# Battle of the LA Cities

IBM Coursera Capstone Project - 2020

Eric Jung

## I. Introduction

### 1. Background

Ice cream is America's favorite dessert. According to [IDFA](#), the average American consumes more than 23 pounds of ice cream per year. With consumption levels this high, an ice cream business is bound to bring in massive profit.

### 2. Problem Statement

Due to its high population and scorching summer heat, Los Angeles (LA) County is an ideal place for entrepreneurs to start an ice cream shop.  However, LA County stretches far, and a subpar business location will result in low revenue at best. This project will demonstrate the best cities within LA county for starting an ice cream business.

### 3. Interest

The project is dedicated to stakeholders who are interested in starting an ice cream business, but are unable to decide on the location. After reviewing the results, stakeholders can confidently take one step forward into starting their business.

# II. Data

## 1. Data Acquisition & Description

The dataset was obtained through web scraping and various APIs.

- **City** – A table with the city name, incorporated date, and population as of 2010 can be found [here](). Only the city name was scraped for the dataset. The incorporated date would not have been a factor in choosing the best city. Although the population would have an impact on the decision, the numbers from 2010 are outdated and can provide inaccurate results.

- **Population** – A densely populated area means larger customer base. Previously mentioned, the data from 2010 is not an accurate representation of a city's population. Therefore, the US Census API was used in place of the 2010 data. The population is derived from a 5-year average, resulting in more reliable estimates. For this project, years ranging from 2014 to 2018 were used to approximate the cities' population. More information on the US Census API can be found [here]().

- **Income** – High-income residents have more to spend on luxurious items. Similar to the population data, the US Census API was utilized to obtain the household median income for each city.

- **Latitude & Longitude** – Google's Geocoding API was used to retrieve the cities' latitude and longitude. The coordinates will be utilized to visualize city location as well as find nearby competitors. More information on Google's Geocoding API can be found [here]().

- **Area** – A table with the city name, land area, and other features can be found [here](). Only city name and land area data were scraped. The table has land area in both square mile and square kilometer. In order to avoid unnecessary future conversions, only the land area in square kilometer was retrieved. The remaining features either had no impact on the business problem or contained outdated information. By itself, city area has little meaning to choosing the optimal location. However, the radius can be calculated from the land area.

- **Radius** – The radius was used to approximate a circle that envelops the city. Due to a city's irregular shape, this method is the best approach to find competitors within city limits.

- **Mean** – The Foursquare API was utilized to search for nearby dessert shops. For each city, the number of ice cream shops are divided by the total number of venues returned. This is to take into account the popularity of dessert shops within the vicinity. For example, a city with only two ice cream shops may seem ideal at first. However, absence of other dessert shops indicates residents' distaste towards desserts. Getting the percentage of the number of ice cream shops in relation to other dessert venues addresses these deceitful cases. More information on Foursquare API can be found here.

| | city | population | income | latitude | longitude | area | radius | mean |
|---|---|---|---|---|---|---|---|---|
| 0 | Agoura Hills | 20636 | 121896 | 34.153340 | -118.761676 | 20.2 | 2535.72 | 0.333333 |
| 1 | Alhambra | 84974 | 57117 | 34.095287 | -118.127015 | 19.8 | 2510.49 | 0.295455 |
| 2 | Arcadia | 58207 | 92102 | 34.139729 | -118.035345 | 28.3 | 3001.36 | 0.428571 |
| 3 | Artesia | 16817 | 63708 | 33.865848 | -118.083121 | 4.2 | 1156.24 | 0.382353 |
| 4 | Avalon | 3763 | 69440 | 33.342819 | -118.328228 | 7.6 | 1555.36 | 0.750000 |

*Table 1: Dataset First Five Rows*