```python
import urllib.request
urllib.request.urlretrieve(
    'http://www.delhimetrorail.com/rightto_infoact/documents/I_19_%20Junior_Engineer_El
    'dmrc.pdf')
```

('dmrc.pdf', <http.client.HTTPMessage at 0x1f477f66a90>)

```python
import tabula
#the pd is the standard shorthand for pandas
import pandas as pd
```

```python
from tabula import read_pdf
dfs = tabula.read_pdf("dmrc.pdf", pages='all')
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
~\anaconda3\lib\site-packages\tabula\io.py in _run(java_options, options, path,
encoding)
     79     try:
---> 80         result = subprocess.run(
     81             args,

~\anaconda3\lib\subprocess.py in run(input, capture_output, timeout, check, *popenargs,
**kwargs)
    492
--> 493     with Popen(*popenargs, **kwargs) as process:
    494         try:

~\anaconda3\lib\subprocess.py in __init__(self, args, bufsize, executable, stdin,
stdout, stderr, preexec_fn, close_fds, shell, cwd, env, universal_newlines, startupinfo,
creationflags, restore_signals, start_new_session, pass_fds, encoding, errors, text)
    857
--> 858             self._execute_child(args, executable, preexec_fn, close_fds,
    859                                 pass_fds, cwd, env,

~\anaconda3\lib\subprocess.py in _execute_child(self, args, executable, preexec_fn,
close_fds, pass_fds, cwd, env, startupinfo, creationflags, shell, p2cread, p2cwrite,
c2pread, c2pwrite, errread, errwrite, unused_restore_signals, unused_start_new_session)
   1310             try:
-> 1311                 hp, ht, pid, tid = _winapi.CreateProcess(executable, args,
   1312                                         # no special security

FileNotFoundError: [WinError 2] The system cannot find the file specified

During handling of the above exception, another exception occurred:
```

```
JavaNotFoundError                          Traceback (most recent call last)
<ipython-input-3-aad92c369546> in <module>
      1 from tabula import read_pdf
----> 2 dfs = tabula.read_pdf("dmrc.pdf", pages='all')

~\anaconda3\lib\site-packages\tabula\io.py in read_pdf(input_path, output_format,
encoding, java_options, pandas_options, multiple_tables, user_agent, **kwargs)
    320
    321     try:
--> 322         output = _run(java_options, kwargs, path, encoding)
    323     finally:
    324         if temporary:

~\anaconda3\lib\site-packages\tabula\io.py in _run(java_options, options, path,
encoding)
     89         return result.stdout
     90     except FileNotFoundError:
---> 91         raise JavaNotFoundError(JAVA_NOT_FOUND_ERROR)
     92     except subprocess.CalledProcessError as e:
     93         logger.error("Error from tabula-
java:\n{}\n".format(e.stderr.decode(encoding)))

JavaNotFoundError: `java` command is not found from this Python process.Please ensure
Java is installed and PATH is set for `java`
```

```
pip install install-jdk
```

```python
import pandas as pd
```

```python
dmrc=pd.read_excel('Book1.xlsx', index_col=0)
```

```
dmrc
```

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
2.0	19RNE0201258666	AMAN GUPTA	35249	UR	N	57.878	10.020	67.898	NaN	Page 1 Table 1
66.0	19RNE0204247093	NaN	36711	OBC	N	56.522	1.030	57.552	JASIM ALAM	Page 2 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
1.0	19RNE0201258398	SUMAN KUMARI	19/03/92	OBC	N	55.410	14.680	70.090	NaN	Page 1 Table 1
6.0	19RNE0201258320	SHUBHAM SHARMA	35556	UR	N	54.943	7.910	62.853	NaN	Page 1 Table 1
15.0	19RNE0202264018	SURAJ SINGH	34614	UR	N	54.725	6.920	61.645	NaN	Page 1 Table 1
...	...	...	...	...	...	...	...	...	...	...
252.0	19RNE0202263387	SANJAY KUMAR	15/11/78	OBC	Y	40.291	11.243	51.534	NaN	Page 6 Table 1
258.0	19RNE0201258854	JYOTI SETH	33765	UR	N	38.949	12.463	51.412	NaN	Page 6 Table 1
246.0	19RNE0202260920	MRITYUNJAY JHA	15/11/93	UR	N	38.915	12.683	51.598	NaN	Page 6 Table 1
248.0	19RNE0202261782	PUNEET KUMAR	35099	OBC	N	36.909	14.680	51.589	NaN	Page 6 Table 1
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	MUTTEWAR LAXMIKANT	Page 2 Table 1

264 rows × 10 columns

```
dmrc_45=dmrc[dmrc['Paper-I (out of 80)']>=45.0]
```

```
dmrc_45
```

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
2.0	19RNE0201258666	AMAN GUPTA	35249	UR	N	57.878	10.020	67.898	NaN	Page 1 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
66.0	19RNE0204247093	NaN	36711	OBC	N	56.522	1.030	57.552	JASIM ALAM	Page 2 Table 1
1.0	19RNE0201258398	SUMAN KUMARI	19/03/92	OBC	N	55.410	14.680	70.090	NaN	Page 1 Table 1
6.0	19RNE0201258320	SHUBHAM SHARMA	35556	UR	N	54.943	7.910	62.853	NaN	Page 1 Table 1
15.0	19RNE0202264018	SURAJ SINGH	34614	UR	N	54.725	6.920	61.645	NaN	Page 1 Table 1
...	...	...	...	...	...	...	...	...	...	...
130.0	19RNE0203247536	RAJU KUMAR	34944	OBC	N	45.024	9.793	54.817	NaN	Page 3 Table 1
216.0	19RNE0210252349	ASHISH KAUSHAL	35349	UR	N	45.024	7.243	52.267	NaN	Page 5 Table 1
201.0	19RNE0210252089	IMRAN KHAN	34158	EWS	N	45.017	7.577	52.594	NaN	Page 5 Table 1
137.0	19RNE0209250060	MANGTU RAM	17/12/89	OBC	N	45.017	9.457	54.474	NaN	Page 3 Table 1
151.0	19RNE0201259963	KAMAL DEEP SINGH	35247	SC	N	45.010	9.127	54.137	NaN	Page 4 Table 1

194 rows × 10 columns

```
dmrc_ur.count()
```

```
Roll No.                                                    77
Candidate Name (S/Sh./Ms.)                                  63
Date of Birth                                               77
Applying Cat.                                               77
EXS                                                         77
Paper-I (out of 80)                                         77
Paper-II (out of 20)                                        77
Total Marks (out of 100)                                    77
Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL         14
```

Table Name
dtype: int64

```
dmrc.describe()
```

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
count	263.000000	263.000000	263.000000
mean	47.071399	8.406814	55.478213
std	3.462149	2.200691	3.375140
min	36.909000	1.030000	51.321000
25%	44.787000	6.913000	52.746500
50%	46.834000	8.463000	54.748000
75%	49.413500	9.797000	57.541000
max	57.878000	14.680000	70.090000

```
dmrc_ur=dmrc[dmrc['Applying Cat.']=='UR']
```

```
dmrc_ur
```

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
2.0	19RNE0201258666	AMAN GUPTA	35249	UR	N	57.878	10.020	67.898	NaN	Page 1 Table 1
6.0	19RNE0201258320	SHUBHAM SHARMA	35556	UR	N	54.943	7.910	62.853	NaN	Page 1 Table 1
15.0	19RNE0202264018	SURAJ SINGH	34614	UR	N	54.725	6.920	61.645	NaN	Page 1 Table 1
47.0	19RNE0210250465	GRIJESH KUMAR RAY	28/02/92	UR	N	54.719	4.133	58.852	NaN	Page 1 Table 1
49.0	19RNE0202263457	NaN	34036	UR	N	52.454	6.127	58.581	AHMED	Page 2 Table 1
...	...	...	...	...	...	...	...	...	...	...

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
237.0	19RNE0208255915	VIKRAM SINGH	34978	UR	N	40.943	10.790	51.733	NaN	Page 5 Table 1
225.0	19RNE0201256739	PRATEEK PANDEY	34344	UR	N	40.725	11.347	52.072	NaN	Page 5 Table 1
238.0	19RNE0202262850	AVANTIKA BHANDARI	20/12/95	UR	N	40.481	11.237	51.718	NaN	Page 5 Table 1
258.0	19RNE0201258854	JYOTI SETH	33765	UR	N	38.949	12.463	51.412	NaN	Page 6 Table 1
246.0	19RNE0202260920	MRITYUNJAY JHA	15/11/93	UR	N	38.915	12.683	51.598	NaN	Page 6 Table 1

77 rows × 10 columns

```
dmrc_ur.describe()
```

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
count	77.000000	77.000000	77.00000
mean	46.940974	8.802286	55.74326
std	3.801853	2.219849	3.68140
min	38.915000	2.710000	51.32100
25%	44.535000	7.353000	52.53700
50%	46.821000	9.020000	54.86500
75%	49.546000	10.230000	58.54800
max	57.878000	13.787000	67.89800

```
dmrc_45[dmrc['Applying Cat.']=='UR'].describe()
```

<ipython-input-87-88528efdb2d3>:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  dmrc_45[dmrc['Applying Cat.']=='UR'].describe()

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
count	55.000000	55.000000	55.000000
mean	48.710309	8.390782	57.101091

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
std	2.834612	2.381302	3.483869
min	45.024000	2.710000	51.619000
25%	46.477500	6.638500	54.271000
50%	48.400000	8.347000	56.820000
75%	50.315500	10.125000	59.578000
max	57.878000	13.787000	67.898000

```
dmrc.loc[1:19].describe()
```

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
count	24.000000	24.000000	24.000000
mean	53.358917	7.587208	60.946125
std	1.110777	2.439203	2.864955
min	51.803000	3.253000	55.274000
25%	52.259750	6.212750	58.916500
50%	53.356000	7.133500	60.791000
75%	54.322500	8.439000	62.311000
max	55.410000	14.680000	70.090000

```
dmrc=dmrc.sort_values(by=['Paper-I (out of 80)'],ascending=False)
```

```
dmrc_new=dmrc.loc[1:200]
```

```
dmrc_new[dmrc['Applying Cat.']=='UR'].describe()
```

```
<ipython-input-114-ce691a688736>:1: UserWarning: Boolean Series key will be reindexed
to match DataFrame index.
  dmrc_new[dmrc['Applying Cat.']=='UR'].describe()
```

	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)
count	59.000000	59.000000	59.000000
mean	48.192373	8.395864	56.588237
std	2.709981	2.296842	3.221449
min	43.905000	2.710000	51.619000
25%	46.020000	6.700000	54.007500
50%	47.729000	8.463000	55.986000
75%	50.088000	9.578500	58.938500
max	54.943000	13.787000	63.314000

```
dmrc_new[dmrc['Applying Cat.']=='UR']
```

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
6.0	19RNE0201258320	SHUBHAM SHARMA	35556	UR	N	54.943	7.910	62.853	NaN	Page 1 Table 1
15.0	19RNE0202264018	SURAJ SINGH	34614	UR	N	54.725	6.920	61.645	NaN	Page 1 Table 1
47.0	19RNE0210250465	GRIJESH KUMAR RAY	28/02/92	UR	N	54.719	4.133	58.852	NaN	Page 1 Table 1
49.0	19RNE0202263457	NaN	34036	UR	N	52.454	6.127	58.581	AHMED	Page 2 Table 1
16.0	19RNE0206246508	RADHESAYAM	33700	UR	N	52.021	9.360	61.381	NaN	Page 1 Table 1
19.0	19RNE0203248195	ALOK SRIVASTAV	16/01/94	UR	N	51.803	9.360	61.163	NaN	Page 1 Table 1
17.0	19RNE0201260017	TABREZ HASAN	34888	UR	N	51.559	9.797	61.356	NaN	Page 1 Table 1
12.0	19RNE0210250538	MANMOHAN TIWARI	35098	UR	N	51.343	10.903	62.246	NaN	Page 1 Table 1
79.0	19RNE0210250524	NaN	34064	UR	N	51.343	5.477	56.820	RAVI PRAKASH GUPTA	Page 2 Table 1
94.0	19RNE0201257083	NaN	34188	UR	N	50.657	5.363	56.020	PRERANA KUMARI	Page 2 Table 1
5.0	19RNE0201256630	NEHA	22/10/94	UR	N	50.637	12.677	63.314	NaN	Page 1 Table 1
25.0	19RNE0202261004	APENDRA RAWAT	20/07/94	UR	N	50.421	10.230	60.651	NaN	Page 1 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
51.0	19RNE0210251067	NaN	35131	UR	N	50.421	8.127	58.548	SHUBHAM MISHRA	Page 2 Table 1
38.0	19RNE0201258342	JAYA LAKSHMI	14/07/94	UR	N	50.210	9.353	59.563	NaN	Page 1 Table 1
46.0	19RNE0201258358	SURAJ KUMAR	35187	UR	N	50.210	8.690	58.900	NaN	Page 1 Table 1
95.0	19RNE0210252104	NaN	16/12/98	UR	N	49.966	6.020	55.986	SHUBHAM KUMAR PANDEY	Page 2 Table 1
208.0	19RNE0201256453	GOVIND SINGH	35125	UR	N	49.783	2.710	52.493	NaN	Page 5 Table 1
64.0	19RNE0210251137	NaN	35431	UR	N	49.756	7.913	57.669	RITVIK SRIVASTAVA	Page 2 Table 1
26.0	19RNE0201257208	SUBODH SOLANKI	20/08/94	UR	N	49.546	11.020	60.566	NaN	Page 1 Table 1
125.0	19RNE0210250426	DEEPAK KUMAR VERMA	25/07/92	UR	N	49.512	5.353	54.865	NaN	Page 3 Table 1
20.0	19RNE0209249982	VIVEK DADHICH	34400	UR	N	49.295	11.790	61.085	NaN	Page 1 Table 1
144.0	19RNE0210251956	REKHA SINGH	34371	UR	N	48.847	5.470	54.317	NaN	Page 3 Table 1
120.0	19RNE0208255963	PRINCE KUMAR KUSHWAHA	29/12/99	UR	N	48.624	6.463	55.087	NaN	Page 3 Table 1
96.0	19RNE0202261725	NaN	22/01/97	UR	N	48.624	7.357	55.981	NITIN AGGARWAL	Page 2 Table 1
157.0	19RNE0201259648	DEEPAK TRIPATHI	29/06/98	UR	N	48.624	5.370	53.994	NaN	Page 4 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
44.0	19RNE0201258214	SUSHIL MAAN	29/12/95	UR	N	48.624	10.353	58.977	NaN	Page 1 Table 1
73.0	19RNE0210250952	NaN	36079	UR	N	48.400	8.683	57.083	CHANDAN KUSHWAHA	Page 2 Table 1
117.0	19RNE0201256522	GAURAV SEMWAL	34245	UR	N	47.932	7.243	55.175	NaN	Page 3 Table 1
32.0	19RNE0201256940	PARVEZ RAFAT	15/08/96	UR	N	47.919	11.897	59.816	NaN	Page 1 Table 1
72.0	19RNE0201258461	NaN	34404	UR	N	47.729	9.357	57.086	DEEPTI SHUKLA	Page 2 Table 1
93.0	19RNE0207254898	NaN	34371	UR	N	47.729	8.347	56.076	MANISH SINGH NEGI	Page 2 Table 1
147.0	19RNE0203247482	SANJEEV KUMAR	23/08/95	UR	N	47.525	6.700	54.225	NaN	Page 4 Table 1
56.0	19RNE0205248431	NaN	34826	UR	N	47.261	10.903	58.164	PARVITA DUHAN	Page 2 Table 1
102.0	19RNE0202263284	JYOTI PATHAK	20/02/96	UR	N	47.044	8.797	55.841	NaN	Page 3 Table 1
22.0	19RNE0208246216	ISHANT SHARMA	34092	UR	N	47.031	13.787	60.818	NaN	Page 1 Table 1
150.0	19RNE0201257224	ANIL KUMAR PRADHAN	21/04/92	UR	N	46.841	7.353	54.194	NaN	Page 4 Table 1
140.0	19RNE0208246291	PRIYANKA CHOUDHARY	33978	UR	N	46.821	7.580	54.401	NaN	Page 3 Table 1
207.0	19RNE0209250207	RAMLAKHAN SHARMA	34760	UR	N	46.821	5.690	52.511	NaN	Page 5 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
178.0	19RNE0210251609	PRAGYA TRIVEDI	19/08/94	UR	N	46.603	6.700	53.303	NaN	Page 4 Table 1
36.0	19RNE0207254896	ARIJIT BHADRA	18/11/94	UR	N	46.576	13.017	59.593	NaN	Page 1 Table 1
106.0	19RNE0201260022	NISHA KUMARI	33699	UR	N	46.379	9.243	55.622	NaN	Page 3 Table 1
55.0	19RNE0201259980	NaN	19/08/92	UR	N	46.379	11.797	58.176	PRIYANKA PRASAD	Page 2 Table 1
194.0	19RNE0210251307	VIJAY KUMAR DIXIT	23/03/93	UR	N	46.359	6.577	52.936	NaN	Page 4 Table 1
115.0	19RNE0207255309	KARAN SINGH MEHTA	34370	UR	N	46.115	9.120	55.235	NaN	Page 3 Table 1
70.0	19RNE0210251032	NaN	21/02/00	UR	N	45.925	11.240	57.165	GAUTAM RAJ	Page 2 Table 1
220.0	19RNE0210250746	RUPAK KUMAR MISHRA	20/11/96	UR	N	45.912	6.253	52.165	NaN	Page 5 Table 1
77.0	19RNE0210254142	NaN	33762	UR	N	45.898	11.013	56.911	DISHA SRIVASTAVA	Page 2 Table 1
244.0	19RNE0210251241	RAMAN KUMAR	35431	UR	N	45.702	5.917	51.619	NaN	Page 6 Table 1
155.0	19RNE0203248235	PUSHPANJALI CHAUHAN	34700	UR	N	45.444	8.577	54.021	NaN	Page 4 Table 1
99.0	19RNE0202261761	AMAN KUMAR	28/11/94	UR	N	45.431	10.460	55.891	NaN	Page 3 Table 1
199.0	19RNE0210253267	ANUSHREE SRIVASTAVA	33822	UR	N	45.254	7.363	52.617	NaN	Page 5 Table 1

S. No.	Roll No.	Candidate Name (S/Sh./Ms.)	Date of Birth	Applying Cat.	EXS	Paper-I (out of 80)	Paper-II (out of 20)	Total Marks (out of 100)	Candidate Name (S/Sh./Ms.) PATHAN ANWAR AHMED JAMIL	Table Name
166.0	19RNE0203247584	DEEPAK UPADHYAY	34345	UR	N	45.234	8.463	53.697	NaN	Page 4 Table 1
182.0	19RNE0201257174	NEHA SACHDEVA	23/11/95	UR	N	45.234	7.907	53.141	NaN	Page 4 Table 1
216.0	19RNE0210252349	ASHISH KAUSHAL	35349	UR	N	45.024	7.243	52.267	NaN	Page 5 Table 1
152.0	19RNE0210250820	ATUL BHAWESH	35221	UR	N	44.766	9.343	54.109	NaN	Page 4 Table 1
165.0	19RNE0205249035	RAMAN MANOCHA	16/07/92	UR	N	44.576	9.130	53.706	NaN	Page 4 Table 1
171.0	19RNE0205248871	AMIT KUMAR GUPTA	34337	UR	N	44.535	9.020	53.555	NaN	Page 4 Table 1
224.0	19RNE0210252903	SHIVA AGRAHARI	17/07/93	UR	N	44.379	7.697	52.076	NaN	Page 5 Table 1
200.0	19RNE0206246702	MANOJ POKHRIYAL	34062	UR	N	43.905	8.693	52.598	NaN	Page 5 Table 1

```python
import matplotlib.pyplot as plt
plt.plot(['(Paper-I (out of 80)','Applying Cat.'])
plt.show()
```