

Data Scientist Case Brief

Customer Segmentation Challenge

Description

With the data provided in the file `customer_data_sample.csv` using any method you deem fit, answer the question:

"What are the most important factors for predicting whether a customer has converted or not?"

Converted customer is represented in the data in the field "converted", and the nature of what this conversion means is (intentionally) unknown in the context of the challenge.

Fields

field	explanation
customer_id	Numeric id for a customer
converted	Whether a customer converted to the product (1) or not (0)
customer_segment	Numeric id of a customer segment the customer belongs to
gender	Customer gender
age	Customer age
related_customers	Numeric - number of people who are related to the customer
family_size	Numeric - size of family members
initial_fee_level	Initial services fee level the customer is enrolled to
credit_account_id	Identifier (hash) for the customer credit account. If customer has none, they are shown as "9b2d5b4678781e53038e91ea5324530a03f27dc1d0e5f6c9bc9d493a23be9de0"
branch	Which branch the customer mainly is associated with

Submission requirements

Submit your work as a git repository (preferred way):

Via git (github or bitbucket):

Submit your answer as a version controlled (git) repository (repo) in github or bitbucket. Make sure your repo is public and submit a link to it via email.

Suggested tools / approaches

- Use summary statistics, visualization or other analytical means to explain your argumentation - it's important that you coherently explain, why you deem certain factors important and why some might be considered more important than others
- You can for example use ipython (jupyter) notebooks or such
- Remember to include your full answer and used visualizations (code and pdfs) in your submission

Reach out to adithya@lenus.io if you have any questions regarding the case brief or dataset.