

# Práctica 1

## 1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

El dataset seleccionado recoge información mensual del número de afiliaciones a la Seguridad Social por régimen (régimen general, autónomos y total) y por sectores de actividad (total, agricultura, industria, construcción, servicios y no clasificables) para toda España y para la comunidad de Catalunya. Las series también muestran las variaciones interanuales.

Esta información se consulta en el Instituto de Estadística de Catalunya y la fuente original de los datos es el Ministerio de Trabajo, Migraciones y Seguridad social.

El Instituto de Estadística de Cataluña (Idescat) es el organismo oficial de estadística de Cataluña cuya misión es proveer información estadística relevante y de alta calidad, con independencia profesional, y coordinar el Sistema estadístico de Cataluña, con el objetivo de contribuir a la toma de decisiones, la investigación y la mejora de las políticas públicas. La información de los datos de afiliación a la Seguridad Social forma parte de los indicadores de coyuntura económica, en concreto de trabajo, recogidos por Idescat.

## 2. Definir un título para el dataset. Elegir un título que sea descriptivo.

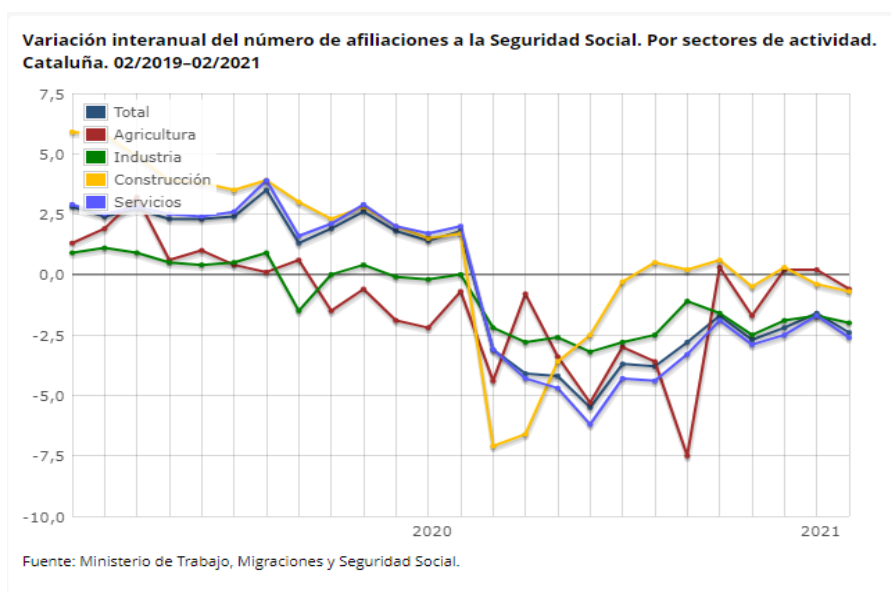
Afiliaciones a la Seguridad Social por tipos de régimen y sectores de actividad.

## 3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

El conjunto de datos generado en esta práctica recoge información sobre el número de afiliaciones a la Seguridad Social y variación interanual, por régimen y sector, en España y en Catalunya. Estos datos se recogen mensualmente y en el momento actual se dispone de datos desde 01/1994 hasta 02/2021, es decir, más de 27 años.

## 4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.

En la siguiente gráfica, se puede observar la evolución temporal (2019-2020) de la variación interanual del número de afiliaciones a la Seguridad Social en Catalunya, según diferentes los sectores: agricultura, industria, construcción, servicios y el total del conjunto.



## 5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y como se ha recogido.

El dataset está constituido por 7 columnas que se describen a continuación:

- **Año:** Año en formato (YYYY). Tipo numérico (entero).
- **Mes:** Mes en formato (MM). Tipo numérico (entero).
- **Región:** Localización a la que hacen referencia los datos ("Catalunya" o "España"). Tipo texto.
- **Régimen.** Tipo de régimen al que hacen referencia los datos ("Régimen general y minería del carbón", "Autónomos" o "Total"). Tipo texto.
- **Sector.** Tipo de sector al que hacen referencia los datos ("Total", "Agricultura", "Industria", "Construcción", "Servicios" y "No clasificables")
- **Valor.** Número de afiliaciones a la Seguridad Social (en miles) para la fecha, región, régimen y sector correspondientes. Tipo numérico (decimal).
- **Variación\_interanual.** Representa, en porcentaje, la tasa de variación interanual del último mes sobre el mismo mes del año anterior. Tipo numérico (decimal).

Los datos se actualizan de forma mensual.

**Última actualización:** 10 de marzo de 2021. **Próxima actualización:** 8 de abril de 2021.

## 6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.

Los datos han sido recolectados de la web del Instituto de Estadística de Catalunya (<https://www.idescat.cat/indicadors/?id=conj&n=10222&lang=es>), que a su vez, según citan obtienen los datos del Ministerio de Trabajo, Migraciones y Seguridad social.

Desde Idescat recomiendan consultar el enlace siguiente, en el que podemos encontrar análisis, estudios y recursos relacionados con el proyecto que hemos escogido.

<https://observatoritreball.gencat.cat/ca/inici/>

## 7. Inspiración. Explique por qué es interesante este conjunto de datos y que preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Se ha seleccionado este dataset ya que en el contexto actual de pandemia, uno de los temas que más preocupan son las repercusiones económicas de la misma. Un buen indicador de en qué medida se van recuperando los puestos de empleo es monitorizar temporalmente el número de afiliaciones a la Seguridad Social. Además, ya que la información en este dataset aparece segregada por sectores y régimen, proporciona información muy interesante acerca de qué sectores son los más afectados o los que presentan más dificultades para la recuperación.

En el enlace siguiente vemos un análisis similar al que ofrece nuestro proyecto pero basado en el balance de contratación del año 2020 en Cataluña. Como son datos relacionados, las conclusiones a las que podemos llegar usando nuestro dataset irá de la mano con las que se llegan en el informe.

<https://observatoritreball.gencat.cat/web/.content/generic/documents/treball/estudis/contractacio/2020/arxiu/Balanc-contractacio-2020.pdf>

## 8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

Se ha seleccionado la licencia **"Released Under CC BY-NC-SA 4.0 License"** ya que permite que se haga uso de los datos de forma no comercial respetando al creador.

De esta forma se ha de reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios.

El motivo de su selección es respetar la política de Idescat, de donde se han tomado los datos que se recogen en el dataset.

Los contenidos y elementos gráficos que constan en el sitio web [www.idescat.cat](http://www.idescat.cat) son de titularidad exclusiva del Idescat o de otros organismos que han autorizado su uso y difusión al Idescat.

Esta web, y todo el material que contiene, está protegida por los derechos de propiedad intelectual e industrial. Por lo tanto, cualquier uso que se haga de sus contenidos está sometido a la normativa reguladora de estos derechos.

El Idescat no permite el uso de logotipos, marcas u otros signos distintivos de esta web y que son propiedad del Idescat, sin que haya tenido conocimiento ni lo haya autorizado previamente.

Con carácter general, el Idescat permite la reutilización de la información que difunde en su sitio web, siempre que se cite la fuente de procedencia, que no se altere ni se desnaturalice el contenido ni el sentido de la información y que se mencione la fecha de la última actualización de la información. Para citar la fuente de procedencia de la información, se puede hacer de la siguiente manera:

- Fuente: Idescat. (Si no hay un tratamiento de los datos.)
- Fuente: elaboración propia a partir de datos del Idescat. (Si los datos se tratan posteriormente.)

El Idescat puede limitar la reutilización de la información por la tutela de otros bienes jurídicos prioritarios, como la protección de los datos personales, la intimidad o los derechos de propiedad intelectual de terceros.

En ningún caso se puede indicar ni sugerir que el Idescat participa, patrocina o apoya la utilización de la información que haga el usuario y sin previo aviso.

#### 9. Código. Adjuntar el Código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código con el que se ha generado el dataset (en Python) con el nombre "web\_scrapper.py" se encuentra accesible en el siguiente repositorio de GitHub:

<https://github.com/taniapvidal/SS-afiliaciones-scrapper>

#### 10. Dataset. publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El dataset se ha generado se encuentra accesible con el nombre "Afiliaciones a la seguridad Social.csv" en el siguiente repositorio de GitHub:

<https://github.com/taniapvidal/SS-afiliaciones-scrapper>

#### Contribuciones de los integrantes

Contribuciones	Firma
Investigación previa	TPV, JSF
Redacción de las respuestas	TPV, JSF
Desarrollo código	TPV, JSF