Predictive Analytics Project

Team: 3 Js 1 Imposter
Team Members: Amber, Javerine, Jonathan, Jess

Members

Amber

Plastic

Jess

Food Waste

Javerine

Energy - Electricity

Jonathan

Carbon Emission

TABLE OF CONTENTS

01

Project Scope

Background Information

Many Singaporeans share the sentiment that addressing climate change might seem futile, given concerns about costs, inconvenience, and limited impact. In 2021, 67% worried about costs, 66% saw it as inconvenient, 27% struggled with sustainable habits, 24% found the status quo insufficient, and 22% deemed individual action insignificant. While reversing climate change demands persistent effort, there's a drive to reduce emissions, plastic, food, and energy wastage. Singapore's Environmental Performance Index (EPI) reflects this effort, with a 50.9 score in 2022, ranking 44th globally, showing a 3.7 increase over a decade. Despite a considered "poor" score, the upward trend signals improving sustainability. The urgency of environmental pollution's impact necessitates global efforts in reducing waste and emissions

Business Scenario

Our project aims to provide actionable insights into current consumption trends, aiding community organizations, non-profits, and grassroots initiatives in resource reduction efforts. Through predictive analysis, we predict resource consumption patterns across sectors, empowering businesses to align with the Green Plan's Objectives. By optimizing resource management and fostering sustainability, we contribute to a greener future and business growth. Our initiative also benefits academia, offering valuable data for studies in sustainability and predictive modeling. Emphasizing awareness, innovation, and sustainable practices, our project supports a more environmentally responsible future. Timely action is crucial for delaying environmental impacts and securing a planet for future generations

BUSINESS Goal

Improving Singapore environment sustainability by reducing food, plastic, energy, and carbon emissions consumption.

Success Criteria

- Develop predictive models of less than 0.05 Validation ASE
- (And/Or) Adjusted R-Square to be greater than 0.7

02

My Objective

Predict the top contributors to **plastic** waste to raise awareness and reduce plastic wastes in those areas.

Reason for choosing Objective Plastic

Since its invention in 1950, plastic has experienced a **remarkable surge** in popularity and usage, growing exponentially over the years. Its appeal stemmed from being a durable, flexible, and cost-effective alternative to traditional materials like wood. This presents a grave concern as its **eternal presence** on our planet **leaves a lasting imprint with devastating consequences for our environment and ecosystems**. Despite its convenience and versatility, the longevity of plastic waste perpetuates a cycle of pollution and harm, exerting a profound toll on our delicate ecosystems, wildlife, and even human health.

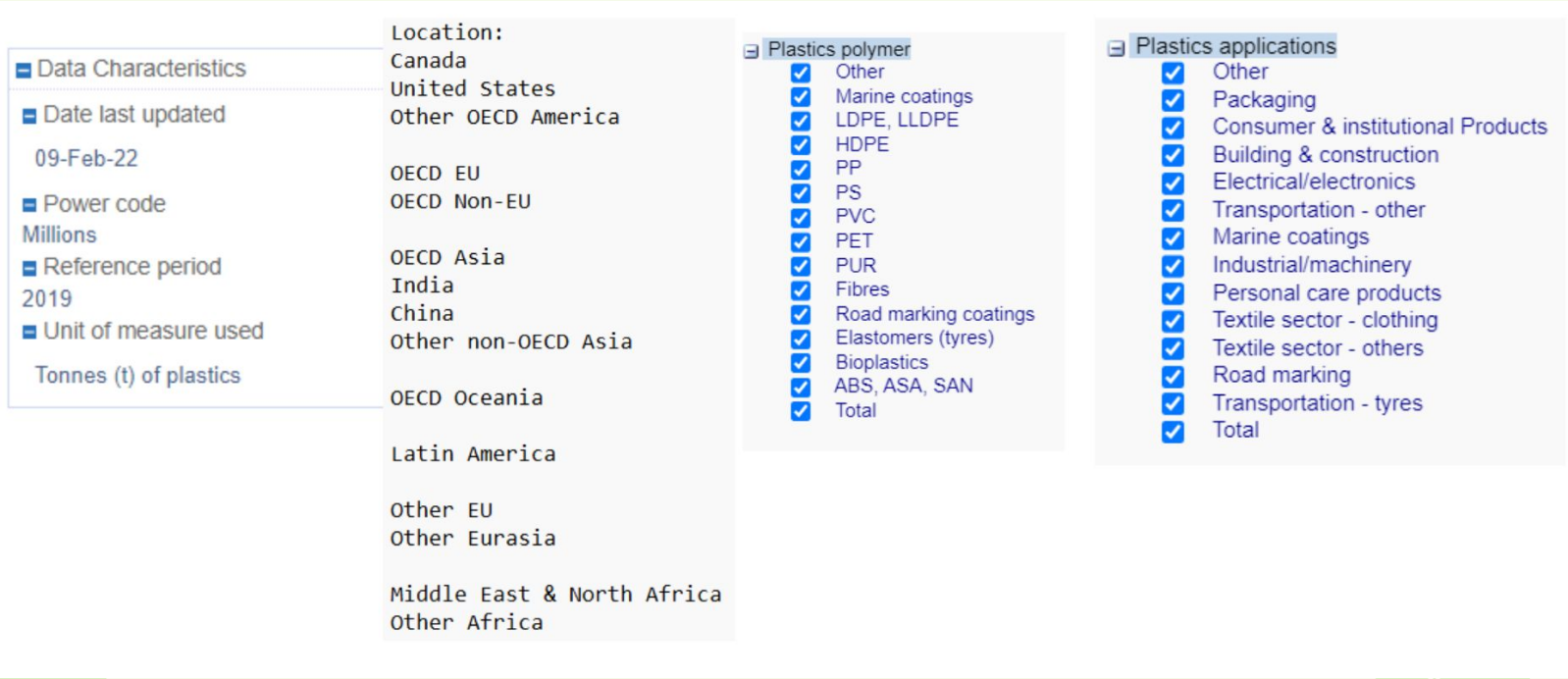
Singapore, as an active participant in environmental conservation, has an essential role to play in reducing plastic waste. By focusing efforts on the sectors responsible for the highest waste production, significant strides can be made in curbing plastic waste. Predictive analysis on the trend of plastic waste generated by specific sectors serves as **a crucial tool for drawing attention to areas that require immediate attention**. Identifying these sectors early on **enables proactive measures to be taken, reducing plastic waste before the numbers escalate to unmanageable levels.**

Data Overview

- Data Source: Organisation for Economic Co-operation and Development (OECD)
  - work together with governments
  - provide a unique forum and knowledge hub for data analysis
  - their data is highly generalised based on location/country and does not infringe on any personal data protection rules.
- From 2019
- Target Column: Plastic Waste (in million tonnes)
- Feature Column(s): Plastic Applications, Plastic Polymer
- Number of rows: 1794 (before) -> 1504 (after)

Data Understanding

Variables/categories in each columns



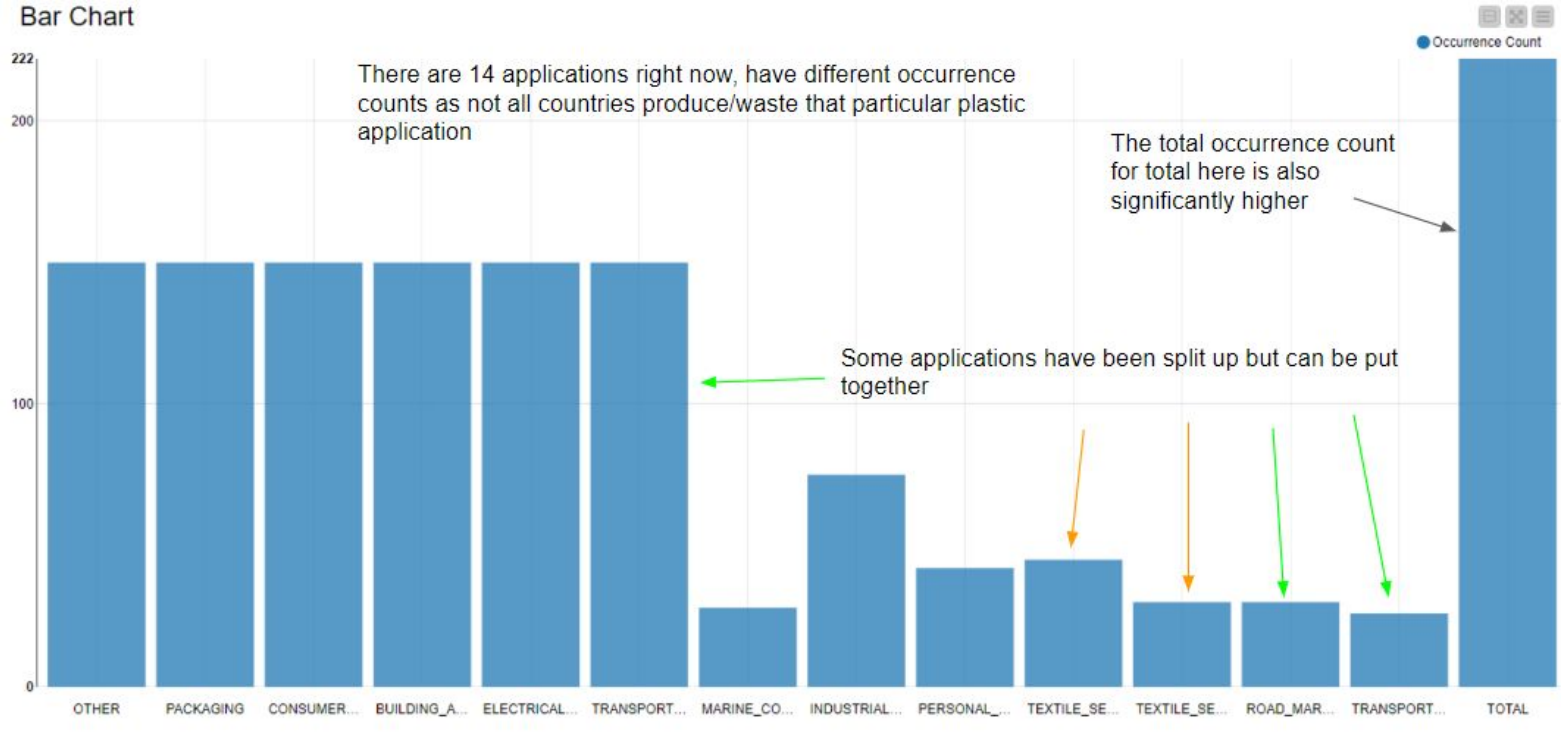
Data Understanding

Plastic **Waste** data (1794 rows)	
Column name	Details
LOCATION	Short forms of the actual location
Location	Full location name
PLASTICS_POLYMER	Different polymer types in all caps
Plastics polymer	Same as PLASTICS_POLYMER but not all caps
PLASTICS_APPLICATION	Different sectors that use/waste the plastic and in all caps
Plastics application	Same as PLASTICS_APPLICATION but not all caps
Value	Plastic **waste** in million tonnes
Flag Codes	Empty column
Flag	Empty column

These are the columns in the original data and the categories in the different columns. There are quite a few duplicated columns and unnecessary columns like Flag and Flag code. Other category in application includes cosmetics & beauty products, toys and games, and medical device etc, while Other category in polymer includes polymers like PLA (commonly used in packaging) and PMMA (also known as acrylic).

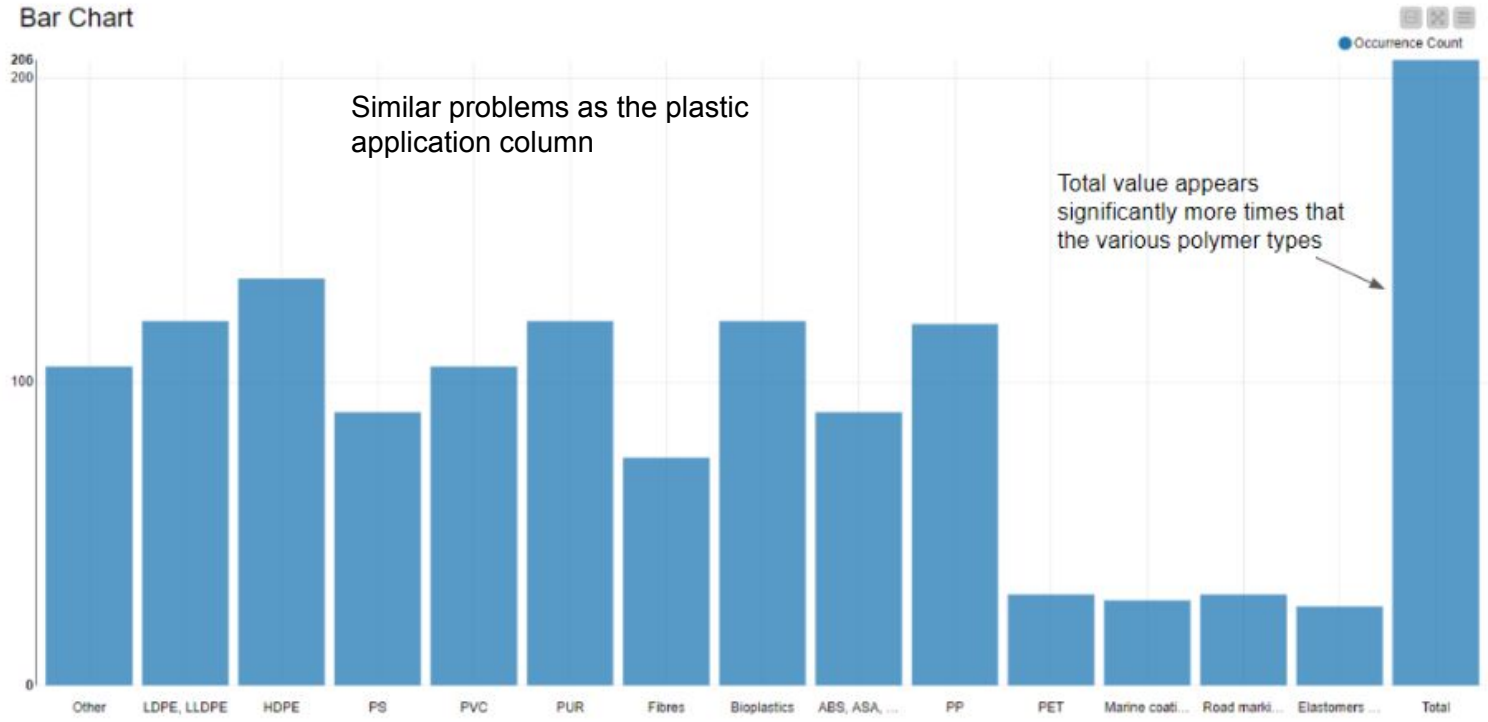
Data Understanding

Occurrence count for plastic applications



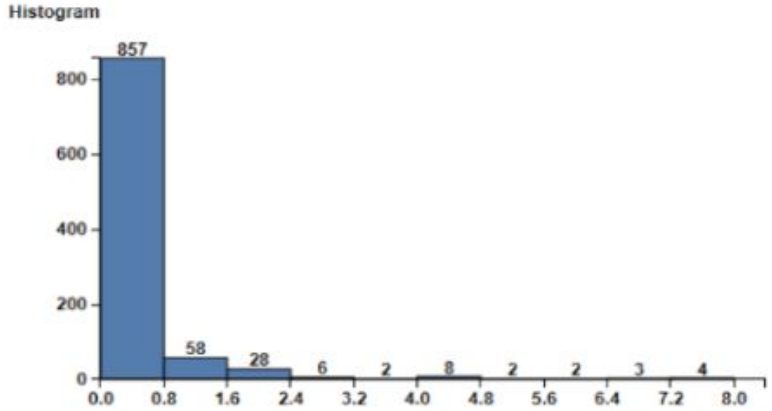
Data Understanding

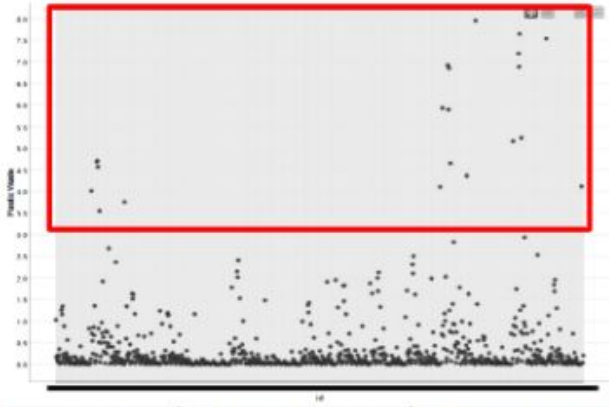
Occurrence count for plastic polymer column



Data Understanding

Outliers



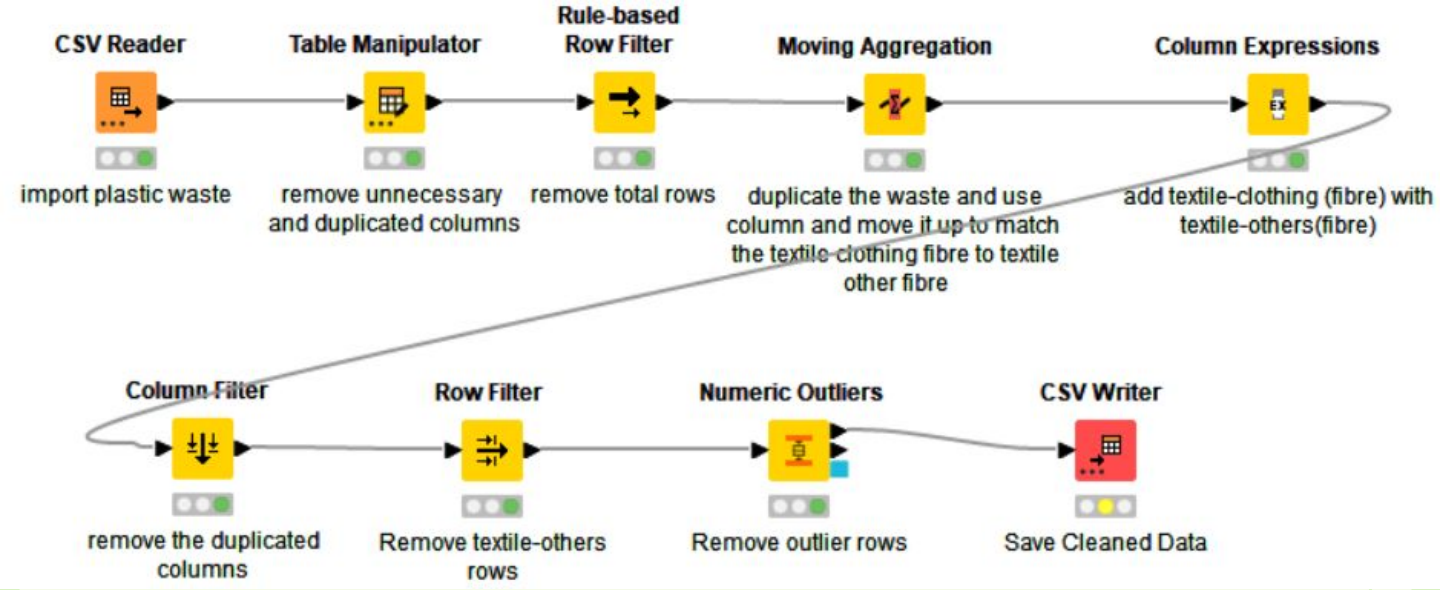


There are outliers because different countries use different amounts of plastic and hence countries like China and US who have a larger population would use more and have more waste

Location	Plastics polymer	Plastics applications	Plastic Waste
China	PP	Packaging	7.642
China	LDPE, LLDPE	Packaging	7.189
United States	LDPE, LLDPE	Packaging	6.918
China	HDPE	Packaging	6.881
United States	PP	Packaging	6.85
United States	Other	Other	5.93
United States	HDPE	Packaging	5.889
China	PET	Packaging	5.238
China	Other	Other	5.161
United States	Fibres	Textile sector - clothing	5.142
China	Fibres	Textile sector - clothing	4.969

Data Cleaning

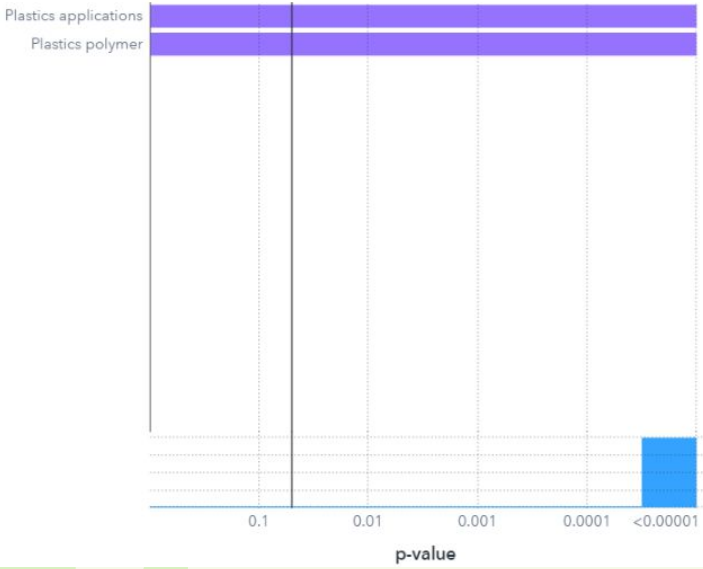
- Combine overlapping applications (14 applications -> 10 applications)
- Remove outliers
- Remove total category
- Remove unnecessary columns



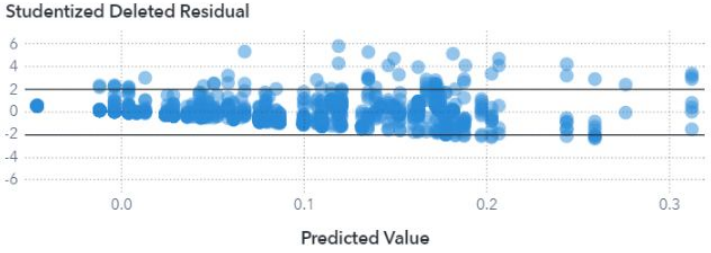
Modelling: Linear Regression

Linear Regression **Plastic Waste** | Validation ASE | **0.0083** | Observations Used **1,504** | Create Pipeline

Fit Summary



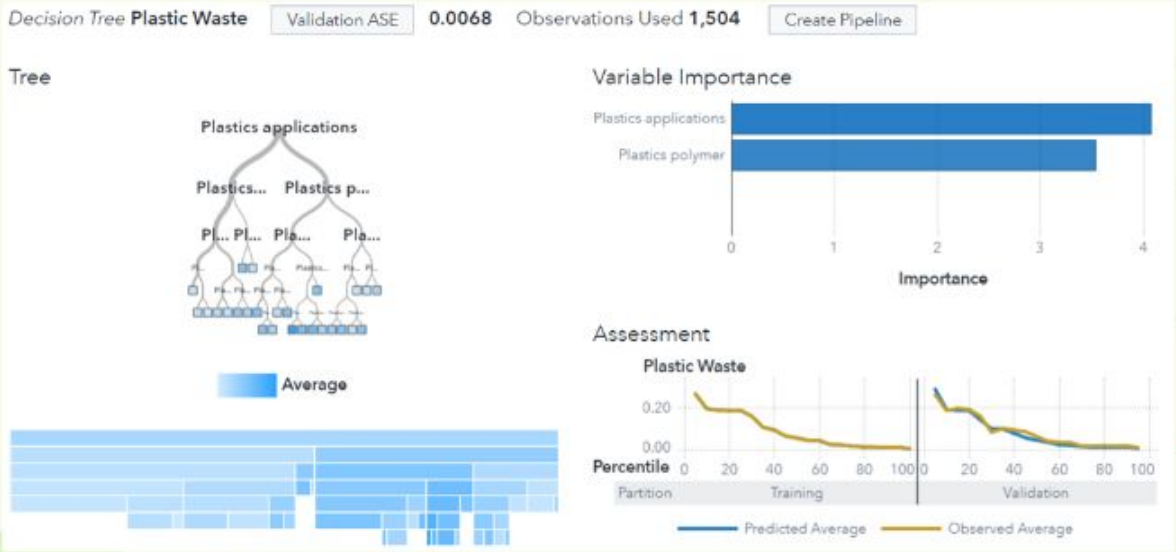
Residual Plot



Assessment

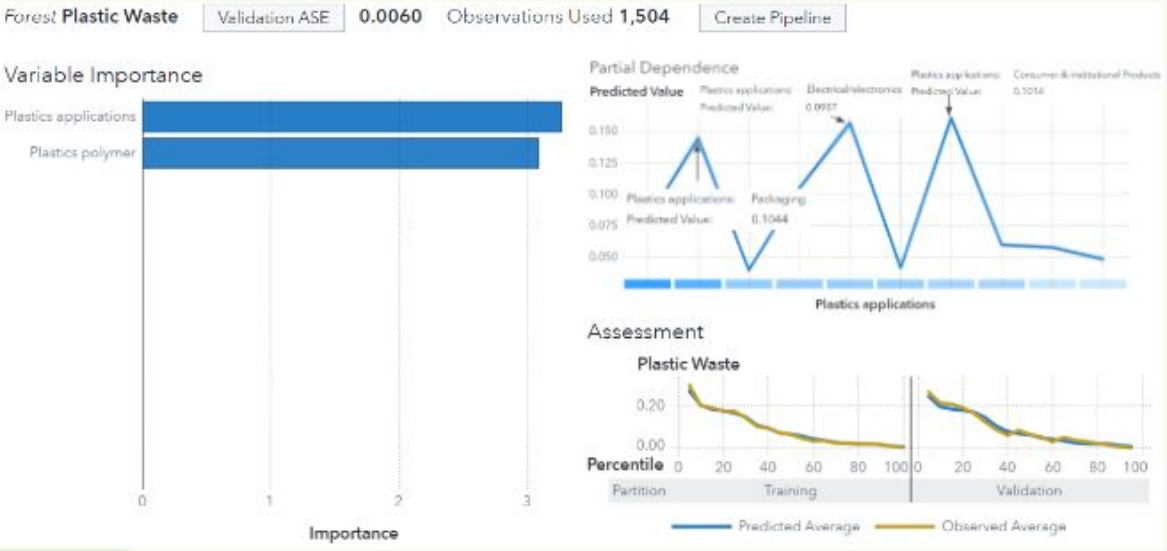
Adjusted R-Square	0.3629
AIC	-4,637.69
AICC	-4,636.76
ASE	0.0075
F Value of Model	33.61
Mean Square Error	0.0076
Observed Average	0.2278
R-Square	0.3740
Root MSE	0.0874
SBC	-5,730.66
SSE	9.0181
✓ Validation ASE	0.0083
Validation Observed Average	0.2581

Modelling: Decision Tree



	Parameters Tuned	ASE
Default	Min value: 1, Max branches: 2, Max levels: 6, Leaf size: 5	Validation: 0.0079 Training: 0.0075
Tuned 1	Min value: 1, Max branches: 2, Max levels: 8, Leaf size: 10	Validation: 0.0067 Training: 0.0065
Tuned 2	Min value: 1, Max branches: 2, Max levels: 7, Leaf size: 11	Validation: 0.00068 Training: 0.00067
Tuned 3	Min value: 1, Max branches: 2, Max levels: 12, Leaf size: 11	Validation: 0.0068 Training: 0.0067

Modelling: Random Forest



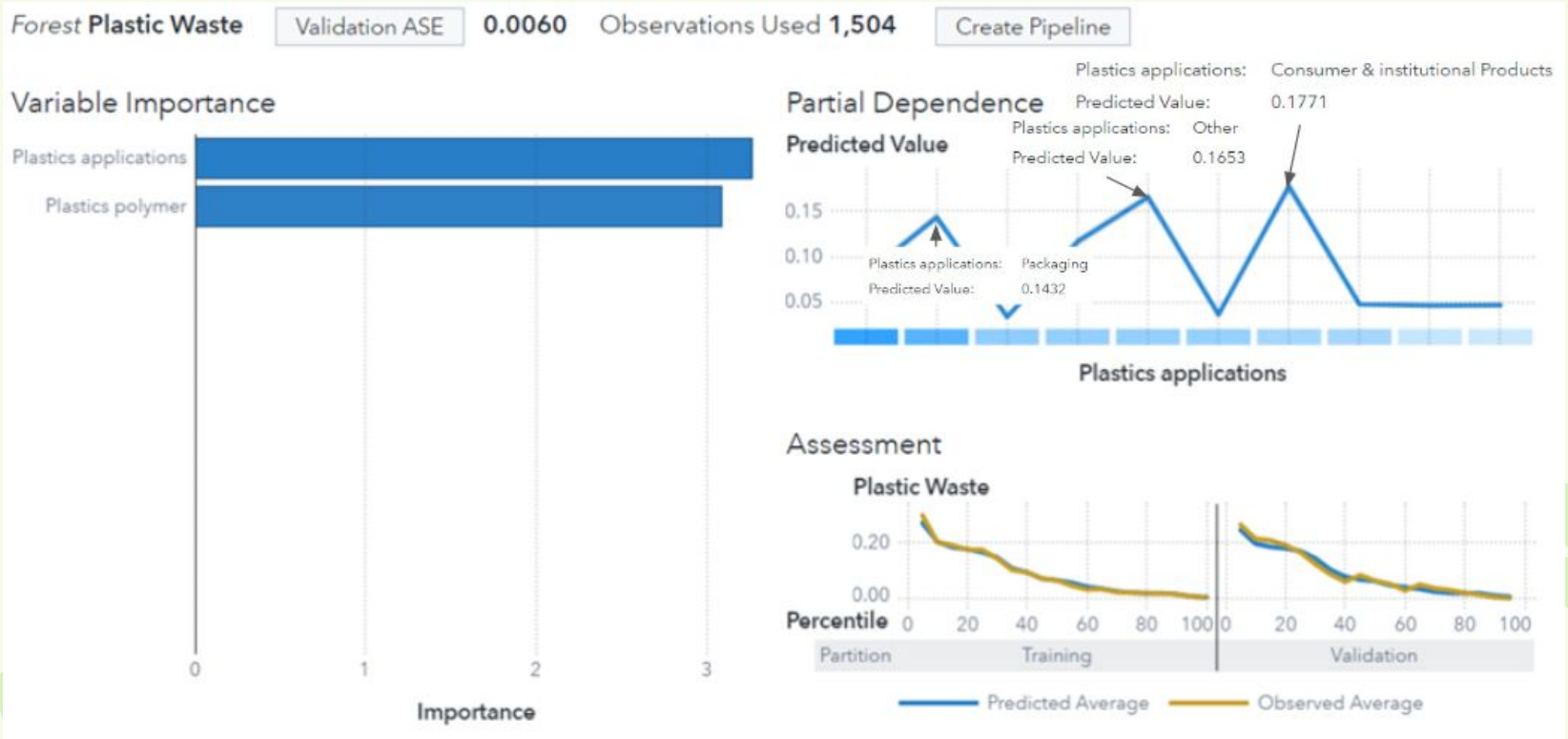
	Parameters Tuned	ASE
Default	Number of trees: 100 , Bootstrap: 0.6, Min value: 1, Max branches: 2, Max levels: 6, Leaf size: 5	Validation: 0.0069 Training: 0.0066
Tuned 1	Number of trees: 150, Bootstrap: 0.8, Min value: 1, Max branches: 2, Max levels: 6, Leaf size: 5	Validation: 0.0063 Training: 0.0059
Tuned 2	Number of trees: 100 , Bootstrap: 0.8, Min value: 1, Max branches: 2, Max levels: 7, Leaf size: 1	Validation: 0.0060 Training: 0.0057
Tuned 3	Number of trees: 100 , Bootstrap: 0.9, Min value: 1, Max branches: 2, Max levels: 12, Leaf size: 1	Validation: 0.0062 Training: 0.0059

Evaluation

As my linear regression model didn't meet my success criteria, I only put my decision tree and random forest models into model comparison. The best model is my random forest model, as it has the lowest validation ASE value compared to the other model.

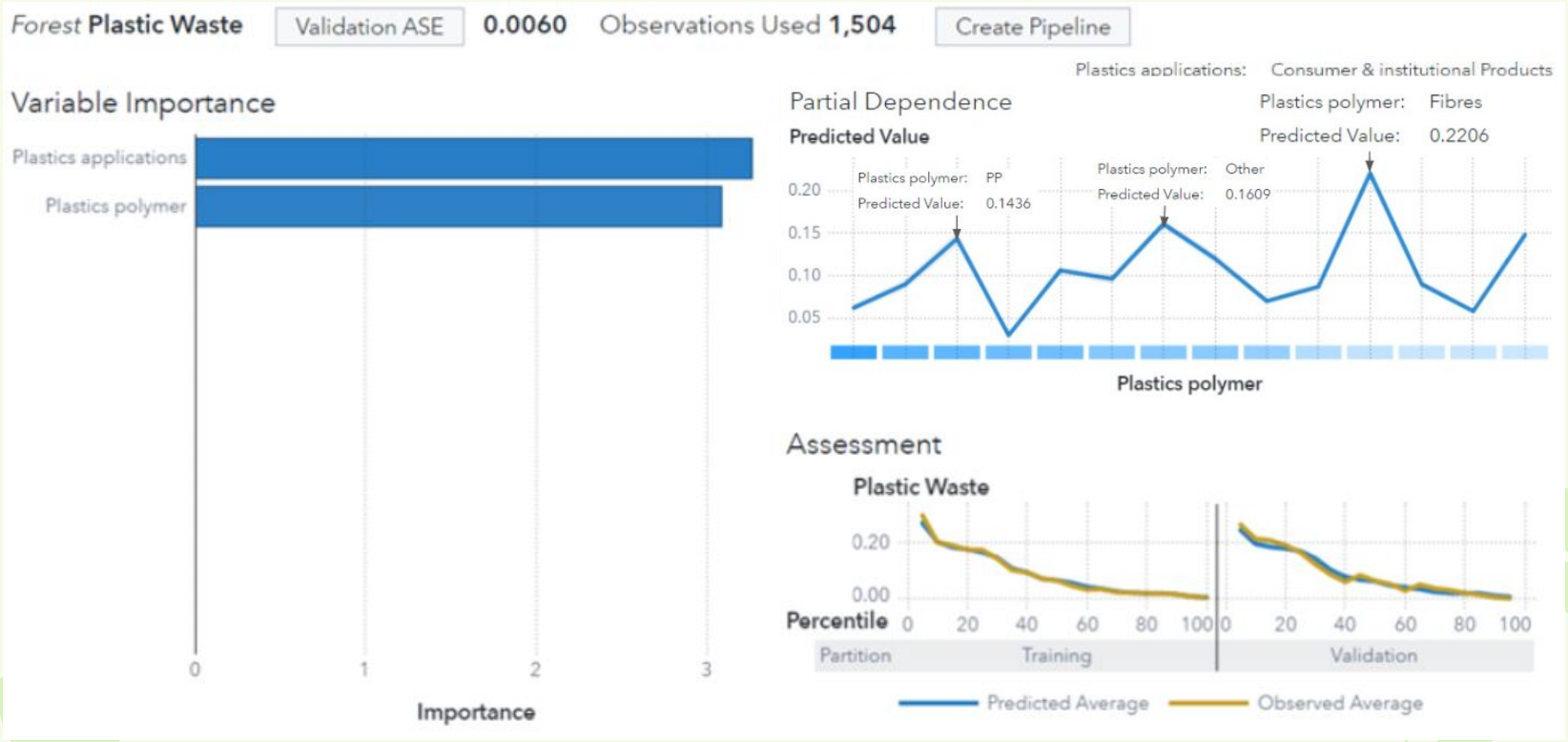
Best Model: Random Forest

Plastic applications partial dependence

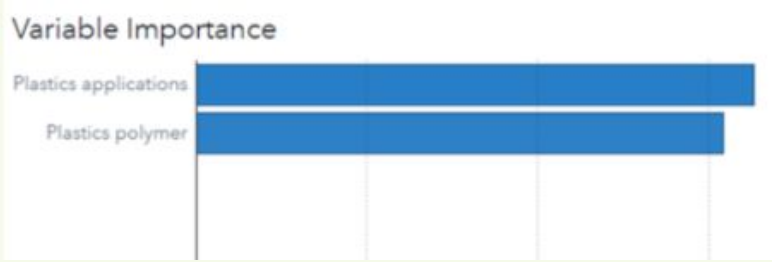


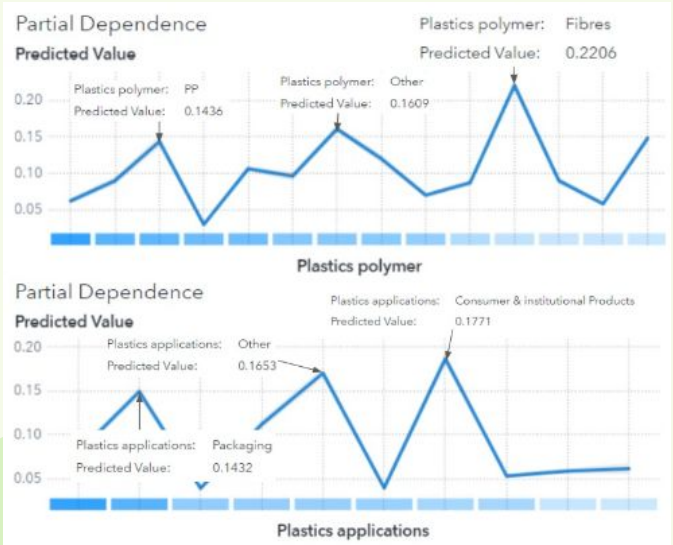
Best Model: Random Forest

Plastic polymer partial dependence



Best Model: Random Forest



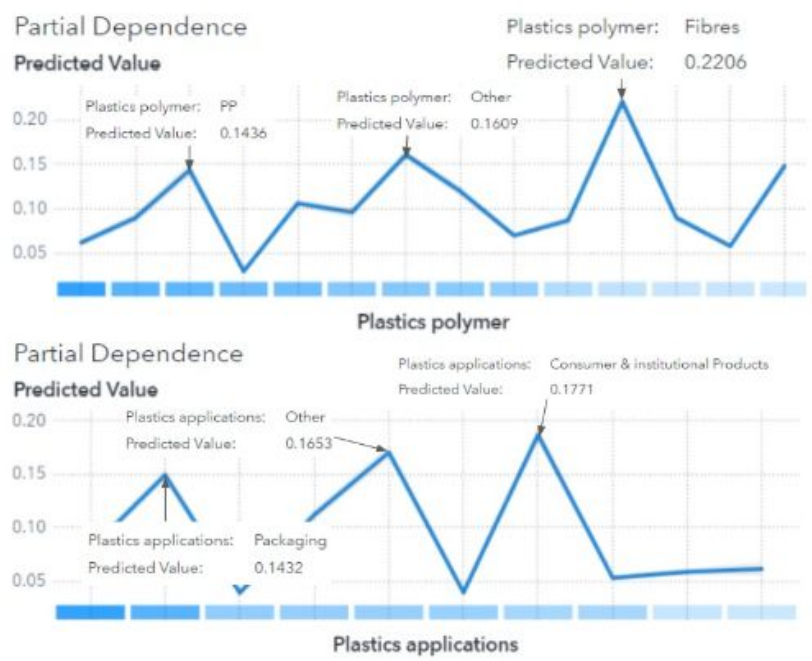


From the variable importance chart, we can see that both my plastic application and polymer columns are important factors. And taking a closer look into the different applications and polymers, the top waste contributor categories with the highest predicted value are packaging (application), other (application), consumer & institutional products (application), fiber (polymer), PP (polymer), other (polymer).

03

Conclusion & Recommendations

Conclusion



Plastic

The top 4 contributors of plastic waste are packaging (application), other (application), consumer & institutional products(application), fiber (polymer), PP (polymer), other (polymer).

With the top contributors I have identified, By focusing efforts on the contributors responsible for the highest waste production, significant strides can be made in curbing plastic waste. We can start taking proactive measures to combat plastic wastage before numbers escalate to unmanageable levels

RESOURCES

Data:

- https://stats.oecd.org/Index.aspx?DataSetCode=PLASTIC_WASTE_7

Tools:

- Tableau
- PowerBi
- Rapid Miner
- SAS

THANKS

Do you have any questions?