# Coursera Capstone
## IBM Applied Data Science Course

## *Opening a new Multiplex in Kolkata*

By:Ritesh Manna
Date:February 2020

# Business Problem

- Aim: To analyse and select the best locations in the city of Kolkata to open a new multiplex.

- Question: In Kolkata if a property developers is looking for a location to open a new multiplex what would you recommend?

# Data

- Data requirements:
  - List of neighbourhoods in Kolkata
  - Latitude and Longitude of the neighbourhoods
  - Venue data, especially data related to multiplexes

- Data sources:
  - Wikipedia of neighbourhoods: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Kolkata
  - Geocode package
  - Foursquare API

# Methodology

- Web scraping of Wikipedia page to get neighbourhoods
- Getting latitude and longitude from Geocoder package
- Using Foursquare API to get venue data
- Group data by neighbourhood and taking mean of frequency of occurrence of each venue category
- Filter venue category by Multiplex
- Perform clustering on data using K-means clustering
- Visualize the clusters in a map using Folium

# Results

- Cluster 0: Neighbourhoods with zero to low number of multiplexes

- Cluster 1: Neighbourhoods with a high number of multiplexes

- Cluster 2: Neighbourhoods with a moderate number of multiplexes

# Conclusion

- The neighbourhoods in cluster 0 should be the preferred option to open a new multiplex.

- Neighbourhoods in cluster 2 are likely to suffer from intense competition and should not be the ideal location in opening a new multiplex.

- Property developers with unique selling propositions can try their hand in opening a new multiplex in regions represented by cluster 1.