New York City - Theaters' visitors Venue Recommendation

Harrison J. Angonga

May 24, 2019

1. Introduction

In this project we will try to find an optimal selection of venues recommendation for travelers visiting in Theaters in New York City. Specifically, this report will be targeted to tour packages companies and travel websites that are looking to recommend venues around in New York that will interest travelers.

Since there are lots of venues in New York we will try to detect locations that are most common and within 100meters of the theaters. We are also particularly interested in areas close to city center as possible.

There are a lot of websites where travelers can check and retrieve recommendations of places to stay or visit. However, most of these websites provides recommendation simply based on usual tourist attractions or key residential areas that are mostly expensive or already known for travelers based on certain keywords like "Hotel", or "Backpackers" etc. The intention on this project is to collect and provide a data driven recommendation that can supplement the recommendation with statistical data. This will also be utilizing data retrieved from New Yorke open data sources and Foursquare API venue recommendations.

The sample recommender in this notebook will provide the following use case scenario:

A person planning to visit New York as a Tourist or a Tour Packaging company looking for an optimal list of venues in New York to recommend to clients. The user wants to receive venue recommendation where he can stay or eat in a restaurant with close proximity to places of interest in this case Theaters. The recommendation should not only present the most viable option, but also present a comparison table of all possible town venues. For this demonstration, this notebook will make use of the following data:

New York Theaters data. Popular venues in the vicinity. Note: While this demo makes use of Foursquare Venue Category, Other possible categories can also be used for the same implementation such as checking categories like:

Outdoors and Recreation Nightlife Nearby Schools, etc. I will limit the scope of this search as Foursquare API only allows 50 free venue query limit per day when using a free user access.

2. Data Acquisition

Based on definition of our problem, factors that will influence our decision are:

- Distance of theaters from venues Popularity (Commonness) of venue on Foursquare

The following data sources will be needed to extract/generate the required information:

- Venue categories around every theater will be obtained using Foursquare API (Foursquare website: www.foursquare.com)
- Coordinate of New York will be obtained using Python geopy.geocoders, Nominatim library
- Coordinates and list of theaters in New York will be obtained from New York City Open Data portal https://data.cityofnewyork.us

I will be using the Foursquare API to explore locations in New York City. The Foursquare explore function will be used to get the most common venue categories around each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

Venue ID Venue Name Coordinates: Latitude and Longitude Category Name Another venue query will be performed to retrieve venue ratings for each location. Note that rating information is a paid service from Foursquare and we are limited to only 50 queries per day. With this constraint, we limit the category analysis with only one type for this demo.

3. Methodology

The source data contains all listings of theaters in New York City as of May 2019. I will retrieve this data. For this demonstration, I will simplify the analysis by using 5 features of the data: Name, Latitudes, Longitudes, Address and City.

Data Cleanup and regrouping. The retrieved table contains some unwanted entries and needs some cleanup.

The following tasks will be performed:

- Drop/ignore cells with missing data
- Use most recent data record
- Fix data types. Convert column data types to relevant ones
- Geographical coordinates for the data

Retrieve and clean up coordinates

The data from New York Open data portal already has geometry coordinates. This simplifies the process to removing unwanted characters and remain with the desired coordinates. For easy plotting, we separate the Latitude and Longitudes to separate columns. We then convert the columns form python objects to float64.

Before:

Load and explore the data

```
In [3]: df = pd.read_csv('theaters.csv') # Read data from csv to a pandas dataframe
        # take a look at the dataset
        df.head()
```

Out[3]:

	the_geom	NAME	TEL	URL	ADDRESS1	ADDRES2	CITY	ZIP
0	POINT (-73.99061840882582 40.75985115447559)	45th Street Theater	(212) 352-3101	http://www.theatermania.com/new-york/theaters/...	354 West 45th Street	NaN	New York	10036
1	POINT (-73.9881059525377 40.76047123447081)	47th Street Theater	(800) 775-1617	http://www.bestofoffbroadway.com/theaters/47st...	304 West 47th Street	NaN	New York	10036
2	POINT (-73.97038450260143 40.76339942774153)	59E59	(212) 753-5959	http://www.59e59.org/	59 East 59th Street	NaN	New York	10022
3	POINT (-73.99332384622063 40.7585366821068)	Acorn Theater	(212) 279-4200	http://www.theatrerow.org/theacorn.htm	410 West 42nd Street	NaN	New York	10036
4	POINT (-73.9892143340222 40.75926091219353)	Al Hirschfeld Theater	(212) 239-6200	http://www.newyorkcitytheatre.com/theaters/alh...	302 W 45th Street	NaN	New York	10036

	the_geom
0	POINT (-73.99061840882582 40.75985115447559)
1	POINT (-73.9881059525377 40.76047123447081)
2	POINT (-73.97038450260143 40.76339942774153)
3	POINT (-73.99332384622063 40.7585366821068)
4	POINT (-73.9892143340222 40.75926091219353)

After:

	Name	Address	City	Longitudes	Latitudes
0	45th Street Theater	354 West 45th Street	New York	-73.990618	40.759851
1	47th Street Theater	304 West 47th Street	New York	-73.988106	40.760471
2	59E59	59 East 59th Street	New York	-73.970385	40.763399
3	Acorn Theater	410 West 42nd Street	New York	-73.993324	40.758537
4	Al Hirschfeld Theater	302 W 45th Street	New York	-73.989214	40.759261

4. Segmenting and Clustering Theaters in New York

Retrieving Foursquare Places of Interest.

Using the Foursquare API. The **explore** API function was used to get the most common venue categories in each location, and then used this feature to group the locations into clusters, the *k*-means clustering algorithm was used for the analysis. Finally, the Folium Library was used to visualize the recommended locations and their emerging clusters.

In the ipynb notebook, the function getNearbyVenues extracts the following information for the dataframe it generates:

- Venue ID
- Venue Name
- Coordinates: Latitudes and Longitudes
- Category Name

The function getVenuesByCategory performs the following:

1. Category based venue search to simulate searches based on certain places of interest. This search extracts the following information:
    - Venue ID
    - Venue Name
    - Coordinates: Latitudes and Longitude
    - Category Name
2. For each retrieved Venue ID, retrieve the venues category rating.

The generated dataframe in the second function contains the following column:

Search Venues with recommendations on: Food Venues (Restaurants, Fast foods, etc.)

To demonstrate user selection of places of interest, we will use this Food Venues category in our further analysis.

- This Foursquare search is expected to collect venues in the following category:
- Category
- Food Courts
- Restaurants
- Cafes
- Coffee Shops
- Other food venues

Data cleanup

- Eliminate possibilities of duplicates
- Improve the quality of our venue selection by removing venues with no ratings or 0.0

	Theater	Theater Latitude	Theater Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	45th Street Theater	40.759851	-73.990618	Kinky Boots at the Al Hirschfeld Theatre	40.759384	-73.989173	Theater
1	45th Street Theater	40.759851	-73.990618	Broadway Dance Center	40.759490	-73.989562	Dance Studio
2	45th Street Theater	40.759851	-73.990618	Birdland	40.758947	-73.989677	Jazz Club
3	45th Street Theater	40.759851	-73.990618	Amy's Bread	40.761323	-73.990414	Bakery
4	45th Street Theater	40.759851	-73.990618	Gyu-Kaku Japanese BBQ	40.759042	-73.990043	Japanese Restaurant

Venue Category	Theater	Theater Latitude	Theater Longitude	Venue	Venue Latitude	Venue Longitude
Accessories Store	2	2	2	2	2	2
American Restaurant	36	36	36	36	36	36
Arcade	1	1	1	1	1	1
Argentinian Restaurant	2	2	2	2	2	2
Art Gallery	7	7	7	7	7	7
Art Museum	6	6	6	6	6	6
Arts & Crafts Store	7	7	7	7	7	7
Asian Restaurant	8	8	8	8	8	8
Australian Restaurant	1	1	1	1	1	1
Austrian Restaurant	3	3	3	3	3	3

Analyze Each Location nearby recommended venues

- Technique: One Hot Encoding

```
# one hot encoding
new_york_onehot = pd.get_dummies(new_york_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
new_york_onehot['Theater'] = new_york_venues['Theater']

# move neighborhood column to the first column
fixed_columns = [new_york_onehot.columns[-1]] + list(new_york_onehot.columns[:-1])
new_york_onehot = new_york_onehot[fixed_columns]

new_york_onehot.head()
```

Analysis of New York city most common venues (Small Excerpt from the analysis)

```
----45th Street Theater----
                venue  freq
0        Burger Joint  0.10
1              Bakery  0.10
2              Resort  0.05
3       Indie Theater  0.05
4  Chinese Restaurant  0.05


----47th Street Theater----
                  venue  freq
0              Jazz Club  0.05
1                 Resort  0.05
2            Coffee Shop  0.05
3  Performing Arts Venue  0.05
4       Sushi Restaurant  0.05


----59E59----
               venue  freq
0              Hotel  0.10
1                Spa  0.10
2      Indie Theater  0.05
3  French Restaurant  0.05
4         Food Truck  0.05


----Acorn Theater----
                    venue  freq
0     Gym / Fitness Center  0.15
1             Pizza Place  0.05
2      Peruvian Restaurant  0.05
3  New American Restaurant  0.05
4              Steakhouse  0.05


----Al Hirschfeld Theater----
                  venue  freq
0          Burger Joint  0.05
1         Indie Theater  0.05
2  Performing Arts Venue  0.05
3                Bakery  0.05
4         Ice Cream Shop  0.05


----Ambassador Theatre----
                          venue  freq
0                   Coffee Shop  0.05
1  Vegetarian / Vegan Restaurant  0.05
2             Mexican Restaurant  0.05
3                            Gym  0.05
4                            Bar  0.05


----Broadhurst Theatre----
           venue  freq
0      Jazz Club  0.05
1        Exhibit  0.05
2   Dance Studio  0.05
3  Indie Theater  0.05
4   Concert Hall  0.05


----Broadway Theatre----
            venue  freq
0  Sandwich Place  0.05
1      Restaurant  0.05
2      Steakhouse  0.05
3      Taco Place  0.05
4   Grocery Store  0.05


----Brooks Atkinson Theatre----
                          venue  freq
0                         Plaza  0.05
1                        Resort  0.05
2                Ice Cream Shop  0.05
3  Vegetarian / Vegan Restaurant  0.05
4                         Hotel  0.05


----Castillo Theater----
                  venue  freq
0                   Gym  0.10
1  Gym / Fitness Center  0.10
2             Wine Shop  0.10
3              Building  0.05
4     Health Food Store  0.05


----Century Center For the Performing Arts----
                     venue  freq
0                 Wine Shop  0.10
1               Yoga Studio  0.05
2                 Bookstore  0.05
3             Indie Theater  0.05
4  Mediterranean Restaurant  0.05


----Cherry Lane Theatre----
                 venue  freq
0          Pizza Place  0.10
1  American Restaurant  0.10
2  Japanese Restaurant  0.10
3         Cheese Shop  0.10
4       Sandwich Place  0.05
```

5. Results: Categorized Results

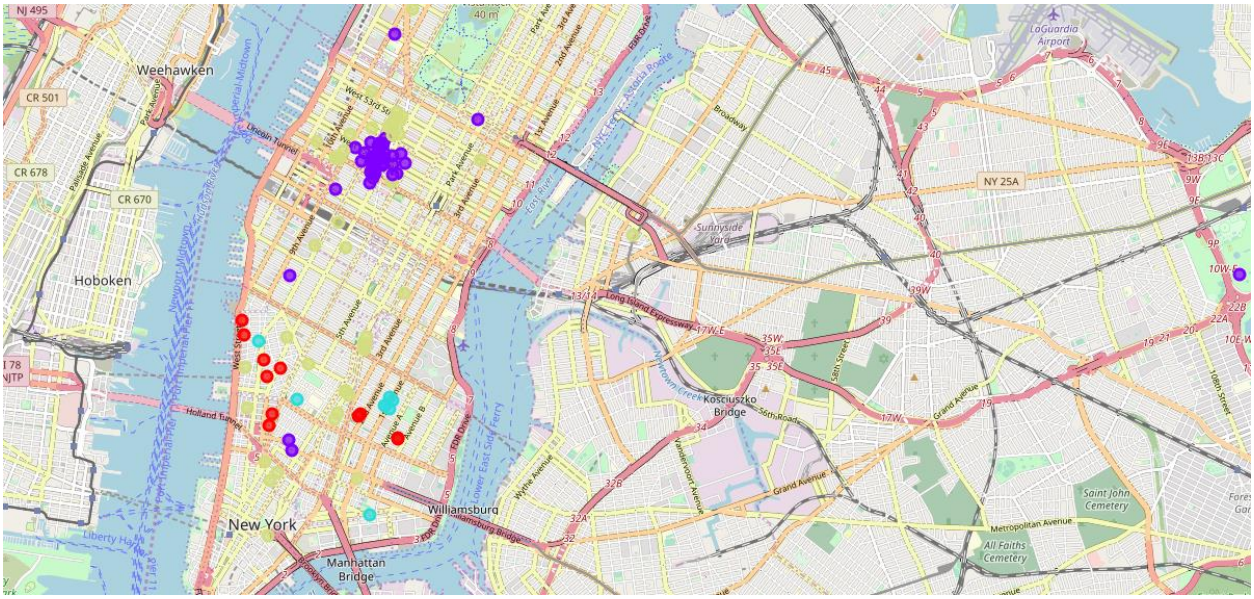
	Theater	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	45th Street Theater	Burger Joint	Bakery	Cosmetics Shop	Sushi Restaurant	Ice Cream Shop
1	47th Street Theater	Italian Restaurant	Jazz Club	Sushi Restaurant	Resort	Coffee Shop
2	59E59	Spa	Hotel	French Restaurant	Shoe Store	Food Truck
3	Acorn Theater	Gym / Fitness Center	Pie Shop	Sandwich Place	Dive Bar	Gift Shop
4	Al Hirschfeld Theater	Dance Studio	Jazz Club	Bakery	Resort	Performing Arts Venue

6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Hotel	Japanese Restaurant	Jazz Club	Juice Bar	Dance Studio
Vegetarian / Vegan Restaurant	Performing Arts Venue	Dance Studio	American Restaurant	Ice Cream Shop
Mediterranean Restaurant	Steakhouse	Salon / Barbershop	Boutique	Liquor Store
French Restaurant	Steakhouse	New American Restaurant	Chinese Restaurant	Peruvian Restaurant
Indie Theater	Burger Joint	Japanese Restaurant	Ice Cream Shop	American Restaurant

5.1. Results: KMeans Cluster

	Theater	Address	City	Longitudes	Latitudes	Cluster Labels	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	45th Street Theater	354 West 45th Street	New York	-73.990618	40.759851	1	Burger Joint	Bakery	Cosmetics Shop	Sushi Restaurant	Ice Cream Shop	Hotel	Japanese Restaurant	Jazz Club	Juice Bar	Dance Studio
1	47th Street Theater	304 West 47th Street	New York	-73.988106	40.760471	1	Italian Restaurant	Jazz Club	Sushi Restaurant	Resort	Coffee Shop	Vegetarian / Vegan Restaurant	Performing Arts Venue	Dance Studio	American Restaurant	Ice Cream Shop
2	59E59	59 East 59th Street	New York	-73.970385	40.763399	1	Spa	Hotel	French Restaurant	Shoe Store	Food Truck	Mediterranean Restaurant	Steakhouse	Salon / Barbershop	Boutique	Liquor Store
3	Acorn Theater	410 West 42nd Street	New York	-73.993324	40.758537	3	Gym / Fitness Center	Pie Shop	Sandwich Place	Dive Bar	Gift Shop	French Restaurant	Steakhouse	New American Restaurant	Chinese Restaurant	Peruvian Restaurant
4	Al Hirschfeld Theater	302 W 45th Street	New York	-73.989214	40.759261	1	Dance Studio	Jazz Club	Bakery	Resort	Performing Arts Venue	Indie Theater	Burger Joint	Japanese Restaurant	Ice Cream Shop	American Restaurant
5	Ambassador Theatre	219 West 49th Street	New York	-73.985001	40.761259	3	Italian Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Restaurant	Mexican Restaurant	Bar	Gym	Food Truck	Pizza Place	Hotel
6	American Airlines Theatre	227 W 42nd Street	New York	-73.988049	40.756779	1	Italian Restaurant	Taco Place	Hotel	Music Store	Exhibit	Comic Shop	Cajun / Creole Restaurant	Burger Joint	Indie Theater	Asian Restaurant
7	Apollo Theater	253 West 125th Street	New York	-73.950040	40.810037	3	Arts & Crafts Store	Yoga Studio	Gym / Fitness Center	Pizza Place	Indian Restaurant	Caribbean Restaurant	Shoe Store	Spanish Restaurant	French Restaurant	Sporting Goods Shop
8	Arclight Theatre	152 W 71st St	New York	-73.981210	40.777120	3	Juice Bar	Italian Restaurant	Wine Bar	Coffee Shop	Gym / Fitness Center	Spa	Bakery	Bookstore	Church	French Restaurant
9	Astor Place Theatre	434 Lafayette St	New York	-73.992344	40.729377	3	Cycle Studio	Gym	Bagel Shop	Cosmetics Shop	Coffee Shop	Ramen Restaurant	Salad Place	Music Venue	Soba Restaurant	Miscellaneous Shop
10	Atlantic Theatre	336 W 20th St	New York	-74.001554	40.743888	1	Ice Cream Shop	American Restaurant	Cupcake Shop	Italian Restaurant	Poke Place	Burger Joint	Shoe Repair	Speakeasy	Beer Bar	Bar
11	August Wilson Theatre	245 W 52nd Street	New York	-73.984200	40.763372	3	Bar	Mexican Restaurant	Restaurant	Steakhouse	Grocery Store	Gym	Ramen Restaurant	Sandwich Place	Karaoke Bar	Food Truck

5.2. Visualization



6. Discussion

On this notebook, Analysis of best venue recommendations based on theater locations has been presented. Recommendations based on other user searches like available restaurants and recreation areas are also available. As New York is a metropolitan City with a whole host of interesting venues scattered around the city, the information extracted in this notebook will be a good supplement to web-based recommendations for visitors to find out nearby venues of interest and be a useful aid in deciding a place to stay or where to go during their visits.

Using Foursquare API, we have collected a good amount of venue recommendations in New York City. Sourcing from the venue recommendations from Foursquare has its limitation, the list of venues is not exhaustive list of all the available venues is the area. Furthermore, not all the venues found in the area has a stored rating. For this reason, the number of analyzed venues are only about 60% of all the available venues initially collected. The results therefore may significantly change, when more information is collected on those with missing data.

7. Conclusion

The generated clusters from our results shows that restaurants are the most common venues around theaters, Italian restaurants. This kind of results may be very interesting for travelers who are looking for a specific type of restaurant to visit.

This information may also be used by investors who are looking to put up a new theater. Knowing the most common venues around highly rated theaters might help position a theater around readily available customers.