GUIDELINES FOR THE DIGITAL MARKETING DATASCIENCE PROJECT:

The purpose of the project is to **present a data-driven communication strategy based on machine learning models**.
You should think at the project in these terms:

You are a data scientist in a retail company. The marketing team wants to increase the return from the digital marketing investments. To do so the Chief of Marketing asks you about those data-driven approaches everybody is talking about. The idea is to address the churning of high-valuable customers.

- The project consists in
    - **a brief business-oriented presentation** (10-15 slides) **of the logic developed**
        - An explanation of the business questions approached
        - A brief overview of the preliminary analyses performed
        - The models chosen, the modelling structures and the algorithm performances
        - Examples of data-driven actions that can be taken using the models
    - **the R script used for the modelling**.
        - The code submitted must run correctly, be replicable and contain a clear and commented exposition of each pipeline step
        - The code must include the commands for loading the necessary packages
        - The pseudorandom generator must be fixed within the script using set.seed(12345)
        - Together with the script, it must be provided a .txt file contained the info on the R session for testing the replicability obtained using the function sessionInfo()
- The project is done individually
- The datasets necessary for the project are available on google drive and are explored during the lectures
    - Part1 https://drive.google.com/open?id=1gPiKUzRUHLbXLiWe2Fe6O91V86ZE2WMX
    - Part2 https://drive.google.com/open?id=1MTrcTzndYhuUOFtaDNk-jWC8UmGdbo3V
- The students are required **to complete the R script for the RFM and CHURN model development started during the lessons** and **to add some personal perspective in the presentation**.
- The final marking is based on:
    - Correct and well-structured R code
    - Right balance of presenting quantitative results avoiding unnecessary technicalities
    - Clear identification of the key-points of the business questions
    - Personal interpretation of the preliminary analyses
    - Original ideas for data-driven actions
    - A creative presentation layout

    A work that covers brilliantly these key points is entitled for the maximum mark

- Optionally a student can develop autonomously one of the following models:
    - Simulation modelling: Buy Till You Die
      (reference https://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf
    - Time series modelling: Purchase Trends Forecast
      (reference https://otexts.com/fpp2/arima-r.html)
    - Probabilistic modelling: Market Basket Analysis for the 100 most purchased items
      (reference https://datascienceplus.com/a-gentle-introduction-on-market-basket-analysis%E2%80%8A-%E2%80%8Aassociation-rules/)

  Completing an additional model will be considered for extra points in the final marking.