Atlanta or Bust

WHERE TO BUILD DURING ATLANTA'S GROWING HOUSING BOOM

DEVON JANKE

Contents

I.   Introduction

Atlanta as had a huge real estate boom over the past decade. These days, building cranes can always be seen across downtown, and it doesn't show any sign of slowing down. With many wanting to capitalize on the rapid growth of Atlanta, the question many are asking is "Where is the best place to build?".

It is common sense that the areas closest to downtown will be able to charge the most. However, this also means a significant up-front investment in order to acquire and build, and the location may not be very desirable since crime rates tend to be higher in large cities, especially Atlanta, which has a safety rating of 2/100 according to [1]. Careful consideration needs to be given to deciding where to build a set of apartments; future landlords will need to balance common factors such as school performance, crime rates, and nearby businesses to ensure they are well suited to please their tenants.

In this report, I will explore data related to important factors that people usually take into account when choosing a place to live. Using this information, I plan to predict the average rating that an apartment complex can expect to receive based on its surroundings. I will then recommend which zip codes in Atlanta would best place to build in order to take part in the booming housing market. The reason I plan to rely on ratings as a good indicator is that if a complex can expect a higher rating based solely on location, their main responsibility would be to increase the rating through customer satisfaction. A high rating will lead to more interest and will allow them to increase their prices over time.

II.   Data

There is a significant amount of data available to the public to tell a story about a location. Usually, police departments keep a record of crimes available for download on their web page, and school performance ratings are also free to access. I will be using the following data for my analysis.

A.   Data Categories

Apartments - Google Places API

Using the Google Places API, I will get information on Atlanta apartments their associated user ratings as a way to estimate the success of the apartment.

Crime - https://www.atlantapd.org/i-want-to/crime-data-downloads

This is a dataset of crimes that the atlanta police department has collected. I will use this data to determine the average crime rates around various apartment buildings in Atlanta.

Schools - https://schoolgrades.georgia.gov/dataset/school-level-data

Georgia has kept a record of performance for each school based on standardized testing and grades. I will be using the overall scores for these schools to see if there is an association between apartment ratings and school ratings.

Average Income - https://www.irs.gov/pub/irs-soi/17zp11ga.xlsx

The IRS collects reported income information for each year and compiles a summary of the date for each state and each zip code. I will use this information to determine if there is a correlation between income and satisfaction with housing.

Venues - https://developer.foursquare.com/docs/api

By associating each apartment complex with the most common venues nearby, we can determine if there is a pattern in how the nearby businesses affect people's perception of an apartment.

GeoJSON zip codes - https://github.com/OpenDataDE/State-zip-code-GeoJSON

This is a collection of GeoJSON files for zip codes across the United States. This may be useful for visualization purposes.

B. Data Filtering

The data that is available online generally covers the entire stat of Georgia, which is much more than we need, but this only available full list of crimes is from the Atlanta Police Department, which only represents 75% of the land area inside the Atlanta I-285 interstate belt loop. In order to keep only the relevant information and also use only areas we have enough information for, we will need to follow a set of steps to obtain a list of apartments to use in our exploration. This will be covered in the next section.

III.   Methodology

A.   Narrowing the Data

We begin with a broad subset of data collected for the state of Georgia, but it is not all applicable to the city of Atlanta. To start narrowing the data, I used the BeautifulSoup library to collect a set of 41 zip codes that cover the city of Atlanta and its surrounding suburbs.

The imported data for Georgia Schools was filtered based on the zip codes for each school, keeping only the schools with zip codes inside the Atlanta belt loop. Using the street address and zip code, the Latitude and Longitude information was obtained; any schools lacking coordinates were discarded.

Using the coordinates from the schools and crimes dataset, the number of crimes that occurred within 2 km of each school was counted, and any schools with less than 10 crimes were discarded, assuming they were not within the Atlanta Police Department's jurisdiction where the crime information was available.

Finally, a list of apartments withing 5 km of each school was collected using the Google Places API, and duplicates were dropped. To ensure that the ratings for these apartments were not overly biased, apartments with fewer than 10 ratings were discarded.

B.   Generating Features

Using the resulting list of apartments, crimes were counted for each apartment using the same procedure for counting crimes for schools. In a similar manner, the average school scores were calculated. Using the data from the IRS, I calculated the average income for each zip code, and associated an average income for each apartment based on the average income for its zip code. Finally, I used the Foursquare API to generate a list of venue categories near each apartment and assigned a percentage for each venue, representing the percentage of venues that fit in that category.

C.   Correlation: what makes people happier?

The first interesting finding in my exploration is the lack of correlation between school performance and apartment ratings; even more surprising was that crime seemed to have no effect as well (Figure 1). With this information, I determined that I had limited my data to crime-only areas unnecessarily. I re-did my data filtering, this time using all metro-Atlanta schools to

find apartments using the Google Placed API with no restrictions based on crime. This resulted in more usable data for analysis.

Figure 1 - Effects of crimes and schools on ratings. The regression plots show that there is very little information to be gained from these factors.

After re-collecting the data, I removed uncommon venues from my list of collected venues (e.g., acupuncturists) and the venues with the non-descriptive category 'building' in order to increase the integrity and usability of the collected data. Using a word cloud (Figure 2), I was then able to visually rank the characteristics of a location that make residents the happiest with their apartment, namely average income, recreational facilities, and coffee shops.

Figure 2 - Word cloud illustrating strongest positive factors on apartment ratings.

D. Predicting Apartment Ratings

My original intention in this exercise was to use the data to try to predict the ratings that people would give an apartment. The major challenge in all of this was a lot of noise and uncertainty based on the attitudes of the tenants, the personalities of the apartment management, and the fact the people are more likely to give a very low or very high rating rather than a precise rating.

Another challenge was the bias of the data (Figure 3). The average rating of most apartments was between 3 and 4 stars, meaning the training data would also be biased and encourage any multi-layer perceptron (MLP) regressors or classifiers to over-fit to the mean or mode of the data, respectively. I attempted to predict apartment ratings using simple and complex MLP structures, but the result was always predicted to be '4' by the classifier and about 3.2 by the regressor, which is not very useful.
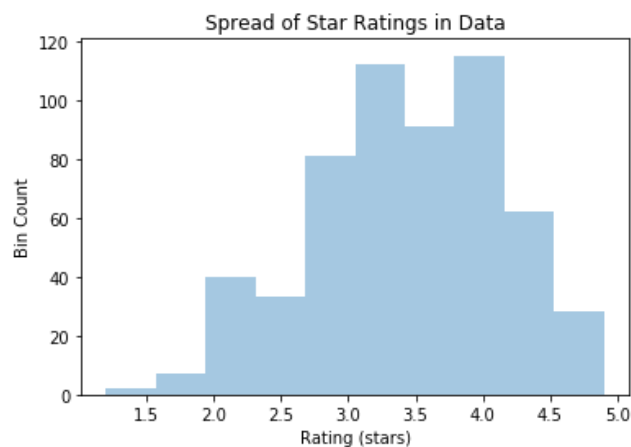

Figure 3 - Spread of apartment ratings in Atlanta, showing a clear bias around 3.5 stars.

Using a simple linear regressor, I was able to generate a model that predicted the apartment rating with good accuracy (Figure 4), with prediction error less than 0.4 stars on 90% of the training data and less than 0.6 stars on 90% of the test data.

Figure 4 - Resulting error from regression model in predicting apartment ratings based on surrounding venues and income.

IV. Results

The first surprising result from the analysis was the lack of correlation between crime and school performance on apartment ratings; normally, these factors are discussed among families when deciding where to live. However, the demographics of a city tend to include more single-person or childless homes. This is because the cost of living close to a city is much higher and is not conducive to families with young children just starting to gain their financial foothold. Since there are fewer families within the city, it makes sense that school performance has a lower bearing on rating.

In many parts of the world, high crime rate tends to increase in lower-valued neighborhoods. This trend is not as applicable near large cities, where property values and crime both increase the closer you get to the center of the city. Because the sample size of our apartments was the whole inside of the Atlanta belt loop, it was much more difficult to find a pattern between crime and ratings.

The generated word cloud pointed to average income as the most significant factor behind higher apartment ratings with gyms and other recreational facilities in close second. This agrees with common reasoning, as people with a stronger financial footing and good health tend to be happier. Other surrounding features, such as coffee shops, massage studios, and shopping provide an opportunity to relax and socialize, further contributing to overall satisfaction. Since we were able to use the correlation data to generate a fairly accurate linear model that predicts apartment ratings, it is safe to assume that the information in this word cloud can be used to find locations with features that correspond with happier residents.

V.   Discussion and Recommendations

Using the Income data from the IRS, a plot of average income per zip code (Figure 5) suggests that the first place to start looking for a place to build an apartment complex would be in Northwest Atlanta, area code 30327, along with areas in central North Atlanta. After surveying the options, a list of surrounding venues can be generated and categorized. Builders should select the area that suggests the highest rating using the linear model represented by the word cloud.
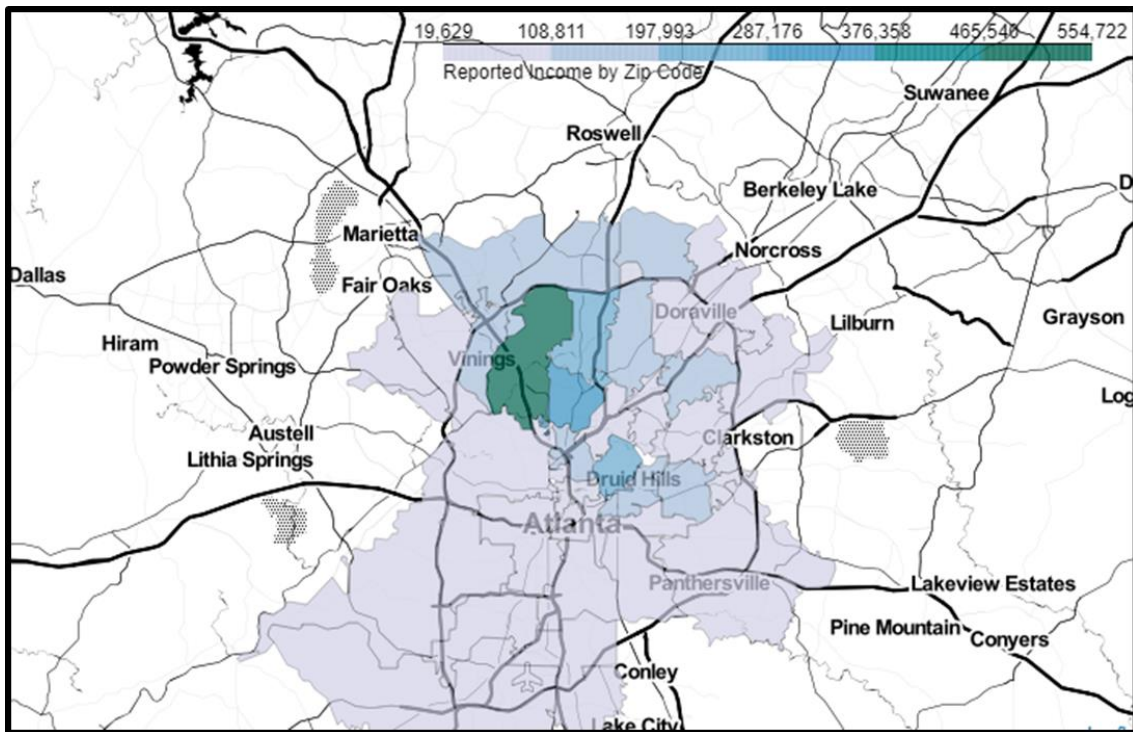

Figure 5 - Colored map of average income in Atlanta, showing Northwest Atlanta as a prime spot to build.

Since not all locations may have the desired features available, builders may choose to include a gym and/or coffee shop inside the apartment complex or contract with these businesses to run one of these facilities close to or on their property.

VI.   Conclusion

While the Atlanta housing market is in full swing, now is the time to take advantage of the opportunities. As businesses continue to grow and jobs multiply, more money will flow into Atlanta, leading to even more opportunities to house higher-income individuals and happier tenants. Coordinating efforts with other like-minded businesses will lead to better revenue for all concerned. With the path clearly shown, the only logical next step is to start.