

The Battle of Neighborhoods – Restaurants in Pittsburgh

Peilin Tian

Capstone Project for IBM Data Science Professional Certificate

Introduction

Pittsburgh is a city at the junction of three rivers with a variety of cultural and historical heritages. The food from most parts of the world is an essential part of them. For an investor, he/she may be interested in the type of restaurant one should open in a specific neighborhood based on the most popular kinds of restaurants in the area. As a consumer, he/she would like to know where to go for a specific type of restaurant. In this project, geo spatial data from the Internet and venue information from Foursquare API would be used combined with K-means clustering to find the most common types of restaurants in Pittsburgh neighborhoods. Each cluster will be examined and analyzed to answer the imposed questions above.

Data

Zip codes and their coordinates of Pittsburgh neighborhoods can be obtained from some internet sources. For this project, the data from *US Zip Code Latitude and Longitude* [<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>], specifically the data of Pennsylvania would be used. Some data cleaning is required to get the data we need in this project.

Restaurant information would be obtained through Foursquare API using the coordinates from the geo spatial information of each neighborhood.

Methodology

In this project, we will use the zip codes of each neighborhood in Pittsburgh and their corresponding coordinates along with the venue data from Foursquare API to cluster them into different clusters based on the patterns of popular restaurant types in each neighborhood using K-means clustering, because K-means clustering is an unsupervised clustering method where data points are clustered based on their similarities. It is an automatic process so no further assumptions or operations are required once we feed the data into the method. After getting all the cluster information, we can analyze each cluster and determine the right type of restaurant to open in the neighborhood for investors or the neighborhood to go for a specific type of restaurant for consumers.

Below are the neighborhoods plotted on the map of Pittsburgh (Fig 1).

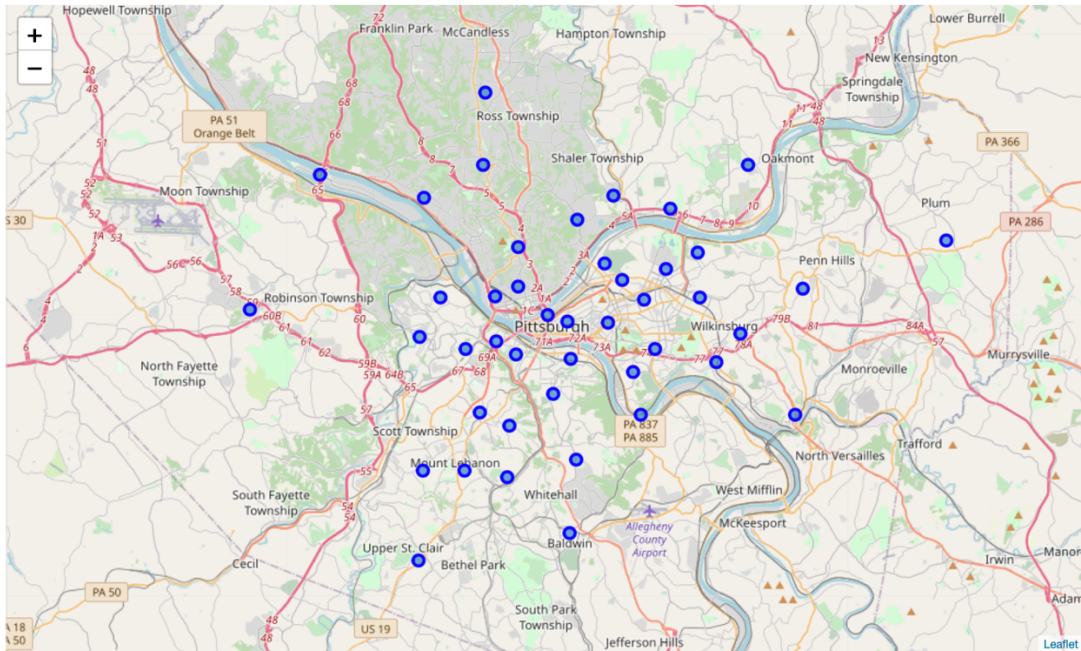


Fig 1. Neighborhoods of Pittsburgh based on zip codes

After cleaning the data, including removing the duplicate coordinates, and merging the data with Foursquare restaurant information, a data frame containing all restaurant information in each neighborhood is formed. Then it is one-hot encoded to make sure all the variables are numerical so that we can use them in K-means clustering. Then the dataframe is sorted to show the top 5 common types of restaurants in each neighborhood. The resulting dataframe as shown below (Fig 2).

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	15112	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
1	15201	Vietnamese Restaurant	Seafood Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Restaurant
2	15202	Italian Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
3	15203	American Restaurant	Sushi Restaurant	Asian Restaurant	Mexican Restaurant	Thai Restaurant
4	15204	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
5	15205	Chinese Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Cuban Restaurant
6	15206	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
7	15207	Seafood Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
8	15208	Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
9	15209	Fast Food Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
10	15210	American Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
11	15211	American Restaurant	Russian Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
12	15212	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
13	15213	New American Restaurant	Asian Restaurant	Chinese Restaurant	Sushi Restaurant	Indian Restaurant
14	15214	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
15	15215	Italian Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
16	15216	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
17	15217	Vietnamese Restaurant	Thai Restaurant	Szechuan Restaurant	Sushi Restaurant	Chinese Restaurant
18	15218	American Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant
19	15219	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
20	15220	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant

Fig 2. Top 5 common types of restaurants in each neighborhood (only showing the first 20 entries)

After this step, the data is fed into K-means clustering and the cluster labels are created for each neighborhood. The data are separated into five clusters. The clusters are shown on the map below (Fig 3).

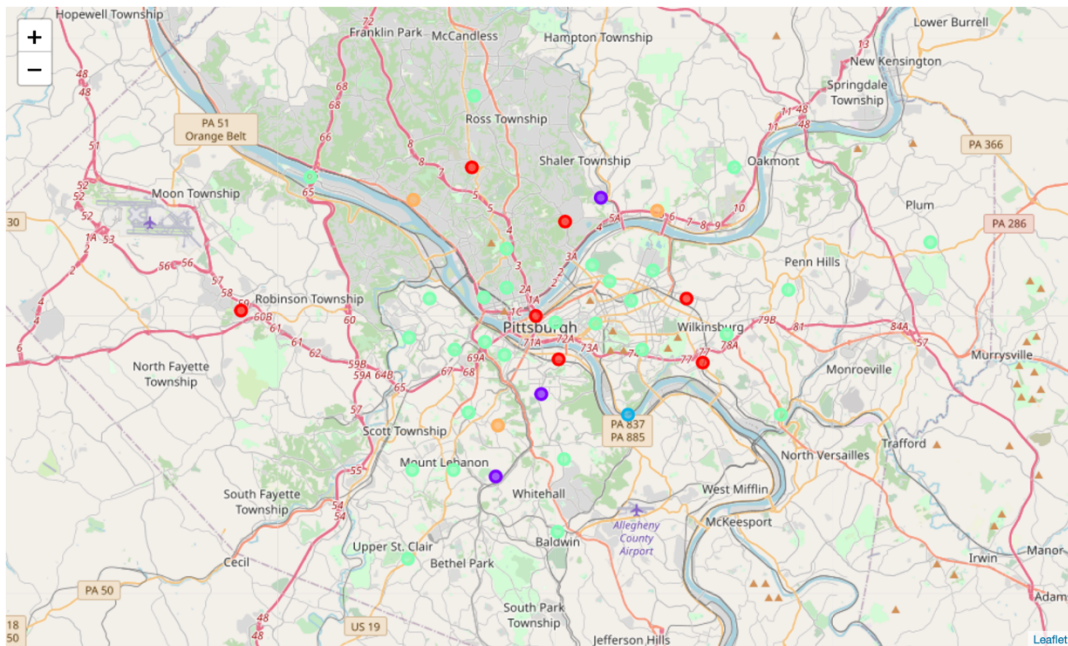


Fig 3. Clustering results for each neighborhood.

Results and Discussions

The resulting clusters are examined and analyzed below.

Cluster 1

Neighborhood		1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
12	15275	Fast Food Restaurant	Mexican Restaurant	Italian Restaurant	Restaurant	Asian Restaurant
35	15203	American Restaurant	Sushi Restaurant	Asian Restaurant	Mexican Restaurant	Thai Restaurant
37	15218	American Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Mediterranean Restaurant
38	15229	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Vietnamese Restaurant
46	15222	American Restaurant	Italian Restaurant	Restaurant	Indian Restaurant	Korean Restaurant
49	15209	Fast Food Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
55	15208	Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant

As the above information shows, in this cluster the most popular restaurants are fast food and American restaurants, if you would like to open one you might want to choose somewhere else to avoid fierce competition. On the other hand, it would be a good choice to go if you would like some fast food or American food.

Cluster 2

Neighborhood		1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
36	15234	American Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
50	15210	American Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
61	15223	American Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant

In this cluster, the most common types are American and Mediterranean restaurants, so avoid open one in these neighborhoods.

Cluster 3

Neighborhood		1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
78	15207	Seafood Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant

This cluster contains only one neighborhood so there is not enough information for us to decide whether it is a good idea to open a specific kind of restaurant but definitely don't try seafood restaurant here. For food seekers the seafood here may be a good choice.

Cluster 4

Neighborhood		1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	15278	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
1	15227	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
2	15238	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
4	15236	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
6	15239	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
9	15112	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
10	15201	Vietnamese Restaurant	Seafood Restaurant	Middle Eastern Restaurant	Ramen Restaurant	Restaurant
16	15216	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
22	15228	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
25	15225	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
28	15237	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
29	15212	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
31	15243	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
33	15205	Chinese Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Cuban Restaurant
39	15206	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
41	15204	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
43	15220	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
47	15213	New American Restaurant	Asian Restaurant	Chinese Restaurant	Sushi Restaurant	Indian Restaurant
51	15224	Thai Restaurant	New American Restaurant	Italian Restaurant	Japanese Restaurant	Chinese Restaurant
53	15241	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
58	15219	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
63	15233	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
65	15211	American Restaurant	Russian Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
68	15221	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
71	15214	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant
72	15217	Vietnamese Restaurant	Thai Restaurant	Szechuan Restaurant	Sushi Restaurant	Chinese Restaurant
74	15232	Japanese Restaurant	Asian Restaurant	Middle Eastern Restaurant	Turkish Restaurant	Korean Restaurant
77	15235	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant	Cuban Restaurant

This cluster contains the greatest number of neighborhoods and they follow basically the same pattern. The most popular types are Vietnamese, Mediterranean and Asian, in this order. So maybe Asian restaurant, or be more specific Chinese restaurant, would be a decent choice since there are less competition compared to the previous two.

Cluster 5

Neighborhood		1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
54	15202	Italian Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
56	15215	Italian Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	Asian Restaurant	Chinese Restaurant
60	15226	Italian Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant

This final cluster is special in that it is the only cluster where an European cuisine, Italian food, is the most popular kind of restaurant in the area.

Conclusion

In this project, the neighborhoods of Pittsburgh are clustered into different clusters based on the popular types of restaurants in it. If someone wants to open a new restaurant, he/she should avoid the popular choices and aim for the less popular kinds, for example the 4th or 5th ones instead. For consumers, this clustering is also inspiring if they want to find the best restaurant for a specific type. They will know where to go to have the best chance finding a good one.