**Do near venues affect students’ performance?**

Pedro Chacón

Applied Data Science Capstone

**Introduction**

Between Colombian parents there is the conception that schools nearby bars, clubs, stores and other recreational places are worst for their children, because it increasing the distractions and the probability of their children to skip classes to go to those places.

That’s why in the present project we will look if there is really a relationship the kind of establishments near schools in Bogotá, Colombia, and the results of its students in the national standardized academic test, ICFES, similar to the American SAT, which measure the competence of students in different academic areas.

**Data**

In order to archive our objective, we need 3 datasets: schools location, score average by school and venues near every schools.

**Schools location**

We will use the dataset of public and private schools from the government page *datos.gov.co* from the next page:

<https://www.datos.gov.co/Educaci-n/LISTADO-COLEGIOS-BOGOTA/qijw-htwa>

This dataset has the follow columns (The dataset is in Spanish):

* calendario
* sector
* nombreestablecimiento
* telefono
* nombre\_Rector
* correo\_Electronico
* direccion
* secretaria
* zona
* tipo\_Establecimiento
* etnias
* genero
* niveles
* jornadas
* caracter
* especialidad
* grados
* modelos\_Educativos
* capacidades\_Excepcionales
* discapacidades
* idiomas
* numero\_de\_Sedes
* estado
* prestador\_de\_Servicio
* propiedad\_Planta\_Fisica
* resguardo
* matricula\_Contratada
* internado
* estrato\_Socio\_Economico
* año

From those columns we only have to care about two, “nombreestablecimiento“ which means “Establishment name” and “direccion” which means “address”. Below is shown an example of this two columns’ information:

|  |  |
| --- | --- |
| nombreestablecimiento | direccion |
| INST PEDAG NACIONAL | AC 127 11 20 |
| COLEGIO CRISTO REY BOGOTA | AC 138 58 D 50 |

As you can see, in Bogotá addresses are given in number, therefore is relative easy to translate it to coordinates.

**ICFES score average by school**

Again from datos.gov.co we get the last ICFES scores dataset from the following page:

<https://www.datos.gov.co/Educaci-n/Saber-11-2018-2/m2nt-jw2h>

This is a rather large dataset with 83 columns and one row for each student that do the test, so let’s just look at the columns that we need to consider:

* COLE\_NOMBRE\_ESTABLECIMIENTO
* PUNT\_GLOBAL

The first row “COLE\_NOMBRE\_ESTABLECIMIENTO” means “Establishment name” and the second one means “Global score”, this is because the test is divide in several areas, so the global score is the final result of all of them.

Bellow you can see an example of these two columns.

|  |  |
| --- | --- |
| COLE\_NOMBRE\_ESTABLECIMIENTO | PUNT\_GLOBAL |
| INSTITUCION EDUCATIVA DISTRITAL PARA EL DESARROLLO HUMANO MARIA CANO | 305 |
| IE ANTONIO DE LA TORRE Y MIRANDA | 242 |

**Venues near every schools**

To get the venues near a school we would use the foursquare API, in which we send the location and the search radius, and it return us information about all the venues near that point. Bellow you can see an example of this API that may be found in the page https://developer.foursquare.com/docs/api/venues/search

{ "meta": { "code": 200, "requestId": "5ac51d7e6a607143d811cecb" }, "response": { "venues": [ { "id": "5642aef9498e51025cf4a7a5", "name": "Mr. Purple", "location": { "address": "180 Orchard St", "crossStreet": "btwn Houston & Stanton St", "lat": 40.72173744277209, "lng": -73.98800687282996, "labeledLatLngs": [ { "label": "display", "lat": 40.72173744277209, "lng": -73.98800687282996 } ], "distance": 8, "postalCode": "10002", "cc": "US", "city": "New York", "state": "NY", "country": "United States", "formattedAddress": [ "180 Orchard St (btwn Houston & Stanton St)", "New York, NY 10002", "United States" ] }, "categories": [ { "id": "4bf58dd8d48988d1d5941735", "name": "Hotel Bar", "pluralName": "Hotel Bars", "shortName": "Hotel Bar", "icon": { "prefix": "https://ss3.4sqi.net/img/categories\_v2/travel/hotel\_bar\_", "suffix": ".png" }, "primary": **true** } ], "venuePage": { "id": "150747252" } } ] } }

The principal attribute for our project is the “categories” which give us the category of the venue.