

Project: Capstone Project 1: Milestone Report

July 5th, 2018

Brittany Bennett

Springboard Data Science Career Track

Problem Statement

Gentrification is becoming more and more of an issue as we see disadvantages communities pushed out of their neighborhoods while housing prices soar. I moved to Denver right after graduation 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.

This all had devastating effects on the black population of Five Points, who were driven out of their homes and further east where the cost of living was cheaper.

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.

Dataset Description

I was primarily interested in how Denver had or had not gentrified since the legalization of marijuana in 2014. Therefore, I decided to look at the change in Denver from 2011 to 2016.

I used the following formula to determine if a census tract had gentrified or not. I compared census data from 2011 with census data from 2016 to make my decision. From Wikipedia:

"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 used this formula to determine which census tracts in Denver have gentrified from 2011 to 2016.

To build a predictive model of gentrification, I theorized some variable I believed were early signs of gentrification. I narrowed down my list to the two variables I believed were the biggest signs: new expensive restaurants and new cafes.

To build upon this study, I suggest also looking at the opening of art galleries and other institutions relating to art.

Data Wrangling

I first read in datasets from the 2011 and 2016 findings from the American Community Surveys and added them to their own dataframes. Data pertained to educational attainment, housing, and income. For both 2011 and 2016, I pulled the relevant columns for the variables in the gentrification formula: "median_income", "percent_bachelors", and "median_household_value".

The next step was to use these data to determine which census tracts in Denver had gentrified from 2011 to 2016 based on the formula presented above. There were four variables to consider:

gent_1: Bottom 40th percentile in median income for 2011 census data

gent_2: Bottom 40th percentile median household value for 2011 census data

gent_3: Top 30th percentile educational attainment for 2016 census data

gent_4: Top third median household value for 2016 census data

I created a dataframe, *gentrification*, that listed the outcomes for each of these variables per census tract. For example, if census tract 17.02 was in the bottom 40th percentile in median income for 2011, it received a 1.

The original formula states that if gent_1 and gent_2 are both 1, and gent_3 and gent_3 are both 1, the census tract can be said to have gentrified from 2011 to 2016. However, running this formula on the data produced poor results. I therefore experimented with variations of the formula to produce results that: a) matched with my understanding of Denver's neighborhoods and b) yielded at least 3 census tracts that were shown to have gentrified.

In the end, I decided to go with a more lenient description of gentrification for Denver. If either gent_1 or gent_2 was a 1, and either gent_3 and gent_4 was a 1, I classified the census tract as gentrified. I believe that this is an acceptable approximation of gentrification for Denver given the small sample size (one city versus an entire nation) and the small time difference (2011 to 2016).

With this calculation, I found that eight census tracts had gentrified from 2011 to 2016 (Figure 1)

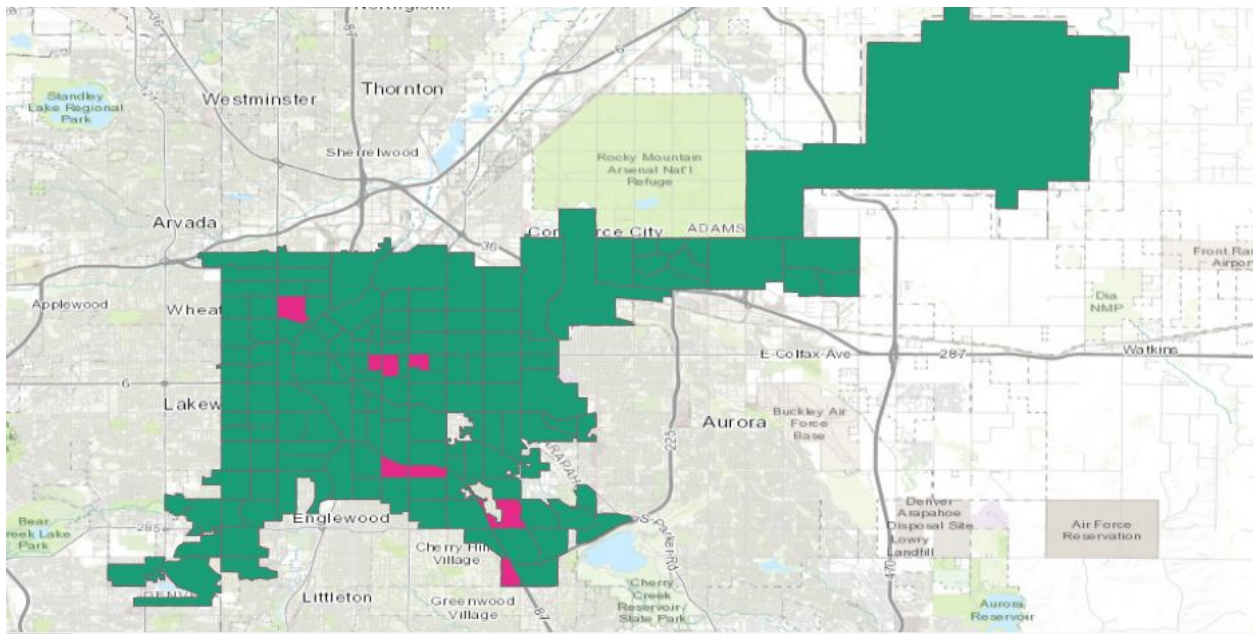


Figure 1: All 2010 census tracts in Denver, CO. Census tracts in pink are those that were determined to have gentrified from 2011 to 2016.

Assuming that this calculation is accurate for the given data, I then brought in data to build a predictive model of gentrification. While housing, education, and income data *define* gentrification, I hypothesized that the establishment of new expensive restaurants and cafes in an area could predict gentrification. In other words, I hypothesized that census tracts that had a lot of new cafes and expensive restaurants pop up around 2011, that census tract was more prone to become gentrified 5 years later.

Data on new cafes and expensive restaurants was gathered using Yelp's API. Given that Yelp's API can only return 50 results at a time, I picked three disparate longitude/latitude points in Denver to query.

I defined expensive restaurants as those rated with a 3 or 4 dollar sign (\$\$\$, \$\$\$\$), which is Yelp's indication of how much items on the restaurants menu cost. My query for cafes in Denver was not limited to expensive cafes, but did exclude all 7/11s.

Determining a restaurant's establishment date proved to be tricky. There is no field for "established on" on Yelp. I decided to take the data of the first review on Yelp as the founding year. However, Yelp's API does not allow on to query for specific reviews. Therefore, I exported the csv of restaurants and cafes and manually added data for a new column: year.

A second point of difficulty came from the need to assign each longitude/latitude point a census tract. I used ArcGIS to build a map of census tract polygon data and point data of new restaurants and cafes. I used the Sample tool in ArcGIS to assign each point the value of it's overlaying census tract.

Once I obtained these data and read it back into Python, I selected the new cafes and restaurants that were said to have established on or before 2014. I mapped the resulting data in ArcGIS (Figure 2).

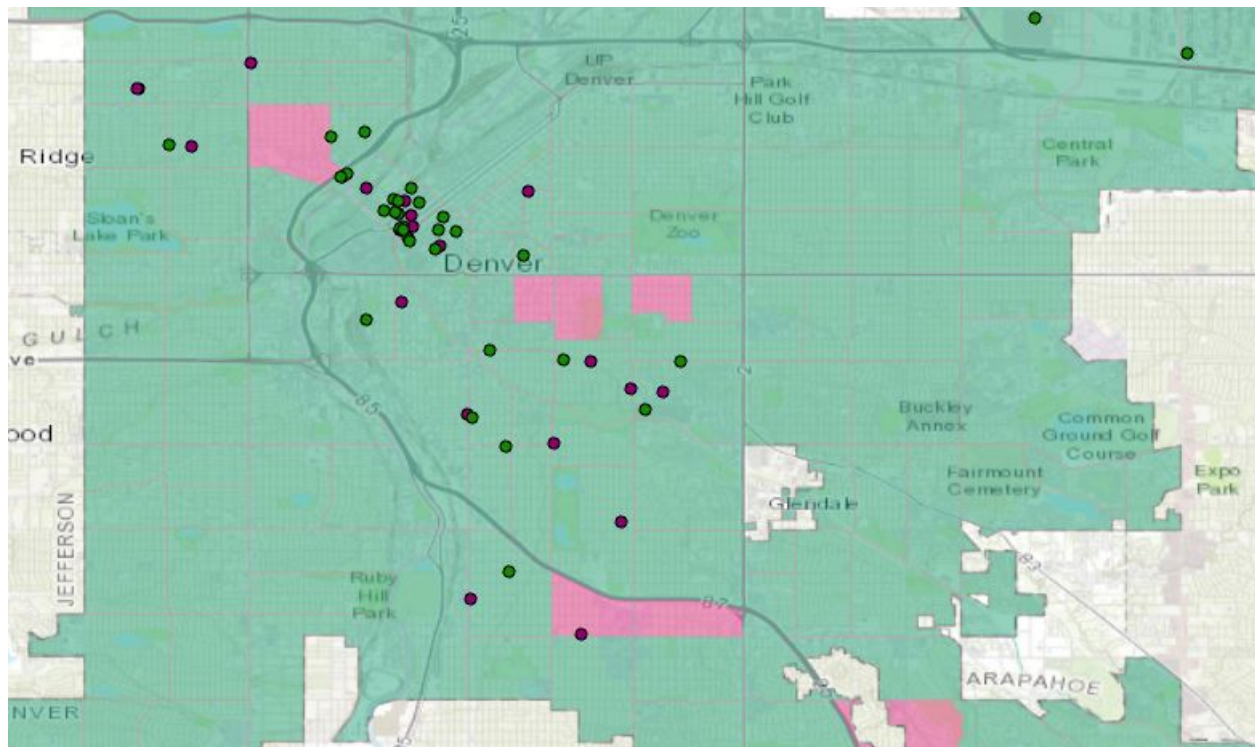


Figure 2: A snapshot of census tracts in Denver, CO, where pink census tracts are those calculated to have gentrified from 2011 to 2016. Each green dot represented a “new” expensive restaurant and each pink dot represented a “new” cafe.

The last step in creating the final data set was to add a column combining the counts for new cafes and restaurants. The head of the resulting data frame is presented in Table 1.

Table 1: The final dataframe summarizing the results from the data wrangling phase. For each census tract in Denver there exists data describing the number of new cafes, the number of new restaurants, whether that census tract was said to have gentrified, and the sum of the cafes and restaurants columns.

	tract	cafes	rest	is_gent	new_places
0	8031000301	2	0	0	2
1	8031000303	1	1	0	2
2	8031000401	1	0	0	1
3	8031001102	1	4	0	5
4	8031001701	7	11	0	18
5	8031001702	1	6	0	7

Initial Findings

From Figure 2 above, there does not seem to be a strong correlation between the number of new places in a census tract and whether it was determined to have gentrified from 2011 to 2016. However, it is worth pointing out that the census tracts that have gentrified near downtown Denver, which experienced many new cafes and restaurants popping up, also were census tracts in the bottom 40th percentile of housing costs (Figure 3).

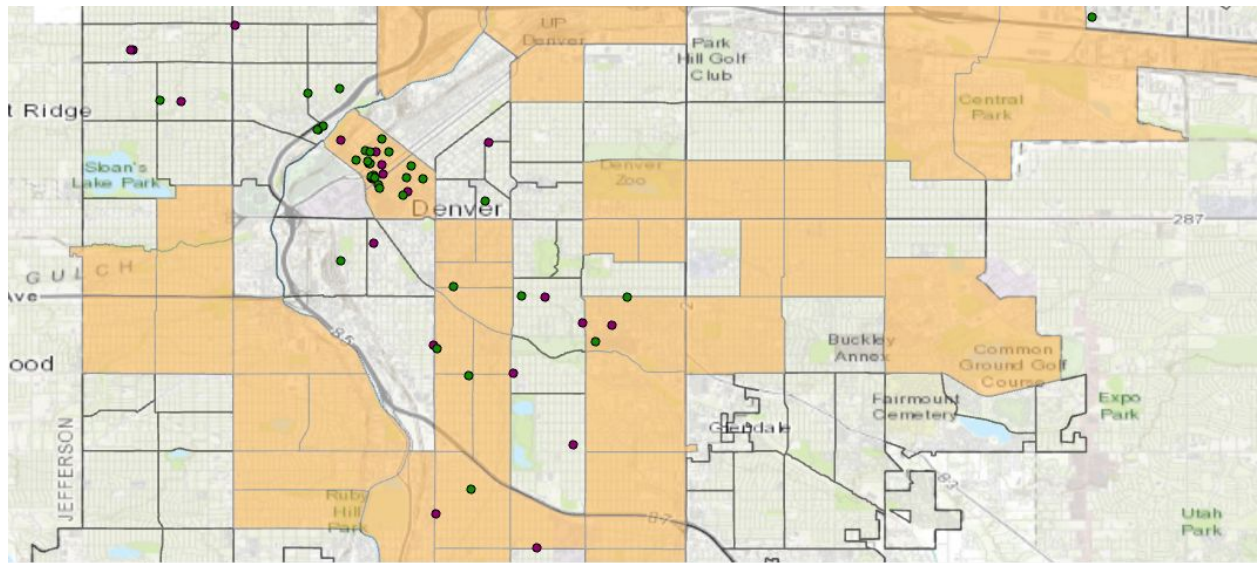


Figure 3: Map of new (opened from before 2011) cafes (purple) and new restaurants (green) overlaid on census tracts with bottom 40th percentile housing costs.

The common thread through the data explorations was the growth centered in downtown Denver. New restaurants and cafes popped up in high numbers concentrated in downtown Denver. In tandem, four census tracts near downtown Denver were classified as gentrified. These census tracts all were in the bottom 40th percentile for median household value.

There is not a clear pattern for why the southern census tracts gentrified.

Furthermore, my neighborhood, Five Points, was not classified as gentrified. I played with the formula for determining if a census tract had gentrified and was not able to classify Five Points as gentrified. It could be Five Points it's too recently gentrified. It could be that the formula used to determine gentrification is not well suited to the Denver region.

From the analysis above, I conclude that an early sign of gentrification for Denver is revitalization and growth downtown. The neighborhoods vulnerable to gentrification are those close to downtown that have cheap housing.

I used inferential statistics to ask the following question:

Are census tracts with more new cafes and restaurants more prone to gentrification 5 years later?

H0: There is not a significant difference in the number of new cafes and restaurants started up around 2011 for census tracts that have been shown to gentrify by 2016 in Denver.

HA: There IS a significant difference in the number of new cafes and restaurants started up around 2011 for census tracts that have been shown to gentrify by 2016 in Denver.

I determined that there is NOT significant difference in the number of new places in a census tract that does gentrify. The alpha for the permutation test I performed is $0.307 \gg 0.05$, so we cannot reject the null hypothesis.