

# Data Science Challenge

This folder contains a time-series data set saved in `transactions.csv`, which contains transaction data of a fictive beverage distributor between January 2022 and June 2024.

## Exploratory Data Analysis

Before starting with modelling experiments, it is important to perform exploratory data analysis (EDA) in order to get familiar with the data and the general patterns it contains.

- Create a basic notebook to explore and visualize the data -- what are your main initial observations?
- What are the top 5 most popular products by revenue?
- How is revenue distributed across different customer segments?
- Can you detect any outliers in the price column? How would you handle these outliers?
- What seasonal patterns can you identify and what might be the reason for this seasonality?
- Have there been any significant shifts in consumption affecting the business in 2023 vs 2022 and how?

N.B: The reviewers are not looking for a specific answer here; just for the ability to explore data, come up with hypotheses and test them appropriately.

## State calculation – practice

### Background

MYWAY is generating next best actions for sales reps. To generate these actions, we need to calculate the current state of each client, defining the situation the client is currently in. Some example states are:

- The client has stopped ordering Coca-Cola PET 1L
- The client has a high churn risk
- There is a new relevant product for the client

Depending on the state, we can then recommend the next best action. As you can see from the examples above, the state calculation requires various models, including anomaly detection, product recommendations, or churn prediction.

### Exercise

For each client, calculate the following states as of June 2024:

1. Dropped products [Product name, Total lost revenue, Total lost quantity]
2. Churn risk [Churn risk, Revenue at risk]

Note: Keep in mind you are trying to capture a potentially very weak signal when it comes to the churn risk. When calculating the revenue and quantity numbers, make sure to clearly state your reference period in the code.

Based on the results, create a brief summary report addressing the following:

- Describe the experiments you ran.
- What is the total revenue impact associated with the states?
- What additional steps would you take to improve the performance of your state calculation?
- What problems did you encounter when calculating the churn risk? How would you address them?
- Do you consider your approach scalable and easily adaptable for other MYWAY clients?
- Based on your exploratory data analysis and the trained model(s), would you consider this distributor a good potential client for MYWAY? Why or why not?

You are free to choose whatever type(s) of model you want to experiment with.

Note: The solution does not necessarily need to formulate model training and inference in the optimal way, as long as the assumptions and limitations of the methods used are clearly stated in the submission.

## Personalizing Recommendations for Individual Reps – practical theory

Answers to this section should be text-based only, but feel free to include diagrams or sketches to illustrate your explanations. Your answers should be in the form of a proposal for one or more of experiments that would implement a method or technique that would try to address the issue presented in the question. Keep in mind that we only have limited resources and that we need to be able to delivery this feature in an adequate time and manner.

### Problem Context

Each sales rep may have their own strengths, preferences, and working style. Some may excel at selling certain product categories, while others may be more skilled at negotiating with specific types of customers. To maximize the effectiveness of the next best actions, it's important to personalize them to each rep's unique profile.

### Assignment Question

Describe how you would personalize NBA recommendations for individual sales reps. What data would you need to collect on the reps? What machine learning techniques would you use to tailor the recommendations? How would you ensure that the recommendations not only align with the rep's strengths but also push them to improve in areas where they may be weaker? Explain how you would measure the success of personalized recommendations and iterate on the model to continuously improve it.

Your submission should be a compressed file containing:

- a notebook (or script) with your code and text-based solutions (this can be in a separate txt or md file)
- a [README.md](#) file with brief instructions on how to install dependencies and reproduce experiments.
- a [product\\_states.csv](#) file containing the calculated product states for all clients and a [churn\\_states.csv](#) file containing the calculated churn states for all clients