FINAL REPORT

>Introduction

-In this final report i will be using the district of chicago property tax and expenses

-Goal of this study is to answer: what is the correlation between the property tax and the expenses?

>Methods(steps)

1)Make a table for each district and sort them by tif number so that they are in order

2)Extract the property tax and the expenses from the table to a object and concat them together

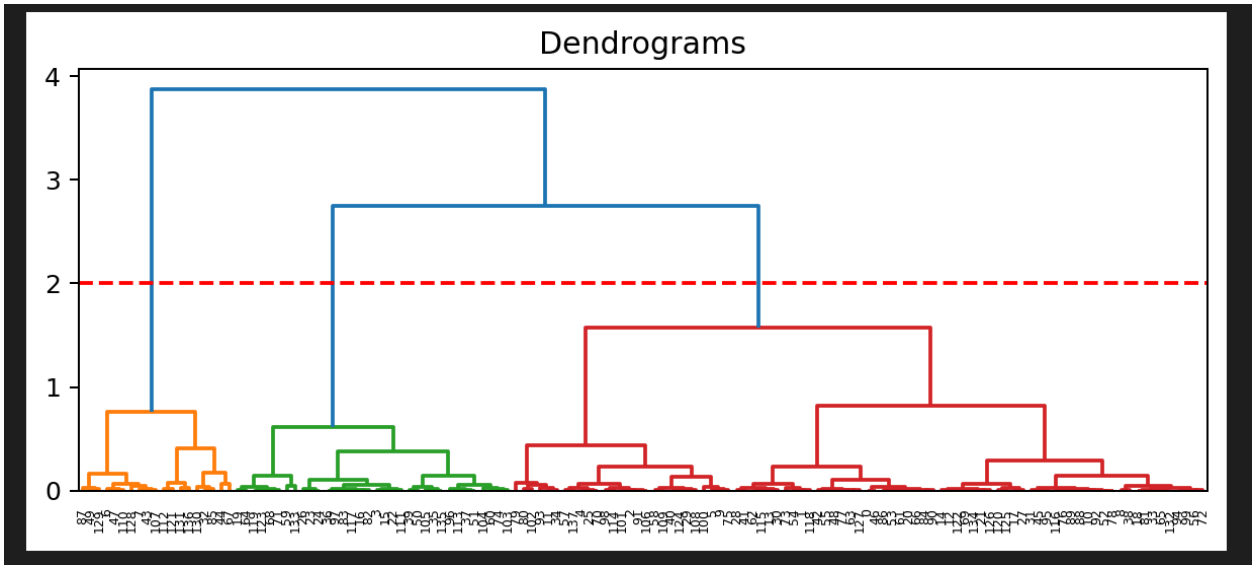
3)Create a filter system where each district takes property tax and expenses with a condition that the district has to have 8 or more observance variables.

4)Now that the filter system has finished, the functional table will contain a lot of NAN values, we will then replace those NAN with 0,run the downcasting method to convert the table back into integer type, then we will replace the zero with the mean of its correspond table.

5)calculate the mean of each district that on the table and store the result in a new table

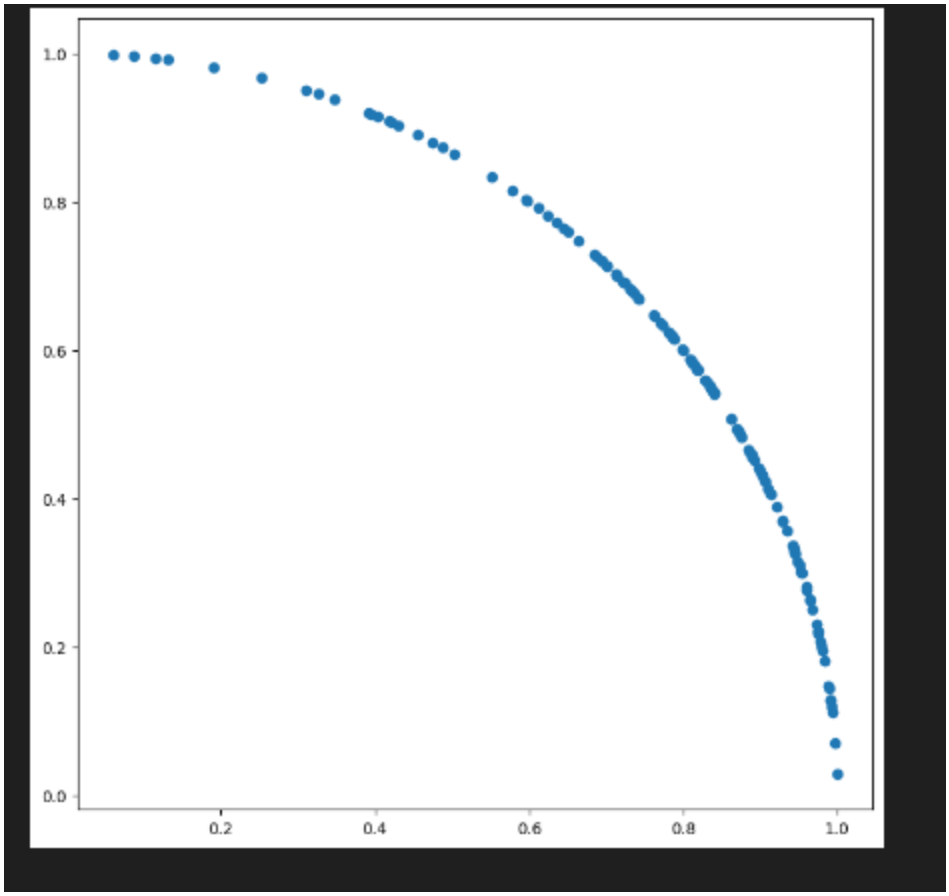
6)normalize the values of the table and graph the dendrogram using the normalized data and pick some cluster and i pick 3

Dendrograms:



7) create another graph using the same table but now with the number of cluster i pick

Graph of cluster:



8)I create and run a filter system that return me the two mean that are closer together in the main table that contain the mean of each district(note* in order fo this to work i make the mean of a each district a X and Y table, where X is contain the district property tax and Y contain the expenses, depend on what i looking for i can drop property tax or expenses with the original table still intact because those are not needed.)

The two district are:

	Little Village Industrial Corridor expenses
2	208211
12	200230
8	169056
10	153168
11	149886
5	149259
9	135582
3	117483
1	116902
13	102873
6	13030
0	11865
4	9543
7	6007

	Armitage/Pulaski expenses
11	306834
8	279652
6	209997
1	160705
13	107182
5	103104
3	91625
4	84383
0	70272
7	62747
9	59280
10	28633
2	24987
12	18341

9)Now that I have the two districts that are closer to each other by comparing their distance from each other, I can now do the two sample T-tests which give me the result of p-value greater than p=0.05 hypothesis.

10)Conclusion if the p-value is greater than p=0.05 then the t-test fails to reject the hypothesis: that means two mean are relatively similar to each other.

>Finding

-Base on the report the graph does show there are some correlation between the property tax and expenses,and these are the conclusion: if the property tax going up then the expenses will going up as well, however also base on the data there are some outlier
As well where this correlation of property tax and expense doesn't apply but because of how small that outlier is, when we can conclude that the outlier data is insignificant and also because we lack necessary information to determine its clause.

-Base on the report the two districts closest means, the two districts are statistically the same based on the T-test case, however their individual districts spending is different from each other therefore we can conclude that they are in fact not the same.

-The reasoning for rejecting null values is because they lack the necessary information to conduct this test, and the requirement of this test is each district has to have 8 or more observances and those Null values are not an observance variable.

-Alternative hypothesis-

The more money the district has spent on self-improvement the more money it will need to survive therefore it will increase the property tax cost of living there. However, if the district expenses go down because of the lack of funds, the property tax still stays the same because that is now the new base line of taxing for that district. (In general tax cannot go down)

Expenses:UP
Property tax:UP
—----------Alternative—----------
Expenses:Down
Property tax:stay the same