**Capstone:** Finding the best neighborhood in Manhattan, NY for opening a Restaurant.

**Data Scientist:** Syed Khaleduzzaman

# Introduction:

A friend of mine just migrated to New York City, NY. He used to be the head chef in a famous restaurant in New Delhi, India. We both like to partner to open an Indian restaurant in Manhattan which is near to our current work place but not sure which neighborhood presents the best opportunity for us to be successful in the restaurant business.

# Problem Description:

Which neighborhood should we open our Indian restaurant in Manhattan, NY? We don’t like to take a chance since almost all of our life savings will be invested in the restaurant business. Before proceeding, we like to make sure our restaurant will be profitable from the get go.

# Requirements:

1. Our Indian restaurants needs to be strategically located with the highest concentration of restaurants in Manhattan, NY area. Patrons generally goes to areas to dine where multiple choices are available. People goes to shopping mall to buy cloths instead of local boutique store because people have more choices in the mall. Similar to that, people go out for lunch or dinner where clusters of restaurants exist instead of a standalone restaurant so they can choose among various restaurant choices.
2. It is the fact that financially affluent people go out frequently to dine at restaurant compare to people with lower income, therefore, we need to consider additional criteria before we can select a neighborhood. Not only we like to select a rich neighborhood but also a populous one so our restaurant can welcome as many patrons as possible.
   1. The average income for the Zip Code should be greater that the national average.
   2. Is density of the population greater than the **NYC’s** density of population of the comparing Zip Codes?
3. We need to confirm all the above assumptions before we can finalize a Zip Code to open our Indian restaurant.
4. In the end, by visualizing clusters of restaurants in the selected Zip Code we will be able to finalize our neighborhood to open the restaurant.

# Data Requirements:

1. **List of NYC (only Manhattan) Zip Codes, Population and Average Household Income:**

List of NYC Zip Codes with Population and Average Household Income can be found in the below webpage using **Beautiful Soup** package. <http://zipatlas.com/us/ny/new-york/zip-code-comparison/median-household-income.htm>

1. **Location Information:**

The location column of the above dataframe is having latitude and longitude in (lat, lng) format. Splitting the column using comma (,) delimiter will allow us to get latitude and longitude information for each Zip Code.

1. **List of Restaurant Venues in all Manhattan Zip Codes:**
2. We use 4SQUARE API to download all Restaurant venues for all Zip Codes.  
   [https://api.foursquare.com](https://api.foursquare.com/)  
   Extracted data will only include restaurants in our Data Set.
3. We will combine all the above information into a working Data Set. We will group by data by count of restaurant and sort the data buy the count as well as by the population and household income and determine the best Zip Code to open the restaurant.

# Machine learning Methods:

1. To finalize the neighborhood in the final Zip Code we will cluster the restaurants for the ZIP Code using KNN and will select the densest cluster as the neighborhood to open the restaurant.
2. We will map the selected cluster and map it to finalize the neighborhood to open the restaurant.

# Results:

1. Based on the data analysis, Zip Code **10021** has the **highest amount of household** income also is one of the most **highly populated neighborhoods** in Manhattan. Not only that, it has also the **highest number of restaurant venues**.

It can be noted that, although we have passed limit of 300, Foursqaure API is only returned limited number of venues.

1. Using K=5 in KNN algorithm, we crated **five clusters** for the venues returned by the Foursquare API for the Zip Code **10021.**
2. We analyzed all five clusters ad found that cluster one has the most concentrated restaurant clusters.
3. We mapped the cluster one and created a boundary line

# Discussion:

The main limitation for the entire analysis was that the Foursquare API returned limited number of venues (same number of venues for all the Zip Codes). If we could get the exact number of restaurants for each Zip Code it might happen that some other ZIP Code may have more restaurants than our selected Zip Code of 10021.

Obviously, Zip Code 10021 has the height amount of household income also very densely populated, therefore, either way the Zip Code 10021 is a solid choice for opening our restaurant.

# Conclusion:

The suggested location: The diagonal corridor between Park Ave/63rd Street & 2nd Ave/80th Street. Below is the map of the preferred location to open the restaurant.

