Introduction

In a real estate project, location selection can be challenging. As one wouldn't have an idea whether one will be able to rent the place one is trying to build. Depending on that location and quality of build, the price of meter is set. We will explore the price in the supply data analysis but the demand data for the pricing is not available. Since those factors data are not readily available, we will focus in this report on rentability of a place based on their location. More specifically, we will study the rentability of a store location in Al Khobar city in Saudi Arabia. Real estate development companies in Al Khobar would benefit from such analysis. Al Khobar is a small city in Saudi Arabia with a total population of 219,679 and a total area of 571 square kilometers. A city with forty-one neighborhoods only. Within those forty-one neighborhoods, we will study the rentability of a store in a given location.

Data

Data Selection

Obtaining the data was not an easy task as most of the needed data was not readily available. I had to find the location data for each neighborhood in my city which was not an easy search but I finally found it here. Using the foursquare data as rented stores, I had to also find the data for the unrented or available stores which I scrapped from one of the most famous sites in Saudi Arabia for renting a place Aqar.

Data Quality

The quality of the data is mediocre at best as the only usable piece of information is the location data which was obtained after a heavy processing and some approximation of the Aqar data especially. Aqar data lacked the location (latitude and longitude) and the foursquare data lacked the price of the rent and the square meter area.

Data Processing

After a long search, I found the coordinates for neighborhoods in Saudi Arabia but it was in SQL files format for cities and neighborhoods which made it really hard to process and open. After downloading a collection of software to open the file, I found a website that opens SQL files but it refused to run the create table query in the file so I had to copy it manually to a text file and turn it into a CSV file. I also had to manipulate it through Excel in order to make it proper to join the cities table with the neighborhood table. Then, I filtered the data in order to obtain the neighborhood data for Al Khobar. For the Aqar data, the only location data that was available is the unofficial address in Arabic. So, I needed to do a fuzzy match between the two data sets in order to obtain the coordinates of the neighborhood the spaces lie in. After that I had to randomize the location through a normal distribution with a mean as the center of the neighborhood and standard deviation as the third of distance between the center and the edges of the neighborhood in order to obtain a unique location for each data point. The points a randomized based on the neighborhood they are located in. Then, I added for each data set a target value of one for rented stores (the foursquare data) and a value of zero for unrented stores (the Aqar data).