Mexican Restauranteurs Ideal Location in Florida

by Brian Beedle

March 14th, 2020

Introduction:

A client would like to relocate their gourmet Mexican restaurant to Florida and would like to locate the most desirable communities for their endeavor. They desire a beach community on the east coast of Florida but are not interested in south Florida. The location needs to be a community that can sustain a new gourmet Mexican restaurant. The client believes an affluent densely populated community, that does not already contain a gourmet Mexican restaurant to be ideal.

The demographics data and restaurants data, could be valuable to any aspiring restauranteur on the east coast of Florida. In addition, the script could quickly and easily be changed to find the ideal location for any style of food establishment.

Data:

A google search of "florida east coast cities list" led to the website https://sciencetrends.com/map-of-florida-east-coast-beaches-and-cities/ , which contains a table with two columns, a column of coastal cities on the east coast of Florida and a region column. The region column has the elements "Northeastern Coast of Florida," "Central East Coast of Florida," and "Southern East Coast of Florida." The table was scraped eliminating cities in the region "Southern East Coast of Florida." The remaining cities were added to a list.

```
CityList = ['American Beach','Atlantic Beach','Fernandina
Beach','Flagler Beach','Jacksonville Beach','Neptune
Beach','Palm Coast','Ponte Vedra','St Augustine','Vilano
Beach','Cocoa Beach','Daytona Beach','Indian Harbour','Jensen
Beach','Melbourne Beach','New Smyrna Beach','Ormond
Beach','Playalinda Beach','Satellite Beach','Stuart','Vero
Beach’]
```

Then a list of zip codes for each city was found using the python package "uszipcode." For each zip code, the location coordinates and boundary coordinates were retrieved. The "uszipcode" python package, also, provides demographic data for each zip code and the population, population density, median income, and median home value for each zip code was retrieved. The land area in square miles for each zip code was considered, but this variable was reflected by the population density variable, which is the zip code population divided by the zip code land area. Therefore, it was decided not to use the land area variable.


```
[SimpleZipcode(zipcode='32233', zipcode_type='Standard',
major_city='Atlantic Beach', post_office_city='Atlantic Beach,
FL', common_city_list=['Atlantic Beach', 'Atlantic Bch',
'Jacksonville', 'Mayport'], county='Duval County', state='FL',
lat=30.36, lng=-81.42, timezone='Eastern', radius_in_miles=3.0,
area_code_list=['904'], population=23673,
population_density=2323.0, land_area_in_sqmi=10.19,
water_area_in_sqmi=3.16, housing_units=10128,
occupied_housing_units=8858, median_home_value=211100,
median_household_income=50338, bounds_west=-81.459709,
bounds_east=-81.390331, bounds_north=30.399843,
bounds_south=30.323343)]
```


The data required some cleaning. As "American Beach" returned a duplicate zip code to a zip code returned for "Fernandina Beach." A google search showed that "American Beach" is a common city name for a part of the major city name "Fernandina Beach." Therefore, "American Beach" is not included when the data was cast to a pandas data frame. "Palm Coast" returned three zip codes one of which insert zip code contained no geographical or population data. Therefore the zip code insert zip code was dropped from the data frame. Finally, "Vilano Beach" returned a zip code on the west coast of Florida. A google search showed that "Vilano Beach" was actually a common city name for part of the major city name "St. Augustine." A google search showed that the zip code for "Vilano Beach" was contained in the zip codes returned for "St. Augustine." Therefore, "Vilano Beach" was removed from the data frame.

Finally, the Foursquare API was used to return the number of restaurants in each zip code. The Foursquare API was, also, used to return the number of Mexican restaurants in each zip code.

Sample on next page.

```
{'meta': {'code': 200, 'requestId': '5e65a93978a484001b948ad5'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'},
    {'name': '$-$$$$', 'key': 'price'}]},
  'headerLocation': 'Jacksonville',
  'headerFullLocation': 'Jacksonville',
  'headerLocationGranularity': 'city',
  'query': 'restaurant',
  'totalResults': 78,
  'suggestedBounds': {'ne': {'lat': 30.399843, 'lng':
-81.390331},
   'sw': {'lat': 30.323343, 'lng': -81.459709}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
```

Methodology:

First, the data returned from the python package “uszipcode” was cast to a pandas data frame, “df.”

	City	Zipcode	Latitude	Longitude	South Boundry	West Boundry	North Boundry	East Boundry	Population	Population Density	Median Income	Median Home Value
0	Atlantic Beach	32233	30.36	-81.42	30.323343	-81.459709	30.399843	-81.390331	23673.0	2323.0	50338.0	211100.0
1	Fernandina Beach	32034	30.60	-81.50	30.506896	-81.574418	30.711258	-81.422796	31008.0	547.0	62932.0	254400.0
2	Flagler Beach	32136	29.47	-81.14	29.401590	-81.192068	29.555170	-81.094828	7080.0	375.0	48767.0	208100.0
3	Jacksonville Beach	32250	30.28	-81.42	30.252243	-81.453815	30.307704	-81.376317	25356.0	2944.0	56466.0	247000.0
4	Neptune Beach	32266	30.32	-81.41	30.306941	-81.438570	30.324647	-81.387568	7037.0	3018.0	67045.0	315300.0

The zip code and zip code boundary coordinates from the “df” data frame were then used to search the Foursquare API to return the number of restaurants and Mexican restaurants for each zip code of interest, this data was cast to a second data frame, “df2.”

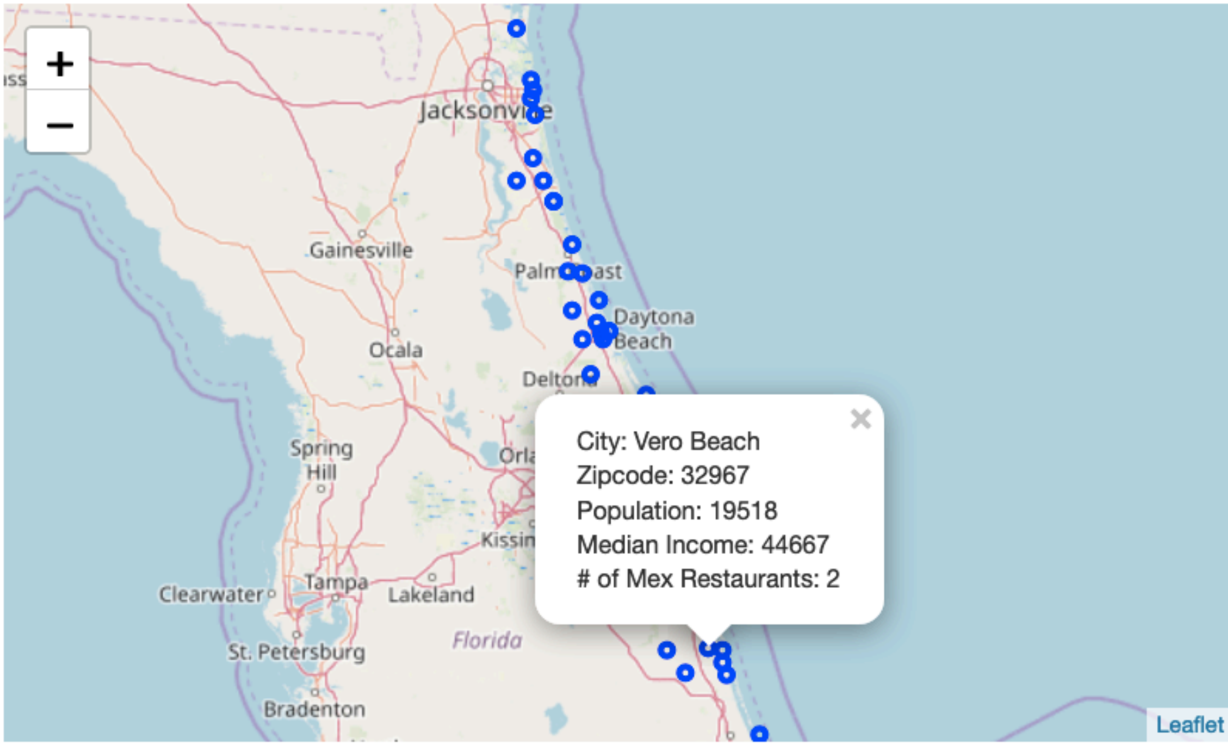
“df2” Sample on next page, full version at the end of the report

	Zipcode2	Restaurant Total	Mexican Restaurants
0	32233	78	9
1	32034	100	15
2	32136	45	3
3	32250	104	20
4	32266	29	3

These two data frames, “df” and “df2” were merged together into one data frame, “df3.” The data frame “df3,” then contained two zip code columns, one from each of the previous data frames. The zip codes were verified to match to make sure the data for each zip code row was correct. The second zip code column was then dropped, along with the zip code boundary coordinate columns were also dropped as they were no longer needed. This resulted in the data frame, “df4.”

	City	Zipcode	Population	Population Density	Median Income	Median Home Value	Restaurant Total	Mexican Restaurants
0	Atlantic Beach	32233	23673.0	2323.0	50338.0	211100.0	78	9
1	Fernandina Beach	32034	31008.0	547.0	62932.0	254400.0	100	15
2	Flagler Beach	32136	7080.0	375.0	48767.0	208100.0	45	3
3	Jacksonville Beach	32250	25356.0	2944.0	56466.0	247000.0	104	20
4	Neptune Beach	32266	7037.0	3018.0	67045.0	315300.0	29	3

The folium python package was then used to generate a map that includes markers for each zip code of interest. Each marker includes a pop up that when clicked upon displays the city name, population, and median income for that zip code. This provided the client an easy way to browse the zip codes locations and corresponding data.

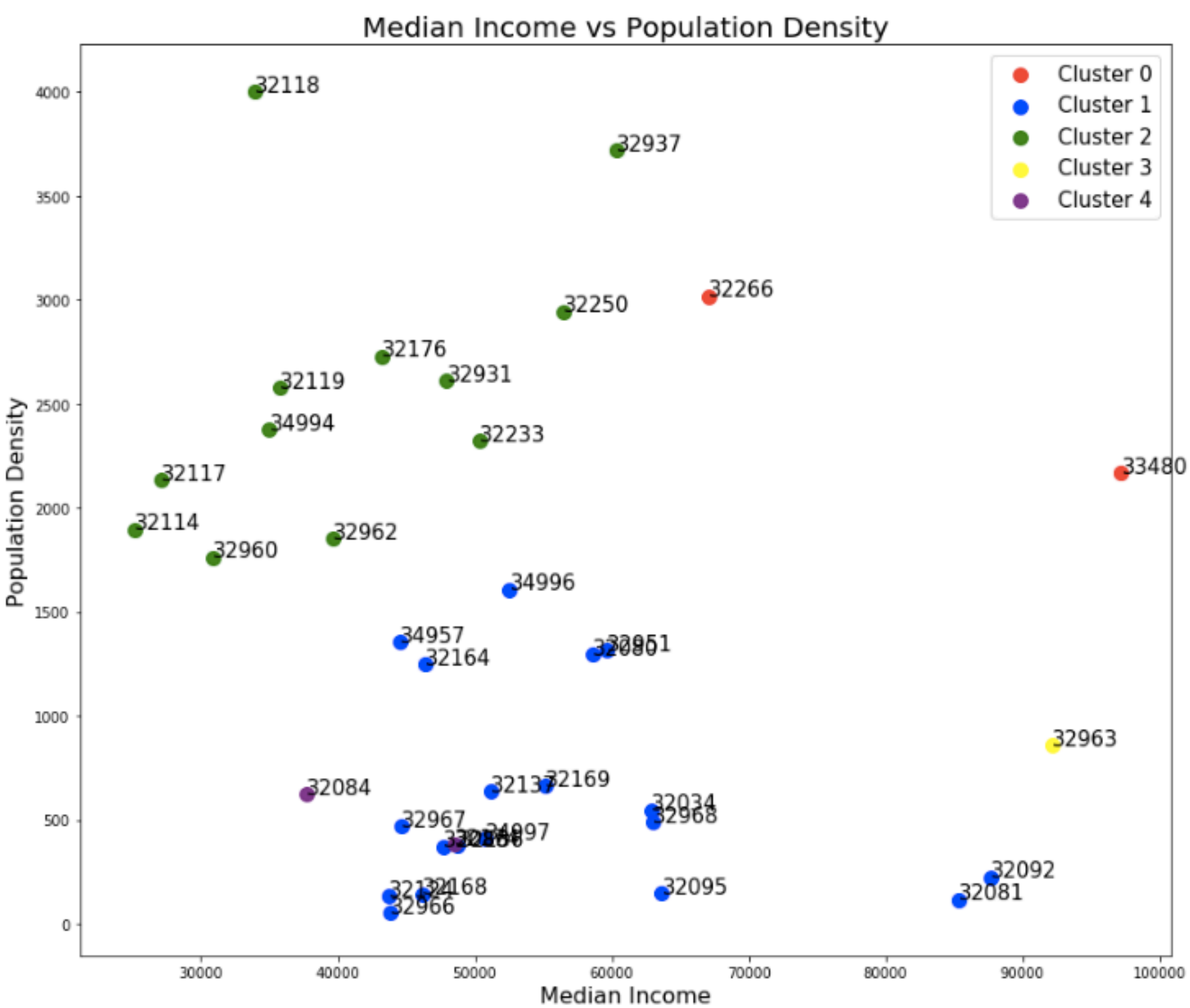


Next, a feature set data frame, “featureset,” was created that contained the variables of interest for agglomerate hierarchical clustering. “feature set”.

	Population	Population Density	Median Income	Median Home Value	Restaurant Total	Mexican Restaurants
0	23673.0	2323.0	50338.0	211100.0	78	9
1	31008.0	547.0	62932.0	254400.0	100	15
2	7080.0	375.0	48767.0	208100.0	45	3
3	25356.0	2944.0	56466.0	247000.0	104	20
4	7037.0	3018.0	67045.0	315300.0	29	3

The feature set included the population and population density, as a higher population and population density would indicate more potential customers and provide an indication of the vicinity of these potential customers. The feature set, also, included the median income and median home value, which were used to identify more affluent communities. A more affluent community would have the monetary resources to support a gourmet restaurant, which tends to be more expensive than a casual Mexican restaurant. Finally, the feature set contained the number of restaurants and number of Mexican restaurants in each zip code. These variables can help the client determine the level of competition and specifically the number of directly competing Mexican restaurants in the zip code.

The feature data set was then clustered using an agglomerative hierarchical approach. The number of clusters was set to five and the linkage attribute was set to average. The cluster label for each zip code was then added back to our data frame, “df4.” This was used to create a scatter plot of zip codes for their median income versus their population density with a legend indicating what cluster the plotted zip code belongs too.



The data frame, “df4,” was then grouped by cluster with the columns from the “featureset” data frame, resulting in the data frame, “zips_clus.” This data frame contained the average population, population density, median income, median home value, restaurant total, and Mexican restaurant total for each cluster. This allowed for the evaluation of each cluster.

cluster_	Population	Population Density	Median Income	Median Home Value	Restaurant Total	Mexican Restaurants
0	8293.000000	2594.500000	82126.5	486800.000000	38.500000	2.000000
1	19668.736842	610.842105	55582.0	210457.894737	55.000000	6.368421
2	21169.000000	2577.916667	40467.0	160525.000000	95.916667	10.750000
3	14911.000000	859.000000	92137.0	489900.000000	100.000000	20.000000
4	38640.500000	504.500000	43107.0	162350.000000	158.500000	24.500000

After the most ideal cluster was identified, “cluster 0.” Each zip code in the cluster was searched again by their bounding coordinates, identifying the existing Mexican restaurants in each zip code and their Foursquare restaurant ID. Then the Foursquare API was used again to search each Mexican restaurant in each zip code in “cluster 0.” For each potential competing Mexican restaurant, the restaurant name, rating, website url, and price tier were retrieved if available. This data was cast to the data frame “df20” and “df22.”
df20:

	Restaurant Name	Restaurant Rating	Restaurant Website	Restaurant Price
0	Flying Iguana	8.9	http://flyingiguana.com	2
1	Tijuana Flats	rating not available	http://www.tijuanaflats.com	1
2	Atomic Flying Fish Taco Grill	rating not available	url not available	1

df22:

	Restaurant Name	Restaurant Rating	Restaurant Website	Restaurant Price
0	Coyo Taco	7.9	http://www.coyo-taco.com	1

Results:
"Cluster 0" was identified as the most ideal cluster to open a gourmet Mexican restaurant. "Cluster 0" had the highest population density of all clusters, it was only 17 more than the next closest cluster, "Cluster 2." However, "Cluster O's" population density was more than three times that of any of the remaining clusters. "Cluster 0" had the second highest median income and was only ten thousand dollars less than the highest median income of "Cluster 3."   "Cluster 0's" median income was more twenty six thousand dollars more than the third highest median income and at least thirty nine thousand more than the remaining clusters. Furthermore, "Cluster 0" had an average of only two Mexican restaurants and thirty nine restaurants all together, these were by far the lowest of all clusters. The median home values supported the resulting interpretation of the median incomes. The population was the lowest of all clusters, but this was deemed to be the least important variable and disregarded in respect to the population density and affluential nature of the cluster.
"Cluster 0" contained two zip codes, the first 32266 which is the city of Neptune Beach. Zip code 32266 contained three Mexican restaurants, with only one returning a rating. That rating was for "The Flying Iguana" and was 8.9 on a scale of 0 to 10. "The Flying Iguana" had a price tier of 2 on a scale of 1 to 4, 1 being the lowest priced and 4 being the highest priced restaurant. The other two Mexican restaurants only had a pricing tier of 1 and would not be considered significant competition for our client's gourmet Mexican Restaurant.
The second zip code in "Cluster 0" was 33480, which is the city of Playalinda Beach. Zip code 33480 had only one Mexican restaurant, "Coyo Taco." "Coyo Taco" had a rating of 7.9 and a pricing tier of 1, which again would not be considered competition for the client's gourmet restaurant.

	City	Zipcode	Population	Population Density	Median Income	Median Home Value	Restaurant Total	Mexican Restaurants	cluster_
4	Neptune Beach	32266	7037.0	3018.0	67045.0	315300.0	29	3	0
25	Playalinda Beach	33480	9549.0	2171.0	97208.0	658300.0	48	1	0

Discussion:

It was important to provide the client with a map to browse all zip codes of interest, and the demographics for all zip codes of interest, which are contained in data frame “df4.” It was important to provide this data to the client as the client will be living in or near the zip code chosen to open the gourmet Mexican restaurant. Therefore, the client’s decision could be influenced by factors beyond what would be considered in just the ideal location for the restaurant.

Conclusion:

Zip code 32266 in the city of Neptune Beach, Florida or zip code 33480 in the city of Playalinda Beach, Florida would be the ideal community to open a gourmet Mexican restaurant, based upon the client’s preferences.

Zip codes of Interest with Variables of interest, df4:

	City	Zipcode	Population	Population Density	Median Income	Median Home Value	Restaurant Total	Mexican Restaurants
0	Atlantic Beach	32233	23673.0	2323.0	50338.0	211100.0	78	9
1	Fernandina Beach	32034	31008.0	547.0	62932.0	254400.0	100	15
2	Flagler Beach	32136	7080.0	375.0	48767.0	208100.0	45	3
3	Jacksonville Beach	32250	25356.0	2944.0	56466.0	247000.0	104	20
4	Neptune Beach	32266	7037.0	3018.0	67045.0	315300.0	29	3
5	Palm Coast	32137	37821.0	638.0	51153.0	198900.0	93	9
6	Palm Coast	32164	41616.0	1248.0	46405.0	157000.0	69	5
7	Ponte Vedra	32081	4524.0	118.0	85354.0	326800.0	46	5
8	St Augustine	32080	20165.0	1296.0	58648.0	293500.0	48	7
9	St Augustine	32084	29729.0	624.0	37670.0	155200.0	133	29
10	St Augustine	32086	24546.0	373.0	47670.0	169800.0	69	12
11	St Augustine	32092	28242.0	224.0	87640.0	240300.0	44	8
12	St Augustine	32095	7302.0	147.0	63651.0	149700.0	42	6

13	Cocoa Beach	32931	13567.0	2610.0	47952.0	234600.0	71	10
14	Daytona Beach	32114	32084.0	1896.0	25150.0	106100.0	174	16
15	Daytona Beach	32117	24170.0	2137.0	27106.0	89400.0	68	3
16	Daytona Beach	32118	16961.0	4004.0	33942.0	166400.0	153	14
17	Daytona Beach	32119	20189.0	2579.0	35727.0	120400.0	74	10
18	Daytona Beach	32124	5986.0	133.0	43698.0	214700.0	45	5
19	Jensen Beach	34957	22257.0	1354.0	44487.0	202600.0	62	6
20	Melbourne Beach	32951	10673.0	1314.0	59604.0	276500.0	20	1
21	New Smyrna Beach	32168	23642.0	144.0	46200.0	173800.0	61	5
22	New Smyrna Beach	32169	9816.0	665.0	55101.0	286600.0	96	6
23	Ormond Beach	32174	47552.0	385.0	48544.0	169500.0	184	20
24	Ormond Beach	32176	14339.0	2729.0	43166.0	187800.0	97	8
25	Playalinda Beach	33480	9549.0	2171.0	97208.0	658300.0	48	1
26	Satellite Beach	32937	24969.0	3719.0	60313.0	222800.0	70	9
27	Stuart	34994	15967.0	2378.0	35000.0	118300.0	149	17
28	Stuart	34996	10649.0	1609.0	52533.0	183300.0	78	4
29	Stuart	34997	40405.0	411.0	50716.0	180500.0	44	10
30	Vero Beach	32960	20037.0	1763.0	30832.0	119700.0	90	11
31	Vero Beach	32962	22716.0	1853.0	39612.0	102700.0	23	2
32	Vero Beach	32963	14911.0	859.0	92137.0	489900.0	100	20
33	Vero Beach	32966	15626.0	54.0	43825.0	112500.0	54	11
34	Vero Beach	32967	19518.0	468.0	44667.0	178700.0	18	2
35	Vero Beach	32968	12830.0	488.0	63007.0	191000.0	11	1