

# TP Data Exploration

April 17, 2018

```
In [1]: import pandas as pd
```

```
In [2]: credits= pd.read_csv('./the-movies-dataset/credits.csv', sep= ',')
```

```
In [3]: credits.head()
```

```
Out[3]:
```

		cast	\
0	[{'cast_id': 14, 'character': 'Woody (voice)',...		
1	[{'cast_id': 1, 'character': 'Alan Parrish', '...		
2	[{'cast_id': 2, 'character': 'Max Goldman', 'c...		
3	[{'cast_id': 1, 'character': "Savannah 'Vannah...		
4	[{'cast_id': 1, 'character': 'George Banks', '...		

  

		crew	id
0	[{'credit_id': '52fe4284c3a36847f8024f49', 'de...		862
1	[{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...		8844
2	[{'credit_id': '52fe466a9251416c75077a89', 'de...		15602
3	[{'credit_id': '52fe44779251416c91011acb', 'de...		31357
4	[{'credit_id': '52fe44959251416c75039ed7', 'de...		11862

```
In [4]: keywords= pd.read_csv('./the-movies-dataset/keywords.csv', sep= ',')
```

```
In [5]: keywords.head()
```

```
Out[5]:
```

	id	keywords
0	862	[{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1	8844	[{'id': 10090, 'name': 'board game'}, {'id': 1...
2	15602	[{'id': 1495, 'name': 'fishing'}, {'id': 12392...
3	31357	[{'id': 818, 'name': 'based on novel'}, {'id':...
4	11862	[{'id': 1009, 'name': 'baby'}, {'id': 1599, 'n...

```
In [6]: ratings= pd.read_csv('./the-movies-dataset/ratings.csv', sep= ',')
```

```
In [7]: fields= ['id', 'genres', 'title', 'original_title', 'vote_average', 'vote_count']
metadata= pd.read_csv('./the-movies-dataset/movies_metadata.csv', sep= ',', low_memory=
```

```
In [8]: credits['cast'][0]
```

```
Out[8]: "[{'cast_id': 14, 'character': 'Woody (voice)', 'credit_id': '52fe4284c3a36847f8024f95
```

```

In [9]: import ast

In [10]: ast.literal_eval(credits['cast'][0])[0]['name']

Out[10]: 'Tom Hanks'

In [11]: credits['crew'][0]

Out[11]: '[{\''credit_id\': \''52fe4284c3a36847f8024f49\'', \''department\': \''Directing\'', \''gend

In [12]: ast.literal_eval(credits['crew'][0])[0]['name']

Out[12]: 'John Lasseter'

In [13]: ast.literal_eval(keywords['keywords'][0])

Out[13]: [{'id': 931, 'name': 'jealousy'},
          {'id': 4290, 'name': 'toy'},
          {'id': 5202, 'name': 'boy'},
          {'id': 6054, 'name': 'friendship'},
          {'id': 9713, 'name': 'friends'},
          {'id': 9823, 'name': 'rivalry'},
          {'id': 165503, 'name': 'boy next door'},
          {'id': 170722, 'name': 'new toy'},
          {'id': 187065, 'name': 'toy comes to life'}]

In [14]: keywords['id'] = keywords['id'].astype('int')
         credits['id'] = credits['id'].astype('int')
         metadata['id'] = metadata['id'].astype('int')

```

-----

ValueError

Traceback (most recent call last)

```

<ipython-input-14-cb20ad256598> in <module>()
    1 keywords['id'] = keywords['id'].astype('int')
    2 credits['id'] = credits['id'].astype('int')
----> 3 metadata['id'] = metadata['id'].astype('int')

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\util\_d
116         else:
117             kwargs[new_arg_name] = new_arg_value
--> 118         return func(*args, **kwargs)
119         return wrapper
120     return _deprecate_kwarg

```

```

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\gen
4002         # else, only a single dtype is given
4003         new_data = self._data.astype(dtype=dtype, copy=copy, errors=errors,
-> 4004                                     **kwargs)
4005         return self._constructor(new_data).__finalize__(self)
4006

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\in
3460
3461     def astype(self, dtype, **kwargs):
-> 3462         return self.apply('astype', dtype=dtype, **kwargs)
3463
3464     def convert(self, **kwargs):

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\in
3327
3328         kwargs['mgr'] = self
-> 3329         applied = getattr(b, f)(**kwargs)
3330         result_blocks = _extend_blocks(applied, result_blocks)
3331

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\in
542     def astype(self, dtype, copy=False, errors='raise', values=None, **kwargs):
543         return self._astype(dtype, copy=copy, errors=errors, values=values,
--> 544                             **kwargs)
545
546     def _astype(self, dtype, copy=False, errors='raise', values=None,

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\in
623
624         # _astype_nansafe works fine with 1-d only
--> 625         values = astype_nansafe(values.ravel(), dtype, copy=True)
626         values = values.reshape(self.shape)
627

c:\users\kimbe\appdata\local\programs\python\python36\lib\site-packages\pandas\core\dt
690     elif arr.dtype == np.object_ and np.issubdtype(dtype.type, np.integer):
691         # work around NumPy brokenness, #1987
--> 692         return lib.astype_intsafe(arr.ravel(), dtype).reshape(arr.shape)
693
694     if dtype.name in ("datetime64", "timedelta64"):

```

```
pandas\_libs\lib.pyx in pandas._libs.lib.astype_intsafe()
```

```
pandas/_libs/src\util.pxd in util.set_value_at_unsafe()
```

```
ValueError: invalid literal for int() with base 10: '1997-08-20'
```

```
In [15]: import string
```

```
In [16]: length= len(metadata['id'])
```

```
In [17]: l=[]
         for i in range(length):
             for j in metadata['id'][i]:
                 if j not in string.digits:
                     l.append(i)
                     break
```

```
In [18]: print (l)
```

```
[19730, 29503, 35587]
```

```
In [19]: metadata= metadata.drop(l)
```

```
In [20]: metadata['id'].shape
```

```
Out[20]: (45463,)
```

```
In [21]: keywords['id'] = keywords['id'].astype('int')
         credits['id'] = credits['id'].astype('int')
         metadata['id'] = metadata['id'].astype('int')
```

```
In [22]: allFeatures = metadata.merge(credits, on='id')
         allFeatures = allFeatures.merge(keywords, on='id')
```

```
In [23]: allFeatures.head(2)
```

```
Out[23]:
```

	adult		belongs_to_collection		budget	
0	False	{'id': 10194, 'name': 'Toy Story Collection', ...			30000000	
1	False				NaN	65000000

  

		genres	
0	{'id': 16, 'name': 'Animation'}, {'id': 35, '...		
1	{'id': 12, 'name': 'Adventure'}, {'id': 14, '...		

  

	homepage	id	imdb_id	original_language	
0	http://toystory.disney.com/toy-story	862	tt0114709		en

```

1          NaN  8844  tt0113497          en

original_title          overview \
0      Toy Story  Led by Woody, Andy's toys live happily in his ...
1      Jumanji  When siblings Judy and Peter discover an encha...

... \
0      ...
1      ...

spoken_languages  status \
0      [{'iso_639_1': 'en', 'name': 'English'}] Released
1      [{'iso_639_1': 'en', 'name': 'English'}, {'iso... Released

tagline  title  video  vote_average \
0      NaN  Toy Story  False  7.7
1  Roll the dice and unleash the excitement!  Jumanji  False  6.9

vote_count          cast \
0      5415.0  [{'cast_id': 14, 'character': 'Woody (voice)',...
1      2413.0  [{'cast_id': 1, 'character': 'Alan Parrish', '...

crew \
0  [{'credit_id': '52fe4284c3a36847f8024f49', 'de...
1  [{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...

keywords
0  [{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1  [{'id': 10090, 'name': 'board game'}, {'id': 1...

[2 rows x 27 columns]

```

```
In [24]: allFeatures['genres'][0]
```

```
Out[24]: "[{'id': 16, 'name': 'Animation'}, {'id': 35, 'name': 'Comedy'}, {'id': 10751, 'name': 'Fantasy'}]"
```

```
In [25]: allFeatures['cast'][0]
```

```
Out[25]: "[{'cast_id': 14, 'character': 'Woody (voice)', 'credit_id': '52fe4284c3a36847f8024f49', 'name': 'Woody', 'order': 1}]"
```

```
In [26]: measures = ['cast', 'crew', 'keywords', 'genres']
allFeatures['genres'] = allFeatures['genres'].apply(ast.literal_eval)
```

```
In [27]: allFeatures['cast'] = allFeatures['cast'].apply(ast.literal_eval)
```

```
In [28]: allFeatures['crew'] = allFeatures['crew'].apply(ast.literal_eval)
```

```
In [29]: allFeatures['keywords'] = allFeatures['keywords'].apply(ast.literal_eval)
```

```
In [30]: allFeatures['keywords'][0]
```

```
Out[30]: [{'id': 931, 'name': 'jealousy'},
          {'id': 4290, 'name': 'toy'},
          {'id': 5202, 'name': 'boy'},
          {'id': 6054, 'name': 'friendship'},
          {'id': 9713, 'name': 'friends'},
          {'id': 9823, 'name': 'rivalry'},
          {'id': 165503, 'name': 'boy next door'},
          {'id': 170722, 'name': 'new toy'},
          {'id': 187065, 'name': 'toy comes to life'}]
```

```
In [31]: def getTop3(feature):
          if isinstance(feature, list):
              lst= [i['name'] for i in feature]
              if len(lst) > 3: return lst[:3]
              else: return lst
          return []
```

```
In [32]: allFeatures['cast']= allFeatures['cast'].apply(getTop3)
          allFeatures['keywords']= allFeatures['keywords'].apply(getTop3)
          allFeatures['genres']= allFeatures['genres'].apply(getTop3)
```

```
In [33]: a= allFeatures['genres'][0]
```

```
In [34]: a
```

```
Out[34]: ['Animation', 'Comedy', 'Family']
```

```
In [35]: def lowerStripSpace(feature):
          l= []
          for i in feature:
              i= i.replace(" ", "")
              l.append(i.lower())
          return l
```

```
In [36]: allFeatures['cast']= allFeatures['cast'].apply(lowerStripSpace)
          allFeatures['keywords']= allFeatures['keywords'].apply(lowerStripSpace)
          allFeatures['genres']= allFeatures['genres'].apply(lowerStripSpace)
```

```
In [37]: allFeatures['keywords']
```

```
Out[37]: 0          [jealousy, toy, boy]
          1    [boardgame, disappearance, basedonchildren'sbook]
          2          [fishing, bestfriend, duringcreditsstinger]
          3    [basedonnovel, interracialrelationship, single...
          4          [baby, midlifecrisis, confidence]
          5          [robbery, detective, bank]
          6    [paris, brotherbrotherrelationship, chauffeur]
          7          []
```

8 [terrorist, hostage, explosive]  
 9 [cuba, falselyaccused, secretidentity]  
 10 [whitehouse, usapresident, newlove]  
 11 [dracula, spoof]  
 12 [wolf, dog-sleddingrace, alaska]  
 13 [usapresident, presidentialelection, watergate...]  
 14 [exoticisland, treasure, map]  
 15 [poker, drugabuse, 1970s]  
 16 [bowling, basedonnovel, servant]  
 17 [hotel, newyear'seve, witch]  
 18 [africa, indigenous, humananimalrelationship]  
 19 [brotherbrotherrelationship, subway, newyorkcity]  
 20 [gambling, miami, basedonnovel]  
 21 [policebrutality, psychology, policeoperation]  
 22 [competition, assassination, cia]  
 23 [deadanimal, shottodeath, teenageboy]  
 24 [individual, prostitute, alcohol]  
 25 [shakespeare, jealousy, friends]  
 26 [1970s, femalefriendship, comingofage]  
 27 [captain, napoleonicwars, oldflame]  
 28 [clone, dream, island]  
 29 [servant, cabaret, shanghai]  
 ...  
 46598 []  
 46599 [love, teenager, lgbt]  
 46600 [biography, singer]  
 46601 [pauladupree]  
 46602 [overweight, independentfilm, prejudice]  
 46603 [militaryschool]  
 46604 [policeraid]  
 46605 [laboratory, madscientist, disembodiedhead]  
 46606 [magic, silentfilm, trickfilm]  
 46607 [silentfilm, acab]  
 46608 [tenant, silentfilm]  
 46609 []  
 46610 []  
 46611 [halloween, winniethepooh]  
 46612 []  
 46613 [wrestling, sport, silentfilm]  
 46614 []  
 46615 [physicalabuse, sexualassault]  
 46616 []  
 46617 []  
 46618 [revenge, murder, serialkiller]  
 46619 [blairwitch]  
 46620 [witch, mythology, legend]  
 46621 []  
 46622 []

```

46623                                     [tragiclove]
46624                                [artist, play, pinoy]
46625                                     []
46626                                     []
46627                                     []
Name: keywords, Length: 46628, dtype: object

```

```

In [38]: def searchString(movie):
          return ' '.join(movie['keywords']) + ' ' + ' '.join(movie['cast']) + ' ' \
          + ' '.join(movie['genres'])

```

```

In [39]: allFeatures['search'] = allFeatures.apply(searchString, axis='columns')

```

```

In [40]: allFeatures['search']

```

```

Out[40]: 0      jealousy toy boy tomhanks timallen donrickles ...
        1      boardgame disappearance basedonchildren'sbook ...
        2      fishing bestfriend duringcreditsstinger walter...
        3      basedonnovel interracialrelationship singlemot...
        4      baby midlifecrisis confidence stevemartin dian...
        5      robbery detective bank alpacino robertdeniro v...
        6      paris brotherbrotherrelationship chauffeur har...
        7      jonathantaylorthomas bradrenfro rachaelleighb...
        8      terrorist hostage explosive jean-claudevandamm...
        9      cuba falselyaccused secretidentity piercebrosn...
       10      whitehouse usapresident newlove michaeldouglas...
       11      dracula spoof leslienielson melbrooks amyyasbe...
       12      wolf dog-sleddingrace alaska kevinbacon bobhos...
       13      usapresident presidentialelection watergatesca...
       14      exoticisland treasure map geenadavis matthewmo...
       15      poker drugabuse 1970s robertdeniro sharonstone...
       16      bowling basedonnovel servant katewinslet emmat...
       17      hotel newyear'seve witch timroth antoniobander...
       18      africa indigenous humananimalrelationship jimc...
       19      brotherbrotherrelationship subway newyorkcity ...
       20      gambling miami basedonnovel johntravolta geneh...
       21      policebrutality psychology policeoperation sig...
       22      competition assassination cia sylvesterstallon...
       23      deadanimal shottodeath teenageboy marysteenbur...
       24      individual prostitute alcohol nicolascage elis...
       25      shakespeare jealousy friends laurencefishburne...
       26      1970s femalefriendship comingofage christinari...
       27      captain napoleonicwars oldflame amandaroot cia...
       28      clone dream island ronperlman dominiquepinon j...
       29      servant cabaret shanghai gongli libao-tian wan...
          ...
46598      laurenholly lukeperry stevebacic action myste...
46599      love teenager lgbt family animation romance
46600      biography singer martijnfischer hadewychminis ...

```



```

46601    pauladupree acquanetta evelynankers j.carrolna...
46602    overweight independentfilm prejudice kaleycuoc...
46603    militaryschool hilaryduff christycarlsonromano...
46604                                     policeraid
46605    laboratory madscientist disembodiedhead george...
46606    magic silentfilm trickfilm georgesméliès fantasy
46607    silentfilm acab georgesméliès fantasy comedy
46608    tenant silentfilm andréméliès georgesméliès fa...
46609                                     georgesméliès
46610
46611    halloween winniethepooh jimmybennett petercull...
46612                                     georgesméliès fantasy action thriller
46613    wrestling sport silentfilm jeanned'alcy george...
46614                                     documentary
46615    physicalabuse sexualassault sridevikapoor saja...
46616    markieadams robertoaguires tinaarning comedy d...
46617                                     giulio brogi renatocestiè vitocipolla
46618    revenge murder serialkiller rondohatton robert...
46619    blairwitch tonyabatemarco andrebrooks mariclar...
46620    witch mythology legend montybane lucybutler da...
46621    lisaboyle kenaland zanetapolard sciencefiction
46622    patrickbergin umathurman davidmorrissey drama...
46623    tragiclove leilahatami kouroshtahami elhamkord...
46624    artist play pinoy angelaquino perrydizon hazel...
46625    erikaeleniak adambaldwin juliedupage action d...
46626    iwanmosschuchin nathalielissenko pavelpavlov
46627
Name: search, Length: 46628, dtype: object

```

```

In [41]: def normalizeTitle(title):
          t= title.lower()
          words= t.split(' ')
          words= "".join(words)
          return words

```

```

In [42]: normalizeTitle('Harry Potter')

```

```

Out[42]: 'harrypotter'

```

```

In [43]: allFeatures['newTitle']= allFeatures['original_title'].apply(normalizeTitle)

```

```

In [44]: allFeatures['newTitle']

```

```

Out[44]: 0                                toystory
          1                                jumanji
          2                        grumpieroldmen
          3                    waitingtoexhale
          4                fatherofthebridepartii

```

5	heat
6	sabrina
7	tomandhuck
8	suddendeath
9	goldeneye
10	theamericanpresident
11	dracula:deadandlovingit
12	balto
13	nixon
14	cutthroatisland
15	casino
16	senseandsensibility
17	fourrooms
18	aceventura:whennaturecalls
19	moneytrain
20	getshorty
21	copycat
22	assassins
23	powder
24	leavinglasvegas
25	othello
26	nowandthen
27	persuasion
28	lacitydesenfantsperdus
29	
	...
46598	thefinalstorm
46599	inaheartbeat
46600	bloed,zweetentranen
46601	junglewoman
46602	tobefatlikeme
46603	cadetkelly
46604	letripotclandestin
46605	l'hommeàlatêtedecaoutchouc
46606	lescartesvivantes
46607	lesaffichesengouette
46608	lelocatairediaboliue
46609	leroidumaquillage
46610	lestransmutationsimperceptibles
46611	pooh'sheffalumphalloweenmovie
46612	l'hommeorchestre
46613	nouvellesluttensextravagantes
46614	deephearts
46615	maa
46616	themorningafter
46617	sanmicheleavevaungallo
46618	houseofhorrors
46619	shadowoftheblairwitch

```

46620          theburkittsville7
46621          cagedheat3000
46622          robinhood
46623
46624          siglongpagluluwal
46625          betrayal
46626          satanalikuyushchiy
46627          queerama
Name: newTitle, Length: 46628, dtype: object

```

```
In [46]: allFeatures[allFeatures['original_title'].duplicated()==True]
```

```

Out[46]:      adult      belongs_to_collection      budget  \
677      False                                NaN          0
678      False                                NaN          0
679      False                                NaN          0
680      False                                NaN          0
681      False                                NaN          0
682      False                                NaN          0
683      False                                NaN          0
846      False                                NaN          0
847      False                                NaN          0
848      False                                NaN          0
849      False                                NaN          0
850      False                                NaN          0
851      False                                NaN          0
852      False                                NaN          0
902      False                                NaN    2238813
944      False                                NaN          0
964      False                                NaN          4
965      False                                NaN          4
966      False                                NaN          4
967      False                                NaN          4
968      False                                NaN          4
969      False                                NaN          4
970      False                                NaN          4
1317     False                                NaN          0
1852     False                                NaN    750000
1970     False                                NaN          0
1989     False  {'id': 97459, 'name': 'The Jungle Book Collect...  4000000
2046     False                                NaN    18000000
2047     False                                NaN          0
2123     False                                NaN          0
...      ...                                ...      ...
46157    False                                NaN    4700000
46158    False  {'id': 2704, 'name': 'The Bible Collection', '...  20000000
46165    False                                NaN          0
46177    False                                NaN    60000000

```

46188	False	NaN	0
46204	False	NaN	0
46227	False	NaN	0
46228	False	NaN	0
46232	False	NaN	0
46236	False	NaN	0
46241	False	NaN	0
46281	False	NaN	0
46285	False	NaN	0
46287	False	NaN	0
46308	False	NaN	0
46309	False	NaN	0
46388	False	NaN	0
46401	False	NaN	0
46408	False	NaN	0
46433	False	NaN	0
46437	False	NaN	0
46488	False	NaN	0
46511	False	NaN	0
46568	False	NaN	0
46576	False	NaN	0
46577	False	NaN	0
46582	False	NaN	0
46616	False	NaN	0
46622	False	NaN	0
46625	False	NaN	0

	genres \
677	[drama, romance]
678	[drama, romance]
679	[drama, romance]
680	[drama, romance]
681	[drama, romance]
682	[drama, romance]
683	[drama, romance]
846	[drama]
847	[drama]
848	[drama]
849	[drama]
850	[drama]
851	[drama]
852	[drama]
902	[comedy, drama, romance]
944	[drama]
964	[drama, romance, war]
965	[drama, romance, war]
966	[drama, romance, war]
967	[drama, romance, war]

968 [drama, romance, war]  
 969 [drama, romance, war]  
 970 [drama, romance, war]  
 1317 [drama, thriller]  
 1852 [drama]  
 1970 [comedy, drama, family]  
 1989 [family, animation, adventure]  
 2046 [adventure, fantasy, comedy]  
 2047 [comedy, family, romance]  
 2123 [drama, thriller]  
 ... ..  
 46157 [drama, comedy]  
 46158 [history, drama]  
 46165 [documentary]  
 46177 [action, western, sciencefiction]  
 46188 [music, family, comedy]  
 46204 [adventure, family, fantasy]  
 46227 [music, drama, romance]  
 46228 [documentary]  
 46232 [sciencefiction]  
 46236 [romance, drama]  
 46241 [fantasy, drama]  
 46281 [comedy, animation]  
 46285 [comedy, crime]  
 46287 [drama, crime, thriller]  
 46308 [action, adventure, crime]  
 46309 [drama, romance, fantasy]  
 46388 [drama]  
 46401 [adventure, animation, comedy]  
 46408 [drama]  
 46433 [action, comedy, crime]  
 46437 []  
 46488 [drama]  
 46511 [drama, war]  
 46568 [comedy, romance]  
 46576 [documentary]  
 46577 [documentary]  
 46582 [drama, family, fantasy]  
 46616 [comedy, drama]  
 46622 [drama, action, romance]  
 46625 [action, drama, thriller]

	homepage	id	imdb_id	\
677	NaN	105045	tt0111613	
678	NaN	105045	tt0111613	
679	NaN	105045	tt0111613	
680	NaN	105045	tt0111613	
681	NaN	105045	tt0111613	

682		NaN	105045	tt0111613
683		NaN	105045	tt0111613
846		NaN	132641	tt0046468
847		NaN	132641	tt0046468
848		NaN	132641	tt0046468
849		NaN	132641	tt0046468
850		NaN	132641	tt0046468
851		NaN	132641	tt0046468
852		NaN	132641	tt0046468
902		NaN	6620	tt0047437
944		NaN	85638	tt0017350
964		NaN	22649	tt0022879
965		NaN	22649	tt0022879
966		NaN	22649	tt0022879
967		NaN	22649	tt0022879
968		NaN	22649	tt0022879
969		NaN	22649	tt0022879
970		NaN	22649	tt0022879
1317		NaN	11349	tt0055824
1852		NaN	23383	tt0040416
1970		NaN	9820	tt0120783
1989	<a href="http://movies.disney.com/the-jungle-book-1967">http://movies.disney.com/the-jungle-book-1967</a>		9325	tt0061852
2046		NaN	16081	tt0061584
2047		NaN	18331	tt0057372
2123		NaN	8208	tt0025452
...		...	...	...
46157		NaN	166624	tt2091427
46158		NaN	2331	tt0199232
46165	<a href="http://www.cityofghosts.com">http://www.cityofghosts.com</a>		428501	tt6333056
46177	<a href="http://thedarktower-movie.com">http://thedarktower-movie.com</a>		353491	tt1648190
46188		NaN	26177	tt0207972
46204		NaN	32593	tt0255605
46227	<a href="http://abc.go.com/shows/dirty-dancing">http://abc.go.com/shows/dirty-dancing</a>		444902	tt5262792
46228		NaN	54988	tt0251613
46232		NaN	401427	tt5769236
46236		NaN	146416	tt0271467
46241		NaN	465044	tt5943940
46281		NaN	130627	tt2095646
46285		NaN	105759	tt1430144
46287		NaN	99229	tt1342408
46308		NaN	463906	tt2569088
46309		NaN	63449	tt0058423
46388		NaN	459950	tt5226512
46401		NaN	423453	tt4466894
46408		NaN	280640	tt3565174
46433		NaN	69727	tt0126301
46437		NaN	76289	tt0092070
46488		NaN	120831	tt0018107

46511	NaN	63990	tt1799110
46568	NaN	98604	tt2147597
46576	NaN	75608	tt0059091
46577	NaN	275331	tt0392137
46582	NaN	18098	tt0181199
46616	NaN	420346	tt4130180
46622	NaN	30840	tt0102797
46625	NaN	67758	tt0303758

	original_language	original_title \
677	de	Das Versprechen
678	de	Das Versprechen
679	de	Das Versprechen
680	de	Das Versprechen
681	de	Das Versprechen
682	de	Das Versprechen
683	de	Das Versprechen
846	ja	Tsuma
847	ja	Tsuma
848	ja	Tsuma
849	ja	Tsuma
850	ja	Tsuma
851	ja	Tsuma
852	ja	Tsuma
902	en	Sabrina
944	en	The Scarlet Letter
964	en	A Farewell to Arms
965	en	A Farewell to Arms
966	en	A Farewell to Arms
967	en	A Farewell to Arms
968	en	A Farewell to Arms
969	en	A Farewell to Arms
970	en	A Farewell to Arms
1317	en	Cape Fear
1852	en	Hamlet
1970	en	The Parent Trap
1989	en	The Jungle Book
2046	en	Doctor Dolittle
2047	en	The Nutty Professor
2123	en	The Man Who Knew Too Much
...	...	...
46157	en	Treading Water
46158	en	Jesus
46165	en	City of Ghosts
46177	en	The Dark Tower
46188	en	Annie
46204	en	Snow White
46227	en	Dirty Dancing

46228	en	Bird on a Wire
46232	en	Elegy
46236	en	
46241	en	Abduction
46281	en	Get Out
46285	en	The Black Hole
46287	cs	Normal
46308	en	The Saint
46309	ru	
46388	en	Feed
46401	fr	Sahara
46408	fr	Geronimo
46433	en	Final Justice
46437	fr	Terminus
46488	en	Love
46511	en	Joanna
46568	ru	Zolushka
46576	en	Dead Birds
46577	en	The Hunters
46582	en	Arabian Nights
46616	en	The Morning After
46622	en	Robin Hood
46625	en	Betrayal

overview \

677	East-Berlin, 1961, shortly after the erection ...
678	East-Berlin, 1961, shortly after the erection ...
679	East-Berlin, 1961, shortly after the erection ...
680	East-Berlin, 1961, shortly after the erection ...
681	East-Berlin, 1961, shortly after the erection ...
682	East-Berlin, 1961, shortly after the erection ...
683	East-Berlin, 1961, shortly after the erection ...
846	Ten years into a marriage, the wife is disappo...
847	Ten years into a marriage, the wife is disappo...
848	Ten years into a marriage, the wife is disappo...
849	Ten years into a marriage, the wife is disappo...
850	Ten years into a marriage, the wife is disappo...
851	Ten years into a marriage, the wife is disappo...
852	Ten years into a marriage, the wife is disappo...
902	Linus and David Larrabee are the two sons of a...
944	In Puritan Boston, seamstress Hester Prynne is...
964	British nurse Catherine Barkley (Helen Hayes) ...
965	British nurse Catherine Barkley (Helen Hayes) ...
966	British nurse Catherine Barkley (Helen Hayes) ...
967	British nurse Catherine Barkley (Helen Hayes) ...
968	British nurse Catherine Barkley (Helen Hayes) ...
969	British nurse Catherine Barkley (Helen Hayes) ...
970	British nurse Catherine Barkley (Helen Hayes) ...



1317 Sam Bowden witnesses a rape committed by Max C...  
 1852 Winner of four Academy Awards, including Best ...  
 1970 Hallie Parker and Annie James are identical tw...  
 1989 The boy Mowgli makes his way to the man-villag...  
 2046 Get ready for the wildest adventure of a lifet...  
 2047 Jerry Lewis directed, co-wrote and starred in ...  
 2123 A man and his wife receive a clue to an immine...  
 ...  
 46157 At first glance, Mica seems a perfectly normal...  
 46158 The film is notable for presenting a more huma...  
 46165 With unprecedented access, this documentary fo...  
 46177 The last Gunslinger, Roland Deschain, has been...  
 46188 Eleven-year-old Annie has been living in an or...  
 46204 Snow White's mother dies during childbirth, le...  
 46227 Spending the summer at a Catskills resort with...  
 46228 Tony Palmer's film, thought lost for almost 40...  
 46232 In the future, nature has evolved to destroy h...  
 46236 A young man and woman meet by chance and exper...  
 46241 A horror comedy spoofing conspiracy theory mov...  
 46281 A doctor unrelentingly tries to convince a pho...  
 46285 Charlie, a sleep-deprived office worker accide...  
 46287 Dreaded serial killer Peter Kurten voluntarily...  
 46308 International master thief, Simon Templar, als...  
 46309 In a romantic and philosophical tale of magic ...  
 46388 Olivia and Matthew Grey, 18-year-old twins bor...  
 46401 Two clans of snakes cohabit in the desert. Bea...  
 46408 South of France. In the sultry August heat, Ge...  
 46433 A cop relies on a young car thief to take down...  
 46437 In a sort of "Mad Max" futuristic adventure, a...  
 46488 In Imperial Russia, Anna Karenina falls in lov...  
 46511 In an era in which it is difficult for the tru...  
 46568 Masha Krapivina - is yet beautiful, and not th...  
 46576 The film's title is borrowed from a Dani fable...  
 46577 An ethnographic film that documents the effort...  
 46582 Scheherezade puts herself in danger to save Su...  
 46616 The Morning After is a feature film that consi...  
 46622 Yet another version of the classic epic, with ...  
 46625 When one of her hits goes wrong, a professiona...

... \

677 ...  
 678 ...  
 679 ...  
 680 ...  
 681 ...  
 682 ...  
 683 ...  
 846 ...

847	...
848	...
849	...
850	...
851	...
852	...
902	...
944	...
964	...
965	...
966	...
967	...
968	...
969	...
970	...
1317	...
1852	...
1970	...
1989	...
2046	...
2047	...
2123	...
...	...
46157	...
46158	...
46165	...
46177	...
46188	...
46204	...
46227	...
46228	...
46232	...
46236	...
46241	...
46281	...
46285	...
46287	...
46308	...
46309	...
46388	...
46401	...
46408	...
46433	...
46437	...
46488	...
46511	...
46568	...
46576	...

46577           ...  
 46582           ...  
 46616           ...  
 46622           ...  
 46625           ...

		tagline \
677	A love, a hope, a wall.	
678	A love, a hope, a wall.	
679	A love, a hope, a wall.	
680	A love, a hope, a wall.	
681	A love, a hope, a wall.	
682	A love, a hope, a wall.	
683	A love, a hope, a wall.	
846		NaN
847		NaN
848		NaN
849		NaN
850		NaN
851		NaN
852		NaN
902	...the chauffeur's daughter who learned her st...	
944		NaN
964	Every woman who has loved will understand	
965	Every woman who has loved will understand	
966	Every woman who has loved will understand	
967	Every woman who has loved will understand	
968	Every woman who has loved will understand	
969	Every woman who has loved will understand	
970	Every woman who has loved will understand	
1317	Now he had only one weapon left - murder! ...T...	
1852		NaN
1970	Twice the Fun, Double the Trouble.	
1989	The Jungle is JUMPIN'!	
2046	Ride across the sea inside the GIANT PINK SEA ...	
2047	What does he become? What kind of monster?	
2123	Public Enemy No. 1 of all the world...	
...		...
46157	Mica is a perfectly normal boy. Well he would ...	
46158		NaN
46165	Our words are stronger than their weapons.	
46177	There are other worlds than these	
46188	Disney makes the classical musical more magica...	
46204	Deception never looked so sweet	
46227	A musical re-imagining	
46228		NaN
46232		NaN
46236		NaN

46241	Horrifically Funny
46281	NaN
46285	NaN
46287	A poignant look at how a family may look
46308	NaN
46309	NaN
46388	NaN
46401	NaN
46408	NaN
46433	NaN
46437	NaN
46488	Love Makes the World Go 'Round. It Will Spin L...
46511	NaN
46568	NaN
46576	NaN
46577	NaN
46582	When Night Falls, the Adventure Begins!
46616	What happened last night?
46622	NaN
46625	A deadly game of wits.

	title	video	vote_average	vote_count	\
677	The Promise	False	5.0	1.0	
678	The Promise	False	5.0	1.0	
679	The Promise	False	5.0	1.0	
680	The Promise	False	5.0	1.0	
681	The Promise	False	5.0	1.0	
682	The Promise	False	5.0	1.0	
683	The Promise	False	5.0	1.0	
846	Wife	False	0.0	0.0	
847	Wife	False	0.0	0.0	
848	Wife	False	0.0	0.0	
849	Wife	False	0.0	0.0	
850	Wife	False	0.0	0.0	
851	Wife	False	0.0	0.0	
852	Wife	False	0.0	0.0	
902	Sabrina	False	7.4	284.0	
944	The Scarlet Letter	False	8.3	4.0	
964	A Farewell to Arms	False	6.2	29.0	
965	A Farewell to Arms	False	6.2	29.0	
966	A Farewell to Arms	False	6.2	29.0	
967	A Farewell to Arms	False	6.2	29.0	
968	A Farewell to Arms	False	6.2	29.0	
969	A Farewell to Arms	False	6.2	29.0	
970	A Farewell to Arms	False	6.2	29.0	
1317	Cape Fear	False	7.2	125.0	
1852	Hamlet	False	7.1	48.0	
1970	The Parent Trap	False	6.8	768.0	

1989	The Jungle Book	False	7.1	1791.0
2046	Doctor Dolittle	False	6.0	42.0
2047	The Nutty Professor	False	6.6	90.0
2123	The Man Who Knew Too Much	False	6.4	73.0
...	...	...	...	...
46157	Treading Water	False	5.6	6.0
46158	Jesus	False	5.4	8.0
46165	City of Ghosts	False	2.0	3.0
46177	The Dark Tower	False	5.7	688.0
46188	Annie	False	6.1	54.0
46204	Snow White: The Fairest of Them All	False	5.3	52.0
46227	Dirty Dancing	False	5.3	58.0
46228	Bird on a Wire	False	6.0	2.0
46232	Elegy	False	0.0	0.0
46236	A Long and Happy Life	False	6.0	2.0
46241	Abduction	False	0.0	0.0
46281	Get Out	False	4.5	4.0
46285	The Black Hole	False	7.2	23.0
46287	Normal	False	5.2	3.0
46308	The Saint	False	5.8	41.0
46309	Ordinary Wonder	False	0.0	0.0
46388	Feed	False	7.3	50.0
46401	Sahara	False	5.4	43.0
46408	Geronimo	False	7.7	9.0
46433	Final Justice	False	4.4	4.0
46437	Terminus	False	5.0	1.0
46488	Love	False	7.6	5.0
46511	Joanna	False	7.0	2.0
46568	Cinderella	False	4.6	6.0
46576	Dead Birds	False	5.3	2.0
46577	The Hunters	False	5.0	1.0
46582	Arabian Nights	False	6.9	21.0
46616	The Morning After	False	4.0	2.0
46622	Robin Hood	False	5.7	26.0
46625	Betrayal	False	3.8	6.0

	cast \
677	[corinnaharfouch, meretbecker, augustzirner]
678	[corinnaharfouch, meretbecker, augustzirner]
679	[corinnaharfouch, meretbecker, augustzirner]
680	[corinnaharfouch, meretbecker, augustzirner]
681	[corinnaharfouch, meretbecker, augustzirner]
682	[corinnaharfouch, meretbecker, augustzirner]
683	[corinnaharfouch, meretbecker, augustzirner]
846	[miekotakamine, kenuehara, rentarômikuni]
847	[miekotakamine, kenuehara, rentarômikuni]
848	[miekotakamine, kenuehara, rentarômikuni]
849	[miekotakamine, kenuehara, rentarômikuni]

850 [miekotakamine, kenuehara, rentarômikuni]  
 851 [miekotakamine, kenuehara, rentarômikuni]  
 852 [miekotakamine, kenuehara, rentarômikuni]  
 902 [humphreybogart, audreyhepburn, williamholden]  
 944 [lilliangish, larshanson, henryb.walthall]  
 964 [helenhayes, garycooper, adolphemenjou]  
 965 [helenhayes, garycooper, adolphemenjou]  
 966 [helenhayes, garycooper, adolphemenjou]  
 967 [helenhayes, garycooper, adolphemenjou]  
 968 [helenhayes, garycooper, adolphemenjou]  
 969 [helenhayes, garycooper, adolphemenjou]  
 970 [helenhayes, garycooper, adolphemenjou]  
 1317 [gregorypeck, robertmitchum, pollybergen]  
 1852 [laurenceolivier, jeansimmons, johnlaurie]  
 1970 [lindsaylohan, dennisquaid, natasharichardson]  
 1989 [brucereitherman, philharris, sebastiancabot]  
 2046 [rexharrison, samanthaeggar, richardattenborough]  
 2047 [jerrylewis, stellastevens, henrygibson]  
 2123 [lesliebanks, peterlorre, pierrefresnay]  
 ... ..  
 46157 [douglassmith, zoëkravitz, ariadnagil]  
 46158 [jeremysisto, arminmueller-stahl, debramessing]  
 46165 []  
 46177 [idriselba, matthewmconaughey, tomtaylor]  
 46188 [kathybates, victorgarber, alancumming]  
 46204 [mirandarichardson, kristinkreuk, verafarmiga]  
 46227 [abigailbreslin, coltprattes, sarahhyland]  
 46228 [leonardcohen]  
 46232 []  
 46236 [innagulaya, kirilllavrov]  
 46241 [karolinaantosik, amelieleroy, tessamcgin]  
 46281 [jacquesbriere, philippepeythieu]  
 46285 [napoleonryan]  
 46287 [milankako, pavelgajdos, dagmarvekrnová-hav...]  
 46308 [adamrayner, elizadushku, ianogilvy]  
 46309 [alekseikonsovsky, ninazorskaya, olegvidov]  
 46388 [troianbellisario, tomfelton, benwinchell]  
 46401 [omarsy, louaneemera, franckgastambide]  
 46408 [célinesallete, rachidyous, davidmurgia]  
 46433 [stephenchow, jamesha, kenlo]  
 46437 [johnnyhallyday, karenallen, jürgenprochnow]  
 46488 [gretagarbo, johngilbert, brandonhurst]  
 46511 [urszulagrabowska, saraknothe, stanislawacelin...]  
 46568 [kristinaasmus, nikitaefremov, artyomtkachenko]  
 46576 [robertgardner]  
 46577 [johnmarshall]  
 46582 [miliavital, dougraycott, jasonscottlee]  
 46616 [markieadams, robertoaguire, tinaarning]

46622 [patrickbergin, umathurman, davidmorrissey]  
 46625 [erikaeleniak, adambaldwin, juliedupage]

```

                                crew \
677  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
678  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
679  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
680  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
681  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
682  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
683  [{'credit_id': '52fe4a44c3a36847f81c463f', 'de...
846  [{'credit_id': '52fe4b9ac3a368484e190d25', 'de...
847  [{'credit_id': '56365ed9925141285701b06e', 'de...
848  [{'credit_id': '56365ed9925141285701b06e', 'de...
849  [{'credit_id': '52fe4b9ac3a368484e190d25', 'de...
850  [{'credit_id': '52fe4b9ac3a368484e190d25', 'de...
851  [{'credit_id': '56365ed9925141285701b06e', 'de...
852  [{'credit_id': '56365ed9925141285701b06e', 'de...
902  [{'credit_id': '52fe445fc3a36847f8091fb3', 'de...
944  [{'credit_id': '52fe49509251416c910a9689', 'de...
964  [{'credit_id': '56d73d3e9251414291002436', 'de...
965  [{'credit_id': '56d73d3e9251414291002436', 'de...
966  [{'credit_id': '56d73d3e9251414291002436', 'de...
967  [{'credit_id': '56d73d3e9251414291002436', 'de...
968  [{'credit_id': '56d73d3e9251414291002436', 'de...
969  [{'credit_id': '56d73d3e9251414291002436', 'de...
970  [{'credit_id': '56d73d3e9251414291002436', 'de...
1317 [{'credit_id': '52fe442b9251416c7502b3f9', 'de...
1852 [{'credit_id': '52fe4467c3a368484e0212a3', 'de...
1970 [{'credit_id': '594852a9c3a3681115033422', 'de...
1989 [{'credit_id': '52fe44e8c3a36847f80b10ab', 'de...
2046 [{'credit_id': '52fe469f9251416c7507e2cb', 'de...
2047 [{'credit_id': '52fe47749251416c75099f1f', 'de...
2123 [{'credit_id': '52fe4496c3a36847f809e889', 'de...
...
46157 [{'credit_id': '52fe4ca0c3a36847f8237687', 'de...
46158 [{'credit_id': '52fe434fc3a36847f804a813', 'de...
46165 [{'credit_id': '586ba10ac3a3685e9e00574f', 'de...
46177 [{'credit_id': '5912cf71c3a36864d40533b7', 'de...
46188 [{'credit_id': '52fe44f2c3a368484e03fd71', 'de...
46204 [{'credit_id': '52fe44dc9251416c9101f469', 'de...
46227 [{'credit_id': '58b9dd05c3a368666b00d387', 'de...
46228 [{'credit_id': '52fe48b6c3a36847f817577b', 'de...
46232 [{'credit_id': '5759ef93c3a3684e30002c43', 'de...
46236 [{'credit_id': '52fe4b809251416c75104eb5', 'de...
46241 [{'credit_id': '595f97249251412251083338', 'de...
46281 [{'credit_id': '52fe4b60c3a368484e18617b', 'de...
46285 [{'credit_id': '52fe4a56c3a36847f81c81a5', 'de...

```

46287 [{'credit\_id': '52fe4a2b9251416c750e00af', 'de...  
 46308 [{'credit\_id': '59551d549251411d98030919', 'de...  
 46309 [{'credit\_id': '52fe46aac3a368484e09d405', 'de...  
 46388 [{'credit\_id': '592ef910c3a368614301005c', 'de...  
 46401 [{'credit\_id': '5815b1bc9251415c5e00a502', 'de...  
 46408 [{'credit\_id': '5732751d925141656b000445', 'de...  
 46433 [{'credit\_id': '52fe47d8c3a368484e0dc9f9', 'de...  
 46437 [{'credit\_id': '52fe4930c3a368484e11edcd', 'de...  
 46488 [{'credit\_id': '52fe4a51c3a368484e14c973', 'de...  
 46511 [{'credit\_id': '52fe46c4c3a368484e0a2b1d', 'de...  
 46568 [{'credit\_id': '52fe4a1c9251416c750de11b', 'de...  
 46576 [{'credit\_id': '5704afc4c3a3683ef8000ce0', 'de...  
 46577 [{'credit\_id': '53973532c3a3684f45003d41', 'de...  
 46582 [{'credit\_id': '5545faee925141499f00288a', 'de...  
 46616 [{'credit\_id': '587626f4c3a3682b33008299', 'de...  
 46622 [{'credit\_id': '52fe44439251416c9100a899', 'de...  
 46625 [{'credit\_id': '52fe4776c3a368484e0c8387', 'de...

	keywords \
677	[anti-communism, political, womandirector]
678	[anti-communism, political, womandirector]
679	[anti-communism, political, womandirector]
680	[anti-communism, political, womandirector]
681	[anti-communism, political, womandirector]
682	[anti-communism, political, womandirector]
683	[anti-communism, political, womandirector]
846	[basedonnovel, wife, marriage]
847	[basedonnovel, wife, marriage]
848	[basedonnovel, wife, marriage]
849	[basedonnovel, wife, marriage]
850	[basedonnovel, wife, marriage]
851	[basedonnovel, wife, marriage]
852	[basedonnovel, wife, marriage]
902	[brotherbrotherrelationship, chauffeur, champa...]
944	[silentfilm]
964	[italy, nurse, loveletter]
965	[italy, nurse, loveletter]
966	[italy, nurse, loveletter]
967	[italy, nurse, loveletter]
968	[italy, nurse, loveletter]
969	[italy, nurse, loveletter]
970	[italy, nurse, loveletter]
1317	[poison, smalltown, boat]
1852	[shakespeare, fathermurder]
1970	[matchmaking, divorced, twinsseparatedatbirth]
1989	[indianlead, narration, humananimalrelationship]
2046	[england, snail, vegetarian]
2047	[professor, chemistry, love]



2123 [assassination, attemptedkidnapping]  
 ...  
 46157 [fish, boy, parenting]  
 46158 [christianity, communion, jesuschrist]  
 46165 []  
 46177 [gunslinger, basedonnovel]  
 46188 [holiday, musical, franklindelamorosevelt]  
 46204 [mirror, snowwhite, evilwitch]  
 46227 [musical, remake, catskills]  
 46228 []  
 46232 []  
 46236 []  
 46241 []  
 46281 []  
 46285 []  
 46287 [serialkiller, truecrime]  
 46308 [thesaint]  
 46309 []  
 46388 [twins, valedictorian, eatingdisorder]  
 46401 [snake, prejudice, talkinganimal]  
 46408 [socialworker, musical]  
 46433 []  
 46437 []  
 46488 [militaryofficer, extramaritalaffair, silentfilm]  
 46511 []  
 46568 []  
 46576 [ethologist, papuanewguinea]  
 46577 [africa, desert, hunting]  
 46582 []  
 46616 []  
 46622 []  
 46625 []

search \

677 anti-communism political womandirector corinna...  
 678 anti-communism political womandirector corinna...  
 679 anti-communism political womandirector corinna...  
 680 anti-communism political womandirector corinna...  
 681 anti-communism political womandirector corinna...  
 682 anti-communism political womandirector corinna...  
 683 anti-communism political womandirector corinna...  
 846 basedonnovel wife marriage miekotakamine kenue...  
 847 basedonnovel wife marriage miekotakamine kenue...  
 848 basedonnovel wife marriage miekotakamine kenue...  
 849 basedonnovel wife marriage miekotakamine kenue...  
 850 basedonnovel wife marriage miekotakamine kenue...  
 851 basedonnovel wife marriage miekotakamine kenue...  
 852 basedonnovel wife marriage miekotakamine kenue...

902 brotherbrotherrelationship chauffeur champagne...  
 944 silentfilm lilliangish larshanson henryb.walth...  
 964 italy nurse loveletter helenhayes garycooper a...  
 965 italy nurse loveletter helenhayes garycooper a...  
 966 italy nurse loveletter helenhayes garycooper a...  
 967 italy nurse loveletter helenhayes garycooper a...  
 968 italy nurse loveletter helenhayes garycooper a...  
 969 italy nurse loveletter helenhayes garycooper a...  
 970 italy nurse loveletter helenhayes garycooper a...  
 1317 poison smalltown boat gregorypeck robertmitchu...  
 1852 shakespeare fathermurder laurenceolivier jeans...  
 1970 matchmaking divorced twinsseparatedatbirth lin...  
 1989 indianlead narration humananimalrelationship b...  
 2046 england snail vegetarian rexxharrison samanthae...  
 2047 professor chemistry love jerrylewis stellastev...  
 2123 assassination attemptedkidnapping lesliebanks ...  
 ...  
 46157 fish boy parenting douglassmith zoëkravitz ari...  
 46158 christianity communion jesuschrist jeremysisto...  
 46165 documentary  
 46177 gunslinger basedonnovel idriselba matthewmcon...  
 46188 holiday musical franklindelanoroosevelt kathyb...  
 46204 mirror snowwhite evilwitch mirandarichardson k...  
 46227 musical remake catskills abigailbreslin coltpr...  
 46228 leonardcohen documentary  
 46232 sciencefiction  
 46236 innagulaya kirilllavrov romance drama  
 46241 karolinaantosik amelieleroy tessamcginn fanta...  
 46281 jacquesbriere philippepeythieu comedy animation  
 46285 napoleonryan comedy crime  
 46287 serialkiller truecrime milankako pavelgajdos...  
 46308 thesaint adamrayner elizadushku ianogilvy acti...  
 46309 alekseikonsovsky ninazorskaya olegvidov drama...  
 46388 twins valedictorian eatingdisorder troianbelli...  
 46401 snake prejudice talkinganimal omarsy louaneeme...  
 46408 socialworker musical célinesallete rachidyous...  
 46433 stephenchow jamesha kenlo action comedy crime  
 46437 johnnyhallyday karenallen jürgenprochnow  
 46488 militaryofficer extramaritalaffair silentfilm ...  
 46511 urszulagrabowska saraknothe stanislawacelinsk...  
 46568 kristinaasmus nikitaefremov artyomtkachenko c...  
 46576 ethologist papuanewguinea robertgardner docume...  
 46577 africa desert hunting johnmarshall documentary  
 46582 miliavital dougraycott jasonscottlee drama f...  
 46616 markieadams robertoaguires tinaarning comedy d...  
 46622 patrickbergin umathurman davidmorrissey drama...  
 46625 erikaeleniak adambaldwin juliedupage action d...

	newTitle
677	dasversprechen
678	dasversprechen
679	dasversprechen
680	dasversprechen
681	dasversprechen
682	dasversprechen
683	dasversprechen
846	tsuma
847	tsuma
848	tsuma
849	tsuma
850	tsuma
851	tsuma
852	tsuma
902	sabrina
944	thescarletletter
964	afarewelltoarms
965	afarewelltoarms
966	afarewelltoarms
967	afarewelltoarms
968	afarewelltoarms
969	afarewelltoarms
970	afarewelltoarms
1317	capefear
1852	hamlet
1970	theparenttrap
1989	thejunglebook
2046	doctordolittle
2047	thenuttyprofessor
2123	themanwhoknewtoomuch
...	...
46157	treadingwater
46158	jesus
46165	cityofghosts
46177	thedarktower
46188	annie
46204	snowwhite
46227	dirtydancing
46228	birdonawire
46232	elegy
46236	
46241	abduction
46281	getout
46285	theblackhole
46287	normal
46308	thesaint
46309	

```

46388          feed
46401          sahara
46408          geronimo
46433      finaljustice
46437          terminus
46488          love
46511          joanna
46568          zolushka
46576          deadbirds
46577          thehunters
46582          arabiannights
46616      themorningafter
46622          robinhood
46625          betrayal

```

```
[3258 rows x 29 columns]
```

```
In [47]: allFeatures.shape
```

```
Out[47]: (46628, 29)
```

```
In [48]: allFeatures.sort_values(by="release_date", ascending= False, inplace= True)
```

```
In [49]: allFeatures.drop_duplicates(subset= "original_title", keep="last", inplace= True)
```

```
In [50]: allFeatures.reset_index(drop=True, inplace=True)
```

```
In [65]: indices = pd.Series(allFeatures.index, index=allFeatures['newTitle'])
```

```
In [66]: indices.shape
```

```
Out[66]: (43370,)
```

```
In [67]: indices
```

```

Out[67]: newTitle
avatar2                                0
theothersideofthewind                  1
badboysforlife                         2
maryshelley                           3
mobilehomes                           4
ironsky:thecomingrace                  5
slycooper                             6
theking'sdaughter                      7
howtotalktogirlsatparties              8
pitchperfect3                         9
machines                             10
sweetvirginia                         11
justiceleague                         12

```

78/52	13
callmebyyourname	14
thor:ragnarok	15
dina	16
beyondskyline	17
resurrectinghassan	18
porto	19
leatherface	20
god'sowncountry	21
thetriptospain	22
sciencefictionvolumeone:theosirischild	23
legodcsuperherogirls:braindrain	24
patticake\$	25
whathappenedtomonday	26
ingridgoeswest	27
goodtime	28
theglasscastle	29
...	
thelastgold	43340
scotthall:livingonarazor'sedge	43341
baddadrehab	43342
elpejesapo	43343
losounddesert	43344
padyatra:agreenodyssey	43345
theadventuresofcinderella'sdaughter	43346
alwaysfaithful	43347
allendeensulaberinto	43348
dolpotulku-heimkehrindenhimalaya	43349
patientzero	43350
winningfavour	43351
whnthedayhadnoname	43352
blindpassasjer	43353
half-life	43354
jedijuniorhigh	43355
cosmos	43356
irwin&fran2013	43357
drzacetraby	43358
supermassiveblackholes	43359
43360	
neitherwolfnordog	43361
saladfingers	43362
mundocão	43363
thegardenofafflictions2017	43364
konstruktorkrasnogotsveta-1993	43365
allsuperheroesmustdie2:thelastsuperhero	43366
thelandwherethebluesbegan	43367
aprel	43368

43369

Length: 43370, dtype: int64

```
In [54]: from sklearn.feature_extraction.text import CountVectorizer
         from sklearn.metrics.pairwise import cosine_similarity
```

```
In [68]: def showCos(title):
         i= indices[title]
         vocab= (allFeatures['search'][i]).split(" ")
         countVector = CountVectorizer(stop_words='english', vocabulary= vocab)
         countData = countVector.fit_transform(allFeatures['search'])
         cosScore = cosine_similarity(countData[i], countData)
         return cosScore
```

```
In [69]: allFeatures['search'][0].split(" ")
```

```
Out[69]: ['sequel',
          'alienplanet',
          'avatar',
          'katewinslet',
          'zoesaldana',
          'samworthington',
          'action',
          'adventure',
          'fantasy']
```

```
In [57]: c= showCos('Toy Story')
```

```
In [58]: c
```

```
Out[58]: array([[0.          , 0.33333333, 0.          , ..., 0.          , 0.          ,
                  0.33333333]])
```

```
In [71]: def similarTitle(word):
         countVector = CountVectorizer(stop_words='english')
         c = countVector.fit(allFeatures['original_title'])
         cosScore = cosine_similarity(c.transform([word]), c.transform(allFeatures['original_title']))
         return cosScore
```

```
In [75]: def getTitleRecs():
         try:
             title= str(input('Enter a movie title: '))
             # normalize input
             normTitle= normalizeTitle(title)
             scores= showCos(normTitle)
             # convert into tuples of movie id and score
             scoresLst= list(enumerate(scores[0]))
             # sort the tuples by the score
             scoresLst.sort(key= lambda x: x[1], reverse= True)
             # get top 10 movie recs
```

```

        for z, score in scoresLst[1:11]:
            print(allFeatures['original_title'][z])
    except:
        print('Movie not found. Did you mean this:')
        getSpellingSuggestion(title)
        getTitleRecs()

In [72]: def getSpellingSuggestion(word):
        simTitle= similarTitle(word)
        sim= list(enumerate(simTitle[0]))
        sim.sort(key= lambda x: x[1], reverse= True)
        for j, val in sim[0:3]:
            print(allFeatures['original_title'][j])

```

```

In [76]: getTitleRecs()

```

```

Enter a movie title: harry potter
Movie not found. Did you mean this:
Harry Potter and the Goblet of Fire
Harry Potter and the Deathly Hallows: Part 2
Harry Potter and the Deathly Hallows: Part 1
Enter a movie title: harry potter and the goblet of fire
Harry Potter and the Half-Blood Prince
Harry Potter and the Deathly Hallows: Part 2
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Chamber of Secrets
Harry Potter and the Philosopher's Stone
Harry Potter and the Deathly Hallows: Part 1
Harry Potter and the Order of the Phoenix
Fantastic Beasts and Where to Find Them
Oz: The Great and Powerful
The Chronicles of Narnia: The Voyage of the Dawn Treader

```