Clustering the Taiwan Market

Determine the top common venues for each area in Taiwan

ABSTRACT

Explore top 3 common venues of area and give customers recommend

Yun Lin

Table of Content

Introduction

Taiwan were divided into four regions (tetrachotomy) scheme corresponds to the prefectures under Qing Dynasty rule. Thus, we are using the ‘present divisions’ of Taiwan for this project.

This project will delve deep into the neighborhoods of Taiwan and, using the Foursquare API, find venues in each neighborhood. After gathering all the required data, analysis will follow to help cluster the different neighborhoods and gain insight into which areas are perfect for opening a new business. In addition, it will help suggest the best businesses to open in those areas.

Business Problem

A client would like to open a new business; however, they do not know what kind of business they would like to open, let alone where to open it. In this project, I will determine the optimal areas in Taiwan to open a business, as well as the perfect business for that area.

Data

The first source will be Wikipedia. My project is about the regions of Taiwan. Next, using Google Maps' API, we will collect the approximate coordinates of each area. With the coordinates and area names collected, we will next be using the Foursquare API to collect venue information for each area, within a designated radius. The venue data collected from Foursquare will then be used to determine the top venues in each area.

Data Methodology

Once the venue information is gathered, the next step will be to cluster areas in Taiwan based on venues categories. This information will allow us to cluster customers in each area before moving on to identify areas within the clusters which are prime candidates for a new venue, as well as identifying which specific venues would be the most lucrative.

Collection

At the beginning of entire proejct, I began wutg collecting area names. Wikipedia is a good souce because they have area names in Taiwan. However, there are more complex case in my project where we see many cities show up in one region.

	Present divisions
0	Taipei, New Taipei, Keelung, Taoyuan, Hsinchu ...
1	Miaoli, Taichung, Changhua, Nantou
2	Yunlin, Chiayi City/County, Tainan
3	Kaohsiung, Pingtung, Penghu
4	Hualien, Taitung

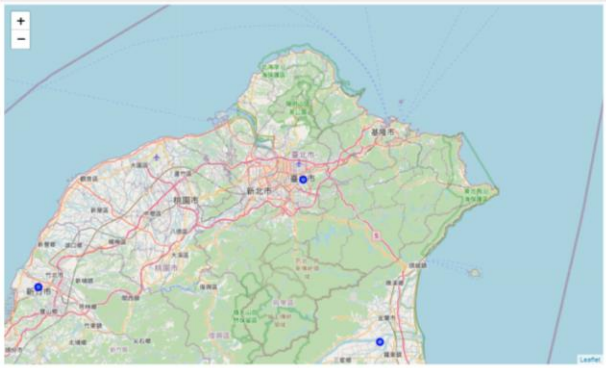
Then we collect all the coordinates for each area. For my case, I utilized the coordinates by Google Maps API. Then assigned they into the previous dataframe.

	Present divisions	Latitude	Longitude
0	Taipei, New Taipei, Keelung, Taoyuan, Hsinchu ...	24.813829	120.967480
1	Miaoli, Taichung, Changhua, Nantou	24.163165	120.674669
2	Yunlin, Chiayi City/County, Tainan	23.480075	120.449111
3	Kaohsiung, Pingtung, Penghu	22.627278	120.301435
4	Hualien, Taitung	23.991073	121.611195

Map

```
# add markers to map
for area, lat, lng in zip(df_tw['Present divisions'], df_tw['Latitude'], df_tw['Longitude']):
    label = '{}'.format(area)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_taiwan)
map_taiwan
```

Out[44]:



Total Venue at all areas in Taiwan

The figure below is showing total numbers of venue of each area in Taiwan, and we are seeing the area here is clearer because each area has been Clustered.

Area	Area Lat	Area Long	Venue	Venue Lat	Venue Long	Category
(Penghu), Kinmen, Matsu (Lienchiang)	8	8	8	8	8	8
(Yilan), Hualien, Taitung	46	46	46	46	46	46
Hualien, Taitung	60	60	60	60	60	60
Kaohsiung, Pingtung, (Penghu)	100	100	100	100	100	100
Kaohsiung, Pingtung, Penghu	100	100	100	100	100	100
Kinmen, Matsu (Lienchiang)	6	6	6	6	6	6
Miaoli, Taichung, Changhua, Nantou	100	100	100	100	100	100
Taichung, Changhua, Nantou	100	100	100	100	100	100
Taipei, New Taipei, Keelung, (Yilan)	100	100	100	100	100	100
Taipei, New Taipei, Keelung, Taoyuan, Hsinchu City/County, Yilan	95	95	95	95	95	95
Taoyuan, Hsinchu City/County, Miaoli	95	95	95	95	95	95
Yunlin, Chiayi City/County, Tainan	94	94	94	94	94	94

Unique Categories

The figure below is showing the total number of unique category of venues at Taiwan.

```
: print('There are {} uniques categories.'.format(len(taiwan_venues['Category'].unique())))
    There are 131 uniques categories.
```

K-Mean

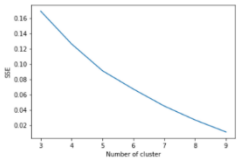
I use k=3 to calculate the SSE between each shop.

```
In [76]: sse = {}
         list_1 = []
         list_2 = []
         sil = pd.DataFrame()

         for k in range(3,10):
             kmeans = KMeans(n_clusters=k, max_iter=10000, random_state=0).fit(taiwan_grouped_clustering)
             taiwan_grouped_clustering["Clusters"] = kmeans.labels_
             #print(data["clusters"])
             sse[k] = kmeans.inertia_ # Inertia: Sum of distances of samples to their closest cluster center
             label = kmeans.labels_
             sil_coeff = silhouette_score(taiwan_grouped_clustering, label, metric='euclidean')
             list_1.append(k)
             list_2.append(sil_coeff)

         sil['k'] = list_1
         sil['sil_coeff'] = list_2
         highest = sil.sort_values(['sil_coeff'], ascending=False).head(1)
         print('The best k is {} with a silhouette score of {}'.format(highest['k'].values, highest['sil_coeff'].values))
         plt.figure()
         plt.plot(list(sse.keys()), list(sse.values()))
         plt.xlabel("Number of cluster")
         plt.ylabel("SSE")
         plt.show()

         The best k is [3] with a silhouette score of [0.68934177]
```




```
In [77]: kclusters = 4

         # run k-means clustering
         kmeans = KMeans(n_clusters=kclusters, random_state=0, max_iter=10000).fit(taiwan_grouped_clustering)

In [78]: kmeans.labels_[0:10]

Out[78]: array([2, 1, 3, 0, 0, 2, 2, 1, 2, 0])
```

Common Venue

In the end, we are looking for top 3 common venue at each area in Taiwan. By that, the business will have better version to know what should they do.

	Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	(Penghu), Kinmen, Matsu (Lienchiang)	Hotel	Department Store	Tourist Information Center
1	(Yilan), Hualien, Taitung	Bakery	Asian Restaurant	Park
2	Hualien, Taitung	Hotel	Chinese Restaurant	Hostel
3	Kaohsiung, Pingtung, (Penghu)	Coffee Shop	Hotel	Taiwanese Restaurant
4	Kaohsiung, Pingtung, Penghu	Coffee Shop	Hotel	Taiwanese Restaurant
5	Kinmen, Matsu (Lienchiang)	Airport	Noodle House	Historic Site
6	Miaoli, Taichung, Changhua, Nantou	Hotel	Chinese Restaurant	Café
7	Taichung, Changhua, Nantou	Hotel	Chinese Restaurant	Café
8	Taipei, New Taipei, Keelung, (Yilan)	Hotel	Café	Bakery
9	Taipei, New Taipei, Keelung, Taoyuan, Hsinchu ...	Chinese Restaurant	Convenience Store	Coffee Shop
10	Taoyuan, Hsinchu City/County, Miaoli	Chinese Restaurant	Convenience Store	Coffee Shop
11	Yunlin, Chiayi City/County, Tainan	Taiwanese Restaurant	Hotel	Park

Conclusion

To conclude, when are company is decided to start a business in Taiwan. They should consider more about their location and venue that surrounding them, because if they start a restaurant and there is similar one nearby then that would have an inevitable competition.

Bibliography

https://en.wikipedia.org/wiki/Regions_of_Taiwan