Data Science Capstone Project

**Where to live in Vancouver?**

1**. Business Problem**

Vancouver, a beautiful west coast city in British Columbia, Canada, is always well-known for its continuous present in the list of top cities in the world for quality of living. As a result, Vancouver is a very attractive destination for both immigrants and tourists. Although the city is quite small comparing to other giant metropolitan cities, it stills has 22 official neighborhoods (or areas) and the **decision of where to live (for immigrants) or where to stay (for tourists) is quite a challenging one**.

The purpose of this project is to use location data from Foursquare to **give recommendations to immigrants and tourists on which neighborhood(s) that best fit to their specific preferences**. For example, an immigrant family with kids may value Education, Shop & Service and Outdoor Activities higher than other criteria. On the other hand, a young tourist couple may give Food, Transportation and Nightlife higher weights than other criteria. This project aims to provide recommendation for such specific preferences.

With such problem and solution, this project will offer values to the following stakeholders:

* Immigrants who want to live in Vancouver
* Tourists who want to visit Vancouver
* Real estate agents who want to give recommendations to their clients
* Real estate renting platforms such as Airbnb to help their users searching for neighborhoods that meet their references

**2. Data**

The following data will be used in this project:

* A list of all neighborhoods (or areas) in Vancouver. This can be retrieved from the city’s official website at <https://vancouver.ca/news-calendar/areas-of-the-city.aspx>. There are 22 of them totally.
* A map of Vancouver splitting into neighborhoods. This is for reference only. Source: <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver#/media/File:Stadtgliederung_Vancouver_2008.png>



* Latitude and longitude of each neighborhood. This can be retrieved by using the Python library ***geopy***. The data is something like this:

Table

Description automatically generated

The result is stored in a CSV file, which is also available on GitHub at: <https://raw.githubusercontent.com/uclinux83/Coursera_Capstone/master/vancouver_neighborhood.csv>

* Foursquare venues data retrieved by calling Foursquare APIs (mainly the ***/venues/explore*** API endpoint). The returned data will be in json format, which will be processed later using Python.

Text

Description automatically generated

* Foursquare venue category hierarchy and ID, which is available at: <https://developer.foursquare.com/docs/build-with-foursquare/categories/>. Categories in the list will be selected and mapped into the following 9 criteria (which will be used by the users to setup their preferences):
  + Arts & Entertainment
  + Education
  + Food
  + Nightlife
  + Outdoors & Recreation
  + Health
  + Spiritual Center
  + Shop & Service
  + Public Transportation

These category IDs will be used as the input parameter ***categories*** when calling the ***/venues/explore*** API

The overall approach is to get all venues within or around each Vancouver neighborhood and group them into the 9 “criteria” as listed above. Users can set their specific preferences by giving weight to each criteria. The system will then recommend the neighborhood basing on such preferences.