Chengdu Borough Analyze

A. Description & Disscusion of the Background

The city of Chengdu is the biggest city in south-east China . Chengdu is the capital of Sichuan Province, a mega city, the core city of Chengdu metropolitan area, an important central city in western China approved by the State Council, an important national high-tech industrial base, trade and logistics center and comprehensive transportation hub.By 2020, the city has jurisdiction over 12 districts, 3 counties and 5 county-level cities, with a total area of 14,335 square kilometers.In 2019, the built-up area covered 949.6 square kilometers, with a permanent resident population of 16,581,000 and an urban population of 12,337,900, representing an urbanization rate of 74.41%.

Because of its economic importance and large population, chengdu's real estate prices remain high.As a commercial investor, he would like to know where the real estate investment cost will be lower and the commercial competition will be less fierce. In such a region, more commercial income will be obtained with less cost.

Therefore, in the following part, I will use the method of data science to visualize the housing price of various regions in Chengdu on the map in the form of thermal chart.In addition, I would use the Forsquare Api to get venue information and cluster them on the map.In this way, we will

consider the conduct of the business on both of the real estate price and the intense of business competition.

B. Data Description

To solve this problem, we need the following data:

1. I found the average real estate price data of each Borough in Chengdu on the website of www.anjuke.com.
2. I used Forsquare API to get the most common venues of given Borough of Chengdu.
3. I used Baidu Map API to obtain the center longitude and latitude of each Borough in Chengdu.
4. I use Aliyun API to get the geojson data of each Borough in Chengdu.

C. Methodology

	Borough	longitude	latitude	EestatePrice
0	ChenHua	104.153661	30.681403	16135
1	WuHou	104.055946	30.645411	17288
2	JinJiang	104.121031	30.632947	20100
3	LongQuanYi	104.313177	30.596178	11577
4	QinYang	104.020544	30.674066	20329
5	PiDu	103.871470	30.852131	10102
6	WenJiang	103.834341	30.726763	11323
7	XinDu	104.157378	30.832607	10794
8	ShuangLiu	103.958189	30.565774	13276

First of all, I obtained the information of all Boroughs of Chengdu from the website of The National Bureau of Statistics, and then called the API

interface of Baidu Map with the names of these Boroughs through the Python program, so as to obtain the longitude and latitude positions of the central points of these regions.

Then I cleaned up the data and spliced it into a Python dataframe object.

After obtaining the names and latitude and longitude of these Boroughs, I used Folium to visualize these regions to see where they are and how they relate to each other.

	name	categories	lat	lng
0	Starbucks Reserve (星巴克甄选)	Coffee Shop	30.662988	104.071605
1	The Ritz-Carlton, Chengdu (成都富力丽思卡尔顿酒店)	Hotel	30.666959	104.067618
2	Niccolo Chengdu	Hotel	30.657128	104.080289
3	The St. Regis Chengdu (成都瑞吉酒店)	Hotel	30.663286	104.072057
4	Taikoo Li (太古里)	Shopping Mall	30.655020	104.081711
5	The Temple House (博舍)	Hotel	30.653936	104.082239
6	Fraser Suites Chengdu (成都仁恒辉盛阁国际公寓)	Hotel	30.654694	104.065605
7	和幸 とんかつ Tonkatsu Wako	Japanese Restaurant	30.656739	104.076326
8	方所 Fangsuo Commune	Bookstore	30.655600	104.080319

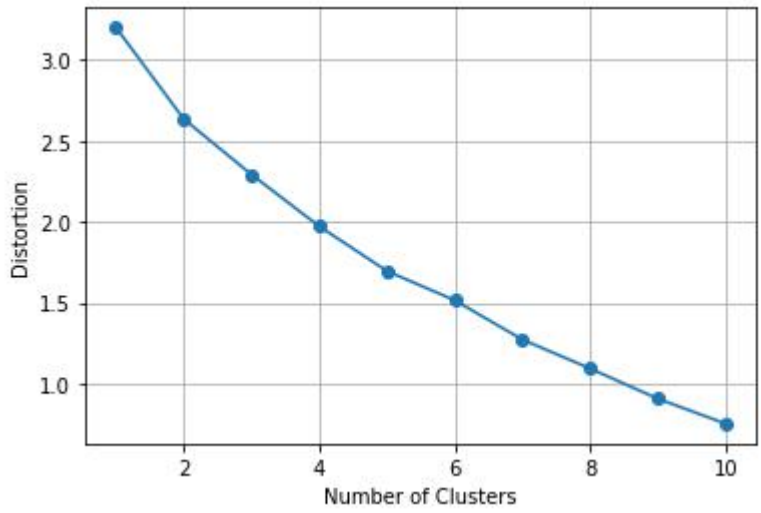
Then I pass these latitude and longitude information to the Forsquare API and use its explore interface.Since the Forsquare API is mainly aimed at North America and contains less information in Asia, I expanded the exploration radius to 5,000 meters and raised the limit of return for a single query to 500.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ChenHua	30.681403	104.153661	Chengdu Eastern Suburb Memory (东郊记忆)	30.671010	104.119535	Music Venue
1	ChenHua	30.681403	104.153661	Ito Yokado (伊藤洋华堂)	30.673877	104.107265	Shopping Mall
2	ChenHua	30.681403	104.153661	Starbucks (星巴克)	30.669571	104.109794	Coffee Shop
3	ChenHua	30.681403	104.153661	Starbucks (星巴克)	30.673792	104.108003	Coffee Shop
4	ChenHua	30.681403	104.153661	Starbucks (星巴克)	30.677556	104.106319	Coffee Shop
5	ChenHua	30.681403	104.153661	SUBWAY (赛百味)	30.670865	104.108734	Sandwich Place
6	ChenHua	30.681403	104.153661	SM City (SM广场)	30.669919	104.110056	Shopping Mall
7	ChenHua	30.681403	104.153661	Pizza Hut (必胜客)	30.673541	104.107753	Pizza Place

I connected the information returned by the Forsquare API to the form of each Borough, so that I could count the categories and quantities of each venue

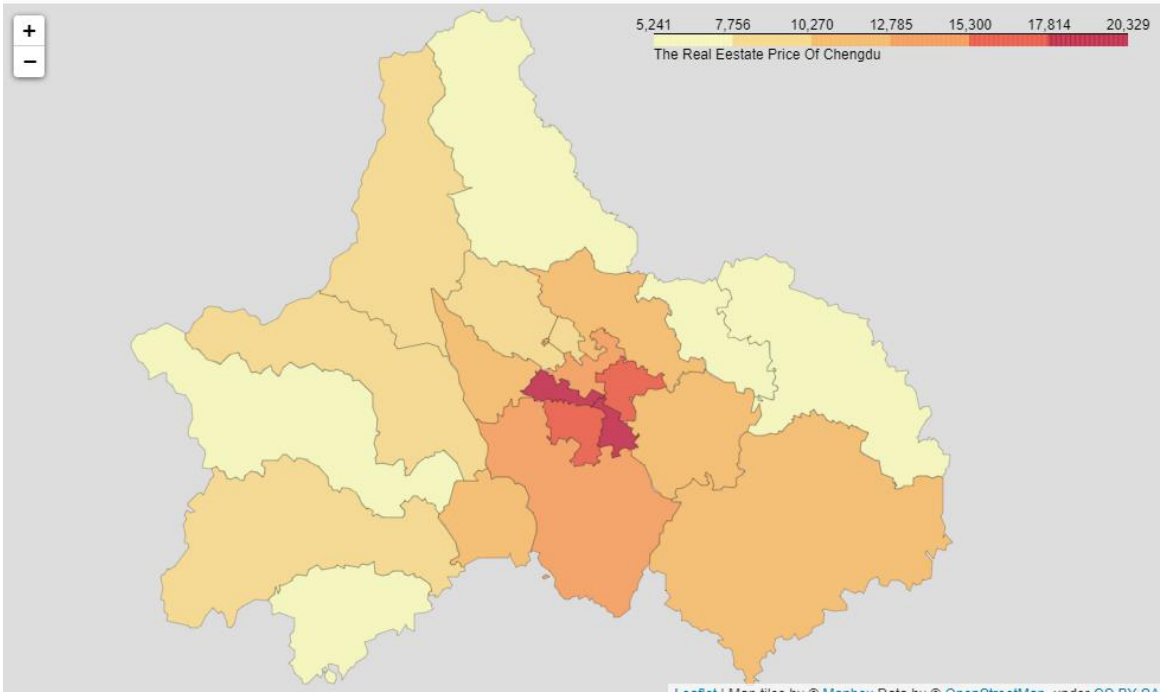
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ChenHua	Coffee Shop	Shopping Mall	Pizza Place	Café	Sandwich Place	Bus Station	Music Venue	Fast Food Restaurant	Food Court	Department Store
1	ChongZhou	Hotel	Women's Store	Chinese Restaurant	Train Station	Hotpot Restaurant	Thai Restaurant	Shopping Plaza	Coffee Shop	Bakery	Zoo Exhibit
2	DaYi	Restaurant	Hotel	Fast Food Restaurant	Fish & Chips Shop	German Restaurant	Convenience Store	Department Store	Donut Shop	Electronics Store	Food Court
3	DuJiangYan	Hotel	Historic Site	Shopping Mall	Asian Restaurant	Bus Station	Beach	Zoo Exhibit	German Restaurant	Donut Shop	Electronics Store

After data processing, I arranged the venues in each Borough according to their occurrence frequency, and displayed the relevant tables to observe the types of venues in each Borough.



Based on the above results, I use one-hot encoding method to encode the venue data and then normalize them.After these treatments, I used the k-means algorithem to cluster them.

How to determine K in the process of classification is an important problem, so I use the elbow-method to determine K.As we can see from the figure on the left, the curve inflection point occurs when K equals 4. Therefore, I choose 4 as my clustering parameter.

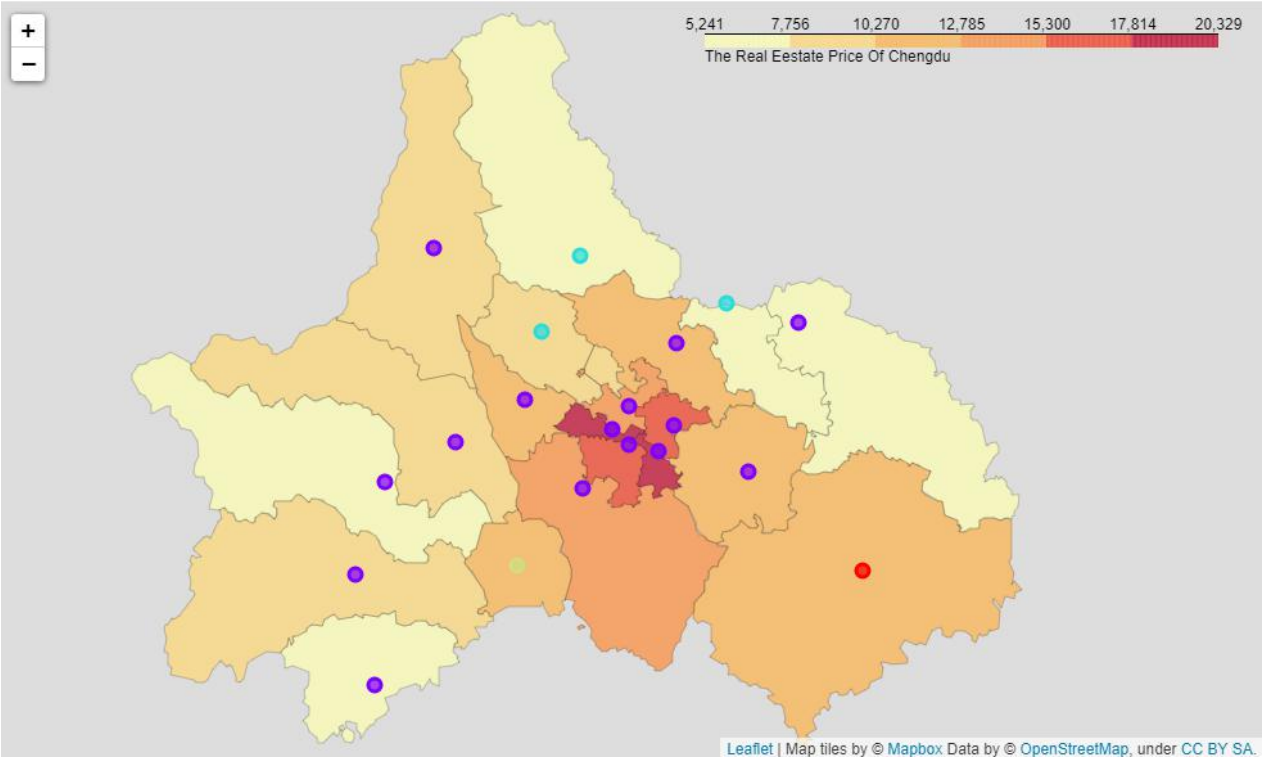


Another purpose of my project is to visualize the housing price of each Borough in Chengdu with thermal diagram.So I used Folium to visualize house prices.I used a Python program to crawl the housing prices of each Borough in Chengdu from Anjuke's website, and then merged these data with the names of Boroughs.In addition, I used Aliyun to obtain geojson data to characterize the boundaries of each Borough.After all this work, The house price thermal diagram will be shown.

D. Results

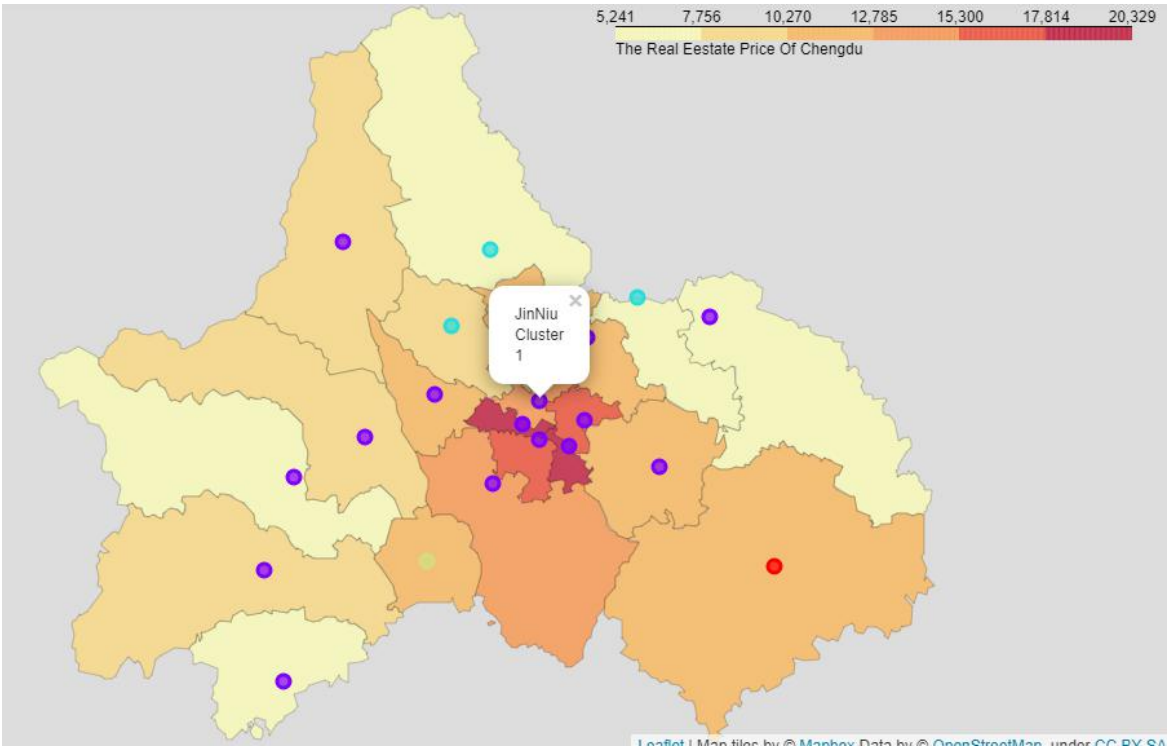
	Borough	longitude	latitude	EestatePrice	chinese_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Con
0	ChenHua	104.153661	30.681403	16135	成华区	1	Coffee Shop	Shopping Mall	Pizza Place	Café	Sandwich Place	Bus Station	Music
1	WuHou	104.055946	30.645411	17288	武侯区	1	Hotel	Coffee Shop	Shopping Mall	Hostel	Metro Station	Park	Szec Resta
2	JinJiang	104.121031	30.632947	20100	锦江区	1	Coffee Shop	Fast Food Restaurant	Shopping Mall	Hotel	Movie Theater	Pizza Place	Cockta
3	LongQuanYi	104.313177	30.596178	11577	龙泉驿区	1	Gym / Fitness Center	Mountain	Art Gallery	Hotel	Bus Station	Historic Site	Ge Resta
4	QinYang	104.020544	30.674066	20329	青羊区	1	Coffee Shop	Hotel	Chinese Restaurant	Hostel	Noodle House	Park	Pizza
5	PiDu	103.871470	30.852131	10102	郫都区	2	Park	Train Station	Art Museum	Asian Restaurant	Zoo Exhibit	Greek Restaurant	Donut

Finally I got four clusters.The first cluster belongs to the outer suburbs, where the frequency of business activity is low.The second cluster belongs to the business district, where there are many cafes and office buildings.The third cluster belongs to tourism and sports areas, where there are more sports venues and parks.The fourth cluster belongs to the transportation hub area, with more stations and an airport.



I visualized all four clusters on the map.The red dots belong to the suburbs.The purple dots belong to the business district.The blue dots belong to sports and tourism areas.The yellow dots belong to the traffic hub area.

E. Conclusion



Through all the above work, we can find that the JinNiu Borough is the best area to carry out business activities.The Borough marked on the left figure is JinNiu Borough.It belongs to the area with low housing price from the perspective of housing price heat diagram, but from the perspective of cluster classification, this area already has a lot of office and commercial facilities and abundant catering facilities. In addition, from the perspective of distance, it is close to other business districts.

Commercial activities in this area will have relatively low housing costs and great business potential.