

New Grocery Store in San Diego

Cong Zhang

1. Introduction

For everyone who wants to start their own business, the most concerned question is where to open their business. In this scenario, I am trying to find a location for a new grocery store in San Diego County, CA. The best location should be in the city that has the most potential customers (population of city/number of grocery stores in the city), also as a potential supplier, it should close to the most restaurants in this city (cluster all the restaurants in the neighborhood, the best location should be close to the centroid of the cluster that has the most restaurants). Hope it could bring some insights to future business owners when choosing their dream locations.

2. Data

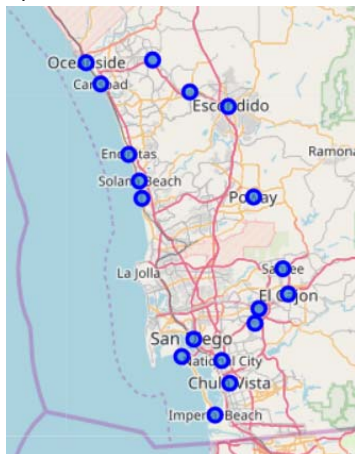
I found two useful datasets: 1) Population of California cities; 2) Geographic coordinates of California cities. From which I can easily extract information of each city in San Diego County, such as geographic coordinates and population of each city and combine as one dataset. With Foursquare API, I could find out the venue information of each city, especially grocery store amount in each city.

3. Methodology

- 1) Population -> figure out the population of each city in San Diego County.
- 2) Geographic coordinates + Foursquare -> find out the number of grocery stores in each city.
- 3) Best city location -> has the most potential customers -> 'New Store Share' (number of potential customers): $\text{population of the city} / (\text{number of grocery stores} + 1 \text{ new store})$ -> sort values -> the city with the largest value would be the best city as the new grocery store location.
- 4) Best location in the city -> close to the most restaurants in the city -> cluster all the restaurants in the city with Foursquare, the best location should be in the center of the largest cluster, i.e. the centroid.

4. Results and Discussion

- 1) There are 18 cities in San Diego County and mapped with **folium** (see below).



- 2) With **Foursquare API** and the geographic coordinates of each city, I fetched the venue information for each city, and extracted all the grocery stores in the whole county (see below).

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
257	Escondido	33.119208	-117.086422	Smart & Final Extra!	33.122506	-117.087514	Grocery Store
280	Imperial Beach	32.583944	-117.113086	Grocery Outlet	32.583659	-117.114633	Grocery Store
299	La Mesa	32.767828	-117.023083	VONS	32.765336	-117.024285	Grocery Store
341	Lemon Grove	32.742553	-117.031417	Sprouts Farmers Market	32.740505	-117.030014	Grocery Store
401	National City	32.678108	-117.099197	Chan Dong Oriental Supermarket	32.677389	-117.100612	Grocery Store
402	National City	32.678108	-117.099197	Mothers nutritional center	32.676662	-117.098464	Grocery Store
567	San Diego	32.715328	-117.157256	Grocery Outlet	32.711764	-117.155108	Grocery Store
586	San Marcos	33.143372	-117.166144	San Marcos Gym	33.142503	-117.166866	Grocery Store
663	Vista	33.200036	-117.242536	Frazier Farms	33.197798	-117.246055	Grocery Store
697	Vista	33.200036	-117.242536	Rodeo's Market	33.200669	-117.246968	Grocery Store

3) The number of grocery stores for the cities are listed (see below), for other cities not listed in, the number is 0.

	City	Grocery Store
0	Escondido	1
1	Imperial Beach	1
2	La Mesa	1
3	Lemon Grove	1
4	National City	2
5	San Diego	1
6	San Marcos	1
7	Vista	2

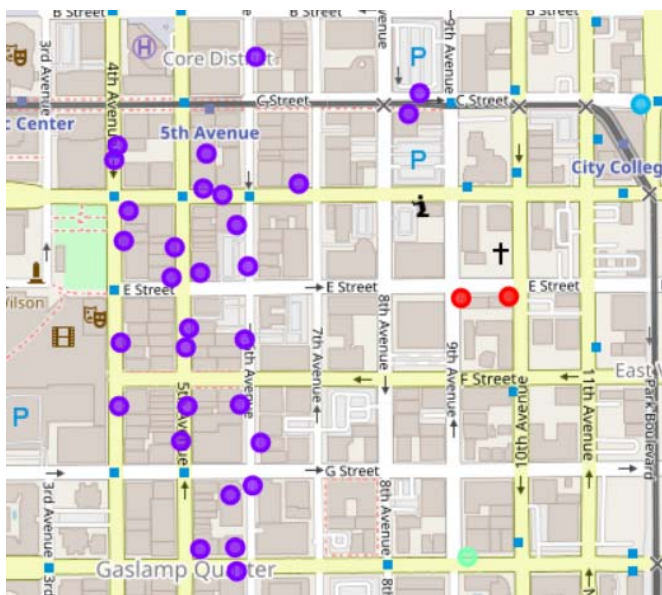
4) Potential customer amount: the 'New Store Share' value was calculated by dividing the population of each city by (the number of grocery stores of each city + 1). The data was sorted by 'New Store Share' in descending order. **The city of San Diego has the largest potential customers, thus it would be the best city for opening a new grocery store.**

	City	pop	Latitude	Longitude	Grocery Store	New Store Share
0	San Diego	1307402	32.715328	-117.157256	1	653701
1	Chula Vista	243916	32.640053	-117.084197	0	243916
2	Oceanside	167086	33.195869	-117.379483	0	167086
3	Carlsbad	105328	33.158092	-117.350594	0	105328
4	El Cajon	99478	32.794772	-116.962528	0	99478
5	Escondido	143911	33.119208	-117.086422	1	71955
6	Encinitas	59518	33.036986	-117.291983	0	59518
7	Santee	53413	32.838383	-116.973917	0	53413
8	Poway	47811	32.962822	-117.035864	0	47811
9	San Marcos	83781	33.143372	-117.166144	1	41890
10	Vista	93834	33.200036	-117.242536	2	31278
11	La Mesa	57065	32.767828	-117.023083	1	28532
12	National City	58582	32.678108	-117.099197	2	19527
13	Coronado	18912	32.685886	-117.183089	0	18912
14	Imperial Beach	26324	32.583944	-117.113086	1	13162
15	Solana Beach	12867	32.991156	-117.271147	0	12867
16	Lemon Grove	25320	32.742553	-117.031417	1	12660
17	Del Mar	4161	32.959489	-117.265314	0	4161

5) Since I decided to open the grocery store in San Diego city, the next question is where exactly the location would be. Ideally, the grocery store should be close to most of the restaurants, as a potential grocery supplier, thus, I gathered all the restaurant venue information of Foursquare from above, extracted all the restaurant locations (total is 34) in San Diego city, and clustered them into 4 groups using DBSCAN, and mapped them in different colors (see below).

(34, 7)

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
477	San Diego	32.715328	-117.157256	Pokez Mexican Restaurant	32.714574	-117.155761	Mexican Restaurant
480	San Diego	32.715328	-117.157256	Tajima Ramen East Village	32.714553	-117.156398	Ramen Restaurant
486	San Diego	32.715328	-117.157256	The Crab Hut	32.715802	-117.159871	Seafood Restaurant
490	San Diego	32.715328	-117.157256	The Melting Pot	32.714844	-117.159930	Fondue Restaurant
491	San Diego	32.715328	-117.157256	Empanada Kitchen	32.716635	-117.157114	Empanada Restaurant



6) From the map above, I could easily tell that the purple dot (cluster 1) is the largest cluster. The best location should be the center of this cluster, i.e. the centroid. Using **NearestCentroid** from **sklearn**, I got the centroid for cluster1 is: 32.71456327 -117.15607955.

5. Conclusion

The best location for the new grocery store is: 32.71456327 -117.15607955, in the city of San Diego.

Data Source:

California cities dataset by Cam Nugent

<https://www.kaggle.com/camnugent/california-housing-feature-engineering>

References:

'Clustering to Reduce Spatial Data Set Size' by Geoff Boeing

<https://geoffboeing.com/2014/08/clustering-to-reduce-spatial-data-set-size/>