Recycling by NYC Neighborhood

CUSP Principles of Informatics - Fall 2020, 15 December 2020
Group members: Danielle Bayer, Sridevi Prasanna T, Brian Newborn, Amber Jiang

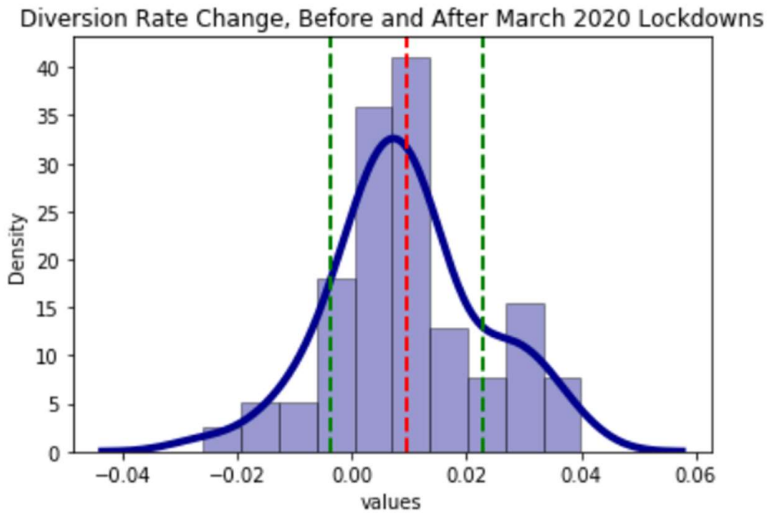
ABSTRACT

Recycling is key to reducing humans' negative impact on the environment. With the COVID-19 pandemic, at-home consumption has increased, and the need to understand which areas are improving recycling is increasingly important. The goal of this paper is to explore the change in the residential monthly recycling rate before and after the COVID-19 induced lockdowns in March 2020. We will be looking for a correlation between demographic data and the change in diversion rates from before to after the lockdowns. We expect to see that recycling rates changed differently across the city since the pandemic lockdowns and working patterns shifted. Identifying variables which correlate with the change in recycling rate will instruct city efforts on targeted outreach to improve recycling in the future. Additionally, should another lockdown be implemented, it is important to understand what areas are likely to see a negative shift in recycling for proactive outreach.

INTRODUCTION

Recycling rates in NYC have remained at 18% for the past several years. Since the onset of the pandemic, a large majority of the New York City workforce began working from home. The increased risk of going out in public led to exponential growth in online shopping for groceries, amenities and restaurant takeout. Disposable cleaning products and packaging from online shopping has greatly increased the amount of refuse generated in homes. The recycling rate has differed greatly across communities, often along demographic and economic dividers. COVID may have exacerbated these differences. Looking at the factors that contribute to this disparity may indicate where recycling education and infrastructure can be further targeted, as well as what communities are potentially in line to continue to increase recycling should there be another long period of lockdowns.

As seen below, the diversion rate change from before and after the March 2020 lockdowns follows a close to normal distribution. This delta was calculated by subtracting the pre-shutdown diversion rate from the post-shutdown diversion rate. A zero in the below chart indicates that a community district recycled after March 2020 at the same rate as before. See appendix for views of waste distributions before and after the Pandemic.



Distribution of change in recycling rates from pre to post March 2020 lockdowns. The average neighborhood had a slight increase in diversion rate from pre to post lockdowns.

LITERATURE REVIEW

New York City residents produce 12,000 tons of trash every day. Of the total trash produced by New Yorkers, only a fifth of it is recycled which includes 18% from homes and 25% from businesses[1]. This remains a problem today since Mayor Michael Bloomberg promised seven years ago to double the residential recycling rate to 30% by 2017. According to the Sanitation Department, if everything recyclable was sorted and recycled, 68% of residential trash and 75% of commercial trash could be kept out of landfills. This would be extremely important in reducing the emission of harmful pollutants into the environment and would save the city millions of dollars every year on trash transportation. Despite the financial and environmental incentives to recycle, recycling diversion rates vary greatly throughout the city's neighborhoods, ranging from 9 to 31% per district[2]. Meanwhile, the city-wide diversion rate averages at about only 20%.

There have been several analyses already done on the possible reasons for the disparity of recycling participation rates among New York City neighborhoods. One paper, published in Science Direct in February 2006, found four variables strongly correlated with low diversion rates: "percentage of residents below poverty level, percentage of adults without a high school diploma, percentage of households led by a single female with children and percentage of minority population[3]." The paper used a linear model to calculate a new measure which relates recycling behavior with demographic and socio-economic variables that may help predict recycling rates.

[1] Barnard, Anne 2020.
[2]Marjorie, Clarke and Juliana Mantay 2006.
[3]Marjorie, Clark and Juliana Mantay 2006.

The demographic and socio-economic variables include percentage of minority population, percentage of persons below the poverty line, percentage of adults without a high school diploma and percentage of households headed by a female with children. The regression analysis concluded that all four variables are strongly associated with diversion rates and therefore each of them could be used to "predict" the diversion rate.

A statewide assessment of Massachusetts' recycling program found several important demographic variables significantly predicted recycling rates. The paper used a stepwise regression analysis to determine which demographic variables were significantly associated with the five year municipal recycling rate. The following demographic variables were significantly correlated with municipal recycling rates: population size, mean household income, number of registered republicans, number of registered libertarians, percentage of white residents and percentage of black residents[4]. The analysis concluded that mean household income, population size, percentage of registered republicans and libertarians significantly predicted variance in municipal recycling rates. While, the number of Caucasian and Black residents did not significantly predict recycling rates.

Another paper published in 2016 by the Pew Research Center found that "the rules, practices and community norms around recycling vary considerably from place to place, contributing to dramatically different local recycling levels[5]." For example, people who live in neighborhoods where social norms encourage recycling are more likely to be aware of recycling rules and recycling options. A survey found that 28% of Americans say their local communities strongly encourage recycling while 22% said their communities don't really encourage recycling. The remaining half fall somewhere in the middle. In contrast to this, the Science Direct paper discusses how recycling rules clearly impact participation rates. For example, the diversion rate dropped from 35% in June 2002 to 21% a year later. This was the exact same time NYC decided to stop collecting plastic, glass and wax paper. This demonstrates that recycling behavior depends to a certain degree on the policies in place and that any program changes can be extremely disruptive, especially long term.

In a different 2017 report, researchers created a model that predicts weekly waste data for each of the 232 geographic areas in NYC[6]. The model can capture slight fluctuations in waste production associated with holidays, special events, seasonal changes and weather related events.

Media sources, in the early days of the lockdown period, compared the change in tons of household waste collected between March 2020 and March 2019. While the data showed a 4.1% citywide increase in collected waste, outliers showed significant decreases and increases by neighborhood, indicating population movement.[7]

DATA COLLECTION AND PREPARATION

The datasets published on NYC Open Data platform were used for our analysis of recycling diversion rates. The Monthly Tonnage Data includes the waste collected in tonnage by the Department of Sanitation from NYC residences and institutions. It contains 21.2K records of Borough-wise RefuseTons, PaperTons, MGPTons, Residential OrganicT ons, School Organics

[4] Seacat, Jason 2018.
[5]Desilvre, Drew 2020.
[6]Johnson, Nicholas 2017.
[7]Sandoval, Gabriel 2020.

Tons, Leaves Organic Tons, and Xmas Tree Tons. The American Community Survey (ACS) Socio-demographic and Economic Data includes detailed population and economic information. For the purpose of this project, the features chosen to analyze diversion rates are: population under 18, population over 65, white-population alone, black-population alone, hispanic, asian population alone, other race alone, number of employed, number of unemployed, individuals with: income less than 10k, income range 10k-14k, income 15k-24k, income range 25k-34k, income range 35-49k, income range 35k-49k, income range 50k-74k, income range 75k-99k, income range 100k-149k, income range 150k-199k, income >200k, individuals with SNAP benefits in past 12 months, number of people employed in: information sector, fire sector, science and waste sector, education-health-social sector, arts-entertainment-food sector, other sectors, public administration sector. These ACS fields ensured accuracy, given that they were from the US Census, and were relevant to possible waste trends in an area.

Data cleaning tasks included formatting date columns, creating separate columns for month and year. No duplicate data records were found. The monthly tonnage data is broken down by Community Board level. ACS data is broken down by Community Board Level.

To maintain commonality among datasets, an index was created for every row in the format Borough ID__Community Board Number where Borough IDs for Manhattan, Bronx, Brooklyn, Queens and Staten Island are 1, 2, 3, 4, 5 respectively and Community Board Numbers are represented by numeric values 01, 02, etc. for each Community District under each Borough. For example, index 3__04 will hold values for Brooklyn-Community District 4.

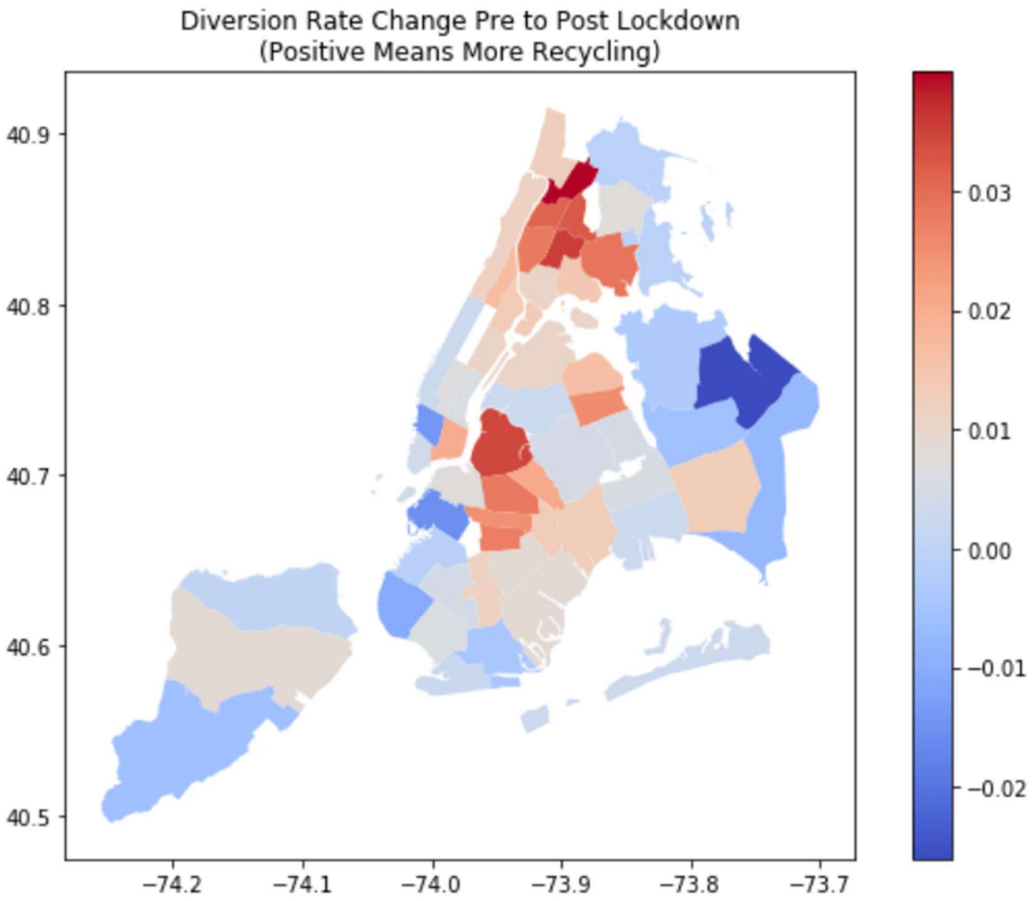
As mentioned above, our target variable, Change in Diversion Rate, is computed using Pre-Lockdown and Post-Lockdown Diversion Rates from Monthly Tonnage Data.

EXPLORATORY DATA ANALYSIS

Waste Tonnage collected in the years 2019 and 2020 is as follows[1]. There is a notable difference in the MGP and Organics collected from School and Residents.

Waste Types	2019	2020
Trash or Refuse	79.61%	79.84%
Paper	9.46%	9.23%
Metal, Glass, Plastic and Beverage Cartons	9.50%	10.47%
Organics from Residents	1.04%	0.03%
Organics from School	0.26%	0.09%
Leaves	0.08%	0.00%
Christmas Trees	0.05%	0.06%

Additionally, we were able to see a wide variance of delta diversion rate across the city. Below you can see how the changes in diversion rates tended to cluster in the Bronx and Western Brooklyn.



Geographic overview of change in recycling rates from pre to post March 2020 lockdowns. Notice how the best improvements tended to cluster in Western Brooklyn and the Bronx

FEATURE ENGINEERING AND MODEL TRAINING

As we are predicting a continuous variable, a linear regression is our ideal model. Linear regression also has the benefit of being highly interpretable, in that we can see which coefficients are most highly correlated with predicting our target variable.

In this instance, the demographic and economic data were all available for each community district in raw numbers. Additionally, we engineered two features. The first being a categorical representation of each district's borough, dropping the Staten Island column to avoid perfect multicollinearity. As borough recycling patterns could differ across the city, this felt like a useful input. The second engineered column was bins of the household income numbers to avoid too granular of household amounts. These were aggregated to 0-49K,50-100K,100-200K and 200K+ bins.

The data was randomly split into test and train sets at a ratio of 75%-25%, with a careful eye to ensuring by-borough splits for each. This meant that in our training set, we'd have 75% of

each borough's districts in training, and the remaining 25% in testing. This ensures that the model picks up all boroughs during training. Since the dataset was relatively small, cross training validation was done by running 50 iterations of the random test-train splitting by borough. For each iteration, the model was trained with the training set and then validated on the test set, with an average r2 value reported at the end.

REGULARIZATION

Because we had so many (25) input variables for ~40 districts (in the training set), we felt regularization was important to include in the pipeline so as to avoid overfitting. Lasso regularization was chosen to aid interpretability by zeroing out unneeded coefficients entirely. This also required scaling all data by a standard scaler so each regressor was on an equal scale. An optimal alpha, the regularization strength parameter, was derived by testing out-of-sample r2 across a whole host of values - ultimately, an extremely low alpha (implying weak but still present regularization) was consistently best at out of sample prediction.

RESULTS

Our regression model ultimately was able to achieve an out-of-sample, cross-validated average r2 of **0.36**, indicating moderate predictability of the change in diversion rate from our input features. Additionally, errors seemed to be somewhat consistent, as the model under-predicted boroughs with higher rates of change and over-predicted boroughs with lower rates of change.

FINDINGS & LESSONS LEARNED

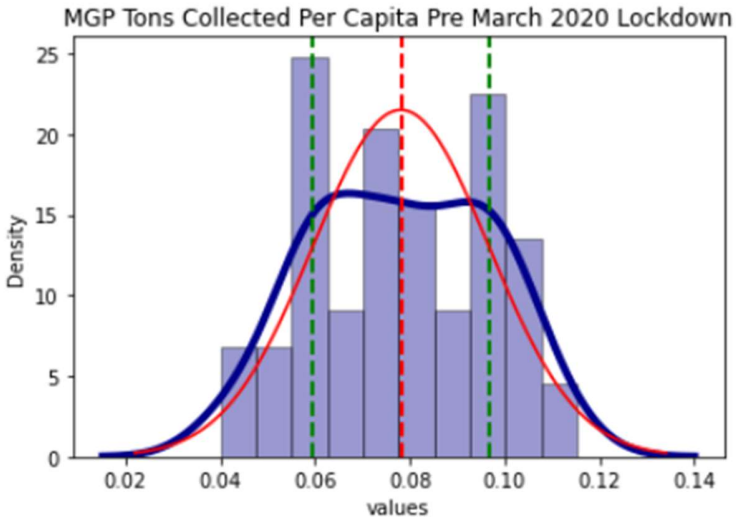
Coefficients that were most predictive of diversion rate included **having more households with 50-99K and 0-49K household income, number of households on SNAP programs, number of Hispanic households and having more residents under the age of 18**. See appendix for graph showing relative normalized strengths of coefficients that remained after regularization. The takeaway here is that characteristics associated with more diverse, lower and middle income districts had higher diversion rates post lockdown than before. There could be numerous explanations for this. One may be that households that previously were consuming more out of the house were now confined to their home and nearby areas, increasing their time at home and desire to recycle. Ultimately, the map of diversion rate changes aligns with this - the highest rates of improvement in diversion rate are seen in the Bronx, in diverse and low-to-middle income community districts.

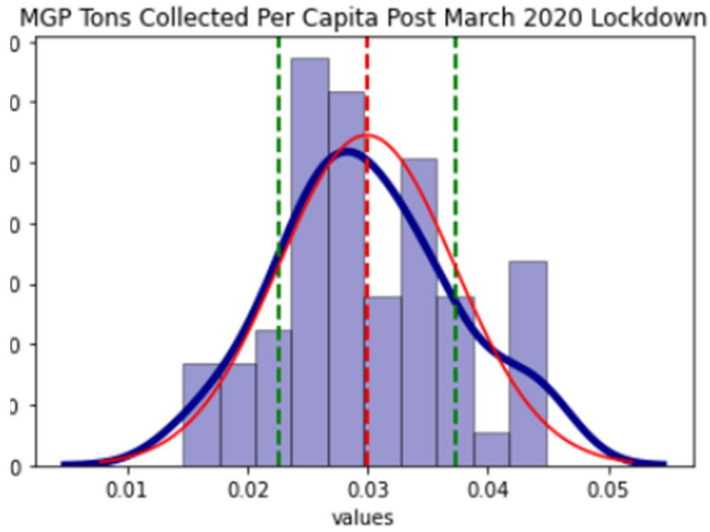
CONCLUSION

This model took a look at community recycling rates post COVID lockdowns from March to November. The short timeframe combined with the unprecedented social and economic lockdowns gives us a model that is aiming to predict large behavioral changes in a changing society. Given the high level granularity of the data that was used, this model gives researchers a targeted direction to apply more research. The data shows which communities increased or decreased recycling during COVID, and the model further illuminates communities that either exceeded or did not meet recycling rate expectations from the model. Further investigation into

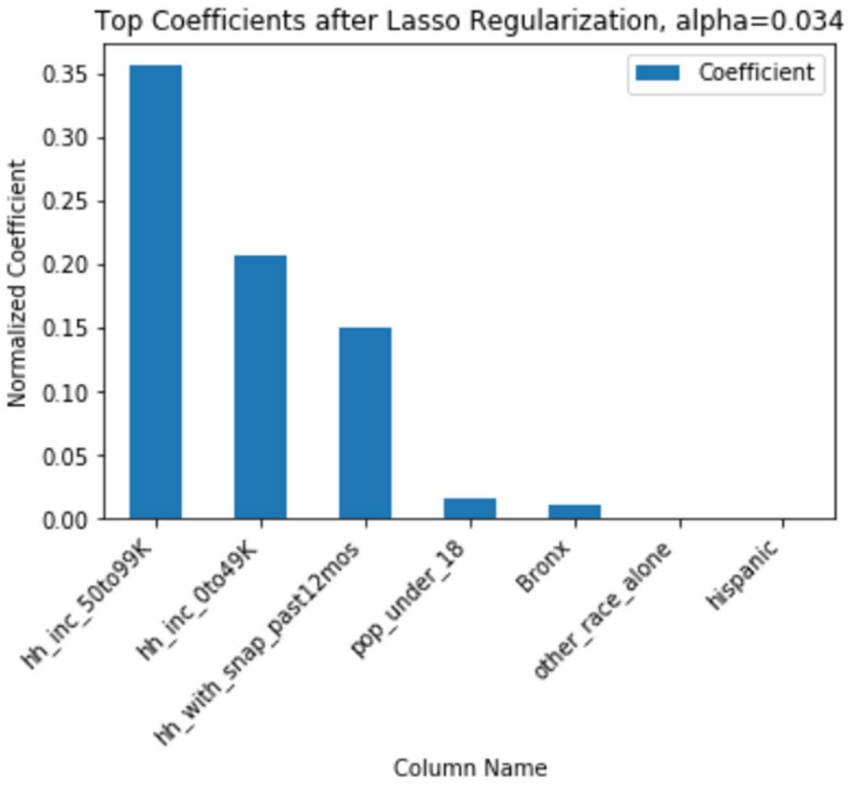
the specific methods, programs, education in place in these communities can shed light onto which methods were successful and which were not. Moving forward past COVID, many white-collar workers will shift from going into an office every day to a hybrid work from home and work from the office situation. The measures that were successful in raising recycling rates in communities can then be applied to other communities with similar socioeconomic attributes. As landfills are becoming fuller and resources becoming scarcer, it is vital to investigate how we can improve recycling in an ever changing society. Can we take advantage of the lockdowns to increase recycling behavior on a mass scale?

APPENDIX - Additional References





Distributions of MGP Tons Collected Per Capita Pre and Post March 2020 Lockdown



Coefficients remaining and their relative strength after lasso regularization. Notice how the number of households with 50-99K income and 0-49K income are the strongest predictors of increases in diversion rate. This suggests that the neighborhoods that were most able to improve diversion rate with 2020's changes in behavior were ones that had the most low-to-middle income families.

ROLES:

- **Data Collection & Cleaning:** Danielle, Sri, Amber, Brian
- **Data Collation/Engineering/Modeling:** Brian and Amber
- **Report Write Up and Data Interpretation:** Danielle and Sri

DATA SOURCES REFERENCES:

[1] https://data.cityofnewyork.us/City-Government/DSNY-Monthly-Tonnage-Data/ebb7-mvp5
[2] https://data.cityofnewyork.us/Environment/Recycling-Diversion-and-Capture-Rates/gaq9-z3hz
[3] https://data.cityofnewyork.us/Social-Services/Borough-Community-District-Report-SNAP-Population/jye8-w4d7
[4] https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Community-Districts/xi7c-iiu2
[5] https://furmancenter.org/coredata/userguide/data-downloads

Research References

Barnard, Anne. "7 Reasons Recycling Isn't Working in New York City." *The New York Times*, The New York Times, 29 Jan. 2020, www.nytimes.com/2020/01/29/nyregion/nyc-recycling.html.

DeSilver, Drew. "Recycling Perceptions, Realities Vary Widely in U.S." Pew Research Center, Pew Research Center, 30 May 2020, www.pewresearch.org/fact-tank/2016/10/07/perceptions-and-realities-of-recycling-vary-widely-from-place-to-place/.

Johnson, Nicholas. "Patterns of Waste Generation: A Gradient Boosting Model for Short-Term Waste Prediction in New York City." *Waste Management*, vol. 62, Feb. 2017, doi:10.1016/j.wasman.2017.01.037.

Marjorie, Clarke, and Juliana Mantay. "Optimizing Recycling in All of New York City's Neighborhoods: Using GIS to Develop the REAP Index for Improved Recycling Education, Awareness, and Participation." *Science Direct*, vol. 46, no. 2, Feb. 2006, pp. 128–148., doi:https://www-sciencedirect-com.proxy.library.nyu.edu/science/article/pii/S0921344905000960?via%3Dihb.

Sandoval, Gabriel, et al. "Garbage Pickups Tell a Tale of Two Cities, With Part of Manhattan Shrinking." *THE CITY*, THE CITY, 12 Apr. 2020, www.thecity.nyc/government/2020/4/12/21247125/garbage-pickups-tell-a-tale-of-two-cities-with-part-of-manhattan-shrinking.

Seacat, Jason D., and Nicholas Boileau. "Demographic and Community-Level Predictors of Recycling Behavior: A Statewide, Assessment." *Journal of Environmental Psychology*, Academic Press, 12 Feb. 2018, www.sciencedirect.com/science/article/pii/S0272494418300963.