Coursework: Programming for Data Science (20COP504)

Dr Bangli Liu(b.liu2@lboro.ac.uk)

Coursework 1 (40%)

Task 1

There are some promotion activities happening in a store. You can get a 10% discount when *the total_price* is between £20 and £40, inclusive. If the *total_price* is greater than £40, you can get a 15% discount.

Write a function to take a list of *quantity* and a list of *unit price* as inputs and return the *discounted_price* (e.g. for a 10% discount, *discounted_price = total_price*(1-0.1)*).

Please use the following lists to test your function: *quantity = [2, 4, 6], unit_price = [5, 6.99, 3.59].*

[8 marks]

Task 2

Go to **LEARN--20COP504--Coursework section** and download **drinks.txt**. **drinks.txt** records the sale status of different drinks (cappuccino, tea, espresso) over a specific date.

2.1

Write a program to read the **drinks.txt** file and calculate the total sale of each kind of drink.

Please note that due to typo errors, the word cappuccino is sometimes recorded as cappuccin. The word cappuccin should be counted as cappuccino.

[15 marks]

2.2

Draw a bar chart to illustrate the popularity of the drinks.

[10 marks]

Task 3

3.1

Generate an array **x** with 100 random integers from 0 to 50.

[5 marks]

3.2

Write a function named **MaxFun** (use **for** loop inside the function) to calculate the maximum value of array **x**.

[10 marks]

3.3

Write a function named **MinFun** (use **for** loop inside the function) to calculate the minimum value of array **x**.

[10 marks]

3.4

Normalize array **x** by using **x_normed = (x - minimum) / (maximum-minimum)**

[5 marks]

Task 4

4.1

Create a DataFrame named **dfRetail** (you can create it from Dictionary, Lists or Numpy arrays). The DataFrame should contain five columns, including the information about **product_name**, **sold_quantity**, **purchase_price**, **selling_price**, and **import_country**. Generate data yourself for each column. The DataFrame must have 10 rows (representing for 10 different products).

[12 marks]

4.2

Add a column named **profit** to **dfRetail**. **profit** is calculated as follows:

*profit = (selling_price - purchase_price) * sold_quantity.*

You need to accomplish this by Python code.

[10 marks]

4.3

Identify the top five profitable products and present the information via a pie chart.

[10 marks]

Format

- Please ensure that detailed comments are embedded inside the code.
- Please ensure that your charts are labelled with a suitable colour scheme, title, and legend.

[5 marks]

Files to submit

- You will submit your coursework in the form of a single Jupyter notebook (i.e. the **.ipynb** file).
  - Please ensure that all your results are presented in the Jupyter notebook.
  - Please ensure that your codes are executable.
  - Wherever necessary, add meaningful comments to your code.
  - Use of additional Python packages is allowed.