Barry's Bootcamp Seattle Expansion

USING K-MEAN TO FIND IDEAL LOCATIONS FOR DEVELOPMENT

KUAN YU BILLY CHEN

Introduction:

As millennials become the major consumer force, companies are tailoring their products and services to the millennial consumers.

Studies done by the Pew Research Center have found that millennials are making more personal improvement commitments than other generations before. According to the 2016 IHRSA Health Club Consumer report, millennial consumers prefer specialized classes over a traditional gym membership with workout machines.

Barry's Bootcamp is one of the most popular boutique workout studios that provides an intense specialized cross-training workout and a community with a cult-like following. As the profit margins continues to soar, Barry's is looking to expand in areas where there is a growing population and consumer wealth.

With Seattle's continuing tech-boom, it's no surprise that it is one of the fastest growing cities in the U.S., and with that more and more wealthy people are moving into the city. After opening its first successful branch in the South Lake Union neighborhood of Seattle, Barry's is looking to open more locations to accommodate the city's demand for specialized and community-based workouts.

After initial research, Barry's is looking to open a second location in the Capitol Hill neighborhood due to its concentration of wealthy millennial consumers. Barry's around the world have been primarily a single location in each major city, except for New York City, Los Angeles, San Francisco, and Chicago. San Francisco is like Seattle, both in the most popular industry (technology) and the size of the cities. I will be examining the four Barry's locations in San Francisco to find the surrounding shops, retailers, and venues that trend highly on Foursquare, and with this information, I will seek to find several possible locations to open the next Barry's Bootcamp branch in Seattle's Capitol Hill neighborhood.

Data Collection and Source:

I gathered location data: coordinates, business categories, and distance to the four Barry's locations in San Francisco.

I used the foursquare API to explore the four Barry's locations:

- Barry's FiDi location: 333 Bush St, San Francisco, CA
- Barry's Marina location: 2246 Lombard St, San Francisco, CA
- Barry's SoMa Location: 236 King St, San Francisco, CA
- Barry's Castro location: 2280 Market Street San Francisco, CA

Barry's Locations	latitude:	longitude:
FiDi	37.79046465	-122.4030519
Marina	37.7996758	-122.4385036
SoMa	37.77756473	-122.392825
Castro	37.7648157	-122.4327172

The column name and data type returned from the foursquare data:

Field Name	Data Type	Description
name	object	Venue name.
categories	object	Venue category.
address	object	Venue address.
cc	object	Venue country code.
city	object	Venue city.
country	object	Venue country.
crossStreet	object	The cross street from the provided location to explore.
distance	int64	The distance from the provided location to explore.
formattedAddress	object	Formatted address displayed on the venue's information page.
labeledLatLngs	object	Formatted latitude and longitude.
lat	float64	Latitude of the venue.
lng	float64	Longitude of the venue.
neighborhood	object	The neighborhood of the location of the venue.
postalCode	object	Venue postal code.
state	object	Venue state.
id	object	Venue unique foursquare identifier.

I used the GeoPy library to locate the latitude and longitude information and the foursquare explore API to obtain the popular locations around the four Barry's locations in San Francisco.

I also gathered identical location data in for Seattle's Capitol Hill neighborhood and the businesses in the area from this same source.

Methodology:

Once I collected the data for the four Barry's locations with the GeoPy library, the foursquare requests to explore the four locations returned a .json file, which I then mapped to dataframes of columns and rows with headers information.

Data Preprocessing:

The data cleanup for the dataframes mapped from the foursquare API output was simple – a few unnecessary columns were dropped from the final dataframe, including address, cc, city, country, crossStreet, formattedAddress, neighborhood, postalCode, state, and labeledLatLngs. I dropped these rows because they are not needed for my analysis given that the same information is retrievable through the coordinate's columns.

name	categories	distance	lat	lng	neighborhood	id
Cafe Okawari	Café	108	37.77817	-122.394	NaN	5cddccf91822230025707c7a
Petit Marlowe	Wine Bar	99	37.7781	-122.394	NaN	5941dd4ae2ead1688f35f615
Cento	Coffee Shop	114	37.77814	-122.394	SoMa	5ca6aa2db25fee0039757530
Tan Bella	Tanning Salon	69	37.77701	-122.393	NaN	4b09762ef964a520a41723e3
District	Wine Bar	111	37.77851	-122.393	NaN	461604b8f964a52045451fe3

Exploratory Analysis:

I combined the dataframes for the four locations into a single dataframe for the ease of analysis.

Popular venues around the 4 Barry's locations: 115.

Then, I employed visualization via folium to observe the venues around the four Barry's location to make sure the venues are within our parameters.

The goal of my research is to understand the best location for a new Barry's branch based on the type of venues around it. I used one-hot encoding to observe the most popular locations by turning each category into a binary column, with 0 meaning no and 1 meaning yes. For

example, a value 0 in the Russian Restaurant column means the venue is not a Russian restaurant.

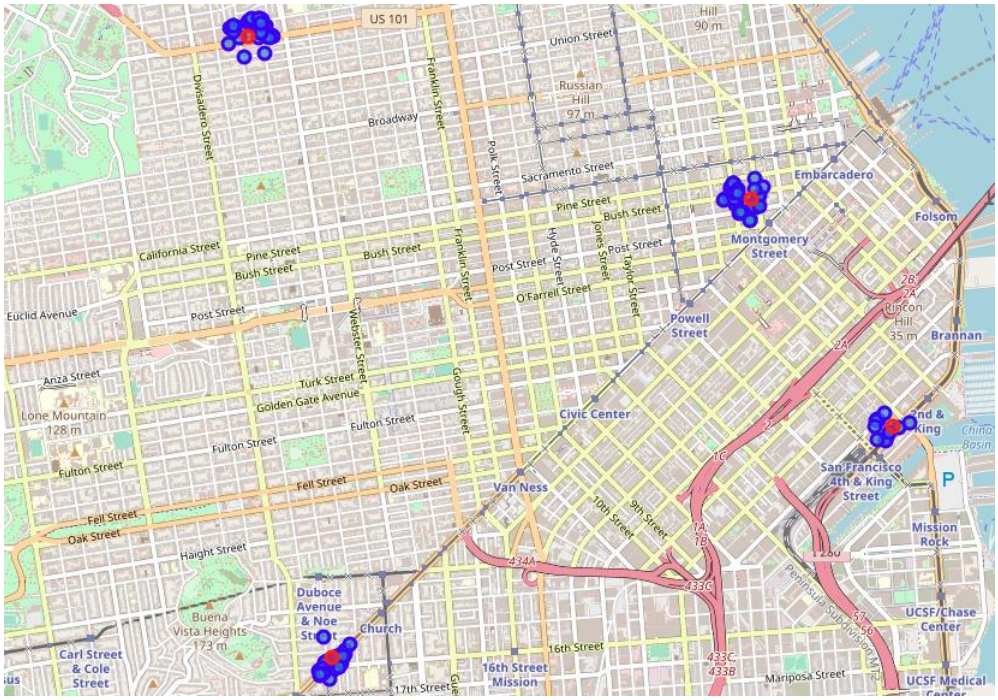
Bakery	Bank	Bubble Tea Shop	Burger Joint	Burmese Restaurant	Burrito Place	Café	...	Spanish Restaurant	Steakhouse	Sushi Restaurant	Tanning Salon	Thai Restaurant	Tiki Bar	Vietnamese Restaurant	Wine Bar	Women's Store	Yoga Studio
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

I ran a quick analysis on the one-hot encoded dataframe to understand the top locations around our Barry's San Francisco locations. After I found the top location, I then gathered the geodata for the Capitol Hill neighborhood in Seattle.

Top Categories	Number of Venus
French Restaurant	6
Gym	6
Coffee Shop	6
Sushi Restaurant	5
Sandwich Place	5
Burger Joint	4
Salad Place	3
Café	3

The Barry's locations in San Francisco are like 4 clusters with the Barry's branches as the centroids of each clusters. With that in mind, I chose to run K-means as the machine learning method to find the centroids for the new Capitol Hill location.

I used visualization with folium again to make sure the algorithm is producing locations based off the top venues.



Results:

Seattle - Capitol Hill Data:

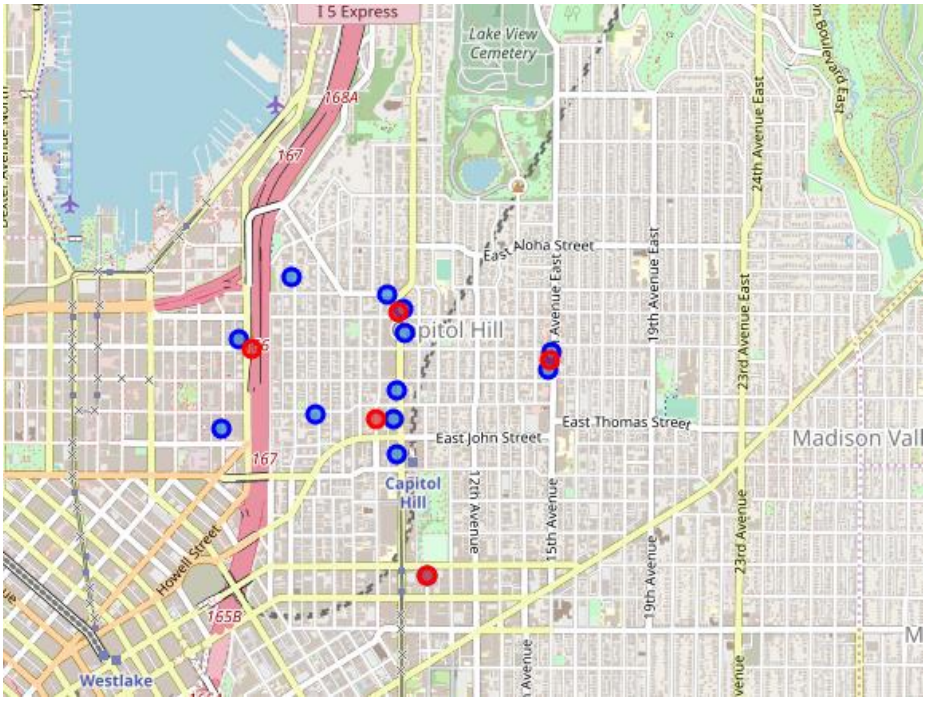
Venues matching the San Francisco top categories: 13

K-Means:

With K = 5, the machine learning algorithm returned 5 centroids for possible Barry's locations.

The centroid locations are:

Clus_km	lat	lng
0	47.620578	-122.32222
1	47.623128	-122.328936
2	47.622689	-122.312836
3	47.614886	-122.319493
4	47.624446	-122.321023



Discussion:

There are many factors at play when opening a new location of a company. Barry's is still looking to expand to more locations around the world, and by opening a second location in the same city, that means the city has potential consumer growth. Like San Francisco, Seattle is growing rapidly similarly to the city of San Francisco due to the continuing boom of technology companies in the area. Because of the similarities between the two cities and their primary industries, it is helpful to get an idea San Francisco Barry's location to pinpoint the popular locations and ideal location for Barry's new studio.

The data gathered from the San Francisco locations seemed to indicate a relationship between Barry's studios and high-end food and drink venues. The Capitol Hill neighborhood has plenty of such locations and putting the new Barry's location around these venues will hopefully produce a similar atmosphere and mood.

Simply using location data, of course, may not produce the best recommendation for a new branch location. We will also want to take in other factors, such as the ease of transportation and commute, rent pricing, average income, city ordinances, etc. As Barry's continues its search, it would be wise to bring in these other features and do additional analysis to identify the perfect location for the new Capitol Hill location.

An oddity I noted during my final clustering is one of the centroids does not seem to have any data points around it. This appears to be an outlier, and I believe with more data I could mitigate this issue.

After data preprocessing, I was also left with a limited number of datapoints for my project. If I were to repeat this project again, I'd like to bring in more data, possibly from other sources, to hopefully create better clusters.

Conclusion:

I believe the cluster centroids produced by the K-means algorithm are a good start in Barry's search for a second Seattle location. This has been an educational project that is applicable in other industries and businesses searching for physical locations to expand or open a new venue.

References:

- https://www.ft.com/content/194cd1c8-6583-11e8-a39d-4df188287fff
- https://www.npr.org/2017/06/04/531051473/the-millennial-obsession-with-self-care
- https://www.marketwatch.com/story/this-is-the-insane-amount-millennials-are-spending-on-fitness-2018-01-21
- https://www.seattletimes.com/pacific-nw-magazine/seattles-newest-boom-and-the-wealth-it-has-created-are-thrilling-and-terrifying/
- https://www.bestplaces.net/compare-cities/seattle_wa/san_francisco_ca/people
- https://developer.foursquare.com/docs
- https://www.barrys.com/studios/