

Coursera Capstone Project

(IBM Applied Data Science)

New Urgent Care Center in Boston, MA

Mrinal Shukla

November 2019



Introduction

The cost of ER care is increasing everyday and becoming unaffordable. As a result more and more Urgent Care Centers are coming up. A 2016 study in the [Annals of Emergency Medicine](#) found that ER treatment costs were about 10 times more (an average of about \$2,200) than in an urgent care center (about \$168) — even for patients with the same diagnosis. Also, the wait time is much less in Urgent Care centers than in ER rooms. The UCA's 2018 Benchmarking Report found that more than 70% of patients waited less than 20 minutes to see a provider at an urgent care center, and nearly 94% were seen in less than 30 minutes. Overall, 85% of urgent care centers patients are taken care of in under an hour. There has been a great increase in the number of Urgent Care Centers. The Urgent Care companies are looking for best possible venues to open more.

Business Problem

This study explores the neighborhoods of Boston to find the best possible locations for Urgent Care Centers. It uses Machine Learning Methodologies like Clustering and Segmentation to find the neighborhoods which would be best suited for opening Urgent Care Centers. This project aims to answer the business question: What locations are best suited for opening an Urgent Care Center in the city of Boston.

Target Audience

This project will be useful for Urgent Care companies who are looking to open new Urgent Care Centers in Boston. This project is time appropriate as the demand for more and more Urgent Care Centers is on the rise due to increasing ER care costs. According to Consumer Reports, the number of urgent care facilities increased from 6,400 in 2014 to 8,100 in 2018, with another 500 to 600 expected to open.

Data

To solve the problem we would need the following data:

- List of all the neighborhoods of Boston. This will define the scope of the project, which is limited to the city of Boston.
- Latitudes and Longitudes of all the neighborhoods of Boston. We will use this data to plot maps and get venue data.
- Venue data, particularly data related to Urgent Care Centers. We will use this data to perform segmentation and clustering.

Data Sources and Extraction Methods

The list of neighborhoods of Boston is obtained from Wikipedia ([https://en.wikipedia.org/wiki/Neighborhoods in Boston](https://en.wikipedia.org/wiki/Neighborhoods_in_Boston)) it has 22 neighborhoods. We get the data by scraping the webpage using the Requests and BeautifulSoup packages of Python.

We get the Longitude and Latitude of the neighborhoods by using the Geocoder package of Python. The venue data is retrieved using the FourSquare API. This returns various types of venues but we are only interested in Urgent Care Centers so we run a query for only these venues.

After retrieving the data we perform various Data Science methodologies like Data Cleaning, Data Wrangling, Visualization, Segmentation and Clustering to analyze the data and find the answer to our business problem: What is the best location to open a new Urgent Care Center in Boston. In the following sections we will elaborate on the Methodologies used and the conclusion from the analysis.

Methodology

First of all we will need to get the list of all neighborhoods of Boston. This is available on the Wikipedia page (https://en.wikipedia.org/wiki/Neighborhoods_in_Boston). We get the list by web scraping using the requests and BeautifulSoup packages of Python. Once we have the list of neighborhoods we need to get the geographical co-ordinates (Latitudes and Longitudes) of all the neighborhoods. These would be needed to get make FourSquare API calls. We get these co-ordinates by using the Geocoder package of Python. Once we get these co-ordinates we plot them on a map of Boston to visualize the data and also to make sure the co-ordinates are correct.

The next process is to analyze the neighborhoods and find out the number of Urgent Care Centers in each neighborhood. We do this by making FourSquare API calls. Usually, the explore API call is used to get top 100 venues and then the data is analyzed but in this case, Urgent Care Center is a sub-category of venues and not one of the top venues it is not returned as one of the venues in an Explore API call. For this we use the Search API call, which searches for all venues within 2000 meters of each neighborhood, which are Urgent Care Centers or similar. Once we have the list of Venues in each neighborhood, we group them together by neighborhood and frequency. Then we compute the mean of the frequency of each venue. Since we are interested in only the Urgent Care Centers we filter out the data for these only.

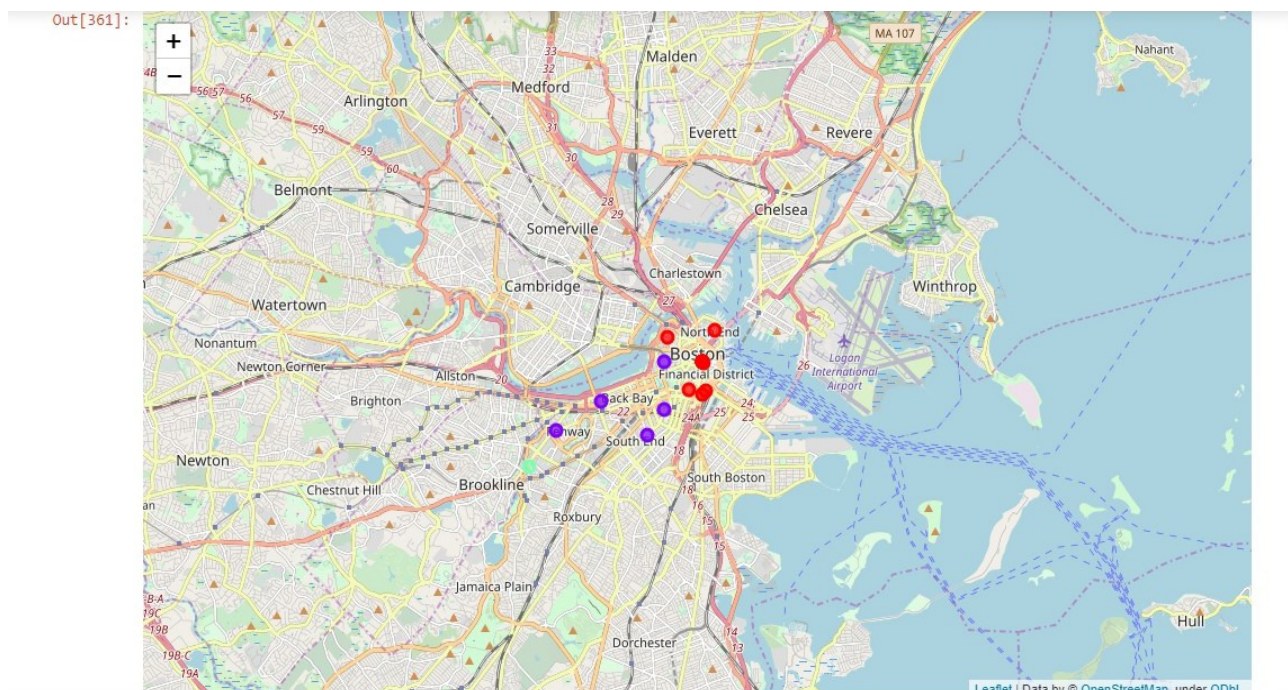
Once we have the data for Urgent Care Centers, we perform clustering using the K-Means algorithm as that is the most common and best suited for our business problem. The K-means algorithm has 3 steps Initialisation - K initial "means" (centroids) are generated at random, Assignment - K clusters are created by associating each observation with the nearest centroid, Update - The centroid of the clusters becomes the new mean. Assignment and Update are repeated iteratively until convergence. We create 3 clusters for the frequency of occurrences of 'Urgent Care Centers'. These clusters will help us determine which neighborhoods have the most or least number of Urgent Care Centers, thus enabling us to answer our question: Which neighborhoods are suitable for opening a new Urgent Care Center.

Results

The data from the k-means clusters shows that we can categorize the clusters by the frequency of occurrences of Urgent Care Centers.

- Cluster 0 – This cluster contains neighborhoods with the most number of Urgent Care Centers.
- Cluster 1 – This cluster contains neighborhoods with a few Urgent Care Centers.
- Cluster 3 – This is the cluster with the neighborhoods with least number of Urgent Care Centers.

The results have been visualized in the map below with cluster 0 shown in red, cluster 1 in purple and cluster 2 in mint green.



Discussion

As observed in the above results, Cluster 0 has the neighborhoods with the most number of Urgent Care Centers, Cluster 2 has the least and Cluster 1 moderate number. Cluster 0 has neighborhoods which are in the North End or Central Part of Boston. As these areas already have a number of Urgent Care Centers it would not be advisable to open more centers in neighborhoods of Cluster 0. Cluster 2 and Cluster 1 have very few Urgent Care Centers so it would be a good idea to open new centers in these neighborhoods. Opening centers in neighborhoods of Cluster 0 would pose a lot of competition and may not be successful as people who are already using other Centers would like to stick to them than go to new a center. As people like to stick to one doctor or facility as much as possible rather than switching doctors and facilities. Thus, we can say that for a new Urgent Care Center to be successful, it should be opened in an area where there are very few or no Urgent Care Centers, which in this case are neighborhoods in Cluster 2 and Cluster 1.

Limitations and Suggestions

Though this project gives the base solution for the Business question: Which neighborhoods are most suitable to open a new Urgent Care Center; this project has some limitations. The first being that it takes only one factor in consideration which is frequency of occurrence. There could be other factors which can influence the success of a new Urgent Care Center like density of population, income groups etc. A suggestion for future research would be that other factors are also considered while clustering and segmenting to get a better result. Also, this project uses the Sandbox tier of the FourSquare API, which has limited number of API calls. A suggestion would be to use the paid tier of the FourSquare API to get unlimited API calls hence getting more detailed data.

Conclusion

In this project all the steps for Data Analysis were taken namely:

- Identifying the business problem
- Understanding the Data
- Data Collection
- Modelling the Data using Machine Learning
- Recommendation to stakeholders, in this case the Urgent Care Companies.

The answer to the business question stated in the Introduction section is: The neighborhoods in Cluster 2 and Cluster 1 would be the best suited for opening new Urgent Care Centers.

References

- Neighborhoods of Boston, *Wikipedia*. Data retrieved from https://en.wikipedia.org/wiki/Neighborhoods_in_Boston
- Foursquare Developers Documentation.
- <https://developers.foursquare.com/docs>
- Statistics about Urgent Care Centers was quoted from the following article
- <https://www.businessinsider.com/urgent-care-industry-trends>