Data Science Post Block Assignment 2: Task B Report

Datasets:

General comments on the dataset used.

* The dataset used is the Concrete Compressive Strength dataset from the UCI machine learning repository (<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>).
* Number of instances: 1030
* Number of Attributes: 9
* Attribute breakdown: 8 quantitative input variables, and 1 quantitative output variable.
* Missing Attribute Values: None
* The data is in raw from, not scaled
* All of the quantitative input variables are in the same unit (where applicable)
* kg in a m3 mixture
* Ager is measured in days
* The output Concrete compressive strength is measured in MegaPascal (MPa)

Possible pre-imputation cleaning steps

Scale data.