



# **Module 2 exercises: Introduction to Python for Data Science**

## **Exercise 1 – Virtual Environments**

1. Connect to the environment called pds using Anaconda Navigator or Anaconda Prompt.
2. The virtual environment will need the following packages to be installed:
  - a. Numpy
  - b. Pandas
  - c. Seaborn
  - d. Matplotlib
  - e. scikit-learn

## **Exercise 2 – Jupyter Notebooks**

1. Create a new notebook.

Experiment with the keyboard shortcuts in Command mode. You might want to try:

- Adding cells above and below.
- Deleting cells
- Changing from markdown to code.
- Verifying, i.e., entering some text or code and running checks.

## **Exercise 3 – Python Review**

### **3.1 – Display a message using variables**

1. Create two variables to hold a person's name and age.



2. Display the person's name and age with a space in between.
3. Alter the code so that . is between the person's name and age.
4. Alter the code so that the person's name is displayed in uppercase.

### 3.2 – Casting variables

1. Capture user input to get the length of the first side of a rectangle.

Use a suitable variable name such as length.

You must cast (convert) the text you input to an integer type (int).

2. Input the length of the second side of the rectangle.

Use a suitable variable name such as width.

Again, cast the input text to an integer type (int).

3. Calculate and display the perimeter of the rectangle.

### 3.3 – Data structures

Using the below data structures, complete the following tasks:

```
ages = [12,18,33,84,45,67,12,82,95,16,10,23,43,29,40,34,30,16,44,69,70,74,38,65,36,83,50,11,79,64,78,37,3,8,68,22,4,60,33,82,45,23,5,18,28,9,17,81,14,88,50,19,59,7,44,93,35,72,25,63,11,69,11,76,10,60,30,14,21,82,47,6,21,88,46,78,92,48,36,28,51]
byron = 'Fare thee well and if forever still forever fare thee well'
desserts = 'sticky toffee pudding', 'affogato', 'syrup sponge'
books = {
    "title": "Hamlet",
    "author": "William Shakespeare",
    "language": "English"
}
```

- Display the length of the ages list.
- Display the ages from the third position to the last.
- Find the position of the first occurrence of the word 'forever' in Byron.
- Print the second of the desserts.



- Check if pancakes are included in desserts.
- Print the title value from books.

### 3.4 – Functions

1. Write a function that takes a string as an argument and returns the number of capital letters in the string. Hint: 'foo'.upper() returns 'FOO'.
2. Write a function that takes two sequences seq\_a and seq\_b as arguments and returns True if every element in seq\_a is also an element of seq\_b, else False.

By “sequence” we mean a list, a tuple or a string.  
Do the exercise without using sets and set methods.

### 3.5 – (Extension) Iteration

Write a piece of code which calculates change to be given from an input amount in £. To start you off, we've given you a list of denominations of UK currency.

```
denoms = [50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, 0.05, 0.02, 0.01]
```

### Exercise 4 – Numpy

```
# import necessary libraries
import numpy as np

# create data to work with
x = np.arange(10)
x
```

1. Display the first 5 elements.
2. Display every other element.
3. Display the elements from index 5 in reverse order.

```
# two dimensional array
y = np.array([[12, 5, 2, 4],
              [ 7, 6, 8, 8],
              [ 1, 6, 7, 7]])
```

4. Display the first 2 rows and the first 3 columns of y.
5. Read the contents of file cdc.csv, containing participant heights, weights, ages (and much more!), into array data. To do this, you can use



the below code:

```
data = np.genfromtxt('data/cdc.csv', delimiter=',',  
skip_header=1)
```

6. Separate the heights (column 5) and the weights (column 6)
7. Calculate the median for the heights and the weights and assign the values to variables.

## Exercise 5 – Pandas

1. Read the file mortgage\_applicants.csv, which sits in the data folder, into a variable called mortgage. **Hint: Your path will need to reflect the location of the file.**
2. Select the Score column of the mortgage DataFrame
3. Select the Balance and Income columns.
4. Select the first row in the mortgage DataFrame.
5. Display all mortgage applicants who have a Balance greater than £1000.
6. Display all mortgage applicants who have a Balance greater than £1000 and a Debt below £50.

## Exercise 6 – Visualisation

1. Load in the dataset train\_viz.csv
2. Create a scatter plot of duration vs. price. Segment by destination.