

Take-Home Challenge: Data Scientist, Product Analytics

Cleo is a personal financial assistant. Users can interact with Cleo to obtain insights into their spending habits and to work towards improving their financial health. We want to understand what leads to users engaging more regularly with the service and being retained over time.

Instructions

You are a data scientist tasked with helping the product team to better understand engagement and retention of Cleo users. The team asks for your help with the following:

1. What key metrics would you propose to monitor over time to measure the success of the team's efforts in improving user engagement/retention and why? Clearly define your metric(s) and explain how each is computed.
2. Can you help us identify which factors are associated with better engagement/retention? What segments are doing well and what could be improved? Can you propose any recommendations (business initiatives or product changes) that could address these opportunities?
3. There is also interest from the CEO about the work you are doing, and a desire to understand the broader framing of the challenge of improving engagement/retention, thinking beyond the data provided. What other research, experiments, or approaches could help the company get more clarity on the problem?

Summarise your findings and recommendations in response to the questions above in a 5-8 slide presentation intended for the Head of Product at Cleo. Please send this as a PDF, Keynote, Google Slides, or PowerPoint document.

Include an organized appendix sharing the details of your analysis that would be useful for the data team to understand your work. We typically see data processed in SQL/Python/R, but you are welcome to use any software you feel comfortable with. If you use Excel, please document the operations used to process the data, and append your spreadsheet.

We suggest spending 2-3 hours on this. We will grade the assignment with the time you had to complete it in mind. Please submit your output within 72 hours of receiving the challenge.



Data

A sample of data is provided in two attached CSV files:

1. cleo_users.csv
2. cleo_user_activity.csv

1. cleo_users

A user table with data on users who joined Cleo between January and August 2019. This table includes:

- id: the user's id
- created_at: date when the user created a Cleo account
- user_country: country of the user. This takes one of three values:
 - 'US' – USA
 - 'CA' – Canada
 - 'GB' – United Kingdom
- age: age of the user
- campaign_source: the marketing channel through which the user came to the product. This takes one of four values:
 - 'Facebook Ads: General' – Facebook ads on general Cleo product
 - 'Facebook Ads: Overdraft Fees' – Facebook ads specifically addressing avoiding overdraft fees
 - 'Influencer' – Promotion by a social media influencer
 - 'Friend Referral' – Referred by a friend (who received £/\$5 for the referral)
- device: the mobile device of the user. This takes one of three values:
 - 'ios' – Cleo app on iOS phone
 - 'android' – Cleo app on Android phone
 - 'messenger' – Facebook messenger
- budget_created_at: the date the user first set up a budget. This will be NULL if the user has never set up a budget.
- autosave_created_at: the date the user first set up autosaving with Cleo. This will be NULL if the user has never set up autosaving.
- subscription_created_at: the date the user upgraded to a premium Cleo Plus subscription. This will be NULL if the user has never upgraded.

2. cleo_user_activity

A usage summary table that has a row for each day a user logged into the product.

- user_id: the user's id
- active_date: date user logged into Cleo

