MOTIVATION (©)



Which is the target group of our mini-project? Who is the end-user?

What are their objectives? What needs do we need to address with our work?

How will they benefit from this proposed solution?

DATA COLLECTION



Title (preliminary):

Which data sources are we planning to

Mention database tables, API methods, websites to scrape, etc.

Which is the data management plan?

PREPROCESSING X



What are the goals of the preprocessing pipeline?

Give some examples of data preprocessing steps.

What are some possible data cleaning/wrangling methods you're planning to use?

What are some possible data transformations that could be useful?

Any feature engineering necessary?

EXPLORATORY DATA ANALYSIS (EDA)

Look at the data!

What steps are you planning to take towards exploring and understanding better the data vou have?

What properties would be meaningful to summarize/visualize in this step?

VISUALIZATIONS I



List any meaningful visualizations you are planning to produce that will be useful to the end user?

Are you planning to produce any interactive visualizations?

If so, which types of interactivity might be useful to the end user?

LEARNING TASK 🐭 (focus on problem definition)

Define the problem setting.

Is this supervised / unsupervised / other ...?

Classification / regression / other...?

What are we planning to learn? E.g. What is the target variable / learning outcome?

What variables are we using as input?

LEARNING APPROACH



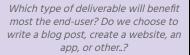
(focus on solution implementation)

Which ML/statistical methods seem more relevant for the defined problem setting and why?

Which evaluation metrics could be relevant?

Is any special treatment relevant regarding how we choose to split the data or how we cross-validate?

COMMUNICATION OF RESULTS 📢



How do we communicate best our results to the predefined target group?

Short description of your interface/workflow (if applicable).

DATA PRIVACY AND ETHICAL CONSIDERATIONS



Are there any fairness constraints that apply to our proposed pipeline?

Is there a need to ask for consent during the data collection process?

Is there a need for data pseudonymization/anonymization?

Any other privacy considerations that come to mind?

ADDED VALUE



Is there a possibility for added value from the data we're planning to use?

What is the added value?

How are predictions turned into added value for the end-user?

LEGEND

WEEK 1: Data collection/preprocessing

WEEK 2: EDA & visualizations

WEEKS 3-4: Machine/deep learning

WEEK 5: Fairness & data privacy