

WHICH NEIGHBORHOODS ARE GOOD CHOICES FOR NEW MEXICAN RESTAURANTS AT HOUSTON



INTRODUCTION

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude
VenueCategory						
Mexican Restaurant	257	257	257	257	257	257
Pizza Place	184	184	184	184	184	184
Fast Food Restaurant	179	179	179	179	179	179
Sandwich Place	154	154	154	154	154	154
Discount Store	145	145	145	145	145	145
Fried Chicken Joint	125	125	125	125	125	125
Coffee Shop	120	120	120	120	120	120
Gas Station	107	107	107	107	107	107
Burger Joint	96	96	96	96	96	96
Bar	95	95	95	95	95	95

- 1. MEXICAN RESTAURANT IS THE 1ST MOST VENUES AT HOUSTON, RANKS HIGHER THAN GAS STATION, BAR, AND ALL OTHER RESTAURANT.**
- 2. IF SOMEONE WANT TO OPEN A MEXICAN RESTAURANT AT HOUSTON, WHERE SHOULD HER CHOOSE THE LOCATION?**

DATA

```
In [152]: # convert the venues list into a new DataFrame
venues_df = pd.DataFrame(venues)

# define the column names
venues_df.columns = ['Neighborhood', 'Latitude', 'Longitude', 'VenueName', 'VenueLatitude', 'VenueLongitude', 'VenueCategory']

print(venues_df.shape)
venues_df.head()
```

(5115, 7)

Out[152]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Willowbrook	29.9524	-95.54463	Bed Bath & Beyond	29.953517	-95.543865	Furniture / Home Store
1	Willowbrook	29.9524	-95.54463	Babin's Seafood House	29.955088	-95.544452	Seafood Restaurant
2	Willowbrook	29.9524	-95.54463	Costco	29.954658	-95.547697	Warehouse Store
3	Willowbrook	29.9524	-95.54463	buybuy BABY	29.953127	-95.543557	Kids Store
4	Willowbrook	29.9524	-95.54463	Pho An 2	29.956606	-95.543805	Vietnamese Restaurant

1. WIKI

2. GEOCODER

3. FOURSQUARE API

METHODOLOGY

Cluster Neighborhoods

```
In [512]: # set number of clusters
kclusters = 5

#original
#kl_clustering = kl_gym.drop(["Neighborhoods"], 1)

#change to:
kl_clustering = kl_r.drop(["Neighborhoods"], 1)

print(kl_clustering.shape)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(kl_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]

(88, 11)

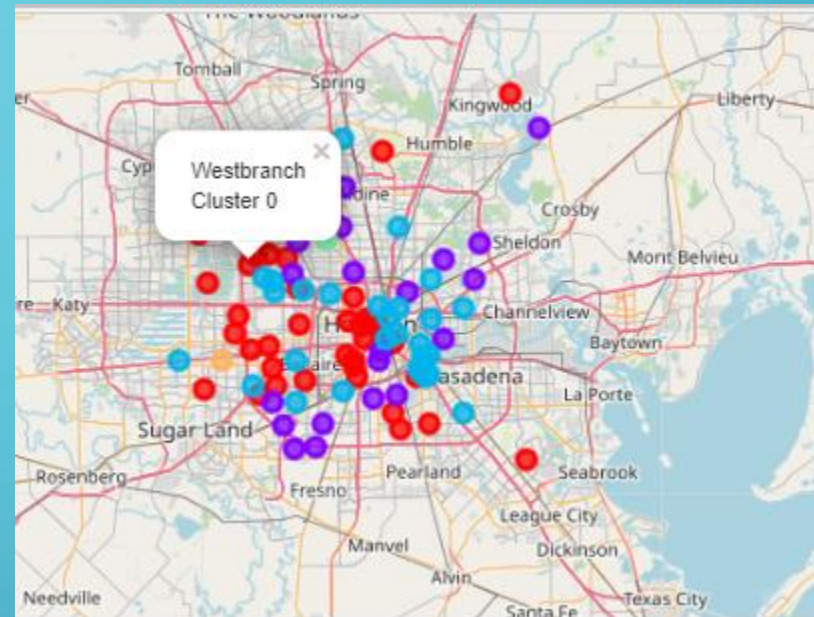
Out[512]: array([3, 0, 2, 4, 0, 0, 0, 1, 0, 0])

In [513]: # create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
kl_merged = kl_r.copy()

# add clustering labels
kl_merged["Cluster Labels"] = kmeans.labels_
```

WE CLUSTER THE NEIGHBOURHOODS INTO 5 CLUSTERS

RESULTS



```
In [521]: k1_merged.groupby('Cluster Labels').mean()
```

Out[521]:

	Mexican Restaurant	Pizza Place	Fast Food Restaurant	Sandwich Place	Fried Chicken Joint	Burger Joint	American Restaurant	Vietnamese Restaurant	Chinese Restaurant	Seafood Restaurant	Italian Restaurant	Latitude	Longit.
Cluster Labels													
0	0.030222	0.025456	0.016077	0.024366	0.010695	0.016468	0.015643	0.008315	0.015032	0.009439	0.012502	29.759653	-95.4371
1	0.008773	0.049978	0.100218	0.050279	0.053853	0.021471	0.010871	0.001826	0.015379	0.009865	0.000560	29.839513	-95.3701
2	0.097117	0.043596	0.039028	0.037571	0.035934	0.021157	0.013641	0.012542	0.008913	0.014417	0.007123	29.755647	-95.3751
3	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000	0.142857	0.000000	0.000000	29.870470	-95.4351
4	0.000000	0.015625	0.031250	0.015625	0.031250	0.000000	0.000000	0.265625	0.000000	0.000000	0.000000	29.710880	-95.5951



DISCUSSION

OPEN A NEW MEXICAN RESTAURANT AT CLUSTER 1, OR NORTH AND EAST-NORTH HOUSTON