



Pipeline Proyecto

Alumno: Apaza Apaza Nelson Jorge

Contexto

¡En Engagement, Ohio, el futuro es ahora! Durante años, esta tranquila comunidad residencial fue una joya secreta en el corazón del estado. ¡Pero ahora el secreto ha salido a la luz y la gente no camina, sino que corre para reclamar su pedazo de este paraíso!

Anticipándose a un rápido crecimiento, la ciudad de Engagement, Ohio (EE. UU.) está realizando un ejercicio de planificación urbana participativa para comprender el estado actual de la ciudad e identificar oportunidades para su desarrollo futuro. Alrededor de 1000 residentes representativos de esta ciudad de tamaño modesto han aceptado proporcionar datos utilizando la aplicación de planificación urbana de la ciudad, la cual registra los lugares que visitan, sus gastos y sus compras, entre otras cosas. A partir de estos voluntarios, la ciudad contará con datos que ayudarán en sus importantes esfuerzos de revitalización comunitaria, incluyendo cómo asignar una gran subvención de renovación urbana que han recibido recientemente. Como experto en análisis visual, te has unido al equipo de planificación urbana para dar sentido a los datos proporcionados por estos residentes.



Análisis rápido de la base de datos (luego están las preguntas):

Proyecto - Base de Dat... > Activity Logs

Tipo Personas Modificado Fuente

Nombre	Propietario
.DS_Store	yo
ParticipantStatusLogs10.csv	yo
ParticipantStatusLogs9.csv	yo
ParticipantStatusLogs8.csv	yo
ParticipantStatusLogs7.csv	yo
ParticipantStatusLogs6.csv	yo
ParticipantStatusLogs5.csv	yo
ParticipantStatusLogs4.csv	yo
ParticipantStatusLogs3.csv	yo
ParticipantStatusLogs2.csv	yo
ParticipantStatusLogs1.csv	yo

Proyecto - Base de Dat... > Journals

Tipo Personas Modificado Fuente

Nombre	Propietario
.DS_Store	yo
TravelJournal.csv	yo
SocialNetwork.csv	yo
FinancialJournal.csv	yo
CheckinJournal.csv	yo

Proyecto - Base de Dat... > Attributes

Tipo Personas Modificado Fuente

Nombre	Propietario	Última modifica...
.DS_Store	yo	14 abr 2022 yo
Jobs.csv	yo	14 abr 2022 yo
Pubs.csv	yo	14 abr 2022 yo
Apartments.csv	yo	5 abr 2022 yo
Buildings.csv	yo	5 abr 2022 yo
Employers.csv	yo	5 abr 2022 yo

ParticipantStatusLogs1.csv: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
timestamp,currentLocation,participantId,currentMode,hungerStatus,sleepStatus,apartmentId,available
2022-03-01T00:00:00Z,POINT (-2724.6277665310454 6866.2081834436985),0,AtHome,JustAte,Sleeping,926,
2022-03-01T00:00:00Z,POINT (-1526.9372331431534 5582.2951345645315),1,AtHome,JustAte,Sleeping,928,
2022-03-01T00:00:00Z,POINT (-1360.9905987829304 2108.804385379679),2,AtHome,JustAte,Sleeping,291,1
2022-03-01T00:00:00Z,POINT (-1558.517200825967 5600.664347152427),3,AtHome,JustAte,Sleeping,1243,1
2022-03-01T00:00:00Z,POINT (976.2409614204214 4574.575079882071),4,AtHome,JustAte,Sleeping,194,-68
2022-03-01T00:00:00Z,POINT (-1525.6957374012197 1994.5285187115571),5,AtHome,JustAte,Sleeping,243,
2022-03-01T00:00:00Z,POINT (1795.1297501295278 3238.4053705049837),6,AtHome,JustAte,Sleeping,183,1
2022-03-01T00:00:00Z,POINT (-1023.8165705255449 1578.3713681439597),7,AtHome,JustAte,Sleeping,97,6
2022-03-01T00:00:00Z,POINT (616.2956028633527 2274.8909931311796),8,AtHome,JustAte,Sleeping,321.54
```

Apartments.csv: Bloc de notas

Archivo Edición Formato Ver Ayuda

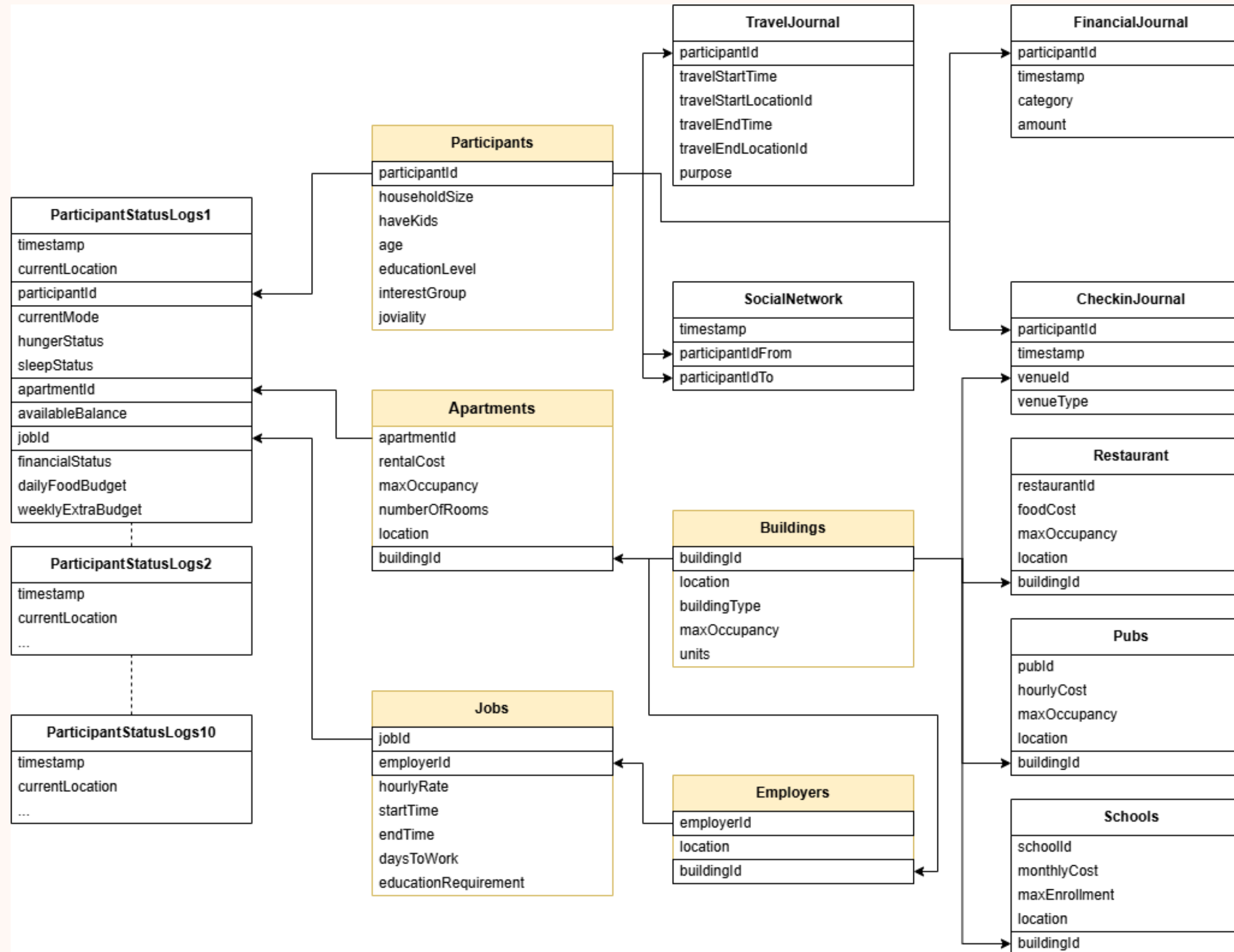
```
apartmentId,rentalCost,maxOccupancy ,numberOfRooms,location,buildingId
1,768.16,2,4,POINT (1077.6979444315298 648.4427163702453),340
2,1014.55,2,1,POINT (-185.9292838076562 1520.3270983045118),752
3,1057.39,4,3,POINT (2123.0141855392585 5126.753457243003),639
4,1259.1,4,3,POINT (2103.6301776944765 4266.932930123476),397
5,411.5,1,4,POINT (7.0589743819342985 79.96163671849988),628
6,859.58,3,2,POINT (2250.85490611142 5251.3368306902885),533
7,982.11,3,4,POINT (486.8811262316384 2251.1260599901484),61
8,980.05,4,1,POINT (1233.4547558395932 1768.6111384755895),360
9,433.45,1,3,POINT (1274.2715913565519 1163.5051209752276),251
10,1104.33,3,4,POINT (-1697.0303105735857 1239.0301787057274),512
11,960.16,3,2,POINT (-341.9635948004018 1222.5966329871515),922
```

Jobs.csv: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
jobId,employerId,hourlyRate,startTime,endTime,daysToWork,educationRequirement
0,379,10,7:46:00 AM,3:46:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",HighSchoolOrColla
1,379,22,21763336,7:31:00 AM,3:31:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Bachelor
2,380,10,8:00:00 AM,4:00:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Bachelors
3,380,15,31207064,7:39:00 AM,3:39:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Bachelor
4,381,21,35540929,7:53:00 AM,3:53:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",HighSch
5,381,12,09382569,8:13:00 AM,4:13:00 PM,"[Monday,Sunday,Thursday,Tuesday,Saturday]",HighSch
6,381,21,84618702,8:36:00 AM,4:36:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",HighSch
7,381,18,71319307,7:47:00 AM,3:47:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Low
8,382,13,37043245,7:45:00 AM,3:45:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Bachelor
9,387,14,52833044,7:10:00 AM,3:10:00 PM,"[Monday,Tuesday,Wednesday,Thursday,Friday]",Graduate
```

Análisis rápido de la base de datos (luego están las preguntas):



¿Qué problemas identifican en el dataset?

- Al observar "ParticipantStatusLogs*", hay algunos archivos que en apartmentID los tiene como tipo de almacenamiento "float", pero en verdad debería ser "int".
- Debido a ello, convertimos la columna 'apartmentId' a tipo entero, capartmentId es un identificador discreto y nominal (nunca debería tener decimales).

```
⇒ Tipos de datos por columna:  
timestamp          object  
currentLocation     object  
participantId       int64  
currentMode         object  
hungerStatus        object  
sleepStatus         object  
apartmentId         int64  
availableBalance    float64  
jobId               int64  
financialStatus      object  
dailyFoodBudget     int64  
weeklyExtraBudget   float64  
dtype: object
```

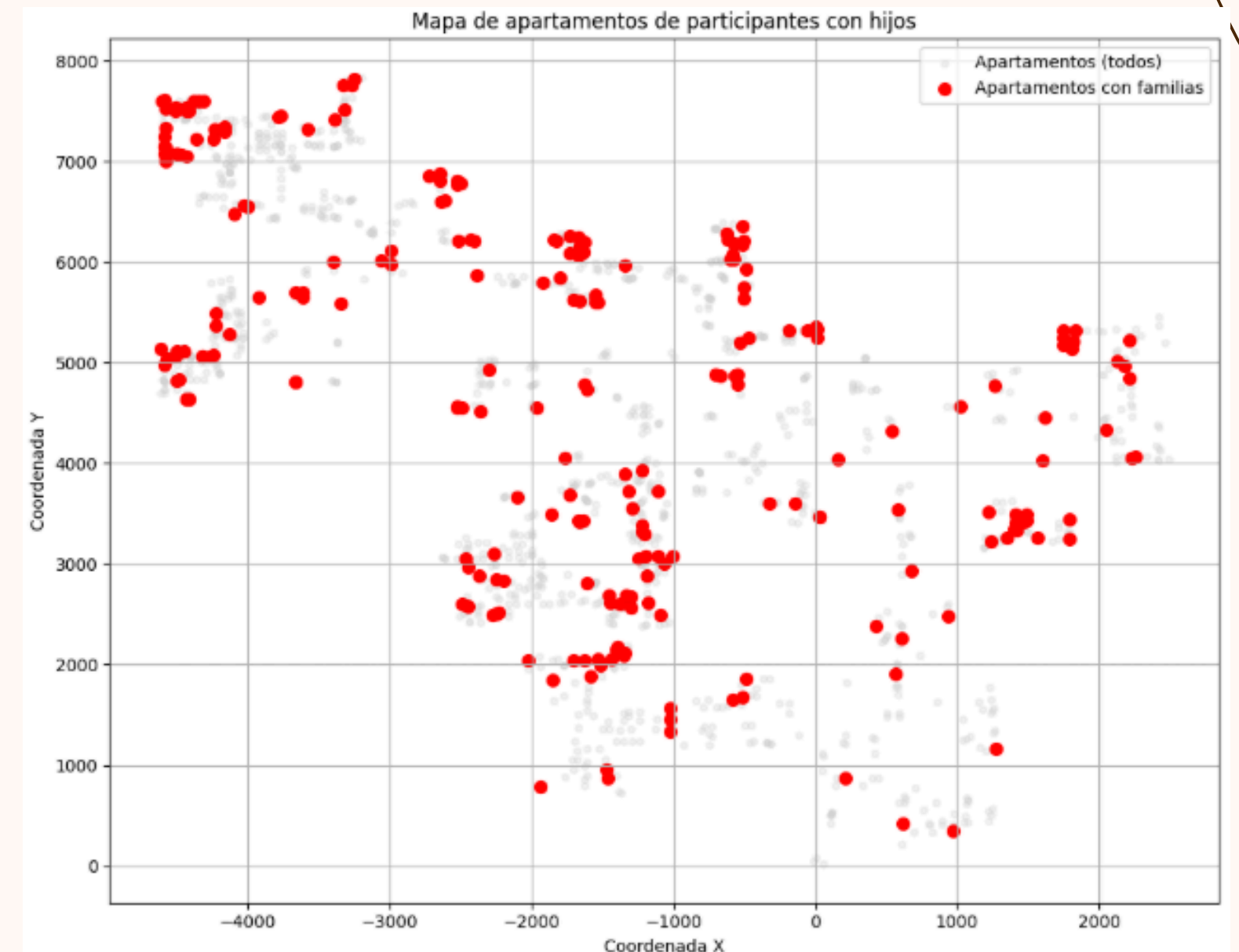

¿Qué problemas identifican en el dataset?

- En la tabla de apartments "location" es una variable cualitativa nominal codificada en una estructura tipo texto, con contenido que representa coordenadas.
- Su forma original "POINT (x y)" no puede ser procesada numéricamente ni graficada sin transformación. Además la ubicación en el espacio geográfico tiene significado métrico: permite comparar distancias, calcular cercanías, concentraciones, proximidades, etc.
- Así que nos conviene transformar a dos variables cuantitativas continuas: x y y.

	apartmentId	rentalCost	maxOccupancy	numberOfRooms	location	build
0	1	768.16	2	4	POINT (1077.6979444315298 648.4427163702453)	
1	2	1014.55	2	1	POINT (-185.9292838076562 1520.3270983045118)	
2	3	1057.39	4	3	POINT (2123.0141855392585 5126.753457243003)	
3	4	1259.10	4	3	POINT (2103.6301776944765 4266.932930123476)	
4	5	411.50	1	4	POINT (7.0589743819342985 79.96163671849988)	

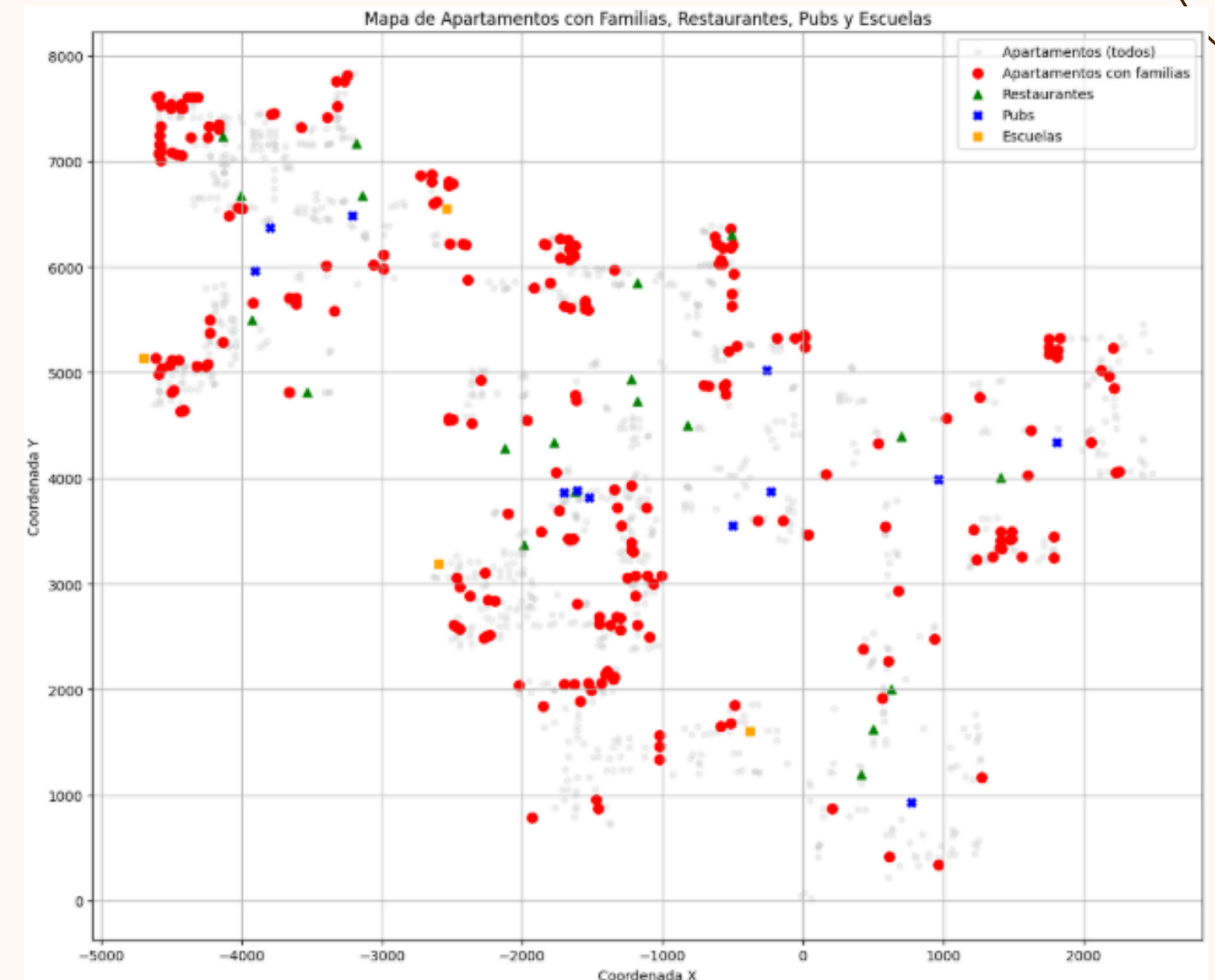
1. ¿Qué descubrieron al analizar los datos?

- Se descubrió que existen zonas preferidas para vivir para las familias que se tienen en el dataset, ellos prefieren tanto la parte céntrica como la parte de la esquina superior izquierda, viendo que las demás zonas no viven muchas de ellas. Esto nos da una nueva pregunta ¿Por qué estas personas(familias), prefieren dichas zonas? ¿Acaso habrá algún local o actividad que los atraiga?

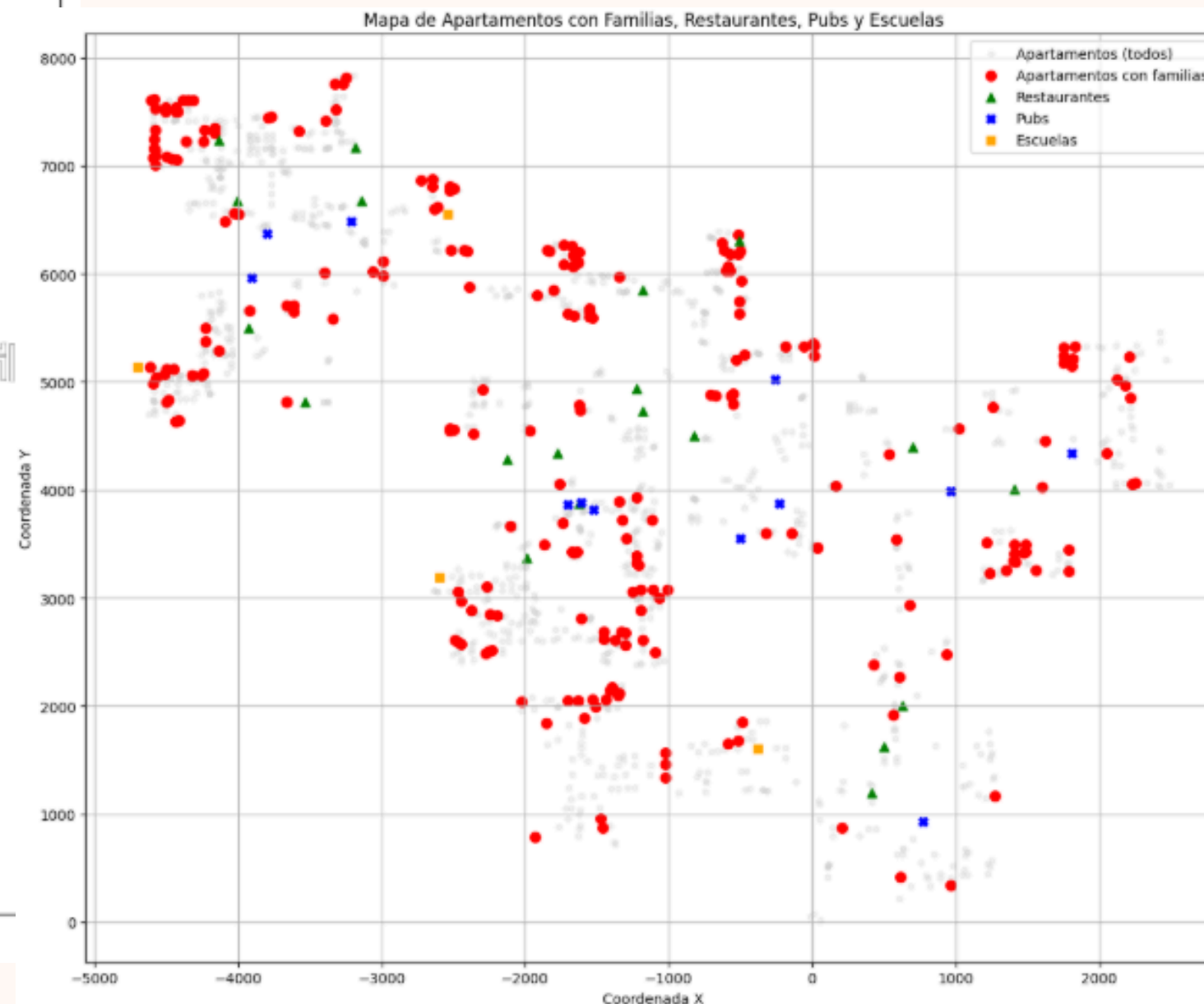
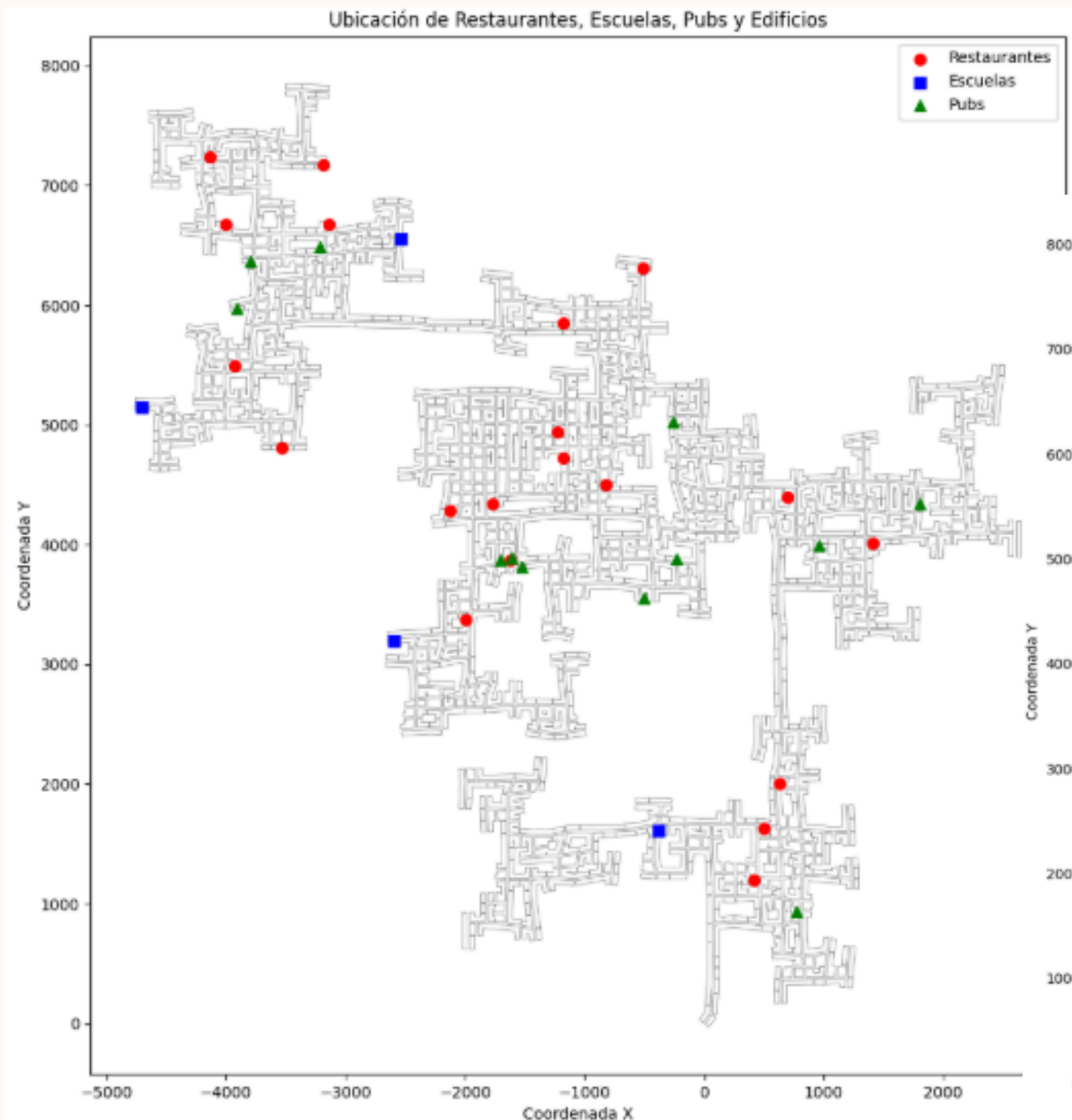


1. ¿Qué reflejan los patrones de tendencia?

- Las zonas que mayor número de habitantes en general tienen son aquellas que tienen mayor presencia de 3 tipos de edificios: restaurantes, escuelas y pubs. Esto incluso responde a nuestra pregunta propuesta en la pregunta anterior, donde se ve que las familias también se inclinan a buscar dichas zonas.
- Se puede apreciar que hay zonas donde hay presencia de un mayor número de restaurantes, pero poca o ninguna presencia de escuelas o pubs. Estas zonas no son muy habitadas, demostrando que las personas prefieren estar en lugares donde puedan acceder fácilmente a distintos lugares.



1. ¿Cómo es afectado el comportamiento humano (movilidad) en relación con la geografía?



- En la imagen de abajo, e incluso revisando las anteriores, el movimiento de las personas se da en torno a la ubicación de los principales edificios: restaurantes, escuelas, y pubs. Donde resalto a los colegios, ya que alrededor de estos, la movilidad de las personas es mayor. Y si se tiene a los 3 tipos de edificios que acabo de mencionar, el movimiento es mayor en dichas zonas.



¡Gracias!