International Journal of Recent Innovations in Academic Research
E-ISSN: 2635-3040; P-ISSN: 2659-1561
Homepage: https://www.ijriar.com/
Volume-9, Issue-4, October-December-2025: 255-272

Research Article

Machine Learning Based Feature Importance Regarding Recycling Rates for Municipal Waste in European Countries

[a]Aaron Gun-Hee Cha and *[b]Bob Nam

[a]Seoul International School, South Korea; [b]Seoul Innovations Research Institute, South Korea
*Corresponding Author Email: bobyep25@gmail.com

Received: October 30, 2025 **Accepted:** November 20, 2025 **Published:** November 27, 2025

Abstract
Gradient Boost Regression reveals that real GDP is the variable that has the highest feature importance, followed by education level, total employment and life expectancy in feature importance ranking. Simultaneously, in the context of feature importance, Slovakia and Germany play a special role in predicting municipal waste recycling rates due to their contrasting positions. Maintaining negative correlations, these two countries account for a significant portion of the first principal component that governs the variation of variables including recycling rates. Moreover, the groupings of European countries in the context of PCA biplot or feature importance ranking can be compared with those by recycling rates only or by Ward's hierarchical cluster analysis used in other studies. Partial Dependence Plots (PDP) more clarify the association of each variable with the recycling rates. Country specific dummy variables enhance model performance measured by RMSE or R^2, which can be useful in facilitating integrated and well-coordinated policy design, not only at EU level but also at each national level.
Keywords: Recycling Rates, Gradient Boost Regression, Feature Importance, Partial Dependence Plot, European Countries, PCA, Multiple Linear Regression.

Introduction
Waste Framework Directive (Directive (EU) 2018/851) mandates a 55% recycling rate for municipal waste, which is binding for all EU member states and must be achieved by 2025. More ambitious goals are: A 60% municipal waste recycling target by 2030 and a 65% municipal waste recycling target by 2035. These regulations are part of the EU's broader efforts under the Circular Economy Action Plan aimed at reducing the environmental impact of waste and fostering a more circular economy. Benefits of meeting EU recycling targets are broad from environmental protection and climate action, resource independence and job creation to improved quality of life. Simultaneously, the penalties of not meeting targets range from infringement procedures to financial disincentives.

Despite these clear policy directives, substantial recycling performance gap persists among EU member states, ranging from 13.3% to 66.7% in 2022, which in turn invokes better understanding for broad socio-economic conditions enabling efficient recycling system. In this study, economic factors such as real GDP, total employment, employment in ICT along with socio-economic factors such as education level and life expectancy are integrated as a set of features to broaden the association structures with recycling rates of member states while special attention is paid to the role of country wise heterogeneity.

Literature Review
A literature review of the determining factors for recycling rates in European countries highlights the multidimensional influences on recycling outcomes, especially within the context of differing economic, institutional, demographic, and policy environments.

Key Economic Determinants
Economic prosperity, commonly measured as real GDP per capita, consistently emerges as a strong positive factor for higher recycling rates [1-3]. Economically wealthy countries tend to have more resources to invest in recycling infrastructure and greater public engagement. A strong positive association between GDP per capita and recycling performance in the EU is reported, a finding echoed across 25 European countries and

across 27 countries from 2000 to 2019 [4-6]. This points to economic wealth as a key factor in recycling and circular performance [3]. Private investment in circular economy sectors also promotes recycling performance [1]. However, higher incomes might lead to greater waste generation, making the relationship complex [1].

Institutional and Policy Drivers
Effective environmental taxation, government spending on environmental protection are significant factors [1,2]. Environmental taxes and targeted R&D funding can foster innovation and efficiency in recycling systems. Several empirical studies have consistently shown that environmental taxation correlates positively with recycling outcomes in the EU [3,4]. Based on this evidence, it is highlighted that such taxes not only encouraged cleaner production but also supported investments in recycling infrastructure [7].

Moreover, government expenditures can directly fund the development of robust recycling infrastructure, including advanced material recovery facilities and efficient collection systems, while also supporting public awareness campaigns that foster pro environmental behaviors [8]. Their efficacy is often amplified when combined with external incentives, such as monetary rewards or social influence, which are well-established predictors of recycling participation [9,10]. However, some research finds that public expenditure does not always translate into better recycling, partly due to inefficiencies or time lag before the benefits of R&D are realized [1].

Socio-Demographic and Educational Factors
Urban population size and population density demonstrate complex associations with recycling rates. High urbanization may negatively impact recycling due to structural and logistical pressures in densely populated areas [1, 2].

Higher levels of educational attainment in the population correlate positively with recycling activity, indicating that more informed citizens are more likely to engage in sustainable behaviors [2,4]. Studies indicate that while urbanization and population growth contribute to higher waste generation, effective community participation in waste management, including recycling, can mitigate these effects [11].

This perspective is further supported by observations that public participation is critical for the success of waste recovery activities, highlighting the dual challenge of motivating participation while sustaining involvement [12]. In the European context, population density is revealed to significantly affect municipal waste management efficiency, while population growth is reported as a key driver of rising recycling volumes in China [13,14].

Sectoral and Technological Influences
The sectoral structure of the economy-such as proportion of agriculture, industry, or services also shapes recycling outcomes. Increased agricultural intensification is linked to lower recycling rates in some contexts, while other studies find positive effects depending on the specific waste stream [2,15]. Also, technological capacity and investments in advanced recycling methods are considered pivotal for high recycling rates [16].

Additional Insights
Cluster analyses show advanced recycling countries such as Germany, France, Italy, and Spain combine high recycling rates with favorable economic and institutional conditions [1]. On the other hand, countries lagging often share characteristics such as lower GDP, limited private investment, and weaker institutional frameworks [1].

Therefore, macro-level factors such as economic strength, investment in R&D and environmental sectors, educational attainment, and political commitment play crucial roles in recycling performance across European countries [3]. The complex interplay of these factors underscores the need for integrative and well-coordinated policy design at both national and EU levels.

Data
Most data sets are from Eurostat (2025). Dependent variable is recycling rate of municipal waste from 2015 to 2022, where some values of 2023 and 2024 are added by other sources such as European Environment Agency (EEA). Figure 1 summarizes dependent variable after classifying the values into three groups (A: above 50%, B: 30%~50%, C: below 30%). Approximately half of recycling rates of countries from 2015 to 2024 belong to Group B.

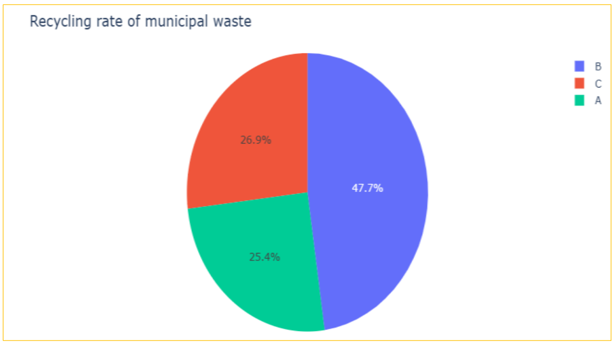

Figure 1. Recycling rate of municipal waste from 2015 to 2024: A: above 50%, B: 30%~50%, C: below 30%.

Summary statistics of features related with recycling rate and thus selected in this study are given in Table 1. Except for real GDP per capita, other features are somewhat indirectly associated with the dependent variable, some of which might be affected by dependent variable. ICT employment is included as it is possible that technological innovation and digitalization can be a driving force in the waste management sectors [17]. Tertiary education can be positively correlated with the recycling rate and thus selected as an explanatory variable [18]. Citizens of countries with higher life expectancy can be more concerned with environmental issues and thus life expectancy is added [19]. Higher recycling rates can reduce domestic net greenhouse gas emissions, suggesting the possibility of negative association [20]. Although straightforward relation is vague, output of the agricultural industry can indicate economic development level, which in turn is closely related with recycling rate of European countries [21]. In 2020, about 1.8 million people were employed in municipal waste recycling and related activities [22], suggesting that total employment and recycling rate can be positively associated. Nominal labor productivity can be a proxy for economic development and technological capacity and thus can be positively associated with the recycling rate [1]. Many other features that are already assessed as significant are not included in this study, while the features that are vague in their link with recycling rate are intentionally added to broaden the scope of possible explanatory variables.

Table 1. Summary statistics of explanatory variables.

	Year	Life expectancy	Employed information and communications technology (ICT) specialists	Population in private households by educational attainment level	Real GDP per capita	Domestic net greenhouse gas emissions (tonnes per capita)	Output of the agricultural industry - basic and producer prices	Employment, domestic concept - Total	Nominal labour productivity per person employed
count	264.000000	264.000000	264.000000	264.000000	264.000000	264.000000	264.000000	264.000000	264.000000
mean	2018.912879	80.298485	283.699242	9909.894697	35291.477273	8.485227	15382.540076	7274.914356	98.946284
std	2.550625	2.955139	392.319677	13459.351790	22793.019938	5.856737	21409.226039	10100.827541	30.628108
min	2015.000000	71.200000	7.500000	205.200000	8130.000000	-0.200000	121.090000	180.500000	44.700000
25%	2017.000000	78.000000	44.650000	1834.300000	17352.500000	5.400000	2303.592500	1380.957500	76.750000
50%	2019.000000	81.500000	154.450000	5537.700000	27085.000000	7.250000	7309.240000	4295.570000	98.184500
75%	2021.000000	82.400000	269.850000	7293.850000	46357.500000	9.800000	12744.530000	5353.310000	110.550000
max	2024.000000	84.500000	2107.600000	53802.300000	107570.000000	38.900000	97344.810000	45935.000000	224.500000

8 rows × 40 columns

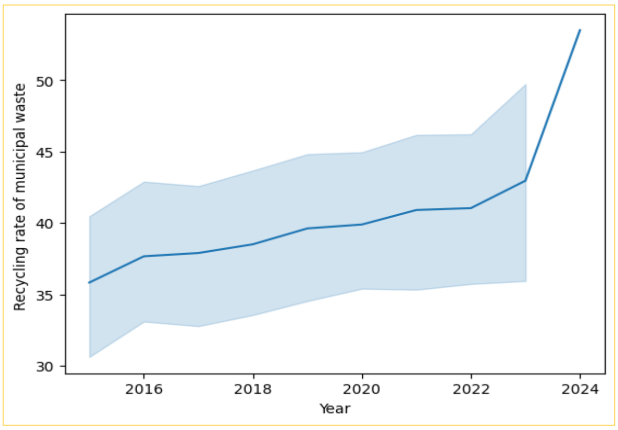

Figure 2. Recycling rate with year.

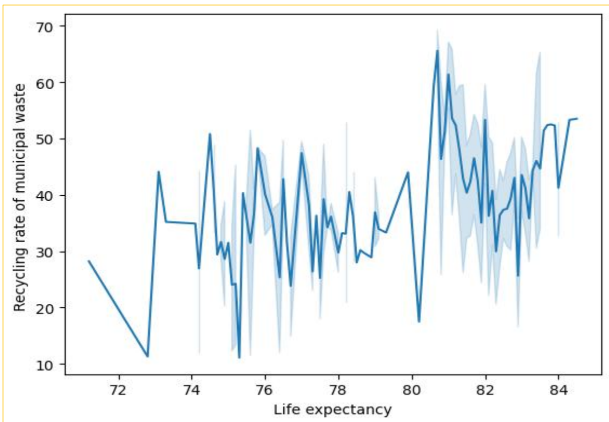

Figure 3. Recycling rate with life expectancy.

Line graph of Figure 2 reveals clearly increasing pattern of recycling rate as time passes. The steep growth after 2023 might be related with a measurement error because the values in this range is substituted from other sources to fill the gap in Eurostat. Figure 3 of line plot shows overall increasing pattern but fluctuation is very high.

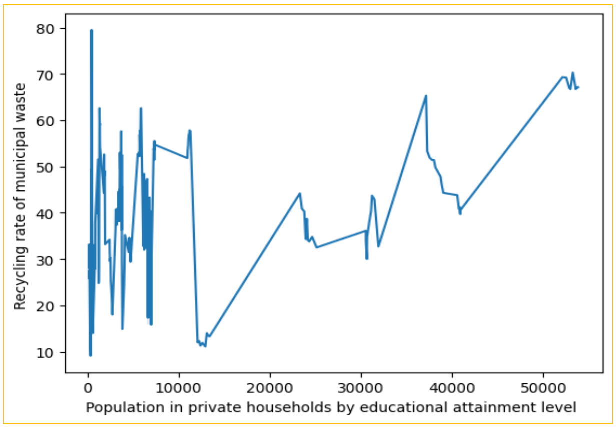

Figure 4. Recycling rate with education level.

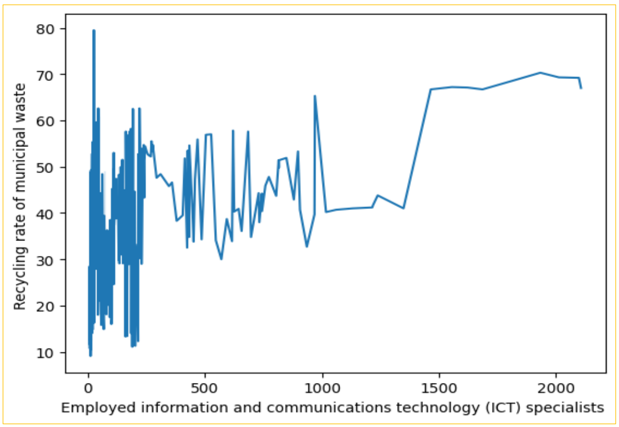

Figure 5. Recycling rate with ICT employment.

Line plot of education levels measured by the population of educational attainment exhibits increasing pattern after some threshold of 12000, whereas the pattern is very unpredictable below this level. In Figure 5, association is more unclear, especially below the threshold level of 1000.

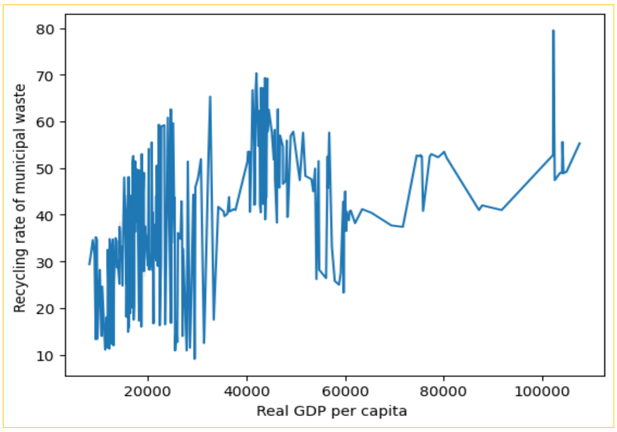

Figure 6. Recycling rate with real GDP.

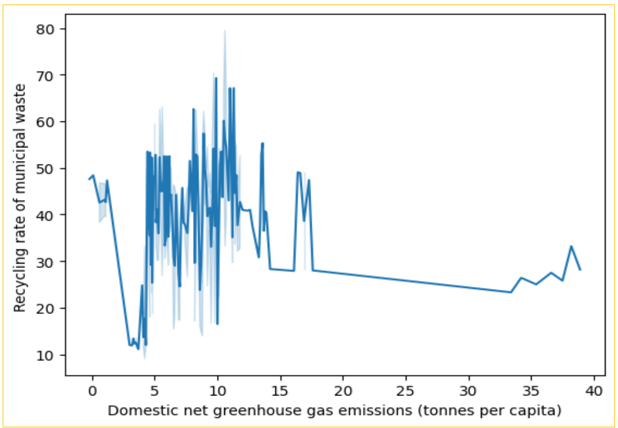

Figure 7. Recycling rate with greenhouse gas emission.

Overall upward sloping in Figure 6 and downward sloping in Figure 7 after some threshold levels can be observed but the pattern is ambiguous before these turning points.

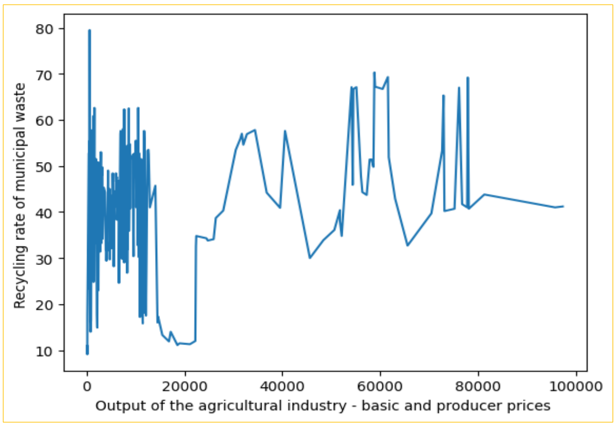

Figure 8. Recycling rate with agriculture output.

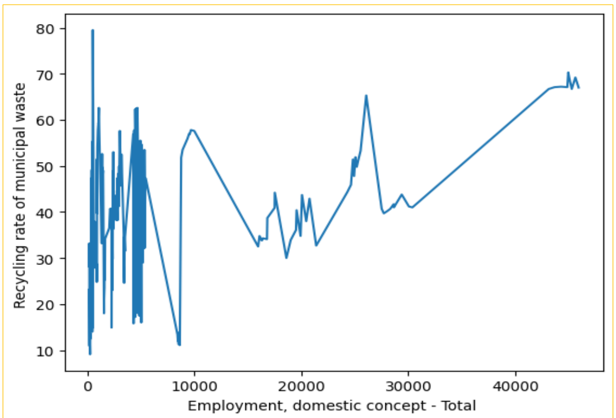

Figure 9. Recycling rate with total employment.

Similar characteristics such as high volatility below certain level and slight increase after that level are repeated in Figure 8, 9 which can in turn contribute to the formation of country clusters or impair the significancy of these variables. In case of nominal labor productivity, the relationship is more complex. Within the range of high volatility, tendency is vague. Beyond this volatile region, both increasing tendency or decreasing tendency is possible given different starting point.

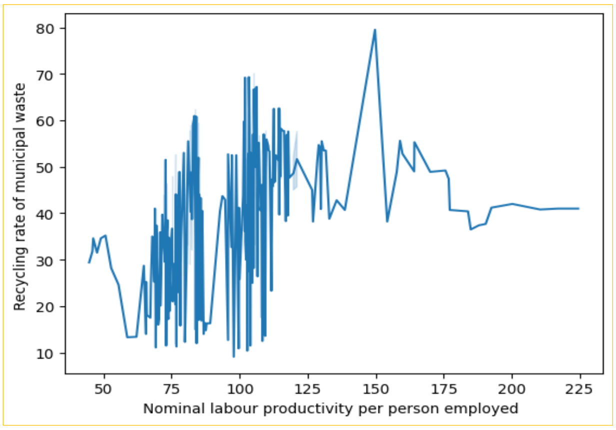

Figure 10. Recycling rate with labor productivity.

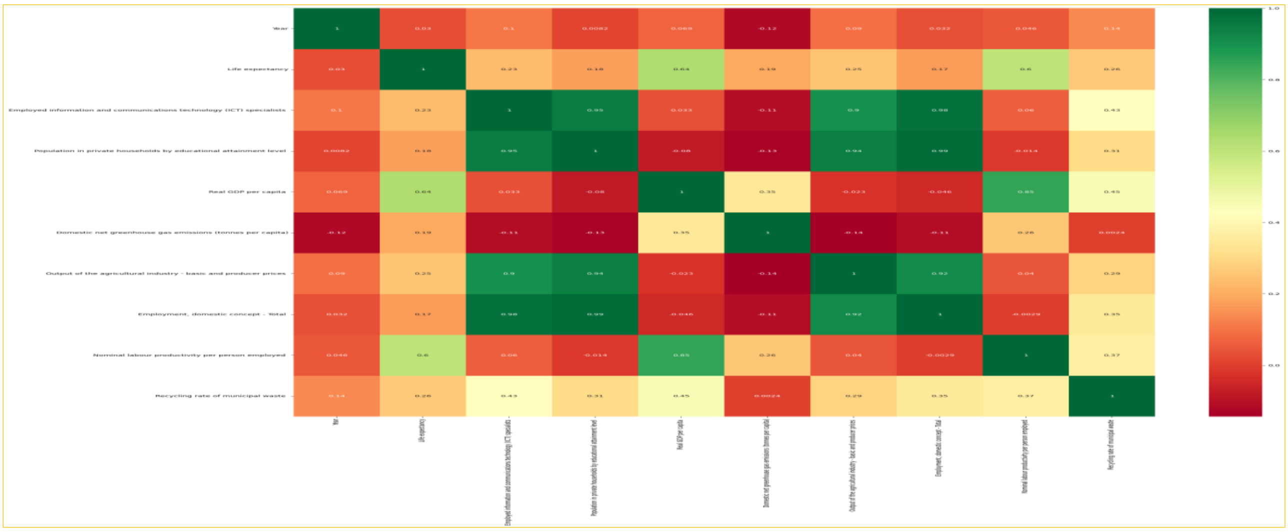

Figure 11. Correlation heatmap of 10 features.

Correlation heat map without country wise dummy variables in Figure 11 shows that relatively higher positive and negative correlation are mingled.

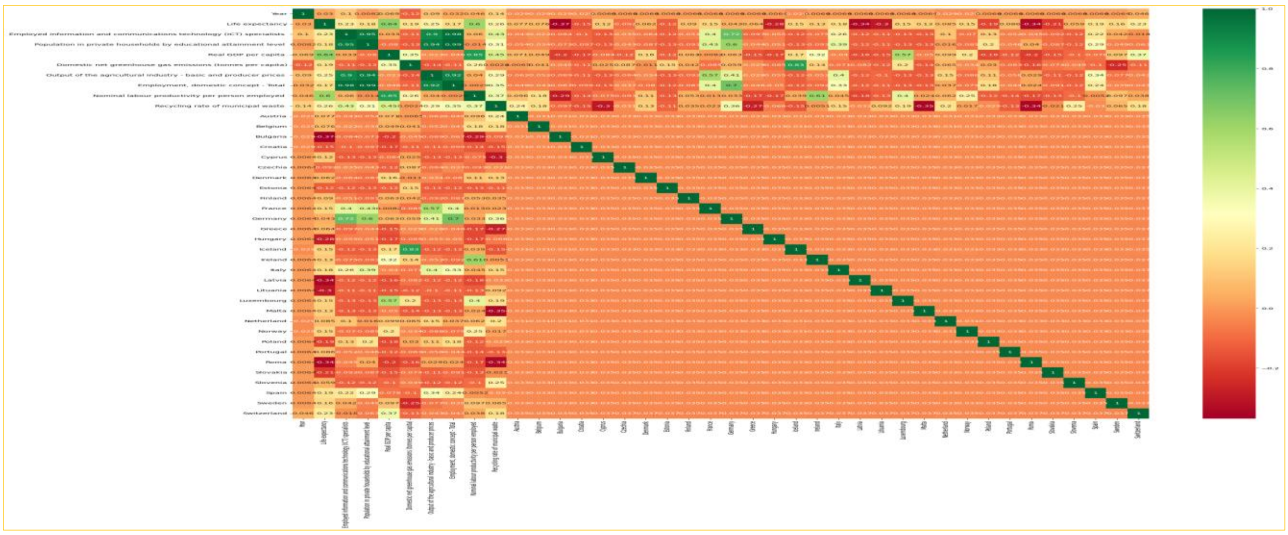

Figure 12. Correlation heatmap of 10 features with country wise dummy variables.

Similar pattern is repeated in Figure 12 with country wise dummy variables. As is expected, the distribution of correlations varies by country.

To more closely look into the interplay, recycling rates are grouped into three categories, integrated in segmented bar chart of each variable from Figure 13 to Figure 30.

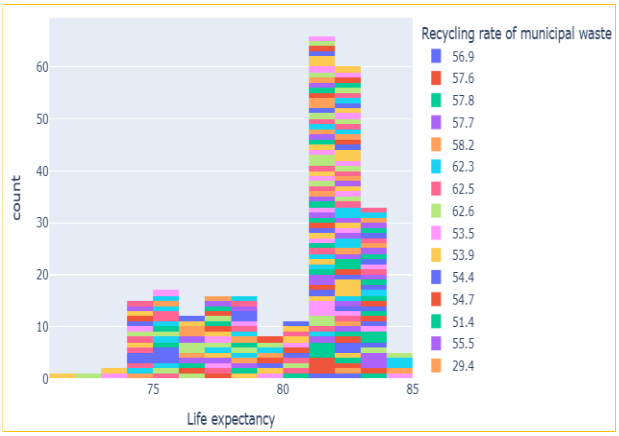

Figure 13. Life expectancy bar chart (i).

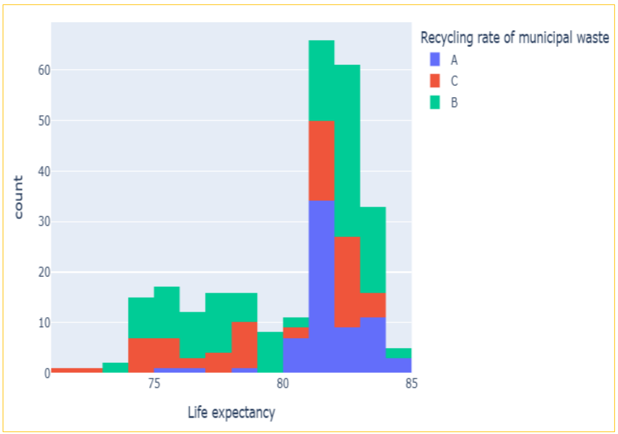

Figure 14. Life expectancy bar chart (ii).

Figure 14 visualizes that the proportion of higher recycling rates increases as life expectancy rises, as is expected since life expectancy is positively associated with economic development level.

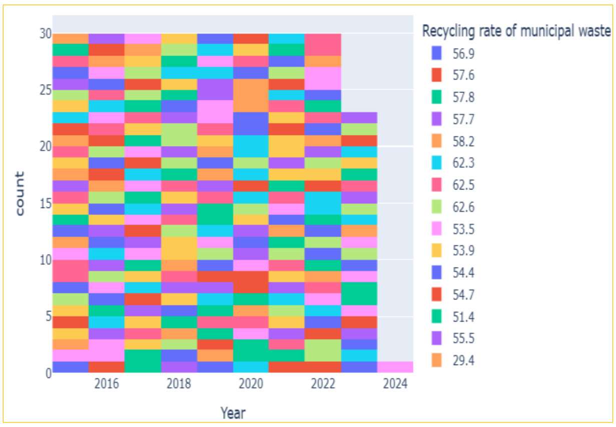

Figure 15. Year bar chart (i).

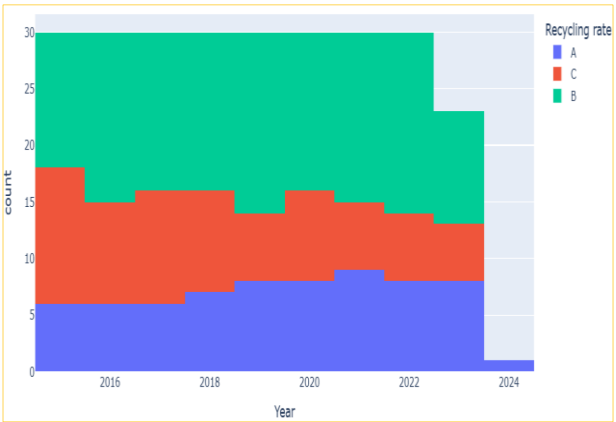

Figure 16. Year bar chart (ii).

The proportion of low recycling rates decreases while that of high rates increases over time. The frequencies of the values of 2023 and 2024 is due to data gap in Eurostat.

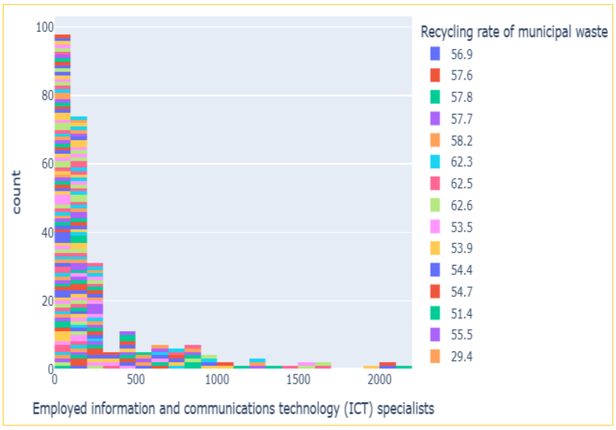

Figure 17. ICT employment bar chart (i).

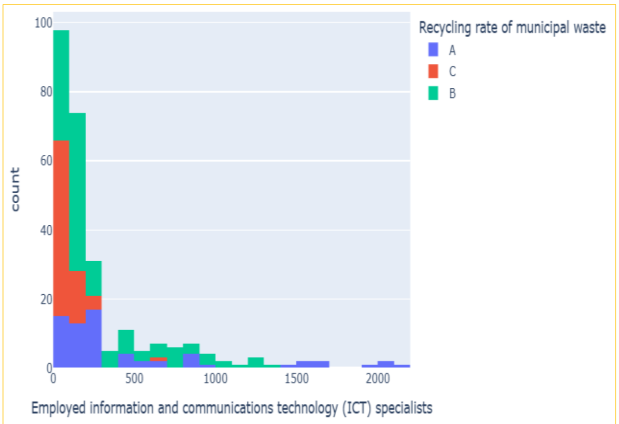

Figure 18. ICT employment bar chart (ii).

Interestingly, the percent of low recycling rates is almost zero when the number of ICT employment is above 300 or the population of educational attainment is above 15k according to Figure 18 and 20.

Figure 19. Education level bar chart (i).

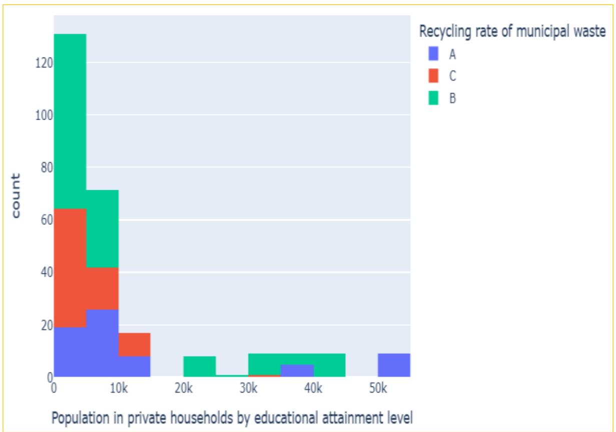

Figure 20. Education level bar chart (ii).

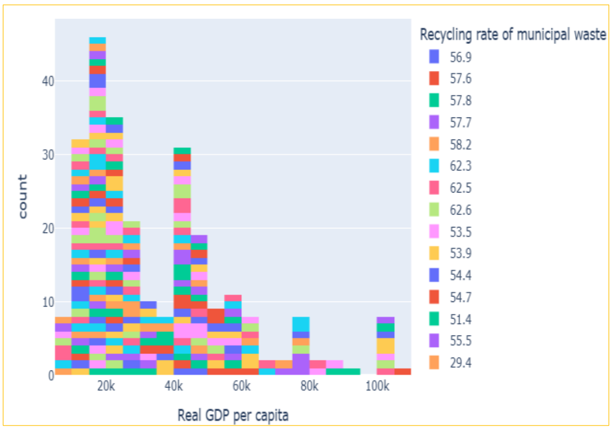

Figure 21. Real GDP bar chart (i).

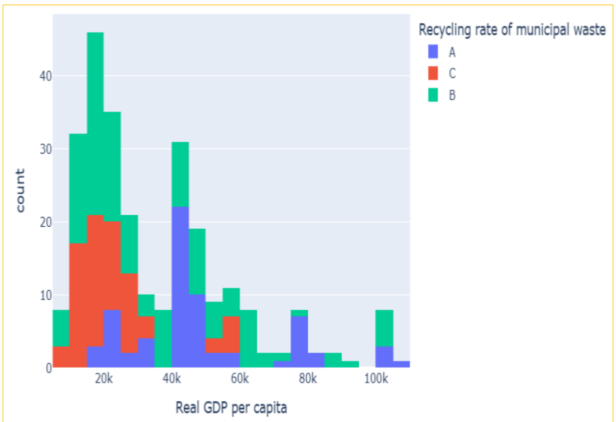

Figure 22. Real GDP bar chart (ii).

Real GDP is a key variable most widely investigated. Figure 21 and 22 shows that as real GDP per capita increases, the portion of high recycling rates increases while that of low rates decreases.

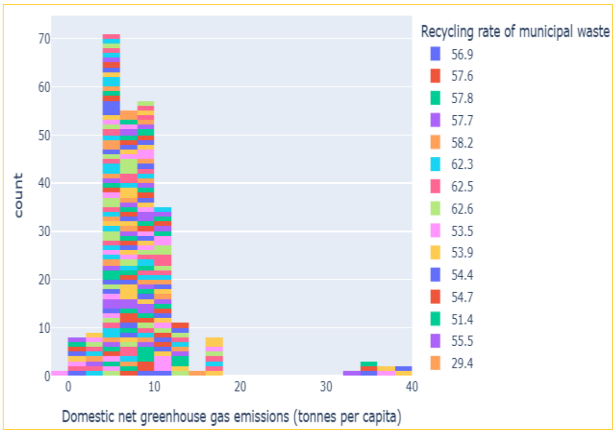

Figure 23. Bar chart of greenhouse gas emission (i).

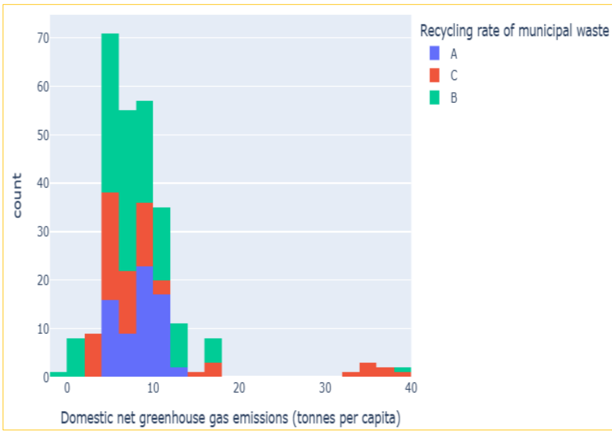

Figure 24. Bar chart of greenhouse gas emission (ii).

The countries with high recycling rates are placed in the range of greenhouse gas emissions from 4 to 13 only, while those with low rates are clustered above 30 tones.

As the output of agricultural industry increases above 25k, the countries with high recycling rates almost disappear as in Figure 25 and 26.

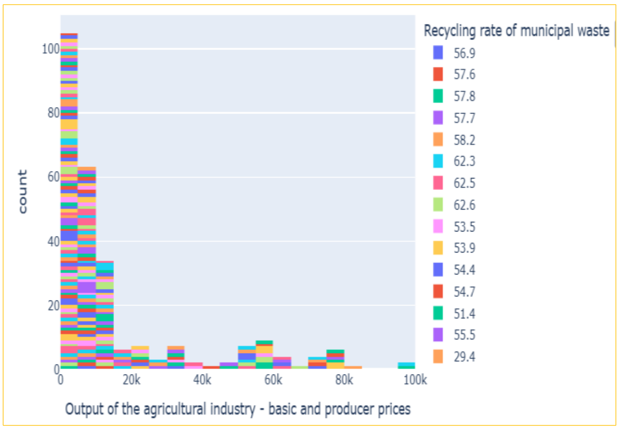

Figure 25. Bar chart of agricultural output (i).

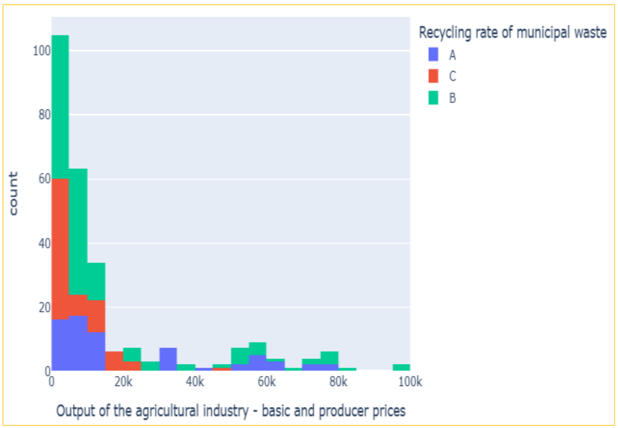

Figure 26. Bar chart of agricultural output (ii).

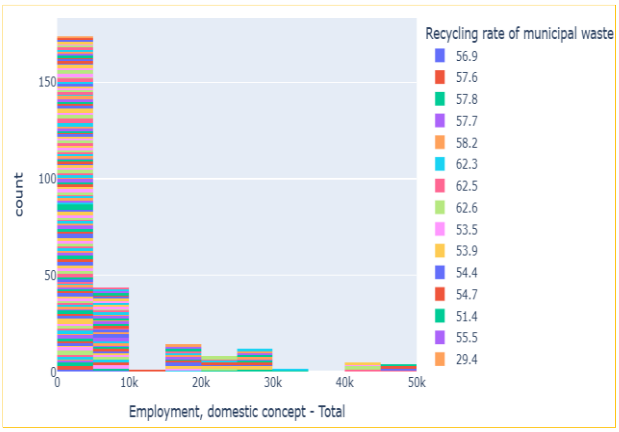

Figure 27. Total employment bar chart (i).

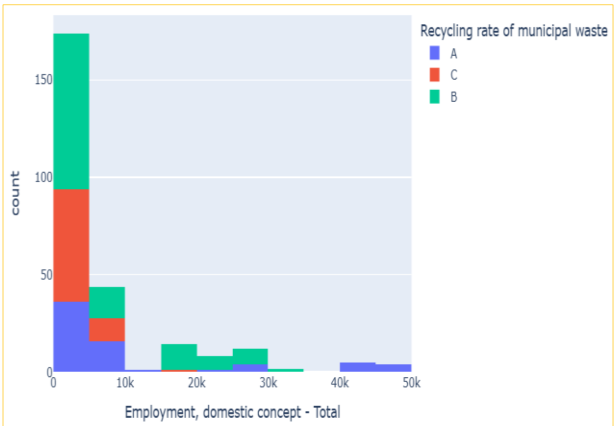

Figure 28. Total employment bar chart (ii).

Likewise, the portion of countries with lower recycling rates becomes almost zero when the total employment increases above the threshold level of 10k.

Figure 29. Labor productivity bar chart (i).

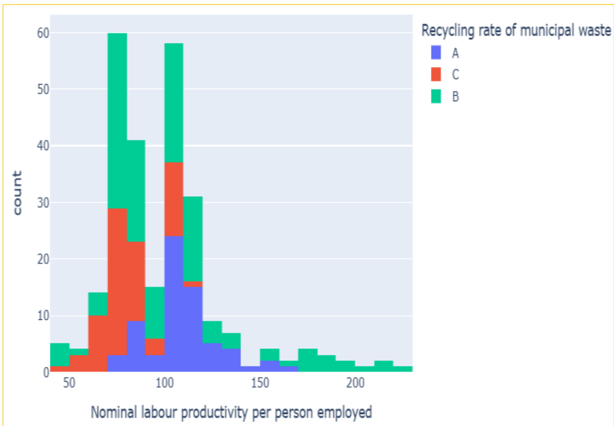

Figure 30. Labor productivity bar chart (ii).

When nominal labor productivity is above 120, the countries with low recycling rates disappear. However, positive correlation between recycling rates and labor productivity is unclear because the Group B dominates above the value of 170.

First Data Inspection: Multiple Linear Regression Model

As the first data inspection, multiple linear regression model without country wise dummy variables is applied. Since adjusted R^2 is only 0.424, some key variables seem to be missing. The sign life expectancy is unexpected, indicating that overall goodness of fit is not satisfactory.

Resulting regression line is given below with the output table and residual plot added in appendix as Table 8, 9 and Figure 44.

Recycling rate = $90.828_{(25.213)}$ - $0.814_{(0.327)}$ *Life_expectancy + $0.043_{(0.009)}$ *Employment_ICT + $3.436 * 10^{-4}{}_{(4.367*10^{-5})}$*RGDP - $0.295_{(0.129)}$*Greenhouse_gas_emissions - $1.723 * 10^{-4}{}_{(8.519*10^{-5})}$*Agricuture_output – $7.719 * 10^{-4}{}_{(3.741*10^{-4})}$*Employment_total

Country wise dummy variables can help to capture the heterogeneity of countries, enhancing model performance measured by adjusted R^2, which is jumped into 0.920 from 0.424, along with the significant drop of RMSE from 11.190 to 4.162. More importantly, the set of significant explanatory variables are remarkably changed. The resulting regression line with country dummy variables is given as:

Recycling rate = $-1548.809_{(205.653)}$ + $0.803_{(0.102)}$ *Year - $0.002_{(2.648*10^{-4})}$ *Education + $1.100 * 10^{-4}{}_{(2.868*10^{-5})}$*RGDP – $0.179_{(0.019)}$*Labor_productivity + $14.281_{(1.726)}$*Austria + $14.709_{(1.975)}$*Belgium - $24.909_{2.015}$ *Bulgaria - $28.869_{(1.961)}$ *Croatia - $43.041_{(1.944)}$ *Cyprus + $64.578_{(9.912)}$ *France - $9.765_{(1.901)}$ *Czechia - $29.423_{(1.973)}$ *Estonia - $9.619_{(1.565)}$ *Finland + $116.115_{(13.153)}$ *Germany - $30.551_{(1.935)}$ *Greece- $15.107_{(1.919)}$ *Hungary - $32.703_{(1.874)}$ *Iceland + $70.722_{(9.225)}$ *Italy - $23.155_{(2.004)}$ *Latvia - $10.716_{(1.901)}$ *Lithuania - $45.111_{(1.917)}$ *Malta + $19.723_{(2.525)}$ *Netherland- $6.244_{(1.602)}$ *Norway + $25.215_{(5.593)}$ *Poland - $18.841_{(1.889)}$ *Portugal - $23.430_{(2.933)}$ *Roma- $16.049_{(1.806)}$*Slovakia + $41.201_{(7.376)}$*Spain

The signs of coefficients of variables such as education and labor productivity are still unpredicted, which comes from the gap between groupings of recycling rates as in the segmented bar chart and raw values in one route or from possible nonlinearity in the other route. Still possible one more aspect is that there can exist time lags in case of the education and nominal labor productivity.

The population size of educational attainment does not directly indicate same year community awareness of environmental protections. Moreover, the speculation that nominal labor productivity is related with the same year technological capacity for recycling is not substantiated.

In case of nominal labor productivity, this ambiguous relation is already observed in line plot and segmented bar chart. Therefore, further investigations are required to more accurately capture the relations.

Second Data Investigation: Factor Analysis

Below Figure 31 is a biplot from principal component analysis (PCA). The clustering of features such as ICT employment, agricultural output, total employment and education is positively correlated with the recycling rate, which forms the main part of the first principal component.

Real GDP, labor productivity and life expectancy also have positive correlation with recycling rates, while greenhouse gas emissions are almost uncorrelated with recycling rates, all of which make up the second principal component.

The relative contributions to the first component by ICT employment, agricultural output, total employment and education are high and similar (0.93~0.97), contrasting to the smallest effect of year (0.05~0.1) in terms of the overall variation of variables. Recycling rate is linked to both of two components. Real GDP forms the main part of the second principal component. Component loadings of these 10 variables are given in Table 2.

Table 2. Component loadings of PCA.

Variable	Principal component 1	Principal component 2
Year	0.1074149	0.05391665
Life expectancy	0.36019388	0.70097293
Employment_ICT	0.97212441	-0.13229856
Education	0.95603946	-0.23925003
RGDP	0.15398861	0.93816334
Greenhouse gas emission	-0.10073901	0.46330619
Agricultural output	0.93850353	-0.17619556
Employment total	0.96341804	-0.21425861
Labor productivity	0.18991947	0.88190638
Recycling rate	0.5011269	0.41019944

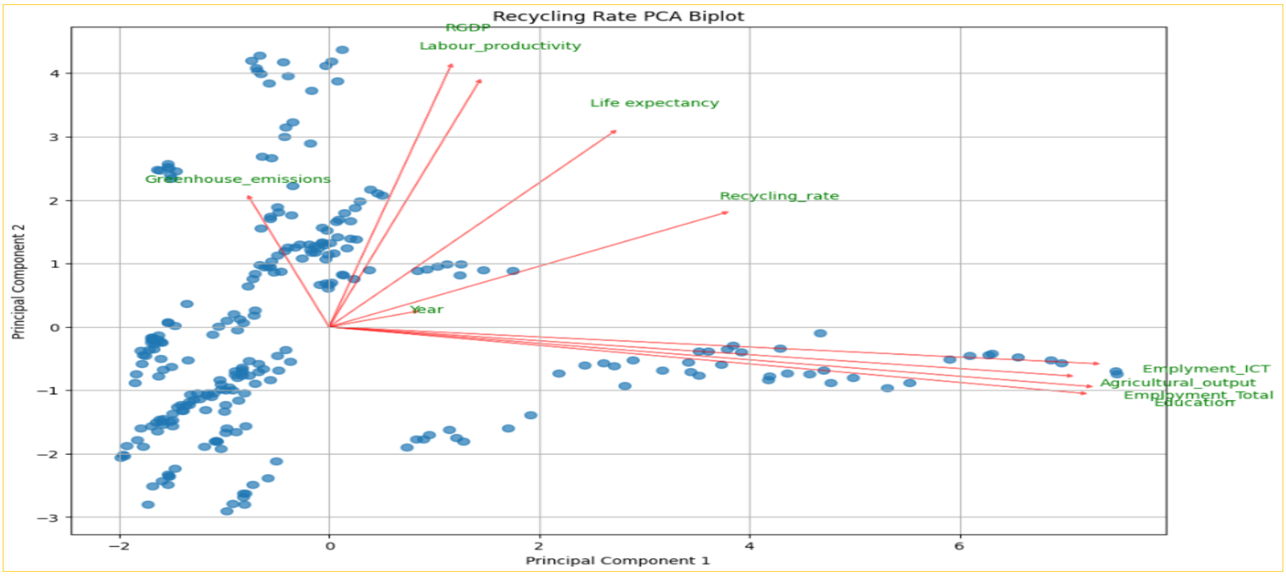

Figure 31. Recycling rate PCA biplot without country wise dummy variables.

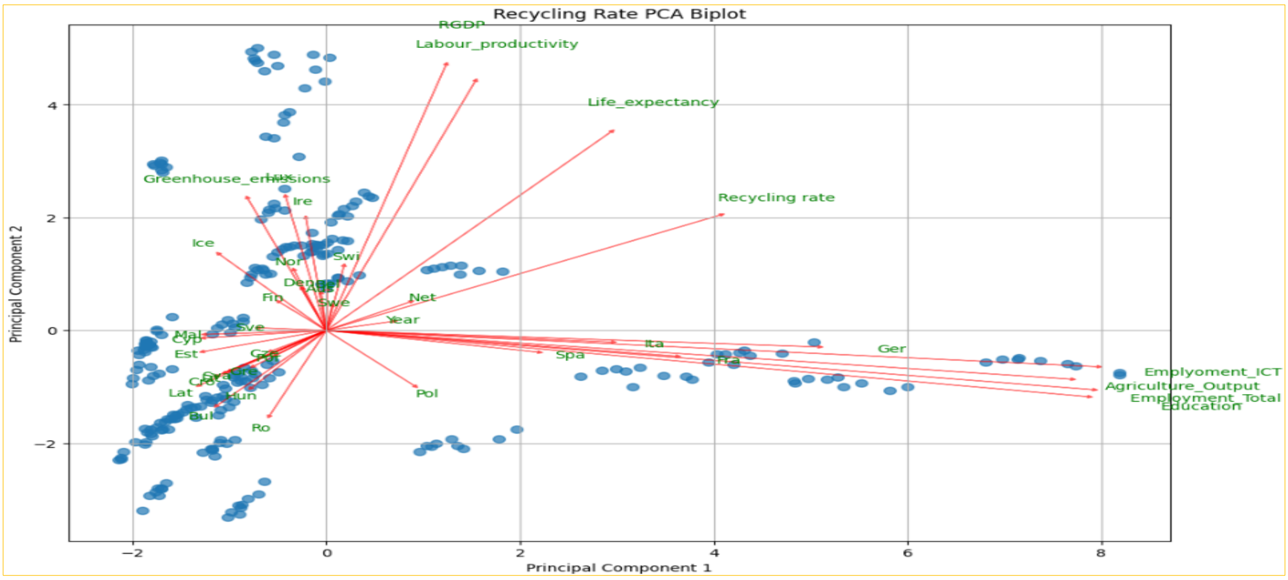

Figure 32. Recycling rate PCA biplot with country wise dummy variables.

The relation of variables is almost unchanged when country wise dummy variables are added as in Figure 32. Interesting observation is that countries are classified into several groups along with 10 features. Groupings of European countries based on this PCA analysis is summarized in the following table.

In the context of the variability of 10 features, countries in Group1 are slightly negatively correlated with both the countries in Group 2 and Group 3. The relation between Group 2 and Group 3 are rather ambiguous, ranging from positive to negative correlations in this context.

Table 3. Groupings of European countries based on clustering on PCA biplot.

Group 1	Italy, France, Spain, Germany
Group 2	Poland, Slovenia, Estonia, Czechia, Greece, Malta, Cyprus, Estonia, Portugal, Slovakia, Croatia, Lavia, Hungary, Bulgaria, Roma, Luxembourg, Lithuania
Group 3	Iceland, Norway, Denmark, Ireland, Netherland, Switzerland, Finland, Sweden, Austria

The countries in Group 3 are more positively correlated with greenhouse gas emissions, real GDP, and life expectancy, while these countries are almost uncorrelated with recycling rate. Simultaneously, the countries in Group 2 are negatively correlated with the features such as greenhouse gas emissions, real GDP, and life expectancy.

To compare groupings by PCA and those by 2022 recycling rate values, Table 4 is added.

Table 4. Groupings of European countries based on 2022 recycling rates.

Group 1 (above 50%)	Germany, Slovenia, Belgium, Netherland, Switzerland, Austria
Group 2 (30%~50%)	Italy, France, Spain, Poland, Hungary, Luxembourg, Lithuania, Norway, Denmark, Ireland, Finland, Sweden
Group 3 (~30%)	Estonia, Czechia, Greece, Malta, Cyprus, Portugal, Slovakia, Croatia, Latvia, Bulgaria, Roma, Iceland

Final Data Inspection: Gradient Boost Regression

Gradient regression usually refers to using the gradient descent algorithm to optimize a regression model, often linear regression, by adjusting model parameters step by step to minimize errors between predictors and actual data [23]. It works by computing the gradient of a loss function-which quantifies prediction error-relative to the model's parameters, then moving these parameters in the direction that reduces error most efficiently, like rolling downhill toward the lowest point of a valley [23].

For linear regression, the loss function commonly used is mean squared error, and the updates to parameter values look like: $\theta_{new} = \theta_{old} - \eta \frac{\partial}{\partial \theta} Loss$ where θ is the parameter η is the learning rate, and the partial derivative is the gradient [13-23]. Gradient descent avoids heavy matrix operations and scales better with high-dimensional data, making it computationally efficient and feasible for big data scenarios [23]. Gradient descent bypasses matrix inversion issues since it updates parameters incrementally without needing matrix inversion. However, interpretability is limited in gradient descent.

To assess gradient boost regression performance, several measures are calculated:

RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$

R^2= 1 - $\frac{\sum_{i=1}^{n}(y_i-\hat{y_i})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

MAE score = $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$

When the measures in Table 5 and Table 6 are compared, noticeable gains of model performance are identified, especially in R^2 improvement from 0.87 to 0.99, which is also demonstrated in two gradient boost regression lines of Figure 33 and 34, where magnitudes of residuals are remarkably reduced when dummy variables are added.

Table 5. Gradient boost regression outcome without country wise dummy variables.

```
MODEL GRADIENT BOOST REGRESSION
RMSE score for test: 5.2472098959446285
RMSE score for train: 0.0004009758973836163
R2 score: 0.8709922597433987
MAE score: 2.9429727674582344
Variance score test: 0.8709922597433987
Variance score train: 0.9999999992511175
```

Table 6. Gradient boost regression outcome with country wise dummy variables.

```
MODEL GRADIENT BOOST REGRESSION
RMSE score for test: 0.014541433863350817
RMSE score for train: 5.416695823707145e-06
R2 score: 0.9913249928696531
MAE score: 0.0016275283641546637
Variance score test: 0.9913249928696531
Variance score train: 0.9999999992944921
```

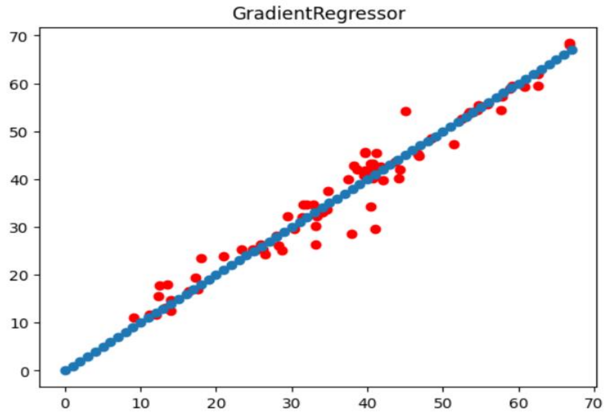

Figure 33. Plot of gradient regression with country wise dummy variables.

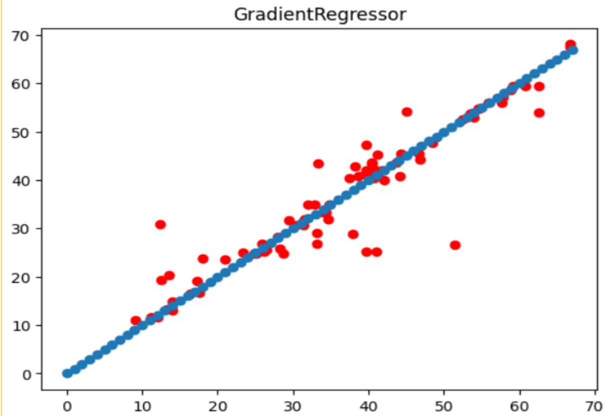

Figure 34. Plot of gradient regression without country wise dummy variables.

Feature importance indicates how valuable each feature is in constructing the boosted decision trees that comprise the model. Feature importance is calculated by summing the amount of loss reduction attributed to splits using a given feature [13-23]:

$$Importance_j = \sum_{m=1}^{M} \sum_{k=1}^{T_m} I\left(v_{m,k} \ = j \ \right) \times \Delta L_{m,k} \left(s_{m,k}\right)$$

Where,

M = The total number of trees
T_m = The number of internal nodes in trees m
I (·) = Indicator function which is 1 if the node k in tree m splits on feature j, and 0 otherwise.
$v_{m,k}$ = Feature used for the split at node k in tree m
$\Delta L_{m,k}$ = Reduction in the loss function such as mean squared error achieved by the split $s_{m,k}$

With these procedures, feature importance is calculated, where features with higher importance scores have more impact on predictions.

Table 7. Feature importance of gradient boost regression.

Feature name	Feature importance	Feature importance ranking	Feature name	Feature importance	Feature importance ranking
RGDP	0.3932	1	Poland	0.0009	21
Education	0.1492	2	Greece	0.0006	22
Employment total	0.1192	3	Ireland	0.0005	23
Slovakia	0.0672	4	Croatia	0.0004	24
Germany	0.0460	5	Denmark	0.0004	25
Life expectancy	0.0361	6	Cyprus	0.0001	26
Portugal	0.0294	7	Hungary	0.0001	27
Greenhouse gas emissions	0.0238	8	Bulgaria	0.0000	28
France	0.0231	9	Czechia	0.0000	29
Latvia	0.0227	10	Slovenia	0.0000	30
Employment ICT	0.0213	11	Netherland	0.0000	31
Year	0.0202	12	Austria	0.0000	32
Roma	0.0126	13	Finland	0.0000	33
Agriculture output	0.0080	14	Malta	0.0000	34
Labor productivity	0.0075	15	Switzerland	0.0000	35
Italia	0.0060	16	Belgium	0.0000	36
Spain	0.0043	17	Estonia	0.0000	37
Labor productivity	0.0043	18	Iceland	0.0000	38
Lithuania	0.0020	19	Norway	0.0000	39
Luxembourg	0.0009	20			

Among the 9 variables, real GDP has the highest feature importance, followed by education, total employment, life expectancy, greenhouse gas emissions and so on. Interesting observation is that Slovakia and Germany play a special role in predicting municipal waste recycling rates in European countries. These two countries with the negatively correlated dummy variables are placed in the counterpart in terms of the first principal component in PCA biplot, thereby governing the important variations of all variables, playing a special role in predicting municipal waste recycling rates. Germany represents a leading group with high recycling rates, while Slovakia exhibits rapid recent improvement and effective policy shifts in enhancing the rates.

For further interpretation, partial dependence plot (PDP) is applied [13-23]. PDP is defined as:

$$\hat{f}_S(x_S) = E_{X_C}[\hat{f}(x_S, X_C)] = \int \hat{f}(x_S, X_C)\,dP(X_C)$$

Where,

x_S = Variable to be plotted as PDP to capture its effect on dependent variable
X_C = Variables to be held fixed except for x_S after learning is completed

By integrating out and thus marginalizing X_C, we can derive the effect of set S on the dependent variable. With these procedures, PDP of features are derived as in Figure 32 to Figure 40. Most of variables reveal complex non-linearity in the interplay with dependent variable. Among these, real GDP demonstrates three regions where rapid growth, slight increase, almost constant after a threshold point of around 33000. That is, at a low level of real GDP, the rise of real GDP moves together with the steep increase, followed by slight increase of recycling rates; however, the marginal effect of increase in real GDP is almost constant after some level. In case of education level, both increasing and decreasing portions coexist, while for the case of nominal labor productivity, most domains are comprised of decreasing or constant portions.

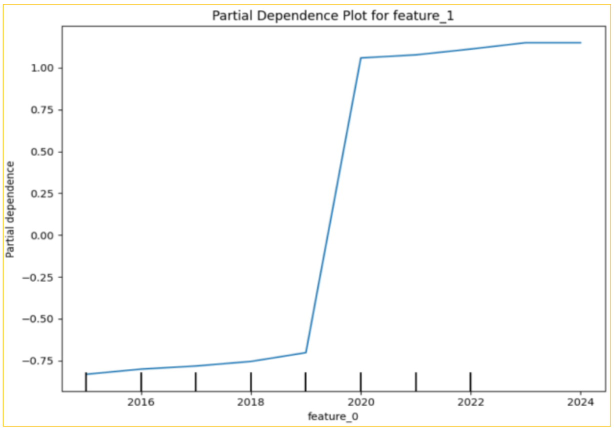

Figure 35. PDP of year.

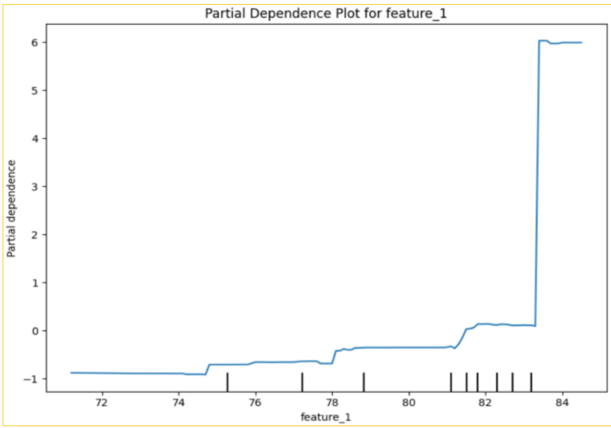

Figure 36. PDP of life expectancy.

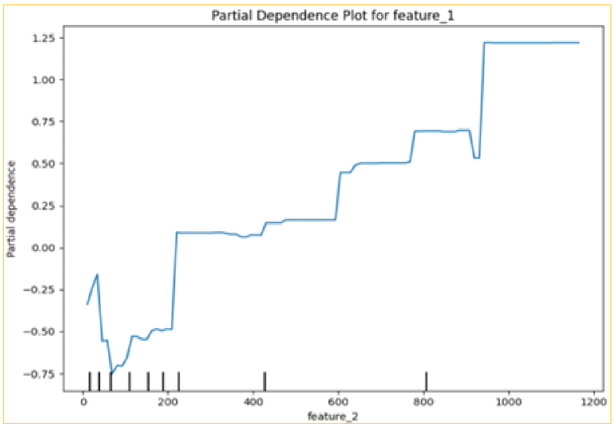

Figure 37. PDP of ICT employment.

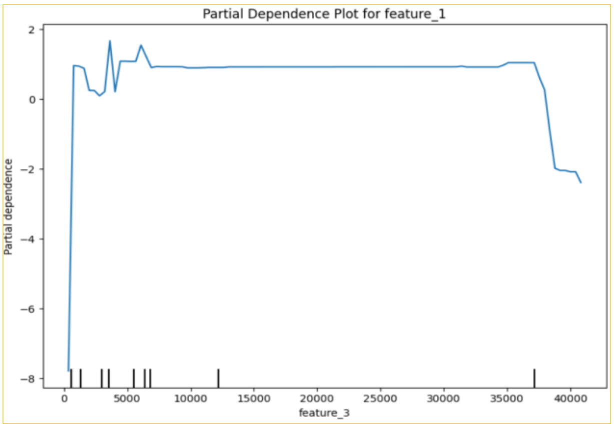

Figure 38. PDP of education level.

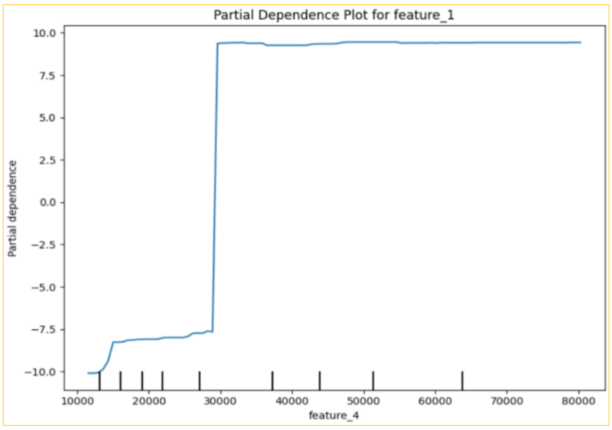

Figure 39. PDP of real GDP.

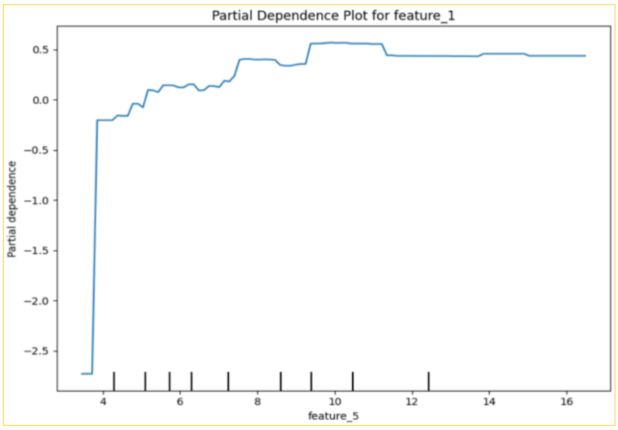

Figure 40. PDP of greenhouse gas emission.

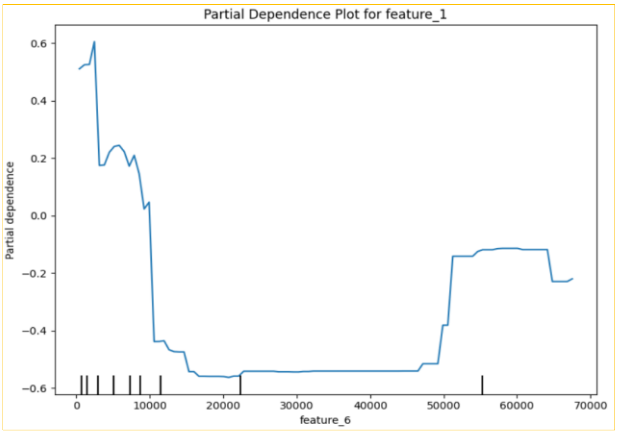

Figure 41. PDP of agriculture output.

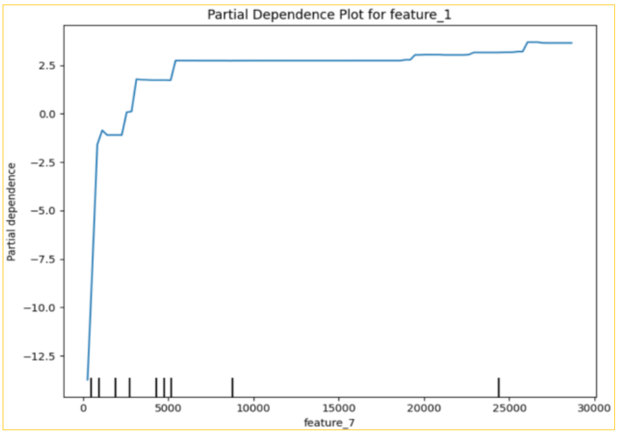

Figure 42. PDP of year employment total.

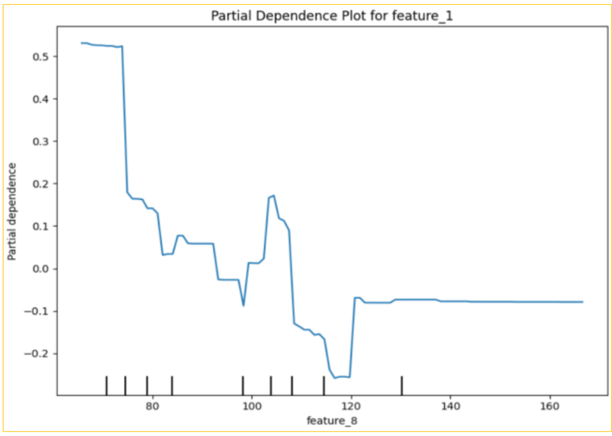

Figure 43. PDP of labor productivity.

Conclusion

From feature importance ranking and partial dependence plots obtained from gradient boost regression, two observations are noticeable. Firstly, real GDP, education level, total employment, life expectancy are key features in explaining and predicting municipal waste recycling rates. Secondly, Slovakia and Germany play a special role because Germany has a leading position among the group of countries with high recycling rates, while Slovakia demonstrates rapid recent improvement and policy shift in enhancing recycling rates.

Careful attention is paid to the improvement induced by dummy variables in terms of measurement scores or dispersions of residual plots. Although a sticking point related with the unexpected signs of coefficients of variables, education level and nominal labor productivity, is mitigated by the feature importance and partial dependence plots, evidence based further clarification is remained.

Appendix

Table 8. Output table of multiple linear regression without country wise dummy variables (i).

Model Summary - Recycling rate

Model	R	R^2	Adjusted R^2	RMSE
M_0	0.667	0.444	0.425	11.181
M_1	0.666	0.444	0.427	11.163
M_2	0.665	0.442	0.427	11.164
M_3	0.661	0.437	0.424	11.190

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
M_0	Regression	25400.016	9	2822.224	22.576	< .001
	Residual	31752.860	254	125.011		
	Total	57152.875	263			
M_1	Regression	25374.537	8	3171.817	25.452	< .001
	Residual	31778.338	255	124.621		
	Total	57152.875	263			
M_2	Regression	25248.650	7	3606.950	28.942	< .001
	Residual	31904.225	256	124.626		
	Total	57152.875	263			
M_3	Regression	24973.574	6	4162.262	33.242	< .001
	Residual	32179.301	257	125.211		
	Total	57152.875	263			

Table 9. Output table of multiple linear regression without country wise dummy variables (ii).

Model		Unstandardized	Standard Error	Standardized	t	p
M_0	(Intercept)	−168.567	599.295		−0.281	0.779
	Year	0.134	0.296	0.023	0.451	0.652
	Life_expectancy	−0.922	0.340	−0.185	−2.714	0.007
	Emplyoment_ICT	0.057	0.013	1.514	4.471	< .001
	Education	0.001	7.214×10^{-4}	1.146	1.740	0.083
	RGDP	4.107×10^{-4}	6.850×10^{-5}	0.635	5.995	< .001
	Greenhouse_emissions	−0.272	0.130	−0.108	−2.091	0.038
	Agriculture_Output	-3.367×10^{-4}	1.286×10^{-4}	−0.489	−2.618	0.009
	Employment_Total	−0.003	0.001	−1.766	−2.224	0.027
	Labour_productivity	−0.045	0.045	−0.093	−1.004	0.316
M_1	(Intercept)	101.732	25.992		3.914	< .001
	Life_expectancy	−0.931	0.339	−0.187	−2.749	0.006
	Emplyoment_ICT	0.058	0.013	1.537	4.596	< .001
	Education	0.001	7.146×10^{-4}	1.108	1.699	0.091
	RGDP	4.108×10^{-4}	6.839×10^{-5}	0.635	6.006	< .001
	Greenhouse_emissions	−0.278	0.129	−0.110	−2.153	0.032
	Agriculture_Output	-3.253×10^{-4}	1.259×10^{-4}	−0.472	−2.583	0.010
	Employment_Total	−0.003	0.001	−1.766	−2.227	0.027
	Labour_productivity	−0.045	0.045	−0.093	−1.005	0.316
M_2	(Intercept)	100.336	25.956		3.866	< .001
	Life_expectancy	−0.946	0.338	−0.190	−2.794	0.006
	Emplyoment_ICT	0.056	0.012	1.484	4.494	< .001
	Education	0.001	6.891×10^{-4}	0.935	1.486	0.139
	RGDP	3.588×10^{-4}	4.476×10^{-5}	0.555	8.016	< .001
	Greenhouse_emissions	−0.274	0.129	−0.109	−2.122	0.035
	Agriculture_Output	-3.080×10^{-4}	1.247×10^{-4}	−0.447	−2.469	0.014
	Employment_Total	−0.002	0.001	−1.568	−2.041	0.042
M_3	(Intercept)	90.828	25.213		3.602	< .001
	Life_expectancy	−0.814	0.327	−0.163	−2.487	0.014
	Emplyoment_ICT	0.043	0.009	1.137	4.855	< .001
	RGDP	3.436×10^{-4}	4.367×10^{-5}	0.531	7.867	< .001
	Greenhouse_emissions	−0.295	0.129	−0.117	−2.290	0.023
	Agriculture_Output	-1.723×10^{-4}	8.519×10^{-5}	−0.250	−2.023	0.044

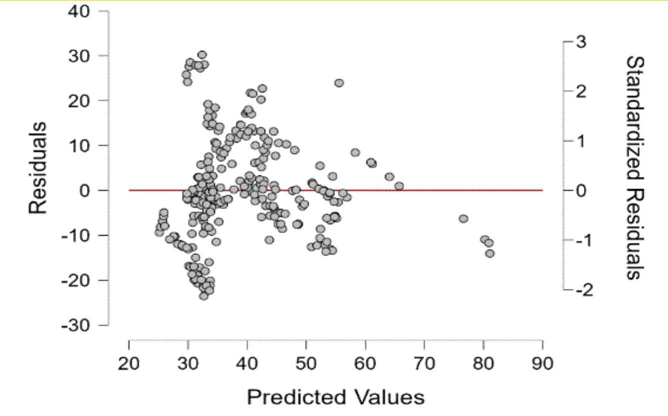

Figure 44. Residual plot of MLR without dummies.

Table 10. Output table of multiple linear regression with country wise dummy variables (i).

Model Summary - Recycling rate

Model	R	R^2	Adjusted R^2	RMSE
M_0	0.965	0.931	0.919	4.199
M_1	0.965	0.931	0.919	4.190
M_2	0.965	0.931	0.920	4.182
M_3	0.965	0.930	0.920	4.175
M_4	0.965	0.930	0.920	4.170
M_5	0.964	0.930	0.920	4.166
M_6	0.964	0.930	0.920	4.164
M_7	0.964	0.930	0.920	4.158
M_8	0.964	0.930	0.921	4.156
M_9	0.964	0.929	0.921	4.156
M-	0.964	0.929	0.920	4.162

Table 11. Output table of multiple linear regression with country wise dummy variables (ii).

M-	(Intercept)	−1548.809	205.653		−7.531	< .001
	Year	0.803	0.102	0.139	7.883	< .001
	Education	−0.002	2.648×10^{-4}	−1.846	−7.635	< .001
	RGDP	1.100×10^{-4}	2.868×10^{-5}	0.170	3.834	< .001
	Labour_productivity	−0.179	0.019	−0.372	−9.563	< .001
	Aus	14.281	1.726	0.166	8.276	< .001
	Bel	14.709	1.975	0.171	7.447	< .001
	Bul	−24.909	2.015	−0.290	−12.360	< .001
	Cro	−28.869	1.961	−0.336	−14.724	< .001
	Cyp	−43.041	1.944	−0.531	−22.142	< .001
	Cze	−9.765	1.901	−0.120	−5.137	< .001
	Est	−29.423	1.973	−0.363	−14.915	< .001
	Fin	−9.619	1.565	−0.119	−6.148	< .001
	Fra	64.578	9.912	0.796	6.515	< .001
	Ger	116.115	13.153	1.432	8.828	< .001
	Gre	−30.551	1.935	−0.377	−15.785	< .001
	Hun	−15.107	1.919	−0.186	−7.871	< .001
	Ice	−32.703	1.874	−0.381	−17.451	< .001
	Ita	70.722	9.225	0.872	7.666	< .001
	Lat	−23.155	2.004	−0.286	−11.557	< .001
	Lit	−10.716	1.901	−0.132	−5.636	< .001
	Mal	−45.111	1.917	−0.556	−23.534	< .001
	Net	19.723	2.525	0.230	7.810	< .001
	Nor	−6.244	1.602	−0.073	−3.898	< .001
	Pol	25.215	5.593	0.311	4.508	< .001
	Por	−18.841	1.889	−0.232	−9.975	< .001
	Ro	−23.430	2.933	−0.289	−7.989	< .001
	Sva	−16.049	1.806	−0.198	−8.888	< .001
	Spa	41.201	7.376	0.508	5.586	< .001

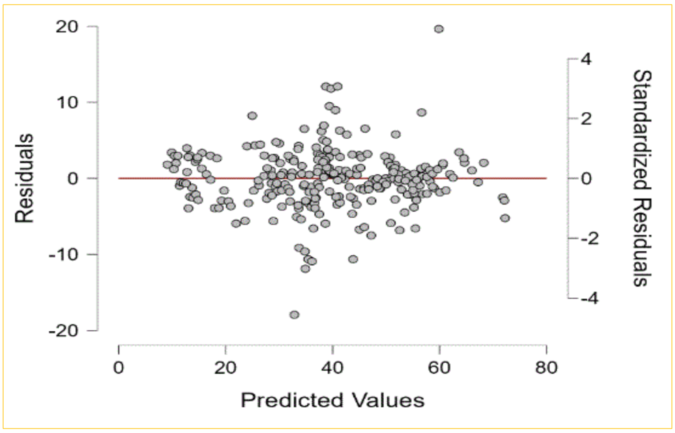

Figure 45. Residual plot of MLR with dummies.

Declarations

Acknowledgments: Authors are thankful to the entire staff of Seoul Innovations Research Institute for insightful ideas and invaluable proofreading.

Author Contributions: AGHC: Design of study, definition of intellectual content, review manuscript, data collection, prepared first draft of manuscript, data analysis, statistical analysis; BN: Statistical interpretation, prepared first draft of manuscript, review manuscript, editing.

Conflict of Interest: The authors declare no conflict of interest.

Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.

Data Availability Statement: The datasets generated and/or analyzed during this study are not publicly available but are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Research Content: The research content of this manuscript is original and has not been published elsewhere.

References

1. Cehlár, M., Taušová, M., Ivanková, V. and Khouri, S. 2025. Municipal waste recycling in the EU: A multi-method analysis of determinants and country profiles (2005–2023). Frontiers in Environmental Science, 13: 1670365

2. Hondroyiannis, G., Sardianou, E., Nikou, V., Evangelinos, K. and Nikolaou, I. 2024. Recycling rate performance and socioeconomic determinants: Evidence from aggregate and regional data across European Union countries. Journal of Cleaner Production, 434: 139877.

3. Kostakis, I. and Tsagarakis, K.P. 2022. Social and economic determinants of materials recycling and circularity in Europe: An empirical investigation. The Annals of Regional Science, 68(2): 263-281.

4. Gabor, M.R., López–Malest, A. and Panait, M.C. 2023. The transition journey of EU vs. Non-EU countries for waste management. Environmental Science and Pollution Research, 30(21): 60326-60342.

5. Georgescu, I., Kinnunen, J. and Androniceanu, A.M. 2022. Empirical evidence on circular economy and economic development in Europe: A panel approach. Journal of Business Economics and Management (JBEM), 23(1): 199-217.

6. Holmen, R.B., Carvelli, G., Razminienė, K. and Tvaronavičienė, M. 2025. Macroeconomic influences on recycling in Europe: An econometric investigation. Circular Economy and Sustainability, 5(1): 573-602.

7. Imran, M., Jijian, Z., Sharif, A. and Magazzino, C. 2024. Evolving waste management: The impact of environmental technology, taxes, and carbon emissions on incineration in EU countries. Journal of Environmental Management, 364: 121440.

8. Chierrito-Arruda, E., Rosa, A.L.M., Paccola, E.A.D.S., Macuch, R.D.S. and Grossi-Milani, R. 2018. Pro-environmental behavior and recycling: Literature review and policy considerations. Ambiente and Sociedade, 21: e02093.

9. Hornik, J., Cherian, J. and Madansky, M. 1995. Determinants of recycling behavior: A synthesis of research results. The Journal of Socio-Economics, 24(1): 105-127.
10. Nikiema, J. and Asiedu, Z. 2022. A review of the cost and effectiveness of solutions to address plastic pollution. Environmental Science and Pollution Research, 29(17): 24547-24573.
11. Santoso, A.N. and Farizal. 2019. Community participation in household waste management: An exploratory study in Indonesia. In: E3S Web of Conferences (Vol. 125, p. 07013). EDP Sciences.
12. Oh, J. and Hettiarachchi, H. 2020. Collective action in waste management: A comparative study of recycling and recovery initiatives from Brazil, Indonesia, and Nigeria using the institutional analysis and development framework. Recycling, 5(1): 4.
13. Osińska, M. 2024. The determinants of municipal solid waste management efficiency in EU countries. Economics and Environment, 88(1): 637.
14. Huang, Q., Chen, G., Wang, Y., Xu, L. and Chen, W.Q. 2020. Identifying the socioeconomic drivers of solid waste recycling in China for the period 2005–2017. Science of the Total Environment, 725: 138137.
15. Peng, X., Jiang, Y., Chen, Z., Osman, A.I., Farghali, M., Rooney, D.W. and Yap, P.S. 2023. Recycling municipal, agricultural and industrial waste into energy, fertilizers, food and construction materials, and economic feasibility: A review. Environmental Chemistry Letters, 21(2): 765-801.
16. Önder, H. 2018. The socio-economic determiners of recycling: An analysis on European countries through a macro perspective. Amfiteatru Economic, 48(20): 405-417.
17. European Environment Agency. 2021, January 27. Digital technologies will deliver more efficient waste management in Europe (Briefing No. 26/2020). European Environment Agency. https://doi.org/10.2800/297122
18. Pelău, C. and Chinie, A.C. 2018. Econometric model for measuring the impact of the education level of the population on the recycling rate in circular economy. Amfiteatru Economic, 20(48): 340–355.
19. European Environment Agency. 2024, December 21. Waste recycling in Europe. https://www.eea.europa.eu/en/analysis/indicators/waste-recycling-in-europe
20. Eurostat. 2020, January 23. Greenhouse gas emissions from waste. European Commission. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200123-1
21. Prakash, C., Roy, A. and Gawdiya, S. 2022. Utilization of municipal wastes in agriculture. Food and Scientific Reports, 3(6): 13-16.
22. Morgan, J. and Mitchell, P. 2015. Employment and the circular economy: Job creation in a more resource efficient Britain. Green Alliance: London. https://doi.org/10.13140/RG.2.1.1026.5049
23. Ruder, S. 2016. An overview of gradient descent optimization algorithms. arXiv. https://arxiv.org/abs/1609.04747

Citation: Aaron Gun-Hee Cha and Bob Nam. 2025. Machine Learning Based Feature Importance Regarding Recycling Rates for Municipal Waste in European Countries. International Journal of Recent Innovations in Academic Research, 9(4): 255-272.