

# CS 210 Homework 4

Spring 2023

## Overview

In this assignment, we'll do some analysis and visualization of EU city temperatures. Please write your code in the provided Jupyter notebook, `eu_city_temps.ipynb`. You can add additional markdown cells as needed so it's clear what part of the problem you're solving. Once complete, submit your notebook on Canvas. You should not include any input or output files.

## Preprocessing/Analysis

1. Load the CSV file `EuCitiesTemperatures.csv` into a Pandas DataFrame. The first few records (of 213 total) are shown below:

city	country	population	EU	coastline	latitude	longitude	temperature
Elbasan	Albania	2.9	no	yes	41.12	20.08	15.18
Andorra	Andorra	0.07	no	no	42.5	1.52	
Innsbruck	Austria	8.57	yes	no		11.41	4.54
Graz	Austria	8.57	yes	no	47.08		6.91
Linz	Austria	8.57	yes	no	48.32	14.29	6.79

2. Fill in the missing latitude and longitude values by calculating the average for that country. Round the average to 2 decimal places.
3. Find the subset of cities that lie between latitudes 40 to 60 (both inclusive) and longitudes 15 to 30 (both inclusive)
4. Find out which countries have the maximal number of cities in this geographical band (from #3). More than one country could have the maximal number of values.
5. Fill in the missing temperature values by the average temperature value of the similar region type. A region type would be a combination of whether it is in EU and whether it has a coastline.

For example, if we have a missing temperature value for Bergen, Norway, which is not in the EU but lies on the coast, we will fill it with the average temperature of cities with `EU='no'` and `coastline='yes'`.

## Visualization

For all plots, make sure to label the axes, and set appropriate tick labels. You can use seaborn instead of or in addition to matplotlib if you like.

1. Plot a bar chart for the number of cities belonging to each of the regions described in Preprocessing/Analysis #5 above.
2. Plot a scatter plot of latitude ( $y$ -axis) vs. longitude ( $x$ -axis) values to get a map-like visual of the cities under consideration. All the cities in the same country should have the same color.
3. The population column contains values based on country, so two cities of the same country will show the same population value. Plot a histogram of the number of countries belonging to each population group. Split the population values into 5 bins (groups).
4. Create subplots in a  $2 \times 2$  grid, with proper titles, one each for the region types described in Preprocessing/Analysis #5 above.

Each subplot should be a scatter plot of Latitude ( $y$ -axis) vs. City ( $x$ -axis), where the color of the plot points should be based on the temperature values: red for temperatures above 10, blue for temperatures below 6, and orange for temperatures between 6 and 10 (both inclusive).

For each subplot, set `xticks` to an array of numbers from 0 to  $n - 1$  (both inclusive), where  $n$  is the total number of cities in each region type. I.e., each city is represented by a number between 0 and  $n - 1$ .