Coursera Capstone Project

IBM Applied DataScience Capstone

Opening a new shopping center in Hyderabad, Telangana

Sathwik Vadlamani

May 2020

Introdution

Hyderabd is famous for its centuries-old monuments, but the rise of numerous malls in Hyderabad has added a stylish edge to the economic powerhouse. Well-heeled travellers can choose from various swanky malls housing world-renowned fashion labels, global restaurants and cinemas. If you have the budget to back you, then get ready to splurge at these malls during your next visit to Hyderabad.

Retailers in Hyderabad have displayed a preference for hosting their brands in high-street outlets over shopping malls of late. The trend is growing on the back of locational advantage, lower rentals and personalised experience that exclusive stores offer to customers.

Business Problem

The Objective of this capstone project is to analyse and select the best locations in the city of Hyderabad, Telangana to open a new shopping center. Using data science methodology and machine learninng techniques like clustering. This project aims to provide solutions to answer the bsiness question: In the city of Hyderabad, Telanagan, if a property developer is looking to open a new shopping center, where would you recommend that they open it?

Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new shopping centers in the capital city of Telangana i.e Hyderabad. A recent article from the *The times of India* states that

> Evolving consumer spending patterns and increasing disposable income are redefining the retail landscape in Hyderabad. Retail brands which earlier preferred shopping malls are now vying for shops in high-street complexes owing to lower rental rates, higher brand visibility and relatively higher sales than their counterparts in malls.

Data

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

- List of neighbourhoods in Hydearabad. This defines the scope of this project which is confined to the city of Hydrabad, the capital city of the state of Telangana in India.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This wikipedia page contains a list of neighbourhoods in Hyderabad, with a total of 220 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 database 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.