# BA711 Lab 1

## Due Date: Sunday January 16th 11:59pm

## Task

The dataset *ToyotaCorolla.csv* contains data on used cars on sale during the late summer of 2004

in the Netherlands. It has 1436 records containing details on 38 attributes, including *Price*, *Age*,

*Kilometers*, *HP*, and other specifications.

Prepare the data for use as follows:

a. The dataset has two categorical attributes, *Fuel Type* and *Color.* Use pandas methods to transform categorical data into dummy variables. (Hint: If you are getting column names for several models of Toyota Corollas, this is an indicator to specify the columns in the get\_dummies function. This is not something covered in the lecture, but the documentation should be pretty clear.)

b. Prepare the dataset (as factored into dummies) for data mining techniques of supervised

learning by creating partitions in Python. Select all the variables and use default values for the

random seed and partitioning percentages for training (50%), validation (30%), and test (20%)

sets.