











<p>MOTIVATION </p> <p>Which is the target group of our mini-project? Who is the end-user?</p> <p>What are their objectives? What needs do we need to address with our work?</p> <p>How will they benefit from this proposed solution?</p>	<p>DATA COLLECTION </p> <p>Which data sources are we planning to use?</p> <p>Mention database tables, API methods, websites to scrape, etc.</p> <p>Which is the data management plan?</p>	<p>PREPROCESSING </p> <p>What are the goals of the preprocessing pipeline?</p> <p>Give some examples of data preprocessing steps.</p> <p>What are some possible data cleaning/wrangling methods you're planning to use?</p> <p>What are some possible data transformations that could be useful?</p> <p>Any feature engineering necessary?</p>	<p>EXPLORATORY DATA ANALYSIS (EDA) </p> <p>Look at the data!</p> <p>What steps are you planning to take towards exploring and understanding better the data you have?</p> <p>What properties would be meaningful to summarize/visualize in this step?</p>	<p>VISUALIZATIONS </p> <p>List any meaningful visualizations you are planning to produce that will be useful to the end user?</p> <p>Are you planning to produce any interactive visualizations?</p> <p>If so, which types of interactivity might be useful to the end user?</p>
<p>LEARNING TASK  (focus on problem definition)</p> <p>Define the problem setting.</p> <p>Is this supervised / unsupervised / other...?</p> <p>Classification / regression / other...?</p> <p>What are we planning to learn? E.g. What is the target variable / learning outcome?</p> <p>What variables are we using as input?</p>	<p>LEARNING APPROACH  (focus on solution implementation)</p> <p>Which ML/statistical methods seem more relevant for the defined problem setting and why?</p> <p>Which evaluation metrics could be relevant?</p> <p>Is any special treatment relevant regarding how we choose to split the data or how we cross-validate?</p>		<p>COMMUNICATION OF RESULTS </p> <p>Which type of deliverable will benefit most the end-user? Do we choose to write a blog post, create a website, an app, or other..?</p> <p>How do we communicate best our results to the predefined target group?</p> <p>Short description of your interface/workflow (if applicable).</p>	<p>DATA PRIVACY AND ETHICAL CONSIDERATIONS  (if applicable)</p> <p>Are there any fairness constraints that apply to our proposed pipeline?</p> <p>Is there a need to ask for consent during the data collection process?</p> <p>Is there a need for data pseudonymization/anonymization?</p> <p>Any other privacy considerations that come to mind?</p>
	<p>ADDED VALUE </p> <p>Is there a possibility for added value from the data we're planning to use?</p> <p>What is the added value?</p> <p>How are predictions turned into added value for the end-user?</p>	<p>LEGEND</p> <p>WEEK 1: Data collection/preprocessing</p> <p>WEEK 2: EDA & visualizations</p> <p>WEEKS 3-4: Machine/deep learning</p> <p>WEEK 5: Fairness & data privacy</p>		