**How to find a place in Shanghai to open a new coffee shop**

# Introduction

## Description and Discussion of the background

Shanghai is one of the four municipalities of the People's Republic of China. It is located on the southern estuary of the Yangtze, with the Huangpu River flowing through it. With a population of 24.28 million as of 2019, it is the most populous urban area in China and the second most populous city proper in the world. Shanghai is a global center for finance, innovation and transportation and the Port of Shanghai is the world's busiest container port.[1]

As you see, Shanghai has a high population. There are a lot of retails in the city. The wealth is different from each district.

A friend of mine wants to open a coffee shop in Shanghai.

The coffee shop he wants to open has a theme about the culture, so he want to the place where he open the shop can have culture background. And around the shop, there is not too much other coffee shops. Of course, the more the people want to drink coffee in the place, the better.

# Data Description

As I consider all the problem above，we can create a map with all the districts’ data which show the following information.

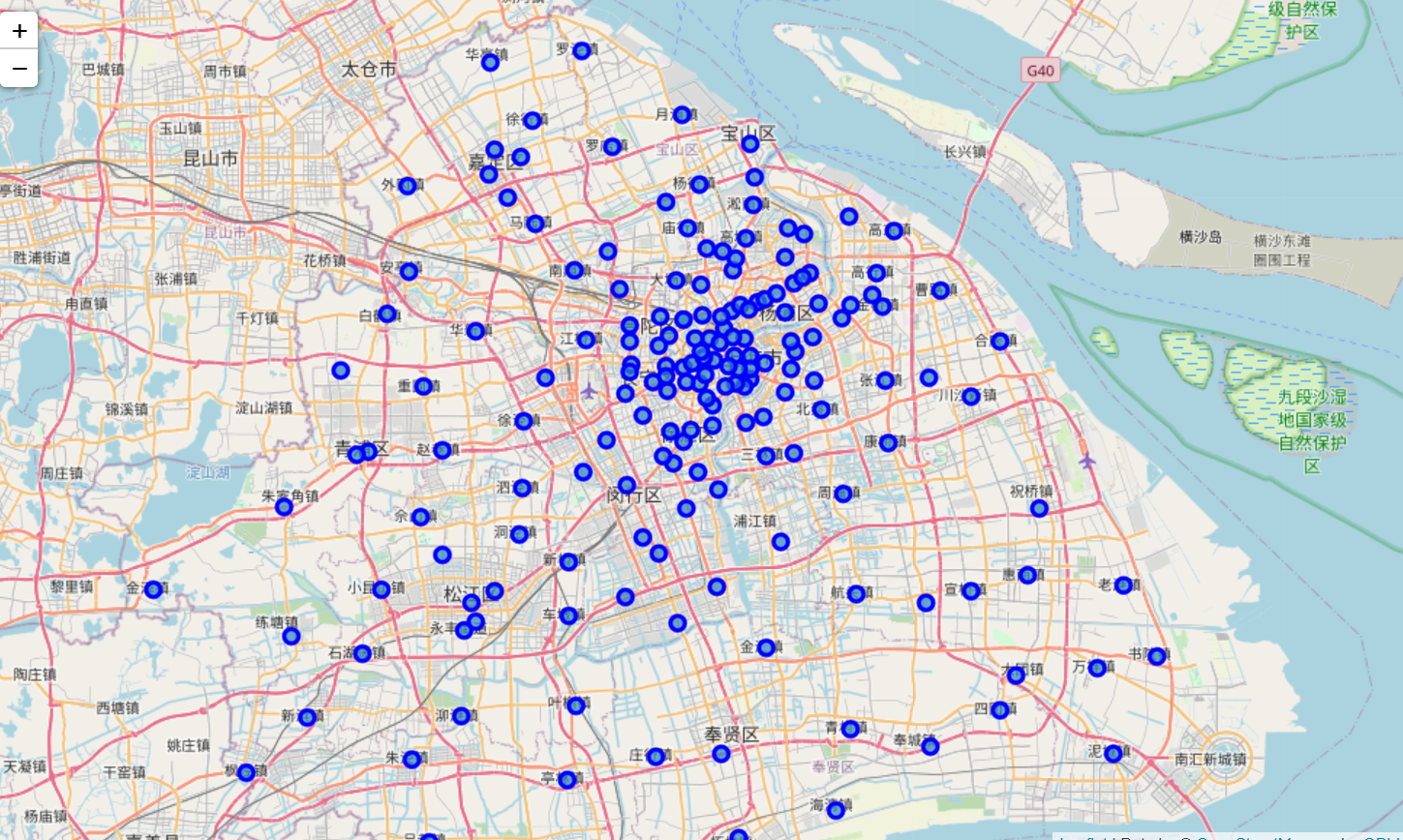
## I get the ‘街道’ district from the wiki page.[2]

First, I want to divide Shanghai City to 街道 district. I find all the 街道 from the wiki page, use the pandas mean to download them. There are some problems about the data. First the tables are organized as 区. I find the first row of every table is 区, so I insert ‘区’ in the dataframe to represent how the ‘街道’ is organized and drop the first row. All do that, I make the 区 sheets one whole Shanghai 街道 pandas’ dataframe.

## Use the Geocoder to turn the 街道’district’ to longitude and latitude.

In order to visualize the 街道Location in the map, I need to get the longitude and latitude of all the 街道. In order to do that, I use the geocoder method and google map API to get the longitude and latitude. And insert longitude and latitude column in the Shanghai pandas’ dataframe. I do not separate the longitude and latitude data. I will explain how I use the data later.

## To see how the 街道 is located in Shanghai city ,I use folium to visualize them ,one dot mean on 街道 as the following map.



As we can see, the 街道 is not located very averagely. The kinds of the nearby shops are different from each other. We need to get the venues around the ‘街道’.

## Use the four-square api to get the venues around the ‘街道’

As I mentioned before, my friend want to find the place that has cultures nearby, so I use the four-square api to get all the nearby culture places. I set the radius as 3000 meter, because I assume that one person can walk around 3000 meters to get to a coffee shop. The data from four-square is Jason form, first to get the name and the categories. We don’t care the name of the categories, but the numbers of the categories. I group the data by categories, and the numbers.

## Download a geojson file of Shanghai from gitbit[3]

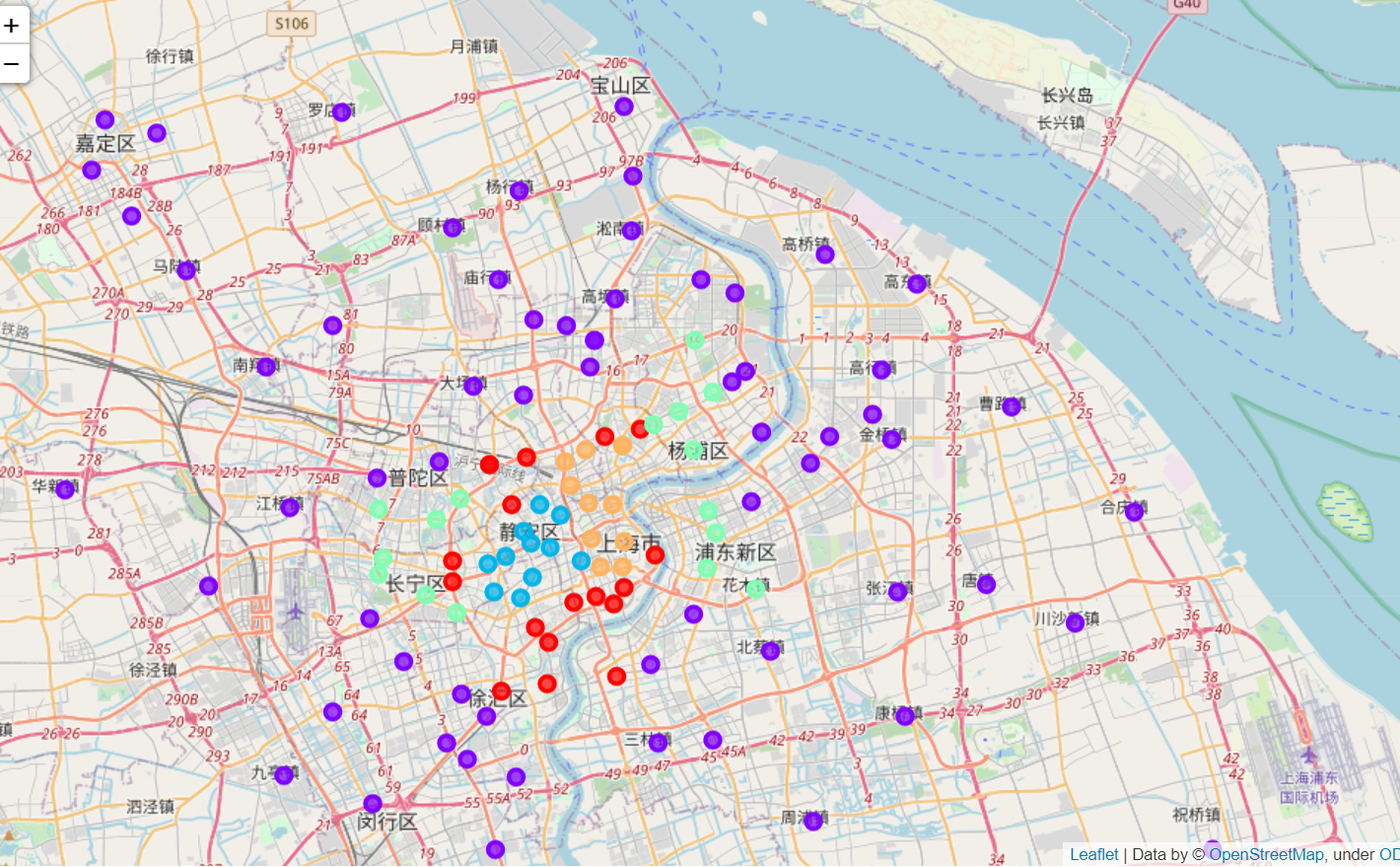
In order to make a heat map of Shanghai’s house-price, we need the to get the geojson file of shanghai. It can be easily got from the internet.

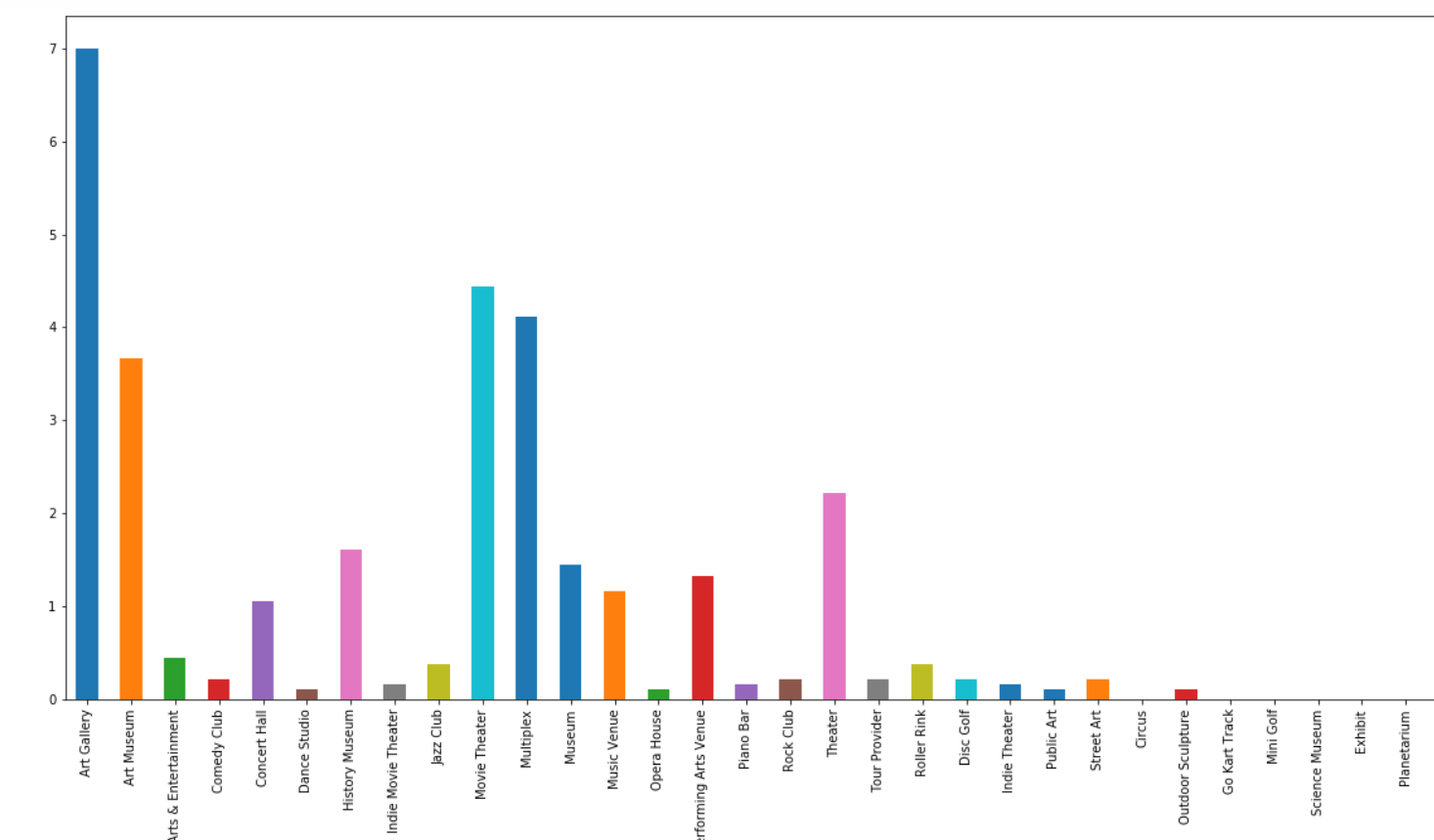
## Get the rent price of Shanghai

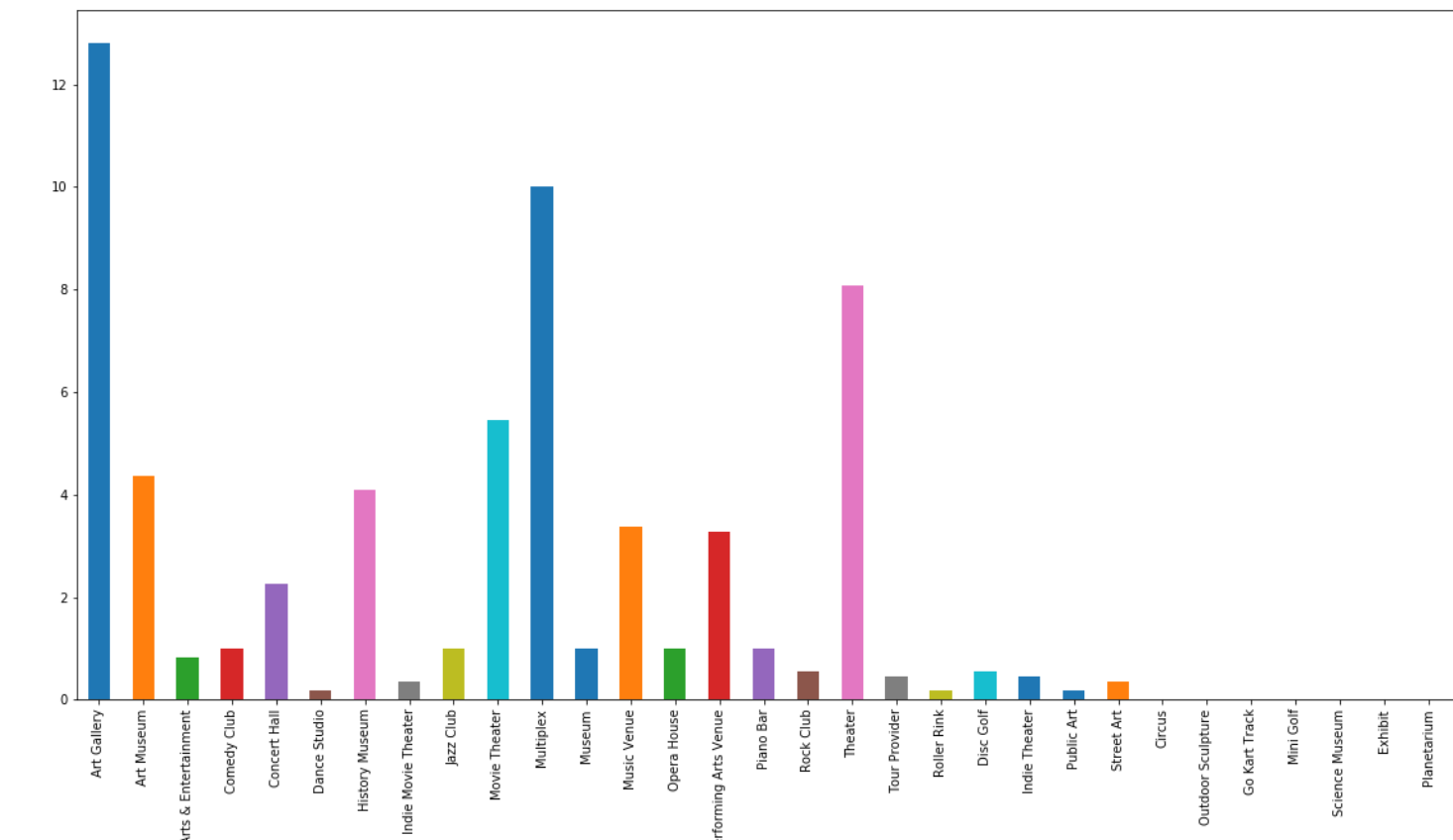
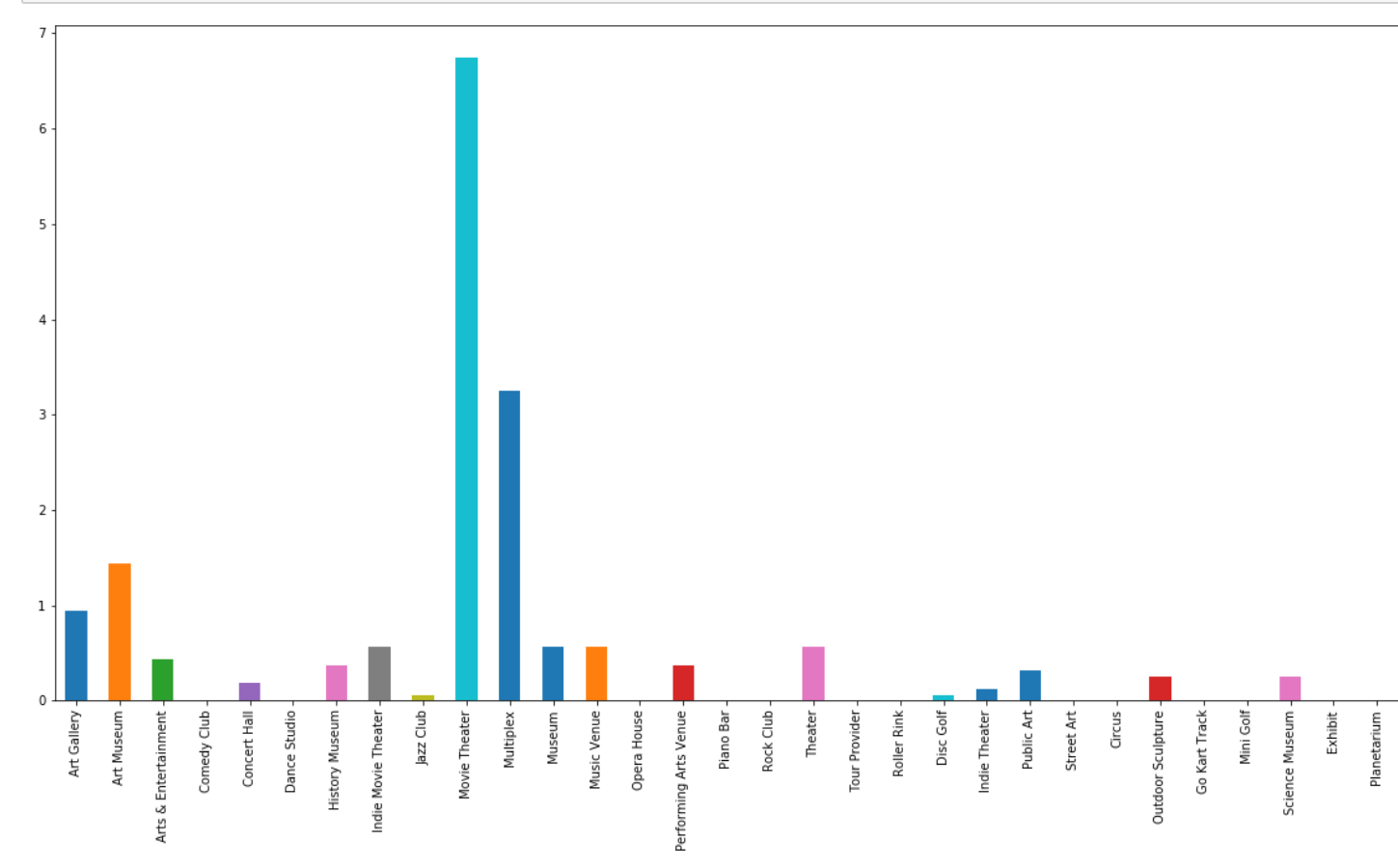
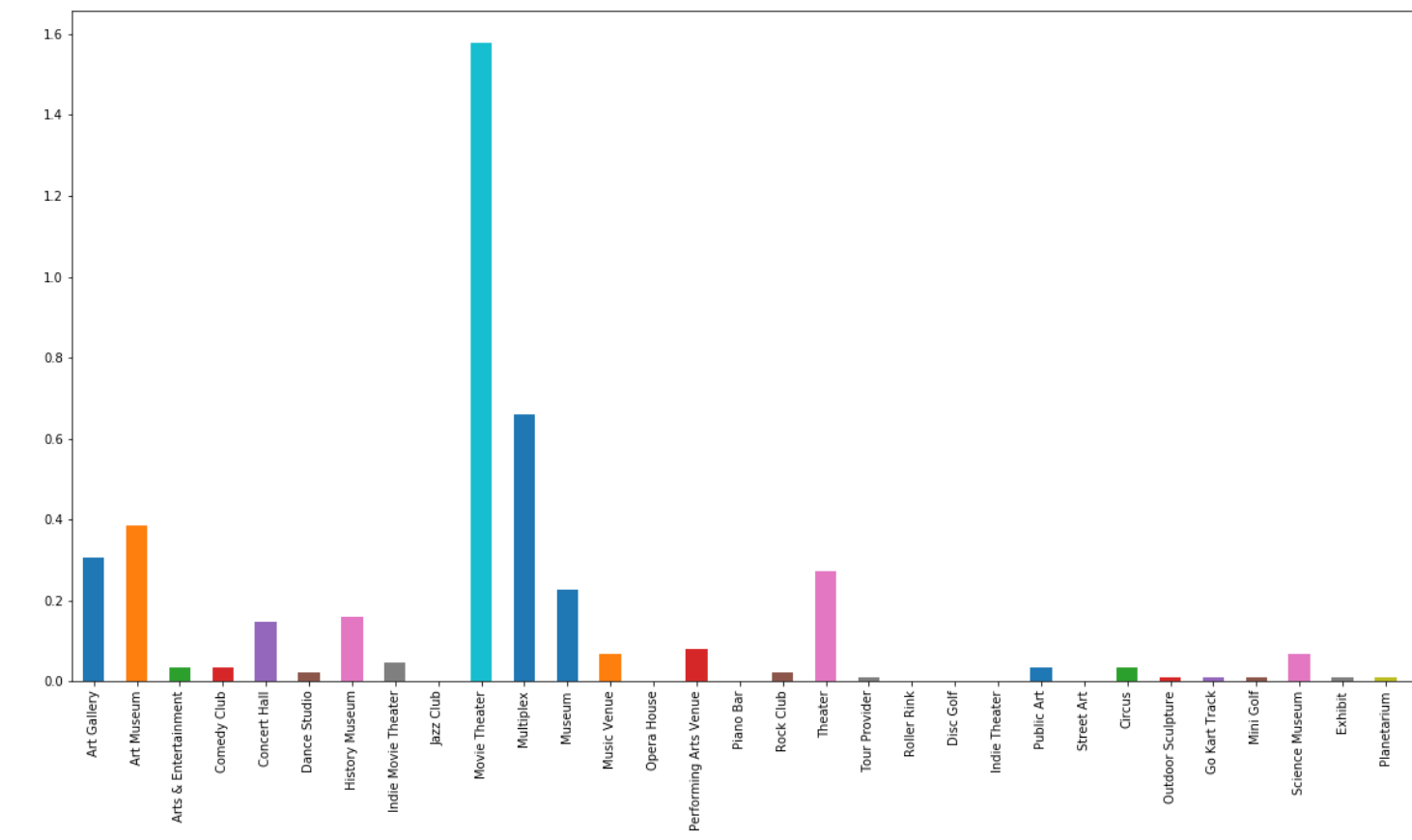
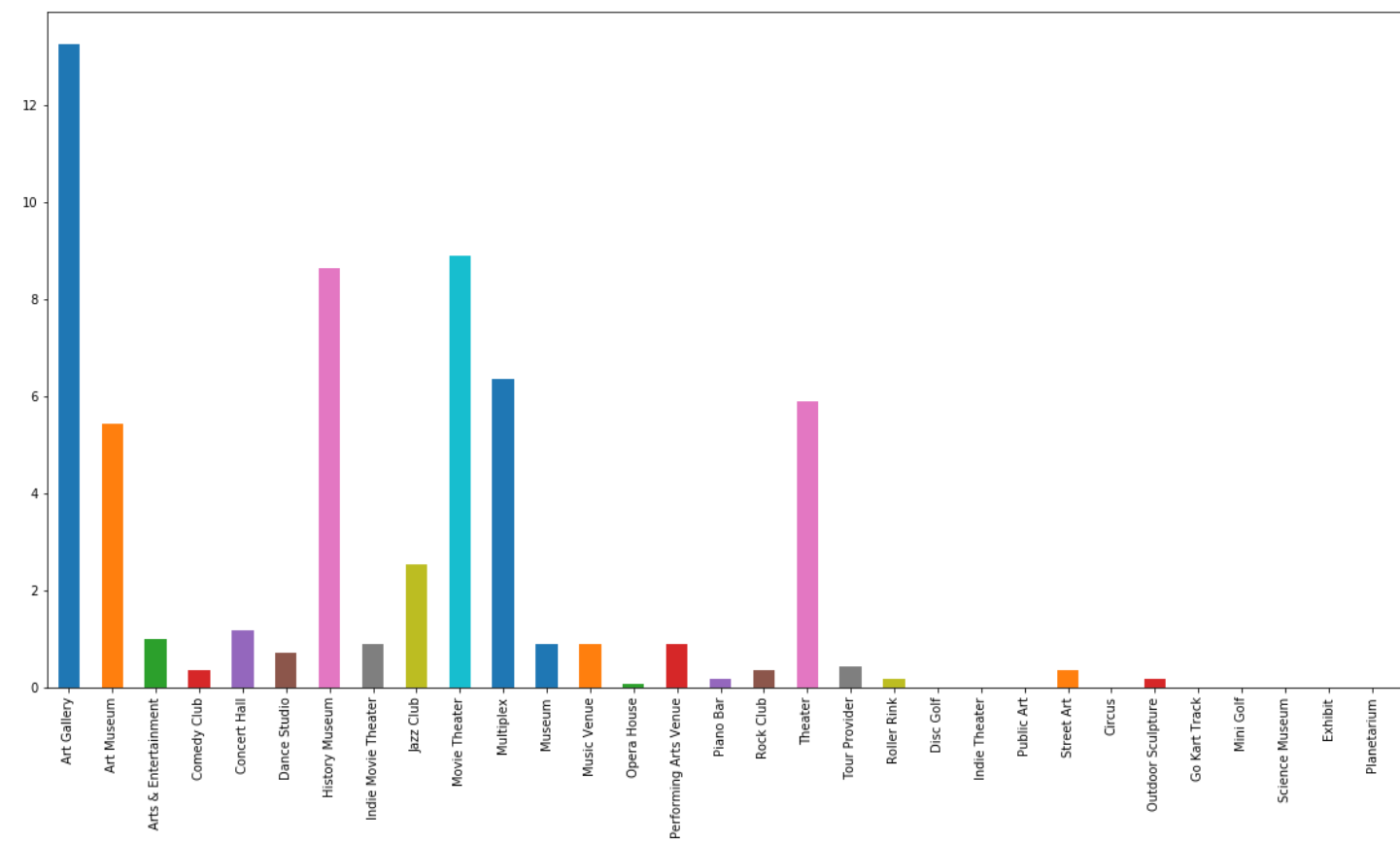
It is hard to get the shop rent price, I use the house price as a indicator how the rent price is in all the 区 in Shanghai.

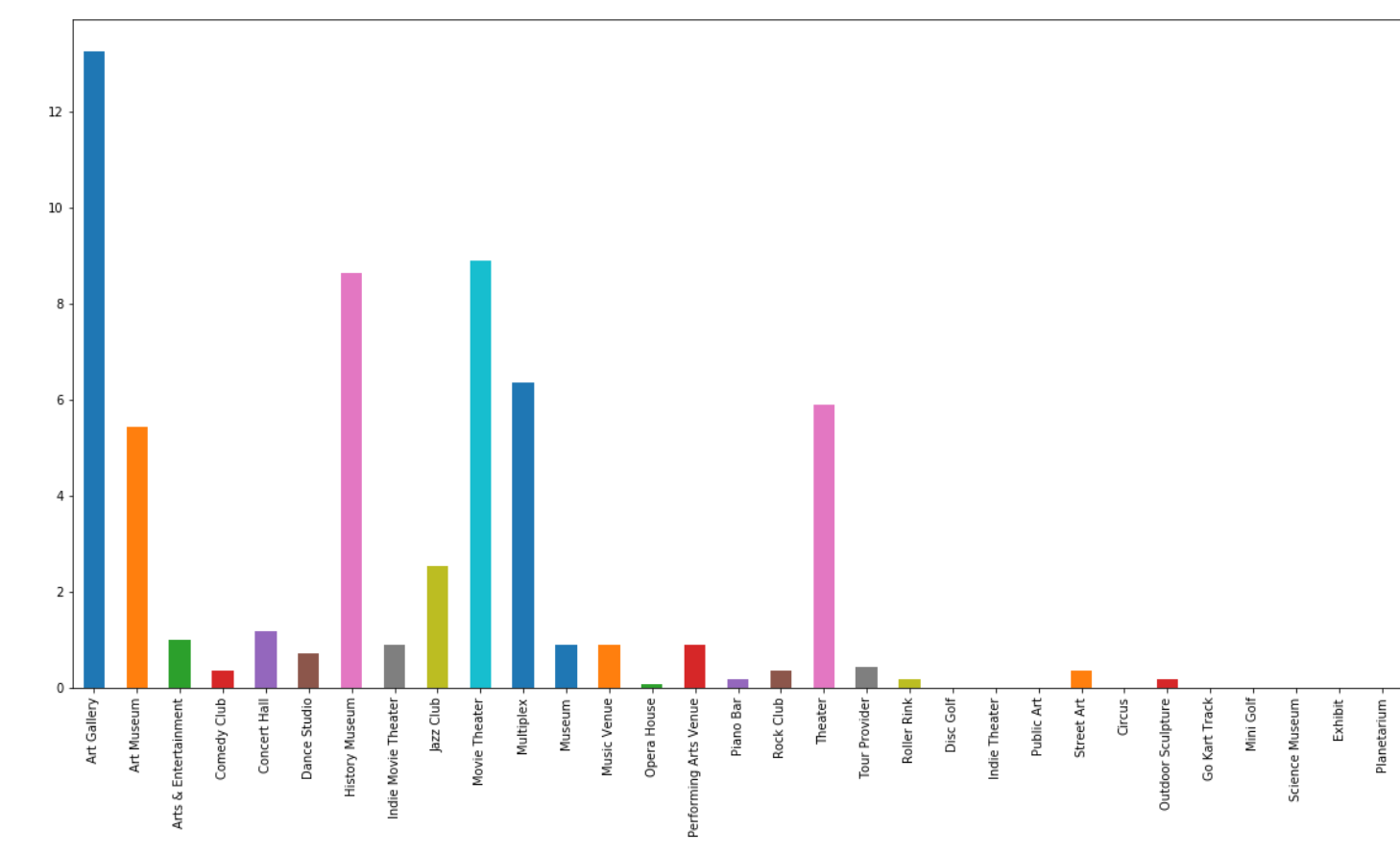
# Methodology

In order to find the 街道 that is the most culture, I use k-means to divide the culture to 5 kinds. Let us check the findings.



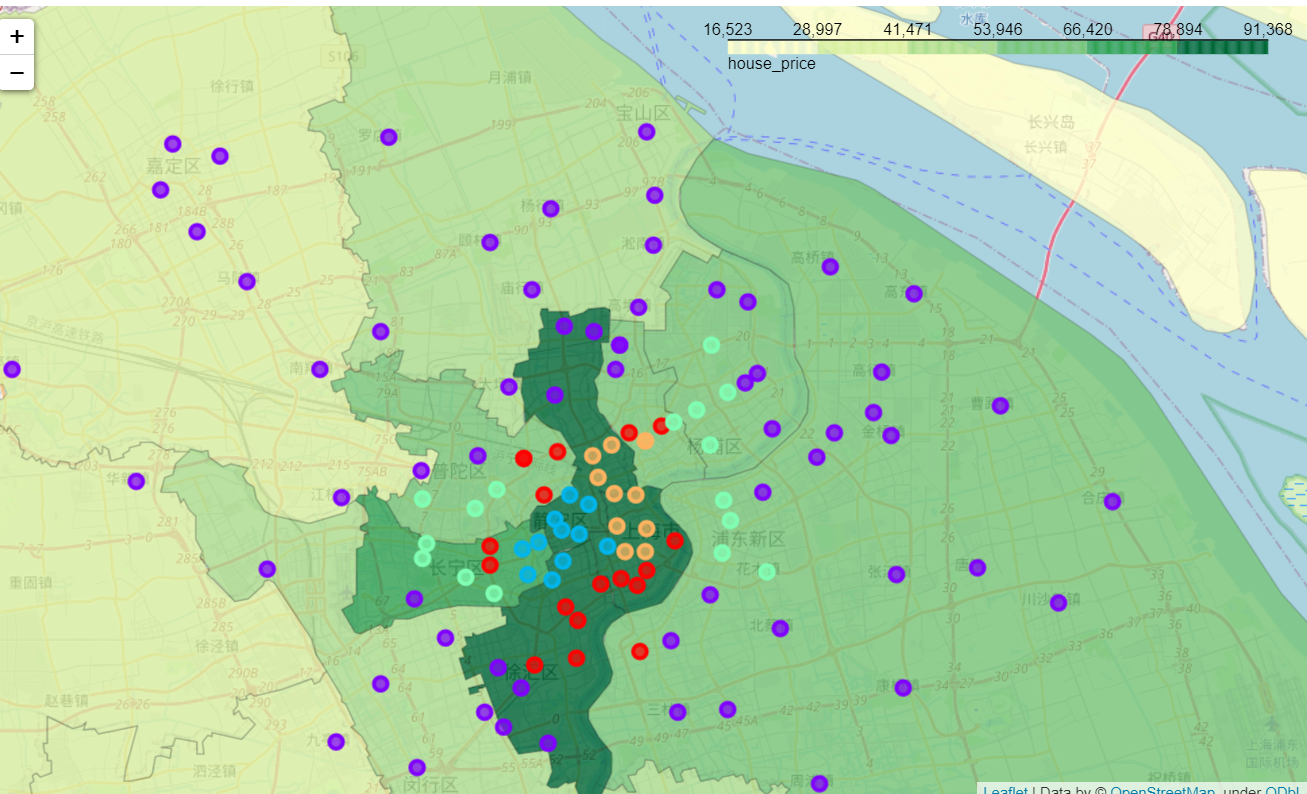
So I plot the bar chart of 5 kinds of 街道 as following. 



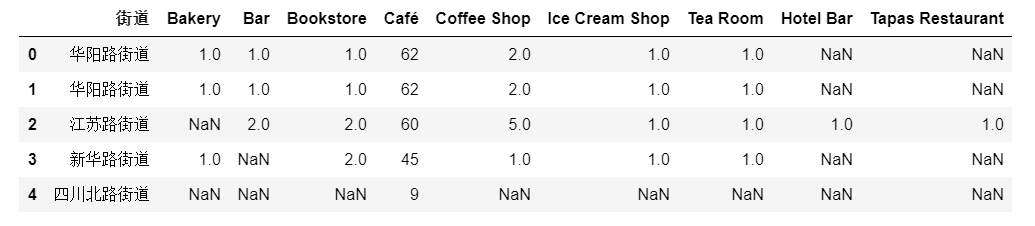
The 2nd and 4th group have the most culture venues nearby. In the center of the city, the 街道 has more culture venues nearby. But I also find that, in some 街道 not in the center city has some culture venues.

In order to find which 区has the most suitable rent place, I create place the 街道to a shanghai house price heat map:



# Result section

Base on the findings in the above, We can find the following 街道names: "四川北路街道","江苏路街道","华阳路街道","新华路街道". The four 街道 locate in the 区which house price is not so high and the nearby culture venues are plentiful. And then, I find the cafe about the four 街道.



Base on this, I recommend the 四川北路街道 to my friend to start his business.

# Discussion

I use the 街道 names in Shanghai as the initial start point to analyze the venues. That is because I can’t get the enough information about the shopping centers in Shanghai. It may be more acuity if I can start from the shopping centers. Second, about the shop rent price, I can’t get it, so I use the house price instead. The third one is that, if we can know how many shoppers around the 街道. I suggest my friend to go the places he’s interested in to find out the true shopper numbers.

Although the data I collect has

# Conclusion

In the city like Shanghai, the city life expand to the outlier of the city. But the culture life and the culture venues do not expand as the city expand. Use the geo map like python folium tool, and the location service api like four square, we can vivid demonstrate the life and atmosphere of the city’s district. Like my friend, I use the python skill to help him to find the candidate place to open the café.

## <https://en.wikipedia.org/wiki/Shanghai>

## <https://zh.wikipedia.org/wiki/%E4%B8%8A%E6%B5%B7%E5%B8%82%E4%B9%A1%E7%BA%A7%E4%BB%A5%E4%B8%8A%E8%A1%8C%E6%94%BF%E5%8C%BA%E5%88%97%E8%A1%A8>

## <http://datav.aliyun.com/tools/atlas/#&lat=31.769817845138945&lng=104.29901249999999&zoom=4>