
Noise in Leuven?

Modern Data Analytics

Exam Project GOZ39a

Introduction

There is no exam for this course, instead a project has to be handed in. The project involves a data analytics project related to the noise problem in the City of Leuven, more in particular the neighborhood of the Naamse Straat¹. The City of Leuven is trying to strike the balance between a vibrant nightlife and people getting a good night's sleep. A trial project used sound monitors to map night-time noise levels, along with behavioural 'nudges' to reduce noise. In this project you are exposed to two sets of data:

- Noise levels during the year 2020 from these monitors (Appendix A)
- Meteo data during the same period (Appendix B)

Open Research Question

There is no particular research question. In other words: **it is up to you to provide insight into the data**. By no means are you restricted to stick to the suggested data. On the contrary, a good data-scientist is capable to "think out of the box" and to retrieve data from other sources and blend these together.

It would be a shame, if the only thing you achieved in this course, is the production of a vanilla Jupyter Notebook where you deliver some standard charts using a technique picked up during the lectures. You should have advanced much further on the learning curve than the topics we handed out during the lectures.

What is expected from each Team?

- In a nutshell, **we want you to think as a data-scientist throughout the whole production pipeline**: retrieving & pre-processing data, creating informative visualizations, exception handling, building a model, hyperparameter optimization, etc...
- We expect that you bring the topics explained during the course into practice. Your team should be able to bring value to the data. You can use techniques that were not covered during the course and can bring other python packages into the project. This is even considered a plus.

¹ <https://cities-today.com/leuven-tackles-night-noise-with-smart-nudging/>

- Make sure you start from the same python environment, used in the course. Of course you can update packages, install new ones, ...
- Make sure that you understand the underlying mathematics in the approach that you use (supervised, unsupervised, nlp, AI).

What do you need to hand in ?

The deliverables shall consist of:

1. A program (python and/or cypher (if any))
Code can be cypher/python scripts, a jupyter notebook or both.
If you managed to deploy your solution as an app, this is considered a big plus.
2. A report (pdf)
3. A presentation (slides)

Deliverable

The project is a group-effort which has to result in the **Three** deliverables mentioned above

- Your Python Code / Cypher scripts have to be shared on a GitHub account.
- Report (pdf) of maximum 2 pages
- Each team will be invited for a presentation on-campus or on-line (teams are free to choose).
- At a later stage you will have to participate in a peer-review, scoring your team-mates on their efforts.

A Google-drive folder will be provided to the team-captains to upload results.

Exam / Presentation

There are 3 dates scheduled for the exam / presentation:

- June 9
- June 22
- June 29.

You should already have been assigned one of these dates. If some of the team-members are not scheduled on the same date, feel free to propose an alternative date (one of the three dates above).

During this presentation the team will present their findings (15 min) and will answer questions (10 minutes). Each team member will receive one or more questions on the project and one or more theoretical questions on some items covered during the lectures. Team members can be asked questions on any code-snippet.

Delivery Date

Before June 5th, 2023

Failure to deliver the report in time, results in a no-pass grade for this course

Grading

- 20% on the presentation and Q&A
- 20% on the written report
- 50% on the work done (project)
- 10% peer evaluation

Grading Criteria

Below is in bullet-point format a **non-exhaustive** list of the criteria that we will take into account when we evaluate your work.

Modeling

- Are you able to reach out beyond the two data-sets and add your own data?
- Visualisation
- Code: Style & Organisation of your Python/Cypher Code.
- Your code needs to be documented and clean! In 2022, we got some very messy notebooks.
- Does the code actually work?
We should be able to clone your code on github and run it on our computer. Make sure that you use a **requirements.txt** file to specify the python packages you require.
- Do you understand the code you wrote ?
- Delivery App
If you deliver an App, the code should be on Github. Great if the app has been deployed.

Content of the report

- Your pipeline: from retrieving data to the actual model
- Introduction and problem statement
- Research method & scientific character of the work done
- Argumentation
- Results: discussion/ interpretation
- General conclusion
- Coherence/ logical composition
- Originality & creativity
- References

Presentation

- Presentation: used language
- Presentation: content/ accessibility
- Presentation: form/ composition/ timing
- Understanding underlying mathematics
- Answering questions on Python, Machine Learning, Cypher Code

Failure

There are three ways to obtain a "no-pass" result:

- Your project did not receive a pass grade
- Your team did not hand in a report in time.
- Your team did receive a pass grade but you failed to answer the questions you were asked during the Q&A

In case of a "no-pass" result, students can participate in the August exam. Here new projects will be made available. The August session will be individual, not in team work.

Q&A

We will schedule a Q&A session where you can ask questions. **Do not contact the owners/ research teams that delivered/assembled the data.** These persons should not be disturbed at all, they were very kind to allow us to use the data. There are 20 different teams in the MDA-course. Hence if every team asks questions to these researchers, this would be very annoying. This is the reason we have a Q&A-session.

Appendix A : Meteo Data

Source

<https://rdr.kuleuven.be/dataset.xhtml?persistentId=doi:10.48804/SSRN3F>

<please do not contact the authors of this paper>

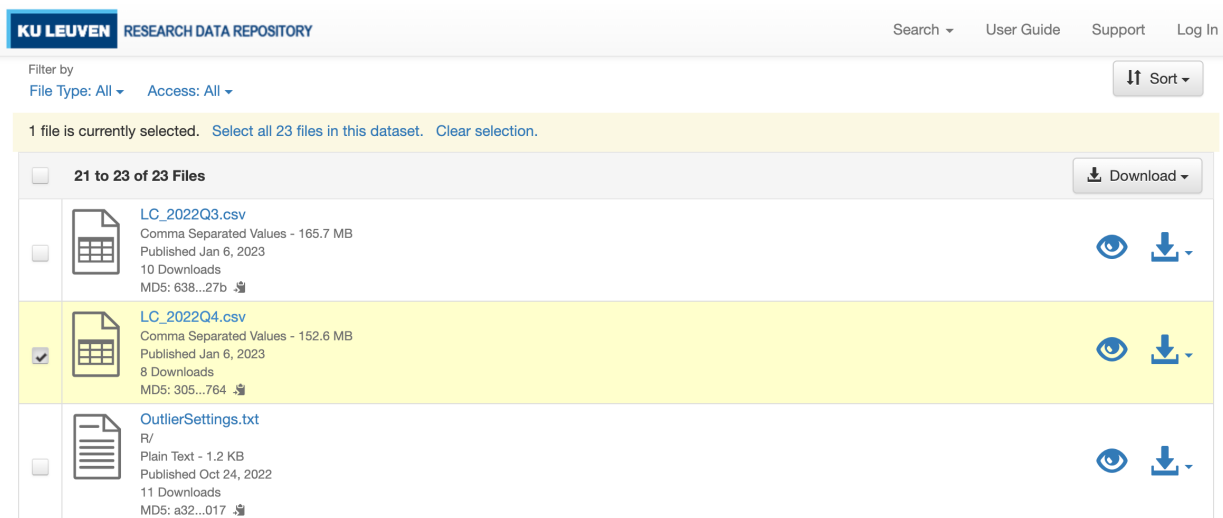
Short Description

This dataset presents crowdsourced data from the Leuven.cool network, a citizen science network of around 100 low-cost weather stations. With the Leuven.cool project, KU Leuven studies the urban climate in Leuven in collaboration with the RMI², the city of Leuven and climate non-profit organization Leuven2030. The project aims, among other things, to map the moderating effect of green elements in the city and in citizens' gardens.

The urban microclimate is created by the interaction between the urban surface and the atmosphere. The physical processes occurring at the heart of this microclimate are the result of energy and water exchange between the urban surface and the atmosphere. The design of the urban landscape plays an important role in this phenomenon. Specifically in this researchproject, KU Leuven, together with its partners, wants to focus on the role of private gardens and public green spaces in the city.

Downloading

Since the noise-level data are limited to 2022, you can limit your research to meteo data corresponding to this period. The URL-endpoints where you can download the data (csv-format) are documented in the source (Figure 1)



Figuur 1: Downloading data for Q4 2022

All the different columns are explained in the paper

² <https://www.meteo.be/en/belgium>

Appendix B : Noise Data

This dataset has 3 different files covering noise data for Jan 1, 2022 till Dec 31, 2022:










Downloading

- **Export_40.zip**
(https://drive.google.com/file/d/1WEUXrwnU2D2gPWz4LfNFDfFaZoFMiParH/view?usp=share_link)
- **Export_41.zip**
(https://drive.google.com/file/d/1SFdfylkGIKfS1ZvcUGgR4zm130iAM9pD/view?usp=share_link)
- **Export_42.zip**
(https://drive.google.com/file/d/1jwxjOZxZQqQs-1rGIzXYIA_t-Z58Psvx/view?usp=share_link)

Each of these zipfiles can be extracted in 9 different datafiles (Fig 2). These extracted files correspond each to a particular location in neighbourhood of the Naamse straat in the city of Leuven:

1. Number 35
2. Number 57
3. Number 62
4. Number 76 ("His & Hears" Hair salon in Leuven)
5. Calvarie Chapel (near the Faculty of Economics)
6. Corner of the Parkstraat & Naamse straat
7. Number 81
8. Kiosk Stadspark
9. Vrijthof

export_40.zip 9 items

Naam	Laatst gewijzigd	Bestandsgro...
 csv_results_40_255439_mp-01-naamsestraat-35-ma... 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255440_mp-02-naamsestraat-57-xi... 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255441_mp-03-naamsestraat-62-tas... 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255442_mp-05-calvariekapel-ku-leu... 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255443_mp-06-parkstraat-2-la-filos... 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255444_mp-07-naamsestraat-81.csv 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_255445_mp-08-kiosk-stadspark.csv 24 mrt 2023	24 mrt 2023	464 bytes
 csv_results_40_280324_mp08bis---vrijthof.csv 24 mrt 2023	24 mrt 2023	1 MB
 csv_results_40_303910_mp-04-his-hears.csv 24 mrt 2023	24 mrt 2023	322 KB

Figuur 2 Extracted Files

Short Description

Export_40 contains percentiled data

Export_41 contains "Noise Events"

In this file, certain noise events are singled out. Each of these events get assigned a certainty:

- 32% that a car is passing by
- 65% that somebody is shouting
- On 15 April (16:25:48), the system picked up a human voice singing. That was a Friday evening, maybe a student happy to go home after a busy week...

Export_42 contains noise level data

On the following link (<https://www.acoustic-glossary.co.uk/definitions-1.htm>), one can find a good definition of what the columns mean:

- LAF = A weighted, sound level, measured in the files in db(A)
- Lamax = Maximum sound level