

Capstone Project - The Battle of Neighborhoods (Week 2)

1. Introduction/Business Problem

Houston, Texas is one of the biggest cities in the U.S. with a high population density and is well known for its energy industry. To capitalize on Houston's success, a major company would like to determine which neighborhood in Houston in Harris County, Texas to open up a Mexican Restaurant. The CEO, investors, and stakeholders would be interested in this project. They specifically would like it:

1. near less other competitive Mexican restaurants,
2. with smooth traffic, and
3. near a great deal of markets and residential buildings

2. Data

The data that was collected included the number of Mexican restaurants in every Houston neighborhood.

- Foursquare data was used to identify and analyze these locations
- The longitude and latitude coordinates as well as population of these neighborhoods was obtained from the U.S. Geological Survey (USGS) website
- The traffic counts data was obtained from the City of Houston GIS Open Data Portal

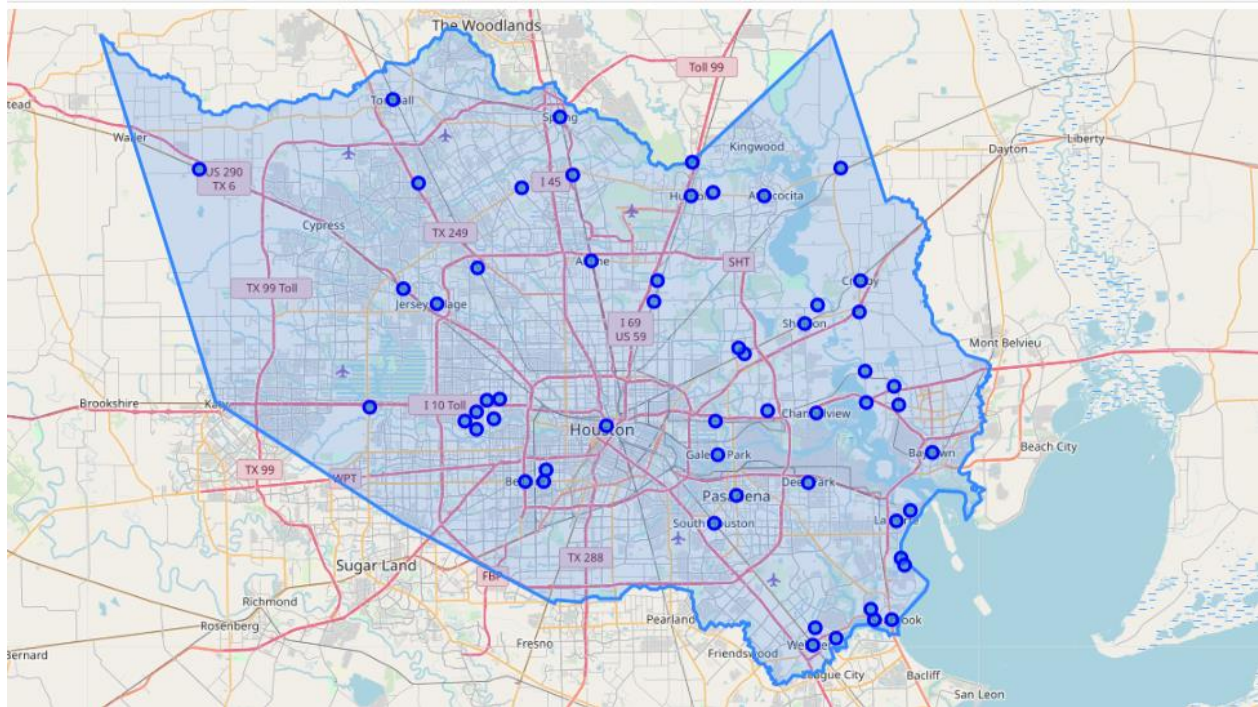
These are important in selecting the best location for a Mexican Restaurant according to all those involved in this decision.

3. Methodology

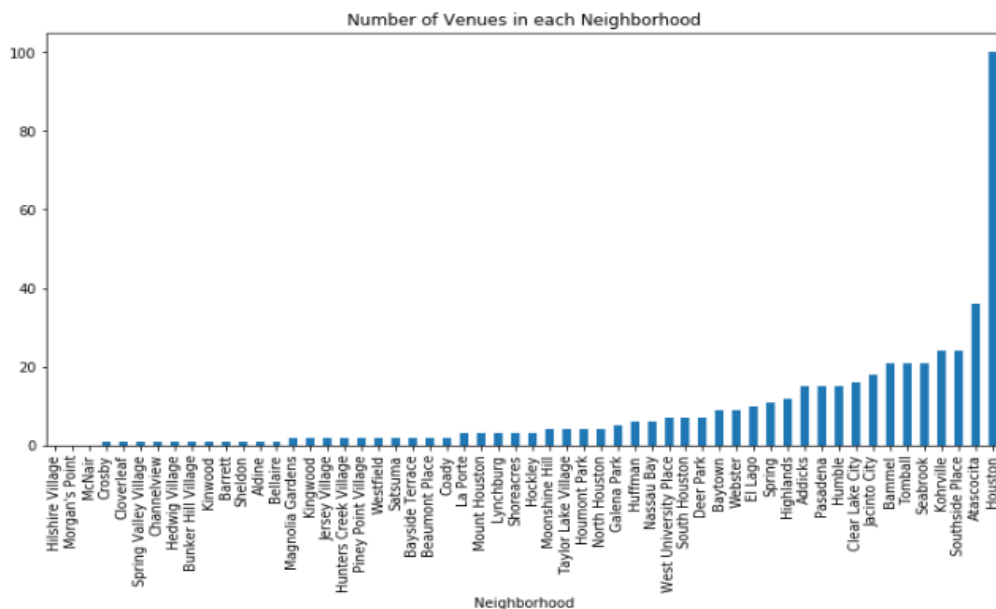
First, we obtained the neighborhoods by scraping data for only Harris County, Texas from the USGS website. We were able to load data for 56 neighborhoods in Harris County from the Feature Collection of this data uploaded.

	Neighborhood	County	State	Latitude	Longitude	Population
0	Addicks	Harris	TX	29.782451	-95.642446	-999
1	Aldine	Harris	TX	29.932445	-95.380215	15869
2	Atascocita	Harris	TX	29.998831	-95.176598	65844
3	Bammel	Harris	TX	30.006888	-95.463550	-999
4	Barrett	Harris	TX	29.879945	-95.062985	3199

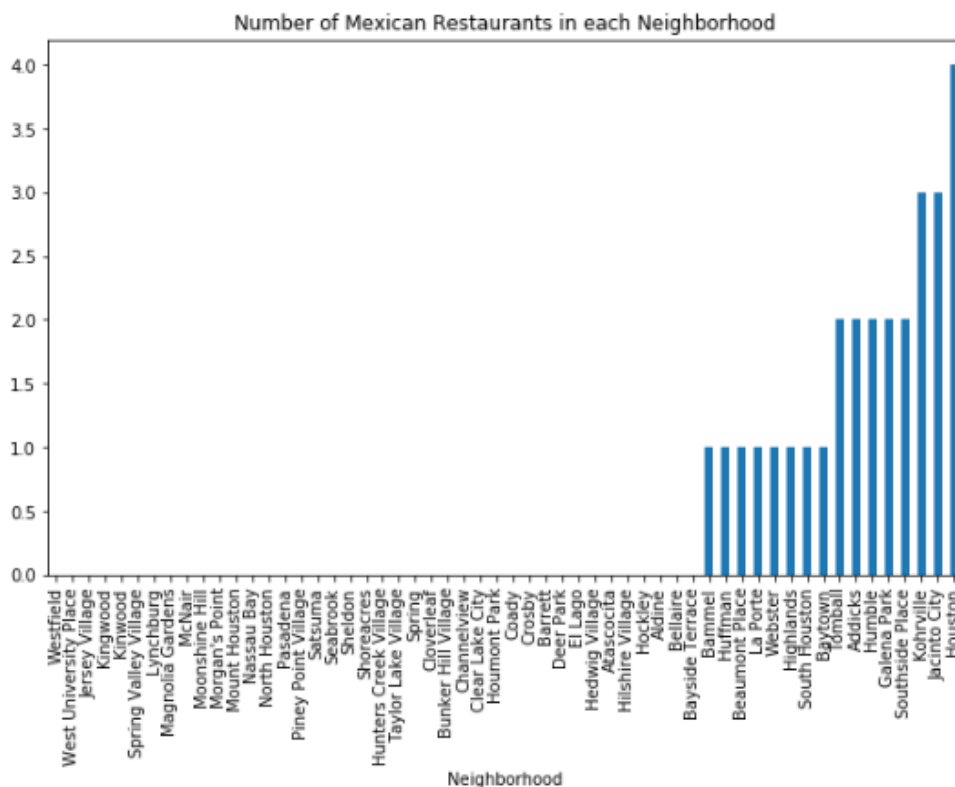
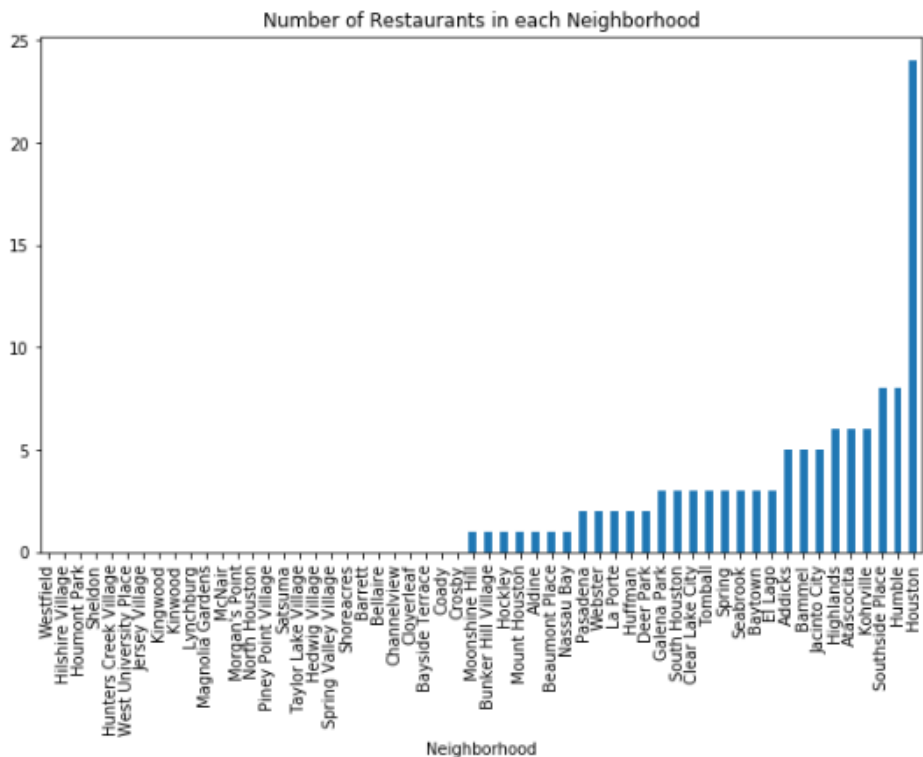
We mapped out these neighborhoods with their geographical coordinates and with Houston's coordinates as the center:



Second, we used Foursquare data to analyze the venues of each neighborhood. Though some neighborhoods do not have any venues, we were able to get venue data for 53 out of the neighborhoods. We used a bar chart to see the venue counts of each neighborhood and included those neighborhoods with no counts:

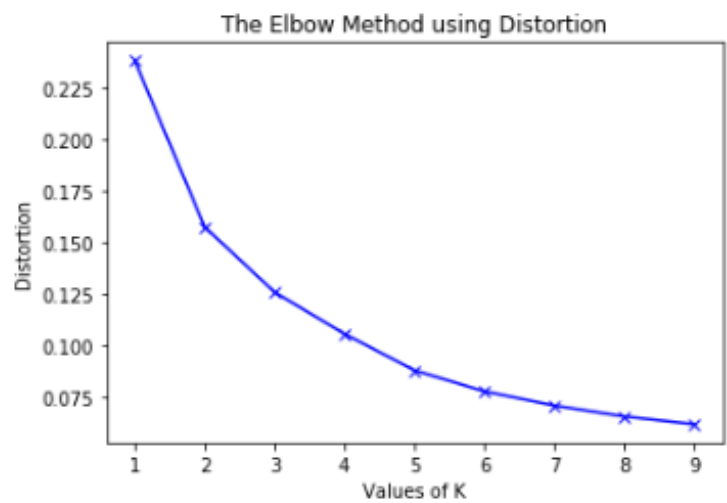


We also used a bar chart to see the restaurant and Mexican restaurant counts of each neighborhood:



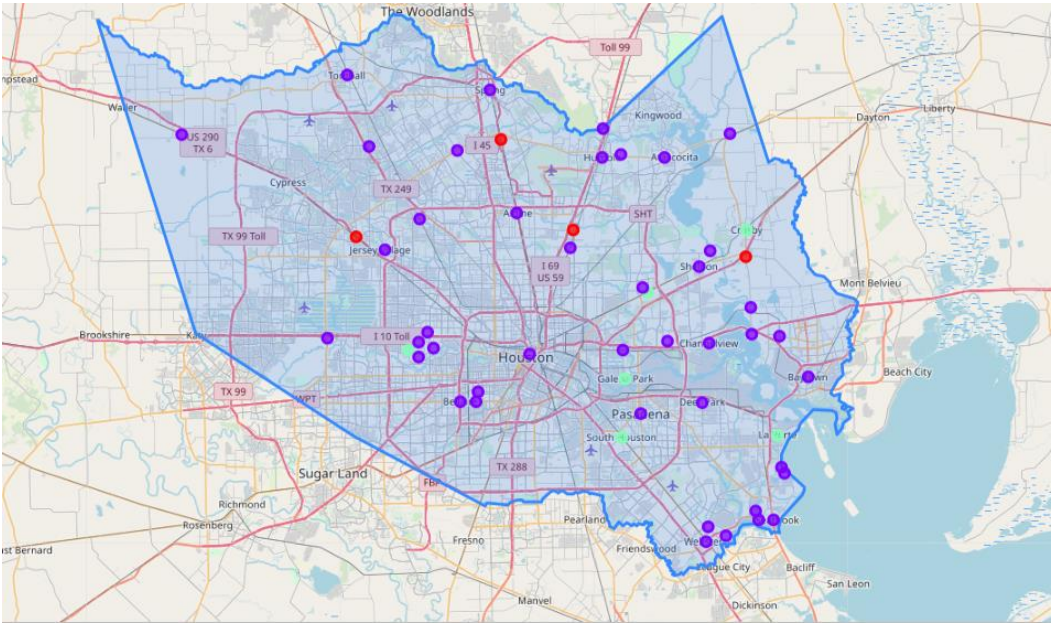
This is important in identifying those neighborhoods with few restaurants and Mexican restaurants and with many venues. Maps were also drawn to identify where these restaurants are located.

Third, we used one hot encoding to obtain the top 10 venues and subsequently cluster the neighborhoods into similar groups. We used K means clustering to group the neighborhoods into 3 clusters based on the elbow method and according to their 1st most common venue.

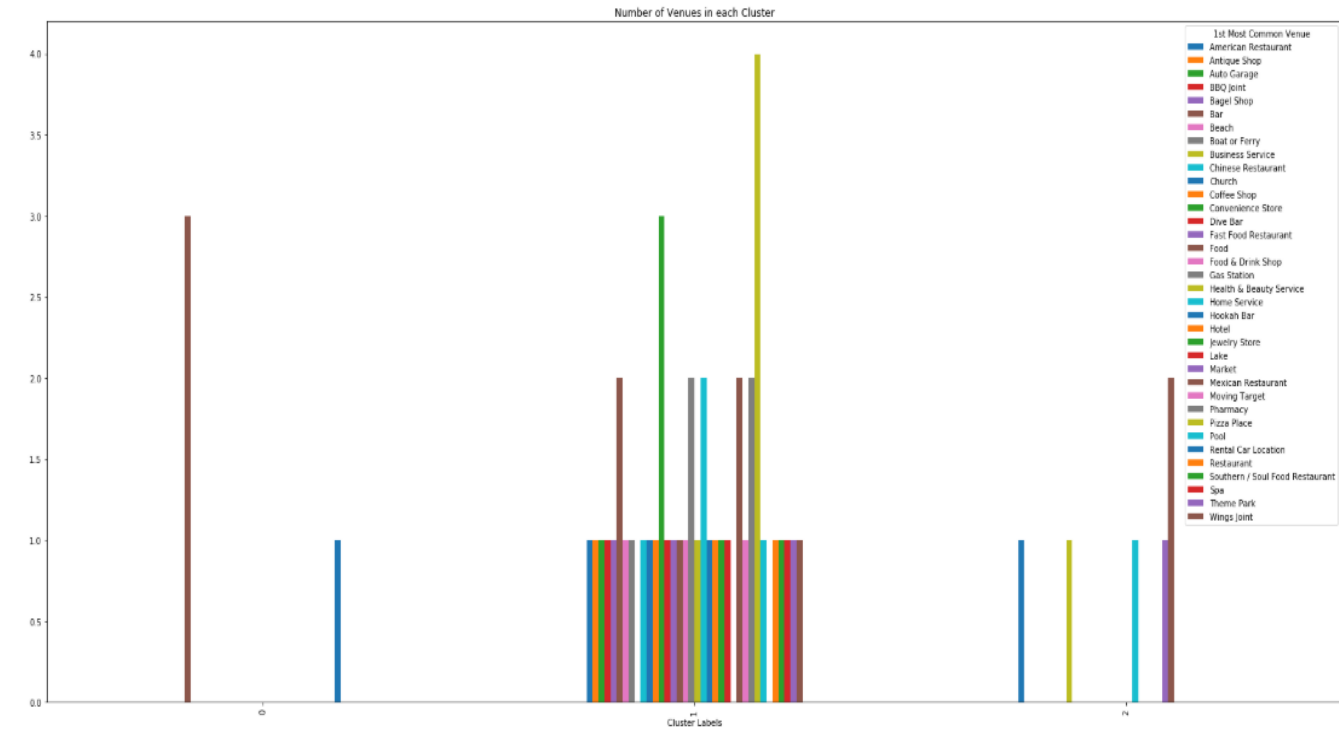


1st Most Common Venue	American Restaurant	Antique Shop	Auto Garage	BBQ Joint	Bagel Shop	Bar	Beach	Boat or Ferry	Business Service	Chinese Restaurant	Church	Coffee Shop	Convenience Store	Dive Bar	Fast Food Restaurant	Food	Food & Drink Shop	Gas Station	Health & Beauty Service	Home Service	Hookah Bar	Hotel
Cluster Labels																						
0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	0.0	1.0	1.0	1.0	3.0	1.0	1.0	1.0	1.0	2.0	1.0	2.0	1.0	1.0
2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

The following is the map of these clusters:



We then used a bar chart to label the clusters:



According to the bar chart, we can group each cluster according to its venues/markets so:

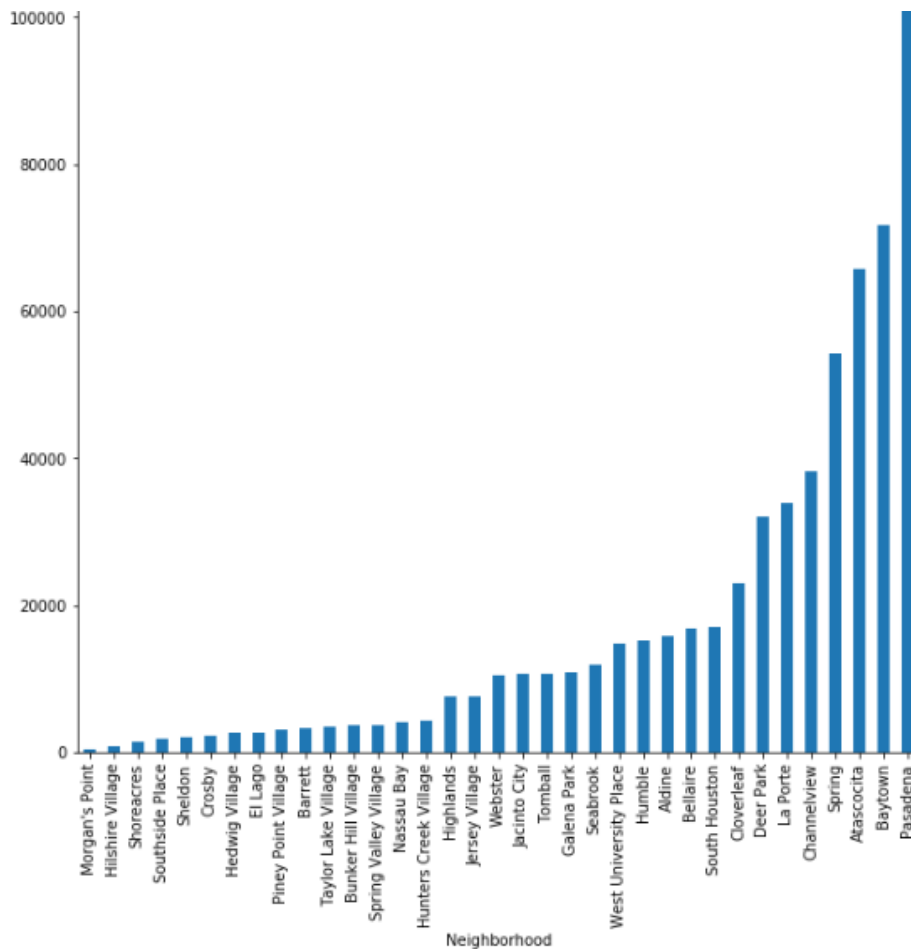
Cluster 0 - Bar and rental venues

Cluster 1 - Multiple venues

Cluster 2 - Service and food venues

Fourth, we gathered the population data of all the neighborhoods from the original data frame/data set and graphed it in a bar chart:

	Neighborhood	Population
0	Cloverleaf	22942
1	Crosby	2299
2	Aldine	15869
3	Atascocita	65844
4	Barrett	3199



Fifth, we scraped the average daily traffic (ADT) counts from a cvs file and shape file from the City of Houston GIS Open Data Portal and then graphed these counts:

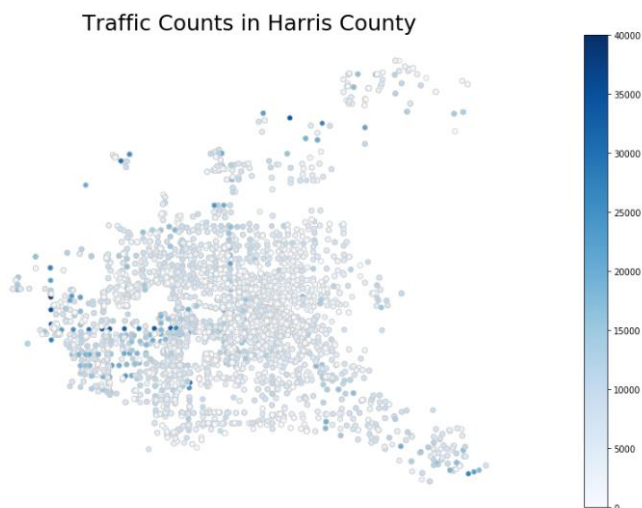
We put our cvs file into a data frame:

OBJECTID	SEGMENT	ADT	ID	STREETNAME	STREETADDR	FILENAME	PERCENTILE	OUTCOME	COLLECTION	X	Y	YEAR	PRE_NTMP_A	POST_NTMP_	PRE_NTMP_S	I
0	1	SPRR TO WEST LOOP	19402.0	129	BEECHNUT	4610	0129	None	SUCCESS	MAJOR THOROUGHFARE ADT	3093465.52	13815159.57	0	None	None	None
1	2	SYNOTT TO W CITY LIMIT	34826.0	158	BELLAIRE	13100	0158	None	SUCCESS	MAJOR THOROUGHFARE ADT	3042291.43	13818867.17	0	None	None	None
2	3	CHIMNEY ROCK TO HILLCROFT	25154.0	176	BELLFORT	5614	0176	None	SUCCESS	MAJOR THOROUGHFARE ADT	3084670.97	13803301.70	0	None	None	None
3	4	FONDREN TO S GESSNER	21667.0	178	BELLFORT	8000	0178	None	SUCCESS	MAJOR THOROUGHFARE ADT	3074100.26	13802378.25	0	None	None	None
4	5	WEST BELT TO SW FWY	18069.0	180	BELLFORT	10200	0180	None	SUCCESS	MAJOR THOROUGHFARE ADT	3061213.60	13800911.51	0	None	None	None

We put our shape file into a data frame as well and merged the two data frames:

OBJECTID	STREETNAME	STREETADDRESS	geometry	ADT
0	1	BEECHNUT	4610 POINT (-95.45627 29.68915)	19402.0
1	2	BELLAIRE	13100 POINT (-95.61704 29.70350)	34826.0
2	3	BELLFORT	5614 POINT (-95.48507 29.65728)	25154.0
3	4	BELLFORT	8000 POINT (-95.51843 29.65561)	21667.0
4	5	BELLFORT	10200 POINT (-95.55912 29.65263)	18069.0

We then graphed these traffic counts:



4. Results

Our analysis looked at the neighborhoods in Harris County, Texas and their venues in relation to restaurants and Mexican restaurants as well as their population and traffic counts data. We were looking at those neighborhoods with none or few Mexican restaurants and none or few restaurants in general. We were also looking at those neighborhoods with high population/residents and smooth/low traffic. Three neighborhoods stood out with these characteristics; namely West University Place, Houmont Park, and Taylor Lake Village. West University Place has a higher population and venue count than the other two neighborhoods and is close to Houston, the center of Harris County. Houmont Park is northwest of the county and has a slightly higher population than Taylor Lake Village, but these two neighborhoods have a similar venue count. Taylor Lake Village is southwest of the county and so has more access to ports.

5. Discussion

Based on these results, I would recommend the West University Place neighborhood. This neighborhood has the benefit of being close to the county's center and has many venues including a pizza place, café, and park which are great tourist/residential venues. It also has a high population and thus will be a good customer base for a Mexican restaurant. The data from the Foursquare showed that it has no Mexican restaurants and no restaurant in general.

6. Conclusion

The purpose of this project was to identify an optimal location in Harris County, Texas to open up a Mexican restaurant. Though there were limitations with the data in that not all the neighborhoods had venues and population counts, we were still able to identify some neighborhoods which can answer our business problem: namely West University Place, Houmont Park, and Taylor Lake Village with West University Place as the seemingly optimal location. We will present these findings to the those involved to make final decision on the optimal location of a Mexican restaurant.