Steps of the Group Project:

The group project follows the same approach as we have done in the classes, but now it's time for you and your group to do it for the **same dataset but for a different target**. To help you, below is a list of the steps needed for the project, with some suggestions.

1st – Data pre-processing

I advise you to create a new diagram for the project.

Use the following steps:

File Import – Now we want **Nutriappoint** as the target. You may copy and paste everything from the file import until transform variables. Still, you need to be aware **that Nutriappoint must be the target and Totcheck rejected**, as shown in the image below:

Nutriappoint	Target	Binary	No		No
Perdeal	Input	Interval	No		No
Recency	Input	Interval	No		No
Skin	Input	Interval	No		No
Teenhome	Input	Binary	No		No
Totcheck	Rejected	Binary	No		Yes
WebPurchase	Input	Interval	No		No

Multiplot; Graph explore; Statexplore – Although the dataset is the same, you should see if there are new visual outcomes of having Nutriappoint as the target. In the report, you should put the main conclusions of this visual exploration (e.g., older clients are more likely to buy the product – with the graph that proves it).

Filter; Replacement; Impute – Same as the class, as the dataset is the same. You can copy what you have done in class and discuss the best approach with your group. State your decisions and why on the report!

Transform Variables – In the class, we saw some examples, but now it's your turn to be creative and create new variables that you think are interesting.

2nd – Clustering – done directly on your project!

Cluster – Done in class! You need to name the clusters. We will use K=3.

Segment Profile– You need to interpret the graphs and characterize the segments you found.

3rd – Prediction

Data Partition – Split the data into 70% to train and 30% for validation.

Decision Tree; Regression; Neural Network – Apply the same as in class but for the new target. You may try new approaches, such as different criteria or the number of hidden units, layers, branches etc. In the report, you must state your interpretation of the outputs/graphs of the models, the same ones we have seen in class.

Model Comparison – Use the ROC chart and cumulative lift to select the best model. State your interpretation of the report and which model you have chosen.

File Import – Import the new dataset for scoring.

Score – Score the new dataset with the model you have discovered to be the best on model comparison.

Structure for the report:

A Report in PDF.

1. **Cover page** – Title, date, and authors.
2. **Executive Summary/ Abstract** - It should not exceed a paragraph (100/200 words)
3. **Introduction** - It should describe the rationale for undertaking the work reported on, the question you are trying to answer, its context (brief), and how you will approach the brief.
4. **Methodology** - A part concerning the methods, describing the data, analysis, steps taken, and why. Please keep it simple; try to use prints, graphs, and tables to help the reader!
5. **Results and Conclusions** – Main conclusions of the work done, limitations, suggestions for future work, suggestions of approaches to clients etc.
6. **References**.
7. **Appendices containing detailed information** (e.g., graphs and tables that are not of central importance for the conclusions of the project)