











<p><b>MOTIVATION</b> </p> <p><i>The target group is real estate speculators who buy properties to renovate and sell/let them.</i></p> <p><i>The target group needs to establish which districts of Helsinki are likely to be subject to gentrification as property prices and rents rise faster in areas which are in the early stages of gentrification.</i></p> <p><i>Being able to buy and renovate properties in areas which are about to be subject to gentrification will help maximize selling/letting prices.</i></p>	<p><b>DATA COLLECTION</b> </p> <p><i>Taxable income in Helsinki by district from 1993 onwards from Helsinki city district data series (stat.hel.fi) together with district border information from Helsinki Region Infoshare service (hri.fi).</i></p> <p><i>As the income data is updated once a year, the data management plan can probably be very simple: Once a year the data will be fetched from the two services mentioned, and stored locally (wherever the single page web application is served from, c.f. "Communication of Results").</i></p>	<p><b>PREPROCESSING</b> </p> <p><i>There appears to be very little need to preprocess data as what the project is using is provided by professional data services and this shows by there being no missing or unclear values.</i></p> <p><i>Since input data consists of yearly series of taxable household income in each district of Helsinki city, it needs to be fitted with the task at hand: Computing for each district how much the yearly change (growth) of income differs from, say, mean or median yearly change provides indicators of gentrification (the higher the income growth, the higher the gentrification should be).</i></p>	<p><b>EXPLORATORY DATA ANALYSIS (EDA)</b> </p> <p><i>Plotting yearly incomes and changes of income for each district is the first step with Python Seaborn or a similar tool would be the first step towards having a better understanding of the data the project will be using.</i></p>	<p><b>VISUALIZATIONS</b> </p> <p><i>Visualizations are an integral part of the communication of results; so, perhaps this section doesn't need to be elaborated further?</i></p>
<p><b>LEARNING TASK</b>  (focus on problem definition)</p> <p><i>Given input variables of taxable income per person and per household in each district of Helsinki, which areas are likely to be subject to gentrification in the near future (i.e. now or next year)?</i></p> <p><i>Should the target variable be a real number on the interval <math>[0,1]</math>, where larger values mean a district is more likely to gentrify than districts with smaller values, this would facilitate communication of results (the target variable can simply be a coefficient for map coloring).</i></p>	<p><b>LEARNING APPROACH</b>  (focus on solution implementation)</p> <p><i>At this point in time it is not at all clear how, if at all, machine learning methods would be appropriate or used.</i></p>		<p><b>COMMUNICATION OF RESULTS</b> </p> <p><i>Since maps are easy to parse and interpret (for a human), a single-page web application with a map or two of Helsinki should make do. Maps might have each district coloured in deepening shades of, say, blue and red, the former indicating how gentrified the area is currently and the latter indicating how likely the area is to be further (or at all) gentrified in the near future.</i></p> <p><i>From the city planning point of view a map showing time series of income changes yearly might be interesting?</i></p>	<p><b>DATA PRIVACY AND ETHICAL CONSIDERATIONS</b> </p> <p><i>There are no fairness constraints this project is subject to, and there is no need to ask for consent during the data collection process as all data used is open and anonymous data.</i></p>
	<p><b>ADDED VALUE</b> </p> <p><i>Knowing which districts of Helsinki are either currently going through gentrification or are likely to do so in the future will allow optimization of real estate procurement for renovation and selling /letting.</i></p> <p><i>The city planners might benefit too (perhaps to the detriment of property speculators, should the planners be able to stop or lessen gentrification in a district where speculators have already been buying real estate).</i></p>			<p><b>LEGEND</b></p> <p><b>WEEK 1:</b> Data collection/preprocessing</p> <p><b>WEEK 2:</b> EDA &amp; visualizations</p> <p><b>WEEKS 3-4:</b> Machine/deep learning</p> <p><b>WEEK 5:</b> Fairness &amp; data privacy</p>

