


**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

## 2.5. Lectura/escritura archivos CSV

### Características de los datos en ficheros CSV en Python:

- Pueden incluir una cabecera con los nombres de las columnas
- Pueden incluir comentarios (# al comienzo de una línea).
- Pueden usar diferentes tipos de delimitadores (el estándar es la coma, pero pueden usar punto y coma, tabulaciones, etc).
- Los valores que contiene el fichero CSV puede que tengan espacios En ese caso, deberían venir entre comillas dobles.

Podemos copiar todos los datos que aparecen en un Excel. Bien guardarlo en este caso lo guardamos como CSV, y también podemos guardarlo como libro de Excel.

### Dataset para pruebas:

- Dataset “Algerian Forest Fires”:

[Algerian forest dataset](#)

- Dataset “Wine Quality”:

[Wine quality dataset](#)

- Dataset “Cars” (aula en el Campus Virtual).

### Cargar CSV desde URL en Python

- Dimensiones: `df.shape()`
- Primeras filas: `df.head(valor)`
- Últimas filas: `df.tail(valor)`

### Características de los datos en ficheros CSV en MATLAB, lectura de datos:

Se usa `readTable()` o `readmatrix()` para datos numericos, antes se usaba `csvread()` y también `xlsread()` pero ya están obsoletas.

```
A = readmatrix(filename)
A = readmatrix(filename,opts)
A = readmatrix(___,Name,Value)
```


### Description

#### Ejemplo:

`A = readmatrix(filename)` creates an array by reading column-oriented data from a file. The `readmatrix` function performs automatic detection of import parameters for your file.

`readmatrix` determines the file format from the file extension:

- `.txt`, `.dat`, or `.csv` for delimited text files
- `.xls`, `.xlsb`, `.xlsm`, `.xlsx`, `.xltm`, `.xltx`, or `.ods` for spreadsheet files

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

For files containing mixed numeric and text data, readmatrix imports the data as a numeric array by default.

Ejemplo:

```
A = readmatrix(filename,opts) additionally uses the import options opts.
```

Podemos copiar todos los datos que aparecen en las bases de datos en un Excel o venir directamente en formato Excel. Podemos guardarlo como CSV, o bien podemos guardarlo como libro de Excel. Una vez en Excel podemos separar esta base de datos en diferentes columnas para poder manejarlo más cómodamente en MATLAB, tratando a esta base de datos como si fuera una matriz, con sus filas y sus columnas.

Hay que tener en cuenta que, al trabajar con una base de datos en Excel, si queremos calcular con números tendremos que quitar las cabeceras de las columnas pues suelen ser letras.

En esta práctica trabajaremos con esta base de datos en CSV.

**Nota:** Se adjuntan estas bases de datos en el Campus Virtual.


#### Ejercicio 75\_01. Cargar datos CSV desde URL en Python

```
import pandas as pd
# Importante: Nos permite continuar La URL en La línea de abajo
url = ("https://archive.ics.uci.edu/ml/machine-learning-
databases/00547"/"Algerian_forest_fires_dataset_UPDATE.csv")
# En La primera fila aparece un título. La saltamos con skiprows = 1
# En La segunda hay un encabezado con Los nombres de Las columnas con lo que no
necesitamos añadirlos
datos = pd.read_csv(url, skiprows = 1)
print('Dimensiones de los datos cargados: ')
print(datos.shape)
print('\n Mostramos las 5 primeras filas: ')
print(datos.head)
print('\n Mostramos las 2 primeras filas: ')
print(datos.head(2))
print('\n Mostramos las 5 últimas filas: ')
print(datos.tail)
print('\n Mostramos las 2 últimas filas: ')
print(datos.tail(2))
```

Dimensiones de los datos cargados:  
(246, 14)

Mostramos las 5 primeras filas:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
0	1	06	2012	29	57	18	0	65.7	3.4	7.6
1	2	06	2012	29	61	13	1.3	64.4	4.1	7.6
2	3	06	2012	26	82	22	13.1	47.1	2.5	7.1
3	4	06	2012	25	89	13	2.5	28.6	1.3	6.9
4	5	06	2012	27	77	16	0	64.8	3	14.2

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](http://www.unicadiz.es)

ISI	BUI	FWI	Classes
1.3	3.4	0.5	not fire
1	3.9	0.4	not fire
0.3	2.7	0.1	not fire
0	1.7	0	not fire
1.2	3.9	0.5	not fire

Mostramos las 2 primeras filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	
0	1	06	2012	29	57	18	0	65.7	3.4	7.6
1	2	06	2012	29	61	13	1.3	64.4	4.1	7.6

ISI	BUI	FWI	Classes
1.3	3.4	0.5	not fire
1	3.9	0.4	not fire

Mostramos las 5 últimas filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	
241	26	9	2012	30	65	14	0	85.4	16	44.5
242	27	9	2012	28	87	15	4.4	41.1	6.5	8
243	28	9	2012	27	87	29	0.5	45.9	3.5	7.9
244	29	9	2012	24	54	18	0.1	79.7	4.3	15.2
245	30	9	2012	24	64	15	0.2	67.3	3.8	16.5

ISI	BUI	FWI	Classes	
241	4.5	16.9	6.5	fire
242	0.1	6.2	0	not fire
243	0.4	3.4	0.2	not fire
244	1.7	5.1	0.7	not fire
245	1.2	4.8	0.5	not fire


Mostramos las 2 últimas filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	
244	29	9	2012	24	54	18	0.1	79.7	4.3	15.2
245	30	9	2012	24	64	15	0.2	67.3	3.8	16.5

ISI	BUI	FWI	Classes	
244	1.7	5.1	0.7	not fire
245	1.2	4.8	0.5	not fire

### Ejercicio 75\_02. Cargar datos CSV desde URL en MATLAB

```
% URL del conjunto de datos
url = 'https://archive.ics.uci.edu/ml/machine-learning-
databases/00547/Algerian_forest_fires_dataset_UPDATE.csv';
% Cargar los datos desde la URL
datos = readtable(url, 'ReadVariableNames', true);
% Mostrar dimensiones de los datos cargados
disp('Dimensiones de los datos cargados:');
disp(size(datos));
% Mostrar las 5 primeras filas
```

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](http://www.unicadiz.es)

```
disp('Mostramos las 5 primeras filas:');
disp(datos(1:5, :));
% Mostrar las 2 primeras filas
disp('Mostramos las 2 primeras filas:');
disp(datos(1:2, :));
% Mostrar las 5 últimas filas
disp('Mostramos las 5 últimas filas:');
disp(datos(end-4:end, :));
% Mostrar las 2 últimas filas
disp('Mostramos las 2 últimas filas:');
disp(datos(end-1:end, :));
```

Dimensiones de los datos cargados:

246 14

Mostramos las 5 primeras filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
1	6	2012	29	57	18	0	65.7	3.4	7.6
2	6	2012	29	61	13	1.3	64.4	4.1	7.6
3	6	2012	26	82	22	13.1	47.1	2.5	7.1
4	6	2012	25	89	13	2.5	28.6	1.3	6.9
5	6	2012	27	77	16	0	64.8	3	14.2

ISI	BUI	FWI	Classes
1.3	3.4	0.5	{'not fire'}
1	3.9	0.4	{'not fire'}
0.3	2.7	0.1	{'not fire'}
0	1.7	0	{'not fire'}
1.2	3.9	0.5	{'not fire'}

Mostramos las 2 primeras filas:


day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
1	6	2012	29	57	18	0	65.7	3.4	7.6
2	6	2012	29	61	13	1.3	64.4	4.1	7.6

ISI	BUI	FWI	Classes
1.3	3.4	0.5	{'not fire'}
1	3.9	0.4	{'not fire'}

Mostramos las 5 últimas filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
26	9	2012	30	65	14	0	85.4	16	44.5
27	9	2012	28	87	15	4.4	41.1	6.5	8
28	9	2012	27	87	29	0.5	45.9	3.5	7.9
29	9	2012	24	54	18	0.1	79.7	4.3	15.2
30	9	2012	24	64	15	0.2	67.3	3.8	16.5

ISI	BUI	FWI	Classes
4.5	16.9	6.5	{'fire' }

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

0.1	6.2	0	{'not fire'}
0.4	3.4	0.2	{'not fire'}
1.7	5.1	0.7	{'not fire'}
1.2	4.8	0.5	{'not fire'}

Mostramos las 2 últimas filas:

day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
29	9	2012	24	54	18	0.1	79.7	4.3	15.2
30	9	2012	24	64	15	0.2	67.3	3.8	16.5

ISI	BUI	FWI	Classes
1.7	5.1	0.7	{'not fire'}
1.2	4.8	0.5	{'not fire'}

### Ejercicio 76\_01. Cargar datos CSV desde fichero en Python


```
import pandas as pd
# Cargamos un primer dataset de calidad de vinos:
nomFichero = 'winequality-red.csv'
# No es necesario añadir nombres de columnas, ya que el csv los trae.
# Se usa como separador, y no la coma por defecto
datosWines = pd.read_csv(nomFichero, sep=";")
print('Mostramos dimensiones: ')
print(datosWine.shape)
print('Imprimimos 3 primeras filas de los datos que hemos cargado: ')
print(datosWine.head(3))
# Cargamos otro dataset de coches:
# Una inspección del csv abriéndolo con el notepad nos indica que la separación es
# por comas (La que read_csv espera por defecto), y que no hay header con los nombres
# de las variables. Los pondremos en código, y son los siguientes: buying, maint,
# doors, persons, lug_boot, safety, category
nomFichero = 'cars.csv'
nombresVars = ['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety',
               'category']
datosCars = pd.read_csv(nomFichero, names=nombresVars)
# datosCars = pd.read_csv(nomFinero)
print('Mostramos dimensiones: ')
print(datosCars.shape)
print('Imprimimos 3 primeras filas de los datos que hemos cargado: ')
print(datosCars.head(3))
```

Mostramos dimensiones del conjunto de datos de vinos:

1599      12

Imprimimos 3 primeras filas del conjunto de datos de vinos:

fixedAcidity	volatileAcidity	citricAcid	residualSugar	chlorides
7.4	0.7	0	1.9	0.076
7.8	0.88	0	2.6	0.098
7.8	0.76	0.04	2.3	0.092

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

freeSulfurDioxide	totalSulfurDioxide	density	pH	sulphates	alcohol
11	34	0.9978	3.51	0.56	9.4
25	67	0.9968	3.2	0.68	9.8
15	54	0.997	3.26	0.65	9.8

quality

5
5
5

Mostramos dimensiones del conjunto de datos de coches:  
1728 7

Imprimimos 3 primeras filas del conjunto de datos de coches:

	bying	maint	doors	persons	lug_boot	safety	category
0	vhigh	vhigh	2	2	small	low	unacc
1	vhigh	vhigh	2	2	small	med	unacc
2	vhigh	vhigh	2	2	small	high	unacc


### Ejercicio 76\_02. Cargar datos CSV desde fichero en MATLAB

```
% Cargar el primer conjunto de datos de calidad de vinos
nomFicheroWines = 'winequality-red.csv';
datosWines = readtable(nomFicheroWines, 'Delimiter', ',');
% Mostrar dimensiones
disp('Mostramos dimensiones del conjunto de datos de vinos:');
disp(size(datosWines));
% Mostrar las 3 primeras filas
disp('Imprimimos 3 primeras filas del conjunto de datos de vinos:');
disp(datosWines(1:3, :));
% Cargar el segundo conjunto de datos de coches
nomFicheroCars = 'cars.csv';
nombresVarsCars = {'buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety',
'category'};
datosCars = readtable(nomFicheroCars, 'Delimiter', ',');
% Mostrar dimensiones
disp('Mostramos dimensiones del conjunto de datos de coches:');
disp(size(datosCars));
% Mostrar las 3 primeras filas
disp('Imprimimos 3 primeras filas del conjunto de datos de coches:');
disp(datosCars(1:3, :));
```

Mostramos dimensiones del conjunto de datos de vinos:

fixedAcidity	volatileAcidity	citricAcid	residualSugar	chlorides
7.4	0.7	0	1.9	0.076
7.8	0.88	0	2.6	0.098
7.8	0.76	0.04	2.3	0.092

freeSulfurDioxide	totalSulfurDioxide	density	pH	sulphates	alcohol
-------------------	--------------------	---------	----	-----------	---------

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

11	34	0.9978	3.51	0.56	9.4
25	67	0.9968	3.2	0.68	9.8
15	54	0.997	3.26	0.65	9.8

**quality**

5
5
5

Mostramos dimensiones del conjunto de datos de coches:  
1728 7

Imprimimos 3 primeras filas del conjunto de datos de coches:

bying	maint	doors	persons	lug_boot	safety	category
{'vhigh'}	{'vhigh'}	2	2	{'small'}	{'low' }	{'unacc'}
{'vhigh'}	{'vhigh'}	2	2	{'small'}	{'med' }	{'unacc'}
{'vhigh'}	{'vhigh'}	2	2	{'small'}	{'high'}	{'unacc'}

### Ejercicio 77\_01. Guardar datos en CSV en Python

```
import pandas as pd
df = pd.DataFrame({'Est': ['1', '2'], 'Contaminante': ['N02', 'S02'], 'Tipo':
['Train', 'Test']})
print(df)
# Guardamos con Las cabeceras (header = True)
df.to_csv("prueba2.csv", header = True, index = False)
# Lo cargamos y comprobamos
nomFichero = 'prueba2.csv'
datos = pd.read_csv(nomFichero)
print(datos)
```


	Est	Contaminante	Tipo
0	1	N02	Train
1	2	S02	Test

	Est	Contaminante	Tipo
0	1	N02	Train
1	2	S02	Test

### Ejercicio 77\_02. Guardar datos en CSV en MATLAB

```
% Crear un DataTable similar a un DataFrame en pandas
data = { '1', 'N02', 'Train';
        '2', 'S02', 'Test'};
nombres_columnas = {'Est', 'Contaminante', 'Tipo'};
df = table(data(:, 1), data(:, 2), data(:, 3), 'VariableNames', nombres_columnas);
% Mostrar el DataTable original
disp('DataTable original:');
disp(df);
% Guardar el DataTable en un archivo CSV con cabeceras
writetable(df, 'prueba2.csv', 'WriteVariableNames', true);
```

**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](#)

```
% Cargar el archivo CSV y comprobar
datos = readtable('prueba2.csv');
disp('Datos cargados desde prueba2.csv:');
disp(datos);
```

DataTable original:

<b>Est</b>	<b>Contaminante</b>	<b>Tipo</b>
{'1'}	{'NO2'}	{'Train'}
{'2'}	{'SO2'}	{'Test' }

Datos cargados desde prueba2.csv:

<b>Est</b>	<b>Contaminante</b>	<b>Tipo</b>
1	{'NO2'}	{'Train'}
2	{'SO2'}	{'Test' }