Coursera Project: Capstone

Where to open a new stadium in Shanghai, China

By: Zhe Chen

Jan. 2020

Introduction

For many residents, stadium is a great place to have tons of fun. They can enjoy the games, do exercise and even find restaurants and do shopping. Nowadays, modern stadiums are more like a shopping mall with sport functions luring all types of people. For the shopkeeper, the attraction of stadium bringing the large crowd provides a great distribution channel to sell their products and services. The stadium developers are also taking advantage of this trend to build more stadiums to cater to the demand. As a result, there are some stadiums in Shanghai, China. Opening stadium allows developers to earn consistent rental income. Of course, as with any business decision, opening a new stadium requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the stadium is one of the most important decisions that will determine whether it will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in Shanghai, China to open a new stadium. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Shanghai, China, where would be recommended, if a property developer is looking to open a new stadium?

Data

To answer the question above, we need the following data:

- List of districts in Shanghai
  We can get it from Wikipedia(https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Shanghai). To do that, we choose to use web scraping techniques to extract the data with Python requests and beautifulsoup packages.

- The coordinates of these districts
  Then we will get the geographical coordinates of the districts using Python Geocoder package which will give us the latitude and longitude coordinates of the districts.

- Venue data related to the current stadium
  We will use Foursquare API to get the venue data for those districts.