# Opening a New Hotel in Taipei City

Angela L

May 11th, 2020

## 1) Introduction/Business problem

This project will help develops decide where to open a hotel in Taipei, Taiwan based on the present competition in the different districts in the city. Taipei, as the capital city, relies heavily on tourism and prominent substitutes of hotels such as hostels and AirBnB are not as popular here. The hotel business is, therefore, a great business opportunity in Taipei if done correctly.

Here are the steps to tackle this problem that will be more closely explained in the latter part of this report:

1. Build a data frame of neighborhoods in Taipei, Taiwan by web scraping the data from Wikipedia page
2. Get the geographical coordinates of the neighborhoods
3. Obtain the venue data for the neighborhoods from Foursquare API
4. Explore and cluster the neighborhood
5. Select the best cluster to open a new hotel

## 2) Data

The data that will be used include Wikipedia raw data on the different districts in Taipei, Foursquare's data on the different categories of known establishments in the city by district, as well as of course the geographic coordinates of the districts and the establishments.

2-1) Raw Wikipedia data on the different districts in Taipei

It is important to have a basics categorization of the city to analyze a good location for a new hotel. The government designed "districts" is a good start because there are defined boundaries as well as different characteristics among the districts. For example, upon analysis, there might a cluster of districts that have a high concentration of hotels and they are all located in a certain part of the city.

The source of the data will be https://en.wikipedia.org/wiki/Category:Districts_of_Taipei

2-2) Foursquare location data

This data is important because it is rich in knowledge about what exactly goes into each district ranging from metro stations to types of restaurants. And "hotel" is highly likely a part of the dataset since it is a metropolitan city. Upon analysis, we can understand how the existing hotels are dispersed throughout the city against the defined district areas that we analyzed previously with Wikipedia data.

2-3) The geographic coordinates

The coordinates help us, later on, create map visualization of the results to understand visually the existing competition of hotels in Taipei and using other data visualization specifications to find out if there are certain clusters of areas that have varying levels of competition. The source of the data will be retrieved from the geocoder library from Python.

## 3) Methodology

3-1) Build a data frame of neighborhoods in Taipei, Taiwan by web scraping the data from Wikipedia page

- Import into a dataframe the data about the districts

3-2) Get the geographical coordinates of the neighborhoods

- Retrieve the district geographic coordinates by using ge_latlng function
- Merge the known coordinates with district names for a dataframe of precise locations
- Display visually on a map of the districts with known coordinates

3-3) Obtain the venue data for the neighborhoods from Foursquare API

- Find out the venue lists by creating a new dataframe organized by the neighborhood/district names and coordinates
- Using the groupby and count function to understand dispersion of venues
- Then using unique names of the venue categories to further analyze the desired solution (hotel competition)

3-4) Explore and cluster the neighborhood

- Using one hot encoding and frequency valuation to transform venue data
- Create a new dataframe with just the hotel data from the transformed set

- Build clusters using kcluster function and merge with the hotel data by using the kmeans function to define the clusters
- Visualize the clusters onto a map with different colors to understand visually where the the clusters are at and if there are unique locations to be identified

3-5) Select the best cluster to open a new hotel (see Results & Discussion section)

**4) Results**

```
# Cluster 0
tpe_merged.loc[tpe_merged['Cluster Labels'] == 0]
```

| | Neighborhood | Hotel | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 3 | Eastern District of Taipei | 0.00 | 0 | 25.26553 | 121.52270 |
| 5 | Jingmei District | 0.00 | 0 | 24.99311 | 121.54135 |
| 6 | Muzha District | 0.00 | 0 | 24.98879 | 121.56373 |
| 8 | Neihu District | 0.02 | 0 | 25.06909 | 121.58847 |
| 12 | Wenshan District | 0.00 | 0 | 24.98974 | 121.56963 |
| 15 | Zhongzheng District | 0.01 | 0 | 25.03247 | 121.51856 |

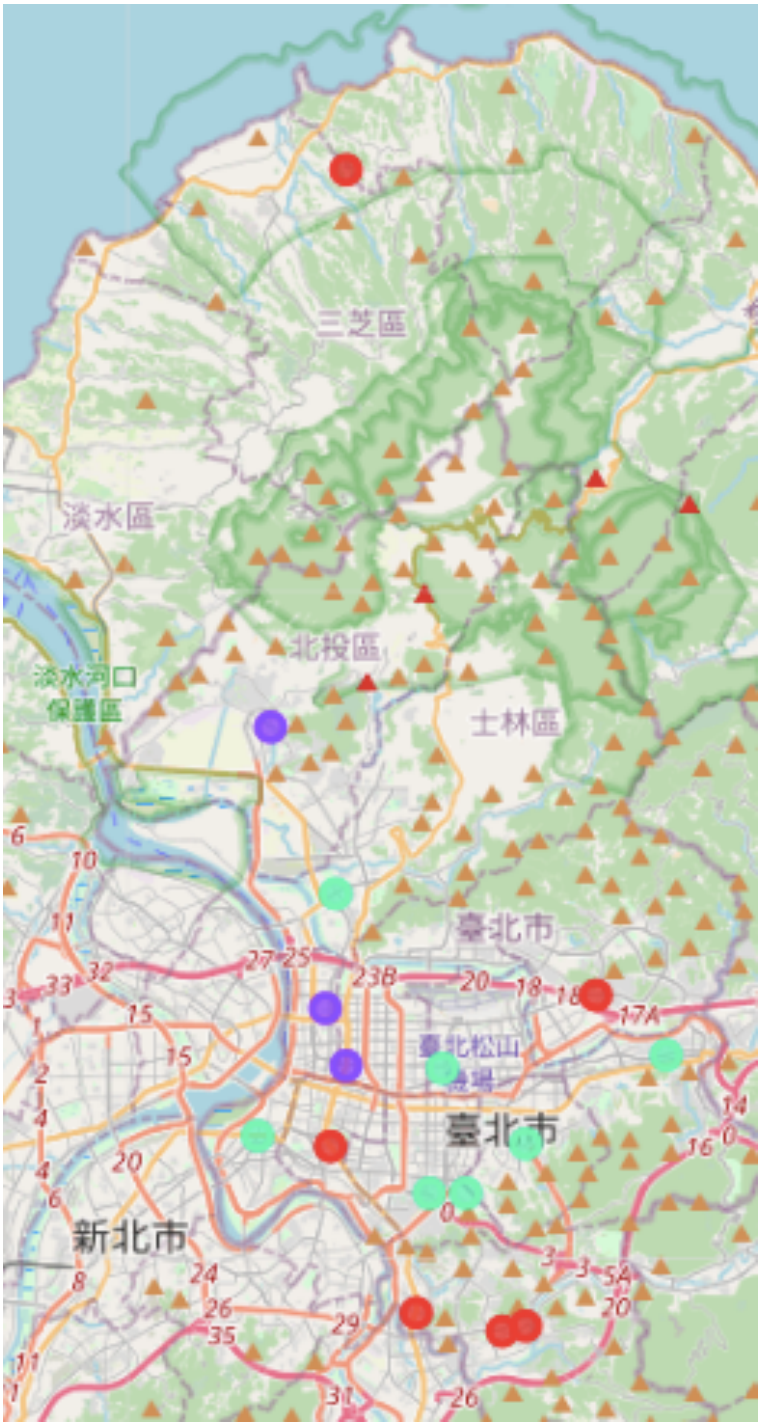
```
# Cluster 1
tpe_merged.loc[tpe_merged['Cluster Labels'] == 1]
```

| | Neighborhood | Hotel | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | Beitou District | 0.149254 | 1 | 25.13289 | 121.50253 |
| 2 | Datong District, Taipei | 0.100000 | 1 | 25.06589 | 121.51670 |
| 14 | Zhongshan District, Taipei | 0.110000 | 1 | 25.05229 | 121.52269 |

```
# Cluster 2
tpe_merged.loc[tpe_merged['Cluster Labels'] == 2]
```

| | Neighborhood | Hotel | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 1 | Daan District, Taipei City | 0.040000 | 2 | 25.02138 | 121.54434 |
| 4 | Guting District | 0.040000 | 2 | 25.02147 | 121.55399 |
| 7 | Nangang District, Taipei | 0.036585 | 2 | 25.05438 | 121.60673 |
| 9 | Shilin District | 0.030000 | 2 | 25.09313 | 121.51976 |
| 10 | Songshan District, Taipei | 0.050000 | 2 | 25.05165 | 121.54774 |
| 11 | Wanhua District | 0.060000 | 2 | 25.03535 | 121.49948 |
| 13 | Xinyi District, Taipei | 0.060000 | 2 | 25.03361 | 121.57002 |

4-1) As seen in the final examination of the created clusters, hotels are mostly located in Cluster 2 then Cluster 1 and barely in Cluster 0. This makes Cluster 0 a high potential location for new hotels if taking in only the consideration of having least competition.



4-2) Similarly, on this map visualization, we can see that the purple cluster (Cluster 1) is the smallest cluster while the other two green and red clusters are equal in amount. It is notable that there is one outlier for the red district that is very far north that shares the same mean value as the other red district down south.

## 5) Discussion

Upon the data analyzed, it is clear that Cluster 0 is the best locations to open up new hotel. These locations are concentrated not necessarily in the city center and are even generally on the side lines of Taipei and the neighboring New Taipei city.

There is a risk of opening new hotel businesses in these locations because they are not the most touristy locations and are mostly residential areas or business parks. This is understandable because this might be the top reason why there is the least existing hotel competition here. Therefore, the locations in Cluster 0 (red) can be very suitable for a well-established developer such as a big hotel chain seeking to expand.

However, developers that are trying to open one hotel and concentrate its effort on managing that one location can consider locations that are more in the city center from Cluster 1 (purple) such as Zhongshan District or the lower competition locations within Cluster 2 (red) such as Shilin District.

Further analyzation on the either the scale of the hotels or related venues such as metro stations or tourist attractions can help paint a better hotel market picture of Taipei.