de 1728.63 y el MAE de 1023.05 proporcionan una medida cuantitativa del error de predicción del modelo, indicando la desviación promedio de las predicciones del modelo respecto a los valores reales. Estas métricas de error, especialmente el RMSE, sugieren que, aunque el modelo puede capturar algunas tendencias generales, hay una variabilidad significativa en las disoluciones empresariales que no está siendo completamente explicada.
- *Confiabilidad del modelo:* La confiabilidad del modelo se ve comprometida por la multicolinealidad entre las variables independientes, como se refleja en el alto número de condición. A pesar de los esfuerzos por mitigar este problema a través del análisis de VIF y la eliminación de variables con alta multicolinealidad, la persistencia de este fenómeno sugiere cautela en la interpretación de los coeficientes. La significancia estadística de ciertas variables, como el indicador ICEA, aporta valor al modelo, indicando su potencial para identificar factores relevantes que afectan las disoluciones empresariales. Sin embargo, la presencia de errores significativos, evidenciados por el RMSE y el MAE, subraya la necesidad de un análisis más detallado y la inclusión de variables adicionales que puedan ofrecer una explicación más completa de las disoluciones empresariales.

- *Análisis:* Este análisis revela que, a pesar de identificar algunas relaciones significativas, la capacidad del modelo para explicar la variabilidad en las disoluciones empresariales es limitada. El indicador ICEA emerge como un factor significativo, pero el bajo R-cuadrado junto con el RMSE y el MAE relativamente altos sugieren que hay aspectos del fenómeno de las disoluciones empresariales que el modelo actual no logra capturar. Esto destaca la necesidad de un enfoque más integral que incluya una gama más amplia de factores económicos, indicadores sectoriales y elementos cualitativos como la confianza empresarial y el entorno político. La evaluación del modelo enfatiza la importancia de abordar la multicolinealidad y otros supuestos estadísticos en el desarrollo de modelos de regresión, asegurando así que las inferencias y predicciones sean válidas y confiables. La inclusión de métricas de error como el RMSE y el MAE proporciona una perspectiva adicional sobre la precisión predictiva del modelo, subrayando áreas para futuras mejoras y refinamientos.

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:

- R-cuadrado (R^2): Representa la proporción de la variabilidad en la variable dependiente que puede ser explicada por el modelo de regresión. Un R^2 cercano a 1 indica que el modelo explica una gran parte de la variabilidad, mientras que un R^2 cercano a 0 sugiere lo contrario.
- R-cuadrado ajustado: Modifica el R^2 para reflejar el número de predictores en el modelo, proporcionando una medida más precisa de la bondad de ajuste, especialmente útil en modelos con múltiples variables independientes.

- Raíz del Error Cuadrático Medio (RMSE): Mide la desviación de los valores predichos por el modelo de los valores reales. Proporciona una estimación de la magnitud del error en las mismas unidades que la variable dependiente.
- Error Absoluto Medio (MAE): Similar al RMSE, pero mide el promedio de los errores absolutos. Es menos sensible a los valores atípicos que el RMSE y proporciona una medida más intuitiva del error promedio.

11.2. Comparación de los resultados obtenidos en los modelos.

Los modelos analizados presentan variaciones significativas en sus métricas de evaluación, reflejando diferencias en su capacidad para explicar y predecir las disoluciones empresariales:

Modelos de Regresión Lineal Simple: Los modelos mostraron R^2 cercanos a cero, indicando una capacidad explicativa muy limitada. Esto sugiere que la relación entre las variables independientes (población y capital desembolsado) y la variable dependiente (número de sociedades o capital) es débil o no lineal.

Modelos de Regresión Lineal Múltiple: Aunque estos modelos incorporaron más variables, los R^2 ajustados permanecieron bajos, lo que indica que la adición de variables no mejoró sustancialmente la capacidad explicativa de los modelos. Sin embargo, el análisis de VIF y la reconfiguración subsiguiente del modelo para minimizar la multicolinealidad resultaron en una mejora marginal de las métricas de error (RMSE y MAE), sugiriendo una ligera mejora en la precisión predictiva.

12. Visualización de Datos y Resultados de Modelos

12.1. Gráficos de Dispersión

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:

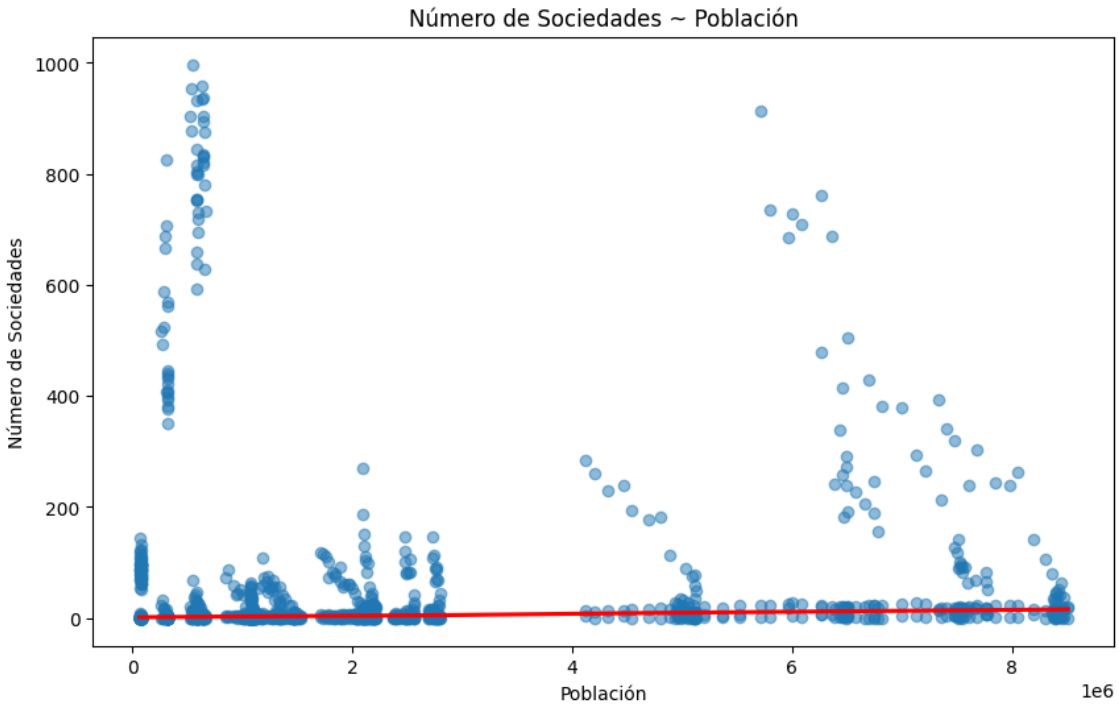


Ilustración 1. Gráfico de dispersión, que muestra la distribución del número de empresas en función del tamaño de la población, con una línea de tendencia suavizada que indica el patrón general de los datos.

El análisis del gráfico de dispersión que compara el número de sociedades con la población sugiere una relación no lineal entre estas variables, con un patrón de crecimiento que se aplana para rangos más amplios de población antes de incrementarse levemente. Este patrón indica una variabilidad considerable en el capital en áreas de baja población y un efecto de "meseta" para valores altos de población, sugiriendo rendimientos decrecientes en la acumulación de capital con el aumento de la población. La dispersión de datos en valores altos de población refleja una diversidad mayor en la acumulación de capital en áreas más pobladas, y el posible error de etiquetado subraya la importancia de una interpretación precisa de los datos.

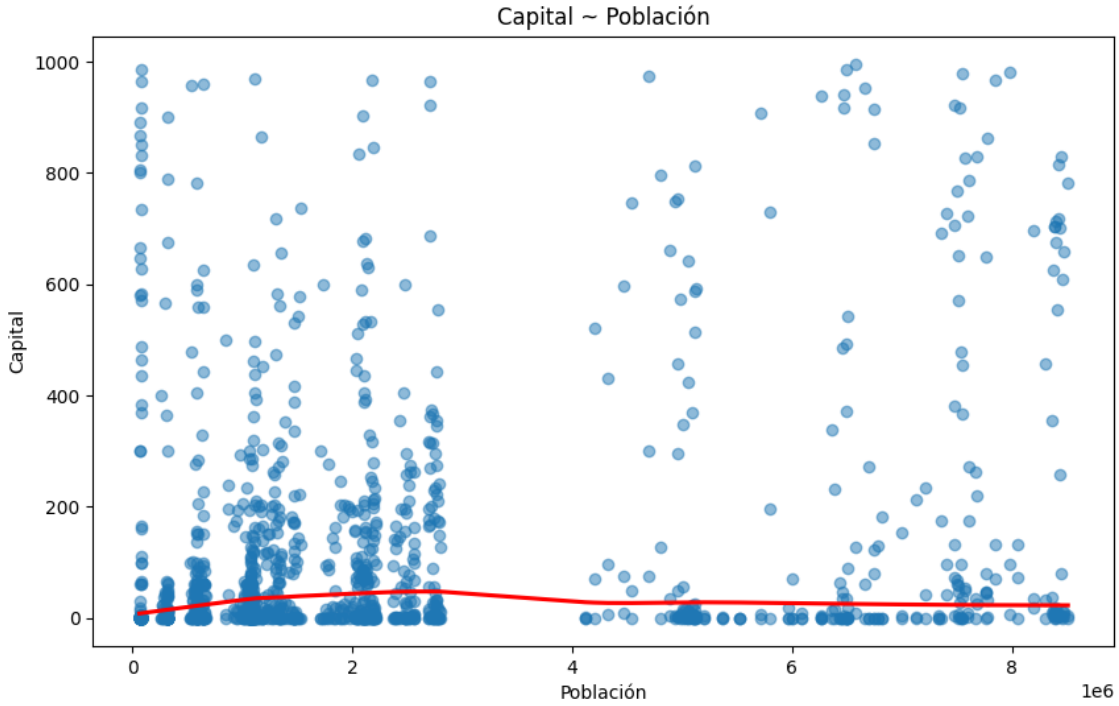


Ilustración 2. Gráfico que muestra la dispersión del capital desembolsado en relación con los distintos tamaños de población, complementado con una línea de tendencia suavizada para ilustrar la tendencia general dentro del conjunto de datos.

El gráfico de dispersión entre el capital desembolsado y la población revela una relación lineal positiva muy débil, caracterizada por una amplia dispersión de datos alrededor de la línea de tendencia y una variabilidad significativa en el número de empresas para poblaciones menores. Se identifican valores atípicos que sugieren la existencia de ciudades o centros con un número elevado de empresas, y se observa una tendencia a la estabilización o "meseta" en el número de empresas a medida que aumenta la población, lo que indica posibles factores limitantes para el crecimiento empresarial en áreas de alta población. La escasez de datos en el cuadrante de alta población y bajo número de empresas sugiere un umbral mínimo de empresas presentes en áreas de gran población.

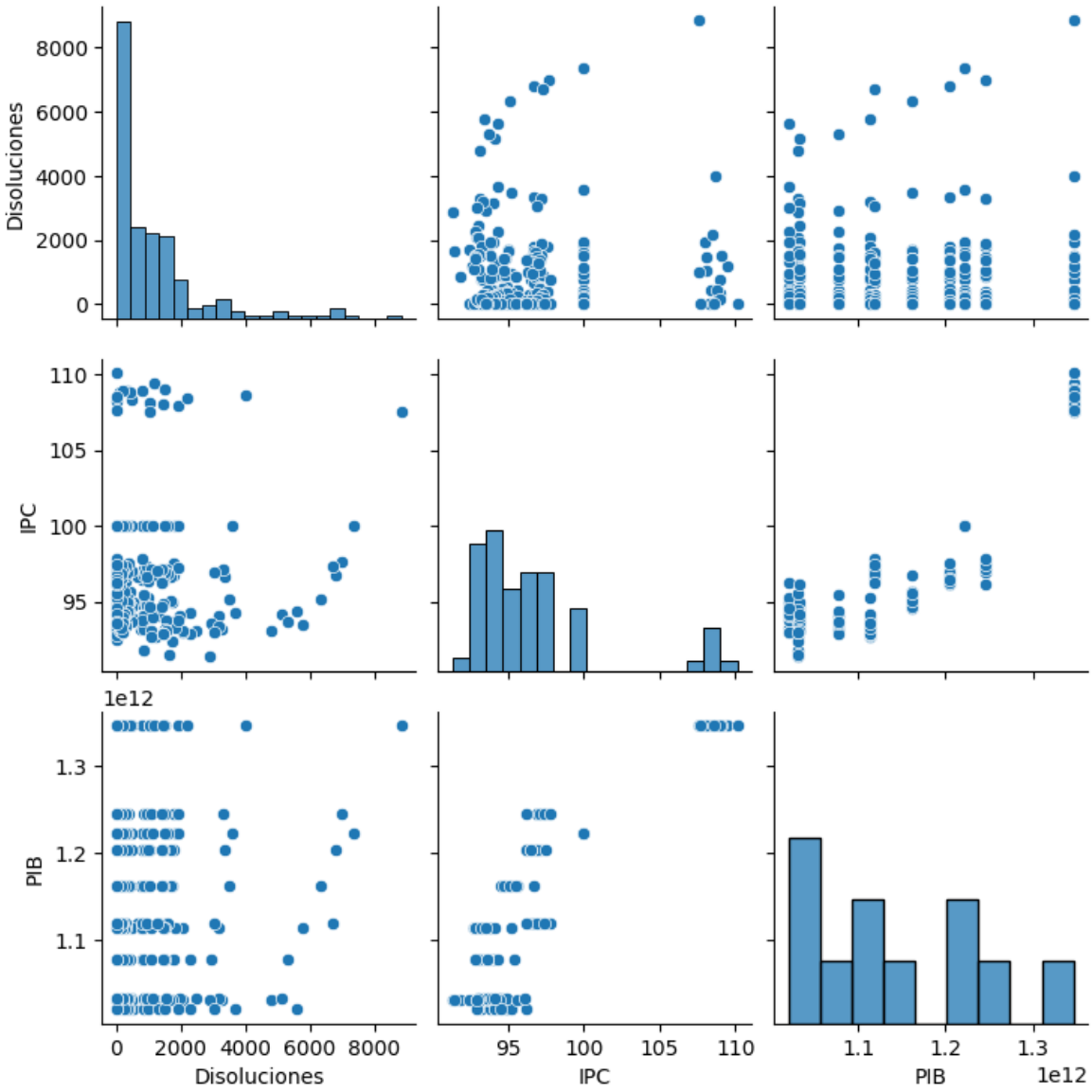


Ilustración 3. Matriz de diagrama de pares de "Disoluciones", "IPC" y "PIB": Esta matriz consiste en histogramas a lo largo de la diagonal y gráficos de dispersión en las celdas fuera de la diagonal que visualizan las relaciones de pares entre estas tres variables.

Los histogramas en la diagonal ilustran la distribución de cada variable, donde "Disoluciones" y "PIB" muestran distribuciones sesgadas hacia la derecha, lo que indica una concentración de observaciones en valores bajos con algunos valores mucho más altos. Por otro lado, el "IPC" parece tener una distribución más simétrica. Los gráficos de dispersión fuera de la diagonal revelan las relaciones entre pares de variables, mostrando patrones no claros entre "Disoluciones" y "IPC" y "PIB", lo que sugiere relaciones débiles o no lineales. Sin embargo, el gráfico de dispersión entre "IPC" y "PIB" sugiere una correlación positiva posible, ya que los puntos parecen seguir una tendencia ascendente.

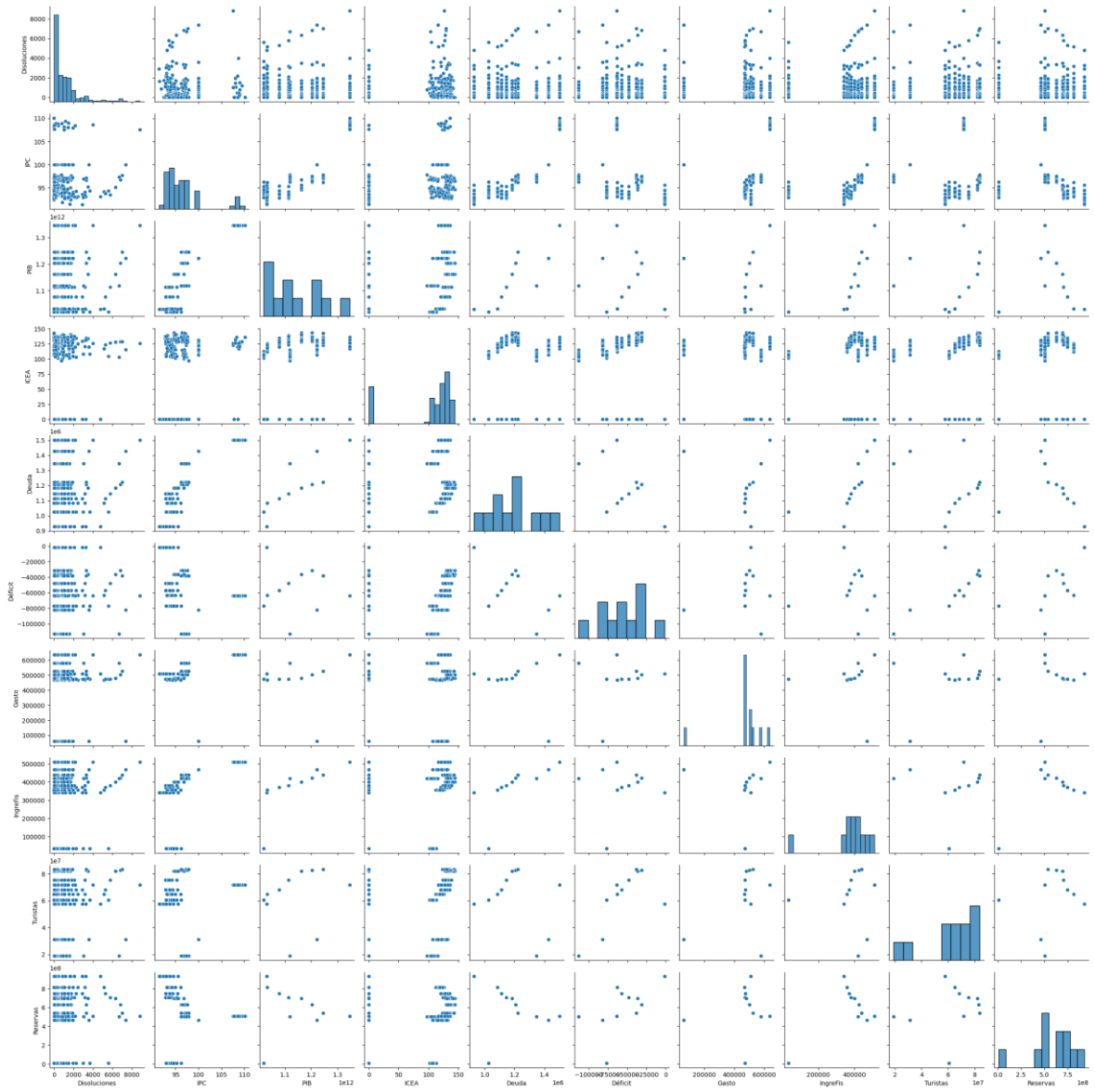


Ilustración 4. Matriz ampliada de diagramas de pares de variables económicas y sociales

El análisis de correlación entre indicadores económicos y sociales revela relaciones significativas, especialmente entre el Índice de Precios al Consumidor (IPC), el Producto Interno Bruto (PIB) y la deuda nacional, sugiriendo que el aumento en el IPC se asocia con un crecimiento del PIB y un incremento en la deuda. Los ingresos fiscales también muestran una correlación positiva con el PIB, indicando que el crecimiento económico incrementa los ingresos gubernamentales. Contrariamente, existe una correlación negativa entre el déficit y la deuda, y una fuerte correlación positiva entre el turismo y la reducción del déficit, posiblemente por el impulso económico del turismo. Las disoluciones de empresas tienen correlaciones débiles con estas variables, con una leve tendencia a aumentar con la actividad económica. Este análisis subraya la interconexión entre estos indicadores y su relevancia para futuras investigaciones y políticas económicas.

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 modelos de regresión lineal múltiple centrados en "disolución" y "dismac". 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.

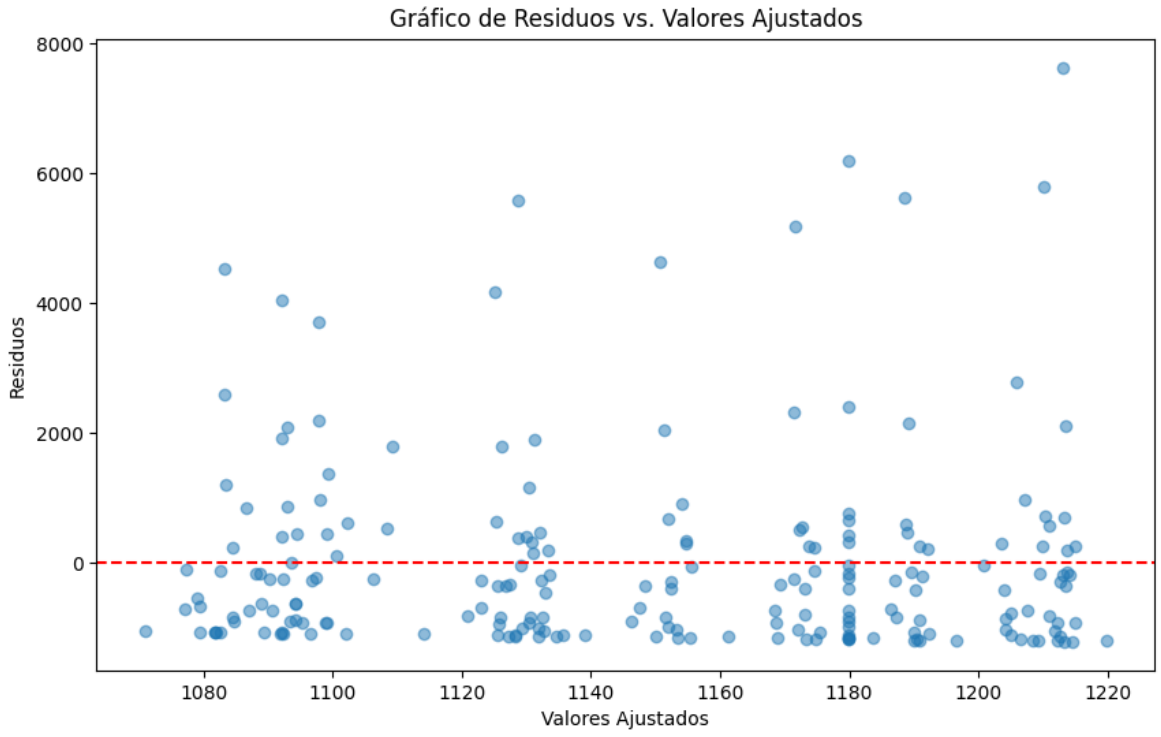


Ilustración 5. Gráfico de residuos correspondiente a la primera regresión lineal múltiple entre la variable "Disoluciones", "PIB", "IPC"

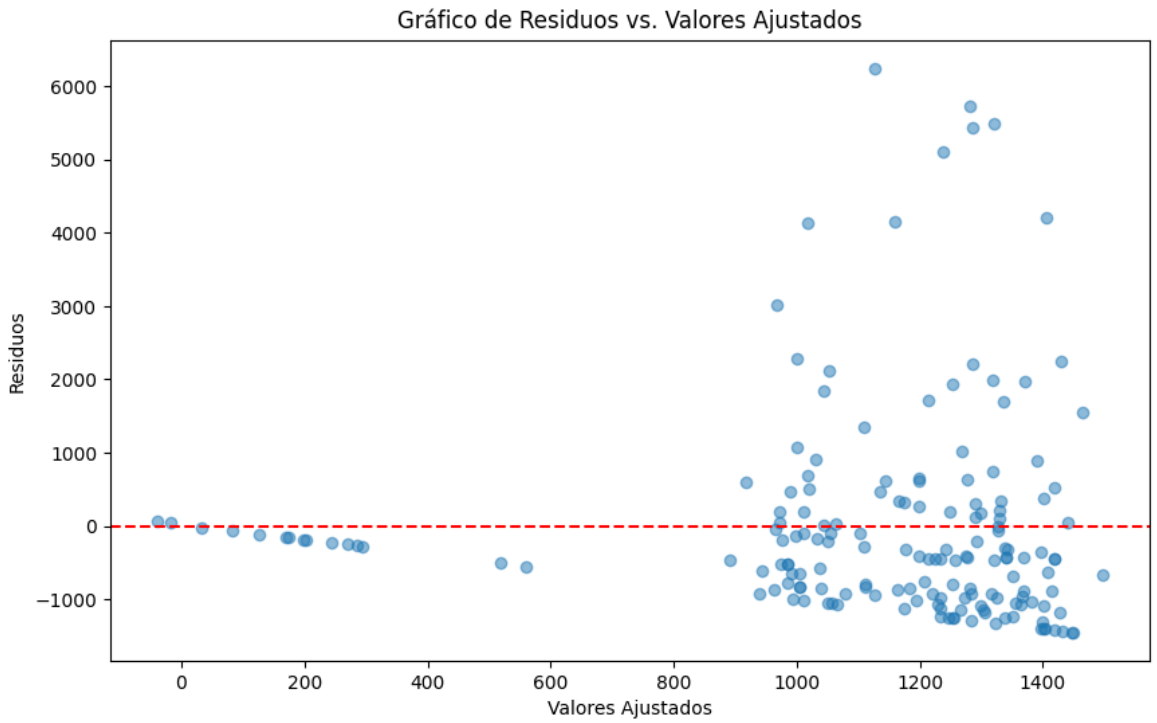


Ilustración 6. Gráfico de residuos correspondiente a la segunda regresión lineal múltiple entre la variable "Disoluciones", y los indicadores macroeconómicos

Los gráficos de residuos analizados ofrecen una perspectiva valiosa sobre los desafíos y complejidades inherentes al modelado de las disoluciones empresariales en España mediante regresiones lineales. A pesar de las dificultades encontradas, como la presencia de patrones sistemáticos en los residuos y la heterocedasticidad la dispersión aleatoria de residuos en ambos gráficos, junto con la identificación de outliers, sugiere que los modelos lineales, aunque útiles, no capturan completamente las complejidades y dinámicas subyacentes de los datos. Este hallazgo resalta la necesidad de explorar especificaciones de modelos alternativos y considerar la estructura específica de los datos, como efectos regionales, para mejorar la precisión y relevancia de los modelos. La riqueza de los conjuntos de datos utilizados, abarcando desde indicadores macroeconómicos hasta variables sociales, refuerza la comprensión de que la estabilidad empresarial es influenciada por una multitud de factores interconectados. A través de estos análisis de residuos, se ha podido concluir que, si bien los modelos proporcionan insights significativos sobre algunas relaciones económicas, aún hay espacio para mejorar la modelización para capturar la complejidad real de cómo el entorno económico afecta las disoluciones empresariales en España.

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

Los modelos de regresión lineal múltiple analizados en este estudio se diseñaron para evaluar cómo diversos factores económicos influyen en las disoluciones empresariales en España. El primer modelo incorporó el Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) como variables independientes, mientras que el segundo modelo amplió el análisis incluyendo variables adicionales como la deuda pública, el déficit, el gasto público, los ingresos fiscales, el turismo y las reservas nacionales. Estos modelos permiten una exploración detallada de las relaciones entre indicadores económicos clave y la estabilidad empresarial, proporcionando insights sobre los factores que pueden contribuir a las disoluciones empresariales.

El análisis reveló que, aunque ciertas variables como el ICEA mostraron una relación estadísticamente significativa con las disoluciones empresariales, la capacidad general de los modelos para explicar la variabilidad en las disoluciones fue limitada, como lo indican los bajos valores de R-cuadrado. Este hallazgo sugiere que, mientras algunos factores económicos tienen un impacto en las disoluciones empresariales, existe una complejidad inherente en este fenómeno que no se captura completamente a través de las variables seleccionadas.

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 interpretación de las medidas de adecuación revela desafíos significativos en la modelización de las disoluciones empresariales utilizando variables macroeconómicas. Los

bajos valores de R-cuadrado y R-cuadrado ajustado indican que los modelos tienen una capacidad limitada para explicar la variabilidad en las disoluciones empresariales, lo que resalta la necesidad de considerar factores adicionales o interacciones complejas que pueden estar influyendo en este fenómeno. Además, la presencia de multicolinealidad, especialmente en el segundo modelo, sugiere que algunas variables independientes están altamente correlacionadas entre sí, lo que puede distorsionar las estimaciones de los coeficientes y complicar la interpretación de los resultados.

La evaluación de las métricas de error, como el RMSE y el MAE, proporciona una medida cuantitativa del error de predicción del modelo. Aunque estas métricas indican que hay una discrepancia significativa entre los valores observados y los predichos por el modelo, también ofrecen una dirección para futuras mejoras. Específicamente, sugieren la importancia de revisar la selección de variables, considerar la inclusión de términos de interacción o variables no lineales, y explorar modelos alternativos que puedan capturar mejor la complejidad de las disoluciones empresariales.

En conclusión, los resultados obtenidos de los modelos de regresión lineal múltiple subrayan la complejidad de modelar las disoluciones empresariales y la influencia de los factores económicos en este proceso. Aunque se identificaron algunas relaciones significativas, la capacidad limitada de los modelos para explicar la variabilidad en las disoluciones empresariales destaca la necesidad de enfoques analíticos más sofisticados y la consideración de una gama más amplia de factores que pueden afectar la estabilidad empresarial en el contexto económico de España.

14. Conclusiones y Recomendaciones

14.1 Conclusiones

Este Trabajo de Fin de Grado ha abordado la compleja relación entre diversos factores económicos y las disoluciones empresariales en España, utilizando modelos de regresión lineal simple y múltiple para analizar cómo variables como el PIB, el IPC, y otros indicadores macroeconómicos afectan la estabilidad empresarial. A través de este análisis, se buscó proporcionar una comprensión más profunda de las dinámicas económicas que afectan a las empresas en el contexto económico actual.

Los resultados obtenidos de los modelos de regresión han revelado una capacidad limitada para explicar la variabilidad en las disoluciones empresariales a través de los

indicadores macroeconómicos seleccionados. Esto sugiere que las disoluciones empresariales son influenciadas por una gama más amplia y compleja de factores, más allá de los indicadores económicos generales. A continuación, se presentan conclusiones clave derivadas del análisis:

Influencia del Clima Empresarial: Uno de los hallazgos más significativos es la influencia del clima empresarial, medido a través del indicador ICEA, en las disoluciones empresariales. Este resultado sugiere que la confianza empresarial y las expectativas económicas juegan un papel crucial en las decisiones de disolución, reestructuración o fusión de empresas. Es posible que, en períodos de alta confianza empresarial, algunas empresas opten por disolverse voluntariamente o fusionarse para reorientar recursos hacia oportunidades más prometedoras, reflejando una dinámica de mercado activa y orientada al futuro.

Complejidad de las Disoluciones Empresariales: La variabilidad limitada explicada por los modelos indica que las disoluciones empresariales son el resultado de un conjunto complejo de factores, incluyendo no solo condiciones macroeconómicas, sino también factores sectoriales específicos, cambios en la legislación, innovaciones tecnológicas, y dinámicas de competencia. Por ejemplo, sectores altamente innovadores pueden experimentar tasas más altas de disolución como parte de un proceso de "destrucción creativa", donde las empresas obsoletas son reemplazadas por nuevas entidades más innovadoras y eficientes.

Impacto de Factores Temporales y Externos: Los modelos también sugieren la influencia de factores temporales y externos en las disoluciones empresariales. Eventos económicos globales, como crisis financieras, cambios en la política comercial internacional, o pandemias, pueden tener efectos profundos y a menudo impredecibles en la estabilidad empresarial. Estos eventos pueden acelerar las disoluciones empresariales al alterar drásticamente las condiciones del mercado y la viabilidad de ciertos modelos de negocio.

Diversidad Regional y Sectorial: La investigación subraya la importancia de considerar la diversidad regional y sectorial dentro de España. Las comunidades autónomas y los sectores económicos pueden experimentar dinámicas de disolución muy diferentes en función de factores locales, como el ecosistema empresarial, la disponibilidad de financiación, la infraestructura, y el apoyo gubernamental. Este aspecto resalta la necesidad de políticas y estrategias diferenciadas que aborden las particularidades regionales y sectoriales.

En resumen, las conclusiones de este estudio destacan la complejidad en la variabilidad de las disoluciones empresariales en España. Aunque los indicadores

macroeconómicos proporcionan cierta visión sobre las tendencias generales, es evidente que las disoluciones empresariales están influenciadas por una constelación de factores económicos, sociales, tecnológicos y políticos. Reconocer esta complejidad es fundamental para desarrollar estrategias efectivas que promuevan la estabilidad y el crecimiento empresarial en el cambiante panorama económico de España.

14.2 Recomendaciones

En el ámbito de la investigación sobre las disoluciones empresariales en España, los estudios actuales han proporcionado una comprensión inicial significativa, destacando la influencia de diversos factores económicos. Sin embargo, la complejidad y la naturaleza multifacética de este fenómeno demandan un enfoque más sofisticado y detallado para capturar con precisión las dinámicas subyacentes. La exploración de modelos analíticos avanzados emerge como una recomendación primordial para futuras investigaciones. Modelos de regresión no lineal, análisis de series temporales y técnicas de machine learning, como árboles de decisión y redes neuronales, presentan una capacidad prometedora para manejar las interacciones complejas y las no linealidades entre múltiples predictores, ofreciendo así una herramienta más robusta para el análisis de las disoluciones empresariales.

La inclusión de variables adicionales en los modelos constituirá otra área crítica para la investigación futura. Variables como indicadores de confianza empresarial, datos sobre startups y emprendimiento, indicadores de innovación, y factores sectoriales específicos podrían proporcionar una visión más holística de las fuerzas que impulsan las disoluciones empresariales. Estas variables permitirían una evaluación más detallada de cómo diferentes aspectos del entorno económico y empresarial contribuyen a la estabilidad y sostenibilidad de las empresas.

Además, se sugiere la realización de análisis segmentados por sector de actividad y tamaño de empresa, lo cual podría revelar diferencias significativas en las tasas de disolución y ofrecer insights más detallados sobre las vulnerabilidades y resiliencias sectoriales. Estos análisis segmentados ayudarían a identificar patrones específicos y a formular estrategias dirigidas a sectores o tipos de empresas particulares.

Los estudios longitudinales y comparativos también podrían representar un enfoque valioso para comprender las tendencias temporales y las variaciones regionales o internacionales en las disoluciones empresariales. Estos estudios podrían iluminar cómo los

cambios en el entorno económico y político afectan la dinámica empresarial a lo largo del tiempo y entre diferentes contextos.

La integración de perspectivas cualitativas, mediante estudios de caso o entrevistas con empresarios y expertos, enriquecería la comprensión de las causas y consecuencias de las disoluciones empresariales desde una perspectiva más profunda y matizada. Esta aproximación cualitativa complementaría los hallazgos cuantitativos, ofreciendo una visión integral de los factores que influyen en las decisiones de disolución.

Finalmente, la evaluación de políticas públicas y programas de apoyo empresarial en relación con su impacto en las tasas de disolución podría ofrecer recomendaciones prácticas para el diseño de intervenciones efectivas. Este enfoque orientado a la política permitiría identificar estrategias basadas en evidencia para fomentar un entorno empresarial resiliente y propicio para el crecimiento sostenible.

En conjunto, estas recomendaciones buscan guiar a futuras investigaciones hacia un análisis más exhaustivo y refinado de las disoluciones empresariales en España. Al adoptar estos enfoques, los investigadores podrán desarrollar un entendimiento más completo de este fenómeno complejo, contribuyendo así a la formulación de estrategias y políticas que apoyen la estabilidad y el desarrollo empresarial en el país.

15. Resumen Narrativo

Este trabajo explora las dinámicas detrás de las disoluciones empresariales en España, un tema de relevancia tanto económica como social. Se emplearon técnicas avanzadas de análisis de datos para identificar patrones, correlaciones y causas potenciales de disoluciones empresariales, apoyándose en un enfoque multidimensional que incluye variables económicas, sociales y de mercado.

La investigación comenzó con la selección y preparación cuidadosa de los datos, seguida por el desarrollo y aplicación de modelos analíticos. Estos modelos, centrados en regresiones lineales simples y múltiples, buscaban aclarar las relaciones entre la creación de empresas, el capital desembolsado y variables macroeconómicas clave como el PIB, el IPC o la Deuda Pública Española. A través de una meticulosa evaluación, se reveló la complejidad de las interacciones entre estos factores y su impacto en la estabilidad empresarial.

Los resultados obtenidos sugieren que, aunque ciertos factores económicos como el clima empresarial tienen un impacto significativo en las disoluciones, la relación es compleja y multifacética. Las disoluciones empresariales en España se ven influenciadas por una combinación de variables, incluidas aquellas que reflejan la salud económica general del país y otras más específicas del entorno empresarial

Este trabajo contribuye al conocimiento académico sobre las disoluciones empresariales en España y proporciona herramientas analíticas que pueden aplicarse en futuras investigaciones y en la práctica empresarial. Las conclusiones y recomendaciones ofrecen insights valiosos para fomentar un entorno económico más estable y resiliente, destacando la importancia de un enfoque integral que considere tanto variables económicas amplias como factores específicos del clima empresarial.

En última instancia, este estudio subraya la necesidad de abordar las disoluciones empresariales desde múltiples perspectivas para comprender completamente las dinámicas en juego. Se recomienda la aplicación de modelos analíticos más sofisticados y la inclusión de variables adicionales en investigaciones futuras para capturar con mayor precisión la complejidad de las disoluciones empresariales en el contexto económico y social de España.