

DATA 4 GOOD CHALLENGE

DOCUMENTATION





Table of Contents

1. Introduction	2
What is Emergent? What is the D4GC?	
2. The Challenge	3
Transforming Waste: A Data-Driven Revolution	
3. The Data	4
Dig in & gain some insights!	
4. Inspiration	6
Where to start?	
5. Submission	9
What is the output? What are the rules?	
6. Evaluation and prizes	10
Which prizes can we win? How do we win them?	
7. FAQ	11
Frequently asked questions by participants	

1. Introduction

What is Emergent? What is the D4GC?

Emergent, the Data Science & AI student community

We believe that Data Science & AI will change every aspect of our future society. We aim to be the leading student community in this field, welcoming people from all backgrounds. We prepare students to reach their full potential and effect positive change with Data Science & AI throughout their lives.

One of our ways of achieving this vision is by letting students apply their skills to real-life problems through competitions like the D4GC.

The Data 4 Good Challenge, our flagship event

The D4GC allows students to solve a real case. Participants work together in multidisciplinary teams. Together, they are tasked with solving a socio-economic problem through the use of data. In a truly Emergent fashion, participants will have to think through all aspects of their solution, not only considering the economical effects but also the social and ethical consequences of their plan.

Since the challenge is two-fold - produce insights from data, then develop a strategy based on the insights - an **interdisciplinary collaboration** is the best way to tackle this challenge successfully.



DATA 4 GOOD CHALLENGE



2. The Challenge

Transforming Waste: A Data-Driven Revolution

In today's world, effective waste management is no longer a choice—it's an imperative. We are faced with pressing environmental challenges that demand innovative solutions. "Transforming Waste: A Data-Driven Revolution" invites you to embark on a journey of innovation and problem-solving.

The objective is clear: **Develop a robust waste management strategy that directly addresses the most critical issues in waste management today.** But how will you accomplish this mission? By harnessing the power of data storytelling and by creating a state-of-the-art waste management chatbot/visualization tool. Your goal goes beyond providing guidance; you aim to craft a captivating story that delves into the challenges we face and the solutions we can implement. This is how you intend to create an effective strategy that will have a lasting impact.

Data is your most potent tool, and you have access to a range of valuable resources. These include access to an API for an LLM-based Chatbot with access to an extensive knowledge base about recycling and waste management, data obtained from the Recycle Me App from Stad Leuven, Belgium's recycling rules and regulations, recent blogs and articles focusing on upcycling trends, and an extensive array of indicators sourced from Belgium related to waste management. These valuable resources serve as the foundation upon which you will build your innovative strategy. Craft a problem statement, find relevant stakeholders, and explain your ideas with data.

Join us in "Transforming Waste: A Data-Driven Revolution." Together, we will pioneer innovative approaches to waste management and create a compelling narrative that drives effective change.

3. The Data

Dig in & gain some insights!

In the Data 4 Good Challenge, we have curated a diverse and comprehensive collection of data sources to empower participants in developing a data-driven waste management strategy. Our data resources are divided into **two main categories: tabular and textual**, each offering unique insights and perspectives on waste management challenges and solutions. This multifaceted data environment will serve as the foundation for participants to craft compelling narratives and innovative solutions.

D4GC web application

3.1 Tabular Data

Our tabular data resources encompass five distinct datasets in time series format, providing a rich array of information:

Dataset 1: Economical Metrics per Cities (2000 to 2020)

The Economical Metrics per Cities dataset is a tabular dataset that contains comprehensive economic metrics for various cities and countries.

Dataset 2: Waste Production in Belgium (2010 to 2020)

The Waste Production in Belgium dataset is a comprehensive tabular dataset from statbel.fgov.be that provides information on waste production across various economic sectors in Belgium.

Dataset 3: Indicators in Belgium (2000 to 2021)

The dataset contains information on various economic indicators with a focus on sustainability and environmental goals, aligning with international initiatives such as the United Nations Sustainable Development Goals (SDGs).

Dataset 4: Industries accross Belgium

The Industries across Belgium data is a tabular dataset extracted from stat.nbb.be containing information on various economic sectors and their performance in different regions of Belgium over a span of five years.

HINT: This dataset may be used to link multiple datasets together for more elaborate data storytelling and more in depth analysis ;)



**DATA 4 GOOD
CHALLENGE**

3. The Data

Dig in & gain some insights!

Dataset 5: Green Areas

The Green Areas dataset is a tabular dataset from stats.oecd.org containing information on green areas in urban centers across various cities and countries.

3.2 Textual Data

In addition to tabular data, we have compiled a collection of scraped articles from diverse sources, conveniently processed into CSV files. These articles offer valuable insights into waste management challenges, recent trends, and best practices. We have harnessed the power of this knowledge base with a **retrieval-augmented Large Language Model** (LLM) relying on Llama 2 as the backbone. We have **integrated this LLM into an API**, crafting a personalized recycling assistant we affectionately call "**RecycLlama**" (pictured below). By experimenting with prompt engineering, we provide the teams the opportunity to interact with this dynamic resource.

The distinctive feature of RecycLlama lies in its ability to provide tailored and up-to-date responses, setting it apart from generic chatbots like ChatGPT. With this resource, teams can access a wealth of recycling knowledge and leverage it to develop innovative waste management solutions.





4. Inspiration

How to approach the problem?

Structure your approach

Instead of analyzing all aspects of the problem, it often makes more sense to dive deep into one or two crucial aspects of the problem. What follows is a **non-exhaustive** list of aspects to the problem coupled with some of the ways you can tackle that aspect of the problem.

Begin by organizing your team's efforts and responsibilities. Allocate tasks based on individual strengths and expertise to ensure a cohesive approach to the challenge.

1. Problem Statement

Clearly define the specific waste management issue you aim to address. This could be related to recycling rates, hazardous waste, or sustainable practices. A well-articulated problem statement sets the foundation for your solution.

Your statement should be formulated concisely. You should deliver your "solution" to the jury which consists of a mix of different profiles and hence it should be understood easily even by non-technical background personnel.

2. Decide a direction

After having built your problem statement, it might be a good idea to **narrow down the scope** of your research. Consider the data at your disposal and decide on the direction of your analysis. Are you focusing on predictive modeling, benchmarking, or a combination of both? Define a clear path to guide your research.

Don't forget that you are given only 5 hours to not only build your solution, but also prepare a slide deck for a presentation.

3. Develop a DATA-DRIVEN solution

We recommend spending around 1-2 hours completing the first two steps. Now, you should carefully analyze the given data to support (or modify if needed) your statement or prove whether your hypothesis is correct or not. Utilize the insights gathered from your analysis to develop a **DATA-DRIVEN** waste management strategy. Leverage the power of **visualization**, **storytelling**, and **interactive chatbots** to create an impactful solution that addresses the identified problem and engages users effectively.



4. Inspiration

Where to start?

To successfully tackle the challenge, you should combine solid data analysis with strategic insight. On which areas you focus is completely up to your team. Here, we have listed a couple of starting points to solve the challenge. **We have prepared a code sample (both in Python and R).**

Note that these are just guidelines. You can solve the challenge any way you want.

Use case 1: Studying the evolution of waste management practices over the years

How does the waste management practices change over the years in Belgium? How do Belgian efforts for recycling and waste management compare to other countries in the EU? Has there been any shifts in domestic material consumption, hazardous waste generation, and municipal waste trends over the years?

Use case 2: Understanding the disparity in recycling trends across regions in Belgium

How does different regions across Belgium compare in recycling rates and waste production? Which areas have lower recycling and upcycling rates? How could you investigate the contributing factors and provide targeted interventions?

Use case 3: Topic modelling and extraction of key themes handled in recycling and upcycling

What are the most prevalent themes in recycling and upcycling articles? What are the most emerging trends and innovative solutions you could extract from RecycLlama? What are the possible challenges and limitations these waste management strategies might uncover?

4. Inspiration

Interesting sources to look at



Early warning assessment related to the 2025 targets for municipal waste and packaging waste

The early warning report assists Member States in meeting 2025 recycling targets for municipal and packaging waste, as well as the 2035 target to limit landfilling to 10%. It builds upon previous Commission support for EU waste management compliance, including the 2018 early warning report.

[<LINK>](#)

How One Belgian City Used Data Analytics To Stop Illegal Waste Dumping

Ghent's tourism surge caused litter and waste problems. The city collaborated with IVAGO, a recycling service, using a mobile reporting app and tablets. They deployed SAP Analytics Cloud for real-time insights, service optimization, and efficiency. SAP tech centralized data and automated KPI analysis, resolving stock and metric issues for IVAGO.

[<LINK>](#)



DATA 4 GOOD CHALLENGE

5. Submission

What is the output? What are the rules?

Create a pitch deck

You will use a slide deck to pitch your solution to the jury. You are free to use any software you want (PowerPoint, Canva...) to create the presentation, but the result must be a file in .pdf format.

Your pitch can be a maximum of 6 minutes long. An Emergent member will be present to signal the time you have left and to cut you off if you go over the time limit.

After the pitch, the jury members will have some time to ask questions. It's a good idea to think about which points of discussion might come up (e.g. limitations of your analysis).

Submit your work by 6:00 PM

At 18:00, you will need to submit everything you worked on. Please collect all your work, put it in a zip file, and upload it. The following needs to be submitted (if applicable):

- Your slide deck/presentation (in **.pdf format**)
- Your python/R/other code
- Your BI tool files (Tableau, Power BI...)
- If you used cloud software for which you can not download the underlying analysis (e.g. AWS), or code that can not be opened with freely available software, please take screenshots of everything and upload them as well.

You will not be able to change anything about your submission (e.g. correcting typos, adding/removing slides, fixing bugs) after uploading. The slide deck you submitted will be the one used for your pitch, and it will be pre-downloaded to the pitch room computer.

[CLICK TO SUBMIT](#)

Rules & anti-cheating measures

Breach of the following rules may result in disqualification from the challenge:

- When using additional resources, figures or data, you must specify a source.
- Every statistic or figure that can be calculated based on the provided data has to be created by the team. You may not copy (partial) solutions from the internet.
- You may not use any data visualisation found online (All data visualisations or analyses should be fully yours)

Anti-cheating measures have been implemented during the creation of the challenge. After months of working on this challenge, we have gotten a very good idea about what can and cannot be found online. Cheating will not be tolerated and results in **immediate disqualification** from the D4GC and all future Emergent events.



**DATA 4 GOOD
CHALLENGE**

6. Evaluation & Prizes

Which prizes can we win? How do we win them?

The solutions that you present will be judged on many different aspects. The exact details about how you will be judged will **not be shared** with the participants. Nonetheless, we have some guidelines that can help you win one of these prizes!

Important notes: the jury consists of both technical and non-technical people. Make sure you communicate in a way that both understand. Consider also that the jury might have only a limited understanding of the topic of the challenge.



Best Data Visualisation - €250

Convince the jury by creating stunning visualizations that contribute to the case you are trying to build. Make sure to present the insights you gained in a clear and understandable way, and link it to impact.



Best Pitch - €250

Focus on delivering a pitch that can convince stakeholders of your solution. Clearly communicate in a structured and comprehensive way. When one of the judges asks questions, answer them in an insightful way.



Best Technical Solution - €500

Your solution should be a technical solution that uses the most appropriate techniques to successfully tackle this challenge. You should not be afraid of using more advanced methods as long as they are the right tool for this problem.



Best Business Strategy - €500

Develop a solution that is feasible, valuable and tackles the problem or opportunity you identified in a complete way. Don't forget to clearly communicate what your solution tries to achieve and why it is valuable.



Best Overall - €1000

Find a balance between all the points above and convince the jury that your solution is the solution your stakeholder(s) need(s).



7. FAQ

Frequently asked questions by participants.

Do we have to use all of the available datasets?

No, it's up to you to decide what data to work with. For instance, you do not need to look at the green area data if it is not of your interest.

Do we have to use a particular software / product?

No, you can choose either Python or R for your analysis. You can of course use both if needed. Additionally you're allowed to use whatever other software you deem necessary or useful

Can we use data from the internet?

*Yes, in this case you **MUST** provide the source of the dataset(s) you used.*

Can we use information from the internet?

Yes, in this case you must provide the source of the information you used.

Do we need to prepare a slideshow presentation?

Yes, you need to prepare a slideshow presentation and upload a PDF version of it along with your code/BI tool files.

Does everyone from the team need to present to the jury?

No, you may decide among yourselves, but at least one person should present.

How much time do we have to present?

You have 6 minutes to present your strategy and 4 minutes to answer the questions from the jury.

Do we have to do a technical analysis?

No, it is not necessary. That is up to you.

You can opt for a purely strategic approach, of course, grounded in data.

If you want to include a technical approach, you can. But, keep in mind that your presentation time is limited to 6 minutes only.