Engineering mixed variable

**Engineering mixed variables**

We've seen that mixed variables are those which values contain **both numbers and labels.** A mixed variable can contain **numbers OR labels** in different observations, or **numbers AND labels** in each observation.

For example, the variable vehicle registration is an example of letters and numbers combined in each observation (**e.g., NK11DGX)**, whereas the variable number of missed payments as we covered in section 2, is an example of numbers or labels, where the values can **be 0, 1, 2 or 3, or A and D.**

See section 2 on Variable types, lectures on mixed variables for a refresher.

**How can we engineer mixed variables?**

We can sometimes **derive meaningful features by extracting the categorical part in one variable** and the **numerical part in a different variable**. Therefore, we obtain 2 variables from the original one. Then we proceed with the derived variables as we would do with any categorical or numerical variable.

Let's explore how we can do that.













