Tabular Data Science – Final Research Project Guidelines

1. The project's goal is to automate, extend, improve and/or combine selected methods/concepts covered in class (or that appear in the syllabus), **in order to assist the user in performing a better/easier data science process.**

2. The usability and effectiveness of your proposed method should be **properly tested**, and demonstrated on at least 4 different datasets. To do so, you will need to use (existing or new) quantitative evaluation metrics. The metrics will provide score for evaluating the effectiveness of your approach. You should compare it to at least one alternative approach (could be a weak, simple or naive version of your solution)
3. The project should refer, when possible, to the tools and techniques shown in class
4. The project should also refer to related academic work and existing tools
5. You will create a folder in your existing github repo, containing:
    a. A LATEX-generated PDF (up to 8 single pages, 1.5 line spacing, font size 11) that describes your results (in English), including the following sections:
        i. Abstract – a short summary of the problem, your solution, and experimental results (up to 200 words)
        ii. Problem description – what element of the DS pipeline are you trying to improve, and *why*? What are the problems it suffers from?
        iii. Solution overview: Describe your solution in detail
        iv. Experimental evaluation: Explain how you prove that your solution "works" and outperform the baseline(s) w.r.t. your suggested metric(s). Provide the details, accompanied with graphs and/or tables. You must include a comparison table or graph between your solution and baseline approach(es). Important: **Note that your grade will not be affected in the case your solution does not outperform the baselines – if you explain why and analyze it properly.**
        v. Related work – include a short discussion on relevant existing tools and techniques. *Cite the works as well as discuss them*. Did they try to solve the same problem? In what manner your solution different? Did you use/ got inspiration from it?
        vi. Conclusion – summarize your finding, and the things you learned from the project
    b. Containing all your (well-documented, organized) code
    c. It also includes an easy to use Jupyter notebook that exemplify your solution The notebook and code should run *seamlessly* on a Linux machine.
        1. Include a requirements.txt file if you use any non-ordinary libraries (i.e.,

that we didn't use in class).

2. IMPORTANT: include in the requirements.txt only non-native python libraries that we didn't use in class. A long requirements.txt that contain all your python libraries on your machine will not be used and therefore points will be deducted.

7. The grading process:
   a. Grades will be given w.r.t. the quality of each of the items mentioned above.
   b. The grade will be affected by the quality of the project proposal submitted, and the relation between the final project to it.
   c. Additional aspect of the grade is based on the refences/discussions/connections in your project to what we saw in class.
   d. Please read this document carefully to see that you understand each item. (Historically, most grade deductions are done because a team did not address one or more of the items above)
   e. If your proposed method eventually did not work so well in practice, but your project answers all items above – you will not lose points in case you explain and discuss potential reasons, as well as perform error analysis.
   f. Remember: the grade is given mainly for the effort, the quality of execution, the experiments plan, and most importantly - the way you applied/extended/inspired by the concepts and ideas learned in class.
8. This document may be further updated/changed, but no later than the proposal submission date: 12/03/25.
9. **If you have any questions about this document (not personal), please raise them *in class*, and not by mail, so everyone can hear as well.**

Good luck!!