# Coursera Capstone (The Battle of Neighborhood)

### IBM Applied Data Science Capstone

***Opening an Ethical Chinese Restaurant in Dallas***

### By: Charles Odozi August 2020



**Introduction**

Dallas County demographics show that it is a large and ethnically diverse metropolis. **Dallas County** is a [county](https://en.wikipedia.org/wiki/County_(United_States)) in the [U.S. state](https://en.wikipedia.org/wiki/U.S._state) of [Texas](https://en.wikipedia.org/wiki/Texas), the state's [second-most populous](https://en.wikipedia.org/wiki/List_of_counties_in_Texas) county, and the [eighth-most populous](https://en.wikipedia.org/wiki/List_of_the_most_populous_counties_in_the_United_States) in the [United States](https://en.wikipedia.org/wiki/United_States). As of the [2010 U.S. census](https://en.wikipedia.org/wiki/2010_United_States_Census), the population was 2,368,139; in 2019 it was estimated to have 2,635,516 inhabitants.

This final project explores the best locations for Chinese restaurants throughout the city of New York. Potentially the owner of the new Chinese restaurant can have great success and consistent profit. However, as with any business, opening a new restaurant requires serious considerations and is more complicated than it seems from the first glance. In particular, the location of the restaurant is one of the most important factors that will affect whether it will have success or a failure. So our project will attempt to answer the questions “Where should the investor open a Chinese Restaurant?” and “Where should I go if I want great Chinese food?”

**Business Problem**

The objective of this Capstone project is to analyze and select the best locations in the city of Dallas to open a new Chinese restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in the city of Dallas, should the investor open a Chinese Restaurant?

## Target Audience of this project and some demographic facts

This project is particularly useful to developers and investors looking to open or invest in a Chinese restaurant in the city of Dallas. Overall, Dallas County is a great place to open a restaurant with an ethical cuisine. As Dallas is one of the most diverse city in the world. With its diverse culture, comes diversity in the food items. There are many restaurants in Dallas city, each belonging to different categories like Mexican, Indian, French, and African etc. Why did we decide to focus on Chinese cuisine in our project? Now when the idea of a healthy lifestyle conquered the minds of people all over the country, Chinese

Restaurants became extremely popular, as they offer a healthy alternative to regular American eating habits.

## Data

#### To solve the problem, we will need the following data:

* Dallas city data containing the neighborhoods and County.
* 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. We are going to use this data to perform further analysis of the neighborhoods.

#### Data Source and methods to extract them

Dallas county data containing the neighborhoods and county will be obtained from the open data source: <https://en.wikipedia.org/wiki/Dallas_County,_Texas> after it we will get the geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package.

Finally, we 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; we are particularly interested in the restaurant data to solve the business problem defined above.

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). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning techniques were used.