In [4]:

```
class ProjectPanAadhar():

    def __init__(self):
        pass
    def pan_card(self,pan_path):
        import pytesseract
        import re

        """
        This function extract texts from PAN card image and provides user in
        """
        if type(pan_path)==str:
            pan_path_new=[]
            pan_path_new.append(pan_path)
            pan_path=pan_path_new
        pytesseract.pytesseract.tesseract_cmd=r"C:\Program Files\Tesseract-O
        dob1=[]
        user=[]
        pan=[]
        for path in pan_path:
            path_list=path.split("\\")
            file_name=path_list[-1]
            file_name=file_name.rstrip(".jpg")
            text=pytesseract.image_to_string(path)
            dob_pattern=r"\b\d{1,2}[-/]\d{1,2}[-/]\d{2,4}\b"
            dob=re.findall(dob_pattern,text)
            dob=dob[0]
            pan_pattern=r"\b[A-Z]{5}[0-9]{4}[A-Z]\b"
            pan_no_list=re.findall(pan_pattern,text)
            try:
                pan_no=pan_no_list[0]


            except:
                pan_no="Not available"

            dob1.append(dob)
            pan.append(pan_no)
            user.append(file_name)
        dict1={"User Name":user,"DOB":dob1,"PAN No.":pan}

        dict1={"User Name":user, "PAN No.":pan, "DOB":dob1}
        return dict1

    def aadhar_card(self,aadhar_path):
        import pytesseract
        import re
        """
        This function extract texts from aadhar card image and provides user
        """
        if type(aadhar_path)==str:
            aadhar_path_new=[]
            aadhar_path_new.append(aadhar_path)
            aadhar_path=aadhar_path_new
        pytesseract.pytesseract.tesseract_cmd=r"C:\Program Files\Tesseract-O
        dob_adhar=[]
```

```
        adhar=[]
        user_aadhar=[]
        for path in aadhar_path:
            aadhar_dict={}
            path_list=path.split("\\")
            file_name=path_list[-1]
            file_name=file_name.rstrip(".jpg")
            text=pytesseract.image_to_string(path)
            dob_pattern=r"\b\d{1,2}[-/]\d{1,2}[-/]\d{2,4}\b"
            dob_list=re.findall(dob_pattern,text)
            if bool(dob_list):
                dob=dob_list[0]
            else:
                dob_pattern=r"\b[: ]{2,4}[0-9]{4}\b"
                dob_list=re.findall(dob_pattern,text)
                dob_string="".join(dob_list)
                dob_new_list=[]
                for char in dob_string:
                    if char.isdecimal():
                        dob_new_list.append(char)
                dob="".join(dob_new_list)

            aadhar_pattern=r"\b[0-9]{4}[ ][0-9]{4}[ ][0-9]{4}\b"
            aadhar_no_list=re.findall(aadhar_pattern,text)
            try:
                aadhar_no=aadhar_no_list[0]

            except:
                aadhar_no="Not available"
                pass
            dob_adhar.append(dob)
            adhar.append(aadhar_no)
            user_aadhar.append(file_name)
            dict2={"User Name":user_aadhar,"DOB":dob_adhar,"Aadhar No.":adha
        return dict2









import glob
import pandas as pd
pan_path_list=glob.glob(r"C:\Users\kannu\OneDrive\Desktop\DataScience\Projec
aadhar_path_list=glob.glob(r"C:\Users\kannu\OneDrive\Desktop\DataScience\Pro
obj=ProjectPanAadhar()
pan_details=obj.pan_card(pan_path_list)
pan_df=pd.DataFrame(pan_details)
pan_df.to_csv("User_PAN_details.csv",index=False)
aadhar_details=obj.aadhar_card(aadhar_path_list)

adhar_df=pd.DataFrame(aadhar_details)
adhar_df.to_csv("User_Aadhar_details.csv",index=False)
print("PAN details are=\n",pan_details)
print("\n\n\n")
print("Aadhar details are=\n",aadhar_details)
```

```
114
115
116
117
118
```

```
PAN details are=
 {'User Name': ['user_aditya', 'user_deepak', 'user_kalu', 'user_manikandan',
'user_monika', 'user_prem', 'user_sanjay', 'user_shrikanth', 'user_soniya', 'us
er_vivek'], 'PAN No.': ['DCQPK8010K', 'DZNPP7962M', 'FMNPM7828G', 'BNZPM2501F',
'EJAPS0276M', 'BJDPP6011M', 'LLSPS8447P', 'GQBPK8700C', 'HWLPS5369L', 'EAKPK120
4G'], 'DOB': ['24/01/1994', '15/07/1983', '17/04/1982', '16/07/1986', '31/10/19
92', '09/07/1986', '28/02/1991', '04/05/1997', '01/01/1993', '17/09/1996']}



Aadhar details are=
 {'User Name': ['user_kanika', 'user_kasif', 'user_khalid', 'user_nazifa', 'use
r_pandey', 'user_panka', 'user_parikh', 'user_santosh', 'user_sira', 'user_sohe
b'], 'DOB': ['11/08/1993', '01/01/2001', '07/08/1985', '06/06/2020', '12/01/201
2', '1994', '09/02/2016', '07/02/1999', '01/02/1991', '1993'], 'Aadhar No.':
['2114 5270 9955', '3140 3351 7159', '3076 6146 3849', '8776 7245 1739', '5548
0381 0828', '2595 2610 4168', '6997 4950 7961', '6562 6029 6757', '4976 5233 26
09', '5177 5394 4789']}
```

In [ ]:
```
1
```