**Fyttlyf Data Science Team Test**

We are pleased to invite you to the interview process for our Decision Science Team! This is a practical exercise that will test your programming and analytical skills, please **include your codes as a PDF** in the submission. The programming language that is acceptable: Python or R

**Instructions: Please read carefully**

* Submit your answers for part 2 in a pdf format and part 1 & part 3 in the word document. Any collab link/txt file would be rejected by default.
* You may not consult with any other person regarding the test.
* You may use internet searches, books, or notes you have on hand.
* The test has three parts all of which are mandatory. Failing to complete any one part would result in the rejection of the submission.
* A thoughtful, clean & commented code is expected as a submission.
* Please do not make the submission public on sites like GitHub. This would help us in maintaining a high standard & regard for the assignment.

**Part 1**

**(Guesstimate)**

1. The government of India is proactively planning to dispatch Covid 19 kits which would have medicines, sanitizers & masks in the historic hot spots of the country. The kits would be delivered to the homes of the citizen who cannot afford them. You, the star analyst of the health ministry, are tasked to find out:
   1. How many such kits are needed?
   2. What is the budget the health ministry should get sanctioned from the national treasury?
   3. What is the approximate time required to complete this whole activity?

(Note: You must back your answers with structured numbers, please refer to the census or any website to get intel to solve this question. Hints & approach : [Link](https://www.analyticsvidhya.com/blog/2014/01/tips-crack-guess-estimate-case-study/) | [Link](https://mconsultingprep.com/case-interview-market-sizing-guesstimate/))

**Part 2**

**(Technical skills)**

1. Please create a new column name called ‘Batches’ with following steps:
   1. Get the min & max ‘Date\_TS’ from the data.
   2. Subtract max\_date\_TS from min\_date\_TS to get the number of days.
   3. Suppose you get 2 days as the output of point ‘b’, you would now write a code to create a tuple like:

{1: <min\_date\_TS>, <min\_date\_TS>+12 hours}

{2: <min\_date\_TS>+12 hours, <min\_date\_TS>+24 hours }

{3: <min\_date\_TS>+24 hours, <min\_date\_TS>+36 hours }

{4: <min\_date\_TS>+36 hours, <min\_date\_TS>+48 hours }

…

Till <min\_date\_TS>+x hours is less than equal to max\_date\_TS from point ‘a’.

* 1. Now create the column Batch where the values would be
     1. 1 where the ‘Date\_TS’ is between <min\_date\_TS>, <min\_date\_TS>+12 hours
     2. 2 where the ‘Date\_TS’ is between <min\_date\_TS>+12 hours, <min\_date\_TS>+24 hours
     3. 3 where the ‘Date\_TS’ is between <min\_date\_TS>+24 hours, <min\_date\_TS>+36 hours
     4. And so on till the tuple ends

1. Find which ‘version’ of the website has the highest reward rate.
   1. Reward rate is defined as Count (distinct Session \_id where reward = 1)/count(distinct Session\_id where reward = 0)
2. Build a graph which would show us the traffic allocation of each version across different batches.
   1. We are looking for a line graph plot for distinct visitors across ‘Batches’
3. What are the top 5 Button\_Names which were clicked?
   1. Find the button\_name which has the greatest number of Session \_id with clicks. To find definition of Clicks refer to the data\_dict sheet.
4. Which Device, Operating system & browsers got the highest reward rate from the winning version you found out in question 1.
   1. Let us say you found out that T3 has the highest reward rate, you would now need to find how does reward rate of T3 performs across different devices and browsers.

**Part 3**

**(Project)**

1. Please produce a brief write up on your last project in the field of Data science in less than 200 words.