CREATED API

from tkinter import \*  
import pyPDF  
  
my\_file = open('sample.pdf', 'rb')  
pdf\_reader = pyPDF.pdfrfilereader(my\_file)  
  
root=Tk()  
root.geometry('300\*300')  
  
b=Button(root, text="cancel")  
b.pack()  
  
b1=Button(root, text="Ok")  
b1.pack()  
  
root.mainloop()

NEXT STEP

rom tkinter import \*

import cv2

import pytesseract

def ocr\_core(img):

text=pytesseract.image\_to\_string(img)

return text

img=cv2.imread('img.jpg')

def get\_greyscale(img):

return cv2.cvtColor(image,cv2.COLOR\_BGR2GRAY)

def remove\_noise(img):

return cv2.medianBlur(image,5)

def thresholding(img):

return cv2.threshold(image,0,255,cv2.THRESH\_BINARY+cv2.THRESH\_OTSU)[1]

img=get\_greyscale(img)

img=get\_thresholding(img)

img=remove\_noise(img)

print(ocr\_core(img))

root=Tk()

root.geometry('300\*300')

b=Button(root, text="cancel")

b.pack()

b1=Button(root, text="Ok")

b1.pack()

root.mainloop()

CONVERTING TEXT FROM IMAGE

import pyautogui

from PIL import Image

from pytesseract import \*

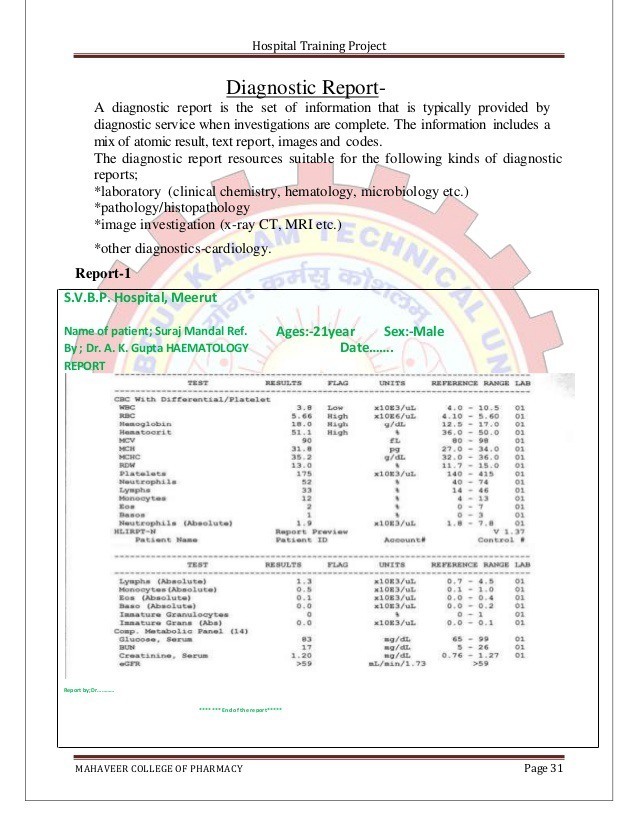
pytesseract.tesseract\_cmd=r"C:\Program Files\Tesseract-OCR\tesseract.exe"

#a = Image.open(r"C:\Program Files\Tesseract-OCR\chandu.png")

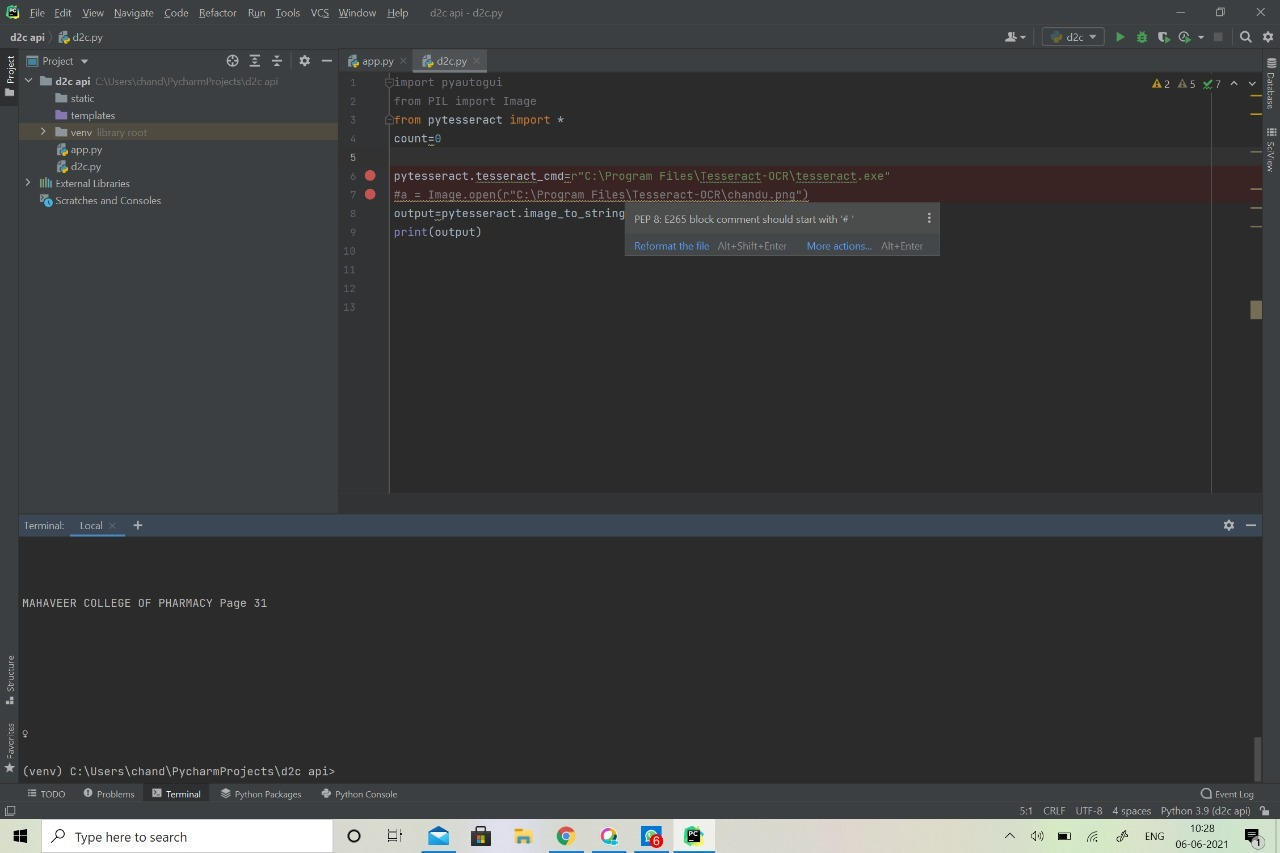
output=pytesseract.image\_to\_string(r"C:\Program Files\Tesseract-OCR\chandu.png")

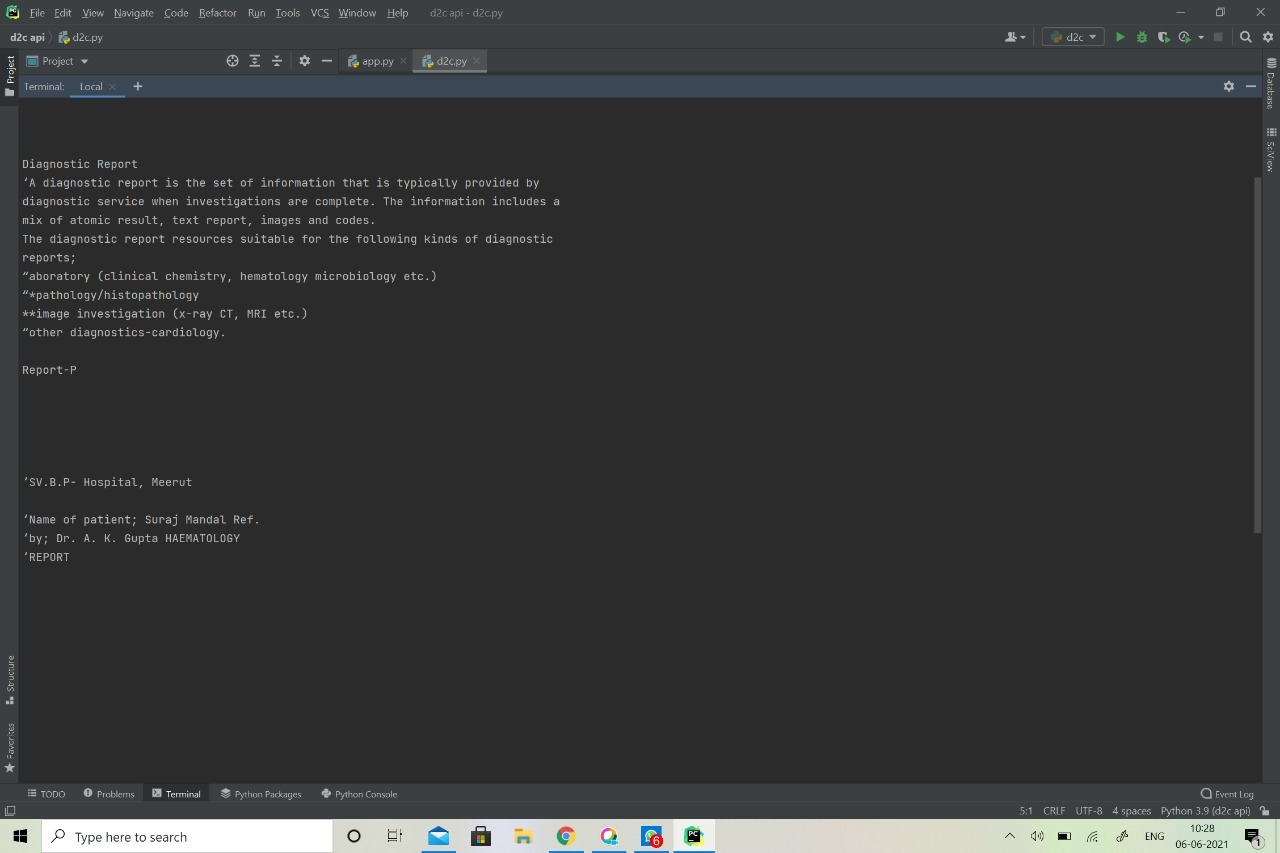
print(output)

INPUT



OUTPUT





THIS IS DONE BY CODELINGUISTS

BALACHANDRASEKHARREDDY

VAIDEHI RAHANGDALE

BERSIKA SWEETY