

Introduction/Business Problem Introduction/Business Problem

For an American citizen who lives in New York city, they decided to go on a vacation to a south eastern country, and specifically they has to decide the venue to be in one of the following cities:

1. Kuala Lumpur
2. Bangkok
3. Tokyo

The citizen preference is Historic places, but they also want to compare which city is more similar to New York City in different categories such as restaurants (different types such as Chinese, Spanish...etc.), hotels, cafes and historic places.

The citizen also needs to know what the frequency of each place in each of those cities is! In addition, the probability that they would find different venues in those cities.

Data

I used data of all the above-mentioned cities, and used Foursquare to explore venues at each place using my free account, which makes the data limited to a specific number of observations.

I then collected the category of each venue in all those cities, latitudes and longitudes to plot a geo-map of those venues, started comparing the categories of each city, counting the occurrence of venues in each city and collected them all in one dataset so that you can see frequency of each place. Then we calculate the probability of each venue and make clustering to find the similar cities.