Logotipo

Descripción generada automáticamente

**UNIVERSIDAD EUROPEA DE MADRID  
ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO**

**MÁSTER UNIVERSITARIO EN**

**ANÁLISIS DE DATOS MASIVOS (BIG DATA)**

TRABAJO FIN DE MÁSTER

**Análisis y Comparativa de las Listas de Sancionados de la Unión Europea y Estados Unidos**

**Alberto Alonso Chana**

**CURSO 2021-2022**

**TÍTULO**: Análisis y Comparativa de las Listas de Sancionados de la Unión Europea y Estados Unidos

**AUTOR**: Alberto Alonso Chana

**TITULACIÓN**: Máster Universitario en Análisis de Datos Masivos (Big Data)

**DIRECTOR DEL PROYECTO**: Jorge Luis Hita y José Javier Ruiz Cobo

**FECHA**: septiembre de 2022

# RESUMEN

El resumen tiene entre 150-250 palabras. Resumir consiste en ofrecer información exhaustiva sobre cómo, dónde, cuándo y por qué se aplica el proyecto. Se realiza al finalizar el trabajo.

El resumen

* Establece el tema tratado en el proyecto
* Resume los problemas planteados y las aportaciones más importantes del proyecto
* Resume las conclusiones más importantes del trabajo realizado

El resumen NO

* Da una información genérica
* Explica porque se realiza ese trabajo
* Se refiere a datos aportados en el texto del proyecto.

**Palabras clave:** hasta un máximo de 6

# ABSTRACT

Resumen en inglés.

**Key words:** Palabras clave en inglés

**AGRADECIMIENTOS**

En ocasiones se incluye este apartado para agradecer a aquellos que han ofrecido su ayuda en el desarrollo del trabajo, ya sea técnica o de otro tipo.

Índice

El índice debe ir paginado (o sea, debe incluir los números de página). Ha de ser un listado de cada una de las secciones o capítulos. Intenta que no haya más de 3 niveles en las secciones. Recuerda que lo mejor es que el procesador de textos que uses lo haga de forma automática.

[RESUMEN 4](#_Toc109668642)

[ABSTRACT 4](#_Toc109668643)

[Capítulo 1. INTRODUCCIÓN 10](#_Toc109668644)

[1.1 Planteamiento del problema 10](#_Toc109668645)

[1.2 Objetivos del proyecto 10](#_Toc109668646)

[1.3 Estructura del proyecto 10](#_Toc109668647)

[Capítulo 2. Estructura y títulos 12](#_Toc109668648)

[2.1 Estilo de redacción 12](#_Toc109668649)

[2.2 Ejemplo de estilo 12](#_Toc109668650)

[2.3 Separaciones 12](#_Toc109668651)

[2.4 Márgenes y sangrados 12](#_Toc109668652)

[2.5 Numeración de páginas 13](#_Toc109668653)

[2.6 Numeración de figuras 13](#_Toc109668654)

[2.7 Numeración de tablas 13](#_Toc109668655)

[2.8 Encabezado de páginas 13](#_Toc109668656)

[Capítulo 3. ESTRUCTURA DEL PROYECTO 14](#_Toc109668657)

[3.1 Memoria 14](#_Toc109668658)

[3.2 Anexos 15](#_Toc109668659)

[3.2.1 Condiciones Generales 15](#_Toc109668660)

[3.2.2 Especificaciones de Materiales y Equipos 15](#_Toc109668661)

[3.2.3 Especificaciones de Ejecución 15](#_Toc109668662)

[3.2.4 Estado de mediciones 16](#_Toc109668663)

[3.3 Presupuesto 16](#_Toc109668664)

[3.3.1 Precios unitarios 16](#_Toc109668665)

[3.3.2 Presupuesto Mediciones 16](#_Toc109668666)

[3.3.3 Estudios con entidad propia 16](#_Toc109668667)

[Capítulo 4. NORMAS DE ESTILO 17](#_Toc109668668)

[Capítulo 5. CONCLUSIONES Y FUTURAS LÍNEAS DE TRABAJO 18](#_Toc109668669)

[ANEXOS 19](#_Toc109668670)

[PRESUPUESTO 20](#_Toc109668671)

[BIBLIOGRAFÍA 21](#_Toc109668672)

Índice de Figuras

Este índice ha de existir tan sólo en el caso de que tengas más de 4 apoyos visuales (gráficos, figuras, ilustraciones, fotos, etc.) en tu memoria del proyecto.

Recuerda esta sección es optativa.

Índice de Tablas

Este índice ha de existir tan sólo en el caso de que tengas más de 4 tablas de datos en tu memoria del proyecto.

Recuerda esta sección es optativa.

# INTRODUCCIÓN

En la INTRODUCCIÓN se plantean las cuestiones generales que se van a abordar en el trabajo, los aspectos más importantes e, incluso, la estructura que va a tener el trabajo. Ocupa entre 1,5 y 3 páginas habitualmente.

La introducción exige:

* Focalizar muy bien el tema, los objetivos, el contexto y el problema de estudio, planteando a grandes rasgos qué es lo que se pretende resolver, estudiar o proponer.
* Justificar la necesidad, idoneidad y pertinencia del trabajo a partir de ciertas necesidades detectadas en un determinado ámbito. Es decir, explicar por qué vale la pena elaborar dicho trabajo.
* Mostrar, sucintamente, qué se sabe hasta el momento sobre la temática y que se ha hecho al respecto.
* Prever estrategias clave en la metodología utilizada o en los pasos a seguir para obtener los datos y planificar el proceso de investigación.
* Evidenciar, someramente cuáles han sido los resultados alcanzados en función de los objetivos planteados.

Puede ser algo parecido a:

## Descripción del Problema

Las sanciones internacionales son medidas coercitivas que los Gobiernos y las entidades supranacionales aplican contra Estados, empresas o individuos que suponen una amenaza para la seguridad global. Abarcan restricciones económicas, diplomáticas, comerciales, militares e incluso deportivas, y sus objetivos son debilitar y reducir la capacidad de maniobra de estos actores sin hacer uso de la fuerza.

## Contexto Histórico

Las sanciones se han utilizado como herramienta diplomática desde la Antigüedad. La primera conocida fue el embargo comercial de Atenas a Mégara en el 432 a. C. Posteriormente, durante el siglo XIX, el uso de sanciones se intensificó a partir del bloqueo continental que Napoleón decretó en 1806 con el fin de excluir cualquier intercambio con el Reino Unido, su rival por el control de Europa. No obstante, el auge definitivo de las sanciones internacionales llegó con el final de la Primera Guerra Mundial. Durante este período, la Sociedad de Naciones (SDN), predecesora de la ONU, fue fundamental en el intento de imponer medidas para boicotear a los países agresores. En los años noventa, la invasión de Kuwait por el Irak de Sadam Huseín y las guerras yugoslavas llevaron a un nuevo repunte de las sanciones internacionales [1].

El uso de las sanciones se ha consolidado en la comunidad internacional en las últimas décadas. Su adopción ha trascendido de la ONU a la **Unión Europea**, en su Política Exterior y de Seguridad Común, o a la Organización para la Seguridad y Cooperación Europea (OSCE), entre otras. A nivel estatal, **Estados Unidos** es el país que más ha utilizado este mecanismo, a través de leyes y órdenes ejecutivas. La Ley para la Democracia Cubana de 1992, por ejemplo, reguló el embargo a la isla.

## Actualidad

Ya en 2022, Tanto Washington como Bruselas decidieron apostar por estas medidas para responder a la invasión rusa de Ucrania. Las sanciones, que incluyen el cierre del espacio aéreo europeo a Rusia o su desconexión parcial del sistema Swift de comunicaciones financieras, pretenden aislar la economía rusa. Las restricciones también alcanzan al ámbito deportivo, como el veto a Rusia en el Mundial de fútbol de Catar.

## Motivación

Para un mejor control y seguimiento de las entidades sancionadas, tanto la Unión Europea como Estados Unidos elaboran unas listas que ponen a disposición de la ciudadanía a través de internet. Las personas, grupos y compañías que aparecen en estas listas tienen sus activos bloqueados y generalmente se prohíbe a las empresas europeas y estadounidenses tratar con ellos.

Los avances en análisis y procesamiento de datos suponen un nuevo paradigma en la forma en que estas sanciones son aplicadas:

* Mediante algoritmos de Machine Learning de aprendizaje no supervisado, se pueden conocer patrones ocultos de los datos que de otra manera pasarían desapercibidos. El aprendizaje supervisado permite conocer la información faltante de algunos registros a partir de los registros en los que sí existe esa información.
* Las herramientas enfocadas a extraer información a partir de los datos mediante la generación de dashboards son cada vez más avanzadas. Estas son capaces de ingerir grandes cantidades de datos y mantenerlos siempre actualizados, además de presentarlos de forma dinámica e interconectada.

Estos avances benefician a las empresas susceptibles de tener entre sus clientes a entidades sancionadas. Existen diversos motivos por los que una compañía querría conocer cuáles son sus clientes sancionados: Mantener la reputación, evitar el pago de multas, cumplir con la normativa para la no financiación del terrorismo, etc. [2]

Este trabajo tiene como objetivo tanto realizar un proceso de investigación de los datos en busca de relaciones e informaciones desconocidas, como de generar valor a las empresas que pudieran beneficiarse del seguimiento de los sancionados mediante dashboards generados con herramientas de visualización accesibles y fáciles de usar. Esto con el objetivo de que cualquier miembro de la compañía en cualquier departamento, incluido el equipo directivo, pueda acceder fácilmente a la información e interpretarla sin necesidad de ser experto en el campo.

Se puede ampliar el dashbpard con un presupuesto de lo que se ahorra la compañía invirtiendo en la herramienta para detectar sancionados entre sus clientes (buscar información sobre leyes y multas respectivas a este campo)

## Objetivos del proyecto (pendiente)

¿Cuáles son los objetivos que persigo? Es bueno desglosarlo en un objetivo general del que se derivan varios objetivos específicos

## Estructura del proyecto (dejar para el final)

Una buena organización de tu memoria ayuda a su lectura y refleja mejor la metodología de trabajo usada. Normalmente consta de objetivos, metodología, resultados y, conclusiones y futuras líneas de trabajo. Cuenta aquí más o menos en dos líneas por cada capítulo, el contenido de dicho capítulo.

NO OLVIDES:

* Numerar las páginas.
* Numerar las figuras
* Numerar las tablas
* Justificar los márgenes.
* Destacar los apartados dentro de cada capítulo
* Cuida el FORMATO: usa siempre el mismo tipo de letra, el mismo tamaño para el texto, el mismo tamaño para los apartados, el mismo tamaño para los títulos de los capítulos,…
* Si te dan una guía de estilo con tamaños de letra, interlineados, márgenes etc, **respétala**
* Destaca lo que creas oportuno utilizando las **negritas** y las *cursivas*.

# Estructura y títulos

Piensa antes que nada qué estructura va a tener el trabajo, para poder dividir la información por capítulos.

Resulta interesante además ordenar cada capítulo, dividiendo cada uno en diferentes apartados. Recuerda que el nombre que le des a cada sección también es importante. Deben ser lo más descriptivos posibles

Piensa también en la conveniencia de incluir ilustraciones (figuras, circuitos, mapas, paisajes,…). Eso sí, la MISIÓN DE LA ILUSTRACIÓN ES ILUSTRAR, o sea, ejemplificar, ni no es mejor no ponerlas.

Quizás te pueda ayudar el manual de escritura técnica de Joe Miró (2010)

## Estilo de redacción

Deberá mantenerse el mismo estilo de redacción a lo largo de todo el texto. Se recomienda utilizar un estilo impersonal (“Se ha desarrollado…”), o en primera persona plural (“Hemos desarrollado un…”).

Cuando sea necesario el uso de siglas o acrónimos, es recomendable aclarar su significado la primera vez que aparecen, siempre que no sean suficientemente conocidos

## Ejemplo de estilo

Todos los textos deberán estar ordenados en capítulos, divididos en apartados y subapartados, y éstos, en párrafos. Se recomienda que estén escritos en formato A4 y mecanografiado a espacio y medio (pero puede que tengáis una guía de estilo, en ese caso se deben seguir las recomendaciones de la misma).

Los capítulos llevarán numeración correlativa, así como apartados y subapartados de cada capítulo. Capítulos, apartados y subapartados utilizarán el sistema de numeración decimal.

El tipo de letra a utilizar será unas de las tipografías de letra comúnmente utilizadas (Times New Roman, Arial,…), en tamaño 11 o 12.

Recuerda que quieres que se lea fácilmente, por tanto:

* el uso de viñetas se agradece,
* usa los títulos de secciones y sub-secciones para dividir la información
* si hay fórmulas utiliza un interlineado de 1,15 como mínimo.

## Separaciones

Los capítulos iniciarán página. La separación, entre nombres de los capítulos y los nombres de los apartados y subapartados y su primer párrafo, será de dos espacios.

La separación entre párrafos será de dos espacios.

## Márgenes y sangrados

Sigue la guía de estilo que te den. Si vas a encuadernar el trabajo tenlo en cuenta a la hora de establecer los márgenes.

Entre cada número o letra y el inicio del texto se dejarán dos espacios.

## Numeración de páginas

Todas las páginas estarán numeradas correlativamente, indicando en la parte central inferior el número de la página.

## Numeración de figuras

Todas las figuras estarán numeradas correlativamente, indicando en la parte central inferior de la misma el número de la figura y el título que le dais a la misma, en un tamaño generalmente de 10 puntos.

Si son gráficos recuerda que el título que le des tiene que ser breve y descriptivo (referirse a los ejes). Por ejemplo “Velocidad en función del Tiempo”. En general, el título que le des a una figura debe ser tal que no necesite explicaciones. Si la figura ha sido extraída de alguna otra fuente, recuerda que has de referenciarlo.

## Numeración de tablas

Todas las tablas estarán numeradas correlativamente, indicando en la parte central inferior el número de la tabla y su título en un tamaño de 10 puntos.

## Encabezado de páginas

Todas las páginas estarán encabezadas con la siguiente estructura:

* Logo de la empresa en la que trabajáis (la universidad en vuestro caso) a la derecha. Se puede incluir además en una segunda línea el nombre del departamento (asignatura en vuestro caso).
* Título del proyecto o del trabajo y en una segunda línea Nombre de los autores, en la izquierda.

# ESTRUCTURA DEL PROYECTO

Más o menos debe contener los siguientes apartados, aunque algunos de ellos no tienen por qué estar siempre:

* Introducción
* Memoria técnica:
  + Objetivos
  + Metodología empleada /Marco teórico/Antecedentes
  + Resultados
* Conclusiones y futuras líneas de trabajo
* Referencias
* Anexos
* Planos
* Pliego de Condiciones
* Presupuesto
* Estudios con entidad propia

Si se trata de un proyecto de ingeniería real recuerda que ha de ser las normas UNE (AENOR, 2010; UNE 157001, 2002)

## Memoria

La Memoria deberá incluir las referencias Bibliográficas y Legales que sean pertinentes.

Las Referencias deben colocarse al final del Proyecto.

En la memoria del Proyecto deben recogerse al menos los siguientes aspectos:

* Descripción concreta y detallada de las tareas, trabajos desarrollados y departamentos de la entidad a los que ha estado asignado (si procede).
* Relación de los problemas planteados y el procedimiento seguido para su resolución (metodología seguida).
* Si procede se puede incluir aquí un apartado con el marco teórico.

La Memoria deberá ser claramente comprensible, no sólo por profesionales especialistas sino por terceros, en particular por el cliente, especialmente en lo que se refiere a los objetivos del Proyecto, las alternativas estudiadas, sus ventajas e inconvenientes, y las razones que han conducido a la solución elegida.

**Puede** incluir los siguientes apartados:

* **Objeto**: En este capítulo de la Memoria se indicará el objetivo del Proyecto y su justificación.
* **Alcance:** En este capítulo se indicará el ámbito de aplicación del Proyecto.
* **Antecedentes:** Aquí se enumerarán todos aquellos aspectos necesarios para la comprensión de las alternativas estudiadas, y la solución final adoptada.
* **Disposiciones legales y normas aplicadas:** En este apartado se contemplará el conjunto de disposiciones legales (leyes, reglamentos, ordenanzas, etc.) y las normas de obligado cumplimiento que se han tenido en cuenta para la realización del Proyecto.
* **Programas de cálculo:** Se indicará la relación de programas, modelos u otras herramientas utilizadas para desarrollar los diversos cálculos del Proyecto.
* **Requisitos de diseño:** Aquí se describirán las bases y datos de partida que quizás derivan de:
  + La legislación, reglamentación y normativas aplicables.
  + El emplazamiento, y su entorno socio-económico y ambiental.
  + Los estudios realizados encaminados a la definición de la solución adoptada.
  + Los interfaces con otros sistemas, elementos externos al Proyecto u otros que condicionan las soluciones técnicas del mismo.
  + Etc.
* **Análisis de las soluciones:** Se indicarán las distintas alternativas estudiadas, qué caminos se han seguido para llegar a ellas, ventajas e inconvenientes de cada una y cuál es la solución finalmente elegida y justificación.
* **Resultados finales:** Aquí se describirá el producto, obra, instalación, servicio o software (soporte lógico) según la solución elegida, indicando cuáles son sus características definitorias y haciendo referencia a los planos y otros elementos del Proyecto que lo definen.

## Anexos

En ciertos proyectos pueden ser necesarios estudios especiales, en forma de Anexos, sobre aspectos concretos del mismo, como los de seguridad, impacto ambiental, etc.

**Planos**

Los planos deben contener toda la información necesaria para ejecutar la etapa constructiva sin necesidad de consultar otros documentos, por lo que su contenido y grado de detalles deben acoplarse a esta necesidad.

Los planos se ordenaran de modo que primero aparecen los planos generales del proyecto, después los aspectos específicos, y finalmente, los planos de detalle que sean necesarios.

Pueden ser circuitos por ejemplo.

**Pliego de condiciones**

Las especificaciones de materiales y equipos y las de ejecución establecen cómo y con qué hay que realizar el proyecto. El Pliego de Condiciones consta de las siguientes partes:

### Condiciones Generales

Descripción general del contenido del proyecto, sus características principales y los aspectos legales y administrativos a tener en cuenta.

### Especificaciones de Materiales y Equipos

En este apartado aparecerán definidos todos los equipos, materiales, máquinas, etc. que formen parte del proyecto.

### Especificaciones de Ejecución

En este apartado se definirá la ejecución material del proyecto a partir de los materiales especificados en el punto anterior.

### Estado de mediciones

Permite definir las unidades de producto, obra, instalación o servicio. Servirá de base para realizar el presupuesto.

## Presupuesto

El Presupuesto, que supone la evaluación económica total del proyecto, consta de dos partes:

### Precios unitarios

En este apartado se recogerán los precios de todos los materiales y equipos que van a integrar el proyecto.

### Presupuesto Mediciones

Resulta de multiplicar las cantidades recogidas en el apartado de mediciones por los precios unitarios de materiales y equipos.

### Estudios con entidad propia

Incluye documentos exigidos por exigencias legales, tales como Prevención de Riesgos Laborales e Impacto Ambiental.

# Memoria Técnica

Los ficheros, scripts, dashboards y demás documentos empleados en este trabajo, incluida esta memoria, están publicados en el siguiente repositorio de Github:

<https://github.com/Chanalber/TFM.git>

A continuación, se exponen las tareas realizadas durante el desarrollo del trabajo y la forma en que fueron abordadas. En esta sección se tratarán los problemas, cambios e imprevistos en la consecución de estos objetivos para finalmente presentar los resultados a los que se ha llegado.

## Extracción de Datos

El notebook de Jupyter para código Python ha sido la aplicación utilizada en la carga de los datos de las listas de sancionados. Se generó un script de nombre *01\_extraer\_datos.ipynb* en la ubicación *TFM/Código y datos/* con los siguientes propósitos generales:

* Acceder a las listas de sancionados de EEUU y la UE vía URL y cargar los datos en Python transformándolos a formato de DataFrame.
* Almacenar los datos en crudo en formato .csv en la ubicación *TFM/Código y datos/UE\_EEUU\_data\_crudo* para tenerlos localizados en caso de que se quieran revisar. Cada vez que este proceso es ejecutado, de borran los archivos antiguos del directorio local y se vuelven a cargar desde la web.
* Selección de los datos de los DataFrame que sean de interés. Habrá dos tablas para la lista de EEUU relacionadas por clave y una para la UE. Los datos seleccionados se centran en tres dimensiones de la información de sancionados: **Nombre, ubicación y tipo de sancionado.**
* Pasar los DataFrame con los datos cargados y seleccionados a formato .csv y almacenarlos en la ubicación *TFM/Código y datos/datos\_entrada\_script\_02.* Esto se hace para que los datos que vayan a transformarse con el script 02 (del que se hablará más adelante) puedan ser tomados directamente desde esta ubicación siempre que se quiera, sin ser necesario ejecutar antes el script 01. De este modo, solo será necesario ejecutar el script *01\_extraer\_datos.ipynb* antes del script 02en caso de que se quiera tener la última versión de los datos de entrada al script 02*.*

Para cargar en el script de Python las url’s con las listas de sancionados de EEUU se utiliza wget.download() dentro de un bucle for que se ejecuta una vez por cada url. Cada vez que este proceso es ejecutado, se borran los archivos antiguos en el directorio local y se vuelven a cargar desde la web [3]. Al principio se intentó cargar la lista de EEUU desde un único .xml con toda la información de los sancionados. Pero esto no fue posible porque, al pasarlo a DataFrame, los elementos anidados del .xml no permitían su correcta representación. Por ello, se optó por acceder a la misma información, pero dividida en varios .csv, que se encontraban en la misma página web que el .xml. De entre estos archivos .csv, se encuentran dos que llevan información relevante para nuestro estudio: *sdn.csv* y *add.csv.* Estos archivos se relacionan por una columna llamada ‘uid’ por lo que, tras ponerlos en formato DataFrame y seleccionar las columnas relevantes para nuestro estudio, se puede hacer un join a las tablas. Las columnas de los .csv no vienen etiquetadas con un nombre. Es necesario buscar el nombre correspondiente a cada columna en el archivo .xml con la información general y escribirlo.

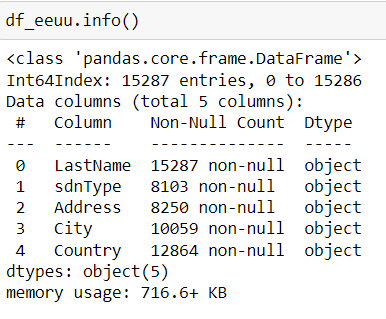


Figura 1: Información de metadatos de la tabla df\_eeuu.

Una vez se tiene un único DataFrame para la lista de EEUU, al que se llama *df\_eeuu*, se estudian los valores de cada variable en busca de posibles incidencias (fig. 1). Se encuentra qué para la columna con la información del tipo de sancionado, llamada ‘sdnType’, hay demasiados valores nulos. El tipo de sancionado puede deducirse a partir del nombre del sancionado (columna de nombre ‘LastName’). Para ello, se crea un bucle que devuelve el nombre de todos los sancionados cuya columna ‘sdnType’ es nula. Una vez hecho, se observa que la mayoría de sancionados se corresponden a una empresa o asociación (fig. 2).

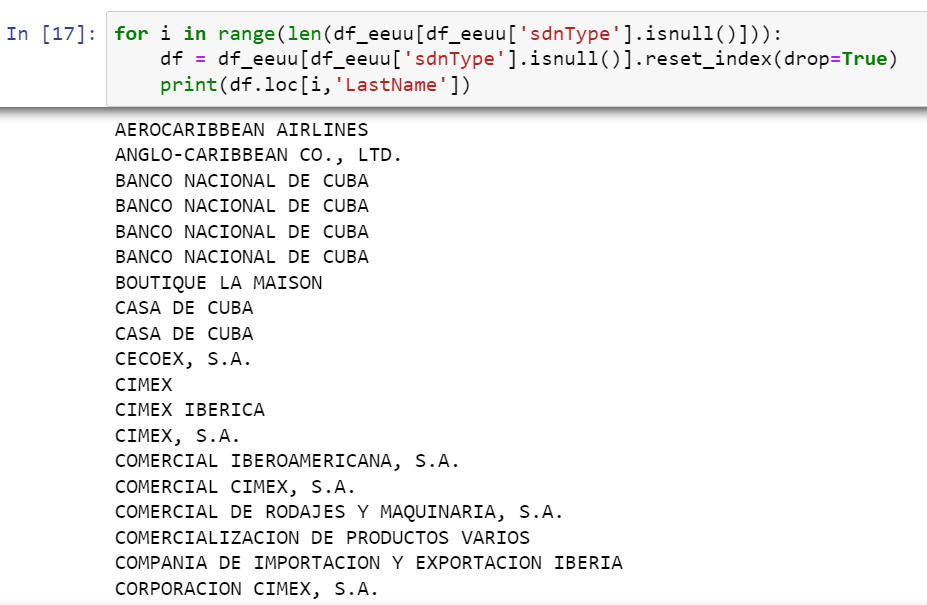


Figura 2: Bucle for para conocer los valores de la columna 'LastName' en los que 'sdnType' tiene valor nulo.

No obstante, existe un número minoritario de nombres en los que no queda claro que tipo de sancionado puede ser. Se optará por pasar todos los valores nulos a valor ‘enterprise’. Haciendo esto se introduce un cierto error en la clasificación de los registros por tipo de sancionado. Pero a cambio se recupera una gran cantidad de información que antes aparecía como nula.

La columna ‘Address’ también presenta gran cantidad de valores nulos. Como los nulos de esta columna no van a tener especial relevancia en los estudios posteriores, podemos dejarla como está.

Para las listas de la UE, la petición a la URL se hace con un token. No sirve el código utilizado con las listas de Estados Unidos porque no acepta el método wget.download(). Es necesario utilizar un proceso alternativo que no almacena la tabla en local, si no que genera directamente en Python un archivo con los datos [4]. Una vez se tiene este archivo, le aplicamos diversas modificaciones para pasarlo a un DataFrame al que se llama *df\_ue*. Cuando se tienen los datos en crudo de la lista de la UE como DataFrame, se pasan como .csv a la ubicación local *TFM/Código y datos/UE\_EEUU\_data\_crudo.* Cada vez que se ejecute el proceso, se elimina el .csv con los datos antiguos y se cargan los nuevos datos*.*

El DataFrame con los datos da la Unión Europea presenta muchas columnas. Por ello, el mejor método para identificar cuáles de estas son relevantes para nuestro estudio es dividir el DataFrame en grupos de 15 columnas y analizar separadamente las columnas de estas subtablas. La información sobre las variables seleccionadas tras realizar este análisis se encuentra en la figura 3.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

*Figura 3: Información de metadatos de la tabla df\_ue.*

Los valores correspondientes a la ubicación (‘Address\_City’ y ‘Address\_CountryDescription’) no aparecen en el mismo registro que el nombre correspondiente a dicha ubicación. La información de nombres viene dada en la columna ‘NameAlias\_WholeName’. Para poder identificar los valores para la ubicación y moverlos al registro con su correspondiente nombre, es necesario agrupar por la columna 'Entity\_EU\_ReferenceNumber'. Esta columna engloba bajo el mismo identificador a un grupo de sancionados que comparten la misma ubicación. Lo que se pretende agrupando por cada valor único 'Entity\_EU\_ReferenceNumber' es generar un nuevo DataFrame, llamado *df*, en el que a cada número de referencia se le asocie una ubicación descrita en términos de país y ciudad (fig. 4). Una vez definido *df*, se hace outer join con *df\_ue* para trasladar su información de la ubicación a la tabla de sancionados principal. Hecho esto se elimina la columna 'Entity\_EU\_ReferenceNumber', pues ya no es necesaria.

Tabla

Descripción generada automáticamente

Figura 4: Tabla con la información de ubicación para cada número de referencia. Se observa que algunos de los sancionados de la UE no van a presentar información sobre su ubicación o va a estar incompleta.

En el momento en que fue realizado este análisis sobre la lista de la UE, la columna con la información sobre el tipo de entidad, 'Entity\_SubjectType\_ClassificationCode', presentaba 16 valores únicos contando el valor nulo. Son demasiados, ya que solo interesan rasgos más generales del tipo de sancionado como si es una empresa, una persona o una embarcación. Representamos 'Entity\_SubjectType\_ClassificationCode' en un gráfico de barras (fig. 5):

Gráfico

Descripción generada automáticamente

Figura 5: Gráfico de barras de la cardinalidad para cada valor de la columna 'Entity\_SubjectType\_ClassificationCode' de la tabla df\_ue

Como se aprecia en la figura 5, para casi todos los registros el valor no nulo de 'Entity\_SubjectType\_ClassificationCode' es 'person' o 'enterprise'. Por lo tanto, se pueden eliminar los otros registros sin que se pierda demasiada información.

Tras finalizar el proceso de extracción, los DataFrames *df\_ue* y *df\_eeuu* se llevan, en formato .csv, a la ubicación *TFM/Código y datos/datos\_entrada\_script\_02.* Que es el fichero desde donde se toman los datos de entrada al script *02\_transformar\_datos.ipynb* para continuar con la depuración de las tablas.

## Transformación de Datos

Para la transformación de los valores de las tablas obtenidas en el script 01 también se utiliza un notebook de Jupyter. Estas transformaciones buscan hacer los valores lo más adecuados posibles para su empleo en la herramienta de visualización y en los modelos de Machine Learning. El script encargado de la transformación de los datos es *02\_transformar\_datos.ipynb.* Este script toma las tablas *df\_ue.csv* y *df\_eeuu.csv* en *TFM/Código y datos/datos\_entrada\_script\_02* y las carga como DataFrame. Una vez aplicadas las transformaciones, los DataFrames resultantes se almacenan en formato .csv en *TFM/Código y datos/datos\_entrada\_script\_03,* que es la ubicación desde la que accede el script 03 a los datos tranformados.

Para las primeras transformaciones sobre los valores de *df\_eeuu* se crea un nuevo DataFrame llamado *df\_eeuu\_t1.* El 't1' en el nombre implica que esta es la primera transformación hecha sobre los datos tras ser cargados. Este DataFrame:

* Presenta una nueva variable llamada 'Origin' de valor 'EEUU' para identificar cual es la lista de origen.
* Se le han eliminado los registros con 'LastName' nulo. Ya que no podemos identificar un sancionado si no tiene un nombre.
* En caso de existir, se eliminan los registros duplicados. Para que se eliminen los registros duplicados pero que se diferencian por las mayúsculas y minúsculas, se ponen todas las variables en mayúsculas.
* Se cambia el nombre de algunas columnas para darles un nombre más conciso (fig. 6). Este será el mismo nombre que se dé a las variables de la lista de la UE para que las filas puedan concatenarse sin problemas. Esta concatenación de las tablas se hará en el script 03, del que se hablará más adelante.

A continuación, se crea una nueva tabla llamada *df\_eeuu\_t2* para aplicar la segunda transformación sobre los datos: en algunos valores de la columna 'City' aparecen códigos y números que no interesan (fig. 6), pues solo interesa el nombre de la ciudad. Se aplica una transformación mediante una expresión regular para eliminar estos códigos [5]. Ha sido necesario cambiar los valores perdidos en 'City' a cadena de caracteres para poder escanearlos por la expresión regular. Una vez aplicada la expresión, se han vuelto a poner como valor ‘Not a Number’ (NaN). Se ha empleado el método re.sub() dentro de un bucle for que selecciona cada registro de la columna ‘City’ y le pasa la expresión regular. Los nuevos valores para ‘City’ se van almacenando en una lista con cada ejecución del bucle. Finalmente, esta lista se sustituye por la columna ‘City’ antigua en *df\_eeuu\_t2.* Las pruebas para crear la expresión regular que eliminase las cadenas que no son de interés se han hecho en la página:

<https://regex101.com/r/3vtVKW/1>

En esta página se puede ver un cuadro con las cadenas de los valores de ‘City’. Las subcadenas que han sido eliminadas aparecen resaltadas a color. La expresión regular que selecciona estas subcadenas aparece encima del cuadro.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Figura 6: Tabla de Estados Unidos tras aplicar la primera transformación. Se aprecia que algunos valores de la columna ‘City’ presentan códigos tras el nombre de la ciudad

Texto

Descripción generada automáticamente

Figura 7: Información de metadatos de la tabla transformada final df\_eeuu\_t2

Las primeras transformaciones sobre los valores de *df\_eu* siguen una estructura similar a las de la tabla de Estados Unidos. Se crea un DataFrame llamado *df\_eu\_t1* sobre el que se aplican las siguientes transformaciones:

* Se añade una nueva variable llamada 'Origin' de valor 'UE' para identificar cual es la lista de origen.
* Se han eliminado los registros con 'NameAlias\_WholeName' nulo. Ya que no podemos identificar un sancionado si no tiene un nombre.
* Se eliminan los registros duplicados. Para que se eliminen los registros duplicados pero que se diferencian por las mayúsculas y minúsculas, se ponen todas las variables en mayúsculas.
* Se cambia el nombre de las columnas para darles un nombre más conciso (fig. 8). Este será el mismo nombre que se ha dado a las variables de la lista de EEUU para que puedan concatenarse sin problemas en el script 03.

Echando un vistazo a los valores de la columna ‘Name’ de *df\_eu\_t1* (fig. 8). Se observan dos problemas con sus valores:

1. Aparecen nombres en la columna ‘Name’ en idiomas que utilizan alfabetos distintos al inglés o al español, dificultando su comprensión. Sería útil poder traducir, en la medida de lo posible, estos nombres al inglés para que sean más comprensibles.
2. Hay registros muy parecidos salvo por pequeñas diferencias en los valores en la columna ‘Name’ que pueden considerarse duplicados. Esto puede deberse a que el sancionado escribió mal su nombre en algún formulario del que luego se pasó esta información a la lista de sancionados, o al uso de un alias por parte del sancionado. También puede ocurrir que, aunque los nombres se parezcan mucho, correspondan a distintos sancionados. En cualquier caso, como este trabajo no busca el seguimiento individual de cada sancionado, si no una visión más general de los tipos de sancionados y sus ubicaciones, consideraremos los nombres similares que tengan el resto de variables iguales como duplicados.

Tabla

Descripción generada automáticamente

Figura 8: Tabla de la Unión Europea tras aplicar la primera transformación. En la columna ‘Name’ aparecen nombres en distintos idiomas y algunos de ellos muy similares entre sí.

Para resolver la problemática del punto 1 se han probado varios módulos de Python, como goslate y googletrans, que hacen una petición a un servicio de traducción en línea. Ambos módulos daban problemas relacionados con la versión y el número límite de peticiones que se podían hacer. Se terminó por utilizar el módulo de deepl. Esta librería de Python da acceso a la API de DeepL. DeepL es un traductor en línea que utiliza tecnología basada en Inteligencia Artificial para obtener mejores resultados. Para poder utilizar DeepL desde Python, es necesario crear una cuenta gratuita que genere una API key que permita la autentificación del cliente. El registro de cuenta se hace desde la web oficial de la API DeepL:

<https://www.deepl.com/docs-api>

Se pasó un bucle for a los registros de la columna ‘Name’ para traducirlos y se llevó cada nombre traducido a una lista. Finalmente se sustituyeron los valores en la columna ‘Name’ en el DataFrame *df\_ue\_t1* por los valores de la lista y se renombró como *df\_ue\_t2.* Algunos de los valores, como los nombres en árabe, no pudieron ser traducidos. No obstante, el resultado general de la traducción es bueno.

-ANALIZAR LOS VALORES UNICOS POR PAIS Y CIUDAD Y ELIMINAR AQUELLOS Q NO TENGAN SENTIDO

-HACER DIAGRAMA DE GANTT

# NORMAS DE ESTILO

En la redacción de los textos incluidos en el proyecto se seguirán los siguientes criterios:

* Títulos directos y completos (ejemplo: “*Interrelación entre secciones del proceso de fabricación”*, en vez de: “*Proceso de fabricación. Interrelación entre secciones”*).
* Párrafos cortos.
* Oraciones directas y completas con el mínimo posible de oraciones intercaladas.
* Estilo impersonal y objetivo (ejemplo: *Posteriormente se estudiará*, en vez de: *Posteriormente estudiaremos*)

Hay que incluir además del índice un índice de figuras y un índice de tablas

¡Cuidado! NO SE PUEDE COPIAR NADA DE NINGÚN SITIO SIN MENCIONAR LA FUENTE, ESO ES **PLAGIO**

# CONCLUSIONES Y FUTURAS LÍNEAS DE TRABAJO

Es más que conveniente terminar un trabajo con unas conclusiones que sirvan de colofón al mismo. Caben en ellas tus impresiones, o destacar la importancia que tiene el tema, lo que has aprendido, o la trascendencia, etc.

Además las conclusiones tienen que referirse siempre a los objetivos generales y específicos empleados.

Es importante destacar en este apartado las líneas con las que se podría continuar el mismo.

# ANEXOS

Sirven para incluir documentación complementaria (planos, circuitos, códigos, especificaciones, hojas características, fichas explicativas, etc.)

Eso sí, cada comentario en una página independiente. Y ES QUE LOS ANEXOS SE ORGANIZAN ASÍ: CADA DOCUMENTO EN UNA PÁGINA INDEPENDIENTE.

# PRESUPUESTO

Si se trata de un proyecto siempre debe incluir este apartado.

El Presupuesto supone la evaluación económica total del proyecto.

No olvides que tu tiempo también vale dinero, no sólo hay que incluir el coste de los materiales empleados.

# BIBLIOGRAFÍA

En este apartado figurará el conjunto de libros, revistas u otros textos que el autor considere de interés para justificar las soluciones adoptadas en el Proyecto. **Cita todas las fuentes** que has utilizado como consulta para elaborar el trabajo.

Sigue el estilo de cita que te indiquen las normas de estilo y respétalo a los largo de todo el proyecto. Recuerda que has de citar todas las fuentes que hayas usado. Los estilos de cita más comunes son:

* ISO
* IEEE
* APPA
* Etc.

En ingeniería se suele usar el ISO o el IEEE.

Si puedes usar un gestor de citas bibliográficas te será más fácil. Si no tendrás que recurrir a las páginas web de las bibliotecas para saber como citar adecuadamente. Por ejemplo lo encuentras en:

<http://biblioteca.uem.es/es/aprendizaje-y-formacion/citas-bibliograficas-documentos>

Referencias usadas en este manual de estilo:

**AENOR. 2010.** AEN/CTN 157 - PROYECTOS. *Normas y Publicaciones.* [En línea] 2010. [Citado el: 25 de abril de 2013.] http://www.aenor.es/aenor/normas/ctn/fichactn.asp?codigonorm=AEN/CTN%20157.

**Miró Julià, José. 2010.** Recursos para aprender a escribir. [En línea] 2010. http://bioinfo.uib.es/~joemiro/RecEscr/manual.pdf.

**UNE 157001. 2002.** Criterios generales para la elaboración de proyectos. *Escuela Universitaria de Ingeniería de Vitoria.* [En línea] 2002.

# Bibliografía

[1] "¿Qué son las sanciones internacionales?" [Online]. Available: https://elordenmundial.com/que-son-sanciones-internacionales/ [Accessed: 2022]

[2]"Reporting de Sancionados"[Online]. Available: https://www.sas.com/content/dam/SAS/es\_es/doc/other1/sas-consulting-web.pdf [Accessed: 2022]

[3]"Overwrite existing files with Python's wget?"[Online]. Available: https://stackoverflow.com/questions/63226700/overwrite-existing-files-with-pythons-wget [Accessed: 2022]

[4]"¿Cómo descargar un archivo csv desde internet con python 3?"[Online]. Available: https://es.stackoverflow.com/questions/120084/c%C3%B3mo-descargar-un-archivo-csv-desde-internet-con-python-3 [Accessed: 2022]

[5]"Operaciones con expresiones regulares"[Online]. Available: https://docs.python.org/es/3/library/re.html [Accessed: 2022]

[Citado el: 25 de abril de 2013.] http://www.coiib.es/coiib/documentos/DocumentosContenidos/Gu%C3%ADa%20de%20elaboraci%C3%B3n%20de%20proyectos/2-Electricidad/5\_PNE\_157701\_Criterios.pdf.