Capstone Project week 5

February 25, 2021

1 1. Introduction

1.0.1 An important Gym Franchise wants to establish gyms in Central America. The have made studies that show that gyms are trending in Central America.

1.0.2 To start, they have decided to inaugurate two gyms in Costa Rica and then expand through the rest of the countries of Central America.

1.0.3 They want to know in wich cities it would be a great idea to stablish without having a great competition at the beginning.

1.0.4 So they decided to look for cities that are the most populated and at the same time, places where gyms are not among the five most common places.

2 2. Data

2.0.1 For this project I'm going to import data from the webpage: https://simplemaps.com/data/world-cities.

2.0.2 This webpage shows the information of many countries, including: city, latitude, longitude, country, country abbreviation, state or province and population, among others.

2.0.3 This webpage lets to download the information in a csv file. So I'm going to use pandas to read that csv file.

2.0.4 Then I'm going to filter the data to use only the information of the available cities from Costa Rica as requested for this project.

2.0.5 After that, I'm going to use Foursquare API to fetch all the information of the common venues in every city.

2.0.6 Then, I will use K-means method to clusterize the cities and focus only in the most populated cities that are close to each other.

2.0.7 Finally, in the most populated clusters, I will look for those two most populated cities where gyms are not among the five most common places, to recomend stablish the gyms in those cities.¶

3 3. Methodology

3.0.1 First, the information was imported from a webpage and then it was read with pandas.

3.0.2 Then the Foursquare API was used to extract information about the most common venues of the cities of interest.

3.0.3 After that, K-means method was used to clusterize the cities and focus only in the most populated cities that are close to each other.

3.0.4 Finally, the clusters were used to look for those two most populated cities where gyms are not among the five most common places, to recomend stablish the gyms in those cities.¶

Import Libraries

```
[1]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't
 ↪completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and
 ↪longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas
 ↪dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you
 ↪haven't completed the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - geopy


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    certifi-2020.12.5          |   py36h5fab9bb_1         143 KB  conda-forge
    geographiclib-1.50         |             py_0          34 KB  conda-forge
```

```
    geopy-2.1.0                |     pyhd3deb0d_0          64 KB  conda-forge
    openssl-1.1.1j             |       h7f98852_0         2.1 MB  conda-forge
    ------------------------------------------------------------
                                           Total:         2.4 MB

The following NEW packages will be INSTALLED:

  geographiclib      conda-forge/noarch::geographiclib-1.50-py_0
  geopy              conda-forge/noarch::geopy-2.1.0-pyhd3deb0d_0

The following packages will be UPDATED:

  certifi                            2020.12.5-py36h5fab9bb_0 -->
2020.12.5-py36h5fab9bb_1
  openssl                                   1.1.1i-h7f98852_0 -->
1.1.1j-h7f98852_0



Downloading and Extracting Packages
geopy-2.1.0          | 64 KB     | ##################################### | 100%
openssl-1.1.1j       | 2.1 MB    | ##################################### | 100%
certifi-2020.12.5    | 143 KB    | ##################################### | 100%
geographiclib-1.50   | 34 KB     | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Libraries imported.
```

Import Data

Import Data from webpage: https://simplemaps.com/data/world-cities

[91]:
```
df=pd.read_csv('worldcities.csv')
df.head()
```

[91]:
```
      city city_ascii      lat       lng      country iso2 iso3   admin_name  \
0    Tokyo      Tokyo  35.6897  139.6922        Japan   JP  JPN        Tōkyō
1  Jakarta    Jakarta  -6.2146  106.8451    Indonesia   ID  IDN      Jakarta
2    Delhi      Delhi  28.6600   77.2300        India   IN  IND        Delhi
3   Mumbai     Mumbai  18.9667   72.8333        India   IN  IND  Mahārāshtra
4   Manila     Manila  14.5958  120.9772  Philippines   PH  PHL       Manila

   capital  population          id
0  primary  37977000.0  1392685764
1  primary  34540000.0  1360771077
2    admin  29617000.0  1356872604
3    admin  23355000.0  1356226629
```

```
4  primary  23088000.0  1608618140
```

Eliminate unnecessary columns

[92]:
```python
df = df.drop(['iso2','iso3','id','city_ascii','capital'], 1)
df.head()
```

[92]:

	city	lat	lng	country	admin_name	population
0	Tokyo	35.6897	139.6922	Japan	Tōkyō	37977000.0
1	Jakarta	-6.2146	106.8451	Indonesia	Jakarta	34540000.0
2	Delhi	28.6600	77.2300	India	Delhi	29617000.0
3	Mumbai	18.9667	72.8333	India	Mahārāshtra	23355000.0
4	Manila	14.5958	120.9772	Philippines	Manila	23088000.0

Rename columns

[93]:
```python
df.columns = ['City','Latitude','Longitude','Country','Province','Population']
df.head()
```

[93]:

	City	Latitude	Longitude	Country	Province	Population
0	Tokyo	35.6897	139.6922	Japan	Tōkyō	37977000.0
1	Jakarta	-6.2146	106.8451	Indonesia	Jakarta	34540000.0
2	Delhi	28.6600	77.2300	India	Delhi	29617000.0
3	Mumbai	18.9667	72.8333	India	Mahārāshtra	23355000.0
4	Manila	14.5958	120.9772	Philippines	Manila	23088000.0

Filter the table to obtain the information of the available cities from Costa Rica

[94]:
```python
df2 = df[df['Country'] == 'Costa Rica'].reset_index(drop=True)
df2.head()
```

[94]:

	City	Latitude	Longitude	Country	Province	Population
0	San José	9.9333	-84.0833	Costa Rica	San José	288054.0
1	Cartago	9.8667	-83.9167	Costa Rica	Cartago	221733.0
2	Puerto Limón	10.0022	-83.0840	Costa Rica	Limón	61072.0
3	Liberia	10.6338	-85.4333	Costa Rica	Guanacaste	45380.0
4	Alajuela	10.0278	-84.2041	Costa Rica	Alajuela	42975.0

Examine number of cities available

[95]:
```python
df2.shape
```

[95]: (18, 6)

Get the geographical coordinates of San Jose, capital of Costa Rica

[96]:
```python
address = 'San Jose, CR'

geolocator = Nominatim(user_agent="cr_explorer")
```

```
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinates of Costa Rica are {}, {}.'.format(latitude,␣
↪longitude))
```

```
The geographical coordinates of Costa Rica are 9.9325427, -84.0795782.
```

Create map of Costa Rica using latitude and longitude values

[97]:
```
map_cr = folium.Map(location=[latitude, longitude], zoom_start=9)

# add markers to map
for lat, lng, label in zip(df2['Latitude'], df2['Longitude'], df2['City']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_cr)

map_cr
```

[97]:
```
<folium.folium.Map at 0x7fe019917f60>
```

Define Foursquare Credentials and Version

[98]:
```
CLIENT_ID = '5OQC1AURHZFISWL3IKHBLWVJWLMZVE4IQ2LFERPPGEV4SJXG' # your␣
↪Foursquare ID
CLIENT_SECRET = 'GAPAK1LZP3W4ZLAYDGHHYJD5KPLATRVLPGXHYQVNPHCKACSO' # your␣
↪Foursquare Secret
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: 5OQC1AURHZFISWL3IKHBLWVJWLMZVE4IQ2LFERPPGEV4SJXG
CLIENT_SECRET:GAPAK1LZP3W4ZLAYDGHHYJD5KPLATRVLPGXHYQVNPHCKACSO
```

Function to extract the information from all the cities

[99]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?
↪&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item␣
↪in venue_list])
    nearby_venues.columns = ['City',
                  'City Latitude',
                  'City Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Code to run the above function on each city

[100]:
```
CR_venues = getNearbyVenues(names=df2['City'],
                                   latitudes=df2['Latitude'],
                                   longitudes=df2['Longitude']
```

```
                                   )
```

```
San José
Cartago
Puerto Limón
Liberia
Alajuela
Puntarenas
San Juan
Heredia
Santa Ana
Buenos Aires
Quesada
Cañas
El Roble
Santiago
Sixaola
La Cruz
Golfito
Ciudad Cortés
```

Check the size of the resulting dataframe

[101]:
```python
print(CR_venues.shape)
CR_venues.head()
```

```
(255, 7)
```

[101]:
```
       City  City Latitude  City Longitude                        Venue  \
0  San José         9.9333        -84.0833    La Sorbetera de Lolo Mora
1  San José         9.9333        -84.0833                Rincón Retana
2  San José         9.9333        -84.0833                  El Tostador
3  San José         9.9333        -84.0833  Mercado Central de San José
4  San José         9.9333        -84.0833                    Soda Tala

   Venue Latitude  Venue Longitude  Venue Category
0        9.934467       -84.081841  Ice Cream Shop
1        9.934561       -84.082022  Sandwich Place
2        9.934511       -84.083321            Café
3        9.934492       -84.081830          Market
4        9.934671       -84.081785      Restaurant
```

Check how many venues were returned for each city

[102]:
```python
CR_venues.groupby('City').count()
```

[102]:
```
               City Latitude  City Longitude  Venue  Venue Latitude  \
City
```

```
Alajuela                       4                4      4                4
Buenos Aires                   4                4      4                4
Cartago                       29               29     29               29
Cañas                          2                2      2                2
Ciudad Cortés                  6                6      6                6
El Roble                       5                5      5                5
Heredia                       28               28     28               28
La Cruz                        4                4      4                4
Liberia                       15               15     15               15
Puerto Limón                   1                1      1                1
Puntarenas                    50               50     50               50
Quesada                        2                2      2                2
San José                      42               42     42               42
San Juan                      38               38     38               38
Santa Ana                     24               24     24               24
Sixaola                        1                1      1                1

               Venue Longitude  Venue Category
City
Alajuela                     4               4
Buenos Aires                 4               4
Cartago                     29              29
Cañas                        2               2
Ciudad Cortés                6               6
El Roble                     5               5
Heredia                     28              28
La Cruz                      4               4
Liberia                     15              15
Puerto Limón                 1               1
Puntarenas                  50              50
Quesada                      2               2
San José                    42              42
San Juan                    38              38
Santa Ana                   24              24
Sixaola                      1               1
```

Let's find out how many unique categories can be curated from all the returned venues

[103]:
```
print('There are {} uniques categories.'.format(len(CR_venues['Venue Category'].
 ↪unique())))
```

```
There are 88 uniques categories.
```

Analyze Each Neighborhood

[104]:
```
# one hot encoding
CR_onehot = pd.get_dummies(CR_venues[['Venue Category']], prefix="",␣
 ↪prefix_sep="")
```

```python
# add city column back to dataframe
CR_onehot['City'] = CR_venues['City']

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

CR_onehot.head()
```

[104]:

```
        City  American Restaurant  Art Gallery  Arts & Crafts Store  \
0  San José                    0            0                    0
1  San José                    0            0                    0
2  San José                    0            0                    0
3  San José                    0            0                    0
4  San José                    0            0                    0

   Asian Restaurant  Athletics & Sports  Auto Garage  Bakery  Bar  \
0                 0                   0            0       0    0
1                 0                   0            0       0    0
2                 0                   0            0       0    0
3                 0                   0            0       0    0
4                 0                   0            0       0    0

   Bed & Breakfast  Beer Garden  Big Box Store  Bistro  Boutique  Boxing Gym  \
0                0            0              0       0         0           0
1                0            0              0       0         0           0
2                0            0              0       0         0           0
3                0            0              0       0         0           0
4                0            0              0       0         0           0

   Brewery  Burger Joint  Burrito Place  Bus Station  Bus Stop  Café  \
0        0             0              0            0         0     0
1        0             0              0            0         0     0
2        0             0              0            0         0     1
3        0             0              0            0         0     0
4        0             0              0            0         0     0

   Caribbean Restaurant  Chinese Restaurant  Church  Coffee Shop  \
0                     0                   0       0            0
1                     0                   0       0            0
2                     0                   0       0            0
3                     0                   0       0            0
4                     0                   0       0            0

   Convenience Store  Creperie  Deli / Bodega  Department Store  Dessert Shop  \
0                  0         0              0                 0             0
```

```
1                  0         0              0                  0             0
2                  0         0              0                  0             0
3                  0         0              0                  0             0
4                  0         0              0                  0             0

   Diner  Electronics Store  Event Space  Falafel Restaurant  \
0      0                  0            0                   0
1      0                  0            0                   0
2      0                  0            0                   0
3      0                  0            0                   0
4      0                  0            0                   0

   Fast Food Restaurant  Food  Food & Drink Shop  Fried Chicken Joint  \
0                     0     0                  0                    0
1                     0     0                  0                    0
2                     0     0                  0                    0
3                     0     0                  0                    0
4                     0     0                  0                    0

   Gift Shop  Grocery Store  Gym  Gym / Fitness Center  Gymnastics Gym  \
0          0              0    0                     0               0
1          0              0    0                     0               0
2          0              0    0                     0               0
3          0              0    0                     0               0
4          0              0    0                     0               0

   Harbor / Marina  Historic Site  Hotel  Ice Cream Shop  Italian Restaurant  \
0                0              0      0               1                   0
1                0              0      0               0                   0
2                0              0      0               0                   0
3                0              0      0               0                   0
4                0              0      0               0                   0

   Japanese Restaurant  Juice Bar  Karaoke Bar  Latin American Restaurant  \
0                    0          0            0                          0
1                    0          0            0                          0
2                    0          0            0                          0
3                    0          0            0                          0
4                    0          0            0                          0

   Market  Mediterranean Restaurant  Mexican Restaurant  Museum  Music Venue  \
0       0                         0                   0       0            0
1       0                         0                   0       0            0
2       0                         0                   0       0            0
3       1                         0                   0       0            0
4       0                         0                   0       0            0
```

	Other Repair Shop	Park	Peruvian Restaurant	Pet Store	Pharmacy
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Pizza Place	Plaza	Pool	Pub	Racetrack	Restaurant	Salad Place
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0

	Sandwich Place	Seafood Restaurant	Shoe Store	Shop & Service
0	0	0	0	0
1	1	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Shopping Mall	Shopping Plaza	Snack Place	Soccer Stadium	Sports Bar
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Steakhouse	Supermarket	Sushi Restaurant	Taco Place	Theater
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Train Station	Tree	Vegetarian / Vegan Restaurant	Video Store
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Wings Joint	Women's Store
0	0	0
1	0	0
2	0	0
3	0	0

```
4              0                0
```

Let's examine the new dataframe size

[105]:
```
CR_onehot.shape
```

[105]:
```
(255, 89)
```

Let's group rows by City and by taking the mean of the frequency of occurrence of each category

[106]:
```
CR_grouped = CR_onehot.groupby('City').mean().reset_index()
CR_grouped
```

[106]:

	City	American Restaurant	Art Gallery	Arts & Crafts Store	\
0	Alajuela	0.00	0.000000	0.000000	
1	Buenos Aires	0.00	0.000000	0.000000	
2	Cartago	0.00	0.000000	0.000000	
3	Cañas	0.00	0.000000	0.000000	
4	Ciudad Cortés	0.00	0.000000	0.000000	
5	El Roble	0.00	0.000000	0.000000	
6	Heredia	0.00	0.035714	0.000000	
7	La Cruz	0.00	0.000000	0.000000	
8	Liberia	0.00	0.000000	0.000000	
9	Puerto Limón	0.00	0.000000	0.000000	
10	Puntarenas	0.02	0.020000	0.000000	
11	Quesada	0.00	0.000000	0.000000	
12	San José	0.00	0.000000	0.000000	
13	San Juan	0.00	0.000000	0.000000	
14	Santa Ana	0.00	0.041667	0.041667	
15	Sixaola	0.00	0.000000	0.000000	

	Asian Restaurant	Athletics & Sports	Auto Garage	Bakery	Bar	\
0	0.000000	0.000000	0.000000	0.000000	0.250000	
1	0.000000	0.000000	0.000000	0.000000	0.000000	
2	0.034483	0.000000	0.000000	0.068966	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.000000	
6	0.000000	0.035714	0.000000	0.035714	0.071429	
7	0.000000	0.000000	0.000000	0.000000	0.500000	
8	0.000000	0.000000	0.000000	0.000000	0.133333	
9	0.000000	0.000000	0.000000	0.000000	0.000000	
10	0.020000	0.000000	0.000000	0.020000	0.000000	
11	0.000000	0.000000	0.000000	0.000000	0.000000	
12	0.000000	0.000000	0.000000	0.023810	0.023810	
13	0.026316	0.000000	0.000000	0.000000	0.052632	
14	0.000000	0.041667	0.041667	0.000000	0.041667	

15	0.000000	0.000000	0.000000	0.000000	0.000000

	Bed & Breakfast	Beer Garden	Big Box Store	Bistro	Boutique	\
0	0.000000	0.000000	0.0	0.000000	0.00000	
1	0.000000	0.000000	0.0	0.000000	0.00000	
2	0.000000	0.000000	0.0	0.000000	0.00000	
3	0.000000	0.000000	0.0	0.000000	0.00000	
4	0.000000	0.000000	0.0	0.000000	0.00000	
5	0.000000	0.000000	0.2	0.000000	0.00000	
6	0.000000	0.000000	0.0	0.035714	0.00000	
7	0.000000	0.000000	0.0	0.000000	0.00000	
8	0.133333	0.000000	0.0	0.000000	0.00000	
9	0.000000	0.000000	0.0	0.000000	0.00000	
10	0.000000	0.040000	0.0	0.000000	0.02000	
11	0.000000	0.000000	0.0	0.000000	0.00000	
12	0.000000	0.000000	0.0	0.000000	0.02381	
13	0.000000	0.000000	0.0	0.000000	0.00000	
14	0.000000	0.041667	0.0	0.000000	0.00000	
15	0.000000	0.000000	0.0	0.000000	0.00000	

	Boxing Gym	Brewery	Burger Joint	Burrito Place	Bus Station	Bus Stop	\
0	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
1	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
2	0.034483	0.000000	0.034483	0.00000	0.000000	0.0	
3	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
4	0.000000	0.166667	0.000000	0.00000	0.166667	0.0	
5	0.000000	0.000000	0.000000	0.00000	0.000000	0.4	
6	0.000000	0.000000	0.035714	0.00000	0.000000	0.0	
7	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
8	0.000000	0.000000	0.066667	0.00000	0.000000	0.0	
9	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
10	0.000000	0.000000	0.020000	0.00000	0.000000	0.0	
11	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	
12	0.000000	0.000000	0.023810	0.02381	0.000000	0.0	
13	0.000000	0.000000	0.052632	0.00000	0.000000	0.0	
14	0.000000	0.041667	0.000000	0.00000	0.041667	0.0	
15	0.000000	0.000000	0.000000	0.00000	1.000000	0.0	

	Café	Caribbean Restaurant	Chinese Restaurant	Church	Coffee Shop	\
0	0.000000	0.000000	0.00	0.00000	0.000000	
1	0.000000	0.000000	0.00	0.00000	0.000000	
2	0.034483	0.034483	0.00	0.00000	0.034483	
3	0.000000	0.000000	0.00	0.00000	0.000000	
4	0.000000	0.000000	0.00	0.00000	0.000000	
5	0.000000	0.000000	0.00	0.00000	0.000000	
6	0.035714	0.035714	0.00	0.00000	0.071429	
7	0.000000	0.000000	0.00	0.00000	0.000000	

8	0.000000	0.000000	0.20	0.00000	0.000000
9	0.000000	0.000000	0.00	0.00000	0.000000
10	0.000000	0.000000	0.06	0.00000	0.000000
11	0.000000	0.000000	0.00	0.00000	0.000000
12	0.023810	0.000000	0.00	0.02381	0.095238
13	0.000000	0.000000	0.00	0.00000	0.000000
14	0.000000	0.000000	0.00	0.00000	0.000000
15	0.000000	0.000000	0.00	0.00000	0.000000

	Convenience Store	Creperie	Deli / Bodega	Department Store	\
0	0.000000	0.00	0.00	0.00	
1	0.000000	0.00	0.00	0.00	
2	0.000000	0.00	0.00	0.00	
3	0.000000	0.00	0.00	0.00	
4	0.000000	0.00	0.00	0.00	
5	0.000000	0.00	0.00	0.00	
6	0.000000	0.00	0.00	0.00	
7	0.000000	0.00	0.00	0.00	
8	0.000000	0.00	0.00	0.00	
9	0.000000	0.00	0.00	0.00	
10	0.000000	0.02	0.02	0.02	
11	0.000000	0.00	0.00	0.00	
12	0.047619	0.00	0.00	0.00	
13	0.000000	0.00	0.00	0.00	
14	0.000000	0.00	0.00	0.00	
15	0.000000	0.00	0.00	0.00	

	Dessert Shop	Diner	Electronics Store	Event Space	Falafel Restaurant	\
0	0.000000	0.00000	0.000000	0.00000	0.00000	
1	0.000000	0.00000	0.000000	0.00000	0.00000	
2	0.000000	0.00000	0.034483	0.00000	0.00000	
3	0.000000	0.00000	0.000000	0.00000	0.00000	
4	0.000000	0.00000	0.000000	0.00000	0.00000	
5	0.000000	0.00000	0.000000	0.00000	0.00000	
6	0.035714	0.00000	0.000000	0.00000	0.00000	
7	0.000000	0.00000	0.000000	0.00000	0.00000	
8	0.000000	0.00000	0.000000	0.00000	0.00000	
9	0.000000	0.00000	0.000000	0.00000	0.00000	
10	0.000000	0.04000	0.000000	0.00000	0.00000	
11	0.000000	0.00000	0.000000	0.00000	0.00000	
12	0.000000	0.02381	0.000000	0.02381	0.02381	
13	0.000000	0.00000	0.000000	0.00000	0.00000	
14	0.000000	0.00000	0.000000	0.00000	0.00000	
15	0.000000	0.00000	0.000000	0.00000	0.00000	

	Fast Food Restaurant	Food	Food & Drink Shop	Fried Chicken Joint	\
0	0.000000	0.00	0.25	0.000000	

1	0.000000	0.00	0.00	0.000000
2	0.000000	0.00	0.00	0.000000
3	0.000000	0.00	0.00	0.000000
4	0.000000	0.00	0.00	0.000000
5	0.000000	0.00	0.00	0.000000
6	0.035714	0.00	0.00	0.000000
7	0.000000	0.00	0.00	0.000000
8	0.000000	0.00	0.00	0.000000
9	0.000000	0.00	0.00	0.000000
10	0.060000	0.02	0.00	0.000000
11	0.000000	0.00	0.00	0.000000
12	0.071429	0.00	0.00	0.000000
13	0.052632	0.00	0.00	0.026316
14	0.000000	0.00	0.00	0.000000
15	0.000000	0.00	0.00	0.000000

	Gift Shop	Grocery Store	Gym	Gym / Fitness Center	Gymnastics Gym	\
0	0.000000	0.000000	0.000000	0.000000	0.0	
1	0.000000	0.000000	0.000000	0.000000	0.0	
2	0.000000	0.034483	0.034483	0.000000	0.0	
3	0.000000	0.000000	0.000000	0.000000	0.0	
4	0.000000	0.166667	0.000000	0.000000	0.0	
5	0.000000	0.000000	0.000000	0.000000	0.0	
6	0.000000	0.000000	0.107143	0.035714	0.0	
7	0.000000	0.000000	0.000000	0.000000	0.0	
8	0.000000	0.000000	0.000000	0.000000	0.0	
9	0.000000	0.000000	0.000000	0.000000	0.0	
10	0.000000	0.000000	0.020000	0.000000	0.0	
11	0.000000	0.000000	0.000000	0.000000	0.5	
12	0.000000	0.023810	0.000000	0.023810	0.0	
13	0.026316	0.000000	0.052632	0.026316	0.0	
14	0.000000	0.000000	0.000000	0.000000	0.0	
15	0.000000	0.000000	0.000000	0.000000	0.0	

	Harbor / Marina	Historic Site	Hotel	Ice Cream Shop	\
0	0.00	0.000000	0.000000	0.000000	
1	0.00	0.000000	0.000000	0.000000	
2	0.00	0.034483	0.000000	0.034483	
3	0.00	0.000000	1.000000	0.000000	
4	0.00	0.000000	0.000000	0.000000	
5	0.00	0.000000	0.000000	0.000000	
6	0.00	0.035714	0.000000	0.107143	
7	0.00	0.000000	0.250000	0.000000	
8	0.00	0.000000	0.133333	0.000000	
9	1.00	0.000000	0.000000	0.000000	
10	0.02	0.020000	0.000000	0.100000	
11	0.00	0.000000	0.000000	0.000000	

12	0.00	0.000000	0.000000	0.047619
13	0.00	0.000000	0.000000	0.000000
14	0.00	0.000000	0.000000	0.041667
15	0.00	0.000000	0.000000	0.000000

	Italian Restaurant	Japanese Restaurant	Juice Bar	Karaoke Bar \
0	0.000000	0.000000	0.000000	0.00
1	0.000000	0.000000	0.250000	0.25
2	0.000000	0.000000	0.034483	0.00
3	0.000000	0.000000	0.000000	0.00
4	0.000000	0.000000	0.000000	0.00
5	0.000000	0.000000	0.000000	0.00
6	0.000000	0.000000	0.000000	0.00
7	0.000000	0.000000	0.000000	0.00
8	0.066667	0.000000	0.000000	0.00
9	0.000000	0.000000	0.000000	0.00
10	0.000000	0.000000	0.000000	0.00
11	0.000000	0.000000	0.000000	0.00
12	0.000000	0.000000	0.023810	0.00
13	0.026316	0.026316	0.026316	0.00
14	0.041667	0.000000	0.000000	0.00
15	0.000000	0.000000	0.000000	0.00

	Latin American Restaurant	Market	Mediterranean Restaurant \
0	0.000000	0.000000	0.000000
1	0.250000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.000000	0.035714	0.000000
7	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.020000	0.020000	0.000000
11	0.000000	0.500000	0.000000
12	0.047619	0.023810	0.000000
13	0.078947	0.000000	0.000000
14	0.000000	0.041667	0.041667
15	0.000000	0.000000	0.000000

	Mexican Restaurant	Museum	Music Venue	Other Repair Shop	Park \
0	0.000000	0.000000	0.000000	0.0	0.000000
1	0.000000	0.000000	0.000000	0.0	0.000000
2	0.068966	0.034483	0.000000	0.0	0.000000
3	0.000000	0.000000	0.000000	0.0	0.000000
4	0.000000	0.000000	0.000000	0.0	0.166667

5	0.000000	0.000000	0.000000	0.2	0.000000
6	0.000000	0.000000	0.035714	0.0	0.000000
7	0.000000	0.000000	0.000000	0.0	0.000000
8	0.000000	0.000000	0.000000	0.0	0.000000
9	0.000000	0.000000	0.000000	0.0	0.000000
10	0.000000	0.000000	0.020000	0.0	0.040000
11	0.000000	0.000000	0.000000	0.0	0.000000
12	0.000000	0.000000	0.000000	0.0	0.023810
13	0.026316	0.000000	0.000000	0.0	0.026316
14	0.000000	0.000000	0.000000	0.0	0.000000
15	0.000000	0.000000	0.000000	0.0	0.000000

	Peruvian Restaurant	Pet Store	Pharmacy	Pizza Place	Plaza	Pool	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.25	
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
2	0.000000	0.000000	0.034483	0.103448	0.034483	0.00	
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
4	0.000000	0.000000	0.166667	0.000000	0.000000	0.00	
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
6	0.000000	0.000000	0.000000	0.035714	0.000000	0.00	
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
8	0.000000	0.000000	0.000000	0.066667	0.000000	0.00	
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
10	0.000000	0.000000	0.000000	0.020000	0.000000	0.00	
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
12	0.023810	0.000000	0.023810	0.023810	0.000000	0.00	
13	0.000000	0.052632	0.000000	0.026316	0.000000	0.00	
14	0.041667	0.041667	0.000000	0.000000	0.000000	0.00	
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	

	Pub	Racetrack	Restaurant	Salad Place	Sandwich Place	\
0	0.000000	0.00	0.250000	0.000000	0.000000	
1	0.000000	0.00	0.000000	0.000000	0.000000	
2	0.034483	0.00	0.034483	0.034483	0.034483	
3	0.000000	0.00	0.000000	0.000000	0.000000	
4	0.000000	0.00	0.000000	0.000000	0.000000	
5	0.000000	0.00	0.000000	0.000000	0.000000	
6	0.035714	0.00	0.000000	0.000000	0.000000	
7	0.000000	0.00	0.250000	0.000000	0.000000	
8	0.000000	0.00	0.133333	0.000000	0.000000	
9	0.000000	0.00	0.000000	0.000000	0.000000	
10	0.020000	0.02	0.060000	0.000000	0.000000	
11	0.000000	0.00	0.000000	0.000000	0.000000	
12	0.023810	0.00	0.071429	0.000000	0.095238	
13	0.078947	0.00	0.052632	0.000000	0.078947	
14	0.000000	0.00	0.166667	0.000000	0.000000	
15	0.000000	0.00	0.000000	0.000000	0.000000	

	Seafood Restaurant	Shoe Store	Shop & Service	Shopping Mall	\
0	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	
2	0.034483	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	0.000000	
6	0.000000	0.000000	0.000000	0.000000	
7	0.000000	0.000000	0.000000	0.000000	
8	0.000000	0.000000	0.000000	0.000000	
9	0.000000	0.000000	0.000000	0.000000	
10	0.140000	0.000000	0.000000	0.000000	
11	0.000000	0.000000	0.000000	0.000000	
12	0.000000	0.000000	0.000000	0.000000	
13	0.026316	0.026316	0.000000	0.026316	
14	0.000000	0.000000	0.041667	0.000000	
15	0.000000	0.000000	0.000000	0.000000	

	Shopping Plaza	Snack Place	Soccer Stadium	Sports Bar	Steakhouse	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	0.250000	
2	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.166667	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.000000	
6	0.000000	0.000000	0.000000	0.000000	0.000000	
7	0.000000	0.000000	0.000000	0.000000	0.000000	
8	0.000000	0.000000	0.066667	0.000000	0.000000	
9	0.000000	0.000000	0.000000	0.000000	0.000000	
10	0.020000	0.020000	0.000000	0.000000	0.000000	
11	0.000000	0.000000	0.000000	0.000000	0.000000	
12	0.000000	0.047619	0.000000	0.000000	0.000000	
13	0.000000	0.000000	0.000000	0.000000	0.000000	
14	0.041667	0.000000	0.000000	0.041667	0.083333	
15	0.000000	0.000000	0.000000	0.000000	0.000000	

	Supermarket	Sushi Restaurant	Taco Place	Theater	Train Station	Tree	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
2	0.000000	0.000000	0.034483	0.000000	0.000000	0.0	
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.2	
6	0.000000	0.000000	0.035714	0.000000	0.035714	0.0	
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	

```
9     0.000000         0.000000    0.000000  0.000000        0.000000   0.0
10    0.020000         0.000000    0.000000  0.000000        0.000000   0.0
11    0.000000         0.000000    0.000000  0.000000        0.000000   0.0
12    0.000000         0.000000    0.000000  0.023810        0.000000   0.0
13    0.026316         0.026316    0.000000  0.026316        0.000000   0.0
14    0.000000         0.000000    0.041667  0.000000        0.000000   0.0
15    0.000000         0.000000    0.000000  0.000000        0.000000   0.0

    Vegetarian / Vegan Restaurant  Video Store  Wings Joint  Women's Store
0                        0.000000         0.00     0.000000       0.000000
1                        0.000000         0.00     0.000000       0.000000
2                        0.000000         0.00     0.000000       0.034483
3                        0.000000         0.00     0.000000       0.000000
4                        0.000000         0.00     0.000000       0.000000
5                        0.000000         0.00     0.000000       0.000000
6                        0.035714         0.00     0.000000       0.000000
7                        0.000000         0.00     0.000000       0.000000
8                        0.000000         0.00     0.000000       0.000000
9                        0.000000         0.00     0.000000       0.000000
10                       0.000000         0.02     0.000000       0.000000
11                       0.000000         0.00     0.000000       0.000000
12                       0.000000         0.00     0.000000       0.000000
13                       0.000000         0.00     0.026316       0.000000
14                       0.000000         0.00     0.000000       0.000000
15                       0.000000         0.00     0.000000       0.000000
```

Let's print each neighborhood along with the top 5 most common venues

[107]:
```
num_top_venues = 5

for City in CR_grouped['City']:
    print("----"+City+"----")
    temp = CR_grouped[CR_grouped['City'] == City].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).
 ↪head(num_top_venues))
    print('\n')
```

```
----Alajuela----
               venue  freq
0  Food & Drink Shop  0.25
1               Pool  0.25
2                Bar  0.25
3         Restaurant  0.25
4  Other Repair Shop  0.00
```

```
----Buenos Aires----
                        venue  freq
0                  Steakhouse  0.25
1                   Juice Bar  0.25
2                 Karaoke Bar  0.25
3  Latin American Restaurant  0.25
4           Other Repair Shop  0.00


----Cartago----
                venue  freq
0         Pizza Place  0.10
1  Mexican Restaurant  0.07
2              Bakery  0.07
3       Women's Store  0.03
4                 Gym  0.03


----Cañas----
         venue  freq
0        Hotel   1.0
1  Art Gallery   0.0
2          Pub   0.0
3         Pool   0.0
4        Plaza   0.0


----Ciudad Cortés----
           venue  freq
0           Park  0.17
1       Pharmacy  0.17
2  Grocery Store  0.17
3    Bus Station  0.17
4    Snack Place  0.17


----El Roble----
                venue  freq
0            Bus Stop   0.4
1   Other Repair Shop   0.2
2                Tree   0.2
3       Big Box Store   0.2
4  American Restaurant  0.0


----Heredia----
```

```
            venue  freq
0  Ice Cream Shop  0.11
1             Gym  0.11
2     Coffee Shop  0.07
3             Bar  0.07
4          Market  0.04


----La Cruz----
               venue  freq
0                Bar  0.50
1              Hotel  0.25
2         Restaurant  0.25
3  Other Repair Shop  0.00
4                Pub  0.00


----Liberia----
                venue  freq
0  Chinese Restaurant  0.20
1               Hotel  0.13
2                 Bar  0.13
3     Bed & Breakfast  0.13
4          Restaurant  0.13


----Puerto Limón----
                 venue  freq
0       Harbor / Marina   1.0
1  American Restaurant   0.0
2    Other Repair Shop   0.0
3                  Pub   0.0
4                 Pool   0.0


----Puntarenas----
                  venue  freq
0    Seafood Restaurant  0.14
1         Ice Cream Shop  0.10
2            Restaurant  0.06
3  Fast Food Restaurant  0.06
4    Chinese Restaurant  0.06


----Quesada----
                 venue  freq
0      Gymnastics Gym   0.5
1              Market   0.5
```

```
2  American Restaurant   0.0
3    Other Repair Shop   0.0
4                  Pub   0.0


----San José----
                        venue  freq
0              Sandwich Place  0.10
1                 Coffee Shop  0.10
2        Fast Food Restaurant  0.07
3                  Restaurant  0.07
4  Latin American Restaurant  0.05


----San Juan----
                        venue  freq
0                         Pub  0.08
1  Latin American Restaurant  0.08
2              Sandwich Place  0.08
3        Fast Food Restaurant  0.05
4                   Pet Store  0.05


----Santa Ana----
                       venue  freq
0                 Restaurant  0.17
1                 Steakhouse  0.08
2  Mediterranean Restaurant  0.04
3                  Pet Store  0.04
4        Peruvian Restaurant  0.04


----Sixaola----
                 venue  freq
0          Bus Station   1.0
1  American Restaurant   0.0
2           Restaurant   0.0
3                  Pub   0.0
4                 Pool   0.0
```

Function to sort the venues in descending order

[108]:
```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)
```

```
    return row_categories_sorted.index.values[0:num_top_venues]
```

Let's create the new dataframe and display the top 5 venues for each City.

[109]:
```
num_top_venues = 5

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['City']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
Cities_venues_sorted = pd.DataFrame(columns=columns)
Cities_venues_sorted['City'] = CR_grouped['City']

for ind in np.arange(CR_grouped.shape[0]):
    Cities_venues_sorted.iloc[ind, 1:] = return_most_common_venues(CR_grouped.
 ↪iloc[ind, :], num_top_venues)

Cities_venues_sorted.head()
```

[109]:
```
            City 1st Most Common Venue 2nd Most Common Venue  \
0       Alajuela            Restaurant     Food & Drink Shop
1   Buenos Aires             Juice Bar           Karaoke Bar
2        Cartago           Pizza Place                Bakery
3          Cañas                 Hotel       Harbor / Marina
4  Ciudad Cortés           Snack Place         Grocery Store

       3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
0                        Bar                  Pool         Women's Store
1  Latin American Restaurant            Steakhouse         Women's Store
2         Mexican Restaurant         Women's Store                   Gym
3                Coffee Shop     Convenience Store              Creperie
4                   Pharmacy           Bus Station                  Park
```

Cluster Neighborhoods

[110]:
```
# set number of clusters
kclusters = 5

CR_grouped_clustering = CR_grouped.drop('City', 1)
```

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

[110]:
```
array([0, 0, 0, 3, 0, 0, 0, 0, 0, 1], dtype=int32)
```

Let's create a new dataframe that includes the cluster as well as the previous information

[111]:
```
# add clustering labels
Cities_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

CR_merged = df2

# merge CR_grouped with df2 to add previous information for each city
CR_merged = CR_merged.join(Cities_venues_sorted.set_index('City'), on='City')

CR_merged.head() # check the last columns!
```

[111]:
```
           City  Latitude  Longitude     Country    Province  Population  \
0      San José    9.9333   -84.0833  Costa Rica    San José    288054.0
1       Cartago    9.8667   -83.9167  Costa Rica     Cartago    221733.0
2  Puerto Limón   10.0022   -83.0840  Costa Rica       Limón     61072.0
3       Liberia   10.6338   -85.4333  Costa Rica  Guanacaste     45380.0
4      Alajuela   10.0278   -84.2041  Costa Rica    Alajuela     42975.0

   Cluster Labels 1st Most Common Venue 2nd Most Common Venue  \
0             0.0           Coffee Shop        Sandwich Place
1             0.0           Pizza Place                Bakery
2             1.0        Harbor / Marina           Wings Joint
3             0.0     Chinese Restaurant                 Hotel
4             0.0            Restaurant     Food & Drink Shop

  3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
0  Fast Food Restaurant            Restaurant     Convenience Store
1    Mexican Restaurant         Women's Store                   Gym
2           Coffee Shop     Convenience Store              Creperie
3            Restaurant                   Bar       Bed & Breakfast
4                   Bar                  Pool         Women's Store
```

[114]:
```
CR_merged.dropna()
```

[114]:
```
           City  Latitude  Longitude     Country    Province  Population  \
0      San José    9.9333   -84.0833  Costa Rica    San José    288054.0
1       Cartago    9.8667   -83.9167  Costa Rica     Cartago    221733.0
```

2	Puerto Limón	10.0022	-83.0840	Costa Rica	Limón	61072.0
3	Liberia	10.6338	-85.4333	Costa Rica	Guanacaste	45380.0
4	Alajuela	10.0278	-84.2041	Costa Rica	Alajuela	42975.0
5	Puntarenas	9.9764	-84.8339	Costa Rica	Puntarenas	41528.0
6	San Juan	9.9609	-84.0731	Costa Rica	San José	24944.0
7	Heredia	9.9985	-84.1169	Costa Rica	Heredia	22700.0
8	Santa Ana	9.9320	-84.1760	Costa Rica	San José	11320.0
9	Buenos Aires	9.1497	-83.3334	Costa Rica	Puntarenas	45000.0
10	Quesada	10.3305	-84.4400	Costa Rica	Alajuela	31106.0
11	Cañas	10.4300	-85.1000	Costa Rica	Guanacaste	20306.0
12	El Roble	9.9771	-84.7443	Costa Rica	Puntarenas	15759.0
14	Sixaola	9.5083	-82.6147	Costa Rica	Limón	10234.0
15	La Cruz	11.0742	-85.6294	Costa Rica	Guanacaste	9195.0
17	Ciudad Cortés	8.9600	-83.5239	Costa Rica	Puntarenas	3850.0

	Cluster Labels	1st Most Common Venue	2nd Most Common Venue \
0	0.0	Coffee Shop	Sandwich Place
1	0.0	Pizza Place	Bakery
2	1.0	Harbor / Marina	Wings Joint
3	0.0	Chinese Restaurant	Hotel
4	0.0	Restaurant	Food & Drink Shop
5	0.0	Seafood Restaurant	Ice Cream Shop
6	0.0	Latin American Restaurant	Pub
7	0.0	Gym	Ice Cream Shop
8	0.0	Restaurant	Steakhouse
9	0.0	Juice Bar	Karaoke Bar
10	4.0	Gymnastics Gym	Market
11	3.0	Hotel	Harbor / Marina
12	0.0	Bus Stop	Big Box Store
14	2.0	Bus Station	Women's Store
15	0.0	Bar	Hotel
17	0.0	Snack Place	Grocery Store

	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Fast Food Restaurant	Restaurant	Convenience Store
1	Mexican Restaurant	Women's Store	Gym
2	Coffee Shop	Convenience Store	Creperie
3	Restaurant	Bar	Bed & Breakfast
4	Bar	Pool	Women's Store
5	Chinese Restaurant	Restaurant	Fast Food Restaurant
6	Sandwich Place	Fast Food Restaurant	Pet Store
7	Coffee Shop	Bar	Bistro
8	Shop & Service	Brewery	Ice Cream Shop
9	Latin American Restaurant	Steakhouse	Women's Store
10	Event Space	Coffee Shop	Convenience Store
11	Coffee Shop	Convenience Store	Creperie
12	Tree	Other Repair Shop	Event Space

```
14                      Church     Convenience Store               Creperie
15                  Restaurant         Women's Store            Event Space
17                    Pharmacy           Bus Station                   Park
```

Examine Clusters

Cluster 1

[118]:
```
CR_merged.loc[CR_merged['Cluster Labels'] == 0, CR_merged.columns[[1] +
 ↪list(range(5, CR_merged.shape[1]))]]
```

[118]:
```
    Latitude  Population  Cluster Labels      1st Most Common Venue  \
0     9.9333    288054.0             0.0                Coffee Shop
1     9.8667    221733.0             0.0               Pizza Place
3    10.6338     45380.0             0.0         Chinese Restaurant
4    10.0278     42975.0             0.0                 Restaurant
5     9.9764     41528.0             0.0         Seafood Restaurant
6     9.9609     24944.0             0.0  Latin American Restaurant
7     9.9985     22700.0             0.0                        Gym
8     9.9320     11320.0             0.0                 Restaurant
9     9.1497     45000.0             0.0                  Juice Bar
12    9.9771     15759.0             0.0                   Bus Stop
15   11.0742      9195.0             0.0                        Bar
17    8.9600      3850.0             0.0                Snack Place

   2nd Most Common Venue      3rd Most Common Venue 4th Most Common Venue  \
0         Sandwich Place      Fast Food Restaurant            Restaurant
1                 Bakery        Mexican Restaurant         Women's Store
3                  Hotel                Restaurant                   Bar
4       Food & Drink Shop                       Bar                  Pool
5          Ice Cream Shop        Chinese Restaurant            Restaurant
6                    Pub            Sandwich Place  Fast Food Restaurant
7          Ice Cream Shop               Coffee Shop                   Bar
8             Steakhouse            Shop & Service               Brewery
9            Karaoke Bar  Latin American Restaurant            Steakhouse
12         Big Box Store                      Tree     Other Repair Shop
15                 Hotel                Restaurant         Women's Store
17         Grocery Store                  Pharmacy           Bus Station

   5th Most Common Venue
0      Convenience Store
1                    Gym
3        Bed & Breakfast
4          Women's Store
5   Fast Food Restaurant
6              Pet Store
7                 Bistro
```

```
8         Ice Cream Shop
9          Women's Store
12           Event Space
15           Event Space
17                  Park
```

Cluster 2

[119]:
```python
CR_merged.loc[CR_merged['Cluster Labels'] == 1, CR_merged.columns[[1] +
 list(range(5, CR_merged.shape[1]))]]
```

[119]:
```
   Latitude  Population  Cluster Labels 1st Most Common Venue  \
2   10.0022     61072.0             1.0       Harbor / Marina

  2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue  \
2           Wings Joint           Coffee Shop     Convenience Store

  5th Most Common Venue
2              Creperie
```

Cluster 3

[120]:
```python
CR_merged.loc[CR_merged['Cluster Labels'] == 2, CR_merged.columns[[1] +
 list(range(5, CR_merged.shape[1]))]]
```

[120]:
```
    Latitude  Population  Cluster Labels 1st Most Common Venue  \
14    9.5083     10234.0             2.0           Bus Station

   2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue  \
14         Women's Store                Church     Convenience Store

   5th Most Common Venue
14              Creperie
```

Cluster 4

[121]:
```python
CR_merged.loc[CR_merged['Cluster Labels'] == 3, CR_merged.columns[[1] +
 list(range(5, CR_merged.shape[1]))]]
```

[121]:
```
    Latitude  Population  Cluster Labels 1st Most Common Venue  \
11     10.43     20306.0             3.0                 Hotel

   2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue  \
11       Harbor / Marina           Coffee Shop     Convenience Store

   5th Most Common Venue
11              Creperie
```

Cluster 5

[122]:
```
CR_merged.loc[CR_merged['Cluster Labels'] == 4, CR_merged.columns[[1] +
↪list(range(5, CR_merged.shape[1]))]]
```

[122]:
```
     Latitude  Population  Cluster Labels 1st Most Common Venue  \
10    10.3305     31106.0             4.0        Gymnastics Gym

   2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue  \
10                Market           Event Space           Coffee Shop

   5th Most Common Venue
10     Convenience Store
```

4 Results

4.0.1 From Cluster 1, it is recommended to establish the gyms in the cities of San Jose and Liberia.

5 Discussion section

5.0.1 The results may not represent the reality, because in the selected country the common venues does not necessarily include all the exact venues that exist in the area. For that reason it is recommended to made a second analysis.

6 Conclusion

6.0.1 According to the requirements defined at the beginning of the project, it was possible to find and recommend the best two places to establish the gyms.