

# **Coursera Capstone**

## **IBM Data Science Course Final Report**

### **Opening A New Restaurant in Shanghai, China**

By: Jeffrey Hu

July 2020

#### **Background Introduction**

As we all know, opening a restaurant of your own is one of the biggest dream for many entrepreneurs. However, as with any business decision, it may require serious consideration and is much complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether it will be a success or not.

#### **Business Problem**

The aim of this capstone project is to analyse and choose the best location in the city of Shanghai, China to open a new restaurant. Using data science methodology and machine learning techniques like clustering, the project's purpose is to provide a solution to answer the question: In the city of Shanghai, China, if someone is looking to open a restaurant, where would you recommend that they open it?

#### **Data Section**

**The following data will be needed for us to solve the problem:**

- \* List of administrative divisions of Shanghai: this defines the scope of this project which is confined to the city of Shanghai, one the of biggest city in East Asia.
- \* Latitude and longitude coordinate of these administrative divisions, which is needed to plot the map and get the venue data.
- \* Venue data, especially data related to restaurants. We will use them to perform clustering algorithm on the neighborhoods to find the best location.

## **Source of data and methods to extract them**

The wikipedia page <[https://en.wikipedia.org/wiki/List\\_of\\_administrative\\_divisions\\_of\\_Shanghai](https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Shanghai)> contains a list of administrative divisions of Shanghai, with a total of 16 administrative divisions. We will use web scraping techniques to extract data from the wikipedia page, with the help of Python requests and other library packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 100+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data and we will be particularly interested in the restaurant category in order to help us to solve the business problem we put forward.

This is a project that will make use of many data science skills, from web scraping, working with API(Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization(Folium). In the next section, we will present the Methodology part where we will discuss the steps taken in this project, the data analysis and the machine learning techniques we've used.

## **Methodology**

Firstly, we need to get the list of administrative divisions in the city of Shanghai. Through wikipedia page <[https://en.wikipedia.org/wiki/List\\_of\\_administrative\\_divisions\\_of\\_Shanghai](https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Shanghai)>, we will do web scraping using Python requests and libraries to extract the list of neighborhoods data. However, this is just a list of names. We need to get the geographical coordinate in the form of latitude and longitude in order to be able to use Foursquare API. Through Geocoder package, we convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Shanghai.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, category, latitude and longitude. With the data, we can check how many venues were returned for each neighborhoods and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighborhood by grouping the rows and taking the mean of frequency of occurrence of each venue category.

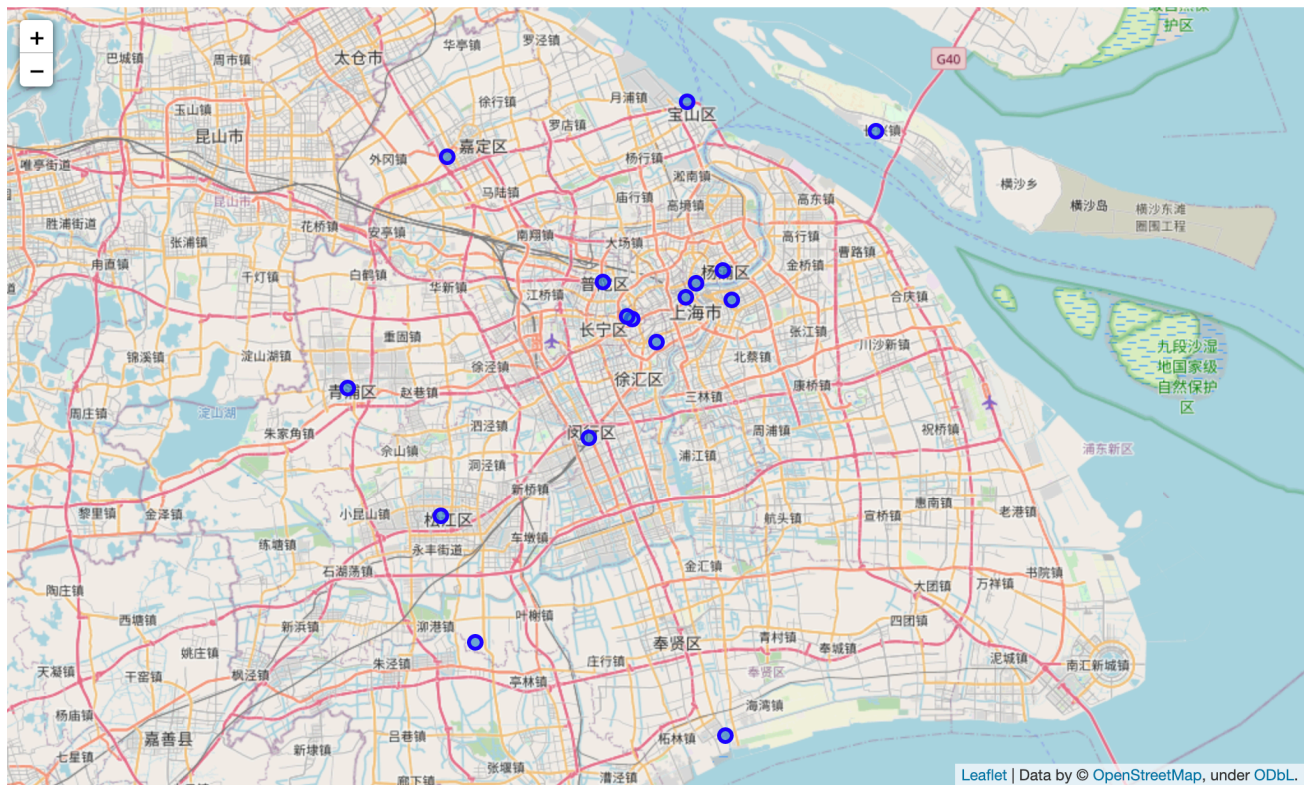
Finally, we will analyse the result and give some useful recommendations for opening a restaurant in Shanghai based on their frequency of occurrence. According to their different frequency rate in different neighborhoods, it will help us to answer the question as to which place is most suitable to open new restaurants.

## Results

The results can be shown on the following frequency chart.

<p>----Baoshan District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Park</td><td>0.4</td></tr> <tr><td>1</td><td>Fast Food Restaurant</td><td>0.2</td></tr> <tr><td>2</td><td>Boat or Ferry</td><td>0.2</td></tr> <tr><td>3</td><td>Shopping Mall</td><td>0.2</td></tr> <tr><td>4</td><td>American Restaurant</td><td>0.0</td></tr> </table>		venue	freq	0	Park	0.4	1	Fast Food Restaurant	0.2	2	Boat or Ferry	0.2	3	Shopping Mall	0.2	4	American Restaurant	0.0	<p>----Huangpu District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Hotel</td><td>0.20</td></tr> <tr><td>1</td><td>Lounge</td><td>0.06</td></tr> <tr><td>2</td><td>French Restaurant</td><td>0.05</td></tr> <tr><td>3</td><td>Café</td><td>0.04</td></tr> <tr><td>4</td><td>Italian Restaurant</td><td>0.04</td></tr> </table>		venue	freq	0	Hotel	0.20	1	Lounge	0.06	2	French Restaurant	0.05	3	Café	0.04	4	Italian Restaurant	0.04	<p>----Putuo District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Fast Food Restaurant</td><td>0.16</td></tr> <tr><td>1</td><td>Metro Station</td><td>0.11</td></tr> <tr><td>2</td><td>Hotel</td><td>0.11</td></tr> <tr><td>3</td><td>Coffee Shop</td><td>0.11</td></tr> <tr><td>4</td><td>Furniture / Home Store</td><td>0.05</td></tr> </table>		venue	freq	0	Fast Food Restaurant	0.16	1	Metro Station	0.11	2	Hotel	0.11	3	Coffee Shop	0.11	4	Furniture / Home Store	0.05
	venue	freq																																																						
0	Park	0.4																																																						
1	Fast Food Restaurant	0.2																																																						
2	Boat or Ferry	0.2																																																						
3	Shopping Mall	0.2																																																						
4	American Restaurant	0.0																																																						
	venue	freq																																																						
0	Hotel	0.20																																																						
1	Lounge	0.06																																																						
2	French Restaurant	0.05																																																						
3	Café	0.04																																																						
4	Italian Restaurant	0.04																																																						
	venue	freq																																																						
0	Fast Food Restaurant	0.16																																																						
1	Metro Station	0.11																																																						
2	Hotel	0.11																																																						
3	Coffee Shop	0.11																																																						
4	Furniture / Home Store	0.05																																																						
<p>----Changning District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Chinese Restaurant</td><td>0.08</td></tr> <tr><td>1</td><td>Hotel</td><td>0.05</td></tr> <tr><td>2</td><td>Bakery</td><td>0.05</td></tr> <tr><td>3</td><td>Coffee Shop</td><td>0.05</td></tr> <tr><td>4</td><td>Japanese Restaurant</td><td>0.05</td></tr> </table>		venue	freq	0	Chinese Restaurant	0.08	1	Hotel	0.05	2	Bakery	0.05	3	Coffee Shop	0.05	4	Japanese Restaurant	0.05	<p>----Jiading District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Bus Station</td><td>0.17</td></tr> <tr><td>1</td><td>Metro Station</td><td>0.17</td></tr> <tr><td>2</td><td>Hotel</td><td>0.17</td></tr> <tr><td>3</td><td>Toll Plaza</td><td>0.17</td></tr> <tr><td>4</td><td>Park</td><td>0.17</td></tr> </table>		venue	freq	0	Bus Station	0.17	1	Metro Station	0.17	2	Hotel	0.17	3	Toll Plaza	0.17	4	Park	0.17	<p>----Qingpu District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Coffee Shop</td><td>0.50</td></tr> <tr><td>1</td><td>Hotel</td><td>0.33</td></tr> <tr><td>2</td><td>Department Store</td><td>0.17</td></tr> <tr><td>3</td><td>Pizza Place</td><td>0.00</td></tr> <tr><td>4</td><td>Performing Arts Venue</td><td>0.00</td></tr> </table>		venue	freq	0	Coffee Shop	0.50	1	Hotel	0.33	2	Department Store	0.17	3	Pizza Place	0.00	4	Performing Arts Venue	0.00
	venue	freq																																																						
0	Chinese Restaurant	0.08																																																						
1	Hotel	0.05																																																						
2	Bakery	0.05																																																						
3	Coffee Shop	0.05																																																						
4	Japanese Restaurant	0.05																																																						
	venue	freq																																																						
0	Bus Station	0.17																																																						
1	Metro Station	0.17																																																						
2	Hotel	0.17																																																						
3	Toll Plaza	0.17																																																						
4	Park	0.17																																																						
	venue	freq																																																						
0	Coffee Shop	0.50																																																						
1	Hotel	0.33																																																						
2	Department Store	0.17																																																						
3	Pizza Place	0.00																																																						
4	Performing Arts Venue	0.00																																																						
<p>----Chongming District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Hotel</td><td>0.25</td></tr> <tr><td>1</td><td>Chinese Restaurant</td><td>0.25</td></tr> <tr><td>2</td><td>Bar</td><td>0.25</td></tr> <tr><td>3</td><td>Boat or Ferry</td><td>0.25</td></tr> <tr><td>4</td><td>American Restaurant</td><td>0.00</td></tr> </table>		venue	freq	0	Hotel	0.25	1	Chinese Restaurant	0.25	2	Bar	0.25	3	Boat or Ferry	0.25	4	American Restaurant	0.00	<p>----Jing'an District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Chinese Restaurant</td><td>0.07</td></tr> <tr><td>1</td><td>Coffee Shop</td><td>0.07</td></tr> <tr><td>2</td><td>Bakery</td><td>0.07</td></tr> <tr><td>3</td><td>Hotel</td><td>0.06</td></tr> <tr><td>4</td><td>Japanese Restaurant</td><td>0.05</td></tr> </table>		venue	freq	0	Chinese Restaurant	0.07	1	Coffee Shop	0.07	2	Bakery	0.07	3	Hotel	0.06	4	Japanese Restaurant	0.05	<p>----Songjiang District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Fast Food Restaurant</td><td>0.17</td></tr> <tr><td>1</td><td>Coffee Shop</td><td>0.17</td></tr> <tr><td>2</td><td>Big Box Store</td><td>0.08</td></tr> <tr><td>3</td><td>Chinese Restaurant</td><td>0.08</td></tr> <tr><td>4</td><td>Grocery Store</td><td>0.08</td></tr> </table>		venue	freq	0	Fast Food Restaurant	0.17	1	Coffee Shop	0.17	2	Big Box Store	0.08	3	Chinese Restaurant	0.08	4	Grocery Store	0.08
	venue	freq																																																						
0	Hotel	0.25																																																						
1	Chinese Restaurant	0.25																																																						
2	Bar	0.25																																																						
3	Boat or Ferry	0.25																																																						
4	American Restaurant	0.00																																																						
	venue	freq																																																						
0	Chinese Restaurant	0.07																																																						
1	Coffee Shop	0.07																																																						
2	Bakery	0.07																																																						
3	Hotel	0.06																																																						
4	Japanese Restaurant	0.05																																																						
	venue	freq																																																						
0	Fast Food Restaurant	0.17																																																						
1	Coffee Shop	0.17																																																						
2	Big Box Store	0.08																																																						
3	Chinese Restaurant	0.08																																																						
4	Grocery Store	0.08																																																						
<p>----Fengxian District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Korean Restaurant</td><td>0.25</td></tr> <tr><td>1</td><td>Hotel</td><td>0.25</td></tr> <tr><td>2</td><td>Beach</td><td>0.25</td></tr> <tr><td>3</td><td>Café</td><td>0.25</td></tr> <tr><td>4</td><td>Pier</td><td>0.00</td></tr> </table>		venue	freq	0	Korean Restaurant	0.25	1	Hotel	0.25	2	Beach	0.25	3	Café	0.25	4	Pier	0.00	<p>----Minhang District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Coffee Shop</td><td>0.22</td></tr> <tr><td>1</td><td>Fast Food Restaurant</td><td>0.15</td></tr> <tr><td>2</td><td>Metro Station</td><td>0.07</td></tr> <tr><td>3</td><td>Bakery</td><td>0.07</td></tr> <tr><td>4</td><td>Café</td><td>0.07</td></tr> </table>		venue	freq	0	Coffee Shop	0.22	1	Fast Food Restaurant	0.15	2	Metro Station	0.07	3	Bakery	0.07	4	Café	0.07	<p>----Xuhui District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Coffee Shop</td><td>0.09</td></tr> <tr><td>1</td><td>Chinese Restaurant</td><td>0.07</td></tr> <tr><td>2</td><td>Bakery</td><td>0.05</td></tr> <tr><td>3</td><td>Cocktail Bar</td><td>0.05</td></tr> <tr><td>4</td><td>Japanese Restaurant</td><td>0.03</td></tr> </table>		venue	freq	0	Coffee Shop	0.09	1	Chinese Restaurant	0.07	2	Bakery	0.05	3	Cocktail Bar	0.05	4	Japanese Restaurant	0.03
	venue	freq																																																						
0	Korean Restaurant	0.25																																																						
1	Hotel	0.25																																																						
2	Beach	0.25																																																						
3	Café	0.25																																																						
4	Pier	0.00																																																						
	venue	freq																																																						
0	Coffee Shop	0.22																																																						
1	Fast Food Restaurant	0.15																																																						
2	Metro Station	0.07																																																						
3	Bakery	0.07																																																						
4	Café	0.07																																																						
	venue	freq																																																						
0	Coffee Shop	0.09																																																						
1	Chinese Restaurant	0.07																																																						
2	Bakery	0.05																																																						
3	Cocktail Bar	0.05																																																						
4	Japanese Restaurant	0.03																																																						
<p>----Hongkou District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Hotel</td><td>0.18</td></tr> <tr><td>1</td><td>Lounge</td><td>0.07</td></tr> <tr><td>2</td><td>Restaurant</td><td>0.06</td></tr> <tr><td>3</td><td>Hotel Bar</td><td>0.05</td></tr> <tr><td>4</td><td>Italian Restaurant</td><td>0.05</td></tr> </table>		venue	freq	0	Hotel	0.18	1	Lounge	0.07	2	Restaurant	0.06	3	Hotel Bar	0.05	4	Italian Restaurant	0.05	<p>----Pudong New Area----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Coffee Shop</td><td>0.24</td></tr> <tr><td>1</td><td>Hotel</td><td>0.09</td></tr> <tr><td>2</td><td>Fast Food Restaurant</td><td>0.06</td></tr> <tr><td>3</td><td>Shopping Mall</td><td>0.06</td></tr> <tr><td>4</td><td>Chinese Restaurant</td><td>0.04</td></tr> </table>		venue	freq	0	Coffee Shop	0.24	1	Hotel	0.09	2	Fast Food Restaurant	0.06	3	Shopping Mall	0.06	4	Chinese Restaurant	0.04	<p>----Yangpu District----</p> <table> <tr><th></th><th>venue</th><th>freq</th></tr> <tr><td>0</td><td>Coffee Shop</td><td>0.25</td></tr> <tr><td>1</td><td>Hotel</td><td>0.18</td></tr> <tr><td>2</td><td>Fast Food Restaurant</td><td>0.05</td></tr> <tr><td>3</td><td>Pier</td><td>0.05</td></tr> <tr><td>4</td><td>Chinese Restaurant</td><td>0.05</td></tr> </table>		venue	freq	0	Coffee Shop	0.25	1	Hotel	0.18	2	Fast Food Restaurant	0.05	3	Pier	0.05	4	Chinese Restaurant	0.05
	venue	freq																																																						
0	Hotel	0.18																																																						
1	Lounge	0.07																																																						
2	Restaurant	0.06																																																						
3	Hotel Bar	0.05																																																						
4	Italian Restaurant	0.05																																																						
	venue	freq																																																						
0	Coffee Shop	0.24																																																						
1	Hotel	0.09																																																						
2	Fast Food Restaurant	0.06																																																						
3	Shopping Mall	0.06																																																						
4	Chinese Restaurant	0.04																																																						
	venue	freq																																																						
0	Coffee Shop	0.25																																																						
1	Hotel	0.18																																																						
2	Fast Food Restaurant	0.05																																																						
3	Pier	0.05																																																						
4	Chinese Restaurant	0.05																																																						

	Name	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	Baoshan District	Park	Shopping Mall	Boat or Ferry	Fast Food Restaurant	Zhejiang Restaurant	Garden	Furniture / Home Store	French Restaurant	Food & Drink Shop	Flower Shop
1	Changning District	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Bakery	Hotel	Italian Restaurant	French Restaurant	Lounge	Beer Bar	Café
2	Chongming District	Chinese Restaurant	Boat or Ferry	Hotel	Bar	Zhejiang Restaurant	General Entertainment	Garden	Furniture / Home Store	French Restaurant	Food & Drink Shop
3	Fengxian District	Korean Restaurant	Beach	Hotel	Café	Farmers Market	Fast Food Restaurant	Flea Market	Flower Shop	Gourmet Shop	Electronics Store
4	Hongkou District	Hotel	Lounge	Restaurant	Italian Restaurant	Hotel Bar	French Restaurant	Chinese Restaurant	Shopping Mall	Dumpling Restaurant	Department Store
5	Huangpu District	Hotel	Lounge	French Restaurant	Coffee Shop	Café	Italian Restaurant	Chinese Restaurant	Shopping Mall	Park	Shanghai Restaurant
6	Jiading District	Park	Asian Restaurant	Toll Plaza	Bus Station	Metro Station	Hotel	Zhejiang Restaurant	Food & Drink Shop	General Entertainment	Garden
7	Jing'an District	Bakery	Chinese Restaurant	Coffee Shop	Hotel	Japanese Restaurant	Shopping Mall	Café	French Restaurant	Lounge	Beer Bar
8	Minhang District	Coffee Shop	Fast Food Restaurant	Shopping Mall	Bakery	Café	Metro Station	Zhejiang Restaurant	Supermarket	Clothing Store	Park
9	Pudong New Area	Coffee Shop	Hotel	Shopping Mall	Fast Food Restaurant	Chinese Restaurant	Pizza Place	Sandwich Place	Café	Sushi Restaurant	Breakfast Spot
10	Putuo District	Fast Food Restaurant	Coffee Shop	Hotel	Metro Station	Flea Market	Grocery Store	Pizza Place	Soccer Field	Movie Theater	Bar
11	Qingpu District	Coffee Shop	Hotel	Department Store	Flower Shop	General Entertainment	Garden	Furniture / Home Store	French Restaurant	Food & Drink Shop	Flea Market
12	Songjiang District	Coffee Shop	Fast Food Restaurant	Hotel	Grocery Store	Supermarket	Metro Station	Toll Booth	Chinese Restaurant	Big Box Store	Hockey Arena
13	Xuhui District	Coffee Shop	Chinese Restaurant	Cocktail Bar	Bakery	Park	Café	Hotpot Restaurant	Japanese Restaurant	French Restaurant	Szechuan Restaurant
14	Yangpu District	Coffee Shop	Hotel	Chinese Restaurant	Supermarket	Fast Food Restaurant	Pier	Plaza	Hotel Bar	Flower Shop	Seafood Restaurant



## Discussion and Observations

For the question that we want to answer, "how to open a restaurant in Shanghai" depends on which kind of restaurant you want to choose from.

Base on the principle of **knock-on effect**, we advise you'd better open a restaurant which is located near other similar ones.

### For this reason, we get some conclusions as follows:

1. Qingpu District will be a best choice if you want to open a coffee shop because its frequency is as high as 0.50; Minhang, Yangpu and Pudong New area can also be considered since their frequency is more than 0.20
2. If you want to open a Chinese restaurant, then Changning and Chongming District should be your first consideration since it is the 1st most common venue in this area.
3. Fengxian District is a good choice for Korean style and Putuo District is suitable for Fast Food restaurant.
4. For westerners, Huangpu and Xuhui district can be considered first because they have the highest rate of frequency for western style restaurant, say, French, Italian, Cocktail Bar, etc.

## Conclusion and Suggestions for Further Research

In this project, we only consider administrative divisions of Shanghai and its frequency rate of occurrence. In fact, deeper and more specific sub-districts can be divided into more details, and there are other factors such as population and income of residents that could also influence the location decision of a new restaurant.

However, to the best knowledge of this research, such data are not available to the neighborhood level required by this project. Further research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant.

In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Further research could make use of paid account to bypass these limitations and obtain more results.

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting the preparing the data, performing machine learning by analyzing the data and give some recommendations to the relevant stakeholders, say, property developers and investors regarding the best location to open a new restaurant in Shanghai.

## **References**

List of administrative divisions of Shanghai. Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/List\\_of\\_administrative\\_divisions\\_of\\_Shanghai](https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Shanghai)

Foursquare Developers Documentation. Foursquare. Retrieved from <https://developer.foursquare.com/docs>