|  |  |
| --- | --- |
| The Lisbon MBA Catolica | Nova - Business school rankings from the  Financial Times - FT.com |  |

3. Predicting Mobile Communications Consumer Churn

File: nextcarrier.csv

PART 1 – Data Cleaning

1. Load the dataset onto RapidMiner
2. Browse the dataset. Does everything seem generally ok with it?
3. Analyze the dataset via the Statistics operator. Do you see anything interesting in terms of customers’ age attribute?
4. Let’s perform the data cleaning steps required to run the logistic regression. A) Change the type of the churn variable to binomial, then set its role to the label variable. Next, remove the attributes churn\_dum1 and churn\_dum2. Finally, transform the age variable to take into account the issue detected in point 3 above. Paste a screenshot of the resulting process below (pc: windows key+shift+s; mac: shift+cmd+4)
5. Move all of the data cleaning steps into a subprocess. You can do this by selecting all of the operators and clicking the right mouse button. Select the “Move into new subprocess” option.

PART 2 – Analysis

1. Add the Logistic Regression operator. Connect the Model output to the Res input to see the estimated coefficients.
2. Use the left “Description” option under the results tab to access additional information. Find the MSE, RMSE, R-squared.
3. Go back to the design view. Use the Apply Model operator to add a prediction of the labeled variable, “churn”. A) Connect the **model** output of the logistic regression to the Apply Model operator. Do the same with the **example output.** Then, connect the outputs of the Apply Model operator to the res inputs. Can you tell how the prediction(churn) column was calculated?