

SENIOR DATA SCIENTIST CHALLENGE

Welcome to the Neural Design challenge for Data Scientists designed for Influur. From your CV/LinkedIn profile, we know that you say that you love data and machine learning.

In this challenge, we want to test your ability to: make decisions regarding what to do and how to do it, not only in terms of machine learning and all the technical topics, but also in terms of impacting the business!

Imagine you're already a crucial part of Influur's team and that many business decisions depend on the input you provide.

Please follow the instructions:

- Download the following files:

[Sales Detail](#): Detail of the sales of an online retailer

[wide_mtrx](#): Aggregated sales (in USD) by customer and month

[wide_qtrx](#): Aggregated sales (in Transactions) by customer and month
(we created the last 2 files to make the analysis easier)

The challenge is to:

- EDA: Explore the datasets and present a thorough analysis. Don't hesitate in doing some preprocessing of the data to generate more valuable insights.
- Propose a categorization of the different customers (most valuable customers, most frequent buyers, ...)
- Build a model that is capable of identifying clients with higher probability to buy in the next month (based only on the sales of windows of information including the last 6 months)
- Put this model into production (locally or preferably in the cloud)
- Explain how good your model is, which are the most important features, and how the values of the variables influence in the model
- Explain how this model could be used to generate value
- Push the project to a repo and share this with us. Include instructions and the packages to make the system replicable



- Share your most valuable insights of the data and the evidence of the model in production

You won't be evaluated by how good the model performs. Don't waste your time there.

Influur's and Neural Design's teams wish you the best of luck!