**Coursera Capstone**

**IBM Applied Data Science Capstone**

***Opening a new Bakery in Delhi, India***

By: Vaishnavi Gupta

July 2020

**INTRODUCTION**

For many people, bakeries are a good place to get the best and exotic delicacies from. They can have various food items that the bakeries provide and satisfies everyone’s sweet tooth. With the improving lifestyle of people, many prefer to opt for good bakeries to satisfy their tummies and with the arrival of new delicacies and sweets from across the world, opening a bakery store has its own brilliant financial returns.

Baking enthusiasts are keener than ever to tap this growing market that has more demand now than ever before. As a result, there are many bakeries in the city of Delhi and many more are being built. Opening bakeries allows property developers to earn rental income as well.

Of course, as with any business decision, opening a new bakery also requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the bakery is one of the most important decisions that will determine whether it will be a success or a failure.

**Business Problem**

The objective of this capstone project is to analyse and select the best locations in the city of Delhi, India to open a new bakery. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question that where we should open a new bakery.

**DATA**

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

* List of neighbourhoods in Delhi. This defines the scope of this project which is confined to the city of Delhi, the capital city of the country of India.
* Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
* Venue data, particularly data related to bakeries. We will use this data to perform clustering on the neighbourhoods.

**Sources of data and methods to extract them**

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods\_in\_Delhi) contains a list of neighbourhoods in Delhi, with a total of 138 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.