# Study project. Clustering world nations by a set of different features

This is a continuation of another project that compared religiousness of population to other factors: life expectancy, corruption, democracy index, Freedom In the World index, GDP per capita, Human Development Index, homicide per capita, proportion of religious people and population.

All data were taken from Wikipedia, collected in a single file (Raw\_data.xlsx) and then prepared and saved in individual csv files (folder raw\_data).

I also removed and renamed some columns for this project (folder prepared\_csvs).

Data were merged into a single data frame and prepared for further analysis: removed rows (countries) with missing data and scaled all variables.

I considered two options: with and without the size of population.

PCA was performed for both options, showing several variables with almost equal contribution to the first principal component.

This was followed by KMeans clustering with silhouette scores used to select the best clustering approach.

Notebooks:

- data\_preparation\_analysis.ipynb – data preparation and PCA

- country\_clusters\_with\_population.ipynb – clustering with data on population

- country\_clusters\_without\_population.ipynb – clustering without data on population