Complete EDA and Preprocessing Workflow

Dataset Description

The **rain_data_mechanical_master.csv** contains different timestamps and respective rainfall recorded in the last 3 minutes. The **rainfall_sound_8k** folder contains different audio files of short length. Each file is named with the timestamp it is started to record the data.

Tasks

1. Explore the data set (https://www.kaggle.com/datasets/sajilck/rain-data-master-8k) using Python and do the following.
    a. What is the sampling rate?
    b. What is the length (in samples) of individual audio files?
    c. Share a line plot of time vs. rainfall & label axis properly
    d. On which date was the maximum and minimum (non-zero) rainfall observed?
2. Create a data preprocessing workflow which does the following
        i. Based on timestamps from **rain_data_mechanical_master.csv** find out relevant audio clips (in the previous 3 min duration).
        ii. Read the audio files corresponding to a timestamp, sort them in chronological order and combine them together to form a single array.
        iii. Save the combined array as a numpy file which can be loaded later.
        iv. Perform feature engineering on the combined array
            1. Compute at least five statistical/other features
        v. Create a preprocessed data file with computed features and respective target values.
        vi. Upload all your preprocessed files to a Kaggle dataset and share the link of the same via Paatshala