Challenge - Data Scientist - Product

Introduction

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 are reflective of cases we have solved in the past.

Deadline: 7 Days
Deliverable:

- 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 (please include .html format of notebook here as well)
    - code (in case you have some Python modules and/or helpers functions)
- Please label the .zip file: <your_first_name>_<your_last_name>_ht.zip., For example: aubrey_graham_ht.zip

---

The Task

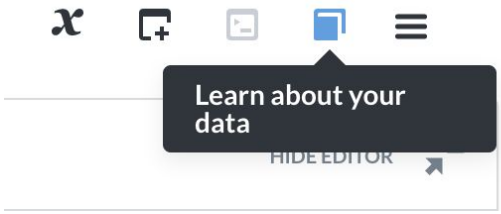
The datasets contains examples of what we often use to solve business problems. Data Scientists who work on product teams are tasked with quantifying success and understanding how to drive teams to success. One of the largest areas we are concerned with is **engagement** and **churn**.

Utilizing this data, your business knowledge and potentially your interests, please accomplish the following tasks:

1. Please write a query and create a visualisation to answer the following question:
    a. For all users that received a notification, what is the difference in average transactions 7 days before the notification arrived vs. 7 days after the notification arrived, by country and age group?
    b. How to execute:
        i. Please navigate to https://bi-challenge.revolutlabs.com and log in using the credentials we have sent to you

1. This is Metabase, an open source reporting tool. It is a simple and free alternative to tools like Tableau, Looker, Microstrategy, PowerBI, etc
2. If you feel confused, you can read their getting started guide here: https://metabase.com/docs/latest/getting-started.html.

ii. In the *Data Science* database, we have pre-loaded tables with synthetic data inspired by real-life data we from our day-to-day. You will find the full description of the table names and variables at the end of the challenge. You can also view the database tables and schema in Metabase (see screenshot below)



iii. You can use Metabase's SQL interface to test SQL and generate different chart types

iv. Please create a new "question", and after you are satisfied with your query and visualisation please save your question as "ds_challenge-q1" in your Personal collection

v. Feel free to save more questions and play around with Metabase if you want to, but only what is included in "ds_challenge-q1" will be considered for assessment

2. Define a target metric to measure user engagement. How would you define an *engaged* vs. *unengaged* user?
    a. Please provide the business justification and associated visualisations / rationale in choosing your definition of engagement
3. Using your logic from above, build a model (heuristic/statistical/ML) to classify *engaged* and *unengaged* users
    a. Note that features which are directly correlated with your target metric could lead to overfitting
4. 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)
    a. How would you set up a test/experiment to check whether we are actually reducing churn?
    b. What metrics and techniques would you use to assess the impact of the business action?

***NB** - Question 1 has to be done on Metabase and questions 2-3-4 have to be done in a Jupyter Notebook using the datasets provided in the 'data' folder as CSV files.*

The Datasets

1. devices.csv
a table of devices associated with a user

- **brand**: string corresponding to the phone brand
- **user_id**: string uniquely identifying the user

2. users.csv
a table of user data

- **user_id**: string uniquely identifying the user
- **birth_year**: integer corresponding to the user's birth year
- **country**: two letter string corresponding to the user's country of residence
- **city**: two string corresponding to the user's city of residence
- **created_date**: datetime corresponding to the user's created date
- **user_settings_crypto_unlocked**: integer indicating if the user has unlocked the crypto currencies in the app
- **plan**: string indicating on which plan the user is on
- **attributes_notifications_marketing_push**: float indicating if the user has accepted to receive marketing push notifications
- **attributes_notifications_marketing_email**: float indicating if the user has accepted to receive marketing email notifications
- **num_contacts**: integer corresponding to the number of contacts the user has on Revolut
- **num_referrals**: integer corresponding to the number of users referred by the selected user
- **num_successful_referrals**: integer corresponding to the number of users successfully referred by the selected user (successfully means users who have actually installed the app and are able to use the product)

3. notifications.csv
a table of notifications that a user has received

- **reason**: string indicating the purpose of the notification
- **channel**: string indicating how the user has been notified
- **status**: string indicating the status of the notification
- **user_id**: string uniquely identifying the user
- **created_date**: datetime indicating when the notification has been sent

4. transactions.csv
a table with transactions that a user made

- **transaction_id**: string uniquely identifying the transaction
- **transactions_type**: string indicating the type of the transaction
- **transactions_currency**: string indicating the currency of the transaction
- **amount_usd**: float corresponding to the transaction amount in USD

- **transactions_state**: string indicating the state of a transaction
    - COMPLETED - the transaction was completed and the user's balance was changed
    - DECLINED/FAILED - the transaction was declined for some reason, usually pertains to insufficient balance
    - REVERTED - the associated transaction was completed first but was then rolled back later in time potentially due to customer reaching out to Revolut
- **ea_cardholderpresence**: string indicating if the card holder was present when the transaction happened
- **ea_merchant_mcc**: float corresponding to the Merchant Category Code (MCC)
- **ea_merchant_city**: string corresponding to the merchant's city
- **ea_merchant_country**: string corresponding to the merchant's country
- **direction**: string indicating the direction of the transaction
- **user_id**: string uniquely identifying the user
- **created_date**: datetime corresponding to the transaction's created date

---

Tips

1. **Tell us a story!** - we expect you to be creative and inquisitive
    a. Presentation, creativity and the ability to dive deep into the problem space is a strong component of the Data Scientist role
    b. Imagine presenting your results to a team of technical and non-technical employees
2. **Show us your data skills** - some of the data skills we look for are:
    a. tie data and analyses to business impact
    b. data preparation, exploration, visualization
    c. target selection (labelling, defining success metrics)
    d. modeling (statistical tests or machine learning)
    e. code (clean, commented, pythonic)
3. **Use Jupyter notebook** - You should use Jupyter with a Python kernel (Python3 preferred)
4. **Communicate your thought process** - we care a lot about the auditability & reproducibility of research/analysis; keep your code clean and commented, and accompany all analyses with explanations (models, coefficients, tests and values)
5. **Back up assumptions with data** - 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