

# **Coursera Capstone**

IBM Applied Data Science  
Capstone

## ***Opening a Seafood Restaurant in Houston***

### ***Part-2: Data Description***

By: Vahid Sabzevari

March 2021



# 1. Data

## Required data to solve the problem:

- Houston city data containing the neighborhoods and districts.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to restaurants from Foursquare location data. I am going to use this data to perform further analysis of the neighborhoods.

## Data source and extract's methods:

Houston data containing the neighborhoods and districts, latitudes, and longitudes will be obtained from the data source:

<https://www.kaggle.com/mrchristolpher/houston-texas-neighborhoods-lat-long-list>. After it, I will get the geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package.

Finally, I will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases of 105+ million places and over 125,000 developers use this application. Foursquare API provides many categories of the venue data; I am particularly interested in the restaurant data to solve the business defined problem.

This project will require using of many data science skills, from web scrapping (open-source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium).