Predicting Gentrification in Denver, CO

overview.

Gentrification is becoming even more of a pressing issue as we see disadvantaged people pushed out of their neighborhoods as housing prices soar. I moved to Denver right after graduating college in 2016 and saw first hand how the city is changing, for better or for worse.

I lived in Denver's historically black neighborhood, Five Points. When I looked around my neighborhood, I saw middle class white families walking their dogs, upscale fried chicken restaurants, and an expensive cafe juxtaposed against a family owned soul food restaurant, and old-school car repair shop, and derelict houses that lined Welton Street.

Digging into the history of Five Points, it became apparent that the neighborhood has undergone a serious change in the past couple of years. The fancy fried chicken joint had replaced a family owned fried chicken joint. What used to be small, low cost shops were now breweries and yoga studios.

From my casual observation, it appeared that this all had devastating effects on the black population of Five Points. It appeared that the black population now existed further east of Five Points, where the cost of living is cheaper.

Specifically, I was curious as to how the legalization of marijuana in 2014 affected the rate of gentrification in Denver. Are we experiencing more gentrification now that marijuana is legal and Denver and other parts of the state are experiencing rapid population growth?

Regardless of your stance on gentrification, it will be valuable for developers and city officials to understand if a certain neighborhood is gentrifying. Being able to predict gentrification will allow appropriate parties to better plan for the future and potentially protect residents from being displaced.

methodology.

The process for this project is broken up into several parts:

1. Define gentrification
2. Determine scale of gentrification from 2011 to 2016
3. Hypothesize predictor variables for the year 2011 that might predict the outcome from the above calculation
4. Gather the data for the predictor variables and wrangle it into Python
5. Determine correlation between predictor variables and outcome variable
6. Create a linear regression model from the data
7. Determine accuracy of linear regression model

1. Gentrification Definition

In an early attempt at this project, I used the following definition from Wikipedia to determine whether a census tract had gentrified from 2011 to 2016:

> "Whether gentrification has occurred in a census tract in an urban area in the United States during a particular 10-year period* between censuses can be determined by a method used in a study by Governing:[50] If the census tract in a central city had 500 or more residents and at the time of the baseline census
>
> - had median household income and median home value in the bottom 40th percentile and at the time of the next 10-year census the
> - tract's educational attainment (percentage of residents over age 25 with a bachelor's degree) was in the top 33rd percentile;
> - the median home value, adjusted for inflation, had increased;
> - and the percentage of increase in home values in the tract was in the top 33rd percentile when compared to the increase in other census tracts in the urban area
>
> then it was considered to have been gentrified.

* I didn't want you to overlook this crucial detail! Standard gentrification models look over a 10-year period. I specifically chose to look at a 5-year window encompassing the time immediately before and immediately after the legalization of weed in Colorado.

While this formula carries some significant gravitas behind it, I ultimately found that it was not a good fit for analyzing Denver's changing urban landscape. From my observations of Denver while living here, this formula did not accurately classify census tracts as having gentrified or not. The above formula had to be tweaked to be more lenient to classify any census tract as having gentrified between 2011 and 2016.

Therefore, I decided to parse out one quantifiable component of gentrification. For the purposes of this project, I chose to define gentrification in terms of the increase in rent from 2011 to 2016.

2) Determine scale of gentrification from 2011 to 2016

Given that I was focused on rent increase as the sole measure of gentrification, the following procedure was conducted to extract the amount, in dollars, of the change in rent from 2011 to 2016.

Data were gathered from the American Community Surveys from 2011 and 2016. I pulled from the following three tables:

- Gross rent
- Median rent
- Gross rent by income

I created two dataframes --2011 and 2016--that contained the data from these three tables respectively.

With some simple math, I created a second dataframe to start building out a predictive model. This dataframe contains the median rent for both 2011 and 2016 as well as the difference between the two years.

data exploration mini game.

Let's look at some initial findings.

Where in Denver did we see the most increase in median rent? What parts of Denver experienced a *decrease* in median rent? Below is a map of Denver showing the increase or decrease of rent in Denver, CO (Figure 1).

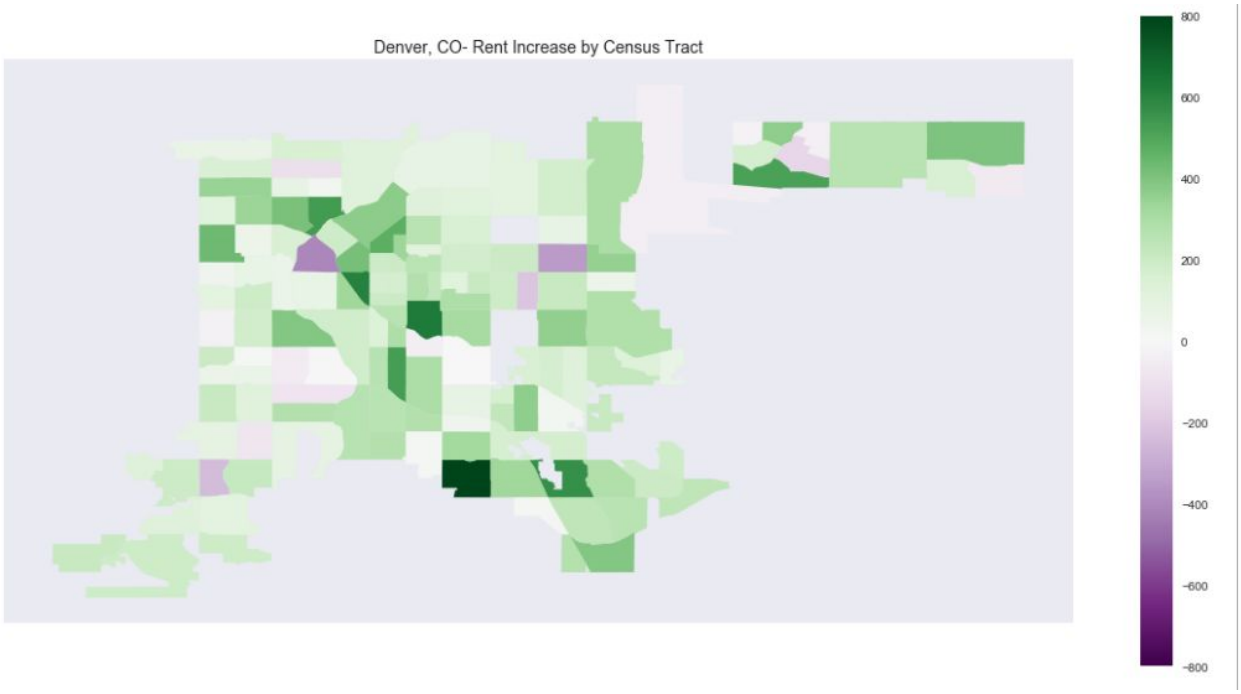

Figure 1. Rent increase between 2011 and 2016 in Denver, CO, broken down by census tracts. Some census tracts experienced a decrease in rent over these 5 years.

There is no strong spatial pattern to rent increase. It seems that overall, most areas had an increase in rent from 2011 to 2016. We observe a few areas where rent increased surrounding the central business district, but we also observe an "outlier" in south Denver.

back to methodology.

Are you still with me? Great, let's dig into the real task of building out our predictive model. First, we need to establish variables that we think might predict the results of the above analysis.

3) Hypothesize predictor variables

First, let's look at this white paper by Peter Koulizos from the University of South Australia titled, "Are there indicators that can be used as predictive precursors to identify gentrification within an area?"

The paper presents several variables that they found to predict gentrification. It should be noted that this model was fitted to the predetermined gentrifying suburb of Torrensville. Two benchmarks cities were used: Brooklyn Park as a non-gentrifying city and Unley as an already-gentrified city. Their findings are presented below:

> "This is when the results are compared to a neighbouring non-gentrifying suburb, a gentrifying suburb and the state average. The four indicators are:
>
> The greatest decrease in the people aged 18 years and under
> The greatest increase in couples without children
> The greatest increases in the people that lived at a different address five years ago.
> The greatest increase in the percentage of females working in professional occupations."

These predictor variables are still time-series data, but for the sake of the analysis we will start here. That is, how can patterns before the start of gentrification impact later gentrifcation?

The white paper also mentions some other interesting research:

> "The study has selected Torrensville and Brooklyn Park as the literature shows that two of the precursors to gentrification are proximity to the Central Business District (CBD) and historical homes (built before World War 2). Galster and Peacock (1986) in their study in Philadelphia identified historical housing and proximity to the CBD as two key elements that must be present for gentrification to take place. Kolka (2010) and Wholey (2009) also recognised historical housing as a prerequisite to gentrification."

Let's add these two historical variables to our analysis as well.

All together, the following are our predictor variables:

- Distance between census tract center and the CBD of Denver
- Number of historical homes
- The change in number of people aged 18 years and under in from 2000 and 2011
- The change in the number of couples without children in from 2000 and 2011

- The change in the number of people that lived at a different address in 2006
- The changed in percentage of women working in professional occupations from from 2000 and 2011

4) Gather data

Predictor Variable	Source	Method	Type
Distance between census tract center and the CBD of Denver	Google Map Census tract shapefiles	Compute Haversine distance between Denver CBD center and census tract centers using PyShp and GeoPy	float
Number of historical homes	Denvergov.org historical site data	Use Nominatim from geopy to determine a (lat,lon) for each historical site given the site's general street address. Use the Census API to map the census tract for each (lat,lon)	int
Change in the number of people aged 18 years and under from 2000 and 2011	Census.gov American Fact Finder	Sum columns for all age brackets for children. Note: one age bracket includes 19 year-olds	float
Change in the number of couples without children from 2000 to 2001	Census.gov American Fact Finder	Assume 2-person households are single family households	float
The number of people that lived at a different address in 2006	Census.gov American Fact Finder	Use "Same house 1-4 years ago" from 2011 survey to determine how many people moved between 2006 and 2010.	float
Change in the percentage of women working in professional occupations from 2000 to 2011	Census.gov American Fact Finder	Use the total estimate for employed females	float

5) Determine correlation

On the following page you can see the spatial patterns of the five predictor variables over Denver, CO (Figure 2).

Figure 2. Five spatial maps of predictor variables for rent increase for Denver, CO. The predictor variables are: the number of historical sites, the change in number of children, the change in household type, how many people had moved recently, and the change in working women for each census tract between 2006 and 2011.

To make things more clear, I created simple linear regression plots for each predictor variable against rent increase for each census tract. You can see the plots below (Figure 3).

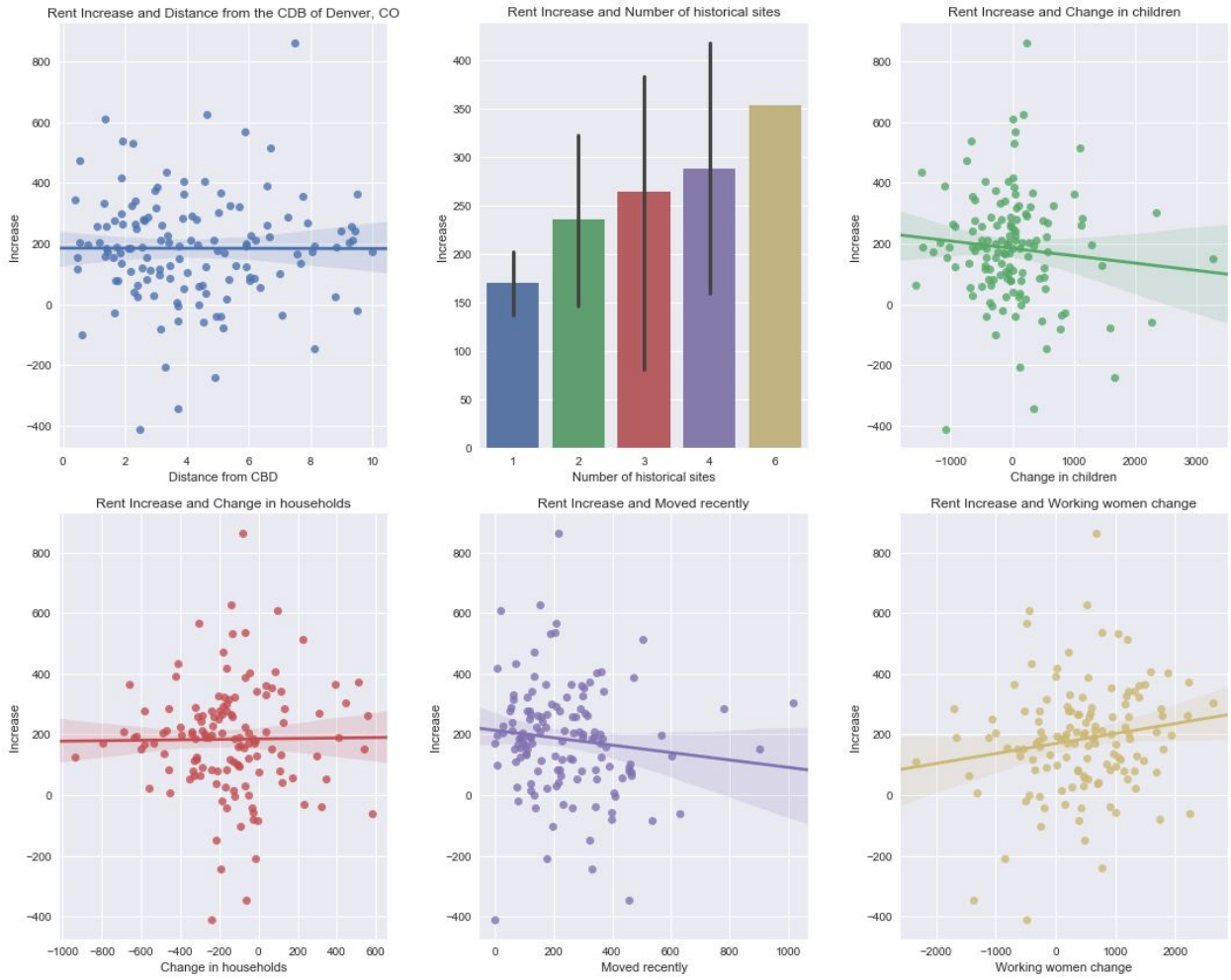

Figure 2. Six linear regression plots of predictor variables versus rent increase.

There is no clear correlation between distance from Denver's downtown and rent. This is surprising to me since my casual observation has led me to believe that areas like LoDo and Capitol Hill, areas near downtown, are some of the more expensive neighborhoods in Denver.

We do see a positive linear pattern between the number of historical sites and an increase in rent.

The remaining plots, unfortunately, have a lot of noise. While we may be able to squint and parse out positive or negative relationships, the truth of the matter is that these data don't seem to be saying much.

before we start modelling, let's pause and think.

The impetus for this investigation was a white paper that concluded that variables such as the change in working women and distance from a central business district might be early indicators of gentrification in cities. However, Koulizos is not clear on what definition of gentrification he is using for this research. He declares one of his three sample cities as "gentrifying," and uses that measure to judge whether the predictor variables are a good fit. The research this analysis was based on focused on whole cities, not census tracts.

The quantitative analysis by Koulizos is muddied. It seems that he compared the percent change between 2001 and 2006 and then 2006 and 2011. It is unclear how Koulizos synthesized these percent changes into a quantitative analysis, or if he chose variables that had the most percent change for the predetermined gentrifying city.

It would be unfair to pick on Koulizos alone. In preparing for this project, I searched for quantitative research on gentrification. I even met with subject matter experts on gentrification in Denver for their input. Trustworthy, quantitative analysis on gentrification is few and far between. I hoped that my project would illuminate some much needed answers on the subject.

But from out quick plots above it seems that there's just not a clear correlation between any of these variables outlined in the white paper and gentrification in Denver.

linear regression.

After playing around with decision trees and various linear regression models, I present below the results of a simple linear regression model for these data.

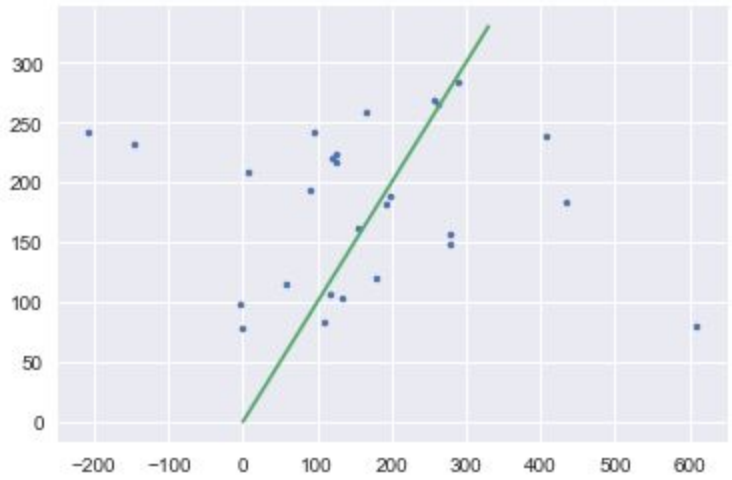

Figure 4. The predicted values for rent increase plotted against the actual values for rent increase.

While I could go on about the varying payoffs for using the L1 or L2 regularization, I hope that from this figure you can see that it would be moot. Every model that I tried produced the same results: an R^2 value of around 0 and a mean squared error of over 100.

conclusion

In the end, you simply can't fit a model to noise.

Maybe we can't simplify the complex process of gentrification to just an increase in rent. It could be that these variables can't predict gentrification at a granularity of a census tract. Maybe these variables don't align with Denver's gentrification process, which so closely follows the legalization of weed. Maybe we can only predict gentrification over longer time spans. It could be that we need a larger sample size, and need to look at census tracts over many cities.

This project was done as soon as I saw that there was no strong correlation between rent increase and any of the chosen predictor variables. You can solve this by attempting to redefine gentrification or by choosing different predictor variables -- or both.

For anyone interested in continuing this work, I suggest the following:

- Play with different definitions of gentrification. I tried two in this analysis and chose to stick with "rent increase" and my main quantitative metric.
- Explore the correlation of gentrification and business permit data. Would an increase in the number of business permits taken out in a neighborhood correlate with more gentrification?
- Try a larger sample size. One method would be to pick something like New York City. Another would be to look at census tracts of the top 10 cities in the United States.

references.

Galster, G & Peacock, S 1986, 'Urban gentrification: evaluating alternative indicators', Social Indicators Research, vol. 18, issue 3, pp. 321-337.
Koulizos, Peter. (2015). Are there indicators that can be used as predictive precursors to identify gentrification within an area?. 10.13140/RG.2.1.4757.4000.
Kolko, J 2010, 'The determinants of gentrification', SSRN Working paper Series.
Wholey, K 2009, 'Inner city revitalization: Mapping and predicting gentrification in the city of Chicago,