

coursera_capstone

May 1, 2019

1 Battle of the Neighborhoods - New Dispensary

2 Introduction

With the legal landscape rapidly changing, regulation and decriminalization of Marijuana is significantly increasing. A substance that was once considered dangerous through urban legends and propaganda has been debunked and is being touted for a combination of its medicinal benefits, increased safety vs alternative substances, plant by-products and more. In the changing environment, it is the perfect time to invest in this growing \$16 Billion industry.

The goal of this project is to determine the optimal location for opening a new dispensary in cities in which marijuana is recreationally legal. The initial location and underlying data may then be used to find a storefront, establish a supply chain, build marketing campaign and develop a fiscal business plan to get investor contribution.

3 Data

3.1 To determine the best location for opening a number of factors will be looked at:

2019 State Laws <https://disa.com/map-of-marijuana-legality-by-state>

2019 Local Venue Data <https://foursquare.com>

2017-2018 Population and Crime Statistics <https://ucr.fbi.gov/crime-in-the-u.s/2018/preliminary-report/tables/table-4/table-4.xls/view>

2019 State Tax Rate <https://taxfoundation.org/sales-tax-rates-2019/>

List of US States https://simple.wikipedia.org/wiki/List_of_U.S._states

- First, the legality will be assessed for each state and city.
- The legality will be cross referenced against a US-population database to find the most densely populated cities with population > 100,000 people.
- Locations of current dispensaries within 1 mile (1.609 km) will be checked.

- If a dispensary exists, the location will be ruled out.
- Another factor that will be assessed is the demographic of the area.
- Cities or neighborhoods with children and families will not be looked at in order to keep them family-friendly.
- Since the goal of this project is to open a recreational dispensary
- I will be looking in areas with active nightlife and adult activities
- I will be looking for places with nearby fast-food, pizza restaurants and convenience stores. Establishments are often frequented by target clientele
- I will be verifying crime rates in cities and aim for lower property crime rate locations
- Many dispensaries have been robbed since installation due to the nature of the product. Try to minimize this risk as much as possible.

Stipulations and Assumptions

1. The following states did not have crime data available and were not included in this assessment. This was not a major problem since none of these states had recreational marijuana legal except for Vermont.
 - Delaware
 - Maine
 - Mississippi
 - Vermont*
 - West Virginia
 - Wyoming
2. Although a legalized area, District of Columbia was not considered in this report since it is not a state and therefore did not have state data available.
3. Only property crime was considered. Violent crimes, although bad, were not taken into account since the risk posed in this endeavor is robbery, burglary or larceny.
4. Since this project is created to get investors and not based on a specific region, only cities with 100,000 residents or more were considered, additionally, anywhere in the United States was considered to be Viable.

3.2 Methodology

In order to gather the appropriate data, significant data cleansing was performed to get information into a usable format, remove bad data, exclude cities that are not in scope and generate a streamlined dataset.

I began with consolidating all data based on state/city. To do this I took the tax information which had abbreviated state names and combined them with a list of the 50 states from wikipedia. Once I had the taxes and state names available in its own dataframe I was able to combine it with the crime statistics dataframe. I also got geospatial data from GeoPy using the city and state names to add to the city database. Finally, I calculated a crime-index which is a ratio of property crime to population for each city to complete the database.

Once the database was established, I used the Geo-coordinate data to query the Foursquare server API and gather local venue data for each city. Cities that had dispensaries were dropped from the dataset. The venue data was then transformed with one-hot encoding and grouped along venue frequency by the counts in each city. This data was then used to cluster the cities into 5 different clusters based on what types of venue and overall culture is present in the area.

I started performing exploratory analysis by creating a map of all of the cities that were in my dataset. After clustering was performed, a new map was generated color coding the clusters on the map. Finally, once the target cluster was identified, the crime rate and sales tax were cross referenced to determine the top choice of city to open the dispensary.

```
In [1]: import numpy as np # library to handle data in a vectorized manner
import matplotlib as mpl
import matplotlib.cm as cm
import matplotlib.colors as colors
import pandas as pd # library for data analysis
import json # library to handle JSON files
import requests # library to handle requests
import folium # map rendering library
from sklearn.cluster import KMeans # import k-means from clustering stage
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
from bs4 import BeautifulSoup
#!conda install beautifulsoup4 --yes
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

print('Libraries imported.')
```

Libraries imported.

3.3 Get updated legalization data

```
In [2]: url = "disa.com/map-of-marijuana-legality-by-state"
r = requests.get("http://" + url)
data = r.text
soup = BeautifulSoup(data)

In [3]: text = []
legalized = []
i=0
for table in soup.find_all('td'):
    text.append([i, table])
    i+=1

for j in range(0,len(text)-1):
    if str(text[j+1][1]) == "<td>Fully Legal</td>":
        s = str(text[j][1])
        state = s[4:-5]
```

```

        legalized.append(state.upper())
print(legalized)

```

```
['ALASKA', 'CALIFORNIA', 'COLORADO', 'DISTRICT OF COLUMBIA', 'MAINE', 'MASSACHUSETTS', 'M
```

3.4 State Sales Tax

```

In [4]: url = "taxfoundation.org/sales-tax-rates-2019/"
        r = requests.get("http://" + url)
        data = r.text
        soup = BeautifulSoup(data)

```

```

In [5]: text = []

        for table in soup.find_all('td'):
            text.append(table)
            text.pop(0)

        state = []
        tax = []

        for i in range(0, len(text)):
            if i % 7 == 0:
                s = str(text[i])
                t = str(text[i+1])
                state.append(s[4:-5])
                tax.append(t[4:-5])

```

```

In [6]: for j in range(0, len(tax)):
        try :
            tax[j] = float(tax[j].strip('%'))/100
        except :
            tax[j] = float(0)

```

```

In [7]: for k in range(0, len(tax)):
        try :
            state[k] = state[k].split(' ', 1)[0]
            state[k] = state[k].split('\xa0', 1)[0]
        except :
            next

```

3.5 Correct State Names

```

In [8]: url = "simple.wikipedia.org/wiki/List_of_U.S._states"
        r = requests.get("http://" + url)
        data = r.text
        soup = BeautifulSoup(data)

```

```

In [9]: text = []

        for table in soup.find_all('area'):
            text.append(str(table))
        statelist = []

        for k in range(0,len(text)):
            try :
                text[k] = text[k][11:].split('"', 1)[0]
                text[k] = text[k].split(' CA', 1)[0]
                statelist.append(text[k].upper())
            except :
                next

In [10]: statelist.sort()
         statelist = list(dict.fromkeys(statelist))

In [11]: taxes = pd.DataFrame([statelist,tax[0:50]]).transpose()

In [12]: taxes.rename(columns={0:'State',1:'Tax'},inplace=True)

```

3.6 Import Crime Statistics

```

In [13]: crime = pd.read_csv(r'crime_stats.csv',skiprows=4)
         crime = crime[:-9]

In [14]: taxes.head()

Out[14]:
   State  Tax
0  ALABAMA  0.04
1  ALASKA   0
2  ARIZONA  0.056
3  ARKANSAS  0.065
4  CALIFORNIA  0.0725

In [15]: states = list(crime['State'].unique())
         del states[1]

In [16]: crime = crime.merge(taxes, on='State', how='left')

In [17]: for i in range(1,len(crime)):
         if i%2 != 0:
             crime.loc[i, 'City'] = crime.loc[i-1, 'City']
         if pd.isnull(crime.loc[i, 'State']):
             crime.loc[i, 'State'] = crime.loc[i-1, 'State']
             crime.loc[i, 'Tax'] = crime.loc[i-1, 'Tax']

In [18]: crime.rename(columns={"Population1":"Population","Property \ncrime":"Property Crime", "Larceny-\n
         crime.drop(columns=["Unnamed: 2","Violent \ncrime","Murder","Rape2", "Aggravated \nassault","Moto
         crime = crime[:-9]

```

```

In [19]: crime = crime.fillna(0)

In [20]: crime["Population"] = crime["Population"].str.replace(",", "").astype(float)
         crime["Property Crime"] = crime["Property Crime"].str.replace(",", "").astype(float)

In [21]: crime = crime[crime['Population'] != 0]

In [22]: crime['Crime Index'] = crime['Property Crime']/crime['Population']

In [23]: crime.dropna(axis=0,inplace=True)

In [24]: crime = crime.reset_index(drop=True)
         crime.head()

Out[24]:
```

	State	City	Population	Property Crime	Tax	Crime Index
0	ALABAMA	BIRMINGHAM	212178.0	6472.0	0.04	0.030503
1	ALABAMA	MOBILE4	248431.0	6493.0	0.04	0.026136
2	ALABAMA	MONTGOMERY	199099.0	4246.0	0.04	0.021326
3	ALABAMA	TUSCALOOSA	101124.0	1953.0	0.04	0.019313
4	ALASKA	ANCHORAGE	296188.0	7708.0	0.00	0.026024

```

In [25]: for i in range(0,len(crime)):
         if crime['State'][i] not in legalized:
             crime.drop([i],axis=0, inplace=True)

```

3.6.1 Find the city with lowest Property Crime City to open a dispensary in each state

```

In [26]: crime = crime[crime.City != 'RIALTO5']
         crime = crime[crime.City != 'LAS VEGAS METROPOLITAN POLICE DEPARTMENT']

In [27]: cities = pd.concat([crime['State'],crime['City'],crime['Tax'],crime['Crime Index']],axis=1)

In [28]: cities.reset_index(inplace=True,drop=True)

In [29]: cities.head()

```

```

Out[29]:
```

	State	City	Tax	Crime Index
0	ALASKA	ANCHORAGE	0.0000	0.026024
1	CALIFORNIA	ANAHEIM	0.0725	0.013098
2	CALIFORNIA	ANTIOCH	0.0725	0.016686
3	CALIFORNIA	BAKERSFIELD	0.0725	0.019861
4	CALIFORNIA	BERKELEY	0.0725	0.023042

3.6.2 Get Geospatial Data using GeoPy for City, State

```

In [30]: lat = []
         lon = []
         for i in range(0,len(cities)):
             address = '{} {}'.format(cities['City'][i],cities['State'][i])
             geolocator = Nominatim(user_agent='capstone')
             location = geolocator.geocode(address)
             lat.append(location.latitude)
             lon.append(location.longitude)

```

```
In [31]: cities = pd.concat([cities,pd.Series(lat),pd.Series(lon)],axis=1)
```

```
In [32]: cities.rename(columns={0:'Latitude',1:'Longitude'},inplace=True)
```

```
In [33]: cities.head()
```

```
Out[33]:
```

	State	City	Tax	Crime Index	Latitude	Longitude
0	ALASKA	ANCHORAGE	0.0000	0.026024	61.216313	-149.894852
1	CALIFORNIA	ANAHEIM	0.0725	0.013098	33.834752	-117.911732
2	CALIFORNIA	ANTIOCH	0.0725	0.016686	38.004921	-121.805789
3	CALIFORNIA	BAKERSFIELD	0.0725	0.019861	35.373871	-119.019464
4	CALIFORNIA	BERKELEY	0.0725	0.023042	37.870839	-122.272864

3.6.3 Get local venue data for each city from Foursquare API

```
In [34]: CLIENT_ID = 'K5HRW4OC5J4EVS14DH2OKQBNZV5GNXIOZKY03QXJ4RDCPTUD' # your Foursq
CLIENT_SECRET = '4ZWZFMIOVAL01GTTVLUJSJOL0YPARUAQB225RMYILABEIQ0Q' # your Fo
VERSION = '20190420'
limit = 100
radius = 1609
latitude = 1
longitude = 1
print('Your credentials:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

Your credentials:

```
CLIENT_ID: K5HRW4OC5J4EVS14DH2OKQBNZV5GNXIOZKY03QXJ4RDCPTUD
CLIENT_SECRET: 4ZWZFMIOVAL01GTTVLUJSJOL0YPARUAQB225RMYILABEIQ0Q
```

```
In [35]: url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{},{}&
CLIENT_ID,
CLIENT_SECRET,
VERSION,
latitude,
longitude,
radius,
limit)
```

```
In [36]: results = requests.get(url).json()
```

```
In [37]: def getNearbyVenues(names, latitudes, longitudes, radius=1609):
```

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

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}
```

```

        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 = ['Neighborhood',
                             'Neighborhood Latitude',
                             'Neighborhood Longitude',
                             'Venue',
                             'Venue Latitude',
                             'Venue Longitude',
                             'Venue Category']

    return(nearby_venues)

```

In [38]: # 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']

```

In [39]: venues = results['response']['groups'][0]['items']

```

nearby_venues = json_normalize(venues) # flatten JSON

```



```

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues = nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[1] for col in nearby_venues.columns]

```

3.6.4 Create map of all viable US cities with population over 100,000 and legalized Recreational Marijuana

```

In [40]: # create map of Cities using latitude and longitude values
cities_map = folium.Map(location=[39.83, -98.58], zoom_start=4)

# add markers to map
for lat, lng, city, state in zip(cities['Latitude'], cities['Longitude'], cities['City'], cities['State']):
    label = '{} {}'.format(state, city)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=10,
        popup=label,
        color='green',
        fill=True,
        fill_color='#228B22',
        fill_opacity=0.7,
        parse_html=False).add_to(cities_map)

cities_map

```

Out[40]: <folium.folium.Map at 0x7f4076071898>

```

In [41]: LIMIT = 100
search_query = 'top'
radius = 1609

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']

```

```

def getNearbyVenues(state, names, latitudes, longitudes, radius=1609):

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

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}&types={}&nearby={}'
        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([(
            state,
            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 = ['State',
                            'City',
                            'City Latitude',
                            'City Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)

```

```

In [43]: city_venues = getNearbyVenues(state=cities['State'],
                                         names=cities['City'],
                                         latitudes=cities['Latitude'],
                                         longitudes=cities['Longitude'],
                                         radius = 1609
                                         )

```

```

venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues = nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[1] for col in nearby_venues.columns]

```

In [44]: city_venues.head()

```

Out[44]:   State   City City Latitude  City Longitude \
0  ALASKA  ANCHORAGE    61.216313   -149.894852
1  ALASKA  ANCHORAGE    61.216313   -149.894852
2  ALASKA  ANCHORAGE    61.216313   -149.894852
3  ALASKA  ANCHORAGE    61.216313   -149.894852
4  ALASKA  ANCHORAGE    61.216313   -149.894852

          Venue Venue Latitude  Venue Longitude \
0          Glacier BrewHouse    61.217719   -149.896839
1  Humpy's Great Alaskan Alehouse    61.216427   -149.894146
2          Crow's Nest    61.217838   -149.899718
3          49th State Brewing    61.219736   -149.895975
4  Apple Anchorage 5th Avenue Mall    61.217140   -149.888671

          Venue Category
0          Brewery
1          Bar
2  Seafood Restaurant
3          Brewery
4  Electronics Store

```

3.6.5 Prepare the venue data to perform K-means clustering

```

In [68]: # one hot encoding
city_onehot = pd.get_dummies(city_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
city_onehot['State'] = city_venues['State']
city_onehot['City'] = city_venues['City']

# move neighborhood column to the first column
fixed_columns = [city_onehot.columns[-2]] + [city_onehot.columns[-1]] + list(city_onehot.columns[:-2])
city_onehot = city_onehot[fixed_columns]

```

```
In [46]: city_group = city_onehot.groupby(['State', 'City']).mean().reset_index()
```

3.6.6 Drop cities that already have dispensaries

```
In [47]: print(city_group.shape)
city_group = city_group[city_group['Marijuana Dispensary'] == 0]
print(city_group.shape)
```

```
(105, 387)
```

```
(99, 387)
```

```
In [48]: def return_most_common_venues(row, num_top_venues=10):
row_categories = row.iloc[1:]
row_categories_sorted = row_categories.sort_values(ascending=False)
return row_categories_sorted.index.values[0:num_top_venues]
```

3.6.7 Find top 10 most common venue types for each city

```
In [49]: num_top_venues = 10
```

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

```
# create columns according to number of top venues
```

```
columns = ['State', '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
```

```
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
```

```
neighborhoods_venues_sorted['State'] = city_group['State']
```

```
neighborhoods_venues_sorted['City'] = city_group['City']
```

```
for row in np.arange(city_group.shape[0]):
```

```
    neighborhoods_venues_sorted.iloc[row, 2:] = return_most_common_venues(city_group.iloc[row, 1:], 10)
```

```
neighborhoods_venues_sorted.head()
```

```
Out[49]:
```

	State	City	1st Most Common Venue	2nd Most Common Venue	\
0	ALASKA	ANCHORAGE	Coffee Shop	Park	
1	CALIFORNIA	ANAHEIM	Mexican Restaurant	Coffee Shop	
2	CALIFORNIA	ANTIOCH	Fast Food Restaurant	Pizza Place	
3	CALIFORNIA	BAKERSFIELD	Mexican Restaurant	Coffee Shop	
4	CALIFORNIA	BERKELEY	Chinese Restaurant	Japanese Restaurant	
			3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue \
0			Clothing Store	Seafood Restaurant	Bar

1	Ice Cream Shop	Indian Restaurant	Convenience Store
2	Mexican Restaurant	Chinese Restaurant	Racetrack
3	Fast Food Restaurant	Sandwich Place	Bar
4	New American Restaurant	Thai Restaurant	French Restaurant

	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue \
0	Accessories Store	Steakhouse	Sporting Goods Shop
1	Taco Place	Liquor Store	Burger Joint
2	Paintball Field	Bakery	Grocery Store
3	Chinese Restaurant	Italian Restaurant	Steakhouse
4	Café	Yoga Studio	Ice Cream Shop

	9th Most Common Venue	10th Most Common Venue
0	Cosmetics Shop	Pizza Place
1	Brewery	Southern / Soul Food Restaurant
2	Sports Bar	Burger Joint
3	Breakfast Spot	General Entertainment
4	Coffee Shop	Pizza Place

3.6.8 Performing K-Means Clustering using 5 clusters

I used SciKit Learn K-Means Clustering unsupervised learning in order to group different cities based on their top 10 most common venues. Once the clustering was performed, I reviewed the venues in the cluster to classify the clusters:

cluster 0: Shops cluster 1:

```
In [50]: k_means = KMeans(init = "k-means++", n_clusters = 5, n_init = 12)
```

```
In [51]: city_group = city_group.merge(cities,on=['City','State'])
```

```
In [52]: # set number of clusters
kclusters = 5
from sklearn.preprocessing import StandardScaler

city_clusters = city_group.drop(['State','City','Latitude','Longitude'], 1)
city_fit = StandardScaler().fit_transform(city_clusters)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(city_fit)

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

```
Out[52]: array([1, 1, 1, 1, 4, 1, 1, 1, 1, 1], dtype=int32)
```

```
In [53]: # add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster', kmeans.labels_)

city_full = cities
```

```
# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
city_full = city_full.merge(neighborhoods_venues_sorted, on=['State', 'City'], how='left')
```

```
In [54]: city_full.dropna(inplace=True, axis=0)
```

```
In [55]: city_full['Cluster'] = city_full['Cluster'].astype(int)
```

```
In [56]: city_full.head()
```

```
Out[56]:
```

	State	City	Tax	Crime Index	Latitude	Longitude
0	ALASKA	ANCHORAGE	0.0000	0.026024	61.216313	-149.894852
1	CALIFORNIA	ANAHEIM	0.0725	0.013098	33.834752	-117.911732
2	CALIFORNIA	ANTIOCH	0.0725	0.016686	38.004921	-121.805789
3	CALIFORNIA	BAKERSFIELD	0.0725	0.019861	35.373871	-119.019464
4	CALIFORNIA	BERKELEY	0.0725	0.023042	37.870839	-122.272864

	Cluster	1st Most Common Venue	2nd Most Common Venue
0	1	Coffee Shop	Park
1	1	Mexican Restaurant	Coffee Shop
2	1	Fast Food Restaurant	Pizza Place
3	1	Mexican Restaurant	Coffee Shop
4	4	Chinese Restaurant	Japanese Restaurant

	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Clothing Store	Seafood Restaurant	Bar
1	Ice Cream Shop	Indian Restaurant	Convenience Store
2	Mexican Restaurant	Chinese Restaurant	Racetrack
3	Fast Food Restaurant	Sandwich Place	Bar
4	New American Restaurant	Thai Restaurant	French Restaurant

	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Accessories Store	Steakhouse	Sporting Goods Shop
1	Taco Place	Liquor Store	Burger Joint
2	Paintball Field	Bakery	Grocery Store
3	Chinese Restaurant	Italian Restaurant	Steakhouse
4	Café	Yoga Studio	Ice Cream Shop

	9th Most Common Venue	10th Most Common Venue
0	Cosmetics Shop	Pizza Place
1	Brewery	Southern / Soul Food Restaurant
2	Sports Bar	Burger Joint
3	Breakfast Spot	General Entertainment
4	Coffee Shop	Pizza Place

```
In [57]: # create map
```

```
map_clusters = folium.Map(location=[39.83, -98.58], zoom_start=4)
```

```
# 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(city_full['Latitude'], city_full['Longitude'], city_full['City'], city_full['Cluster']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=10,
        popup=label,
        color=rainbow[cluster],
        fill=True,
        fill_color=rainbow[cluster],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters

```

Out[57]: <folium.folium.Map at 0x7f4074366630>

3.7 Review Clusters

Cluster 0

In [58]: city_full.loc[city_full['Cluster'] == 0, city_full.columns[[1] + list(range(2, city_full.shape[1]))]]

```

Out[58]:
      City  Tax  Crime Index  Latitude  Longitude  Cluster \
51  SAN BERNARDINO  0.0725    0.018655  34.108345 -117.289765    0

      1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue \
51  Convenience Store      Clothing Store      Discount Store

      4th Most Common Venue 5th Most Common Venue 6th Most Common Venue \
51  Fast Food Restaurant    Grocery Store      Pizza Place

      7th Most Common Venue 8th Most Common Venue 9th Most Common Venue \
51      Nightclub      Shoe Store      Department Store

      10th Most Common Venue
51  Mexican Restaurant

```

Cluster 1

In [59]: city_full.loc[city_full['Cluster'] == 1, city_full.columns[[1] + list(range(2, city_full.shape[1]))]]

```

Out[59]:
      City  Tax  Crime Index  Latitude  Longitude  Cluster \
0  ANCHORAGE  0.0000    0.026024  61.216313 -149.894852    1
1  ANAHEIM    0.0725    0.013098  33.834752 -117.911732    1

```

2	ANTIOCH	0.0725	0.016686	38.004921	-121.805789	1	
3	BAKERSFIELD	0.0725	0.019861	35.373871	-119.019464	1	
5	BURBANK	0.0725	0.014175	34.181648	-118.325855	1	
6	CARLSBAD	0.0725	0.008791	33.158093	-117.350597	1	
7	CHULA VISTA	0.0725	0.007034	32.640054	-117.084196	1	
8	CLOVIS	0.0725	0.013983	36.825228	-119.702919	1	
9	CONCORD	0.0725	0.016411	37.976852	-122.033562	1	
10	CORONA	0.0725	0.010682	33.875295	-117.566445	1	
11	COSTA MESA	0.0725	0.020571	33.663339	-117.903317	1	
12	DALY CITY	0.0725	0.007722	37.705767	-122.461921	1	
13	DOWNEY	0.0725	0.014721	33.942215	-118.123565	1	
14	EL CAJON	0.0725	0.011556	32.794773	-116.962527	1	
15	ELK GROVE	0.0725	0.006830	38.408799	-121.371618	1	
17	ESCONDIDO	0.0725	0.008303	33.121675	-117.081485	1	
18	FAIRFIELD	0.0725	0.012314	38.249358	-122.039966	1	
19	FONTANA	0.0725	0.008688	34.092233	-117.435048	1	
20	FREMONT	0.0725	0.011516	37.548270	-121.988572	1	
21	FRESNO	0.0725	0.019484	36.729529	-119.708861	1	
22	FULLERTON	0.0725	0.014008	33.870821	-117.929417	1	
23	GARDEN GROVE	0.0725	0.012589	33.774629	-117.946372	1	
24	GLENDALE	0.0725	0.007674	34.192912	-118.246249	1	
25	HAYWARD	0.0725	0.015630	37.668821	-122.080796	1	
26	HUNTINGTON BEACH	0.0725	0.010547	33.678334	-118.000017	1	
27	INGLEWOOD	0.0725	0.012327	33.956200	-118.353132	1	
28	IRVINE	0.0725	0.006508	33.685697	-117.825982	1	
29	JURUPA VALLEY	0.0725	0.012886	33.979847	-117.451575	1	
30	LANCASTER	0.0725	0.009717	34.698106	-118.136615	1	
31	LONG BEACH	0.0725	0.013322	33.785389	-118.158049	1	
33	MODESTO	0.0725	0.020026	37.639097	-120.996878	1	
34	MORENO VALLEY	0.0725	0.015924	33.937517	-117.230594	1	
35	MURRIETA	0.0725	0.007406	33.577752	-117.188454	1	
36	NORWALK	0.0725	0.009315	33.909280	-118.084917	1	
37	OAKLAND	0.0725	0.030660	37.804456	-122.271356	1	
38	OCEANSIDE	0.0725	0.011877	33.195870	-117.379483	1	
39	ONTARIO	0.0725	0.012694	34.065846	-117.648430	1	
41	OXNARD	0.0725	0.012863	34.197631	-119.180382	1	
42	PALMDALE	0.0725	0.008084	34.579313	-118.117111	1	
44	POMONA	0.0725	0.014517	34.055381	-117.751750	1	
45	RANCHO CUCAMONGA	0.0725	0.011890	34.103319	-117.575174	1	
46	RICHMOND	0.0725	0.019259	37.935758	-122.347749	1	
47	RIVERSIDE	0.0725	0.016063	33.953355	-117.396162	1	
48	ROSEVILLE	0.0725	0.012634	38.752124	-121.288006	1	
50	SALINAS	0.0725	0.014825	36.674412	-121.655037	1	
55	SAN MATEO	0.0725	0.009554	37.496904	-122.333057	1	
57	SANTA CLARA	0.0725	0.012059	37.354113	-121.955174	1	
58	SANTA CLARITA	0.0725	0.007566	34.391664	-118.542586	1	
59	SANTA MARIA	0.0725	0.011338	34.953130	-120.435858	1	
60	SANTA ROSA	0.0725	0.008885	38.440467	-122.714431	1	

61	SIMI VALLEY	0.0725	0.007241	34.269447	-118.781482	1
62	STOCKTON	0.0725	0.018623	37.957702	-121.290780	1
63	SUNNYVALE	0.0725	0.008075	37.368830	-122.036350	1
64	TEMECULA	0.0725	0.012463	33.494635	-117.147366	1
65	THOUSAND OAKS	0.0725	0.006476	34.171427	-118.910588	1
66	TORRANCE	0.0725	0.010483	33.835849	-118.340629	1
67	VALLEJO	0.0725	0.019390	38.104086	-122.256637	1
69	VICTORVILLE	0.0725	0.013789	34.536107	-117.291156	1
70	VISALIA	0.0725	0.016111	36.330228	-119.292058	1
71	VISTA	0.0725	0.007242	33.200037	-117.242536	1
72	WEST COVINA	0.0725	0.011708	34.068621	-117.938953	1
75	COLORADO SPRINGS	0.0290	0.015192	38.833958	-104.825349	1
77	FORT COLLINS	0.0290	0.012820	40.550853	-105.066808	1
78	GREELEY	0.0290	0.011897	40.423314	-104.709132	1
79	PUEBLO	0.0290	0.029836	38.254447	-104.609141	1
80	WESTMINSTER	0.0290	0.015327	39.836653	-105.037205	1
83	LOWELL	0.0625	0.010297	42.633425	-71.316172	1
84	SPRINGFIELD	0.0625	0.014286	42.101483	-72.589811	1
85	WORCESTER	0.0625	0.010178	42.262593	-71.802293	1
86	ANN ARBOR	0.0600	0.007332	42.268157	-83.731229	1
87	CLINTON TOWNSHIP	0.0600	0.007386	42.584852	-82.934824	1
89	GRAND RAPIDS	0.0600	0.009319	42.963240	-85.667864	1
90	LANSING	0.0600	0.014394	42.733771	-84.555380	1
91	STERLING HEIGHTS	0.0600	0.005351	42.580312	-83.030203	1
92	WARREN	0.0600	0.010428	42.493257	-83.006275	1
93	HENDERSON	0.0685	0.008818	36.039146	-114.981923	1
94	NORTH LAS VEGAS	0.0685	0.012398	36.200837	-115.112096	1
96	EUGENE	0.0000	0.018294	44.050505	-123.095051	1
97	GRESHAM	0.0000	0.017143	45.506741	-122.436706	1
98	HILLSBORO	0.0000	0.009960	45.522894	-122.989827	1
100	SALEM	0.0000	0.021874	44.939157	-123.033121	1
101	BELLEVUE	0.0650	0.015553	47.614422	-122.192337	1
102	EVERETT	0.0650	0.022710	47.967306	-122.201400	1
103	KENT	0.0650	0.026390	47.382690	-122.227027	1
106	SPOKANE	0.0650	0.036404	47.657942	-117.421227	1

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue \
0	Coffee Shop	Park	Clothing Store
1	Mexican Restaurant	Coffee Shop	Ice Cream Shop
2	Fast Food Restaurant	Pizza Place	Mexican Restaurant
3	Mexican Restaurant	Coffee Shop	Fast Food Restaurant
5	Mexican Restaurant	Sandwich Place	American Restaurant
6	Beach	Mexican Restaurant	Café
7	Mexican Restaurant	Grocery Store	Convenience Store
8	Mexican Restaurant	Sandwich Place	Coffee Shop
9	Mexican Restaurant	Sandwich Place	Discount Store
10	Mexican Restaurant	Convenience Store	Rental Car Location
11	Sushi Restaurant	Mexican Restaurant	Playground

12	Chinese Restaurant	Sandwich Place	Fast Food Restaurant
13	Mexican Restaurant	Coffee Shop	Fast Food Restaurant
14	Mexican Restaurant	Clothing Store	Coffee Shop
15	Pizza Place	Pharmacy	Fast Food Restaurant
17	Mexican Restaurant	Fast Food Restaurant	American Restaurant
18	Coffee Shop	Mexican Restaurant	Convenience Store
19	Fast Food Restaurant	Mexican Restaurant	Convenience Store
20	Grocery Store	Chinese Restaurant	Pizza Place
21	Fast Food Restaurant	Chinese Restaurant	Sandwich Place
22	Sushi Restaurant	Italian Restaurant	Burger Joint
23	Vietnamese Restaurant	Fast Food Restaurant	Korean Restaurant
24	Trail	Park	Golf Course
25	Mexican Restaurant	Pizza Place	Coffee Shop
26	Coffee Shop	Pizza Place	Mexican Restaurant
27	Mexican Restaurant	Fast Food Restaurant	Grocery Store
28	Sandwich Place	Bakery	Ice Cream Shop
29	Fast Food Restaurant	Convenience Store	Pizza Place
30	Mexican Restaurant	Diner	Fast Food Restaurant
31	Mexican Restaurant	Pizza Place	Thai Restaurant
33	Coffee Shop	Park	American Restaurant
34	Fast Food Restaurant	Mexican Restaurant	Pizza Place
35	Mexican Restaurant	Pizza Place	Fast Food Restaurant
36	Fast Food Restaurant	Mexican Restaurant	Coffee Shop
37	Bar	Coffee Shop	Chinese Restaurant
38	American Restaurant	Beach	Ice Cream Shop
39	Mexican Restaurant	Burger Joint	Convenience Store
41	Mexican Restaurant	Grocery Store	Pharmacy
42	Fast Food Restaurant	Mexican Restaurant	Discount Store
44	Mexican Restaurant	Pizza Place	Coffee Shop
45	Mobile Phone Shop	Mexican Restaurant	Burger Joint
46	Mexican Restaurant	Convenience Store	Pharmacy
47	Fast Food Restaurant	Mexican Restaurant	Pizza Place
48	Mexican Restaurant	Park	Pizza Place
50	Hotel	Mexican Restaurant	Fast Food Restaurant
55	Trail	Intersection	Yoshoku Restaurant
57	Korean Restaurant	Fast Food Restaurant	Mexican Restaurant
58	Pizza Place	Pharmacy	Convenience Store
59	Mexican Restaurant	Fast Food Restaurant	Convenience Store
60	Mexican Restaurant	Coffee Shop	Pizza Place
61	Mexican Restaurant	Sushi Restaurant	Sandwich Place
62	Mexican Restaurant	Fast Food Restaurant	Coffee Shop
63	Coffee Shop	Indian Restaurant	Burger Joint
64	American Restaurant	Mexican Restaurant	Italian Restaurant
65	Hotel	Martial Arts Dojo	Park
66	Clothing Store	Cosmetics Shop	Japanese Restaurant
67	Mexican Restaurant	Harbor / Marina	Seafood Restaurant
69	Pizza Place	Convenience Store	Mexican Restaurant
70	Mexican Restaurant	Fast Food Restaurant	Pizza Place

71	Fast Food Restaurant	Mexican Restaurant	Sandwich Place
72	Clothing Store	Mexican Restaurant	Bubble Tea Shop
75	Bar	Coffee Shop	Brewery
77	Coffee Shop	Clothing Store	Cosmetics Shop
78	Bar	Mexican Restaurant	Convenience Store
79	Mexican Restaurant	Pizza Place	Italian Restaurant
80	Mexican Restaurant	Convenience Store	Fast Food Restaurant
83	Sandwich Place	Pizza Place	Asian Restaurant
84	Donut Shop	Sandwich Place	American Restaurant
85	Italian Restaurant	Café	Bar
86	Coffee Shop	Pizza Place	Bar
87	Convenience Store	Food	Liquor Store
89	Coffee Shop	Museum	Bar
90	Bar	Coffee Shop	Bakery
91	Sandwich Place	Fast Food Restaurant	Shipping Store
92	Fast Food Restaurant	American Restaurant	Fried Chicken Joint
93	Fast Food Restaurant	Coffee Shop	Convenience Store
94	Fast Food Restaurant	Convenience Store	Mexican Restaurant
96	Brewery	Coffee Shop	Pizza Place
97	Coffee Shop	Pizza Place	Bar
98	Mexican Restaurant	Coffee Shop	Fast Food Restaurant
100	Coffee Shop	American Restaurant	Bar
101	Coffee Shop	Steakhouse	Spa
102	Coffee Shop	Mexican Restaurant	Gym
103	Coffee Shop	Mexican Restaurant	Pub
106	Bar	Pizza Place	American Restaurant

	4th Most Common Venue	5th Most Common Venue \
0	Seafood Restaurant	Bar
1	Indian Restaurant	Convenience Store
2	Chinese Restaurant	Racetrack
3	Sandwich Place	Bar
5	Pizza Place	Burger Joint
6	Coffee Shop	Hotel
7	Clothing Store	Taco Place
8	Pizza Place	Burger Joint
9	Japanese Restaurant	Coffee Shop
10	Fast Food Restaurant	Indian Restaurant
11	Hotel	Concert Hall
12	Pizza Place	Playground
13	Pizza Place	Burger Joint
14	Convenience Store	Fast Food Restaurant
15	American Restaurant	Coffee Shop
17	Convenience Store	Pizza Place
18	American Restaurant	Thai Restaurant
19	Gas Station	Pizza Place
20	Coffee Shop	Bakery
21	Pharmacy	Pizza Place

22	Mexican Restaurant	Ice Cream Shop
23	Coffee Shop	Convenience Store
24	Scenic Lookout	Yoshoku Restaurant
25	Bar	Fast Food Restaurant
26	Fast Food Restaurant	Gym / Fitness Center
27	Convenience Store	Southern / Soul Food Restaurant
28	Café	Coffee Shop
29	Golf Course	Pharmacy
30	Convenience Store	Bar
31	Bar	Fast Food Restaurant
33	Sandwich Place	Italian Restaurant
34	Pharmacy	Sandwich Place
35	Pharmacy	Park
36	Pizza Place	Pharmacy
37	Mexican Restaurant	Sandwich Place
38	Mexican Restaurant	Seafood Restaurant
39	Pizza Place	Sandwich Place
41	Convenience Store	Italian Restaurant
42	Pizza Place	Thai Restaurant
44	Bar	Pharmacy
45	Asian Restaurant	Coffee Shop
46	Liquor Store	Food Truck
47	Coffee Shop	Sandwich Place
48	Bar	Grocery Store
50	Coffee Shop	American Restaurant
55	Farm	Empanada Restaurant
57	Thai Restaurant	Bubble Tea Shop
58	Park	Coffee Shop
59	Burger Joint	Sushi Restaurant
60	Hotel	American Restaurant
61	Coffee Shop	Fast Food Restaurant
62	American Restaurant	Deli / Bodega
63	Sandwich Place	Grocery Store
64	Coffee Shop	Sushi Restaurant
65	Gym	Sandwich Place
66	Lingerie Store	American Restaurant
67	Coffee Shop	Food Truck
69	Sandwich Place	Platform
70	Sandwich Place	Italian Restaurant
71	Coffee Shop	Convenience Store
72	Vietnamese Restaurant	Korean Restaurant
75	Italian Restaurant	Pizza Place
77	Mexican Restaurant	Seafood Restaurant
78	Fast Food Restaurant	Sandwich Place
79	Fast Food Restaurant	Bakery
80	Sushi Restaurant	Grocery Store
83	Donut Shop	Pharmacy
84	Burger Joint	Discount Store

85	Coffee Shop	Mexican Restaurant
86	Sandwich Place	Korean Restaurant
87	Baseball Field	Diner
89	Brewery	Hotel
90	Mexican Restaurant	Pharmacy
91	Shopping Mall	Discount Store
92	Coffee Shop	Intersection
93	Pizza Place	Mexican Restaurant
94	Fried Chicken Joint	Casino
96	Sushi Restaurant	Café
97	American Restaurant	Furniture / Home Store
98	Sandwich Place	Pizza Place
100	Sandwich Place	Pizza Place
101	Grocery Store	Vietnamese Restaurant
102	Hockey Arena	Asian Restaurant
103	Clothing Store	Fast Food Restaurant
106	Coffee Shop	Brewery

	6th Most Common Venue	7th Most Common Venue \
0	Accessories Store	Steakhouse
1	Taco Place	Liquor Store
2	Paintball Field	Bakery
3	Chinese Restaurant	Italian Restaurant
5	Diner	Bakery
6	Breakfast Spot	Pizza Place
7	Sandwich Place	Italian Restaurant
8	Hotel	Ice Cream Shop
9	Café	Chinese Restaurant
10	Sandwich Place	Diner
11	Convenience Store	Flea Market
12	Convenience Store	Rental Car Location
13	Sushi Restaurant	Restaurant
14	Middle Eastern Restaurant	Sandwich Place
15	Rental Car Location	Sports Bar
17	Coffee Shop	Sandwich Place
18	Gas Station	Sushi Restaurant
19	American Restaurant	Café
20	Pet Store	Intersection
21	Mobile Phone Shop	Mexican Restaurant
22	Pizza Place	Café
23	Bar	Café
24	Electronics Store	Empanada Restaurant
25	Chinese Restaurant	Sandwich Place
26	Spa	Grocery Store
27	Burger Joint	Pizza Place
28	Japanese Restaurant	Burger Joint
29	Ice Cream Shop	Movie Theater
30	Vegetarian / Vegan Restaurant	Coffee Shop

31	Convenience Store	Sandwich Place
33	Fast Food Restaurant	Café
34	American Restaurant	Grocery Store
35	Chinese Restaurant	Coffee Shop
36	Park	Chinese Restaurant
37	Cocktail Bar	Japanese Restaurant
38	Surf Spot	Coffee Shop
39	Seafood Restaurant	ATM
41	Fast Food Restaurant	Thrift / Vintage Store
42	Train Station	Convenience Store
44	Convenience Store	Music Venue
45	Japanese Restaurant	Furniture / Home Store
46	Park	Metro Station
47	Convenience Store	American Restaurant
48	Convenience Store	Coffee Shop
50	Convenience Store	Breakfast Spot
55	Entertainment Service	Ethiopian Restaurant
57	Coffee Shop	Asian Restaurant
58	Fast Food Restaurant	Bank
59	Pharmacy	Pizza Place
60	Thai Restaurant	Italian Restaurant
61	Grocery Store	Burger Joint
62	Baseball Stadium	Chinese Restaurant
63	Park	Mexican Restaurant
64	Steakhouse	Brewery
65	Sushi Restaurant	Thrift / Vintage Store
66	Coffee Shop	Mexican Restaurant
67	Pizza Place	Rental Car Location
69	Performing Arts Venue	Donut Shop
70	American Restaurant	Coffee Shop
71	Chinese Restaurant	Ice Cream Shop
72	Noodle House	Pizza Place
75	Gastropub	Mexican Restaurant
77	Breakfast Spot	Gym
78	Pizza Place	Pharmacy
79	American Restaurant	Bar
80	Sandwich Place	Vietnamese Restaurant
83	Discount Store	Coffee Shop
84	Hotel	Shipping Store
85	Pizza Place	Sandwich Place
86	Thrift / Vintage Store	College Auditorium
87	Sandwich Place	Coffee Shop
89	Music Venue	Steakhouse
90	Convenience Store	Middle Eastern Restaurant
91	Rental Car Location	American Restaurant
92	Discount Store	Pharmacy
93	Grocery Store	Gym / Fitness Center
94	Bakery	Pizza Place

96	Breakfast Spot	Indie Movie Theater
97	Burger Joint	Mexican Restaurant
98	Convenience Store	Grocery Store
100	Restaurant	Italian Restaurant
101	Sushi Restaurant	Sandwich Place
102	Sushi Restaurant	Bakery
103	Sandwich Place	Bakery
106	Sushi Restaurant	Lounge

	8th Most Common Venue	9th Most Common Venue \
0	Sporting Goods Shop	Cosmetics Shop
1	Burger Joint	Brewery
2	Grocery Store	Sports Bar
3	Steakhouse	Breakfast Spot
5	Donut Shop	Deli / Bodega
6	Italian Restaurant	American Restaurant
7	Seafood Restaurant	Cosmetics Shop
8	American Restaurant	Fast Food Restaurant
9	Pizza Place	Italian Restaurant
10	Discount Store	Sushi Restaurant
11	Italian Restaurant	Video Store
12	Park	Dive Bar
13	Pharmacy	American Restaurant
14	Chinese Restaurant	Cosmetics Shop
15	Salon / Barbershop	Sandwich Place
17	Vietnamese Restaurant	Cosmetics Shop
18	Burger Joint	Chinese Restaurant
19	Discount Store	Fried Chicken Joint
20	Café	Mexican Restaurant
21	Salon / Barbershop	Grocery Store
22	Coffee Shop	Sports Bar
23	Chinese Restaurant	Asian Restaurant
24	Entertainment Service	Ethiopian Restaurant
25	Shipping Store	Italian Restaurant
26	Japanese Restaurant	Sandwich Place
27	Coffee Shop	Pharmacy
28	Mediterranean Restaurant	Mexican Restaurant
29	Fried Chicken Joint	Park
30	Plaza	Fried Chicken Joint
31	Cosmetics Shop	Grocery Store
33	New American Restaurant	Mexican Restaurant
34	Video Store	Discount Store
35	Sandwich Place	Gym
36	Cosmetics Shop	Donut Shop
37	Vietnamese Restaurant	Café
38	Breakfast Spot	Beer Garden
39	Candy Store	Fried Chicken Joint
41	Chinese Restaurant	Park

42	Burger Joint	Bus Station
44	Nightclub	Gay Bar
45	Pizza Place	Bakery
46	Performing Arts Venue	Gym
47	Pet Store	Grocery Store
48	Fast Food Restaurant	Gas Station
50	Shipping Store	Thai Restaurant
55	Event Service	Event Space
57	Sandwich Place	Convenience Store
58	Mexican Restaurant	Martial Arts Dojo
59	Sandwich Place	Steakhouse
60	Brewery	Clothing Store
61	Breakfast Spot	Park
62	Sandwich Place	Rental Car Location
63	Chinese Restaurant	Bubble Tea Shop
64	Wine Bar	Pizza Place
65	Liquor Store	Beer Store
66	Chinese Restaurant	Shoe Store
67	Pet Store	Southern / Soul Food Restaurant
69	Bakery	Fast Food Restaurant
70	Salon / Barbershop	Breakfast Spot
71	Pizza Place	American Restaurant
72	Asian Restaurant	Ice Cream Shop
75	Sandwich Place	Steakhouse
77	Grocery Store	Vietnamese Restaurant
78	Discount Store	Bank
79	Café	Sandwich Place
80	Rental Car Location	Pizza Place
83	Café	Chinese Restaurant
84	Sports Bar	Gas Station
85	American Restaurant	Breakfast Spot
86	Park	Gourmet Shop
87	Clothing Store	Park
89	American Restaurant	Pizza Place
90	Intersection	Irish Pub
91	Pub	Bank
92	Video Store	Chinese Restaurant
93	Big Box Store	Shoe Store
94	Pharmacy	Coffee Shop
96	Japanese Restaurant	Burger Joint
97	Fast Food Restaurant	Thai Restaurant
98	Flea Market	Ice Cream Shop
100	Park	Fast Food Restaurant
101	Shopping Mall	Food Truck
102	Burger Joint	Tattoo Parlor
103	Gym / Fitness Center	Chinese Restaurant
106	Italian Restaurant	Concert Hall

	10th Most Common Venue
0	Pizza Place
1	Southern / Soul Food Restaurant
2	Burger Joint
3	General Entertainment
5	Pet Store
6	Bar
7	Coffee Shop
8	Italian Restaurant
9	Thai Restaurant
10	Furniture / Home Store
11	Coffee Shop
12	Supermarket
13	Department Store
14	Bar
15	New American Restaurant
17	Thai Restaurant
18	Sandwich Place
19	Latin American Restaurant
20	Fried Chicken Joint
21	Coffee Shop
22	Breakfast Spot
23	Arts & Crafts Store
24	Event Service
25	Sushi Restaurant
26	Pharmacy
27	Cosmetics Shop
28	Asian Restaurant
29	Sporting Goods Shop
30	American Restaurant
31	Thrift / Vintage Store
33	Convenience Store
34	Convenience Store
35	Grocery Store
36	Burger Joint
37	Vegetarian / Vegan Restaurant
38	Brewery
39	Mobile Phone Shop
41	Ice Cream Shop
42	Sushi Restaurant
44	Taco Place
45	Sushi Restaurant
46	Fried Chicken Joint
47	Sushi Restaurant
48	Sandwich Place
50	Chinese Restaurant
55	Exhibit
57	Indian Restaurant

58	Chinese Restaurant
59	Shoe Store
60	French Restaurant
61	Gym
62	Cosmetics Shop
63	Hotel
64	Gastropub
65	Italian Restaurant
66	Brewery
67	Chinese Restaurant
69	History Museum
70	Brewery
71	ATM
72	Grocery Store
75	American Restaurant
77	Brewery
78	Steakhouse
79	Hotel
80	Storage Facility
83	Gym
84	Rental Car Location
85	Middle Eastern Restaurant
86	Sushi Restaurant
87	Bar
89	Ice Cream Shop
90	Sandwich Place
91	Bar
92	Clothing Store
93	Cosmetics Shop
94	Discount Store
96	Beer Store
97	Women's Store
98	Sports Bar
100	Mexican Restaurant
101	Mexican Restaurant
102	Beer Store
103	Hockey Arena
106	Hotel

Cluster 2

```
In [60]: city_full.loc[city_full['Cluster'] == 2, city_full.columns[[1] + list(range(2, city_full.shape[1]))]]
```

```
Out[60]:
```

	City	Tax	Crime Index	Latitude	Longitude	Cluster	\
74	CENTENNIAL	0.029	0.008221	39.568064	-104.977831	2	
	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	\			
74	Fast Food Restaurant	Sandwich Place	Grocery Store				

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
74	American Restaurant	Pizza Place	Mexican Restaurant

	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
74	Spa	BBQ Joint	Coffee Shop

	10th Most Common Venue
74	Furniture / Home Store

Cluster 3 This is our ideal cluster. Active nightlife, pubs, breweries and trendy areas and activities.

```
In [61]: cluster3 = city_full.loc[city_full['Cluster'] == 3, city_full.columns[[1] + list(range(2, city_full.shape[1]))]
cluster3
```

```
Out[61]:
```

	City	Tax	Crime Index	Latitude	Longitude	Cluster
32	LOS ANGELES	0.0725	0.012505	34.053691	-118.242767	3
40	ORANGE	0.0725	0.009155	33.750038	-117.870493	3
43	PASADENA	0.0725	0.010903	34.147645	-118.144478	3
52	SAN DIEGO	0.0725	0.009250	32.717421	-117.162771	3
54	SAN JOSE	0.0725	0.012075	37.336191	-121.890583	3
56	SANTA ANA	0.0725	0.011227	33.749495	-117.873221	3
81	BOSTON	0.0625	0.009736	42.360253	-71.058291	3
82	CAMBRIDGE	0.0625	0.008119	42.375100	-71.105616	3
88	DETROIT	0.0600	0.022378	42.331551	-83.046640	3
95	RENO	0.0685	0.014195	39.529270	-119.813674	3
105	SEATTLE	0.0650	0.025904	47.603832	-122.330062	3

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
32	Sushi Restaurant	Coffee Shop	Japanese Restaurant
40	Mexican Restaurant	Convenience Store	Bar
43	American Restaurant	Coffee Shop	Pizza Place
52	Hotel	Mexican Restaurant	Italian Restaurant
54	Mexican Restaurant	Coffee Shop	Cocktail Bar
56	Mexican Restaurant	Fast Food Restaurant	Bar
81	Italian Restaurant	Seafood Restaurant	Historic Site
82	New American Restaurant	Coffee Shop	Pub
88	Coffee Shop	Bar	American Restaurant
95	Bar	Pub	Mexican Restaurant
105	Coffee Shop	Cocktail Bar	Seafood Restaurant

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
32	Plaza	Mexican Restaurant	Ramen Restaurant
40	Restaurant	Fast Food Restaurant	Coffee Shop
43	Italian Restaurant	Bar	Steakhouse
52	Café	Bar	American Restaurant
54	Bar	Sushi Restaurant	Sandwich Place

56	Convenience Store	Restaurant	Pharmacy
81	Park	Coffee Shop	Pizza Place
82	Café	Gastropub	Brewery
88	Restaurant	Diner	Park
95	Brewery	Coffee Shop	Café
105	Vietnamese Restaurant	American Restaurant	Sushi Restaurant

	7th Most Common Venue	8th Most Common Venue \
32	Ice Cream Shop	Bookstore
40	Pizza Place	Sandwich Place
43	Bakery	Cosmetics Shop
52	Coffee Shop	Seafood Restaurant
54	Theater	Pizza Place
56	Sandwich Place	Pizza Place
81	Sandwich Place	Market
82	Vegetarian / Vegan Restaurant	Portuguese Restaurant
88	Lounge	Steakhouse
95	Hotel	Breakfast Spot
105	Italian Restaurant	Breakfast Spot

	9th Most Common Venue	10th Most Common Venue
32	Bar	Mediterranean Restaurant
40	American Restaurant	Diner
43	Pub	Beer Garden
52	New American Restaurant	Burger Joint
54	Pub	Ice Cream Shop
56	American Restaurant	Coffee Shop
81	Bakery	Hotel
82	Ice Cream Shop	Tapas Restaurant
88	Burger Joint	Hotel
95	Steakhouse	Pizza Place
105	Hotel	Art Museum

Cluster 4

In [62]: city_full.loc[city_full['Cluster'] == 4, city_full.columns[[1] + list(range(2, city_full.shape[1]))]]

Out[62]:

	City	Tax	Crime Index	Latitude	Longitude	Cluster \
4	BERKELEY	0.0725	0.023042	37.870839	-122.272864	4

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue \
4	Chinese Restaurant	Japanese Restaurant	New American Restaurant

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue \
4	Thai Restaurant	French Restaurant	Café

	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue \
4	Yoga Studio	Ice Cream Shop	Coffee Shop

```

10th Most Common Venue
4      Pizza Place

```

In [63]: # create map

```
map_clusters = folium.Map(location=[39.83, -98.58], zoom_start=4)
```

```
# 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(cluster3['Latitude'], cluster3['Longitude'], cluster3['City'], cluster3['Cluster']):
```

```
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
```

```
    folium.CircleMarker(
```

```
        [lat, lon],
```

```
        radius=10,
```

```
        popup=label,
```

```
        color=rainbow[cluster],
```

```
        fill=True,
```

```
        fill_color=rainbow[cluster],
```

```
        fill_opacity=0.7).add_to(map_clusters)
```

```
map_clusters
```

Out[63]: <folium.folium.Map at 0x7f407436e4a8>

In [64]: cam = cluster3[cluster3['City'] == 'CAMBRIDGE']

```
cam
```

Out[64]:

	City	Tax	Crime Index	Latitude	Longitude	Cluster \
82	CAMBRIDGE	0.0625	0.008119	42.3751	-71.105616	3

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue \
82	New American Restaurant	Coffee Shop	Pub

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue \
82	Café	Gastropub	Brewery

	7th Most Common Venue	8th Most Common Venue \
82	Vegetarian / Vegan Restaurant	Portuguese Restaurant

	9th Most Common Venue	10th Most Common Venue
82	Ice Cream Shop	Tapas Restaurant

In [65]: cluster3.sort_values('Tax').head(5)

```

Out[65]:      City  Tax  Crime Index  Latitude  Longitude  Cluster \
88  DETROIT 0.0600   0.022378 42.331551 -83.046640      3
81  BOSTON 0.0625   0.009736 42.360253 -71.058291      3
82  CAMBRIDGE 0.0625   0.008119 42.375100 -71.105616      3
105 SEATTLE 0.0650   0.025904 47.603832 -122.330062      3
95  RENO 0.0685   0.014195 39.529270 -119.813674      3

      1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue \
88      Coffee Shop      Bar  American Restaurant
81      Italian Restaurant  Seafood Restaurant  Historic Site
82  New American Restaurant      Coffee Shop      Pub
105      Coffee Shop      Cocktail Bar  Seafood Restaurant
95      Bar      Pub  Mexican Restaurant

      4th Most Common Venue 5th Most Common Venue 6th Most Common Venue \
88      Restaurant      Diner      Park
81      Park      Coffee Shop      Pizza Place
82      Café      Gastropub      Brewery
105 Vietnamese Restaurant  American Restaurant  Sushi Restaurant
95      Brewery      Coffee Shop      Café

      7th Most Common Venue 8th Most Common Venue \
88      Lounge      Steakhouse
81      Sandwich Place      Market
82  Vegetarian / Vegan Restaurant  Portuguese Restaurant
105      Italian Restaurant      Breakfast Spot
95      Hotel      Breakfast Spot

      9th Most Common Venue 10th Most Common Venue
88      Burger Joint      Hotel
81      Bakery      Hotel
82      Ice Cream Shop      Tapas Restaurant
105      Hotel      Art Museum
95      Steakhouse      Pizza Place

```

```

In [66]: cluster3.sort_values('Crime Index').head(5)

```

```

Out[66]:      City  Tax  Crime Index  Latitude  Longitude  Cluster \
82  CAMBRIDGE 0.0625   0.008119 42.375100 -71.105616      3
40  ORANGE 0.0725   0.009155 33.750038 -117.870493      3
52  SAN DIEGO 0.0725   0.009250 32.717421 -117.162771      3
81  BOSTON 0.0625   0.009736 42.360253 -71.058291      3
43  PASADENA 0.0725   0.010903 34.147645 -118.144478      3

      1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue \
82  New American Restaurant      Coffee Shop      Pub
40  Mexican Restaurant      Convenience Store      Bar
52      Hotel      Mexican Restaurant      Italian Restaurant

```

81	Italian Restaurant	Seafood Restaurant	Historic Site
43	American Restaurant	Coffee Shop	Pizza Place
4th Most Common Venue 5th Most Common Venue 6th Most Common Venue \			
82	Café	Gastropub	Brewery
40	Restaurant	Fast Food Restaurant	Coffee Shop
52	Café	Bar	American Restaurant
81	Park	Coffee Shop	Pizza Place
43	Italian Restaurant	Bar	Steakhouse
7th Most Common Venue 8th Most Common Venue \			
82	Vegetarian / Vegan Restaurant	Portuguese Restaurant	
40	Pizza Place	Sandwich Place	
52	Coffee Shop	Seafood Restaurant	
81	Sandwich Place	Market	
43	Bakery	Cosmetics Shop	
9th Most Common Venue 10th Most Common Venue			
82	Ice Cream Shop	Tapas Restaurant	
40	American Restaurant	Diner	
52	New American Restaurant	Burger Joint	
81	Bakery	Hotel	
43	Pub	Beer Garden	

3.8 Results

3.8.1 City Selection

Of the 11 cities in the target cluster, the top 5 safest and 5 lowest state tax are cross referenced to select Cambridge, MA as the top choice to open a Dispensary followed by Boston, MA.

```
In [67]: # create map
cambridge = folium.Map(location=[42.375100,-71.105616], zoom_start=10)

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(cam['Latitude'], cam['Longitude'], cam['City'], cam['Cluster']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=20,
        popup=label,
        color='darkgreen',
        fill=True,
        fill_color='green',
        fill_opacity=0.7).add_to(cambridge)

cambridge
```

Out[67]: <folium.folium.Map at 0x7f40741b1a20>

3.9 Discussion

Upon review of the 5 different cluster datasets, I was able to determine which clusters contain the target demographic and culture that is becoming of a new dispensary such as an active nightlife scene from bars, pubs, liquor stores, activities, lots of restaurants such as fast food and convenience stores. The recommendation would be to focus your investors on opening a dispensary in Cambridge, MA followed by Boston, MA as a backup option or expansion location. These cities are both highly populated, very trendy, have an active nightlife, are relatively safe, and have a decent sales tax compared to other cities assessed.

3.10 Conclusion

Recreational marijuana is a rapidly materializing industry, estimated to surpass \$16 Billion in 2019 and exponentially growing. With the influx of new users and change of legal status, there is a huge demand for recreational marijuana with very limited supply. This is the prime chance to seize this opportunity and open a new legal marijuana supply chain in the United States. Based on the data reviewed and analyzed, Cambridge is the top city in the US to open the next Marijuana Dispensary based on its population, safety and culture.