

2023

MDSS DATATHON COMPETITION



mdss.monash@gmail.com



[Monash Data Science Society](#)



www.linkedin.com/company/monashdss/

{mdss}

Monash Data Science Society

DATATHON CHALLENGE



ANALYZING NATIONAL WASTE DATABASE

National waste management is a critical issue that demands attention, and Data Science can play a pivotal role in addressing this challenge. Waste generation and mismanagement have far-reaching environmental, economic, and social implications. This datathon competition is an opportunity to harness the power of data analysis and prediction to tackle the issue of national waste effectively.

WHY IS THIS SIGNIFICANT?



National waste encompasses a wide range of materials, from household and industrial waste to electronic and hazardous waste. In its most basic form, waste management involves the collection, transportation, disposal, and recycling of these materials.

Data Dictionary

Year : Financial year. Data is presented for each year between 2006-07 and 2020-21 except 2007-08, 2011-12 and 2012-13, for which years a national data set was not developed.

Jurisdiction : State or territory in which the waste was generated.

Category: A broad classification of waste material.

Type: A more detailed classification of waste material. For example, the category 'Metals' may be split into : 'Aluminium', 'Non-ferrous metals (ex. aluminium)', and 'Iron and steel'.

Classification : A reference field stating whether the particular row represents a 'type', 'category' or 'total' (totals collate categories in different ways).

Total type: Describes which category each data point classified as 'Total' refers to.

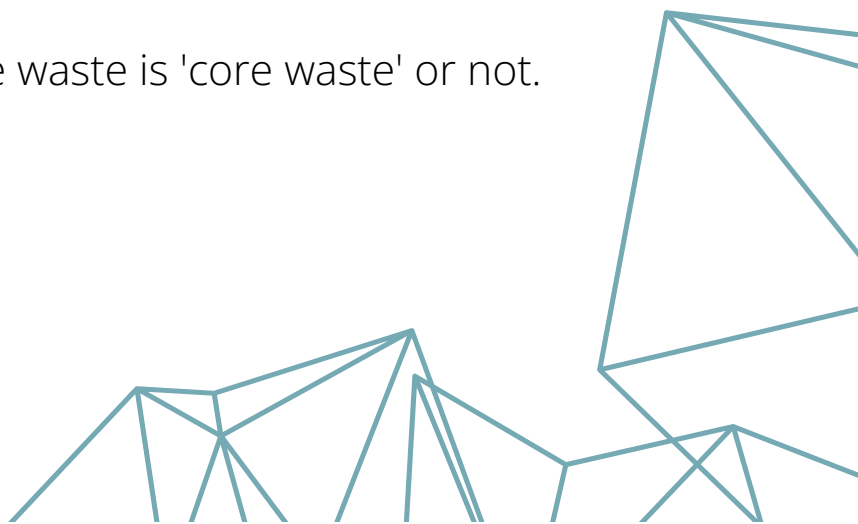
Stream : Describes the source of waste, comprising three options: municipal solid waste (MSW) from households and council operations; commercial and industrial (C&I) waste; and construction and demolition (C&D) waste (plus a total collating all three).

Management: Refers to the infrastructure that receives waste (e.g. landfill, compost facility, alternative waste treatment facility).

Fate: The ultimate destination of the waste, comprising five options: disposal; recycling; energy recovery; long-term storage; and waste reuse.

Tonnes: The quantity of waste.

Core or non-core : Lists whether the waste is 'core waste' or not.



Competition Structure

1 Deriving insights about national waste management database

- Clean and preprocess data and make sure its ready for analysis
- Identify potential correlations or dependencies in the data
- Visualize using appropriate data visualization techniques to identify patterns and trends
- --- and many more methods

2 Predicting future waste generation - Optional

- Utilize predictive modelling to forecast future waste generation using 'Tonnes' column as a target variable
- Evaluate your model using appropriate evaluation metric

Sample Insights/Suggestions

Sample questions to ask from the data

Q1. Is there a seasonal trend in waste generation & disposal?

Q2.What factors or features are most influential in affecting waste generation rates?

Q3. What are the different waste management practices in different jurisdiction?

Q4.Is there any noticeable difference in core or non-core waste?

Evaluation Criteria

Component	Weightage
Inference & Insights	20%
Quality of Code/Comments	5%
Presentation	75%

Presentation Outline

Each team has to present their findings to the judges panel for a maximum of 4 minutes.

- 1 Describe the national waste generation problem
- 2 Insights from data through
Visualizations - Mandatory
Hypothesis testing - Optional
Correlation analysis - Optional
Predictive Modelling - Optional
- 3 Recommendations for better waste management

RESOURCES

The purpose of this guide is to assist all teams in preparing for the competition. Some really great resources to give you a head start into the project.

USEFUL LINKS

- <https://www.data-to-viz.com/>
- <https://www.statology.org/hypothesis-testing-real-life-examples/>
- <https://www.epa.nsw.gov.au/~media/EPA/Corporate%20Site/resources/wastestrategy/SKM-waste-model.ashx>
- <https://www.projectpro.io/article/predictive-modelling-techniques/598>
- [Principles of Effective Data Visualization - https://www.sciencedirect.com/science/article/pii/S2666389920301896](https://www.sciencedirect.com/science/article/pii/S2666389920301896)

Prize Money



Beginner Stream

- First position: \$100
- Second position: \$50

Advanced Stream

- First position: \$250
- Second position: \$150
- Third position: \$100

Best Code and Report Structure

- \$50

TIMELINE



October 6th

- **4:00-4:15:** Welcome Ceremony
- **4.15 - 4.45 pm:** Keynote speech
- **4.45 - 5.45 pm:** Workshop



October 10th

- **4:00pm:** Finish!
- **4:15-6:45pm:** Presentations
- **7:00pm:** Code Submission

Deadline



October 17th

- Results Finalised and Announced

BROUGHT TO YOU BY:

{mdss}

Monash Data Science Society