

Data Science Capstone Report:

The Battle of the Neighborhoods

By : Uday Gupta

July , 2019

Introduction

Shopping malls serve as a great place for recreational activities and thus for various businesses to flourish, develop and make profits. There are restaurants, cafes, multiplexes, shopping outlets, grocery shops, arcades and much more. Shopping malls often target the people of one or more localities by centralizing their location amongst the localities to get the right amount of inflow of customers with low competition with other vendors or shopping malls. Therefore one should consider the location to be the most influencing factor when building a new shopping mall for it to be a success or failure.

Business Problem

The objective of this capstone project is to analyze all neighbourhoods of Delhi, India and select the best ones to open a new shopping mall. We will use data science methodology and machine learning techniques like clustering. The problem statement that can be solved using this project is:

“In the city of Delhi, India if a property developer wants to open a new shopping mall, what locations can be recommended for it?”

Target Audience

This project is very useful for property dealer and investors looking for new profitable localities to open new shopping malls in the city of Delhi, India. There is significant overcrowding, oversupply and unwanted competition among malls in some significant neighborhoods of Delhi while others don't have enough for basic supply.

Data

We shall need the following data:

- A dataset containing the names of all the neighborhoods in Delhi along with their latitude and longitude coordinates.
- We will use the coordinates to find venue data for each locality and to plot them on the Delhi map.
- We will use the venue data to cluster the neighborhoods based on shopping malls.

Data sources:

- The Delhi neighborhood data was downloaded from the Kaggle website: <https://www.kaggle.com/shaswatd673/delhi-neighborhood-data/data#>
- After getting the neighborhoods with their respective coordinates from the link above, we'll use the Foursquare API to get the venue data for each neighborhood. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. The API will provide many venue categories but we require only the shopping mall category for each neighborhood in order to proceed.