Food Deserts in Pasadena, California

Table of Contents

Introduction

The city of Pasadena, California is approximately 10 miles northeast of downtown Los Angeles, in the foothills of the San Gabriel mountains. It has an estimated population of 141,000 people. Wikipedia.

Pasadena has many, many restaurants. However, there are large parts of the city that do not have grocery stores. This report looks to identify those regions with an eye toward recommending locations for new grocery stores.

In other words, we are looking for "food deserts".

A food desert is "... an area that has limited access to affordable and nutritious food, in contrast with an area with higher access to supermarkets or vegetable shops with fresh foods". Wikipedia

This project is intended for entrepreneurs who might be interested in opening a new retail outlet. It is also aimed at city planners and related government officials who may be in a position to incentivize these kinds of establishments. The goal is to identify neighborhoods that need closer access to fresh food.

Data

Sources

The data for this project comes from several sources:

1. Wikipedia, for the definition of a food desert other background details.
2. The Python “folium” library, which provides maps of Pasadena and surrounding areas.
3. The Python “nominatum” library, which is used to look up longitude and latitude of Pasadena.
4. The City of Pasadena's Open Data web site which has GEOJson files that precisely define the city's borders. This project would have been much less precise without this data.
5. The Foursquare venues database which was used to find the list of existing grocery stores, and their locations.
6. Some original research, in the form of site visits. This was partially to confirm some of the data in Foursquare.

Cleaning and Other Issues

For the most part the data used in this project was already fairly clean, although gathering it proved a challenge. The curvature of the earth also made distance calculations slightly more complicated.

Foursquare’s API returns data in a JSON format that is easily parsed and traversed in Python. However, the API is limited to return only 50 venues per call. If a search would have more than 50 results, only 50 venues are returned and there is no indication when that limit has been encountered. The city of Pasadena is fairly large and although it is possible to set the parameters of a Foursquare search to include all venues within its borders, that search would hit the 50 venue limit. To get around this limitation, a grid of more than 200 points was overlaid on the city, and each search of the Foursquare API was centered on those points. The list of “search points” was used during analysis. The results of the searches were added to a pandas data frame to eliminate duplicates.

To further refine the search results, the search was limited to Foursquare’s “Food & Drink Shop” category. A manual review of the data found that some restaurants

were included in that category. Categories were further refined to remove from consideration restaurants and other establishments that were not of interest to this report, such as liquor stores. Farmers markets were also excluded, because they are only once a week. One venue was removed because it is closed.

The problem of the curvature of the earth affected how distances were calculated. As a part of the analysis, the distance form each point to every venue was calculated. The location of every search point and the location of every venue was specified by longitude and latitude. The center of Pasadena is at roughly 34 degrees North latitude and -118 degree longitude. 1 degree of latitude is roughly 110.5 kilometers at 34 degrees North; 1 degree of longitude is roughly 88 kilometers. To calculate the distance between a search point and venue, the difference in degrees was found and converted to kilometers for longitude, and then for latitude. Then these numbers were used in the Pythagorean theorem:

$$\text{distance} = \sqrt{\left(\left(\text{long}_s - \text{long}_v\right) * 88\right)^2 + \left(\left(\text{lat}_s - \text{lat}_v\right) * 110.5\right)^2}$$

Where

long_s : longitude of the search point

long_v : longitude of the venue

lat_s : latitude of the search point

lat_v : latitude of the search point

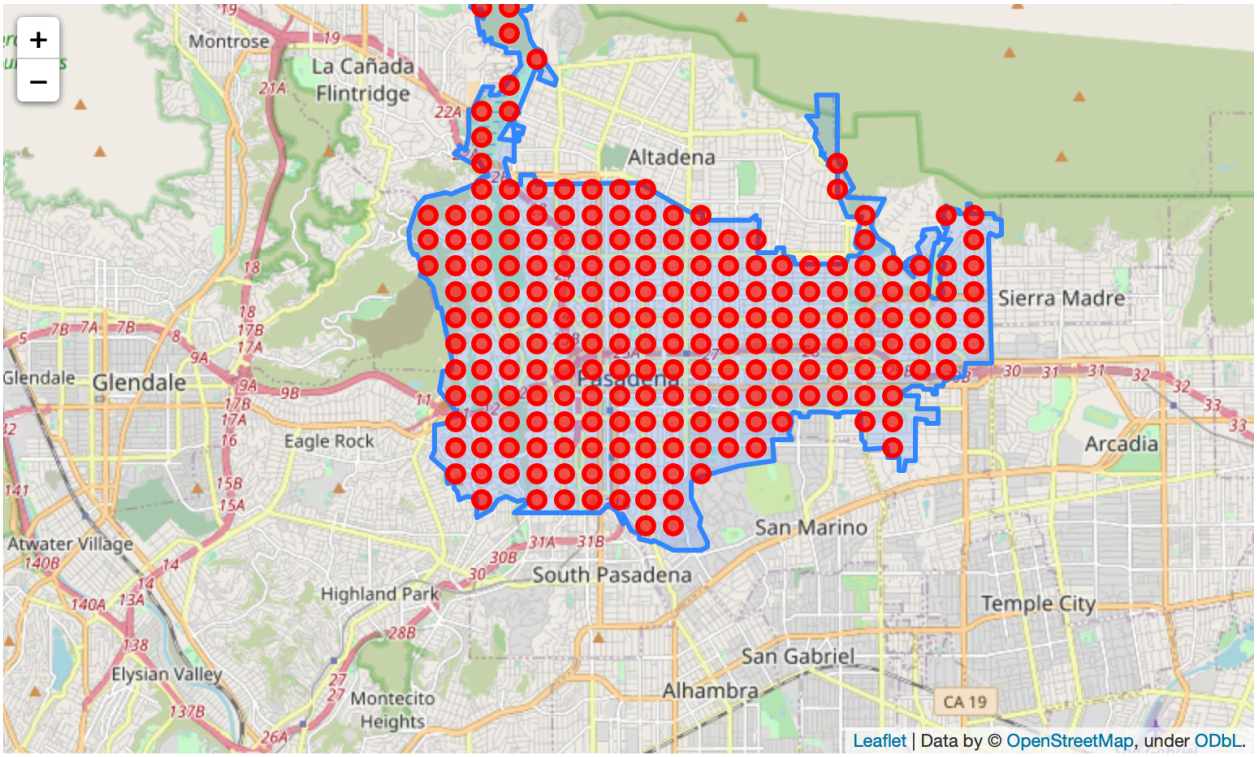
Feature Selection

Per the Wikipedia article, the United States Department of Agriculture reported a resident of the U.S. is in a food desert if “they live more than one mile from a supermarket in urban or suburban areas”. The only feature considered in this analysis is distance from a search point to a grocery store.

Methodology

To begin, a map of the city was generated using boundary data from the city’s web site and the python folium library

A grid of search points was overlaid on the city. These point became the basis for further analysis:



The longitude and latitude of each search point was placed in a pandas dataframe:

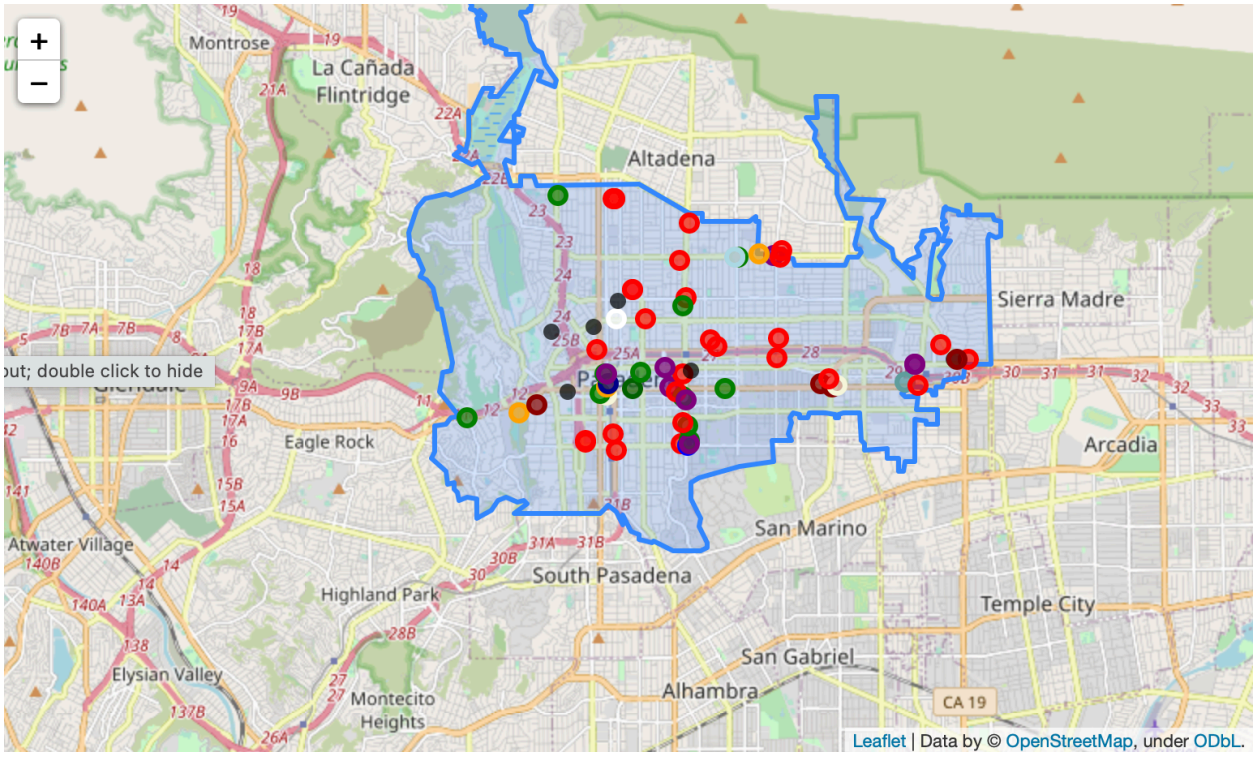
	SP Latitude	SP Longitude
0	34.121789	-118.140670
1	34.121789	-118.134896
2	34.126319	-118.175311
3	34.126319	-118.163764
4	34.126319	-118.157990

For each search point, a search was run against the Foursquare database, using the longitude and latitude of search search point as the center, a radius of 500 meters, and a limit of 50 venues per search (the maximum supported at the time). The venue information was saved in a dataframe:

	Venue Name	Venue Latitude	Venue Longitude	Venue Category	Venue ID
0	Raymond Restaurant	34.124552	-118.150038	American Restaurant	4b78620bf964a52012ca2ee3
1	Gerlach's Drive-In Liquor	34.127421	-118.150592	Liquor Store	4bd790dd09ecb713e940477c
2	Raymond Restaurant	34.124552	-118.150038	American Restaurant	4b78620bf964a52012ca2ee3
3	BevMo!	34.130553	-118.147717	Liquor Store	4a79bd12f964a52081e71fe3
4	Gerlach's Drive-In Liquor	34.127421	-118.150592	Liquor Store	4bd790dd09ecb713e940477c

Several venues were the removed: duplicate entries, a venue that is closed, and venues that were in categories that are not of interest such as “Sushi Restaurant”. The end result was that the dataframe of venues was reduced from 333 rows to 74.

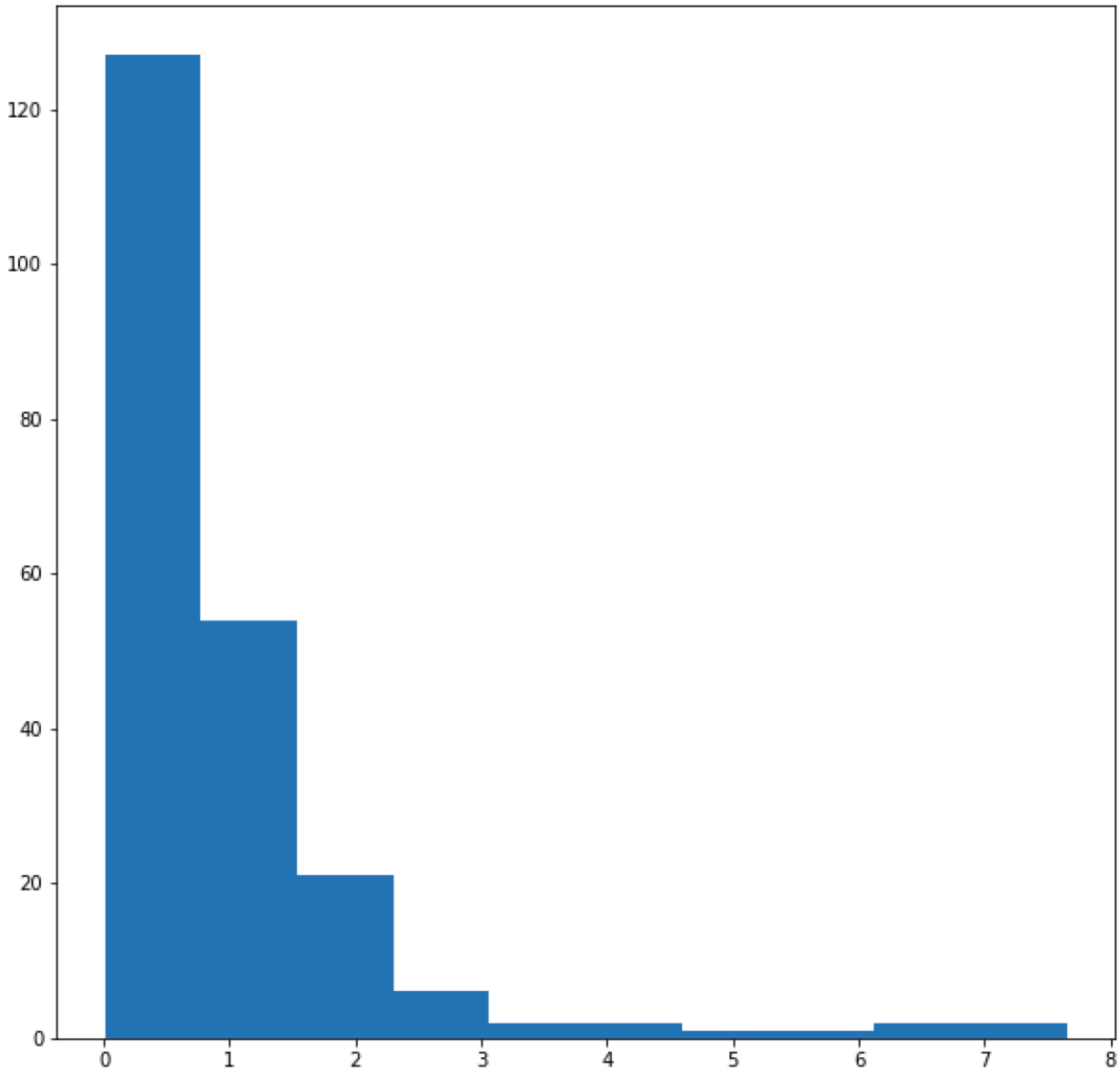
The list of venues was then compared to the boundaries of the city, with those outside the city were removed. this resulted in a collection of 70 venues:



A new dataframe was created. Each search point was matched with data for the closest grocery store, along with the calculated distance, in kilometers.

	SP Latitude	SP Longitude	CV Name	CV Latitude	CV Longitude	CV Category	CV ID	CV Distance
0	34.121789	-118.140670	Trader Joe's	34.135222	-118.147142	Grocery Store	4ae277bdf964a5207a8e21e3	1.583004
	ll output; double click to hide		Papa Georges Gary's Mini Market	34.136091	-118.131883	Deli / Bodega	4b8a2870f964a520406232e3	1.601040
2	34.126319	-118.175311	The Kitchen For Exploring Foods	34.140728	-118.178710	Food & Drink Shop	4bad4fd6f964a52003453be3	1.618181
3	34.126319	-118.163764	Asia World	34.136640	-118.153615	Grocery Store	5d278c4a9d81e3002380b813	1.429979
4	34.126319	-118.157990	Asia World	34.136640	-118.153615	Grocery Store	5d278c4a9d81e3002380b813	1.199572

At this point, analysis could begin. The first step was to create a histogram of the distances:

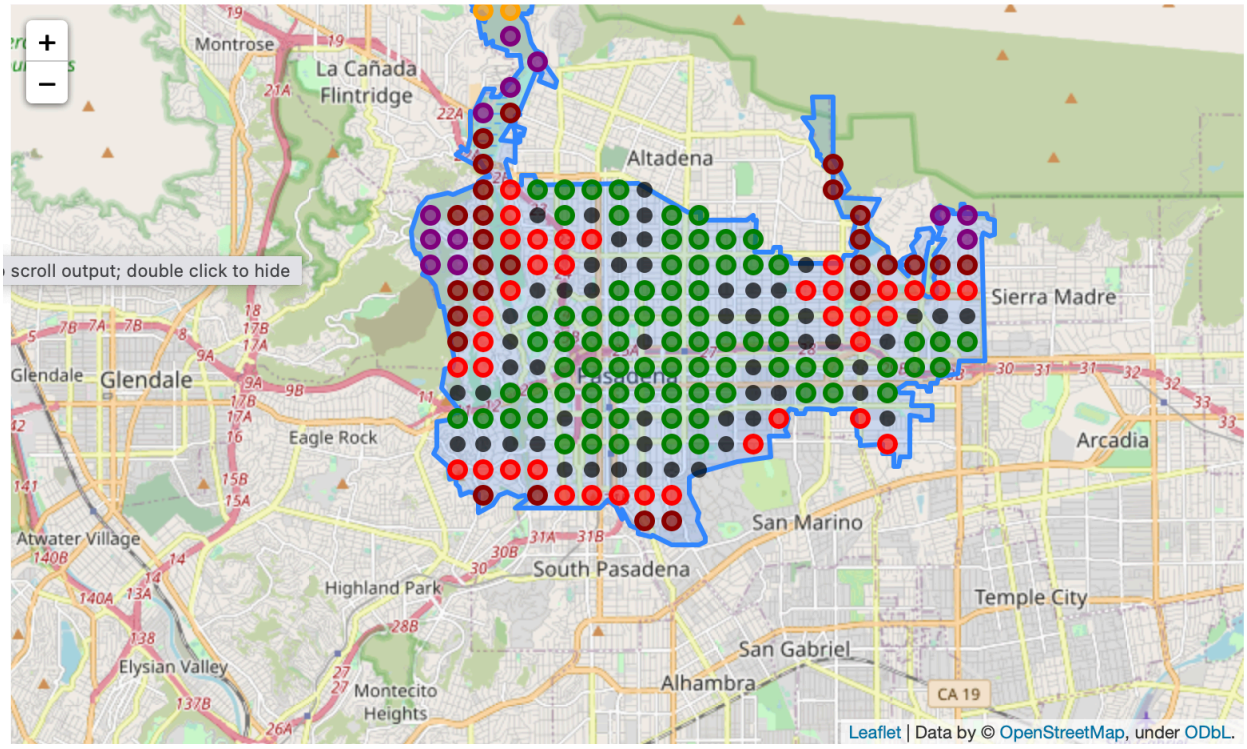


Clearly, most search points are within 2 kilometers of a grocery store, while some are not.

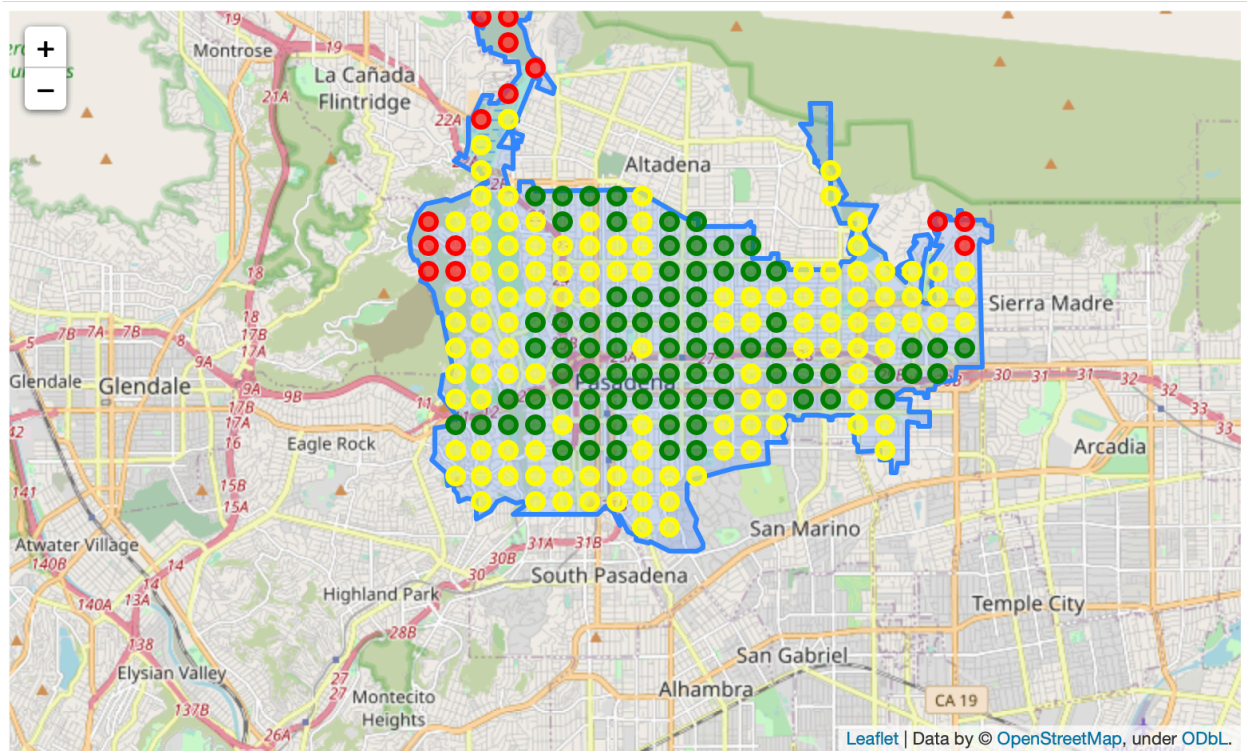
Using the values for distance, several statistics were generated:

Statistic	Value (in km)
Mean	1.002586
Standard of Deviation	1.182493
Minimum	0.008193
25%	0.328292
50%	0.651370
75%	1.243295
Maximum	7.653132

Finally, K-Means clustering was used to grade each of the search points. After some experimentation, 7 clusters were used. This is somewhat higher than commonly used, but he higher number revealed an interesting result, discussed in the next section. The values of the clusters were used to generate a new map Pasadena, with the search points color coded by their cluster value:



The colors were adjusted to improve readability, yielding this:



In this final map. areas in green are close to a grocery store, areas in yellow are a little bit farther, and areas in red are the farthest.

Results and Discussion

The core of the city of Pasadena (colored green) has adequate access to fresh food available at grocery stores. However, there are significant segments of the city that could have better access.

The area of the city farthest to the west (colored in red) does not have nearby access to a grocery store. However, this area would not be a good place to open a store, as it is dominated by nature parks or residential neighborhoods that are fairly wealthy and opposed to any kind of commercial development, even if it means a little bit of travel to buy necessities.

The slender region that extends north from the western part of the city is mostly taken up by NASA Jet Propulsion Laboratory. There is no need for a grocery store in this area.

Large portions of the city, colored yellow, are what we might call borderline underserved. Residents in these neighborhoods do not fit the technical definition of a food desert, but do have to travel farther than residents of the neighborhoods coded green. The western area that is colored yellow has many liquor stores and convince stores, but there are few to no places to buy fresh ingredients.

The area in the north east corner of the city is under served. Several years ago there was a store in this neighborhood, but it is now closed. If it were still open, this region would be adequately served.

The neighborhoods in the southern part of the city are also underserved, though the situation is a little more complicated than this analysis shows. If the analysis had included the city of South Pasadena (which borders Pasadena) it is possible that the southern neighborhoods of Pasadena would have been found to be well served.

The most interesting search point in the grid is right in the middle. As it happens this point is in the center of a busy transit corridor and commercial district. But it is also a neighborhood with dense housing, and a grocery store would probably do well there.

Conclusion

The purpose of this project was to identify neighborhoods in Pasadena, California, USA that are underserved by grocery stores, with the goal of identifying locations where a new one could do well. Although the city of Pasadena is, for the most part, well served by a variety of places to buy food, there are significant portions that are not. And, a single area in the center of town that is underserved.

The area in the north east corner of the city used to have a grocery store, but it was closed after a complicated corporate merger was completed (see Vons, https://en.wikipedia.org/wiki/Vons). A new store in this location could do well, as the neighborhood is underserved. By any definition, a resident of this neighborhood lives in a food desert.

A somewhat less obvious location is the center of town. At the intersection of two major thoroughfares there are no grocery stores.

Lastly, the area to the north west would be well served by another grocery store. Although the area is well served by liquor stores and and convenience stores, there is a distinct lack of stores selling fresh food.