In [11]:
```
import numpy as np #linear algebra
import pandas as pd #data processing
# step 1 get top line
from bs4 import BeautifulSoup as bs
import bs4
import requests as rq # get url
import re
from datetime import date, time, datetime

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.expand_frame_repr', False)
pd.set_option('max_colwidth', None)

# https://www.the-numbers.com/person/41500401-Robert-Downey-Jr#tab=acting
df_starsAct = pd.read_excel ('./Prj_Data/ImdbScrapingData/StarEarningsv1_NoM
df_starswrter = pd.read_excel ('./Prj_Data/ImdbScrapingData/StarEarningsv1_N
df_starsDir = pd.read_excel ('./Prj_Data/ImdbScrapingData/StarEarningsv1_NoM
df_starsProd = pd.read_excel ('./Prj_Data/ImdbScrapingData/StarEarningsv1_No
```

In [ ]:
```
1
```

In [74]:

```
dfmovies = pd.DataFrame(columns=['Actor', 'Born','ReleaseDate', "Title", "Rc

# df_starsAct_1050_1500 = df_starsAct.iloc[1051:]
# tab=technical
#  "#tab=acting"

for index, row in df_starswrter.iterrows():
    row["Link"]
    currentname = row["Name"]
    currentlink = row["Link"]
    currentlink = currentlink + "#tab=technical"

    # ['ReleaseDate','Title','Role','DomesticBox Office','InternationalBox C
#     the_getString = 'https://www.the-numbers.com/person/41500401-Robert-Do
    the_getString = currentlink
#     print(the_getString)
    Thecols = []
    Themovies = []

    r=rq.get(the_getString)
    # 'html.parser'
    # p=bs(r.text,'lxml')
    p=bs(r.text,'html.parser')


    # BornOn=p.find("table", id="all_acting_credits")
    bornOns = p.find_all("a", href=re.compile("/on-this-day/"))
    for bornon in bornOns:
    #     print(bornon.get_text())
        CurrentBornOn = str(bornon.get_text())



#     tbody=p.find("table", id="all_acting_credits")
    tbody=p.find("table", id="all_technical_credits")
    trs = tbody.find_all("tr")
    # print(tbody)
    for tr in trs:
    #     print(tr)
    #     print("loop")
    #     if tr.find_all("th"):
        ths = tr.find_all("th")
        for th in ths:
            Thecols.append(th.text)
    #             print(th.text)
            Thecols = Thecols


        Themovies.append(currentname)
        Themovies.append(CurrentBornOn)
        if tr.find_all("td"):
            tds = tr.find_all("td")
    #         print(tds)

    #         print(f' lend of td {len(tds)}')
            counter = 1
```

```
57
58              for td in tds:
59                  if counter <len(tds):
60
61      #                 print(f' the counter = {counter}')
62      #                 print(td.text)
63                      if not td.text:
64                          text = "NA"
65                      elif td.text:
66                          text = td.text
67      #                 print(text)
68                      Themovies.append(text)
69      #                 print(f' the counter = {counter}')
70
71                  elif counter == len(tds):
72
73      #                 print("inside >6")
74                      counter = 0
75      #                 print(f' the LENGTH {len(Themovies)}')
76
77
78                      additionalColsToAdd = len(dfmovies.columns) - len(Themov
79
80
81                      if additionalColsToAdd == 1:
82      #                     print("appending1")
83                          Themovies.append("NA")
84      #
85
86                      if additionalColsToAdd == 2:
87      #                     print("appending2")
88                          Themovies.append("NA")
89                          themovies.append("NA")
90
91
92                      if len(Themovies) == len(dfmovies.columns):
93      #                     print(Themovies)
94      #                     dfmovies.loc[len(dfmovies)] = ['Dec 22, 2021', 'Si
95      #                     themovies = ['Dec 22, 2021', 'Sing 2', 'Baxter', '
96      #                     res = dict(zip(Thecols, Themovies))
97                          x = pd.Series(Themovies, index = dfmovies.columns)
98      #                   dfmovies = dfmovies.append(res, ignore_index=True)
99      #                     dfmovies = dfmovies.append(x, ignore_index=True)
100                         dfmovies.loc[len(dfmovies)] = Themovies
101     #                     print(dfmovies)
102                     Themovies = []
103                 counter  = counter +1
104     #           print("loop next")
105     #             dfmovies["Actor"] =
106 # print(dfmovies)
```

...

In [75]:
```
1 dfmovies.info()
```

...

In [72]:
```
1 dfmovies.tail(100)
```

...

In [ ]:
```
1
```

In [77]:
```
1 dfmovies.to_excel("starActorMovies_Writer.xlsx")
```

In [52]:
```
1 dfmovies.tail()
```

...

In [ ]:
```
1
```

In [ ]:
```
1
```