

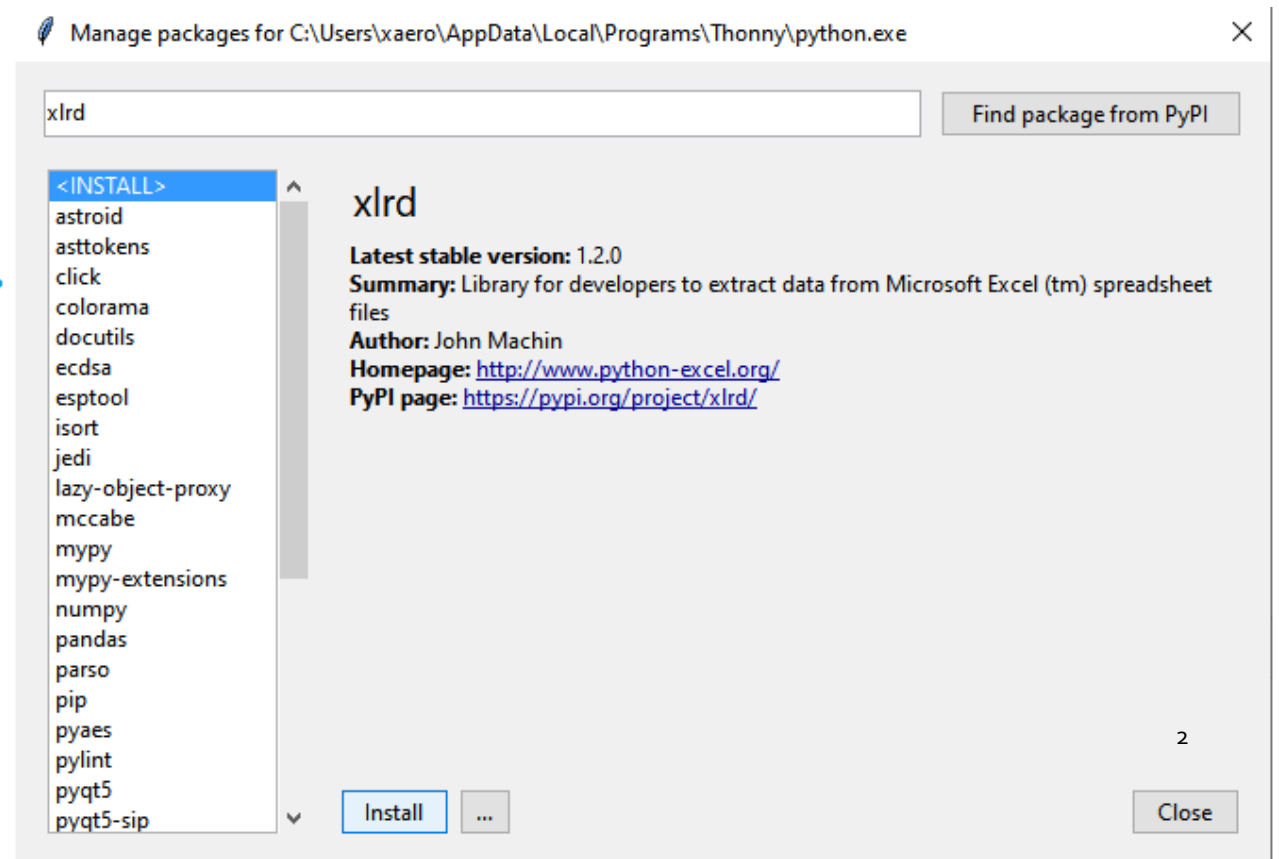
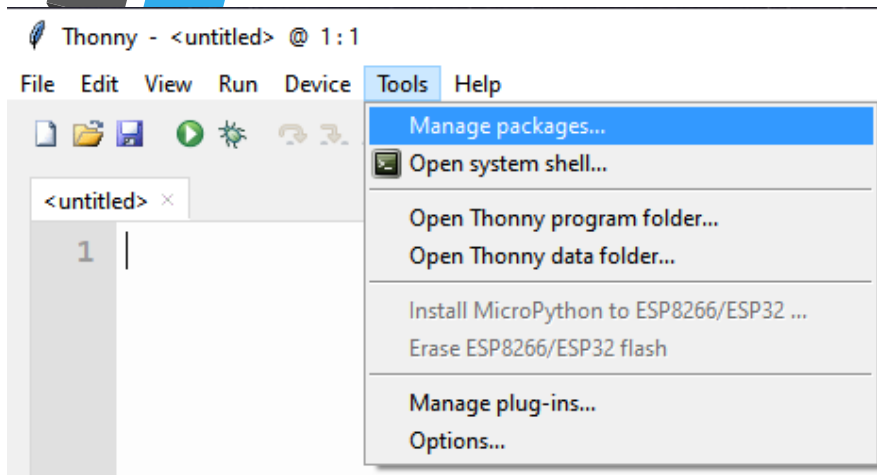


PROCESAMIENTO DE ARCHIVOS DE EXCEL

PREPARATIVOS

Para poder manipular archivos de Excel mediante Python, debes tener instalado ***pandas*** y la librería ***xlrd***.

En Thonny, la instalación se hace con los siguientes pasos:



PREPARATIVOS

También descarga la tabla [vendedores.xlsx](#), la cual muestra información de los vendedores en una empresa, dividida por regiones (norte, sur, este, oeste).

	A	B	C	D	E	F	G	H
1	REGION	ID	NOMBRE	APELLIDO	SALARIO	UNIDADES VENDIDAS	VENTAS TOTALES	PORCENTAJE DE VENTAS
2	NORTE	8725	Arturo	Bustos González	12041	212	4554	0.0693
3	NORTE	4352	Jesús Said	Caro Ibarra	11504	488	3298	0.0502
4	NORTE	5185	Ana Valeria	Cruz Montesinos	10046	429	7238	0.1102
5	NORTE	6918	Miranda	Doring Romo	10266	109	5635	0.0858
6	NORTE	8607	Bruno Guissepe	Garcini García	10431	222	5405	0.0823
7	NORTE	5158	Donovan Hiram	García Cruz	11773	438	3577	0.0545
8	NORTE	5397	Ixchel	García Galán	8654	291	10020	0.1526
9	NORTE	2353	Sebastián	Garduño Pabello	11777	280	7922	0.1206
10	NORTE	5467	Gustavo	Guadarrama García	12071	184	8972	0.1366
11	NORTE	2523	Juan Pablo	Herrera Moro Salcedo	12072	143	9053	0.1378
12	SUR	3969	Héctor	Hidalgo Sepúlveda	8438	431	6626	0.0920
13	SUR	3397	Xareni	Lara Roldán	9041	451	4373	0.0607
14	SUR	9252	Abigail	Luna Pineda	8561	330	4158	0.0577
15	SUR	9091	Diego	Mancha Mendoza	9446	366	5022	0.0697
16	SUR	2445	Eric	Martin Monterrubio	12141	476	6234	0.0865
17	SUR	5305	Carlos Emmiliano	Medina Rojo	8044	340	6984	0.0969
18	SUR	3796	Jimena Sarai	Medina Sedano	11431	231	7522	0.1044
19	SUR	6032	Fermín	Méndez García	9953	448	3612	0.0501
20	SUR	6413	Sergio Arturo	Nava Rodríguez	9572	218	5157	0.0716
21	SUR	9619	Óscar Eduardo	Nieto Espitia	8421	145	5951	0.0826
22	SUR	2323	Isis Renata	Reyes Rosales	12003	341	16408	0.2277
23	ESTE	7537	Orson Nekros Oswaldo	Reyes Téllez	8152	299	6742	0.1195
24	ESTE	5453	Carlos Eduardo	Riosmena Castañón	9837	494	8234	0.1460
25	ESTE	6930	Natalia	Rivera Albarrán	10520	472	6301	0.1117
26	ESTE	3346	Andrés Alejandro	Rodríguez López	8204	180	9387	0.1664
27	ESTE	5832	Angélica Jazmín	Román Rodríguez	8776	420	8831	0.1565
28	ESTE	3696	Erick David	Rosales Aldana	10188	102	3284	0.0582
29	ESTE	7240	Marien	Ruenes Amarante	11402	358	5668	0.1005

Carga de un archivo de Excel

Podemos utilizar la librería de Pandas para poder leer la información contenida en un archivo de Excel.

Para ello, utiliza el método `read_excel`("nombre del archivo")

La información será guardará en
una tabla también conocida como
DataFrame

```
excelPandas.py * ×
1 import pandas as pd
2
3 #Lee la información del archivo de Excel
4 #y se guarda en un Dataframe llamado tabla.
5 tabla = pd.read_excel("vendedores.xlsx")
6 #Imprime la tabla cargada.
7 print(tabla)
```

```
Shell ×
[52 rows x 8 columns]
>>> %Run excelPandas.py
```

	REGION	ID	...	VENTAS TOTALES	PORCENTAJE DE VENTAS
0	NORTE	8725	...	4554	0.069343
1	NORTE	4352	...	3298	0.050218
2	NORTE	5185	...	7238	0.110211
3	NORTE	6918	...	5635	0.085803
4	NORTE	8607	...	5405	0.082300
5	NORTE	5158	...	3577	0.054466
6	NORTE	5397	...	10020	0.152572
7	NORTE	2353	...	7922	0.120626
8	NORTE	5467	...	8972	0.136614
9	NORTE	2523	...	9053	0.137848
10	SUR	3969	...	6626	0.091968
11	SUR	3397	...	4373	0.060696
12	SUR	9252	...	4158	0.057712
13	SUR	9091	...	5022	0.069704

Asegúrate de que
escribiste
correctamente el
nombre del
archivo

Puedes usar *print*
para visualizar la
tabla, aunque
aparecerá
recortada en el
Shell.

Extracción de información por columna

Si requieres guardar la información de una columna entera en una lista, utiliza la siguiente sintaxis:

```
lista = tabla['Columna']
```

```
1 import pandas as pd
2
3 #Lee la información del archivo de Excel
4 #y se guarda en un Dataframe llamado tabla.
5 tabla = pd.read_excel("vendedores.xlsx")
6
7 #Obtener los nombres de los vendedores
8 nombres = tabla['NOMBRE']
9 print(nombres)
10
11
```

Shell x

```
>>> %Run excelPandas.py
```

```
0          Arturo
1       Jesús Said
2     Ana Valeria
3         Miranda
4   Bruno Guissepe
5   Donovan Hiram
6         Ixchel
7     Sebastián
8         Gustavo
9     Juan Pablo
10        Héctor
11        Xareni
12      Abigail
```

La lista guarda los nombres de los vendedores de las cuatro regiones.

Medidas descriptivas

Al guardar la información en listas, puedes obtener diversas medidas descriptivas, tales como el valor más grande, el más pequeño, el promedio, la cantidad de registros, etc.

```
lista.max()  
lista.min()  
lista.mean()  
lista.count()
```



```
1 import pandas as pd
2
3 #Lee la información del archivo de Excel
4 #y se guarda en un Dataframe llamado tabla.
5 tabla = pd.read_excel("vendedores.xlsx")
6
7 #Obtener el salario de todos los vendedores
8 salarios = tabla['SALARIO']
9
10 #Guarda el salario más grande en la variable salarioMax
11 salarioMax = salarios.max();
12 print("El salario más grande es: $", salarioMax)
13
```

Shell ×

```
>>> %Run excelPandas.py
El salario más grande es: $ 12369
```

Una forma más directa es
hacerlo en una sola sentencia

`tabla['SALARIO'].max()`

Extracción de subtablas

Para poder obtener una subtabla de la tabla principal, debes de agruparla por alguna columna, y después especificar el valor.

Para ello, utiliza el método `groupby('Columna').get_group('Valor')`

Este método es utilizado por el Dataframe creado anteriormente (la tabla principal)

```

1 import pandas as pd
2
3 #Lee la información del archivo de Excel
4 #y se guarda en un Dataframe llamado tabla.
5 tabla = pd.read_excel("vendedores.xlsx")
6
7 #Obtén una subtabla que tenga únicamente
8 #la información de la Región Sur.
9 tablaSur = tabla.groupby('REGION').get_group('SUR')
10 print(tablaSur)

```

Shell x

[52 ROWS X 8 COLUMNS]

>>> %Run excelPandas.py

	REGION	ID	...	VENTAS	TOTALES	PORCENTAJE DE VENTAS
10	SUR	3969	...	6626		0.091968
11	SUR	3397	...	4373		0.060696
12	SUR	9252	...	4158		0.057712
13	SUR	9091	...	5022		0.069704
14	SUR	2445	...	6234		0.086527
15	SUR	5305	...	6984		0.096937
16	SUR	3796	...	7522		0.104404
17	SUR	6032	...	3612		0.050134
18	SUR	6413	...	5157		0.071578
19	SUR	9619	...	5951		0.082599
20	SUR	2323	...	16408		0.227740

El nombre de la columna y del valor deben de coincidir con la tabla de Excel

tablaSur es la subtabla con la información de la Región Sur.

¡A programar!

Realiza un programa en Python con las siguientes indicaciones:

1. Carga la información del archivo [vendedores.xlsx](#) en un Dataframe llamado **reporte**.
2. Extrae una subtabla llamada **reporteNorte** con la información de la región **Norte**.
3. Guarda la información de las **Ventas Totales** de la subtabla **reporteNorte** en una lista llamada **ventasNorte**.
4. Muestra el promedio de las **Ventas Totales** de la región **Norte**.

Continúa

¡A programar!

5. Guarda los nombres de los vendedores de la región Norte en una lista llamada **vendedoresNorte**
6. Guarda los salarios de los vendedores de la región Norte en una lista llamada **salariosNorte**
7. Mediante la librería **matplotlib** realiza una gráfica de barras de los nombres de los vendedores y sus salarios de la región Norte. Agrega un título y etiquetas a los ejes.