THE LAST JUDGEMENT

IMPORTANT: READ THESE INSTRUCTIONS BEFORE YOU START

- This is an alternative assessment form for this course which we haven't tried before, therefore many things are new to all of us. Adding this to the current worldwide crisis, it goes without saying that we will need (as my legal friends say) *good faith* to get through: I trust that you are going to do your best to complete this assignment by yourselves (no cheating) and you trust me that I will complete the assessment the best fitting way.
- The deadline to submit your solutions is *fixed* for 09.00 CEST (Dutch Time) 29th April.
- Make sure that you check your deliverables after you submit them (e.g. download them again and check that they are the correct ones and that they are readable).
- I will do a short technical check to your submissions *right after the deadline* but since you are a lot, I rely on you to do that check as well and of course we all rely on technology.
- I will be available via email for any technical issues (e.g. if portal is down or if your home situation suddenly changes) and any obvious misconceptions in the assignment (that I will communicate to all if needed).
- This assignment gives you maximum 50 points and is replacing your final written exam, thus contributes 50% to your final grade. The rest 50% comes from the practical grade (which you already know).
- There is a bonus question (COVID-19 version) that can give you up to 0.1 points in your final grade. You are reminded that you get 0.5 bonus for Kaggle and 0.4 for annotations.
- You do not need a complex algorithm or an algorithm beyond the scope of the course. However, any scientifically valid method or process will be accepted.
- That being said, try to limit your internet search activity to debugging (e.g. stackoverflow is a friend of ours) and instead try to get inspiration from what you have learned throughout the course.
- Follow the data analysis process and make reasonable and justified decisions on every step.
- Remember that the correct data analysis process includes verbal, scientific and visual elements. You need to include all of these in your final report.

Your deliverables (PAY ATTENTION, no .zip files, JUST the TWO files mentioned below):
a) A Python notebook (**.ipnyb**) with your code and all your work organized per question. Some documentation for the code is required with short explanations of the steps. You do not need to be detailed. Make sure you document properly code that you found in the class notebooks, your clinics and the internet.
b) A report (around 4 pages) in **PDF** where you explain your work/methodology. You are expected to use proper (English) language and it's obvious that you should include relevant plots, graphs and tables (that are produced in your notebook). Kindly use the templates I provide you (one in latex if you are comfortable with it and one in word). Obviously, the form of your report is not that important (e.g. it's fine if you don't have perfectly aligned captions, or your tables look a bit off) but try to stick to what is given in terms of structure.

The Assignment

If you think that COVID-19 was the worst thing that could have happened in 2020, then you are wrong. Today, Monday 27th at 09.00 it was announced that MASA (Maastricht Space Agency) radars discovered a small, 3 meter base meteorite, which just entered the Earth's atmosphere.

The meteorite is small and is not expected to cause big devastation, however, a governing authority who cares about their citizens has to be proactive and take early measures. Therefore, Maastricht Gemeente would like to know the location of the impact point, so that people can be warned in advance and that resources to be allocated on the population that is affected.

The Chief Scientist of MASA (let's call that person Jeronimo) has sought out the best data scientist in Limburg, that is you, to help save the day!

You are given the following three datasets:

   a) **radar.csv:** Radar position estimates (x,y,z; z being the altitude) of the meteorite at various times which are in the accompanying file. x,y and z are coordinates in kilometers (3 **last** columns respectively) and time is in seconds (**first** column). This is the main file to be used in Questions 1 & 2.
   b) **mbuildings.csv:** Locations and other details of every building in the town (to be used in Question 3)
   c) **radar2.csv:** Additional radar position estimates (to be used in Question 4). The columns follow the same order as the ones of the radar.csv file.

The Questions (in parentheses the total points)

Q1 (20). Using (any) methods you learned in this (wonderful) class estimate the expected point of impact. The impact is obviously determined when altitude of the meteorite is 0.

Hint 1: You can (and maybe you should) try different methods, before you pick the optimal framework for determining the impact point.
Hint 2: There is no best method here, so make sure that you show your work (in the notebook) and justify any selection (in the report).

Q2 (10). Jeronimo is weird and does not like results that come in the form of simple numbers without interpretation or some uncertainty, so he additionally asks you to implement the following process in order to provide the Gemeente with not just an impact point but an area that has to be evacuated:

"You create multiple (at least 1000, feel free to go for more) new datasets based on *resampling (with replacement)* from the original dataset (note that the original dataset has 130 data points). For each one of these 1000 new datasets, you pick 130 data points (from the 130 original ones) using the following process: You pick one data point randomly, then you "put it back" (that is the meaning of resampling with replacement) and you repeat the same process 129 times. Obviously, in this new dataset, some data points are going to be picked multiple times and some data points are not going to be picked at all. The ones that were picked, form your training set, on which you train your optimal framework from Question 1 and the ones that are not picked are going to be used for testing on this framework. Using all the predictions (from the multiple datasets (e.g. 1000) you created), you can compute the mean prediction, as well as a 90% range for the predicted values for x and y."

For this question you need to implement the above methodology and conclude what is the area (in terms of x,y) that will be affected with a 90% probability by the meteorite.

Hint: If you want to implement another methodology, it's fine. In any case, show and justify your work accordingly.

Q3 (10). Using the building database, estimate the total number of people that will *most likely* be affected. Use the region you defined in Question 2.

Q4 (10). Additional measurements from another radar are released in the file **radar2.csv**. The accuracy of this radar is approximately 5 times higher than the first radar. How would you modify your answer if you were to

take into account this additional information? Provide an answer for the impacted region and comment on the result.

Bonus question (2020-COVID19-version): For an extra 0.1 point on your final, design a meme (read here https://en.wikipedia.org/wiki/Internet_meme) about the course. It can be about the content (e.g. regression), the process (e.g. workload) or anything you find related (e.g. something stupid that I said during classes).

For a static meme (e.g. image) you get up to 0.05 points but if you opt for an animated meme (e.g. gif or tiktok video) you get up to 0.1 point. For some inspiration around tiktok check the following (and hopefully you can understand them):
https://twitter.com/PhDemetri/status/1250909795462365186?s=20
https://twitter.com/ChelseaParlett/status/1249390163379142658?s=20

Submit your memes separately via the special assignment on the portal.

The best memes (all of them?) will be featured (upon your permission) on Twitter and Instagram. If you are not okay with this, do let me know upon your submission.

I wish you success! Break a keyboard!