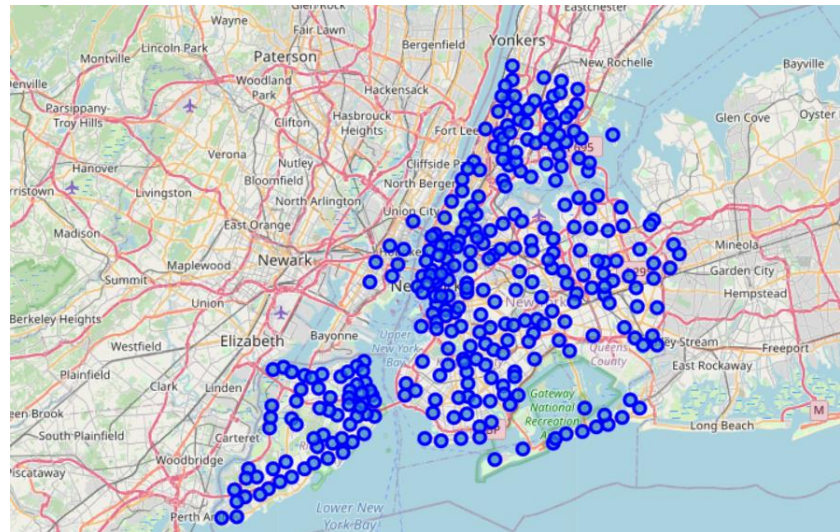# Estimating the Best New Venue

A million dollar question:
What is the best new business to start?

- Finding the best new businesses to start in a given neighborhood is not an easy task. It could be a restaurant, a gym, a bar, or anything.

- Clearly this would be an interesting topic to many investors and entrepreneurs.

- In addition, such information would also be useful to the local government and decision markers, as it can help them with better future city design.

Data Sources

1. Venue data for New York City from Foursquare.com.

2. Venue categories definition data from Foursquare.com.

3. New York Neighborhoods Latitude and Longitude data from Class Lab - Segmenting and Clustering Neighborhoods in New York City.

Data Cleaning

1. Merge the venue category table with original venue data table to retrieve the higher-level (level 1-3) venue information for each row of the original venue data.

```
nearby_venues_full.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category	Level I	Level II	Level III	Level IV
0	Manhattan_Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place	Food	Pizza Place	Pizza Place	Pizza Place
1	Manhattan_Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio	Outdoors & Recreation	Athletics & Sports	Gym / Fitness Center	Yoga Studio
2	Manhattan_Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner	Food	Diner	Diner	Diner
3	Manhattan_Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop	Food	Coffee Shop	Coffee Shop	Coffee Shop
4	Manhattan_Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop	Food	Donut Shop	Donut Shop	Donut Shop

2. Reformat the venue data for each neighborhood using one hot encoding, then aggregate the venue data within each of the neighborhood by taking average (the reformatting is performed on Level II)

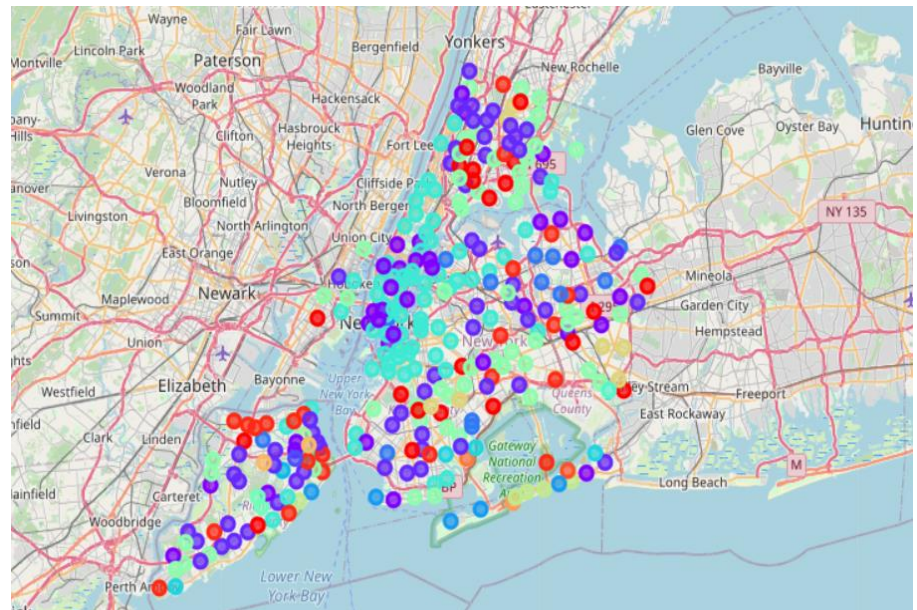|	Neighborhood	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	A
0	Bronx_Allerton	0.0	0.0	0.0	0.034483	0.0	0.00	0.0	0.0	0.0	0.068966	0.034483	
1	Bronx_Baychester	0.0	0.0	0.0	0.000000	0.0	0.05	0.0	0.0	0.0	0.000000	0.000000	
2	Bronx_Bedford Park	0.0	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	0.062500	0.062500	
3	Bronx_Belmont	0.0	0.0	0.0	0.010526	0.0	0.00	0.0	0.0	0.0	0.010526	0.010526	
4	Bronx_Bronxdale	0.0	0.0	0.0	0.000000	0.0	0.00	0.0	0.0	0.0	0.071429	0.071429	

Most Popular Venue in Allerton

◦ We will take Allerton as an example in our study.

◦ The most popular venue in Allerton is Pizza Place

◦ But that does not mean the opening a pizza place is the best choice for new business in Allerton as there might be already too many pizza places in Allerton.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronx_Allerton	Pizza Place	Deli / Bodega	Food & Drink Shop	Department Store	Asian Restaurant
1	Bronx_Baychester	Donut Shop	Bank	Convenience Store	Sandwich Place	Electronics Store
2	Bronx_Bedford Park	Diner	Mexican Restaurant	Pizza Place	Deli / Bodega	Sandwich Place
3	Bronx_Belmont	Italian Restaurant	Food & Drink Shop	Pizza Place	Deli / Bodega	Bakery
4	Bronx_Bronxdale	Food & Drink Shop	Italian Restaurant	Paper / Office Supplies Store	Bank	Mexican Restaurant
5	Bronx_Castle Hill	Southern / Soul Food Restaurant	Bus Station	Market	Pharmacy	Bank
6	Bronx_City Island	Spanish Restaurant	Seafood Restaurant	Thrift / Vintage Store	Jewelry Store	Food & Drink Shop
7	Bronx_Claremont Village	Food & Drink Shop	Asian Restaurant	Bus Station	Bakery	Pizza Place
8	Bronx_Clason Point	Park	Pool	Food & Drink Shop	Boat or Ferry	Latin American Restaurant
9	Bronx_Co-op City	Bus Station	Food & Drink Shop	Discount Store	Salon / Barbershop	Bagel Shop

Comparing with Other Similar Neighborhoods - Clustering

1. To perform a more comprehensive analysis, let's take a look at other neighborhoods that are similar to Allerton so we would have more data to analyze.

2. In order to find the other neighborhoods that are similar to Allerton, we will perform an unsupervised clustering analysis using the K-Means algorithm.

3. We will start with K = 20

Most Popular Venue in the Entire Cluster that Allerton Belongs to

1. There are 60 neighborhoods that are grouped into the same cluster as Allerton.

2. Pizza place is indeed the most popular venue category with in the neighborhoods that are similar to Allerton, followed by Asian Restaurant.

	Venue	Most Popular	2nd Popular	3rd Popular
0	Pizza Place	15.0	13.0	7.0
1	Asian Restaurant	11.0	8.0	7.0
2	Food & Drink Shop	11.0	7.0	8.0
3	Italian Restaurant	5.0	4.0	2.0
4	Deli / Bodega	5.0	3.0	1.0
5	Bus Station	2.0	2.0	1.0
6	Bank	2.0	2.0	2.0
7	Spanish Restaurant	1.0	1.0	0.0
8	Coffee Shop	1.0	0.0	0.0
9	Professional & Other Places	1.0	0.0	0.0
10	Pharmacy	1.0	6.0	2.0
11	Diner	1.0	1.0	1.0
12	Bagel Shop	1.0	0.0	0.0
13	Dessert Shop	1.0	1.0	2.0

The Business that Allerton Needs the Most: sensitivity test on each venue

- Let's view the problem in a different way: what new business/venues does Allerton need the most?

- We can solve this problem using the intra-cluster similarity of the cluster and Allerton belongs to.

- We will perform a sensitivity analysis on each of the venue category in Allerton by a positive shock of +1 (assuming we will start a venue).

  - For example, for the sensitivity on American restaurants, we will just add one more American restaurant to the count of American restaurants. That will slightly change the distribution of the venue categories in Allerton.

Original distribution of venue categories

```
American Restaurant      0.026316
Antique Shop             0.000000
Arcade                   0.000000
Art Gallery              0.000000
Arts & Crafts Store      0.000000
Arts & Entertainment     0.000000
Asian Restaurant         0.052632
Name: 127, dtype: float64
```

Shock the venue "American Restaurants" by adding one more American restaurant to Allerton

Post-shock distribution of venue categories

```
American Restaurant      0.051939
Antique Shop             0.000000
Arcade                   0.000000
Art Gallery              0.000000
Arts & Crafts Store      0.000000
Arts & Entertainment     0.000000
Asian Restaurant         0.051247
Name: 127, dtype: float64
```
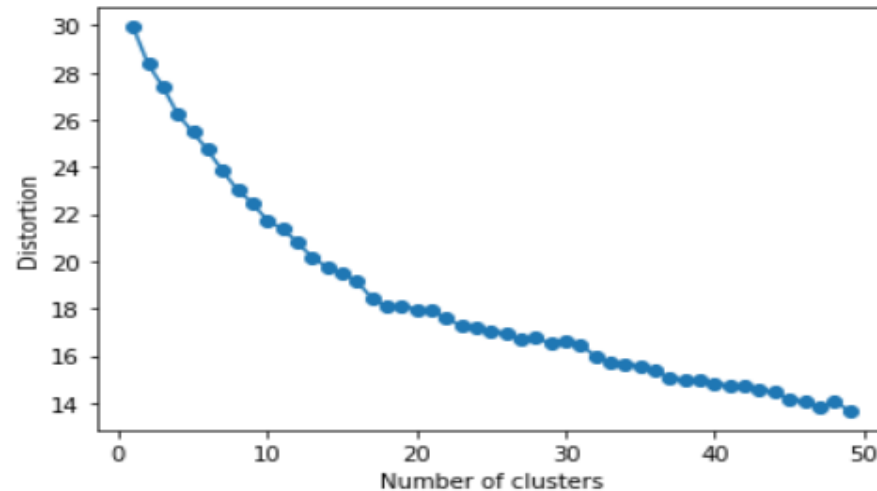
The Business that Allerton Needs the Most: intra-cluster similarities

- For each shock. an increment in the intra-cluster similarity means that by starting a new business in that venue, Allerton is getting more similar to other neighborhoods in the same cluster, which means that venue is what Allerton needs.

- We will repeat the above shocking-testing process on all of the venue categories, and the one result in the largest increment of intra-cluster similarity is the venue category that Allerton needs the most.

- We will calculate the intra-cluster similarity using the distance matrix. Note that they are inversely correlated: higher distance means lower similarity.

	index	Venue	Distance	Delta in Distance
0	21	Bank	15.549499	-0.372283
1	141	Italian Restaurant	15.576119	-0.345663
2	20	Bakery	15.632218	-0.289564
3	230	Sandwich Place	15.635771	-0.286011
4	163	Mexican Restaurant	15.666826	-0.254956
5	19	Bagel Shop	15.676339	-0.245443
6	22	Bar	15.681945	-0.239837
7	78	Diner	15.712208	-0.209574
8	59	Coffee Shop	15.719045	-0.202737
9	46	Café	15.727091	-0.194691
10	191	Park	15.733023	-0.188759

Different choices of K in K Means

Let's take a look at the clustering performance with different K.

Unfortunately, we were unable to find any clear elbow point.

In such scenario, in order to improve the robustness of our model, we will repeat the above analysis for all choices of K, then aggregate the results.

The Best Business Estimated by K Means with Different K (K = 5 to 40)

	K	Neiborhoods in the Group	Best Business	2nd Best Business	3rd Best Business
0	5	135	Bank	Sandwich Place	Bar
1	6	69	Food & Drink Shop	Caribbean Restaurant	Bakery
2	7	157	Bank	Sandwich Place	Bakery
3	8	101	Bank	Sandwich Place	Beach
4	9	72	Food & Drink Shop	Caribbean Restaurant	Bakery
5	10	72	Bank	Bus Stop	Sandwich Place
6	11	116	Bank	Sandwich Place	Bakery
7	12	45	Bank	Bar	Sandwich Place
8	13	67	Bank	Sandwich Place	Italian Restaurant
9	14	46	Bank	Bar	Sandwich Place
10	15	97	Bank	Sandwich Place	Park
11	16	46	Bank	Italian Restaurant	Bar
12	17	61	Bank	Caribbean Restaurant	Clothing Store

The Best Business Estimated by K Means with Different K (K = 5 to 40) Cont.

13	18	36	Bank	Bar	Sandwich Place
14	19	79	Bank	Sandwich Place	Mexican Restaurant
15	20	61	Bank	Italian Restaurant	Bakery
16	21	38	Bank	Sandwich Place	Bar
17	22	41	Food & Drink Shop	Bakery	Sandwich Place
18	23	30	Bank	Bar	Pizza Place
19	24	23	Pizza Place	Bar	Coffee Shop
20	25	43	Bank	Latin American Restaurant	Sandwich Place
21	26	25	Bar	Coffee Shop	Bakery
22	27	48	Bank	Sandwich Place	Bar
23	28	43	Bank	Bakery	Bar
24	29	21	Deli / Bodega	Park	Bank
25	30	22	Bank	Pizza Place	Bagel Shop

Final Scores

1. The outcomes on the last two pages (most recommended venue categories) are overall quite stable with different Ks.

2. Now we aggregate the results above by summing over different Ks. Let's define a score for each venue category:

$$Score_{venue(i)} =$$

$$\sum_{k} \{[3 * I(if\ venue(i)\ is\ the\ most\ recommended\ venue)]$$

$$+ [2 * I(if\ venue(i)\ is\ the\ 2nd\ most\ recommended\ venue)]$$

$$+ [1 * I(if\ venue(i)\ is\ the\ 3rd\ most\ recommended\ venue)]\}$$

Results

- The top 4 best choices for new business in Allerton are Bank, Sandwich Place, Bar and Bakery.

- In particular, Bank is the top one choice, with its score much higher than all other venue categories.

- There is currently no bank in Allerton, while there are banks in most of its similar neighborhoods. Therefore, bank is indeed one of the most needed business in Allerton – the output of our model is consistent with intuition.

	Venue	Best	2nd	3rd	Final Score
0	Bank	25.0	1.0	2.0	79.0
1	Sandwich Place	0.0	10.0	8.0	28.0
2	Bar	3.0	5.0	6.0	25.0
3	Bakery	0.0	5.0	6.0	16.0

Conclusions

- According to the summary table on the last page, we found the top 4 best choices for new business in Allerton are Bank, Sandwich Place, Bar and Bakery.

- In particular, Bank is the top one choice, with its score much higher than all other venue categories.

- Of course, opening a bank is not a practical choice to most people. Therefore, the top three best choices of new businesses in Allerton would be Sandwich Place, Bar and Bakery.

- In addition, the fact that bank is the most needed business in Allerton is quite useful information -- it might be a good choice for banks to open a branch in Allerton