Dear members of the California Farm Bureau Federation,

We have created a report that provides an extensive overview of waste collection approaches in San Francisco County. We will give details on our model, list all relevant assumptions, and provide our analysis below.

As you might know, the problem of efficient waste collection is complex. The optimal solution suggests that the truck (or a few of them) needs to pick up waste from all farms without visiting any farm twice and return to the starting point. This reminded us of the Travelling Salesman Problem, where the order of the visited locations is unimportant, and all solutions try to minimize the traveling distance (Laporte, 1992).

In this report, we would like to narrow the scope of the problem down by applying two waste collection strategies and observing their impact on several key metrics related to the process.

Many factors can directly or indirectly influence waste collection efficiency. In this simulation, we have varied the number of farms, the maximum waste (maximum amount of waste on a farm as a proportion of the truck capacity), and the number of drop-off sites to see how they affect the working time, drop-off site visits, driving time, and the amount of fuel used. We believe that all outcome metrics could help you gain better insights into the design of a new waste collection network and the appropriate waste collection routes and schedules.

Model Assumptions and Design

Based on our research and system requirements, we have developed a set of rules and assumptions incorporated into the model design.

Rules:

- The average refuse truck speed is 16 miles/hour (Sandhu et al., 2015).
- The average fuel economy is 4.4 miles/gallon (0.23 gallons/mile) (Sandhu et al., 2015).

- The average tank size of a refuse truck is 50 gallons (Kelly, 2022).
- If the refuse truck is full, it needs to go to the drop-off site, dump the waste, and continue operating.
- The refuse truck stops operating once the waste from all farms has been collected.
- Headquarters is the final destination at the end of every working day.
- The truck visits every farm once. If the next farm on the list has more waste when available in the refuse truck, it needs to go to the drop-off site and then head towards the next farm.
- The amount of waste on a single farm cannot be larger than the truck capacity.
- If the truck does not have enough fuel to go to the next farm and return to headquarters to refuel, it needs to go to the headquarters first and then proceed according to the collection route.

Assumptions:

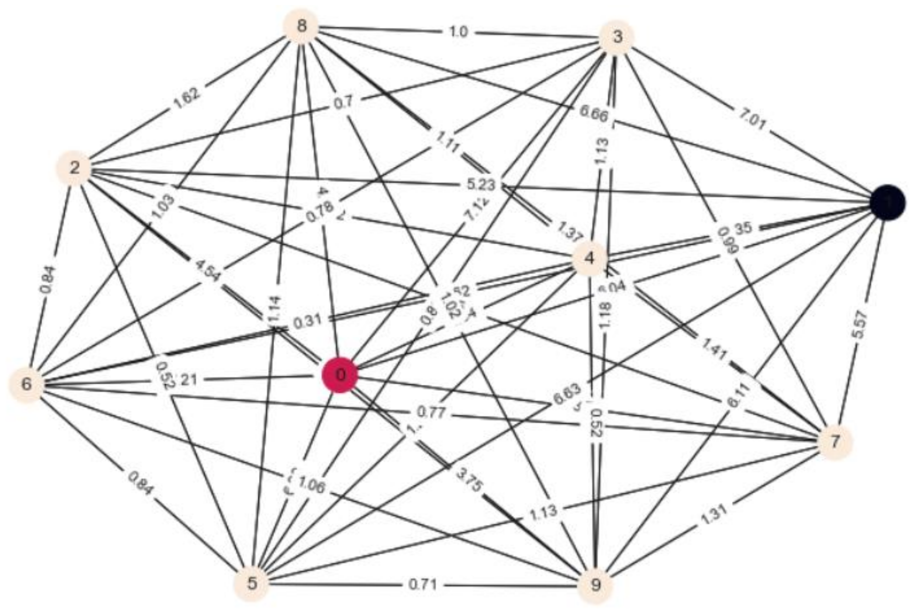
- It takes 20 minutes to fill up the refuse truck with waste that equals its full capacity (Based on our observations, it takes 5 minutes on average to empty the trash cans standing near apartments. Assuming the amount of waste produced on the farms, the value should be higher).
- It takes 10 minutes to unload the refuse truck with waste that equals its full capacity.
- Values from the truncated normal distribution determine the distances between the farms with bounds of 0 and 2 miles (min/max distance), a mean of 1 mile, and a standard deviation of 0.3 miles.
- The amount of waste on every farm is determined by values from the uniform distribution with bounds of 0 and 1 (as a proportion of truck capacity).
- The network represents a complete graph meaning that all farms, drop-off site(s), and headquarters are connected.

After finalizing all the rules and assumptions, we have implemented the model in Python using an Object-Oriented Programming approach. The simulation has three core classes. The first one is called **Truck**. It is a helper class that implements the refuse truck functionality with all relevant attributes. The class contains various methods that help calculate and update four metrics of interest considered dependent variables in the simulation. **WasteRemovalSimulation** is the main class in the simulation. It initializes the network and all related components, runs every step of the waste collection, and visualizes the system's state. **Test** class allows the users to vary multiple parameters in the system (designed in a way where the users can modify the values one at a time) and visualize the outputs through graphs and histograms.

Attempted strategies

The next step after model design and implementation consisted of multiple tests to check if all simulation aspects were working as expected. During the variation of the three parameters mentioned above, we displayed the results for two waste collection strategies — random and greedy. Following the random strategy, the refuse truck chooses a random farm to visit in every simulation step. Since we model a network as a complete graph in this scenario, the truck can go to any farm (or other destinations if necessary) from the current location. The second strategy is based on the greedy approach. It chooses the next farm to visit based on the shortest distance from the current one (Abdulkarim & Alshammari, 2015). The greedy algorithm is considered suboptimal because it makes the best decision based on the current state. Another potential approach is brute force. It suggests iterating through all possible waste collection routes and choosing the shortest one. However, the time complexity of this strategy is O(n!), where n is the number of nodes. Thus, it is not suitable for any networks besides very small ones.

We started the tests by running the simulation once with a random strategy to see whether the outcomes made sense (see Fig. 1).




```
Total waste collected: equivalent to 1.94 trucks.
Total amount of gas used: 14.67 gallons.
Truck was operating for 3.64 hours (including 2.78 hours of driving).
Truck visited the drop-off site 3 times.
```

Fig. 1. Waste collection network with the application of the random strategy (other model parameters included: n=10, max_waste=0.5, dropoff=1) and four key metrics based on 1 iteration. The red node is the headquarters; the black node indicates the drop-off site, rest of the nodes are farms. The edges values indicate the distances between 2 nodes.

The results seemed plausible given the low number of farms in the system and truck specifications. Thus, we proceeded by experimenting with the number of farms in the network.

Number of farms

We ran the simulation with the number of farms ranging from 3 to 30 (100 trials for each value; the trial finished when all waste was collected). Looking at the outputs (see Fig.2), we notice the positive correlation between the number of farms and every outcome variable. This is expected because as the number of farms grows, the truck spends more time picking up the waste, driving, and, as a result, spends more gas.

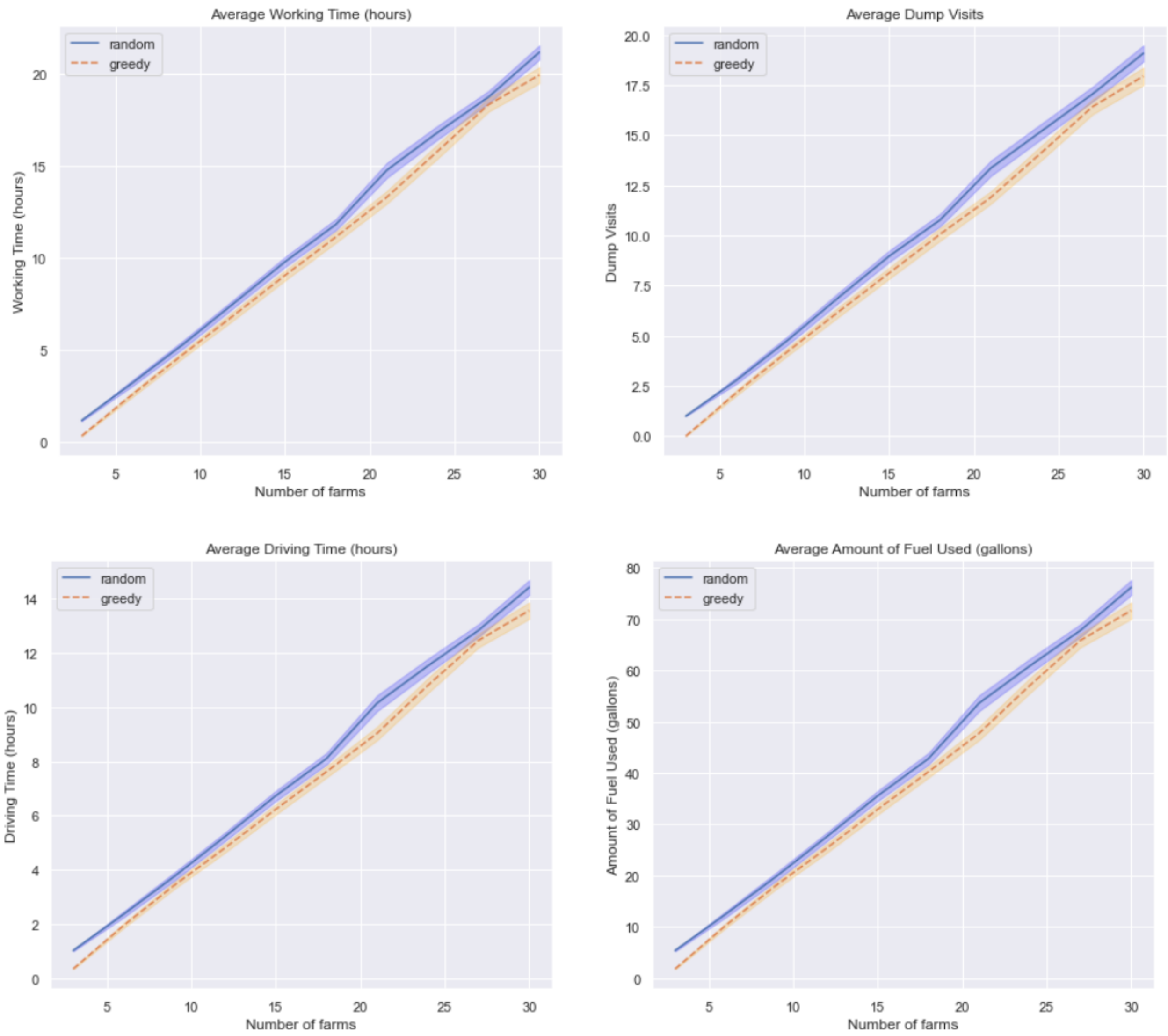


Fig. 2. Working time, dump visits, driving time, and amount of gas used based on the varying number of farms. The solid line indicates a random strategy; the dashed line shows the greedy approach.

The dominance of the random greedy strategy over the random strategy is visible on the graphs and by looking at the first set of tables (see Appendix A) containing average values for every outcome variable with the corresponding 95% confidence intervals. The most significant difference is the amount of used fuel. The greedy approach saves around 5 gallons of fuel when the number of farms in the network equals 30. Generally, the difference between the strategies becomes bigger for all key metrics once there are more farms in the network. Since the number of farms can realistically be even higher, the greedy approach seems more efficient for waste collection. The number of working and driving hours is also worth mentioning. We can observe how the average working time exceeds the typical working day of 8 hours once the number of farms in the network reaches 15. Thus, the waste collection company management should consider this data. For instance, they can hire more drivers and have multiple shifts or allocate more trucks for the given system so that they operate simultaneously.

Maximum waste amount

In our next experiment, we varied the upper bound of the maximum waste on a farm. Since we previously agreed that the amount of waste on a single farm is equal to or less than the capacity of one refuse truck, we varied the values from 0.1 to 1 as proportions of truck capacity (see Fig. 3).

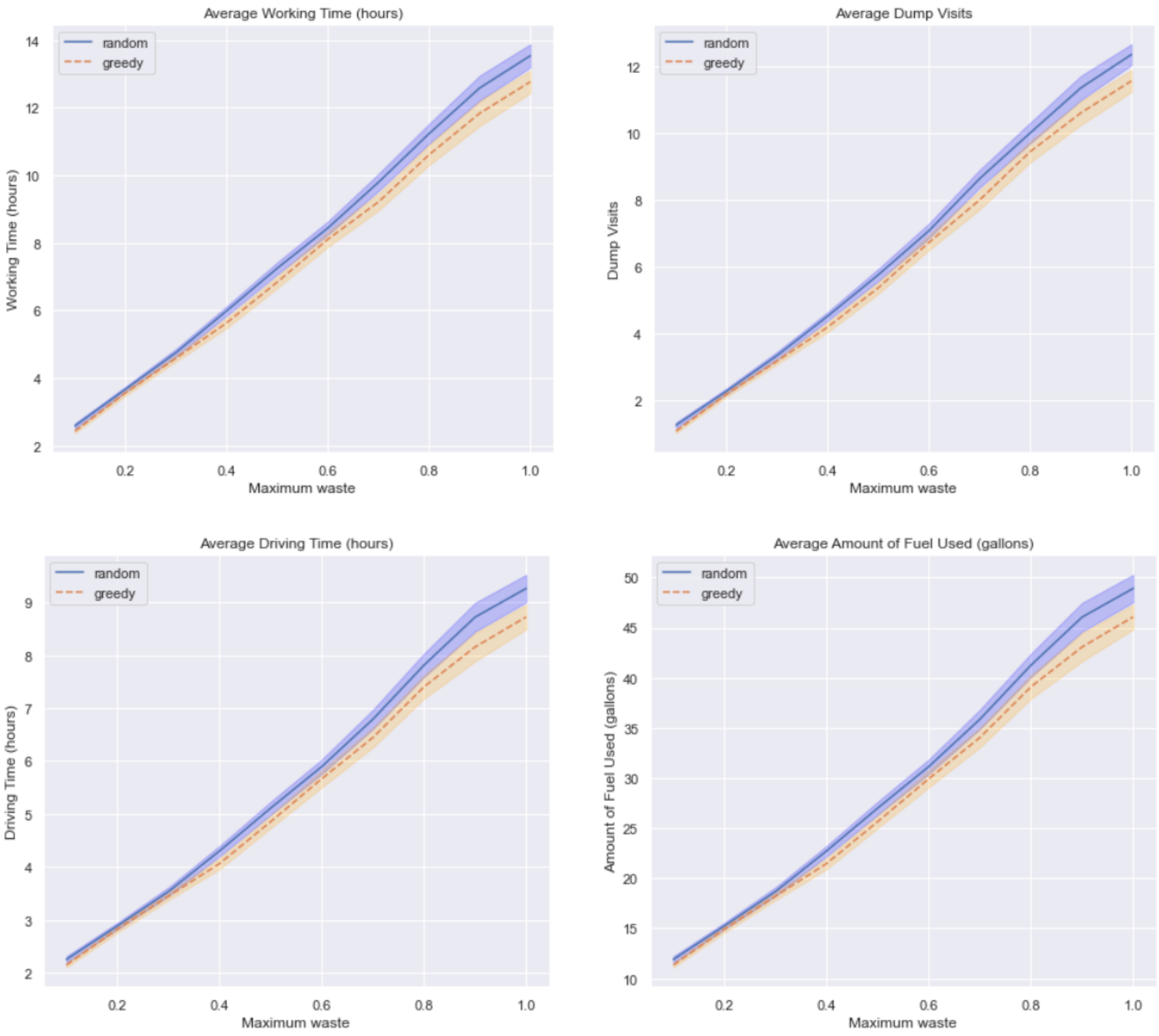
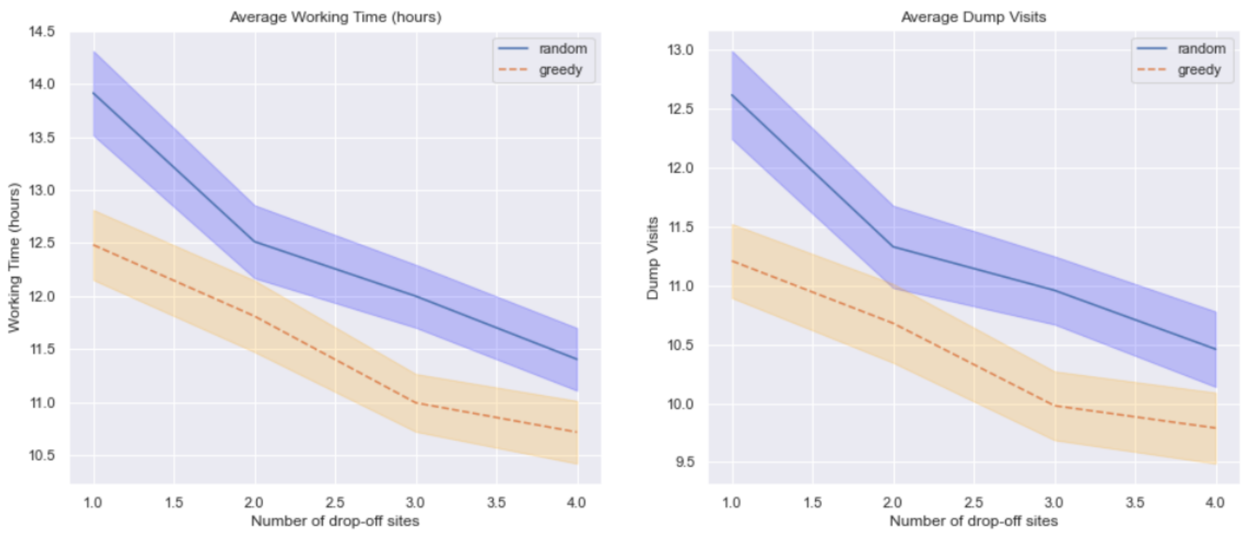


Fig. 3. Working time, dump visits, driving time, and amount of gas used based on the varying the maximum amount of waste. The solid line indicates a random strategy; the dashed line shows the greedy approach.

We notice the same pattern as in the previous experiment — the greedy strategy is better with respect to every relationship represented on the four graphs above. Compared to the previous experiment, the outcomes contain higher variability and lower certainty. As a result, the confidence intervals on all graphs are wider than those on their counterparts in Fig. 2. However, the data from the greedy approach exhibits a slightly higher certainty than the data from the random strategy (see Appendix B). While we varied the waste amounts, other parameters in the simulation remained default (20 nodes in the network, one drop-off site). Comparing the values from the tables, we notice that the difference between strategies is smaller. Thus, we conclude that the outcomes are more robust to the changes in the waste amount than to the changes in the number of farms. This makes sense in real-world scenarios because adding a new farm has a higher impact on the network of farms than the amount of waste on every farm. Additional connections (roads) are created between the new farm and the existing ones, which affects the network structure.

Number of drop-off sites

In the third experiment, we tested different values of drop-off sites, while the number of farms and the maximum waste amount remained default (20 and 1, respectively). The outputs (see Fig. 4) show a negative relationship between the number of drop-off sites and the outcome variables. This aligns with the expectations because the increasing number of places to unload waste indicates that the refuse truck needs to make fewer trips to and from the drop-off sites, which saves time and gas.



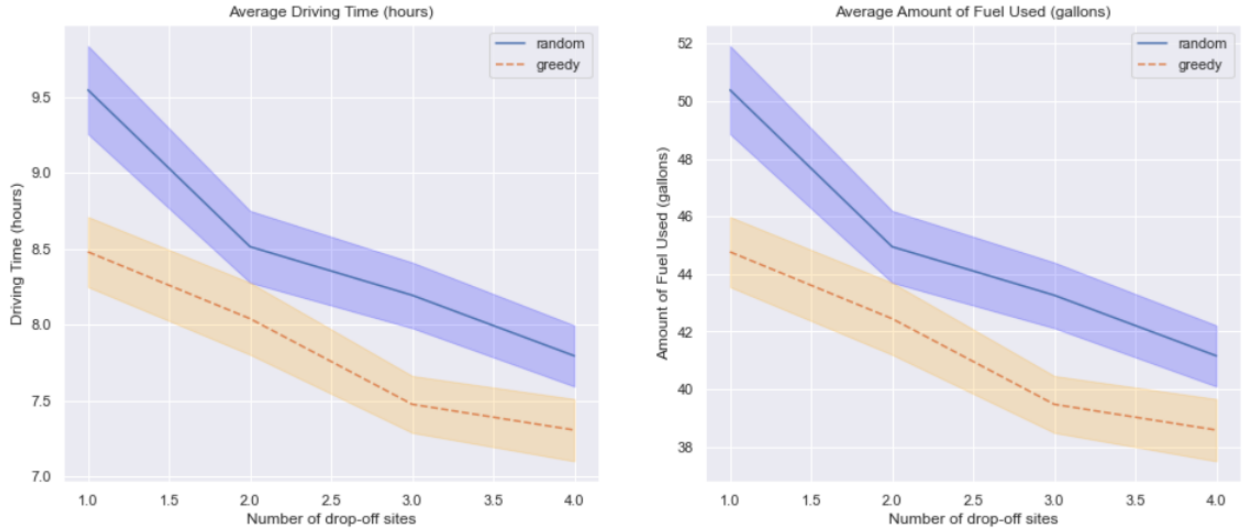


Fig. 4. Working time, dump visits, driving time, and amount of gas used based on the varying the number of drop-off sites. The solid line indicates a random strategy; the dashed line shows a greedy approach.

The superiority of the greedy strategy is the most noticeable here out of all experiments. However, the outcomes also have the highest level of uncertainty in this case, which makes the confidence intervals so wide (see Appendix C). To increase the certainty about the outcomes, we recommend running the simulations with different parameter variations for a bigger number of trials. We believe the intervals will become thinner once the difference in trials is within at least one order of magnitude (e.g., 1000 trials instead of 100 trials used in the three experiments above).

Result distributions

Finally, we ran the simulation 1000 times (all parameters were default) and examined the differences in the distributions for both strategies (see Appendix D). The distribution tails were thicker for all outcome variables in the case of using the greedy approach. This shows that the greedy strategy is suboptimal and can sometimes lead to worse results than the random strategy. Additionally, the support of distributions was also smaller for working time, driving time, and the fuel used. While this might indicate lower outcome variability, we recommend running more tests with a bigger number of trials to prove the statement.

Conclusion

To qualitatively verify the dominance of the greedy approach over the random one, we compared the average values for all metrics of interest after 1000 simulation iterations (see Fig. 5). The greedy strategy is around 6% more efficient than the random strategy based on the measured parameters.

Table 1: Mean values for all metrics of interest.

Metric	Random Avg	Greedy Avg
Working Time (hours)	13.62	12.77
Dump Visits	12.43	11.49
Driving Time (hours)	9.32	8.7
Amount of Fuel Used (gallons)	49.22	45.95

Table 2. Comparison of 2 strategies for all metrics of interest.

Metric	Greedy over Random (%)
Working Time (hours)	6.2
Dump Visits	7.57
Driving Time (hours)	6.64
Amount of Fuel Used (gallons)	6.64

Fig. 5. Set of tables with the average values for working time, dump visits, driving time, and amount of gas under two strategies. The ratio in the second table shows how much better the greedy approach is compared to the random approach.

Thus, we recommend California Farm Bureau Federation consider our findings when partnering with the waste collection company to organize the workflows. Assuming the network of around 20 farms and the maximum waste amount equal to the truck capacity, we suggest hiring at least two drivers who will either work during different shifts or utilize two refuse trucks simultaneously (depending on the resources and the work schedule of the waste collection company). We also believe that setting up a second drop-off site will be beneficial for the waste collection happening in the area with at least 15-20 farms. If California Farm Bureau Federation decides to combine the existing farm districts into larger areas, more than two trucks and two drop-off sites are required for waste collection (assuming that it happens every day). The waste collection workers should follow the greedy approach. Thus, we recommend California Farm Bureau Federation provide the data about the distances between all farms to the waste collection company to optimize the collection route.

Numerous tests proved the model setup and the simulation results to be valid. The main reasons are the realistic rules, assumptions, and observations (some of them are backed up by real data). The outcomes make sense when we assess them through the lenses of everyday life. To improve the model further, we suggest incorporating multiple trucks into the model to determine their impact on the overall efficiency. Running the simulation for a long time (e.g., a week instead of finishing when all trash is collected) can uncover more sophisticated relationships since one would need to incorporate the fact that the amount of waste renews every day. Another idea is to experiment with the distributions for waste allocation and distance values. For instance, the practical evidence might point out that the log-normal distribution is more suitable for modeling the distances between the farms than the modified version of the truncated normal distribution we have used. Finally, testing the model on a random graph (e.g., Erdos-Renyi) where not all farms are connected to one another will lead to the development of a waste collection strategy that is a better fit for the given scenario and more resilient to the system changes.

References

Abdulkarim, H., Alshammari, I., (2015). Comparison of Algorithms for Solving Traveling Salesman Problem. *International Journal of Engineering and Advanced Technology* 4(6).

Kelly, K. (2022). *Alternative fuel trucks case studies: Running refuse haulers on compressed natural gas*. Academia.edu. Retrieved from https://www.academia.edu/73284007

Laporte, G. (1992). The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, *59*(2), 231–247. https://doi.org/10.1016/0377-2217(92)90138-y

Sandhu, G., Frey, C., Bartelt-Hunt, S., & Jones, E. (2015) In-use activity, fuel use, and emissions of heavy-duty diesel roll-off refuse trucks. *Journal of the Air & Waste Management Association*, *65*(3), 306-323.

10.1080/10962247.2014.990587

Appendices

Appendix A. Means and 95% confidence intervals of 4 key metrics based on the varying number of farms.

Table 1. Outcome variable: Working Time (hours). Varied parameter: Number of farms.

Number of farms	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
3	1.18	0.03	0.34	0.01
6	3.22	0.16	2.6	0.13
9	5.33	0.22	4.79	0.18
12	7.57	0.22	6.92	0.24
15	9.81	0.27	9.07	0.27
18	11.84	0.31	11.16	0.3
21	14.79	0.41	13.32	0.34
24	16.84	0.37	15.82	0.38
27	18.77	0.34	18.35	0.37
30	21.2	0.39	19.97	0.43

Table 2. Outcome variable: Dump Visits. Varied parameter: Number of farms.

Number of farms	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
3	1.0	0.0	0.0	0.0
6	2.79	0.16	2.17	0.13
9	4.77	0.22	4.21	0.19
12	6.9	0.25	6.19	0.24
15	8.97	0.28	8.13	0.28
18	10.77	0.31	10.06	0.32
21	13.36	0.38	11.9	0.31
24	15.21	0.38	14.14	0.36
27	17.05	0.37	16.42	0.38
30	19.08	0.4	17.94	0.43

Table 3. Outcome variable: Driving Time (hours). Varied parameter: Number of farms.

Number of farms	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
3	1.02	0.02	0.34	0.01
6	2.36	0.11	1.95	0.09
9	3.76	0.15	3.43	0.13
12	5.23	0.16	4.81	0.16
15	6.73	0.19	6.23	0.19
18	8.11	0.21	7.62	0.21
21	10.16	0.29	9.04	0.25
24	11.53	0.27	10.81	0.27
27	12.84	0.23	12.47	0.25
30	14.43	0.28	13.57	0.3

Table 4. Outcome variable: Amount of Fuel Used (gallons). Varied parameter: Number of farms.

Number of farms	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
3	5.38	0.13	1.79	0.07
6	12.47	0.59	10.29	0.46
9	19.84	0.8	18.12	0.67
12	27.61	0.84	25.4	0.87
15	35.56	1.0	32.92	0.99
18	42.8	1.13	40.23	1.09
21	53.62	1.53	47.74	1.31
24	60.87	1.41	57.06	1.45
27	67.8	1.24	65.84	1.32
30	76.18	1.46	71.67	1.6

Appendix B. Means and 95% confidence intervals of 4 key metrics based on the varying maximum waste amount.

Table 1. Outcome variable: Working Time (hours). Varied parameter: Maximum waste.

Maximum waste	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
0.1	2.59	0.08	2.43	0.05
0.2	3.67	0.08	3.56	0.08
0.3	4.76	0.12	4.59	0.09
0.4	5.99	0.14	5.63	0.15
0.5	7.25	0.18	6.84	0.17
0.6	8.44	0.2	8.11	0.22
0.7	9.79	0.25	9.21	0.25
0.8	11.23	0.29	10.61	0.3
0.9	12.58	0.37	11.83	0.37
1.0	13.53	0.34	12.76	0.33

Table 2. Outcome variable: Dump Visits. Varied parameter: Maximum waste.

Maximum waste	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
0.1	1.26	0.09	1.08	0.05
0.2	2.28	0.09	2.19	0.08
0.3	3.35	0.12	3.18	0.08
0.4	4.53	0.13	4.2	0.14
0.5	5.77	0.18	5.38	0.17
0.6	7.09	0.22	6.74	0.22
0.7	8.65	0.26	8.01	0.28
0.8	10.02	0.31	9.46	0.32
0.9	11.37	0.36	10.62	0.36
1.0	12.37	0.31	11.58	0.33

Table 3. Outcome variable: Driving Time (hours). Varied parameter: Maximum waste.

Maximum waste	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
0.1	2.25	0.06	2.15	0.04
0.2	2.89	0.06	2.83	0.06
0.3	3.54	0.08	3.45	0.07
0.4	4.3	0.1	4.07	0.1
0.5	5.11	0.12	4.86	0.12
0.6	5.9	0.14	5.67	0.16
0.7	6.79	0.18	6.45	0.18
0.8	7.81	0.22	7.4	0.22
0.9	8.72	0.27	8.16	0.27
1.0	9.26	0.26	8.72	0.23

Table 4. Outcome variable: Amount of Fuel Used (gallons). Varied parameter: Maximum waste.

Maximum waste	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
0.1	11.89	0.33	11.33	0.22
0.2	15.24	0.34	14.92	0.32
0.3	18.67	0.43	18.2	0.37
0.4	22.72	0.52	21.46	0.55
0.5	26.99	0.65	25.64	0.62
0.6	31.14	0.72	29.92	0.84
0.7	35.86	0.96	34.03	0.96
0.8	41.25	1.17	39.09	1.16
0.9	46.05	1.44	43.07	1.41
1.0	48.91	1.37	46.06	1.23

Appendix C. Means and 95% confidence intervals of 4 key metrics based on the varying number of drop-off sites.

Table 1. Outcome variable: Working Time (hours). Varied parameter: Number of drop-off sites.

Number of drop-off sites	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
1	13.92	0.4	12.48	0.33
2	12.51	0.34	11.81	0.34
3	12.0	0.3	10.99	0.27
4	11.4	0.3	10.72	0.3

Table 2. Outcome variable: Dump Visits. Varied parameter: Number of drop-off sites.

Number of drop-off sites	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
1	12.62	0.37	11.21	0.31
2	11.33	0.35	10.68	0.33
3	10.96	0.29	9.98	0.29
4	10.46	0.32	9.79	0.3

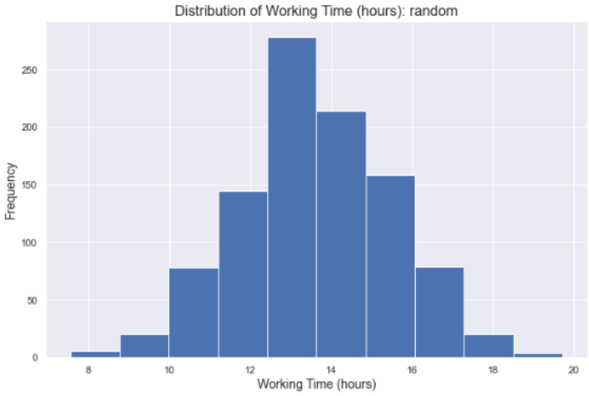
Table 3. Outcome variable: Driving Time (hours). Varied parameter: Number of drop-off sites.

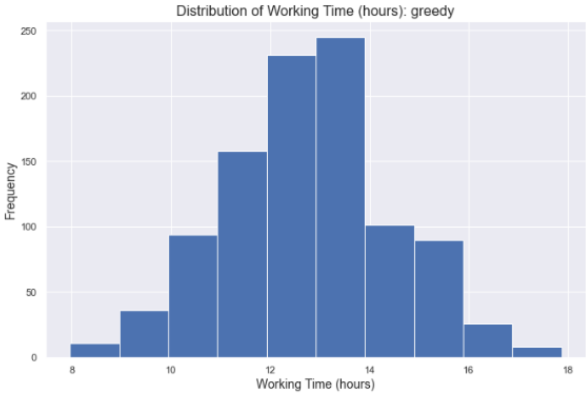
Number of drop-off sites	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
1	9.54	0.29	8.48	0.23
2	8.51	0.24	8.04	0.24
3	8.19	0.22	7.48	0.19
4	7.8	0.2	7.31	0.2

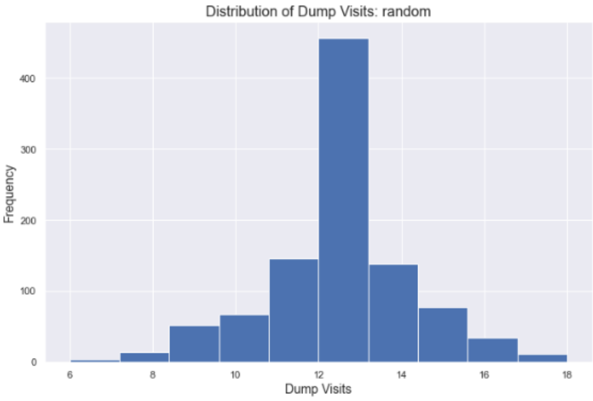
Table 4. Outcome variable: Amount of Fuel Used (gallons). Varied parameter: Number of drop-off sites.

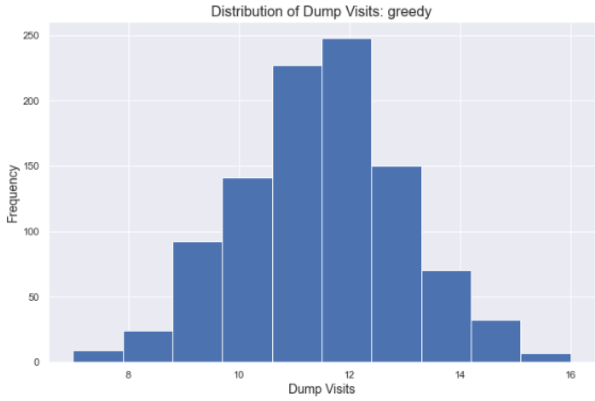
Number of drop-off sites	Random Avg	Random CI (+-)	Greedy Avg	Greedy CI (+-)
1	50.39	1.53	44.77	1.22
2	44.95	1.25	42.45	1.25
3	43.26	1.14	39.48	0.99
4	41.16	1.06	38.58	1.08

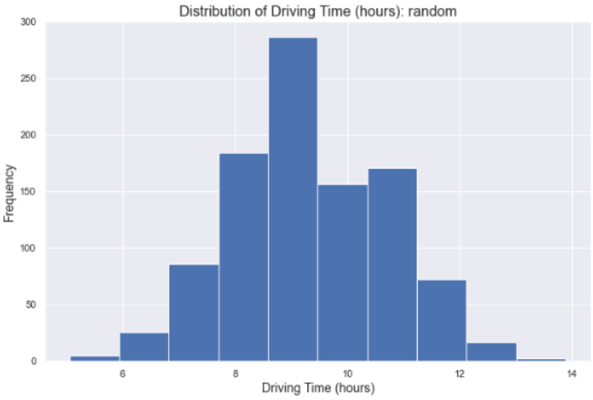
Appendix D. Distributions of 4 outcome variables (random and greedy strategy).

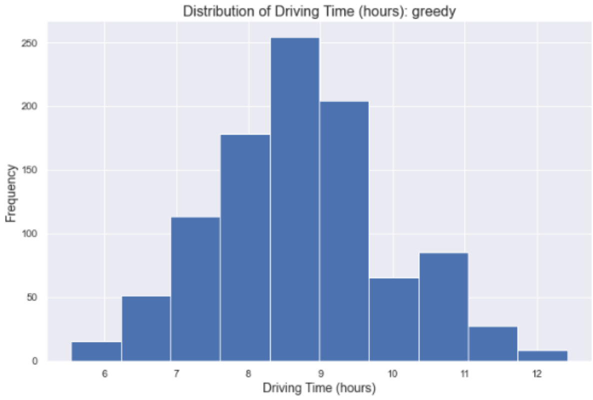


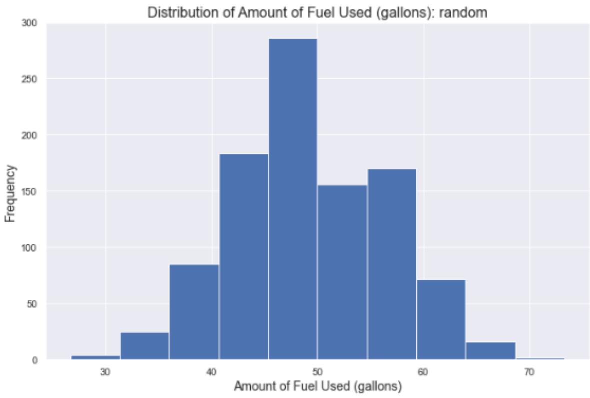


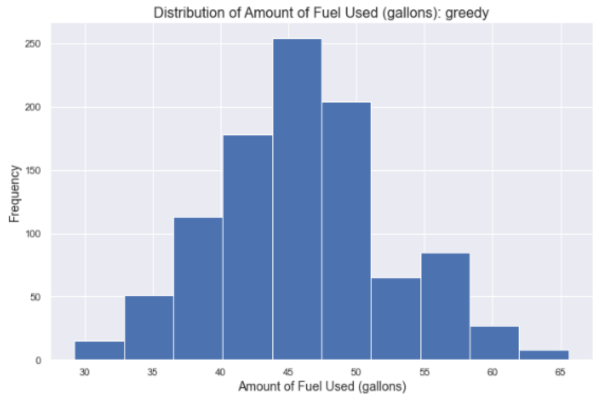

Distribution of Dump Visits: random
Frequency
Dump Visits


Distribution of Dump Visits: greedy
Frequency
Dump Visits


Distribution of Driving Time (hours): random
Frequency
Driving Time (hours)


Distribution of Driving Time (hours): greedy
Frequency
Driving Time (hours)


Distribution of Amount of Fuel Used (gallons): random
Frequency
Amount of Fuel Used (gallons)


Distribution of Amount of Fuel Used (gallons): greedy
Frequency
Amount of Fuel Used (gallons)