Olabode Alamu 1498663 A guide to engineering data science Fun-work 1

1a

In [39]:
```
# Data sources
# Web sources
# CSV - Comma Separated Values
# Excel
# Text files
# json files
# XML files
# binary files
```

1c Install missing packages

In [40]:
```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
```

2a Data types in python

In [41]:
```
# 1 Strings
# 2 Integer
# 3 Lists
# 4 Dictionary
# 5 Tuples
```

2b Basic operations Strings

In [42]:
```
z = 'eggs'
```

In [43]:
```
v = 'chicken'
```

In [44]:
```
z + ' ' + v # String concatenation
```

Out[44]:
```
'eggs chicken'
```

In [45]:
```
z * 3
```

Out[45]:
```
'eggseggseggs'
```

Integer

In [46]:
```
c = 2  # Integer type
d = 4
```

In [47]:
```
c + c  # addition
```

Out[47]: 4

In [48]:
```
c * d  # multiplication
```

Out[48]: 8

In [49]:
```
c / d # division
```

Out[49]: 0.5

In [50]:
```
d - c # Substraction
```

Out[50]: 2

List

In [51]:
```
b = ['new','clothes','are','good']  # Lists
f = ['and','watches','too']
Continent = ['Europe','America','Asia']
```

In [52]:
```
b
```

Out[52]: ['new', 'clothes', 'are', 'good']

In [53]:
```
b * 2
```

Out[53]: ['new', 'clothes', 'are', 'good', 'new', 'clothes', 'are', 'good']

In [54]:
```
b + f  # List concatenation
```

Out[54]: ['new', 'clothes', 'are', 'good', 'and', 'watches', 'too']

Dictionary

In [55]:
```
dict = {'Country':['Germany','USA','Japan'], 'Capital':['Berlin', 'Washington
 DC', 'Tokyo']}
```

In [56]:
```
dict
```

Out[56]:
```
{'Capital': ['Berlin', 'Washington DC', 'Tokyo'],
 'Country': ['Germany', 'USA', 'Japan']}
```

In [57]:
```
dict['Capital']
```

Out[57]: ['Berlin', 'Washington DC', 'Tokyo']

2c

In [58]:
```
df = DataFrame(data = dict, index = Continent ) # Pandas DataFrame object
```

In [59]:
```
df
```

Out[59]:

	Capital	Country
Europe	Berlin	Germany
America	Washington DC	USA
Asia	Tokyo	Japan

In [60]:
```
df.to_csv('new.csv')
```

Import data from different files can be done with the pandas library as shown below

In [61]:
```
df1 = pd.read_csv('new.csv')
```

In [62]:
```
df1
```

Out[62]:

	Unnamed: 0	Capital	Country
0	Europe	Berlin	Germany
1	America	Washington DC	USA
2	Asia	Tokyo	Japan

In [63]:
```
# Import from excel files
df2 = pd.read_excel('p.xlsx')
```

In [64]:
```
df2
```

Out[64]:

						WEEK 1	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unname 4
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
				Name	M	T	W	TH	F	TOTAL
				Jacob	114	51	0	0	15	180
				Giselle	113	53	0	0	15	181
				McKenzie	130	57	0	0	5	192
				Ava	126	53	0	0	15	194
				Mekhi	130	44	0	0	5	179
				Brooklyn	131	40	0	0	15	186
				Ethan	126	53	0	0	15	194
				Avery	116	34	0	0	15	165
				Micheal	131	74	0	0	15	220
				Janae	100	34	0	0	5	139
				Akeem	100	38	0	0	5	143
				Laila	100	19	0	0	15	134

In [ ]:
```

```