Point-of-Dispense Optimization for Allegheny County

DECISION ANALYTICS FOR BUSINESS AND POLICY- FALL 2020

Ajay Valecha,Akshay Oza, Chetna Bakhshi, Lara Haase

All members contributed equally to the project

Decision Analytics for Business and Policy

Ajay Valecha (avalecha), Akshay Oza (aoza), Chetna Bakhshi (cbakhshi), and Lara Haase (lhaase)

Final Project Report

Background

Our model is inspired by a research study by Boonmee et.al (2017) that examined facility location optimization models in case of emergency. They ran a deterministic model to locate centers based on input parameters such as capacity, individuals impacted, transportation costs, and fixed costs. However we faced the problem of not having access to exact costs. Additional research found a study by Rawls & Turnquist (2009) which developed a tool to determine location and quantities of various emergency supplies using stochastic integer programming. We attempted to use their approach to calculate costs, but faced many challenges, and decided instead to conduct a cost sensitivity analysis.

Problem Definition

With our focus on Allegheny County, we aim to build a model to help choose Point-of-Dispense (POD) sites which reach all residents in the county in case of a public health emergency. Two models based on a deterministic approach will be created for the same aim with varying objectives: first, minimizing total distance from all census tracts to their assigned PODs, and second, minimize the maximum distance between any census tract and its assigned POD. Further, both models will be run in two scenarios: first, when PODs could be run on full capacity, and second, when PODs cannot be run on full capacity due to special measures such as social distancing.

Modeling Assumptions

- A communicable disease is a threat to the community, and testing kits for it are available.
- Mass dispensing of the testing kits for the disease is required to get community members tested with the aim to detect and contain the spread of the disease.
- Emergency incidence and facility construction occur within a fixed single time period
- Testing kits are self-administered, at the homes of the test takers.
- One testing kit is only for one individual
- All testing kits are of good quality, so an individual will not require additional kits
- All testing kit needs for all PODs are easily and quickly met without constraints
- All staffing and testing kit needs for all PODs are easily and quickly met without constraints
- All households from the same Census Tract will be assigned to the same POD
- Every census tract is only assigned to one POD
- PODs can have multiple Census Tracts assigned to it.
- One testing kit is only for one individual
- All testing kits are of good quality, so an individual will not require additional kits
- All households have access to a vehicle

Formulation

Parameters	
Indices	
Census Tracts (census tracts in Allegheny County)	i = 1,..., 402
Potential POD sites (high schools available for setting up PODs)	j = 1,...., 47
Set	
distance from population centroid of Census Tract$_i$ to POD$_j$	d_{ij}

number of households in Census Tract_i	h_i
max number of PODs county can afford to set up = 1, ... , 47 (for Cost Sensitivity Analysis)	n
capacity of POD j	c_j
Decision Variables	
whether to set up site j as a POD (binary)	x_j
whether to send households in Census Tract_i to POD_j (binary)	y_{ij}
max distance of all Census Tracts to their assigned POD ***(for objective 2 only)***	m

Formulation based on the following two objectives:

1. Minimize expected total travel distance of the population for Census Tract i to assigned PODj
2. Minimize expected maximum travel distance for any Census Tract i to its assigned PODj

Objective	Min $(\sum_i \sum_j x_j * y_{ij} * d_{ij} * h_i)$	Min m
Constraints		
Limit number of PODs created	$\sum_j x_j \leq n$	$\sum_j x_j \leq n$
Capacity of PODs	$\sum_i y_{ij} * h_i \leq C_j x_j \quad \forall_j$	$\sum_i y_{ij} * h_i \leq C_j x_j \quad \forall_j$
Do not assign any i to j if j does not exist	$y_{ij} \leq x_j \quad \forall_{i,j}$	[1] $y_{ij} \leq x_j \quad \forall_{i,j}$
Every census tract is assigned to only 1 POD	$\sum_j y_{ij} = 1 \quad \forall_i$	$\sum_j y_{ij} = 1 \quad \forall_i$
Minimax constraint: the distance between any census tract and its assigned POD must be less than or equal to the overall max distance	----	$\sum_i \sum_j y_{ij} * d_{ij} \leq m$

Data Sources & Assumptions

Data Related Assumptions:

- Average Household Size in Allegheny County is approximately 2.21 (Source) .
- The population centroid coordinates from 2010 are still accurate for the current (2018) population data (this is the best/most recent data we can find)
- The centroids in the census population tracts closely mirror the household centroids.
- Population centroids are better than geographic centroids as it helps us to show the true population distribution more accurately.
- Household proportions across census tracts in 2018 are still accurate for 2020.

[1] We have used a modified constraint in MinMax Distance model for normal scenario without social distancing:

$\sum_i \sum_j y_{ij} \geq 402$

- The schools' square foot capacities divided by 6 will provide us with a close enough estimate of the largest hall/gym/area that can be used for dispensing within the POD.
- The estimated POD capacities when POD can run on full capacity were further divided by a factor of 30 sq ft per person and again by 2.21 people per household (as our population count is households) to get the number of people that could be served by the POD hourly. (Source)
- The estimated POD capacities for the social distancing scenario were divided by a factor of 113 sq ft per household (but not by 2.21 per person as households would be expected to stand together) to get the number of people that could be served by the POD hourly. (Source)
- Families travel together to PODs for getting tested/vaccinated together (so distance counts only once per household)
- Using a distance API matrix is better than the aerial/spatial distance as the latter does not give us an accurate estimate of the actual distance between population centroids and the POD sites.

Data Sources:

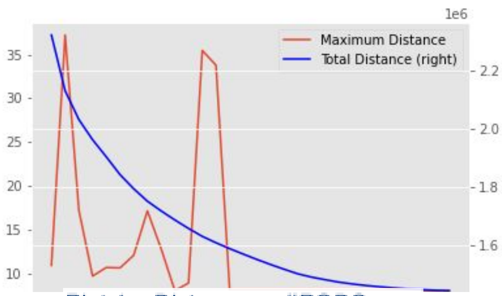
S.no.	Data	Source	Description
1	Household population counts at Census tract level for Allegheny County	ACS: Link	Data on Census Tracts was retrieved to get the total households in each Census Tract in Allegheny County. In a public health emergency, all household members are assumed to get tested/ vaccinated together. Basis for this assumption: using total households helps us plan the setting up of PODs efficiently.
2	Population Centroids for Census Tracts	Baruch College: Link	The population centroids for census tracts are geographic coordinates accounting for density of the population in a census tract and based on the density distribution finds a centre/balance point within the census tract. We assume this is better than using geographic central points since it will help us calculate distance more accurately showing the true population distribution.
3	POD (high schools) Long/Lat coordinates	Google Maps	The latitudes and longitudes were manually retrieved from Google Maps, then used in the distance matrix to calculate exact road distance between census tracts and potential PODs sites.
4	Distance Matrix	API Google	The API was used to calculate the road distance between coordinates of potential POD sites and population centroids of census tracts. Using the distance matrix and API, we calculate road distance based on origin and destination coordinates.
5	POD Capacities	Google Earth	Google Earth was used to find the building area of each school. This was used to derive the area available for setting up PODs.

Model Implementation

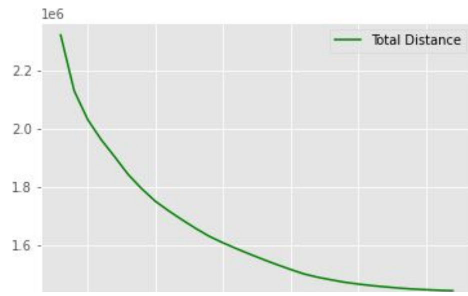
We implemented our model in Python, using the Gurobi software for optimization. Our code closely follows the model formulation. The only significant issue we met was while running our Cost Sensitivity Analysis: we chose to write a for-loop that iterates through all possible numbers of PODs (1-47) and added a constraint limiting the maximum PODs to the iteration number. When running the loop from 1 through 47 however, the results would find **all models infeasible.** But when only running one iteration at a time, we found at least some scenarios with higher numbers of max PODs in fact were **feasible**. Finally we realized the first constraint restricting the model to only have a maximum of 1 POD was still part of the model. We were then able to run the loop successfully by starting with 47 max PODs and iterating backwards.

Analysis - (Social Distancing Scenario)

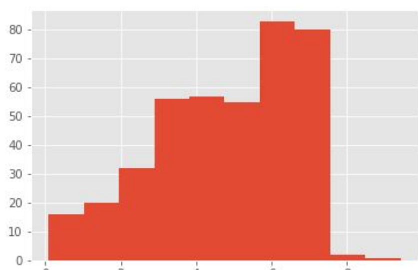
The two major decisions a policy maker would be concerned about is how many POD sites to build and which objective strategy should be used to assign census tracts to POD sites. Here we provide analysis to make it easier for policy/decision makers to answer these questions. (**Note: All distance values are in miles**)

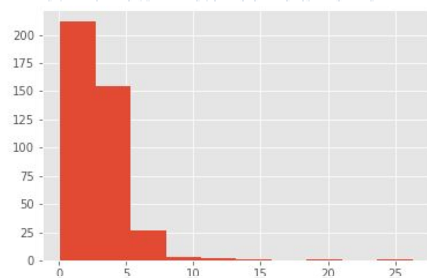


Plot 1.a Distance vs #PODS



Plot 1.b Total Distance vs #PODs


Plot 2.a Histogram- Minimzied Maximum Distance

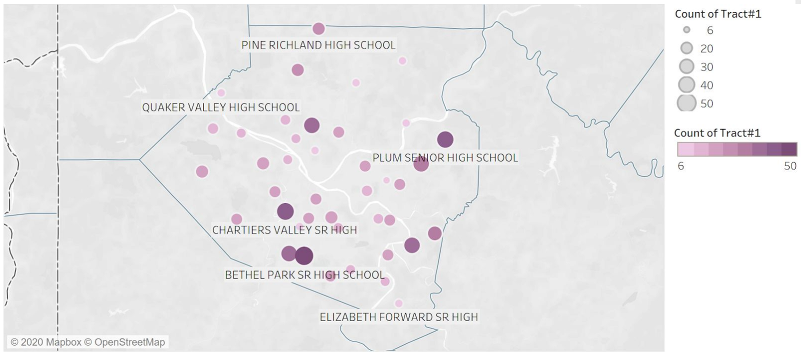

Plot 2.b Histogram - Minimized Total Distance

From plot 1.a we see the model becomes feasible only when at least 18 PODs are built due to our capacity constraints.

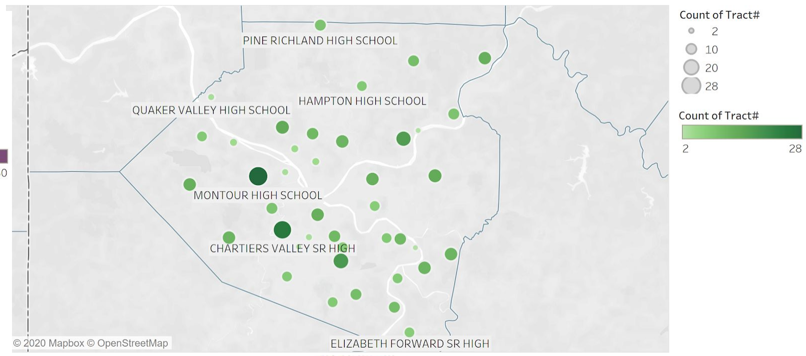
Minimized maximum distance becomes constant after 31 PODs, meaning no matter how many additional POD sites are built the maximum distance minimized would not be further minimized . Taking into account this constant behavior, minimized total distance is the factor deciding the number of POD sites. Plot 1.b shows the minimized total distance becomes almost constant after 41 PODs. So in the social distancing scenario, we decide initially to build 41 PODs as optimized by both models, then the assignment of census tracts prescribed by both the objectives is analyzed further.

Plot 2.a shows the minimax strategy assignments require most census tracts to travel a distance of 4-7 miles. In plot 2.b with minimized total distance, although the majority of census tracts are travelling less than 5 miles, there are few that have to travel 20 miles or more. This is unfair for households in these tracts. A public health emergency requires fairness and equity to be made a priority over efficiency. Further our spatial analysis provides more evidence to support the minimax strategy.

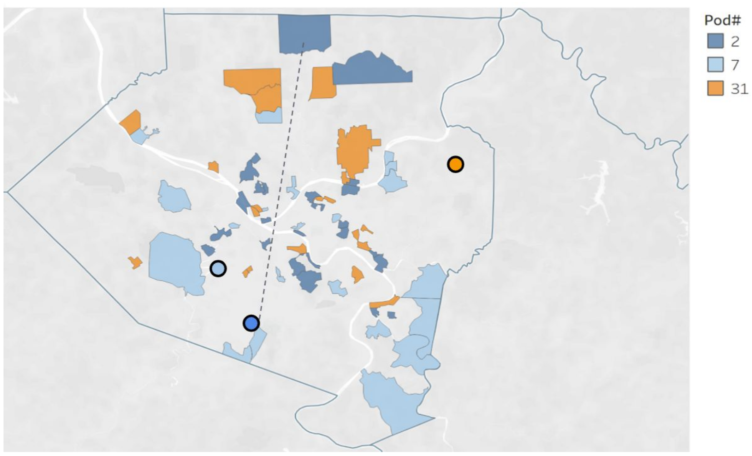
In Map 1 (below) we see assignment of census tracts using the minimized total distance objective. Here concentrated POD pockets formed and four major POD sites end up serving almost 50% of the total census tracts. Whereas, in Map 2 (below), assigned by the minimax objective, we see census tracts more evenly assigned to POD sites with a maximum of 28 census tracts assigned to a POD.



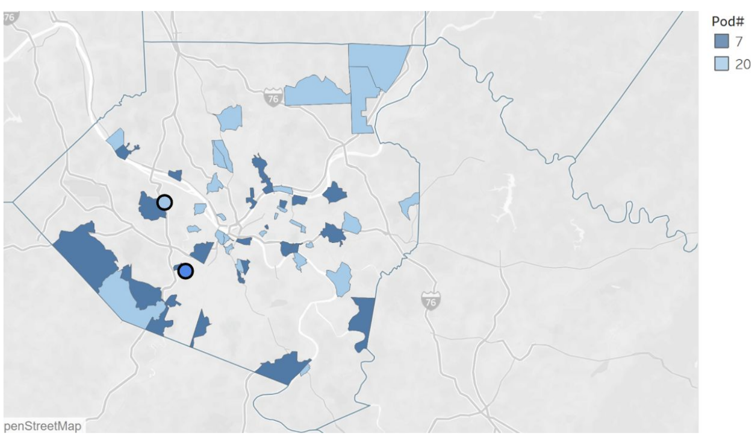
Map 1



Map 2



Map 3



Map 4

Further if we map the top 2-3 POD sites (highest number of census tracts assigned to them) we see in Map 3 that with the minimize total distance objective, there are census tracts assigned to PODs located at the opposite end of the county. On the other hand, Map 4 shows that although not perfect, assignment of tracts is more even and tracts are assigned in such a way that the highways and expressways provide direct connectivity to PODs.

We ran a similar analysis for a normal, non-social distancing scenario and found building 35 PODs and using assignments from the minimax strategy gave us similar inference as the social distancing scenario. In this case the minimized maximum distance becomes constant after 20 POD sites.

Recommendations

The main goal for decision makers would be to make sure the impact of this task is maximized. One way to measure impact is the coverage of the number of households that received testing kits from POD sites. In this context, since residents have to travel to POD sites, they would do so only if POD sites are within a distance bound. Also, as we mentioned before, fairness and equity would take priority over efficiency, so we suggest the census tracts to be assigned using the minimax strategy. Further, if decision makers still care about total distance traveled, a combination of the two objective strategies will provide an optimally fair and efficient allocation. In the social distancing scenario the minimized maximum distance becomes constant after building 31 PODs, so decision makers can use assignments from the minimized total strategy for the range of PODs above 31, and with respective assignments, backtrack and calculate the total distance travelled to choose the exact number of PODs that fits their threshold of total distance traveled. Similarly for the normal scenario this backtracking analysis can be done for a range of 20 -47 POD sites.

An additional recommendation is that in the case of such scenarios presented here, the county could make a contract with rideshare companies in order to assure that all residents have access to vehicle transport to get to their assigned POD.