|  |  |
| --- | --- |
| **A view of a city with a mountain in the background  Description automatically generated** | **Moving to Palm Beach County Florida**  **Data Analysis** |

## As I find myself in the process of moving to Palm Beach County, Florida, I decided to use this as my Capstone Project - Battle of the Neighborhoods – for the IBM Data Science Professional Certificate.

## Introduction: Business Problem

Currently trending, many active seniors, especially those who live in states like New York and New Jersey, once retired, look to relocate to a warmer, less expensive location like Florida. They would like to move to an area where they can still enjoy many of the same activities in their new home as they enjoyed in their northern one, while maintaining a comparable quality of life in terms of affordability and demographics.

An important consideration for active seniors is to move to an area that is close to family and friends, as well as an area that their guests, including grandchildren, can enjoy visiting. For this study, that area will be limited to Palm Beach County, Florida. Palm Beach County is a large area - 2,386 square miles - on the east coast of Florida consisting of 39 municipalities.  
  
Although there are many factors to consider when moving to a new neighborhood such as available social activities, recreational activities, access to healthcare, as well as the crime rate, taxes, affordability, and public amenities including transportation, this study will focus on venue data collected by FOURSQUARE.  
  
FOURSQUARE provides a variety of categories of venues such as restaurants, shopping, and leisure activities. This study will cluster neighborhoods (municipalities) in Palm Beach County, Florida by venues collected from FOURSQUARE. In addition, median income, median property value, population density, and area (square miles) data will be collected and analyzed.  
  
The goal of the analysis is to provide seniors with information that will help them make a better determination as to which areas they should explore further when considering relocation to Palm Beach County, Florida.

## Data

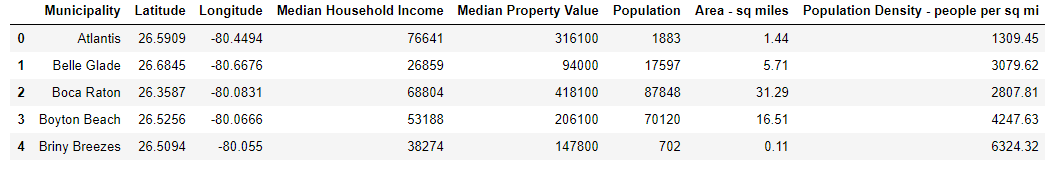
I collected data from the following sources:

* geopy.geocoders Nominatim from which I retrieved latitude and longitude for Palm Beach County and each municipality in order to display Folium maps of Palm Beach County and its municipalities.

* I retrieved Palm Beach County municipality demographic information from the homefacts.com website. Data retrieved includes Median Income, Median Property Value, Population, Population Density, and Area. These values are from 2016.
* I called the FOURSQUARE API to retrieve venues trending at the time of execution including restaurants, shopping, leisure activities, etc. for Palm Beach County neighborhoods (municipalities).

I used pandas and python to perform the analysis for my project.

The data was read from an Excel file into the following pandas DataFrame.



*Figure 1. pandas DataFrame of data read from an Excel file (first 5 rows)*

## Methodology

**To visualize Palm Beach County and its municipalities,** I generated two Folium maps. The first marks the location of Palm Beach County within the state of Florida; the second marks each municipality within Palm Beach County. The second map also incorporates popups with demographic information.

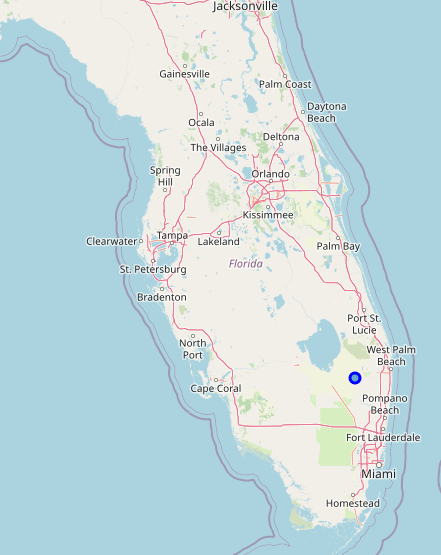
I created histograms and bar charts to visualize demographics of each of the municipalities as they relate to each other and to the whole.

The goal of this project is to group municipalities that have similar venues. Due to the nature of the municipality data, labeling is not possible; and therefore, requires an unsupervised method of analysis. I chose KMeans for my analysis because it is an unsuperivised algorithm that clusters data into groups based upon their similarities.

### Display Maps of Palm Beach County and Neighborhoods

## Palm Beach County Map

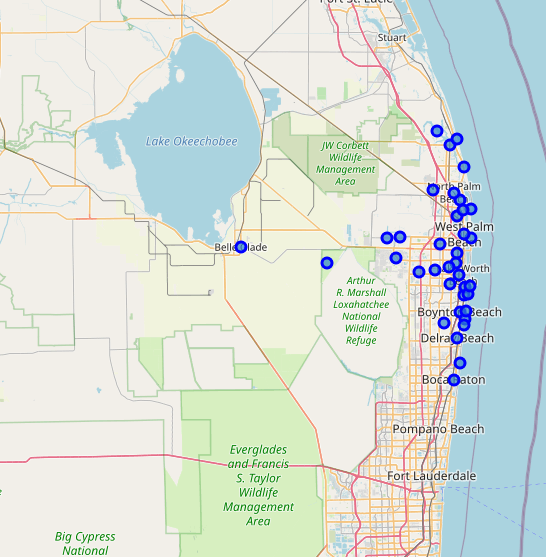
The first of two maps places Palm Beach County in context to its location in Florida. I called Nominatim to get the latitude and longitude of Palm Beach County which were then used to generate the following Folium map.



#### Figure 2. Florida map with Palm Beach County identified by the blue dot

## Palm Beach County Neighborhoods Map

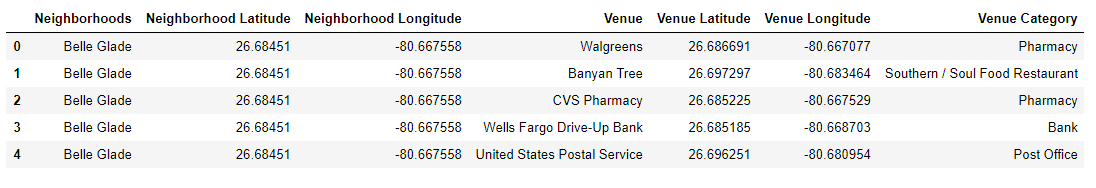
The second map places neighborhoods (by municipality) in context to their locations within Palm Beach County. In the same manner as the first map, I obtained the latitude and longitude of each municipality in Palm Beach County from Nominatim so that Folium could generate the following map.



*Figure 3. Palm Beach County Neighborhoods Map with each municipality identified by a blue dot*

### FOURSQUARE API

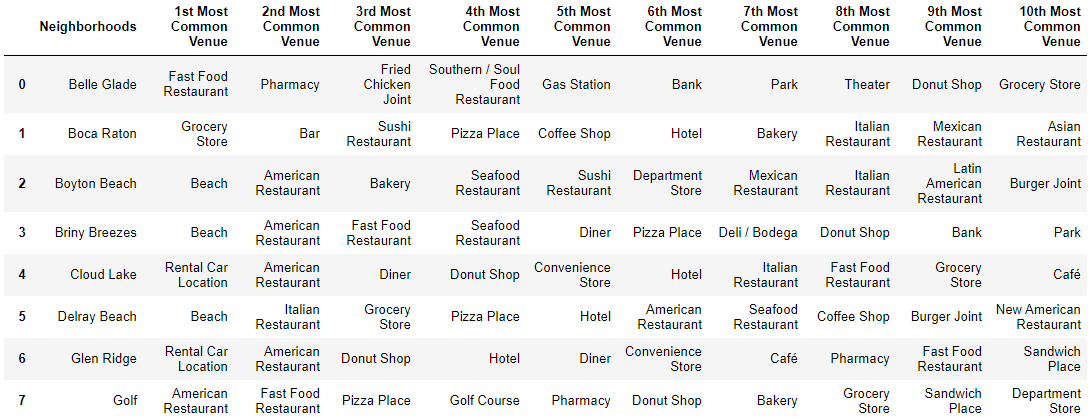
I called the FOURSQUARE API to retrieve up to 100 venues with a radius of one and a half miles of the latitude and longitude of each neighborhood. The radius was calculated as the average square mile area of all municipalities. The API retrieved 248 unique venue categories. The first five rows of the retrieved data:



*Figure 4. pandas DataFrame of neighborhood venues retrieved by FOURSQUARE API (first 5 rows)*

Most of the municipalities reached the 100 venue limit or came close to it, but a quarter of them did not. Those that did not are at the lower end in terms of population density. One municipality; however, Atlantis, did not return any venues.

The FOURSQUARE API results are dependent upon FOURSQUARE user input at the time the API is run. The pandas DataFrame below shows the ten most common venues for each municipality. At the time the API was run, Florida had recently opened from the lockdown from COVID-19. Palm Beach County was a hot spot in Florida at the time; and therefore, was more restricted in terms of opening businesses. The venues listed reflect this situation as grocery stores, restaurants with takeout, banks, and parks and beaches predominate.

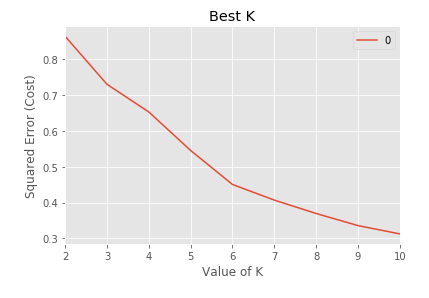
 



*Figure 5. pandas DataFrame of top ten venues retrieve from FOURSQUARE API for municipalities*

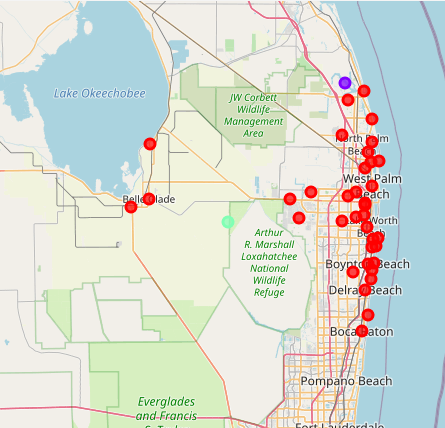
### K Nearest Neighbor

I chose sklearn.cluster.KMeans in the scikit-learn library to perform a K Nearest Neighbor unsupervised algorithm to cluster neighborhoods based upon their shared top venue types. First, I found the best k by executing KMeans from 2 to 10 values for k, calculated the squared error and graphed the result. The best value for k is 3 as indicated by the first ‘elbow’ on the graph.



*Figure 6. Graph indicating best K as 3*

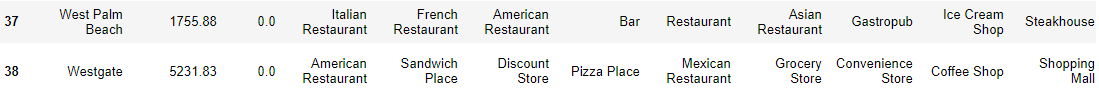
I then ran KMeans again with k set to 3 to produce Cluster Labels for each neighborhood. The resulting three clusters can be characterized by their main venues. The clusters are visualized with the following Folium map.



*Figure 7. Cluster Map: Cluster 0 in red, Cluster 1 in purple, Cluster 3 in green*

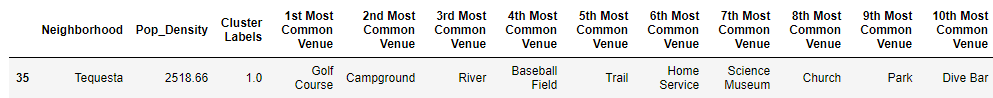
By examining the top ten venues in each cluster, I defined them as follows:

**CLUSTER 0** - Residential - with a variety of restaurants and other eateries

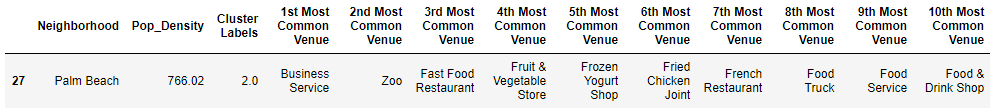
*Figure 8. Top nine venues for Cluster 0*

**CLUSTER 1** - Outdoor leisure activities - golfing, camping, boating - and parks



*Figure 9. Top ten venues for Cluster 1*

**CLUSTER 2** - Food and Business



*Figure 10. Top 10 venues for Cluster 2*

Results

### KMeans

The best value for K is 3 according to the calculation of cost or squared error when plotted.

Cluster 0 includes 36 neighborhoods which includes all neighborhoods except for the two that are in Clusters 1 and 2. I characterized this cluster as residential. It is made up of a wide variety of restaurants and other eateries as well as grocery stores, banks and beaches.

Cluster 1 includes only the municipality of Tequesta. I called this cluster outdoor leisure activities because the most common venues are golfing, camping, boating, and baseball. Also included are science museum, church, and parks.

Cluster 2 includes only the municipality of Palm Beach. I characterized this cluster as business and food. The top venues are business services, zoo, and eateries.  
  
Venues were not retrieved for the municipality of Atlantis; and therefore, was not included in a cluster.

### Histograms and Bar Charts

I created histograms and bar charts with matplotlib.pyplot to visualize median income, median property values, and population density in Palm Beach County.

Median Income

The histogram of Median Income shows three groups.

|  |  |  |
| --- | --- | --- |
| **Groups** | **Median Income** | **Neighborhoods** |
| 1 - Lowest | $27,000 to $81,000 | Remaining 32 Neighborhoods |
| 2 – Medium | $83,000  $100,000  $115,000 | Wellington  Highland Beach  Palm Beah |
| 3 – High | $163,000  $183,000  $189,000  $209,000 | Jupiter Inlet Colony  Gulf Stream  Golf  Manalapan |

Median Property Values

The histogram of Median Property Values shows four groups.

|  |  |  |
| --- | --- | --- |
| **Groups** | **Median Property Values** | **Neighborhoods** |
| 1 - Lowest | $77,000 to $370,000 | Remaining 28 Neighborhoods |
| 2 - Medium | $418,000 to $687,000 | Boca Raton  Palm Beach Shores  Highland Beach  Ocean Ridge |
| 3 – Medium-High | $956,000 to $1.1 million | Palm Beach  Golf  Jupiter Inlet |
| 4 - High | over $1.6 million | Gulf Stream  Manalapan |

Population Density

The histogram of Population Density shows three groups.

|  |  |
| --- | --- |
| **Group** | **Neighborhoods** |
| Group 1 – Low | Manalapan  Loxahatchee Groves  Golf |
| Group 2 – Medium | Remaining Neighborhoods |
| Group 3 – High | Westgate  Lake Worth  Palm Springs  Briny Breezes  Greenacres |

Area

The histogram of Area shows there are four major neighborhoods in terms of square mile area.

|  |  |  |
| --- | --- | --- |
| **Groups** | **Area (sq mil)** | **Neighborhoods** |
| 1 – Very Small | Under 7 | Remaining 27 Neighborhoods |
| 2 – Small | 10 to about 17 | Riviera Beach  Manalapan  Palm Beach  Royal Palm Beach  Loxahatchee Groves  Delray Beach  Boynton Beach  Jupiter |
| 3 – Medium | 31 | Boca Raton |
| 4 – Medium-High | 45 | Wellington |
| 5 – Large | 55 – 58 | Palm Beach Gardens  West Palm Beach |

## Discussion

Palm Beach County is diverse in terms of income levels, and property values. Palm Beach County median income ranges from below the national average of $57,617 to well above it. Similarly, Palm Beach County median property values ranges from below the national average of $205,000 to well above it. As a result, Palm Beach County offers home affordability for all income levels.  
  
Most municipalities are relatively small in terms of area in square miles. And, most of Palm Beach County has a higher population density than the national average of 92.9 residents per square mile of land area.  
  
According to K Means, Palm Beach County can be segregated into three main clusters. Two of these clusters have only one municipality (Tequesta, and Palm Beach), and the third cluster has the rest of the municipalities.  
  
Tequesta (Cluster 0) is the northernmost municipality in Palm Beach County, situated between the Atlantic Ocean and the Loxahatchee River, with an active recreational department, lots of golf courses, and many campgrounds.  
  
Palm Beach (Cluster 1) is an island centrally located in Palm Beach County with the Atlantic Ocean to the east and the Intracoastal Waterway to the west. It is just across the waterway to West Palm Beach; and as such, the venues cited are more related to West Palm Beach than Palm Beach. For example, zoo is the second most common venue, but there is not a zoo on Palm Beach; rather, it is in West Palm Beach. In addition, fast food restaurants, the third most common venue are also not found on Palm Beach. Palm Beach is mostly residential and home to the wealthy and upscale dining and shopping. Notable examples are Mar-a-Lago and Worth Avenue.  
  
The third cluster (Cluster 2) groups municipalities that are inland with those that are on the coast. Inland municipalities are mainly residential whereas coastline municipalities, while also residential, contribute to Florida’s tourism economy with their beaches and hotels. The most popular venues in this cluster are restaurants, grocery stores, and banks.  
  
The clustering is skewed by COVID-19. Although Florida was one of the first states to open from the lockdown, people mainly patronize grocery stores, and banks as well as restaurants and other eateries, and the beach. The tourist industry in Florida has greatly suffered as people are not traveling in general, and specifically to Florida. Once the economy has recovered, the K Means clustering should be more representative of Palm Beach County municipalities.

## Conclusion

This project is meant to give active seniors who plan to relocate an overview of Palm Beach County, Florida information that will help them determine which areas they should explore. The data shows that Palm Beach County allows affordability to a wide range of incomes and desired property values. According to K Means, the county is fairly homogeneous in terms of the most common venues which center around grocery stores, restaurants, banks, and beaches. Because of COVID-19, I believe the data is skewed. Although the state has opened, people are not yet shopping, and taking part in activities as they were before the onset of the pandemic. Also, the tourist industry has suffered greatly since people are not traveling. Once life gets back to normal, the clustering data will be more helpful.

## References

## Coursera, IBM Professional Data Science Certificate, Course 9, Applied Data Science Capstone, <https://www.coursera.org/learn/applied-data-science-capstone/home/welcome>

## FOURSQUARE API <https://api.foursquare.com/v2/venues/explore>

Homefacts website <https://www.homefacts.com/demographics/Florida/Palm-Beach-County.html>

The data for each municipality was retrieved by following the links for each municipality where *Palm-Beach-County* was replaced with municipality name. For example, to obtain data for the municipality *Atlantis*, the URL becomes <https://www.homefacts.com/demographics/Florida/Palm-Beach-County/Atlantis.html>

Palm Beach County logo, <http://discover.pbcgov.org/Pages/default.aspx>

Palm Beach County photo, Wikipedia, <https://en.wikipedia.org/wiki/West_Palm_Beach,_Florida>