# Introduction

The city of Boston, MA is one of the older and large cities in United States that has a history of over 400 years. Many people and cultures passed through generations and it still kept some of the old charm and traditions. Either due to religion or culture or community life, at some point in history, alcohol consumption was a restricted activity. For distributing and using alcohol, local government needs to provide license and track the effects of such license in the neighborhood. Though the license generates income, it still needs to weigh to local community sentiments, possible impact on traffic or crimes related to alcohol. With the city of many universities and younger generation of students, it requires considerable amount of evaluation to grant one such license.

# Business Problem

For this project, I am planning to analyze the number of license per zip code in Boston area and potential place to be considered as a better place among the zip codes of Boston based on the count of licenses and restaurants in that zip code.

This analysis might help a new restaurant owner to make a decision on serving liquor or expand business in other parts of city with that option.

*However, other data such as population with age and median income can provide more insights, starting with small data set and potentially add more data columns as analysis progresses.*

# Data Needed

For this project, need data on liquor permits issued per each zip code and top venues (restaurant) in that zip code within the area of Boston, MA:

* Zip code
* Number of licenses issued
* Venue Name
* Venue ID
* Venue Location
* Venue Category

*Possible data of demographics, population, age, median income, etc.*

# Data Acquisition Approach

The data can be obtained as below

* Liquor licenses data is taken from Boston city data site (boston.data.gov).
  + One geo location per each zip code (instead of one per license).
* Use Foursquare API to get a list of top 100 venues in each zip code
  + Get venue name, venue ID, location, and category

# Methodology

## Get data and clean-up

With the dataset obtained from boston city data site per each license, grouped it by the zip code and one geo location from that zip code is used.

For that each geo location, used Foursquare API to get the top venues to identify the restaurants. Assuming more restaurants has more possibility of liquor license.

Clustered the data within zip code to identify less licenses and less restaurants.

A potential location is identified with the above assumptions.

# Results

The data results showed neighborhoods with higher licenses have top venues and they are clustered together near the center of the city. Locations that are farther from city have less licenses issued.

# Recommendations

From the visualization of the map, the results seemed to be encouraging from license issuer perspective, such as higher in city with more population, not just residents, but, tourists, etc. However, it does not provide a good recommendation from an entrepreneur perspective. More data points such as population and other metrics will give a more closer prediction.

# Conclusion

This is a good starting point with data collection and analysis, but, more factors and parameters can improve this.