EMERGENT

NOVEMBER 24th

DATA 4 GOOD CHALLENGE

DOCUMENTATION

dataroots McKinsey & Company aws Belfius datacamp

Table of Contents

EMERGENT

DATA 4 GOOD CHALLENGE

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.

EMERGENT

DATA 4 GOOD CHALLENGE

2. The Challenge

Urban Planning: Leuven City

Effective urban planning is essential to deliver a good living environment to inhabitants. Urban planners in both the public and private sectors employ data-driven methods to address a wide array of issues that have long-term implications for communities and the surrounding landscape.

Stad Leuven, the Leuven city hall, has been working on various campaigns which aim to create a sustainable and smart city presented by the Leuven 2030 roadmap. In addition, motivated by 'De Zilverroutes' in Ostende and 'Red De Stoep' campaign by Mobiel 21, Stad Leuven purchased 10 public benches to provide better infrastructure to seniors in Leuven. However, a lack of solid urban planning and covid outbreak intervened in the initiation of the bench installation. Now, Stad Leuven is eager to develop its own urban planning which provides initial assistance in finding proper locations for the new benches, and further, suggests how the city of Leuven can benefit from the planning in the long term.

During this challenge, you will build your own data-driven urban planning. The planning affects your daily life directly, and hence it is of utmost importance to closely look at the demographic data generated from the city. After having built your urban planning, you find the proper location(s) to install the 10 benches. We emphasize that your urban planning should be data-driven: What data did you utilize to develop yours? Who will benefit from your planning? Think carefully about your targeted stakeholder(s). Build a problem statement, find relevant stakeholders, and explain your ideas with data.

There is no specific condition to be taken into account in your urban planning. Hence, please be creative! Note that the winning solutions will be delivered to Stad Leuven and will be taken into account for actual implementation.

EMERGENT

DATA 4 GOOD CHALLENGE

3. The Data

Dig in & gain some insights!

You are provided with 4 datasets. Every dataset is available via the D4GC web dashboard that we provide. Datasets 1 and 2 are the main datasets that you must utilize to come up with your solution. Datasets 3 and 4 (Traffic and Temperature) are publicly available and the web dashboard provides guidelines on how to utilize these data. The datasets are complementary to each other. You can get limited insights from the individual datasets, but to properly tackle the challenge, you should combine some of them.

Dataset 1: Leuven Demographic Data

This dataset contains demographic information for each statistical sector in the city of Leuven. It is assumed to be the most recent data, and therefore your analysis should be based on the records even if there is little uncertainty. The code snippets (written in Python and R) are available on DataCamp Workspace.

Dataset 2: Leuven Location Data (Tree, bench, etc)

This dataset shows the locations of the currently available public trees, benches, trash cans, bicycle parking, bus stops, and food stores in the city of Leuven. The data is provided on the D4GC web application that is given as default. We have visualized all the locations of the data on the interactive map. Note that the total number of trees, benches, and trash cans in each statistical sector is included in Dataset 1.

Dataset 3: Leuven Traffic Data (Telraam)

This dataset shows the traffic (pedestrians, bicycles, cars, and heavy vehicles) of some streets in the city center of Leuven. The data is publicly available on Telraam, and be approachable via the D4GC dashboard. Note that Postman Telraam API is optionally available and integrated into the D4GC application.

Dataset 4: Leuven Temperature Data (Leuven.cool)

This dataset tracks the real-time temperature (and other environmental features too) in the city of Leuven. The data is publicly available on Leuven.cool (and WOW-BE), and be approachable via the D4GC dashboard.

3. The Data

Dig in & gain some insights!

Data 1: Variable Explanation

area_in_ha	[amount] area of the territory (in hectares)
total amount of residents according to National Register	[amount] residents according to the National Register of Belgium (incl. asylum seekers)
population density [per km^2]	[ratio] population density: inhabitants per km^2
men	[amount] men (incl. asylum seekers)
women	[amount] women (incl. asylum seekers)
0-17 year olds (w.r.t. all residents) [%]	[ratio] residents with an age between 0 and 17 years w.r.t. all residents
65-... year olds (w.r.t. all residents) [%]	[ratio] residents over the age of 65 w.r.t. all residents
births	[amount] births
deaths	[amount] deaths
natural growth	[amount] natural growth: births - death
total growth	[amount] total growth: population on 31/12 - population on 01/01, also the sum of natural balance and migration balances
elderly dependency ratio [per 100]	[ratio] ratio of the over-65s to the 20-64 year-olds * 100
youth dependency ratio [per 100]	[ratio] ratio of 0-19 year olds to 20-64 year olds * 100
caregiving ratio	[ratio] ratio of 40-79 year-olds to over-80s *100
dependency coefficient [per 100]	[ratio] dependency coefficient or dependency ratio: ratio of the number of 0-19 year olds and over-65s to 20-64 year olds * 100
family care index [per 100]	[ratio] family care index: ratio of over-80s to over-50-59 year-olds * 100
non-Belgian current nationality (w.r.t. to residents)	[percentage] non-Belgian current nationality (w.r.t. inhabitants), incl. asylum seekers
non-Belgian origin (w.r.t. to residents)	[percentage] non-Belgian origin (w.r.t. inhabitants), incl. asylum seekers

EMERGENT

DATA 4 GOOD CHALLENGE

3. The Data

Dig in & gain some insights!

Data 1: Variable Explanation

0-19 year old	[amount] 0-19 year old
20-39 year old	[amount] 20-39 year old
40-59 year old	[amount] 40-59 year old
60-79 year old	[amount] 60-79 year old
80+ year old	[amount] 80+ year old
private households	[amount] private households (incl. asylum seekers)
households without (LIPRO) children	[amount] households without (LIPRO) children (incl. asylum seekers)*
multi-person households with at least 1 minor (LIPRO) child	[amount] multi-person households with at least 1 minor (LIPRO) child (incl. waiting register)*
multi-person households with only minor (LIPRO) child(ren)	[amount] multi-person households with only adult (LIPRO) child(ren) (incl. waiting register)*
households 1 person	[amount] household 1 person
living alone (w.r.t. to private households)	[percentage] household type 'living alone' (w.r.t. total private households) (incl. asylum seekers)
household without (LIPRO) children (w.r.t. private households)	[percentage] households without (LIPRO) children (compared to total private households) (incl. asylum seekers)
multi-person households with at least 1 minor (LIPRO) child (compared to total private households)	[percentage] multi-person households with at least 1 minor (LIPRO) child (compared to total private households) (incl. asylum seekers)
single-parent household with min. 1 minor (LIPRO) child (w.r.t. household with min. 1 minor (LIPRO) child	[percentage] family consisting of a single parent with at least 1 minor (LIPRO) child (compared to households with at least 1 minor (LIPRO) child)*
unemployed jobseekers (w.r.t. residents 18-64 year old)	[percentage] unemployed jobseekers (w.r.t. residents aged 18-64)

3. The Data

Dig in & gain some insights!

Data 1: Variable Explanation

short-term unemployed jobseekers (w.r.t. residents 18-64 year old)	[percentage] short-term (< 1 year) unemployed jobseekers (w.r.t. unemployed jobseekers)
long-term unemployed jobseekers (w.r.t. residents 18-64 year old)	[percentage] long-term (1-2 years) unemployed jobseekers (w.r.t. unemployed jobseekers)
very long-term unemployed jobseekers (w.r.t. residents 18-64 year old)	[percentage] very long-term (> 2 years) unemployed jobseekers (w.r.t. unemployed jobseekers)
commercial properties	[amount] commercial property
commercial properties per 1.000 residents	[ratio] commercial property per 1.000 residents
stores: food products	[amount] stores: food products
stores: personal care	[amount] stores: personal care
stores: clothing and fashion	[amount] stores: clothing and fashion
stores: household goods	[amount] stores: household goods
stores: leisure	[amount] stores: leisure
stores: in and around the house	[amount] stores: in and around the house
stores: brown and white goods	[amount] stores: brown and white goods
stores: car and bicycle	[amount] stores: car and bicycle
stores: do-it-yourself	[amount] stores: do-it-yourself
stores: others	[amount] stores: others
hospitality business	[amount] hospitality business
consumer-oriented services: culture	[amount] consumer-oriented services: culture
consumer-oriented services: transportation and fuels	[amount] consumer-oriented services: transportation and fuels

3. The Data

Dig in & gain some insights!

Data 1: Variable Explanation

consumer-oriented services: recreation	[amount] consumer-oriented services: recreation
consumer-oriented services: others	[amount] consumer-oriented services: others
commercial properties - vacant (incl. renovation up to and including 2019)	[amount] commercial properties - vacant (incl. renovation up to and including 2019)
vacant commercial properties (w.r.t. commercial properties)	[percentage] vacant commercial properties (w.r.t. commercial properties)
stores: daily goods (w.r.t. stores)	[percentage] daily goods (w.r.t. stores)
commercial properties 'other retail' (w.r.t. retail)	[percentage] commercial properties 'other retail' (w.r.t. retail)
stores: periodic goods (w.r.t. stores)	[percentage] stores: periodic goods (w.r.t. stores)
stores: periodic goods (w.r.t. stores)	[percentage] stores: periodical goods (w.r.t. stores)
residential units in apartments	[amount] residential units with residential function in apartment buildings
residential units in community housing	[amount] residential units with residential function in community housing
residential units in individual housing	[amount] residential units with residential function in individual housing
renters (w.r.t. households with known title of ownership)	[percentage] renters (relative to households with known title of ownership)
social rental housing (social housing company + social rental office)	[amount] social housing units
persons with an increased benefit	[amount] persons with increased health insurance benefits
persons with an increased benefit (w.r.t. persons on health insurance)	[percentage] persons with increased health insurance benefits (w.r.t. total health insurance beneficiaries)
average net taxable income per resident	[average] net taxable income per capita (in €)

3. The Data

Dig in & gain some insights!

Data 1: Variable Explanation

preschool students	[amount] preschool students - according to the residence of student
elementary school students	[amount] elementary school students - according to the residence of student
secondary school students	[amount] secondary school students - according to the residence of student
indicator pupils primary education	[percentage] indicator elementary school pupils (w.r.t primary school pupils) - according to pupil's place of residence
indicator pupils secondary education	[percentage] an indicator pupil in secondary education is one who receives a scholarship and/or has a low-educated mother, according to the pupil's place of residence
education deprivation indicator value primary education	[ratio] OKI (Education Opportunity Indicator) values primary education - according to the student's place of residence
education deprivation indicator value secondary education	[ratio] OKI (Education Opportunity Indicator) value secondary education

3. The Data

Dig in & gain some insights!

Data 2: Leuven Map

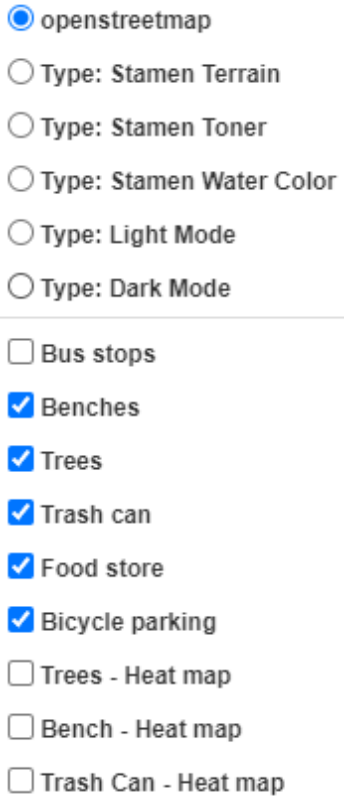
MARKER TYPES

Bicycle Parking

Public Bench

Foodstore

Trashcan



Select a type of map you want

Select what you would like to see on the map

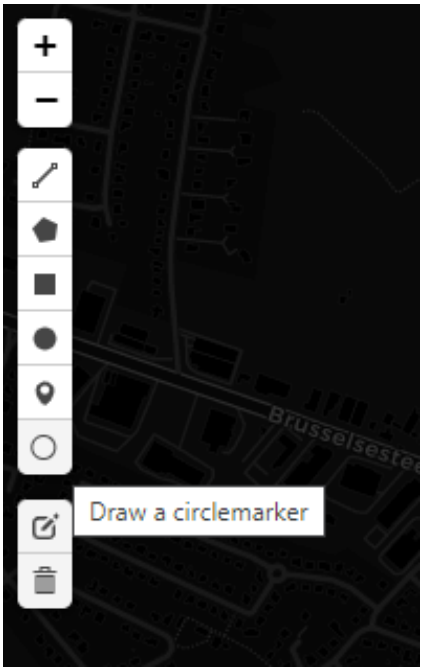
EMERGENT

DATA 4 GOOD CHALLENGE

3. The Data

Dig in & gain some insights!

Data 2: Leuven Map



Draw a marker on the map

EX: circle

Click a sector to see its name

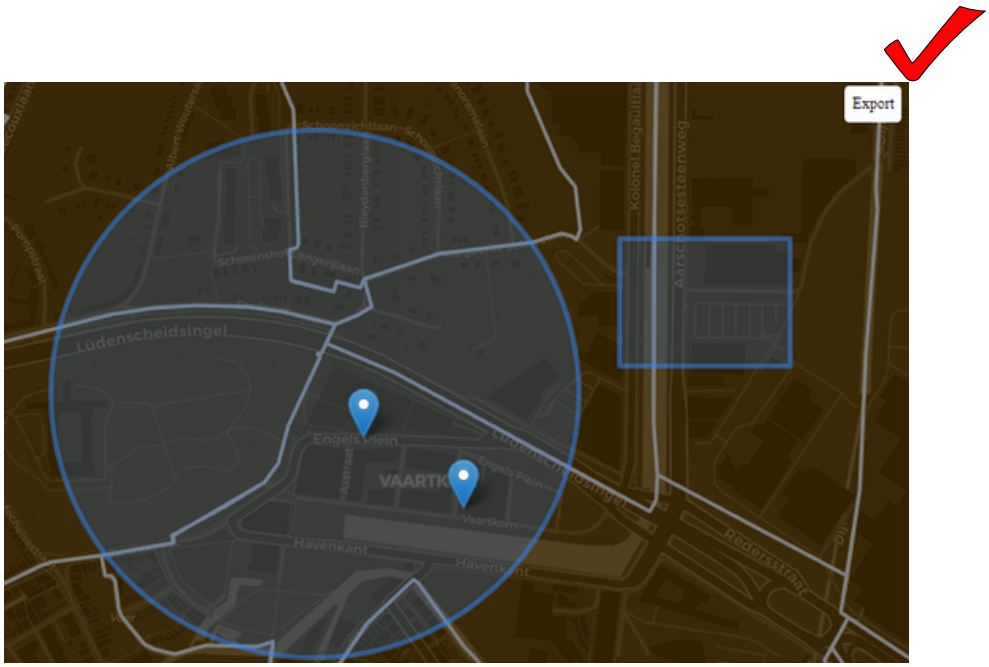
EMERGENT

DATA 4 GOOD CHALLENGE

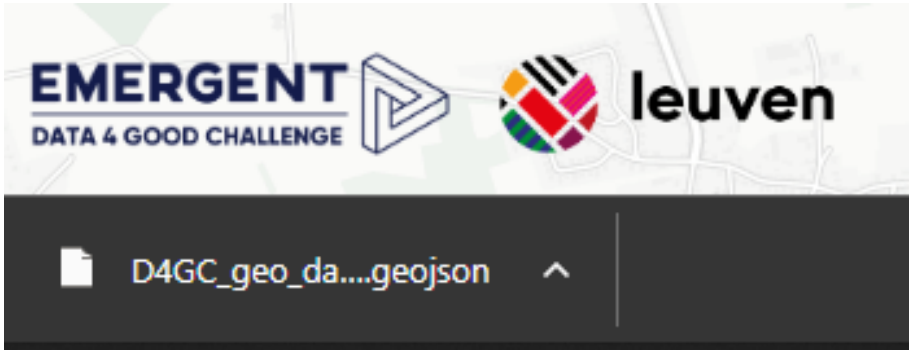
3. The Data

Dig in & gain some insights!

Data 2: Leuven Map



1. Draw markers in the area you want
2. Click the "Export" in the top right corner



3. The coordinates of the markers are automatically saved in a geojson file
4. Feel free to use the coordinates for further analysis

EMERGENT

DATA 4 GOOD CHALLENGE

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.

Note that how you implement these aspects might differ based on the stakeholder you have chosen. Always remember to link back your research to the end user of your recommendations!

1. Problem Statement

A first part of the challenge is defining a concrete problem statement. Look at the current Leuven urban planning. Which characteristics (demographic, geographic, ...) make the planning vulnerable? Which population groups are not targeted by the planning? How do these affect the stakeholder you have chosen?

Your statement should be formulated concisely. You should deliver your "solution" to the city hall, 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 (e.g. limit yourself to a certain target or a set of targets with similar characteristics/risks).

Don't forget that you are given only 4 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 an hour 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. Note that your urban planning should be DATA-DRIVEN. Anybody can pick the location(s) to place the benches on the map. However, you should provide data-driven reasoning to support your solution. In other words, you should be able to persuade others why and how you used data to support your urban planning.

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) in the DataCamp Workspace to help you get started.
Note that these are just guidelines. You can solve the challenge any way you want.

Focus on ONE stakeholder

Your team can consider any of these (and other!) stakeholders and build a solid case for them. Think about what is important for these stakeholders. Who does your urban planning target? How are they impacted by it and what potential do they have to develop a response? Once you identified these points, start looking for answers in the data we provided. Don't forget to clearly communicate your stakeholder to the judges!

Citizens

Travelers

Leuven citizens (Travellers) will be affected by the urban planning. Who is your target? Which group do you think needs to be benefited the most?

Policymakers

Stad Leuven is eager to faciliate the usage of data to create suitable policies which add value to our daily life. What is your suggestion to successfully implement this plan?

Industry

Leuven-based business may need effective urban planning in order to boost their business performance. How would you assist their needs?

Someone else

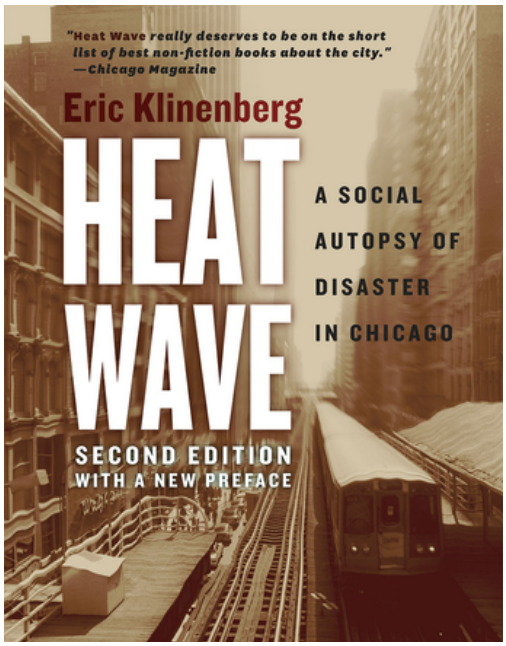
Feel free to come up with your own stakeholder!

EMERGENT

DATA 4 GOOD CHALLENGE

4. Inspiration

Interesting sources to look at



1995 Chicago Heat Wave

Robust social infrastructure saved more life

Eric Klinenberg, an American sociologist, explained in his book, "Palaces for the People: How Social Infrastructure Can Help Fight Inequality, Polarization, and the Decline of Civic Life" that the importance of social infrastructure is far more significant than we think. In 1995, the tragedic heat wave in Chicago caused 739 deaths over a period of 5 days. The research found out, after the tragedy, that a district with having more robust public infrastructure had much fewer casualties than the rest. The reason was that this public facility enabled citizens to let engage with their neighbors more, and then the citizens were able to check the status of their people in the community.

Check: "346 - Palaces for the people" 99% invisible

Sitting Together: How Benches Help Build Community

Bench offers the opportunity to participate, connect with neighbors, and share a sense of belonging in public spaces. The placement of your bench is going to determine how successfully it invites people to linger and enliven your public space.

For the placement, you may consider including shade, shelter from the wind, and other elements. Often, benches should be placed close to other public infrastructure, such as trashcans.

Check more information HERE

EMERGENT

DATA 4 GOOD CHALLENGE

4. Inspiration

Interesting sources to look at

Social Life Project

How Seating Shapes Welcoming Cities

Benches are not just objects, they determine social behavior. Lots of cities in the modern era are built with different types of benches with different styles, and it is certain that those benches influenced the image creation of the cities. In other words, benches are not only for those who live in the city but also to attract people from outside.

Check more information HERE

Let's Put A Bench on Every Street Corner

"If we make street corners more comfortable, starting with something like a bench, we can make our streets safer, more sociable, and more inviting places. Chance encounters and long conversations can become the norm."

Check more information HERE

EMERGENT

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 7 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 (notebooks downloaded from DataCamp workspace)
- 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.

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.

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).

EMERGENT

DATA 4 GOOD CHALLENGE

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 temperature data if it is not of your interest.

Do we have to use a particular product?

Yes, you should work on DataCamp workspace, but you can choose either Python or R for your analysis. You can of course use both if needed.

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 7 minutes to present your strategy and 3 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 7 minutes only.