



Universidad  
Francisco de Vitoria  
**UFV** Madrid

---

# 1ª Parte del Trabajo de Fin de Grado

## Ingeniería del Dato

### **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

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.

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 recopilieron 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.

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 y manteniendo las consideraciones éticas y de privacidad. 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.

## Índice del Documento

---

<b>1. Introducción .....</b>	<b>6</b>
<b>2. Origen de los Datos.....</b>	<b>6</b>
<b>3. Descripción de los Datos .....</b>	<b>8</b>
3.1. Número total de observaciones y variables en los datasets (limpias).....	8
3.2. Descripción del tipo de datos de cada variable .....	10
3.3. Formato de los datos en bruto y cualquier transformación aplicada. ....	12
3.3.1) Base Datos 1:.....	12
3.3.2) Base Datos 2:.....	13
3.3.3) Base Datos 4:.....	14
3.3.4) Base Datos 5:.....	15
3.3.5) Base Datos 6 y 7: .....	16
3.3.6) Base Datos 8:.....	17
3.4. Periodicidad y temporalidad de los datos. ....	17
<b>4. Justificación de la Elección de Variables .....</b>	<b>18</b>
4.1. Explicación de la selección de variables.....	18
4.2. Relación con los objetivos del TFG. ....	19
<b>5. Análisis Exploratorio de Datos .....</b>	<b>19</b>
5.1. Descripción de medidas de tendencia central y dispersión para las variables cuantitativas. ....	19
5.1.2 Medidas de tendencia central.....	19
5.1.3 Medidas de dispersión .....	21
5.2. Frecuencias y proporciones para las variables categóricas. ....	23
<b>6. Gráficos Descriptivos .....</b>	<b>24</b>
6.1. Gráficos sobre variables y relaciones. ....	24
6.2. Patrones, tendencias o correlaciones observadas en los gráficos. ....	26

<b>7. Almacenamiento de Datos .....</b>	<b>28</b>
<b>8. Resumen Narrativo.....</b>	<b>29</b>

## 1. Introducción

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 recopilieron 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.

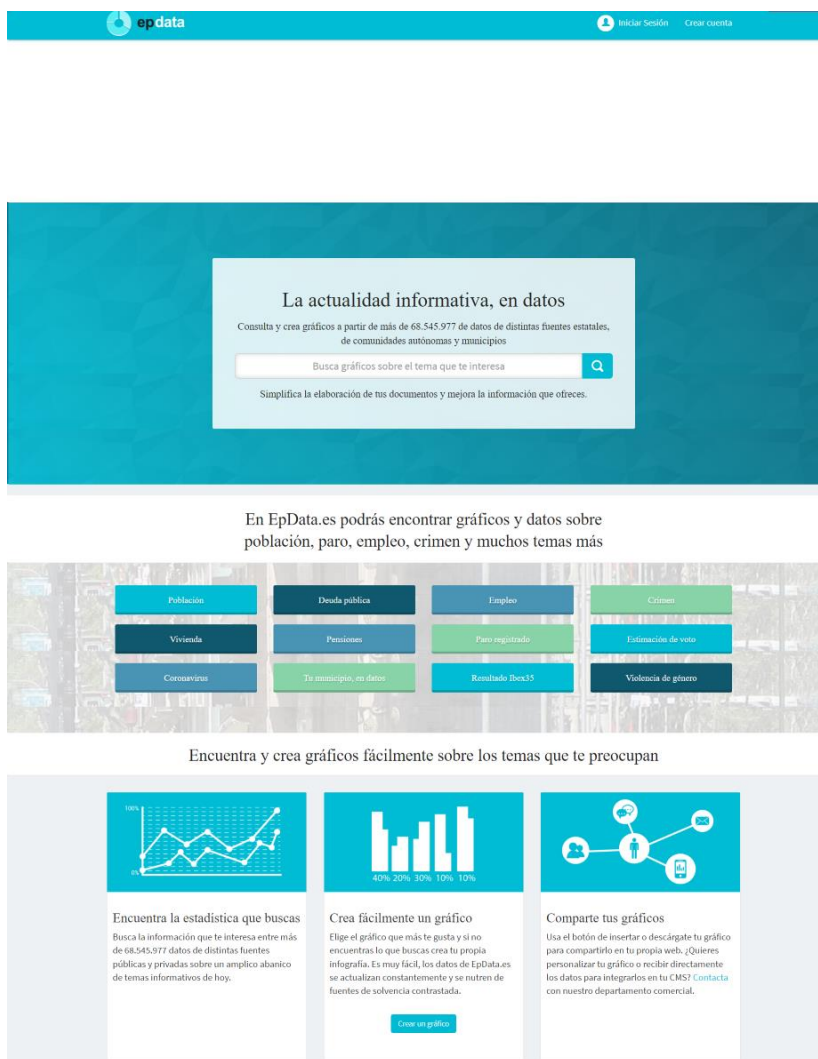
El INE, como organismo oficial en España, proporciona una amplia gama de datos estadísticos que abarcan diversos aspectos socioeconómicos y demográficos del país. La información obtenida es crucial para comprender las tendencias y patrones a nivel nacional, ofreciendo una perspectiva integral y actualizada. 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.	Índice de las filas.
Comunidad Autónoma	String	Nombre de las 19 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 posibles: "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 de la comunidad autónoma correspondiente.	Según el año indicado.

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 posibles: "Constituidas", "Disueltas".
Actividad económica	String	Actividad económica según la CNAE-2009.	
Año	Integer	Año de la observación, 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 19 Comunidades Autónomas de España con prefijo numérico.	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 19 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

<b>Valor Indicador</b>	124,36	126,5	132
------------------------	--------	-------	-----

**Tabla 5.** Medidas de tendencia central de la quinta base de datos.

<b>Nombre</b>	<b>Media</b>	<b>Mediana</b>	<b>Moda</b>
<b>Año</b>	2021,5	2021,5	2021
<b>Número de ERTES</b>	3934,27	1777	17

**Tabla 6.** Medidas de tendencia central de la quinta base de datos.

<b>Nombre</b>	<b>Media</b>	<b>Mediana</b>	<b>Moda</b>
<b>Afiliados a la Seguridad Social</b>	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.

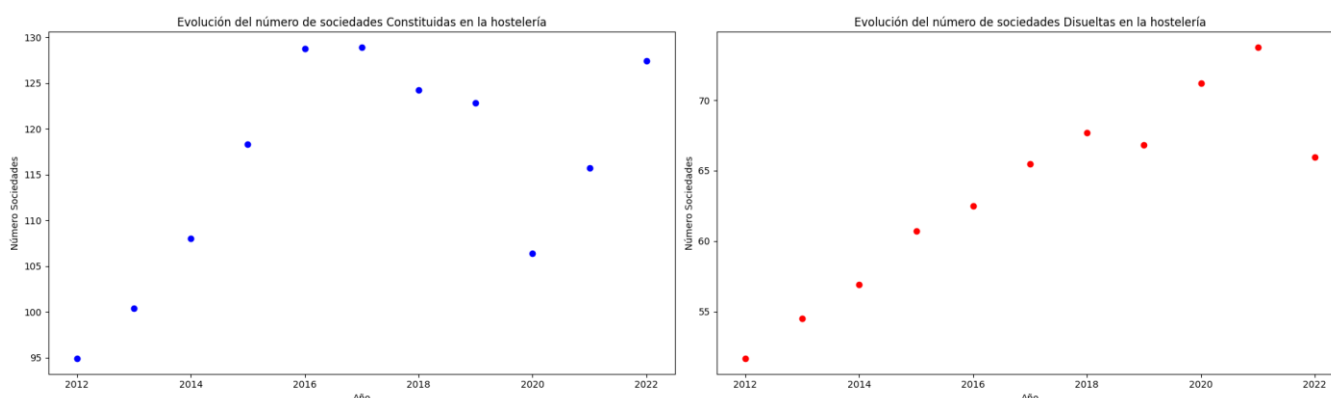
Este fenómeno se debe a la naturaleza del estudio, que se enfoca en términos totales y anuales, recopilando información específica para cada año sin variaciones intraanuales en la clasificación de las variables mencionadas. Dado que el análisis se realiza a nivel agregado por año, cada variable categórica 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.

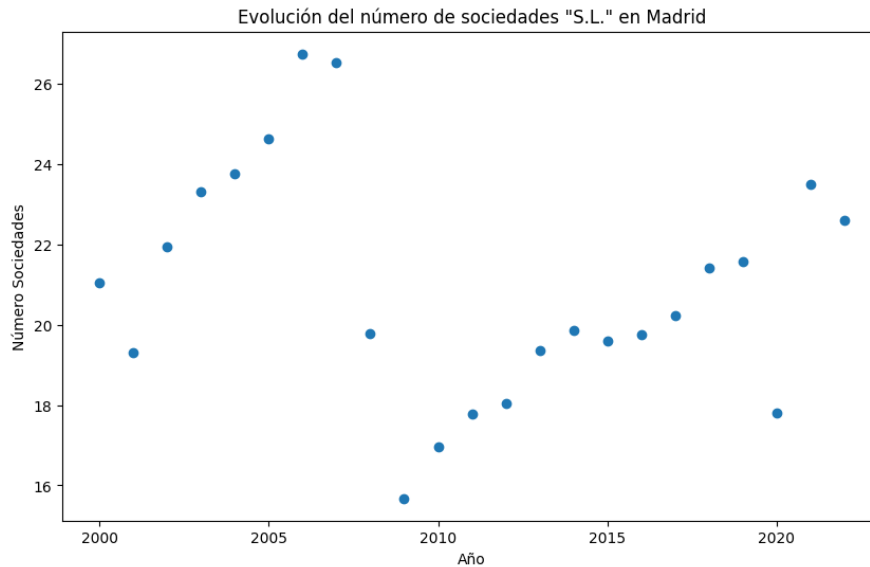
En la elaboración de este Trabajo de Fin de Grado, se adopta un enfoque detallado y minucioso para analizar el entramado empresarial en España. Para facilitar este análisis, se recurre a la utilización de gráficos descriptivos, herramientas visuales que permiten una interpretación clara y directa de 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.



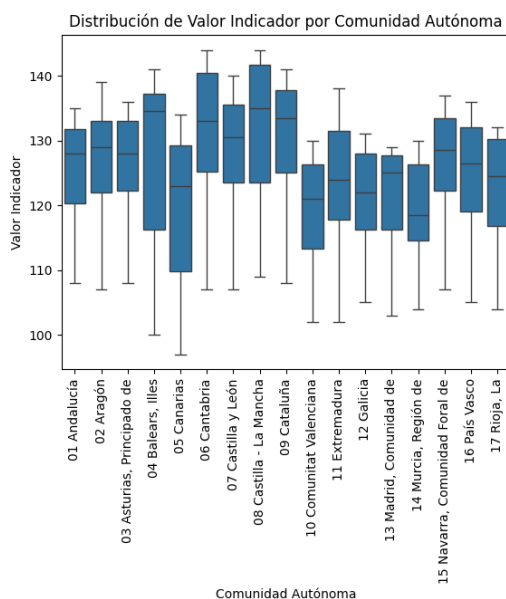
**Ilustración 3.** Captura de pantalla de los diagramas de dispersión correspondientes a las evoluciones de las sociedades Constituidas y disueltas en el sector de la hostelería desde el año 2012 al 2022.





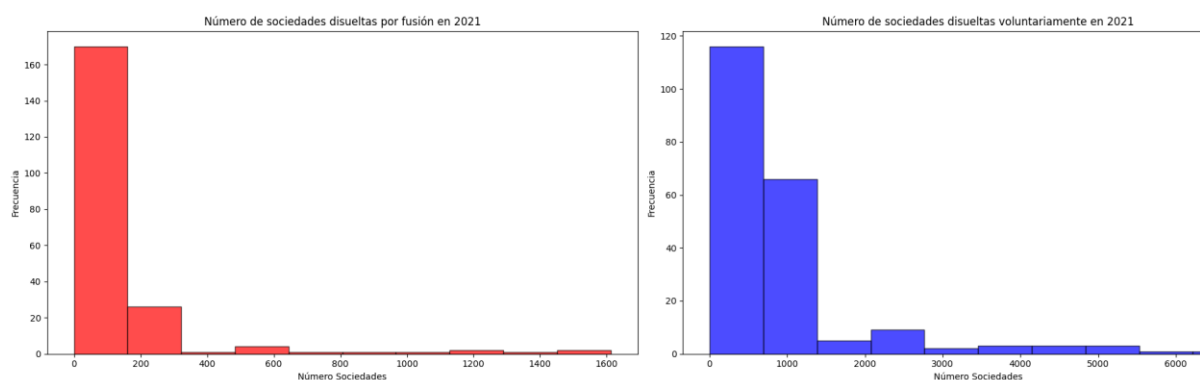
**Ilustración 4.** Captura de pantalla del diagrama de dispersión correspondiente a la evolución de sociedades de forma jurídica "S.L." constituidas en Madrid desde el año 2000 al 2022.

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



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

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



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

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

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

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

El análisis detallado en la 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.

## **8. Resumen Narrativo**

El desarrollo de este Trabajo de Fin de Grado ha permitido realizar un análisis integral en el tejido empresarial español. A través de un meticuloso proceso ETL, se han limpiado, transformado y almacenado datos provenientes de fuentes fiables como el INE y epdata, abarcando desde el año 2000 hasta 2022. Este proceso ha sido esencial para garantizar la calidad y la precisión del análisis posterior.

La selección de variables ha sido crucial para alinear el estudio con los objetivos específicos del TFG, permitiendo evaluar la tasa de creación de empresas, las transformaciones sectoriales, el impacto de las medidas gubernamentales y la resiliencia empresarial. Las medidas de tendencia central y dispersión han revelado patrones y tendencias, junto con el análisis de frecuencias y proporciones que ha mostrado como no cualquier tipo de análisis sirve para sacar conclusiones o información relevante. Además, los gráficos descriptivos han facilitado la visualización de estas dinámicas.

Los hallazgos obtenidos subrayan la complejidad del entorno empresarial español en el contexto de la pandemia, evidenciando cambios significativos en la creación de empresas y en la estructura sectorial.