

# Análisis avanzado de datos con Python. Ejemplo de carga de datos

Iker Martín, Maribel Castillo y Vicente R. Tomás

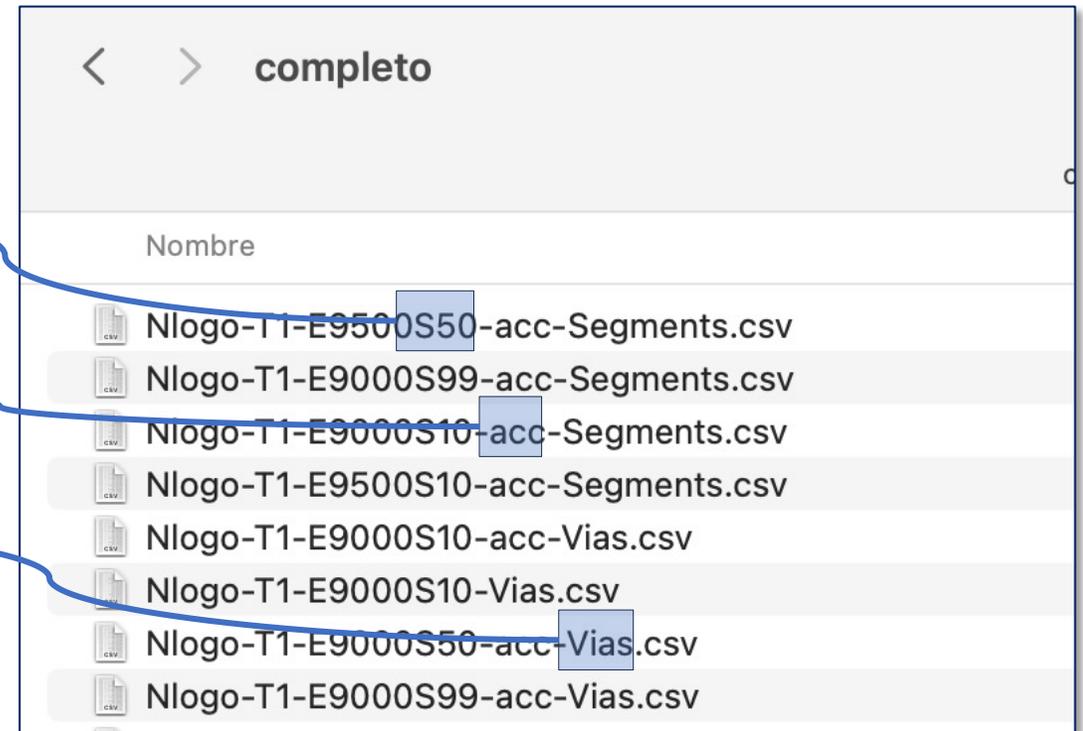
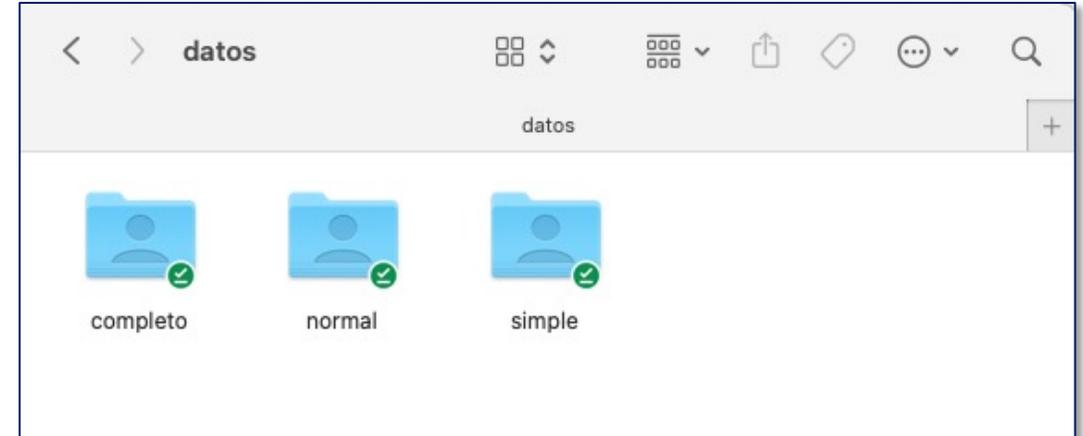
Noviembre 2023

# Ficheros de datos

- Directorio ***datos*** que contiene 3 carpetas:
  - Completo, normal y simple
  - Múltiples ficheros por carpeta

El nombre del fichero sirve para identificar el contenido del mismo.

- % del tipo de vehículos: 10,50 y 99
- Acc. Simulación con accidente
- Indicación de segmentos o vías



# Ficheros de datos

## 1.- Estructuras de datos:

Por cada directorio, tenemos 4 listas:

- sg\_accFiles\_com
- sg\_noaccFiles\_com
- sg\_accData\_com
- sg\_noaccData\_com

- sg\_accFiles\_nor:
- sg\_noaccFiles\_nor

- sg\_accData\_nor
- sg\_noaccData\_nor

- sg\_accFiles\_sim
- sg\_noaccFiles\_sim
- sg\_accData\_sim
- sg\_noaccData\_sim

completo	
Nombre	
E9000S99-acc-Segments.csv	
E9000S50-acc-Segments.csv	
E9000S10-acc-Segments.csv	
E9000S99-Segments.csv	
E9000S50-Segments.csv	
E9000S10-Segments.csv	

Contiene, de forma ordenada, el nombre de los ficheros. Una para los ficheros con accidente o otra para los que no tienen.

Contiene, de forma ordenada, los dataframes resultantes de cargar los ficheros csv.

# Ficheros de datos

## Ejemplo 1

completo	
Nombre	
	E9000S99-acc-Segments.csv
	E9000S50-acc-Segments.csv
	E9000S10-acc-Segments.csv
	E9000S99-Segments.csv
	E9000S50-Segments.csv
	E9000S10-Segments.csv

sg\_accFiles\_com

E9000S10-acc-Segments.csv	E9000S50-acc-Segments.csv	E9000S99-acc-Segments.csv
---------------------------	---------------------------	---------------------------

sg\_accdata\_com

--	--	--

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1	120	6					
SA73-TST-17	20:2	120	8					
SA73-TST-17	20:3	120	7					
SA73-TST-17	20:4	120	9					

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1	120	6					
SA73-TST-17	20:2	120	8					
SA73-TST-17	20:3	120	7					
SA73-TST-17	20:4	120	9					

**sg\_dataFiles\_com[2]**

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1	120	6	0	3.00	0.002260	0.310186	0.000048
SA73-TST-17	20:2	120	8	2	5.50	0.001233	0.188882	0.000011
SA73-TST-17	20:3	120	7	3	5.75	0.000982	0.222482	0.000045
SA73-TST-17	20:4	120	9	2	6.00	0.003586	0.519242	0.000107

# Ficheros de datos

## 2.- Estructuras de datos:

### 2 diccionarios

- 1.- lista con los nombres de los ficheros
- 2.- lista con el contenido de los dataframes

completo	
Nombre	
 E9000S99-acc-Segments.csv	
 E9000S50-acc-Segments.csv	
 E9000S10-acc-Segments.csv	
 E9000S99-Segments.csv	
 E9000S50-Segments.csv	
 E9000S10-Segments.csv	

Dict\_sg\_accFiles:

Dict\_sg\_accData

Contiene el diccionario con la **clave** y su **valor**.  
La clave permite identificar el fichero al que nos referimos  
El valor es el nombre del fichero csv.

Contiene el diccionario con la **clave** y su **valor**.  
La clave permite identificar el fichero al que nos referimos  
El valor es el dataframe.

# Ficheros de datos

## Ejemplo 2

completo	
Nombre	
	E9000S99-acc-Segments.csv
	E9000S50-acc-Segments.csv
	E9000S10-acc-Segments.csv
	E9000S99-Segments.csv
	E9000S50-Segments.csv
	E9000S10-Segments.csv

Dicc\_sg\_Files

Completa_S10_acc	Completa_S50_acc	Completa_S99_acc	Completa_S10_noacc	...
Completa_S10_acc.csv	Completa_S50_acc.csv	Completa_S99_acc.csv		

Dicc\_sg\_data

Completa_S10_acc	Completa_S50_acc	Completa_S99_acc	Completa_S10_noacc	...

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1							
SA73-TST-17	20:2							
SA73-TST-17	20:3							
SA73-TST-17	20:4							

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1	120	6	0	3.00	0.002260	0.310186	0.000048
SA73-TST-17	20:2	120	8	2	5.50	0.001233	0.188882	0.000011
SA73-TST-17	20:3	120	7	3	5.75	0.000982	0.222482	0.000045
SA73-TST-17	20:4	120	9	2	6.00	0.003586	0.519242	0.000107

Dicc\_sg\_data[completa\_s10\_noacc]

Name	Time	CurrentSpeed	car-intensity	truck-intensity	eq-intensity	NoxS	CO2S	PPMS
SA73-TST-17	20:0	120	7	1	4.25	0.005792	0.838231	0.000170
SA73-TST-17	20:1	120	6	0	3.00	0.002260	0.310186	0.000048
SA73-TST-17	20:2	120	8	2	5.50	0.001233	0.188882	0.000011
SA73-TST-17	20:3	120	7	3	5.75	0.000982	0.222482	0.000045
SA73-TST-17	20:4	120	9	2	6.00	0.003586	0.519242	0.000107

Iker Martín, Maribel Castillo y Vicente R. Tomás  
martini@uji.es; castillo@uji.es; vtomas@uji.es