

***Destination New England!***

***What new Students are coming to!***

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

Youssou Fall  
March 2019



# 1. Introduction/Business Problem

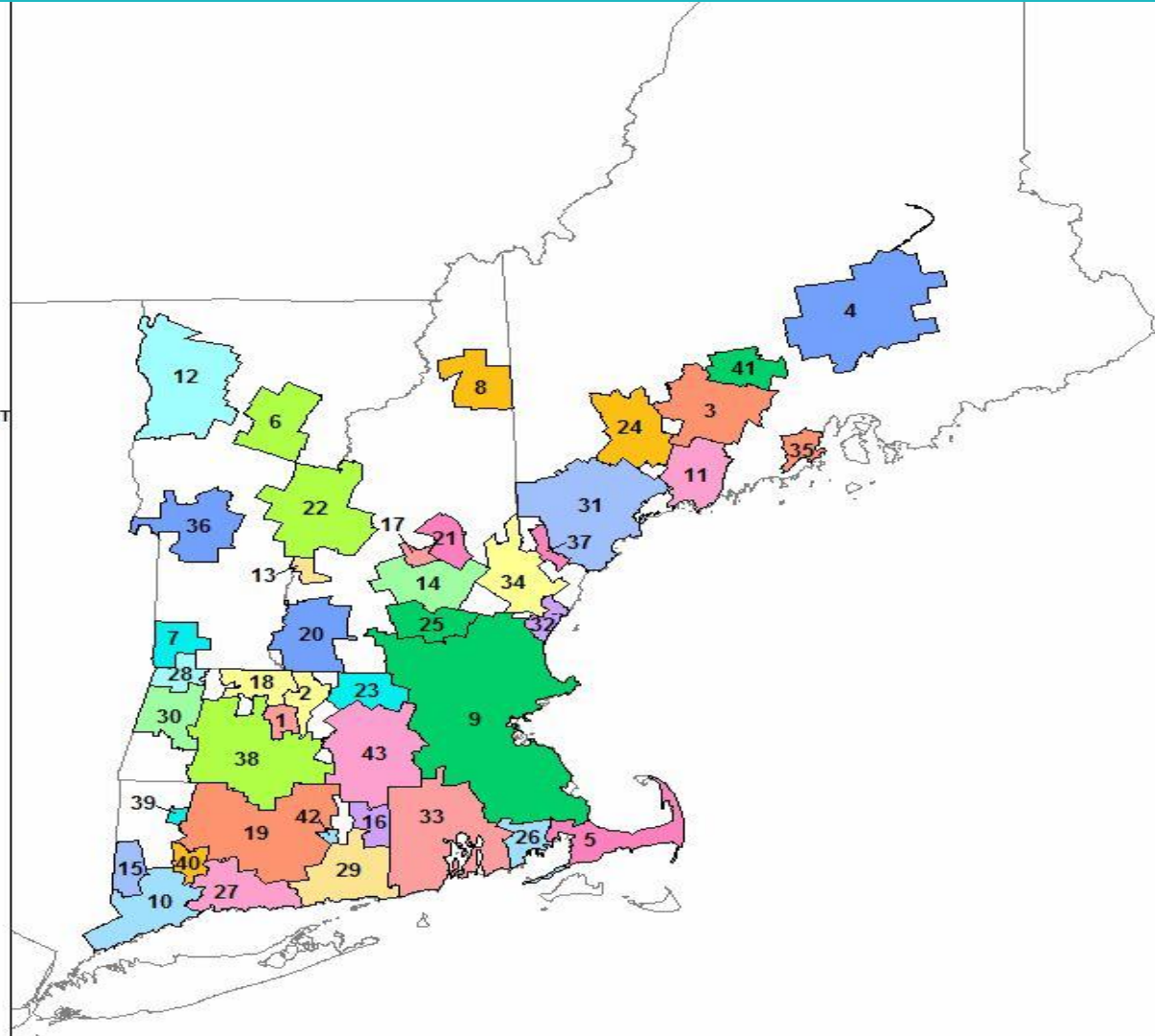
## 1.1. Problem definition:

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 metropolitan 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 metropolitan NECTAs. A metropolitan 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.

# *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



# 1. Introduction/Business Problem

## 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. 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](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,

## 2. Data to support this project

**- Nomination from geopy.geocoders for geocoding the County names and get their coordinates.**

With Nomination 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.

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%

# 3. Methodology

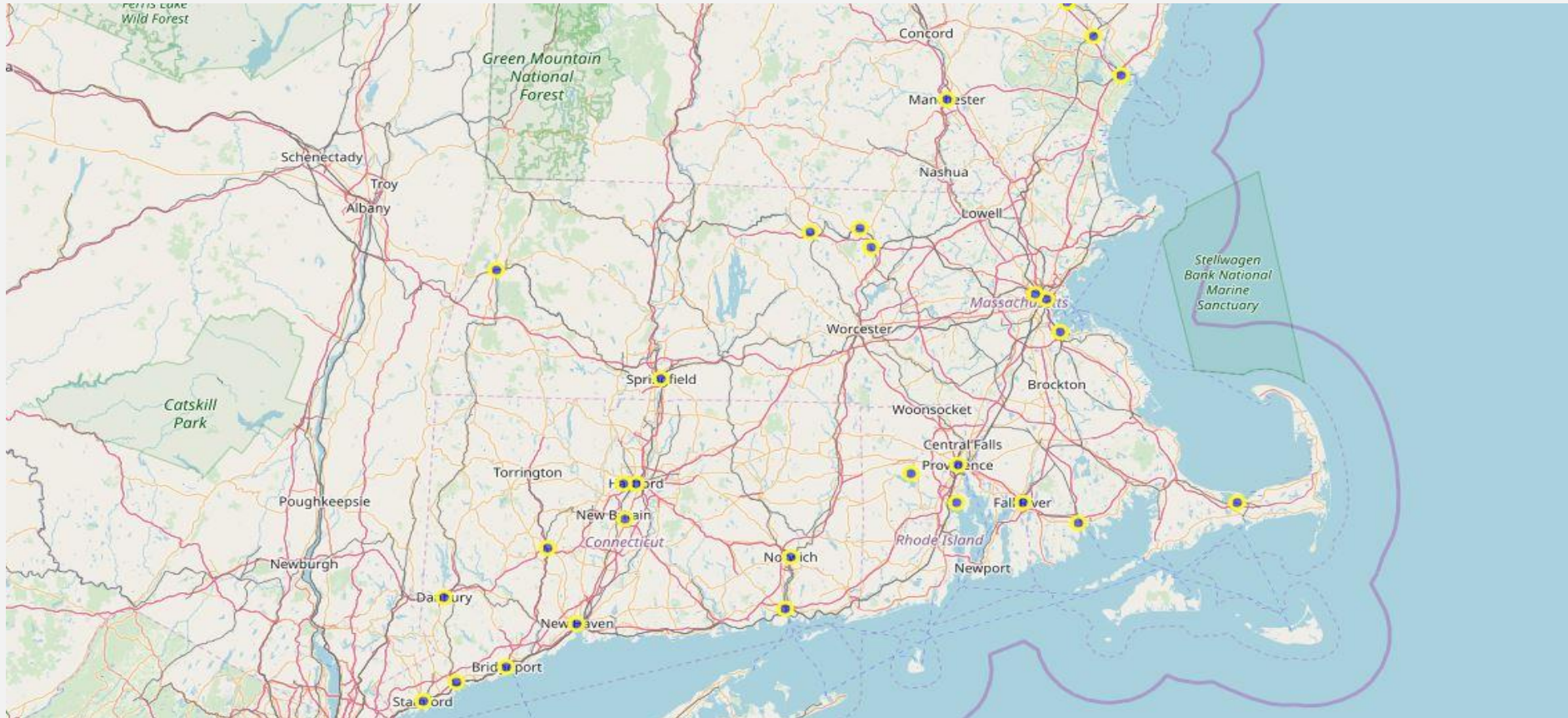
*The methods used in this work were:*

1. Web Scraping with the BeautifulSoup library
2. Geocoding with Nominalism 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](https://en.wikipedia.org/wiki/New_England_city_and_town_area) that contains a list of all towns in New England.

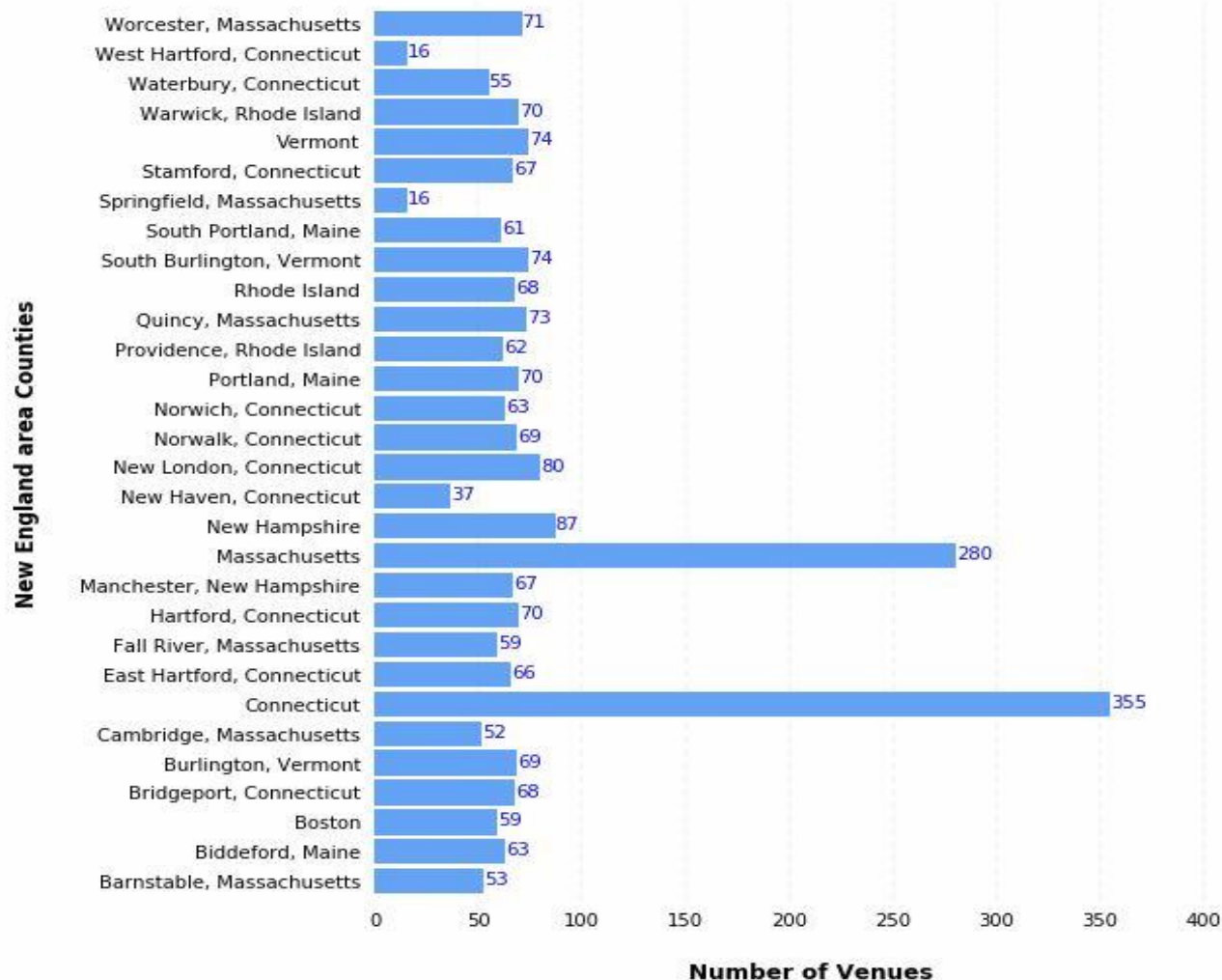




## 4. Results

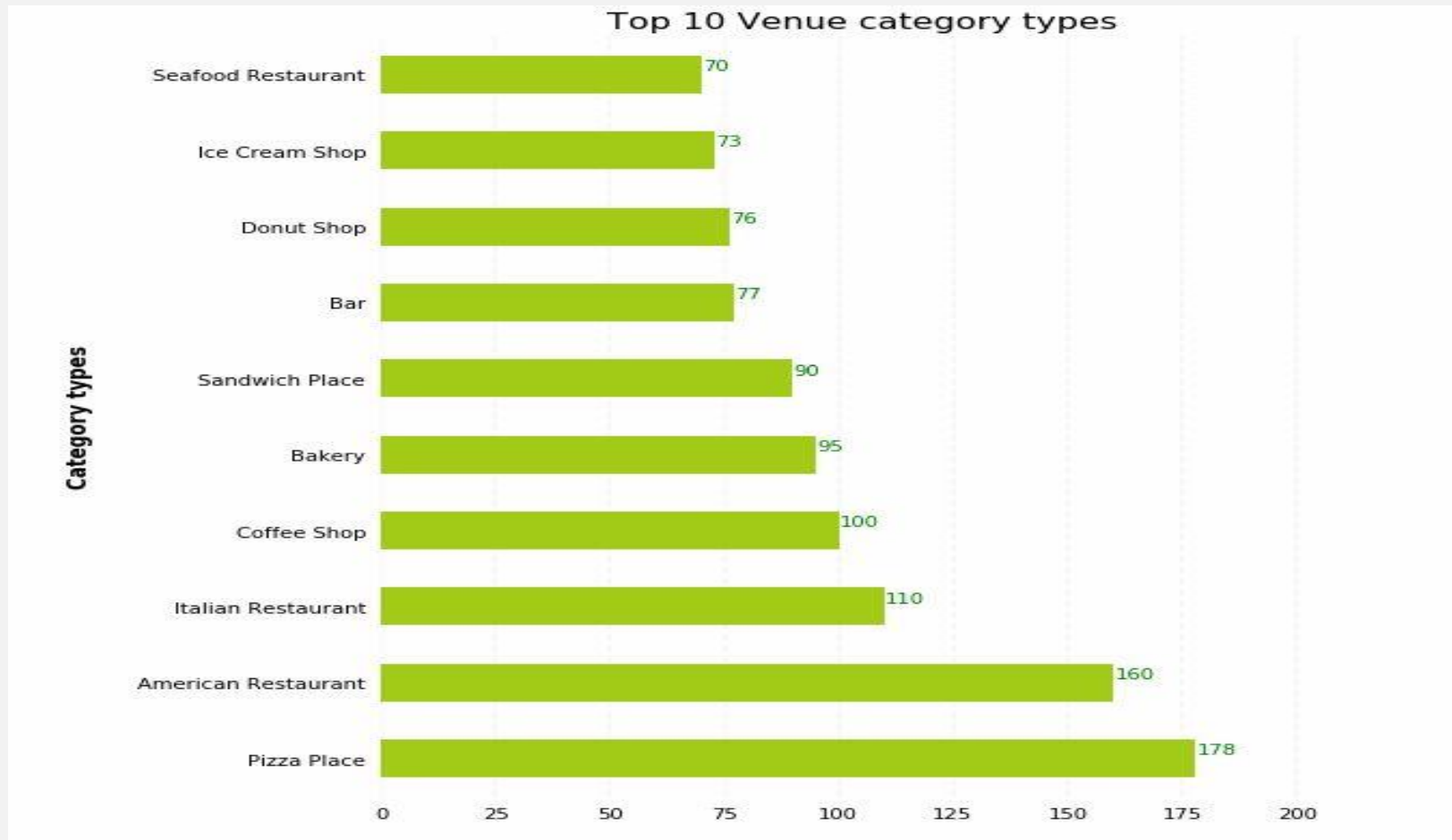
### 4.2. Mapping the NECTAs with clusters.

- 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



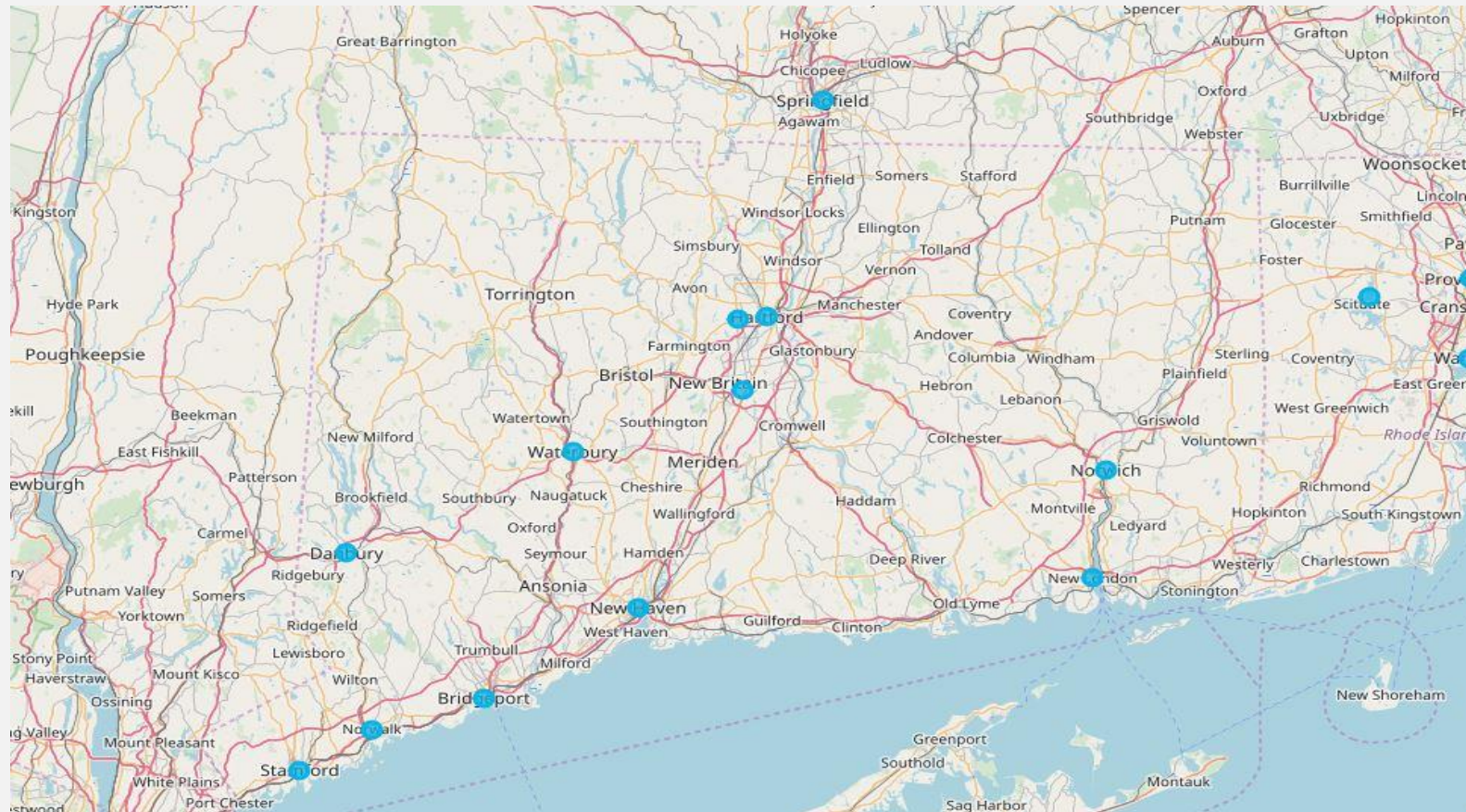
# 4. Results

Second look at the top venues

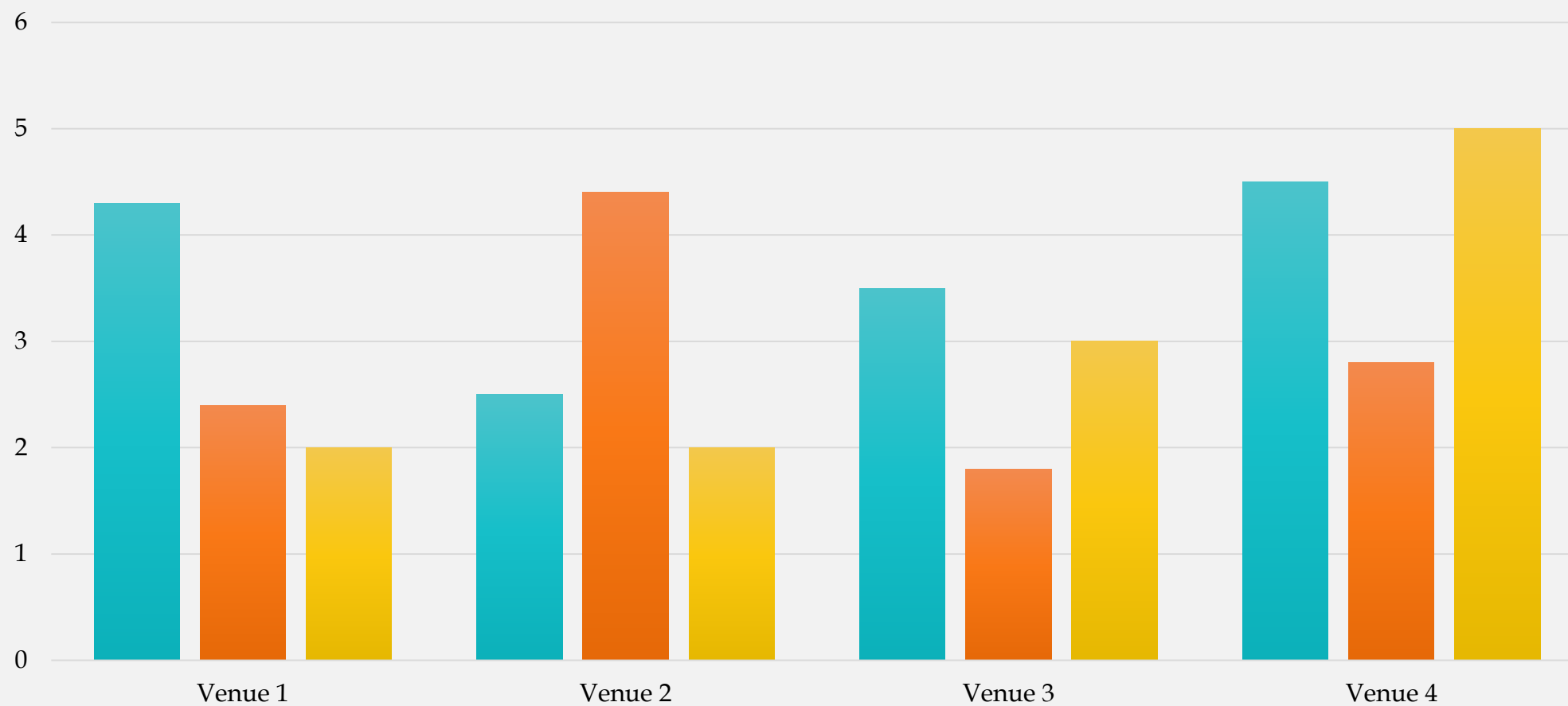


# 4. Results

## Clustered NECTAs



## 4. Results



Venues strengths on top NECTAs



## 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.