Capstone Project - The Battle of Neighborhoods - Taichung version

This project is a part of IBM data science; you will find in this post an overview of my final capstone project.

In this assignment, I will go through the problem description (**Introduction/Business Problem**), data set preparation and final to analysis and overview these data step by step. Detailed code is given in Github and link can be found at the end of the post.

1 Introduction/Business Problem

1.1 Background

Taiwan is a famous nation in the world. It was very small (**36,197** km^2) but with more than 23,600,000 population. Taiwan with the high metropolitan GDP in the world and it belonged a developed country.

There was COVID-19 spread in the world but it seems not impact to Taiwan. Here with the best epidemic prevention concept and medical technology so the infection rate and mortality rate is the lowest in the world.

With estimated 11,840,000 overseas visitors every year. It's famous with delicious foods and beautiful landscapes. Many travelers needed to know some leisure and entertainment places when they first visit to Taiwan.
These characteristics also attract many immigrations come to Taiwan.

Taichung is a second big city in Taiwan, its' urban scale was smaller than Taipei only. Taichung retained the advantage of Taipei like: convenient transportation/convenient life/good job opportunity/medical system…etc. Many immigrations will choose settle down in Taichung.

1.2 Business Problem

Taichung with 29 different districts, some districts are commercial and some are convenient and some are leisure. These districts are so difference!!!

In this article will help immigrations to understand and overview the Taichung city. Let them to choose the district which they want to live or travel.

2 Data and Data Preparation

2.1 Data Clean and Preprocess:

2.1.1 Define CSV Process Class

In [4]:
```
class CSVprocess(object):

    def __init__(self):
        #self.address = "C:\\Users\\Brian\\Desktop\\ibm_coursera\\Coursera_Capstone\week4\\"
        self.address = "C:\\Users\\brian\\Desktop\\Coursera_Capstone\\week4\\"

    def load_from_csv(self, filename):
        df = pd.read_csv(self.address + filename, encoding='UTF-8')
        return df

    def write_to_csv(self, filename, dictionary):
        colume_list = []

        for key, value in dictionary.items():
            colume_list.append(key)

        df = pd.DataFrame(dictionary, columns=colume_list)
        df.to_csv(self.address + filename, index=False, encoding="utf_8_sig")
```

2.1.2 Define JSON Process Class

In [5]:
```
class JsonProcess(object):

    def __init__(self):
        #self.address = "C:\\Users\\Brian\\Desktop\\ibm_coursera\\Coursera_Capstone\week4\\"
        self.address = "C:\\Users\\brian\\Desktop\\Coursera_Capstone\\week4\\"

    def load_from_json(self, filename):
        json_file = open(self.address + filename, "r", encoding="utf-8")
        j = json_file.read()
        json_file.close()
        dict_data = json.loads(j)
        return dict_data

    def dictionary_to_json(self, dictionary):
        json_data = json.dumps(dictionary)
        return json_data
```

2.1.3 Define Get Data from Web Class

In [6]:
```
class GetDateFromWeb(object):

    def get_taichung_info(self):
        response = requests.get('https://zh.wikipedia.org/wiki/臺中市#人口')
        content = response.content
        df = pd.read_html(content, encoding='utf-8')
        taichung_info_df = df[11]

        return taichung_info_df
```

2.1.4 Define DataFrame Process Class

In [7]:
```
class DataProcess(object):

    def __init__(self):
        pass

    def merge_dataframe(self, source_df, target_df, field_name):
        df = pd.merge(source_df, target_df, on=field_name)
        return df

    def drop_rename_dataframe(self, df, row_drop_list, colume_drop_list, rename_list):
        df = df.drop(row_drop_list)
        df = df.drop(columns=colume_drop_list)
        for item in rename_list:
            for key, value in item.items():
                df = df.rename(columns={key: value})

        return df
```

2.2 Taichung wiki

Data source: https://zh.wikipedia.org/wiki/臺中市#人口

Description: This data set included 29 districts in Taichung city. There are some fields we need like: Chinese Name, Area, Density and Postal Code.

```
In [61]: row_drop_list = [29]
colume_drop_list = ["下轄里數", "下轄鄰數", "人口消長", "地理分區"]
rename_list = [{"區名": "Chinese Name"},{"面積（km²）": "Area"},{"人口數": "Population"},{"人口密度（人/km²）": "Density"},{"郵遞區號": "Postal

df = getDataFromWeb.get_taichung_info()
df = dataProcess.drop_rename_dataframe(df, row_drop_list, colume_drop_list, rename_list)
df
```

Out[61]:

	Chinese Name	Area	Population	Density	Postal Code
0	中區	0.8803	18173	20644	400.0
1	東區	9.2855	75769	8160	401.0
2	南區	6.8101	126585	18588	402.0
3	西區	5.7042	114618	20094	403.0
4	北區	6.9376	146676	21142	404.0
5	北屯區	62.7034	286208	4564	406.0
6	西屯區	39.8467	230978	5797	407.0
7	南屯區	31.2578	175439	5613	408.0
8	太平區	120.7473	194835	1614	411.0
9	大里區	28.8759	212957	7375	412.0
10	霧峰區	98.0779	65144	664	413.0
11	烏日區	43.4032	76249	1757	414.0
12	豐原區	41.1845	166199	4035	420.0
13	后里區	58.9439	54374	922	421.0
14	石岡區	18.2105	14659	805	422.0
15	東勢區	117.4065	49250	419	423.0
16	和平區	1037.8192	10815	10	424.0
17	新社區	68.8874	24108	350	426.0
18	潭子區	25.8497	109274	4227	427.0
19	大雅區	32.4109	95703	2953	428.0
20	神岡區	35.0445	65550	1870	429.0
21	大肚區	37.0024	56759	1534	432.0
22	沙鹿區	40.4604	95135	2351	433.0
23	龍井區	38.0377	77998	2051	434.0
24	梧棲區	16.6049	59199	3565	435.0
25	清水區	64.1709	87863	1369	436.0
26	大甲區	58.5192	76116	1301	437.0
27	外埔區	42.4098	31989	754	438.0
28	大安區	27.4045	18925	691	439.0

2.3 English - Chinese glossary of districts in Taiwan

Data source: http://gn.moi.gov.tw/geonames/Translation/Translation.aspx

Description: Government help to translate the city or districts name between English and Chinese

```
In [12]: taichung_english_name_df = csvProcess.load_from_csv("Taichung_English_Name.csv")
         print(taichung_english_name_df.shape)
         taichung_english_name_df.head(30)
```

```
(29, 2)
```

Out[12]:

	Chinese Name	Name
0	南屯區	Nantun District
1	西屯區	Xitun District
2	北屯區	Beitun District
3	西區	West District
4	北區	North District
5	霧峰區	Wufeng District
6	豐原區	Fengyuan District
7	龍井區	Longjing District
8	潭子區	Tanzi District
9	新社區	Xinshe District
10	清水區	Qingshui District
11	梧棲區	Wuqi District
12	神岡區	Shengang District
13	烏日區	Wuri District
14	東勢區	Dongshi District
15	和平區	Heping District
15	和平區	Heping District
16	沙鹿區	Shalu District
17	后里區	Houli District
18	石岡區	Shigang District
19	外埔區	Waipu District
20	太平區	Taiping District
21	大雅區	Daya District
22	大里區	Dali District
23	大肚區	Dadu District
24	大安區	Da'an District
25	大甲區	Dajia District
26	中區	Central District
27	東區	East District
28	南區	South District

2.4 Latitude and longitude in Taiwan

Data source:
https://www.astrocode.net/%E5%8F%B0%E7%81%A3%E5%90%84%E7%B8%A3%E5%B8%82%E5%9C%B0%E5%8D%80%E7%B6%93%E7%B7%AF%E5%BA%A6/

Description: This is a XML file and we can transfer to JSON file. The file included all districts of latitude and longitude in Taiwan.

```
In [13]: taiwan_geo_dict = jsonProcess.load_from_json("Taiwan_Geospatial_Coordinates.json")
         taiwan_geo_dict
```


```
Out[13]: {'dataroot': {'@_noNamespaceSchemaLocation': '1050429_行政區經緯度(toPost).xsd',
           '@_generated': '2016-08-12T12:06:09',
           '_x0031_050429_行政區經緯度_x0028_toPost_x0029_': [{'行政區名': '臺北市中正區',
             '_x0033_碼郵遞區號': '100',
             '中心點經度': '121.5198839',
             '中心點緯度': '25.03240487',
             'TGOS_URL': 'http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_MetaData_View.aspx?MID=9C715A5CD330360D355AE105F908B29E&amp;SHOW_BACK_BUTTON=false'},
            {'行政區名': '臺北市大同區',
             '_x0033_碼郵遞區號': '103',
             '中心點經度': '121.5130417',
             '中心點緯度': '25.06342433',
             'TGOS_URL': 'http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_MetaData_View.aspx?MID=9C715A5CD330360D355AE105F908B29E&amp;SHOW_BACK_BUTTON=false'},
            {'行政區名': '臺北市中山區',
             '_x0033_碼郵遞區號': '104',
             '中心點經度': '121.5381597',
             '中心點緯度': '25.06969917',
             'TGOS_URL': 'http://tgos.nat.gov.tw/tgos/Web/MetaData/TGOS_MetaData_View.aspx?MID=9C715A5CD330360D355AE105F908B29E&amp;SHOW_BACK_BUTTO
```


2.5 Venues in each neighborhood of Taichung City

Data source: Foursquare APIs

Description: We will get all the venues in each neighborhood by using this API and filter these venues to get restaurants and other data.

3 Visualization and Data Exploration

3.1 Get Merged Dataframe

By data pre-process and data clean, we will store the data into Taichung_merge_Info.csv

Load dataframe from Taichung_merge_Info.csv

```
In [19]: address = csvProcess.address + "Taichung_Merge_Info.csv"
         df = pd.read_csv(address)
         print(df.shape)
         df.head()
```

```
(29, 8)
```

Out[19]:

	Chinese Name	Area	Population	Density	Postal Code	Name	Latitude	Longitude
0	中區	0.8803	18173.0	20644.0	400.0	Central District	24.141686	120.680598
1	東區	9.2855	75769.0	8160.0	401.0	East District	24.137332	120.697086
2	南區	6.8101	126585.0	18588.0	402.0	South District	24.121141	120.664618
3	西區	5.7042	114618.0	20094.0	403.0	West District	24.143911	120.664758
4	北區	6.9376	146676.0	21142.0	404.0	North District	24.158640	120.680952

3.2 Explore the neighborhoods in Taichung (By folium map)

Use geopy library to get the latitude and longitude values of Taichung and create a map of Taichung with neighborhoods superimposed on top.

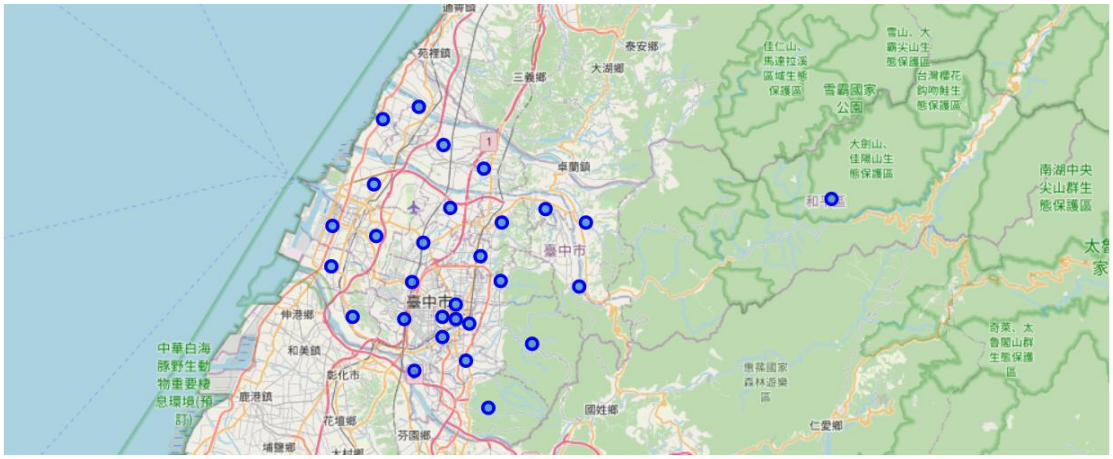

Fig. 1. All districts location in Taichung

3.3 Explore the neighborhoods in Taichung (By Foursquare API)

Use Foursquare API and get the top 100 venues that are in Taichung within a radius of 1000 meters. We found there are total 288 venues in this table.

In [33]:
```
print(taichung_venues.shape)
taichung_venues.head(50)
#print ('{} unique categories in Taichung'.format(taichung_venues['Venue Category'].value_counts().shape[0]))
#print (taichung_venues['Venue Category'].value_counts()[0:50])
```

(288, 7)

Out[33]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central District	24.141686	120.680598	台中市第四信用合作社	24.138987	120.681956	Ice Cream Shop
1	Central District	24.141686	120.680598	台中市公有第二零售市場	24.142094	120.679246	Market
2	Central District	24.141686	120.680598	范記金之園草燒飯	24.141308	120.682276	Chinese Restaurant
3	Central District	24.141686	120.680598	洪瑞珍正宗招牌三明治	24.140132	120.680565	Sandwich Place
4	Central District	24.141686	120.680598	阿斗伯冷凍芋頭	24.145315	120.677813	Dessert Shop
5	Central District	24.141686	120.680598	瑪露蓮	24.141186	120.681042	Dessert Shop
6	Central District	24.141686	120.680598	宮原眼科 摘星樓	24.138062	120.683413	Vegetarian / Vegan Restaurant
7	Central District	24.141686	120.680598	Miyahara (宮原眼科)	24.138036	120.683466	Ice Cream Shop
8	Central District	24.141686	120.680598	沁園春	24.139649	120.682850	Chinese Restaurant
9	Central District	24.141686	120.680598	Taichung Park (台中公園)	24.142873	120.684247	Park
10	Central District	24.141686	120.680598	紅點文旅 RedDot	24.143205	120.675442	Hotel
11	Central District	24.141686	120.680598	阪神本舖長崎蛋糕	24.144448	120.675576	Cupcake Shop
12	Central District	24.141686	120.680598	翁記泡沫廣場	24.146167	120.683408	Snack Place
13	Central District	24.141686	120.680598	寶島53行館 53 Hotel	24.137132	120.683358	Hotel
14	Central District	24.141686	120.680598	老賴紅茶	24.142012	120.679195	Bubble Tea Shop

3.4 Visualization the Taichung City dataframe information

3.4.1 Rank of Venue Category in Taichung City

In [70]:
```
taichung_venues_category_top10 = taichung_venues['Venue Category'].value_counts()[0:10].to_frame(name='frequency')
taichung_venues_category_top10=taichung_venues_category_top10.reset_index()
#Taichung Venues Top10
taichung_venues_category_top10.rename(index=str, columns={"index": "Venue Category", "frequency": "Frequency"}, inplace=True)
taichung_venues_category_top10
```

Out[70]:

	Venue Category	Frequency
0	Chinese Restaurant	23
1	Hotel	17
2	Café	16
3	Coffee Shop	13
4	Convenience Store	13
5	Dessert Shop	12
6	Tea Room	11
7	Asian Restaurant	10
8	Taiwanese Restaurant	9
9	Japanese Restaurant	9

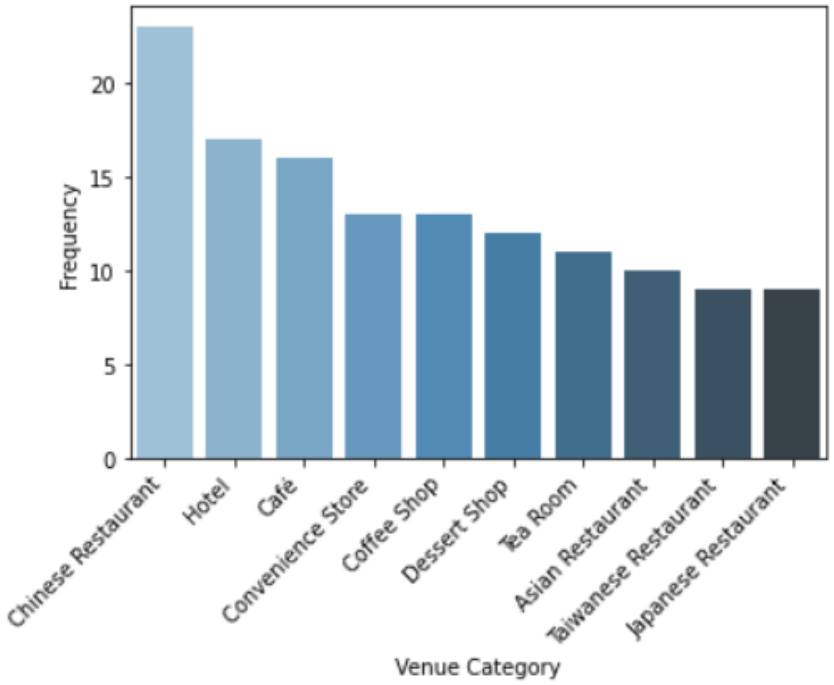

Fig. 2. Top 10 Venue Category in Taichung City

3.4.2 Number of every type of restaurant in Taichung City

In [75]:
```
# create a dataframe of top 10 categories
dist_venues_top10 = taichung_venues_only_restaurant['Venue Category'].value_counts()[0:10].to_frame(name='frequency')
dist_venues_top10=dist_venues_top10.reset_index()

#Taichung Venues Top10
dist_venues_top10.rename(index=str, columns={"index": "Venue Category", "frequency": "Frequency"}, inplace=True)
dist_venues_top10
```

Out[75]:

	Venue Category	Frequency
0	Chinese Restaurant	23
1	Asian Restaurant	10
2	Japanese Restaurant	9
3	Taiwanese Restaurant	9
4	Fast Food Restaurant	7
5	Italian Restaurant	4
6	Dumpling Restaurant	3
7	Restaurant	3
8	Hotpot Restaurant	2
9	American Restaurant	2

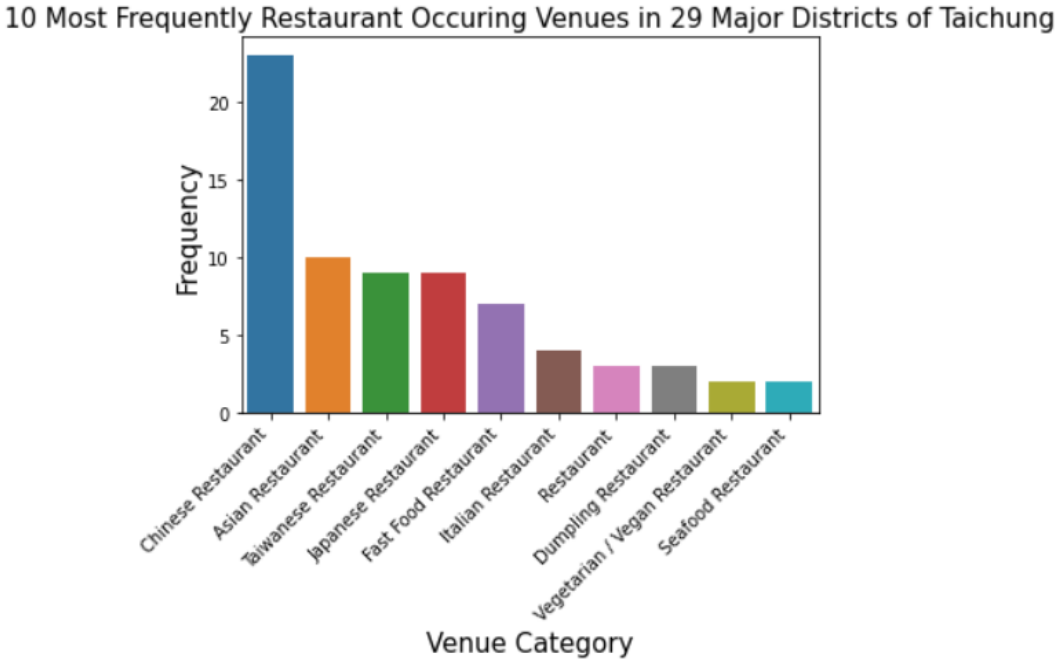

Fig. 3. Top 10 restaurant in Taichung City

3.4.3 Number of venues in each Neighborhood

```
#Taichung Venues Top10
taichung_neighborhoods_top10.rename(index=str, columns={"index": "Neighborhood", "frequency": "Frequency"}, inplace=True)
taichung_neighborhoods_top10
```

Out[77]:

	Neighborhood	Frequency
0	West District	62
1	Central District	62
2	North District	31
3	Beitun District	16
4	East District	16
5	Xitun District	13
6	Houli District	11
7	Tanzi District	9
8	Fengyuan District	8
9	Dali District	7

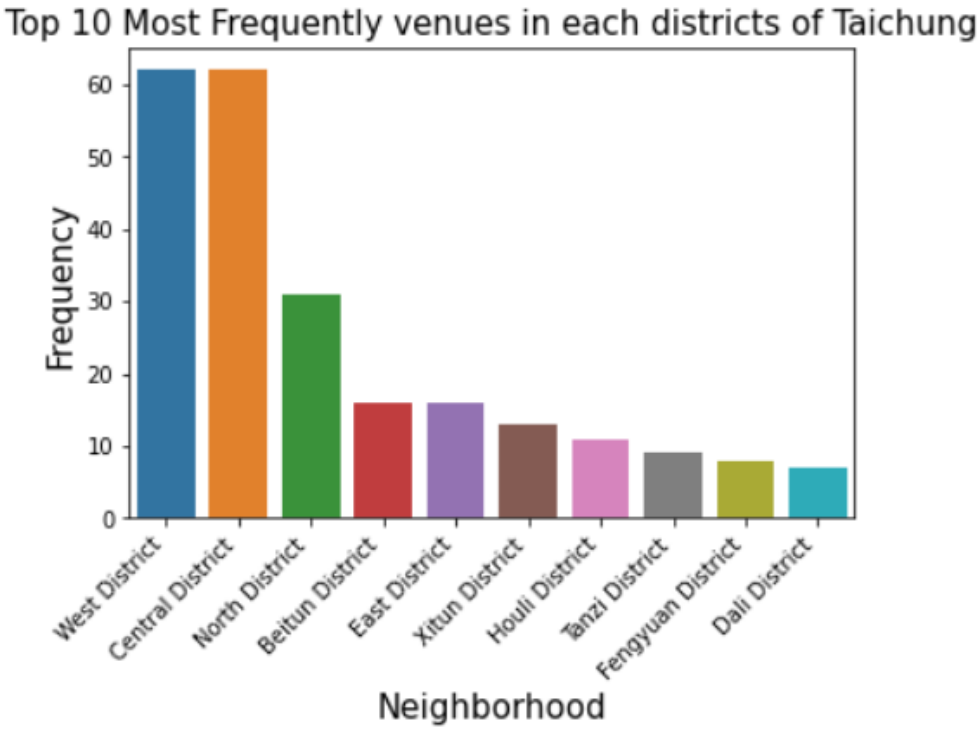

Fig. 4. Top 10 most frequently venues in each districts of Taichung City

The Folium Map correspond with above bar chart. The map displayed 29 districts in Taichung City. The red points present top 10 districts and the blue points present other 19 districts.

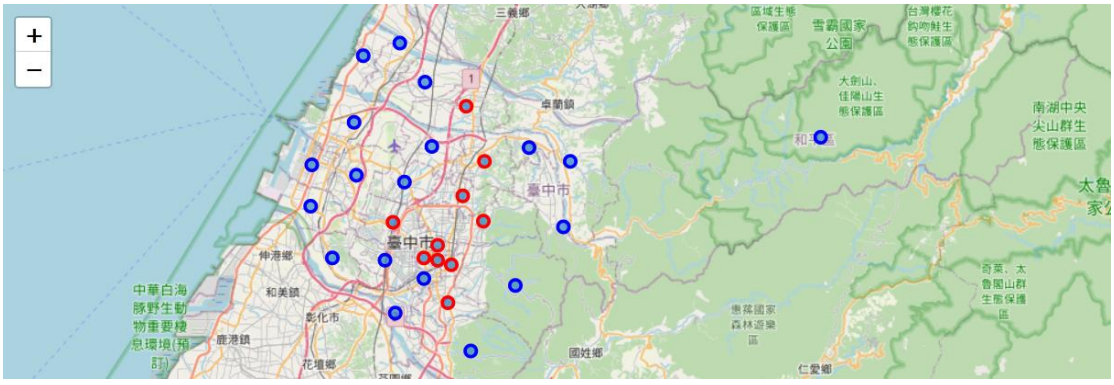

Fig. 5. Top 10 most frequently venues in each districts locations

3.4.4 Layer of district density in Taichung city

There are three kind colors in the map. The red point means the density larger than 4000, the blue point means the density is from 1000 to 4000 and the green point present the density is smaller than 1000.

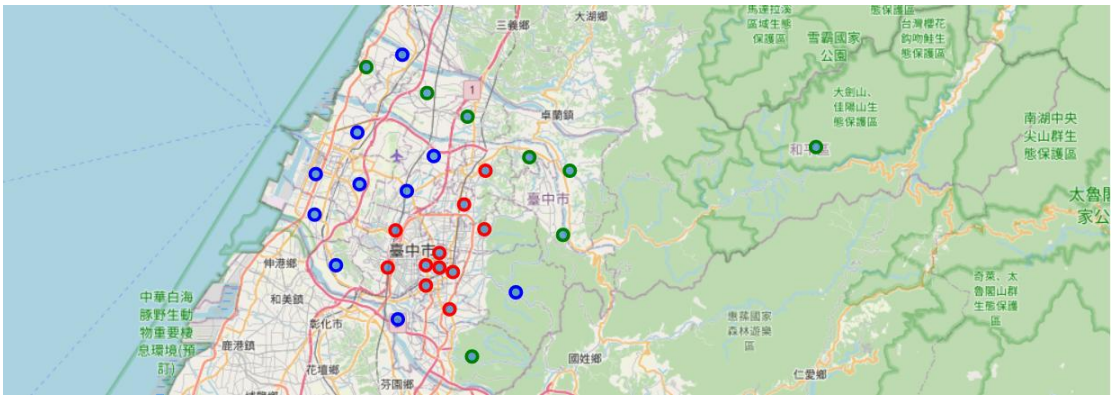

Fig. 6. Layer of district density locations in Taichung city

3.5 Results & Discussion

3.5.1 Rank of Venue Category in Taichung City

The bar chart (Fig. 2.) is the top 5 venue category Venue Category in Taichung City, there are restaurant/hotel/cafe/convenience store/dessert ship.

That indicate Taichung is a vary convenience city, you can eat any food and buy anything everywhere. Travers can eat paradise in Taichung.

3.5.2 Number of every type of restaurant in Taichung City

The bar chart (Fig. 3.) is number of every type of restaurant in Taichung city, the top 5 restaurants are Chinese restaurant, Asian restaurant, Taiwanese restaurant, Japanese restaurant, Fast Food restaurant

We can get some analysis from this bar chart.

The First: Chinese restaurant/Asian restaurant/Taiwanese restaurant are top 3, that percent the food culture still prefer traditional Chinese favor and Taiwanese local favor in Taichung.

The Second: Japanese restaurant /Fast Food restaurant are fourth and fifth, that percent the food culture was impact by Japan and Western culture (USA) in recent twenty years.

That means there are more and more food favors in Taichung. If someone want to create restaurant and I will suggest Italian or French cuisine or Indian food. These favors are rare in Taichung city and these restaurants will be popular if their meals vary are delicious or special.

3.5.3 Number of venues in each Neighborhood

The bar chart (Fig. 4) is number of venues in each neighborhood in Taichung city. There are 29 districts in Taichung and here we list top 10 districts West District, Central District, North District, Beitun District, East District, Xitun District, Houli District, Tanzi District, Fengyuan District and Dali District

The map (Fig. 5.) is mapping with bar chart. The red points are top 10 districts with venues number and the blue points are other 19 districts

We can get some analysis from this bar chart. Some districts are more urbanization and prosperity and some place are still rural and original.

We also get other analysis from the map. The top 10 are concentrated on the central of Taichung and others are located on mountain or coastal.

3.5.4 Layer of district density in Taichung city

This map is corresponding with the map of “Number of venues in each Neighborhood”.

This map (Fig. 6.) is similar with previous map (Fig. 5.). The red points located in the same districts in these two maps. That means the number of venues and density with high relationship.

The property of red point is convenience and business development but with large population and noisy in the night.

The property of green point is quiet and with natural environment and small population but is not convenience.

The property of blue point is between red point and green point. With leisure and moderate population.

This analysis can recommend immigrations or travers which districts they want to journey or settle down.

4 Analyze Each Neighborhood and Clustering the

neighborhoods

4.1 Analyze Each Neighborhood

4.1.1 One hot encoding

In [47]:
```
# one hot encoding
taichung_onehot = pd.get_dummies(taichung_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
taichung_onehot['Neighborhood'] = taichung_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [taichung_onehot.columns[-1]] + list(taichung_onehot.columns[:-1])
taichung_onehot = taichung_onehot[fixed_columns]

taichung_onehot.head()
```

Out[47]:

	Neighborhood	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Garage	Auto Workshop	BBQ Joint	Bakery	...	Szechuan Restaurant	Taiwanese Restaurant	Tea Room	Trail	Train Station	Use Bookstor
0	Central District	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	Central District	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Central District	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	Central District	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	Central District	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

5 rows × 97 columns

4.1.2 Analyze top 10 venues

In [88]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = taichung_grouped['Neighborhood']

for ind in np.arange(taichung_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(taichung_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[88]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beitun District	Trail	Restaurant	Snack Place	Gift Shop	Dessert Shop	Park	Pizza Place	Convenience Store	Chinese Restaurant	Café
1	Central District	Hotel	Convenience Store	Chinese Restaurant	Dessert Shop	Coffee Shop	Snack Place	Ice Cream Shop	Tea Room	Asian Restaurant	Japanese Restaurant
2	Da'an District	Smoke Shop	Yunnan Restaurant	Cupcake Shop	Dessert Shop	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant	Fast Food Restaurant
3	Dadu District	Gourmet Shop	Market	Supermarket	Yunnan Restaurant	Food Truck	Dessert Shop	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop

4.2 Clustering the neighborhood

4.2.1 k-means

Try to cluster these 29 districts based on the venue categories and use K-Means clustering. We expect the similarities of venue categories will be clustered.

In [53]:
```
# set number of clusters
kclusters = 5
taichung_grouped_clustering = taichung_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[53]: `array([0, 0, 2, 0, 0, 0, 4, 0, 0, 1])`

Create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

In [54]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

taichung_merged = df

taichung_merged.rename(columns={'Name':'Neighborhood'}, inplace=True)

# merge manhattan_grouped with manhattan_data to add latitude/longitude for each neighborhood
taichung_merged = taichung_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

taichung_merged = taichung_merged.dropna()

taichung_merged.head() # check the last columns!
```

Out[54]:

	Chinese Name	Area	Population	Density	Postal Code	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Commo Venu
0	中區	0.8803	18173.0	20644.0	400.0	Central District	24.141686	120.680598	0.0	Hotel	Convenience Store	Chinese Restaurant	Dessert Shop	Coffee Shop	Snac Plac

4.2.2 Visualize the resulting clusters

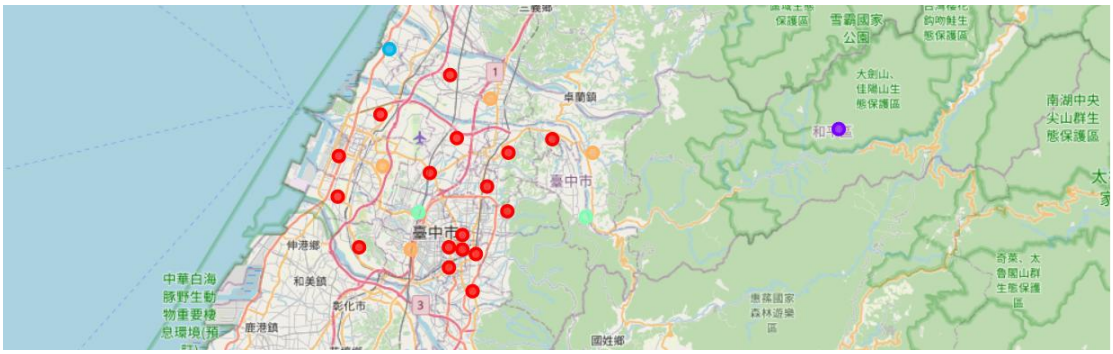

Fig. 7. The districts locations of the clustering result

4.3 Examine Clusters

Cluster 1

In [56]:
```
taichung_merged.loc[taichung_merged['Cluster Labels'] == 0, taichung_merged.columns[[1] + list(range(5, taichung_merged.shape[1]))]]
```

Out[56]:

	Area	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	
0	0.8803	Central District	24.141686	120.680598	0.0	Hotel	Convenience Store	Chinese Restaurant	Dessert Shop	Coffee Shop	Snack Place	Ice Cream Shop	Tea Room	F
1	9.2855	East District	24.137332	120.697086	0.0	Market	Chinese Restaurant	Shopping Mall	Bookstore	Coffee Shop	Museum	Shopping Plaza	Movie Theater	
2	6.8101	South District	24.121141	120.664618	0.0	Coffee Shop	Train Station	Tea Room	Skating Rink	Burger Joint	Fast Food Restaurant	Food Court	Dessert Shop	F
3	5.7042	West District	24.143911	120.664758	0.0	Café	Dessert Shop	Tea Room	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Plaza	Coffee Shop	F
4	6.9376	North District	24.158640	120.680952	0.0	Coffee Shop	Hotpot Restaurant	Hotel	Japanese Restaurant	Diner	Taiwanese Restaurant	Chinese Restaurant	Café	
5	62.7034	Beitun District	24.184003	120.736232	0.0	Trail	Restaurant	Snack Place	Gift Shop	Dessert Shop	Park	Pizza Place	Convenience Store	F
9	28.8759	Dali District	24.095757	120.692626	0.0	Food Truck	Convenience Store	Art Museum	Japanese Restaurant	Hakka Restaurant	Chinese Restaurant	Fast Food Restaurant	Yunnan Restaurant	
12	41.1845	Fengyuan District	24.249903	120.737572	0.0	Japanese Restaurant	Italian Restaurant	Taiwanese Restaurant	Supermarket	Fried Chicken Joint	Mountain	Bubble Tea Shop	Yunnan Restaurant	
14	18.2105	Shigang District	24.264933	120.790382	0.0	Dessert Shop	Park	Café	Garden	Taiwanese Restaurant	Noodle House	Food & Drink Shop	Dim Sum Restaurant	
18	25.8497	Tanzi District	24.211711	120.710997	0.0	Convenience Store	Historic Site	Japanese Restaurant	Auto Workshop	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Bus Station	F

19	32.4109	Daya District	24.227042	120.641182	0.0	Hotel	Asian Restaurant	Chinese Restaurant	Café	Fast Food Restaurant	Dessert Shop	Food Truck	Diner
20	35.0445	Shengang District	24.265680	120.673332	0.0	Convenience Store	Auto Garage	Tea Room	Food Truck	Dessert Shop	Dim Sum Restaurant	Diner	Dive Bar
21	37.0024	Dadu District	24.144675	120.554324	0.0	Gourmet Shop	Market	Supermarket	Yunnan Restaurant	Food Truck	Dessert Shop	Dim Sum Restaurant	Diner
23	38.0377	Longjing District	24.200629	120.528373	0.0	Asian Restaurant	Trail	BBQ Joint	Yunnan Restaurant	French Restaurant	Dim Sum Restaurant	Diner	Dive Bar
24	16.6049	Wuqi District	24.245524	120.530126	0.0	Seafood Restaurant	BBQ Joint	Szechuan Restaurant	Bus Stop	Supermarket	Food & Drink Shop	Yunnan Restaurant	Food Court
25	64.1709	Qingshui District	24.292057	120.580909	0.0	Dessert Shop	Bookstore	French Restaurant	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant
27	42.4098	Waipu District	24.335511	120.665064	0.0	Dive Bar	Food Truck	Yunnan Restaurant	Water Park	Dessert Shop	Dim Sum Restaurant	Diner	Donut Shop

Cluster 2

```
In [57]: taichung_merged.loc[taichung_merged['Cluster Labels'] == 1, taichung_merged.columns[[1] + list(range(5, taichung_merged.shape[1]))]]
```

Out[57]:

	Area	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	Con V
16	1037.8192	Heping District	24.276203	121.140185	1.0	Water Park	Yunnan Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant	Fast Food Restaurant	F

Cluster 3

```
In [58]: taichung_merged.loc[taichung_merged['Cluster Labels'] == 2, taichung_merged.columns[[1] + list(range(5, taichung_merged.shape[1]))]]
```

Out[58]:

	Area	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th M Comm Ven
28	27.4045	Da'an District	24.365096	120.591441	2.0	Smoke Shop	Yunnan Restaurant	Cupcake Shop	Dessert Shop	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant	Fast Fo Restaur

Cluster 4

```
In [59]: taichung_merged.loc[taichung_merged['Cluster Labels'] == 3, taichung_merged.columns[[1] + list(range(5, taichung_merged.shape[1]))]]
```

Out[59]:

	Area	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
6	39.8467	Xitun District	24.183089	120.627013	3.0	Hotel	Taiwanese Restaurant	Gym / Fitness Center	Japanese Restaurant	Furniture / Home Store	Fast Food Restaurant	Cantonese Restaurant	Yunnan Restaurant	Tea Room
17	68.8874	Xinshe District	24.177693	120.831323	3.0	Hotel	Café	Department Store	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant	Fast Food Restaurant

Cluster 5

```
In [60]: taichung_merged.loc[taichung_merged['Cluster Labels'] == 4, taichung_merged.columns[[1] + list(range(5, taichung_merged.shape[1]))]]
```

Out[60]:

	Area	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
7	31.2578	Nantun District	24.141655	120.617738	4.0	Vegetarian / Vegan Restaurant	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Shopping Mall	Yunnan Restaurant	Food & Drink Shop	Dessert Shop	Dim Sum Restaurant
13	58.9439	Houli District	24.309625	120.714613	4.0	Chinese Restaurant	Restaurant	Asian Restaurant	Jazz Club	Juice Bar	Dumpling Restaurant	Breakfast Spot	Buffet	Yunnan Restaurant
15	117.4065	Dongshi District	24.249526	120.840140	4.0	Chinese Restaurant	Shoe Store	Taiwanese Restaurant	Dumpling Restaurant	Mountain	Yunnan Restaurant	Food Court	Dessert Shop	Dim Sum Restaurant
22	40.4604	Shalu District	24.234252	120.583863	4.0	Breakfast Spot	Chinese Restaurant	French Restaurant	Dim Sum Restaurant	Diner	Dive Bar	Donut Shop	Dumpling Restaurant	Fast Food Restaurant

4.4 Results & Discussion

Clusters 1 have the greatest number of neighborhoods, cluster 2 and 3 has only one, Cluster 4 has two and Cluster 5 has four.

Cluster 1 is a group of neighborhoods that has the commonly known for good Hotel, Convenience Store, Café and Coffee Shop

Cluster 2 recommend Water Park

Cluster 3 recommend Smoke Shop
Cluster 4 is main Hotel
Cluster 5 are almost Chinese Restaurant

We can understand the similarity and difference on these districts from the map and clustering experiment in first step.

People could base on their favorite or consideration to filter the special districts which they can settle down or visit in Taichung city.

5 Conclusion

We used the previous skills and technology like: Data collection, Date pre-process, Data clean, web crawler, Foursquare API, Supervisor machine learning, Unsupervised learning, Folium Map in this final project.

By these skills and knowledge, we can analysis data and recommend some suggestions to solve problem or more understand business mode. This project let me learned a lot of machine learning technology and understand what is data science and how to do and complete it.

In this final project, we can add house price and the job opportunity in each district that can raise the data integrity. Use the more detail data, we can get clearly analysis report and improved the recommendation to tell immigrations and travers how to choose the districts more clearly.

Link to Github (no map):
https://github.com/prophetstorbrianchen/Coursera_Capstone/blob/main/week5/Capstone%20Project%20-%20The%20Battle%20of%20Neighborhoods%20-%20Taichung%20vision.ipynb

Link to Github (with map):
https://nbviewer.jupyter.org/github/prophetstorbrianchen/Coursera_Capstone/blob/main/week5/Capstone%20Project%20-%20The%20Battle%20of%20Neighborhoods%20-%20Taichung%20vision.ipynb