IMPORTANT TIP: Please don't use ChatGPT or similar AI engines for your answers. We will analyse the results and automatically reject any submissions flagged as AI-generated.

Q1: You are given a table user_activity with the following schema:

user_id	activity_date
INT	DATE

Write a query to calculate the Day 1 retention rate. Day 1 retention is defined as the percentage of users who returned to the platform exactly 1 day after their first activity.
Sample table user_activity data:

user_id	activity_date
1	2023-09-10
2	2023-09-10
1	2023-09-11
3	2023-09-11
2	2023-09-12
4	2023-09-12
3	2023-09-12
4	2023-09-13

For each activity_date, calculate:

- The number of new users (users whose first activity was on that date).
- The number of those users who returned exactly on the next day (Day 1).
- The Day 1 retention rate, which is the ratio of users who returned on the next day to the number of new users on the activity_date.

Sample Expected Output:

activity_date	new_users	returned_users	day_1_retention_rate
2023-09-10	2	1	0.50
2023-09-11	1	1	1.00
2023-09-12	1	1	1.00

Q2. You are given a table user_events that tracks user activity with the following schema:

user_id	event_type	event_time
INT	STRING	TIMESTAMP

Task:

1. Identify user sessions. A session is defined as a sequence of activities by a user where the time difference between consecutive events is less than or equal to 30 minutes. If the time between two events exceeds 30 minutes, it's considered the start of a new session.
2. For each session, calculate the following metrics:
    - session_id (a unique identifier for each session).
    - session_start_time (the timestamp of the first event in the session).
    - session_end_time (the timestamp of the last event in the session).
    - session_duration (the difference between session_end_time and session_start_time).
    - event_count (the number of events in the session).

Sample Table user_events data:

user_id	event_type	event_time
1	'click'	2023-09-10 10:00:00
1	'scroll'	2023-09-10 10:10:00
1	'click'	2023-09-10 10:50:00
1	'scroll'	2023-09-10 11:40:00
2	'click'	2023-09-10 09:00:00
2	'scroll'	2023-09-10 09:20:00
2	'click'	2023-09-10 10:30:00

Sample Expected Output:

user_id	session_id	session_start_time	session_end_time	session_duration	event_count
1	1	2023-09-10 10:00:00	2023-09-10 10:10:00	00:10:00	2
1	2	2023-09-10 10:50:00	2023-09-10 10:50:00	00:00:00	1
1	3	2023-09-10 11:40:00	2023-09-10 11:40:00	00:00:00	1
2	1	2023-09-10 09:00:00	2023-09-10 09:20:00	00:20:00	2
2	2	2023-09-10 10:30:00	2023-09-10 10:30:00	00:00:00	1

Q3. You are tasked with increasing the Day-on-Day (DoD) retention of users on the Seekho app. Currently, many users sign up and engage with content initially, but their activity drops off after the first day. What changes would you suggest to improve Day-on-Day retention? Consider both product features and data-driven strategies in your response.