A Guide to Living in Shanghai

7/30/2019

Section 1: Introduction

Shanghai (上海), the most populous city in China as well as one of the most prosperous and modern cities in the world, is a splendid gem carved by generations of Chinese people and city planners with their remarkable wisdom. Over the past hundred years, Shanghai has thrown off shackles of pain and indignity imposed by feudalists, warlords and imperialists, and has gradually transitioned itself into the economic and financial hub in the world. Known as "Pearl of the Orient" (东方之珠) or "Paris of the East" (东方巴黎), Shanghai at present days has attracted millions of adventurous people from the globe to study, tour, and live in this city.

Nevertheless, for people—especially "Westerners"—who are strangers to both Shanghai and mainland China at large, starting a new life on this mysterious piece of land is difficult. For them, questions like which neighborhood to live in, basic statistics of each neighborhood, where prominent venues are located at, among others, remain unanswered.

There is a growing body of articles and blogposts on the Internet providing rudimentary city-wise information to the audience who is unfamiliar with the nuts and bolts of Shanghai. Nevertheless, they are mainly qualitative and subjective. In essence, they are not quantitatively rigorous in terms of methodology. Therefore, they run the risk of being unreliable, at least from a data analyst's perspective. This guide, namely *A Guide to Living in Shanghai*, is deemed to answer the questions above and to provide an alternative, big data and machine learning-based solution to the potential trailblazers who are about to embark on a new chapter of life in Shanghai.

Section 2: Data

The project uses two datasets: *shanghai_econ* and *shanghai_data*.

shanghai_econ includes economic and geographic data of all 16 districts (区) in Shanghai. It has 6 variables: *District*, *Population Density*, *Salary*, *Home Price*, *GDP* and *GDP Per Capita*.

District is a character variable. This column lists Shanghai's 16 districts. *Population Density* (person/square kilometer) is density of population in 2017 (data source: Shanghai Statistical Yearbook 2018). *Salary* (RMB/month) stands for personal monthly salary in 2019 (data source: Sohu). *Home Price* (RMB/square meter) is from the same source (Sohu). *GDP* (billion RMB) column has district level GDP in 2018 (data source: Sohu). Finally, *GDP Per Capita* (RMB) (year 2017 data) is sourced from Haojingui Finance.

Note that all data are from different sources and are in different years. This is because it is extremely hard to find one single, reliable source providing related district-level data, which turns out to be the most severe fallacy of this project. Nevertheless, it is believed that this would not affect the results too greatly given that the data usually does not change too much within 1 to 2 years if its accuracy can be fully guaranteed.

The full dataset of *shanghai_econ* is:

Table 1: District Level Economic and Demographic Information

	District	Population Density	Salary	Home Price	GDP	GDP PP
0	Pudong	4567	8170	48713	1046.009	175448
1	Huangpu	32004	7160	81375	227.030	320701
2	Xuhui	19874	7640	71064	167.000	144983
3	Changning	18112	8030	68491	142.800	191305
4	Jing'an	28910	8380	66228	184.700	159550
5	Putuo	23431	7720	55738	100.170	72796
6	Hongkou	34058	7970	58927	83.801	96955
7	Yangpu	21627	7220	59443	184.770	130074
8	Minhang	6836	8030	47381	201.360	88089
9	Baoshan	7494	7910	38860	139.206	56506
10	Jiading	3408	7350	34425	236.270	136218
11	Jinshan	1367	7360	17850	94.231	121515
12	Songjiang	2892	7670	40546	127.960	63518
13	Qingpu	1799	7400	30578	100.120	83071
14	Fengxian	1681	7530	27555	94.231	66758
15	Chongming	586	5940	25007	34.571	47623

shanghai_data lists prominent neighborhoods in Shanghai (in both English and Chinese) as well as the districts they are located in. It has three variables: *District*, *Neighborhood* and *Neighborhood Chinese Name*.

Honestly, "neighborhood" (社区) is essentially a "western" concept and is not used to indicate the same thing in China. In the country, a "neighborhood" is no more than a "residential community" (小区) that comprises residential buildings only rather than a spacious area that has shopping malls, stores, restaurants and attractions (and of course residential communities) in the western world. An equivalent concept in Shanghai is "subdistrict" (街道).

"Subdistrict," however, is still not entirely the same as "neighborhood." For instance, Wujiaochang (五角场) in Yangpu District is a subdistrict (五角场街道) in Shanghai's township-level divisions hierarchy and can be treated as a neighborhood to an extent. In contrast, the famous Xintiandi (新天地), an area full of delicacy, decent food and fashion in Huangpu District, is not a subdistrict but can be regarded as a neighborhood.

The report adopts the western convention and focuses on a total of 47 neighborhoods (subdistricts/towns) across 16 districts in Shanghai. The author built *shanghai_data* based on his own discretion.

The top 10 rows of raw *shanghai_data* is:

Table 2: Prominent Neighborhoods in Shanghai (Top 10 Rows)

	District	Neighborhood	Neighborhood Chinese Name
0	Pudong	Lujiazui	陆家嘴
1	Pudong	Century Park	世纪公园
2	Pudong	Zhoujiadu	周家渡
3	Pudong	Zhangjiang	张江
4	Huangpu	People's Square	人民广场
5	Huangpu	Huaihai Road	淮海路
6	Huangpu	The Bund	外滩
7	Huangpu	Former French Concession	旧法租界
8	Huangpu	Xintiandi	新天地
9	Huangpu	Dapuqiao	打浦桥

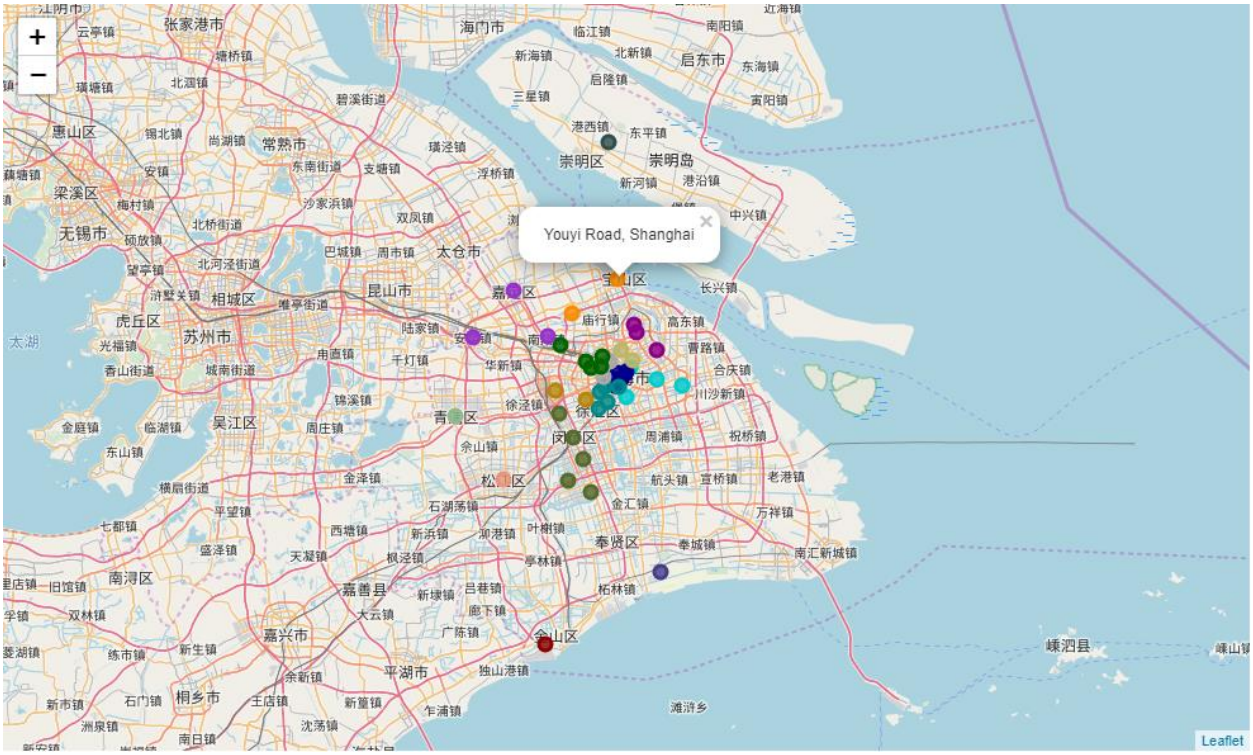
shanghai_data does not have longitude/latitude information for each neighborhood. Thus, further manipulation with the help of Python's *geopy* library is conducted. The resulting data frame with longitude/latitude information (top 10 rows) is:

Table 3: Prominent Neighborhoods in Shanghai (with Geographic Information)

	District	Neighborhood	Neighborhood Chinese Name	Latitude	Longitude
0	Pudong	Lujiazui, Shanghai	陆家嘴	31.240168	121.497945
1	Pudong	Century Park, Shanghai	世纪公园	31.218700	121.554338
2	Pudong	Zhoujiadu, Shanghai	周家渡	31.187146	121.489237
3	Pudong	Zhangjiang, Shanghai	张江	31.207347	121.610182
4	Huangpu	People's Square, Shanghai	人民广场	31.231926	121.471535
5	Huangpu	Huaihai Road, Shanghai	淮海路	31.220936	121.467353
6	Huangpu	The Bund, Shanghai	外滩	31.234038	121.488921
7	Huangpu	Former French Concession, Shanghai	旧法租界	31.211806	121.464982
8	Huangpu	Xintiandi, Shanghai	新天地	31.217936	121.469819
9	Huangpu	Dapuqiao, Shanghai	打浦桥	31.208286	121.463941

With the geographic information of all neighborhoods in Shanghai at handy, we can create a map of Shanghai with all 47 prominent neighborhoods highlighted.

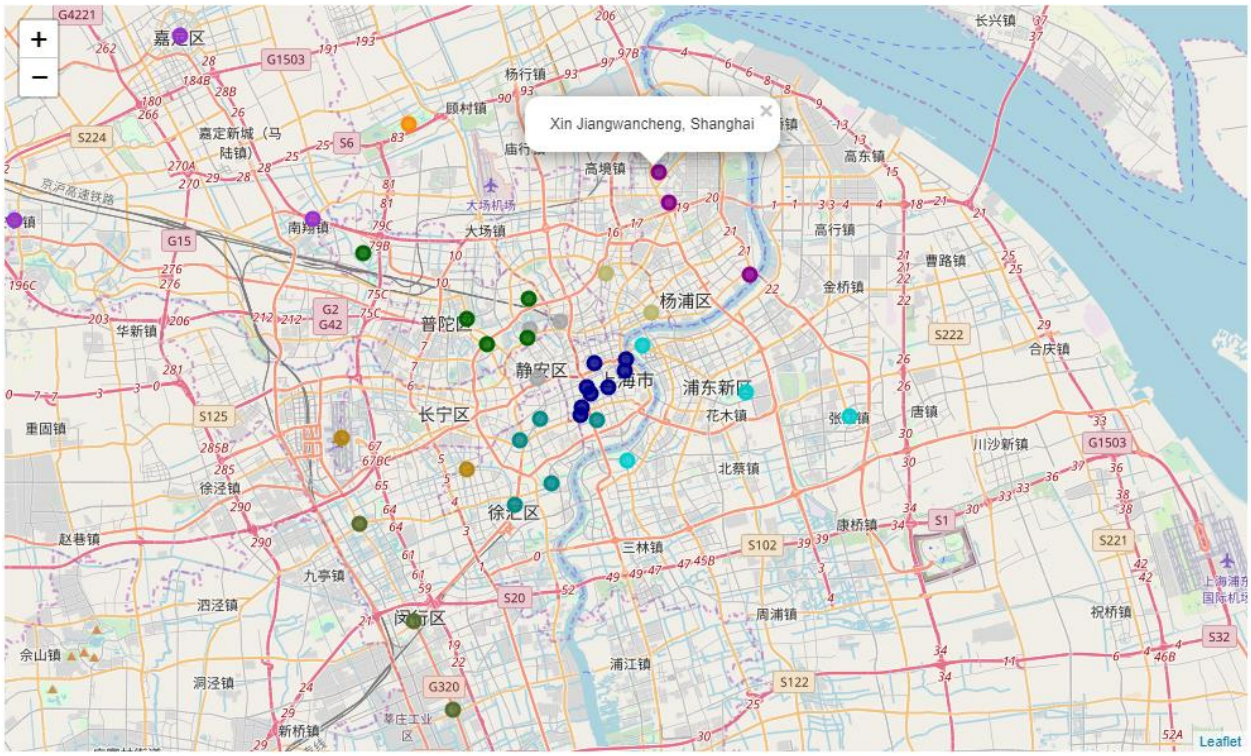
Figure 1: All Neighborhoods in Shanghai



From Figure 1, it is clear that Shanghai is huge. 47 neighborhoods ar at different corners of Shanghai (the northernmost point is Jianshe Zhen in Chongming District, and the southernmost point is Jinshanwei in Jinshan District). In Jupyter Notebook, the map is interactive and includes pop-ups (in Figure 1, Youyi Road in Baoshan District is highlighted). Furthermore, all neighborhoods are colored by their corresponding districts. Same color indicates that neighborhoods are in the same district.

Shanghai’s breadth and reach make itself astonishing, but it also is not a bad idea to have a glance at Shanghai Proper (上海市区), where its beauty is hidden. Shanghai Proper comprises 7 districts: Huangpu, Xuhui, Changning, Jing’an, Putuo, Hongkou, and Yangpu. Figure 2 shows that most prominent neighborhoods are in fact located in Shanghai Proper (especially in Huangpu District).

Figure 2: All Neighborhoods in Shanghai (Zoomed In)



Section 3: Methodology

In order to answer questions that interest tourists and newcomers to the city, methods such as exploratory data analysis and machine learning (*k-means* clustering) are used.

3.1 Exploratory Data Analysis

The first step is to use Foursquare API to explore neighborhoods in Shanghai. Foursquare API enables developers to get trending venues around a location (in this case, around a neighborhood in Shanghai). This answers the key questions of newcomers: what venues can I access in the neighborhood I live in? Later in this section, various venues’ relative frequency data in each neighborhood would also add value to the machine learning algorithm.

After constructing a URL, setting parameters (number of venues returned by Foursquare API is 200, radius is 500 meters), and sending a request to the API, the following results are returned:

Table 4: Venues Information within each Neighborhood (Top 5 Rows)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	District	Neighborhood2
0	Lujiazui, Shanghai	31.240168	121.497945	The Ritz-Carlton Shanghai, Pudong (上海浦东丽思卡尔顿酒店)	31.238699	121.496778	Hotel	Pudong	Lujiazui
1	Lujiazui, Shanghai	31.240168	121.497945	Flair	31.238750	121.496684	Hotel Bar	Pudong	Lujiazui
2	Lujiazui, Shanghai	31.240168	121.497945	Apple Pudong (Apple 浦东)	31.239282	121.497060	Electronics Store	Pudong	Lujiazui
3	Lujiazui, Shanghai	31.240168	121.497945	IFC Mall (国际金融中心商场)	31.238492	121.497902	Shopping Mall	Pudong	Lujiazui
4	Lujiazui, Shanghai	31.240168	121.497945	Oriental Tower-Space Module 350m (太空舱)	31.240493	121.495593	Scenic Lookout	Pudong	Lujiazui

Foursquare API returns recorded venues (upper limit is 200 venues) within the 500-meter-radius of a certain location. For instance, Table 4 shows 5 venues that are within 500-meter-radious of Lujiazui neighborhood in Pudong New District.

Notice that venues’ geographic information is also generated by Foursquare API. Therefore, we can create a venue map, as is put below. Again, all venues are colored by the corresponding districts they are affiliated with.

Figure 3: All Venues in Shanghai

Next, with venues data, we can check top 5 neighborhoods that have the highest number of venues as well as top 5 neighborhoods that have the lowest number of venues. Unsurprisingly, the most touristy neighborhoods are renowned Huaihai Road (淮海路), Xujiahui (徐家汇), People’s Square (人民广场), Yu Garden (豫园) and Lujiazui (陆家嘴). These are places where normal Shanghainese people hangout during their free time. On the other hand, for newcomers, it is advised to avoid living in the neighborhoods that have very few venues.

Table 5: Top 5 Neighborhoods Having the Highest Number of Venues

Neighborhood	Number of Venues
Huaihai Road, Shanghai	88
Xujiahui, Shanghai	61
People's Square, Shanghai	49
Yu Garden, Shanghai	47
Lujiazui, Shanghai	45

Table 6: Top 5 Neighborhoods Having the Lowest Number of Venues

Neighborhood	Number of Venues
Songjiang, Shanghai	2
Caohejing, Shanghai	2
Haiwan Zhen, Shanghai	1
Century Park, Shanghai	1
Maqiao, Shanghai	1

In addition, what are the venue categories a person most/least frequently finds in Shanghai? Table 7 and Table 8 are aimed to answer this question.

Table 7: Most Popular Venue Types

Venue Category	Number of Venues
Coffee Shop	63
Chinese Restaurant	50
Café	38
Hotel	37
Fast Food Restaurant	32

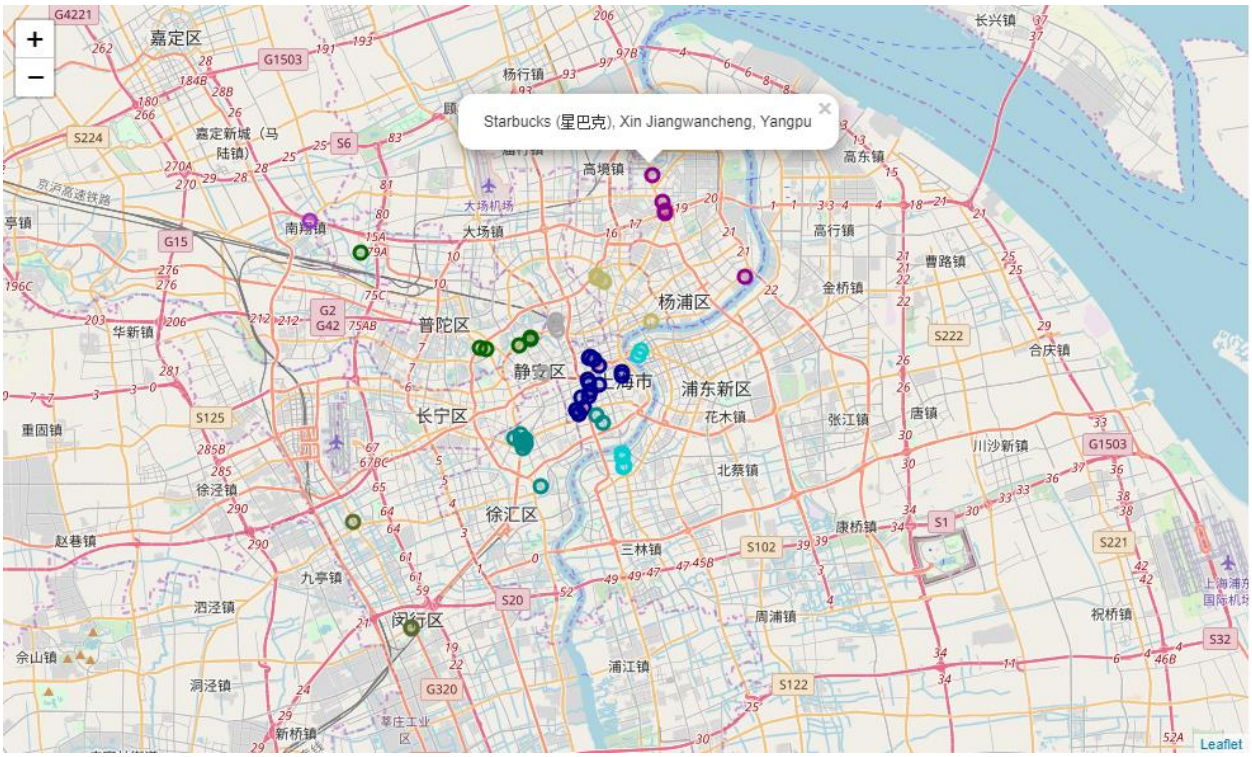
Table 8: Least Popular Venue Types

Venue Category	Number of Venues
Frozen Yogurt Shop	1
Shopping Plaza	1
Skating Rink	1
Soccer Field	1
Zhejiang Restaurant	1

The outside world usually regards Shanghainese people enjoying a “yuppie” (小资) lifestyle. The big data technique seems to have proved this point. Among all venue types, coffee shops (including café) are most commonly seen in the city. Why not take your time and chill in one of the coffee shops on Hengshan Road on a warm Sunday afternoon?

Hengshan Road Neighborhood in Xuhui District (Source: Internet)

Figure 4: All Coffee Shops in Shanghai



Other destinations such as shopping malls and parks are also visualized.

Figure 5: All Shopping Malls in Shanghai

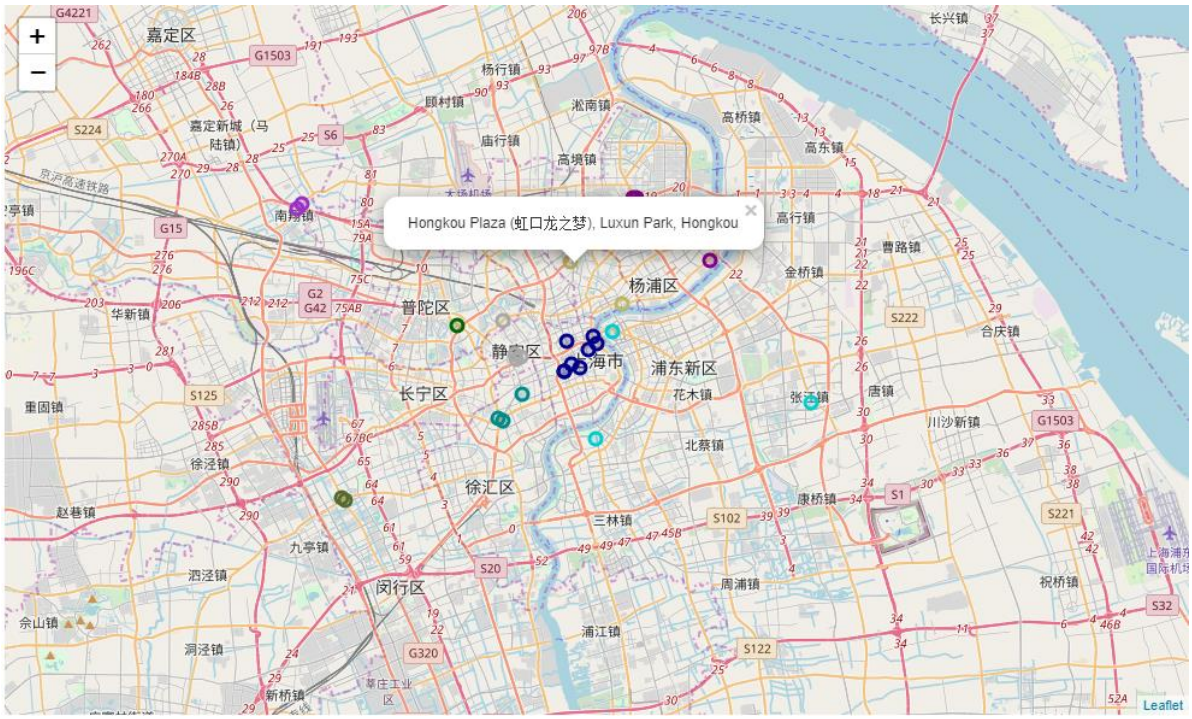
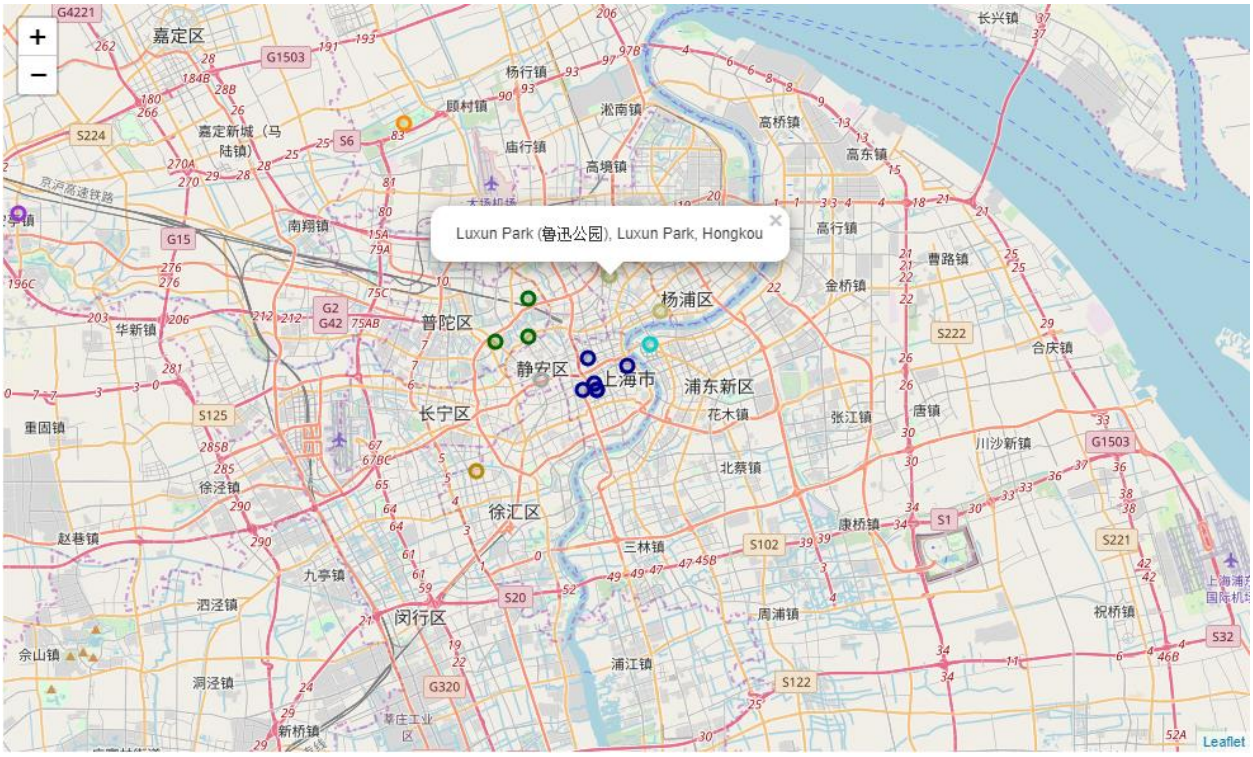


Figure 6: All Parks in Shanghai



The final part of exploratory data analysis is the most popular venue type within each neighborhood. To get the result, we transform the venue information table (Table 4) from long form to wide form. Table 9 is the new data frame (wide form) that has relative frequency of each venue type for each neighborhood. For instance, 0.074074 at the intersection of the 5th row and *Bakery* column means that of all venue types in Changshou Road neighborhood, about 7.4% are bakeries. Table 10 is the table of most popular venue category within each neighborhood.

Table 9: Relative Frequency of Each Venue Type for All Neighborhoods

	Neighbourhood	Airport Service	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bakery	Bar	...	Vegetarian / Vegan Restaurant	Veterinarian	Video Store	Vietnamese Restaurant	Whisky Bar
0	Anting, Shanghai	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
1	Caohejing, Shanghai	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
2	Caoyang, Shanghai	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
3	Century Park, Shanghai	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
4	Changshou Road, Shanghai	0.0	0.0	0.0	0.0	0.0	0.037037	0.0	0.074074	0.0	...	0.0	0.0	0.0	0.0	0.0

Table 10: Most Popular Venue Type within Each Neighborhood (Top 20 Rows)

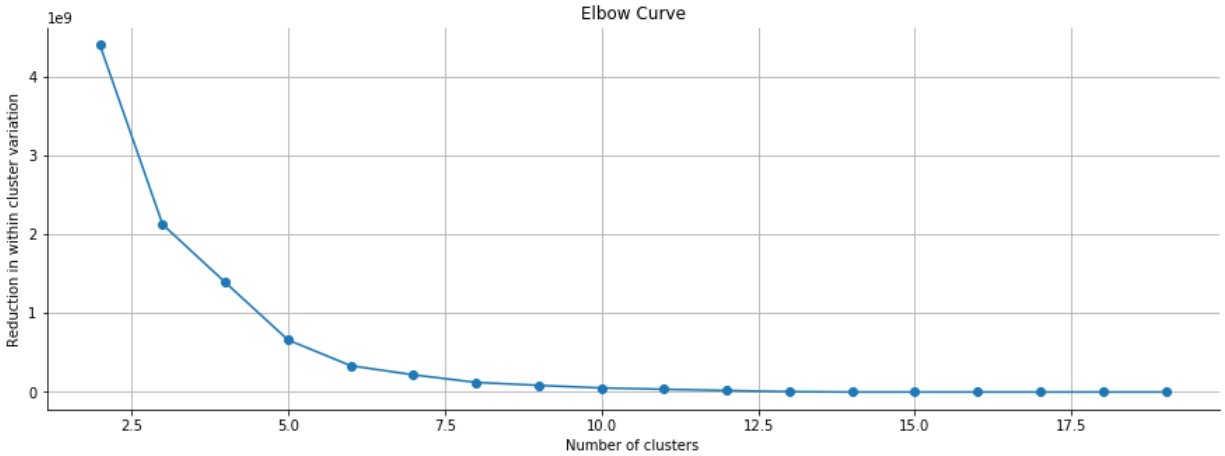
	Neighbourhood	Neighborhood Chinese Name	Most Popular Venue
0	Anting, Shanghai	安亭	Chinese Restaurant
1	Caohejing, Shanghai	漕河泾	Convenience Store
2	Caoyang, Shanghai	曹杨	Coffee Shop
3	Century Park, Shanghai	世纪公园	Chinese Restaurant
4	Changshou Road, Shanghai	长寿路	Chinese Restaurant
5	Dapuqiao, Shanghai	打浦桥	Café
6	Dinghai Road, Shanghai	定海路	Coffee Shop
7	Former French Concession, Shanghai	旧法租界	Café
8	Gubei, Shanghai	古北	Japanese Restaurant
9	Gucun Park, Shanghai	顾村公园	Convenience Store
10	Haiwan Zhen, Shanghai	海湾镇	Hotel
11	Hengshan Road, Shanghai	衡山路	Bar
12	Hongqiao, Shanghai	虹桥	Airport Service
13	Huaihai Road, Shanghai	淮海路	Café
14	Jiading, Shanghai	嘉定	Bakery
15	Jiangchuan Road, Shanghai	江川路	Asian Restaurant
16	Jiangning Road, Shanghai	江宁路	Coffee Shop
17	Jing'an Temple, Shanghai	静安寺	Hotel
18	Laoximen, Shanghai	老西门	Noodle House
19	Longhua, Shanghai	龙华	Cantonese Restaurant

3.2 Machine Learning

We use machine learning to categorize all 47 neighborhoods in Shanghai. The dependent variable is GDP per capita, and all remaining demographic and economic variables as well as relative frequency of various venue types are independent variables used for predicting clusters.

First, the elbow curve in Figure 7 tells that 5 is the optimal number of clusters.

Figure 7: Elbow Curve

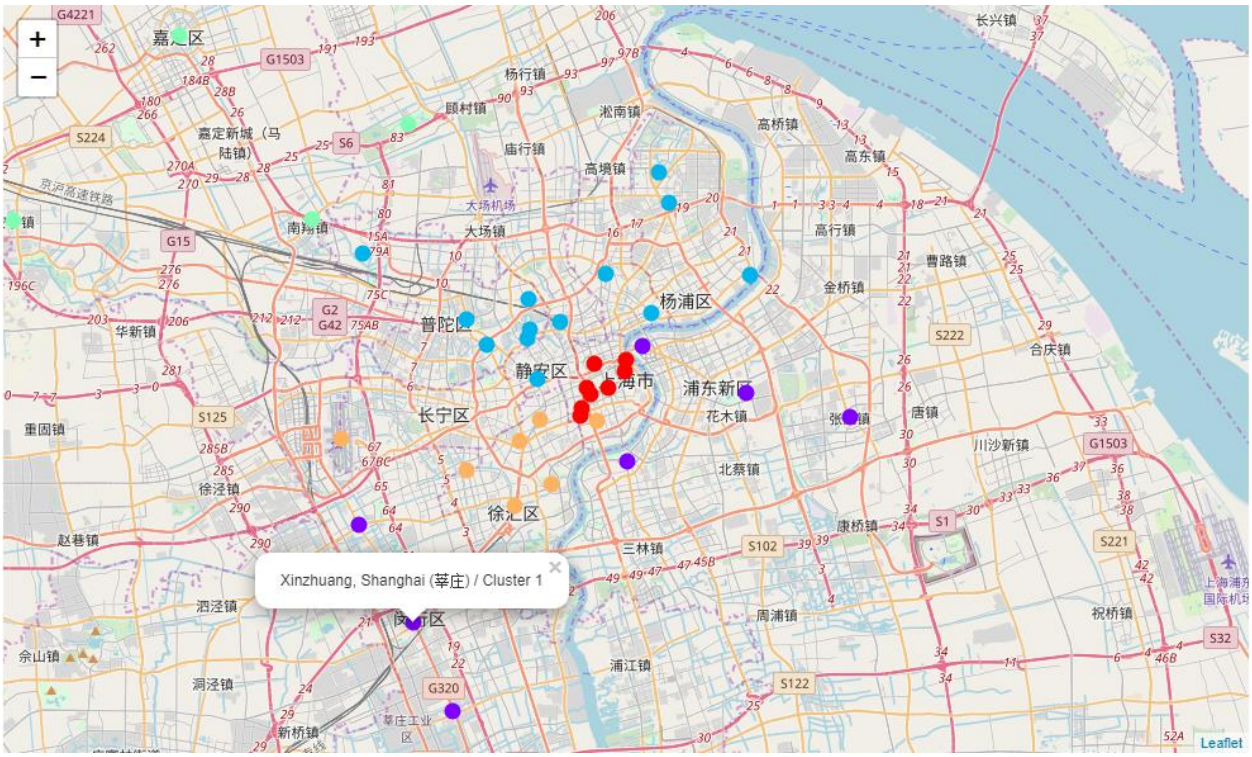


Then, we fit a *k-means* clustering model with 5 as the number of clusters. The results are illustrated in the next section.

Section 4: Results

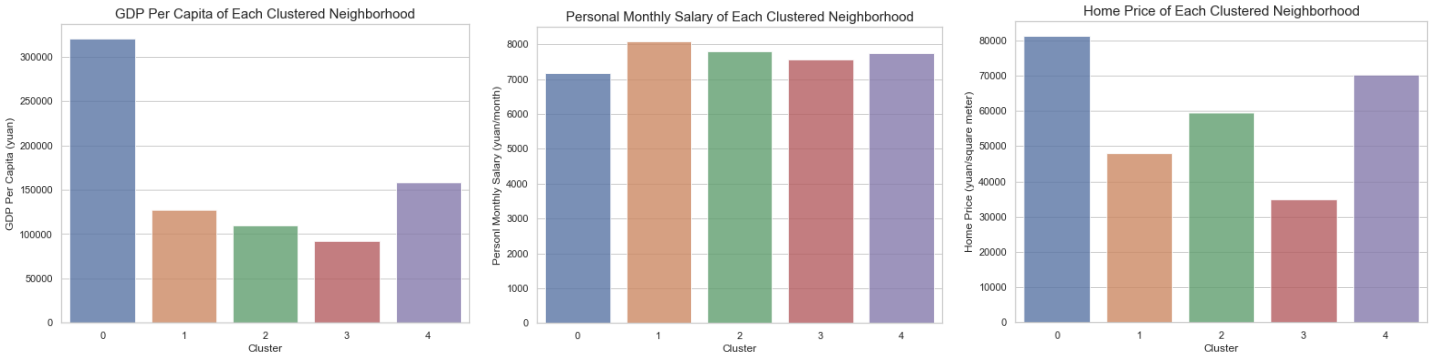
Now that all neighborhoods are assigned specific clusters, we create a new map of neighborhoods colored by clusters.

Figure 8: All Neighborhoods in Shanghai (Clustered)



The results are expected and intuitive. The very central part of Shanghai (Huangpu District, red color) can be interpreted as the center cluster. It is Shanghai’s political and economic center surrounded by the rest of four clusters. As one can imagine, the farther the cluster is away from the center, the more scattered and less developed it is. Therefore, for newcomers, they are recommended to tour in the center cluster (red, orange and blue clusters). The pros is that neighborhoods in these clusters are full of attractions and decent dining places, and are “richer” to an extent. The cons, on the flip side, is that it might be costly to live. The bar charts below help illustrate this point. For instance, Cluster 0/Huangpu District (i.e. red-color cluster in Figure 8) has the highest GDP per capita. Meanwhile, it has the highest home price among all five clusters (and surprisingly, lowest monthly salary), meaning that it would be extremely hard for a starter to afford the cost of daily living in this cluster.

Figure 9: Bar Charts



Section 5: Discussion and Conclusion

From the exploratory data analysis part in Section 4, one can easily locate her favorite venues in Shanghai and have a rough idea on which neighborhoods are touristy. The machine learning part, on the other hand, mainly suggests which district is suitable for a newcomer to live in.

We already know that Cluster 0/Huangpu District (i.e. red-color cluster) is unsuitable. Based on the statistics, it turns out that Cluster 1/Minhang District and Pudong New District (i.e. purple-color cluster) might be an appropriate option.

Table 11: Statistics of Cluster 1/Minhang District and Pudong New District

```
statistics(shanghai_merged,1)
```

```
Average GDP PP of this cluster is 126915.2 yuan.
Average personal monthly salary of this cluster is 8092.2 yuan/month.
Average home price of this cluster is 47973.0 yuan/square meter.
```

	Neighbourhood	Neighborhood Chinese Name	District
3	Century Park, Shanghai	世纪公园	Pudong
15	Jiangchuan Road, Shanghai	江川路	Minhang
20	Lujiazui, Shanghai	陆家嘴	Pudong
22	Maqiao, Shanghai	马桥	Minhang
25	Qibao Zhen, Shanghai	七宝	Minhang
36	Xinzhuang, Shanghai	莘庄	Minhang
41	Zhangjiang, Shanghai	张江	Pudong
43	Zhoujiadu, Shanghai	周家渡	Pudong
44	Zhuanqiao, Shanghai	航桥	Minhang

This area is not rich in terms of GDP per capita ranking, but people in this cluster tend to earn the most and have relatively less stress to live. Honestly, it is not a big deal for newcomers who want to live in this cluster while having fun during the weekend. The reason is: public transportation is convenient in Shanghai; say if a person lives in Minhang District (purple points at bottom left in Figure 8), she is able to arrive at the red cluster (Cluster 0/Huangpu District) in just 30 minutes by taking Shanghai Metro Line 1 (上海地铁 1 号线), or arrive at Cluster 4/Xuhui and Changning District (i.e. orange-color cluster in Figure 8) in just 15 minutes by taking the same metro line.

As the report is about to conclude, it is essential to mention potential problems of this project. First, like has already been discussed in the Data section, the economic data of each district is not consistent; they are not in the same year and are not from the same source. Unfortunately, only *Population Density* variable is from the authority.

Second, venues in Shanghai generated by Foursquare API are problematic and are limited in numbers. In plain words, it only returns renowned venues (e.g. KFC and McDonald's for fast food category, and Starbucks for café category). It fails to return less renowned or "Eastern" venues. For example, it is impossible that there is only Maqiao Stadium and nothing else in Maqiao neighborhood, Minhang District. This weird finding is probably due to the reason that remote areas in Shanghai has never been a focus of Foursquare API, and it fails to take into account location information of many less well-known venues.

Third, double counting. Foursquare API is unable to aggregate venues belonging to the same category. Take findings in Table 7 as an example. Coffee shops and café are essentially the same; it is inconceivable to separate a coffeehouse into two distinct types. Similarly, Zhejiang Restaurant in Table 8 is a subset of the larger group Chinese Restaurant in Table 7. Consequently, it should not be independent from the latter.

To conclude, this guide presents venues information for each neighborhood across Shanghai, categorizes all prominent neighborhoods, and visualizes them in the map of Shanghai on a big data basis. The guide also recommends newcomers to begin their new lives in either Minhang District or Pudong New District.