



HERRAMIENTAS DE PROCESADO Y VISUALIZACIÓN DE DATOS

Edición 2021

Este documento ha sido elaborado en el marco de la Iniciativa Aporta (datos.gob.es), desarrollada por el Ministerio de Asuntos Económicos y Transformación Digital a través de la Entidad Pública Empresarial Red.es.

Aviso legal: Esta obra está sujeta a una licencia Atribución 4.0 de Creative Commons (CC BY 4.0). Está permitida su reproducción, distribución, comunicación pública y transformación para generar una obra derivada, sin ninguna restricción, siempre que se cite al titular de los derechos (Ministerio de Asuntos Económicos y Transformación Digital a través de la Entidad Pública Empresarial Red.es). La licencia completa se puede consultar en:

<https://creativecommons.org/licenses/by/4.0/>

CONTENIDOS

INTRODUCCIÓN	4
HERRAMIENTAS	5
1. HERRAMIENTAS DE WEB SCRAPING	6
1.1. ParseHub	6
1.2. Scrapy	7
1.3. Table capture	7
1.4. Función IMPORTHTML (Google Spreadsheet)	8
2. HERRAMIENTAS DE DEPURACIÓN DE DATOS	9
2.1. Open Refine (anterior Google Refine)	10
2.2. Talend Open Studio	11
3. HERRAMIENTAS DE CONVERSIÓN DE DATOS	12
3.1. Mr. Data Converter	12
3.2. Beautify Converters	12
3.3. Tabula	13
4. HERRAMIENTAS DE ANÁLISIS DE DATOS	14
4.1. <i>Herramientas para no programadores</i>	14
4.1.1. Weka	14
4.1.2. Knime	15
4.1.3. Orange	15
4.2. <i>Herramientas para programadores</i>	17
4.2.1. R (The R Project for statistical computing)	17
4.2.2. Python	18
4.2.3. GNU Octave	19
5. HERRAMIENTAS DE VISUALIZACIÓN DE DATOS	20
5.1. <i>Herramientas de visualización genéricas</i>	20
5.1.1. Kibana	20
5.1.2. Tableau Public	21
5.1.3. SpagoBI	22
5.1.4. Grafana	22
5.2. <i>Librerías y APIs de visualización</i>	24
5.2.1. Google Chart Tools	24
5.2.2. JavaScript InfoVis Toolkit	24
5.2.3. Data-Driven Documents (D3.js)	25
5.2.4. Matplotlib	26
5.2.5. Bokeh	26
5.3. <i>Herramientas de visualización geoespacial</i>	27
5.3.1. Carto	28
5.3.2. OpenLayers	29
5.3.3. OpenStreetMap	29
5.3.4. Geocoder y Geopy	30
5.3.5. GDAL	30
5.3.6. PROJ.4 y PROJ4.JS	31
6. HERRAMIENTAS DE ANÁLISIS DE REDES	32
6.1. Gephi	33
6.2. NodeXL	33
COMPARATIVA	35
CONCLUSIONES FINALES	37

Introducción

En los últimos años se ha producido una enorme proliferación de datos en bruto, que deben ser procesados y preparados en una forma comprensible para el usuario con el objetivo de extraer de ellos el máximo conocimiento. Todos estos datos sin procesar suelen ser difíciles de obtener, transformar y comprender, de ahí que se hayan desarrollado distintas herramientas de tratamiento y visualización que facilitan su interpretación.

El esquema clásico del tratamiento de datos, previo a la ejecución de cualquier análisis o visualización, establece que hay que prestar atención a las técnicas de adquisición de los datos y llevar a cabo un estudio de los mismos para asegurar que representan adecuadamente el universo de discurso. Una vez asegurados ambos, se puede someter al conjunto de datos en cuestión a diferentes [análisis de tipo descriptivo, predictivo o prescriptivo](#) y obtener los resultados y la visualización que mejor se ajuste a la información que se quiere transmitir. En la actualidad, la necesidad de procesar grandes y variadas cantidades de datos, es decir, *big data*, ha impulsado la oferta de herramientas de procesamiento y visualización, hacia un escenario donde la computación distribuida irrumpe para aportar soluciones a crecientes necesidades de ingesta de datos, almacenamiento y recuperación sobre bases de datos no relacionales, *edge computing* u otros requisitos de los proyectos actuales.

Para analizar y visualizar de manera eficiente los datos, es importante identificar las herramientas disponibles, conocer el alcance de su funcionalidad y valorar su aplicación en las tareas que implica la Ciencia de Datos. Existen muchas herramientas para estos fines, pero en numerosas ocasiones su uso puede conllevar un elevado coste. Por tanto, diferentes comunidades y empresas desarrollan herramientas de libre disposición para permitir que cualquier usuario tenga acceso a ellas.

Esta guía es una actualización del [Análisis de herramientas de procesado y visualización de datos](#) publicado en datos.gob.es en 2016, donde se describen diferentes herramientas útiles para abordar este tipo de tareas. Este documento actualiza la recopilación inicial de herramientas gratuitas o aquellas que en su modelo de comercialización incluyen alguna versión gratuita, para la extracción, procesamiento, análisis y visualización de datos.

Esta nueva edición desglosa una muestra de herramientas tomando como criterio fundamental la popularidad que les otorga la Comunidad de usuarios que aborda este tipo de actividades vinculadas con la Ciencia de Datos. En cualquier caso, se quedan fuera de esta selección una ingente cantidad de herramientas de excelente calidad que para muchos usuarios merecerán una posición preeminente en la relación que en esta compilación se expone. Somos conscientes de ello y el riesgo que supone la selección y publicación de un pequeño conjunto de herramientas, pero también pensamos que una guía con este alcance es un punto de referencia y orientación para aquellas personas que se inician en el mundo de la reutilización de datos abiertos.

Herramientas

A continuación, se presenta una selección de herramientas útiles para la extracción, procesamiento, análisis y visualización de datos desarrolladas en el ámbito del software abierto y que disponen de versiones gratuitas en sus diferentes modos de comercialización. Esta selección se presenta agrupada en base a su funcionalidad principal.

Funcionalidad principal	Herramienta
Web scraping	Parsehub
	Scrapy
	Table Capture
	Función IMPORTHTML
Depuración de datos	Open Refine
	Talend Open Studio
Conversión de datos	Mr. Data Converter
	Beautify Converter
	Tabula
Análisis de datos y Lenguajes de programación	Weka
	Knime
	Orange
	R
	Python
	GNU Octave
Visualización	Kibana
	Tableau Public
	SpagoBI
	Grafana
APIS y librerías de visualización	Google Chart Tools
	JavaScript InfoVis Toolkit
	D3.js
	Matplotlib
	Bokeh
Visualización y tratamiento de datos geoespaciales	Carto
	OpenLayers
	OpenStreetMap
	Geocoder
	Geopy
	GDAL
	PROJ.4 y PROJ.4S
Análisis de redes	Gephi
	NodeXL

1. Herramientas de web scraping

El [web scraping o extracción de datos de la web](#) es un conjunto de técnicas de programación con las cuales es posible extraer información o datos de la web, sea cual sea su presentación. Esta información se recopila y se exporta a un formato procesable por el usuario, como puede ser una hoja de cálculo o ficheros tipo texto. Esta acción se puede hacer manualmente, pero es preferible usar herramientas automatizadas para extraer datos de una web, ya que se ahorra tiempo y quebraderos de cabeza debido a que cada sitio web puede tener elementos constructivos distintos. La elección de la herramienta para realizar esta tarea depende de cuales sean las necesidades del usuario.

Cuando se utiliza este tipo de herramientas se deben tener en cuenta dos aspectos relevantes. En primer lugar, algunos sitios web pueden implementar medidas para bloquear *scrapers* y por otro lado, es importante saber que esta práctica puede rozar los límites de la legalidad, especialmente si los datos objetivo del web scraping contienen información personal sujeta a restricciones recogidas en el Reglamento General de Protección de Datos (RGDP). En todo caso, siempre hay que prestar atención a los términos recogidos en el aviso legal del sitio web al objeto de no vulnerar derechos de autor, de propiedad intelectual, el uso de marcas registradas o evitar cometer un acto de competencia desleal. En caso de duda, la solución está en obtener la autorización por parte del propietario del sitio web para evitar ir en contra de los términos de uso del sitio web.

1.1. ParseHub

Tipo: Aplicación de escritorio o aplicación web

Licencia: [MIT](#)

Coste: Versión gratuita / diferentes versiones de pago

Enlaces:

Sitio web: <https://www.parsehub.com/>

Tutoriales: <https://help.parsehub.com/hc/en-us/categories/202725858-Videos>

Esta herramienta permite extraer datos de cualquier sitio web, trabajando con una o varias URLs.

Algunas de las características que presenta esta herramienta son:

- Admite la recopilación de datos de sitios web que utilizan tecnologías [AJAX](#), [Javascript](#) o [cookies](#).
- Vista en tiempo real del sitio web, lo que facilita su manejo.
- Puede manejar diferentes elementos web, como mapas interactivos, calendarios, búsquedas en foros, comentarios anidados, imágenes, entre otros.
- Los resultados obtenidos pueden descargarse en formatos como Excel, CSV, JSON y API.

- Dispone de una versión gratuita y una versión comercial. La versión gratuita está limitada en algunos aspectos como la utilización de únicamente 200 páginas por ejecución, por lo que está recomendada para la realización de proyectos pequeños, donde no intervengan una gran cantidad de sitios web.

1.2. Scrapy

Tipo: Librería Python

Licencia: [BSD](#)

Coste: Librería gratuita

Enlaces:

Sitio web: <https://scrapy.org/>

Repositorio GitHub: <https://github.com/scrapy/scrapy>

Tutorial: <https://docs.scrapy.org/en/latest/intro/tutorial.html#intro-tutorial>

Scrapy es una librería de código abierto, diseñada para automatizar la extracción de datos de páginas web.

Esta herramienta presenta las siguientes características:

- Para su manejo es necesario disponer de conocimientos avanzados en [Python](#), lo que hace que sea una herramienta compleja de manejar.
- Tiene un soporte incorporado para extraer datos mediante expresiones XPath o CSS.
- Permite exportaciones en formatos JSON, CSV y XML.
- Flexible, rápida y fácilmente extensible.

1.3. Table capture

Tipo: Extensión de navegador

Licencia: [Términos del Servicio de cada Navegador](#)

Coste: Extensión gratuita

Enlaces:

Sitio web: <https://www.georgemike.com/tablecapture/>

Sitio de descarga: <https://chrome.google.com/webstore/detail/table-capture/>

Table Capture es una extensión para Chrome, Firefox y Microsoft Edge, que nos permite capturar tablas en HTML, tablas dinámicas, tablas de varias páginas y listas a nuestro

portapapeles, para poder trabajar con esos datos posteriormente en diferentes formatos como Excel, CSV o [Markdown](#).

1.4. Función *IMPORTHTML* (Google Spreadsheet)

Tipo: Función de Google Spreadsheet

Licencia: [Términos del Servicio de Google](#)

Coste: Función gratuita

Enlaces:

Sitio web: <https://support.google.com/docs/answer/3093339?hl=es>

Se trata de una fórmula aplicable sobre una hoja de cálculo creada con Google Spreadsheet que importa y procesa datos estructurados en XML, CSV o HTML.

Para esta función necesitamos:

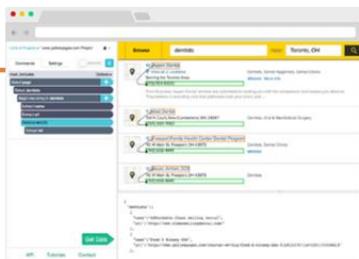
1. La URL del sitio del cuál queremos extraer los datos
2. Especificar si queremos un elemento de tipo lista ("list") o tipo tabla ("table")
3. Índice de tabla o lista importada

```
= IMPORTHTML ("URL"; "table"; 1)
```

Otras funciones útiles son, [IMPORTXML](#), [IMPORTFEED](#), o [IMPORTDATA](#) que permiten Importar datos de una URL en determinados formatos.

Herramientas de Web Scraping

ParseHub



Scrapy

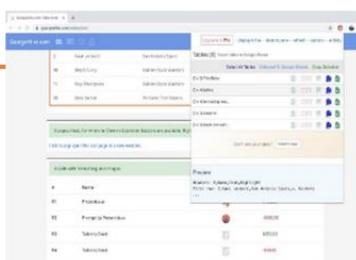
```
Terminal
$ pip install scrapy
$ cat > myspider.py <<EOF
import scrapy

class BlogSpider(scrapy.Spider):
    name = 'blogspider'
    start_urls = ['http://blog.scrapinghub.com']

    def parse(self, response):
        for title in response.css('.post-header h2'):
            yield {'title': title.css('a::text').extract_first()}

        for next_page in response.css('div.prev-post > a'):
            yield response.follow(next_page, self.parse)
EOF
$ scrapy runspider myspider.py
```

Table Capture



2. Herramientas de depuración de datos

La [depuración de los datos](#) es el conjunto de tareas de preparación de datos extraídos de un origen de datos y su transformación a un formato concreto, adecuadamente estructurado y libre de errores, normalmente denominada *vista minable* de datos. Este proceso mejora la calidad de los datos, ya que, entre otras acciones, elimina datos duplicados, incorrectos, incompletos o mal formateados.

La depuración de datos juega un papel muy importante en el procesamiento de los datos. Los datos sin procesar provienen de fuentes dispares, con diferentes estructuras y formatos y las herramientas de depuración limpian esos datos entrantes, estandarizándolos y formateándolos antes de introducirlos en el sistema de destino. Es necesario mantener una calidad y coherencia de los mismos durante el proceso de depuración, para que la estructura se mantenga y no haya duplicidad en el destino.

2.1. Open Refine (anterior Google Refine)

Tipo: Aplicación Java

Licencia: [BSD-3](#)

Coste: Aplicación gratuita

Enlaces:

Sitio web: <https://openrefine.org/>

Repositorio GitHub: <https://github.com/OpenRefine>

Documentación: <https://github.com/OpenRefine/OpenRefine/wiki/Doc>

Es una herramienta gratuita que se creó con el objetivo de entender la estructura y calidad de los datos, pudiendo organizarlos, transformarlos y mejorarlos corrigiendo determinados tipos de errores comunes en ellos.

Admite diferentes fuentes de datos: archivos locales, URLs, enlaces a documentos de Google Docs o incluso fragmentos de datos almacenados en el *clipboard* del sistema operativo. Además, soporta un número significativo de formatos, entre los que destaca, JSON, XML, XLS, RDF, CSV y HTML. Igualmente, la exportación de los datos depurados puede realizarse en CSV, XML, JSON, ODF, o [RDF mediante la extensión adecuada](#), entre otros formatos.

Open Refine posee varios aspectos a destacar:

- Limpieza de datos. Entre otras tareas, permite modificar el contenido de las celdas y unificar campos, de manera manual o de forma automática aplicando determinadas sugerencias que proporciona OpenRefine.
- Transformación de datos. Mediante el lenguaje GREL es posible aplicar un importante número de funciones que permite, entre otras tareas: dividir columnas, crear nuevas columnas a partir de los valores de otras, o aplicar expresiones regulares para extraer subcadenas de datos.
- Enriquecer datos mediante el uso de servicios externos, para entre otras funciones, georeferenciar información o vincular datos de la DBpedia generando así datos enlazados.

2.2. Talend Open Studio

Tipo: Aplicación de escritorio

Licencia: [Apache V2](#)

Coste: Aplicación gratuita

Enlaces:

Sitio web: <https://www.talend.com/es/products/talend-open-studio/>

Repositorio GitHub: <https://github.com/Talend/tdi-studio-se>

Documentación: <https://www.talend.com/es/resources/>

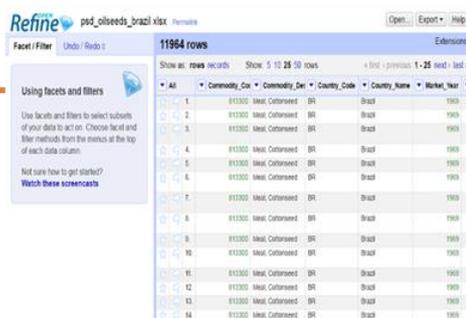
Talend Open Studio es un conjunto de herramientas [ETL \(Extraer, Transformar y Cargar\)](#) de código abierto, diseñada para extraer información de diversos orígenes de datos, normalizar, depurar y transformar en el formato adecuado para ser utilizados posteriormente por otras herramientas o aplicaciones.

Algunas de las características de esta aplicación son:

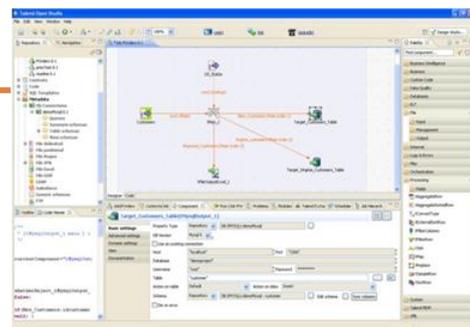
- Se basa en la programación por componentes, consiste en unir procesos con funcionalidades diversas (migración, transformación, eliminación de datos, entre otros), mediante flujos de entrada y salida.
- Genera código estándar utilizando motores optimizados (Java o Perl).
- Permite conexiones estandarizadas a bases de datos como MySQL, PostgreSQL o SQLite.

Herramientas de depuración de datos

Open Refine



Talend Open Studio



3. Herramientas de conversión de datos

Aunque las herramientas anteriormente descritas incluyen habitualmente funcionalidades para conversión de formatos de datos, existen igualmente varias que realizan esta misión de forma exclusiva facilitando al usuario la realización de esta rutinaria tarea. La conversión se lleva a cabo para que los datos puedan ser leídos por herramientas de análisis o visualización que solo aceptan determinados formatos.

3.1. Mr. Data Converter

Tipo: Aplicación web

Licencia: [MIT](#)

Coste: Aplicación gratuita

Enlaces:

Sitio web: <https://shancarter.github.io/mr-data-converter/>

Repositorio GitHub: <https://github.com/shancarter/Mr-Data-Converter>

Aplicación web que acepta datos en formato CSV o Excel (XLS o XLSX) y los convierte en XML, JSON o HTML, entre otros formatos. Una peculiaridad de esta herramienta es que no tendremos que seleccionar y subir ningún archivo, ya que los datos se incorporan copiando y pegando sobre la interfaz web de la aplicación. Para exportar debemos hacerlo de la misma manera, copiando y pegando el código generado.

3.2. Beautify Converters

Tipo: Aplicación web

Licencia: [MIT](#)

Coste: Aplicación gratuita

Enlaces:

Sitio web: <http://beautifytools.com/>

Repositorio GitHub: <https://github.com/rennan/beautify-tools>

Beautify es el nombre de una amplia colección de herramientas online gratuitas entre las que se incluye la conversión de datos.

A diferencia de la herramienta anterior, esta admite un número de formatos significativamente mayor, y permite cargar archivos de datos, locales y remotos indicando la URL correspondiente.

3.3. Tabula

Tipo: Aplicación web

Licencia: [MIT](#)

Coste: Aplicación gratuita

Enlaces:

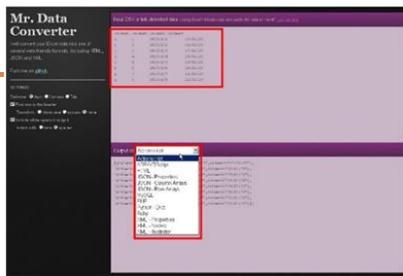
Sitio web: <https://tabula.technology/>

Repositorio GitHub: <https://github.com/tabulapdf/tabula>

Tabula es una herramienta que permite convertir informes PDF en hojas de cálculo de Excel (formato XLS o XLSX), CSV o archivos JSON, para ser usadas en bases de datos, aplicaciones de análisis o herramientas de visualización.

Herramientas de conversión de datos

Mr Data Converter



Beautify Converter



Tabula



4. Herramientas de análisis de datos

El [análisis de datos](#) es un proceso por el cual obtener conocimiento de la información subyacente a los propios datos. Este tipo de análisis implica la aplicación de técnicas y procesos cuantitativos y cualitativos utilizados para extraer información de valor que permita, entre otros objetivos, mejorar la productividad, responder a preguntas o tomar decisiones informadas. La analítica de datos se puede describir como [el análisis de datos estadísticos y matemáticos que agrupa, segmenta, clasifica y predice qué escenarios son más probables que sucedan](#).

Las metodologías de analítica incluyen tareas ligadas a los siguientes tipos de análisis: exploratorio de datos, que somete a las muestras de datos a un tratamiento estadístico; análisis descriptivo, cuyo objetivo es la exploración de los datos desde diferentes perspectivas, como agrupaciones o segmentaciones, entre otras y análisis predictivo, que pone el foco en modelos que permitan predecir valores futuros de las variables de interés.

La creciente disponibilidad de datos en bruto y su facilidad de acceso, ha permitido el desarrollo de un significativo número de herramientas que facilitan dichos análisis. Este apartado se ha dividido en dos categorías principales, herramientas orientadas a usuarios que no van a realizar tareas de programación y herramientas que aportan una mayor versatilidad gracias al manejo de lenguajes de programación.

4.1. Herramientas para no programadores

4.1.1. Weka

Tipo: Aplicación de escritorio

Licencia: [GNU - GLP V3](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://www.cs.waikato.ac.nz/ml/weka/>

Repositorio GitHub: <https://github.com/Waikato/weka-3.8>

Documentación: <https://waikato.github.io/weka-wiki/documentation/>

Weka es un software multiplataforma de aprendizaje automático de código abierto al cual se puede acceder a través de una interfaz gráfica de usuario apta para no programadores, mediante línea de comandos o a través de una API Java.

Las características principales de Weka son:

- Presenta una gran cantidad de herramientas integradas para tareas estándar de aprendizaje automático y permite el acceso a otras herramientas como son [scikit-learn](#), [R](#) y [Deeplearning4j](#).

- Preprocesamiento de los datos, esta herramienta soporta tareas estándar de minería de datos como, *clustering* o agrupamiento, clasificación, regresión y visualización, entre otros.
- Soporta el acceso a bases de datos SQL gracias a una conexión Java Database Connectivity (JDBC).

4.1.2. *Knime*

Tipo: Aplicación de escritorio

Licencia: [GNU – GPL V3](#)

Coste: Software gratuito (a partir de la versión 2.1) / diferentes versiones de pago

Enlaces:

Sitio web: <https://www.knime.com/>

Repositorio GitHub: <https://github.com/knime>

Knime es un software de minería de datos implementado sobre la plataforma de desarrollo Eclipse, que permite el análisis de datos y la realización de visualizaciones mediante una interfaz gráfica sobre la que se modelan los flujos de análisis de datos utilizando nodos, que representan a diferentes algoritmos y flechas que representan el flujo que siguen los datos en el pipeline de procesamiento.

Las propiedades principales que presenta esta herramienta son:

- Manipulación de datos, realización de muestreos, transformaciones, agrupaciones, o evaluación de modelos.
- Diferentes tipos de visualizaciones básicas, como histogramas o matrices de dispersión, para ver la disponibilidad y comportamiento de los datos.
- Aplicación de análisis estadísticos, utilizando, entre otros métodos, arboles de decisión, clasificaciones, clustering o regresiones.
- Igualmente, permite interacción con Weka e incorporar de manera sencilla código desarrollado en R o Python.

4.1.3. *Orange*

Tipo: Aplicación de escritorio

Licencia: [GNU - GPL V3](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://orangedatamining.com/>

Repositorio GitHub: <https://github.com/biolab/orange3>

Orange es software abierto de aprendizaje automático y minería de datos, similar a Knime.

Las características principales que presenta este software son:

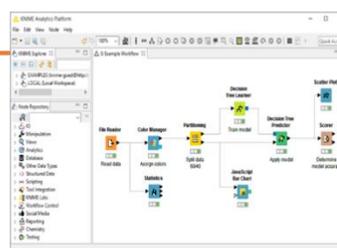
- Crea los análisis y las visualizaciones de datos utilizando el paradigma *drag and drop* a partir de un catálogo de herramientas o *widgets* que simbolizan diferentes tareas.
- Permite crear flujos de trabajo interactivos con el fin de analizar y visualizar los datos adaptándolos a las necesidades del usuario.
- Los widgets de visualizaciones incluyen gráficos estándar como pueden ser histogramas, diagramas de dispersión u otros específicos como dendrogramas, diagramas de silueta, arboles de decisión y redes, entre otros.
- Puede ser instalado como una librería de Python.

Herramientas de analítica de datos para no programadores

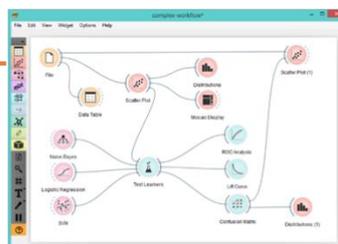
Weka



Knime



Tabula



4.2. Herramientas para programadores

4.2.1. R (*The R Project for statistical computing*)

Tipo: Lenguaje de programación

Licencia: [GNU - GPL V2](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://www.r-project.org/>

Guía para principiantes: https://cran.r-project.org/doc/contrib/rdebuts_es.pdf

R es un lenguaje de programación orientado a objetos e interpretado diseñado inicialmente para computación estadística y creación de representaciones gráficas. Es uno de los lenguajes de programación más usados en investigación científica, siendo muy popular en los campos de aprendizaje automático, minería de datos, investigación biomédica, bioinformática y matemáticas financieras.

Las principales características y ventajas de este lenguaje de programación son:

- Una de sus peculiaridades es la disponibilidad del entorno de programación, [R Studio](#).
- Aunque inicialmente se desarrolla para el contexto estadístico es una herramienta muy poderosa para todo tipo de procesamiento y manipulación de datos, que está permanentemente actualizada.
- Permite crear gráficos de alta calidad que pueden ser exportados en diversos formatos.
- Dispone de un notable número de herramientas estadísticas (modelado lineal y no lineal, test estadísticos, algoritmos de aprendizaje automático, clasificación, regresión, agrupamiento y redes neuronales, entre otros).
- Está formado por un conjunto de funciones que pueden ampliarse fácilmente mediante librerías o definiendo nuestras propias funciones.
- Desde sus inicios, una extensa [comunidad de usuarios y programadores](#) de alto nivel contribuyen a desarrollar nuevas funciones, paquetes y actualizaciones que están rápidamente disponibles para todos los usuarios de forma libre y gratuita.

4.2.2. Python

Tipo: Lenguaje de programación

Licencia: [Python Software Foundation \(PSF\) License Agreement](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://www.python.org/>

Repositorio de GitHub: <https://github.com/python>

Tutorial: <https://www.learnpython.org/es/>

Python es un lenguaje de programación interpretado, dinámico y multiplataforma. Se trata de un lenguaje de programación [multiparadigma](#) ya que soporta parcialmente programación orientada a objetos, programación estructurada, programación imperativa y en menor medida, programación funcional. Responde, además, a una [colección de principios básicos para su escritura](#) que afecta directamente en el diseño del lenguaje de programación.

Algunas de las ventajas y características de Python son:

- Lenguaje de programación sencillo, que presenta una curva de aprendizaje moderada, ya que su filosofía hace hincapié en ofrecer una sintaxis de código legible y, sobre todo, fácil de usar y aprender.
- Es de tipado dinámico, por lo que una variable puede tomar diferentes valores de distintos tipos en diferentes momentos. Las variables son declaradas por su contenido y no por su contenedor, lo que permite cambiar el valor y el tipo de variable durante la ejecución.
- La integración de diferentes librerías permite realizar avanzados análisis estadísticos y crear gráficos interactivos complejos.
- Dispone de potentes librerías para realizar visualizaciones ([Matplotlib](#) o [Bokeh](#)), cálculo numérico y análisis de datos ([Numpy](#), [SciPy](#) o [Pandas](#)) o para el procesamiento de lenguaje natural ([spaCy](#)), entre otras numerosas tareas.

4.2.3. GNU Octave

Tipo: Lenguaje de programación

Licencia: [GNU - GLP V3](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://www.gnu.org/software/octave/>

Repositorio de GitHub: <https://github.com/gnu-octave>

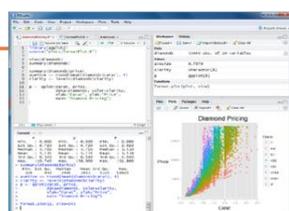
GNU Octave es un lenguaje de programación escrito en C++, diseñado principalmente para resolver tareas de álgebra computacional. Es la alternativa más conocida a la solución comercial [MATLAB](#), pero se trata de un software libre, no dispone de interfaz gráfica, el consumo de memoria es mucho menor pero la velocidad de ejecución puede resultar más lenta.

Las características principales que presenta esta herramienta son:

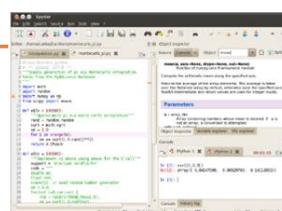
- Potentes funciones matemáticas integradas (ecuaciones diferenciales, cálculo de matrices, álgebra lineal, entre otras) y es posible ampliarlas con la incorporación de bibliotecas de funciones.
- Compatible con muchos scripts de Matlab.
- El intérprete de Octave tiene un motor gráfico basado en la librería [OpenGL](#) para crear gráficos 2D y 3D. Alternativamente se puede utilizar la librería [gnuplot](#) con el mismo propósito.
- Dispone del [paquete index](#) con numerosas extensiones que enriquecen notablemente la funcionalidad de la herramienta.

Herramientas de analítica de datos para programadores

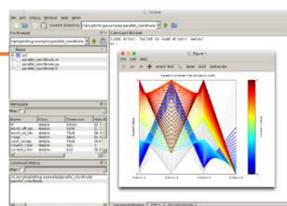
R



Python



GNU Octave



5. Herramientas de visualización de datos

La visualización de datos es una tarea vinculada al análisis de datos cuyo objetivo es la representación gráfica para interpretar la información subyacente a los datos, como puede ser un gráfico, un mapa y un diagrama, entre otros. Además de contribuir a una mayor comprensión e interpretabilidad de los datos, permite detectar patrones, tendencias, datos anómalos, proyectar predicciones o comunicar inferencia derivada del análisis de datos.

Las visualizaciones de datos no son algo nuevo, pero la incorporación de nuevas herramientas ha logrado que esta disciplina sea una parte esencial de la Ciencia de Datos. La aplicabilidad de las técnicas de visualización trasciende a ámbitos como el marketing, la salud, la ciencia, la economía e incluso tienen una presencia permanente en los medios de comunicación. Otras disciplinas, a priori, alejadas del contexto de los datos como el arte, incluyen estas [técnicas generando obras de arte, piezas musicales e incluso esculturas](#).

Las posibilidades de las visualizaciones son innumerables según sea la naturaleza de los datos. Desde el uso de elementos básicos para representar datos, como son los gráficos de líneas, barras, sectores, de dispersión o mapas de calor, entre otros, pasando por visualizaciones complejas configuradas sobre cuadros de mando o *dashboard* que pueden incorporar la información de diferentes indicadores en un espacio abarcable en un solo golpe de vista. Los dashboards incorporan habitualmente la interactividad necesaria para explotar la información mediante el uso de controles, filtros o funcionalidades como *drill-downs* que permiten navegar con facilidad de lo general a lo específico.

A continuación, veremos una serie de herramientas que realizan todo tipo de visualizaciones.

5.1. Herramientas de visualización genéricas

5.1.1. Kibana

Tipo: Aplicación de escritorio y Servicio en la nube

Licencia: [Dual: licencia comercial \(Elastic\) y SSPL](#)

Coste: Software gratuito / diferentes versiones de pago

Enlaces:

Sitio web: <https://www.elastic.co/es/kibana>

Repositorio GitHub: <https://github.com/elastic/kibana>

ELK Elastic: <https://www.elastic.co/es/what-is/elk-stack>

Primeros pasos: <https://www.elastic.co/es/webinars>

Kibana es una aplicación de código abierto que forma parte del paquete de productos [Elastic Stack](#) ([Elasticsearch](#), [Beats](#), [Logstash](#) y [Kibana](#)) que proporciona capacidades de visualización y exploración de datos indexados sobre el motor de analítica Elasticsearch.

Las características principales de Kibana son:

- Presentación visual a través de cuadros de mando o dashboards personalizables mediante la integración de filtros facetados por rangos, categorías de información, cobertura geoespacial o intervalos temporales, entre otros.
- Creación de informes gráficos personalizables ([Canvas](#)) mediante la combinación de gráficos, texto e imágenes.
- Dispone de un catálogo de herramientas de desarrollo ([Dev Tools](#)) para interactuar con los datos almacenados en Elasticsearch.
- Kibana cuenta con una versión básica gratuita y una versión enterprise que se desarrolla en un cloud propio de Elastic u otras infraestructuras en cloud como Amazon Web Services (AWS).

5.1.2. Tableau Public

Tipo: Aplicación de escritorio y Servicio en la nube

Licencia: [Comercial EULA](#)

Coste: software gratuito / diferentes versiones de pago

Enlaces:

Sitio web: <https://public.tableau.com/es-es/s/download>

Primeros pasos: <http://downloads.tableau.com/quickstart>

Galería: <https://public.tableau.com/en-us/gallery>

Herramienta de visualización de datos mediante gráficos que combina algunos elementos habituales en las herramientas de [Business Intelligence](#) como pueden ser el modelo de organización de variables mediante el uso de dimensiones y medidas o la conexión con otros sistemas de gestión de información como las bases de datos u hojas de cálculo con el uso de una interfaz gráfica atractiva, eficiente y rápida.

Algunas de las características más relevantes de esta aplicación son:

- Integración rápida y sencilla de datos: trabaja con bases de datos y hojas de cálculo de cualquier tamaño.
- Trabaja con una gran variedad de gráficos: líneas, barras, sectores, tablas, mapas, etc.
- Publicación de gráficos interactivos.
- Combinación de diversas fuentes de datos en una sola vista.
- Permite la descarga de datos desde las visualizaciones creadas.
- Para poder usar la versión gratuita, los datos deben ser publicados en su web, por tanto, no hay confidencialidad de la información.

5.1.3. SpagoBI

Tipo: Aplicación de escritorio

Licencia: [Mozilla Public License V2 \(MPL\)](#)

Coste: Software gratuito

Enlaces:

Sitio web: <https://www.spagobi.org/>

Repositorio GitHub: <https://github.com/SpagoBILabs/SpagoBI>

Primeros pasos: <https://spagolabs.wordpress.com/>

SpagoBI es una plataforma diseñada principalmente para aplicaciones de Inteligencia de negocios. Presenta una estructura modular basada en componentes ([Server](#), [Studio](#), [meta](#) y [SDK](#)) y ofrece soluciones para la presentación de informes, análisis multidimensional, minería de datos, dashboards y consultas ad-hoc.

Algunas de las características más relevantes de esta herramienta son:

- Crear informes personalizados y exportables en diferentes formatos (HTML, PDF, CSV, etc.) que pueden contener tablas, tablas cruzadas y gráficos.
- Incluye SpagoBI Studio como entorno de desarrollo que permite diseñar y editar las plantillas de los distintos motores de análisis, tales como informes, gráficos y análisis geográficos.
- Explora los datos con diferentes dimensiones y niveles de detalle.
- Creación de gráficos personalizados e interactivos.
- Permite extraer conocimiento y descubrir patrones a partir de grandes volúmenes de datos.
- Capacidad de añadir nuevos registros o modificar los existentes en el dataset directamente a través de la interfaz de usuario.

5.1.4. Grafana

Tipo: Aplicación de escritorio y Servicio en la nube

Licencia: [Apache V2](#)

Coste: Software gratuito / diferentes versiones de pago

Enlaces:

Sitio web: <https://grafana.com/>

Repositorio de GitHub: <https://github.com/grafana/grafana>

Grafana es una herramienta de código abierto que permite crear visualizaciones interactivas y dinámicas. Originalmente empezó siendo un componente de Kibana, pero en la actualidad son herramientas completamente independientes.

Las características más destacadas de Grafana son:

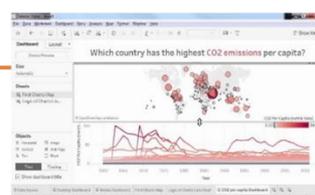
- Permite la creación de dashboards dinámicos y reutilizables.
- Permite una colaboración entre usuarios para el intercambio de datos y dashboards.
- Visualización de múltiples tipos de gráficos (histogramas, mapas de calor, series temporales, entre otros) que se pueden enriquecer y extender con la inserción de diferentes plugins.
- Integra múltiples bases de datos como PostgreSQL, MySQL y Elasticsearch, de las que se pueden obtener métricas personalizadas, filtrar datos o realizar anotaciones en tiempo real.

Herramientas de visualización

Kibana



Tableau Public



SpagoBI



Grafana



5.2. Librerías y APIs de visualización

Existe una amplia gama de librerías y APIs que implementan funcionalidades relacionadas con la visualización de datos, para diferentes lenguajes de programación o plataformas. Entre otras, reseñamos las siguientes:

5.2.1. *Google Chart Tools*

Tipo: API JavaScript

Licencia: [Términos del Servicio de Google](#)

Coste: Versión gratuita / Versión comercial

Enlaces:

Sitio web: <https://developers.google.com/chart>

Repositorio de GitHub: <https://github.com/google/charts>

Galería: <https://developers.google.com/chart/interactive/docs/gallery>

API de Google que permite la creación de visualizaciones gráficas. Es una herramienta muy fácil de usar e intuitiva.

Algunas características de esta herramienta son:

- Dispone de una gran variedad de gráficos interactivos.
- Los gráficos se pueden personalizar fácilmente.
- Permite la creación de dashboards utilizando diferentes tipos de widgets con los que se puede interactuar, como selectores de categoría, rangos y autocompletadores, entre otros.
- Las visualizaciones pueden ser integradas en sitios webs utilizando tecnología HTML5/SVG.
- El uso gratuito de la API está limitado a 250.000 peticiones por día.

5.2.2. *JavaScript InfoVis Toolkit*

Tipo: Librería JavaScript

Licencia: [MIT](#)

Coste: Librería gratuita

Enlaces:

Sitio web: <http://philogb.github.io/jit/>

Repositorio de GitHub: <https://github.com/philogb/jit>

Librería de JavaScript que proporciona herramientas para crear visualizaciones interactivas en aplicaciones web (mapas estratégicos, árboles jerárquicos o gráficos de líneas) a partir de estructuras de datos en JSON.

Algunas de las características más relevantes de esta librería son:

- Permite combinar las visualizaciones para crear nuevas formas de visualización.
- Eficiente en el manejo de estructuras de datos complejas.
- Extensible a través de módulos, pudiendo realizar compilaciones personalizables.
- Debido a la gran diversidad de opciones de visualización disponibles, se adapta a cualquier necesidad del desarrollador.

5.2.3. *Data-Driven Documents (D3.js)*

Tipo: Librería JavaScript

Licencia: [BSD V3](#)

Coste: Librería gratuita

Enlaces:

Sitio web: <https://d3js.org/>

Repositorio de GitHub: <https://github.com/d3/d3>

Galería: <https://www.d3-graph-gallery.com/>

Librería de JavaScript que permite crear visualizaciones complejas y gráficos interactivos. Básicamente la librería permite manipular documentos basados en datos usando estándares abiertos de la web, fundamentalmente SVG, CSS y HTML, de forma que los navegadores pueden interpretarlos para crear visualizaciones complejas sin depender de software propietario.

Algunas características relevantes de esta librería son:

- Configuración sencilla de gráficos encadenando funciones.
- Sus posibilidades de creación de gráficos son muy amplias, gráficos de burbujas, diagramas, matriciales y links de nodos, entre otros.
- Versatilidad prácticamente ilimitada.
- Es modular. No es necesario cargar toda la librería a la vez y se puede descargar e invocar únicamente el módulo que se desea utilizar.
- Permite una manipulación de un DOM (Modelo en Objetos para la Representación de Documentos) aplicando las transformaciones necesarias sobre el documento.
- El modelo de implementación permite la reutilización de código disponible en un amplio catálogo de [módulos D3](#).

5.2.4. Matplotlib

Tipo: Librería Python

Licencia: [Python Software Foundation \(PSF\)](#)

Coste: Librería gratuita

Enlaces:

Sitio web: <https://matplotlib.org/>

Repositorio de GitHub: <https://github.com/matplotlib/matplotlib>

Matplotlib es una de las librerías de Python más populares para crear visualizaciones estáticas, dinámicas e interactivas. Esta librería tiene una base de código muy extensa que puede resultar abrumadora para nuevos usuarios, pero la mayor parte de su marco conceptual es simple y con el conocimiento de algunos conceptos básicos se puede manejar con soltura. Para que esto ocurra, presenta una organización jerárquica que va desde el nivel más general, como puede ser el contorno de una matriz 2D, hasta un nivel muy específico, como puede ser colorear un pixel determinado.

Las características más comunes de esta librería son:

- Capacidad para crear interfaces de usuario interactivos.
- Es muy flexible, aunque parte de ajustes de valores predeterminados que ayudan al usuario a la creación de gráficos.
- Permite la exportación de imágenes en los formatos más populares (PNG, JPG, TIFF o GIF) o en formato PGF para LaTeX y su inclusión en artículos científicos.

5.2.5. Bokeh

Tipo: Librería Python

Licencia: [BSD](#)

Coste: Librería gratuita

Enlaces:

Sitio web: <https://bokeh.org/>

Repositorio de GitHub: <https://github.com/bokeh/bokeh>

Bokeh es una librería de código abierto para Python. Permite el desarrollo de representaciones gráficas y dashboards versátiles e interactivos de manera rápida y sencilla.

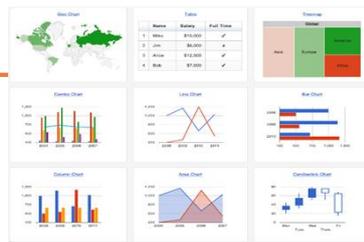
Esta herramienta presenta diferentes características relevantes:

- Permite la creación de una amplia variedad de formas visuales.

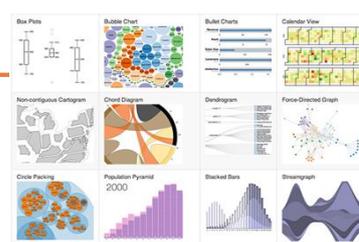
- Simplicidad en la implementación, con pocas líneas de código se pueden crear visualizaciones interactivas complejas, pero a la vez posee flexibilidad para personalizar cualquier gráfico.
- Los gráficos o los dashboards creados con esta herramienta, puede ser publicados en sitios web o cuadernos de [Jupyter](#).
- La librería permite embeber código JavaScript para implementar funcionalidad específica.

Librerías y APIs de visualización

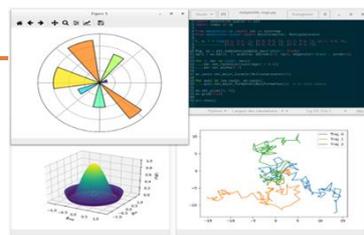
Google Chart Tools



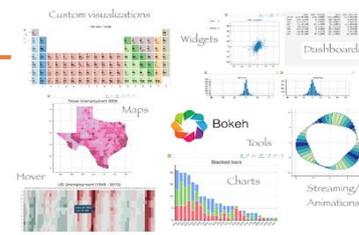
D3.js



Matplotlib



Bokeh



5.3. Herramientas de visualización geoespacial

Muchos de los datos disponibles para su [reutilización están georeferenciados o se pueden georeferenciar maximizando así su valor](#). Para manejar adecuadamente este tipo de datos es esencial entender conceptos clave vinculados a la representación visual de la información geográfica, como los sistemas de referencia de coordenadas, las proyecciones cartográficas, los diferentes modelos de representación con los que se trabaja, raster o vectoriales, sin olvidar la disponibilidad de muchos de estos datos a través de servicios online geográficos de visualización o descarga de mapas.

Sin lugar a dudas, se trata de una disciplina que en si misma combina el conocimiento del contexto de la información geoespacial, con el proceso e integración de datos en capas de información y la visualización e interacción sobre plataformas de servicios geográficos.

Muchos de estos elementos se combinan sobre los Sistemas de Información Geográfica (GIS o SIG), no obstante, a continuación, se relaciona un conjunto de herramientas útiles para abordar las tareas necesarias para desarrollar visualizaciones de datos geoespaciales.

5.3.1. Carto

Tipo: Aplicación web y Servicio en la Nube

Licencia: [BSD](#)

Coste: Versión gratuita / Versión de pago

Enlaces:

Sitio web: <https://carto.com/>

Repositorio de GitHub: <https://github.com/CartoDB/cartodb>

Plataforma de análisis de datos geoespaciales con una interfaz muy amigable, orientado a desarrolladores sin experiencia previa en sistemas de información geoespacial. La finalidad fundamental de Carto es facilitar la creación de aplicaciones interactivas geolocalizadas.

Entre otras características, cabe destacar:

- Diseño de mapas: es posible utilizar CartoCSS en las capas de datos para editar fácilmente el formato y la apariencia que presentan.
- Permite diseñar y desarrollar mapas en tiempo real que funcionan en plataformas web y dispositivos móviles.
- Vinculación con servicios cartográficos como Google Maps o MapBox, aprovechando de éstos las funciones típicas de zoom o desplazamiento.
- Geocodificación: permitiendo obtener coordenadas geográficas a partir de información alfanumérica.
- Capacidad de importar colecciones de datos en múltiples formatos fácilmente permitiendo introducir datos directamente en las tablas, mediante consultas SQL o vinculando datos remotos invocando su URL.
- Mediante el uso de la librería [PostGIS](#), Carto permite consultar y combinar conjuntos de datos geoespaciales.
- Es un proyecto Open Source disponible como aplicación local o como SaaS, aportando la escalabilidad, flexibilidad y elasticidad que aporta este tipo de servicios.

5.3.2. OpenLayers

Tipo: API

Licencia: [BSD](#)

Coste: API gratuita

Enlaces:

Sitio web: <https://openlayers.org/>

Repositorio de GitHub: <https://github.com/openlayers/openlayers>

Librería de JavaScript de código abierto que permite la inclusión de componentes tipo mapa en cualquier página web.

Algunas características:

- Funciona como un visor de mapas desde el navegador, por lo tanto, los mapas se descargan directamente del servidor de mapas, por lo que no genera tráfico en el servidor Web.
- Permite superponer distintas capas y añadir diferentes características como puntos, líneas, polígonos e iconos sobre los que vincular una leyenda.
- Renderiza cualquier elemento DOM en cualquier lugar del mapa.
- Incorpora un set de controles básicos y una barra de herramientas de controles avanzados permitiendo embeber la funcionalidad necesaria haciendo uso del API.

5.3.3. OpenStreetMap

Tipo: Aplicación web y API

Licencia: [Open DataBase License \(ODbL\)](#)

Coste: Aplicación web y API gratuitas

Enlaces:

Sitio web: <https://www.openstreetmap.org/>

Repositorio de GitHub: <https://github.com/openstreetmap>

Proyecto colaborativo de creación de mapas libres y editables. Los mapas se crean utilizando información geográfica capturada con dispositivos GPS, ortofotos y otras fuentes de dominio público. Tanto la cartografía, como los datos vectoriales almacenados en su base de datos, se distribuyen bajo licencia abierta Open Database License (ODbL).

Las características principales que presenta esta herramienta son:

- Los usuarios registrados pueden subir sus trazas desde el GPS y crear y corregir datos vectoriales mediante herramientas de edición creadas por la comunidad OpenStreetMap.
- Utiliza una estructura de datos topológica que se almacena en el [datum WGS84 lat/lon \(EPSG:4326\)](#) como sistema de referencia de coordenadas.
- Los elementos básicos de la cartografía son los nodos, vías y relaciones entre grupos de nodos y/o vías con características similares que se mantienen editando la [wiki](#) del proyecto.

5.3.4. Geocoder y Geopy

Tipo: Librerías Python

Licencia: [MIT](#)

Coste: Librerías gratuitas

Enlaces:

Geocoder:

Página web: <https://geocoder.readthedocs.io/>

Repositorio de GitHub: <https://github.com/DenisCarriere/geocoder>

GeoPy:

Página web: <https://geopy.readthedocs.io/en/stable/>

Repositorio de GitHub: <https://github.com/geopy/geopy/>

Aunque no se trata de herramientas de visualización geoespacial propiamente dichas, conviene destacar en este apartado la disponibilidad de librerías para Python que resuelven una tarea habitual en el tratamiento de información geográfica como es el proceso de geocodificación para convertir direcciones postales en coordenadas espaciales o viceversa, incorporando además la capacidad de calcular la distancia entre puntos geolocalizados.

5.3.5. GDAL

Tipo: Librería

Licencia: [MIT](#)

Coste: Librería gratuita

Enlaces:

Página web: <https://gdal.org/>

Repositorio de GitHub: <https://github.com/OSGeo/gdal>

GDAL es una librería de código abierto útil para la traslación entre datos geoespaciales vectoriales y raster. Un buen número de herramientas que incorporan funciones de Sistema de Información Geográfica (SIG o GIS), como [PostGIS](#), [Carto](#) o [ArcGIS](#), entre otros, tienen integrada esta librería para realizar este proceso.

Algunas de sus características son:

- Soporta numerosos formatos raster (GeoTIFF, ECW, DETD o AIG), formatos vectoriales (Shapefile, GML, GeoJson o KML) tanto de entrada como de salida.
- Soporta archivos de gran tamaño, más de 4GB.
- Permite acceder a ellas con diferentes lenguajes de programación como son Python, Java, Ruby, VB6 y Perl.

5.3.6. PROJ.4 y PROJ4.JS

Tipo: Librerías

Licencia: [MIT](#)

Coste: Librerías gratuitas

Enlaces:

PROJ.4:

Página web: <https://proj.org/>

Repositorio de GitHub: <https://github.com/OSGeo/PROJ>

PROJ4.JS:

Página web: <http://proj4js.org/>

Repositorio de GitHub: <https://github.com/proj4js/proj4js>

Otra tarea habitual en la representación de datos geográficos es la transformación de coordenadas geoespaciales de un sistema de referencia de coordenadas a otro. Esta librería de manejo de proyecciones cartográficas permite abordar este tipo de transformaciones.

Algunas de las características principales que presenta estas librerías son:

- Invocable desde línea de comandos para una fácil conversión de coordenadas en archivos de texto.
- PROJ4.JS es la implementación de PROJ.4 para JavaScript.

Herramientas de visualización de datos espaciales

Carto



OpenLayers



OpenStreetMap



6. Herramientas de análisis de redes

La minería de medios sociales es el proceso de representar, analizar y extraer patrones con significado a partir de datos de medios sociales. Este tipo de análisis permite observar relaciones entre entidades de un sistema, para poder describir comportamientos, estructura o mecanismos de influencia. Es una aproximación metodológica y teórica que enfatiza en el estudio de las relaciones entre personas, cosas, países u organizaciones entre otros elementos que configuran un sistema. La aplicación del análisis de redes ha aumentado en los últimos años y sigue aumentando debido a la proliferación de contextos relacionales y su utilización en múltiples disciplinas (sociología, psicología, física, matemáticas, antropología, salud, etc.). Algunos usos destacables son: el descubrimiento y caracterización de comunidades, la monitorización de tendencias culturales, el análisis de mercados y productos, la predicción o monitorización de epidemias o la minería de opiniones.

Este tipo de herramientas permite estudiar características intrínsecas de las redes sociales como el grado de centralidad, la densidad, la intermediación o cercanía entre miembros de una red.

6.1. Gephi

Tipo: Aplicación de escritorio

Licencia: [CDDL 1.0](#) y [GNU General Public License V3](#)

Coste: Aplicación gratuita

Enlaces:

Página web: <https://gephi.org/>

Repositorio de GitHub: <https://github.com/gephi/gephi>

Plataforma de exploración y visualización interactiva para todo tipo de redes y grafos complejos, dinámicos y jerárquicos. Funciona en cualquier entorno y es una plataforma de propósito general. La arquitectura de Gephy es modular y no son necesarias habilidades de programación.

Las características principales que presenta son:

- Permite manejar redes de gran tamaño con hasta 50.000 nodos y un millón de aristas mediante acciones de particionado y ordenación de los grafos.
- Admite un número significativo de formatos de entrada comunes para el análisis de redes: GDF (GUESS), GraphML (NodeXL), GML, Pajek NET, GEXF, entre otros.
- Ofrece un conjunto de estadísticos fruto del análisis de redes muy completo e incorpora un potente laboratorio de datos que permite una útil exportación de los informes gráficos generados.

6.2. NodeXL

Tipo: Extensión de Excel

Licencia: [Comercial](#)

Coste: Versión gratuita / Diferentes versiones de pago

Enlaces:

Página web: <https://www.smrfoundation.org/nodexl/>

Galería: <https://nodexlgraphgallery.org/Pages/Default.aspx>

Herramienta de análisis y representación de grafos que se integra como una extensión de Microsoft Excel. Trabaja mostrando gráficos de red a partir de una lista dada de conexiones aportando capacidades de análisis y descubrimiento de patrones y relaciones en los datos.

Algunas de las características más relevantes son:

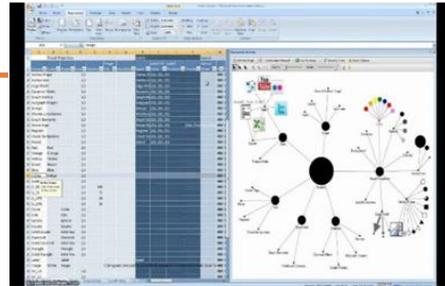
- Importación y exportación flexible de múltiples fuentes dentro del entorno Microsoft Excel.
- Conexiones directas para analizar medios sociales en línea, incluyendo conexiones para hacer consultas a las API de Twitter, Flickr y YouTube.
- Cálculo de métricas (cálculo de la centralidad, densidad, modularidad, entre otras) y análisis de grafos, aunque la orientación principal es la visualización.
- Automatización de tareas.

Herramientas de análisis de redes

Gephi



NodeXL



Comparativa

En esta tabla se muestran algunas de las características de las herramientas, para que puedan ser identificadas y comparadas con sus homónimos de un solo vistazo.

CATEGORÍA	HERRAMIENTA	TIPO	LICENCIA
Herramientas de Web Scraping	ParseHub	App de escritorio y Web	MIT
	Scrapy	Librería	BSD
	Table Capture	Extensión de navegador	Términos de servicio de Google
	IMPORTHTML()	Función Google SpreadSheet	Términos de servicio de Google
Herramientas de Depuración de datos	Open Refine	App Web	BSD-3
	Talend Open Studio	App de escritorio	Apache V2
Herramientas de Conversión de datos	Mr Data Converter	App Web	MIT
	Beautify Converter		MIT
	Tabula		MIT
Herramientas de Analítica de datos	Weka	App de escritorio	GNU-GPL
	Knime		GNU-GPL V3
	Orange		GPL
	R	Lenguaje de programación	GNU-GPL
	Python		PSF
	GNU Octave		GNU-GPL
Herramientas de visualización	Kibana	App de escritorio y Servicio en la Nube	Comercial + SSPL
	Tableau Public	App de escritorio y Servicio en la Nube	Comercial EULA
	SpagoBI	App de escritorio	MPL
	Grafana	App de escritorio y Servicio en la Nube	Apache V2
Librerías y APIs de visualización	Google Chart Tools	API	Términos de Servicios de Google
	JavaScript InfoVis Toolkit	Librería	MIT
	D3-js	Librería	BSD
	Matplotlib	Librería	PSF
	Bokeh	Librería	BSD

CATEGORÍA	HERRAMIENTA	TIPO	LICENCIA
Visualizaciones y tratamiento de datos espaciales	CartoDB	App de escritorio y Servicio en la Nube	BSD
	Openlayers	API	BSD
	OpenStreetMap	API App Web	ODbL
	Geocoder	Librería	MIT
	GeoPy	Librería	MIT
	GDAL	Librería	MIT
	PROJ.4 y PROJ4.JS	Librería	MIT
Herramientas de Análisis de redes	Gephi	App de escritorio	CDDL 1.0 GNU-GLP
	NodoXL	Extensión de Excel	Comercial

Conclusiones finales

Como hemos observado, el procesamiento, análisis y visualización de datos es el mecanismo necesario para explotar el potencial de conocimiento subyacente de los datos. Actualmente existe una notable cantidad y diversidad de herramientas disponibles en el mercado que abordan las tareas que conlleva. En esta guía hemos reseñado una selección de ellas muy utilizadas por la comunidad de la reutilización de datos abiertos. Afortunadamente la disponibilidad es muy amplia y son los usuarios quienes pueden juzgar la bondad de cada una en función de la utilidad que aportan.

La mejor herramienta será la que cubra las necesidades de tu proyecto y a la vez, mejor se adapte a tus competencias profesionales. En algunos casos, los usuarios optan por complejas y completas *suites* de desarrollo multipropósito y en otros la preferencia se focaliza en herramientas específicas que cumplen provechosamente su misión. En cualquier caso, las que hoy son herramientas de referencia para determinadas tareas, irán dejando espacio a otras que presentarán características evolucionadas por incorporar nuevas o actualizadas funcionalidades, o proponer un nuevo modelo de funcionamiento que aporta potencia o mejora considerablemente la eficiencia.

En este sentido, la computación en la nube ha modificado significativamente el escenario de explotación de este tipo de herramientas, de tal forma que muchas de las propuestas vigentes, o son plataformas nativas en *cloud*, o disponen de modalidades de comercialización como servicio. De ahí que las versiones gratuitas de un buen número de las herramientas disponibles, aunque en su concepción se hayan implementado sobre el paradigma *open source*, incorporan algún tipo de limitación funcional, de capacidad de almacenamiento o de importación o exportación de formatos, entre otras.

Otro aspecto a resaltar del conjunto de herramientas disponibles es la importancia que cobran los estándares, sobre todo en lo referente a entrada y salida de datos. Aunque, la existencia y manejo de formatos propietarios es un hecho y algunos pueden pasar por tener la consideración de estándares de facto, prácticamente ninguna herramienta impide trabajar con archivos en formatos JSON, XML o CSV, lo cual es muy de agradecer, aunque siga haciendo falta contar con herramientas que faciliten la conversión de formatos, como reseñamos en esta propuesta.

Este escenario cambiante impulsa la evolución de las herramientas para hacerlas más útiles y eficientes, pero no hay que olvidar que, en el terreno de las visualizaciones de datos, la necesidad a cubrir sigue siendo la misma: lograr comunicar a las personas toda la información contenida en los datos con la máxima posibilidad de comprensión. Para ello vamos a seguir utilizando gráficos de líneas, barras o sectores combinados con otros más sofisticados, pero sobre potentes *dashboards* interactivos que sin lugar a dudas enriquecen notablemente la experiencia de cada usuario.

¿QUIERES SABER MÁS SOBRE
LA **INICIATIVA APORTA**?

Visita www.datos.gob.es

Twitter: [@datosgob](https://twitter.com/datosgob)

Linkedin: [datos.gob.es](https://www.linkedin.com/company/datos-gob-es)

Suscríbete a nuestro [boletín](#)

Escribe a contacto@datos.Gob.es

Puedes identificar los
espacios de **datos abiertos**
gracias a este logo



**datos
abiertos**