The Battle of Neighborhoods -
Opening A Restaurant in Philadelphia

NOVEMBER 3

Applied Data Science Capstone Project

Authored by: Alvis Wu

Introduction

This report is for the final project in the IBM Applied Data Science Capstone course. It aims to leverage the Foursquare (a location data and intelligence company) location data to explore or compare neighborhoods or cities of my choice or to come up with a problem that I can use the Foursquare location data to solve. In this project I plan to analyze the different neighborhoods in Philadelphia and give recommendations to investors planning to open restaurants in Philadelphia.

Philadelphia, also known as Philly, is the largest city in the US State of Pennsylvania, and the sixth-most populous US city with a 2018 census-estimated population of 1,584,138. It has 5 Fortune 1000 companies and is the headquarter of companies such as Comcast, Cigna and Urban Outfitters. It’s also the home of more than 80 colleges, universities, trade, and specialty schools including University of Pennsylvania, Temple University, Drexel University, and Thomas Jefferson University. There are more than 120,000 college and university students enrolled within the city and nearly 300,000 in the metropolitan area. Moreover, Philadelphia is also a city full of history that attracts more than 40 million domestic tourists annually.

Without a doubt, there are huge demands for restaurants in Philadelphia and nearby area. However, there are more than 430 full-service restaurants inside just the central business district, and close to 4,000 restaurants in the city. To successfully launch a business venture to open a restaurant in Philadelphia requires a comprehensive analysis of the competitions in all the neighborhoods in the city. In this project, the goal is to utilized information about restaurants in Foursquare and machine learning methods to group the neighborhoods into different clusters with distinct characteristics so that the investors can make better decisions.

Data

Even though Philadelphia has neighborhood names such as Old City, Pint Breeze and Bella Vista, people in real estate more often use zip codes when they refer to a neighborhood. It's different than how people use neighborhood names in places such as NYC. Therefore, in this project I will use zip codes rather than neighborhood names. Data in this project are mainly from three sources:

1. A list of zip codes of Philadelphia. This information is available in numerous websites online. I used http://www.ciclt.net/sn/clt/capitolimpact/gw_ziplist.aspx?FIPS=42101. The table inside is turned into a data frame using Python package Pandas.
2. Latitudes and longitudes of the zip codes. Website https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/ provides such information. A table is downloaded and uploaded to the project. The latitudes and longitudes are required when using the Foursquare APIs.
3. Restaurant categorical information from Foursquare. A limited number (100) of restaurants are searched and summarized within a radius around the coordinates of each neighborhood.

Methodology

Summary

The following steps are involved to produce the results I need:

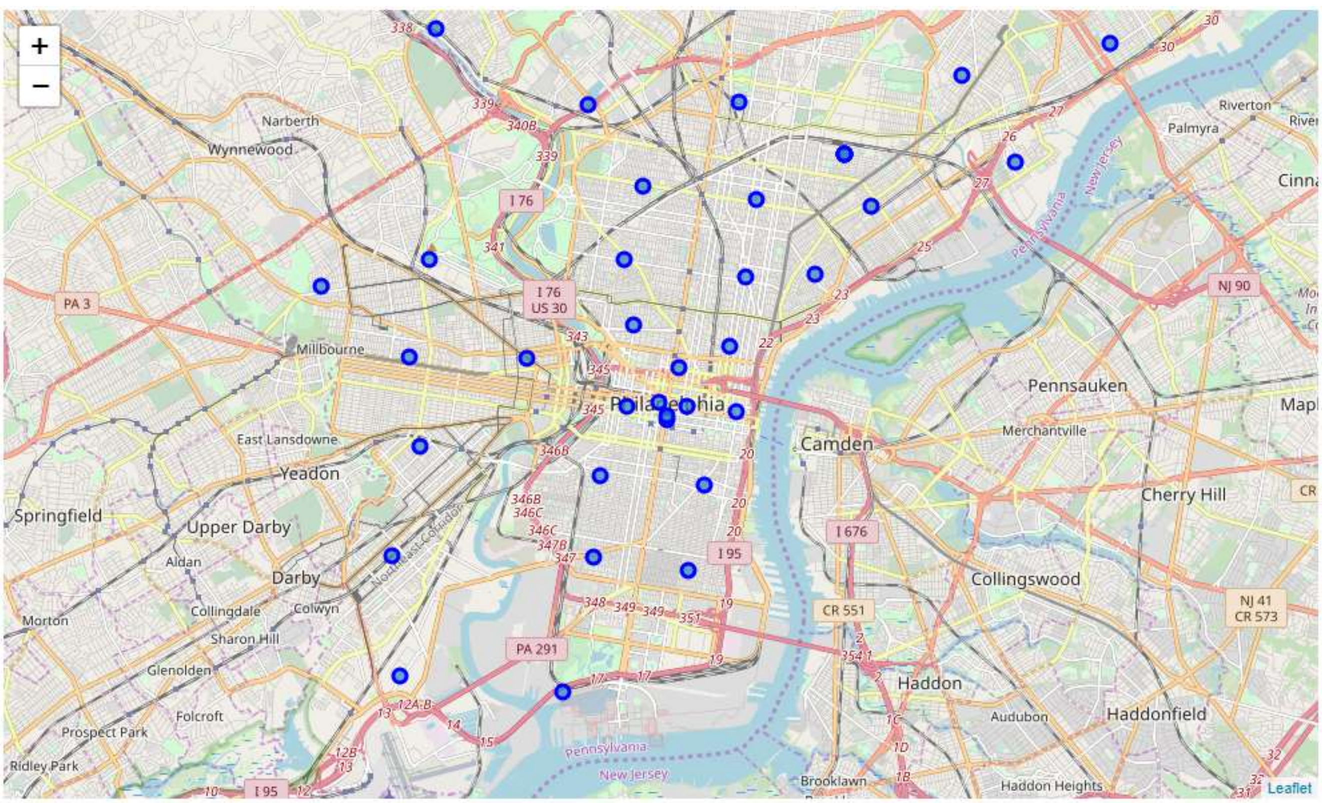
1. A list of zip codes of Philadelphia is scrapped from a table stored in a website.
2. Latitudes and longitudes of all Pennsylvania zip codes are downloaded from a website.
3. The coordinates are added to the list of Philadelphia zip codes.
4. Searches of restaurants are done in Foursquare for all neighborhoods.
5. K-means machine learning method is run on the data to cluster the neighborhoods into clusters.

Details

After getting the list of zip codes and coordinates, I merged them into a data frame contains zip code, city, latitude and longitude.

	Zipcode	City	Latitude	Longitude
0	19019	Philadelphia	40.001811	-75.11787
1	19101	Philadelphia	40.001811	-75.11787
2	19102	Mid City East, Middle City East, Philadelphia	39.952962	-75.16558
3	19103	Mid City West, Middle City West, Philadelphia	39.952162	-75.17406
4	19104	Philadelphia	39.961612	-75.19957

Let's first use the python package folium to generate a map with all these neighborhoods marked out.



Now, let's utilized Foursquare API to explore the neighborhoods. I used the query "restaurants" with limit of 100 venues and radius 500 meters to search around the coordinates. There are in total 1,142 results returned for all the neighborhoods. Here are the first 5 records.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	19019	40.001811	-75.11787	Teplitzky's	40.004032	-75.118782	American Restaurant
1	19019	40.001811	-75.11787	Classic Pizza	40.001532	-75.114627	Pizza Place
2	19019	40.001811	-75.11787	Kitchen Express	39.999921	-75.115444	Food
3	19019	40.001811	-75.11787	Hoagies Plus	40.001568	-75.114342	Deli / Bodega
4	19019	40.001811	-75.11787	EL COQUI	39.998548	-75.117013	Spanish Restaurant

There are 3 zip codes (19102, 19107 and 19110) reach the 100 limits. Most of the zip codes returned less than 50 records with 19112, 19114 and 19115 have the least of 2 records.

Foursquare returned 84 different types of venue categories, such as “American Restaurant”, “Bakery” and “Pizza Place”.

	Neighborhood	African Restaurant	American Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Brazilian Restaurant	Breakfast Spot	Burger Joint	Burmese Restaurant	Cafeteria	
0	19019	0.0	0.200000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.00	0.0	0.000000	0.0
1	19101	0.0	0.200000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.00	0.0	0.000000	0.0
2	19102	0.0	0.080000	0.020000	0.0	0.010000	0.060000	0.0	0.0	0.010000	0.01	0.0	0.000000	0.0
3	19103	0.0	0.092784	0.010309	0.0	0.030928	0.051546	0.0	0.0	0.000000	0.00	0.0	0.000000	0.0
4	19104	0.0	0.058824	0.000000	0.0	0.000000	0.058824	0.0	0.0	0.058824	0.00	0.0	0.058824	0.0

To show the characteristics of the restaurants in the neighborhoods, I ranked the most common types of restaurants in them. The percentages of the different types are also calculated. Below are a few examples of the 5 most common restaurants in a few neighborhoods.

```
----19019----
                venue  freq
0         Pizza Place   0.2
1 American Restaurant   0.2
2       Deli / Bodega   0.2
3                Food   0.2
4  Spanish Restaurant   0.2
```

```
----19101----
                venue  freq
0         Pizza Place   0.2
1 American Restaurant   0.2
2       Deli / Bodega   0.2
3                Food   0.2
4  Spanish Restaurant   0.2
```

```
----19102----
                  venue  freq
0   American Restaurant  0.08
1                  Café  0.07
2           Salad Place  0.06
3                Bakery  0.06
4    Italian Restaurant  0.05
```

```
----19104----
                  venue  freq
0           Pizza Place  0.18
1    Chinese Restaurant  0.12
2     Indian Restaurant  0.06
3             Cafeteria  0.06
4          Deli / Bodega  0.06
```

From the examples above we can already see that pizza place is the most common type in 19019, 19101 and 19104. Such information can tell investors opening a restaurant that pizza place might be already over-saturated in these neighborhoods. Are there other neighborhoods also have a lot of pizza place? What can we say about the rest of the neighborhoods that pizza place is not the most common?
To answer the questions, we need to apply a machine learning algorithm called K-means. It is an unsupervised machine learning algorithm that group unlabeled data (in our case, neighborhoods) into a few clusters with different features. I first used 5 clusters. You can see below the first few neighborhoods and the clusters they belong.

	Neighborhood	City	Latitude	Longitude	Cluster Labels	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 Co
0	19019	Philadelphia	40.001811	-75.11787	3	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Crep
1	19101	Philadelphia	40.001811	-75.11787	3	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Crep
2	19102	Mid City East, Middle City East, Philadelphia	39.952962	-75.16558	0	American Restaurant	Café	Salad Place	Bakery	Pizza Place	Italian Restaurant	Mexican Restaurant	New Ame Rest
3	19103	Mid City West, Middle City West, Philadelphia	39.952162	-75.17406	0	American Restaurant	Italian Restaurant	Deli / Bodega	Café	New American Restaurant	Bakery	Pizza Place	Seaf Rest
4	19104	Philadelphia	39.961612	-75.19957	0	Pizza Place	Chinese Restaurant	Food Truck	Indian Restaurant	Deli / Bodega	Caribbean Restaurant	Sandwich Place	Mexi Rest

Results

Firstly, let's list the neighborhoods in the 5 different clusters:

Cluster 1 (only the first portion of the list due to quantity)

	Neighborhood	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
2	19102	American Restaurant	Café	Salad Place	Bakery	Pizza Place	Italian Restaurant	Mexican Restaurant	New American Restaurant	Steakhouse	Mediterranean Restaurant
3	19103	American Restaurant	Italian Restaurant	Deli / Bodega	Café	New American Restaurant	Bakery	Pizza Place	Seafood Restaurant	Bagel Shop	Restaurant
4	19104	Pizza Place	Chinese Restaurant	Food Truck	Indian Restaurant	Deli / Bodega	Caribbean Restaurant	Sandwich Place	Mexican Restaurant	Cafeteria	Mediterranean Restaurant
6	19106	Italian Restaurant	New American Restaurant	Café	American Restaurant	Gastropub	Sushi Restaurant	Pizza Place	Sandwich Place	French Restaurant	Indian Restaurant
7	19107	Bakery	Chinese Restaurant	Sandwich Place	Deli / Bodega	Burger Joint	Donut Shop	Breakfast Spot	Snack Place	Pizza Place	Asian Restaurant
8	19108	Food Truck	Pizza Place	Breakfast Spot	Chinese Restaurant	Sandwich Place	Deli / Bodega	Gastropub	Vietnamese Restaurant	Donut Shop	Italian Restaurant
9	19109	Pizza Place	American Restaurant	Mexican Restaurant	Italian Restaurant	Sandwich Place	Bakery	Steakhouse	Seafood Restaurant	New American Restaurant	Mediterranean Restaurant
10	19110	Pizza Place	Italian Restaurant	American Restaurant	Sandwich Place	Bakery	Mexican Restaurant	Seafood Restaurant	Steakhouse	Deli / Bodega	New American Restaurant
11	19111	Deli / Bodega	Pizza Place	Snack Place	Food Truck	Chinese Restaurant	Bakery	Italian Restaurant	Diner	Sandwich Place	Donut Shop
15	19116	Chinese Restaurant	Pizza Place	Indian Restaurant	Diner	Spanish Restaurant	Donut Shop	English Restaurant	Creperie	Cuban Restaurant	Deli / Bodega
16	19118	Bakery	American Restaurant	French Restaurant	Noodle House	Snack Place	Mexican Restaurant	Chinese Restaurant	Korean Restaurant	Caribbean Restaurant	Pizza Place
17	19119	Pizza Place	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Food	Bakery	Deli / Bodega	Breakfast Spot	Café	Cuban Restaurant
		Donut	Korean	Seafood	Fast Food	Pizza	Sandwich		Chinese	Thai	Falafel

Cluster 2

	Neighborhood	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
12	19112	Food Truck	Wings Joint	English Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant

Cluster 3

	Neighborhood	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
13	19114	Pizza Place	Donut Shop	English Restaurant	Comfort Food Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Eastern European Restaurant
22	19124	Pizza Place	Chinese Restaurant	Deli / Bodega	Fast Food Restaurant	Breakfast Spot	Donut Shop	Wings Joint	English Restaurant	Cuban Restaurant	Dim Sum Restaurant
24	19126	Pizza Place	Korean Restaurant	Chinese Restaurant	Deli / Bodega	Donut Shop	Japanese Restaurant	Wings Joint	Cuban Restaurant	Dim Sum Restaurant	Diner
32	19134	Pizza Place	Donut Shop	American Restaurant	Chinese Restaurant	Deli / Bodega	Food	Wings Joint	English Restaurant	Cuban Restaurant	Dim Sum Restaurant
33	19135	Pizza Place	Food	American Restaurant	Chinese Restaurant	Fast Food Restaurant	Donut Shop	Wings Joint	Eastern European Restaurant	Cuban Restaurant	Deli / Bodega
35	19137	Pizza Place	American Restaurant	Restaurant	Café	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant
38	19140	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Diner	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant
40	19142	Pizza Place	Breakfast Spot	African Restaurant	American Restaurant	Sandwich Place	Eastern European Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant
50	19152	Pizza Place	Chinese Restaurant	Burger Joint	Seafood Restaurant	Deli / Bodega	Restaurant	Brazilian Restaurant	Diner	Italian Restaurant	Wings Joint
52	19154	Pizza Place	Deli / Bodega	English Restaurant	Comfort Food Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant

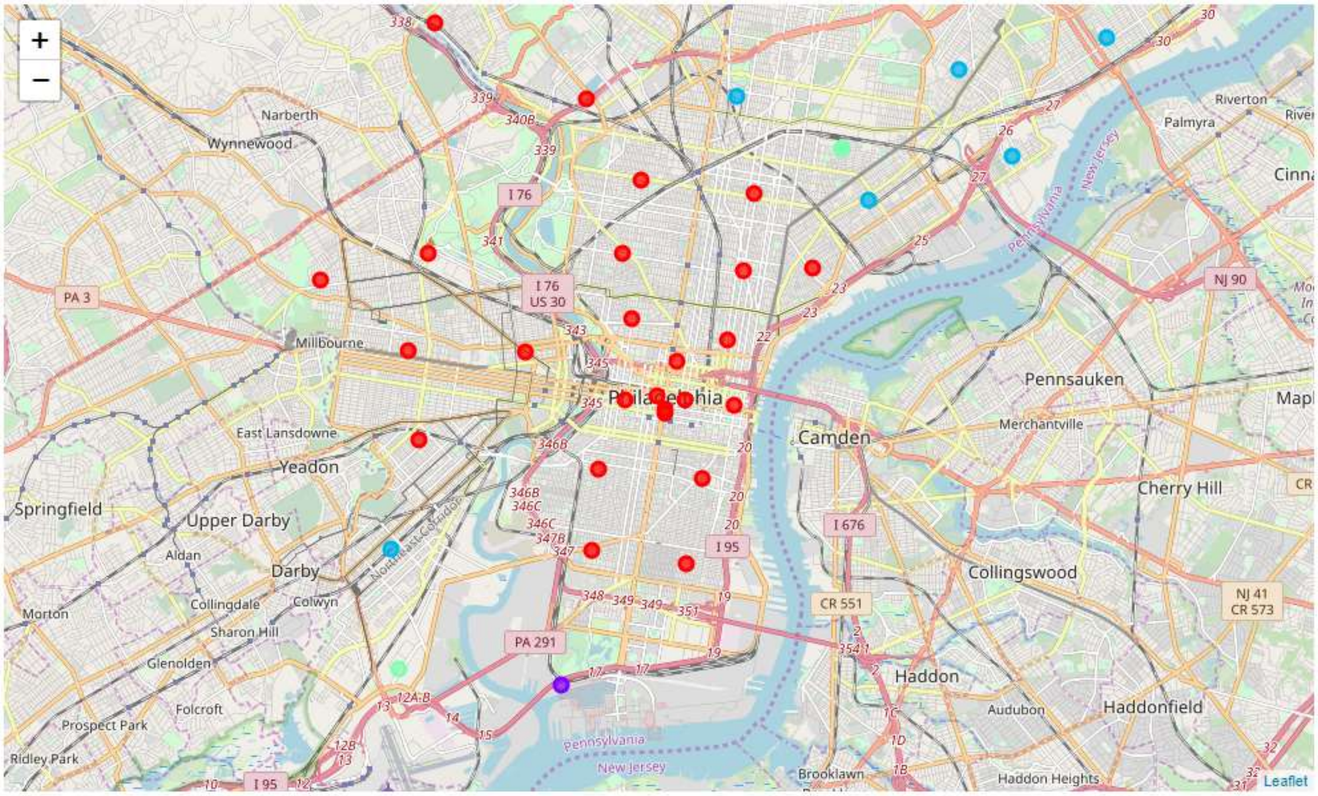
Cluster 4

	Neighborhood	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	19019	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant
1	19101	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant
5	19105	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant
51	19153	American Restaurant	Breakfast Spot	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant
53	19155	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant
54	19160	American Restaurant	Pizza Place	Food	Deli / Bodega	Spanish Restaurant	Wings Joint	Eastern European Restaurant	Creperie	Cuban Restaurant	Dim Sum Restaurant

Cluster 5

	Neighborhood	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
14	19115	Bakery	Italian Restaurant	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant

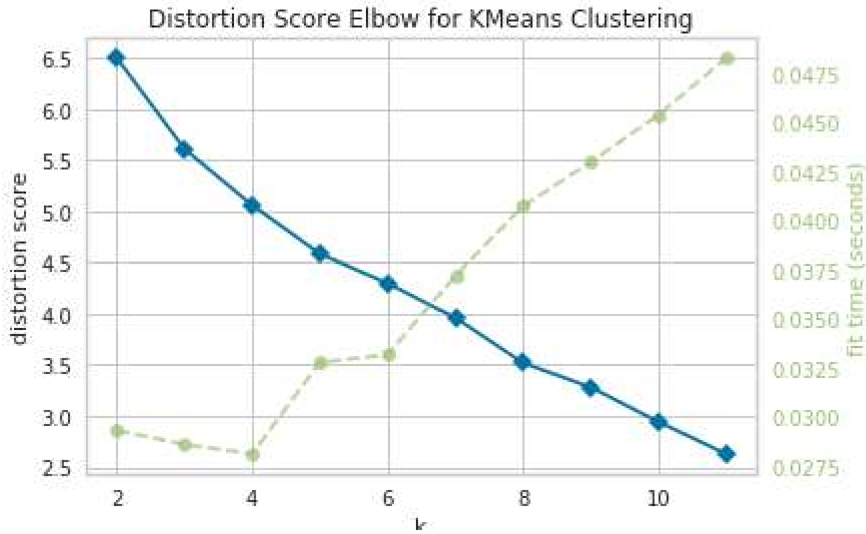
I now use folium to generate a map to show the different clusters:



Discussion

From the results in the section above, we can see that the neighborhoods in cluster 3 all have pizza place as their most common venue while those in cluster 4 has American restaurant as their most common. Neighborhoods in cluster 1 have pizza place and American restaurant in the top 5 but mostly not the most common. Cluster 2 and 5 have 1 neighborhood each with very different restaurant mixes.

Now it's natural to ask: is 5 the best number to cluster the neighborhoods? We can use the elbow method to find out.



In the graph above, the x-axis shows the number of clusters and the y-axis shows the distortion score. The lower the score the closer the data in a cluster to the "center" of the cluster. Keep in mind that the distortion score always decreases as the number of clusters increase. The best number of clusters is at the point beyond which the decrease of the distortion score significantly slows down. We can't find such a point in the graph above. However, with 5 different clusters, I consider I already have enough information about the restaurants in the neighborhoods.

For an investor trying to open a pizza place in Philadelphia, he/she should check out the neighborhoods in cluster 1. In these neighborhoods, pizza places are common, showing that there are demands but not as over-saturated as in neighborhoods in cluster 3.

Conclusion

Every neighborhood has its own unique restaurant mix that opening a certain type of restaurant may or maybe be suitable. In this project, I was able to categorize the neighborhoods into a few clusters with distinct features to better assist restaurant investors to make educated decisions.

References

1. https://en.wikipedia.org/wiki/Philadelphia
2. http://www.ciclt.net/sn/clt/capitolimpact/gw_ziplist.aspx?FIPS=42101
3. https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/
4. https://foursquare.com