COMP6235: CW 1 Specification

Module:	Foundations of Data Science			Lecturers:	ES, MB, CP, HF, RT
Assignment:	Data processing			Weight:	15%
Deadline:	18/11/2016	Feedback:	16/12/2016	Version	1.0

Description

This coursework will assess your understanding of the early stages of the Data Science pipeline and will cover collection, storage, and retrieval of data. You will be working with data from the Food Standards Agency http://ratings.food.gov.uk/open-data/en-GB

You will be required to run code to answer the given questions in the Jupyter notebook, and write a report describing an approach to a given problem.

Instructions

Task 1 (75% of the marks)

Download the Jupyter notebook provided, and upload it to the Jupyter Hub instance you have running on your VM or other computer or your choice. Refer to the instructions in the tutorial to setup your machine. Note, that Jupyterhub is not supported on Windows.

Write code to answer the questions in the notebook. The following marks are available for each question:

Question 1: 15 marks split equally between (a), (b), and (c)

Question 2: 15 marks split as follows

- A. 1
- B. 1
- C. 3
- D. 5
- E. 5

Task 2 (25% of the marks)

Write a document of up to 2 pages describing the manner in which you would store the data you downloaded in question 1, for use in a Data Science application, and make mention of the following:

1. The difference between a relational database and a NoSQL database 20%
2. The costs and benefits of using different forms of storage, both in general, and with the particular data here. 20%
3. Conclude which method you would use for the data described in questions 1 and 2, and justify your choice 30%.
4. Presentation, grammar, and good academic practice 20%

You may make any reasonable assumptions, provided that they are documented.

Academic Integrity

You must include the following statement in your report, whose spirit you must adhere to:

I am aware of the requirements of good academic practice, and the potential penalties for any breaches.

Details of the possible penalties for breaches of academic integrity and how to avoid them can be found in the Academic Integrity Regulations in Section IV of the University Calendar: http://www.calendar.soton.ac.uk/sectionIV/ academic-integrity-regs.html

This is not included as a part of your page limit.

Download

Jupyter notebook http://www.edshare.soton.ac.uk/id/document/292461

Handin

You will hand in two files to the ECS submission system:

- The completed Jupyter notebook in .ipynb format. Please use this exact notebook, as it contains additional metadata which will help us with the marking. To download it, open the notebook in Jupyter and click File -> Download As -> Notebook (.ipynb)
- A document in either Word (doc or docx), Libre office (odt), HTML, or PDF format which answers question 3