Course Era Capstone

IBM Professional DATA SCIENCE Capstone

Jan 2020

*“Want to open a New Multiplex in Tampa? Here's what data says”*

Submitted by

Sai Suraj Argula



**Introduction**

In the era of fast-moving world, Movie theatres play an important role of providing relaxation to General people. Though online streaming services like Netflix, Hotstar and Amazon prime taken over a part of finances and profits in the entertainment Industry, movie theatres still be able to bring crowds who prefer experience compared to convenience. Movie theatres become one stop destination for entertainment, food and beverages. Increasing number of movies, along with increase in millennials who prefer theatres, demand for new movie theatres in the neighborhood which is accessible increased.

**Business Problem**

Foreseeing increased demand of New Movie theatres, company XYZ Entertainment wants to expand by investing in construction of new theatres. As a pilot project XYZ Entertainment choose City of Tampa as a start point to identify possible places to construct theatre. XYZ Entertainment needs the theatres to be constructed in the neighborhoods which are away from existing theatres, so that they can avoid possible competition.

**Approach**

As company wants to identify the neighborhoods which doesn’t have and have existing theatres. Analytical team decided to use K means clustering to cluster the neighborhoods into three groups of having more, moderate and less concentrations.

**Data Requirement**

For the data analysis team has come to brainstom the data needed for project, identify neighborhoods of Tampa along with its location data (longitude and latitude). Data of the existing theatres with its location.

**Data Collection**

Data required regarding the neighborhoods has been collected from Wikipedia website of all the residential neighborhoods in the city of Tampa (<https://en.wikipedia.org/wiki/Neighborhoods_in_Tampa,_Florida>). Python libraries packages like requests and Beautiful Soup helped in web scraping the page. For locating the neighborhoods location data i.e., latitude and longitude python library of Geocoder is used. All the information is tabled into a pandas DataFrame with neighborhood, latitude and longitude.

Foursquare City square API(<https://foursquare.com/city-guide>) has been used to locate the nearby venues (movie theatres in particular) in and around each of the neighborhoods.

This project has been built from scrap, using many data science skills like webscraping, data cleaning, machine learning, data visualisations.