# Prescriptive Analytics- Assignment 3 (2022)

This assignment is a group assignment, and you should only collaborate with the students within your group. Do not copy or share code from/to other groups. The same holds for your report. Collaborate solely with your group members.

Hand in a report in .pdf on Canvas containing your answers to the questions and submit your R-code in a separate file. Do not put them together in a zip-file but upload the documents separately. You will mainly be graded based on your report. Make sure that it is understandable what you did without consulting the R code. Please make sure to explain and justify all of your answers. Your reasoning and interpretation of the results is as important for your grade as the technical correctness of your analysis! Answer the questions below in the main body of your report. Make sure you use a random seed to be able to reproduce your results.

The Deadline is July 7th, 2022. To pass this course you need at least a grade of 4.5 for this assignment. Thus, if your grade is lower than 4.5, you must go for the improvement option to pass the course. If your grade for this assignment is lower than 5.5 (but higher than 4.5), you have the option to improve your assignment. The deadline for handing in the improvement option is 7 days after the grades of this assignment are released.

# Promotion folder

You are asked to develop a promotion folder. Given the limited space in the folder (40 pages), you need to determine which items to put in the folder. You can choose from 200 items, in 10 categories, with 20 different items in each category. You want to select those items that fit into the folder such that the total revenue that is realized by the folder is maximized. The revenue corresponding to selecting an item for the folder does not only depend on the item itself, but also on the other items that you choose to display in your folder. For some pairs of items there is a positive effect of including them together in the folder, these items are complementary (e.g. hamburgers and hamburger buns). For other combinations of items, items from the same category, including them together has a negative effect on the revenue (e.g. Pepsi and Coca Cola).

Selecting items for this promotion folder corresponds to solving a quadratic knapsack problem, the knapsack problem with a quadratic objective function:

$$max \sum_{i=1}^{200} a_i x_i + \sum_{i=1}^{200} \sum_{j=i+1}^{200} b_{ij} x_i x_j$$

$$\sum_{i=1}^{200} x_i \leq 40$$

$$x_i \in \{0,1\} \, for \, i = 1, \ldots, 200$$

trường hợp item 1 được chọn : a1*11 + b12*11*11
0.265 +

Here, $x_i$ represents whether item $i$ is included in the folder, $a_i$ is the revenue (in thousands of Euros) to be earned by including this item and $b_{ij}$ is the additional revenue (in thousands of Euros) to be achieved when both item $i$ and item $j$ are included in the folder (which can be negative). Because we cannot use an LP solver for the quadratic knapsack problem, you are going to use heuristics to solve this problem.

**Question 1 [65/100]**

We want to find good solutions by trying different heuristics for the presented quadratic knapsack problem. The coefficients of the objective function are provided in the dataset workspace_coefficients.RData: the matrix 'value' contains on the diagonal the $a_i$ values and on the off-diagonals the $b_{ij}$ values.

a. [20pts] Please implement a heuristic that starts with an initial solution where 40 items are randomly selected to be included in the folder. Then, the heuristic swaps in each iteration the values of the decision variable for two of the items. For instance, if $x_2 = 1$ and $x_{36} = 0$, a swap between item 2 and item 36 would result in the following values: $x_2 = 0$ and $x_{36} = 1$. Perform 7.000 iterations of swapping. Provide the heuristic steps, the best solution, the total revenue with the initial solution and the total revenue with the best solution in the report.

b. [10pts] Now consider the same heuristic structure with 7.000 iterations. Instead of swapping two random items, remove the item with the lowest $a_i$ and add a random item instead. Does your solution improve? Why or why not? Provide the heuristic steps, the best solution and the total revenue in the report.

c. [15pts] Now come up with your own improved heuristic with the same number of iterations. Please provide the heuristic steps (in pseudocode) in the report, together with the best solution and the total revenue.

d. [20pts] So far, we have assumed that each item takes up exactly one page in the folder. In practice, it is possible to include an item with 1 page or with 2 pages in the folder. How would this change your model? Would the optimal solution value increase or decrease with this change? Change your heuristic of question 1.c according to this adaptation and perform 7.000 iterations. Please provide the heuristic steps (in pseudocode) in the report, together with the best solution and the total revenue. *Note: If you did not manage to solve question 1.c, you can use the heuristic of 1.a instead.*

**Question 2 [35/100]**

Of course it cannot be precisely known how much revenue including an item in the folder will generate. Now assume that the revenue that you obtain with the different items is stochastic. You have information on the distribution of these revenue values, but you don't know the specific coefficient values in the objective function. In this exercise you will apply Monte Carlo simulation to take this stochasticity into account in your solution approach. Assume that the coefficients follow a triangular distribution. The coefficients used in Question 1, correspond to the most likely realizations of this distribution. Now you get 3 matrices for the coefficients, corresponding to the minimum, the maximum and the most likely value of the distribution for each coefficient (matrices 'value_min', 'value_max' and 'value_mode').

a. [35pts] Use the same heuristic as in question 1a., but now instead of using the most likely coefficients when evaluating the objective value, you will perform Monte Carlo simulation to evaluate the average objective value over 100 realizations of the coefficients. You will evaluate each solution by looking at the average objective value over the different realizations instead of looking at the objective value obtained with the most likely coefficients. You can limit the number of heuristic iterations to 200. Compare the solution from question 1a. to the new solution. You can evaluate both solutions by calculating the average objective values over 100 realizations.