**IBM Data Science Professional Certification**

**Capstone Coursera Project Week 4. Part #1**

**Marco Saenz**

# Introduction

A new customer is wanting to invest in a Mexican cuisine restaurant in New York, a Mexican restaurant may prove to be a great investment opportunity if the location is right.

We will help the customer to find an optimal location area for his restaurant to be able address the opportunity of covering a need in a certain area.

# Problem Statement

We need to understand where the Mexican Restaurants and all other restaurants are in New York, then use a clustering algorithm to find similar areas in New York considering demographic data of each borough.

# Data description and use

We will be using 2 different datasets to merge into one dataset for the exercise.

1. New York Neighborhoods and Boroughs. <https://cocl.us/new_york_dataset> <https://geo.nyu.edu/catalog/nyu_2451_34572>
2. All related locations to Mexican restaurants will be obtained via the foursquare API.

We start by using the New York database and doing some Data Wrangling into a dataframe.

We proceed to create a map of Manhattan to analyze the neighborhoods and the most common venues.

We finally use K-means Clustering technique to cluster the neighborhoods into 5 clusters and map it with the most common venues per neighborhood related to Mexican Food restaurants.

Finally, we should be able to identify the best spot to open a new Mexican restaurant based on the Foursquare venue analysis and the New York location data.