WEEK 4 PART 1

Introduction/Business Problem Introduction/Business Problem

This problem seems very interesting as we need to take care of multiple parameters into account for selecting any neighborhood region. As a person I would prefer the region where I would go is more familiar to me, or within a week it should be second homeplace to me. There are many primary things to consider here like, demographics of region like weather conditions, clock time difference, food and culture, some government data about how favored it was in past. But these things are just for so called secondary thing which is business idea.

Usually, citizens prefer places with moderate climate situation, having historical sites, city which is more similar to New York city (considering American Citizen), city having variety of restaurants like continental, sub-continental, language preferably English. So, for a guy living in New York city, if they want some outing so preferable locations might be **Bangkok, Tokyo.**

Some important points:

- People prefer language and culture most according to me.
- Worldwide reputation of the city also important, say ease of doing business ranking.
- How similar the location is when compared to their native place.
- Different tourists' attractions and sites to visit in that place.
- Budget.

WEEK 4 PART 2

Data

I used data of all the above-mentioned cities, and used various techniques as listed in project itself to explore venues at each place. Also, data cleaning and selection is performed to narrow down the search space and get more accurate result. So, I listed out proposed steps below:

- I will be using Dataset which contain all the required geographical data about New York City.

- To be more specific I am using 'Borough', 'Neighborhood', 'Latitude', and ‘Longitude’ kind of fields in dataset. The dataset is already gathered in week 3 of this course but need few fine tunings.
- Geo-coordinates of districts will be obtained with the help of the geo-coder tool in the notebook.
- Counting the occurrence of venues in each city and collected them all in one dataset so that we can see frequency of each place. Then we calculate the probability of each venue and make clustering to find the similar cities.