QUESTION 1

1. Data description

The data provided shows records per activity. (One row is an activity)

An activity is defined a series of actions in a game that complete a word (Nonsensical or otherwise; a full word or simply a letter) as requested by the game. If the child abandons play or never completes the action correctly as the game is expecting; that is not recorded.

As you might expect in real life; the data might not be perfect. We've also left some mistakes / anomalies stay in the data – to simulate what you might see if you came to work with us!

The table provided has the following fields:

Id: Identifier of the activity
Childid: Identifier for the child
Userid: Identifier for the adult associated with the child

Gamename: Name of the game being played in the activity (One of our 10 games)
Nonsensewords: Is the word being played a nonsense word (Some games allow nonsense games)
Word: Word being played
Wordtype: type of word (Word or letter)

Createdat: Time stamp of the activity

Starttime: Start time of the activity in seconds (UNIX time)
Endtime: End time of the activity in seconds (UNIX time)
start_ms_time: start time of the activity in millisecond resolution
end_ms_time: end time of the activity in millisecond resolution

Duration: The duration (in seconds) of the activity. This is the difference in of end_ms_time and start_ms_time when these are available (Divided by 1000); else difference of endtime and starttime.

platform: platform of the tablet used with the playset

dnc: “Do Not Consider”. You should only use rows where dnc = 0.
Gender: Gender of the child. Blank or null indicates the data was not recorded
Dateofbirth: The child’s birthday if the adult recorded it

2. Questions:

1. Sessions: As you can see, children can have multiple activities on the same day. Some activities are close in time; while some are further apart.
    a. ‘Session-ize’ the activities: Can you group activities in a way to create sessions out of activities that are close together?
    b. How many activities per session on average (Over time)?
    c. How many different games do children play per session?
2. Retention:
    a. Do children spend more time per day; or less once they get started?
    b. What’s a player’s lifetime? In other words, how many days does a child play our games?
    c. Do children play the same game they started playing with? What other games do they play?
    d. Per session; do they spend more time on the game(s) they started with or on other games?
3. Modeling:
    a. Can you create a model to predict if a child is likely to play our games in the next 7 days?
    b. Don’t need to create a model for this one, simply discuss: If you were to create a model to predict if the child will play our games in the next 14 days? Next 30 days? Which of the 3 models (7 day, 14 day, 30 day) would you say be the most accurate one? (How do you define accuracy) Why?
4. Open ended, “bonus” questions: (Only do this if you have time)
    a. “Interesting insights”: Are there any insights from this data that you found interesting? If you did have unlimited time; how would you think about this question?
    b. “Clusters”: Given the data – can you create personas / clusters of the children? The personas could have different uses, for example:
        i. Help our marketing team promote different games for different types of children

ii. Help our education and product team understand what kind of games are played by which children