

Choosing a Location for a Medical Center

Breeann Flesch

June 2020

Contents

1	Introduction	3
2	Literature Review	3
3	Methodology	3
4	Results	8
5	Conclusion	11
6	References	13

1 Introduction

It is a pandemic and the state of Colorado may not have enough hospital beds (1). The government of the state of Colorado has decided to add another medical center to their state, to ensure that all of their citizens will be served. They want to place this medical center optimally - in a location that will best serve the state. However, they only have enough funding to build one new medical center. So the question is: where is the best location to place a new medical center during the pandemic?

2 Literature Review

To determine the best place to put a medical center, we need to know about the disease that is causing the pandemic. This pandemic is hitting people of color particularly hard, so they want to prioritize communities with a higher than average percent of folks of color (2). Of those locations, they also want to choose a community with less access to medical centers than other communities. This disease also seems to be harder on men than it is on women, so the gender breakdown of the communities may be a consideration (3). We also know that the outbreak has hit densely populated urban areas hardest, so population of the communities should be considered when reviewing the areas (4). Lastly, we want to ensure that all of the citizens of Colorado have access to medicine, so we want to prioritize communities that do not have as many medical centers.

3 Methodology

“A community is a social unit (a group of living things) with commonality such as norms, religion, values, customs, or identity. Communities may share a sense of place situated in a given geographical area (e.g. a country, village, town, or neighbourhood) or in virtual space through communication platforms” (5). When defining the communities of Colorado, the ‘norms’ part of the definition is important for our context, because when folks are sick and vulnerable they want a medical center to visit that makes them feel comfortable. However, since we seeking a location for our medical center, the communities need to be geographically near as well.

There are some natural, built-in ways to define communities, especially with the emphasis on geography, like postal codes and cities. However, there are 521 different postal codes (also known as zip codes) (6) and 451 different cities (7) in the state of Colorado, which is too many to analyze effectively. Furthermore, some of the towns are quite close together, especially in the Denver Metro Area. Thus we look towards heuristics to define the communities, and we are looking for no more than 12 communities to analyze.

To start this analysis, we need the postal codes and their location, as latitude and longitude. To get this, we used the United States zip code database (8). We imported this data as a .csv file and created a dataframe.

	zip	primary_city	state	county	latitude	longitude
0	80001	Arvada	CO	Jefferson County	39.82	-105.10
1	80002	Arvada	CO	Jefferson County	39.79	-105.11
2	80003	Arvada	CO	Jefferson County	39.83	-105.06
3	80004	Arvada	CO	Jefferson County	39.82	-105.10
4	80005	Arvada	CO	Jefferson County	39.85	-105.13
5	80006	Arvada	CO	Jefferson County	39.82	-105.10
6	80007	Arvada	CO	Jefferson County	39.86	-105.19
7	80010	Aurora	CO	Arapahoe County	39.74	-104.86
8	80011	Aurora	CO	Arapahoe County	39.74	-104.78
9	80012	Aurora	CO	Arapahoe County	39.69	-104.81

Figure 1: The first 10 rows of the Colorado zip codes dataframe.

Next, we need the demographic information for the various zip codes, which we gather from the state of Colorado (9). Again, we create a dataframe from the downloaded .csv file.

	zip	pop	hispanic	white_nh	ntvam_nh	black_nh	asian_nh	hawai_nh	other_nh	twoplus_nh	male	female	ageless5
0	80002	18439	4742	12900	28	164	411	0	29	165	9121	9318	887
1	80003	36518	9024	25104	382	534	888	5	12	569	17695	18824	2088
2	80005	37070	4641	30749	172	435	414	2	6	651	17623	19447	2212
3	80005	27662	2486	23740	21	139	779	0	48	449	13795	13867	1671
4	80007	7537	368	6793	0	17	187	0	5	167	3674	3863	341
5	80010	43987	23767	8747	274	6734	3254	0	120	1091	22840	21147	4294
6	80011	48360	23395	13804	531	7928	1492	48	68	1094	23385	24975	4384
7	80012	49010	15408	16620	123	12016	2571	275	64	1933	23978	25032	3942
8	80013	72348	14534	40415	117	9164	4313	80	469	3256	35726	36622	5356
9	80014	37617	5584	23781	75	3916	2726	67	110	1358	17702	19915	1683

Figure 2: The first 10 rows and 13 columns the Colorado demographic information dataframe.

Then we merge the two dataframes based on the zip code.

	zip	primary_city	state	county	latitude	longitude	pop	hispanic	white_nh	ntvam_nh	black_nh	asian_nh	l
0	80002	Arvada	CO	Jefferson County	39.79	-105.11	18439	4742	12900	28	164	411	
1	80003	Arvada	CO	Jefferson County	39.83	-105.06	36518	9024	25104	382	534	888	
2	80004	Arvada	CO	Jefferson County	39.82	-105.10	37070	4641	30749	172	435	414	
3	80005	Arvada	CO	Jefferson County	39.85	-105.13	27662	2486	23740	21	139	779	
4	80007	Arvada	CO	Jefferson County	39.86	-105.19	7537	368	6793	0	17	187	
5	80010	Aurora	CO	Arapahoe County	39.74	-104.86	43987	23767	8747	274	6734	3254	
6	80011	Aurora	CO	Arapahoe County	39.74	-104.78	48360	23395	13804	531	7928	1492	
7	80012	Aurora	CO	Arapahoe County	39.69	-104.81	49010	15408	16620	123	12016	2571	
8	80013	Aurora	CO	Arapahoe County	39.66	-104.77	72348	14534	40415	117	9164	4313	
9	80014	Aurora	CO	Arapahoe County	39.66	-104.84	37617	5584	23781	75	3916	2726	

Figure 3: The first 10 rows and 12 columns the merged dataframe containing both geographic and demographic information for Colorado.

Our first attempt at defining communities starts with the foursquare API (10). We procure the most common venue categories in each neighborhood using the explore function.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	80002 Mexican Restaurant	Mobile Phone Shop	Sandwich Place	Coffee Shop	Chinese Restaurant	Marijuana Dispensary	Salon / BarberShop	Big Box Store	Sporting Goods Shop
1	80003 Convenience Store	Pub	Park	Baseball Field	Fabric Shop	Food Truck	Food Service	Food Court	Food
2	80004 Convenience Store	Nature Preserve	Bus Stop	Women's Store	Food Truck	Food Service	Food Court	Food	Flower Shop
3	80005 Playground	Arcade	Basketball Court	Business Service	Lake	Women's Store	Fabric Shop	Food Service	Food
4	80007 Outdoor Sculpture	Women's Store	Eye Doctor	Food Truck	Food Service	Food Court	Food	Flower Shop	Flea Market

Figure 4: The first 5 rows of the most common venue categories by Colorado zip code.

Then we use the k-means clustering algorithm on the venue data to group the zip codes into clusters. Using the Folium library, we can color-code the clusters and display them visually.

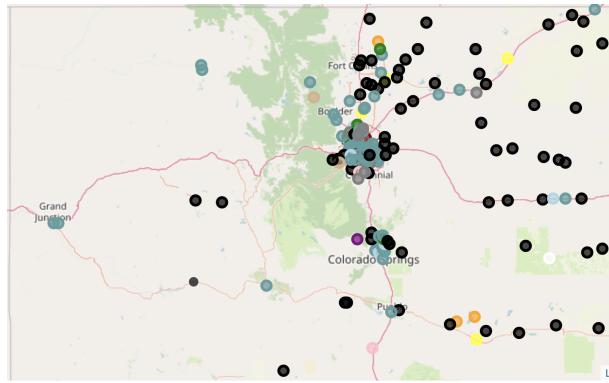


Figure 5: Color-coded clusters of Colorado based on the common venues and k-means clustering.

In Figure 5, notice the yellow dots. While they are a cluster based on how rural they are, they are not near each other geographically. So this definition for communities will not work for our purposes. Thus we turn to k-means clustering based on the geographic and demographic information. Then we move the clustering using all of the numerical data from the dataframe and get Figure 6.

The clustering is better for our purposes, but it is pretty wide spread vertically also not meeting our needs for communities. We give k-means clustering one more attempt using only the latitude and longitude and get Figure 7.

The clustering in Figure 7 has the same issues Figure 6 (wide spread vertically). It seems that k-means clustering is not the tool to use for this situation, which may be because it does not perform well using Euclidean distances, which is what we have here (11). So we turn to using density-based spatial clustering of applications with noise (DBSCAN), which will hopefully deal better with our situation (12).

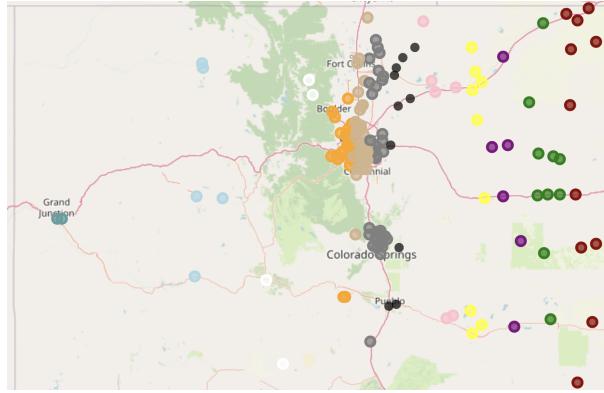


Figure 6: Color-coded clusters of Colorado based k-means clustering using all of our numerical information for the zip codes.

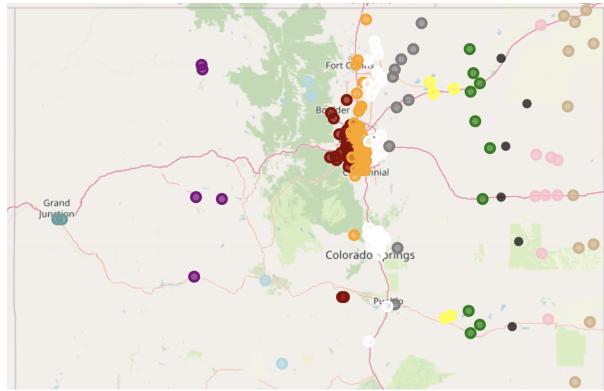


Figure 7: Color-coded clusters of Colorado based k-means clustering using the latitude and longitude for the zip codes.

Notice in Figure 8 that we have outliers, which are colored black. That does make sense with our definition of community, since there are many zip codes in Colorado that are small in population do not cluster with other zip codes. However now all of Denver Metro Area is one community, but that doesn't make sense either, having done business in the community for a decade. It can take hours to drive from one side of the Denver Metro Area to the other, so we need it to be more fine grained than that.

Now we add zip codes to the input data. “ZIP Codes are numbered with the first digit representing a certain group of U.S. states, the second and third digits together representing a region in that group (or perhaps a large city) and the fourth and fifth digits representing a group of delivery addresses within that region” (13). Thus this numerical data may help us to create communities.

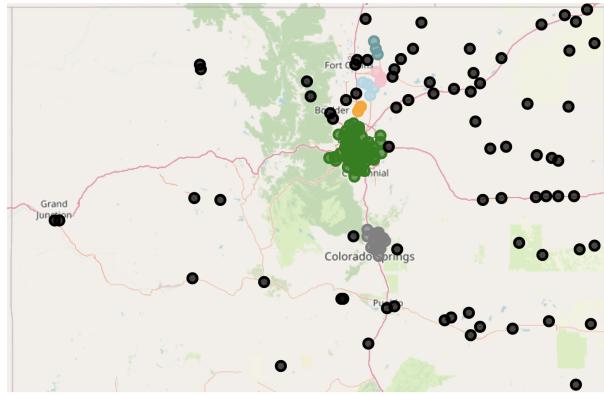


Figure 8: Color-coded clusters of Colorado using DBSCAN with the latitude and longitude for the zip codes.

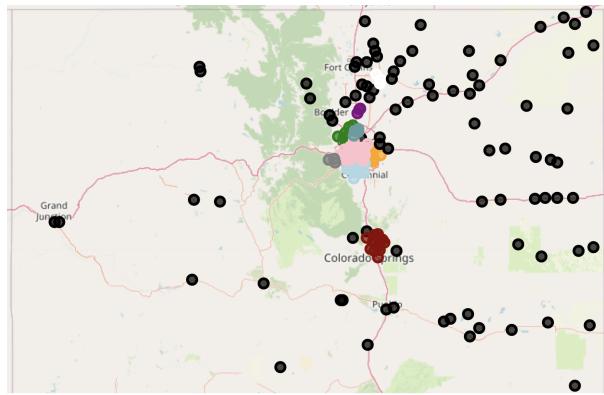


Figure 9: Color-coded clusters of Colorado using DBSCAN with the latitude, longitude and zip codes.

Figure 9 also has outliers, but now we can see that the Metro area seems to be divided up geographically. This affirms our understanding of the area, so we will go with this definition of community.

Now that the communities are defined, we will look at the details of the communities. We need to know how many medical centers each community has. In order to do that, we first need a list of the medical centers in Colorado, which we will get using the foursquare API again (10). Notice in the dataframe below that all types of medical centers are listed, and we will treat them equally. That is based on our assumption that in a pandemic, if a facility has medical supplies, then it can be repurposed to help in the treatment or detection of the disease.

Now we need to determine how these medical centers are distributed amongst

	0	1	2	3	4	5
0	80002	39.79	-105.11	Kid Focus Dentistry	39.789936	-105.111984
1	80002	39.79	-105.11	Home Care Assistance Denver	39.786776	-105.108766
2	80002	39.79	-105.11	Thrive Health Systems	39.786102	-105.110630
3	80002	39.79	-105.11	Kingdom Family Chiropractic	39.786669	-105.111077
4	80002	39.79	-105.11	Just Wisdom teeth	39.786427	-105.111109

Figure 10: The first five rows of the dataframe created for medical centers of Colorado.

our defined communities. To do this, we use the nearest neighbor algorithm (14). We take the zip codes’ information and train the nearest neighbor algorithm to predict the community assigned. Then we take the medical centers information and use the algorithm to assign each to a community. Figure 11 is a visualization of the zip codes and the medical center community assignments (again color-coded). To differentiate them, the medical centers are larger disks with no fill.

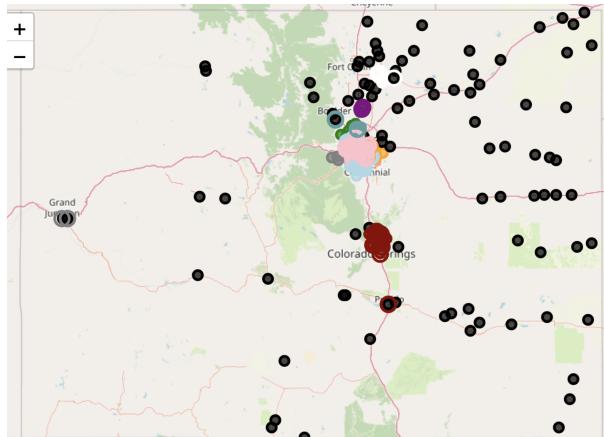


Figure 11: Color-coded zip codes and medical centers. The medical centers are larger disks with no fill

Figure 12 is the same map zoomed in on the Denver Metro Area, for clarity.

4 Results

Now we examine the communities and medical centers based on the metrics described in the literature review. We start with total population of the communities, which we have renamed based on their location and near-by towns and is Figure 13.

While Central Denver Metro area has the largest population and Evergreen the least, we want to know how that compares to the number of medical centers in each area. Communities with a large population may also have many medical

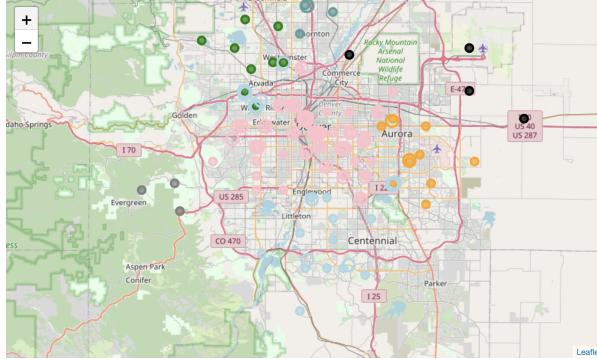


Figure 12: Color-coded zip codes and medical centers. The medical centers are larger disks with no fill.

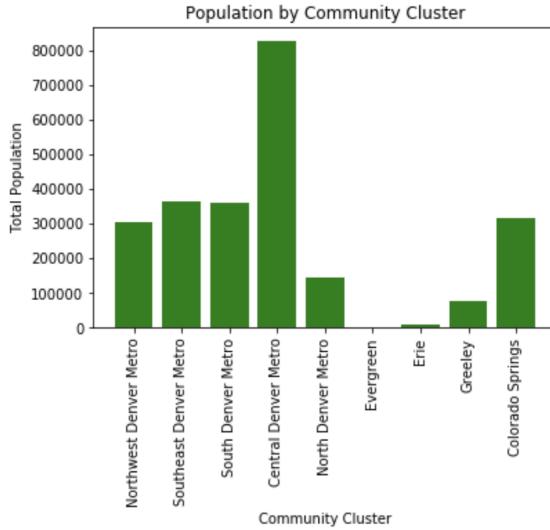


Figure 13: A bar chart of the total population of the community clusters.

centers, and therefore not need access to a new center. So we count how many medical centers are in each community and examine the percent of medical centers per population in a bar chart, which is Figure 14.

We can see from Figure 14 above that Central Denver Metro area has the fourth highest percentage of medical centers. So while the population is high, it may not be the most advantageous place to put a new center. Also note that Northwest Denver Metro area seems to have no medical centers and the Southeast Denver Metro, North Denver Metro, and Colorado Springs areas have very few. Now we examine the percent of folks of color for each community.

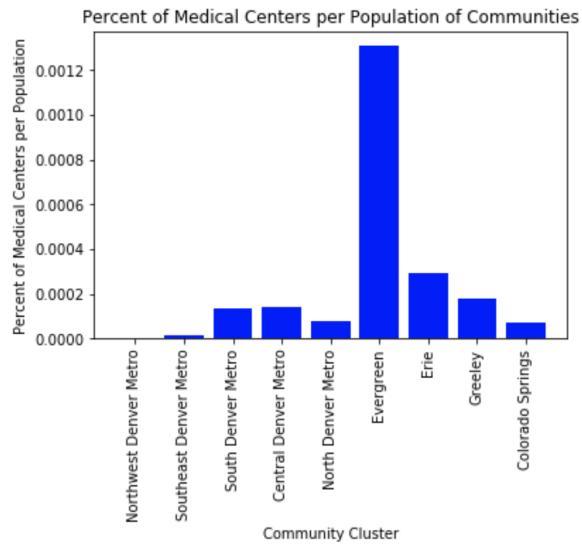


Figure 14: A bar chart of the percent of medical centers per total population of the community clusters.

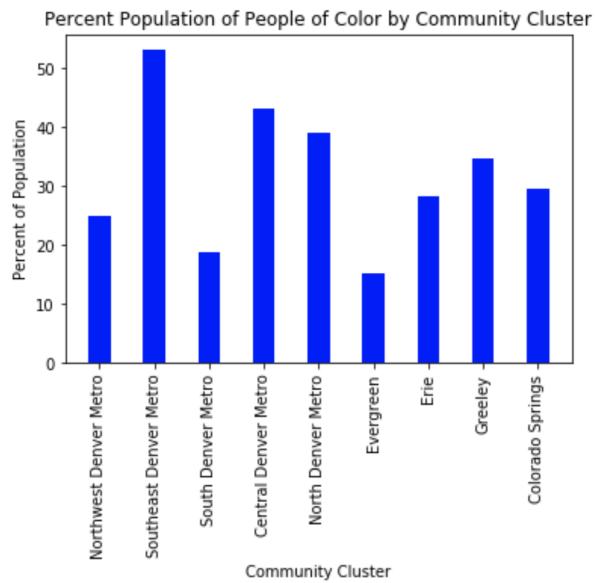


Figure 15: A bar chart of the percent of people of color per total population of the community clusters.

Examine Figure 15. While we would have predicted that the Central Denver Metro area has the highest percentage of folks of color, and they do have the second highest, the Southeast Denver Metro area has the highest with over 50%. The other two communities in the top of of percentage of people of color are the North Denver Metro and Greeley areas. Now we investigate the percent of folks 50 and over in our communities which is in Figure 16.

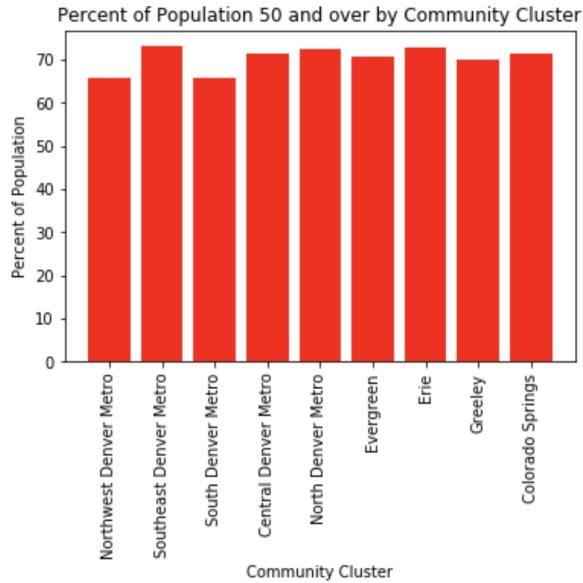


Figure 16: A bar chart of the percent of people of age 50 and older per total population of the community clusters.

The ages represented in Figure 16 are all quite similar, with the Southeast Denver Metro area with a slight lead. Lastly, we investigate the gender breakdown in Figure 17.

The gender breakdown indicates only small fluctuations in percent of men, and the differences are not big enough to use as a metric for new medical center location.

5 Conclusion

Each metric, percent of medical centers, percent of folks of color, percent of people or 50 and the gender breakdown, points to a variety of communities that may benefit from a new medical center. However, they all point to the Southeast Denver Metro area as the best location for the new medical center. This area has the second lowest percentage of medical centers per population, indicating a real need for access to medicine. This area also had the highest percentage

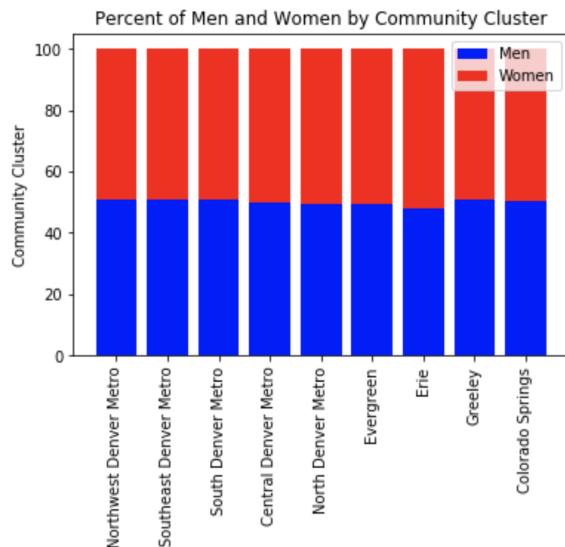


Figure 17: A bar chart of the gender breakdown of the community clusters.

of folks of color and folks 50 and over, indicating that this virus could cause much damage in this area. Lastly, this is the second most populated area within our communities. So our professional recommendation is to place the new medical center in the Southeast Denver Metro area during this pandemic.

6 References

References

- [1] <https://kdvr.com/news/coronavirus/will-colorado-have-enough-hospital-beds-to-accommodate-coronavirus-patients/>
- [2] <https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/covid19-racial-disparities>
- [3] <https://www.theguardian.com/commentisfree/2020/apr/07/coronavirus-hits-men-harder-evidence-risk>
- [4] <https://www.healthsystemtracker.org/brief/urban-and-rural-differences-in-coronavirus-pandemic-preparedness/>
- [5] <https://en.wikipedia.org/wiki/Community>
- [6] [https://www.colorado-demographics.com/zip_codes_by_population#:~:text=Colorado%20Zip%20Codes%20by%20Population,\(there%20are%20some%20ties\).](https://www.colorado-demographics.com/zip_codes_by_population#:~:text=Colorado%20Zip%20Codes%20by%20Population,(there%20are%20some%20ties).)
- [7] https://www.colorado-demographics.com/cities_by_population
- [8] <https://www.unitedstateszipcodes.org/zip-code-database/>
- [9] [https://data.colorado.gov/Demographics/Census-Zip-Codes-in-Colrado-2015/jd2b-7gyb/about](https://data.colorado.gov/Demographics/Census-Zip-Codes-in-Colorado-2015/jd2b-7gyb/about)
- [10] <https://foursquare.com/>
- [11] https://en.wikipedia.org/wiki/K-means_clustering
- [12] <https://en.wikipedia.org/wiki/DBSCAN>
- [13] https://en.wikipedia.org/wiki/ZIP_Code
- [14] [https://en.wikipedia.org/wiki/Nearest_neighbor_search#:~:text=Nearest%20neighbor%20search%20\(NNS\)%2C,the%20larger%20the%20function%20values.](https://en.wikipedia.org/wiki/Nearest_neighbor_search#:~:text=Nearest%20neighbor%20search%20(NNS)%2C,the%20larger%20the%20function%20values.)