

City/neighborhood similarity

July 7, 2020

Introduction

Last year, a poll found that US citizens desire to relocate to Canada[1] does hit a new record. 16% of US citizens said that they would like to permanently move to another country. Another poll [2] studies the worldwide sentiment to migrate to another country, and the favorites are the USA, Canada, Germany, France, and Australia.

Right now, we are suffering the slashes of a pandemic, COVID-19. Admittedly, this will redefine the migratory movements for individuals and company relocations around the world. Now, we can use the time to redefine this exhausting and determining kinds of projects.

We are not interested in the factors that drive people to migrate, which is a complex subject [3]. We want to help people to decide the city of destination based on the city of origin.

In this brief study, we would like to compare cities or neighborhoods, using an arbitrary set of criteria to give a rough idea of similarity between towns. For instance, we can use

the foursquare API to classify one location given a set of features, and add other ones like services and cost of living.

Data

1. **Foursquare:** We will characterize the city/neighborhood main venues and services by using the foursquare API.
2. **Wikipedia/Web:** We will collect other features by web scraping, such as services available, airports, etc. Wikipedia will be a resource.
3. **expatistan.com:** The main feature will be the cost of living, which will narrow the cities of destination to compare. The site expatistan.com offers such a service.

Libraries and tools

1. Pandas and numpy for data manipulation.
2. BeautifulSoup and requests for web scrapping.
3. Foursquare API, bing! API, and geocoder. For venues and geolocation.
4. Folium for maps.
5. Matplotlib for visualization.
6. Dotenv for api keys manipulation.
7. Scikit-learn for k-means clustering.

Methods

First, we need to establish the country and city of origin. Second, we define the target country, i.e., the destination. With this information, we will use a recommender engine to find similarities between the cities.

When a couple of suitable destination cities are found, we will use k-means clustering to group neighborhoods with similar features, using the Foursquare API. Also, the same analysis we will perform in the community of origin. We will locate the clusters in a Folium map.

With all this data, we will report the destination neighborhoods that are similar to the neighborhood of origin.

References

1. <https://news.gallup.com/poll/245789/record-numbers-americans-leave.aspx?>
2. <https://news.gallup.com/poll/245255/750-million-worldwide-migrate.aspx>
3. Francesco Castelli, Drivers of migration: why do people move?, Journal of Travel Medicine, Volume 25, Issue 1, 2018, tay040, <https://doi.org/10.1093/jtm/tay040>