

# Location Selection for New Business

Prasad K  
January 1, 2020

## 1. Introduction

### 1.1 Background

Many people can't imagine starting their day without a cup of coffee in the morning. 66% of American women drink coffee every day compared to 62% of American men. An average American drinks 3.1 cups of coffee per day.

New York City has more coffee shops and cafes than any place else in the U.S. Manhattan's daytime population is approximately 4 million and keeps growing. So there is still lot of potential to open new coffee outlets.

### 1.2 Problem and Interests

A coffee house chain has three coffee retail stores in New York city. They wanted to expand their business by opening more stores in various locations in the city. The location selection depends on the population working and living in the area along with venues/activities around the location. They wanted to make a study of the existing stores' data and neighborhoods to determine best locations for their new stores and to predict the sales.

Other retailers can also be interested in similar information.

## 2. Data acquisition and cleaning

### 2.1 Data sources

Store information and neighborhood information are the two types of datasets required for this analysis.

#### 2.1.1 Store and sales information

There are not many retail store level datasets available for public use, but found a small data set in Kaggle. This contains existing 3 store locations and 5 possible new locations. The current operating store data consists of store locations and sales for each transaction for one month. But the problem is that the store sales in the dataset are almost equal. I still used this dataset

with unavailability of any other sales data set.

Sample Data:

sales\_outlet.csv:

```
sales_outlet_id,sales_outlet_type,store_square_feet,store_address,store_city,store_state_province,store_telephone,store_postal_code,store_longitude,store_latitude,manager,Neighborhood
2,warehouse,3400,164-14 Jamaica Ave,Jamaica,NY,972-871-0402,11432,-73.795168,40.705226,,Jamaica
3,retail,1300,32-20 Broadway,Long Island City,NY,777-718-3190,11106,-73.924008,40.761196,6,Astoria
4,retail,1300,604 Union Street,Brooklyn,NY,619-347-5193,11215,-73.983984,40.677645,11,Gowanus
5,retail,900,100 Church Street,New York,NY,343-212-5151,10007,-74.01013,40.71329,16,Lower Manhattan
6,retail,1000,122 E Broadway,New York,NY,613-555-4989,10002,-73.992687,40.713852,21,Lower East Side
7,retail,1200,224 E 57th Street,New York,NY,287-817-2330,10021,-73.96,40.77,26,Upper East Side
8,retail,1500,687 9th Avenue,New York,NY,652-212-7020,10036,-73.990338,40.761887,31,Hell's Kitchen
9,retail,1700,175 8th Avenue,New York,NY,242-212-0080,10011,-74.000502,40.74276,36,Chelsea
10,retail,1600,183 W 10th Street,New York,NY,674-646-6434,10014,-74.002722,40.734367,41,Greenwich Village
```

sales\_receipts.csv:

```
"transaction_id","transaction_date","transaction_time","sales_outlet_id","staff_id","customer_id","instore_yn","order","line_item_id","product_id","quantity","line_item_amount","unit_price","promo_item_yn"
7,2019-04-01,12:04:43,3,12,558,N,1,1,52,1,2.50,2.50,N
11,2019-04-01,15:54:39,3,17,781,N,1,1,27,2,7.00,3.50,N
19,2019-04-01,14:34:59,3,17,788,Y,1,1,46,2,5.00,2.50,N
32,2019-04-01,16:06:04,3,12,683,N,1,1,23,2,5.00,2.50,N
33,2019-04-01,19:18:37,3,17,99,Y,1,1,34,1,2.45,2.45,N
39,2019-04-01,18:54:46,3,17,664,Y,1,1,32,1,3.00,3.00,N
50,2019-04-01,13:03:49,3,12,316,N,1,1,49,2,6.00,3.00,N
53,2019-04-01,11:21:14,3,12,38,N,1,1,60,1,3.75,3.75,N
```

### 2.1.2 Neighborhood information

Neighborhood information can be obtained from Foursquare. It provides the most popular nearby venues and their categories. It's good to have living and working population statistics within 500 meters of each store location, but couldn't get it as most of these store locations are within Manhattan.

Used the API <https://api.foursquare.com/v2/venues/explore?> to get the top 100 venues within 500 meters of each location.

Venue	Venue_Latitude	Venue_Longitude	Venue_Category
Astoria Bier & Cheese	40.760581	-73.922542	Cheese Shop
Yoga Agora	40.761200	-73.923862	Yoga Studio
Lockwood	40.760928	-73.924028	Gift Shop
Brooklyn Bagel & Coffee Co.	40.760408	-73.921967	Bagel Shop
King Of Falafel & Shawarma	40.762041	-73.925098	Middle Eastern Restaurant

## 2.2 Data pre-processing / cleaning

Stores data:

- Spelling is corrected in the column header for neighborhood
- Deleted the data for warehouse where the manager information is not available
- Changed the manager id to integer from float

store_square_feet	store_address	store_city	store_state_province	store_telephone	store_postal_code	store_longitude	store_latitude	manager	Neighborhood
1300	32-20 Broadway	Long Island City	NY	777-718-3190	11106	-73.924008	40.761196	6	Astoria
1300	604 Union Street	Brooklyn	NY	619-347-5193	11215	-73.983984	40.677645	11	Gowanus
900	100 Church Street	New York	NY	343-212-5151	10007	-74.010130	40.713290	16	Lower Manhattan
1000	122 E Broadway	New York	NY	613-555-4989	10002	-73.992687	40.713852	21	Lower East Side
1200	224 E 57th Street	New York	NY	287-817-2330	10021	-73.960000	40.770000	26	Upper East Side
1500	687 9th Avenue	New York	NY	652-212-7020	10036	-73.990338	40.761887	31	Hell's Kitchen
1700	175 8th Avenue	New York	NY	242-212-0080	10011	-74.000502	40.742760	36	Chelsea
1600	183 W 10th Street	New York	NY	674-646-6434	10014	-74.002722	40.734367	41	Greenwich Village

Sales data:

- Replaced the store number with store neighborhood name from stores dataset
- Grouped the data by store by day to understand the sales trend
- Grouped the data by store only to use in the modeling

	Store_Neighborhood	transaction_date	line_item_amount
0	Astoria	2019-04-01	2571.40
1	Astoria	2019-04-02	2701.50
2	Astoria	2019-04-03	2759.05
3	Astoria	2019-04-04	2511.75
4	Astoria	2019-04-05	2669.55

	Store_Neighborhood	line_item_amount
0	Astoria	77213.23
1	Hell's Kitchen	79528.25
2	Lower Manhattan	76894.47

Neighborhood information:

- There are total of 800 venues in 203 categories in our dataset. These are too many categories and some store locations missing many of these categories, so it may skew the results. Analyzed the data and decided to merge some of these categories into one. For example: categories that contain the word 'Restaurant', 'Diner', 'Steak', 'Bistro', 'BBQ' are grouped into one category called 'Restaurant'. The assumption here is that in general there will not be much difference of coffee drinking habits between the people who goes to different restaurants. Now the categories are reduced to 24.
- After the above consolidation of categories, there are four categories in only 2 or 3 out of 8 store neighborhoods. Eg.: Children store. Removed data for these four categories.

There are 20 unique categories.

```
['Bakery',  
'Bar',  
'Clothing',  
'Coffee',  
'Dessert',  
'Food',  
'Grocery',  
'Gym',  
'Lounge',  
'Medical',  
'Miscellaneous Store',  
'Museum',  
'Music Place',  
'Outdoors',  
'Plaza',  
'Restaurant',  
'Shopping Mall',  
'Spa',  
'Theater',  
'Women Store']
```