

Coursera Capstone Project
IBM Data Science Professional



**New Movie Theaters in San Francisco,
California**

By: Kyle Tran
August, 2019

Introduction/ Background

Movie theaters have become an integral part of the way people live since 1905. The first theater originated in Pittsburgh, Pennsylvania, and it revolutionized the way we spend our free time around the world. Short silent films eventually evolved into the motion pictures, which led to significant advancements in film production became the foundation in many technological equipments and techniques that are being used today. As the popularity of movies increased the market, companies began to invest money into expanding more locations around the world. The idea of choosing the proper location for a business is time-consuming, stressful, and tedious. A company needs to know the amount of foot traffic, competitors, and etc. in order to ensure the success and future of the business. Owners need to know the location of movie theaters in order to separate themselves from others. The goal of opening a new movie theater is to avoid nearby competition and increase the amount of foot traffic, because it is hard to research which area in San Francisco would be acceptable.

Business Problem

The purpose of my capstone project is to examine which locations in San Francisco, California would be fit to open a movie theater.

Question:

Which area in San Francisco, California would be recommended for project developers to open up a new movie theater?

Data

The data that would be needed for this project would be the locations of each area(along with latitude and longitude), latitude and longitude of movie theaters nearby, and various information of movie theaters. The source is from a wikipedia

page(https://en.wikipedia.org/wiki/San_Francisco_Bay_Area) that contains tables of 10 areas. Web scraping techniques from before will be used to extract the data we need using BeautifulSoup and Python. Afterwards, the location(latitude/ longitude) will be provided for us through geocoder.

Foursquare API provides users social location services that provides information of the venues to 150,000 developers. The project will use the service to obtain the data of movie theaters and their respective locations for plotting each area. We will combine the data extracted by BeautifulSoup to retrieve the locations through geocode. After cleaning/ wrangling the data, the FourSquare API will plot locations of movie theaters, and it will be clustered using KMeans. Perspectively, this process will portray the Data Science methodology and present an accurate analysis of the data.