Capstone Project

Problem Discussion & Background

- With the advances in globalization and remote work, it is increasingly frequent that employees, entrepreneurs and self-employed professionals have opportunities to relocate to different cities across the world to work
- However, one does not always have the ability to visit and experience the cities they will live in before they actually move. This can be stressful for families as they move into different cities they do not know and may have a hard time to get adapted to.
- This project aims to provide insights on how similar certain cities across the world are, from the point of view of availability of commercial and services venues, according to the local population.

Target Audience

- The target audience for this project is the current and future expat community, composed by professionals who have relocated or want to relocate to different countries. For such professionals, it is always important to learn more about their prospects of personal life in the place they are about to move to.
- Obtaining information on the commonly offered commercial, service and leisure options will allow the audience to build an initial perception of the top cities in the country they intend to live in, imagine how their personal life could be compared to their home city, and effectively prioritize their job search.

Data Description

- The external data sources used for this project are:
  - Foursquare venue review data – using Foursquare API (https://developer.foursquare.com)
  - Simple Maps World Cities open database – database of the world's cities and towns (https://simplemaps.com/data/world-cities)
- Below are some examples of features we can extract from each dataset:
  - **Foursquare –** Extract venue information for a selected venue; view user scores and reviews; view user comments; get nearest places to a given location
  - **Simple Maps –** City or town name, country, isocode, population, capital city status, and geo localization

Methodology

The methodology approach to solve the presented problem is the following:

1. Capture user input as to the current city they live in;
2. Capture user input as to the current country they want to move to;
3. Use Simple Maps database to get coordinates for venue review search;
4. Use Foursquare data to obtain the current city profile and for the top 20 cities in the destination country;
5. Use an Euclidian distance function to calculate distance between cities;
6. Rank the top 5 cities most similar to the input city, and present results to user.

Results

User inputs: current city and destination country (for comparison) ¶

```
In [ ]:  current_city = 'Paris'

         destination_country = 'United States'
```

```
In [276]: x = euclidian_distance(venues_grouped, 'Paris')
          x['lat'] = destination_df[['city', 'lat', 'lng']].sort_values('city').rese
          x['lng'] = destination_df[['city', 'lat', 'lng']].sort_values('city').rese
          print('List of top 20 cities by highest similarity (lowest Euclidian dista
          x[['city','distance']].sort_values('distance')
```

```
List of top 20 cities by highest similarity (lowest Euclidian distance):
```

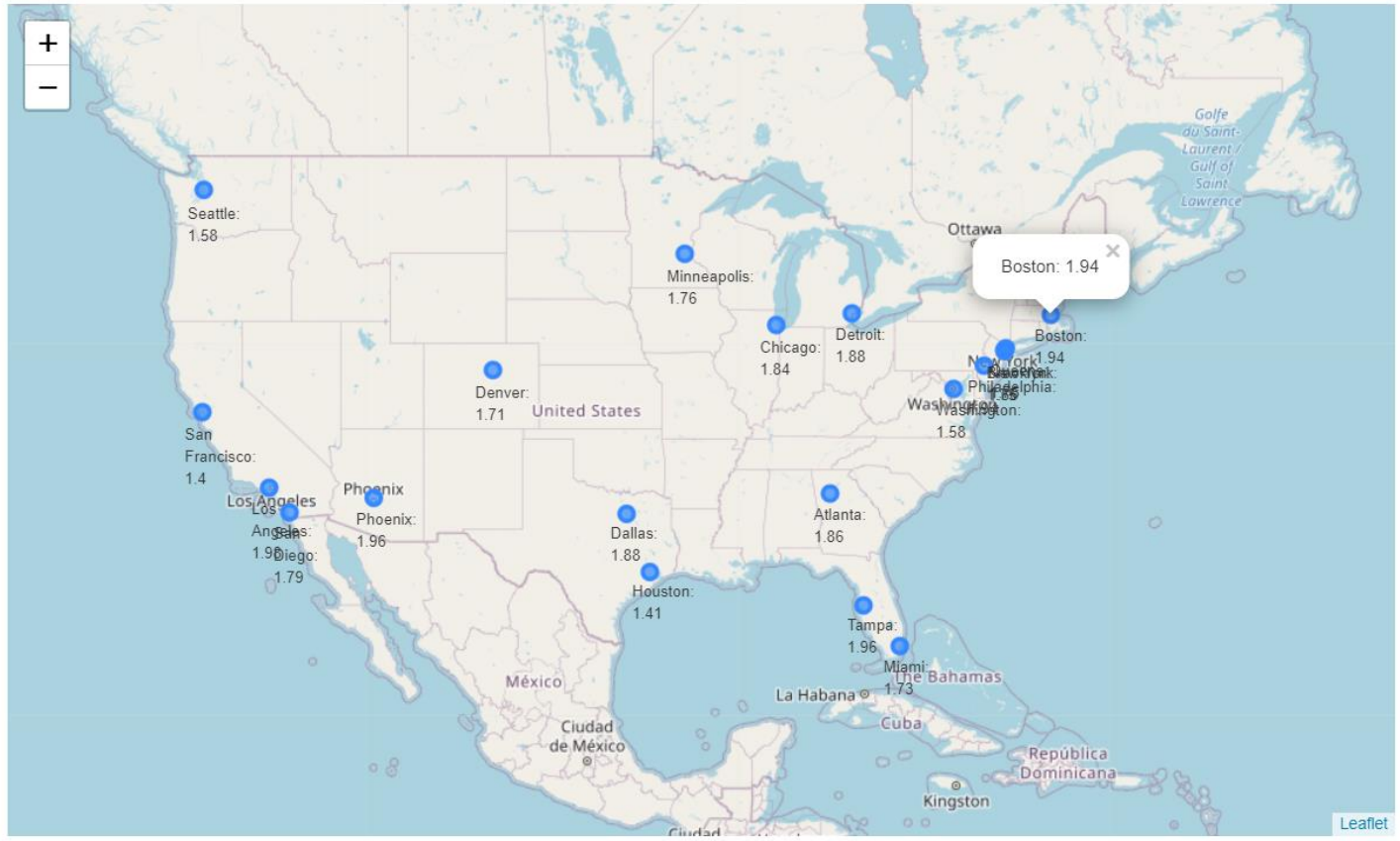
Out[276]:

	city	distance
16	San Francisco	1.400000
7	Houston	1.415000
19	Washington	1.580000
17	Seattle	1.580000
11	New York	1.660000
5	Denver	1.713043
9	Miami	1.733043
2	Brooklyn	1.750112
14	Queens	1.755000
10	Minneapolis	1.760000
15	San Diego	1.794545
3	Chicago	1.840000
0	Atlanta	1.860000
4	Dallas	1.877436
6	Detroit	1.880000
1	Boston	1.940000
12	Philadelphia	1.940000
18	Tampa	1.960000
13	Phoenix	1.960000
8	Los Angeles	1.980000

Additional Visualization – plotting with Folium

```
map_clusters
```

Out[282]:



Discussion

- The results obtained are according to expectations, as among the top cities in the US, cities well known for having high immigrant influx and international influence are ranking highest.
- Interestingly, cities with smaller geographical areas also tend to rank high due to the dense number of options found within a radius of their city centers. This is another feature that can be further explored in future versions of this project.
- The model also provides a map view of the same results, plotting the candidate destination cities and the respective distances to the home city.
- On the East Coast, the data is cluttered as many cities are very close to each other. Therefore, the map also has a tooltip menu which shows the city name and score upon clicking.