The Battle of Neighborhoods
Where To Open Bakery?

BY
RAJAT GUPTA

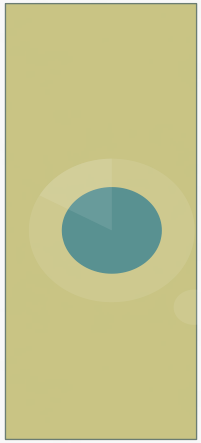

TABLE OF CONTENTS

BUSINESS PROBLEM

- Find the best location for opening a new Bakery in Delhi (India).
- Decision making for Investment into Bakery and real estate business.
- This project is particularly useful to property developers and investors.
- This project is timely as the city is currently suffering from oversupply of Bakery
- This project help the Investor to locate the best suitable well connected location to open the bakery

DATA DESCRIPTION

- 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

TECHNOLOGIES USED

List of tools and technologies used:

- KMeans clustering algorithm used for categorizing the venues.
- Python programming for Implementation of the project.
- Pandas library for preprocessing of the data.
- Scikit-learn for applying machine learning.
- Geocoder for fetching latitude and longitude of the venues.
- Folium for the visualization of venues on the map of Delhi

IMPLEMENTATION

Steps involved in the end-to-end Implementation of the project:

- Exploratory Data Analysis for understanding the dataset.
- Preprocessing for cleaning of the data.
- Feature Selection for selecting the best feature for the model.
- Machine learning model development.
- Developed KMeans clustering model.
- Cluster analysis for selecting best places for opening a new Bakery.

OBSERVATIONS

Cluster Name	Suitable for new Bakery	Detail Description
First Cluster	✗	• Lower to moderate availability of Bakery. • Mid size cluster
Second Cluster	✓	• Very Low/None availability of Bakery. • Big size Cluster. • Good well connected location.
Third Cluster	✗	• Availability score of the Bakery is very high. • Cluster consists of only two places.
Forth Cluster	✗	• Availability of Bakery is good if not the best. • Mid size Cluster
Fifth Cluster	✗	• Availability score of the Bakery is very high. • Cluster consists of only two places

FUTURE ENHANCEMENTS

- Collect more data (demographic, income, investments etc)
- Combine multiple models for better decision making
- Consult from domain experts of Real estate and Investments
- Better chance Revenue generation for Investor.