




# Clustering of Potential areas to invest on services for college students

Christian Vilca

# Selecting the appropriate location is fundamental when opening a branch office

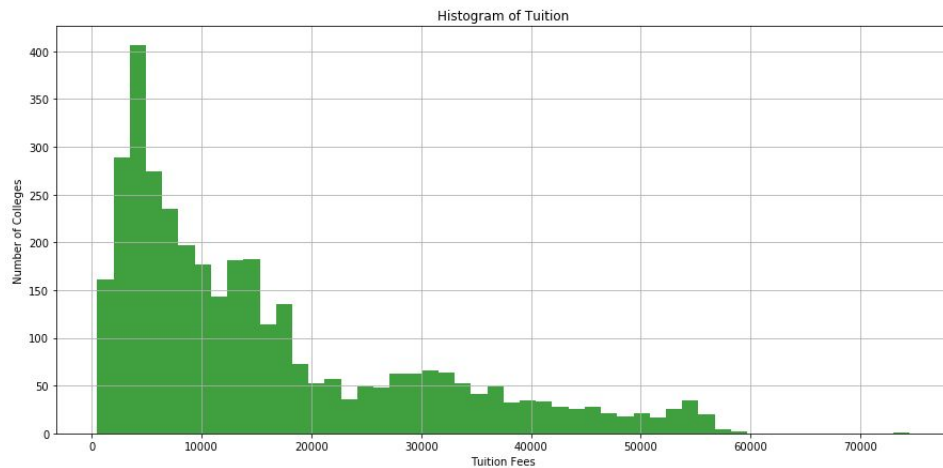
- In the specific case of an investor whose target customers are college students, there are three main characteristics for a location
    - 1.The quantity of students that surround the area, as possible customer.
    - 2.An approximation of the economic status of students to check if they would pay premium services or products, identify potential customers.
    - 3.The kind of venues that are most common around each location.
  - Students that are planning to make an exchange will find valuable to know comon venues surrounding each university as an additional characteristic.
- 

# Data acquisition and Cleaning

- Information and statistics about colleges (March 30, 2020) from College Scoreboard: <https://collegescorecard.ed.gov>
- After preprocessing, 3551 rows and 7 features in the data frame.
- “Tuition fee in” has more consistency than “tuition fee out”, so it was dropped
- Information about venues from FOURSQUARE API
- Analysis of 100 venues in a radius of 1000 meters surrounding each college.
- Time estimated to walk 1000 meters = 12 min

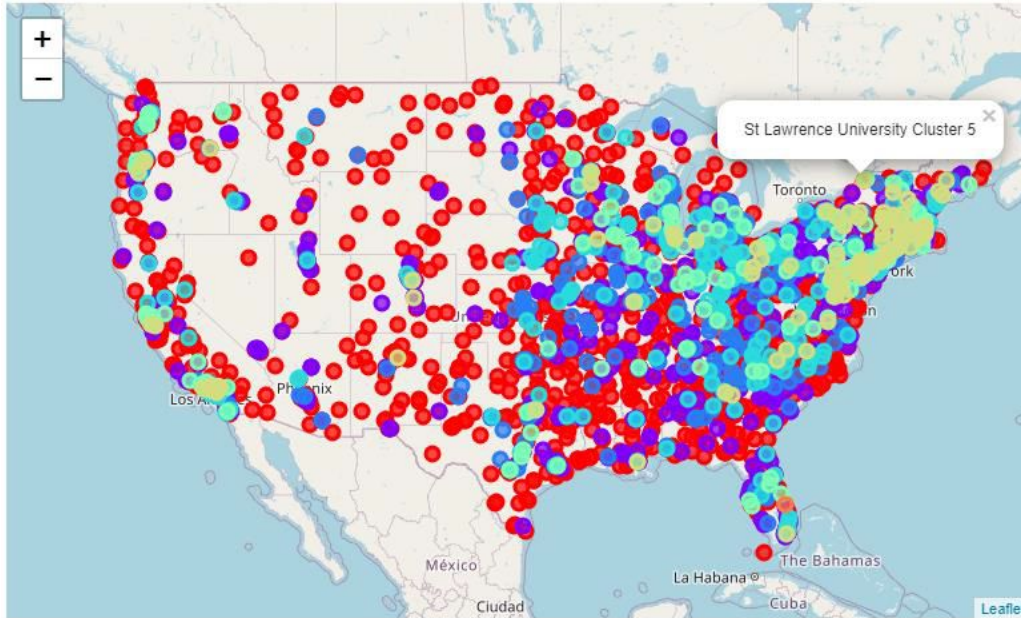


# Clustering per Tuition Fee



To approximate the socioeconomic status of students, I used the tuition fee per university to set ranges (USD 10,000) and segmented the data.

# Comparing Tuition Fees and household income.



The colleges with higher tuition fees are located in states that have the higher household income:

- Boston
- Philadelphia
- New York
- Los Angeles
- San Francisco

# Using Foursquare API

	Univ. 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	Arizona College-Las Vegas	Sandwich Place	Pool	Casino	Mexican Restaurant	Seafood Restaurant	Pizza Place	Pub	Bridal Shop	Buffet	Café
1	Arizona College-Mesa	Convenience Store	Mexican Restaurant	Grocery Store	Gas Station	Donut Shop	Bar	Coffee Shop	Sandwich Place	Electronics Store	Asian Restaurant
2	Avila University	College Gym	Sandwich Place	Park	Baseball Field	Farm	Falafel Restaurant	Eye Doctor	Event Space	Doctor's Office	Electronics Store
3	California Jazz Conservatory	Coffee Shop	Theater	Yoga Studio	Sushi Restaurant	Brazilian Restaurant	Asian Restaurant	Music Venue	American Restaurant	Café	Sandwich Place
4	Calumet College of Saint Joseph	Bar	Mexican Restaurant	Dive Bar	Pizza Place	Bus Station	Eye Doctor	Dry Cleaner	Eastern European Restaurant	Electronics Store	Event Space

Got the ten most common venues that surround each college.

Each cluster of tuition fee (6) is analyzed

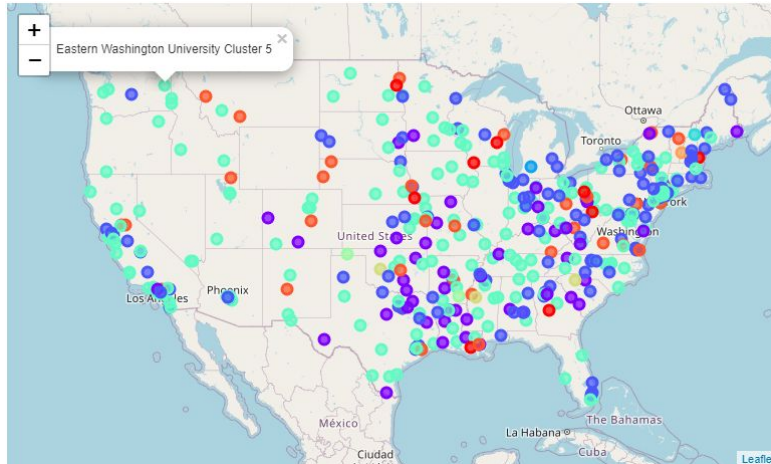


# k-means

- To find the similarity among colleges by the venues that are close to them
- Clustering is repeated to each range (6) of tuition fee
- $k=10$

# Kmeans - Tuition Fee (USD 0 - 10,000)

- Categorized mainly in 5 clusters.
- In cluster 1 and 2, there are 220 colleges surrounded mainly by Pizza Places and Fast Food Restaurants, while in cluster 0,5,9 are 234 colleges surrounded mainly by Coffee Shops, bar and Parks.

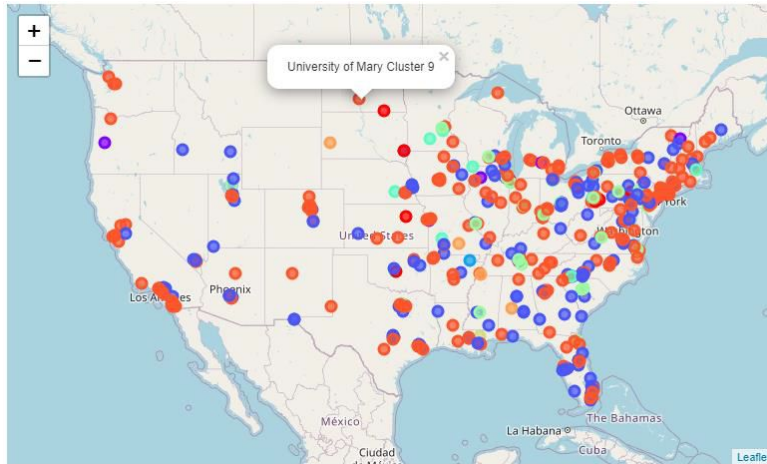


	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	7309.9	7975.7	8250.6	7300.0	7970.0	8271.5	7696.0	7147.7	7298.0	8043.9
Avg. Underg. Students	4005	7604	4092	534	1527	12070	2599	2936	1659	5017
Colleges Clustered	11	62	158	1	1	214	2	3	1	41
1 <sup>st</sup> Common Venue	0.64 Bar	0.64 Fast Food Rest.	0.22 Pizza Place	Athletics & Sports	Restaur.	0.18 Coffee Shop	Football Stadium	Discount Store	New American Rest.	0.31 Park
2 <sup>nd</sup> Common Venue	0.09 IT Service	0.06 Bakery	0.16 Fast Food Rest.			0.11 Sandwich Place		Convenie. Store		0.09 Trail
3 <sup>rd</sup> Common Venue	0.09 Diner	0.03 Mexican Rest.	0.04 Grocery Store			0.10 Bar				0.07 Golf Course
4 <sup>th</sup> Common Venue	0.09 Café	0.03 Discount Store	0.04 Hotel			0.10 Pizza Place				0.07 Café
5 <sup>th</sup> Common Venue	0.09 Pizza Place	0.03 Pizza Place	0.04 Discount Store			0.05 Hotel				0.05 Grocery Store



# Kmeans - Tuition Fee (USD 10,000 - 20,000)

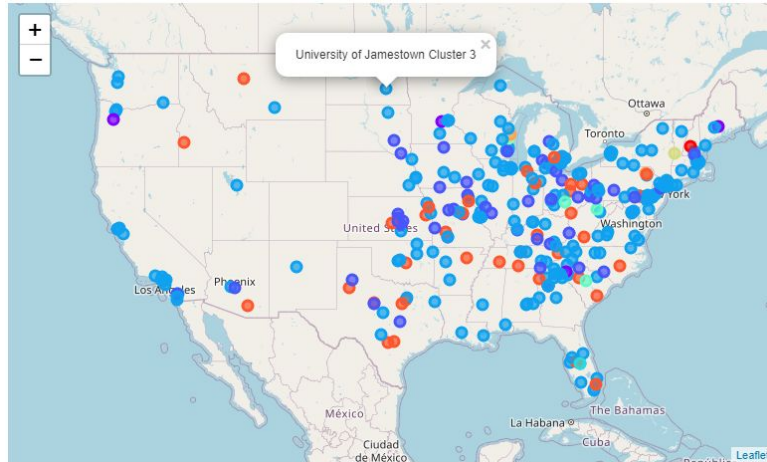
- Categorized mainly in 5 clusters.
- Cluster 0,1,5 and 6 gather venues related to social places like bar, hotel and parks, while the cluster 2 gather venues related to Fast Food Restaurant.



	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	16,942.5	17,375.7	16,602.8	17,320.0	17,665.0	16,906.6	16,173.1	18,230.0	16,710.8	16,391.4
Avg. Underg. Students	486	504	2069	539	157	2357	1957	149	2628	3595
Colleges Clustered	10	7	154	1	1	10	41	1	4	268
1 <sup>st</sup> Common Venue	0.7 Bar	0.29 Park	0.29 Fast Food Rest.	Baseball Field	Intersect.	0.2 Park	0.83 Hotel	Bookstore	0.25 Jewelry Store	0.09 Pizza Place
2 <sup>nd</sup> Common Venue	0.1 General Entertainment	0.14 Construc. & Landsc.	0.10 Conveni. Store			0.2 Construc. & Landsc.	0.05 Fast Food Rest.		0.25 American Rest.	0.09 Clothing Store
3 <sup>rd</sup> Common Venue	0.1 Grocery Store	0.14 Trail	0.10 Pizza Place			0.2 History Museum	0.02 College Cafeteria		0.25 Pizza Place	0.09 Coffee Shop
4 <sup>th</sup> Common Venue	0.1 Italian Rest.	0.14 Gym	0.04 Grocery Store			0.1 Scenic Lookout	0.02 American Rest.		0.25 Airport Terminal	0.07 Sandwich Place
5 <sup>th</sup> Common Venue		0.14 Café	0.04 Pharmacy			0.1 Golf Course	0.02 Café			0.06 Hotel

# Kmeans - Tuition Fee (USD 20,000 - 30,000)

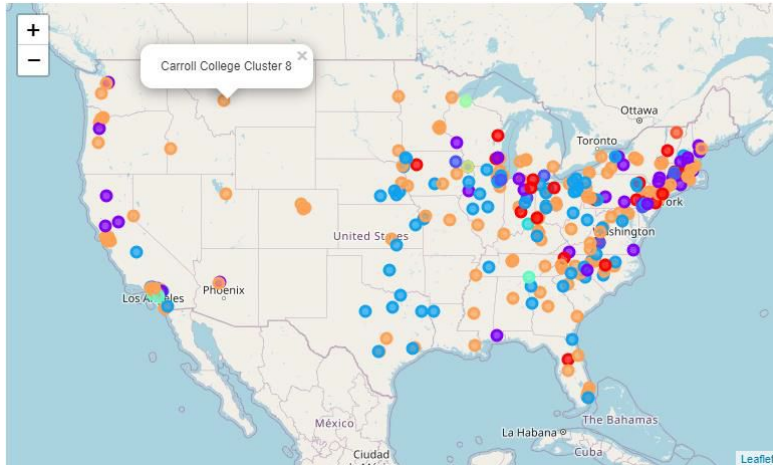
- Categorized mainly in 4 clusters.
- Cluster 1 and 3 gather 70% of all colleges and are mainly surrounded by social places like Coffee Shops, while the cluster 2 and 9 gather venues related to Fast Food Restaurants and Pizza Places.



	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	24,000.0	26,284.4	25,662.0	25,097.0	21,694.0	24,135.3	28,870.0	25,511.0	27,555.0	25,765.4
Avg. Underg. Students	71	804	1663	1952	1123	2167	2900	404	1048	2913
Colleges Clustered	1	5	52	240	1	3	1	1	2	39
1 <sup>st</sup> Common Venue	Lounge	0.6 Coffee Shop	0.33 Pizza Place	0.09 Coffee Shop	Lake	IT Services	Pub	Art Gallery	College Residence Hall	0.61 Fast Food Rest.
2 <sup>nd</sup> Common Venue		0.2 Disc Golf	0.08 Grocery Store	0.08 Pizza Place		Football Stadium			Italian Rest.	0.05 Conveni. Store
3 <sup>rd</sup> Common Venue		0.2 Brewery	0.08 American Rest.	0.07 Hotel		Convenie. Store				0.05 Baseball Field
4 <sup>th</sup> Common Venue			0.06 Gym	0.06 Park						0.05 Discount Store
5 <sup>th</sup> Common Venue			0.06 Mexican Rest.	0.05 Italian Rest.						0.03 Mexican Rest.

# Kmeans - Tuition Fee (USD 30,000 - 40,000)

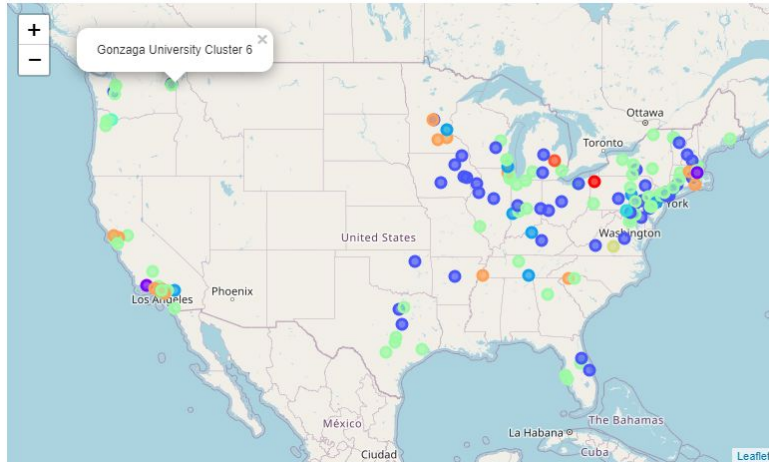
- Categorized mainly in 5 clusters.
- Cluster 0,1 and 8 gather most of students and their colleges are mainly surrounded by social places like Bar, golf place and most important Coffee Shops, while the cluster 2 and 3 gather venues related to Fast Food Restaurants and Pizza Places.



	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	34,274.3	34,220.2	34,167.4	33,525.1	37,670.0	34,518.3	36,183.0	33,350.0	34,782.7	38,300.0
Avg. Underg. Students	2090	4876	2642	2281	1113	1164	685	817	2454	234
Colleges Clustered	23	38	8	78	1	6	1	1	179	1
1 <sup>st</sup> Common Venue	0.39 Bar	0.18 Golf Course	0.75 Pizza Place	0.22 Pizza Place	Bus Station	Trail	Gym Fitness	Dog Run	0.17 Coffee Shop	Bookstore
2 <sup>nd</sup> Common Venue	0.09 Baseball Field	0.08 Ice Cream Shop	0.13 College Cafeteria	0.19 Fast Food Rest.					0.08 Hotel	
3 <sup>rd</sup> Common Venue	0.09 American Rest.	0.08 Baseball Field	0.12 Bagel Shop	0.06 Sandwich Place					0.07 American Rest.	
4 <sup>th</sup> Common Venue	0.04 Athletics & Sports	0.08 Café		0.06 Bank					0.05 Bar	
5 <sup>th</sup> Common Venue	0.04 Diner	0.05 Pub		0.06 Discount Store					0.04 Pizza Place	

# Kmeans - Tuition Fee (USD 40,000 - 50,000)

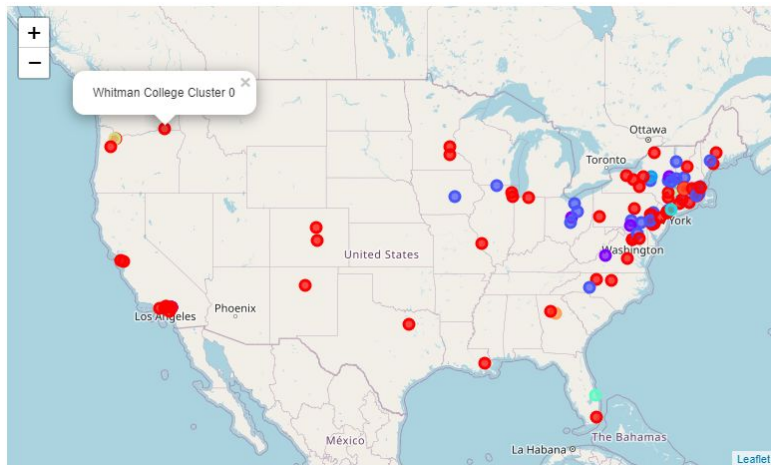
- Categorized mainly in 4 clusters.
- Cluster 3,6,8 and 9 gather most of students, and their colleges are mainly surrounded by social places like Coffee Shop, Parks and Bar, while the cluster 2 gather Pizza Places.



	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	47,540.0	42,687.0	44,136.5	45,559.4	41,350.0	45,904.0	44,732.6	45,746.0	44,382.3	43,490.0
Avg. Underg. Students	1859	2191	3446	2976	1961	4241	3307	1061	2483	2139
Colleges Clustered	1	2	45	9	1	1	94	1	15	1
1 <sup>st</sup> Common Venue	Pharmacy	Breakfast Spot	0.53 Pizza Place	0.33 Coffee Shop	College Basketball Court	Cruise	0.16 Coffee Shop	Food Court	0.27 Park	Bar
2 <sup>nd</sup> Common Venue		Baseball Field	0.11 Bar	0.33 Baseball Field			0.12 American Rest.		0.13 Trail	
3 <sup>rd</sup> Common Venue			0.06 Fast Food Rest.	0.22 Park			0.07 Bar		0.07 Food Court	
4 <sup>th</sup> Common Venue			0.04 American Rest.	0.11 College Theater			0.06 Café		0.07 Beach	
5 <sup>th</sup> Common Venue			0.44 Sandwich Place				0.06 Italian Rest.		0.07 Pub	

# Kmeans - Tuition Fee (USD 50,000 - Max)

- Categorized mainly in 3 clusters.
- Cluster 0 gather most of students, and their colleges are mainly surrounded by social places like Coffee Shop, Parks and Bar, while the cluster 1 and 2 gather American Restaurant and Pizza Places.



	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Avg. Tuition Fee (USD)	53,755.8	54,760.0	53,781.8	54,620.0	50,175.0	74,514.0	55,082.0	50,934.0	51,306.0	51,668.0
Avg. Underg. Students	6495	1750	3210	2039	98	432	432	2240	1002	1304
Colleges Clustered	71	9	28	1	1	1	1	1	1	1
1 <sup>st</sup> Common Venue	0.25 Coffee Shop	0.56 American Rest.	0.39 Pizza Place	Pub	Trail	Airport	Soccer Field	ATM	Trail	Bus Stop
2 <sup>nd</sup> Common Venue	0.08 Bar	0.11 Coffee Shop	0.11 Italian Rest.							
3 <sup>rd</sup> Common Venue	0.07 Park	0.11 Hotel	0.07 Pub							
4 <sup>th</sup> Common Venue	0.07 American Rest.	0.11 College Theater	0.07 American Rest.							
5 <sup>th</sup> Common Venue	0.04 Café	0.11 Theater	0.07 Bar							



# Distribution of social and food places

- Clusters can be categorized as food places and social places.
- Social Places: most common are Coffee Shops, parks and bars.
- Food Places: most common are Pizza places and Fast Food Restaurants.
- The categorizing of venues around colleges shows how much differ the preference of students and gives an insight to an investor the kind of venues that are more profitable per range of tuition fee.

Range Tuition Fee	Social Places		Food Places		Mixed	
	# Cluster	Proportion	# Cluster	Proportion	# Cluster	Proportion
0 – 10,000	0,3,5,9	55 %	1,2,4,8	45 %	6,7	0 %
10,000 – 20,000	0,1,5,6	14 %	2	31 %	3,4,7,8,9	55 %
20,000 – 30,000	1,6,7,3	70 %	2,9	26 %	0,4,5,8	4 %
30,000 – 40,000	0,1,8	71 %	2,3	26 %	4,5,6,7,9	3 %
40,000 – 50,000	3,6,8,9	70 %	2,7	27 %	0,1,4,5	3 %
50,000 - Max	0,3	63 %	1,2	32 %	4,5,6,7,8,9	5 %