

CSE 6748

Applied Analytics Practicum

Summer 2019

Carlos White

## **Final Report**

In an effort to aid Habitat for Humanity's mission, we have written this report that identifies parcels of lands that would be ideal for purchase in order to build affordable housing in Fulton County.

This report has three sections. In the first section we will do a market analysis by area where we determine which areas of Fulton are most likely to have an increase in real estate value in the following years. In the second section we will examine the constraints that properties should have, and how we filtered out those constraints. Finally, in the third section we will analyze our final list of properties.

### **Market Analysis**

In order to determine which areas Habitat for Humanity should prioritize we developed a time series forecasting model. We used this model to predict the percentage change in aggregated value of the parcels by zip code in five years, that is in 2023. We used the historical data from the tax parcels from 2007 to 2018.

The literature states that univariate Autoregressive Moving Average (ARIMA) models are the very well suited for predicting future property prices:

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.4606&rep=rep1&type=pdf>
- <https://pdfs.semanticscholar.org/8d25/21124a36d1270477b8138bc950555a23699a.pdf>

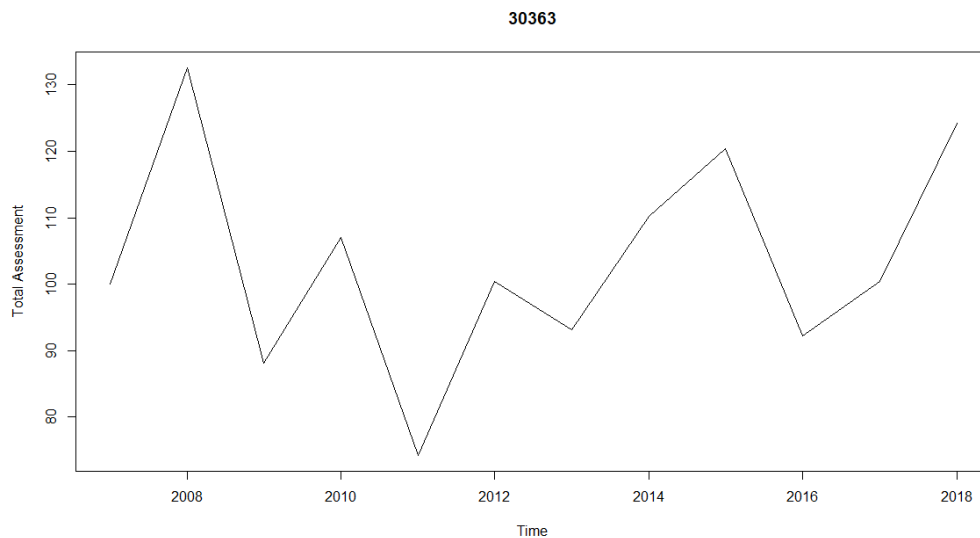
### **Methodology**

The only variable for our time series model was the total assessment field found in the [tax parcels data](#). But first we needed to find the zip codes associated with each parcel of land. We loaded the tax parcels 2018 shape file into the QGIS software. We also loaded this [zip codes](#) shape file into QGIS and did a spatial join with the tax parcels layer and then exported it into a csv file. This gave us the zip code for each tax parcel of land in Fulton.

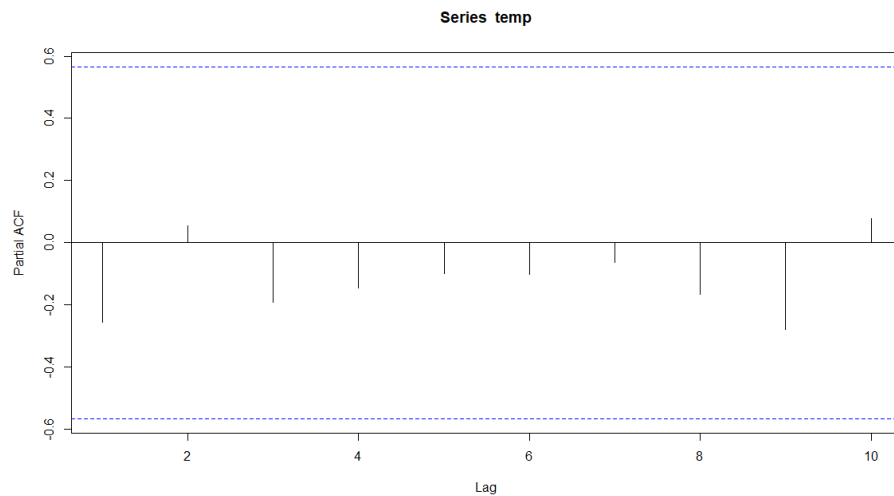
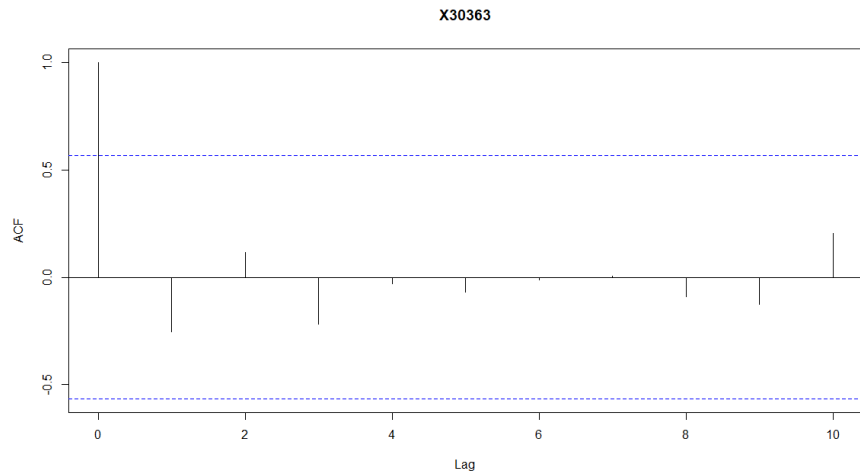
We then downloaded the tax parcels tables for each year until 2007. We then joined the tables only on the Parcel Id and total assessment columns in SQL to ensure that we were looking at the same parcels

for all the years. We then deleted records that had value of zero for any year in order to obtain a “cleaner” dataset. Finally, we aggregated the total assessment by zip code for every year.

We then indexed the data, that is we set the initial value of our index equal to 100 in 2007 (the base year) and then every year we updated our index based on the percentage change between the years. So for example if the aggregated total assessment changed by 30% from 2007 to 2008 then the index value in 2008 would be equal to 130. We then transformed this data into a time series format and plotted it. For example, the plot for zip code 30363 would be:



We then plotted the auto correlation (ACF) and partial autocorrelation (PACF) graphs of the data. These graphs allow us to determine the stationarity of the data. Data must be stationary in order to successfully forecast it using an ARIMA model. For all the zip codes the data appeared to be stationary. Here is the example of the ACF and PACF plots of the data for the 30363 zip code:



Then using R code we select the p and q orders of the ARIMA models. We used Aikake Information Criterion to select these orders. I will not delve into specifics but these are the parameters of the model, and we did this process manually for every zip code. I have attached the R code used to build these models for reference. After having created the model we predicted the future prices in five years, that is from 2019 to 2023.

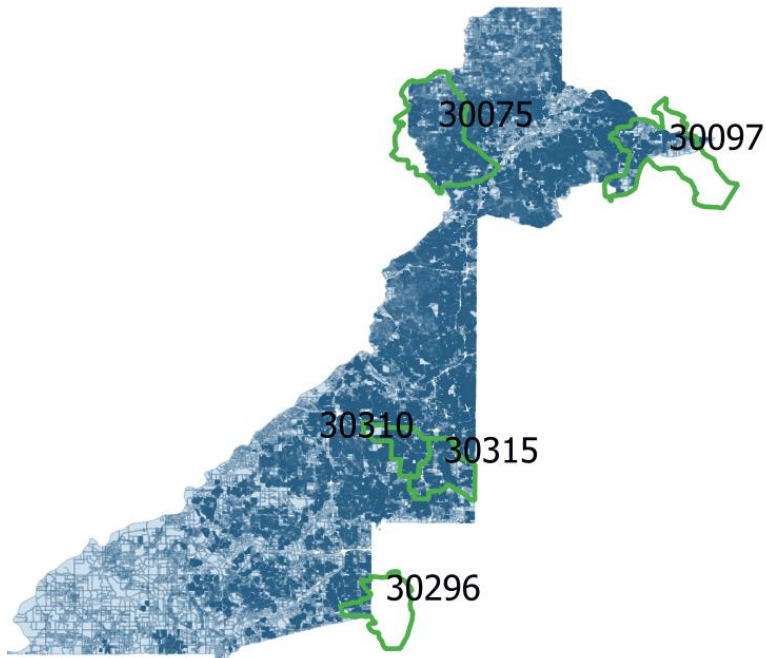
## Results

The table below contains the % change in total assessment from 2018 to 2023 by zip code. It is organized in descending order from highest percentage change to lowest percentage change:

ZIP	% change from 2018 -2023
30310	45%
30296	37%
30097	15%
30315	14%
30075	12%
30314	0%
30331	-1%
30336	-1%
30337	-1%
30338	-1%
30339	-1%
30342	-1%
30344	-1%
30346	-1%
30349	-1%
30350	-1%
30354	-1%
30360	-1%
30363	-1%
30040	-1%
30290	-1%
30022	-9%
30080	-9%
30306	-9%
30313	-9%
30076	-9%
30317	-9%
30318	-9%
30319	-9%
30320	-9%
30324	-9%
30326	-9%
30327	-9%
30328	-9%
30330	-9%
30005	-10%
30311	-10%
30092	-11%
30004	-11%
30316	-11%
30024	-11%
30303	-12%

30303	-12%
30268	-12%
30309	-13%
30115	-13%
30308	-13%
30305	-14%
30307	-14%
30188	-15%
30213	-15%
30096	-16%
30291	-19%
30062	-21%
30312	-21%
30263	-33%
30068	-35%

According to our results, Habitat for Humanity should focus on purchasing properties in Zip codes 30310, 30296, 30097, 30315, 30075. All of these zip codes have an expected increase in value by more than 10%. Here is a visual representation of these zip codes. The green lines represent the zip code boundaries, the blue fills represent the all the parcels of land inside Fulton county. Those parcels inside the green boundaries should be given the highest priority:



The Zip codes 30310 and 30315 correspond to neighborhoods south of Downton Atlanta. The zip code 30075 corresponds to Roswell and zip code 30097 corresponds to Duluth. Finally, zip code 30296 corresponds to Riverdale.

A more general comment seems to be that the majority of zip codes have a predicted decrease in price. This is pure speculation, but we suspect that since that the last recession was more than ten years ago, it is very likely that a new recession will happen in the upcoming five years. These ARIMA models seem to capture the seasonality of boom and bust cycles in the economy. In the upcoming years, this will cause real estate prices to drop. It seems that the zip codes for which we have predicted an increase in price will not be affected by this recession. The rising trend in the property prices in these areas will offset the negative effect caused by a future recession.

### **Property Analysis**

In order to determine the lots that fill Habitat for Humanity's criteria we need to start with a base table. For this study we will use the [Tax Parcels 2018](#). This is the same file that we used for our market analysis.

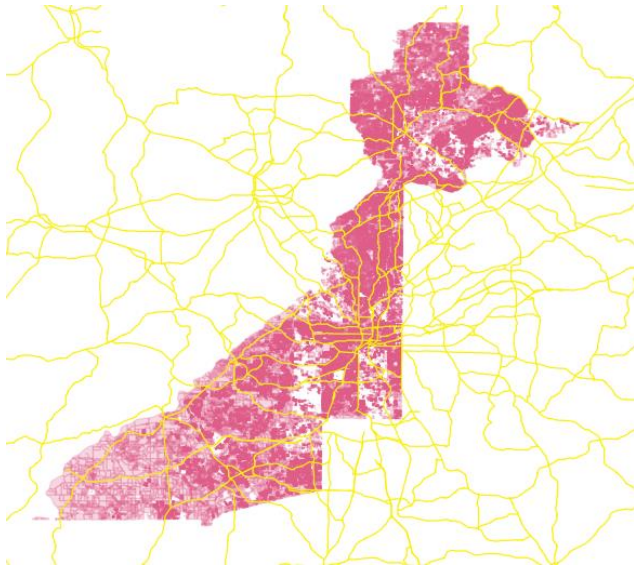
We will now examine how we filtered the different requirements that the lots must have.

#### **Flat**

The [Tax Cama](#) folder has a table called Tax\_ParcelSupplement with each tax parcel and its corresponding characteristics that have proved quite useful for our analysis. To filter the parcels with flat roads we did a JOIN operation in SQL with the Tax\_ParcelSupplement table and our base table and selected those parcels that had attribute 'Topo' equal to 'level'

#### **Not on a Major Road**

The following image is the Tax Parcels 2018 data loaded into QGIS in the shape format. Also loaded into QGIS is the [Major Roads](#) data set in shape format. The pink area corresponds to the tax parcels and the yellow lines correspond to the major roads



To obtain the parcels that are on a major road we first created a buffer layer from the major roads layer. This means that we created a new layer that corresponds exactly to the major roads layer, the difference is that the width of the layer is 250 meters. Then we did a spatial join of the buffered layer with the parcels. We reasoned that buffered layer with a width of 250 meters would touch all the lots that are right next to a major road. With this spatial join we were able to determine which lots are adjacent major roads and we filtered these out of our base file.

#### On a paved road

We did the same procedure as 'not on a major' road require. We did an INNER JOIN operation with our base file table and `Tax_ParcelSupplement` table, selecting those parcels that had attribute 'Street1' equal to 'Paved'

#### Not in a Flood Plain

In QGIS we were able to do a spatial join of our base file and the [Floodplains](#) dataset. Then we selected those parcels with attribute 'Zone' equal to 'X'. These are the areas that are outside of floodplains.

#### 75 feet away from a stream

We imported the [Bluelines Streams](#) dataset into QGIS and created a vector layer. We created a buffer vector layer from the blue lines streams vector that was 150 feet wide (so 75 feet to each side) and then did a spatial join with the tax parcels layer. Those parcels that touched the buffered layer were filtered out of our base file.

### Not in College Park

We imported the [City Limits](#) dataset into QGIS and created a vector layer. We did a spatial join with the tax parcels vector layer. With our joined layer we selected those attributes where the field City was not equal to 'College Park'.

### Lot frontage must be at least 50 feet and lot depth must be at least 80 feet

We reasoned that all lots that are 6400 square feet (80x80 feet) or greater would satisfy both requirements. If we have a square that is 80 feet on each side, then it will automatically satisfy the 50 feet of frontage requirement as well as the lot frontage of 80 feet. 6400 square feet is equal to 0.1469237, so selected those parcels where the field 'LandAcres' was equal or greater than 0.1469237.

### Sewer Connection

We joined our base table with the Tax\_ParcelSupplement table in SQL we selected those tables where the attribute Sewer was equal to 'All public' or 'All underground' or 'Public sewer' or 'Public water'. We determined that these criteria correspond to sewer connection.

### Not in a neighborhood association

In our base table if the attribute 'LUcode' was equal to 111 or 188 then the parcel is part of a neighborhood association. We filtered from our base files those parcels that met these criteria.

### If there is a house on the lot should we rehab or demolish i

Although we did not exactly determine this, we did determine which lots were vacant, the 'LUCode' (Land Use) field was equal to 100, 200, 300, 400, 500, 600 or 700 then the lot was vacant. We believe that the decision to rehab or demolish it requires further study of a particular plot and is outside the scope of this project.

### If the lot is in non-conforming zoning

From our base table, if the ClassCode field is equal to R1, R3, R4 or R5 then the parcel would be in a conforming zone based on this [information](#).

### Undetermined constraints

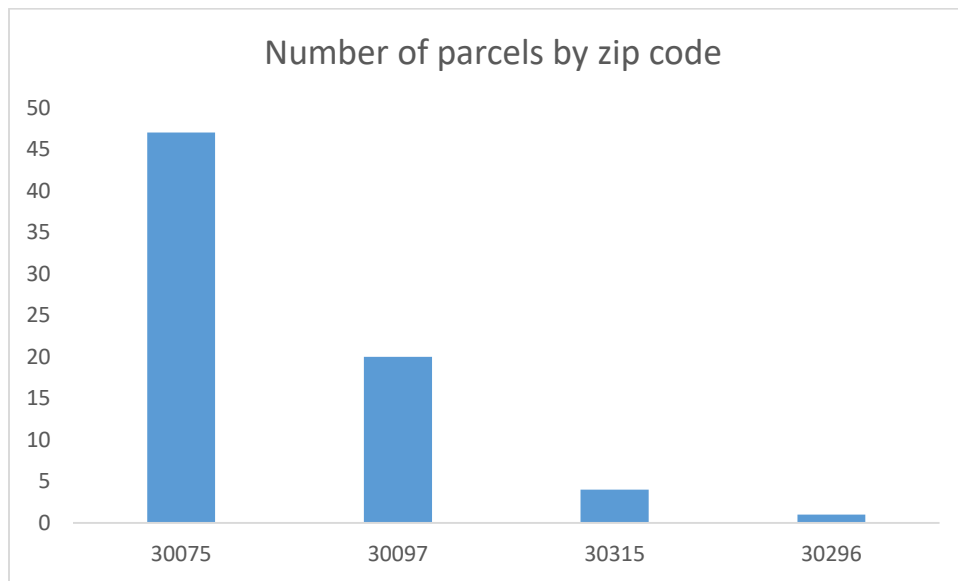
We were unable to determine the constraints No easements and no sewer setback easement



## Final List of Parcels

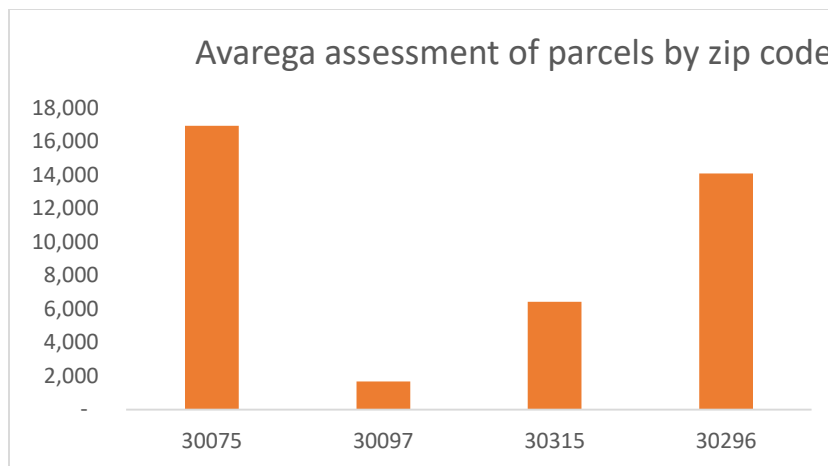
### Analysis

We have attached an csv file called 'final parcels' that contains a list of parcels with their attributes. These parcels fulfill all the criteria that we have described above as well as the zip codes where we have recommended focusing. In total the number of parcels is equal to 72. Here is a bar chart of our final list by zip code:



The zip codes with the highest amount of lot, 47, is 30075. The second one, 30097 has 20 lots and correspond to Dultuh. Zip code 30315 has 4 and 30296 only one.

The following graph shows the average assessment of the lots by zip code:



### Conclusion

Based on the previous two graphs it would make sense to first focus on those parcels found in zip code 30097, that is those parcels that are in Duluth. On our list there are 20 parcels and the total average assessment value is only \$2,000. Which means these parcels have a great price. In addition, it is in North Fulton an area that is easier for Habitat for Humanity to transport to.

The second area to focus on would be zip code 30075. There are 37 parcels to focus which gives more options, however the average assessment value is approximately \$17,000 which means that the parcels will be more expensive. This area is also in North Fulton.

The next step in this process would be for Habitat for Humanity to individually examine these parcels and research if these are for sale, and they do in fact meet the requirements.

We hope that this report can further aid Habitat for Humanity's mission of helping those in need of Affordable Housing.

We have also attached a csv file called 'parcels' which lists all the parcels based on our selection criteria.

