Assignment - Data Scientist

Data Background :-

There is a csv file (Toppr_DS_Assignment_Data.csv) present in the zip folder. The data has 5 columns (session_id, subject_id, tutor_id , tutor_joined_on, exited_on). This is a dummy data created from the Doubts on Chat platform provided by Toppr where the user can ask doubts by clicking pictures of the questions and he will be assisted with the solution by a tutor via a chat enable platform.

Every session_id corresponds to a doubt asked by the user, tutor_id corresponds to the tutor who is solving the doubt, subject_id corresponds to the subject to which the doubt belongs to. The HT (handling time) of the doubt is defined as (exited_on - tutor_joined_on)

Tasks :-

1. Find the AHT (Average Handling time) at a monthly level for all the subjects
2. For every session_id find the number of doubts the tutor is solving at that instant. The tutor is allowed to solved multiple doubts simultaneously
3. Establish a hypothesis between the HT's when tutors are solving 1,2,3,4,5 and 6 doubts simultaneously (overall as well as subject wise)
4. From the above hypothesis please suggest a limit on the number of doubts the tutor should be able to pick at an instant to reduce the AHT

Submission :- Please submit the ipynb notebook as well the html version of the same. Include analysis / thought process in the notebook itself