***Combine two files in python:***

import pandas

import csv

# reading my new file

newfilename = "\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertification.txt"

fields = []

newrows = []

with open(newfilename, 'r') as txtfile:

tsvreader = csv.reader(txtfile, delimiter="\t")

fields = next(tsvreader)

for row in tsvreader:

newrows.append(row)

# reading old file

oldfilename = "\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertificationDetails1.csv"

rows = []

with open(oldfilename, 'r') as csvfile:

csvreader = csv.reader(csvfile)

for row in csvreader:

rows.append(row)

# combining both files

filename = "\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertificationDetailsMerge.csv"

with open(filename, 'w' , newline='') as csvfile:

csvwriter = csv.writer(csvfile)

csvwriter.writerow(fields)

csvwriter.writerows(rows)

csvwriter.writerows(newrows)

***Rename entire data 'Splunk' to 'Big Data':***

text = open("\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertificationDetailsMerge.csv", "r")

text = ''.join([i for i in text]).replace("Splunk", "BigData")

x = open("\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertificationDetailsMerge.csv","w")

x.writelines(text)

x.close()

***Create new attriubute using python:***

from csv import writer

from csv import reader

def add\_column\_in\_csv(input\_file, output\_file, transform\_row):

with open(input\_file, 'r') as read\_obj, \

open(output\_file, 'w', newline='') as write\_obj:

csv\_reader = reader(read\_obj)

csv\_writer = writer(write\_obj)

for row in csv\_reader:

transform\_row(row, csv\_reader.line\_num)

csv\_writer.writerow(row)

add\_column\_in\_csv('\\Users\\anujm\OneDrive\Desktop\\big data\Project\EmployeeCertificationDetails1.csv', '\\Users\\anujm\OneDrive\Desktop\\big data\Project\Certification\_Details.csv', \

lambda row, line\_num: row.append(row[0] + '\_\_' + row[1][-3] + row[1][-2] + row[1][-1]))