

## **Challenge - Data Scientist - Product**

Congratulations for making it to the Challenge stage of the application process! The goal of this task is to objectively assess your technical prowess and your ability to solve business problems using data exploration, statistical analysis and potentially machine learning. Please note that the topic, data and problem set are reflective of cases we solved in the past.

**Deadline:** 7 Days

### **Deliverable: REQUIRED**

Your submission should be in a .zip file with contents grouped into the following sub-directories (you may omit empty directories):

- data (only include new data files, exclude raw data)
- notebooks (include .html format of notebook here as well)
- artifacts
- code
- misc

Please label the .zip file: `<your_first_name>_<your_last_name>_ht.zip`.

For example: `aubrey_graham_ht.zip`.

*Before proceeding to the task, please go through the README.md file and complete all steps included in the preparation section. We have also included additional information and tips to help you! The data provided is synthetic and a simplified form of our real data.*

## **The Task**

### PART I - SQL

We expect you to solve the following questions using SQL. Please, comment your queries and explain your findings.

- a) Daily sign-ups is an important metric to quantify the success of growth strategies and campaigns led by our country managers. **Write a query to calculate the mean, the standard deviation and the median of weekly sign-ups per country.**

- b) We care about growth and also if users are engaged with our product. One of the many ways to track if they are “active” is from their card payment transactions.  
**Write a query to calculate the volume of monthly card payment in USD per age group in the interval of 10?**
- c) We are interested in the spending patterns of non standard vs standard users.  
**Write a query to show the number of days where the average card payments volume of non standard users is less than standard users.**

## PART II - DATA EXPLORATION & ANALYSIS

The dataset provided consists of examples of data points we often use to solve business problems. Data Scientists who sit with our product teams are typically tasked with understanding our user behaviour. Usually, we are concerned with **engagement** and **churn**. Utilizing this data, your business knowledge and potentially your interests, answer the following questions.

- a) Define a target metric to measure user engagement. How would you define an *engaged* vs. *unengaged* user?
- b) Using your logic from above, build a model (heuristic/statistical/ML) to classify *engaged* and *unengaged* users. Note that features which are directly correlated with your target metric could lead to overfitting.
- c) Let's assume an *unengaged* user is a churned user. Now suppose we use your model to identify unengaged users and implement some business actions try to convert them to engaged users (commonly known as reducing churn). How would you set up a test/experiment to check whether we are actually reducing churn?
- d) In the past, one business action we took to reduce churn was to re-engage inactive users. Our engagement team designed a marketing campaign, **REENGAGEMENT\_ACTIVE\_FUNDS** to remind inactive users about funds on their Revolut account. Define a metric to measure the effectiveness of the campaign. Under this metric, was the campaign effective?

**Guidelines:**

1. We expect you to be creative and inquisitive, so tell us a story!
2. Some skills we look for are:
  - a. data preparation, exploration, visualization
  - b. target selection (labelling, defining success metrics)
  - c. modeling (statistical tests or machine learning)
  - d. code (clean, commented, pythonic)
3. You should use Jupyter (or R for some exceptional cases)
4. You may use any packages to aid your work
5. Assumptions can be made but should be stated and backed up with data where possible
6. Play to your strengths; spend more time on visualization, statistics, machine learning, product ideas, or business insights depending on your skills
7. Keep it simple; complex solutions will require us to ask you complex questions during the interview