Timothy Abramov

Project Proposal

Data Mining

As I’m slowly approaching my graduation date, I’m starting to think more and more where would I like to live after college. There are a few criterias I want to explore:

* First of all, since my area of work most likely will be remote computer-based work, the cost of living or the price index of some sort is pretty important to me. The lower the better.
* Another criteria for me would be the amount of light pollution in the area. The less the light pollution - the prettier is the night sky and potentially better sleep quality for me.
* There are some other parameters that I’m considering, such as: population density, weather throughout the year, and probably landscape features (mountains, hills, rivers, ocean, lakes, forests, etc.). Also, primary language in the region/country might be important too.

In terms of datasets, I know that there exist cost-of-living datasets out there. And I would assume that most of the data I need can also be found.

By the end of it I would like to have a list of countries that qualify by parameters set by me.

In terms of Data Mining techniques, I’m not exactly sure what will suit the problem the best, and if this problem is actually worth a DM treatment, but it’s definitely an interesting question for me to explore.

Satisfactory

Seems like a god topic. You might also think about enlarging it to focus on region (cities instead of countries). Also, better to label data as “good / not good” and then try to learn patterns with e.g. decision tree to get more DM treatment.