In [2]:

```
#importing libraries
from urllib.request import Request, urlopen
from bs4 import BeautifulSoup as soup
import pandas as pd
import matplotlib as plt
import seaborn as sns
import time
```

Scrapping 150 movies data from given url

In [61]:

```python
#Scrapping 150 movies data from following url
movies_name = []
movies_rating = []
movies_genre = []
movies_release_date = []
movies_runtime = []
movies_director = []
movies_links = []
movies_budget = []
movies_revenue = []
for i in range(1,151):
    main_url = 'https://www.themoviedb.org/movie?page=' + str(i)
    req = Request(main_url , headers={'User-Agent': 'Mozilla/5.0'})
    webpage = urlopen(req).read()
    page_soup = soup(webpage, "html.parser")
    time.sleep(0.2)
    soup_body = page_soup.body
    print(i,main_url)
    for j in range(0,1):
        a_tag_movie_link = soup_body.find_all('a',class_='image')
        href_data = a_tag_movie_link[j].get('href')
        movie_title = a_tag_movie_link[j].get('title')
        movie_url = 'https://www.themoviedb.org/' + str(href_data)
        print(movie_url)
        req_jloop = Request(movie_url , headers={'User-Agent': 'Mozilla/5.0'
        webpage_jloop = urlopen(req_jloop).read()
        page_soup_jloop = soup(webpage_jloop, "html.parser")
        soup_body_jloop = page_soup_jloop.body
        page_wrap_class = soup_body_jloop.find_all('div',class_='page_wrap m
        try :
            release_span = page_wrap_class[j].find_all('span',class_='releas
            release_text = release_span[0].get_text()
            genres_span = page_wrap_class[j].find_all('span',class_='genres'
            genres_text = genres_span[0].get_text()
            runtime_span = page_wrap_class[j].find_all('span',class_='runtim
            runtime_text = runtime_span[0].get_text()
            li_profile = page_wrap_class[0].find_all('div',class_='user_scor
            rating = li_profile[0].get('data-percent')
            money_data = page_wrap_class[0].find_all('section',class_='facts
            money_text = money_data[0].find_all('p')
            budget_value=money_text[2].text
            budget = budget_value.split()[1]
            #print(budget)
            revenue_value =money_text[3].text
            revenue = revenue_value.split()[1]
            #print(revenue)
            for k in range(0,1):
                li_profile = page_wrap_class[k].find_all('li',class_='profil
                dr = li_profile[k]
                director_text = (dr.find('a').text)
        except :
            pass
        print(j,movie_url,movie_title,release_text,genres_text,runtime_text,
        movies_name.append(movie_title)
        movies_rating.append(rating)
        movies_genre.append(genres_text.strip())
        movies_release_date.append(release_text.strip())
        movies_runtime.append(runtime_text.strip())
        movies_director.append(director_text)
        movies_links.append(movie_url)
        movies_budget.append(budget)
        movies_revenue.append(revenue)
        time.sleep(0.1)
#creating dataFrame
df = pd.DataFrame({
    'Name' : movies_name,
    'Rating' : movies_rating,
    'Genre' : movies_genre,
    'Release date' : movies_release_date,
    'Runtime' : movies_runtime,
    'Director' : movies_director,
    'Budget ($)' : movies_budget,
    'Revenue ($) ' : movies_revenue,
    'Url' : movies_links
})
#removing special character
df['Genre'] = df['Genre'].map(str).apply(lambda x: x.encode('utf-8').decode(
#print(df)
#converting dataframe into CSV
time.sleep(0.1)
df.to_csv('Movies scrapped data.csv',index = False)
print('CSV Successfully Created')

```

In [ ]:

```

```