Given a dataset with users features and emails sent/open times:

https://drive.google.com/file/d/18qenbF2j0S0UMj8nxfYqPjdvYYQ0lAZr/view?usp=sharinghtt
ps

N users (N=100.000), every user data lays in separate row (i.e. you have 100K rows for 100K users)

X[i] - features of i-th user, i in [1..N]
 X[i][0] - feature 1 (numerical)
 X[i][1] - feature 2 (numerical)
 X[i][2] - feature 3 (categorical)

M[i] - number of emails sent to i-th user (1 <= M[i] <= 15), Mi is not the same for every user i, (for instance M[1] could be 10, M[2] could be 5)
TS[i][j] - time of j-th email sending to i-th user, j in [1..M[i]]
TO[i][j] - time of j-th email opening by i-th user, j in [1..M[i]]

NOTE: If TS[i][j] > TO[i][j], then TO[i][j] belongs to the next day (for instance if sending time is 14:00 and opening time is 10:00, then it's 10:00 of the next day)

Goal: given features of new user X predict the sending time TS so that opening time TO will be as early as possible, i.e
(TO-TS) -> min

Hint: Users can open emails immediately or some time during a day (like some users prefer to read their email after lunch, some of them do it mornings, some do both)

Task:
1) Build and train your model
2) Discuss how your model can be improved
3) Publish your results on GitHub