**Fayetteville Middle School Zoning Optimization Details**

**Data Collection:**

The data was collected using GIS products found on the City of Fayetteville website, created by the City of Fayetteville. To obtain all the potential addresses we could need, we used the GIS feature, Residential Occupant. This file includes all occupied residential addresses both inside and outside of Fayetteville. Although it was last updated in 2018, with over 50,000 addresses, we felt it would provide ample insight into where the population of Fayetteville resides. We used the GIS image layer, School Attendance Zones, to add the boundaries of school zones for the Fayetteville Public School District to the map. This was last updated 2/13/2024, so it is the most up to date version of the boundaries we could find.

**Data Extraction and Preparation:**

To use our data, we had to extract it from ARCGIS Pro. For this project, we ended up creating three excel files of data before coding, and two within the code.

*We created each initial excel file like so:*

All Fayetteville Middle Schools:

For this we didn't have to do anything inside of ARCGIS, we just had to find the addresses of the schools and input them into their own excel file. We used this file to create the set ‘model.schools’ for our model.

All Fayetteville addresses + outside of Fayetteville but in the school district addresses:

Using select by attribute, we selected all addresses that listed their city as Fayetteville. Then using the school zone map layer, we ran a selection for all addresses that weren't in Fayetteville, and individually selected the address points that we could see were within the district. We then exported all selected addresses into an excel file.

All Addresses Outside of Fay District:

Using the school zone layer as a guide, we selected all points on the map that fell outside the school district and exported them into an excel file.

*We created each subsequent excel file like so:*

All Addresses in Fay School District:

Because the Fayetteville Public School District excludes some Fayetteville addresses and includes some non-Fayetteville addresses, we had to find which addresses were in the district. Using the priorly created files (all Fay addresses & all addresses outside of district), we ran an inner join on the unique object IDs of each address and filtered out any addresses that appeared in both data sets. If it occurred in each data set, it meant that it was both a Fayetteville address and outside of district, so it would not help us in our optimization.

Sample Addresses:

To create this dataset, we took the addresses found to be in the district and took a random sample of 4000. This was so we would be able to work with them later on. We used this dataset to create the set ‘model.student’ for the model.

*Geocoding Process:*

To apply the optimization model to the data, we needed to add the latitude and longitude of each address to the data frame. We did so using geopy’s Nominatim and RateLimiter. To apply the geocode to the data, we had to convert the full address from an object to a string.

With the amount of data, it proved to be impossible to apply the geocode to every address in Fayetteville. Because of this, we randomly sampled 4000 rows from the data frame to make the additions easier to handle. There are a little less than 2000 middle school students in total at the three schools in Fayetteville, so we doubled that number to ensure our bases were covered.

From there, we created a new column from the full address containing the address of each data point in the geocode format. This column was then used to add two more columns, latitude and longitude. This process was applied to the filtered addresses data frame and the middle school addresses data frame so we could calculate the distance between them.

*Distance from Address to each School:*

Using the OSMnx and NetworkX libraries, we were able to calculate the shortest driving distance (in miles) from each address to each of the three middle schools. We created a for loop that looped through each row of the data set for each school. For each address-school pair, the code calculated the shortest driving distance from the address to the school. Each pair was saved as a key in a dictionary, with the distance from point A to point B as the value. This created the distances dictionary that was then used to populate the ‘model.distances’ parameter.

**Parameters, Sets, and Distance Dictionary:**

For our parameters, we used the capacities of each school and the numbers of students and schools. We used ‘model.students’ which was the set that contained an index for each student, and ‘model.schools’ which is a set that contains an index for each school. The distance dictionary contained the computed distances between every student and each school, which is what we used to minimize the total distance traveled by students.

**Model:**

Our model is an LP designed to optimize the assignment of students to middle school based on proximity, while also using capacity constraints of each school. The objective was to minimize the total distance students would have to travel to get to their assigned schools, while also respecting the capacity of each school to keep from creating an overcrowded environment. For the decision variable, each student and school pair have a binary decision variable that indicates whether the student is assigned to that particular school (1 if assigned, 0 otherwise).

In terms of constraints, we have two - one for assignment, and another for capacity. The assignment constraint says that each student must be assigned to exactly one school. The capacity constraint says that each school cannot exceed its maximum capacity. This means that when one school fills up, the student will be assigned to their second closest school.

We selected the pyomo CBC solver for its flexibility and ability to handle various types of variables. It is customizable enough to work with our parameters.

*Objective Function:*

Minimize the total distance traveled by all students:

min

Where:

*N* = Total number of students

*M* = Total number of schools

= Distance from the student *i*’s address to school *j*

= Binary decision variable that is 1 if student *i* is assigned to school *j*, and 0 otherwise

*Constraints:*

1. Assignment Constraint (Each student is assigned to exactly one school):
2. Capacity Constraint (The number of students assigned to each school must not exceed its capacity):

Where: = Capacity of school *j*

**Challenges and Future Work:**

Due to the size of the data sets, with nearly 50,000 addresses, we were limited in how much of the data we could efficiently run through our model. To address this issue, we randomly sampled 4,000 addresses (as mentioned before). While this approach was the best to use with our resources, it inherently introduces some degree of sampling bias and won’t fully capture the intricacies of the city. Ideally, we would have more powerful machines for this project to analyze every address. These computational resources would allow for refined optimization of the rezoning that considers special characteristics and distribution of the locations, yielding a more precise zoning solution.

Since the Fayetteville Public School District is considering creating another middle school, if we had the opportunity in the future, we would investigate how adding that to the model could further optimize the zoning in a balanced way. This would also include figuring out where the best placement of that new school would be. This part of the model would be especially important if the district decided to go through with changing John L. Colbert Middle School into a junior high school. Since this process would be complicated, as it would involve finding plots of land to potentially build a new school on, we decided not to consider this in our project.

Another aspect of the rezoning that would be beneficial to consider for future work is how the process would affect the demographics of the schools. Naturally, rezoning the district would change the makeup of the schools, and it’s important to ensure that these changes promote inclusivity. Looking at the demographic shifts after rezoning can help make sure that the middle school students of Fayetteville have equal access to quality education. This kind of understanding would also allow for a more effective tailoring of support programs and allocation of resources. Ultimately, analyzing the shifts in demographics would allow for more informed decisions by the district. This was not within our capabilities as it includes demographic data we do not have access to.

While we have acknowledged the complexities that incorporating these factors into the current scope of our project would bring, they are valuable considerations for future research and refinement by those with the ability.

**Resources:**

<https://fayettevilleflyer.com/2023/01/23/fayetteville-school-board-holds-middle-school-rezoning-discussion-until-august/>

<https://www.4029tv.com/article/fayettevilles-john-l-colbert-middle-school-honors-superintendent-with-47-years-in-the-district/44207655#:~:text=Fayetteville%20Public%20Schools%20said%20about,L%20Colbert%20Middle%20School%20home>.

<https://www.usnews.com/education/k12/arkansas/holt-middle-school-275105#:~:text=The%20student%20population%20of%20Holt,the%20school%20serves%205%2D6>.

<https://adedata.arkansas.gov/statewide/ReportList/Schools/EnrollmentCount.aspx?year=34&search=&pagesize=10>

<https://www.fayetteville-ar.gov/384/GIS-Interactive-Maps>

<https://district.fayar.net/page/school-locator>