MOTIVATION 🎯

Target group:
People who are interested in importing electric cars from Germany (and from other parts of EU) to Finland. Also this project benefits people who are about to buy a used electric car from Finland.

Objectives:
Car importers: To sell imported cars with maximal profit.
Car buyers: Not to pay overprice from a used electric car.

Benefits:
Car importers: Find cars that have the best resale margins.
Car buyers: More informed to negotiate about the price.

DATA COLLECTION 🧩

The data will be *scraped* from nettiauto.com and mobile.de. We will be using API's when possible

The data will be as three JSON documents - two for raw data (both platforms) and one for processed

We will be limiting the collected car data to EV's newer than October 2021 (no import tax) to simplify the import costs.

Also, we will be further limiting the collected car data to non-damaged cars.

PREPROCESSING 🛠️

The **goal** is to normalize Nettiauto and Mobile.de data. We aim to get as many relevant attributes normalized as possible.

The preprocessing starts mapping the car make, model, price, kms's... etc that same in the both platforms.

Since the prices vary a lot based on the features, we need to do a mapping table to be able to normalize them. These features include things like air conditioning, sunroof, winter tyres etc.. We will most likely do the mapping rules by hand since it's faster than implementing an automated solution.

We will not include processing the freetext right now (unless it's the only source for battery size).

Most of the cars in Finland include winter packages (tyres, heating), so we might need to come up with some solution to estimate the impact on price.

After normalizing the data, we will train a predictive model based on the Nettiauto data. This model will be used to predict Finland price for each Mobile.de listings.

EXPLORATORY DATA ANALYSIS (EDA) 🔍

Calculate summary statistics of used electric cars sold abroad and in Finland.

Study correlation of various features with asked price. Also do this analysis visually with plots.

Study pairwise and multivariate relationship of features. Use e.g. seaborn pairplot.

VISUALIZATIONS 📊

Bar charts (or other kind of useful diagrams) about price differences in Finland and abroad. Preferably at least by brand. Also the relationship of other features such as driven kms with price delta could be shown (as scatter plot for example).

LEARNING TASK 🐭
(focus on problem definition)

This is clearly a *supervised learning regression* task.

The target variable is the estimated Price in Finland based on car make, model, year, km's and features.

LEARNING APPROACH 📦
(focus on solution implementation)

A simple multivariate linear regression should yield a close enough model for the task. We might try other algorithms too.

We acknowledge the fact that some cars are more popular than others i.e have more listings. We can leverage this by constructing the test and validations sets of car models that are more popular.

COMMUNICATION OF RESULTS 📣

Ideally there would be a **website** with an interactive interface. This would allow e.g. filtering the shown data based on the user's own preferences. Data could be presented in tables and methods mentioned in cell "Visualizations".

DATA PRIVACY AND ETHICAL CONSIDERATIONS 🔐
(if applicable)

Anonymization: Remove any contact information included in the scraped sales ads. (Although, for the end user it may be beneficial to keep some identifying information to find the original ad).

ADDED VALUE 🎁

The added value is the **potential savings** when deciding which EV to buy and sell.t

LEGEND

WEEK 1: Data collection/preprocessing

WEEK 2: EDA & visualizations

WEEKS 3-4: Machine/deep learning

WEEK 5: Fairness & data privacy