**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.

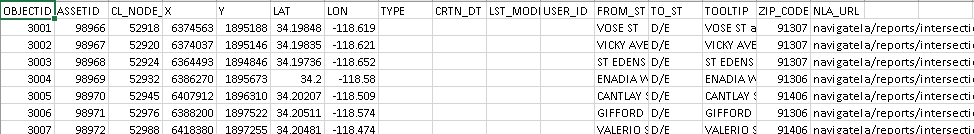
We can use foursquare to locate the plazas, shopping centers, and recreational centers to spot the best possible places to start an ice-cream business while we should have an enough distance from the existing ice-cream shops as possible.

**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>



**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.

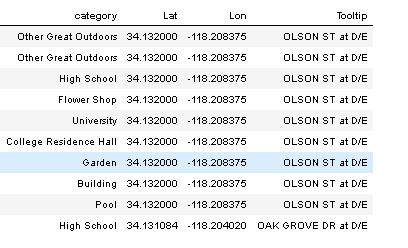
****

**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.

****

**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.

****

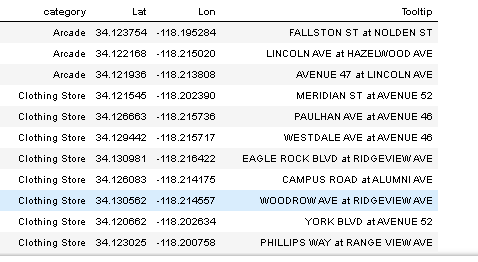
**Large number of businesses**

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.

****

**Examples of categories that we can remove in “Red”, and example of desired business in “Green”**



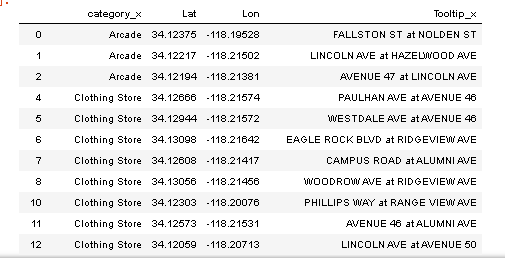
**Arcade, Clothing Stores, and Schools are desired (for this assignment)**

**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.

****

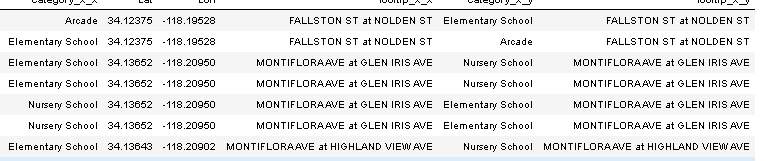
**Joining two data-frames to remove the interest points that are close to Ice-cream shops**



**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.



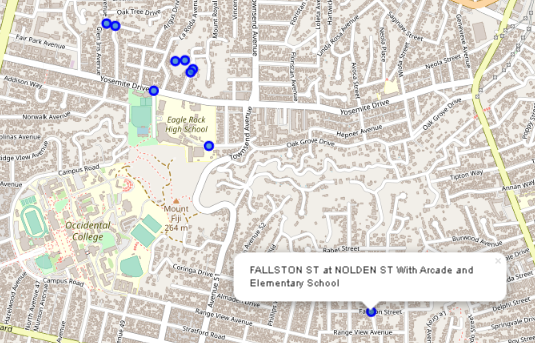
**Joining the desired intersections to find the best locations**

**5. Conclusion**

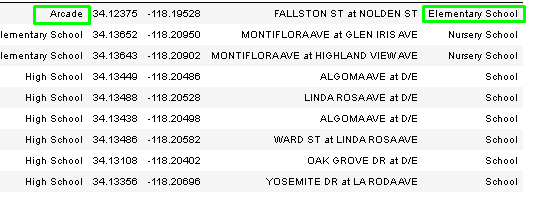
For this assignment, I used a data set for all the intersections of the city of Los Angeles, and selected an area of it in one of its sections since Los Angeles is a very large city. I limited the area since I could send limited numbers of call requests to Foursquare, and it happened couple of times 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**



**The best locations coordinates**