The Battle of Neighbourhoods
Where To Open Bakery?

1. Business Problem

The Objective of this Capstone Project is to located the best Location to open new Bakery Outlet in City New Delhi, India. Using Data Science and Machine Learning Technique, this Project aim to answer this Business problem. With this help of Machine learning we are able to sorted out the location which is well connected to Airport/Bus Station and other public places. The Solution recommend best neighbourhood to open the Bakery.

2. Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new Bakery in New Delhi, India. This project is timely as the city is currently suffering from oversupply of Bakeries. This project will lead to proper availability of Bakery througout the city. It will result in the benefit of the Investors and the customers.

3. Data Description

The data set that I have used for solving the problem is:

- A complete list of neighborhoods in New Delhi, India. Source of the data is Wikipedia.org (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi))
- Geographical coordinates (latitude and longitude) of those neighborhoods. Source of the data will be FourSquare.
- FourSquare provided Venue data which is related to Bakery. Machine Learning Technique called Clustering will be used for solving the problem.
- Clustering using K-Means clustering algorithm

This wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi) contains a list of neighborhoods in New Delhi, with a total of 137 neighborhoods. 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 neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbors. After that, we will use the Foursquare API to get the venue data for those neighborhoods.

4. Methodology

Methodology for finding a suitable location for opening a new Bakery in Delhi, India is based on Clustering of venues and places in Neighbourhoods of Delhi. I have grouped the similar venues together on the basis of availability of venues of the different categories. I have used Machine Learning technique called Clustering for the analysis of venues and places of different categories in the Neighbourhoods of Delhi. First step is pulling and preprocessing of the data. In have used web scraping python library for pulling the data from wikipedia pages and then I have performed the preprocessing of the data using python data munging library called pandas.
Second step is all about Exploratory Data Analysis, where I have used few statistical measures and data visualization techniques for understanding the internal structure of the data and the relationship between different parameters of the data. Exploratory Data Analysis has helped me to decide which machine learning technique will be suitable for solving the business problem. I have also used map visualization library (folium) and python library geocoder for fetching the geographical coordinates of different venues and places. I have used Machine Learning technique called Clustering.
Clustering unsupervised machine learning where have unlabelled dataset.