

# **Finding the best location to open a restaurant in Williamsburg**

Yuke Luo

2020.04.06

## **Introduction**

As a Chinese student who is studying at College of William & Mary in Williamsburg, VA, I feel this place is extremely lack of Chinese food. Hotpot is a traditional type of food in China; we have different kinds of Hotpot. The most famous one is combining half of spicy soup and half of tomato soup, separately but in one pot, and put anything you like into the boiling soup until it's cooked.

To open a restaurant, one of the most important factors to consider is location. A perfect location helps the restaurant get benefit much more than normal ones. Therefore, this project is to predict and find the best location to open a Hotpot restaurant in Williamsburg. To clarify, the results in this project don't necessarily only serve for a hotpot restaurant, but for any kind of restaurant, because normally we have to compare between Chinese restaurants in certain regions to examine the validation of the results. However, there are too few Chinese restaurants (approximately only one or two in Williamsburg), comparison is not meaningful.

## **Data acquisition**

All data are from Foursquare developer. Luckily, I don't need to find data anywhere else because all information needed can be found on Foursquare. To access data from Foursquare, we need our own id and secret generated from the website.

For data cleaning, since the data is collected by web scraping, there are too much information that is redundant. Therefore, the first thing is to print out what we have and choose the data I need. In this project, most of the data are in the list of venue. The reason that I print the result of web scraping is that sometimes I get the list called "venue", but sometimes it's called "venues", and this makes me

run in mistakes several time. Also, it's lucky that on Foursquare, I won't get any missing value, so I don't need to worry about omit any "NA".

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	cour
0	The Trellis Restaurant	American Restaurant	403 W Duke of Gloucester St	N Henry St	37.271081	-76.706156	[{"label": "display", "lat": 37.27108130431542...	499	23185	US	Williamsburg	VA	Un Sta
1	Berret's Seafood Restaurant and Taphouse Grill	Seafood Restaurant	199 S Boundary St	Merchant's Square	37.270027	-76.707125	[{"label": "display", "lat": 37.27002662899438...	643	23185	US	Williamsburg	VA	Un Sta
2	Peking Restaurant	Japanese Restaurant	120 Waller Mill Rd Ste J	Bypass Rd	37.288596	-76.717956	[{"label": "display", "lat": 37.28859584124185...	2122	23185	US	Williamsburg	VA	Un Sta
3	The Veranda Restaurant (Fort Magruder Hotel)	American Restaurant	NaN	NaN	37.263629	-76.679951	[{"label": "display", "lat": 37.26362902781627...	2296	NaN	US	Williamsburg	VA	Un Sta
4	The Jefferson Restaurant	American Restaurant	1453 Richmond Rd	NaN	37.289970	-76.722190	[{"label": "display", "lat": 37.28997, "lng": ...	2491	23185	US	Williamsburg	VA	Un Sta

The graph above shows the first data frame including names of restaurants, categories, address, coordinates and so on. These are very important information I need for later data processing. This data frame mainly provides me a general idea of the structure of data I need.

## Data analysis

In this part, we are going to find neighborhoods around the center of Williamsburg, also, we will find the number of restaurants in the neighborhoods. The final locations for our restaurant are going to be around the center of Williamsburg. First, we need to know the address of the center of Williamsburg. I create a get address function. The function will be used later to access not only for address of center Williamsburg, but also for other locations. For now, we are able to define latitude and longitude because we got the coordinate in data acquisition part. And then I create three functions to transfer coordinate into distance and transfer it back to coordinate. This sounds redundant but it's actually important when calculating distances for candidate neighborhood centers.

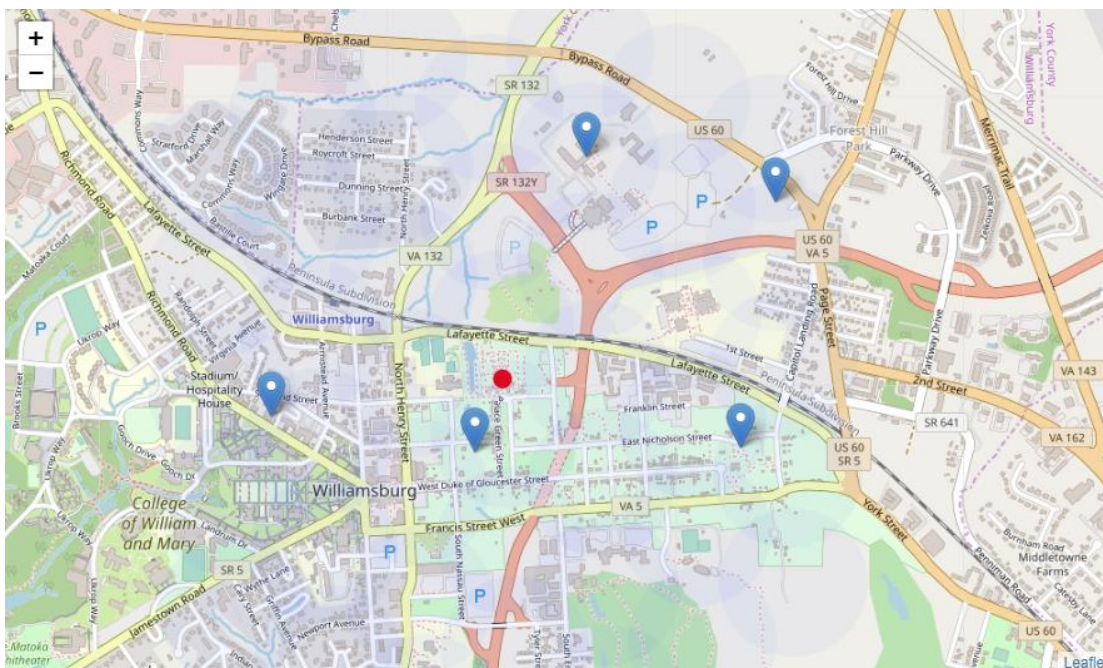
Next and the last data frame is the most important one since we have the number of restaurants in area which is one of the factors that will influence the restaurant opening. To get this information, I

create a function to grab data from Foursquare again and choose the data that are needed. This time is much more complicated because there are a series of dictionaries and lists in “requests.get(url).json()”, what I need to do is to print out the result of this requests function to find a way to access data. After that, the work left is just to count the number of restaurants in a specific area. The graph shown below is the final data frame for this project.

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	403 W Duke of Gloucester St (N Henry St), Will...	37.268093	-76.691090	-6.424049e+06	1.024451e+07	1967.231557	0
1	403 W Duke of Gloucester St (N Henry St), Will...	37.271378	-76.691289	-6.423449e+06	1.024451e+07	1670.329309	1
2	403 W Duke of Gloucester St (N Henry St), Will...	37.274664	-76.691488	-6.422849e+06	1.024451e+07	1558.845727	0
3	403 W Duke of Gloucester St (N Henry St), Will...	37.277950	-76.691687	-6.422249e+06	1.024451e+07	1670.329309	1
4	403 W Duke of Gloucester St (N Henry St), Will...	37.281237	-76.691886	-6.421649e+06	1.024451e+07	1967.231557	1
5	403 W Duke of Gloucester St (N Henry St), Will...	37.266312	-76.694550	-6.424349e+06	1.024503e+07	1824.828759	0
6	403 W Duke of Gloucester St (N Henry St), Will...	37.269598	-76.694750	-6.423749e+06	1.024503e+07	1374.772708	5
7	403 W Duke of Gloucester St (N Henry St), Will...	37.272883	-76.694949	-6.423149e+06	1.024503e+07	1081.665383	2
8	403 W Duke of Gloucester St (N Henry St), Will...	37.276169	-76.695149	-6.422549e+06	1.024503e+07	1081.665383	0
9	403 W Duke of Gloucester St (N Henry St), Will...	37.279456	-76.695348	-6.421949e+06	1.024503e+07	1374.772708	0

Next thing to do is from this data frame, find the neighborhoods that have more than 0 restaurants (meaning there are people come and go around the area) and fewer than 2 (we don't want too many competitors since we are new here).

The last step is to print the map showing spots that fit the requirements for opening a restaurant.



The red circle in this map is the very first point we find, which is the center of Williamsburg. To decide the coordinate of center, I do not use any function or package to find it, instead, I manually choose Governor's Palace since this is one of the most famous location in Williamsburg, everyone will go there, and everyone lives near the place.

## **Conclusion**

In this project, I analyzed ideal locations to open a restaurants in Williamsburg, VA, according to the distance from center and number of restaurants in certain areas. Foursquare, indeed, help me save a lot of time since the data are comprehensive with no missing value or wrong value. This project may help someone who actually want to open a restaurant. They can consider about the locations in the result. I really wish if there are more Chinese restaurants in Williamsburg because people in Williamsburg deserve to try Chinese famous food.

## **Discussion**

Although this project is done well, there are some defections. For example, I manually set the center of Williamsburg as Governor's Palace instead of asking machine to do it. This may cause the results shift towards wrong directions. Also, the choice of Williamsburg may not be a perfect decision since it's a small city with not so many restaurants or restaurants are not dense, leading some neighborhoods have 0 restaurant. These problems can be fit in the future research.