Checklist

1. you must use git to track your work and publish a repository of your project to github.
2. Secondly, you should use Python, the command line, or SQL (or various of these if necessary) for any data cleaning, and use Python for any web-scraping, modelling and generation of graphs and output
3. it is expected that some element of your project either uses web-scraping to generate data, or works with analysis using tools such as those laid out in unit 5 (regression modelling, causal forests, etc.)
4. additionally ensuring that all steps can be followed and replicated by someone else.
5. Please submit your project as a single compressed .zip file named as STUDENTNUMBER.zip (replacing STUDENTNUMBERwithyourstudentnumber).
6. AREADMEfilecontaining instructions on how to replicate your work
7. Awell-structured project folder including all required source code, data, and so fort
8. The output from running the code when you run it on your machin
9. Thegitfolder which tracks this code (provided that you have used git you will not need to do anything, as this should be available in the .git folder)
10. Alink to the blog post.
11. If you include your “blog post” simply as a jupyter notebook, this can just be an .ipynb file. Please call this blog.ipynb., If you publish your blog post on some web platform (which will likely provide a nicer output), please include the link to this URL ensuring that I can access it from the web. Please include this address in the file called blog.txt (which just requires the URL that I can then copy and paste)
12. Alink to the github repository– If you include your “blog post” simply as a jupyter notebook, please include the github link at the end of this file– Ifyoupublish your blog post on some webplatform, please include the link to your github folder in the in the file called blog.txt
13. Yourscripts must be sufficient to reproduce your work, and by reading the READMEandscriptsIshould have agoodideaofwhatIneedtodotoreplicateeverythingyoudo.
14. If you wish to be sure that the Jupyter file can be read easily, you could also hand in the output of your Jupyter file as a separate pdf (but please also include the .ipynb file too).
15. Length: Your post should have between 1,000–2,500 words, and it should contain 4-7 pieces of output (e.g., plots, tables, summary statistics, and so forth).
16. Insightfulness: istheanalysisinsightfulandcompelling? Aretheresomeinteresting, thought-provoking, and/or surprising findings?
17. Soundness: Is the analysis sound? Are all comparisons fair (not comparing apples to oranges)? Was proper filtering done? Are plots used appropriately based on data types (e.g., using bar plots for categorical data points, scatter plots for independent data points, lines for time series)? • Presentation: Is the narrative coherent? does it all read as a one story? Is it easy to understand the post? Is it easy to follow the ideas and the main story? Are there appropriate (and readable) labels, legends, and captions provided for plots?
18. Visual appeal: are the plots visually appealing (beautiful, appropriate colours, non-trivial yet simple plots)?
19. Pre-processing: Has there been an appropriate level of pre-processing of data? Is the code easy to follow? Is the code efficient? Did the student put some effort in cleaning or re-shaping data?