

1 Introduction/Business problem

1.1 Problem

The restaurant business is highly competitive. To be successful in this area, one needs a carefully thought business plan. The major aspect of the business plan is to find a convenient place to host the business. I will be focusing on the Houston metro area. The goal is to find the 5 best neighborhoods in the Houston metro area for opening a new restaurant.

1.2 Audience

This study and results could be very helpful to anyone seeking to open a business in the Houston metro area. I am thinking here about individual or family own restaurant. Also corporations looking to open new restaurants in this part of the country could also benefit from this work. More generally, anyone looking to open a restaurant anywhere in the country can use the method developed here and just feed in the corresponding data.

2. Data description

To complete this work, we will be using 2 datasets: The Foursquare dataset and census data.

2.1 Foursquare data

Foursquare is a company that provides information about venues in many locations. Foursquare will help us find restaurants in Houston neighborhoods. Their online website is at www.foursquare.com. Once a venue is located, it is also possible to look deeper into the venue to have more information about it.

2.2 Statistical Atlas

To find Houston county and Neighborhood, I use the statisticalatlas web page at www.statisticalatlas.com. By selecting the state (Texas) and the Metro area (Houston metro), I was able to get all the county and Neighborhood in the metro area. I then feed the information in the census data tool to get the statistics about the city or neighborhood.

2.3 census data

The US Government conducts census every 10 years to gather information about US population. The US government then makes these data public so that anyone interested can use it. I use a specific tool called QuickFacts from the census Bureau to get some statistical information about Houston neighborhoods. All the data I got was compiled on a csv file. You can see the first lines of the file below.

```
Out[3]:
```

```
(64, 64)
```

3 Methodology

3.1 Exploratory Data Analysis

3.1.1 Missing values

What are missing values in this dataset

It appears from the output above that each column has 64 elements with no missing value. But in this dataset, missing values have been replaced by letters for referencing and explaining why there are no values. The letters are

"-" Either no or too few sample observations were available to compute an estimate

D Suppressed to avoid disclosure of confidential information

D Suppressed to avoid disclosure of confidential information
F Fewer than 25 firms
FN Footnote on this item in place of data
N Data for this geographic area cannot be displayed because the number of sample cases is too small.
NA Not available
S Suppressed; does not meet publication standards
X Not applicable
Z Value greater than zero but less than half unit of measure shown

Missing values strategy

All column with more that 90% of missing values will be deleted. The rest of missing value will be set as corresponding column average

looks like MontBelvieu and MissouriCity are duplicate. let remove one of each so as to have just unique city names

Out[5]:

```
Int64Index([15, 25], dtype='int64')
```

Out[6]:

```
Int64Index([16, 26], dtype='int64')
```

Out[7]:

```
(62, 64)
```

Out[8]:

```
(62, 64)
```

```
PopulationEstimates2019
PopulationEstimates2010
PopulationPercentChange2010TO2019
HousingUnitsJuly2019
BuildingPermits, 2019
TotalAccommodationAndFoodServicesSales2012T
TotalHealthCareAndSocialAssistanceReceiptsRevenue2012T
TotalManufacturersShipments2012T
TotalMerchantWholesalerSales2012T
TotalRetailSales2012T
TotalEmployerEstablishments2017
TotalEmployment2017
TotalAnnualPayroll2017T
TotalEmploymentPercentChange2016-2017
TotalnonemployerEstablishments2018
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:3:
FutureWarning: elementwise comparison failed; returning scalar instead, but in the future will per
form elementwise comparison
This is separate from the ipykernel package so we can avoid doing imports until
```

Out[14]:

	PopulationEstimates2019	PopulationEstimates2010	PopulationPercentChange2010TO2019	HousingUnitsJuly2019	BuildingPermits,201
0	18,971	16,907	0.12	X	
1	77,192	71,605	0.08	X	
2	19,431	18,829	0.03	X	
3	26,723	24,161	0.11	X	
4	22,222	22,222	0.00	X	

4	8389	7364	0.14	X	
...	PopulationEstimates2019	PopulationEstimates2010	PopulationPercentChange2010TO2019	HousingUnitsJuly2019	BuildingPermits,201
...
57	33474	32010	0.05	X	
58	7914	6967	0.14	X	
59	8355	5768	0.45	X	
60	57602	54635	0.05	25433	26
61	50446	47741	0.06	X	

62 rows × 15 columns



Out [19] :

2

Out [20] :

3

Out [21] :

5

Out [22] :

4

Out [23] :

Int64Index([25, 37, 38, 39, 40, 41, 43, 44, 45, 47, 48, 49, 50, 51], dtype='int64')

Out [24] :

Int64Index([25, 37, 38, 39, 40, 41, 43, 44, 45, 47, 48, 49, 50, 51], dtype='int64')

Replacing "X" in columns 'PopulationEstimates2019' and 'PopulationEstimates2010' by the corresponding row value in column 'PopulationCensus2010'

Out [26] :

	PopulationEstimates2019	PopulationEstimates2010
25	54298	54298
37	38289	38289
38	8619	8619
39	65844	65844
40	15869	15869
41	93847	93847
43	7522	7522
44	19069	19069
45	12382	12382
47	22942	22942
48	18274	18274
49	13721	13721
50	15963	15963
51	36501	36501

Replacing "X" in columns 'PopulationPercentChange2010TO2019' by 0. this assume there was no population change for the city where we have x

Out[27]:

```
Int64Index([25, 37, 38, 39, 40, 41, 43, 44, 45, 47, 48, 49, 50, 51], dtype='int64')
```

Out[29]:

```
25    0
37    0
38    0
39    0
40    0
41    0
43    0
44    0
45    0
47    0
48    0
49    0
50    0
51    0
```

```
Name: PopulationPercentChange2010TO2019, dtype: object
```

converting to float and replacing all string as NaN values

some of the columns values has to be multiplied by 1000 to reflect the correct values as in the census data they were entered in tausends

some of the columns have a lot of Null and may have to be deleted

```
'TotalAccommodationAndFoodServicesSales2012T', 'TotalHealthCareAndSocialAssistanceReceiptsRevenue2012T',  
'TotalManufacturersShipments2012T', 'TotalMerchantWholesalerSales2012T', 'TotalRetailSalesPerCapita2012'
```

replacing the remaining null value in each column by the average value of the corresponding column

Geolocalization of the city on map

Importing the libraries necessary for plotting and clustering

```
Collecting package metadata (current_repodata.json): done  
Solving environment: done
```

```
# All requested packages already installed.
```

```
Collecting package metadata (current_repodata.json): done  
Solving environment: done
```

```
# All requested packages already installed.
```

```
Libraries imported.
```

Out[46]:

	County	City	latitude	longitude
18	Brazoria	Lake Jackson	29.033857	-95.434386

Out[47]:

	County	City	latitude	longitude
50	Fort Bend	Pecan Grove	29.626068	-95.731616

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/pandas/core/frame.py:3997:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
errors=errors,
```

Out[49]:

County	City	latitude	longitude
--------	------	----------	-----------

Out[50]:

County	City	latitude	longitude
--------	------	----------	-----------

Out[52]:

```
0      False
1      False
2      False
3      False
4      False
...
57     False
58     False
59     False
60     False
61     False
Name: location, Length: 62, dtype: bool
```

create a map of Houston city

Out[53]:

Make this Notebook Trusted to load map: File -> Trust Notebook

3.2 Machine Learning

3.2.1 Using Foursquare to find the venues in Houston Neighborhood

Define Foursquare Credential and Version

The function define below will be used to get the venues in Houston Neighborhood

Getting the nearby venues for Houston Neighborhoods

```
Bellaire
Baytown
Angleton
Alvin
Dayton
Conroe
Clute
Cleveland
Brookshire
LaPorte
LaMarque
Katy
Jersey Village
Jacinto City
Humble
Manvel
Liberty
League City
Lake Jackson
Richmond
Prairie View
Pearland
Pasadena
Mont Belvieu
Missouri City
Spring
South Houston
Sealy
Seabrook
Santa Fe
Rosenberg
West University Place
Webster
Tomball
Texas City
Sugar Land
Stafford city
Channelview
Bacliff
Atascocita
Aldine
The Woodlands
Willis
Highlands
Fresno
Four Corners
Crosby County
Cloverleaf
Ginger Branch
```

CINCO RANCH
Sienna Plantation
Pecan Grove
Mission Bend
Galena Park
Fulshear
Friendswood
Freeport
Dickinson
Deer Park
Hitchcock
Hempstead
Hardin County
Galveston

(4801, 7)

how many venues where return for each neighborhood

Out[58]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Aldine	100	100	100	100	100	100
Alvin	79	79	79	79	79	79
Angleton	40	40	40	40	40	40
Atascocita	86	86	86	86	86	86
Bacliff	79	79	79	79	79	79
...
The Woodlands	100	100	100	100	100	100
Tomball	100	100	100	100	100	100
Webster	100	100	100	100	100	100
West University Place	100	100	100	100	100	100
Willis	60	60	60	60	60	60

62 rows × 6 columns

let extract just the restaurant from the venues

Out[61]:

```
array(['Middle Eastern Restaurant', 'Grocery Store',  
      'Fast Food Restaurant', 'Gym', 'Gymnastics Gym', 'Ice Cream Shop',  
      'Cuban Restaurant', 'Coffee Shop', 'Convenience Store', 'Café',  
      'Furniture / Home Store', 'Trail', 'Women's Store', 'Bakery',  
      'Park', 'Shopping Mall', 'Donut Shop', 'Health & Beauty Service',  
      'Mediterranean Restaurant', 'Department Store', 'Eye Doctor',  
      'Clothing Store', 'Steakhouse', 'Cosmetics Shop',  
      'Japanese Restaurant', 'Burger Joint', 'French Restaurant',  
      'Movie Theater', 'Gourmet Shop', 'Hotel', 'Churrascaria',  
      'Electronics Store', 'Warehouse Store', 'Seafood Restaurant',  
      'Sushi Restaurant', 'Taco Place', 'Food Truck', 'Hotel Bar',  
      'New American Restaurant', 'Wine Shop', 'American Restaurant',  
      'Deli / Bodega', 'Restaurant', 'Sandwich Place',  
      'Chinese Restaurant', 'Frozen Yogurt Shop', 'Smoothie Shop',  
      'Juice Bar', 'Dessert Shop', 'Gym / Fitness Center',  
      'Cupcake Shop', 'Gift Shop', 'Farmers Market',  
      'Italian Restaurant', 'Bagel Shop', 'Big Box Store', 'Wine Bar',  
      'Comfort Food Restaurant', 'Thai Restaurant', 'Chocolate Shop',  
      'Kitchen Supply Store', 'Creperie', 'Smoke Shop',  
      'Accessories Store', 'Breakfast Spot', 'Mexican Restaurant', 'Pub',  
      'Fabric Shop', 'Sports Bar', 'BBQ Joint', 'Supermarket',  
      'Discount Store', 'Pharmacy', 'Skate Park', 'Gas Station',
```

```
'Laundry Service', 'Hardware Store', 'Pizza Place',
'Sporting Goods Shop', 'Video Store', 'Bar', 'Supplement Shop',
'Outdoors & Recreation', 'Wings Joint', 'Fried Chicken Joint',
'Pet Store', 'Arts & Crafts Store', 'Rental Car Location', 'Bank',
'Mobile Phone Shop', 'Storage Facility', 'Liquor Store',
'General Entertainment', 'Sculpture Garden', 'Diner',
'Tex-Mex Restaurant', 'Airport', 'Airport Terminal', 'Nightclub',
'Plane', 'Antique Shop', 'Farm', 'Market', 'Zoo',
'Salon / Barbershop', 'Thrift / Vintage Store',
'Paper / Office Supplies Store', 'Fireworks Store',
'Garden Center', 'Bed & Breakfast', 'Construction & Landscaping',
'Spa', 'Carpet Store', 'Post Office', 'Theater', 'Brewery',
'Shoe Store', 'Golf Course', 'Video Game Store', 'Buffet',
'Sports Club', 'Southern / Soul Food Restaurant',
'Asian Restaurant', 'German Restaurant', 'Karaoke Bar',
'Lingerie Store', 'Latin American Restaurant', 'Insurance Office',
'Automotive Shop', 'Aquarium', 'Truck Stop', 'Intersection',
'Museum', 'Athletics & Sports', 'Other Repair Shop', 'Lake',
'Beach', 'High School', 'Baseball Field', 'Burrito Place',
'Bistro', 'Surf Spot', 'Motel', 'Other Great Outdoors',
'Bowling Alley', 'Racetrack', 'Pool', 'Music Store', 'Beer Bar',
'Water Park', 'Arcade', 'Football Stadium', 'Irish Pub',
'Bookstore', 'Candy Store', 'Greek Restaurant', 'Playground',
'Vietnamese Restaurant', 'Salad Place', 'Food Service',
'Indian Restaurant', 'South American Restaurant',
'Eastern European Restaurant', 'Motorcycle Shop',
'Cajun / Creole Restaurant', 'Outdoor Supply Store',
'Volleyball Court', 'Snack Place', 'Music Venue', 'Rental Service',
'Beer Garden', 'Cocktail Bar', 'Church', 'Airport Lounge',
'Multiplex', 'Korean Restaurant', 'Flower Shop', 'Moving Target',
'Food', 'Butcher', 'Business Service', 'Martial Arts Dojo',
'Golf Driving Range', 'Polish Restaurant', 'Chiropractor',
'Gastropub', 'Ramen Restaurant', 'Government Building',
'Boxing Gym', 'Science Museum', 'Kids Store', 'Jewelry Store',
'Brazilian Restaurant', 'Rock Club', 'Gun Range', 'Massage Studio',
'Hobby Shop', 'Student Center', 'Paintball Field', 'Campground',
'Dog Run', 'Caribbean Restaurant', 'Shop & Service',
'Spanish Restaurant', 'Airport Service', 'Bubble Tea Shop',
'Pakistani Restaurant', 'Lounge', 'Shipping Store', 'Nail Salon',
'Hot Dog Joint', 'Camera Store', 'Resort', 'Mini Golf',
'Harbor / Marina', 'Event Service', 'History Museum',
'State / Provincial Park', 'Design Studio', 'Other Nightlife',
'Theme Park', 'Tiki Bar', 'Lighthouse', 'Fish Market',
'Neighborhood', 'Nature Preserve', 'Monument / Landmark',
'Outlet Mall', 'Toy / Game Store', 'Boutique', 'Exhibit',
'Outlet Store', 'Record Shop', 'Dive Bar', 'Zoo Exhibit',
'Art Museum', 'Historic Site', 'Tea Room', 'Fishing Store',
'Baseball Stadium', 'Shopping Plaza', 'Noodle House',
'Boat or Ferry', 'Flea Market', 'Pier', 'Plaza', 'Soccer Field',
'Home Service', 'Concert Hall', 'Elementary School', 'River',
'Mongolian Restaurant', 'Theme Park Ride / Attraction', 'Gym Pool',
'Pie Shop', 'Bike Shop', 'Recreation Center', 'Hookah Bar',
'Kebab Restaurant', 'Speakeasy', 'Stadium', 'Social Club',
'General Travel', 'Opera House', 'Board Shop', 'Library',
'Scenic Lookout', 'Gay Bar', 'Hotel Pool'], dtype=object)
```

Out[62]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bellaire	29.705786	-95.458830	Fadi's Meyerland Mediterranean Grill	29.689590	-95.460313	Middle Eastern Restaurant
2	Bellaire	29.705786	-95.458830	Chick-fil-A	29.688802	-95.465120	Fast Food Restaurant
6	Bellaire	29.705786	-95.458830	Cafe Piquet	29.700276	-95.481887	Cuban Restaurant
18	Bellaire	29.705786	-95.458830	Mary'z Lebanese Cuisine	29.731227	-95.482201	Mediterranean Restaurant
26	Bellaire	29.705786	-95.458830	Roka Akor	29.738590	-95.441658	Japanese Restaurant
...
4779	Galveston	29.299328	-94.794588	Porch Cafe	29.319570	-94.750232	Restaurant
4780	Galveston	29.299328	-94.794588	Cajun Greek - Seafood	29.276625	-94.831449	Seafood Restaurant

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4787	Galveston	29.299328	-94.794588	Benno's	29.281570	-94.801496	Seafood Restaurant
4788	Galveston	29.299328	-94.794588	Joe's Crab Shack	29.281570	-94.801496	Seafood Restaurant
4789	Galveston	29.299328	-94.794588	Salsas Mexican & Seafood Restaurant	29.275570	-94.811700	Mexican Restaurant

1236 rows × 7 columns

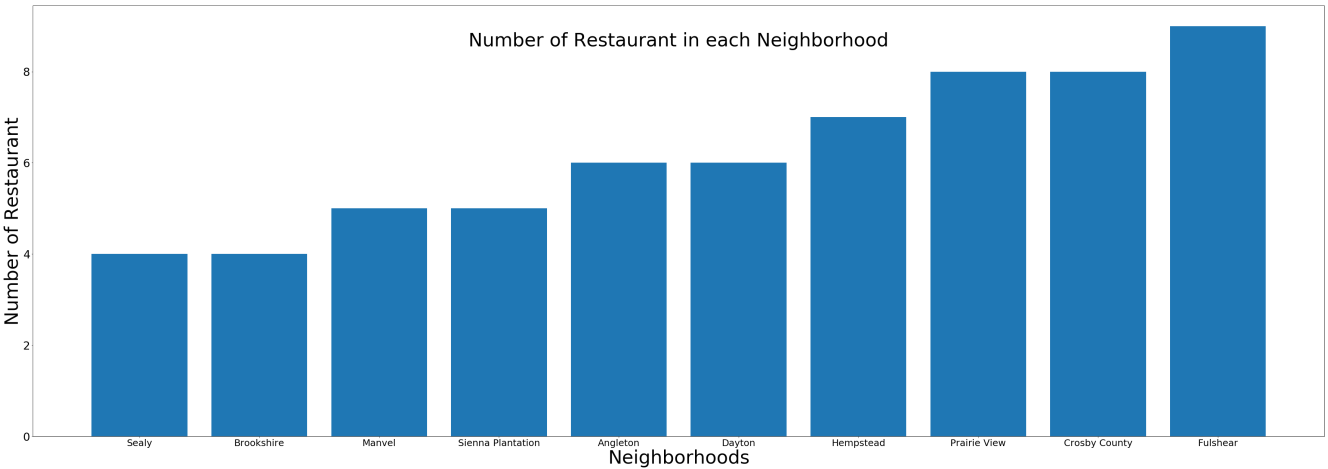
Out [63]:
(1236, 7)

Out [65]:

	Venue Latitude	Venue Longitude
Neighborhood		
Aldine	29.898604	-95.370529
Alvin	29.422301	-95.236024
Angleton	29.179284	-95.429897
Atascocita	30.012310	-95.167041
Bacliff	29.529418	-95.008115
...
The Woodlands	30.174297	-95.473803
Tomball	30.107958	-95.596545
Webster	29.542335	-95.127965
West University Place	29.733029	-95.429870
Willis	30.404391	-95.488813

61 rows × 2 columns

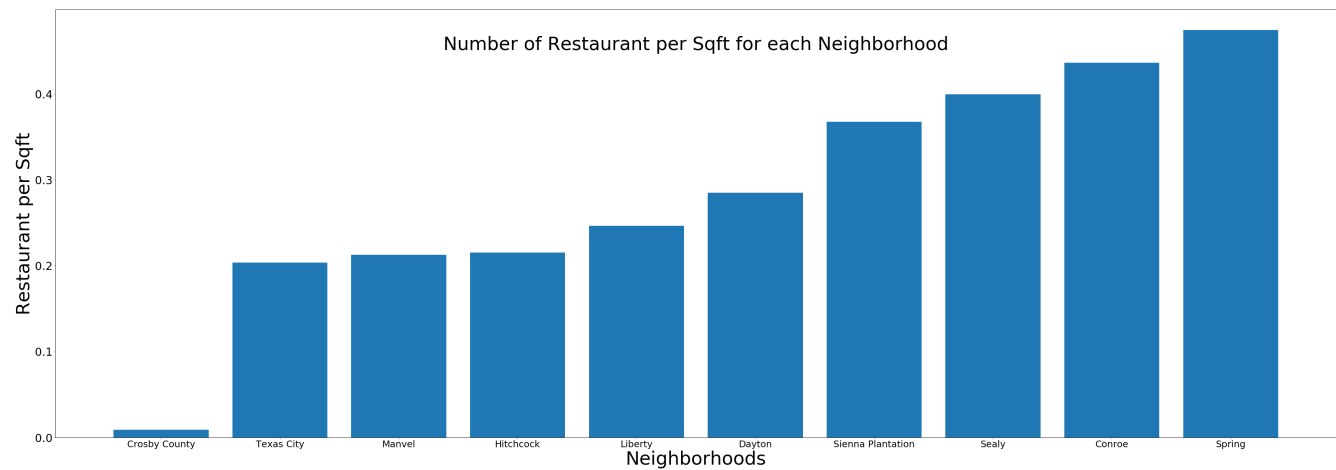
Number of Restaurants in each neighborhood



The 6 neighborhoods with the least number of restaurants are: Sealy, Brookshire, Manvel, Sienna Plantation, Angleton and Dayton

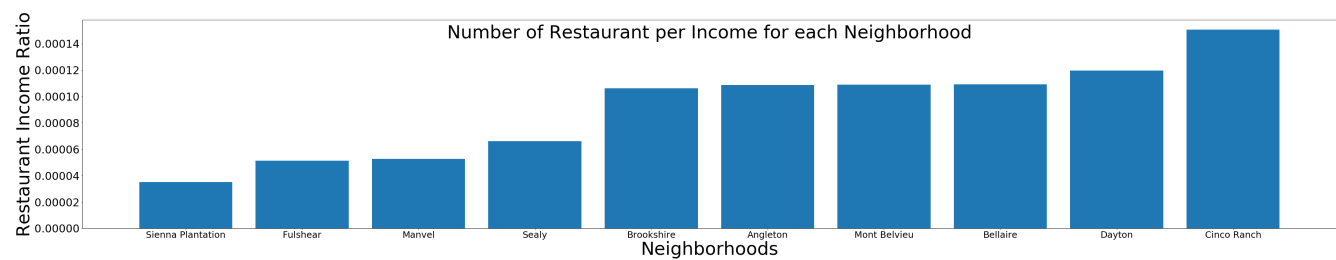
let add some more columns to our data

Number of restaurants per neighborhood sqft



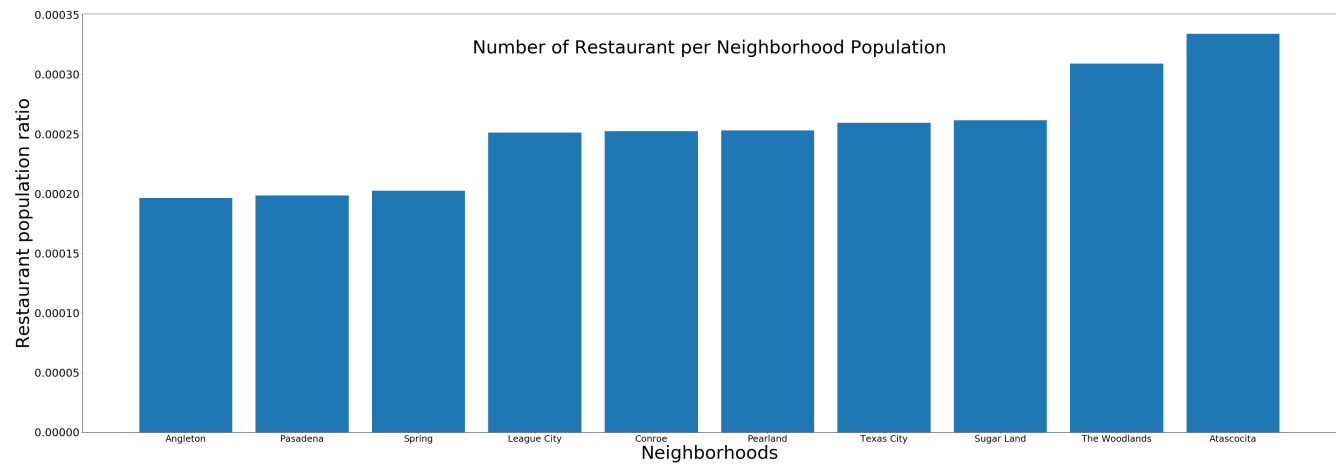
Crosby, Texas City, Manvel, Hitchcock, Liberty and Dayton are the 5 Neighborhoods with less restaurant per sqft

Number of restaurants per family income in each neighborhood



Sienna Plantation, Fulshear, Manvel, Sealy, Brookshire and Angleton are the neighborhoods with less restaurants per family income

Number of restaurants per neighborhood population



Angleton, Pasadena, Spring, Conroe, League City and portland are the neighborhood with less restaurants per neighborhood population

3.2.2 Clustering Houston Neighborhood and identify their main characteristics

There are 269 unique categories.

Analyze each neighborhood

There are 4801 venues in Houston Neighborhood

Out[84]:
(4801, 269)

Let's print each neighborhood along with the top 5 most common venues

Dataframe with top 10 venues for 5 neighborhoods

Out[89]:

	Neighborhood	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	Aldine	Mexican Restaurant	Hotel	Burger Joint	Pharmacy	Sandwich Place	Seafood Restaurant	Fried Chicken Joint	Rental Car Location	Park	Discount Store
1	Alvin	Discount Store	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Donut Shop	Gas Station	Fried Chicken Joint	Seafood Restaurant	American Restaurant	Pizza Place
2	Angleton	Fast Food Restaurant	Discount Store	Pharmacy	Rental Car Location	Pizza Place	Airport	Taco Place	Fried Chicken Joint	Tex-Mex Restaurant	Donut Shop
3	Atascocita	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Fried Chicken Joint	Gym	Discount Store	Italian Restaurant	Golf Course	Gas Station	Park
4	Bacliff	Seafood Restaurant	Bar	Mexican Restaurant	Pizza Place	Café	Playground	Fried Chicken Joint	Burger Joint	Greek Restaurant	Theme Park

Cluster Neighborhoods

Run k-means to cluster the neighborhood into 5 clusters.

Out[90]:
array([0, 4, 4, 1, 1, 1, 1, 2, 0, 1], dtype=int32)

Create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Out[91]:

	County	Neighborhood	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
52	Harris	Galena Park	29.733562	-95.230212	0	Mexican Restaurant	Pharmacy	Seafood Restaurant	BBQ Joint	Wings Joint	Burger Joint	Re
53	Fort Bend	Fulshear	29.690292	-95.899626	4	Fast Food Restaurant	Grocery Store	Pet Store	American Restaurant	BBQ Joint	Sandwich Place	Re

	Neighborhood	Neighborhood	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
54	Galveston	Freeport	28.954137	-95.359662	4	Pizza Place	Discount Store	Seafood Restaurant	Fast Food Restaurant	American Restaurant	Beach	Re
55	Brazoria	Freeport	28.954137	-95.359662	4	Pizza Place	Discount Store	Seafood Restaurant	Fast Food Restaurant	American Restaurant	Beach	Re
56	Galveston	Dickinson	29.460788	-95.051317	1	Mexican Restaurant	Clothing Store	Burger Joint	Coffee Shop	Pizza Place	Sandwich Place	A Re
57	Harris	Deer Park	29.705228	-95.123820	1	Mexican Restaurant	Coffee Shop	Burger Joint	Fast Food Restaurant	Seafood Restaurant	Ice Cream Shop	Co
58	Galveston	Hitchcock	29.348291	-95.016037	4	Sandwich Place	Discount Store	Fast Food Restaurant	Mexican Restaurant	Burger Joint	Pizza Place	Conv
59	Waller	Hempstead	30.092949	-96.080655	2	Fast Food Restaurant	Seafood Restaurant	Ice Cream Shop	Big Box Store	Farmers Market	Farm	S
60	Liberty	Hardin County	30.152435	-94.739089	3	Athletics & Sports	Discount Store	Pet Store	Zoo	Flower Shop	Fireworks Store	Fis
61	Galveston	Galveston	29.299328	-94.794588	1	Seafood Restaurant	Bar	Beach	Coffee Shop	Ice Cream Shop	Burger Joint	

Finally, let's visualize the resulting clusters

Out[92]:

Make this Notebook Trusted to load map: File -> Trust Notebook

3.2.3 Examine Clusters

We can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster. I will leave this exercise to you.

Cluster 1

Out[127]:

	Neighborhood	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
13	Jacinto City	Mexican Restaurant	Discount Store	Seafood Restaurant	Sandwich Place	Pharmacy	Chinese Restaurant	Fast Food Restaurant	Pizza Place	BBQ Joint	Burger Joint
14	Galveston	Mexican Restaurant	Burger Joint	Fast Food Restaurant	Discount Store	Cosmetics	Discount Store	Sporting Goods	Pizza Place	BBQ Joint	Coffee Shop

22	Pasadena	Mexican Restaurant	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
38	Bacliff	Seafood Restaurant		Bar	Mexican Restaurant	Pizza Place	Café	Playground	Chicken Joint	Burger Joint	BBQ Joint	Theme Park
40	Aldine	Mexican Restaurant		Hotel	Burger Joint	Pharmacy	Sandwich Place	Seafood Restaurant	Fried Chicken Joint	Rental Car Location	Park	Discount Store
47	Cloverleaf	Mexican Restaurant		Discount Store	Pharmacy	Seafood Restaurant	Sandwich Place	Burger Joint	BBQ Joint	Pizza Place	Chinese Restaurant	Ice Cream Shop
52	Galena Park	Mexican Restaurant		Pharmacy	Seafood Restaurant	BBQ Joint	Wings Joint	Burger Joint	Chinese Restaurant	Diner	Park	Sandwich Place
61	Galveston	Seafood Restaurant		Bar	Beach	Coffee Shop	Ice Cream Shop	Burger Joint	Museum	Sandwich Place	Hotel	Mexican Restaurant

Cluster 2

Out [128]:

	Neighborhood	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
60	Hardin County	Athletics & Sports	Discount Store	Pet Store	Zoo	Flower Shop	Fireworks Store	Fish Market	Fishing Store	Flea Market	Food

Cluster 3

Out [129]:

	Neighborhood	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
0	Bellaire	Burger Joint	Seafood Restaurant	Ice Cream Shop	Steakhouse	Café	Shopping Mall	Department Store	Dessert Shop	American Restaurant
5	Conroe	Mexican Restaurant	Fast Food Restaurant	Burger Joint	Breakfast Spot	Steakhouse	Park	Cosmetics Shop	Fried Chicken Joint	Sandwich Place
11	Katy	Mexican Restaurant	Grocery Store	Sporting Goods Shop	Fast Food Restaurant	American Restaurant	Burger Joint	Clothing Store	Coffee Shop	Hotel
12	Jersey Village	Burger Joint	Sandwich Place	Coffee Shop	American Restaurant	Asian Restaurant	Fast Food Restaurant	Mexican Restaurant	Hotel	Seafood Restaurant
14	Humble	Mexican Restaurant	Coffee Shop	Fast Food Restaurant	Seafood Restaurant	Gym	Burger Joint	Airport Lounge	Furniture / Home Store	American Restaurant
17	League City	Mexican Restaurant	Park	Coffee Shop	American Restaurant	Gym	Science Museum	Bakery	Fast Food Restaurant	Grocery Store
19	Richmond	Mexican Restaurant	Fast Food Restaurant	Coffee Shop	Burger Joint	Pharmacy	Discount Store	Café	BBQ Joint	Pizza Place
21	Pearland	Mexican Restaurant	Coffee Shop	Fast Food Restaurant	Grocery Store	Donut Shop	Burger Joint	Asian Restaurant	Park	Pizza Place
24	Missouri City	Mexican Restaurant	Park	Pharmacy	Fast Food Restaurant	Cajun / Creole Restaurant	Fried Chicken Joint	Grocery Store	Chinese Restaurant	Ice Cream Shop
26	South Houston	Burger Joint	Mexican Restaurant	Pizza Place	Park	Fast Food Restaurant	Vietnamese Restaurant	Pharmacy	Diner	Sporting Goods Shop
28	Seabrook	Mexican Restaurant	Seafood Restaurant	Burger Joint	Science Museum	Coffee Shop	Park	Bar	Hotel	Pizza Place
30	Rosenberg	Mexican Restaurant	Fast Food Restaurant	Burger Joint	Coffee Shop	Gym / Fitness Center	Hardware Store	Cosmetics Shop	Bar	Convenience Store
31	West University Place	Ice Cream Shop	Grocery Store	Café	Shopping Mall	Food Truck	Pizza Place	New American Restaurant	Italian Restaurant	Burger Joint
32	Webster	Mexican Restaurant	Coffee Shop	American Restaurant	Cosmetics Shop	Science Museum	Park	Pizza Place	Hotel	Furniture / Home Store

33	Tomball Neighborhood	Mexican Restaurant 1st Most Common Venue	Fast Food Restaurant 2nd Most Common Venue	Fast Food Restaurant 3rd Most Common Venue	Pizza Place 4th Most Common Venue	Burger Joint 5th Most Common Venue	Coffee Shop 6th Most Common Venue	Fast Food Restaurant 7th Most Common Venue	American Restaurant 8th Most Common Venue	Italian Restaurant 9th Most Common Venue
35	Sugar Land	Mexican Restaurant	Park	Grocery Store	Bakery	American Restaurant	Fast Food Restaurant	Burger Joint	Coffee Shop	Ice Cream Shop
36	Stafford city	Mexican Restaurant	Fast Food Restaurant	Department Store	Grocery Store	Pharmacy	Park	Liquor Store	Gym	Cajun / Creole Restaurant
39	Atascocita	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Fried Chicken Joint	Gym	Discount Store	Italian Restaurant	Golf Course	Gas Station
41	The Woodlands	Park	Pizza Place	Grocery Store	New American Restaurant	Mexican Restaurant	Fast Food Restaurant	Seafood Restaurant	Golf Course	American Restaurant
44	Fresno	Fast Food Restaurant	Mexican Restaurant	Burger Joint	Coffee Shop	Pharmacy	Donut Shop	Supermarket	Cosmetics Shop	Seafood Restaurant
45	Four Corners	Pharmacy	Coffee Shop	Mexican Restaurant	Grocery Store	Supermarket	Chinese Restaurant	Pet Store	Gym / Fitness Center	Fast Food Restaurant
48	Cinco Ranch	Grocery Store	Mexican Restaurant	Coffee Shop	Fast Food Restaurant	BBQ Joint	Liquor Store	Supermarket	Bakery	Sushi Restaurant
51	Mission Bend	Supermarket	Burger Joint	Pharmacy	Coffee Shop	Fast Food Restaurant	Park	Bar	Breakfast Spot	Discount Store
53	Fulshear	Fast Food Restaurant	Grocery Store	Pet Store	American Restaurant	BBQ Joint	Sandwich Place	Mexican Restaurant	Fried Chicken Joint	Burger Joint
54	Friendswood	Mexican Restaurant	Coffee Shop	Burger Joint	Park	Pizza Place	American Restaurant	Fast Food Restaurant	BBQ Joint	Grocery Store
56	Dickinson	Mexican Restaurant	Clothing Store	Burger Joint	Coffee Shop	Sandwich Place	Pizza Place	American Restaurant	Supermarket	Grocery Store
57	Deer Park	Mexican Restaurant	Coffee Shop	Burger Joint	Fast Food Restaurant	Seafood Restaurant	Ice Cream Shop	Cosmetics Shop	Spa	Breakfast Spot

Cluster 4

Out [130]:

	Neighborhood	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
1	Baytown	Mexican Restaurant	Discount Store	Steakhouse	American Restaurant	Coffee Shop	Ice Cream Shop	Fast Food Restaurant	Pizza Place	Chinese Restaurant
2	Angleton	Fast Food Restaurant	Discount Store	Pharmacy	Rental Car Location	Pizza Place	Airport	Taco Place	Fried Chicken Joint	Tex-Mex Restaurant
3	Alvin	Discount Store	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	Donut Shop	Gas Station	Fried Chicken Joint	Seafood Restaurant	American Restaurant
6	Clute	Pizza Place	Discount Store	Fast Food Restaurant	Mexican Restaurant	Fried Chicken Joint	Gas Station	Convenience Store	Pharmacy	American Restaurant
9	LaPorte	Sandwich Place	Gas Station	Pizza Place	Mexican Restaurant	Fast Food Restaurant	Discount Store	BBQ Joint	Fried Chicken Joint	Convenience Store
10	LaMarque	Mexican Restaurant	Sandwich Place	Discount Store	Burger Joint	Fast Food Restaurant	Bar	Pharmacy	Pizza Place	Italian Restaurant
25	Spring	American Restaurant	Pizza Place	Convenience Store	Burger Joint	Discount Store	Mexican Restaurant	Pharmacy	Grocery Store	BBQ Joint
29	Santa Fe	Clothing Store	Fast Food Restaurant	Shoe Store	Accessories Store	Gas Station	Discount Store	Pizza Place	Grocery Store	Fried Chicken Joint
34	Texas City	Discount Store	Fast Food Restaurant	Sandwich Place	Mexican Restaurant	Burger Joint	Convenience Store	Grocery Store	Fried Chicken Joint	Bar
37	Channelview	Mexican Restaurant	Discount Store	Pizza Place	Fast Food Restaurant	Fried Chicken Joint	Burger Joint	BBQ Joint	Pharmacy	Sandwich Place
42	Highlands	Sandwich	Discount	Fast Food	Fried Chicken	Hotel	Rest or Eat	Pizza Place	Convenience	Gas Station

43	Highlandus	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
46	Crosby County	Pizza Place	Fast Food Restaurant	Restaurant Mexican	Chicken Joint	Gas Station	Bar	Discount Store	Pharmacy	Sandwich Place
49	Sienna Plantation	Pizza Place	Sandwich Place	Pharmacy	Gym	Mexican Restaurant	Pet Store	Bank	Sushi Restaurant	Supermarket
55	Freeport	Pizza Place	Discount Store	Seafood Restaurant	Fast Food Restaurant	American Restaurant	Beach	Restaurant	Gas Station	Convenience Store
58	Hitchcock	Sandwich Place	Discount Store	Fast Food Restaurant	Mexican Restaurant	Burger Joint	Pizza Place	Convenience Store	Pharmacy	Gas Station

Cluster 5

Out[131]:

	Neighborhood	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
4	Dayton	Discount Store	Fast Food Restaurant	Mexican Restaurant	Ice Cream Shop	Gym	Pizza Place	Convenience Store	Pharmacy	Sandwich Place	Gas Station
7	Cleveland	Fast Food Restaurant	Mexican Restaurant	Pizza Place	Pharmacy	Convenience Store	Fried Chicken Joint	Gas Station	Discount Store	Grocery Store	Bar
8	Brookshire	Fast Food Restaurant	Fried Chicken Joint	Hotel	Athletics & Sports	Pizza Place	Discount Store	Construction & Landscaping	Park	Sandwich Place	Other Repairs Shop
15	Manvel	Fast Food Restaurant	Farm	Gas Station	Fried Chicken Joint	Seafood Restaurant	Sandwich Place	Hotel	Cajun / Creole Restaurant	Flower Shop	Athletics Sports
16	Liberty	Fast Food Restaurant	Burger Joint	Department Store	Gym / Fitness Center	Business Service	Fried Chicken Joint	Pizza Place	Mobile Phone Shop	Gas Station	Mexican Restaurant
20	Prairie View	Fast Food Restaurant	BBQ Joint	Discount Store	Gas Station	Seafood Restaurant	Pizza Place	Post Office	Hotel Bar	Sandwich Place	Business Service
23	Mont Belvieu	Fast Food Restaurant	Sandwich Place	Discount Store	Mexican Restaurant	Grocery Store	Fried Chicken Joint	Ice Cream Shop	Burger Joint	Gas Station	BBQ Joint
27	Sealy	Gas Station	Fast Food Restaurant	Sandwich Place	Discount Store	Pizza Place	Pharmacy	Burger Joint	BBQ Joint	State / Provincial Park	Market
42	Willis	Fast Food Restaurant	Gas Station	Golf Course	Sandwich Place	Pizza Place	American Restaurant	Mexican Restaurant	Resort	BBQ Joint	Coffee Shop
59	Hempstead	Fast Food Restaurant	Seafood Restaurant	Ice Cream Shop	Big Box Store	Farmers Market	Farm	Sandwich Place	Fried Chicken Joint	Bank	BBQ Joint

Sumarizing Clusters Characteristics

Cluster 1: Mexican Restaurant

Cluster 2: Athletics & Sports

Cluster 3: A combination of Mexican restaurants \$ Burger Joint

Cluster 4: Pizza Place & Sandwich Place

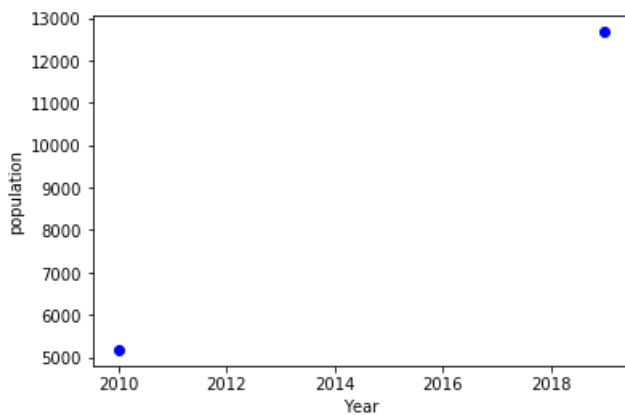
Cluster 5: Fast Food Restaurant

3.3 Inferial Statistical Testing`

we will find out what is the expected population growth in Manvel by 2025

	Year	population
0	2010	5179
1	2019	12671

Scatterplot of Manvel population in 2010 and 2019



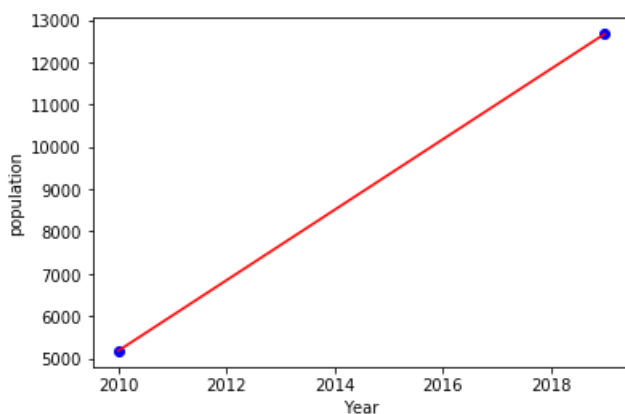
Modeling Using sklearn package

```
Coefficients:  [[832.44444444]]
Intercept:    [-1668034.33333333]
```

plot the fitted line

```
Out[100]:
```

```
Text(0, 0.5, 'population')
```



estimated Manvel population in 2025

```
17665.6666666666744
```

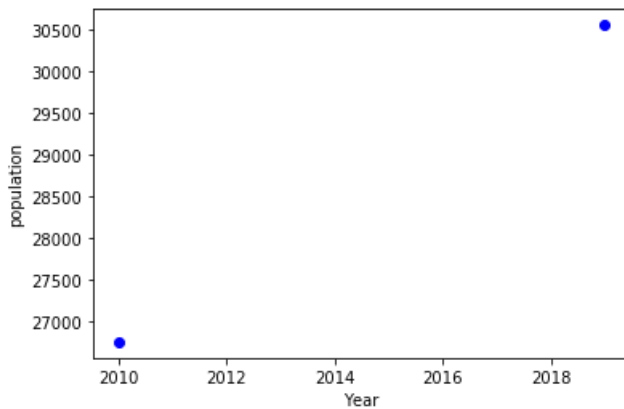
The Manvel population is expected to increase at a rate of 39.4% from 2019 to 2025

Expected population growth in Angleton by 2025

Angleton population in 2010 and 2019

	Year	population
0	2010	26748
1	2019	30550

Scatterplot of Angleton population in 2010 and 2019



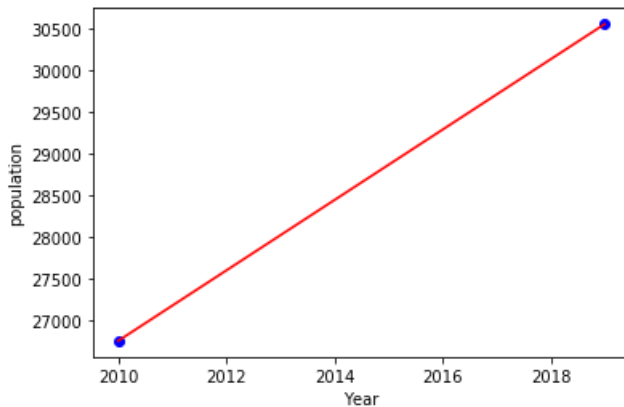
Modeling Using sklearn package

Coefficients: `[[422.44444444]]`
 Intercept: `[-822365.33333333]`

plot the fitted line

Out[109]:

Text(0, 0.5, 'population')



Angleton population in 2025

33084.66666666663

Out[113]:

8.297872340425531

The Angleton population is expected to increase at a rate of 8.3% from 2019 to 2025

4 Results

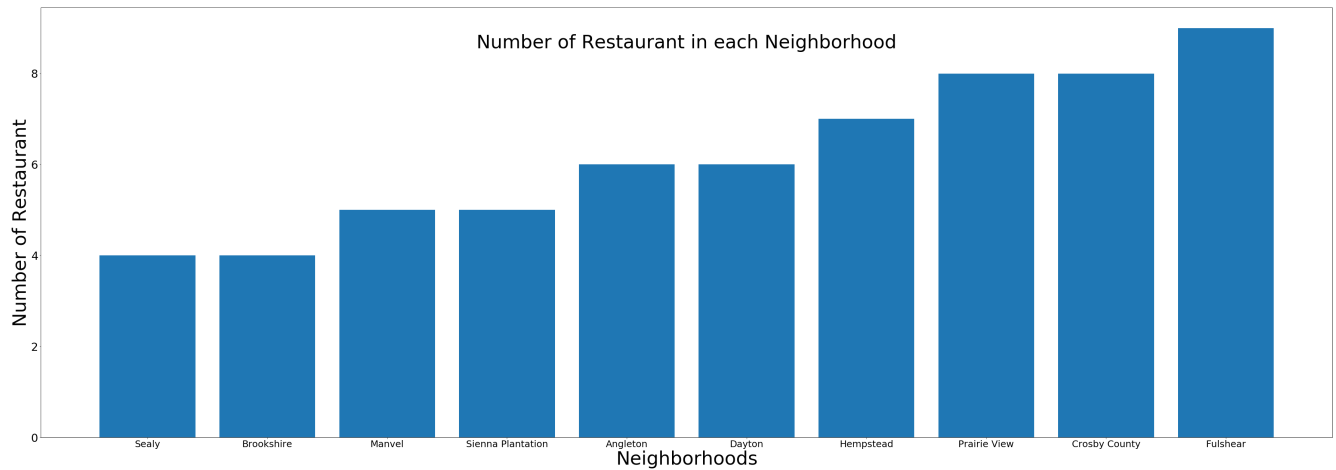
4 RESULTS

The question we are trying to answer by this project is what is the best neighborhood to open a restaurant in Houston? To answer the question we will first summarize the result gain from the study. The results are presented in form of Key Performance Indicators (KPI)

4.1 Key Performance Indicators

4.1.1 Number of Restaurant in each Neighborhood

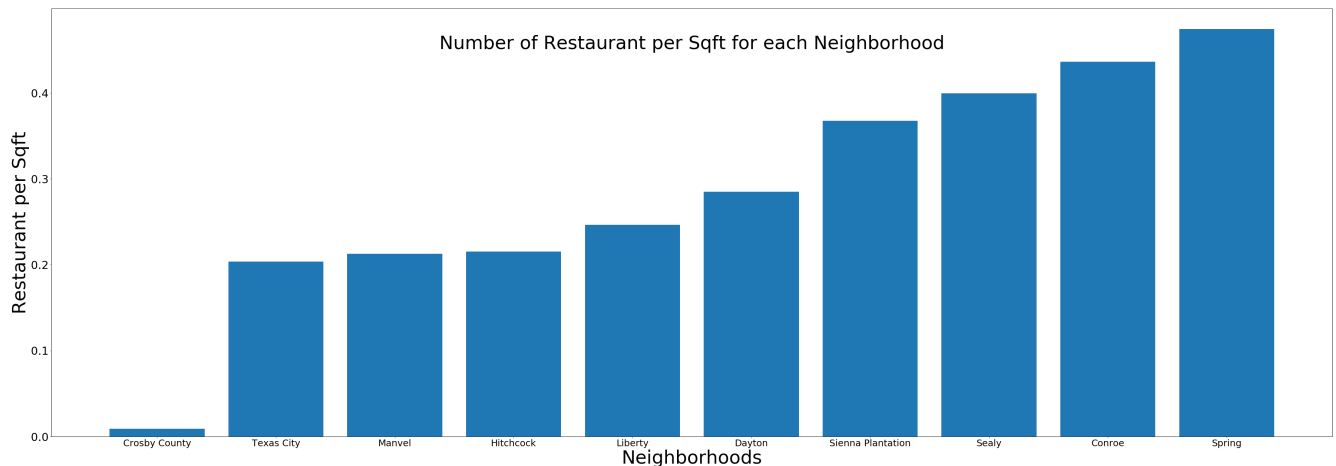
What are the Neighborhoods with less number of restaurants



the six Neighborhood with less reataurants are:

sealy, Brookshire, Manvel, Sienna Plantation, Angleton and Dayton

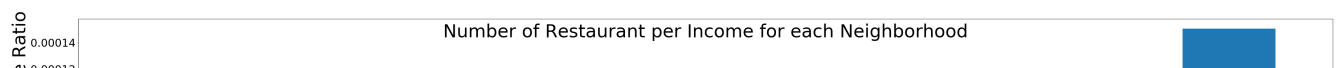
4.1.2 Number of Restaurants per sqft in each Neighborhood

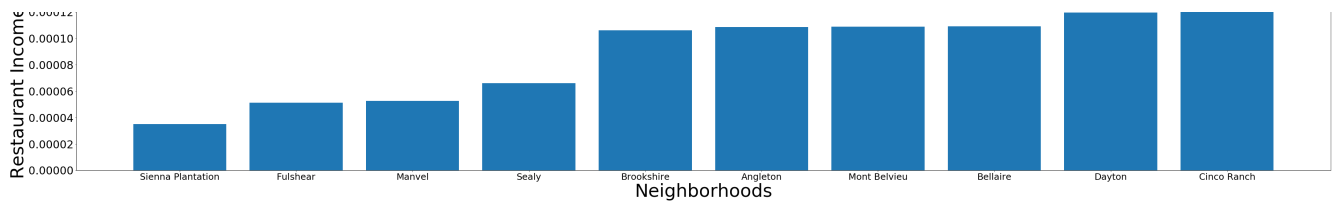


The less dense Neighborhood in terms of restaurant per sqft are:

Crosby County, Texas City, Manvel, Hitchcock, Liberty and Dayton

4.1.3 Number of Restaurants per family income in each Neighborhood

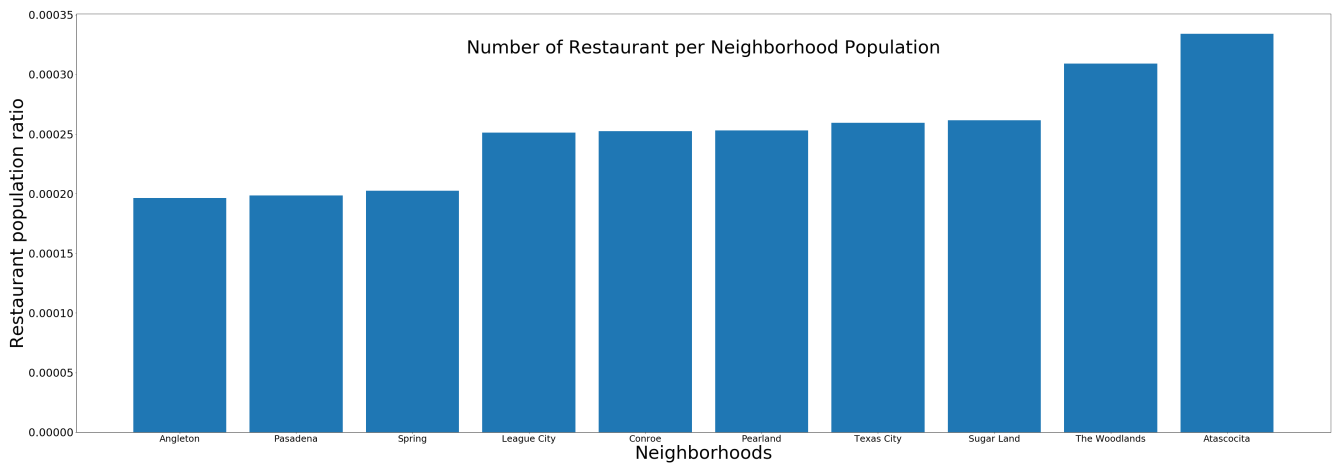




The Neighborhood with less restaurants per income are:

Sienna Plantation, Fulshear, Manvel, Sealy, Brookshire and Angleton

4.1.4 Number of Restaurants per population in each Neighborhood



Neighborhoods with less Restaurant per population are:

Angleton, Pasadena, Spring, Conroe, League City and Pearland

4.2 Sumarizing Clusters Characteristics

Cluster 1: Mexican Restaurant

Cluster 2: Athletics & Sports

Cluster 3: A combination of Mexican restaurants \$ Burger Joint

Cluster 4: Pizza Place & Sandwich Place

Cluster 5: Fast Food Restaurant

4.3 Population in Manvel and Angleton Neighborhood by 2025

Base on the population of 2010 and 2019, the population of Manvel was estimated to be 17665 in 2025. A 39.4% increase from 2019

Base on the population of 2010 and 2019, the population of Angleton was estimated to be 33085 in 2025. A 8.3% increase from 2019

5 Discussion

From the 4 KPI's above, we can all agreed that the family income is very important for where we want to establish the restaurant.

people who have higher income will probably eat out more often. But the income alone is not enough. We want to capture the neighborhood with less restaurant per income ratio.

The other KPI are mainly to find out where the competition is low.

Neighborhood with less number of restaurant: Sealy, Brookshire, Manvel, Sienna Plantation, Angleton and Dayton

Neighborhood with less restaurant per income: Sienna Plantation, Fulshear, Manvel, Sealy, Brookshire and Angleton

Neighborhood with less restaurant per sqft: Crosby County, Texas City, Manvel, Hitchcock, Liberty and Dayton

Neighborhood with less restaurant per population: Angleton, Pasadena, Spring, Conroe, League City and Pearland

Manvel and Angleton are the neighborhoods appearing in 3 of the 4 KPI's categories at the position of interest (low value of the KPI's).

Manvel and Angleton are the Neighborhoods I proposed we should open the new Restaurants at.

Since Angleton has a lower number of restaurant per population than Manvel, which means the competition might be better at Angleton, I will suggest to open the restaurant at Angleton.

Also Angleton has a larger population than Manvel. Although Manvel population is increasing at a faster rate, it will take more than 10 years to catch up.

In order of priority I will suggest Angleton, Manvel and then other neighborhood appearing at least in 2 of the KPI's in the right level like Sealy, Brookshire and Dayton.

6 Conclusion

I was able to find the Houston Neighborhood for opening the new restaurant at. In order of priority Angleton and Manvel.

Since according to the data Angleton and Manvel came out as the best neighborhoods to create a restaurant, another question is where exactly in Angleton or Manvel should the restaurant be opened?

This can be the subject of another project. Also the type of restaurant to be opened is also very important.

The k nearest neighbor help us to localize the restaurant by type on Houston map. This will guide to choose a type which is not already abundant in the neighborhood