

Capstone Project

The battle of the neighborhoods

1.Introduction

Problem Background

Public Schools in Madrid had experimented a big increase of people demanding running activities for their children out of school premises. Madrid City Hall is creating a new project to cover those requests and, in order to better aim the activities to be done, Madrid City Hall asked for a way to group Public Schools taking into account the kind of activities that could take place near them since creating a program per school is not feasible due lack of resources.

Success Criteria

Create groups of Public Schools attending to the close amusement venues.

2.Data

Madrid City Public School Data

- Source: Madrid City Hall - <https://datos.madrid.es/egob/catalogo/202311-0-colegios-publicos.csv>
- Data of interest: School name and coordinates
- Data overview:
-

PK	NOMBRE	DESCRIPCION-ENTIDAD	HORARIO	EQUIPAMIENTO	TRANSPORTE	DESCRIPCION	ACCESIBILIDAD	CONTENT-URL	NOMBRE-VIA	...	DISTRITO	COORDENADA-X	COORDENADA-Y	LATITUD	LONGITUD	TELEFONO	FAX	EMAIL	TIPO	Unnamed: 39
0	176950	Colegio Público Adolfo Suárez	NaN	NaN	Comedor - Horario ampliado - Planes de mejora.	BUS: 174	ENSEÑANZA: Segundo ciclo de E. Infantil - Pri...	0	http://www.madrid.es/sites/v/index.jsp?vgnext...	...	INFANTA CATALINA MICHELA	444607	4482402	40.488603	-3.652554	917 502 058 / 676 562 782	917 502 058	NaN	/contenido/entidades/organismos/CentrosEduca...	
1	5075	Colegio Público Agustina Díaz	NaN	NaN	Comedor - Horario ampliado	METRO: Alto del Arenal BUS: 54, 58, 103, 143	ENSEÑANZA: Educación Infantil - Primaria	0	http://www.madrid.es/sites/v/index.jsp?vgnext...	...	VESUBIO	445176	4471253	40.389191	-3.647237	917 770 073	917 770 073	NaN	/contenido/entidades/organismos/CentrosEduca...	
2	469338	Colegio Público Agustín Rodríguez Saragat	NaN	NaN	Horario ampliado - Comedor - Planes de mejora	BUS: 142	ENSEÑANZA: Educación Infantil - Primaria - Ens...	0	http://www.madrid.es/sites/v/index.jsp?vgnext...	...	LA GRAJAL DE SAN ILDEFONSO	447288	4469000	40.368040	-3.622168	913 324 941 / 348 475	913 324 941	NaN	/contenido/entidades/organismos/CentrosEduca...	
3	5076	Colegio Público Alberto Alcocer	NaN	NaN	Comedor - Transporte - Horario ampliado - Plan...	METRO: Las Musas BUS: 38, 48, 140, 153	ENSEÑANZA: Educación Infantil - Primaria - Ens...	0	http://www.madrid.es/sites/v/index.jsp?vgnext...	...	CANILLEJAS A VICALVARO	448087	4476035	40.433338	-3.612052	913 202 200 / 446	913 202 446	NaN	/contenido/entidades/organismos/CentrosEduca...	
4	5077	Colegio Público Alcalde de Morales	NaN	NaN	Transporte - Comedor - Horario ampliado - Plan...	BUS: 34, 39, 139, 117, 17, 18	ENSEÑANZA: Educación Infantil - Primaria	0	http://www.madrid.es/sites/v/index.jsp?vgnext...	...	ALDEANUEVA DE LA VERA	434201	4470414	40.379837	-3.776448	917 055 465 / 095 129	917 055 129	NaN	/contenido/entidades/organismos/CentrosEduca...	

Foursquare Venues Data

- Source: Foursquare API - <https://foursquare.com/>
- Data of interest: Near venues, category and coordinates
- Data Overview:

Venue	Venue Latitude	Venue Longitude	Venue Category
Parque Sanchinarro	40.485844	-3.656469	Park
Sanchinarro	40.491610	-3.656731	Neighborhood
Valdefuentes	40.488199	-3.651074	Neighborhood
Parque Infantil Pirata Pata de Lata	40.490574	-3.649876	Park
Centro Cultural Sanchinarro	40.490163	-3.655297	Cultural Center

3.Methodology

Analytic Approach

These are the steps to create the groups of Public Schools:

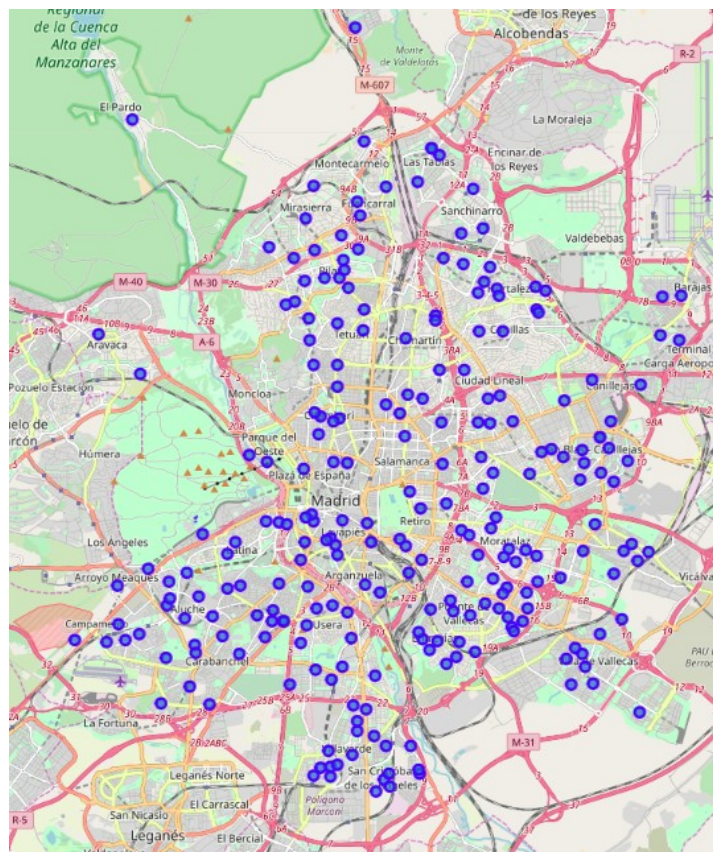
- Gather and clean Schools data
- Select coordinates per school and look for close venues of certain categories
- Create a list of the most common kind of venues of those categories per school
- Create clusters of schools attending to most common venues
- Review information about generated clusters to better understand the schools categories

4.Exploratory Data Analysis

Madrid City Public School Data

After some data cleaning and filtering, it can be seen how the different Public Schools are distributed in Madrid:

	PK	NOMBRE	DISTRITO	LATITUD	LONGITUD
0	176850	Colegio Público Adolfo Suárez	HORTALEZA	40.488603	-3.652554
1	5075	Colegio Público Agustina Díez	PUENTE DE VALLECAS	40.388191	-3.647237
2	4693138	Colegio Público Agustín Rodríguez Sahagún	VILLA DE VALLECAS	40.368040	-3.622168
3	5076	Colegio Público Alberto Alcocer	SAN BLAS-CANILLEJAS	40.433338	-3.612052
4	5077	Colegio Público Alcalde de Móstoles	LATINA	40.379837	-3.776448



Foursquare Venues Data

Foursquare API provides a way to get close venues of certain categories. Reviewing all categories, some of them are interesting for obtaining possible locations to run activities with children, such "Outdoors & Recreation", "Museum", "Music Venue" and "Performing Arts Venue" for each school: <https://developer.foursquare.com/docs/resources/categories>

Those categories have some subcategories, and some of them are out of the interest for the propose of this document so were removed from the study. Those categories are the ones including these words in their names: 'Gym', 'Neighborhood', 'Fitness', 'Yoga', 'Bar', 'Dog', 'Restaurant', 'Club', 'Park', 'Plaza'

Using the search endpoint of the Forusquare API and matching with the location of the Public Schools it's possible to get a list of the most common venue categories close to each school:

	School	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Colegio Público Adolfo Suárez	Pool	Other Great Outdoors	Outdoors & Recreation
1	Colegio Público Agustina Díez	Pool	Garden	Other Great Outdoors
2	Colegio Público Agustín Rodríguez Sahagún	Dance Studio	Playground	Embassy / Consulate
3	Colegio Público Alberto Alcocer	Playground	Pool	Field
4	Colegio Público Alcalde de Móstoles	Dive Spot	River	Pool

5.Results

After some tests, grouping Public Schools in 3 types seems to be a good approach. These are some of the most common venues in those groups:

Group 0:

	PK	NOMBRE	DISTRITO	LATITUD	LONGITUD	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	5075	Colegio Público Agustina Díez	PUENTE DE VALLECAS	40.388191	-3.647237	0	Pool	Garden	Other Great Outdoors
2	4693138	Colegio Público Agustín Rodríguez Sahagún	VILLA DE VALLECAS	40.368040	-3.622168	0	Dance Studio	Playground	Embassy / Consulate
9	5081	Colegio Público Amador de los Ríos	SALAMANCA	40.426280	-3.666558	0	Dance Studio	Theater	Museum
12	5087	Colegio Público Antonio Machado	CARABANCHEL	40.367425	-3.754198	0	Pool	Music School	Music Venue
13	5088	Colegio Público Antonio Moreno Rosales	CENTRO	40.411471	-3.701275	0	Theater	Other Great Outdoors	Performing Arts Venue

Group 1:

	PK	NOMBRE	DISTRITO	LATITUD	LONGITUD	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	5075	Colegio Público Agustina Díez	PUENTE DE VALLECAS	40.388191	-3.647237	0	Pool	Garden	Other Great Outdoors
2	4693138	Colegio Público Agustín Rodríguez Sahagún	VILLA DE VALLECAS	40.368040	-3.622168	0	Dance Studio	Playground	Embassy / Consulate
9	5081	Colegio Público Amador de los Ríos	SALAMANCA	40.426280	-3.666558	0	Dance Studio	Theater	Museum
12	5087	Colegio Público Antonio Machado	CARABANCHEL	40.367425	-3.754198	0	Pool	Music School	Music Venue
13	5088	Colegio Público Antonio Moreno Rosales	CENTRO	40.411471	-3.701275	0	Theater	Other Great Outdoors	Performing Arts Venue

Group 2:

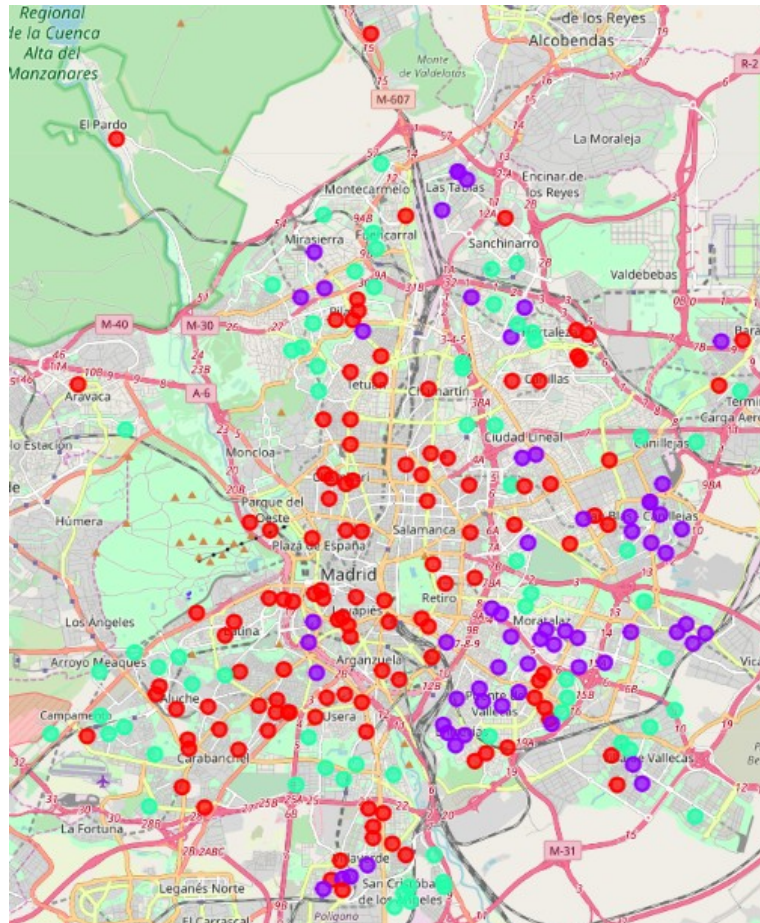
	PK	NOMBRE	DISTRITO	LATITUD	LONGITUD	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	5075	Colegio Público Agustina Díez	PUENTE DE VALLECAS	40.388191	-3.647237	0	Pool	Garden	Other Great Outdoors
2	4693138	Colegio Público Agustín Rodríguez Sahagún	VILLA DE VALLECAS	40.368040	-3.622168	0	Dance Studio	Playground	Embassy / Consulate
9	5081	Colegio Público Amador de los Ríos	SALAMANCA	40.426280	-3.666558	0	Dance Studio	Theater	Museum
12	5087	Colegio Público Antonio Machado	CARABANCHEL	40.367425	-3.754198	0	Pool	Music School	Music Venue
13	5088	Colegio Público Antonio Moreno Rosales	CENTRO	40.411471	-3.701275	0	Theater	Other Great Outdoors	Performing Arts Venue

It's possible to have 3 differentiated groups of Public Schools in Madrid attending to their close venues than can be used for running activities with children:

- Type 0: School with close water activities
- Type 1: School with close sport centers and (possible) outdoor activities
- Type 2: School with close music and arts venues

Regarding location of each type, there is a clear distinction between the distribution of type 2 and type 0/1.

Type 2 schools are mainly located downtown while types 0 and 1 are in the suburbs. Type 1 is also mainly distributed among south-east of Madrid city.



6. Conclusion

From the results it's clear that Madrid City Hall should promote art/music activities for Public Schools located in Downtown, since it this kind of venues are more easily to access from them. For Public Schools located in the suburbs the recommendation would be to promote sports activities, and it could be a good idea to specifically focus in water activities (pools) since those ones are close to the majority of Public Schools in that areas.

Another option is to have two different programs for Public Schools in the suburbs, one focused in indoor/outdoor sports and another one focused just in pool activities.

Even though we end up with three categories of Public Schools, due the kind of the most common categories of closest venues, it could make sense to have to big programs one focused in arts and music for downtown Public Schools and another one focused in sports for suburbs Public Schools.