Predicting Best City for Luxury Apartment Development

Justin Clark

14 May 2020

# ****Introduction****

## Background

Realty is one of the most consistently profitable business ventures with average returns that outpace even the best on Wall Street; however, there a large investment is necessary to realize those returns. By predicting where demand of realty will be highest, property developers can better plan where to build such that they can obtain the largest return on their investment.

## Problem

Data may be used to predict areas and cities where a development would garner the most interest and therefore return on investment. This project looks at one case of a luxury apartment developer. Luxury apartments are extremely popular in the “Millennial Generation” due to the lack of requirement for a large down payment, low maintenance, and access to large amounts of amenities not normally found in homes in similar price range. By analyzing the venues in a city for diversity and how well liked they are, this project will comparatively score cities for determination of which is best for luxury apartment development.

## Interest

Developers would be the primary candidate for using this tool as it may give insight into where to invest. Additionally, city governments and people looking to move may find the tool useful for deciding on zoning and a fast way to compare cities to each other, respectively.

# Data

## Data Sources

Venue information was obtained through Foursquare, an application that collects venue information through its users, venue internet presence, and venue operators. Foursquare is a platform that has existed for many years and is trusted by many to provide unbiased information about each individual venue. City information was obtained for each through Wikipedia and verified by a number of different sources more specific to each city, such as the official city government website of each and the Google Maps application.

## Data Usage

The coordinates for each city center are input to the tool as an area to explore for potential development. The coordinates are necessary to use with the Foursquare Application Program Interface (API) which is the mainstay of the tool in determining the final results. The Foursquare API is used to collect information about venue locations in the city and further to detail individual venue information used in scoring the city.

# Methodology

## Exploratory Data Analysis

To determine how to score an area for developmental purposes, I used my own city as a control sample to determine which data sets were relevant to the scoring and how to find those within the larger available sets collected from the Foursquare API. Using ratings of luxury apartments, vacancies, and price points – I found the similarities of areas in which major luxury apartment complexes had higher ratings, lower vacancies, and higher price points compared to other areas. I then selected those factors which were measurable over a large and diverse test set. Using these factors, I created and tested code that could measure and score the different factors – again using my own city control sample to determine the efficacy.

## Method and Model

### Assumptions

A number of assumptions and estimations were made in the building of this tool to simplify the code and the interface. These include a one kilometer radius in which venues would influence a potential tenant’s decision to live in an specific area, the two venue factors that would significantly influence a potential tenant is rating and if other venues of the same category exist in the area, and that a potential development could be created in the location – ignoring all zoning and geographical concerns. Because the Foursquare API uses coordinates to explore an area and distance is a function of Latitude, I assumed the cities would be between then 25th and 40th parallels and that the small differences between the Latitudes would be negligible in the calculations.

### Process

The tool user will input the areas’ names and coordinates in which development is being considered. The tool requests venue information using the Foursquare API in a four kilometer radius of the inputted coordinates for each area. The tool, using K-Means clustering, adds points in different directions from the input and divides the venues among the points to determine which coordinate set would lead to the best score within the four kilometer radius. The coordinate set, with the most venues classified to it, is sent through the Foursquare API with a one kilometer radius range limit for a final venue set for each original area inputted. The final venue set is then iteratively requested within the Foursquare API for more details. By this process, the tool determines the customer rating and the venue type for each venue within the area. The score for an area is determined by the mean customer rating and the diversity of venues. This process is repeated for each area and the scores saved for review. The tool, at the end, outputs a statement of which area has the best score.

# Results

## Test Set

To test and exemplify the efficacy of the tool, a sampling of major cities was used to include Birmingham, Atlanta, Alexandria, Denver, and Seattle. During the tool’s exploration of the near environment of each city center, the K-Means Clustering correctly classified the points and found a centroid that had a plurality of points upon ocular and quantitative inspection.

## Findings

The five cities’ score were within five percent of each other, with Seattle having the largest score of 9.12 out of 20 possible points. The lowest score of the five cities was Atlanta with 8.79 points, and the mean score was 9.014 points. Upon closer inspection of the data, Atlanta’s score dramatically impacted by the diversity of venues portion of the score. Additionally, the Birmingham score was dramatically different among the different points explored during the K-Means Clustering section of the code.

# Discussion

## Observations

The large differences in the Birmingham score based on approximately one kilometer variations in point spacings shows evidence of two things. One, the K-Means Clustering portion of the code is vital to a city receiving the best score possible given the user does not have vast, detailed knowledge of the city; and two, some cities may receive a higher or lower than representative score based on changes as small as a thousandth of a degree in input.

Turning towards the Atlanta score, Atlanta is known to have many sub-regions within even the urban center which suggests that instead of picking the city center, it would be wise to choose a specific sub-region. Compared to Atlanta; Alexandria, the third highest scorer, the venues are not as spread throughout much of the wider city and many different examples are found on four to five blocks in particular. Whereas fives blocks in Atlanta seem to contain a number of venues of two or three types suggesting diverse types of spaces in Atlanta are found by travelling through the city instead of walking down the block.

## Recommendations

Due to limited time lack of need to be thorough, there are a number of improvements that can be made to the tool and process as well as ways to use the tool in its current state or an improved state more effectively. Any user of the tool should have moderate to great experience or knowledge in areas in which they are studying to more effectively choose a centroid to start with. This knowledge should include zoning, potential clients, costs to build and cost of living versus potential revenue. Users should also take into account the different factors that go into a potential renter or client’s decision to live somewhere. The code currently treats venue diversity and rating as equivalent and the only factors; however, that is not true. The tool may also be improved to score a number of different factors and score differently to more accurately reflect the likelihood someone would be willing to pay a higher rent to live somewhere. These factors could be weather, job market, and quality of infrastructure.

# Conclusion

Seattle, in terms of best selection of venues to attract potential renters, is the best city center to build a luxury apartment complex when compared against Alexandria, Atlanta, Birmingham, and Denver. Using the tool, a better location may be found by testing other cities or including more factors; however, at a first iteration level, this result has statistical significance compared to the other cities tested.