**House Prices and Venues analysis of Guangzhou**

Xiaogang Zhu

1. **Introduction**

Guangzhou is at the heart of the most-populous built-up metropolitan area in mainland China, which extends into the neighboring cities of Foshan, Dongguan, Zhongshan and Shenzhen, forming one of the largest urban agglomerations on Earth, the Pearl River Delta Economic Zone. Administratively, the city holds sub-provincial status and is one of China's nine National Central Cities. At the end of 2018, the population of the city's expansive administrative area was estimated at 14,904,400 by city authorities, up 3.8% from the previous year.

I live in Foshan which is near Guangzhou. Nowadays lots of people heading Guangzhou for its rich job opportunities. Guangzhou is split to 11 districts and all of them have their own features. One important question is that if you want to work and live in Guangzhou which districts should you choose. It is may depend on the venues in that area, house prices, population densities and so on.

So after thinking about this question, I will try to discover the similarities and dissimilarities of the 11 districts in Guangzhou. Also the house prices of each district will be considered too. We will create a map include all the information which helps people easily decide which districts they would choose.

1. **Data**

To consider the question we can list the data below:

1. I use geopy package in python to find the center coordinate of Guangzhou as well as the center coordinate of 11 districts.

2. From http://datav.aliyun.com/ I find the geojson file of Guangzhou city which is used to create the choropleth map.

3. From http://gz.cityhouse.cn/market I find the average house prices of 11 districts of Guangzhou.

4. I use Forsquare API to get the most common venues of given districts of Guangzhou

1. **Methodology**

First of all I use geocoder Nominatim to obtain the exact enter coordinate of Guangzhou as well as its districts. There are 11 districts in Guangzhou: Liwan, Yuexiu, Haizhu, Tianhe, Baiyun, Huangpu, Panyu, Huadu, Nansha, Conghua and Zengcheng. We get all their coordinate and transfer to a dataframe.

A screenshot of a cell phone

Description automatically generated

A picture containing text, map

Description automatically generatedWe can create a map to visualize them.

Noticed that some of the districts are very close to each other and some are far from the city center.

Next we use Foursquare API to explore the districts and segment them. The limit of venues return is set to 100 since I only get the personal account. And also the radius is set to 2500 meters which means we would like to find the venues which locate within 2500 meters from districts center. After setting all these parameters we get 433 venues of all 11 districts. Head of the venues table is given below:

A screenshot of a cell phone

Description automatically generated

We also group the venues by districts to see how many venues each districts have.

A screenshot of a cell phone

Description automatically generated

From above Baiyun, Liwan and Yuexiu reach the 100 limit venues. This indicate that these districts are very popular. And if you check out their location you will find they locate in the city center. This explains why they are so popular. On contrary Nansha and Huangpu only receive 4 venues which may due to their large area and not locate in city center.

A screenshot of a cell phone

Description automatically generatedThere are 97 unique venues categories. In order to cluster the districts I created a table which shows list of top 10 venue category for each districts in below table.

And then we use K-means to cluster districts. First I want to find the optimize K so I use elbow method to plot and find the best k to be 2.

A close up of a map

Description automatically generated

After running the K-means algorithms we can see the result. Also we want to construct the choropleth map so I download the average house price data from the Internet and combine them with the district names. The price unit is yuan/m2. We also download the geojson file of Guangzhou.

A screenshot of a cell phone

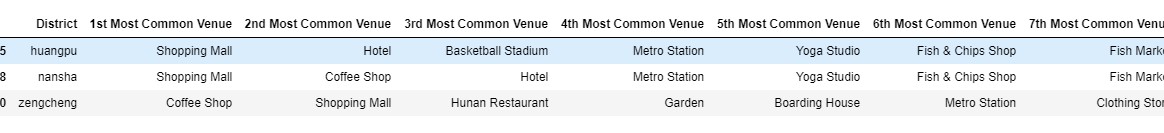
Description automatically generated

1. **Result**

Here we get our cluster result.

A screenshot of a social media post

Description automatically generatedCluster1:

Cluster2:

Huangpu, Nansha, Zengcheng are classify to one cluster while all the other 9 districts are cluster to cluster 1. Looking at the 1st most common venue we can label the cluster as:

Cluster 1: “Hotel, fast food and coffee shop venues”

Cluster 2: ”Shopping mall and coffee shop venues”

We can then visualize the cluster on map:

A picture containing text, map

Description automatically generated

Finally, we want to add average house price to create a choropleth map. The choropleth map includes the districts’ name, their clusters and average house prices.

A picture containing text, map

Description automatically generated

The house prices of the city center is higher than other, which is not surprising.

**5.Discussion**

After doing all this job we can easily visualize the feature of each districts in Guangzhou. It seems that the districts located in city center such as Tianhe have very high house prices, if we want to live in there we would probably cannot afford. But after cluster the districts we find that some districts such as Huadu shares the same venues categories with Tianhe. So in this case if you want to enjoy the city center’s rich venues but cannot afford the high house price, why not choose to live in Huadu where the house price is lower. So from the result someone may get what they want. But also there are some limitations of the analysis. Since the area of each districts are different and we choose their center coordinate, we cannot guarantee that the venues densities are the same within each district. We can repeat this analysis by changing the coordinate to what we want to get specific result.

**6.Conclusion**

As a result, more and more people are heading to their ideal cities for working and living. So it might be very helpful to analyze the city before you go and determine where do you want to stay. For those investor this kind of analysis is very helpful too.

Hope we all have a good future!