Capstone Project
# Battle of Neighborhoods: Philadelphia Region

Introduction

- This presentation is an addendum to the report addressed to the Management of "Burger Chain, Inc."

- In this project, we explore and analyze towns surrounding greater Philadelphia in order to help evaluate potential locations for future Burger Chain locations

- We leverage several Python packages, the Foursquare API, interactive plots, and unsupervised machine learning (i.e, clustering analysis) to develop potential insights about the neighborhoods and towns
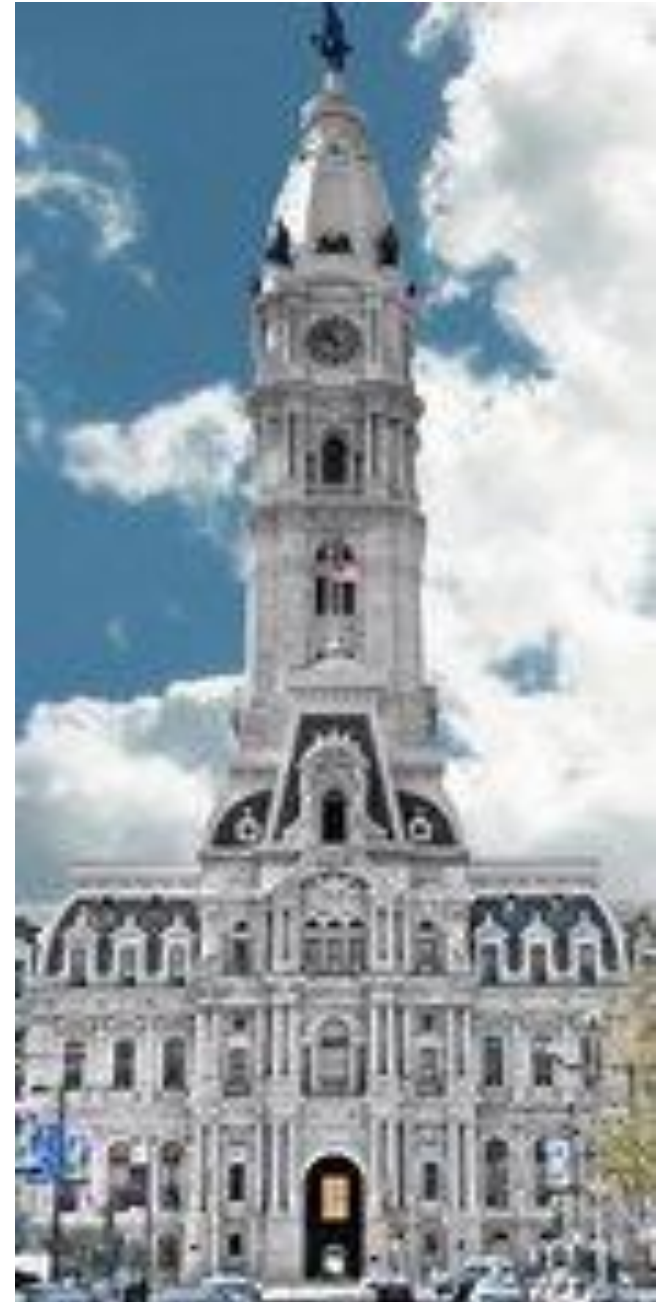
Data
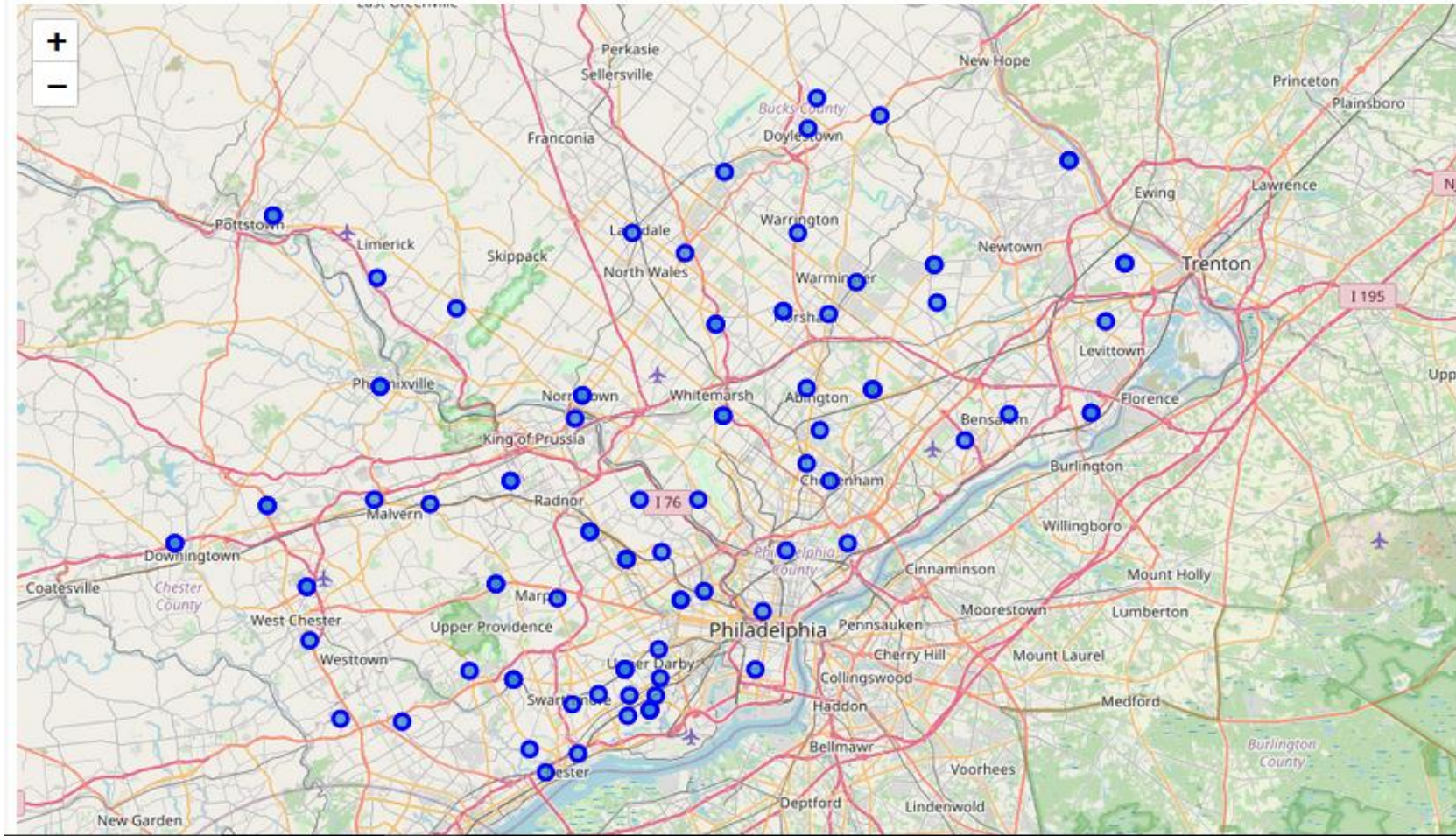
- The data used in this project are sourced from Open Data Philly (the official repository of data for the City of Philadelphia), Kaggle, and Foursquare

- The final dataset consists of 98 observations and 10 variables, representing features of the towns and neighborhoods
  - House Prices
  - Crime Rate
  - Population Change
 - Zip code
  - and others

Data

- The data set was pre-processed, visualized and analyzed in Python using the pandas (analysis), seaborn and matplotlib (visualization), folium (maps) and sklearn (machine learning) libraries

- Latitude and Longitude were extracted using the geopy (geo-coding) library

- A portion of the final dataset is show in the table below:

|	HousePrice	HsPrc ($10,000)	CrimeRate	MilesPhiladelphia	PopChg	Name	County	ZipCode	Latitude	Longitude
0	140463	14.0463	29.7	10.0	-1.0	Abington	Montgomery	19001	40.126918	-75.126680
1	113033	11.3033	24.1	18.0	4.0	Ambler	Montgomery	19002	40.174003	-75.212356
2	124186	12.4186	19.5	25.0	8.0	Aston	Delaware	19014	39.887240	-75.563472
3	110490	11.0490	49.4	25.0	2.7	Bensalem	Bucks	19020	40.108163	-74.938024
4	79124	7.9124	54.1	19.0	3.9	Bristol Borough	Bucks	19007	40.110056	-74.860923

Exploratory Data Analysis

- In our exploration of 98 towns, we find a negative correlation between House Prices and Crime Rates…

- … modest correlation between Population Growth and House Prices, and

- Median House price of $140k, distance from center city Phil. of 20 miles, and population growth of 1.6%
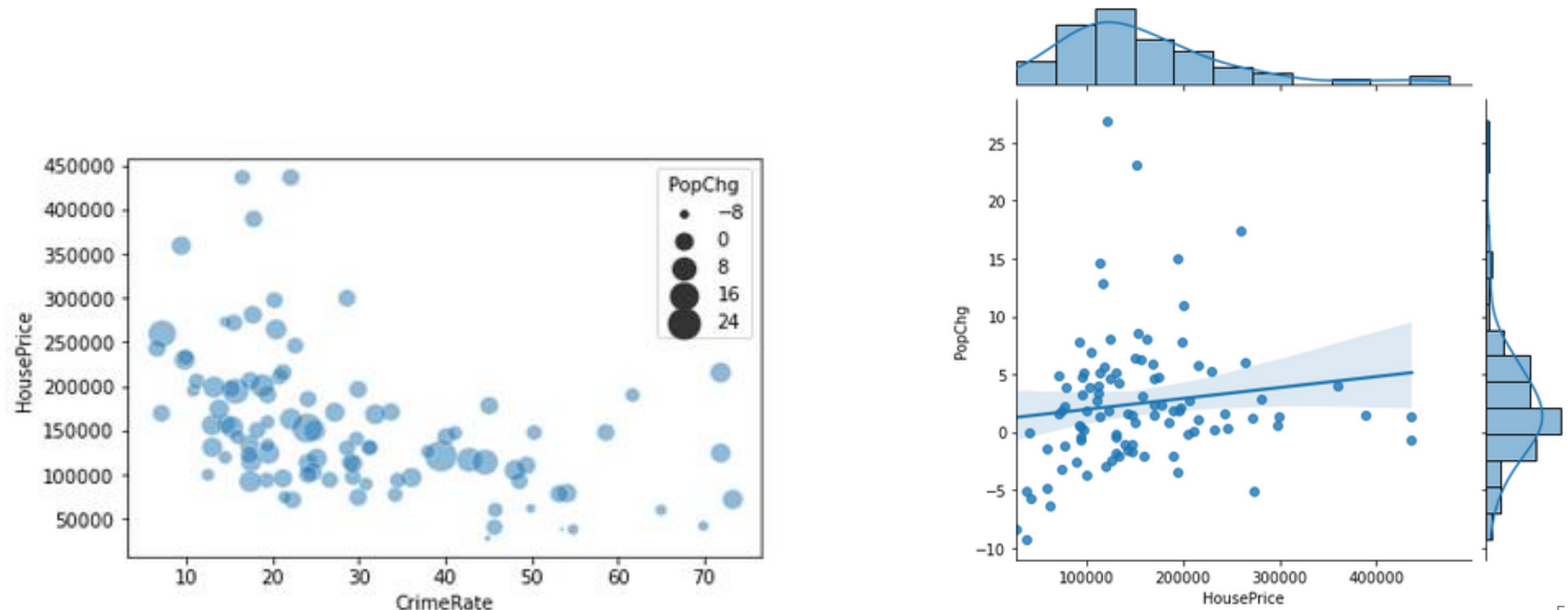
Methodology

- Foursquare is a leading, independent data platform and location search information service

- After obtaining coordinate data using geopy, we used the Foursquare API to obtain the most prominent local businesses, venues and features for each town
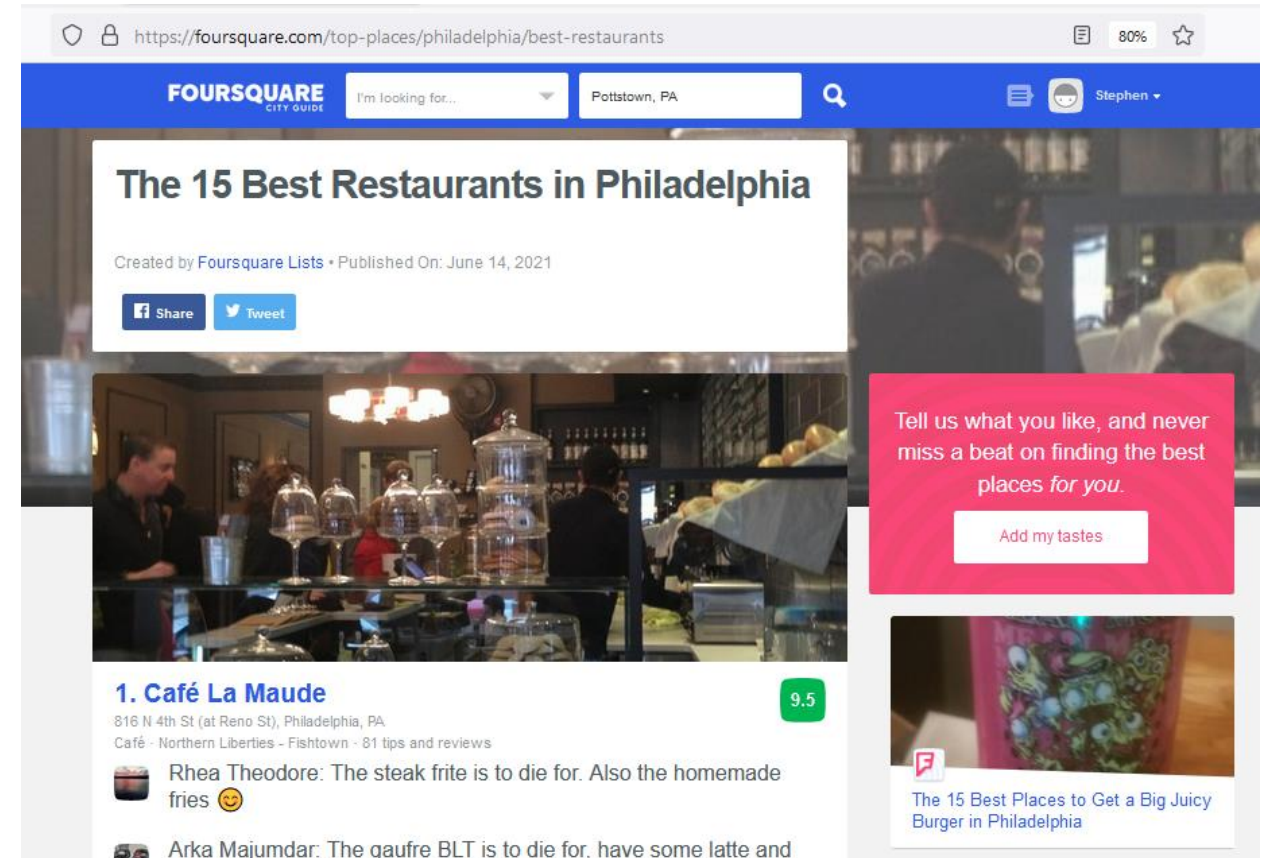
Clustering Analysis

- The data returned from Foursquare were wrangled into a dataframe showing the 10 most common type of venue for each town

- These 10 features were used as the basis for clustering analysis, which grouped the towns by similarity into nine clusters

	Name	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	Abington	Pizza Place	Fried Chicken Joint	Pharmacy	Convenience Store	Sandwich Place	Smoke Shop	Brewery	Deli / Bodega	Insurance Office	Home Service
1	Ambler	Other Repair Shop	Home Service	Construction & Landscaping	Building	Dive Bar	Doctor's Office	Food	Flower Shop	Fish Market	Fish & Chips Shop
2	Aston	Construction & Landscaping	Yoga Studio	Event Service	Food	Flower Shop	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm
3	Bensalem	Pizza Place	Sandwich Place	Bank	Pharmacy	Music Venue	Pub	Convenience Store	Bagel Shop	Playground	Japanese Restaurant
4	Bristol Borough	Liquor Store	Discount Store	Italian Restaurant	Gas Station	Park	Eye Doctor	Flower Shop	Fish Market	Fish & Chips Shop	Fast Food Restaurant

Results & Discussion

- The nine clusters are illustrated on an interactive map of Philadelphia area

- Clusters can be used to rule out or 'rule in' groups of towns as candidates for further analysis and potential location expansion
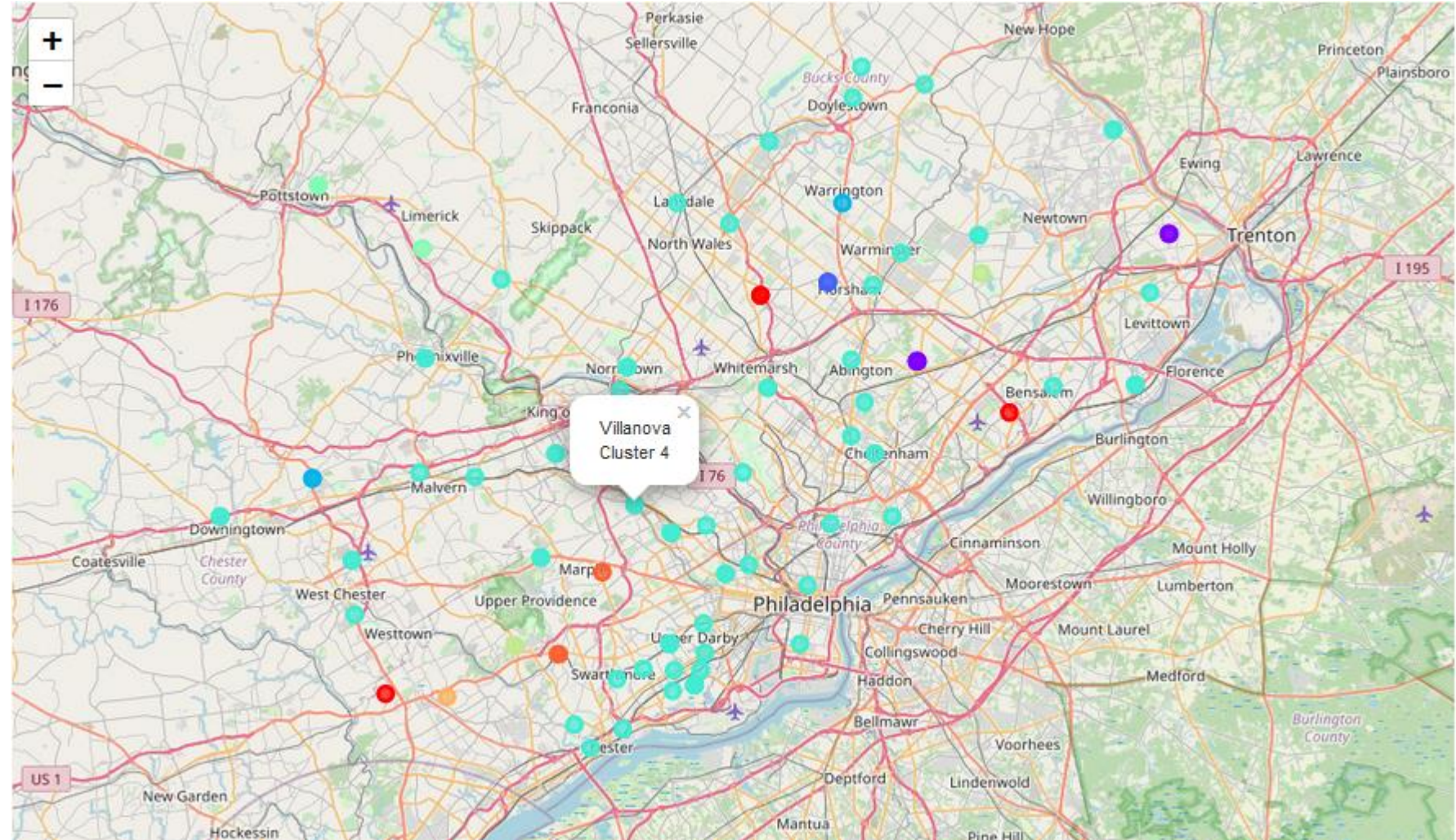
Conclusion

- In this project, we investigated traditional metrics such as median House Price, Crime Rate and Population growth, to develop a big picture of surrounding towns

- We obtained detailed 'street level' information from Foursquare to develop a granular understanding of the types of businesses and venues in each town

- Finally, we used clustering analysis to group the 98 towns into 9 clusters with the goal of discovering similarities between towns that could help in the process of accepting or rejecting candidate locations

- Based on the clustering analysis, there are several potentially interesting locations, particularly in clusters 3, 4 and 8, that would warrant further investigation and could be suitable for future expansion