

Finding the Ideal Neighborhood for Relocating to a New City

IBM Data Science Capstone Project

1. Introduction

Often times people need to move to a new place due to various different reasons, whether it is due to a change in job, school, or others. This new journey in life can be exciting and scary at the same time. You will be busying cleaning out your old house, saying farewell to old friends, and packing up for the new place. On top of all these, you will also need to decide on a new neighborhood to live in the new city. Needless to say, this can be very challenging, especially if you don't know the new city too well. You may be wondering if you can find a neighborhood that is similar to your current neighborhood. And if you are looking to buy a property, you would be interested in knowing what type of housing would be a better investment in the region. Luckily with the advance in technology, we can rely on information from the internet to aid in this process.

In this project, we will assume that our client will be relocating from **San Francisco, CA to Cincinnati, OH** due to a change in job assignment. The client currently lives in **Sunset District** and really enjoys the area -- therefore the client would like to find a similar neighborhood in Cincinnati. The client would also be looking into buying a property and would like to understand if buying a condo or a single-family house would be a better option in terms of price growth potential.

2. Data

Data Sources:

In order to determine a similar neighborhood in Cincinnati, we will need:

- A list of all neighborhoods in Cincinnati, which can be obtained from <https://www.zillow.com/cincinnati-oh/home-values/>
- Types of venues in each neighborhood in Cincinnati, which can be obtained using **Foursquare API**
- Types of venues in current neighborhood in San Francisco, which can be obtained using **Foursquare API**

And to understand what type of housing to buy, we will need:

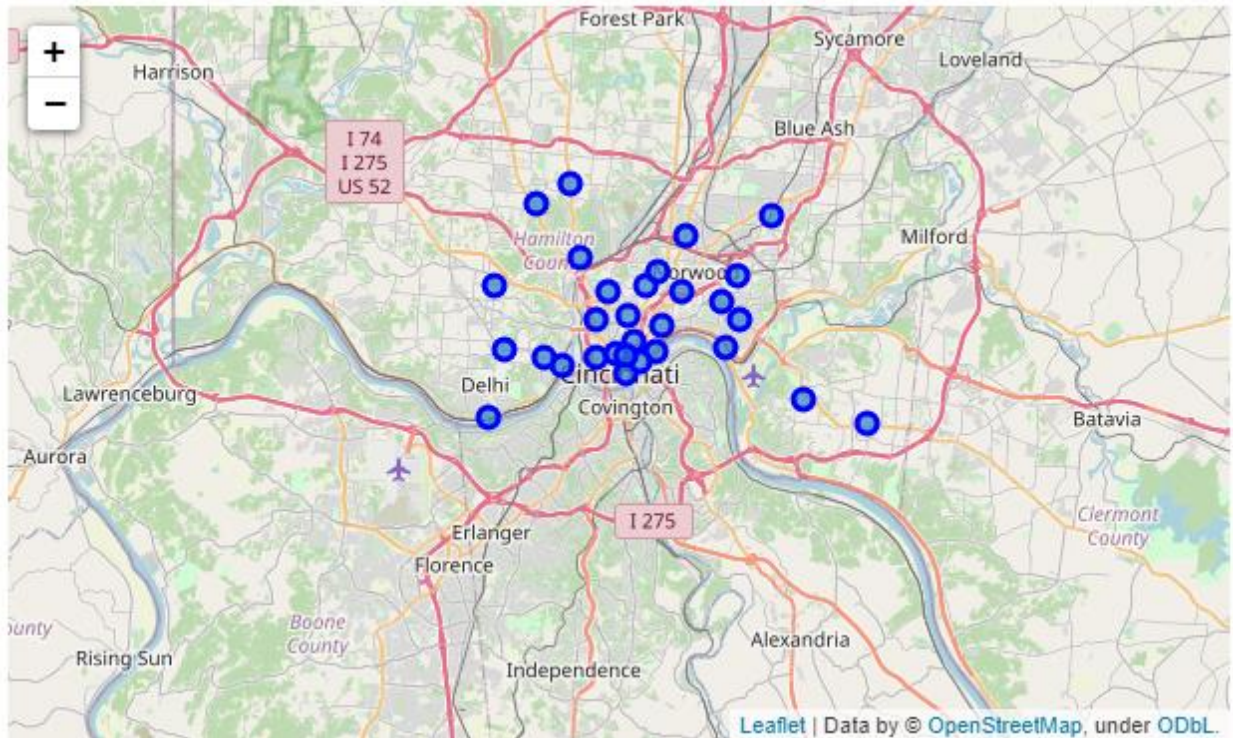
- housing price trend based on housing type for neighborhoods of interest in Cincinnati, which can be obtained from <https://www.zillow.com/research/data/>

Data Cleaning:

All data were downloaded in the mentioned websites and imported as separate dataframes. The table containing neighborhoods were cleaned so only neighborhoods in

Cincinnati remained. To confirm that the list was correct, the geospatial coordinates of each neighborhood were obtained using geocoders. All neighborhoods were graphed on the map using folium. The map shows all neighborhoods are indeed around Cincinnati, confirming that the list is correct.

Figure 1. Map of all neighborhoods in Cincinnati



Then, using Foursquare API and the table with all neighborhoods in Cincinnati plus Sunset District (along with the corresponding geospatial coordinates), a list of nearby venues was obtained for each neighborhood. All data was stored in one data for analysis later.

Lastly, the tables containing time series price data for condo and single-family house originally contained data for most of the neighborhoods in the United States. They were first filtered to containing only neighborhoods in Cincinnati. A close examination of the tables showed that they were structured and ready to be used for analysis later.

3. Data Analysis/Methodology

There are two tasks for this project – identifying neighborhoods in Cincinnati that are similar to Sunset District in San Francisco and determining whether buying a condo or single-family house in the selected neighborhood is a better choice.

Generally what defines a neighborhood would be the surrounding venues, or more specifically, the venue category. For example, if the client's current neighborhood is surrounded by restaurants, a similar neighborhood in Cincinnati should also be surrounded by restaurants.

Based on this, k-means clustering was used to divide all neighborhoods including Sunset District into a few groups based on similarities of nearby venues. The cluster containing Sunset District would be the one that we would further analyze. Recall that from the data preparation step, a table with all neighborhoods and corresponding nearby venues' detailed information such as geospatial coordinates, venue name, and venue category were obtained through Foursquare API. First of all, the "venue category" data was extracted from the table to be used for k-means clustering analysis. One hot encoding was used to transform the category data into numeric values and the density of venue category in each neighborhood was obtained. A range of k-values from 1 to 24 were used for the k-means clustering analysis and the corresponding cost function did not indicate a clear elbow point. In this case, k-value of 17 was used for the final analysis as this was where an obvious change in cost was noticed. Out of the 17 clusters, Sunset District belonged to cluster 0, along with 8 other neighborhoods in Cincinnati. This essentially completed the first task of this project.

For the second task where the client would like to understand the housing market in the potential neighborhoods, historic price of single-family house and condo were used for the analysis. The dataset was cleaned and restructured so that it only contained information for the interested neighborhoods identified by previous k-means clustering analysis. The house price was plotted vs time for all neighborhoods, both for condo and single-family house. This provides an overview of the trend for each neighborhood and a comparison between neighborhoods. Then, the house price of condo and single-family house was plotted for each neighborhood to provide a comparison for the impact of property type over time.

4. Results

From the k-means clustering analysis, a list of neighborhoods in Cincinnati that are similar to Sunset District was obtained, as they all belonged to the same cluster. Table 1 below shows all neighborhoods in that cluster, as well as the top 5 common venue categories in each neighborhood. From Table 1, we can see that indeed these neighborhoods are very similar. All neighborhoods are heavily populated with restaurants and bars. These are all potential neighborhoods in Cincinnati that the client can relocate to, with a similar vibe as Sunset District.

Table 1. Cluster containing Sunset District and neighborhoods in Cincinnati

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central Business District	Sandwich Place	Bar	Hotel	Coffee Shop	Mexican Restaurant
1	Columbia-Tusculum	Bar	Yoga Studio	American Restaurant	Baseball Field	Cajun / Creole Restaurant
2	Mt. Adams	Bar	American Restaurant	Trail	Ice Cream Shop	Bakery
3	Mt. Lookout	Bar	Sushi Restaurant	Pizza Place	Thai Restaurant	Bank
4	Northside	Bar	Dive Bar	Coffee Shop	Bakery	Brewery
5	Oakley	Furniture / Home Store	Pizza Place	Art Gallery	Bar	Rock Club
6	Over-The-Rhine	Bar	Bakery	American Restaurant	Cocktail Bar	Coffee Shop
7	Pendleton	Bar	Coffee Shop	Bakery	American Restaurant	Sports Bar
8	Sunset District(SF)	Chinese Restaurant	Bubble Tea Shop	Dim Sum Restaurant	Bakery	Middle School

The housing market for these neighborhoods was also analyzed and the price trend was visualized below. A first glance shows that all neighborhoods followed a similar trend in terms of increase/decrease in price. Overall, there was an increasing trend for house price from 1996 to 2008, followed by a decline from 2008 to 2012. Then, from 2012 onward, there was an increasing trend. These trends are similar regardless of whether it is a condo or single-family house. For condo price, Pendleton has seen the highest increase since 2012, while Central Business District has seen the lowest increase. For single-family house, Pendleton also has the highest increase since 2012, while other neighborhoods have a similar increase.

Figure 2. Price Trend for Condo in Interested Neighborhoods

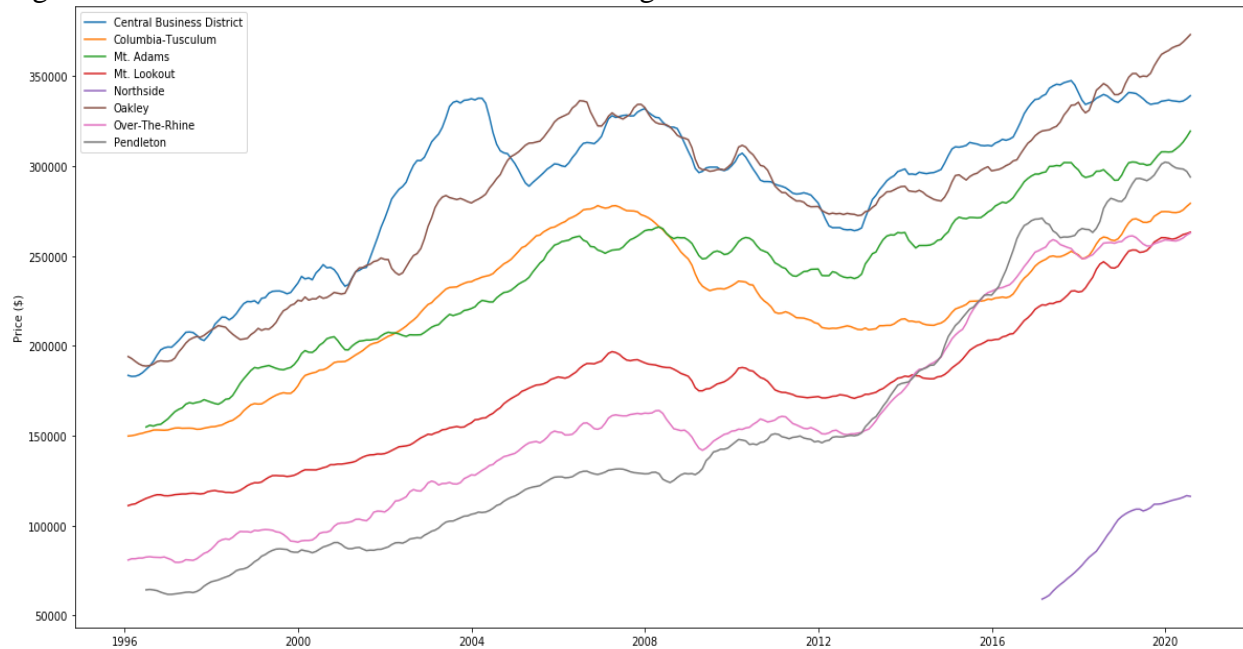
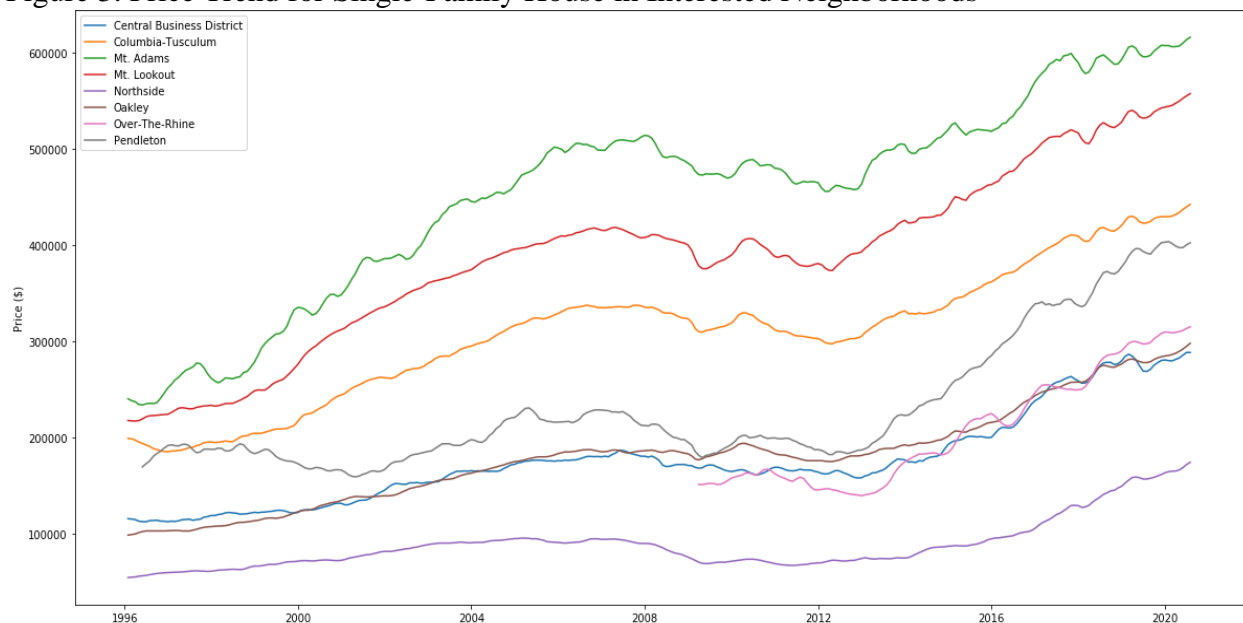
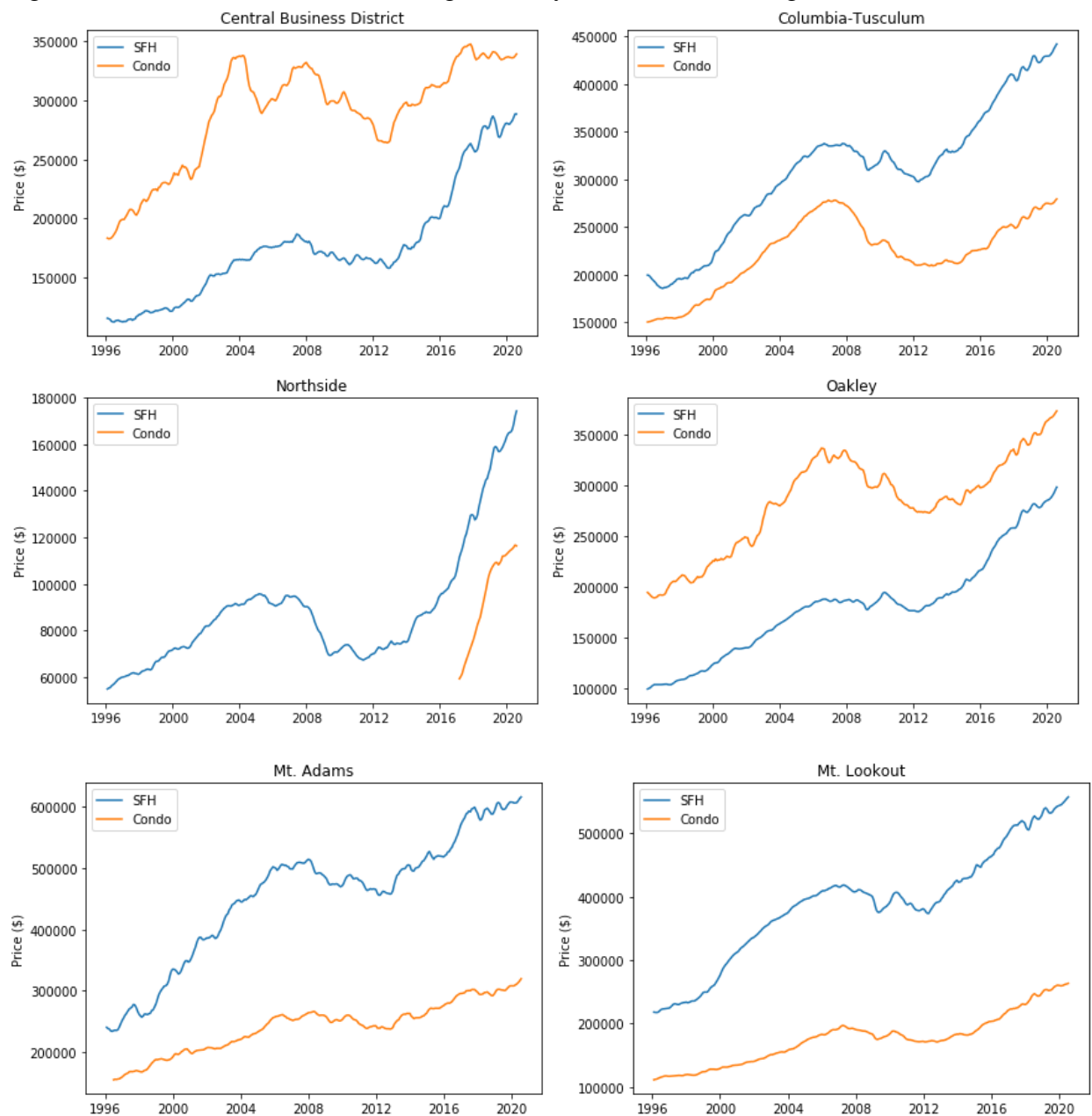


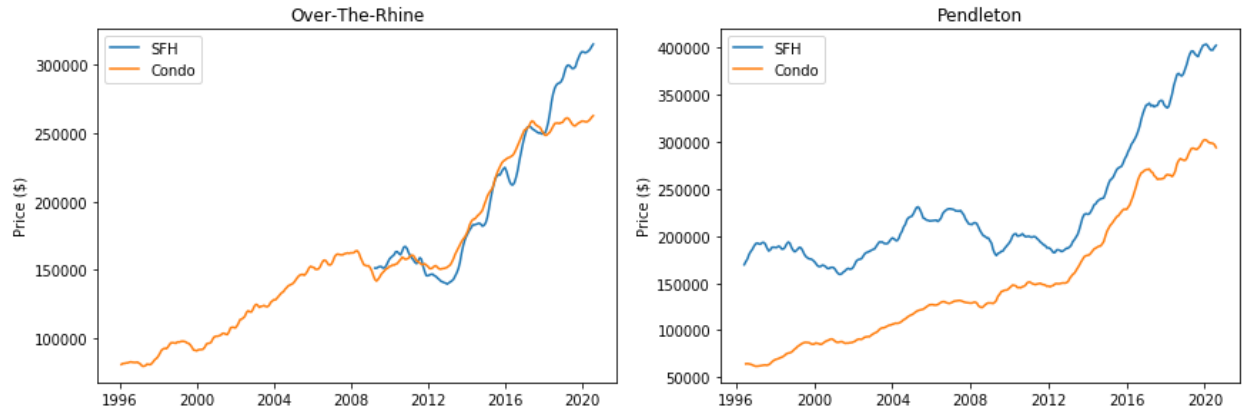
Figure 3. Price Trend for Single-Family House in Interested Neighborhoods



The impact of property type on price trend was also investigated by looking at the trend of condo vs single-family house in each neighborhood. From Figure 4, we can see that there is no neighborhood in which the rate of increase in price is higher for condo compared to single-family house. Most of the neighborhoods have higher rate of increase in single-family house, except Northside, Oakley, and Over-The-Rhine, where the rate of increase is similar for both property types. In terms of the actual price, the price of single-family house is always higher than price of condo in 5 neighborhoods (Columbia-Tusculum, Mt.Adams, Mt.Lookout, Northside, Pendleton), while the price of condo is always higher in only 2 neighborhoods (Central Business District, Oakley).

Figure 4. Price Trend for Condo vs Single-Family House in Each Neighborhood





5. Discussion

Based on the analysis and results, there are 8 neighborhoods in Cincinnati that the client can consider relocating to. The housing market analysis also provides additional information about the neighborhoods. First of all, the fact that all neighborhoods follow a similar trend in price throughout the years, regardless of property type, suggests that all neighborhoods have similar tolerance to changes in economy. For the initial increasing trend from 1996 to 2008, this is probably a reflection of the prospering economy during the same time. From 2008 to 2012, the housing market saw a decline due to the Great Recession. During this time period, however, the rate of decline is higher in condo vs single-family house, suggesting that single-family house is better at holding up its values during economy downturns. Lastly, from 2012 onward, the housing market started to recover from the Great Recession, and an increasing trend was observed on all neighborhoods. In addition, the rate of increase in single-family house price is always higher or similar compared to that of condo price, again making single-family house a better option for investing.

6. Conclusion

In conclusion, there are 8 neighborhoods in Cincinnati that are similar to the current neighborhood that the client lives in, and hence giving the client a few options to consider for relocation. The additional information from the housing market analysis will also influence the client's decision in choosing the final neighborhood, as well as what type of property (condo vs single-family house) to buy. In terms of growth potential, single-family house is always a better option, regardless of the neighborhood, as it can hold up its value during economy downturns as well as seeing higher rate of increase when the economy is performing well. Besides the factors mentioned in the analysis, the client might also need to consider other factors such as commute time, budget for housing, or others when choosing the final neighborhood. This report will serve as the first step in helping the client narrow down house search options.