# 6CS030 Worksheet Two – 10%

# Hand-out: Week 7. Due: Week 8

|  |  |
| --- | --- |
| **Student Number:** |  |
| **Student Name:** |  |

This worksheet is based on the MongoDB Wiki and Workbook.

1. On Canvas there are three JSON exports from Twitter.

You need to analyse just one of the JSON datasets.

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

|  |  |  |
| --- | --- | --- |
| **Remainder Value** | **JSON Dataset to use** | **Dataset Generated From** |
| 0 | obama.json | Barrack Obama |
| 1 | bhx.json | Birmingham Airport |
| 2 | kermode.json | Mark Kermode (Film Review) |

For example, if your student number is *1712345*, *1712345/3= 2* so you would use the *kermode.json* dataset.

|  |  |
| --- | --- |
| **My Remainder Value:** |  |

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

|  |  |  |
| --- | --- | --- |
| **Task no** | **Task** | **Marks** |
| a | Import the data into your own MongoDB database:   * Show the command to do this (you can just show your password as *myPassword*.) * Write a command to show how many documents are in your collection   Do ensure you include the results from the query too. | 20 |

|  |  |  |
| --- | --- | --- |
| b | Analyse the data  Write a command to:   * Show one document * Show the unique values in one field * Shows a set of documents based on some criteria. Output just two fields from the document. * Use a regular expression to search for some criteria. The search should be case insensitive   Do ensure you include the results from each query too. | 40 |

|  |  |  |
| --- | --- | --- |
| c | Reshape the collection  Write a command to:   * Update a field within the collection * Create a new collection based on a subset of the dataset. Include a query to show a document from the new collection   Do ensure you include the results from these commands too. | 20 |

|  |  |  |
| --- | --- | --- |
| d | Name one advantage to using this approach for handing Big Data and include brief explanation of why you think this is an advantage.  Include:   * Advantage * Short explanation | 10 |

|  |  |  |
| --- | --- | --- |
| e | Name one disadvantage to using this approach for handing Big Data and include brief explanation of why you think this is a disadvantage.  Include:   * Disadvantage * Short explanation | 10 |

Remember:

* This is an individual assessment
* You must use the correct dataset – 0 marks will be given if you pick the wrong one!
* Upload this document on Canvas when complete.
  + You can create a Python Notebook instead, ensure you include all the above information.