Coursera - IBM Data Science Capstone Project

The Battle of Neighborhoods

Introduction

Chennai, the capital city of the state of Tamil Nadu, India, is the largest industrial and commercial center of South India. Recent estimates of the economy of the Chennai Metropolitan Area have ranged from US$79 to US$86 billion, ranking it from fourth to sixth most productive metro area of India and the third highest by GDP per capita.

Chennai remains the Chief Retail Industry and Shopping Centre in South India, with some of its suburbs serving as exclusive shopping districts. Since the formation of the city in the seventeenth century, George Town remains one of the chief commercial neighborhoods of the city. However, with the centuries passing, the central business district of the city started shifting towards the south of Fort St. George and moving to its present location at Gemini Circle. The city's retail industry is concentrated chiefly in T. Nagar, which is by far the largest shopping district of India, generating more than twice the revenue of Connaught Place in New Delhi or Linking Road in Mumbai, even by conservative estimates.

In Terms, One-stop destination for all shop solution, people move on to shopping malls in order to cover wide range of products. Retailers and Property Developers eyes for the central location and large crowd at the shopping mall to provide the retail channel to market their products and services. Opening shopping malls is herculean task which requires serious consideration and need an effective modeling and planning to come out of effective solution. Core point of the solution, the location of the shopping mall.

Business Problem

Key objective of the project is to focus on the target audience and help them to bring out better solution for their problems. This project is particularly helpful for property developers, retailers and investors to open a new shopping mall in the neighborhoods of Chennai City. This project is a timely solution for future developers, according to the 2019 Cushman & Wakefield report *Main Streets Across the World*, Khader Nawaz Khan Road at Nungambakkam ranked 10th position in the list of 'Top 10 Global Highest Retail Rental Growth Markets 2019', with 36.7 percent increase in rents.

Using Data Science Methodology and Machine Learning Techniques, it provides business solutions by analyze and cluster the neighborhoods and bring out the central location to build a shopping mall.

Data

The data for this project has been retrieved and processed through multiple sources by carving out exact considerations to the accuracy of the methods used.

Data Acquisition and Cleaning

1. Data Acquisition

The data acquired for this project is a combination of two major sources
First source of the project uses a **Wikipedia** Page
(https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Chennai)
Use of Web scraping techniques to extract data from Wikipedia page with help of Python requests and package. With help of Python Geocoder packages, the geographical coordinates of Chennai neighborhoods along with latitudes and longitudes.

Second Source, **FOURSQUARE** API to get the venue data of the Chennai City by mainly focusing on the Shopping mall category by data cleaning, wrangling and Machine Learning techniques (K- Means Clustering) and Map Visualization (Folium).