






<p>MOTIVATION </p> <p><i>Which is the target group of our mini-project? Who is the end-user?</i></p> <p><i>What are their objectives? What needs do we need to address with our work?</i></p> <p><i>How will they benefit from this proposed solution?</i></p> <ul style="list-style-type: none"> - End users are Wolt customers - The problem is variance in the delivery times and we want to make better estimates by using additional data. 	<p>DATA COLLECTION </p> <p><i>Which data sources are we planning to use?</i></p> <p><i>Mention database tables, API methods, websites to scrape, etc.</i></p> <p><i>Which is the data management plan?</i></p> <ul style="list-style-type: none"> - Historical delivery data provided by Wolt - Historical and real-time traffic data from Digitraffic - Weather data? 	<p>PREPROCESSING </p> <p><i>What are the goals of the preprocessing pipeline?</i></p> <p><i>Give some examples of data preprocessing steps.</i></p> <p><i>What are some possible data cleaning/wrangling methods you're planning to use?</i></p> <p><i>What are some possible data transformations that could be useful?</i></p> <p><i>Any feature engineering necessary?</i></p> <ul style="list-style-type: none"> - Fill in any missing values - Check granularity of the location data and possibly group/cluster the data by area 	<p>EXPLORATORY DATA ANALYSIS (EDA) </p> <p><i>Look at the data!</i></p> <p><i>What steps are you planning to take towards exploring and understanding better the data you have?</i></p> <p><i>Plot the delivery patterns and traffic patterns on a map to see if there's some regularities (or irregularities)</i></p> <p><i>Find the the shortest paths from revenue and user coordinate points</i></p> <p><i>Merging given data from Wolt, and HRI to have the final data</i></p> <p><i>What properties would be meaningful to summarize/visualize in this step?</i></p>	<p>VISUALIZATIONS </p> <p><i>List any meaningful visualizations you are planning to produce that will be useful to the end user?</i></p> <p><i>Bar chart to compare real time delivery with estimate time of delivery with uncertainty of the estimate also shown</i></p> <p><i>A map which shows estimates where the orders will be delayed.</i></p> <p><i>Are you planning to produce any interactive visualizations?</i></p> <p><i>no</i></p> <p><i>If so, which types of interactivity might be useful to the end user?</i></p>
---	---	--	--	---

<p>LEARNING TASK  (focus on problem definition)</p> <p><i>Define the problem setting. Is this supervised / unsupervised / other...?</i></p> <p><i>Supervised learning from historical data</i></p> <p><i>Classification / regression / other...?</i></p> <p><i>Using linear regressions and mean square error as loss function</i></p> <p><i>What are we planning to learn? E.g. What is the target variable / learning outcome?</i></p> <p><i>What variables are we using as input?</i></p> <p>This would be a regression based model</p> <p>We aim to discover delivery times depending on variables such as weather and traffic</p> <p>The input would be the estimated delivery time, and pickup and delivery locations, current weather and current traffic.</p>	<p>LEARNING APPROACH  (focus on solution implementation)</p> <p><i>Which ML/statistical methods seem more relevant for the defined problem setting and why?</i></p> <p><i>Which evaluation metrics could be relevant?</i></p> <p><i>Is any special treatment relevant regarding how we choose to split the data or how we cross-validate?</i></p> <ul style="list-style-type: none"> - Uncertain about the methods so far 		<p>COMMUNICATION OF RESULTS </p> <p><i>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..?</i></p> <p><i>How do we communicate best our results to the predefined target group?</i></p> <p><i>Short description of your interface/workflow (if applicable).</i></p> <p>The main interface could be a web page, which visualizes if the delivery times are likely to be reliable currently.</p> <p>If the user enters a delivery estimate, we will make a prediction of our own estimate, and give a score how reliable the Wolt estimate is likely to be.</p>	<p>DATA PRIVACY AND ETHICAL CONSIDERATIONS  (if applicable)</p> <p><i>Are there any fairness constraints that apply to our proposed pipeline?</i></p> <p><i>Is there a need to ask for consent during the data collection process?</i></p> <p><i>Is there a need for data pseudonymization/anonymization?</i></p> <p><i>Any other privacy considerations that come to mind?</i></p> <p>Our Wolt data does not contain personal data, only coordinates of general areas although addresses maybe should be thought about.</p> <p>Traffic and weather data are also public</p> <p>Since you can't identify individuals with the data, no pseudonymization or anonymization is needed</p>
	<p>ADDED VALUE </p> <p><i>Is there a possibility for added value from the data we're planning to use?</i></p> <p><i>What is the added value?</i></p> <p><i>How are predictions turned into added value for the end-user?</i></p>	<ul style="list-style-type: none"> - Wolt customers have more reliable service and can make more informed decisions on how to use their time. - Customers would decide whether ordering food or cooking for one's self is a better alternative depending on delivery time 		<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>