

# CHALLENGE

## Customers' income

Figuring out who are the wealthy customers is any company's dream. The "whales" have more to spend and are arguably less discount affine. This challenge will use Radom Forest to predict which customer

### PREPARE DATA SET

# 01

Remove variables that don't make sense to be there, and transform some that need transformation (e.g., Birth Year). Additionally, make sure you only have numeric variables and remove NA's. Finally create a Training and Test set

---



# 02

### RANDOM FOREST MODEL

You prepared everything and now you are ready to create your random forest model. Is your dependent variable continuous or binary? That will impact the type of Random Forest model

---

# 03

### ACCURACY ASSESSMENT

How good is your model? Predicting income is an important task and thus being good at it is pivotal. Of course, it is difficult to assess a model based on just one iteration, but you will have a benchmark for the future

---



# 04

### MODEL TUNING

Build a process to tune the optimal number of trees. This is an introduction to parameter tuning. The goal is for to have the skill to understand advanced analytical projects.

---



# 05

### GENERATE INSIGHTS

Explainable Artificial Intelligent (XAI) is one of the most exciting topics in analytics. The idea is to extract insights from advanced Machine Learning models, which are traditionally black boxes.

---



Diogo Resende

