Best Locations for a Restaurant in Quito Ecuador. Based on location data provided by Foursquare API

Introduction/Business Problem

There is a group of stake holders that would like to know in what neighborhoods it would be profitable to locate a restaurant in the city of Quito, Ecuador, based on the existence of other restaurants in the zone. There are some assumptions about some neighborhoods having significantly less quantity of restaurants than other similarly populated neighborhoods. So, in this case, I will use foursqare data to identify restaurant venues by neighborhood and help stake holders decide what neighborhoods would represent a better opportunity for locating a restaurant.

Data

The data I used for this analysis was an excel file that contained the coordinates of each neighborhod of the city of Quito. that I have downloaded from official city sites and use them as input for the forsquare API queries to get the venues of each neighborhood in the city.

Example data source of Quito Neighborhood's Coordinates

	Código	Lugar	Latitud	Longitud
0	170101	Alfaro (Chimbacalle)	-0.23333	-78.51667
1	170102	Benalcazar	-0.18262	-78.48122
2	170103	Cotocollao	-0.11757	-78.49729
3	170104	Chaupicruz (La Concepcion)	-0.15749	-78.48697
4	170105	Chillogallo	-0.31887	-78.57307
5	170106	El Salvador	-0.16563	-78.51045
6	170107	Gonzalez Suarez	-0.22197	-78.51239
7	170108	Guapulo	-0.20474	-78.47835
8	170109	La Floresta	-0.20909	-78.48348
9	170110	La Libertad	-0.27739	-78.57991
10	170111	La Magdalena	-0.24360	-78.52990

Example of the Venues returned by the Query to Foursquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alfaro (Chimbacalle)	-0.23333	-78.51667	Museo Interactivo de Ciencia	-0.236313	-78.516309	Science Museum
1	Alfaro (Chimbacalle)	-0.23333	-78.51667	MIC	-0.231879	-78.515172	Science Museum
2	Alfaro (Chimbacalle)	-0.23333	-78.51667	Mesón de la Recoleta	-0.231333	-78.512786	South American Restaurant
3	Benalcazar	-0.18262	-78.48122	Corfú	-0.181578	-78.481249	Ice Cream Shop
4	Benalcazar	-0.18262	-78.48122	Chicberry	-0.181021	-78.480118	Ice Cream Shop
5	Benalcazar	-0.18262	-78.48122	Europa Café	-0.183643	-78.480667	Coffee Shop
6	Benalcazar	-0.18262	-78.48122	Lucía Pie House & Grill	-0.183717	-78.481240	Pie Shop
7	Benalcazar	-0.18262	-78.48122	Cyril	-0.183930	-78.479529	Dessert Shop
8	Benalcazar	-0.18262	-78.48122	Mister Bagel	-0.181844	-78.479324	Bagel Shop
9	Benalcazar	-0.18262	-78.48122	Cyrano	-0.181446	-78.481109	Bakery
10	Benalcazar	-0.18262	-78.48122	Cosa Nostra Trattoria	-0.184815	-78.480762	Italian Restaurant
11	Benalcazar	-0.18262	-78.48122	Restaurante Ópera	-0.183607	-78.481155	French Restaurant
12	Benalcazar	-0.18262	-78.48122	Di Serggio	-0.181351	-78.480177	Italian Restaurant
13	Benalcazar	-0.18262	-78.48122	Hotel Dann Carlton	-0.183539	-78.480667	Hotel
14	Benalcazar	-0.18262	-78.48122	Jurgen	-0.183540	-78.481715	Bakery

Methodology

Using pandas, I manipulated the data to know exactly how many venues there are in each neighborhood and what type o them they are. After that I generated a data structure in which it is very easy to check what are most common type of venues in each neighborhood.

.

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

```
There are ['Science Museum' 'South American Restaurant' 'Ice Cream Shop'
 'Coffee Shop' 'Pie Shop' 'Dessert Shop' 'Bagel Shop' 'Bakery'
 'Italian Restaurant' 'French Restaurant' 'Hotel' 'Burger Joint'
 'Creperie' 'Peruvian Restaurant' 'Pizza Place' 'Diner' 'Sushi Restaurant'
 'Food Stand' 'Park' 'Japanese Restaurant' 'Pharmacy'
 'Empanada Restaurant' 'Movie Theater' 'Argentinian Restaurant'
 'Multiplex' 'Sandwich Place' 'Middle Eastern Restaurant' 'BBQ Joint'
 'Hotel Bar' 'Gym / Fitness Center' 'Cupcake Shop' 'Gastropub'
 'Snack Place' 'Supermarket' 'Playground' 'Karaoke Bar'
 'Korean Restaurant' 'Fast Food Restaurant' 'Convenience Store'
 'Cajun / Creole Restaurant' 'Butcher' 'Mediterranean Restaurant'
 'Seafood Restaurant' 'Food Truck' 'Grocery Store' 'Restaurant'
 'Mexican Restaurant' 'Bar' 'Asian Restaurant' 'Chinese Restaurant'
 'Sculpture Garden' 'Fried Chicken Joint' 'Latin American Restaurant'
 'Deli / Bodega' 'Pakistani Restaurant' 'Electronics Store'
 'Paintball Field' 'Church' 'Art Gallery' 'Historic Site' 'Plaza'
 'Breakfast Spot' 'History Museum' 'Bed & Breakfast' 'Cafeteria'
 'Shopping Mall' 'Chocolate Shop' 'Café' 'Museum' 'Wine Bar' 'Bus Station'
 'Southern / Soul Food Restaurant' 'Scenic Lookout' 'Wings Joint'
```

```
Quito_venues.groupby('Neighborhood')[["Venue"]].count().sort_values(by=['Venue'])
```

Neighborhood	Venue
La Ferroviaria	1
Eloy Alfaro	1
La Argelia	1
Nayón	1
El Beaterio	1
Nono	1
Ponceano	1
Chilibulo	1
Puembo	1
Belisario Quevedo	1
Amaguaña	1

The next step was to generate clusters to help classify Neighboordos in different groups based on their characteristics, so the first step to do this was to perform "one hot encoding" and generate new data structrures that later served for the clustering algorithm to work correctly.

```
[39]: # one hot encoding
quito_onehot = pd.get_dummies(Quito_venues[['Venue Category']], prefix="", prefi

# add neighborhood column back to dataframe
quito_onehot['Neighborhood'] = Quito_venues['Neighborhood']

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

quito_onehot.shape

[39]: (632, 148)

[48]: quito_grouped = quito_onehot.groupby('Neighborhood').mean().reset_index()
quito_grouped
```

Using machine learning unsupervised algorithm "k-means", I clustered the neighborhoods, to help differentiate which neighborhoods represent the best opportunities for locating a business.

```python
quito_merged = postal_codesdf

# merge quito_grouped with quito_data to add latitude/longitude for each neighborhood
quito_merged = quito_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Lugar')
quito_merged.head() # check the last columns!
quito_merged=quito_merged.dropna().reset_index()

quito_merged['Cluster Labels']=kmeans.labels_
```

Lugar	Latitud	Longitud	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	Cluster Labels
Alfaro (Chimbacalle)	-0.23333	-78.51667	Science Museum	South American Restaurant	Farm	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	Farmers Market	Event Space	3
Benalcazar	-0.18262	-78.48122	Italian Restaurant	Coffee Shop	Bakery	Hotel	Ice Cream Shop	Movie Theater	Japanese Restaurant	Pizza Place	Sandwich Place	Argentinian Restaurant	0
Cotocollao	-0.11757	-78.49729	Burger Joint	Grocery Store	Park	Food Truck	Food Stand	Food Court	Food & Drink Shop	Food	Flea Market	Fast Food Restaurant	0
Chaupicruz (La	[illegible]	[illegible]	Convenience	Fried	[illegible]	Pakistani	Electronics	Coffee	Chinese	[illegible]	Sculpture	Seafood	[illegible]

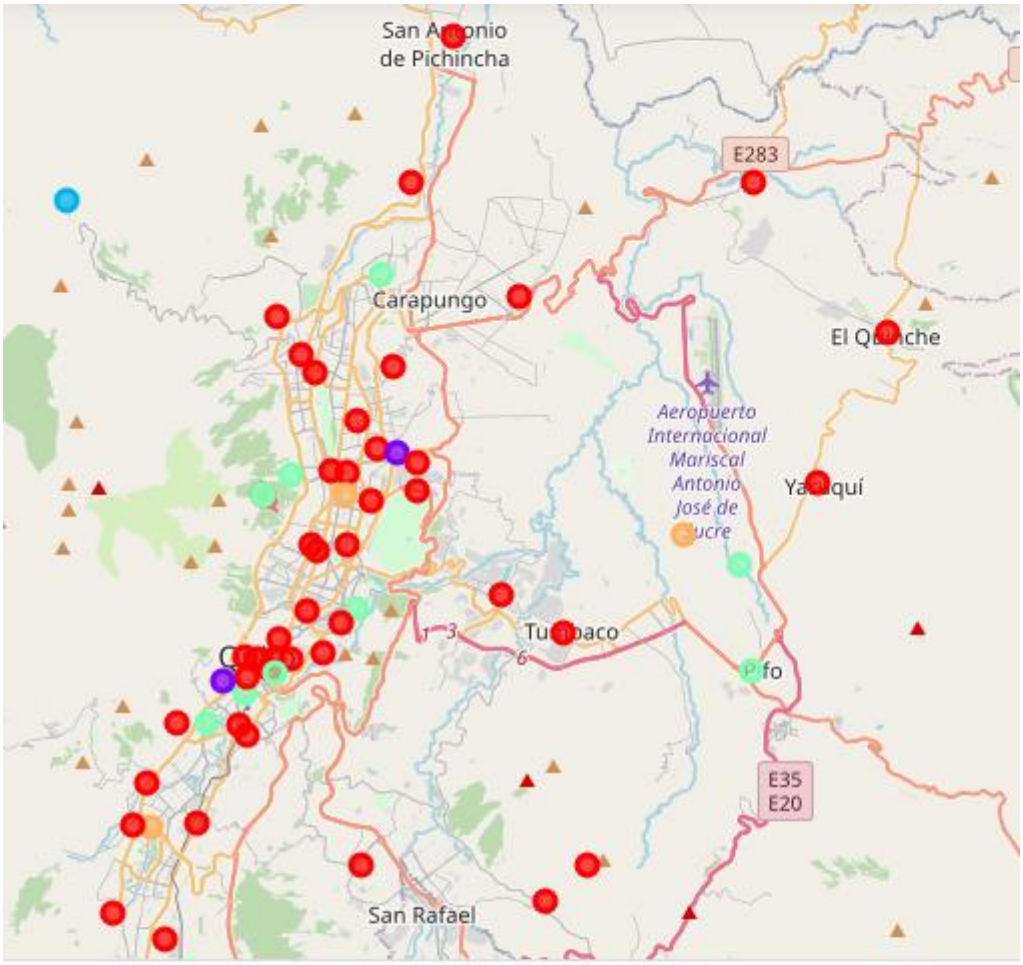
Finally, I used Folium to display the location of each Neighborhood along with the representation of the cluster that they were assigned.

```python
address = 'Quito, Ecuador'
geolocator = Nominatim(user_agent="to_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Quito are {}, {}.'.format(latitude, longitude))

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(quito_merged['Latitud'], quito_merged['Longitud'], quito_merged['Lugar'], quito_merged['Cluster
    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)
```



Results

Based on what I got out of manipulating the data, I have Identified Neighborhoods in which less venues are located, and therefore neighborhoods that represent better opportunities for locating a business.

So neighborhoods like la Ferroviaria, Eloy Alfaro, LA argelia or Ponceano, among others represent a good opporuntiy to locate a restaurant.

This results can also be supported by what I from clustering the data. These neighborhoods where identified with the label "3".

All neighborhoods with this label represent a good opportunity as we can see in the map.

Conclussion

The purpose of this project was to identify Quito with low concentration of restaurants. This was done to help stakeholders to make an informed decision on what location for a new restaurant would represent a better opportunity. By getting venue information of each neighborhood in Quito, Ecuador from Foursquare we have first identified the neighborhoods in which there is less concentration of venues and to justify that analysis further we have performed a cluster analysis to group those neighborhoods that could represent an optimal opportunity to locate a restaurant. We have also represent the result in a map.

Based on this information, decision making by stake holders will be much simplified, and they could be more certain that their investment will be less risky and have better returns.