

Similar Neighborhoods in two Cities

Mriganka Medhi

September 8, 2019

1- Introduction

1.1- Background

Moving from one city to another can be a very tedious and time-consuming task. On moving, people often tend to find it difficult to adjust to the new surroundings. They definitely prefer to find areas that are similar to their previous areas of residence. This can make it much easier for them to adapt to life in a new city. Also, they would not need to change their habits and lifestyle if they can manage to find similar neighborhoods. Hence, it could be very beneficial for people planning to move between cities to gain valuable information about the neighborhoods in the new city that are somewhat identical to their current city.

1.2- Business Problem

In this project, data related to the neighborhoods of the two cities of Adelaide, Australia and Melbourne, Australia will be used. The aim of this project is to find the similar groups of

neighborhoods in the two cities and cluster them based on their location data, i.e. data about the types of venues present in the vicinity of the neighborhood. If the common venues around a neighborhood are similar, then the neighborhoods are likely to be similar to each other.

1.3- Interest

The similarity of neighborhoods in different cities is of interest to people who have to move from one city to another (specifically, in this case, from Adelaide to Melbourne or vice-versa). It may also be useful for travel agencies and consultancies that advice people on the suitability of neighborhoods.

2- Data Acquisition, Wrangling and Utilization

2.1- Data Sources

The list of neighborhoods/suburbs, their postal codes and local government areas for Adelaide and Melbourne are present in the Wikipedia pages

https://en.wikipedia.org/wiki/List_of_Adelaide_suburbs and https://en.wikipedia.org/wiki/List_of_Melbourne_suburbs respectively. The latitude and longitude values are obtained

using the python library geopy. The location data for the neighborhoods is obtained using the Foursquare API.

2.2- Data Wrangling

Initially, the two tables containing the neighborhood details are converted to pandas dataframes. Only the columns corresponding to suburb, local government areas and postal codes are kept. The suburbs with the same postal codes are combined into single rows, such that their names are separated by commas. Two new columns for latitude and longitude of each postal code are added. The two dataframes are concatenated, one after the other so that the details of all neighborhoods in both cities could be obtained in a single dataframe.

2.3 –Utilization of Data

The latitude and longitude of each postal code is used to obtain the corresponding location data using Foursquare. Using the location data, the most common venues for each venue is obtained. This data is used to cluster the neighborhoods into several disjoint clusters. The neighborhoods (of both cities) that fall in the same cluster can be considered to be similar to each other. The clusters can be further visualized on a map using Folium library. Hence, this data can be very useful to gain

insights about similarity of neighborhoods among the two cities.