TMA 02

Question 2 (35 marks)

You should be able to answer this question when you have completed Parts 8 to 17 of the module.

This question is designed to get you started on a data investigation that will be developed into a larger investigation for the end-of-module assessment (EMA).

This question tests the following learning outcomes:

- Use data to answer a practical question.
- Use appropriate software packages to explore a dataset.
- Write a report detailing a systematic approach to analysing a dataset.

Ensure that you have made a copy of the TMA02_Question2b.ipynb Notebook and renamed it so that it has your personal identifier (PI) at the front of the Notebook filename (i.e. *yourPI*_TMA02_Question2b.ipynb).

You should spend no more than three and a half hours on this question.

Write your answers to Questions 2(a) and 2(c) directly into your *Solution Document*. Question 2(b) requires you to work in the *yourPI*_TMA02_Question2b.ipynb Notebook. Write the filename of your Question 2(b) Notebook in your *Solution Document* under the heading 'Question 2(b)'.

Scenario for Question 2

You are to investigate patterns in votes cast in the EU (Brexit) referendum.

For this question you are provided with an Electoral Commission dataset:

http://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/eu-referendum/electorate-and-count-information

The dataset you will be using is the full results data. It is contained in the file EU-referendum-result-data.csv in the TMA02_2016J/data/ folder.

Note that you will be investigating the referendum results in this TMA and the EMA. When you discuss this work with your tutor and other students, please limit your discussions to the data and not observations about the correctness or otherwise of anyone's vote in the referendum or opinion on EU membership.

a. You should spend no more than half an hour on this part. This part is designed to give you a feel for the data you are investigating.

- Read the introduction to 'The 2016 EU Referendum Report' 2016-EU-referendum-report.pdf (in the TMA02_2016J/ folder) downloaded from http://www.electoralcommission.org.uk/__data/assets/pdf_file/0008/215279/2016-EU-referendum-report.pdf (accessed 13 October 2016). This gives some background to the referendum and results.

- Visit the BBC's summary of the referendum results at http://www.bbc.co.uk/news/politics/eu_referendum/results. Browse the website. Use the 'interactive map' and list of local results to view selected parts of the data.

- In your *Solution Document*, use no more than 100 words to write bullet points to briefly observe what you have been able to find out from the PDF and website about how the referendum results were reported.

The purpose of writing this short summary is to demonstrate that you have explored the background and got some feel for the data you will be working with. We do not want you to write about the pattern of voting: you will be carrying out your own data analysis in the next part of the question.

(5 marks)

b. You should spend no more than two hours on this part. In this part you will work in the
*yourPI*_TMA02_Question2b.ipynb Notebook to investigate the EU referendum data.
Treat this Notebook as a lab notebook: keep all the work you do and don't tidy it up or delete work that turns out to be
a dead end. Use level 1 headings in Markdown cells in the Notebook to help your tutor identify regions in the Notebook
that demonstrate you have performed the required steps.

- Open your copy of the Notebook. Import the dataset EU-referendum-result-data.csv from the
  TMA02_2016J/data/ folder into a DataFrame. Use some simple *pandas* commands such as head(),
  describe() and unique() to explore the content of the dataset and decide which columns are of interest
  to you. Add explanatory comments to your code.
- The data is organised by region and district. Investigate patterns in the votes between regions.

- Use the matplotlib facility provided in the Notebook to create and label at least three plots to visualise some
  aspects of the data. You may choose to create plots for different regions and/or for different turnout rates.

  Make sure you label any plots you decide to use in your report for part (c).

- Include notes in your lab Notebook critically evaluating what you think your visualisations tell you.

(20 marks)

c. You should spend no more than one hour on this part. In this part you will use your findings from parts (a) and (b) to
write a report using the following outline structure:

    Aims and objectives
    Background
    Sources of data
    Analysis pipeline
    Findings
    Conclusions
    References

Present your report in your *Solution document*. Your report should be no more than 600 words. Some sections may be
very short. Include evidence in the form of screenshots and plots, as appropriate. You should include at least two
visualisations. For every visualisation you include, critically evaluate what it tells you about the data.

Document any uncertainties that you have about the data. If you have no uncertainties, then state this. Comment on
whether there are risks to the report linked to the uncertainties you have documented. If there are risks, state what
those risks are.

You should use references in your report, as appropriate, to support your conclusions and give a context for your
investigation. Include a reference to the Notebook you used in your investigation so that your results may be
independently verified.

(10 marks)