Jaselle Brower

GIS 5653-999 Spatial Programming

Dr. Jennifer Koch

19 August 2022

Term Project Report

Residential Areas for Families based on Proximity and Income Requirements

Abstract:

The project is based on the idea that most families may be unsure of what areas they can afford and which schools they want to send their kids to. It is an important informative project based on the idea that, most families when they are searching for an area to live, they want to find a good school and an area they can afford. With this project, it is intended to help families in finding the best area for them to live in, based on their current income. It can give them a general range of schools that are beneficial for them with their current salary range. This will allow them to research the areas and the schools based on their inputs to aid in deciding where they want to reside. Using census block group data, compiled with the City of Oklahoma's school's shapefile, this project combines the data to create a concise dataset for user input to search and find the best places for a potential family to reside at. It takes into account distance from schools and all the census block groups within a general 5-mile area around each school. This gives a short distance of driving, assuming most families prefer the schools be close to them to drop off their kids at. This project's aim is to aid potential families in finding the best area of residence for them based on their familial needs. Although it does not take into account everything a family may consider

when choosing an area of residence, it takes into account areas they can afford and their family size and kids ages, to generally help find areas they may be interested in living in.

Introduction

The general research question of this project is, what is the best possible area for three different family types to reside in based on their income and family needs in the Oklahoma City Metro?

Research done within suburban and urban areas within Malaysia and various surveys taken from parents, a majority of parents prefer personal vehicle travel due to convenience and quickness, with some areas having an element of safety involved in the desired mode of transportation for their kids to get to school (Nasrudin & Nor, 2013). Another factor, is that "Forty-seven per cent of public school parents indicated that they sent their child to their designated school without first seeking information…[with] Approximately 20% of both private and alternative school parents [indicating] that they first chose to send their child to their designated public school; but eventually changed schools" (Bosetti, 2007, pg.394). Indicating that more often than not, most parents are ill-informed of their options and choices when it comes to schools surrounding them and their options for schools. This may not be due to their own lack of choice, but due to the current districting regulations, zoning their kids to specific schools based on where they live. They base their information instead off of social media, friends, family, and neighbors are the biggest influences in the schools and where parents may choose to send their child. Bosetti concludes that with this study, it indicates the need for a way to allow those who may be socially disadvantaged, to find a place to exercise the freedom of choice in schools that middle and upper middle-class families have when selecting schools. Finally, the need for selecting schools based on socioeconomic requirements to assist parents in

finding the best place to provide education for their kids comes from the need for "…the best opportunity for a robust market to thrive where the supply of quality schools can increase and families are able to find the best educational match" (DeAngelis & Erickson, 2018). With these previous studies comes a need for a program that can allow parents to choose and find the best areas to reside in for their children to get the best possible education they can, that can provide for their adequate lifestyle needs. It is important for parents to find the best possible area because not only do most parents do limited research on areas and school districts, but most may be unaware of the options of schools they have and can possibly have to give their kids the best educational opportunity.

In general, the motivation behind this work is that most families want to find the best areas to live in within an area. It is important for them to find an area they can generally afford and a good area they can send their kids to. With parents wanting the best possible chance for their kids to gain an education, it is important for them to know all their options when it comes to possible residential areas. Initially I was motivated to create a project based on this because my husband and I are looking into good areas for us to afford with good schools for our future child. I wanted to have a program that would list possible areas that I could look into to determine the best possible area for us to live in based on our income and the schools surrounding the area so I can do more targeted research into the areas we can live in. The Oklahoma City metro was chosen as the area of interest because the only schooling areas that have a good reputation are in Edmond and Moore. But there are so many other schools that families can choose from based on their income requirements and what they need/want for their kids in terms of their education. There are a lot of schools that get written off due to their area generalizations from the local people, but with a program like this, families can do their own research on the schools in areas

they can afford. For other families, I know this could be beneficial in their research and decision making for finding good areas they are thinking of living in. Although this project does not fully represent the area fully, it will give an idea of where one can live based on their family needs and allow for the chance to look deeper into areas that fit their needs rather than going into looking into too many areas blindly.

This research is significant in that, it will allow a general overview of family needs and requirements for an area to aid in further research for a family to decide where they may want to live. It is meant to create a user-friendly application for potential parents to use to see exactly what areas may fit their particular needs. This type of research is important because most families may pick areas based on the things they know, and they may generally get stuck in an area they may be unhappy with. With this application and project, it will allow for the chance to find an area within the Oklahoma City metro, that can guide them to other areas to live in that they may not have thought to look at. This will provide the general median income of the area and schools associated with the area and their distance from the current area of interest. It is significant because parents want the best possible opportunity for their child to excel in school. By knowing their options, they can find the best places they can live to provide their children with the best possible opportunity for them to thrive.

Materials and Methods

1. Materials
    a. Using the US Census Bureau data, the table used was an overall income excel sheet for all the census block groups within Oklahoma. Census block groups were used rather than tracts or blocks as in general, a block group may more accurately

represent neighborhoods within the Oklahoma City Metro. The website the table was downloaded from was: https://data.census.gov/cedsci/table

b. Again, using the US Census Bureau, through the TIGER/Line Shapefile download, the vector shapefile downloaded was the census block groups within Oklahoma. The same justification was used for deciding between census blocks, tracts, and block groups. The website the vector shapefile was downloaded from was: https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=Block+Groups

c. Finally, on the Oklahoma City GIS data download website, the shapefile available for download for the schools was a vector polygon shapefile. The OKC metro was chosen because this is where I currently reside and I wanted to look into the general area I reside in to test and run the product for an eventual possible implementation across the United States. The website the vector shapefile was downloaded from was: https://data.okc.gov/portal/page/catalog

d. All these sources were chosen based on the need for a generalized area for a family to look in. Census blocks and tracts were a bit too small to be generalized and were more focused areas, but census block groups allow for a general area overview to allow for families to look more specifically into areas within the census block groups to decide where they may want to reside. As for the schools, the Oklahoma City metro area was selected due to it being my current area of residence to test out the first implementation of the project.

2. Methods:

a. The initial project began with selecting various geoprocessing tools for use within the idle interface to output the answers and information needed for the various user input.
b. Geoprocessing portion
    i. The data gathered was from various sources such as the Oklahoma City Government portal and the US Census Bureau. The shapefiles downloaded were in completely different projections and needed to be altered to reflect the same coordinate system and produce the answers needed.
        1. Define Projection
            a. This would help me in defining the coordinate system I would like the rest of my shapefiles to be in
            b. Arcpy management function define projection of both the census block group shapefile and OK Schools shapefile
    ii. With the table provided in the shapefile and the table downloaded from the US Census Bureau website, the GEOID of the shapefile was a string data type, while the CSV displayed the GEOID as a double. In order to allow the CSV information to match with that of the shapefile, the shapefile needed an additional column which converted the GEOID from a string to a double.
        1. Add and Calculate Field
            a. The values of the table may come up as null, so I want to match the block group data type to the data type of the CSV file, so this will help in creating the join in the next step

b. Arcpy management function calculate field of the census block group shapefile to create an incomeName column with the data type double

iii. Once the column was added, the table needed to be joined to the shapefile so the information can be parsed based on user input. The initial join takes the data from the stand-alone table and matches the GEOID from the shapefile with that of the table to create new columns and populates the information from the table into the shapefile.

1. Add Join

a. This will help in combining the census tract data with the income data downloaded to be used for the intersect later on

b. Arcpy management function join field of census block group shapefile and median income CSV by the columns incomeName and name

iv. The school data downloaded from the Oklahoma City government website was provided as a polygon. For simplicity, these were converted to points to create a general area surrounding the points within the Oklahoma City metro that may be affiliated with the census block groups.

1. Feature to Point

a. The OK Schools shapefile is a polygon file. In order to use the generate drive time trade areas, this has to be converted to a point feature

b. Arcpy management function feature to point, converting the OK Schools shapefile from a polygon to point to allow for next function to determine distance from a central point

v. Some of the block groups within the US Census Bureau shapefile did not contain any of the schools, nor were they within a close distance. Using the near function, a 5-mile buffer was created to allow the study to focus on a smaller area within the Oklahoma City metro that did not span far, and allow for distance from a school to be generally calculated based on if a possible school was near the block group a family could be in or far away from a possible desired school.

1. Near

a. This is used to calculate the distance each census block is away from the schools to further help families decide which schools they may want to bring their kid to based on income

b. Arcpy analysis function near to add into the census block shapefile the distance and name of the nearest school that are within a specified 5-mile distance of the census block group

vi. Finally, in order to join these two shapefiles into one that can be parsed and selected based on user income, a spatial join is needed to allow the school names and distances from certain block groups to create one final shapefile that contains the area, income, and possible schools a kid can

attend based on user input. It further removes the unwanted surrounding areas and leaves only the Oklahoma City metro for the program to choose from.

1. Spatial Join
    a. This is used to add in the school names and remove all the other census block groups outside of the OKC metro area, for the final shapefile to be used for user input
    b. Arcpy analysis spatial join function with the OK Schools Points shapefile and census block group shapefile, joining one to many to allow the schools to be applied to each individual census block group based on varying proximities to schools and allows for multiple census block groups to be assigned to one school

c. Coding Portion
    i. Creating the initial main file, the first thing that was done was creating the user interactive questionnaire that would allow the best answers to parse the GIS shapefile that was created above. The first step was to import all the modules such as arcpy and the user defined function module. These modules first allowed the setting of the environment and workspace that the program would run the shapefiles out of.
        1. Pseudocode:
            a. import module to run arcpy
            b. from the module arcpy import the env module

c. append the system path to a specific file path

d. import module with functions

e. set the env workspace to a specific file path

f. allow the env to overwrite outputs

ii. The second portion of this was to generate the questionnaire and set the user input into variables to call them within the functions. By creating the various print statements and input statements, it gives the user a bit of interactive material to work through to help the program determine areas best for their living situation. This was also where the general functions were called to run with specific user input.

1. Pseudocode:

a. Print “Hello and welcome to the OKC Metro ideal neighborhood finding software. Before we begin we need to gather information.” Message

b. Get user input for household income as an integer using print "Please enter your household income, rounded to the nearest thousand. EX: if your income is $52,568, please input 52000" message

c. Get user input for if they have kids or not as a string using print "Do you have any children? Please answer with 'Yes' or 'No'." message

d. Create a list of the kids ages based on user input

e. If the user has kids:

i. Get user input for private or public selection as a string using print "Do you want to live near the private or public school? Please answer with 'Private' or 'Public'." message
ii. If the above user input was private:
    1. Run defined function to find listed private schools
    2. Run defined function to narrow down the listed private schools based on median income of the area
iii. Else/if the above user input was public:
    1. Run defined function to find listed public schools
    2. Get user input for number of kids as an integer using print "How many kids do you have who are in school? Please enter with a number only." Message
    3. If the user has more than 1 kid:
        a. Get user input for the ages of the kids as an integer using print "What are their ages? Please answer with the ages separated by a comma. EX: '7,9,13' to indicate their ages are 7

years, 9 years, and 13 years old." Message, and split function to split the ages into respective numbers

b. For each kid in kids ages list from user input, set each individual value to an integer

c. Run defined function to narrow down the listed public schools based on the kids ages

4. Anything other than a number greater than 1:

a. Get user input for the kids age as an integer using "How old is your child? Please answer with a number." Message

b. Run defined function to narrow down the listed public school based on kid age

5. Run defined function to narrow down the listed public schools based on median income of the area

f. If the user does not have kids:

i. Run defined function to select areas based on median income alone

g. Run defined function to print out statements of the areas selected from the shapefile

iii. The final portion of the coding section was to develop the functions that would allow the answers the user gave to select and develop a final shapefile that would be presented to the user. This was where the general functions were defined and developed.

1. Pseudocode:

a. Define a column called “School_Nam” as a variable

b. Define function to narrow down the listed private schools based on median income of the area

i. Select rows from the specified shapefile by the school names that do not end with the text ES, MS, and HS

c. Define function to find listed public schools

i. Select rows from the specified shapefile by the school names that do end with the text ES, MS, or HS

d. Define function to narrow down the listed public schools based on median income of the area

i. Define the oldest elementary school age as 11

ii. Define the oldest middle school age as 14

iii. Define the oldest high school age as 18

iv. Define the original selection as none

v. For each kid in kids ages list from user input:

    1. If the kids age is less than the oldest elementary school age:

        a. If the original selection is none:

            i. Define the original selection as the arcpy function select rows from the specified shapefile by the school names that end with the text ES

        b. If the original selection is anything else:

            i. Add to the original selection as the arcpy function select rows by the school names that end with the text ES

    2. If the kids age is less than the oldest middle school age:

        a. If the original selection is none:

            i. Define the original selection as the arcpy function select rows from the specified

shapefile by the school names that end with the text MS

b. If the original selection is anything else:

i. Add to the original selection as the arcpy function select rows by the school names that end with the text MS

3. If the kids age is less than the oldest high school age:

a. If the original selection is none:

i. Define the original selection as the arcpy function select rows from the specified shapefile by the school names that end with the text HS

b. If the original selection is anything else:

i. Add to the original selection as the arcpy function select rows by the school names that end with the text HS

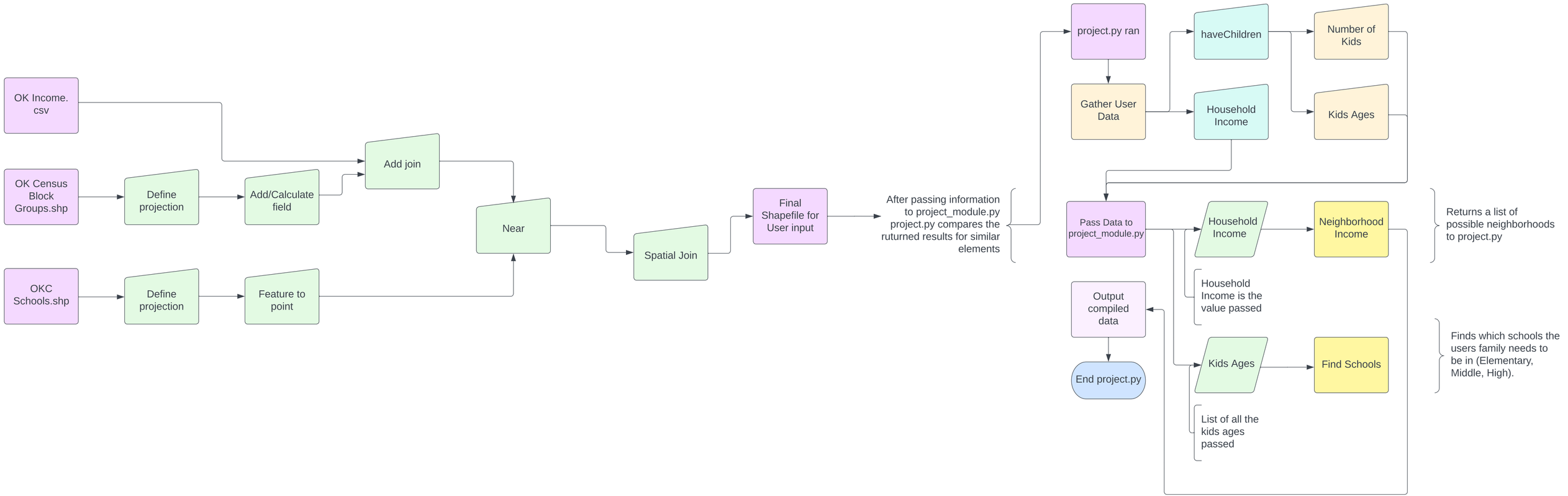
vi. Copy features from the original selection into a separate layer

e. Define function to narrow down the listed private schools based on median income of the area
    i. Select rows from the specified shapefile where the median income is less than or equal to the user inputted household income
f. Define function to narrow down the listed public schools based on median income of the area
    i. Select rows from the specified shapefile where the median income is less than or equal to the user inputted household income
g. Define function to select areas based on median income alone
    i. Select rows from the specified shapefile where the median income is less than or equal to the user inputted household income
h. Define function to print out statements of the areas selected from the shapefile
    i. Create arcpy search cursor for the specified shapefile, using the columns “Geographic; median_inc; NEAR_DIST; School_Nam”
    ii. For all the rows within the specified shapefile:

1. Print the formatted text, "Geographic area: {0}, Median Income: {1}, Distance, in miles, from school: {2}, School Name: {3}"
    a. Formatted as:
    b. Get the value of index position 0 from the column “Geographic”
    c. Get the value of index position 1 from the column “median_inc”
    d. Get the value of index position 2 from the column “NEAR_DIST”
    e. Get the value of index position 3 from the column “School_Nam”

iii. Delete the cursor

FLOW CHART DIAGRAM


OK Income. csv
OK Census Block Groups.shp
Define projection
Add/Calculate field
Add join
OKC Schools.shp
Define projection
Feature to point
Near
Spatial Join
Final Shapefile for User input
After passing information to project_module.py project.py compares the ruturned results for similar elements
project.py ran
Gather User Data
haveChildren
Household Income
Number of Kids
Kids Ages
Pass Data to project_module.py
Household Income
Neighborhood Income
Returns a list of possible neighborhoods to project.py
Household Income is the value passed
Output compiled data
End project.py
Kids Ages
Find Schools
Finds which schools the users family needs to be in (Elementary, Middle, High).
List of all the kids ages passed

Results

The results of this project were interesting. A majority of the schools for various specified criteria came up with thousands of entries. These entries did vary however, based on distance, income, and exactly what types of schools a family could be looking for. The research project was based on the idea that most families when choosing an area to reside in do not conduct in-depth research about the area they are moving into and the schools available for their kids. Based on this idea, the project was developed to help families to select the best area based on their inputted criteria.

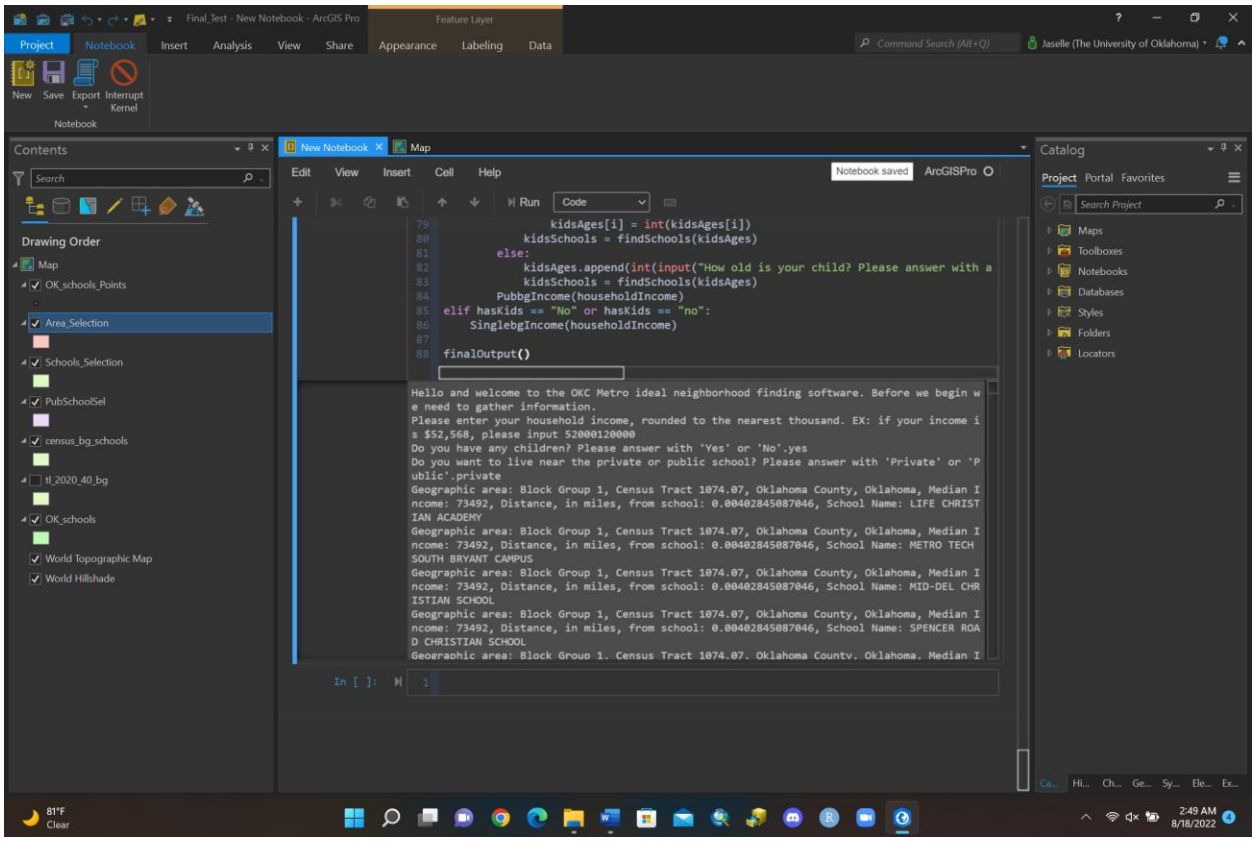
The project followed three types of family groups:

1. A family looking into alternative schools, such as private or charter schools
2. A family looking into public schools, for multiple children within the family of all different ages
3. And, a family looking into public schools, who have a single child they want to send to a school within an area

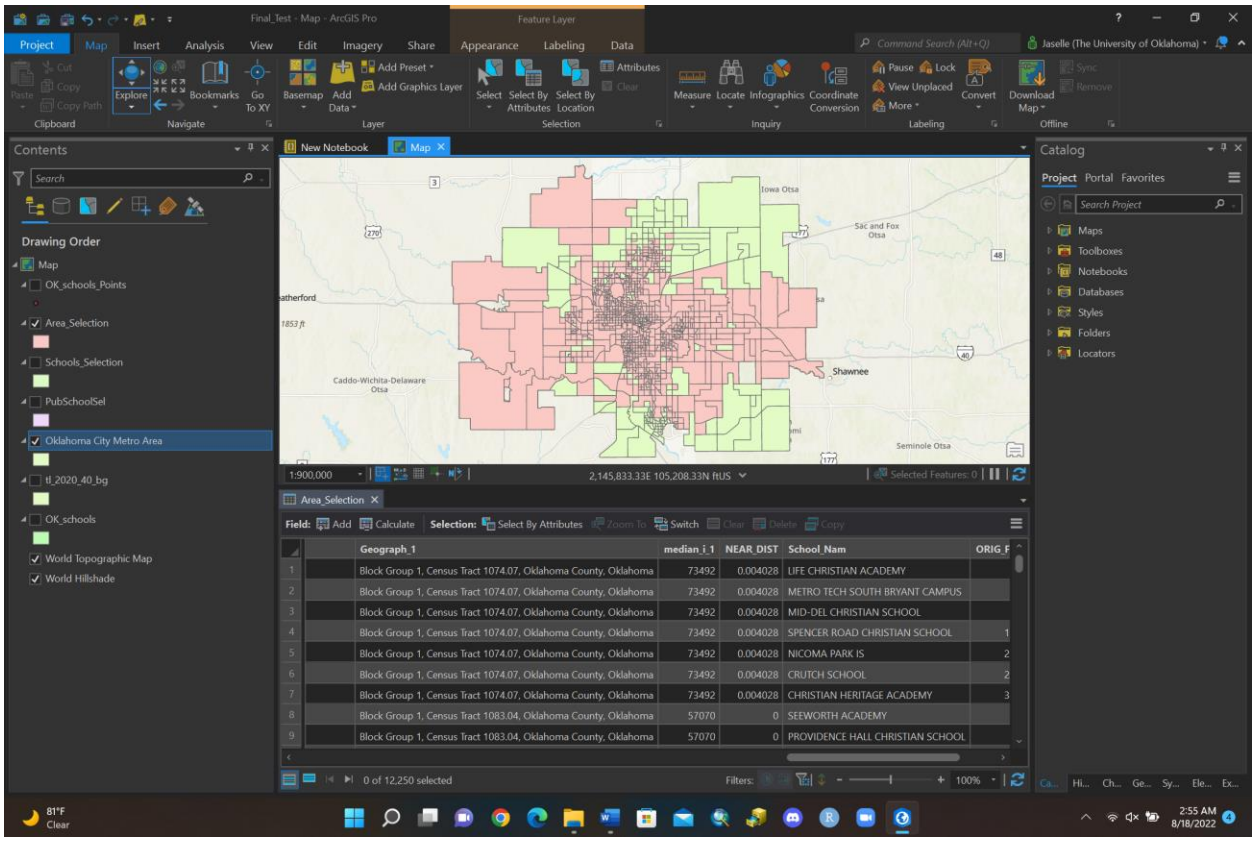
The results for each family type were very thorough. They provided each family type with a few thousand entries which are demonstrated below.

FAMILY TYPE 1

For family type number 1, the alternative school family, their average income was around $120,000. For the user questionnaire it included any census block group within the income range that was listed as a private or charter school.



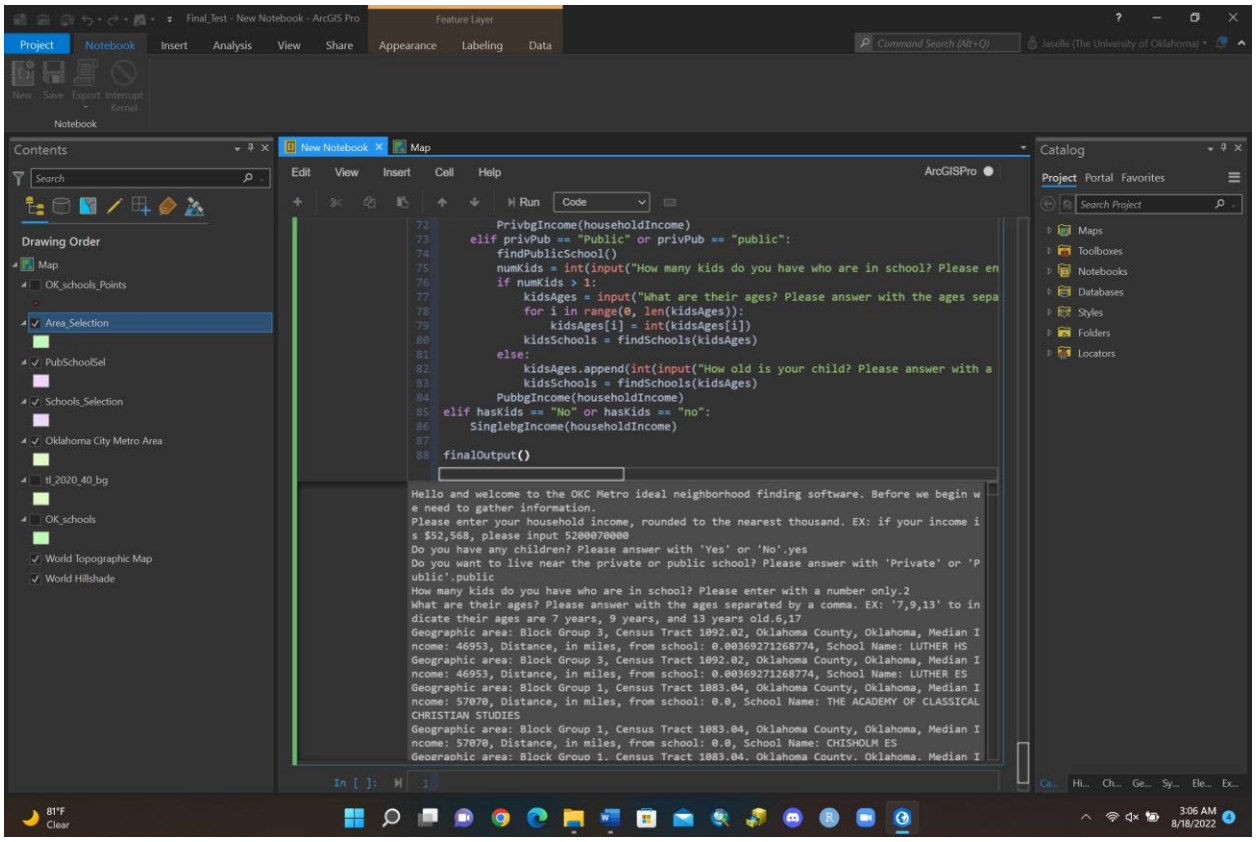
This particular family type produced about 12,000 different entries of block groups and their respective census tracts that were within a 5-mile distance of a particular school. It printed out the geographic area, median income, distance in miles from the school listed, and the school listed for the geographic area.



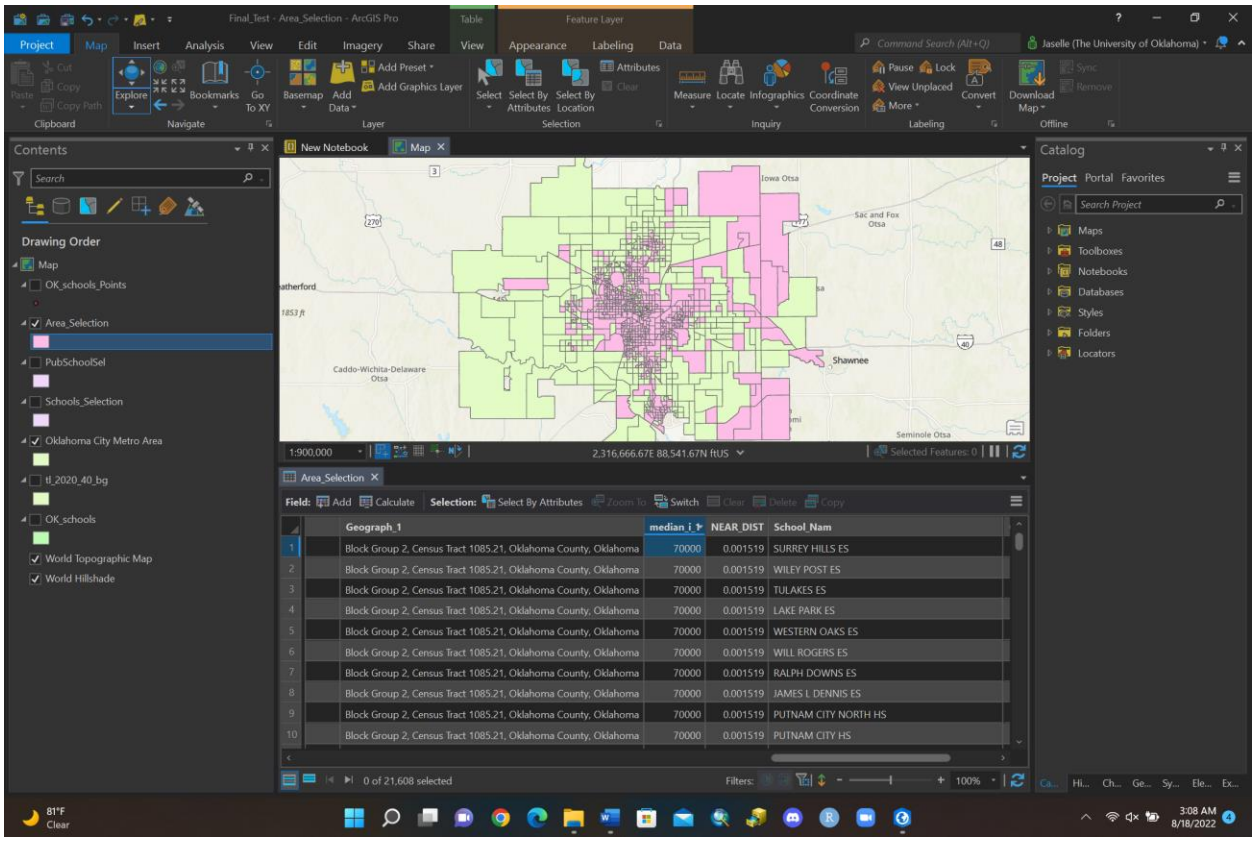
The map and table demonstrate the total amount of entries found for a private school or charter school interested family, and provides the general overview of where within the Oklahoma City metro they can reside, compared to the overarching metro area. According to the map, there is a lot of different block groups these families can live in within the Oklahoma City metro area. Based on their income and family needs, this family type can reside within a good amount of area in Oklahoma City.

FAMILY TYPE 2

For family type number 2, a family looking to send their multiple kids to various public schools in the area, their average income was around $70,000. For the user questionnaire it included any census block group within the income range that was listed as a general Elementary School, Middle School, and/or High School.



This particular family type produced about 21,000 different entries of block groups and their respective census tracts that were within a 5-mile distance of a particular school. It printed out the geographic area, median income, distance in miles from the school listed, and the school listed for the geographic area.

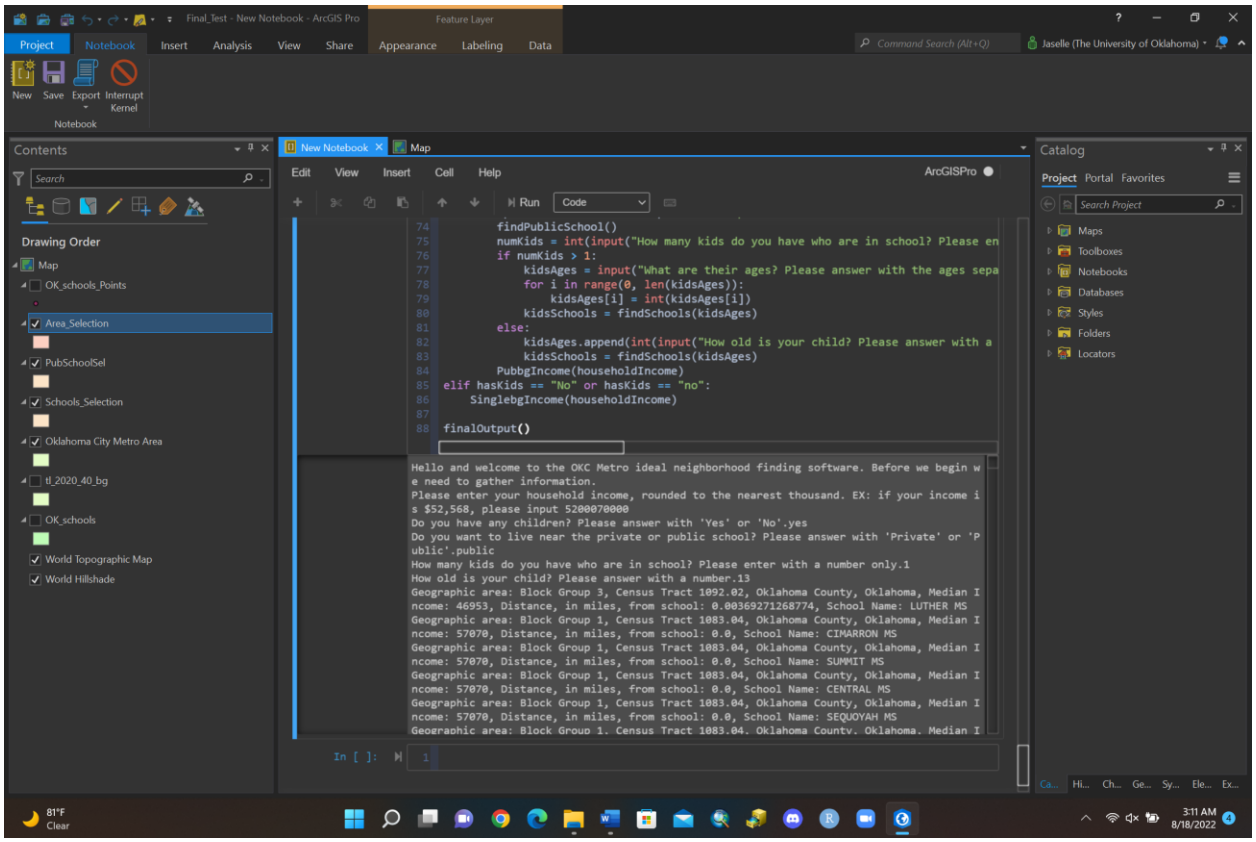


The map and table demonstrate the total amount of entries found for a public school interested family with multiple children, and provides the general overview of where within the Oklahoma City metro they can reside, compared to the overarching metro area. According to the map, there is a significantly less amount of different block groups these families can live in within the Oklahoma City metro area. Based on their income and family needs, this family type can reside within a small amount of area in Oklahoma City.

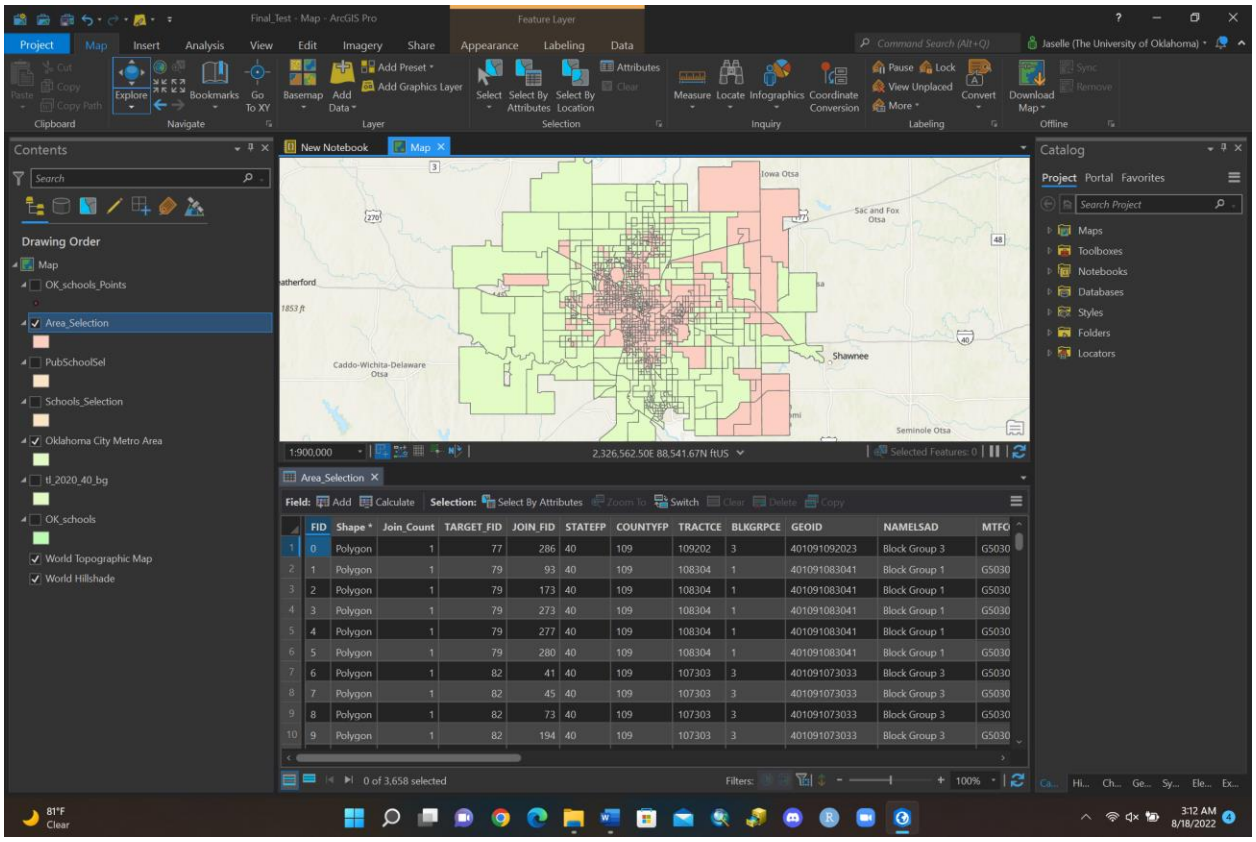
This user questionnaire was based on a family having a younger child, possibly in elementary school and a generally older child, who may be entering high school. The ages were selected at random, based on my curiosity to see the various familial combinations that could produce different combinations of block groups. This was the final choice and was the tester ages for the code.

FAMILY TYPE 3

For family type number 3, a family looking to send their only child to a public school in the area, their average income was generally also around $70,000. For the user questionnaire it included any census block group within the income range that was listed as a general Elementary School, Middle School, and/or High School.



This particular family type produced about 3,000 different entries of block groups and their respective census tracts that were within a 5-mile distance of a particular school. It printed out the geographic area, median income, distance in miles from the school listed, and the school listed for the geographic area.



The map and table demonstrate the total amount of entries found for a public school interested family with a single child, and provides the general overview of where within the Oklahoma City metro they can reside, compared to the overarching metro area. According to the map, there is a significantly less amount, as compared to the earlier private school block and multiple children public school block, of different block groups these families can live in within the Oklahoma City metro area. Based on their income and family needs, this family type can reside within a small amount of area in Oklahoma City.

The user questionnaire was based on a family that would possibly have a more pre-teen aged child. This assumption was based purely on my own curiosity to again, see how the various familial combinations could produce different combination of block groups. The age of 13 was the final choice and was the tester for the age in this particular code.

Discussion and Conclusions

Discussions

The results were really interesting. I tried a varying combination of types of families but ultimately stuck with the general tester ages I had selected when testing and correcting the code. The results are purely based on varying degrees of user input, rather than being generalized to one specific answer. It was a really interesting project to work on to develop and use a variety of user input answers to develop and let the code run and choose the areas of best residence for a family based on what they entered.

Although results for this particular research can be generally considered inconclusive because it requires many types of user inputs, it can be beneficial to know what kinds of families would use this type of software to research and discover their options to send their kids to school. Due to this, it does not yield much of a finalized answer I can discuss critically and analyze more in depth. But it does provide the option for critical discussion to see if families are not only limited financially, but education wise in terms of the opportunities they can provide for their kids.

This type of research and development can bring about a new perspective on how families can and will choose areas to reside in because with this program they have a simple and easy way to determine the general areas they can afford to live in, while also providing them with schools they can potentially send their child to. By creating this tool for families to use, it may encourage more in-depth research about education opportunities for their children.

Limitations

A major limitation of this particular project was that it was only focused on the Oklahoma City metro area. By limiting the study to a single area, it was hard to get some possible concise answers out of the particular research and fully answer the research question. The capabilities of this study were only limited by income and distance. It did not allow a choice for parents to decide how far or close they wanted to live to a particular school, but rather assumed it would be within a 5-mile distance of the school. Another limitation was that of the study including no additional socioeconomic other than income. This could also have been an important factor in families deciding where to live because they may choose one area over another based on the environment and other social factors that were not included in the study.

In all, the study is a starting point for the development or more products similar to this that encourage and inspire the addition of varying social and environmental factors for more accurate representations of areas, based on what particular families are looking for at the moment.

Based on my own limited knowledge of python, I was unable to print the final answer in a better format. Though this presents a very rough draft of the kind of program and product tis could be, there is a lot of things that need to be improved both code wise and Arcpy tool function wise. The project uses very simple and basic python/ArcGIS techniques which can lead to the project having limitations in terms of the things it can achieve. With greater knowledge of the subject and python coding, this project could service and help out so many families if developed correctly. But due to a foundational level knowledge of python and basic datasets containing one particular type of information, the project was limited in area and in capability, while still providing the foundational step for development and progression.

References:

Bosetti, L. (2007, June 27). Determinants of school choice: understanding how parents choose elementary schools in Alberta. Journal of Education policy. Retrieved August 8, 2022, from https://www.tandfonline.com/doi/figure/10.1080/02680930420002274 65?scroll=top&needAccess=true

City of Oklahoma City. (2022, August 1). Schools Dataset, Schools. Oklahoma City; https://data.okc.gov/portal/page/catalog.

DeAngelis, C. A., & Erickson, H. H. (2018). What Leads to Successful School Choice Programs? A Review of the Theories and Evidence. Cato.org. Retrieved August 8, 2022, from https://www.cato.org/cato-journal/winter-2018/what-leads-successful-school-choice-programs-review-theories-evidence

Nasrudin, N., & Nor, A. R. M. (2013, March 15). Travelling to school: Transportation selection by parents and awareness towards Sustainable Transportation. Procedia Environmental Sciences. Retrieved August 8, 2022, from https://www.sciencedirect.com/science/article/pii/S1878029613000546

US Census Bureau. (2020). 2020 TIGER/Line Shapefiles. Suitland; https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=Block+Groups

US Census Bureau. (2020). US Census Bureau Tables. Suitland; https://data.census.gov/cedsci/table.