

Capstone Project: Food Venues Data Analysis of Knoxville using Python

A. Introduction

A.1 Business Problem

Knoxville is the **3rd largest city** in Tennessee, second only to Nashville and Memphis. It is also home to the University of Tennessee, which also drives many businesses to move into the area. Food Venue Data Analysis of the city would be of particular interest to someone who is looking to open a food restaurant in the area. If someone would like to open an Asian restaurant, or a Pizza place etc utilizing the data from Foursquare and Unsupervised learning clusters would help them discover which neighborhoods would be better options for this placement. Perhaps there are neighborhoods which already have too many pizza places, and some that have so few that they are very rare. This information will allow the business user to see where it is better to spend their resources.

A.2 Data

The first data set that we would need is the list of Neighborhoods in Knoxville, TN. Next, we will need to get the latitude and longitude locations for each of the neighborhoods. We can get location data from `geolocator.geocode()` in **python**. We will create a visual map, just to make sure that the data is valid. Next, we are going to start using **Foursquare API** to look for the top 100 venues in a 750 meter radius of each neighborhood. These calls will be specific for the category of Food which has the categoryId of **4d4b7105d754a06374d81259**. This Id captures all categories of restaurants and it would be best for our use.

Once we have the data for each neighborhood, and their top food venues by category we are going to put this data through **k-means unsupervised learning** in order to create clusters. These clusters will be visualized using **Folium**. Post-visualization of these clusters, we will study what are the main differences between the clusters, and if we can discover a pattern which allows us to help answer our initial business problem.

B. Methodology

B.1 Gathering and Preparing Data:

I get the names of all of the Neighborhoods in Knoxville from the following: https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Knoxville,_Tennessee. Importing this into Pandas data frame to populate Latitude and Longitudinal co-ordinates for each Neighborhood.

	Neighborhood	City	Latitude	Longitude
0	Bearden	Knoxville	NaN	NaN
1	Cedar Bluff	Knoxville	NaN	NaN
2	Chilhowee Park	Knoxville	NaN	NaN
3	Colonial Village	Knoxville	NaN	NaN
4	Downtown Knoxville	Knoxville	NaN	NaN

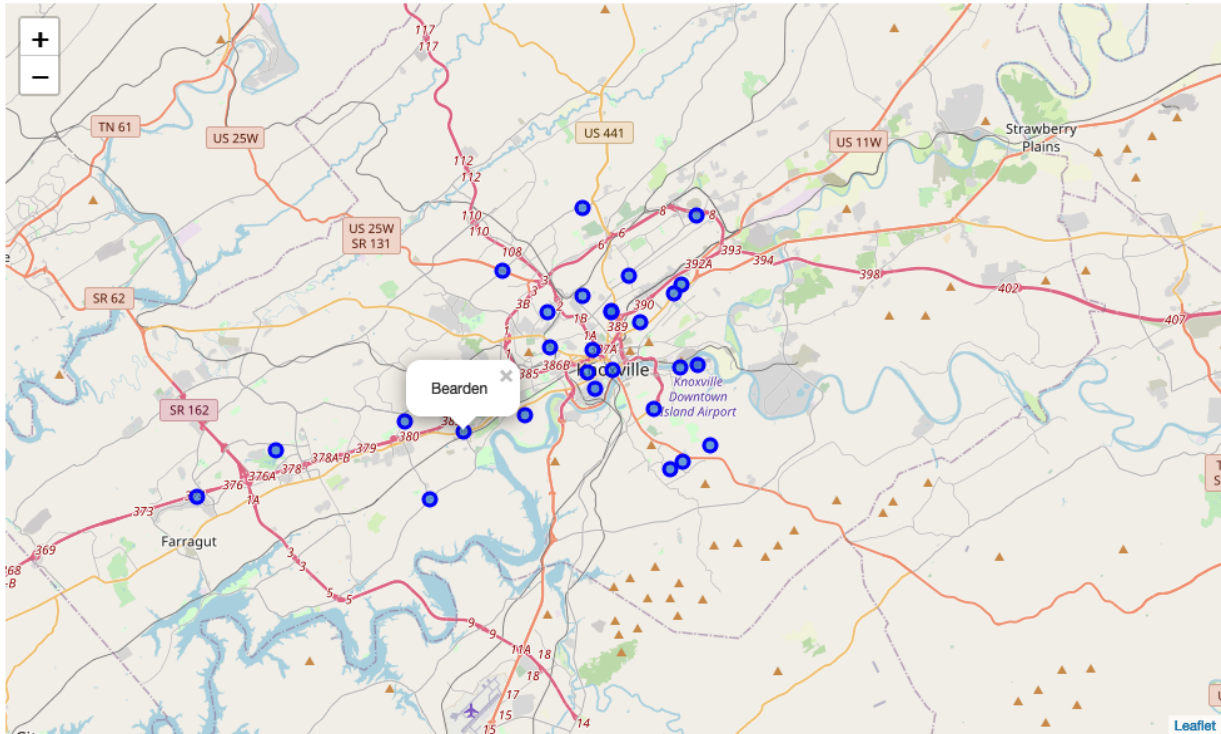
B.2 Add Latitude/Longitude Co-ordinates for each Neighborhood.

Using Forloop and geolocator to fill the dataframe.

	Neighborhood	City	Latitude	Longitude
0	Bearden	Knoxville	35.932859	-84.002686
1	Cedar Bluff	Knoxville	35.924885	-84.105837
2	Chilhowee Park	Knoxville	35.998107	-83.883183
3	Colonial Village	Knoxville	35.919528	-83.882681
4	Downtown Knoxville	Knoxville	35.962600	-83.874223

B.3 Plotting Neighborhoods in Knoxville, TN using Folium

We want to plot the dataframe in order to make sure that our information is correct. Secondly, we can also visualize the Neighborhoods and see how spread they are.



B.4 Gathering Data from Foursquare for each Neighborhood in Knoxville, TN

We use our credentials to create an access to FourSqaure, and then from there we will be using our collected Latitude and Longitude co-ordinates for each Neighborhood to list 100 venues in a 750 meter radius.

```
nearby_venues.head()
```

	name	categories	lat	lng
0	Sitar Indian Cuisine	Indian Restaurant	35.932890	-84.001477
1	Bistro by the Tracks	American Restaurant	35.934955	-84.003747
2	Surin of Thailand	Thai Restaurant	35.932803	-84.005040
3	Aubrey's	American Restaurant	35.937712	-84.005495
4	Wasabi Japanese Steakhouse & Sushi Bar	Japanese Restaurant	35.933088	-84.008790

B.5 Iterating through the entire list of Neighborhood names, and using the categoryId for Food

Using this categoryId will allow us to get all of the different types of food in Knoxville, TN and that which are contained in the FourSquare database.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bearden	35.932859	-84.002686	Sitar Indian Cuisine	35.932890	-84.001477	Indian Restaurant
1	Bearden	35.932859	-84.002686	Bistro by the Tracks	35.934955	-84.003747	American Restaurant
2	Bearden	35.932859	-84.002686	Surin of Thailand	35.932803	-84.005040	Thai Restaurant
3	Bearden	35.932859	-84.002686	Aubrey's	35.937712	-84.005495	American Restaurant
4	Bearden	35.932859	-84.002686	Wasabi Japanese Steakhouse & Sushi Bar	35.933088	-84.008790	Japanese Restaurant

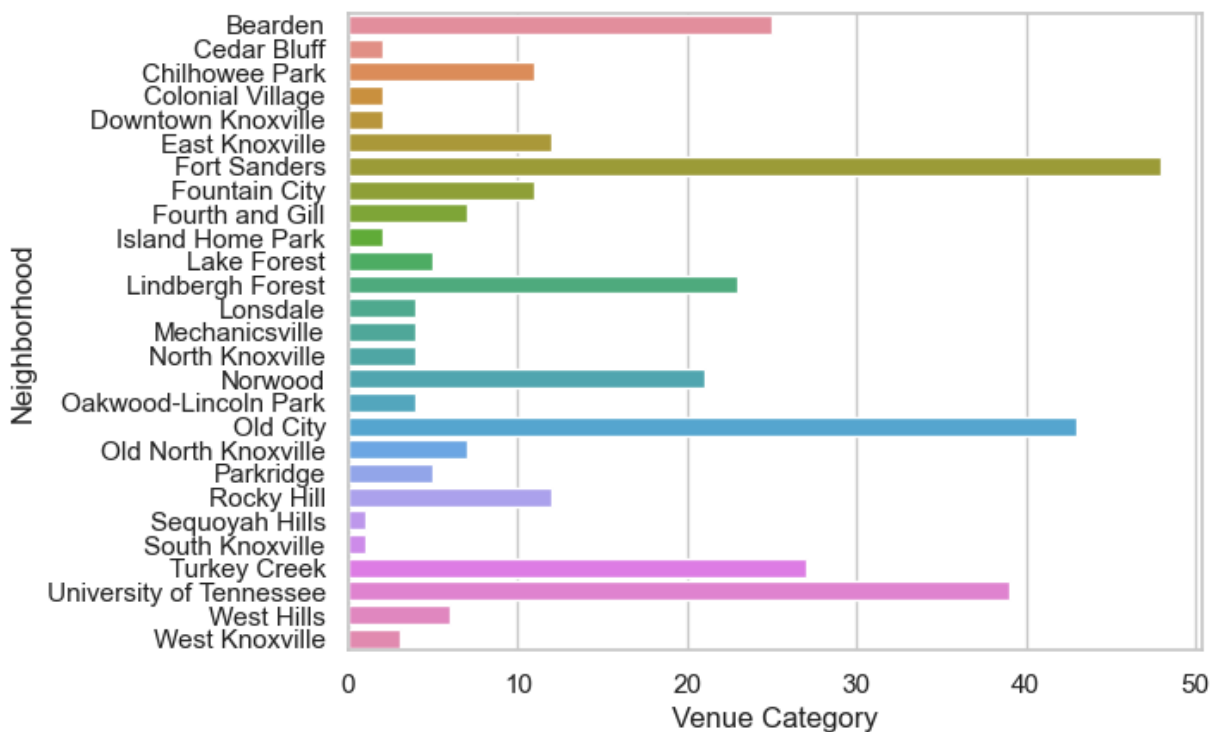
C. Results

C.1 Analysis of how many different venues are there in each Neighborhood

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Bearden	25	25	25	25	25	25
Cedar Bluff	2	2	2	2	2	2
Chilhowee Park	11	11	11	11	11	11
Colonial Village	2	2	2	2	2	2

C.2 Plot of how many different venues exist in each Neighborhood

Fort Sanders has the most number of venues in Knoxville, followed closely by the Old City and University of Tennessee.

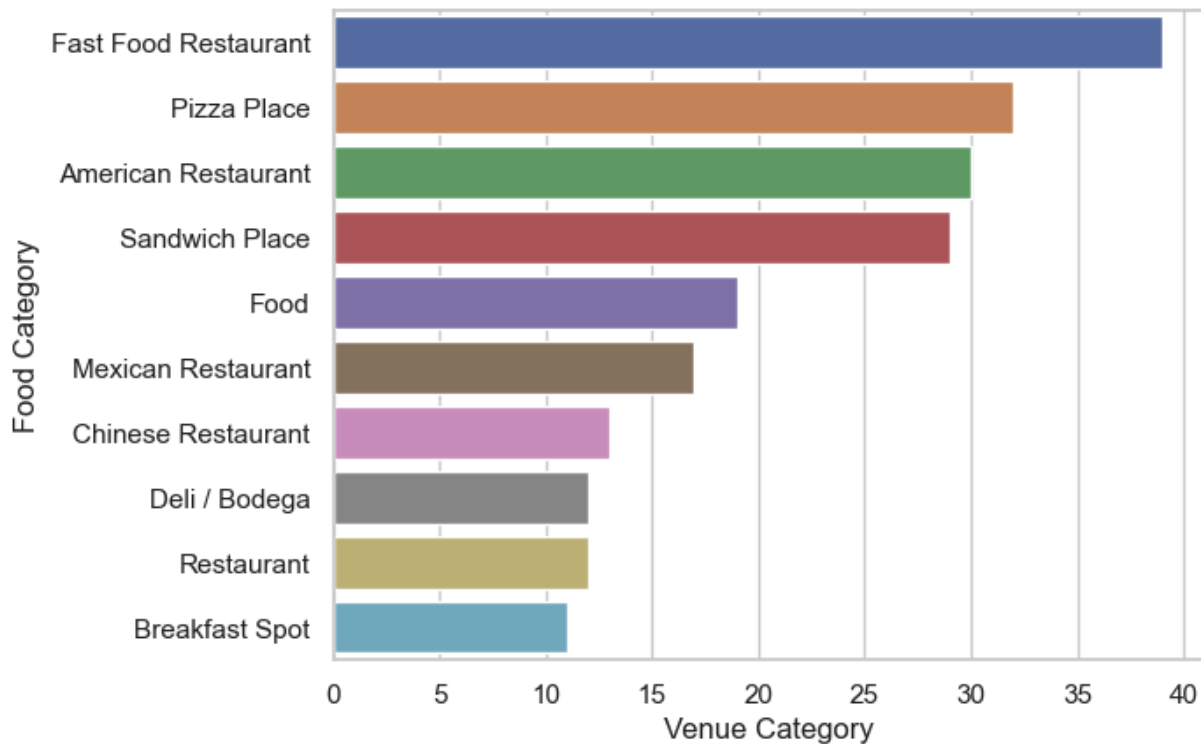


C.3 Analysis of what food categories exist in Knoxville

	Food Category	Venue Category
0	Fast Food Restaurant	39
1	Pizza Place	32
2	American Restaurant	30
3	Sandwich Place	29
4	Food	19
5	Mexican Restaurant	17
6	Chinese Restaurant	13
7	Deli / Bodega	12
8	Restaurant	12
9	Breakfast Spot	11

C.4 Plot: Food Categories in Knoxville

When look a little deeper to see what are the most popular types of food in Knoxville, and it appears Fast Food Restaurant is at the top. The next top category is Pizza.

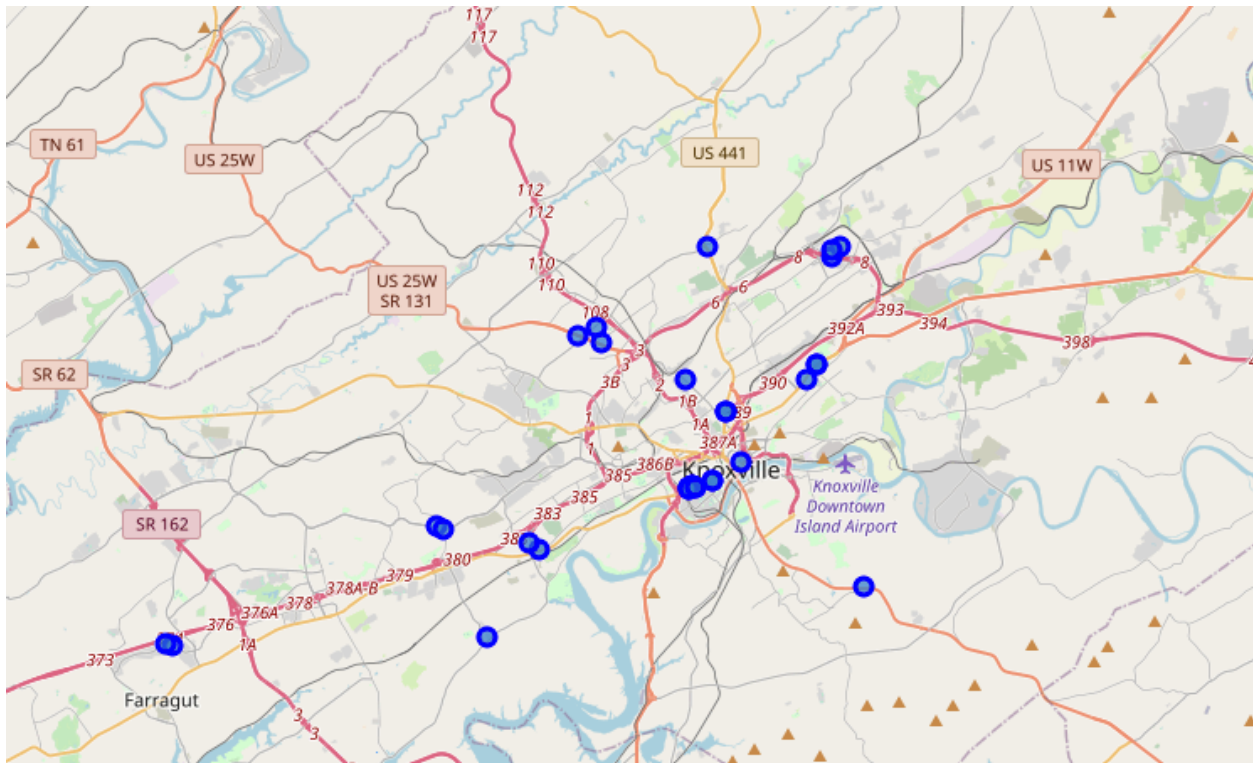


Getting this information of the popularity of Pizza in Knoxville, we can take a closer look at what are the most common Pizza places in the city.

C.5 Analysis of Pizza Places in Knoxville

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
19	Bearden	35.932859	-84.002686	Snappy Tomato Pizza	35.933990	-83.999573	Pizza Place
21	Bearden	35.932859	-84.002686	Rosati's	35.936238	-84.003548	Pizza Place
27	Chilhowee Park	35.998107	-83.883183	Pizza Palace	35.994964	-83.885873	Pizza Place
42	East Knoxville	35.994200	-83.887200	Pizza Palace	35.994964	-83.885873	Pizza Place
47	East Knoxville	35.994200	-83.887200	Little Caesars Pizza	35.990312	-83.890161	Pizza Place

C.6 Map of all of the Pizza places in Knoxville

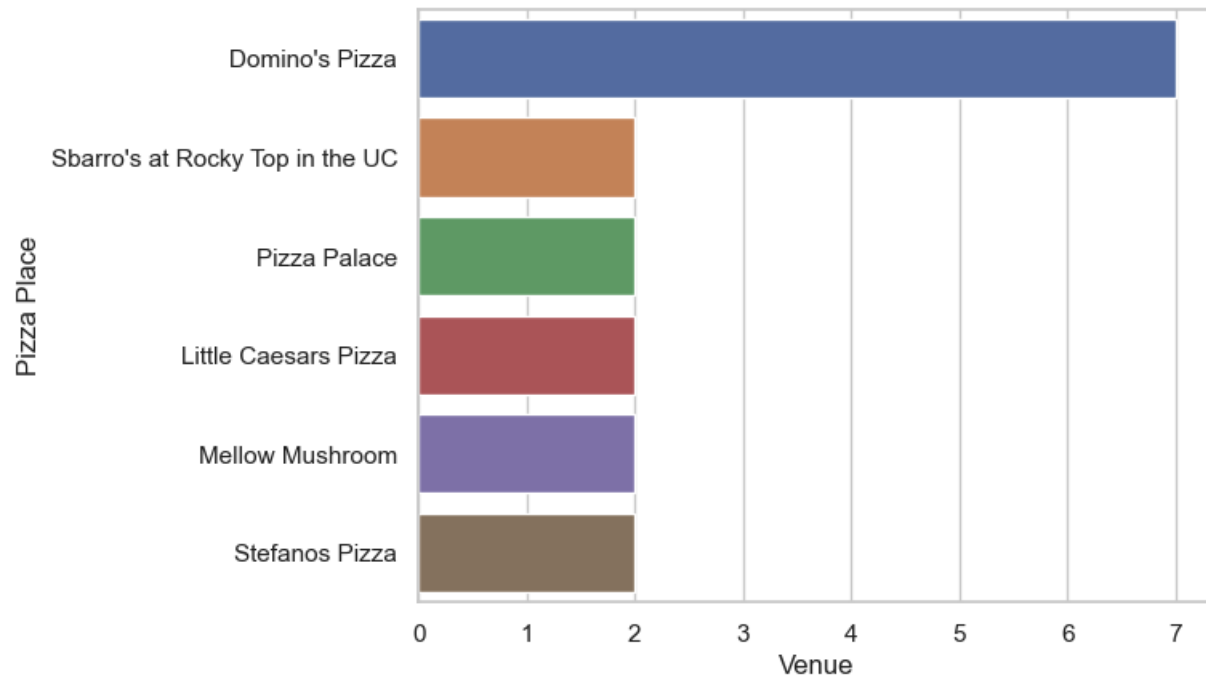


C.7 Most common Pizza places

	Pizza Place	Venue
0	Domino's Pizza	7
1	Sbarro's at Rocky Top in the UC	2
2	Pizza Palace	2
3	Little Caesars Pizza	2
4	Mellow Mushroom	2
5	Stefanos Pizza	2
6	Papa John's Pizza	2

C.8 Plot: Names of Most common Pizza places

Domino's franchise has the most locations in Knoxville, TN out of all of the different Pizza locations.

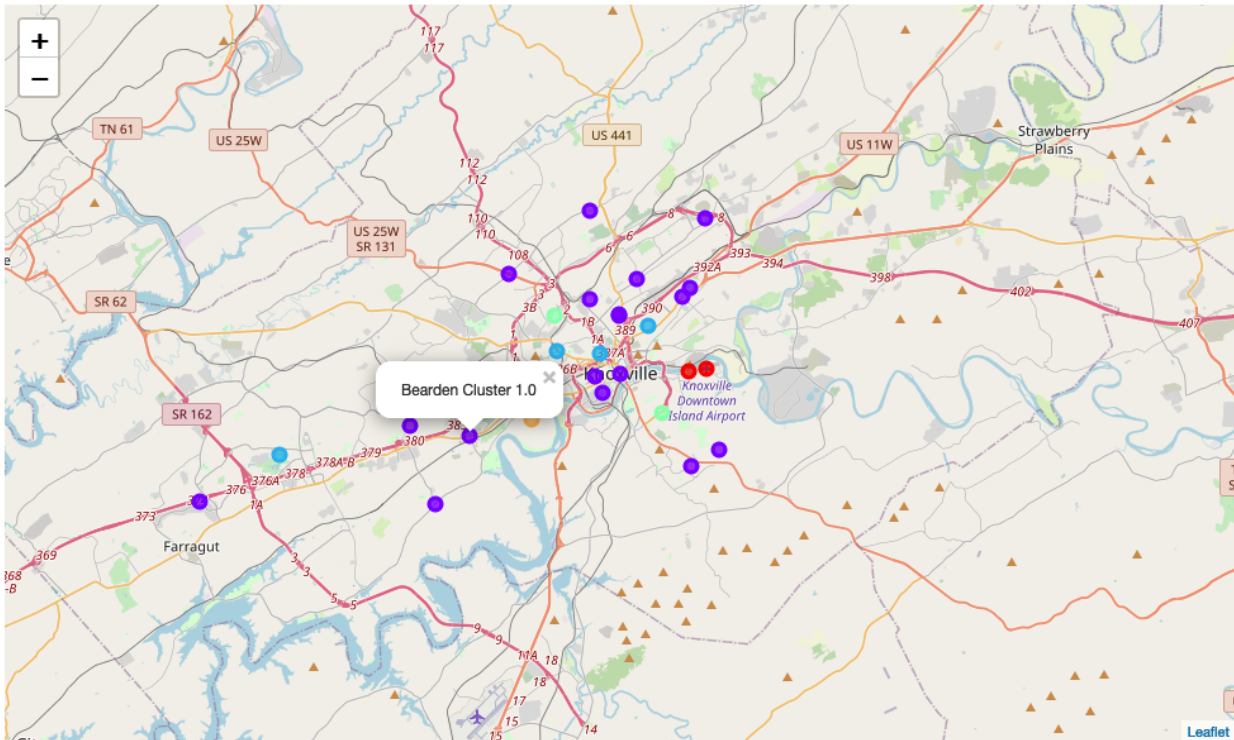


C.9 Analysis of top 5 unique categories by Neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bearden	American Restaurant	BBQ Joint	Thai Restaurant	Fast Food Restaurant	Food
1	Cedar Bluff	Food	Café	Wings Joint	Donut Shop	Gastropub
2	Chilhowee Park	American Restaurant	Southern / Soul Food Restaurant	Food	Pizza Place	Asian Restaurant
3	Colonial Village	Fast Food Restaurant	Café	Wings Joint	Donut Shop	Gastropub
4	Downtown Knoxville	Chinese Restaurant	Restaurant	Wings Joint	Donut Shop	Gastropub

C.10 Creating Neighborhood Clusters using K-means Unsupervised Learning

We are going to create **5 clusters**, and allow the algorithm to group neighborhoods together by their analytical similarity.



C.11 Neighborhood listing by unique cluster

```
tn_merged[tn_merged['Cluster Labels'] == 0.0]
```

	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Downtown Knoxville	Knoxville	35.962600	-83.874223	0.0	Chinese Restaurant	Restaurant	Wings Joint	Donut Shop	Gastropub
10	Island Home Park	Knoxville	35.961106	-83.884009	0.0	Chinese Restaurant	Restaurant	Wings Joint	Donut Shop	Gastropub

```
tn_merged[tn_merged['Cluster Labels'] == 1.0]
```

	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bearden	Knoxville	35.932859	-84.002686	1.0	American Restaurant	BBQ Joint	Thai Restaurant	Fast Food Restaurant	Food
2	Chilhowee Park	Knoxville	35.998107	-83.883183	1.0	American Restaurant	Southern / Soul Food Restaurant	Food	Pizza Place	Asian Restaurant
3	Colonial Village	Knoxville	35.919528	-83.882681	1.0	Fast Food Restaurant	Café	Wings Joint	Donut Shop	Gastropub
5	East Knoxville	Knoxville	35.994200	-83.887200	1.0	American Restaurant	Pizza Place	Southern / Soul Food Restaurant	Food Truck	Food

7	Fort Sanders	Knoxville	35.959170	-83.934895	1.0	Pizza Place	Sandwich Place	Deli / Bodega	Mexican Restaurant	American Restaurant
8	Fountain City	Knoxville	36.032026	-83.937408	1.0	Fast Food Restaurant	Sandwich Place	American Restaurant	Pizza Place	Breakfast Spot
9	Fourth and Gill	Knoxville	35.985916	-83.921851	1.0	Fast Food Restaurant	Deli / Bodega	Mexican Restaurant	Chinese Restaurant	Sandwich Place
11	Lake Forest	Knoxville	35.926751	-83.867681	1.0	Fast Food Restaurant	Mexican Restaurant	Food	Pizza Place	Wings Joint
12	Lindbergh Forest	Knoxville	36.028778	-83.875321	1.0	Fast Food Restaurant	American Restaurant	Pizza Place	Fried Chicken Joint	Asian Restaurant
15	North Knoxville	Knoxville	36.002027	-83.912407	1.0	American Restaurant	Restaurant	Mexican Restaurant	Sandwich Place	Fried Chicken Joint
16	Norwood	Knoxville	36.004248	-83.981576	1.0	Fast Food Restaurant	Pizza Place	Wings Joint	Sandwich Place	Asian Restaurant
17	Oakwood-Lincoln Park	Knoxville	35.993138	-83.937407	1.0	Restaurant	Breakfast Spot	Sandwich Place	Pizza Place	Wings Joint
18	Old City	Knoxville	35.960395	-83.921026	1.0	American Restaurant	Restaurant	Sandwich Place	New American Restaurant	Café
19	Old North Knoxville	Knoxville	35.986074	-83.921704	1.0	Fast Food Restaurant	Deli / Bodega	Mexican Restaurant	Chinese Restaurant	Sandwich Place
21	Rocky Hill	Knoxville	35.902777	-84.021437	1.0	Asian Restaurant	American Restaurant	Fast Food Restaurant	Breakfast Spot	Mexican Restaurant
24	University of Tennessee	Knoxville	35.951635	-83.930882	1.0	Pizza Place	Sandwich Place	Fast Food Restaurant	Mexican Restaurant	American Restaurant
25	Turkey Creek	Knoxville	35.904191	-84.148873	1.0	Fast Food Restaurant	Sandwich Place	American Restaurant	Breakfast Spot	Sushi Restaurant
26	West Hills	Knoxville	35.937382	-84.035405	1.0	Pizza Place	Sandwich Place	Deli / Bodega	Food	Café

tn_merged[tn_merged['Cluster Labels'] == 2.0]

	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Cedar Bluff	Knoxville	35.924885	-84.105837	2.0	Food	Café	Wings Joint	Donut Shop	Gastropub
14	Mechanicsville	Knoxville	35.969145	-83.931845	2.0	Food	Sandwich Place	BBQ Joint	Deli / Bodega	Gastropub
20	Parkridge	Knoxville	35.981203	-83.906144	2.0	Food	BBQ Joint	Southern / Soul Food Restaurant	Deli / Bodega	Wings Joint
27	West Knoxville	Knoxville	35.970058	-83.955371	2.0	BBQ Joint	Food	Snack Place	Wings Joint	Donut Shop

tn_merged[tn_merged['Cluster Labels'] == 3.0]

	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
13	Lonsdale	Knoxville	35.985638	-83.957130	3.0	American Restaurant	Food	Indian Restaurant	Gastropub	Fried Chicken Joint
23	South Knoxville	Knoxville	35.943139	-83.898794	3.0	Food	Wings Joint	Indian Restaurant	Gastropub	Fried Chicken Joint

tn_merged[tn_merged['Cluster Labels'] == 4.0]

	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	Sequoyah Hills	Knoxville	35.940194	-83.969397	4.0	Bakery	Wings Joint	Donut Shop	Gastropub	Fried Chicken Joint

D. Discussion

Knoxville, TN has a very diverse group of neighborhoods, some areas are heavily populated with food venues (e.g. Fort Sanders, Old City, University of Tennessee) whereas others could use more food places within a 750 meter radius of the neighborhood (e.g. Colonial Village and South Knoxville). As we studied the distribution of which categories of food are in Knoxville, then they appear to have a little bit of everything i.e. Chinese Restaurants, Mexican Restaurants, Indian Restaurants, however top of the list are Fast Food Restaurants. The second most popular food category in Knoxville are Pizza Places. Upon further examination of the data for Pizza Places, we discover that although they are spread all over Knoxville, the most common Pizza Place in Knoxville is Domino's Pizza with the most franchises placed closest to the neighborhood locations.

As we next look at the k-means cluster data, we notice that almost all of the Pizza places in Knoxville are concentrated in Cluster 1.0. For a business that is looking to open a Pizza place in Knoxville there are going to be a few questions that we will have to answer before doing so, and other facts to keep in mind:

Fort Sanders has the most food venues in Knoxville, and the most popular venue is Pizza. They should gather data of population density by neighborhood, and that would allow them to see if it even makes sense to open another Pizza place in this area. It might be better for them to select another location from Cluster 1.0, however where Pizza is perhaps the least common venue, that may prove to be an excellent location as long as we also have population density.

Another thing to keep in mind would be that clusters other than 1.0, seem to have no Pizza places as their most common place. This perhaps may be a prime location to open a place up, however this would require more analysis as they would still have to wonder, 'How come no one else decided to place more Pizza options there?'

E. Conclusion

Knoxville is a growing city, and the diversity of food, number of venues gives an indication towards this. In order for someone to open a Pizza restaurant, they would have to do further analysis keeping population density in mind to make a better decision. K-means clusters show an excellent differentiation of where all of the pizza places in the city have been concentrated, this would be quite powerful for someone to study this deeper, in order to make a location selection for their restaurant.

The results of this post are specific to the location co-ordinates used from geolocator in python. Secondly, we could also get more data for review by expanding the search radius in Foursquare past 750 meters.