Housing Sales Prices & Venues Data Analysis of Taipei City

A. Introduction

A.1. Description & Discussion of the Background

Taipei is the 40th most-populous urban area in the world—roughly one-third of Taiwanese citizens live in the metro district. Taipei city is home to an estimated population of **2,646,204** (2019). Taipei City is divided up into 12 administrative districts. I have been living in Taipei city for 10 years and had good experience there, then I decided to use Taipei City in my project. Taipei City is an enclave of the municipality of New Taipei City that limits cover an area of **271.7997** square kilometers [1].

Taipei is a densely populated urban areas which continued to increase from year to year as well as the surrounding cities. Taipei is one of the world's most expensive cities and crowded which make it harder for investors to do business around the city, unaffordability remains a serious issue. Most investors would prefer to have access to relevant information to invest in their preferable district at a lower real estate cost and the type of business that is more popular in this area. Obviously, most people are looking for lower price real estate value and less dense area as well. It is difficult for investor to get this information in one place.

All these problems give an opportunity to create a map and information chart with real estate index marked in Taipei city and each district depending of the venue density.

A.2. Data Description

To provide a possible solution you will need data listed below:

I used Foursquare API to get the most common venues of given District of Taipei City [2].

I found the Second-level Administrative Divisions of the Taipei City from Spatial Data Repository of NYU [3]. The .json file has coordinates of the all city of Taiwan. I cleaned the data and reduced it to city of Taipei where I used it to create choropleth map of Housing Sales Price Index of Taipei.

I used Google Map, 'Search Nearby' option to get the center coordinates of each District. [4].

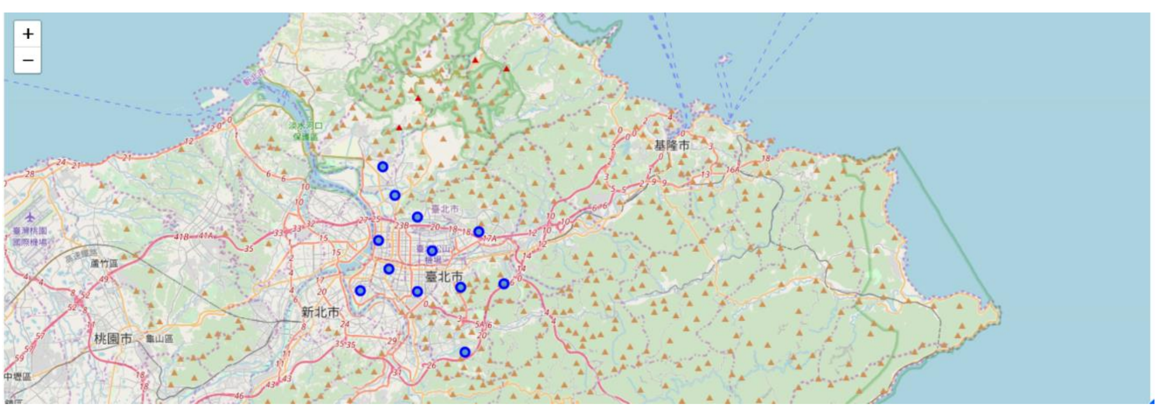
I collected latest per square meter Housing Sales Price (HSP) Averages for each District of Taipei from housing retail web page [5]

B. Methodology

As a database, I used GitHub repository in my study. My master data df which has the main components District, Average House Price, Latitude and Longitude information of the city.

	District	Avg-housePrice	Latitude	Longitude
0	Beitou	45026000	25.115176	121.515018
1	Daan	65344000	25.026158	121.542709
2	Datong	66177125	25.062724	121.511306
3	Nangang	43702353	25.031235	121.611195
4	Neihu	107481500	25.068942	121.590903

I used python **folium** library to visualize geographic details of Taipei and its districts and I created a map of Taipei with districts superimposed on top. I used latitude and longitude values to get the visual as below:



I applied the Foursquare API to analyze the districts and divide them. I arranged the limit as **100 venue** and the radius **750 meter** for each district from their given latitude and longitude information. Here is a head of the list

Venues name, category, latitude and longitude information from Foursquare API.

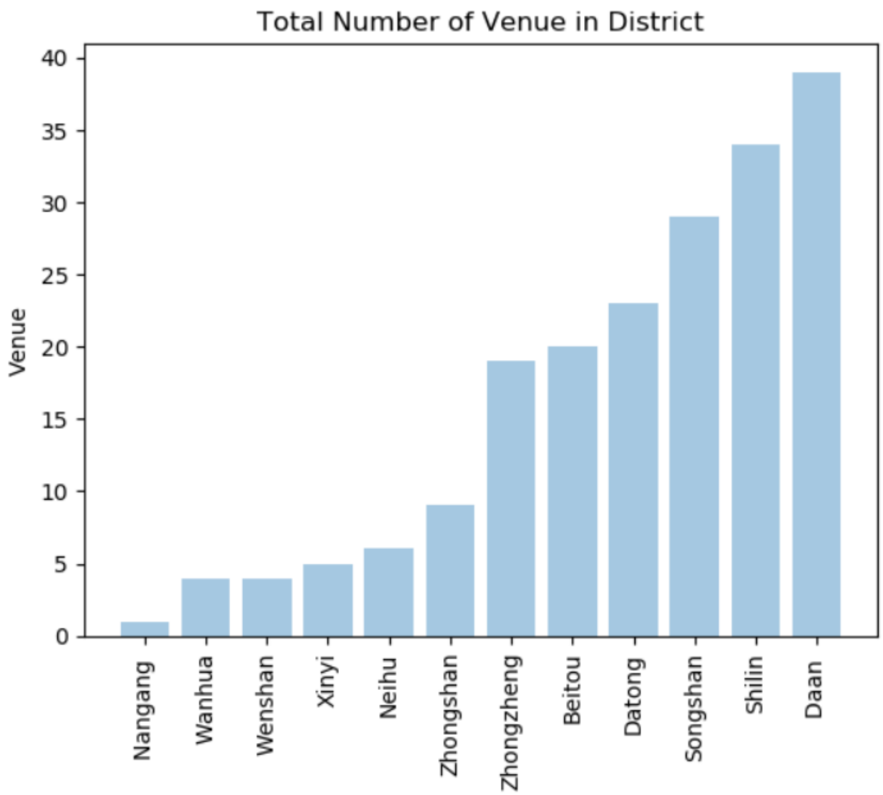
	name	categories	lat	lng
0	水龜伯古早味	Dessert Shop	25.116794	121.515918
1	阿二麻辣旗艦店	Hotpot Restaurant	25.116681	121.516192
2	石牌夜市 Shipai Nightmarket	Night Market	25.116622	121.516702
3	東方泰國小館	Thai Restaurant	25.114670	121.515385
4	蕭記大餛飩	Chinese Restaurant	25.116001	121.517358

In summary of this data **27** venues were returned by Foursquare. Here is a merged table of districts and venues.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beitou	25.115176	121.515018	水龜伯古早味	25.116794	121.515918	Dessert Shop
1	Beitou	25.115176	121.515018	阿二麻辣旗艦店	25.116681	121.516192	Hotpot Restaurant
2	Beitou	25.115176	121.515018	東方泰國小館	25.114670	121.515385	Thai Restaurant
3	Beitou	25.115176	121.515018	石牌夜市 Shipai Nightmarket	25.116622	121.516702	Night Market
4	Beitou	25.115176	121.515018	蕭記大餛飩	25.116001	121.517358	Chinese Restaurant

From the graph it's clearly show that all the districts are bellow 40 venues in our given coordinates with latitude and longitude. From this graph not all the possible results in each district are process.

It's all depends on a given latitude and longitude data available, it's boils down on a single latitude and longitude pair for each borough. For future studies we can look for more data of neighborhood with more latitude and longitude information.



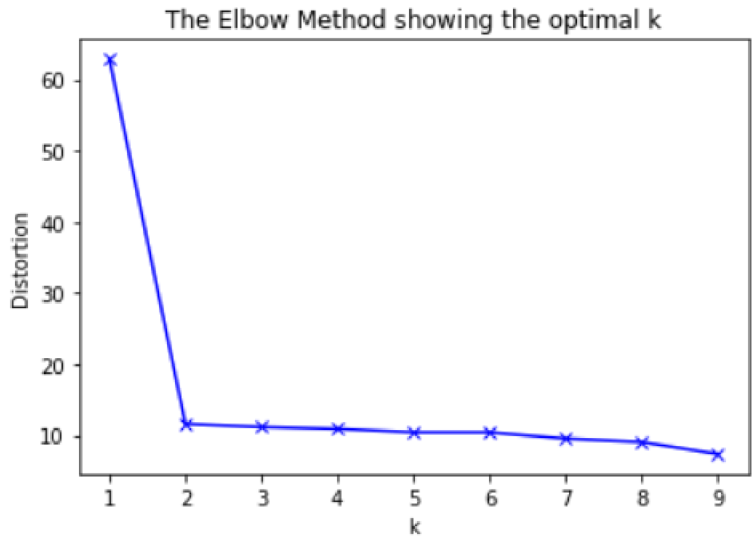
In summary of this graph **68** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

:

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Beitou	Chinese Restaurant	Coffee Shop	Convenience Store	Dumpling Restaurant
1	Daan	Coffee Shop	Café	Taiwanese Restaurant	Noodle House
2	Datong	Dessert Shop	Noodle House	Taiwanese Restaurant	Convenience Store
3	Nangang	Boat or Ferry	Convenience Store	Vietnamese Restaurant	Department Store
4	Neihu	Convenience Store	Food Court	Bus Station	Coffee Shop

K-Means algorithm is one of the most common cluster methods of unsupervised learning. I will use K-Means algorithm for my study in this project.

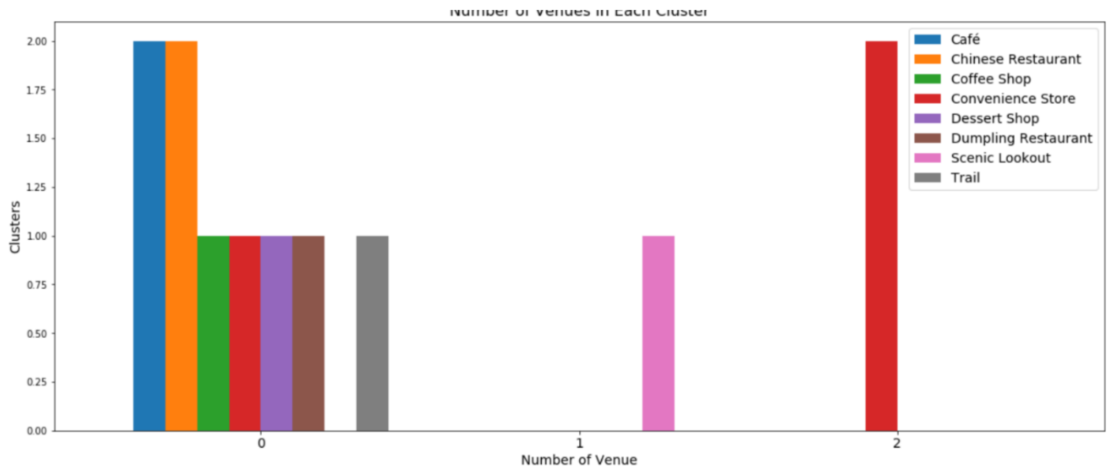
First, I will run K-Means to cluster the districts into 3 clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means



Here is my merged table with cluster labels for each district.

	District	Avg-housePrice	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	Beitou	45026000	25.115176	121.515018	1	Chinese Restaurant	Coffee Shop
1	Daan	65344000	25.026158	121.542709	1	Coffee Shop	Café
2	Datong	66177125	25.062724	121.511306	1	Dessert Shop	Noodle House
3	Nangang	43702353	25.031235	121.611195	2	Boat or Ferry	Convenience Store
4	Neihu	107481500	25.068942	121.590903	1	Convenience Store	Food Court

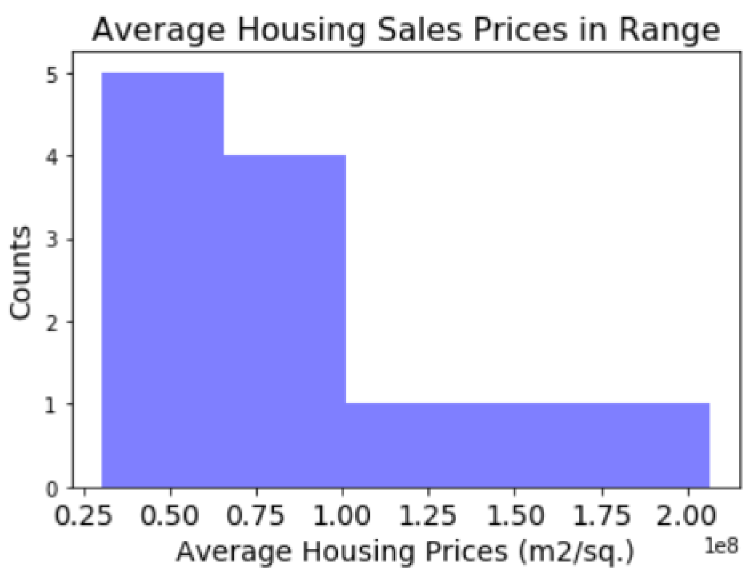
We can also estimate the number of **1st Most Common Venue** in each cluster. Then, we can create a bar chart which may help us to find proper labels for each cluster.



When we examine above graph, we can label each cluster as follows:

- Cluster 0: "Cafe Venues"
- Cluster 1: "Multiple Social Venues"
- Cluster 2: "Accommodation & Intensive Cafe Venues"

We can also examine the frequency of average housing sales prices in different ranges. The histogram can help for visualization:



As it seems in above histogram, we can define the ranges as below:

- 50000000 AHP: "Low Level HSP"
- 75000000-100000000 AHP: "Mid-Level HSP"
- '>' 100000000 AHP: "High Level HSP"

One of the goals was also to show the number of top 3 venues information for each district on the map. Then, I grouped each district by the number of top 3 venues, and I combined this information in Join column.

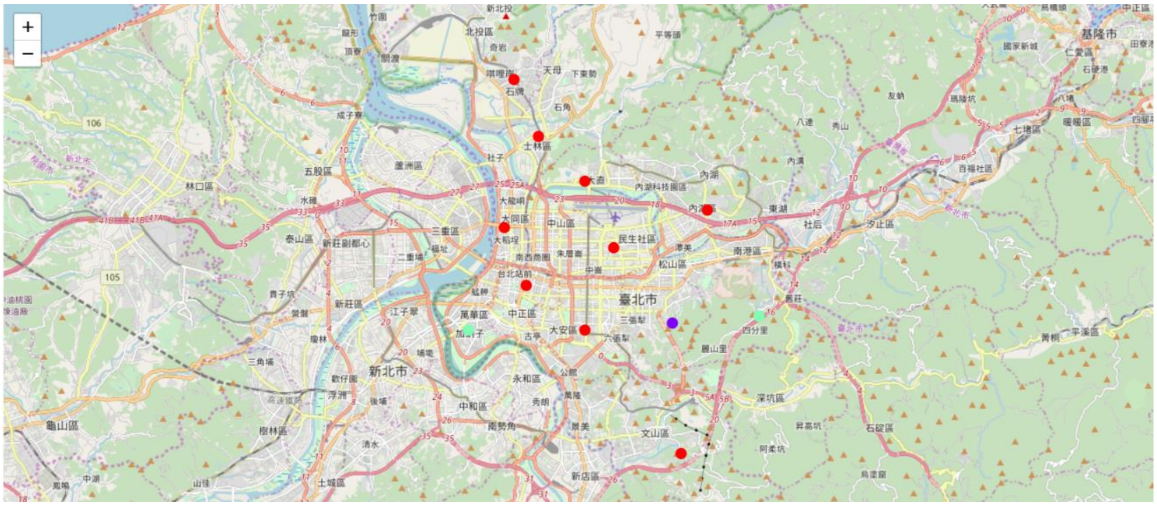
	District	Join
0	Beitou	6 Chinese Restaurant, 2 Coffee Shop, 2 Conveni...
1	Daan	4 Café, 4 Coffee Shop, 3 Noodle House
2	Datong	4 Dessert Shop, 2 Convenience Store, 2 Noodle ...
3	Nangang	1 Boat or Ferry, 1 Convenience Store
4	Neihu	2 Convenience Store, 1 Asian Restaurant, 1 BBQ...

C. Results

Let's merge those new variables with related cluster information in our main **master table**.

6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Join	Labels	Level_labels
Korean Restaurant	Breakfast Spot	Dessert Shop	Hotpot Restaurant	Fried Chicken Joint	6 Chinese Restaurant, 2 Coffee Shop, 2 Conveni...	Multiple Social Venues	Low Level HSP
Farmers Market	Pizza Place	Ice Cream Shop	Department Store	Hotpot Restaurant	4 Café, 4 Coffee Shop, 3 Noodle House	Multiple Social Venues	High Level HSP
History Museum	Hotel	Hotpot Restaurant	Coffee Shop	Café	4 Dessert Shop, 2 Convenience Store, 2 Noodle ...	Multiple Social Venues	High Level HSP

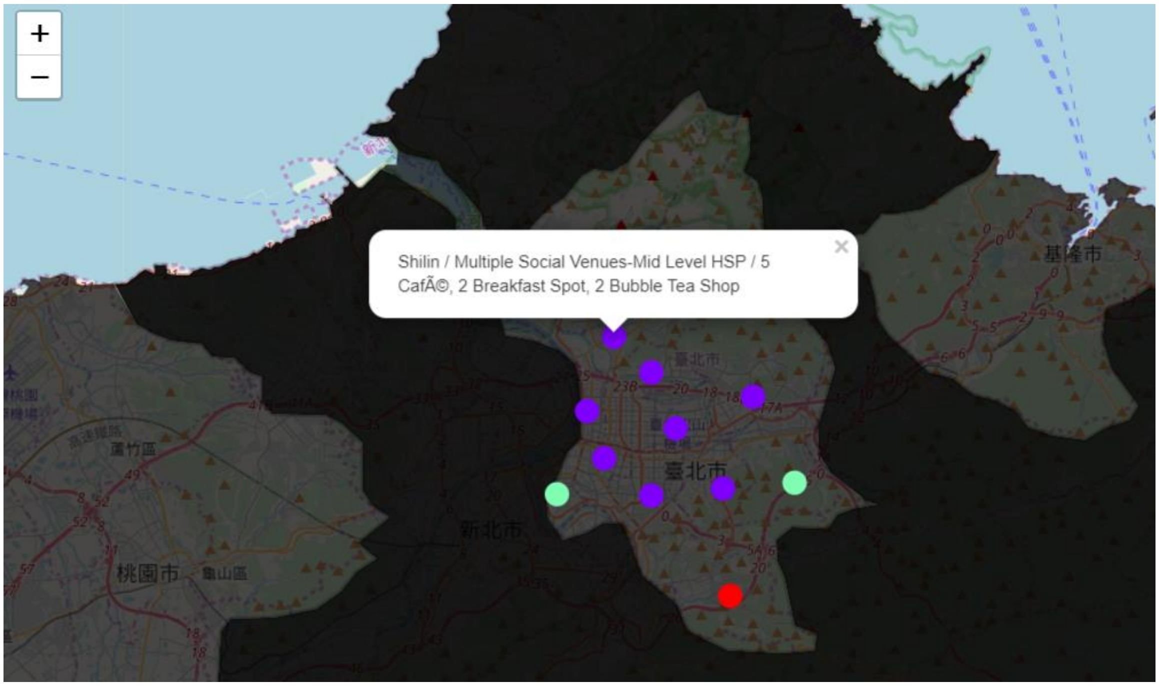
You can now see Join, Labels and Level labels columns as the last three ones in above table. You can also see a clustered map district of Taipei in the below.



Another goal of this project was to visualize the Average Housing Sale Prices per square meter with choropleth style map. A json file of Second-lever Administrative Divisions of Taiwan from Spatial Data Repository of NYU. I cleaned the json file and pull out only Taipei city.

In final section, I created choropleth map which also has the below information for each borough:

- District name,
- Cluster name,
- Housing Sales Price (HSP) Levels,
- Top 3 number of venues



D. Discussion

As outlined above, Taipei has a very dense population compare to the other cities in Taiwan. For this project I focus on clustering to get the best result, however, population densities of the 12 districts can vary. In future project different approaches can be tried in clustering and classification studies.

To process the data, I used Kmeans algorithm as part of the clustering study. I set the optimum k value to 3 during the elbow method, even though there were only 12 districts. For future research, the data can be expanded with more details of the neighborhood or street for more accurate result.

I also add coordinates of districts and home sales price averages as static data on GitHub. For other studies, these data can be used.

One of the aims was to visualize the data and clustering information on the Taipei map. However, other info can be added like, web or telephone applications for investors.

F. Conclusion

Nowadays, a lot of entrepreneur are willing to start their own business by moving to big cities. From this platform, they can get available information to achieve better outcomes and make smart decision.

This info can also help government representatives or city managers if in need of data analysis.

References:

[1] [Taipei_Wikipedia] (https://en.wikipedia.org/wiki/Taipei)

[2] [Forsquare API] https://developer.foursquare.com/

[3][Second-level Administration Division of Taiwan] https://geo.nyu.edu/catalog/stanford-fn648mm8787

[4] [Google map] https://www.google.com/maps/

[5] [Housing Sales Prices of each District from "Century21global" for 2020"] https://www.century21global.com/taiwan