Eli Mendels and Matt Catalano

10/13/2019

Our primary plan is to use a combination of datasets to predict which cities or locations would be the happiest places to live. The process will be broken up in four major sections. The first is combining data from the world happiness index with demographic and socioeconomic data about each country on the list. To do this we will combine a variety of datasets so that we have a multitude of information about each country. Our next step will be to analyze the data and find out which metrics correspond the most with happiness. We plan to use some machine learning to create a model that will weigh each metric so that we can determine how much each statistic affects a country’s happiness. We will also try and look at the data to find tiers, if they exist, amongst the happiness of each country so that the model behaves slightly differently for very happy countries as opposed to very unhappy countries. Finally, once we have created this model we will attempt to apply it to databases of cities or other locations to determine, based on demographic and socioeconomic metrics, which locations are likely to be happiest.

Our main research questions are whether there are certain metrics of a country that can accurately predict or signify how happy that country is and then whether those same metrics can be applied to other locations such as cities or states. Our hypothesis is that there will be some factors such as income, health care, and crime rates that will correlate strongly with happiness and can therefore be used to predict happiness on a smaller scale.

We have linked three relevant datasets that we will be utilizing in our project at the bottom of this page. One is the world happiness report which attributes happiness scores to countries based on answers to the main life evaluation question asked in the Gallup World Poll (this question being to rank one’s life from a 0 to a 10 where a zero would be the worst possible life and a ten would be the best possible life). The second dataset we will be using comes from the CIA’s World Factbook and consists of country profile data for all the countries in the world. This dataset includes country stats on GDP, adult obesity prevalence, unemployment rate, education expenditures, and health expenditures which will all surely be useful in determining what makes citizens of certain countries happier than others. The last dataset is taken from data.gov and provides socioeconomic indicators like poverty rate, unemployment rate, education levels at a county-level in the United States. This last dataset will be the final piece of our puzzle to connect the world happiness rankings to the happiest cities in the United States.

Hopefully, through this process, we will be able to create a system that can be used to determine the happiness of any location given certain statistics on the area. We believe that this could be incredibly useful to city governments as a high happiness ranking for a city could lead to more residents settling in the city. On the flip side, knowing what the cause of high versus low happiness is could lead to some cities with low happiness making adjustments to fix the problem.

[https://kaggle.com/unsdsn/world-happiness](https://www.kaggle.com/unsdsn/world-happiness)

<https://github.com/thewiremonkey/factbook.csv>

<https://catalog.data.gov/dataset/county-level-data-sets>