Use linear regression to predict COVID-19 mortality in the countries worldwide. For each country, let us use the following two features:

1) percentage of people over the age of 65 in the population. This information can be found at https://en.wikipedia.org/wiki/List_of_countries_by_age_structure
2) the number of hospital beds per 1,000 people in the most recent reported year. This information can be found at https://data.worldbank.org/indicator/SH.MED.BEDS.ZS

The response variable for each country will be the number of COVID-19 deaths per 100K population. This information can be found at

https://coronavirus.jhu.edu/data/mortality

To train the model, use the data from 10 countries of your choice. Use the 5-fold cross-validation to validate the model. It means that you should split the countries into 5 groups, each containing 2 countries. For each group, train the model on the remaining 8 countries, apply the trained model to 2 countries from the group and calculate the prediction error for them. An average error over all iterations will be an assessment of the quality of linear regression model.

You can use Linear Regression functions from **sklearn** package: https://scikit-

learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

Please submit

1) Your source code written in Python. Do NOT archive your **py-** or **ipynb-files**. You may need to add the extension **txt** to it to be able to upload it on iCollege dropbox.
2) All external files with the data used in your python script. Please make sure that the paths in your script are set to be the current directory. To NOT use your own paths!
3) A text file that contains
    a) the list the countries that you used to train the model;
    b) the results of 5-fold cross-validation.