In [1]:
```
### >> Table Extraction << ###
```

In [2]:
```
# Importing Modules
import os
import tabula
import camelot
import pymongo
import pandas as pd
```

In [3]:
```
# Connecting to MongoDB
client = pymongo.MongoClient('localhost', 27017)
db = client.PDF_Extraction
```

In [4]:
```
# a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf #
```

In [5]:
```
# Extracting

File = 'Files/a6b29367-f3b7-4fb1-a2d0-077477eac1d9.pdf'
x = tabula.read_pdf(File,pages="all",multiple_tables=True,guess=True)
```

In [6]:
```
# Cleaning

x[0].iloc[0,0]+=" "+x[0].iloc[1,0]
x[0].dropna(inplace=True)
```

In [7]:
```
# Storing

collection_name = File[6:] + "_Tables"
col = db[collection_name]
col.insert_one(x[0].to_dict('records')[0])
```

Out[7]:
```
<pymongo.results.InsertOneResult at 0x128100840>
```

In [8]:
```
# 1c1edeee-a13e-4b2e-90be-eb1dd03c3384.pdf #
```

In [9]:
```
# Extracting

File = 'Files/1c1edeee-a13e-4b2e-90be-eb1dd03c3384.pdf'

x = tabula.read_pdf(File,pages="all",multiple_tables=True,guess=True)
y = camelot.read_pdf(File,strip_text='\n')
```

In [10]:
```
# Cleaning

z = y[0].df[:1]
z.columns = ['0','1']
t = [x[0],z]
```

In [11]:
```
# Storing

collection_name = File[6:] + "_Tables"
col = db[collection_name]
col.insert_one(t[1].to_dict('records')[0])
```

Out[11]:
```
<pymongo.results.InsertOneResult at 0x1285b92c0>
```

In [12]:
```
# d9f8e6d9-660b-4505-86f9-952e45ca6da0.pdf
```

In [13]:
```
# Extracting

File = 'Files/d9f8e6d9-660b-4505-86f9-952e45ca6da0.pdf'


y = camelot.read_pdf(File,strip_text='\n')
```

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

In [14]:
```
# Cleaning

z = y[0].df
z.columns = ['0','1','2','3']
z.iloc[1,1]=z.iloc[1,0][15:]+" "+z.iloc[1,1]
z.iloc[1,0]=z.iloc[1,0][:15]
z.iloc[2,0]=z.iloc[1,0]
z.iloc[3,0]=z.iloc[3,1][:15]
z.iloc[3,1]=z.iloc[3,1][15:]
z.iloc[3,3]="M"+z.iloc[3,3]
z.iloc[2,3]="-"
```

In [15]:
```
# Storing

collection_name = File[6:] + "_Tables"
col = db[collection_name]
col.insert_one(z.to_dict('records')[0])
```

Out[15]:
```
<pymongo.results.InsertOneResult at 0x128100980>
```

In [16]:
```
# EICHERMOT.pdf #
```

In [17]:
```
# Extracting

File = 'Files/EICHERMOT.pdf'
x = tabula.read_pdf(File,pages="all",multiple_tables=True,guess=True)
y = camelot.read_pdf(File,strip_text='\n')
```

In [18]:
```
# Cleaning
```

In [19]:
```
# Storing

collection_name = File[6:] + "_Tables"
col = db[collection_name]
col.insert_one(x[0].to_dict('records')[0])
```

Out[19]:
```
<pymongo.results.InsertOneResult at 0x1285fd400>
```

In [20]:
```
### END OF NOTEBOOK ###
```