

Pull Data from OMDB API

June 20, 2022

1 Final Project - Movie Reviews Analysis

- Team 13: Jimmy Nguyen, Dallin Munger, Tyler Wolff

2 Packages

```
[3]: 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
```

3 Pulling Data from API

```
[4]: # 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, [])
```

```
[5]: # 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)
```

```
[5]:
```

	title	year	rated	\
4893	One Nation Under Mike	2015	N/A	
2163	Before One More Day	2019	N/A	
3108	Secret Access: Air Force One	2008	N/A	

4884	ONE Fighting Championship 3: War of Lions	2012	N/A
4931	One Liners	2008	N/A
4803	One Last Love Song	2009	N/A
2263	Reclamation - chapter one - a Star Trek fan pr...	2019	N/A
4157	Batman: Year One	2019	N/A
3262	A Thousand and One Nights	1941	N/A
3769	Five times smile plus one	2010	N/A
2040	One Minute for Conductors	2013	N/A
400	One Last Shot	1998	TV-MA
4777	Table for One	2020	N/A
190	One Piece: Episode of Alabasta - The Desert Pr...	2007	PG-13
2638	Yum Ciao News: The Revenge of the Number One Fan	2011	N/A

	released	runtime	genre \
4893	N/A	11 min	Short, Comedy
2163	06 Aug 2019	16 min	Short, Drama
3108	N/A	N/A	Documentary
4884	31 Mar 2012	N/A	Sport
4931	09 Nov 2008	2 min	Short
4803	N/A	N/A	Short, Drama
2263	22 Dec 2019	N/A	Short, Sci-Fi
4157	01 Nov 2019	N/A	Action, Horror, Mystery
3262	N/A	N/A	Adventure, Comedy, Drama
3769	03 May 2010	N/A	Short
2040	23 Nov 2013	87 min	Documentary, Music
400	30 Oct 2017	30 min	Short, Comedy, Drama
4777	20 Aug 2020	N/A	Documentary, Short
190	03 Mar 2007	90 min	Animation, Action, Adventure
2638	01 Feb 2011	7 min	Short, Comedy

	director \
4893	Vera Abrams, August Hartwell, Jessica Miner
2163	Renata Abreu
3108	Peter Schnall
4884	N/A
4931	Tim Best
4803	Keith Mackin, John Reck
2263	Greg Ogles
4157	Jose Luis Garcia Baylon
3262	Togo Mizrahi
3769	Loris Arduino
2040	Angel Esteban, Elena Goatelli
400	Mike Clattenburg
4777	Jenny Gao
190	Takahiro Imamura
2638	Harrison Berenger

	writer \
4893	Jillian Sanders
2163	Renata Abreu
3108	Don Campbell
4884	N/A
4931	N/A
4803	Keith Mackin, John Reck
2263	Greg Ogles
4157	Darren Aronofsky (based on the screenplay by),...
3262	N/A
3769	Loris Arduino
2040	Angel Esteban (original idea), Elena Goatelli ...
400	Mike Clattenburg, John Paul Tremblay, Robb Wells
4777	N/A
190	Eiichiro Oda, Hirohiko Uesaka
2638	Harrison Berenger

	actors \
4893	Phil Abrams, John Glouchevitch, Jessica Miner,...
2163	Thais Castro, Flávio Dias, Thalita Franco
3108	N/A
4884	Kevin Belington, Jan Kai Chee, Nicole Chua
4931	Tim Best
4803	Allan Richardson, Matt Watkins, Rholda Wilson
2263	Haley Barnes, Crimson Kromer, Hassel Kromer, R...
4157	Alan Acosta, Alejandro Fierro
3262	Ali Al-Kassar, Aqila Ratib
3769	Luigi Caso, Andrea Ciardulli, Nicola Cuomo
2040	N/A
400	Robb Wells, John Paul Tremblay, John Dunsworth
4777	N/A
190	Charles Baker, Troy Baker, Anthony Bowling
2638	Harrison Berenger, Samuel Berenger, Tom Crane

	plot ... metascore \
4893	N/A ... N/A
2163	One family suffers from the death of their mot... ... N/A
3108	N/A ... N/A
4884	N/A ... N/A
4931	Movies have permeated the consciousness of Ame... ... N/A
4803	N/A ... N/A
2263	A diabolical foe resurfaces during the Federat... ... N/A
4157	The story recounts the beginnings of James Gor... ... N/A
3262	N/A ... N/A
3769	N/A ... N/A
2040	More than 130 young conductors participate in N/A
400	Two best friends, Rob and GW, spend a night ou... ... N/A

4777	Against all odds, an aging table tennis icon r...	...	N/A
190	A re-telling of the Alabaster Arc from One Pie...	...	N/A
2638		N/A ...	N/A

	imdb_rating	imdb_votes	imdb_id	type	dvd	box_office	\
4893	N/A	N/A	tt5541410	movie	N/A	N/A	
2163	N/A	8	tt11597840	movie	N/A	N/A	
3108	N/A	N/A	tt1567417	movie	N/A	N/A	
4884	N/A	N/A	tt6217432	movie	N/A	N/A	
4931	N/A	N/A	tt1337526	movie	N/A	N/A	
4803	N/A	N/A	tt1387245	movie	N/A	N/A	
2263	7.7	7	tt11534274	movie	N/A	N/A	
4157	N/A	N/A	tt7829938	movie	N/A	N/A	
3262	N/A	N/A	tt0280411	movie	N/A	N/A	
3769	N/A	N/A	tt9878362	movie	N/A	N/A	
2040	8.5	10	tt3314488	movie	N/A	N/A	
400	6.5	523	tt0380601	movie	N/A	N/A	
4777	N/A	N/A	tt12788300	movie	N/A	N/A	
190	6.9	1,631	tt1037116	movie	03 Dec 2020	\$6,587	
2638	N/A	N/A	tt11651744	movie	N/A	N/A	

	production	website	response
4893	N/A	N/A	True
2163	N/A	N/A	True
3108	N/A	N/A	True
4884	N/A	N/A	True
4931	N/A	N/A	True
4803	N/A	N/A	True
2263	N/A	N/A	True
4157	N/A	N/A	True
3262	N/A	N/A	True
3769	N/A	N/A	True
2040	N/A	N/A	True
400	N/A	N/A	True
4777	N/A	N/A	True
190	N/A	N/A	True
2638	N/A	N/A	True

[15 rows x 25 columns]

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

```
[6]: (5000, 25)
```

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

```
[7]: 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')
```

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

Created in Deepnote