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

Introduction/Business problem

- 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.
- 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?

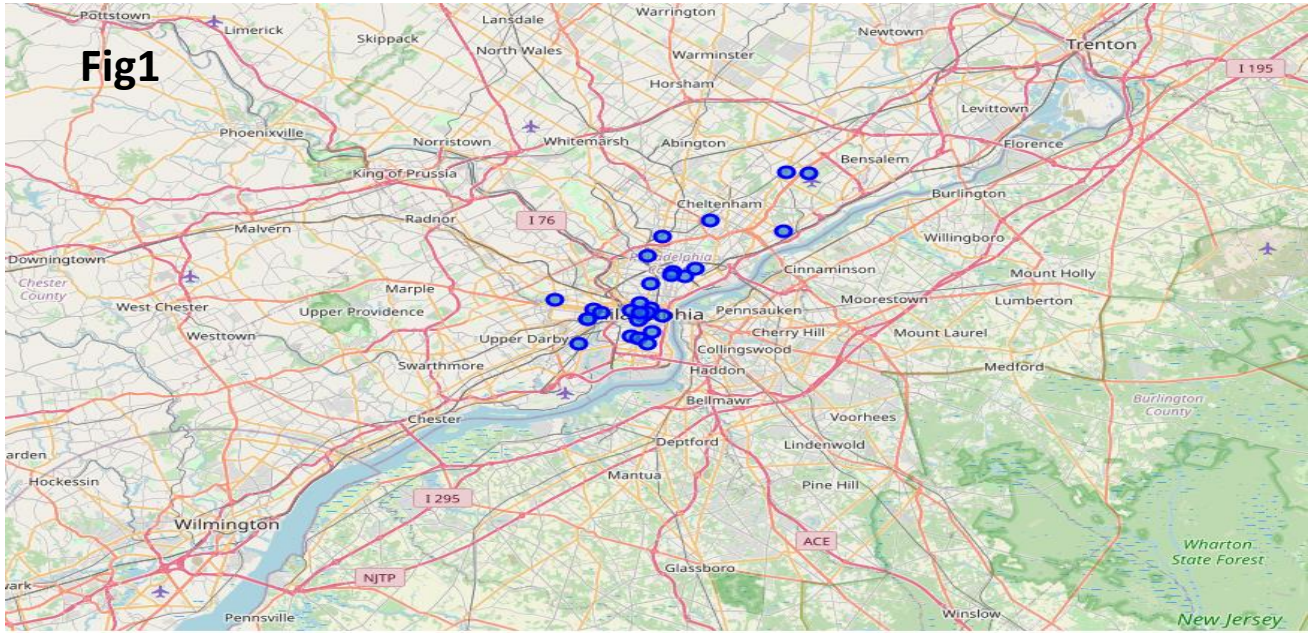
Data

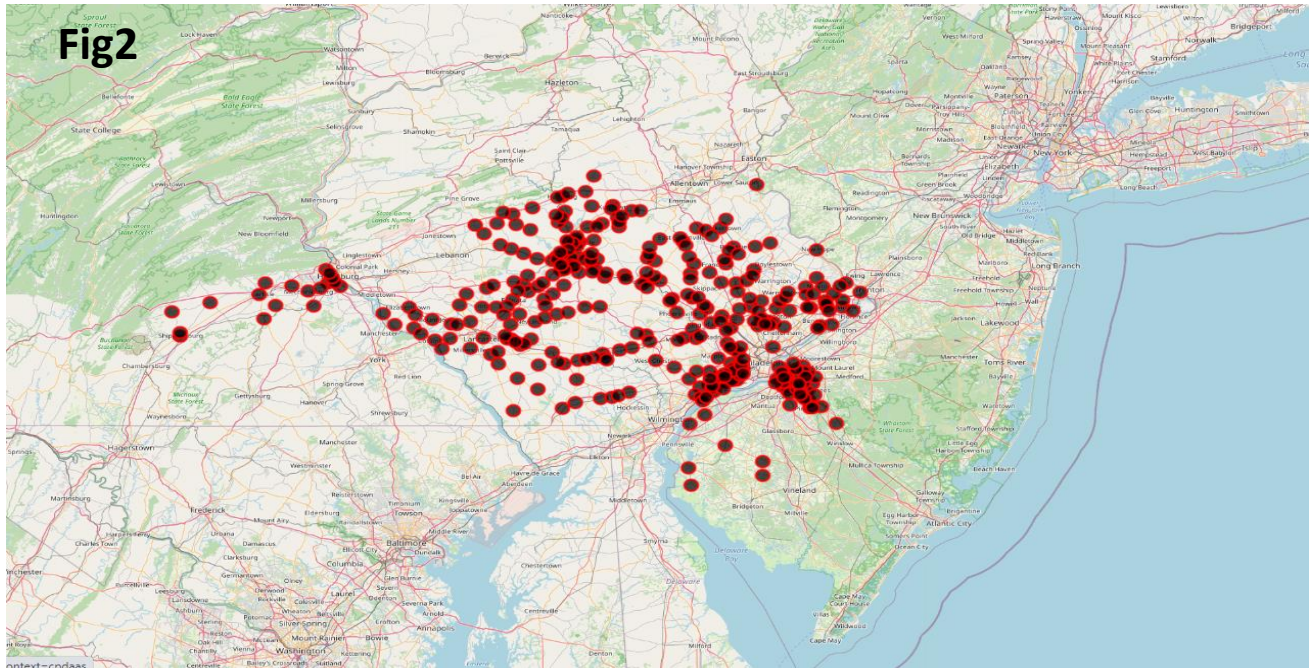
- Data required
  - Philadelphia ESL class locations
  - Ten counties including and surrounding Philadelphia
  - Latitude and Longitude coordinates of the neighborhoods

- Sources of data
  - https://www.opendataphilly.org.
  - Geocoder package for latitude and longitude coordinates
  - Foursquare API for venue data

Methodology

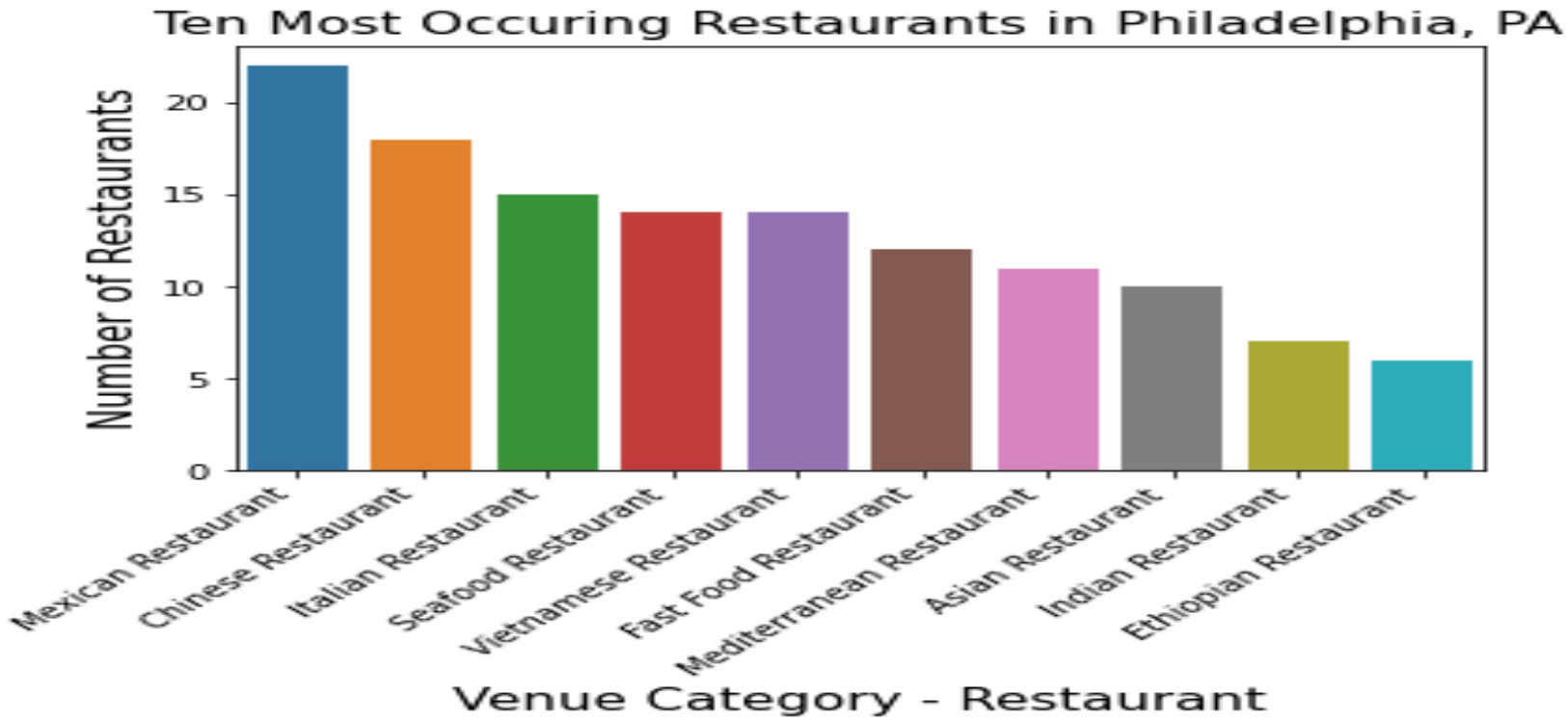
- Clean and process the data (https://www.opendataphilly.org) into a data frame.
- Using Foursquare, locate all venues and then filter by restaurants.
- Clustering method is used to retrieve the ten most common cuisine type in ten different counties.
- Map of Philadelphia providing ESL classes (Fig1)
- Map of ten counties in and around Philadelphia, Berks, Bucks, Chester, Cumberland, Delaware, Lancaster, Montgomery, Camden, Salem (Fig2)





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, the most type of cuisine available is Mexican.



Results

The counties with the most common 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.

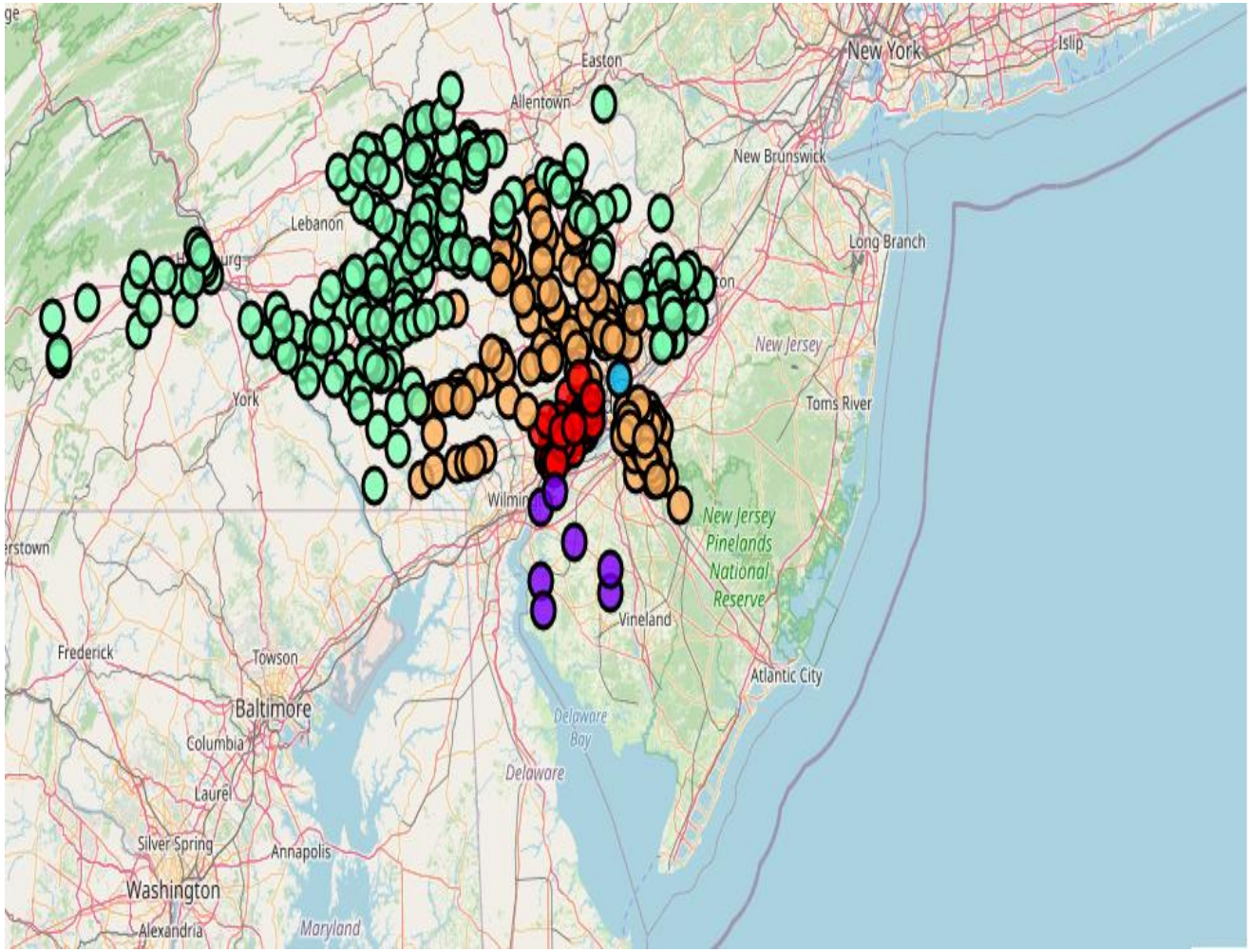
Cluster0: contains Delaware, PA is indicated by red color in the map. Fast food and Chinese cuisines are more prevalent.

Cluster1: contains Salem in NJ is indicated by Violet color in the map. This cluster shows high density American cuisine

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

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

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



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.