**Lab Session 4 Customer Analytics – Questions after Lab**

***Questions received via email***

1. [Konstantinos];
2. It kind of confuses me why we need to use as.factor() in an already binary variable. For instance, I have noticed that under glmnet() and tree() you use as.factor() for the Churn variable, while under glm(), cv.glmnet(), ranger() you take Churn naturally.
3. How do we calculate that label='yprob'?
4. [Noa];
5. Random Forests – variable importance: I do not understand the values of the following table

Graphical user interface, text, application, email

Description automatically generated

1. Can you go over the comprehension check questions and share the code so we can check our own answers?
2. [Mieke];
3. Why do we use lambda.min and what exactly are we implying here?

pred<-predict(lasso\_cv, newx=x, s = "lambda.min", type = "response")

1. When determining how many non-zero coefficient there are in a model, do you also have to count the intercept if it is non-zero (since it is also a coefficient)? Or should you look at the plot and check which value is above the vertical line?
2. If we want to prune a tree, do we always have to start with the most complex tree and prune that one or are we supposed to prune the tree we already have from preceding questions?