

# Determining the lowest priced neighborhood amongst similar Denver neighborhoods

Piyush Mehta

January 16, 2020.

## 1. Introduction

### 1.1 Background

Denver is a [fast growing city](#). It was named the best place to live in the United States by [U.S. News & World Report](#).

Denver comprises of [78 neighborhoods](#). As is to be expected, each neighborhood has its own flavor, with certain neighborhoods exhibiting similar characteristics. Prospective home buyers may benefit with information that identifies neighborhoods with the same set of characteristics and the median prices for houses in such neighborhoods. The prospective home buyer may then choose the neighborhood with the preferred characteristics with the lowest median price..

### 1.2 Problem

How does a prospective home buyer in Denver determine the least expensive neighborhood to buy a home in amongst neighborhoods with similar characteristics?

### 1.3 Interest

Clearly, a prospective homebuyer would be very interested in knowing the characteristics of neighborhoods, neighborhoods with similar characteristics and the median home prices in these neighborhoods. Insight into this information would be one of many inputs that the home buyer will consider when narrowing the neighborhood in which to purchase their home.

## 2. Data acquisition and cleaning

### 2.1 Data sources

- Names of the 78 neighborhoods (scraping Wikipedia)
- Latitude and longitude of each neighborhood (using Google Maps API)
- Venue data for the 78 neighborhoods (using FourSquare API)
- Median housing price for each neighborhood (using Zillow API)
- Neighborhood GeoJson data to plot the choropleth map (derived from sources from the Denver Open Data Catalog)

The names of the 78 neighborhoods are scraped form [List of neighborhoods in Denver](#). From that list of neighborhood names, the Google Maps API is used to determine the latitude and longitude of each neighborhood. The latitude and longitude information is next used to get venue data from the FourSquare API. The venue data is used to cluster the 78 neighborhoods around the venue category types (restaurant, coffee shop, park, etc.). The number of clusters is determined by evaluating the results of the KMeans cluster algorithm. Next, the Zillow APIs are used to get the median housing prices for the neighborhoods. The clusters are plotted on a choropleth map to give the home buyer a visual representation of median home prices in clusters of similar neighborhoods. The GeoJson data for the choropleth map is derived from the shapefile obtained at [Statistical Neighborhoods](#). The shapefile is loaded in a program such as [QGIS](#) and the GeoJson data is then exported.

### Example workflow:

An example neighborhood is 'Central Business District'. We use this in the Google Maps API '<https://maps.googleapis.com/maps/api/geocode/json?address=>'. The response is parsed to determine the latitude and longitude. These in turn are used as inputs to the FourSquare API '<https://api.foursquare.com/v2/venues/explore>' to determine the venues of interest within a 750 meter radius of the location. The venue types will be extracted and the neighborhoods will be clustered around the venue types. The neighborhood name is also used in the Zillow API '<http://www.zillow.com/webService/GetDemographics.htm>' to retrieve the median home prices in the neighborhood. Finally, the clustered neighborhoods are visualized against a choropleth map of the neighborhoods based on their median prices.

## 2.2 Data cleaning

Some of the neighborhood names have special characters in them. The names must be URL encoded prior to invoking the Google Maps API. When retrieving the venue information from the FourSquare API, the radius needs to be tuned to ensure that each neighborhood returns at least one venue result. The Google Maps API and the Zillow API both cue off of the neighborhood name. It is possible however that certain neighborhood names exist in one and not the other. Data from the two APIs must be consolidated. The geographic (GeoJson) data used for the choropleth map is derived from a shapefile representing the neighborhoods. Care must be taken to ensure that this data is consistent with the neighborhood information from other sources.

## 3. Exploratory Data Analysis

### 3.1 Plotting the 78 neighborhoods in Denver on a Folium map

The neighborhoods are retrieved from [List of neighborhoods in Denver](#). The coordinates are next obtained using the Google Maps API. The neighborhoods are then plotted on a Folium map:

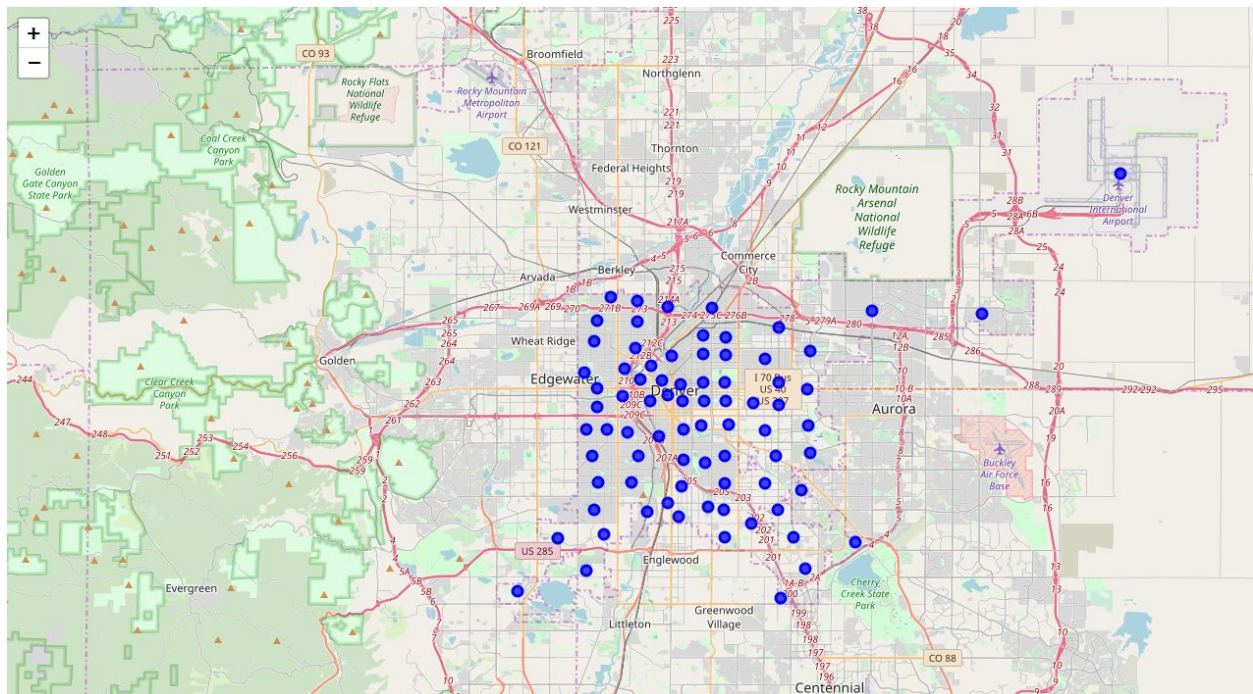


Figure 1. The 78 Denver neighborhoods plotted on a Folium map

### 3.2 Clustering neighborhoods based on the Venue Category Type

Given the coordinates of each neighborhood, the Zillow APIs are used to retrieve the venues in a 700 meter radius. Care was taken to choose a radius size that would ensure that at least one venue was obtained per neighborhood. In all, across 78 locations, 2205 venues across 299 venue categories were retrieved. An abbreviated count of venue categories per neighborhood is

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Athmar Park, Denver	4	4	4	4	4	4
Auraria, Denver	39	39	39	39	39	39
Baker, Denver	30	30	30	30	30	30
Barnum West, Denver	6	6	6	6	6	6
Barnum, Denver	8	8	8	8	8	8
Bear Valley, Denver	4	4	4	4	4	4
Belcaro, Denver	1	1	1	1	1	1
Berkeley, Denver	60	60	60	60	60	60
Capitol Hill, Denver	89	89	89	89	89	89
Central Business District, Denver	100	100	100	100	100	100
Chaffee Park, Denver	14	14	14	14	14	14
Cheesman Park, Denver	16	16	16	16	16	16

Figure 2. Count of venue categories per neighborhood

The venue categories are one hot encoded. The 2205 venues are grouped by neighborhood by taking the mean of the frequency of occurrence of each category. An abbreviated view of the top 5 venue category types per neighborhood is visualized in Figure 3.

Denver_Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0 Athmar Park, Denver	Park	Vietnamese Restaurant	Home Service	Warehouse Store	---
1 Auraria, Denver	Theme Park	Theme Park Ride / Attraction	Light Rail Station	American Restaurant	Fast Food Restaurant
2 Baker, Denver	Coffee Shop	Marijuana Dispensary	Chinese Restaurant	Intersection	Residential Building (Apartment / Condo)
3 Barnum West, Denver	Mexican Restaurant	Convenience Store	Discount Store	Home Service	American Restaurant
4 Barnum, Denver	Mexican Restaurant	Hookah Bar	Marijuana Dispensary	Market	Burger Joint
5 Bear Valley, Denver	Park	Gym / Fitness Center	Pool	---	---
6 Belcaro, Denver	Park	---	---	---	---
7 Berkeley, Denver	Brewery	Coffee Shop	Pizza Place	Breakfast Spot	Bar
8 Capitol Hill, Denver	Pizza Place	Coffee Shop	Sandwich Place	Nightclub	Pub
9 Central Business District, Denver	Hotel	American Restaurant	Theater	Italian Restaurant	Coffee Shop
10 Chaffee Park, Denver	Grocery Store	Pizza Place	Hobby Shop	Park	Thrift / Vintage Store
11 Cheesman Park, Denver	Botanical Garden	Garden	Park	Café	Ice Cream Shop

Figure 3. Top 5 venue category types per neighborhood

Next, The KMeans clustering algorithm (from sklearn.cluster) is used. Since the optimal number of clusters is not known, we try each value between 1 and 14. We use the elbow method using inertia (also known as sum of squared distance) to determine the optimal number of clusters. From the image below, the optimal number of clusters could be deemed as 6

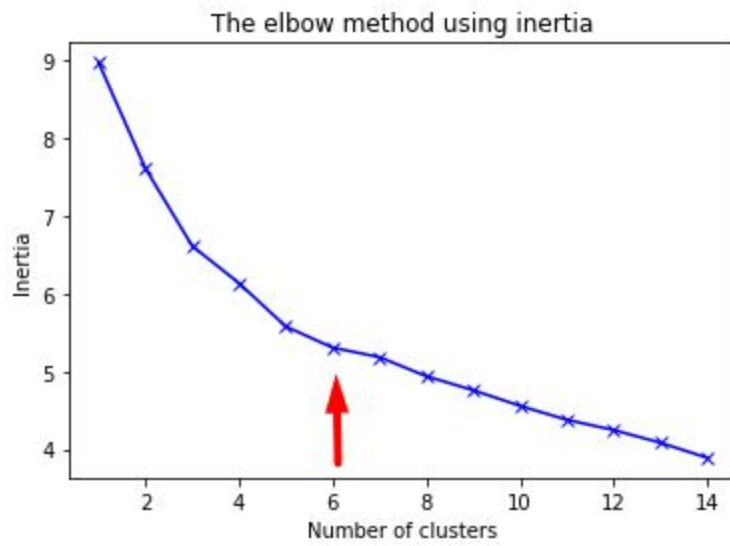


Figure 4. The elbow method using inertia to determine the optimal number of clusters

Figure 5. displays the number of neighborhoods in each cluster.

Neighborhood	
Cluster Label	
0	7
1	64
2	2
3	1
4	2
5	2

Figure 5. The number of neighborhoods in each cluster

Figure 6. displays the neighborhoods plotted as clusters on a Folium map.

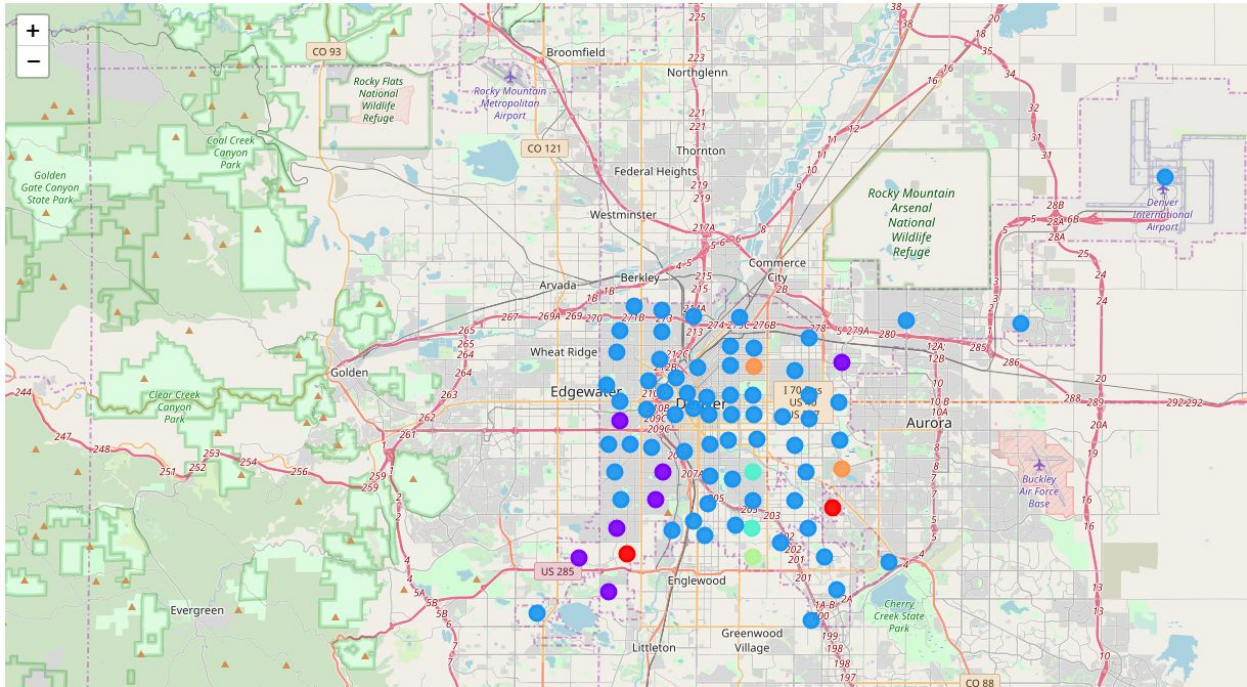


Figure 6. Neighborhoods plotted as clusters on a Folium map



#### 4. Results

The Zillow APIs are used to determine the median housing prices for each of the neighborhood. The housing prices are used to plot a choropleth map of Denver displaying the neighborhoods in varying shades based on the median housing price in the neighborhood. This information allows us to pick a neighborhood with the lowest median housing prices in a cluster of neighborhoods.

From Figure 7., for cluster 0, the neighborhood with the lowest median housing prices would be the neighborhoods of Harvey Park (on the pop-up below - \$354,100), Ruby Hill (\$352,100) and Athmar Park (\$381,300) and the highest would be the neighborhood of Stapleton (\$616,900). From Figure 8. below, it may be concluded that cluster 0 is a cluster of neighborhoods that predominantly has parks. A prospective homebuyer, primarily interested in the outdoors (parks), would benefit most by purchasing her home in the Ruby Hill neighborhood.

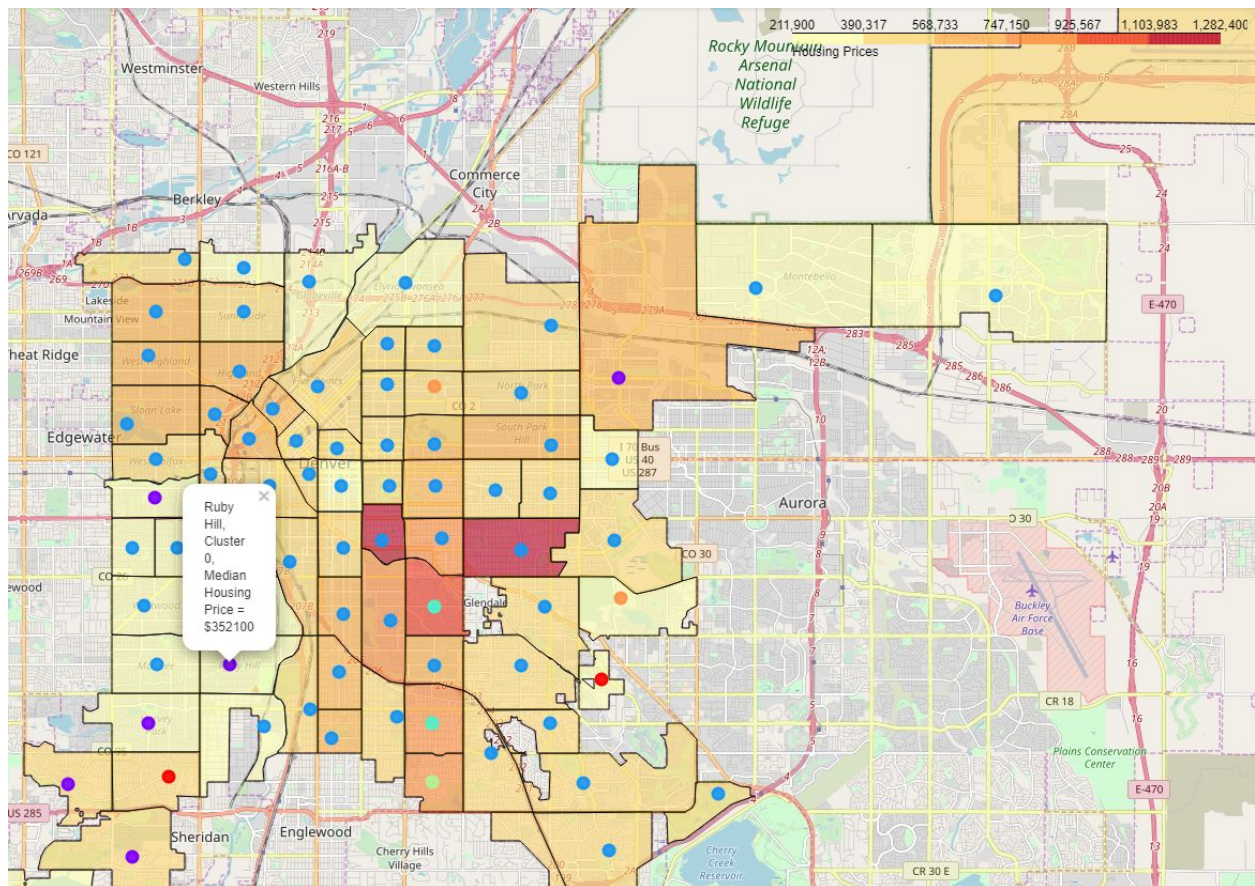


Figure 7. A choropleth map based on housing prices showing the neighborhood clusters

Cluster Label	Denver_Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	House Price
0	Athmar Park, Denver	Park	Vietnamese Restaurant	Home Service	Warehouse Store	---	354100
	Bear Valley, Denver	Park	Gym / Fitness Center	Pool	---	---	415400
	Fort Logan, Denver	Park	BBQ Joint	Music Venue	Historic Site	---	428900
	Harvey Park, Denver	Historic Site	Construction & Landscaping	Health Food Store	Lake	Mexican Restaurant	381300
	Ruby Hill, Denver	Flower Shop	Gym / Fitness Center	Park	Veterinarian	Home Service	352100
	Stapleton, Denver	Park	Wine Shop	Breakfast Spot	Pool	Gym	616900
	Villa Park, Denver	Park	Liquor Store	Food & Drink Shop	Pizza Place	---	370100
1	Auraria, Denver	Theme Park	Theme Park Ride / Attraction	Light Rail Station	American Restaurant	Fast Food Restaurant	706800
	Baker, Denver	Coffee Shop	Marijuana Dispensary	Chinese Restaurant	Intersection	Residential Building (Apartment / Condo)	501100
	Barnum West, Denver	Mexican Restaurant	Convenience Store	Discount Store	Home Service	American Restaurant	346200
	Barnum, Denver	Mexican Restaurant	Hookah Bar	Marijuana Dispensary	Market	Burger Joint	333100
	Berkeley, Denver	Brewery	Coffee Shop	Pizza Place	Breakfast Spot	Bar	597400
	Capitol Hill, Denver	Pizza Place	Coffee Shop	Sandwich Place	Nightclub	Pub	327700
	Central Business District, Denver	Hotel	American Restaurant	Theater	Italian Restaurant	Coffee Shop	423000
	Chaffee Park, Denver	Grocery Store	Pizza Place	Hobby Shop	Park	Thrift / Vintage Store	377800
	Cheesman Park, Denver	Botanical Garden	Garden	Park	Café	Ice Cream Shop	401600
	Cherry Creek, Denver	Clothing Store	Cosmetics Shop	American Restaurant	Coffee Shop	Furniture / Home Store	904700
	Cherry Creek West, Denver	Disc Place	Mexican Restaurant	Park	American Restaurant	Bookstore	544700
	Washington Park, Denver	American Restaurant	Gift Shop	Massage Studio	Boutique	Coffee Shop	916700
	Washington Virginia Vale, Denver	Liquor Store	Pizza Place	Mexican Restaurant	Discount Store	Track	433400
	West Colfax, Denver	Ice Cream Shop	Coffee Shop	American Restaurant	Convenience Store	Mexican Restaurant	562500
	West Highland, Denver	Coffee Shop	Pizza Place	American Restaurant	Brewery	Grocery Store	619000
	Westwood, Denver	Mexican Restaurant	Museum	Food	Nightclub	Hobby Shop	310400
	Whittier, Denver	Liquor Store	Grocery Store	Café	Bar	Baseball Field	564000
	Belcaro, Denver	Park	---	---	---	---	1101000
	University Park, Denver	Park	Cosmetics Shop	---	---	---	816600
3	Wellshire, Denver	Smoke Shop	---	---	---	---	815100
4	Skyland, Denver	Gym	Shop & Service	---	---	---	505800
	Windsor, Denver	Gym	Pool	Grocery Store	Dog Run	---	211900
5	Harvey Park South, Denver	Indie Theater	Construction & Landscaping	Disc Golf	Park	---	417400
	Indian Creek, Denver	Trail	Park	Construction & Landscaping	---	---	298100

Figure 8. An abbreviated view of each cluster listing the top 5 venue categories and the median housing prices

A limitation of this approach is that the clustering may be lop-sided in that a significant number of neighborhoods are grouped in one cluster. For instance, cluster 1 has 64 neighborhoods whereas cluster 3 has only 1. With 64 out of 78 neighborhoods in cluster 1, the prospective home buyer does not have a wider breakdown of neighborhoods to make a finer-grained selection.

## 5. Conclusion

This Capstone Project has the potential to assist a prospective homebuyer purchase a home in the least expensive neighborhood among a group of neighborhoods with similar characteristics. The data is presented both visually and numerically so that the homebuyer may make an informed decision.