

BOBA Drink exploration in Bay Area

Introduction: The problem

Bay Area in California knowing for attracting more talented individuals from all over the world. The tech-culture expanding which bringing many different people and culture from all over the world into the Bay area. Boba is a new trending topic that is very popular in the Asian community. Boba is a type of gelly balls called pearl that served either hot or cold. Boba milk tea is basically as the name describing it. Mixing milk with tea, and add boba. However, there are already so many boba places in the Bay area, and it is difficult for us to know which one or where is the best location to get boba. So my project goal is to locate the boba shops within the bay area. Find out which city has the most shop. The consumer can navigate to those shops or boba entrepreneur avoid a location with too many boba shops.

The Data

The solution for our problem would require the data consisted of: - Number of boba shop within the area - Boba shop rating and their location positions

```
In [1]: import pandas as pd
import io
import requests
import lxml
import numpy as np
import folium
from sklearn.cluster import KMeans
from pandas.io.json import json_normalize # tranform JSON file into a panda
from geopy.geocoders import Nominatim # convert an address into latitude and longitude

from IPython.display import Image
from IPython.core.display import HTML

import matplotlib.cm as cm
import matplotlib.colors as colors
```

```
In [*]: # Hidden cell
```

```
In [3]: # Specify the location of our data point
address = 'Sunnyvale, CA'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Bay area, California {}, {}'.format(l
```

The geograpical coordinate of Bay area, California 37.3688301, -122.0363496.

```
In [4]: # Generate a foursquare query for RESTAURANTS in the area
search_query = 'boba'
radius = 2000000
# url = 'https://api.foursquare.com/v2/venues/search?client_id={} & client_se

# Get our result return from the foursqaure dbase
results = requests.get(url).json()
v=results['response']['venues'] # extract the result from json file
dframe = json_normalize(v) # normalize our result with json normalize func
```

<ipython-input-4-9552f13398e9>:11: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead
dframe = json_normalize(v) # normalize our result with json normalize function

```
In [5]: # keep only columns that include venue name, and anything that is associate
filtered_columns = ['name', 'categories'] + [col for col in dframe.columns
dataframe_filtered = dframe.loc[:, filtered_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

# filter the category for each row
dataframe_filtered['categories'] = dataframe_filtered.apply(get_category_ty

# clean column names by keeping only last term
dataframe_filtered.columns = [column.split('.')[ -1] for column in dataframe

# dataframe_filtered
```

```
In [6]: df = dataframe_filtered[['name', 'categories', 'address', 'lat', 'lng', 'postalCode', 'formattedAddress']]
```

```
In [7]: df.head()
```

```
Out[7]:
```

	name	categories	address	lat	lng	postalCode	formattedAddress	
0	Boba Bar Teahouse & Eatery	Bubble Tea Shop	310 S 3rd St	37.332368	-121.884731	95112	[310 S 3rd St (at E San Carlos St), San Jose, ...	53cf2ef34
1	Boba Guys	Bubble Tea Shop	855 El Camino Real #120	37.438476	-122.159122	94301	[855 El Camino Real #120, Palo Alto, CA 94301]	5c3518fd2
2	Boba Drive	Bubble Tea Shop	NaN	37.403202	-122.008931	94089	[Sunnyvale, CA 94089]	5c561c556
3	Boba Guys	Bubble Tea Shop	1002 16th St	37.766448	-122.397042	94107	[1002 16th St (at Missouri St), San Francisco, ...	58d451f49
4	Boba	Coffee Shop	1710 N Milpitas Blvd	37.455524	-121.910233	95035	[1710 N Milpitas Blvd, Milpitas, CA 95035]	4f32367a1

Method of analysis

- Use of folium map to locate all of the boba shop within the bay area
- Use of clustering analysis to separate those the boba shop locations of clustering

```

In [138]: sanjose_map = folium.Map(location=[latitude, longitude], zoom_start=12)

incidents = folium.map.FeatureGroup()

for lat, lng, in zip(df.lat, df.lng):
    incidents.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            radius=5,
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )

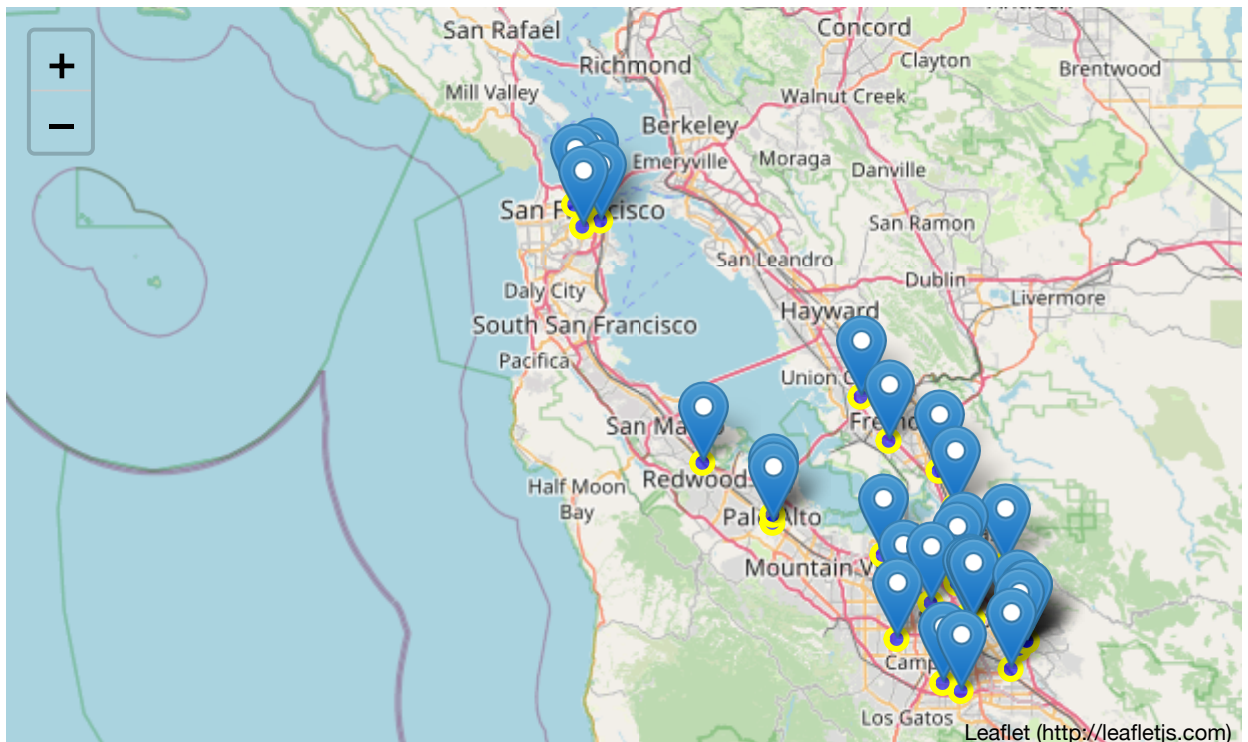
# add pop-up text to each marker on the map
latitudes = df.lat
longitudes = df.lng
label = df.categories

for lat, lng, label in zip(latitudes, longitudes, label):
    folium.Marker([lat, lng], popup=label).add_to(sanjose_map)

# add incidents to map
sanjose_map.add_child(incidents)

```

Out[138]:



```
In [139]: df.columns
```

```
Out[139]: Index(['name', 'categories', 'address', 'lat', 'lng', 'postalCode',
               'formattedAddress', 'id', 'distance', 'city'],
              dtype='object')
```

```
In [140]: # set number of clusters
kclusters = 4

group_cluster = df.drop(['name', 'categories', 'address', 'postalCode', 'form

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(group_cluster)

# check cluster labels generated for each row in the dataframe
kmeans.labels_

# Insert k cluster as column into df

df.insert(0, 'cluster label', kmeans.labels_)
```

```
In [ ]: plt.plot(range(1,Kcluster),mean_acc,'g')
```

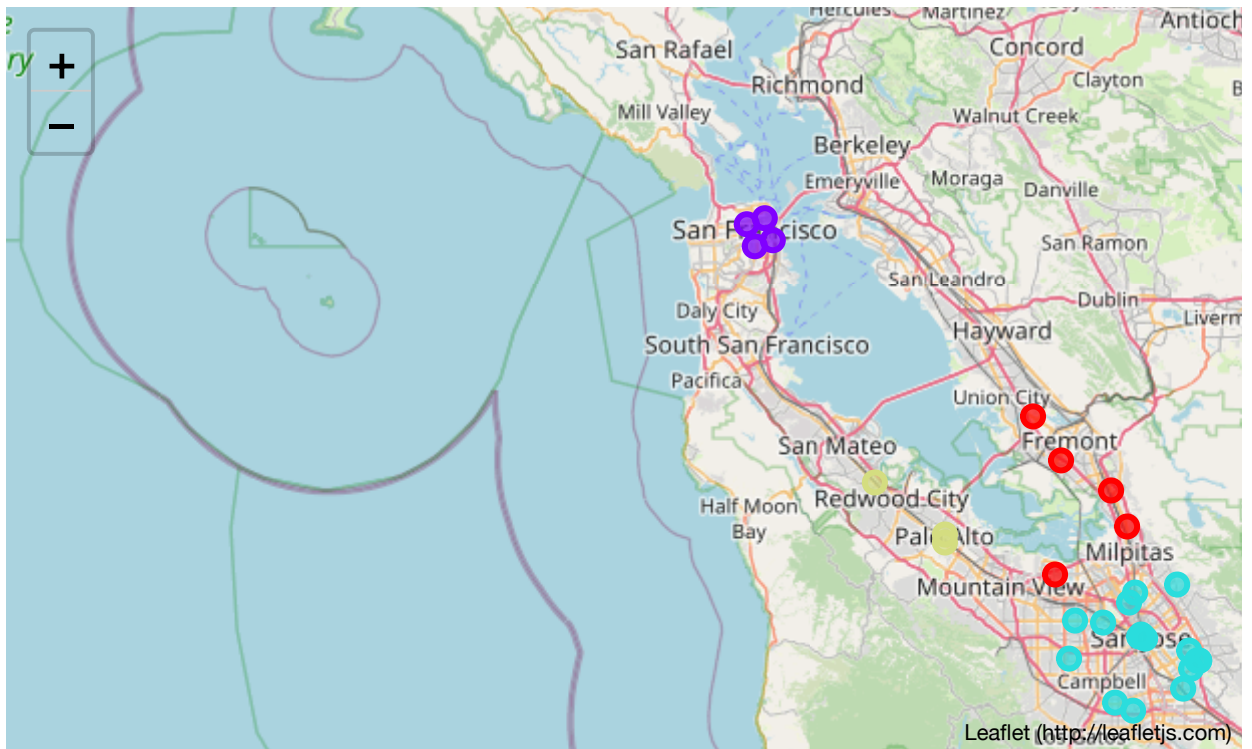
```
In [141]: map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(df['lat'], df['lng'], df['city'], df['cluster label']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)
```

```
In [142]: map_clusters
```

```
Out[142]:
```



```
In [150]: cluster_2 = df.loc[df['cluster label'] == 2]
cluster_2
```

```
Out[150]:
```

	cluster label	name	categories	address	lat	lng	postalCode	formattedAddress
0	2	Boba Bar Teahouse & Eatery	Bubble Tea Shop	310 S 3rd St	37.332368	-121.884731	95112	[310 S 3rd St (at San Carlos S San Jose,
6	2	Oh Boba	Bubble Tea Shop	NaN	37.350555	-121.944010	NaN	[Santa Clara, C.
7	2	Boba Pub	Coffee Shop	NaN	37.253480	-121.901566	NaN	[San Jose, C.
9	2	Bobaholics	Bubble Tea Shop	1055 E Brokaw Rd #40	37.384296	-121.897496	95131	[1055 E Brokaw F #40, San Jose, C 9513
10	2	Tiger Milk Boba	Bubble Tea Shop	72 N Almaden Ave	37.336055	-121.894399	95110	[72 N Almaden Ave, San Jose, C 9511
11	2	Bob & Karen's	None	NaN	37.310632	-121.988374	NaN	[Californi
12	2	WeBoba	Bubble Tea Shop	3030 El Camino Real	37.351721	-121.981472	95051	[3030 El Camir Real, Santa Clar CA 9505
13	2	Cafe Boba	Coffee Shop	110 E San Fernando St	37.335346	-121.886551	95112	[110 E S Fernando St, S Jose, CA 9511
14	2	Boba Fitt	Bubble Tea Shop	1051 E Capitol Expy	37.301310	-121.822820	95121	[1051 E Capit Expy, San Jos CA 9512
16	2	Boba Tea Express	Café	4100 Monterey Hwy	37.279468	-121.834097	95111	[4100 Monter Hwy (Marina), S Jose, CA 9511
18	2	Bob & Steve's Auto & Truck Repair	Automotive Shop	NaN	37.372134	-121.908033	NaN	[San Jose, C.
19	2	Oh My Boba	Bubble Tea Shop	NaN	37.392014	-121.842310	95132	[San Jose, C 9513
21	2	BOBATEANI	Bubble Tea Shop	75 E Santa Clara St	37.337212	-121.889275	95113	[75 E Santa Cla St, San Jose, C 9511
24	2	Simply Boba	Bubble Tea Shop	3005 Silver Creek Rd Ste 192	37.309512	-121.813647	95121	[3005 Silver Cree Rd Ste 192, S Jose, CA 95
25	2	Bob and Sue's	None	NaN	37.264193	-121.927576	NaN	[San Jose, C.
26	2	Pho 21 & Boba 21	Vietnamese Restaurant	NaN	37.319470	-121.823760	95122	[San Jose, C 9512

	cluster label	name	categories	address	lat	lng	postalCode	formattedAddress
27	2	OooH Boba Tea and Desserts	Dessert Shop	1783 E Capitol Expy	37.309635	-121.810071	95121	[1783 E Capit Expy, San Jos CA 9512
28	2	Joy Boba Tea	Bubble Tea Shop	1783 E Capitol Expy	37.309663	-121.810100	95121	[1783 E Capit Expy, San Jos CA 9512

Cluster result

In comparison with other county and city within the area, We can observe San Jose attracted majority of the boba locations


```

In [151]: sanjose_map = folium.Map(location=[latitude, longitude], zoom_start=12)

map_feature = folium.map.FeatureGroup()

for lat, lng, in zip(cluster_2.lat, cluster_2.lng):
    map_feature.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            radius=5,
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )

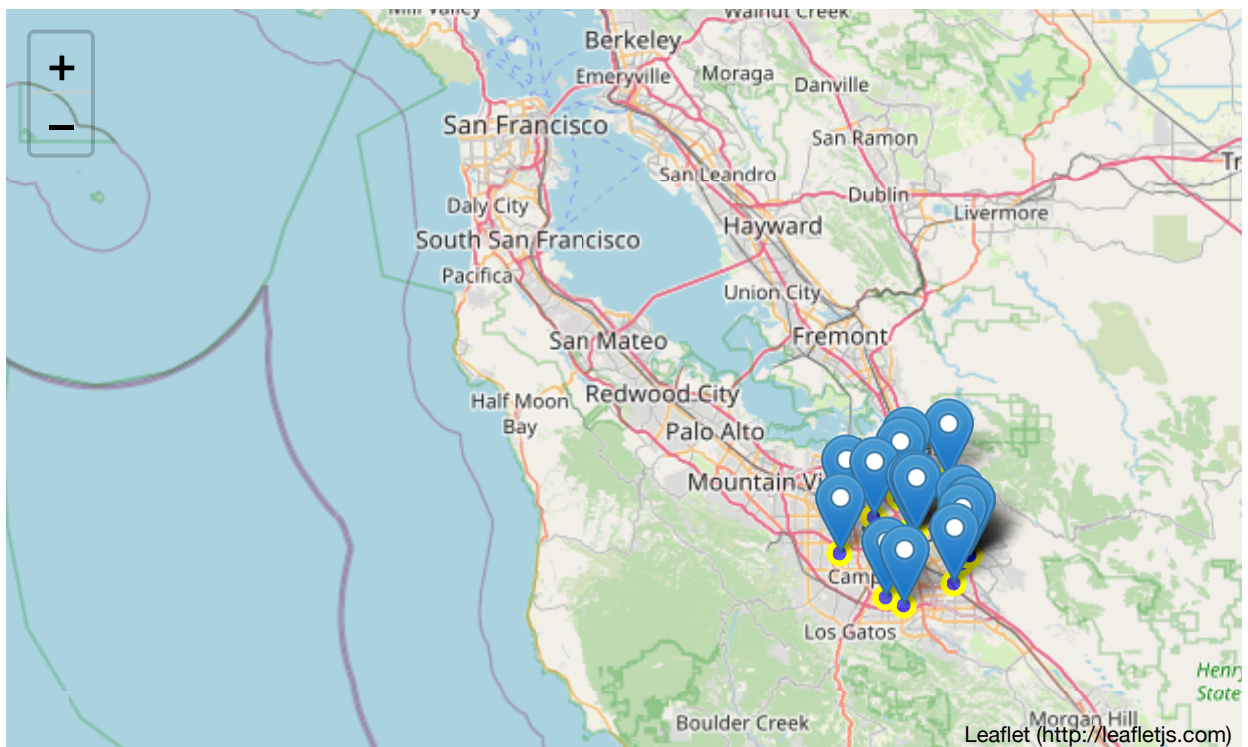
# add pop-up text to each marker on the map
latitudes = list(df.lat)
longitudes = list(df.lng)
labels = list(df.name)

for lat, lng, label in zip(cluster_2.lat, cluster_2.lng, df.categories):
    folium.Marker([lat, lng], popup=label).add_to(sanjose_map)

# add incidents to map
sanjose_map.add_child(map_feature)

```

Out[151]:



CLuster with most Boba shop and what are those shop

Tiger Milk Tea

```
In [152]: venue_id = '53cf2ef3498e1e5b6248251d'
          radius = 200

          # Get our result return from the foursqaure dbase
          results = requests.get(url).json()
          print('total like: ', results['response']['venue']['likes']['count'])
          print('price: ', results['response']['venue']['price'])

total like: 51
price: {'tier': 1, 'message': 'Cheap', 'currency': '$'}
```

```
In [153]: venue_id = '5ec8592ba823280008d502fb'
          radius = 200

          # Get our result return from the foursqaure dbase
          results = requests.get(url).json()
          print('total like: ', results['response']['venue']['likes']['count'])

total like: 0
```

The last one doesnt look too promising because it only have 2 rating

```
In [154]: boba_shop_in_SJ = cluster_2[cluster_2['city'] == 'San Jose'].city.count()
          boba_shop_in_SJ
```

Out[154]: 15

```
In [155]: boba_shop_in_Bay = df.city.count()
          boba_shop_in_Bay
```

Out[155]: 29

```
In [156]: percentage_boba_sj = (boba_shop_in_SJ/boba_shop_in_Bay) * 100
          percentage_boba_sj
```

Out[156]: 51.724137931034484

Discussion

By using folium and K-mean cluster analysis, we was able to figure out which region within the Bay area contain the most boba shop. We also check out two of those boba shop within the San Jose city which is a variable of cluster 0. We can see that Tiger Milk Tea have higher rating and also have more information about price.

Conclusion

There is a total of 29 boba shops in the Bay area, and 15 of these boba shops located in San Jose City. That is 52% if the boba shop locate in San Jose. SO, for those boba lover considering moving to the Bay Area, I would highly recommend for you to stay near San Jose.