Opening A Café In Philadelphia Report

Tierney Miller

August 3, 2020

1. Introduction/Business Problem

My project is centered on the city of Philadelphia, Pennsylvania. As with any expansive city, there is certainly no shortage of dining options, in quantity or variety. By leveraging available data, we can study where restaurants already exist, which areas of the city are more popular than other, or see more foot traffic, and analyze or predict certain areas that might benefit from a new restaurant or cafe. Some initial questions that I will address to make identifying a viable location for opening a cafe are:

1) What areas/neighborhoods of Philadelphia would be a promising, safe spot for a cafe?
2) What neighborhoods already have breakfast restaurants and cafes throughout them?
3) What neighborhoods have available retail space?
4) Are some neighborhoods more expensive than others to open a cafe in?

2. Data

The underlying data required to tackle this problem is location data on the city of Philadelphia. That includes names of different regions, neighborhoods, zip codes, etc. The first portion of this data is obtained here, which lists the 12 different "regions" of Philadelphia, and all neighborhoods within each.

The rest of the necessary location data- latitude and longitude coordinates- for each neighborhood is not readily available, thus I opted to employ Python's GeoPy software to ascertain these coordinates for all of Philadelphia's neighborhoods.

Upon acquisition of the location data for Philadelphia's neighborhoods, the Foursquare API will be useful to explore the neighborhoods of Philadelphia for what kind of cafes, coffee shops, restaurants, and diners are located in different neighborhoods. This will be the primary information for further analysis to address at least questions 1&2 listed in Section 1.

The last subset of data that will be explored contains crime rate and population information on neighborhoods throughout Philadelphia (http://data.philly.com/philly/crime/?) to provide further insight into potential locations for opening a café.

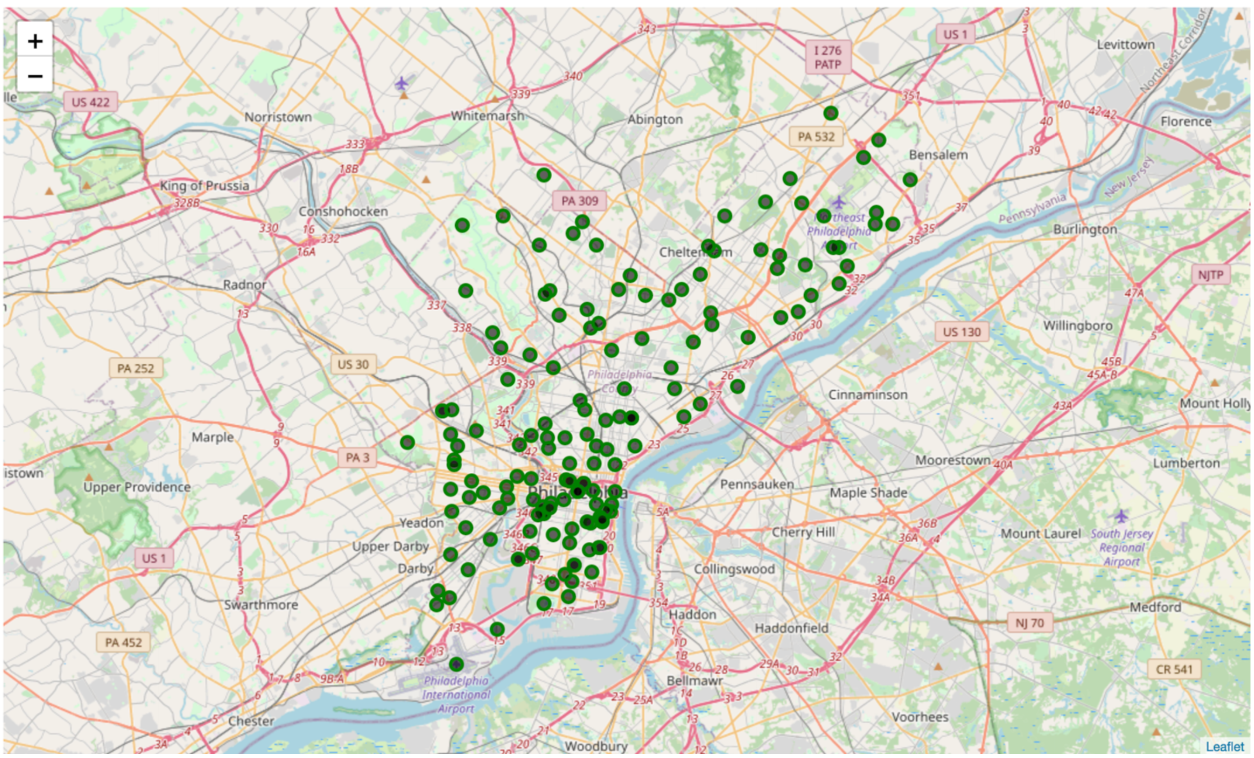
3. Methodology

From the initial list of Philadelphia regions and neighborhoods in combination with the GeoPy software to obtain central latitude and longitude coordinates of each neighborhood, the cleaned data set appears as follows:

	Region	Neighborhood	Latitude	Longitude
0	Center City	Callowhill	39.967005	-75.236701
1	Center City	Chinatown	39.953446	-75.154622
2	Center City	Elfreth's Alley	39.952712	-75.141980
3	Center City	Logan Square	39.958125	-75.170560
4	Center City	Naval Square	39.943959	-75.184315

From the list of 164 neighborhoods, the cleaned data set contains 154 neighborhoods. Removed from this are neighborhoods whose coordinates were unable to be obtained using GeoPy, as well as neighborhoods whose obtained coordinates were not correctly identified within Philadelphia, Pennsylvania.

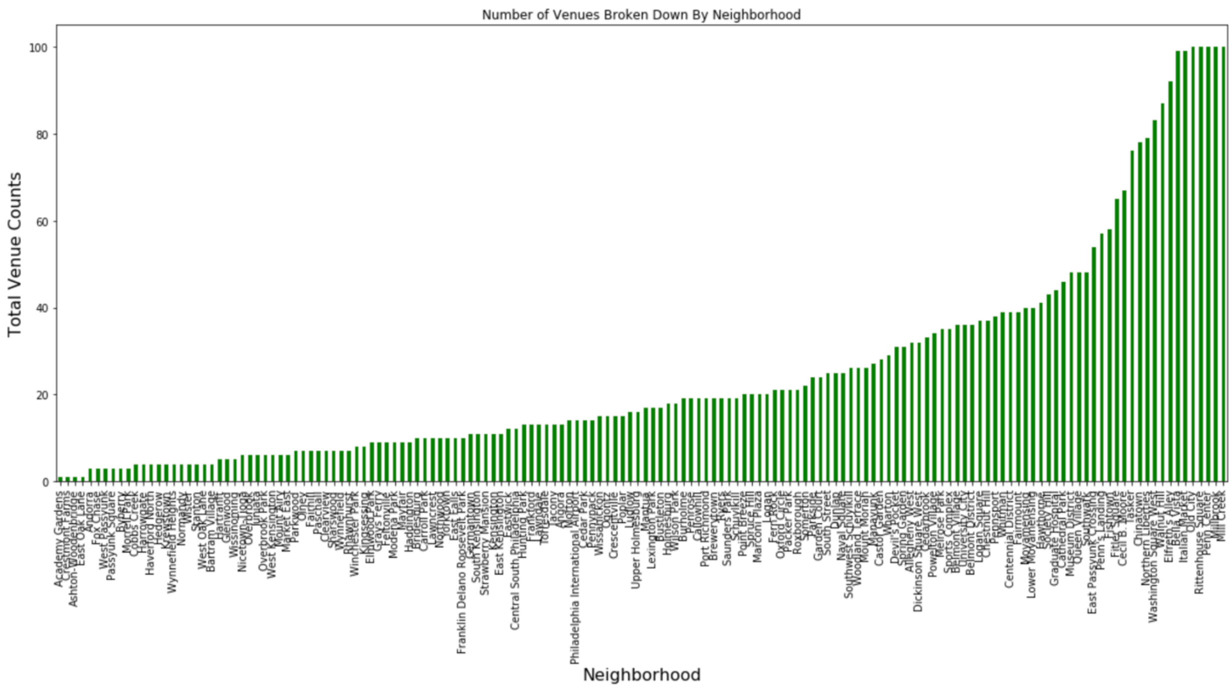
I used the Python Folium library to visualize these neighborhoods in Philadelphia: I overlaid a map of the Philadelphia region with the individual neighborhood locations from their (latitude, longitude) coordinates.



I then made a series of Foursquare API calls to obtain nearby venue information relative to each neighborhood shown on the map above. I applied certain criteria to these API calls, including a limit of 100 venues per call, and a radius of 750 meters from the central coordinate of each neighborhood. As evidenced by the table below, the breadth of the types of venues found by the Foursquare call ("Venue Category" column) are not limited to only food venues, let alone specifically breakfast spots or cafes.

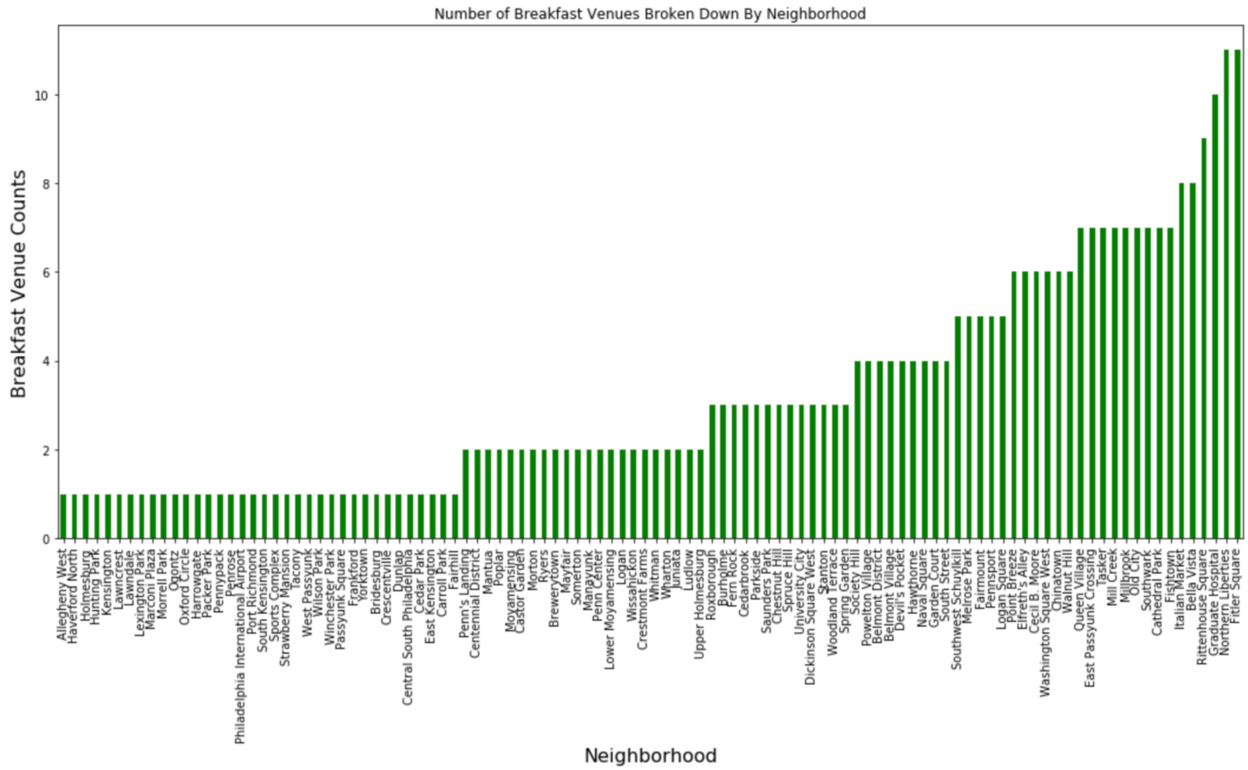
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Callowhill	39.967005	-75.236701	GiGi's Soul Food	39.969707	-75.238903	Southern / Soul Food Restaurant
1	Callowhill	39.967005	-75.236701	Just To Serve You	39.970552	-75.233362	Southern / Soul Food Restaurant
2	Callowhill	39.967005	-75.236701	Citizens Bank	39.966597	-75.232675	Bank
3	Callowhill	39.967005	-75.236701	Family Dollar	39.966712	-75.232872	Discount Store
4	Callowhill	39.967005	-75.236701	Save-A-Lot	39.966479	-75.233145	Grocery Store

Additionally, a distribution of venue counts per neighborhood shows that every neighborhood does not return the maximum number of venues for any given call made with the Foursquare API. In fact, less than a dozen neighborhoods returned the maximum number of venues, with close to two-thirds of the neighborhood calls not locating more than 50 venues each.



I further refined the collected venue data by limiting the results from all obtained venues (301 unique categories) down to a select few relevant to this inquiry (6 unique categories):

Coffee Shop, Donut Shop, Café, Breakfast Spot, Diner, Bagel Shop



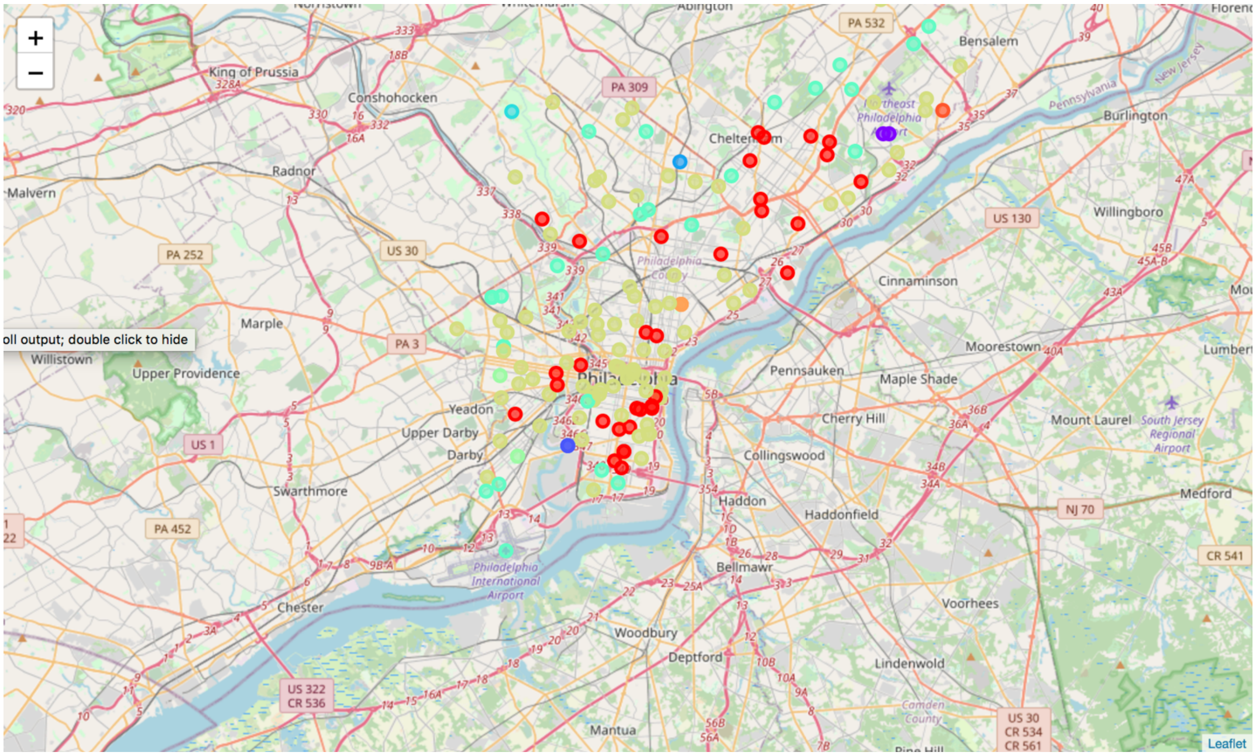
At this point, I determined the most common venue for each individual neighborhood, limited only to these six unique breakfast-related categories:

	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
0	Allegheny West	Donut Shop	Diner	Coffee Shop	Café	Breakfast Spot	Bagel Shop
1	Bella Vista	Coffee Shop	Diner	Breakfast Spot	Donut Shop	Café	Bagel Shop
2	Belmont District	Breakfast Spot	Donut Shop	Coffee Shop	Diner	Café	Bagel Shop
3	Belmont Village	Breakfast Spot	Donut Shop	Coffee Shop	Diner	Café	Bagel Shop
4	Brewerytown	Café	Donut Shop	Diner	Coffee Shop	Breakfast Spot	Bagel Shop

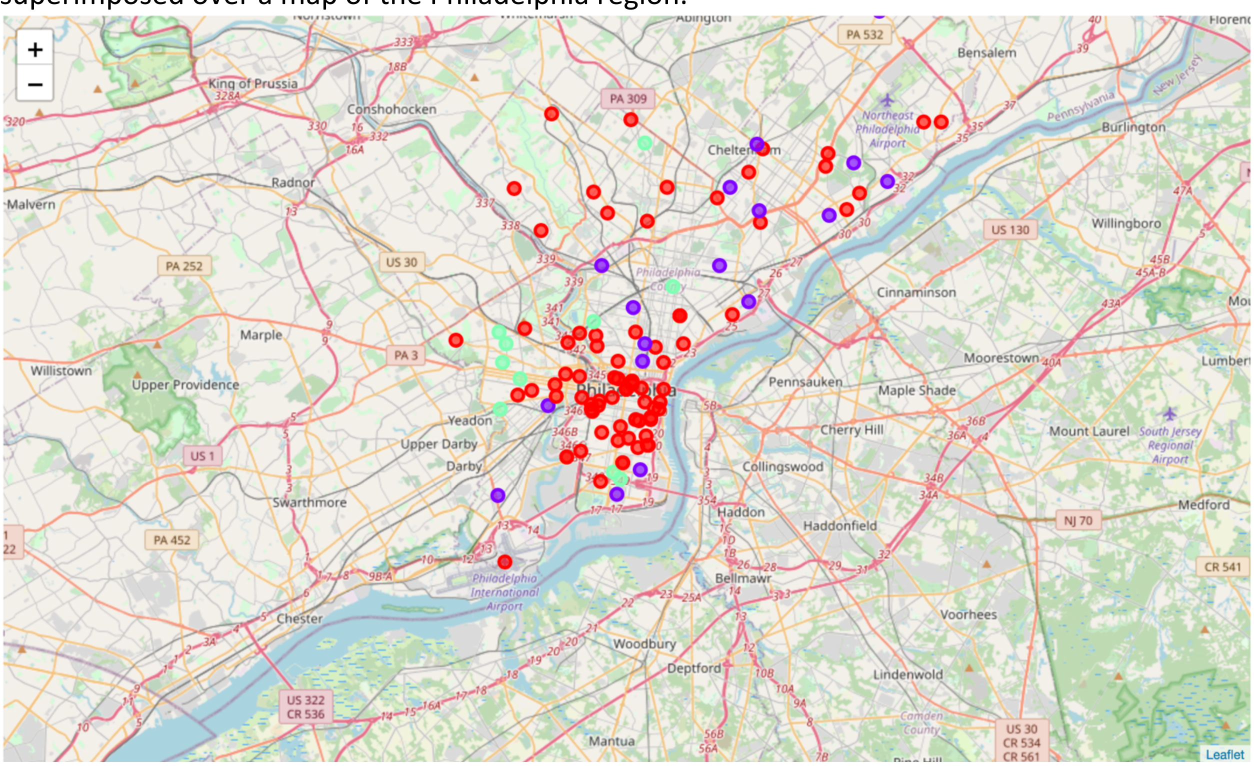
Given the select subset of venue types I am focusing on here, there is significant commonalities between all of the included neighborhoods. I chose to implement an unsupervised learning technique, K-means clustering, to group the various neighborhoods.

4. Results

Using the K-means clustering algorithm, I segmented the neighborhoods based on all of the venues identified with the Foursquare API. As there are a variety of venue types, I used a large number of clusters for this implementation of the algorithm (10 clusters).



I segmented the neighborhoods based upon their most common breakfast-type venues into three clusters. The results of this clustering can be observed in the map below, superimposed over a map of the Philadelphia region.



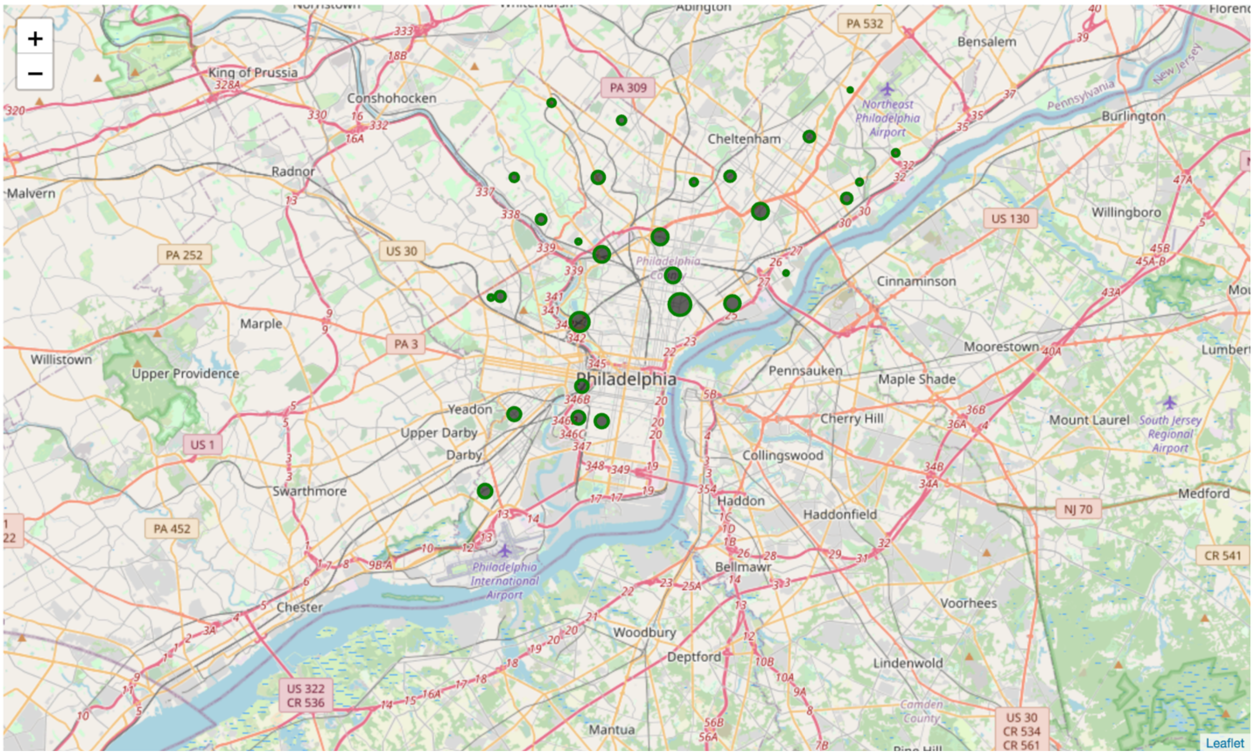
Individual clusters were further explored to understand which types of venues primarily compose each cluster. As shown in the following table, neighborhoods in the red cluster mainly had coffee shops as the most common venue, purple neighborhoods were commonly frequented by donut shops, while the neighborhoods in the green cluster almost exclusively possess breakfast spots as the most common breakfast venue.

Cluster	Venue Type
Red	Coffee Shops
Purple	Donut Shops
Green	Breakfast Spots

The final piece of analysis came from crime data obtained for some of the neighborhoods throughout Philadelphia:

	Neighborhood	Population	Violent Crime Rate*	Property Crime Rate*	Drug Crime Rate*
0	Allegheny West	17022	0.82	2.23	0.23
1	Bella Vista/Southwark	28191	0.18	1.31	0.00
2	Bridesburg	6289	0.00	0.64	0.00
3	Bustleton	40731	0.05	0.59	0.00
4	Cedarbrook/Stenton	33535	0.54	1.16	0.00
5	Center City	57668	0.80	4.14	0.33

From this data set, I focused on the property crime rate data, in conjunction with neighborhoods identified as containing breakfast venues from prior portions of this study. This covered a smaller subset of the data, which I visualized as a bubble map overlaid on the map of the Philadelphia region to gain a sense of the property crime rates in different regions of Philadelphia.



5. Discussion

The map of breakfast venue clustering overlaid on the Philadelphia region provides a few interesting insights. Firstly, we see that neighborhoods in the downtown Philadelphia area are populated by a large number of coffee shops. The north/northeast region has a much higher density of donut shops, but still some neighborhoods with more breakfast spots or coffee shops.

It should be noted that these predominant venue types represent the composite most popular venue for the top breakfast venue type of the various neighborhoods.

Property crime data did not exist for all neighborhoods analyzed in this study, so only general recommendations can be made about safer locations due to this lack of data. However, the trend observed is that the further you go from the downtown area, the lower the property crime rate. That being said, the downtown area has the highest concentration of neighborhoods with a large number of coffee shops, so if property safety isn't a high concern, a café could thrive well in certain downtown neighborhoods.

6. Conclusion

The insights derived in this report provide an overview of the predominant types of breakfast-related venues in operations in the various neighborhoods throughout Philadelphia. Specifically focusing on cafes, donut shops, coffee shops, diners, breakfast spots, and bagel shops, overall neighborhoods boast a number of donut shops, breakfast spots, and coffee shops. Additional commercial crime rate information coupled with general neighborhoods lacking in cafes narrows down the possible safe neighborhoods that might support and help a new café succeed in a suburb outside of downtown Philadelphia.