**Ideal Places to Work**

**Segmenting and Clustering Universities in Manhattan**

*Ryan Brown*

*February 17, 2020*

Introduction

*Background*

Manhattan provides abundant opportunities for New Yorkers to pursue administrative careers in higher education, including positions at community colleges, research universities, and professional schools. And while Manhattan is the largest metropolis in the United States, its neighborhoods are diverse and each one spans the spectrum of dining, entertainment, and fitness options (among other activities). At some universities, countless restaurants, coffee shops, bars, and gyms are located just outside the campus gates. At others, faculty, students, and employees are stricken to very limited venue options located within walking distance.

*Problem*

We expect that prospective employees will consider the venue options located in the surrounding neighborhoods when searching for university administrative positions. However, it is unrealistic to expect that applicants will have familiarity with all of Manhattan’s universities and surrounding neighborhoods when submitting their applications. This paper will address this shortcoming by clustering 21 of Manhattan's private and public universities by the most common venues located within walking distance of each campus. The goal is to identify universities that share similar neighborhood venue categories by partitioning them into groups with similar characteristics.

*Interest*

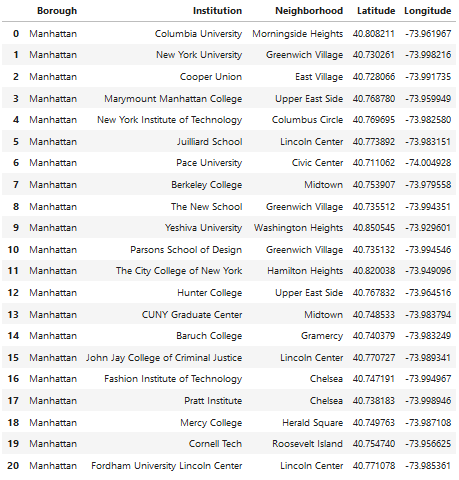
Prospective employees will benefit from this analysis as it will provide additional information about a particular university that is not otherwise available. And while selecting to work at a particular university should be based on many factors, including workplace fit, career progression, and employee benefits, it can be argued that nearby venues contribute to an employee's work-life balance and, ultimately, job satisfaction.

Data Description

*University Coordinates*

A dataset containing the geographic coordinates of Manhattan’s universities was created to identify the location of 21 Manhattan universities. Using Google Map, the latitudinal and longitudinal coordinates of the universities were recorded and stored in a *pandas* dataframe, along with the names of the corresponding neighborhood and borough (see Figure 1). The 21 universities were selected because they encompass most areas of Manhattan.

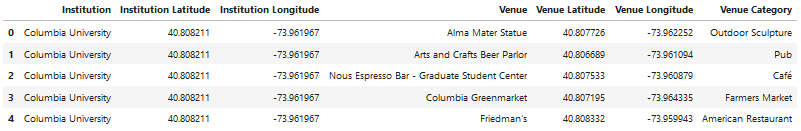
Figure 1



*Foursquare Venues*

The Foursquare API was utilized to obtain data on the venues located near each university, including the venue’s name, latitudinal coordinate, longitudinal coordinate, and category. In order to capture the walking distance of a typical employee, a radius of 500 meters was established from each campus’s coordinates. Moreover, since Manhattan is such a large metropolis, a limit of the top 100 venues was set in place for each university. The Foursquare information was merged with the university’s information from Figure 1. The resulting dataframe contained 1841 venues for the 21 universities (see Figure 2 for a sample).

Figure 2

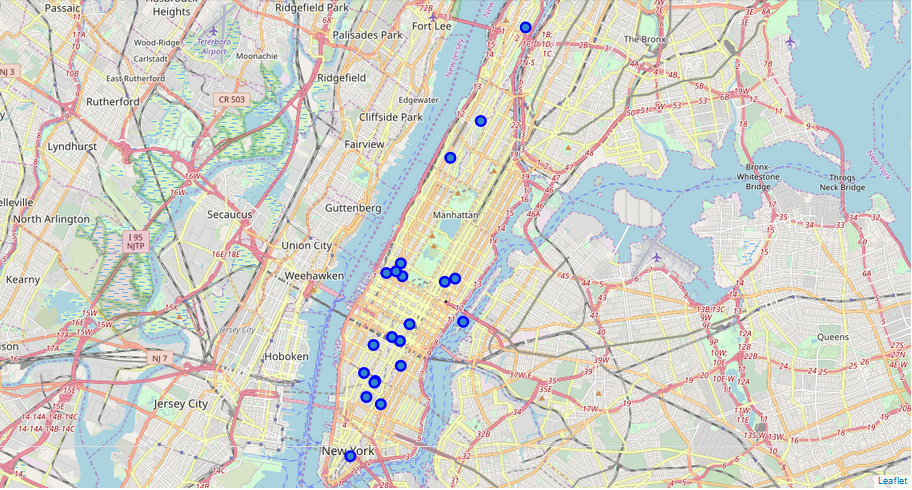


Methodology

*Exploratory Data Analysis*

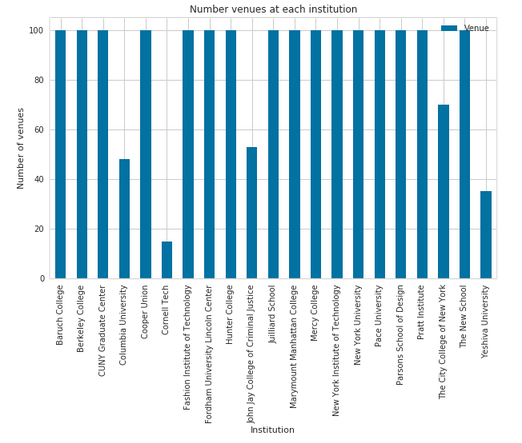
After confirming that each university’s longitudinal and latitudinal coordinates were properly loaded into thedataframe, a *folium* map was created to visually explore each university’s location. As seen in Figure 3, the 21 universities are spread across Manhattan. Although efforts were made to select geographically diverse universities, the resulting map shows that majority of universities are located in the central and southern part of the borough.

Figure 3



The Foursquare API returned 249 unique venue categories and revealed that most of the 21 universities reached the 100-venue limit (see Figure 4). This suggests that there are more venues options for these universities that are not captured in this analysis. However, it also shows that five universities—Columbia University, Cornell Tech, John Jay College of Criminal Justice, the City College of New York, and Yeshiva University—have far fewer venue options when compared to the other universities. In fact, Cornell Tech only has 15 venues listed within walking distance!

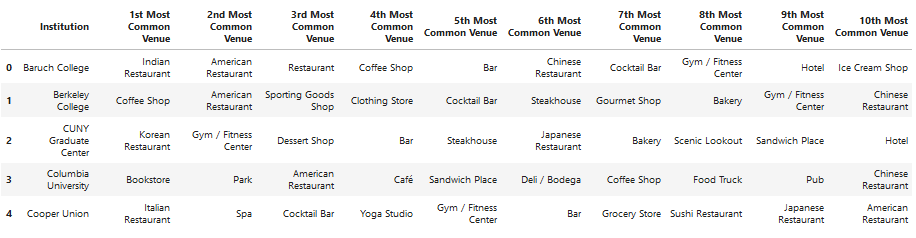
Figure 4



*k-means Clustering*

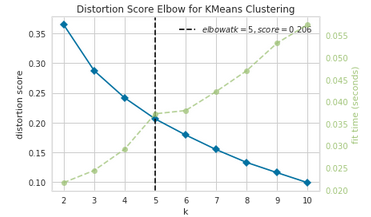
k-means clustering was selected to cluster the 21 universities according to their 10 most common venues. The purpose of k-means clustering is to divide the data into k non-overlapping clusters where objects within a cluster are very similar and objects across different cluster are very dissimilar. This method was chosen due to its simplicity as a machine learning algorithm and its applicability across science applications. Before running the k-means algoritm, it was necessary to 1) perform “one hot encoding;” and 2) group the dataframe’s rows by each university and take the mean of the frequency of occurrence of each venue category. The resulting dataframe lists the top 10 venue categories for each university (see Figure 5).

Figure 5



With the top ten common venues by university established, the universities were then clustered using k-means. The Elbow Method was selected to find the optimal number of clusters by fitting the model with a range of values for k. Figure 6 shows that “5” lies at the elbow of the curve and is a good indication that the underlying model fits best at this point.

Figure 6



The results of the k-means clustering were inputted into a new dataframe, which merged the new cluster labels with the previous dataframe (see Figure 7 for a sample). The resulting five clusters were examined to determine the discriminating venue categories that distinguish each cluster. The difference across the clusters were apparent.

Figure 7



Results

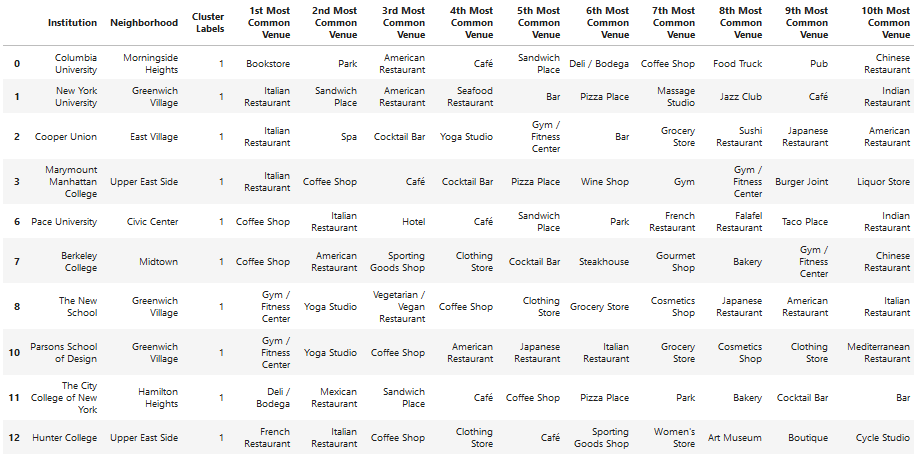
*Cluster 1 – Koreatown*

The universities in Cluster 1 are located near Manhattan’s Koreatown, which accounts for Korean restaurants occupying the first most common venue for both universities. The other categories listed in this cluster are similar to those listed in Cluster 2.



*Cluster 2 – Restaurants, Coffee Shops, and Bars Everywhere*

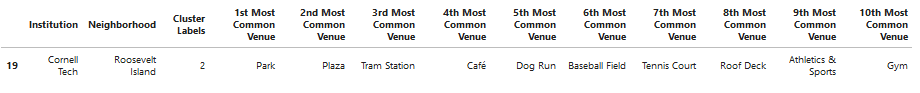
The universities in Cluster 2 offer the most varied venues. They are primarily restaurants, coffee shops, and fitness centers. 14 of the 21 universities selected for analysis fall in this cluster.





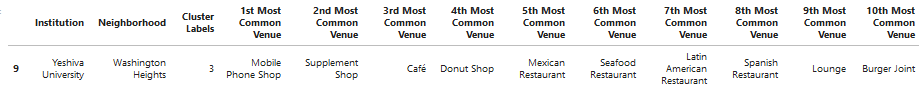
*Cluster 3 – Enjoying the Outdoors Due to Minimal Options*

Cluster 3 consists of Cornell Tech, which is located on the relatively isolated Roosevelt Island (still part of Manhattan). It holds the fewest venue options of the 21 universities, most of which involve the outdoors. Its common venue types contrast those found in the other clusters.



*Cluster 4 – Consumer Stores and Hispanic Community Offerings*

Cluster 4 consists of Yeshiva University, which is located in the Washington Heights neighborhood on the northern end of Manhattan. This neighborhood is home to a large percentage of Hispanic residents, which is reflected in the venue categories.



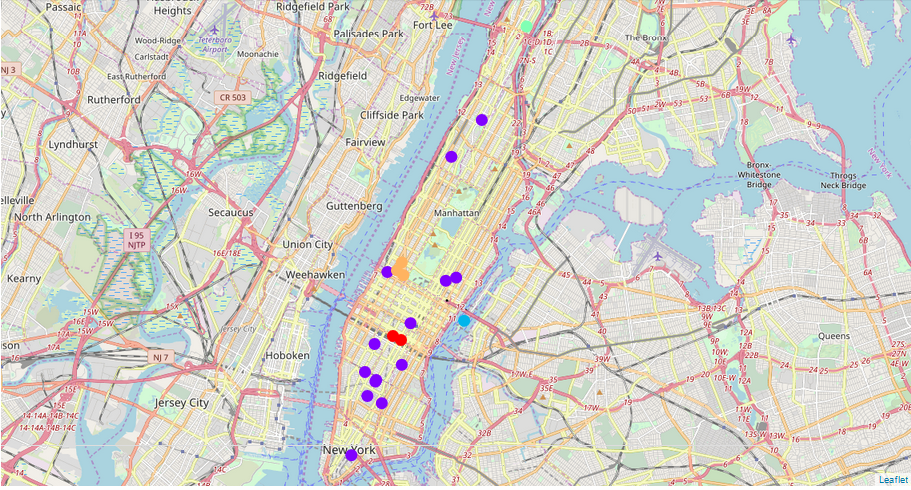
*Cluster 5 – Performing Arts Mecca*

The universities in this cluster are located near Lincoln Center, which home to the New York Philharmonic, the Metropolitan Opera, the New York City Ballet, and other performing arts organizations. Consequently, the most common venue options for these universities involve theater and performing arts. These categories of venues were not found in the other clusters.



After clustering and segmenting the universities, it is helping to visualize the clusters on a map. Figure 8 shows the location of the universities in Cluster 1 (red dots), Cluster 3 (light blue dot), Cluster 4 (teal dot), and Cluster 5 (yellow dots). The remaining purple dots belong to Cluster 2 and are found throughout Manhattan.

Figure 8



Discussion

The resulting clustering shows clear differences in the venues surrounding 21 of Manhattan’s universities. The most common venues for 14 of the universities include restaurants, coffee shops, and fitness centers. Other universities, notable Cornell Tech, lacks surrounding venues. In fact, most of its common venues are dissimilar to all other universities. Other universities, such as CUNY Graduate Center, Mercy College, and Yeshiva University, are in distinct neighborhoods, which are reflected in their common venue categories.

These universities are dispersed throughout Manhattan and represent many of the universities that applicants apply to when looking for administrative positions. While it is no surprise that many universities are similar to one another with regards to their surrounding venues, it is interesting to see that certain universities are very dissimilar from all others. Applicants should use caution when applying to these universities if there are certain expectations of what the surrounding neighborhood should offer.

Conclusion

This report employed k-means clustering to identify universities that share venues within walking distances of their campus gates. There are other methods that could have achieved this clustering but they were outside the scope of this paper. Additionally, many other institutions of higher education in Manhattan were not included in this report, notably community colleges and professional schools. A possible extension of this report would be to determine how these institutions fit into the clusters.