

Data Visualisation: Homework

(3 points)

Blerina Sinimeri and Lorenzo Madeddu

March 7, 2022

Deadline: Submit your work on luis.learn by **23:59 (CET) on Monday, March 14.**

For this assignment we will use the following dataset available in <https://www.kaggle.com/prasertk/healthy-lifestyle-cities-report-2021> but also in the course website on luis.learn. The dataset contains 10 healthy living metrics in top 44 cities measured on 2021.

Each row in the assignment datafile corresponds to a city. The following variables are provided to you: (1) City Rank, (2) Sunshine hours, (3) Cost of a bottle of water, (4) Obesity levels, (5) Life expectancy (years per Country), (6) Pollution (Index score), (7) Annual avg. hours worked, (8) Happiness levels (Country) (9) Outdoor activities (City), (10) Number of take out places(City), (11) Cost of a monthly gym membership (City).

Familiarize yourself with the dataset, then write a Python code for the following questions:

- (a) Check if the dataset has missing values and choose a way to correct them. Motivate your choice. (0.25 points)
- (b) Plot a bar chart describing the Happiness levels for each city. The bars should be colored in gray except for the bars corresponding to Italian cities which should be colored in dark blue. (0.25 points)
- (c) Plot the correlation between any two pair of variables as a heatmap with a custom palette. What can you say about the result? (0.5 points)
- (d) Divide the cities into 6 groups based on the first letter of their name (A-D, E-H, I-L, M-P, Q-U, V-Z). Plot a bar chart for the “Number of take out places” variable with six bins (A-D, E-H, I-L, M-P, Q-U, V-Z). Each bin contains the sum of the values of “Number of take out places” for the cities that fall in that bin. (0.75 point)
- (e) Plot the graphics as described in Figure 1 (0.75 point).

Make the visual nice and select wisely what kind of palette apply for each figure! Leverage principles from what we have seen in this course when developing your solution. Consider issues such as legends, labels, and chart junk. We will also take into account the quality of your code.

Format: You have two possibilities: (1) You submit in a single zip folder: the python code and all files corresponding to the figures produced (2) You submit a single Jupyter Notebook with your code.

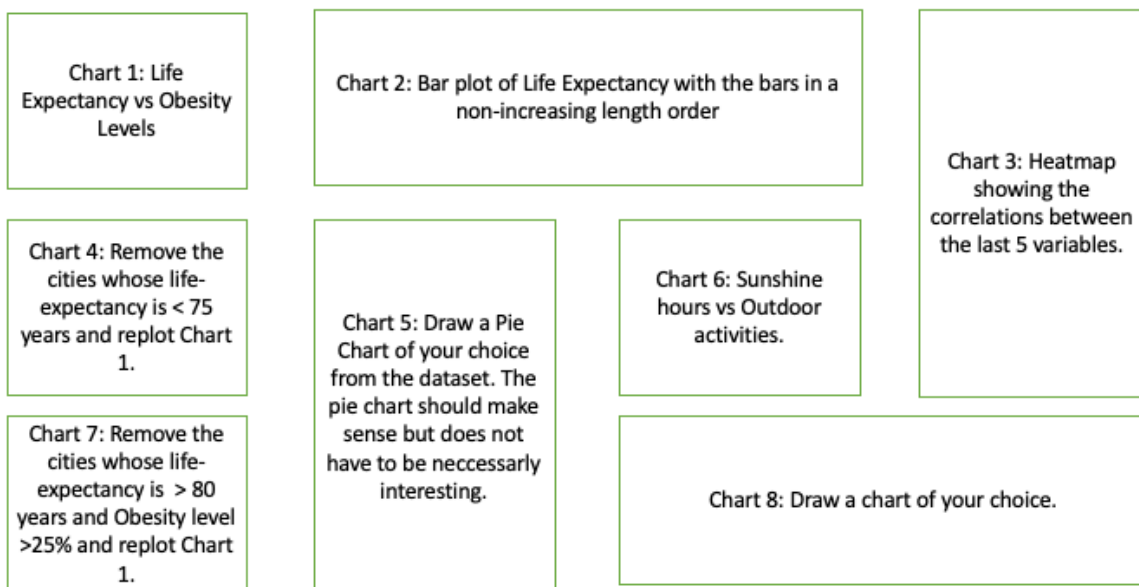


Figure 1: Create the charts according to this figure.