**Problem 1 Assumptions**: We make the fundamental assumption that the scenario of childcare in NYS can be formulated into a linear program with a finite, feasible, and convex region of solutions. This assumption is necessary as it forms the basis for our linear program. Another assumption we make is that if an existing facility currently has a total capacity of 0, it will not be expanded. This makes sense because the maximum number of slots a facility can have is the minimum of 500 and 1.2 times its current capacity. A 1.2 multiplier on 0 is still 0, so it cannot be expanded. This assumption is necessary because there are a handful of existing facilities with a total capacity of zero.

**Data Preparation:** The data is provided in the project instructions and needs only a little cleaning after downloading. We first drop unnecessary data from our population, such as the number of people aged 40-45 in a zip code. Then, we calculate the 0-12 population using proportions as follows, rounding down at the end:

Then, we use a similar methodology to calculate the 0-5 capacity of an existing facility.

Finally, to wrap up the data cleaning, we drop existing facilities that have a total capacity of zero, because they are of no use. They cannot be expanded, as per our assumptions.

**Problem 1 Decision Variables**: The first set of decision variables are non-zero integers that represent how many small, medium, or large facilities to build in each zip code. Hence, we have 3 \* number of zip codes, as the quantity of these decision variables.

For expansions, we make two more sets of decision variables: one representing the number of 0-5 slots we add to each existing facility in each zip code, and one representing the number of 5-12 slots we add to each existing facility in each zip code. It can be defined as follows:

**Problem 1 Cost Function**: The cost function is built piece by piece. First, we will add the cost of building new buildings. This is simple, as the cost is fixed for each size and given to us:

For expanding existing facilities, the cost is more involved. It considers the current capacity of the existing facility as well as the number of slots added of each type. The relationship for a single facility is described below:

To get our final cost function, we simply add our cost of building new facilities in zip code with the cost of expanding existing facilities in that same zip code. Then, we sum over all zip codes in NYS to get the total cost. That results in our final objective (second summation is still inside the first one):

**Problem 1 Constraints**: The first constraints we can add are on the decision variables. Since we can only build a positive integer number of new buildings in each zip code, we must restrict that variable to be non-negative and integer. Same goes for expanding slots in existing facilities, they must be integer slots greater than zero.

Next, we add the constraint that a facility can only be expanded by up to 20%, or until it reaches 500 slots. This is represented as below:

Then we set demand constraints, where depending on whether the zip code is high demand or normal, our facilities must provide the required number of slots. To do this, we first calculate the new number of slots available in each zip code. This is simply the existing slots plus the slots we gain from building new facilities plus the slots we gain from expanding existing facilities, all in a particular zip code. The number of slots we gain by making new buildings are provided in the instructions.

Then, we obtain the corresponding populations in that particular zip code:

Then, we make the final demand constraints, changing based on high demand status for total capacity but keeping the same 0-5 capacity constraint regardless of high demand status. High demand in a zip code is defined as having employment over 60% or income under 60,000:

**Results**: After optimization, we can successfully meet all demands and constraints by using $316,248,854. To analyze our data we first begin by grouping our zip codes into groups of 11 by their proximity. This makes the data less granular and easier to view in charts. Once the data is aggregated into groups of 11, we first look at how often each size of building was built in each group:

A graph of a number of buildings

Description automatically generated

From the chart, we can see that the overwhelming majority of the time, large buildings are built in our zip codes. This makes sense as they provide a very large number of slots for less than double the cost of small facilities that are built. We also see that the amount of small and medium buildings is mostly flat across all the zip code groups, while the large buildings frequency has a discernible shape. One unique point is that the peak of the large buildings being built happens in zip code group 25, which corresponds to zip codes in the Northern Poughkeepsie area. This area may be susceptible to childcare deserts due to its low socioeconomic status in its resident population. Next, we look at expansion vs building new facilities. The results are seen below:

A graph of different numbers

Description automatically generated with medium confidence

In this chart we can see that most slots are added by building new facilities rather than expanding current facilities. Looking deeper, we can see that both distributions roughly have the same shape. This implies that when slots are added, usually you must build new facilities and expand existing ones, you just obtain more slots by building. It also implies that there is no zip code or geography where it is disproportionally better to expand vs build new slots, the shapes are roughly the same. Finally, we look at expansion in detail, analyzing which type of slots are usually added when facilities are expanded. The results are below:

A graph of numbers and numbers

Description automatically generated with medium confidence

In this chart, we see a density of points on the x-axis, implying that most of the time you will be adding 0-5 slots. This makes sense as most existing facilities had a shortage of 0-5 slots. We also see that the range of the 0-5 slots is a lot higher, reaching 4000 in some zip code groups. Whereas for 5-12 slots, we add at most 250 in a zip code group. The last deduction that can be made from the data is that 5-12 slots are only really built when not a lot of 0-5 slots are built. This indicates that perhaps facilities may have a shortage in one type of slot or the other.

**Mathematical Formulation:**

**Notation Guide**

**Decision Variables:**

**Handy Expressions:**

**Mathematical Formulation: (2nd summation is still inside first summation)**

**Subject to**