**Rental Property Cap Rate and Venue Data Analysis of the Greater Boston Area**

1. Introduction - Business problem and interested audience

The Greater Boston Area is well known for its highly demanded real estate and rental markets. Savvy Investors of all kinds face a challenge when attempting to make an entrance to Greater Boston due to the area’s highly competitive nature. Interested parties commonly seek to gain an edge by taking a neighborhood targeted approach. As such the question beholds, which neighborhoods could take priority? Should venue offerings be in the forefront when selecting neighborhoods to take into focus? In this analysis, we will help to answer such questions. We will analyze average capitalization rates of condos and multifamily properties by neighborhoods. We will also analyze venue data for the possibility of gaining insight regarding a neighborhood’s venue character vs expected returns. We will utilize Kmeans algorithm to identify if clustering of neighborhoods exist in Greater Boston. Further we will compare and determine if any relationships exist between neighborhood venue clusters vs neighborhood capitalization rates. Finally, we will also compare venue density vs rental days on the market for possible relationship or insights towards rental desirability.

1. Data – Description and source of data that was used to solve the problem
2. Average rent and property price by neighborhood and number of bedrooms. We start the analysis by sourcing MLS data, (Multiple Listing Service – utilized by licensed realtors), of both recently sold properties and recently rented properties. MLS is a reliably accurate source of data to retrieve home prices and rental values. It is in fact used by licensed appraisers as a source of valid comparable when performing real estate valuation. It is also a good source for “days on the market” data points.

In this analysis we captured 4,202 sold listings and 8,875 rental listings. It is important to understand that not all of these data points will make it into the analysis. We will further elaborate on this in the Methodology section.

1. Operations cost data and estimates. In order to calculate capitalization rate (cap rate), we must estimate operations costs. Operations cost are the total cost required to operate a rental property. These include property tax, insurance rates, utilities (if any, normally paid by owner of a multiple unit property with a single water meter or common spaces), HOA fees for condos with an HOA, and we will also include maintenance and vacancy cost.

The property tax rates that are used was sourced from web site: <https://joeshimkus.com/MA-Tax-Rates.aspx>. This site lists current property tax rates for all towns/cities of Massachusetts. We multiply the property tax rate by the average price per neighborhood and reduce the total by (2 x 7.5%) . The reduction is to factor property assessment lag by 2 years and assumes property value increased by 7.5% per year. Normally towns will be two years behind in their assessed value used to calculate property taxes. 7.5% is an average of the greater area. This number is ever changing. When continuing this study this number must be updated. As we collect sold price averages we will use it to calculate property value % changes.

For insurance rate, we simple borrow the insurance rate found on Redfin.com property listings, currently 0.4% x property price. We also confirmed these values with our insurance office. It’s a good estimate.

For utilities, we multiply multi-family properties per bedroom by an expected water cost per person. The data point of the cost of water per person is sourced from: <http://archive.boston.com/yourtown/specials/water/massachusetts_water_usage_map/>.

For maintenance and vacancy, we assign 10% of average rent to cost. Condos should actually have smaller maintenance cost percentage as it is likely that some maintenance costs are partially covered by the HOA. Therefore, we simple exclude HOA fees from our calculations and keep the rate the same as we would do for multi-family properties.

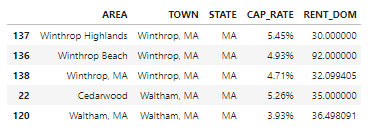
1. Geo location data was partially sourced from geocoder geolocator. Unfortunately, we found that geolocator was pulling in some erroneous locations. We then manually corrected and stored the neighborhood geo locations into a data frame. Since it is always the same neighborhoods, this needs to be stored only once or only updated when new neighborhoods are introduced. If this code was to be scaled to a larger area, perhaps in a nationwide scale, we surely would recommend a better method to source geo locator data, perhaps a paid API service.
2. Finally, we used Foursquare API as a data source to explore venues of each neighborhood.
3. Methodology – Main components discussion, description of exploratory data analysis, inferential statistic testing, machine learnings used.
4. MLS data methodology - It is important to understand that not all of the 4k+ sold listing and 8k+ rental listings data points will make it into the analysis. First, we calculate averages of each sold home prices and rental prices broken down by neighborhoods and number of bedrooms. Each neighborhood and number of bedrooms rent price average and sold price average must have a pair with one another or it will be excluded from the analysis. In considering reliability of the averages, we look to sample size by counting the number of data points used for each paired average. The strength behind the averages is the source of data being MLS, some neighborhoods may end up with singular or only minimal data points. MLS being a reliable data source makes even singular data point worth considering as an indicator. However, we need to remain mindful of results with minimal data.
5. Calculating Capitalization Rates (cap rate)

Cap Rate = NOI / Property Acquisition Cost

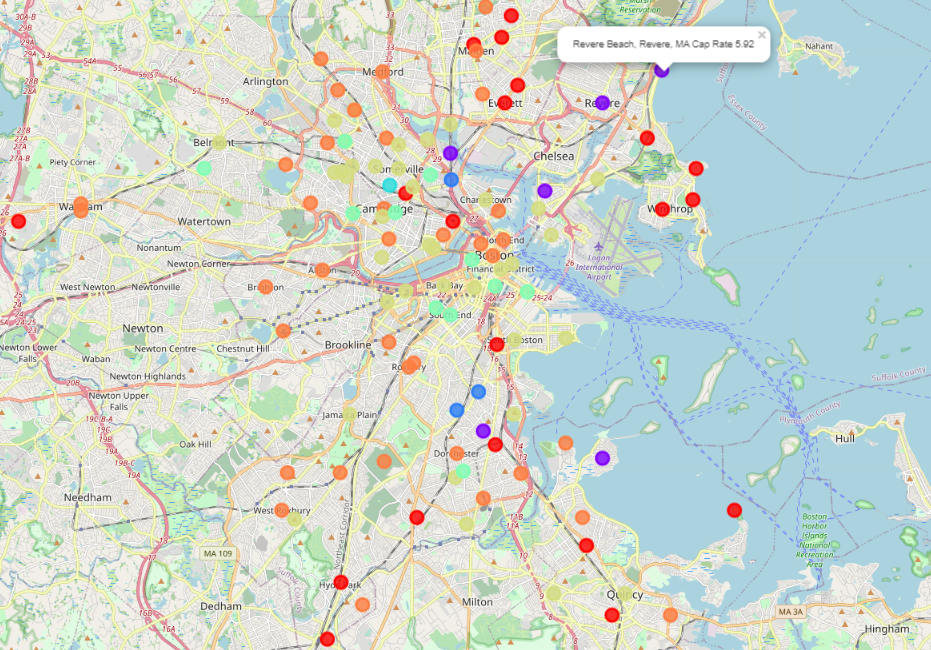
NOI (Net Operating Income) = Total Rent – Operation Cost

Operation Cost = Total annual: Property tax, Insurance, Utilities, Maintenance & Vacancy Cost

We first use the sourced MLS data to form averages of rent and property prices by neighborhoods and number of bedrooms. We keep all data points that pair rent and sold prices by number of bedrooms per neighborhood. Next, we use each neighborhood and number of bedrooms to calculate individual cap rates. We then calculate the mean cap rate for each neighborhood by taking the average cap rate of each neighborhood defined by number of bedrooms. Note that if an average cap rate was directly calculated to neighborhoods without considering number of bedrooms the results would likely be skewed, either price heavy or rent heavy, depending on the distribution of the MLS data sample.

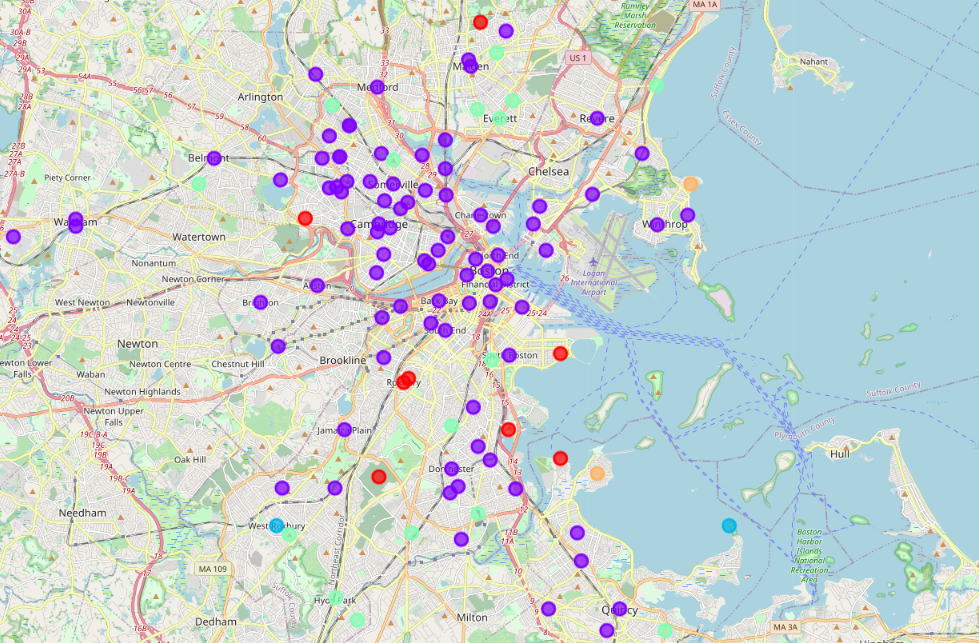


2b. Visualizing cap rates per neighborhood. We combine geo locator to our cap rate data frame. Then we use folium to display each level of cap rate in increments of 100 basis points. Further we will use this visualization in comparison to clustered venues of each neighborhood to observe any similarity or patterns.

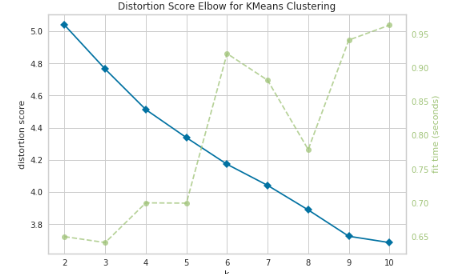
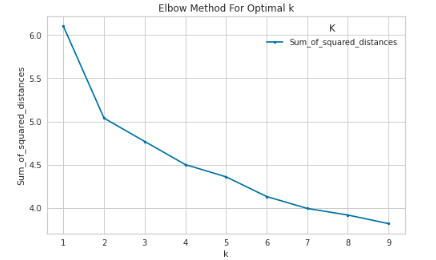


1. Foursquare venue data and Kmeans clustering. We run through a method of collecting venue data for each neighborhood’s geo location. We define our search to a 700 meter radius and request max of up to 100 top venues. We then use the Kmeans algorithm to identify possible clusters. These clusters could identify trends of venue types and similarity between neighborhoods.

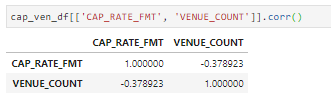
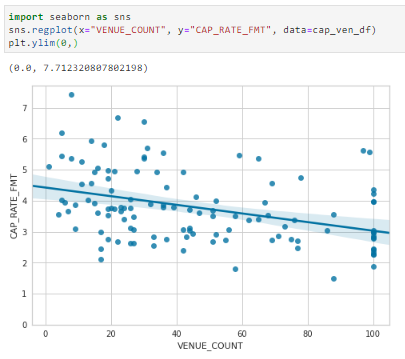
Once the neighborhoods are clustered Folium is used to visualize each neighborhood clusters on to a map. In comparison to the above cap rate map we can see there are no observed similarity or likeness in patterns. This could lead us to conclude that cap rates and neighborhood’s venue types are not co-dependent.



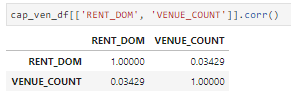
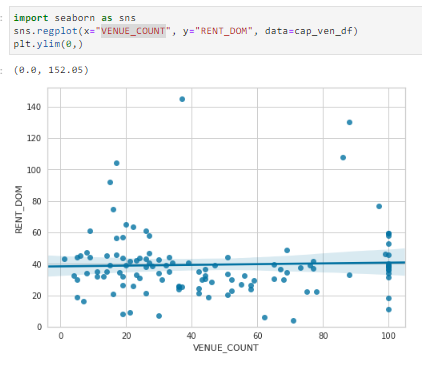
When we visualized the kmeans elbow method, we discovered that the curve is mostly flat. This is an indicator that clustering of venue types are not well defined or even exist at all. We observed a slight bend and intersection at 5 clusters so we define 5 clusters as most optimal.

1. Results – Discussion regarding results
2. Regarding cap rates by neighborhood, we are able to identify estimated cap rates of 121 separate neighborhoods. This information is very insightful for any interested investors and certainly gives an edge when making an attempt to take a targeted approach.
3. Comparing both cap rates and clustered venues maps, clearly there was no relationship between neither. As seen in the maps above. We also run correlation analysis which return 0.37 again a clear indicator that venue clusters and cap rates are not correlated. This is useful to know as one may understand that venue types offered in a given area of Greater Boston do not necessarily have a direct effect on potential investment property returns.

1. We also review number of venues as a measure of venue density vs cap rates per neighborhoods and again we observe independent factors. Correlation observed was 0.03, again there was no relationship between venue densities vs cap rate observed. However, venue density is still a valuable data point to potential investors. Understanding density of venues would help an investor become more familiar with the neighborhood’s profile.

1. Discussion – Discussion of observations noted and recommendations
2. In our observation we have identified several neighborhoods which have resulted with cap rates of 6% or higher. Generally any property with an estimated cap rate of 6% or higher is one that should be taken into consideration. Of course investors’ profiles differ and so do investment objectives so all should consider their own investment return requirements. I would say confidently I myself would certainly be interested in any property with a cap rate of a 6% level or higher.
3. Considering venues as an indicator of targeted approach has proven as an undependable method. We recommend that investors do not look towards a neighborhood’s venue profile as a means to select investable neighborhoods. Clearly we can see that no relationship exist between expected returns vs venue mix in the Greater Boston area. We do of course recommend taking observation of the 5 neighborhood clusters as a way to becoming more familiarized with neighborhood offerings of the area. I also recommend becoming familiar with venue density. Investors should consider expected returns, however in part of due diligence it is also imperative to them to gain an insight of a neighborhood’s venue characters and density of venues.
4. Conclusion

In conclusion of the current state of Greater Boston’s real estate market that was observed and analyzed. Investors looking to take a targeted approach should consider all neighborhoods with 6% cap rate or higher. They could also utilize the venue analyses result as a means to become familiarized with neighborhoods venues types and density. The following are the neighborhoods with 6% cap rate or higher.

**Neighborhood Town Cap Rate Avg.**

Dorchester's Grove Hall Boston, MA 7.42%

Dorchester's Uphams Corner Boston, MA 6.81%

Charlestown's Sullivan Square Boston, MA 6.54%

Squantum Quincy, MA 6.20%

Hyde Park Boston, MA 6.16%

1. References

* <https://mlspin.com/>
* <http://archive.boston.com/yourtown/specials/water/massachusetts_water_usage_map/>
* <https://joeshimkus.com/MA-Tax-Rates.aspx>
* <https://www.redfin.com/>
* <https://pypi.org/project/foursquare/>