Data Scientist Coding Challenge

# Objective

Create a model that can predict the response variable y given the explanatory x variables.

# Data

We have provided a CSV file containing the y target variable and the x features. Use this to build your model.

# Model

You may use the libraries and tools that you are most comfortable with. If you use anything outside of well-known Python, R, or Scala libraries, please provide everything needed to execute your model. If we are not able to reproduce your results, your solution will not be accepted.

# Expected Output

1. Provide a write up of your approach and solution
2. The code you used to build your model
3. Instructions to run your model on a related hold out dataset

# Report Requirements

1. Any exploratory analysis you did of the data
2. An overview of the approach you used to tackle the problem
3. An overview of the algorithms you used
4. A description of the tools used to build your model
5. An evaluation of your model

# Model Evaluation

1. We will be using root mean squared error as the primary metric
2. Provide a plot showing your predictions against the actual y values
3. You may include any other metrics you feel are appropriate