DATA 621 Final Project Proposal

For the final project Critical Thinking Group 3 will use the dataset from the CoIL 2000 Challenge. It contains 86 variables on customers including product usage data and socio-demographic data derived from zip code data. The data was supplied by the Dutch data mining company Sentient Machine Research and is based on a real world business problem.

The training set contains over 5000 rows, including the information of whether or not a customer has a caravan insurance policy (prediction target). The test set contains 4000 customers. The data set is available through UCI's Machine Learning Repository at:
https://archive.ics.uci.edu/ml/datasets/Insurance+Company+Benchmark+%28COIL+2000%29.