Battle of Neighborhoods for setting up

Diagnostic and Treatment center in New York State

Gauri Priya Saran
February 07, 2020

Introduction:

❖ Problem Background:

In New York State the health care landscape is evolving. It is marked in particular by an overall shift toward the provision of primary care and medical care services in neighborhood settings. This movement has been supported by federal and state initiatives aimed at improving overall population health and outcomes of care, and creating more efficient delivery by promoting value throughout the health care delivery system.
In New York State, someone is looking to open a Diagnostic and Treatment Center. Which location should we recommend that they open it?

❖ Problem Description:

This project is an initiative to help provide valuable information to entrepreneurs who want to start a venture in the health care industry by setting up a **Diagnostic and Treatment center** in the borough of Brooklyn, county-level administrative division coterminous with Kings County, New York State. (https://en.wikipedia.org/wiki/Boroughs_of_New_York_City) It is the New York City's most populous borough, with an estimated 2,504,700 residents in 2010.

What is a Diagnostic and Treatment center?

Under the statutory authority of Article 28, Section 3401 of the Public Health Law (PHL), and Title 10 of the New York Codes of Rules and Regulations (NYCRR), Section 405, providers who do not qualify or choose not to operate as private practices may be licensed by the State to operate free-standing clinics are also known as **Diagnostic and Treatment Centers** (Clinics). These free-standing clinics are separately-owned and are not operated by a hospital. By contrast, clinics that are owned and operated by a hospital are known as Hospital Extension Clinics. (https://profiles.health.ny.gov/clinic/)

To find a great location for a Diagnostic and Treatment Center requires intensive research, knowledge gathering of health care sector, and appropriate dataset. A great location in the business plan helps in preparing more accurate earnings estimate for the business and also in fundraising.

The health care industry is a highly controlled sector, with regulation from the administration, hence business plan should address particular issues and help in resolving any queries of prospective investors including **location accessibility, facilities nearby**, and permissibility. A number of businesses are linked with hospitals and health care, from treatment providers and

pharmacy services to eateries, residential areas for health care staff and so on and hence it is critical that the health care business plan deals with the locational issues appropriately.

Data:

❖ Data Sources:

To figure out the best location to set up a Diagnostic and Treatment Center, logically, we need 2 things:

1. Its geographical coordinates (latitude and longitude) to find out where exactly it is located.
2. Population and facilities of the neighborhood where the facility is located.

The data from different health care facilities with their geographical coordinates in the borough of Brooklyn, New York will be used for analyzing different neighborhoods that can be conducive for establishing a new Diagnostic and Treatment Center.

The hospital **data** for New York State **with their geographical coordinates** is available on the website (https://health.data.ny.gov/Health/Health-Facility-General-Information/vn5v-hh5r/data).

The data related to existing diagnostic and treatment centers to is also verifiable at https://health.data.ny.gov/Health/HFIS-Diagnostic-and-Treatment-Centers-General-Outp/tx6m-mpjb

Data Format:

Data is available in the form of **'.csv' files**, there are 36 columns in the dataset of interest, and some of the column names are following:

- Facility ID
- Facility Name
- Short Description
- Description
- Facility Open Date
- Facility Address 1
- .....
- Ownership Type
- Facility Latitude
- Facility Longitude
- Facility Location

Data Cleaning:

Data source was found to include errors and missing values – hence data cleaning was used to address these anomalies. Not cleaning data can lead to problems such as linking errors, model

misspecification, errors in parameter estimation and incorrect analysis leading users to draw false conclusions.

After the data was imported into a Pandas data frame, rows having **Nan values** in columns of interest, specifically **Facility Latitude** and **Longitude**, were dropped.

How Data will be used?

The data is imported into a Pandas dataframe and has been cleaned to perform analysis using KMeans Algorithm to form clusters of different health care facilities that are Diagnostic and treatment centers. Foursquare API was used to gain information about the surrounding neighborhoods. The information gained from surrounding neighborhoods in these clusters has been used to make important conclusions regarding the set up of a new diagnostic and treatment center.