

DATA VIZ 2025 Strasbourg Take Home

Consider the context described in this kaggle dataset, which you also find in the folder.

<https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign>

The Objective (From Kaggle)

The objective of a marketing team is to build a predictive model that will produce the highest profit for the next direct marketing campaign, scheduled for the next month. The new campaign, sixth, aims at selling a new gadget to the customers. To build the model, a pilot campaign involving 2.240 customers was carried out. The customers were selected at random and contacted by phone regarding the acquisition of the gadget. During the following months, customers who bought the offer were properly labeled. The total cost of the sample campaign was 6.720MU (monetary unit) and the revenue generated by the customers who accepted the offer was 3.674MU. Globally the campaign had a profit of -3.046MU. The success rate of the campaign was 15%. The objective is of the team is to develop a model that predicts customer behavior and to apply it to the rest of the customer base. Hopefully the model will allow the company to cherry pick the customers that are most likely to purchase the offer while leaving out the non-respondents, making the next campaign highly profitable. Moreover, other than maximizing the profit of the campaign, the CMO is interested in understanding to study the characteristic features of those customers who are willing to buy the gadget.

Todos

1. Analyze the data and provide insights with key descriptive statistics. 20 points
2. Propose a market segmentation, using Self Organizing Maps (Kohonen Maps). 20 points
3. Compare two different prediction models for the variable response. Use different graphical techniques to compare the model. 30 points.
4. Compare the two models with the use of the profit curve and answer the following question: what is the best percentage of customers to be contacted to maximize profits? 30 points.

HELP:

Target variable: Response

Cost per contact variable: Z_CostContact

Revenue per positive response variable: Z_Revenue

OUTPUT:

1. File/notebook python/R with the analysis.
2. Business Presentation in PDF with nice tables and graphs answering the 4 points. I will evaluate ONLY the Business Presentation.

EVALUATION: I will evaluate the completeness and clarity of the answers (50%), but also the quality of the graphs (overall aesthetic, use of meaningful channels and appropriate for both attributes and clarity, right balance of the ink/space ratio, good level of self-explainability of the graph...) (50%).

Pay attention to: clarity, information content, ability to convey a message.

BONUS: 10 extra points for the most astonishing VIZ!

File naming: yourteam.pdf

DEADLINE: 20.02.2025

SUBMISSION:

EMAIL to marco.guerzoni@unimib.it with subject: DATAVIZ PROJECT.