

---

# **Unpacking Before the Move: Investigating the Social and Commercial Pulse of a Town Before a Big Move**

---

**By Colette Campbell**  
**Data Science Capstone Project**  
**May 31, 2020**

# 1 Introduction

Moving is one of the top five most stressful life events.<sup>1</sup> On average, Americans move approximately 11.4 times over a lifetime.<sup>2</sup> However, once retirement age is reached, or quickly approaching, people are looking for their forever home. They don't want to face the stress of moving again. Needless to say, when looking for that last house to call home, you want to get it right.

Recently, a friend of mine who lives in California mentioned to me that she and her husband are looking to move in the next year or two. They have already identified a town in Connecticut, Avon, that seems appealing to them. The problem is, they have never been to this town. They have questions like: What type of venues are in close proximity to the town? Are there a variety of restaurants? Which stores are available for shopping? Are there other towns similar to Avon that offer a similar social and cultural experience?

To help answer these questions, I used data science to explore Avon, as well as the other towns located in Hartford County, Connecticut.

## 2 Data

Primarily, two types of data were needed to carry out the investigation. First, I needed to compile a list of towns in Hartford County and their geographical coordinates. Second, I needed to know the types of venues surrounding each town.

I was able to find a list of counties and towns within the state of Connecticut by using web scraping.<sup>3</sup> I opted to use this larger dataset to leave room for possible future investigations of other counties adjacent to Hartford County. Latitude and longitude values were found using *geopy* ("geopy is a Python 2 and 3 client for several popular geocoding web services; geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources."<sup>4</sup>). Foursquare was then implemented to provide data of the various types of venues surrounding each town.

---

<sup>1</sup> [https://www.healthstatus.com/health\\_blog/depression-stress-anxiety/top-5-stressful-situations/](https://www.healthstatus.com/health_blog/depression-stress-anxiety/top-5-stressful-situations/)

<sup>2</sup> <https://www.census.gov/newsroom/blogs/random-samplings/2017/01/mover-rate.html>

<sup>3</sup> [https://en.wikipedia.org/wiki/List\\_of\\_towns\\_in\\_Connecticut](https://en.wikipedia.org/wiki/List_of_towns_in_Connecticut)

<sup>4</sup> <https://geopy.readthedocs.io/en/stable/>

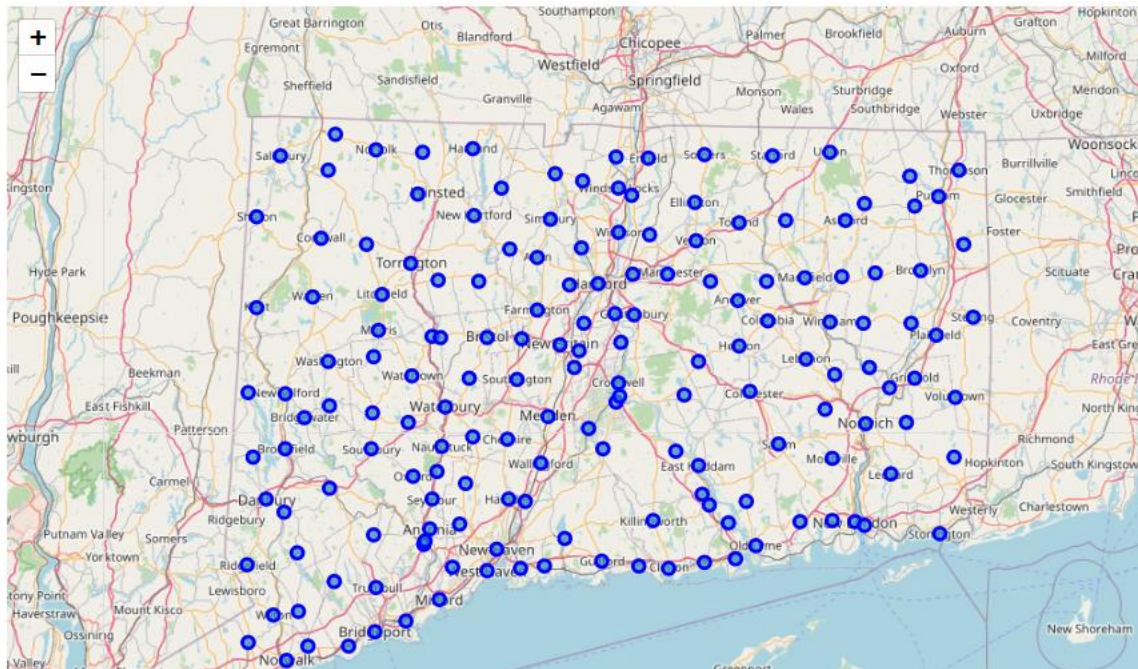
### 3 Methodology

Once the data for the towns in Connecticut were acquired and the geographical coordinates were assigned, I generated a map showing the various towns and their locations using Folium. As mentioned in the previous section, I wanted to leave room for possible future investigations.

#### 3.1 Map of Connecticut

From the map of Connecticut, it is easy to see that the towns are spaced out relatively evenly over the state (This is a good time to state that there is a clear distinction in Connecticut between a city and a town. It is beyond the scope of this report to explain the nuances of the distinction).

In addition, should my friend want to investigate towns along the water, for example, she can easily zoom in or out of the map and identify them visually (this can be done using the link to the Jupyter notebook, but is not able to be done in the image below).



Fig, 1 - Map of Connecticut and its 170 towns

### 3.2 Map of Hartford County

Next, I focused on the county of interest, Hartford County. From the dataframe containing all of the towns in Connecticut, I filtered to created a new dataframe containing only the towns within Hartford County; then, I created a map showing all of the towns (29) within Hartford County (see below).

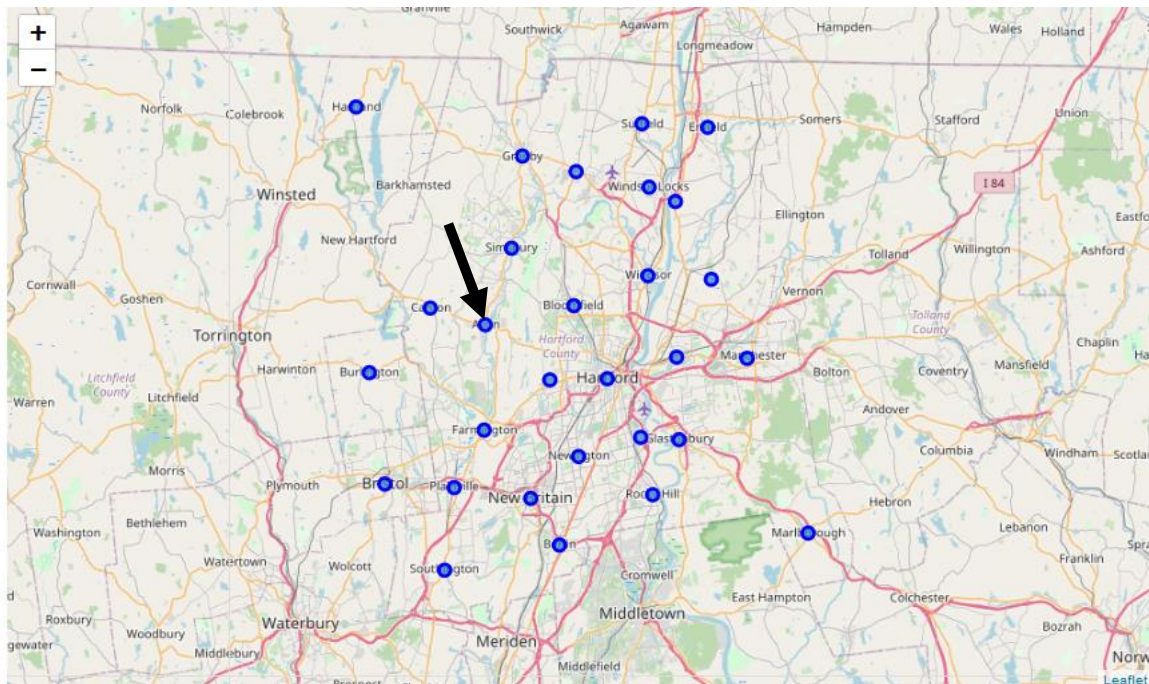


Fig. 2 - Map of Hartford County and its 29 towns (Avon indicated by arrow)

At the center of the county, where the major roads (in red) converge, is the city of Hartford. Northwest of the city of Hartford, and directly north of the city of Farmington, is Avon (indicated by the arrow). This map is useful to zoom in on towns of interest based on the results of top venues associated with a town.

### 3.3 Using Foursquare to Explore Avon

After creating the maps above, I used Foursquare to explore the venues within a 500-meter radius (approximately 1/3 of a mile) of Avon. This is appropriate given that it is the town of primary interest for my friend and her husband. Following are the results:

	name	categories	lat	lng
0	Dom's Coffee	Coffee Shop	41.810382	-72.832328
1	Elephant Trail	Thai Restaurant	41.808008	-72.828137
2	First & Last Tavern	Italian Restaurant	41.810954	-72.832292
3	Da Capo	Italian Restaurant	41.809174	-72.829948
4	The UPS Store	Shipping Store	41.807955	-72.827703
5	SUBWAY	Sandwich Place	41.808727	-72.828960
6	Bruegger's Bagel Bakery	Bagel Shop	41.807357	-72.827398
7	The Coffee Trade Inc.	Coffee Shop	41.809926	-72.832480
8	Residence Inn Hartford Avon	Hotel	41.812420	-72.831833
9	Pick and Mix	Korean Restaurant	41.807818	-72.827272
10	Country Curtains	Furniture / Home Store	41.811023	-72.833113
11	Liquid Nirvana	Juice Bar	41.809069	-72.827679
12	Welcome Wine & Liquor	Wine Shop	41.807975	-72.827813
13	barre 44	Athletics & Sports	41.807639	-72.826891
14	Carmen Anthony Fishhouse	Seafood Restaurant	41.807446	-72.826834
15	Lacrosse Unlimited of Avon CT	Sporting Goods Shop	41.807446	-72.826834

Fig. 3 - Venues within a 500-m radius of Avon, CT

### 3.4 Using Foursquare to Explore Towns in Hartford County

Foursquare was used to explore all towns located in Hartford County, similar to the example given above for the town of Avon. Once again, a 500-m radius was used. The results were then grouped by town and a count of the number of venues per town was returned as a dataframe (see next page).

Town	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Avon	16	16	16	16	16	16
Berlin	5	5	5	5	5	5
Bloomfield	9	9	9	9	9	9
Bristol	14	14	14	14	14	14
Burlington	11	11	11	11	11	11
Canton	11	11	11	11	11	11
East Granby	7	7	7	7	7	7
East Hartford	14	14	14	14	14	14
East Windsor	11	11	11	11	11	11
Enfield	4	4	4	4	4	4
Farmington	4	4	4	4	4	4
Glastonbury	38	38	38	38	38	38
Granby	23	23	23	23	23	23
Hartford	16	16	16	16	16	16
Manchester	17	17	17	17	17	17
Marlborough	16	16	16	16	16	16
New Britain	15	15	15	15	15	15
Newington	25	25	25	25	25	25
Plainville	6	6	6	6	6	6
Rocky Hill	10	10	10	10	10	10
Simsbury	23	23	23	23	23	23
South Windsor	8	8	8	8	8	8
Southington	32	32	32	32	32	32
Suffield	15	15	15	15	15	15
West Hartford	1	1	1	1	1	1
Wethersfield	14	14	14	14	14	14
Windsor	16	16	16	16	16	16
Windsor Locks	2	2	2	2	2	2

Fig. 4 – Number of venues within 500-m radius of each town, Hartford County

## 3.5 Analysis of Each Town in Hartford County

### 3.5.1 One Hot Encoding

To compare the various towns in Hartford County, they were grouped according to the top ten most common venues in each town. In total, there were 117 unique venues returned for towns in Hartford County. These are categorical variables (as opposed to numerical). In order to do further analysis, one hot encoding was employed. One hot



encoding is a method used to map variables from categorical to numerical (more specifically, binary). This way, algorithms are able to process information to give a prediction or outcome. Once one hot encoding was completed, the resultant dataframe was grouped by town with the mean of occurrence of each category indicated.

### 3.5.2 K-Means Clustering

K-means clustering is a method used to take a data set and group similar cases together; in this specific investigation, similar towns will be grouped together based on venues. When employing this technique, it is necessary to specify the number of clusters, “k”. To find the optimal number of clusters, a method called “The Elbow Method” is most popular. In this method, one wants to minimize the distance between points in a cluster (measured using the total within-cluster sum of squares, or WCSS). The total WCSS is a measure of the compactness of a cluster. When a graph of WCSS versus number of clusters is made, an “elbow” is formed at the optimal value of “k”.

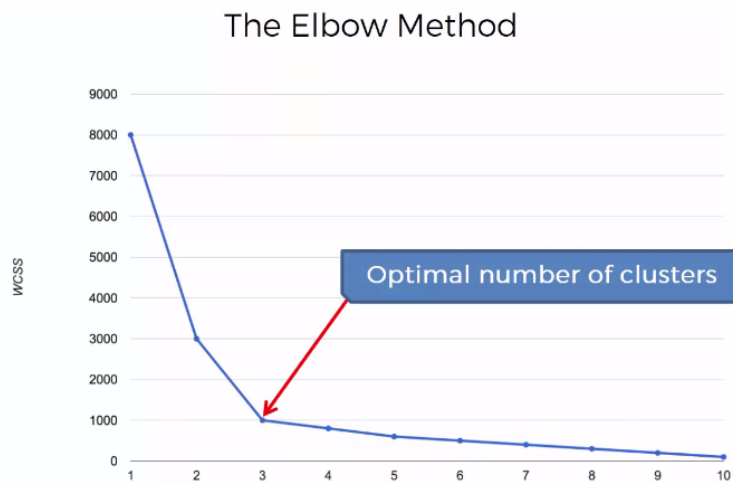


Fig. 5 - Example of Graph of WCSS versus number of clusters<sup>5</sup>

Another less rigorous method to approximate k is to take the square root of half the total number of cases:

$$k \approx \sqrt{\left(\frac{n}{2}\right)}$$

---

<sup>5</sup> <https://arun-aiml.blogspot.com/2017/07/k-means-clustering.html>

Using this formula, gave me a value of  $k = 4$  (I also tested  $k = 3$  and  $k = 5$ ;  $k = 4$  gave the best results). The towns in Hartford County were clustered based on the top 10 most common venues in each town (based on the frequency of occurrence). The results were then visualized on a map of Hartford County (see next section).

## 4 Results

The maps below once again show the towns in Hartford County, but this time each town is assigned a colored marker according to the assigned cluster (maps are identical, but at different zoom levels). The red marker indicates **Cluster 1**; the purple marker indicates **Cluster 2**; the aqua marker indicates **Cluster 3**; and the light green marker indicates **Cluster 4**.

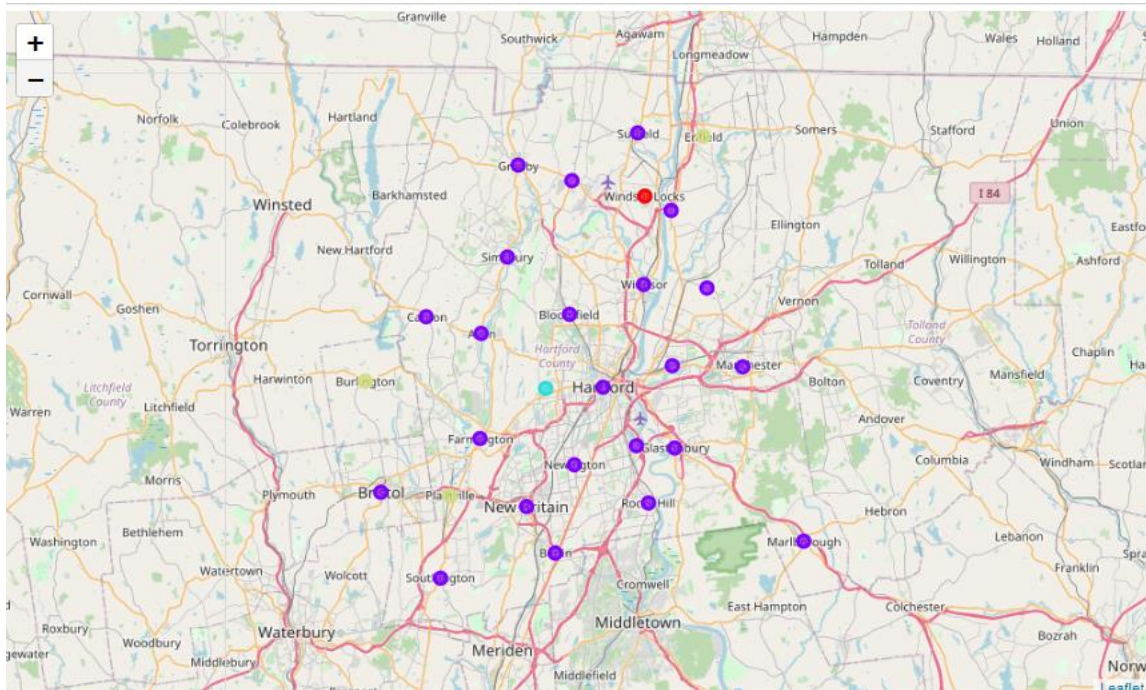


Fig. 6a – Map of Hartford County indicating cluster assignment of each town



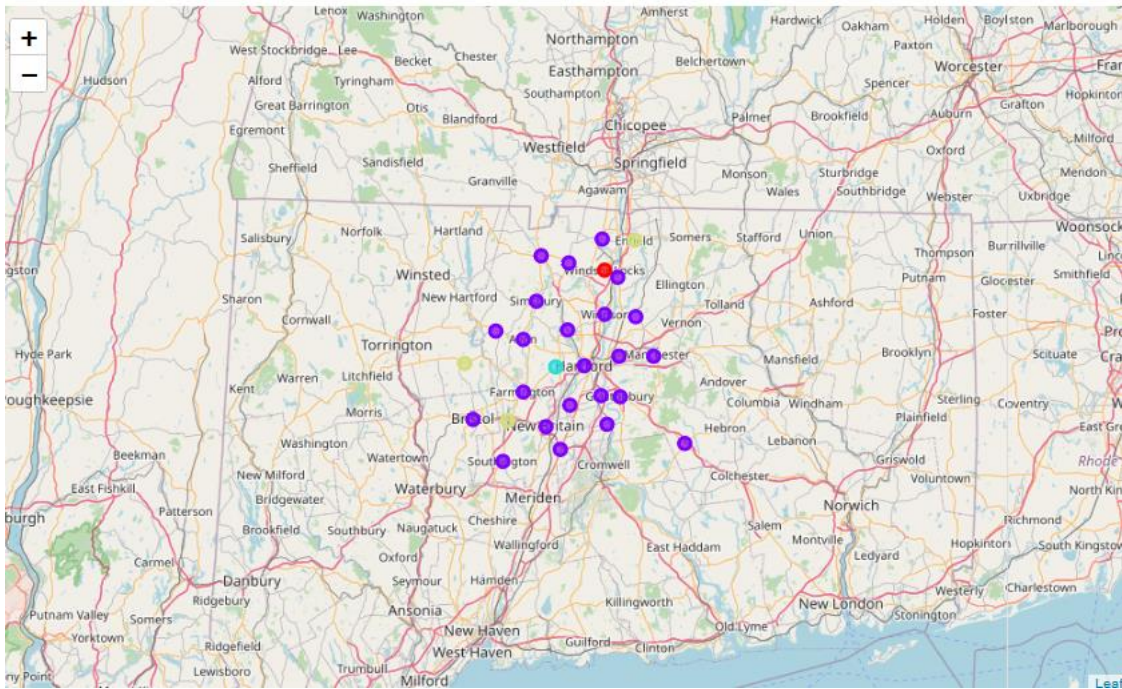


Fig. 6b – Map of Hartford County indicating cluster assignment of each town

More detail of the clusters (towns with their top 10 venues) can be seen in the dataframes below.

#### Cluster 1

```
hartford_merged.loc[hartford_merged['Cluster Labels'] == 0, hartford_merged.columns[[1] + list(range(5, hartford_me
```

	Town	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
28	Windsor Locks	American Restaurant	Deli / Bodega	Yoga Studio	Event Space	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio	Dessert Shop	Diner

Fig. 7 – Cluster 1 (1 town)

## Cluster 2

```
hartford_merged.loc[hartford_merged['Cluster Label'] == 1, hartford_merged.columns[[1] + list(range(5, hartford_merged.n
```

	Town	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	East Hartford	Discount Store	Liquor Store	Shipping Store	Dessert Shop	Pizza Place	Home Service	Pharmacy	Chinese Restaurant	Automotive Shop	Caribbean Restaurant
1	Rocky Hill	ATM	Steakhouse	Liquor Store	Business Service	Dance Studio	Pharmacy	Deli / Bodega	Bridal Shop	Sandwich Place	Donut Shop
2	East Windsor	Fast Food Restaurant	Hotel	Fried Chicken Joint	Asian Restaurant	Diner	Sushi Restaurant	Donut Shop	Sporting Goods Shop	Pharmacy	Yoga Studio
3	Wethersfield	Farm	Yoga Studio	Bar	Deli / Bodega	Pizza Place	Event Space	Museum	Big Box Store	Farmers Market	Ice Cream Shop
4	Newington	American Restaurant	Pharmacy	Hobby Shop	Bank	Bar	Chinese Restaurant	Shoe Store	Market	Nail Salon	Burger Joint
6	Farmington	Elementary School	Food	Home Service	Lawyer	Yoga Studio	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio	Deli / Bodega
7	Glastonbury	Italian Restaurant	Pizza Place	Bakery	Ice Cream Shop	Salon / Barbershop	Coffee Shop	Thai Restaurant	Grocery Store	Cocktail Bar	Jewelry Store
8	Granby	Ice Cream Shop	Grocery Store	Chinese Restaurant	Sandwich Place	Liquor Store	Coffee Shop	Mexican Restaurant	Dessert Shop	Italian Restaurant	Park
11	Hartford	Coffee Shop	Café	Bar	Bakery	Gas Station	Bus Station	Discount Store	Donut Shop	Speakeasy	New American Restaurant
12	Berlin	Bike Shop	Pizza Place	Flower Shop	Donut Shop	Breakfast Spot	Event Service	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio
13	East Granby	Pizza Place	Massage Studio	Sandwich Place	American Restaurant	Chinese Restaurant	Breakfast Spot	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio
16	South Windsor	Pet Store	Pizza Place	Chinese Restaurant	Food Truck	Ice Cream Shop	Beer Garden	Hardware Store	Donut Shop	Discount Store	Diner
18	Windsor	Park	Bakery	Grocery Store	Pharmacy	Diner	Coffee Shop	Sandwich Place	Bar	Pizza Place	Indian Restaurant
17	Canton	Coffee Shop	Thai Restaurant	Pizza Place	Fast Food Restaurant	Bank	Bagel Shop	Outlet Store	Thrift / Vintage Store	Trail	Furniture / Home Store
18	Marlborough	Pizza Place	American Restaurant	Gift Shop	Chinese Restaurant	Sandwich Place	Convenience Store	Café	Donut Shop	Bank	Grocery Store
20	Suffield	American Restaurant	Italian Restaurant	Liquor Store	Greek Restaurant	Pizza Place	Business Service	Bank	Gas Station	Sushi Restaurant	Chinese Restaurant
21	Southington	Pizza Place	Gym	American Restaurant	Pharmacy	Gastropub	BBQ Joint	Bank	Sandwich Place	Museum	Brewery
22	New Britain	Business Service	Pizza Place	Grocery Store	Salon / Barbershop	Hot Dog Joint	Performing Arts Venue	Gym / Fitness Center	Donut Shop	Thai Restaurant	Latin American Restaurant
23	Manchester	Pizza Place	Gift Shop	Spa	Chinese Restaurant	Credit Union	Park	Pharmacy	Discount Store	Rental Car Location	Bookstore
24	Bloomfield	Bank	Convenience Store	Liquor Store	Restaurant	Pharmacy	Donut Shop	Caribbean Restaurant	American Restaurant	Dance Studio	Elementary School
26	Simsbury	American Restaurant	Restaurant	Burger Joint	Pharmacy	Pizza Place	Playground	Dog Run	Donut Shop	Café	Sandwich Place
28	Bristol	Pizza Place	Sandwich Place	Grocery Store	Brewery	Pharmacy	Dessert Shop	Bank	Discount Store	Donut Shop	Thai Restaurant
27	Avon	Coffee Shop	Italian Restaurant	Furniture / Home Store	Juice Bar	Sandwich Place	Seafood Restaurant	Shipping Store	Sporting Goods Shop	Bagel Shop	Athletics & Sports

Fig. 8 – Cluster 2 (23 towns)

### Cluster 3

```
hartford_merged.loc[hartford_merged['Cluster Labels'] == 2, hartford_merged.columns[[1] + list(range(5, hartford_merged.aha
```

	Town	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
18	West Hartford	Clothing Store	Event Service	Construction & Landscaping	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio	Deli / Bodega	Dessert Shop	Diner

### Cluster 4

```
hartford_merged.loc[hartford_merged['Cluster Labels'] == 3, hartford_merged.columns[[1] + list(range(5, hartford_merged.aha
```

	Town	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
6	Enfield	Chinese Restaurant	Garden Center	Food	Flower Shop	Yoga Studio	Elementary School	Convenience Store	Cosmetics Shop	Credit Union	Dance Studio
9	Plainville	Mexican Restaurant	Convenience Store	Pizza Place	Bakery	Burger Joint	Diner	Yoga Studio	Elementary School	Cosmetics Shop	Credit Union
14	Burlington	Construction & Landscaping	Flower Shop	Irish Pub	Pizza	Pizza Place	Pharmacy	Farmers Market	Italian Restaurant	Campground	Asian Restaurant

Fig. 9 – Cluster 3 (1 town) and Cluster 4 (3 towns)

## 5 Discussion

When examining the results of the clustering performed on the towns in Hartford County, it is clear there is an imbalance; Cluster 2 contains just under 83% of the total number of towns. Why is this?

Clusters are divided according to similarities. The most obvious answer is that there is not much variation in places of interest (venues) from town to town. Indeed, the most common venues in Cluster 2 seem to follow a theme of having mostly restaurants, pizza places, coffee shops, Perhaps, a better question is what makes the towns in Cluster 1, Cluster 3, and Cluster 4 different.

In Fig. 6a, showing the towns in Hartford County by Cluster, notice that the towns in Cluster 1 and Cluster 4 are located on the outer perimeter of Hartford County; is that a coincidence? I would assume that towns further from the center, or seat of the county,

would be smaller. This assumption can be verified by examining Fig. 4, which shows the total number of venues returned within a 500-m radius (Note: Clusters were formed by the top 10 most common venues in each town; clusters are not limited by radius).

The towns in Cluster 1, 3, and 4 include:

Cluster 1: Windsor Locks

Cluster 3: West Hartford

Cluster 4: Enfield, Plainville, Burlington.

According to Fig. 4 the number of venues within a 500-m radius of the towns above are as follows:

<b>Town</b>	<b>Number of venues within a 500-m radius of the town center</b>
Windsor Locks	2
West Hartford	1
Enfield	4
Plainville	6
Burlington	11

Compare this to an average of 16 venues within a 500-m radius for Cluster 2. Although there may be a difference in town size and density of venues, overall, there is not much

differentiating the towns based on the top 10 most common venues. How could we get better clustering?

Better clustering could be achieved by adding demographic information of each town (for example, population, crime rate, etc.) as additional features would improve clustering and create more distinction between towns.

The initial question

## **6 Conclusion**

In this project, K-Means Clustering was used to analyze towns in Hartford County, Connecticut based on the top 10 most common venues in each town. This analysis was done to answer the following questions for my friend and her husband who are interested in moving to Hartford County, Connecticut; more specifically, Avon.

### **What type of venues are in close proximity to the town (Avon)?**

Within a small radius of 500-m of the town center, there were restaurants, coffee shops, a shipping store, a hotel, a furniture store, a juice bar, a wine shop, a barre studio, and a sporting goods shop. Due to the variety of venues within a small area, Avon appears to have quite a variety of options.

### **Are there a variety of restaurants?**

Avon offers a good variety when it comes to restaurant including the following cuisine: Italian, American, Thai, Korean, and seafood.

### **Which stores are available for shopping?**

The only stores returned in the small area of exploration were a shipping store and a sporting goods store. When examining the top 10 most common venues, the results practically mirror the stores found within 500-m of the town center. This would suggest that a resident of Avon may have to travel outside of the town limits for a variety when shopping.

### **Are there other towns similar to Avon that offer a similar social and cultural experience?**

Based on the top 2 most common venues, the town of Canton would be considered most similar to Avon. However, Avon appears to be unique in that no other town has a juice bar or furniture store listed in their top ten most common venues.

## 7 Further Investigation and Application

To refine the results, I could do further exploration of the towns in Hartford County using demographic information, which tends to be more distinctive. Again, I could use K-Means Clustering to group towns that are similar based on the features.

This type of analysis can be used on a grander scale to find similar states (or on a smaller scale to find homes in a neighborhood that are similar). In fact, out of curiosity, I used K-means Clustering and demographic information to find states that are considered similar to Connecticut (Cluster label = 1).

	State	MHI	Population	Crime Index	Cost Index	Percent Sun	Pop. In Urban Areas (%)	Cluster label
8	Florida	55462.0	21477737	5.2	97.9	66.0	91.2	0
31	New York	67844.0	19453561	2.9	139.1	46.0	87.9	0
27	Nevada	58646.0	3080156	6.7	108.5	79.0	94.2	1
43	Utah	71414.0	3205958	1.9	98.4	66.0	90.6	1
23	Mississippi	44717.0	2976149	5.7	86.1	61.0	49.4	1
3	Arkansas	47062.0	3017825	7.2	86.9	61.0	56.2	1
30	New Mexico	47169.0	2096829	8.0	87.5	76.0	77.4	1
6	Connecticut	76348.0	3565287	2.3	127.7	56.0	88.0	1
15	Kansas	58218.0	2913314	3.9	89.0	65.0	74.2	1
14	Iowa	59955.0	3155070	1.7	90.1	59.0	64.0	1

Overall, using data science can help to alleviate the stress of moving by unpacking useful information and trends about an area before starting to pack for a big move.