**Data Science Case - Value Management Systems**

**About Gympass**

Gympass is a corporate wellness benefit that offers unlimited access to thousands of gyms, studios, and wellbeing activities worldwide.

We are a discovery platform that allows employers to engage their workforce in wellbeing activities by facilitating access to our global network of partners. With Gympass, employees can access hundreds of services across thousands of studios, fitness centers, and apps, so each person can find an activity to love.

Gympass’s programs are very flexible with a wide range of activities near work, home, and online. Employees can sign up for monthly memberships priced at 50% to 70% off of market prices. Companies benefit from higher engagement, as well as reduced absenteeism, turnover, and healthcare costs. And partners benefit from a wide promotion of their services and a new stream of revenue.

**About the case**

The case presentation is another step in the recruitment process of Data Scientists at Gympass. The goal is to simulate a real-world problem to assess the candidate’s skills - both technical (e.g. coding, problem solving, technical approach) and non-technical (e.g. presentation, business sense, recommendations).

We expect you to present very clearly the steps you have taken and the rationale used to structure your code and your solution.

Please read carefully and follow the instructions. If there are any doubts, please talk to your recruiter.

There are two main deliverables:

1. Code: Please share your script 7 days after you receive this case. This 1st step is offline/asynchronous, serving mostly for us to have enough time to review your code.
2. Presentation: To be scheduled after code sharing. You can use any preferred media to present it and you will have 25 minutes to do the presentation. After this time, there will be up to 25 minutes of Q&A when Gympassers will explore more topics related to your material.

**The assignment**

Gympass sells a subscription service to companies that offer it to their employees as a benefit. One of the greatest challenges that we have is charging the right price for users to visit each gym in our network. To do that we have this plan structure:

| **Plan** | **Price ($)** |
| --- | --- |
| Basic I | 39.90 |
| Basic II | 59.90 |
| Silver | 99.90 |
| Silver+ | 149.90 |

When we move a gym to a more expensive plan (e.g. gym that was available to Silver and Silver+ becomes available only on Silver+), we call it an uptier. When Gympass plans to perform an uptier, it communicates impacted users 15 days before the actual change. In this context we consider impacted users the ones who i) visited the product 60 days before the communication and ii) had the cheapest plan accepted by that gym (i.e. the ones that will lose access to the gym after the uptier).

After an uptier users may:

* Stay at the same plan (**keep**)
* Buy a more expensive plan (**upgrade**)
* Cancel their subscription (**churn**).

We want you to build a model that helps us decide which gyms (from “application.csv”) should be uptiered without a significant impact in churn rates. You may use any library, language or tool of your preference.

To do that you are receiving a table (“df.csv”) with past data on user (prefix “user\_”) and gyms (prefix “gym\_”) that had an uptier and were visited by this user. We also provide what the behavior of each user was at the end: keep (“is\_keep”), churn (”is\_churn”) or upgrade (“is\_upgrade”). A simple description of the columns can be found on a separate file called “variables description.csv”.

We expect you to explain:

1. Did you spot any interesting patterns?
2. What are your assumptions?
3. What model have you used?
   1. Overall concept
   2. Model output
   3. Interpretation of output
4. How do you measure performance?
5. To which gyms (from “application.csv”) do you recommend an uptier?