

Julia Versinina

Applied Data Science Capstone, IBM Data Science Certification, Coursera
September 10, 2019

Capstone Project

Battle of Neighborhoods: Examining venues of Palm Beach county in sunny Florida

1. Introduction

1.1 Background

Many people in the United States move from one state to another for many reasons including but not limited to family bonds, professional opportunities, lower cost of living and overall pursuit of happiness. When people make the decision to move, they must gather relevant information about the new location. Most likely, they would like to keep some things they liked while living in an existing (old) location. Also, hopefully, they can solve some of the problems that were not desirable in an old setting.

1.2 Problem Definition

There are a lot of reasons for many people to love Chicago for cultural diversity, beautiful architecture, convenient public transportation, lots of professional opportunities and attractions such as museums, theaters and great dining options. There are several challenges of living in Chicago such as long cold winters and high property costs, including high property taxes. One of the ways to achieve a better life balance would be to seek for places in the United States that have warmer weather and lower cost of living. Florida combines both, warm weather and generally lower cost of living, when compared to Chicago. On the other hand, Florida has a completely different culture and style of living compared to the Midwest. People who consider moving to Florida should do their research to find the best fit for them. This project will be limited to analysis of venues data of neighborhoods of Palm Beach County in Florida, but a similar approach can be used to analyze any other county of Florida.

1.3 Target Audience

Target audience are people who consider moving to warmer states, such as Florida. This analysis will help target audience to learn about distribution of surrounding venues within Palm Beach County such as restaurants, health clubs, parks, and other venues.

2. Data

Following data will be used to solve defined problem:

- Zip Codes for the county of Palm Beach, Florida:
<https://www.zipcodestogo.com/Palm%20beach/FL/>
- US Zip Code Latitude and Longitude values:
<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=FL>
- Foursquare API to get the most common venues of given zip code of Palm Beach, Florida

3. Methodology & Analysis

We will use Pandas Python library to work with data. We will use collected data from various public web sources to convert it into Pandas dataframes for further clean up and manipulation. We will clean up Pandas dataframes by using existing methods built into pandas. We will combine several dataframes into one dataframe for final analysis. We will use several libraries for additional data retrieval and analysis. Specifically, we will use Geopy library to retrieve Latitude and Longitude values of Palm Beach county for each zip code. We will use Matplotlib and Folium libraries for visualizing data as maps. We will use Foursquare API to locate closest venues per any desired zip code. We will use Sklearn library to compute k-means and to identify clusters of areas per zipcode of one chosen city of Palm Beach county.

Step 1

Collect data with zip codes and cities of Palm Beach county of Florida. We ended up retrieving 74 entries total that are in Palm Beach county in Florida with either different zip code or city name.

	Zip Code	City	State	Zip Code Map
0	33401	West Palm Beach	Florida	View Map
1	33402	West Palm Beach	Florida	View Map
2	33403	West Palm Beach	Florida	View Map
3	33404	West Palm Beach	Florida	View Map
4	33405	West Palm Beach	Florida	View Map
5	33406	West Palm Beach	Florida	View Map
6	33407	West Palm Beach	Florida	View Map
7	33408	North Palm Beach	Florida	View Map
8	33409	West Palm Beach	Florida	View Map
9	33410	Palm Beach Gardens	Florida	View Map
10	33411	West Palm Beach	Florida	View Map
11	33412	West Palm Beach	Florida	View Map
12	33413	West Palm Beach	Florida	View Map
13	33414	West Palm Beach	Florida	View Map
14	33415	West Palm Beach	Florida	View Map
15	33416	West Palm Beach	Florida	View Map
16	33417	West Palm Beach	Florida	View Map
17	33418	Palm Beach Gardens	Florida	View Map
18	33419	West Palm Beach	Florida	View Map
19	33420	West Palm Beach	Florida	View Map
20	33421	West Palm Beach	Florida	View Map
21	33422	West Palm Beach	Florida	View Map
22	33424	Boynton Beach	Florida	View Map
23	33425	Boynton Beach	Florida	View Map
24	33426	Boynton Beach	Florida	View Map

Step 2

Retrieve Latitudes and Longitudes for each zip code:

	Unnamed: 0	Zipcode	Latitude	Longitude
0	NaN	33446	26.452473	-80.165090
1	NaN	33499	26.645895	-80.430269
2	NaN	33415	26.659344	-80.127040
3	NaN	33431	26.381304	-80.096230
4	NaN	33434	26.382408	-80.166990
5	NaN	33484	26.458152	-80.135600
6	NaN	33421	26.645895	-80.430269
7	NaN	33413	26.668944	-80.147210
8	NaN	33468	26.645895	-80.430269
9	NaN	33411	26.719596	-80.220770
10	NaN	33404	26.782114	-80.065280
11	NaN	33405	26.669744	-80.058500
12	NaN	33493	26.663097	-80.718270
13	NaN	33437	26.522249	-80.150310
14	NaN	33466	26.645895	-80.430269
15	NaN	33401	26.711192	-80.060430
16	NaN	33430	26.685125	-80.662790
17	NaN	33487	26.409953	-80.088410
18	NaN	33448	26.645895	-80.430269
19	NaN	33459	26.645895	-80.430269
20	NaN	33416	26.665394	-80.092864
21	NaN	33422	26.645895	-80.430269
22	NaN	33463	26.609226	-80.128740
23	NaN	33417	26.715596	-80.126200
24	NaN	33420	26.645895	-80.430269

Step 3

Combine two dataframes:

	City	Zipcode	Latitude	Longitude
0	West Palm Beach	33401	26.711192	-80.060430
1	West Palm Beach	33402	26.645895	-80.430269
2	West Palm Beach	33403	26.802139	-80.070320
3	West Palm Beach	33404	26.782114	-80.065280
4	West Palm Beach	33405	26.669744	-80.058500
5	West Palm Beach	33406	26.659294	-80.091180
6	West Palm Beach	33407	26.750991	-80.072960
7	North Palm Beach	33408	26.840684	-80.063120
8	West Palm Beach	33409	26.709575	-80.094430
9	Palm Beach Gardens	33410	26.839588	-80.088240
10	West Palm Beach	33411	26.710506	-80.220770

Step 4

Download dependencies for Geopy, Json, Matplotlib, sklearn KMeans and Folium:

```
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
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 # transform 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
import folium # map rendering library

print('Libraries imported.')
```


Step 5

Use Geopy library to get the latitude and longitude values of Palm Beach county:

```
address = 'Palm Beach county, FL'
geolocator = Nominatim(user_agent="palmbeach_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Palm Beach county are {}, {}'.format(latitude, longitude))
```

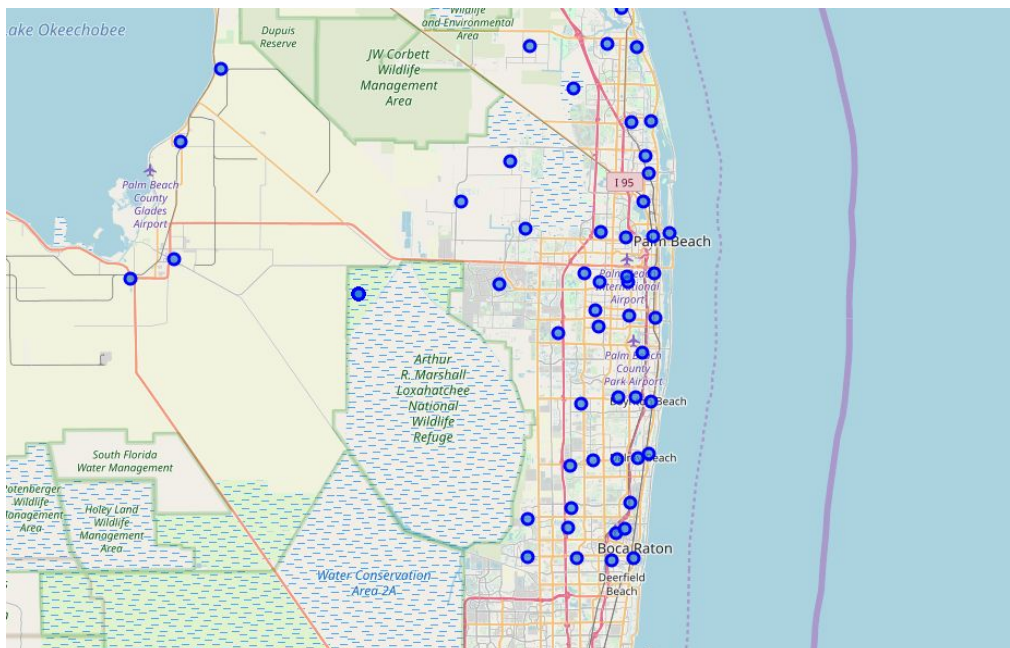
The geograpical coordinate of Palm Beach county are 26.6279798, -80.4494174.

Step 6

Create a map of Palm Beach county:

```
# create map of Palm Beach county using Latitude and Longitude values
map_palmbeach = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, city, zipcode in zip(df['Latitude'], df['Longitude'], df['City'], df['Zipcode']):
    label = '{} {}'.format(city, zipcode)
    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_palmbeach)
```



Step 7

Examine first zip code in the West Palm Beach town:

Get first zipcode:

```
In [17]: west_palm_beach_df.loc[0, 'Zipcode']
```

```
Out[17]: 33401
```

Get latitude and longitude values:

```
In [18]: zipcode_latitude = west_palm_beach_df.loc[0, 'Latitude'] # zipcode Latitude value
zipcode_longitude = west_palm_beach_df.loc[0, 'Longitude'] # zipcode Longitude value

zipcode_name = west_palm_beach_df.loc[0, 'Zipcode'] # zipcode name

print('Latitude and longitude values of {} are {}, {}'.format(zipcode_name,
                                                              zipcode_latitude,
                                                              zipcode_longitude))
```

Latitude and longitude values of 33401 are 26.711191999999997, -80.06043000000001.

Step 8

Get the top 100 venues that are in West Palm Beach within a radius of 500 meters:

nearby_venues

	name	categories	lat	lng
0	The Cheesecake Factory	American Restaurant	26.708346	-80.057215
1	Brightline West Palm Beach	Train Station	26.711554	-80.055798
2	Brio Tuscan Grille	Italian Restaurant	26.708645	-80.057064
3	Uptown Art	Art Gallery	26.711106	-80.055754
4	Cityplace	Shopping Mall	26.708155	-80.056851
5	Kravis Center for the Performing Arts, Inc.	Performing Arts Venue	26.706785	-80.059656
6	Blue Martini	Bar	26.709093	-80.056826
7	Rita's Italian Ice & Frozen Custard	Ice Cream Shop	26.709735	-80.057185
8	Publix	Grocery Store	26.710622	-80.057529
9	Sloan's Ice Cream	Ice Cream Shop	26.708334	-80.056669
10	City Pizza Italian Cuisine	Pizza Place	26.708656	-80.057837
11	Kapow! Noodle Bar	Asian Restaurant	26.713568	-80.056160
12	O'Shea's Irish Pub	Pub	26.713573	-80.056571
13	SEPHORA	Cosmetics Shop	26.708814	-80.056927
14	Ristorante Santucci	Italian Restaurant	26.713268	-80.057119
15	Wine Scene	Wine Bar	26.710628	-80.055783
16	Jamba Juice	Juice Bar	26.709642	-80.057043
17	Lost Weekend	Bar	26.713327	-80.056426
18	Keese's	Mediterranean Restaurant	26.713404	-80.056903
19	Victoria's Secret PINK	Lingerie Store	26.706917	-80.060949
20	H&M	Clothing Store	26.707889	-80.057319
21	AMC CityPlace 20	Movie Theater	26.709087	-80.056499
22	Brother Jimmy's BBQ	BBQ Joint	26.708445	-80.056643

Step 9

Repeat the same process to all the zipcodes in West Palm Beach. Create a new dataframe called `west_palm_beach_venues`. Check how many venues were returned for each zip code. Check how many unique categories can be curated from all the returned venues:

Zipcode	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
33401	23	23	23	23	23	23
33403	9	9	9	9	9	9
33404	4	4	4	4	4	4
33405	5	5	5	5	5	5
33407	8	8	8	8	8	8
33409	30	30	30	30	30	30
33411	1	1	1	1	1	1
33412	2	2	2	2	2	2
33413	2	2	2	2	2	2
33414	4	4	4	4	4	4
33415	4	4	4	4	4	4
33416	4	4	4	4	4	4
33417	4	4	4	4	4	4

How many unique categories can be curated from all the returned venues:

```
int('There are {} uniques categories.'.format(len(west_palm_beach_venues['Venue Category'].unique())))
```

There are 66 uniques categories.

Step 10

Analyze each zip code. Group rows by zip code and by taking the mean of the frequency of occurrence of each category:

Zipcode	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Bar	Bus Stop	Business Service	Cafe	Clothing Store	Coffee Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop	Dance Studio	Deli / Bodega	Diner	Discount Store	Electronics Store	Fast Food Restaurant	Food Truck	Furniture / Home Store	
0	33401	0.043478	0.043478	0.000000	0.043478	0.000000	0.043478	0.00	0.000000	0.086957	0.0	0.00	0.000000	0.043478	0.000	0.0	0.000	0.043478	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
1	33403	0.111111	0.111111	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.111111	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
2	33404	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.25	0.000000	0.000000	0.0	0.25	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
3	33405	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.200000	0.000000	0.000000	0.400000	0.00	0.000000	0.000000	0.00	0.000000
4	33407	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.125	0.0	0.125	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.125000	0.00	0.000000
5	33409	0.000000	0.000000	0.033333	0.000000	0.033333	0.033333	0.00	0.033333	0.000000	0.0	0.00	0.000000	0.033333	0.000	0.0	0.000	0.000000	0.033333	0.033333	0.033333	0.00	0.033333	0.033333	0.00	0.133333
6	33411	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	1.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
7	33412	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	1.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
8	33413	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
9	33414	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000
10	33415	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.25	0.000000	0.000000	0.25	0.000000
11	33416	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.250000	0.00	0.250000
12	33417	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000	0.0	0.000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000

Step 11

Print each zip code along with the top 5 most common venues. Put that into a pandas dataframe. Create function to sort the venues in descending order. Create the new dataframe and display the top 10 venues for each neighborhood:

	Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	33401	Italian Restaurant	Bar	Ice Cream Shop	Performing Arts Venue	Art Gallery	Asian Restaurant	BBQ Joint	Clothing Store	Cosmetics Shop	Grocery Store
1	33403	American Restaurant	Café	Gastropub	Market	Pizza Place	Gym / Fitness Center	Vietnamese Restaurant	Art Gallery	Thrift / Vintage Store	Asian Restaurant
2	33404	Light Rail Station	Grocery Store	Bakery	Business Service	Yoga Studio	Food Truck	Diner	Discount Store	Electronics Store	Fast Food Restaurant
3	33405	Diner	Yoga Studio	Cosmetics Shop	Park	Gym / Fitness Center	Grocery Store	Golf Course	Gastropub	Gas Station	Furniture / Home Store
4	33407	Pizza Place	Coffee Shop	Fast Food Restaurant	Sandwich Place	Convenience Store	Seafood Restaurant	Grocery Store	Furniture / Home Store	Food Truck	Gas Station

Step 12

Cluster Neighbourhoods by Zip Code. Run k-means to cluster the zip codes into 5 clusters. Create a new dataframe that includes the cluster as well as the top 10 venues for each zip code:

	City	Zipcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	West Palm Beach	33401	26.711192	-80.060430	0.0	Italian Restaurant	Bar	Ice Cream Shop	Performing Arts Venue	Art Gallery	Asian Restaurant
1	West Palm Beach	33402	26.645895	-80.430269	NaN	NaN	NaN	NaN	NaN	NaN	
2	West Palm Beach	33403	26.802139	-80.070320	0.0	American Restaurant	Café	Gastropub	Market	Pizza Place	Gym / Fitness Center
3	West Palm Beach	33404	26.782114	-80.065280	0.0	Light Rail Station	Grocery Store	Bakery	Business Service	Yoga Studio	Food Truck
4	West Palm Beach	33405	26.669744	-80.058500	3.0	Diner	Yoga Studio	Cosmetics Shop	Park	Gym / Fitness Center	Grocery Store

Step 13

Examine Clusters

Cluster 1:

Cluster 1											
west_palm_beach_merged.loc[west_palm_beach_merged['Cluster Labels'] == 0, west_palm_beach_merged.columns[[1] + list(range(5, west_palm_beach_merged.shape[1]))]]											
0	33401	Italian Restaurant	Bar	Ice Cream Shop	Performing Arts Venue	Art Gallery	Asian Restaurant	BBQ Joint	Clothing Store	Cosmetics Shop	Grocery Store
2	33403	American Restaurant	Café	Gastropub	Market	Pizza Place	Gym / Fitness Center	Vietnamese Restaurant	Art Gallery	Thrift / Vintage Store	Asian Restaurant
3	33404	Light Rail Station	Grocery Store	Bakery	Business Service	Yoga Studio	Food Truck	Diner	Discount Store	Electronics Store	Fast Food Restaurant
6	33407	Pizza Place	Coffee Shop	Fast Food Restaurant	Sandwich Place	Convenience Store	Seafood Restaurant	Grocery Store	Furniture / Home Store	Food Truck	Gas Station
7	33409	Furniture / Home Store	Pizza Place	Sports Bar	Italian Restaurant	Home Service	Fast Food Restaurant	Electronics Store	Latin American Restaurant	Diner	Deli / Bodega
11	33414	Pet Store	Grocery Store	Park	Playground	Electronics Store	Dance Studio	Deli / Bodega	Diner	Discount Store	Fast Food Restaurant
12	33415	Food Truck	Discount Store	Playground	Pawn Shop	Golf Course	Gastropub	Gas Station	Furniture / Home Store	Cosmetics Shop	Fast Food Restaurant
13	33416	Italian Restaurant	Fast Food Restaurant	Furniture / Home Store	Sandwich Place	Dance Studio	Deli / Bodega	Diner	Discount Store	Electronics Store	Food Truck
14	33417	Pharmacy	Music Venue	Pool	Pool Hall	Yoga Studio	Discount Store	Dance Studio	Deli / Bodega	Diner	Electronics Store

Cluster 2:

Cluster 2

```
west_palm_beach_merged.loc[west_palm_beach_merged['Cluster_Labels'] == 1, west_palm_beach_merged.columns[[1] + list(range(5, west_palm_beach_merged.shape[1]))]]
```

0]:

Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	33412	Construction & Landscaping	Yoga Studio	Food Truck	Deli / Bodega	Diner	Discount Store	Electronics Store	Fast Food Restaurant	Furniture / Home Store

Cluster 3:

Cluster 3

```
west_palm_beach_merged.loc[west_palm_beach_merged['Cluster_Labels'] == 2, west_palm_beach_merged.columns[[1] + list(range(5, west_palm_beach_merged.shape[1]))]]
```

1]:

Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	33411	Bus Stop	Yoga Studio	Food Truck	Deli / Bodega	Diner	Discount Store	Electronics Store	Fast Food Restaurant	Furniture / Home Store

Cluster 4:

Cluster 4

```
west_palm_beach_merged.loc[west_palm_beach_merged['Cluster_Labels'] == 3, west_palm_beach_merged.columns[[1] + list(range(5, west_palm_beach_merged.shape[1]))]]
```

42]:

Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	33405	Diner	Yoga Studio	Cosmetics Shop	Park	Gym / Fitness Center	Grocery Store	Golf Course	Gastropub	Gas Station

Cluster 5:

Cluster 5

```
west_palm_beach_merged.loc[west_palm_beach_merged['Cluster_Labels'] == 4, west_palm_beach_merged.columns[[1] + list(range(5, west_palm_beach_merged.shape[1]))]]
```

1]:

Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	33413	Golf Course	Other Repair Shop	Yoga Studio	Cosmetics Shop	Gym / Fitness Center	Grocery Store	Gastropub	Gas Station	Furniture / Home Store

4. Results & Discussion

Based on the results, we can observe that Cluster 1 returned much more zip codes than Cluster 2, Cluster 3, Cluster 4 and Cluster 5. We can see that Cluster 1 includes nine unique zip codes: 33401, 33403, 33404, 33407, 33409, 33414, 33415, 33416, and 33417. Cluster 1 returned a lot of fun and convenient venues such as various restaurants, entertainment places and retail stores. Cluster 2, Cluster 3, Cluster 4 and Cluster 5 only have one zip code each with just a few venues such as construction & landscaping, discount store and park venues. Based on these observations, if I would consider moving from the city of Chicago to Florida, West Palm Beach could be a good candidate. I would specifically look into zip codes from Cluster 1 since it seems to offer a lot of things that I already like in Chicago such as great dining and access to health clubs.

5. Conclusion

Similar calculations and analysis can be performed on other counties and zip codes in Florida or other states to examine clusters based on venues. This sort of analysis can help people to decide what specific locations and zip codes have ideal set of venues in order to plan to move to another state.