Coursera Capstone Project Report

Opening a New Movie Theatre in Boston, MA

Thomas Abowd

Introduction:

Going to the movies is a beloved past time. Many decisions go into which movie theater people will choose to go see the latest film. One of the most important factors is location. Questions one asks themselves before picking which movie theater to go to may include: Is the theater near my home? Easy to get to via public transit? Is the theater going to be too crowded? For a property developer in Boston, MA, the task of finding the perfect location for a new movie theater is quite overwhelming.

Business Problem:

The question this project aims to solve is: Where is the best location for a property developer to open a brand new movie theater in Boston, MA?

Target Audience:

The target audience is Boston property developers interested in opening new movie theaters in the Boston area.

Data:

Data Needed:

- Neighborhoods of Boston
- Latitude and Longitude values for Boston Neighborhoods
- Venue Data from Foursquare

Data Sources and Methods for Extraction:

I will be using data from the following web page which contains a list of the 23 neighborhoods in Boston. (https://theculturetrip.com/north-america/usa/massachusetts/articles/a-guide-to-bostons-23-neighborhoods/). I will use web scraping to extract the data. Next, I will use the Python Geocoder package to get both the latitude and longitude coordinates. With these coordinates I will use Foursquare API to get the relevant venue data. During analysis, I will be using k-means clustering and map visualization with Folium.

Methodology:

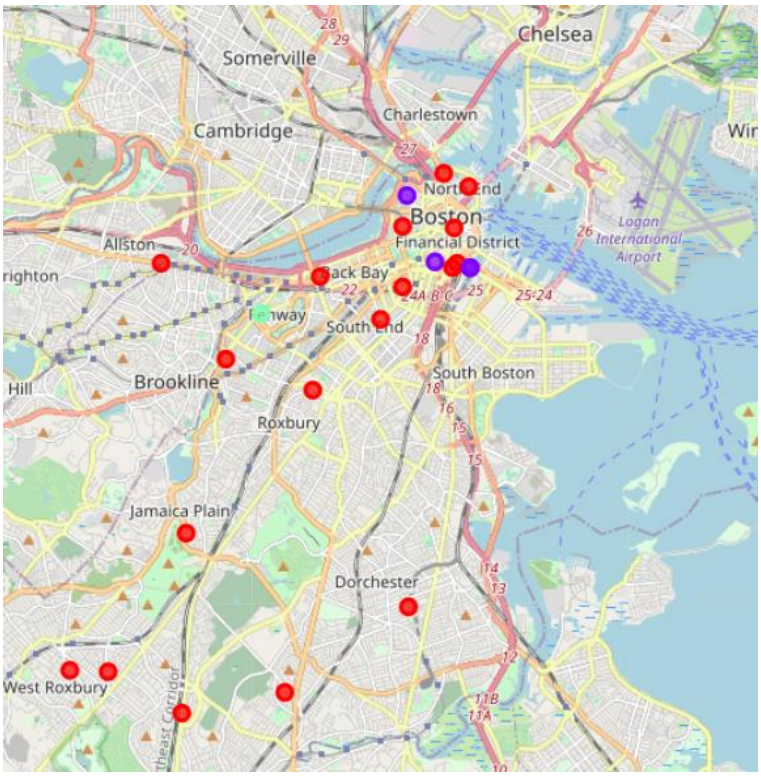
First, I needed to get the list of all neighborhoods in Boston. I used data from the following web page which

contains a list of the 23 neighborhoods in Boston (https://theculturetrip.com/north-america/usa/massachusetts/articles/a-guide-to-bostons-23-neighborhoods/). I scraped the data from the web page using Python requests and beautiful soup packages. Next, I got the geographical coordinates for each neighborhood in Boston using the Geocoder package which converts an address into latitude and longitude. Using this data, I created a pandas DataFrame and visualized the data in the form of a Folium map. Mapping the neighborhoods allowed me to ensure that the data was correct. Next, using the Foursquare API I examined the top 100 venues in the Boston neighborhoods. Then, I looked at how many unique venues there were in each Boston neighborhood. I also checked to make sure that 'Movie Theater' was a result inside of the boston_venues DataFrame that I created. Next, I analyzed each neighborhood using one hot encoding by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. After checking out the top 10 venues in each neighborhood, I made a new DataFrame called boston_movie which contained only movie theater data. Lastly, I performed clustering on the data through the use of k-means clustering. K-means clustering is a form of unsupervised machine learning. K-means works by identifying k number of centroids, then allocating each

data point to the nearest, while keeping the centroids as small as possible. Each neighborhood was clustered based on their frequency of occurrence based on the value ‘Movie Theater’. From the results of clustering I was able to select which cluster would be best suited for a new movie theater.

Results of K-means Clustering:

- Cluster 0: Neighborhoods with no movie theaters
- Cluster 1: Neighborhoods with one movie theater
- Cluster 2: Neighborhood with the highest concentration of movie theaters



Cluster 0:

Cluster 0

```
In [30]: b_movie_merged.loc[b_movie_merged['Cluster Label'] == 0]
```

Out[30]:

	Neighborhood	Movie Theater	Cluster Label	Latitude	Longitude
0	Allston	0.0	0	42.293046	-71.067169
20	South End	0.0	0	42.342560	-71.073580
19	South Boston	0.0	0	42.352250	-71.055691
18	Roxbury	0.0	0	42.330294	-71.089499
17	Roslindale	0.0	0	42.281813	-71.137111
16	North End	0.0	0	42.365490	-71.052970
15	Mission Hill	0.0	0	42.335780	-71.109810
13	Mattapan	0.0	0	42.278225	-71.096083
12	Jamaica Plain	0.0	0	42.305850	-71.119095
11	Hyde Park	0.0	0	42.274864	-71.119856
8	Downtown	0.0	0	42.358290	-71.056630
5	Charlestown	0.0	0	42.367770	-71.059017
4	Brighton	0.0	0	42.352131	-71.124924
3	Beacon Hill	0.0	0	42.358420	-71.068600
2	Bay Village	0.0	0	42.348168	-71.068471
1	Back Bay	0.0	0	42.349990	-71.087650
9	East Boston	0.0	0	42.351414	-71.056713
22	West Roxbury	0.0	0	42.282193	-71.145995

Cluster 1:

Cluster 1

```
In [31]: b_movie_merged.loc[b_movie_merged['Cluster Label'] == 1]
```

Out[31]:

	Neighborhood	Movie Theater	Cluster Label	Latitude	Longitude
21	West End	0.01087	1	42.363940	-71.06739
7	Dorchester	0.01000	1	42.351349	-71.05285
14	Mid-Dorchester	0.01000	1	42.351349	-71.05285
6	Chinatown – Leather District	0.01000	1	42.352510	-71.06090

Cluster 2:

Cluster 2

```
In [33]: b_movie_merged.loc[b_movie_merged['Cluster Label'] == 2]
```

Out[33]:

	Neighborhood	Movie Theater	Cluster Label	Latitude	Longitude
10	Fenway-Kenmore	0.018868	2	42.34355	-71.10157

Discussion:

The concentration of movie theaters is in descending order among the clusters from Cluster 2 down to Cluster 0. The Fenway-Kenmore neighborhood is the most concentrated movie theater location out of all the Boston neighborhoods, as well as the only neighborhood in Cluster 2. Cluster 1 contains four neighborhoods that all have at least one theater. Cluster 0 contains 18 neighborhoods that all have no movie theaters. My recommendation would be to open up a movie theater in Cluster 0. Any movie theater opening up in Allston, South End, South Boston, Roxbury, Roslindale, North End, Mission Hill, Mattapan, Jamaica Plain, Hyde Park, Downtown, Charlestown, Brighton, Beacon Hill, Bay Village, Back Bay, East Boston, or West Roxbury would face no competition. Overall, Boston has very few movie theaters. Opening up a theater in Cluster 0 would be a great opportunity for any property developer to take on. It should be noted that a limitation of the analysis was that only one factor was considered, frequency of occurrence of movie theaters. Further research, would help shine a light into other import factors for consideration in choosing a prime movie theater location.

Conclusion:

In this capstone project, I completed the process of identifying a business problem, determining what data would be needed, extracting and cleaning this data, using machine learning to create clusters for analysis, and providing a recommendation for an important stakeholder. The data analysis carried out was used to answer the question asked in the introduction: Where is the best location for a property developer to open a brand new movie theater in Boston, MA? The answer to this question is that the opening a new movie theater should be done in a neighborhood located in cluster 0. The findings of my research will be helpful for any Boston property developer.