

Identifying the best neighborhood in the city of Philadelphia to setup an Indian restaurant

Abhishek Aiyangar

Introduction/Business Problem: The problem I have chosen for my data science project is to determine the best neighborhood in Philadelphia to open an Indian restaurant.

Background

Being home to universities such as University of Pennsylvania, Temple, and Drexel, several international students call Philadelphia their home. The largest ethnic minority of international students in Philadelphia are of Chinese and Indian descent. With a steady influx of international students in universities across Philadelphia, there is an obvious demand for quality restaurants to cater to these students, along with the city's longtime residents.

Currently, Philadelphia has a few Indian restaurants scattered across various neighborhoods of the city. However, the general consensus on the quality and service of these restaurants is abysmal. Specifically, the following are some of the problems associated with Indian restaurants in the city:

- Quality of food: *lack of authentic Indian cuisine*
- Price (\$\$\$): *unaffordable to students*
- Distance: *located away from public transportation*
- Ambience: *very messy to extremely lavish*

Therefore, the city is need of a quality affordable Indian restaurant with authentic cuisine having the ability to cater to students, city residents and tourists alike.

Audience

The idea described above will resonate with entrepreneurs and businesses who intend to serve quality authentic Indian food to the entire community of Philadelphia, focused specifically on the student population of the city.

Since one of the problems associated with the current Indian restaurants in Philadelphia is location, the solution of this data science problem will help potential stakeholders determine the ideal neighborhood to establish an Indian restaurant. Selecting the right neighborhood for a restaurant is as important as the price and quality of food. Choosing the right location provides a restaurant with a great opportunity to be successful having a steady inflow of customers throughout the year. With Philadelphia's brutal winter, it becomes even more crucial to select a location that is convenient for people using different modes of transport.

Data

Following are the datasets that will be used to solve this problem:

- 1) Foursquare data (Venues, universities, public transport locations, surrounding business competition)

Foursquare location data will be used for the city of Philadelphia to obtain the following:

- Neighborhoods: an inventory of all the different neighborhoods in the city
- Using Venues,
- Universities: an inventory of all the universities and their locations in the city
- Public transport locations: (if available) locations of all bus and train stops in the city
- Yoga studios
- Indian restaurants: A large number of Indian restaurants in the neighborhood indicates greater competition

The above data will help better understand the different neighborhoods in Philadelphia. One or more of the above data elements may be used to better determine the ideal neighborhood for an Indian restaurant.

For example: Penn's Landing in Philadelphia, being a tourist attraction, has a number of bus and train stops in its vicinity. However, due to its high real estate prices and lack of universities in the neighborhood, it may not be the best location to establish an Indian restaurant.

- 2) Crime rates by neighborhood

Philadelphia has a higher crime rate compared to other cities in the North East area. Therefore, it is critical for a commercial venture to have low incidents of crime in its neighborhood to attract more customers. For example, Bridesburg in Philadelphia has reported 34¹ incidents of crime between Jan 22, 2020 to Feb 19, 2020. Therefore, it may not be the wisest decision to establish a restaurant in this neighborhood.

- 3) Population

A densely populated neighborhood improves the chances of more customers visiting a new restaurant. Therefore, independent of other variables, population is a great variable to be used for determining the ideal neighborhood to setup a new Indian restaurant.

¹ Source: <http://data.philly.com/philly/crime/?&nType=crime&dNeigh=Bridesburg>

Methodology

The following are the activities performed sequentially in order to solve the data science problem described above:

Research

The first step in the process involved better understanding the variables that will help determine the best neighborhood to open a restaurant. Based on research online², some of the factors that should be considered to locate a restaurant are:

- Parking
- Safety
- Visibility
- Competition
- Property Rental prices

However, since we are interested in setting up an Indian restaurant, the variables were tailored to solve this specific problem. Additional consideration was made towards the availability of data for the city of Philadelphia.

As a result, the final data points that were finalized to be used for research were:

- Number of Indian restaurants in the neighborhood
- Number of Public Transport options
- Number of culturally relevant vendors (i.e. Yoga studios)
- Population density
- Number of universities
- Crime rate

Data Preparation

After identifying the variables to be used for analysis, the next step in the process required for obtaining the data from external sources (or in some instances preparing the data manually). Below is a detail breakdown of the approach to prepare the dataset:

1. Creating a dataframe containing the list of the most popular neighborhoods in the city of Philadelphia and their corresponding zipcodes.
Source: <https://www.visitphilly.com/areas/philadelphia-neighborhoods/>
2. Since there were many instances of the same zipcode having multiple neighborhoods, these neighborhoods were grouped together separated by a comma.

² <https://www.thebalancesmb.com/choosing-restaurant-location-2888543>

	Zipcode	Neighborhood
0	19102	Logan Square
1	19103	Fairmount, Rittenhouse Square
2	19104	Powelton Village, Spruce Hill and Cedar Park, ...
3	19106	Old City, Society Hill
4	19107	Callow Hill, Chinatown, Market East, Midtown V...
5	19123	Northern Liberties
6	19125	Fishtown
7	19130	Spring Garden
8	19146	Graduate Hospital
9	19147	Bella Vista, Pennsport, Queen Village, Washing...
10	19148	East Passyunk

- The latitudes and longitudes corresponding to each zipcode were then concatenated into the dataframe using the Nominatim package.

	Zipcode	Neighborhood	Latitude	Longitude
0	19102	Logan Square	39.946212	-75.165018
1	19103	Fairmount, Rittenhouse Square	39.955870	-75.171830
2	19104	Powelton Village, Spruce Hill and Cedar Park, ...	39.949252	-75.209377
3	19106	Old City, Society Hill	39.947500	-75.146213
4	19107	Callow Hill, Chinatown, Market East, Midtown V...	39.947072	-75.154648
5	19123	Northern Liberties	39.948016	-75.223406
6	19130	Spring Garden	39.964310	-75.166057
7	19146	Graduate Hospital	39.940422	-75.177791
8	19147	Bella Vista, Pennsport, Queen Village, Washing...	39.937486	-75.155404
9	19148	East Passyunk	39.926922	-75.167059

- The next step involved using the Foursquare package to obtain venues information. For our problem, we have considered a limit of 500 venues within a radius of 500 feet.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Logan Square	39.946212	-75.165018	Kimmel Center for the Performing Arts	39.946785	-75.165234	Concert Hall
1	Logan Square	39.946212	-75.165018	Writer's Block Rehab	39.946450	-75.163602	Cocktail Bar
2	Logan Square	39.946212	-75.165018	Verizon Hall	39.946598	-75.165976	Concert Hall
3	Logan Square	39.946212	-75.165018	Vetri	39.946761	-75.163183	Italian Restaurant
4	Logan Square	39.946212	-75.165018	Sweet Box Cupcakes & Bake Shop	39.945457	-75.162746	Cupcake Shop

- The venue categories were then used to filter the data set such that only those categories containing the desired feature variable keywords were selected. The feature variables used are Indian restaurants, Yoga studios, and Public transport. The intention was to get a count of these venues by each neighborhood.

	Neighborhood	Indian Restaurant	Public Transport	Yoga Studio
0	Callow Hill, Chinatown, Market East, Midtown V...	0	0	1
1	East Passyunk	0	0	1
2	Fairmount, Rittenhouse Square	1	0	0
3	Logan Square	2	0	1
4	Northern Liberties	0	1	0
5	Old City, Society Hill	1	0	0
6	Powelton Village, Spruce Hill and Cedar Park, ...	1	0	1
7	Spring Garden	0	1	0

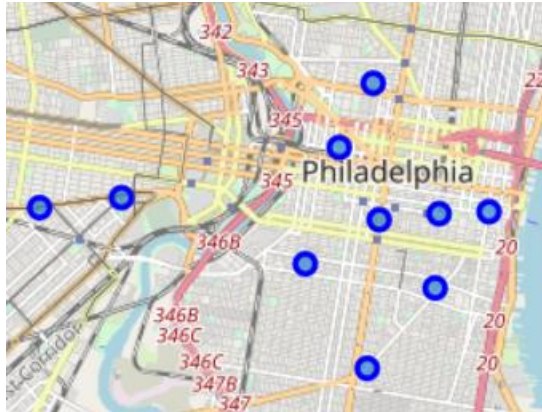
- Additional feature variables were then identified and added to the data set such as Population, Crime rate and Number of universities for the neighborhoods. In the absence of a freely available data set, the data was searched manually online and then appended to the dataframe.

	Neighborhood	Indian Restaurant	Public Transport	Yoga Studio	Population	Number of universities	Crime rate
0	Callow Hill, Chinatown, Market East, Midtown V...	0	0	1	24248.0	4.0	0.41
1	East Passyunk	0	0	1	46532.0	12.0	0.92
2	Fairmount, Rittenhouse Square	1	0	0	11140.0	5.0	1.26
3	Logan Square	2	0	1	12232.0	6.0	0.92
4	Northern Liberties	0	1	0	34112.0	3.0	0.65
5	Old City, Society Hill	1	0	0	19092.0	7.0	0.23
6	Powelton Village, Spruce Hill and Cedar Park, ...	1	0	1	20421.0	15.0	1.56
7	Spring Garden	0	1	0	NaN	NaN	NaN

Exploratory Analysis

The following exploratory analysis were used:

1. Map of Philadelphia with neighborhoods superimposed



2. Number of Venues and categories by each neighborhood

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Bella Vista, Pennsport, Queen Village, Washington Square West	89	89	89	89	89	89
Callow Hill, Chinatown, Market East, Midtown Village	93	93	93	93	93	93
East Passyunk	73	73	73	73	73	73
Fairmount, Rittenhouse Square	78	78	78	78	78	78
Graduate Hospital	33	33	33	33	33	33
Logan Square	100	100	100	100	100	100
Northern Liberties	23	23	23	23	23	23
Old City, Society Hill	100	100	100	100	100	100
Powelton Village, Spruce Hill and Cedar Park, University City	26	26	26	26	26	26
Spring Garden	48	48	48	48	48	48

Machine Learning

Since we were dealing with an exploratory having no training data, a clustering algorithm was used for prediction. The rationale behind using clustering was to cluster the neighborhoods based on similarity and then select the cluster containing the criteria closest to our liking.

	Neighborhood	Indian Restaurant	Public Transport	Yoga Studio	Population	Number of universities	Crime rate	Clus
0	Callow Hill, Chinatown, Market East, Midtown V...	0	0	1	24248	4	0.41	2
1	East Passyunk	0	0	1	46532	12	0.92	1
2	Fairmount, Rittenhouse Square	1	0	0	11140	5	1.26	0
3	Logan Square	2	0	1	12232	6	0.92	0
4	Northern Liberties	0	1	0	34112	3	0.65	1
5	Old City, Society Hill	1	0	0	19092	7	0.23	2
6	Powelton Village, Spruce Hill and Cedar Park, ...	1	0	1	20421	15	1.56	2
7	Spring Garden	0	1	0	16203	3	0.60	0

	Indian Restaurant	Public Transport	Yoga Studio	Population	Number of universities	Crime rate
Clus						
0	1.000000	0.333333	0.000000	12178.333333	4.333333	0.930000
1	0.000000	1.000000	0.500000	40322.000000	7.500000	0.785000
2	0.666667	0.000000	0.666667	21253.666667	8.666667	0.733333

Results

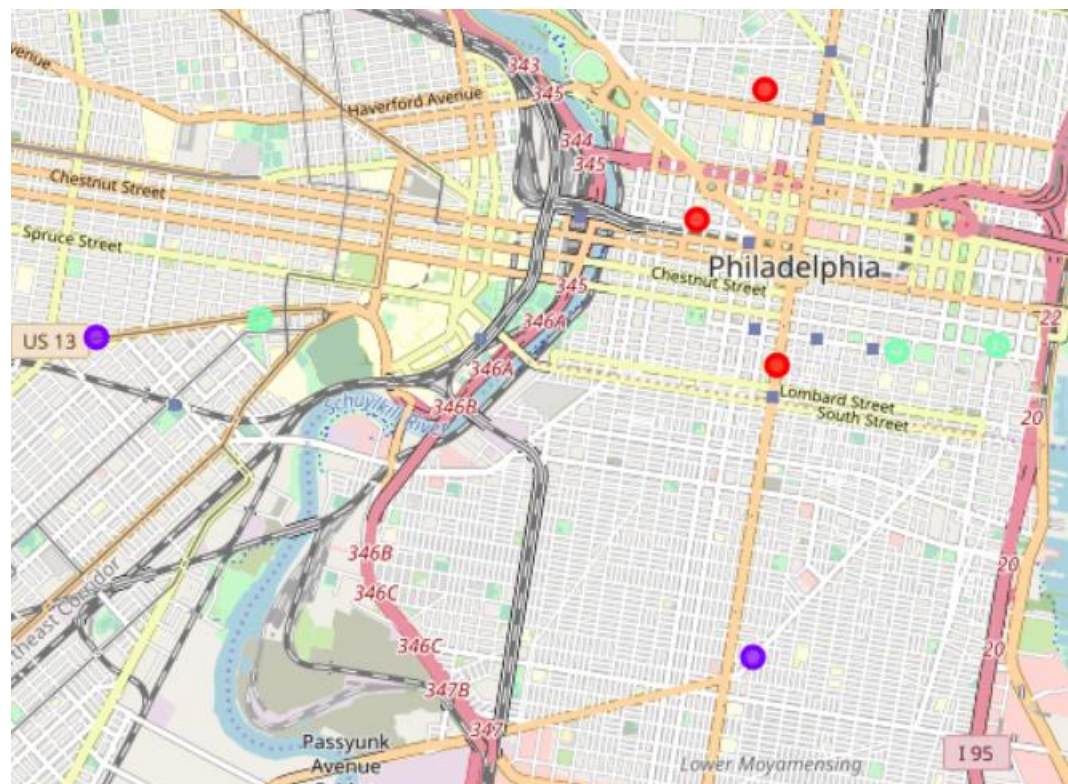
Based on the above data, we can conclude the following about each cluster:

Cluster 0: The least number of universities, highest crime rate and the most number of indian restaurants. This easily makes it the weakest choice for an Indian restaurant.

Cluster 1: Best public transport options, highest population, and a high number of universities. However, the most population in the cluster and also no other Indian restaurants i.e. no competition, makes it the best choice for an Indian restaurant

Cluster 2: No public transport, but the highest number of universities and the least crime rate makes this an alternative option.

	Neighborhood	Indian Restaurant	Public Transport	Yoga Studio	Population	Number of universities	Crime rate	Clus	Zipcode	Latitude	Longitude
0	Callow Hill, Chinatown, Market East, Midtown V...	0	0	1	24248	4	0.41	2	19107	39.947072	-75.154648
1	East Passyunk	0	0	1	46532	12	0.92	1	19148	39.926922	-75.167059
2	Fairmount, Rittenhouse Square	1	0	0	11140	5	1.26	0	19103	39.955870	-75.171830
3	Logan Square	2	0	1	12232	6	0.92	0	19102	39.946212	-75.165018
4	Northern Liberties	0	1	0	34112	3	0.65	1	19123	39.948016	-75.223406
5	Old City, Society Hill	1	0	0	19092	7	0.23	2	19106	39.947500	-75.146213
6	Powelton Village, Spruce Hill and Cedar Park, ...	1	0	1	20421	15	1.56	2	19104	39.949252	-75.209377
7	Spring Garden	0	1	0	16203	3	0.60	0	19130	39.964310	-75.166057



Discussion

The cluster containing the neighborhoods of East Passyunk and Northern Liberties would be the best neighborhoods to setup an Indian restaurant. NOTE: Due to the lack of other additional feature variables such as Property prices, we cannot confidently conclude this is the best option. However, the assertion is being made on the limited availability of features.

Conclusion

In conclusion, we were able to determine the best neighborhoods in the city of Philadelphia to setup an Indian restaurant which would help stakeholders in the decision making process.