In [ ]:

```
import os
import pytesseract
import re
import pandas as pd

# class
class ExtractDocData():
    def __init__(self,file_path,pytesseract_path):
        print("Init method of ExtractDocData class")
        self.file_path = file_path
        self.pytesseract_path = pytesseract_path
        self.image_name = self.file_path.split(".")[0]

    def get_extract_text(self):
        pytesseract.pytesseract.tesseract_cmd = self.pytesseract_path
        self.text  = pytesseract.image_to_string(self.file_path)
    def get_save_document_data(self):
        folder_name = "Bank KYC Data"
        if not os.path.exists(folder_name):
            os.mkdir(folder_name)

        self.file_name = self.file_path.split(".")[0]
        with open(self.file_name,"w") as file:
            file.write(self.text)
class ExtractPanDetails(ExtractDocData):
    def __init__(self,file_path,pytesseract_path):
        print("Init method of ExtractPanDetails derived class")
        super().__init__(file_path,pytesseract_path)

    def get_pan(self):
        pan_pattern = re.search(r"\b[A-Z]{5}\d{4}[A-Z]\b",self.text)
        self.pan_number = pan_pattern.group() if pan_pattern else ""
    def get_dob(self):
        dob_pattern = re.search(r"\b\d{2}[/]\d{2}[/]\d{4}\b",self.text)
        self.dob = dob_pattern.group() if dob_pattern else ""
    def get_pan_details(self):
        self.get_extract_text()
        self.get_pan()
        self.get_dob()
        pan_data = {"Image Name": self.image_name,
                   "Date of Birth":self.dob,
                   "Pan Number":self.pan_number }
        # df = pd.DataFrame(pan_data)
        # print(pan_data)
        return pan_data

if __name__ == "__main__":
    file_path = r"pan1.png"
    pytesseract_path = r"C:\Users\Admin\AppData\Local\Tesseract-OCR\tesseract.exe"
    obj = ExtractPanDetails(file_path,pytesseract_path)
    obj.get_pan_details()
```