# Capstone Final Project

Schools and their environment

Schools and enviroment

- Based on a dataset(geojson)of the locations of the schools on my city (Bogotá, Colombia) I decided to cross that information with the foursquare data to know if there are some venues preeminent over others.

- Thanks to foursquare we were able to generate a list of venues from the location (coordinates) of the schools

- With cluster we can see if there is some relation of the near venues with the school location.

- It will be a nice complement crossing the data with the official test of Colombia (Pruebas Saber 11)

Data
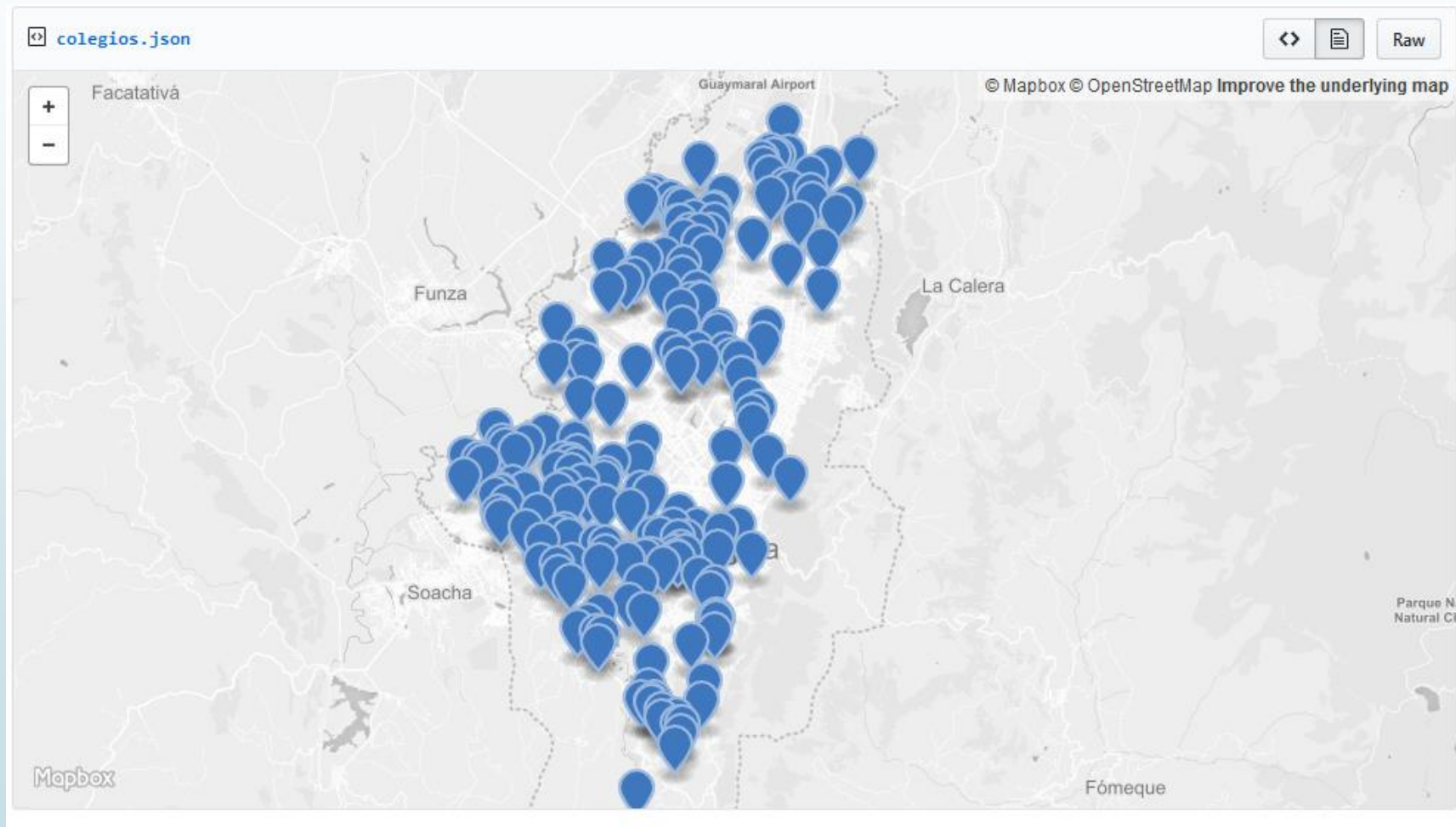
Method

- Transpose the GEOJSON to a new pandas data frame
- Correlate the coordinates with the information retrieved from foursquare
- Visualize the 10 most common venues
- Plot in a map
- Cluster the schools looking for correlation

Import Dataset found here: https://gist.github.com/leonardobareno /ba093f8d2ba5e67a9fc4d17bad860d66"

```python
In [2]:   1  data = pd.read_json('C:\\Users\\David\\Desktop\\IBM\\Capstone\\colegios.json')
```

```python
In [3]:   1  with open('colegios.json') as json_data:
          2      bogota_data = json.load(json_data)
```

```python
In [4]:   1  bogota_data
```

{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.16702222, 4.600930556]},
 'properties': {'title': '(C006) COL COLOMBO JAPONES', 'id': 'C006'}},
{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.08384444, 4.684352778]},
 'properties': {'title': '(C007) COL INST INSCAP SEDE FERIAS',
  'id': 'C007'}},
{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.10326944, 4.582227778]},
 'properties': {'title': '(C008) CENT PANAMERICANO DE CAPACITACION SEDE RESTREPO',
  'id': 'C008'}},
{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.16375278, 4.638091667]},
 'properties': {'title': '(C009) COL CIUDAD PATIO BONITO', 'id': 'C009'}},
{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.06759167, 4.6469]},
 'properties': {'title': '(C010) EXTERN SIMON BOLIVAR', 'id': 'C010'}},
{'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [-74.12639444, 4.708158333]},
 'properties': {'title': '(C011) COL DE FORMACION INTEGRAL VIRGEN DE LA PEÃ'A',

Extract the features into another dataset

```
In [5]:    1  neighborhoods_data = bogota_data['features']
```

```
In [6]:    1  # define the dataframe columns
           2  column_names = ['Title', 'id', 'Latitude', 'Longitude']
           3
           4  # instantiate the dataframe
           5  neighborhood = pd.DataFrame(columns=column_names)
```

```
In [8]:    1  for data in neighborhoods_data:
           2      neighborhood_name = data['properties']['title']
           3      neighborhood_id = data['properties']['id']
           4      neighborhood_latlon = data['geometry']['coordinates']
           5      neighborhood_lat = neighborhood_latlon[0]
           6      neighborhood_lon = neighborhood_latlon[1]
           7
           8      neighborhood = neighborhood.append({'Title': neighborhood_name,
           9                                          'id':neighborhood_id ,
          10                                          'Latitude': neighborhood_lon,
          11                                          'Longitude': neighborhood_lat}, ignore_index=True)
```

```
In [9]:    1  neighborhood.head()
```

Out[9]:

	Title	id	Latitude	Longitude
0	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	C001	4.748233	-74.058667
1	(C002) CENT EDUC PARA JOVENES Y ADULTOS LEPANTO	C002	4.571531	-74.092625
2	(C003) NUEVO LIC GRANADINO	C003	4.605258	-74.182075
3	(C004) COL BRITANICO SEDE KENNEDY	C004	4.623572	-74.152997
4	(C005) CENT EDUC WILFREDO LEHNER	C005	4.598536	-74.184717

Add markets of the schools to the map

```
In [91]:   1  map_bogota
```

Out[91]:
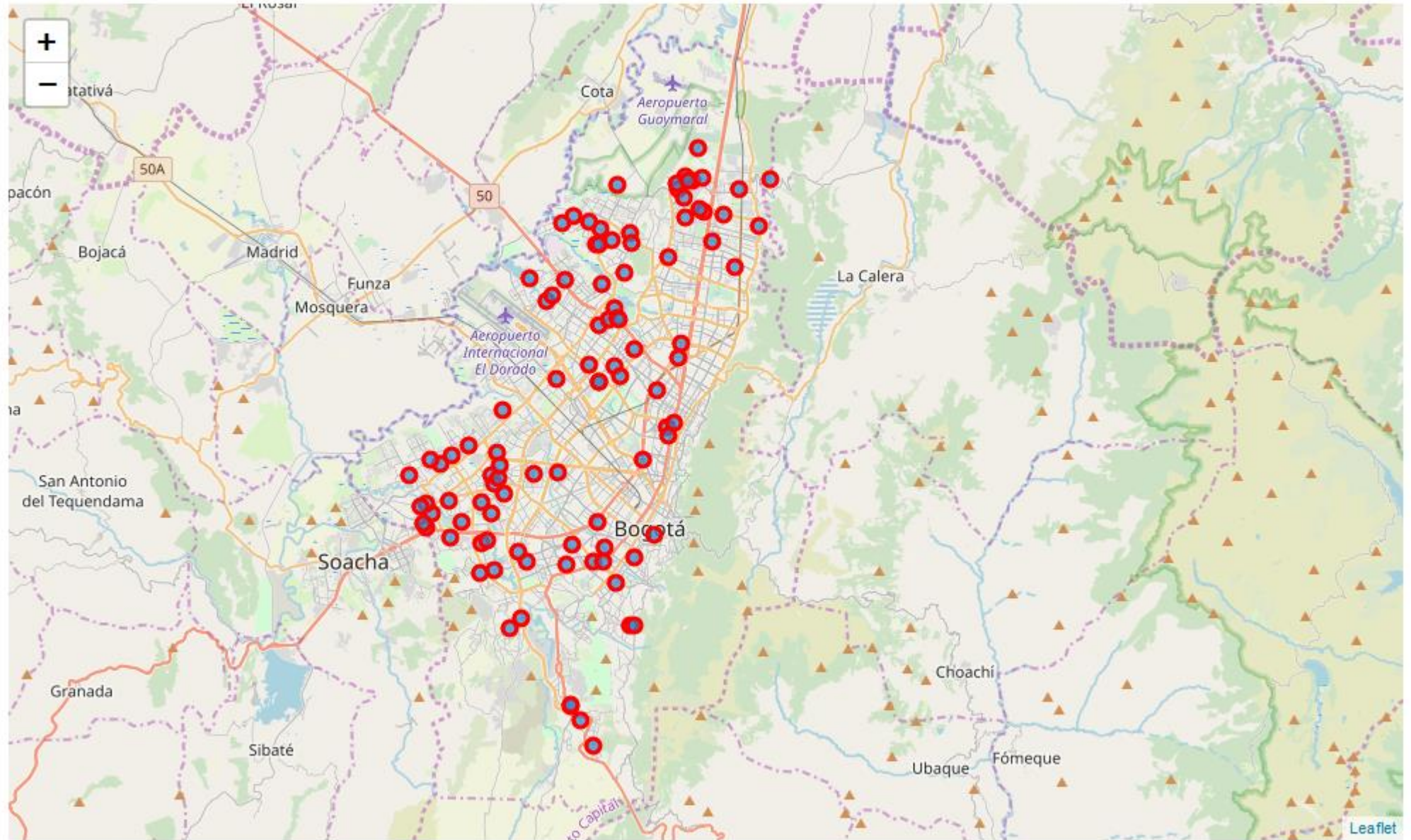
Foursquare Data

Load Foursquare credentials

```
In [19]:  1  CLIENT_ID = 'ITAKD2F1QR5R1ATL5OUFDSZQLV5RDZ5SGW32KWKR3D5ZS4BQ' # your Foursquare ID
          2  CLIENT_SECRET = 'J2B1PEJVSY5SA00LPM4MELOLTHECMYXQAS1LITXL3FIMR22T' # your Foursquare Secret
          3  VERSION = '20180605' # Foursquare API version
          4
          5  print('Your credentails:')
          6  print('CLIENT_ID: ' + CLIENT_ID)
          7  print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: ITAKD2F1QR5R1ATL5OUFDSZQLV5RDZ5SGW32KWKR3D5ZS4BQ
CLIENT_SECRET:J2B1PEJVSY5SA00LPM4MELOLTHECMYXQAS1LITXL3FIMR22T
```

```
In [21]:  1  neighborhood_latitude = df.loc[0, 'Latitude'] # neighborhood latitude value
          2  neighborhood_longitude = df.loc[0, 'Longitude'] # neighborhood longitude value
          3
          4   # neighborhood name
          5
          6  print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
          7                                                                 neighborhood_latitude,
          8                                                                 neighborhood_longitude))
```

```
Latitude and longitude values of (C200) CENT EDUC DIST MANUEL ELKIN PATARROYO are 4.748233333, -74.05866667.
```

```
In [22]:  1  LIMIT = 100 # limit of number of venues returned by Foursquare API
          2  radius = 500 # define radius
          3
          4  url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.form
          5      CLIENT_ID,
          6      CLIENT_SECRET,
          7      VERSION,
          8      latitude,
          9      longitude,
         10      radius,
         11      LIMIT)
         12  url # display URL
```

```
Out[22]:  'https://api.foursquare.com/v2/venues/explore?&client_id=ITAKD2F1QR5R1ATL5OUFDSZQLV5RDZ5SGW32KWKR3D5ZS4BQ&client_secret=J2B
          1PEJVSY5SA00LPM4MELOLTHECMYXQAS1LITXL3FIMR22T&v=20180605&ll=4.59808,-74.0760439&radius=500&limit=100'
```

Retrieve results from foursquare

	name	categories	lat	lng
0	Centro Cultural Gabriel García Márquez	Cultural Center	4.597694	-74.074285
1	La Puerta de la Catedral	Latin American Restaurant	4.597848	-74.074611
2	La Puerta Falsa	Café	4.597977	-74.074878
3	Teatro Colón	Theater	4.596540	-74.074538
4	Juan Valdez Café	Coffee Shop	4.597782	-74.074299

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	4.748233	-74.058667	Leonardelos	4.750154	-74.061461	Pizza Place
1	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	4.748233	-74.058667	La Gallina Tropical	4.745729	-74.056596	Restaurant
2	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	4.748233	-74.058667	Tiendas D1	4.749758	-74.059247	Convenience Store
3	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	4.748233	-74.058667	Surtifruver	4.749473	-74.057417	Fruit & Vegetable Store
4	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	4.748233	-74.058667	Juan Valdez Unicentro	4.744055	-74.059476	Breakfast Spot

Locate Near Venues and sort

----(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL----

	venue	freq
0	Restaurant	0.12
1	Convenience Store	0.12
2	Breakfast Spot	0.12
3	Fast Food Restaurant	0.12
4	Food Truck	0.12

----(C002) CENT EDUC PARA JOVENES Y ADULTOS LEPANTO----

	venue	freq
0	Park	0.25
1	Shopping Mall	0.25
2	Bar	0.25
3	BBQ Joint	0.25
4	Mountain	0.00

----(C003) NUEVO LIC GRANADINO----

	venue	freq
0	Movie Theater	0.25
1	Shopping Mall	0.25
2	Mexican Restaurant	0.25
3	Restaurant	0.25
4	Mountain	0.00

----(C004) COL BRITANICO SEDE KENNEDY----

	venue	freq
0	Pizza Place	0.29
1	Department Store	0.14
2	Restaurant	0.14
3	Fried Chicken Joint	0.14
4	Latin American Restaurant	0.14

----(C005) CENT EDUC WILFREDO LEHNER----

	venue	freq
0	Mexican Restaurant	0.5
1	Auto Garage	0.5
2	Women's Store	0.0
3	Movie Theater	0.0
4	Men's Store	0.0

----(C006) COL COLOMBO JAPONES----

	venue	freq
0	Supermarket	0.33
1	Burger Joint	0.33
2	BBQ Joint	0.33
3	Movie Theater	0.00
4	Men's Store	0.00

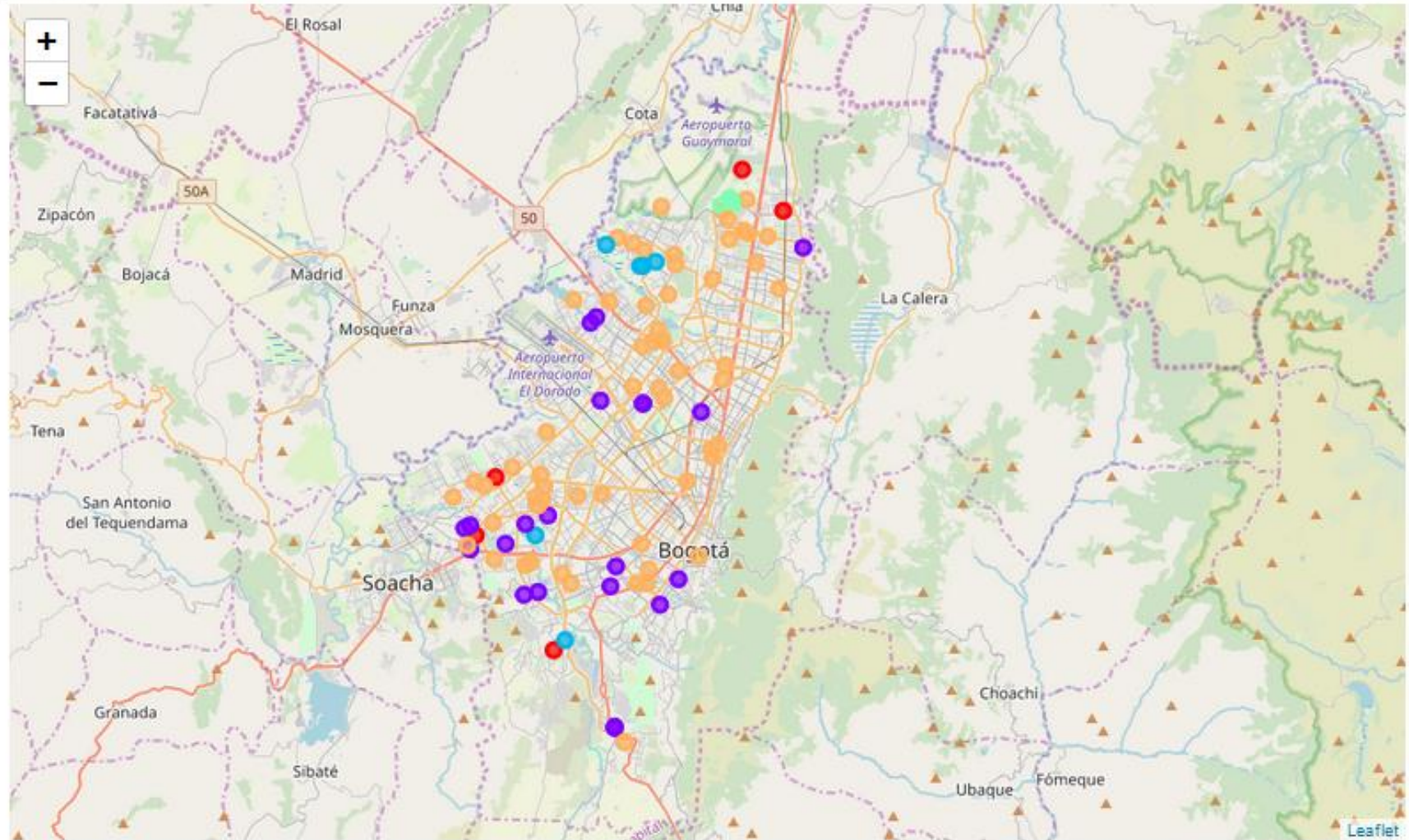|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	Fruit & Vegetable Store	Gym	Fast Food Restaurant	Breakfast Spot	Convenience Store	Pizza Place	Restaurant	Food Truck	Dance Studio	Diner
1	(C002) CENT EDUC PARA JOVENES Y ADULTOS LEPANTO	Park	Bar	BBQ Joint	Shopping Mall	Deli / Bodega	Ethiopian Restaurant	Electronics Store	Drugstore	Donut Shop	Diner
2	(C003) NUEVO LIC GRANADINO	Shopping Mall	Mexican Restaurant	Movie Theater	Restaurant	Dance Studio	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
3	(C004) COL BRITANICO SEDE KENNEDY	Pizza Place	Burger Joint	Fried Chicken Joint	Latin American Restaurant	Department Store	Restaurant	Donut Shop	Diner	Dessert Shop	Deli / Bodega
4	(C005) CENT EDUC WILFREDO LEHNER	Mexican Restaurant	Auto Garage	Wings Joint	Deli / Bodega	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store

Visualize first 20 Schools

Neighborhood and their Venue Category
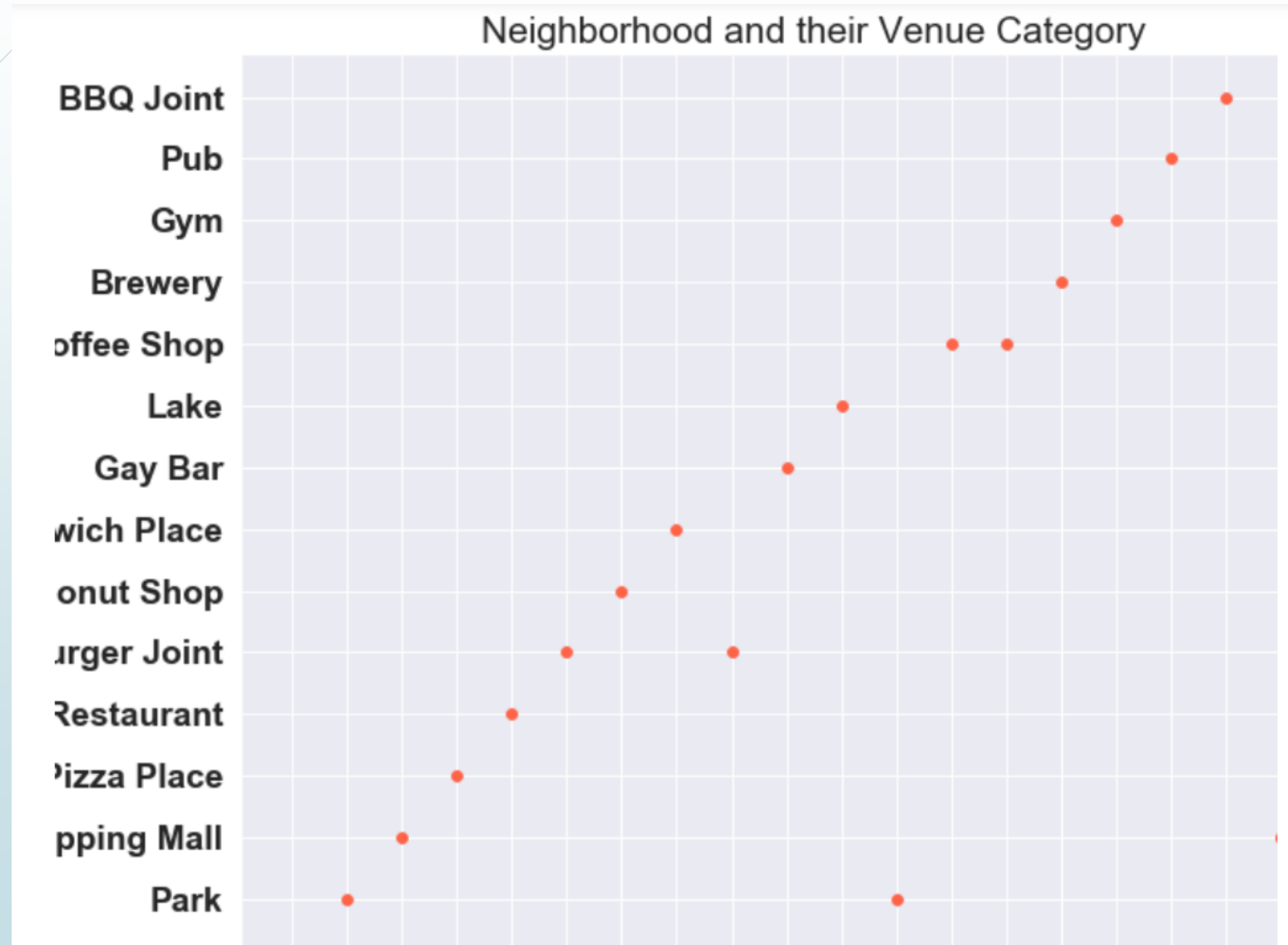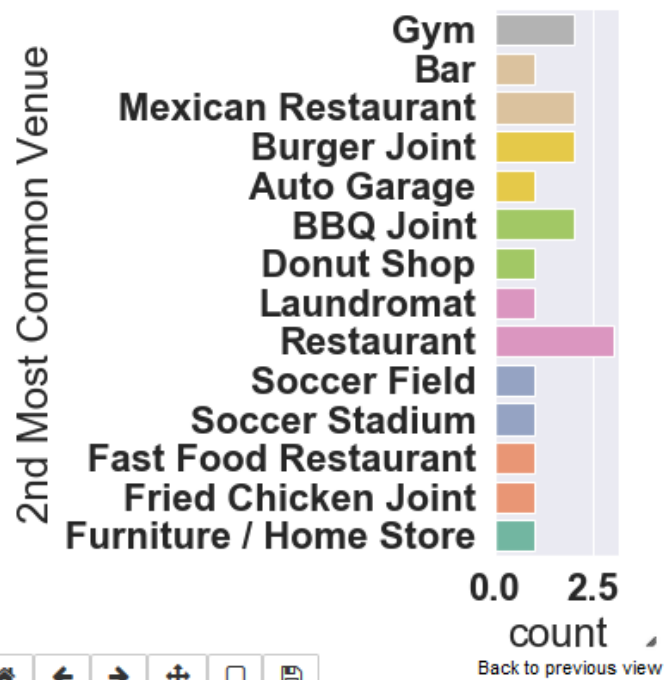
Cluster Organization

Cluster Distribution

Cluster 1

```
In [106]:   1  cluster1 = dff.loc[dff['Cluster Labels'] == 0, dff.columns[[0] + list(range(5, dff.shape[1]))]]
            2  cluster1
```

Out[106]:

	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
2	(C003) NUEVO LIC GRANADINO	Shopping Mall	Mexican Restaurant	Movie Theater	Restaurant	Dance Studio	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
18	(C019) COLEGIO ERNESTO GUHL - SEDE PRINCIPAL	Shopping Mall	Restaurant	Burger Joint	Dance Studio	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
44	(C045) NUEVO COL LUSADI	Restaurant	Wings Joint	Dance Studio	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
62	(C063) COL MAYOR JOSE CELESTINO MUTIS	Mexican Restaurant	Wings Joint	Dance Studio	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
74	(C075) COL DE LA ENSEÃ'ANZA	Basketball Court	Restaurant	Wings Joint	Fast Food Restaurant	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store

Cluster 2

```
In [115]:   1  cluster2 = dff.loc[dff['Cluster Labels'] == 1, dff.columns[[0] + list(range(5, dff.shape[1]))]]
            2
```

```
In [116]:   1  cluster2.shape
```
Out[116]: (21, 11)

```
In [117]:   1  cluster2
```
Out[117]:

	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
1	(C002) CENT EDUC PARA JOVENES Y ADULTOS LEPANTO	Park	Bar	BBQ Joint	Shopping Mall	Deli / Bodega	Ethiopian Restaurant	Electronics Store	Drugstore	Donut Shop	Diner
4	(C005) CENT EDUC WILFREDO LEHNER	Mexican Restaurant	Auto Garage	Wings Joint	Deli / Bodega	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
5	(C006) COL COLOMBO JAPONES	Burger Joint	BBQ Joint	Supermarket	Dance Studio	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
10	(C011) COL DE FORMACION INTEGRAL VIRGEN DE LA ...	Lake	BBQ Joint	Mexican Restaurant	Supermarket	Fast Food Restaurant	Wings Joint	Department Store	Electronics Store	Drugstore	Donut Shop
11	(C012) COL CLARETIANO EL LIBERTADOR - SEDE PR...	Park	Mexican Restaurant	Fast Food Restaurant	Notary	College Bookstore	Comfort Food Restaurant	Drugstore	Donut Shop	Diner	Dessert Shop
17	(C018) INST ACAD BETHEL	BBQ Joint	Furniture / Home Store	Wings Joint	Fish & Chips Shop	Ethiopian Restaurant	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop

Cluster 3

```
1  dff.loc[dff['Cluster Labels'] == 2, dff.columns[[0] + list(range(5, dff.shape[1]))]]
```

	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
45	(C046) COL POPULAR BOLIVARIANO	Soccer Field	Dance Studio	Ethiopian Restaurant	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
53	(C054) INST CULT RAFAEL MAYA	Park	Ice Cream Shop	Dance Studio	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
54	(C055) LIC SIGLO XXI	Park	Soccer Field	Bar	Pizza Place	Cupcake Shop	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
57	(C058) INST COPESAL	Park	Soccer Field	Auto Garage	Fast Food Restaurant	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
73	(C074) CENT DE INTEGRACION EDUCACION DEL NORTE...	Park	Tea Room	Rental Car Location	Cupcake Shop	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
86	(C087) INSTITUTO COMERCIAL LORETO - SEDE PRINC...	Park	Pizza Place	Fast Food Restaurant	Grocery Store	Greek Restaurant	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega

Cluster 4

```
In [87]:  1  dff.loc[dff['Cluster Labels'] == 3, dff.columns[[0] + list(range(5, dff.shape[1]))]]
```

Out[87]:

	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
64	(C065) COL STELLA MATUTINA	Tennis Court	Hot Dog Joint	Wings Joint	Cupcake Shop	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store
75	(C076) INST ALBERTO MERANI	Tennis Court	Wings Joint	Cupcake Shop	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
83	(C084) COL SANTIAGO MAYOR	Tennis Court	Wings Joint	Cupcake Shop	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store	Deli / Bodega
92	(C093) COL EMILIO VALENZUELA - SEDE PRINCIPAL	Tennis Court	Hot Dog Joint	Wings Joint	Cupcake Shop	Electronics Store	Drugstore	Donut Shop	Diner	Dessert Shop	Department Store

Cluster 5

```
In [111]:  1  dfff  =dff.loc[dff['Cluster Labels'] == 4, dff.columns[[0] + list(range(5, dff.shape[1]))]]
```

```
In [113]:  1  dfff.shape
```

Out[113]: (60, 11)

```
In [114]:  1  dfff
```

Out[114]:

	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	(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL	Fruit & Vegetable Store	Gym	Fast Food Restaurant	Breakfast Spot	Convenience Store	Pizza Place	Restaurant	Food Truck	Dance Studio	Diner
3	(C004) COL BRITANICO SEDE KENNEDY	Pizza Place	Burger Joint	Fried Chicken Joint	Latin American Restaurant	Department Store	Restaurant	Donut Shop	Diner	Dessert Shop	Deli / Bodega
6	(C007) COL INST INSCAP SEDE FERIAS	Donut Shop	Burger Joint	Shopping Mall	Creperie	Fast Food Restaurant	Soccer Stadium	Mexican Restaurant	Food Court	Supermarket	Sandwich Place
7	(C008) CENT PANAMERICANO DE CAPACITACION SEDE	Sandwich Place	Donut Shop	Convenience Store	Drugstore	Bowling Alley	Soccer Field	Steakhouse	Pizza Place	Miscellaneous Shop	Seafood Restaurant