Predicting Future Impacts from Plastic Waste on Oceans

Sarah Theriot

DSC630

December 4th, 2024

***Predicting Future Impacts from Plastic Waste on Oceans***

**Introduction**

Plastic pollution has become a growing environmental crisis, impacting oceans and marine wildlife across the globe. As someone working firsthand in conservation education, I have witnessed the devastating effects of plastic waste on our oceans. These effects include dangers to marine wildlife and ecosystem degradation, all of which are increasingly felt worldwide. However, the issue extends beyond environmental concerns. Plastic waste also poses a significant business problem, as companies face rising pressure from consumers, local governments, and environmental organizations to adopt sustainable practices. One of the most crucial shifts involves transitioning from single-use plastics to reusable products. This project aims to leverage predictive analytics to forecast future trends in plastic waste production, with the goal of encouraging companies to make this switch. By adopting more sustainable practices, businesses can not only contribute positively to fragile ecosystems but also enhance consumer satisfaction and public perception.

**Project Overview**

The core of this project is to use predictive models to analyze global plastic waste production trends and forecast the potential impact of switching to reusable products. I began with a linear regression analysis to identify key factors influencing plastic waste production across the globe. This allowed me to examine relationships between variables such as national recycling rates, per capita plastic consumption, and waste management practices. The Global Plastic Waste 2023 dataset, which contains data on waste production, recycling rates, and risk management practices, served as the primary source of information. I planned to use this dataset to identify correlations and how these variables might affect one another, particularly in regions with high coastal risk. Additionally, I explored time series forecasting to project future plastic waste trends, relying on historical data from this dataset, along with potential additional datasets on global plastic waste trends and reusable product impacts.

**Evaluation and Data Accuracy**

To evaluate the accuracy of my models, I used two essential metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics are vital for assessing the reliability of the regression and time series models, ensuring that the models could accurately forecast future trends. Before running these models, I performed an exploratory data analysis (EDA) to clean and inspect the data, removing any missing or inaccurate data points. Ultimately, the goal was to determine if predictive models could reliably forecast plastic waste production trends and guide businesses in transitioning to more environmentally friendly alternatives.

**Ethical Considerations and Risks**

As I developed this project, it was crucial to address potential risks and ethical challenges, particularly regarding the quality of the dataset. The accuracy of the models is directly tied to the quality of the data; therefore, identifying and using high-quality, complete datasets is essential. Any inaccuracies or missing data could lead to misleading results. To mitigate this risk, I planned to clean the data thoroughly and cross-check it against multiple sources. Another ethical consideration was the risk of oversimplification. Predictive models can provide insights, but they cannot capture all the nuances of the plastic waste issue, such as cultural practices, societal norms, and accessibility to waste management systems. While the models can offer valuable insights, real-world solutions will require a more comprehensive understanding of these factors. Maintaining transparency throughout the project was critical for credibility and ensuring that these ethical concerns were addressed adequately.

**Contingency Plan**

In case the models did not yield reliable results, I had a contingency plan in place. If my initial regression models were ineffective, I planned to explore alternative datasets or adjust my methodology by incorporating decision tree models or random forests. These models are better suited to capturing non-linear relationships between variables that linear regression might overlook. While I hoped everything would go as planned, I knew it was essential to remain flexible and ready to pivot if necessary.

**Conclusion**

By completing this project, I aimed to gain a deeper understanding of predictive analytics and its real-world applications, particularly in the context of environmental issues that I am passionate about. I was excited to explore the complexities of plastic waste management and its potential solutions, and I looked forward to learning from both the challenges and successes that came along the way.

**Revised Approach and Dataset Review**

As I continued working on this project, I revisited my approach and reassessed the dataset to ensure that I could answer all of the critical questions regarding plastic waste reduction, recycling rates, and waste management practices. After delving deeper into the Global Plastic Waste 2023 dataset, I felt confident that it contained key variables such as per capita plastic consumption, national recycling rates, and waste management practices. These variables were directly aligned with the research questions I sought to explore, enabling me to draw meaningful conclusions. However, the complexity of the issue required me to adjust some of my initial expectations, which I will explain further in this section.

**Visualizations and Analysis Techniques**

To effectively communicate my findings, I planned to use various types of visualizations, which would aid in explaining the data in an accessible manner. Scatterplots would help illustrate the relationship between plastic consumption and recycling rates, allowing for the identification of any trends or correlations. Additionally, I planned to use line graphs for time series forecasting to show projected trends in plastic waste production. By using these visualizations, I could better understand how these variables change over time and present the findings in a clear and compelling way.

**Data Inconsistencies and Adjustments**

As I continued exploring the dataset, I realized that some inconsistencies existed in how different countries reported their recycling rates and waste management practices. Fortunately, these inconsistencies were not widespread, and I was able to manually correct the minor errors. With these adjustments made, I was able to proceed with my analysis using an updated and more accurate dataset.

**Modeling Decisions and Potential Adjustments**

As I entered the data preparation phase, I revisited my initial modeling approach. I remained committed to using linear regression and time series forecasting as my primary methods for analysis, but I also anticipated that incorporating decision tree models or random forests could be beneficial. These models could capture non-linear relationships that linear regression might miss, ensuring that I didn't overlook important patterns in the data. In terms of model evaluation, I continued to rely on RMSE and MAE to assess the accuracy of the predictions. However, I remained flexible and open to adjusting my methodology as necessary throughout the process.

**Anticipating Complexities**

Although my original expectations were reasonable, further analysis revealed the complexity of the data and the broader factors influencing plastic waste management. I recognized that cultural practices, local regulations, and waste management infrastructure play crucial roles in how plastic waste is handled, but these factors are not always reflected in the dataset. Therefore, while predictive models could provide valuable insights, they would not capture the full spectrum of variables that influence waste management. It was important to keep these broader, non-quantifiable factors in mind as I moved forward with the project.

**Interpretation of Linear Regression Results**

As I ran the linear regression model, I obtained some insightful results that helped me refine my understanding of the relationships between plastic waste, mismanaged waste, and recycling rates. The R-squared value of 0.06472 suggested that the model could only explain 6.5% of the variability in mismanaged plastic waste per capita. This low R-squared value indicated that other factors—such as local waste management infrastructure, public behavior, or governmental policies—likely influenced mismanaged plastic waste but were not captured in the model.

The coefficients from the regression model revealed a positive relationship between total plastic waste and mismanaged plastic waste (p-value = 0.00436), suggesting that as total plastic waste increases, so does the amount of mismanaged waste per capita. Conversely, the relationship between recycling rates and mismanaged waste was negative but not statistically significant (p-value = 0.38437). This result was surprising, as I had anticipated that higher recycling rates would significantly reduce mismanaged plastic waste. However, the lack of significance for recycling rates indicated that other factors might be at play in addressing the plastic waste issue.

**Reflection on the Hypothesis**

Looking back at my hypothesis, I found that it held true for part of the prediction. The correlation between total plastic waste and mismanaged waste was significant, supporting the idea that as plastic waste increases, so does mismanagement. However, my assumption that higher recycling rates would lead to reduced mismanagement did not materialize as expected. This finding suggests that simply increasing recycling rates may not be enough to tackle the broader plastic waste problem. In reality, addressing plastic waste will require a multifaceted approach that includes better waste management infrastructure, public education, and stronger policies.

**Limitations of the Analysis**

Several limitations of this analysis must be acknowledged. The low R-squared value indicates that the model does not explain much of the variation in mismanaged plastic waste, suggesting the presence of many unaccounted-for factors. Additionally, the dataset was cross-sectional, meaning it only captured data from one point in time. A longitudinal study would provide more insights into how trends evolve and how policies or initiatives impact waste management over time. Finally, the issue of missing data was a recurring challenge, and the removal of observations with missing values could have introduced bias into the analysis.

**Reflection on Data Collection and Cleaning**

The data collection and cleaning process presented several challenges. In particular, dealing with missing data and inconsistencies in column names required significant effort. Despite these obstacles, I was able to clean the dataset and standardize country names, ensuring that the data was usable for analysis. However, the removal of rows with missing values meant that some observations were excluded, which may have impacted the representativeness of the dataset.

**Suggestions for Future Research**

Based on the findings and limitations of this study, I have several suggestions for future research. Including additional variables, such as economic development, government policies, or public awareness programs, would likely improve the model’s ability to explain variance in mismanaged plastic waste. A longitudinal study could also provide more clarity on how recycling and waste management strategies evolve over time. Finally, focusing on specific regions or countries—especially developing countries—could reveal valuable insights into how different economic and social contexts influence waste management practices.

**Conclusion**

This analysis provided valuable insights into the relationship between total plastic waste and mismanaged waste, highlighting the importance of better waste management systems. However, the lack of significance for recycling rates suggested that increasing recycling alone may not be enough to address the issue. Future research should explore the broader factors that influence waste management and consider a more comprehensive approach that includes infrastructure, policies, and public engagement. This study underscored the need for a multifaceted solution to plastic waste and opened up many avenues for further research.

**References**

Dongre, P. (2023). Global Plastic Waste 2023: A Country-wise Analysis. Kaggle. https://www.kaggle.com/datasets/prajwaldongre/global-plastic-waste-2023-a-country-wise-analysis