OBJECTIVE AND PURPOSE:

The overarching objective of this procedure is to develop a real-time data processing and analysis application using Spark streaming. This application aims to provide actionable recommendations based on the analysis of sensor data (like motion sensors), which could be applicable in various scenarios, including health and fitness monitoring, ergonomic assessments, or even in gaming and interactive environments.

1. **Upload Data to S3 Bucket**: Store the streaming data in an S3 bucket on AWS, which will serve as the data source for the Spark application.

2. **Load Data with PySpark as Structured Streaming**: Utilize PySpark to load the data as Spark structured streaming data, allowing for real-time processing and analysis.

3. **Estimate Key Metrics in Different Time Windows**: - Group the data by the "gt" (activity type) and estimate the count of "gt" and mean values of "x", "y", and "z" in various tumble window sizes (5 minutes, 15 minutes, 30 minutes).

   - Append these aggregated results to a results table for further analysis.

   - Set `maxFilesPerTrigger` to 5, controlling the number of files read in each trigger.

   - Implement a loop with a sleep value to ensure all data is read and processed.

4. **Visualize Time-Varying and Static Values**:

   - (Optional) Plot a chart showing these time-varying values against static values to analyze trends. This visualization can provide insights into how the activity data changes over time.

5. **Activity Recommendations Based on 15-Minute Tumble Windows**:

   - If during any 15-minute tumble window the count of "sit" activity is more than "stand", print "standing recommended". This suggests a recommendation for the user to stand, presumably for health or ergonomic reasons.

   - Calculate the average distance moved between consecutive tumble windows using the formula: \[sqrt((x2-x1)² + (y2-y1)² + (z2-z1)²)\], where x1, y1, z1 are the mean values of x, y, z in one window, and x2, y2, z2 are those in the next window.

   - If the average distance moved in any two consecutive windows is smaller than in the previous two windows, print "move recommended". This could be a prompt for the user to engage in more physical activity.