# CAPSTONE PROJECT COURSERA : FINDING LOCATION IN BROOKLYN TO OPEN UP A BAKERY

**By: Vaibhav Arora**

# INTRODUCTION

New York city is the financial capital of USA. On a daily basis many new business keep on opening and closing. A lot of factors need to be kept in mind when opening  a business. One of them is the appropriate location because numerous business fail due to poor location choice.

## PROBLEM STATEMENT

My client wishes to open up a Bakery in New York particularly in the Brooklyn area. As Brooklyn has a lot of neighbourhoods, he wishes to get the most appropriate location to open up his bakery.

After careful analysis we came upon a decision that we need to find the location that does not have many bakery but that that alone is not sufficient. For a bakery to perform successfully the area must show potential that opening a bakery there would be profitable. This potential of the areas will be determined by the top venues of that area. We want the top venues of that area to be something in the line of bakery or compliments it. So we will select those areas which have Café, Coffee shops and bars as their top venues keeping in mind that there must not be lot of bakery in that area as it would result in a tough competition.

# DATA

## DATA SOURCE

To perform this analysis, we will need the following data:

- A dataset that contains information about neighbourhoods of Brooklyn
- Venues's information of each neighbourhood.

Neighbourhoods data of Brooklyn will be obtained from dataset link: https://geo.nyu.edu/catalog/nyu_2451_34572

Top venues data will be obtained from Foursquare through an API.

# DATA EXTRACTION AND CLEANING

The data collected from New York csv file will be converted to a dataframe using pandas. Then from that dataframe a new dataframe will be extracted which contains neighbourhood information of Brooklyn.

Once it is done the co-ordinates of each neighbourhood will be extracted using geolocator and a dataframe will be created having the neighbourhood of Brooklyn and its latitude and longitude.

Then we will use Foursquare API to extract top venues of each neighbourhood and merge it with the dataframe containing neighbourhood information of Brooklyn.