Software Developer Interview Assignment

Background

We have an SMS-based service where pregnant women and new mums can ask questions about their health, pregnancy, and health of their baby. The questions are answered by a team of helpdesk agents.

Our helpdesk uses a ticketing (Q&A) platform, called Freshdesk, in order to respond to questions that mums ask via our SMS service. The platform is not designed to work with SMS and we have built workarounds in order to make Freshdesk work with our SMS program. Because of this, the analytics within Freshdesk are not useful to us and we need to build our own analytics capabilities. We do this by extracting the source data from the Freshdesk API (in JSON format) and put it into a Google Sheet for further analysis. For this exercise, instead of a Google Sheet, we will use Excel or CSV as the output file format.

1) When a question is asked from a new phone number (someone who has not asked a question before), a *new ticket* is created in Freshdesk for that phone number.
2) When we get a subsequent message from someone who *has* previously asked a question, a new *entry* is created *within the existing ticket* for that phone number.
3) All the entries (inbound and outbound messages) in each ticket are collectively referred to as a *conversation*.
4) Each phone number is linked with only one ticket and phone number.

We use the Freshdesk Conversations API end-point to extract the relevant details for every inbound message (question asked by a mum) and every outbound message (response sent by an agent), each question or response is a single *entry* in the conversation, and the data is in JSON format.

You will find attached the following text files:

1) The JSON output for a three tickets (ticket number 70039, 70063, and 71047) from the Conversations endpoint in text format.
2) An example output file.

Task

Write **Python** code to do the following:

Read the JSON data from each of the 3 files and, for *each entry in each ticket*, process the data for output to an Excel or CSV file (your choice) as per the instructions below. Remember that each message in or out consists of a single entry and the entire ticket thread is a conversation.

1) `Ticket_number` (column 1): This corresponds to the `'ticket_id'` field

2) `Entry_direction` (column 2): This is defined by the `'user_id'` key.
    a. The following User IDs correspond to the direction of `'in'`:
        i.   "43019547057"
    b. The following User IDs correspond to the direction of `'out'`:
        i.   "43050067402"
        ii.   "43050067304"
        iii. "43050067221"
        iv.   "43050067495"
        v.   "43049279159"
        vi.   "43068095002"
        vii. "43038614851"
        viii. "43067338910"
        ix.   "43019546306"
        x.   "43050067351"

3) `Entry_datetime` (column 3): This corresponds to the `'created_at'` key

4) `Timedelta_hours` (column 4): This is the time (in hours) between the first incoming message and the next outgoing message, should be sorted according to datetime, and is arrived at as follows:
    a. Start the timer at the first inbound message
    b. Stop the timer at the next outbound message
    c. Calculate the time difference between (a) and (b) above
    d. Ignore every other inbound message in between (a) and (b) above
    e. Repeat for the next inbound message
    f. See *Table 1* below as well as the example output file for an example:

Tkt_number	Entry_datetime	Direction	Timer
257	2020-20-05 09:15	in	<- start timer
257	2020-20-05 10:30	in	
257	2020-20-05 21:15	in	
257	2020-21-06 08:00	out	<- stop timer
257	2020-22-05 13:25	in	<- start timer
257	2020-22-05 16:40	out	<- stop timer

Table 1: Illustration on how to calculate Timedelta. In this example, we want to calculate the time difference between rows 1 and 4 and between rows 5 and 6.

Output

1) Write Python code to process each json file according to the directions above.
2) Process the combined output from all three tickets into a single file (Excel or CSV), sorted by ticket number
3) Document your code along with any assumptions you made
4) Send an email to jpatel@jacarandahealth.org containing the following:
    a. Your code and any accompanying documentation.
    b. The output file, either Excel or CSV, produced by your code.

Note: If you are unable to complete the assignment within the assigned time, please send in whatever you have completed along with an explanation of what challenges you faced. Please feel free to ask for clarifications at any time.