**Assessing the Effectiveness of US Plastic Legislations through 5 Gyres' Audits**

Group 8

Northeastern University

ALY 6080 - Integrated Experiential Learning

Professor Ajit Appati

June 19th, 2023

**Group Members:**

Mariana Aguilera

Dhairyav Jatin Shah

Likhitha Varakala

Abstract

Plastic pollution is a critical environmental issue with adverse impacts on ecosystems and human health. This report investigates the efficacy of plastic legislation and 5 Gyres' audits in addressing plastic pollution in the United States. The study analyzes trash production rates in states with different plastic pollution laws, conducts statistical tests to examine variations in total trash per million inhabitants, and develops a prediction model to assess the impact of various factors on plastic waste pollution. The data utilized in this study combines information from 5 Gyres' brand audits, legislative initiatives from the National Conference of State Legislatures (NCSL), and population data from the United States Census Bureau. Data cleaning and preparation were performed, including the imputation of missing values using text classification models. The report's findings can be summarized as follows: Inland audits in the US have a higher total count of trash per million habitants compared to other audit types. Single-layer food packaging products are the most common trash item found during the audits, but multi-layered smoking materials and food packaging are also prevalent. The number of recorded trash items varies per year, with 2018 having the highest count and 2019 the lowest. States with plastic bans demonstrate a lower total count of recorded trash per population, indicating their effectiveness in reducing plastic waste. However, states with plastic taxes exhibit higher trash counts. A comparative analysis using a two-sample t-test does not provide sufficient evidence to conclude that the mean trash per million habitants is lower in states with plastic bans compared to those with preemption laws prohibiting plastic.

Efficacy study of US Plastic Legislations and 5 Gyres’ Audits

# Introduction

Plastic pollution is a pressing environmental issue that poses significant threats to ecosystems and human health. It is characterized by the accumulation of plastic waste in various environments, including oceans, rivers, and landfills. Plastic pollution has detrimental effects on wildlife, marine organisms, and ecosystems, leading to habitat destruction, entanglement, and ingestion of plastics.

In response to the plastic pollution problem, numerous organizations, including 5 Gyres, have taken initiatives to raise awareness, conduct research, and implement solutions. 5 Gyres is a prominent organization that focuses on reducing plastic pollution through research, education, and advocacy. They have conducted extensive research on plastic pollution and organized cleanup efforts to quantify the plastic waste found in bodies of water around the world.

We aim to understand trash production rates across different US states with varying plastic pollution laws. We perform statistical tests to examine the variation in total trash per million inhabitants between states with plastic bans and those with preemption laws that prohibit plastic bans. Additionally, we develop a prediction model to investigate the impact of various factors on plastic waste pollution and make forecasts regarding the quantity of plastic waste in locations with different plastic legislation.

# Methodology

This study includes an interactive dashboard that compares trash production rates across US states with different plastic pollution laws. We also present a prediction model that investigates the relationship between plastic waste legislation, product type, material type, type of audit, layers, and time spent per volunteer, with the response variable being the total trash per million habitants. Lastly, we perform statistical tests to examine the variation in total trash per US state population between states with plastic bans and those with preemption laws that prohibit plastic bans.

*About the Data*

The study utilized data from three distinct sources. The initial dataset was obtained from 5 Gyres, encompassing comprehensive details on brand audits conducted in North America. This dataset encompasses the overall tally of polluted items discovered during cleanups organized by 5 Gyres, along with supplementary information such as location (longitude, latitude, continent, country, province, city), product type, material type, number of plastic layers, item description, brand, and collection date (spanning from 2018 to 2022).

The second data set was created by gathering information on legislative and regulatory initiatives, allowing us to categorize states based on their actions concerning plastic pollution. These actions included implementing bans on plastic, imposing taxes on plastic, having no specific legislation, or preempting bans on plastic. To gather this information, we relied on the website of the National Conference of State Legislatures (NCSL), which provides details on state legislatures' efforts to reduce plastic bag usage in grocery stores and other commercial establishments (Brief state plastic bag legislation 2021).

The third dataset was obtained from the United States Census Bureau website, which provided us with the population figures for each state from 2018 to 2022. The population values were all divided by 1 million to simplify our comprehension. Upon merging the three data sets, we computed a new variable representing the total trash per million habitants by state and year. This calculation allowed us to compare the rate of total count of trash recorded per population for each state, considering the distinct types of plastic legislation mentioned earlier.

*Relevant Variables*

We focused our analysis on the following variables shown in table 1.

Table : Description of Most Relevant Variables

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Source** |
| Year | Year audit was conducted (2018-2022) | 5 Gyres |
| Type of Product | The type of product that the trash object is by use or sector. | 5 Gyres |
| Type of Material | The type of material that the object is by polymer type. | 5 Gyres |
| Layer | The number of layers that the plastic material has. | 5 Gyres |
| Time Spent (hours) | The amount of time spent on the audit by each volunteer in hours. | 5 Gyres |
| Specifics of Audit | What land type the audit took place on, including coastal, freshwater, inland or other. | 5 Gyres |
| Plastic Legislation | The actions taken by different US states, including implementation of plastic bans, imposition of taxes on plastic, absence of any action, or preemption of plastic bans. | NCSL |
| Trash per million habitants | The total amount of waste recorded per state per year, expressed in terms of million inhabitants. | 5 Gyres, US Census |

*Data Cleaning and Preparation*

Prior to commencing the analysis, we conducted data preparation tasks that involved rectifying spelling inconsistencies and imputing missing values. Firstly, we addressed Spanish words that originally contained accents, which were causing typos. We corrected words like Los Angeles. Secondly, we tackled inconsistencies in the formatting of city or province names. For instance, we encountered entries like "St Peterburg" and "St. Peterburg," so we standardized all applicable entries to "St. Peterburg" for consistency. Thirdly, we noticed that some cities had "Florida" mentioned at the end but were not included in the province column. We rectified this by adding "Florida" to the province field. Fourthly, we made sure that all state abbreviations were converted to their full names. This involved changing "Fl" to "Florida," "MN" to "Minnesota," and so on. Fifthly, we verified the consistency between city and province entries. In cases where the province was missing, we cross-checked the location using latitude and longitude values online to fill in the blanks accurately.

To address the missing values, a text classification approach was employed to predict the missing values based on the available data. We observed the following number of missing values for the following variables to be of concern shown in table 2:

Table : Number of Missing Values in each Column

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | Product Type | Material Type | Layer | Audit Specifics |
| **Number of NA’s** | 5690 | 2258 | 4348 | 220 |

To predict the missing type of products, a text classification model was trained using the item description, layer, and material type column as the predictor variable. The data set was filtered to include only rows with non-null values in the predictor and target columns. A pipeline was constructed, consisting of a TF-IDF vectorizer for text representation and a linear support vector classifier (LinearSVC) as the classifier. The pipeline was trained on a portion of the dataset, and the accuracy of the model was evaluated on a test set. The trained model was then used to predict the missing values of the product type in the test set.

A similar process was followed to predict the missing material type. The predictor variable was the item description, layer, and product type columns, and the target variable was the material type. The dataset was filtered to include only non-null values in the predictor and target columns. The text classification pipeline was again used, and the model was trained on a subset of the data. The accuracy of the model was evaluated on a test set, and the missing material types were predicted using the trained model.

The process was repeated to predict the missing "layer" values. The item description, material type, and product type columns were used as the predictor variable, and the layer column was the target variable. The dataset was filtered to include only non-null values in the predictor and target columns. The text classification pipeline was applied, and the model was trained and evaluated on a test set. The missing layer values were then predicted using the trained model.

After predicting the missing values in the product type, material type and layer columns, the original dataset was updated with the predicted values. The missing values were replaced with the predicted values obtained from the corresponding models. To assess the performance of the prediction models, accuracy scores were calculated for each model by comparing the predicted values with the actual values in the test sets. The accuracy scores for the 3 columns product type, material type and layer were 85.67% ,66.96% and 71.69%. The counts of the original values and predicted values were compared for each column: product type, material type and layer. The differences in counts were calculated to observe the impact of the prediction on the dataset as shown in figures 1-3, respectively. The table for the above given calculation is presented in the appendix.

Figure : Type of Material Imputation Results

A picture containing screenshot, software, multimedia software, design

Description automatically generated

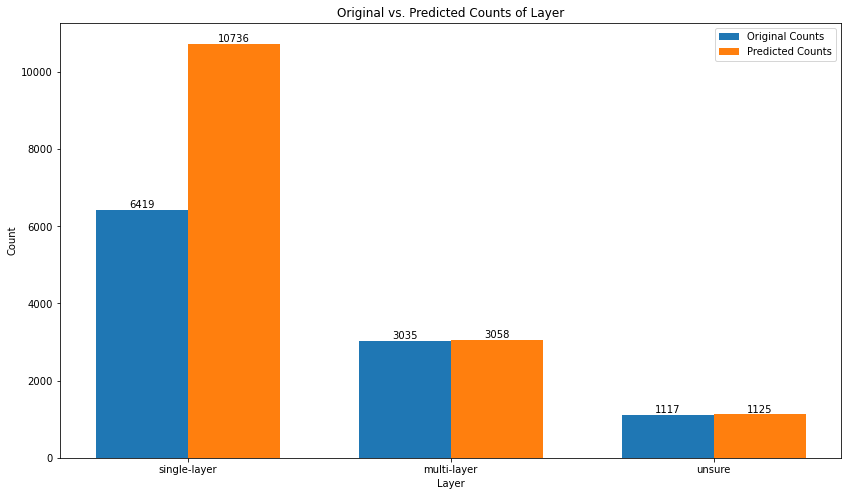
For context, the types of materials mentioned in figure 1 correspond to the following abbreviations: PP (Polypropylene), HDPE (High-Density Polyethylene), LDPE (Low-Density Polyethylene), PET (Polyethylene Terephthalate), PS (Polystyrene), PVC (Polyvinyl Chloride), and O (Other).

Figure : Type of Product Imputation Results

A picture containing screenshot, multimedia software

Description automatically generated

Figure : Layer Imputation Results



# Exploratory Data Analysis

Our Exploratory Data Analysis (EDA) results can be categorized into three main themes: location analysis, time analysis, and audit-specific analysis. These findings were subsequently consolidated into an interactive output that incorporates filters, allowing you to observe how observations would be affected by examining specific aspects of the data.

*Audit-Specific Analysis of Plastic Pollution in the US*

*A picture containing text, screenshot, diagram, rectangle

Description automatically generated* Figure 4 was created to show the total trash per million habitants by audit type. We observe that the aggregate of the total count of trash with respect to the states’ population is higher for the audits conducted inland compared to the rest.

Figure

A screen shot of a graph

Description automatically generated with low confidence Within this theme, figure 5 depicts a heatmap illustrating the total trash per million habitants, categorized by product type and layers. Based on our findings, we can observe that the predominant characteristic of the trash discovered during the 5 Gyres' audits was single-layer food packaging products. However, multi-layered smoking materials and multi-layered food packaging were also prevalent in the audits.

Figure

*Time Trend Analysis of Plastic Pollution in the US*

A picture containing text, screenshot, diagram, rectangle

Description automatically generatedHaving explored the association between the total trash count and audit specifics, we did an examination of the quantity of recorded trash for all audits per year. To do so, we created figure 6. The plot presents the total trash per million habitants and number of events by year. The goal is to see how the number of trashes recorded for all states varies per year. Based on the plot, it is evident that 2018 had the highest recorded number of trashes, while 2019 had the lowest. In general, the number of events is also higher for the years with higher number of total trash per million habitants. It is important to note that this does not necessarily indicate a decrease in the amount of plastic found in 2019 compared to other years. Rather, it reflects the quantity of trash recorded by 5 Gyres during their audits in different years.

Figure

*Geospatial Distribution of Plastic Pollution in the US*

Finally, we generated figures 6 and 7 to visually represent the data we gathered regarding the plastic bans legislations implemented in different states across the United States.

A map of the united states

Description automatically generated with medium confidenceFigure 7 presents a US map that distinguishes the plastic bag legislation in each state. It is important to note that not all states have been colored in the map. This arises from the fact that audits were not conducted in some US states. However, it is essential to understand that the absence of audits in these states does not imply an absence of trash being found within them.

Figure

A picture containing text, screenshot, diagram, line

Description automatically generatedFigure 8, shown below, is directly related to the map, given that it also presents information related to the plastic legislations in the US. We created the following boxplot of the log (total trash per million habitants) by plastic legislation type. This connects all the information presented above to understand how effective each of the plastic legislation types have been with respect to the information collected by 5 Gyres. Providing context, the negative log values arise because the ratio between the total count and population is less than 1. Notably, we observe that states in the US that have implemented plastic bans demonstrate the lowest total count of recorded trash per population. This finding suggests that banning plastic has proven effective in reducing the quantity of plastic identified in 5 Gyres' audits. However, it is noteworthy that states implementing taxes against plastic exhibit the highest median value. It is important to emphasize that these findings solely pertain to the trash items recorded during audits conducted by 5 Gyres. Consequently, if more audits were conducted in states with tax legislation on plastic, more trash items would be documented.

Figure

*Interactive Dashboard*

The finalized dashboard containing the plots described can be found in figure 9:

Figure

A picture containing text, map, diagram, screenshot

Description automatically generated

The dashboard (figure 9) includes various filters that provide interactive functionality. You can click on the colored states displayed on the map to filter the data accordingly. Additionally, you have the option to click on any bar plot associated with the years to further refine the information. Finally, you can filter the type of audit by clicking on the relevant bar, allowing for a more targeted analysis. Observe figure 10 to see how the dashboard changes when we click on California only:

Figure

*A screenshot of a graph

Description automatically generated with low confidence*

For California, the total trash count per million habitants was significantly higher in 2022 compared to other years.

Instead, if we click the year 2022, you can observe how the plots change in figure 11.

Figure

A screenshot of a graph

Description automatically generated with low confidence

In 2022, some states, such as New Mexico, did not conduct any 5 Gyres’ audits.

Finally, we can observe how the dashboard changes when we click on coastal only in figure 12.

Figure

A screenshot of a graph

Description automatically generated with low confidence

In this scenario, only the states with a coastline are colored on the map. Furthermore, we can observe a substantial disparity in the trash count per million habitants for single-layered food packaging products compared to other categories.

# Regression Analysis: Predicting Plastic Waste Quantities in US States

For the prediction model, we have opted for a regression analysis. We utilized the logarithm of the total trash per million habitants as our response variable and examined its relationship with predictors such as plastic waste legislation, product type, and type of audit, layers, and time spent per volunteer. To clarify the calculation process, let's revisit the method for determining the total trash per million habitants. This metric is derived by dividing the total amount of trash found in a particular US state and year by the corresponding population value of that state during the same period. By employing regression analysis, we aim to estimate the impact of these predictors on plastic waste pollution and make forecasts regarding the quantity of plastic waste in locations with different plastic legislations.

Before we ran the model, we divided our data set into training and test set. The model was first trained on the training data set with 70% of the observations. Afterwards, we tested the model using the remaining 30% of the observations. Additionally, we performed a log transformation of our response variable given its skewness to the right. In table 3 below you can observe our regression results for the train set.

Table



From the regression summary, several observations can be made. The train set regression’s R-squared value of 0.2558 suggests that the model explains approximately 25.58% of the variance in the target variable. Additionally, to determine the performance of our model, we compare the root mean square error (RMSE) values of the train and test set. The RMSE for the train set was 12.86229 and for the test set 10.67417. The RMSE of the test set is smaller than the RMSE for the train set, indicating that there is a small likelihood of overfitting.

Important observations from our train model coefficient results include:

* Compared to states with plastic ban legislation, in trash cleanup events held during 2018-2022, volunteers found about 3.71 more units of product items per million habitants in states with no plastic legislation.
* Compared to states with plastic ban legislation, in trash cleanup events held during 2018-2022, volunteers found about 2.99 more units per million habitants in states with preemptive plastic legislation.
* Compared to states with plastic ban legislation, in trash cleanup events held during 2018-2022, volunteers found about 7.82 more units of product items per million habitants in states with no tax statute for plastic use.

These findings suggest that banning plastic has proven effective in reducing the quantity of plastic identified in 5 Gyres' audits. However, it is noteworthy that states implementing taxes against plastic exhibit the highest trash number per state population.

# Comparative Analysis of Trash Levels Across States: Two-Sample t-Test

Considering our emphasis on the effectiveness of plastic legislation, we employed statistical analysis techniques to examine whether there was a difference in the average trash per million habitants between states with plastic bans and those with preemption laws that prohibit plastic bans. The hypothesis we examined aimed to determine if the average amount of waste per million habitants in states with plastic bans is lower compared to those with preemption laws prohibiting plastic. To investigate this, we performed a two-sample t-test in R and obtained the results displayed in Table 4.

Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Comparison of States with Bans vs Preemption Using Welch Two-Sample T-Test | | | | | |
| **Groups** | **Means** | **Difference** | **t-test** | **df\*** | **p-value** |
| Bans | 0.56 | -0.36 | -1.43 | 4250.5 | 0.07 |
| Preemption | 0.92 |  |  |  |  |

\*df: degrees of freedom

We resulting t-statistic was -1.43 and the corresponding p-value was 0.07. Since the p-value (0.07) is greater than our significance level α = 0.05, we fail to reject the null hypothesis. Hence, we do not have sufficient evidence to conclude that the mean trash per million habitants for the states with plastic bans is lower compared to those with preemption laws prohibiting plastic.

# Recommendations

Based on our report findings, we recommend 5 Gyres to keep refining their plastic audits in the US and worldwide. For instance, we found that audits were not conducted in some US states, which does not imply an absence of trash. It is recommended for the company to expand their audit coverage to include states where audits have not been conducted. This will provide a more comprehensive understanding of plastic pollution across the United States and enable more accurate comparisons between states.

Secondly, the analysis indicates that states with plastic bans demonstrate the lowest total count of recorded trash per population. This finding suggests that banning plastic has proven effective in reducing plastic pollution. The company should consider using their findings to advocate for the implementation of plastic bans in states that do not currently have such legislation. This can be done through partnerships with environmental organizations, engaging with policymakers, and raising public awareness about the benefits of plastic bans.

Finally, the report reveals that states implementing taxes on plastic exhibit higher median values of recorded trash per population. It is recommended for the company to further investigate the reasons behind this observation. Conducting additional research and analysis to assess the effectiveness of tax legislation on reducing plastic pollution would provide valuable insights. This can involve studying the implementation and enforcement of existing tax laws and identifying potential areas for improvement or alternative approaches to reducing plastic waste.

References

*Brief state plastic bag legislation*. National Conference of State Legislatures. (2021, February 8). Retrieved May 1, 2023, from https://www.ncsl.org/environment-and-natural-resources/state-plastic-bag-legislation

Bureau, U. C. (2021, December 16). *State Population Totals: 2010-2019*. Census.gov. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html

Bureau, U. C. (2023, March 23). *State population totals and components of change: 2020-2022*. Census.gov. https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html

Appendix

1. **Data Cleaning (NLP) Results**

