COMP5122M Data Science Coursework (2019-20)

Coursework 1 (marked out of 50 and worth 25% of your final module grade)

For this part of the coursework you need to produce the data tables that are specified below and write a report.

Data tables [30 marks]
Submit an Excel workbook that is named "cw1.xlsx" and contains each of the data tables that are specified below in a separate worksheet. The name of each worksheet should be the same as that specified below (e.g., "Referendum"). Each table should contain the columns that are specified below Marks will be awarded according to the workbook content and adherence to the specification.

Referendum table
This data table should show the referendum result for each constituency.

Column name	Description
ONS code	The ONS parliamentary codes
Voted leave (%)	The percentage of people who voted "leave"

Petitions table
This data table should show the number of signatures from each constituency for six petitions that each gained at least 10,000 signatures. You may choose any petitions, provided that three are pro-leave and the other three are pro-remain.

Column name	Description
ONS code	The ONS parliamentary codes
1st petition title	The number of signatures
2nd petition title	The number of signatures
3rd petition title	The number of signatures
4th petition title	The number of signatures
5th petition title	The number of signatures
6th petition title	The number of signatures

Age table
This data table should show the number of people of each age (in years) in each constituency.

Column name	Description
ONS code	The ONS parliamentary codes
Age 0	The number of people aged 0 years
Etc.	
Age 89	The number of people aged 0 years
Age 90+	The number of people aged 0 years

Jobs table
This data table should show the employment rate and median monthly earnings for each constituency.

Column name	Description
ONS code	The ONS parliamentary codes
Employment rate	The percentage of the economically active population who are in employment
Median monthly earnings	The median earnings (£), according to where people are resident

Report [20 marks]

Your report should be a maximum of four sides of A4 paper (plus a coursework header sheet), and should have a title and separate sections for each data table, which are subdivided into the following sections:

- Method (The method should be described in sufficient detail for it to be replicated by another student. State the tools/software that you used. Your method is expected to be efficient)
- Data quality (Describe any data quality issues that affect the table)

Note: Pay close attention to detail, including the clarity of your writing. Use an 11pt font (or larger). Check your spelling & grammar (e.g., using Word).

Submission procedure & deadline

Each group should submit <u>one copy</u> of the coursework, and submit it both on paper and electronically.

- Via Minerva, hand in your Excel workbook (.xlsx format).
- Hand in a paper copy of your report at the School of Computing SSO (Level 9, Long Room) <u>and</u> as a PDF via Minerva.
- Both the paper and electronic copy must include a coursework header sheet that states the names and student IDs of all the group members.
- **Optionally:** Submit signed copy of the COMP5122M Data Science *Group Work Form* for your group with the <u>paper version</u> of your coursework.

For both paper and electronic submission the deadline is 10am Tue 19th Nov 2019. Late submissions will be penalised in accordance with departmental guidelines.

Roy Ruddle, October 2019