KAMCORD ANALYTICS CODING CHALLENGE

Instructions

One of the metrics we use to measure the performance of our app is **Day-7 retention**. We define Day-7 retention as the percent of users who re-open the Kamcord app *exactly* 7 days after the date of their first Kamcord app open. Note that whether a user re-opens Kamcord on Day 6 or Day 8 is irrelevant to this metric.

For example, suppose we want to measure Day-7 retention over the first two days of October 2016. 20 users opened Kamcord for the first time on October 1st. On October 8th, 4 of these 20 users re-opened Kamcord. On October 2nd, 5 users opened Kamcord for the first time and 2 of these users returned on October 9th. The app's Day-7 retention from October 1st through October 2nd would then be (4+2) / (20+5) = 24%.

Attached is a CSV file with sample app data from a subset of users who first launched the app in September 2016. The fields provided are:

- user_id
- event_name
    - Name of the event
    - There are two events that can be used to track app opens:
        - APP_OPEN is logged when the app is opened or foregrounded
        - APP_CLOSED is logged when the app is closed or backgrounded
        - Due to poor internet connection, errors, or app/phone crashes, it is **not** guaranteed that we will receive both events for every app session
- event_time
    - Time the event occurred
- os_name

- app_version

We would like you to perform the following:

1. Design a model for the data that enabled efficient querying of Day-7 retention over a given date range, as well as the ability to filter for a specific OS and app version.
2. Write a script that reads the data from the CSV file into your model.
3. Use your model to answer the following questions:
    a. What was the overall Day-7 retention over the month of September?
    b. What was the Day-7 retention from September 8 through September 10 for Android users?
    c. What was the Day-7 retention over the month of September for iOS users using version 6.5?

Please submit the code for your script and model, as well as the answers to the above questions.

If you make any significant assumptions on the data, please state them.

<u>Programming language and restrictions</u>

Please use Python for this coding challenge. You may import modules to help read and parse the CSV file, but the internal model should be a Python class of your own design.

<u>Evaluation</u>

We'll evaluate your submission on the following criteria:

1. Correctness
2. Code quality, readability, and extensibility
3. Design of internal model
4. Time is not a major factor, but we would like to know how long you spent