

# Coursera Data Science Capstone – Week 4 Report:

## The Battle of the Worcester Neighborhoods

### 1. Introduction:

#### A description of the background.

In Worcester County, MA. I recently get a dream job in Boylston and the new office is located almost 15 miles away from home. This means my daily travel time is close to 1.5 hours. I am considering moving closer to the office to avoid wasting time. I and my family currently stay in the neighborhood of the city of Worcester, which is in the west of Worcester County, and the job office is in Boylston, which is in the east of Worcester County. Since I don't have to consider to stay in Boylston where has a relatively high rent price. But I prefer moving to a neighborhood closer to my workplace and similar to our current neighborhood of Worcester.

#### A discussion of the problems.

- Which neighborhoods in Worcester County are good choices for our family to move to?
- These candidate neighborhoods should be both close to my new office and similar to our current living environment.

### 2. Data and Methodology:

#### To address this problem, we need:

- Geographical Data of Worcester County neighborhoods - latitude and longitude details of the cities in Worcester County
- Neighborhoods Environment Data - data describing interests and venues within a given radius

**Sources of Data:** For the latitude and longitude coordinates of neighborhoods, our primary source of information is the following Wikipedia page:

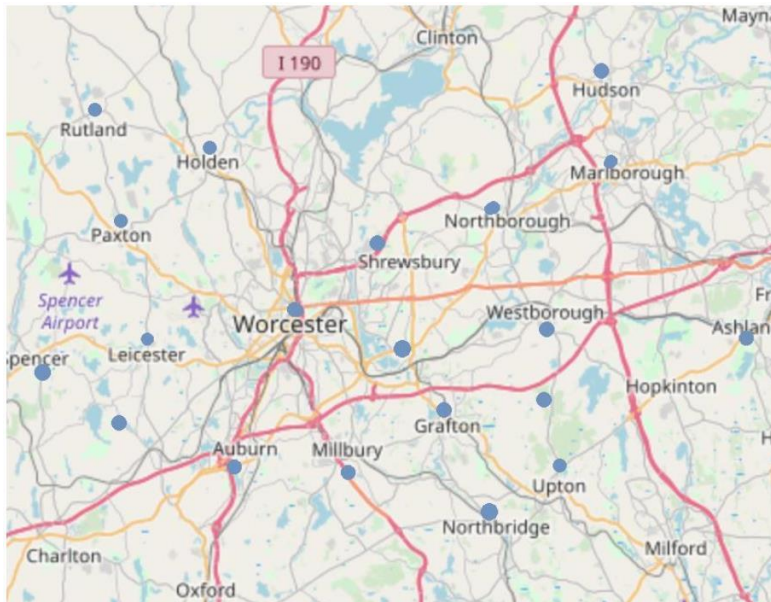
[https://en.wikipedia.org/wiki/Worcester\\_County,\\_Massachusetts](https://en.wikipedia.org/wiki/Worcester_County,_Massachusetts)

- This wiki page comprises a list of links to 60 cities/towns neighborhoods. Each link contains the latitude and longitude of that particular location. Hence, we can compile a database of 88 neighborhoods in Worcester County from these wiki pages, using a web-scraping tool i.e. BeautifulSoup.
- For neighborhood environment details, we use the Foursquare database. Since we are interested in the character and composition of each neighborhood, we use the 'explore' endpoint to learn about the popular venues in each neighborhood.
- In the end, we will run an unsupervised k-means clustering model to look for similar locations as our current neighborhood.

### 3. Data Description:

We crawled 57 records of cities in the Worcester county with the name of the city and its geographic location.

	City	Latitude	Longitude
0	Southborough	42.305556	-71.525000
1	Bolton	42.433333	-71.608333
2	Harvard	42.500000	-71.583333
3	Northborough	42.319444	-71.641667
4	Westborough	42.269444	-71.616667



Then, from the Foursquare Database, we obtained 461 records of the Neighborhoods Environment Data – the data describing interests and venues within a given radius (10km) of each city.

We then get the venues for every neighborhood ranked by popularity and proximity.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Southborough	42.305556	-71.525	Mauro's	42.305840	-71.524282	Breakfast Spot
1	Southborough	42.305556	-71.525	Southborough House Of Pizza	42.305947	-71.524667	Pizza Place
2	Southborough	42.305556	-71.525	Walgreens Pharmacy at Southboro Medical Group	42.310020	-71.525548	Pharmacy
3	Southborough	42.305556	-71.525	Phaidra Nail Salon	42.305849	-71.524878	Spa
4	Southborough	42.305556	-71.525	Santander Bank	42.306135	-71.526181	Bank

After grouping by the neighborhood and calculating the mean value each type of venue, we are able to identify the top ten venues for each neighborhood.

	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	Ashburnham	Pizza Place	Grocery Store	Convenience Store	Coffee Shop	Baseball Field	Liquor Store	Gourmet Shop	Fast Food Restaurant	Discount Store	Dive Bar
1	Athol	Pizza Place	Baseball Field	Breakfast Spot	Chinese Restaurant	Laundromat	Gastropub	Gym	American Restaurant	Doctor's Office	Donut Shop
2	Auburn	Clothing Store	Shoe Store	Kids Store	Women's Store	Jewelry Store	Lingerie Store	Snack Place	Cosmetics Shop	Department Store	Doctor's Office
3	Barre	Pizza Place	Massage Studio	Flower Shop	Bistro	Restaurant	Gas Station	Bakery	Italian Restaurant	Breakfast Spot	Dumpling Restaurant
4	Berlin	Deli / Bodega	Performing Arts Venue	Design Studio	Diner	Discount Store	Dive Bar	Doctor's Office	Donut Shop	Dumpling Restaurant	Electronics Store

## 4. Results

In the end, we adopted the K-means Clustering method to group similar neighborhoods, in a hope of finding a similar neighborhood which locate near Boylston and meanwhile having an identical nearby venues and interests as Worcester city.

We are able to cluster the 57 cities into 5 groups according to their various characteristics.

### Cluster 4:

	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
2	Harvard	Farmers Market	Hobby Shop	Gourmet Shop	Moving Target	Food	Diner	Discount Store	Dive Bar	Doctor's Office	Donut Shop
5	Sterling	Pizza Place	Grocery Store	Mexican Restaurant	Flower Shop	Antique Shop	American Restaurant	Diner	Discount Store	Dive Bar	Food
14	Shrewsbury	Pizza Place	Donut Shop	Pharmacy	Steakhouse	Plaza	Convenience Store	Coffee Shop	Liquor Store	Flower Shop	Italian Restaurant
15	Sturbridge	Park	Hotel	Arts & Crafts Store	Bar	Gift Shop	Bakery	Doctor's Office	Donut Shop	Dumpling Restaurant	Food & Drink Shop
34	Clinton	Café	Sandwich Place	Gym / Fitness Center	Pizza Place	Liquor Store	Mediterranean Restaurant	Italian Restaurant	Bar	Bank	Food & Drink Shop
44	Leominster	Pizza Place	Gym / Fitness Center	Sandwich Place	Spa	Mexican Restaurant	Donut Shop	Liquor Store	Pharmacy	Plaza	Chinese Restaurant
45	Phillipston	Home Service	Department Store	Design Studio	Diner	Discount Store	Dive Bar	Doctor's Office	Donut Shop	Dumpling Restaurant	Electronics Store
46	Warren	Bakery	Bar	Women's Store	French Restaurant	Discount Store	Dive Bar	Doctor's Office	Donut Shop	Dumpling Restaurant	Electronics Store
47	Petersham	Scenic Lookout	Convenience Store	Nature Preserve	Women's Store	Flower Shop	Diner	Discount Store	Dive Bar	Doctor's Office	Donut Shop
52	Gardner	Liquor Store	Gas Station	Diner	Breakfast Spot	Doctor's Office	Pizza Place	Donut Shop	Convenience Store	Pub	Restaurant
53	Worcester	American Restaurant	Hotel	New American Restaurant	Bakery	Art Gallery	Pizza Place	Mexican Restaurant	Italian Restaurant	Electronics Store	Fast Food Restaurant

We found that Worcester city, where we live in now, is categorized in cluster 4 with the other 10 cities. These are a list of good candidate cities that may meet our criterion.

The neighborhood of Gardner, Petersham, and Warren come closest to our current neighborhood in terms of nearby venues. But these neighborhoods are still far away from My workplace. Since our first consideration was to reduce his travel time, we might want to assess a few more neighborhoods in the same Cluster as Worcester.