***Csv\_combine program instructions:***

**Assumption:**

1. There will be csv files stored in a folder( code is generic to pick files from any folder)
2. The folder name can be anything where the csv files are stored
3. Folder will also contain a file name “Combined.csv”
4. All other files except “Combined.csv” need to be added to the combined.csv
5. If the data in the files are duplicate then take the last updated data
6. The file name should be appended to the “Environment” column of Combined.csv when running the code
7. If a file name contains numbers or starting and ending spaces or some special characters, we will remove them
8. Once the files are processed through the code they will be moved to a another folder ( please enter path for processed files in the code). The folder structure created will be similar to Datalake processed structure ( e.g YEAR --🡪 MONTH-🡪 DAY)
9. If there are duplicate files coming in again for the same day then they should be moved to another folder (please enter path for bad files/duplicate file in the code). ). The folder will be created based on minute with a date structure.

**Instructions:**

To run this code (**csv\_combine.ipynb**) you only need to input below three paths

1. Please input path to your folder where the csv files are stored.
2. Input the path where you want to store your processed files
3. Input the path where you want to store your duplicate files
4. There are 4 test cases, you can see 4 different scripts
5. Please execute them in order from test\_case1 to testcase4
6. Test\_case1 will validate the path and the count from the initial files
7. Test\_case2 will validate the count from “Prod 4.csv” file
8. Test\_case3 will validate the count from “NA Preview.csv” file
9. Test\_case4 will validate that there should be any change in the counts when we process a blank file without data