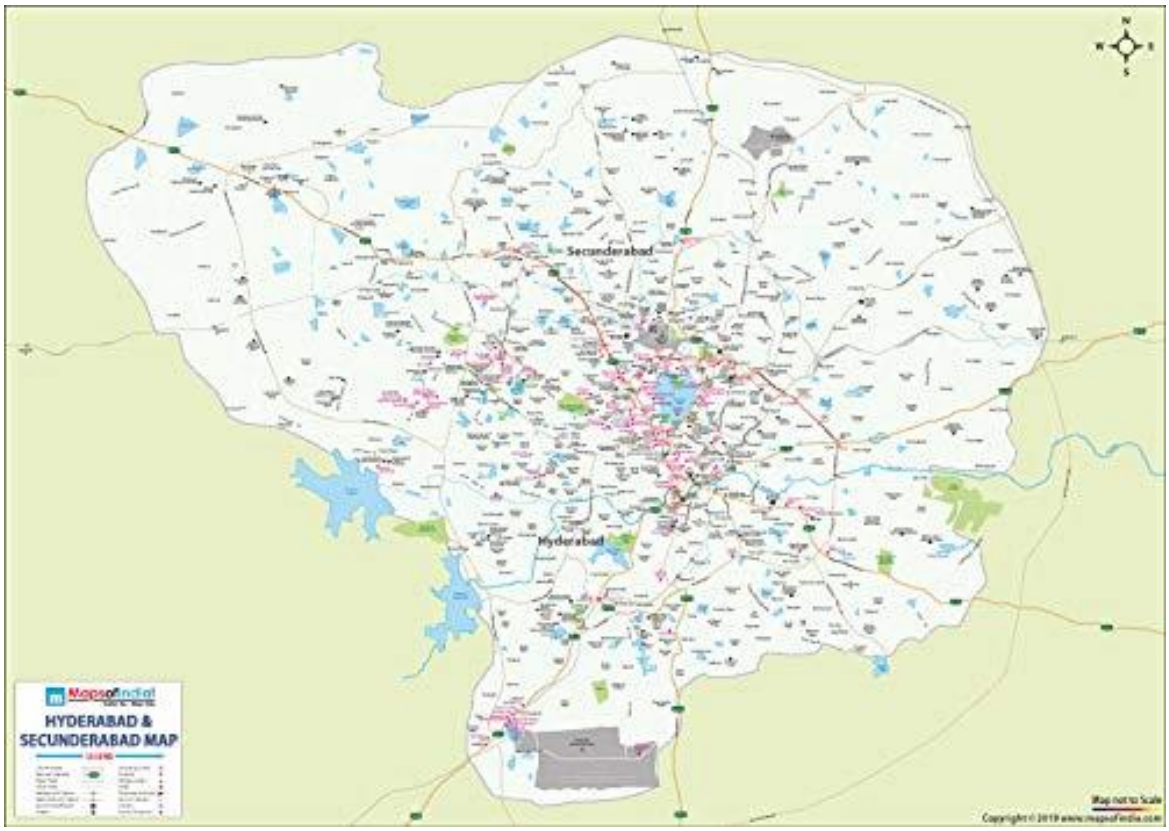
Coursera Capstone IBM Applied Data Science Capstone

Opening a New Multiplex in Hyderabad, India

By: Sharath Kumar V

Nov 2019



Introduction

Multiplex segment is growing while single screen segment is declining. As of March 2005, there were approximately 13000 cinemas in India of which 73 were multiplexes with a total of 276 screens. Multiplexes constitute only 0.6% of about 12000 cinemas halls, but account for28% to 34% of box office collection for top 50films in 2004(Yes bank, 2004). Growing film industry is the key driver for generating more footfalls for film exhibition industry growing faster (IBEF,2013). The modern shopping malls offer variety of entertainment services, life style products, gaming hubs, food courts and cinemas(Ibrahim and Ng, 2002; Friedberg, 1993). Shopping trips can have many purposes (O'Kelly,1981). If we consider the view point of Davies (1995) people enter in a theatre or cinema for leisure. People are turning towards multiplexes due to various reasons, some of them are safety, better ambience, eateries, security etc. (ooi and sim,2007). One will be surprised to see the number of women walking out of multiplexes after nightfall but saying that with the boom of multiplexes it is not a happy time for single screen theatres. When it comes to enjoying a movie with a woman companion most of the people prefer multiplexes (Eliashberg etal.,2005).

The purpose of this is to highlight the various aspects which lead to the preference of multiplexes over single screen theatres. The survey would be conducted across a minimum of ten malls across Hyderabad.

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. To enhance the attraction of the shopper malls investors will start multiplexes also promote the many products as adds as part of the movie show. The location of the 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 Hyderabad, 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: In the city of Hyderabad, India, if a property developer is looking to open a new multiplex, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new multiplex in the one of the fastest growing capital of Telangana, India i.e. Hyderabad. This project is timely as the city is currently suffering from oversupply of Multiplexes due to most of the film cities around the city. Some research gone to understand where to start the multiplex and how to enhance the attraction in the city, but many are failed to create the attraction, this problem statement may help to the financial investors and developers to select the right place to start the multiplexes.

Data

To solve the problem, we will need the following data:

• List of neighbourhoods in Hyderabad. This defines the scope of this project which is confined to the city of Hyderabad, the capital city of the Telangana and one of the fastest growing cities in India

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and to get the venue data.

• Venue data, particularly data related to Multiplexes. We will use this data to perform clustering on the neighbourhoods. Sources of data and methods to extract them This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) contains a list of neighbourhoods in Hyderabad, with a total of 200 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used