Introduction to Python (practice)

Programa de Doctorat en Economia, Universitat de Barcelona

David Moriña

Python practice 1

Things from Session 1:

- Web scraping library: *Scrapy*
- help(pandas.DataFrame.unstack)
- Get a list of objects in current session: *%whos*

Exercise 1

- Download all *csv* files in the *practice/data* folder and define the proper working directory
- Import all the files to the JupyterLab session

```
import pandas as pd
import os

os.chdir('/home/dmorina/Insync/dmorina@ub.edu/OneDrive
 ↪  Biz/Docència/UB/2023-2024/PyEcon/1. Introduction
 ↪  to Python/practice/data/')

df1 = pd.read_csv('SiniestrosUAC2020ENE.csv')
df2 = pd.read_csv('SiniestrosUAC2020FEB.csv')
df3 = pd.read_csv('SiniestrosUAC2020MAR.csv')
df4 = pd.read_csv('SiniestrosUAC2020ABR.csv')
df5 = pd.read_csv('SiniestrosUAC2020MAY.csv')
df6 = pd.read_csv('SiniestrosUAC2020JUN.csv')
df7 = pd.read_csv('SiniestrosUAC2020JUL.csv')
```

Exercise 1

- Download all *csv* files in the *practice/data* folder and define the proper working directory
- Import all the files to the JupyterLab session

```
import pandas as pd
import os

os.chdir('/home/dmorina/Insync/dmorina@ub.edu/OneDrive
 ↪  Biz/Docència/UB/2023-2024/PyEcon/1. Introduction
 ↪  to Python/practice/data/')
vec = [] * len(os.listdir())
for i in range(len(os.listdir())):
  df = pd.read_csv(os.listdir()[i])
  vec.append(df)

%whos
```

```
Variable     Type          Data/Info
```

Exercise 2

- Combine all the files in a single DataFrame
- Sort the resulting file by date ('fecha Siniestro Acto')
- Rename the column *fecha Siniestro Acto* to *Date*

```
total = pd.concat([df1, df2, df3, df4, df5, df6, df7,
                   df8, df9, df10, df11, df12])

total = total.sort_values(by='fecha Siniestro Acto')
total = total.rename(columns={'fecha Siniestro Acto':
 ↪  'Date'})
total.head(2)
```

Exercise 2

- Combine all the files in a single DataFrame
- Sort the resulting file by date ('fecha Siniestro Acto')
- Rename the column *fecha Siniestro Acto* to *Date*

```
total = pd.concat(vec)

total = total.sort_values(by=['fecha Siniestro
 ↪  Acto'], ascending=[True])
total = total.rename(columns={'fecha Siniestro Acto':
 ↪  'Date'})
total.head(2)
```

	Identificador CIS	Date	Código Especialidad Rea
352	28554095	2020-01-01 00:00:00	68
286	29604337	2020-01-02 00:00:00	51

Exercise 3

- Group by week number (Hint: Use first *to_datetime* function from *pandas* library to make the column *datetime* and then *isocalendar().week* from *datatable* library) and sum the values of 'Unidades Acto'.

```
import datatable as dt
total['Date'] = pd.to_datetime(total['Date'])
total['WeekNum'] =
 ↪  total['Date'].dt.isocalendar().week
total_grouped = total.groupby(["WeekNum"])["Unidades
 ↪  Acto"].sum()
total_grouped.head(3)
```

```
<IPython.core.display.HTML object>
```

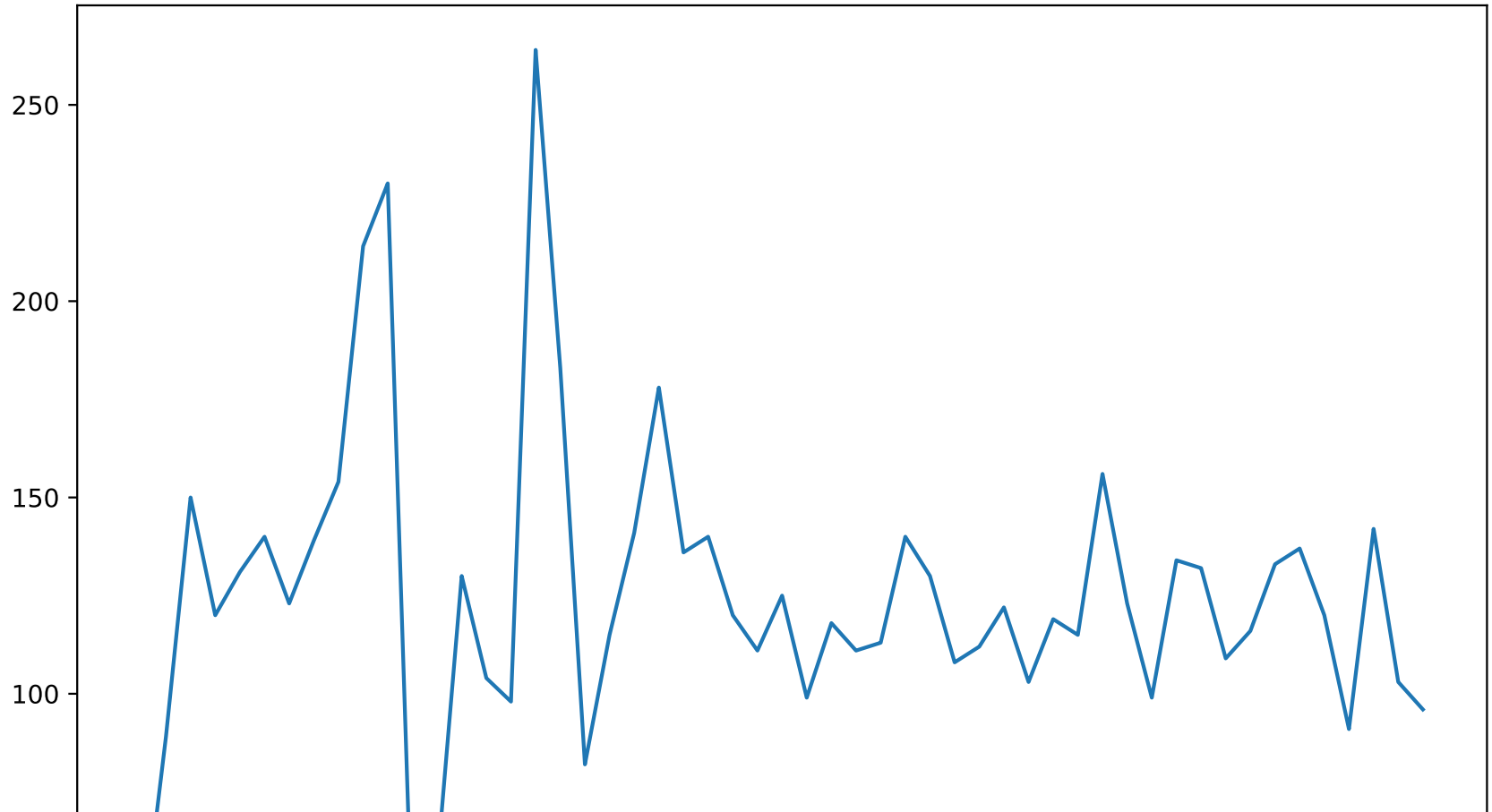
```
WeekNum
1      37
2      89
3     150
```

Exercise 4

- Plot the evolution of the variable ‘Unidades Acto’.

```
total_grouped.plot()
```

```
<Axes: xlabel='WeekNum'>
```



Exercise 5

- Export the final DataFrame to a *csv* file

```
total_grouped.to_csv('Total.csv')
```