Austin Texas: Determining the Best Neighborhoods to Locate a Restaurant Business

Solomon Green Ejukorlem

July 10, 2020

1. Introduction

1.1 Background

There are several important factors to consider in making the decision on where to locate a restaurant business, having in mind that location can affect cost and profit, and also taking note of the fact that profit maximization is the primary goal of most businesses.

It is important to undertake a surrounding business and competitor analysis in determining the location of the business. Are businesses in the area doing well? Are people and businesses in the area affluent? What is the average income of people in the area? What types of restaurants are in the area? If, for example, there are already six barbeque restaurants in the neighborhood, it may not make business sense to establish another in the neighborhood.

If there is a night club in the neighborhood it might be expedient to situate a restaurant that caters for after hour crowd, after they are done with night drinking and clubbing.

Visibility and accessibility are also vital considerations; one would want not a location in a side street with very little foot traffic and one would prefer a location accessible to car traffic, parking space should, for instance be available.

What about the cost of the space? Is it affordable? Will it sustain or consume profit? Is there enough space to accommodate your vision?

What is the crime rate in the neighborhood? You would not want crime to chase away your customers. What is the population density of the neighborhood?. What are the demographics in terms of age, gender, employment, unemployment, underemployment, income bracket? How accessible are the sources of food supply?

1.2 Problem

The above questions and considerations are within the purview of location data and other data sources to answer. This is precisely what this project is about, using data to analyze, provide insight and provide solution. This project aims to provide a guide to entrepreneurs seeking to establish a restaurant business in a neighborhood the best location to site the business. Austin Texas in the USA has been chosen as the case study.

1.3 Target Audience

This project is for investors who are looking for prospects to invest in, particularly for entrepreneurs who want to invest in restaurant business in the United States of America but are curious of where best to locate it. The information here, though targeted at restaurant business are also useful for other businesses.

Startups would find this information also invaluable.

Data scientist, particularly beginners will gain insights into real world application of data science.

2. Data acquisition and cleaning

2.1 Data sources

The major dataset used for this project is sourced from data.austintexas.gov.

Google maps are used to derive the latitude and longitude coordinates of the Neighborhoods in Austin, Texas derived from the above dataset.

The Foursquare API was used to provide location data of venues and allied information in the neighborhoods.

This project uses an unsupervised learning approach, which uses data to identify patterns and information about major trends in the very dynamic restaurant industry in Austin neighborhoods, this is with a view to also provide insight into some atypical behaviors that are unique to Austin restaurant business; which are more evident in view of the new realities introduced by the Covid-19 pandemic.

The K-means analytic clustering algorithm will be used to segment restaurants into clusters with a view to identifying the best clusters to locate what type of restaurant as prescribed by the analysis.

2.2 Import Python Libraries

Prior to data wrangling and data exploration I had to import some important python libraries, many of which had been previously installed, these libraries include Numpy, which is a library to handle data in a vectorized manner, Pandas is another library, and this is used for data analysis.

JSON was also imported to handle JSON files. Geopy is a geospatial library that is used in conjunction with another library Nominatim for converting street addresses to latitude and longitude coordinate values.

Matplotlib and associated plotting modules were also imported, these libraries are used for plotting data visualization objects like bar charts, histogram and scatter plots. Also imported is Sklearn library which contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

2.3 Read Austin Neighborhood Dataset

```
In [2]: df = pd.read_excel("Neighborhood.xlsx")
        df.head(20)
```

Out[2]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed 1
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
4	NaN	Population and Housing, Table I	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
5	NaN	by Neighborhood Reporting Area	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
6	NaN	Census 2010 Data	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
7	NaN	Map of Neighborhood Reporting Areas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
8	NaN	http://www.austintexas.gov/sites/default/files...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
10	NaN		NaN	NaN	NaN	non-Hispanic	African	NaN	NaN	NaN	NaN	Tota
11	NaN	NaN	NaN	NaN	Total	White	American	Hispanic	Asian	Other	NaN	Housin
12	NaN	Neighborhood	Latitude	Longitude	Population	Percentage	Percentage	Percentage	Percentage	Percentage	NaN	Unit
13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
16	NaN	ALLENDALE	29.6855	-95.2132	6643	0.831703	0.0189673	0.10974	0.0189673	0.0206232	NaN	361
17	NaN	ANDERSON MILL	30.4558	-97.8071	28473	0.647455	0.0332947	0.131107	0.163277	0.0248657	NaN	1150
18	NaN	AVERY RANCH--LAKELINE	30.4901	-97.8249	14785	0.58485	0.0395671	0.140615	0.206628	0.0283395	NaN	610
19	NaN	BARTON CREEK MALL	30.2577	-97.8094	5147	0.793472	0.00816009	0.0922868	0.0866524	0.0194288	NaN	219

I had to read the major dataset for this project using the pandas library read function, the file had previously been downloaded to my local device from where I had to read it with the command shown in the figure below. A data frame was derived from this operation.

2.4 Data Wrangling

The data is raw and needs some cleaning. The process of cleaning the data is also known as data wrangling.

2.4.1 Drop Unnecessary rows

One of the data wrangling operations I performed is to drop some unnecessary rows, drop some unnecessary columns and rename some columns to more meaningful names as well as reset the data frame index.

In [3]:
```
df.drop(df.index[[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]], inplace=True)
df
```

Out[3]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unna
16	NaN	ALLENDALE	29.6855	-95.2132	6643	0.831703	0.0189673	0.10974	0.0189673	0.0206232	NaN	3612	
17	NaN	ANDERSON MILL	30.4558	-97.8071	28473	0.647455	0.0332947	0.131107	0.163277	0.0248657	NaN	11507	1
18	NaN	AVERY RANCH--LAKELINE	30.4901	-97.8249	14785	0.58485	0.0395671	0.140615	0.206628	0.0283395	NaN	6108	
19	NaN	BARTON CREEK MALL	30.2577	-97.8094	5147	0.793472	0.00816009	0.0922868	0.0866524	0.0194288	NaN	2195	
20	NaN	BARTON HILLS	30.2532	-97.8207	8022	0.78372	0.0138369	0.135627	0.0433807	0.0234356	NaN	4965	
21	NaN	BERGSTROM	30.1975	-97.6685	179	0.513966	0.273743	0.212291	0	0	NaN	2	
22	NaN	BLUFF SPRINGS	30.1726	-97.7708	23000	0.175696	0.0798696	0.71587	0.0106957	0.0178696	NaN	7947	

Drop Unnecessary Columns and RenameColumns

In [4]:
```
df.columns = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19']
```

In [5]:
```
df.drop(['1','5','6','7','8','9','10','11','12','13','14','15','16','17','18','19'],axis=1,inplace=True)
```

In [6]:
```
df.rename({'2':'Neighborhood', '3':'latitude','4':'longitude'},axis=1, inplace=True)
```

In [7]:
```
df.head(2)
```

Out[7]:

	Neighborhood	latitude	longitude
16	ALLENDALE	29.6855	-95.2132
17	ANDERSON MILL	30.4558	-97.8071

3.0 Data Exploration

Data exploration are methods data scientist use to better understand the datasets they are working on, it involves the use of visual tools to understand what is in a dataset and the features of the data, rather than through traditional data management systems.

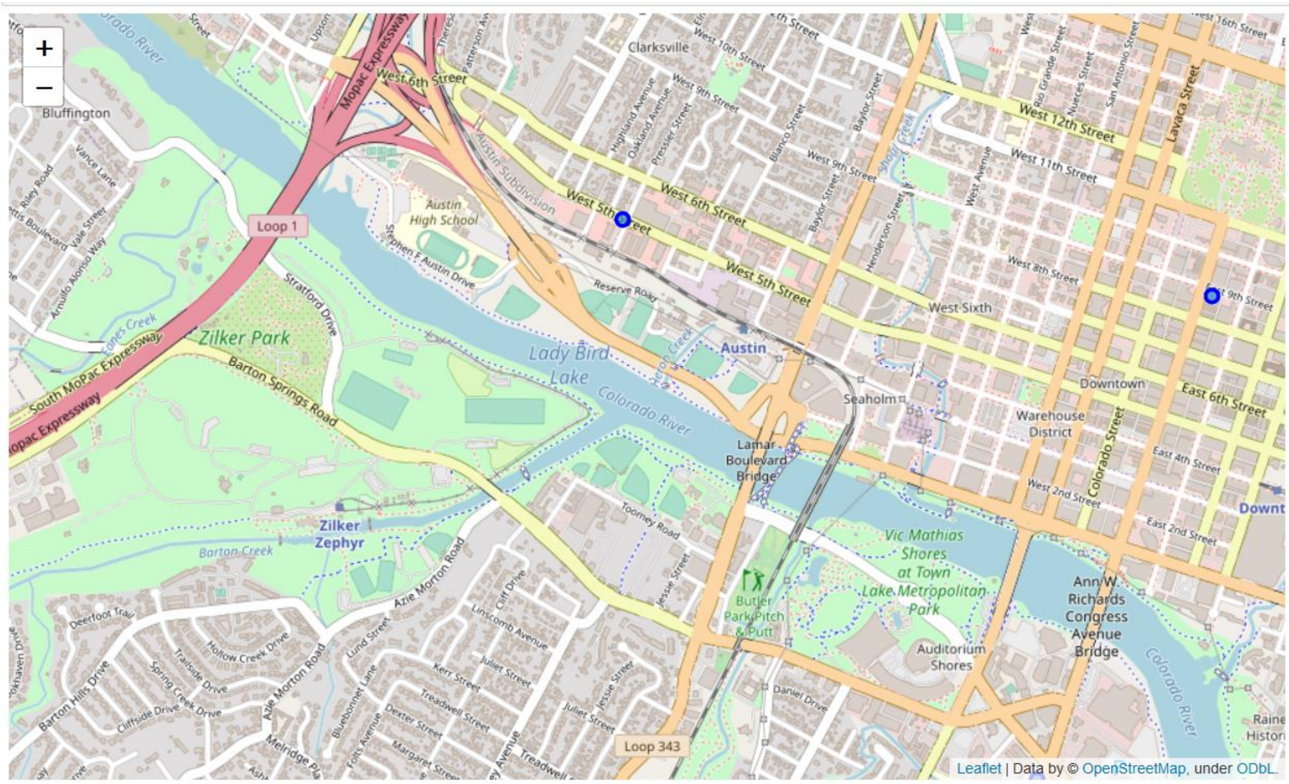
3.1 Use Folium To Plot OpenStreetMap of Austin Texas

Folium is a python library that makes it easy to visualize data that's been

In [11]:
```
# create map of Austin using latitude and longitude values
map_Austin = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng,neighborhood in zip(df['latitude'], df['longitude'], df['Neighborhood']):
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_Austin)
map_Austin
```

Out[11]: Make this Notebook Trusted to load map: File -> Trust Notebook manipulated in Python on an interactive leaflet map. It used on a map for visualizations.



3.2 Using Foursquare API to obtain Neighborhood Location Data

Foursquare is an API used to explore the world around us. It is used to obtain location data of venues such as churches, hotels, restaurants and gas stations. In this project we are going to use it to obtain data of venues, particularly about restaurants.

To use Foursquare we are required to register and obtain an Identity number, a secret API key and other credentials such as the version number of the API. With the credentials we will send a query to the API requesting to be furnished with location data. The API will respond with the requested data if the query is successful.

In this project I used Foursquare to obtain location data about venues in each neighborhood, the result of the query was returned in JSON format, I then convert the response to a data frame using Pandas.

Process the response from JSON into Pandas dataframe

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

	name	categories	lat	lng
0	Shell	Gas Station	29.686708	-95.209602
1	The Cuban Spot	Food Truck	29.686374	-95.215741
2	JR Foundation Repair	Construction & Landscaping	29.682518	-95.212793

```
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
3 venues were returned by Foursquare.
```

I used the Foursquare API to explore venues in all the neighborhoods in Austin, Texas and proceeded to explore venues in all neighborhoods that are specifically restaurants. I used Pandas to categorize the restaurants by type.

Categorize Restaurants by Type

```
print('There are {} unique categories.'.format(len(Austin_restaurants['Venue Category'].unique())))
```

```
There are 31 unique categories.
```

Create a dataframe of top 10 categories

```
Austin_Top_10_Restaurants = Austin_restaurants['Venue Category'].value_counts()[0:10].to_frame(name='frequency')
Austin_Top_10_Restaurants = Austin_Top_10_Restaurants.reset_index()
Austin_Top_10_Restaurants.rename(index=str, columns={"index": "Venue_Category", "frequency": "Frequency"}, inplace=True)
Austin_Top_10_Restaurants
```

	Venue_Category	Frequency
0	Mexican Restaurant	64
1	American Restaurant	16
2	Fast Food Restaurant	14
3	Thai Restaurant	10
4	Seafood Restaurant	10
5	Sushi Restaurant	9
6	Tex-Mex Restaurant	8
7	Italian Restaurant	8
8	Chinese Restaurant	8
9	Vietnamese Restaurant	7

3.3 Visualize the Top Ten Categories of Restaurants in the Neighborhoods

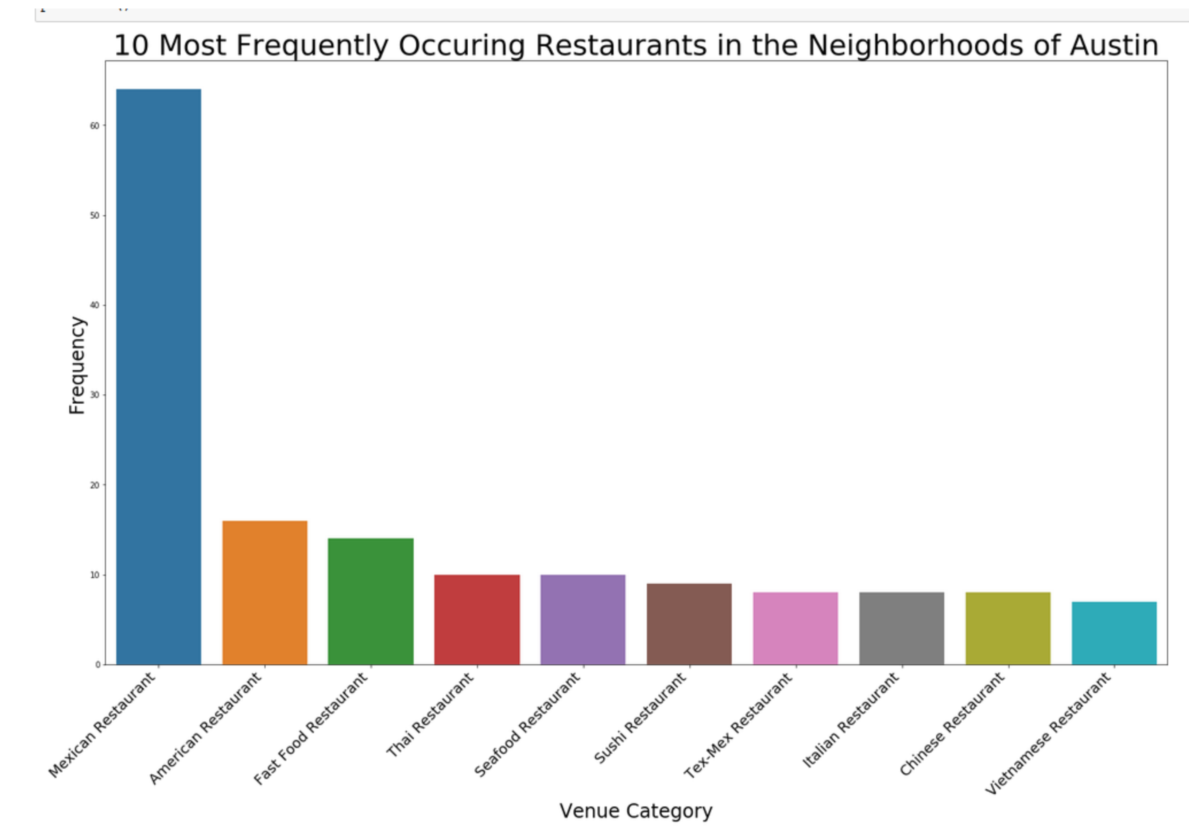
Having categorized the restaurant in the neighborhoods by type, I proceeded to visualize the result for better understanding. For the visualization I used Matplotlib in conjunction with another python library called Seaborn. Seaborn is a python statistical data visualization library efficient in the drawing of statistical graphics.

```
In [31]: import seaborn as sns
         from matplotlib import pyplot as plt

         s=sns.barplot(x="Venue_Category", y="Frequency", data=Austin_Top_10_Restaurants)
         s.set_xticklabels(s.get_xticklabels(), rotation=45, horizontalalignment='right')

         plt.title('10 Most Frequently Occuring Restaurants in the Neighborhoods of Austin', fontsize= 36)
         plt.xlabel("Venue Category", fontsize=24)
         plt.ylabel ("Frequency", fontsize=24)
         plt.xticks(fontsize=18)
         plt.savefig("Most_Freq_Restaurants.png", dpi=300)
         plt.rcParams["figure.figsize"] = [24,14]

         plt.show()
```



3.4 One Hot Coding

One hot coding is a data preprocessing method used to convert categorical data into numerical binary format with the aim of eliminating weighting bias amongst the categories by the computer.

```
# one hot encoding
Austin_onehot = pd.get_dummies(Austin_restaurants[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
Austin_onehot['Neighborhood'] = Austin_restaurants['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [Austin_onehot.columns[-1]] + list(Austin_onehot.columns[:-1])
Austin_onehot = Austin_onehot[fixed_columns]

Austin_onehot.head()
```

```
Austin_onehot.head()
```

	Neighborhood	American Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Gluten-free Restaurant	Greek Restaurant	Hal Restaura
1	BARTON CREEK MALL	1	0	0	0	0	0	0	0	0	0	0	
2	BARTON CREEK MALL	0	0	0	0	0	0	0	0	0	0	0	
3	BARTON CREEK MALL	0	0	0	1	0	0	0	0	0	0	0	
4	BARTON CREEK MALL	0	0	0	0	0	0	0	0	0	0	0	
5	BARTON CREEK MALL	0	0	0	0	0	0	0	0	0	0	0	

3.5 Grouping of Restaurants by Neighborhood and Display Mean Frequency of Occurrence of Each Category

After One Hot Coding, I had to group the restaurants by neighborhood and show the mean frequency of each category

```
In [34]: Austin_grouped = Austin_onehot.groupby('Neighborhood').mean().reset_index()
         Austin_grouped
```

Out[34]:

	Neighborhood	American Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Gluten-free Restaurant	Greek Restaurant	H Restau
0	BARTON CREEK MALL	0.200000	0.000000	0.0	0.200000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
1	CENTRAL EAST AUSTIN	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
2	CORONADO HILLS	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
3	DAVENPORT--LAKE AUSTIN	0.250000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
4	DOWNTOWN	0.142857	0.000000	0.0	0.142857	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
5	EAST CESAR CHAVEZ	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000	0.000000	0.000
6	EAST CONGRESS	0.200000	0.000000	0.0	0.200000	0.000000	0.000000	0.200000	0.000000	0.0	0.000	0.000000	0.000

3.6 Obtain the Top Ten Most Common Restaurant per Neighborhood.

For our analysis, we need to know the top most common restaurant in each neighborhood; these data are invaluable in our goal in determining which neighborhood is suitable or not in locating a new restaurant.

```
num_top_venues = 10

for hood in Austin_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = Austin_grouped[Austin_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----BARTON CREEK MALL----
                        venue  freq
0         American Restaurant   0.2
1          Chinese Restaurant   0.2
2             Thai Restaurant   0.2
3            Sushi Restaurant   0.2
4          Mexican Restaurant   0.2
5         Japanese Restaurant   0.0
6  Vegetarian / Vegan Restaurant   0.0
7          Turkish Restaurant   0.0
8          Tex-Mex Restaurant   0.0
9          Seafood Restaurant   0.0


----CENTRAL EAST AUSTIN----
                    venue  freq
0      Italian Restaurant  0.18
1  New American Restaurant  0.18
2         Thai Restaurant  0.09
3        Sushi Restaurant  0.09
4      Seafood Restaurant  0.09
5              Restaurant  0.09
6  Middle Eastern Restaurant  0.09
7      Mexican Restaurant  0.09
8  Latin American Restaurant  0.09
9     Japanese Restaurant  0.00


----CORONADO HILLS----
                        venue  freq
0          Tex-Mex Restaurant  0.25
1            Sushi Restaurant  0.25
2          Seafood Restaurant  0.25
3          Mexican Restaurant  0.25
4         American Restaurant  0.00
5         Japanese Restaurant  0.00
6  Vegetarian / Vegan Restaurant  0.00
7          Turkish Restaurant  0.00
8             Thai Restaurant  0.00
9                  Restaurant  0.00
```

3.6 Create a Pandas DataFrame of the Top Ten Most Common Restaurants Per Neighborhood

```python
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

```python
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Austin_grouped['Neighborhood']

for ind in np.arange(Austin_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Austin_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head(30)
```

	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	BARTON CREEK MALL	American Restaurant	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Chinese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant	Filipino Restaurant
1	CENTRAL EAST AUSTIN	Italian Restaurant	New American Restaurant	Seafood Restaurant	Latin American Restaurant	Mexican Restaurant	Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Thai Restaurant	Cajun / Creole Restaurant
2	CORONADO HILLS	Mexican Restaurant	Tex-Mex Restaurant	Sushi Restaurant	Seafood Restaurant	Vietnamese Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant
3	DAVENPORT--LAKE AUSTIN	Italian Restaurant	Restaurant	Mexican Restaurant	American Restaurant	Tex-Mex Restaurant	Hawaiian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
4	DOWNTOWN	American Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Sushi Restaurant	Seafood Restaurant	Latin American Restaurant	Mexican Restaurant	Filipino Restaurant	Greek Restaurant	Gluten-free Restaurant
5	EAST CESAR CHAVEZ	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant
6	EAST CONGRESS	Mexican Restaurant	American Restaurant	Chinese Restaurant	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Cuban Restaurant	Dim Sum Restaurant	Filipino Restaurant
7	FOUR POINTS	Seafood Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
8	FRANKLIN PARK	Mexican Restaurant	Tex-Mex Restaurant	Sushi Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant	Filipino Restaurant	Dim Sum Restaurant
9	GALINDO	Italian Restaurant	Seafood Restaurant	New American Restaurant	Fast Food Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant
10	GATEWAY	Seafood Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
11	GEORGIA ACRES	Mexican Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
12	GOVALLE	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	Thai Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant

3.7 Clustering the Neighborhoods

At this juncture, clustering process will be used to identify relationships in our data and to segment the restaurants based on the relationships. Amongst the various clustering algorithms available, we shall use the K means algorithm.

```
In [39]: # set number of clusters (I choose 5)
         kclusters = 5

         Austin_grouped_clustering = Austin_grouped.drop('Neighborhood', 1)

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

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

Out[39]: array([0, 2, 0, 0, 0, 4, 0, 0, 0, 2])
```

```
In [40]: Austin_merged = df
         Austin_merged.head(10)
```

Out[40]:

	Neighborhood	latitude	longitude
0	ALLENDALE	29.6855	-95.2132
1	ANDERSON MILL	30.4558	-97.8071
2	AVERY RANCH--LAKELINE	30.4901	-97.8249
3	BARTON CREEK MALL	30.2577	-97.8094
4	BARTON HILLS	30.2532	-97.8207
5	BERGSTROM	30.1975	-97.6685
6	BLUFF SPRINGS	30.1726	-97.7708
7	BOULDIN	30.1726	-97.7708
8	BRENTWOOD	30.3295	-97.7481
9	BRODIE LANE	30.1902	-97.8477

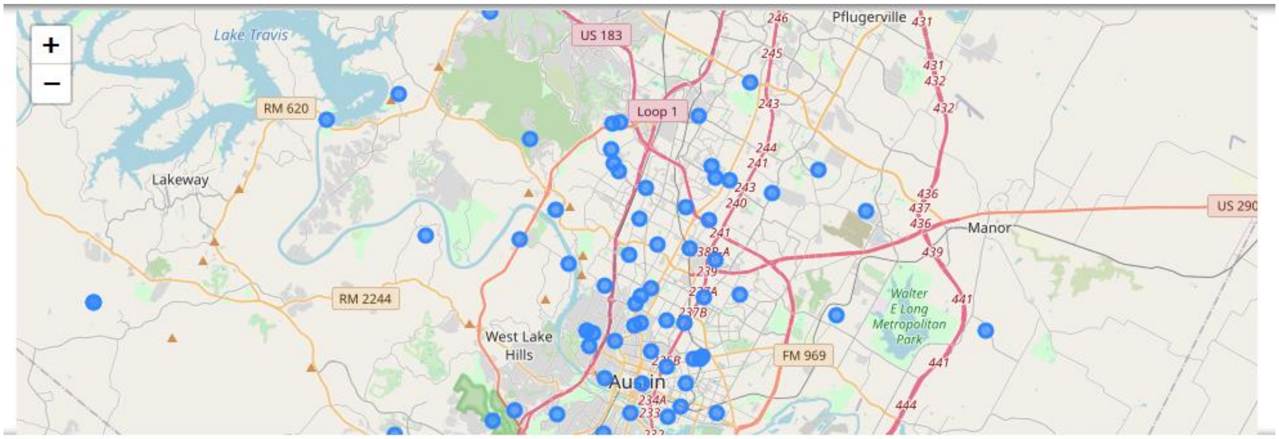
3.7 Visualize the Resulting Clusters

```
In [49]: # create map
         map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

         # set color scheme for the clusters
         x = np.arange(kclusters)
         ys = [i + x + (i*x)**2 for i in range(kclusters)]
         colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
         rainbow = [colors.rgb2hex(i) for i in colors_array]

         # add markers to the map
         markers_colors = []
         for lat, lon, poi, cluster in zip(Austin_merged['latitude'],Austin_merged['longitude'],Austin_merged['Neighborhood'], Austin_r
             label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
             folium.CircleMarker(
                 [lat, lon],
                 radius=5,
                 popup=label,
                 #color=rainbow[cluster-1],
                 fill=True,
                 #fill_color=rainbow[cluster-1],
                 fill_opacity=0.7).add_to(map_clusters)

         map_clusters
```



3.7.2 Examine Clusters

Cluster 1

```
Austin_merged.loc[Austin_merged['Cluster Labels']==0]
```

	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	8th Most Common Venue
3	BARTON CREEK MALL	30.2577	-97.8094	0.0	American Restaurant	Mexican Restaurant	Thai Restaurant	Sushi Restaurant	Chinese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant
16	CORONADO HILLS	30.3272	-97.702	0.0	Mexican Restaurant	Tex-Mex Restaurant	Sushi Restaurant	Seafood Restaurant	Vietnamese Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
18	DAVENPORT--LAKE AUSTIN	30.3366	-97.8063	0.0	Italian Restaurant	Restaurant	Mexican Restaurant	American Restaurant	Tex-Mex Restaurant	Hawaiian Restaurant	Asian Restaurant	Cajun / Creole Restaurant
24	DOWNTOWN	30.2728	-97.761	0.0	American Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Sushi Restaurant	Seafood Restaurant	Latin American Restaurant	Mexican Restaurant	Filipino Restaurant
26	EAST CONGRESS	30.2107	-97.781	0.0	Mexican Restaurant	American Restaurant	Chinese Restaurant	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Cuban Restaurant
28	FOUR POINTS	30.4038	-97.8711	0.0	Seafood Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
29	FRANKLIN PARK	30.1973	-97.7661	0.0	Mexican Restaurant	Tex-Mex Restaurant	Sushi Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
32	GATEWAY	30.39	-97.7573	0.0	Seafood Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
34	GOVALLE	30.2595	-97.7209	0.0	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	Thai Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant
36	HANCOCK	30.2981	-97.7417	0.0	Thai Restaurant	New American Restaurant	Mexican Restaurant	Fast Food Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
42	HOLLY	30.2549	-97.7276	0.0	Mexican Restaurant	Thai Restaurant	Cajun / Creole Restaurant	Korean Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant
45	JOHNSTON TERRACE	30.2569	-97.7012	0.0	Mexican Restaurant	American Restaurant	French Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
49	MCKINNEY	30.1984	-97.7664	0.0	Mexican Restaurant	Tex-Mex Restaurant	Sushi Restaurant	Vietnamese Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
53	MONTOPOLIS	30.2296	-97.7328	0.0	Vietnamese Restaurant	Mexican Restaurant	Asian Restaurant	Seafood Restaurant	Filipino Restaurant	Indian Chinese Restaurant	Cajun / Creole Restaurant	Chinese Restaurant
56	NORTH BURNET	30.3909	-97.753	0.0	Japanese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Mexican Restaurant	Vietnamese Restaurant	Filipino Restaurant	Greek Restaurant	Gluten-free Restaurant
58	NORTH LOOP	30.314	-97.7366	0.0	Mexican Restaurant	Italian Restaurant	Gluten-free Restaurant	Tex-Mex Restaurant	Japanese Restaurant	Restaurant	Mediterranean Restaurant	Greek Restaurant
59	NORTH SHOAL CREEK	30.368	-97.7534	0.0	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Cuban Restaurant	Dim Sum Restaurant

Cluster 2

```
Austin_merged.loc[Austin_merged['Cluster Labels']==1]
```

	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	8th Most Common Venue	9th Most Common Venue	10th M Comi Ve
65	PARKER LANE	30.2231	-97.7487	1.0	American Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant	Fre Restau
71	ROGERS HILL	30.2814	-97.7137	1.0	Italian Restaurant	American Restaurant	Turkish Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast F Restau
72	ROSEDALE	30.3157	-97.7609	1.0	American Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant	Fre Restau
88	UPPER BOGGY CREEK	30.1859	-97.7728	1.0	American Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant	Fre Restau

Cluster 3

```
Austin_merged.loc[Austin_merged['Cluster Labels']==2]
```

	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	8th Most Common Venue	9th Most Common Venue
11	CENTRAL EAST AUSTIN	30.2706	-97.7415	2.0	Italian Restaurant	New American Restaurant	Seafood Restaurant	Latin American Restaurant	Mexican Restaurant	Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Thai Restaurant
30	GALINDO	30.2363	-97.7856	2.0	Italian Restaurant	Seafood Restaurant	New American Restaurant	Fast Food Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
35	GRACY WOODS	30.3936	-97.7107	2.0	Fast Food Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant
39	HERITAGE HILLS	30.3457	-97.7057	2.0	Seafood Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
52	MLK-183	30.283	-97.7091	2.0	Seafood Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant
54	MUELLER	30.2985	-97.7183	2.0	Fast Food Restaurant	Mediterranean Restaurant	Restaurant	Mexican Restaurant	Filipino Restaurant	Halal Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
55	NACA	30.3649	-97.7023	2.0	Halal Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
82	ST. EDWARDS	30.2274	-97.7702	2.0	Chinese Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant
85	TECH RIDGE	30.4092	-97.6833	2.0	Indian Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
87	UNIVERSITY HILLS	30.3112	-97.6889	2.0	Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant
89	UT	30.2849	-97.7362	2.0	Fast Food Restaurant	American Restaurant	Asian Restaurant	New American Restaurant	Indian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant
95	WESTGATE	30.2249	-97.8153	2.0	Asian Restaurant	Vietnamese Restaurant	Indian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Filipino Restaurant

Cluster 4

```
Austin_merged.loc[Austin_merged['Cluster Labels']==3]
```

	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	8th Most Common Venue	9th Most Common Venue	10t Co
63	OLD WEST AUSTIN	30.2944	-97.771	3.0	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Res
66	OLD WEST AUSTIN	30.2944	-97.771	3.0	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Res
100	WINDSOR ROAD	30.2936	-97.7672	3.0	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Res

Cluster 5

```
Austin_merged.loc[Austin_merged['Cluster Labels']==4]
```

	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	8th Most Common Venue	9th Most Common Venue	10t Co
25	EAST CESAR CHAVEZ	30.2567	-97.7473	4.0	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest
33	GEORGIA ACRES	30.3513	-97.7178	4.0	Mexican Restaurant	Korean Restaurant	Vietnamese Restaurant	Indian Chinese Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fas Rest
40	HIGHLAND	30.3341	-97.7328	4.0	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest
43	HYDE PARK	30.3073	-97.7445	4.0	Mexican Restaurant	American Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest
57	NORTH LAMAR	30.3703	-97.7043	4.0	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest
77	SOUTH LAMAR	30.2381	-97.8002	4.0	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest
78	SOUTH MANCHACA	30.2194	-97.801	4.0	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Cuban Restaurant	Dim Sum Restaurant	Fast Food Restaurant	F Rest

3.7.3 Determine the number of Neighborhoods per Cluster

```
Austin_merged['Cluster Labels'].value_counts()
```

```
0.0    29
2.0    12
4.0     7
1.0     4
3.0     3
Name: Cluster Labels, dtype: int64
```

4.0 Results

The above analysis shows that Cluster 1 has a high concentration of neighborhoods where the restaurant business thrives with Mexican restaurants dominant, closely followed by American restaurant, Cluster 3 has next higher density of neighborhoods with thriving restaurants, and interestingly Cluster 3 has a better blend of restaurant types than Cluster 1, Cluster 5 has a fair share of neighborhoods with restaurants, it can be conviniently be called 'The Mexican Restaurant Cluster' as practically all the neighborhood have Mexican restaurants as the first most common restaurant, closely folloewd by vietnamese restaurants. Clusters 2 and 4 seem to lag behind in the number of neighborhoods with 4 and 2 neighborhoods respectively.

5.0 Discussion

The above results have far-reaching implications; Cluster 1 with the highest number of neighborhoods with restaurants coincidentally happens to inhabit about 80% of the richest neighborhoods in Austin. Neighborhoods like Downtown with a median household income of $98,49; South River City with a median income of $83,666; Barton Hill with median household income of $72,514 just to mention a few. Locating a restaurant in this cluster will be precarious, not only will the potential investor face stiff competition from the already saturated environment but also such an investor must have deep pockets.

Clusters 3 and 5 provide better promises than cluster 1 for potential investors particularly because they proffer lesser density of neighborhoods with restaurants, though they are also inhabited by high profile neighborhoods such as East Cesar Chavez and South lamar in Cluster 5 with a median household income of $62,881 and $54,142 respectively; Galindo and Central East Austin neighborhoods in Cluster 3 with a median household income of $53,509 and $54,415 respectively. An investor wishing to site a new restaurant in cluster 5 is very likely to face stiff competition if the restaurant type is Mexican or even Vietnamese. The most common analogy in the Cluster data frames is a useful guide in this respect.

Of all the clusters, cluster 3 is best suited for potential investors for several reasons, firstly it consists of neighborhoods of the median household income class who can afford services of restaurants without being too expensive, secondly, the cluster consists of neighborhoods of moderate density that can accommodate new entrants to the industry, and thirdly there is a good blend of restaurant types in the cluster providing an environment that can accommodate almost any restaurant type.

An investor investing in clusters 2 and 4 may experience lower patronage than in the other clusters; however these neighborhoods may be suitable for startups that may not be not be able to weather the storm of competition.

Seasoned investors who are competition savvy and have deep pockets can still make tremendous in-roads in Cluster 1

This project has some limitations particularly in sourcing recent datasets dealing with neighborhoods in Austin, such data as the average cost of housing units per neighborhood, the average rent per neighborhood, the crime rate per neighborhood and demographic data such as number family households per neighborhood, gender distribution, age distribution. Where some of these datasets are available they are not recent and often expressed per zip code rather than per neighborhood names that are readily understandable to the average user.

We wish to recommend that data scientist take the limitations expressed here as a challenge and fill up the gap, we also wish to recommend to organizations engaging in dataset design and development to consider expressing datasets per neighborhood in addition to per zip code

6.0 Conclusion

This project is a veritable tool to guide investors in the restaurant, despite the limitations described in the last paragraph, its efficacy as a guide to potential investors cannot be undermined. This project is a work in progress and we intend to make it more robust in the very near future.

7.0 References

Jimmy Im, These are the 10 best places to live in the US in 2019, [online], Published Mon, Apr 15 20193:34 PM EDT Updated Mon, Apr 29 201912:42 PM

EDT https://www.cnbc.com/2019/04/15/us-news-world-report-best-places-to-live-in-the-us-in-2019.html

Chris Kolmar, Richest Neighborhoods In Austin, TX For 2020, [online], January 12, 2020, https://www.homesnacks.net/richest-neighborhoods-in-austin-128966/