[94867] Decision Analytics for Business and Policy Final Project

The Expansion of the Pit Stop Program in San Francisco

(Jamie Lim, Shun Tomita, Dai Ling Wu, Lisa Yeung)

Resources Used & Links to other Deliverables:

Software packages	- Gurobi - Folium
Code samples from other sources	1. Codes for creating hexagon heatmaps (https://jens-wirelesscar.medium.com/lhexagone-in-hexagons-uber-h3-map-1566bc412172) 2. Codes for visualizing grids with results on maps (https://www.jpytr.com/post/analysinggeographicdatawithfolium/)
Link to datasets	https://github.com/Shun-Tomita/PitStopOptimization/tree/main/dataset
Link to codes	https://github.com/Shun-Tomita/PitStopOptimization/tree/main/code

Work Distribution:

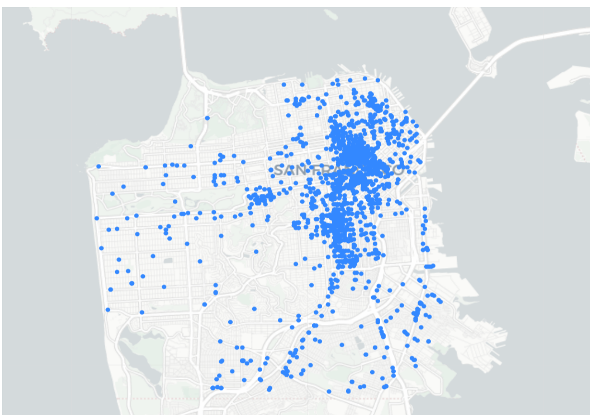
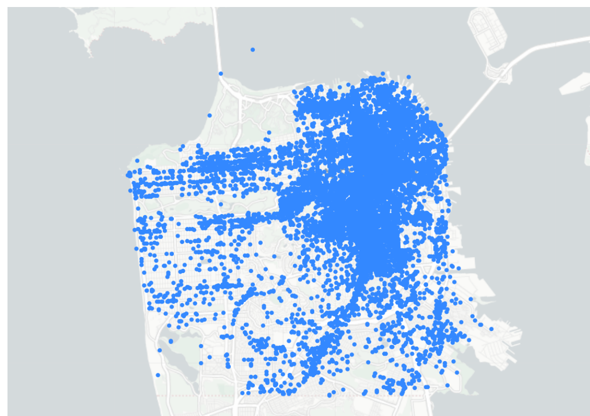
Jamie Lim	Data preprocessing, EDA, model outcome visualization
Shun Tomita	Model formulation, model building
Dai Ling Wu	Data collection, model building, outcome visualization
Lisa Yeung	Data collection, model building, outcome visualization

I. Problem Statement

There was an increasing number of human and animal waste cases reported to the 311 hotlines in the city of San Francisco. In 2012, there were approximately 6,500 cases reported and it rose to a significant number of over 30,000 cases in 2019[1].

To address the growing concerns of public hygiene and to meet the demand of public toilets, the city started the Pit Stop public toilet program in 2014, which aims to provide clean public toilets for everyone, and provide a safe spot to dispose of used needles and bags to dispose dog's waste.

However, Despite the installation of Pit Stops, the city is still experiencing a continued public hygiene problem. This is evidenced by the significant increase in the number of human and animal waste cleaning 311 cases over the last decade as shown below.

2011	2021
	

To meet the increasing need for public toilets, in addition to the already installed 31 Pit Stops, we want to find out the optimal number and the location to install additional Pit Stops in the San Francisco area. To achieve this, we focus on areas that are likely to have high demand for public toilets. We believe that installing Pit Stops in these areas will improve public hygiene and ultimately public utility. In identifying the candidate locations and the number of Pit Stops, we hypothesize the following:

1) Areas with many 311 cleaning cases are unhygienic and thus installing Pit Stops in these areas will increase public utility.

2) Areas with large homeless populations are susceptible to human waste and thus installing Pit Stops in these areas will increase public utility.

---

[1] Ting, E. (2021, July 14). *Don't look now but SF's poop problem seems to be getting better*. SFGATE. Retrieved September 12, 2022, from https://www.sfgate.com/bay-area-politics/article/San-Francisco-poop-problem-stats-streets-feces-new-16311073.php

II. Data Summary

1. SF311 Cases dataset[2]

With regard to our two hypotheses, we mainly used the SF311 Cases dataset which is provided by the San Francisco city government. The dataset contains information about each 311 case collected since 2008. The variables of our interest are 'CaseID', 'Opened', 'Category', 'Request Type', 'Request Details', 'Latitude', and 'Longitude'. The coordinate data is important as we are focusing on optimizing geographical locations for additional Pit Stops. We will only use cases from 2020 to 2022 to better focus on the recent data.

In our project, we are interested in human and animal waste cleaning cases as well as encampment-related cases which we will use as a proxy for the homeless population. We created two subsets by filtering on 'Opened', 'Category', and 'Request Details.' As a result, we obtained 'Cleaning_request_dataset.csv' and 'Encampments_dataset.csv' from the original 311 cases dataset '311_Cases_1121.csv.'

	CaseID	Opened	Category	Request Type	Request Details	Latitude	Longitude	Opened_year
0	14606209	11/12/2021 03:19:00 PM	Street and Sidewalk Cleaning	Human or Animal Waste	Human or Animal Waste	37.718404	-122.398264	2021
1	14581748	11/07/2021 08:00:00 AM	Street and Sidewalk Cleaning	Human or Animal Waste	Human or Animal Waste	37.796925	-122.408737	2021
2	14583313	11/07/2021 01:47:00 PM	Street and Sidewalk Cleaning	Human or Animal Waste	Human or Animal Waste	37.796371	-122.436874	2021
3	14597742	11/10/2021 04:11:00 PM	Street and Sidewalk Cleaning	Human or Animal Waste	Human or Animal Waste	37.755657	-122.414314	2021
4	14604192	11/12/2021 09:52:00 AM	Street and Sidewalk Cleaning	Human or Animal Waste	Human or Animal Waste	37.708672	-122.405441	2021

2. Existing Pit Stops dataset[3]

In addition, we manually retrieved the coordinate data of 31 existing Pit Stops from Google Map and created the 'Existing_Pit_Stop_Locations.xlsx' dataset (SF's Public Works' website mentions 33 Pit Stops currently in operation but we were able to obtain location data for only 31 Pit Stops from their google map link) We need to take into account the existing Pit Stop locations because we are interested in the optimal number of all Pit Stops in an area which includes the existing and the newly installed Pit Stops.

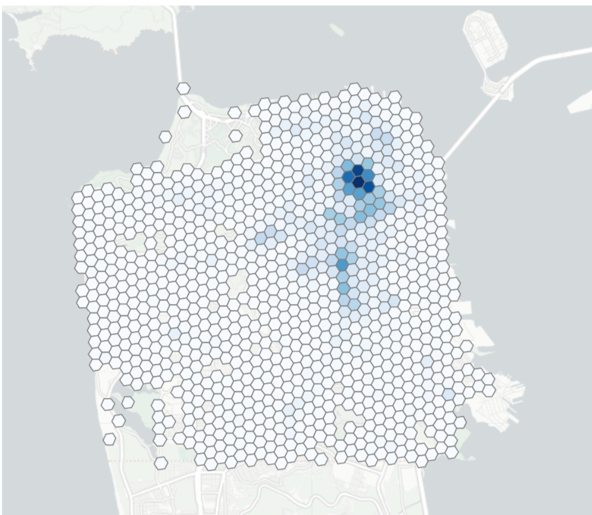
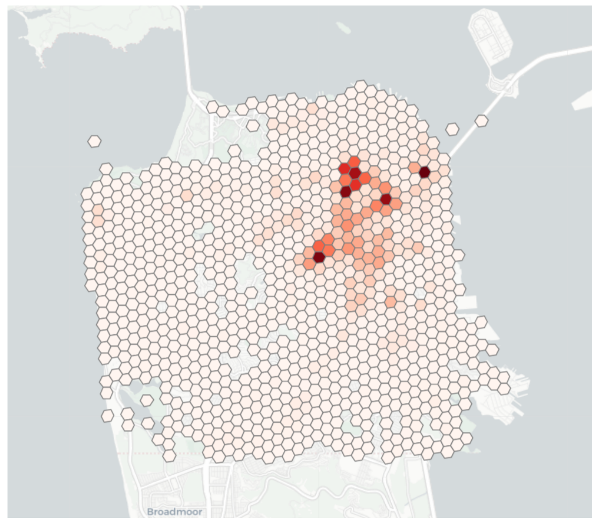
	Name	Latitude	Longitude
0	Castro Pit Stop	37.769203	-122.428220
1	Haight Ashbury Pit Stop	37.774078	-122.452130
2	Haight Pit Stop	37.771040	-122.440369
3	Mission Pit Stop	37.758046	-122.417815
4	Mission Pit Stop	37.749694	-122.411635

[2] https://data.sfgov.org/City-Infrastructure/311-Cases/vw6y-z8j6
[3] San Francisco Pit Stop Locations - Google My Maps

III. Exploratory Data Analysis

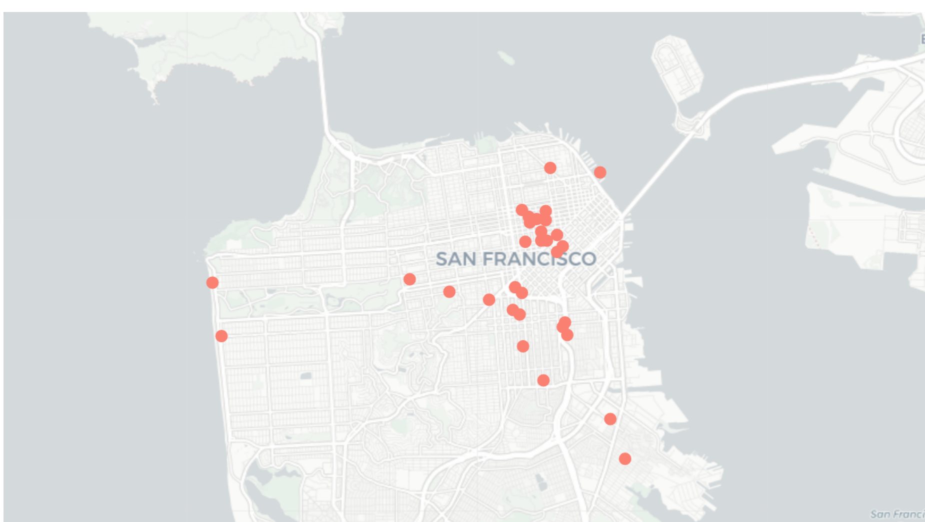
1. 311 Cases of our Interest

We plotted heatmaps to visualize the geographical distribution of the 311 cases of our interest. The number of cases are normalized and the darker the color, the more cases in that area. We notice that both types of cases are concentrated in the northeastern parts of San Francisco.

Human and Animal Waste Cleaning 311 Cases (2020-2022)	Encampment-related 311 Cases (2020-2022)
	

2. Existing Pit Stop Locations

We also plotted the locations of the existing 31 Pit Stops. We see that their locations show a similar pattern with the darker areas in the heatmaps above.



IV. Model Formulation

To solve the Pit Stops allocation problem, we implemented the Integer Linear Programming (ILP) problem by dividing the San Francisco region into 20 x 30 grids and tried to find the optimal number of Pit Stops in each cell. Each cell is as large as 8 blocks in practice and it would take 10 minutes to walk from one end to the other.

1. Objective function

Our objective function in the ILP formulation is utility function. According to the hypotheses mentioned above, our utility function is a function of uncleanliness and susceptibility and our model computes the optimal number of Pit Stops that maximizes the utility function. Since marginal utility diminishes and our utility function needs to be linear in terms of our decision variables, a piecewise linear function is used to implement concave utility function with three different slopes, which is equivalent to the minimum of three different linear utility functions as shown in ***Figure 1***.

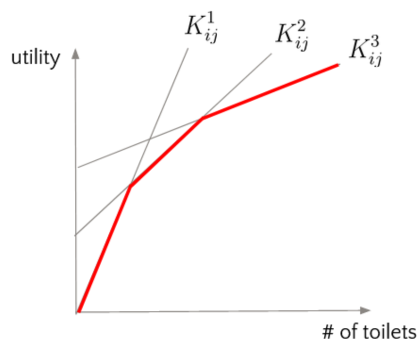


Figure 1. Our objective function is defined as the minimum of three different linear functions, which enables us to model concave objective function without any nonlinear function.

Weights and intercepts in these linear utility functions reflect assumptions about value systems and the utility function, such as how much uncleanliness and susceptibility would affect utility of people and communities in San Francisco and how quickly their marginal utility diminishes. The model also assumes the spillover effects that a Pit Stop in area i,j improves the utility in areas contiguous to area i,j.

2. Constraints

A main constraint in our model is a budget constraint. The city government allocates $8.6 million for Pit Stops projects yearly[4], and our model considers two types of expenditure: installation costs and maintenance costs. The model assumes that the installation cost is $100 thousands per Pit Stop[5]. Also, dividing the San Francisco region into 100 cells, the maintenance cost is $60

[4] Brinklow, A (May 31, 2019) *SF pledges $8.6 million for more toilets*. CURBED San Francisco. Retrieved December 14, 2022, from https://sf.curbed.com/2019/5/31/18647600/san-francisco-toilets-poop-pit-stop-budget-breed

[5] Controller's Office Geospatial Analysis of Accessibility to Public Toilets, *Pit Stop Pilot Program Expansion.* Retrieved December 14, 2022, from

thousands per cell with at least one Pit Stop[6]. The rationale behind this calculation is that the city government needs to hire one person to clean up all toilets in each cell.

Also, the model has upper bound constraints for the number of Pit Stops in each area and each contiguous 9 areas, and also has lower bound constraints assuming the city government will not change the location of existing Pit Stops. The detailed mathematical formulation of our model is in the appendix.

V. Model Implementation and Interpretation

In our objective function, there are three tunable hyperparameters, a, b and e. We next take a deep dive into the optimization results of pit stop allocation with several combinations of hyperparameter settings.

1. Weights

For the weight setting, we explore four types of scenarios with fixed intercepts: weights only on uncleanliness, weights only on susceptibility, higher weight on uncleanliness and higher weight on susceptibility. According to the graphs of optimization results below, we can conclude that the model allocates pit stops differently when the weights change. Some areas would only be installed with pit stops when certain extreme weight is applied. We can compare the results of two extreme weights by overlaying the two graphs: blue represents the optimization result from the model with weights only on uncleanliness, red is the result from the model with weights only on susceptibility.

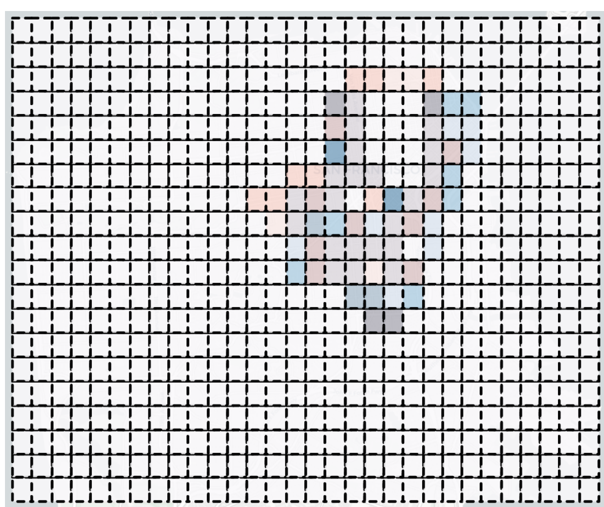
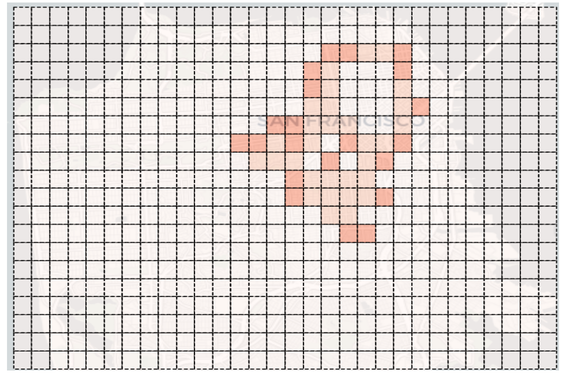
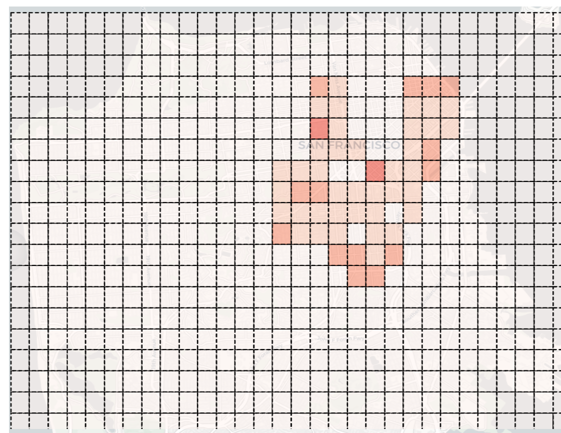
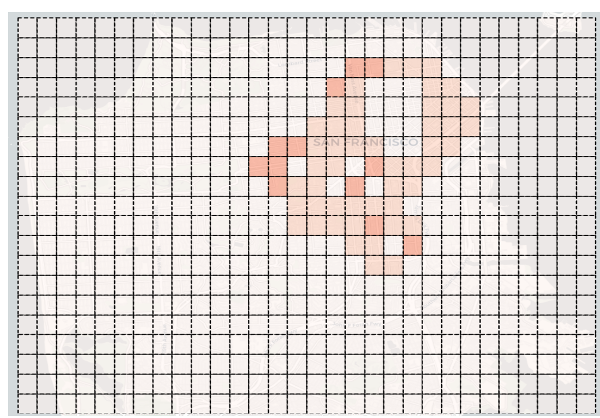
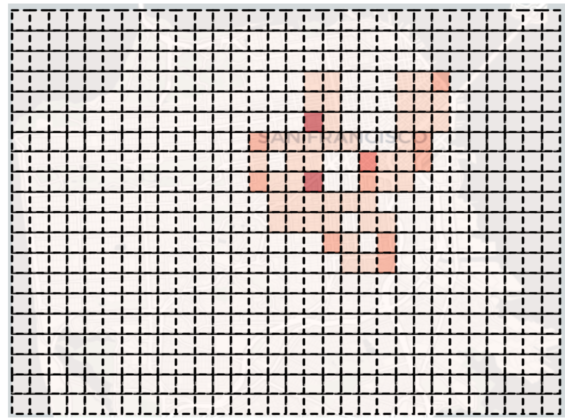


Figure 2. Overlay of Two Extreme Cases

https://sfcontroller.org/sites/default/files/FileCenter/Documents/6948-Pit%20Stop%20Pilot%20Program%20Expansion%20-%20Geospatial%20Analysis%20of%20Accessibility%20to%20Public%20Toilets.pdf

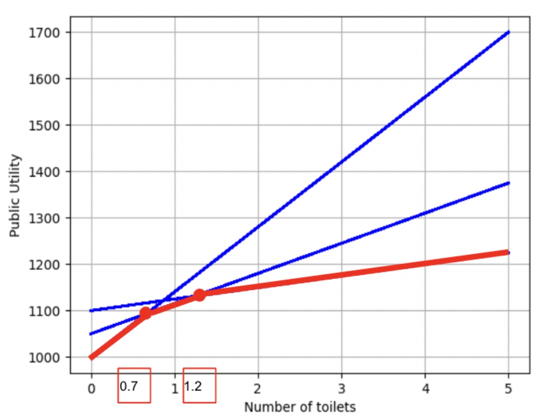
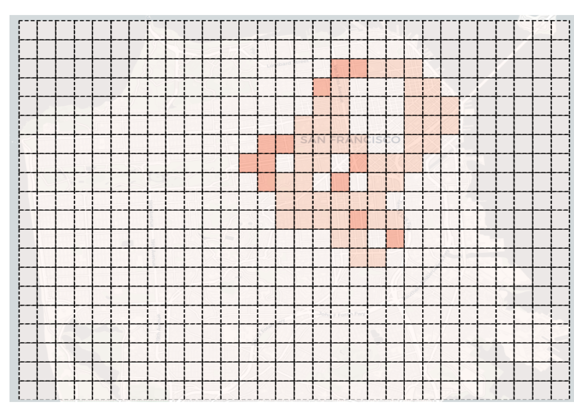
[6] Ibid.

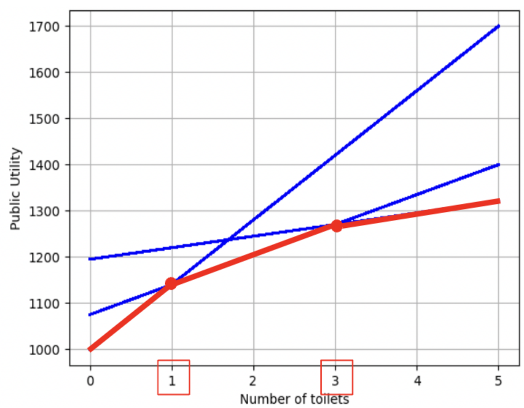
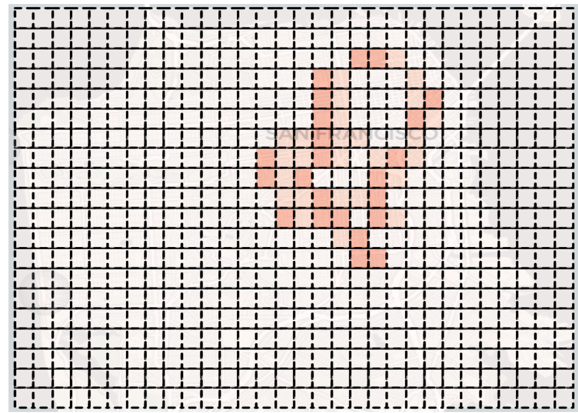
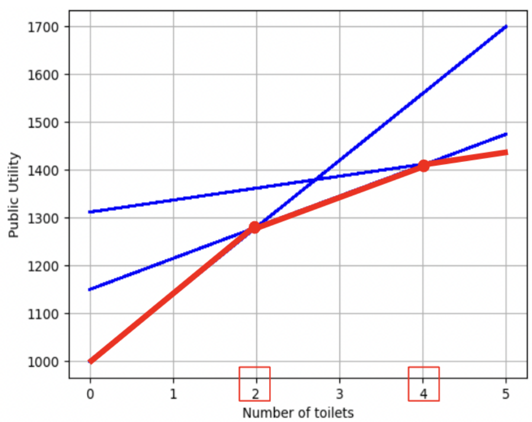
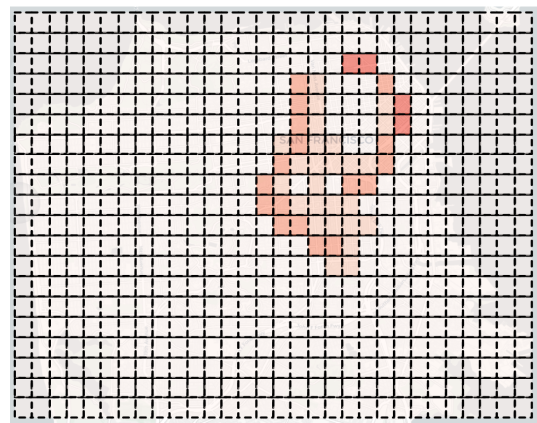
	Hyperparameter setting						Model Result
Weights only on uncleanliness	a	100	b	0	e	1000	
	a’	50	b’	0	e’	1050	
	a”	20	b”	0	e”	1100	
Weights only on susceptibility	a	0	b	100	e	1000	
	a’	0	b’	50	e’	1050	
	a”	0	b”	20	e”	1100	
Higher weight on uncleanliness	a	100	b	40	e	1000	
	a’	50	b’	15	e’	1050	
	a”	20	b”	5	e”	1100	

Higher weight on susceptibility	a	40	b	100	e	1000	
	a'	15	b'	50	e'	1050	
	a''	5	b''	20	e''	1100	

2. Intercept

To explore how concavity of utility function affects optimization of pit stops installation, we investigate three types of scenarios: quickly, moderately and slowly diminishing marginal utility. We first implement quickly diminishing marginal utility in our model by adjusting the intercepts. As shown in the first graph of utility function below, the utility functions intersect at 0.7 toilets and 1.2 toilets, which means that the public utility diminishes quickly after these two points. Therefore, as we can see in the first graph of the model result, our model installs two toilets in twelve grids. Next, to implement moderately diminishing marginal utility, intercepts are set to match the intersection point at one and three, meaning the public utility diminishes quickly after one and three toilets. In the optimization result, there are twenty-three grids installed with two toilets. Last, with relatively slowly diminishing marginal utility, the intersection point of utility functions are at two and four, meaning the public utility diminishes quickly after two and four toilets. The model result shows there are seventeen grids installed with two toilets and four grids installed with three toilets. Therefore, we can conclude that the more slowly the marginal utility diminishes, the larger the possible maximum number of pit stops in one grid that our model installs will become.

	Utility Function	Model Result
Quickly diminishing marginal utility		

Moderately diminishing marginal utility		
Slowly diminishing marginal utility		

VI. Policy Implication and Recommendation

Our model is robust to changes in hyperparameters: 1) the concavity of utility function and 2) the weights assigned to uncleanliness and susceptibility. The San Francisco Department of Public Works should take a set of hyperparameters according to the urban planning philosophy. The city adopted "The Better Street Plan" in 2010. The plan seeks to balance the needs of all street users, with a particular focus on the pedestrian environment and how streets can be used as public space.

In the case of installing Pit Stops around the city, we have identified the stakeholders, their specific needs and the relevant hyperparameters in the following table:

Stakeholders	Needs	Hyperparameters
City Cleaning Crews	Reduce working hours on street cleaning	Allocate higher weights on uncleanliness
Pedestrians and	Accessible roads without	More concave utility function

Bikers	disruptions	
Residents	Improvement in neighborhood	* Sufficient number of toilets
Small Business	Business friendly neighborhood	* Sufficient number of toilets
Homeless Community	Enough toilets to use with dignity	Allocate higher weights on Susceptibility

** indicates not a hyperparameter*

For City Cleaning Crews, if the model attends more weights on uncleanliness, the model would allocate more toilets in areas with higher number of 311-call requests. Similarly, if the model attends more weights on susceptibility, the model would allocate more toilets in areas with higher homeless populations. This might sound like a competing weights allocation but from the model output, we can see there is a high proportion of overlapping areas even with extreme weights allocation cases as shown in ***Figure 2.***

For pedestrians and bikers, they would like to walk and bike without sudden disruptions. Therefore, the model should have a rapidly diminishing utility function that would allow less concentration of toilets in one grid. This would also reduce the space usage in one grid by Pit Stops and give more space for other construction that serves bikers and pedestrians, such as bike racks and public art.

For residents and small businesses, they would have similar needs, which are a better environment that is welcoming. They would not want to have human waste in their neighborhood but at the same time they do not want an excess number of toilets. To have just the sufficient number of toilets is handled by the model design, where toilets are located according to demand and are bound by budget constraints. Therefore, the department does not need to tune hyperparameters for this purpose.

Aside from utilizing hyperparameters tuning to achieve urban planning policy goals, we also suggest the department to keep track of and to evaluate the usage of each Pit Stop from time to time.

VII. Limitations and Future Work

We would like to extend our model in the following three directions.

1) Incorporate different types of Pit Stop
   There are two types of Pit Stop, one is sponsored by the media company JCDecaux, where the sponsored Pit Stop will be used as a billboard for print advertising. The annual cost for this type of toilet is $0. The other type, which is the “Portable”, is the only type of toilet we are considering in the model.

	JCDecaux	**Portable**
Estimated Annual Cost	$0	$110,000
Staffing Costs (8 hours)	$60,000	$60,000

Since the number and location of Pit Stop that will be sponsored by JCDecaux is a business decision, we cannot hypothesize a deterministic criteria and incorporate that into our model. In the future, the department of Public Work can communicate with JCDecaux and understand their selection criteria and incorporate that into the model.

2) <u>Incorporate maintenance optimization into overall planning</u>
   In our model, the maintenance cost is rather simplistic as we assign 1 person to staff a region if $Y_{p,q} = 1$. However, we could also create an optimized cleaning route to minimize maintenance costs if we have data such as the number of city crews and the number of cleaning vehicles, etc.

3) <u>Extends over one time stamp</u>
   Currently, our model only optimizes the allocation of Pit Stops in one time stamp given an annual budget. If we can extend the model over time and consider the installation cost of each Pit Stop as an initial fixed cost and incorporate maintenance cost as a variable cost, our model would be able to find the optimal number of toilets given budgets over multiple years.

4) <u>Evaluation of our model's performance</u>
   Our model is based on the hypothesis that installing Pit Stops in areas with many 311 cleaning cases and areas with large homeless populations improves public utility. However, our hypothesis might not always be true and it is possible that Pit Stops installed in these areas are not used as much as expected. We can measure the usage of newly installed Pit Stops to improve our model.

Appendix

1. Notations in our mathematical formulation
   $K_{ij} =$ utility in area i, j
   $X_{ij} =$ the number of Pit Stops in area i, j (existing + new)
   $U_{ij} =$ the average of the number of human and animal waste cleaning cases in 311 calls in 9 contiguous blocks
   $S_{ij} =$ the average of the number of encampment-related cases in 311 calls in 9 contiguous blocks
   $Y_{pq} = 1 \, if \, X_{ij} \geq 1$ in at least one of areas that belong to region p, q,
   $0 \, otherwise$
   $a, a', a'', b, b', b'', e, e', e''$: coefficients and intercepts of utility function
   P_{ij} = # of existing pit stops in area i, j
   $M =$ some large number

2. Mathematical formulation of our model

$$\max_{K_{ij}, X_{ij}, Y_{pq}} U = \sum_{i,j} K_{ij}$$

$$s.t. \, K_{ij} \leq aX_{ij}U_{ij} + bX_{ij}S_{ij} + e \quad \forall i, j,$$

$$K_{ij} \leq a'X_{ij}U_{ij} + b'X_{ij}S_{ij} + e' \quad \forall i, j,$$

$$K_{ij} \leq a''X_{ij}U_{ij} + b''X_{ij}S_{ij} + e'' \quad \forall i, j,$$

$$(\sum_{i,j} X_{ij} - \sum_{i,j} P_{ij}) * \$100K + \sum_{p,q} Y_{pq} * \$60K \leq \$8.6M,$$

$$M * Y_{pq} \geq \sum_{i=2p}^{2p+1} \sum_{j=3q}^{3q+2} X_{ij} \quad \forall p, q,$$

$$X_{ij} \leq 5 \quad \forall i, j,$$

$$\sum_{l=i-1}^{i+1} \sum_{m=j-1}^{j+1} X_{lm} \leq 13 \quad \forall i, j,$$

$$X_{ij} \geq P_{ij} \quad \forall i, j$$