Marking Scheme

The marking scheme for this submission has two parts:

1. The *Executive Briefing* (60% of total mark for this submission) will be assessed as an essay incorporating analytical elements, with consideration given to the language, presentation, and content of the essay as befits a data-led briefing for a busy politician (the Mayor of London) and their policy-makers.

   - *Research Question and Framing* (20% of total mark): the briefing develops a set of research questions and/or research problem within a clearly-defined frame that is relevant to the specified audience.
   - *Results and Interpretation* (30% of total mark): the briefing's results and interpretation are relevant to the specified audience.
   - *Structure and Presentation* (10% 0f total mark): the briefing is written and presented in a manner that is appropriate to the specified audience.

2. The *Reproducible Analysis* (40% of total mark for this submission) will be assessed on the following criteria:

   - *Reproducibility* (20% of total mark): the code is portable and robust.
   - *Quality* (20% of total mark): the code and its outputs are efficient and legible.

The evaluation criteria span the skillset expected of a practicing analyst or data scientist: it is *not* just about writing code, but about selecting, developing, and communicating evidence with your 'end user' in mind, as well as demonstrating an understanding of how to give others within your organisation or team confidence in the results that you present.

Executive Briefing

Research Question and Framing (20% of total mark)

This aspect is focussed on your ability to select and frame a research question or problem that is relevant to a specified audience.

Mark	Guidance
Distinction (70+%/A)	The research problem has been well-developed and -framed for the specified audience. There is a broad range of background research entailing the use of diverse sources to produce a convincing framing of the topic. The briefing is clearly of compelling interest to the audience.
Merit (60-70%/B)	The research problem has been developed for the specified audience. There is a range of background research used to produce a relevant framing of the topic. The briefing is clearly of interest to the audience, but makes some assumptions that suggest more limited reasoning about the audience.
Pass (50-60%/C)	The research problem is of limited interest to the specified audience. There is a limited range of background research used to support the framing of the topic. Parts of the briefing are demonstrably of interest to the audience but this is only weakly-communicated or largely implicit.
Fail (<50%)	The research problem is not obviously of interest to the specified audience. There is a very limited range of background research used to support the framing of the topic. Few, if any, efforts are mode to make the briefing relevant to the audience, and the overall framing is weak and/or illogical.

Results and Interpretation (30% of total mark)

This aspect is focussed on your ability to choose and employ the analytical tools covered in this and other modules (you are free to draw on knowledge developed in *Quantitative Methods* and *GIS*) to support your investigation of the research question.

Mark	Guidance
Distinction (70+%/A)	The analysis is entirely appropriate to the problem. It demonstrates a comprehensive technical understanding of how to select and deploy analytical approaches in support of an analysis. Interpretations of results are very well constructed, clear and focused in relation to the research problem/framing. Maps, figures, and tables are excellent additions to the text and all work together to communicate effectively with the specified audience.
Merit (60-70%/B)	The analysis is appropriate to the problem. It demonstrates a sound technical understanding of how to select and deploy analytical approaches in support of an analysis, but with some opportunities for additional learning. Interpretations of results are logical and contain a well constructed discussion with linkage to the research problem. Maps, figures, and tables are appropriate, but could be improved upon or may contain minor errors.

Mark	Guidance
Pass (50-60%/C)	The analysis is relevant to the problem but lacks focus and/or clarity. It demonstrates sound technical understanding of individual analytical approaches, but those selected may not be the most relevant to the research problem or could be better-specified. Interpretations of results are largely correct but need to be better-connected to the research problem. Maps, figures, and tables are adequate but more appropriate visualisations could have been implemented or may containt multiple errors that detract from their usefulness.
Fail (<50%)	An attempt at analysis is made but does not address the problem effectively and demonstrates that the material covered in class has not been fully understood. It demonstrates limited technical understanding of individual analytical approaches, and those selected may not address the research problem or may be incorrectly-specified. Interpretations of results may not be correct or may not have a bearing on the research problem. Maps, figures, and tables are of poor quality and do not aid understanding in any way.

Structure and Presentation (10% of total mark)

This aspect focusses on the attention you've given to communicating your results in a manner that is appropriate to the target audience:

Mark	Guidance
Distinction (70+%/A)	The briefing demonstrates an excellent awareness of its audience, and excellently balances the elements of a model data-led briefing. The linkage between sections is obvious and there is a strong narrative throughout.
Merit (60-70%/B)	The briefing demonstrates a good awareness of its audience, and there is a clear and logical structure to the data-led briefing. Attempts to link sections as part of an overall narrative are evident, but could be improved by further reflection about the audience and topic.
Pass (50-60%/C)	The report has an obvious structure but demonstrates a limited awareness of its audience. Connections between sections are weakly-articulated and the overall narrative is lost.
Fail (<50%)	The report is weakly structured and demonstrates a very limited awareness of its audience. There is no obvious linkage between sections and little narrative throughout.

The models provided should assist in determining whether you will do well in this area.

Reproducible Notebook

Reproducibility (20% of total mark)

This aspect focusses on the attention you've given to ensuring that your analysis can be run on another system without changes being necessary to the code. So this requires thinking about both 'portability' and how to ensure that data and libraries are made available to end-users.

Mark	Guidance
Distinction (70%+/A)	Your notebook runs without errors and you have clearly thought through the ways that your code might fail on another system as well as the processes that could mitigate this.
Merit (60-70%/B)	Your notebook runs without errors.
Pass (50-60%/C)	Your notebook runs with minor errors, such as to a path variable or a missed library import from the Docker image.
Fail (50-%)	Your notebook does not run without significant edits to the code, such as installing a missing library or a failed download.

If you have, for instance, a NLP analysis that would take hours to run then it is acceptable to provide a partially-processed data file (i.e. at the point where the time-consuming analysis is complete) and to 'comment out' the cells that generated this data, but you *must* include the code used to generate this data.

Quality (20% of total mark)

This aspect focusses on the attention you've given to creating an analysis that is intelligible and well-presented such that another user doesn't just 'run the code' but actually understands and has confidence in the how and why of what you've done. This is about making maximal use of the tools at your disposal such that the code is efficient, even elegant, and the outputs are clear and legible.

Mark	Guidance
Distinction (70+%/A)	The code is well-commented and each stage of the analysis is organised in a logical manner as part of a coherent and well-structured whole. Another data scientist could understand and use your code with minimal effort demonstrating a comprehensive understanding of the practices taught on the course. The code is efficient, including, for example, the creation of functions where appropriate, the use of iteration to work efficiently with data (loops), and the demonstration of more advanced programming techniques, and clever and creative use of the packages taught on the course. Outputs from the analysis are appropriate, clear, and well-designed.
Merit (60-70%/B)	The code is well-commented and each stage of the analysis is organised in a logical manner. Another data scientist could understand and use your code with some effort, demonstrating a good understanding of the practices taught on the course. The code is also efficient, including, for example, the creation of functions where appropriate, and good use of the packages taught on the course. Outputs are mostly appropriate even if they rely largely on sensible defaults.
Pass (50-60%/C)	The code is well-commented, but is unstructured in places. Another data scientist could understand and use your code, but it would require a lot of effort on their part, suggesting more limited engagement with the course material. While the code is efficient in places there are often more effective and 'legible' ways that could have been used to generate the results, including, for example, repeated code that could have been made into a function, or failing to take advantage of the packages taught in the course. Outputs are adequate for the analysis but more appropriate visualisations could have been implemented.
Fail (<50%)	The code is uncommented/poorly commented and largely unstructured. Another data scientist could not understand and use your code, suggesting almost no engagement with the fundamentals of the course. The code is generally of poor quality, highly inefficient, and riddled with errors. Attempts are made at outputs but they do not add to understanding in any way.