Data-Driven Waste Management Solutions

Seattle Public Utilities

Sakshi Agrawal, Josh Breiger, Chen Song,
Kielan Lemoine-Kowalski

Table of Contents

01

Project Overview

Executive Summary

Seattle Public Utilities' (SPU) Solid Waste Division has in place approximately 900 public place trash cans for pedestrian use throughout the city in an effort to keep Seattle free of litter. The pick-up frequencies of these cans have been historically determined primarily by empirical observation, where staff and citizen reports govern scheduling updates. While this method worked to a degree, there were many cans that were empty or overflowing at the time of collection.

Thus, SPU sought a data-driven solution to better estimate waste levels, thereby optimizing waste bin placement and service scheduling. Our team collaborated with the relevant stakeholders to create an appropriate data collection system and created an optimization algorithm to minimize costs. We also provided recommendations with regard to collection frequency changes, both on a per-can and neighborhood basis. Our work aimed to reduce SPU's expenses, support a cleaner city, and highlight shortcomings of SPU's current information management practices.

Background

Seattle Public Utilities

Seattle Public Utilities provides three major utilities to the city: Water utility, drainage and wastewater utility, and solid waste utility.

Seattle Public Utilities (SPU) is a department within the city of Seattle that works to keep the city clean, both environmentally and aesthetically. SPU manages residential, commercial, and public place can waste management; monitors and abates issues such as illegal dumping and graffiti; and oversees and tests water and waste-water systems.

SPU's Public Place Litter and Recycling program falls under a broader set of Clean City programs -- a set of programs aimed at making Seattle a cleaner and safer place to live in addition to fostering citizen environmental awareness. The Public Place Litter and Recycling Program is an initiative intended to reduce pedestrian littering, with approximately 900 trash cans and 300 paired recycling bins to date. The program also intends to support Seattle's goal of a 75% recycling rate by 2022.

Problem

SPU pays their contractors a fixed amount for each can collected, regardless of how full the can is at the time of pick-up. When there are many empty cans and/or cans far below capacity on many routes, this is very costly to the organization.

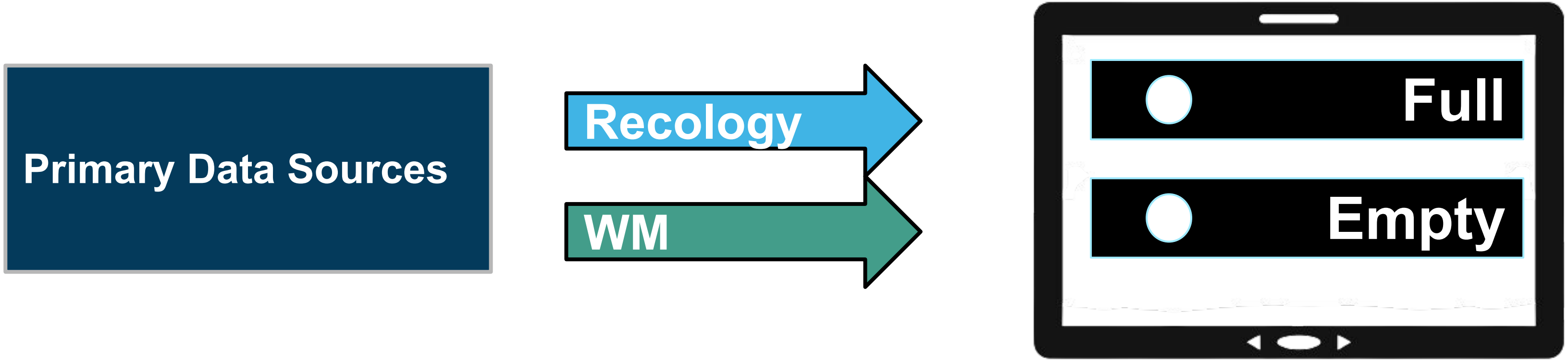
SPU also was experiencing problems with can overflow, which is also costly, as clean-up crews must be deployed to pick up and properly dispose of the litter. Sometimes can overflow could be a product of illegal dumping, special events, or unanticipated crowds. However, much of the time, cans that were beyond capacity were a result of normal pedestrian waste disposal practices and too few scheduled pick-ups.

SPU wanted a data-driven solution to reduce the number of empty cans and overflowing cans serviced.

02

Our Approach

Data Collection - Primary



SPU did not have any historical data of relevance. As such, we met with their two contractors -- Recology and Waste Management -- to determine whether they had any data that may be useful for creating a scheduling optimization algorithm. Neither contractor had any data applicable to our interests, but each Recology and Waste Management each agreed to collect data pertaining to waste levels for 6 and 10 weeks, respectively. However, the contractors had pre-programmed tablets with only two limited can reporting options -- full or empty. Any cans not marked as such would be assumed to be optimal. We operationally defined empty as 0-20% capacity, full as 80-100% capacity, and optimal at any level between the two.

Data Collection - Secondary Sources

SPU informed us that the main drivers of trash disposal are pedestrian traffic. Data on this did not exist, however, and it was infeasible to manually record pedestrian traffic over the entire Seattle area for many weeks. As such, we looked through King County public datasets and identified several variables that were related to -- or could work as a proxy of -- foot traffic.

The most pertinent variables we identified were restaurants, parks, bus stops (and associated ridership), and grocery stores.

Using the latitude and longitude coordinates of each can, we calculated “buffers zones” -- areas within a 2-block radius of each can. Then, using the coordinates of our features, we calculated how many restaurants, parks, grocery stores, and bus stops were in each of these “buffer zones”.

While we did not use these features for modeling their current problem, this data will be very useful if SPU obtains the necessary data to run machine learning models in the future.

Optimization Model

We developed an optimization model to minimize the cost function associated with each route

Calculate average daily fill rate of each waste bin

Perform simulations to minimize cost function for each route

Determine optimal collection frequency

To calculate the average daily fill rate of each bin, we took the average collection volume for each can and multiplied it by (collection frequency / 7)

For each route, for each potential collection interval, a simulation was run 2,000 times to optimize collection frequency with regard to cost function minimization

Our output provided, for each route, optimal collection frequency, average waste level per can at that frequency, and the number of cans expected to be empty and or full

Optimization Model

```
Route Number: 5
Neighborbood Name:  Greenwood
Average Waste Level: 0.8793%
Route Metric[empty, overfull] :[0, 0]
Suggested Collections Per Weeks:2.8
Current Collection Frequency: 5
```

Average waste level per can associated with this collection frequency

The suggested collection frequency and current collection frequency compared. The former was often as much 50% less than the latter

Cost Function Design

To develop the cost function, we first had to identify the relative costs associated with empty and overfull cans. SPU deemed an empty can to be about twice as costly as an overfull can; thus, our cost function weighted these costs in a 2:1 ratio. Additionally, we wanted cans to be closer to full at the time of collection. As such, we added another statement that assigned a cost of (1 minus the % fullness of the can) to each can, penalizing less full cans more harshly. To avoid our model returning results with many overflowing cans due to the relatively low cost associated with them, we assigned a very high cost to cans with significant overflow (> 125% capacity).

By allowing cans to fill to a greater average waste level, we are able to reduce weekly collection frequency

03

Project Outcomes

Solution Deployment

Phase 1 - Deploy

Conservatively pilot the revised collection frequencies provided by our optimization algorithm in a limited number of neighborhoods.

Phase 2 - Evaluate

If suggested frequencies result in increased costs (or few improvements), this suggests data collected was *not accurate* or is *not reflective* of current conditions (COVID-19's effect on pedestrian traffic may change necessary collections, as may seasonality/weather).

Phase 3 - Refine

Collect new data that is reflective of current environmental conditions and run the optimization algorithm with the new data. Also adjust cost function if necessary.

Additional Recommendations

Data Collection

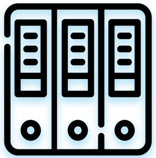
Expand current data collection practices. This may include contract renegotiation, technology adoption, or some combination thereof.

In the future, waste level data should be collected across more bins (more precise ranges than those defined for this project) to reduce the number of inaccuracies introduced via approximation.

The total amount of data collected should be significantly increased. Additionally, data should be collected across different seasons and environmental conditions, as these variables affect fills rates.

Additional Recommendations Cont’d

Data Infrastructure



Consider a centralized database to house data. This prevents data “siloing” in different departments by combining and allowing for the contextualization of data.

An efficient database allows for easy and fast data retrieval, better data organisation, and also supports high end data visualization and reporting.

Considering exploring waste management technology that will allow for more accurate data and will lend itself to strong predictive analytics with the use of machine learning models.

Benefits

- Our model will reduce SPU's collection frequencies in each neighborhood, thus reducing their overall expenditures.

- Optimizing collection frequency will result in fewer overflowing cans and consequently reduce litter. Furthermore, it will lower Seattle's environmental footprint by reducing carbon emissions associated with truck driving.

- Working with SPU highlighted information-related shortcomings with respect to both technology and processes and will shape future information management practices. We provided SPU with a blueprint they can use in order to more efficiently collect data in the future. We also helped to incorporate some practices for data storage and management.

Benefits

Neighborhood	Current Collections	Suggested Collections	Projected Difference
International Dis..	14.00	7.00	-7.00
Pioneer Square	14.00	7.00	-7.00
Roosevelt	7.00	2.80	-4.20
Adams	7.00	3.50	-3.50
Belltown	7.00	3.50	-3.50
Broadway	7.00	3.50	-3.50
Columbia City	7.00	3.50	-3.50
East Queen Anne	7.00	3.50	-3.50
Fremont	7.00	3.50	-3.50
Lower Queen An..	7.00	3.50	-3.50
Loyal Heights	7.00	3.50	-3.50
Mid-Beacon Hill	7.00	3.50	-3.50
Pike-Market	7.00	3.50	-3.50
Stevens	7.00	3.50	-3.50
University Distri..	7.00	3.50	-3.50
Montlake	4.00	1.27	-2.73
Wallingford	5.00	2.33	-2.67
Alki	7.00	4.67	-2.33
Genesee	7.00	4.67	-2.33

The following figure shows SPU's current collection frequencies alongside our model's suggested collection frequencies. The "Projected Difference" column shows that most routes will be collected about half as often.

These changes to route collection frequency have the potential to save SPU tremendous amounts of money and resources.

Project Limitations

During the data collection phase of the project, we had to rely on the contractors to collect data for us. The data fed into our calculation will be reflected in the success of the model. The following assumptions could significantly impact model performance:

1

The contractors agreed to collect data for 6-10 weeks, and the detail of collection was limited. The contractors recorded cans as “over-filled” (80-100% capacity), empty (0-20% capacity), or “optimal” (20% - 80% capacity). To calculate the daily rate of change, we used the middle numbers from each “bin” (10%, 50%, and 90% for empty, optimal, and full, respectively). However, since most of the cans were recorded as optimal, this number was 50% the majority of the time, and thus the average trash volumes at collection were almost always close to 50%. The daily rate of change was a function of this number, and any inaccuracies with respect to rate of change will be reflected in model output.

2

The second assumption made is that each route for which data was provided was serviced every collection interval of every week for the entire duration of the pilot data collection period. The consequences of failing to meet this requirement skew the ratio of empty and full to “optimal” cans, which in turn results in the average can volume at pick-up being closer to ~50% capacity, thus affecting the fill rates, and by extension, the optimization algorithm’ s efficacy.

Future Work

SPU has many possibilities for improving their waste management solutions using our project as a foundation

- Due to budget and time constraints, the accuracy of our solution has limitations. SPU should continue to collect more data and data of greater detail that will maximize performance of the current model.

- Using the secondary data sources and buffers that we built, SPU has a starting point from which to build features for complex machine learning models. However, they first must acquire additional data regarding the “ground truth,” which can be done by acquiring large sets of accurate data on waste levels.

- Exploration of smart waste management solutions for the most accurate data and associated route optimization -- with regard to collection frequency and can placement, and more!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Thanks!