

Good afternoon!

I hope your week is going well

I have just finished up my analysis on the client data, below I have outlined the steps I took as well as my outputs and findings

First off, I found that roster 2 was the only roster that formatted its dates using MM/DD/YYYY, so I reformatted the date of birth column in that roster to be YYYY-MM-DD to match with the rest of the rosters.

I was able to create a new table that found all members who were eligible in april 2022 and I named this table std_member_info. Using the std_member_info table I used sql queries to uncover the answers to the questions laid out- my outputs are below:

- How many distinct members are eligible in April 2022? **71,115 members**
- How many members were included more than once? **13,775 members**
- What is the breakdown of members by payer? **Madv: 31878 members Mdccl: 53054 members**
- How many members live in a zip code with a food_access_score lower than 2? **6608 members**
- What is the average social isolation score for the members? **3.064**
- Which members live in the zip code with the highest algorex_sdoh_composite_score? Here is a list of the member ID's [('15404143'), ('15418054'), ('15537550'), ('15391076'), ('15363143'), ('15375855'), ('15456560'), ('15462579'), ('15340102'), ('15346807'), ('15389381'), ('15414827'), ('15482255'), ('15484216'), ('15486908'), ('15489640'), ('15514403'), ('15532243'), ('15539805'), ('15359217'), ('15447186'), ('15487938'), ('15493624'), ('15497844'), ('15506131'), ('15387928'), ('15389759'), ('15403845'), ('15417526'), ('15459547'), ('15495072'), ('15499581'), ('15525968'), ('15363143'), ('15375855'), ('15456560'), ('15340102'), ('15346807'), ('15389381'), ('15482255'), ('15484216'), ('15486908'), ('15489640')]

For my analysis, I decided to do a cluster analysis which takes a large data set and finds clusters of similar data points. I used a method where I could pre-set the number of clusters I wanted to find and I chose to run the analysis on 3 clusters. Each of the three clusters contains members whose zip codes have similar scores (the scores were taken from the model_scores_by_zip table). Here is the breakdown of the average scores for each cluster as well as the percent of members in each cluster

Cluster 1 - 47 % of members

neighborhood_stress_score **0.265307**

social_isolation_score **3.330170**

transportation_access_score	4.309297
food_access_score	3.525602
unstable_housing_score	3.022741
state_govt_assistance	0.768979
homeless_indicator	7.058801

Cluster 2 - 47% of members

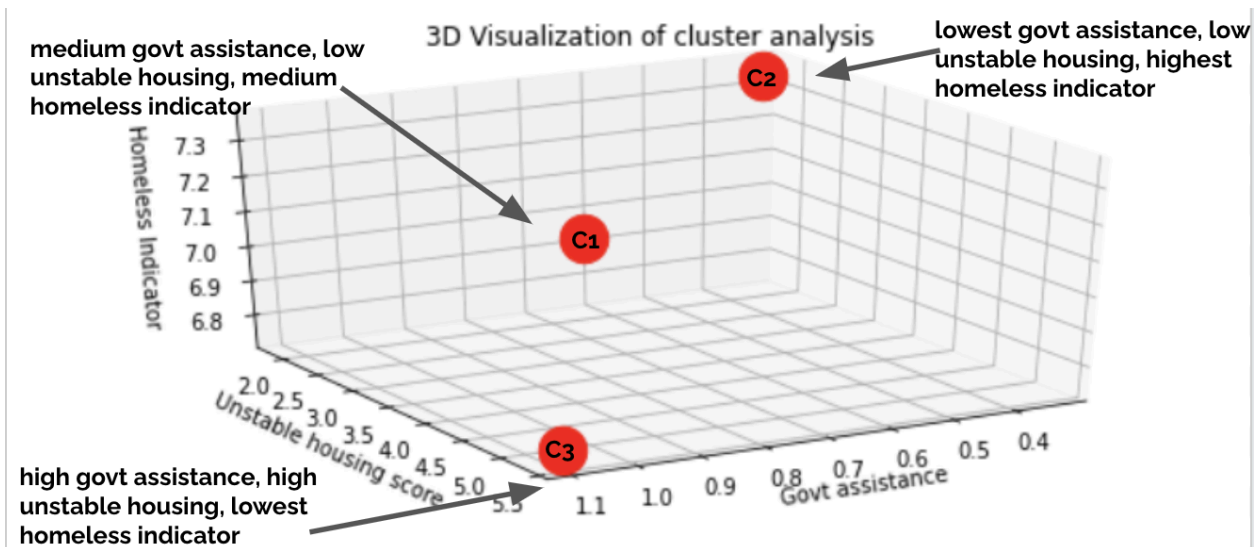
neighborhood_stress_score	-0.491657
social_isolation_score	2.630329
transportation_access_score	4.143515
food_access_score	2.784939
unstable_housing_score	1.875563
state_govt_assistance	0.338415
homeless_indicator	7.339112

Cluster 3 - 6% of members

neighborhood_stress_score	0.702471
social_isolation_score	4.333782
transportation_access_score	4.400072
food_access_score	3.285652
unstable_housing_score	5.344900
state_govt_assistance	1.084788
homeless_indicator	6.745485

The three scores I have highlighted in yellow are the ones I decided to visualize using a three dimensional graph.

From the graph below we can see that the clusters each take up a unique space in this three dimensional graph where government assistance is on the x axis, unstable housing is on the y axis, and homeless indicator is on the z axis- each point is marked with what cluster it represents.



Visualizing these three data points shows us that neighborhoods with low government assistance and low unstable housing are struggling with homelessness- this was a surprising outcome as one would think that areas with low government assistance and low unstable housing would not have high homelessness, but we can see from the visualization that the lowest homeless indicator actually exists in the cluster in the bottom left with the highest unstable housing and highest government assistance

A more important takeaway from this has to do with cluster three, we can assume that the government assistance is probably most needed in areas with more unstable housing like cluster 3 hence why cluster 3 has the highest assistance by far, and the assistance given was able to provide these residents more stable housing, which in turn has reduced homelessness in those neighborhoods- which we can hypothesize from the fact that cluster three has the lowest homeless indicator. In other words, the government assistance seems to be working, based on my analysis!

Please let me know if you have any questions or need clarification,

Looking forward to hearing what you think!

Best,

Divya