Code for Sagility Assignment : I've completed the assignment and submitted the same. As stated in the assignment, I've used Tesseract for OCR, but I didn't use openCV instead I used PIL to edit/crop, preprocess images and obtained the desired results.

!pip install pytesseract

import pytesseract

from PIL import Image

import matplotlib.pyplot as plt

import pandas as pd

image=Image.open('cheque-sample.jpg')

plt.imshow(image)

image.size

print(image.width,image.height)

img1=image.crop((1,1,1192,500))

img2=image.crop((1,501,1192,1100))

img3=image.crop((1,1101,1192,1700))

img4=image.crop((1,1701,1192,2292))

img1.save(r'C:\Users\sag\Desktop\sagar\_assignment\_sagility\output\_folder\Syndicate\_bank.jpg')

img2.save(r'C:\Users\sag\Desktop\sagar\_assignment\_sagility\output\_folder\ICICI\_bank.jpg')

img3.save(r'C:\Users\sag\Desktop\sagar\_assignment\_sagility\output\_folder\Canara\_bank.jpg')

img4.save(r'C:\Users\sag\Desktop\sagar\_assignment\_sagility\output\_folder\Axis\_bank.jpg')

dates=[]

bank=[]

amount=[]

text1=pytesseract.image\_to\_string(img1)

l1=[i.split(' ') for i in text1.split('\n')]

l1

dates.append(l1[1][6])

bank.append(l1[9][8]+' '+l1[9][9])

amount.append(l1[7][3])

text2=pytesseract.image\_to\_string(img2)

l2=[i.split(' ') for i in text2.split('\n')]

l2

dates.append(l2[16][0])

amount.append(l2[21][1])

bank.append(l2[29][0].split(']')[0]+' '+l2[29][1])

text3=pytesseract.image\_to\_string(img3)

l3=[i.split(' ') for i in text3.split('\n')]

l3

dates.append(l3[4][2])

amount.append(l3[19][1])

bank.append(l3[21][0]+' '+l3[21][1])

text4=pytesseract.image\_to\_string(img4)

l4=[i.split(' ') for i in text4.split('\n')]

l4

dates.append(l4[2][6])

amount.append(l4[7][1])

bank.append(l4[9][2]+' '+l4[9][3])

df=pd.DataFrame([dates,amount,bank]).transpose()

df.columns=['dates','amount','bank\_name']

df

df.to\_csv(r'C:\Users\sag\Desktop\sagar\_assignment\_sagility\output\_folder\sagar\_assignment\_output.csv')