Overview

After lots of hard work you are almost done with this data science bootcamp! Now you will put many of the pieces together and create a showcase project for your portfolio. The project will involve elements of data exploration, preprocessing, and machine learning.

The high-level goal of this project is straightforward – build a predictive model. You’ll be given some guidelines (see below), but we’ve left a lot of room for flexibility. Be creative!

The Datasets

For this project you get to choose one of four data sets. Two are for regression and two are for classification.

* [Census Income](https://archive.ics.uci.edu/ml/datasets/Census%20Income) – Predict whether an individual’s income exceeds $50K/year based on census data (classification)
* [California Housing](http://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html) – Predict media house value (regression)
* [Thyroid Disease](https://archive.ics.uci.edu/ml/datasets/Thyroid%20Disease) – Determine whether an individual has thyroid disease from various medical attributes (classification)
* [Insurance Company](https://archive.ics.uci.edu/ml/datasets/Insurance%20Company%20Benchmark%20%28COIL%202000%29) – Predict number of insurance policies (regression)