

Task3

December 22, 2020

```
[4]: import pymongo
from pymongo import MongoClient
from pprint import pprint
client = pymongo.MongoClient("mongodb://user1:123@cluster0.ygzg9.mongodb.net/
↳db1?retryWrites=true&w=majority")
mydb=client["db1"]
```

```
[6]: import camelot
files=["d9f8e6d9-660b-4505-86f9-952e45ca6da0.pdf",
↳"a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf",
↳"1c1edeee-a13e-4b2e-90be-eb1dd03c3384.pdf"]
for file in files:
    tables=camelot.read_pdf(file)
    print("FILE ",file,"-----",tables.n)
    for table in tables:
        print(table.df)
```

PdfReadWarning: Invalid stream (index 16) within object 41 0: Stream has ended unexpectedly [pdf.py:1572]

FILE d9f8e6d9-660b-4505-86f9-952e45ca6da0.pdf ----- 1

	0	\
0		Date
1	April 4, 2018	Motilal Oswal Asset Management
2		
3		

	1	2	\
0	Name of the analyst/investor		Type
1	Company Limited	One-on-One	\nmeeting
2	Credit Suisse	\nI	Voice call
3	April 5, 2018	Maybank Eng Securities India Pr...	One-on-One \nmeeting

	3
0	Location
1	Mumbai
2	- M
3	umbai

```
FILE a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf ----- 0
FILE 1c1edeee-a13e-4b2e-90be-eb1dd03c3384.pdf ----- 1
```

0 \

```
0 To,\nThe Manager,\nListing Department,\nBSE le...
1
```

1

```
0 To,\nThe Manager,\nListing Department,\nThe Na...
1
```

```
[6]: import tabula
import pandas as pd
import json
import pymongo
from pymongo import MongoClient
from pprint import pprint
client = pymongo.MongoClient("mongodb+srv://user1:123@cluster0.ygzg9.mongodb.
    ↳net/db1?retryWrites=true&w=majority")
mydb=client.test
files=["d9f8e6d9-660b-4505-86f9-952e45ca6da0.
    ↳pdf","a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf",
    ↳"1c1edeee-a13e-4b2e-90be-eb1dd03c3384.pdf","EICHERMOT.pdf"]
for file in files:
    tables = tabula.read_pdf(file, output_format='dataframe', pages='all')
    print("FILE ",file,"-----")
    for table in tables:
        print(table)
        d = json.loads(table.T.to_json()).values()
        print(d)
        col=mydb[file]
        col
        col.insert_many(d)
```

Got stderr: Dec 22, 2020 8:27:16 PM

org.apache.pdfbox.contentstream.PDFStreamEngine operatorException

SEVERE: Cannot read JBIG2 image: jbig2-imageio is not installed

Dec 22, 2020 8:27:16 PM org.apache.pdfbox.contentstream.PDFStreamEngine
operatorException

SEVERE: Cannot read JBIG2 image: jbig2-imageio is not installed

The output file is empty.

```
FILE d9f8e6d9-660b-4505-86f9-952e45ca6da0.pdf -----
```

```
FILE a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf -----
```

	Day& Date	Name ofOrganisation	Type ofMeeting
0	Friday.	HDFCSecurities Limited	One-on-one Call
1	April 06, 2018	NaN	NaN

```
dict_values([{'Day& Date': 'Friday.', 'Name ofOrganisation': 'HDFC Securities
Limited', 'Type ofMeeting': 'One-on-one Call'}, {'Day& Date': 'April 06, 2018',
'Name ofOrganisation': None, 'Type ofMeeting': None}])
```

FILE 1c1edeeee-a13e-4b2e-90be-eb1dd03c3384.pdf -----

```

Date                Particulars Type of Interaction
0 April 06, 2018 SBICAP Securities Ltd          One on One
```

```
dict_values([{'Date': 'April 06, 2018', 'Particulars': 'SBICAP Securities Ltd',
'Type of Interaction': 'One on One'}])
```

FILE EICHERMOT.pdf -----

STATUTORY REPORTS \

```

0 (ii) Recommending the amount of expenditure
1 to be incurred on the activities referred to in
2 CSR policy.
```

```

3 (iii) Monitoring the CSR Policy of the Company
4 from time to time.
```

```
5 NaN
```

```
6 6.2 Meetings and Attendance
```

```
7 NaN
```

```
8 T wo Meetings of the Corporate Social
```

```
9 Responsibility Committee of the Company were
```

```
10 held during the Financial Year 2016-17 on May
```

```
11 5, 2016 and October 28, 2016. The names of the
```

```
12 members, chairperson and attendance details are
```

```
13 as under:
```

```
14 NaN
```

```
15 NaN
```

```
16 Sl. Name Chairman/ No. of No. of
```

```
17 No. Member meetings meetings
```

```
18 held attended
```

```
19 1. Mr S. Sandilya Chairman 2 2
```

```
20 2. Mr Siddhartha Lal Member 2 2
```

```
21 NaN
```

```
22 3. Mr Prateek Jalan Member 2 2
```

```
23 NaN
```

```
24 M r Manhar Kapoor, General Counsel & Company
```

```
25 Secretary acts as the Secretary to the Corporate
```

```
26 Social Responsibility Committee.
```

```
27 NaN
```

```
28 7. RISK MANAGEMENT COMMITTEE
```

```
29 In terms of Regulation 21 of the SEBI (LODR)
```

Unnamed: 1

```
0 7.1 Major Terms of Reference
```

```
1 (i) To assist the Board in formulating the Risk
```

```
2 Management Plan and practices.
```

```
3 (ii) To monitor and review risk management plan
```

```
4 and practices of the Company as approved
```

```
5 by the Board.
```

6 NaN

7 7.2 Members of the Committee

8 NaN

9 NaN

10 Sl. Name Chairman/

11 No. Member

12 1. Mr Siddhartha Lal (Managing Director & Cha...

13 Chief Executive Officer)

14 2. Mr S. Sandilya (Chairman and Non-Executive ...

15 Independent Director)

16 3. Mr Lalit Malik (Chief Financial Officer) Me...

17 NaN

18 8. SHARES COMMITTEE

19 The Shares Committee of the Company consists of

20 three members i.e. Mr Siddhartha Lal - Managing

21 Director & Chief Executive Officer, Mr Lalit M...

22 NaN

23 Financial Officer and Mr Manhar Kapoor - General

24 Counsel & Company Secretary to look after and

25 approve transfer/transmission of equity shares,

26 dematerialisation, issue of duplicate certific...

27 All valid requests for transfer & transmission...

28 physical form, duplicate issue of share certif...

29 processed in time.

dict_values([{'STATUTORY REPORTS': '(ii) Recommending the amount of
expenditure', 'Unnamed: 1': '7.1 Major Terms of Reference'}, {'STATUTORY
REPORTS': 'to be incurred on the activities referred to in', 'Unnamed: 1': '(i)
To assist the Board in formulating the Risk'}, {'STATUTORY REPORTS': 'CSR
policy.', 'Unnamed: 1': 'Management Plan and practices.'}, {'STATUTORY REPORTS':
'(iii) Monitoring the CSR Policy of the Company', 'Unnamed: 1': '(ii) To monitor
and review risk management plan'}, {'STATUTORY REPORTS': 'from time to time.',
'Unnamed: 1': 'and practices of the Company as approved'}, {'STATUTORY REPORTS':
None, 'Unnamed: 1': 'by the Board.'}, {'STATUTORY REPORTS': '6.2 Meetings and
Attendance', 'Unnamed: 1': None}, {'STATUTORY REPORTS': None, 'Unnamed: 1': '7.2
Members of the Committee'}, {'STATUTORY REPORTS': 'T wo Meetings of the
Corporate Social', 'Unnamed: 1': None}, {'STATUTORY REPORTS': 'Responsibility
Committee of the Company were', 'Unnamed: 1': None}, {'STATUTORY REPORTS': 'held
during the Financial Year 2016-17 on May', 'Unnamed: 1': 'Sl. Name Chairman/'},
{'STATUTORY REPORTS': '5, 2016 and October 28, 2016. The names of the',
'Unnamed: 1': 'No. Member'}, {'STATUTORY REPORTS': 'members, chairperson and
attendance details are', 'Unnamed: 1': '1. Mr Siddhartha Lal (Managing Director
& Chairman'}, {'STATUTORY REPORTS': 'as under:', 'Unnamed: 1': 'Chief Executive
Officer)'}, {'STATUTORY REPORTS': None, 'Unnamed: 1': '2. Mr S. Sandilya
(Chairman and Non-Executive Member'}, {'STATUTORY REPORTS': None, 'Unnamed: 1':
'Independent Director)'}, {'STATUTORY REPORTS': 'Sl. Name Chairman/ No. of No.
of', 'Unnamed: 1': '3. Mr Lalit Malik (Chief Financial Officer) Member'},
{'STATUTORY REPORTS': 'No. Member meetings meetings', 'Unnamed: 1': None},
{'STATUTORY REPORTS': 'held attended', 'Unnamed: 1': '8. SHARES COMMITTEE'},

```
{'STATUTORY REPORTS': '1. Mr S. Sandilya Chairman 2 2', 'Unnamed: 1': 'The
Shares Committee of the Company consists of'}, {'STATUTORY REPORTS': '2. Mr
Siddhartha Lal Member 2 2', 'Unnamed: 1': 'three members i.e. Mr Siddhartha Lal
- Managing'}, {'STATUTORY REPORTS': None, 'Unnamed: 1': 'Director & Chief
Executive Officer, Mr Lalit Malik - Chief'}, {'STATUTORY REPORTS': '3. Mr
Prateek Jalan Member 2 2', 'Unnamed: 1': None}, {'STATUTORY REPORTS': None,
'Unnamed: 1': 'Financial Officer and Mr Manhar Kapoor - General'}, {'STATUTORY
REPORTS': 'M r Manhar Kapoor, General Counsel & Company', 'Unnamed: 1': 'Counsel
& Company Secretary to look after and'}, {'STATUTORY REPORTS': 'Secretary acts
as the Secretary to the Corporate', 'Unnamed: 1': 'approve transfer/transmission
of equity shares,'}, {'STATUTORY REPORTS': 'Social Responsibility Committee.',
'Unnamed: 1': 'dematerialisation, issue of duplicate certificates, etc.'},
{'STATUTORY REPORTS': None, 'Unnamed: 1': 'All valid requests for transfer &
transmission of shares in'}, {'STATUTORY REPORTS': '7. RISK MANAGEMENT
COMMITTEE', 'Unnamed: 1': 'physical form, duplicate issue of share certificate
were'}, {'STATUTORY REPORTS': 'In terms of Regulation 21 of the SEBI (LODR)',
'Unnamed: 1': 'processed in time.'}]})
```