

Introduction/Problem

Problem

In what neighborhood would you recommend opening a café in Manhattan, NY?

Multiple factors need to be considered when considering a business venture that involves a café. Chief among these are the demand, the competition, operating cost, and time before the investment start yielding returns. The demand can be estimated from businesses and residents in a given neighborhood. The competition comprises of other eateries in the area. There needs to be a clear differentiator from the competition for a new business to gain traction. Operating costs include renting space, hiring and employing staff, raw materials, appliances, publicity charges. In this project we are going to focus on honing in which neighborhood(s) would be good candidates for investment in a café in Manhattan. This data will be useful to potential investors, prospective small business owners, or restaurant-chains looking to expand.

Data

We will primarily be using Foursquare location data for this project. We will examine neighborhoods in Manhattan by looking at fixed number of venues. For analysis we will compute the frequency of cafes and other types of venues in each neighborhood. We will then look for neighborhoods where cafes do not form the majority of venues. This indicates there being a need for cafe. We will use regression analysis to determine what surrounding venues best complement a café. We will then predict which neighborhood is a likely candidate for investment in a café.

Methodology

Data wrangling

A publicly available dataset (https://geo.nyu.edu/catalog/nyu_2451_34572) was used to get the list of neighborhoods and their coordinates in New York City. Initially a data frame was built to include the borough, neighborhood, latitude and longitudes. This data was then trimmed to include neighborhoods from Manhattan, NY only (Fig. 1).

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508

Fig 1. Table with all neighborhoods in Manhattan with their coordinates.

Foursquare API call was made to retrieve up to 100 venues within a 500-meter radius of each neighborhood. Following data cleaning, a table was obtained with the following categories: neighborhood, neighborhood latitude, neighborhood longitude, venue, venue category, venue latitude, and venue longitude (Fig. 2).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
5	Marble Hill	40.876551	-73.91066	Rite Aid	40.875467	-73.908906	Pharmacy
6	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium
7	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
8	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	Coffee Shop
9	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

Fig. 2. Venues and their categories from Marble Hill neighborhood in Manhattan

An assumption was made and the venue category “Coffee Shop” was merged with Café for the purposes of this analysis.

Exploratory data analysis

Frequency of each type of venue was computed for each neighborhood as shown in Fig. 3.

	venue	freq
0	Chinese Restaurant	0.09
1	Café	0.07
2	Bakery	0.04
3	Cocktail Bar	0.04
4	Art Gallery	0.04
----Manhattan Valley----		
	venue	freq
0	Café	0.12
1	Yoga Studio	0.05
2	Bar	0.05
3	Pizza Place	0.05
4	Mexican Restaurant	0.05

Fig. 3. Most frequent venue types in Lower East Side and Manhattan Valley

Machine learning and inferential statistical testing

A correlation matrix was generated using frequencies of each type of venue category across all neighborhoods in Manhattan. A multivariable linear model was used with frequency of café as the dependent variable and all other venues as independent variable. The goal was to determine whether surrounding venues could be used as a predictor of presence of cafes. A distribution plot was used to determine in what frequency range of cafes did the model perform the best.

Results

A map was generated to demonstrate exactly what neighborhoods were included in this analysis Fig. 4.

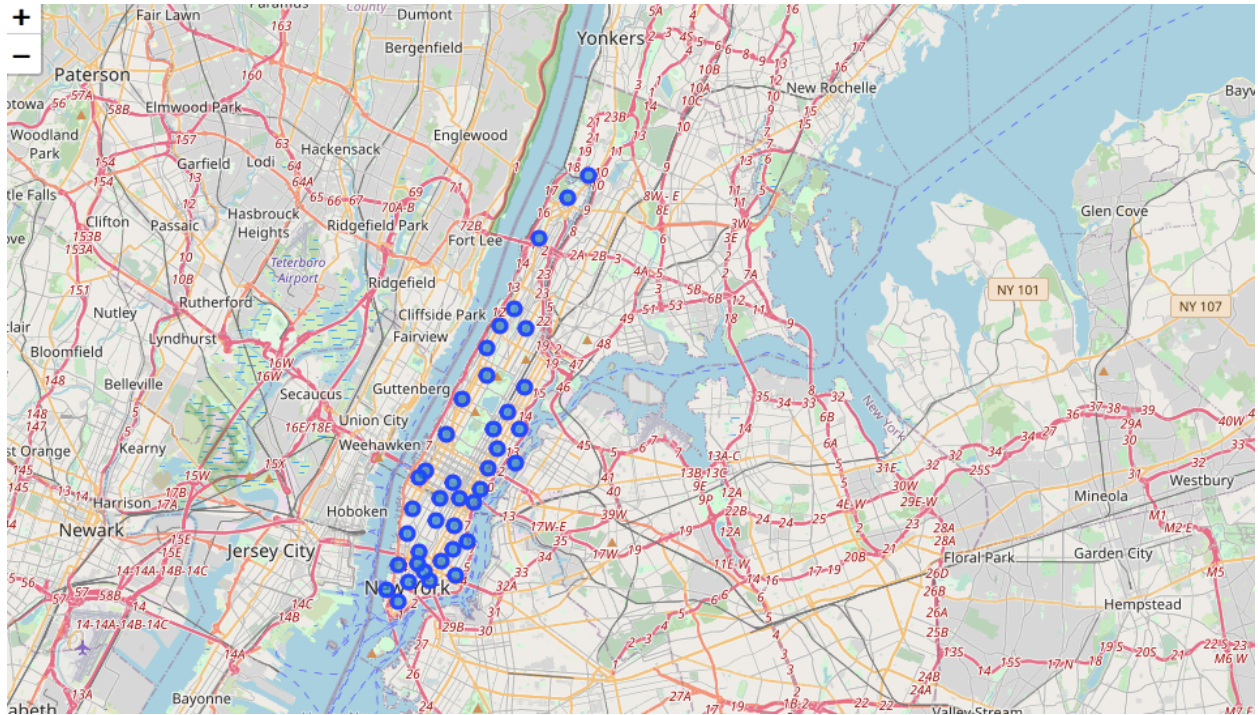


Fig. 4. Map showing what neighborhood of Manhattan were included in this analysis

A total of 40 different neighborhoods were analyzed and there were 328 unique venue categories. There were 224 Cafés distributed across the 40 neighborhoods. Incredibly in 33 of the 40 neighborhoods Café was among the top 3 of the most frequent venue categories.

Among venue categories that had at least 10 entries Yoga Center and Pizza Place were most correlated with Cafes (Fig. 5 and Fig. 6).

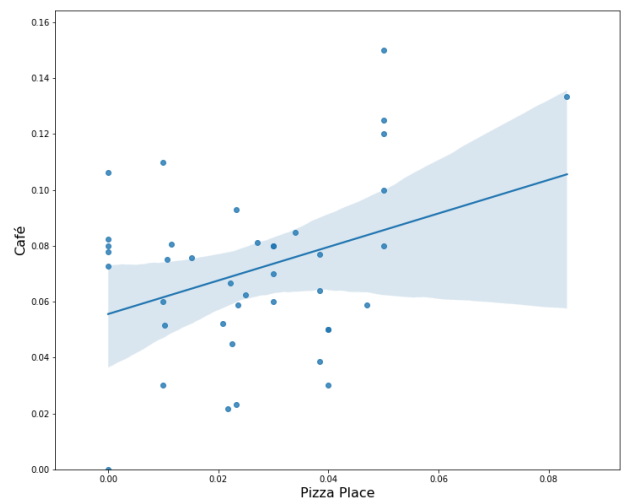
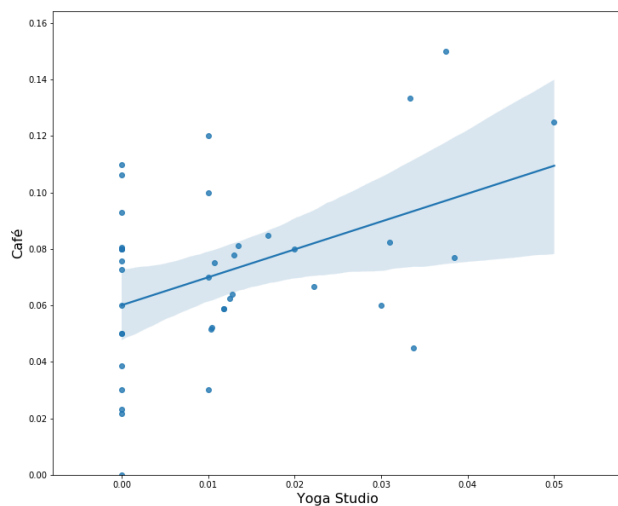


Fig 5 and 6. Correlation of Café vs Yoga Studio and Café vs Pizza Place

A multiple linear regression model was fit with Café frequency as the dependent variable and frequencies of Yoga Studio, Pizza Place, and Sushi Restaurant as independent variables. A distribution plot of fitted vs observed café frequency showed the model performed reasonably well in cases where café frequency was between 0.5 and 0.1 (Fig. 7).

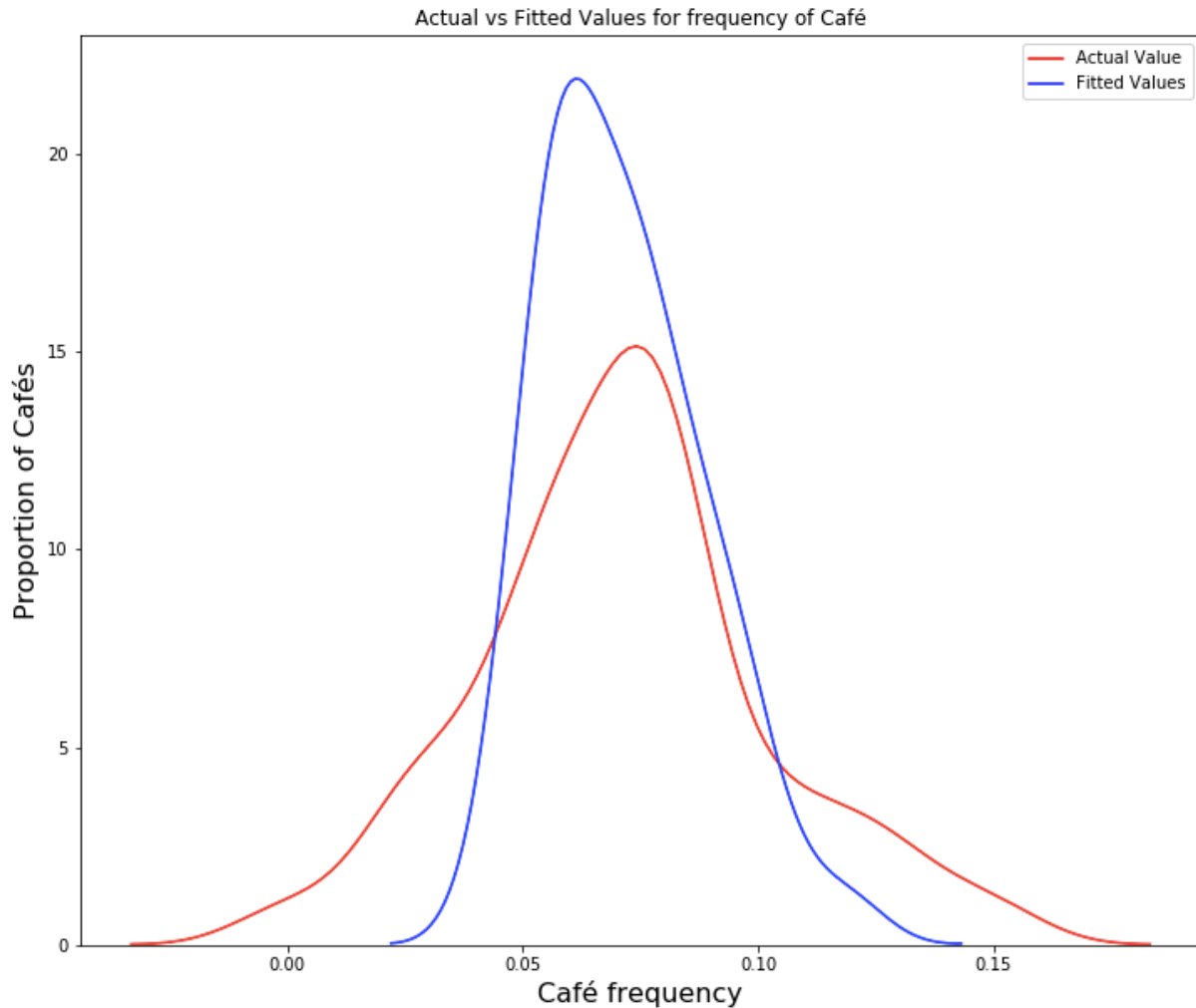


Fig. 7. Distribution plot of observed café frequency and café frequency predicted by the model

The 7 neighborhoods where Café was not in the top 3 most frequent venue types are shown in Fig 8.

	Neighborhood	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	Central Harlem	African Restaurant	Cosmetics Shop	Bar	Seafood Restaurant	French Restaurant	Chinese Restaurant	American Restaurant	Market	Southern / Soul Food Restaurant	Caribbean Restaurant
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Salon / Barbershop	Spa	Ice Cream Shop	American Restaurant	Optical Shop	Vietnamese Restaurant
2	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Latin American Restaurant	Deli / Bodega	Spa	Sandwich Place	Beer Bar	Liquor Store	Gas Station
3	Little Italy	Bubble Tea Shop	Mediterranean Restaurant	Bakery	Italian Restaurant	Ice Cream Shop	Chinese Restaurant	Pizza Place	Spa	Thai Restaurant	Café
4	Roosevelt Island	Park	Bubble Tea Shop	School	Liquor Store	Metro Station	Supermarket	Food & Drink Shop	Soccer Field	Café	Farmers Market
5	Stuyvesant Town	Park	Bar	Harbor / Marina	Helipoint	Farmers Market	Pet Service	Gas Station	Cocktail Bar	Baseball Field	Gym / Fitness Center
6	Upper East Side	Italian Restaurant	Bakery	Gym / Fitness Center	Café	Yoga Studio	Wine Shop	Juice Bar	French Restaurant	Exhibit	Spa

Fig. 8. Neighborhoods where Cafés are less popular than other neighborhoods

We used our model to predict café frequency in these 7 neighborhoods and plotted the difference in observed frequency and predicted frequency (Fig. 9). We propose the greater the discrepancy the higher is the need for a Café.

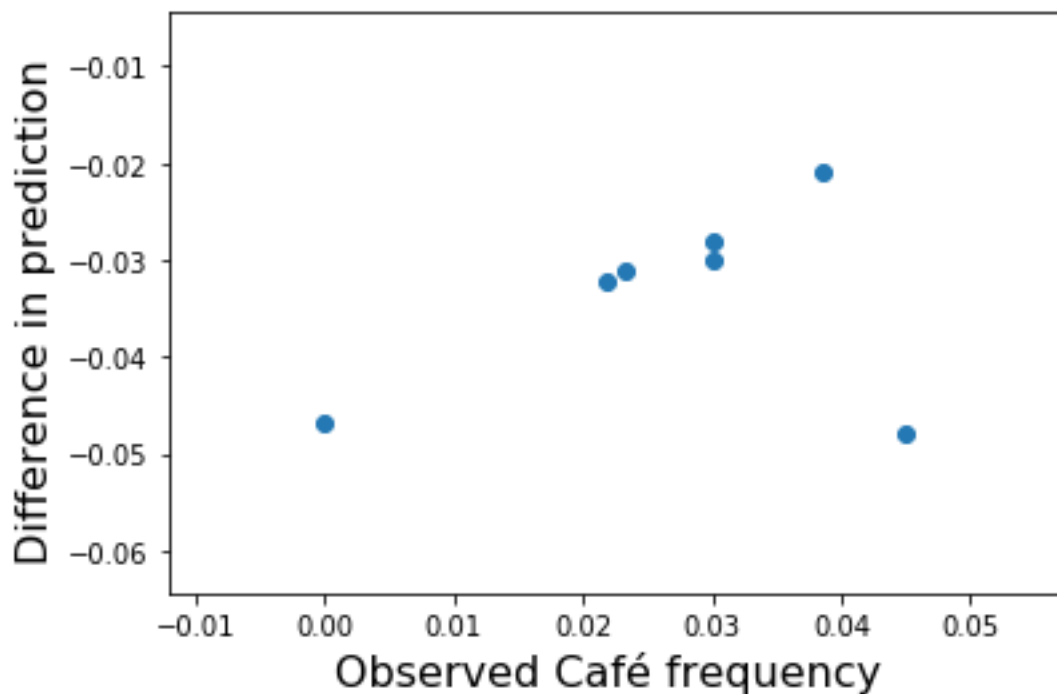


Fig. 9. The difference in prediction in the 7 neighborhoods where Cafés are less popular

Discussion

Multiple variables need to be considered when deciding on whether to invest in a Café. This analysis is limited to determining the optimal location for a new Café in Manhattan. Our model assumes that data from nearby venues could potentially predict an optimal location for a new Café. Our analyses revealed that in 33 out of 40 neighborhoods Cafes were really popular. Our analysis also shows that presence of cafes is highly correlated with those of Yoga Studios which is quite intuitive.

Conclusion

We recommend the following neighborhoods for investment in a new Café:

- A) Upper East Side
- B) Stuyvesant Town