THE CURIOSITY CUP 2024
A Global SAS® Student Competition

Dynamic Impact Of Food Waste In The World
THE GREEN TEAM

ABSTRACT

Waste generation has exponentially increased and is increasingly becoming a critical issue bothering the countries and the world at large, in this report, our team sourced 2 datasets; A World Bank dataset from Kaggle and OECD Waste data. We conducted exploratory data analysis, clustering and forensic forecasting, we examined the trends on primarily food waste generation over time.

INTRODUCTION

In response to the escalating waste crisis, the United Nations General Assembly convened during its seventy-seventh session on 14 December 2022, to proclaim 30 March as the International Day of Zero Waste, signifying a pivotal moment in the global commitment to waste reduction and sustainable management practices. Championed by Türkiye alongside 105 other nations, this resolution underscores the urgent need to confront the multifaceted challenges posed by waste and catalyze collective action towards a zero-waste future.

Waste, in its various forms, engenders a host of detrimental effects that permeate nearly every aspect of human existence. From the pollution of air, water, and soil to the depletion of natural resources and the proliferation of hazardous substances, if left unchecked, the consequences are alarming. It aggravates the climate crisis, nature and biodiversity crisis. It causes water pollution which in turn makes people sick and affects the economy.

Against this backdrop, our report delves into the dynamic impact of waste generation, and leveraging insights gleaned from extensive data analysis and forecasting techniques. Through an intensive exploration, we aim to bring light to the evolving trends, challenges, and opportunities on waste in general, while underscoring the imperative of concerted action towards a more sustainable and resilient future.

THE DATA

There were two datasets used in this research. The primary dataset comes from The OECD Waste Dataset, available on the OECD website, which offers comprehensive insights into waste generation patterns across countries spanning 21 years from 1993 to 2013. This dataset encompasses 17 dimensions or attributes, which provide detailed categorization of waste types and their sources, such as food loss, municipal waste, agricultural waste, and pre-processed waste, among others. The primary focus of this dataset is to examine the volume of waste generated by countries.

Complementing the primary dataset, the research also incorporated a secondary dataset obtained from the World Bank (which was available on Kaggle), comprising data from 215 countries and featuring 25 dimensions or attributes. This secondary dataset includes crucial demographic information such as population figures, as well as various economic indicators for each country under study.

By integrating both datasets, the research aims to explore the interplay between waste generation, population dynamics, and economic factors, in an attempt to explore possible drivers influencing waste generation globally.

PROBLEM

1. What is the most prominent category of waste?
2. Time series trend of waste generation
3. Which Country has the most waste generated?
4. Relationship between the economic indicators and waste generated

DATA CLEANING

A quick dive and clear look into the waste data set showed an inconsistent appearance of some countries in the early ages of the data collection, following this observation a quick filter through for years in which countries were consistent was done to reduce the imbalance data risk posed by this inconsistency, hence a more consistent appearance was filtered which happened to be between 2000 to 2013.

Having this in mind a left join was performed where the second dataset from the World Bank available in Kaggle was joined to the already filtered waste data ON country code and year, this joins the worldbank dataset to the filtered waste data only when it matches the country code and exact year pre-existing in filtered waste for each row in the filtered waste dataset.

This additionally provides us with no null value dataset as both flag and flag codes columns were also dropped after this join.

In SAS studio, we prepared the data for visualisation by renaming columns and also changing varchar date columns to numeric columns

```
data {{_dp_outputTable}} (caslib={{_dp_outputCaslib}} promote="no");

set {{_dp_inputTable}} (caslib={{_dp_inputCaslib}} );


 time_date = input(time, mmddyy10.);

    format time_date date9.;


run;
```
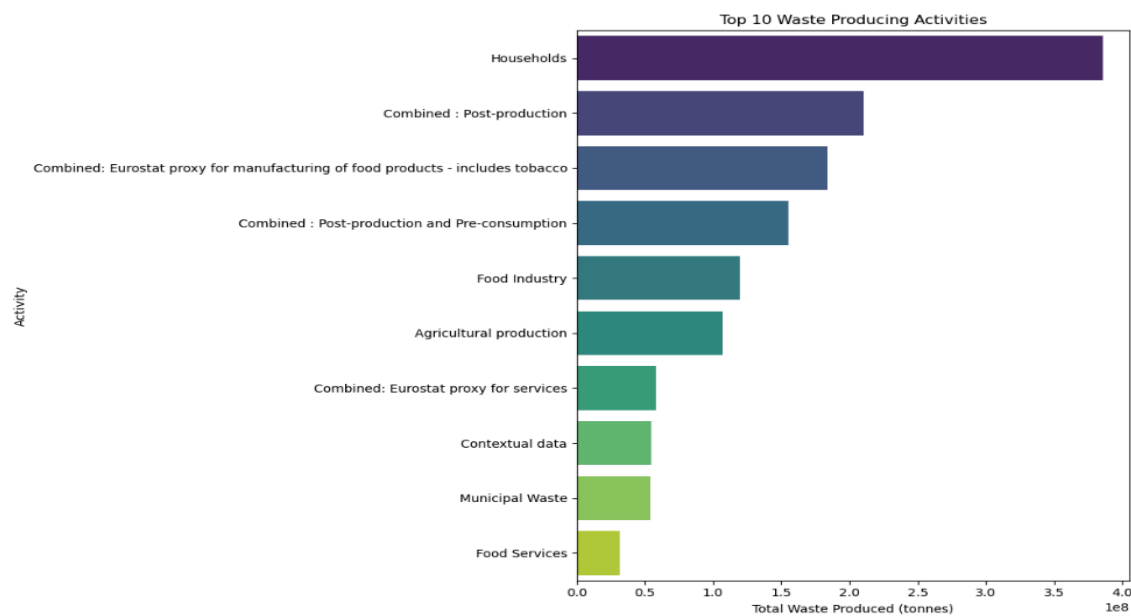
ANALYSIS

The SAS tool that was used was the SAS® Viya® for Learners - SAS studio, SAS Jupyter notebook extension and SAS visual analytics to visualize some results obtained from our analysis. Our analysis discovered that household waste is the most prominent waste.

Figure 2: shows the top 10 waste-producing activities

Figure 2: Top 10 waste-producing activities

The team conducted an analysis of waste-producing activities over the period from 2000 to 2013. We observed that the trend followed a relatively stable pattern, with a slight increase noted in 2005. According to the dataset, " combined post-production stood out as an outlier, while it didn't appear at the beginning of the 10 years, it consistently maintained a significantly higher level on the chart throughout the entire period of its occurrence, while household maintained its appearance over the 10 years, this indicates that post-production food waste should be kept an eye on to mitigate a surge in it.

Figure 3 shows the time-trend of the waste-producing activity over a 10-year
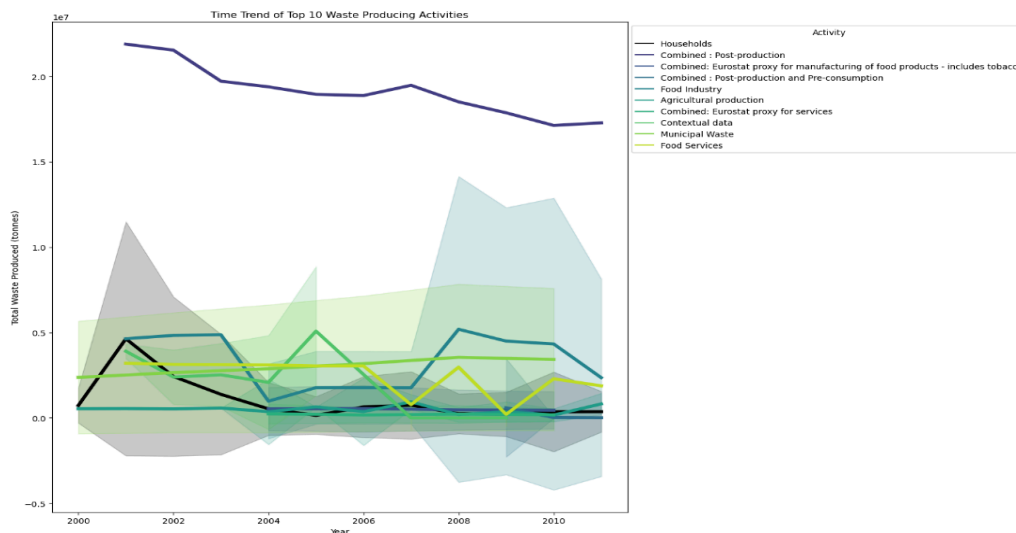
Figure 3: Time Trend of waste-producing activity

The team decided to put more focus on household waste and analysed which country produced the most household waste. It was discovered that "Japan" stood out once again as the country with not only the highest waste per ton but also the highest household waste.

Figure 4 illustrates the distribution of household waste across various countries.
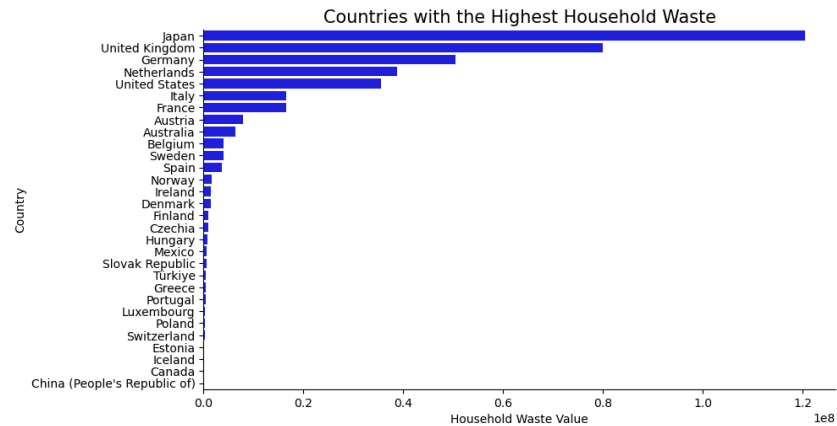
Figure 4: Distribution of household waste across various countries

FURTHER ANALYSIS

The team decided to conduct further analysis of the dataset by using the K-Means clustering modelling approach. Clustering offers a systematic approach to identifying patterns and relationships within complex datasets, enabling the exploration of potentially hidden structures and correlations.

Clustering techniques have been widely utilized in various fields, including environmental science, economics, and healthcare, to uncover meaningful patterns and associations within datasets (Omran et al., 2007). In the context of waste management and resource allocation, clustering can aid in identifying distinct groups of countries based on their waste production, economic indicators, and healthcare expenditure.

We decided to cluster countries based on economic indicators and their relationship to waste generation. Figure 5 shows the outcome of our clustering analysis.
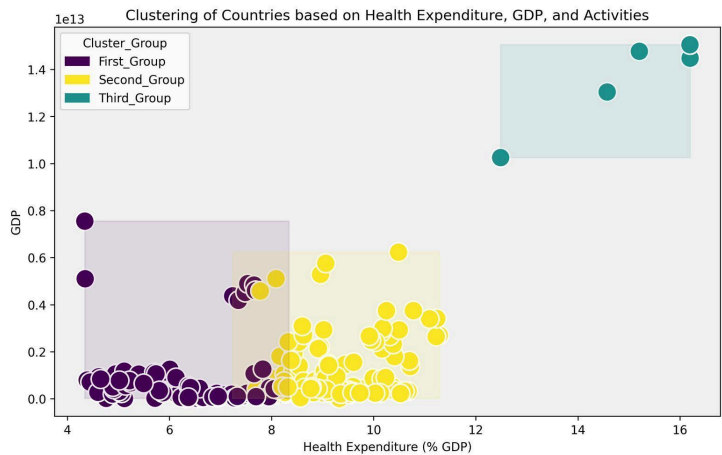
Figure 5: The outcome of our clustering analysis.

4

1. Three distinct cluster groups emerged from the analysis:

    ● The first group comprised countries with low GDP and a correspondingly low percentage (4% to 8%) of GDP allocated to healthcare expenditure.

    ● The second group consisted of countries with low GDP but a relatively higher percentage (8% to 12%) allocated to healthcare expenditure.

    ● The third group demonstrated strong performance, characterized by high GDP and a significant percentage (13% to 17%) allocated to healthcare expenditure. Notably, the United States was the sole member of this group, suggesting a minimal impact of waste activities on overall waste management.

2. Japan consistently appeared in underperforming clusters (first and second groups), indicating a persistent association with top waste-generating activities identified in previous analyses, such as household waste and combined post-production waste.

SUGGESTION FOR FURTHER STUDY

Since this research focused on a limited number of countries, I propose conducting further studies that encompass a broader range of nations with relevant comparative analyses. Additionally, the team intends to undertake more extensive research on ocean waste management, particularly in countries situated along these oceanic regions

CONCLUSION

After extensive data analysis, Japan was identified as the leading producer of waste. It's noteworthy that the team observed the negligible role of population size in waste generation, indicating that effective waste management practices can mitigate societal impacts compared to indiscriminate waste disposal.

Consequently, the study emphasizes the importance of waste management over population size. Additionally, the analysis highlighted "household waste" as the primary contributor to waste generation, emphasizing the necessity for increased attention and education targeting individuals and families regarding proper food waste management techniques.

REFERENCES

Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, *11*(6), 583–605. https://doi.org/10.3233/ida-2007-11602

APPENDIX

The team also used SAS Visual Analytics to create a responsive report on this dataset as well. In one of the pages, The team includes features such as filtering options based on countries through a drop-down menu. Figure 6 displays waste per ton by activity, and users can conveniently filter the information by selecting different countries from the drop-down column.
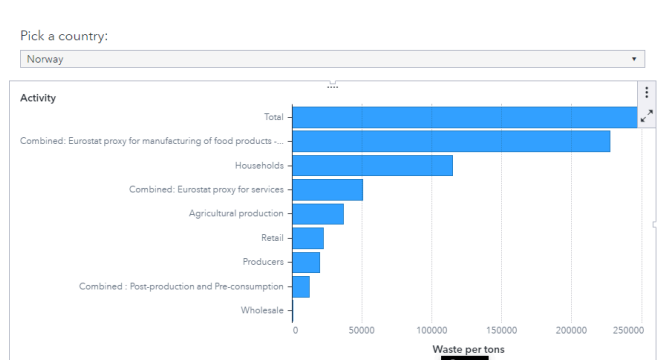
**Figure 6: Waste per ton by category filtered by country**

Figure 7 shows waste per ton by category filtered by country(using a drop-down column)
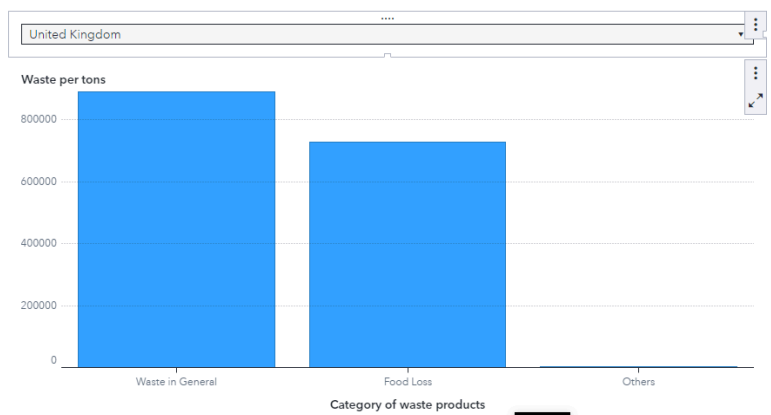
Figure 7: Waste per ton by category of waste filtered by Country

The team visualised a map of the distribution of waste around the world, from the map we see that much concentration of waste can be found in North America region and Japan.

Figure 8 shows the waste distribution on the map
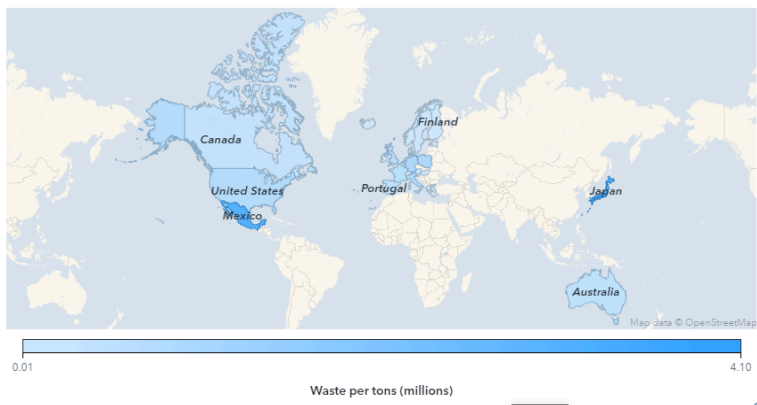
Figure 8: The sum volume of waste generated by countries within the dataset