CAPSTONE PROJECT

THE BATTLE OF THE NEIGHBORHOODS

Applied Data Science Capstone by IBM/Coursera

- Taiwan is the 17th most densely populated country in the world. Its capital city, Taipei, and New Taipei City are the two the most densely populated areas in the territory. Taiwan has a culture of overworking. Dining out becomes the most preferred leisure activity of Taiwanese and resulting in the blossom of the restaurantion.

- Our mission is to help shareholders to find a optimal location to open a new venue

INTRODUCTION

- When it comes to open a new venue, different factors need to take into consideration. A diverse economic environment will be more ideal for a restaurant business. This report attempts to provide some semblance to these questions based on the analytics of the data collected from public platforms.

PROBLEM AND DISCUSSION I

- The questions have been narrowed to the follows:

1. Which districts have a greater population density?

2. Which districts have a better economic diversity?

3. Which district has the most potential to open a venue?

4. What business can be a good idea to open.

PROBLEM AND DISCUSSION II

To understand better the business activity in Greater Taipei Area, we need to collect not only basic demography data, but also geographic data:

▶ Taipei and New Taipei geographic coordinate data can be found on Taiwan opening government data platform: https://data.gov.tw/en

▶ New Taipei City Municipal administration data: https://en.wikipedia.org/wiki/New_Taipei_City

▶ Taipei City Municipal administration data: https://en.wikipedia.org/wiki/Taipei

▶ Taipei and New Taipei Geojson data: https://github.com/littlebtc/geodata4appy/tree/master/simplified

DATA COLLECTION

- Pandas.read_html() function is used as a scrapper to obtain the municipal administration data of Taipei and New Taipei City from Wikipedia.

- After cleaning, population density is calculated.

- Append() function is used to merge the Taipei dataset and New Taipei City dataset.

- The geographic coordinate data then eventually joined with the Dataframe using the join() function.

DATA PROCESSING

DATA

	District	Postal Code	Longitude	Latitude	Population	Area_sq.km	Density_sq.km
0	Zhongzheng	100	121.519884	25.032405	162549	7.6071	21368.0
1	Datong	103	121.513042	25.063424	131029	5.6815	23062.0
2	Zhongshan	104	121.538160	25.069699	231286	13.6821	16904.0
3	Songshan	105	121.557588	25.059991	209689	9.2878	22577.0
4	Daan	106	121.543445	25.026770	312909	11.3614	27541.0
5	Wanhua	108	121.497986	25.028590	194314	8.8522	21951.0
6	Xinyi	110	121.571670	25.030621	229139	11.2077	20445.0
7	Shilin	111	121.550847	25.125467	290682	62.3682	4661.0
8	Beitou	112	121.517799	25.148068	257922	56.8216	4539.0
9	Neihu	114	121.592383	25.083706	287726	31.5787	9111.0
10	Nangang	115	121.609757	25.036009	122296	21.8424	5599.0
11	Wenshan	116	121.573608	24.988579	275433	31.5090	8741.0
12	Wanli	207	121.643931	25.175725	22634	63.3766	357.0
13	Jinshan	208	121.605264	25.217146	22273	49.2132	453.0
14	Banqiao	220	121.457968	25.011865	554008	23.1373	23944.0
15	Xizhi	221	121.654699	25.073313	196150	71.2354	2754.0
16	Shenkeng	222	121.620062	24.997675	23614	20.5787	1147.0
17	Shiding	223	121.647228	24.947141	7857	144.3498	54.0
18	Ruifang	224	121.823202	25.098129	40922	70.7336	579.0
19	Pingxi	226	121.757882	25.026071	4872	71.3382	68.0
20	Shuangxi	227	121.832982	24.996984	9233	146.2484	63.0
21	Gongliao	228	121.918247	25.024856	12858	99.9734	129.0
22	Xindian	231	121.531656	24.930390	300283	120.2255	2498.0
23	Pinglin	232	121.724223	24.910971	6503	170.8350	38.0
24	Wulai	233	121.541481	24.788243	6182	321.1306	19.0

GEOJSON DATA
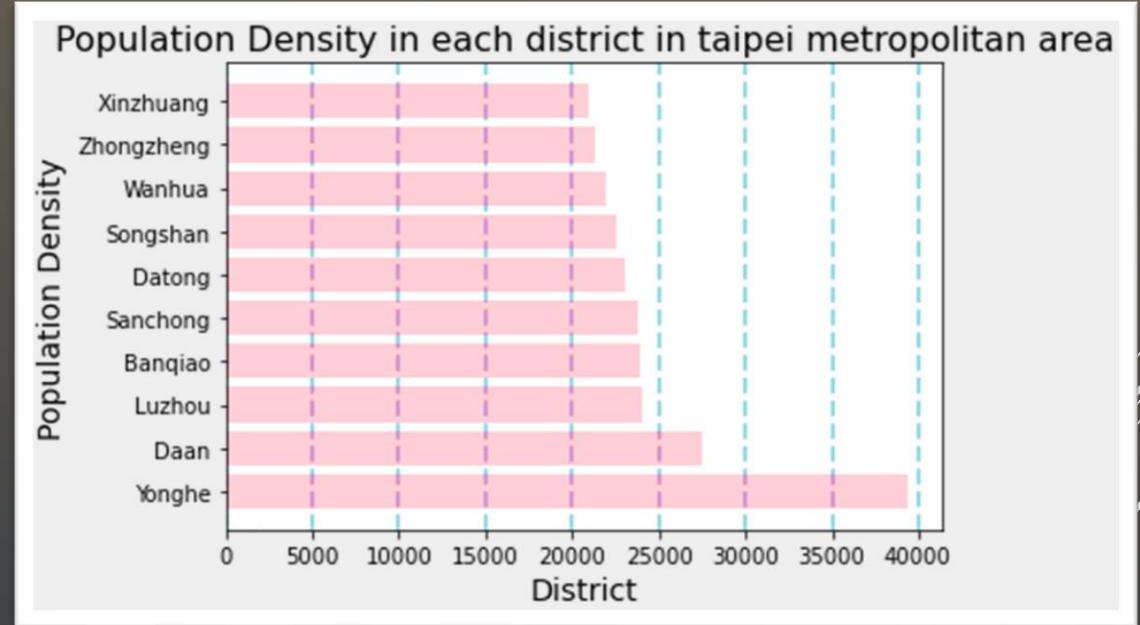
- Descriptive statistics analyse is used to answer the population density and the economic diversity question.

- Foursquare API is used to explore the activities in Greater Taipei City

- K-means clustering then used to discover the optimal location for the venue

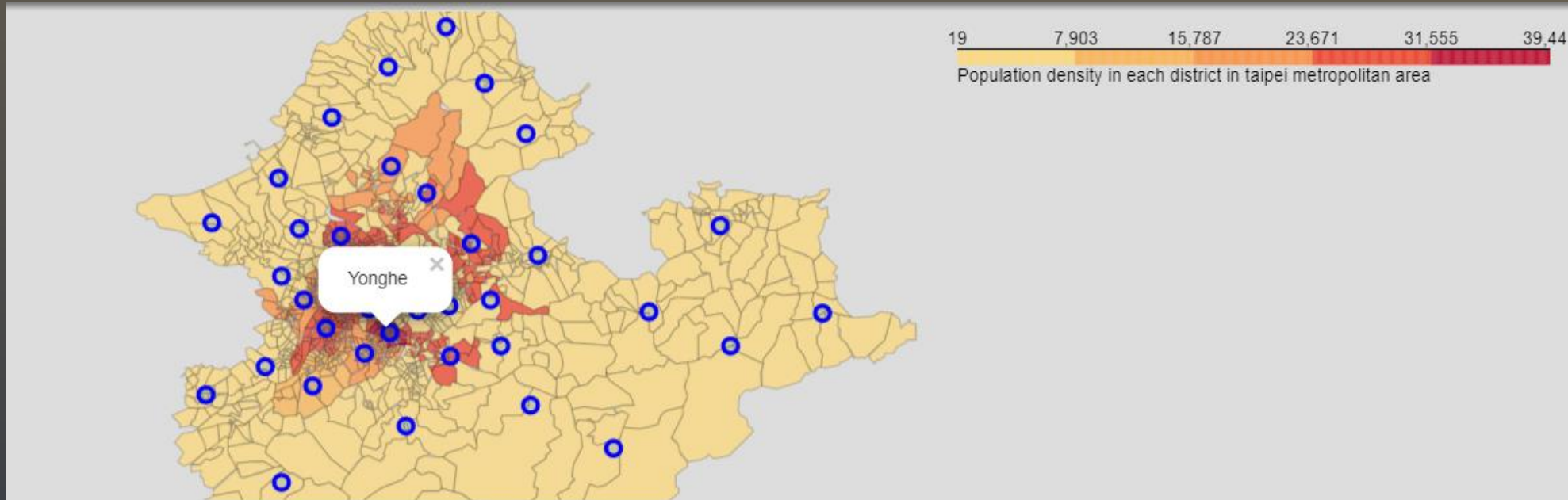# METHODOLOGY

The population density

The most densely populated district Yonghe, Daan, Luzhou, Banqiao

	District	Postal Code	Longitude	Latitude	Population	Area_sq.km	Density_sq.km
25	Yonghe	234	121.516745	25.008102	225353	5.7138	39440.0
4	Daan	106	121.543445	25.026770	312909	11.3614	27541.0
35	Luzhou	247	121.471246	25.089272	200055	8.3210	24042.0
14	Banqiao	220	121.457968	25.011865	554008	23.1373	23944.0
31	Sanchong	241	121.487098	25.062817	388386	16.3170	23803.0
1	Datong	103	121.513042	25.063424	131029	5.6815	23062.0
3	Songshan	105	121.557588	25.059991	209689	9.2878	22577.0
5	Wanhua	108	121.497986	25.028590	194314	8.8522	21951.0
0	Zhongzheng	100	121.519884	25.032405	162549	7.6071	21368.0
32	Xinzhuang	242	121.436753	25.035830	413443	19.7383	20946.0

Population Density in each district in taipei metropolitan area

RESULTS

The population density

the population density in a geographical framework

Population density in each district in taipei metropolitan area

RESULTS

▶ Using Foursquare API for the exploration

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Zhongzheng	25.032405	121.519884	Kinfen Braised Pork Rice (金峰魯肉飯)	25.032194	121.518534	Taiwanese Restaurant
1	Zhongzheng	25.032405	121.519884	Chiang Kai-Shek Memorial Hall (中正紀念堂)	25.034555	121.521835	Monument / Landmark
2	Zhongzheng	25.032405	121.519884	樂田麵包屋 Gakuden Boulangerie	25.032757	121.517534	Bakery
3	Zhongzheng	25.032405	121.519884	National Theater (國家戲劇院)	25.035197	121.518188	Theater
4	Zhongzheng	25.032405	121.519884	臻味赤肉胡椒餅 烤地瓜	25.033022	121.518246	Bakery
...	...	...	...	...	...	...	...
336	Luzhou	25.089272	121.471246	三民路肯德雞	25.087815	121.471632	Fast Food Restaurant
337	Luzhou	25.089272	121.471246	肯德基 KFC（蘆洲三民餐廳）	25.087043	121.471592	Fried Chicken Joint
338	Luzhou	25.089272	121.471246	FamilyMart_蘆洲民族店	25.091190	121.474040	Deli / Bodega
339	Luzhou	25.089272	121.471246	捷運三民高中站 MRT Sanmin Senior High School Station	25.085741	121.473080	Metro Station
340	Sanzhi	25.231599	121.515558	八連溪水車	25.233907	121.514404	Park

RESULTS

► Top ten most common venues ini Greater Taipei area

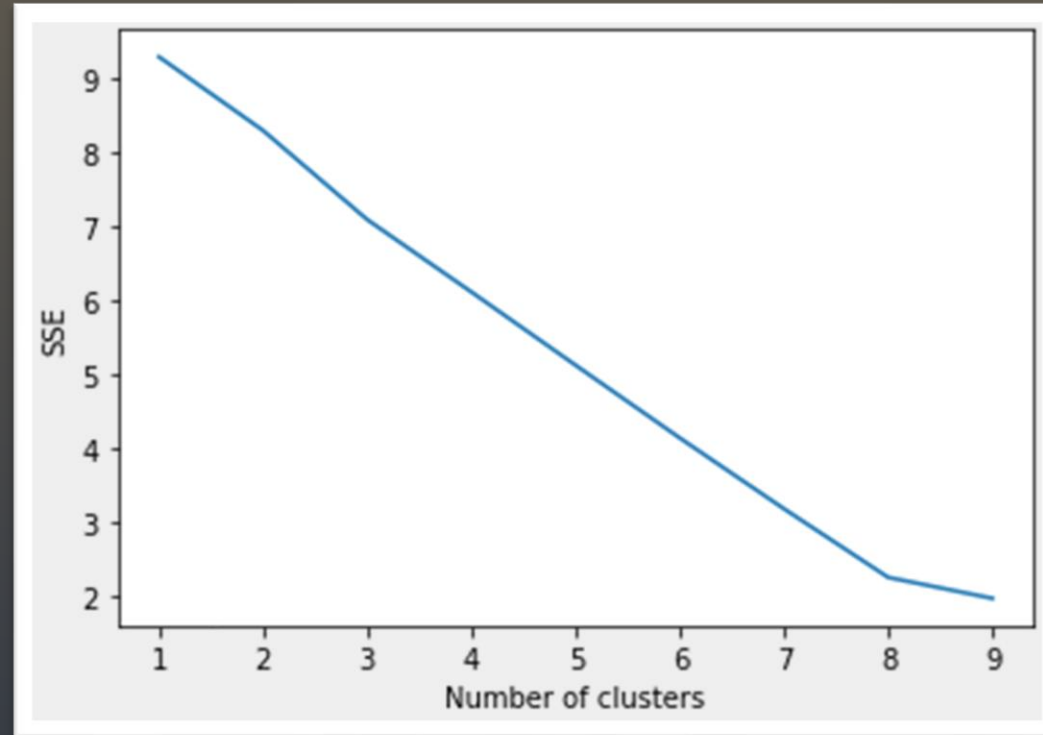|	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Banqiao	Japanese Restaurant	Bookstore	Taiwanese Restaurant	Hotel	Bus Stop	Café	Chinese Restaurant	Multiplex	Fast Food Restaurant	Convenience Store
1	Daan	Café	Coffee Shop	Noodle House	Taiwanese Restaurant	Chinese Breakfast Place	Farmers Market	Brewery	Ice Cream Shop	Dessert Shop	Cocktail Bar
2	Datong	Taiwanese Restaurant	Convenience Store	Food Truck	Asian Restaurant	Dessert Shop	Coffee Shop	Noodle House	Gym / Fitness Center	Bus Station	Flea Market
3	Jinshan	Farm	Hotel	Dumpling Restaurant	Concert Hall	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
4	Luzhou	Fried Chicken Joint	Deli / Bodega	Fast Food Restaurant	Metro Station	Women's Store	Discount Store	Concert Hall	Convenience Store	Cosmetics Shop	Department Store
5	Nangang	Convenience Store	Bus Station	Supermarket	Sake Bar	Breakfast Spot	Women's Store	Dive Bar	Cosmetics Shop	Deli / Bodega	Department Store
6	Neihu	Convenience Store	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Italian Restaurant	Breakfast Spot	Bookstore	Noodle House	Metro Station	Market
7	Sanchong	Convenience Store	Chinese Restaurant	Restaurant	Breakfast Spot	Women's Store	Coffee Shop	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop
8	Sanzhi	Park	Women's Store	Dive Bar	Concert Hall	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
9	Shenkeng	Taiwanese Restaurant	Farmers Market	Women's Store	Dive Bar	Concert Hall	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop
10	Shiding	Tea Room	Women's Store	Dumpling Restaurant	Concert Hall	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant

RESULTS

▶ The most popular and less popular business.

	Venue Category	Venue
28	Convenience Store	31
20	Café	29
26	Coffee Shop	22
92	Taiwanese Restaurant	17
69	Noodle House	13
14	Breakfast Spot	12
22	Chinese Restaurant	11
36	Dumpling Restaurant	8
7	Bakery	8
57	Italian Restaurant	7

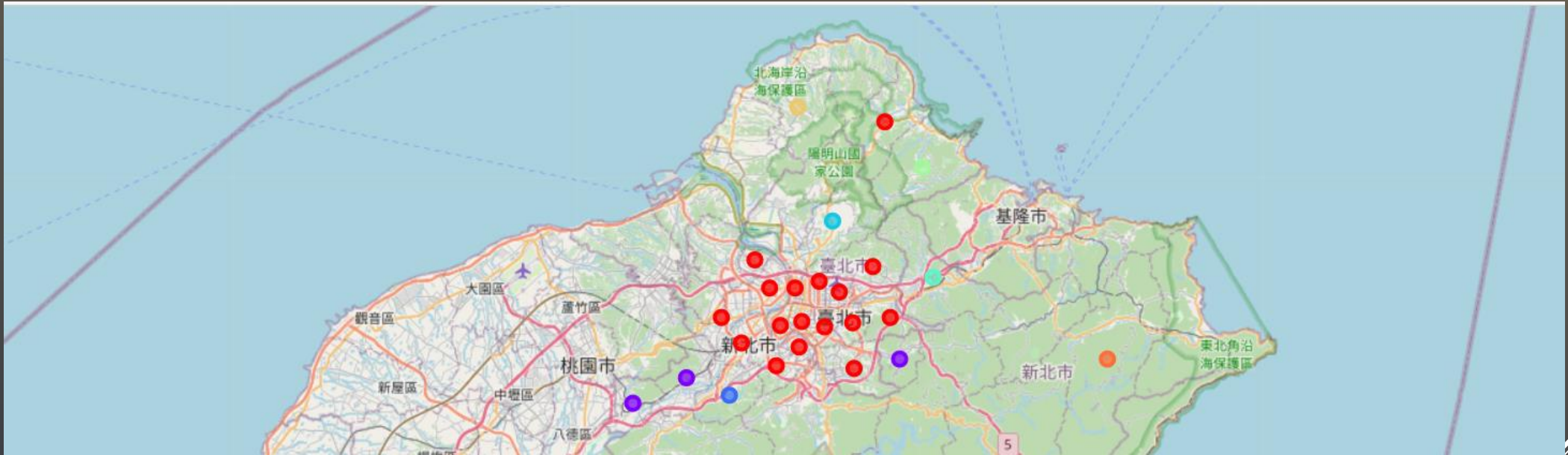|	Venue Category	Venue
48	Indie Movie Theater	1
51	Jewelry Store	1
53	Karaoke Bar	1
55	Latin American Restaurant	1
92	Women's Store	1

RESULTS

- K-means Clustering – the determination of the optimal K. Elbow method.

RESULTS

▶ Clustering on the map

RESULTS

- The most densely populated district is Yonghe, followed by Daan, Luzhou, Banqiao and Sanchong. Among them, Dann and Banqiao have a good diversity of economic activities

- Business type : Offering an obscure product/service like a Latin American restaurant can be an idea.

- Location: Cluster 1 has the most stable market with enough population to develop business. Opening a Latin American brunch place in Banqiao or Daan.

DISCUSSION & CONCLUSION

- Some aspects need to be considered. The more detailed demographic data is needed, the income, education level, age range…etc. Also the rent, transportation facility, and the lifestyle of the population.

- Not every business/venue has been listed on Foursquare. An official statistic data is recommended if available.

LIMITATION