# Destination New England! What new Students are coming to!

Top Colleges including few IVY Leagues, Great sports teams, friendly environment and what more?

Let's explore the New England city and town area (NECTA)

NECTA Divisions

- 1 – Amherst Center, MA
- 2 – Athol, MA
- 3 – Augusta, ME
- 4 – Bangor, ME
- 5 – Barnstable Town, MA
- 6 – Barre, VT
- 7 – Bennington, VT
- 8 – Berlin, NH
- 9 – Boston-Cambridge-Quincy, MA-NH
- 10 – Bridgeport-Stamford-Norwalk, CT
- 11 – Brunswick, ME
- 12 – Burlington-South Burlington, VT
- 13 – Claremont, NH
- 14 – Concord, NH
- 15 – Danbury, CT
- 16 – Danielson, CT
- 17 – Franklin, NH
- 18 – Greenfield, MA
- 19 – Hartford-West Hartford-East Hartford, CT
- 20 – Keene, NH
- 21 – Laconia, NH
- 22 – Lebanon, NH-VT
- 23 – Leominster-Fitchburg-Gardner, MA
- 24 – Lewiston-Auburn, ME
- 25 – Manchester, NH
- 26 – New Bedford, MA
- 27 – New Haven, CT
- 28 – North Adams, MA-VT
- 29 – Norwich-New London, CT-RI
- 30 – Pittsfield, MA
- 31 – Portland-South Portland, ME
- 32 – Portsmouth, NH-ME
- 33 – Providence-Fall River-Warwick, RI-MA
- 34 – Rochester-Dover, NH-ME
- 35 – Rockland, ME
- 36 – Rutland, VT
- 37 – Sanford, ME
- 38 – Springfield, MA-CT
- 39 – Torrington, CT
- 40 – Waterbury, CT
- 41 – Waterville, ME
- 42 – Willimantic, CT
- 43 – Worcester, MA-CT
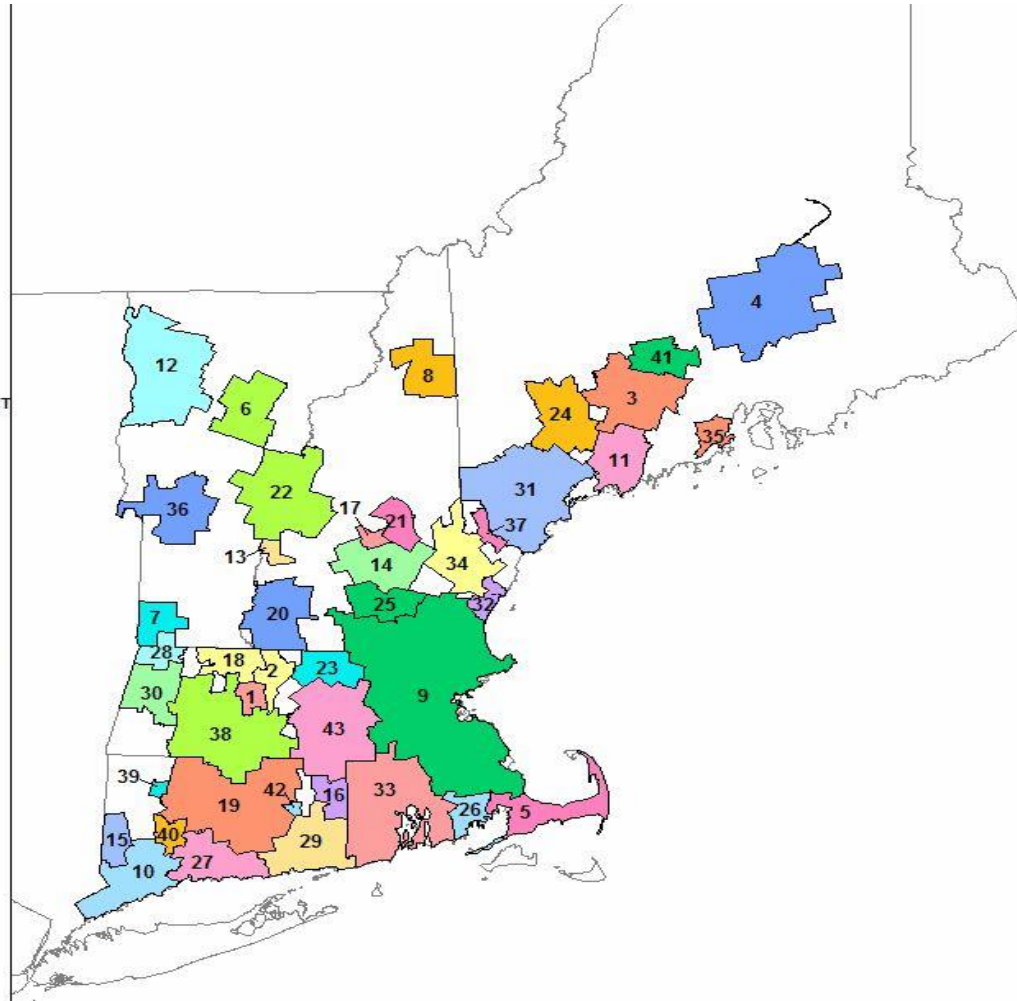
Introduction/Business Problem

1.1. Problem definition:

Every year thousands of new students call New England their home for the next 4 years. Finding a new place to live for the many who opt to leave off campus, could be a bit tricky if you only rely on few web search here and there, not to mention such process is time consuming and can be costly if you land at the wrong neighborhoods. This project is alleviate the hard tasks and eliminate the guessing game for the thousand of students and families different city, state, and/or country moving to New England this Fall. I will cover all of New England city and town area (NECTA). NECTA is a geographic and statistical entity defined by the U.S. federal government for use in the six-state New England region of the United States. NECTAs are analogous to metropolitan statistical areas and micropolitan statistical areas and are defined using the same criteria, except that they are defined on the basis of New England towns instead of entire counties. NECTAs are classified as either metropolitan or micropolitan NECTAs. A micropolitan NECTA has an urban core with a population of at least 10,000 but less than 50,000, whereas a metropolitan NECTA has an urban core with a population of at least 50,000.

1.2. Targeted audience:

I propose a way of shortening the new students' search process by using data about the most popular venues of each area to help find their "venue profile". Potentially there are few groups of people interested in this type of insights:

- The some real estate companies knowing what are the requirements – venue profile of the client, this analysis will limit the search area to the preferred venues of clients, and therefore the need of real estate resources and time.
- The New students who want to move or relocate to this NECTA. This type of analysis can speed up the process of searching a place to live saving time and money. It will also help investors who see an opportunity to invest in apartments.

1.3. Search requirements:

Students are known to consume lot of Pizza, they love coffee shops, bugger joins and restaurants, gyms to name a few. Therefore my highlights will be focused on:

- Coffee shop
- Book store
- Restaurants
- Theater
- Pizza
- Parks

2. Data to support this project

2.1. Data description

To capture the "venue profiles" of the NECTAs, I use:

- Web scraping with BeautifulSoup to obtain a list of New England city and town area (NECTA).

The Wiki page: https://en.wikipedia.org/wiki/New_England_city_and_town_area has a table with:

- List of all the NECTA,
- 2010 Population,
- 2000 Population,
- Percent change,

- Area density.

Rank	NECTA	State(s)	2000 Pop	2010 Pop	%Change
1	Boston–Cambridge–Quincy	MA-NH	4,503,683	4,703,187	4.43%
2	Providence–Fall River–Warwick	RI-MA	1,292,942	1,301,595	0.67%
3	Hartford–West Hartford–East Hartford	CT	1,059,878	1,121,463	5.81%
4	Bridgeport–Stamford–Norwalk	CT	892,283	926,465	3.83%
5	Springfield	MA-CT	660,837	683,800	3.47%
6	New Haven	CT	571,310	597,172	4.53%
7	Worcester	MA-CT	539,828	577,537	6.99%
8	Portland–South Portland–Biddeford	ME	333,624	357,412	7.13%
9	Norwich–New London	CT-RI	265,288	278,598	5.02%
10	Barnstable Town	MA	244,257	239,675	-1.88%
11	Waterbury	CT	195,540	204,451	4.56%
12	Burlington–South Burlington	VT	187,105	198,627	6.16%
13	Manchester	NH	176,663	187,596	6.19%
14	New Bedford	MA	170,161	175,502	3.14%
15	Danbury	CT	155,304	163,260	5.12%
16	Rochester–Dover	NH-ME	135,367	149,471	10.42%
17	Leominster–Fitchburg–Gardner	MA	143,905	147,818	2.72%
18	Bangor	ME	124,906	135,632	8.59%
19	Lewiston–Auburn	ME	101,778	106,216	4.36%
20	Portsmouth	NH-ME	71,232	73,274	2.87%
21	Pittsfield	MA	73,297	72,051	-1.70%

- Nominatim from geopy.geocoders for geocoding the County names and get their coordinates.
With Nominatim we will be able to concatenate the list of New England city and town area (NECTA) with their coordinates.
- Foursquare API to get all venues in the
With Foursquare we will get the top venues that are within certain radius of each other.
3. Methodology
The methods used in this work were:
   1. Web Scraping with the BeautifulSoup library
   2. Geocoding with Nominatim from geopy.geocoders
   3. Data acquisition from Foursquare's API

4. Feature reduction by considering most common venue categories
5. Machine learning: k-Means clustering because it is the most simple clustering algorithm and it was capable of meeting the proposed objective

4. Results

4.1. Web scraping

This is the Wikipedia page https://en.wikipedia.org/wiki/New_England_city_and_town_area that contains a list of all towns in New England.
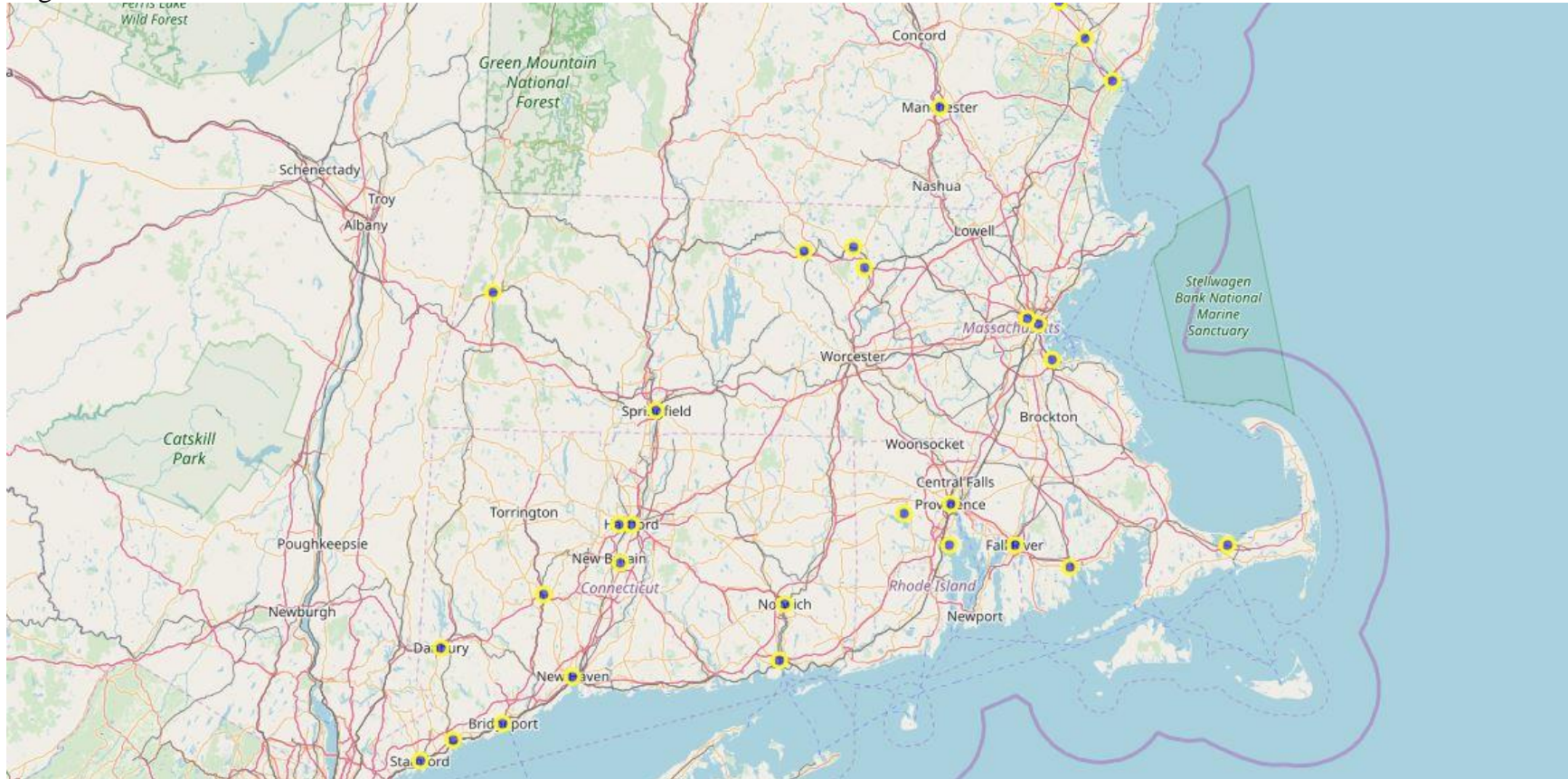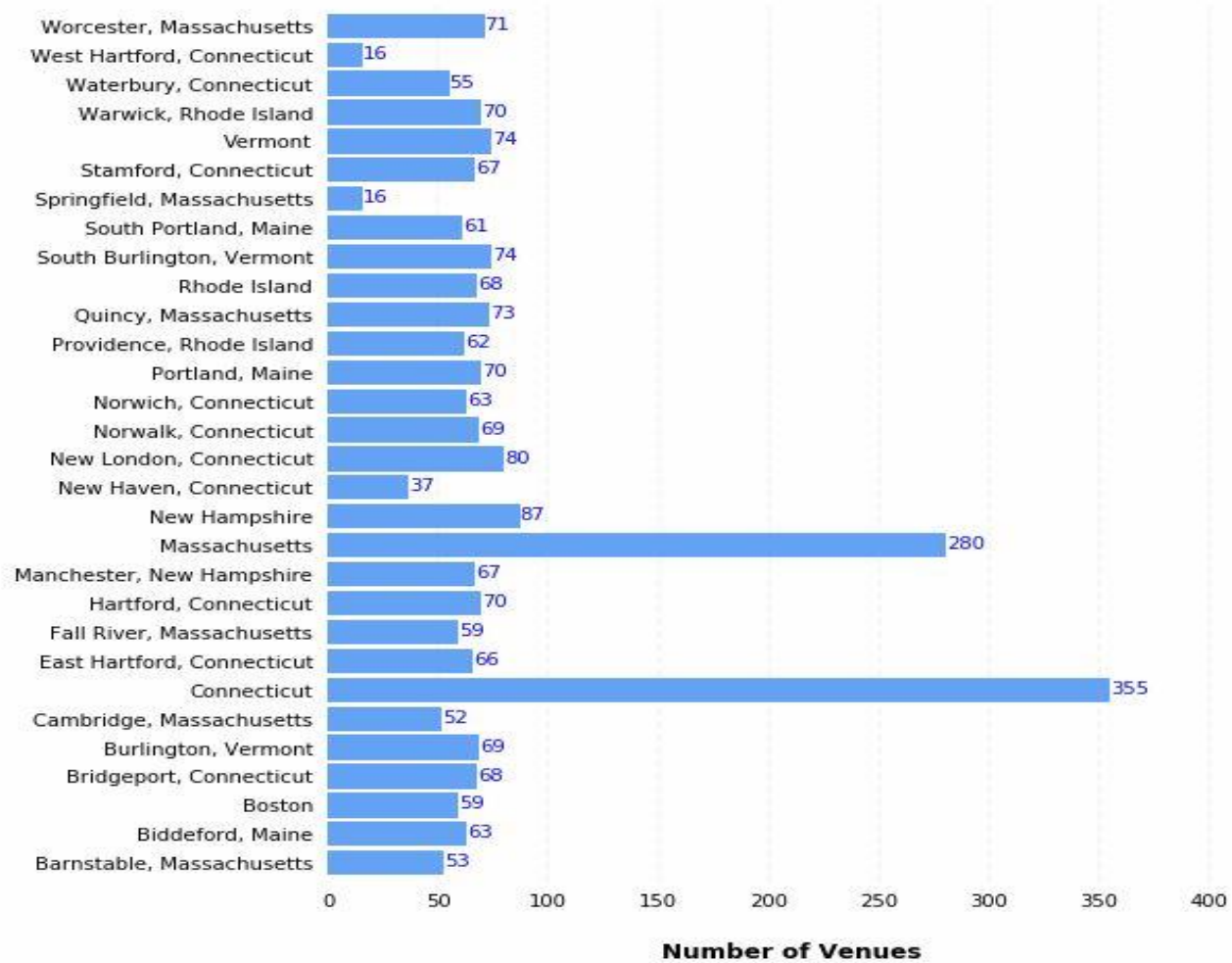

4.2. Mapping the NECTAs with clusters.

The goal is This project is alleviate the hard tasks and eliminate the guessing game for the thousands of students and families different city, state, and/or country moving to New England this Fall, to do that I will cluster based on the NECTAs.
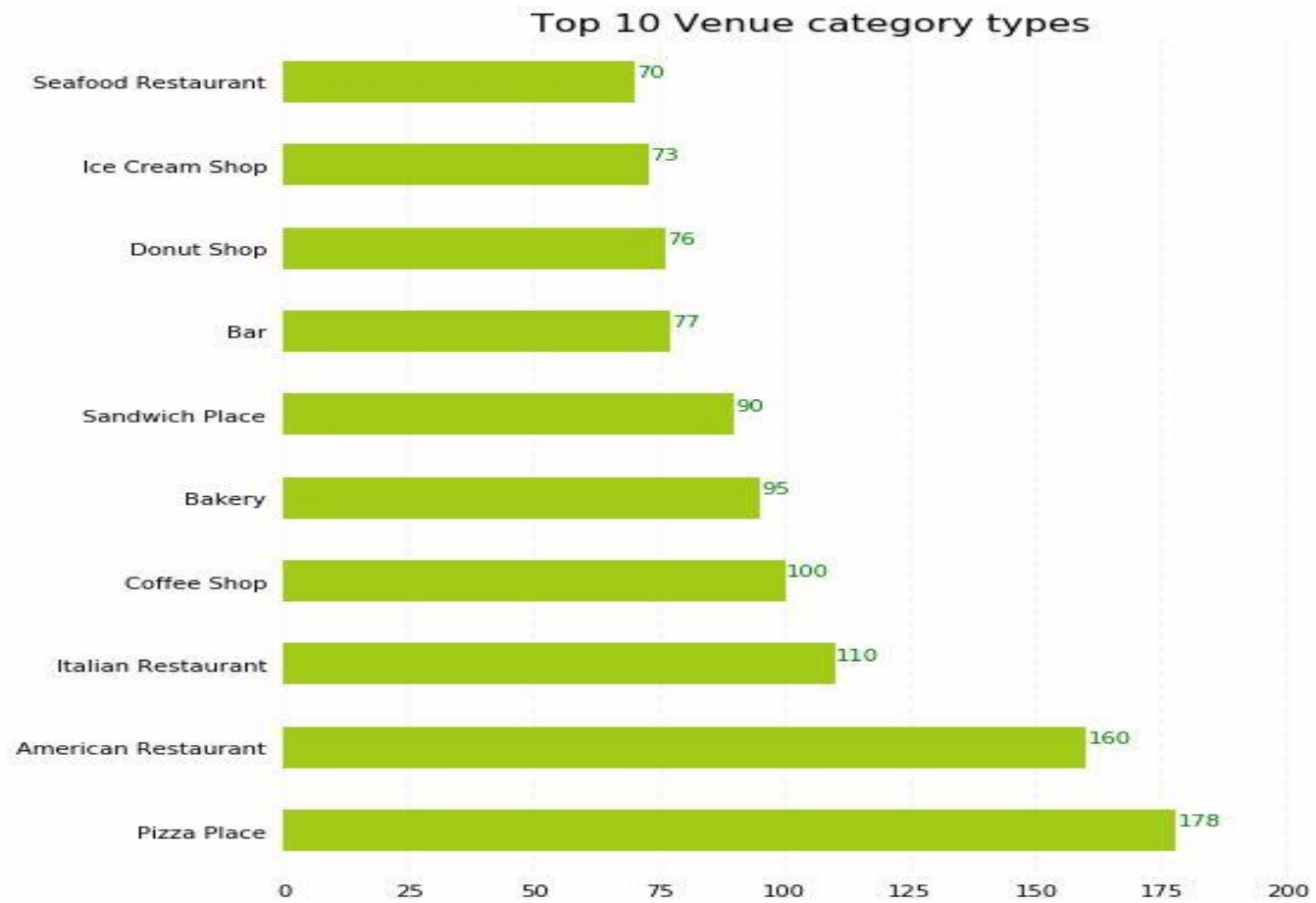
First I evaluate the venues by NECTAs

Second look at the top venues

Top 10 Venue category types

Then third proceed to the clustering process. Below are some of the outputted clusters

Profile of cluster with ID=0:
  It has 5 members (16.7% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Coffee Shop	6.750512	13.2
1	Bakery	5.787722	11.3
2	Pizza Place	5.424646	10.6
3	Café	5.392715	10.5
4	American Restaurant	5.371898	10.5
5	Park	5.242297	10.2
6	Brewery	5.096072	9.9
7	Breakfast Spot	4.173758	8.1
8	Hotel	4.155511	8.1
9	Italian Restaurant	3.857003	7.5

Profile of cluster with ID=1:
  It has 2 members (6.7% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Fast Food Restaurant	31.25	31.2
1	Discount Store	18.75	18.8
2	Pizza Place	6.25	6.2
3	Grocery Store	6.25	6.2
4	Hotel	6.25	6.2
5	Italian Restaurant	6.25	6.2
6	Diner	6.25	6.2
7	Pharmacy	6.25	6.2
8	Mexican Restaurant	6.25	6.2
9	BBQ Joint	6.25	6.2

Profile of cluster with ID=2:
  It has 4 members (13.3% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	American Restaurant	8.221167	16.6
1	Donut Shop	6.573545	13.2
2	Pizza Place	5.650388	11.4
3	Italian Restaurant	5.119704	10.3
4	Coffee Shop	4.764175	9.6
5	Sandwich Place	4.087022	8.2
6	Breakfast Spot	4.058336	8.2
7	Fast Food Restaurant	4.039829	8.1
8	Bar	3.616937	7.3
9	Bakery	3.517591	7.1

Profile of cluster with ID=3:
  It has 2 members (6.7% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency i
0	Donut Shop	9.692513	
1	Pizza Place	8.783422	
2	Coffee Shop	6.403743	
3	Pharmacy	6.403743	
4	American Restaurant	6.229947	
5	Convenience Store	5.842246	
6	Bar	3.288770	
7	Asian Restaurant	3.288770	
8	Ice Cream Shop	3.288770	
9	Sandwich Place	3.288770	

Profile of cluster with ID=4:
  It has 3 members (10.0% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Italian Restaurant	11.675864	23.2
1	Pizza Place	9.259584	18.4
2	Deli / Bodega	4.406330	8.8
3	American Restaurant	4.356369	8.7
4	Sandwich Place	3.901501	7.8
5	Mexican Restaurant	3.467954	6.9
6	Donut Shop	3.425310	6.8
7	Bakery	3.411305	6.8
8	Park	3.396672	6.8
9	Ice Cream Shop	2.956436	5.9

Profile of cluster with ID=5:
  It has 1 members (3.3% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Bar	14.516129	20.0
1	Café	11.290323	15.6
2	New American Restaurant	8.064516	11.1
3	Coffee Shop	8.064516	11.1
4	Italian Restaurant	8.064516	11.1
5	Restaurant	6.451613	8.9
6	American Restaurant	4.838710	6.7
7	Pizza Place	4.838710	6.7
8	Asian Restaurant	3.225806	4.4
9	Bakery	3.225806	4.4

Profile of cluster with ID=6:
It has 10 members (33.3% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Pizza Place	8.635783	18.9
1	American Restaurant	8.627573	18.9
2	Sandwich Place	4.753187	10.4
3	Pharmacy	3.876381	8.5
4	Bakery	3.678704	8.1
5	Italian Restaurant	3.615008	7.9
6	Ice Cream Shop	3.324730	7.3
7	Donut Shop	3.060439	6.7
8	Coffee Shop	3.046263	6.7
9	Mexican Restaurant	2.998403	6.6

Profile of cluster with ID=7:
It has 2 members (6.7% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Pizza Place	7.174658	12.8
1	Seafood Restaurant	7.174658	12.8
2	American Restaurant	6.549658	11.6
3	Chinese Restaurant	6.549658	11.6
4	Bar	5.864726	10.4
5	Diner	5.804795	10.3
6	Breakfast Spot	5.239726	9.3
7	Coffee Shop	4.049658	7.2
8	Bakery	3.929795	7.0
9	Italian Restaurant	3.929795	7.0

Profile of cluster with ID=8:
It has 1 members (3.3% of total venues)

	Venue Category	Venue Mean Frequency	Venue Frequency in top-10
0	Bakery	15.384615	22.2
1	Café	9.615385	13.9
2	Pizza Place	7.692308	11.1
3	Bookstore	5.769231	8.3
4	Brewery	5.769231	8.3
5	Sandwich Place	5.769231	8.3
6	Park	5.769231	8.3
7	New American Restaurant	5.769231	8.3
8	Theater	3.846154	5.6
9	Mexican Restaurant	3.846154	5.6

5. Discussion

The objective of this analysis was to shorten the search process of finding a place to live in New England by analyzing the most popular "venue profiles" for clusters of counties

The ten most common venue categories of each cluster that we identified are enough to reveal differences between them and allow students to imagine which "client profile" fits best

based on venues like:
- Coffee shop
- Book store
- Restaurants
- Theater
- Pizza
- Parks

Cluster venues and matches based on the venues above.
1. Cluster 0 - Has about 17% of the desired venues above
2. Cluster 1 - Has about 7% of the desired venues above
3. Cluster 2 - Has about 13% of the desired venues above
4. Cluster 3 - Has about 7% of the desired venues above
5. Cluster 4 - Has about 10% of the desired venues above
6. Cluster 5 - Has about 3% of the desired venues above
7. Cluster 6 - Has about 33% of the desired venues above
8. Cluster 7 - Has about 7% of the desired venues above
9. Cluster 8 - Has about 3% of the desired venues above

Cluster 6 is the strongest with about 33% of the prime venues

6. Conclusion

The analysis showed here, albeit simple, successfully identified 9 clusters of counties in New England with different venues profiles that can be mapped to different students wishes.

The cluster that best matches our targeted "venue requirements" is cluster 6 with 33%. we can set new desirable venues and cluster based on them.

Many enhancements to this analysis can be done to help students pick an ideal area to live:

- How close is the NECTA to the college
- What are the types of public transportation exist between the NECTAs
- College ranking and locations and admission rate.