Harnessing Machine Learning Insights for Effective Policy Projections: A Study of Single-Use Plastic Bag Bans in Indonesia

Sandya Wijaya

ABSTRACT

Plastic pollution poses a significant environmental challenge globally, prompting governments to implement policies aimed at mitigating its adverse effects. In this study, I employ predictive modeling techniques to predict the effectiveness of plastic policies in cities of Indonesia where such interventions have not been previously employed. Specifically, I explored single-use plastic bag bans. I explored 9 different models but the best performing one was a XGBoost model with default parameters. I also performed causal inference analysis, using outcome regression and inverse propensity weighting to confirm that there is a causal link between plastic bag bans and plastic pollution. My findings reveal population as a key determinant influencing the impacts of plastic bag bans on pollution levels, that there is a 2 year delay between when a ban is implemented to when reported numbers start to change, and that coastal cities are more likely to see decreases in plastic if a ban is implemented. By elucidating the impacts of plastic bag bans effectiveness, this study informs strategic resource allocation and policy formulation efforts, contributing to the advancement of environmental governance practices in Indonesia and beyond.

KEYWORDS

Indonesia, policy, single-use plastic bag ban, predictive modeling, causal inference

INTRODUCTION

Plastic pollution has become a global environmental challenge. Past research has predicted that there will be 69.14 megatons of mismanaged plastic waste by 2025, and that mismanaged waste per capita will be highest in countries in the middle of the global income spectrum (Jambeck et al. 2015), showing that Indonesia stands out as an extreme case among developing nations facing disproportionate impacts. Yet, in the past, based on World Bank 2019 income categories, high- and upper-middle-income countries drove growth in the number of national policies adopted (Diana et al. 2022).

According to Indonesia Statistics, in 2016, Indonesia produced 65,200,000 tons of waste each year, 14% of which was plastic – totalling up to 85,000 tons per year of plastic waste. The country dumps 3.2 million tons of plastic waste into the sea, and a comparative study of previously published field observations ranked drains in Jakarta, Indonesia's capital, among the highest polluting rivers globally (van Calcar and van Emmerik 2019). High plastic pollution has endless adverse environmental health and biodiversity impacts: acting as a vector for disease in marine habitats (e.g. coral reefs) and causing the entanglement of marine creatures which restrict movement that can lead to injury and death (Watt et al. 2021), among other issues. Considering Indonesia is a collection of islands surrounded by water and located within the Coral Triangle – a hotspot for marine biodiversity highly susceptible to adverse effects of plastic pollution (Tomascik et al., 1997; Spalding et al. 2001; Lasut et al. 2018) – it is even more pressing that we move toward a feasible solution for plastic waste from Indonesia.

While scientific research has been steadily increasing over the past decade, there is a lack of literature on anti-plastic activism in Indonesia. Based on a search of the Scopus database on April 25, 2023, "Indonesia, plastic*, and pollution" yielded 203 publications (with 87% since 2019). Adding the word "politic*" to these search terms yielded only 1 article. Replacing the search term "politic*" with "activism," "advocacy," "civil society," "nongovernmental," or "nonprofit" yields only 1 result for "nonprofit" in Indonesia (Hermawan et al. 2021). Such results reflect the novelty of the issue in public discourse, and call for more research specifically on the politics of plastic pollution solutions.

To this end, my thesis aims to employ both statistical/machine learning techniques to predict how plastic pathways in various regions of Indonesia shift under specific policies, thus filling this scientific research gap of plastic pollution policies and its blockers. To answer this question of the effectiveness of plastic policies in Indonesia, I first did a literature review to understand the current plastic landscape in Indonesia – factors, causes, past solutions – then moved on to analyzing the causal effect and predictive modeling the effect of a chosen policy, and lastly answered what model is best to do this with. By predicting how an existing policy will affect a new area's landscape, this thesis can provide decision-making bodies with information that can help them better prepare for more effective outcomes. These findings can be applied not only to Indonesia but other parts of the world as well – especially developing nations grappling with environmental degradation amid resource constraints – to conclude what policies to implement, where, and what tools are best to utilize when doing so.

RESEARCH FRAMEWORK

Study site description

Indonesia is an archipelago with over 17,000 islands and a population of over 270 million. Some major islands include Java, Sumatra, Borneo, and Sulawesi, with cities like Jakarta, Surabaya, Bandung, and Medan, and the geography for each region varies from urban to rainforests to coastal. The country houses many environmental challenges, but especially within waste; Indonesia is ranked second in the world for its plastic waste contribution (van Calcar and van Emmerik 2019).

Over 80% of annual marine input of plastic around the world comes from land-based sources – the largest one being larger plastic litter such as everyday items like plastic bottles (Sherrington 2016). Studies show that inadequate municipal solid waste (MSW) treatment systems are the primary cause of marine litter in the Global South (Napper & Thompson 2018). Zooming into Indonesia, the same trend of large influx of land-based sources falls true. However, what exacerbates the problem in Indonesia is problems with waste collection and disposal. Informal collection of waste is legal in Indonesia, with roughly 50% collected informally through waste pickers, who are considered to be the last link in the waste management chain. In addition, separation at source is limited with only 7488 waste banks throughout the country (Wang, Y. and R. Karasik. 2022), limiting waste separation.

A brief on plastic solutions

Interview with Bustar Maitar (Founder of EcoNusa)

To gain insight on plastic solutions and the complexities tied to it, I interviewed Bustar Maitar in the summer of 2023 back in Indonesia. Maitar has 20 years of experience campaigning for social justice and environmental protection in Indonesia, working closely with communities on natural resources management – from the grassroots level to the global movement. He founded the Papua-based NGO, PERDU, in 1998, and was subsequently recruited as the first staff member of Greenpeace Indonesia office in 2005, and also founded EcoNusa Foundation in 2017 focused on forest conservation in Eastern Indonesia. The interview questions focused on getting first hand insights on what makes a solution effective in Indonesia's current landscape, and what blockers still exist.

One of the first things he shared to structure my conversation is that in general, plastic solutions come in 3 levels: local/consumer-level, business-level, national government-level. The complexities surrounding plastic in Indonesia encompass many categories and political intricacies.

Past and existing solutions

On the government level, Indonesia's timeline of plastic policies reveals a series of legislative efforts aimed at addressing the plastic waste crisis. Initiatives include the 2008 Solid Waste Management Act prohibiting open dump sites, though the 2013 target was not met (OECD marine plastic pollution in INDONESIA) shown by how the Ministry of Environment and Forestry recorded 167 open-dump waste disposal facilities that are still in operation as of 2018. Subsequent laws, such as the 2009 Environmental Protection Act and the 2012 Waste Bank Guidelines, emphasize waste reduction and recycling. Various presidential regulations from 2017 to 2019 outline national waste management policies, reducing marine litter, and implementing Extended Producer Responsibility (EPR). Most recently in 2019, Indonesia launched a collaboration with National Plastic Action Partnership (NPAP) aiming to achieve a 70% reduction in the country's marine plastic debris by 2025 (Wilson Center 2021).

On a business level, private companies have contributed successful waste management strategies. Unilever has assisted neighborhoods across 18 cities since 2008 in developing nearly 4,000 waste banks, providing training, tools, and systems to collect and sell plastic waste, creating an additional income stream for its members. In 2019, these waste banks generated $1.2 million for residents from 12,500 tons of recyclable waste, including 4,000 tons of plastic. Additionally, more companies are reducing plastic use through refill programs and changing the composition of their product packaging. Unilever launched its refill program in Jakarta's Saruga Bulk Store in 2020, and their products – Rinso, Bango, Sunlight, Love & Beauty Planet – now use less plastic through lighter packaging or 100% recycled materials.

Private investors have also caught on to anti-plastic awareness. Bintari Foundation has strengthened recycling policies in Semarang City, building profitable waste banks and neighborhood facilities for processing collected waste and recovering recyclable materials. Circulate Capital, an investment management firm focusing on circular plastic supply chains, has also invested in sustainable waste value chains, contributing to business-driven solutions for managing plastic waste in Indonesia (On the Frontline of Indonesia's Plastic Waste Crisis | Wilson Center).

At the local level, there are various actions showcasing regional commitment. Some examples of these actions include promoting circular economy practices through waste sorting management (Gerakan Pilah Sampah in 2019 by MoEF), implementing recyclables collection models, carrying out ocean plastic waste observations (MoEF 2021), and issuing local regulations like Regent Regulation No.13/2019 concerning Plastic Styrofoam Usage Reduction by Bogor. The ADIPURA Program, a clean city initiative, has been implemented as an incentive for municipalities excelling in environmental management and city cleanliness (MoEF 2020). Currently, 35 agencies, 35 cities, and 2 provinces have implemented single-use plastic bans at the local/regional level (OECD marine plastic pollution in INDONESIA) to discourage the use of plastic bags.

Complexities of plastic solutions

With respect to business-level interventions, business-level interventions are a smart target as private companies are able to implement changes without requiring confirmation from legislative bodies. However, challenges arise as profit-driven motives often take precedence over sustainability (Interview with Maitar, 07/2023).

For instance, companies like Danon and Unilever introduced smaller product sizes to cater to consumer preferences, prioritizing profit over more environmentally friendly options. Maitar shared a past experience when he recognized this pattern of private companies often only taking sustainability campaigns seriously when profit is at risk. He shared a specific case study of Barbie packaging sourced from natural forest timber in Indonesia, which he tackled by creating an eye-catching video campaign calling for Mattel to assess their excessive use of natural resources. The campaign increased awareness and this pressure from the public led to mobilization, causing Mattel to pull out of their contract with supplier Sinar Mas Pulp and Paper.

However, counterforces to anti-plastic activism exist, with powerful oil, gas, chemical, and plastics corporations influencing patrimonial politics. Such politics shift compliance costs onto consumers and marginalized communities – regular citizens who cannot afford basic needs now also have to carry the weight of overconsumption, recycling, and littering (Dauvergne 2018; Mah 2021, 2022; Omeyer et al. 2022). Scrap industries and informal cartels further complicate the landscape as powerful players that can lobby decision-makers and hinder policies to pass. The effects of these are magnified but go undetected with low government accountability and corporate transparency. A unique challenge especially present in informal settings is how regulators sometimes fail to enforce bans on single-use plastics (Dauvergne, P., and S. Islam. 2023) thus leading to inequitable and inconsistent implementation of environmental rules.

Choosing a policy

The ideal policy should integrate education and government-level interventions. Education plays a crucial role in fostering self-motivation, as being environmentally conscious is more challenging in random settings. Government-level interventions need to address both past plastic issues (e.g. past landfill) and current/future plastic challenges (Interview with Maitar, 07/2023). In determining the policy to be assessed in this paper, a careful consideration of various factors is crucial.

- Ambition vs feasibility – The selected policy should strike a balance between ambition and feasibility, both in terms of the paper's methodology and on-the-ground implementation. While it should be sufficiently ambitious to address plastic pollution effectively, it should also be practical and implementable within the existing social, economic, and political context.
- Substantial references – To build an accurate predictive model, the policy being assessed needs to have been adopted by a substantial number of places, as this means there will be

a greater number of references/training data for the model and thus better model performance.

- Longevity – Enacting the policy is not enough, maintaining it is the difficult part. Increasing transparency and expressing expected outcomes ensures that all stakeholders are committed for the long term. In tangent, a visible policy where individuals can see progress being made day-to-day encourages behavioral change among communities.

To formulate the best policy, an understanding of what the government is capable of is also essential. In the past, national policies primarily used regulatory bans against macroplastics and bags, which suggest governments may prioritize focused approaches with low economic enforcement costs – though geographic trends also play a role (Diana et al. 2022). However, other actions within their control include stopping the production and distribution of problematic plastics, supporting plastic-free business models, and introducing progressive legislation such as higher taxes on oil and petrochemical companies. However, the substantial investment required for Indonesia's waste management goals, estimated at $5 billion, presents a challenge, with the national government only able to fund about 10%. Private sector investment is crucial, but obtaining it is difficult due to factors like prevalent informal waste collection which generates inconsistent and unspecific numbers. (Wilson Center 2021).

More on chosen policy: single-use plastic bag bans

For this paper, I have chosen to focus on single-use plastic shopping bags (SUPBs) bans – encompassing full bans, partial bans or restrictions (allowing consumers to pay extra for a bag), and other classifications. SUPBs are a significant source of environmental pollution that can clog waterways resulting in flooding, degrade the visual and recreational appeal of landscapes as well as seashores. Such effects have resulted in anti-plastic bag sentiment and led to plastic bag taxes and/or ban being commonly employed tools (Muposhi et al. 2021).

Bans are an ambitious solution as it assesses the root of the plastic pollution problem by reducing the amount from the start, rather than justifying mass production. In addition, it also meets the requirement of having a large training set for this paper's predictive modeling methodology as a large proportion of Indonesia's provinces have a ban of some sort – the results from this paper will only work to make this action more widespread (other provinces enact it) and extensive (partial ban to full ban).

Lastly, bans can be coupled with other policies to increase its magnitude and balance its upstream impacts. Specifically, comprehensive regulation and economic instruments incentivizing behavioral change and associated information measures can tackle challenges that often arise from partial bans. For example, environmental education (Latinopoulos et al. 2018) for increased public acceptance and compliance with bans, reducing occurrences of exemptions. Top-down market-based interventions such as cleanups (Schnurr et al. 2018), and this increased tracking places pressure on enforcing restrictions on plastic production.

History of bans in Indonesia

Notably, Indonesia has experimented with plastic bag taxes, such as the trial of a IDR 200 per bag tax (~USD 0.01 per bag) across 23 cities in 2016 and the approval of an excise tax ranging from IDR 450-500 per bag (~USD 0.02 per bag) by the House of Representatives in 2020. Several cities in Indonesia have implemented single-use plastic bag bans, including Jakarta (specifically targeting transparent bags made from various plastics), Bali, Denpasar, Balikpapan, Bogor City, and Banjarmasin. Different areas of Indonesia enact plastic bans in various forms – degree of informality, availability of alternatives, public knowledge and awareness, and monitoring and enforcement mechanisms (Wang, Y. and R. Karasik. 2022).

Public Opinion

A 2023 study on plastic bans conducted interviews to shine light on the socio-economic and environmental complexities involved in implementing ban policies. Ban opposition responses reveal concerns about potential adverse impacts on economic revenue and employment, the cost of alternatives to single-use plastic products (SUPPs), the lack of waste segregation systems, and a limited uniform definition and understanding of biodegradable and compostable materials. Notably, critical voices point out that middle-class segments often prefer SUPPs, while poorer segments opt for traditional and cheaper alternatives, adding complexity to ban oppositions (Nøklebye et al. 2023). Other critical opinions talk about how bans often act as a "bandaid" as it simply temporarily prevents more plastic from being produced, but does not solve the past inventory of paste that has already been accumulated (Interview with Maitar, 07/2023).

METHODS

Data Collection

For my study, I collected mainly 3 types of data. The first type of data is plastic data from the Indonesia National Waste Management Database, *Sistem Informasi Pengelolaan Sampah Nasional (SIPSN)*. Specifically, I collected data from SIPSN on trash collected, and composition of trash. I then multiplied the proportion of plastic from each year with the amount of trash from each year to get the amount of plastic each year. SIPSN is a waste management system launched by The Ministry of Environment and Forestry (KLHK) only recently in 2021 on National Waste Awareness Day (HPSN). This platform manages data regarding the management of household waste and household-like waste in all districts and cities in Indonesia, and was created to make accurate, up-to-date data easily accessible by the public. However, something to note is that their numbers sometimes do not cover every scenario – in 2023, though it managed nearly 15,500,000 tons of plastic, this number only makes up ~68% of trash; 32% of trash was not processed (SIPSN 2024).

The second type of data I collected were indicators from the *Badan Pusat Statistik* (BPS). BPS is the Central Agency of Statistics in Indonesia, and is a non-departmental government institute responsible for conducting statistical surveys, annually investigating national and provincial socio-economics, manufacturing establishments, population and the labor force. Specific indicators that I used were demographics such as population and land area, economic indicators such as Gross domestic product (GDP), and socioeconomic information such as Gini Index and Human Development Index (HDI). I chose these specific features as indicators because the combination of these factors would cover as many aspects of Indonesia and create a comprehensive picture of the nation.

Gross domestic product (GDP) is a monetary measure of the market value of all the final goods and services produced and rendered in a specific time period by a country or countries. GDP is more often used by the government of a single country to measure its economic health. Gini index is a measure of statistical dispersion intended to represent the wealth inequality within a nation or social group. It is a coefficient calculated from the Lorenz curve with the cumulative percent of income against the cumulative percent of population; by dividing the area of inequality over the area under the perfect equality reference line. The larger the Gini index, the less equitable a group is – 0 represents perfect income equality; 1 represents perfect income inequality where one person has all the income and everyone else has none. The Human Development Index (HDI) is a summary measure of human development. It measures the average achievements in a country in three basic dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. The HDI is the geometric mean of normalized indices measuring achievements in each dimension (World Bank).

The last data I collected was information pertaining to bans – specifically, on which areas of Indonesia enacted bans and when such implementation started. Since there was no pre-existing dataset with this information, I built my own dataset with 2 columns 1) highest restriction that regency/city has ever implemented, and 2) start date. When that restriction was implemented. I collected data from various literature, specifically *PlasticDiet*'s collection of documents and bills pertaining to these bans. I also read news articles dated from around the time it was implemented to find out any needed additional context.

I collected each source in its own CSV files. With this, the final 8 datasets I will work with are plastic, population, land area, GDP, Gini, HDI, and bans. Each file covered the time range of 2017-2022 because 2017 was the oldest year with total overlap from every source for each feature, and some features did not have data for 2023 so 2022 was the most recent comprehensive year. Each row in the CSV represents information for each city in each year – for example, DKI Jakarta's GDP in 2020.

Cleaning & Merging of Data

The methodology for creating the final dataset I was working with involved 3 main steps: 1) preparing indicator dataset 2) preparing plastic dataset and 3) merging all datasets together. All data cleaning, merging and analysis was done in python.

First, I urge my various features into one cleaned indicator dataset with all the indicators (representing the estimators in my model). In order to merge the various features into one indicator dataset, the name of the regency/city represented in one dataset had to *exactly* match how it is represented in a different dataset. However, this was not the reality for my datasets initially – each regency/city was represented differently through different spellings ('D I YOGYAKARTA', 'DAERAH ISTIMEWA YOGYAKARTA'), spacings ('BANYU ASIN', 'BANYUASIN'), alternative names ('MALUKU TENGGARA BARAT / KEPULAUAN TANIMBAR', 'KEPULAUAN TANIMBAR'), and so on.

To address these disparities, a matching dictionary was created using the *difflib* library, where each key in the dictionary represents the standardized name and the value in the dictionary was all the different variations of the name. Some issues arose such as duplicate matches, and there were some instances where towns listed in some dataset lacked corresponding entries in other datasets. These discrepancies were identified and resolved manually. The solution code and record of initial missing value as well as duplicate issues can be found in Appendix A.

After addressing all identified issues and ensuring 514 unique matches, the merged dataset with all the features as columns now consists of 3,084 rows, covering the span of six years from 2017 to 2022 without any missing values. Quality control measures were implemented to detect and address any occurrences of missing values, duplicates, and column reordering to enhance user experience.

The second step is to prepare my plastic dataset (representing the target variable) by cleaning out any missing values. The first run at removing missing values involved imputing them using the mean of all other data points in that province for the given year. For example, since the Regency/City 'Kab. Aceh Selatan' had a missing value in 2022, this missing value would be filled by the 2022 average of all the regency/cities in the province it belongs to; Aceh.

There were still some missing values that persisted – for certain cities that had no province data in that year, therefore there was nothing to average. This included *DKI Jakarta in 2021, Papus Barat Daya in 2022, Papus Barat Daya in 2021, Papua Pegunungan in 2021, Papua Tengah in 2022, Sulawesi Barat in 2022, Sulawesi Barat in 2021***.** For these cases, I imputed missing values using the average value across all available years for the respective province. For example, to impute Regency/City 'Kab. Adm. Kep. Seribu' missing value in 2021, since I could not use its province DKI Jakarta's 2021 average, I imputed this missing value instead by averaging DKI Jakarta's 2019, 2020, 2022 data. These two runs at imputing missing values eliminated all the empty data except one row – 'Kab. Lanny Jaya' in province 'Papua Pegunungan,' which only had data for 1 year (2021) available and because that one year has no value, there is no reference available for imputation. This row was dropped for further analysis.

The final step involved merging the indicators with the plastic dataset as well as the ban dataset I created for the overlapping time frame of 2018-2022. I also checked for any duplicates and missing regency/cities that resulted from merging, then addressed them in a similar way as described above.

Exploratory Data Analysis and Feature Engineering

The first thing I investigated in exploring the data is the effect of various restrictions on plastic. To do so, I plotted plastic per year in tons against years since the start of the restriction.

Note that it is vital to use years since start as the time x-axis instead of year because not all bans start at the same time. If years is used as the x-axis, some years will have significantly larger total plastic simply because there are more cities that get categorized as having bans – as the years pass, more cities enact bans, so the number of cities included in the average ban cities' plastic will increase, and thus this total count will increase. To prevent misunderstanding these increases to be attributed to an ineffective ban, I will instead use years since the start of the restriction as the x-axis, so the starting point is fair – only then can I do any quantification.

Before proceeding with modeling, it is essential to ensure two key aspects: 1) the presence of relationships between variables, and 2) the absence of collinearity. I plotted my variables against the target variable to indicate the potential existence of linear relationships among features, necessitating data transformation for clearer interpretation (Figure 1).

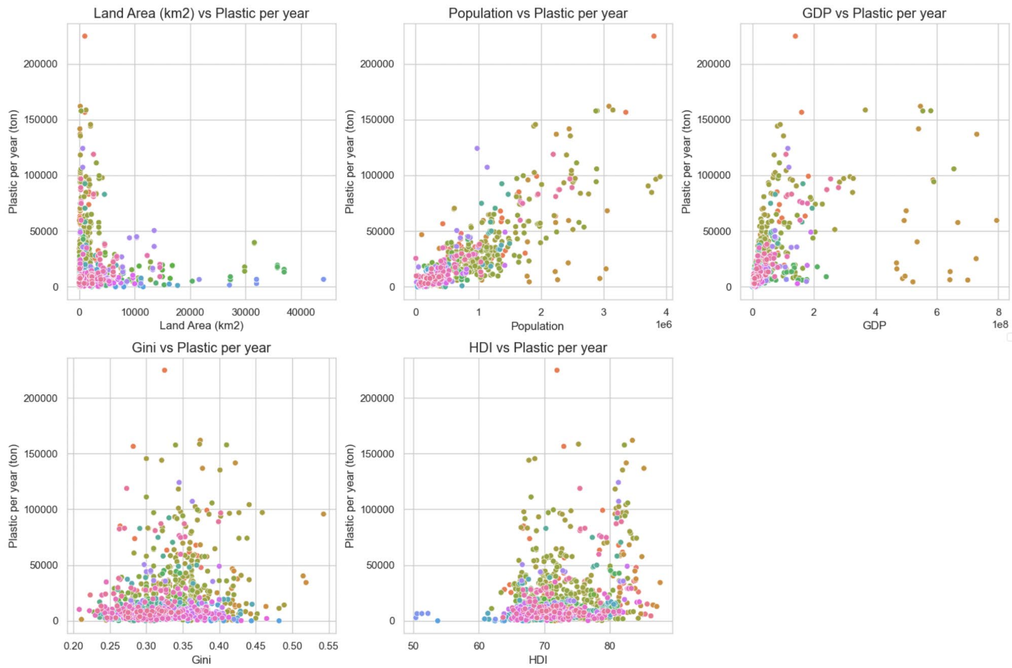

Figure 1: Each indicator against the plastic target variable. Created using Seaborn library in Python.

The shape of these graphs guides my feature engineering efforts. A correlation matrix is employed to reveal relationships between variables (Figure 2).

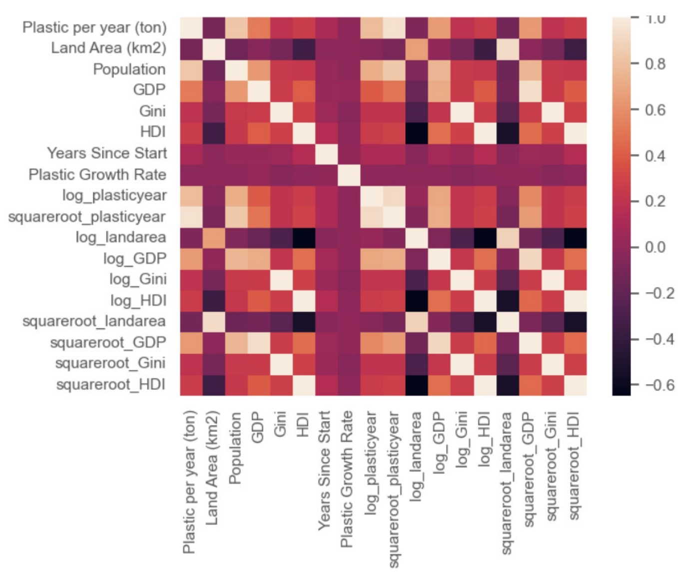

Figure 2: Correlation matrix of all features. Created using Seaborn library in Python.

Upon inspection, certain relationships, such as the correlation between log population and log plastic, exhibit higher coefficients, suggesting the utilization of transformed variables for improved model performance. Collinearity, or high correlation between features, is assessed to avoid redundancy in my model. Notably, the variables with the highest correlation are population and plastic per year, indicating their suitability as features for estimating the target variable. However, all other combinations demonstrate correlation coefficients of 0.7 or lower, indicating the absence of collinear features. The relationship between indicators and plastic pollution is further explored, with the identification of optimal combinations for each indicator: land area undergoes square transformation, while population and GDP are log-transformed. Finally, HDI remains untransformed due to its inherent nature.

Causal Inference

One can measure the causal effect of a binary treatment Z on an outcome Y by considering the potential outcomes $Y(0)$ and $Y(1)$, which represent thought experiments about what would happen if the treatment was or was not applied.

Average treatment effect (ATE) represents the causal effect of a treatment Z on an outcome Y. In general, one cannot estimate this without making assumptions. ATE is represented by the Greek letter tau (τ), is defined as:

$$\tau = E[Y(1) - Y(0)]$$

In the special case where the data come from a randomized experiment, one can leverage the fact that the participants were randomly assigned to the treatment group, i.e has a ban ('treat' = 1) and the control group, i.e. does not have a ban ('treat' = 0). If this was the case, there will be an unbiased estimate of the true average treatment effect. However, in reality, most cases are observational studies, meaning we want to estimate the causal effect of a treatment without randomization. One issue with this is that because one can never observe the counterfactual potential outcomes, in an observational study we cannot check if our answer is correct.

To ensure the observational causal inference tools work on real data, one can replace the randomized control group with non-random observational data, then apply causal inference methods to the true treated group and the observational control group – similar to what is done in a normal observational study.

I attempted 3 techniques that utilize unconfoundedness to estimate the treatment effect in an observational study like my dataset – Outcome Regression, Matching, and Inverse Propensity Weighting. The unconfoundedness assumption means one observes all the relevant confounding variables (like years of education); there are no unobserved confounders.

Before investigating unconfoundedness using these techniques, I first identified possible confounders. A variable is a confounder if it is correlated with both the treatment (ban) and also the outcome (plastic). By comparing histograms of a variable in the treated (cities with bans) and untreated (cities without bans) groups, I visually inspected whether there are noticeable differences in the distributions. Noticeable differences in the histograms between the two groups suggest that the variable may be associated with both the treatment and the outcome – and the variable may be a confounding variable. Some aspects to consider when comparing distributions are range and spread, height and density, shape and distribution, and alignment. First, I visually inspected the distribution of data and used statistical tests to assess normality (Figure 3).

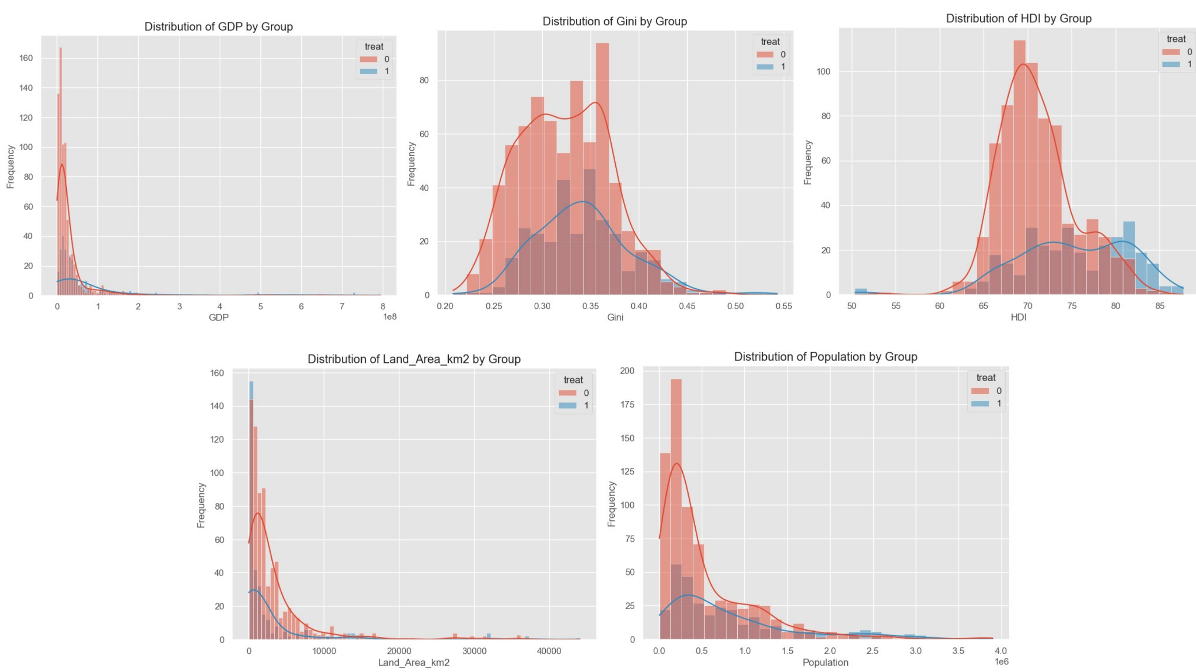

Figure 3: Distribution of each indicator in ban group and non-ban group. Created using Seaborn and Matplotlib library in Python.

I also confirmed this formally through the Shapiro-Wilk test, which defines whether or not a dataset is normally distributed. After confirming that my data is not normally distributed –

meaning it does not meet the assumptions of the t-test – I then used the Mann-Whitney U test to assess whether two independent samples come from the same distribution. If the p-value is less than my chosen significance level of 0.05, I reject the null hypothesis and conclude that there is a statistically significant difference between the groups. This acts as a quantitative way to help confirm whether the observed differences between the treatment and control groups are statistically meaningful – and in effect determine if it is a possible confounder. I found that all of them have statistically significant differences between the ban group and no ban group – so they are all possible confounders. The results of these tests can be found in Appendix B.

Technique 1: Outcome Regression

Now that I know the full set of variables (population, land area, GDP, Gini, HDI) are confounders in this problem, I can make the unconfoundedness assumption, where X represents the collection of all 6 confounding variables. I fit a linear model of the following form:

$$Plastic = \tau * Z + a * Population + b * LandArea + c * GDP + d * Gini + e * HDI$$

Then, under the following assumptions, the estimated coefficient of treatment from OLS, $\hat{\tau}$, is an unbiased estimate of the ATE:

1) Assume unconfoundedness given this set of 6 variables.
2) Assume this new linear model correctly describes the interaction between the variables.

```
                            OLS Regression Results
==========================================================================================
Dep. Variable:     Plastic_per_year_tons   R-squared (uncentered):                  0.804
Model:                               OLS   Adj. R-squared (uncentered):             0.803
Method:                    Least Squares   F-statistic:                             841.1
Date:                   Sat, 20 Apr 2024   Prob (F-statistic):                       0.00
Time:                           10:22:09   Log-Likelihood:                        -11357.
No. Observations:                   1033   AIC:                                 2.272e+04
Df Residuals:                       1028   BIC:                                 2.275e+04
Df Model:                              5
Covariance Type:               nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Population        0.0336      0.001     36.741      0.000       0.032       0.035
Land_Area_km2     0.0400      0.087      0.458      0.647      -0.131       0.211
GDP            -1.18e-06   6.22e-06     -0.190      0.850   -1.34e-05     1.1e-05
Gini          -1.922e+04   9537.231     -2.016      0.044   -3.79e+04    -508.509
HDI              75.2695     43.085      1.747      0.081      -9.275     159.814
==============================================================================
Omnibus:                      327.858   Durbin-Watson:                   1.445
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5823.860
Skew:                           0.980   Prob(JB):                         0.00
Kurtosis:                      14.466   Cond. No.                     2.20e+09
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 2.2e+09. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 4: Ordinary Least Squares Regression Summary. Created using statsmodels library in Python.

Let's assume that I had correctly specified my linear model (Figure 4), and that both assumption 1 and assumption 2 do hold. Even then, there is still a chance that my estimated treatment effect $\hat{\tau}$ won't exactly line up with the true treatment effect, and this is because of randomness in the data (according to the Frequentist perspective). Depending on the observational data that I have, I can calculate different estimates of the average treatment effect.

One way to account for this uncertainty in my analysis is to create an interval that will, with 95% confidence, contain the true average treatment effect. To do this, I employed the bootstrap. I generated 1000 bootstrapped estimates of the ATE and found the 95% confidence interval to be a range from 3968.0 to 3968.0. These results indicate that the estimated average treatment effect (ATE) is 3968.0, and the 95% confidence interval for this estimate ranges from 3968.0 to 3968.0. This means that, based on the bootstrap sampling, the ATE is estimated to be 3968.0 tons of plastic per year.

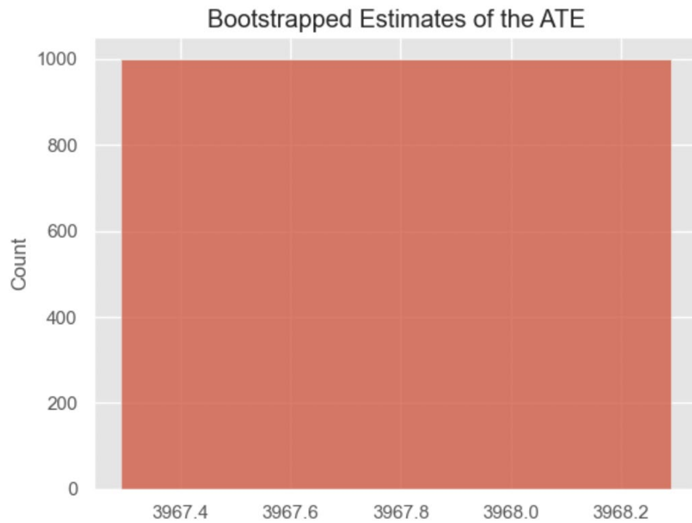

Figure 5: Bootstrapped Estimates of the ATE. Created using Seaborn library in Python.

My histogram (Figure 5) does not include 0, making my results statistically significant. My narrow confidence interval typically indicates high precision, and the absence of 0 in the histogram of bootstrapped ATE estimates indicates strong evidence for a significant and consistent treatment effect. All this reinforces the validity and reliability of the findings from my analysis – providing strong evidence for the effectiveness of the treatment in influencing the outcome variable.

Technique 2: Matching

Seen above, it is clear that a simple linear regression model is not ideal - even if I add all the variables as controls. Lalonde used these findings to argue that linear regression for causal inference is highly unreliable. So next, I considered a different technique called matching.

If we assume unconfoundedness, then for these two people, there should be no other variables that have an effect on both the treatment and the outcome. So, by subtracting their outcomes, we should be able to estimate the causal effect of the ban for this particular X (specifically, a regency/city with a population, b land area, etc.). If we do this for every possible set of values for the confounders X, then we can take all of them and compute the expectation (weighting each by the probability of seeing that corresponding value of X). Empirically, this corresponds to just taking the average of all the data points. Here is the matching algorithm:

1. For each treated row:
    - Find all untreated rows that have the exact same values of all confounders.

- Take those untreated rows and average their outcome
- Subtract the average above from the treated row's outcome

2. For each untreated row:
   - Find all treated rows that have the exact same values of all confounders.
   - Take those treated rows and average their outcome
   - Subtract the untreated row's outcome from the average above
3. Average all the results from steps 1 and 2.

However, in my case, *exact matching* will not work because I would not get many exact matches with the number of variables to consider. There are solutions such as *approximate matching* which matches people if they have similar features (not identical), but I turned to using propensity scores instead. Propensity scores are a dimensionality-reduction technique that map all 5 confounders down to a single value per observation, and are more effective than matching.

Technique 3: Inverse Propensity Weighting

The definition of the propensity score: it is the probability that a unit was treated, conditioned on a particular set of confounders x:

$$e(x) = P(Z = 1 \mid X = x)$$

Inverse Propensity Weighting (IPW) is also sometimes called a "selection model" because it uses the covariates to model how observations are selected into treatment or control. The simplest and most common way to compute propensity scores is using logistic regression. In particular, I regressed the confounders (Population, GDP, etc.) on a binary indicator for whether the observation was treated or not (0 for untreated, 1 for treated).

After generating the propensity scores, I plotted a histogram of propensity scores, grouped by dataset (Figure 6).

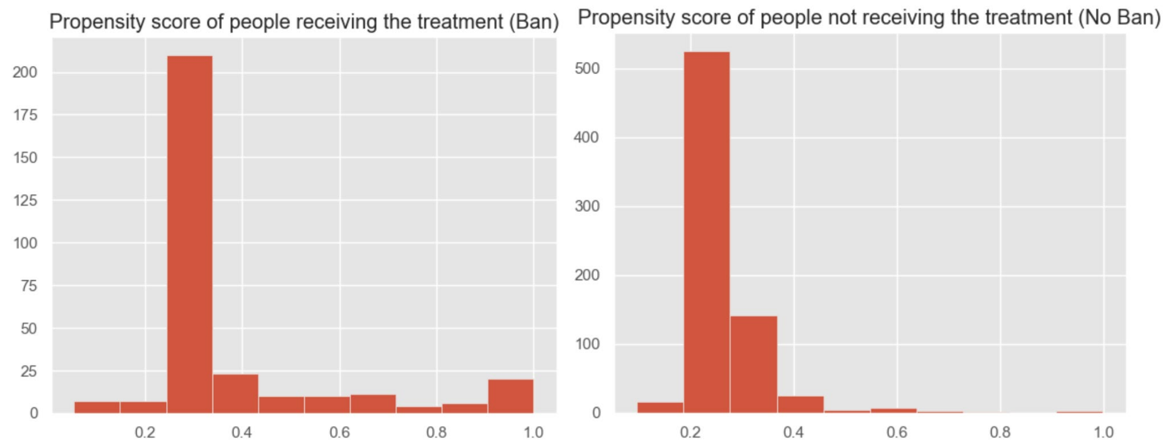

Figure 6: Propensity scores of ban and no ban group. Created using Seaborn library in Python.

The mode of the histogram for the ban group being around 0.3, and the mode for the no ban group is around 0.20, which indicates a lower likelihood of receiving the ban among individuals in the control group. The taller bars in the ban histogram (around 200) indicate a higher concentration of observations around certain propensity score values compared to the no ban histogram (around 500). The distribution of propensity scores for cities without bans appeared to be narrower, with fewer cities having high propensity scores above 0.5. This tells us that cities with bans tended to have certain characteristics that made them more likely to implement bans, while cities without bans had different characteristics.

Finally, I computed the IPW estimate for the ATE using the formula below. Note that the weights are different for the two groups. Intuitively, the weights decrease the importance of points that have a high probability of being in the group that they're in.

$$\hat{\tau}_{IPW} = \frac{1}{n} \sum_{i:Z_i=1} \frac{Y_i}{e(X_i)} - \frac{1}{n} \sum_{i:Z_i=0} \frac{Y_i}{1-e(X_i)}$$

Using this formula, I got the value ~1430.950. This means that if I assume that there are no confounding variables – features such as population, land area, Gini, GDP, HDI index – then the estimated effect of a regency/city using bans is that the ban causes there to be 1430 more tons of plastic than they would have.

Now that I have determined that ignoring my indicators will lead to a misleading insight that bans are tied to an increase in plastic, and also that characteristics differ for cities with bans and cities without bans – I can take advantage of this to forecast the effect of bans on plastic through predictive modeling.

Predictive Modeling

Process & Experimentation

To prepare the data for model building, the dataset was first split into training, validation, and testing subsets. I first divided the data into cities that have implemented bans, and cities without bans. The set of cities that have implemented bans is split up into a training set and validation set – a training set that comprises 80% of the data to fit the model, and a validation set comprising the remaining 20% of the data to provide an unbiased evaluation of a model fit. The test set comprises all the cities without bans; this is the subset of data I apply my model to after the training process and the model now knows the patterns to look for in forecasting. It is only used once a model is completely trained (using the train and validation sets).

I first experimented with different models before settling on the best one, considering both model performance (accuracy, error metrics) as well as interpretability (how easy it is to understand, how simple it is) in my decision. My 2 main error metrics were MSE (Mean Squared Error) where a lower MSE indicates that the model's predictions are closer to the actual values, and R-squared (R2) which measures the proportion of the variance in the dependent variable that is explained by the independent variables, so a higher value indicates better performance as it indicates that the model explains as much variability of the response data around its mean.

I looked to past literature on modeling environmental impacts to see what specific methods have worked favorably in the past. Linear and multiple regression models were used as a baseline, though recognized as less favorable due to their comparatively limited ability to capture complex relationships within the data. Recommended methods included advanced neural networks such as Long Short-Term Memory (LSTM) networks for their ability to capture intricate temporal patterns, random forest models for their adaptability to complex datasets, as well as both Least squares Support Vector Machines (SVM) and standard SVM models.

I first trained a Linear Regression model to use as a baseline, before moving on to models that allow for non-linear relationships between predictors and the response variable, such as LSTM Neural Networks, SVMs, and Gradient Boosting model XGBoost. Long Short-Term Memory (LSTM) consists of layers of neurons, each one takes in information, understands patterns over time, and output predictions. Least Squares Support Vector Machines (LS-SVM) with a linear kernel can be viewed as a regression technique that finds a linear hyperplane to approximate the relationship between input features and target variables, while minimizing the prediction error using least squares optimization. Some other common kernels used in SVMs are RBF and Polynomial, but in my case I used a Linear Kernel since it provides the lowest MSE. Linear kernels are best suited for datasets that exhibit clear linear separability or are relatively simple. They are computationally efficient and less prone to overfitting, making them suitable for datasets with a large number of features. Extreme Gradient Boosting (XGBoost) creates a collection of decision trees and combines their predictions to make a final decision.

Generalized linear models were also implemented, both Frequentist approach and Bayesian approach were attempted. Generalized linear models (GLMs) use X to predict Y by estimating weights B. GLMs include 2 functions 1) linear function which multiplies and/or adds the predictions and coefficients, 2) inverse link function which generates the likelihood function. The linear function is what makes this GLM model linear.

The first step involves determining the likelihood distribution for the specific problem, with common choices including Gaussian, Poisson, or gamma distributions. I plotted the kernel density estimate to see the shape, and determined the distribution it suits best (Figure 7).

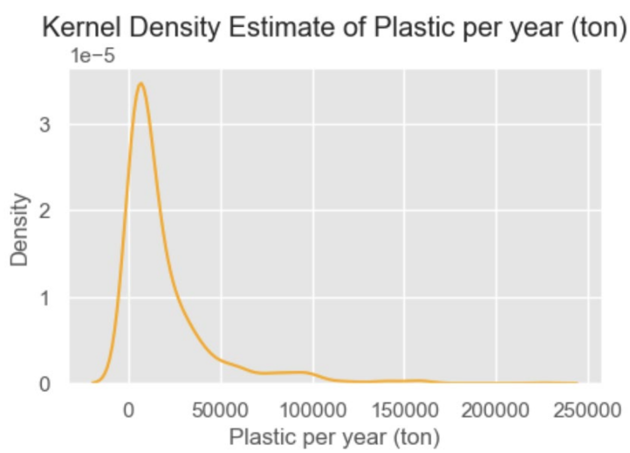

Figure 7: Kernel Density Estimate of Plastic per year. Created using Seaborn and Matplotlib library in Python.

It is definitely not a normal distribution as it is not centered in the middle; it is right-skewed. Between poisson and gamma distribution, the main difference is that poisson is for discrete variables and gamma is for continuous variables. In this case, plastic values most likely are continuous variables as it can take on any real value within a range, and there are no gaps or interruptions between the values. So, I chose the Gamma distribution as the likelihood distribution. Subsequently, the appropriate link function is identified; for continuous response variables, the identity link function is typically employed.

Model building and evaluation follow, wherein both Frequentist and Bayesian GLMs are attempted. Frequentist and Bayesian are two opposing perspectives in statistics. In the Frequentist approach, the data is random but the unknowns are fixed – we are finding a specific number. In Bayesian logic, both the data and the unknowns are random, therefore needing us to make more informed assumptions and affecting interpretability more.

Different GLMs with different distributions and link functions are compared using evaluation metrics such as likelihood-based metrics like AIC. AIC (Akaike Information Criterion) is a measure of the relative quality of a statistical model. It balances the goodness-of-fit of the model with its complexity, penalizing for the number of parameters. A lower AIC value indicates a better model.

Frequentist Regression was attempted using statsmodels.api. The best frequentist GLM I came up with was a Gamma GLM, which had a lower AIC (~7738) than the AIC for a Poisson GLM (~3226420). Bayesian regression via sampling using Bambi is attempted but ultimately causes a SamplingError due to uninformative priors. Kernel density plots are utilized to approximate distributions (Figure 8).

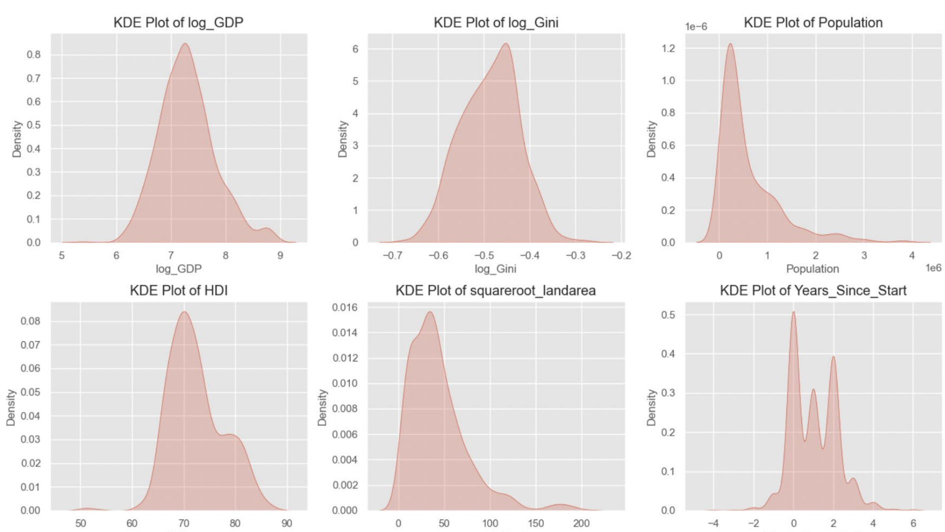

Figure 8: KDE Plots of each variable. Created using Seaborn and Matplotlib library in Python.

However, I noticed some variables – namely Years since the start date of ban – are multi-modal, meaning it has many modes. Such observation shows it is not appropriate to assume a single probability distribution for the entire variable, and indicates the need for non-parametric methods, and that my problem is better suited to a Frequentist point of view instead of Bayesian.

Nonparametric methods are statistical techniques that do not make explicit assumptions about the functional form or distributional shape of the underlying population. Instead of estimating parameters based on specific parametric models, nonparametric methods seek to directly estimate patterns or relationships from the data itself. I tried the nonparametric methods K-Nearest Neighbors, Decision trees, and random forest. With this, I have now implemented 9 different models and can compare their performance to settle on which one is the best for my problem. All the model's performance visuals such as plots of actual versus predicted values, and interpretation of model inner workings are in Appendix C.

Model Selection & Improvement

Based solely on MSE, the 2 best models are XGBoost (Mean Squared Error = ~309108544) and LSTM (Mean Squared Error = ~310311952). Since the 2 best MSE are so close, the deciding factor can come from things like complexity, interpretability, and computational resources, or other error metrics. XGBoost models are relatively less complex compared to deep learning models like LSTM. XGBoost also tends to perform well on small to medium-sized datasets while LSTM models might require more data to effectively capture temporal patterns and dependencies. Considering all of this, let's move forward with XGBoost.

I enhanced model performance through hyperparameter tuning, which involves adjusting the model's internal settings. K-fold cross-validation is a common method used for this purpose, allowing us to utilize more data for training and ensuring better generalization to unseen data. However, employing k-fold cross-validation demands increased computational resources as it requires fitting multiple models per hyperparameter choice. Instead of reserving a portion of the data for model selection, k-fold cross-validation employs the training set for this purpose by splitting it into multiple temporary train and validation sets, known as "folds". The average validation error across all k folds guides us in making optimal choices regarding features, models, and hyperparameters. The general steps are:

1. Define a parameter grid specifying the hyperparameters you want to tune.

2. Use GridSearchCV or RandomizedSearchCV from scikit-learn to search the parameter grid for the best combination of hyperparameters
3. Measuring the goodness of this combination by minimizing MSE (mean squared error).
4. Evaluate each combination using cross-validation to find the optimal hyperparameters.

Some specific hyperparameters I tuned include:

- Learning Rate (learning_rate): This hyperparameter controls the step size at each iteration while moving toward a minimum of the loss function. Lower values make the model more robust, but require more boosting rounds. Typical range: [0.01, 0.3].
- Maximum Depth of a Tree (max_depth): This parameter controls the maximum depth of each tree. Deeper trees can capture more complex relationships but are more prone to overfitting. Typical range: [3, 10].
- Minimum Sum of Instance Weight (min_child_weight): This is the minimum sum of weights of all observations required in a child. It's used to control over-fitting. Higher values prevent the model from learning too specific patterns. Typical range: [1, 10].
- Subsample (subsample): This is the fraction of observations to be randomly sampled for each tree. Lower values make the algorithm more conservative and prevent overfitting but too low values might lead to underfitting. Typical range: [0.5, 1.0].

In an effort to expedite the process while sacrificing some level of thoroughness, RandomizedSearchCV is employed instead of GridSearchCV, with the early_stopping_rounds parameter set to specify the number of rounds with no improvement after which training will be stopped. In the RandomizedSearchCV approach, the parameter grid includes the following distributions: learning_rate (uniform distribution between 0.01 and 0.3), max_depth (discrete uniform distribution between 3 and 10), min_child_weight (discrete uniform distribution between 1 and 5), and subsample (uniform distribution between 0.5 and 1.0).

On the other hand, the GridSearchCV method explores a predetermined parameter grid, which includes options for learning_rate (0.01, 0.1, 0.3), max_depth (6, 7, 10), min_child_weight (1, 3, 5), and subsample (0.5, 0.7, 0.9, 1.0). GridSearch is a very brute force method that takes a lot of time and computational power, so will just try a few educated guesses as the options to choose hyperparam from. This may not be the best one, but it is the best out of the limited options I gave it.

After tuning the hyperparameters, I applied regularization to prevent overfitting. Regularization helps penalize large coefficients and simplify the model, reducing overfitting. I tuned parameters like gamma, alpha, and lambda to control the complexity of the model.

- Alpha (alpha): L1 regularization term on weights. Typical range: [0, 1].
- Lambda (lambda): L2 regularization term on weights. Typical range: [0, 1].

For L1 regularization (alpha) and L2 regularization (lambda), a parameter grid is defined to explore different values covering a range of potential regularization strengths. The parameter grid includes values such as alpha = [0, 0.1, 0.2, 0.3] for L1 regularization and lambda = [0, 0.1, 0.2, 0.3] for L2 regularization.

All data and code notebooks are available on github.com/sandyawijayaa/ES-thesis-24. Researchers interested in replicating or further examining the analysis conducted in this thesis are encouraged to refer to the code notebooks provided.

RESULTS

Bans and Plastic over time

Plastic only decreases around 2 years after the start of a ban, seen by how that's when a decreasing slope begins (Figure 9). Upon investigating this on a granular level where there is 1 line for each regency/city with a ban (Figure 10), I observe that while overall there is generally a decreasing trend, some lines in the individual graph are increasing. In other words, some regency/cities have ineffective bans, and this may be a source of discrepancy as this 'outlier' may be dragging the total up.

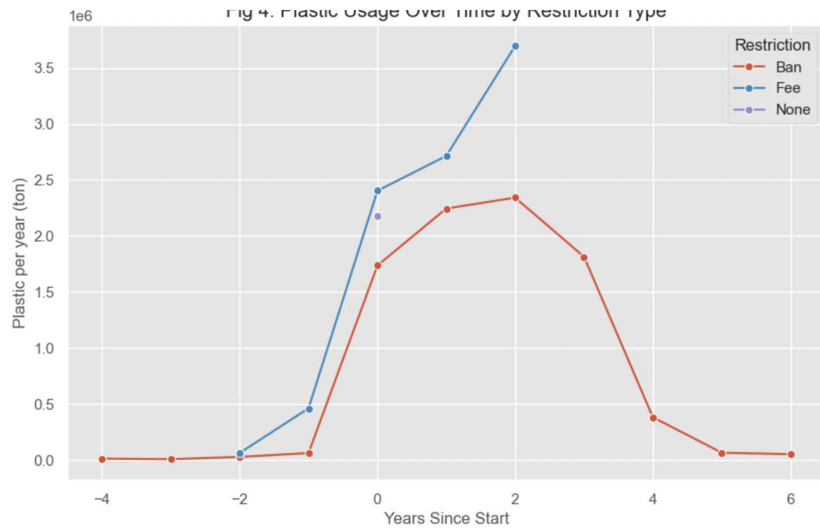

Figure 9: Plastic Usage Over Time by Restriction type. Created using Seaborn and Matplotlib library in Python.

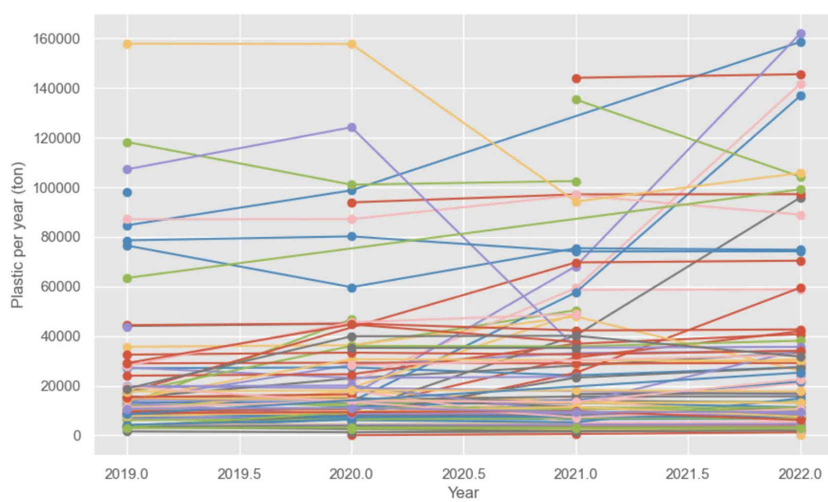

Figure 10: Plastic Usage Over Time for every regency/city that has enacted a ban. Created using Seaborn and Matplotlib library in Python.

Linear relationships between indicators and plastic

Assuming that all other features remain constant:

- As population increases by 1, linear regression predicts that the plastic per year increases by 0.034 tons.
- If the GDP doubles (since it's a log scale, so take exponential which is roughly 2.72), the model predicts a decrease of about 1050 tons of plastic per year.

- For every point increase in the Gini index, the model predicts an increase of about 6483 tons of plastic per year.
- If the land area increases by 1 square root unit (whatever that may be), the model predicts an increase of about 83 tons of plastic per year.
- For every point increase in the HDI, the model predicts an increase of about 729 tons of plastic per year.
- For each additional year since the ban started, the model predicts an increase of about 2550 tons of plastic per year.

Comparison of Models

I plotted predicted vs. actual values to see visually how each model's predictions looked against the true values (Figure 11, Figure 12).

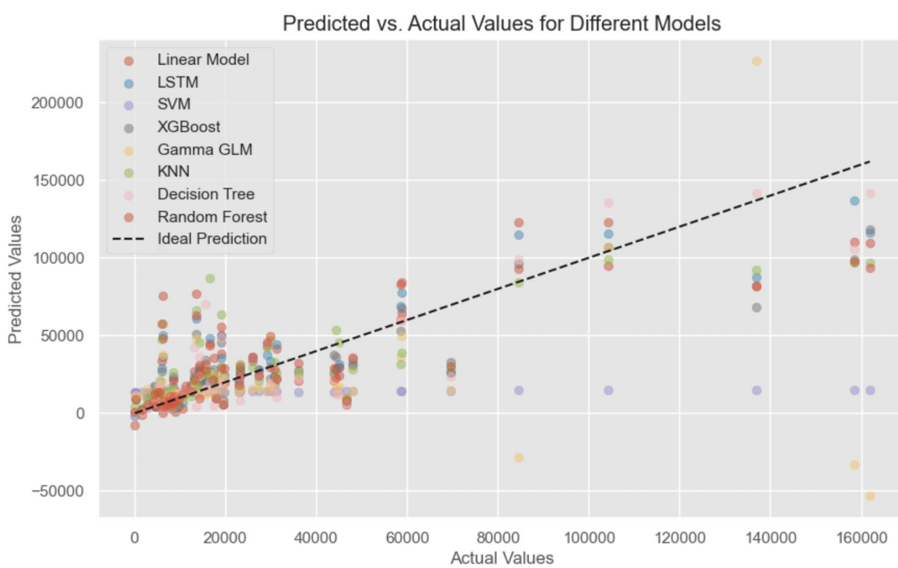

Figure 11: Predicted against Actual Values for the 9 different models. Created using Seaborn and Matplotlib library in Python

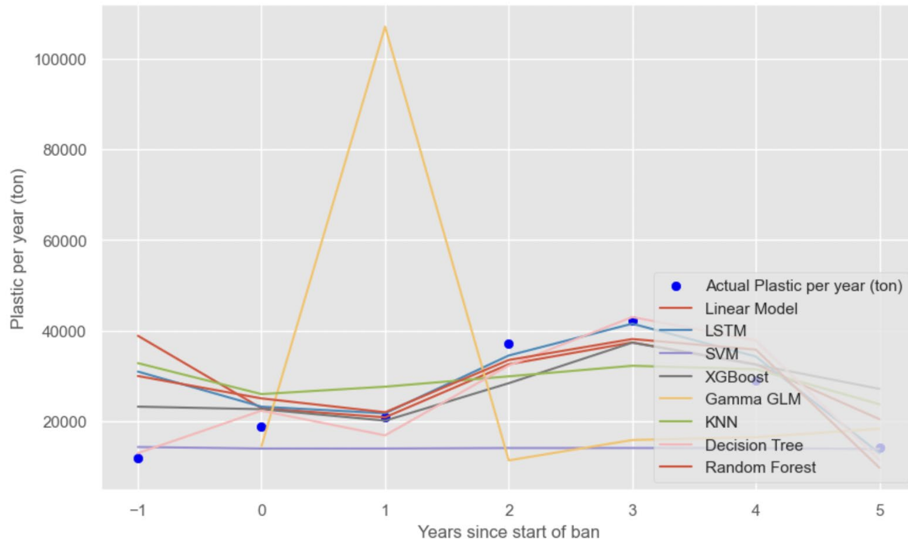

Figure 12: Actual and Predicted Values of Plastic over time. Created using Seaborn and Matplotlib library in Python.

To understand the model performance quantitatively, I calculated the MSE on the validation set for each model below (Table 1).

Table 1: Error metrics of each model. Created using Google Docs.

Model	Mean Squared Error (MSE)
Linear Regression	~ 442438137.691
Long Short Term Memory (LSTM)	~ 310311952.595
Least-Squares Support Vector Machine (SVM)	~ 1322360629.867
XGBoost	~ 309108544.768
Gamma Generalized Linear Model (GLM)	~ 1801448477.971
K-Nearest Neighbors (KNN)	~ 502097431.422
Decision Tree	~ 404472588.035
Random Forest	~ 387778126.416

Because XGBoost had the lowest MSE at a fairly low model complexity, I chose XGBoost as the final model I will use and improve.

Tuning XGBoost

In general, a low MSE and a high R2 are desirable. In general R2 of 0.7 might be considered good, and MSE depends on the scale of my problem and numbers. Despite the attempt to regularize the model, both L1 and L2 regularization techniques result in decreased performance (Table 2). While L1 regularization shows slightly better performance, there is a very significant decrease in performance after both forms of regularization, indicating that both L2 regularization and L1 regularization might be too strong for my dataset. With this, I decided not to move forward with regularization. With how the hyperparameters are tuned, I see that surprisingly, MSE is higher and R2 is lower – the worst of both cases.

Table 2: Actual and Predicted Values of Plastic over time. Created using Google Docs.

Process	learning_rate	max_depth	min_child_weight	subsample	alpha	lambda	MSE	R2 Score
Default	0.3	6	1	1.0	0	0	~309108544	~0.735
Randomized Search	~0.254	3	2	~0.924	0	0	~389183950	~0.667
Grid Search	0.1	7	5	0.9	0	0	~377677816	~0.676
L1 (Alpha) Regularization	0.1	7	5	0.9	0.2	0	~571869135	~0.510
L2 (Lambda) Regularization	0.1	7	5	0.9	0	0.2	~577278709	~0.505

In some cases, a slightly higher MSE or lower R-squared might be acceptable if the model is more interpretable or easier to deploy in practice. Lower MSE can sometimes also be a sign of a model working to prevent overfitting. However, in this case the difference is quite significant before and after hyperparameter tuning. The initial model without hyperparameter tuning outperforms the model without regularization in terms of MSE and R-squared, so I decided to continue with my initial model not only because of better error metrics, but also because it uses default values for its hyperparameters, thus leading to a less complex model. Less complex models are easier to interpret, and require less computational resources for training and inference, so they might be easier to deploy in production environments.

Final XGBoost model

The best model is an XGBoost model with default parameters and no regularization. I visualized its predicted values against actual values of the training set to see its performance (Figure 13).

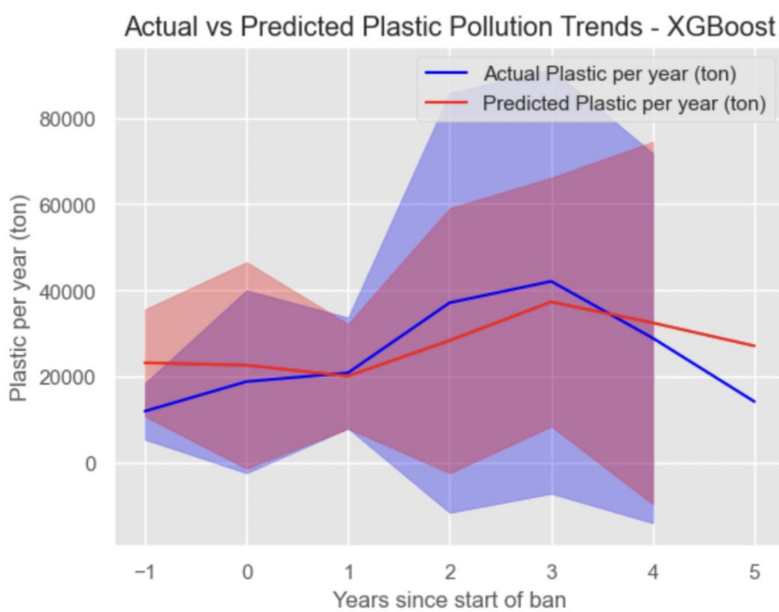

Figure 13: Actual and Predicted Values by final XGBoost. Created using Seaborn and Matplotlib library in Python.

I visualized both the feature importance scores and SHAP values of this final XGBoost model to provide insights into the relative importance and impact of features on the model's predictions. XGBoost provides a feature importance score for each feature, indicating its relative importance in the model's predictions. I visualized this using a bar plot to identify the most influential features (Figure 14).

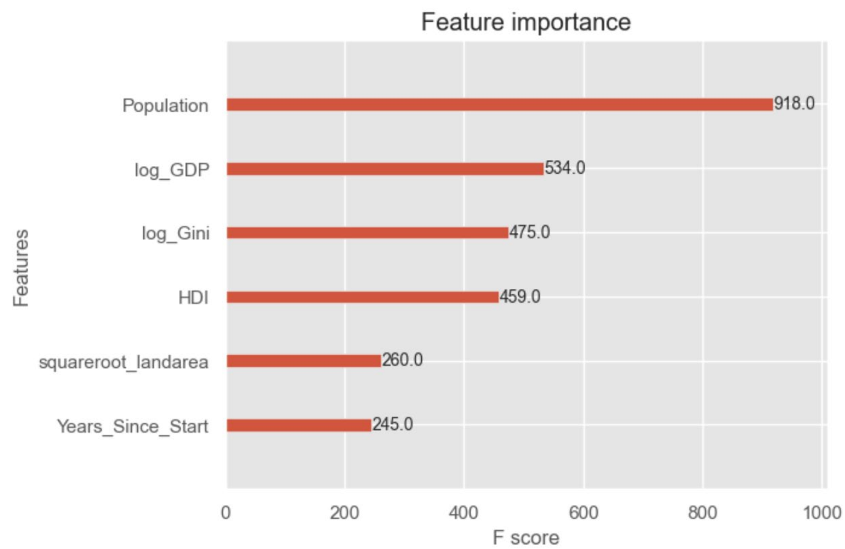

Figure 14: Feature Importance of features in final XGBoost model. Created using Seaborn and Matplotlib library in Python.

SHAP (SHapley Additive exPlanations) values provide a way to explain the output of any machine learning model. They quantify the impact of each feature on the model's predictions for individual instances. The ranges of SHAP values for each feature provide additional context on the variability and impact of each feature on the model's predictions (Figure 15).

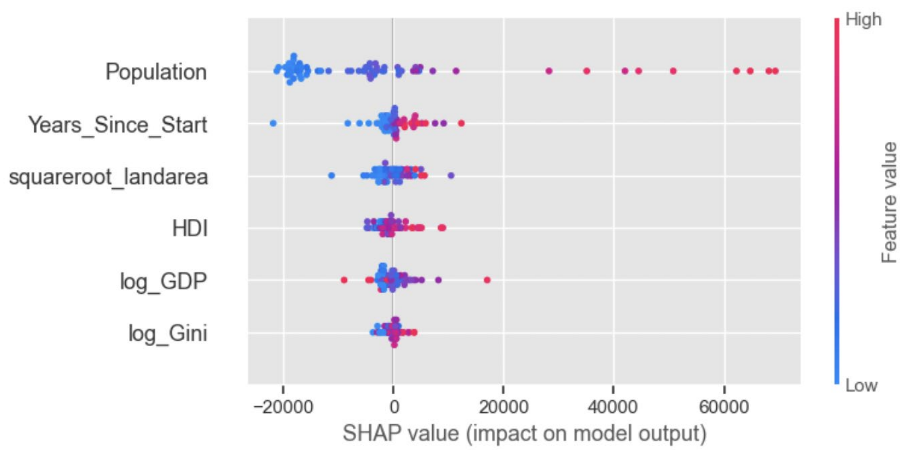

Figure 15: SHAP values of features in final XGBoost model. Created using Seaborn and Matplotlib library in Python.

Feature importance scores are based on the overall contribution of each feature to the model's predictive performance, while SHAP values provide a more granular understanding of how each feature influences individual predictions. Comparing the two, I observed similarities or differences in the ranking and impact of features. One notable insight is that Population and Years Since Start appear to be important features according to both feature importance scores and SHAP values. Additionally, based on these provided feature importance scores and SHAP values, none of the features consistently rank lower in importance across both metrics. All of them are deemed critical based on domain knowledge, so none of the features will be deleted from the model.

Predicted plastic for cities without bans given they implemented bans in 2019

Now that I have my final model, I applied this trained XGBoost model to my test set; the cities without bans. I specified that a ban is implemented in 2019 and plot how the predicted plastic pollution changes over time, averaging overall as a first step (Figure 16).

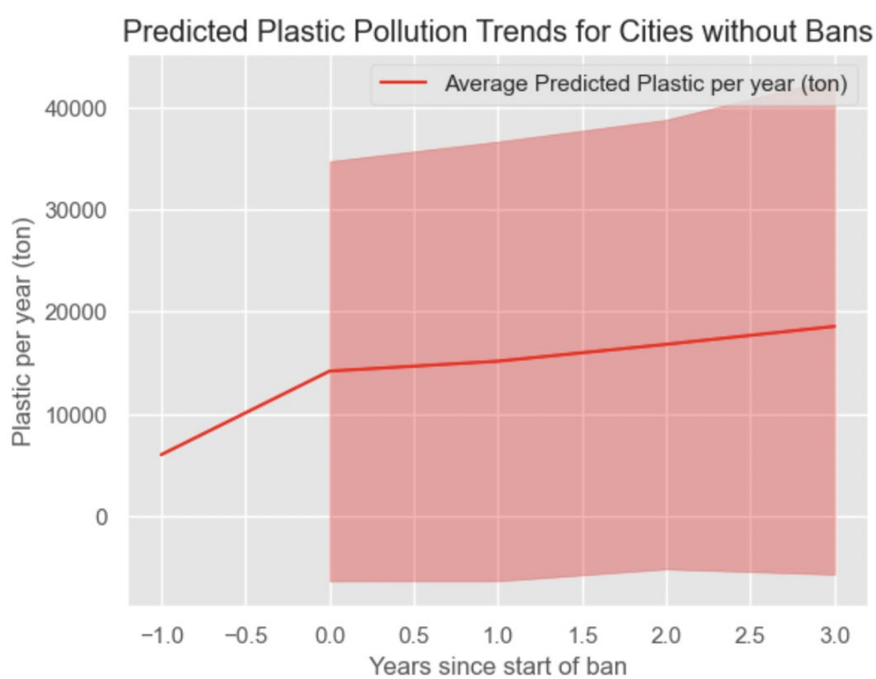

Figure 16: Predicted plastic pollution trends for cities without bans - average in general. Created using Seaborn and Matplotlib library in Python.

This is unexpected that there is no projected decrease in plastic after ban is implemented, especially after ban has been implemented for 2 years (x-axis = 2.0) because as I learnt at the start of this process, there is usually a decrease in plastic after this cutoff of 2 years. Such results may be due to the possibility that more cities have increases in plastic than cities that have decreases in plastic.

To investigate further, I plotted it on a granular scale of provinces instead of overall (Figure 17). Upon visualizing the predicted plastic pollution trends by province; averaging the predicted values for all the regency/city in that province, I saw that there is a decrease in plastic after the ban was implemented for 2 years, but only for some provinces.

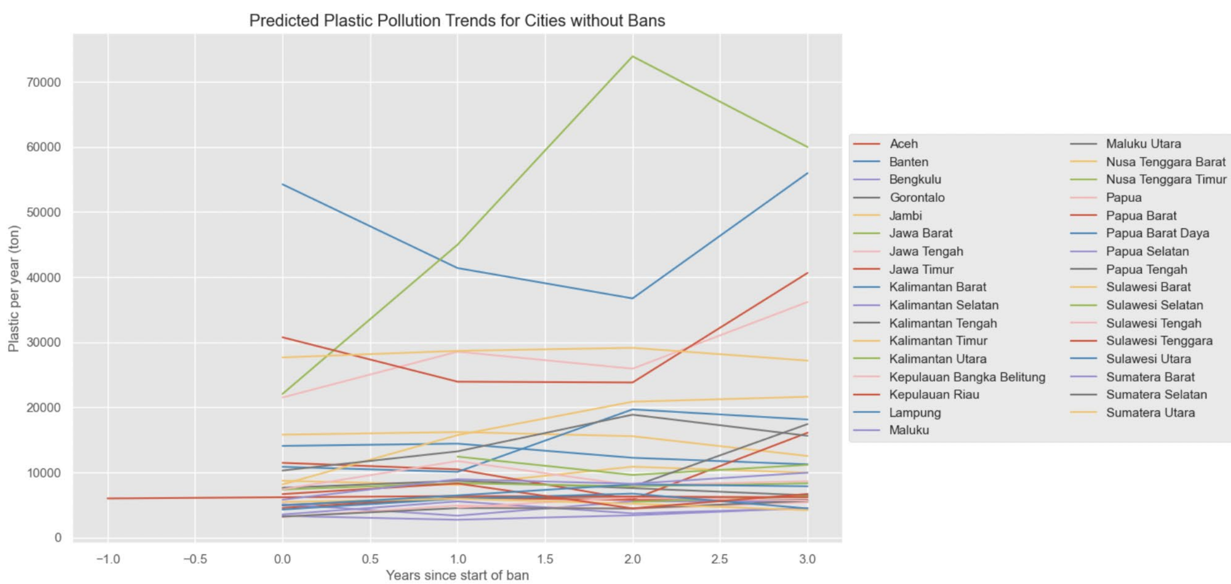

Figure 17: Predicted plastic pollution trends for cities without bans - by province. Created using Seaborn and Matplotlib library in Python.

Upon investigating on a more granular scale of provinces, I found that there is a decrease in plastic after the ban was implemented for 2 years, but only for 13 provinces (out of the 33 provinces that do not have bans). These provinces with a decrease in plastic pollution 2 years after a ban begins being implemented are as follows:

- Gorontalo: Decrease by ~571.929 tons per year
- Jambi: Decrease by ~927.537 tons per year
- Jawa Barat: Decrease by ~13934.851 tons per year

- Kalimantan Barat: Decrease by ~1014.660 tons per year
- Kalimantan Tengah: Decrease by ~1098.928 tons per year
- Kalimantan Timur: Decrease by ~3041.210 tons per year
- Lampung: Decrease by ~1555.326 tons per year
- Papua Barat: Decrease by ~94.685 tons per year
- Papua Barat Daya: Decrease by ~2276.838 tons per year
- Sulawesi Barat: Decrease by ~994.112 tons per year
- Sulawesi Utara: Decrease by ~289.392 tons per year
- Sumatera Selatan: Decrease by ~3244.038 tons per year
- Sumatera Utara: Decrease by ~1961.791 tons per year

DISCUSSION

Linear relationships between indicators and plastic

In examining the linear relationships between various socio-economic factors and plastic pollution levels, our analysis uncovers intriguing insights that merit further exploration and discussion. One interesting result is the inverse relationship between GDP and plastic pollution. Such observation aligns with the theoretical framework of the Kuznets Curve, which suggests that as countries experience economic growth, environmental degradation initially worsens before eventually improving as societies become wealthier and allocate resources towards environmental protection measures (Science Direct).

The positive association between Gini index and plastic pollution levels underscores the intricate interplay between socio-economic factors and environmental governance. The middle class is the engine of plastic bans. Affluent segments of society may contribute more significantly to overall plastic consumption through higher consumption patterns. Meanwhile, marginalized or lower-income communities may lack resources to purchase sustainable alternatives and lack access to waste management services, leading to increased environmental burden in these areas. The more inequality, the higher proportion of the middle class, the more push for recycling and bans. As middle-class efforts towards sustainability intersect with the labor of marginalized communities, inequalities in consumption patterns and waste management practices become apparent (Anantharaman 2024).

Varying places, varying needs

Beyond politics, many other factors influence the effectiveness of plastic pollution solutions. A major blocker is how different areas have different physical, demographic, socioeconomic factors, which lead to varying stress points and needs (Figure 18).

Framework of City Categorization on Waste Management

	CITIES	PRIORITY NEEDS	EFFECTIVE MEASURES	DONOR PROJECT EXAMPLES
Category I: Big cities with severe land limitation	Jakarta, Surabaya, Bandung & Bekasi	Waste volume reduction	Treatment facility that can significantly reduce volume, such as thermal treatment	Waste treatment facility development in Bandung region (JICA and IFC)
Category II: Island and beach side, lakeside cities	Bali, Lombok, Pulau Seribu & Municipalities around Lake Toba	Cleaning up, reducing plastic waste	Plastic waste management, clean-up activities	Plastic waste improvement in Lake Toba (Government of Japan and UNEP)
Category III: Middle/small cities / countryside	Balikpapan, Palembang, Sorong, Kupang, Palu, Ruteng & Buol	Improving sanitation & waste collection rates	Efficient waste collection, landfill management, trial projects for new technologies or systems	3R project in Balikpapan and Palembang focusing on source segregation and collection improvement (JICA)

Source: made by the author.

Figure 18: Framework of City Characterization on Waste Management. Created by authors of "Turning the Tide: How Can Indonesia CLOSE the LOOP on Plastic Waste?" by Wilson Center, 2021.

For large cities where population density is high, more waste is generated while the price of land for landfill cannot keep up. Since the economic and social value of land in urban areas can be quite high, reduction of waste volume to save space for landfill is the priority. For island and beachside & lakeside cities, removing waste, especially plastic waste, is the priority. For middle and small cities, they suffer from the opposite problem of large urban areas – these cities have available land but no budget, so the priority of waste management policy lies in better waste management. Specifically, initiatives to properly handle and sanitize waste are needed in order to increase the collection rate (Wilson Center 2021).

I assigned each "successful ban province" – provinces that saw decreases in plastic after 2+ years of ban implementation – to each different category.

- Category 1 - Dense: Jawa Barat (West Java). Jawa Barat has a population of over 49 million people, making it one of the most densely populated provinces in Indonesia (BPS).
- Category 2 - Coastal: Lampung, Papua Barat (West Papua), Papua Barat Daya (Southwest Papua), Sulawesi Barat (West Sulawesi), Sulawesi Utara (North Sulawesi), Sumatera Selatan (South Sumatra), Sumatera Utara (North Sumatra). All of these provinces are situated along coastlines.
- Category 3 - Small/Rural: Gorontalo, Jambi, Kalimantan Barat (West Kalimantan), Kalimantan Tengah (Central Kalimantan), Kalimantan Timur (East Kalimantan). All of these provinces generally have lower GDP per capita and economic development compared to urban centers (BPS).

I then represented these different categories with different colors on a map (Figure 19). Since most of the "successful ban provinces" fall under the second category of coastal areas, I inferred that Coastal areas are more likely to see decreases in plastic if bans are implemented (after 2 years). Such cities have the motivation to reduce plastic as they are usually tourist attractions that need to keep attracting visitors with a clean environment to maintain the economy, but they are not as high density as big cities so it is easier to keep rules in place.

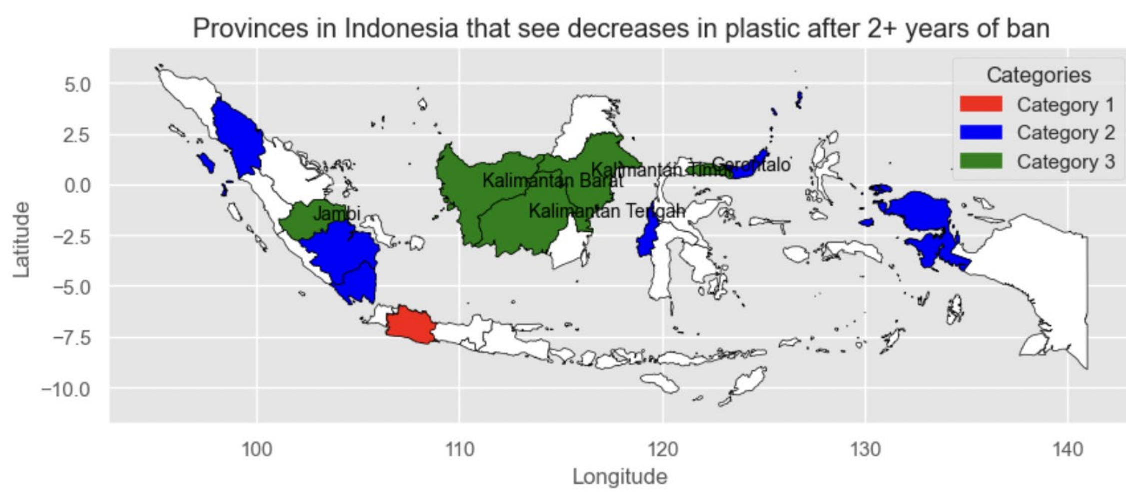

Figure 19: Provinces in Indonesia that see decreases in plastic after 2+ years of ban, categorized into groups by needs. Created using Seaborn and Matplotlib library in Python.

Synthesis

By employing various unconfoundedness techniques such as outcome regression and inverse propensity weighting, I found that there is a substantial efficacy of ban interventions in mitigating plastic pollution, shown by the large estimated average treatment effect (ATE) of bans. The difference in the distribution of data for cities with bans and without tells us that cities with bans tended to have certain underlying characteristics that made them more likely to implement bans. Additionally, I discovered the misleading insight that bans caused an increase in plastic, if confounders – such as population, land area, Gini, GDP, HDI index – are not considered at all in my analysis. Given these insights, I now have a solid foundation that tells us there is indeed a causal relation between bans and plastic pollution, and the varying characteristics in areas with and without bans tells us that it is indeed possible to forecast the effect of bans on plastic through predictive modeling.

The most successful model to forecast the effect of bans on plastic pollution was XGBoost with default parameters. My modeling results highlighted the significant roles played by population and the number of years since the ban's implementation in predicting plastic pollution levels. Furthermore, I observed that not all bans have the same impact, with trends varying by geographic area and time since implementation. Coastal cities, in particular, are more likely to experience decreases in plastic pollution following the implementation of bans of 2 years or more. This 2 year threshold represents the delay period of bans; plastic pollution numbers start to change 2 years after bans are implemented.

Limitations & Future Directions

Despite the comprehensive nature of my analysis, several limitations hindered the depth and scope of my investigation.

Firstly, this study aimed to delve beyond the binary classification of bans versus no bans and also collect data that could convey the nuanced intricacies of ban implementation across different regions. I aspired to gather detailed information such as the extent of ban enforcement, the provision of alternatives such as paper bags, return policies, responsible governing entities, funding sources, sustainability measures post-implementation, and potential loopholes in the enforcement process. However, the availability and accessibility of such granular data were limited by the disparate data collection infrastructure across various regions in Indonesia. While some cities may have maintained comprehensive records, the majority lacked systematic tracking mechanisms for these specific details.

Additionally, discrepancies between official policy documentation and real-world implementation posed a challenge, as certain areas may purportedly enforce a full ban on plastic usage while encountering localized instances of non-compliance, such as the unauthorized distribution of plastic bags for a fee.

In terms of causal analysis, a deeper exploration is warranted to delve into the assumptions underlying our model. Scrutinizing alternative model specifications and conducting sensitivity analyses will contribute to more robust conclusions. It is also vital to widen the scope of confounders accounted for. While we identified and controlled for certain confounders associated with plastic bans using 5 indicators, this set might not encompass the full spectrum of relevant variables. Future research that can collect and analyze such additional data can enhance the precision and reliability of our findings.

Furthermore, resource constraints, including time and computational power, constrained the extent of my model optimization efforts. While my analysis employed hyperparameter tuning via GridSearchCV, limitations in computational resources restricted the granularity of parameter grid exploration. There is potential to enhance model performance by expanding the parameter grid and exhaustively exploring the optimal combination of hyperparameters. Additionally, model performance could also be further improved if other hyper parameters were tuned, such as Number of Estimators (n_estimators) which determines the number of boosting rounds or trees to build. Higher values can lead to overfitting, so it is good to tune this with other parameters.

Moreover, while my study pioneers the application of machine learning algorithms to this niche problem domain, it is important to remember that not all problems are meant for machine learning – at times, the infrastructure and past documentation do not support the assumptions for such strategies. Spatial methods such as suitability analysis have traditionally been employed in these contexts. A comparative analysis between machine learning-based predictions and those derived from traditional methods could offer valuable insights into the relative accuracy and efficacy of different predictive approaches, informing future research directions and policy implementation strategies.

These limitations underscore the need for future research endeavors to address these constraints and further elucidate the complexities of policy interventions in mitigating environmental challenges.

Broader implications

The predictive insights generated by this study hold immense value for governmental decision-makers, offering a glimpse into the potential outcomes of implementing novel environmental policies in new areas – is it feasible, and if so, will it yield positive outcomes. By answering these critical questions surrounding policy implementation, governments can maximize the impact of their environmental initiatives. These predictions also serve as invaluable tools for guiding resource allocation decisions, and for countries like Indonesia that face limited budgets to address pressing environmental concerns, this ability to streamline their efforts becomes paramount. The overarching insight of this study that not all bans are equal underscores the importance of tailoring policy interventions to local contexts and underscores the need for adaptive governance frameworks capable of accommodating regional nuances.

While this study focuses specifically on the implementation of bans on single-use plastics, the methodology and predictive framework developed herein hold broader applicability across various policy domains. Whether addressing issues related to recycling initiatives or other environmental interventions, the adaptable nature of this modeling approach enables utilization in diverse policy contexts, provided that sufficient data and defined policy criteria are available.

Lastly, this study not only contributes to advancing environmental governance practices in Indonesia but also serves as a blueprint for addressing plastic pollution challenges in countries worldwide. The methodology and findings of this research can be extrapolated to similar settings globally. One example of this are other countries in Southeast Asia, which have similarly emerging economies and also analogous socio-economic structures as well as environmental challenges to Indonesia. Furthermore, nations with coastal areas and popular tourist destinations may find particular relevance in understanding that this type of geography is one that can see success in plastic bans. By leveraging the lessons learned from this study, stakeholders across the globe can work independently as well as collaboratively to implement effective policy interventions and safeguard the planet's environmental health for future generations.

ACKNOWLEDGEMENTS

Thank you to my mentor Kate O'Neill, for giving me ideas and resources for not only completing my thesis but improving it. Thank you to Pak Bustar for allowing me to interview you and gain your insight. Thank you to my graduate student instructor Jessica Craigg and reader Annie Miller for giving such useful feedback on my writing. Thank you to my peers and the rest of the staff of ESPM100ES and ESPM175A/B for their guidance and sympathy these past few semesters. Thank you to my parents, Cassandra and Eddy, and my sister, Lindya, for their limitless support throughout my college career.

REFERENCES

Anantharaman, Manisha. *Recycling Class: The Contradictions of Inclusion in Urban Sustainability.* The MIT Press, 2021.

Badan Pusat Statistik Indonesia. http://bps.go.id.

Dauvergne, Peter. 2018. "Why Is the Global Governance of Plastic Failing the Oceans?" *Global Environmental Change* 51 (July): 22–31. https://doi.org/10.1016/j.gloenvcha.2018.05.002.

Dauvergne, P., & Islam, S. (2023). The politics of anti-plastics activism in Indonesia and Malaysia. Cambridge Prisms: Plastics, 1: e1.

Diana, Zoie, Tibor Vegh, Rachel Karasik, Janet Bering, Juan D. Llano Caldas, Amy Pickle, Daniel Rittschof, Winnie Lau, and John Virdin. 2022. "The Evolving Global Plastics Policy Landscape: An Inventory and Effectiveness Review." *Environmental Science & Policy* 134 (August): 34–45. https://doi.org/10.1016/j.envsci.2022.03.028.

Dietplastik. 2024. Dokumen Peraturan pelarangan plastik Sekali Pakai di Indonesia. https://plasticdiet.id/dokumen-peraturan-pelarangan-plastik-sekali-pakai-di-indonesia/.

Environmental Kuznets curve. ScienceDirect. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/environmental-kuznets-curve.

Hermawan, M, P Heriyati, and N Andrew. 2021. "Exploring Program on Ocean Plastic Pollution Management: Case of NGO in Jakarta." *IOP Conference Series: Earth and Environmental Science* 729 (1): 012122. https://doi.org/10.1088/1755-

1315/729/1/012122.

Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., & Law, K. L. (2015). Plastic waste inputs from land into the ocean. Science, 347, 768–771.

Lasut, M. T., Weber, M., Pangalila, F., Rumampuk, N. D. C., Rimper, J. R. T. S. L., Warouw, V., Kaunang, S. T., & Lott, C. (2018). From Coral Triangle to Trash Triangle-How the Hot Spot of Global Marine Biodiversity Is Threatened by Plastic Waste. Proceedings of the International Conference on Microplastic Pollution in the Mediterranean Sea, 107–113.

Latinopoulos, Dionysis, Charalampos Mentis, and Kostas Bithas. 2018. "The Impact of a Public Information Campaign on Preferences for Marine Environmental Protection. The Case of Plastic Waste." *Marine Pollution Bulletin* 131 (June): 151–62. https://doi.org/10.1016/j.marpolbul.2018.04.002.

Mah, Alice. 2021. "Future-Proofing Capitalism: The Paradox of the Circular Economy for Plastics." *Global Environmental Politics* 21 (2): 121–42. https://doi.org/10.1162/glep_a_00594.

Muposhi A, Mpinganjira M, Wait M. "Considerations, benefits and unintended consequences of banning plastic shopping bags for environmental sustainability: A systematic literature review." *Waste Management & Research*. 2022; 40(3): 248-261. https://doi.org/10.1177/0734242X211003965

Napper, I. E., & Thompson, R. C. (2018). Micro-and macro plastics in aquatic ecosystems. Encyclopedia of Ecology, 1: 116.

Nøklebye, E., Adam, H. N., Roy-Basu, A., Bharat, G. K., & Steindal, E. H. (2023). Plastic bans in India – Addressing the socio-economic and environmental complexities. Environmental Science & Policy, 139, 219–227.

OECD. "Marine Plastics Pollution INDONESIA." www.oecd.org/ocean/topics/ocean-pollution/marine-plastics-pollution-Indonesia.pdf.

"On the Frontline of Indonesia's Plastic Waste Crisis | Wilson Center." 2021. Www.wilsoncenter.org. March 30, 2021. https://www.wilsoncenter.org/event/frontline-indonesias-plastic-waste-crisis.

Omeyer, Lucy C.M., Emily M. Duncan, Kornrawee Aiemsomboon, Nicola Beaumont, Sujaree Bureekul, Bin Cao, Luis R. Carrasco, et al. 2022. "Priorities to Inform Research on

Marine Plastic Pollution in Southeast Asia.” *Science of The Total Environment* 841 (October): 156704. https://doi.org/10.1016/j.scitotenv.2022.156704.

Schnurr, Riley E.J., Vanessa Alboiu, Meenakshi Chaudhary, Roan A. Corbett, Meaghan E. Quanz, Karthikeshwar Sankar, Harveer S. Srain, Venukasan Thavarajah, Dirk Xanthos, and Tony R. Walker. 2018. “Reducing Marine Pollution from Single-Use Plastics (SUPs): A Review.” *Marine Pollution Bulletin* 137 (December): 157–71. https://doi.org/10.1016/j.marpolbul.2018.10.001.

Sistem Informasi pengelolaan sampah Nasional. https://sipsn.menlhk.go.id/sipsn/.

Spalding, M., Spalding, M. D., Ravilious, C., & Green, E. P. (2001). World Atlas of Coral Reefs. Univ of California Press.

Tomascik, T., Mah, A. J., Nontji, A., & Moosa, K. M. (1997). The Ecology of the Indonesian Seas. Oxford University Press.

“Turning the Tide: How Can Indonesia CLOSE the LOOP on Plastic Waste?” 2021. https://www.wilsoncenter.org/sites/default/files/media/uploads/documents/Insight%20Out_April2021_FINALr4.pdf.

United Nations HDR Technical Notes. Technical note 1: Calculating the Human Development Index. https://data.un.org/_Docs/HDR%20Technical%20Notes.pdf.

Van Emmerik, Tim, Michelle Loozen, Kees Van Oeveren, Frans Buschman, and Geert Prinsen. "Riverine plastic emission from Jakarta into the ocean." *Environmental Research Letters* 14, no. 8 (2019): 084033.

Wang, Yifan, Rachel Karasik, and Key Takeaways. 2022. “Policy Brief Plastic Pollution Policy Country Profile: Indonesia.” https://nicholasinstitute.duke.edu/sites/default/files/projects/Plastic-Pollution-Policy-Country-Profile-Indonesia.pdf.

Watt, E., Picard, M., Maldonado, B., Abdelwahab, M. A., Mielewski, D. F., Drzal, L. T., Misra, M., & Mohanty, A. K. (2021). Ocean plastics: environmental implications and potential routes for mitigation – a perspective. RSC Advances, 11, 21447–21462.

APPENDIX A

```
Duplicate Matches:
('SABANG', 'SUBANG')
('LANGSA', 'LANGKAT')
('LANGKAT', 'LANGKAT')
('DKI JAKARTA', 'KOTA SURAKARTA')
('SUBANG', 'SUBANG')
('KOTA SURAKARTA', 'KOTA SURAKARTA')
('D I YOGYAKARTA', 'KOTA YOGYAKARTA')
('KOTA YOGYAKARTA', 'KOTA YOGYAKARTA')
```

Remember first value is the melted_pop value, aka the correct/reference one; in other words, it should be present. Out of these duplicate matches,

- ('SABANG', 'SUBANG'): WRONG! should be SABANG, Kota Sabang
- ('LANGSA', 'LANGKAT'): WRONG! should be LANGSA, Kota Langsa
- ('LANGKAT', 'LANGKAT') -- correct
- ('DKI JAKARTA', 'KOTA SURAKARTA'): WRONG! should be DKI JAKARTA, Daerah Khusus Ibukota Jakarta
- ('SUBANG', 'SUBANG') -- correct
- ('KOTA SURAKARTA', 'KOTA SURAKARTA') -- correct
- ('D I YOGYAKARTA', 'KOTA YOGYAKARTA'): WRONG! should be DI YOGYAKARTA, Daerah Istimewa Yogyakarta
- ('KOTA YOGYAKARTA', 'KOTA YOGYAKARTA') -- correct

Figure A1: Duplicate regency/cities from the first iteration of using difflib library to match population dataset and area dataset.

```python
# Get the unique town names from melted_pop
unique_towns_melted_pop = melted_pop['Province/Regency/City'].unique()

# Find unmatched towns
unmatched_towns = [town for town in unique_towns_melted_pop if town not in matches_melted_to_area]

# Print unmatched towns
print("Unmatched towns:")
for town in unmatched_towns:
    print(town)
```

```
Unmatched towns:
TOBA SAMOSIR / TOBA
KEP SERIBU
MAMUJU UTARA / PASANGKAYU
INDONESIA
```

```python
# Define the mappings
new_mappings = {
    "TOBA SAMOSIR / TOBA": "TOBA",
    "KEP SERIBU": "ADMINISTRASI KEPULAUAN SERIBU",
    "MAMUJU UTARA / PASANGKAYU": "PASANGKAYU"
}

# Add new mappings
matches_melted_to_area.update(new_mappings)

# Delete specified entries
entries_to_delete = ["INDONESIA"]
for entry in entries_to_delete:
    if entry in matches_melted_to_area:
        del matches_melted_to_area[entry]

print("New mappings added and specified entries deleted successfully.")
```

```
New mappings added and specified entries deleted successfully.
```

Figure A2: Unmatched regency/cities from first iteration of using difflib library to match population dataset and area dataset.

APPENDIX B

```
Shapiro-Wilk test for Population: p-value = 1.1579689429558345e-35
   Conclusion: The data is not normally distributed.

Shapiro-Wilk test for Land_Area_km2: p-value = 1.0773762988049873e-45
   Conclusion: The data is not normally distributed.

Shapiro-Wilk test for GDP: p-value = 3.1347821150092257e-49
   Conclusion: The data is not normally distributed.

Shapiro-Wilk test for Gini: p-value = 1.2641536713438465e-07
   Conclusion: The data is not normally distributed.

Shapiro-Wilk test for HDI: p-value = 2.068597574863771e-15
   Conclusion: The data is not normally distributed.
```

Figure B1: Results of Shapiro-Wilk Test to see if data for each indicator is normally distributed

```
Mann-Whitney U test for Population:
   U statistic: 143718.0
   p-value: 2.6601625912953413e-13
   Conclusion: There is a statistically significant difference between the groups.

Mann-Whitney U test for Land_Area_km2:
   U statistic: 75760.5
   p-value: 2.796110865498724e-16
   Conclusion: There is a statistically significant difference between the groups.

Mann-Whitney U test for GDP:
   U statistic: 159419.0
   p-value: 1.286503180694726e-27
   Conclusion: There is a statistically significant difference between the groups.

Mann-Whitney U test for Gini:
   U statistic: 135075.5
   p-value: 9.255049737149148e-08
   Conclusion: There is a statistically significant difference between the groups.

Mann-Whitney U test for HDI:
   U statistic: 159933.0
   p-value: 3.5285835315978754e-28
   Conclusion: There is a statistically significant difference between the groups.
```

Figure B2: Results of Mann-Whitney U Test to check if there is a statistical significant difference between cities with bans vs cities without bans

APPENDIX C

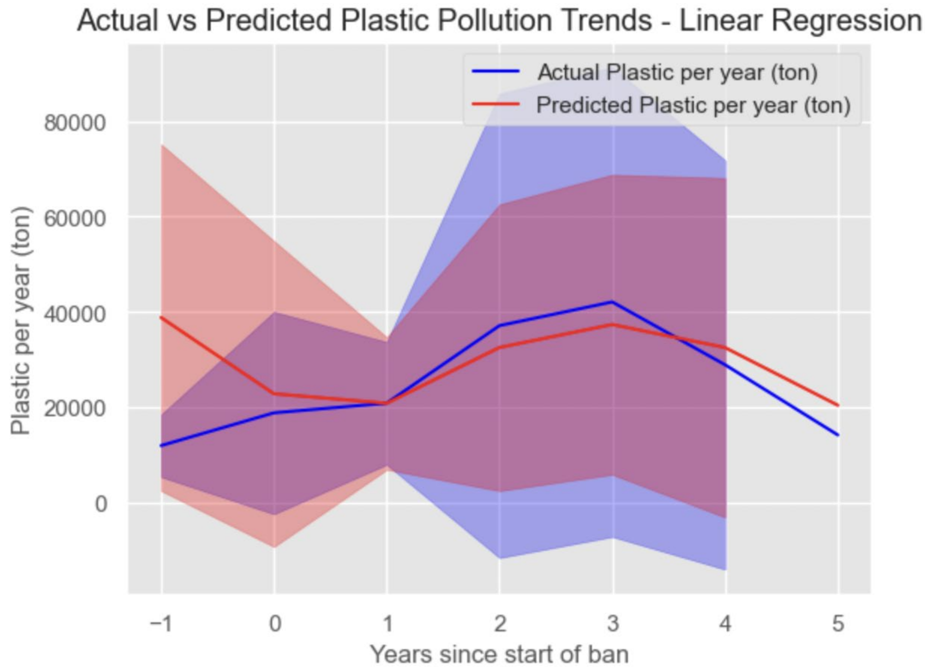

Figure C1: Actual and Predicted Plastic pollution values by Linear Regression.

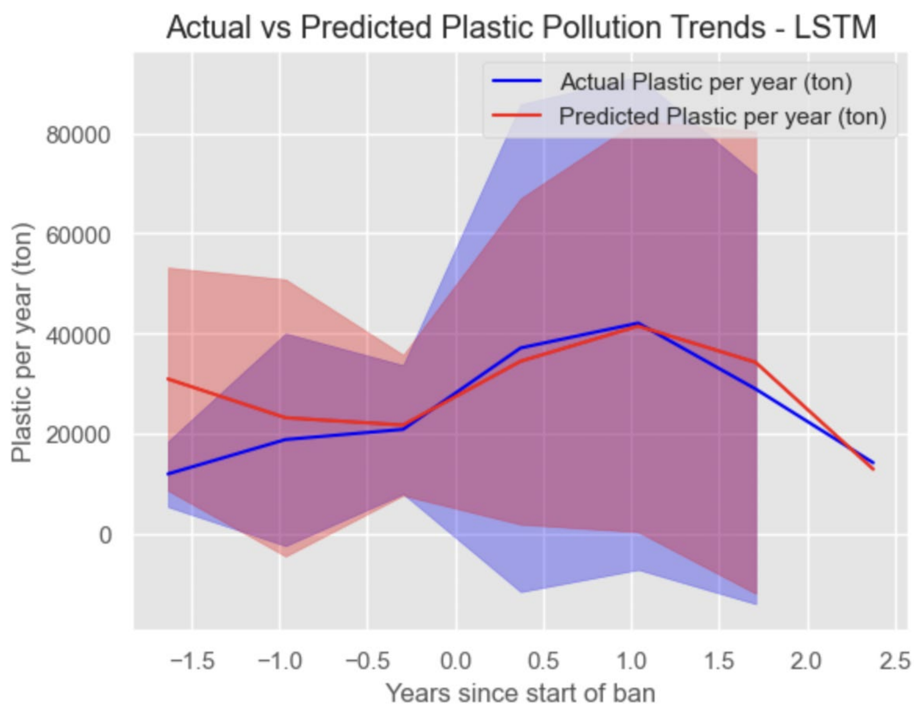

Figure C2: Actual and Predicted Plastic pollution values by LSTM.

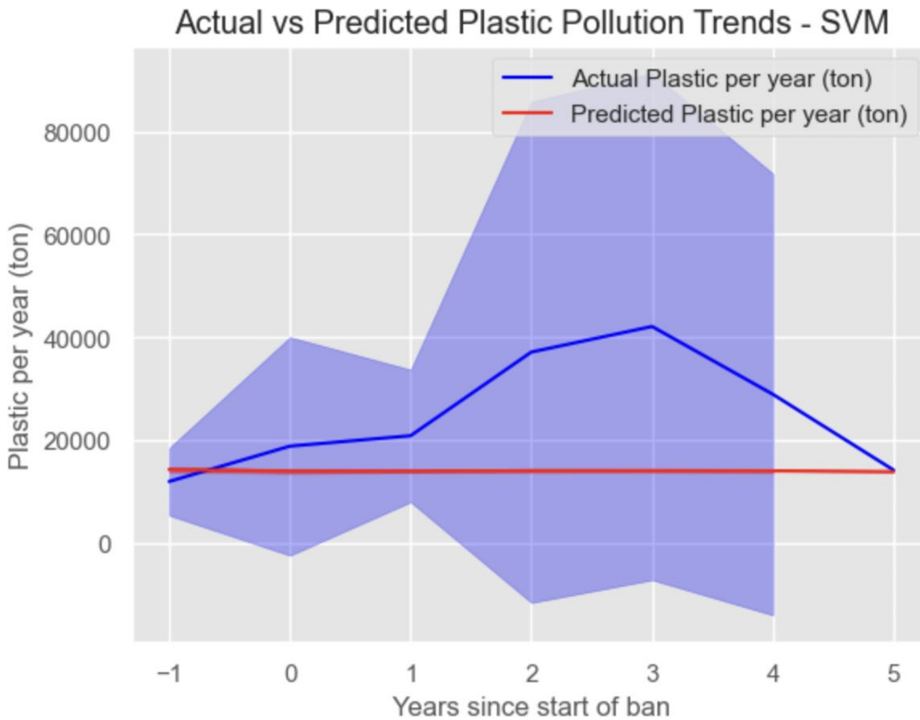

Figure C3: Actual and Predicted Plastic pollution values by SVM.

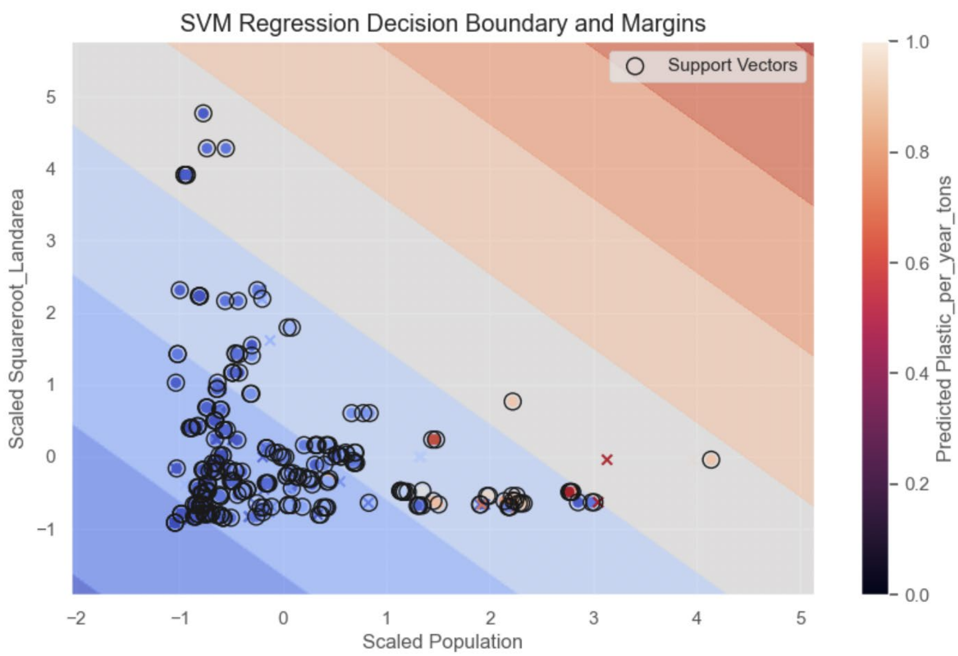

Figure C4: SVM Regression Decision Boundary and Margins.

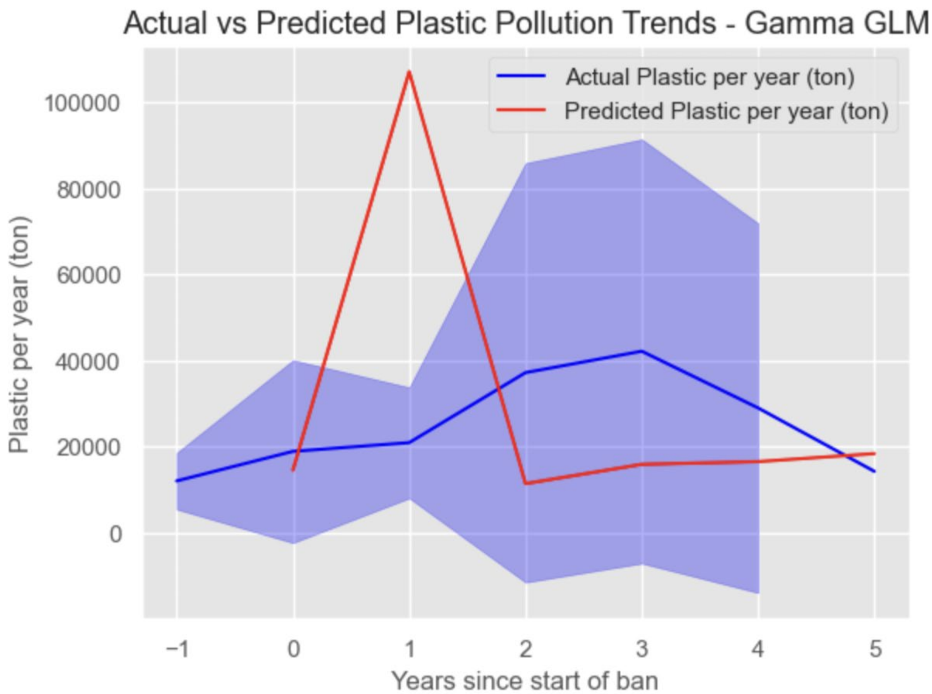

Figure C5: Actual and Predicted Plastic pollution values by Gamma GLM.

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:     Plastic_per_year_tons   No. Observations:              324
Model:                               GLM   Df Residuals:                  317
Model Family:                    Poisson   Df Model:                        6
Link Function:                       Log   Scale:                      1.0000
Method:                             IRLS   Log-Likelihood:        -1.6132e+06
Date:                   Thu, 18 Apr 2024   Deviance:               3.2227e+06
Time:                           14:06:43   Pearson chi2:             3.48e+06
No. Iterations:                        6   Pseudo R-squ. (CS):          1.000
Covariance Type:               nonrobust
======================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                  5.6602      0.010    580.248      0.000       5.641       5.679
Population          6.552e-07   5.98e-10   1096.136      0.000    6.54e-07    6.56e-07
log_GDP                0.3045      0.001    243.346      0.000       0.302       0.307
log_Gini               0.1277      0.007     18.015      0.000       0.114       0.142
squareroot_landarea    0.0021   1.66e-05    124.781      0.000       0.002       0.002
HDI                    0.0165      0.000    160.828      0.000       0.016       0.017
Years_Since_Start      0.1289      0.000    521.597      0.000       0.128       0.129
======================================================================================
```

Figure C6: Generalized Linear Model Regression Results.

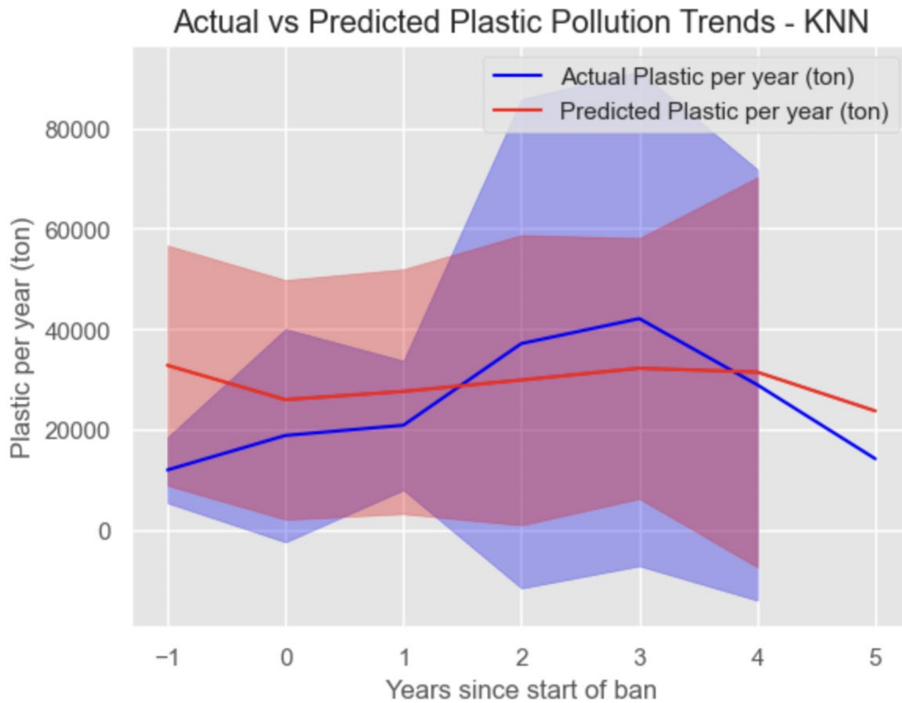

Figure C7: Actual and Predicted Plastic pollution values by KNN.

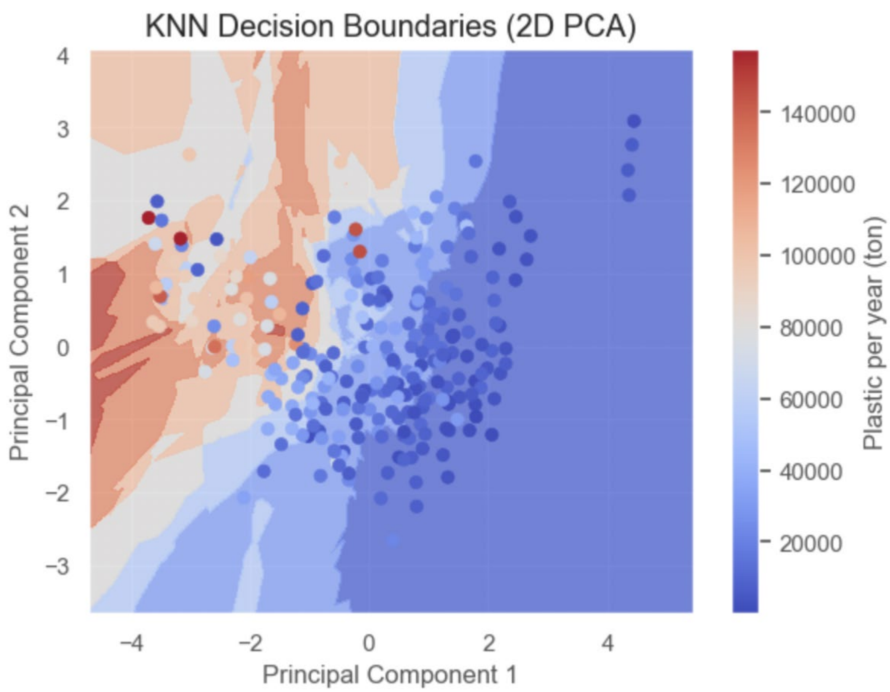

Figure C8: KNN Decision Boundaries using 2-Dimensional PCA.

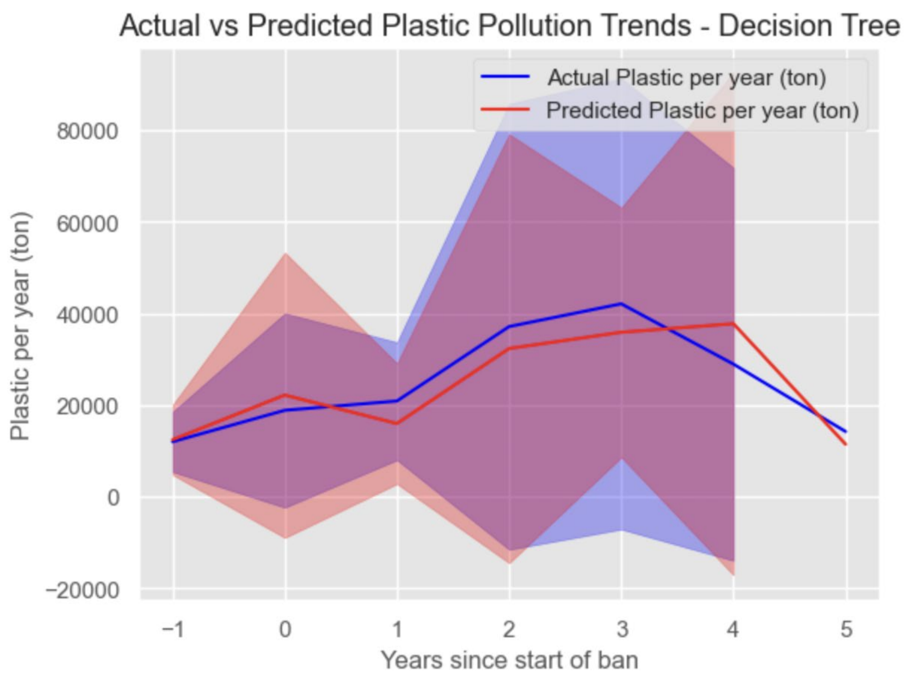

Figure C9: Actual and Predicted Plastic pollution values by Decision Tree.

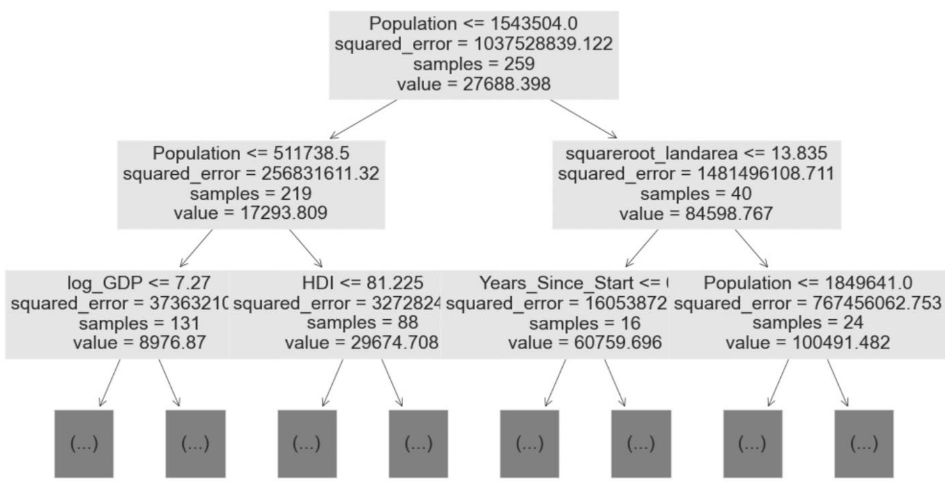

Figure C10: Top nodes of Decision Tree.

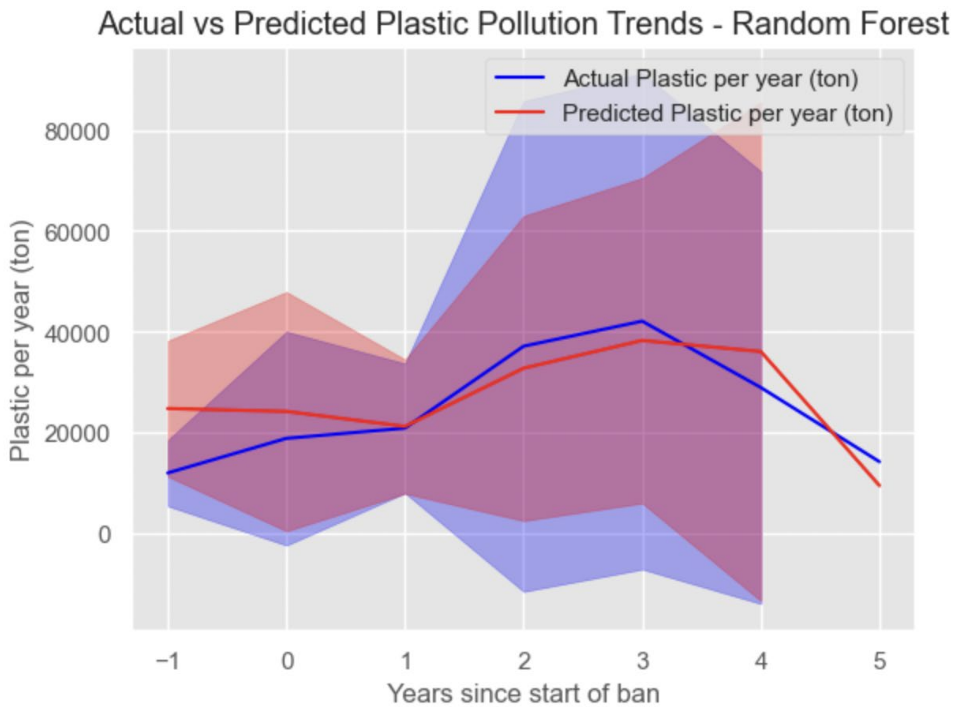

Figure C11: Actual and Predicted Plastic pollution values by Random Forest.