We need to talk about NYC's trash

Walking around New York city you are almost immediately mesmerized by its sky high buildings and bustling cultural scene. However, down at eye level, it's equally hard to miss the overflowing trash cans and the garbage bags. In fact, trash in NYC can be somewhat mesmerizing in its own sense: the city produces 24 million pounds of trash daily, enough to fill the Empire State Building every two weeks. That translates to 2-3 pounds of trash per person per day.

While the fate of trash can seem like that of something entering a blackhole, *once it's out of our hands and into the garbage truck, it's as if the trash has escaped out of existence, gone forever, making no contact with us whatsoever again.*

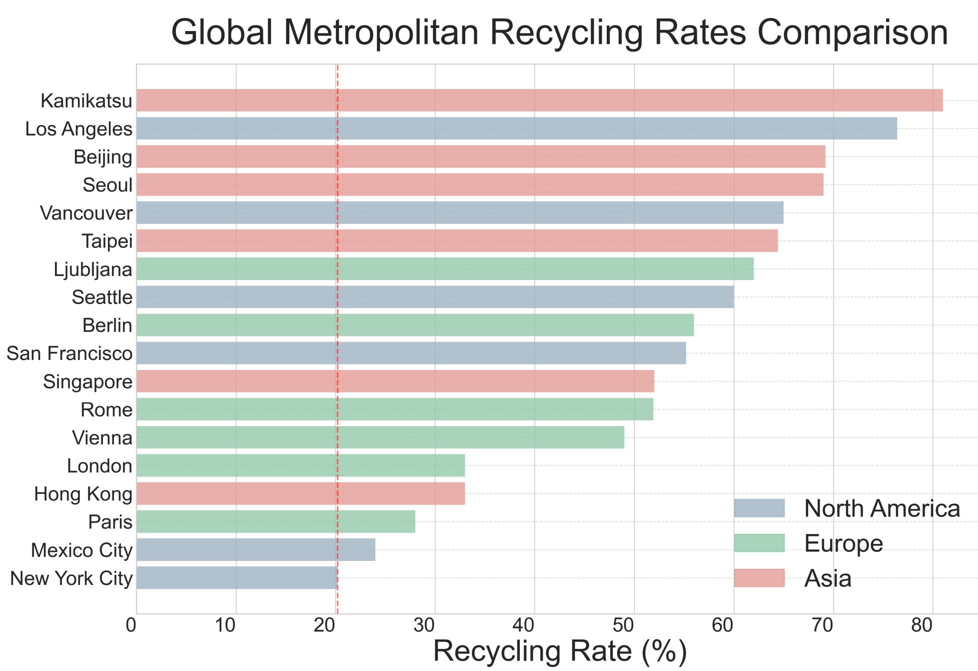
The reality is far from that.

The journey that trash takes, while less exciting than entering a blackhole, is a vastly complex one often depending on where it was generated, reflecting the area's vast population, diverse communities, and unique infrastructure.

As someone who has lived in various parts of the world–India, Japan, UK, Germany and now the US, I find waste and how we manage it a very interesting window into understanding the culture. This inclination to trash made me want to learn about the state of trash in the city and so I did.

I wasn't sure how the overflowing garbage cans and bags of trash translate onto recycling data on paper. To get a feel for things I looked up what typical recycling rates look like.

I was taken aback when I saw how low NYC's recycling rates were. While cities globally, Kamikatsu, Japan and San Francisco lead with impressive rates of over 80%. New York City was nowhere to be found, lagging behind significantly at just 18%.

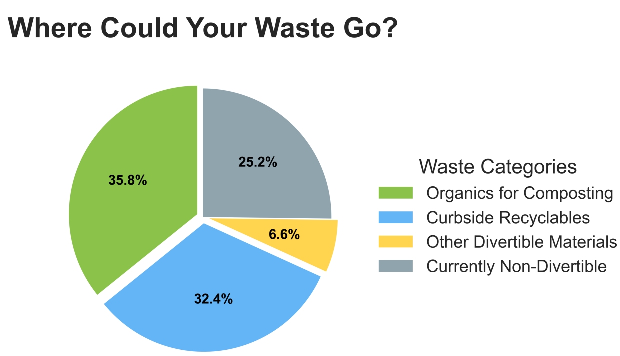


This disparity raised fundamental questions about what was happening with the trash in America's one of the most densely populated cities. New Yorkers generate approximately 14 million tons of waste annually, with about 3 million tons coming from residential sources - roughly 2,000 pounds per household per year being managed by thousands of trucks managed by the Department of Sanitation New York City (DSNY).

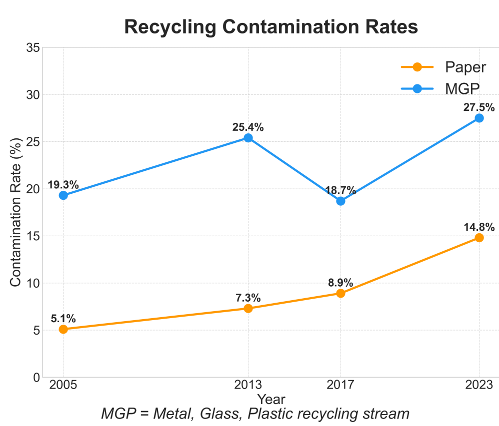
To understand the problem at hand, I looked the 2023 waste chracterization study put out by the department of Sanitaition NYC. The city conducts these detailed waste characterization studies every few years, methodically analyzing what's in our trash. These studies directly inform policy decisions and give precise breakdowns of the waste stream.

I learned some key facts, such as that the diversion rate–how much NYC could be recyling– currently at 18% could in principle be closer to 75%, which was promising.

But I also learned that within the city there was a huge disparity in terms of how the community districts performed.



Furthermore, even if materials were put out for recycling, they did not always get recycled. That is, sometimes materials would be contaminated and would lead to throwing away bags of potentially recyclable items to landfill. The data shows that contamination in the metal, glass, and plastic stream has increased from 19.3% in 2005 to 27.5% in 2023, while paper contamination has nearly tripled from 5.1% to 14.8%.



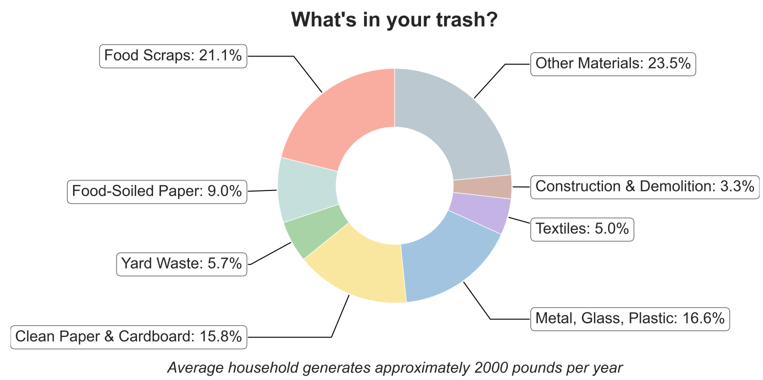
These apparent disparities are more than a waste problem–an emergent system begging for scientific analysis. Critical questions emerge: What fraction of trash is recyclable? In reality, what fraction of those recyclable trash actually get recycled? Why are some districts performing better than the others? and most importantly, what can we learn from these differences for the future?

These were some of the thousands of questions that were circling my brain.

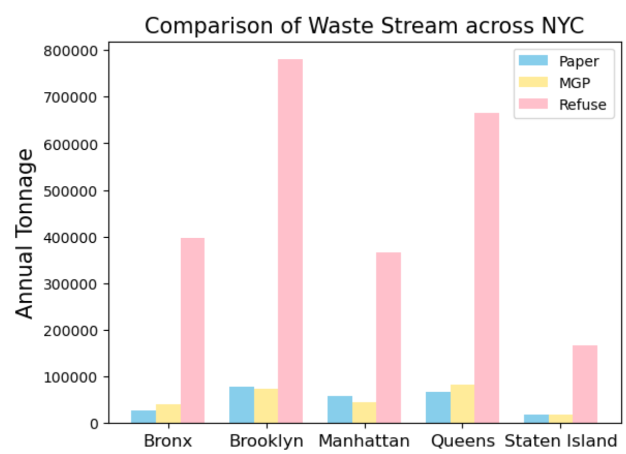
What did I do? I used data to converge on some questions and answers.

Thankfully there's tons of good data out there. NYC OpenData has ample data on recyling. Department of Sanitation New York (DSNY) publishes monthly data on waste collected per community district in the city. NYC Pop FactFinder is versatile tool with more than enough data on socio-economic features within each of the 59 community districts like median income, racial composition, gender composition, number of family house holds etc. And finally NYC.gov has a PLUTO tool that can be used to extract infrastructure data about the city. The links are all in the references.

First and foremost I extracted what does the trash of an average new-yorker look like?



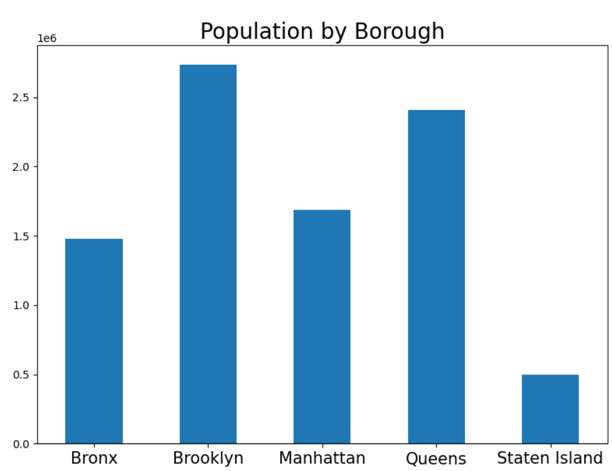
If we break this at the level of the 5 borough of NYC, this trash that is generated per day and collected by DSNY trucks looks something like this.



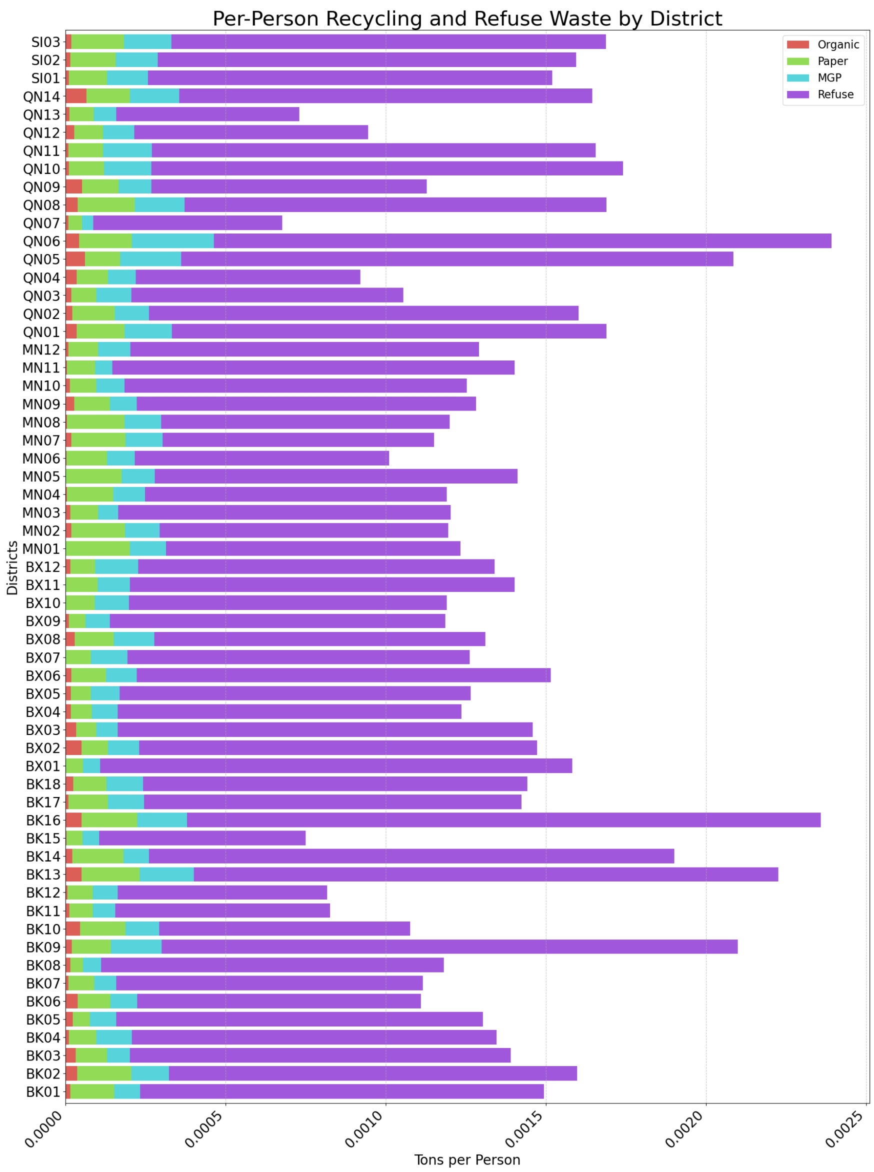
Total trash (Refuse) in Brooklyn is the highest followed by Queens, Bronx, manhattan and then Staten Island. The volume of recycling also followed a similar trend. My guess was that this difference in numbers comes from the sheer population. So I looked into the 2020 population data on the NYC population FactFinder.

I was surprised to see that my guess was not entirely right. Comparing Bronx and Manhattan, we can see that even though the population of the former is smaller, the volume of trash produced is more and the waste diverted towards recycling is lesser.

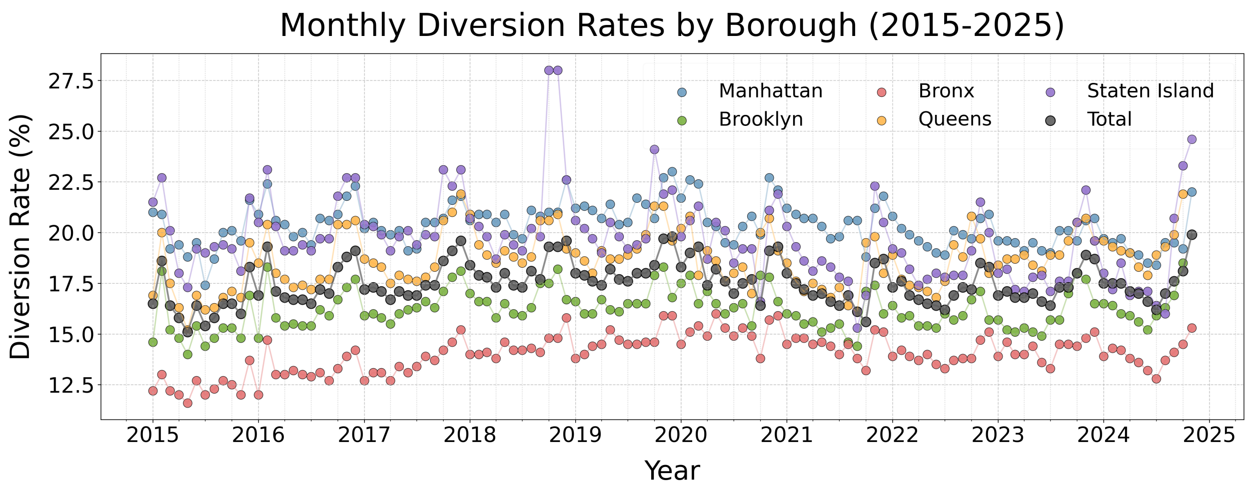
In fact, the waste produced and tonnage sent for recycling looks like per district per person is also highly varied. These results imply that everyone in the city is not producing trash uniformly and that the recycling habits are also different depending on the neighbourhood in discussion.



This was my first introduction to how disparate things are. This also helped me develop a more concrete question: *What drives recycling success in some communities but not others, and how can we quantify these factors?* Just like with any science problem, half of the answer lies in having good data. The challenge was that data on this topic was available in fragments and across different databases. The first step required mining, cleaning and compiling data that would allow exploration of the nexus of three things: waste generation and recycling data across the past decade, socio-economic and infrastructure data per district, and geometric data for each community district. This approach enabled slicing data across location, time and demographics.



The analysis began with examining how different districts perform over time. Two key metrics were chosen as target variables: diversion rate–the rate at which people are recycling–and capture rate–the amount of material that actually gets collected for recycling after contamination checks. The higher these variables, the better the performance.



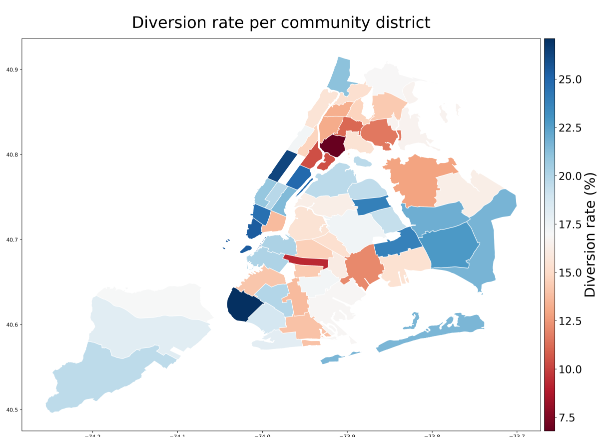
Looking at the data, we can see that the diversion rate has not changed significantly for the city.

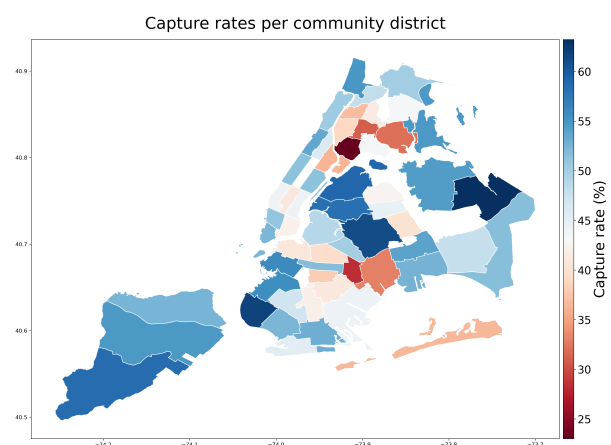
Certain districts performed better over time in recycling, with rates varying dramatically across different neighborhoods while the city's overall recycling rate hovered around 20%–well below the national average of 32%.

Manhattan and Staten Island generally perform better. Queens and Bronx have steadily been improving while Brooklyn has stayed about the same in the past decade.

If we pull out just the recent years, the data from 2019-2020 on the map of NYC looks like this

We could further look into the capture rate which represents what percentage of the materials could be properly recycled:

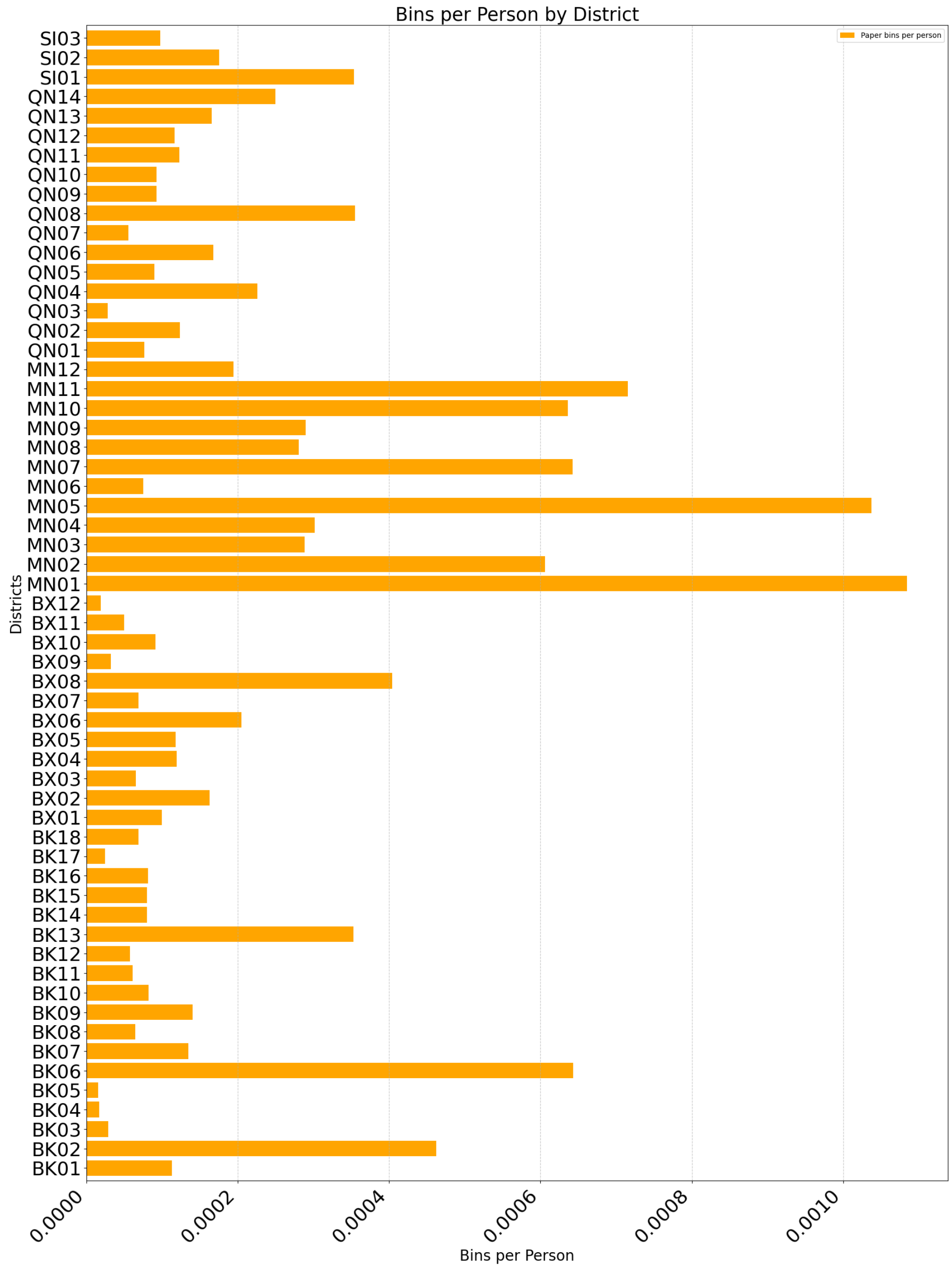




What maybe causing this?

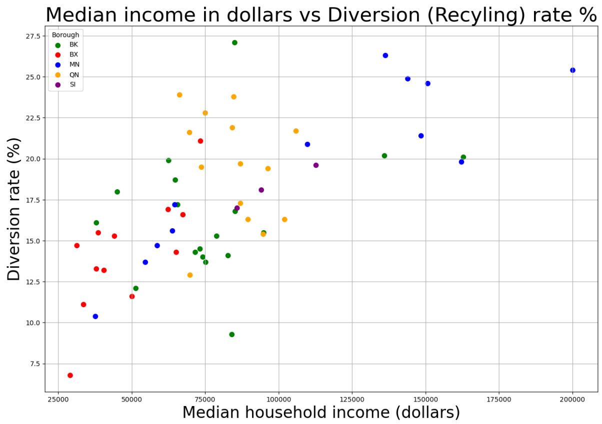
As a first guess I looked at the infrastructure. Does having access to more public recyling bins initiate people to adopt better recyling habits?

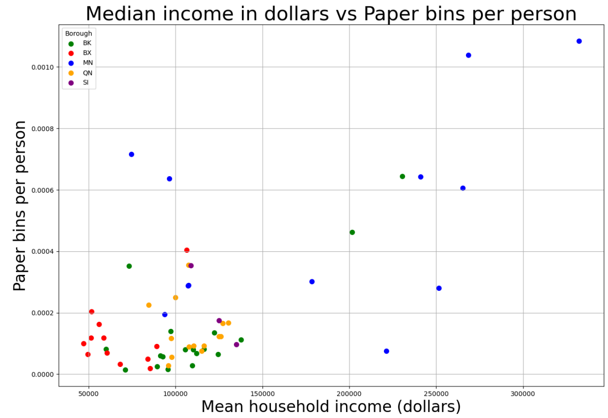
After extracting data on the location and number of bins per district, I learned a stiking thing: the number of bins per person varies dramatically across NYC districts. Manhattan districts like MN01 and MN05 have nearly ten times more bins per person than areas in the Bronx and Staten Island. Could this infrastructure disparity have real consequences for recycling outcomes? Potentially.

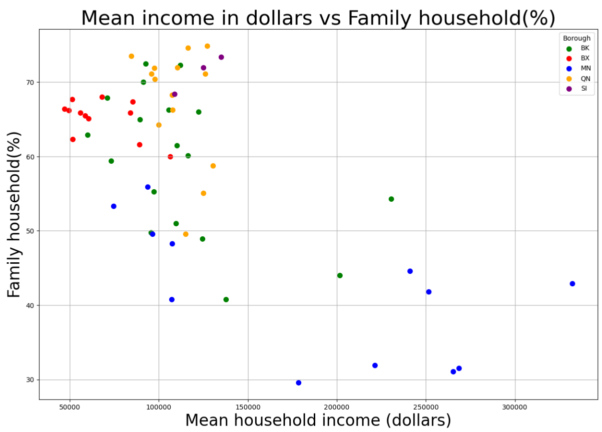


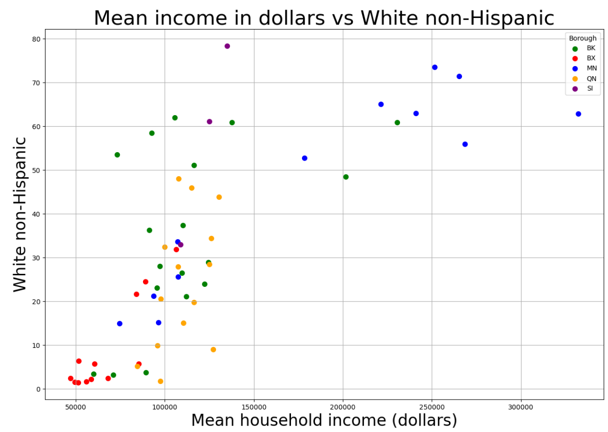
Infrastructure isn't the only factor. Some districts in Brooklyn and Manhattan achieve nearly 25% recycling rates, while others barely reach 12%.

When recycling rates are analyzed alongside socioeconomic factors, interesting patterns emerge through correlation analysis between recycling rates and factors like median income, education levels, population density, and building age.









One strong correlation stands out: as median household income increases, so does paper recycling. Higher-income districts consistently recycle more paper per person than lower-income areas. Through feature engineering–similar to identifying dominant forces in physics–the analysis identified which socioeconomic factors most strongly influence recycling outcomes.

We can combine these learnings into a multi-variate regression model. While scatter plots allow us to pul out correlations, thety do not tell us how to quantify what we are seeing and to run statistics on how confident we are about the correlations we see. Regression allow us to put a number to how confident we are about one feature affecting the other. We will demonstrate this with OLS regression stats below.

The model explains about 53.4% of the variation in diversion rates (R-squared: 0.534) The overall model is statistically significant (F-statistic: 8.335, p-value: 9.43e-07) Four variables have a statistically significant relationship with diversion rates:

Mean household income (positive effect, p¡0.001) Family household percentage (positive effect, p=0.006) Paper bins per person (negative effect, p=0.005) MGP (Metal Glass and Plastic) bins per person (positive effect, p=0.004)

1)Mean household income: For each additional dollar in mean household income, the diversion rate increases by about 0.00006621 percentage points. While this seems small, it means a $10,000 increase in income would be associated with about a 0.66 percentage point increase in diversion rate.

2)Family household percentage: A 1% increase in family households is associated with a 0.1335 percentage point increase in the diversion rate.

3)Paper bins per person: The negative coefficient (-584,500) suggests that more paper bins per person is associated with lower diversion rates. This seems counterintuitive but might indicate that areas with many paper bins have poor recycling practices or that the metric is capturing something else.

4)MGP bins per person: The positive coefficient (587,900) indicates that more mixed recycling bins per person is associated with higher diversion rates.

5)Non-significant variables: Hispanic percentage, White non-Hispanic percentage, and Housing units do not have statistically significant relationships with diversion rates in this model.

Assuming that waste and recycling practices follow trends in regions, I tried running a k-means clustering algorithm that. K-means clustering is an unsupervised method which means that based on these data points only the algorithm will select community districts that have similar relation ships betweeen variables and put them in a cluster.

I learned that the algorithm was putting districts with higher income into similar clusters as can be seen. We can also visualize how these clusters show up on the map of NYC.

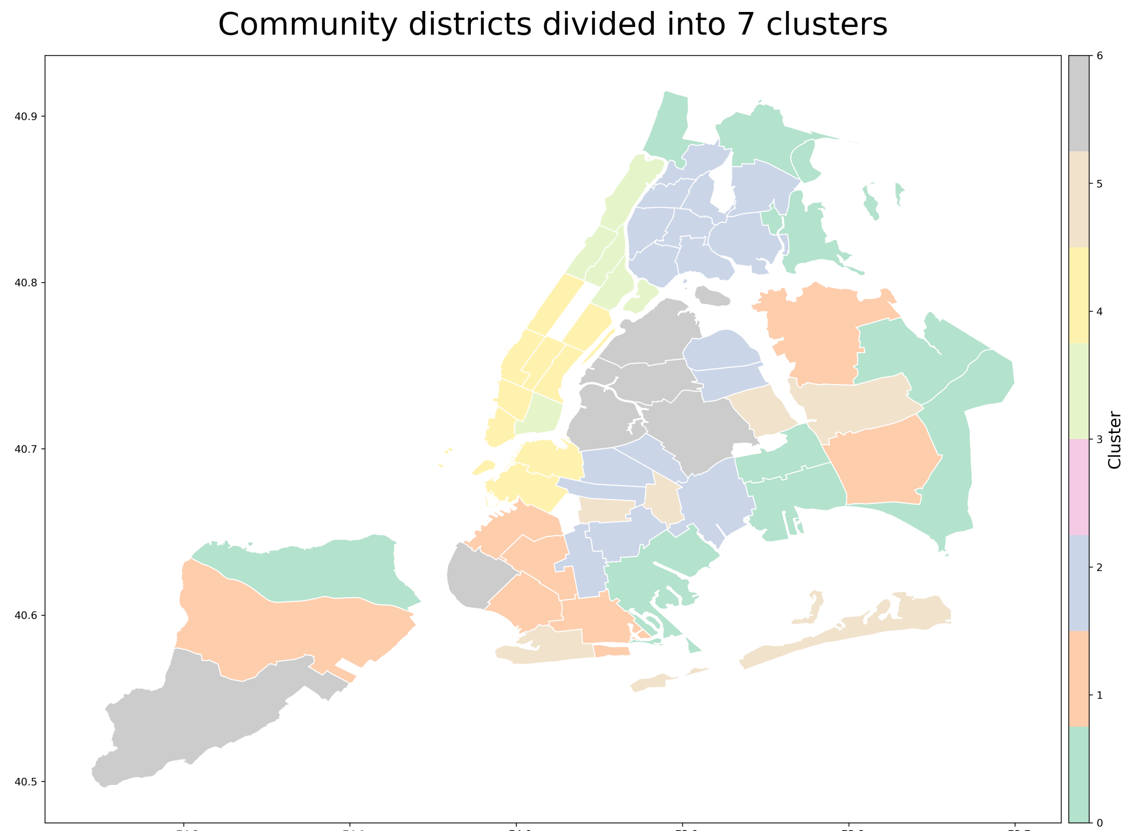
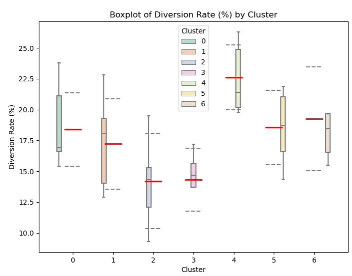
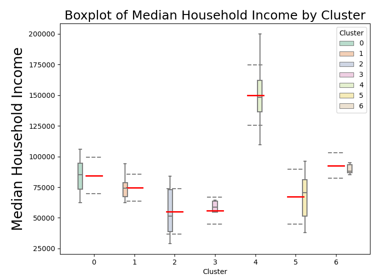


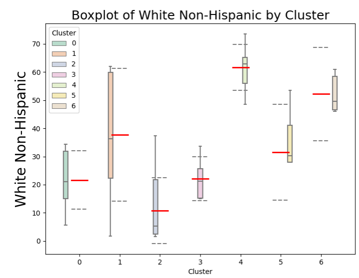
Figure 4: Spatial autocorrelation analysis showing neighborhood clustering and recycling islands

This analysis was exciting because it revealed neighborhood form recycling islands–communities that perform differently than surrounding areas.

After testing clustering patterns, unsupervised clustering algorithms, without explicitly identifying recycling rates as target variables, recognized districts with similarly grouped values. These represent neighborhoods that, despite being in different boroughs, show remarkably similar patterns in their recycling habits.



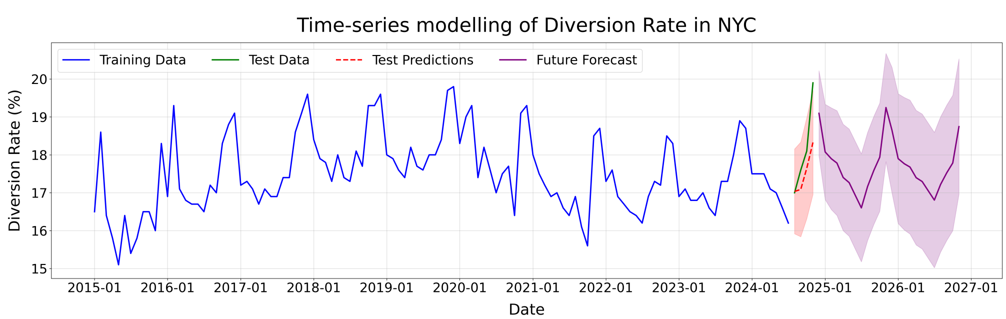




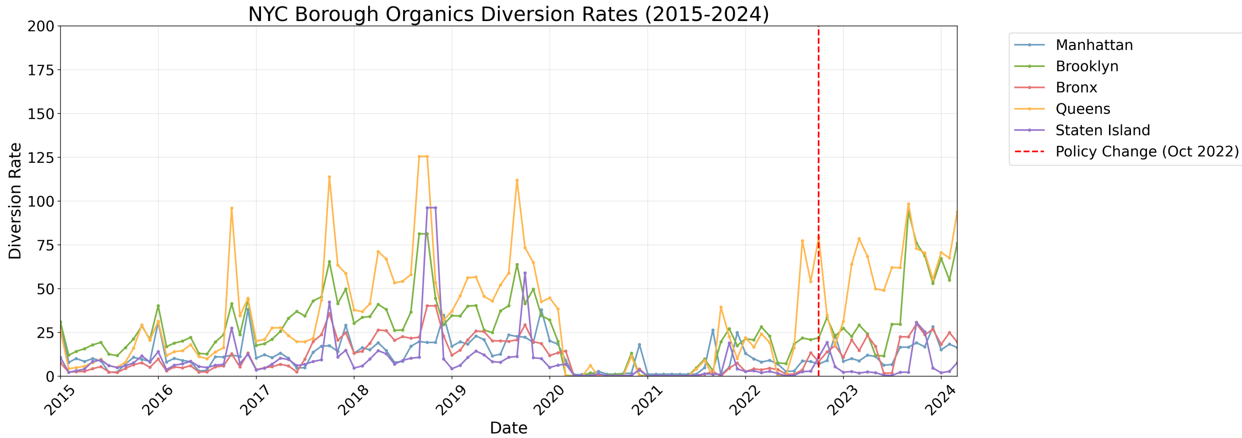
The 2023 waste characterization data claimed that policies work. After implementing plastic bag and foam product bans, NYC saw dramatic reductions. Plastic bag waste dropped by 68%, from 38.2 pounds per household in 2017 to just 12.3 pounds in 2023. Similar success occurred with foam products, which fell by 54%.

Time series analysis reveals how recycling patterns change over time, both understanding how numbers in the city have evolved and whether recent policies have had measurable impact. Slicing data over time showed how recycling rates across districts changed over the years.

I decided to use a SARIMA model to make predictions more concrete, the analysis captured seasonal trends in recycling rates and predicted rates for coming years. Validating the model by testing against recent months of 2024-25 data I could dvelop confidence in th model.



To analyze whether changes could be detected since the pilot composting program post-COVID in Queens was implemented, Queens served as base data compared to composting rates in other boroughs like Staten Island and Manhattan. While regression revealed correlations, it couldn't determine causality. To isolate policy change impacts, I used a quasi-experimental technique from econometrics called the Difference-in-Differences analysis.



The data and the DiD analysis confirm that with statisticaly significance we can conclude that the organics program had an order of magnitude improvement effect in Queens compared to the othr boroughs.

This is promising because all NYC households will have curbside organics collection by April 2025, representing a major step toward realizing this potential.

From a physicist's perspective, the waste management problem can be understood as an entropy problem. We use resources that are low entropy, derive work from them, and then discard high entropy garbage. Physics tells us that unless active intervention occurs, high entropy waste will continue to accumulate.

The story of NYC's trash is complicated, varying widely by location and influenced by countless factors. But through data science, unprecedented insights into these patterns are emerging. This data doesn't just inform large-scale policy–it can help individuals understand their community's recycling patterns, recognize the impact of personal habits, and identify opportunities for community organization and advocacy.

Data equips citizens to participate more effectively in local government, asking informed questions about waste management investments and priorities. By understanding where our trash goes and why, we can make more informed choices as individuals, communities, and policymakers–working toward a future where less waste ends up in landfills and more resources are recovered through recycling and composting.

The analysis reveals that systematic change is possible. Infrastructure investments, targeted policies, and community engagement all have measurable impacts on waste diversion. Most importantly, the data shows that with coordinated effort, NYC could transform from a recycling laggard into a leader, joining cities like San Francisco and Kamikatsu in achieving high diversion rates.

The physics of urban waste management may be complex, but the path forward is clear: combine robust data analysis with targeted interventions, measure results, and iterate toward better outcomes. In a city that generates enough trash to fill the Empire State Building every two weeks, even small improvements in recycling rates translate to massive environmental benefits.

1 Acknowledgements

UA thanks Dr. Azadeh Keivani and Dr. Amir Khosroshahi for the valuable guidance and discussions throughout this work.

All codes used in th analysis are available on github and can be freely used.

2 References

1. Climate Chance. Kamikatsu social project: zero waste objective." Accessed 2025. https://www.climate-chance.org/en/card/kamikatsu-social-project-zero-waste-objective/
2. South China Morning Post. "How Beijing's circular economy policy is driving recycling in carbon-intensive sectors." March 2023. https://www.scmp.com/business/article/3204001/how-beijings-circular-economy-policy-driving-recycling-carbon-intensive-sectors
3. Environmental Protection Department, Hong Kong. "Municipal Solid Waste Statistics 2023." 2024. https://www.wastereduction.gov.hk/sites/default/files/resources_centre/waste_statistics/msw2023_eng.pdf
4. Seoul Metropolitan Government. "Seoul to increase recycling rate to 79% by 2026." 2024. https://english.seoul.go.kr/seoul-to-increase-recycling-rate-to-79-by-2026/
5. Department of Environmental Protection, Taipei. "Waste Management News." Accessed 2025. https://english.dep.gov.taipei/News_Content.aspx?n=010EDE53B1F4A4AF&sms=85FBABE70858A8D4&s=98C0E979E6454330
6. National Environment Agency, Singapore. "Waste Statistics and Overall Recycling." 2024. https://www.nea.gov.sg/our-services/waste-management/waste-statistics-and-overall-recycli
7. European Environment Agency. "Municipal waste management in Austria." Accessed 2025. https://www.eea.europa.eu/publications/austria-municipal-waste-management
8. Statista. "Municipal waste recycling rate in Italy." 2024. https://www.statista.com/statistics/632923/municipal-waste-recycling-italy/
9. CPRE London. "Recycling Scorecard 2024." 2024. https://www.cprelondon.org.uk/news/recycling-scorecard-2024/
10. What is Green Living. "Paris Trash Problem: A Comprehensive Guide." 2024. https://whatisgreenliving.com/paris-trash-problem-a-comprehensive-guide/
11. European Commission. "Circular Economy Platform - Berlin Report." November 2024. https://circulareconomy.europa.eu/platform/sites/default/files/2024-11/459defe7-en.pdf
12. Statistical Office of the Republic of Slovenia. "Municipal waste statistics." 2024. https://www.stat.si/StatWeb/en/News/Index/11481
13. City of San Francisco. "Data - Recovery Rate." 2025. https://www.sf.gov/data--recovery-rate
14. Metro Vancouver. "Draft Integrated Solid Waste and Resource Management Plan Biennial Report 2022." 2022. https://metrovancouver.org/services/solid-waste/Documents/draft-iswrmp-biennial-report-2022.pdf

15. Global Plastic Action. “Mexico City.” Accessed 2025. https://www.globalplasticaction.org/mexico-city

16. Seattle Public Utilities. “Recycling Rates.” 2024. https://www.seattle.gov/utilities/your-services/accounts-and-payments/rates/collection-and-disposal/recycling-rates

17. City of Los Angeles Bureau of Sanitation. “Recycling Information.” Accessed 2025. https://sanitation.lacity.gov/san/faces/home/portal/s-l

18. NYC OpenData. “Recycling Data.” Accessed 2025. https://data.cityofnewyork.us/browse?q=recycling&sortBy=relevance&pageSize=20

19. Department of Sanitation New York (DSNY). “Monthly Curbside Collection Statistics.” Accessed 2025. https://www.nyc.gov/site/dsny/resources/statistics/dsny-monthly-curbside-collec page

20. NYC Planning. “Population FactFinder.” Accessed 2025. https://popfactfinder.planning.nyc.gov/#8.91/40.6934/-73.6286

21. NYC Planning. “PLUTO Data.” Accessed 2025. https://nycplanning.github.io/db-pluto/

22. NYC Department of Sanitation. “Waste Characterization Study 2023.” 2023. https://www.nyc.gov/assets/dsny/downloads/resources/reports/waste-characterization-studies/2023/wcs-2023.pdf