7143CEM Programming for Data Science — Portfolio (draft version)

Task 3. Critically assess, select and apply data science tools (ILO3)
This task involves creating interesting exploratory graphical plots of a dataset using appropriate Python libraries and then critically assessing/comparing the capabilities of those plotting libraries.

(1) Select one or two publically available datasets that have some interesting features, e.g., from Kaggle, fivethirtyeight, or the UCI Machine Learning Repository. Consider building two polished nontrivial graphical plots that each illustrate something interesting about the dataset. The plots must be substantially different, and one of them should involve some complexity, e.g., a Mekko chart, a circular barplot, or a hexagonal heatmap. For each of the two plots, build that plot in three separate ways, each using a different Python plotting library. For example, you might use Matplotlib (alone) to build a stacked barchart and then also build a similar stacked barchart using each of two other Python plotting libraries separately. *You must use at least two of seaborn, plotly and plotnine plotting libraries. The Python Graph Gallery (https://python-graph-gallery.com/) is a good place to start. Aim to make the plots of the same type as similar as possible using the different plotting libraries. You could have two plots using the same dataset or each plot using different datasets. Clearly show where each dataset comes from and give a small snippet to show the structure of the dataset.*

[10 marks]

(2) Drawing on your experience from part (1) and additional sources (e.g. articles, blogs, documentation or tutorials), critically assess the libraries you used to produce these particular plots in terms of: (a) difficulty of coding, (b) adaptability of the code, (c) level of control over the plot layout or labelling, and (d) quality of the graphical plot produced. *Ensure you clearly reference any sources you use. You may find it useful to present your critical assessment in a table.*

[10 marks]

(3) Suppose you wish to visualise geospatial data. Briefly describe the capabilities of one Python library that you would recommend and give an example (code and output) using UK data.

[5 marks]