Report to the Executive Management of “Burger Chain, Inc.”

From: Stephen Choi

Date: June 17, 2021

Subject: Philadelphia Region Towns Exploration

Table of Contents

Image of Philadelphia Waterworks and Art Museum

1. Introduction

This project will be an exploration and analysis of suburban Philadelphia, Pennsylvania, USA. The intended audience for the final report is the management of Burger Chain, Inc. who are looking to expand their locations in PA. In this project we will leverage the Foursquare API, along with interactive maps, clustering analysis and other exploratory techniques, to learn about the neighborhoods and the associated market opportunity. Identifying the right expansion locations for Burger Chain, Inc. could make the difference toward a successful expansion and therefore to the future of the business.

2. Data

As part of this project, we will use the Foursquare API to obtain data on types of restaurants, retail venues and other industry, demographic and other features in the towns and regions around Philadelphia and surrounding area. The coordinates of towns and regions of suburban Philadelphia are publicly available. Certain key features will be explored, such as new upscale housing and commercial real estate developments, and competitive and complementary food and restaurant locations. We will seek to understand where successful branded companies are

located, and what the demographic trends are in different neighborhoods. Nearby golf courses, parks, trails and other green spaces, good schools, colleges and universities and economic and population growth may be factors to consider.

Additional datasets may also be used such as the Crime Dataset from Kaggle (see https://www.kaggle.com), and one or more data sets from Open Data Philly, the official repository for the City of Philadelphia(https://www.opendataphilly.org). Dataset addressing one or more of the following topics may also be explored:

- Population density, regional wealth indices
- Parks, Playgrounds, Picnic sites
- Tobacco retailer permits, Air quality index
- Choice neighborhoods, Neighborhood food retail and neighborhood resources
- Land use, Vacant property indicators
- Crime incidents, Shooting victims, Fatal crashes

The objective will be to explore and analyze the various data in new ways in order to develop an insightful picture of the Philadelphia region broadly and potential specific areas where branches could be located.

3. Methodology

In this project, we started by exploring several key features of 98 suburban towns in the Philadelphia and surrounding area: median house price, crime rate, and population growth. These features were sorted and visualized through plots (using the Seaborn package) and through interactive maps (using the Folium package).

Latitude and Longitude coordinates were obtained using the Geopy package by using the zip codes of each town. Once the coordinates were checked, data concerning local businesses were then retrieved using the Foursquare API for each suburban town. A 1.0 km radius was used as the search radius. The retrieved records were counted to get a sense for the density of establishments at each location, and top ten records were inspected for each town.

Using the information about venues and local businesses, a clustering analysis of the town was performed to explore similarity of the towns according to the Foursquare data. A range of k-values from 3:10 were tried and ultimately a value of k=9 was used based on an inspection of the results.

Finally, the seven clusters were examined to see what types of features stood out as salient types of businesses for each location. These features could be mapped and used as a starting opint for a 'street level' exploration of possible expansion locations for the client, Burger Chain Inc.

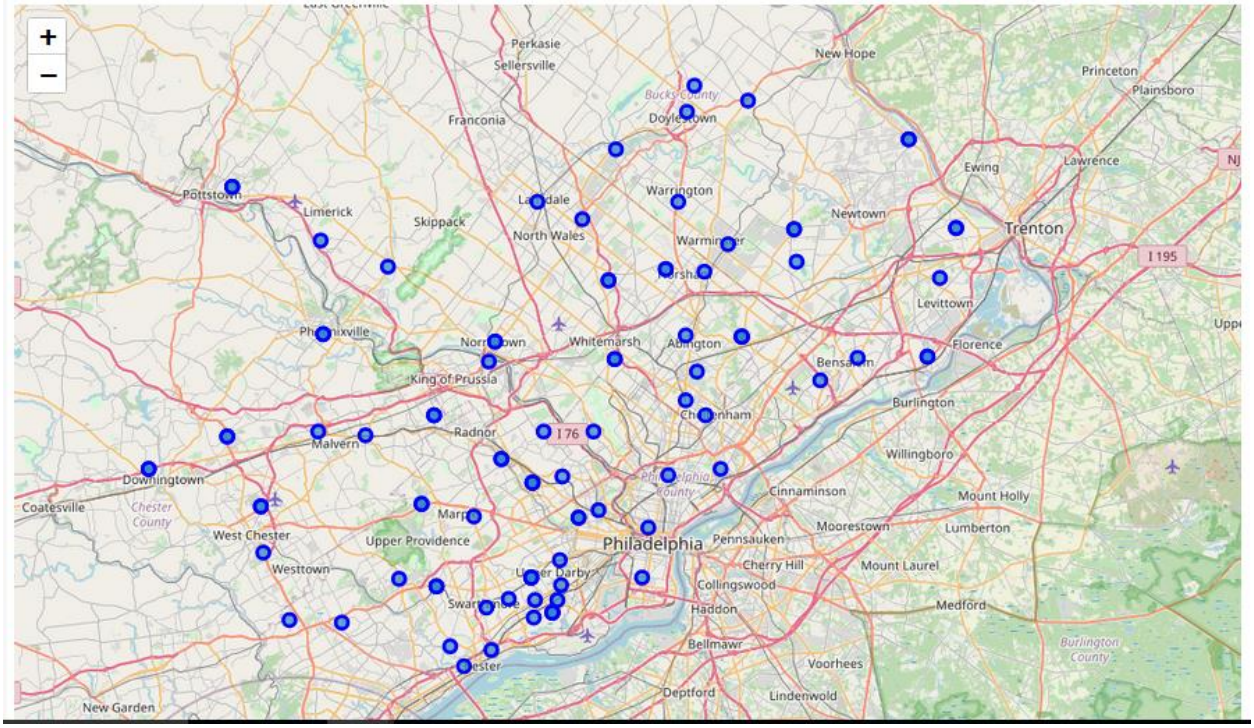


Figure 1. Map of Philadelphia, PA with 98 towns highlighted in blue

4. Analysis

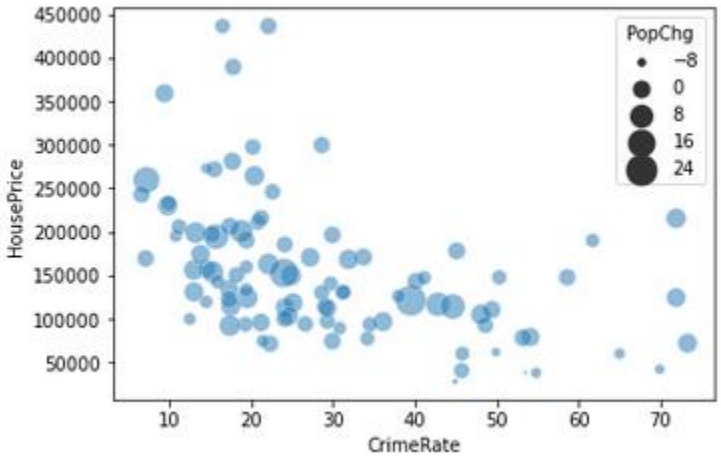
We first performed an exploratory data analysis to understand the original eight variables and 98 observations contained in the data set. A partial numerical summary of key variables is provided below:

Table 1. Numerical summary of the Philadelphia-region-towns data set

	HousePrice	CrimeRate	MilesPhiladelphia	PopChg
count	99.000000	99.000000	99.000000	98.000000
mean	157835.606061	32.576768	19.282828	2.479592
std	86644.159973	37.540908	9.099552	5.632307
min	28000.000000	6.600000	0.000000	-9.200000
25%	98505.000000	17.400000	13.000000	-0.625000
50%	140463.000000	24.100000	20.000000	1.600000
75%	195830.000000	39.800000	24.500000	4.775000
max	475112.000000	366.100000	50.000000	26.900000

We then generated several plots and graphs to examine the variable distributions and their correlations. A bubble plot showing median House Price versus Crime Rate with bubble size representing Population Change is shown in figure 2 below. A negative correlation between House Price and Crime Rate is evident.

Figure 2. Bubble plot of House Price, Crime Rate, and Population Change



A scatterplot of Population Change (%) versus median House Price is displayed in Figure 3. Both variable distributions are slightly skewed, as shown by their histograms in the margin. A slight positive correlation exists between the two variables.

Figure 3. Scatter plot of Population versus House price

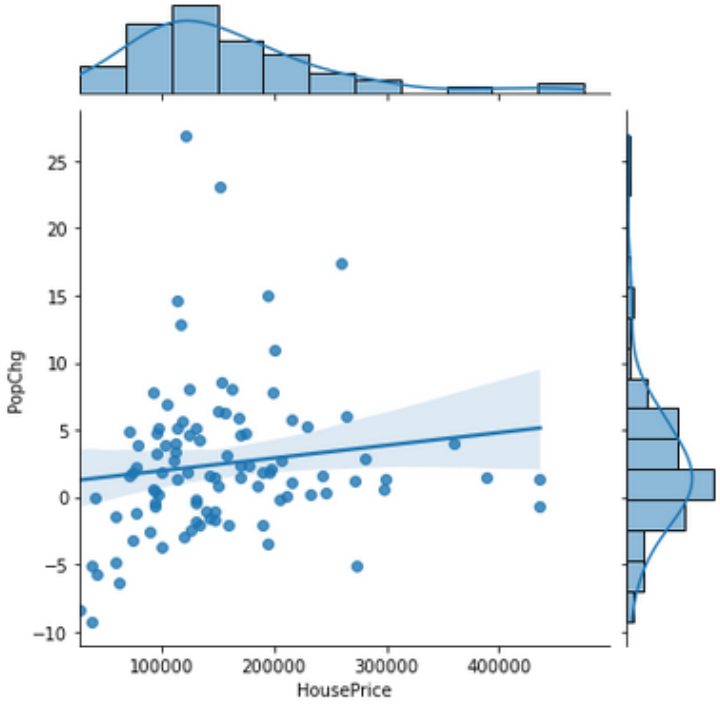
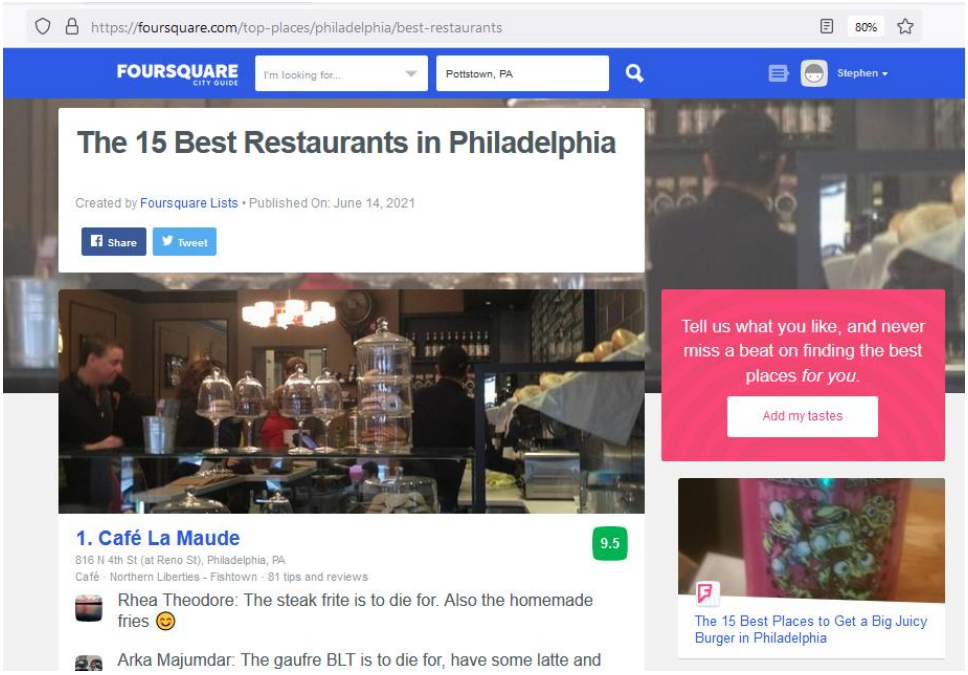


Figure 4. Foursquare landing webpage



The Foursquare API

After exploring the data, latitude and longitude coordinates were extracted using zip codes for each of the 98 Philadelphia region towns using the Geopy module in Python. We then used the Foursquare application programming interface (API) to obtain detailed information about local businesses and venues in each of the towns.

We used the following GET request using the Foursquare API, using our developer credentials, a limit of 100 records per request, and a search radius of 1000 meters:

https://api.foursquare.com/v2/venues/explore?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&ll=LATITUDE,LONGITUDE&v=VERSION&limit=LIMIT

The data returned from Foursquare were then analyzed by type and frequency. A total of 167 types of business establishments and venues were found, and the top ten venue categories were reported for each town. A sample of the resulting data is shown in Table 2 below:

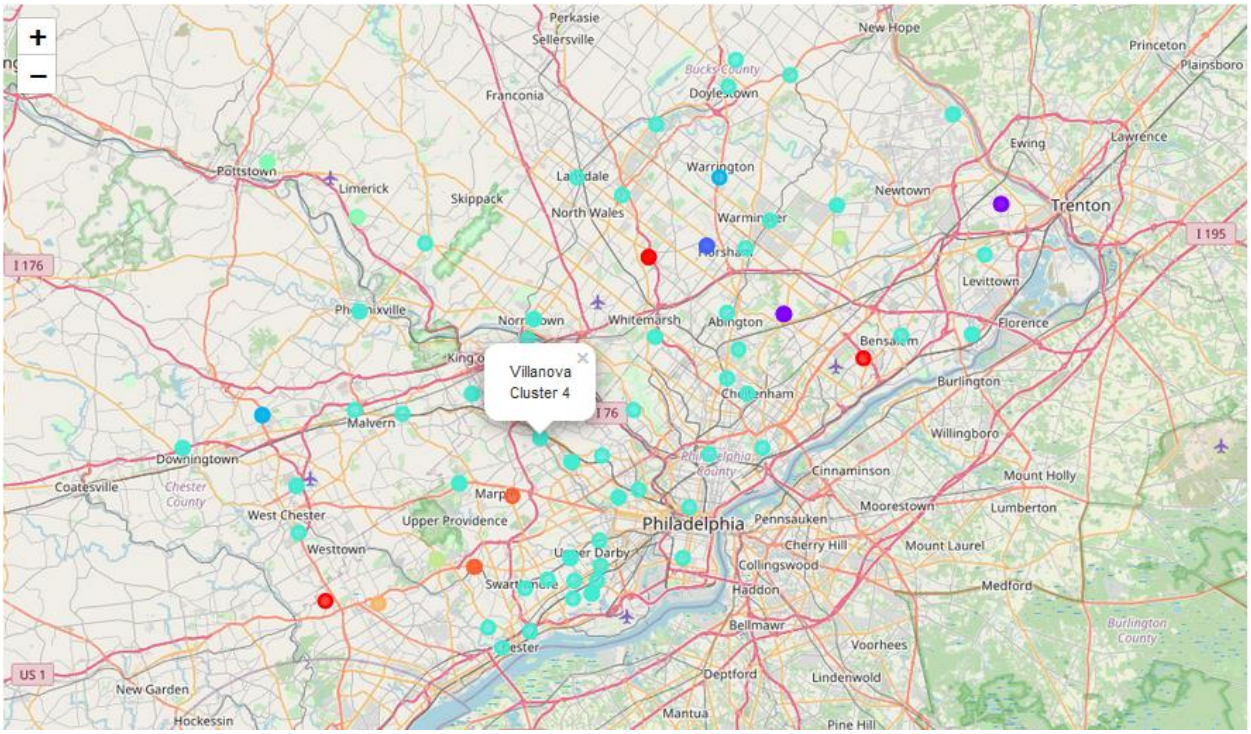
Table 2: Top ten venues for each of 98 Philadelphia region towns

	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

Clustering Analysis

The Foursquare venue data were cleaned and pre-processed and used in a clustering analysis of the towns. Clustering analysis is an unsupervised machine learning method that calculates the distance between each observation, using a distance (or dissimilarity) metric such as Euclidean distance. In k-means clustering, the algorithm then groups the observations into a pre-defined number of clusters. While a range of k-values was used here, and techniques such as the elbow method were applied, it was decided that the optimal number of clusters was nine. A map showing the nine clusters is illustrated in Figure 5.

Figure 5. Map of Philadelphia region showing the 9 clusters.



5. Results and Discussion

The clusters resulting from the analysis were in some cases informative, and in other cases, uninformative. In most cases, the clusters showed obvious similarity in the type and number of establishments. There were also cases where very dissimilar towns, based on other measures such as crime rate or house prices, were grouped into the same cluster.

Where a single town was returned as belonging to its own 'cluster', these cases were evaluated on a case-by-case basis and interpreted to be significantly different from other clusters. In some cases, groupings were found to be useful in terms of highlighting groups of towns that could be ruled out as possible expansion locations because they lacked a critical mass of establishments, implying low foot traffic and below median commercial opportunity.

Clusters 3, 4 and 8 may be the most interesting and potentially useful. They appear to show towns with higher than average disposal income levels and a large of number and variety of complementary venues such as doctor's offices, farmer's markets, nightlife locations, breweries, coffee shops, etc. These clusterings represent good starting points for additional analysis of potential expansion locations.

Overall, the results provide a different perspective from more traditional approaches based on investigating house prices, crime rates and population growth. This alternative perspective is potentially valuable in that it could reveal information as to feasibility or lack of feasibility that a traditional approach might fail to detect. In reality, the first few locations decision would likely be based on obvious criteria and the usual analysis of economic, real estate and demographic factors. However, the clustering analysis could open our eyes to less obvious similarities and patterns. For example, in Cluster 8 we see that Media PA is a town with below average House Prices; but a high density of dining and other commercial options make it perhaps more similar to two other towns with much higher House Prices. This might suggest that Media, while not as mature and wealthy, but may be growing quickly.

Figure 6. "Cluster 8" as decided by the clustering unsupervised learning method

Cluster 8:

```
philly_suburbs_merged.loc[philly_suburbs_merged['Cluster Labels'] == 8, philly_suburbs_merged.columns[[1] + list
```

	HsPrc ($10,000)	Name	County	ZipCode	Latitude	Longitude	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	8 C
41	18.5404	Marple	Delaware	19008	39.974778	-75.360338	8	Moving Target	Park	Beer Garden	Yoga Studio	Eye Doctor	Flower Shop	Fish Market	
42	12.6160	Media	Delaware	19063	39.916052	-75.401804	8	Moving Target	Home Service	Nightlife Spot	Coffee Shop	Brewery	Yoga Studio	Farm	
82	24.2821	Upper Providence	Delaware	19063	39.916052	-75.401804	8	Moving Target	Home Service	Nightlife Spot	Coffee Shop	Brewery	Yoga Studio	Farm	

6. Conclusion

The goal of this project was to help the management of Burger Chain, Inc. to explore and analyze the greater Philadelphia suburbs to identify potential expansion locations. 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 also obtained detailed 'street level' information to develop a more granular understanding of the types of businesses in each town. Finally, we used clustering analysis to group the 98 towns into 9 clusters with the goal of potentially finding similarities between towns that could help in the process of accepting or rejecting candidate locations. Based on the clusterings, there are several potentially interesting locations, particularly in clusters 3, 4 and 8, that warrant further investigation and could be suitable for future expansion.