**USER GUIDE**

1. The inputs of the code are 3 files:   
   one the NAV (.csv) file which is the format in which we need the data

second is a zip folder containing all the folders of AMC

third is the scheme code mapping excel ( .xlsx )

1. Some exceptions AMC’s are there which shouldn’t be included in the NAV extraction process such as adityabirla and geojit because they do not include “ NAV” header in their data which might lead to wrong extraction of data as it wont find any nav data for respective date column.Similarly due to wrong/improper format of excel data can result to no data being copied into first excel output hence should be manually done.
2. Refer to the comments in the code where I have mentioned where manually changes need to be made for excluding a folder or changing the month of data to be extracted.
3. The scheme code mapping excel has to be prepared manually once only which would include all the scheme code and scheme names of all funds of all AMC’s. It also has to have another column named folder name ie AMC name which has to be labelled exactly how the folders are named in zip file containing all folders.

This is because while matching scheme code, it first matches folder name… if folder name is matched it searches all those scheme names that match to folder name in the scheme code .xlsx and matches if scheme name found

1. I have also shared the original scheme code list vs my mapping excel on how its to be made ( have added around 100 folders on those I was working on). For example, if scheme name in original scheme code list is “Philip Emerging India Portfolio” where Philip is AMC/folder name, then new mapping excel will have only “Emerging India” or “Emerging India Portfolio” and folder name will be “Philip”.
2. So only after matching folder name inside zip and inside mapping excel it will search the scheme name “emerging india” in the excel inside that folder.A substring match takes place and if match found near where date nav data block is extracted only then it will assign respective scheme code.  
   Refer to mapping excel on how to complete it.(I cannot because I don’t have all the folders data to know exact scheme name)
3. The outputs will be two csv excels, one having data and second having a list of scheme name,codes that did not match due to any reason
4. Firstly, there is a folder not available column in the excel, which means that folder is not available in the zip or is wrongly labelled so didn’t match
5. Second, is the excluded folder column which includes those folder names which you have written in the code to exclude and really have been excluded during the run ie folders that are skipped by code. (commented in code)
6. Rest there remains unmatched scheme name , codes and folder names…If a scheme code hasn’t been assigned or matched to any nav block, it falls back to folder name so its easier to know which It belongs to  
   This fallbacks could have many possibilities like

Either the scheme name inside input mapping excel is spelled wrong or extra and does not match hence need to be corrected and changed.. because it searches for entire substring , so even an extra unnecessary word would lead to match not found

OR there are multiple nav data blocks (multiple funds) in one excel and scheme names are written as short forms in those excels and hance won’t match.

IN BOTH THESE CASES THE NAV DATA IS COPIED INTO THE FIRST CSV FILE,

But just instead of schemecode, it gets the folder name and requires manual edition to change the schemecode only

1. So by refering to unmatched scheme excel and going through each row, one can check and get to know whether schemecode is not matched or data has not come into output excel one and can manually do it.