Using Space to Identify Opportunity

Census Administration

Pew Charity and Equity Institute

The Key Idea...

Use granular statistical information and spatial data to identify high-value (and low adversity) housing for lower income families

Data Sources

1. Census (Structure)
2. San Diego Police Department (Structured)
3. Reddit Data (Unstructured)
4. Google Geocode API (Unstructured)
5. Housing Data (Unstructured)
6. GeoJson Mapping Polygons

Statistical info
Census Tract data which is localized (1 tract per every 8000 people)

Geo Spatial Data
Housing: use Realtor Data to get individual housing costs

Crime: Use Open Portal Police Data to get individualized crimes

Open Street Map to look at cafes, shopping, etc and identify patterns of cost and value

one single, fully connected graph of the place

https://taginfo.osm.org/node/9 Osma Data

graph model - but not a graph database

node is a geographic location

a way is a connection of nodes in order, can be closed (buildings, streets as well as polygon - park or walking)

relation is an arbitrary grouping of nodes, ways and other relations, could be multiple times

people have to tell you what the data means - the column is not that

query the data from Amazon Athena

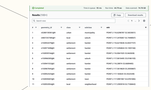
Get polygon of San Diego County

List of available polygons for id = 396482

parent *	timestamp	NPoints	Length	WKT	GeoJSON	poly	temp
0	2022-11-26 04:23:34.126976	407	1.0027564717783	WKT	GeoJSON	poly	temp
0.020000-0.005000-0.005000	2022-12-02 19:28:21.560000	39	1.7347109974174	WKT	GeoJSON	poly	temp

Refresh original geometry

Run a query with these polygon coordinates to filter only San Diego County



create nodes and edges

https://drive.neo4j.com/r/1MGqNVRbbGbqqq-RZ0YKcvnbqwmcVqwJ9ccJ8Ldl_Ns/edit?usp=sharing&folder=1ZwTwbVmcqbwcUDCL1_Ysm4-2qnfcNdbf&sharecode=eyJhaXQiOiAgIiAgIiwgIiAgIiAgIiA