Source Data Access and Analysis

Contributors: Alan Li, Blake Crowther, Wesley Schiller

Data Sources

Google Place API

Resource Type: Google API

Description: The Google Places API is a service that allows developers to query geographical locations, retrieve detailed information about places, and interact with a wide variety of location data. It supports features like searching for places by name, address, or type (such as restaurants, parks, or shops), retrieving reviews, photos, business hours, and more.

API Endpoints:

Name	Description	Parameters	Usage
Text Search (New)	Lets you specify a text string on which to search for a place. Supports various search criteria like place types, keywords, location, price range, and more. Returns a list of places that match the search criteria.	`query`: Search query (e.g., "pizza near me") `location`: Latitude and longitude of the location `radius`: Radius in meters `type`: Place type (e.g., "restaurant") `keyword`: Keyword to search for `language`: Language of the results `opennow`: Indicates if the place is open now `price_level`: Price level of the place (1-4) `place_id`: Place ID of the place `name`: Name of the place `types`: Types of places to search for	https://maps.googleapis.com/maps/api/place/textsearch/json?query=pizza+near+me&key=YOUR_API_KEY

Nearby Search (New)	Returns places located within a specified radius of a given location. Supports filtering by place types, keywords, and other criteria. Ideal for finding nearby restaurants, attractions, or businesses.	`location`: Latitude and longitude of the location `radius`: Radius in meters `type`: Place type (e.g., "restaurant") `keyword`: Keyword to search for `language`: Language of the results `opennow`: Indicates if the place is open now `price_level`: Price level of the place (1-4) `place_id`: Place ID of the place `name`: Name of the place `types`: Types of places to search for	https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=40.7127837,-74.0059413&radius=1500&type=restaurant&key=YOUR_API_KEY
Place Details (New)	Provides detailed information about a specific place using its place ID. Includes address, phone number, website, reviews, photos, and other relevant data.	`placeid`: Place ID of the place `fields`: Fields to return (e.g., "address", "phone_number", "website", "reviews", "photos") `language`: Language of the results	https://maps.googleapis.com/maps/api/place/details/json?placeid=ChIJ0Qm52F9i44R_9094k23Q&key=YOUR_API_KEY
Autocomplete (New)	Returns place predictions and query predictions in response to an HTTP request. In the request, specify a text search string and geographic bounds that control the search area.	`input`: Search query `location`: Latitude and longitude of the location `radius`: Radius in meters `types`: Types of places to search for `language`: Language of the results `components`: Components to filter results by (e.g., "country:US")	https://maps.googleapis.com/maps/api/place/autocomplete/json?input=pizza&key=YOUR_API_KEY

Knowledge Graph Usage: This API can be used to identify and represent businesses as nodes in the graph. Each business can be a node, and their types (grocery, fast food, etc.) will help classify and analyze the distribution of businesses across neighborhoods. We can then

establish relationships (edges) between business nodes and administrative regions (neighborhoods, block groups, etc.) based on their geographical locations. We can also enrich this mapping with socioeconomic context. By combining the Google Place API with the other datasets our graph structure could provide meaningful insights to our users. One such use case might be identifying food deserts in lower-income areas by querying the business type density of lower-income regions.

Geo-enriched Blockgroup Data for San Diego and Imperial Counties (Geographic and Demographic Information)

Dataset: bgs_sd_imp

Resource Type: NDP Postgres Database

Description: This geo-enriched blockgroup dataset for San Diego and Imperial Counties includes detailed geographic boundaries, population statistics, socioeconomic indicators, and diverse demographic metrics. With 2172 rows x 588 columns, this dataset is very detailed and contains a huge amount of information. We will use it for geographic and demographic analysis. By integrating this dataset with other relevant datasets, such as the arcgis_geoenrichment dataset discussed next, we will conduct a comprehensive analysis to inform business and spatial planning decisions.

Field Name	Data Type	Description
ogc_fid	Int	Primary Key
statefp, countyfp, tractce, blkgrpce	text	These columns represent geographic identifiers for states, counties, census tracts, and block groups
source_cou	text	Source country identifier (for this dataset: USA)
population	float	Population of the block group
apportionm	float	Population apportionment
has_data	boolean	Flag indicating data availability (1 = has data, 0 = no data).
From x2001_a to sei_cy, etc	float	Total of 578 columns, these columns indicate numercial values aggregated by the block group, e.g: some financial values in different fiscal year bands (e.g: sei_cy indicates Socioeconomic index for the current year)
wkb_geometry	binary	This column contains spatial data in Well-Known Binary (WKB) format, which is used to represent geographic shapes or areas, corresponding to the block groups

Dataset: arcgis_geoenrichment

Resource Type: NDP Postgres Database

Description: This dataset has 18123 rows × 5 columns, it contains information to further explain the column variables of the bgs_sd_imp dataset, we will use it for decoding the meaning of column variables from the bgs_sd_imp dataset, allowing for filtering and identifying the variables we want to use for analysis. By integrating this dataset with other relevant datasets, we will conduct a comprehensive analysis to inform business and spatial planning decisions.

Field Name	Data Type	Description
datacollectionid	text	Source for the data collection
analysisvariable	text	represents the coded identifier for the variable
alias	text	A more user-friendly description or label for the variable
fieldcategory	text	represents a category for the variables
vintage	numeric	represents year or time peroid

screen shot of the Bgs_sd_imp dataset, what do these column variables mean ?

	K	L	M	N	O	P
	x2001_a	x2002_a	x2003_a	x2004_a	x2005_a	x2006_a
1	148.41	101.33	42.97	7.55	39.9	10.9
1	668.78	445.64	162.52	37.22	203.79	42.11
1	662.82	489.24	238.36	38.89	164.16	47.84

By referencing the arcgis_geoenrichment dataset, for example, x2001_a represents average alcoholic beverage spending.

food.X2001_A	2023 Avg: Alcoholic Bevs	2023 Alcoholic Beverages (Consumer Spending)

Knowledge Graph Usage: The knowledge graph nodes represent block groups (regions) with attributes like population, income, and spending patterns etc. The edges link these regions to spatial boundaries and demographic attributes. By connecting the bgs_sd_imp and arcgis_geoenrichment datasets, the knowledge graph enables comprehensive spatial planning and business analysis. Relationships between regions and socioeconomic indicators, or regions and lifestyle comparison (e.g., average alcohol spending in one block group vs neighbor block group) provide insights for business opportunities. It helps for making data-driven decisions for targeted business planning.

Administrative Topology for City and Community

Resource Type: NDP Postgres Database

Description: This dataset contains detailed information about the administrative topology, providing users with a comprehensive view of the hierarchical relationships between states, counties, and cities. It also specifies the neighboring cities or communities for each location, allowing for an understanding of geographical and jurisdictional proximity across different administrative regions.

Knowledge Graph Usage: For the knowledge graph construction, the city neighborhoods and community neighborhoods data points can be represented by nodes. The neighborhoods data can be integrated with the geo enrichment data to develop relationships/edges between the nodes. The knowledge graph could be used to determine business opportunities by analyzing demographic information such as age, income, education level, local economic environment, and market trends by a selected region (city, community, neighborhoods). Analytics could provide business value by identifying gaps in services and demand for specific products within a particular area, allowing to strategically pinpoint potential business ventures.

Tables:

city_neighborhoods	
id	primary key
state_name	text
county	text
city	text
metro_area	text
is_unincorporated	boolean
zipcodes	text
neighboring_cities	text
neighboring_unincorporated_places	text
nearby_unincorporated_places	text
neighborhoods	text
nearby_cities	text

Sample tuple from table:

id	state_name	county	city	metro_area	is_unincorporated_place	zipcodes	neighboring_cities
16	California	San Diego	La Mesa	San Diego Area	False	[91941, 91942, 92020, 92120]	[El Cajon, Lemon Grove, San Diego]

neighboring_unincorporated_places	nearby_unincorporated_places	neighborhoods	nearby_cities
[Casa de Oro-Mount Helix, Spring Valley]	[Bonita, Bostonia, Eucalyptus Hills, Granite H...	[Lake Murray, College East, Lake Murray, San C...	[National City, Santee]

community_neighborhoods	
id	primary key
state_name	text
county	text
city	text
metro_area	text
community	text
zipcodes	text
neighboring_cities	text
neighboring_unincorporated_places	text
nearby_unincorporated_places	text
neighboring_communities	text
nearby_communities	text
nearby_cities	text

Sample tuple from table:

id	state_name	county	city	metro_area	community	zipcodes	neighboring_cities	neighboring_unincorporated_places
441	California	San Diego	San Marcos	San Diego Area	La Costa Oaks	[92009, 92024, 92078]	None	None

nearby_unincorporated_places	neighboring_communities	nearby_communities	nearby_cities
None	[La Costa Ridge, Rancho La Costa]	[Aviara, Bressi Ranch, Carlsbad Ranch, Hediond...	None

arcgis_geoenrichment	
datacollectionid	text
analysisvariable	text
alias	text
fieldcategory	text
vintage	numeric

Sample from table:

datacollectionid	analysisvariable	alias	fieldcategory	vintage
Wealth	Wealth.VALBASE_FY	2028 Home Value Base	2028 Home Value (Esri)	2028.0
Wealth	Wealth.SEI_CY	2023 Socioeconomic Status Index	2023 Key Demographic Indicators (Esri)	2023.0

Knowledge Graph Usage: The datacollectionid column is an id that points to a data collection. A data collection is a preassembled list of attributes or groupings of variables that can be used to enrich the input features. These groupings can speed up analysis by offering a selection of variables, i.e. diversity index, wealth index, median age, total population, etc. Collection attributes can describe various types of information, such as demographic characteristics and geographic context by location. Once the data collection is selected, the analysisvariable/alias columns can be looked at for variables of interest. Notably, each data collection and analysis variable has a unique id. This id can be used to find geo-enrichment data for an area with the given analysis variable. In summary, the values in this table can be useful for filtering to identify the variables used for analysis and also for matching to existing datasets.

SANDAG Road Network Data

Resource Type: Regional GIS Data

Description: The San Diego Association of Governments (SANDAG) provides a single source of reliable GIS data, accessible from the interactive SANDAG Regional Data Warehouse. The Regional Data Warehouse consists of 18 data categories and provides over 330 GIS layers such as roads/freeways, property, city boundaries, business zones, and more.

Data Formats: FileGDB, Shapefile, CSV, GeoJSON, JSON

Relevant Datasets:

Census_Block_Groups**:** Comprises polygons of census block groups for San Diego County.

Field	Type	Description	Addl. Info & Constraints
OBJECTID	OID	Internal feature number.	
Shape	Geometry	Feature geometry.	
CTBLOCKGROUP	Integer	Census Tract and Blockgroup (combined).	
TRACT	Integer	Census Tract	- Contains at least one BG. - Typically contains between 1,200 to 8,000 people.

BLOCKGROUP	SmallInteger	Blockgroup ID	- Range: 0 - 9 - Block Groups (BG) are clusters of blocks within the same census tract. - Uniquely numbered within census tracts. - Have the same first digit of their 4-digit census block - Blockgroups coded 0 are intended to include only water areas. - Generally covers a contiguous area but never crosses county or census tract boundaries. - Typically contains between 600 and 3,000 people.
Shape_Length	Double	Length of feature in internal units.	
Shape_Area	Double	Area of feature in internal units squared.	

Knowledge Graph Usage: This dataset provides polygon boundaries of census block groups, which can be used to define nodes representing geographical areas. These nodes could serve as key geographic units that can be connected to other data sources such as businesses or demographic information.

Land_Use_2023: SANDAG performs an annual land use and housing unit inventory in the interest of maintaining a robust and accurate catalog of the existing conditions for any given year. This catalog of snapshots are the base year inputs to SANDAG's Regional Demographic, Economic, and Land Use Models.

Field	**Type**	**Description**	**Addl. Info & Constraints**
OBJECTID	OID	Internal feature number.	
Shape	Geometry	Feature geometry.	
lu	SmallInteger	Existing land use.	SANDAG four-digit land use code.
DESCRIPTION	String	Description of land use codes.	
Shape_Length	Double	Length of feature in internal units.	
Shape_Area	Double	Area of feature in internal units squared.	

Knowledge Graph Usage: This dataset can be connected to block group or business nodes to correlate land usage zoning and business density. By doing so we might be able to identify regions with a shortage of grocery stores despite zoning approval for commercial activity.

Road_segments: This dataset comprises road segments (both active and inactive) of roads maintained by the County of San Diego or in the inventory because other County assets exist on it. Created to assist in the management of the county's road-related assets.

Field	Type	Description	Addl. Info & Constraints
OBJECTID	OID	Internal feature number.	
ACCELA_ID	String	Unique ID for Public Works asset management system.	
ROAD_SEG_ID	String	ID comprised for the beginning intersection number and the ending intersection.	
ROAD_NAME	String	Road name.	
ROAD_ID	String	Road ID.	
ROAD_SEQ_NUMBER	String	Road sequence number.	q
FROM_STREET	String	Intersecting road name at beginning of road segment.	
TO_STREET	String	Intersecting Road name at end of road segment.	
GIS_FROM_POSITION	Double	Location of the beginning of the asset from beginning of road segment (in feet).	For road segments, it is always 0.
GIS_TO_POSITION	Double	Location of the end of the asset from end of road segment (in feet).	
GIS_LENGTH	Double	Length of asset (in feet).	
INVENTORY_DATE	Date	Date asset was accepted into County of San Diego inventory.	
EXPIRATION_DATE	Date	Date asset was removed from County of San Diego inventory.	
RECORD_MAP_NO	String	Map number.	
JURISDICTION	String	Maintenence responsibility.	

MOBILITY_ELEMENT	String	County of San Diego Mobility Element designation.	2.1 Community Collector 2.2 Light Collector 2.3 Minor Collector 4.1 Major Road 4.2 Boulevard 6.1 Expressway 6.2 Prime Arterial LPR Local Public Road
STREET_DIST	String	Public Works Street District the asset is located.	
NEIGHBORHOOD	String	Neighborhood the asset is located in.	
ASSET_STATUS	String	Asset status.	
Shape	Geometry	Feature geometry.	
Shape_Length	Double	Length of feature in internal units.	

Knowledge Graph Usage: Connecting road segment data to business and block group nodes can help analyze accessibility—identifying areas with poor road connectivity. Edges between road segments and nearby businesses or residential areas could be leveraged to assess transportation infrastructure's impact on service availability.