REFERRAL Statistics and Trends (20%)

* **Due** 17 Jun by 12:00

* **Points** 20

* **Submitting** a file upload

* **File types** pdf

* **Available** 24 May at 12:00 - 17 Jun at 12:00

|  |  |  |  |
| --- | --- | --- | --- |
| **Weighting %:** | 20 | **Submission deadline (for students):** | 17/6/24 at 12pm (Midday) |
| **Authorship:** | Individual | **Target date for returning marked coursework:** | 24/6/24 |
| **Tutor setting the work:** | Dr. William Cooper | **Number of hours you are expected to work on this assignment:** | 10 |

|  |
| --- |
| **This Assignment assesses the following module Learning Outcomes (from Definitive Module Document):**  1. Be able to maintain and develop code using the git version control system.  2. Be able to apply different techniques for cleaning data and preparing it for analysis.  3. Be able to communicate their findings to others, including a critical assessment of performance.  4. Demonstrate knowledge and understanding of the concepts of version control for code development.  5. Demonstrate knowledge and understanding of key data manipulation techniques for data preparation.  6. Understand how to approach a range of different data science problems to obtain an efficient solution. |
| **Assignment Tasks:**  You will create a well-written report exploring and explaining any relations within a dataset. You can download any dataset from Kaggle/Worldbank/etc. Be sure to include your name, student number and a link to your GitHub repository in the report. There will be at least three plots: a histogram/bar chart/pie chart; a line/scatter graph; a confusion matrix/heatmap/corner/box/violin plot. The code will contain evidence of the creation of any displayed graphs (one graph per function) and the creation of any quoted statistic. The minimum expected statistic is the use of a pandas dataframe *describe*and *corr*. |
| **Submission Requirements:**   A two page PDF report, including a functional link to your GitHub repository containing your python code (either notebooks or plain python). Check that your repository link is both clickable and links to a **public** repository. |
| **Marks awarded for:**  See rubric. |
| **Type of Feedback to be given for this assignment:**  Written feedback within the rubric. |
| **Additional information:**   * Regulations governing assessment offences including Plagiarism and Collusion are available from [https://www.herts.ac.uk/\_\_data/assets/pdf\_file/0007/237625/AS14-Apx3-Academic-Misconduct.pdfLinks to an external site.](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct.pdf) (UPR AS14) . * Guidance on avoiding plagiarism can be found here: <https://herts.instructure.com/courses/61421> (see the Referencing section) * For postgraduate modules:   + a score of 50% or above represents a pass mark.   + late submission of any item of coursework for each day or part thereof (or for hard copy submission only, working day or part thereof) for up to five days after the published deadline, coursework relating to modules at Level 7 submitted late (including deferred coursework, but with the exception of referred coursework), will have the numeric grade reduced by 10 grade points until or unless the numeric grade reaches or is 50. Where the numeric grade awarded for the assessment is less than 50, no lateness penalty will be applied. |