**G489/589 Final Project Report and Code Instructions and Rubric**

In the last four weeks of class you will complete your final project. The final submission will contain a written report (see rubric below) plus all the scripts you have used in your project in a Jupyter Notebook (with comments!). The aim of this project is to use many, if not all, of the Python libraries, methods and functions you have learned during this class. I expect you to go back over the assignments we have done to remind yourself of the methods used. You may find it helpful to fill in the Python Methods Table Template, which you can also find in the course GitHub repository in the “Final Project” folder (and in the Final Project assignment on Canvas).

NOTE: We have learned three uses of coding: 1) cleaning/tidying your data and preparing it for analysis; 2) Exploratory data analysis; 3) analyzing the data (often using statistical methods) to answer a specific question about the patterns contained in the data and what we can learn from them.

To achieve full marks for this project it must contain all three elements. You may strongly focus on one aspect, with slightly reduced focus on the second aspect, and a smaller focus on the third aspect.; or, you may present an equal focus on all three. If you strongly focus on one aspect there must be substantial work presented. If you are focusing on cleaning and tidying your data, you must have a good and clearly explained reason for doing so. For example, if you have a lot of data to process (many, large files with multiple issues of missing data, incorrect units etc). If you only have one dataset that does not require a lot of pre-processing, you must instead focus on extensive EDA (that leads you a long way to answering your question or posing new analyses) or an extensive statistical data analysis.

Please write all your code in a Jupyter Notebook with extensive comments for each task you complete. Points will be deducted if your code is not commented. You may choose to write the sections of your report (outlined in the table below) in the Jupyter Notebook (instead of a Word doc or equivalent) by switching the input boxes to markdown text. However, note that I do expect paragraphs – not just sentences – for each section so you can adequately describe the purpose of your project, the methods used, and an analysis and discussion of your results.

**Grading rubric**

|  |  |  |
| --- | --- | --- |
| **Section** | **Points** | **Marks for:** |
| **Introduction** | **10** | * Clear out line of the purpose of your proposed project with aims you wish to address and questions you wish to answer. For example, an aim might be to tidy and clean your data in a certain way. The questions will be related to something you want to learn from the data by analyzing the patterns in the data. * Supporting literature or references (including websites) if any |
| **Methods** | **25** | * Appropriate analysis for your project design 🡪 adequate to meet the aims/questions you defined in the introduction * *Explain which of the three elements you will focus on (see above)* * Clear explanation of all the methods and functions you will use, and why * The number of different methods, functions and libraries that you use (appropriate for your project) * Description of the data you are using, including where you got it from, its attributes/characteristics, which sites, why these data are appropriate. |
| **Results** | **60** | * Clear, efficient and commented code (for example if you are reading in multiple files you should do so in a loop). * Analysis that corresponds to the methods you have proposed. * Appropriate and accurate implementation of your analysis methods * Well formatted plots and analysis outputs that clearly show all the results of your analysis. * *Print-outs/plots demonstrating the results of your data cleaning/tidying and exploratory data analysis to show the steps taken. I need to see examples of every bit of the code you are using so I can better understand what you are doing and to see that it works. E.g. if you merge two dataframes – print the .head() before and after the merge. If you calculate a linear regression but do not plot it, print the results out instead.* * *Descriptions of your results based on your analysis and outputs (figures, tables etc).* |
| **Summary and discussion** | **5** | * A discussion of what you have learned from your analysis * Explanation of any pitfalls you encountered and how you addressed them. |
| **TOTAL** | **100** |  |