**“THE” Place to Live**

**Introduction**

In this study I will implement an application for identifying “good” and “bad” countries for the person to live in, as well as provide a more granular suggestions by ranking countries based on person’s individual preferences (in terms of acceptable values for parameters and priority set to them), such as: cost of living, population density, light pollution and others.

**Methods**

To be used as parameters, datasets on cost of living, population density, and light pollution (current list, more will be added soon) were gathered from \_\_\_\_\_\_\_\_\_\_\_.

Since datasets for this study aren’t that big – in most cases couple hundreds of countries, they were manually cleaned from any unnecessary data and checked for having the same number of countries and that the countries are the same.

To make my system work, necessary parameters will be chosen by system’s user in the beginning and acceptable value ranges will be chosen to create a custom(personalized) decision tree to help with categorizing our data into “good” and “bad”.

To further refine our results, two separate systems will be used – priority list and k-means clustering algorithm. When user will select the parameters for their research, the user will be prompted to assign a priority to each of the parameters they chose. It will allow to favor certain parameters compared to others when time comes to calculate countries “priority points”, which will calculate a value for each of the countries based on priorities set for different parameters, thus allowing to rank countries from “best” to “worst” in their respective categories (“good” or “bad”).

To categorize countries in a vaguer way, excluding the priorities of the parameters, k-means clustering will be used to categorize countries in three distinct groups: “best fit”, “medium fit”, “worst fit”. This will create more of the suggestive categories for people to observe. For example: let’s take our theoretical “good” group of countries and run it through k-means algorithm; this will yield three subgroups of “best fit”, “medium fit”, and “worst fit” countries from our master-group “good countries”.