

# Pull Data from OMDb API

June 20, 2022

## 1 Packages

```
[ ]: import pandas as pd
import numpy as np
import requests
from requests import TooManyRedirects
import re
import omdb
import time
from collections import Counter, defaultdict
```

## 2 API Set-up

```
[ ]: api_key = "2831f563" # Jimmy's key - limit 1000 calls a day
omdb.set_default('apikey', api_key)

# include tomatoes data by default
#omdb.set_default('tomatoes', True)
```

## 3 Pulling Data from API

```
[ ]: # set timeout of 5 seconds for this request
# Pull 500 pages of movies (5000 movies) with the word 'one' in the title
imdb_ids = []
for i in range(1,501):
    year_df = pd.DataFrame(omdb.search_movie('one', page=i, timeout=5))
    # Store the ids in a list
    imdb_ids.append(year_df['imdb_id'].tolist())
imdb_ids = sum(imdb_ids, [])
```

```
[ ]: # Use ids to obtain movie information -
movies_info = pd.DataFrame([omdb.imdbid(i) for i in imdb_ids])

# View the dataframe
movies_info.sample(15)
```

		title	year	rated \
3465		One More River	1964	N/A
4669	The Mysterious Death of the Man with One Hand		2020	N/A
2459		One of These Days	1997	N/A
2029		Murder Act One	2003	N/A
2787		One More Spring	1992	N/A
3710		The Verse: One Night Man	2016	N/A
3008		One Kiss	1929	N/A
195	Zatoichi and the One-Armed Swordsman		1971	Not Rated
3977		One Way Single	2013	N/A
3054		From One Place to Another	2019	N/A
3545		One Step at a Time	2017	N/A
3836	Dream with Me: One Year in America's Public Sc...		2012	N/A
1048		WWE: DX: One Last Stand	2011	PG
4161		A Song for No One	2011	N/A
131		One Step Behind the Seraphim	2017	N/A

	released	runtime	genre \
3465	24 Apr 1963	60 min	Documentary
4669	N/A	8 min	Animation, Short
2459	N/A	10 min	Short
2029	27 Sep 2003	11 min	Short, Crime
2787	N/A	N/A	Comedy
3710	02 Dec 2016	4 min	Short, Music
3008	01 Jun 1929	3 min	Short, Music
195	13 Jan 1971	94 min	Action, Adventure, Drama
3977	N/A	7 min	Short, Comedy, Drama
3054	N/A	N/A	Documentary, Short
3545	30 Oct 2017	N/A	N/A
3836	N/A	23 min	Documentary, Short, Biography, News
1048	05 Apr 2011	118 min	Action, Sport
4161	05 Aug 2011	15 min	Short, Musical
131	22 Sep 2017	150 min	Drama

	director \
3465	Beryl Fox, Douglas Leiterman
4669	L. Michael Mansfield
2459	Scott Umsteadt
2029	Steven Warrington
2787	Kirk Browning
3710	Mathieu Perrault Lapierre, Simon Plante
3008	David Aylott, E.F. Symmons
195	Kimiyoshi Yasuda
3977	Andrew Mitrak
3054	Michael Robayo
3545	Lara Goodman
3836	Kimberly Goodman, Henry Hebert

1048	Kevin Dunn
4161	Ryan Glista
131	Daniel Sandu

	writer \
3465	N/A
4669	L. Michael Mansfield
2459	Scott Umsteadt
2029	Steven Warrington
2787	Robert Nathan (novel), J. Ranelli (teleplay)
3710	N/A
3008	N/A
195	Kan Shimozawa, Takayuki Yamada, Kimiyoshi Yasuda
3977	Robert J. Lee
3054	N/A
3545	N/A
3836	N/A
1048	N/A
4161	Ryan Glista
131	Daniel Sandu

	actors \
3465	Ralph Bunche, James Meredith, James R. Venable...
4669	Bobby Butler, Moe Rock
2459	Scott Umsteadt, Julie Staley, Dennis Bingham
2029	Mark Intrieri, Anna Bartos, Michael N. Malfita...
2787	Kimberli Colbourne, Jason Robards
3710	Pierre-François Bouffard, Damien Robert, Cathe...
3008	N/A
195	Shintarô Katsu, Jimmy Wang Yu, Watako Hamaki
3977	Jay Myers, Hope Shanthi
3054	Anne Humanfeld
3545	N/A
3836	N/A
1048	Bob Barker, Mark Calaway, John Cena, Mike Chioda
4161	Sophie Dillon, Rita Frenzel, Ryan Glista
131	Stefan Iancu, Vlad Ivanov, Toto Dumitrescu

	plot ... metascore \
3465	A film essay, filmed in the cinema vérité styl... .. N/A
4669	When a lonely detective investigating an unsol... .. N/A
2459	N/A ... N/A
2029	A desperate young man hatches a dark plan to a... .. N/A
2787	N/A ... N/A
3710	N/A ... N/A
3008	N/A ... N/A
195	Two of Japan and China's greatest heroic sword... .. N/A

3977	Two lone travelers cross paths on a red-eye co...	...	N/A
3054	A short documentary focused on the process of ...	...	N/A
3545		N/A ...	N/A
3836	Dream with Me uncovers the deepest cracks of h...	...	N/A
1048	They are perhaps the most influential duo in s...	...	N/A
4161	The story of a lonely musician and the day his...	...	N/A
131	A group of freshmen in an orthodox college are...	...	N/A

	imdb_rating	imdb_votes	imdb_id	type	dvd	box_office	\
3465	N/A	N/A	tt0236552	movie	N/A	N/A	
4669	N/A	N/A	tt12343664	movie	N/A	N/A	
2459	8.8	5	tt0119835	movie	N/A	N/A	
2029	6.9	10	tt0384360	movie	N/A	N/A	
2787	N/A	N/A	tt2245035	movie	N/A	N/A	
3710	N/A	N/A	tt6295054	movie	N/A	N/A	
3008	N/A	N/A	tt1321450	movie	N/A	N/A	
195	7.2	1,540	tt0066602	movie	30 Nov 2016	N/A	
3977	N/A	N/A	tt2619772	movie	N/A	N/A	
3054	N/A	N/A	tt10671708	movie	N/A	N/A	
3545	N/A	N/A	tt9066062	movie	N/A	N/A	
3836	N/A	N/A	tt2418206	movie	N/A	N/A	
1048	7.3	68	tt1853732	movie	N/A	N/A	
4161	N/A	N/A	tt7543174	movie	N/A	N/A	
131	8.2	3,070	tt5335314	movie	05 Jan 2017	N/A	

	production	website	response
3465	N/A	N/A	True
4669	N/A	N/A	True
2459	N/A	N/A	True
2029	N/A	N/A	True
2787	N/A	N/A	True
3710	N/A	N/A	True
3008	N/A	N/A	True
195	N/A	N/A	True
3977	N/A	N/A	True
3054	N/A	N/A	True
3545	N/A	N/A	True
3836	N/A	N/A	True
1048	N/A	N/A	True
4161	N/A	N/A	True
131	N/A	N/A	True

[15 rows x 25 columns]

```
[ ]: # Dimensions of the data frame
movies_info.shape
```

```
[ ]: (5000, 25)
```

```
[ ]: # Columns of the data frame  
movies_info.columns
```

```
[ ]: Index(['title', 'year', 'rated', 'released', 'runtime', 'genre', 'director',  
         'writer', 'actors', 'plot', 'language', 'country', 'awards', 'poster',  
         'ratings', 'metascore', 'imdb_rating', 'imdb_votes', 'imdb_id', 'type',  
         'dvd', 'box_office', 'production', 'website', 'response'],  
         dtype='object')
```

```
[ ]: # Write to a csv  
movies_info.to_csv('Raw Movie Data.csv', index= False)
```

Created in Deepnote