Capstone Project 2020

23 MARZO

Schools and enviroment
**Creado por: David Charif**

Schools and enviroment

Which are the most common venues near schools?

The relationship between the enviroment and the schools could affect the students, but since we are not looking for GPA of each one, we are going to look how the venues are distributed in 100 schools, to see if some share some values and with Kcluster we are going to plot those results in a map.

Area of interest
-Parents who would like to know what venues are near to the school in order to select better

-Cluster the schools for insings in venues
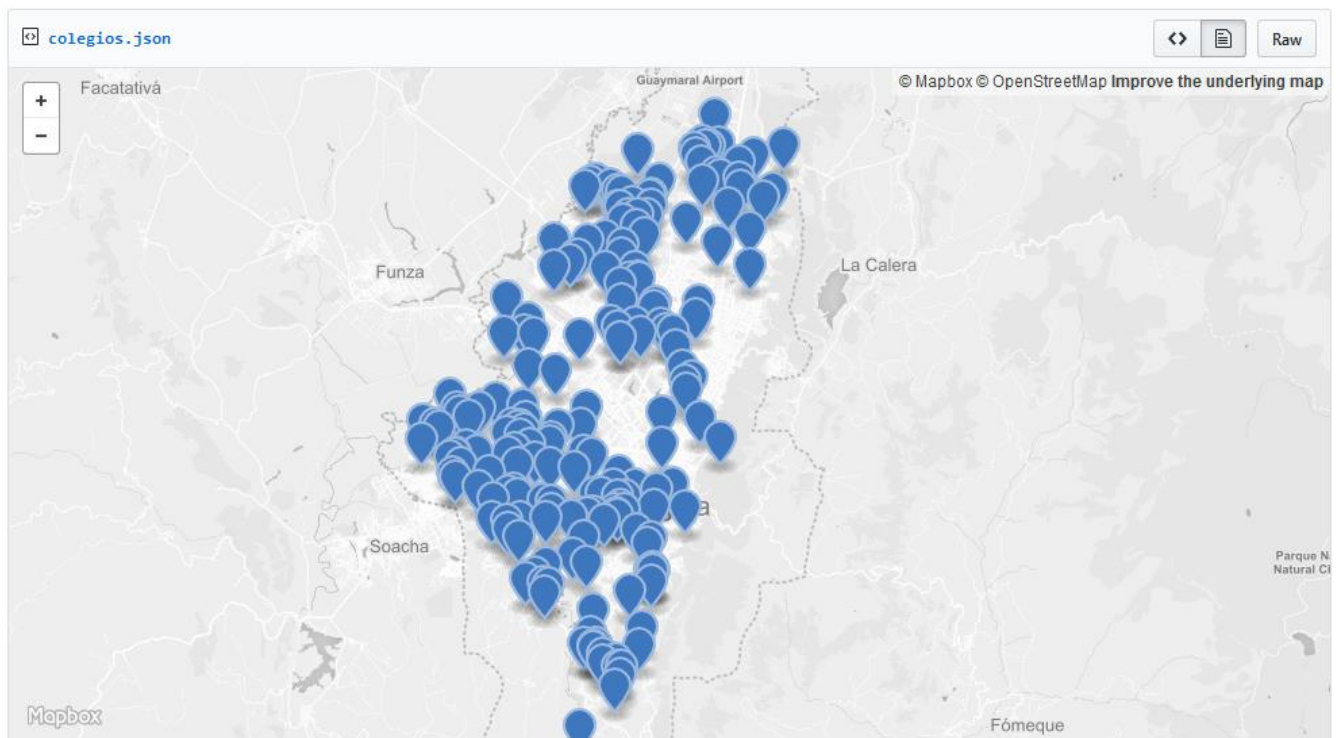
Since the data set has 200 schools we are going to limit the result to 100 schools.

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

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

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

```
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',
```

Add markets of the schools to the map

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

Out[91]:
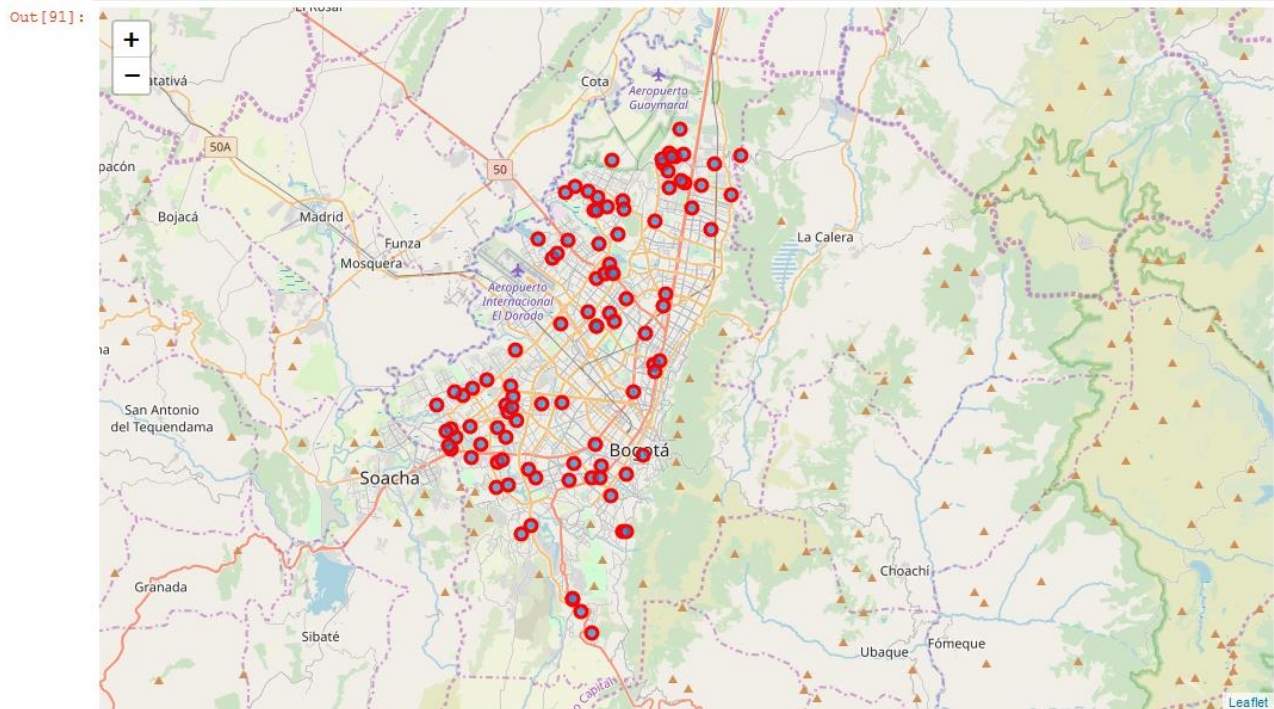
```
----(C001) COL COLOMBO BRIGHTON - SEDE PRINCIPAL----        -(C002) CENT EDUC PARA JOVENES Y ADULTOS L
                   venue  freq                                          venue  freq
0              Restaurant  0.12                             Park   0.25
1       Convenience Store  0.12                    Shopping Mall   0.25
2          Breakfast Spot  0.12                              Bar   0.25
3   Fast Food Restaurant   0.12                        BBQ Joint   0.25
4              Food Truck  0.12                         Mountain   0.00


      ----(C003) NUEVO LIC GRANADINO----             ----(C004) COL BRITANICO SEDE KENNEDY----
                   venue  freq                                          venue  freq
0           Movie Theater  0.25                    0              Pizza Place  0.29
1           Shopping Mall  0.25                    1        Department Store   0.14
2       Mexican Restaurant 0.25                    2               Restaurant  0.14
3              Restaurant  0.25                    3       Fried Chicken Joint 0.14
4                Mountain  0.00                    4   Latin American Restaurant 0.14


  ----(C005) CENT EDUC WILFREDO LEHNER----              ----(C006) COL COLOMBO JAPONES----
                   venue  freq                                          venue  freq
0   Mexican Restaurant   0.5                     0        Supermarket   0.33
1          Auto Garage   0.5                     1       Burger Joint   0.33
2        Women's Store   0.0                     2          BBQ Joint   0.33
3        Movie Theater   0.0                     3      Movie Theater   0.00
4          Men's Store   0.0                     4         Men's Store  0.00
```

Looking this preliminar data, show us how differents are the first 6 values, but we can notice how the food its always near.

	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

Grouping the results from 30 schools we can saw more detailed data
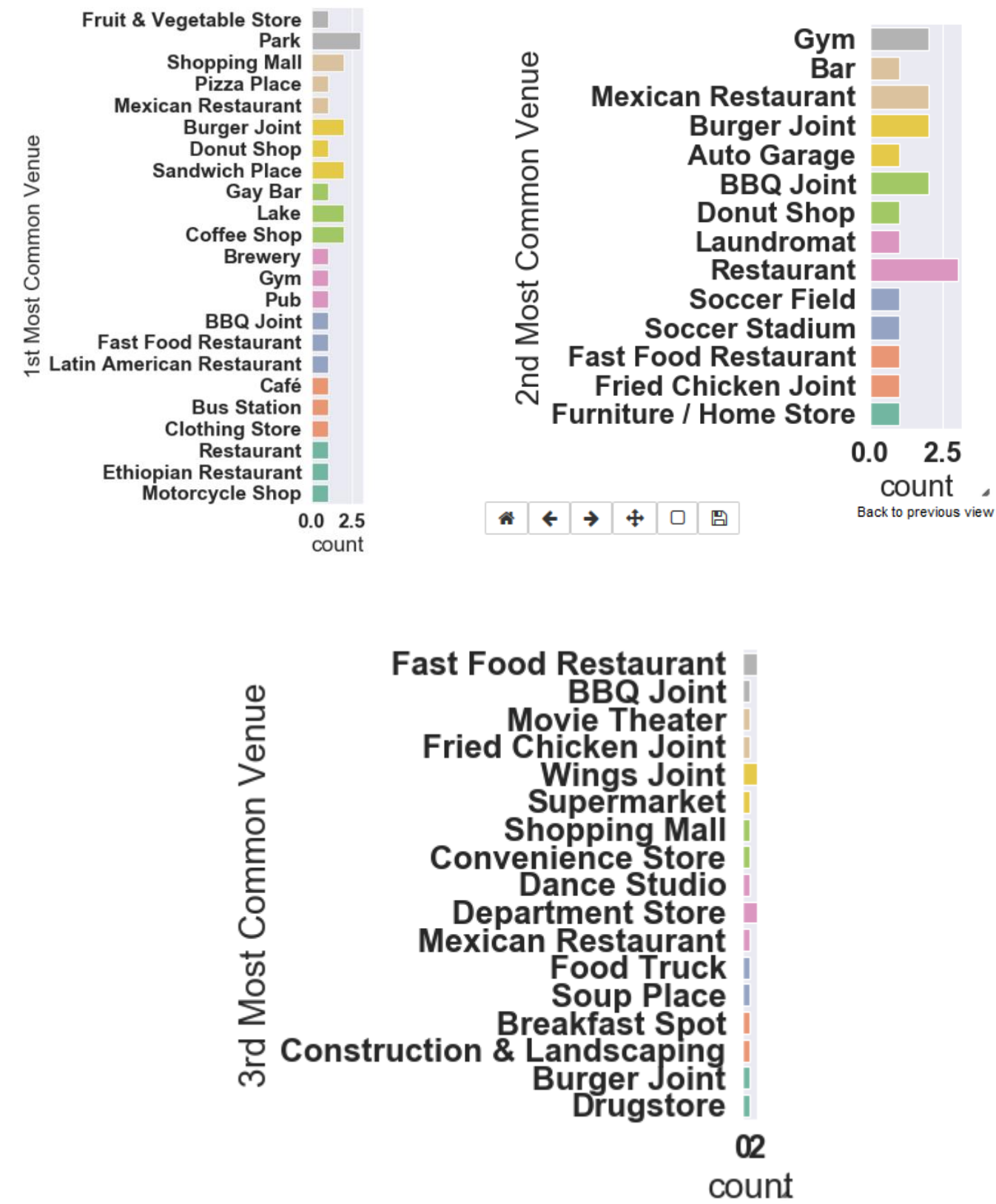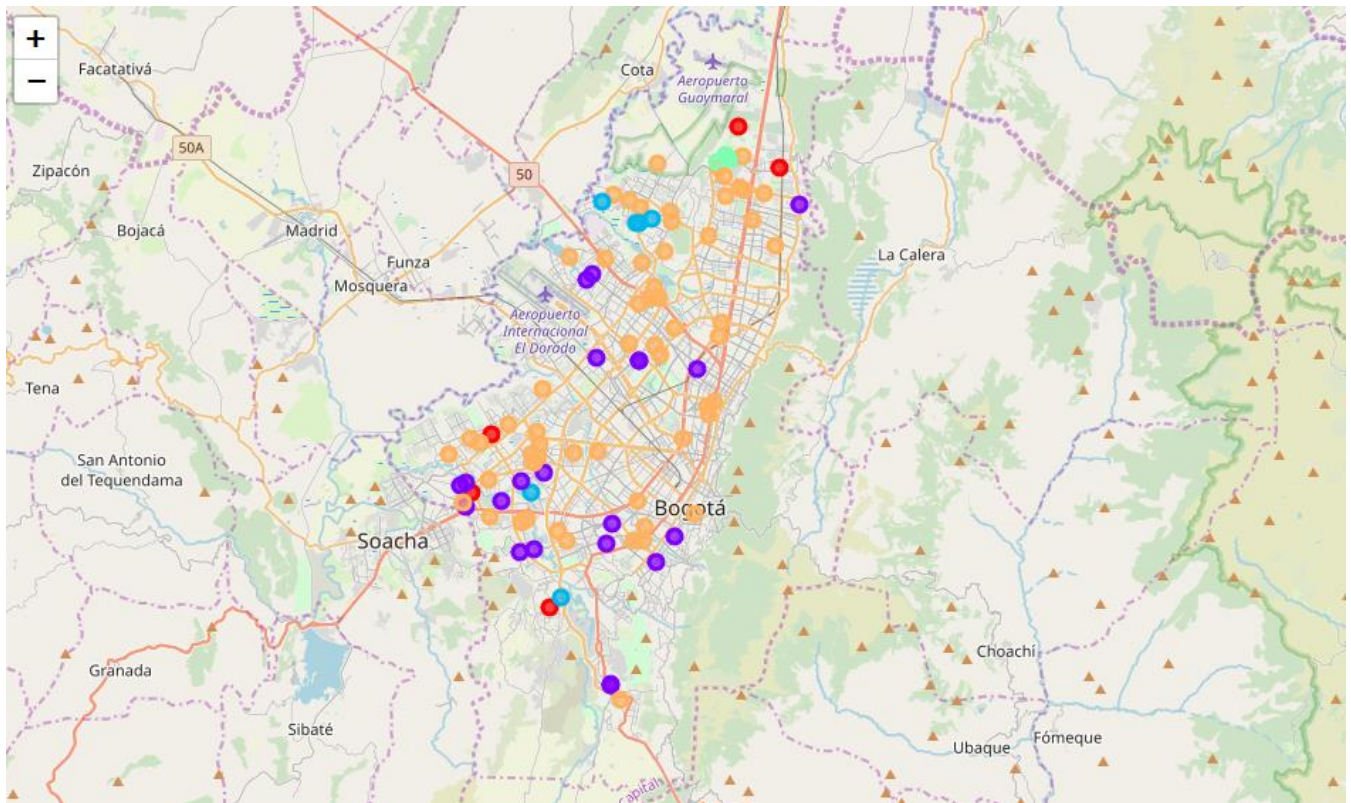
Ploting the results we can notate that the orange cluster (5) has the mayority of schools, along with the purple cluster (2)

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

6

```
1  # 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

The cluster 5 groups the food related venues, this means that the schools are located near a commercial area, not very quiet, the cluster 2 represent new venues like Electrics store, Notary, HomeStore this cluster is even more commercial since has more variety inside.

The blue dots cluster(3) represent a more ideal area to study, even when is mixed with local bussiness the park is the most common venue.

Cluster 3

```
In [86]:  1  dff.loc[dff['Cluster Labels'] == 2, dff.columns[[0] + list(range(5, dff.shape[1]))]]
Out[86]:
```

	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

The red dots cluster(1) one of the littles with only 5 schools are related with big commerce, Shopping malls, dance studio and departament store, which could indicate the delevopment near the schools

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

The green dots cluster(4) represent the most expensive schools in relation with the other ones, since the four schools are near each other all share a Tennis Court, in this cluster we can not say that is a local commercial area

```
1  # 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

Conclusion

The near venues of the schools are related with the neighborhood and with the bussiness around, some areas are residential, which could be more peaceful for the students. The cluster organization groups the schools according to the related area, while we could expect that the parks were directly related with schools we can see that only 6 schools of 100 have one near.

We cannot negate that in bogota, a lot of schools are in a commertial area, exposing the students to not recommendable venues, like bars and fast food.