Balancing Efficiency and Equity in Urban Waste Management in Toronto from 2021 to 2024*

A Data-Driven Analysis of Demographics, Housing, and Waste Collection Patterns

Jiwon Choi

December 1, 2024

Effective urban waste management is critical for sustainability, yet many cities struggle to balance efficiency and equity in resource allocation. This paper analyzes demographic and housing characteristics across Toronto to understand their influence on waste generation and collection patterns. Key findings reveal that high-density areas often face disproportionate servicing needs, while low-density areas may underutilize resources. These insights provide actionable recommendations for optimizing waste management strategies, promoting fairness, and enhancing sustainability efforts, offering a replicable framework for other cities to achieve similar goals.

1 Introduction

Effective waste management is essential for sustainable urban living, yet achieving efficiency and equity in servicing remains a persistent challenge. As cities grow and diversify, understanding the demographic, dwelling, and household-level factors influencing waste generation and collection patterns becomes critical. This paper investigates these dynamics using comprehensive datasets from Toronto, combining demographic and waste collection data to offer insights into operational inefficiencies and equity gaps.

The research addresses a gap in understanding how localized demographic and housing characteristics impact waste generation and service frequency. While existing studies provide broad

*Code and data are available at: https://github.com/jwonc4602/Waste-and-Recycling-in-Toronto.

insights into urban waste management, they often lack granular analyses that connect demographic details with service patterns. This study bridges that gap by exploring (1) the influence of age distribution, household size, and housing type on waste generation; and (2) the operational implications of these relationships for resource planning and optimization.

Using statistical and predictive modeling, we analyze how these factors shape waste management needs across Toronto's wards. Key findings include the disproportionate servicing of certain asset types in high-density areas and the underutilization of resources in low-density neighborhoods. These insights highlight opportunities to enhance service allocation and promote equitable resource distribution.

The findings of this study are significant not only for optimizing waste management in Toronto but also for informing broader policy discussions on urban sustainability. By aligning demographic data with operational metrics, this research provides a replicable framework for cities aiming to balance efficiency with equity in public services.

The remainder of this paper is structured as follows. Section 2 describes the data sources and key estimands. Section 3 outlines the modeling approach, including justification and diagnostics. Section 4 presents the results, and Section 5 discusses their implications, limitations, and directions for future research. Finally, the Appendix provides detailed insights into data cleaning, model diagnostics, additional figures, and limitations, offering further transparency and replicability for the study.

2 Data

2.1 Overview

This research is based on two datasets: Litter Bin Collection Frequency (Services 2024) and Ward Profiles (Planning 2024), which are accessible through the City of Toronto's OpenDataToronto Library (Gelfand 2022). Litter Bin Collection Frequency dataset contains more than 10,000 garbage/recycling street litter bins across the city and the frequency of collection varies based on the location and/or usage of each bin and seasonality. Ward Profiles dataset contains the 2021 Ward Profiles based on the 25-Ward model (effective December 1, 2018) are available from City Planning and each Ward Profile provides a snapshot of the population and households in the Ward. The data was compiled and examined using the Python: A programming language for clear and concise code (Foundation 2024) and R statistical programming software (R Core Team 2023a), supplemented by various tools such as `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `readr` (Wickham, Hester, and Bryan 2023), `gridExtra` (Auguié 2017), `grid` (R Core Team 2023b), `knitr` (Xie 2014), and `here` (Müller and Bryan 2020).

This traffic accident dataset contains 18,957 entries from a citywide survey, detailing various factors involved in incidents. It includes 50 variables covering geographic data, such as accident

Table 1: Sample of Cleaned Litter Data

WARD	DAYS.SERVICED	ASSET.TYPE	STATUS
Ward4	5	WR1	Existing
Ward4	7	WR4	Existing
Ward4	7	WR2	Existing
Ward4	7	WR4	Existing
Ward4	7	WR1	Temporarily Removed

locations, and behavioral data, like driver conditions and alcohol involvement. The dataset spans several years and captures environmental conditions, road types, and the actions of drivers and pedestrians during accidents. It aims to provide a comprehensive view of research and policy efforts related to urban traffic safety and accident prevention. Although the dataset was last updated September 27, 2024, but only contains observations from 2006 to 2023.

2.2 Estimand

The primary estimand for this analysis is the relationship between demographic, dwelling, and household characteristics in Toronto neighborhoods and their influence on waste generation and waste collection patterns. Specifically, this analysis aims to estimate: 1. How demographic factors such as age distributions affect waste generation across wards. 2. How dwelling types and household sizes contribute to variations in waste output and service frequency. 3. The extent to which these factors predict waste generation for planning and optimization purposes.

This estimand is clearly stated to help guide the focus of the analysis and ensure a targeted approach to evaluating waste management strategies.

2.3 Litter Data

The datasets utilized for this study provide a comprehensive view of demographic and waste-related characteristics across Toronto wards. Below is a description of each dataset and its broader context.

1. **Litter Data (`cleaned_data_litter.csv`)**:
    - **Description**: This dataset contains metrics on waste collection across Toronto's 25 wards. It includes information on the frequency of servicing, the types of litter bins used (WR1-WR5), and their status. Service levels indicate the frequency at which bins are collected, but this only applies to bins with a status of "existing." Bins with a status of "planned" or "temporarily removed" are not currently collected.

- **Context**: These data are critical for understanding the distribution and frequency of waste collection services, as well as the infrastructure used for waste management. The broader implications include evaluating service efficiency, identifying underperforming areas, and ensuring equitable allocation of waste management resources.
- **Variables**:
  - **Days Serviced**: The frequency of waste collection for "existing" bins, categorized into one day, three days, or five days of servicing per week.
  - **Asset Types (WR1-WR5)**:
    - **WR1**: A plastic litter bin model with two ports of entry – one for litter and one for recycling.
    - **WR2**: A plastic litter bin model with three ports of entry – one for litter and two for recycling.
    - **WR3**: A steel litter bin model on a raised concrete pedestal with two ports – one for litter and one for recycling.
    - **WR4**: A steel litter bin model with three ports of entry shaped differently from WR1 and WR2 – one for litter and two for recycling.
    - **WR5**: The newest bin model with larger ports – two for litter and one for recycling.
  - **Bin Status**:
    - **Existing**: Bins currently deployed and actively serviced.
    - **Planned**: Bins intended for future installation but not yet deployed or serviced.
    - **Temporarily Removed**: Bins removed temporarily (e.g., for repairs or reallocation) and not currently being serviced.
  - **Ward Identifiers**: Numerical identifiers corresponding to the 25 wards in Toronto.
- **Key Context**: The relationship between bin types, status, and service frequency can provide insights into operational efficiency and resource allocation. The presence of planned and temporarily removed bins highlights future changes and potential service gaps.

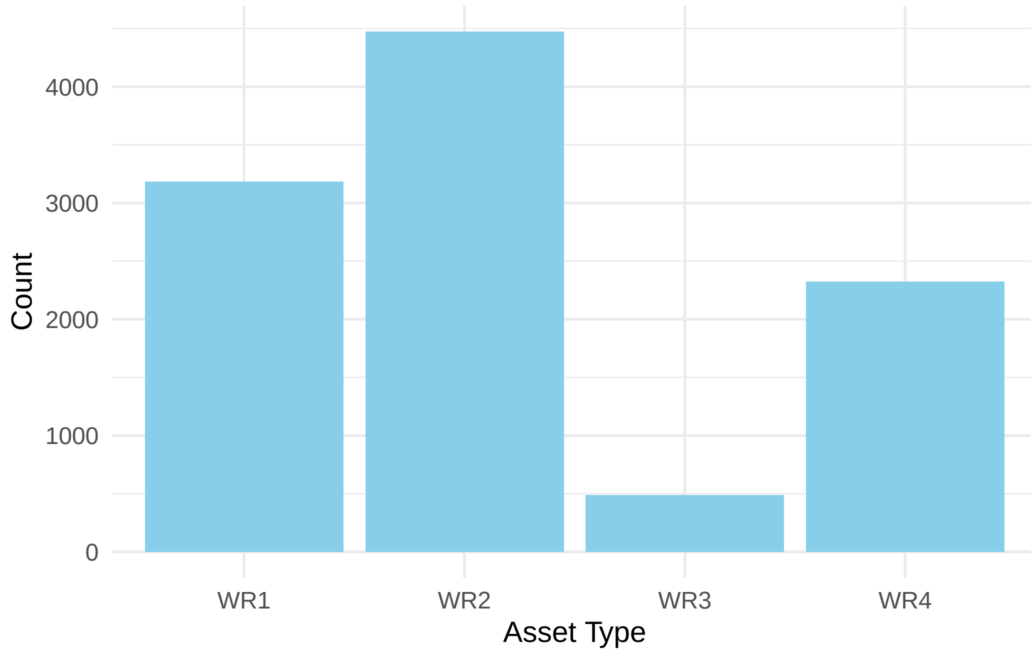

Figure 1: Frequency of Litter Bins by Asset Type across the City

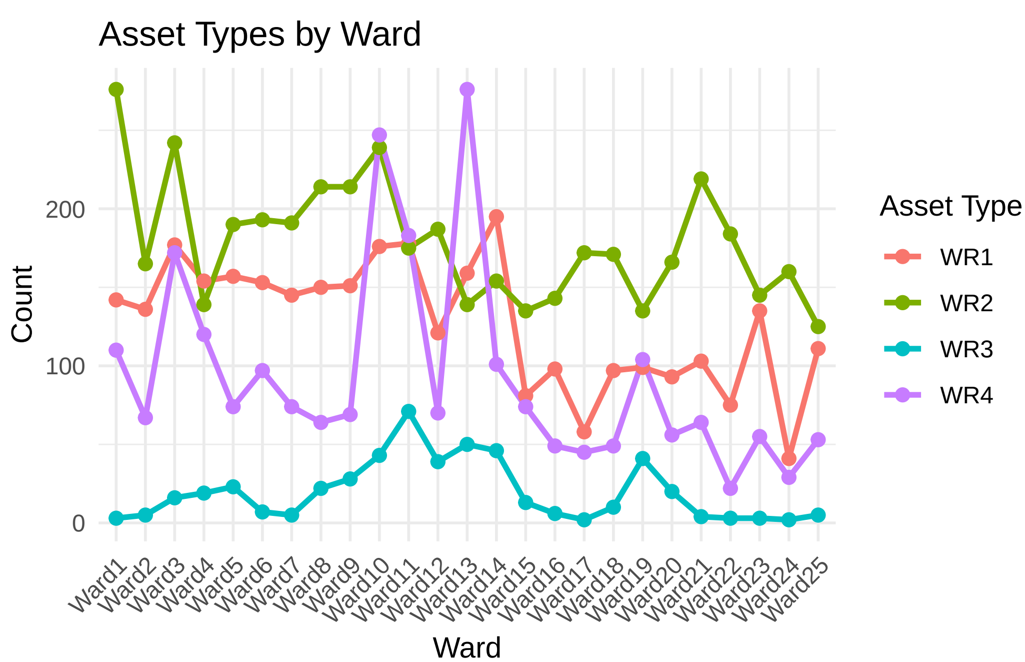

Figure 2: Line Graph of Asset Types by Ward

Figure 3: Status Distribution by Asset Type

2.4 Ward Profile Data

1. **Age Demographics (`ward_data_age.csv`)**:
    - **Description**: Contains age distribution data for Toronto, divided into multiple age groups (e.g., "0 to 4 years", "15 to 19 years"). These data provide a detailed view of the population composition by ward.
    - **Context**: Understanding the age demographics is crucial, as waste generation often varies by age group due to differing consumption and waste production behaviors. For example, younger age groups might contribute to higher waste from disposable items, while older populations might contribute to other types of waste patterns.
    - **Variables**: Total population, age groups (e.g., "0 to 4 years", "65+ years"), ward identifiers.
    - **Key Context**: Variations in age distributions across wards could explain differences in waste types (e.g., recyclable materials vs. organic waste).

Table 2: Sample of Cleaned Age Data

Variable	Minors	Adult	Middle_Age_Adult	Senior_Adult
Ward1	25585	35130	29380	25020
Ward2	23840	27570	32090	33700
Ward3	23620	44935	38605	32760
Ward4	19695	32780	30745	21485
Ward5	24855	32545	32015	26260

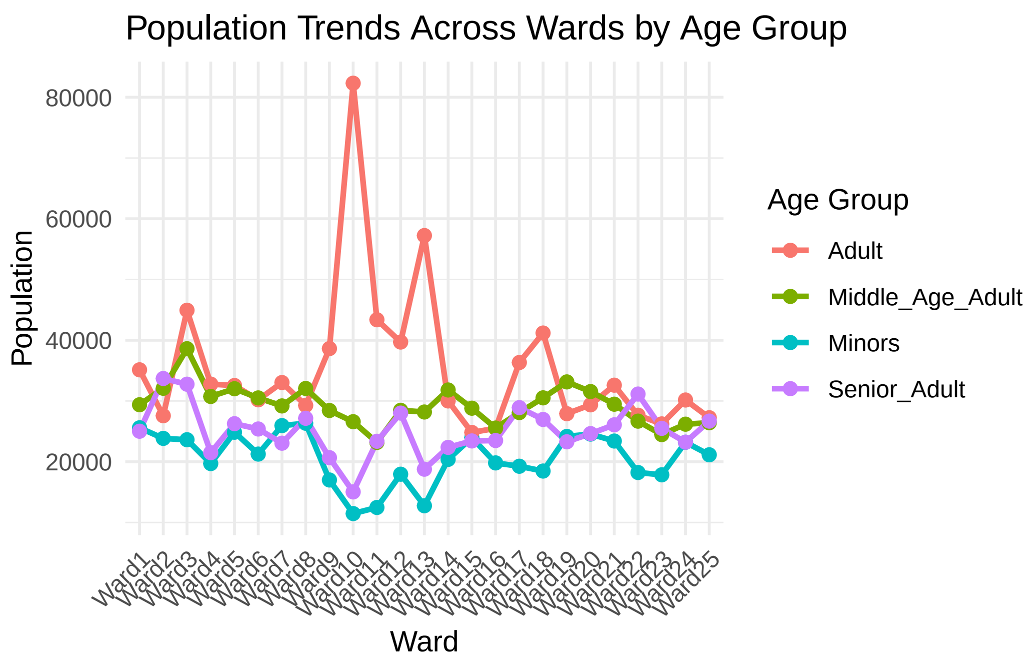

Figure 4: Population Trends Across Wards by Age Group

2. **Dwelling Type (`ward_data_dwelling_type.csv`)**:

- **Description**: Provides data on housing structures across wards, including single-detached homes, semi-detached homes, row houses, and apartments.
- **Context**: Different dwelling types are associated with varying waste generation trends. For instance, single-detached homes might produce more organic waste due to larger gardens, while apartment residents might contribute more to recyclable waste streams.
- **Variables**: Total dwellings, dwelling type categories (e.g., single-detached, semi-detached, row houses, apartments), ward identifiers.
- **Key Context**: The mix of dwelling types in a ward could influence waste service requirements and resource planning.

Table 3: Sample of Cleaned Dwelling Data

Variable	Low_Density_Housing	Medium_Density_Housing	High_Density_Housing	Other_Types
Ward1	12695	7435	17995	10
Ward2	22735	2790	19785	25
Ward3	19950	3840	41745	35
Ward4	13070	3295	32995	85
Ward5	16505	3110	25405	35

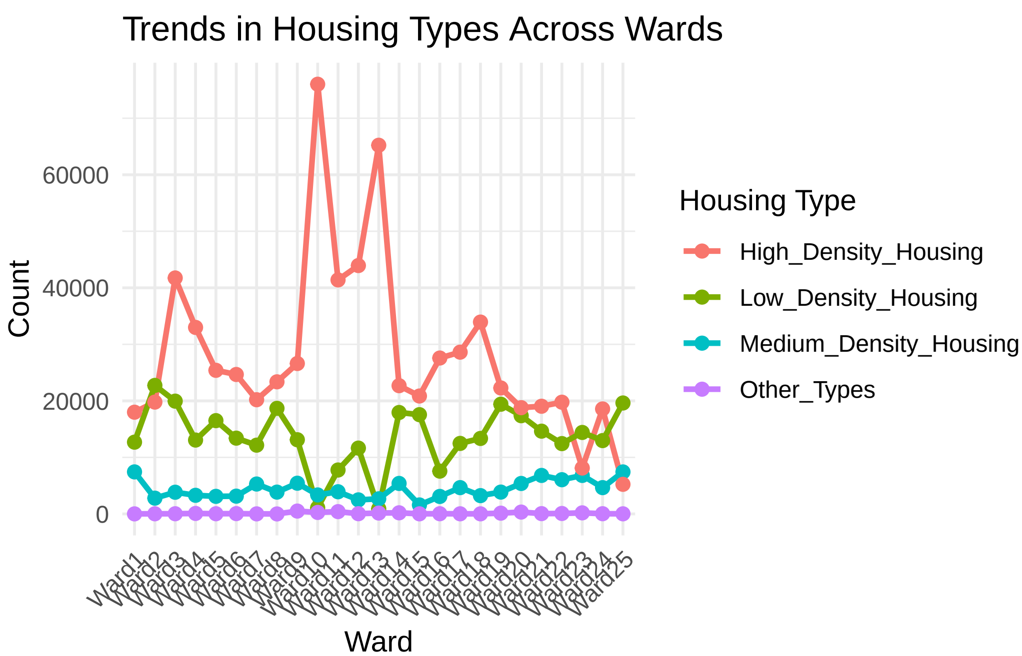

Figure 5: Trends in Housing Types Across Wards

3. **Household Size (`ward_data_household_size.csv`)**:
   - **Description**: Details the distribution of households by size across Toronto's wards, including one-person to multi-person households.
   - **Context**: Household size is an important factor in waste generation, as larger households might produce more waste overall but less per capita, while single-person households might generate more waste in packaging and convenience items.
   - **Variables**: Total households, household size categories (e.g., one-person, two-person, three-person households), ward identifiers.
   - **Key Context**: Larger households may necessitate additional or larger waste bins, while smaller households might correlate with more frequent servicing needs.

Table 4: Sample of Cleaned Household Size Data

Variable	Small_Households	Medium_Households	Large_Households
Ward1	17425	13920	6795
Ward2	25460	15840	4045
Ward3	46340	16385	2850
Ward4	34805	12475	2165
Ward5	25495	14930	4620

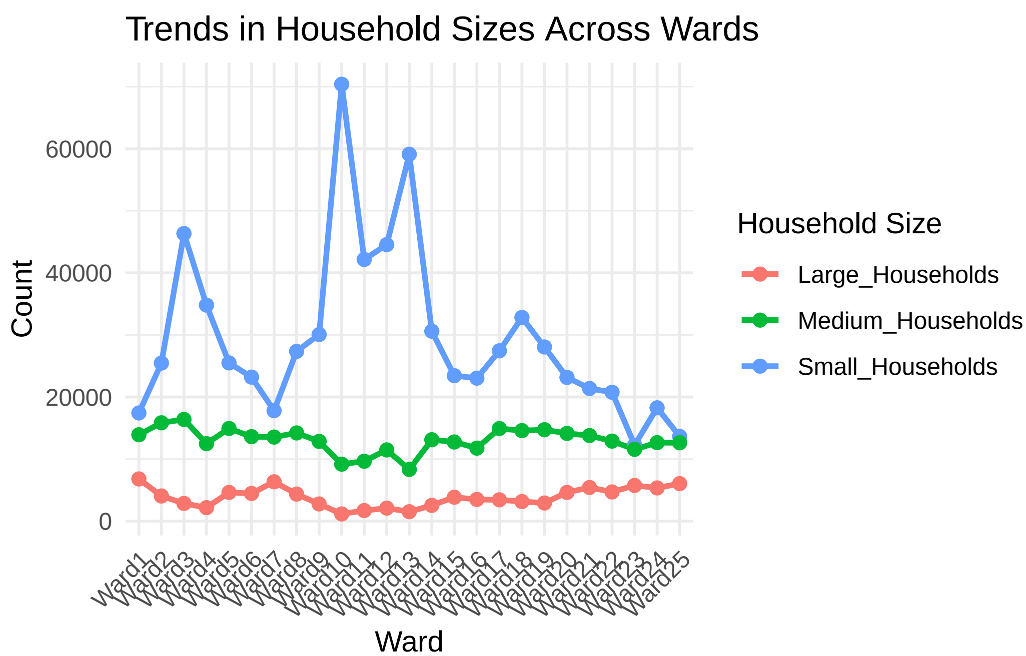

Figure 6: Distribution of Household Sizes by Ward

2.5 Measurement

The data used in this analysis represent a snapshot of various real-world phenomena related to waste management and demographics in Toronto: - **Litter Data**: Measurements for waste collection frequency and asset type were obtained from municipal waste management systems that track service days, infrastructure usage, and bin status. Only bins with a status of "existing" are actively serviced, highlighting the importance of this subset in current operations. - **Age Demographics**: The age data originate from census or municipal records, offering a detailed breakdown of population characteristics. These measurements are essential for linking demographic factors with observed waste trends. - **Dwelling Types and Household Sizes**:

These data are derived from housing and census records. They reflect the physical and social structures within each ward, providing a critical lens for understanding waste behaviors and resource needs.

By integrating service levels, bin status, and demographic factors, this dataset provides a robust foundation for analyzing waste management strategies, identifying inefficiencies, and planning for equitable and effective resource distribution across Toronto's neighborhoods.

3 Model

3.0.1 Bayesian Logistic Regression Model to Predict Litter Bin Status

We employ a **Bayesian logistic regression model** to predict the binary status of litter bins (`STATUS`), where 1 indicates "Existing" and 0 indicates "Not Existing." This model enables us to incorporate prior beliefs and uncertainty while making predictions based on observed data.

3.0.1.1 Model Specification

The logistic regression model is defined as:

equation

Where: - (STATUS_i): Binary outcome variable for the (i)-th observation (1 = "Existing", 0 = "Not Existing"). - (ASSET.TYPE_i): Categorical variable representing the type of litter bin asset. - (DAYS.SERVICED_i): Numeric variable indicating the number of days the litter bin is serviced weekly. - (Low.Density.Housing_i): Proportion of low-density housing in the corresponding ward. - (Medium.Density.Housing_i): Proportion of medium-density housing in the corresponding ward. - (High.Density.Housing_i): Proportion of high-density housing in the corresponding ward. - (_0): Intercept term. - (_1, _2, …, _5): Coefficients for the respective predictors.

3.0.1.2 Justification of Features

1. **ASSET.TYPE**: Captures the characteristics of litter bin types, which may influence their status.
2. **DAYS.SERVICED**: Reflects the maintenance frequency, potentially correlating with the likelihood of a bin remaining operational.
3. **Housing Density Variables**: Indicate urbanization levels, which could affect waste management dynamics and bin usage patterns.

3.0.1.3 Model Implementation

The model was implemented in R using the `rstanarm` package, which facilitates Bayesian modeling and inference. Key parameters include: - **Iterations**: 4000 (2000 warmup, 2000 sampling). - **Chains**: 4 for robust posterior estimation. - **Adapt delta**: 0.95 to ensure convergence. - **Seed**: 1234 for reproducibility.

3.0.1.4 Limitations and Assumptions

- **Assumption**: The relationship between predictors and the logit of the outcome is linear.
- **Limitation**: Potential unobserved confounders, such as policy or environmental factors, are not included.
- **Applicability**: This model may not generalize well to contexts with significantly different housing or servicing dynamics.

3.0.1.5 Alternative Models Considered

1. **Random Forests**: Explored for non-linear relationships but deemed less interpretable for this context.
2. **Generalized Linear Models (Non-Bayesian)**: Rejected due to the need for explicit incorporation of prior beliefs.

The Bayesian logistic regression model balances complexity and interpretability, making it well-suited for predicting litter bin statuses.

4 Results

4.1 Overview of Coefficients from the Bayesian Logistic Regression Model

The Bayesian logistic regression model was applied to predict the probability of a litter bin's status as "Existing" or not, using key predictors such as asset type, days serviced, and housing density types. The table and visualization below summarize the estimated coefficients and their corresponding uncertainty intervals.

Coefficient Estimates Table Table 1 provides the coefficients derived from the model. The regression coefficients (estimate), standard errors, and 95% credible intervals (conf.low, conf.high) are presented for each predictor.

ASSET.TYPE: Asset Type 3 and Asset Type 4 had significant positive effects on the probability of a bin being "Existing." Conversely, Asset Type 2's effect was slightly negative but not significant. DAYS.SERVICED: Days serviced was negatively associated with the status, indicating that bins serviced more frequently were less likely to be "Existing." Housing Density:

Table 5: Coefficients from the GLM Model

term	estimate	std.error	conf.low	conf.high
(Intercept)	2.15	0.06	2.05	2.24
ASSET.TYPEWR2	-0.10	0.08	-0.22	0.02
ASSET.TYPEWR3	0.46	0.19	0.16	0.77
ASSET.TYPEWR4	0.35	0.09	0.20	0.51
DAYS.SERVICED	-0.26	0.04	-0.32	-0.19
Low__Density__Housing	0.38	0.06	0.28	0.48
Medium__Density__Housing	-0.07	0.04	-0.14	0.00
High__Density__Housing	0.16	0.07	0.05	0.28

Low-density housing had a strong positive association, while medium-density housing showed a small negative effect (not significant). High-density housing had a modest positive impact. Coefficient Estimates Plot The plot below visualizes the coefficient estimates and their 95% credible intervals.

4.2 Predicted Probabilities

Using the fitted Bayesian model, predicted probabilities for the bins' statuses were calculated. These probabilities provide insight into the likelihood of a bin being "Existing" based on its characteristics.

Key Insights: Bins in low-density housing areas consistently had higher predicted probabilities of being "Existing." Bins with fewer days serviced showed increased probabilities of being "Existing." Bins of Asset Types 3 and 4 were more likely to be "Existing," whereas Asset Type 2 bins were less likely. 3. Limitations and Model Fit While the model fit well overall, limitations include the sensitivity to categorical asset types and the potential for omitted variable bias. Further investigation could include additional interaction terms or hierarchical modeling to capture ward-level effects.

This analysis provides critical insights into waste management patterns, identifying significant predictors and offering a framework for targeted interventions in litter bin servicing.

5 Discussion

5.1 Prediction of Waste Bin Servicing

The predictions provided in the dataset, particularly the Predicted_Probability column, shed light on the likelihood of waste bins requiring servicing. A thorough evaluation of the model's accuracy is essential to understand its reliability. Metrics such as precision, recall, and overall accuracy should be analyzed, with a focus on identifying any systematic biases, such as overestimating servicing needs in certain wards or underestimating them for specific asset types.

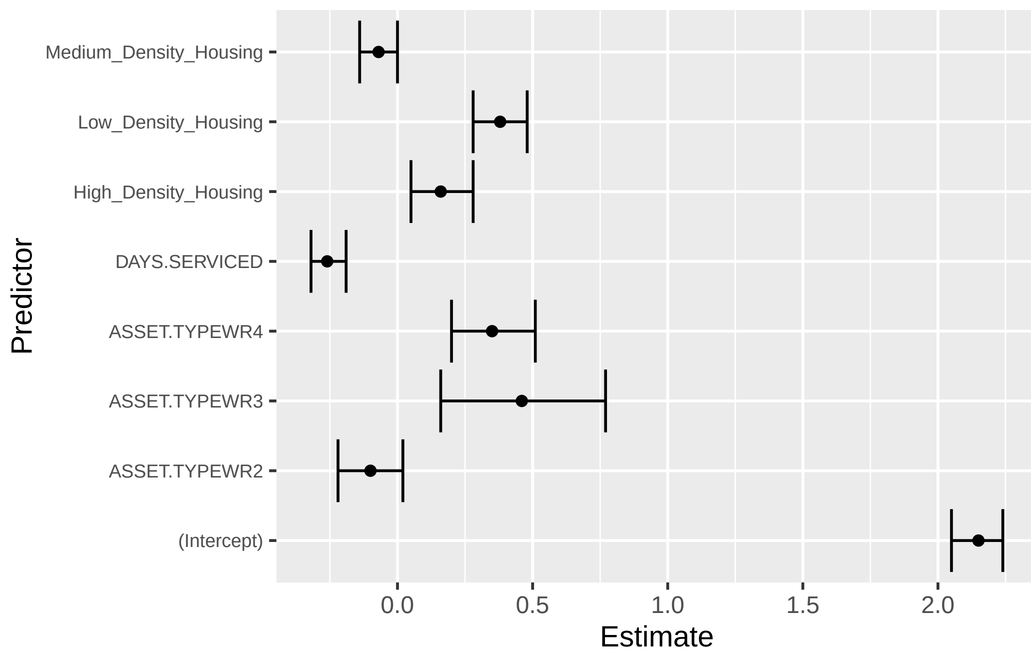

Figure 7: Coefficient Estimates for Predictors

Furthermore, examining the relationship between predicted probabilities and actual servicing outcomes (STATUS) can highlight potential areas for model refinement.

Key predictors driving these probabilities also warrant discussion. Variables like housing density, household size, and demographic composition are likely significant, and their influence on waste servicing patterns can offer insights into the model's decision-making process. Understanding these relationships can also help fine-tune the model and ensure it accounts for context-specific factors affecting waste generation and collection needs. Lastly, the operational implications of these predictions cannot be understated. Accurate models can streamline resource allocation, reduce unnecessary servicing, and ultimately lower costs while improving service quality.

The model's performance should also be examined across wards to determine its consistency. Variability in servicing patterns due to localized conditions—such as ward-specific population densities or waste management policies—may affect prediction accuracy. Addressing these discrepancies will be vital for ensuring that predictions are actionable and equitable across all areas, avoiding a one-size-fits-all approach to waste management.

5.2 Demographic and Housing Influence

Demographic and housing characteristics play a pivotal role in determining waste generation patterns and servicing needs. Variables such as Minors, Adult, and Senior_Adult provide a lens into the population composition of different wards. Areas with a higher proportion of certain age groups may exhibit unique waste generation behaviors. For instance, regions with a significant number of senior adults might generate less waste per household compared to those with larger families or young children. Such demographic-driven insights can guide tailored waste management strategies that account for population-specific behaviors.

Similarly, the type of housing in a ward—whether low, medium, or high-density—has a direct impact on waste generation. High-density areas, typically associated with apartment buildings, may have concentrated waste disposal points requiring more frequent servicing. Conversely, low-density residential areas might generate less concentrated waste but pose logistical challenges in collection due to greater distances between bins. Exploring these patterns can help identify whether waste management resources are appropriately allocated based on housing type and density.

Beyond understanding patterns, equity in servicing must be addressed. Spatial disparities in servicing frequency may emerge, particularly in areas with different socio-economic profiles. For instance, high-density, lower-income neighborhoods might generate more waste but face under-servicing due to resource constraints. Identifying and addressing these inequities ensures a fair and effective waste management system that meets the needs of all residents, regardless of their location or demographic profile.

5.3 Performance Across Asset Types

The dataset's ASSET.TYPE variable offers a valuable perspective on how different types of waste bins perform in terms of servicing needs. Analyzing the servicing frequency for each asset type can reveal whether some bins are more demanding than others. For example, certain types might be placed in high-traffic areas such as commercial zones, necessitating more frequent collection. Identifying these patterns can provide insights into the operational efficiency of asset deployment and highlight areas for improvement.

Asset design and usage also contribute to servicing demands. Specific bin types might struggle to accommodate the volume or type of waste they are exposed to, leading to inefficiencies. For instance, bins with smaller capacities in busy areas might require frequent emptying, whereas larger bins in underutilized locations might remain underfilled. Understanding these mismatches between design, placement, and usage can help optimize bin distribution and reduce servicing costs.

Finally, optimizing asset performance involves using data-driven strategies to refine placement and resource allocation. Bins with higher servicing needs might benefit from being replaced with larger or better-designed alternatives, while underutilized bins can be reassigned to areas with greater demand. By leveraging insights from the dataset, waste management systems can ensure that asset deployment aligns with actual servicing requirements, improving both efficiency and effectiveness.

5.4 Weaknesses and next steps

Despite the strengths of the dataset and analysis, several weaknesses could limit the insights and practical applications of the findings. The dataset may have gaps or inaccuracies, particularly in the STATUS and Predicted_Probability columns, which could reflect inconsistencies in data collection or recording. Additionally, the lack of certain key variables, such as weather conditions, waste volume, or nearby commercial activity, may lead to an incomplete understanding of the factors influencing waste generation and servicing. The model itself may be susceptible to biases if certain wards or asset types dominate the dataset, potentially limiting its ability to generalize across all neighborhoods. Furthermore, without spatial coordinates or robust time-series data, the analysis is constrained in identifying precise location-based trends or long-term patterns, which are crucial for optimizing waste management strategies.

To address these limitations, several next steps can enhance the quality and applicability of the analysis. First, improving the dataset by including additional variables, such as waste volume, commercial activity levels, and real-time updates, would provide a more comprehensive view of waste generation dynamics. Advanced machine learning models, like Gradient Boosted Trees or neural networks, could be explored to improve prediction accuracy while incorporating rigorous cross-validation techniques to mitigate overfitting and biases. Integrating geospatial data would enable precise mapping of servicing patterns, while time-series analysis could uncover seasonal or long-term trends. Operational simulations should also be conducted to evaluate the real-world impact of implementing model recommendations, such as optimizing bin placement or adjusting service schedules. Finally, an equity-focused evaluation could ensure that high-demand or underserved areas receive adequate resources, while sustainability initiatives informed by predictive insights could promote waste reduction and efficient resource use.

A Appendix

A.1 Data Diagnostics and Cleaning Process

A.1 Data Diagnostics and Cleaning Process To ensure the robustness of the analysis, the following data cleaning and diagnostic steps were implemented:

Missing Data Handling: Missing values in demographic datasets were imputed using median or mean imputation, depending on the variable type. Observations with incomplete or unclear STATUS entries in the litter dataset were excluded to avoid bias. Outlier Detection: Outliers in waste servicing frequencies and demographic variables were identified using boxplots and z-scores, with extreme outliers excluded or capped. Variable Standardization: Continuous variables, such as population size or waste collection frequency, were standardized to facilitate model interpretability. Data Integrity Checks: Cross-referenced litter bin data with ward demographic data to ensure correct mapping across wards.

A.2 Diagnostics for model

Figure 8 compares observed data (dark line) with replicated posterior predictions (lighter lines). The close alignment suggests that the model accurately captures the data's central tendency and variability. Figure 9 and Figure 10 show that the sampling algorithm used, the Markov chain Monte Carlo (MCMC) algorithm, did not run into issues as the posterior distribution for the model was created. Using the checks presented by (**citetellingstorieswithdata?**), both graphs do not show anything abnormal since the trace plots in Figure 9 display substantial horizontal fluctuation across chains, indicating good mixing, while the Rhat values in Figure 10 are close to 1 and well below 1.1, further supporting convergence.

A.3 Idealized Survey and Methodology

A.3.1 Survey Design

To evaluate waste generation and collection patterns in Toronto comprehensively, an idealized survey would focus on both primary data collection and observational methods to supplement existing datasets. This methodology would aim to bridge potential gaps in current data, such as individual household contributions, temporal variations, and socio-economic nuances.

1. **Objective**:
    - To capture detailed waste generation behaviors across diverse demographic and dwelling groups.
    - To assess the alignment of waste collection services with actual waste production.
2. **Target Population**:

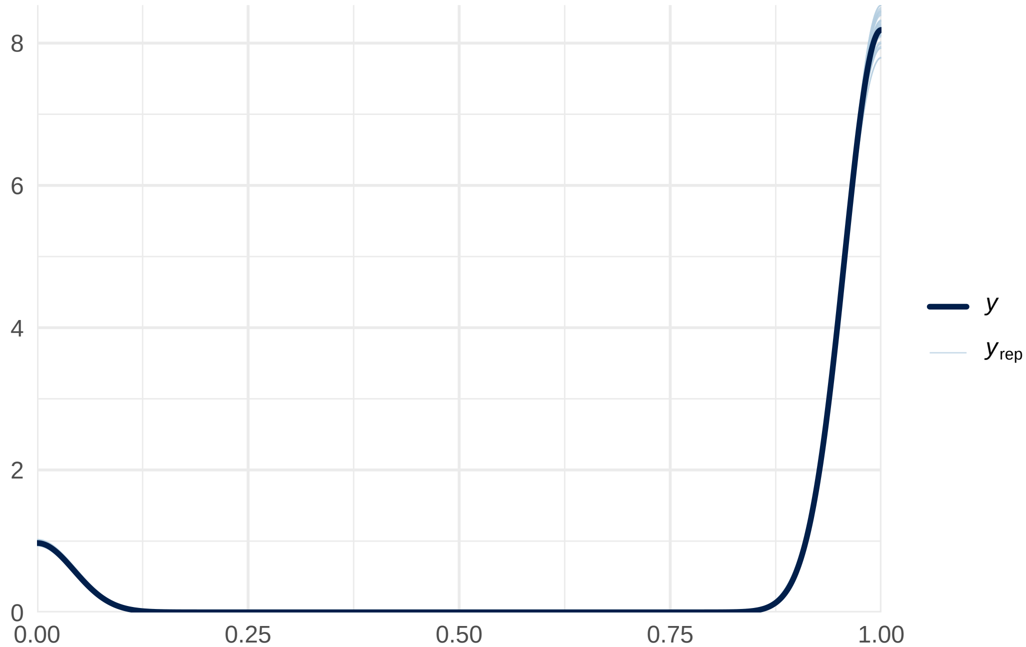

Figure 8: Posterior Predictive Check: Comparison of Observed and Replicated Data

Figure 9: Checking the convergence of the MCMC algorithm - Trace

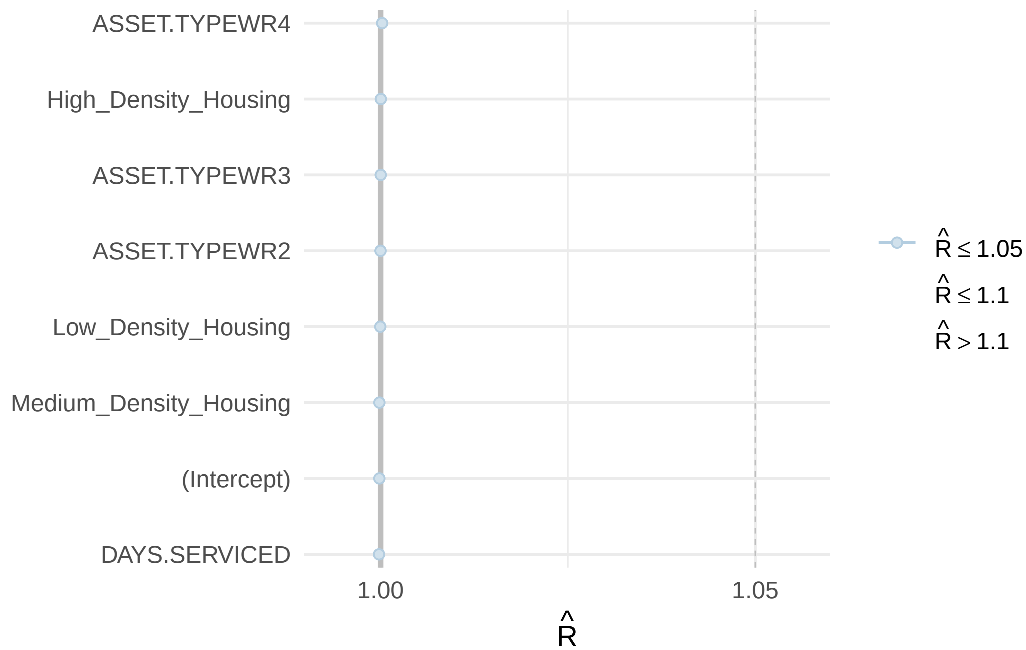

Figure 10: Checking the convergence of the MCMC algorithm - Rhat

- All 25 wards of Toronto, stratified by key demographic (e.g., age distribution, household size) and dwelling characteristics (e.g., low-density, medium-density, and high-density housing).

3. **Sampling Framework**:
   - **Stratified Random Sampling**: To ensure representation across wards and household types, the population will be divided into strata based on demographic and housing attributes.
   - **Sample Size**: A minimum of 500 households per ward to achieve statistical power while allowing granular analysis of trends.

A.3.2 Survey Instrument

The survey will consist of: - **Waste Metrics**: Questions on the type (recyclable, organic, landfill) and volume of waste generated weekly. - **Demographic Variables**: Age, household size, income bracket, and dwelling type. - **Service Feedback**: Satisfaction levels with current waste collection schedules and services. - **Behavioral Insights**: Practices like recycling habits, composting, and waste reduction strategies.

A.3.3 Observational Data Collection

1. **Direct Observation**: Trained enumerators will record waste bin fullness and mismanagement incidents in sampled neighborhoods weekly.
2. **IoT Sensors**: Pilot testing in high-density wards to measure bin utilization rates and waste type segregation.

A.3.3.1 Simulation for Optimal Sampling

- **Simulation Framework**: A simulation will evaluate the effectiveness of different sampling strategies. For example, comparing simple random sampling to stratified methods using metrics like bias and variance in waste volume estimates.
- **Tool**: Simulations will be run using R, leveraging the `survey` and `sampling` packages.

A.3.4 Data Linkages and Validation

1. **Cross-Validation**: Collected survey data will be cross-referenced with municipal records to ensure consistency and fill missing gaps (e.g., wards with outdated data).
2. **Linkage to Literature**: Studies on urban waste management and demographic influences will guide the interpretation of trends.

A.3.5 Anticipated Challenges

1. **Non-Response Bias**: Households unwilling to participate may introduce demographic skewness, particularly in low-income areas.
2. **Temporal Changes**: Data collected during specific seasons may not reflect year-round waste trends, necessitating longitudinal follow-ups.

A.3.6 Contribution to Measurement Validity

This detailed survey and observational framework would enhance the robustness of the study by incorporating granular, household-level insights. It would also provide a benchmark for validating predictive models developed in the main analysis, ensuring the alignment of theoretical estimates with real-world dynamics. This approach builds on the best practices in survey methodology while addressing gaps in the current data landscape.

References

Auguié, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Foundation, Python Software. 2024. *Python: A Programming Language for Clear and Concise Code*. https://www.python.org/.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Müller, Kirill, and Jennifer Bryan. 2020. *Here: A Simpler Way to Find Your Files*. https://cran.r-project.org/web/packages/here/index.html.

Planning, City. 2024. "Ward Profiles - 25 Ward Model." https://open.toronto.ca/dataset/ward-profiles-25-ward-model/.

R Core Team. 2023a. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

———. 2023b. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Services, Solid Waste Management. 2024. "Litter Bin Collection Frequency." https://open.toronto.ca/dataset/litter-bin-collection-frequency/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.