Capstone Project - Report for finding Ice cream shop location

10/11/2019

1. Introduction

1.1 Background

Most people enjoy eating ice cream in different occasions. Each group of customers from each group of age favor different types of ice-cream products. For example, some people would like to have one type of ice-cream after shopping, or some people would like to meet in an ice cream shop to enjoy a refreshing ice cream with a group of friends and enjoy another type.

1.2 Business Problem

There are many ice-cream shops that many of them are not successful. They start their businesses, but they are not able to continue and close their businesses or change their direction. One of the main reasons of loss of these businesses is not selecting the right location.

1.3 Interest

The goal of this project will be to use data analysis techniques to help the investors to find the right location for an ice-cream shop in a neighborhood in an area in the city of Los Angeles.

2. Data acquisition and cleaning

The City of Los Angeles is very big, and it has very diverse neighborhoods. Additionally, there are many plazas, shopping centers, and recreational centers in the city.

We can use foursquare to locate the plazas, shopping centers, and recreational centers to spot the best possible places to start an ice-cream business while we should have an enough distance from the existing ice-cream shops as possible.

2.1 Gathering initial data

- For the process, first we will get all the longitudes and latitudes coordinates of the neighborhoods of city of Los Angeles by their street intersections using the data source from https://opendata.arcgis.com/datasets/0372aa1fb42a4e29adb9caadcfb210bb_9.csv.

Please note: to guarantee the existence of the data, the data has been downloaded and uploaded to github. Therefore, the data in the assignment will be loaded from below url.

https://raw.githubusercontent.com/Rafik-Ke/ForCapstone/master/Intersections.csv

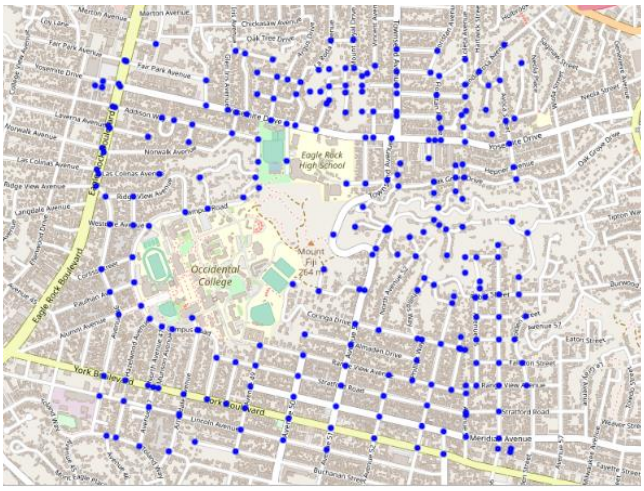
OBJECTID	ASSETID	CL_NODE_	X	Y	LAT	LON	TYPE	CRTN_DT	LST_MODI	USER_ID	FROM_ST	TO_ST	TOOLTIP	ZIP_CODE	NLA_URL
3001	98966	52918	6374563	1895188	34.19848	-118.619					VOSE ST	D/E	VOSE ST a	91307	navigatela/reports/intersecti
3002	98967	52920	6374037	1895146	34.19835	-118.621					VICKY AVE	D/E	VICKY AVE	91307	navigatela/reports/intersecti
3003	98968	52924	6364493	1894846	34.19736	-118.652					ST EDENS	D/E	ST EDENS	91307	navigatela/reports/intersecti
3004	98969	52932	6386270	1895673	34.2	-118.58					ENADIA W	D/E	ENADIA W	91306	navigatela/reports/intersecti
3005	98970	52945	6407912	1896310	34.20207	-118.509					CANTLAY S	D/E	CANTLAY S	91406	navigatela/reports/intersecti
3006	98971	52976	6388200	1897522	34.20511	-118.574					GIFFORD	D/E	GIFFORD	91306	navigatela/reports/intersecti
3007	98972	52988	6418380	1897255	34.20481	-118.474					VALERIO S	D/E	VALERIO S	91406	navigatela/reports/intersecti

Raw data

2.2 Data Cleaning

After getting the data, we will normalize the data according to the range of longitudes and latitudes that we desire to analyze.

For that purpose, we will drop the columns that we would not need in our analysis; and we will drop the rest of the data that is outside of the boundary of the area.



Part of Eagle Rock in Los Angeles city to

Please note: to have minimum reasonable calls to Foursquare the area is limited.

2.3 Finding business categories around each intersection using Foursquare

After normalizing, we will pass the coordinates to Foursquare. We will use a for loop to loop through all the intersections of the area and pass each intersection to Foursquare and get the categories of search result and save it in a data-frame. For example, passing an intersection to Foursquare will return all the categories in 200meter radius of the intersection. The goal is to find locations closer to shopping or recreational centers AND having the largest distances from the existing ice-cream shops.

category	Lat	Lon	Tooltip
Other Great Outdoors	34.132000	-118.208375	OLSON ST at D/E
Other Great Outdoors	34.132000	-118.208375	OLSON ST at D/E
High School	34.132000	-118.208375	OLSON ST at D/E
Flower Shop	34.132000	-118.208375	OLSON ST at D/E
University	34.132000	-118.208375	OLSON ST at D/E
College Residence Hall	34.132000	-118.208375	OLSON ST at D/E
Garden	34.132000	-118.208375	OLSON ST at D/E
Building	34.132000	-118.208375	OLSON ST at D/E
Pool	34.132000	-118.208375	OLSON ST at D/E
High School	34.131084	-118.204020	OAK GROVE DR at D/E

Business Categories around each intersection using Latitudes and Longitudes of each intersection

2.4 Cleaning and Normalizing Foursquare data

Large number of business categories we will get from Foursquare.

```
df_result.shape
```

```
]: (3268, 5)
```

Large number of businesses

We will remove the categories that we are not interested in opening an ice cream shop around it. For example, we will remove ATM since they might be neutral categories.

Please note: this is only for the assignment and business experts may decide, for example, to not remove ATMs.

category	
ATM	2
Advertising Agency	7
American Restaurant	63
Arcade	3
Art Gallery	37
Arts & Crafts Store	7
Asian Restaurant	4
Auto Dealership	1
Automotive Shop	123
BBQ Joint	11

Examples of categories that we can remove in "Red", and example of desired business in "Green"

category	Lat	Lon	Tooltip
Arcade	34.123754	-118.195284	FALLSTON ST at NOLDEN ST
Arcade	34.122168	-118.215020	LINCOLN AVE at HAZELWOOD AVE
Arcade	34.121936	-118.213808	AVENUE 47 at LINCOLN AVE
Clothing Store	34.121545	-118.202390	MERIDIAN ST at AVENUE 52
Clothing Store	34.126663	-118.215736	PAULHAN AVE at AVENUE 46
Clothing Store	34.129442	-118.215717	WESTDALE AVE at AVENUE 46
Clothing Store	34.130981	-118.216422	EAGLE ROCK BLVD at RIDGEVIEW AVE
Clothing Store	34.126083	-118.214175	CAMPUS ROAD at ALUMNI AVE
Clothing Store	34.130562	-118.214557	WOODROW AVE at RIDGEVIEW AVE
Clothing Store	34.120662	-118.202634	YORK BLVD at AVENUE 52
Clothing Store	34.123025	-118.200758	PHILLIPS WAY at RANGE VIEW AVE

Arcade, Clothing Stores, and Schools are desired (for this assignment)

3. Analyzing the categories

We group the categories to see what categories we have collected by Foursquare. We create two sets of data-frame, one for the categories that desirably we want to be closer, and one set that we want our business to be as far as possible, which are existing ice-cream shops. We join the two data-frame and eliminate the intersections that have ice cream shops around them.

7	Clothing Store	34.12608	-118.21417	CAMPUS ROAD at ALUMNI AVE	NaN	NaN
8	Clothing Store	34.13056	-118.21456	WOODROW AVE at RIDGEVIEW AVE	NaN	NaN
9	Clothing Store	34.12066	-118.20263	YORK BLVD at AVENUE 52	Ice Cream Shop	YORK BLVD at AVENUE 52
10	Clothing Store	34.12303	-118.20076	PHILLIPS WAY at RANGE VIEW AVE	NaN	NaN
11	Clothing Store	34.12573	-118.21531	AVENUE 46 at ALUMNI AVE	NaN	NaN
12	Clothing Store	34.12059	-118.20713	LINCOLN AVE at AVENUE 50	NaN	NaN
13	Clothing Store	34.12195	-118.20451	MERIDIAN ST at AVENUE 51	Ice Cream Shop	MERIDIAN ST at AVENUE 51
14	Clothing Store	34.12808	-118.21460	CAMPUS ROAD at CORLISS ST	NaN	NaN
15	Clothing Store	34.12240	-118.20215	STRATFORD ROAD at AVENUE 52	NaN	NaN
16	Clothing Store	34.12236	-118.20664	AVENUE 50 at MERIDIAN ST	NaN	NaN
17	Clothing Store	34.12937	-118.21380	WESTDALE AVE at CAMPUS ROAD	NaN	NaN
18	Clothing Store	34.12940	-118.21455	WESTDALE AVE at WOODROW AVE	NaN	NaN
19	Clothing Store	34.13006	-118.21089	CAMPUS ROAD at ESCARPA DR	NaN	NaN
20	Clothing Store	34.12019	-118.20500	LINCOLN AVE at AVENUE 51	NaN	NaN
21	Clothing Store	34.12759	-118.21616	AVENUE 46 at CORLISS ST	NaN	NaN
22	Elementary School	34.13351	-118.21165	ADDISON WAY at NORTH MAYWOOD AVE	Ice Cream Shop	ADDISON WAY at NORTH MAYWOOD AVE
23	Elementary School	34.12282	-118.19508	RANGE VIEW AVE at NOLDEN ST	NaN	NaN
24	Elementary School	34.13652	-118.20950	MONTIFLORA AVE at GLEN IRIS AVE	NaN	NaN

Joining two data-frames to remove the interest points that are close to Ice-cream shops

	category_x	Lat	Lon	Tooltip_x
0	Arcade	34.12375	-118.19528	FALLSTON ST at NOLDEN ST
1	Arcade	34.12217	-118.21502	LINCOLN AVE at HAZELWOOD AVE
2	Arcade	34.12194	-118.21381	AVENUE 47 at LINCOLN AVE
4	Clothing Store	34.12666	-118.21574	PAULHAN AVE at AVENUE 46
5	Clothing Store	34.12944	-118.21572	WESTDALE AVE at AVENUE 46
6	Clothing Store	34.13098	-118.21642	EAGLE ROCK BLVD at RIDGEVIEW AVE
7	Clothing Store	34.12608	-118.21417	CAMPUS ROAD at ALUMNI AVE
8	Clothing Store	34.13056	-118.21456	WOODROW AVE at RIDGEVIEW AVE
10	Clothing Store	34.12303	-118.20076	PHILLIPS WAY at RANGE VIEW AVE
11	Clothing Store	34.12573	-118.21531	AVENUE 46 at ALUMNI AVE
12	Clothing Store	34.12059	-118.20713	LINCOLN AVE at AVENUE 50

Dropping data that intersect with Ice cream shops and removing redundant columns

4. Optimizing the desired location

We will join the desired table with itself to find maximum number of our desired location in a location. For example, with having a school and Arcade game room and more entertainment in the area.

Category_^_^	Lat	Lon	tooltip_^_^	Category_^_y	tooltip_^_y
Arcade	34.12375	-118.19528	FALLSTON ST at NOLDEN ST	Elementary School	FALLSTON ST at NOLDEN ST
Elementary School	34.12375	-118.19528	FALLSTON ST at NOLDEN ST	Arcade	FALLSTON ST at NOLDEN ST
Elementary School	34.13652	-118.20950	MONTIFLORAAVE at GLEN IRIS AVE	Nursery School	MONTIFLORAAVE at GLEN IRIS AVE
Elementary School	34.13652	-118.20950	MONTIFLORAAVE at GLEN IRIS AVE	Nursery School	MONTIFLORAAVE at GLEN IRIS AVE
Nursery School	34.13652	-118.20950	MONTIFLORAAVE at GLEN IRIS AVE	Elementary School	MONTIFLORAAVE at GLEN IRIS AVE
Nursery School	34.13652	-118.20950	MONTIFLORAAVE at GLEN IRIS AVE	Elementary School	MONTIFLORAAVE at GLEN IRIS AVE
Elementary School	34.13643	-118.20902	MONTIFLORAAVE at HIGHLAND VIEW AVE	Nursery School	MONTIFLORAAVE at HIGHLAND VIEW AVE

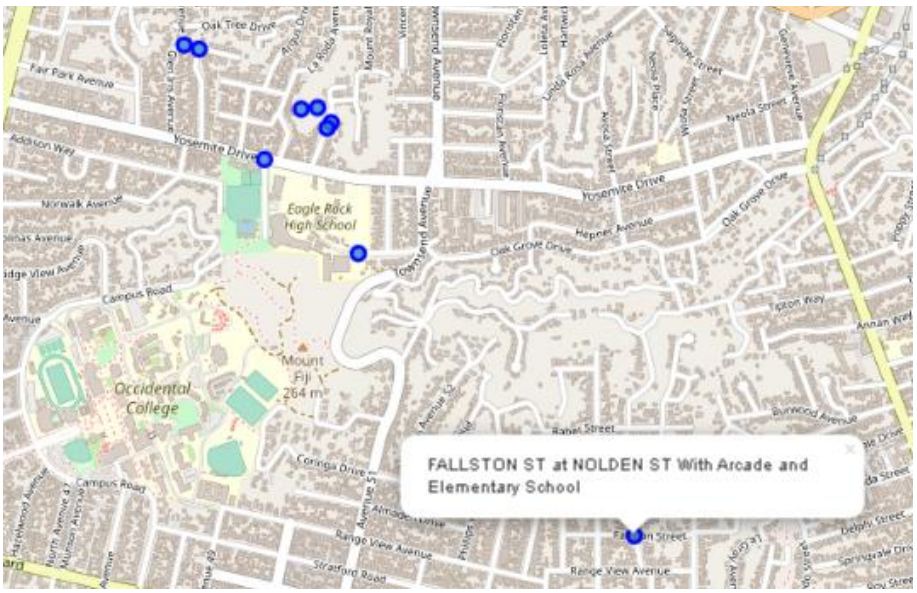
Joining the desired intersections to find the best locations

5. Conclusion

For this assignment, I used a data set for all the intersections of the city of Los Angeles, and selected an area of it in one of its sections since Los Angeles is a very large city. I limited the area since I could send limited numbers of call requests to Foursquare, and it happened couple of times that I exceeded the daily limit and I had to postpone the exercise to the next day. After I getting data from Foursquare, I saved the result in a file so I could use the saved data-frame.

I selected to have the ice cream shop closer to the locations that younger people would go more, such as Arcades, schools, and cloth shopping businesses.

Below are the points with the best location.



The Best locations shown on the map with detailed tooltips

Arcade	34.12375	-118.19528	FALLSTON ST at NOLDEN ST	Elementary School
lementary School	34.13652	-118.20950	MONTIFLORAAVE at GLEN IRIS AVE	Nursery School
lementary School	34.13643	-118.20902	MONTIFLORAAVE at HIGHLAND VIEW AVE	Nursery School
High School	34.13449	-118.20486	ALGOMAAVE at D/E	School
High School	34.13488	-118.20528	LINDA ROSAAVE at D/E	School
High School	34.13438	-118.20498	ALGOMAAVE at D/E	School
High School	34.13486	-118.20582	WARD ST at LINDA ROSAAVE	School
High School	34.13108	-118.20402	OAK GROVE DR at D/E	School
High School	34.13356	-118.20696	YOSEMITE DR at LA RODAAVE	School

The best locations coordinates