Elucidata Assignment

You are given with a xlsx file. This file captures the output of a mass spec machine after an experiment run. In a typical experiment, multiple samples are inserted (sequentially) into a mass spec machine. Each sample contains a bunch of metabolites (chemicals compounds). The machine captures signal of each metabolite in each sample.

The first column in excel file corresponds to m/z ratio of the metabolite. The second column corresponds to the retention time of the metabolite. The third column corresponds to the name of this metabolite. Columns after 3rd column (from Column D) contains data for various metabolites.

For example, cell D1 captures signal value for metabolite "001_C24:0 PC" in "2788e_WHI_BAA_LIP_PrefB01" sample.
Cell E1 captures signal for the same metabolite ("001_C24:0 PC") in "2788f_WHI_BAA_LIP_PrefA01" sample.

Your first task is to write a flask/Django server where a user uploads this file.

After this, you have to perform a couple of operations on the input file. For these operations, pandas is a preferable library, but if you want to use something else, you can use that library as well. Following operations should be done on the input file:

1. In third column “Accepted Compound ID”, you need to filter out all the data for metabolite ids ending with: ‘PC’, ‘LPC’ and ‘plasmalogen’, and create 3 child datasets (1 for each compound id) from the data in input file.

2. Add a new column in the parent dataset with the name “Retention Time Roundoff (in mins)”. This column should have rounded-off values of the corresponding retention time. Retention time should be rounded-off to the nearest natural number.

3. After this, you should find the mean of all the metabolites which have same "Retention Time Roundoff" across all the samples. The resultant of above operation should be a new data-frame where you have to include the "Retention Time Roundoff" column and all samples. You don't have to include columns like m/z, Accepted Compound Id and Retention time.
You should create a new data-frame for each operation defined above, and for each operation, you should define a different API. For example, an API should be defined for uploading the file, another API for the first task. APIs should download the output after the tasks have been completed at the backend. (It is expected that no file will be downloaded for upload API).

You can use a very minimalistic frontend for this problem. You should also add test cases for functions you define.

Please note that this assignment has different grades for different parts of the assignment.