IBM / Coursera Applied Data Science Capstone Project – The Battle of Neighborhoods
Philadelphia, PA – In and out Food Choice

Introduction/Business Problem

English can be a difficult language for immigrants coming to United States. The county Philadelphia in PA provides English speaking classes for the immigrants. Immigrants would be interested to taste different types of food in Philadelphia and its surrounding counties. Immigrants who prefer to eat their preferred cuisine, finding the right spot to eat in Philadelphia can be difficult and frustrating.
This project provides information regarding most dense/frequent cuisines in the following counties in and around Philadelphia: Philadelphia, Berks, Bucks, Chester, Cumberland, Delaware, Lancaster, Montgomery in PA. Camden, Salem counties in NJ. It gives in which county will you find large number or even concentration of which type of restaurants. Where to eat Chinese food? Where to eat Italian food?

Description of the data

For this project, the following datasets are collected:

1) Philadelphia ESL class locations data with latitude and longitude coordinates is publicly available at https://www.opendataphilly.org.

   The first 5 rows of Philadelphia neighborhood data is below:

	Neighborhood	PostalCode	Latitude	Longitude
0	Philadelphia	19104	39.957033	-75.210202
1	Philadelphia	19145	39.930332	-75.174720
2	Philadelphia	19106	39.950638	-75.146764
3	Philadelphia	19107	39.957720	-75.157264
4	Philadelphia	19125	39.989663	-75.126142

2) Ten counties including and surrounding Philadelphia with latitude and longitude coordinates, is publicly available at https://www.opendataphilly.org.

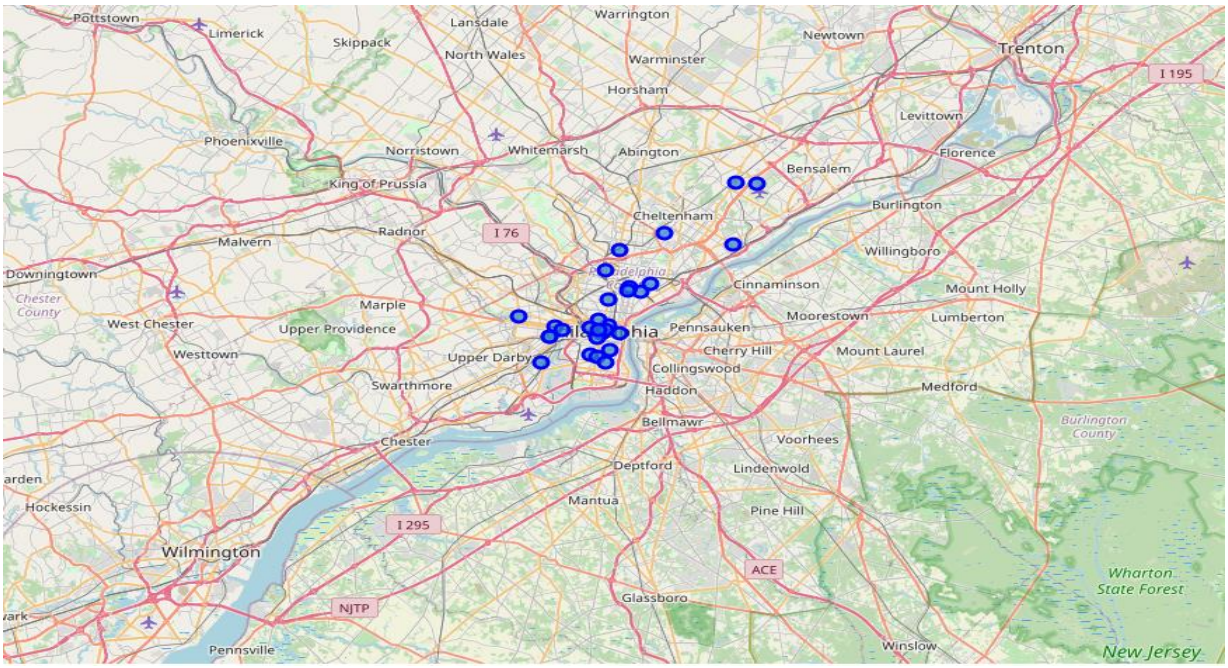
   The first 5 rows of ten counties data is below:

	City	State	Neighborhood	Latitude	Longitude	PostalCode
0	Philadelphia	Pennsylvania	Philadelphia	40.0077	-75.1339	19154
1	Lancaster	Pennsylvania	Lancaster	40.0421	-76.3012	17603
2	Reading	Pennsylvania	Berks	40.3400	-75.9267	19604
3	Pottstown	Pennsylvania	Montgomery	40.2507	-75.6444	19464
4	Levittown	Pennsylvania	Bucks	40.1537	-74.8530	19054

3) Foursquare API about restaurants in and around Philadelphia. By passing the latitude and longitude into Foursquare API, we can explore venue information like Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category. The number of places per neighborhood parameter was set to 30 and the radius parameter to 700. The restaurants data within and around Philadelphia is extracted to show the density of restaurants in each county, type of cuisine.

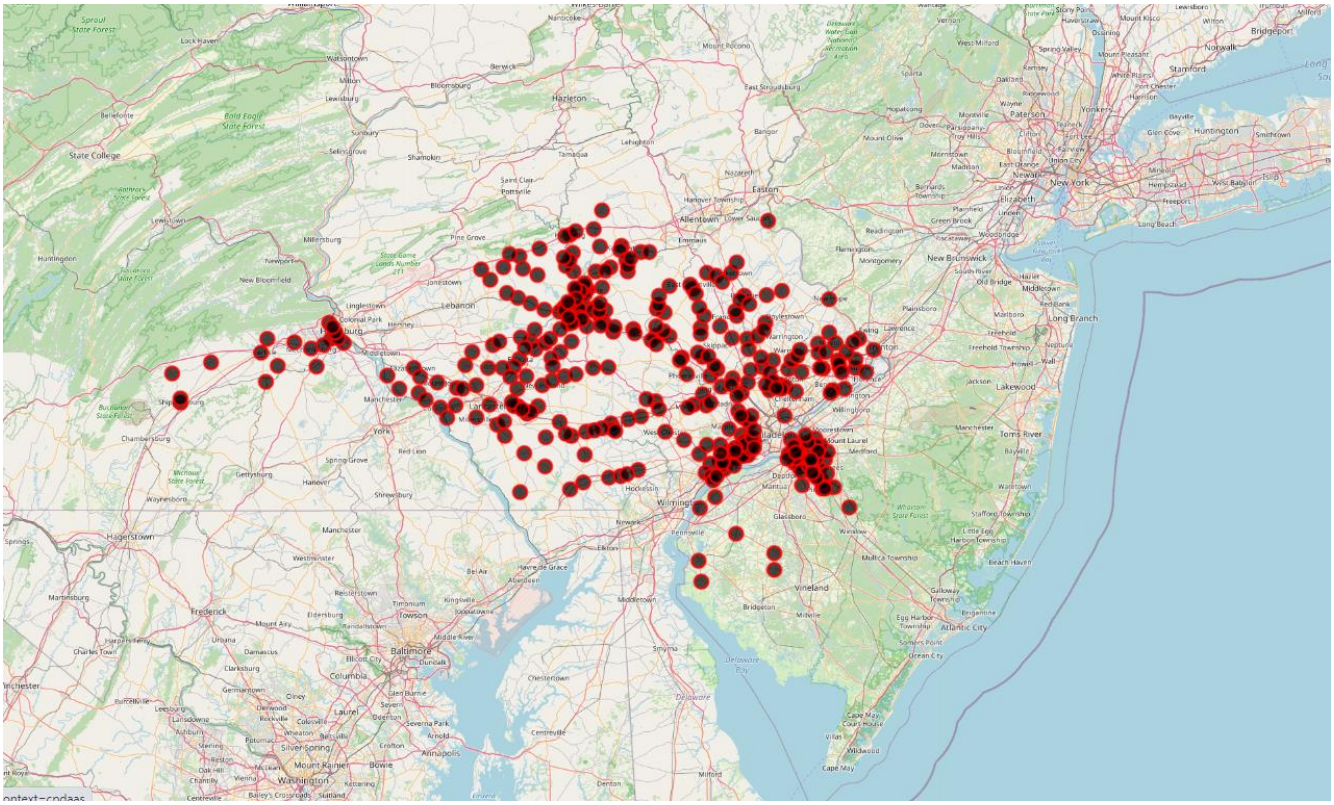
Methodology

In this section, I will describe the data analysis and how I used the data to yield the results. Data is collected from the source, Opendataphilly (https://www.opendataphilly.org). Information was filtered out with web scraping and data wrangling. Data frame of Philadelphia providing ESL classes with the following parameters: postal code, neighborhood, latitude, and longitude was created. Using the folium package, ESL classes (blue circles) in Philadelphia is represented on a map.



Next, FourSquare API is used to retrieve the restaurants within Philadelphia County. There are 204 restaurants which can be grouped into 41 unique categories (such as Mexican, Chinese, Italian and so on). Using seaborn/matplotlib packages, the data of ten most frequented restaurants in Philadelphia was plotted on a bar graph. Immigrants attending ESL classes will be able to decide the most common cuisine type available in Philadelphia.

Additionally, restaurant density in counties around Philadelphia are also reported in this project so that people can compare the type of cuisine available in and around counties of Philadelphia. Folium- Python visualization library is used to visualize ten counties around Philadelphia over an interactive leaflet map. Red circles on the map are indicative of the following counties: Philadelphia, Berks, Bucks, Chester, Cumberland, Delaware, Lancaster, Montgomery, Camden(NJ), Salem(NJ).



FourSquare API and K-means clustering methods are used to retrieve the ten most common restaurant type in these different counties. To find the clusters of restaurant types, the data frame is transformed with the restaurant venues associated to the neighborhoods, by one-hot encoding (0/1). Then, grouping is used to show the frequency of each restaurant category. Following which, a data frame is created to show the most common restaurant types for each county. Finally, an unsupervised machine learning algorithm, k-means, k to be 5 are selected to derive the most common cuisine type.

Results

People can compare the type of cuisine available in and around counties of Philadelphia. Based on the latitude and longitude of ESL classes centers (hence making the program more catered to English learning speakers or immigrants), the most type of cuisine available is Mexican. This is followed by Chinese, Italian, Seafood, Vietnamese, Fast Food, Mediterranean, Asian, Indian, and Ethiopian, in that respective order. The following data is shown on the bar graph, below.

Top ten cuisine category of the first 5 neighborhoods is in the table below.

	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	Berks	American Restaurant	Fast Food Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Sushi Restaurant
1	Bucks	American Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Mexican Restaurant	Fast Food Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
2	Camden	Italian Restaurant	Chinese Restaurant	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Japanese Restaurant	Asian Restaurant	Restaurant	Indian Restaurant	Seafood Restaurant
3	Chester	Italian Restaurant	American Restaurant	Mexican Restaurant	Chinese Restaurant	Fast Food Restaurant	Sushi Restaurant	Restaurant	Japanese Restaurant	Mediterranean Restaurant	Seafood Restaurant
4	Cumberland	American Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Korean Restaurant	Fast Food Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant

K Means and clustering and visualization The counties with the most common venues (restaurants) are assigned into five different cluster labels of 0 to 4. These clusters are identified with five different colors, each showing the type of cuisine available.

Cluster0: contains Delaware, PA is indicated by red color in the below map. Fast food and Chinese cuisines are more prevalent, in the respective order.

Neighborhood	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
Delaware	0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	New American Restaurant
Delaware	0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	New American Restaurant
Delaware	0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	New American Restaurant
Delaware	0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	New American Restaurant
Delaware	0	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Southern / Soul Food Restaurant	New American Restaurant

Cluster1: contains Salem in NJ is indicated by Violet color in the below map. This cluster shows high density American cuisine, followed by Mexican and Italian.

Neighborhood	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
Salem	1	American Restaurant	Mexican Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant
Salem	1	American Restaurant	Mexican Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant
Salem	1	American Restaurant	Mexican Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant
Salem	1	American Restaurant	Mexican Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant
Salem	1	American Restaurant	Mexican Restaurant	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant

Cluster2: contains Philadelphia, PA is indicated by blue color in the map. This cluster shows a high density of Fast food restaurants followed by Spanish cuisine.

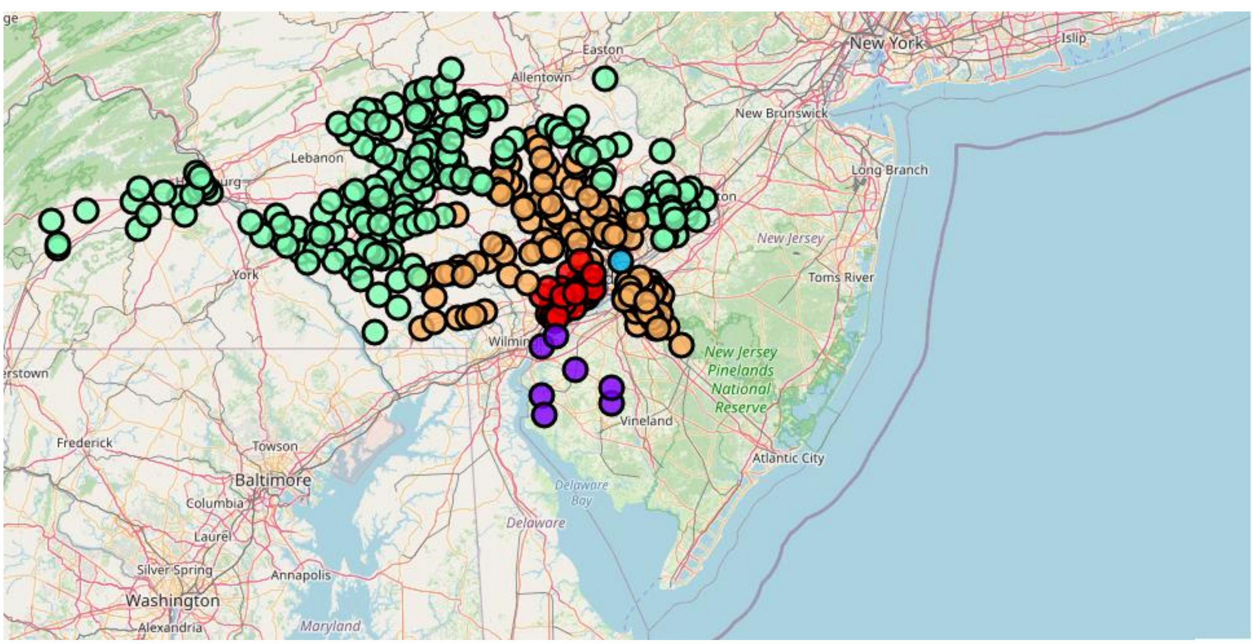
Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Philadelphia	2	Fast Food Restaurant	Spanish Restaurant	South American Restaurant	Mexican Restaurant	Middle Eastern Restaurant	New American Restaurant	Polish Restaurant	Ramen Restaurant	Restaurant	Seafood Restaurant

Cluster3: contains Berks, Bucks, Cumberland and Lancaster counties in PA as indicated by green color in the map below. This cluster shows a high density of American followed by Italian restaurants.

Neighborhood	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
Lancaster	3	American Restaurant	Italian Restaurant	Fast Food Restaurant	Restaurant	Chinese Restaurant	French Restaurant	Asian Restaurant	Mexican Restaurant	Indian Restaurant	Vietnamese Restaurant
Berks	3	American Restaurant	Fast Food Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Sushi Restaurant
Bucks	3	American Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Chinese Restaurant	Mexican Restaurant	Fast Food Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
Cumberland	3	American Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Korean Restaurant	Fast Food Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant
Cumberland	3	American Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Korean Restaurant	Fast Food Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Southern / Soul Food Restaurant

Cluster4: contains Montgomery, Chester, Camden(NJ) is indicated by orange color in the map below. It has a high density of Italian cuisine followed by American.

Neighborhood	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
Montgomery	4	Italian Restaurant	American Restaurant	Chinese Restaurant	Mexican Restaurant	Restaurant	Fast Food Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Seafood Restaurant
Montgomery	4	Italian Restaurant	American Restaurant	Chinese Restaurant	Mexican Restaurant	Restaurant	Fast Food Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Seafood Restaurant
Montgomery	4	Italian Restaurant	American Restaurant	Chinese Restaurant	Mexican Restaurant	Restaurant	Fast Food Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Seafood Restaurant
Chester	4	Italian Restaurant	American Restaurant	Mexican Restaurant	Chinese Restaurant	Fast Food Restaurant	Sushi Restaurant	Restaurant	Japanese Restaurant	Mediterranean Restaurant	Seafood Restaurant
Montgomery	4	Italian Restaurant	American Restaurant	Chinese Restaurant	Mexican Restaurant	Restaurant	Fast Food Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Seafood Restaurant



Discussion

All the tools are available for free, we just have to get to know the available open-source packages and learn to use them for scraping, cleaning, handling, transforming and visualizing the data. With these tools, there are many exciting data science uses. Since the data can be collected online, we should always be double check the sources and use the above project as a starting point.

Conclusion:

This analysis will provide a robust comprehensive tool to access and compare cuisine types available in and around Philadelphia, PA. All the above is dependent on the adequacy and accuracy of the Foursquare data.