



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI



comillas.edu

Universidad Pontificia Comillas ICAI

Analítica Social y de la Web *Social and Web Analytics*

1.3 Entorno de trabajo

Javier Ruiz de Ojeda

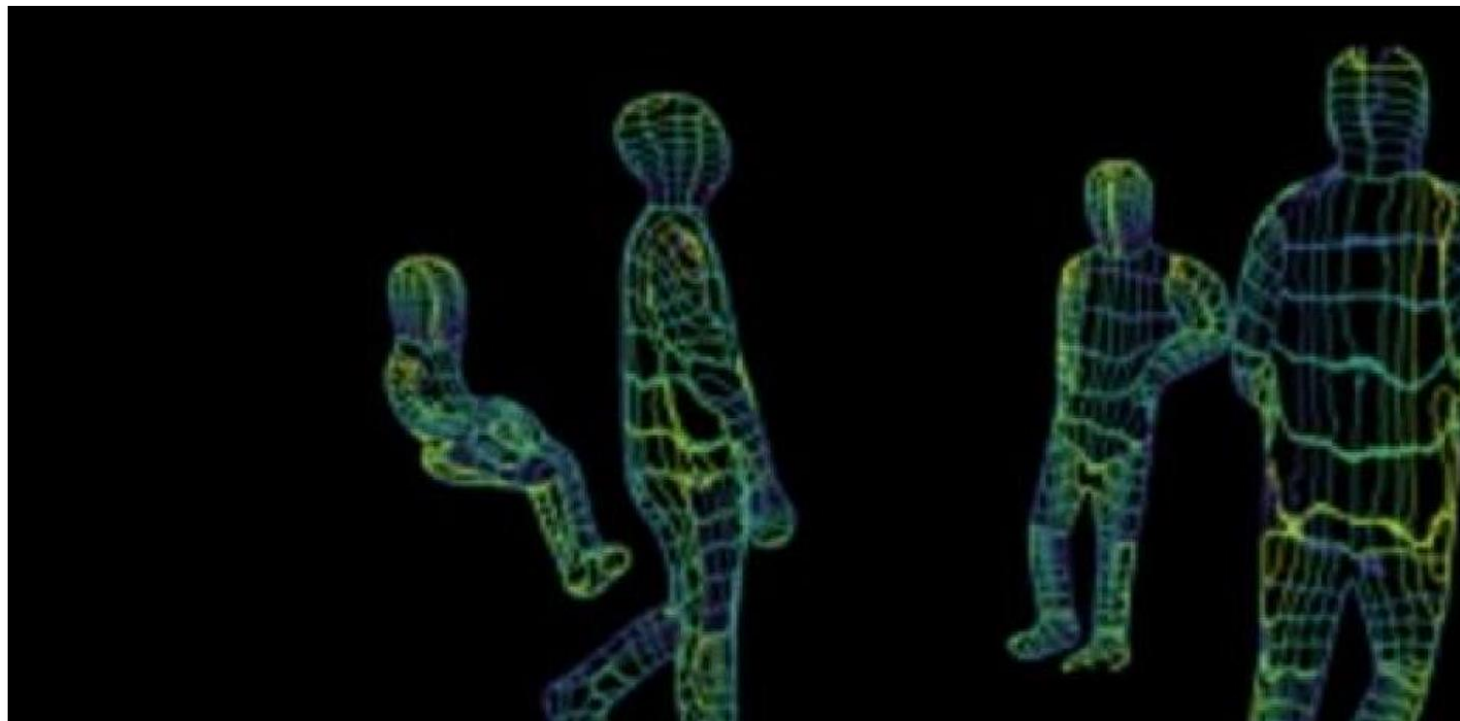
Curso 2023-24
Segundo semestre



Nuevo jaque a la privacidad: llegan los sistemas que usan el wifi para ver a través de las paredes



- Los investigadores logran mapear en detalle el cuerpo humano o identificar dispositivos en las casas mediante los routers



Las señales de WiFi se pueden usar para mapear un cuerpo humano en detalle a través de las paredes (Jiaqi Geng/ Universidad Carnegie Mellon)



Mayte Rius

20/01/2023 06:00 | Actualizado a 20/01/2023 12:00



36

A DAY IN THE LIFE OF A TYPICAL GOOGLE USER

MORNING ACTIVITY



- 1 Gets ready in the morning while listening to Google Play Music
- 2 Drops kids off at daycare before commuting to work
- 3 Checks the news while commuting to work on the subway
- 4 Searches for cold medicine while on the subway

cnn.com
Visited CNNPolitics - Political News, Analysis and Opinion

WSJ - wsj.com
Visited California Takes Big Step to Require Seller on New Homes - WSJ

Search
Searched for cold symptoms

Search
Searched for cold vs allergies

Search
Searched for cold medicine

AFTERNOON ACTIVITY



- 5 Walks from subway to work
- 6 Uses Maps to find a new lunch spot
- 7 Gets coffee from Starbucks using her Starbucks app
- 8 Schedules doctors appointment, Google creates a Calendar event from the confirmation email
- 9 Walks to Walgreens and purchases cold medicine using Google Pay

← DUANE READE #14480
\$12.99
DATE: May 9 TIME: 6:50 PM
CARD USED: Visa ****1041

Web site tracking tools

```

<script async src="https://tag-73101.appspot.com/gtag/js?id=UA-12345678-1"></script>
<script>
  (function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
    new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName(s)[0],
    j=d.createElement(s),dl='//www.googletagmanager.com/gtm.js';j.async=true;j.src=
    dl;f.parentNode.insertBefore(j,f);
})(window,document,script,analytics,analytics);
  
```

EVENING ACTIVITY



- 10 Takes an Uber home from work
- 11 Looks for hotels on Expedia for a weekend trip
- 12 Uses Google Home to play music for her kids
- 13 Watches late night TV videos on YouTube

Assistant
Said play lullaby
9:25 PM • Details
PLAY

Assistant
Said lower the volume
9:25 PM • Details
PLAY

Item details

YouTube
Watched Jimmy Fallon Recaps the Met Gala
The Tonight Show Starring Jimmy Fallon

Drake
Today at 10:19 PM

Logística y datos de interés

Bibliografía de la asignatura

- **Siddhartha Chatterjee, Michal Krystyanczuk, Python Social Media Analytics, Packt Publishing 2017**
- Krishna Raj P.M., Ankith Mohan, Srinivasa K.G., Practical Social Network Analysis with Python, Springer 2018
- Matthew A. Russell, Mikhail Klassen, Mining the Social Web, 3rd Edition, O'Reilly 2019
- *(Ryan Mitchell, Web Scraping with Python, 2nd Edition, O'Reilly 2018)*
- *(Gabor Szabo, Gungor Polatkan, P. Oscar Boykin, Antonios Chalkiopoulos, Social Media Data Mining and Analytics, Wiley 2018)*



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

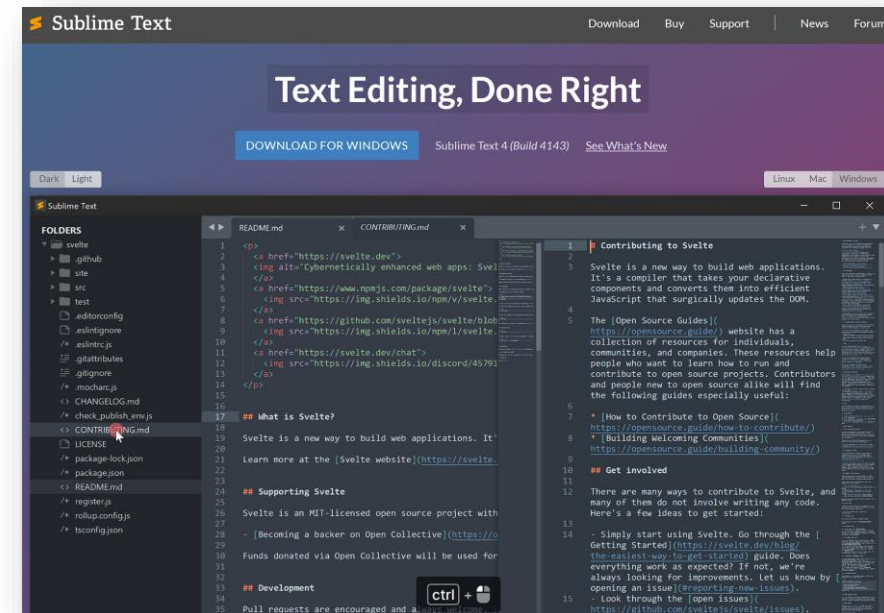
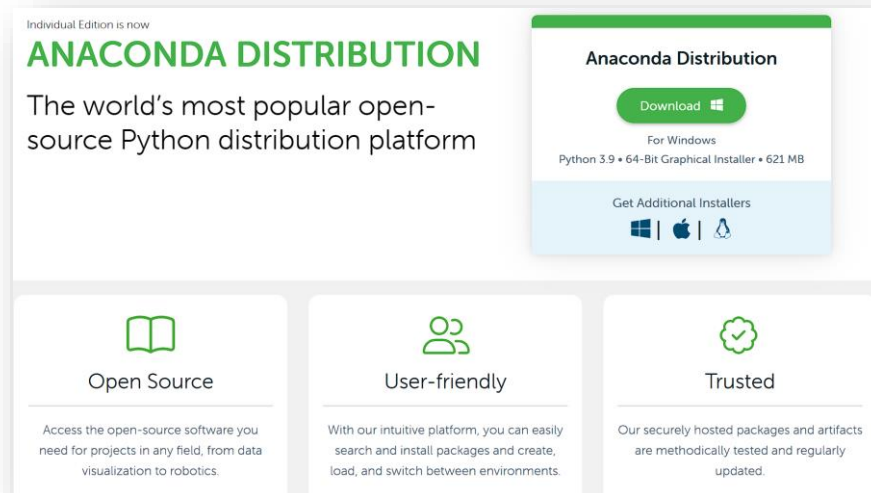
Entorno de trabajo



Entorno de trabajo

Python y entornos

- Versión 3.X
- Gestor de Paquetes: `pip`
- Entornos:



Entorno de trabajo

Gestor de paquetes: `pip`

- Permite la instalación de paquetes no incluidos en las librerías estándar
- Emplea el repositorio PyPI (Python Package Index) como fuente
- Instalación de paquetes: `pip install`

```
$ python3 -m pip install package_name
```

- Desinstalación de paquetes: `pip uninstall`

```
$ python3 -m pip uninstall package_name
```

Entorno de trabajo

Gestor de paquetes: pip

- Listar paquetes: `pip list`

```
In [1]: pip list
```

Package	Version	Note: you may need to restart the kernel to use updated packages.

alabaster	0.7.12	
anaconda-client	1.11.0	
anaconda-navigator	2.3.1	
anaconda-project	0.11.1	
anyio	3.5.0	
appdirs	1.4.4	
argon2-cffi	21.3.0	
argon2-cffi-bindings	21.2.0	
arrow	1.2.2	
astroid	2.11.7	
astropy	5.1	
atomicwrites	1.4.0	
attrs	21.4.0	
Automat	20.2.0	
autopep8	1.6.0	
Babel	2.9.1	
bokeh	2.3.0	



COMILLAS
UNIVERSIDAD PONTIFICIA

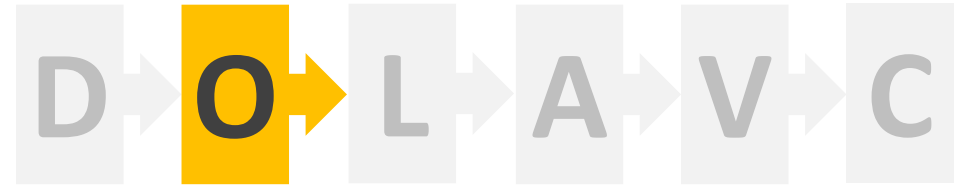
ICAI

Obtención de datos



Obtención de datos

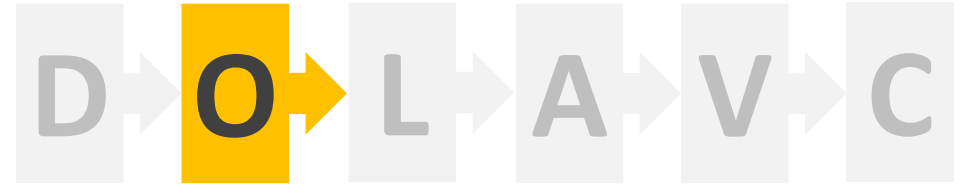
Data harvesting



- *Data harvesting* (cosecha) es la obtención de datos sociales, y puede hacerse de dos formas principales:
 - Mediante APIs
 - Mediante *Scraping* o *Crawling* de redes sociales
- Saber de dónde provienen nuestros datos es crucial pues nos permitirá anticiparnos a sesgos o vacíos que pueda haber

Obtención de datos

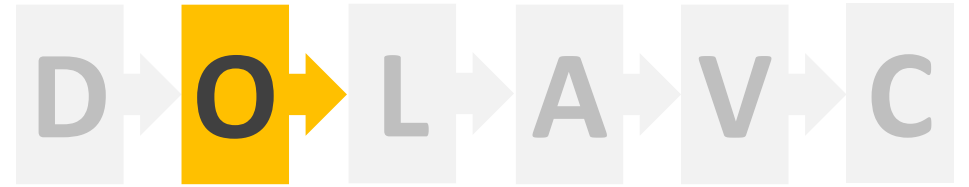
APIs



- API (Application Programming Interface): Conjunto de instrucciones y estándares que permiten acceder a aplicaciones software basadas en web.
- Mediante una API podemos enviar peticiones de recursos a una web y recibir datos en respuesta. Casi todos los fabricantes establecen límites al tipo y cantidad de datos que podemos obtener.
- Algunas APIs permiten incluso que ciertas operaciones se procesen en remoto, como Alchemy API, que nos devuelve datos ya procesados.
- Acceso a las APIs: `pip3 install urllib3`
 - También: `pip3 install requests`

Obtención de datos

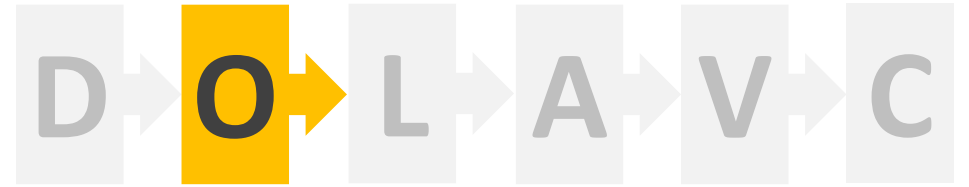
Scraping y Crawling



- *Scraping* (rascar): Cuando no se dispone de una API solo podemos obtener información visible del código HTML generado en una página. Para ello necesitamos un *scraper* que sea capaz de obtener la información que necesitamos y estructurarla en un formato predefinido.
- *Crawling* (rastrear): Un *crawler* es una herramienta capaz de seguir todos los enlaces de una página web y de sus sub-webs.
- `pip3 install bs4, scrapy`

Obtención de datos

Consideraciones



- Antes de realizar una operación de *scraping* debemos confirmar que los términos y condiciones de la web lo permiten.
- La forma más sencilla suele ser mediante el archivo `robots.txt` de la página web, por ejemplo:

```
< > ↺ 🏠 facebook.com/robots.txt  
Start  
# Notice: Collection of data on Facebook through automated means is  
# prohibited unless you have express written permission from Facebook  
# and may only be conducted for the limited purpose contained in said  
# permission.  
# See: http://www.facebook.com/apps/site_scraping_tos_terms.php
```

Indica que el sitio web no permite y no desea ser *scrapeado*:

```
User-agent: *  
Disallow:/
```



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

Ejercicio:

¿Permiten scraping las RRSS?





COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

Análisis de datos

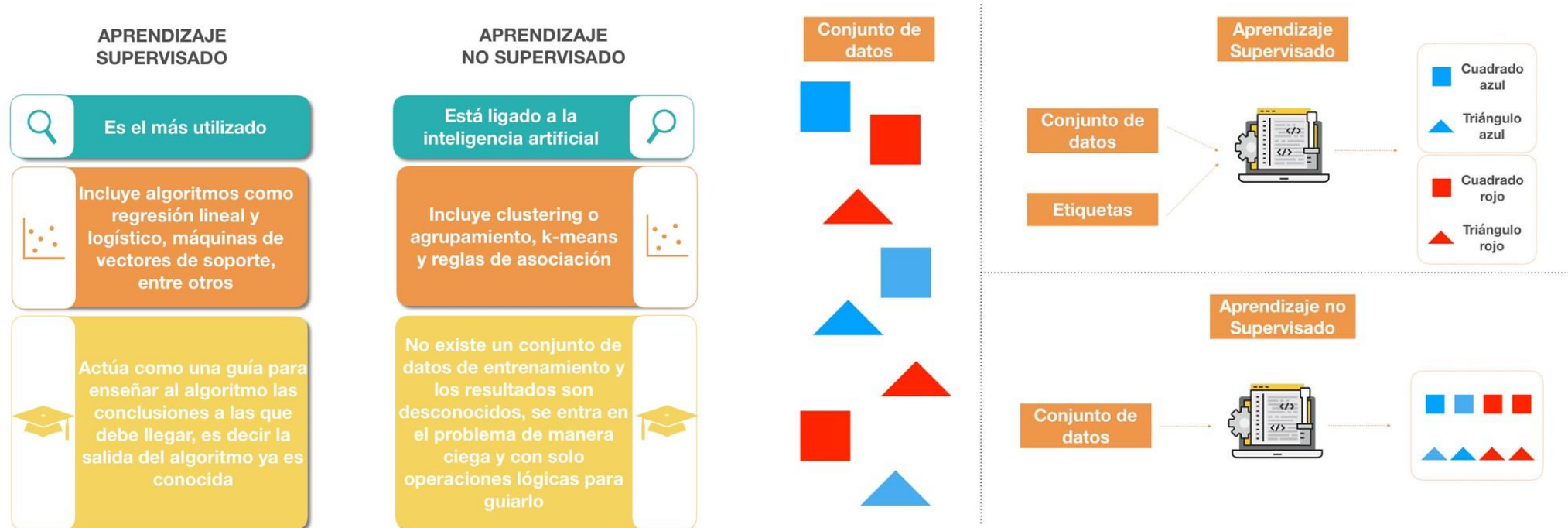


Análisis de datos

Machine Learning



- Disciplina que da a los ordenadores la capacidad de aprender sin ser específicamente programados para ello.

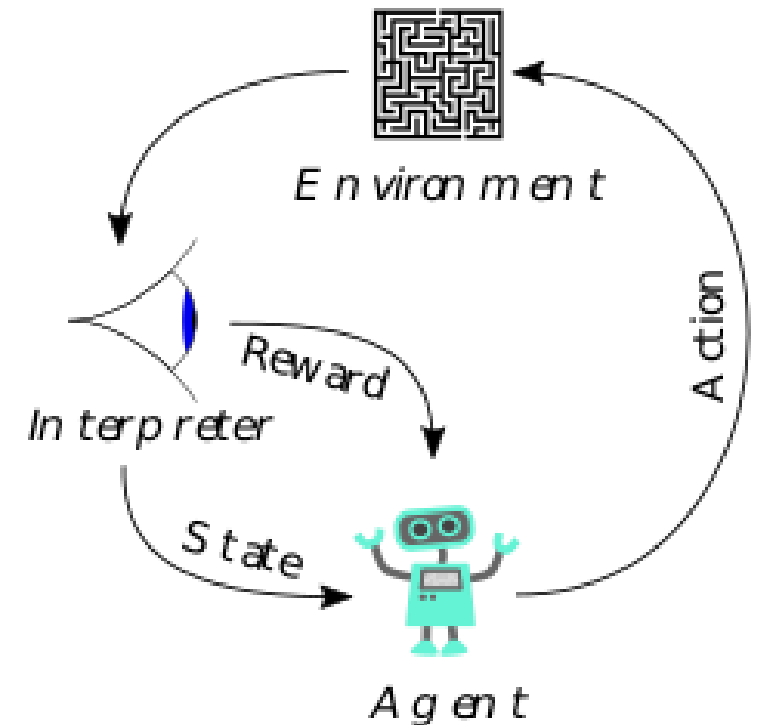


Análisis de datos

Machine Learning



- Aprendizaje por refuerzo (IA): es un área del aprendizaje automático inspirada en la psicología conductista, cuya ocupación es determinar qué acciones debe escoger un agente de software en un entorno dado con el fin de maximizar alguna noción de "recompensa" o premio acumulado.



Análisis de datos

Aplicado al Análisis Social



- **Text Analytics:** permite la obtención de información no trivial de datos textuales, como marcas, nombres propios, relaciones entre palabras, números de teléfono, URLs, *hashtags*, etc.
- **Natural Language Processing:** pretende encontrar el sentido del texto analizando su estructura, semántica y conceptos entre otros.
- **Graph Mining:** Sabiendo que una red social puede representarse en forma de grafo, este método ayuda a descubrir relaciones, aristas, conexiones y agrupaciones de personas, marcas, asuntos, etc.

Análisis de datos

Librerías de estructura de datos



- Emplearemos **pandas**, **sframe** (que permite acceder a conjuntos de datos mayores que de otra forma no cabrían en la memoria RAM), y **MongoDB**:
- `pip3 install pandas, sframe, pymongo`





COMILLAS
UNIVERSIDAD PONTIFICIA

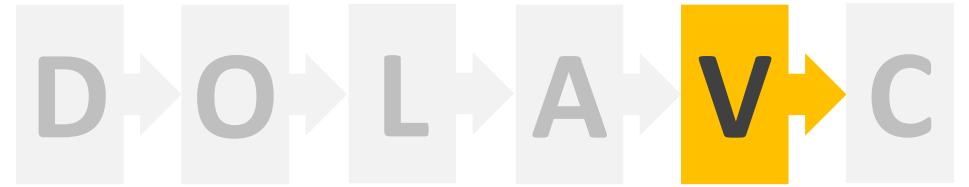
ICAI

Visualización de datos



Visualización de datos

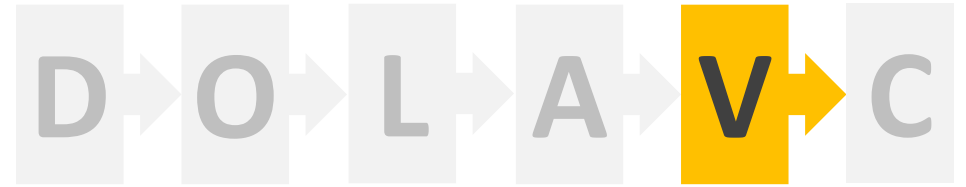
Representación e interpretación



- La visualización de datos se puede entender como la presentación gráfica de información.
- Tiene dos objetivos: interpretar y construir un significado a partir de los datos y la comunicación.
- Se trata de un mecanismo muy importante a la hora de descubrir y entender la lógica que existe tras un conjunto de datos, y también a la hora de compartirla.
- `pip3 install plotly, matplotlib, seaborn`

Visualización de datos

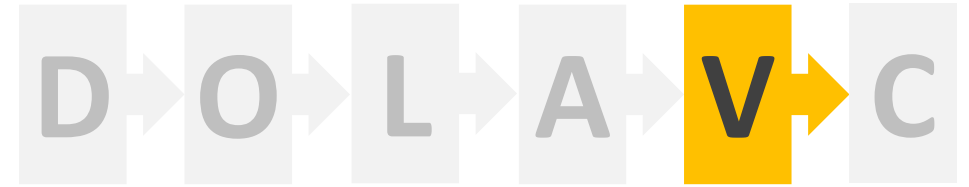
Beneficios



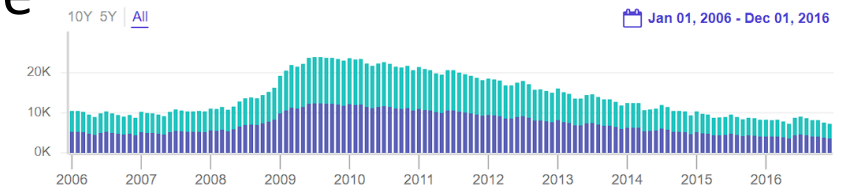
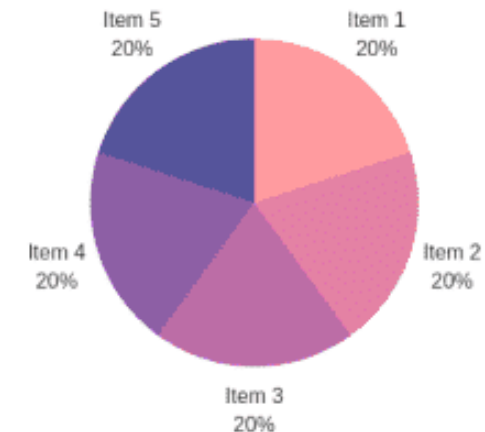
- Gracias a las gráficas podemos interpretar grandes cantidades de datos de forma clara y coherente. Esto nos permite ver la información desde diferentes perspectivas, sacar conclusiones, o crear métricas importantes para nuestros objetivos. También facilita el reconocimiento de tendencias emergentes y poder responder rápidamente en función de lo que detectamos.
- Las imágenes y los diagramas nos ayudan a identificar parámetros correlacionados y tienen más sentido cuando se muestran gráficamente. Algunas relaciones suelen ser obvias, pero solo emergen al visualizarse.
- Ver un gráfico es más cómodo para el cerebro que entender un texto. Por eso, la visualización de los datos mejorará automáticamente la velocidad de los procesos de toma de decisiones.

Visualización de datos

Tipos de gráficos

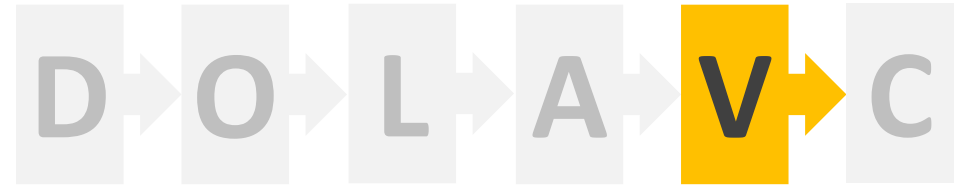


- Siempre antes de escoger lo más importante es tener claro a quién te diriges y cuál es el propósito de la imagen, pues cada gráfico tiene una función diferente:
 - Gráfico Circular: Muestran la división de varios elementos y son muy útiles si el destinatario se lee los datos por encima o si quieres mostrar cómo distintas partes representan un total.
 - Gráfico de Barra: Su función es hacer énfasis en la comparación entre elementos, aunque el formato no se presta para una comprensión de grandes datos.

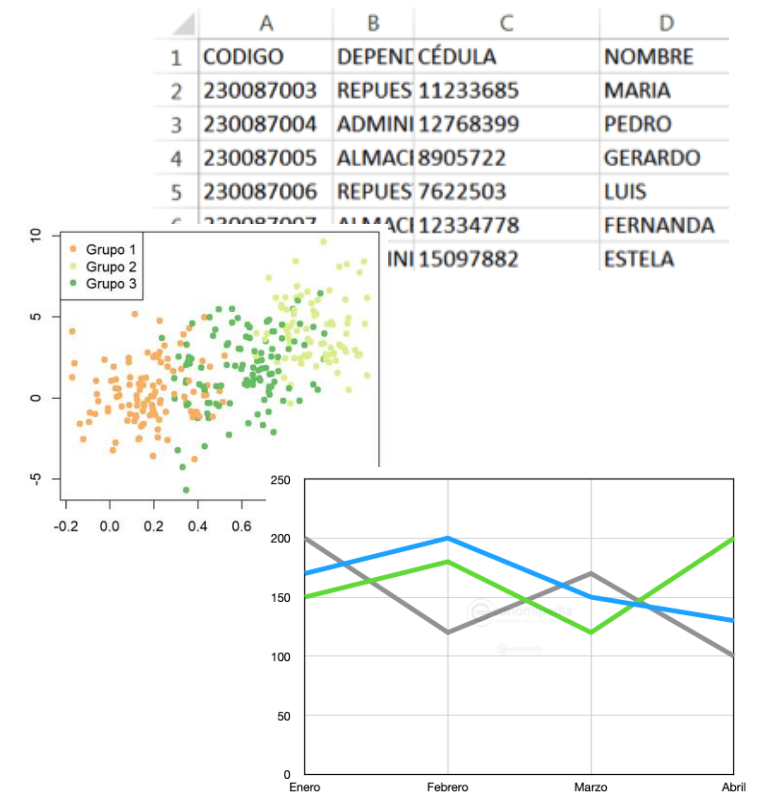


Visualización de datos

Tipos de gráficos

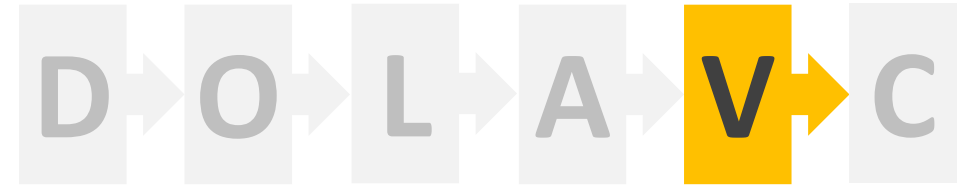


- Tabla: Incluye desde un par de números en una hoja de Excel hasta cientos de filas y columnas. Sirven para mostrar los números casi a secas.
- Gráfico de dispersión: Interesantes si el objetivo es mostrar la relación entre diferentes puntos de datos. Utiliza valores numéricos para ambos ejes.
- Gráfico de Línea: Son útiles para enseñar tendencias, especialmente si están al alza. Muestra las relaciones de los cambios en los datos en un período de tiempo.

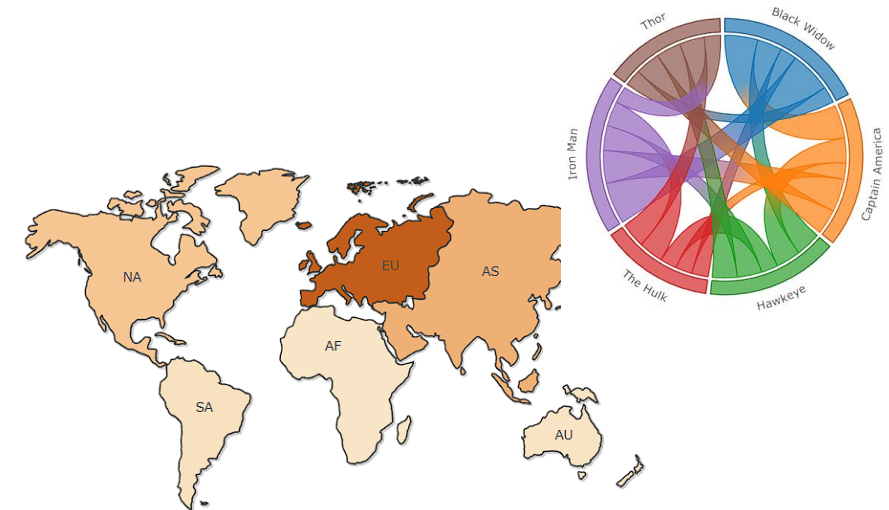
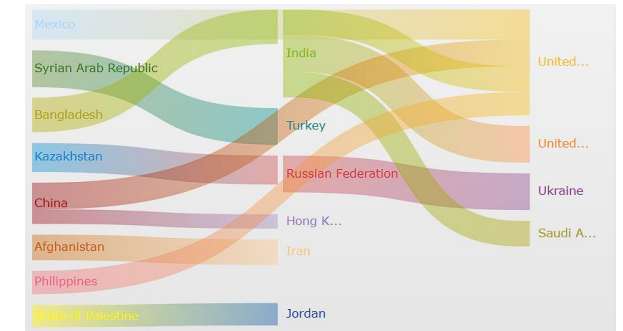


Visualización de datos

Tipos de gráficos

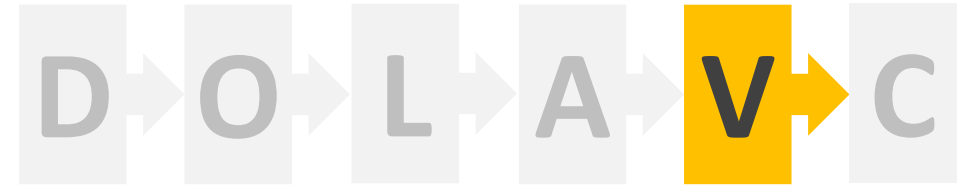


- Sankey: Muestra la evolución temporal de un conjunto de datos, reflejando relevancia, transformaciones, etc.
- Cuerdas: Otro tipo de gráfico social, muestra las relaciones entre individuos, así como su número e intensidad
- Mapa: Asigna valores a distintas regiones geográficas (países, estados, comunidades, ciudades, etc.)

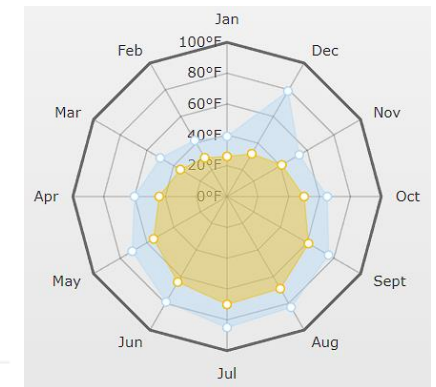
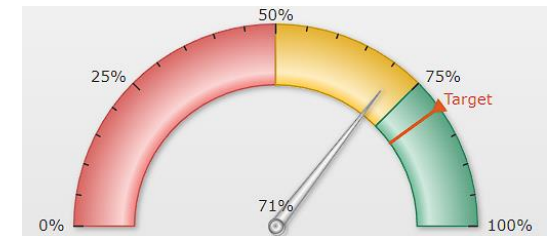


Visualización de datos

Tipos de gráficos

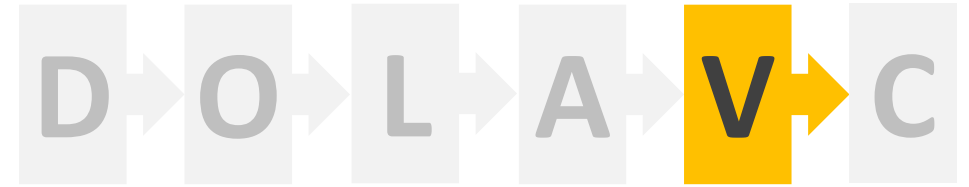


- Angular/Dial: Muestra el rendimiento de una determinada variable, y lo sitúa en una región en función de una serie de umbrales.
- Radar/Araña: Mezcla cualidades de los diagramas de líneas/barras con las de los gráficos circulares.
- Acciones: Empleados generalmente para analizar los rangos de datos experimentados por acciones en el mercado de valores a lo largo del tiempo.

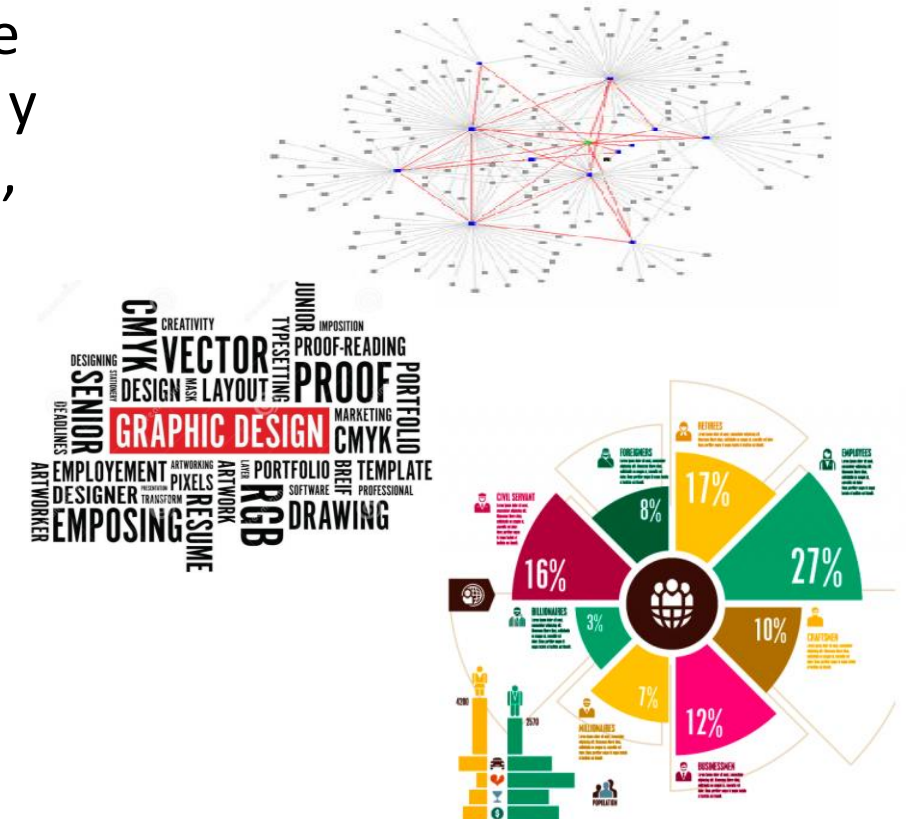


Visualización de datos

Tipos de gráficos

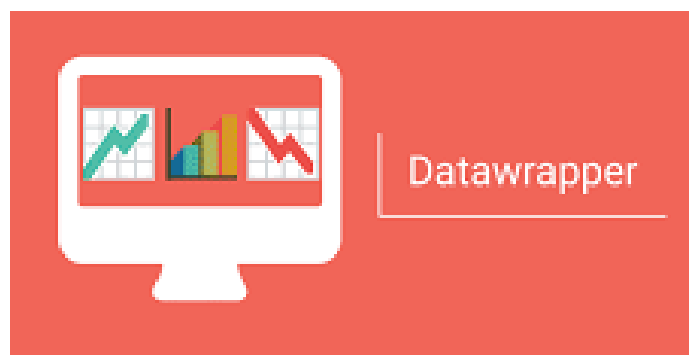
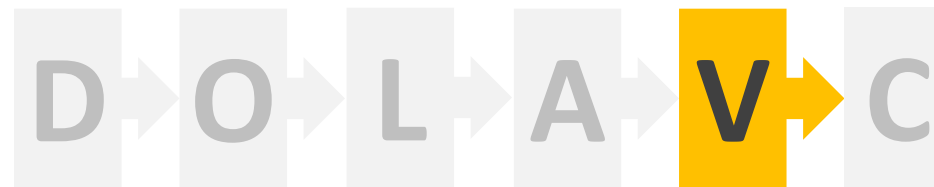


- Gráficos sociales: Es como un mapa global que muestra con quién se relacionan las personas y constan de nodos, que son personas y flechas, que son relaciones que conectan los nodos.
- Palabras: Son nubes de palabras o tags para descubrir tendencias.
- Infografía: Utiliza los datos para compartir y difundir información, así como para generar discusión. Suelen utilizarse con el objetivo de generar tráfico y enlaces a una página web.



Visualización de datos

Algunas herramientas





COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

Ejercicio:

*Escoger una gráfica para los
objetivos diseñados*





COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

Ejercicio en grupo:

*Diseñar una infografía que
represente 2024 en una empresa*





COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

¡Muchas gracias!

