

</

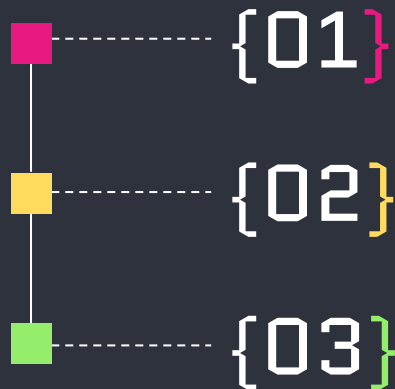
# Uso de weka

/>

Equipo 9

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

# </Indice



Instalación

..... Demostramos la  
instalación de  
weka

Preparación

..... Convertimos un  
archivo csv a arff

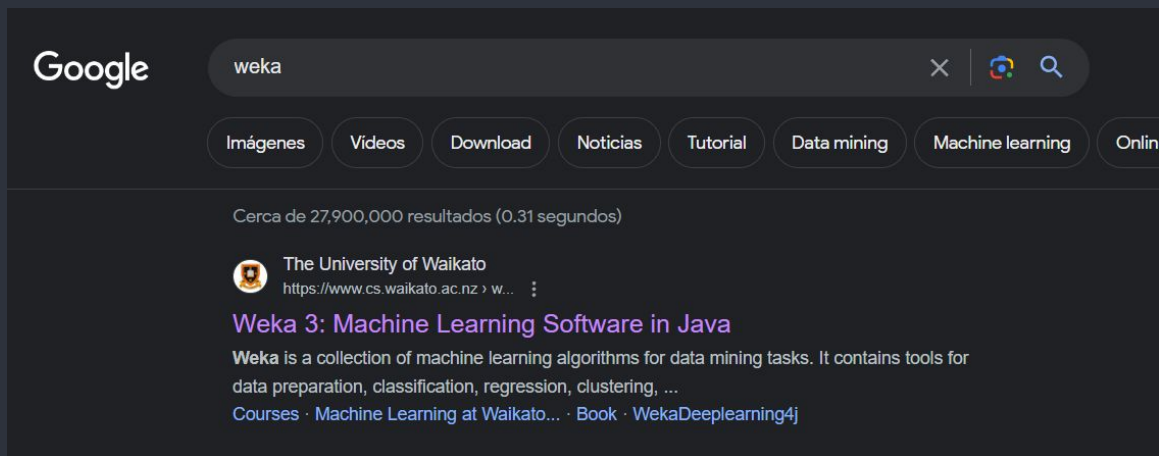
Ejemplo

..... Mostramos un  
ejemplo de uso en  
weka

# </Instalando la herramienta

- Escribimos “weka” en nuestro navegador

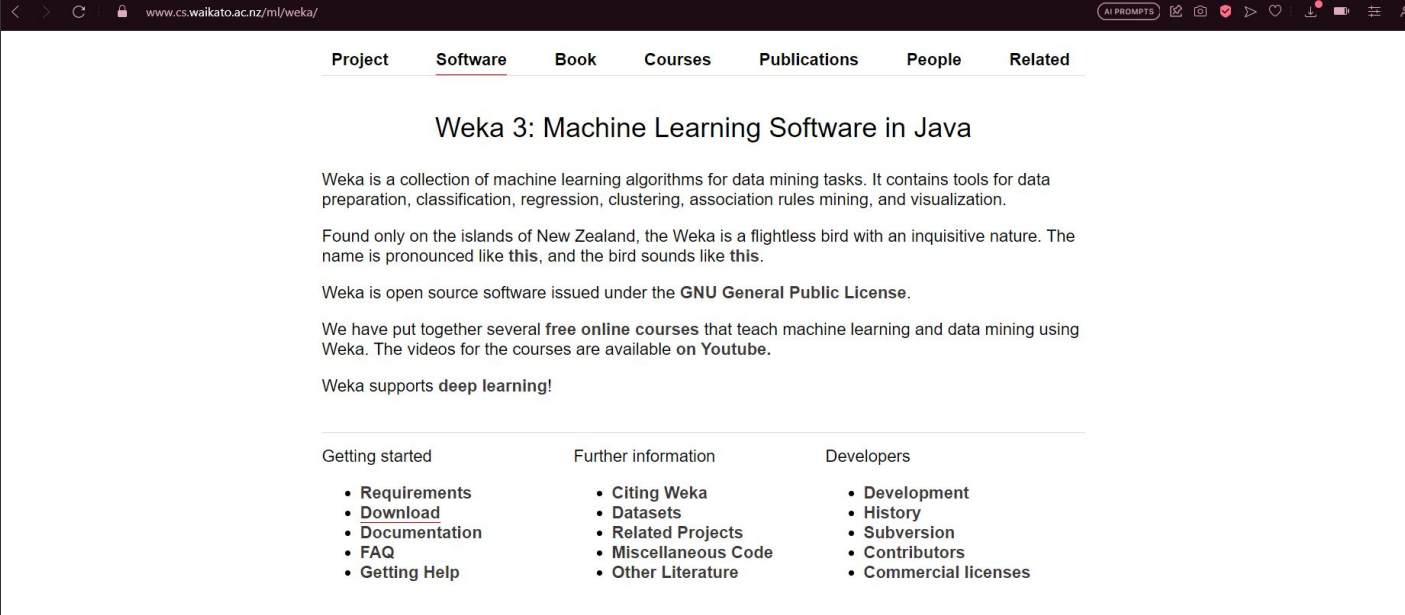
- 



1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 1 0 1

# </Instalando la herramienta

- Nos vamos a la tercera sección de la página, en la columna de “Getting started” seleccionamos la opción “Download”.



The screenshot shows the Weka 3 website. The browser address bar displays 'www.cs.waikato.ac.nz/ml/weka/'. The navigation menu includes 'Project', 'Software' (which is underlined), 'Book', 'Courses', 'Publications', 'People', and 'Related'. The main heading is 'Weka 3: Machine Learning Software in Java'. Below this, there is descriptive text about Weka as a collection of machine learning algorithms and its origin in New Zealand. It also mentions that Weka is open source software under the GNU General Public License and provides links to free online courses and a deep learning support page. At the bottom, there are three columns of links: 'Getting started' (Requirements, Download, Documentation, FAQ, Getting Help), 'Further information' (Citing Weka, Datasets, Related Projects, Miscellaneous Code, Other Literature), and 'Developers' (Development, History, Subversion, Contributors, Commercial licenses).

Project Software Book Courses Publications People Related

## Weka 3: Machine Learning Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like this, and the bird sounds like this.

Weka is open source software issued under the **GNU General Public License**.

We have put together several **free online courses** that teach machine learning and data mining using Weka. The videos for the courses are available **on Youtube**.

Weka supports **deep learning**!

---

Getting started	Further information	Developers
<ul style="list-style-type: none"><li>• Requirements</li><li>• <u>Download</u></li><li>• Documentation</li><li>• FAQ</li><li>• Getting Help</li></ul>	<ul style="list-style-type: none"><li>• Citing Weka</li><li>• Datasets</li><li>• Related Projects</li><li>• Miscellaneous Code</li><li>• Other Literature</li></ul>	<ul style="list-style-type: none"><li>• Development</li><li>• History</li><li>• Subversion</li><li>• Contributors</li><li>• Commercial licenses</li></ul>

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 1 0 1

# </Instalando la herramienta

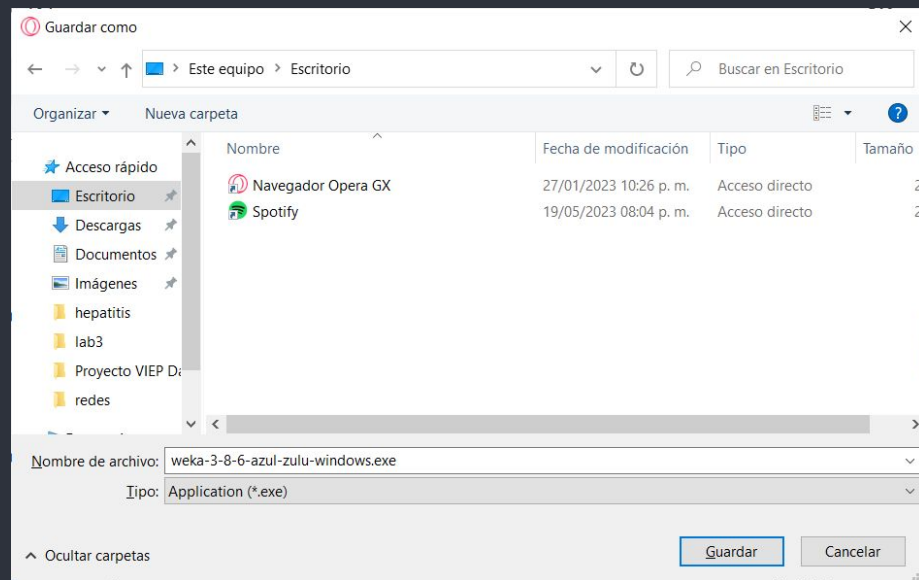
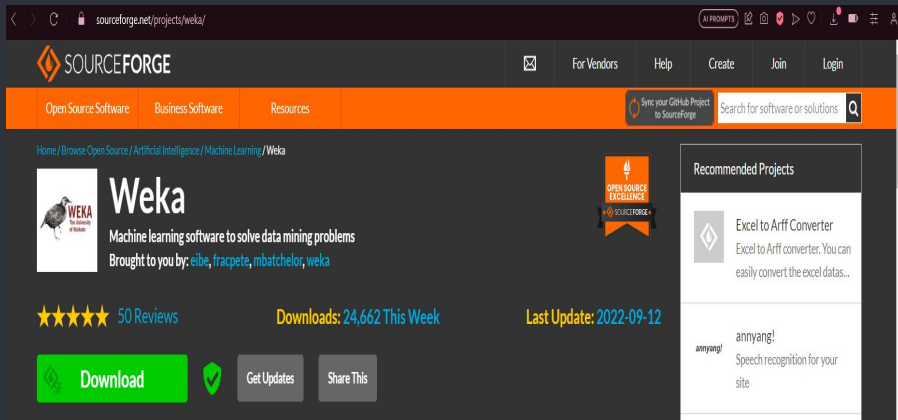
- Nos abrirá una nueva página en la cual tenemos que ir al apartado de “Stable version”, aquí escogeremos la opción acorde a nuestro sistema operativo.

The screenshot shows a web browser window with the URL `waikato.github.io/weka-wiki/downloading_weka/`. The page title is "Downloading and installing Weka". The left sidebar contains a "Weka Wiki" menu with links: Home, Downloading and installing Weka (active), Requirements, Documentation, Getting help, Citing Weka, Literature, Development, History, and Resources. The main content area is titled "Stable version" and contains the following text: "Weka 3.8 is the latest stable version of Weka. This branch of Weka only receives bug fixes and upgrades that do not break compatibility with earlier 3.8 releases, although major new features may become available in packages. There are different options for downloading and installing it on your system:". Below this, under the heading "WINDOWS", there is a bullet point: "Click [here](#) to download a self-extracting executable for 64-bit Windows that includes Azul's 64-bit OpenJDK Java VM 17 (weka-3-8-6-azul-zulu-windows.exe; 133.2 MB)". At the bottom of the main content area, it states: "This executable will install Weka in your Program Menu. Launching via the Program Menu or shortcuts will automatically use the included JVM to run Weka." The right sidebar contains a "Table of contents" with links: Snapshots, Stable version (active), Windows, Mac OS - Intel processors, Mac OS - ARM processors, Linux, Other platforms, Developer version, Windows, and Mac OS - Intel processors. The browser's address bar and various icons are visible at the top.

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1

## </Instalando la herramienta

- En nuestro caso escogimos la opción de windows, por lo tanto realizamos la descarga como cualquier aplicación en este SO.



1 0 1 1    0 1 1    0 1    1 0 1 1 0 0 1    1 0    1 1 0 1 1    0 1 1    0 1    1 1 0 1 1 0    1 1 0 1 1 1    1 1 0 1

# </Instalando la herramienta

- Abrimos el archivo .exe, se nos abrirá una ventana emergente para la instalación, daremos next durante todo el proceso de instalación(Si deseas modificar algo puedes hacerlo).
- Damos next a todo.



1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1

# </Conversión de formato

- Primero tenemos que descargar un dataset para trabajar.

The screenshot shows the UC Irvine Machine Learning Repository page for the BuddyMove Data Set. The page is titled "BuddyMove Data Set" and includes a "DONATED" badge. The dataset is described as "User interest information extracted from user reviews published in holidayiq.com about various types of point of interests in South India". It is a "Multivariate, Text" dataset with "249" instances and "0" features. The associated tasks are "Classification, Clustering". The dataset is licensed under a "Creative Commons Attribution 4.0 International (CC BY 4.0) license". The page also includes a "DOWNLOAD" button, a "CITE" button, and a "Login" link.

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate, Text	Other	Classification, Clustering

Feature Type	# Instances	# Features
Real	249	-

**Dataset Information**

**Additional Information**  
This dataset was populated from destination reviews published by 249 reviewers of holidayiq.com till October 2014. Reviews falling in 6 categories among destinations across South India were considered and the count of reviews in each category for every reviewer (traveler) is captured.

**Has Missing Values?**  
No

**Creators**  
Shini Renjith

**DOI**  
10.24432/C5N316

**License**  
This dataset is licensed under a **Creative Commons Attribution 4.0 International (CC BY 4.0)** license.  
This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given

<http://archive.ics.uci.edu/dataset/476/buddymove+data+set>

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 0 1



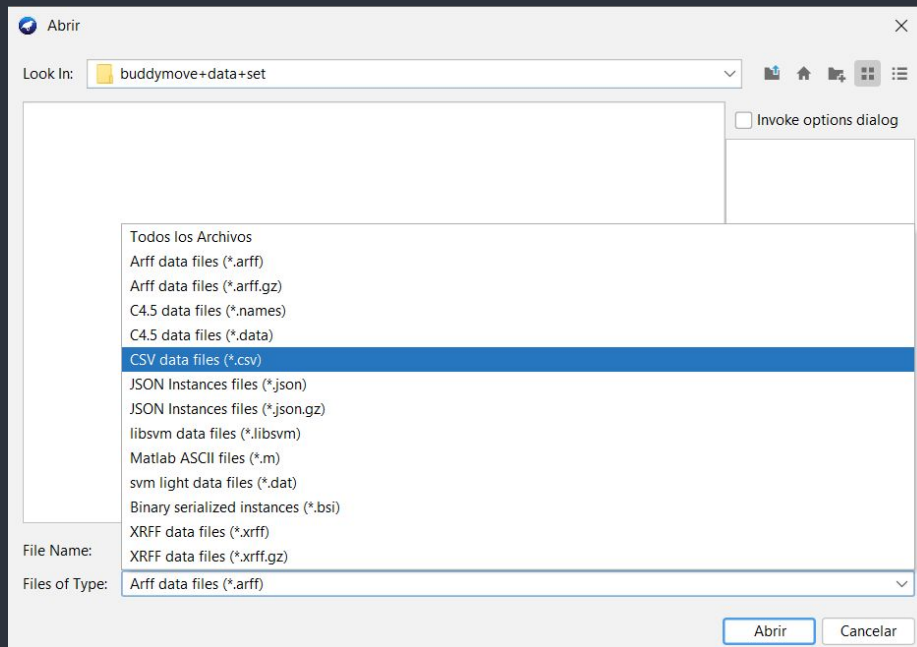
# </Conversión de formato

- Weka admite varios tipos de archivos, pero el formato propio de la aplicación (y con el que mejor se entiende) es .arff (Attribute-Relation File Format).
- Por lo tanto procederemos a explicar cómo pueden transformar un archivo csv a un arff con la misma herramienta.



# </Conversión de formato

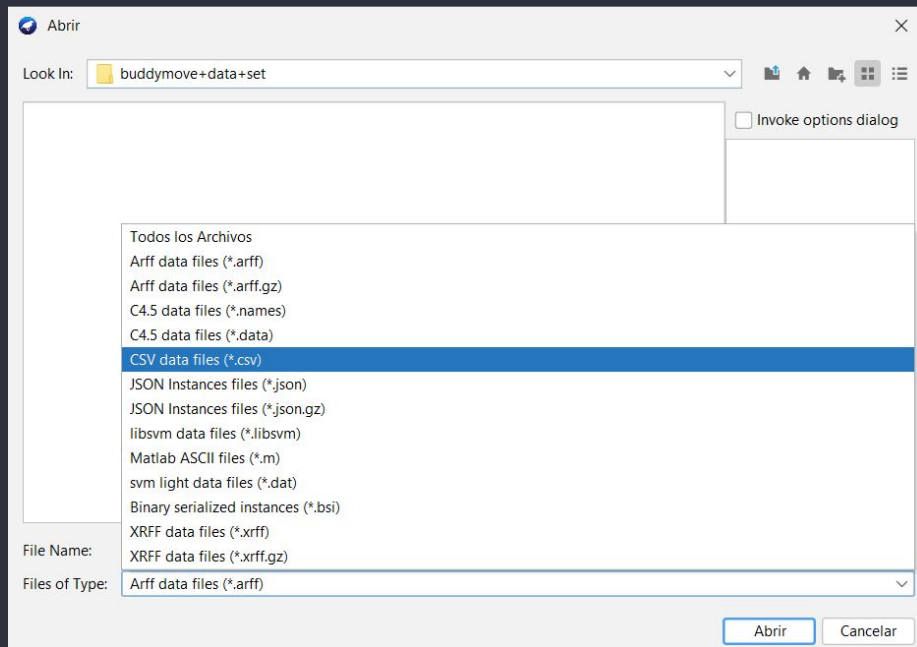
- Cambiamos el tipo de datos a CSV



1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

# </Conversión de formato

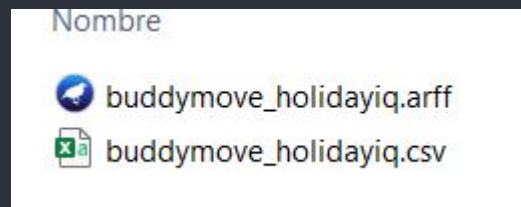
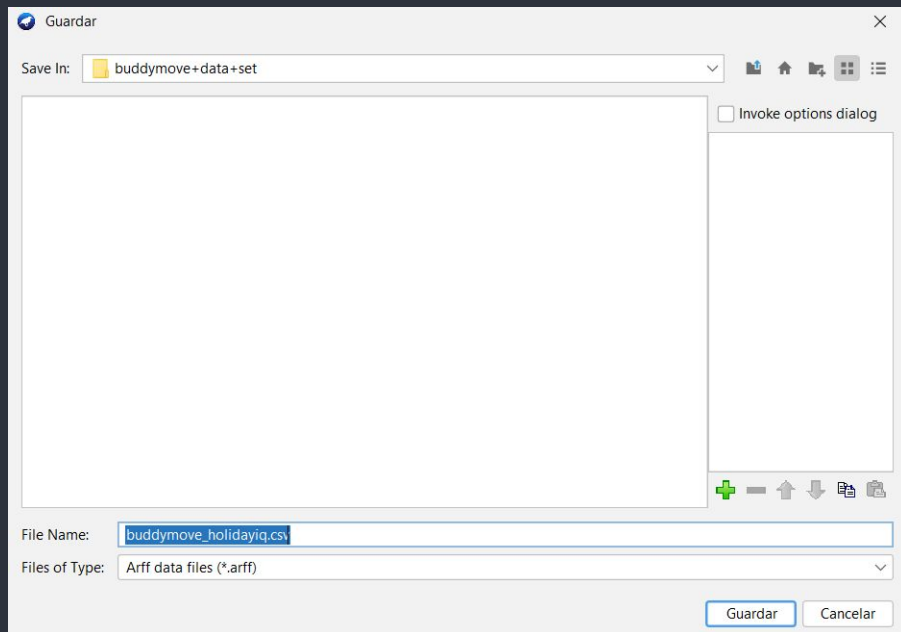
- Cambiamos el tipo de datos a CSV y seleccionamos nuestro archivo.



1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

# </Conversión de formato

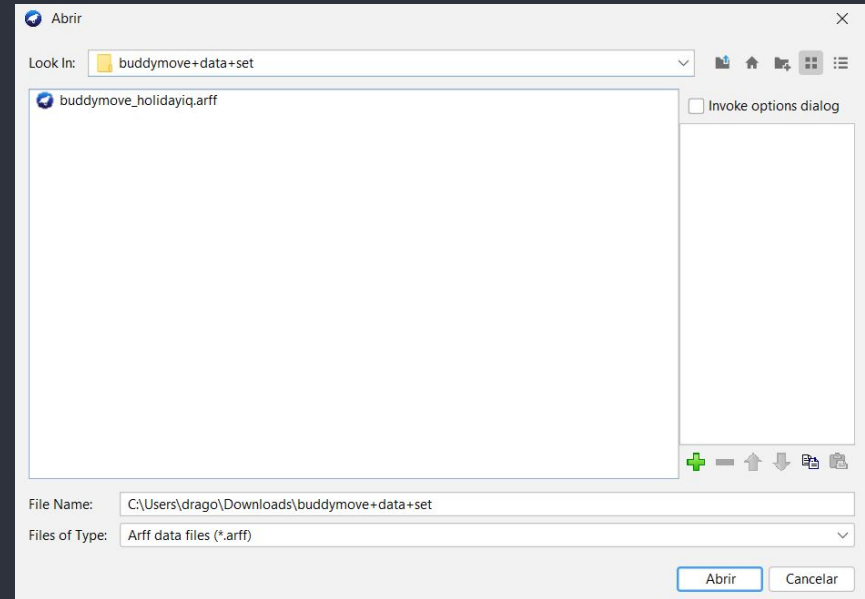
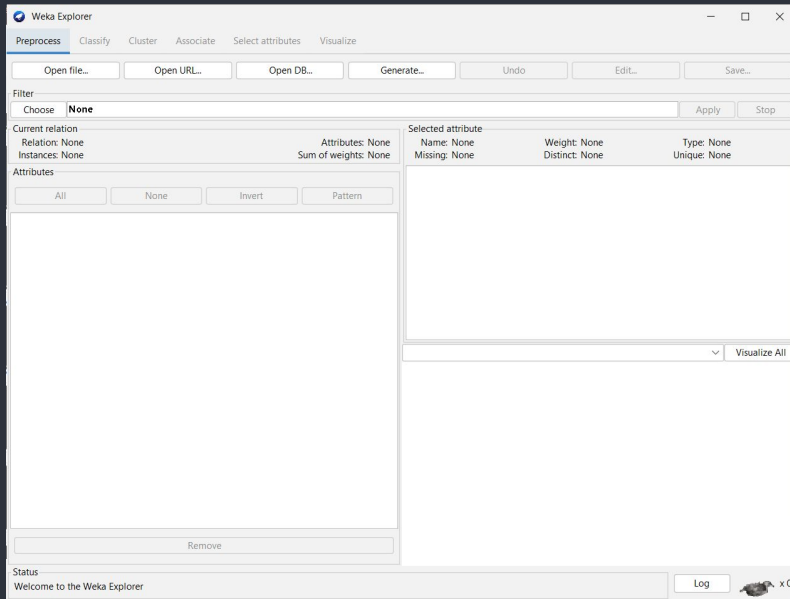
- Guardamos el archivo como ARFF



1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 1 1 0 1

# </Ejemplo de uso

- Primero damos en la opción “Explorer” en el menú principal, después abrimos nuestro archivo ARFF, dando a la opción “Open File” dentro de la ventana “Explorer”

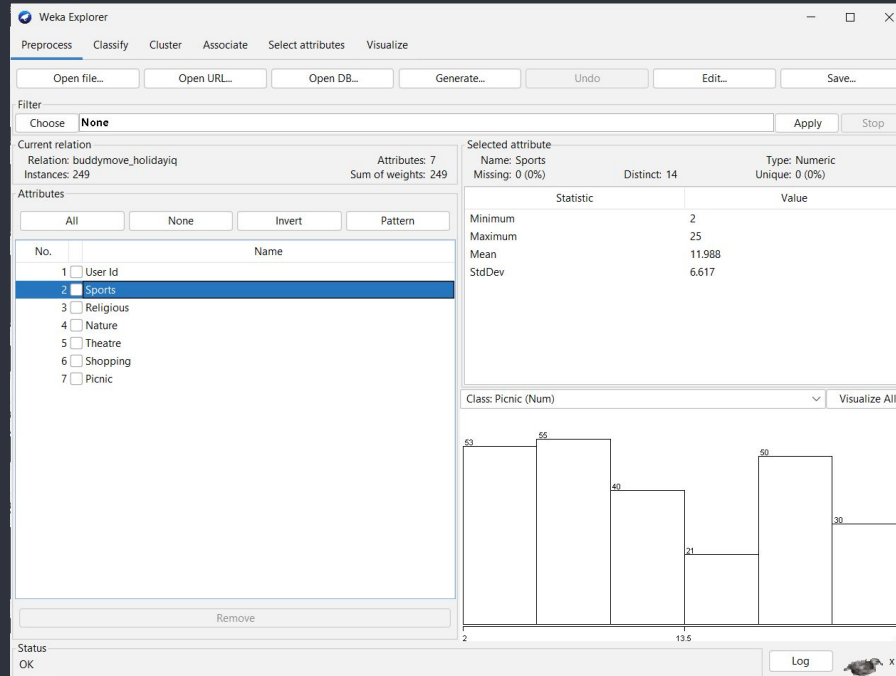


1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1

0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

# </Ejemplo de uso

- Podemos visualizar nuestras características o atributos.



1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 1 1 1 0 1

# </Ejemplo de uso

- En la pestaña “Preprocess” podemos manipular nuestros atributos

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** | Apply | Stop

Current relation: buddymove\_holidayiq  
Instances: 249  
Attributes: 7  
Sum of weights: 249

Attributes: All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> User Id
2	<input type="checkbox"/> Sports
3	<input type="checkbox"/> Religious
4	<input type="checkbox"/> Nature
5	<input type="checkbox"/> Theatre
6	<input type="checkbox"/> Shopping
7	<input type="checkbox"/> Picnic

Selected attribute:  
Name: User Id  
Missing: 0 (0%)  
Distinct: 249  
Type: Nominal  
Unique: 249 (100%)

No.	Label	Count	Weight
1	User 1	1	1
2	User 2	1	1
3	User 3	1	1
4	User 4	1	1
5	User 5	1	1
6	User 6	1	1
7	User 7	1	1
8	User 8	1	1
9	User 9	1	1
10	User 10	1	1

Class: Picnic (Num) | Visualize All

Remove selected attributes.

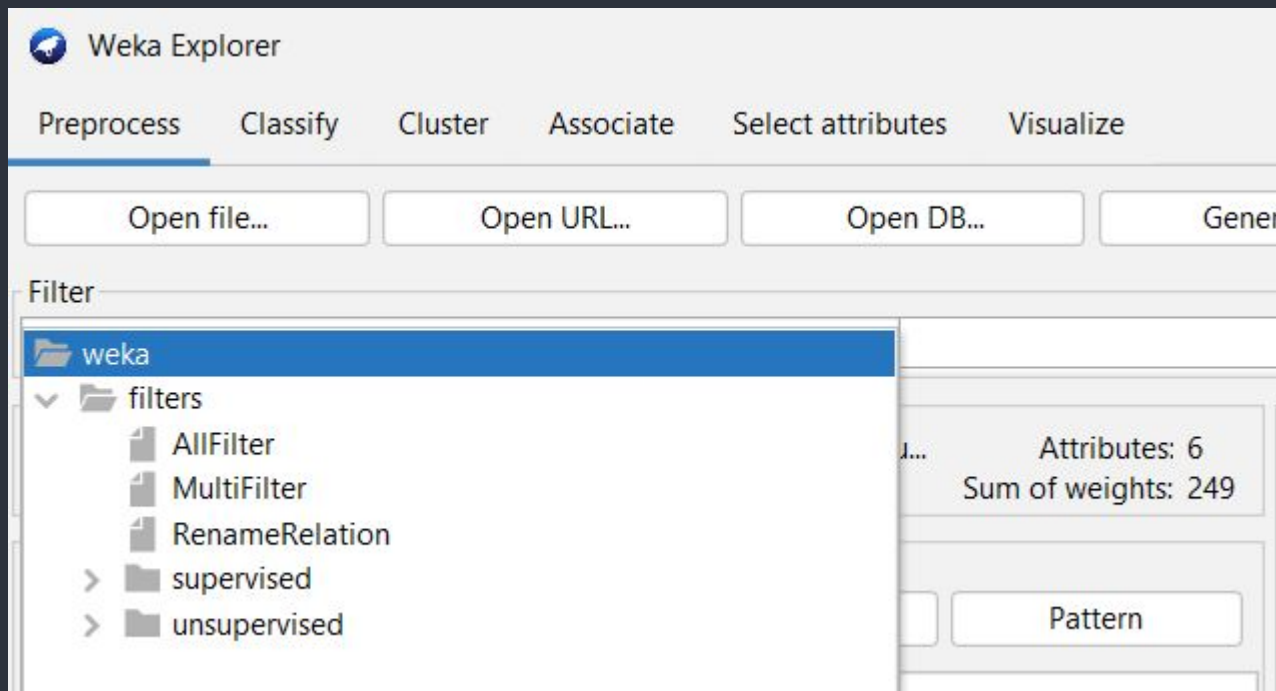
Status: OK | Log | x 0

No.	Name
1	<input type="checkbox"/> Sports
2	<input type="checkbox"/> Religious
3	<input type="checkbox"/> Nature
4	<input type="checkbox"/> Theatre
5	<input type="checkbox"/> Shopping
6	<input type="checkbox"/> Picnic

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 1 1 1 0 1

# </Ejemplo de uso

- Podemos aplicar una grana variedad de filtros para poder preparar nuestros datos para clasificarlos o agruparlos.

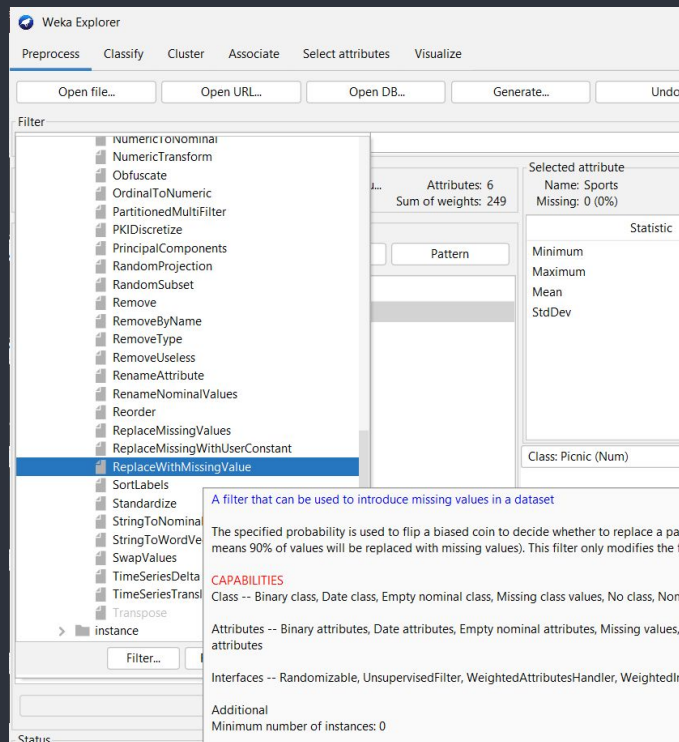


1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1



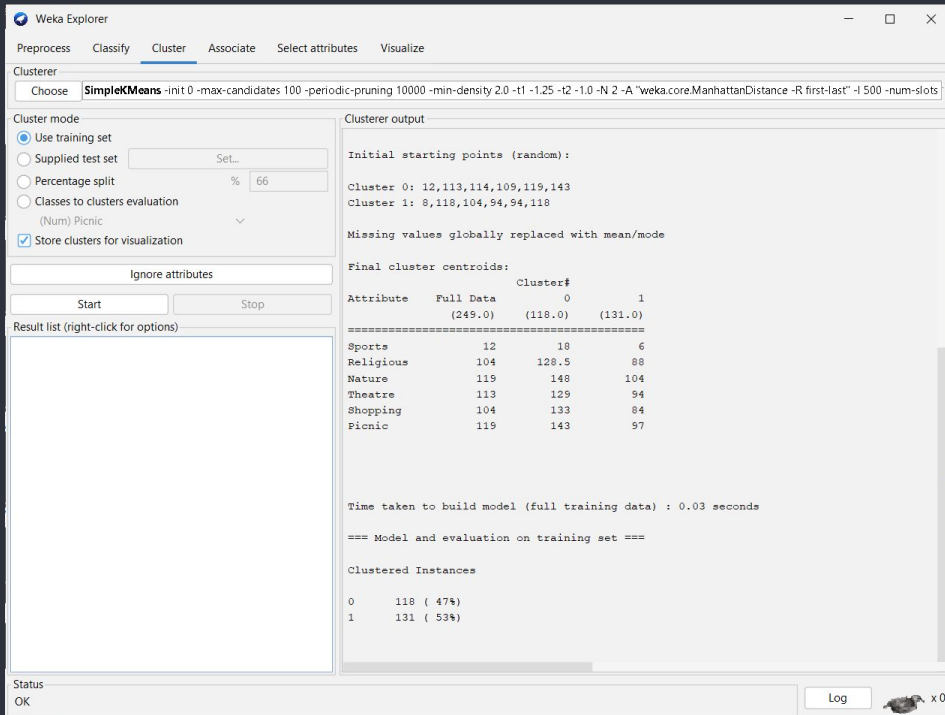
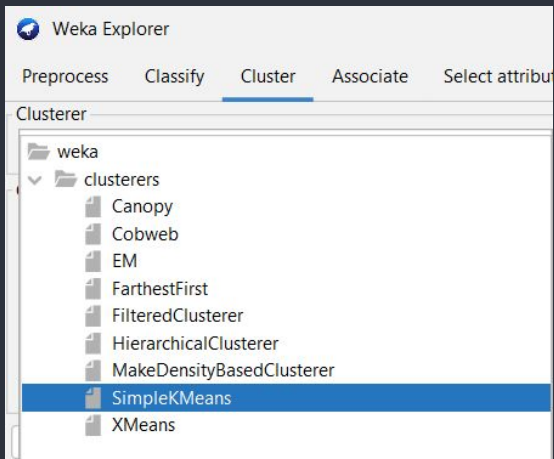
## </Ejemplo de uso

- En este caso vamos a agregar un filtro que coloca valores faltantes dentro de nuestro dataset


$$1 \ 0 \ 1 \ 1 \quad 0 \ 1 \ 1 \quad 0 \ 1 \quad 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \quad 1 \ 0 \quad 1 \ 1 \ 0 \ 1 \ 1 \quad 0 \ 1 \ 1 \quad 0 \ 1 \quad 1 \ 1 \ 0 \ 1 \ 1 \ 0 \quad 1 \ 1 \ 0 \ 1 \ 1 \ 1 \quad 1 \ 1 \ 0 \ 1$$

## </Ejemplo de uso

- Nos pasamos a la pestaña de “Cluster”.
- En el botón “Choose” seleccionamos “SimpleKMeans”


$$1 \ 0 \ 1 \ 1 \quad 0 \ 1 \ 1 \quad 0 \ 1 \quad 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \quad 1 \ 0 \quad 1 \ 1 \ 0 \ 1 \ 1 \quad 0 \ 1 \ 1 \quad 0 \ 1 \quad 1 \ 1 \ 0 \ 1 \ 1 \ 0 \quad 1 \ 1 \ 0 \ 1 \ 1 \ 1 \quad 1 \ 1 \ 0 \ 1$$

# </Ejemplo de uso

- Damos clic en la barra donde aparece la opción que seleccionamos

```
SimpleKMeans -init 0 -max-candidates 100 -periodic-
```

- De este modo aparecerá una pestaña con las opciones para el algoritmo que seleccionamos

The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.clusterers.SimpleKMeans' class. The 'About' tab is active, displaying the description 'Cluster data using the k means algorithm.' and buttons for 'More' and 'Capabilities'. Below this, various configuration parameters are listed with their current values:

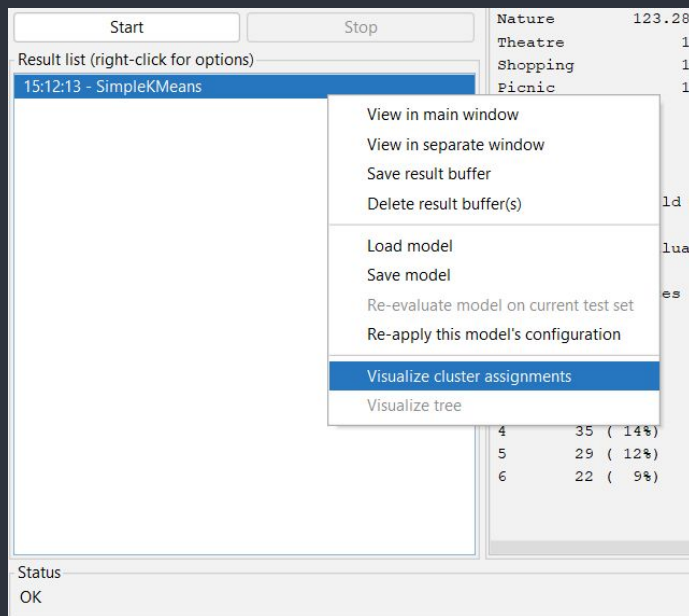
Parameter	Value
canopyMaxNumCanopiesToHoldInMemory	100
canopyMinimumCanopyDensity	2.0
canopyPeriodicPruningRate	10000
canopyT1	-1.25
canopyT2	-1.0
debug	False
displayStdDevs	False
distanceFunction	Choose <b>ManhattanDistance</b>
doNotCheckCapabilities	False
dontReplaceMissingValues	False
fastDistanceCalc	False
initializationMethod	Random
maxIterations	500
numClusters	2
numExecutionSlots	1
preserveInstancesOrder	False
reduceNumberOfDistanceCalcsViaCanopies	False
seed	10

At the bottom of the window are four buttons: 'Open...', 'Save...', 'OK', and 'Cancel'.

1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1

# </Ejemplo de uso

- Hacemos click en start y después click derecho sobre nuestra ejecución que queremos analizar, seguido de este seleccionamos la opción “Visualize cluster assignments”



# </Ejemplo de uso

- Se nos abrirá una nueva ventana donde podremos ver una gráfica a la cual podemos ir seleccionando los atributos que queremos visualizar, además de otras opciones para una buena visualización de los datos.

