Tampa Bay Neighborhood

Introduction

Tampa bay is known for various things like vacation spots, beautiful beaches, water sports, theme parks, lively entertainment areas like Sundial, AMCs etc. For people who are searching and analyzing for short time trips, relocate to Tampa bay from other locations this project can be a starting point to identify the county/city which will be suitable for their own needs like Restaurants, Movie theaters, clubs, Gym, Parks etc.

Objective

This project will provide a brief idea about few counties and located in Tampa bay like Pasco, Hillsborough, and Pinellas. This looks at neighborhoods and locations nearby and helps anyone to choose the best place to live within any given city based on the preferences satisfied.

Using segmentation and clustering based on the common places retrieved from Foursquare Determine

- Similarity or dissimilarity of areas between counties
- Classification of areas located around the Neighborhood.

Description about the Data:

The data set includes the coordinates of the cities/neighborhoods in the USA in JSON format with the county, latitude, longitude. With the available data the project is designed to use location coordinates and tag each data point into a neighborhood in few Counties in Florida - Tampa bay - Pasco, Hillsborough and Pinellas County. The algorithm used is k-means clustering. The main idea is to determine neighborhood with venues clustered around each other so that one can make a decision on the right neighborhood to chose based on the person's requirements.

Data set: https://raw.githubusercontent.com/natarajarumuga/Capstone/master/us_cities.json

Methodology

1. Analysis

Load the data from json file, read into pandas dataframe to perform the analysis on the data by counties using Foursquare. Analysis was performed using the following

a. Segment the Data into Location from Tampa bay Counties – Pinellas, Hillsborough and Pasco.
b. Filter missing values and columns

c. Using K-means, Cluster the data into multiple levels for comparison.

Segmenting and Slicing:

Based on the dataset values, segment the data into Neighborhood, County, State, City Name, Latitude, Longitude, and Density. Using Foursquare API, collect the venues for each county and city Name. There are some challenges with Foursquare with missing data for some of these locations and K-means process needs to be carefully analyzed to filter those data which will not be providing any value for comparison and proceeding with Analysis. Slicing of data was performed to identify only those columns required for our analysis and filter other columns like Time Zone, State name, Country etc which are not considered or relevant for this analysis.

Neighborhood Analysis

Using Folium package, data for the counties – Pinellas, Pasco and Hillsborough were plotted with the locations collected from Foursquare API and coordinates from the data. For this analysis we have considered top 100 locations within a limit of 500m radius. One hot encoding was performed on the locations to get a comparable value for analysis like 0 and 1. Results were grouped by neighborhood locations by mean of the frequency of the occurrence for each category.

Clustering

K-Means clustering was used to cluster the locations in the neighborhood for each County. We have used a predefined number of clusters for each county as part of the dataset. On the basis of K –means model, the following is performed

1. Define Cluster center as arithmetic mean of the all the points to the Cluster
2. Each point is closer to its own cluster.

To visualize the data better, Data was sliced for each county by neighborhood and locations around the neighborhood within the county.

RESULT

Using the source data for Tampa Bay Counties, we have analyzed the locations in the counties Pasco, Pinellas and Hillsborough as shown below in table

Pasco County Data Sliced

index	Neighborhood	County	Density	Latitude	Longitu
6138	Blanton	Pasco	94.2	28.4117	-82.2
6139	Lumberton	Pasco	102.2	28.2661	-82.1
6141	Drexel	Pasco	416.6	28.2428	-82.4
13260	Crystal Springs	Pasco	62.3	28.1830	-82.1
13261	Odessa	Pasco	75.6	28.1820	-82.5
13262	Aripeka	Pasco	200.8	28.4308	-82.6
13263	Trilby	Pasco	239.1	28.4565	-82.1
13264	Shady Hills	Pasco	247.3	28.4043	-82.5
13265	Wesley Chapel	Pasco	345.2	28.2106	-82.3
13266	Hudson	Pasco	631.1	28.3594	-82.6
13267	Land O' Lakes	Pasco	793.1	28.2075	-82.4
13268	Holiday	Pasco	955.1	28.1864	-82.7
13269	Lacoochee	Pasco	1055.5	28.4655	-82.1
13270	Trinity	Pasco	1197.7	28.1809	-82.6
13271	Elfers	Pasco	1338.2	28.2140	-82.7
13272	Beacon Square	Pasco	1847.4	28.2118	-82.7
13273	Bayonet Point	Pasco	2099.4	28.3255	-82.6
13274	Jasmine Estates	Pasco	3047.9	28.2930	-82.6
31076	San Antonio	Pasco	397.0	28.3396	-82.2

Pinellas County Sliced Data

index	Neighborhood	County	Density	Latitude	Longitu
6142	Gandy	Pinellas	431.7	27.8685	-82.6
6143	Highpoint	Pinellas	649.1	27.9170	-82.7
6145	Crystal Beach	Pinellas	1285.1	28.0914	-82.7
6146	Baskin	Pinellas	1558.6	27.8970	-82.7
13275	Feather Sound	Pinellas	2.8	27.9062	-82.6
13276	Bay Pines	Pinellas	442.1	27.8144	-82.7
13277	South Highpoint	Pinellas	649.1	27.9085	-82.7
13278	East Lake	Pinellas	1037.2	28.1205	-82.6
13279	Tierra Verde	Pinellas	1269.6	27.6685	-82.7
13280	Ridgecrest	Pinellas	1327.1	27.8952	-82.8
13281	Harbor Bluffs	Pinellas	1334.0	27.9083	-82.8
13282	Palm Harbor	Pinellas	1345.0	28.0847	-82.7
13283	Lealman	Pinellas	1992.2	27.8197	-82.6
31082	Belleair Shores	Pinellas	733.0	27.9173	-82.8
31083	North Redington Beach	Pinellas	2153.0	27.8214	-82.8
31084	Redington Beach	Pinellas	1598.0	27.8127	-82.8
31085	Indian Shores	Pinellas	1710.0	27.8535	-82.8
31086	Belleair Beach	Pinellas	1314.0	27.9240	-82.8
31087	Belleair Bluffs	Pinellas	1817.0	27.9200	-82.8
31088	Redington Shores	Pinellas	2647.0	27.8294	-82.8
31089	Belleair	Pinellas	902.0	27.9363	-82.8
31090	Indian Rocks Beach	Pinellas	1979.0	27.8963	-82.8
31091	Madeira Beach	Pinellas	1731.0	27.7986	-82.7

Hillsborough County Sliced Data

index	Neighborhood	County	Density	Latitude	Longitu
6058	Knights	Hillsborough	87.1	28.0764	-82.1
6059	Keysville	Hillsborough	152.8	27.8692	-82.0
6060	Sun City	Hillsborough	229.7	27.6784	-82.4
6061	Orient Park	Hillsborough	1315.2	27.9725	-82.3
13062	Keystone	Hillsborough	27.6	28.1312	-82.5
13063	Balm	Hillsborough	94.6	27.7542	-82.2
13064	Fish Hawk	Hillsborough	413.3	27.8511	-82.2
13065	Westchase	Hillsborough	437.7	28.0597	-82.6
13066	Lutz	Hillsborough	456.4	28.1396	-82.4
13067	Wimauma	Hillsborough	460.6	27.6965	-82.3
13068	Northdale	Hillsborough	467.2	28.1058	-82.5
13069	Ruskin	Hillsborough	539.7	27.7065	-82.4
13070	Dover	Hillsborough	544.1	27.9927	-82.2
13071	Brandon	Hillsborough	576.2	27.9367	-82.3
13072	Cheval	Hillsborough	645.6	28.1459	-82.5
13073	Riverview	Hillsborough	735.3	27.8227	-82.3
13074	Sun City Center	Hillsborough	772.6	27.7150	-82.3
13075	Citrus Park	Hillsborough	826.8	28.0730	-82.5
13076	Thonotosassa	Hillsborough	874.0	28.0464	-82.2
13077	Lake Magdalene	Hillsborough	890.0	28.0874	-82.4
13078	Bloomingdale	Hillsborough	1017.0	27.8784	-82.2
13079	Mango	Hillsborough	1024.2	27.9915	-82.3
13080	Valrico	Hillsborough	1335.0	27.9193	-82.2

After retrieving the neighborhood data for each county using Foursquare API, there were various different neighborhood venues like Gym, Park, restaurants by cuisine type etc are collected. I used a radius if 500 there were 89 unique categories for Pasco, 73 unique categories for Hillsborough and 108 unique categories for Pinellas County. Using one hot

encoding there were 164 rows and 90 columns for Pasco, 112 rows and 74 columns for Hillsborough, 241 rows and 108 columns for Pinellas.
Below are the top 3 most common venues by neighborhood location for each county

Pasco County

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd M
Baskin	Intersection	Breakfast Spot	Food
Bay Pines	Pet Store	Smoke Shop	
Belleair	Park	Clothing Store	
Belleair Bluffs	Steakhouse	Breakfast Spot	Ic
Belleair Shores	Beach	Pharmacy	
Crystal Beach	Pier	Beach	
Dunedin	Beach	Park	H
Gandy	Gas Station	Furniture / Home Store	
Gulfport	Park	Flower Shop	
Harbor Bluffs	Park	Beach	
Highpoint	Bar	Sports Bar	Mexic
Indian Rocks Beach	Pizza Place	Seafood Restaurant	Ba
Indian Shores	Park	Resort	

Hillsborough County

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd M
Apollo Beach	Athletics & Sports	Gymnastics Gym	
Bloomingdale	Golf Course	Wings Joint	
Brandon	American Restaurant	Department Store	
Cheval	Home Service	Tennis Court	
Citrus Park	Lake	Dog Run	Rec
Dover	Taco Place	BBQ Joint	
Fish Hawk	Gym	Pool	
Gibsonton	Gym	Wings Joint	
Knights	Construction & Landscaping	Farm	
Lake Magdalene	Park	Gym / Fitness Center	
Mango	American Restaurant	Home Service	
Northdale	American Restaurant	Golf Course	
Orient Park	Bar	Convenience Store	
Pebble Creek	Gym	Playground	
Plant City	Bar	Deli / Bodega	
Progress Village	Construction & Landscaping	Dance Studio	Gr
Riverview	Farm	Travel Agency	
Ruskin	Martial Arts Dojo	Wings Joint	Con
Sun City	Construction & Landscaping	Greek Restaurant	C

Pinellas County

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd M
Baskin	Intersection	Breakfast Spot	Food
Bay Pines	Pet Store	Smoke Shop	
Belleair	Park	Clothing Store	
Belleair Bluffs	Steakhouse	Breakfast Spot	Ic
Belleair Shores	Beach	Pharmacy	
Crystal Beach	Pier	Beach	
Dunedin	Beach	Park	H
Gandy	Gas Station	Furniture / Home Store	
Gulfport	Park	Flower Shop	
Harbor Bluffs	Park	Beach	
Highpoint	Bar	Sports Bar	Mexi
Indian Rocks Beach	Pizza Place	Seafood Restaurant	Ba
Indian Shores	Park	Resort	
Kenneth City	Gay Bar	Park	C
Largo	Golf Course	Chinese Restaurant	

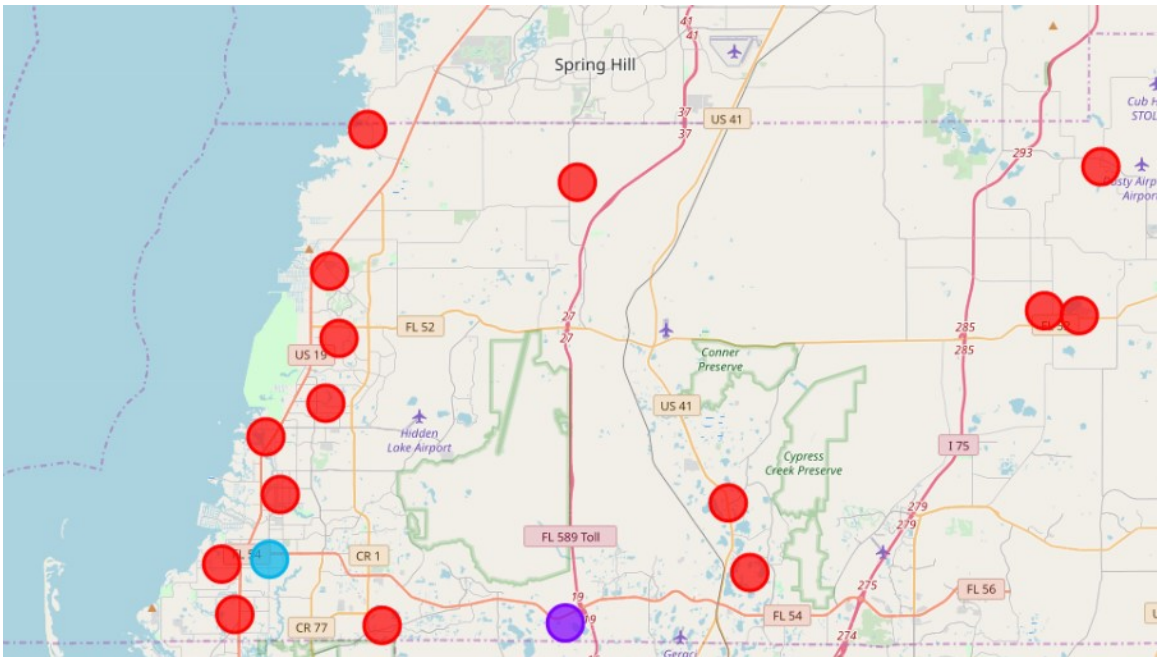
The main objective of the project is to cluster the neighborhoods; k-means algorithm is applied to the one-hot encoded venue dataset. Below is the table that shows the neighborhood, cluster labels assigned to it after K-means algorithm was applied to the data. Cluster are labeled in the below order

1. 1st Cluster = Cluster label 0
2. 2nd Cluster = Cluster label 1
3. 3rd Cluster = Cluster label 2
4. 4th Cluster = Cluster label 3

Pasco County Clusters

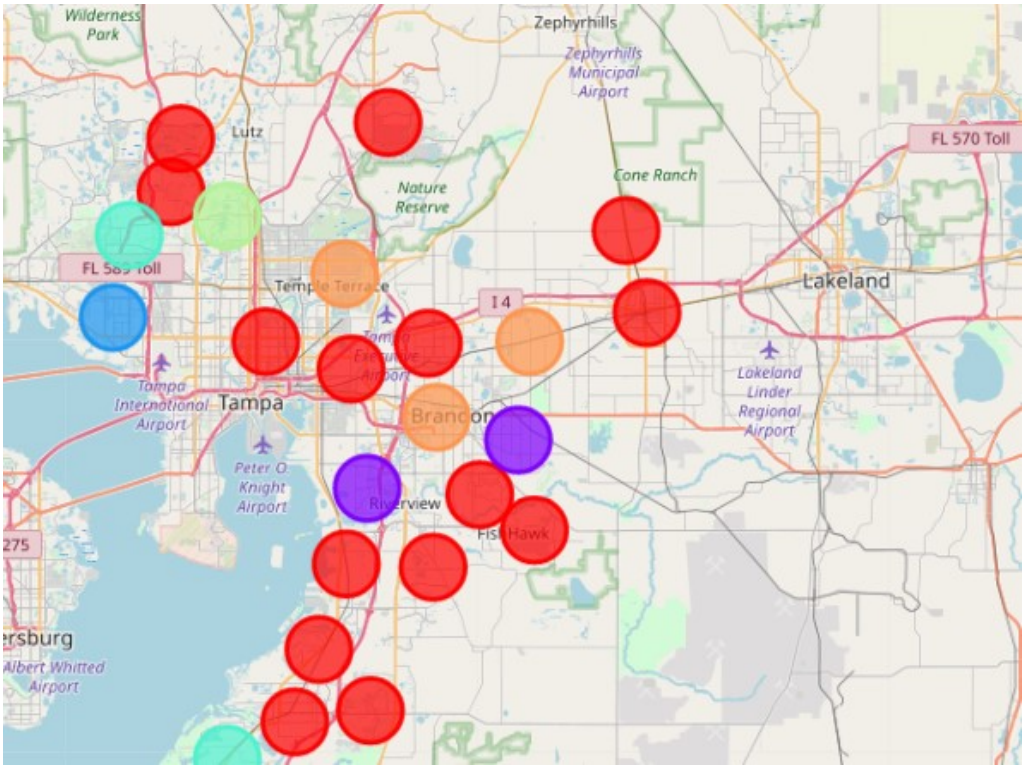
index	Neighborhood	County	Density	Latitude	Longitude	State	Cluster Labels
6138	Blanton	Pasco	94.2	28.4117	-82.2465	FL	0
6139	Lumberton	Pasco	102.2	28.2661	-82.1365	FL	0
6141	Drexel	Pasco	416.6	28.2428	-82.4598	FL	0
13260	Crystal Springs	Pasco	62.3	28.183	-82.1539	FL	3
13261	Odessa	Pasco	75.6	28.182	-82.553	FL	1
13262	Aripeka	Pasco	200.8	28.4308	-82.6669	FL	0
13263	Trilby	Pasco	239.1	28.4565	-82.194	FL	0
13264	Shady Hills	Pasco	247.3	28.4043	-82.5468	FL	0
13266	Hudson	Pasco	631.1	28.3594	-82.6888	FL	0
13267	Land O' Lakes	Pasco	793.1	28.2075	-82.4476	FL	0
13268	Holiday	Pasco	955.1	28.1864	-82.7429	FL	0
13269	Lacoochee	Pasco	1055.5	28.4655	-82.1698	FL	4
13270	Trinity	Pasco	1197.7	28.1809	-82.6584	FL	0
13271	Elfers	Pasco	1338.2	28.214	-82.723	FL	2
13272	Beacon Square	Pasco	1847.4	28.2118	-82.7504	FL	0
13273	Bayonet Point	Pasco	2099.4	28.3255	-82.6834	FL	0
13274	Jasmine Estates	Pasco	3047.9	28.293	-82.6907	FL	0
31076	San Antonio	Pasco	397	28.3396	-82.279	FL	0
31077	Saint Leo	Pasco	403	28.3372	-82.2586	FL	0
31078	Port Richey	Pasco	489	28.2758	-82.7251	FL	0
31079	Dade City	Pasco	461	28.3564	-82.1942	FL	0
31080	Zephyrhills	Pasco	614	28.2404	-82.1796	FL	0
31081	New Port Richey	Pasco	1377	28.2468	-82.7169	FL	0

Cluster 0 is marked in Red Color
Cluster 1 is marked in Purple Color
Cluster 2 is marked in Blue Color
Cluster 3 is marked in Green Color
Cluster 4 is marked in Orange Color



Hillsborough County – Clusters:

index	County	Density	Latitude	Longitude	State	Cluster Labels
6058	Hillsborough	87.1	28.0764	-82.1376	FL	0
6060	Hillsborough	229.7	27.6784	-82.4787	FL	3
6061	Hillsborough	1315.2	27.9725	-82.3729	FL	0
13064	Hillsborough	413.3	27.8511	-82.2164	FL	0
13068	Hillsborough	467.2	28.1058	-82.5263	FL	0
13069	Hillsborough	539.7	27.7065	-82.421	FL	0
13070	Hillsborough	544.1	27.9927	-82.2198	FL	5
13071	Hillsborough	576.2	27.9367	-82.3	FL	5
13072	Hillsborough	645.6	28.1459	-82.5184	FL	0
13073	Hillsborough	735.3	27.8227	-82.3023	FL	0
13074	Hillsborough	772.6	27.715	-82.3569	FL	0
13075	Hillsborough	826.8	28.073	-82.5628	FL	3
13077	Hillsborough	890	28.0874	-82.4791	FL	4
13078	Hillsborough	1017	27.8784	-82.2624	FL	0
13079	Hillsborough	1024.2	27.9915	-82.307	FL	0
13080	Hillsborough	1335	27.9193	-82.2293	FL	1
13081	Hillsborough	1374.3	27.8257	-82.3766	FL	0
13083	Hillsborough	1515.7	28.0108	-82.576	FL	2
13084	Hillsborough	1677.5	27.8832	-82.3593	FL	1
13086	Hillsborough	1723.1	28.1582	-82.3411	FL	0
13087	Hillsborough	1816	27.7619	-82.4003	FL	0
30927	Hillsborough	1450	28.044	-82.378	FL	5
30928	Hillsborough	542	28.0145	-82.1202	FL	0
30929	Hillsborough	1283	27.9937	-82.4454	FL	0

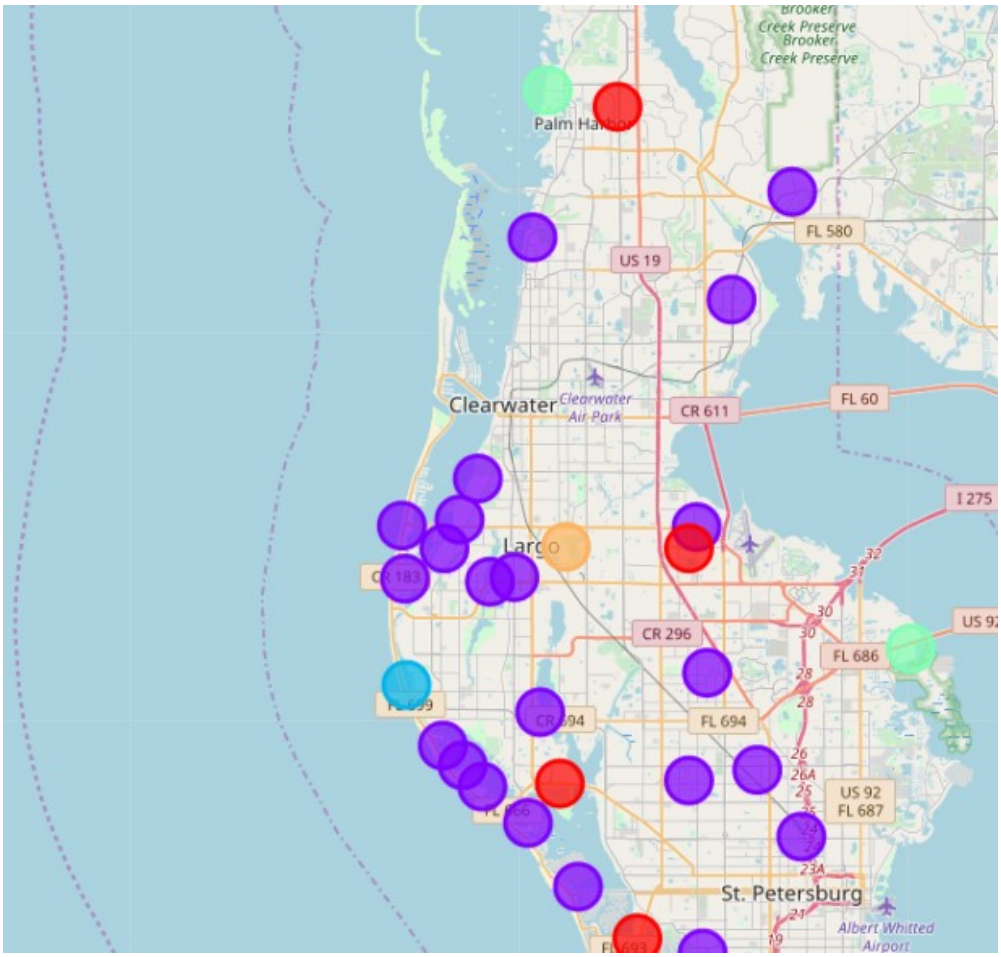


Cluster 0 is marked in Red Color
Cluster 1 is marked in Purple Color
Cluster 2 is marked in Blue Color
Cluster 3 is marked in Green Color
Cluster 4 is marked in Pale Green Color
Cluster 5 is marked in Orange Color

Pinellas County Clusters

index	County	Density	Latitude	Longitude	State	Cluster Labels
6142	Pinellas	431.7	27.8685	-82.6161	FL	3
6143	Pinellas	649.1	27.917	-82.7129	FL	1
6145	Pinellas	1285.1	28.0914	-82.7798	FL	3
6146	Pinellas	1558.6	27.897	-82.7951	FL	1
13276	Pinellas	442.1	27.8144	-82.7744	FL	0
13277	Pinellas	649.1	27.9085	-82.7162	FL	0
13279	Pinellas	1269.6	27.6685	-82.73	FL	0
13280	Pinellas	1327.1	27.8952	-82.8063	FL	1
13281	Pinellas	1334	27.9083	-82.8269	FL	1
13282	Pinellas	1345	28.0847	-82.7481	FL	0
13283	Pinellas	1992.2	27.8197	-82.6849	FL	1
31082	Pinellas	733	27.9173	-82.8455	FL	1
31083	Pinellas	2153	27.8214	-82.8186	FL	1
31084	Pinellas	1598	27.8127	-82.8094	FL	1
31085	Pinellas	1710	27.8535	-82.8439	FL	2

31087	Pinellas	1817	27.92	-82.8195	FL	1
31088	Pinellas	2647	27.8294	-82.8274	FL	1
31089	Pinellas	902	27.9363	-82.8118	FL	1
31090	Pinellas	1979	27.8963	-82.8443	FL	1
31091	Pinellas	1731	27.7986	-82.7889	FL	1
31092	Pinellas	2699	27.8156	-82.7162	FL	1
31093	Pinellas	3248	27.7524	-82.7395	FL	0
31094	Pinellas	1733	27.7731	-82.766	FL	1
31095	Pinellas	1753	27.7215	-82.7383	FL	3
31096	Pinellas	1736	27.7463	-82.7099	FL	1
31097	Pinellas	621	28.0506	-82.6696	FL	1
31098	Pinellas	1385	28.008	-82.6965	FL	1
31099	Pinellas	1378	27.8429	-82.7835	FL	1
31100	Pinellas	1054	28.1492	-82.7622	FL	1
31101	Pinellas	1351	28.0328	-82.7865	FL	1
31102	Pinellas	1281	27.8586	-82.7075	FL	1
31103	Pinellas	1760	27.9088	-82.7714	FL	4
31105	Pinellas	1631	27.7931	-82.6652	FL	1



Cluster 0 is marked in Red Color

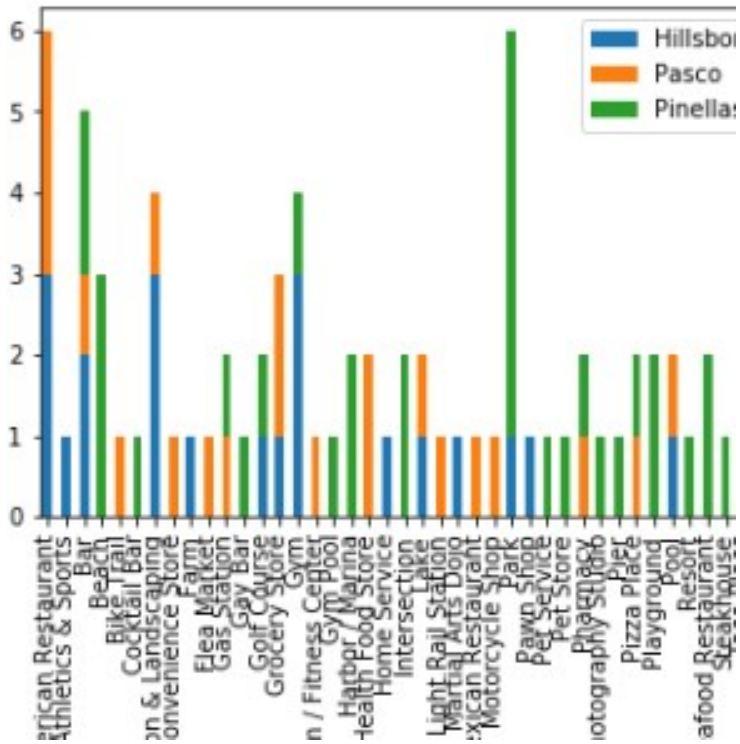
Cluster 1 is marked in Purple Color
Cluster 2 is marked in Blue Color
Cluster 3 is marked in Green Color
Cluster 4 is marked in Pale Green Color
Cluster 5 is marked in Orange Color

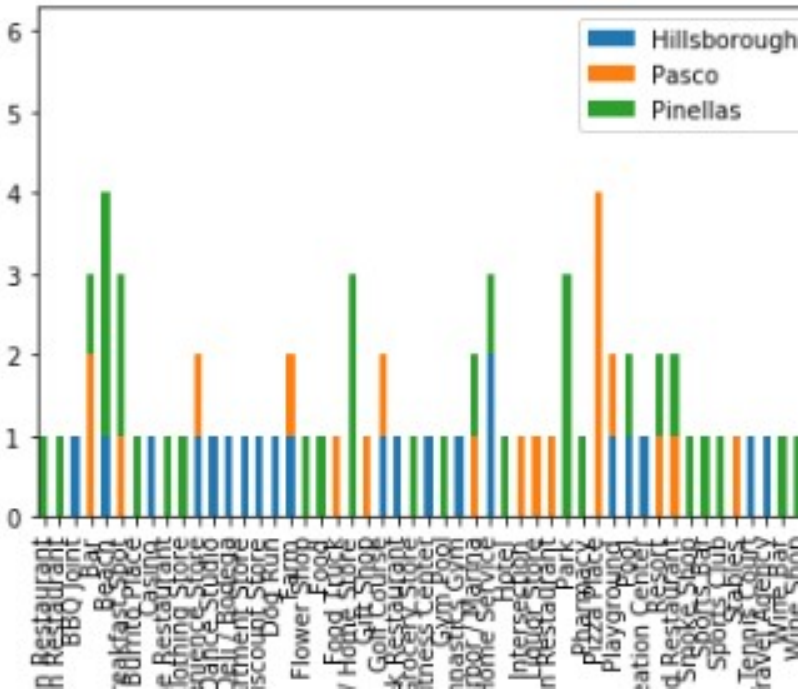
Most Common Venue location by County

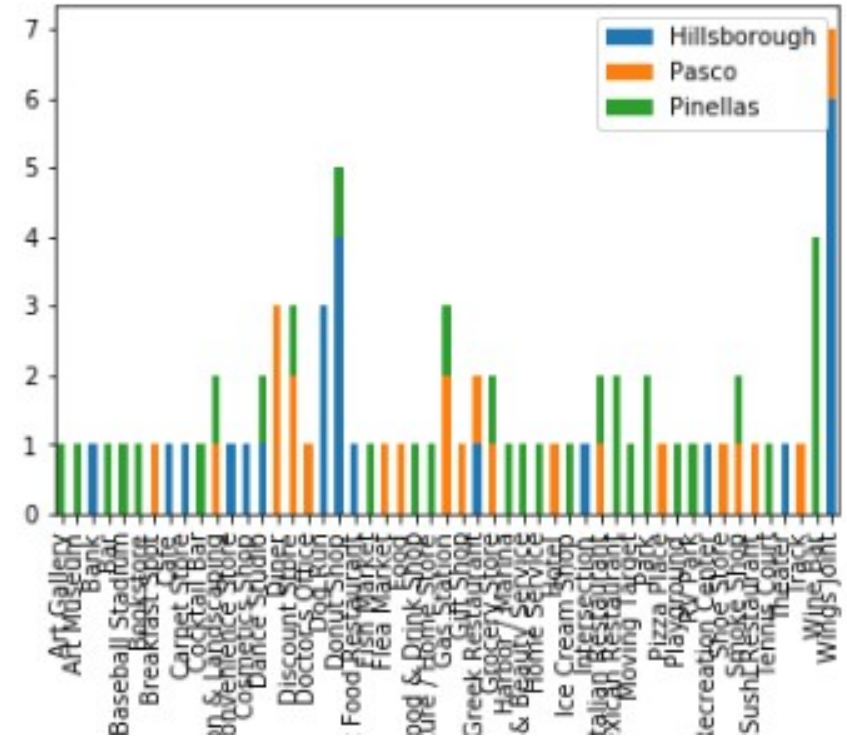
After examining most common venues in each county, it can be determined that the below the are the top 5 common venues in each county

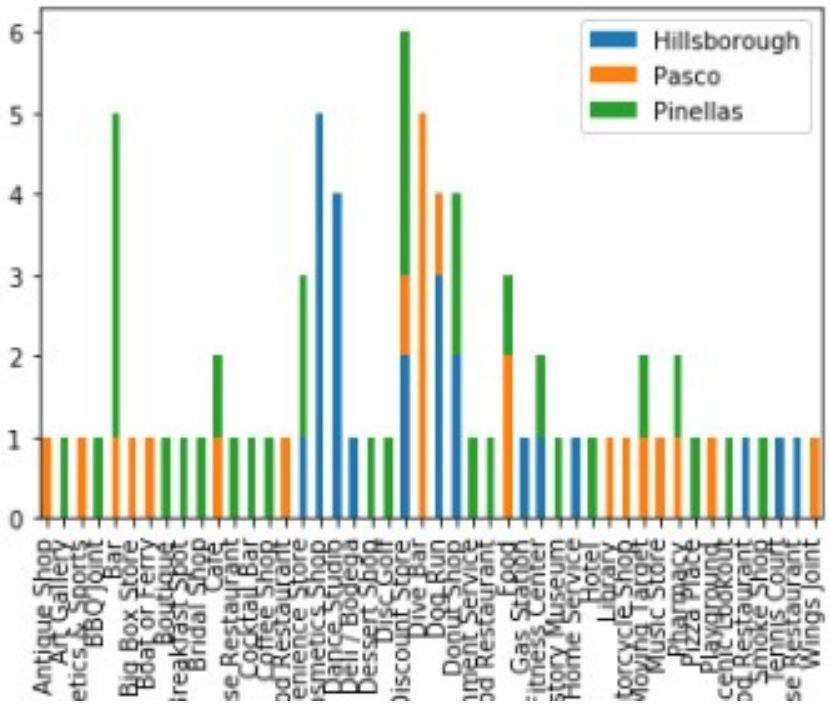
	Pasco	Hillsborough	Pinellas
1	Gym	Restaurants	Gas Station
2	American Restaurant	Gym	Pier
3	Gas Station	Park	Bar
4	Farm	Department Store	Beach
5	Farmers market	Dance Studio	Resort
6	Doctor office	Cosmetics Shop	Hotel

Below plot shows how the Most common venue for each county:




Hillsborough
Pasco
Pinellas
0
1
2
3
4
5
6


Hillsborough
Pasco
Pinellas
0
1
2
3
4
5
6
7



DISCUSSION

Based on the data and the results of the cluster, the neighborhood venues in Red cluster have more venues in Pasco and Hillsborough Counties and Purple cluster in Pinellas. Analyzing the venues, Pasco and Hillsborough County has more similar category of venues as most common ones like Gas Station, Doctor Office, Bar, Gym, Restaurants etc Whereas Pinellas County has Beach, Hotels, Resort, Pier as the most common Venus. People, who are searching to move-in or plan trips in Tampa bay, can select any of these neighborhoods based on the preference. For eg: People who likes to live near ocean, visit beaches frequently can choose Pinellas county where as people who likes amenities like Gym, Gas station, Restaurants can choose Pasco/Hillsborough county.

Similarly for people visiting Tampa bay can decide the place in Pinellas County as this county has the Hotel, Resort and Beach as most common Venues.

CONCLUSION

This project is designed to provide an insight into locations near Tampa bay for help people look and decide the best place to choose to rent, live, visit. Using Information available for public datasets for free I chose three different counties in Tampa bay to analyze the neighborhoods based on the spatial distribution of Venues. This analysis was performed using Python libraries like Folium, matplotlib for visualizing the data, Foursquare API to get the venue information. K-means Algorithm was applied to cluster the data for these neighborhoods. There were some challenges with missing data for few of the neighborhoods which could have provided additional information and affected the results. Also this project has scope to improve more to compare different counties/cities/States in United States with this data.