Code snippet to run:

Step1: Import required class:

from Process\_data import Process\_data

import pandas as pd

Step2: Read excel file as DataFrame

df = pd.read\_excel('cik\_list.xlsx')

Step3: Create object of the class

obj = Process\_data(df)

Step4: Download files in ‘input\_dir’ folder. If the parameter **num** is None then whole 152 records are downloaded, Else specified number of records will be download.

ex.download\_files(num=None)

Step5: Compute the required parameters and convert it into DataFrame form. The parameter **num** works same as above step.

res\_df = ex.compute(num=None)

Step6: Convert dataframe to output csv file

res\_df.to\_csv(‘output\_data.csv’)

Note: Since the website [https://www.sec.gov/](https://www.sec.gov/Archives/) has threshold capacity, so downloading all the files will take lot of time.

Example Code Snippet:

from Process\_data import Process\_data

import pandas as pd

df = pd.read\_excel('cik\_list.xlsx')

obj = Process\_data(df)

ex.download\_files(num=5)

res\_df = ex.compute(num=5)

res\_df.to\_csv(‘output\_data.csv’)