**6CS030 Big Data**

**2018/9**

**Portfolio – Part 1**

# Worksheet One – 10%

# Hand-out: Week 3. Due: Week 4 (See Canvas for date)

This worksheet is based on Workbooks 1 and 2.

1. Download the *Employment Rate & Qualifications Profile of Adults* Excel spreadsheet from Canvas

This spreadsheet consists of 3 worksheets: E*mployment Rate, Degree-Level Qualification* and *No Qualifications.*

You need to examine just one of the worksheets.

First take your student number and divide it by 3. Use the ***remainder*** value to pick one of the following worksheets:

|  |  |
| --- | --- |
| **Remainder Value** | **Worksheet to use** |
| 0 | Employment Rate |
| 1 | Degree-Level Qualification |
| 2 | No Qualifications |

For example, if your student number is *1712345*, *1712345/3= 2* so you would use the *No Qualifications* worksheet.

1. Examine your worksheet and carry out the following tasks:

|  |  |  |
| --- | --- | --- |
| **Task no** | **Task** | **Marks** |
| a | Identify any cleaning issues in the worksheet and take steps to rectify the issues.  Document what changes have been made. | 40 |
| b | Convert the worksheet so that it is in a suitable format for a relational table and import into Oracle.  Choose appropriate names for the table and column names.  Document what changes you have made. | 20 |
| c | Write two SQL commands that demonstrate an OLAP query. | 20 |
| d | Name one advantage to using this approach for handling Big Data and include brief explanation of why you think this is an advantage. | 10 |
| e | Name one disadvantage to using this approach for handling Big Data and include brief explanation of why you think this is a disadvantage. | 10 |

See Marksheet for further marks breakdown.

Note this is an individual assessment. Any group answers will be classed as plagiarism.

Upload a Word Document on Canvas which includes evidence of the above tasks.

You must use the worksheet allocated otherwise 0 marks will be allocated.