IBM Capstone Final

PREPARED BY: STEPHAN G MAHER

Question:

CAN WE USE ECONOMIC INDICATORS, POPULATION DENSITY, AND INFORMATION ON COMPETITORS FOR EACH ZIP-CODE, TO FIND AN OPTIMAL LOCATION FOR OUR FRANCHISE?

1. Problem Background

- My children work for a local delicatessen. Besides providing locally sourced cold- cuts, the deli also makes sandwiches and hoagies and either a soup or hot sandwich selection ranging daily from hot roast beef, pork, or meatballs. If I were to secure franchise rights, one of the first problems would be where would I locate the franchise? Before creating a financial profile, I would like to get the economic outlook and demographic data to see if I can optimize a location. The original store is in the Roxborough section of Philadelphia, specifically the “19128” zip code. I am also located in the same zip code.

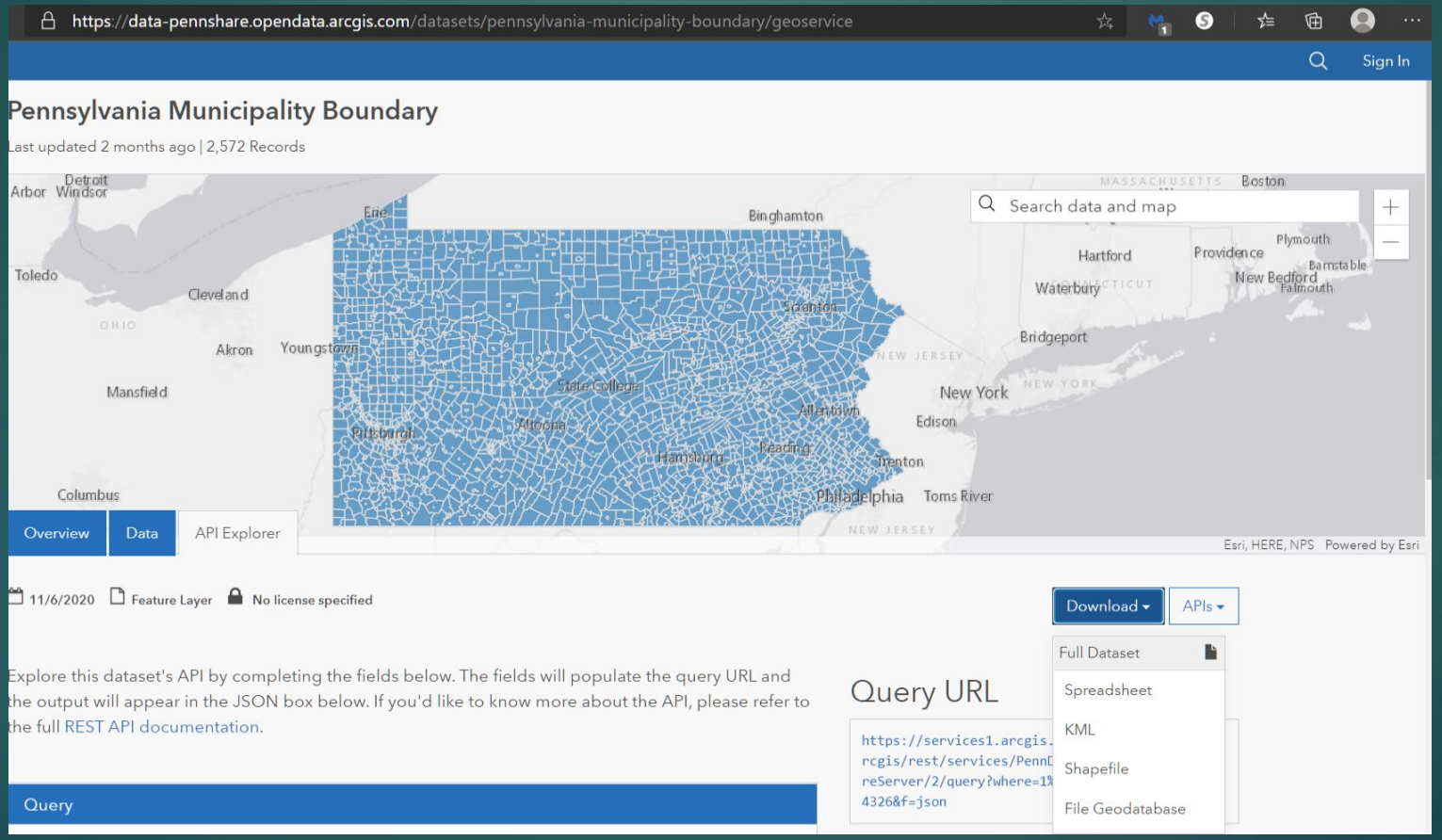
2 Uses for data

- Location Location Location
- This type of research is invaluable to
- Homebuyers and businesses alike.
- For homebuyers, you could add data like crime stats
- Businesses may be interested in Tax deals and other area incentives to relocate
- Police can use this data to adequately reposition staff
- Amazon used it to find a second headquarters
- Municipalities can adequately review construction and street closures, traffic patterns and a host of other functions

3. Description of Data

- This project consists of the following information
  - Zip code information for Pennsylvania including center Longitude and Latitude, zip code shapes and population size and density.
  - Financial information by zip code including:
    - Average Salary
    - Median Housing Costs
    - Avg Income for business by zip code
    - Number of returns filed by individual and business by zip code
    - Counts of small food service establishment by zip code

Shape file for Pennsylvania

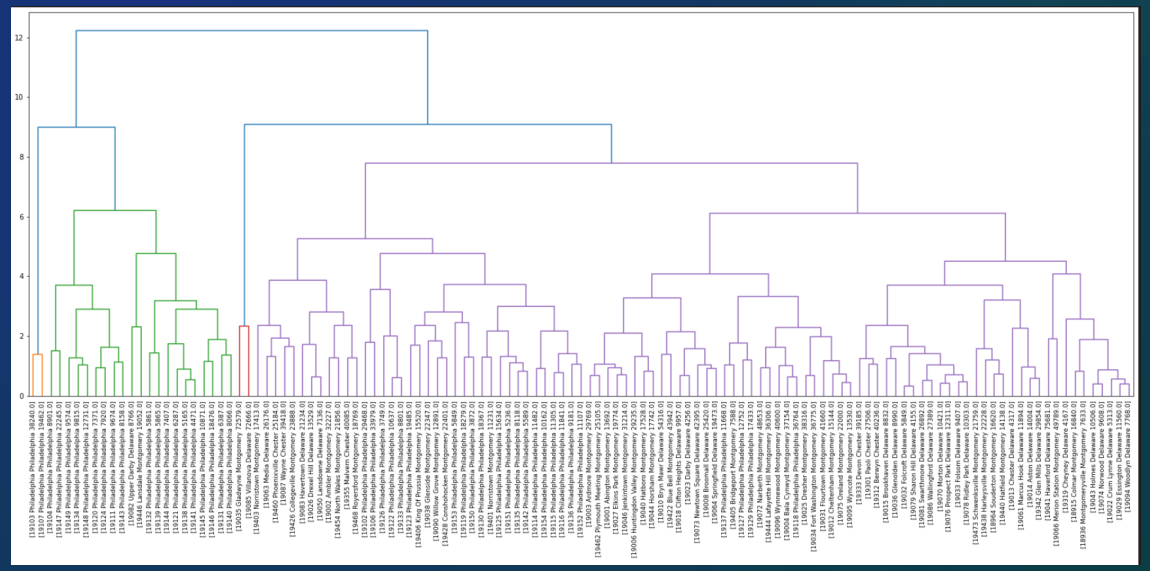


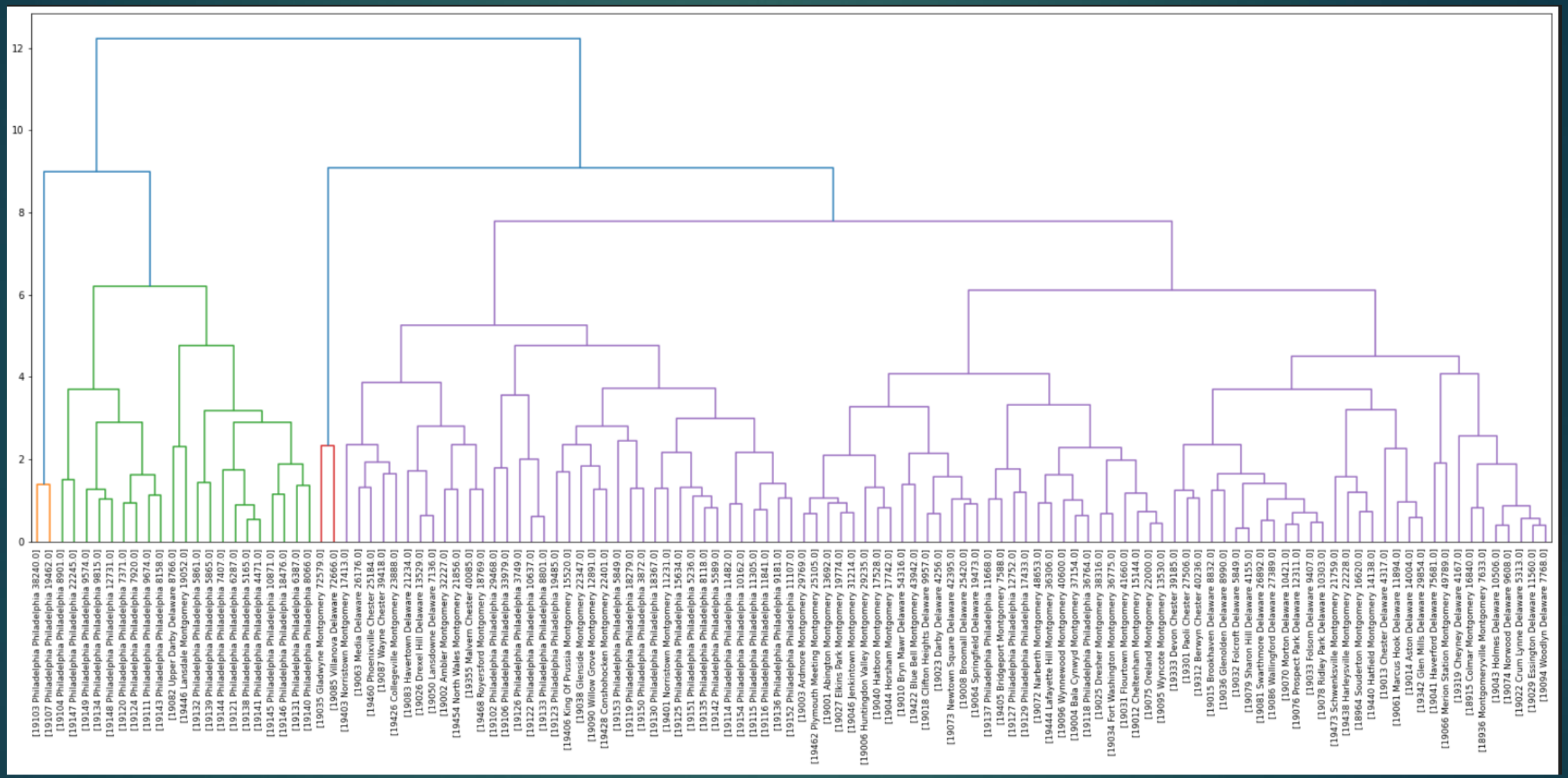
4. Methodology

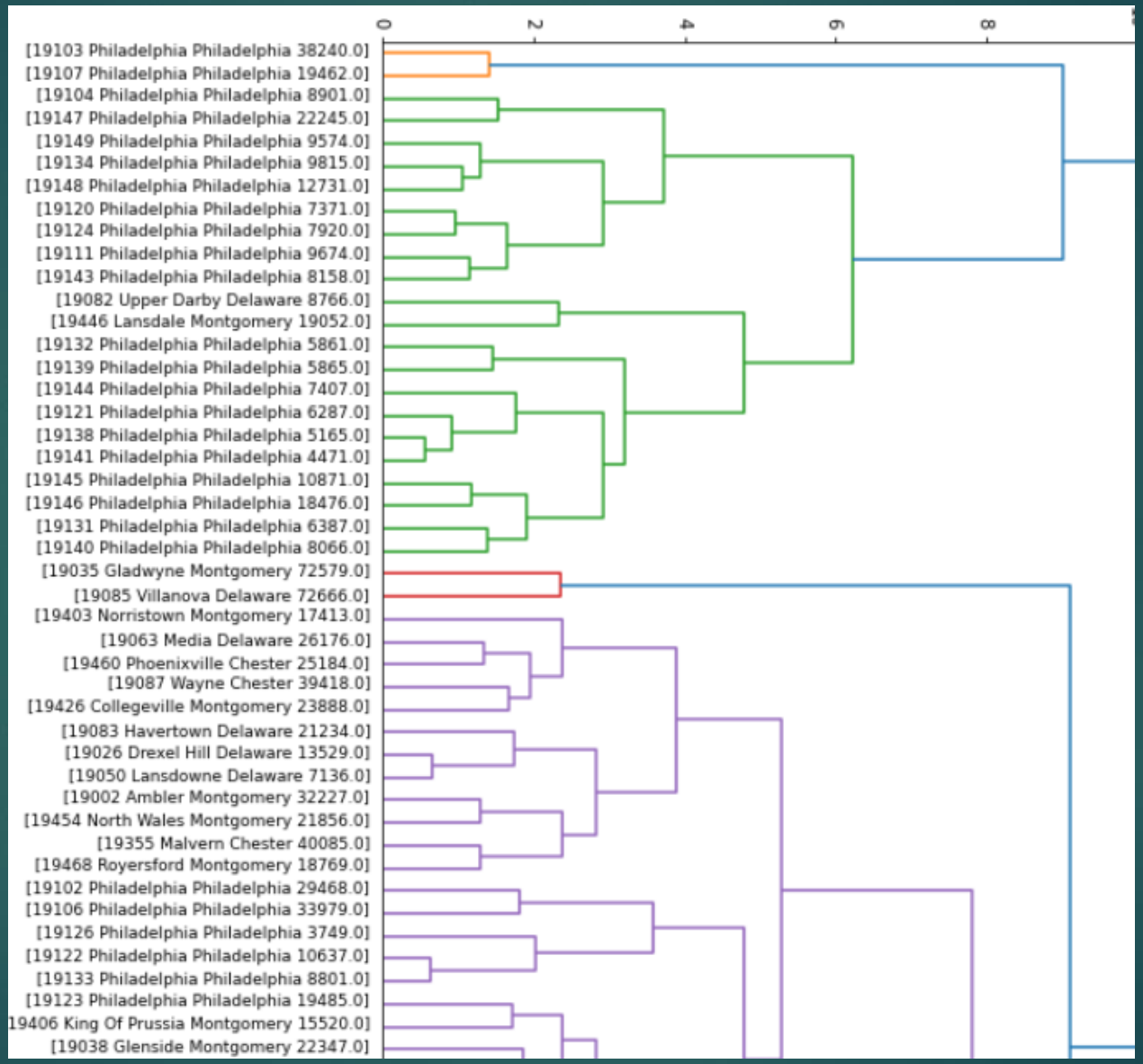
- 1. Agglomerative Hierarchical Clustering
- 2. K-means
- 3. Visualization

Hierarchical Clustering

- We can see from this diagram that there are 2 clusters in red and orange that concentrate the wealth in the city. These are Center city and the areas surrounding Bryn Mahr.

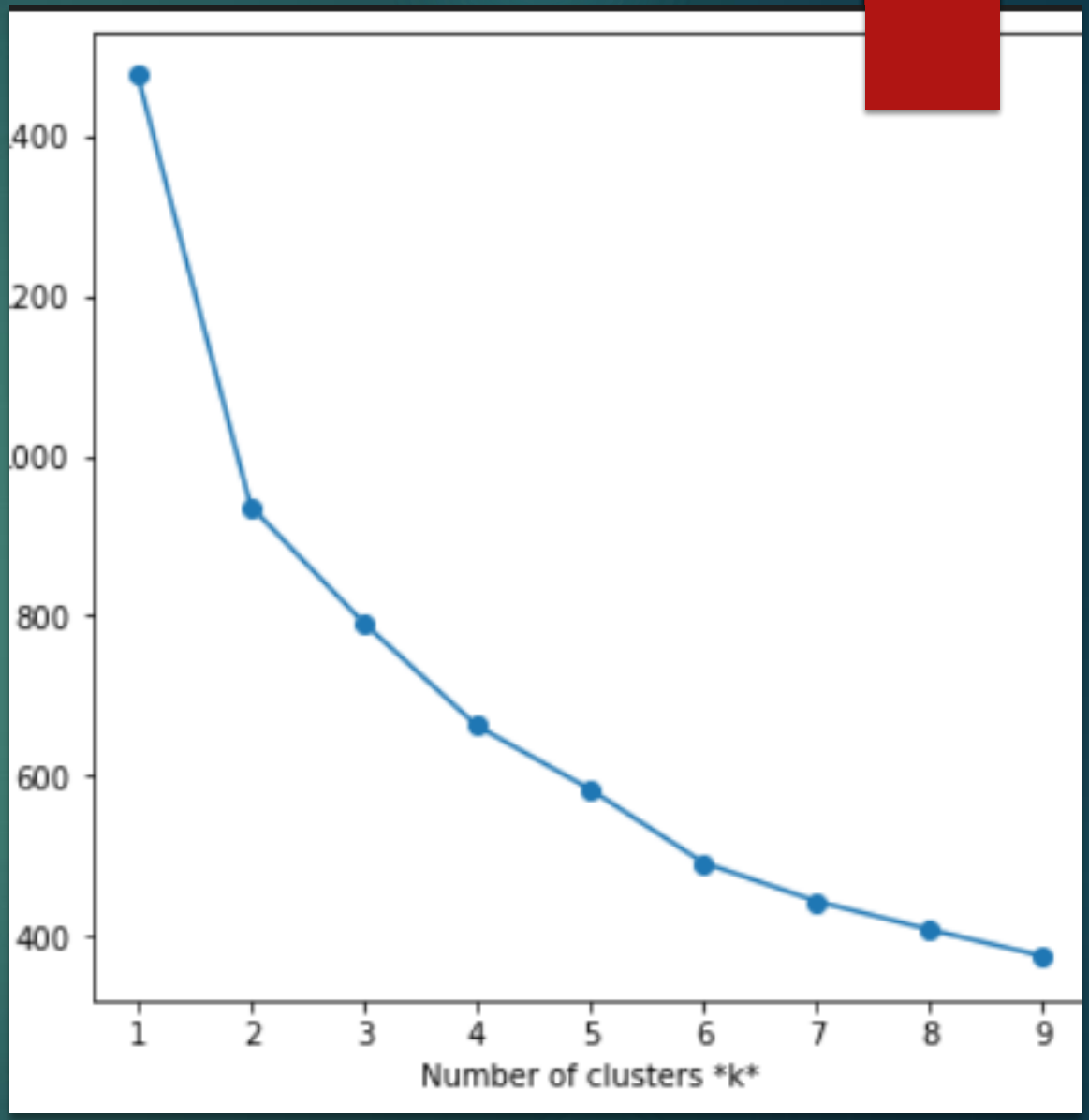



12
10
8
6
4
2
0
[19103 Philadelphia Philadelphia 38240.0]
[19107 Philadelphia Philadelphia 19462.0]
[19104 Philadelphia Philadelphia 8901.0]
[19147 Philadelphia Philadelphia 22245.0]
[19149 Philadelphia Philadelphia 9574.0]
[19134 Philadelphia Philadelphia 9815.0]
[19148 Philadelphia Philadelphia 12731.0]
[19120 Philadelphia Philadelphia 7371.0]
[19124 Philadelphia Philadelphia 7920.0]
[19111 Philadelphia Philadelphia 9674.0]
[19143 Philadelphia Philadelphia 8158.0]
[19082 Upper Darby Delaware 8766.0]
[19446 Lansdale Montgomery 19052.0]
[19132 Philadelphia Philadelphia 5861.0]
[19139 Philadelphia Philadelphia 5865.0]
[19144 Philadelphia Philadelphia 7407.0]
[19121 Philadelphia Philadelphia 6287.0]
[19138 Philadelphia Philadelphia 5165.0]
[19141 Philadelphia Philadelphia 4471.0]
[19145 Philadelphia Philadelphia 10871.0]
[19146 Philadelphia Philadelphia 18476.0]
[19131 Philadelphia Philadelphia 6387.0]
[19140 Philadelphia Philadelphia 8066.0]
[19035 Gladwyne Montgomery 72579.0]
[19085 Villanova Delaware 72666.0]
[19403 Norristown Montgomery 17413.0]
[19063 Media Delaware 26176.0]
[19460 Phoenixville Chester 25184.0]
[19087 Wayne Chester 39418.0]
[19426 Collegeville Montgomery 23888.0]
[19083 Havertown Delaware 21234.0]
[19026 Drexel Hill Delaware 13529.0]
[19050 Lansdowne Delaware 7136.0]
[19002 Ambler Montgomery 32227.0]
[19454 North Wales Montgomery 21856.0]
[19355 Malvern Chester 40085.0]
[19468 Royersford Montgomery 18769.0]
[19102 Philadelphia Philadelphia 29468.0]
[19106 Philadelphia Philadelphia 33979.0]
[19126 Philadelphia Philadelphia 3749.0]
[19122 Philadelphia Philadelphia 10637.0]
[19133 Philadelphia Philadelphia 8801.0]
[19123 Philadelphia Philadelphia 19485.0]
[19406 King Of Prussia Montgomery 15520.0]
[19038 Glenside Montgomery 22347.0]
[19090 Willow Grove Montgomery 12891.0]
[19428 Conshohocken Montgomery 22401.0]
[19153 Philadelphia Philadelphia 5849.0]
[19119 Philadelphia Philadelphia 18279.0]
[19150 Philadelphia Philadelphia 3872.0]
[19130 Philadelphia Philadelphia 18367.0]
[19401 Norristown Montgomery 11231.0]
[19125 Philadelphia Philadelphia 15634.0]
[19151 Philadelphia Philadelphia 5236.0]
[19135 Philadelphia Philadelphia 8118.0]
[19142 Philadelphia Philadelphia 5589.0]
[19114 Philadelphia Philadelphia 11482.0]
[19154 Philadelphia Philadelphia 10162.0]
[19115 Philadelphia Philadelphia 11305.0]
[19116 Philadelphia Philadelphia 11841.0]
[19136 Philadelphia Philadelphia 9181.0]
[19152 Philadelphia Philadelphia 11107.0]
[19003 Ardmore Montgomery 29769.0]
[19462 Plymouth Meeting Montgomery 25105.0]
[19001 Abington Montgomery 13692.0]
[19027 Elkins Park Montgomery 19774.0]
[19046 Jenkintown Montgomery 31214.0]
[19006 Huntingdon Valley Montgomery 29235.0]
[19040 Hatboro Montgomery 17528.0]
[19044 Horsham Montgomery 17742.0]
[19010 Bryn Mawr Delaware 54316.0]
[19422 Blue Bell Montgomery 43942.0]
[19018 Clifton Heights Delaware 9957.0]
[19023 Darby Delaware 4256.0]
[19073 Newtown Square Delaware 42395.0]
[19008 Broomall Delaware 25420.0]
[19064 Springfield Delaware 19473.0]
[19137 Philadelphia Philadelphia 11668.0]
[19405 Bridgeport Montgomery 7588.0]
[19127 Philadelphia Philadelphia 12752.0]
[19129 Philadelphia Philadelphia 17433.0]
[19072 Narberth Montgomery 48653.0]
[19444 Lafayette Hill Montgomery 36306.0]
[19096 Wynnewood Montgomery 40600.0]
[19004 Bala Cynwyd Montgomery 37154.0]
[19118 Philadelphia Philadelphia 36764.0]
[19025 Dresher Montgomery 38316.0]
[19034 Fort Washington Montgomery 36775.0]
[19031 Flourtown Montgomery 41660.0]
[19012 Cheltenham Montgomery 15144.0]
[19075 Oreland Montgomery 22000.0]
[19095 Wyncote Montgomery 13530.0]
[19333 Devon Chester 39185.0]
[19301 Paoli Chester 27506.0]
[19312 Berwyn Chester 40236.0]
[19015 Brookhaven Delaware 8832.0]
[19036 Glenolden Delaware 8990.0]
[19032 Folcroft Delaware 5849.0]
[19079 Sharon Hill Delaware 4155.0]
[19081 Swarthmore Delaware 26892.0]
[19086 Wallingford Delaware 27389.0]
[19070 Morton Delaware 10421.0]
[19076 Prospect Park Delaware 12311.0]
[19033 Folsom Delaware 9407.0]
[19078 Ridley Park Delaware 10303.0]
[19473 Schwenksville Montgomery 21759.0]
[19438 Harleysville Montgomery 22228.0]
[18964 Souderton Montgomery 16620.0]
[19440 Hatfield Montgomery 14138.0]
[19013 Chester Delaware 4317.0]
[19061 Marcus Hook Delaware 11894.0]
[19014 Aston Delaware 14004.0]
[19342 Glen Mills Delaware 29854.0]
[19041 Haverford Delaware 75681.0]
[19066 Merion Station Montgomery 49789.0]
[19319 Cheyney Delaware 4167.0]
[18915 Colmar Montgomery 16840.0]
[18936 Montgomeryville Montgomery 7633.0]
[19043 Holmes Delaware 10506.0]
[19074 Norwood Delaware 9608.0]
[19022 Crum Lynne Delaware 5313.0]
[19029 Essington Delaware 11560.0]
[19094 Woodlyn Delaware 7768.0]


0
2
4
6
8
[19103 Philadelphia Philadelphia 38240.0]
[19107 Philadelphia Philadelphia 19462.0]
[19104 Philadelphia Philadelphia 8901.0]
[19147 Philadelphia Philadelphia 22245.0]
[19149 Philadelphia Philadelphia 9574.0]
[19134 Philadelphia Philadelphia 9815.0]
[19148 Philadelphia Philadelphia 12731.0]
[19120 Philadelphia Philadelphia 7371.0]
[19124 Philadelphia Philadelphia 7920.0]
[19111 Philadelphia Philadelphia 9674.0]
[19143 Philadelphia Philadelphia 8158.0]
[19082 Upper Darby Delaware 8766.0]
[19446 Lansdale Montgomery 19052.0]
[19132 Philadelphia Philadelphia 5861.0]
[19139 Philadelphia Philadelphia 5865.0]
[19144 Philadelphia Philadelphia 7407.0]
[19121 Philadelphia Philadelphia 6287.0]
[19138 Philadelphia Philadelphia 5165.0]
[19141 Philadelphia Philadelphia 4471.0]
[19145 Philadelphia Philadelphia 10871.0]
[19146 Philadelphia Philadelphia 18476.0]
[19131 Philadelphia Philadelphia 6387.0]
[19140 Philadelphia Philadelphia 8066.0]
[19035 Gladwyne Montgomery 72579.0]
[19085 Villanova Delaware 72666.0]
[19403 Norristown Montgomery 17413.0]
[19063 Media Delaware 26176.0]
[19460 Phoenixville Chester 25184.0]
[19087 Wayne Chester 39418.0]
[19426 Collegeville Montgomery 23888.0]
[19083 Havertown Delaware 21234.0]
[19026 Drexel Hill Delaware 13529.0]
[19050 Lansdowne Delaware 7136.0]
[19002 Ambler Montgomery 32227.0]
[19454 North Wales Montgomery 21856.0]
[19355 Malvern Chester 40085.0]
[19468 Royersford Montgomery 18769.0]
[19102 Philadelphia Philadelphia 29468.0]
[19106 Philadelphia Philadelphia 33979.0]
[19126 Philadelphia Philadelphia 3749.0]
[19122 Philadelphia Philadelphia 10637.0]
[19133 Philadelphia Philadelphia 8801.0]
[19123 Philadelphia Philadelphia 19485.0]
19406 King Of Prussia Montgomery 15520.0]
[19038 Glenside Montgomery 22347.0]

Hierarchical Clustering

- This model is maximized by the elbow method which finds where the slope stops descending rapidly.
- The elbow compares the Sum of the Squared Distance with the Number of Clusters
- Our optimum is 4 or 5, but since we had 4 counties, I went with 4.

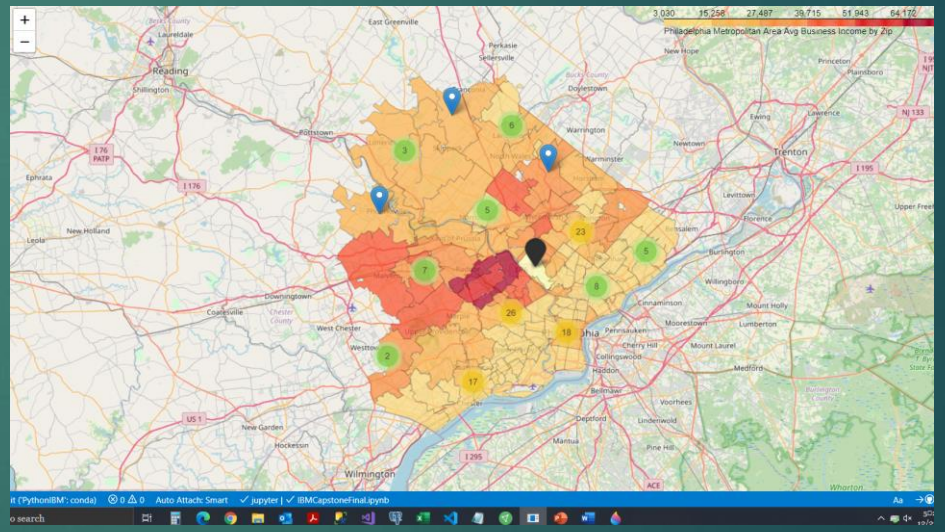


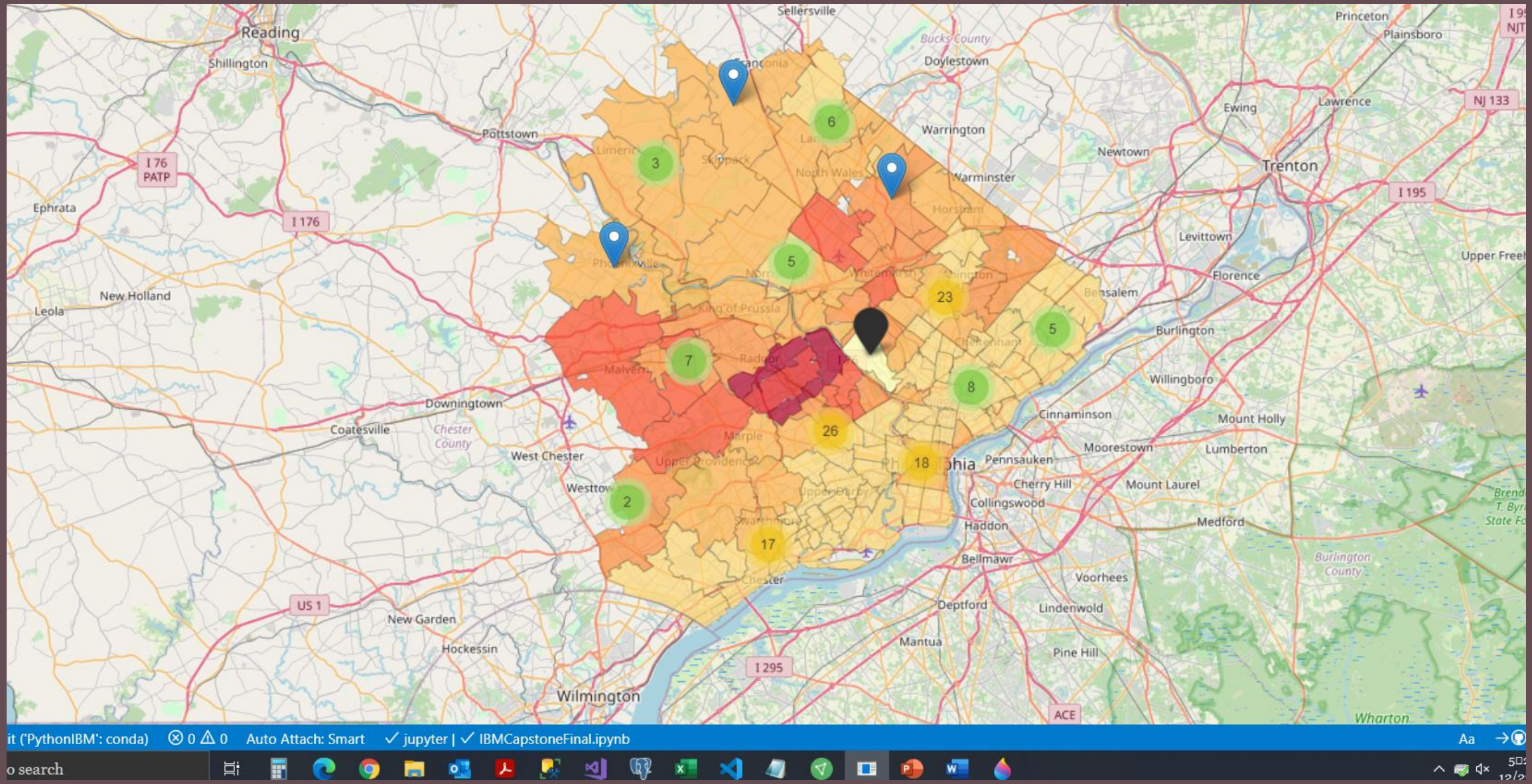
K Means Clustering

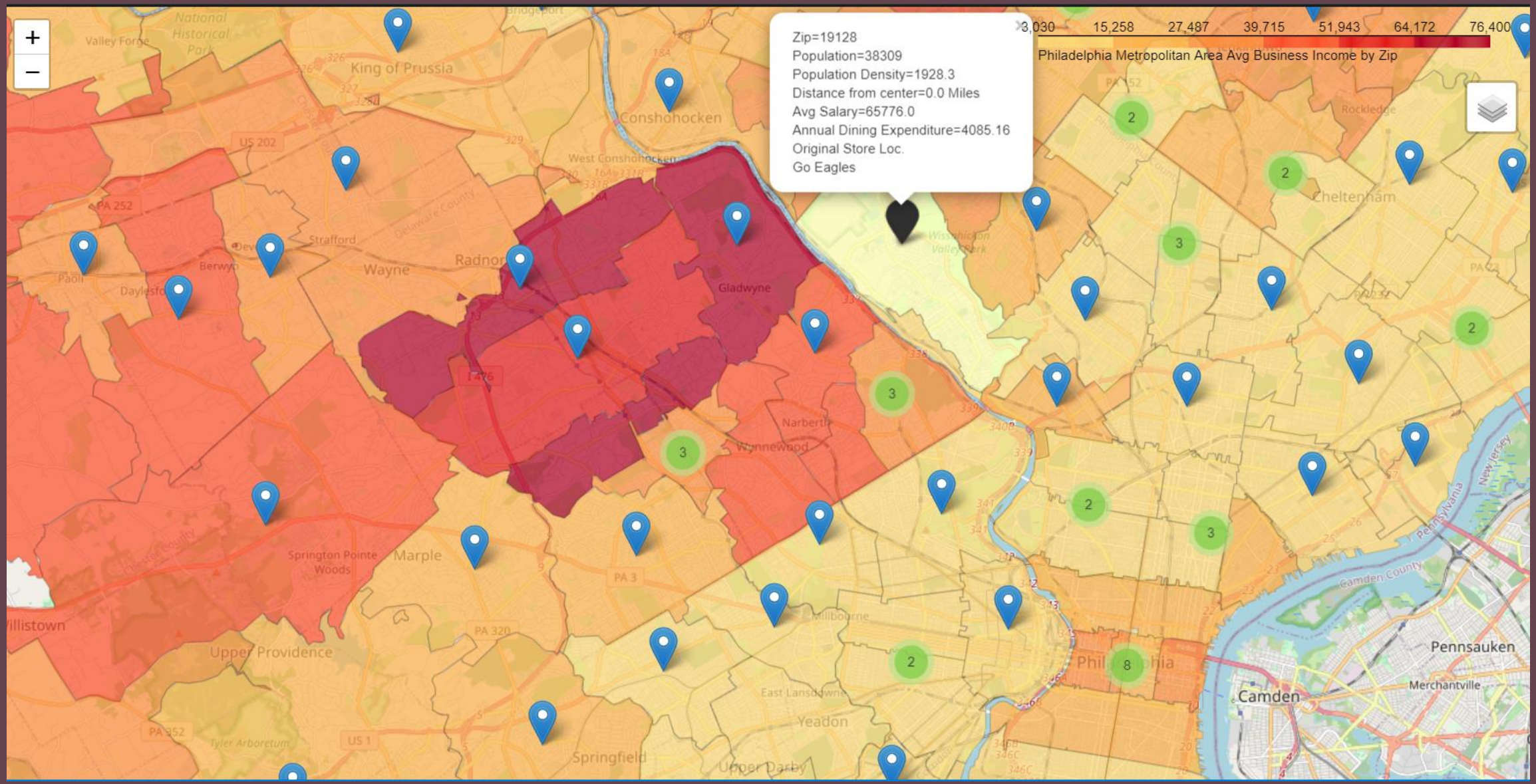
- Due to the nature of k-means clustering, each time it is run it will provide a different result
- I set up clustering for 4 clusters and 10 to 12 cycles. One thing I noticed is that either the 4 zip codes that make up center city Philadelphia are grouped into one cluster, or all of the wealthier zip codes get grouped
- The visualization while important does not add anything that can't be visualized on a map

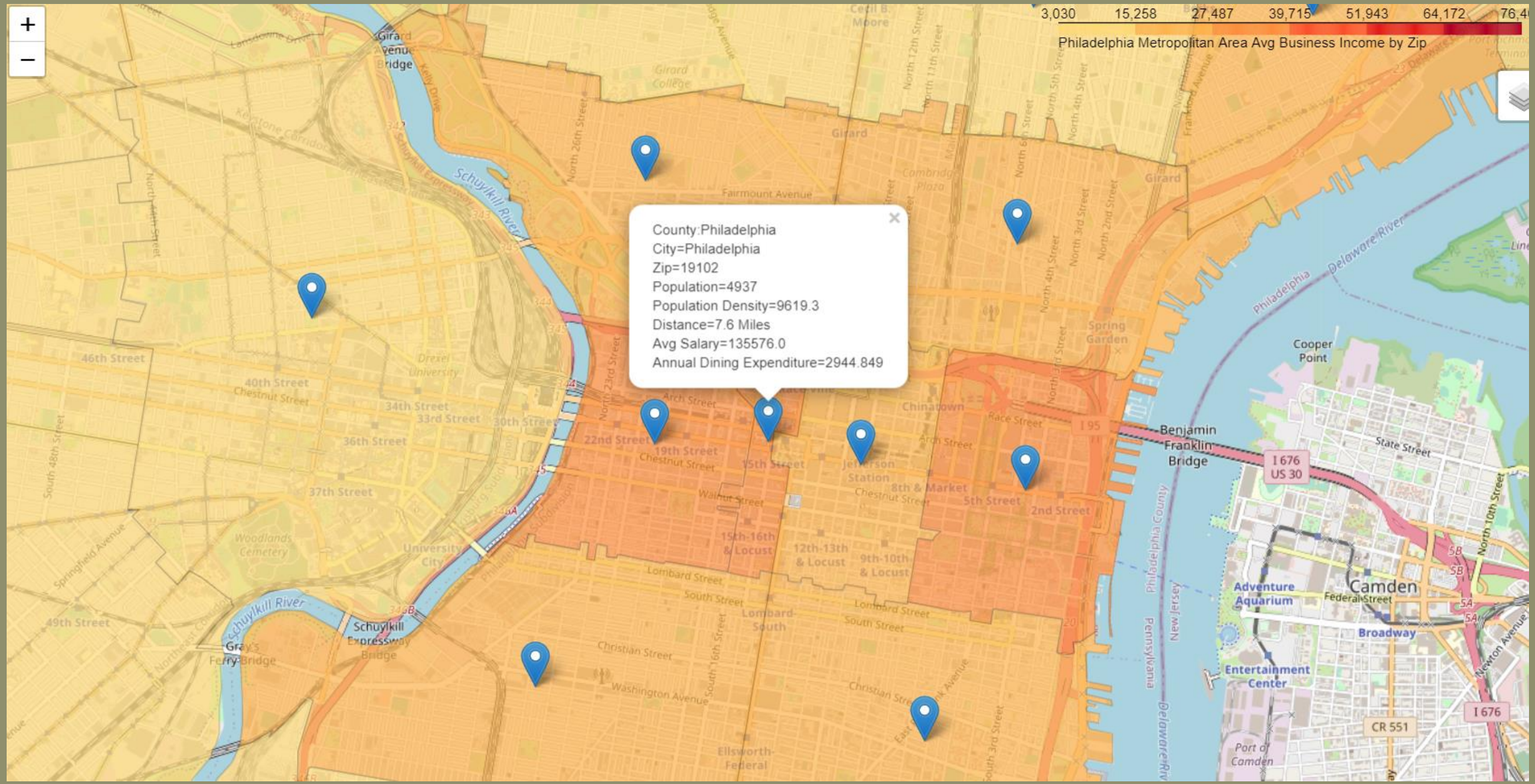
Here are some visualizations

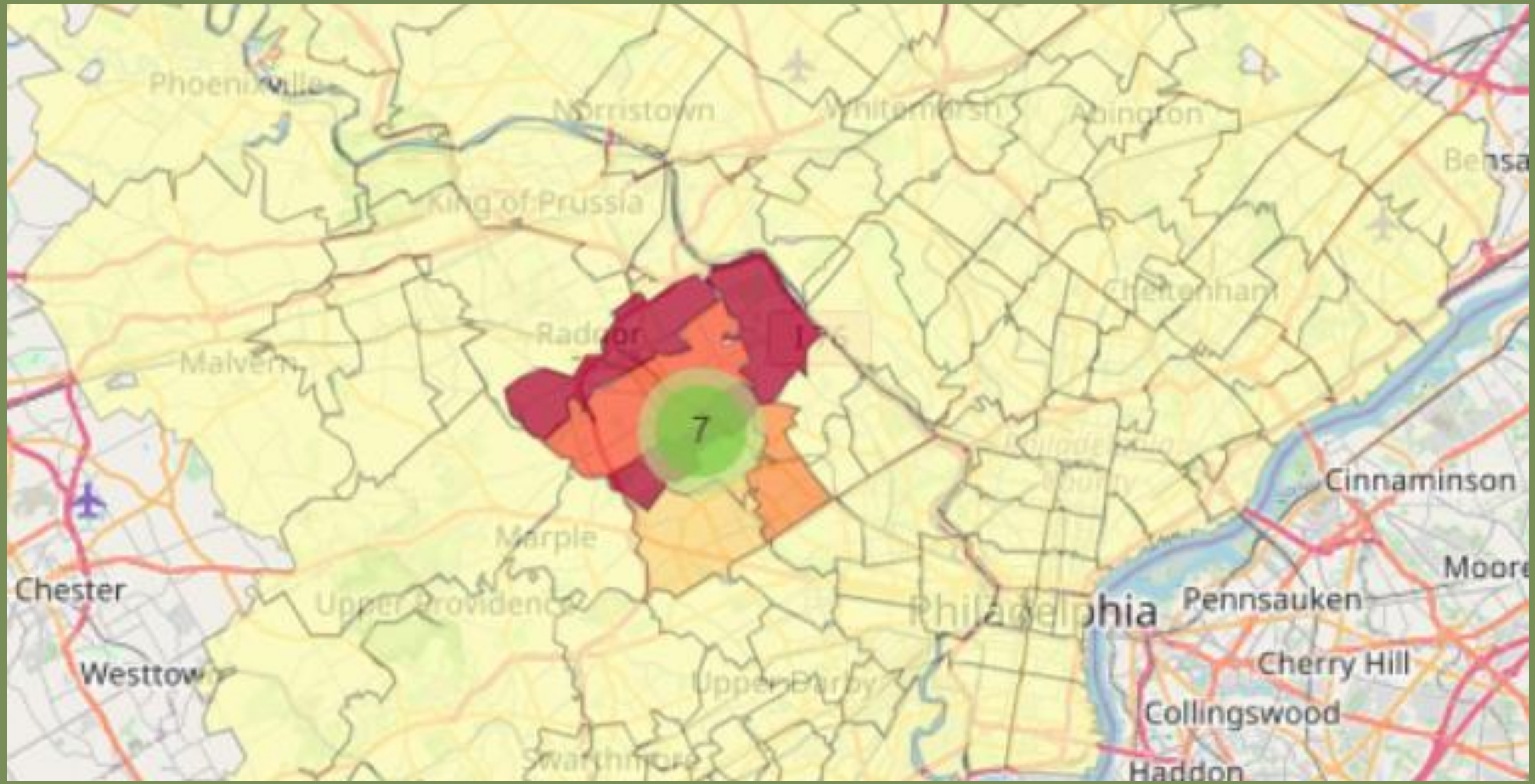
- This is the 20 mile radius within the 4 counties that I set up. You can already see the red horseshoe configuration that encompasses the wealthiest areas of the Greater Philadelphia Metropolitan area.
- The next few slides show:
  - A bigger version of this photo
  - Zoom of the central area
  - Zoom of Center city Philadelphia
  - Closeup of our Target Area

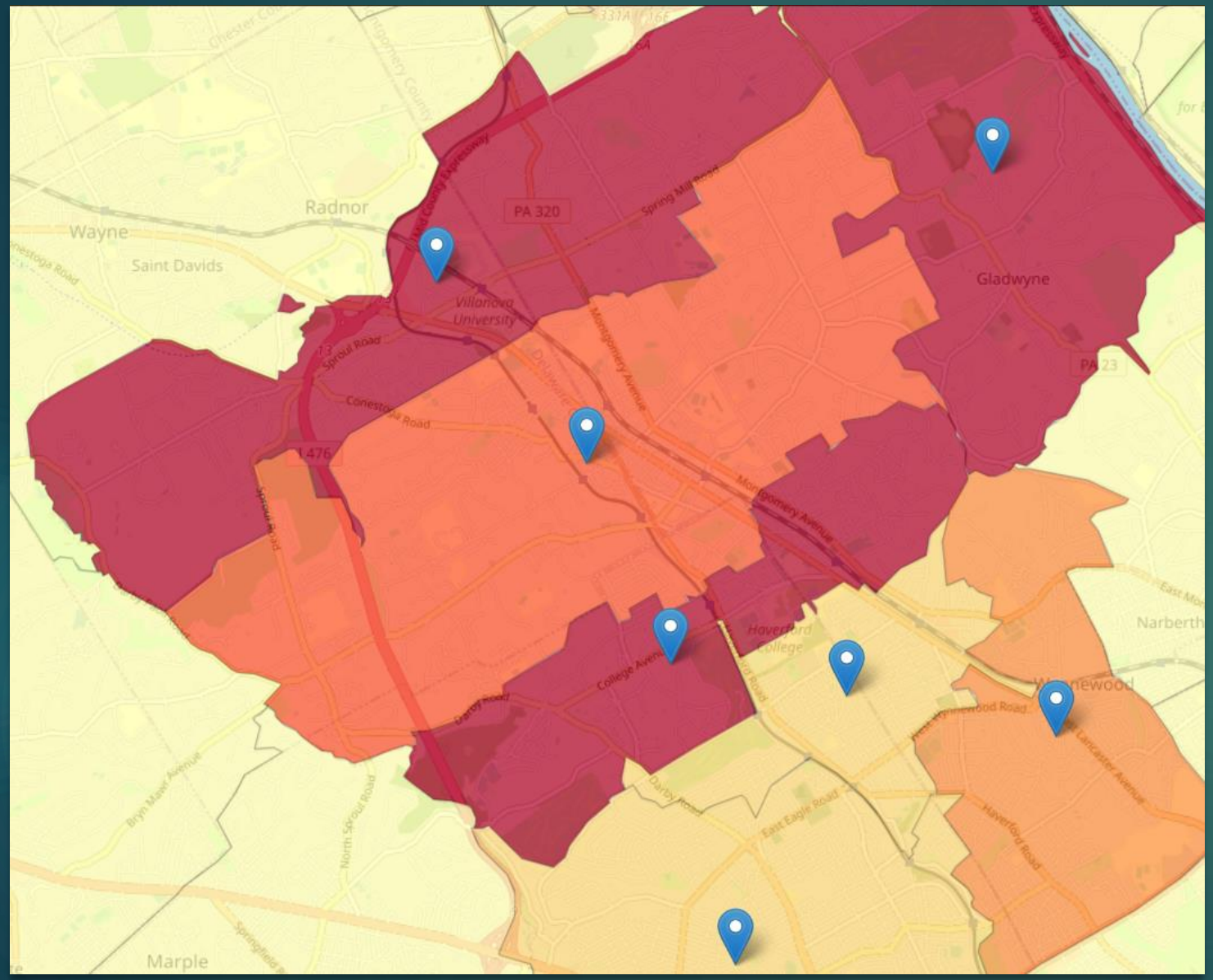



Reading
Shillington
Pottstown
Limerick
Skippack
North Wales
Warrington
Doylestown
Bucks County
Sellersville
Princeton
Plainsboro
Ewing
Lawrence
NJ 133
Trenton
Newtown
Warminster
Horsham
I 195
Levittown
Florence
Bensalem
Burlington
Willingboro
Mount Holly
Cinnaminson
Moorestown
Lumberton
Pennsauken
Cherry Hill
Mount Laurel
Collingswood
Haddon
Medford
Bellmawr
Voorhees
Deptford
Lindenwold
Mantua
Pine Hill
Burlington County
I 76
PATP
I 176
Ephrata
New Holland
Leola
King of Prussia
Radnor
Malvern
Downingtown
Coatesville
Chester County
West Chester
Marple
Upper Providence
Upper Darby
Swarthmore
Chester
Wilmington
US 1
New Garden
Hockessin
I 295
ACE
Wharton
6
3
5
23
5
7
8
26
18
2
17
it ('PythonIBM': conda)
0 0
Auto Attach: Smart
jupyter | IBMCapstoneFinal.ipynb
Aa
o search


Zip=19128
Population=38309
Population Density=1928.3
Distance from center=0.0 Miles
Avg Salary=65776.0
Annual Dining Expenditure=4085.16
Original Store Loc.
Go Eagles
3,030
15,258
27,487
39,715
51,943
64,172
76,400
Philadelphia Metropolitan Area Avg Business Income by Zip
King of Prussia
Conshohocken
West Conshohocken
Wayne
Radnor
Gladwyne
Narberth
Wynnewood
Cheltenham
Rockledge
Marple
Springton Pointe Woods
Upper Providence
Springfield
Upper Darby
Yeadon
East Lansdowne
Camden
Pennsauken
Merchantville
Camden County
Pennsylvania
New Jersey
US 202
PA 252
PA 3
PA 320
PA 352
US 1


3,030
15,258
27,487
39,715
51,943
64,172
Philadelphia Metropolitan Area Avg Business Income by Zip
County:Philadelphia
City=Philadelphia
Zip=19102
Population=4937
Population Density=9619.3
Distance=7.6 Miles
Avg Salary=135576.0
Annual Dining Expenditure=2944.849
Schuylkill River
Delaware River
Benjamin Franklin Bridge
I 676
US 30
Camden
Cooper Point
Adventure Aquarium
Entertainment Center
Broadway
Federal Street
State Street
Spring Garden
Chinatown
Girard
Fairmount Avenue
Arch Street
Race Street
Chestnut Street
Walnut Street
Lombard Street
South Street
Christian Street
Washington Avenue
Schuylkill Expressway
Gray's Ferry Bridge
Drexel University
University City
Woodlands Cemetery
Ellsworth-Federal
Port of Camden
CR 551


Phoenixville
Norristown
Whitemarsh
Abington
King of Prussia
Radnor
Malvern
7
Marple
Upper Providence
Chester
Westtown
Upper Darby
Swarthmore
Philadelphia
Cinnaminson
Pennsauken
Cherry Hill
Collingswood
Haddon



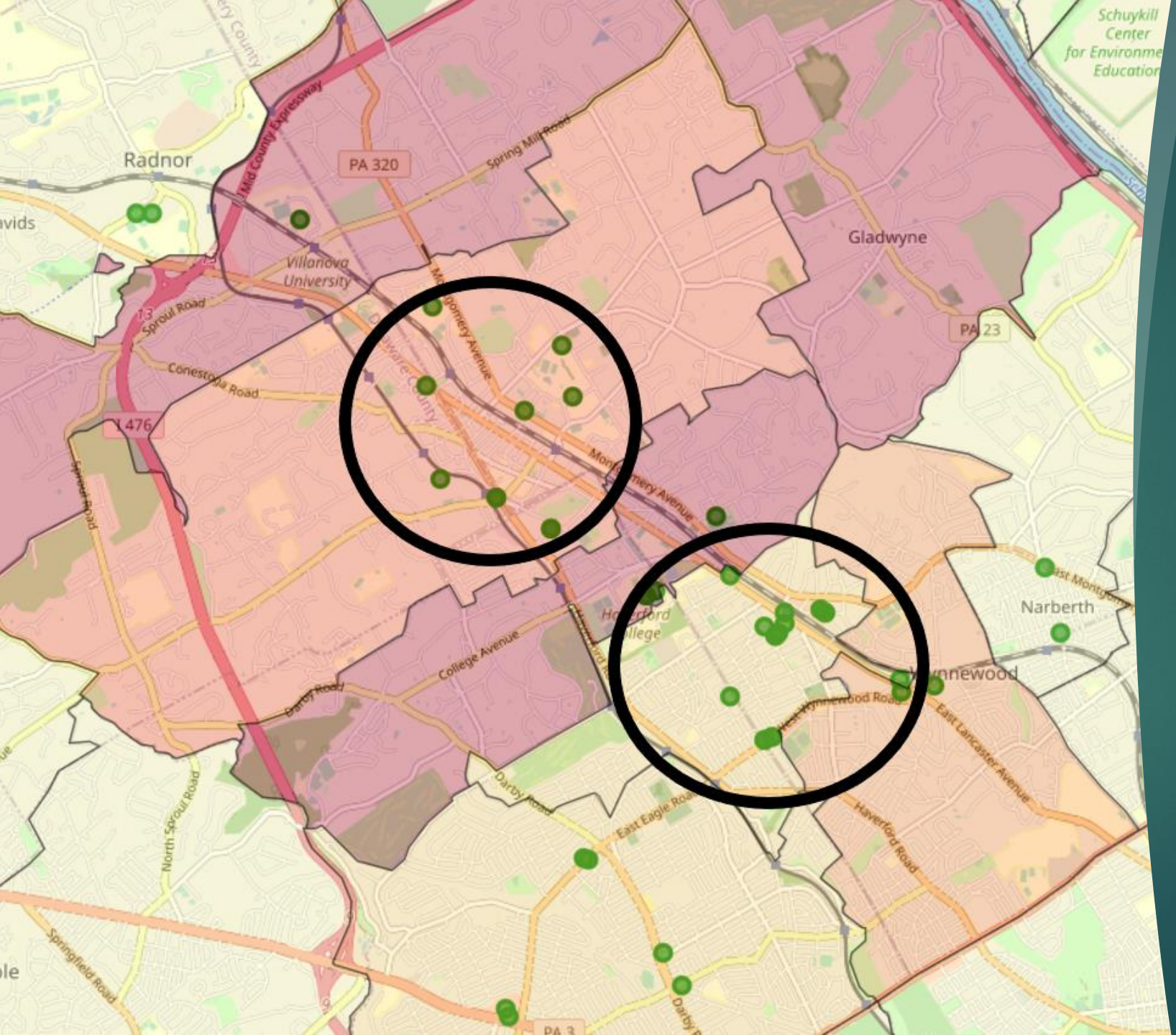
Seven Zip codes in our concentrated location

These are Villanova, Gladwyne, Bryn Mawr, Haverford, Ardmore, Wynnewood, and Havertown.

The 3 counties with the most wealth are Villanova, Gladwyne and Haverford

Two areas of interest are Ardmore which is light yellow and Bryn Mawr which is orange.
Both are close to wealthier neighborhoods but are not themselves.

Both also have higher population densities



The lower circle is our ideal location. This is next to an affluent location, can support more food establishments, and will probably generate profits.

The next location is similarly located. Bryn Mawr is slightly more affluent, but it is surrounded on 3 sides by the wealthiest areas of Philadelphia.

The choice would depend on rental costs and availablility