Dealing with CSV Files

In this lesson, we discuss how to read from and write to CSV files.

We'll cover the following

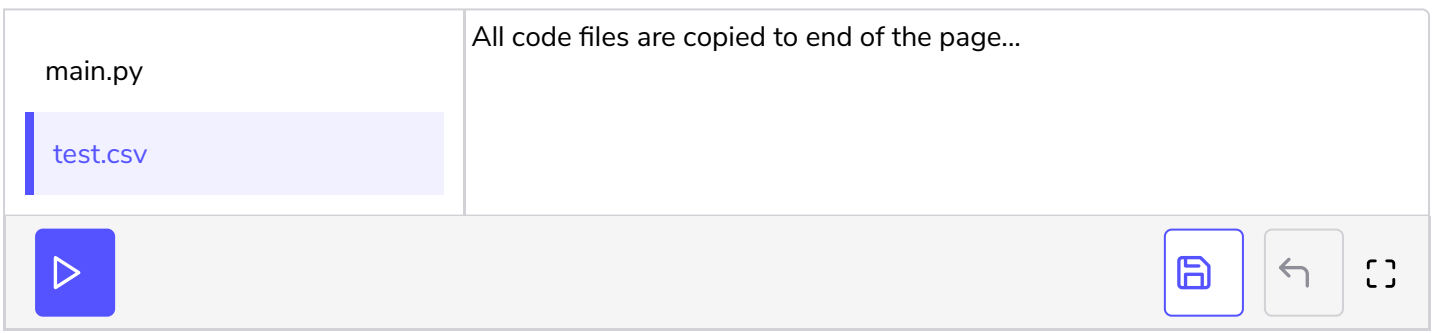
CSV

CSV stands for comma separated values. These are the most common and widely used files with a pandas `DataFrame` object. It is a kind of a text file that uses the extension **.csv** instead of **.txt**. Text is separated by *commas* `,` in this file, and when a `DatFrame` reads this file, values between commas become *cell* values of the `DataFrame`.

The `pandas` package provides both *reading* and *writing* functions for *csv* files.

Reading CSV files

The `read_csv(file_name)` function is used to read from a *csv* file. The `file_name` should be replaced with the name of your file. Make sure that the file is in the same directory as your program file or clearly define the path to that file, e.g., `folder/folder/file.csv`. If no path is defined with the filename, then the program looks for the file in the current directory.



If `header = None` parameter is not passed to the `read_csv` function, the first row in the file is considered to be the column names. If it is passed as a parameter then, it

treats the first row as data and automatically generates the column names as done for *indexes*.

Writing CSV files

The `to_csv(file_name)` function is used from the `DataFrame` object to save the data from the `DataFrame` object to the *csv* file. If the file does not exist, this function first creates the file and then writes data to it.

```
import numpy as np
import pandas as pd

df = pd.DataFrame(np.arange(25).reshape(5,5), columns = ['A','B','C','D','E'])

df.to_csv('output/myfile.csv')
```

Download and open the generated file. You'll be able to see that the `DataFrame` is copied inside the `myfile.csv` file.

> Change your path according to your system. The `output/` path is only to be used on this platform.

In the next lesson, data merging techniques are discussed.

Code Files Content !!!

⤋ ⤋

```
------------------------------------------------------------------------
|  main.py [1]
------------------------------------------------------------------------


import pandas as pd

df = pd.read_csv('test.csv') # reading csv file
print(df)
```

```
df = pd.read_csv('test.csv', header = None) # reading csv file without headers
print('\n',df)



-------------------------------------------------------------------------
|  test.csv [1]
-------------------------------------------------------------------------


C1,C2,C3,C4,C5,C6
A,B,C,D,E,F
G,H,I,J,K,L
M,N,O,P,Q,R
S,T,U,V,W,X
Y,Z,27,28,29,30




*********************************************************************************
```