Battle of the Neighborhoods

Exploring New York - A Brewhaus for Pizza enthusiast

Eddie Sosa

Introduction

- I am an expert in selecting locations for new business venues
- My services have been retained to assist in locating the perfect neighborhood for a new beer garden concept
    - Beerizza Inc. has perfected a brew whose full bouquet of aromas and flavors are best enjoyed alongside a pizza.
    - Beerizza Inc., while master brewers, the expertise in pizza making is lacking
    - Owners would like to open a new beer garden, but don’t want to open a kitchen
- New York is a major metropolitan city famous for its pizza
- Here we find a solution to the Business Problem, Where to open up the new beer with high exposure to pizza restuarants

Data to be used

To build our model we will be using the following data sources

- Foursquare API venue data for exploring venues in various neighborhoods and boroughs of NYC
- New York City neighborhood data that includes boroughs and longitude and latitude data to assist in exploring neighborhoods from https://cocl.us/new_york_dataset
- New York City Geospace data for visualization fromm https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Methodology

- The first step was to collect all of the data and create the relevant data frames using the sources previously listed
- Next we mapped the city
- We then tried to identify additional features that may be useful in location selection (proximity to public transport for example, or nightlife)
- The questions that we will resolve to meet Beerizza Inc. requirements
  - 1. What is the best neighborhood for the new beer garden with respect to pizza density
  - 2. Where is restaurant density highest (in case the beer customer prefers other foods)
  - 3. where is the pizza shop graveyard that should be avoided completely.
  - 4. Can this be on Staten Island
- We will use clustering to segment the data to help us find the best location.

Approach

1. Import appropriate libraries
2. Collect the New York City data from https://cocl.us/new_york_dataset
3. Use FourSquare API to will find all venues for each neighborhood.
4. Inspect and explore the data
5. Visualize the data and restaurants across the city
6. Model data using clustering to find locations with key venues to support our business

Collecting, processing and inspecting the data

The following libraries were used for this project

- pandas, numpy, import, geocoder, os, folium, geopy, matplotlib, sklearn, wget, json

Collecting, processing and inspecting the data

We then downloaded the nyc_data set as a json and converted it into a pandas df

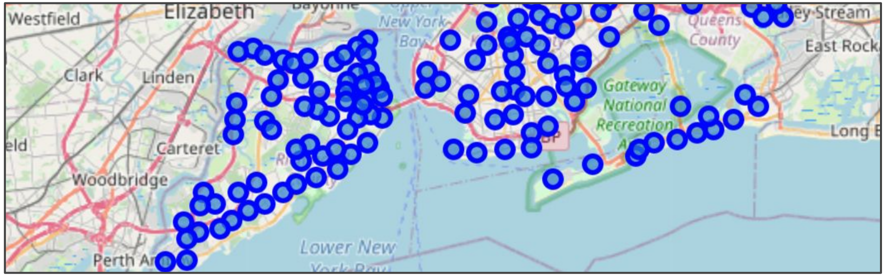
- The df contained information for the 5 Boroughs of New York City as well as the L & L

```
The dataframe has 5 boroughs and 306 neighborhoods.
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Collecting, processing and inspecting the data

We then took the steps to map the city to get a general layout of neighborhoods and density, using 40.71, -74.006 (L&L) as the center point of NYC for our analysis. The city is very dense with neighborhoods. The client decided to focus on Staten Island as they would also like to make the beer garden Wu-Tang Clan themed



Collecting, processing and inspecting the data

Next we work on importing the FourSquare Venue Data. We imported all venues within a 6 mile radius of the center and included up to 1000 venues. In the end we have 830 venues to work with. Because of the radius, some of the venues might be in neighboring boroughs, but that is ok as they will not meet our borough criteria in later steps.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	St. George	40.644982	-74.079353	A&S Pizzeria	40.643940	-74.077626	Pizza Place
1	St. George	40.644982	-74.079353	Beso	40.643306	-74.076508	Tapas Restaurant
2	St. George	40.644982	-74.079353	Staten Island September 11 Memorial	40.646767	-74.076510	Monument / Landmark
3	St. George	40.644982	-74.079353	Richmond County Bank Ballpark	40.645056	-74.076864	Baseball Stadium
4	St. George	40.644982	-74.079353	Shake Shack	40.643660	-74.075891	Burger Joint

Collecting, processing and inspecting the data

We can perform some sanity checks at this juncture by examining the number of venues in each neighborhood. We see that some have as few as 1 venue (not shown below). These are probably poor sites and if they end up in our final model we may have an issue

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Annadale	14	14	14	14	14	14
Arden Heights	4	4	4	4	4	4
Arlington	8	8	8	8	8	8
Arrochar	18	18	18	18	18	18
Bay Terrace	10	10	10	10	10	10
Bloomfield	4	4	4	4	4	4
Bulls Head	45	45	45	45	45	45
Butler Manor	6	6	6	6	6	6
Castleton Corners	19	19	19	19	19	19
Charleston	31	31	31	31	31	31
Chelsea	7	7	7	7	7	7
Clifton	21	21	21	21	21	21
Concord	7	7	7	7	7	7
Dongan Hills	22	22	22	22	22	22
Egbertville	4	4	4	4	4	4

Collecting, processing and inspecting the data

After creating a data frame from the json outputs we explored the data in more detail. It was determined that pizza places are spread out amongst the Island and there are 57 different pizza places. It was also noted that one venue did not have any venues and was thus dropped from analysis.

Pizza Place	57	57	57	57	57	57

Collecting, processing and inspecting the data

Next we evaluated venues at the neighborhood level and performed 1 hot encoding which then also allowed us to group our neighborhoods and venues and then sorted neighborhoods by their top 5 venues. Just on first glance there are a few neighborhoods where a pizza place is the most common venue. This gives us a great start!

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Annadale	Pizza Place	American Restaurant	Sports Bar	Food	Park
1	Arden Heights	Coffee Shop	Pizza Place	Bus Stop	Pharmacy	Yoga Studio
2	Arlington	Bus Stop	American Restaurant	Boat or Ferry	Home Service	Playground
3	Arrochar	Deli / Bodega	Italian Restaurant	Bus Stop	Supermarket	Mediterranean Restaurant
4	Bay Terrace	Supermarket	Insurance Office	Italian Restaurant	Home Service	Train Station
5	Bloomfield	Recreation Center	Bus Stop	Burger Joint	Theme Park	Diner
6	Bulls Head	Pizza Place	Bus Stop	Sandwich Place	Gift Shop	Pharmacy
7	Butler Manor	Baseball Field	Pool	Convenience Store	Bus Stop	Yoga Studio
8	Castleton Corners	Pizza Place	Bank	Ice Cream Shop	Go Kart Track	Sandwich Place
9	Charleston	Big Box Store	Coffee Shop	Cosmetics Shop	Irish Pub	Grocery Store

Building the Model - Clustering

Then we used our new grouped data frame to k-means cluster our data set and help determine a good venue for our beer spot and mapped the clusters on our map. On the first pass, we used k = 5. On review of the cluster groups we appeared to have good separation for neighborhoods that had venues of interest vs not (3 vs 4 for example).

Cluster 4

[139]:
```
staten_island_merged.loc[staten_island_merged['Cluster Labels'] == 3, staten_isla
```

t[139]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Comr
10	Port Ivory	Bus Station	Yoga Studio	Gas Station	French

Cluster 3

n [140]:
```
rged.loc[staten_island_merged['Cluster Labels'] == 2, staten_island_merged.columns[[2] + list(range(5, staten_island_m
```

ut[140]:

	Latitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	40.644982	Clothing Store	Park	Sporting Goods Shop	Italian Restaurant	Bar
2	40.626928	Pizza Place	Discount Store	Mexican Restaurant	Bank	Bar
3	40.615305	Italian Restaurant	Grocery Store	Cosmetics Shop	Discount Store	Café
4	40.631879	Bank	Coffee Shop	Italian Restaurant	Music Store	Diner
8	40.633669	Rental Car Location	Food	Donut Shop	Pizza Place	Bus Stop
11	40.613336	Pizza Place	Bank	Ice Cream Shop	Go Kart Track	Sandwich Place
12	40.594252	Pizza Place	Mobile Phone Shop	Bagel Shop	Chinese Restaurant	Coffee Shop
13	40.586314	Hotel	Deli / Bodega	Bowling Alley	Gym / Fitness Center	Spanish Restaurant
14	40.572572	Italian Restaurant	Pizza Place	Dessert Shop	Bakery	Sandwich Place
15	40.558462	Bar	Nightlife Spot	Lawyer	Playground	Farmers Market
16	40.549480	Italian Restaurant	Pizza Place	Bar	Dessert Shop	Liquor Store
17	40.542231	Pizza Place	Sushi Restaurant	Pharmacy	Italian Restaurant	Sandwich Place
18	40.538114	Pizza Place	American Restaurant	Sports Bar	Food	Park

Results Interpreting the Model - Clustering

Cluster 3 groups a few neighborhoods and the most common venues of interest

- Our most important item is pizza places, followed by other restaurant types, absence of other breweries and potentially public transportation.
- Group 3 is the only group of neighborhoods with pizza being the most common venue
  - Of the neighborhoods with pizza as the most popular venue only three have restaurants as the top two venues.
  - Eltington and Annandale are two of the neighborhoods that meet the most criteria
  - It is recommended that either of these to neighborhoods be selected with the nod going to Annandale as it is close to a park that has the potential to host the beer garden depending on zoning.
  - Additionally, both neighborhoods are in proximity thus regardless of the neighborhood selected, the same pizza places will likely be frequented

Discussion

- The tools of data science and ever-increasing data sets make decision making much easier.
- With the power of the data science approach we identified a business need, identified a problem solving approach and the data that would be required to solve the problem We iterated through the process reviewed our outputs and were able to deliver a business recommendation with the data and thought process to back up our recommendation.

Conclusion

In this project we worked with a new company called Beerizza, Inc to determine the best location for a new beer garden. New York City was picked by the client for it's pizza history. We used location data for New York City Neighborhoods and venue data from Foursquare to explore neighborhoods and venues that would meet our criteria - namely neighborhoods that have high density of pizza restaurants and second to pizza restaurants, other fast food places and an absence of Beer Gardens. An additional request was that the search be limited to Staten Island. We used k-means clustering to build our model. The model showed us that Eltingville and Annadale are both likely to be a good location for new beer garden based on the criteria that was given. The neighborhoods themselves have high density of pizza places, fast food restaurants and other activities and are neighboring so either choice would adequate based on the criteria given.