Python/Web Data Course

Sandboarding in Iquique, Chile in 2018

Bookmark course website / Topics

https://mdjhoel.github.io/datacourse/

Python file input

Python can open **text** files of different formats

1. .txt
2. .csv
3. .json

View ...

1. names.txt
2. data.csv
3. youtube.csv

Python is the best for data science

Easy syntax

```
>>> myfile = open(“names.txt”,”r”)
>>> mydata = myfile.read( )
>>> myfile.close( )
```

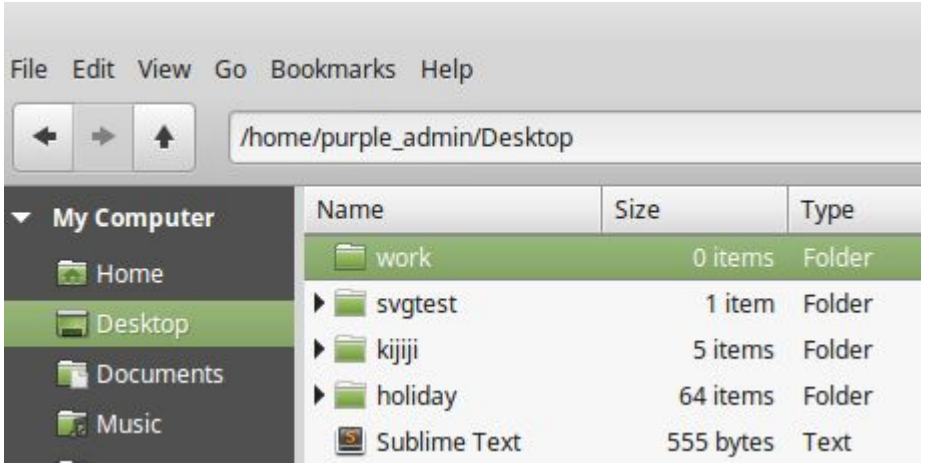
1. Creates a FILE OBJECT called myfile, containing data from names.txt
2. Reads data from file object to a variable as a string
3. Close the file so there will be no memory leaks

Questions

1. What is this course about?
2. What text formats can Python read?
3. Why use Python for reading and writing data?
4. When opening an external file in Python, what data type is created?

Ex. 1 - Prepare your workspace

1. Open a file explorer
2. Navigate to Desktop
3. Create a folder called work



Ex. 1 - Download student names (names.txt)

1. Navigate to https://mdjhoel.github.io/datacourse/
2. Right click and download names.txt to your new “work” folder

Ex. 1 - Read in and print contents of names.txt

1. Start Idle3
2. Type the following commands into the Python Interpreter

```
>>> myfile = open(“names.txt”,”r”)
>>> for name in myfile:
        print(name)
>>> myfile.read( )
>>> myfile.close( )
```

3. What do you think will happen at each stage?

Ex 1. Results and Questions

```
>>> myfile = open("names.txt","r")
>>> for name in myfile:
        print(name)


scott

eva

roy

jonkon

leo

alex

vic

frankie

>>> myfile.read()
''
>>> myfile.close()
>>>
```

1. What does “r” mean?
2. What is happening in the for loop?
3. What appears to be happening in the output?
4. What does myfile.read( ) do?
5. Why does myfile.read( ) return an empty str
6. Why should you ALWAYS close files?

Ex. 1b - Clean data and store it for later

1. Start Idle3
2. Type the following commands into the Python Interpreter

```
>>> myfile = open("names.txt","r")
>>> cleanlist = []
>>> for name in myfile:
        cleanlist.append(name.strip())
>>> myfile.close( )
>>> for name in cleanlist:
        print(name)
```

3. What do you think will happen at each stage?

Ex. 1b - Results and Questions

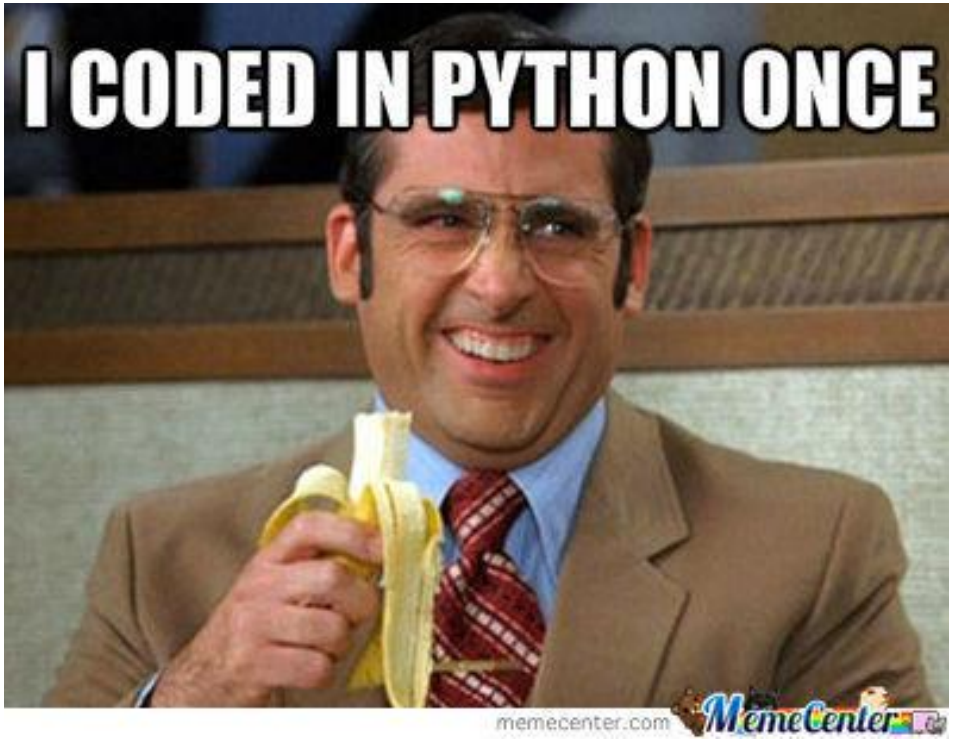
```
>>> myfile = open("names.txt","r")
>>> cleanlist = []
>>> for name in myfile:
        cleanlist.append(name.strip())


>>> cleanlist
['scott', 'eva', 'roy', 'jonkon', 'leo', 'alex', 'vic', 'frankie']
>>> for name in cleanlist:
        print(name)


scott
eva
roy
jonkon
leo
alex
vic
frankie
>>> myfile.close()
>>>
```

1. What is the variable cleanlist initially = [ ] ?
2. What does the list method .append( ) do?
3. What does the str method .strip( ) do?
4. How can individual names be recalled from cleanlist?

Good Job! Now it time for a harder one!



Ex. 2 - Download student data (data.csv)

1. Navigate to https://mdjhoel.github.io/datacourse/
2. Right click and download data.csv to your new “work” folder

Ex. 1 - Read in and print contents of data.csv

1. Start Idle3
2. Type the following commands into the Python Interpreter

```
>>> myfile = open(“data.csv”,”r”)
>>> for line in myfile:
        print(line.strip())
>>> myfile.close( )
```

3. What do you think will happen at each stage?

Ex 2. Results and Questions

```
>>> myfile = open("data.csv","r")
>>> for line in myfile:
        print(line.strip())


scott,2005
eva,2006
roy,2006
jonkon,2007
leo,2007
alex,2007
vic,2006
frankie,2007
>>> myfile.close()
```

1. What does .strip( ) do?
2. Why do you think data.csv has the .csv file extension?

Ex. 2 - Accessing individual parts of a csv line

1. Start Idle3
2. Type the following commands into the Python Interpreter

```
>>> myfile = open("data.csv","r")
>>> namelist = []
>>> yearlist = []
>>> for line in myfile:
        cleanline = line.strip()
        templist = cleanline.split(",")
        namelist.append(templist[0])
        yearlist.append(templist[1])
>>> myfile.close( )
```

3. What do you think will happen at each stage?

Ex. 2 - Results and Questions

```
>>> myfile = open("data.csv","r")
>>> namelist = []
>>> yearlist = []
>>> for line in myfile:
        cleanline = line.strip()
        templist = cleanline.split(",")
        namelist.append(templist[0])
        yearlist.append(templist[1])


>>> namelist
['scott', 'eva', 'roy', 'jonkon', 'leo', 'alex', 'vic', 'frankie']
>>> yearlist
['2005', '2006', '2006', '2007', '2007', '2007', '2006', '2007']
>>> myfile.close()
>>>
```

1. What does .split( ) do?
2. What does parallel lists mean?
3. How can parallel lists be used to find out the birth year of a student?
4. What code do we need to print out “frankie” and “”2007”?

Ex. 2b - Problem

Write a program that will **input** data.csv, **process** the data to find students who were born in years equal or less than 2006 and **output** only these names to the screen.

HINT:

You need to use an ***if statement*** and maybe use int( ) to convert string data to numbers to do the math

Ex. 2b Results and Questions

```
>>> myfile = open("data.csv","r")
>>> for line in myfile:
        cleanline = line.strip()
        temp = cleanline.split(",")
        print(temp)
```

Answer under here

Questions under here

```
['scott', '2005']
scott
['eva', '2006']
eva
['roy', '2006']
roy
['jonkon', '2007']
['leo', '2007']
['alex', '2007']
['vic', '2006']
vic
['frankie', '2007']
```

Good Job! Now it time for writing files!

Writing to files in Python is EASY!

1. To open a file object for reading and write to it ...

```
>>> myfile = open(“mynewfile.csv”,”w”)
>>> myfile.write(“name,type,power”)
>>> myfile.write(“Bulbasor,grass,poison”)
>>> myfile.close( )
```

2. Do not forget to close your file, or NOTHING WILL WRITE
3. Visit https://pokemondb.net/pokedex/national
4. Create a program that writes your 5 favourite Pokemon to a file.

Overwrite or Append to existing files?

```
>>> myfile = open("poke.csv","w")
>>> myfile.write("name,type,power\n")
16
>>> myfile.write("Ivysaur,grass,poison\n")
21
>>> myfile.close()
>>>
>>> myfile = open("poke.csv","a")
>>> myfile.write("Bulbasor,grass,poison\n")
22
>>> myfile.close()
>>>
```

1. What is the “\n” character?
2. What happens when you do NOT use it when writing?
3. What is the difference between the open mode “w” and “a”?

Homework

1. Navigate to https://mdjhoel.github.io/datacourse/
2. Right click and download youtube.csv to your new “work” folder
3. This data is from https://www.kaggle.com/

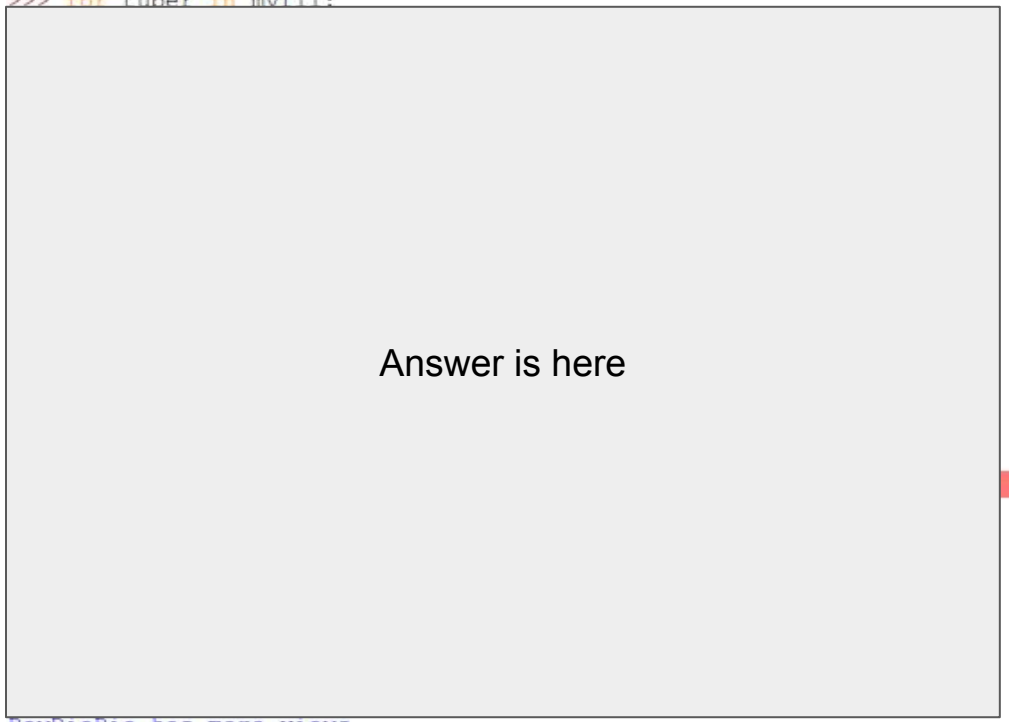
Homework problem

“PewDiePie” and “Logan Paul Vlogs” are rival Youtube channels. **Input** in the youtube.csv and **process** the data to find each Youtuber. Store their data in list variables and then compare who is more popular based on Video Views. **Output** who wins based on Views.

Is there a problem with the ranking system?

Homework answer

```
>>> myfil = open("youtube.csv","r")
>>> for tuber in myfil:
```




```
PewDiePie has more views
>>>
```