

**Opening a new cafe in Syracuse,  
New York, United States**

By: Jianing Cao  
July 2020

## Introduction

Syracuse is a city in Onondaga County, New York, United States. It is the fifth-most populous city in the state of New York. It has 38 neighborhoods covering an area of 25.6 square miles where around 142,729 people live here. More and more people nowadays prefer to buy a cup of coffee before work and school. While in some areas, the distribution of cafes is very unbalanced. To provide more convenience to more residents, we decided to open a new cafe.

The aim of this project is to analyze and select a location to open a new cafe, based on the existing cafes and neighborhood in Syracuse. Using data science methodology and machine learning techniques like clustering, this project is going to solve the business problem: where is the best location to open a new cafe in Syracuse?

To open a new cafe in this area, we need to know which area is highly populated, where are other cafes located. And also, for this type of business, it's better to be located near a school, company or shopping mall. It is very important to know if this area is overwhelmed by competitors. Therefore, the following aspects should be considered before making the decision: types of the neighborhood, the population of the neighborhood, and the number and distribution of competitors.

## Target of Audience

With more requirements of cafes, this project is aimed to provide information to developers and investors who want to open a new cafe in Syracuse, New York, United States to help them avoid oversupply of cafes and to find a high potential area.

## Data

In order to determine the best location to open a new cafe, the following information is necessary:

- **Syracuse neighborhood location** information data form Wikipedia:  
[https://en.wikipedia.org/wiki/Category:Neighborhoods\\_in\\_Syracuse,\\_New\\_York](https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Syracuse,_New_York)  
We will web scraping the neighborhoods data using Python requests and beautifulsoup packages

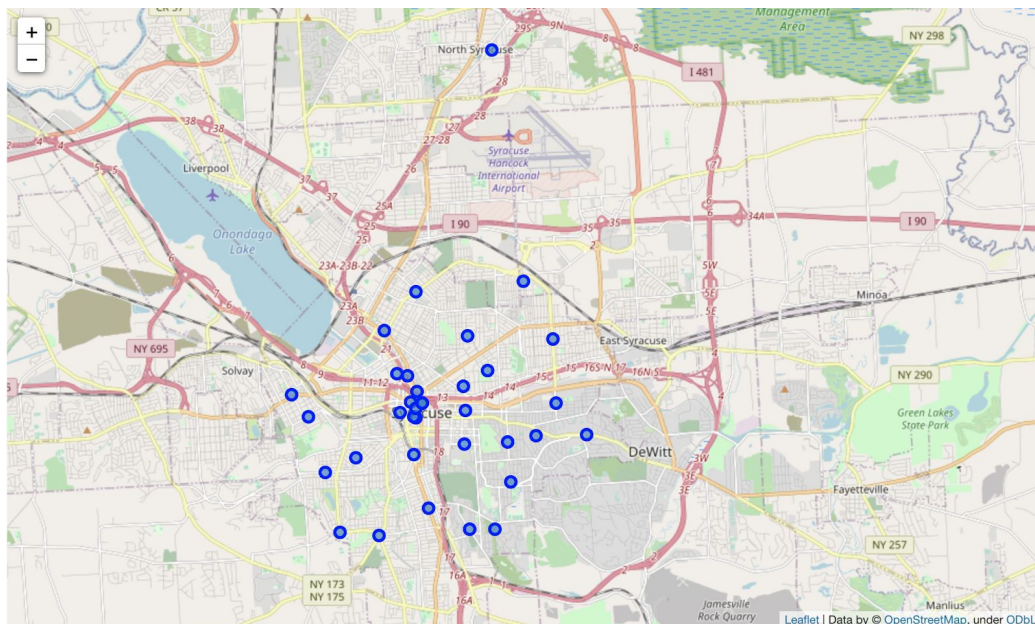
- The **longitude and latitude** of these neighborhoods: After web scraping the neighborhood data, we will use the Python Geocoder package, which will give us the latitudes and longitudes of the neighborhoods.
- **Foursquare location data:** Using the Foursquare API to get the venue data of these neighborhoods. Foursquare API will provide many categories of the venue data, we will focus on the cafe data

## Methodology

First, obtain the data of neighborhoods in Syracuse through wikipedia, then use the Python Geocoder package to get the information of the neighborhoods' latitudes and longitudes and stored them into a pandas dataframe:

	Neighborhood	Latitude	Longitude
0	Armory Square	43.048090	-76.154410
1	Brighton	43.025210	-76.145140
2	Columbus Circle	43.046820	-76.149320
3	Downtown Syracuse	43.047040	-76.150070
4	Eastwood	43.065670	-76.104840
5	Elmwood	43.018640	-76.161420
6	Far Westside	43.052410	-76.189680
7	Franklin Square	43.057352	-76.155646
8	Hanover Square	43.050500	-76.150990
9	Hawley	43.054263	-76.133955

Using Geopy library to plot all the neighborhoods on the map:



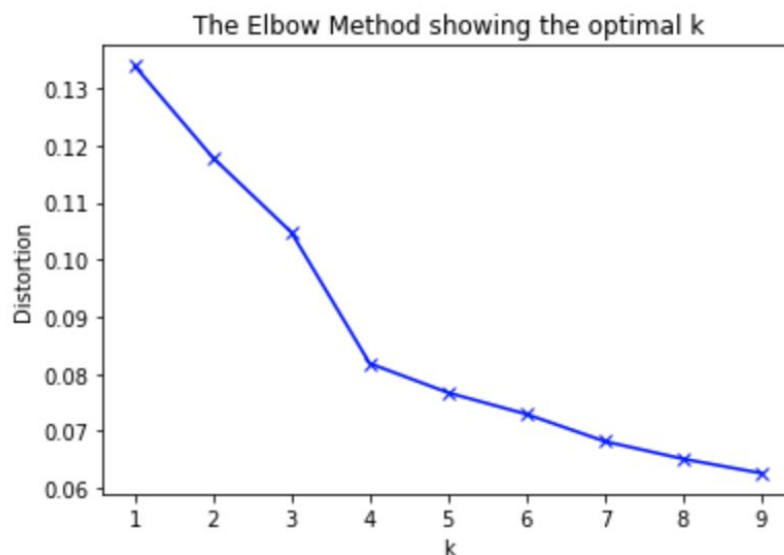
Through using the Foursquare API, get the information of the top 100 venues within a 2000 meters radius of Syracuse. Then built a python loop and passed in the neighborhoods' geographic coordinates to find the venue name, latitudes, longitude and category. Structured them into a pandas dataframe:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Armory Square	43.04809	-76.15441	Kitty Hoynes	43.048586	-76.155651	Pub
1	Armory Square	43.04809	-76.15441	Original Grain	43.048669	-76.152490	Restaurant
2	Armory Square	43.04809	-76.15441	Armory Square	43.047471	-76.155531	Plaza
3	Armory Square	43.04809	-76.15441	Al's Wine And Whiskey Lounge	43.048153	-76.153343	Wine Bar
4	Armory Square	43.04809	-76.15441	Cafe Kubal Downtown Cafe	43.046872	-76.152258	Café
5	Armory Square	43.04809	-76.15441	The Sound Garden	43.047497	-76.156094	Record Shop
6	Armory Square	43.04809	-76.15441	Urban Outfitters	43.047890	-76.156670	Clothing Store
7	Armory Square	43.04809	-76.15441	Cafe Kubal at Creekwalk Commons	43.050850	-76.155818	Coffee Shop
8	Armory Square	43.04809	-76.15441	Landmark Theatre	43.047167	-76.152553	Theater
9	Armory Square	43.04809	-76.15441	Funk 'n Waffles	43.048340	-76.153364	Breakfast Spot

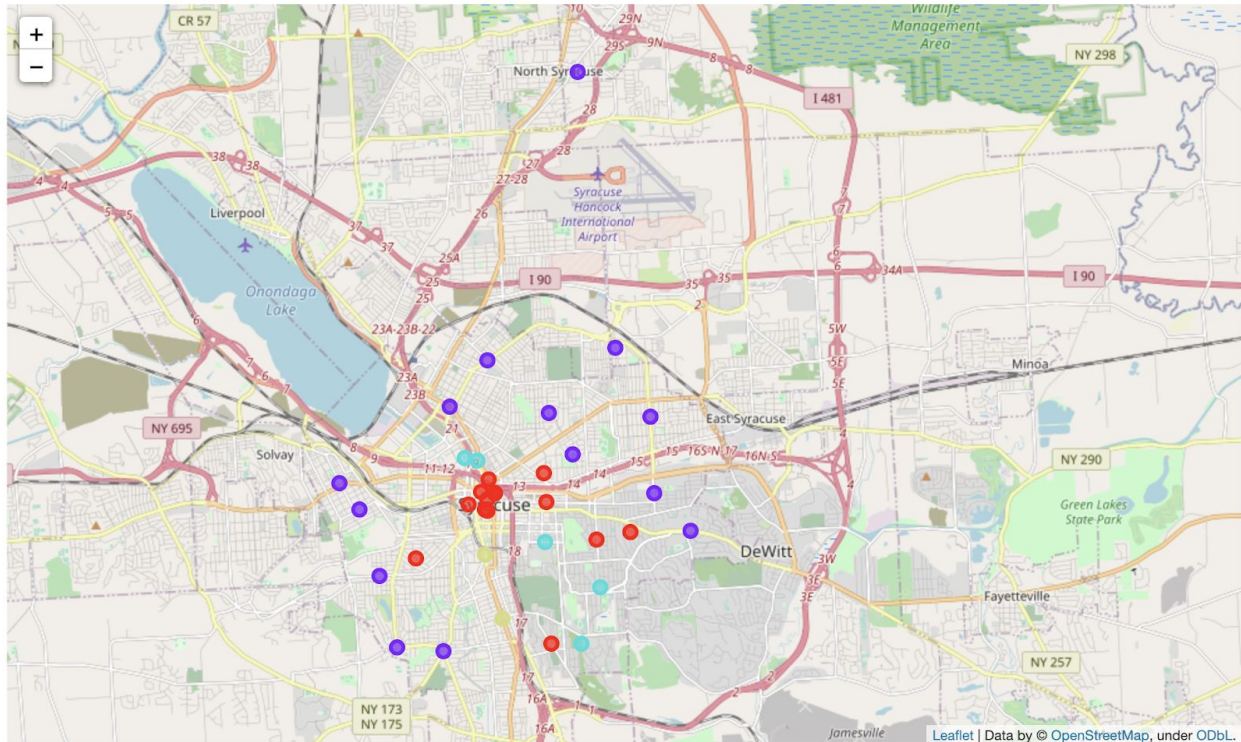
By grouping by the data based on the neighborhood, it is easy to see that there are 193 unique categories. Through one hot method, we selected the category that only contains cafes and saved it into syr\_cafe dataframe.

## Results

To find the best numbers of clusters, we used the Elbow method. We can see after point 4 the distortion/inertia starts decreasing in a linear fashion. Thus for the given data, we concluded that the optimal number of clusters for the data is **4**.



So, define the k=4 and make the clusters and draw the results on the map:



From the map:

- Cluster 0 (Red Points) most distributed in the downtown Syracuse
- Cluster 1 (Purple Points) distributed in the north-east and south-west of the Syracuse
- Cluster 2 (Blue Points) near the center and south-east of Syracuse
- Cluster 3 (Green Points) Brighton and Southwest neighborhoods

## Discussion

We separate Syracuse into four parts, given the  $k=4$ . Through the map we can see, cluster 1 has the highest number and cluster 0 has moderate number, which is most distributed in the central, south-east and north-east of Syracuse. On the other hand, cluster 2 and 3 have very low numbers of cafes in the neighborhoods. This represents a great opportunity and high potential areas to open new cafes as there is very little to no competition from existing multiplexes. Meanwhile, cafes in cluster 0 and 1 are likely suffering from intense competition due to oversupply and high concentration of cafes. Therefore, this project recommends property developers to capitalize on these findings to open new cafes in neighborhoods in cluster 2 or 3 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new cafes in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 1 which already have high concentration of cafes and suffering from intense competition.

## **Conclusion**

Starting a new business requires a lot of steps and considerations before it becomes official, one of the most important things that should be carefully explored is its location. For this project, the goal was to examine the best location for opening a new cafe in Syracuse, New York, United States, based on the information of the neighborhood, the distribution of competitors in each neighborhood. By understanding the business goal, whether we want less competition or we have unique selling propositions to stand out from the competition, we can pick the most ideal location through this kind of analysis.