

DATA 401: Data Science
Alex Dekhtyar
Dennis Sun

Lab 2: Feature Engineering

Due: Wednesday, October 5, 1:10pm

Submission instructions.

Prepare your code (see below) as a Jupyter notebook. Please submit the notebook as a `.ipynb` file. Use the `handin` tool on CSL to submit from one of the unixNN servers:

```
unix11$ handin dekhtyar lab02-401 FeatureEngineering.ipynb
```

Assignment Specification.

For this assignment, download the Jupyter Notebook `FeatureEngineering.ipynb`, which provides a template for this assignment.

1. Select two different numerical variables as independent variables.
2. **For each independent variable**, construct the scatterplot of the car price against it.
3. Using the constructed scatterplot of the guide, create a non-linear fit model for that you consider to be “good”, and visualize it on a plot. You can use the polynomial expansion (add x^2 , x^3 , and so on to the model for the independent variable x), as well as any other functions of x .
4. Once you discover the model you consider to be the best fit (or better fit), include in your notebook a **Markdown** cell with your explanations/thoughts on why you selected this model specifically. (You can visualize other, lesser models in your notebook to make your points, if you want to).