Using k-Means Clustering to Determine Ideal Locations to Open an Art Gallery in Research Triangle, North Carolina

IBM Data Science Professional Certificate: Capstone Project

Sam Neubauer

December 15th, 2020

Introduction

○ This project came to life with a real desire to solve a question me and my partner have been trying to answer for a while:

Question:

Where is the best place to open an art gallery in Research Triangle area of NC?

○ My partner is at a scaling growth stage with her business and wants to open a gallery in the right location of the Research Triangle. This analysis will provide a helpful insight in deciding where to take the next step in her career.

Data (North Carolina Zip Codes)

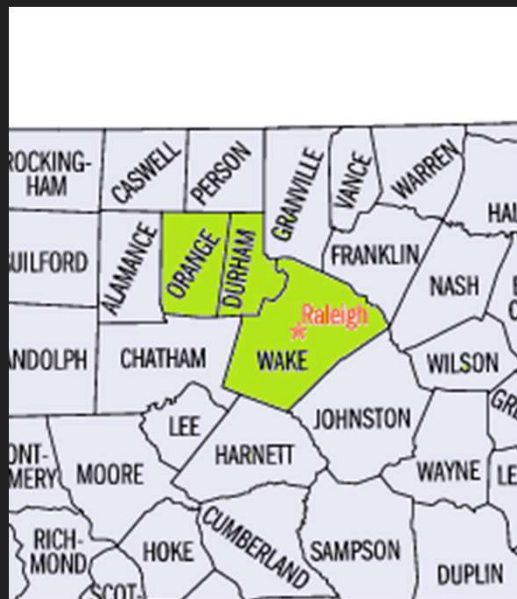○ **CSV file with all zip codes in North Carolina source:**
https://data.opendatasoft.com/explore/dataset/geonames-postal-code@public-us/download/?format=csv&q=&refine.country_code=US&refine.admin_name1=North+Carolina&timezone=America/New_York&use_labels_for_header=true&csv_separator=%3B

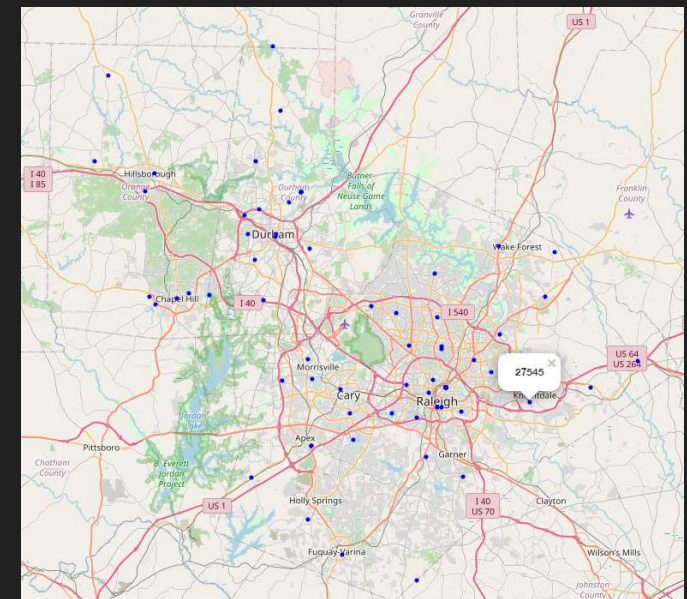|	country code	postal code	place name	admin name1	admin code1	admin name2	admin code2	admin name3	admin code3	latitude	longitude	accuracy	coordinates
0	US	27349	Snow Camp	North Carolina	NC	Alamance	1	NaN	NaN	35.9066	-79.4279	4.0	35.9066,-79.4279
1	US	28678	Stony Point	North Carolina	NC	Alexander	3	NaN	NaN	35.8661	-81.0641	4.0	35.8661,-81.0641
2	US	28675	Sparta	North Carolina	NC	Alleghany	5	NaN	NaN	36.5089	-81.1384	4.0	36.5089,-81.1384
3	US	28626	Fleetwood	North Carolina	NC	Ashe	9	NaN	NaN	36.2814	-81.5140	4.0	36.2814,-81.514
4	US	28640	Jefferson	North Carolina	NC	Ashe	9	NaN	NaN	36.4090	-81.4396	4.0	36.409,-81.4396
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1076	US	27530	Goldsboro	North Carolina	NC	Wayne	191	NaN	NaN	35.3683	-78.0929	4.0	35.3683,-78.0929
1077	US	27863	Pikeville	North Carolina	NC	Wayne	191	NaN	NaN	35.4929	-77.9578	4.0	35.4929,-77.9578
1078	US	28578	Seven Springs	North Carolina	NC	Wayne	191	NaN	NaN	35.2105	-77.9146	4.0	35.2105,-77.9146
1079	US	28651	Millers Creek	North Carolina	NC	Wilkes	193	NaN	NaN	36.2119	-81.2485	4.0	36.2119,-81.2485
1080	US	27894	Wilson	North Carolina	NC	Wilson	195	NaN	NaN	35.7158	-77.9043	4.0	35.7158,-77.9043

1081 rows × 13 columns

Data (North Carolina Zip Codes)

○ Data was then cleaned, removing NaN/unneeded columns, and filtered to only include the counties within the Research Triangle area: Wake, Durham, Orange

	zip	place	county	latitude	longitude
0	27704	Durham	Durham	36.0383	-78.8764
1	27710	Durham	Durham	36.0512	-78.8577
2	27711	Durham	Durham	36.0512	-78.8577
3	27529	Garner	Wake	35.6813	-78.5975
4	27615	Raleigh	Wake	35.8887	-78.6393
...	...	...	...	...	...
86	27523	Apex	Wake	35.7225	-78.8408
87	27560	Morrisville	Wake	35.8344	-78.8466
88	27571	Rolesville	Wake	35.9156	-78.4658
89	27602	Raleigh	Wake	35.7587	-78.6711
90	27695	Raleigh	Wake	35.7977	-78.6253

91 rows × 5 columns

Data (Foursquare API Venues)

Foursquare was utilized to obtain search results for the following queries within a 3-mile radius at each zip code:
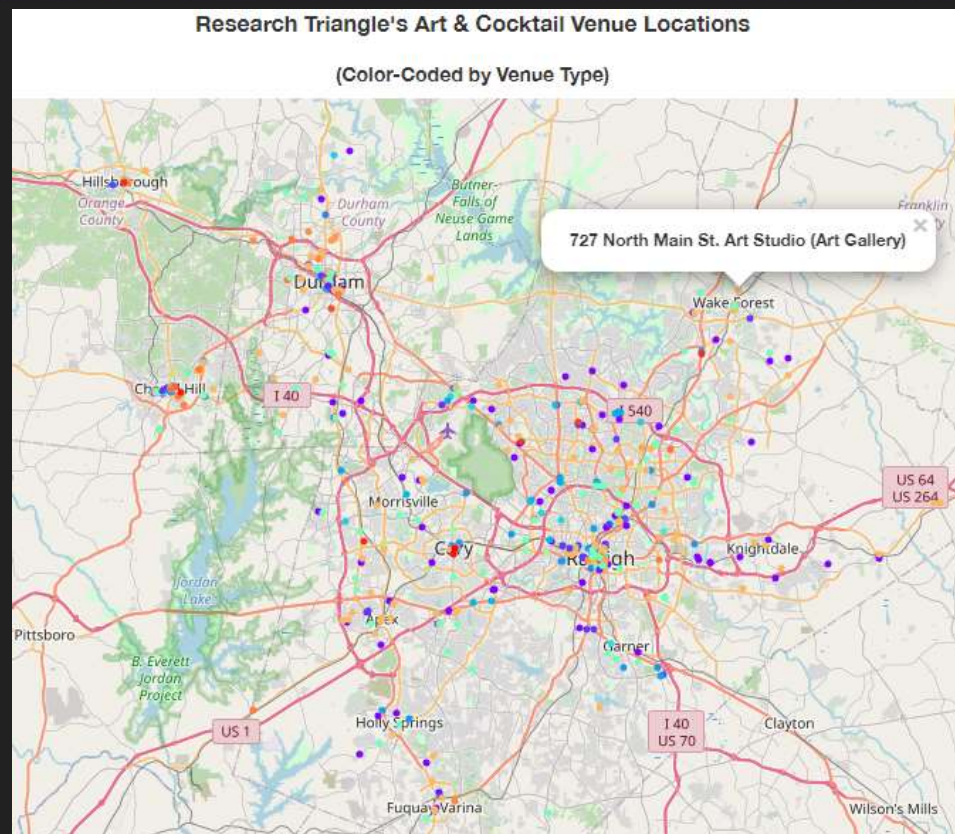
- brewery
- cocktails
- distillery
- gallery
- museum
- music venues

	Zip	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Code
0	27704	36.0383	-78.8764	Fullsteam Brewery	36.003166	-78.899773	Brewery	0
1	27704	36.0383	-78.8764	Durham Distillery	36.004452	-78.902913	Brewery	0
2	27704	36.0383	-78.8764	Durty Bull Brewing Co.	36.001353	-78.898207	Brewery	0
3	27704	36.0383	-78.8764	Geer Street Garden	36.003601	-78.901057	Gastropub	1
4	27704	36.0383	-78.8764	Motorco Music Hall	36.003599	-78.900325	Concert Hall	2
...	...	...	...	...	...	...	...	...
5868	27712	36.0918	-78.9299	Choir Ready Room	36.063362	-78.926900	Music Venue	30
5869	27712	36.0918	-78.9299	Barry & Claire Ragin's Home	36.071176	-78.961081	Music Venue	30
5870	27588	35.9731	-78.4508	Hater's Ball	35.941644	-78.475700	Music Venue	30
5871	27607	35.8014	-78.6877	Museum Park Theater	35.809629	-78.703345	Music Venue	30
5998	27560	35.8344	-78.8466	NC Violinist	35.810209	-78.816562	Music Venue	30

567 rows × 8 columns

Methodology

- The venues were color-coded and mapped for some exploratory data analysis to visually determine where densities of certain venue types exist within the area

- There is a noticeable concentration of art galleries (orange) in the Durham, Wake Forest, and Apex towns.

- There are also many breweries (purple) in the Research Triangle area

Research Triangle's Art & Cocktail Venue Locations

(Color-Coded by Venue Type)

727 North Main St. Art Studio (Art Gallery)

Methodology

- One-hot coding functions were utilized to return the top 10 venue types within each zip code

- To our excitement we notice many zip codes have art galleries and other social opportunities in the top 3 most common venue types.

	Zip	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	27278	Art Gallery	History Museum	Bar	Museum	Music Venue	Rock Club	Cocktail Bar	BBQ Joint	Beer Garden	Historic Site
1	27502	Art Gallery	Brewery	Music Venue	Pub	Irish Pub	Italian Restaurant	Bar	Dive Bar	Concert Hall	Deli / Bodega
2	27510	Bar	Brewery	Art Gallery	American Restaurant	Music Venue	Museum	Cocktail Bar	Asian Restaurant	Sandwich Place	Distillery
3	27511	Brewery	Art Gallery	Pub	Music Venue	Sports Bar	American Restaurant	Breakfast Spot	Italian Restaurant	Performing Arts Venue	History Museum
4	27512	Music Venue	Pub	Art Gallery	New American Restaurant	Dive Bar	College Basketball Court	Concert Hall	Deli / Bodega	Distillery	Event Space
...	...	...	...	...	...	...	...	...	...	...	...
82	27712	Music Venue	Brewery	Art Gallery	Art Museum	Speakeasy	Wings Joint	Dive Bar	Concert Hall	Deli / Bodega	Distillery
83	27713	Art Gallery	Music Venue	Brewery	Pub	Beer Store	Dive Bar	Concert Hall	Deli / Bodega	Distillery	Event Space
84	27715	Science Museum	Arts & Crafts Store	Brewery	Wings Joint	Event Space	Concert Hall	Deli / Bodega	Distillery	Dive Bar	Gastropub
85	27717	Science Museum	Arts & Crafts Store	Brewery	Wings Joint	Event Space	Concert Hall	Deli / Bodega	Distillery	Dive Bar	Gastropub
86	27722	Art Gallery	Brewery	Bar	Cocktail Bar	Music Venue	Concert Hall	Sports Bar	College Arts Building	Museum	Lounge

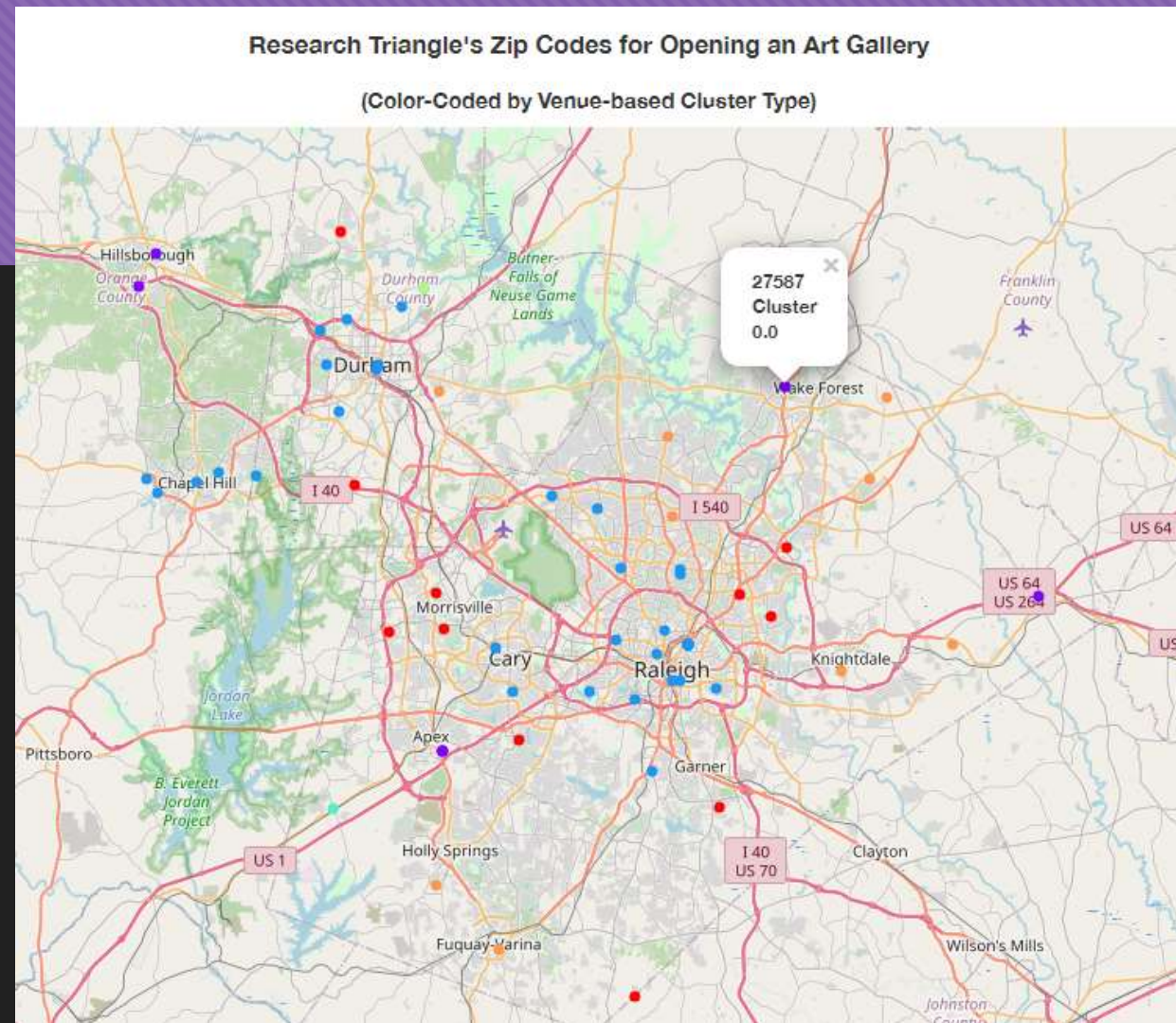87 rows × 11 columns

Methodology

- We ran the k-means from the SKLearn library to cluster the neighborhood into 6 clusters based on most common venue types.

- This machine learning algorithm was chosen due to its speed, ease, and the fact that we must use an unsupervised classification method to group the zip codes since there is not an easy and obvious way to characterize the zip codes via explicit (supervised) training methods.

	zip	place	county	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	27704	Durham	Durham	36.0383	-78.8764	1.0	Art Gallery	Concert Hall	Brewery	Science Museum	Bar	Music Venue	Gastropub	History Museum	Historic Site	Arts & Crafts Store
1	27710	Durham	Durham	36.0512	-78.8577	3.0	Science Museum	Arts & Crafts Store	Brewery	Wings Joint	Event Space	Concert Hall	Deli / Bodega	Distillery	Dive Bar	Gastropub
2	27711	Durham	Durham	36.0512	-78.8577	3.0	Science Museum	Arts & Crafts Store	Brewery	Wings Joint	Event Space	Concert Hall	Deli / Bodega	Distillery	Dive Bar	Gastropub
3	27529	Garner	Wake	35.6813	-78.5975	5.0	Music Venue	Wings Joint	Sports Bar	Art Gallery	Brewery	Restaurant	American Restaurant	Steakhouse	Speakeasy	General Entertainment
4	27615	Raleigh	Wake	35.8887	-78.6393	4.0	Brewery	Art Gallery	Pizza Place	Pub	American Restaurant	Bar	Sandwich Place	Japanese Restaurant	Beer Store	Speakeasy

Results

The zip codes were clustered into 6 segments using the K-means algorithm

Cluster 0 : •   Cluster 3 : •

Cluster 1 : •   Cluster 4 : •

Cluster 2 : •   Cluster 5 : •

Research Triangle's Zip Codes for Opening an Art Gallery

(Color-Coded by Venue-based Cluster Type)

Discussion

- **Cluster 0 Zip Codes**: are by far the *most ideal* locations to open an art gallery since galleries are the most common venue found in this cluster. The 4 runner-up venue types condone many social opportunities for the perfect day/night out on the town enjoying music, art, and drinks with friends and/or family.

```
1st Common: Art Gallery
2nd Common: Brewery
3rd Common: Music Venue
4th Common: Pub
5th Common: Irish Pub
```

- **Cluster 1 Zip Codes**: are still *ideal* locations to open an art gallery since galleries are the 2nd most common venue found in this cluster. The other top venue types are rich in socializing and live music options.

```
1st Common: Bar
2nd Common: Art Gallery
3rd Common: Music Venue
4th Common: Brewery
5th Common: Cocktail Bar
```

- **Cluster 2 Zip Codes:** is unfortunately the *least ideal* location to open an art gallery since the common venue options lack art, entertainment, and social options. While there is historical destinations, there is not as many cultural opportunities as the other clusters.

```
1st Common: Museum
2nd Common: Wings Joint
3rd Common: Hotel
4th Common: History Museum
5th Common: Historic Site
```

- **Cluster 3 Zip Codes**: are *not ideal* locations for an art gallery since there are not entertainment or other cultural common venue options. The event space may hold event that attract customers for an art gallery, but there's less cultural venue density than **Clusters 0, 1,** & **4**.

```
1st Common: Science Museum
2nd Common: Arts & Crafts Store
3rd Common: Brewery
4th Common: Wings Joint
5th Common: Event Space
```

- **Cluster 4 Zip Codes**: like **Cluster 1**, this cluster still highlights *ideal* locations to open an art gallery since galleries are the 2nd most common venue found in this cluster. With eating, drinking, historic locations and a vicinity to college art buildings, these locations may attract a diversity of customer demographics.

```
1st Common: Brewery
2nd Common: Art Gallery
3rd Common: Wings Joint
4th Common: College Arts Building
            Wings Joint
5th Common: Event Space
            History Museum
```

- **Cluster 2 Zip Codes:** are *acceptable* locations to open an art gallery. The most common venue type is a music venue which suggests local patrons might appreciate art. However, most of the common venue types include food & drink which doesn't provide much assurance that people going to these venues would frequent an art venue. They are more likely to want to socialize than yearn to experience art.

```
1st Common: Music Venue
2nd Common: Brewery
3rd Common: Brewery
4th Common: American Restaurant
5th Common: Speakeasy
```

Conclusion

- Analyzing the art, music, and social venues of the Research Triangle area has provided a much clearer understanding of the culture and commerce that occurs within the region. We now know that the 6 zip codes within Cluster 0 are the most ideal zip codes to open an art gallery, while the zip codes in Cluster 1 & Cluster 4 are still ideal, providing ample entertainment and socializing opportunities. The 18 zip codes within Clusters 2, 3, & 5 are not as ideal candidates to open an art gallery due to having less culturally rich venue opportunities.

- With this, it would be wise to become more familiar with each of these zip codes in person to help get a realistic idea of what each area feels like, and what kind of overhead costs must be endured to open a store-front location.

- It was very exciting to give insight towards a problem I wanted to solve for my partner's business, and I seriously look forward to making this investigation a reality! Thank you for viewing my project!

References

- NC Zip Code Geospatial Locations: https://developer.foursquare.com/developer/

- NC County Map: https://www.ereferencedesk.com/resources/counties/images/north-carolina-county-map.gif

- Foursquare API: https://developer.foursquare.com/developer/