|  |
| --- |
|  |
| Recommendation System for Places to Open a Food Court Chain in New York City |
|  |

|  |
| --- |
| S M Azizul Hakim  5-17-2019 |

**1. Introduction**

**1.1 Background**

The City of New York is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km2). New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world.

**1.2 Business Problem**

New York is the most diverse city in the United States. The city manages to bring dishes from all over the world and make them it's own. Over the past few decades, food courts have become an important way of experiencing the cities diverse food culture at a reasonable price. This project aims to help any business person interested in opening a food court chain in New York City.

New York is also rich in educational institutions. A lot of colleges and universities are located in this city. Our target customer will be the students of colleges and universities. Based on the location of different college places like administrative building, a college library, college gym, etc. this project will recommend the best places to open a food court in different neighborhoods.

**2. Data**

**2.1 Data Collection**

Situated on one of the world's largest natural harbors, New York City consists of five boroughs, each of which is a separate county of the State of New York. The five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island – were consolidated into a single city in 1898.

This 5 boroughs consist of 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This dataset exists for free on the web here- <https://geo.nyu.edu/catalog/nyu_2451_34572> We will then use Foursquare to get the college plases location. College & University ID for foursquare is: “4d4b7105d754a06372d81259”.

**2.2 Data Preprocessing**

The given data was initially in json format. So, it is converted into a pandas dataframe consisting of Borough, Neighborhood, Latitude and Longitude.

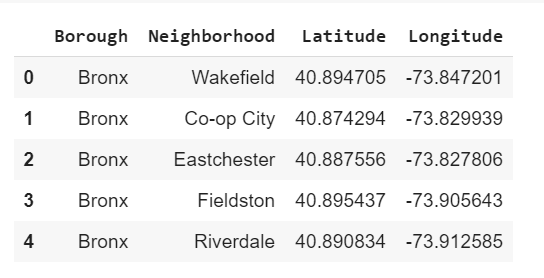


Figure 1: Initial dataframe consisting of neighborhood data

This dataframe has all the neighborhoods of 5 Boroughs and their geolocation. It is very easy therefore to insert them into a map using folium.

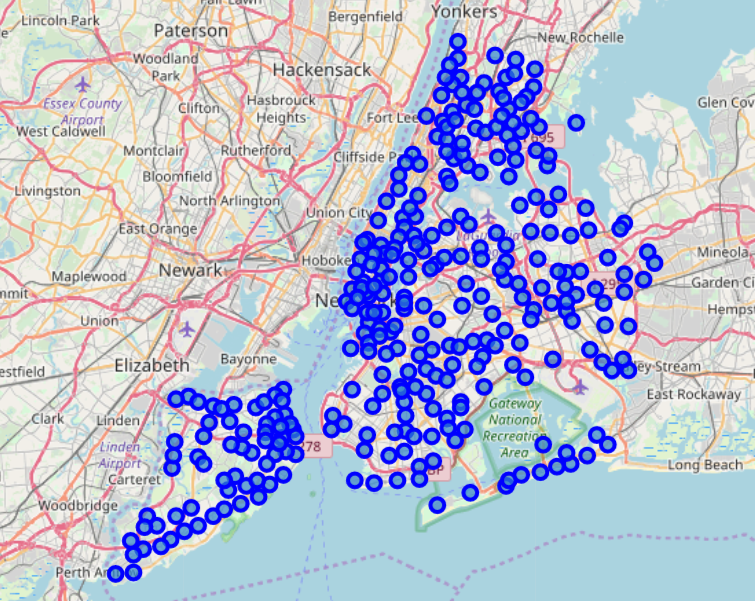


Figure 2: A map showing all the neighborhoods in New York City.

After that we used foursquare to find all the college places location in the New York City. Foursquare venue ID for colleges & universities is: “4d4b7105d754a06372d81259”. We acquired the geolocation of all these places and added them to the initial dataframe according to the neighborhood they are located in.

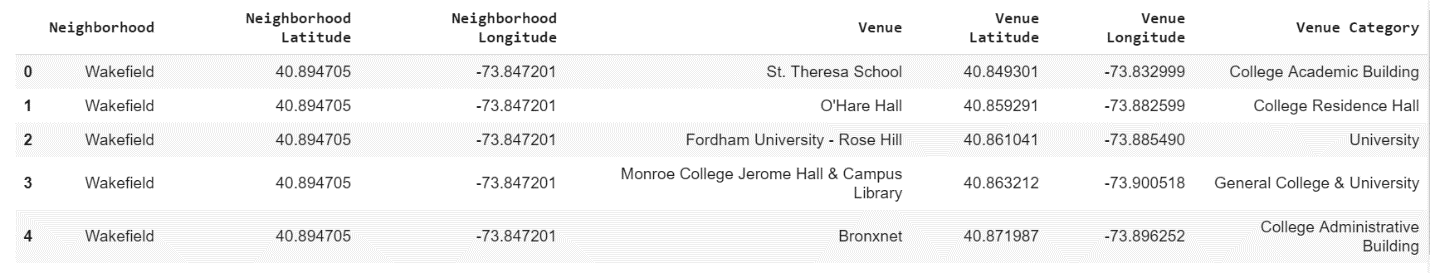


Figure 3: Final dataframe consist of the geolocation the neighborhoods and desired venues located in them.

**3. Methodology**

**3.1 Exploratory Data Analysis**

Final dataframe is grouped by ‘neighborhood’ attribute. This shows how many desired venues located in each neighborhoods.

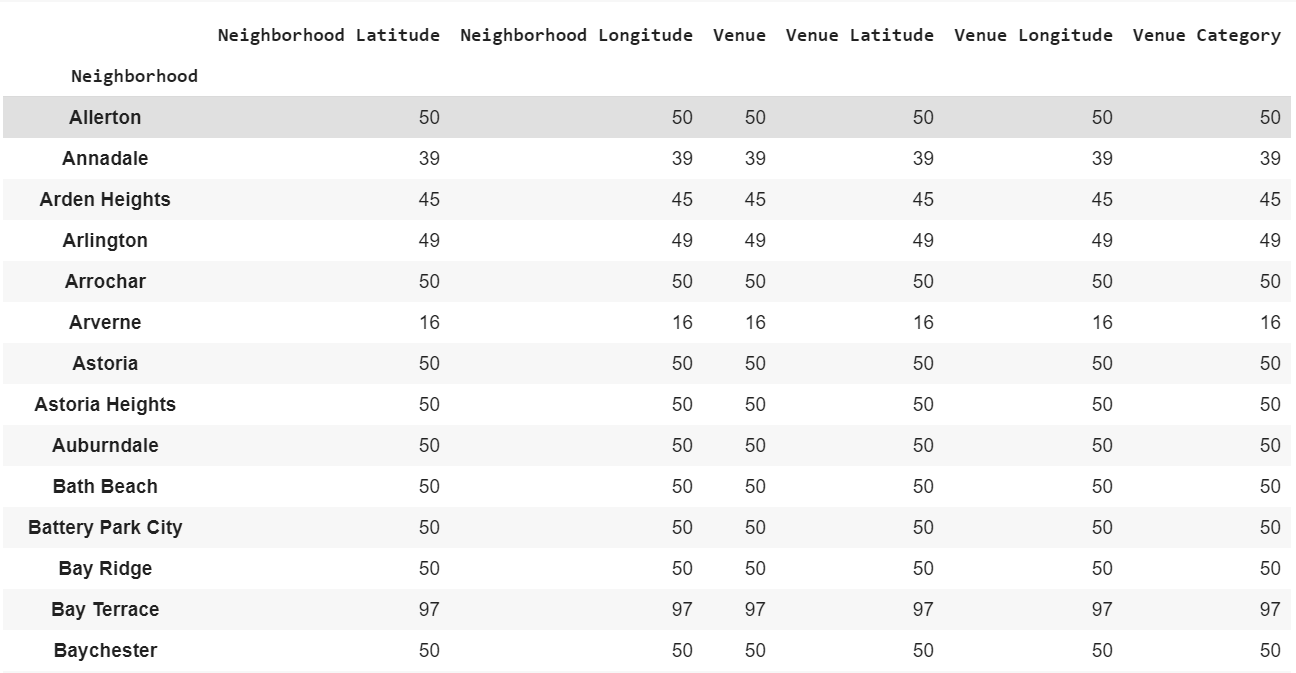


Figure 4: Number of desired venues in each neighborhoods.

There are 71 unique categories in the dataframe. They were then one hot encoded.

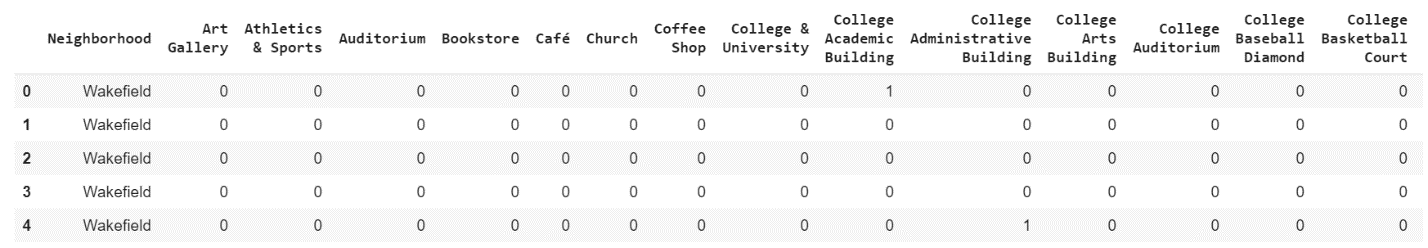


Figure 5: Dataframe with all the categories one hot encoded.

Then we grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

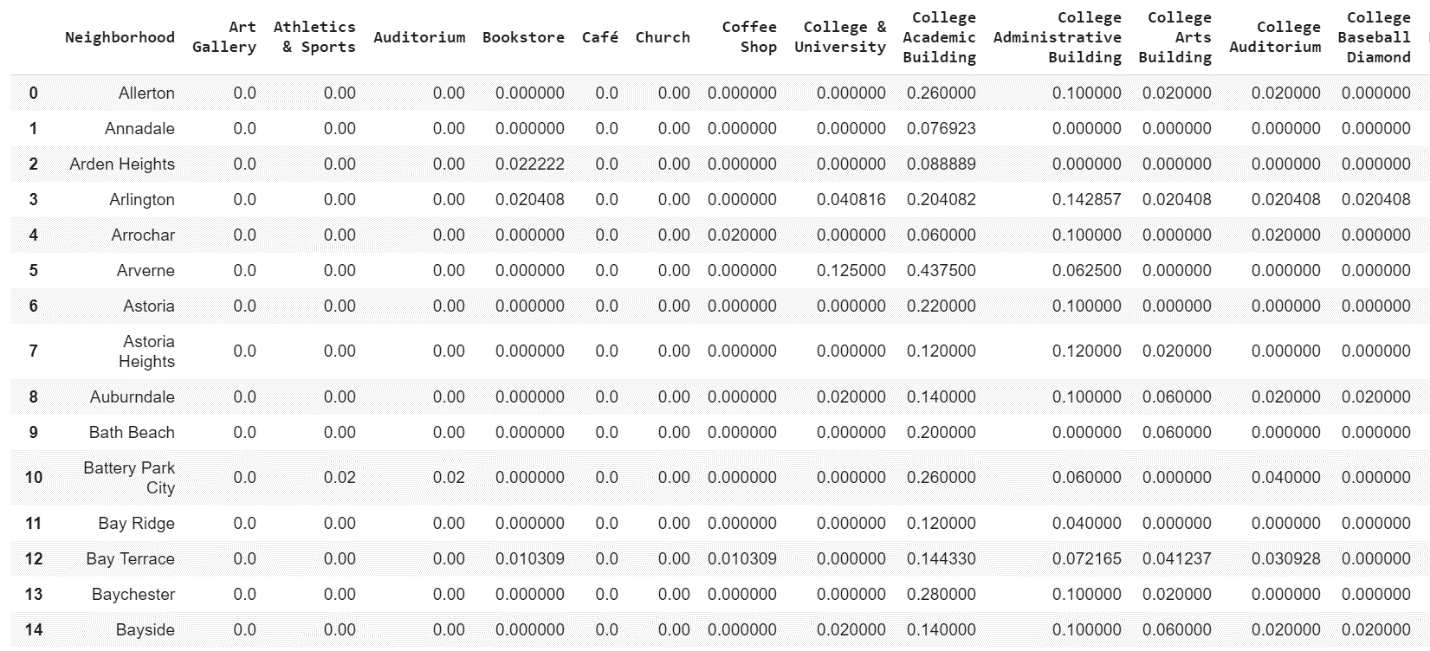


Figure 6: Dataframe with rows grouped by neighborhood.

We sorted the venues in descending order. Then we created a new dataframe displaying the top 10 venues for each neighborhood.

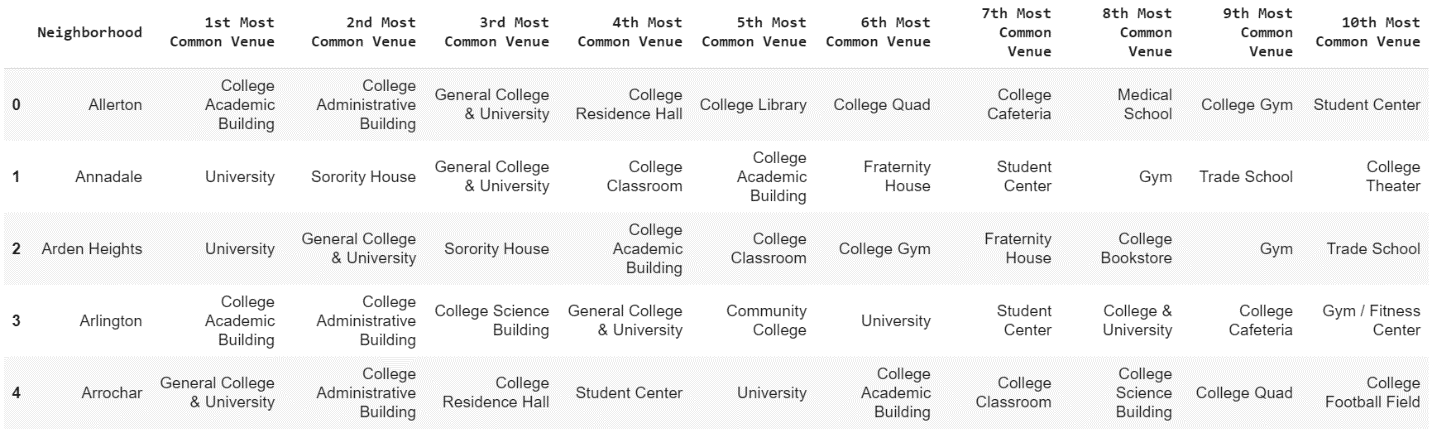
****

Figure 7: A new dataframe displaying the top 10 venues for each neighborhood.

**3.2 Clustering**

Finally, we clustered the venues to find the suitable locations for the food court chain. We used K-means clustering and the number of clusters is set to 10.

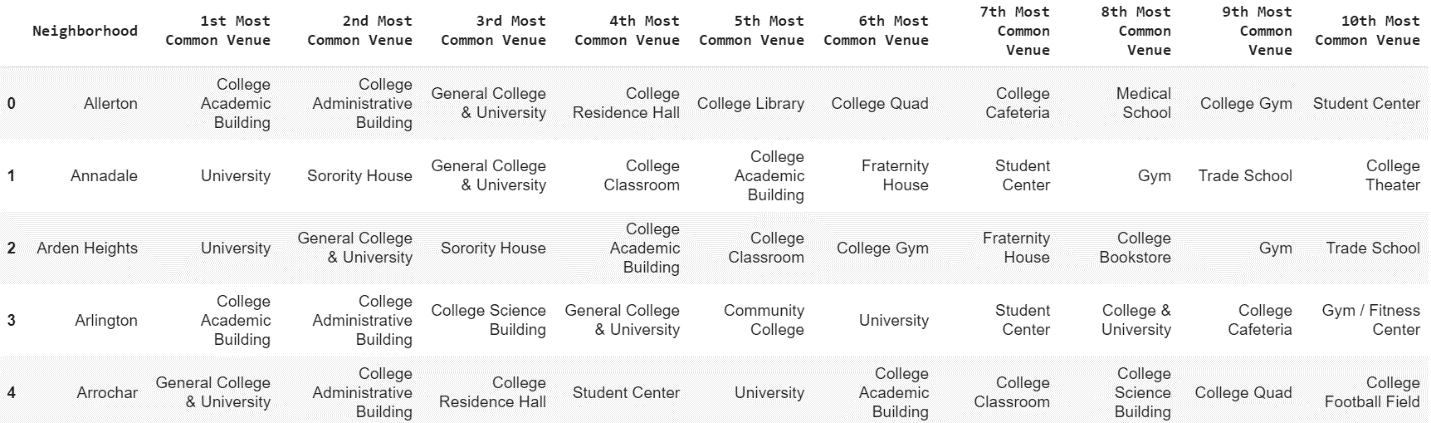


Figure 8: venues clustered to get the most suitable places for food courts.

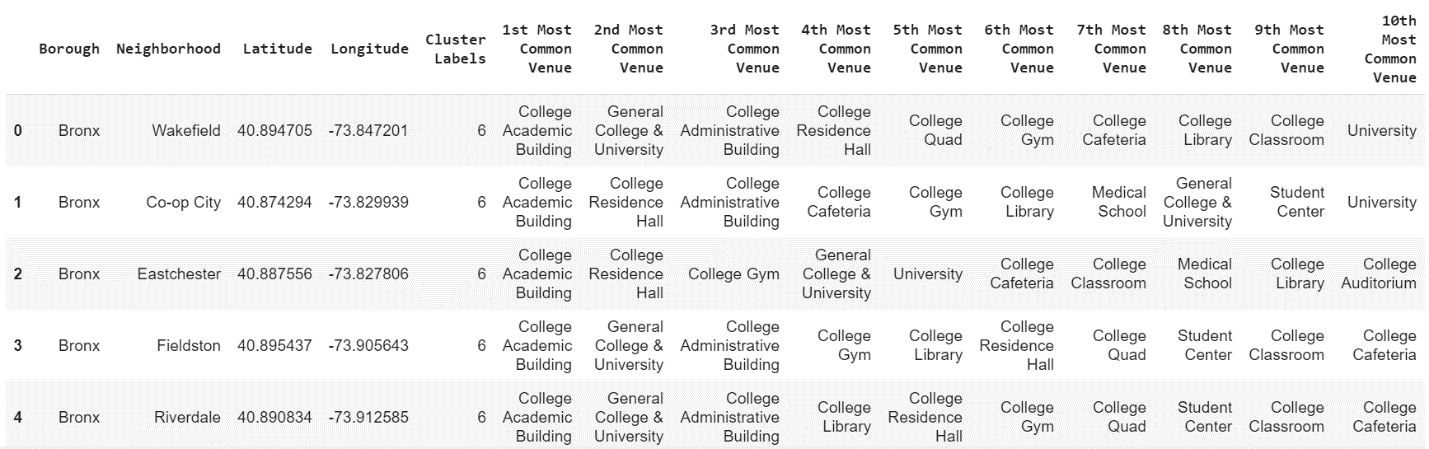


Figure 9: Clustered venues with the neighborhood they are located in.

**4. Results**

We finally visualize the resulting clusters by adding them to a map of New York City using folium. It is the final results that shows our recommended places to open food courts. Apparently, there are a lot of places suitable for opening food courts. So we ranked them by the density of venues shown in figure-9 and provided top 10 places best suited as a place to open food court chain.

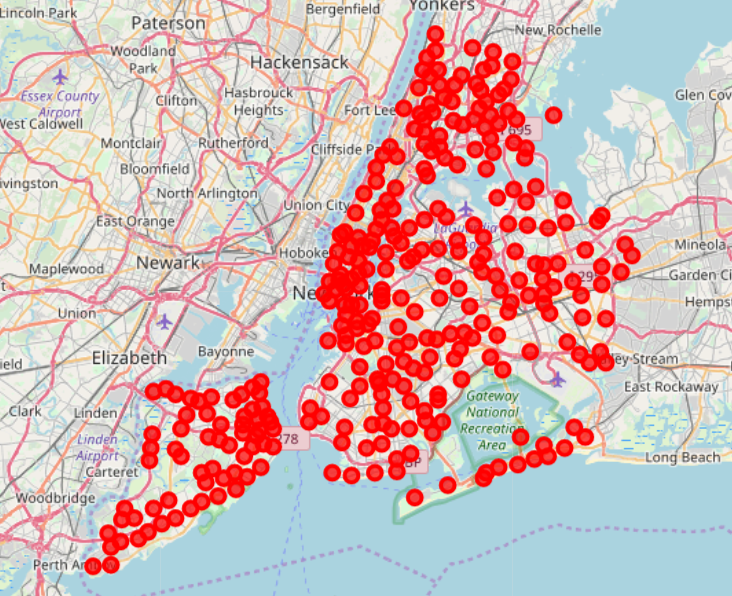
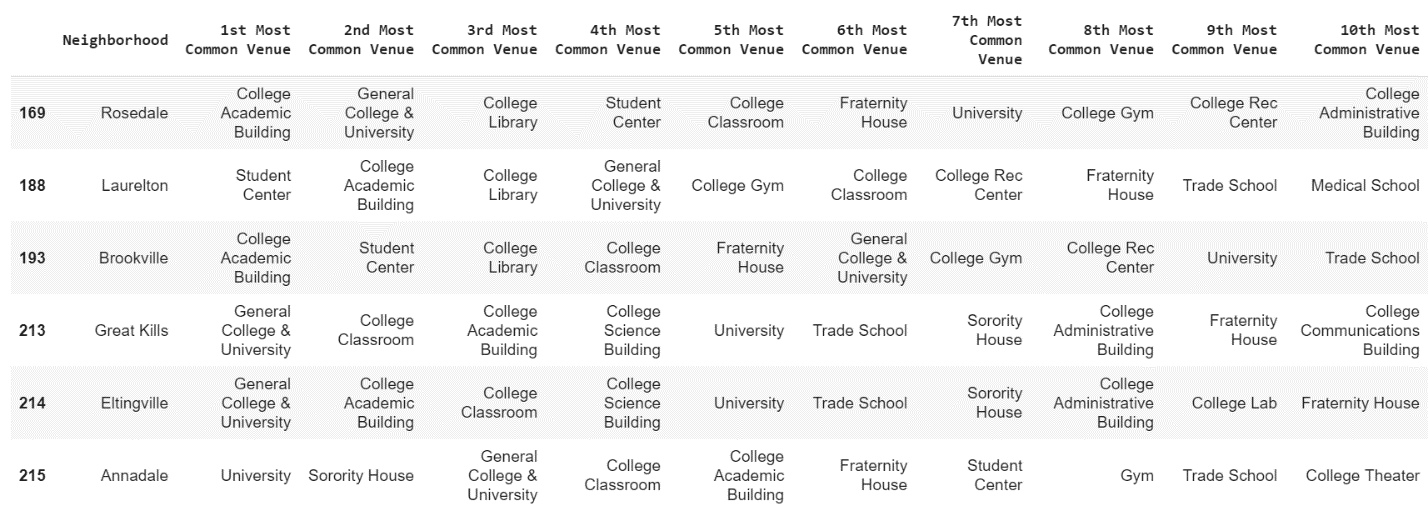


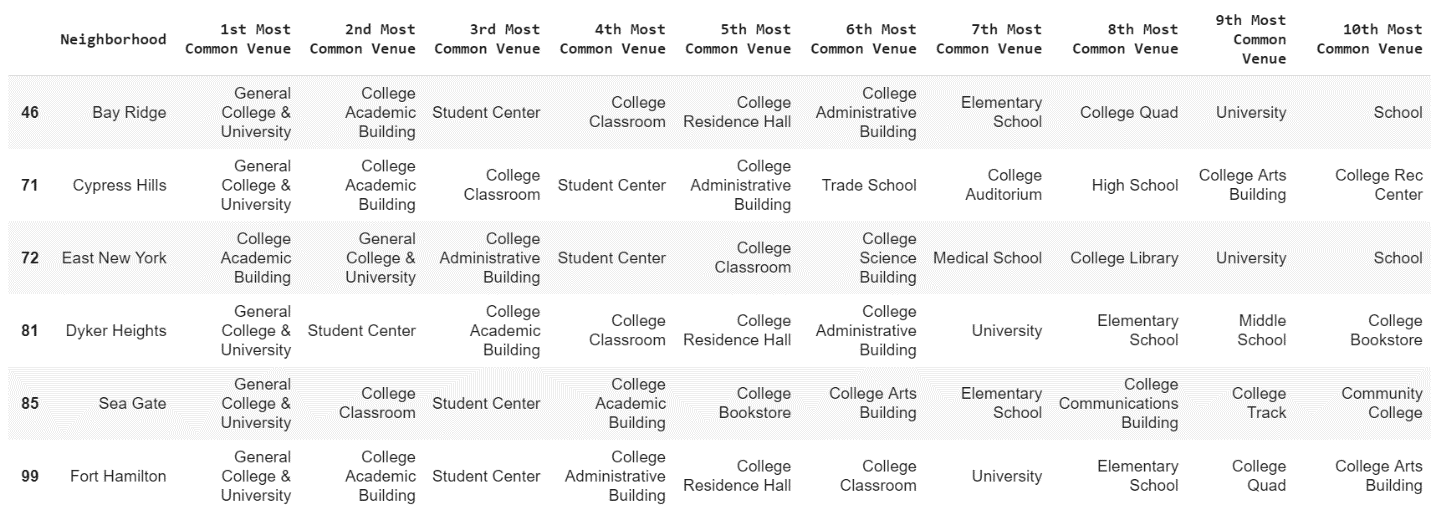
Figure 10: Recommended place to open a food court shown in map.

Here are the results of 10 clusters.

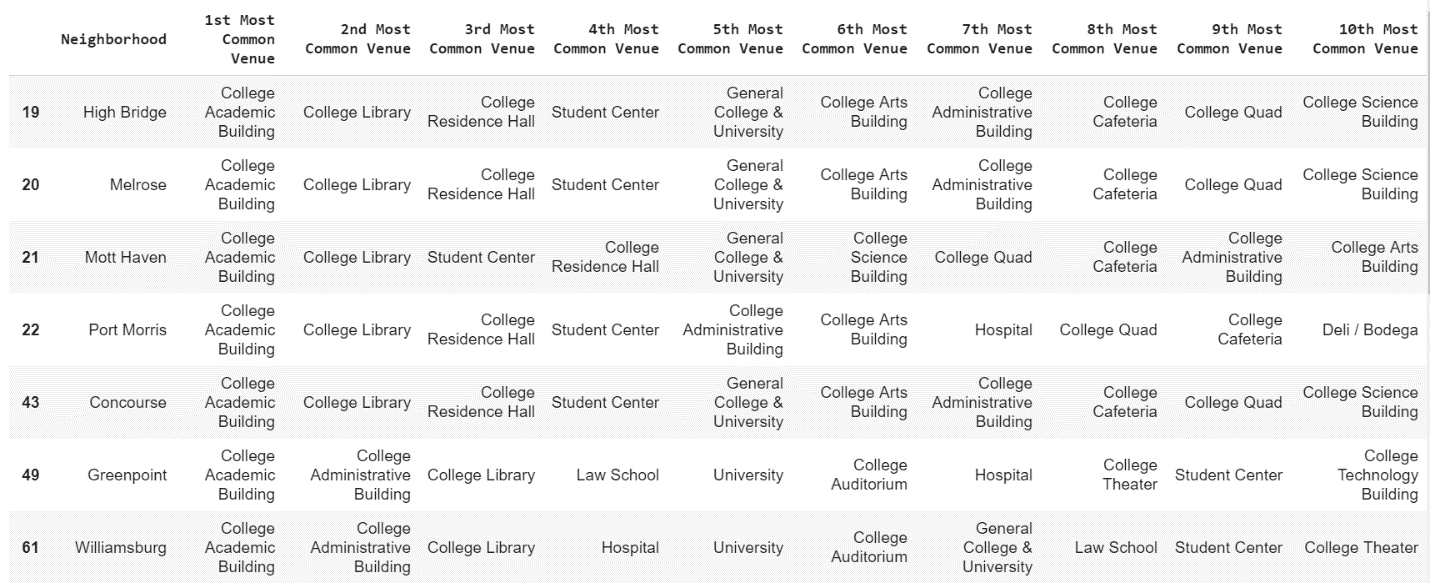
Cluster-1:



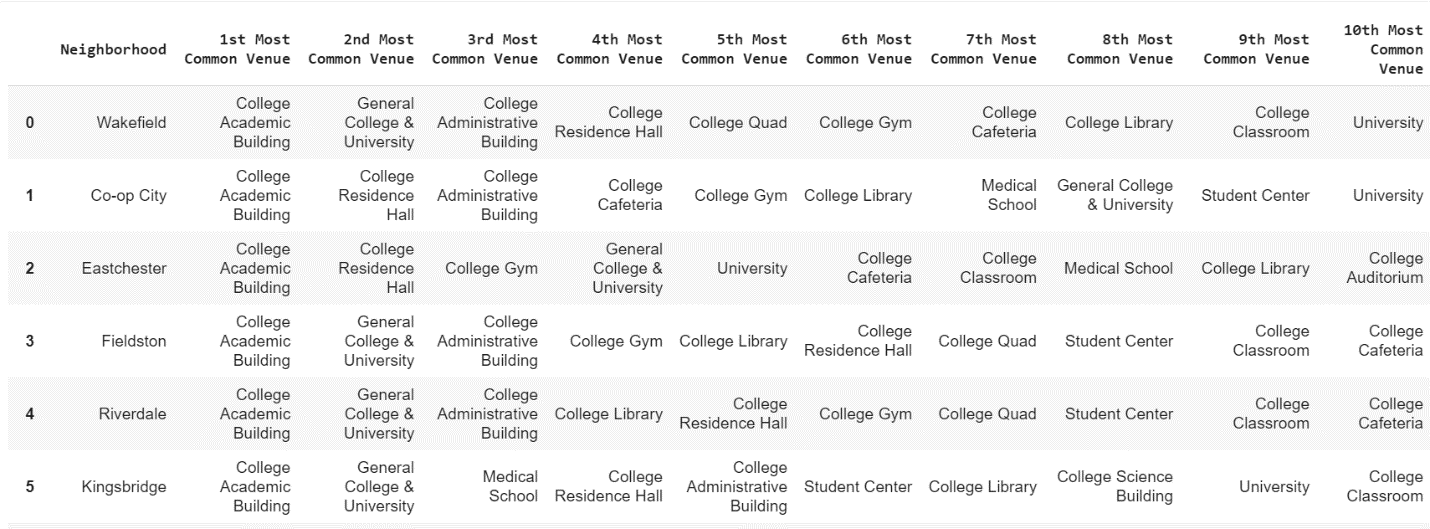
Cluster-2:

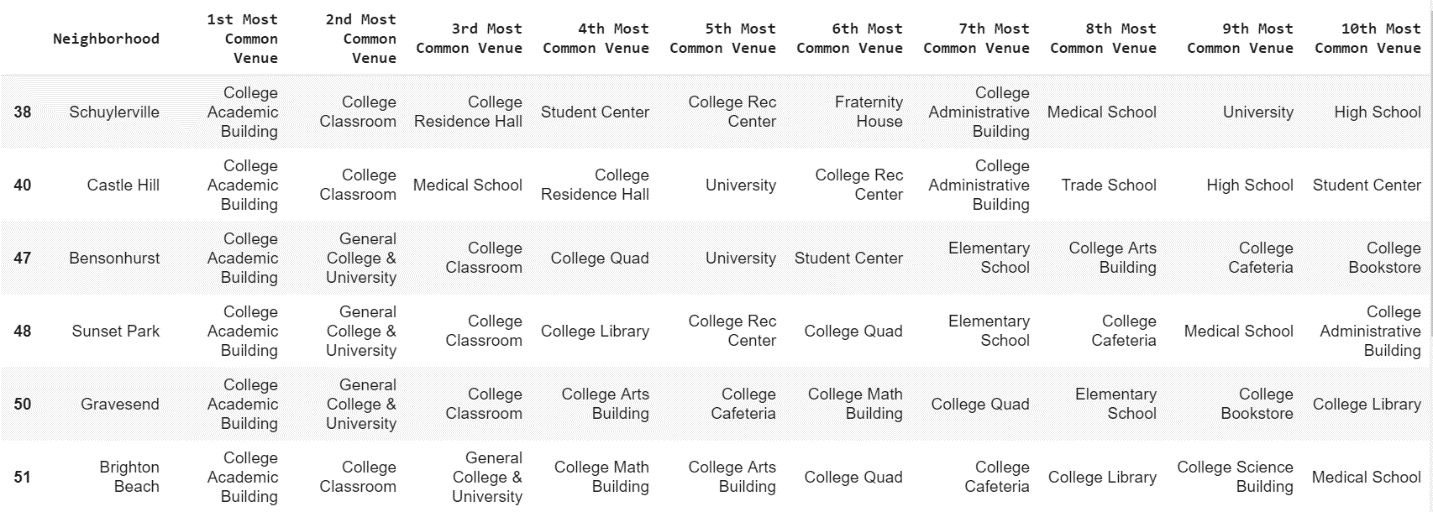


Cluster-3:

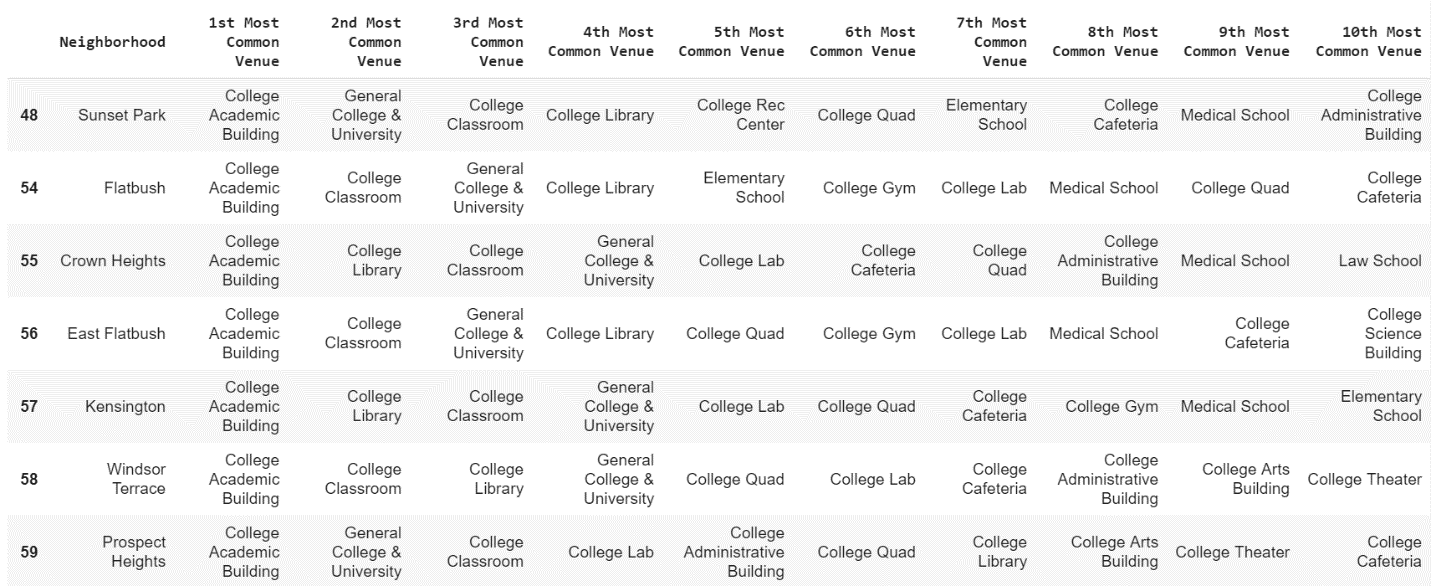


Cluster-4:

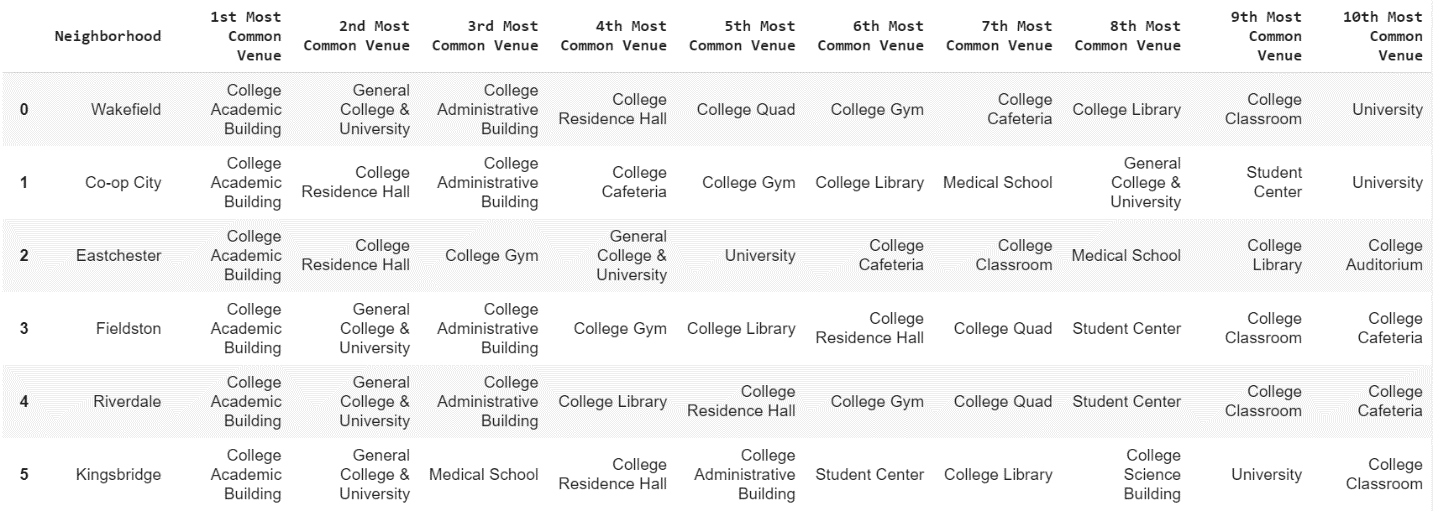


Cluster-5:

Cluster-6:



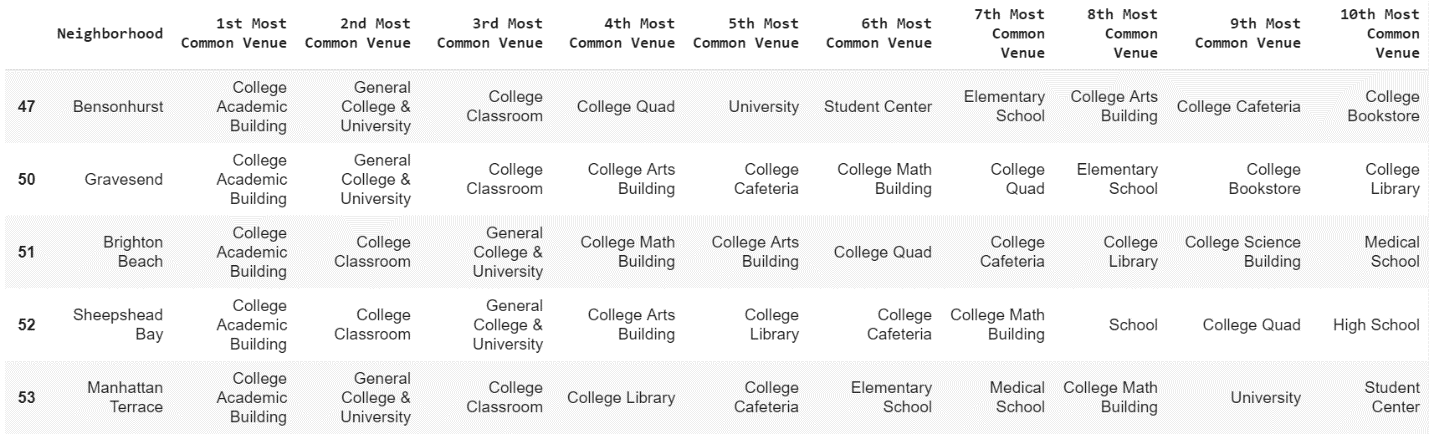
Cluster-7:



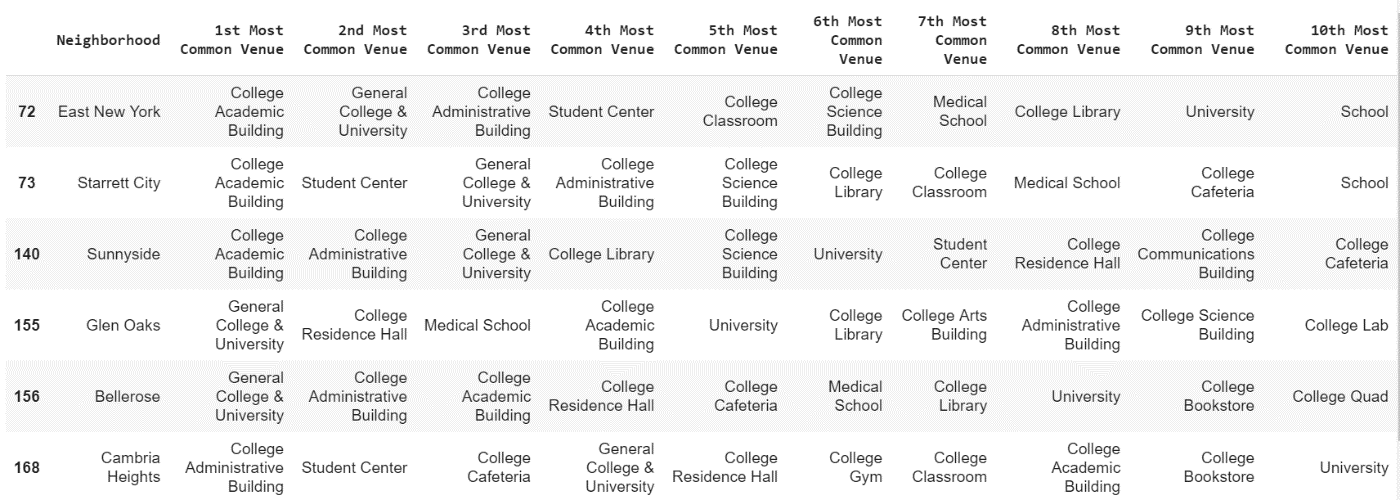
Cluster-8:



Cluster-9:



Cluster-10:



**5. Discussion**

Foursquare sandbox developer account is used to collect the venue data. So, the number of calls can be made per day is limited. With this limited resource we tried our best to make a practical recommendation. We see from the map that Manhattan area has the most dense suitable places to open a food court. Although the accuracy of our recommendation cannot be evaluated easily Any businessperson use these finding and use it with other practical observations to make better decisions.

**6. Conclusion**

This recommendation system can be enhanced in many ways. We can use the approximate number of students in a given venue to make better decision. Also, as food courts are mobile shops, we can move them from time to time to better exploit the recommendation. Obviously, these will require additional timeframe attribute to figure out places are most occupied by students between which time. This will increase our recommendation result more viable and robust.