Alternative areas to Asheville, North Carolina

IBM Data Science Capstone Project

Meadows, Jonathan
6-16-2021

Table of Contents

Introduction / Business Problem

Asheville, North Carolina (NC) has been a tourist destination since the middle of the 19th century when people came to the area seeking relief from tuberculosis. At the time it was thought the climate promoted healing and relief of tuberculosis and numerous boarding houses and sanitariums were established to support this industry. In more recent years, Asheville has seen double-digit population growth since the 1990s and is commonly featured on lists of top destinations. Because of this continued growth and demand, the cost of housing in Asheville is higher than most other areas in NC.

The objective of this project is to compare Asheville with other metropolitan areas in NC to identify similar areas in terms of businesses and venues but with lower cost housing. There are factors that we are not considering that can be considered unique to Asheville which could ultimately influence a buyer's decision such as the geography and climate.

The audience for this project is people who would like to purchase a single-family home in Asheville but either can't afford the cost or choose not to pay the prices. The intent of this project is to help these people find alternative locations in NC that are similar in terms of businesses/venues to Asheville but with lower cost housing.

Data

This business problem required the use of business/venue data and housing price data by location. The format of the location data was dictated by these respective data sources.

Business / Venue Data Source

Foursquare was chosen as the source of business/venue data due to its large dataset that is easily accessed via an Application Programing Interface (API). The Venues endpoint that contains the business/venue data relies on latitude and longitude coordinates for the basis of query.

Housing Price Data Source

The Federal Housing Finance Agency's House Price Index (HPI) was used to determine how the price of housing compares across US Census Bureau Metropolitan Statistical Areas (MSAs). HPI data from the first quarter of 2021 was used for analysis as this was the most recent data available. The HPI "all-transactions" type was used which includes both refinance mortgages and purchase-only data. This was chosen because it was the only type for which recent data was available that included Asheville, NC. The non-seasonally adjusted index was used since the seasonally adjusted index is not available for the "all-transactions" type.

NOTE: the HPI focuses on single-family house prices and primarily on mortgages that are purchased and/or securitized by Fannie Mae or Freddie Mac.

There are several MSAs listed for NC, however the following subset was selected for comparison: Asheville, Charlotte-Concord-Gastonia, Durham-Chapel Hill, Greensboro-High Point, Hickory-Lenoir-Morganton, Raleigh-Cary, Wilmington, and Winston-Salem.

Location Data Mapping

While Foursquare relies on latitude and longitude location data, the HPI is reported by MSA. As such, locations would need to be mapped between latitude and longitude and MSA. The US Department of Housing and Urban Development (HUD) publishes a crosswalk between MSAs and zip-codes. NOTE: zip-codes can sometimes be located partially inside and outside of an MSA. Zip-codes were then mapped to latitude and longitude using the pgeocode Python library.

Methodology

The HPI and HUD data was loaded, cleaned, and merged resulting in a dataframe of 588 zip-codes across the 8 MSAs. Then pgeocode was used to acquire the latitude and longitude for each zip-code.

	MSA_NAME	MSA_ID	PRICE_INDEX	ZIP	CITY	Latitude	Longitude
0	Asheville, NC	11700	332.44	28655	MORGANTON	35.7346	-81.7042
1	Asheville, NC	11700	332.44	28701	ALEXANDER	35.7064	-82.6311
2	Asheville, NC	11700	332.44	28704	ARDEN	35.4637	-82.5354
3	Asheville, NC	11700	332.44	28709	BARNARDSVILLE	35.7748	-82.4567
4	Asheville, NC	11700	332.44	28710	BAT CAVE	35.4515	-82.2871

The Foursquare API was used to acquire up to 100 venues within 500 meters of each latitude and longitude pair. This resulted in a total of 4,633 venues.

	MSA_NAME	MSA Latitude	MSA Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Asheville, NC	35.7346	-81.7042	Moondog Pizza	35.734903	-81.708535	Pizza Place
1	Asheville, NC	35.7346	-81.7042	fred's Super Dollar	35.735161	-81.709251	Discount Store
2	Asheville, NC	35.7346	-81.7042	New China	35.735888	-81.708612	Chinese Restaurant
3	Asheville, NC	35.7064	-82.6311	Pro-Landscape & Service	35.705105	-82.629976	Business Service
4	Asheville, NC	35.7064	-82.6311	Inspired Ts Co	35.707388	-82.630797	Cosmetics Shop

Each venue category was then aggregated to the MSA.

	MSA_NAME	ATM	Accessories Store	African Restaurant	American Restaurant	Amphitheater	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports
0	Asheville, NC	0.000000	0.004566	0.000000	0.018265	0.0	0.004566	0.004566	0.000000	0.000000	0.000000	0.000000
1	Charlotte-Concord-Gastonia, NC-SC	0.005889	0.000000	0.000000	0.044433	0.0	0.000535	0.001606	0.000535	0.000535	0.002141	0.000000
2	Durham-Chapel Hill, NC	0.006565	0.000000	0.000000	0.026258	0.0	0.002188	0.004376	0.002188	0.004376	0.004376	0.004376
3	Greensboro-High Point, NC	0.001199	0.001199	0.001199	0.058753	0.0	0.000000	0.004796	0.000000	0.000000	0.001199	0.000000
4	Hickory-Lenoir-Morganton, NC	0.000000	0.000000	0.000000	0.076923	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

The top 10 venue categories were evaluated for each MSA.

	MSA_NAME	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Asheville, NC	Hotel	Fast Food Restaurant	Home Service	Construction & Landscaping	Discount Store
1	Charlotte-Concord-Gastonia, NC-SC	American Restaurant	Pizza Place	Italian Restaurant	Sandwich Place	Steakhouse
2	Durham-Chapel Hill, NC	Fast Food Restaurant	Hotel	Pizza Place	Mexican Restaurant	Pharmacy
3	Greensboro-High Point, NC	Bar	American Restaurant	Sandwich Place	Brewery	Plaza
4	Hickory-Lenoir-Morganton, NC	American Restaurant	Post Office	Pizza Place	Discount Store	Home Service

K-Means clustering was used to group the MSAs based on their venue categories. K-Means clustering was repeated for each cluster size from 2 to 7. A cluster size of 6 and 7 both resulted in the most granular cluster where Asheville and Durham-Chapel Hill were the only two MSAs in the same cluster. Refer to the Appendix for the results of each cluster size 2 – 7.

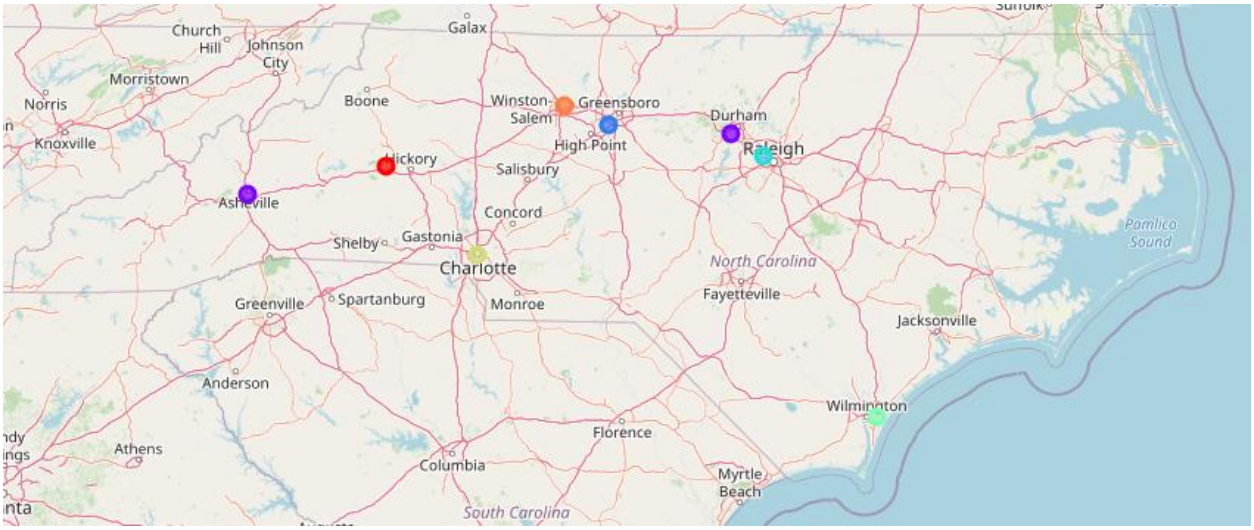
MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	4
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	5
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	6
Greensboro-High Point, NC	24660	190.91	2

Results

The analysis revealed that Durham-Chapel Hill was the closest neighbor to Asheville in terms of the businesses/venues in the MSAs. The Durham-Chapel Hill HPI for the first quarter of 2021 was 246.64 compared with Asheville at 332.44. A cluster size of 6 and 7 both resulted in Asheville and Durham-Chapel Hill being the only two MSAs grouped together.

Results of K-Means cluster size of 7:

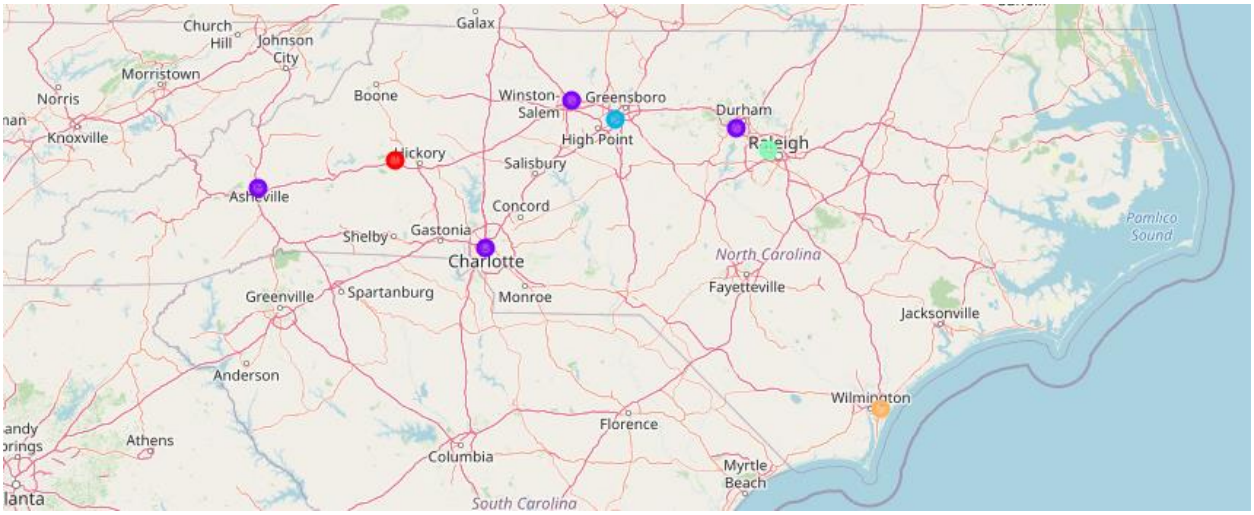
MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	4
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	5
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	6
Greensboro-High Point, NC	24660	190.91	2



A cluster size of 5 resulted in a cluster containing Asheville, Charlotte-Concord-Gastonia, Durham-Chapel Hill, and Winston-Salem. Within this larger group the lowest HPI was for Winston-Salem at 196.06.

Results of K-Means cluster size of 5:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	4
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	2



Discussion

The results reveal that for anyone wanting to purchase a single-family home in Asheville should consider looking in the Durham-Chapel Hill area where prices are lower, but business/venues offer similar products and services. While this analysis focused on business/venues, additional research should take into consideration other attributes including, but not limited to, geography, climate, etc.

Conclusion

This project set out to identify areas in NC that are similar to Asheville in terms of business/venues but with lower cost housing. The results revealed that Durham-Chapel Hill is the most similar area with lower cost housing. Thus, for anyone wanting to purchase a single-family home in Asheville should consider looking in the Durham-Chapel Hill area. Recall that this analysis focused solely on business/venues. Future research should take other attributes into consideration including geography, climate, etc. and consider expanding the analysis to areas beyond NC.

References

Foursquare documentation: https://developer.foursquare.com/docs/places-api/

The Federal Housing Finance Agency documentation:
https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx

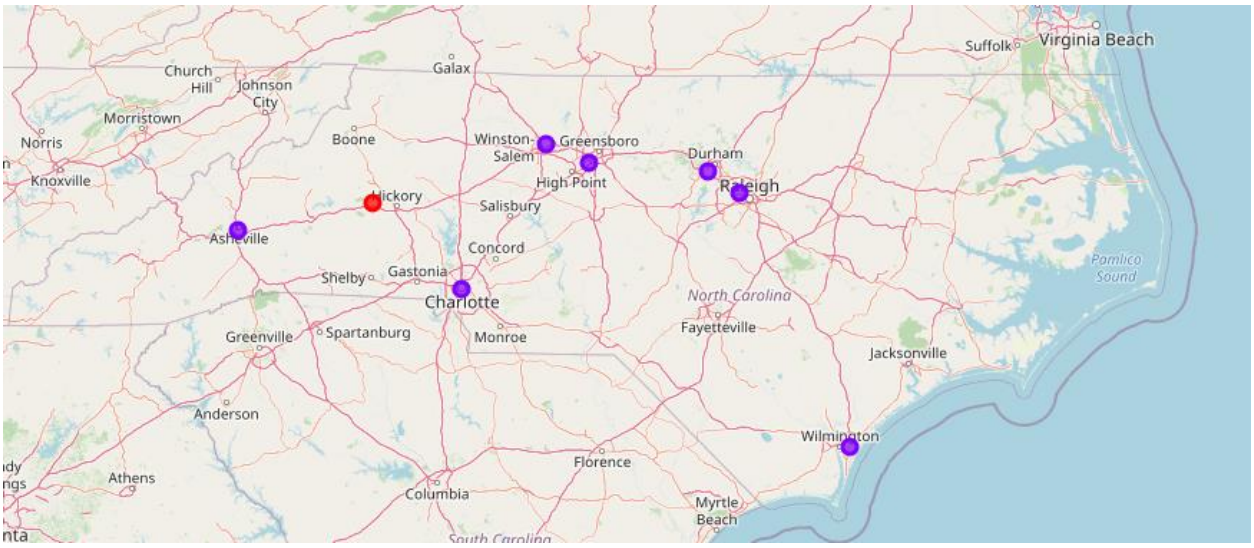
The Department of Housing and Urban Development documentation:
https://www.huduser.gov/portal/datasets/usps_crosswalk.html#data

pgeocode library: https://pypi.org/project/pgeocode/

Appendix

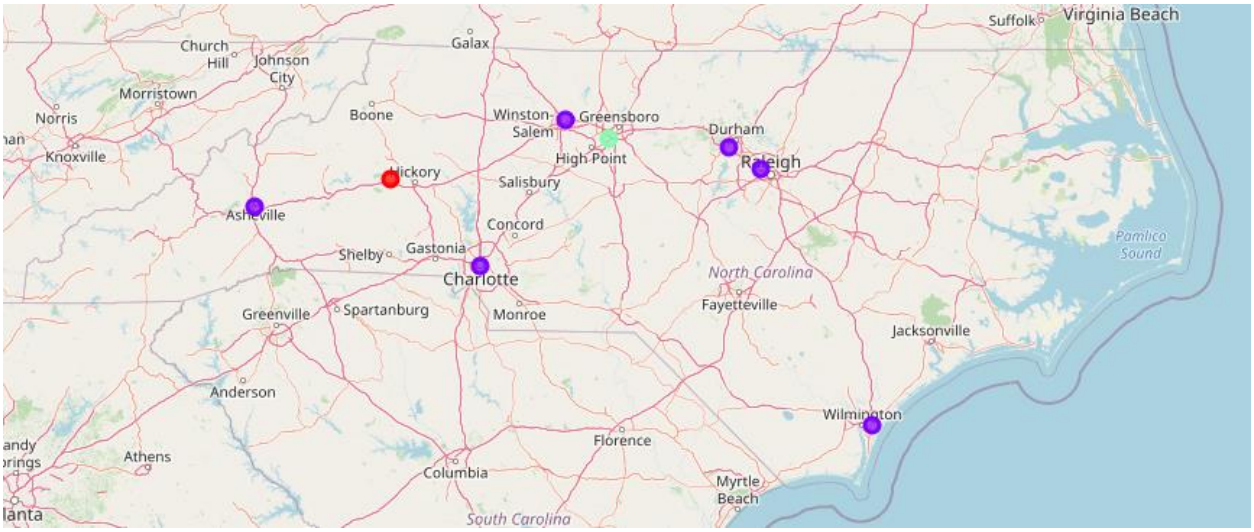
Results from K-Means cluster size of 2:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	1
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	1
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	1



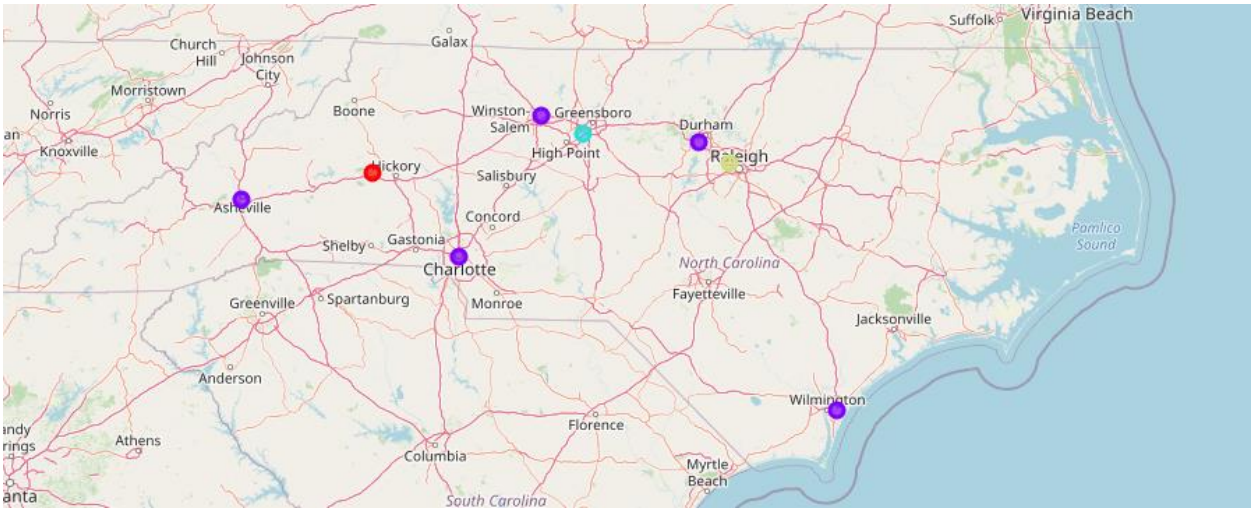
Results from K-Means cluster size of 3:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	1
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	1
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	2



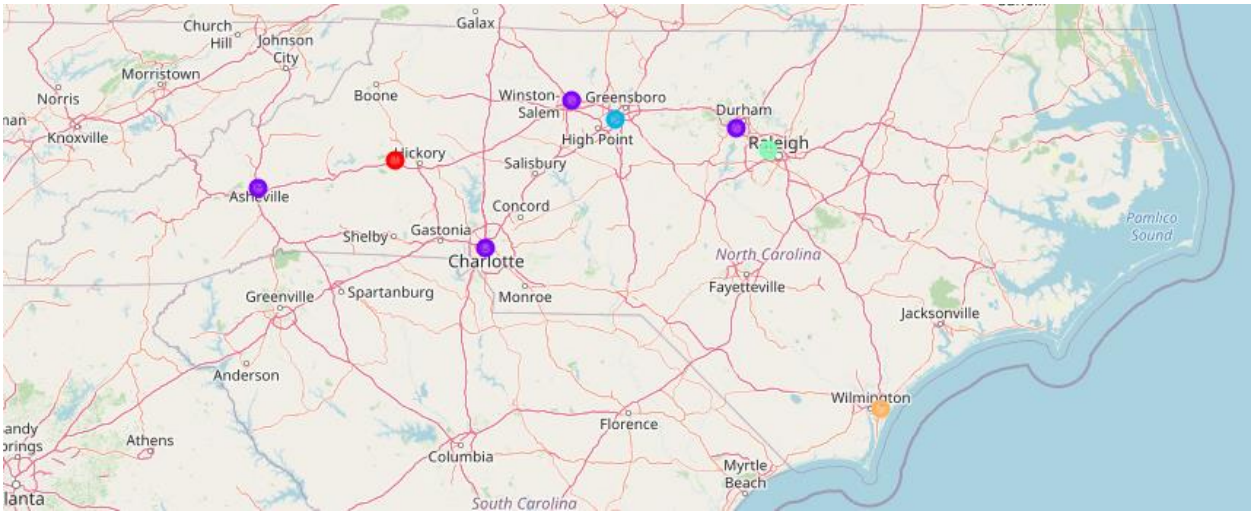
Results from K-Means cluster size of 4:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	1
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	2



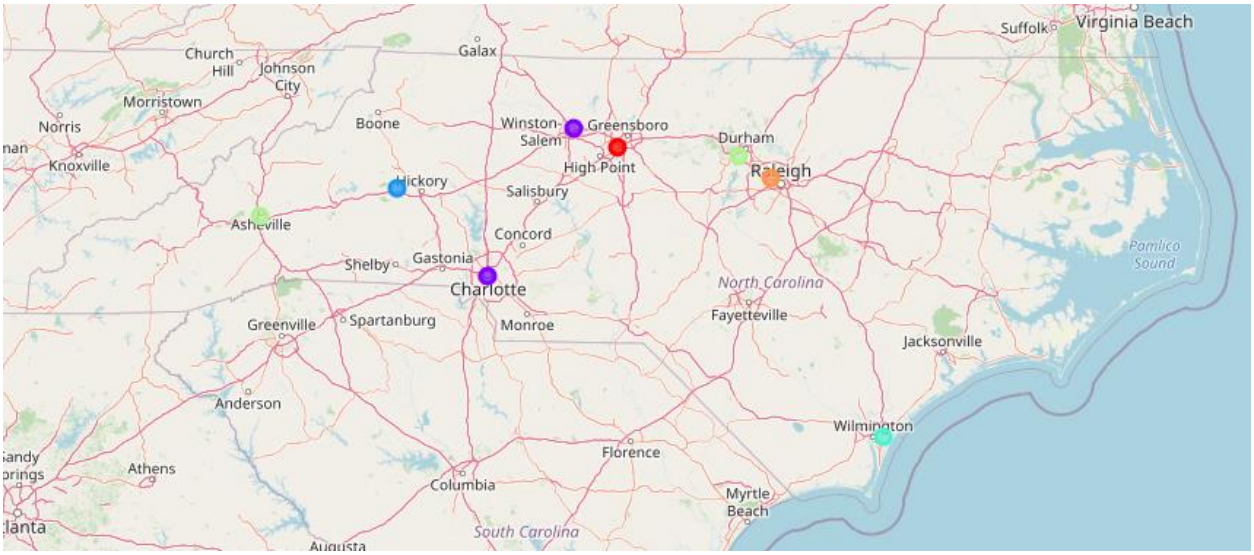
Results from K-Means cluster size of 5:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	4
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	2



Results from K-Means cluster size of 6:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	4
Wilmington, NC	48900	266.57	3
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	1
Durham-Chapel Hill, NC	20500	246.64	4
Raleigh-Cary, NC	39580	236.30	5
Hickory-Lenoir-Morganton, NC	25860	217.67	2
Winston-Salem, NC	49180	196.06	1
Greensboro-High Point, NC	24660	190.91	0



Results from K-Means cluster size of 7:

MSA_NAME	MSA_ID	PRICE_INDEX	Cluster Labels
Asheville, NC	11700	332.44	1
Wilmington, NC	48900	266.57	4
Charlotte-Concord-Gastonia, NC-SC	16740	257.81	5
Durham-Chapel Hill, NC	20500	246.64	1
Raleigh-Cary, NC	39580	236.30	3
Hickory-Lenoir-Morganton, NC	25860	217.67	0
Winston-Salem, NC	49180	196.06	6
Greensboro-High Point, NC	24660	190.91	2