Data Team
Data Scientist
Task

Home task

To complete this assignment you will need the publicly available dataset of Online Retail provided by the UCI, which you can find at http://archive.ics.uci.edu/ml/datasets/Online+Retail+II .
The dataset contains transactions (including cancellations) carried out by an online retailer (mainly focused on gift items) between 01/12/2009 and 09/12/2011.

Please complete the tasks below and provide a clear answer, including code and a walkthrough of the reasoning/ conclusions of each step below, preferably in runnable notebook format (.html and .ipynb — Python or Scala are recommended, but feel free to use whatever language and technology you're comfortable with).

In the notebook you should:

  1. Perform a detailed EDA of the dataset, providing the most relevant business insights.
  2. Split the customers into groups according to their purchase patterns and product purchases, and characterise/ quantify the obtained customer personas.
  3. Have any users churned during the second half of 2011? Are there any users in high risk of churning by the end of 2011?

Let's build the future together!
Thank you!

Good luck!