

**PREDICTING THE BEST NEIGHBORHOOD  
TO ESTABLISH A NEW ITALIAN  
RESTAURANT IN HOUSTON, TX**



# INTRODUCTION

- Finding the right neighborhood to establish a new Italian Restaurant will be key in the profitability of the Restaurant.
- Houston, Texas is the largest city in Texas and is one of the largest Cities in the US and it has a total area of 637.4 square miles.
- Houston is divided into 88 different Super Neighborhoods
- Data can might contribute to determining the ideal Super Neighborhood to establish a new Italian Restaurant
- This analysis of this data will help predict what Super Neighborhood will have the lowest number of competitors and have a larger number of potential customers (ie. Tourists).



# DATA ACQUISITION AND CLEANING

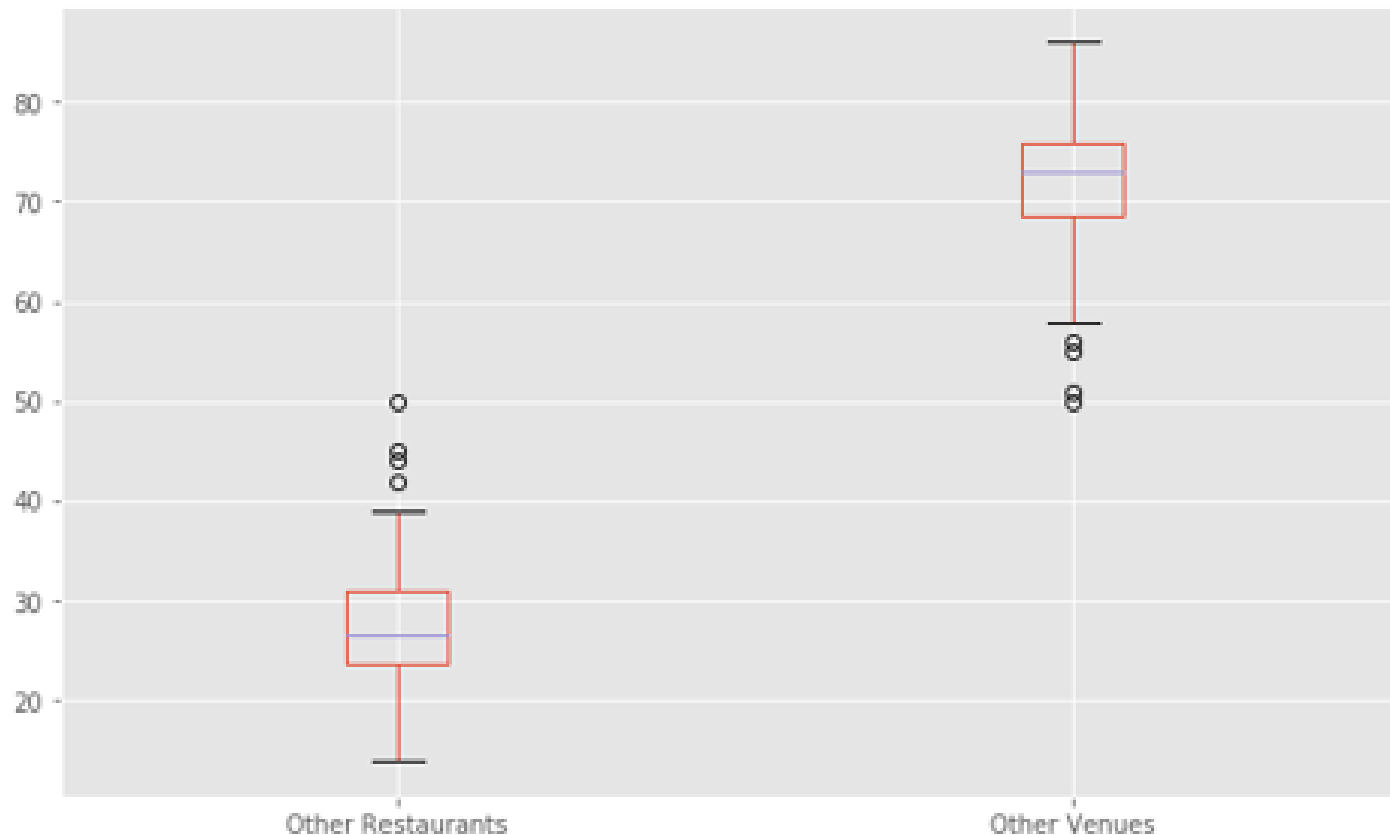
- The geo data for the Super Neighborhoods boundaries will come from <https://cohgis-mycity.opendata.arcgis.com/datasets/coh-super-neighborhoods>
- The restaurant and Venue data will come from the Foursquare API
- The Super Neighbors were clean and center points were added for each of the 88 Super Neighborhoods
- The center points were used to query the Four Square API and returned 8,756 venues total
- An additional dataset was created for counts of each of the macro categories (Italian, Other Restaurants, and Other Venues).

	Neighborhood	Radius	Latitude	Longitude	Italian	Other Restaurants	Other Venues	sum	Italian percent
0	CENTRAL NORTHWEST	41925.603578	29.828763	-95.444862	2	27	73	102	1.960784
1	BRAESWOOD	20043.380925	29.692534	-95.431955	2	23	77	102	1.960784
2	MIDTOWN	13764.248611	29.741398	-95.374541	2	22	78	102	1.960784
3	GREATER INWOOD	56834.519261	29.867343	-95.477039	2	24	76	102	1.960784
4	MID WEST	35099.216395	29.731516	-95.509366	2	28	72	102	1.960784



# FINDING THE RELATIONSHIP TO ITALIAN RESTAURANTS

Effects of Other Restaurants and Other Venues on number of Italian Restaurants



The Box plot depicted below shows a correlation that a Italian restaurant is more likely in Neighborhoods with more Other Venues and less Other Restaurants.



# CLUSTERING THE NEIGHBORHOODS

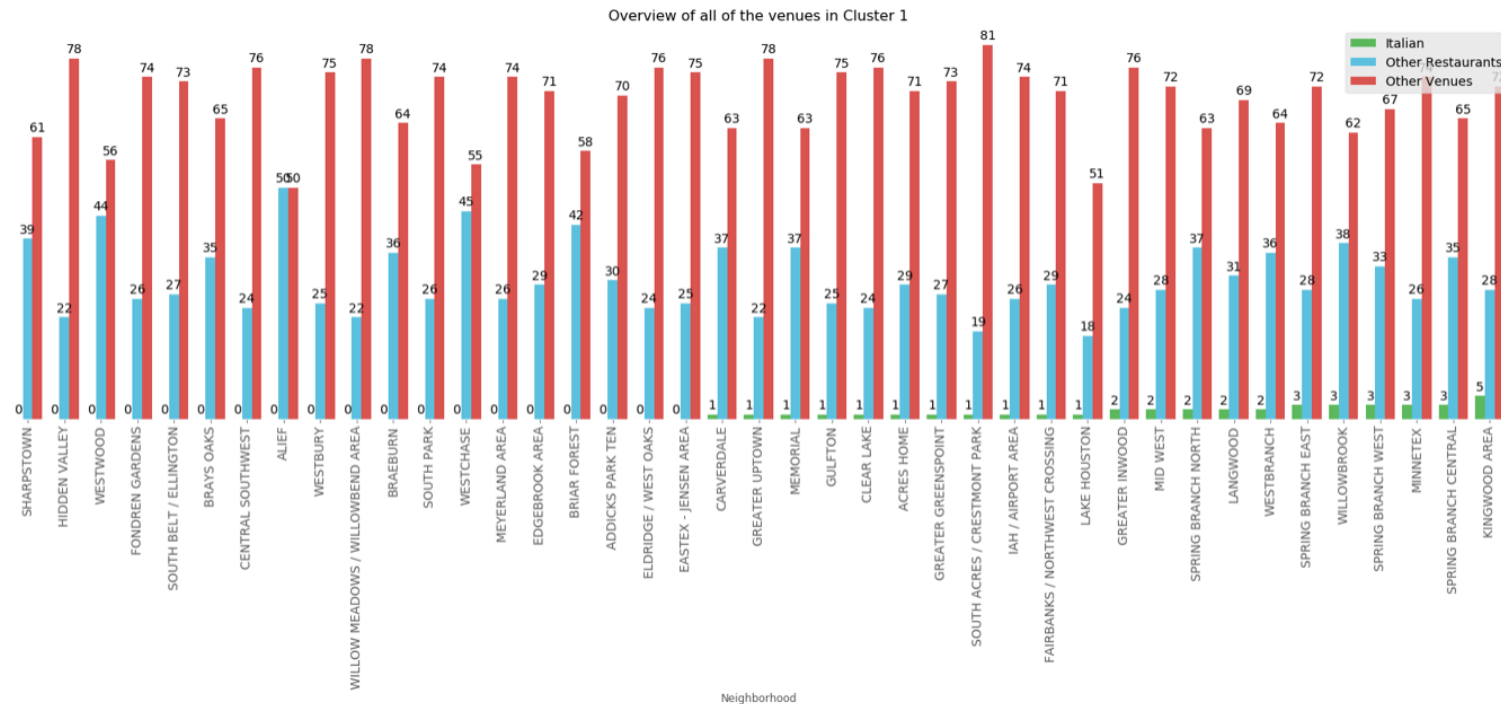
- Created a new dataset with the 10 most common venues per Super Neighborhood
- This helps us see similarities between the different Super Neighborhoods
- I then used the K-means clustering technique to create 5 Clusters of Similar Neighborhoods.

Neighborhood	Radius	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
FOURTH WARD	8286.013012	29.757069	-95.381041	1	Coffee Shop	Park	American Restaurant	Grocery Store	Mexican Restaurant	Trail	Baseball Stadium	Taco Place	Italian Restaurant	New American Restaurant
SECOND WARD	21559.385008	29.748903	-95.326587	1	Park	Coffee Shop	Vietnamese Restaurant	Brewery	Mexican Restaurant	Bar	Sushi Restaurant	Trail	Burger Joint	Theater
DOWNTOWN	19628.193746	29.757909	-95.359908	1	Coffee Shop	Park	Mexican Restaurant	American Restaurant	Baseball Stadium	Burger Joint	Brewery	Trail	Beer Garden	Concert Hall



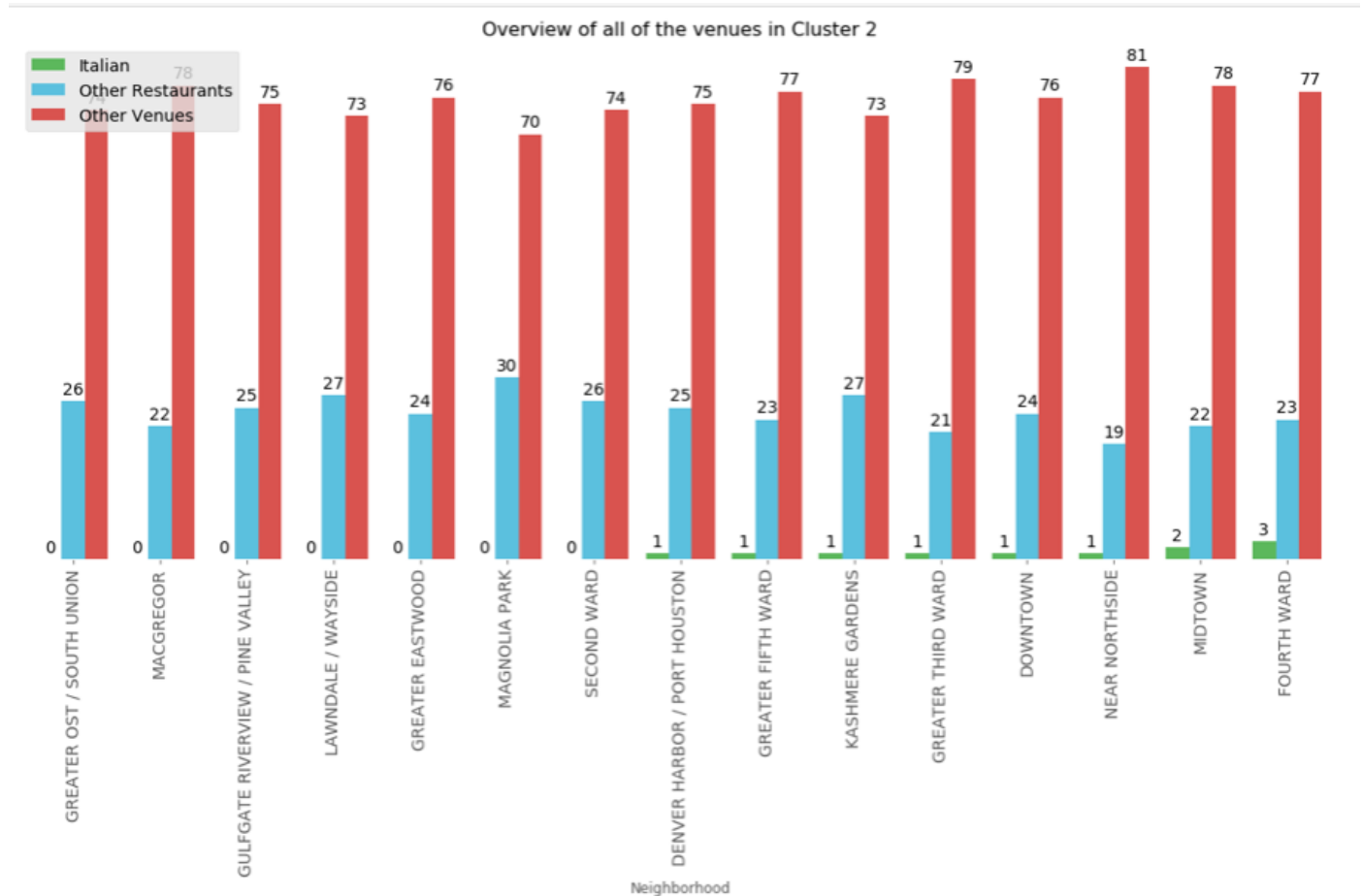
# CLUSTER 1 COMPARISON

- Cluster 1 is the largest cluster and almost half of the Neighborhoods within Cluster 1 have no Italian Restaurants. The 2 most common venues in cluster 1 are Burger Joint and Mexican Restaurants.



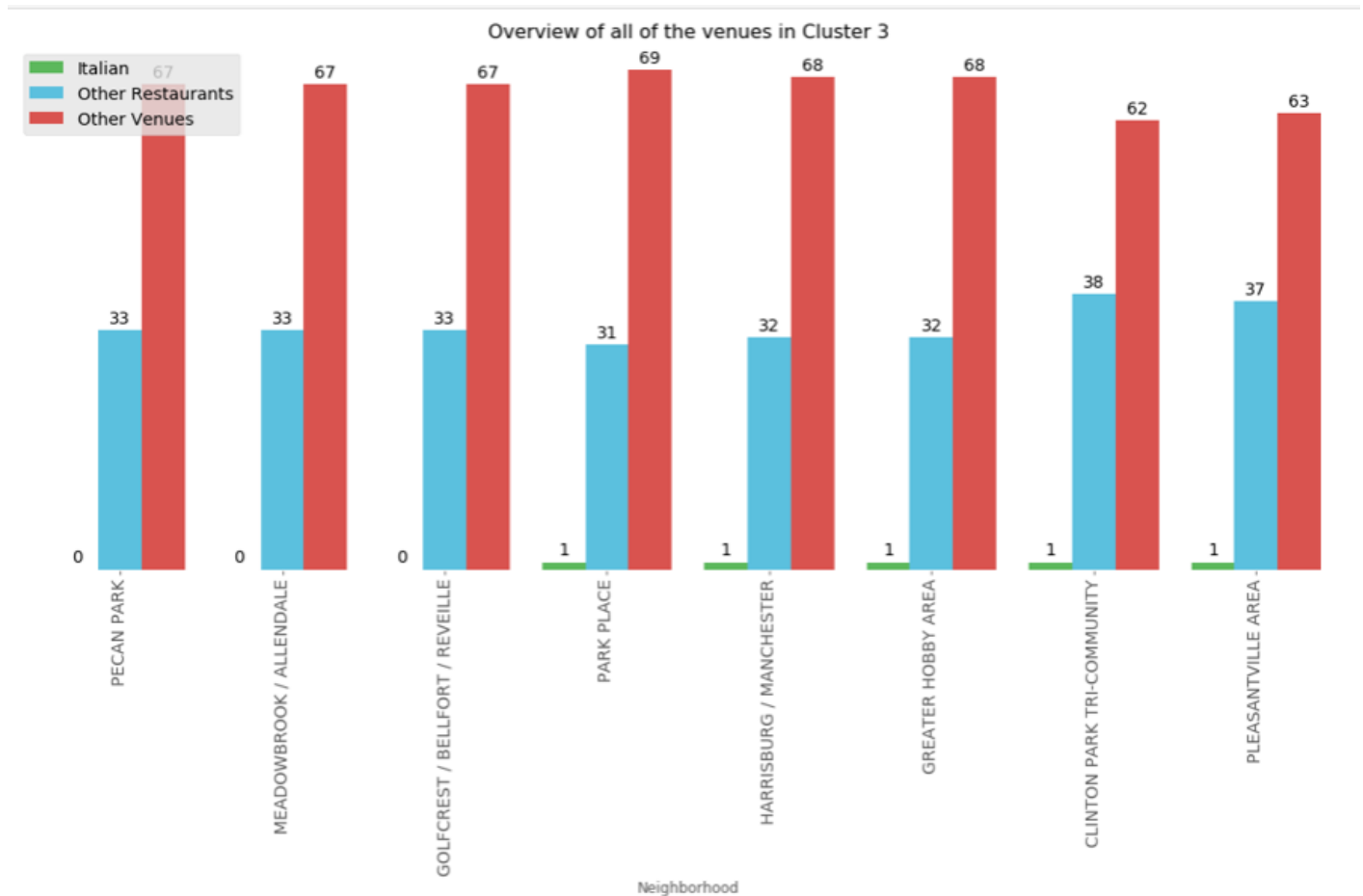
# CLUSTER 2 COMPARISON

- Cluster 2 is a smaller cluster with almost half of the Neighborhoods in Cluster 2 having no Italian Restaurants. The 2 most common Venues in Cluster 2 are Park and Coffee Shop.



# CLUSTER 3 COMPARISON

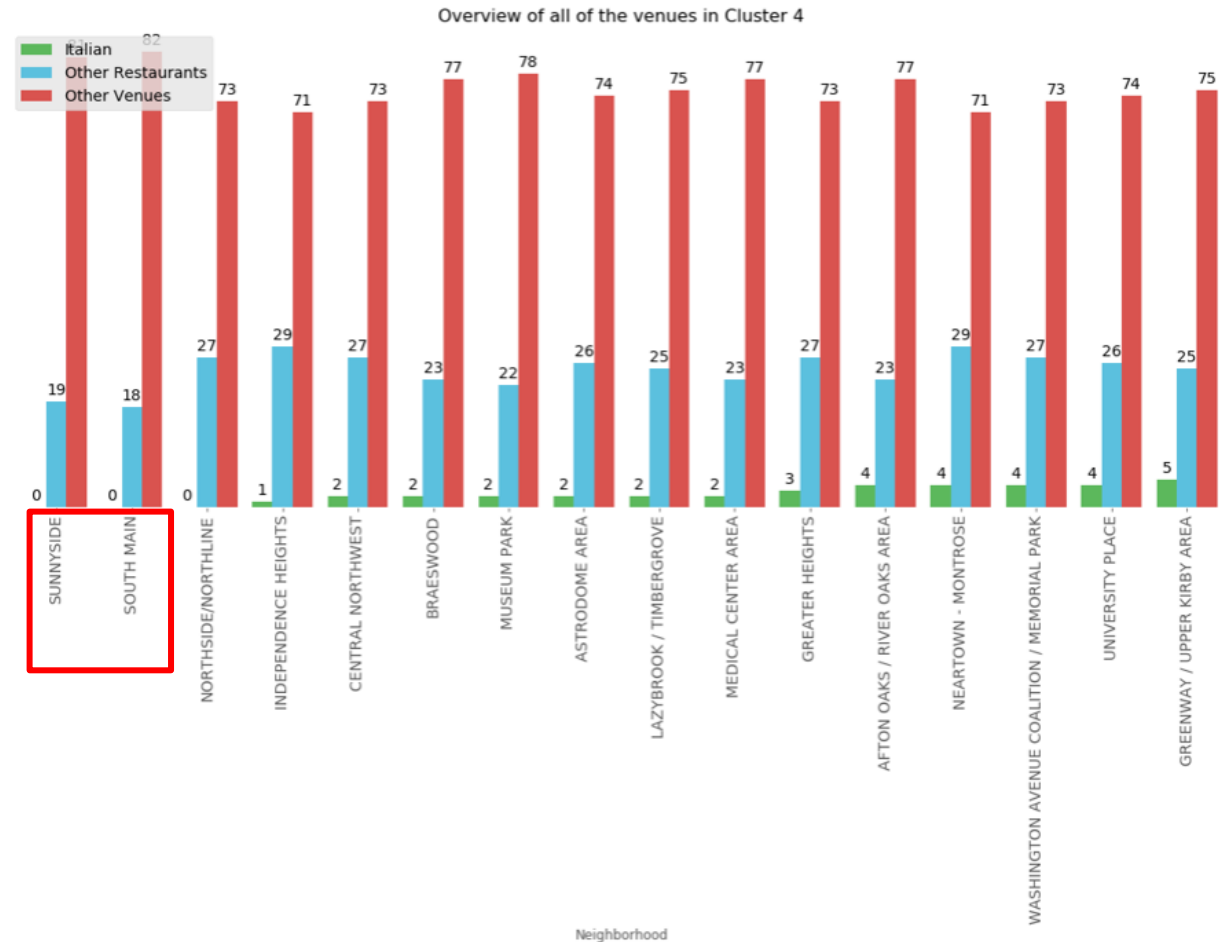
- Cluster 3 is the smallest of the 5 clusters and the majority of the clusters have Italian Restaurants, but the Neighborhoods that do have Italian Restaurants only have one Italian Restaurant. The most common venues in Cluster 3 are Mexican Restaurants and Burger Joints.





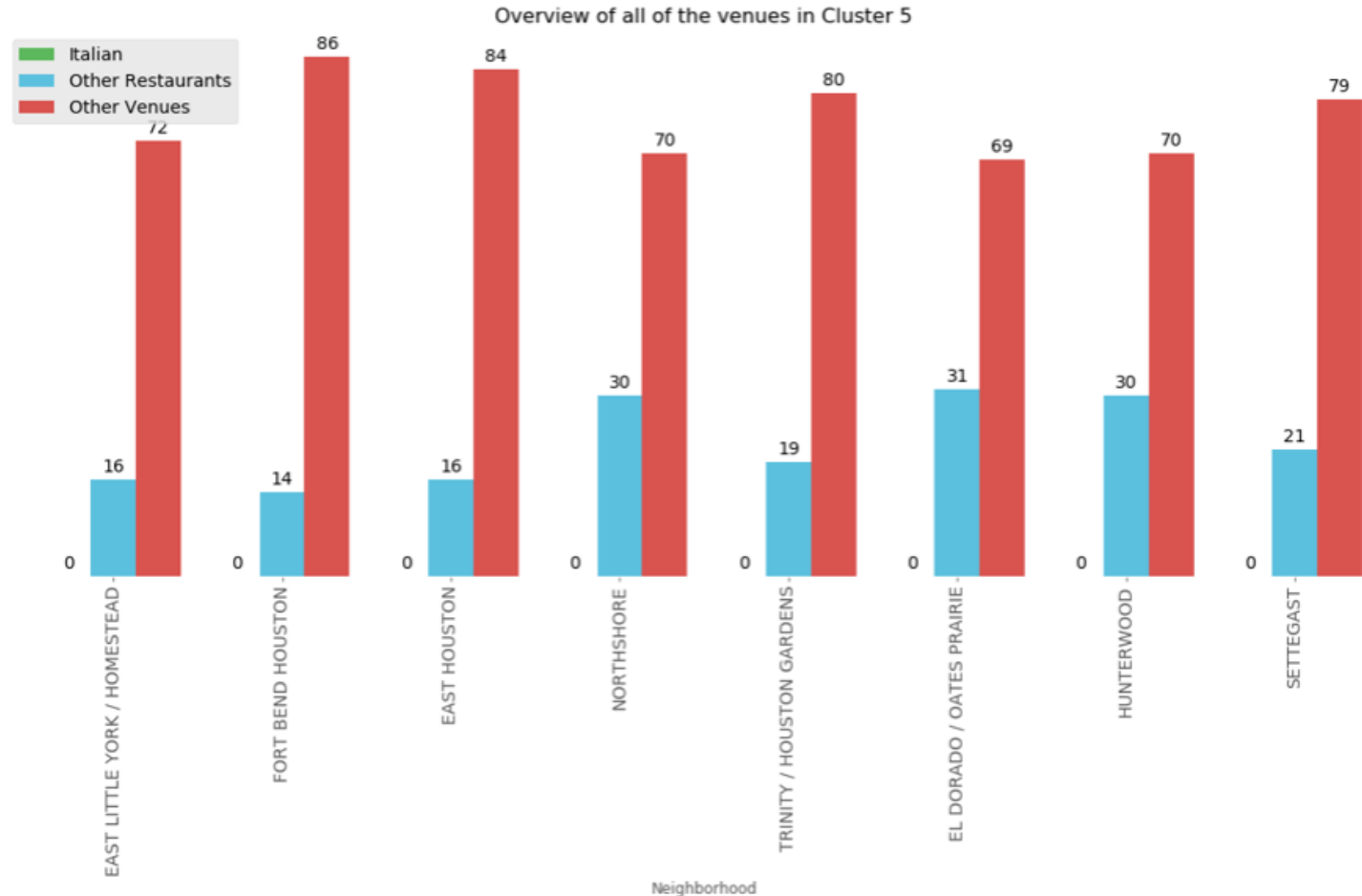
# CLUSTER 4 COMPARISON

- Cluster 4 is primarily comprised of Neighborhoods with Multiple Italian Restaurants. Only three of the Restaurants in Cluster 4 do not have any Italian Restaurants. The most common venues in Cluster 4 consist of Trail, Zoo Exhibit, Parks, and Grocery stores.
- The two neighborhoods in the highlighted box are the recommended best neighborhoods for a new Italian Restaurant.



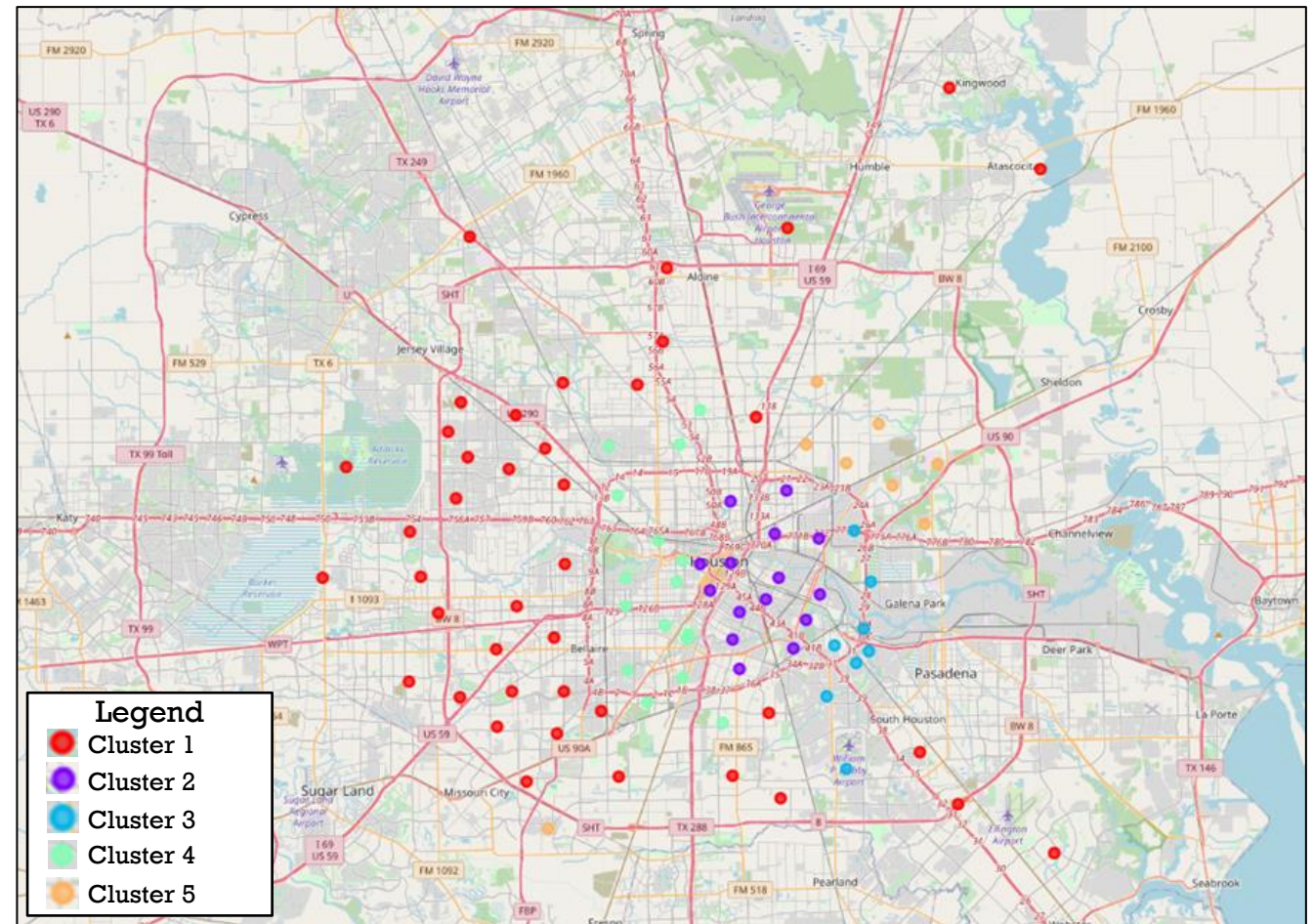
# CLUSTER 5 COMPARISON

- Cluster 5 is a small cluster and none of the Neighborhoods in cluster 5 have any Italian Restaurants. The most common venues in Cluster 4 are Discount Stores and Mexican Restaurants.



# GEOGRAPHIC CLUSTERS

- I then plotted all of the clusters on a map to display where the clusters are geographically. In the map below we can see all 5 clusters plotted on the Folium map.



# CONCLUSIONS

- In Conclusion we can see that the two Super Neighborhoods that are most likely to respond well to a new Italian Restaurant are Sunnyside and Southmain. These two neighborhoods are both within Cluster 4 and they both have a low number of other restaurants and a high number of other venues.
- The reason why I'm not recommending Northside, which is also in Cluster 4 and has no Italian Restaurants, is because of the high number of other Restaurants in the neighborhood.

