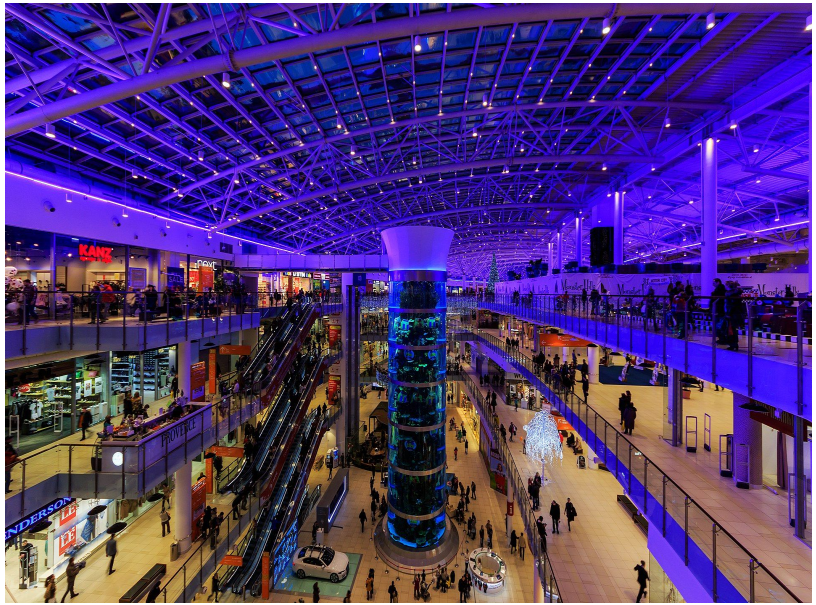
# IBM Applied Data Science Coursera Capstone Project

Opening a new Shopping Mall in Delhi NCR - India

Yashwanth P
March - 2019

# Shopping Mall



Source - Wikipedia

A *Shopping Mall* is a modern term for a multi level Shopping Precinct or Shopping Center in which one or more buildings form a complex of shops representing merchandisers.

Typical Components include

a) Food Courts
b) Department Stores
c) Movie Theaters

Yashwanth P

# Delhi NCR



Source - Wikipedia

*Delhi*, officially the National Capital Territory of Delhi is the capital of India. The National Capital Region (NCR) is Delhi's urban area which also include the satellite cities of *Faridabad, Gurgoan, Ghaziabad & Noida*.

NCR has an area of around 1,484 sqaure KM (573 sq mi). The population of NCR is estimated to be over 26 Million people, making it the 2nd largest Urban area according to the United Nations.

It's also the 2nd most productive metro area of India - home to 18 Billionaires and 23k Millionaires

Yashwanth P

# Problem Statement & Target Audience

Delhi being a very vast location and with a very huge population, it makes it difficult to place the location of a new Shopping Mall, given that it might already have a few.

The project aims to leverage Clustering and Machine Learning to find different locations where a new mall or malls would actually be a great fit in North Capital Region (Delhi and its surrounding cities included).

***Target Audience*** - Any property investor who would want to build any type of Public Entertainment. The analysis would also give us an understanding of the locations for other venues such as restaurants, pubs and nightclubs etc.

# Data Sources

1. Postal Code - The Indian Postal Data is present at (https://www.indiapost.gov.in/vas/pages/findpincode.aspx) as a csv file which would enable us to identify all the postal codes in NCR. Postal Codes in India are a 6 - Digit Code (Ex. 122002)
2. Foursquare API - The foursquare data would enable us to understand the postal code area venues which would become the base for our clustering
3. Geocoder data - To get the Latitude and Longitude of the Postal Codes given


Thank
You