[Project name]

[Fred Christiansen / Group 5], [13.11.23]

# **DESCRIBE THE PROBLEM**

## **SCOPE**

* *The goal of the project is to allow easier price predictions on houses based on a number of features in the dataset.*
* *The predicted solutions could be used to give an estimation of the value of a home which could be of use to multiple kinds of people, like tax workers, real estate agents and etc. Today there likely exists more models likely better and faster, but there is likely still some doing it manually by hand. To do pricing predictions manually you would need a team of experts that discuss the price based on the features of the house.*
* *The performance would be measured by Net promoter score and the change in cost of customer acquisition.*
* *A timeline of the project would look something like this: 1. Choose dataset and task 2. Explore the data and define the issue 3. Develop prediction models and fine tune them 4. Launch the prediction model and system. (3 and 4 would be the primary milestones)*

## **METRICS**

* *Any decrease in cost of customer acquisition would be a positive contributor for the project to be considered a success, the actual numbers would have to be based on how great the change is, and the size of the company. The second metric Net Promoter Score would also play a vital part in whether the project is a success or not as it would give a good indication of how accurate the solution is compared to real life.*
* *The project utilizes mainly MRSE and computation time to decide on the workings of the solution.*

# **DATA**

The solution ended up using 10 features, based on the correlation matrix score, OverallQual, GrLivArea, GarageCars, GarageArea, TotalBsmtSF, 1stFlrSF, FullBath, TotRmsAbvGrd, YearBuilt and YearRemodAdd. The data is provided as a part of the project issue. As mentioned the project is based on the top 10 correlating features as opposed to the 235 features in total, originally the idea was to convert text values to number features to then decide on the correlation and see if there was any major differences, however it proved rather difficult to achieve so the regular values were utilized.

# **MODELING**

The models being explored are common ones that was utilized in the previous compulsory, we decided to keep it simple, so we only utilized the linear regression and random forest models. As described earlier we utilized the correlation matrix to decide on what features we decided to make use of, in this case the top 10 results of the matrix. Furthermore, we performed cross validation and feature tuning on the selected model.

# **DEPLOYMENT**

We will be using flask, and the models will be deployed to Heroku. The predictions could be used to perform a rough, quick, and basic housing price check.

# **REFERENCES**

* <https://github.com/alu042/DAT158-2022/tree/main/a_quick_flask_tutorial>
* <https://skaliy.no/DAT158-ML-21/jupyter/2021/10/09/machine-learning-in-practice.html>