Vi - Data Scientist Home Assignment

Dear candidate, thank you for taking the time to complete this home assignment. The purpose of this assignment is to test both your analytics and ML skills, and to give you a glimpse on our day-to-day challenges. Attached are two data files:

Events data:

Fields:

- Member_id - a uuid of a member
- Dt - date of the event
- Event_type - as described below:

Event types:

- App_interaction - a member using the club's application (for example , a recorded workout)
- Personal_appointment_scheduled - an indicator that a personal appointment between the member and the staff has been scheduled (to sometime in the near future)
- Human_communication_event - a phone call or some other face to face interaction between the member and the staff
- Sms_sent - an sms message sent from the staff to the member
- Chat_message_sent - a message sent from the staff to the member in the club's messaging platform
- Manual_email_sent - an email composed by a (human) staff representative and sent to a member
- Automated_email_sent - an email composed by the clubs messaging system and sent to a member
- Fitness_consoltation - a consultation meeting between the member and a professional personal trainer
- Pt_sale - a successful sale of a personal training package (pt) to a member
- Pt_usage - a personal training taking place in the club (a PT sale will be followed by multiple pt_usage events)
- Usage - a member visiting the club (for a "regular" workout, does not refer to pt)

Subscribers data:

Fields:

- Member_id - a uuid of a member
- Effective_date - the date of which the following data is relevant
- Segment_id - the club's branch id
- Age
- Gender (Male , Female, Non-Binary, NA)
- Country
- District
- City
- Subscription_type:
    - Term: fixed short time (i.e. monthly)
    - Open: long time (i.e. yearly)
    - CashTerm: fixed short time pre paid in cash
- Subscription_status (active/ in active), note that a member can have an inactive club membership but still attend personal training
- Weeks_from_last_activation
- Weeks_from_subscription_start
- Weeks_to_subscription_end

Goal:

We are interested in enhancing our personal training sales. For this matter, once a week, we want to predict the most likely members to buy a personal training package (event_type='pt_sale'). We want to deliver up to 10 of these predictions per segment_id.

- We want to suggest members that do not have a current personal training package
- We can assume that the features in the attached data sets will be available for future predictions
- As all real world problems, the data will not be "clean" and can have some missing value/discrepancies that have to be dealt with
- Ideally, a complete solution includes a fully documented code

Do not hesitate to reach out if something is not clear - allon@vitrainer.com

Best of luck!