**Big 10 Schools Environment Study**

**Introduction**

According to the American Psychological Association, mental health within college students is a growing concern, since the number of students with significant psychological problems has increased during the last years.

Anxiety is the top presenting concern among college students (41.6 percent), followed by depression (36.4 percent). On average, 24.5 percent of students take psychotropic medications. Additionally, 40 percent present mild mental health concerns.

Is necessary that students keep a healthy life where they can be involved in activities outside of campus according to their interests. Yet, in some cases, college towns don't have as much diversity on the activities available for students and therefore are not a good fit for some individuals. As a result, incoming undergraduates and graduates should be able to study their options weighting the kind of activities available.

In this study, we will analyze the venues in the Big 10 member colleges, and we will cluster them to analyze what are the most prominent venues and how would that affect life of students.

**Data**

Since the study is applied to Big 10 Schools, they are extracted from an online list. Then, the location of each is found through geopy. Therefore, we obtain a list as shown below.

| **School** | **Latitude** | **Longitude** |
| --- | --- | --- |
| Northwestern University | 42.055116 | -87.675811 |
| University of Michigan | 42.294214 | -83.710039 |
| University of Minnesota | 44.973086 | -93.237088 |
| University of Maryland | 38.992030 | -76.946103 |
| University of Wisconsin | 44.851978 | -92.616756 |
| The Ohio State University | 40.005709 | -83.028663 |
| Penn State University | 40.803448 | -77.859119 |
| Rutgers University | 40.500076 | -74.445792 |
| Purdue University | 40.431958 | -86.923911 |
| University of Illinois | 40.141058 | -88.255202 |
| Michigan State University | 42.719204 | -84.477920 |
| University of Nebraska-Lincoln | 40.820739 | -96.700491 |
| Indiana University | 39.169100 | -86.519790 |
| University of Iowa | 41.665901 | -91.573188 |

For this project Foursquare databases will be used as the main source to find the venues on each school. They will be accessed by explore requirements on the location of each university. From there, we will find the name, category and location of the venues. This data will be augmentated by adding the broad category of each venue with the use of the Foursquare Venue Category Hierarchy JSON file, that will be flattened to obtain all categories and the corresponding parent (broad) category. A sample of the data is shown below.

|  | **School** | **School Latitude** | **School Longitude** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Specific Venue Category** | **Broad Category** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Northwestern University | 42.055116 | -87.675811 | Centennial Park | 42.047476 | -87.674032 | Park | Outdoors & Recreation |
| **1** | Northwestern University | 42.055116 | -87.675811 | Bennison's Bakery | 42.046891 | -87.685028 | Bakery | Food |
| **2** | Northwestern University | 42.055116 | -87.675811 | Segal Visitors Center | 42.050249 | -87.673408 | College Administrative Building | College & University |

The broad categories are Arts & Entertainment, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Shop & Service and Travel & Transport. For each category, different types of specific venue are included, for instance:

-Arts and Entretainment has specific categories as Aquariums, Art Galeries, Bowling Alleys, Museums, Concert Halls, Stadiums and Theme parks.

-Food, has all food related venues as restaurants of all types, coffee shops, ice cream shops, bakeries, iries pubs, and many others.

-Nightlife Spots, include bars, breweries, lounges, nightclubs and others.

-Outdoors and Recreations, includes all kind of sports courts and fields, beaches, bike trails, forests, marina, gyms, lakes, national parks, pools and rivers.

-Professional and other places, includes auditoriums, ballrooms, buildings, business centers, convention centers, factories, event venues, Government buildings, libraries, medical centers, prisons, schools, spiritual centers, waste facilities and others.

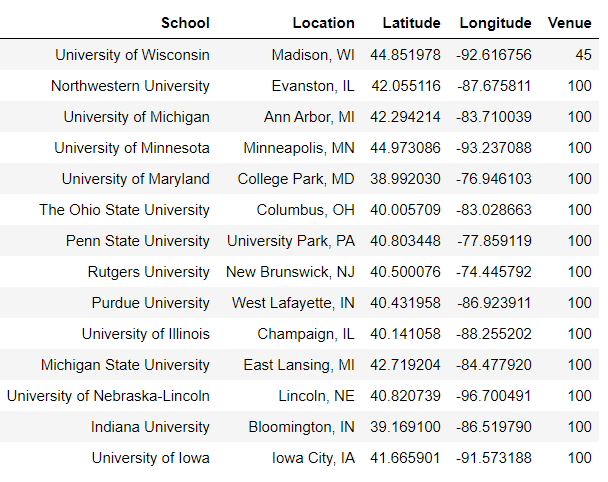
-Shop and Services, include banks, auto workshops, all kind of stores, pharmacies, grocery stores, markets, gas stations, it services, insurance offices, laundromats, nail salons and many others.

- Travel and transport, include airports, bike or boat rentals, bus stations, hotels, ports, RV Parks and similar.

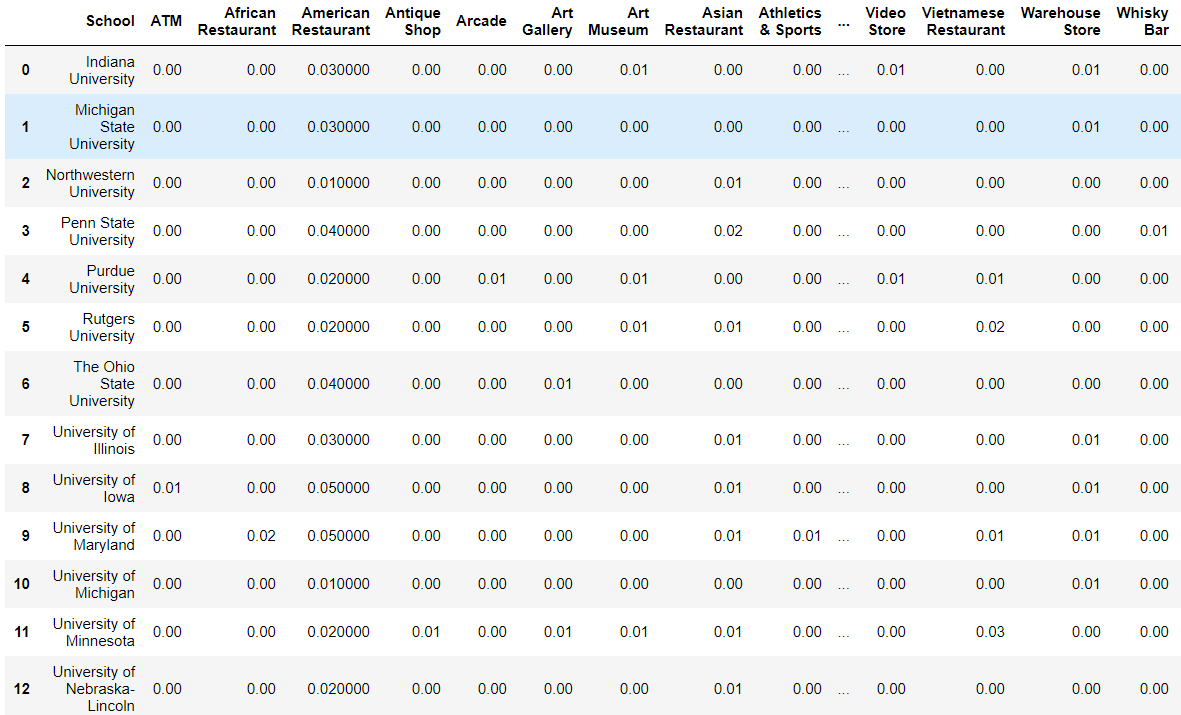
They venues will be analyzed in both specific and broad venue categories to profile each university. The percentage of each of the categories in the town will be used to cluster and analyze the advantages and disadvantages of each location for different types of personalities, and explore possible business opportunities in the area.

**Methodology**

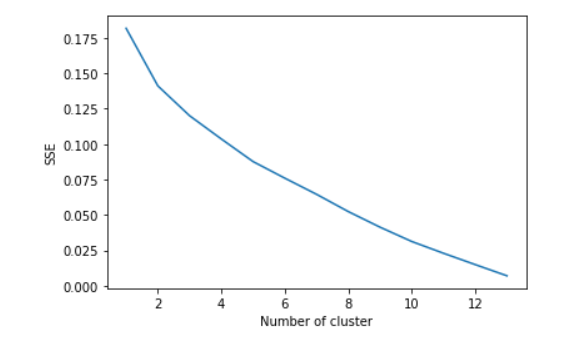
Once the data is cleaned and organized, the count of venues by university is done. The results are shown below.



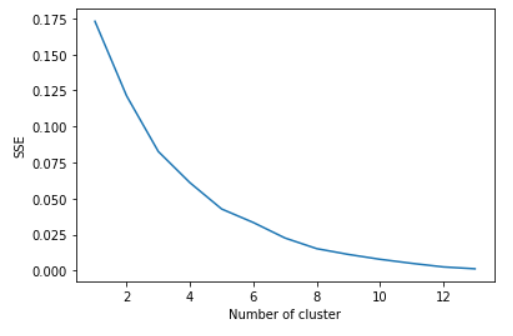
Then the analysis is done for the specific categories first. There were 230 distinct venue categories. In order to understand the distribution on this venues in each city, the venue categories are translated in dummy variables were each location is 1 in the variable correspondent to the category it belongs and 0 in all the others. Then, the venues are grouped by university, and then normalized to find the percentage of venues in the respective category, as shown below:



We want to find the college towns that are similar to each other to analyze them as a group and give students and investors different options, instead of a single choice. As a result, we cluster the schools through K-means. Then, we use the elbow method to find the number of groups the data should be divided in. A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered. The Elbow Method is one of the most popular methods to determine this optimal value of k. To determine the optimal number of clusters, we have to select the value of k at the “elbow” ie the point after which the distortion/inertia start decreasing in a linear fashion. The distortion, computes the sum of squared distances from each point to its assigned center.



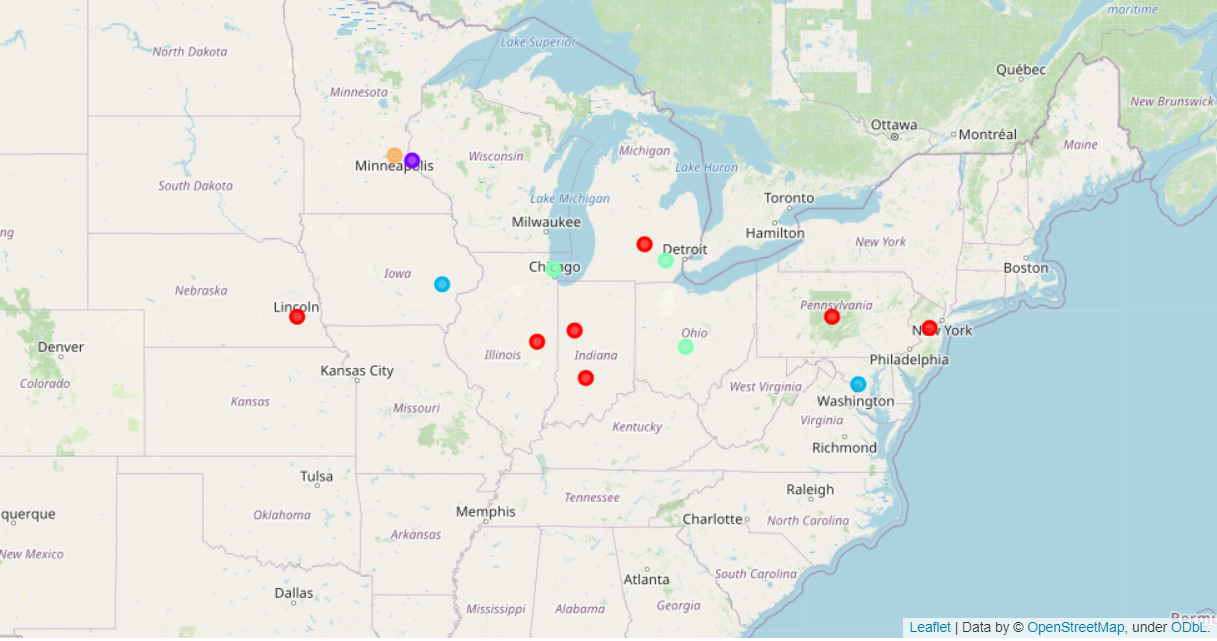
Thus for the given data, we conclude that the optimal number of clusters for the data is 5. The same process was repeated for the broad categories.



We find again an inflection at 8 and 5 clusters. Yet, 8 clusters would single too many universities and therefore, 5 clusters still describes the relationship between the samples, and is chosen to work with.

**Results**

The clusters are visualized in the next map.

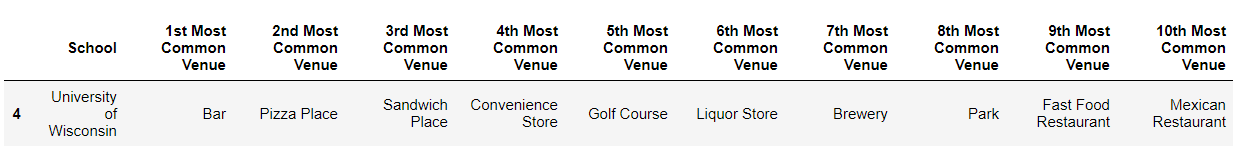


And we select each cluster and show the most common 10 venues of each university to understand the group.

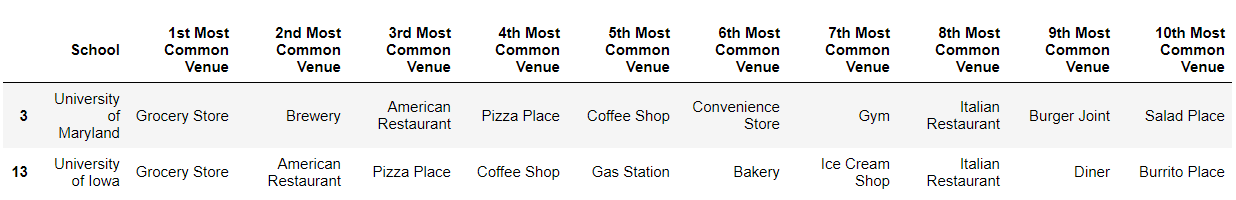
For group or cluster 0, the results shown below show a very heterogeneous group, which has a mixed environment between services and food places. Most of them, have coffee shops and fast food places.



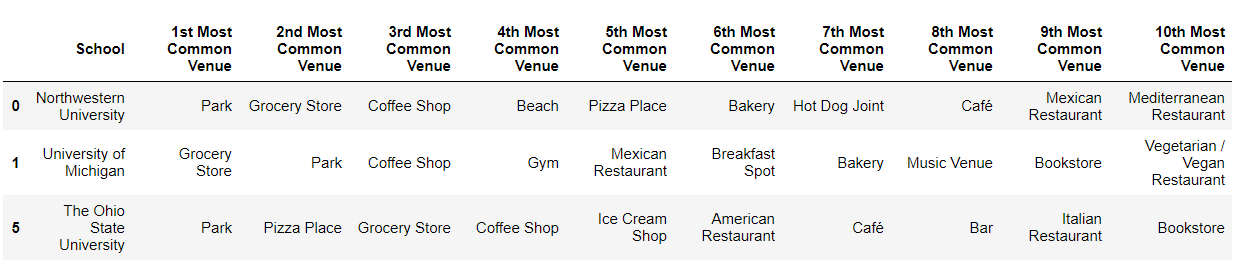
In cluster 1, we find only university of Wisconsin, that had the least number of venues, and the most common were bars in comparison to other universities.



In cluster 2, we have a common layout, with Grocery stores, american restaurants and pizza places being popular.



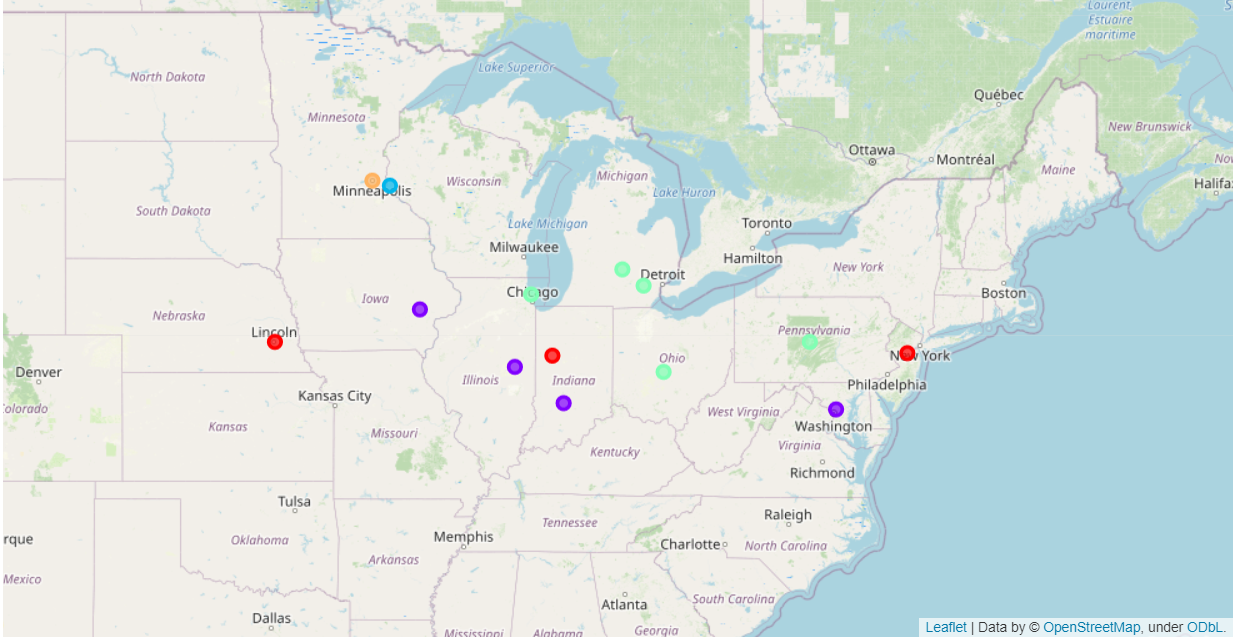
Cluster 3, shows more green places, were parks, coffee shops and grocery stores are relevant.



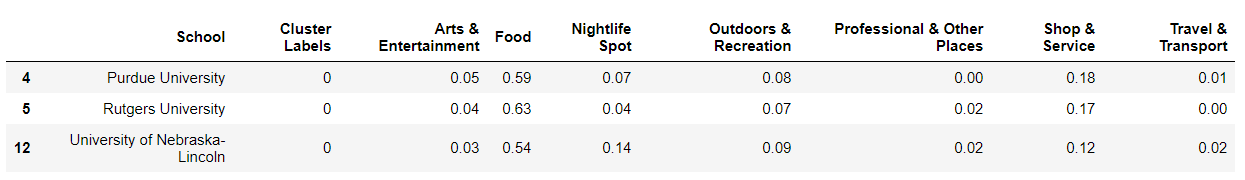
U of Minnesota is the last cluster, and it shows that breweries are the most popular venue followed by parks and theaters, showing a variety of interests.



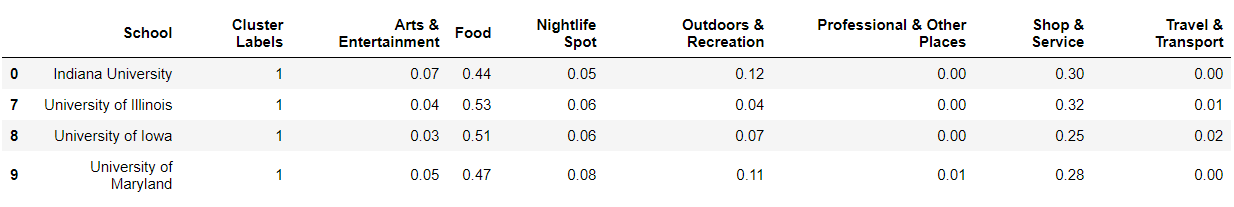
Similarly, we develop the same analysis for the broad categories. We start by showing the clusters on a map to visualize the data:



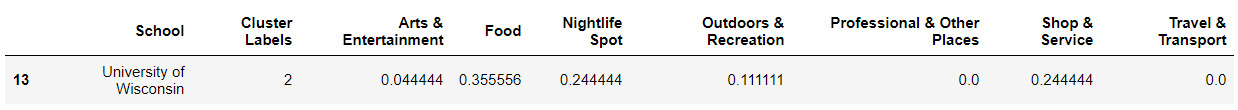
Then, we explore the fractions of each cluster since the order of the clusters is similar.



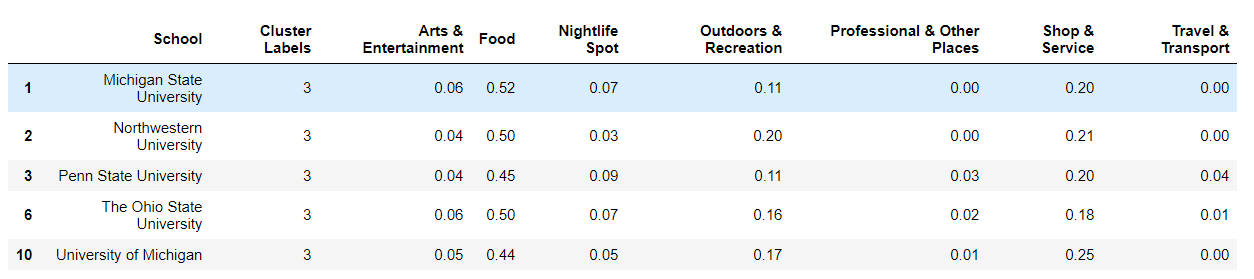
Cluster 0 shows a bigger ammount of food venues.



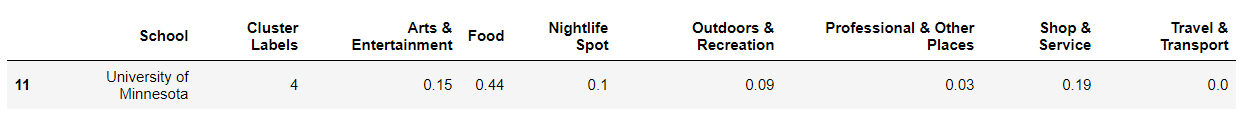
Cluster 1 is similar to cluster 0, but us shows lower values of food and higher in shops and services.



The clustering for Wisconsin continues to be unique for the university, showing a high interest on Night Life compared to most other universities.



Cluster 3, has high values for outdoors and recreation, as well as shop and service.



Cluster 4 continues to be unique and has a higher value for art and entertainment venues and a lower amount of food venues.

To visualize this distribution, we included a final graph.

**Analysis**

The clusters obtained by both specific and broad categories are able to give us insights in the type of life of each place.

Analyzing both clusters, it is obvious that University Minnesota, since its located in Minneapolis, a big city, has a unique profile that is adjusted to all kind of lives, therefore has a higher amount of arts and entertainment venues and professional places, in comparison to all other universities. This type university is for all types of personalities, it allows introverted people the opportunity to visit cultural venues, and outdoor activities, while at the same time, having a high amount of breweries, and theaters for the more extroverted and social.

Purdue University, Rutgers University and University of Nebraska Lincoln, also share clusters in both categories. These are all small cities, were the Universities are big drivers of population, therefore, the market is more oriented to students and there is less diversity on the venues. These universities tend to have higher amounts of food venues. These places are for students that can adapt to a lower amount of choices on activities and are not as interested in cultural activities.

University of Wisconsin showed the highest amount of nightlife spots. Therefore, it is a city for people who like to go out to drink and dance, and probably wont be a good fit for introverted people.

Indiana, Illinois, Iowa and Maryland universities, tend to have a higher amount of services each in its own unique way, yet, it seems that their location in or close to big cities, allows them to have a more balanced offer of venues, yet, they still have lower nightlife spots. These types of cities are more agreeable to all types of personalities.

Michigan State, Northwestern, Penn State, Ohio state and University of Michigan, have a higher amount of outdoors activities, as well as shop and services. This type of cities would not be the best fit for extroverted students.

**Conclusion**

The analysis of venues in each city allows for the aggregation of cities according to the profile of the venues close to the university, and allows an insight in the type of life each of the Universities provides according to their location.