

IBM Capstone Final Project

Opening a new Cinema Hall in Mumbai, India

By: Shreya P. Jain

October 2020



Introduction

Cinema is a major source of recreation in most countries around the world, even in India. Multiplex segment is growing while single screen segment is declining. As of March 2005, there were approximately 13000 cinemas in India out of which 73 were multiplexes with total 276 screens.

Multiplexes constitute only 0.6% of about 12000 cinema halls, but account for 28% to 34% of box office collection for the top 50 films.

People are turning towards multiplexes due to various reasons, some of which are safety, better ambience, eateries, etc.

For many movie lovers, visiting multiplexes is a great way to relax and enjoy themselves during weekends and holidays. Multiplexes are like a one-stop destination for all types of movie-watchers. The location of multiplex is one of the most important decisions that will determine whether the multiplex will be a success or a failure.

Business problem

The objective of this capstone project is to analyse and select the best locations in the city of Mumbai, India to open a new

multiplex mall. Using data science methodology and machine learning techniques like clustering. This project aims to provide solutions to answer the business question: Where should a property developer or investor open a multiplex in the city of Mumbai, India?

Data

To solve the problem, we will need the following data

- List of Neighbourhoods in Mumbai. This defines the scope of the project, which is confined to the city of Mumbai.
- The latitude and longitudes of those neighbourhoods, required to plot the map, as well as get the venues data.
- Venues data, particularly data related to multiplexes, cinema halls, theater, etc.

The Wikipedia page

(https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai#Mumbai_neighbourhood_coordintes) contains the list of neighbourhoods and their latitude and longitude. We will use pandas for web scraping.

After that, we will use Foursquare API to get the venue data for those neighbourhoods.