******

### QQI

### Higher Diploma (HDip) in Data Analytics

### Part-Time March 2018 Intake - Sample

### SAMPLE EXAMINATIONS

*Module Code:* **B8IT105**

*Module Description:* **Programming for Big Data**

*Examiner:*  **Darren Redmond**

*Internal Moderator:*

*External Examiner:* **XXXXXXX**

*Date:*

*Time:*

**INSTRUCTIONS TO CANDIDATES**

**Time allowed is 2 hours.**

**Answer all questions in the asked for scripting language.**

**The marks available for each part are shown clearly: [X marks]**

**Comment your answer script appropriately. Put your student number as a comment at the top of each script.**

**At the end of the exam, submit your script to the Final\_Exam object on Moodle.**

1. The dataset **Boston** is available from the MASS package in R.
   1. Install the MASS package, load the library and access the **Boston** dataset contained in it. Load the help file and read about the dataset. How many observations are there in the dataset?

[5 marks]

* 1. Look at the structure of the dataset and describe it briefly. What unit is the *nitrogen oxide* variable measured in?

[5 marks]

* 1. Find the mean and standard deviation for the room variable in the dataset.

[5 marks]

* 1. For the Charles river variable which group has the highest mean value medv and what is the mean value for this treatment?

[5 marks]

* 1. Create a linear regression for Boston with medv as the response variable against the other variables. Which variables are statistically significant variable to predict the medv?

[5 marks]

**(Total: 25 Marks)**

1. The dataset **mtcars** is available by default in R.
   1. Access the **mtcars** dataset. Load the help file and read about it. In what year was the data collected?

[3 marks]

* 1. Select only those row which have complete records (i.e., those rows which have no NAs). Call this dataframe **mtcars2**.

[4 marks]

* 1. Use **mtcars2** to produce a boxplot of wt against cyl.

You should:

* Include different colours for the three boxplots
* Include a main title, and relevant x- and y-axis labels
* Label the boxplots with the names of the cyl
* Rotate the numbers on the left axis so that they appear horizontal

Comment on the resulting graph.

[18 marks]

**(Total: 25 Marks)**

1. This question tests your ability to manipulate strings in Python.

A file based document analyser application is one that reads a file and

analyses the text in the file while providing summary analysis.

1. Create a class called DocumentAnalyser.

[4 marks]

1. In your DocumentAnalyser class, create methods to calculate the:
   1. number of lines in a file
   2. number of words in a file
   3. number of characters in a file
   4. average number of characters per line in a file
   5. average number of words per line in a file
   6. average number of characters per word in a file

[12 marks]

1. Write some code to test your new DocumentAnalyser class and print out results of your testing to the user. Give some consideration to what sort of strings you might want to use for your testing.

[9 marks]

**(Total: 25 Marks)**

1. This question tests your understanding of Big O notation and programming in Python use lambda, map, and reduce functionality.
2. Create a class in Python called Calculator.

In your Calculator class, implement methods called **triple(),** **third()**, **fourth\_order()**, and **fourth\_order\_root()** which take a list of numbers as an argument and return the appropriate result.

These should be implemented with the map and lambda functions.

For example, if the argument is [1, 2, 3] for the triple() function it will return [3, 6, 9].

[6 marks]

1. Evaluate the Big‐O classification for the following functions.
   1. f(n) = 4\*n ‐ 1
   2. f(n) = 6\*log n – 2
   3. f(n) = 2\*n^4 + 9\*n^3 + 5
   4. f(n) = 6\*n^2 + (1+9\*n)\*3n^2
   5. f(n) = 5\*n + 8\*n
   6. f(n) = 21

[6 marks]

1. Write a program to make a “Guess the Letter” game.

The computer will think of a random letter from ‘a’ to ‘z’, and ask you to guess it. The computer will tell you if each guess is too high or too low in the alphabet. You win if you can guess the number within six tries.

[13 marks]

**(Total: 25 Marks)**

**END OF EXAMINATION**