Compressed Archival Of S3 Data

Problem Statement

Extend the Week 1 Mentor Notebook Lab Walkthrough to implement a feature for representative archival. The task is to create an archive of monthly sensor data that is stored in S3. This is in addition to the raw data.

Generally, this has two purposes:

- To store really old data but only at an aggregate level
- To allow easy trend analysis or further processing at a daily or higher level.

The Python script reading the DynamoDB database is configured to store the sensor data for each calendar month in the S3 bucket, in a separate folder.

- Add a function to the script which:
    - Triggers only at the beginning of a month,
    - Computes the average of sensor values for each day of the previous month, and
    - Collects those 28/29/30/31 values for archiving.

- Store those values in a single CSV file to be saved in the same S3 bucket, in a separate folder structure with yearly folders for example.

The older (say anything over a year) raw data and monthly aggregates can be moved to Amazon S3 Glacier based storage by configuring S3 bucket storage transition policies. This saves further on cost by shifting older data to cheaper storage.