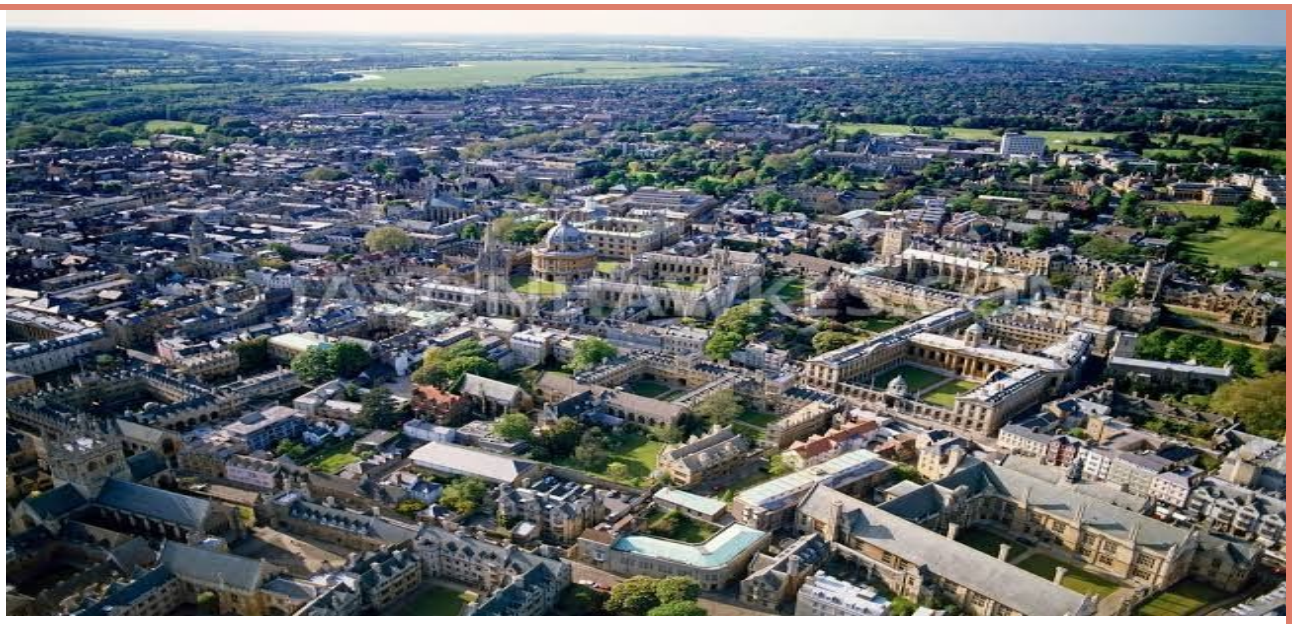

Recommending Business ventures that are Most Lucrative Around a College or University Campus in the United States

08.12.2020

Gospel Saloka-odenma
IBM Data Science Professional

Introduction

Sometimes intending business owners may have the capital but still do not know what kind of business to set up in a location that would yield good profit for their investment. A friend of mine had a land close to a university campus, and he was looking for what kind of building or business that would be best suited around a university campus that would be lucrative and worth his investment.

What is the viability of a campus environment's business investment? The United States alone has over 230 million people. According to the United States Department of Education[1], there are over 124,000 public and private schools in the United States; over 2,000 postsecondary non-degree career and technical schools; and over 4,000 degree-granting institutions of higher education. Of the higher education institutions, over 1,600 award associate degrees and some 2,400 award bachelor's or higher degrees. Over 400 higher education institutions award research doctorates. As an investor, consider having a few businesses around a few of these institutions, wouldn't that be worth your while? Well, that is what this work is about.

Aim:

This work seeks to recommend lucrative businesses one could establish around a college or university campus in the United States.

Objective:

1. What kind of businesses are located around a university campus
2. How do we know the most available businesses
3. How do we cluster them
4. What businesses can we recommend

Data

- Data of a few United States university campuses were obtained from data.world[2] in geojson format. The dataset had the geographical coordinates of each campus nested within the geojson file. The geojson file was cleaned and formatted into a dataframe.

- The **Foursquare API**[3] was used to get the common venues around the universities within 1000 mile radius
- Folium package was used to create maps and place markers on leaflet maps within each campus latitude and longitude location.

Methodology

United states college and university campuses data as obtained from data.world[2] was loaded into a jupyter notebook in cognitive lab server[4], where the data was cleaned and analysed. The resultant dataframe after data cleaning is shown in table 1.

Table 1: US college and university campus data

GID_id	Name	Web_url	Address	Shape_length	Shape_area	Latitude	Longitude
Univ_016	Dominican House of Studies	http://www.dhs.edu	487 MICHIGAN AVENUE NE	753.638132	31647.864270	-76.997774	38.929879
Univ_023	Stanford in Washington	https://siw.stanford.edu/	2661 CONNECTICUT AVENUE NW	105.168729	448.839365	-77.051823	38.925304
Univ_011	Trinity College	http://www.trinitydc.edu/	125 MICHIGAN AVENUE NE	1573.000495	139712.005500	-77.001796	38.928818
Univ_004	George Washington University at Mount Vernon C...	https://www.gwu.edu/mount-vernon-campus	2100 FOXHALL ROAD NW	1439.241902	91182.284910	-77.088793	38.919120
Univ_010	Strayer University	http://www.strayer.edu/	1133 15TH STREET NW	397.966306	4375.199012	-77.013982	38.973461

Folium package was used to create a map with each college and university name and address superimposed on each geographical coordinate as shown in figure 1.

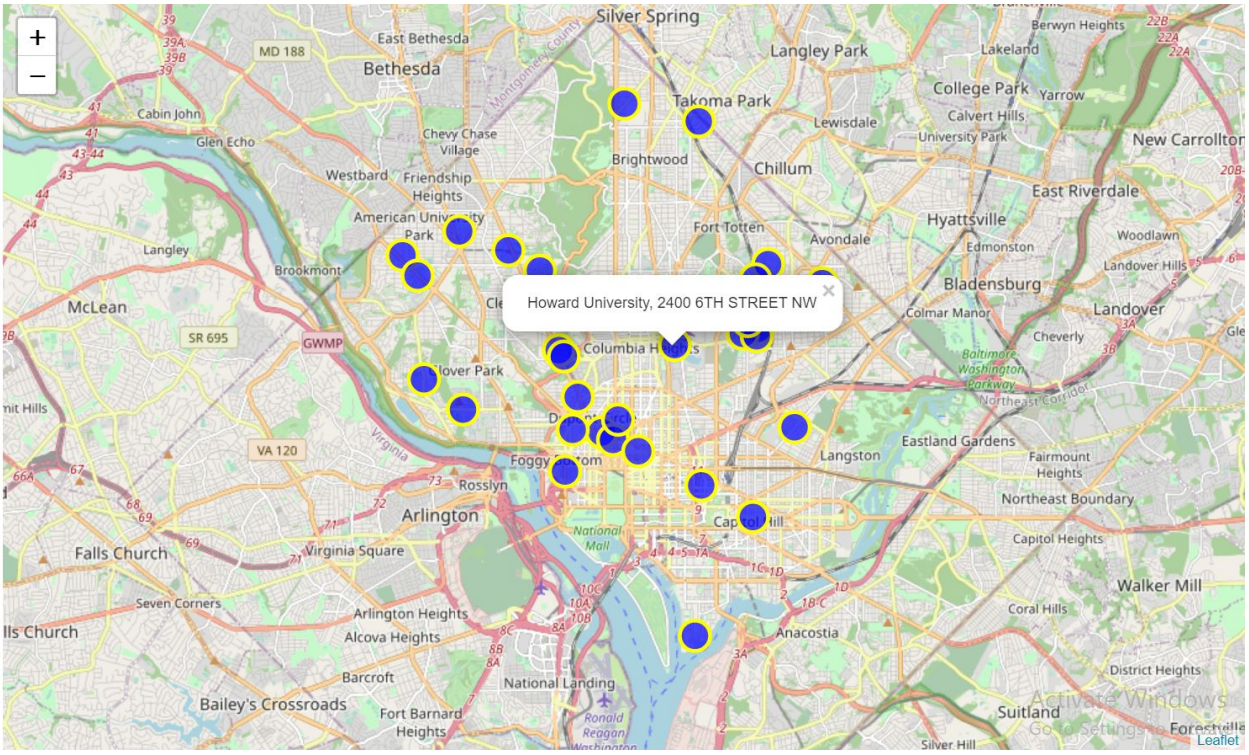

Figure 1: Map of US colleges and university campuses

Exploring venues:

The Foursquare API[3] was used to explore the venues within 1000 mile radius around these 30 campuses and segmenting the venues and their coordinates. A json file was returned by the Foursquare API[3] which was also cleaned, giving us the venues dataframe in table 2. The table has 1931 rows and 7 columns. Grouping the resultant venues dataframe gave 264 unique venue categories including dance studio, restaurant, pizza place, coffee shop etc.

Table 2: US colleges and universities common venues table

	UScampuses	Campus Latitude	Campus Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dominican House of Studies	38.929879	-76.997774	Dance Place	38.930370	-76.994780	Dance Studio
1	Dominican House of Studies	38.929879	-76.997774	Busboys and Poets	38.932117	-76.997640	American Restaurant
2	Dominican House of Studies	38.929879	-76.997774	&pizza	38.932582	-76.996696	Pizza Place
3	Dominican House of Studies	38.929879	-76.997774	Starbucks Reserve	38.932484	-76.997172	Coffee Shop
4	Dominican House of Studies	38.929879	-76.997774	Dew Drop Inn	38.925949	-76.994817	Bar

Analyses of venues Around each campus:

The top 5 venues around each campus were obtained and then all top 5 venues were put into a single dataframe and then sorted in descending order to obtain the top 10 common venues around all campuses. See table 3.

Table 3: Table showing top 10 venues around US college and university campuses

	Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ASU in Washington, DC	Coffee Shop	Park	Cocktail Bar	Hotel	Ice Cream Shop	Gym / Fitness Center	Bar	Taco Place	Bookstore	Pizza Place
1	American University	Coffee Shop	Bus Stop	Bank	Park	Italian Restaurant	Pharmacy	Shopping Mall	Student Center	Burger Joint	Liquor Store
2	American University Washington College of Law	Coffee Shop	Bank	Bakery	Bus Stop	Market	Tea Room	Tennis Court	Seafood Restaurant	Furniture / Home Store	Spa
3	Catholic University of America	Sandwich Place	Liquor Store	Coffee Shop	Bar	Art Gallery	Bookstore	Dance Studio	College Rec Center	Recreation Center	Rental Car Location
4	Cornell in Washington	Hotel	Park	Hotel Bar	Pizza Place	French Restaurant	Sandwich Place	Café	Gym / Fitness Center	Grocery Store	Gym

Results

Haven gotten the top 10 venues there is need to cluster these data. To get how many clusters that we need, the elbow method was used. The elbow method is a graph that suggests the number of possible clusters that should be created by looking at the point where the graph first shows an elbow sign. See figure 2.

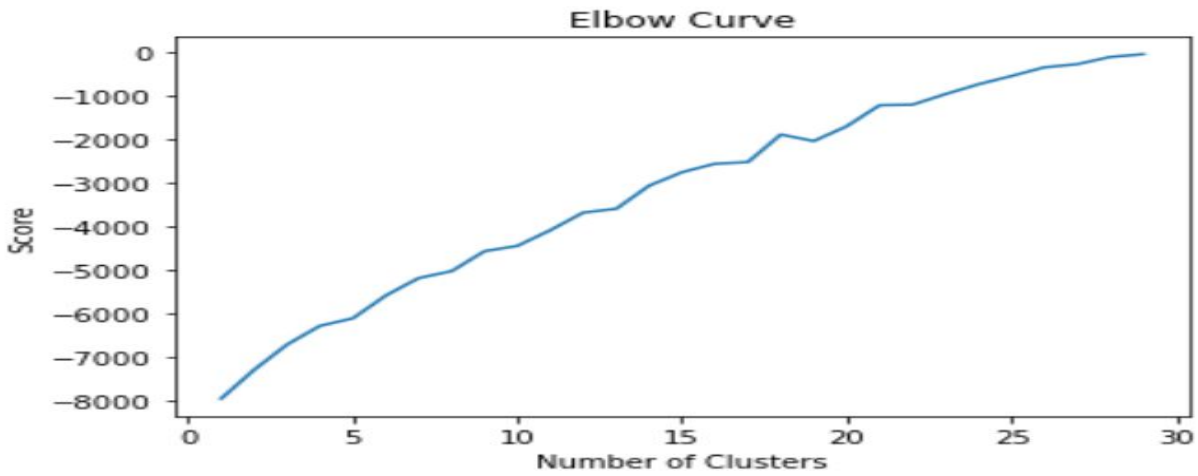

Figure 2: The elbow suggests 5 clusters

Five (5) clusters were generated. The principal components of two columns were also generated from converting the top 10 most common venues. See table 4 for the PCA table.

Table 4: PCA showing four categories of clusters and two PCAs

UScampuses	PC1	PC2	cluster
ASU in Washington, DC	6.569903	0.859129	0
American University	-3.491352	-0.988638	4
American University Washington College of Law	-2.839379	-1.216420	4
Catholic University of America	-5.738632	-1.807857	1
Cornell in Washington	6.454435	-2.484877	0

Then a scatter plot of the PCA was created to visualize the clusters. The plot showed an overlapping of the different clusters in some way. See figure 3.

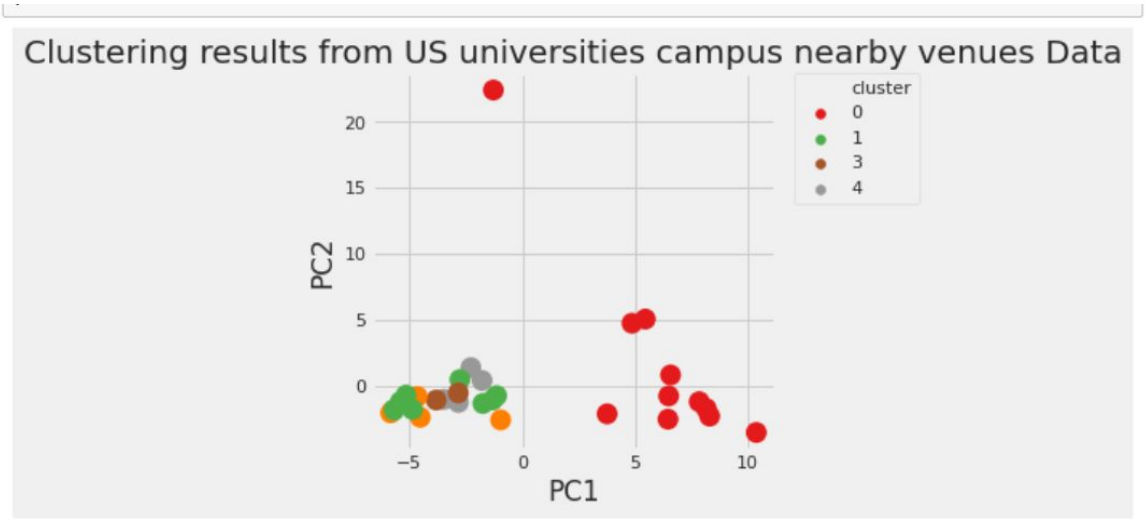

Figure 3: scatter plot of the PCA

Finally, wordclouds were used to show the most prominent venues of each cluster.

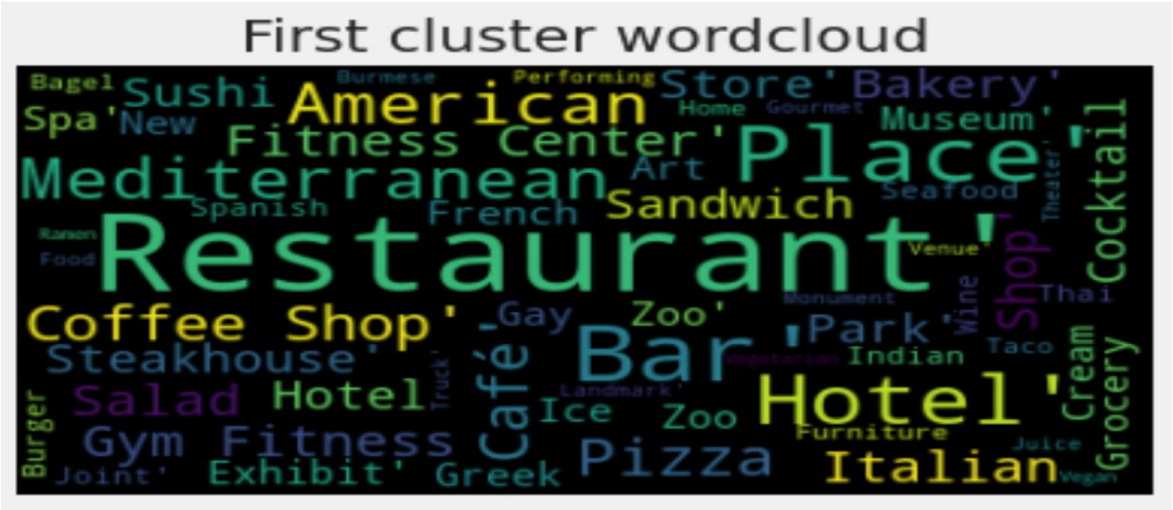

Figure 4: The first cluster

The first cluster shows that eateries, lodges, and relaxation spots are good business ventures because it shows Restaurants, Bar, Hotel, coffee shop and fitness center

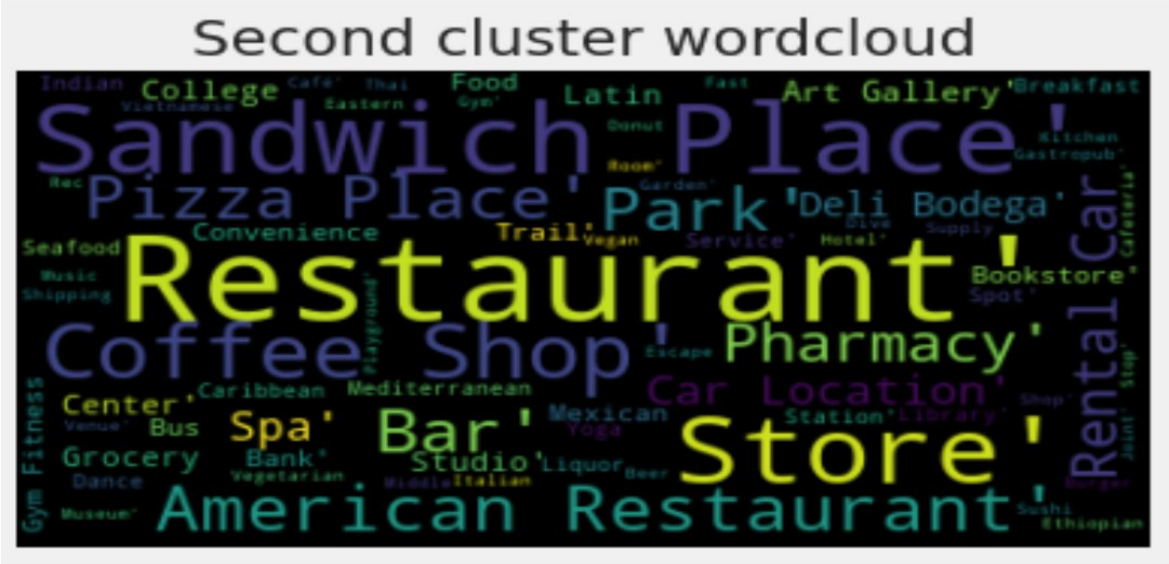

Figure 5: Second cluster

The second cluster shows that eateries, shopping centers, relaxation spots and medical facilities could be a good business because Restaurants, stores, bars, coffee shops and pharmacies are more prominent.

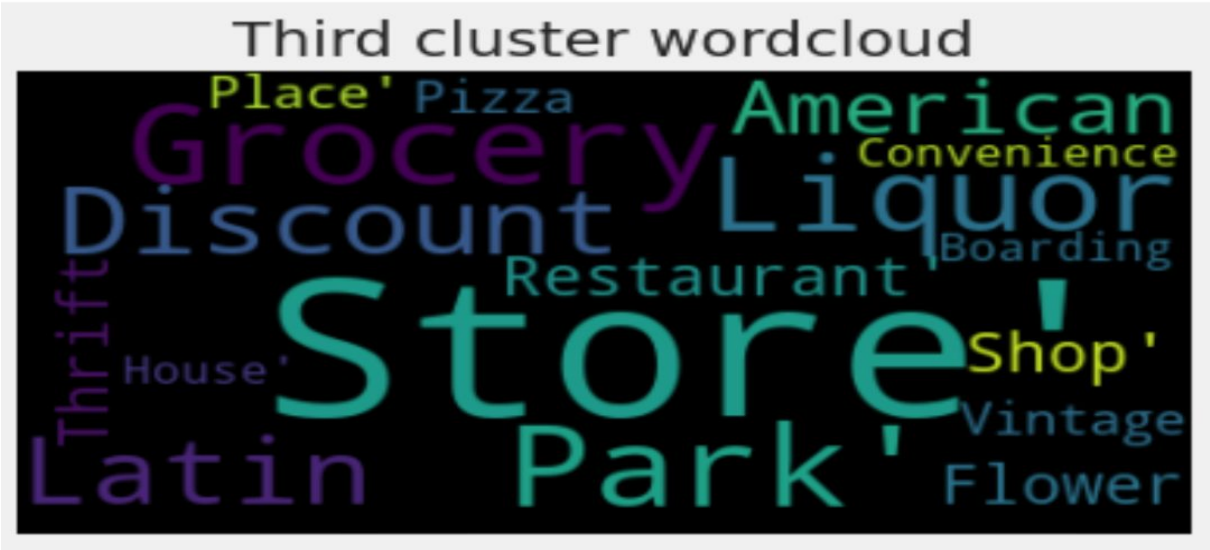

Figure 6: Third cluster

The third cluster shows shopping centers, relaxation spots and convenience facilities could be good business ventures because stores, parks, discount shops, convenience shops and restaurants appear more prominent.

Figure 6: Fourth cluster

The fourth cluster shows a shopping center, fitness center, relaxation spots and sporting facility could be a good business investment because it shows a store, gym, pizza, and sporting facilities.

Figure 6: Fifth cluster

The fifth cluster shows shopping centers, relaxation spots, financial institutions and educational facilities could be good businesses because it shows coffee, shops, banks, furniture shops, colleges and restaurants.

Let us find out why restaurants keep showing up in each cluster by looking at how they compare with one another. See table 5 and figure 7.

Table 5: displaying the top 20 most common venues

	Word	Count
22	restaurant'	58
11	shop'	26
10	'coffee	19
3	store'	19
4	'sandwich	16
31	'pizza	15
28	'park'	12
64	bar'	10
24	'hotel'	10
39	'american	9
9	'bar'	8
1	'pharmacy'	7
82	'food	7
58	'bank'	7
32	'café'	6

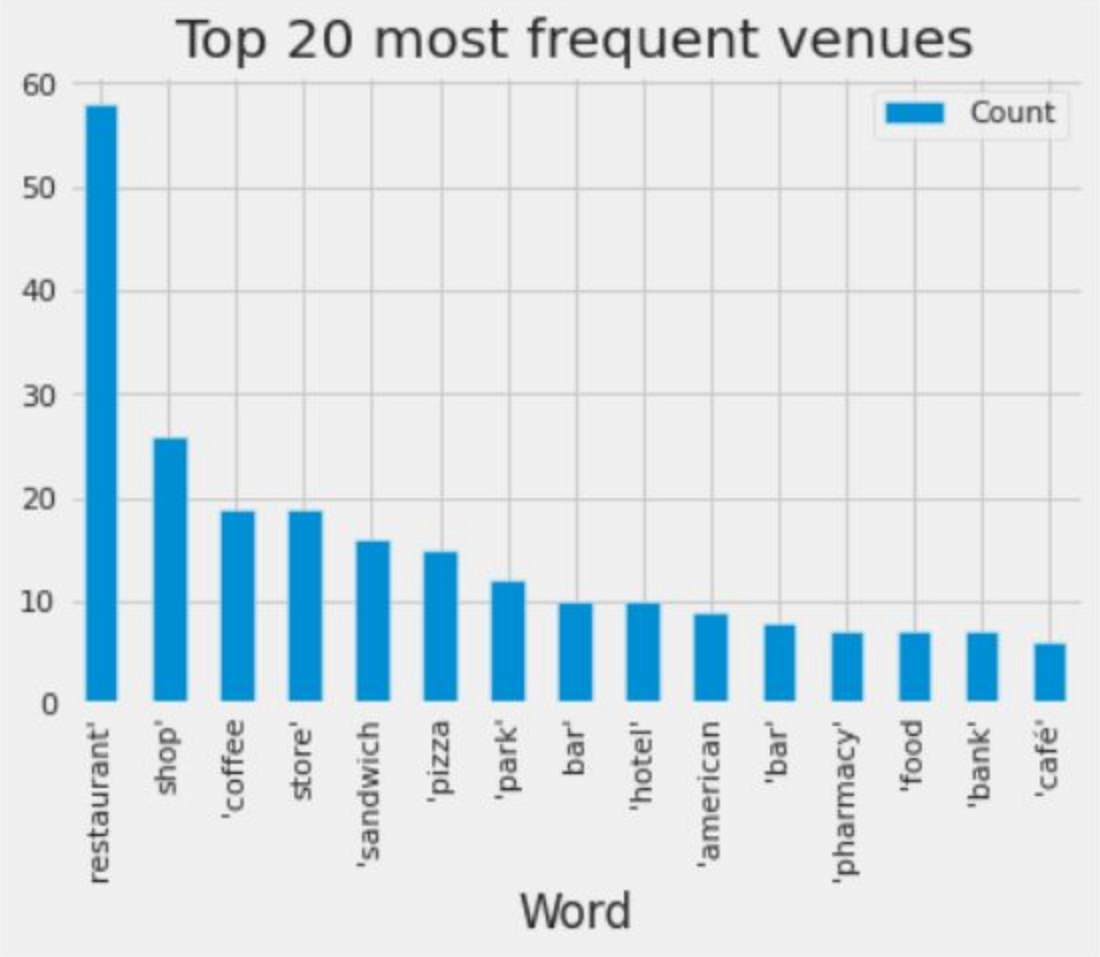

Figure 7: Bar chart of top 20 most frequent venues

Discussion

Results as seen from venues around United States colleges and University campuses shows that:

- Any kind of relaxation spot is a lucrative business around a university campus.
- The campuses are almost over loaded with restaurants which tells us that one is bound to face huge competitions. so you could decide to go for less competitive business venture like coffee shop
- Next to relaxation spots is eateries and shopping centers
- Some ventures that may not be common but not less lucrative could be financial institutions (bank), medical facilities (pharmacy) and fitness facilities (gym)

- Coffee shops appear to be an easy and less capital intensive business venture to invest in.
- Further analysis could be embarked on to have two words processing so we could have something like ‘coffee shops’ not just ‘coffee’ to see if one word processing had any negative effect on the results.

Conclusion

This work recommends what business venture is most suited for a United States college or university campus which can easily be repeated across any university campus across any country. It reiterates the power of analytics in helping business owners understand what business venture to go into depending on the geographical location. The interesting part is that it does not really have to be for multinational organisations. It is an easy and simple approach that small to mid-size enterprises can embrace.

References:

1. https://www2.ed.gov/about/offices/list/ous/international/usnei/us/edlite-institutions-us.html
2. https://data.world/codefordc/college-university-campuses
3. https://foursquare.com/
4. https://labs.cognitiveclass.ai/login?next=https%3A%2F%2Flabs.cognitiveclass.ai%2F